Università degli Studi di Milano

Decision methods and models

Roberto Cordone

– Academic year 2018/19 –

Contents

Foreword

The label *Decision methods and models* is applied in Italian universities to courses with a rather various content. Quite often, it is another name for Operations Research courses, dealing with optimization models and algorithms, or for collections of practical applications of Operations Research. This course belongs to the same general environment, but with a fundamental difference.

The common theme of these courses is the concept of *decision*, that is a choice among nonindifferent alternatives. It is a situation we all face everyday, but that in some cases is particularly difficult because the relevant data to take into account are large, the number of possible alternatives is large and the cost of a wrong choice are large. In these cases, the correct strategy is not to resort to common sense, to past experience or to chance, but to model the problem, compute its solution, interpret it and, finally, decide. This has been known for decades (at least since World War II). Nowadays, however, new tools have put again into fashion the modellistic approach to decision: *Big Data* provides huge quantities of precise, structured and inexpensive data, which are ready to be exploited for extracting information; *Cloud Computing* provides everywhere access to low-cost data and computing power; *Business Analytics* supports the competitive advantage of a business culture open to the use of models in order to extract useful information on one's own company, customers, competitors and business environment, in order to define one's operational strategies. This has stimulated in the academy the development of new theoretical fields, such as online optimization, stochastic optimization, robust optimization, multilevel optimization, etc...

The fundamental difference between this course and most of the other ones with the same name is that this course does not deal with the technical aspect of solving decision models, that is of their mathematical properties and solving algorithms. By contrast, it deals with:

- the factors that make the design of such models complicated on principle;
- the mathematical models proposed to treat such complications and the resulting mathematical models;
- the intrinsic limitations and errors of such methods and models, including some impossibility results about the design of satisfactory solution approaches.

This course, in other words, means to show why it is advantageous to use models as a support to decisions and why the results produced by such models must be used with care. The focus moves from the core of the decision problem to everything that surrounds it, and that in many cases is unavoidable to consider in order to reach the core and to interpret in a meaningful way the results of the algorithms applied to it.

This course has two original sources. The former is the course in *Metodi e modelli per il supporto alle decisioni* which was given in the site of Como of the Politecnico di Milano, for students in environmental engineering and computer engineering and for which years ago I provided some supporting activities. This source provides most of the subjects and the choice of the case studies. These concern decisions about public works, which are probably the situations in which it is clearest how hard it is to turn a practical problem into a standard optimization problem and to turn its numerical results into an operational strategy. On the other side, these cases do not exhaust the application fields in which decision model can provide a useful guide to action. Other examples can be found in:

- finance, where models are used to define investment strategies;
- marketing, where models are used to define advertising campaigns or pricing strategies;
- natural resource management, where models are used to define exploitation strategies;
- videogames, where models are used to define game strategies.

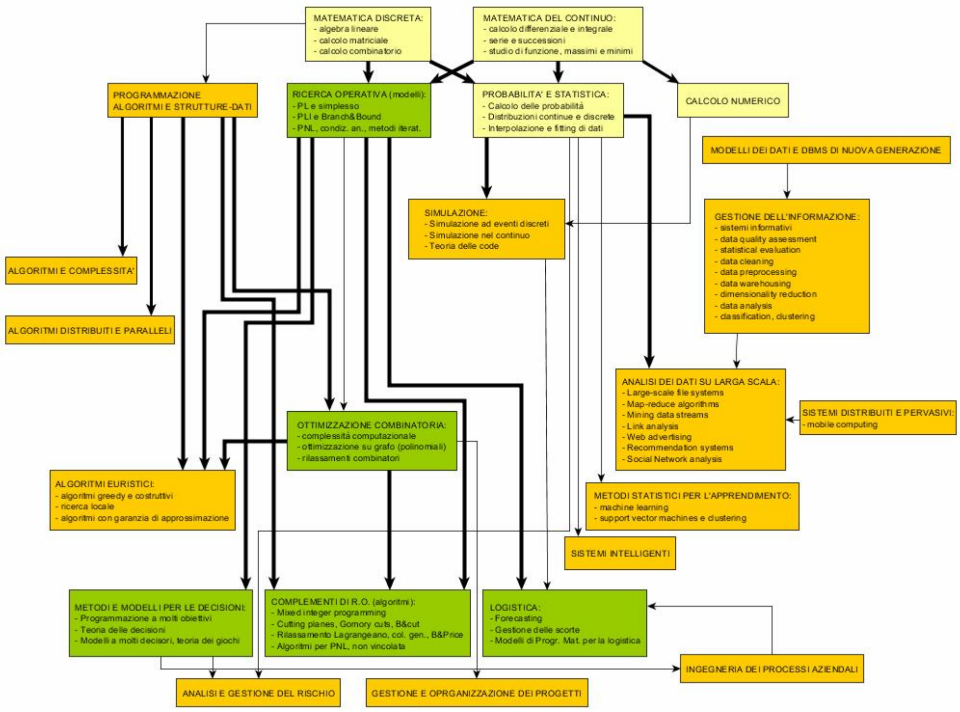

MATEMATICA DISCRETA:
- algebra lineare
- calcolo matriciale
- calcolo combinatorio
MATEMATICA DEL CONTINUO:
- calcolo differenziale e integrale
- serie e successioni
- studio di funzione, massimi e minimi
PROGRAMMAZIONE
ALGORITMI E STRUTTURE-DATI
RICERCA OPERATIVA (modelli):
- PL e simplesso
- PLI e Branch&Bound
- PNL, condiz. an., metodi iterat.
PROBABILITA' E STATISTICA:
- Calcolo delle probabilità
- Distribuzioni continue e discrete
- Interpolazione e fitting di dati
CALCOLO NUMERICO
MODELLI DEI DATI E DBMS DI NUOVA GENERAZIONE
SIMULAZIONE:
- Simulazione ad eventi discreti
- Simulazione nel continuo
- Teoria delle code
GESTIONE DELL'INFORMAZIONE:
- sistemi informativi
- data quality assessment
- statistical evaluation
- data cleaning
- data preprocessing
- data warehousing
- dimensionality reduction
- data analysis
- classification, clustering
ALGORITMI E COMPLESSITA'
ALGORITMI DISTRIBUITI E PARALLELI
ANALISI DEI DATI SU LARGA SCALA:
- Large-scale file systems
- Map-reduce algorithms
- Mining data streams
- Link analysis
- Web advertising
- Recommendation systems
- Social Network analysis
SISTEMI DISTRIBUITI E PERVASIVI:
- mobile computing
OTTIMIZZAZIONE COMBINATORIA:
- complessità computazionale
- ottimizzazione su grafo (polinomiali)
- rilassamenti combinatori
ALGORITMI EURISTICI:
- algoritmi greedy e costruttivi
- ricerca locale
- algoritmi con garanzia di approssimazione
METODI STATISTICI PER L'APPRENDIMENTO:
- machine learning
- support vector machines e clustering
SISTEMI INTELLIGENTI
METODI E MODELLI PER LE DECISIONI:
- Programmazione a molti obiettivi
- Teoria delle decisioni
- Modelli a molti decisori, teoria dei giochi
COMPLEMENTI DI R.O. (algoritmi):
- Mixed integer programming
- Cutting planes, Gomory cuts, B&cut
- Rilassamento Lagrangeano, col. gen., B&Price
- Algoritmi per PNL, non vincolata
LOGISTICA:
- Forecasting
- Gestione delle scorte
- Modelli di Progr. Mat. per la logistica
INGEGNERIA DEI PROCESSI AZIENDALI
ANALISI E GESTIONE DEL RISCHIO
GESTIONE E OPRGANIZZAZIONE DEI PROGETTI

Figure 1: Structure of the track on *Analytics and optimization*

The second source of the course is the track on *Analytics and optimization* that the degrees in Computer Science and Mathematics of the University of Milan have been offering in recent years, in which this course plays a well determined role. This source provides some subjects, but especially a more rigorous and formal approach, as well as a strong complementarity relation with the other courses of the same track. I start by saying that the course is designed so that is can be attended and it can be interesting and useful also for students of other tracks and other degrees, but the following paragraph will focus on the track of *Analytics and optimization* to better clarify the basic purpose and approach of this course. Figure 1 shows the structure of the track: the courses in yellow boxes deal with basic subjects, those in orange boxes with affine subjects from Computer Science[1], those in green boxes (plus *Algoritmi euristici*, that is *Heuristic algorithms*) are the fundamental courses of the track. In short:

- *Operations Research* (*Ricerca operativa*) develops the concept of decision model and the use of general-purpose solvers;
- *Combinatorial Optimization* (*Ottimizzazione combinatoria*) focuses on the problems for which polynomial algorithms have been devised and discusses the best known algorithms and data structures for a collection of fundamental problems;
- *Complements of Operations Research* (*Complementi di ricerca operativa*) focuses on the problems for which no polynomial algorithm is known and discusses the design of exponential algorithms which in the average case can solve medium-size instances of practical use;
- *Heuristic algorithms* (*Algoritmi euristici*) focuses on the problems for which polynomial algorithms have been devised and discusses the design of polynomial algorithms which do not guarantee to find the optimal solution, but on average compute good quality solutions;
- *Logistics* (*Logistica*) presents a collection of models and algorithms for a specific application field, that is the prodution, storage and transportation of goods;
- *Decisions methods and models* (*Metodi e modelli per le decisioni*) surrounds them all, discussing how to build the decision models presented in the other courses and to what extent such models correctly represent the real world.

From this scheme, it should be clear that the course in *Decision methods and models* does not require strong skills on the properties of optimization models and algorithms. The student who haven't got those skills will simply accept the existence of such models and algorithms as black boxes. The students who know more, for example because they have attended the other courses of the track, will use this course to enlarge their point of view to what precedes and follows the algorithmic solution phase. These notes include several sections marked by a star which should be considered as additional material, not required to pass the exam: those sections provide connections to the other courses of the track.

As a further proof of the interdisciplinarity of the course and of the possibility to attend it starting from very different backgrounds, besides being open to students of Computer Science and Mathematics, since 2018 the course has been included in the degree of *Data Science and Economics*, as the module on *Optimization* in the course on *Graph Theory, Discrete Mathematics and Optimization.* This necessarily makes the approach to the subject much wider than deeper and implies that every group of students from a different background will find subjects that are already known or that are apparently explained in a simplistic way. The students in Mathematics will feel that the analytic methods are too informal, the students in Computer Science will feel that data management and computational complexity are only hinted at, the students in Data Science and Economics will feel that the economic concepts are too superficial. This is unfortunately very hard to avoid, if all students must have the possibility to attend the course and pass the exam in the standard times. On the other hand, it is likely that all students will find interesting the occasion to

[1] Except for *Algoritmi euristici* which is in an orange box for beaurocratic reasons.

enlarge their perspective through the combined presentation of the different aspects that must necessarily interact in a complicated real decision.

Part I

Decision problems

Chapter 1

Introduction

1.1 Terminology

Informally speaking, a decision is a choice among alternatives that are not reciprocally indifferent. Let us introduce some terms that will be used in the following in order to be more precise and nonambiguous. It is not yet a formal definition, but just the introduction of some keywords: Section 1.6 and Chapter 2 will provide practical examples of the usage of such terms, while Chapter 3 will provide a formal definition and discuss several methodological problems related to their usage. Unfortunately, the subjects of this course are studied in many different research fields (mathematics, economy, engineering, etc. . .) and each field has adopted its own terminology and conventions, usually different from those of the other fields. I have decided to report all terms of common use and to adopt in each chapter those typical of the field that has studied in more depth the specific subject of the chapter itself, recalling from time to time the equivalence with other terms. This violated the classical rule to use a single word for each concept, but it should allow the students to access more easily the literature. Some concepts, moreover, have no standard name to the best of my knowledge, so that I have been forced to invent a term *ad hoc* for this course.

We will define *system* the portion of the world which the decision will affect[1]. A system is not given once for all, otherwise no decision would be possible. We will denote as *alternative* or *solution* the combination of all controllable aspects of the system[2], as *scenario* or *outcome* the combination of all its uncontrollable aspects[3]. A system, therefore, combines controllable and uncontrollable aspects into a *configuration*[4]. A configuration is associated to an *impact*, that describes all aspects which are relevant for the decision[5]. *Decision-maker* or *stakeholder* is everybody who contributes to the choice of the alternative: the first term indicates who takes part to the choice, while the second term also includes who does not participate, but has interests at stake and could react to a disagreeable choice, thus exerting an indirect influence on the choice. Finally, by *preference* we denote the description of the relative satisfaction between different impacts[6].

A *decision problem* requires to *choose an alternative so as to move the system*

[1] *System* is typical term of the engineering or mathematical jargon.

[2] *Alternative* is more used in the economical jargon, *solution* in the mathematical and engineering one.

[3] *Scenario* is more used in the economical jargon, *outcome* in the statistical one.

[4] *Configuration* is an *ad hoc* term, though it is sometimes used in engineering.

[5] *Impact* is another *ad hoc* term; the literature often adopts circumlocutions.

[6] *Preference* derives from the economic jargon, but it is commonly used in everyday language.

into a configuration whose impact is preferred by the decision-makers to those of the other configurations, keeping into account that the actual configuration does not depend only on the chosen alternative, but also on the possible scenarios. Two fundamental elements for the existence of a decision problem are:

- freedom, that is the existence of several available alternatives;
- rationality, that is the existence of preferences between different impacts.

The former grants the possibility of a choice, the latter a criterium of choice.

Some confusion can derive from the fact that in theoretical Computer Science a *decision problem* is a problem which admits only two possible solutions (*yes* and *no*, or *true* and *false*). This is a completely different concept. Looking for a relation with theoretical Computer Science problems, the decision problems treated in these notes include somehow as special cases the *optimization/search problems*, whose solution is an object with the maximum value (or minimum cost) among all objects in a suitale family.

1.2 The modelling approach to decision

Figure 1.1 gives a scheme of the modelling approach to decision. Instead of directly moving from the practical problem to the operational strategy, this approach requires a series of intermediate passages:

- building a *model* of the problem, with suitable methods;
- solving the model with *algorithms*, that is formal methods;
- *interpreting* the solution, with suitable methods.

Methods occur in every phase of the process. Sometimes they are formal methods, that is algorithms. This especially concerns the model resolution, that is the passage from the model to its abstract solution. The correctness of this passage could even be guaranteed by a mathematical proof, if the algorithm is an exact one. If it is heuristic, the guarantee is only empirical and experimental. Even more informal are the passage from the problem to the model, that is from the real world to the world of mathematics, and the passage from the solution to the operational strategy, that is from the world of mathematics to the real world. Such passages are particularly tricky because they are intrinsically creative and intuitive, and therefore prone to many sources of potential errors. A common risk is that the decision-maker might mistake the model for the reality and blindly trust its results, which could be formally correct, but impossible to apply. Another, subtler, risk, is that the decision-makers might distort the numerical results of the algorithms by applying them directly to build an operational strategy (for instance, they might adopt an "optimal" solution very sensitive to data variations in a situation with slightly different data, thus obtaining very poor results in practice). In this case, both the model and the algorithm are correct, but the interpretation of the results is wrong.

For this reason, the decision process is not a one-pass process, but works by subsequent corrections. Figure 1.2 shows the process sketched by H. A. Simon[7] in his fundamental work "Models of bounded rationality" (1982). The scheme presents the following phases, in which it is possible to recognize rather easily the concepts introduced above:

[7] H. A. Simon () was one of the founders of Artificial Intelligence in the Fifties and won the Nobel Prize for Economy in 1978 for his studies on bounded rationality.

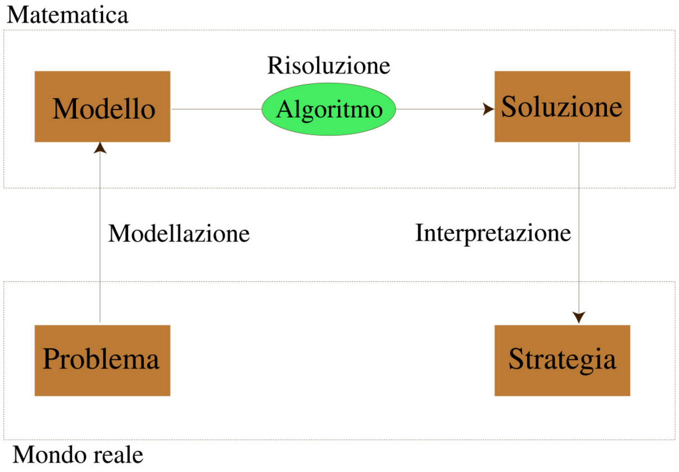


Figure 1.1: Scheme of the modellistic approach to decision

1. *problema formulation*: limit the system, identifying impacts and preferences on one hand (*Objectives*), decision-makers and scenarios on the other hand (*Context*);

2. *identification of the alternatives*: define the set of the feasible alternatives;

3. *evaluation of the alternatives and choice*:

   - evaluate the impact associated to each configuration (that is, alternative and scenario);
   - choose an alternative based on the preferences of the decision-makers;

4. *implementation of the decision*:
   turn into practice the chosen alternative or make a simulation of its application;

5. *monitoring and verification*:

   - observe the consequences of the decision;
   - if unsatisfactory, make corrections and repeate the process introducing new scenarios, new objectives, new alternatives, new evaluation methods.

In all these phases, specific experience and the support of information technology are usually essential.

1.3 Why a formal approach?

It is appropriate to discuss the advantages of formalizing a decision problem. The formal approach is profitable because it allows to:

1. predict in a more certain and exact way the impact of the decision, using descriptive models, instead of intuition and experience;

2. accelerate the decision using algorithms and information technology, which allow to handle a much larger set of possible alternatives;

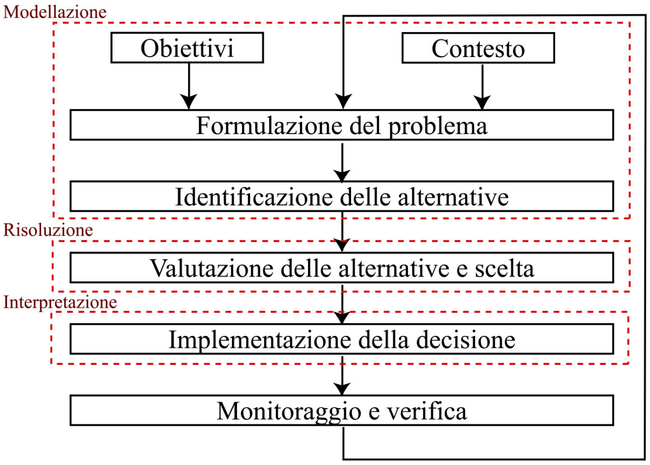

Modellazione
Obiettivi
Contesto
Formulazione del problema
Identificazione delle alternative
Risoluzione
Valutazione delle alternative e scelta
Interpretazione
Implementazione della decisione
Monitoraggio e verifica

Figure 1.2: Scheme of the cyclic decision process according to Simon (1982)

3. assess and certify the decision process:
    - making explicit assumptions on the alternatives, the scenarios, the preferences the decision-makers and their reciprocal relations;
    - guaranteeing the repeatibility of the process, so that it obtains always the same results;
    - make specific modifications to the process without repeating it from scratch every time.

This holds in particular for the public decisions, or decisions in large private organizations: the decision-maker usually has not an absolute power, but must *prove* to other stakeholders that the decision has been taken based on all the available data and on a transparent process. In these situations, moreover, the idea to repeat the whole process from scratch every time that some secondary aspect has changed is perhaps intuitive, but inefficient and unreasonable.

1.4 Prescriptive models and descriptive models

A decision model is usually a composite object. It includes, in fact, a number of submodels which can be classified into two large classes based on the data they require and the results they produce:

1. *prescriptive models* (or *normative models)*
    - use as data the impact for each possible configuration and the preferences of the decision-makers between the impacts;
    - yield as result a "suggested" alternative;

    In short, *if this is the situation, the best which can be done is this.*

2. *descriptive models* (or *predictive models*[8])

[8] The main difference between them is that predictive models concern the future, descriptive models concern the present.

- use as data the description of each configuration (alternative and scenario) of the system;
- yield as result the corresponding impacts.

In short, *if one does this and if that occurs, then the situation will become this.*

The descriptive models can be trivial, but also sophisticated. A descriptive model that is so trivial that it often escapes notice is the model that provides the total cost c_{tot} spent to buy two products given their unitary prices c_1 e c_2 and the acquired quantities x_1 e x_2: the model simply consists of the equation $c_{\text{tot}} = c_1 x_1 + c_2 x_2$. On the other hand, other very sophisticated models allow to compute data which is too expensive or even impossible to measure (for example, future forecasts, etc...). Part VI is fully devoted to the presentation of descriptive models of some relevance for complicated decision problems.

The two classes of models can have subtle interactions. For example, any prescriptive model uses a set of descriptive models to obtain the impacts of the possible alternatives and scenarios. On the other hand, some descriptive models can include prescriptive models. For example, in order to estimate the traffic on a street, it is necessary to predict the choices of drivers in the whole street network, with given physical characteristics, prohibitions and semaforic cycles. These choices result from decisions, which can be describe by prescriptive models. The same occurs for customers' choices depending on the price and selling conditions of a product, for competitors' choices reacting to our commercial strategies, etc...

1.5 Business Intelligence

Business Intelligence has become a trendy buzzword in recent studies on business management. Books and magazine articles glorify it as an essential tool to win competition and be successful. It denotes the *set of all processes and technologies that are used to collect, manage and account for decision-oriented data.*

The architecture of a Business Intelligence system aims to:

- provide data, contents and analysis to the right person in the right moment;
- help the decision-makers to choose a solution, but also to justify their choice, to correct or update it, to interact with each other, etc...;
- conform to the laws and rules on process accounting.

Therefore, it includes technological systems, computer applications and algorithmic or methodologic procedures to perform the following activities:

1. *Data management*: to acquire and manage data;
2. *Transformation tools and processes*: to extract, correct, transmit and store the data (the expression *Extract-Transform-Load* is commonly used);
3. *Repository*: to store data and metadata;
4. *Analytics*: to perform optimizations simulations, estimations, predictions, inferences, pattern matching and data mining;
5. *User interface*: to present and manipulate data and analysis without corrupting the data archives;

6. *Administrative procedures*: to manage security, errors, auditing, privacy.

Each of these points would deserve a detailed discussion, but these notes will focus on the fourth one (*Analytics*), in particular considering the methodological aspect and neglecting the technological one.

Sistemi di questo genere sono ovviamente molto complicati e contengono al loro interno sottosistemi più semplici, con compiti specifici, molti dei quali hanno una lunga storia, che risale a ben prima della moda presente. Per esempio, i sistemi di *Electronic Data Processing* (EDP) da decenni vengono usati per conservare i dati, accedervi, e risolvere semplici problemi standard. Analogamente, i *Management Information Systems* (MIS) consentono elaborazioni statistiche per aggregare dati in un formato più leggibile e immediato. Infine, i *sistemi di supporto alle decisioni* o *Decision Support Systems* (*DSS*) sono l'immediato predecessore dei sistemi di Business Intelligence, dato che combinano l'accesso a modelli descrittivi e prescrittivi, l'uso di sistemi logici per rivelare incoerenze, l'uso di sistemi esperti per elaborare strategie.

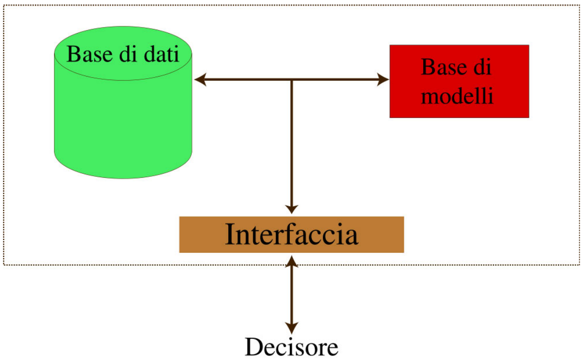


Figure 1.3: General scheme of a Decision Support System

Figure 1.3 shows a general scheme of a *DSS* with its main functional components:

- a *data base* with the corresponding information system that provides the user with the data without introducing biases in their meaning;
- a *model base* with the corresponding management system that uses the data to predict the impacts of every possible decision and to suggest one or more decisions;
- a *user interface* that allows to the decision-maker to rule the process, providing the input data to the models, retrieving their results and possibly turning them into data for other models.

During the whole process, the user is in charge of every step of the system, based on its intermediate results. In particular, the user must:

- define decision-makers and preferences as data;
- define alternatives and scenarios as data;
- control the data flow among the models, choosing which data to use and which models to feed with them:
  - transfer data, hypothetical alternatives and scenarios to the descriptive models, so as to obtain simulated impacts;

  - transfer measured or simulated impacts to the prescriptive models, so as to obtain suggested alternatives;
  - evaluate whether to keep the results of the models as new data;
- use the suggested alternatives as profitable information, and not as instructions to obey.

The fundamental aspect of the process is that the support systems do not create decisions to be applied, but only information to generate decisions.

1.6 Examples of complicated decision problems

This section presents some examples of complicated decision problems, in order to link the concepts introduced above with realistic situations and to suggest what elements make a decision complicated in practice. To the same purpose, Chapter 2 will show in much deeper detail two large-size examples. From these examples, we will obtain a list of the main complicating aspects, which will lead to classify complicated decision problems and to build the index of these notes and of the course.

Example 1 ***(The search for parking)*** *In a rather congested town, we are looking for a parking to leave our car and reach the place of an important meeting; we would prefer to park quickly and not to walk for long to reach our destination.*

- *System is the local street network, with the set of all potential parking places*
- *Alternative is every possible trajectory of the car (path and time)*
- *Scenario is every possible distribution of the free parking places over space and time*
- *Impact are the driving time and the walking time after parking*
- *Decision-maker is the driver (or also the passengers?)*

Notice that the impact reduces to the specific aspects that actually matter for the decisione, but it depends on a complex combination (that is the configuration) of the trajectory followed by the car and the availability of free parking slots in all places and times: the same trajectory, in fact, can lead to park immediately in front of the meeting place if a parking slot is free when the car reaches it or to long useless tours if none is.

Example 2 ***(The thermostat regulation)*** *We want to tune the classroom's thermostat so that the temperature be pleasant for the teacher and the students.*

- *System is the classroom*
- *Alternative is the position of the thermostat knob*
- *Scenario is the external temperature and the exposition of the classroom to the sun*
- *Impact is the internal temperature of the classroom (but probably also its humidity)*
- *Decision-makers are the people dwelling in the classroom (all of them, or just the teacher?)*

Though this problem is easier than the previous ones, notice that the definition of the impact and of the decision-makers is far from trivial.

Example 3 ***(Buying a car)*** *We want to buy a car with a good performance, a high level of comfort, a nice design and a low cost throughout its life-cycle.*

- *System is the local market of cars, petrol, repairs, ecc...*
- *Alternative is the car bought*
- *Scenario are the stock and the prices of the car dealers, the occurrence of accidents, the prices of petrol and car repairs, etc...*
- *Impact are the characteristics of the car throughout its life-cycle*
- *Decision-maker is the buyer (and possibly other family members)*

Notice that a possible alternative, that it is easy to forget, could be to give up buying the car, with an impact that is obviously very bad for some aspects, but very good for other ones.

Example 4 ***(Risiko round)*** *We want to play a round of Risiko, considering in turn all of the players.*

- *System is the map with the distribution of territories, armies, cards*
- *Alternative are the territories from which and to which each attack is launched and the corresponding number of attacking and defending armies*
- *Scenario is the dices' outcome at each attack*
- *Impact is the number of armies destroyed for each player*
- *Decisione-makers are the players*

With respect to other example, in which multiple decision-makers were possible, here the presence of several decision-makers is intrinsic and unavoidable. Moreover, the decision-makers do not act together on the same levers (in the case of the car, they all contribute to choose a single car), but everyone makes his own choice autonomously.

1.7 Classification of decision problems

The previous examples show various factors that can complicate a decision problem. Let us start by stating that here we do not deal with the so called *computational complexity*, that is the existence of perfectly clear and well-defined problems for which not polynomial algorithm is known and it is conjectured that only exponential algorithms exist[9]. In these problems it is expensive to move from the model to the solution, but it is not complicated to define which is the correct solution.

The factor we are interested in concern the decision-makers and their relations with the system. More precisely:

[9] In order to mark the difference from this concept, we will use almost everywhere the expressions "complicated" and "complication" instead of "complex" and "complexity". Unfortunately a few lines below we will talk about "complex preferences", because "complicated preferences" sounds quite bad.

1. *complex preferences*, that is insufficient to define a concept of optimality;
2. *multiple scenarios*, and hence an *uncertain environment*;
3. *multiple decision-makers*, with potentially conflicting preferences.

There is a further factor that can complicate a decision problem: it is an *insufficient formalization* of the system, that is the inability to determine univocally the impact for a given configuration, that is for a fixed alternative and scenario. This situation is often included in the case of uncertain environment by assuming that, besides the known components of the scenario, additional unknown components exist (often denoted as *perturbations*), whose knowledge would allow to compute exactly the impact. The resulting models are denoted as *black-box models*, or *grey-box models*, based on the importance of the perturbations with respect to the measurable components of the configuration.

At this point, the decision problems can be classified based on the occurrence or not of each of the three main complicating factors, obtaining $2^3 = 8$ families of possible decision problems. Figure 1.4 describes this classification with a three-axes scheme.

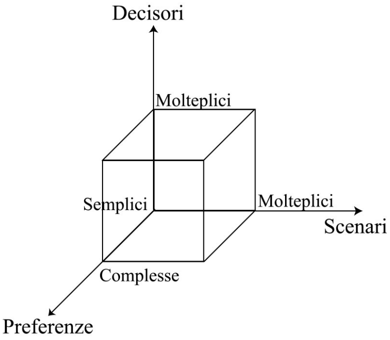


Figure 1.4: Classification of decision problems based on the three complicating factors

On this three-axes scheme is based the course syllabus and the index of these notes. Instead of discussing all 8 classes, we will only consider the 4 fundamental families in which none or only one of the 3 complicating factors occurs, discussing possible important subfamilies. The other 4 families can be treated by suitable combinations of the concepts adopted for the basic families. The course and these notes therefore present in the first four parts the corresponding families of basic prescriptive models:

1. models with simple preferences, a single scenario and a single decision-maker (*Mathematical Programming*)
2. models with complex preferences, a single scenario, a single decision-maker:
   - multiple objectives (*Multiobjective Programming*)
   - preference not modelled by objectives (*Analytic Hierarchy Process*)
   - nontransitive preferences (*Electre methods*)
3. models with simple preferences, multiple scenarios, a single decision-maker:
   - complete uncertainty (*Decisions in conditions of ignorance* or *Robust Programming*)

  - stochastic uncertainty (*Decisions in conditions of risk* or *Stochastic Programming*)

4. models with simple preferences, a single scenario, multiplie decision-makers:
   - independent decision-makers (*Game theory*)
   - coordinating decision-makers (*Group decisions*)

and in the fifth part some sample families of sophisticated descriptive models:

- queuing theory
- discrete-event simulation
- transportation and discrete choice models
- dynamic systems

Chapter 2

Case studies

This chapter is devoted to present two case studies, in order to clarify the concept of complicated decision problem and to understand the practical meaning of the keywords introduced in Chapter 1. The purpose of the case studies is not to be studied in detail to pass the exam, but to build a practical idea of the abstract definitions on which the course is based, and to learn to recognize such concepts in other, very different, situations. It must be specified that these studies concern works that have not yet been implemented, and that therefore could potentially undergo several variations and corrections. The former case, in particular, is a very preliminary study, performed during a master's thesis with the support of the township of Como, whereas the latter is an official feasibility study entrusted by the township of Milan to the Dipartimento di architettura e studi urbani of the Politecnico di Milano.

Both case studies concern large public works. This is just one of the applications fiels in which complicated decisions can occur, but it is particularly suitable to provide sophisticated examples for teaching, because:

1. detailed information is often publicly available;
2. the three complicating factors listed in the introduction (complex preferences, uncertainty and multiple decision-makers) are present at the highest degree.

Concerning the first point, usually the law prescribes that sufficient information be publicly spread so that the citizens could develop a complete and correct idea about the works. Even if this is not strictly required, the government often decides to spread information spontaneously in order to inform the public opinion and acquire its support. This is much rarer in the case of business or private decisions. Concerning the second point, the decisions on large public works have by nature manifold impacts (economic, social, environmental, etc...). Hence, it is hard to express preferences as the simple maximization of a profit or the minimization of a cost. On the other hand, these decisions also depend on events that are not under the control of the decision-makers (wide-range social and economical phenomena, political events, etc...). Finally, these decisions involve large masses of people, associations and institutions which, even if not in charge of the decision, can exert a role in it or put pressure on the decision-makers. This justifies the choice of examples drawn from that application field.

2.1 The tramway of Como

2.1.1 The context

Como has been for centuries at the centre of commercial exchanges and people transfers. It is in fact at the crossing between a main north-south line, which connects Italy (in particular, Milan) to Switzerland and Germany, and a secondary east-west line, which connects the main towns at the end of the alpine valleys (Varese, Lecco, Bergamo, etc. . .). Figure 2.1 shows these two lines.

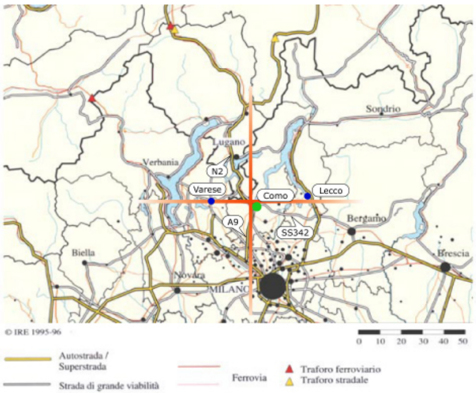


Figure 2.1: The two main traffic lines meeting in Como

The geographical position that has favoured the development of the town originated also some of the problems that now affect it. These problems are mainly related to the saturation of the access ways, with the following natural consequences:

- congestion (see Figures 2.2 and 2.3),
- economic losses due to slow transfers,
- pollution (see Figure 2.4),
- accidents.

Private cars are the main responsible for these problems, because they concentrate on few rather streets of limited width (see Figure 2.5). The main one is the access way from the south-west, through Napoleona Street, that collects two state roads (Briantea and Giovi) and a provincial road (Canturina). The following ones are the state roads Regina and Lariana, from the north, which go along the two narrow shores of the lake, and the state road Briantea from the east.

Enlarging these roads or opening other access ways is complicated and expensive due to the orography of the site: Como lies in a narrow plain along the lake, surrounded by high hills. Only the narrow "breach" of Grandate links it to the southern plain (see Figure 2.6).

A possible improvement could be to transfer people from the private car to the public transport, that is more efficient from the point of view of transportation capacity. However, the analyses show that the fraction of transfers serviced by the public transport is clearly smaller, and above all very uneven. Buses service only a minority of users, with a decreasing trend. Moreover, they are subject to the same access limitations of the private car. As for the train service (see Figure 2.7), the state railway stopping in Como S. Giovanni station (formerly, FS) is rather

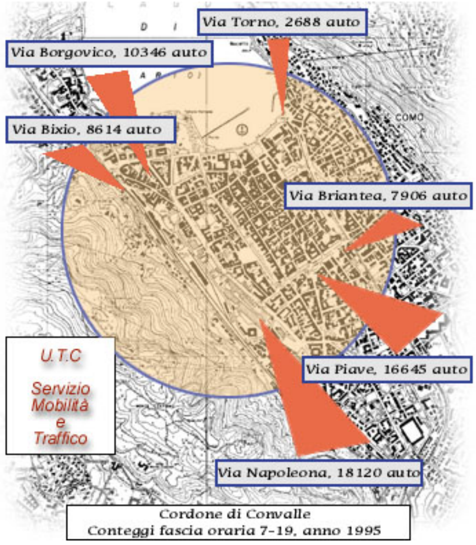
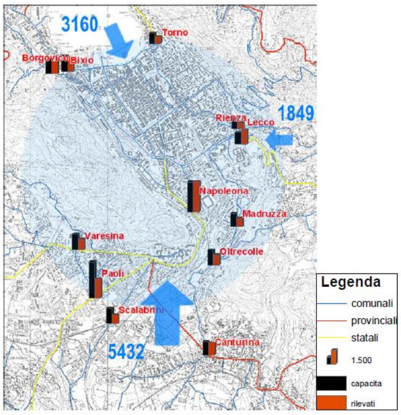
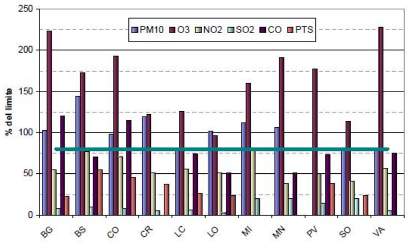


Figure 2.2: Vehicle flows entering the town of Como every day



Veicoli e gradi di saturazione delle direttrici in ingresso a Como (Conteggi 2002 – ore 8.00 9.00 giorno

Figure 2.3: Saturation levels of the streets entering Como during the morning peak-hour



ARPA Lombardia (2002), *Segnali ambientali della Lombardia, rapporto sullo stato dell'ambiente 2002.*

Figure 2.4: Pollution levels of the main towns in Lombardy

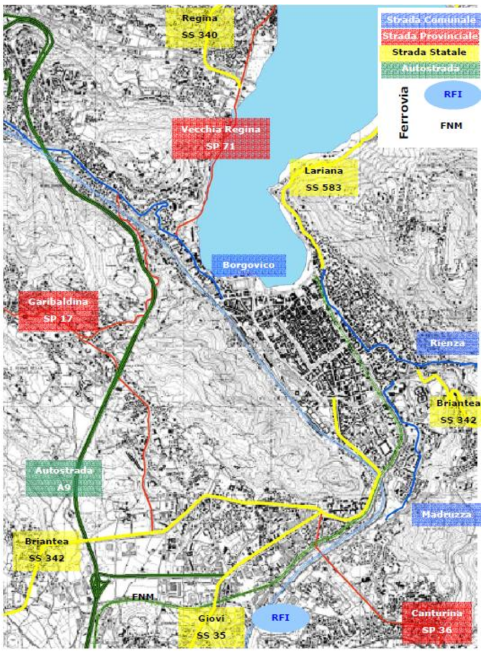


Figura 1.1: Le direttrici automobilistiche di accesso.

Figure 2.5: The main roads through which cars enter Como

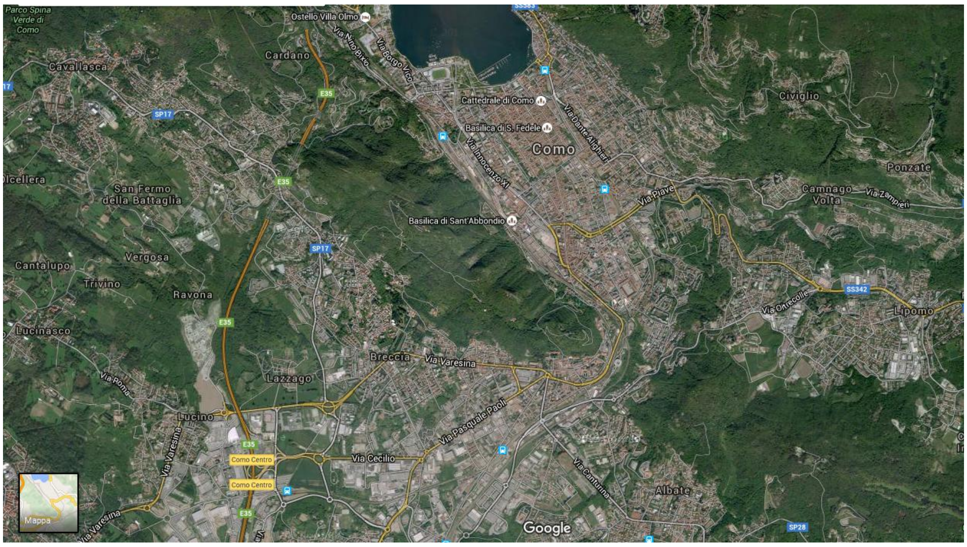
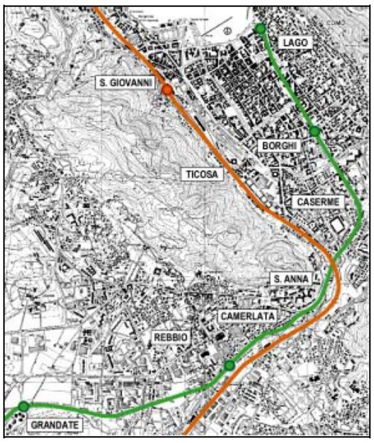


Figure 2.6: Orography of the area around Como

unloaded, whereas the regional railway stopping in Como Lago station (formerly, FNM) is nearly saturated[1] It is therefore spontaneous to think of increasing the capacity of the FNM line, possibly moving some of its passengers on the FS line. But how?



Servizio ferroviario attuale in Como (in verde la linea FNM e in arancione quella FS).

Figure 2.7: The two railways servicing Como: in orange the state line, in green the regional line

In order to better discuss the potential interactions between the two lines, it is necessary to list some critical points:

1. the crossing between the FS and FNM lines occurs in the breach, with an acrobatic system of bridge: the Napoleona street overcomes the FNM line, which overcomes the FS line, which overcomes a small local street named Mulini street (see Figures 2.8 and 2.9);

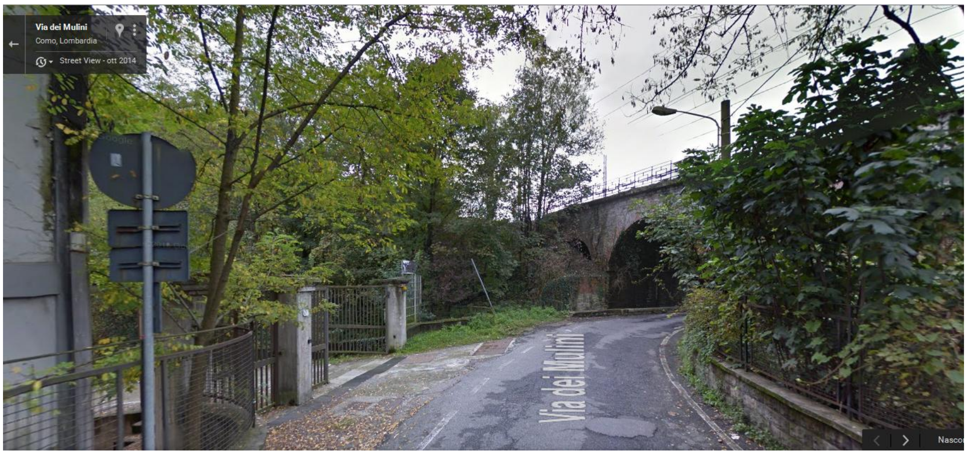


Figure 2.8: The FS line overcomes Mulini street

[1] Since the present study, the regional railways have started servicing also the state line, providing trains between Milan and Como San Giovanni. This has slightly improved the situation, better exploiting the residual capacity of the unloaded line. The two companies have also changed their names. For the sake of simplicity, we will use their historical names.

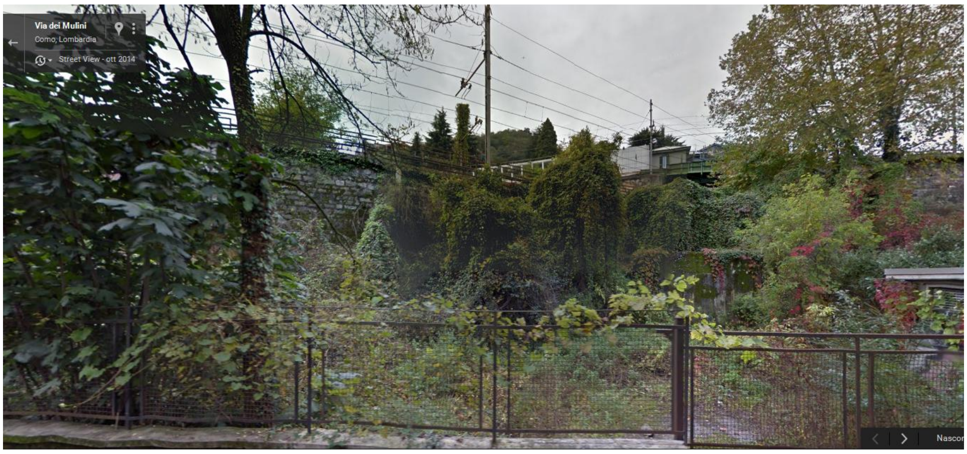


Figure 2.9: The FNM line overcomes the FS line

2. between the stations of Camerlata and Borghi, the FNM line has a single track, as it runs on a narrow hill with steep slopes on both sides;
3. between the stations of Borghi and Lago, the FNM line has a single track, and it runs along a narrow alley between houses (see Figure 2.10);
4. joining the FS and FNM lines through the town plain would require first to cross either the walls (in a historical area) or along the lakefront (in a tourist area, occupied by a main street and flooded from time to time), and then to climb the steep slope on top of which lies the FS station of Como S.Giovanni.

All these factors make a solution based on railways impossible to adopt, unless it is limited to the use of the existing tracks. Increasing the frequency of the trains on the FNM line, however, has two problems:

- the single-track line forbids two trains to run along the Camerlata-Borghi-Lago track (unless a perfectly synchronized meeting point is built in Como Borghi station, which is technically possible, but prone to big problems at the slightest unexpected event;
- the level crossings cut the town in two for several minutes every time a train passes along them: if a train passes every 10-15 minutes, instead of 30 minutes, the impact on the traffic congestion becomes very heavy.

Hence the idea of a tramway line, which could:

- have a double track, thanks to the shorter gauge;
- have fast crossings with traffic lights, instead of slow level crossings;
- be prolonged to the inside of the town.

But a classical tramway involves a big problem. It requires to replace the whole track, and therefore:

- it force the train passengers arriving from outside Como to change at an external station;
- it completely blocks the traffic during the works.

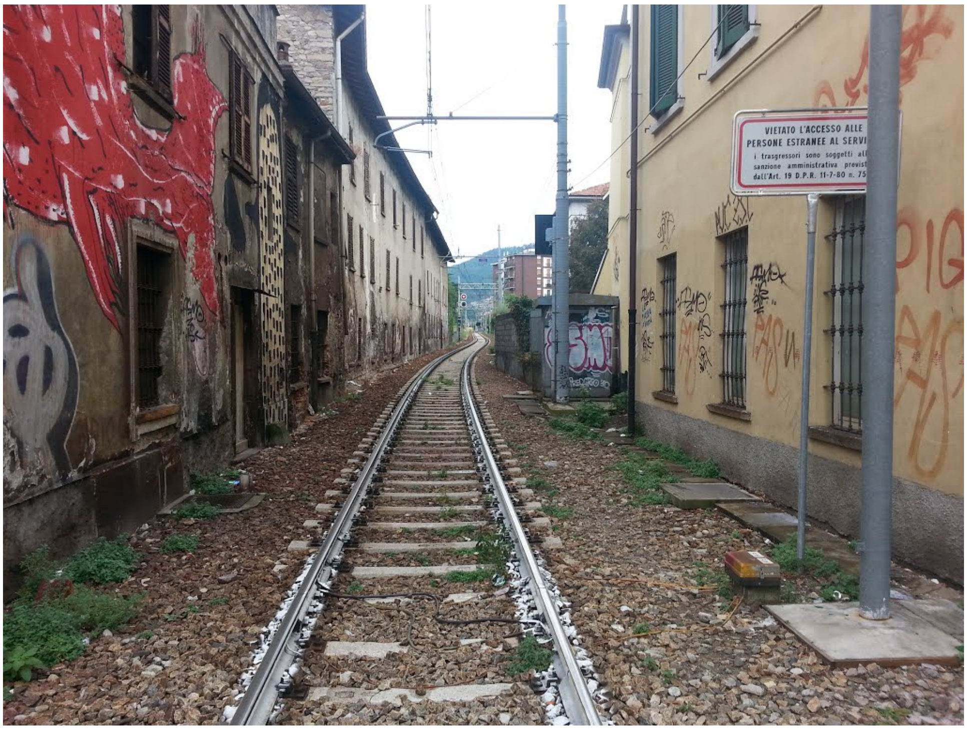

VIETATO L'ACCESSO ALLE
PERSONE ESTRANEE AL SERVI
i trasgressori sono soggetti all
sanzione amministrativa previst
dall'Art. 19 D.P.R. 11-7-80 n. 75

Figure 2.10: The FNM line between the Borghi and Lago stations

2.1.2 The generation of the alternatives

The basic models of Operations Research, as well as the optimization problems (even complicated ones) which arise in well-defined situations have a rather obvious set of possible alternatives. In the present case, this set is far from being clear. One would say that it is infinitely wide. More correctly, it is *indefinite*, and it must be built step by step investigating the whole context. Indeed, it can be built by subsequent iterations. As Section 1.2 pointed out:

- some alternatives are proposed;
- their impacts are evaluated;
- if no alternative is satisfactory, new alternatives are generated, under the guide of the information obtained from the study of the previous ones, and the process repeats.

For instance, if one realizes that modifying the gauge of the tracks is too expensive, one can wonder whether there exist tramways which are able to use a standard train track. Surprise surprise, they exist[2].

The process of generating the alternatives becomes easier if one decomposes the problem. identifying the basic *alternative elements*, which concern different fields and are (at least partly) independent. In this study, three fundamental elements have been identified, in order to base on them the generation of the alternatives:

1. the technology used to implement the new line;

[2] See `https://it.wikipedia.org/wiki/Tram-treno`

2. the route of the new line;
3. the management of the FNM trains with respect to the new line.

Let us discuss each element in further detail.

Technology Previous studies on similar cases suggest three possible technologies:

1. standard railway service: add a shuttle-train between Grandate and Como Lago, so as to increase the frequency of the current railway service;
2. tramway: replace the railway with a classical tramway;
3. interoperabile: add to the current railway service a special tramway which is able to work with both systems (facing the resulting problems about power supply, track gauge, etc. . .).

Route The number of possible routes is huge and not well-defined *a priori*. To start, four possible routes have been identified (see Figure 2.11):

1. keeping the current FNM route, ending at the station of Como Lago;
2. linking the station of Como Lago (FNM) to Como S. Giovanni (FS) through new tracks crossing the town centre;
3. linking the station of Como Lago (FNM) to Como S. Giovanni (FS) through new tracks passing along the lakeside;
4. building a ring around the town centre through a new track from Como Lago (FNM) to Como S. Giovanni (FS), plus a track from Como S. Giovanni (FS) to Como Borghi (FNM).

Management of the FNM trains This element of the alternatives describes the relation between the two rail transportation systems:

1. keep the FNM service on its current route;
2. move the FNM service to the FS station of Como S. Giovanni, through the construction of an interchange between the FNM e FS lines (probably in the station of Camerlata).

These are very extreme possibilities, with a whole range of intermediate solutions, either during the works (the FNM service could be temporarily deviated on Como S. Giovanni when the FNM track is unavailable for the works) or permanently. For example, presently, the regional company also manages the local traffic from Milan to Como S. Giovanni, even if the two lines remain completely unrelated.

Each of these aspects provides an *alternative element*. The single possible choices for an element can be associated to numerical values: sometimes, they will be a proper quantitative measure; other times, they will be simple arbitrary indices. For example, one can use 0 to denote the railway, 1 the tramway, 2 the interoperabile. The purpose of this association, in general, is not to make some computations, but only to describe effectively the set of alternatives.

Once the elements of the alternatives and the possible values for each element have been identified, one can proceed to enumerate their combinations, which in

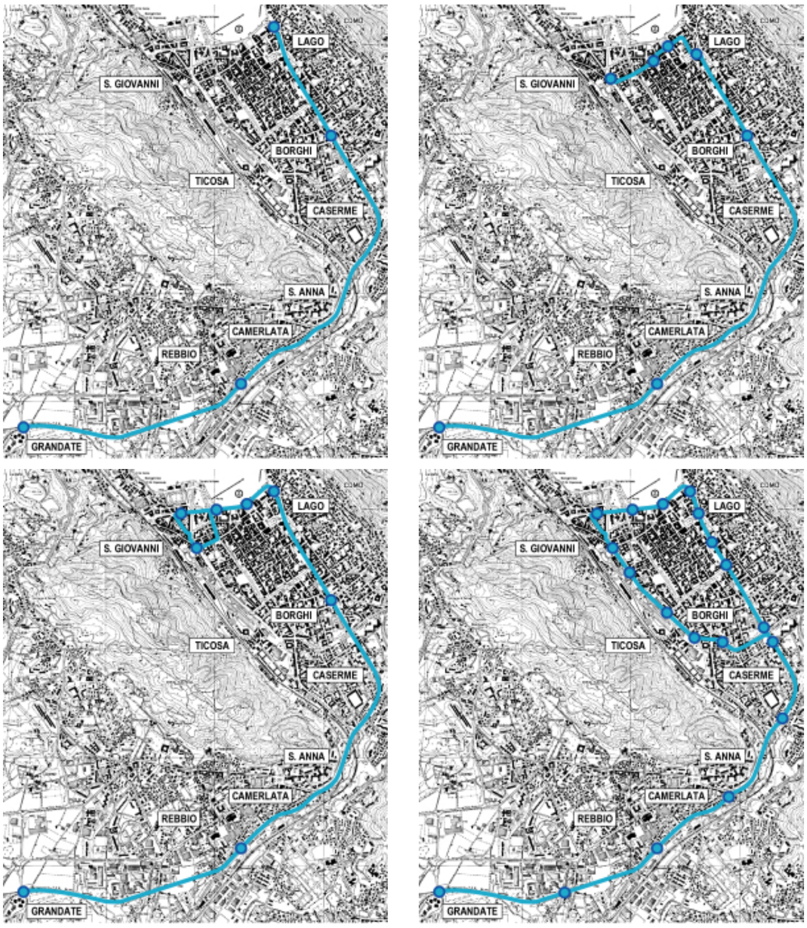


Figure 2.11: The four alternative routes for the project of Como's tramway

general are not all feasible. In the present case, the possible combinations are $3 \times 4 \times 2 = 24$, but only 7 combinations are feasible. The other ones can be excluded owing to:

- obvious contradictions;
- technical impossibility or excessive costs;
- common sense.

For example, the railway technology is in practice incompatible with any route different from the present one: nowadays, trains cannot cross a town as it was customary until the beginning of the twentieth century, unless through a tunnel, that is here impossible to propose. Figure 2.12 shows the possible combinations and the feasible alternatives for the case study.

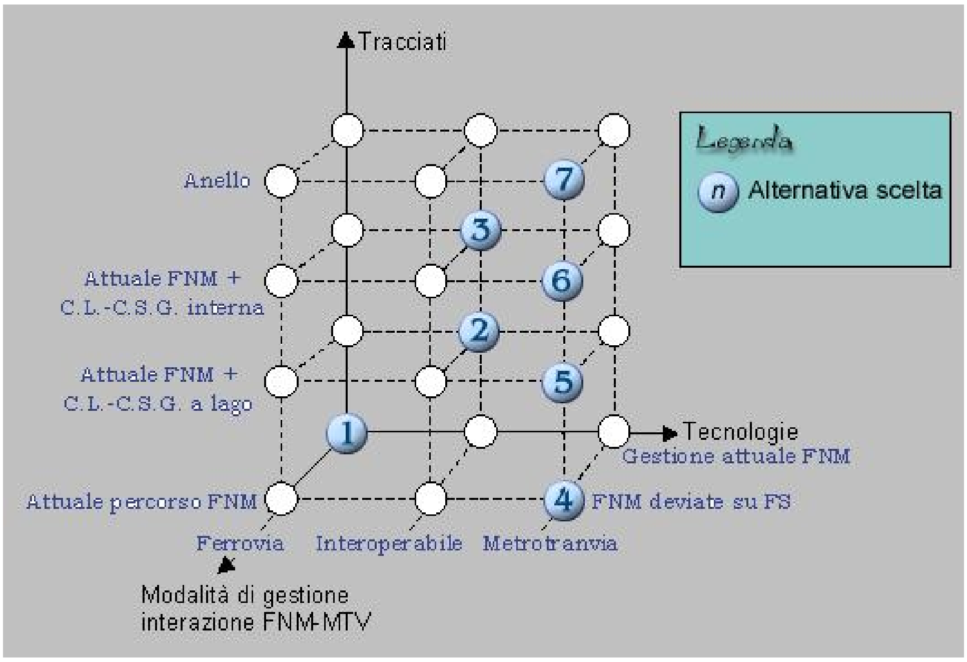


Figure 2.12: The 24 combinations and the 7 feasible alternatives for the case study on Como's tramway; the three axes represent the elements of the alternatives

A fundamental remark on the generation of the alternatives is that *there always exists an alternative that consists in doing nothing and keeping the current situation.* This alternative is conventionally denoted with the name of *alternative zero*[3] This remark is important, because quite often the heat of the public debate leads to forget that new alternatives should at least not worsen the current situation. Since "worsen" is a complicated concept, the remark is far from being trivial.

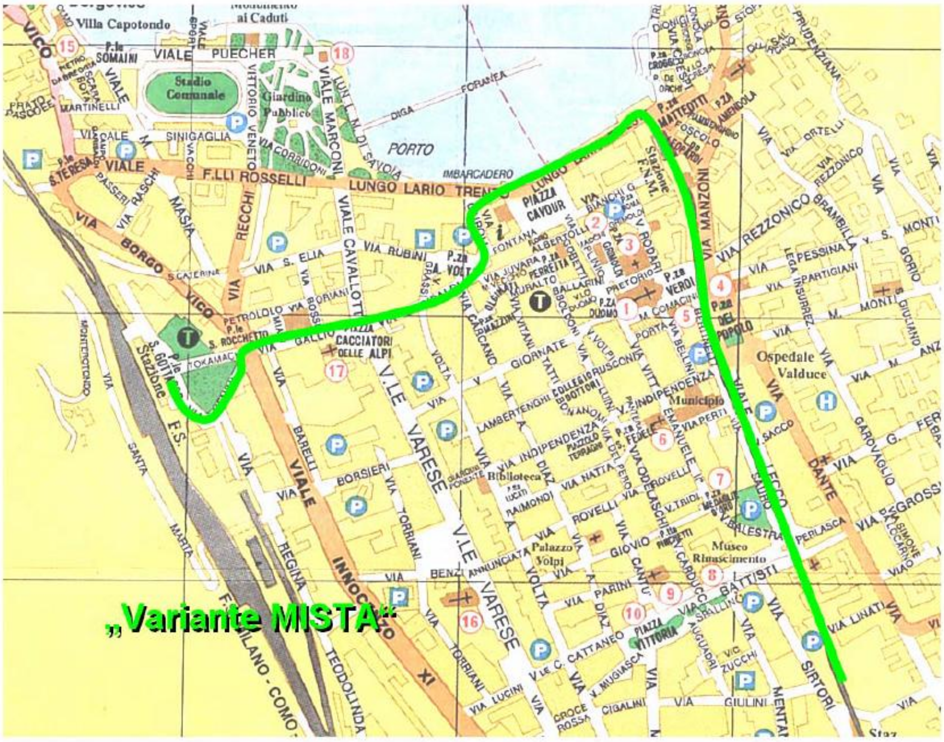
In this discussion, we have neglected several other potential elements of alternative, such as:

- the location of interchange parkings (in various sites);
- the implementation of double-track rails along parts of the current single-track rail;

[3] In this case study, unfortunately, it is denoted with index 1.

- the construction of new stations along the current route and the possible new routes;
- possible transfers of the current route;
- the extension of the service out of Como, beyond the station of Grandate;
- the frequency of the new service;
- the tariffs of the new service;
- ...

In particular, some mobility studies suggest that there could be a strong request for a railway servicing parts of the province currently badly connected to Como, both estward (Cantù) and westward (Olgiate Comasco). Other feasibility studies following the one we are describing, in fact, have focused on these aspects when generating the alternatives: they have rejected the most extreme solutions (current route and town ring) and merged the intermediate ones (centre crossing and lakeside route) in a mixed variant, first passing along the lakeside and then crossing part of the centre (see Figure 2.13).



Variante Mista, il collegamento Como San Giovanni (FS) - Como Lago (FNM).

Figure 2.13: The mixed route alternative passing first along the lakeside and then across the centre

As well, we have neglected the possible *mitigation measures*, that is all those interventions that must be added to the alternatives in order to correct the negative impacts that they produce (traffic or railway service deviations, urban quality improvements to reduce the negative impact of the new line along its route, etc...)

or to convince possible stakeholders strongly opposing the project to accept it or at least not to fight it fiercely (for example, the parks, schools, sport complexes that are built for the towns whose territory hosts a highway, a dump, a power plant, etc...), or, finally, to modify the impacts of the works (for example, limited access zones to discourage the use of the private car and interchange parkings to encourage the travellers to use the new public service more than they would otherwise spontaneously do).

All this confirms that even listing the possible solutions of a decision problem can be a preliminary problem in itself, that is not to be solved in a single phase, and actually is never ultimately solved.

2.1.3 The generation of scenarios

The same complications faced to generate the alternatives affects the definition of the scenarios. In addition, this requires to predict events which are uncontrollable by the decision-makers, but have a potential impact on the problem.

The generation of the scenarios follows the same rationale as the generation of the alternatives: first, one identifies the *scenario elements* and their possible values; then, one lists their combinations and removes those that are considered impossible. Of course, this excludes from the study any unpredicted element and any removed combination, which could undermine the basis of the whole model. On the other hand, the alternative is to proceed at random or by intuition, which could be worse. As already said, if new scenarios emerge or old scenarios disappear, the study should be simply updated and corrected, possibly with the result to suggest a different alternative for the final decision.

In the present case, the scenario elements considered are the following ones:

1. closure of the lakeside to the private traffic: it consists in forbidding the access of vehicles to the street that runs between the city centre and the lake: this interacts with the project because it frees-up the space used by cars, thus making it easier to build a tramway along the lakeside;

2. International Como-Chiasso Station: it consists in the merging of the international railway stations of Como (Italy) and Chiasso (Switzerland) in a single stop, half-way (see Figure 2.14), thus accelerating the trips between Italy and central Europe: this interacts with the project because the station of Como S. Giovanni would be reduced to the level of local station, thus introducing the need for a service taking travellers to Como S. Giovanni and possibly onward to the new international station[4]

3. Anti-flood barriers: it consists in building barriers (partly mobile, partly fixed) that protect the centre of Como from the floodings of the lake (see Figure 2.15): this interacts because a tramway service passing along the lakeside would be blocked by the floodings[5]

4. Borgovico tunnel: it consists of a toll tunnel that should run from the north-west to the south-west of Como, allowing to cross the town without congesting its western side[6]

[4] This is still a project; recently, the trend seems to favour the suppression of Como directly in favour of Chiasso.

[5] The barriers are currently under construction, with strong protest on their visual impact. A stretch of fixed barrier has even been built and destroyed because it blocked the view on the lake to the passers-by...

[6] This project has been apparently put aside because it was too expensive.

5. Underlake tunnel: it consists of a tunnel which should run under the lake (see Figure 2.16) to replace the street that currently connects the north-west ad the north-east of Como[7]

In theory, each combination of these elements is possible, with the exception of those that combine the underlake tunnel with keeping the lakeside open to the privat cars (a meaningless idea) and those that combine both tunnels (too expensive).

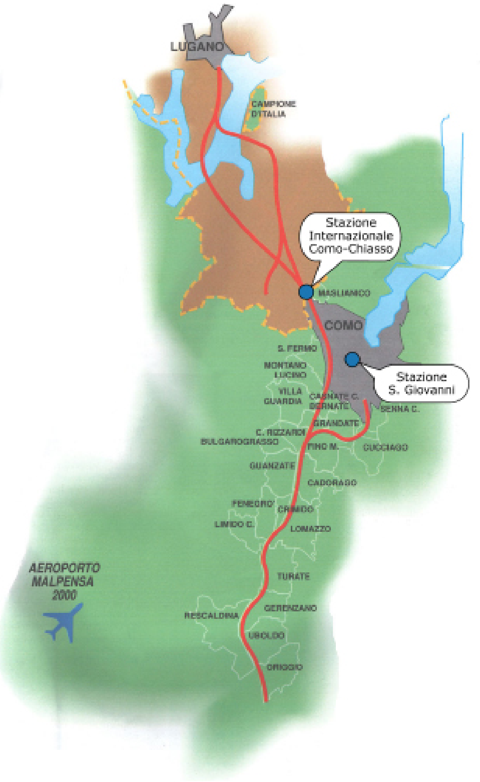


Figure 2.14: La Stazione Internazionale Como-Chiasso

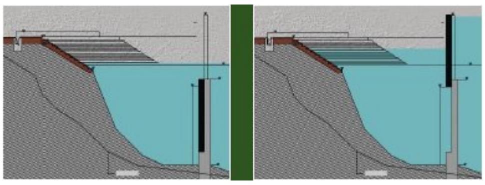


Figure 2.15: The anti-flooding barriers along the lake

This discussion neglects important aspects such as:

- the anticipated varations in the residential and economic structure of the area under study;
- the variations in the *origin-destination matrix* (*O/D-matrix*) of the potential transportation demand (that is, the number of trips from place to place that will take place in the town and could be captured by the new service);

[7] This project has been apparentely put aside because it was too expensive.

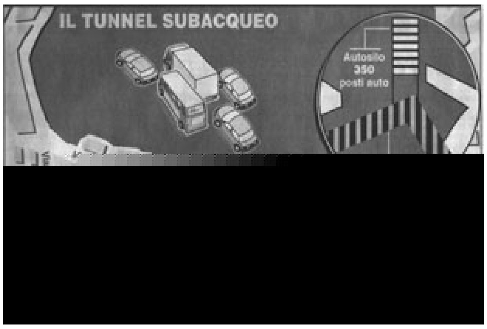


Figure 2.16: The underlake tunnel

- the amount of European, state, region financing.

Some of these scenario elements correspond to numerical values, which could fall inside predictable ranges. Some, on the other hand, correspond to events that could either occur or not, or that could occur in different ways. For each numerical value, for each event or for each way in which an event can occur, it could be possible to estimate a probability, or at least provide a qualitative estimation of its likelyhood. All this information concurs to describe the set of possible scenarios and its auxiliary information.

2.1.4 The definition and computation of the indicators

As the alternatives and the scenarios, also the impacts are combinations of different elementary quantities, which are usually not named "impact elements", but *indicators.* They characterize the satisfaction of the decision-makers for a configuration of the system, and determine their preferences. The indicators are typically much more numerous and diversified than the elements of the alternatives and the scenarios. They are so numerous and diversiied that, in order to avoid forgetting any, their generation adopts a hierarchical process that progressively details the different sectors of the impact:

- first, one identifies very general macrosectors;
- then, each macrosector is decomposed into sectors, and possibly into subsectors, progressively more and more specific;
- finally, the elementary indicators are identified.

This mechanism produces an *indicator tree.*

In the present case, the three standard macrosectors used in *Environmental Impact Assessment* (*EIA*) have been adopted:

1. *Environment*, subdivided into Air pollution, Noise, Vibrations, Landscape and Territorial structure (that is, the compatibility with the current management plans of the area)[8];
2. Economy, subdivided into Costs, Revenues and House values;

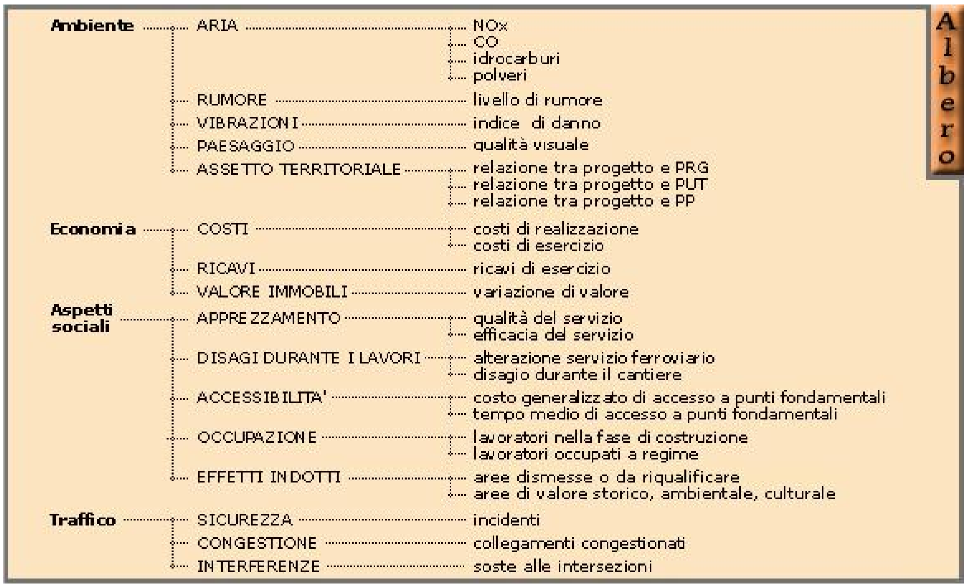
[8] Water and Land pollution have been neglected, given that the impact of the tramway will probably be very limited under those regards.

3. Society, subdivided into Appreciation, Discomfort, Accessibility[9], Employment, Induced effects;

but a fourth macrosector has been added:

4. Transportation, subdivided into Security, Congestion and Interferences.

The last macrosector usually is included in the Society macrosector, but in this case it has been attributed an autonomous role as the study specifically evaluates a large transportation service. This has yielded the indicator tree reported in Figure 2.17.



- **Ambiente**
  - ARIA
    - NOx
    - CO
    - idrocarburi
    - polveri
  - RUMORE
    - livello di rumore
  - VIBRAZIONI
    - indice di danno
  - PAESAGGIO
    - qualità visuale
  - ASSETTO TERRITORIALE
    - relazione tra progetto e PRG
    - relazione tra progetto e PUT
    - relazione tra progetto e PP
- **Economia**
  - COSTI
    - costi di realizzazione
    - costi di esercizio
  - RICAVI
    - ricavi di esercizio
  - VALORE IMMOBILI
    - variazione di valore
- **Aspetti sociali**
  - APPREZZAMENTO
    - qualità del servizio
    - efficacia del servizio
  - DISAGI DURANTE I LAVORI
    - alterazione servizio ferroviario
    - disagio durante il cantiere
  - ACCESSIBILITA'
    - costo generalizzato di accesso a punti fondamentali
    - tempo medio di accesso a punti fondamentali
  - OCCUPAZIONE
    - lavoratori nella fase di costruzione
    - lavoratori occupati a regime
  - EFFETTI INDOTTI
    - aree dismesse o da riqualificare
    - aree di valore storico, ambientale, culturale
- **Traffico**
  - SICUREZZA
    - incidenti
  - CONGESTIONE
    - collegamenti congestionati
  - INTERFERENZE
    - soste alle intersezioni

Figure 2.17: The indicator tree

The aim of the hierarchical structure is also to organize the evaluation of the single indicators during the phase of data collection. The estimation, the prediction and the measurement of the values of each indicator, in fact, will be assigned to experts of the sector to which they belong: environmental experts will evaluate the pollution levels predicted for each configuration; accounting experts will evaluate the implementation and management costs; town planning experts will evaluate the visual quality of the buildings; transportation experts the impact of the service on urban congestion and transfer times, etc...

As well, the construction of preferences in the actual decision phase will be organized based on the specific skills of the experts of the various sectors. The choice whether to consider carbon oxide pollution more or less important than nitrogen oxide pollution will be assigned to health and environment experts; the choice of the relative weight between implementation and management costs (the so called ammortamento) is assigned to economy experts. The tree structure implies that, at the upper levels, the decision-makers will determine the relative importance of air pollution with respect to noise, or security with respect to traffic congestion, and finally between the macrosectors Environment, Economy, Society and Transportation. These more and more abstract concepts require more and more "political" skills. Therefore, nobody will ever be asked to compare completely heterogeneous

[9] This is the time and cost required to access important sites in the town, such as hospitals, monuments, schools, etc...

indicators, such as CO_2 pollution and time savings in the trips to Como Lago stations, or visual quality of the tramway along the lakeside with the discomfort induced by the building sites during its construction. Someone will compare CO_2 and NO_x, trip costs and trip times, the consistency with the Urban Traffic Plan and the consistency with the Piano Regolatore Generale. Let us postpone the problem of how to compare abstract and vague concepts such as those denoting the macrosectors.

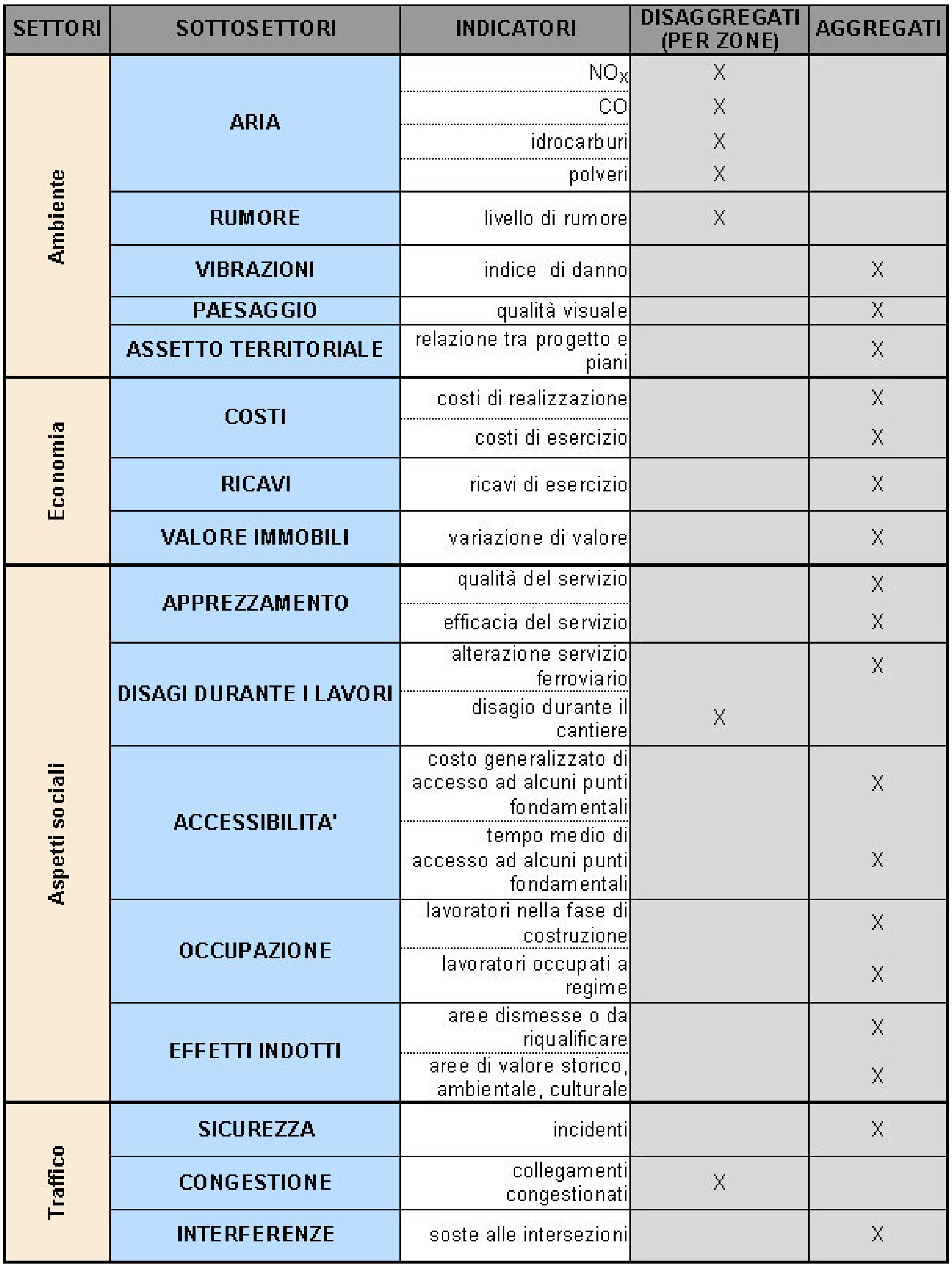
Each indicator must be associated with a tool which will provide its values for each possible configuration (that is, alternative and scenario). In most cases, these values will not be directly measurable, because they will refer to a hypotetical situation (one can directly measure the CO_2 pollution in the current configuration, that is with alternative zero and the current scenario, but it is necessary to estimate somehow the CO_2 pollution induced by the ring tramway in the scenario in which the lakeside is closed to the private cars, the anti-flooding barriers are built and no other scenario element is implemented. Quite often, therefore, the value of an indicator is the result of a descriptive model, for which data and computing tools must be identified.

The indicators can be numerical values that express physical measures, but also qualitative values expressed on an ordinal scale. For example, the visual quality of a building can be bad, poor, sufficient, fairly good, good, very good. These values can be translated into a numerical scale, for the sake of compactness. The purpose, at least at first, is not to make computations, but just to describe the situation. In practice, we will see that some evaluation methods actually use such numerical values to make computations.

Time decomposition (phases) and space decomposition (zones) Some indicators have an intrinsical *space* or *time* feature, which must be taken into account. In other words, the values of an indicator might be meaningless if referred to the whole territory or the whole time horizon of the project. In that case, in order to correctly describe the situation the system must be subdivided into distinct geographical zones, on which it is meaningful to define (at least approximately) a value of the considered indicator. For instance, the time required to reach the station of Como Lago will strongly depend on the starting point, even if it can be considered approximately uniform for a given zone of the town. Of course, this increases the amount of information to manage, and it requires to evaluate a relative importance of the different zones on the decision, dealing with the values of the indicator in the zones as if they were values of different indicators. Other indicators, even if related to a specific geographical position, can remain aggregated, because their effect concerns the whole town (for example, the damages to monumental buildings, which are a common heritage). See Figure 2.18 for the subdivision of the indicators into aggregated and e disaggregated into zones for the current case study.

In this case study, the town of Como has been subdivided into 5 zones which had already been used for other previous works on the same territory (see Figure 2.19):

1. Como centro
2. Borghi
3. Camerlata
4. Lora
5. Tavernola



SETTORI	SOTTOSETTORI	INDICATORI	DISAGGREGATI (PER ZONE)	AGGREGATI
Ambiente	ARIA	NO_X	X	
		CO	X	
		idrocarburi	X	
		polveri	X	
	RUMORE	livello di rumore	X	
	VIBRAZIONI	indice di danno		X
	PAESAGGIO	qualità visuale		X
	ASSETTO TERRITORIALE	relazione tra progetto e piani		X
Economia	COSTI	costi di realizzazione		X
		costi di esercizio		X
	RICAVI	ricavi di esercizio		X
	VALORE IMMOBILI	variazione di valore		X
Aspetti sociali	APPREZZAMENTO	qualità del servizio		X
		efficacia del servizio		X
	DISAGI DURANTE I LAVORI	alterazione servizio ferroviario		X
		disagio durante il cantiere	X	
	ACCESSIBILITA'	costo generalizzato di accesso ad alcuni punti fondamentali		X
		tempo medio di accesso ad alcuni punti fondamentali		X
	OCCUPAZIONE	lavoratori nella fase di costruzione		X
		lavoratori occupati a regime		X
	EFFETTI INDOTTI	aree dismesse o da riqualificare		X
		aree di valore storico, ambientale, culturale		X
Traffico	SICUREZZA	incidenti		X
	CONGESTIONE	collegamenti congestionati	X	
	INTERFERENZE	soste alle intersezioni		X

Figure 2.18: Aggregated and disaggregated indicators

Only the first three zones are directly affected by the project, whereas Lora and Tavernola have been maintained to check the indirect effects on the whole town.

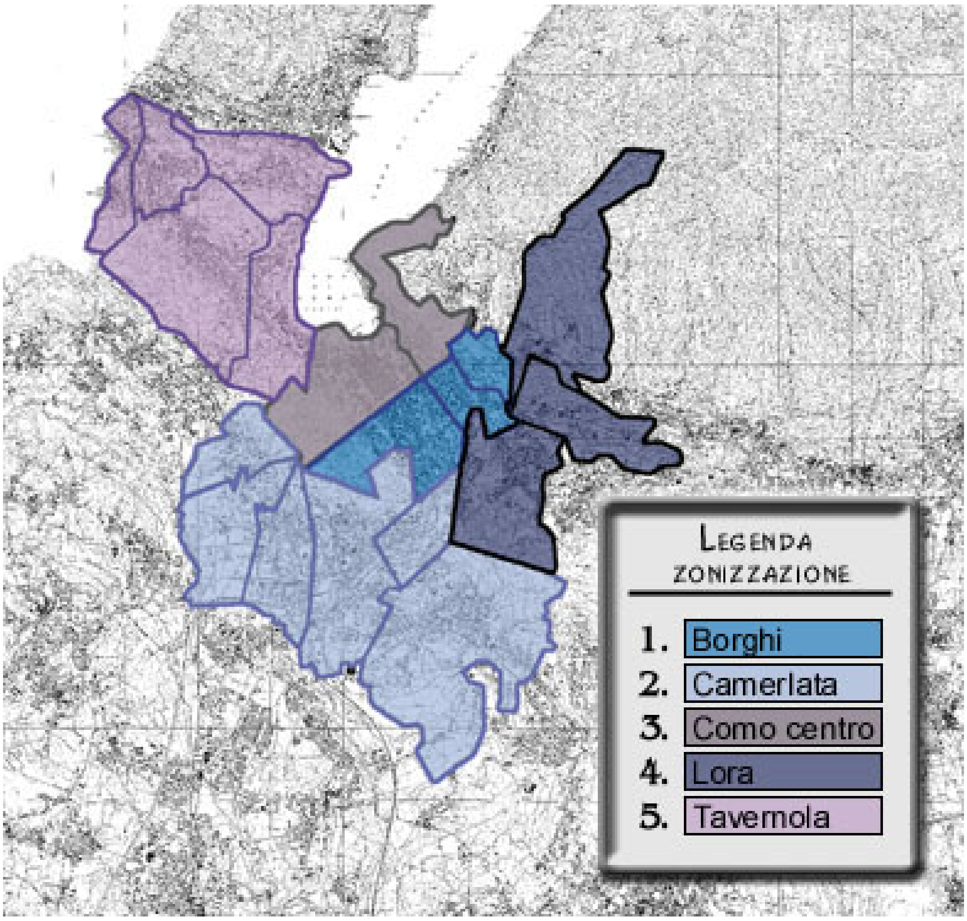


Figure 2.19: Subdivision of the town of Como into zones

2.1.5 The definition of the stakeholders

We denote as *stakeholder* every person, organization, category or institution that is involved, directly or indirectly, in the project, even if it plays no role in the decision, since its interests are affected by the project, and it is therefore presumable that it could react to defend them in a more or less favourable way for the project. The fundamental aspect in the description of the stakeholders is to characterize their *preferential structure*. We will discuss later its definition and how to determine and describe it. However, the stakeholders could also point out neglected alternatives, scenario elements and indicators.

In the present case, four classes of stakeholders have been identified and further subdivided:

1. Institutional stakeholders (Municipality, Province, Region)
2. Society
    - Citizens

- Environmentalist associations
- Category associations

3. Users

   - regular (commuters, students, etc. . .)
   - irregular

4. Transportation companies

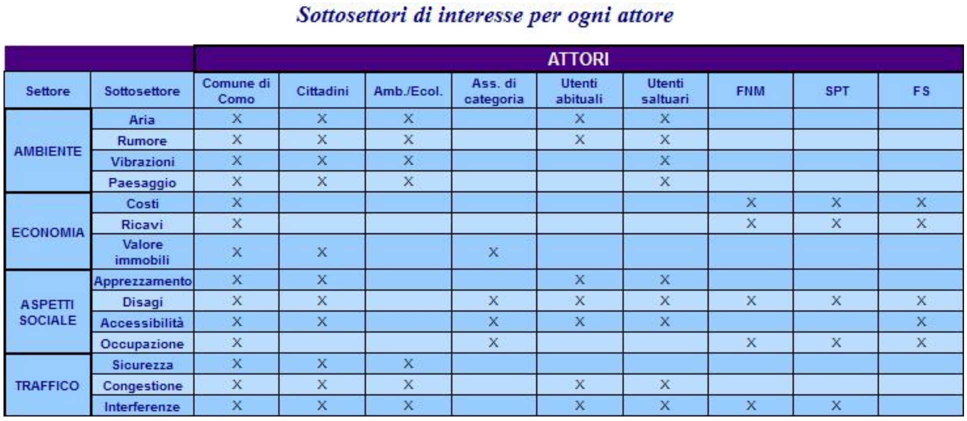
   - FNM (the regional railway company, whose track between Grandate and Como Lago should be used by the new service);
   - SPT (the local public transport company managing the bus lines, whose demand would be disrupted by the new service, requiring a radical revision of the routes and schedules)[10];
   - FS (the state railway company) whose station of Como S. Giovanni could host part or all the trains currently servicing the FNM line.

Each stakeholder has limited interests ad skills, often related to one or few subtrees of the indicator tree (this is another reason for the hierarchical modelling of the impact). Hence, it will usually be involved only (or mainly) in the phases of the decision process which concern those specific indicators. Figure 2.20 describes a hypothetical relation between the stakeholders and the subsectors in which they are likely to be interested:

1. the local government stakeholders (in particular, the municipality of Como) will be interested in the whole system of indicators;
2. concerning society stakeholders:

   - the citizens will probably focus their interest on the Environment, Society and Transportation macrosectors;
   - the environmental associations will focus on the Environment macrosector;
   - the category associations (Camera di Commercio, Industria, Artigianato e Agricoltura - Confesercenti - . . .), who support the economic interests of the traders and entrepreneurs, will focus on the Economy macrosector and on the discomfort induced by the works;

3. the users will focus on the Society macrosector, and in particular on the quality and the accessibility of the service;
4. concerning the transportation companies:
5. FNM will focus on the Society macrosector (in particular, Employment and Discomfort) and the Economy macrosector (Costs and Revenues);
6. SPT is particularly interested in the Society, Economy and Transportation macrosectors, and in particular in Employment, Discomfort, Traffic congestion and its own Costs and Revenues);
7. FS is interested in the management of the train service potentially moved to the station of Como S. Giovanni.



Sottosettori di interesse per ogni attore

		ATTORI								
Settore	Sottosettore	Comune di Como	Cittadini	Amb./Ecol.	Ass. di categoria	Utenti abituali	Utenti saltuari	FNM	SPT	FS
AMBIENTE	Aria	X	X	X		X	X			
	Rumore	X	X	X		X	X			
	Vibrazioni	X	X	X			X			
	Paesaggio	X	X	X			X			
ECONOMIA	Costi	X						X	X	X
	Ricavi	X						X	X	X
	Valore immobili	X	X		X					
ASPETTI SOCIALE	Apprezzamento	X	X			X	X			
	Disagi	X	X		X	X	X	X	X	X
	Accessibilità	X	X		X	X	X			X
	Occupazione	X			X			X	X	X
TRAFFICO	Sicurezza	X	X	X						
	Congestione	X	X	X		X	X			
	Interferenze	X	X	X		X	X	X	X	

Figure 2.20: Relation between the stakeholders and the subsectors in which they are interested

Once the stakeholders have been identified, a reference person (or group of persons) should be defined for each stakeholder, and involved in all phases of the decision process: problem definition, identification of the alternatives, of the scenarios, of the indicators, computation of the impacts, description of the preferences. These reference persons should be also involved in the subsequent iterations in which alternatives, scenarios, indicators and preferences are updated based on the partial results obtained from the previous iterations of the study. The tools to interact with these persons depend on the specific case:

1. the interactions with the local government stakeholders could take place with the mayor of Como, the governor of the Province of Como and the governor of Lombardy, or their delegates, through meetings and detailed interviews;
2. as for society stakeholders:
   - the interactions with the citizens could exploit forms, websites, public meetings and interviews with the local representatives (consiglieri di circoscrizione);
   - the interactions with environmental or category associations could involve forms, meetings and interviews with their spokesmen;
3. the interactions with the users could require polls among the passengers and meetings with commuters' committees;
4. the interactions with the transportation companies could take place with managers specifically charged with this task.

It can be notice that environmentalists, shopkeepers and users are also citizens, and therefore exert a double role. Isn't this a useless redundancy? Not actually, since the preferences of common citizens will be rather different from those of specific categories, and the ways and tools to express such preferences will also be different. Therefore, it is correct to represent them as different stakeholders, even if they are partly overlapping. Keeping all this into account should guarantee a more fluid development of the whole decision process, avoiding that some subjects might feel neglected and decide to oppose the project on principle.

[10] An interesting development is that the bus lines in Como are now managed by ATM, after a (very rare) call for tenders.

2.1.6 The decision process

For each possible configuration of the system, i. e. for each alternative and each scenario, it is necessary to evaluate each indicator. This will require in some cases physical measurements, but most of the time the application of suitable descriptive models in order to compute the expected value of the indicator in the given configuration. Speaking in qualitative terms:

- the alternatives exploiting the railway technology will have:
  - low implementation costs and a low discomfort;
  - a limited effect of the accessibility of the sites not currently serviced;
  - limited environmental improvements;
  - nearly no impact on the employment;
  - ...
- the alternatives exploiting the "interoperabile" technology will have:
  - higher economic costs;
  - a stronger penetration of the service in the urban fabric, mainly depending on the chosen route;
  - an intermediate discomfort;
  - ...
- the alternatives exploiting the classical tramway technology will have:
  - much higher implementation costs and times;
  - a very detailed service;
  - ...

The resulting huge amount of quantitative numerical values and qualitative ordinal values will have to be reaggregated depending on the preferences of the various stakeholders. It is necessary to model somehow the preferential structure of each stakeholder, in order to be able to determine, for each given pair of possible impact, which is preferred by the stakeholder. This requires a basic methodological choice about the kind of model used to describe the preferential structure.

Once this has been done, it is necessary to aggregate the preferences of the different stakeholders in a final ordering of the alternatives, or at least in the selection of one alternative. This can be done in several different ways, since the specific values associated to the elementary triplets (alternative, scenario, indicator) can be recombinated with very different methods. For example, some methods start from the indicators, followed by the scenarios, providing a specific ordering of the alternatives for each stakeholder; only in the end the orderings of the various stakeholders are merged in a final ordering with a negotiation process. Other methods take into account the preferences of the stakeholder since the beginning. The study should possibly also provide information on the *sensitivity* of the final ordering or choice, that is on the bounds within which it remains valid when the data change (scenarios, impacts, preferences, relative importance of the stakeholders). We will discuss these subjects in detail in the following chapters.

2.2 The reopening of the Navigli in Milan

2.2.1 The context

Though few today perceive it, Milan is born as a river town, rich in natural rivers (the Olona, the Vepra, the Nirone, the Seveso, the Southern Lambro and the Lambro), torrents and brooks, and later further enriched by an infinite network of canals. Modernity led to covering nearly the whole network to favour the vehicle traffic and, more recently, to a periodic proposal and discussion of project to reopen part of it. Figures 2.21 and 2.22 show the historical and current situation of Milan and its hydrographic network.

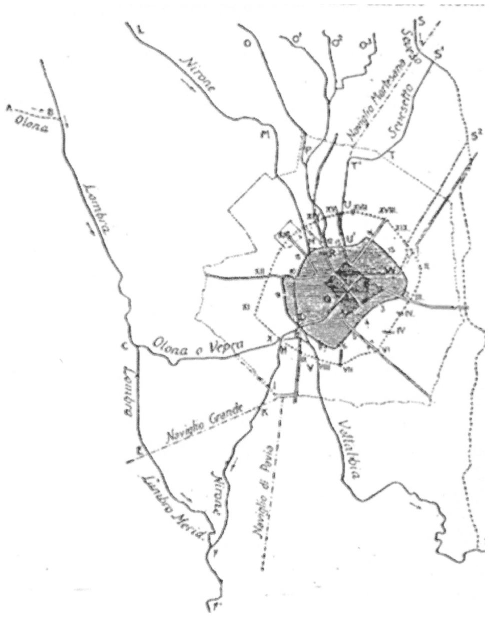


Figure 2.21: Historical map of the gradual enlargement of the city of Milan and its hydrographic network

The consultive referendum of 2011 has led to the development of a detailed project that is available in the internet[11]

Usually, the subject is laid out and tackled from the urbanistic, historic and aesthetic point of view, that is as an *improvement of the landscape and urban quality* and an *enhancement of a cultural heritage* nearly forgotten, though still quite recent. In fact, there are other relevant motivations that are entwined with the ones above and that concur to the interest of a potential reopening. Summarising, the main motivations are:

- the *hydraulic continuity*, that is the possibility to better control and to limit the flooding of water streams in the northern part of Milan, which currently flow through pipes under the city with the resulting limitations on the capacity and the troubles in managing their cleansing;
- the *touristic and commercial navigability*, with their social, cultural and economic outcomes (see Figure 2.23 for the current situation);

[11] See the URLs `http://www.comune.milano.it/wps/portal/ist/it/servizi/territorio/riapertura_navigli_2016` and `http://www.cuoredimilano.org/index.html`.

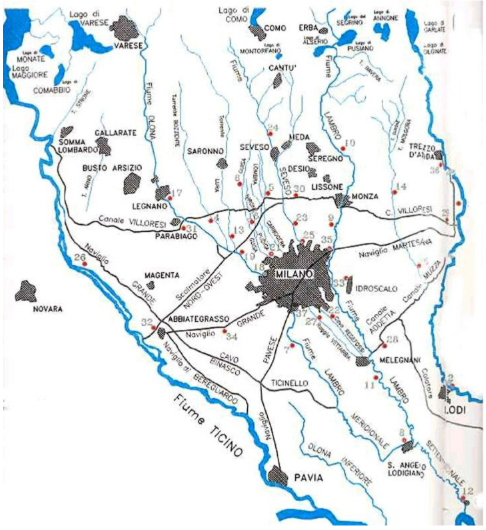


Figure 2.22: Current hydrographic network of the area surrounding the city of Milan

- the *production of electricity*, not by directly installing turbines in the Navigli, but with the increase of the water flow in the Naviglio Pavese, which has several small hydroelectric plants south of Milan;
- the *feeding of heat-pump plants* in the area of the Darsena using nappe water (undergound water);
- the *decrease in the nappe level*, thanks to the extraction of water from wells which are currently close; this would reduce the ceaseless use of pumps in several underground stations that defend the tunnels from flooding (the level of the nappe is strongly increased since the Seventies, due to the deindustrialization of the city).

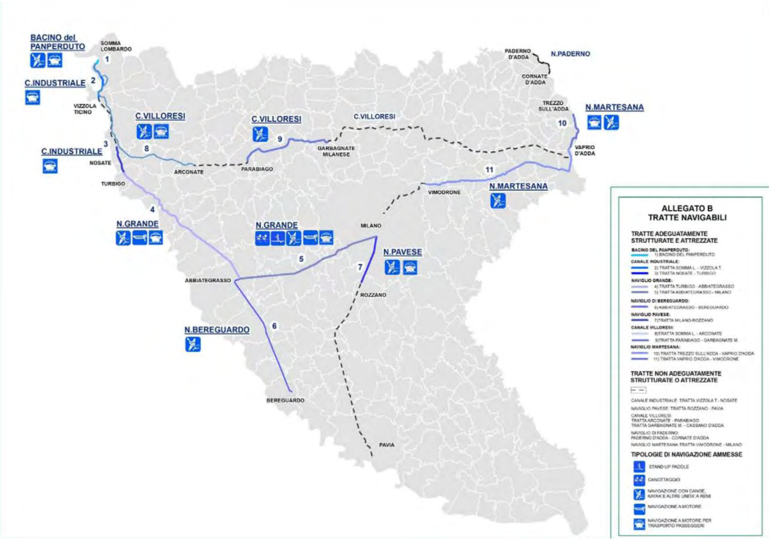

BACINO del PANPERDUTO
SOMMA LOMBARDO
C.INDUSTRIALE
VIZZOLA TICINO
C.VILLORESI
C.INDUSTRIALE
NOSATE
TURBIGO
ARCONATE
PARABIAGO
GARBAGNATE MILANESE
C.VILLORESI
PADERNO D'ADDA
N.PADERNO
CORNATE D'ADDA
TREZZO SULL'ADDA
N.MARTESANA
VAPRIO D'ADDA
VIMODRONE
N.MARTESANA
N.GRANDE
N.GRANDE
MILANO
N.PAVESE
ROZZANO
ABBIATEGRASSO
N.BEREGUARDO
BEREGUARDO
PAVIA
1
2
3
4
5
6
7
8
9
10
11
ALLEGATO B
TRATTE NAVIGABILI
TRATTE ADEGUATAMENTE STRUTTURATE E ATTREZZATE
BACINO DEL PANPERDUTO:
1) BACINO DEL PANPERDUTO
CANALE INDUSTRIALE:
2) TRATTA SOMMA L. - VIZZOLA T.
3) TRATTA NOSATE - TURBIGO
NAVIGLIO GRANDE:
4) TRATTA TURBIGO - ABBIATEGRASSO
5) TRATTA ABBIATEGRASSO - MILANO
NAVIGLIO DI BEREGUARDO:
6) ABBIATEGRASSO - BEREGUARDO
NAVIGLIO PAVESE:
7) TRATTA MILANO-ROZZANO
CANALE VILLORESI:
8) TRATTA SOMMA L. - ARCONATE
9) TRATTA PARABIAGO - GARBAGNATE M.
NAVIGLIO MARTESANA:
10) TRATTA TREZZO SULL'ADDA - VAPRIO D'ADDA
11) TRATTA VAPRIO D'ADDA - VIMODRONE
TRATTE NON ADEGUATAMENTE STRUTTURATE O ATTREZZATE
CANALE INDUSTRIALE: TRATTA VIZZOLA T. - NOSATE
NAVIGLIO PAVESE: TRATTA ROZZANO - PAVIA
CANALE VILLORESI:
TRATTA ARCONATE - PARABIAGO
TRATTA GARBAGNATE M. - CASSANO D'ADDA
NAVIGLIO DI PADERNO:
PADERNO D'ADDA - CORNATE D'ADDA
NAVIGLIO MARTESANA: TRATTA VIMODRONE - MILANO
TIPOLOGIE DI NAVIGAZIONE AMMESSE
STAND UP PADDLE
CANOTTAGGIO
NAVIGAZIONE CON CANOE, KAYAK E ALTRE UNITA' A REMI
NAVIGAZIONE A MOTORE
NAVIGAZIONE A MOTORE PER TRASPORTO PASSEGGERI

Figure 2.23: The current situation of navigability in the area surrounding Milan

2.2.2 The definition of the alternatives

The definition of the alternatives is apparently much easier in this case than in the case of Como's tramway, given that the Navigli had a well-defined historical route. In fact, things are far from being obvious. Starting from a basic dichotomy, one can distinguish:

- "virtual " alternatives, in which the Navigli are not actually reopened, but made enjoyable by the public with visual aids (information signs, vintage photographs, pavings, etc. . .); see Figures 2.24 and 2.25 for some examples that have been proposed as a first step in the detailed project we are describing;
- "physical" alternatives, in which the Navigli are actually reopened.

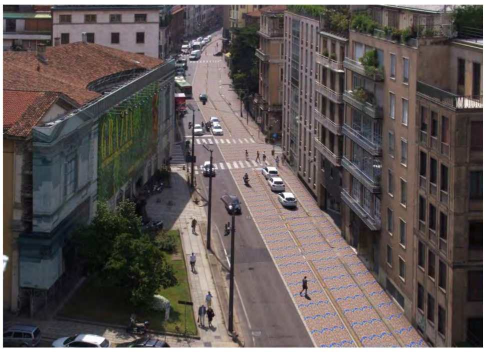


Figure 2.24: A first example of paving, with stone cubes, to revive the historical route of the Navigli

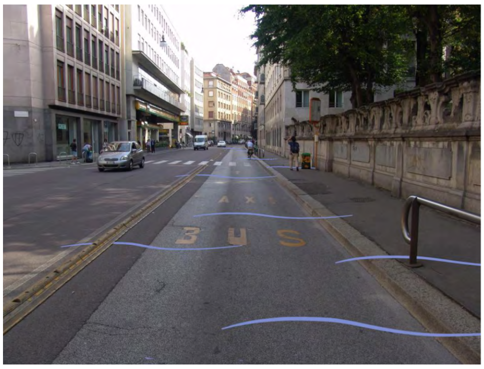


Figure 2.25: A second example of paving, with painted waves, to revive the historical route of the Navigli

One might imagine that all physical alternatives share the same route, that is fixed by history (see Figure 2.26). In fact, many alternatives are possible also from this point of view:

- a partial reopening, keeping some covered stretches;
- the complete reopening of the classic circle north-east-south, which includes a number of minor variations concerning:

  - Porta Nuova park (see Figure 2.27);
  - Cavour Square (see vol. 1 pag. 202 o 276 e 279/281 in the project documentation);
  - the Vallone Naviglio (see vol. 1 pagg. 337 e segg. in the project documentation);
  - the Viarenna basin (see vol. 1 pag. 206, 342, 344-350 in the project documentation);
- the reopening of the whole inner circle, also on the western side: some tecnical reasons make this alternative quite impractical (the main one is the interference between the original route of the Naviglio and the route of the underground line 2 near Sant'Ambrogio);
- additional works with respect to the historical route:
  - the Vettabbia canal (see vol. 2 pag. 249 in the project documentation);
  - the "Darsena 2" project, that is a brand new basin near the railway station of Porta Genova (see vol. 2 pag. 254 in the project documentation).

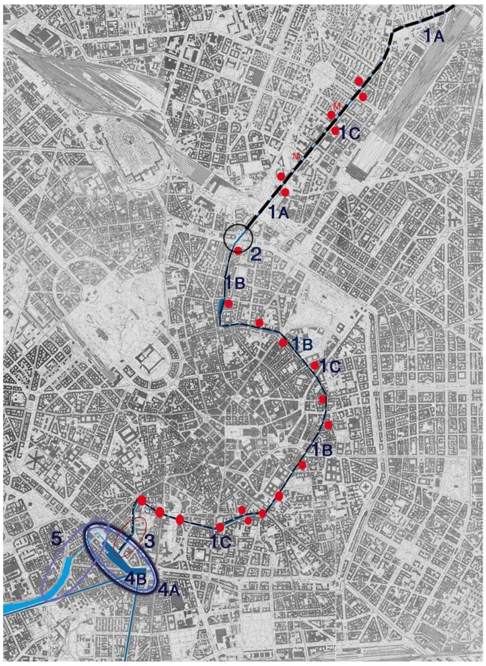


Figure 2.26: The historical route of the Navigli in Milan

Another fundamental element of alternative is navigability: the new canal could be open or close to boat trips. The elements of alternative (route, navigability and other possible ones) must then be combined to build the alternatives, removing the combinations that are impossible. For instances, the virtual alternatives or the alternatives adopting a very limited partial reopening obviously forbid navigability.

2.2.3 The definition of the scenarios

This point is not developed in the detailed project we are describing. We therefore make only a superficial analysis. The most likely elements of scenario should be identified in: Gli aspetti fondamentali riguardano

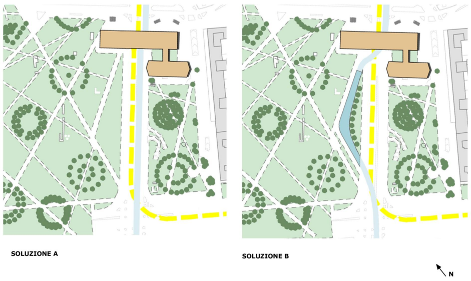


Figure 2.27: The two solutions proposed at the end of Melchiorre Gioia Street: A) full covering of the Naviglio; B) an open stretch of the Naviglio in Porta Nuova park

- the availability of public funding to build the canal and manage it in the long term;
- the (more or less strict) closure or the city centre to private cars;
- the construction of the underground line 4, which would provide an easy access on foot to most of the city centre even if private cars were no longer allowed and some bus lines were hampered by the canal (Figure 2.28 shows the areas covered by the current underground lines and by the future line 4); moreover, the underground works could be combined with the building of an underground canal under the route of the new Naviglio, that would guarantee the hydraulic in shorter time and would reduce the total cost[12]

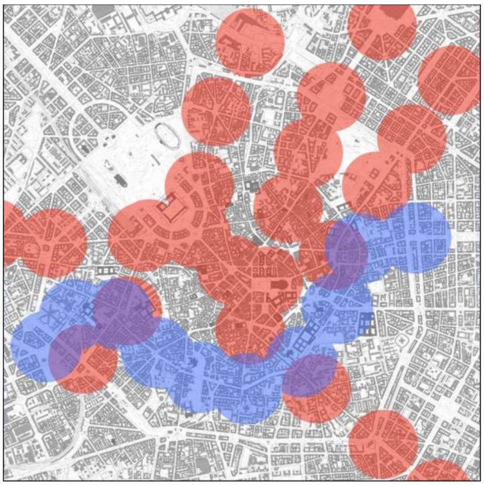


Figure 2.28: The areas at short walk distance from the current underground stations (in red) and those of the future line 4 (in blue)

[12] This is actually planned in the current works for the underground. Some supporters of the physical alternatives are afraid that the construction of an underground canal, by solving the continuity problem, could reduce the interest for a true reopening of the Navigli.

2.2.4 The definition and computation of the indicators

Once again, it is possible to proceed in a hierarchical way, by first identifying large sectors and then dividing them into more and more specific subsectors. To give a nonexhaustive list:

- the impact on pollution;
- the variation of the travel times to various sites of the town;
- the impact on traffic congestion (see vol. 2 pag. 79, 88 in the documentation of the project)
- the impact on the public transport (see vol. 2 pag. 80, 85in the documentation of the project), and in particular with tramways, that forbid slopes along the bridges;
- the construction costs (see vol. 1, da pag. 123 a pag. 138 in the documentation of the project);
- the hedonic impact on the real estate prices (see vol. 2 pag. 145-146 in the documentation of the project)
- the impact on the commercial activities (see vol. 2 pag. 150 in the documentation of the project).

Some of these impacts could be so negative to suggest a feedback loop on the definition of the alternatives, that is a correction and update of the alternatives through the introduction of devices to reduce those impacts. For example, reopening the canals would affect the car traffic along the canal and in some points it would nearly forbid the access of vehicles to the nearby buildings. This requires some local solutions to gain space, such as "balconies" over the canal (see vol. 2 pagg. 292-293-295 in the documentation of the project). As well, the bridges used by tramways must be at ground level, which implies, in order to allow the navigability, that the level of the water be kept low enough and that the boats be suitably designed to have a low height.

2.2.5 The time and space organization

Several indicators (in particular those concerning traffic congestion, accessibility, real estate prices, pollution, etc. . .) are naturally related to geographical positions. Some indicators refer to zones, whereas others, to linear stretches of the Naviglio (e.g., the real estate prices). The detailed project divides the route of the canal into 16 stretches.

Concerning the time phases, some of them correspond, somehow, to alternatives, meaning that it is possible to implement a specific phase of the project and then leave the following phases "frozen" for a long time, and even indefinitely, waiting for favourable conditions to complete them. For example, the virtual alternatives with pavings and signs can be seen as a preliminary phase for a global reopening (vol. 2, pag. 238). On the one hand, they allow to achieve a partial result at low cost and to gradually move towards the final result; on the other hand, however, they could increase the overall cost.

The construction of a double Naviglio, with an underground pipe running under the canal (vol. 2, pag. 239) is obviously an additional cost, but it also allows to reach in short time the important result of hydraulic continuity and to spread over the time the following phases, dividing the expense.

2.2.6 The definition of the stakeholders

The definition of the stakeholders is much similar to what has been described for the tramway of Como. The mayor of Milan can be considered as the decision-maker, but other subjects must be taken into account, such as the transportation companies (mainly ATM, which manages the bus and tramway lines; the railway lines are not affected by the project), the (environmental and category) associations, the citizens. Among the citizens, the owners of real estate along the route of the canal could be considered specifically.

2.2.7 The decision process

The evaluation and decision process should follow the same lines adopted for the tramway of Como. First, it is necessary to evaluate each indicator in each possible configuration, that is in each alternative and each scenario. This implies the application of suitable descriptive models. One could expect that:

- the "virtual" alternatives will be particularly cheap and have a negligible impact on traffic congestion and pollution, but will not offer any advantage with respect to hydraulic continuity (unless they are combined with the construction of an underground pipe), to the production of electricity and to the management of the water nappe, and they have a limited impact on tourism and commerce; the aesthetic impact is harder to evaluate;
- the "physical" alternatives will be much more expensive, that they will reach the desider objective for hydraulic continuity, electricity production and the management of the water nappe, that they will favour tourism, commerce and real estate prices and have a good aesthetic impact.

Moreover, one can expect that:

- the alternatives involving a navigable canal will have touristic, commercial and accessibility advantages, but a higher cost, because they require to manage a system of gatehouses to get over the height difference between the subsequent stretches of the canal (see vol. 2 pag. 8 in the project documentation);
- the alternatives involving a nonnavigable canal will have will be cheaper, but less advantageous for accessibility, commerce and tourism, even if they could have a better aesthetic impact, since a navigable canal requires to keep the level of the water low enough to let the boats pass under the bridges, and therefore reduces the view of the water from the shores.

After all indicators have been evaluated, the aggregation phase can start. In this phase, the preferential structure of all the stakeholder identified must be taken into account. This can be done in several ways, that will be described in the following chapters, and should conclude with an ordering of the alternatives, or at least the choice of one, as well as with information on the sensitivity of such an ordering or choice.

Chapter 3

Fundamental definitions and conceptual problems

Definition 1 *A* decision problem *is defined by a* 6*-uple*

$$P = (X, \Omega, F, f, D, \Pi)$$

where

- X *is the* feasible region*, that is the set of all possible* alternatives*, or* feasible solutions*;*
- Ω *is the* sample space*, that is the set of all possible* scenarios*, or* outcomes*;*
- F *is the* indicator space*, that is the set of all possible* impacts*;*
- $f : X \times \Omega \rightarrow F$ *is the* impact function*:*
  - *every pair* $(x, \omega) \in X \times \Omega$ *describes a* configuration *of the system on which a decision should be taken;*
  - *function* f *associates to each configuration* (x, ω) *of the system an impact* $f(x, \omega) \in F$*;*
  - *impact* $f(x, \omega)$ *describes all the features of the configuration* (x, ω) *that are relevant for the decision (costs, profits, quality levels, etc. . .);*
- D *is the set of all* decision-makers*, or* stakeholders*;*
- $\Pi : D \rightarrow 2^{F \times F}$ *is a function that associates to each decision-maker* $d \in D$ *a subset of impact pairs,* $\Pi(d) \subseteq F \times F$*; this subset is interpreted as a binary relation representing the* preference *of decision-maker* d*.*

The aim of the problem is to identify a solution $x^* \in X$ or a subset of solutions $X^* \subseteq X$ that the decision-makers consider satisfactory based on their preferences between the impacts $f(x^*, \omega)$ with $\omega \in \Omega$ and the other impacts $f \in F$.

3.1 Alternatives

The alternatives formally describe the events that are under the control of the decision-makers. Then, set X includes all the choices that are possible for the decision-makers, because they are compatible with all constraints on the system:

physical, economic, legal, organizational, etc.... The word “alternative” stresses the idea that exactly one solution must be applied. If it is possible to combine more than one, one should denote as “alternative” each possible combination.

In the following, we will assume that $X \subseteq \mathbb{R}^n$, that is we will describe the alternatives quantitatively as vectors of n real numbers (by convention, column vectors):

$$x = [x_1, \ldots, x_n]^T \qquad \text{with } x_i \in \mathbb{R} \text{ for } i = 1, \ldots, n$$

Each component x_i of alternative x is denoted as *element of alternative*, or *decision variable* and it is an elementary quantity that defines an aspect of the alternative.

The assumption that $X \subseteq \mathbb{R}^n$ is actually a limitation, because:

- it excludes infinite-dimension spaces X, where the available choices are functions depending on infinitely many parameters (e. g., we do not consider optimal control problems).

On the other hand, this assumption admits problems whose solutions are:

- families of functions depending on a finite number of parameters (e. g., X could include all polynomials of degree $\leq n - 1$);
- qualitatively described: it is enough to assign to the decision variable x_i a conventional numerical value for each degree of the qualitative scale (e. g., from 1 to 10);
- trivially enumerated: it is enough to associate a binary decision variable to each alternative and impose that exactly one of them be equal to 1 (in this way, alternative i^* corresponds to the *incidence vector* x, where $x_{i^*} = 1$ and $x_i = 0$ for all $i \neq i^*$).

Notice that, so far, the notation is purely descriptive: vector x is not used to make computations.

It is also possible to classify the decision problems based on the cardinality of X, dividing them into *discrete* or enumerable (that is, in one-to-one correspondence with a subset of $\mathbb{N}$) and *continuous.* The discrete problems can, on their turn, be divided into *infinite* and *finite.* To conclude, the finite problems are denoted as *combinatorial* if their cardinality, though finite, is exponential with respect to the number of decision variables n. This usually happens when the alternatives are obtained by combining a finite number of possible values for each decision variable. This classification influences the methods used to solve the problems: each category, in fact, has its own possible (or at least preferred) methods.

3.2 Scenarios

The scenarios formally describe the events that are out of the control of the decision-makers, but have a nonindifferent effect on the system. Then, set Ω defines all the uncontrollable events that could affect the system during the time horizon interested by the decision and that are relevant to the decision.

The scenarios also include every source of uncertainty on the behaviour of the system, such as:

- really exogenous phenomena, that affect the system and could be predictable by suitably enlarging the boundaries of the system (e. g., when flipping a coin, a perfect knowledge of the wind could perhaps allow to predict the outcome of the flip, but the behaviour of the wind is out of the system, and therefore the outcome of the coin is described as part of the scenario);

- effects of an incomplete or partially incorrect model of the system; in that case, the exogeneous variables are often called *disturbances* (e. g., when tuning a hydraulic valve, small losses can let some water pass even if the valve is close; these losses are not due to a exogeneous phenomenon, but they are included in the scenario to allow a simplified perfect valve model).

In the following, we assume that $\Omega \subseteq \mathbb{R}^r$, that is we describe the scenarios quantitatively as (column) vectors of r real numbers.

$$\omega = [\omega_1, \dots, \omega_r]^T \qquad \text{with } \omega_k \in \mathbb{R} \text{ and } k = 1, \dots, r$$

Each component ω_k of the scenario ω is denoted as *element of scenario*, or *exogenous variable*, and it is an elementary quantity that defines an aspect of the scenario.

As for the alternatives, this assumption excludes infinite-dimension scenarios, but admits scenarios described by parametric functions, qualitative scales, or simple enumerations.

3.3 Impacts and impact function

The impacts formally describe all aspects that are relevant for the decision. For instance, in order to regulate the temperature in a room, one can decide the position of the thermostat knob (that is part of the alternative) and we know that the temperature will also depend on the external temperature (that is part of the scenario), but the relevant point for the decision is the internal temperature of the room. This is the impact, which depends both on the position of the thermostat knob and on the external temperature.

Since the decision is taken based on the impact, two different configurations (x, ω) and (x', ω') with the same impact $(f(x, \omega) = f(x', \omega'))$ are absolutely indifferent. It is therefore fundamental for the model to explicitate all the elements that compose the impact, in order to avoid generating a false preference relation. This sounds obvious, but in practice it often happens that solutions considered acceptable, when implemented, prove to be mediocre because some elements of the impact had not been considered in the model. This is the case, for example, of public works stopped lawsuits or hampered by the public opinion: who modelled the decision process did not take into account aspects such as the reaction of the citizens, which is part of the impact. On the one hand, introducing new elements in the model makes it more complicated and costly to solve (data must be measured or estimated); on the other hand, an incorrect model leads to a wrong, and costly, solution.

In the following, we assume that $F \subseteq \mathbb{R}^p$, that is we describe the impacts quantitatively as (column) vectors of p real numbers:

$$f = [f_1, \dots, f_p]^T \qquad \text{with } f_l \in \mathbb{R} \text{ and } l = 1, \dots, p$$

with the usual warning about infinite-dimension impacts (excluded by this assumption) and qualitative or enumerated impacts (included). In some cases it will be necessary to refer to some components of the impact, but not all of them. In order to make the notation smoother, we will denote as $P = \{1, \dots, p\}$ the set of all their indices.

Each component $f_l \in \mathbb{R}$ of impact f is denoted as *indicator*, or *criterium*, *attribute*, *objective* (in the latter cases, usually, it is assumed to represent a benefit or a cost, so that the decision-makers prefer higher or lower impacts). Each indicator is an elementary quantity that defines an aspect of the impact. It should be clearly understandable by each decision-maker, univocal (i. e., non mixing up different aspects), irredundant (i. e., not repeating information already expressed by other indicators). Each indicator depends on the chosen alternative and on the scenario in a way that is computable through function $f_l = f_l(x, \omega)$.

While the other definitions introduced so far all derive from (some field of) the literature, the expression "impact" is not commonly used: it has been invented on purpose for this course, in order to distinguish the single indicators f_l from their combination f, wherever this is useful.

3.3.1 Representations of the impact in the finite case

In the finite case, the impact function f can be represented by an *evaluation matrix*, whose rows correspond to the configurations (x, ω) and whose columns correspond to the indicators f_l. In the example of Table 7.1, the problem has two solutions ($X = \{x', x''\}$) and two scenarios ($\Omega = \{\omega', \omega''\}$), therefore four configurations listed on the rows. It also has an impact consisting of $p = 4$ indicators, reported on the columns. Ogni cella della matrice fornisce il valore dell'indicatore di colonna nella configurazione di riga. Each cell of the matrix provides the value assumed by the column indicator in the row configuration.

$f(x, \omega)$	f_1	f_2	f_3	f_4	f_5	f_6
(x', ω')	10	5	40	20	24	180
(x', ω'')	16	10	60	16	20	190
(x'', ω')	20	6	23	8	17	230
(x'', ω'')	24	8	50	12	10	100

Table 3.1: An evaluation matrix representing the impact function for a finite problem with two alternatives, two scenarios and six indicators

In the finite case, the impact function f can also be represented by a *radar chart*, that is a two-dimension graph with an axis for each indicator f_l, representing a configuration (x, ω) with a p-vertex polygon which links the points on the axes that correspond to the values of the indicators in the given configuration (see Figure 3.1). Notice that each indicator has its own measure unit, independent from the other ones.

3.4 Decision-makers

A decision-maker is whoever takes part in a decision. The subjects who do not directly contribute to a decision, but anyway play a role in it with their preferences are often denoted as stakeholders.

The set of all decision-makers is just a simple finite set D, with no further characterization. In particular, it is not necessary to assume that it is embedded in a Euclidean space as the other sets introduced above.

Definition 1 does not explicitly discuss the relation between decision-makers and decision variables. In fact, this relation depends on the specific problem:

- in some problems, each decision-maker is fully in charge of a subset of decision variables, disjoint from the subsets controlled by the other ones;

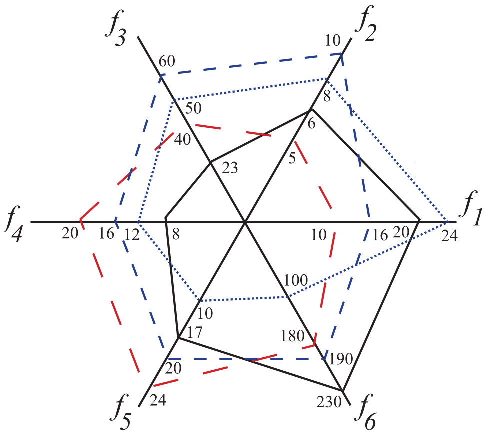


Figure 3.1: A radar chart representing the impact function for a finite problem with two alternatives, two scenarios and six indicators

- in other problems, all decision-makers control all decision variables and must coordinate before assigning their values.

Of course, intermediate problems can also be devised, but they have rarely (if ever) been discussed in the literature. From the point of view of notation, the distinction is not very important:

- if the decision variables are independently managed, usually the set of all feasible solutions is the cartesian product of the feasible solutions sets for the single decision-makers: $X = X^{(1)} \times \ldots \times X^{(d)} \times \ldots X^{(|D|)}$.

- if the decision variables are collectively managed, usually, the set of all feasible solutions is the intersection of the feasible solutions sets for the single decision-makers: $X = X^{(1)} \cap \ldots \cap X^{(d)} \cap \ldots X^{(|D|)}$.

In both cases, scenarios and impacts work as already described, and the preferences are defined independently on each decision-maker as described in the next section. The methods used to tackle and solve the two classes of problems will, on the contrary, be very different.

3.5 Preference relation

For each decision-maker, *preferring impact* $f \in F$ *to impact* $f' \in F$ means *considering acceptable the replacement of* f' *with* f, that is the replacement of the latter impact with the former. This corresponds to establishing a directed relation between pairs of impacts. In fact, function $\Pi : D \rightarrow 2^{F \times F}$ associates to each decision-maker d a subset $\Pi(d)$ of ordered pairs of impacts. To simplify the notation, this subset will be denoted as Π_d in the following. The meaning of this function is that:

- for each pair $(f, f') \in \Pi_d$, decision-maker d prefers f to f', that is accepts to exchange the latter impact f' with the former impact f;

- the list of all pairs in Π_d fully describes the preferences of decision-maker d.

Set Π_d can be interpreted as a *binary relation* on the impact set F:

$$\Pi_d = \{(f, f') \in F \times F : d \text{ prefers } f \text{ to } f'\}$$

When comparing two impacts f and f', if decision-maker d prefers f to f', instead of writing $(f, f') \in \Pi_d$, it is customary to use the notation[1]

$$f \preceq_d f'$$

In the following, every time there is a single decision-maker or a generic decision-maker is taken into account, therefore every time a single preference relation is considered instead of a whole family depending on D, we will suppress index d for the sake of simplicity, writing Π and $\preceq$ instead of Π_d and $\preceq_d$[2]

Moreover, given a binary relation Π, we define:

- the *complementary relation* $\bar{\Pi}$, that relates all the pairs that are not related by Π: $\bar{\Pi} = (F \times F) \setminus \Pi$, where, instead of $(f_1, f_2) \in \bar{\Pi}$, it is customary to write $f_1 \npreceq f_2$;
- the *inverse relation* Π^{-1}, that reverses the order of all the pairs that are related by Π: $\Pi^{-1} = \{(f_1, f_2) \in F \times F : (f_2, f_1) \in \Pi\}$, where, instead of $(f_1, f_2) \in \Pi^{-1}$, it is customary to write $f_1 \succeq f_2$.

3.5.1 Representations of the preference in the finite case

If the impact set F is finite, a generic preference relation Π can be represented by an *incidence matrix* of $|F|$ rows and columns (one for each impact), whose element (f, f') is equal to 1 when $f \preceq f'$, to 0 otherwise. Table 3.2 represents a preference relation defined on a set F with five impacts.

	f	$f^{'}$	$f^{''}$	$f^{'''}$	$f^{''''}$
f	1	0	1	1	1
$f^{'}$	0	1	1	1	1
$f^{''}$	0	0	1	1	1
$f^{'''}$	0	0	0	1	0
$f^{''''}$	0	0	1	1	1

Table 3.2: Incidence matrix for a preference relation Π on a set of five impacts

Alternatively, a preference relation on a finite set can be represented by a directed graph $G = (F, \Pi)$, whose nodes correspond to the impacts, whereas the arcs correspond to the pairs of impacts belonging to the relation. Figure 3.2 reports the graph associated to the same preference relation represented by the incidence matrix of Table 3.2.

3.5.2 Derived relations

Given a preference relation Π, three other relations with a relevant meaning can be derived:

[1] Economics textbooks sometimes use $f \succeq_d f'$; this is because they tend to think of the impact as a profit. We are adopting the opposite view. Of course, the choice is fully arbitrary.

[2] When denoting by Π a generic relation $\Pi_d \subseteq F \times F$, one must pay attention not to misunderstand it for the overall function $\Pi : D \to 2^{F \times F}$. The meaning of symbol Π should be clear from the context. The use of the same symbol is justified by the definition $\Pi_d = \Pi(d)$.

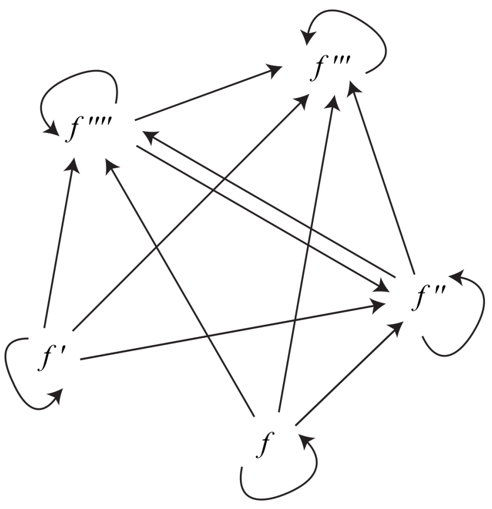


Figure 3.2: Graph associated to a preference relation Π on a set of five impacts

1. the *indifference* relation $\text{Ind}_\Pi = \Pi \cap \Pi^{-1} = \{(f, f^{'}) \in F \times F : f \preceq f^{'}, f^{'} \preceq f\}$ collects all pairs of impacts that the decision-maker accepts to exchange in both directions; when $(f, f') \in \text{Ind}_\Pi$, we will write more simply $f \sim f^{'}$;

2. the *strict preference* relation $\text{Str}_\Pi = \Pi \cap \bar{\Pi}^{-1} = \Pi \setminus \Pi^{-1} = \{(f, f^{'}) \in F \times F : f \preceq f^{'}, f^{'} \not\preceq f\}$ collects all pairs of impacts that the decision-maker accepts to exchange only in the indicated direction; when $(f, f') \in \text{Str}_\Pi$, we will write more simply $f \prec f^{'}$;

3. the *incomparability* relation $\text{Inc}_\Pi = \bar{\Pi} \cap \bar{\Pi}^{-1} = \{(f, f^{'}) \in F \times F : f \not\preceq f^{'}, f^{'} \not\preceq f\}$; collects all pairs of impacts that the decision-maker does not accept to exchange in any direction; when $(f, f') \in \text{Inc}_\Pi$, we will write more simply $f \bowtie f^{'}$.

The meaning of indifference and strict preference are rather obvious. Incomparability models the situations in which the decision-maker is not indifferent between two impacts, but is unable or unwilling to choose between them, refusing both the direct and the opposite exchange. Typical examples in the literature are moral choices such as killing one person or another, killing an animal or destroying an art masterpiece, etc...

Considering the relation represented in Figure 3.2, the indifference relation links impacts $f^{''}$ and $f^{''''}$, and every impact with itself; the strict preference relation links the other pairs between which the original preference relation existed; the incomparability relation links impacts f e $f^{'}$. The three derived relations are represented in the three parts of Figure 3.3.

For any pair of impacts (f, f'), four alternative cases may hold:

1. $f \prec f'$: f is strictly preferable to f';

2. $f' \prec f$: f' is strictly preferable to f;

3. $f \sim f'$: f e f' are indifferent;

4. $f \bowtie f'$: f e f' are incomparable.

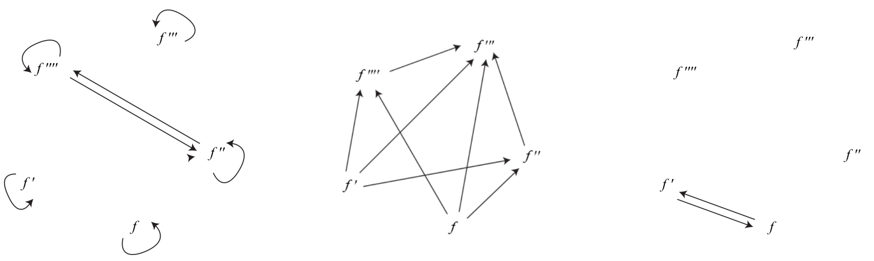


Figure 3.3: Graphs associated to the relations of indifference (on the left), strict preference (in the middle) and incomparability derived from the preference relation of Figure 3.2

Then, strict preference, its opposite, indifference and incomparability form a partition of $F \times F$. As well, strict preference and indifference form a partition of the original preference. This can be easily verified comparing the three parts of Figure 3.2 with Figure 3.3.

It is worth noticing that other texts adopt an alternative system of axioms, that is equivalent to the one introduced above. Instead of starting from a preference relation and deriving three auxiliary relations, one can start from two disjoint relations (subsets of $F \times F$), assign them the meaning of strict preference and indifference, and derive the two remaining relations: preference (in this case usually denoted as *weak* preference) will be defined as the union of strict preference and indifference, whereas incomparability will be defined as the set of all pairs belonging neither to preference nor to its opposite.

3.5.3 Properties of the preference relation

So far, a preference relation is a generic binary relation on F, that is a generic collection of subsets extracted from $F \times F$. Not all binary relations, however, express reasonable preference relations for a decision-maker. In principle, the preference relation should be:

- *realistic*, that is correctly modelling the aims of the decision-maker;
- *effective*, that is allowing algorithms that yield a safisfactory choice.

Quite surprisingly, the two needs of likelyhood and effectiveness often tend to conflict.

Both needs impose suitable additional properties on the preference relation, besides being binary. There is, however, a certain degree of freedom in the choice of which properties to impose, and several properties involve very important conceptual problems. In the end, the choice of the preference model should be motivated based on the aims of the model itself. First, let us introduce some possible properties; then, we will discuss the conceptual or techical problem they imply.

A binary relation Π on a set F is:

- *reflexive* when $I \subseteq \Pi$, where I is the identity relation, that is $f \preceq f$ for all $f \in F$: each impact is preferable to itself (weakly, that is actually indifferent);
- *antisymmetric* when $\Pi \cap \Pi^{-1} \subseteq I$, that is $f \preceq f^{'}$ e $f^{'} \preceq f \Rightarrow f = f^{'}$ for all $f, f^{'} \in F$: two impacts are indifferent only when they are exactly the same;

- *complete* when $\Pi \cup \Pi^{-1} = F \times F$, that is $f \npreceq f' \Rightarrow f' \preceq f$ for all $f, f' \in F$: for every pair of impacts, one is certainly preferable to the other (possibly, they are indifferent);
- *transitive* when $f \preceq f'$ e $f' \preceq f'' \Rightarrow f \preceq f''$ for all $f, f', f'' \in F$: if an impact is preferable to another, and this to a third one, the first impact is preferable to the third one;
- *simmetric* when $f \preceq_\Pi f' \Rightarrow f' \preceq f$ for all $f, f' \in F$; in general, a preference relation does not enjoy this property (strong preference is even asymmetric: $f \prec f' \Rightarrow f' \nprec f$ for all $f, f' \in F$), but we mention it because the derived relations of indifference and incomparability enjoy it by definition.

Reflexivity In general, it is undisputed that a preference relation should be reflexive: it would be rather meaningless to consider a decision-maker unable to compare an impact to itself. As a consequence, indifference is always reflexive, whereas incomparability and strict preference are never reflexive, that is they are irreflexive.

Antisymmetry Antisymmetry is a rather strong condition: it imposes that two impacts are indifferent only in the specific case in which they are perfectly identical. The preference model proposed by Pareto accepts this assumption (see Chapter 6), whereas other models reject it.

Completeness Completeness is also a rather strong condition: it excludes incomparability, that is it requires from the decision-maker to be always able to sort the impacts from the best one to the worst one, though allowing ties. A typical example is given by the final results of sports' leagues. The classical preference models based on objective functions implicitly accept this assumption, whereas other models reject it.

Transitivity If the preference is transitive, the derived relations of strict preference and indifference are also transitive by consequence. Transitivity is by far the most delicated and disputed property: some scholars consider it unavoidable, nearly a sort of equivalent for *rationality*; others question it. Several authors tried to prove by thought experiments that decision problems cannot be solved in a meaningful way without requiring transitive preferences. Other authors tried to dismantle the consequences of such experiments by showing experimentally that the preferences expressed by human beings in the real world are often not transitive. We shall discuss this point in detail in Section 3.5.5.

3.5.4 The most common preference relations

The most commonly used preference relations are:

- *preorder* relations: they enjoy *reflexivity* and *transitivity*;
- *partial order* relations: they enjoy *reflexivity*, *transitivity* e *antisymmetry*;
- *weak order* relations: they enjoy *reflexivity*, *transitivity* and *completeness*;
- *total order* relations: they enjoy *reflexivity*, *transitivity*, *antsymmetry* and *completeness.*

The total orders combine the properties of partial and weak orders. In that case, therefore, the decision-maker is able to sort the impacts from the best one to the worst one without ties. A typical example is given by real numbers.

Concerning the derived relations, when the preference is a preorder (reflexive and transitive), indifference is an *equivalence relation*: reflexive, transitive and symmetric. This relation partitions the impact set F into classes, each composed of reciprocally indifferent impacts.

3.5.5 Conceptual problems on the transitivity assumption

The transitivity assumption deserves a detailed discussion. Transitivity is the most common assumption for a preference relation, and is often interpreted as an equivalent for rationality. According to the supporters of this thesis, rejecting transitivity leads to paradoxical effects.

First, let us show with an example what it means to reject transitivity. Let us assume that a decision-maker, facing the choice between A and B (an apricot and a banana, for example), state to prefer A and, facing the choice between B and C (a banana and a coconut, for example), state to prefer B. A transitive decision-maker should necessarily prefer A to C. A nontransitive decision-maker, on the contrary, could state to be unable to choose between A and C (the apricot and the coconut), if not even to strictly prefer C to A. Let us consider these two cases and their paradoxical implications.

1) Incomparability contradicted by chains of exchanges If the decision-maker states to prefer A to B and B to C, and yet to be unable to choose between A and C, it would be quite easy to induce him or her to give up C (the coconut) in favour of A (the apricot) through a chain of two exchanges using B (the banana) as an intermediate step. This suggests that the decision-maker is actually able to choose between the two options, contrary to what stated. According to its proponents, this thought experiment should prove that such a kind of intransitivity is not very convincing.

2) Money pump If, on the contrary, the decision-maker states to prefer A to B and B to C, and to strictly prefer C to A, this could be even exploited to make him or her accept a cyclic chain of exchanges and even lose money. One could give B to the decision-maker instead of C, then A instead of B and finally, exploiting the strict preference, C instead of A plus a sum of money small enough not to reverse the preference. At that point, the decision-maker would be back to the starting point (the coconut C) having lost money for nothing. This is what the economists use to call a *money pump*. This thought experiment presents various issues to analyse in depth. In fact, it introduces the concept of money, that is it complicates the impact including in it an additional element, or indicator. Moreover, it assumes that money could be split into quantities sufficiently small to guarantee that adding such a quantity of money to a one of the compared impacts might not modify the direction of the preference between them. These assumptions must be carefully evaluated.

The supporters of the possibility to use nontransitive preferences propose on the one hand alternative thought experiments to give them a reasonable interpretation, on the other hand practical experiments showing that human behaviour is nontransitive.

1) The influence of time The different impacts compared by the preference relation are usually interpreted as nontemporal, that is fixed once for all. In practice, a decision often consists of several choices made in different instants along time; the preference could change over time, giving rise to violations of transitivity. For example, perhaps today the decision-maker prefers A to B, tomorrow B to C and the day after C to A. This makes the *money pump* cycle not only possible, but likely, so much that it could be the basis of financial activities, such as the currency market. These violations of transitivity can be arranged in a transitive model by assigning to the impacts a time index, and defining the preference relation between the extended impacts, that is the impact-time pairs. This, however, would make the preference model much more complicated. As the aim of the model is to support decisions, using a very complicated model is in general not desirable.

2) Limited discriminating capacity The decision-maker is not actually able to perfectly discriminate between all impacts. There are extremely similar impacts between which the decisin-maker is indifferent because unable to discriminate them. However, a chain of such impacts could connect very different impacts. For example, let us assume that the decision-maker prefer unsugared coffee. Facing the choice between two cups, respectively with n and $n+1$ sugar grains, the decision-maker could be unable to discriminate, and therefore could consider them as indifferent. Through a chain of cups with a progressively increasing number of sugar grains, it is possible to prove by transitivity that the decision-maker is indifferent between an unsugared cup of coffee and a heavily sugared one. In short, either one assumes an infinite capacity of discrimination or one gives up preference transitivity.

3.5.6 *Framing* effects*

What is even worse, the way in which impacts are presented can heavily affect the preferences declared by the decision-makers, even in the case of completely identical impacts.

In a famous psychology experiment, a sample of individuals was asked to impersonate the Secretary of Health of a country in which an epidemic has outbroken and to evaluate the impacts of three possible alternatives:

1. do nothing, and let 600 people die (impact f);

2. enact programme A, saving everybody with probability 1/3 and nobody with probability 2/3 (impact $f^{'}$);

3. enact programme B, saving 200 people (impact $f^{''}$).

The majority of the sample chose programme B, showing that they preferred the corresponding impact. The expected impact is statistically the same of programme A, but committing to probabilistic games is not the same as choosing a certain outcome, so it is reasonable that some decision-makers could have a strict preference such as $f^{''} \prec f^{'}$[3]. Another sample of decision-makers, considered equivalent to the former, was asked to evaluate the impacts of the three following alternatives:

1. do nothing, and let 600 people die (impact f);

[*] This section presents advanced concepts, that are not part of the course syllabus.

[3] The opposite is also reasonable, only less common in practice. We will discuss this dealing with the decision under risk. Now the probabilistic aspect is irrelevant for the experiment and only helps to hide its fundamental aspect.

2. enact programme A, saving everybody with probability 1/3 and nobody with probability 2/3 (impact $f^{'}$);
3. enact programme B, letting 400 people die (impact $f^{''}$).

Everybody can see that the three alternatives are exactly the same as before, but the third alternative has been described with different words ("letting die" instead of "saving"). The majority of the second sample chose programme A, thus expressing the preference $f^{'} \prec f^{''}$. The context, that is the idea of "saving" people with some probability, instead of "letting die" people with certainty, modifies the perception of the impacts.

3.5.7 Preference models: faithful or effective?

We can now draw some conclusions on the problem of how to model the preferences of the decision-makers. The choice largely depends on the purpose of the model itself. The aim to have a realistic description of these preferences leads to the idea of a faithful model, whereas the aim to have an effective description in order to reach a useful choice leads to the idea of an effective model, that constrains the freedom of the decision-maker, "teaching" him or her how to take a good decision. From time to time, it will be preferable to favour either approach, but in general both the extremes are dangerous.

3.6 Exercises

Exercise 1

Given a decision problem with impact set $F = \{a, b, c, d, e, f\}$ and preference relation $\Pi = \{(a, a), (a, b), (a, c), (a, d), (b, b), (b, c), (b, d), (c, b), (c, c), (c, d), (d, d), (e, b), (e, c), (e, d), (e, e), (f, b), (f, c), (f, d), (f, f)\}$

a) list the main properties enjoyed by Π;

b) state whether Π is an order relation of some kind;

c) build the derived relations of indifference Ind_Π, strict preference Str_Π and incomparability Inc_Π.

Solution

Figure 3.4 reports the graph representation of relation Π.

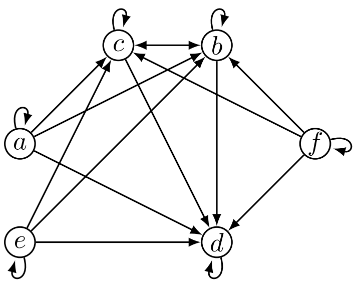


Figure 3.4: Graph representation of the preference relation proposed by Exercise 1

From the graph, it is apparent that Π is reflexive (all impacts have a self-loop). It is not antisymmetric, because impacts b and c are indifferent, but not identical. It is not complete, because impacts a, e and f are incomparable. It is transitive, because every triplet of impacts with a pair of weak preferences corresponds to a weak preference between the two extreme impacts. Consequently, Π is a preorder.

The derived relations are:

- $\text{Ind}_\Pi = \{(a, a), (b, b), (c, c), (d, d), (e, e), (f, f), (b, c), (c, b)\}$
- $\text{Str}_\Pi = \{(a, b), (a, c), (a, d), (b, d), (c, d), (e, b), (e, c), (e, d), (f, b), (f, c), (f, d)\}$
- $\text{Inc}_\Pi = \{(a, e), (a, f), (e, a), (e, f), (f, a), (f, e)\}$

Exercise 2

Given a decision problem with impact set $F = \{a, b, c, d, e\}$ and preference relation $\Pi = \{(a, a), (a, b), (a, d), (a, e), (b, b), (c, a), (c, b), (c, c), (c, d), (c, e), (d, b), (d, d), (e, a), (e, b), (e, d), (e, e)\}$

a) list the main properties enjoyed by Π;

b) state whether Π is an order relation of some kind and what this allow to assume about the decision problem;

c) build the derived relations of indifference Ind_Π, strict preference Str_Π and incomparability Inc_Π.

Solution

Figure 3.5 reports the graph representation of relation Π.

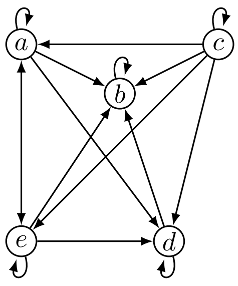


Figure 3.5: Graph representation of the preference relation proposed by Exercise 2

From the graph, it is apparent that Π is reflexive (all impacts have a self-loop). It is not antisymmetric, because impacts a and e are indifferent, but not identical. It is complete, because no impacts are incomparable. It is transitive, because every triplet of impacts with a pair of weak preferences corresponds to a weak preference between the two extreme impacts. Consequently, Π is a weak order.

The derived relations are:

- $\mathrm{Ind}_\Pi = \{(a,a),(a,e),(b,b),(c,c),(d,d),(e,a),(e,e)\}$
- $\mathrm{Str}_\Pi = \{(a,b),(a,d),(c,a),(c,b),(c,d),(c,e),(d,b),(e,b),(e,d)\}$
- $\mathrm{Inc}_\Pi = \emptyset$

Exercise 3

Given a decision problem with impact set $F = \{a,b,c,d,e\}$ and preference relation $\Pi = \{(a,a),(a,b),(a,c),(a,d),(a,e),(b,b),(b,c),(b,d),(b,e),(c,b),(c,c),(c,d),(c,e),$ $(d,d),(e,e)\}$

a) list the main properties enjoyed by Π;

b) state whether Π is an order relation of some kind and what this allow to assume about the decision problem;

c) build the derived relations of indifference Ind_Π, strict preference Str_Π and incomparability Inc_Π.

Solution

Figure 3.6 reports the graph representation of relation Π.

From the graph, it is apparent that Π is reflexive (all impacts have a self-loop). It is not antisymmetric, because impacts b and c are indifferent, but not identical. It is not complete, because impacts d and e are incomparable. It is transitive,

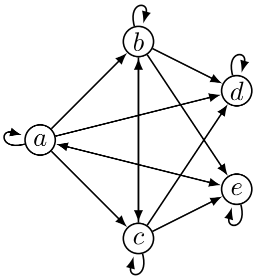


Figure 3.6: Graph representation of the preference relation proposed by Exercise 3

because every triplet of impacts with a pair of weak preferences corresponds to a weak preference between the two extreme impacts. Consequently, Π is a preorder.

The derived relations are:

- $\text{Ind}_\Pi = \{(a,a),(b,b),(c,c),(d,d),(e,e),(b,c),(c,b)\}$
- $\text{Str}_\Pi = \{(a,b),(a,c),(a,d),(a,e),(b,d),(b,e),\ (c,d),(c,e)\}$
- $\text{Inc}_\Pi = \{(d,e),(e,d)\}$

Exercise 4

Given a decision problem with impact set $F = \{a,b,c,d,e\}$ and preference relation $\Pi = \{(a,a),(a,b),(a,c),(b,b),(b,c),(d,d),(e,a),(e,b),(e,c),(e,d),(e,e)\}$:

a) list the main properties enjoyed by Π;

b) state whether Π is an order relation of some kind;

c) build the derived relations of indifference Ind_Π, strict preference Str_Π and incomparability Inc_Π.

Solution

Figure 3.7 reports the graph representation of relation Π.

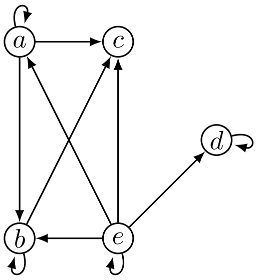


Figure 3.7: Graph representation of the preference relation proposed by Exercise 4

From the graph, it is apparent that Π is not reflexive, because impact c has no self-loops. It is antisymmetric, because the impacts are indifferent only to themselves. It is not complete, because d is incomparable with a, b and c. Is it transitive? Transitivity requires every triplet of impacts with a pair of weak preferences to correspond to a weak preference between the two extreme impacts. Since there is no such triplet, this property holds. Consequently, Π is not an order of any kind, even though it would be a partial order if a self-loop on impact c were added.

The derived relations are:

- $\text{Ind}_\Pi = \{(a,a),(b,b),(d,d),(e,e)\}$
- $\text{Str}_\Pi = \{(a,b),(a,c),(b,c),(e,a),(e,b),(e,c),(e,d)\}$
- $\text{Inc}_\Pi = \{(a,d),(b,d),(c,c),(c,d),(d,a),(d,b),(d,c)\}$

Exercise 5

Given a decision problem with the preference relation Π described by the following table:

Π	a	b	c	d	e
a	1	0	1	0	0
b	1	1	0	1	1
c	0	0	1	1	1
d	0	0	0	1	0
e	0	0	1	0	1

a) list the main properties enjoyed by Π;

b) state whether Π is an order relation of some kind;

c) build the derived relations of indifference Ind_Π, strict preference Str_Π and incomparability Inc_Π.

Solution

Figure 3.8 reports the graph representation of relation Π.

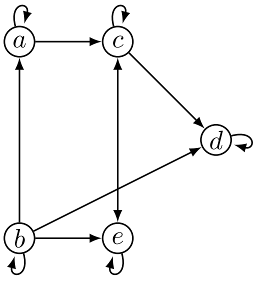


Figure 3.8: Graph representation of the preference relation proposed by Exercise 5

From the graph, it is apparent that Π is reflexive (all impacts have a self-loop). It is not antisymmetric, because impacts c and e are indifferent, but not identical. It is not complete, because impacts a, d and e are incomparable. It is not transitive, because (among other triplets) the weak preferences (b, a) and (a, c) hold, whereas (b, c) does not hold. Consequently, Π is not an order of any kind.

The derived relations are:

- $\text{Ind}_\Pi = \{(a, a), (b, b), (c, c), (d, d), (e, e), (c, e), (e, c)\}$
- $\text{Str}_\Pi = \{(a, c), (b, a), (b, d), (b, e), (c, d)\}$
- $\text{Inc}_\Pi = \{(a, d), (d, a), (a, e), (e, a), (b, c), (c, b), (d, e), (e, d)\}$

Exercise 6

Given a decision problem with impact set $F = \{a, b, c, d, e\}$ and preference relation $\Pi = \{(a, a), (a, b), (a, c), (a, e), (b, b), (b, d), (b, e), (c, b), (c, c), (c, d), (c, e), (d, a), (d, d),$ $(d, e), (e, e)\}$:

a) list the main properties enjoyed by Π;

b) state whether Π is an order relation of some kind;

c) build the derived relations of indifference Ind_Π, strict preference Str_Π and incomparability Inc_Π.

Solution

Figure 3.9 reports the graph representation of relation Π.

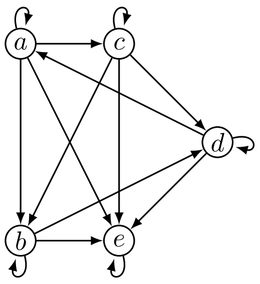


Figure 3.9: Graph representation of the preference relation proposed by Exercise 6

From the graph, it is apparent that Π is reflexive (all impacts have a self-loop). It is antisymmetric, the impacts are indifferent only to themselves. It is complete, because no impacts are incomparable. It is not transitive, because the preferences (a, c), (c, d) and (d, a) form a circuit of strict preferences, whereas (a, d) does not hold. Consequently, Π is not an order of any kind.

The derived relations are:

- $\text{Ind}_\Pi = \{(a, a), (b, b), (c, c), (d, d), (e, e)\}$
- $\text{Str}_\Pi = \{(a, b), (a, c), (a, e), (b, d), (b, e), (c, b), (c, d), (c, e), (d, a), (d, e)\}$
- $\text{Inc}_\Pi = \emptyset$

Exercise 7

Given a decision problem with impact set $F = \{a, b, c, d, e\}$ and preference relation $\Pi = \{(a,a), (b,a), (b,b), (b,c), (b,d), (c,a), (c,c), (c,d), (d,a), (d,d), (e,a), (e,b), (e,c), (e,d), (e,e)\}$:

a) list the main properties enjoyed by Π;

b) state whether Π is an order relation of some kind;

c) build the derived relations of indifference Ind_Π, strict preference Str_Π and incomparability Inc_Π.

Solution

Figure 3.10 reports the graph representation of relation Π.

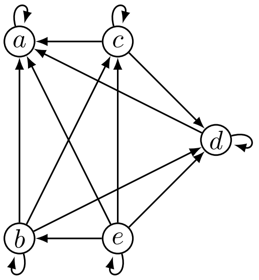


Figure 3.10: Graph representation of the preference relation proposed by Exercise 7

From the graph, it is apparent that Π is reflexive (all impacts have a self-loop). It is antisymmetric, the impacts are indifferent only to themselves. It is complete, because no impacts are incomparable. It is transitive, because every triplet of impacts with a pair of weak preferences corresponds to a weak preference between the two extreme impacts. Consequently, Π is a total order.

The derived relations are:

- $\text{Ind}_\Pi = \{(a,a), (b,b), (c,c), (d,d), (e,e)\}$
- $\text{Str}_\Pi = \{(b,a), (b,c), (b,d), (c,a), (c,d), (d,a), (e,a), (e,b), (e,c), (e,d)\}$
- $\text{Inc}_\Pi = \emptyset$

Exercise 8

Given a decision problem with impact set $F = \{a, b, c, d, e\}$ and preference relation $\Pi = \{(a,a), (b,b), (c,a), (c,b), (c,c), (d,a), (d,b), (d,d), (d,e), (e,e)\}$:

a) list the main properties enjoyed by Π;

b) state whether Π is an order relation of some kind;

c) build the derived relations of indifference Ind_Π, strict preference Str_Π and incomparability Inc_Π.

Solution

Figure 3.11 reports the graph representation of relation Π.

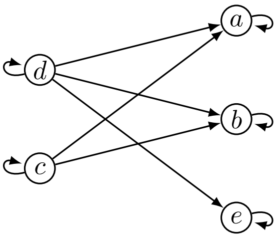


Figure 3.11: Graph representation of the preference relation proposed by Exercise 8

From the graph, it is apparent that Π is reflexive (all impacts have a self-loop). It is antisymmetric, because the impacts are indifferent only to themselves. It is not complete, because impacts a, b and e are incomparable, c and d are incomparable, c and e are incomparable. Is it transitive? Transitivity requires every triplet of impacts with a pair of weak preferences to correspond to a weak preference between the two extreme impacts. Since there is no such triplet, this property holds. Consequently, Π is a partial order.

The derived relations are:

- $\mathrm{Ind}_\Pi = \{(a,a),(b,b),(c,c),(d,d),(e,e)\}$
- $\mathrm{Str}_\Pi = \{(c,a),(c,b),(d,a),(d,b),(d,e)\}$
- $\mathrm{Inc}_\Pi = \{(a,b),(a,e),(b,a),(b,e),(c,d),(c,e),(d,c),(e,a),(e,b),(e,c)\}$

Part II

Basic decision models

Chapter 4

Structured preferences

In this chapter we start to describe how the concepts introduced in the previous one can be applied to the choice of a solution for a decision problem. We consider the particular case of a problem with a single scenario (therefore deterministic), and a single decision-maker: $|\Omega| = |D| = 1$. Additionally, we also require the preference relation of the single decision-maker to be a weak order (reflexive, transitive and complete). We will first show that the first two conditions allow to introduce a well-posed definition of solutions that can be justifiably selected, and the third one improves upon it by allowing the choice of a single solution. Then, we will discuss the relation between decision problems with all three conditions and classical optimisation problems. We will describe some preference models which are weak orders. These models make very strict assumptions on the preference of the decision-maker. The final section deals with the problem of deriving a value function (if one exists) from the preference relation. The technique it presents has a remarkable degree of sophistication and is the basis of the classical economical theory, but it requires rather strong assumptions and a very laborious process.

4.1 Dominance between solutions

If the problem has a single scenario and a single decision-maker ($|\Omega| = |D| = 1$), the impact only depends on the chosen alternative ($f : X \to F$) and the preference relation induces a relation between solutions: namely, a solution dominates another one when the impact of the former is preferable to the impact of the latter.

$$x \preceq x' \Leftrightarrow f(x) \preceq f(x') \qquad x, x' \in X$$

Definition 2 *We denote as* dominated solution *a solution $x \in X$ such that $\exists x' \in X : f(x') \prec f(x)$,* nondominated solution *any other solution. Finally, we denote as $X^* \subseteq X$ the set of all nondominated solutions.*

Let us assume that the preference Π be a preorder, that is a reflexive and transitive relation. It can be proved that the induced relation on the solutions also enjoys these properties. We observe that also completeness is preserved when moving from Π to the induced relation on X, whereas antisymmetry could not be preserved, because the same impact can correspond to different solutions, and therefore two indifferent impacts are identical, but two indifferent solutions are not necessarily identical.

It is reasonable to think that, when an impact admits other impacts that are strictly preferable, all the solutions that generate such an impact can be rejected, limiting the problem to the solutions whose impacts do not admit strictly preferable impacts. We consider a strict preference and not a weak one, because we do not want to reject two solution having indifferent impacts (at least one of them should be saved). Hence, *a rational decision-maker always chooses a nondominated solution.*

Theorem 1 *If the solution set is finite and nonempty ($X \neq \emptyset$) and the preference relation Π is reflexive and transitive, the nondominated solution set X^* is nonempty.*

This is important because it guarantees that the decision problem admits reasonable solutions. In some cases, X^* will contain a single solution or several solutions reciprocally indifferent; in other cases, it will include incomparable solutions. In general, therefore, determining set X^* simplifies the problem, but does not solve it completely. When X is infinite, every solution could admit a strictly dominating one, so that the useful property is no longer verified. However, this would imply an infinite chain of ever-improving solutions. Since in practice that is quite unlikely, it remains reasonable to restrict the search to nondominated solutions whenever the preference relation is reflexive and transitive.

Identification of the nondominated solutions

If X is a finite set, it is possible to build the *strict preference graph*, whose nodes correspond to solutions, while the arcs correspond to solution pairs whose impacts are related by a strict preference. In this graph there are no indifferent pairs. The nondominated solutions correspond to *nodes with no ingoing arc*. In order to identify them, it is enough to scan the arc, determine the indegree for each node and find the nodes with zero indegree. Let $O(\gamma)$ be the complexity of computing the preference between two given impacts (not necessarily constant, given that th two impacts must be first computed and then compared component by component). The overall complexity of the search for nondominated solutions is $O(\gamma\,|X|^2)$. Of course, this procedure cannot be applied to infinite sets and could be impractical in the case of combinatorial sets.

Example 5 *Let us assume to make a trip and to decide which means of transport to use. There are five alternatives: $X = \{Airplane, Car, Coach, Train, Taxi\}$ and the preference relation is $\Pi = \{(Airplane, Airplane), (Airplane, Coach), (Airplane, Taxi), (Car, Car), (Car, Coach), (Car, Taxi), (Car, Train), (Taxi, Taxi), (Train, Taxi), (Train, Train)\}$. By definition, the preference relation should consist of pairs of impact, not of alternatives, but, assuming that each alternative have its own impact, different from the other ones, for the sake of simplicity we can use the same name for the alternative and the corresponding impact. The strict preference graph is reported in Figure 4.1: there are two nondominated solutions, that form set $X^* = \{Car, Airplane\}$.*

4.2 Weak orders and value functions

Theorem 2 *If the solution set is finite and nonempty ($X \neq \emptyset$) and the preference relation Π is reflexive, transitive and complete, the nondominated solution set X^* is nonempty and consists of reciprocally indifferent solutions.*

This allows to choose any of such solutions as the overall solution of the problem. That is a very satisfactory result, as the main aspiration of a decision-maker is actually to “make the right choice”. For infinite sets, the same remark made above still applies.

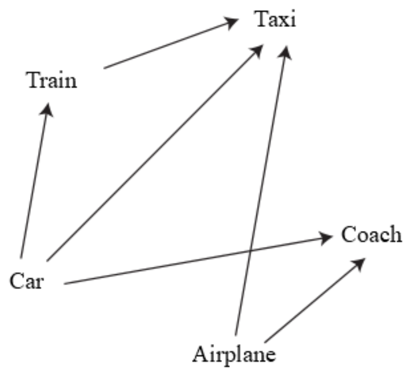


Figure 4.1: Strict preference graph for Example 5

It has already been stated that weak orders allow to sort the impacts on a line, with possible ties, as if each impact were associated to a degree. If a degree could actually be defined for each impact, the choice of a good impact would correspond to the choice of the highest or lowest degree, that is to the maximization of a utility function, or the minimization of a cost function. This would open the way to the application of optimization algorithms to solve decision problems.

First, let us verify that the existence of a degree actually implies the properties of weak orders. Then, we will investigate whether the opposite also holds, that is if any decision-maker whose preference relation is a weak order can be handled as if he or she implicitly assigned a degree to each impact. We will see that things are a bit more complicated, but not very far from this situation.

Definition 3 *A* value function *is a function $v : F \to \mathbb{R}$ that associates a real value to each impact. Function v is* consistent *with preference relation* Π *when*

$$f \preceq f' \Leftrightarrow v(f) \geq v(f') \text{ per ogni } f, f' \in F \tag{4.1}$$

or, equivalently

$$\Pi = \{(f, f') \in F \times F : v(f) \geq v(f')\}$$

The value functions are also called *utility functions*: the former expression is more used in mathematics, the latter in economics.

It is easy to see that, if a preference relation admits a consistent value function, the derived relations of indifference and strict preference correspond to identity and strict inequality between the values of the value function. The relation with real numbers guaranteed by a consistent value function induces strong properties on the preference relation.

Theorem 3 *If a preference relation admits a consistent value function, then the preference relation is a weak order.*

The proof is simple, and it is a good exercise to verify one's understanding. The point is to show that the preference enjoys reflexivity (because Π includes pair (f, f) for all $f \in F$), transitivity (because for all triplets $f, f^{'}, f^{''} \in F$ such that relation Π include pairs $(f, f^{'})$ and $(f^{'}, f^{''})$, it also includes pair $(f, f^{''})$), and completeness (because for each pair $(f, f^{'})$ not included in Π, pair $(f^{'}, f)$ belongs to it).

The fact that a value function implies a weak order is interesting, but in practice not very useful, because in a complicated decision problems the decision-maker usually has no known explicit value function. On the contrary, it is necessary to interview the decision-maker in order to investigate the properties of the preference relation. It is nearly certain that it will be reflexive, and probably it will be transitive. If the decision-maker has a clear idea, the preference relation could even be complete. In that case, it would be nice to deduce that there exists a consistent value function (though still unknown), because the decision problem could be reduced to the maximization of the value function. As this is not always true, it is important to understand when it is not.

4.2.1 Weak orders not reducible to a consistent value function

The lexicographic order is an example of weak order relation (actually, of a strong order relation) which does not admit a consistent value function. Let us consider the simplest case, that is the two-dimensional one: the impacts are vectors with two real components ($F = \mathbb{R}^2$) and the preference relation is defined as:

$$\begin{bmatrix} f_1 \\ f_2 \end{bmatrix} \preceq \begin{bmatrix} f'_1 \\ f'_2 \end{bmatrix} \Leftrightarrow f_1 < f'_1 \text{ or } f_1 = f'_1 \text{ e } f_2 \leq f'_2$$

that is, the decision-maker prefers the impact with the smaller first component, for any value of the second one; in case of a tie (and only in that case), the decision-maker prefers the impact with the smaller second component. It can be proved that this relation does not admit any value function $v(f)$ satisfying Condition (4.1), that is assigning to each impact in F a real value such that the preference between two impacts correspond to an inequality between the function values.

The intuitive reason is that lexicographic order requires total incommensurability between the two components: no improvement of the second can compensate for a worsening, even very small, of the first. A value function depending on both components, on the contrary, would allow some reciprocal compensation. We will see that, if the first component assumed discrete values, instead of continuous ones, it would become possible to build a consistent value function respecting this incommensurability, because there no variations of the first component would be small enough to be compensated by large variations of the second.

4.2.2 Weak orders reducible to a consistent value function*

Most weak orders enjoy additional technical conditions that allow to build a consistent value function. In order to introduce such conditions, the following auxiliary definition is required.

Definition 4 *Let F be a set and Π a weak order on F. A subset $\bar{F} \subseteq F$ is* dense *in F when every pair $f, f' \in F \setminus \bar{F}$ such that $f \prec f'$ admits an element $\bar{f} \in \bar{F}$ such that $f \prec \bar{f} \prec f'$.*

In other words, a subset is dense in the set if it contains, for every pair of elements of the given set, an element with intermediate preference.

Theorem 4 *A weak order Π on F admits a consistent value function if and only if there exists a subset $\bar{F} \subseteq F$ enumerable and dense in F.*

*This section provides advanced concepts, that are not part of the course syllabus.

The intuitive idea of this theorem is that it is possible to build a value function with rational values on $\bar{F}$ and to extend it to the whole of F by exploiting the density property, according to which any pair of impacts in F has an intermediate impact in $\bar{F}$, for which the value function has a value.

Let us consider again the example of the lexicographic order under the new light shed by the concept of dense and enumerable subset. If $F = \mathbb{R}^2$, no subset $\bar{F}$ can be at the same time enumerable and dense in F. By contradiction, let us assume that such a subset $\bar{F}$ exist: the projection of $\bar{F}$ on axis f_1 would be a set of real numbers formed by the first components of all the vectors in $\bar{F}$. Since $\bar{F}$ is enumerable, since every element has a projection and some projections could coincide, the set of all projections is also enumerable. This means that there are infinite values of f_1 that do not belong to the set of projections. Let $\tilde{f}_1$ be one of those values, and let $\tilde{f}_2$ and $\tilde{f}'_2$ be two real values chosen *ad libitum*. With these values, let us build the two impacts $[\tilde{f}_1\ \tilde{f}_2]^T$ and $[\tilde{f}_1\ \tilde{f}'_2]^T$, which by definition belong to $F \setminus \bar{F}$. If $\bar{F}$ is dense, it must contain an element with intermediate preference between the two impacts. By definition of lexicographic order, all elements with intermediate preference must have the first component equal to $\tilde{f}_1$ and the second between $\tilde{f}_2$ and $\tilde{f}'_2$. On the other hand, by construction no impact in $\bar{F}$ can have its first component equal to $\tilde{f}_1$. By contradiction, there exist no subset that is both enumerable and dense in F, and therefore there exists no value function on F consistent with Π.

By slightly changing the definition of F, this becomes possible. Let us assume that the impact set be $F = \mathbb{N} \times \mathbb{R}$, that is that the first component of the impact be an integer. In that case, it is possible to build enumerable subsets dense in F, and therefore to find value functions consistent with the lexicographic order. An example is function $v(f_1, f_2) = f_1 + \tanh(f_2)/2$: term f_1 has integer values, whereas term $\tanh(f_2)/2$ maps the real values of f_2 into values limited in the open interval $(-1/2, +1/2)$.

4.2.3 Multiplicity of the consistent value functions

The existence of a way to build a consistent value function does not imply its unicity.

Theorem 5 *Given a value function $v : F \to \mathbb{R}$, consistent with a preference relation Π on F, for any strictly increasing function $\phi : \mathbb{R} \to \mathbb{R}$ the composite function $\phi(v(\cdot))$ is also consistent with Π.*

Therefore is a weak order admits a consistent value function, it admits infinite equivalent ones, that associate to the same impacts different values, but sorted in the same way. Table 4.3 reports two value functions that are consistent with the preference described in Example 5: the former is Borda's count, whereas the latter is completely independent, but corresponds to the same preference relation, the same graph (the one in Figure 4.2) and the same ordering of the alternatives.

4.3 The finite case

It is always possible to solve the finite case by introducing and maximising (by exhaustive enumeration) a value function known as *Borda count*[1]:

$$v(f) = |\{f' \in F : f \preceq f'\}| \qquad f \in F \tag{4.2}$$

[1] From the name of Jean-Charles de Borda (1733-1799), French mathematician, physicist and admiral, who invented it.

X	$v(f(x))$	$v'(f(x))$
Airplane	4	20
Car	5	30
Coach	3	10
Taxi	1	5
Train	3	10

Table 4.1: Two equivalent value functions consistent with the preference relation of Example 5: v is Borda's count

The value of an impact is set equal to the number of impacts to which it is preferred, including itself.

Example 6 *Let us consider another example concerning a trip with five alternative means of transport:* $X = \{Airplane, Car, Coach, Train, Taxi\}$. *The preference relation is different from that of Example 5:* $\Pi = \{(Airplane, Airplane), (Airplane, Coach), (Airplane, Taxi), (Car, Car), (Car, Coach), (Car, Taxi), (Car, Train), (Taxi, Taxi), (Train, Taxi), (Train, Train)\}$.

Relation Π *is represented by the graph in Figure 4.2. Borda's count allows to build a consistent value function: the alternative Taxi is preferred only to itself, so that its value is* $v(f(\text{Taxi})) = 1$; *the alternatives Train and Coach are reciprocally indifferent and have value* 3; *the alternative Airplane has value* 4 *and the alternative Car, which is preferred to all, has value* 5 *(see Table 4.2).*

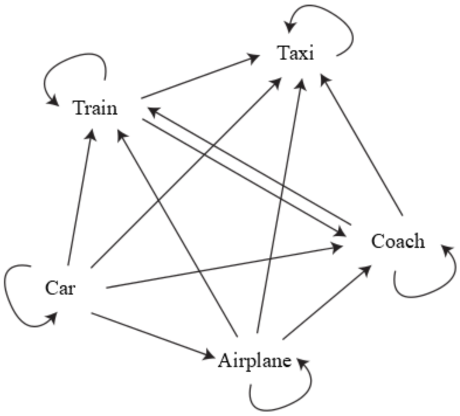


Figure 4.2: Preference graph for Example 6

X	$v(f(x))$
Airplane	4
Car	5
Coach	3
Taxi	1
Train	3

Table 4.2: Borda's count for Example 6

4.4 The one-dimensional case

When the impact is one-dimensional ($F \subseteq \mathbb{R} \Leftrightarrow p = 1$), it is often easy to build a consistent value function:

- if f represents a profit, $u(f) = f$ is a consistent value function; valore conforme;

- if f represents a cost, $u(f) = -f$ is a consistent value function;

- if the preferred values of f are the ones closer to a given value $\bar{f}$ (for instance, f is a desired temperature or an ideal level of irrigation for a field), $u(f) = -\left(f - \bar{f}\right)^2$ is a consistent value function.

Even in the one-dimensional case, however, there can be complex situations[2].

Example 7 *Let us assume that we have to decide the speed x with which to run in order to catch a bus that passes at regular intervals. The feasible speed is between a minimum (the slowest pace allowing to arrive at the bus stop in a reasonable time) and a maximum (that corresponds to the fastest acceptable run). Impact $f(x)$ is the arrival time at the bus stop. By interviewing the decision-maker, one can discover that he or she prefers to arrive as soon as possible to the destination, but hates to wait for the bus. Consequently, given two very close values of f, in general the preferred one will be the larger, because arriving later at the stop reduces the waiting time. For some critical pairs of impact, however, the smaller value will be clearly preferred, because it allows to catch the previous bus, with a zero waiting time. Therefore, the value function $v(f)$ has a andamento a dente di sega: it gradually grows with f, suddenly decrease, increases again, suddenly decreases, etc... Sometimes, a deeper analysis can help to simplify the model. For example, in this case, it is possible to replace the one-dimensional impact (arrival time at the bus stop) with a two-dimensional one, composed by the arrival time at destination and the waiting time at the bus stop. Such an impact is no longer expressed by a single number*

4.5 Lexicographic order

The decision-maker is required to fully sort the indicators. The unordered set $P = \{1, \ldots, p\}$, therefore, becomes a *sequence*, which we will denote with the same symbol, for the sake of simplicity:

$$P = (\pi_1, \ldots, \pi_p)$$

and that will represent a *hierarchy* on the indicators. Once this has been done, the following operations are performed:

- determine all solutions in which the first indicator assumes an optimal value:

$$X^\circ_{\pi_1} = \arg\min_{x \in X} f_{\pi_1}(x)$$

If only one solution remains ($\left|X^\circ_{\pi_1}\right| = 1$), the process terminates returning it.

[2] I am not satisfied with this example, because it is not really one-dimensional, but I am keeping it until I find a more convincing one.

- If more solutions remain ($\left|X^\circ_{\pi_1}\right| > 1$), determine among them the optimal ones with respect to the second indicator:

  $$X^\circ_{\pi_1\pi_2} = \arg \min_{x \in X^\circ_{\pi_1}} f_{\pi_2}(x)$$

  If only one solution remains ($\left|X^\circ_{\pi_1\pi_2}\right| = 1$), the process terminates returning it.

- ...

- If even at the p-th step more solutions remain ($|X^\circ_{\pi_1\dots\pi_{p-1}}| > 1$), choose any optimal solution with respect to the last indicator:

  $$x^\circ_{\pi_1\dots\pi_p} \in \arg \min_{x \in X^\circ_{\pi_1\dots\pi_{p-1}}} f_{\pi_p}(x)$$

Notice that the first $p-1$ steps of the method require to find all optimal solutions, not a single one, as it usually occurs in classical optimisation problems. Violating this requirement could yield the wrong solution. This makes the method potentially demanding from the computational point of view, but in practice it is usually enough to consider the first two or three indicators in order to identify a single solution.

The assumption that the decision-maker be able to provide a total order on the indicators, considering each of them as absolutely dominating with respect to the following ones is in general a stretch, but it allows to obtain anyway a somewhat reasonable decision. Of course, the resulting solution basically depends on the chosen order.

Example 8 *An advanced GPS navigation device offers the possibility to select not only the fastest path or the cheapest one, but to apply a lexicographic ordering method on the two indicators (time and cost). If the user specifies the order (cost, time), the GPS device enumerates all minimum-cost paths (e. g., three paths by* 10 *Euros, one by* 24 *hours, one by* 36 *hours and one by* 28 *hours). Then, among these paths it chooses one of those that require the minimum time (in this case, the optimal path is only one, and requires* 24 *hours). Slightly more expensive paths (e. g.,* 15 *or* 20 *Euros) are ignored even if they are much shorter (e. g.,* 2 *hours).*

4.6 Lexicographic order with aspiration levels*

The idea that an indicator be absolutely dominating with respect to the other ones is often too extreme, leading to very unbalanced solutions: optimal with respect to the first indicator, but very bad with respect to the following indicators. A human decision-maker often does not search for optimality, but for a satisfactory level of performance (Herbert Simon called such solutions “satisficing” from “satisfy” and “sufficing”). This line of thought leads to modify the lexicographic order method asking the decision-maker to totally order the indicators, but also to set an *aspiration level* ϵ_{π_ℓ} for all of them except the first one ($\ell \in \{1, \dots, p\} \setminus \{\pi_1\}$). This determines a restricted feasible region

$$X_\epsilon = \{x \in X : f_{\pi_\ell}(x) \leq \epsilon_{\pi_\ell} \text{ for } \ell = 2, \dots, p\}$$

For the first indicator, no aspiration level needs to be fixed, because it is still to be simply minimised. If the restricted feasible region X_ϵ is empty, the decision-maker must be asked to relax one or more of the aspiration levels.

*This section presents advanced concepts, that are not part of the course syllabus.

Then, the process described for the lexicographic order method is applied, operating in the restricted feasible region:

- determine all solutions in which the first indicator assumes the optimal value and the following ones assume values not worse than the corresponding aspiration levels:
$$X^\circ_{\epsilon,\pi_1} = \arg \min_{x \in X_\epsilon} f_{\pi_1}(x)$$
If a single solution remains ($\left|X^\circ_{\epsilon,\pi_1}\right| = 1$), the process terminates returning it.
- If more solutions remain ($\left|X^\circ_{\epsilon,\pi_1}\right| > 1$), determine among them the optimal ones with respect to the second indicator:
$$X^\circ_{\epsilon,\pi_1\pi_2} = \arg \min_{x \in X_{\epsilon,\pi_1}} f_{\pi_2}(x)$$
If a single solution remains ($\left|X^\circ_{\epsilon,\pi_1\pi_2}\right| = 1$), the process terminates returning it.
- ...
- If even at the p-th step more solutions remain ($|X^\circ_{\epsilon,\pi_1\dots\pi_{p-1}}| > 1$), choose any optimal solution with respect to the last indicator:
$$x^\circ_{\epsilon,\pi_1\dots\pi_p} \in \arg \min_{x \in X^\circ_{\epsilon,\pi_1\dots\pi_{p-1}}} f_{\pi_p}(x)$$

The basic idea of this method is that each indicator dominates the following ones, but only as far as their aspiration levels are respected. As soon as one of them reaches its aspiration level, it becomes more important and it is forbidden to worsen it further.

Of course, the final solution depends on the chosen order, but also on the $p-1$ chosen aspiration levels ϵ_{π_l} (for $l = 2, \dots, p$). Since all these choices are potentially arbitrary, it is once again a stretch, that allows to reach in a simple way a somewhat reasonable decision. The aspiration levels, however, allow to make the final solution less unbalanced in favour of the first indicator.

Example 9 *An advanced GPS navigation device offers the possibility to apply a lexicographic ordering method with aspiration levels. The users specifies the order (time, cost), and a threshold equal to $\epsilon = 20$ Euros on the cost (the method requires $p-1$ thresholds, that is a single one in this case, because the first indicator is optimised). The GPS device, therefore, excludes all paths with a cost larger than* 20 *Euros (for example, the three paths by* 10 *Euros listed in Example 8, that is those by* 24*,* 36 *and* 28 *hours, and the paths by* 15 *and* 20 *Euros seen above). Then, the device enumerates, among the residual paths, those that require the minimum time: two paths require* 2 *hours (and, respectively,* 15 *and* 20 *Euros). At this point, the device minimises the cost, choosing the path by* 2 *hours and* 15 *Euros. The threshold of* 20 *Euros is obviously arbitrary and implies that slightly more expensive paths are ignored, even if they could be much shorter.*

4.7 Utopia point

Each of the indicators has an optimal value, obtained by completely neglecting all other indicators
$$f^\circ_l = \min_{x \in X} f_l(x) \qquad l = 1, \dots, p$$

Determining this value is a classical optimisation problem, sometimes computationally hard, but in general possible to solve.

The vector composed by the optimal values of all p indicators forms an impact $f^U = \left[f_1^\circ \dots f_p^\circ\right]^T$. If $f^U \in F$, it is necessarily the solution to choose (see Example 44). In general, however, such in impact does not fall within F, that is, it cannot be obtained choosing a feasible solution. Such an impact is denoted as *utopia point*, because it provides an unreachable ideal.

A heuristic idea to select a solution is to determine one of those whose implied impact is closer to the utopia point. Thid definition is seemingly harmless and reasonable. However, it is based on a notion of *distance* in the space of the indicators that is quite not obvious. The most frequently used distances are:

- the Manhattan distance (L_1):

$$d\left(f, f'\right) = \sum_{l \in P} |f_l - f_l'|$$

- the Euclidean distance (L_2):

$$d\left(f, f'\right) = \sqrt{\sum_{l \in P} \left(f_l - f_l'\right)^2}$$

- the maximum of the distances referring to the single indicators (L_∞):

$$d\left(f, f'\right) = \max_{l \in P} |f_l - f_l'|$$

but infinite different distances can be defined.

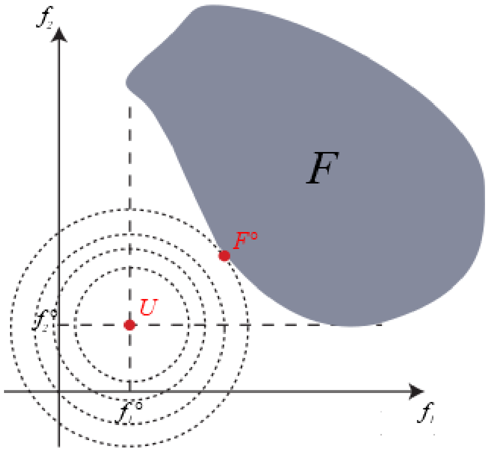


Figure 4.3: Utopia point and closest impact according to the L_2 metric

Example 10 *Apply the utopia point method to the decision problem with impact set $F = \{(f_1, f_2) \in \mathbb{R}^2 : f_1 \geq 0, f_2 \geq 0, f_1 + 2f_2 \geq 3\}$, represented in Figure 4.4, in which both indicators are costs to be minimised. For the sake of simplicity, we operate directly in the indicator space, without considering the feasible region, that is, we determine the desired impact, from which in theory one can deduce the solution that generates it.*

The utopia point is obtained considering the minimum value for each indicator, independently from the other one. In both cases, this value is zero, so that $f^\circ = (0,0)$*. If we consider the Manhattan distance* L_1*, the problem reduces to*

$$\min_{f \in F} |f_1 - 0| + |f_2 - 0| = \min_{f \in F} f_1 + f_2 \Rightarrow f^*_{L_1} = (0, 3/2)$$

If we consider the Euclidean distance L_2*, the problem reduces to*

$$\min_{f \in F} \sqrt{(f_1 - 0)^2 + (f_2 - 0)^2} = \min_{f \in F} \sqrt{f_1^2 + f_2^2} \Rightarrow f^*_{L_1} = (5/6, 4/3)$$

Finally, if we consider the L_∞ *distance, the problem reduces to*

$$\min_{f \in F} \max\left(|f_1 - 0|, |f_2 - 0|\right) = \min_{f \in F} \max\left(f_1, f_2\right) \Rightarrow f^*_{L_\infty} = (1, 1)$$

TO BE ADDED

Figure 4.4: According to the definition adopted for the distance, Example 10 has different solutions with minimum distance from the same utopia point

Moreover, most definitions of distance combine in a single quantity values associated with different indicators, expressed in heterogeneous units of measure. Before combining them, it is necessary to standardise them by multiplying them by suitable coefficients. The choice of the coefficients is complex, and at least partly arbitrary, as we shall see in Section 4.8.9.

4.8 Multi Attribute Utility Theory

The *Multi Attribute Utility Theory* (*MAUT*) assumes that the preference relation of the decision-maker be actually a weak order, admitting a consistent value function, but that the decision-maker be unable to make the value function explicit without help. It poses therefore the problem to derive from the preference relation Π the consistent value function $u : F \to \mathbb{R}$. In the following, we will adopt the economical notation, denoting $u(f)$ as *utility function*.

First, we will tackle the problem in its most general form, introducing a graphical tool and a procedure that in theory should allow to solve the problem, but this procedure will prove very complicated, expensive and error-prone. We will therefore focus on a specific family of utility functions, whose characteristics make it simpler (even if nontrivial) to derive them from the preferences revealed by the decision-maker. We will then discuss the necessary and sufficient conditions under which the preference relation can be expressed through a function of such family, and finally we will describe a procedure to derive the function itself.

4.8.1 Indifference curves

Definition 5 *Given an impact set* F *in the space of the indicators* $\mathbb{R}^p$ *and a preference relation* Π*, we denote as* indifference curve *every subset of impacts that are reciprocally indifferent.*

Under suitable continuity and regularity assumptions, impacts close to each other are also similar with respect to preference[3]. Every subset of reciprocally indifferent impacts yields a regular hypersurface in $\mathbb{R}^p$. As the indifference curves are infinitely many, one for each real value $u(f)$, in general each curve is a $(p-1)$-dimension hypersurface. The various curves are linked by a total order relation; in fact, there are no incomparable impacts and all indifferent impacts belong to the same curve, so that impacts on different curves are linked by a strict preference, and all impacts on a curve are strictly preferable to all impacts on the other one. Finally, the completeness of the relation guarantees that the family of all indifference curves covers the whole set F (see Figure 4.5).

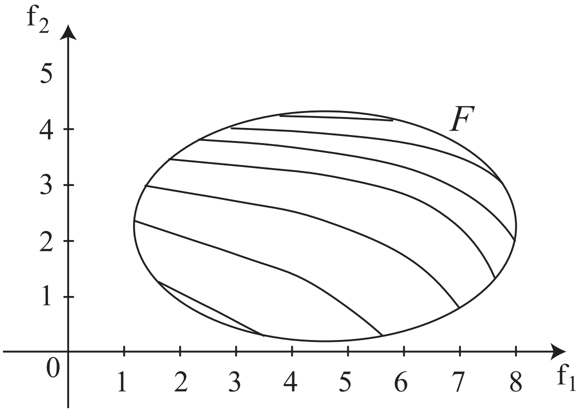


Figure 4.5: Indifference curves: there are ∞ totally ordered curves that cover the whole set F; each curve includes ∞^{p-1} reciprocally indifferent impacts.

If a value function is known, the family of all indifference curves admit an analytic representation in the indicator space F through the equation $u(f) = c$, where c is a constant parameter that identifies each single curve. This corresponds to an analytic representation in the solution space X through the parametric equation $u(f(x)) = c$.

Indifference curves and utility function

From Theorem 5 in Section 4.2.3 we know that, if a preference relation admits a consistent value function, it admits infinitely many. The indifference curves of such functions, however, are always the same. The correspondence between preference relations and indifference curve families is one-to-one, that between preference relations and value functions is one-to-many.

Remark 1 *All value functions u consistent with a preference relation Π have the same indifference curves $u(f) = c$, where $c \in \mathbb{R}$ is a real parameter identifying each curve. Each value function u associates different values of c to each curve, but such values are ordered in the same way.*

[3] We will not go into details about these assumptions, but we remark that they require a continuous impact set F. Many of the concepts developed in the following, in fact, make no sense for discrete problems. In these problems, the impact set must be extended, turning the indicators that assume discrete values to real values with some form of interpolation (for example, passing from an integer number of "sweets" to a real number measuring "portions of sweet"). The indifference curves are defined on this extended impact set, and later restricted to the original discrete set.

Example 11 *Let us consider the utility functions $u\left(f_1, f_2\right) = f_1^2 f_2^3$ and $u'\left(f_1, f_2\right) = 2\log f_1 + 3\log f_2$. The indicators f_1 and f_2 used in this example express a benefit, but the concept holds for any kind of indicator. The two functions are equivalent, because they are related by the invertible transformation $u' = \log u$. Figure 4.6 reports the corresponding indifference curves: they coincide perfectly, but the same curve in the two cases is associated to different utility values.*

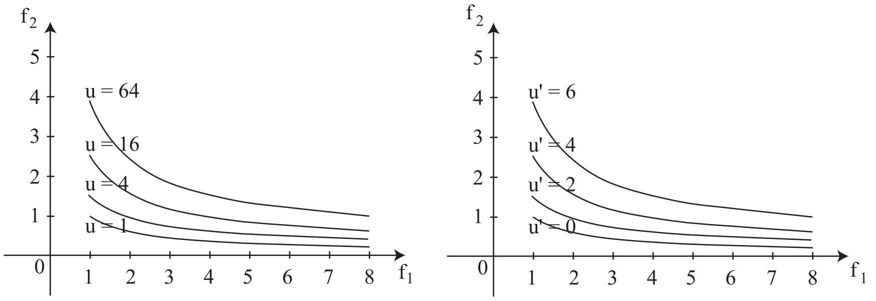


Figure 4.6: Two different, but equivalent, utility functions: $u\left(f_1, f_2\right) = f_1^2 f_2^3$ and $u'\left(f_1, f_2\right) = 2\log f_1 + 3\log f_2$. The indifference curves coincide, even though they correspond to different values.

4.8.2 Determining the utility function

The indifference curves help to estimate a consistent value function for the preference relation through the following process:

1. propose a sample of impact $\tilde{F} \subseteq F$ to the decision-maker, requiring to compare them by pairs;

2. based on the comparisons, classify the impacts:

    - collect the reciprocally indifferent impacts into equivalence classes;
    - order the equivalence classes based on the preference;

3. based on the shape of the indifference curves in the indicator space, assume a family of utility functions $u_\alpha\left(f_1, \ldots, f_p\right)$, defined up to a a vector of numerical parameters α;

4. determine the values of the parameters α_r $(r = 1, \ldots, s)$ imposing that reciprocally indifferent impacts have the same utility value;

5. make consistency checks, generating new impacts, computing the preferences between such impacts through the estimated utility function and asking the decision-maker to confirm or disprove the preferences obtained;

6. in case of errors, modify the family of utility functions and repeat the process.

The process is clearly fragile. If one considers a small sample, it is quite likely that the estimated curves be incorrect; if the sample is large, the workload required from the decision-maker quickly becomes huge (if one evaluates k different values for each indicator, the sample includes k^p different impacts, a number that quickly becomes unmanageable). Moreover, the arbitrariness in the choice of the family of utility functions opens the way to long trial-and-error loops.

The only possibility is that the preference relation of the decision-maker be very simple, and therefore the indifference curves have a very simple analytic form, depending on few parameters, easy to estimate. We are therefore interested to evaluate special cases that allow an easier estimate, hoping that they could be close enough to the practical case to be used as approximation of the reality. The economy courses discuss several families of utility functions, that correspond to different assumptions on the preference relation of the decision-maker. Properties often postulated are:

1. *invertibility*: for each fixed value of $p-1$ indicators, there exist a single value of the remaining indicator that produces a given utility; under this assumption, an indifference curve $u(f) = c$ can be written in explicit form as $f_l = f_l(c, f_1, \dots, f_p)$;
2. *monotony*: in order to compensate for the variations of an indicator, the other ones must vary in a well-determined direction; for example, when all indicators represent costs, in order to compensate for an increase of one, at least another one must decrease, and therefore the curve is decreasing (the same occurs when all indicators are benefits).
3. *convexity* (or *concavity*): the indifference curves compensate for the increase of an indicator by a certain amount with variations of the other ones that increase (or decrease) with the value of the first indicator; the basic idea is that the indicators represent resources: if a resource is scarse, increasing it brings a large utility, compensated by a strong decrease of other resources; if on the contrary it is abundant, increasing it by the same amount brings a small utility, compensated by a weak decrease of other resources; in this case, the curves are convex (they would be concave in the case of indicators expressing costs).

Example 12 *An individual must decide in which town to live based on two (very simple) indicators: the pollution level and the local tax level. Two towns with different pollution levels will be indifferent only if the more polluted one has a lower tax level. So, the indifference curves are monotone decreasing and invertible. Moreover, if the pollution level in a town is small, a given increase of pollution can be compensated by a small tax reduction, whereas if it is high, the same increase will require a strong fiscal incentive.*

Example 13 *A family of utility functions frequently used in economics are the* Cobb-Douglas functions*, defined as:*

$$u_\alpha(f_1, \dots, f_p) = \prod_{l=1}^{p} {f_l}^{\alpha_l}$$

These functions consider the indicators f_l as benefits (they have in fact been proposed by economists), require a single parameter for each indicator and their indifference curves are similar to hyperboloids. The utility function $u(f_1, f_2) = f_1^2 f_2^3$, whose indifference curves are reported in Figure 4.6, belongs to this family: its parameters are, in fact, $\alpha_1 = 2$ and $\alpha_2 = 3$. These functions are invertible ($f_1 = \sqrt{c/f_2^3}$ e $f_2 = \sqrt[3]{c/f_1^2}$ for $f_1 \geq 0$ and $f_2 \geq 0$), monotone strictly decreasing (compute the first derivative) and convex (compute the second derivative).

Notice that the infinite multiplicity of the value functions that are equivalent to a given preference relation implies that the parameters that define a given family of value functions cannot be fully fixed based on the empirical analysis of the indifferent impacts. They must keep a degree of freedom. This degree can be removed only with an arbitrary condition of *normalisation.*

Example 14 *The preliminary analysis of preference of a decision-maker suggests that it corresponds to a Cobb-Douglas function. Moreover, the decision-maker states that the impacts $f = (8,1)$ and $f' = (1,4)$ are indifferent. As a consequence:*

$$(8,1) \sim (1,4) \Rightarrow u_\alpha (8,1) = u_\alpha (1,4) \Rightarrow 8^{\,\alpha_1} 1^{\,\alpha_2} = 1^{\,\alpha_1} 4^{\,\alpha_2} \Rightarrow 8^{\,\alpha_1} = 4^{\,\alpha_2} \Rightarrow 3\alpha_1 = 2\alpha_2$$

Since the utility function remains equivalent when it is raised to any positive exponent, $(u_\alpha (f))^\beta$ is equivalent to $u_\alpha (f)$ for any $\beta > 0$. But this corresponds to multiplying all parameters α_l by β, and therefore in order to normalise the Cobb-Douglas functions it is enough to impose the additional condition that $\sum_{l=1}^{p} \alpha_l = 1$. In conclusion, $3\alpha_1 = 2\alpha_2$ and $\alpha_1 + \alpha_2 = 1$ imply that $\alpha_1 = 2/5$ and $\alpha_2 = 3/5$. Of course, in practice a single pair of indifferent impacts is certainly not enough to determine with reasonable certainty the value function to use.

4.8.3 Additive utility functions

There are utility functions which are easier to estimate starting from a sampling of the preference relation on the impact set.

Definition 6 *We denote a utility function as* additive *when it can be expressed as the sum of functions of the single indicators:*

$$u (f_1, \dots, f_p) = \sum_{l=1}^{p} u_l (f_l)$$

Since every utility function admits infinitely many equivalent ones, it is quite complicated to evaluate whether an additive equivalent function exists or not in a given situation. A classical example is given by the Cobb-Douglas' functions, which are not themselves additive, but which, as already observed in the example of Figure 4.6, admit equivalent functions which are additive:

$$u' (f) = \log u (f) = \log \prod_{l=1}^{p} {f_l}^{\alpha_l} = \sum_{l=1}^{p} \alpha_l \log f_l = \sum_{l=1}^{p} u_l (f_l)$$

where $u_l (f_l) = \alpha_l \log f_l$ for all $l \in P$.

If the utility function is additive, the problem to estimate it can be reduced to the estimation of p single-variable functions. Recalling the case studies of Chapter 2, the number of indicators can be huge, and their nature can be extremely heterogeneous, so that the claim to compare impacts extracted from distant point of set F could become pointless. If one could deal with single-variable functions, the task to estimate the utility would be much easier; in particular, one could assign the estimation of each component of the utility function to a different field expert. Of course, there are intermediate situations, in which the utility function is not completely additive, but at least it can be decomposed in a sum of functions depending on small reciprocally independent blocks of few indicators. That would be a step forward, anyway.

Since *a priori* the utility function is unknown, but the preference relation is known, we must investigate what properties of the preference relation allow to guarantee that an additive utility function actually exist. Next section presents the property that allows to solve the problem, while the following one describes its relation with additivity.

4.8.4 Preferential indipendence

Given the set of attributes $P = \{1, \dots, p\}$, preferential independence is a property of a subset of attributes $L \subset P$ with respect to its complement $\bar{L} = P \setminus L$. If, for the sake of simplicity, we order the attributes, collecting in the first positions those of subset L, we can write that $f = \begin{bmatrix} f_L \\ f_{\bar{L}} \end{bmatrix}$, where f_L and $f_{\bar{L}}$ are the subvectors of impact f corresponding, respectively, to the indicators of L and of $\bar{L}$.

Definition 7 *A proper subset of indicators $L \subset P$ is* preferentially independent *from the complementary subset $\bar{L}$ when, given two impacts with identical values of the indicators in $\bar{L}$, the preference relation between them does not depend on such values:*

$$\begin{bmatrix} f_L \\ \phi \end{bmatrix} \preceq \begin{bmatrix} f'_L \\ \phi \end{bmatrix} \Leftrightarrow \begin{bmatrix} f_L \\ \psi \end{bmatrix} \preceq \begin{bmatrix} f'_L \\ \psi \end{bmatrix}$$

for all f_L, f'_L, ϕ, ψ such that

$$\begin{bmatrix} f_L \\ \phi \end{bmatrix}, \begin{bmatrix} f'_L \\ \phi \end{bmatrix}, \begin{bmatrix} f_L \\ \psi \end{bmatrix}, \begin{bmatrix} f'_L \\ \psi \end{bmatrix} \in F$$

Example 15 *A decision-maker must choose the menu for a simple lunch, pairing a single course and a glass of wine. The available courses are: stew, roast, meatballs, salmon and swordfish; the available wines are Barolo, Nebbiolo, Erbaluce, Arneis. We have therefore a feasible region X made up of 20 alternatives, that are the combinations of 5 courses and 4 wines (see the left side of Figure 4.7). The decision-maker is rather rough and only perceives the difference between meat and fish, and the difference between red and white wine. Then, the impact set reduces to 4 elements: $F =$ (meat, red wine), (meat, white wine), (fish, red wine), (fish, white wine), as is apparent in Figure 4.7, where the impact function $f(x)$ associates two subsets of 6 solutions and two subsets of 4 soluzioni on the left side to four single impacts on the right side. Assuming that the decision-maker follows the traditional use of pairing meat courses with red wine and fish courses with white wine, one obtains the preferences indicated by black arrows on the right side. The double arrows in grey represent the fact that meat courses with red wine are indifferent with respect to fish courses with white wine, and that meat courses with white wine are indifferent with respect to fish courses with red wine. These preferences could be different, without any consequence on the property we are discussing: we report them only to stress that the overall preference relation is compete. The same holds for the self-loops on the single impacts. It is easy to show that the indicators "course" (f_{course}) and "wine" (f_{wine}) are not preferentially independent. In fact, fixing the red wine, meat courses are preferable to fish courses; fixing the white wine, fish courses are preferable to meat courses. The same holds in the opposite direction: fixing a meat course, in fact, red wine is preferable; fixing a fish course, white wine is preferable.*

Example 16 *A decision-maker must define an industrial process to produce pots and lids, evaluating it with respect to the attributes "cost", "number of pots" and "number of lids". The preferable process have a smaller cost and a number of pots and lids as close as possible to each other. The pair ("cost","number of pots") is not independent from the attribute "number of lids", because for each fixed number of lids, it is preferable that the number of pots be equal to that of lids. As well, the pair ("cost","number of lids") is not independent from the attribute "number of pots". By constrast, the pair ("number of pots","number of lids") is independent from the attribute "cost", because for each fixed cost the preference relation between different combinations of pots and lids is always the same.*

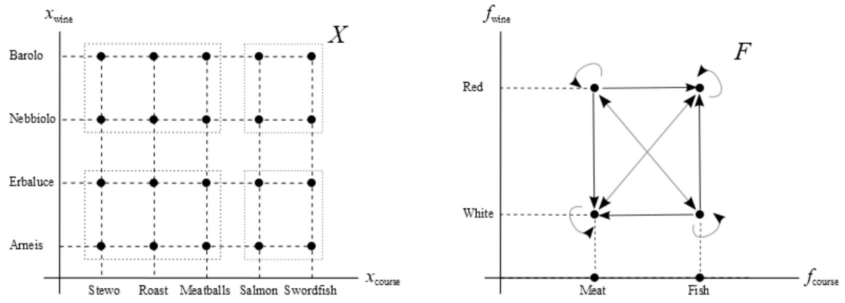


Figure 4.7: A preference relation that does not enjoy preferential independence between the two indicators: if the wine is red, meat is preferable to fish; if the wine is white, fish is preferable to meat; the same holds for the courses with respect to the wine

One could assume independence to be a symmetrical property. In other words, one could assume that when L is independent from $\bar{L}$, conversely $\bar{L}$ be independent from L. The following example shows the contrary.

Example 17 *Given the impact set $F = \left\{ f \in \mathbb{R}^2 : f_1 \geq 0, f_2 \geq 1 \right\}$ and the utility function $u(f) = (f_1 - 5) f_2$, the indicator f_1 is preferentially independent from the indicator f_2 because, fixing $f_2 = \bar{f}_2 \geq 1$, it is always*

$$\left(f_1, \bar{f}_2\right) \preceq \left(f'_1, \bar{f}_2\right) \textit{ per } f_1 \geq f'_1$$

On the contrary, f_2 is preferentially dependent on f_1 because, fixing $f_1 = \bar{f}_1 \geq 1$:

$$\begin{cases} \textit{when } 0 \leq \bar{f}_1 < 5, & \left(\bar{f}_1, f_2\right) \preceq \left(\bar{f}_1, f'_2\right) \textit{ for } f_2 \leq f'_2 \\ \textit{when } \bar{f}_1 = 5, & \left(\bar{f}_1, f_2\right) \sim \left(\bar{f}_1, f'_2\right) \textit{ for all } f_2, f'_2 \\ \textit{when } \bar{f}_1 > 5, & \left(\bar{f}_1, f_2\right) \preceq \left(\bar{f}_1, f'_2\right) \textit{ for } f_2 \geq f'_2 \end{cases}$$

One could also assume that, when an indicator is independent from the other ones, then all subsets of indicators be reciprocally independent. Also this is not true, as shown by the following example.

Example 18 *Given the impact set $F = \left\{ f \in \mathbb{R}^3 : f_1 \geq 0, f_2 \geq 0, f_3 \geq 1 \right\}$ and the utility function $u(f) = 1/\left[(f_1 + f_3)(f_2 + f_3)\right]$, each attribute is independent from the other ones: they are all costs. However, the pair of attributes (f_1, f_2) depends preferentially on the attribute f_3. In fact, it is:*

$$\begin{bmatrix} 1 \\ 3 \\ 1 \end{bmatrix} \succ \begin{bmatrix} 4 \\ 1/2 \\ 1 \end{bmatrix} \quad \textit{since } u(1,3,1) = \frac{1}{8} < \frac{2}{15} = u(4, 1/2, 1)$$

$$\begin{bmatrix} 1 \\ 3 \\ 3 \end{bmatrix} \prec \begin{bmatrix} 4 \\ 1/2 \\ 3 \end{bmatrix} \quad \textit{since } u(1,3,3) = \frac{1}{24} > \frac{2}{49} = u(4, 1/2, 3)$$

A troubling aspect of this property is the time required to verify it. When L includes a single indicator, it is often easy to prove its independence from $\bar{L}$. For example, when the indicator represents a cost, or a benefit, it is always preferentially independent from the other ones, because the decision-maker will always prefer an impact with smaller cost (or larger benefit), if the other indicators are unchanged, whatever value they assume. Unfortunately, even when all indicator, considered one by one, are independent, it is not guaranteed that every subset L of indicators be independent from its complement. Then, it is necessary to prove preferential independence for each of the $2^p - 2$ proper subsets (excluding only the empty subset $\emptyset$ and the whole set of indicators P), and this must be done empirically sampling the impacts and asking many tedious questions to the decision-maker. Luckily, it is not necessary to consider all subsets.

Theorem 6 *A decision problem with $p \geq 3$ indicators enjoys mutual preferential independence if and only if there exists an index $\bar{l} \in P$ such that every pair $\{\bar{l}, l\}$ with $l \neq \bar{l}$ is preferentially independent from its complement $P \setminus \{\bar{l}, l\}$.*

Therefore, in order to verify mutual preferential independence, if is not required to consider all $2^p - 2$ proper subsets of indicators, but only the $p - 1$ pairs that one indicator (chosen *ad libitum*) forms with the others: as soon as a non independent pair is found, the check terminates with a negative result; if all pairs are independent, the check terminates with a positive result.

4.8.5 Preferential independence and additivity

Preferential independence of relation Π and additivity of function $u(f)$ are strictly connected, even if not rigorously equivalent.

Definition 8 *We say that a problem enjoys* mutual preferential independence *when every subset of indicators $L \subset P$ is preferentially independent from its complement $\bar{L}$.*

Mutual preferential independence is a necessary condition for additivity.

Theorem 7 *If a preference relation admits an additive utility function, then it enjoys mutual preferential independence.*

Proof. It is enough to apply the definition. For any quadruplet of vectors f_L, f'_L, ϕ, ψ producing valid impacts:

$$\begin{bmatrix} f_L \\ \phi \end{bmatrix} \preceq \begin{bmatrix} f'_L \\ \phi \end{bmatrix} \Rightarrow u(f_L, \phi) \geq u(f'_L, \phi)$$

from which by additivity

$$\sum_{l \in L} u_l(f_l) + \sum_{l \in \bar{L}} u_l(\phi_l) \geq \sum_{l \in L} u_l(f'_l) + \sum_{l \in \bar{L}} u_l(\phi_l)$$

where the two identical sums in the two members of the inequality can be replaced by other two identical sums

$$\sum_{l \in L} u_l(f_l) + \sum_{l \in \bar{L}} u_l(\psi_l) \geq \sum_{l \in L} u_l(f'_l) + \sum_{l \in \bar{L}} u_l(\psi_l) \Rightarrow$$

$$\Rightarrow \begin{bmatrix} f_L \\ \psi \end{bmatrix} \preceq \begin{bmatrix} f'_L \\ \psi \end{bmatrix}$$

■

However, we need a sufficient condition, because in practice we can verify mutual preferential independence, and not additivity, as the utility function is unknown. In this verse, the implication holds nearly always, but not always.

Now, it is timely to ask two questions:

1. is mutual preferential independence also a sufficient condition for additivity?
2. in order to verify mutual preferential independence is it strictly necessary to verify the preferential independence of every subset L?

Theorem 8 *A decision problem with $p \geq 3$ indicators enjoying mutual preferential independence admits an additive utility function $u(f)$.*

So, for problems with at least three indicators, mutual preferential independence is a necessary and sufficient condition for additivity. Unfortunately, in the case of two indicators, this is no longer true.

Example 19 *Figure 4.8 reports the indifference curves of a decision problem with $p = 2$ indicators that represent benefits. Mutual preferential independence holds, because for any fixed value of f_2 it is preferable to increase f_1, and for each fixed value of f_1 it is preferable that f_2 increase. However, no additive utility function is able to determine such indifference curves.*

In fact, one can observe that:

$$(2,0) \sim (1,1) \sim (0,2) \text{ and } (1,0) \sim (0,1)$$

By contradiction, let us suppose that an additive utility function exist. Since $(2,0)$ and $(0,2)$ are indifferent impact, they have the same utility; the same occurs for $(1,0)$ and $(0,1)$:

$$\begin{cases} u(0,2) = u(2,0) \Rightarrow u_1(0) + u_2(2) = u_1(2) + u_2(0) \\ u(0,1) = u(1,0) \Rightarrow u_1(0) + u_2(1) = u_1(1) + u_2(0) \end{cases}$$

Subtracting the corresponding terms, one obtains

$$u_2(2) - u_2(1) = u_1(2) - u_1(1)$$

and therefore

$$u_2(2) + u_1(1) = u_1(2) + u_2(1) \Rightarrow u(1,2) = u(2,1) \Rightarrow (1,2) \sim (2,1)$$

but the two impacts $(1,2)$ e $(2,1)$ are not indifferent.

Example 19 and Figure 4.8 do not only suggest that mutual preferential indipendence is insufficient to guarantee additivity when $p = 2$, but also that the issue concerns the indifference curves: the presence of two straight lines, followed by other curved lines (a perfectly possible case as long as the curves do not intersect) makes it impossible to build an additive utility function.

4.8.6 Marginal rate of substitution

The missing condition to guarantee additivity concerns the steepness of the indifference curves. We therefore need a measure of steepness.

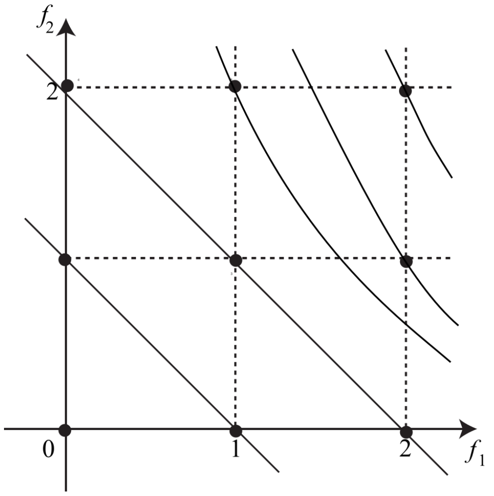


Figure 4.8: A decision problem with $p = 2$ indicators that are preferentially independent, but with no additive utility function

Definition 9 *We denote as* marginal rate of substitution *(*MRS*) of f_1 with f_2 in a given impact f the limit*

$$\lambda_{12}(f) = \lim_{\delta f_1 \to 0} -\frac{\delta f_2(f, \delta f_1)}{\delta f_1}$$

where $\delta f_2(f, \delta f_1)$ is such that: $f + \begin{bmatrix} \delta f_1 \\ \delta f_2(f, \delta f_1) \end{bmatrix} \sim f$.

In other words, the marginal rate of substitution is the limit, for infinitesimal variations of f_1, of the ratio between the variations of f_2 and f_1 that produce impacts indifferent with respect to f. The sign of the limit is reversed because it is quite often negative (for example, when the two indicators are both costs or both benefits). The definition assumes to start from an impact f, to slightly modify the value of indicator f_1 and to determine the corresponding variation of indicator f_2 that allows to stay on the initial indifference curve[4].

It is possible to give three different, but equivalent, expressions of the marginal rate of substitution. The first one is not particularly meaningful, but it is useful to obtain the other two. Let us assume that the indifference curves be regular arcs and let us represent them in parametric form:

$$\begin{cases} f_1 = f_1(\alpha) \\ f_2 = f_2(\alpha) \end{cases}$$

The variations δf_1 e δf_2 used in the definition of $\lambda_{12}(f)$ correspond to a variation $\delta\alpha$ of the parameter; in this way, in fact, it is guaranteed that the impact remain on the indifference curve.

$$\lambda_{12}(f) = \lim_{\delta f_1 \to 0} -\frac{\delta f_2(f, \delta f_1)}{\delta f_1} = \lim_{\delta\alpha \to 0} -\frac{f_2(\alpha + \delta\alpha) - f_2(\alpha)}{f_1(\alpha + \delta\alpha) - f_1(\alpha)} = -\frac{\frac{df_2}{d\alpha}}{\frac{df_1}{d\alpha}} \tag{4.3}$$

[4] In problems with more than two indicators, a marginal rate of substitution is defined for each pair of indicators (l, l'), and it is computed assuming that all other indicators remain constant while f_l and $f_{l'}$ are varied.

Remark 2 *From Equation* (4.3) *it follows immediately that the marginal rates of substitution of f_1 with respect to f_2 and of f_2 with respect to f_1 are reciprocals:*

$$\lambda_{12}(f) = -\frac{\frac{df_2}{d\alpha}}{\frac{df_1}{d\alpha}} \text{ } e \text{ } \lambda_{21}(f) = -\frac{\frac{df_1}{d\alpha}}{\frac{df_2}{d\alpha}} \Rightarrow \lambda_{12}(f) = \frac{1}{\lambda_{21}(f)} \text{ per ogni } f \in F$$

Let us see now other two possible expressions for the marginal rate of substitution.

Marginal rate of substitution and utility function Along an indifference curve, the utility function is constant: $u(f(\alpha)) = c$ for all α. Therefore, it has a first derivative equal to zero with respect to α:

$$\frac{du(f_1(\alpha), f_2(\alpha))}{d\alpha} = 0 \Rightarrow \frac{\partial u}{\partial f_1}\frac{df_1}{d\alpha} + \frac{\partial u}{\partial f_2}\frac{df_2}{d\alpha} = 0 \Rightarrow -\frac{\frac{df_2}{d\alpha}}{\frac{df_1}{d\alpha}} = \frac{\frac{\partial u}{\partial f_1}}{\frac{\partial u}{\partial f_2}}$$

Equation (4.3) implies that:

$$\lambda_{12}(f) = \frac{\frac{\partial u}{\partial f_1}}{\frac{\partial u}{\partial f_2}} \tag{4.4}$$

that is, the marginal rate of substitution of f_1 with f_2 is the ratio of the partial derivatives of the utility function with respect to f_1 and f_2. If the utility depends strongly on f_1 and weakly on f_2, the marginal rate of substitution is large, that is, a large variation of f_2 is required to compensate for a small variation of f_1[5].

Marginal rate of substitution and indifference curves If the indifference curves are invertible, then each value of parameter α corresponds to a different value of f_1 and f_2, and one can define function $\alpha = \alpha(f_1)$, which implies an explicit expression for the indifference curve: $f_2 = f_2(\alpha(f_1))$. The first derivative of that expression is:

$$\frac{df_2}{df_1} = \frac{df_2}{d\alpha}\frac{d\alpha}{df_1} = \frac{\frac{df_2}{d\alpha}}{\frac{df_1}{d\alpha}}$$

from which

$$\lambda_{12}(f) = -\frac{df_2}{df_1} \tag{4.5}$$

The marginal rate of substitution is the steepness of the indifference curves, with a reverse sign.

4.8.7 Additivity and marginal rate of substitution

Let us go back to the problem of which preference relations with $p = 2$ indicators are additive. By combining two distinct values for indicator f_1 (respectively, f_1' and f_1'') and two distinct values for indicator f_2 (f_2' and f_2''), one can build four different impacts. In general, the marginal rates of substitution λ_{12} in these four impacts are fully independent. However, it can happen that they are related.

[5] Equation (4.4) also holds for $p \geq 3$ indicators, given that the rate of substitution is defined keeping all indicators constant, except for two.

Definition 10 *We denote as* corresponding trade-off condition *the following property:*

$$\lambda_{12}(f_1', f_2')\lambda_{12}(f_1'', f_2'') = \lambda_{12}(f_1'', f_2')\,\lambda_{12}(f_1', f_2'')$$

for every quadruplet of impacts (f_1', f_2'), (f_1'', f_2''), (f_1'', f_2'), $(f_1', f_2'') \in F$.

For any rectangle in F, the corresponding trade-off condition requires that the products of the marginal rates of substitution in opposite vertices be the same. The relation can be expressed also in the following equivalent ways:

$$\frac{\lambda_{12}(f_1', f_2')}{\lambda_{12}(f_1'', f_2')} = \frac{\lambda_{12}(f_1', f_2'')}{\lambda_{12}(f_1'', f_2'')} \quad \text{and} \quad \frac{\lambda_{12}(f_1', f_2')}{\lambda_{12}(f_1', f_2'')} = \frac{\lambda_{12}(f_1'', f_2')}{\lambda_{12}(f_1'', f_2'')}$$

In other words, when moving from f_1' to f_1'', the marginal rate of substitution varies by the same multiplying factor for any value of f_2; as well, when moving from f_2' to f_2'', the marginal rate of substitution varies by the same multiplying factor for any value of f_1.

Figure 4.9 shows the condition: in the four points P, Q, R and S the marginal rates of substitution are, respectively, equal to $-a/c$, $-b/c$, $-a/d$ and $-b/d$. The product of the rates in P and S is ab/cd, and it coincides with the product of the rates in points Q and R. In fact, moving from P to Q the marginal rate of substitution changes by a multiplying factor equal to b/a; moving from R to S the two substitution rates are still different, but related by the same multiplying factor b/a. The same occurs moving along axis f_2, that is from P to R and from Q to S.

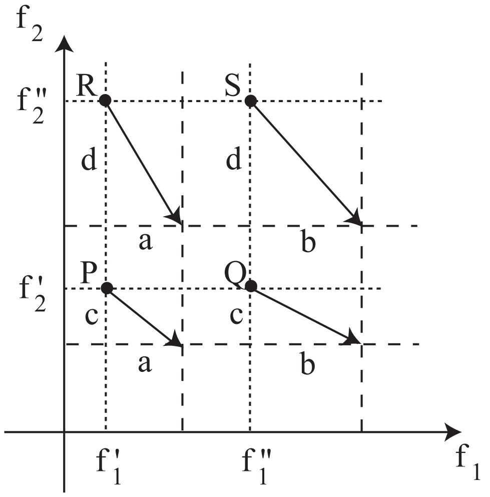


Figure 4.9: The corresponding trade-off condition holds: the ratio between the variations of f_2 and f_1 increases with f_1 in a regular way (independent from f_1) and decreases as f_2 increases, also in a regular way (independent from f_2)

Example 20 *Let us consider the preference relation of Figure 4.8:*

$$(2,2) \prec (2,1) \prec (1,2) \prec (2,0) \sim (1,1) \sim (0,2) \prec (1,0) \sim (0,1) \prec (0,0)$$

which enjoys mutual preferential independence, but not additivity. Let us consider the four points $(1,0)$, $(2,0)$, $(1,1)$ *and* $(2,1)$*: in the first three points, the rate of substitution is* 1 *(in order to counterbalance an increase in* f_1*, it takes an identical increase in* f_2*; in* $(2,1)$*, on the contrary, the rate of substitution is strictly larger* *than 1 (in order to counterbalance an increase in f_1 it is not enough to increase f_2 by the same amoung). Therefore, the corresponding trade-off condition does not hold.*

Example 21 *Let us consider the preference relation whose indifference curves are described by in Figure 4.6 by Cobb-Douglas' utility function $u(f_1, f_2) = f_1^2 f_2^3$. The marginal rate of substitution is:*

$$\lambda_{12}(f) = \frac{\dfrac{\partial u}{\partial f_1}}{\dfrac{\partial u}{\partial f_2}} = \frac{2 f_1 f_2^3}{3 f_1^2 f_2^2} = \frac{2 f_2}{3 f_1}$$

which enjoys the corresponding trade-off condition:

$$\lambda_{12}(f_1', f_2')\,\lambda_{12}(f_1'', f_2'') = \frac{2f_2'}{3f_1'}\frac{2f_2''}{3f_1''} = \frac{2f_2'}{3f_1''}\frac{2f_2''}{3f_1'} = \lambda_{12}(f_1', f_2'')\,\lambda_{12}(f_1'', f_2')$$

In fact, this relation can also be expressed with an additive utility function, as already shown.

Now, let us explore the relation between the corresponding trade-off condition and additivity.

Theorem 9 *If a preference relation admits an additive utility function, it enjoys the corresponding trade-off condition with respect to any pair of indicators.*

Proof. Once again, it is a matter of applying the definition:

$$\lambda_{12}(f_1', f_2')\lambda_{12}(f_1'', f_2'') = \frac{\dfrac{\partial u}{\partial f_1'}}{\dfrac{\partial u}{\partial f_2'}}\frac{\dfrac{\partial u}{\partial f_1''}}{\dfrac{\partial u}{\partial f_2''}}$$

Since $u(f_1, f_2) = u_1(f_1) + u_2(f_2)$:

$$\lambda_{12}(f_1', f_2')\lambda_{12}(f_1'', f_2'') = \frac{\dfrac{\partial u_1}{\partial f_1'}}{\dfrac{\partial u_2}{\partial f_2'}} \cdot \frac{\dfrac{\partial u_1}{\partial f_1''}}{\dfrac{\partial u_2}{\partial f_2''}} = \frac{\dfrac{\partial u_1}{\partial f_1'}}{\dfrac{\partial u_2}{\partial f_2''}} \cdot \frac{\dfrac{\partial u_1}{\partial f_1''}}{\dfrac{\partial u_2}{\partial f_2'}}$$

and getting back to function $u(f_1, f_2)$ one obtains:

$$\lambda_{12}(f_1', f_2')\lambda_{12}(f_1'', f_2'') = \frac{\dfrac{\partial u}{\partial f_1'}}{\dfrac{\partial u}{\partial f_2''}}\frac{\dfrac{\partial u}{\partial f_1''}}{\dfrac{\partial u}{\partial f_2'}} = \lambda_{12}(f_1', f_2'')\lambda_{12}(f_1'', f_2')$$

■

Therefore, the corresponding trade-off condition is necessary for additivity. One can wonder whether it is also sufficient, at least under suitable conditions. The answer is positive, exactly in the missing case.

Theorem 10 *If $p = 2$ and the two indicators enjoy the corresponding trade-off condition, there exists an additive utility function.*

Example 22 *The Cobb-Douglas' functions enjoy the corresponding trade-off condition even though they are not additive utility function. However, they are equivalent to the additive functions $u'(f) = \sum_{l=1}^{p} \alpha_l \log f_l$, obtained by applying the simple transformation $u'(f) = \log u(f)$.*

Uniform marginal rate of substitution

A remarkable special case which enjoys the corresponding trade-off condition is that in which the marginal rate of substitution does not depend on f, but is uniform ($\lambda_{12}(f) = \bar{\lambda}$), that is when the decision-maker is always willing to replace a unit of f_1 with a fixed number $\bar{\lambda}$ of units of f_2. In that case, there exist the additive utility function:

$$u(f_1, f_2) = w_1 f_1 + w_2 f_2$$

and the marginal rate of substitution is $\lambda_{12} = \dfrac{w_1}{w_2}$. The indifference curves form a family of parallel straight lines.

4.8.8 Building an additive utility function

Let us assume that theory guarantee the existence of an additive utility function $u(f) = \sum_{l=1}^{p} u_l(f_l)$. It does not, however, provide the terms $u_l(f_l)$ and it is not enough to find general utility functions that work for each attribute; it is also necessary to combine them correctly. The process to do that can be decomposed into two phases:

1. build functions $\tilde{u}_l(f_l)$ that respect the corresponding trade-off condition, but do not specify the absolute scale of the utility values;
2. estimate the substitution rates, so as to rescale the $\tilde{u}_l$ functions before summing them.

In other words, first one determines a single-variable function $\tilde{u}_l(f_l)$ for each indicator; then, one determines scaling coefficients w_l in order to sum them:

$$u(f) = \sum_{l=1}^{p} w_l \tilde{u}_l(f_l)$$

where coefficients w_l yield a convex combination: $w_l \geq 0$ for each $l \in P$ and $\sum_{l=1}^{p} w_l = 1$. In general, the $\tilde{u}_l(f_l)$ functions are *normalized*, that is, they assume values that:

1. are pure adimensional numbers;
2. fall within the interval $[0, 1]$.

The main advantage of such functions is that they remove the *scale*, that is the starting value (*offset*) and the unit of the measure.

The bisection method

Building a utility function for a one-dimension impact ($F \subseteq \mathbb{R}$) is much easier than for a multidimensional one. In fact, it only takes to compare pairs of numbers and wondering which of the two is better: it can be boring, but not impossible, to give consistent answers to all the required questions.

However, the function we want to build must respect a condition stronger than the simple requirement to sort all impacts. We want the difference between the utilities of two impacts to measure the strength of the relative preference: nearly indifferent impacts should have very similar utility values, even though the values of the indicators are very different; conversely, impacts well distinguished from the preferential point of view should have very different utility values, even though the values of the indicators are very close.

The bisection method builds such a function with a dicothomic approach, by subsequent bisections. Let us suppose that we must tune the air conditioning of a room, and let us focus on the temperature. It is neither a cost nor a benefit, since the preferred value is neither the lowest nor the highest one. First, it is necessary to determine the projection on the f_1 axis (reporting the temperature) of the impact set F, fixing the extreme feasible values (for example, between $f = 16$ and $f = 30$ degrees). Now, asking a decision-maker to directly assign utility values to all impacts in F would require an excessive optimism on human capacities: we would obtain random numbers. On the other hand, if the preference relation is a weak order and the impacts do not form an infinite sequence more and more preferable (very rare in practice), the decision-maker is certainly able to indicate the best and worst impacts, that is the temperatures corresponding to the maximum and the minimum utility. The worst temperatures (that can be several) are conventionally associated to a zero utility, the best ones to a value equal to 1. For example, let $u(30) = 0$ and $u(22) = 1$. To proceed, we assume that the decision-maker is able to indicate temperatures having a utility exactly intermediate between the extreme ones. This is stronger than requiring to sort the impacts, but weaker than requiring to assign them absolute utility values.0 We assign the intermediate impacts a utility value equal to 0.5. Let us suppose that $u(27) = u(18) = 0.5$. The process continues asking to indicate impacts intermediate between the best and the intermediate ones and assigning them value 0.75. We assign the impacts halfway between the worst and the intermediate with value 0.25, and so on, progressively halving the utility intervals (0.125, 0.375, 0.625, 0.875…) When enough points have been obtained, we attempt an interpolation with an analytic expression $\tilde{u}(f)$ drawn from a suitable family of functions.

Utility proportional to the indicator

When the indicator represents a benefit, sometimes it is possible to assume that the utility grow uniformly with the value of the indicator. A common way to generate a normalized utility function in this case is the following one:

$$\hat{u}(f) = \frac{f - \min\limits_{x \in X} f(x)}{\max\limits_{x \in X} f(x) - \min\limits_{x \in X} f(x)}$$

For indicators that represent costs, one can do the same, keeping into account the opposite behaviour:

$$\hat{u}(f) = \frac{\max\limits_{x \in X} f(x) - f}{\max\limits_{x \in X} f(x) - \min\limits_{x \in X} f(x)}$$

Example 23 *Consider the decision problem with two indicators (time and cost) discussed in Example 35 of Chapter 6. Let us assume the utility to be additive and to have a linear dependence both on the time and the cost.*

Table 4.3 reports the original values of the indicators and their corresponding normalized utilities for the five alternatives of the problem. In order to build the overall utility function of the problem, it remains to determine the coefficients w_l with which to combine the normalized utilities.

	f_1 (ore)	f_2 (Euro)	$\tilde{u}_1$	$\tilde{u}_2$
Train	5.5	100	0.00	1.00
Car	4.0	150	0.33	0.83
Airplane	1.0	300	1.00	0.33
Coach	5.0	180	0.11	0.73
Taxi	4.0	400	0.33	0.00

Table 4.3: Times and costs associated to different means of transport which can be used for a trip, and their normalized utilities under the assumption of an additive and linear utility

Problems related to normalization

If the feasible region is infinite or combinatoric, quite often it is not trivial to determine the worst and the best impacts, because it requires to solve optimisation problems. In order to handle this complication, one can replace the extreme values with an underestimate of the minimum and an overestimate of the maximum. This also yields a normalized utility function, that is different from the above mentioned one, because it does not cover the whole interval $[0; 1]$. If the estimates are loose, the values of this normalized utility end up being “compressed” in a smaller range, which can create problems when combining the function with those associated to other indicators.

To be absolutely rigorous, a worst and a best impact could even not exist: the situation could degrade or improve indefinitely (which does not mean unboundedly, as the utility could anyway converge to a limit value without ever reaching it). In these cases, conventional limits could be fixed, beyond which any further improvement or worsening is simply neglected. For example, in the case of the concentration of pollutants, there will be a minimum value under which it is assumed (presumably on the basis of historical investigations) that the pollutant be innocuous, and a maximum value above which the concentration is declared absolutely unacceptable. The choice of these values, as we shall see, can influence the result of the decision process.

Also notice that the normalized utility intrinsically depends on the definition of the impact set F. This could look like an innocuous truism, as F is given. However, we know from the case studies that in complicated decision processes quite often the alternatives are not known *a priori*. New alternatives, and therefore, new impacts can appear in following phases of the decision process. If the values of the indicator in the new impacts exceed the previous maximum or minimum in F, the normalized utility must be recomputed, and this can influence the following computations. However, many other properties can be established on this point.

4.8.9 Determining the weights

Given the single components of an additive utility function, the weigths w_l with which to combine them are still to be determined. This requires to identify a sufficient number of pairs of indifferent impacts. Equalling the utilities of the two impacts imposes a constraint on the weights w_l. When the constraint system admits a single solution, this provides the values of the weights. This process is the same described in Section 4.8.2 for general utility functions, but there is a strong difference: the equation system to be solved is linear, instead of general and dependent on the family of utility functions chosen. Moreover, the coefficients w_l have a much more intuitive meaning than the parameters α_r required by the general method.

The pairwise comparison matrix

If the single-variable utility functions $\tilde{u}_l(f_l)$ have been correctly estimated, the overall utility fuction $u(f)$ has a very useful property. The indifference curves in the indicator space f are general curves, but their images in the utility space $\tilde{u}(f)$ are straight lines. This is trivially due to the fact that $u(f) = c$ corresponds to $\sum_{l=1}^{p} w_l \tilde{u}_l = c$: the transformations $\tilde{u}_l = \tilde{u}_l(f_l)$ rescale the axes so as to incorporate all nonlinearities and make them disappear.

Following this line, it is possible to define marginal rates of substitution $\tilde{\lambda}_{lm}$ between the components of the utility function, $\tilde{u}_l$, instead of between the indicators f_l. These rates of substitution are uniform and each one coincides with the ratio of the weights of the two components:

$$\tilde{\lambda}_{lm} = \frac{\dfrac{\partial u}{\partial \tilde{u}_l}}{\dfrac{\partial u}{\partial \tilde{u}_m}} = \frac{w_l}{w_m}$$

Definition 11 *We denote as* pairwise comparison matrix $\tilde{\Lambda} = \{\tilde{\lambda}_{lm}\}$ *the matrix reporting all rates of substitution between the normalized utilities.*

A very simple way to determine every rate $\tilde{\lambda}_{lm}$ is to find a pair of indifferent impacts, having the same values for all indicators except for f_l and f_m. The equation that forces their utility to be the same relates w_l and w_m by fixing their ratio, that is the marginal rate of substitution $\tilde{\lambda}_{lm} = w_l/w_m$

In $p-1$ steps, all the weights can be determined up to a multiplying factor, that will be determined by the normalization condition $\sum_l w_l = 1$. This requires to avoid pairs (l, m) for which the substitution rate is implicitly determined by other ratios. For example, once $\tilde{\lambda}_{12} = w_1/w_2$ and $\tilde{\lambda}_{23} = w_2/w_3$ are known, it is not necessary to determine $\tilde{\lambda}_{13} = w_1/w_3$, since $\tilde{\lambda}_{13} = w_1/w_3 = (w_1/w_2) \cdot (w_2/w_3) = \tilde{\lambda}_{12}\tilde{\lambda}_{23}$. In order to avoid such pairs, it is enough to build an auxiliary graph, whose vertices represent the attributes and whose edges represent the marginal substitution rates. Since every cycle in the graph corresponds to a subset of substitution rates whose overall product is equal to 1, and in which any rate coincides with the reciprocal of the product of the others, cycles should be avoided: the subset of $p-1$ computed rates should form a spanning tree. The following example applies such a process.

Example 24 *Suppose that the interviews with the decision-maker have established the validity of the mutual preferential independence assumption for a problem with four attributes. Also suppose that the following normalized utility functions have been determined with the bisection method:*

$$\tilde{u}_1 = 2f_1 + \frac{1}{2}, \text{ with } f_1 \in \left[-\frac{1}{4}; \frac{1}{4}\right] \qquad \tilde{u}_2 = \frac{\log_2 f_2 - 2}{8}, \text{ with } f_2 \in [4; 1024]$$

$$\tilde{u}_3 = \frac{20 - f_3}{15}, \text{ with } f_3 \in [5; 20] \qquad \tilde{u}_4 = \frac{\sqrt{f_4}}{6}, \text{ with } f_4 \in [0; 36]$$

Finally, suppose that the following $p - 1 = 3$ *pairs of indifferent impacts have been identified:*

$$A = (0, 8, 10, 9) \sim B = \left(\frac{1}{4}, 8, 10, 4\right)$$

$$C = \left(-\frac{1}{4}, 256, 10, 0\right) \sim D = \left(-\frac{1}{4}, 256, 20, 16\right)$$

$$E = (0, 4, 5, 25) \sim F = \left(-\frac{1}{4}, 64, 5, 25\right)$$

Notice that the two impacts of each selected pair differ only for two of the four indicators: A and B differ for the first and fourth indicator, C and D for the third and fourth, E and F for the first two. This is not rigorously necessary, but it simplifies the computations.

First of all, we have to convert the impact into the corresponding utility functions.

$$\tilde{u}_A = \left(\frac{1}{2}, \frac{1}{8}, \frac{2}{3}, \frac{1}{2}\right) \quad \tilde{u}_B = \left(1, \frac{1}{8}, \frac{2}{3}, \frac{1}{3}\right)$$
$$\tilde{u}_C = \left(0, \frac{3}{4}, \frac{2}{3}, 0\right) \quad \tilde{u}_D = \left(0, \frac{3}{4}, 0, \frac{2}{3}\right)$$
$$\tilde{u}_E = \left(\frac{1}{2}, 0, 1, \frac{5}{6}\right) \quad \tilde{u}_F = \left(0, \frac{1}{2}, 1, \frac{5}{6}\right)$$

Additivity allows to express the overall utility as a convex combination of the single components:

$$u = w_1\tilde{u}_1 + w_2\tilde{u}_2 + w_3\tilde{u}_3 + w_4\tilde{u}_4$$

from which the three equalities

$$\begin{cases} u\,(A) = u\,(B) \Rightarrow \frac{1}{2}w_1 + \frac{1}{8}w_2 + \frac{2}{3}w_3 + \frac{1}{2}w_4 = 1w_1 + \frac{1}{8}w_2 + \frac{2}{3}w_3 + \frac{1}{3}w_4 \\ u\,(C) = u\,(D) \Rightarrow 0w_1 + \frac{3}{4}w_2 + \frac{2}{3}w_3 + 0w_4 = 0w_1 + \frac{3}{4}w_2 + 0w_3 + \frac{2}{3}w_4 \\ u\,(E) = u\,(F) \Rightarrow \frac{1}{2}w_1 + 0w_2 + 1w_3 + \frac{5}{6}w_4 = 0w_1 + \frac{1}{2}w_2 + 1w_3 + \frac{5}{6}w_4 \end{cases} \Rightarrow$$

$$\Rightarrow \begin{cases} \frac{1}{6}w_4 = \frac{1}{2}w_1 \\ \frac{2}{3}w_3 = \frac{2}{3}w_4 \\ \frac{1}{2}w_1 = \frac{1}{2}w_2 \end{cases} \Rightarrow \begin{cases} w_4 = 3w_1 \\ w_3 = w_4 \\ w_1 = w_2 \end{cases}$$

Since the impacts have been chosen with only two different indicators, the equalities relate two weights at a time. In this way, each one determines a rate of substitution $\tilde{\lambda}_{lm} = w_l/w_m$. For example, $\tilde{\lambda}_{41} = 3$. Notice that weight w_4 is larger than weight w_1, suggesting that the utility $\tilde{u}_4$ is more relevant than utility $\tilde{u}_1$. In fact, moving from impact A to impact B, utility $\tilde{u}_4$ varies less than utility $\tilde{u}_1$. If we directly considered indicators f_4 and f_1, we would notice exactly the opposite (the value of f_4 varies strongly with respect to that of f_1), but the indicators depend on the units of measure and they influence the utility in a different way in different impacts.

Adding the normalization condition to the equalities determined by the pairs of indifferent impacts allows to determine the weights:

$$w_1 + w_2 + w_3 + w_4 = 1 \Rightarrow w_2 + w_2 + 3w_2 + 3w_2 = 1 \Rightarrow w_2 = \frac{1}{8}$$

which implies

$$w = \left[\ \frac{1}{8}\ \frac{1}{8}\ \frac{3}{8}\ \frac{3}{8}\ \right]^T$$

Since the pairwise comparison matrix is composed of the ratios w_l/w_m of the weights of the normalized utilities, it necessarily enjoys the following properties:

1. *positivity*: the rates of substitution between normalized utilities are positive
$$\tilde{\lambda}_{lm} > 0 \quad \text{per ogni } l, m \in P$$

2. *reciprocity*: the rate of substitution of a utility with respect to another one is the reciprocal of the rate of the second with respect to the first one
$$\tilde{\lambda}_{lm} = \frac{1}{\tilde{\lambda}_{ml}} \quad \text{per ogni } l, m \in P$$

3. *consistency*: two rates of substitution with "sequenced" indices determine the third one
$$\tilde{\lambda}_{ln} = \tilde{\lambda}_{lm}\tilde{\lambda}_{mn} \quad \text{per ogni } l, m, n \in P$$

The definition builds the elements of matrix $\tilde{\Lambda}$ starting from the weights. In practice, the converse occurs: the elements of the matrix are empirically measured as described in the example, whereas the weights are unknown, and must be reconstructed starting from the former. The following properties allow to do that without solving a linear system of equations.

Proposition 1 *Matrix $\tilde{\Lambda}$ has:*

- *rank 1 and $p-1$ eigenvalues equal to zero;*
- *a main diagonal of elements equal to 1 and a dominant eigenvalue $\mu^{\max} = p$;*
- *a dominant eigenvector $x^{\max}$ proportional to the weight vector.*

Proof. The ratio between corresponding elements of columns m and n is $\tilde{\lambda}_{mn}$, that is the same for all elements. In fact, consistency implies:
$$\frac{\tilde{\lambda}_{ln}}{\tilde{\lambda}_{lm}} = \tilde{\lambda}_{mn} \quad \text{for each } l \in P$$

Since all columns are proportional, the rank of the matrix is 1. Moreover, it coincides with the number of nonzero eigenvalues, so that $p-1$ eigenvalues are null. The dominant eigenvalue $\mu^{\max}$ coincides therefore with the sum of all eigenvalues, which on its turn coincides with the *trace* of the matrix, that is with the sum of the p elements of the main diagonal. These are all equal to 1, because reciprocity imposes $\tilde{\lambda}_{lm}^2 = 1$ and positivity forbids the negative solution. Finally, the dominant eigenvector $x^{\max}$ solves the equation:
$$\tilde{\Lambda} x^{\max} = \mu^{\max} x^{\max} \Rightarrow \sum_{m=1}^{p} \frac{w_l}{w_m} x_m^{\max} = p x_l^{\max} \text{ for each } l \in P \Rightarrow$$
$$\Rightarrow \frac{x_l^{\max}}{w_l} = \frac{1}{p} \sum_{m=1}^{p} \frac{x_m^{\max}}{w_m} \text{ for each } l \in P$$

so that $x_l^{\max}/w_l$ does not depend on l: vector $x^{\max}$ is proportional to w. ■

Remark 3 *In order to find the weight vector w, it is not necessary to determine the dominant eigenvector: it is enough to normalize any column of $\tilde{\Lambda}$.*

Proof.

$$\frac{\tilde{\lambda}_{lm}}{\sum\limits_{l=1}^{p} \tilde{\lambda}_{lm}} = \frac{\frac{w_l}{w_m}}{\sum\limits_{i=1}^{p} \frac{w_i}{w_m}} = \frac{w_l}{\sum\limits_{i=1}^{p} w_i} = w_l$$

■

The process that computes the weights by normalizing the elements of column l corresponds to using as a spanning tree the star centred in the vertex corresponding to indicator f_l.

Example 25 *A decision-maker has been asked to indicate the rates of substitution between three normalized utility functions $\tilde{u}_1$, $\tilde{u}_2$ and $\tilde{u}_3$. The answers of the decision-maker generate the following matrix.*

$$\tilde{\Lambda} = \begin{array}{c|ccc|} & \tilde{u}_1 & \tilde{u}_2 & \tilde{u}_3 \\ \cline{2-4} \tilde{u}_1 & 1 & 2 & 6 \\ \tilde{u}_2 & 1/2 & 1 & 3 \\ \tilde{u}_3 & 1/6 & 1/3 & 1 \\ \cline{2-4} \end{array}$$

It can be verified that the matrix is positive, reciprocal and consistent. Therefore, it is possible to build from it an additive utility function $u = \sum_{l=1}^{p} w_l \tilde{u}_l$. The weight vector is $w = [0.6\ 0.3\ 0.1]^T$, and it coincides, up to a multiplying factor, with each of the columns of $\tilde{\Lambda}$ (the constant is $6/10$ for the first column, $3/10$ for the second one and $1/10$ for the third).

Studying the characteristic polynomial of $\tilde{\Lambda}$

$$\left|\mu I - \tilde{\Lambda}\right| = \begin{vmatrix} \mu - 1 & -2 & -6 \\ -1/2 & \mu - 1 & -3 \\ -1/6 & -1/3 & \mu - 1 \end{vmatrix} = (\mu-1)^3 - 1 - 1 - (\mu-1) - (\mu-1) - (\mu-1) =$$

$$= \mu^3 - 3\mu^2 + 3\mu - 1 - 1 - 1 - 3\mu + 3 = \mu^3 - 3\mu^2 = \mu^2 (\mu - 3)$$

we verify that two eigenvalues are equal to zero, and that the dominant one is equal to $p = 3$. The corresponding eigenvector solves the system of equtions:

$$\left(3I - \tilde{\Lambda}\right) x = 0 \Rightarrow \begin{bmatrix} 3-1 & -2 & -6 \\ -1/2 & 3-1 & -3 \\ -1/6 & -1/3 & 3-1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 2x_1 - 2x_2 - 6x_3 \\ -1/2x_1 + 2x_2 - 3x_3 \\ -1/6x_1 - 1/3x_2 + 2x_3 \end{bmatrix} = 0$$

and is proportional to the weight vector w.

4.8.10 The process in summary

Summarising, the process to determine the utility function is composed of the following steps:

1. interview the decision-maker in order to understand whether mutual preference independence holds (asking him/her to compare different pairs of indicators with the complementary subset);
2. in the positive case, focus on each attribute, fixing the other ones to any value, and build a single-variable component of the utility function with a method respecting the entity of the relative preference gaps;
3. determine the weights with which to combine such functions, sampling the impacts until a sufficient number of pairs of indifferent impacts has been found;

4. check *a posteriori* that the obtained utility function is valid, testing it on impacts not used before.

In any moment, it can be necessary to go back to correct the single components of the utility or even rejecting the additivity assumption, in which case the whole process is invalidated.

4.9 Exercises

Exercise 1

A community centre is looking for a new location: there are four alternatives (A, B, C and D) besides alternative 0 (staying in the present location). It has been decided that the choice among the five alternative should be definitive and that it shall be taken based on three factors: space, accessibility and prestige of the location. The following table provides the values of the indicators associated to each alternative, that must be considered as benefits.

Indicators	A	B	C	D	0
Space	90	90	90	0	100
Accessibility	12	13	100	100	37
Prestige	30	0	5	100	10

Sort the alternatives based on:

- the lexicographic method with respect to sequence $P = (2, 3, 1)$;
- the lexicographic method with aspiration levels, with respect to sequence $P = (2, 3, 1)$ with levels $\epsilon_1 = 5$ and $\epsilon_3 = 10$;
- the utopia point method with the Manhattan distance;
- the utility function $u(f) = w_1 f_1 + w_2 f_2 + w_3 f_3$ with weights obtained assuming that $(50, 100, 0) \sim (50, 20, 40)$ and $(40, 60, 20) \sim (60, 40, 20)$.

For each of the previous cases, build the associated Borda count.

Solution

The lexicographic method sorts all the alternatives with respect to the first indicator in the given sequence. Ties are solved considering the second indicator, then the third, and so on. Contrary to the assumption made in Section 4.5, the indicator here must be considered as benefits, instead of costs. In the specific case, indicator 2 yields the nontotal order $C \sim D \prec 0 \prec B \prec A$. Indicator 3 allows to order the indifferent alternatives, obtaining total order: $D \prec C \prec 0 \prec B \prec A$. The corresponding Borda count is: $B(A) = 1$ (because A is weakly preferred only to itself), $B(B) = 2$ (because B is weakly preferred to itself and to A), $B(C) = 4$, $B(D) = 5$ and $B(0) = 3$.

Since the indicators are benefits, the given aspiration levels require $f_1(x) \geq \epsilon_1$ e $f_3(x) \geq \epsilon_3$. This removes alternatives B and C because they are unsatisfactory with respect to ϵ_3 and alternative D because it is unsatisfactory with respect to ϵ_1. The lexicographic order of the remaining alternatives is determined only by the second indicator and is: $0 \prec A$. The corresponding Borda count is: $B(A) = 1$ and $B(0) = 2$.

The utopia point for the given problem is $f^\circ = (100, 100, 100)$. The Manhattan distances for the alternatives are reported in the following table.

	A	B	C	D	0
Distance	168	197	105	100	153

For example, the distance of A from f° is $d(A, f^\circ) = |90 - 100| + |12 - 100| + |30 - 100| = 10 + 88 + 70 = 168$. As a consequence, the order of the alternatives is $D \prec C \prec 0 \prec A \prec B$, because better alternatives have smaller distance from the utopia point. The corresponding Borda count is: $B(A) = 2$, $B(B) = 1$, $B(C) = 4$, $B(D) = 5$ e $B(0) = 3$.

In theory, the components of the utility function should be normalised before combining them, but the text does not require it, probably because the values reported in the table are already normalised. That they are normalised between 0 and 100, instead of between 0 and 1, of course, does not change anything. The two pairs of indifferent impacts are sufficient (and strictly necessary) to determine the weights, together with the normalisation condition on the weights themselves:

$$\begin{cases} u\,(50, 100, 0) = u\,(50, 20, 40) \\ u\,(40, 60, 20) = u\,(60, 40, 20) \\ w_1 + w_2 + w_3 = 1 \end{cases} \Rightarrow \begin{cases} 50w_1 + 100w_2 + 0w_3 = 50w_1 + 20w_2 + 40w_3 \\ 40w_1 + 60w_2 + 20w_3 = 60w_1 + 40w_2 + 20w_3 \\ w_1 + w_2 + w_3 = 1 \end{cases}$$

from which

$$\begin{cases} 100w_2 = 20w_2 + 40w_3 \\ 40w_1 + 60w_2 = 60w_1 + 40w_2 \\ w_1 + w_2 + w_3 = 1 \end{cases} \Rightarrow \begin{cases} 2w_2 = w_3 \\ w_2 = w_1 \\ w_1 + w_2 + w_3 = 1 \end{cases}$$

which implies $w_3 = 2w_1$, $w_2 = w_1$ and $w_1 + w_1 + 2w_1 = 1$, that is $w_1 = w_2 = 1/4$ and $w_3 = 1/2$. The following table reports the values of the utility function for all alternatives.

	A	B	C	D	0
u	40.5	25.75	50	75	39.25

For example, $u\,(f\,(A)) = 1/4 \cdot 90 + 1/4 \cdot 12 + 2/4 \cdot 30 = 162/4$. As a consequence, the order of the alternatives is $D \prec C \prec A \prec 0 \prec B$. The corresponding Borda count is: $B(A) = 3$, $B(B) = 1$, $B(C) = 4$, $B(D) = 5$ e $B(0) = 2$. Notice that the Borda count and the utility function u are equivalent, even if the transformation that relates them has no simple analitic form.

Exercise 2

Given the following decision problem

$$\begin{aligned} \min f_1(x) &= x^2 - 4x \\ \min f_2(x) &= -x^2 \\ 0 \leq x &\leq 3 \end{aligned}$$

find and represent the utopia point U in the space of the indicators and determine which between solutions $x' = 0$ e $x'' = 1$ is preferable with respect to the L_1 (Manhattan), L_2 (Euclidean) and L_∞ distance from U.

Solution

In order to determine the utopia point, one must solve separately the problems $\min f_1(x) = x^2 - 4x$ con $0 \leq x \leq 3$ and $\min f_2(x) = -x^2$ con $0 \leq x \leq 3$. Since they are problems in a single variable, the student is expected to be able to solve them, obtaining $f_1^\circ = -4$ (for $x = 2$) and $f_2^\circ = -9$ (for $x = 3$). Therefore, the utopia point is $f^\circ = (-4, -9)$.

The impacts corresponding to x' and x'' are $f(x') = (0, 0)$ e $f(x'') = (-3, -1)$. The distances indicated in the text are reported in the following table.

Distance from f°	$f(x')$	$f(x'')$
L_1	13	9
L_2	$\sqrt{97}$	$\sqrt{65}$
L_∞	9	8

According to all three distances, the preferable solution is x''.

Exercise 3

Given the following decision problem:

$$\begin{aligned} &\min f_1(x) = x_1^2 + x_2^2 - 2x_1 \\ &\min f_2(x) = -x_2 \\ &x_1^2 + 4x_2^2 \leq 8 \\ &x_1 - 2x_2 \geq 0 \end{aligned}$$

indicate the solution preferred by a decision-maker with the lexicographic preference determined by the index sequence $P = (2, 1)$.

Apply again the method assuming an aspiration level equal to $\epsilon_1 = 1/4$.

Determine the utopia point U and the preferred solution chosen according the utopia point method based on the Manhattan distance (L_1), the Euclidean distance (L_2) and the $L_{+\infty}$ distance, not in the entire feasible region, but only among the three following alternatives: $A = (0, 0)$, $B = (2, 1/2)$ e $C = (1/2, -1)$.

Solution

The problem can be solved graphically or applying the techniques described in Chapter 5. Figure 4.10 shows the feasible region X.

FIGURE TO BE ADDED: ELLIPSIS CENTRED IN THE ORIGIN WITH SEMIAXES OF LENGTH $2\sqrt{2}$ AND $\sqrt{2}$, LINE THROUGH THE ORIGIN WITH SLOPE 1/2, f_1 POINTS TOWARDS $(1, 0)$, f_2 UPWARDS

Figure 4.10: Feasible region for the problem of Exercise 3

The lexicographic method orders completely the indicators and determines all the optimal solutions with respect to the first indicator. In this case, $\min f_2(x) = -x_2$ corresponds to maximising x_2. Graphically, we obtain a single globally optimal point: $x^\circ = C = (2, 1)$, so it is not necessary to proceed with the optimisation of $f_1(x)$ in set X_2°.

The method with aspiration levels adds to the problem constraints that guarantee acceptable values for the secondary indicators. Since f_1 must be minimised, the associated constraint is $f_1(x) = x_1^2 + x_2^2 - 2x_1 \leq \epsilon_1 = 1/4$, that can be rewritten as $(x_1 - 1)^2 + x_2^2 \leq 1/4$, and therefore imposes to the solutions to have a distance not larger than 1/2 from point $(1, 0)$. The solution obtained graphically is point $(1, 1/2)$.

The utopia point is obtained optimising separately the indicators. The two optimal values are $f_1^\circ = 1$, obtained in $(1,0)$, and $f_2^\circ =$, obtained in $(2,1)$. It is not actually necessary to compute the optimal solutions: the value of the optimum is enough, because the utopia point is in the indicator space. Therefore, the utopia point is $U = (-1,-1)$. The three solutions indicated in the text generate the following impacts: $f(A) = (0,0)$, $f(B) = (1/4,-1/2)$ and $f(C) = (1/2,-1)$. Their distances from the utopia point are reported in the following table.

Distance from f°	$f(A)$	$f(B)$	$f(C)$
L_1	2	7/4	3/2
L_2	$\sqrt{2}$	$\sqrt{29}/4$	3/2
L_∞	1	5/4	3/2

The preferred solution for L_1 is C, for L_2 is B and for L_∞ is A.

Exercise 4

The following table represents the performance of five alternatives with respect to four decision criteria (all of them to be maximised), in a scale from 0 to 100.

	a_1	a_2	a_3	a_4	a_5
f_1	100	70	60	40	20
f_2	60	45	40	100	80
f_3	60	25	20	80	100
f_4	20	100	90	50	40

Which alternative is the best one if the indifference curves belong to the family:

$$w_1 f_1 + w_2 f_2 + w_3 f_3 + w_4 f_4 = c \text{ with } w_i = 0.25 \text{ for } i = 1, \dots, 4$$

How much should the value of w_1 increase (keeping the values of the other weights constant) so that a_1 become the best alternative? And the value of w_4?

Solution

Indicator f_1 is considered a benefit, contrary to the assumption made in Section 4.5. With respect to it, the alternatives follow the order: $a_1 \sim a_2 \sim a_3 \prec a_4 \prec a_5$. This order can be refined considering first indicator f_4 ($a_2 \sim a_3 \prec a_1$), then indicator f_3 ($a_2 \prec a_3$), obtaining the lexicographic order: $a_2 \prec a_3 \prec a_1 \prec a_4 \prec a_5$.

The aspiration levels impose the following constraints: $f_4 \geq \epsilon_4 = 30$, $f_3 \geq \epsilon_3 = 30$ e $f_2 \geq \epsilon_2 = 30$, removing solutions a_1, a_2 e a_3. This generates the order: $a_4 \prec a_5$.

The given indifference curves (hypersurfaces) imply that $f_1 + f_2 + f_3 + f_4 = c$. Since by definition an indifference curve is the set of impacts in which the utility function assumes a constant value, one can write $u(f) = f_1 + f_2 + f_3 + f_4$, that is already additive. In order to normalise this expression, it is enough to normalise the weights of the single components, that are all equal at first: $u(f) = \sum_{l=1}^{p} w_l f_l$ con $\sum_{l=1}^{p} w_l = 1$. Since $w_l = w$ for all $l \in P$, we have $w_l = 1/4$, and therefore

$$u(f) = \frac{1}{4} f_1 + \frac{1}{4} f_2 + \frac{1}{4} f_3 + \frac{1}{4} f_4$$

This function determines the following utility values for the alternatives:

	a_1	a_2	a_3	a_4	a_5
$u(f(a))$	60	61.25	65	67.5	60

so that the order is $a_4 \prec a_3 \prec a_2 \prec a_1 \sim a_5$.

Exercise 5

Given the following decision problem:

$$\begin{aligned} \max f_1(x) &= x_1 - 3x_2 \\ \max f_2(x) &= -4x_1 + x_2 \\ -2x_1 + 2x_2 &\leq 7 \\ 2x_1 + 2x_2 &\leq 11 \\ x_1 &\leq 4 \\ x_1, x_2 &\geq 0 \end{aligned}$$

determine its optimal assuming that the decision-maker has a uniform marginal rate of substitution of 4 units of f_1 for 1 unit of f_2.

Solution

The marginal rate of substitution between two indicators is the ratio between the variation of the second indicator and the corresponding variation of the first one along an indifference curva (keeping all the other possible indicators constant). In the specific case, therefore $\mu_{12} = 1/4$. A uniform rate corresponds to a linear utility function $u(f) = w_1 f_1 + w_2 f_2$ con $\mu_{12} = \frac{\partial u}{\partial f_1} / \frac{\partial u}{\partial f_2}$, that is

$$\mu_{12} = \frac{\frac{\partial u}{\partial f_1}}{\frac{\partial u}{\partial f_2}} = \frac{w_1}{w_2} = \frac{1}{4}$$

from which $w_2 = 4w_1$. The normalisation condition $w_1 + w_2 = 1$ implies $w_1 = 1/5$ e $w_2 = 4/5$, that is $u(f) = f_1 + 4f_2$. Graphically, one can obtain the optimal solution $x^* = (0, 7/2)$.

Exercise 6

Given the following decision problem:

$$\begin{aligned} \min f_1(x) &= x_1^2 + x_2^2 \\ \max f_2(x) &= x_2 \\ 0 \leq x_2 &\leq 10 \end{aligned}$$

determine the utopia point, considering the two indicators as independent.

Determine the optimal solution with respect to an additive utility function whose component associated to $f_1(x)$ is:

$$u_1(f_1) = \begin{cases} 1 - \dfrac{f_1}{200} & \text{per } 0 \leq f_1 \leq 200 \\ 0 & \text{per } f_1 \geq 200 \end{cases}$$

that is, linearly decreasing to zero until $f_1 = 200$ and constant after that. The component associated to f_2 is equal to $u_2(f_2) = f_2/10$. The two components have equal weight.

Solution

Considering the two indicators as independent, the utopia point corresponds to the minimum feasible value for f_1 and the maximum one for f_2, therefore $f^\circ = (0, 10)$.

The utility function is $u(f) = \frac{1}{2}u_1(f_1) + \frac{1}{2}u_2(f_2)$. In order to deal with it, it is advisable to solve the problem separately in the two parts in which its definition divides the feasible region:

- for $f_1(x) = x_1^2 + x_2^2 \geq 200$, the utility function is:

  $$u(x) = \frac{1}{2} \cdot 0 + \frac{1}{2} \cdot \frac{x_2}{10}$$

  that has maximum value in $x_2 = 10$ for any feasible x_1, that is $x_1 \leq -10$ or $x_1 \geq 10$. The corresponding value is equal to $u(x') = \frac{1}{2}$.

- for $f_1(x) = x_1^2 + x_2^2 \leq 200$, the utility function is:

  $$u(x) = \frac{1}{2} \cdot \left(1 - \frac{f_1}{200}\right) + \frac{1}{2} \cdot f_2 = \frac{1}{2} \cdot \left(1 - \frac{x_1^2}{200} - \frac{x_2^2}{200}\right) + \frac{1}{2} \cdot \frac{x_2}{10}$$

  Since x_1 is unconstrained, we can maximise the utility setting $x_1 = 0$, obtaining

  $$u(x) = \frac{1}{2} \cdot \left(1 - \frac{x_2^2}{200}\right) + \frac{1}{2} \cdot \frac{x_2}{10}$$

  that has maximum value in $x_2 = 2$, equal to $u^* = \frac{3}{4}$.

The globally optimal point, therefore, is $x^* = (0, 10)$.

Exercise 7

The following decision problem has three alternatives and four indicators:

Indicatori	a_1	a_2	a_3
f_1	100	0	20
f_2	10	10	100
f_3	30	80	0
f_4	60	0	80

Determine the best alternative with the lexicographic method with respect to the sequence of indicators $P = (2, 1, 3, 4)$, considering them as costs.

Apply again the method introducing the aspiration levels $\epsilon_1 = 90$, $\epsilon_3 = 85$ and $\epsilon_4 = 70$.

Determine the best alternative with respect to the additive utility function that combines components equal to the indicators ($u_l = f_l$) or to their squares ($u_l = f_l^2$) with the following weights: $w_1 = 0.25$, $w_2 = 0.30$, $w_3 = 0.40$ e $w_1 = 0.05$. For the sake of simplicity, we use nonnormalised utilities.

Solution

The lexicographic method first uses indicator f_2 to sort the alternatives: $a_1 \sim a_2 \prec a_3$. Since there are equivalent alternatives, we use f_1 to distinguish them: $a_2 \prec a_1$. Overall, the final order is$a_2 \prec a_1 \prec a_3$.

Since the indicators represent costs, the aspiration levels turn into constraints: $f_1 \leq \epsilon_1 = 90$ (that forbids alternative a_1), $f_3 \leq \epsilon_3 = 85$ and $f_4 \leq \epsilon_4 = 70$ (that forbids alternative a_3). Only alternative a_2 remains: in this case, the aspiration levels do not modify the choice.

The following table reports the utility values for the three alternatives based on the two indicated functions.

Utility	a_1	a_2	a_3
$u(f) = \sum_{l \in P} w_l f_l$	-43	-35	-39
$u(f) = \sum_{l \in P} w_l f_l^2$	-3070	-2590	-3420

Hence, the first function determines the order $a_2 \prec a_3 \prec a_1$, while the second determines the order $a_2 \prec a_1 \prec a_3$.

Exercise 8

Given the following evaluation matrix:

Indicatori	a_1	a_2	a_3
f_1	80	20	60
f_2	10	90	50

determine the ordering of the alternatives with the utopia point method using the L_1, L_2 and L_∞ distances and considering the values of the indicators as costs.

Consider an additive utility function in which components associated to the two indicators are the indicators themselves with the opposite sign. Determine the order of the alternatives normalising the two components with respect to the extreme values assumed by the indicators in the feasible region and combining them with weights $w_1 = 0.7$ and $w_2 = 0.3$.

Solution

The utopia point is $f^\circ = (20, 10)$ and its distances from the alternatives are indicated in the next table.

Distances	a_1	a_2	a_3
L_1	60	80	80
L_2	60	80	56.6
L_∞	60	80	40

The order of the alternatives is therefore:

- $a_1 \prec a_2 \sim a_3$ for L_1;
- $a_3 \prec a_1 \prec a_2$ for L_2 and L_∞.

The normalised components of the utility function:

$$\tilde{u}_l = \frac{\max\limits_{x \in X} f_l(x) - f_l(x)}{\max\limits_{x \in X} f_l(x) - \min\limits_{x \in X} f_l(x)}$$

are reported in the following table:

	a_1	a_2	a_3
$\tilde{u}_1(x)$	0	1	1/3
$\tilde{u}_2(x)$	1	0	1/2

so that the alternatives have utility $\tilde{u}(a_1) = 0.3$, $\tilde{u}(a_2) = 0.7$ e $\tilde{u}(a_3) = 23/60 = 0.38\overline{3}$, and their order is $a_2 \prec a_3 \prec a_1$.

Chapter 5

Mathematical Programming

This chapter deals with the simplest class of decision problems:

- the system is deterministic, that is there is a single scenario:

  $$\Omega = \{\bar{\omega}\}$$

  and, since the exogeneous variables are constant ($\omega_k = \bar{\omega}_k$), the impact f only depends on the alternative x and can be indicated simply as $f(x)$;

- there is a single decision-maker:

  $$|D| = 1$$

- the preference relation Π admits a known consistent value function:

  $$\exists v : F \to \mathbb{R} : \Pi = \{(f, f') \in F \times F : v(f) \geq v(f')\}$$

This class of problems is mainly treated in the mathematical literature, where it is more common to express preference with a cost function, instead of a value function. The only difference is that such a function associates the preferred impacts to smaller values, instead of larger ones and it must therefore be minimized, instead of maximized. Since the cost can be defined as the opposite of the value ($c = -v$), we could formulate the problem as the minimisation of $c(x) = -v(f(x, \bar{\omega}))$ with respect to $x \in X$. For the sake of briefness, we will directly denote with f the cost (which is not in general equal to the impact f: this is an abuse of notation), and we will write

$$\begin{array}{c} \min f(x) \\ \quad\quad x \in X \end{array}$$

We will also make an additional assumption on the structure of the solution set X, that is, that it can be described through a finite system of inequalities:

$$X = \{x \in \mathbb{R}^n : g_j(x) \leq 0 \text{ for } j = 1, \ldots, m\}$$

where $g_j : X \to \mathbb{R}$ e $g_j \in C^1(X)$ for $j = 1, \ldots, m$, that is, the constraint functions g_j are real-valued continuous functions with a continuous first derivative in the feasible solution set of the problem. Finally, we will assume that also the cost function be continuous with its first derivative in the same set: $f \in C^1(X)$.

This class of problems (sometimes with more or less strict assumptions on the continuity of functions and derivatives) are denoted under the large label of *Mathematical Programming* problems. The reason is that describing both the preference and the feasible region through real-valued functions allows to apply to these decision problems the whole machinery of mathematical analysis.

5.1 Basic concepts

Definition 12 *Given a set $X \subseteq \mathbb{R}^n$ and a function $f : X \rightarrow \mathbb{R}$, we denote as* global optimum point *a point $x^\circ \in X$ such that*

$$f(x^\circ) \leq f(x) \quad \textit{for all } x \in X$$

As well, we denote by X° the set of all global optimum points:

$$X^\circ = \arg\min_{x \in X} f(x)$$

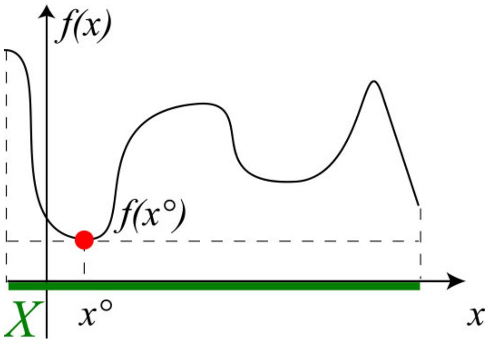


Figure 5.1: Graph of a function $f(x)$ with a feasible region X and the global optimum point x°

Figure 5.1 shows an example of function defined in an interval, with a global optimum point. From the remarks of Section 4.1, it clearly derives that the solutions of a decision problem falling under the class of Mathematical Programming are the global optimum points. As it is not easy to identify such points, we consider a second, larger, set of points, for which it is possible to develop computational methods.

Definition 13 *Given a set $X \subseteq \mathbb{R}^n$ and a function $f : X \rightarrow \mathbb{R}$, we denote as* local optimum point *a point $x^* \in X$ such that*

$$\exists \epsilon > 0 : f(x^*) \leq f(x) \quad \textit{for all } x \in X \cap \mathcal{U}_{x^*,\epsilon}$$

where $\mathcal{U}_{x^*,\epsilon} = \{x \in \mathbb{R}^n : \|x - x^*\| < \epsilon\}$ is a neighbourhood *of x^* of width ϵ. As well, we denote by X^* the set of all local optimum points.*

Remark 4 *A global optimum point is by definition also a local optimum point:*

$$X^\circ \subseteq X^*$$

Figure 5.2 shows an example of function defined in an interval, with several local optimum points, one of which is also a global optimum point.

Since there is no general method to identify neither X^* nor X°, we will search for necessary conditions for local optimality. These conditions identify points that are candidate to being local optimum points, and therefore global optimum points. In this way, we weaken twice our request:

$$\begin{array}{ccccc} \text{Global optimum} & \Rightarrow & \text{Local optimum} & \Rightarrow & \text{KKT conditions} \\ X^\circ & \subseteq & X^* & \subseteq & X^{\text{KKT}} \end{array}$$

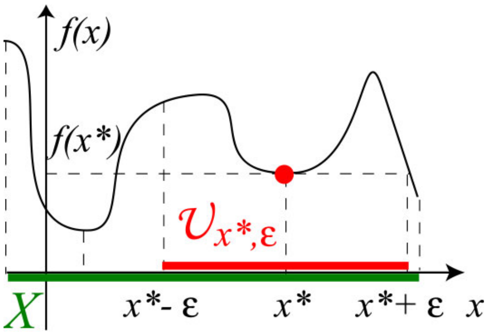


Figure 5.2: Graph of a function $f(x)$ with a feasible region X and a local optimum point x^* (the other two points are x° and the maximum of set X)

The conditions we will tackle are known as *Karush-Kuhn-Tucker conditions* (in short, *KKT conditions*, from the names of their discoverers[1].

We will proceed as follows:

1. lay out the conditions as a system of equalities and inequalities;
2. solve the system, to find the set of candidate points X^{KKT};
3. evaluate the points of X^{KKT} one-by-one, keeping only the best ones.

The best points of X^{KKT} will yield X°. The method requires X^{KKT} to be a finite set, or at least a set which can be analytically described, in order to determine X°.

The KKT conditions are based on approximating the objective function and the constraint functions with linear functions of the decision variables, and on deducing necessary conditions that such approximating functions must satisfy in a given point $\tilde{x}$ so that $\tilde{x}$ be a local optimum point. Appendix A recalls the basic concepts on mathematical analysis that are necessary to follow the derivation of the KKT conditions.

5.1.1 Taylor's series expansion

Every sufficiently regular function (that is, a continuous function with continuous derivatives up to a suitable order k) can be locally approximated with a polynomial of degree k.

Theorem 11 *Let $f \in C^k(\mathcal{U}_{\tilde{x},\epsilon})$ be a function of a real variable $x \in \mathbb{R}$. For all $x \in \mathcal{U}_{\tilde{x},\epsilon}$,* Taylor's series expansion *holds:*

$$f(x) = \sum_{i=0}^{k} \frac{f^{(i)}(\tilde{x})}{i!}(x-\tilde{x})^i + R_k(x-\tilde{x})$$

with $f^{(0)} = f$, $f^{(i)} = \dfrac{d^i f}{dx^i}$ for all $i \in \mathbb{N}^+$, and $\lim\limits_{x \to \tilde{x}} \dfrac{R_k(x-\tilde{x})}{\|x-\tilde{x}\|^k} = 0$.

[1] William Karush (1917-1997) discovered the conditions in 1939 in his Master's degree thesis, without publishing them; Harold William Kuhn (1925-2014) and Albert William Tucker (1905-1995) rediscovered them in 1951, when nobody yet knew Karush' work.

The condition on the limit of R_k means that the approximation error incurred considering only the first $k+1$ terms, that is the polynomial of degree k, is small for $x \approx \tilde{x}$, and moreover tends to zero more quickly than x tends to $\tilde{x}$. This property will allow to ignore such an error in many cases when using such a limit.

We will only use the expansion up to $k = 1$. Any function that is regular up to the first order admits a linear approximation:

$$f(x) = f(\tilde{x}) + f'(\tilde{x})(x - \tilde{x}) + R_1(|x - \tilde{x}|)$$

which can be generalized from a single-variable function to multiple-variable functions ($x \in \mathbb{R}^n$) writing:

$$f(x) = f(\tilde{x}) + (\nabla f(\tilde{x}))^T (x - \tilde{x}) + R_1(\|x - \tilde{x}\|) \qquad (5.1)$$

5.1.2 Directions

Given a point $\tilde{x} \in \mathbb{R}^n$, a vector $d \in \mathbb{R}^n$ is named *direction* if it is interpreted as a term to add to $\tilde{x}$ in order to move to other points; its length can be tuned with a suitable coefficient α so as to decide how much to move.

Definition 14 *A* feasible direction *in $\tilde{x} \in X$ is a vector $d \in \mathbb{R}^n$ such that*

$$\exists \bar{\alpha} > 0 : \tilde{x} + \alpha d \in X \quad \forall \alpha \in [0; \bar{\alpha})$$

In other words, moving from $\tilde{x}$ little enough (less than $\bar{\alpha}$) towards d, the solution reached will always be feasible. Figure 5.3 shows that the same direction $d = [-1\ 2]^T$ is feasible in point $(0, 2)$ and unfeasible in point $(2, 0)$.

Definition 15 *An* improving direction *in $\tilde{x} \in X$ for $f(\cdot)$ is a vector $d \in \mathbb{R}^n$ such that*

$$\exists \bar{\alpha} > 0 : f(\tilde{x} + \alpha d) < f(\tilde{x}) \quad \forall \alpha \in (0; \bar{\alpha})$$

In other words, moving from $\tilde{x}$ little enough (less than $\bar{\alpha}$) towards d, the solution reached will always be strictly better than the starting one. Referring again to Figure 5.3, if the objective function is $f\left(x_1^2 + x_2^2\right)$ that is the distance (squared) from the origin, direction d is improving in point $(2, 0)$ and not improving in point $(0, 2)$.

5.2 Necessary conditions for local optimality

This section introduces a necessary condition for local optimality, derives from it an algorithmic intuition on how to find candidate points and reformulates it in order to make the algorithmic intuition implementable in practice.

Theorem 12 *Let $\tilde{x}$ be a local optimum point and d a feasible direction in $\tilde{x}$. Then, d is not an improving direction in $\tilde{x}$.*

Proof. If $\tilde{x}$ is a local optimum point, then $f(x) \geq f(\tilde{x})$ for all $x \in \mathcal{U}_{\tilde{x},\epsilon} \cap X$. By contradiction, let us assume that there exists a direction d both feasible and improving. Then, there exist two coefficients $\bar{\alpha}_1$ e $\bar{\alpha}_2 > 0$ such that:

- $\tilde{x} + \alpha d \in X$ for all $\alpha \in [0; \bar{\alpha}_1)$
- $f(\tilde{x} + \alpha d) < f(\tilde{x})$ for all $\alpha \in (0; \bar{\alpha}_2)$

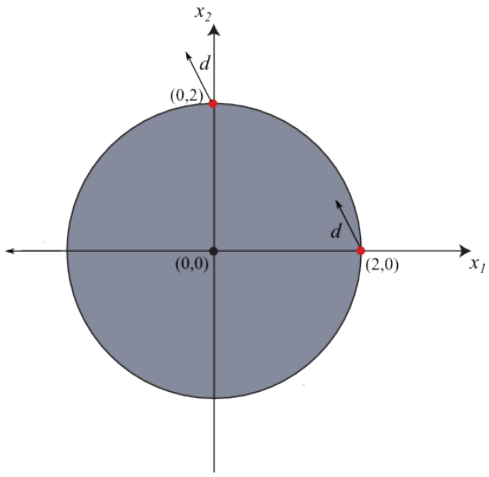


Figure 5.3: Given the objective function $f(x) = \left(x_1^2 + x_2^2\right)$ and the feasible region $X = \{x \in \mathbb{R}^2 : x_1^2 + x_2^2 \geq 4\}$, direction $d = [-1\ 2]^T$ is feasible and nonimproving in point $(0, 2)$, unfeasible and improving in point $(2, 0)$

Now, let us set $\alpha = \frac{1}{2} \min(\epsilon / \|d\|, \bar{\alpha}_1, \bar{\alpha}_2) > 0$ and $x = \tilde{x} + \alpha d$. The three terms in the minimum operator imply tha:

- $\|x - \tilde{x}\| = \|\alpha d\| = \alpha \|d\| \leq \frac{1}{2}\epsilon < \epsilon \Rightarrow x \in \mathcal{U}_{\tilde{x},\epsilon}$
- $x \in X$
- $f(x) < f(\tilde{x})$

but, since $x \in X \cap \mathcal{U}_{\tilde{x},\epsilon}$, necessarily $f(x) \geq f(\tilde{x})$, which is a contradiction. ■

5.2.1 An algorithmic approach

Theorem 12 suggests an algorithmic approach (see also Figure 5.4):

1. consider all feasible points as candidates: $C := X$;
2. scan the set of candidate points C;
3. for each candidate point $\tilde{x}$, scan all feasible directions;
4. for each direction d feasible in $\tilde{x}$, check whether the direction is improving for $f(\cdot)$: if it is, remove $\tilde{x}$ from the candidate set;
5. in the end, scan the remaining candidate points, keeping only the best ones.

This is a fake algorithm, because the set of feasible points and feasible directions are in general infinite; moreover, in order to check whether a direction is improving, one must analyse function $f(\tilde{x} + \alpha d)$ on a whole interval of values of α. This method is even worse than the exhaustive method, that would simply scan all feasible points keeping the best ones (another infinite process). In order to make this process practical, the various conditions that compose it must be replaced with conditions that can be computed in finite time. These conditions should be equivalent, but, since this is not always possible, we will need a number of additional remarks. The computable conditions will be obtained exploiting the linear approximation of the

```
Algorithm $\mathit{FindCandidates}(f, X)$
$X^{\text{KKT}} := X$;
For each $\tilde{x} \in X^{\text{KKT}}$ do
    { Reject the points that violate the necessary conditions }
    For each direction $d$ *feasible in* $\tilde{x}$ do
        If $d$ is *improving respect to* $f$ then $X^{\text{KKT}} := X^{\text{KKT}} \setminus \{\tilde{x}\}$
{ Return the remaining candidate points }
Return $X^{\text{KKT}}$;
```

Figure 5.4: Pseudocode of the process to reduce the feasible set X to the candidate set X^{KKT}.

objective function f and of the constraint functions g_j provided by Taylor's series expansion.

Before doing that, however, we must face a further obstacle: the approach described is a total failure in a very important special case, that is when the problem has nonlinear equality constraints.

Example 26 *Let us assume that the feasible region be the circumference of radius* 2 *with its centre in the origin (see Figure 5.5):*

$$X = \left\{x \in \mathbb{R}^2 : x_1^2 + x_2^2 = 4\right\}$$

and let us focus on the feasible point $\tilde{x} = (0, 2)$.

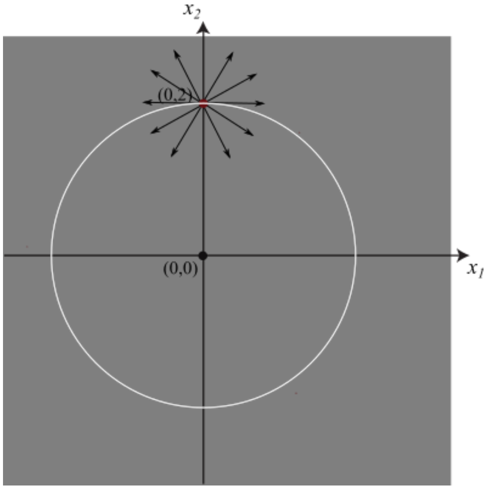


Figure 5.5: Given the feasible region $X = \{x \in \mathbb{R}^2 : x_1^2 + x_2^2 = 4\}$, in all feasible points no direction d is feasible

There is no feasible direction in this point (nor in any other feasible one). Assume by contradiction that d *is a feasible direction:*

$$\tilde{x} + \alpha d = \begin{bmatrix} 0 + \alpha d_1 & 2 + \alpha d_2 \end{bmatrix}^T \in X \textit{ for all } \alpha \in [0; \bar{\alpha})$$

This requires $(0 + \alpha d_1)^2 + (2 + \alpha d_2)^2 = 4 \Rightarrow 4\alpha d_2 + \alpha^2 \left(d_1^2 + d_2^2\right) = 0$, *from which* $\alpha = 0$ *or* $\alpha = -4d_2/\left(d_1^2 + d_2^2\right)$, *that are two fixed values. But the condition should* *hold for any α in a suitable interval. As this is not possible, no direction d is feasible in $\tilde{x}$, and that point cannot be rejected. The same holds for all other point in X: the process, therefore, ends with $X^{\mathrm{KKT}} = X$.*

The consequence on the algorithmic approach is that at step 3 no feasible direction exists, and therefore no candidate point is removed at step 4: in the end, all feasible points must still be explored. Luckily, it is possible to generalise the concept of feasible direction so as to remove most feasible points even in the case of nonlinear equality constraints.

Feasible arcs

Fortunately, it is possible to extend Theorem 12, considering not only linear trajectories from $\tilde{x}$ in a fixed direction, but also curved trajectories (for example, along an arc of circumference).

Definition 16 *We denote as* feasible arc *in a point $\tilde{x} \in X$ a parametric line $\xi : \mathbb{R}_0^+ \to \mathbb{R}^n$, starting from $\tilde{x}$ and belonging to X:*

$$\begin{array}{lll} \xi(0) & = \tilde{x} & \\ \xi(\alpha) & \in X & \textit{for all } \alpha \in [0; \bar{\alpha}) \end{array}$$

Definition 17 *We denote as* tangent direction *of an arc $\xi(\alpha)$ feasible in $\tilde{x}$ the vector composed of the first derivatives of the components ξ_i with respect to parameter α, evaluated in the starting point of the arc:*

$$p_\xi = \left[\left.\frac{d\xi_1}{d\alpha}\right|_0 \dots \left.\frac{d\xi_n}{d\alpha}\right|_0 \right]^T$$

Example 27 *Given $X = \left\{x \in \mathbb{R}^2 : x_1^2 + x_2^2 = 4\right\}$, arc*

$$\xi(\alpha) = \begin{bmatrix} 2\sin\alpha \\ 2\cos\alpha \end{bmatrix}$$

is feasible in point $\tilde{x} = (0, 2)$ because $\xi(0) = \tilde{x}$ and $\xi(\alpha) \in X$ for all $\alpha \in [0; +\infty)$.

The tangent direction of this arc is

$$p_\xi = \begin{bmatrix} 2\cos 0 \\ -2\sin 0 \end{bmatrix} = \begin{bmatrix} 2 \\ 0 \end{bmatrix}$$

Definition 18 *We denote as* improving arc *in a point $\tilde{x} \in X$ for function $f(\cdot)$ a parametric line $\xi : \mathbb{R}_0^+ \to \mathbb{R}^n$ starting from $\tilde{x}$ and for which there exists a value $\bar{\alpha} > 0$ such that*

$$f(\xi(\alpha)) < f(\xi(0)) = f(\tilde{x}) \quad \textit{for all } \alpha \in (0, \bar{\alpha})$$

Parametric arcs are an obvious generalization of directions. In fact, any direction d in $\tilde{x}$ can be associated to an arc $\xi(\alpha) = \tilde{x} + \alpha d$, that is a half-line. The tangent to this arc is the direction itself: $p_\xi = d$. Thus, directions (in particular, feasible, improving, etc...) are a (very small) subset of the arcs, or more precisely of their tangent vectors.

Now, the algorithmic approach can be reformulated by replacing the exploration of all feasible directions in each feasible point, at step 3, with the exploration of all feasible arcs. This makes the task even more impossible to perform directly. In the next sections, we will proceed as follows:

1. replace the nonimprovement condition with a condition on the gradient of the objective function;
2. replace the feasibility condition with a condition on the gradients of the constraint functions (this will yield a first geometric interpretation of the modified conditions);
3. replace these conditions in the KKT conditions, applying a suitable lemma so as to relate the gradient of the objective and those of the constraints (also these conditions will have a geometric interpretation).

5.2.2 Nonimprovement condition

The first step concerns the objective function: as it is difficult to verify that a given direction d is nonimproving in $\tilde{x}$, we will write a condition much easier to verify. This condition is not equivalent, but only necessary, as it is a consequence of the former. Therefore, the points which satisfy it are not local optimum point, but simply candidate points.

Theorem 13 *Let $f \in C^1(X)$, $\tilde{x} \in X$ be a feasible point and $\xi : \mathbb{R} \to \mathbb{R}^n$ an arc feasible and nonimproving in $\tilde{x}$. Then:*

$$(\nabla f(\tilde{x}))^T p_\xi \geq 0$$

Proof. If $\xi(\alpha)$ is a feasible arc, there exists a coefficient $\bar{\alpha}$ such that $\xi(\alpha) \in X$ for all $\alpha \in [0; \bar{\alpha})$ and

$$\begin{aligned} f(\xi(\alpha)) \geq f(\tilde{x}) \Rightarrow \\ \Rightarrow f(\xi(0)) + \left.\frac{df}{d\alpha}\right|_{\alpha=0} \alpha + R_1(\|\xi(\alpha) - \xi(0)\|) \geq f(\tilde{x}) \Rightarrow \\ \Rightarrow (\nabla f(\tilde{x}))^T p_\xi + \frac{R_1(\|\xi(\alpha) - \xi(0)\|)}{\alpha} \geq 0 \end{aligned}$$

where the derivation exploits the properties of derivatives for composite functions of multiple variables (see Appendix A).

If α converges to 0, by continuity the inequality is preserved:

$$\lim_{\alpha \to 0} \left((\nabla f(\tilde{x}))^T p_\xi + \frac{R_1(\xi(\alpha) - \xi(0))}{\|\xi(\alpha) - \xi(0)\|} \frac{\|\xi(\alpha) - \xi(0)\|}{\alpha} \right) \geq 0 \Rightarrow (\nabla f(\tilde{x}))^T p_\xi \geq 0$$

■

Condition $(\nabla f(\tilde{x}))^T d \geq 0$ has a simple geometric interpretation: direction d forms an angle $\leq 90°$ with the gradient of the objective function, that is the direction of quickest worsening (i. e., increase). Therefore, d is nonimproving.

Notice that, in order to verify this condition, it is enough to compute $\nabla f(\tilde{x})$, which is a constant vector, and compute its scalar product with d, that is given. On the contrary, applying the definition to verify that d is nonimproving is impossible in practice, because it requires to verify an inequality for infinite values of α. On the other hand, the condition on the gradient holds also for directions that are actually improving. This happens when $(\nabla f(\tilde{x}))^T d = 0$ and the improvement is due to the rest R_1. This means that in some limit cases, some points will turn out to be candidate even though they are not local optimum points.

Example 28 *Let us consider once again Example 26 about problems with nonlinear equality constraints. Let $X = \left\{x \in \mathbb{R}^2 : x_1^2 + x_2^2 = 4\right\}$ be the feasible region and $f(x) = -x_1 + x_2$ the objective function, with a uniform gradient equal* *to* $(\nabla f(\tilde{x})) = [-1\ 1]^T$. *If one considers point* $\tilde{x} = (0,2)$ *and the feasible arc* $\xi(\alpha) = [2\sin\alpha\ 2\cos\alpha]^T$, *with tangent vector* $p_\xi = [2\ 0]^T$, *it can be verified that:*

$$(\nabla f(\tilde{x}))^T p_\xi = [-1\ 1]^T \begin{bmatrix} 2 \\ 0 \end{bmatrix} = -2 < 0$$

and therefore the point can be rejected. On the contrary, point $x^\circ = \left(\sqrt{2}, -\sqrt{2}\right)$ *admits two feasible arcs:* $\xi(\alpha) = \left[2\cos\left(\alpha - \frac{\pi}{4}\right)\ 2\sin\left(\alpha - \frac{\pi}{4}\right)\right]^T$, *with tangent vector* $p_\xi = \left[\sqrt{2}\ \sqrt{2}\right]^T$, *and* $\bar{\xi}(\alpha) = \left[2\cos\left(\alpha - \frac{\pi}{4}\right)\ 2\sin\left(\alpha - \frac{\pi}{4}\right)\right]^T$, *with tangent vector* $p_{\bar{\xi}} = \left[-\sqrt{2}\ -\sqrt{2}\right]^T$, *and it can be verified that:*

$$(\nabla f(\tilde{x}))^T p_\xi = [-1\ 1]^T \begin{bmatrix} \sqrt{2} \\ \sqrt{2} \end{bmatrix} = 0 \qquad (\nabla f(\tilde{x}))^T p_{\xi'} = [-1\ 1]^T \begin{bmatrix} -\sqrt{2} \\ -\sqrt{2} \end{bmatrix} = 0$$

This is a candidate point in X, *and it is actually the global optimum point.*

Example 29 *Let us consider the following problem:*

$$\begin{aligned} \min f(x) &= (x_1 - 1)^2 + x_2^2 \\ g_1(x) &= -x_1^2 - x_2^2 + 4 \le 0 \\ g_2(x) &= x_1 - 3/2 \le 0 \end{aligned}$$

whose feasible region is reported in Figure 5.6.

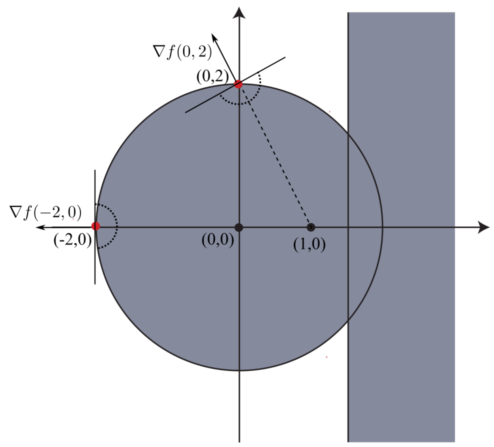


Figure 5.6: Feasible region of the problem reported in Example 29

The gradient of the objective function is $\nabla f(x) = [2(x_1 - 1)\ 2x_2]^T$ *and in every point it is directed to move away from point* $(1,0)$. *For example*[2], *in point* $\tilde{x} = (0,2)$ *the gradient is* $\nabla f(0,2) = [-2\ 4]^T$, *while in point* $\tilde{x} = (-2,0)$ *the gradient is* $\nabla f(-2,0) = [-6\ 0]^T$. *A necessary condition for a direction to be improving is that it be confined in the half-space identified by the dotted semicircumference. Moving from* $\tilde{x} = (0,2)$ *rightwards in direction* $p = [1\ 0]^T$, *the objective function improves. In fact:*

$$(\nabla f(0,2))^T p = [-2\ 4] \begin{bmatrix} 1 \\ 0 \end{bmatrix} = -2 < 0$$

[2] For the sake of simplicity, the figure represents all gradients with the same length, because only the direction in which they point is relevant for the identification of candidate points.

On the contrary, moving upwards in direction $p = [0\ 1]^T$*, the objective function does not improve. In fact:*

$$(\nabla f(0,2))^T p = [-2\ 4] \begin{bmatrix} 0 \\ 1 \end{bmatrix} = 4 \geq 0$$

A limit case can be observed in point $\tilde{x} = (-2, 0)$*: moving downwards along the tangent vector* $p = [0\ -1]^T$*, the condition becomes:*

$$(\nabla f(-2,0))^T p = [-6\ 0] \begin{bmatrix} 0 \\ -1 \end{bmatrix} = 0$$

In fact, this direction is worsening, but there are objective functions with the same gradient for which it would become improving (for instance, $f(x) = -6x_1 + x_2^2$*).*

And there are curved trajectories that have the same tangent vector and are improving for the same objective function: just consider the arc that runs downwards along the circumference of radius equal to 2 *marked in grey. Along that arc, the distance from point* $(1, 0)$ *strictly decreases, and therefore the arc is improving, but the tangent vector is the same, and therefore the nonimprovement condition is satisfied to equality.*

The conclusion is that the conditions are only necessary, and that limit directions cannot be used to reject candidate points, since in such points the conditions can hold both for local optimum points and for points that are not locally optimal.

The resulting "algorithm" (which is still fake because it still contains two infinite nested loops) is reported in Figure 5.7.

```
Algorithm $FindCandidates(f, X)$
$X^{\text{KKT}} := X$;
For each $\tilde{x} \in X^{\text{KKT}}$ do
    { Reject the points that violate the necessary conditions }
    For each $d \in \mathbb{R}^n$ feasible in $\tilde{x}$ do
        If $(\nabla f(\tilde{x}))^T d < 0$ then $X^{\text{KKT}} := X^{\text{KKT}} \setminus \{\tilde{x}\}$
{ Return the best candidate point }
Return $X^{\text{KKT}}$;
```

Figure 5.7: Pseudocode of the process to reduce the feasible set X to the candidate set X^{KKT}.

5.2.3 Feasibility condition

It is possible to replace with an algebraic condition also the condition that d be a feasible direction, but the process is more complicated, because the feasibility condition is a hypothesis of Theorem 12, whereas the nonimprovement condition was the thesis. Therefore, one cannot just replace it with a weaker condition, as we have done with the other one, but only with equivalent or stronger conditions.

First, we need to distinguish the constraints satisfied with an equality in a given point from those strictly satisfied. The reason is that, since the conditions are based on local approximations, it is as though the latter constraints did not exist; in fact, they do not forbid any sufficiently short trajectory.

Definition 19 *We denote as* active constraint *in a point $\tilde{x} \in X$ any constraint $g_j(x) \leq 0$ such that $g_j(\tilde{x}) = 0$. We will indicate by $J_{\mathrm{a}}(x) = \{j \in \{1, \ldots, m\} : g_j(x) = 0\}$ the set of indices of the active constraints.*

Example 30 *Let us consider the example of Figure 5.6. In point $(-2, 2)$ no constraint is active ($J_a(-2,2) = \emptyset$); in fact, $g_1(-2,2) = -4$ and $g_2(-2,2) = -7/2$. In point $(0,2)$ only the first constraint is active ($J_a(0,2) = \{1\}$), since $g_1(0,2) = 0$ and $g_2(0,2) = -3/2$. In point $(3/2, 2)$ only the second constraint is active ($J_a(3/2,2) = \{2\}$), since $g_1(3/2,2) = -9/4$ and $g_2(3/2,2) = 0$. In point $\left(3/2, \sqrt{7}/2\right)$ both constraints are active ($J_a\left(3/2, \sqrt{7}/2\right) = \{1,2\}$), as $g_1\left(3/2, \sqrt{7}/2\right) = 0$ e $g_2\left(3/2, \sqrt{7}/2\right) = 0$.*

Theorem 14 *Let $\xi(\alpha)$ be an arc feasible in $\tilde{x} \in X$, p_ξ its tangent vector. Then*

$$(\nabla g_j(\tilde{x}))^T p_\xi \leq 0 \qquad \textit{for all } j \in J_a(\tilde{x})$$

Proof. If $\xi(\alpha)$ is an arc feasible in $\tilde{x}$, there exists a coefficient $\bar{\alpha} > 0$ such that

$$g_j(\xi(\alpha)) \leq 0 \quad \text{for all } \alpha \in [0; \bar{\alpha}) \text{ and } j = 1, \ldots, m$$

that is

$$\begin{aligned} g_j(\xi(\alpha)) &= g_j(\xi(0)) + \left.\frac{dg_j}{d\alpha}\right|_0 \alpha + R_1(\|\xi(\alpha) - \xi(0)\|) = \\ &= g_j(\tilde{x}) + \alpha (\nabla g_j(\tilde{x}))^T p_\xi + R_1(\|\xi(\alpha) - \xi(0)\|) \leq 0 \end{aligned}$$

The inequality is obvious for all constraints which are not active in $\tilde{x}$, because $g_j(\tilde{x}) < 0$ and the other two terms converge to zero as $\alpha \to 0$. Therefore, by continuity, when α is sufficiently small, they cannot reverse the inequality. For the constraints which are active in $\tilde{x}$, on the contrary, the inequality introduces a strict condition. In fact, for such constraints $g_j(\tilde{x}) = 0$, so that

$$g_j(\xi(\alpha)) = \alpha (\nabla g_j(\tilde{x}))^T p_\xi + R_1(\|\xi(\alpha) - \xi(0)\|) \leq 0$$

Dividing both terms by α and considering the limit, by continuity the thesis follows:

$$\lim_{\alpha \to 0} \left[(\nabla g_j(\tilde{x}))^T p_\xi + \frac{R_1(\|\xi(\alpha) - \xi(0)\|)}{\|\xi(\alpha) - \xi(0)\|} \frac{\|\xi(\alpha) - \xi(0)\|}{\alpha} \right] = (\nabla g_j(\tilde{x}))^T p_\xi \leq 0$$

■

This property is interesting, but cannot replace the requirement that p_ξ be the tangent vector of a feasible arc in the hypothesis of Theorem 13, because it is not an equivalent or sufficient condition for such a requirement, but only a necessary condition. Luckily, it can be proved that, in suitable situations, the condition is actually equivalent.

Regular points

Definition 20 *We denote as* regular point *for a given system of constraints $g_j(x) \leq 0$ a point in which all active constraints have linearly independent gradients (*constraint qualification *condition).*

In a regular point, the conditions on the gradients of the active constraints are not only necessary, but also sufficient to guarantee that there exists a feasible arc with the given tangent vector.

Theorem 15 *Let $\tilde{x}$ be a regular point. There exists an arc $\xi(\alpha)$ feasible in $\tilde{x}$ with tangent direction $p_\xi = p$ if and only if*

$$\left(\nabla g_j(\tilde{x})\right)^T p \leq 0 \quad \textit{for all } j \in J_a(\tilde{x})$$

This implies that in a regular point we can rewrite Theorem 13 replacing the assumption of having a feasible arc with the conditions on the gradients of the active constraints. In other words, we can use the analytical conditions on the gradients to identify the directions that are useful for sifting the candidate points. The nonregular points must be dealt with specifically. Luckily, they are in general a small minority of the feasible points, and we will therefore include them directly in the candidate set withouth further analysis.

Corollary 1 *Let $f \in C^1(X)$, $\tilde{x} \in X$ be a regular and local optimum point and p a vector such that*

$$\left(\nabla g_j(\tilde{x})\right)^T p \leq 0 \quad \textit{for all } j \in J_a(\tilde{x})$$

Then:

$$\left(\nabla f(\tilde{x})\right)^T p \geq 0$$

Example 31 *Consider the following problem:*

$$\begin{aligned}\min f(x) &= x_2\\ g_1(x) &= (x_1-1)^3 + (x_2-2) \leq 0\\ g_2(x) &= (x_1-1)^3 - (x_2-2) \leq 0\\ g_3(x) &= -x_1 \leq 0\end{aligned}$$

whose feasible region is represented in Figure 5.8.

In point $A = (1,2)$, the first two constraints, $g_1(x) = (x_1-1)^3 + (x_2-2) \leq 0$ and $g_2(x) = (x_1-1)^3 - (x_2-2) \leq 0$ are active. Their gradients are $\nabla g_1(A) = [0\,1]^T$ and $\nabla g_2(A) = [0\ -1]^T$. Direction $p = [1\,0]^T$ is orthogonal to both, and thus satisfies the feasibility conditions. In spite of that, no feasible arc admits p as a tangent vector. This is possile because point A is non regular. In fact, the two gradients $\nabla g_1(A)$ and $\nabla g_2(A)$ are opposite, and therefore linearly dependent. In intuitive terms, describing the constraints with their gradients corresponds to linearizing them, that is to treating them as straight lines, which makes it look like it were possible to move rightwards remaining inside the feasible region.

The case of equality constraints

A further complication is introduced by equality constraints, but it can be easily solved by making an exception. In general, an equality constraint can always be replaced by two inequality constraints expressed by opposite functions:

$$h(x) = 0 \Leftrightarrow \begin{cases} h(x) & \leq 0 \\ -h(x) & \leq 0 \end{cases}$$

The complication arises from the fact that, of course, the gradients of these two constraints are exactly opposite ($\pm\nabla h(\tilde{x})$), and therefore linearly dependent. One could conclude that in the problems with equality constraints all points are nonregular, and therefore candidate, which would make the algorithmic approach useless. Fortunately, this is too strict: each single equality constraint, being $h(x) \in C^1(X)$, allows the existence of feasible arcs in every feasible point. Such arcs run

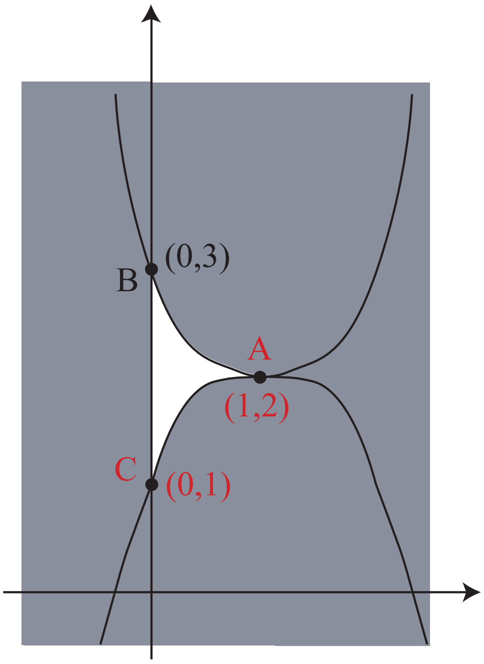


Figure 5.8: A feasible region with a nonregular point: the gradients of the two constraints active in A point in opposite directions

"along" the constraint, following exactly its profile. Only other constraints can make such arcs unfeasible. Therefore, the whole derivation above remains valid with two warnings:

1. *the definition of regular point considers just one gradient for each equality constraint*;
2. the condition on tangent vector p_ξ for each equality constraints, merges the conditions on the two associated inequalities into a simpler single one:

$$\begin{cases} \left(\nabla h_i\left(\tilde{x}\right)\right)^T p_\xi & \leq 0 \\ -\left(\nabla h_i\left(\tilde{x}\right)\right)^T p_\xi & \leq 0 \end{cases} \Leftrightarrow \left(\nabla h_i\left(\tilde{x}\right)\right)^T p_\xi = 0$$

   that is, *the tangent vector p_ξ for a feasible arc is orthogonal to the gradients of equality constraints.*

The algorithmic approach in algebraic form

Now we can summarise the conclusions reached so far, before making the last step. The process we are defining is reported in Figure 5.9 (compare it with the one described in Figure 5.7). It consists of the following steps:

1. find the set X_{nr} of nonregular points and save them aside;
2. consider all regular points as candidates: $C := X \setminus X_{nr}$;
3. scan the set of candidate points C;
4. for each candidate point $\tilde{x} \in C$

5. for each vector p that satisfies the feasibility conditions on the gradients of the active constraints, if the condition on the gradient of the objective function is violated, remove $\tilde{x}$ from the candidate set;
6. add the nonregular points to the candidate set;
7. scan the candidate points to find the best ones.

```
Algorithm $FindCandidates(f, g_1, \dots, g_m)$
$X_{nr} := NonRegular(g)$;
$C := X \setminus X_{nr}$;
For each $\tilde{x} \in C$ do
    { Remove the points that violate the conditions }
    For each $p \in \mathbb{R}^n : (\nabla g_j(\tilde{x}))^T p \leq 0$ for all $j \in J_a(\tilde{x})$ do
        If $(\nabla f(\tilde{x}))^T p < 0$ then $C := C \setminus \{\tilde{x}\}$;
{ Return the candidate set }
$X^{\mathrm{KKT}} := C \cup X_{nr}$;
Return $X^{\mathrm{KKT}}$;
```

Figure 5.9: Pseudocode of the process to reduce the feasible region X to the candidate set X_{KKT}.

The only remaining problem is given by the two infinite loops on x and p. We will solve it replacing the corresponding conditions with a system of equalities and inequalities, whose solutions are the candidate points.

5.2.4 A first geometric interpretation

The property of Corollary 1 has a geometric interpretation that makes it quite intuitive. For each point x taken into account, two sets of vectors are considered:

1. the *feasible cone* $C_{\text{feas}}(x)$ is the set of vectors that form angles ≥ 90° with the gradients of all active constraints:

$$(\nabla g_j(x))^T p \leq 0 \quad \text{for all} j \in J_a(x)$$

   It is a cone because each active constraint identifies a half-space of directions pointing “on the opposite side” of its gradient, and intersecting them yields a cone. In regular points, these are the directions tangent to feasible arcs.

2. the *improving half-space* $C_{\text{impr}}(x)$: is the set of vectors that form angles > 90° with the gradient of the objective function:

$$(\nabla f(x))^T p < 0$$

Corollary 1 states that *the feasible cone must not intersect the improving half-space.* Notice that the former is close, whereas the latter is open, so that the two sets can touch, but not share internal points. The reason is that the directions in which the two cones touch have zero scalar product with the gradient of the objective and with the gradient of some active constraints; in that case, the first-order information is insufficient to determine whether the direction is actually improving or not and actually feasible or not. Therefore, such points cannot be rejected and must remain candidate.

Example 32 *Consider the following mathematical programming model:*

$$\begin{aligned}\min f(x) &= (x_1-1)^2 + x_2^2 \\ g_1(x) &= -x_1^2 - x_2^2 + 4 \leq 0 \\ g_2(x) &= x_1 - 3/2 \leq 0\end{aligned}$$

whose feasible region is represented in Figure 5.10. We now discuss the geometric interpretation of the necessary conditions for local optimality in four significant points. First, we compute the gradients of the objective function and of the two constraint functions:

$$\nabla f(x) = \begin{bmatrix} 2(x_1-1) \\ 2x_2 \end{bmatrix} \qquad \nabla g_1(x) = \begin{bmatrix} -2x_1 \\ -2x_2 \end{bmatrix} \qquad \nabla g_2(x) = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

Then, let us focus on point $(0,2)$. Only the first constraint is active and the feasible cone C_{feas} consists of the directions that form angles $\geq 90°$ with $\nabla g_1(0,2) = [0\ -4]^T$, that form the half-plane pointing upwards, marked with a green half-circumference in Figure 5.12. The improving half-space C_{impr} consists of the directions that form angles $> 90°$ with $\nabla f(0,2) = [-2\ 4]^T$, that form the half-plane pointing downwards and on the right, marked with a blue half-circumference. They clearly intersect (check the scalar products with vector $p = [3\ 1]^T$), and therefore point $(0,2)$ is not candidate.

Now, consider point $\left(3/2, -\sqrt{7}/2\right)$. Both constraints are active and the feasible cone C_{feas} consists of the directions that form angles $\geq 90°$ with $\nabla g_1\left(3/2, -\sqrt{7}/2\right) = \left[-3\ \sqrt{7}\right]^T$ and with $\nabla g_2\left(3/2, -\sqrt{7}/2\right) = [1\ 0]^T$ that form the cone pointing downwards and on the left, marked with a green arc in Figure 5.12. The improving half-space C_{impr} consists of the directions that form angles $> 90°$ with $\nabla f\left(3/2, -\sqrt{7}/2\right) = \left[1\ -\sqrt{7}\right]^T$, that form the half-plane pointing upwards and on the left, marked with a blue half-circumference. Since they do not intersect, point $\left(3/2, -\sqrt{7}/2\right)$ is candidate (it is actually one of the two global optimum points).

Then, consider point $(-2,-2)$. No constraint is active and the feasible cone C_{feas} consists of the whole plane (marked with a green circumference), because every vector satisfies any arbitrary condition with respect to an empty set. The improving half-space C_{impr} consists of the directions that form angles $> 90°$ with $\nabla f(-2,-2) = [-6\ -4]^T$, that form the half-plane pointing upwards and on the right, marked with a blue half-circumference. The two sets clearly intersect, and point $(-2,-2)$ cannot be a candidate.

Finally, consider point $(-2,0)$. Only the first constraint is active and the feasible cone C_{feas} consists of the directions that form angles $\geq 90°$ with $\nabla g_1(-2,0) = [4\ 0]^T$, that form the half-plane pointing leftwards, marked with a green half-circumference in Figure 5.12. The improving half-space C_{impr} consists of the directions that form angles $> 90°$ with $\nabla f(-2,0) = [-6\ 0]^T$, that form the half-plane pointing rightwards, marked with a blue half-circumference. They touch, but do not intersect (remember that the improving half-space is open), and therefore point $(-2,0)$ is a candidate. Notice that this point is not globally optimal, but not even locally optimal, because an arc moving along the circumference would be feasible and also strictly improving. However, the first-order conditions approximate the circumference as a line and therefore cannot reveal this possibility. Indeed, the candidate set X_{KKT} can be strictly larger than the set of local optimum points X^*.

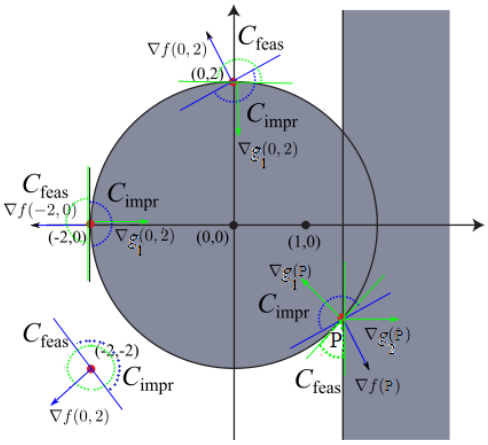


Figure 5.10: Feasible cones and improving cones in four sample points: two non-candidate points, a global optimum point and a candidate point that is not even locally optimal

5.3 Karush-Kuhn-Tucker conditions

This section turns the conditions of Corollary 1 into equivalent conditions that are computationally tractable, that is into the solution of a system of equalities and inequalities. In order to do that, it exploits a theoretical result that is apparently very far from the problems here considered, as it concerns the properties of families of vectors in multidimensional Euclidean spaces. It is just a matter of correctly mapping such objects onto the ones we are manipulating. In order to express the lemma briefly, the concept of conical combination of vectors is required. For this concept (and other ones in linear algebra), we address the reader to Appendix B.

5.3.1 Farkas' Lemma

Theorem 16 *(Farkas' Lemma) Let $f \in \mathbb{R}^n$ and $g_j \in \mathbb{R}^n$ with $j = 1, \ldots, m$ be a vector and a family of vectors in the Euclidean n-dimensional space. Then:*

$$\exists \mu_j \geq 0 : f = \sum_{j=1}^{m} \mu_j g_j \Leftrightarrow p^T f \leq 0 \text{ for all } p : p^T g_j \leq 0 \text{ per } j = 1, \ldots, m$$

In words, vector f is a conical combination of vectors g_j if and only if all vectors pointing on the opposite side of all g_j also point on the opposite side of f.

Proof. It is trivial to prove the direct implication of this theorem (the "only if" part): replacing f with $\sum_{j=1}^m \mu_j g_j$, one can observe that $p^T f = p^T \sum_{j=1}^m \mu_j g_j = \sum_{j=1}^m \mu_j \left(p^T g_j\right)$. Any vector p such that $p^T g_j \leq 0$ guarantees that all terms in the sum be nonpositive, and therefore also the sum be nonpositive. The thesis follows.

The inverse implication, on the contrary, that we will exploit, is much more involved, and we will not provide the proof. ■

Also this lemma has a geometric interpretation, which can make it more intuitive. Figure 5.11 provides an example with $m = 3$ vectors g_j in a $n = 2$-dimensional space:

- for each vector g_j, the set of vectors p that form scalar products $g_j^T p \leq 0$ is the half-plane opposite to g_j;

- the intersection of such half-planes is the cone marked in grey in the lower left;
- the set of vectors p that form scalar products $f^T p \leq 0$ is the half-plane opposite to f, that is in the lower left with respect to the red oblique line;
- the set of vectors that are conical combinations of vectors g_j is the cone marked in grey in the upper right.

Farkas's lemma states that:

- if f falls within the cone of vectors g, then the cone that is opposite to vectors g_j falls in the half-plane opposite to f;
- if the cone that is opposite to vectors g_j falls in the half-plane opposite to f, then f falls within the cone of vectors g.

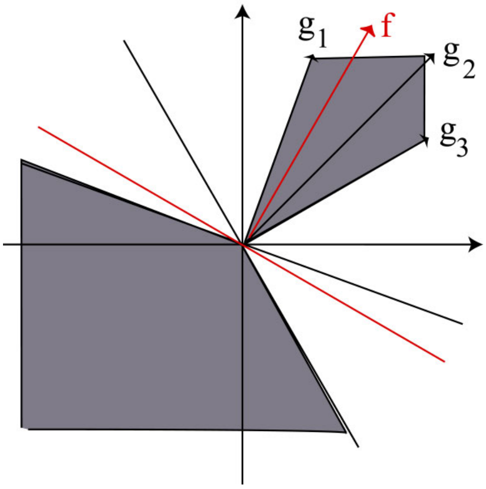


Figure 5.11: Example of application of Farkas' lemma: f falls within the cone of vectors g_j if and only if the cone opposite to vectors g_j falls within the half-plane opposite to f.

5.3.2 Standard form of Karush-Kuhn-Tucker conditions

Now, let us apply Farkas' lemma to the vectors we are using to sift the candidate points. Specifically, let us map:

- the vectors g_j onto the gradients of the active constraints $\nabla g_j(\tilde{x})$;
- vector f onto the *antigradient* of the objective function $-\nabla f(\tilde{x})$.

Notice the negative sign applied to vector ∇f: the antigradient is the opposite of the gradient. The statement of Farkas' lemma becomes:

$$\exists \mu_j \geq 0 : \nabla f(\tilde{x}) + \sum_{j \in J_a(\tilde{x})} \mu_j \nabla g_j(\tilde{x}) = 0$$

if and only if

$$(\nabla f(\tilde{x}))^T p \geq 0 \quad \text{for all } p : (\nabla g_j(\tilde{x}))^T p \leq 0 \text{ for all } j \in J_a(\tilde{x})$$

The second member of the implication means that the feasible directions are nonimproving. In the local optimum points, the second member of Farkas' lemma holds; therefore, also the first member holds: if x^* is a regular local optimum point, then the antigradient of the objective function falls in the cone of the gradients of the active constraints:

$$\exists \mu_j \geq 0 : \nabla f\left(x^*\right) + \sum_{j \in J_a(x^*)} \mu_j \nabla g_j\left(x^*\right) = 0$$

In general, the KKT-conditions are reported in an equivalent standard form, that has two small modifications. First, the conical combination of the active constraints can be extended to a conical combination of all constraints, provided that the multipliers of the nonactive constraints are set to zero. This can be done by introducing additional conditions which relate the multipliers and the constraints, the so called *complementarity conditions*:

$$\sum_{j \in J_a(x^*)} \mu_j \nabla g_j\left(x^*\right) = \sum_{j=1}^{s} \mu_j \nabla g_j\left(x^*\right) \text{ with } \mu_j g_j(x) = 0, \quad \text{for all } j$$

Second, usually the equality constraints are not decomposed into pairs of inequality constraints, because two opposite gradients with multipliers $\mu^+ \geq 0$ and $\mu^- \geq 0$ can always be replaced by a single gradient with a multiplier equal to the difference of the original two, that is a real-valued multiplier:

$$h_i(x) = 0 \Leftrightarrow \begin{cases} g_{j_i'}(x) = h_i(x) & \leq 0 \\ g_{j_i''}(x) = -h_i(x) & \leq 0 \end{cases}$$

$$\begin{aligned} \ldots + \mu_{j_i'} \nabla h_i(x) + \mu_{j_i''}\left(-\nabla h_i(x)\right) + \ldots \qquad & \text{with } \mu_{j_i'}, \mu_{j_i''} \geq 0 \\ \ldots + \lambda_i \nabla h_i(x) + \ldots \qquad & \text{with } \lambda_i = \mu_{j_i'} - \mu_{j_i''} \in \mathbb{R} \end{aligned}$$

Theorem 17 (KKT-conditions) *Let* $X = \{x \in \mathbb{R}^n : h_i(x) = 0, g_j(x) \leq 0, \text{ with } i = 1, \ldots, s \text{ e } j = 1, \ldots, m\}$ *and* $f, h_i, g_j \in C^1(X)$ *for* $i = 1, \ldots, s$ *and* $j = 1, \ldots, m$. *If* x^* *is a regular point in* X *and a local optimum point for* f *in* X, *then there exist free multipliers* λ_i *and nonnegative multipliers* $\mu_j \geq 0$ *such that:*

$$\nabla f\left(x^*\right) + \sum_{i=1}^{s} \lambda_i \nabla h_i\left(x^*\right) + \sum_{j=1}^{m} \mu_j \nabla g_j\left(x^*\right) = 0 \tag{5.2a}$$

$$\mu_j g_j\left(x^*\right) = 0 \qquad j = 1, \ldots, m \tag{5.2b}$$

$$h_i\left(x^*\right) = 0 \qquad i = 1, \ldots, s \tag{5.2c}$$

$$g_j\left(x^*\right) \leq 0 \qquad j = 1, \ldots, m \tag{5.2d}$$

$$\mu_j \geq 0 \qquad j = 1, \ldots, m \tag{5.2e}$$

The sifting algorithm, therefore, does not consist in scanning all points $x \in X$ and all directions $p \in \mathbb{R}^n$, verifying whether they satisfy suitable conditions, but in solving a system of equalities and inequalities, identifying the points that satisfy it. It can be remarked that the system binds $n + s + m$ variables (x, λ, μ) with $n + s + m$ equalities (5.2a, 5.2c, 5.2b), plus $2m$ inequalities (5.2d, 5.2e). The system is therefore balanced, and it can be expected to have in general a finite number of solutions (not a single one, because it is nonlinear).

As a concluding remark, the conditions in standard form are also sometimes described introducing an auxiliary function, named *generalized Lagrangean function*

$$\ell(x, \lambda, \mu) = f(x) + \sum_{i=1}^{s} \lambda_i h_i(x) + \sum_{j=1}^{m} \mu_j g_j(x)$$

Given this definition, the Equations (5.2) can be interpreted stating that they set to zero the partial derivatives of $\ell(x, \lambda, \mu)$ with respect to the x variables, while the Equations (5.2c) set to zero the partial derivatives of $\ell(x, \lambda, \mu)$ with respect to the λ variables. The complementarity conditions (5.2d) and the sign conditions on the multipliers (5.2e), however, have no interpretation related to the generalized Lagrangean function.

5.3.3 A second geometric interpretation

From a geometric point of view, the antigradient of the objective function, $-\nabla f(x)$, is the direction in which the objective f decrease most quickly, that is the most improving direction. The gradients of the active constraints, $\nabla g_j(x^*)$, are the directions in which the constraint functions g_j increase most quickly, that is the directions of maximum violation of the constraints. When the antigradient of the objective falls within the cone of the gradients of the active constraints, the objective can improve only violating at least one constraint.

Example 33 *Consider again the problem of Example 32 and Figure 5.10. In point* $(0, 2)$*, only the first constraint is active and the cone of its gradient consists of the half-line going downwards (that is, the green arrow). The antigradient of the objective function points downwards and on the right, as it is opposite to the gradient (blue arrow). It does not belong to the cone, and therefore the point is not candidate.*

In point $\left(3/2, -\sqrt{7}/2\right)$*, both constraints are active, and the cone of the two gradients points upwards, on the right and partly on the left (between the two green arrows). The antigradient of the objective function points upwards and on the left (opposite to the blue arrow). Therefore, it falls inside the cone and the point is candidate.*

In point $(-2, -2)$*, no constraint is active and the cone of the gradients is empty. The antigradient of the objective function points upwards and on the right, as it is opposite to the gradient (blue arrow). Of course, it does not belong to the empty cone, and therefore the point is not candidate.*

Finally, in point $(-2, 0)$*, only the first constraint is active and the cone of its gradient consists of the half-line going rightwards (that is, the green arrow). The antigradient of the objective function also points rightwards (opposite to the blue arrow). Therefore, it belongs to the cone, and the point is candidate.*

5.4 Interesting special cases*

Several special cases are interesting for various reasons.

Unconstrained problems If the problem has no constraints, the KKT-conditions reduce to

$$\nabla f(x^*) = 0$$

that is, setting the gradient of the objective to zero. This conditions simply generalizes the classical condition of setting the first derivative to zero in order to compute the local optimum points for single-variable functions. The candidate points thus determined can be global minimum points, simply local minimum points, inflection points, and even local maximum points.

*This section presents advanced concepts, that are not part of the course syllabus.

Linear problems If the problem is linear ($f(x) = \sum_{i=1}^{n} c_i x_i$ and $g_j(x) = \sum_{i=1}^{n} a_{ij} x_i - b_j \leq 0$, the KKT-conditions become:

$$\begin{cases} c_i + \mu_j a_{ij} = 0 & i = 1, \dots, n \\ \mu_j \left(\sum_{i=1}^{n} a_{ij} x_i - b_j \right) = 0 & j = 1, \dots, m \\ \mu_j \geq 0 & \text{per } j = 1, \dots, m \\ \sum_{i=1}^{n} a_{ij} x_i - b_j \leq 0 & j = 1, \dots, m \end{cases}$$

The first family of constraints is part of the feasibility conditions for the dual problem, provided that the multipliers μ_j be interpreted as dual variables with a reversed sign (they are equality constraints because the primal problem has free variables). The second family of constraints are the complementary slackness conditions. The third family of constraints are the remaining feasibility conditions for the dual problem: since the primal constraints are of the $\leq$ type and the problem is a minimization problem, the dual variables should be nonpositive, but here they are nonnegative because we have reversed their sign. Finally, the fourth family of constraints are the feasibility conditions for the primal problem. A fundamental theorem of Linear Programming guarantees that these conditions determine the optimal solution, both of the primal and the dual problem.

Discrete problems

When the feasible region of the problem consists of isolated point (for example, points with integer coordinates: $X \subseteq \mathbb{Z}^n$), it is anyway possible to describe it through inequalities. For example, an integrality constraint on a decision variable x_i can be formulated as $h_i(x) = \sin(\pi x_i) = 0$, a binarity constraint as $h_i(x) = x_i (1 - x_i) = 0$. As such formulations are regular, they allow to apply the KKT-conditions, which however prove useless in practice. In fact, the equality constraints $h_i(x) = 0$ introduce in every equality of the system an additional term $\lambda_i \frac{\partial h_i}{\partial x_i}$, which appears only in that equality. This means that in any point x it is possible to assign each multiplier λ_i a value that satisfies the corresponding equation: *in discrete problems, all points are candidates.* This is intuitively obvious: in a discrete problem, every point is isolated, and therefore admits a sufficiently small neighbourhood such that it is the only feasible point belonging to it. therefore, in a discrete problem every point is a local optimum point.

5.5 Applications of the KKT-conditions*

5.5.1 Big-Data system planning

Big-Data management systems such as Hadoop/YARN deal with computational processes whose characteristics and required levels of service are very heterogeneous. This makes it more complex to decide how many and which resources to assign to the computation requests, and how to distribute along time the operations, so as to obtain both a high level of resource use and a good satisfaction of the computation requests. In the following, we describe a small management subproblem to which the KKT-conditions can be directly applied[3].

Let J be a set of $n = |J|$ *jobs* to perform. Each *job* $j \in J$ is associated to an overall workload w_j, that can be sudivided among several resources so that the work proceed in parallel. We will assume perfectly *splittable jobs*, that is, divisible completely *ad libitum* at every time, provided that a minimum and maximum length ($t_j^{\min}$ and $t_j^{\max}$) of the computation, as well as a minimum and maximum resource occupation ($r_j^{\min}$ and $r_j^{\max}$) in every phase of the computation are respected. Of course, this is an approximation, which can be justified for heavily parallelizable jobs. The jobs of J must be performed in a preassigned sequence, so that they form a completely ordered chain. The chain of jobs must be performed minimizing the overall variation of the resource use during the time horizon T. The variation of the resource use is defined at each time step as the absolute value of the variation in the resource use with respect to the previous time step. The idea is that, even if we neglect the time required to activate and deactivate the resources, these operations have a cost that we want to minimize.

We can represent the jobs are two-dimensional geometric figures, where the horizontal axis corresponds to time, while the vertical one corresponds to the amount of resource used at each time. It is quite intuitive that the best solution with respect to the variation of resource use correspond to a rectangular figure, where the job j is divided among r_j resources working for t_j time steps, after which the job terminates. This is because, once the workload (that is, the figure's area) is fixed, increasing and decreasing several times the height of the figure, that is the amount of resources used, would only produce a larger variation. On the other hand, slowly increasing or decreasing the resource use would only produce an extension of the time required to perform the whole job.

At this point, we want to find the basis and the height of each rectangle in J so as to minimize the sum of their heights (counting them twice, because the resourced used have equal (but opposite) variations at the beginning and at the end of each job. The rectangles representing the jobs must be sequentially ordered along time (the horizontal axis) and the total width of the chain must not exceed L. The

*This section presents advanced concepts, that are not part of the course syllabus.

[3] This application derives from R. Cordone, G. M. Fumarola, M. Mazzucchelli, M. Rabozzi, M. Santambrogio, *Preemption-aware planning on Big-Data Systems*, PPoPP '16 Proceedings of the 21st ACM SIGPLAN Symposium on Principles and Practice of Parallel Programming, Article No. 48.

following model is implied:

$$
\begin{aligned}
\min_{r,t} f\left(r,t\right) &= \sum_{j\in J} 2r_j \\
r_j t_j &= w_j && j \in J \\
\sum_{j\in J} t_j &\le T \\
r_j^{\min} \le r_j &\le r_j^{\max} && j \in J \\
t_j^{\min} \le t_j &\le t_j^{\max} && j \in J
\end{aligned}
$$

The problem has $2|J|$ variables, $4|J|+1$ inequality constraints and $|J|$ equality constraints.

The direct application of KKT-conditions yields a far-from-trivial system. It is, however, possible to make some simplification:

1. divide by 2 the objective function, which does not affect the optimal solution;
2. use constraints $r_j t_j = w_j$ to compute variables r_j as a function of variables t_j; this removes $|J|$ variables and all $|J|$ equality constraints;
3. remark that constraints $r_j^{\min} \le r_j \le r_j^{\max}$, after the replacement, become $w_j/r_j^{\max} \le t_j \le w_j/r_j^{\min}$; these new constraints can be combined with the old ones on t_j, choosing case-by-case the tighter constraint.

This simplified formulation follows:

$$
\begin{aligned}
\min f\left(t\right) &= \sum_{j\in J} \frac{w_j}{t_j} \\
\sum_{j\in J} t_j &\le T \\
\bar{t}_j^{\min} \le t_j &\le \bar{t}_j^{\max} && j \in J
\end{aligned}
$$

where $\bar{t}_j^{\min} = \max\left(t_j^{\min}, w_j/r_j^{\max}\right)$ and $\bar{t}_j^{\max} = \min\left(t_j^{\max}, w_j/r_j^{\min}\right)$. This formulation has $|J|$ variables and $2|J|+1$ inequality constraints.

Assigning a multiplier $\mu \ge 0$ to constraint $\sum_{j\in J} t_j \le T$ and multipliers $\nu_j^{\min} \ge 0$ and $\nu_j^{\max} \ge 0$ to the constraints which impose bounds on t_j, one obtains the following system of KKT-conditions:

$$
\begin{aligned}
-\frac{w_j}{t_j^2} + \mu - \nu_j^{\min} + \nu_j^{\max} &= 0 && j \in J \\
\mu\left(\sum_{j\in J} t_j - T\right) &= 0 \\
\nu_j^{\min}\left(t_j - \bar{t}_j^{\min}\right) &= 0 && j \in J \\
\nu_j^{\max}\left(t_j - \bar{t}_j^{\max}\right) &= 0 && j \in J \\
\sum_{j\in J} t_j &\le T \\
\bar{t}_j^{\min} \le t_j &\le \bar{t}_j^{\max} && j \in J \\
\mu, \nu_j^{\min}, \nu^{\max} &\ge 0 && j \in J
\end{aligned}
$$

which is still not trivial.

Now we should consider the complementarity conditions and divide the problem into subproblems. Each subproblem corresponds to a different region of the solution space, in which different families of constraints are active or inactive.

We will apply a very practical approach: we will analyse only one subproblem, that appears to be the most likely. Most inequality constraints require the length of a job to fall within a given range $\left[t_j^{\min}; t_j^{\max}\right]$. If the minimum values $t_j^{\min}$ are small enough and the maximum values are large enough, that is, if the jobs are sufficiently malleable, it is likely that all these constraints will be inactive in the optimal solution. In fact, jobs with very small lengths t_j imply a strong use of resources r_j, which strongly affects the objective. On the other hand, jobs with very long lengths t_j exhaust the overall available time T, and therefore strongly affect the remaining constraint $\sum_{j \in J} t_j \leq T$. In summary, if instead of analysing all $2|J|+1$ complementarity conditions, producing (in the worst case) $2^{2|J|+1}$ subproblems, we will analyse just one condition. The risk is to neglect candidate solutions, and possibly the optimal one.

If we restrict the analysis to the subproblems in which $\nu_j^{\min} = \nu_j^{\max} = 0$ for every $j \in J$, the simplified system becomes:

$$\begin{aligned}
-\frac{w_j}{t_j^2} + \mu &= 0 \qquad && j \in J \\
\mu\left(\sum_{j \in J} t_j - T\right) &= 0 && \\
\sum_{j \in J} t_j &\leq T && \\
t_j^{\min} \leq t_j &\leq t_j^{\max} && j \in J \\
\mu &\geq 0 && j \in J
\end{aligned}$$

The first constraint imposes $\mu = 2w_j/t_j^2 > 0$, and therefore $\sum_{j \in J} t_j = T$: the time horizon is fully exploited (this is rather intuitive, since it allows to use less resources, and therefore to limit the variations of resource use). Solving with respect to the t_j variables, we compute their values as a function of μ:

$$\mu = \frac{w_j}{t_j^2} \Rightarrow t_j^2 = \frac{w_j}{\mu} \Rightarrow t_j = \sqrt{\frac{w_j}{\mu}}$$

Now, the constraint on the time horizon provides the value of μ:

$$\sum_{j \in J} t_j = \sum_{j \in J} \sqrt{\frac{w_j}{\mu}} = T \Rightarrow \sqrt{\mu} = \frac{\sum\limits_{j \in J} \sqrt{w_j}}{T}$$

from which

$$t_j = \frac{\sqrt{w_j}}{\sum\limits_{j \in J} \sqrt{w_j}} T \quad \text{for all} j \in J$$

that is a very elegant solution, as it distributes the time horizon T among the jobs proportionally not to the workload (that is an “area”, in a sense), but to its square root (that is, sort of a “length”).

What about the resources?

$$r_j = \frac{w_j t_j = \sqrt{w_j} \sum\limits_{j \in J} \sqrt{w_j}}{T} \quad \text{for all} j \in J$$

They are also proportional to the square root of the workload. In a sense, the workload of each task affects in a balanced way the amount of resources and the amount of time allotted.

The corresponding value of the objective is:

$$f^* = \sum_{j \in J} 2r_j^* = 2\sum_{j \in J} \frac{w_j}{t_j^*} = 2T \sum_{j \in J} \sqrt{w_j} \sum_{j \in J} \sqrt{w_j} = 2T \left(\sum_{j \in J} \sqrt{w_j} \right)^2$$

This solution is feasible if each t_j is actually included between $t_j^{\min}$ e $t_j^{\max}$. If even one of these values exceeds the corresponding interval, the solution must be rejected. Moreover, even if it is feasible, there could still be other solutions (worse, or better ones) in the other subproblems: in order to know, we should analyse them all. To be complete, it is possible to prove, exploiting second-order conditions (that are not treated in these notes) that the problem is convex, and therefore admits a single candidate solution, that is globally optimal. Therefore, if the solution found is feasible, it is also globally optimal. If it is not, there exist efficient algorithms (which we do not describe here) to extend the simplex algorithm from linear programming to the problems with a convex objective function and linear constraints, so as to find the global optimum without exploring all the subproblems.

5.6 Exercises

Exercise 1

A nuclear waste dump must be located as close as possible to the plant that produces the waste (point $\bar{x} = (1, 0)$). On the other hand, the dump must be at least at 2 km from a town in point $\tilde{x} = (0, 0)$ and must be outside a park: the buffer zone is parallel to line $x = 3/2$ and the plant must be on the left of this line.

Solution

Model and graphical representation

Let $x = (x_1, x_2)$ be the position of the dump. The problem admits the following mathematical programming model:

$$\begin{aligned} \min f(x) &= (x_1 - 1)^2 + x_2^2 \\ g_1(x) &= -x_1^2 - x_2^2 + 4 \leq 0 \\ g_2(x) &= x_1 - 3/2 \leq 0 \end{aligned}$$

Figure 5.12 represents the feasible region.

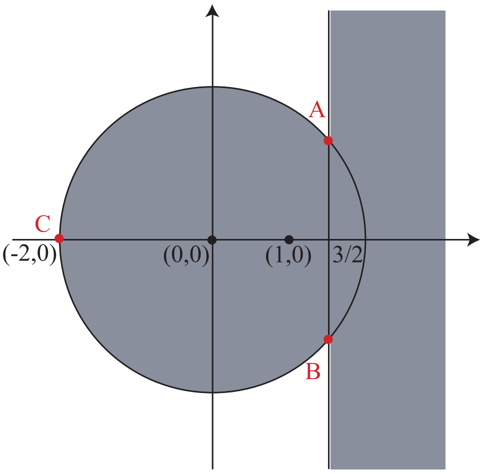


Figure 5.12: Feasible region

Nonregular points

All points in which the gradients of the active constraints are linearly independent are regular. Notice that only the active constraints must be considered. Therefore, the interior of the feasible region is fully composed by regular points. The points through which a single constraint passes are regular when its gradient is nonzero, and so on.

The gradients of the two constraints are:

$$\begin{aligned} \nabla g_1(x) &= [\, -2x_1 \ \ -2x_2 \,]' \\ \nabla g_2(x) &= [\, 1 \ 0 \,]' \end{aligned}$$

The first gradient is zero only in the origin, where the constraint is nonactive ($g_1(0,0) = 4 > 0$). The second gradient is never zero. We are therefore interested only in the points in which both constraints are active (such points are denoted as A and B in the figure):

$$\begin{cases} g_1(x) = -x_1^2 - x_2^2 + 4 = 0 \\ g_2(x) = x_1 - 3/2 = 0 \end{cases} \Rightarrow \begin{cases} x_1 = 3/2 \\ x_2^2 = 4 - x_1^2 = 7/4 \end{cases}$$

from which $A = \left(3/2, \sqrt{7}/2\right)$ e $B = \left(3/2, -\sqrt{7}/2\right)$.

In A, $\nabla g_1(x) = \left[-3 \; -\sqrt{7}\right]'$ and the gradients of the two constraints are linearly independent. This can be verified in two ways: applying the definition or evaluating the rank of the matrix whose columns are the gradients. The first way consists in imposing

$$\alpha_1 \nabla g_1(A) + \alpha_2 \nabla g_2(A) = 0$$

from which

$$\begin{cases} -3\alpha_1 + \alpha_2 = 0 \\ -\sqrt{7}\alpha_1 = 0 \end{cases} \Rightarrow \alpha_1 = \alpha_2 = 0$$

Since the only way to fix to zero a linear combination of the two vectors is to set both coefficients to zero, the two vectors are linearly independent by definition.

On the other hand, matrix

$$\begin{bmatrix} -3 & 1 \\ -\sqrt{7} & 0 \end{bmatrix}$$

has rank equal to 2, because its determinant is nonzero: $-3\;0 + \sqrt{7}\;1 = \sqrt{7} \neq 0$

Karush-Kuhn-Tucker conditions

The KKT-conditions (also known as necessary first-order conditions) state that, if a point is regular and locally minimal, the

1. the partial derivatives of the generalized Lagrangean function with respect to the x variables ($\partial \ell / \partial x_i = 0$) are equal to zero;

2. the partial derivatives of the generalized Lagrangean function with respect to the λ multipliers ($\partial \ell / \partial \lambda_j = h_j = 0$) are equal to zero, that is, the equality constraints are respected;

3. the products of the functions expressing the inequality constraints, times the corresponding multipliers ($\mu_k g_k = 0$) are equal to zero;

4. all inequality constraints are satisfied ($g_k \leq 0$);

5. all multipliers of the inequality constraints are nonnegative ($\mu_k \geq 0$).

Let us use these conditions to reject the points that do not satisfy them, hoping that few *candidate points* remain. The nonregular points shall be added to the candidate set, since the KKT-conditions are not necessary in these points. In the present case, however, all points are regular.

The generalized Lagrangean function is $\ell(x) = f(x) + \mu_1 g_1(x) + \mu_2 g_2(x) = (x_1 - 1)^2 + x_2^2 + \mu_1\left(-x_1^2 - x_2^2 + 4\right) + \mu_2(x_1 - 3/2)$, for which

$$\begin{aligned}
\partial\ell/\partial x_1 &= 2(x_1 - 1) - 2\mu_1 x_1 + \mu_2 = 0 \\
\partial\ell/\partial x_2 &= 2x_2 - 2\mu_1 x_2 = 0 \\
\mu_1 g_1 &= \mu_1\left(-x_1^2 - x_2^2 + 4\right) = 0 \\
\mu_2 g_2 &= \mu_2(x_1 - 3/2) = 0 \\
g_1 &= -x_1^2 - x_2^2 + 4 \leq 0 \\
g_2 &= x_1 - 3/2 \leq 0 \\
& \mu_1 \geq 0 \\
& \mu_2 \geq 0
\end{aligned}$$

In order to solve this system without exhaustively exploring all possible cases, one can set up a search tree, whose nodes divide the feasible region in disjoint parts. In order to do that, we build on the stricter conditions, that are the products $\mu_k g_k = 0$. Given a constraint, we can distinguish three cases:

1. $\mu_k = 0$ and $g_k < 0$
2. $\mu_k = 0$ and $g_k = 0$
3. $\mu_k > 0$ and $g_k = 0$

but in general the computations are much simpler if the second case is merged with the first one, instead of treating it alone or merging it with the third one. We therefore set:

1. $\mu_k = 0$ and $g_k \leq 0$
2. $\mu_k > 0$ and $g_k = 0$

As for the constraint, it is better to choose the simplest one, that is the second, splitting the original problem P^0 into P^1 ($\mu_2 > 0$ and $g_2 = 0$) and P^2 ($\mu_2 = 0$ and $g_2 \leq 0$).

$\mathbf{P^1}$ ($\mu_2 > 0$ and $g_2(x) = 0$)

Since $g_2(x) = 0 \Rightarrow x_1 = 3/2$:

$$\begin{aligned}
1 - 3\mu_1 + \mu_2 &= 0 \\
x_2(1 - \mu_1) &= 0 \\
\mu_1\left(7/4 - x_2^2\right) &= 0 \\
7/4 - x_2^2 &\leq 0 \\
\mu_1 &\geq 0
\end{aligned}$$

As $\mu_1 = (\mu_2 + 1)/3 \geq 1/3 > 0$, it is $x_2^2 = 7/4$. This yields two candidate points: $A = \left(3/2, \sqrt{7}/2\right)$ with $\mu_A = (1, 2)$ and $B = \left(3/2, -\sqrt{7}/2\right)$ with $\mu_B = (1, 2)$.

$\mathbf{P^2}$ ($\mu_2 = 0$ and $g_2 \leq 0$)

$$\begin{aligned} x_1 (1 - \mu_1) &= 1 \\ x_2 (1 - \mu_1) &= 0 \\ \mu_1 \left(-x_1^2 - x_2^2 + 4\right) &= 0 \\ x_1^2 + x_2^2 &\geq 4 \\ x_1 &\leq 3/2 \\ \mu_1 &\geq 0 \end{aligned}$$

The first constraint guarantees that $\mu_1 \neq 1$, so that $x_2 = 0$

$$\begin{aligned} x_1 (1 - \mu_1) &= 1 \\ x_2 &= 0 \\ \mu_1 \left(4 - x_1^2\right) &= 0 \\ x_1^2 &\geq 4 \\ x_1 &\leq 3/2 \\ \mu_1 &\geq 0 \end{aligned}$$

As $x_1^2 \geq 4$ e $x_1 \leq 3/2$, it is $x_1 \leq -2$, and therefore $1 - \mu_1 = 1/x_1 < 0$, which implies $\mu_1 > 1$. Then $4 - x_1^2 = 0 \Rightarrow x_1 = -2$. This yields the candidate point $C = (-2, 0)$ with $\mu_C = (3/2, 0)$.

We therefore have overall three candidate points, among which we must choose the one in which the objective function assumes the minimum value.

$$\begin{cases} A = \left(3/2, \sqrt{7}/2\right) \Rightarrow f(A) = 2 \\ B = \left(3/2, -\sqrt{7}/2\right) \Rightarrow f(B) = 2 \\ C = (-2, 0) \Rightarrow f(C) = 9 \end{cases}$$

which implies that A and B are both globally optimal points.

Exercise 2

Solve the following problem with the KKT-conditions:

$$\begin{aligned}
\min f(x) &= x_1 + x_2 \\
g_1(x) &= -x_1^2 + x_2^2 + 1 \leq 0 \\
g_2(x) &= x_1^2 - 4 \leq 0 \\
g_3(x) &= -x_2 \leq 0
\end{aligned}$$

Solution

Figure 5.13 represents the feasible region.

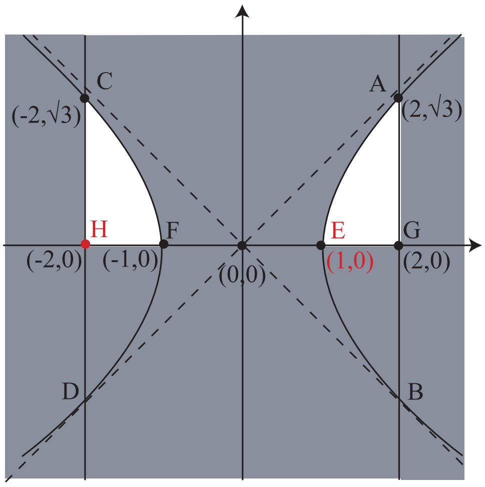


Figure 5.13: Feasible region

Nonregular points

The regular points are all the points in which the gradients of the active constraints are linearly independent. Notice that only the active constraints must be taken into account. Therefore, the internal points of the feasible region are all regular, those in which a single constraint is active are regular when the gradient is nonzero, and so on.

The gradients of the three constraints are:

$$\begin{aligned}
\nabla g_1(x) &= [\, -2x_1 \; 2x_2 \,]' \\
\nabla g_2(x) &= [\, 2x_1 \; 0 \,]' \\
\nabla g_3(x) &= [\, 0 \; -1 \,]'
\end{aligned}$$

The first and second gradient are zero only in the origin, where the corresponding constraints are nonactive. The intersections of the constraints two-by-two are:

- $g_1(x) = g_2(x) = 0$

$$\begin{aligned}
g_1(x) &= -x_1^2 + x_2^2 + 1 \leq 0 \\
g_2(x) &= x_1^2 - 4 \leq 0
\end{aligned}$$

which yields the points $A = \left(2, \sqrt{3}\right)$, $B = \left(2, -\sqrt{3}\right)$, $C = \left(-2, \sqrt{3}\right)$ and $D = \left(-2, -\sqrt{3}\right)$. The gradients ∇g_1 and ∇g_2, evaluated, in these four points, are linearly independent. Let us verify it for A (the process is similar for the other three points):

$$\nabla g_1 = \left[\ -4\ 2\sqrt{3}\ \right]' \quad \nabla g_2 = [\ 4\ 0\ ]'$$

which are linearly independent because the matrix obtained lining them up is nonsingular:

$$\begin{bmatrix} -4 & 4 \\ 2\sqrt{3} & 0 \end{bmatrix} \Rightarrow \begin{vmatrix} -4 & 4 \\ 2\sqrt{3} & 0 \end{vmatrix} = -8\sqrt{3} \neq 0$$

- $g_1(x) = g_3(x) = 0$

$$\begin{aligned} g_1(x) &= -x_1^2 + x_2^2 + 1 \leq 0 \\ g_3(x) &= -x_2 \leq 0 \end{aligned}$$

which yield the points $E = (1, 0)$ and $F = (-1, 0)$. The gradients ∇g_1 e ∇g_3, evaluated in these two points, are also linearly independent.

- $g_2(x) = g_2(x) = 0$

$$\begin{aligned} g_2(x) &= x_1^2 - 4 \leq 0 \\ g_3(x) &= -x_2 \leq 0 \end{aligned}$$

whichi yield the points $G = (2, 0)$ and $H = (-2, 0)$, where the gradients ∇g_2 e ∇g_3 are linearly independent.

Finally, there is no intersection of all three constraints. Hence, all points are regular.

Karush-Kuhn-Tucker conditions

The generalized Lagrangean function is $\ell(x, \mu) = x_1 + x_2 + \mu_1\left(-x_1^2 + x_2^2 + 1\right) + \mu_2\left(x_1^2 - 4\right) + \mu_3(-x_2)$. Therefore, the KKT conditions are:

$$\begin{aligned} \partial\ell/\partial x_1 &= 1 - 2\mu_1 x_1 + 2\mu_2 x_1 = 0 \\ \partial\ell/\partial x_2 &= 1 + 2\mu_1 x_2 - \mu_3 = 0 \\ \mu_1 g_1 &= \mu_1\left(-x_1^2 + x_2^2 + 1\right) = 0 \\ \mu_2 g_2 &= \mu_2\left(x_1^2 - 4\right) = 0 \\ \mu_3 g_3 &= \mu_3 x_2 = 0 \\ g_1 &= -x_1^2 + x_2^2 + 1 \leq 0 \\ g_2 &= x_1^2 - 4 \leq 0 \\ g_3 &= -x_2 \leq 0 \\ & \mu_1 \geq 0 \\ & \mu_2 \geq 0 \\ & \mu_3 \geq 0 \end{aligned}$$

The simplest constraint is $\mu_3 g_3(x) = 0$, so that we split the original problem P^0 into P^1 ($\mu_3 = 0$ and $g_3 \leq 0$) and P^2 ($\mu_3 > 0$ and $g_3 = 0$).

$\mathbf{P^1}$ ($\mu_3 = 0$ and $g_3 \leq 0$)

$$\begin{aligned} 2\left(\mu_1 - \mu_2\right) x_1 &= 1 \\ 1 + 2\mu_1 x_2 &= 0 \\ \mu_1\left(-x_1^2 + x_2^2 + 1\right) &= 0 \\ \mu_2\left(x_1^2 - 4\right) &= 0 \\ -x_1^2 + x_2^2 + 1 &\leq 0 \\ x_1^2 &\leq 4 \\ x_2 &\geq 0 \\ \mu_1 &\geq 0 \\ \mu_2 &\geq 0 \end{aligned}$$

Since $\mu_1 \geq 0$ e $x_2 \geq 0$, it is $1 + 2\mu_1 x_2 \geq 1 > 0$. Then, the second constraints cannot be satisfied and this subproblem does not include candidate points.

$\mathbf{P^2}$ ($\mu_3 > 0$ and $g_3 = 0 \Rightarrow x_2 = 0$)

$$\begin{aligned} 2\left(\mu_1 - \mu_2\right) x_1 &= 1 \\ 1 - \mu_3 &= 0 \Rightarrow \mu_3 = 1 \\ \mu_1\left(-x_1^2 + 1\right) &= 0 \\ \mu_2\left(x_1^2 - 4\right) &= 0 \\ -x_1^2 + 1 &\leq 0 \\ x_1^2 - 4 &\leq 0 \\ \mu_1 &\geq 0 \\ \mu_2 &\geq 0 \end{aligned}$$

Let us further decompose the subproblem P^2, using constraint $\mu_1 g_1(x) = 0$: we obtain P^{11} ($\mu_1 = 0$ and $g_1 \leq 0$) and P^{12} ($\mu_1 > 0$ and $g_1 = 0$).

$\mathbf{P^{11}}$ ($\mu_1 = 0$ and $g_1 \leq 0$)

$$\begin{aligned} -2\mu_2 x_1 &= 1 \\ \mu_3 &= 1 \\ \mu_2\left(x_1^2 - 4\right) &= 0 \\ 1 \leq x_1^2 &\leq 4 \\ \mu_2 &\geq 0 \end{aligned}$$

Since $-2\mu_2 x_1 = 1$, $\mu_2 \neq 0$, which implies $x_1^2 = 4$. On the other hand, $-2\mu_2 x_1 = 1$ and $\mu_2 > 0$ means that $x_1 < 0$. Then, the only candidate point is $H = (-2, 0)$, which corresponds to $\mu_H = (0, 1/4, 1)$.

$\mathbf{P^{12}}$ ($\mu_1 > 0$ and $g_1 = 0 \Rightarrow x_1^2 = 1$)

$$\begin{aligned} 2\left(\mu_1 - \mu_2\right) x_1 &= 1 \\ 1 - \mu_3 &= 0 \Rightarrow \mu_3 = 1 \\ x_1^2 &= 1 \\ \mu_2\left(x_1^2 - 4\right) &= 0 \Rightarrow \mu_2 = 0 \\ x_1^2 - 4 &\leq 0 \\ \mu_2 &\geq 0 \end{aligned}$$

Conditions $2\left(\mu_1-\mu_2\right)x_1=1$ and $\mu_2=0$ imply $2\mu_1x_1=1$, which implies $x_1>0$. Then, the only candidate point is $E=(1,0)$, with $\mu_E=(1/2,0,1)$.

Of the two candidate points, the globally optimal point is H:

$$\begin{cases} E=(1,0)\Rightarrow f\left(E\right)=1 \\ H=(-2,0)\Rightarrow f\left(H\right)=-2 \end{cases}$$

Esercise 3

Solve the following problem with the KKT-conditions:

$$\begin{aligned}\min f(x) &= x_1^2 + x_2^2 \\ g_1(x) &= x_1^2 + x_2^2 - 4 \leq 0 \\ g_2(x) &= -x_1 - x_2 - 2 \leq 0\end{aligned}$$

Solution

Figure 5.14 represents the feasible region.

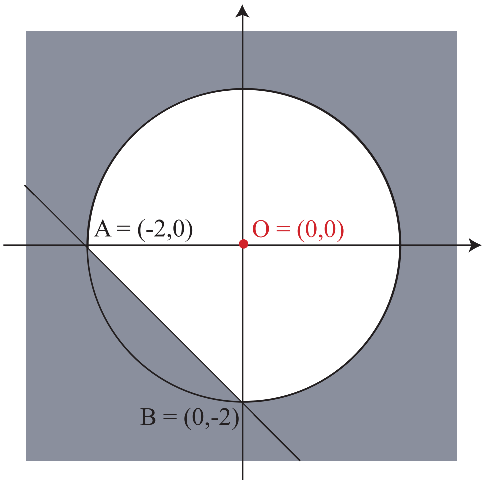


Figure 5.14: Feasible region

Nonregular points

The internal points of the feasible region are all regular. The points through which a single points passes are regular wherever the gradient is nonzero. The gradients of the two constraints are:

$$\begin{aligned}\nabla g_1(x) &= [\,2x_1\; 2x_2\,]' \\ \nabla g_2(x) &= [\,-1\; -1\,]'\end{aligned}$$

The former is zero only in the origin, where the constraint is nonactive; the latter is never zero. The two constraints intersect in

$$\begin{aligned}g_1(x) &= x_1^2 + x_2^2 - 4 = 0 \\ g_2(x) &= -x_1 - x_2 - 2 = 0\end{aligned}$$

that is, in $A = (-2, 0)$ and $B = (0, -2)$. Their gradients are

$$\nabla g_1(A) = [\,-4\; 0\,]' \; \nabla g_2 \equiv [\,-1\; -1\,]' \qquad \nabla g_1(B) = [\,0\; -4\,]' \; \nabla g_2 \equiv [\,-1\; -1\,]'$$

linearly independent. Then, all points are regular.

Karush-Kuhn-Tucker conditions

The generalized Lagrangean function is $\ell(x,\mu) = x_1^2 + x_2^2 + \mu_1\left(x_1^2 + x_2^2 - 4\right) + \mu_2(-x_1 - x_2 - 2)$. Therefore, the first-order necessary conditions are:

$$\begin{array}{rcl} \partial\ell/\partial x_1 & = & 2x_1 + 2\mu_1 x_1 - \mu_2 = 0 \\ \partial\ell/\partial x_2 & = & 2x_2 + 2\mu_1 x_2 - \mu_2 = 0 \\ \mu_1 g_1 & = & \mu_1\left(x_1^2 + x_2^2 - 4\right) = 0 \\ \mu_2 g_2 & = & \mu_2(-x_1 - x_2 - 2) = 0 \\ g_1 & = & x_1^2 + x_2^2 - 4 \leq 0 \\ g_2 & = & -x_1 - x_2 - 2 \leq 0 \\ & & \mu_1 \geq 0 \\ & & \mu_2 \geq 0 \end{array}$$

Subtracting the second constraint from the first one, one obtains

$$2(x_1 - x_2)(1 + \mu_1) = 0$$

Since $\mu_1 \geq 0 \Rightarrow 1 + \mu_1 \geq 1 > 0$, it is $x_1 = x_2 = x$.

$$\begin{array}{rcl} 2x(1+\mu_1) & = & \mu_2 \\ \mu_1\left(x^2 - 2\right) & = & 0 \\ \mu_2(x+1) & = & 0 \\ x^2 & \leq & 2 \\ x & \geq & -1 \\ \mu_1 & \geq & 0 \\ \mu_2 & \geq & 0 \end{array}$$

The simplest constraint is $\mu_2(x+1) = 0$, based on which the original problem P^0 can be split into P^1 $(\mu_2 = 0)$ and P^2 $(\mu_2 > 0 \text{ e } x = -1)$.

$\mathbf{P^1}$ $(\mu_2 = 0)$

$$\begin{array}{rcl} 2x(1+\mu_1) & = & 0 \\ \mu_1\left(x^2 - 2\right) & = & 0 \\ x^2 & \leq & 2 \\ x & \geq & -1 \\ \mu_1 & \geq & 0 \end{array}$$

Since $\mu_1 \geq 0$, it is $x = 0$. This corresponds to the candidate point $O = (0,0)$, with $\mu_0 = (0,0)$. Notice that no constraint is active in O, and correspondingly all multipliers are zero.

$\mathbf{P^2}$ $(\mu_2 > 0$ **e** $x_1 = x_2 = x = -1)$

$$\begin{array}{rcl} -2(1+\mu_1) & = & \mu_2 \\ \mu_1 & = & 0 \\ x_1 = x_2 & = & -1 \\ x^2 & \leq & 2 \\ x & \geq & -1 \\ \mu_1 & \geq & 0 \end{array}$$

which implies $\mu_2 < 0$, and is therefore unfeasible. In fact, in point $(-1, -1)$ the gradient of the objective function is $\nabla f\,(-1, -1) = [\,-2 \;\; -2\,]'$, and it is parallel to the gradient of the only active constraint $\nabla g_2 \equiv [\,-1 \;\; -1\,]'$, but it has the same orientation, instead of the opposite one. This corresponds to the wrong sign of μ_2.

There is, therefore, a single candidate point, that is the globally optimal point: $O = (0, 0)$, with $f\,(O) = 0$.

Exercise 4

Solve the following problem with the KKT-conditions:

$$\begin{aligned}\min f(x) &= x_1 + x_2 \\ h_1(x) &= x_1^2 - x_2 = 0 \\ g_1(x) &= x_1 \leq 0\end{aligned}$$

Solution

Figure 5.15 represents the feasible region.

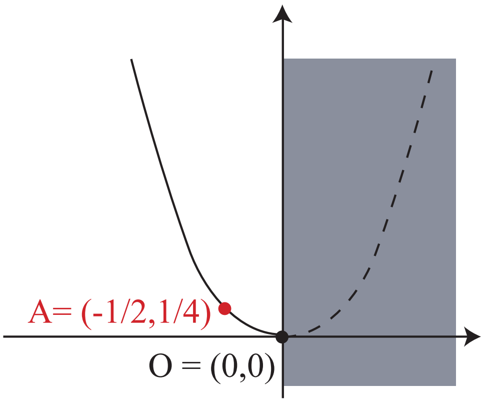


Figure 5.15: Feasible region

Nonregular points

The gradients of the constraints are:

$$\nabla h(x) = [2x_1 \ -1]' \qquad \nabla g(x) = [1 \ 0]'$$

and are never equal to zero. The only potentially nonregular point, therefore, is the intersection of the two constraints, that is the origin, where both constraints are active. The two vectors, however, are linearly independent in the origin:

$$\nabla h(O) = [0 \ -1]' \qquad \nabla g(O) = [1 \ 0]'$$

Hence, all points are regular.

Karush-Kuhn-Tucker conditions

The generalized Lagrangean function is $\ell(x, \lambda, \mu) = x_1 + x_2 + \lambda\left(x_1^2 - x_2\right) + \mu x_1$.

$$\begin{aligned}\partial\ell/\partial x_1 &= 1 + 2\lambda x_1 + \mu = 0 \\ \partial\ell/\partial x_2 &= 1 - \lambda = 0 \Rightarrow \lambda = 1 \\ \partial\ell/\partial\lambda = h &= x_1^2 - x_2 = 0 \\ \mu g &= \mu x_1 = 0 \\ g &= x_1 \leq 0\end{aligned}$$

which implies $\mu = -2x_1 - 1$

$$\begin{aligned} x_2 &= x_1^2 \\ (2x_1 + 1)\, x_1 &= 0 \\ x_1 &\leq 0 \end{aligned}$$

so that $x_1 = -1/2$ e $x_1 = 0$. The first solution is feasible, and corresponds to the candidate point $A = (-1/2, 1/4)$ with $\lambda = 1$ and $\mu = 0$. The second solution is unfeasible, because point $x = (0, 0)$ corresponds to multipliers $\lambda = 1$ and $\mu = -1 < 0$, which is unfeasible.

Hence, the only candidate point is $A = (-1/2, 1/4)$, and it is globally optimal.

Exercise 5

Solve the following problem with the KKT-conditions:

$$\begin{aligned} \min f(x) &= x_2 \\ g_1(x) &= (x_1 - 1)^3 + (x_2 - 2) \leq 0 \\ g_2(x) &= (x_1 - 1)^3 - (x_2 - 2) \leq 0 \\ g_3(x) &= -x_1 \leq 0 \end{aligned}$$

Solution

Figure 5.16 represents the feasible region.

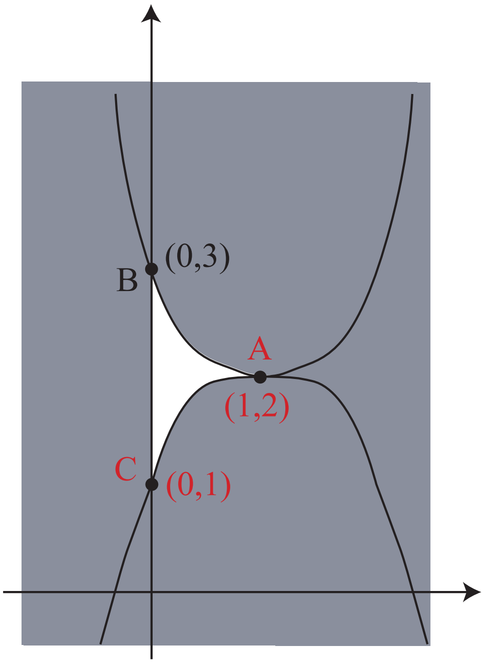


Figure 5.16: Feasible region

Nonregular points

The regular points are the points in which the gradients of the active constraints are linearly independent. Notice that only the active constraints must be considered. Then, all internal points of the feasible region are regular; the points through which a sigle constraint passes are regular wherever its gradient is nonzero, and so on.

The gradients of the three constraints are:

$$\begin{aligned} \nabla g_1(x) &= \left[\, 3(x_1 - 1)^2 \;\; 1 \,\right]' \\ \nabla g_2(x) &= \left[\, 3(x_1 - 1)^2 \;\; -1 \,\right]' \\ \nabla g_3(x) &= [\, -1 \; 0 \,]' \end{aligned}$$

Since no gradient ever becomes zero, we are only interested in the points in which at least two constraints are active. There are three such points, labelled as A, B and C in Figure 5.16.

$$\begin{array}{lll} g_1(A) = 0 & \Rightarrow & (x_1-1)^3 = -(x_2-2) \\ g_2(A) = 0 & \Rightarrow & (x_1-1)^3 = +(x_2-2) \end{array}$$

from which $-(x_2-2) = (x_2-2)$, that is $x_2 = 2$ and $(x_1-1)^3 = 0$, that is $x_1 = 1$. In short, $A = (1,2)$.

$$\begin{array}{lll} g_1(B) = 0 & \Rightarrow & (x_1-1)^3 = -(x_2-2) \\ g_3(B) = 0 & \Rightarrow & x_1 = 0 \end{array}$$

from which $-1 = 2 - x_2$, that is $x_2 = 3$ and therefore $B = (0,3)$. Finally,

$$\begin{array}{lll} g_2(C) = 0 & \Rightarrow & (x_1-1)^3 = (x_2-2) \\ g_3(C) = 0 & \Rightarrow & x_1 = 0 \end{array}$$

which implies $1 = 2 - x_2$, that is $x_2 = 1$ and therefore $C = (0,1)$.

In A, the two gradients $\nabla g_1(A) = [\,0\ 1\,]'$ and $\nabla g_2(A) = [\,0\ \ -1\,]'$ are opposite. Therefore, they are linearly dependent: A is nonregular. In the nonregular points, the Karush-Kuhn-Tucker conditions lose validity, that is they provide no information. Therefore, we are forced to consider all nonregular points as candidates.

In B, the two gradients $\nabla g_1(B) = [\,3\ 1\,]'$ and $\nabla g_3(B) = [\,-1\ 0\,]'$ are linearly independent. Instead of evaluating the determinant of the matrix composed by lining up the two vectors, one can also use the definition of linear independence, and search for a linear combination of the two vectors equal to zero. In other words, setting $\alpha_1 \nabla g_1(B) + \alpha_2 \nabla g_3(B) = 0$, one looks for the values of α_1 e α_2 that satisfy the condition: the result is $[\,3\alpha_1 - \alpha_2\ \alpha_1\,]'$, that is zero only for $\alpha_1 = \alpha_2 = 0$. Hence, the two vectors are linearly indipendent.

Finally, in C the two vectors $\nabla g_2(C) = [\,3\ \ -1\,]'$ and $\nabla g_3(C) = [\,-1\ 0\,]'$ are linearly indipendent. In fact, setting $\alpha_1 \nabla g_2(C) + \alpha_2 \nabla g_3(B) = 0$ one obtains $[\,3\alpha_1 - \alpha_2\ \ -\alpha_1\,]'$, that is zero only for $\alpha_1 = \alpha_2 = 0$.

Karush-Kuhn-Tucker conditions

The generalized Lagrangean function is $\ell(x,\mu) = x_2 + \mu_1 (x_1-1)^3 + \mu_1 (x_2-2) + \mu_2 (x_1-1)^3 - \mu_2 (x_2-2) - \mu_3 x_1$, per cui

$$\begin{array}{rcl} \partial\ell/\partial x_1 & = & 3\mu_1 (x_1-1)^2 + 3\mu_2 (x_1-1)^2 - \mu_3 = 0 \\ \partial\ell/\partial x_2 & = & 1 + \mu_1 - \mu_2 = 0 \\ \mu_1 g_1 & = & \mu_1 \left[(x_1-1)^3 + (x_2-2)\right] = 0 \\ \mu_2 g_2 & = & \mu_2 \left[(x_1-1)^3 - (x_2-2)\right] = 0 \\ \mu_3 g_3 & = & -\mu_3 x_1 = 0 \\ g_1 \leq 0 & = & (x_1-1)^3 + (x_2-2) \leq 0 \\ g_2 \leq 0 & = & (x_1-1)^3 - (x_2-2) \leq 0 \\ g_3 \leq 0 & = & -x_1 \leq 0 \\ & & \mu_1 \geq 0 \\ & & \mu_2 \geq 0 \\ & & \mu_3 \geq 0 \end{array}$$

Before selecting a complementarity constraint to split the problem, we observe that the second equation allows to prove the strict positivity of a multiplier: since $1 + \mu_1 - \mu_2 = 0$, it is $\mu_2 = \mu_1 + 1 \geq 1 > 0$. This implies that μ_2 can be removed from its complementarity constraint, and therefore $(x_1 - 1)^3 - (x_2 - 2) = 0$. The system simplifies to:

$$\begin{aligned}
3\,(2\mu_1 + 1)\,(x_1 - 1)^2 - \mu_3 &= 0 \\
\mu_2 &= 1 + \mu_1 \\
\mu_1\,(x_2 - 2) &= 0 \\
(x_1 - 1)^3 &= (x_2 - 2) \\
\mu_3 x_1 &= 0 \\
2\,(x_2 - 2) &\leq 0 \\
x_1 &\geq 0 \\
\mu_1 &\geq 0 \\
\mu_2 &\geq 0 \\
\mu_3 &\geq 0
\end{aligned}$$

Now it advantageous to select the simplest complementarity constraint, that is the third one, and to split the original problem P^0 into P^1 ($\mu_3 = 0$ and $g_3 \leq 0$) and P^2 ($\mu_3 > 0$ and $g_3 = 0$).

$\mathbf{P^1}$ ($\mu_3 = 0$ **and** $g_3 \leq 0$)

$$\begin{aligned}
(2\mu_1 + 1)\,(x_1 - 1)^2 &= 0 \\
\mu_2 &= 1 + \mu_1 \\
\mu_1\,(x_2 - 2) &= 0 \\
(x_1 - 1)^3 &= (x_2 - 2) \\
x_2 &\leq 2 \\
x_1 &\geq 0 \\
\mu_1 &\geq 0 \\
\mu_2 &\geq 0
\end{aligned}$$

In the first constraint, term $(2\mu_1 + 1)$ is clearly positive, and therefore implies that $x_1 = 1$, and consequently $x_2 = 2$. The remaining constraints do not allow to find a single value for all multipliers: any nonnegative value $\mu_1 \geq 0$ and the corresponding value $\mu_2 = \mu_1 + 1$ are good. Solution $A = (1, 2)$ satisfies Karush-Kuhn-Tucker conditions. Strictly speaking, however, this point should be excluded from the computation, because it is nonregular, and should be considered as candidate for that reason, and not because it satisfies the conditions.

$\mathbf{P^2}$ ($\mu_3 > 0$ and $g_3 = 0$, that is $x_1 = 0$)

$$\begin{aligned}
3\mu_1 + 3\mu_2 - \mu_3 &= 0\\
1 + \mu_1 - \mu_2 &= 0\\
\mu_1\left[-1 + (x_2 - 2)\right] &= 0\\
\mu_2\left[-1 - (x_2 - 2)\right] &= 0\\
-1 + (x_2 - 2) &\le 0\\
-1 - (x_2 - 2) &\le 0\\
\mu_1 &\ge 0\\
\mu_2 &\ge 0\\
\mu_3 &> 0
\end{aligned}$$

Since $\mu_2 = 1 + \mu_1 \ge 1 > 0$, the fourth constraint ($\mu_2\,(1 - x_2) = 0$) becomes $x_2 = 1$, which implies $\mu_1\left[-1 + (x_2 - 2)\right] = -2\mu_1 = 0$ for $\mu_1 = 0$, that implies $\mu_2 = 1$ and $\mu_3 = 3$. Solution $C = (0, 1)$ satisfies Karush-Kuhn-Tucker conditions, and therefore is a candidate point.

Out of the two candidate points, the globally optimal one is C because

$$\begin{cases} A = (1, 2) \Rightarrow f\,(A) = 2\\ C = (0, 1) \Rightarrow f\,(C) = 1 \end{cases}$$

Exercise 6

Solve the following problem with the KKT-conditions:

$$\begin{aligned} \min f(x) &= -x_1 \\ g_1(x) &= (x_1-1)^3 + (x_2-2) \leq 0 \\ g_2(x) &= (x_1-1)^3 - (x_2-2) \leq 0 \\ g_3(x) &= -x_1 \leq 0 \end{aligned}$$

Solution

The problem has the same constraints, and therefore the same feasible region of Exercise 6. Figure 5.17 represents it.

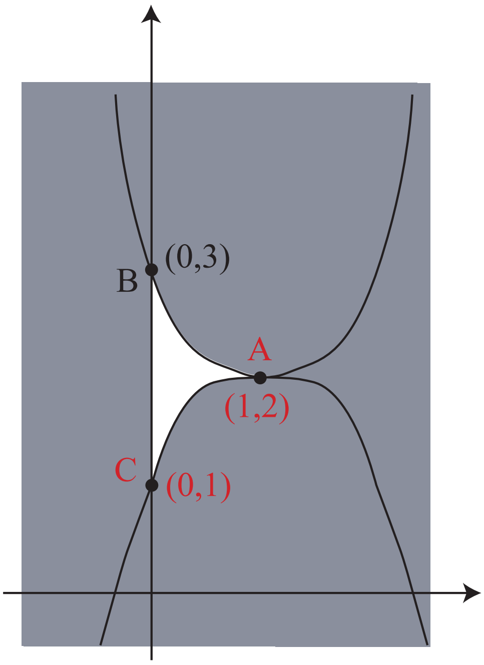


Figure 5.17: Feasible region

Nonregular points

Since the feasible region is the same as that of Exercise 6, the nonregular points are also the same: the only nonregular point is $A = (1, 2)$.

Karush-Kuhn-Tucker conditions

The generalized Lagrangean function is $\ell(x,\mu) = -x_1 + \mu_1 (x_1-1)^3 + \mu_1 (x_2-2) + \mu_2 (x_1-1)^3 - \mu_2 (x_2-2) - \mu_3 x_1$, so that

$$\begin{aligned}
\partial\ell/\partial x_1 &= -1 + 3\mu_1 (x_1-1)^2 + 3\mu_2 (x_1-1)^2 - \mu_3 = 0 \\
\partial\ell/\partial x_2 &= \mu_1 - \mu_2 = 0 \\
\mu_1 g_1 &= \mu_1 \left[(x_1-1)^3 + (x_2-2) \right] = 0 \\
\mu_2 g_2 &= \mu_2 \left[(x_1-1)^3 - (x_2-2) \right] = 0 \\
\mu_3 g_3 &= -\mu_3 x_1 = 0 \\
g_1 \leq 0 &= (x_1-1)^3 + (x_2-2) \leq 0 \\
g_2 \leq 0 &= (x_1-1)^3 - (x_2-2) \leq 0 \\
g_3 \leq 0 &= -x_1 \leq 0 \\
& \mu_1 \geq 0 \\
& \mu_2 \geq 0 \\
& \mu_3 \geq 0
\end{aligned}$$

Since $\mu_1 = \mu_2 = \mu$, the first constraint becomes $6\mu (x_1-1)^2 = \mu_3 + 1 \geq 1 > 0$, which implies that $\mu > 0$. Computing the sum and the difference of the third and fourth constraints, one obtains that $x_1 = 1$ and $x_2 = 2$. Therefore, the first constraint would require $\mu_3 = -1$ and the fourth would require $\mu_3 = 0$, a contradiction. Consequently, no point satisfies the Karush-Kuhn-Tucker conditions.

On the other hand, a globally minimum point certainly exists, because the feasible region is close and limited, and the objective function is continuous. The globally optimal point is A, that can be optimal and violate the KKT-conditions because it is nonregular. Notice that in the previous exercise the same nonregular point actually satisfied the KKT-conditions: both cases are possible. This is why it is always necessary to keep into account the nonregular points as candidates.

Exercise 7

Solve the following problem with the KKT-conditions:

$$\begin{aligned} \min z = (x_1+1)^2 + \left(x_2+\frac{1}{2}\right)^2 & \\ x_1^2 - x_2^2 &\leq 0 \\ x_1 - x_2 &\leq 0 \end{aligned}$$

Solution

This problem has a very peculiar feasible region, that consists in the upper quadrant included between the bisectors of the axes, plus the half-line $x_2 = x_1$ with $x_1 \leq 0$. Figure 5.6 represents the feasible region.

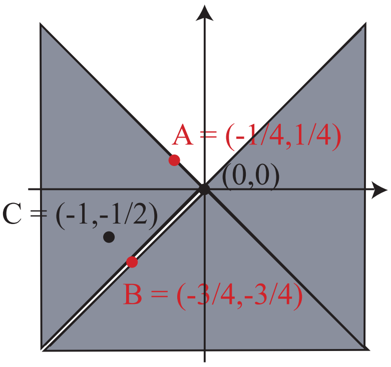


Figure 5.18: Regione ammissibile

Nonregular points

The gradients of the constraints $g_1(x) = x_1^2 - x_2^2 \leq 0$ and $g_2(x) = x_1 - x_2 \leq 0$ are

$$\nabla g_1 = \begin{bmatrix} 2x_1 \\ -2x_2 \end{bmatrix} \qquad \nabla g_2 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

The points in which neither constraint is active are regular by definition. Those in which only g_1 is active, that is those of the bisector of the second and fourth quadrant, excluding the origin ($x_2 = -x_1$, with $x_1 \neq 0$) are regular as long as the gradient is nonzero. This would require $x_1 = x_2 = 0$, that is impossible. Then, these points are all regular. There is no point in which only g_2 is active. The points in which both constraints are active, that is those of the bisector of the first and third quadrant ($x_2 = x_1 = \xi$) are all nonregular, because the two gradients are proportional: $\nabla g_1 = [2\xi \; -2\xi]^T$ and $\nabla g_2 = [1 \; -1]^T$.

Karush-Kuhn-Tucker conditions

The generalized Lagrangean function $\ell\left(x,\mu\right)=\left(x_1+1\right)^2+\left(x_2+\frac{1}{2}\right)^2+\mu_1\left(x_1^2-x_2^2\right)+\mu_2\left(x_1-x_2\right)$ yields the following conditions:

$$\begin{aligned}
\frac{\partial\ell}{\partial x_1} &= 2\left(x_1+1\right)+2\mu_1 x_1+\mu_2=0\\
\frac{\partial\ell}{\partial x_1} &= 2\left(x_2+\frac{1}{2}\right)-2\mu_1 x_2-\mu_2=0\\
\mu_1 g_1\left(x\right) &= \mu_1\left(x_1^2-x_2^2\right)=0\\
\mu_2 g_2\left(x\right) &= -\mu_2\left(x_1-x_2\right)=0\\
\mu_1 &\geq 0\\
\mu_2 &\geq 0\\
g_1\left(x\right) &\leq 0\\
g_2\left(x\right) &\leq 0
\end{aligned}$$

We set apart all points of the bisector of the first and third quadrant, because they are nonregular, and consequently candidate: $x_1-x_2<0$. As a result, the fourth constraint implies that $\mu_2=0$: the multiplier of a nonactive constraint (g_2) is always zero.

$$\begin{aligned}
2\left(x_1+1\right)+2\mu_1 x_1 &= 0\\
2\left(x_2+\frac{1}{2}\right)-2\mu_1 x_2 &= 0\\
\mu_1\left(x_1+x_2\right) &= 0\\
0 &= 0\\
\mu_1 &\geq 0\\
\mu_2 &= 0\\
x_1+x_2 &\geq 0\\
x_1-x_2 &< 0
\end{aligned}$$

We split the problem into two subproblems based on the third constraint.

Problem $\mu_1=0$ The first two constraints yield point $C=(-1,-1/2)$. This violates constraint $g_1\leq 0$, so that is must be rejected. On the other hand, it is a reasonable result, given that it would be the point of minimum of the objective function without the constraints.

Problem $\mu_1>0$ In this subproblem $x_2=-x_1$, so that

$$\begin{aligned}
2\left(x_1+1\right)+2\mu_1 x_1 &= 0\\
2\left(-x_1+\frac{1}{2}\right)+2\mu_1 x_1 &= 0
\end{aligned}$$

The difference of the two constraints yields $x_1=-1/4$ and $x_2=1/4$, while $\mu_1=3$. This is the only points suggested by the Karush-Kuhn-Tucker conditions.

Now, it is necessary to compare point $A=(-1/4,1/4)$ and the nonregular points of line $x_2=x_1=\alpha$. In order to do that, first we determine the point of minimum among the latter, and then we compare it to A.

$$\min f(\alpha) = (\alpha + 1)^2 + \left(\alpha + \frac{1}{2}\right)^2$$

The solution is obtained setting to zero the derivative of f with respect to parameter α: $f'(\alpha) = 2(\alpha + 1) + 2\left(\alpha + \frac{1}{2}\right) = 0$, which implies $\alpha = -3/4$, and therefore $B = (-3/4, -3/4)$, where the objective function value is $f(B) = 1/8$. The value of the objective function in A, on the other hand, is $f(A) = 9/8$. Hence, the globally optimal point is the nonregular point $B = (-3/4, -3/4)$.

Part III

Models with complex preferences

Chapter 6

The Paretian preference

This chapter considers a more complex decision framework: we still assume that there is a single decision-maker and a single scenario, but the preference relation is no longer a weak order described by a consistent value (or cost) function.

In order to state something useful, we must make some assumptions on the structure of the preference relation. Of course, different assumptions will yield different models. The first model proposed in history, that is still among the most important ones is that proposed in the second half of the 19th century by Pareto[1]. We remind that we have assumed the impact to be a vector in a p-dimension space ($F \subseteq \mathbb{R}^p$).

Definition 21 *We denote as* Paretian preference *the relation:*

$$\Pi = \{(f, f') : f_l \leq f'_l \textit{ for each } l \in \{1, \ldots, p\}\}$$

that is

$$f \preceq f' \Leftrightarrow f_l \leq f'_l \textit{ for each } l \in \{1, \ldots, p\}$$

In Pareto's theory, *an impact is preferred to another one when all the elements of the first impact do not exceed the corresponding elements of the second.* Substantially, we assume that *all indicators f_i represent costs.* Of course, there is also a Paretian case in which all indicators represent benefits and the preference relation is defined consequently; here we adopt the same convention adopted to deal with Mathematical Programming.

The Paretian case is a rather reasonable model of practical situations: if the indicators represent costs, it is natural to prefer impacts with small values to impacts with large values. However, this model only allows to compare impacts with elements that dominate each other all in the same orientation. This is called *incomparable indicators assumption*, meaning that the decision-maker is unable to compare two impacts when different indicators give opposite suggestions. This limitation can make the Paretian model unrealistic.

At first sight, the model is also not very useful to reach the actual choice of a solution, given that it terminates proposing several incomparable ones. In practice, the decision-maker could be able to say something more, besides the fact that the indicators, taken one-by-one are costs (or benefits), but could also find hard to provide a complete preference relation. At this point, presenting the Paretian region (as a list of solutions or with a suitable graphic representation), instead of the

[1] Vilfredo Pareto (1848-1923), Italian economist, mathematician and engineer, who was born in France and lived in Italy and Switzerland.

whole feasible region, could help the decision-maker to clarify his/her ideas and to refine the preference relation in a following step, enriching it with new comparable pairs while keeping the ones defined by the Paretian assumption, without retracting from his/her actual feelings. It is even possible to arrive in this way to a complete relation without facing the complications of the multiple-attribute utility theory.

6.1 Formal properties of the Paretian preference

Theorem 18 *The Paretian preference relation is a* partial order*.*

Proof. In fact, it is

- reflexive: $f \preceq f$ for all $f \in F$

$$f_l = f_l \text{ for all } l \in \{1, \ldots, p\} \Rightarrow f_l \leq f_l \text{ for all } l \in \{1, \ldots, p\} \Rightarrow f \preceq f$$

- transitive: $f \preceq f'$ and $f' \preceq f'' \Rightarrow f \preceq f''$ for all $f, f', f'' \in F$

$$\begin{cases} f \preceq f' \\ f' \preceq f'' \end{cases} \Rightarrow \begin{cases} f_l \leq f'_l \text{ for all } l \in \{1, \ldots, p\} \\ f'_l \leq f''_l \text{ for all } l \in \{1, \ldots, p\} \end{cases} \Rightarrow$$

$$\Rightarrow f_l \leq f''_l \text{ for all } l \in \{1, \ldots, p\} \Rightarrow f \preceq f''$$

- antisymmetric: $f \preceq f'$ and $f' \preceq f \Rightarrow f = f'$ for all $f, f' \in F$

$$\begin{cases} f \preceq f' \\ f' \preceq f \end{cases} \Rightarrow \begin{cases} f_l \leq f'_l \text{ for all } l \in \{1, \ldots, p\} \\ f'_l \leq f_l \text{ for all } l \in \{1, \ldots, p\} \end{cases} \Rightarrow$$

$$\Rightarrow f_l = f'_l \text{ for all } l \in \{1, \ldots, p\} \Rightarrow f = f'$$

■

The proofs are very simple, because they exploit the properties of real numbers, and the indicator, considered one-by-one, enjoy the same properties. They can appear as trivial notation tricks, but they are not, because the translation from the relation on the impacts (denoted as $\preceq$) to the relation on the indicators (denoted as $\leq$) is crucial, and is allowed only by Pareto's definition.

Remark 5 *In general, the Paretian preference relation is not complete.*

Example 34 *A trivial counterexample to completeness is provided by the two following impacts:*

$$f = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \qquad f' = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$$

where f is preferred to f' according to the first indicator, but it is worse according to the second. Then, the Paretian preference does not hold in any of the two orientations: the two impacts are incomparable ($f \bowtie f'$).

6.2 Paretian dominance

As discussed in Section 4.1, the strict preference relation between impacts induces a dominance relation between solutions, in which a solution dominates another one if and only if the impact of the former is strictly preferred to the impact of the latter. We use strict preference because in general there can be several solutions with the same impact, which are reciprocally indifferent; stating that such solutions dominate each other would require to reject them all, and this does not appear reasonable.

In the Pareto case, dominance between solutions is therefore expressed as follows:

$$x' \prec x \Leftrightarrow f(x') \prec f(x) \Leftrightarrow \begin{cases} f_l(x') \leq f_l(x) \text{ for all } l \in \{1, \dots, p\} \\ \exists \bar{l} \in \{1, \dots, p\} : f_{\bar{l}}(x') < f_{\bar{l}}(x) \end{cases}$$

and this allows to divide the feasible solutions $x \in X$ into two groups, one of which shall be rejected in every reasonable decision, whereas the other one shall be taken into account.

Definition 22 *We denote as* dominated solution *every feasible solution $x \in X$ that admits another solution $x' \in X$ which dominates it:*

$$\exists x' \in X, \bar{l} \in \{1, \dots, p\} : \begin{cases} f_l(x') \leq f_l(x) \textit{ for all } l \in \{1, \dots, p\} \\ f_{\bar{l}}(x') < f_{\bar{l}}(x) \end{cases}$$

Definition 23 *We denote as* Paretian solution *every feasible solution $x \in X$ such that no other solution x' dominates it:*

$$\forall x' \in X, \exists \bar{l} \in \{1, \dots, p\} : f_{\bar{l}}(x) < f_{\bar{l}}(x') \textit{ or } f(x') = f(x)$$

Definition 24 *We denote as* Paretian region X° *the* set of all Paretian solutions*.*

The Paretian region corresponds to the set of globally optimal points in Mathematical Programming, with two small, but not negligible, differences, due to the incompleteness of the Paretian preference:

1. the Paretian solutions are not preferable to all other solutions, whereas the globally optimal points are;

2. the Paretian solutions are not reciprocally indifferent, whereas the globally optimal points are.

This means that, while in Mathematical Programming usually it is enough to find a globally optimal point and the other ones can be simply ignored, in the Paretian case it is appropriate to find all Paretian solutions.

6.3 Identification of the Paretian region

The fundamental problem now is how to determine the Paretian region. There are several ways to obtain it, which have different application fields, different advantages and disadvantages and in general provide different results. This is because, instead of providing exactly the Paretian region X°, they usually generate an underestimate (i. e., a subset) or an overestimate (i. e. a superset). There is no method that applies in general and yields exactly X°. In the following, we describe:

1. the direct application of the *definition*, that holds only for the finite case;

2. the *inverse transformation method*, that holds only for problems with two indicators ($f \in \mathbb{R}^2$);

3. the *Karush-Kuhn-Tucker conditions*, that hold only for the continuous case and provide an overestimate of X°;
4. the *weighted-sum method*, that holds in all cases, but provides an underestimate of X°;
5. the *ϵ-constraint method*, that holds in all cases, but provides an overestimate of X°.

Table 6.1 summarises the methods here considered, the problems to which they can be applied, their disadvantages and the kind of estimate they provide.

Method	Generality	Disadvantages	Estimate
Definition	Finite problems	Inefficient for combinatorial problems	Exact
Inverse transformation	$p = 2$ indicators	Human intervention	Exact
KKT conditions	Regualar problems	Solving a system	Overestimate
Weighted sum	Complete	Parametric problem	Underestimate
ϵ-constraint	Complete	Parametric problem	Underestimate

Table 6.1: Some methods to determine the Paretian region: generality, disadvantages and kind of estimate obtained

6.3.1 The application of the definition

In the finite case, the Paretian region can be found in finite time by applying the definition that is by building the solution dominance graph, whose nodes correspond to the solutions and whose arcs correspond to the dominance relation, as done in Section 4.1 dealing with general preference relations. In this graph, the Paretian solutions are the nodes with no ingoing arc.

Example 35 *Let us consider once again the problem about a trip introduced in Example 5 of Chapter 4, and let us assume that the impact f be completely described by the travel time f_1 (measured in hours) and by the travel cost f_2 (measured in Euros), with the values reported in Table 6.2. The indicators are both interpreted as cost functions and are reciprocally incomparable.*

	Time (hours)	Cost (Euros)
Train	5.5	100
Car	4.0	150
Airplane	1.0	300
Coach	5.0	180
Taxi	4.0	400

Table 6.2: Travel times and costs associated to different means of transport available for a trip

The resulting preference relation is

$$\Pi = \{(\mathit{Car}, \mathit{Coach}), (\mathit{Car}, \mathit{Taxi}), (\mathit{Airplane}, \mathit{Taxi})\}$$

The Paretian solutions are therefore Train, Airplane and Car. Let us remind that Example 5 in Section 4.1 assumed that Car and Train could be compared, so that *the latter was dominated by the former. The difference here is that the preference relation here is mechanically built on the basis of the values of the two indicators, whereas in that case it was given explicitly listing each pair of impacts.*

Example 36 *Consider the following evaluation matrix, in which the indicators should all be maximised, instead of minimized.*

	Indicators		
Alternatives	f_1	f_2	f_3
a_1	*100*	*60*	*60*
a_2	*70*	*75*	*25*
a_3	*60*	*40*	*20*
a_4	*40*	*100*	*80*
a_5	*20*	*80*	*100*

Performing pairwise comparison, it can be verified that alternative a_3 is dominated by alternatives a_1 and a_2. The other four alternatives are all Paretian.

6.4 Inverse transformation method

This method also consists in applying the definition, but it can be used also on continuous problems, provided that they can be graphically represented in the indicator space. In fact, the method consists in a graphical solution of the problem, and requires to:

- determine the inverse $\phi : F \to X$ of the transformation $f : X \to F$;
- replace $x = \phi(f)$ in the expression of the constraints to obtain an analytic expression of region F in the indicator space;
- determine the subset F° of the nondominated impacts, that is the image of the Paretian region X° in the indicator space; these impact are those whose lower left quadrant does not intersect F;
- determining through the inverse transformation $\phi : F \to X$ the Paretian region X° in the variable space.

Since in general set F° must be determined graphically, this method can be applied only in the case in which there are two indicators ($p = 2$).

Example 37 *Consider the following problem:*

$$\begin{aligned} \min f_1(x) &= x_1 + x_2 \\ \min f_2(x) &= -x_1 \\ g_1(x) &= 3x_1^2 + 4x_2 - 12 \leq 0 \\ g_2(x) &= -x_2 \leq 0 \end{aligned}$$

Figure 6.1 shows the feasible region and the antigradients of the two indicators. First, we determine the inverse transformation $\phi : F \to X$:

$$\begin{cases} f_1(x) = x_1 + x_2 \\ f_2(x) = -x_1 \end{cases} \Rightarrow \begin{cases} x_1(f) = -f_2 \\ x_2(f) = f_1 - x_1 = f_1 + f_2 \end{cases}$$

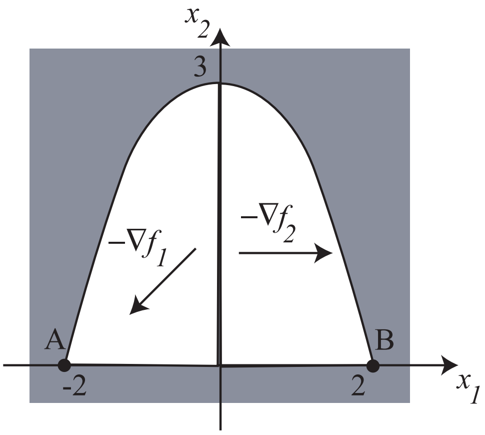


Figure 6.1: Feasible region

which identifies region F (represented in Figure 6.2):

$$\begin{cases} g_1\left(x\left(f\right)\right) = 3f_2^2 + 4\left(f_1 + f_2\right) - 12 \leq 0 \\ g_2\left(x\left(f\right)\right) = -f_1 - f_2 \leq 0 \end{cases} \Rightarrow \begin{cases} f_1 \leq -\frac{3}{4}f_2^2 - f_2 + 3 \\ f_1 + f_2 \geq 0 \end{cases}$$

It is easy to identify graphically the image of the Paretian region: it is enough to find the impacts whose lower left quadrant include no other impact. These points form segment CD, with $C = (-2, 2)$ and $D = (2, -2)$.

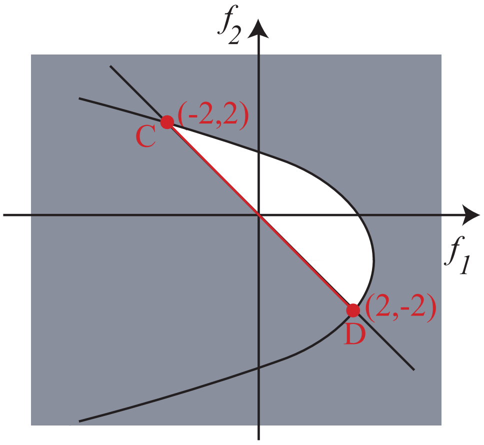


Figure 6.2: The impact set of the problem (in red, the image of the Paretian set)

Then, we apply the inverse transformation to get back to the decision variable space. Since the inverse transformation is linear, is is possible to make a simplified discussion: the line segment CD in the indicator space certainly corresponds to another line segment in the decision variable space. In order to find this segment, it is enough to transform the two extreme points: the Paretian region X° is the segment identified by $x\left(C\right) = A = (-2, 0)$ and $x\left(D\right) = B = (2, 0)$.

In general, we should express the region F° in parametric form as $f\left(\alpha\right)$ and apply the inverse transformation to its points so as to obtain the parametric expression

$x(f(\alpha)) = x(\alpha)$ *of the Paretian region. In the present case*

$$\begin{cases} f_1(\alpha) = \alpha \\ f_2(\alpha) = -\alpha \end{cases} \Rightarrow \begin{cases} x_1(\alpha) = -f_2(\alpha) = \alpha \\ x_2(\alpha) = f_1(\alpha) + f_2(\alpha) = \alpha - \alpha = 0 \end{cases}$$

with $\alpha \in [-2; 2]$.

6.5 Karush-Kuhn-Tucker conditions

These conditions extend the corresponding conditions defined for Mathematical Programming. They are, once again, local conditions, because they exploit a local definition of Paretian dominance. And they are necessary, but not sufficient, conditions, so that they yield a set X^{KKT} of solutions that are candidate to be Paretian, but not certainly such. This set, therefore, is in general larger than the Paretian region: $X^{\text{KKT}} \supseteq X^\circ$. The candidate set needs to be refined in a following step, by sifting away the dominated points.

In the following, we summarise the process, that follows the lines of the process already described in Chapter 5. First, an analytic description of the feasible region X is introduced, that exploits the same inequality constraints and regularity assumptions on the functions that define impact and constraints:

$$X = \{x \in \mathbb{R}^n : g_j(x) \leq 0 \text{ for } j \in \{1, \dots, m\}\} \subseteq \mathbb{R}^n$$

and

$$f = [f_1 \; f_2 \; \dots \; f_p]^T$$

where $g_j(x) \in C^1(X)$ are regular functions for all $j \in \{1, \dots, m\}$ and $f_l \in C^1(X)$ are regular functions for $l \in \{1, \dots, p\}$.

Locally Paretian solutions

Definition 25 *We denote as* locally Paretian solution *a feasible solution* $x \in X$ *that admits a neighbourhood* $\mathcal{U}_{x,\epsilon}$ *in which no other solution dominates it:*

$$\exists \mathcal{U}_{x,\epsilon} : x' \not\preceq x \text{ for all } x' \in \mathcal{U}_{x,\epsilon} \cap X$$

A Paretian solution is always locally Paretian, but the converse is not always true, just as the globally optimal points are always locally optimal, but not the other way round.

Figure 6.3 marks in white on a grey background the set F of the impacts of a decision problem in the indicator space $\mathbb{R}^p$. In this space it is easy to check whether an impact f corresponds to a dominated or a Paretian point x: it is enough to build the lower left quadrant based on f (that is, the set of points with coordinates $\leq f_l$ for each $l \in \{1, \dots, p\}$) and to verify that this quadrant does not include other point of F besides f. For example, in Figure 6.3 solutions A and B are Paretian. Solutions C and D are dominated, since every neighbourhood of the impacts $f(C)$ and $f(D)$ contains points with coordinates not larger, and at least one coordinate strictly smaller (for instance, the points of the segment between $f(B)$ and $f(C)$ are better than $f(C)$ with respect to indicator f_1 and equally good with respect to f_2[2]. Finally, point E is locally Paretian, given that no impact in the intersection of the impact set F and the image of the neighbourhood $\mathcal{U}_{E,\epsilon}$ represented by the

[2] Point C is denoted as *weakly Paretian* because it is not possible to strictly improve all indicator remaining in set F. This concept is out of the scope of the course.

circle falls within the lower left quadrant of $f(E)$, delimited by the dashed lines. However, E is not a globally Paretian point, because impact $f(E)$ is dominated by many other impact, among which $f(A)$, $f(B)$, $f(C)$ and $f(D)$.

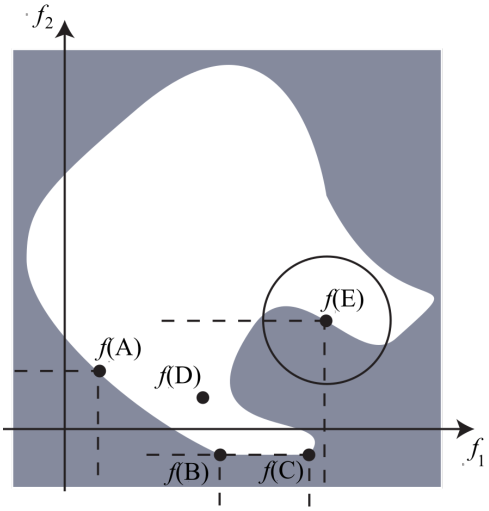


Figure 6.3: Points A and B are globally Paretian points; C is dominated (more precisely, it is weakly Paretian); D is dominated; E is locally, but not globally Paretian (the other four points dominate it).

If X is discrete (see Figure 6.4), all feasible solutions are isolated points. This means that, assigning suitably small values to ϵ, the neighbourhood $\mathcal{U}_{x,\epsilon}$ intersects X only in point x. Therefore, in a discrete problem, every solution is locally Paretian. The same occurred in Mathematical Programming. This is consistent with the fact that in general the KKT-conditions return an overestimate of the Paretian solution set, but in the discrete case they return the whole of X, and are therefore completely useless.

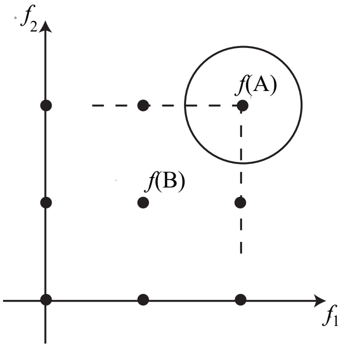


Figure 6.4: In discrete problems, all feasible points are locally Paretian

Sketch of the proof

The development of the conditions is very similar to that of the single-objective case. As in Mathematical Programming, the KKT-conditions are only necessary and provide

1. Definition If a point x is locally Paretian, all the arcs ξ feasible in x are nonimproving.

2. Nonimproving condition Now, we replace all indicators with their first-order approximations through Taylor's series. The directions in which indicator f_l is nonimproving satisfy condition $(\nabla f_l(x))^T p \geq 0$. As in Mathematical Programming, we admit the equality because the higher-order terms could worsen the indicator also along directions orthogonal to the gradient. These conditions are only necessary and could classify as candidates also dominated points (we shall see an example further on). By contrast, the directions that violate all such conditions are strictly improving for all indicators, and therefore sufficient to identify an improving arc and filter out a point from the candidate set.

3. Feasibility condition The feasible arcs are replaced by the associated tangent vectors p_ξ, that (in the regular points) are exactly the directions such that $(\nabla g_j(x))^T p \leq 0$ for all active constraints ($g_j(x) = 0$). Once again, we consider all nonregular points as candidates, because in such points the conditions are unreliable.

4. First geometric interpretation Combining feasibility and nonimprovement implies that, if point $\tilde{x}$ is regular and locally Paretian and p is a vector such that $(\nabla g_j)^T p < 0$ for all constraints active in $\tilde{x}$ (that is, a vector tangent to a feasible arc), then there exists at least one indicator f_l such that $(\nabla f_l)^T p \geq 0$ (that is, an indicator that does not strictly improve along the arc).

From the geometric point of view the vectors p such that $(\nabla g_j)^T p \leq 0$ for all active constraints form the feasible cone, while the vectors such that $(\nabla f_l)^T p \leq 0$ for all indicators also form a cone, composed by the directions in which all the indicators strictly improve. In the single-indicator case, these directions formed a half-space, that we have denoted as improving half-space. By extension, we will denote the new set *improving cone*. The theorem states that no direction of the former cone falls in the interior of the latter: *the feasible cone and the improving cone share no internal points*; they can touch each other, because the KKT-conditions are unreliable in the extreme directions, where some scalar products are zero.

Example 38 *Consider the following problem:*

$$\begin{aligned} \min f_1 &= -2x_1 - x_2 \\ \min f_2 &= -x_1 - 2x_2 \\ g_1(x) &= -x_1 \leq 0 \\ g_2(x) &= -x_2 \leq 0 \\ g_3(x) &= x_1^2 + x_2 - 4 \leq 0 \end{aligned}$$

Let us compute the gradients of the indicators and of the constraint functions:

$$\nabla f_1(x) = \begin{bmatrix} -2 \\ -1 \end{bmatrix} \qquad \nabla f_2(x) = \begin{bmatrix} -1 \\ -2 \end{bmatrix}$$

$$\nabla g_1(x) = \begin{bmatrix} -1 \\ 0 \end{bmatrix} \qquad \nabla g_2(x) = \begin{bmatrix} 0 \\ -1 \end{bmatrix} \qquad \nabla g_3(x) = \begin{bmatrix} 2x_1 \\ 1 \end{bmatrix}$$

Figure 6.5 shows two typical situations: in point $A = (2, 0)$, the feasible cone intersects the improving cone. In fact, moving from A upwards and leftwards one

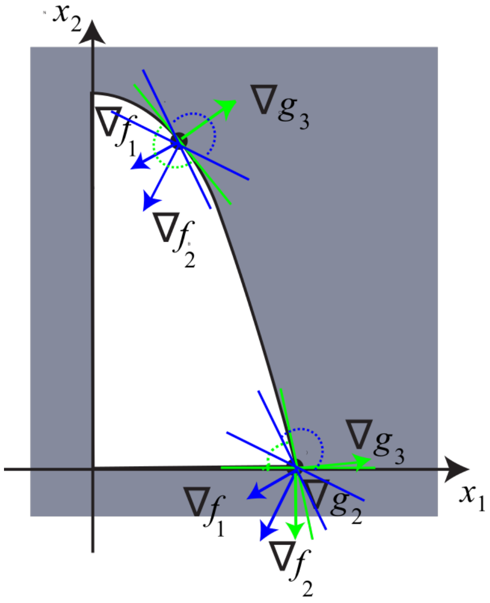


Figure 6.5: Interpretazione geometrica delle condizioni di KKT: il cono ammissibile e il cono migliorante non devono avere direzioni interne comuni

can find a cone of directions which allow to keep in X while decreasing both the value of f_1 and of f_2. In point $B = (1/2, 15/4)$, on the contrary, the two cones do not intersect. In fact, moving from B in any direction, either one gets out of X or one of the two indicators worsens.

5. Farkas' lemma The application of Farkas' lemma is somewhat more complicated than in Mathematical Programming, because there is no longer a single vector on one hand and a cone on the other. The workaround is to introduce an auxiliary vector and apply the lemma twice: the first time to the auxiliary vector and the feasible cone, the second to the opposite of the auxiliary vector and the improving cone.

The existence of two cones that do not intersect, but can touch each other, is equivalent to the existence of a hyperplane separating them. Let γ be one of the two vectors normal to the separating hyperplane: one of the two cones lies on the same side of γ with respect to the hyperplane, and the other lies on the opposite side. With no loss of generality, we assume that the feasible cone lie on the opposite side of γ, that is

$$p^T \gamma \leq 0 \text{ for all } p : p^T \nabla g_j(x) \leq 0$$

By Farkas' lemma, γ lies inside the cone of the gradients of the active constraints:

$$\exists \mu_j \geq 0 : \gamma = \sum_{j=1}^{m} \mu_j \nabla g_j(x) \text{ for all } j : g_j(x) = 0$$

Conversely, the improving cone lies on the side of vector $-\gamma$, and therefore the gradients of the indicators lie on the opposite side:

$$p^T(-\gamma) \leq 0 \text{ for all } p : p^T \nabla f_l(x) \leq 0 \text{ for all } l$$

Therefore, by Farkas' lemma, $-\gamma$ lies inside the cone of the gradients of the indicators:

$$\exists w_l \geq 0 : -\gamma = \sum_{l=1}^{p} w_l \nabla f_l$$

Summing the two equations, one obtains the thesis:

$$\exists w_l \geq 0, \mu_j \geq 0 : \sum_{l=1}^{p} w_l \nabla f_l + \sum_{j=1}^{m} \mu_j \nabla g_j = 0$$

These are the KKT-conditions for the Paretian case.

6. Second geometric interpretation From the geometric point of view, the KKT-conditions for Mathematical Programming were interpreted stating that the antigradient of the objective function must fall inside the cone of the gradients of the active constraints. In the Paretian case, the KKT-conditions can be interpreted stating that *the cone of the antigradients of the indicators intersects the cone of the gradients of the active constraints.*

Example 39 *Taking again into account Figure 6.5, it can be remarked that in A the two cones do not intersect, whereas in B they do. In fact,...*[3]

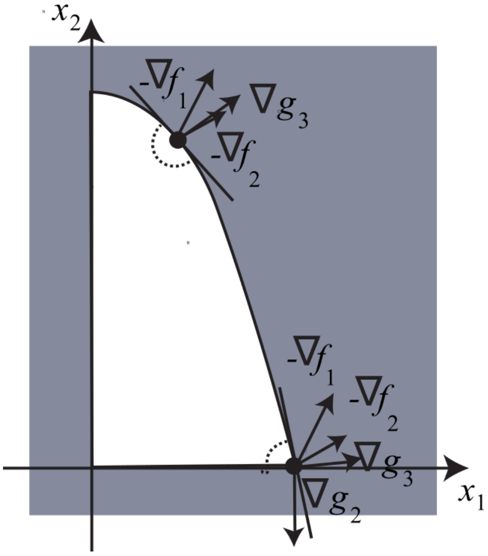


Figure 6.6: Second geometric interpretation of KKT-conditions: the cone of the antigradients of the indicators must intersect the cone of the gradients of the active constraints

7. Standard form of the KKT-conditions As in Mathematical Programming, the nonactive constraints can be reintroduced in the system of conditions by imposing that their multipliers be zero, thanks to the usual complementarity conditions $\mu_j g_j(x) = 0$, and the equality constraints can be treated without explicitly replacing them with pairs of inequalities through the use of free multipliers.

[3] THIS EXAMPLE MUST BE COMPLETED: COEFFICIENTS μ and w CONFIRMING THE INTERSECTION ARE (for example) $\mu_3 = 1$ and $w_1 = w_2 = 1/3$.

Notice that, given a point x and a set of multipliers w_l and μ_j that satisfy the conditions, the same point x satisfies them for multipliers αw_l and $\alpha \mu_j$ for any factor $\alpha > 0$. In other words, the multipliers nave a degree of freedom that does not affect the solution. In order to remove this useless degree of freedom, a normalization condition can be imposed. Usually, this is done by dividing all multpliers by $\sum_{l=1}^{p} w_l$, so that

$$\sum_{l=1}^{p} w_l = 1$$

Of course, this requires to prove that $\sum_l w_l > 0$, but this is true because the gradients of the active constraints are linearly independent in all regular points, and therefore $\sum_j \mu_j \nabla g_j \neq 0$.

Example 40 *APPLY THE NORMALIZATION TO POINTS A AND B*

The Paretian region as a parametric hypersurface

In Mathematical Programming, applying the KKT-conditions means to determine n decision variables x_i and m multipliers μ_j, solving a system of $n+m$ equations: the n equation that set to zero the partial derivatives of the generalized Lagrangian function and the m equations that impose the complementarity conditions. Since the number of equations and variables is the same[4], the system has in general a finite number of solutions. The $2m$ inequalities that impose the feasibility of the candidate points and the nonnegativity of the multipliers μ_j can affect that number reducing it.

The KKT-conditions for the Paretian case yield a system with n decision variables x_i, m multipliers μ_j and p multipliers w_l, which are constrained by $n+m+1$ equations (the last one is the normalization condition on the w_l multipliers). In general, the system has ∞^{p-1} solutions, which can be interpreted as a parametric $(p-1)$-dimension hypersurface expressed in the form $x = x(w)$.

Example 41 *Consider the following problem:*

$$\begin{array}{rcl}\min f_1 & = & -2x_1 - x_2 \\ \min f_2 & = & -x_1 - 2x_2 \\ g_1(x) & = & -x_1 \leq 0 \\ g_2(x) & = & -x_2 \leq 0 \\ g_3(x) & = & x_1^2 + x_2 - 4 \leq 0\end{array}$$

Figure 6.7 represents the feasible region. The two indicators f_1 and f_2 are cost functions, to be minimised, and their gradients are

$$\nabla f_1(x) = \begin{bmatrix} -2 \\ -1 \end{bmatrix} \quad \nabla f_2(x) = \begin{bmatrix} -1 \\ -2 \end{bmatrix}$$

The arrows in the figure represent the antigradients, that is the improvement directions. The gradients of the constraints are:

$$\nabla g_1(x) = \begin{bmatrix} -1 \\ 0 \end{bmatrix} \quad \nabla g_2(x) = \begin{bmatrix} 0 \\ -1 \end{bmatrix} \quad \nabla g_3(x) = \begin{bmatrix} 2x_1 \\ 1 \end{bmatrix}$$

[4] It remains the same if equality constraints are treated with free multipliers, since there are s additional variables λ and s equality constraints $h_i(x) = 0$

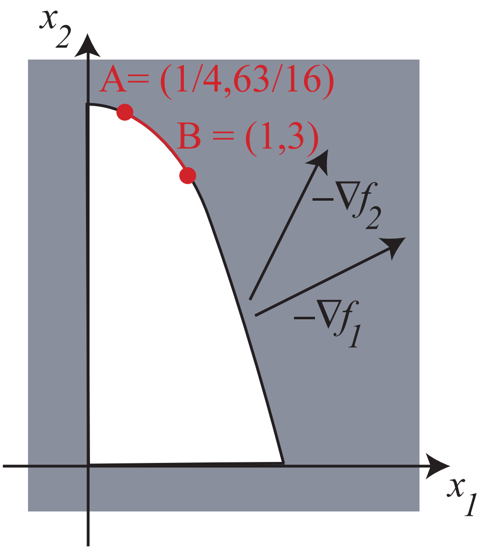


Figure 6.7: Identification of the Paretian region

Since $w_1 + w_2 = 1$, we can set $w_1 = w$ e $w_2 = 1 - w$. The inequalities $w_1 \geq 0$ and $w_2 \geq 0$ therefore reduce to $0 \leq w \leq 1$. The KKT-condition system reduces to:

$$\begin{aligned}
-2w - 1 + w - \mu_1 + 2\mu_3 x_1 &= 0 \\
-w - 2 + 2w - \mu_2 + \mu_3 &= 0 \\
\mu_1 g_1(x) = -\mu_1 x_1 &= 0 \\
\mu_2 g_2(x) = -\mu_2 x_2 &= 0 \\
\mu_3 g_3(x) = -\mu_3 \left(x_1^2 + x_2 - 4\right) &= 0 \\
0 \leq w &\leq 1 \\
\mu_1 &\geq 0 \\
\mu_2 &\geq 0 \\
\mu_3 &\geq 0 \\
g_1(x) = -x_1 &\leq 0 \\
g_2(x) = -x_2 &\leq 0 \\
g_3(x) = x_1^2 + x_2 - 4 &\leq 0
\end{aligned}$$

From the first condition, $2\mu_3 x_1 = w + 1 + \mu_1 \geq 1 > 0$. Then, $x_1 > 0$ and $\mu_3 > 0$. The third condition guarantees that $\mu_1 = 0$ and the fifth that $x_2 = 4 - x_1^2$. Now we know that the Paretian region is an arc of the parabola $g_3(x) = 0$.

$$\begin{aligned}
x_1 &= \frac{w+1}{2\mu_3} \\
\mu_2 &= w-2+\mu_3 \\
\mu_1 &= 0 \\
\mu_2\left(4-x_1^2\right) &= 0 \\
0 &\leq w \leq 1 \\
\mu_2 &\geq 0 \\
\mu_3 &> 0 \\
0 &< x_1 \leq 2
\end{aligned}$$

Let us split problem P_0 into subproblems P_1 ($\mu_2 = 0$) and P_2 ($\mu_2 > 0$ and $4 - x_1^2 = 0$)

P_1 ($\mu_2 = 0$) *From the second constraints, we obtain $\mu_3 = 2 - w$, and therefore the parametric line in (x_1, x_2) (that is an hypersurface in $p - 1 = 1$ dimension):*

$$\begin{cases} x_1 = \dfrac{w+1}{2\,(2-w)} \\ x_2 = 4 - \left[\dfrac{w+1}{2\,(2-w)}\right]^2 \end{cases} \quad \text{with } w \in [0;1]$$

The Paretian region is an arc of the parabola $x_2 = 4 - x_1^2$. To identify the arc, we have to determine which points correspond to the values of w included between 0 and 1. For $w = 0$, it is $x\,(0) = A = (1/4, 63/16)$. For $w = 1$, it is $w\,(1) = B = (1,3)$. In order to be sure that intermediate values of w correspond to intermediate points along the arc, it is enough to observe that $x_1\,(w)$ is increasing, given that

$$\frac{dx_1}{dw_1} = \frac{1 \cdot 2\,(2-w) + 2\,(w+1)}{4\,(2-w)^2} = \frac{3}{2\,(2-w)^2} > 0$$

P_2 ($\mu_2 > 0$ **and** $4 - x_1^2 = 0 \Rightarrow x_1 = 2$)

$$\begin{aligned}
\mu_3 &= \frac{w+1}{4} \\
\mu_2 &= w - 2 + \frac{w+1}{4} = \frac{5w-7}{4}
\end{aligned}$$

Since $w \in [0;1]$, $\mu_2 < 0$, which is unfeasible. Hence, this problem does not provide any candidate locally Paretian point.

From the geometric point of view, it can be remarked that, in all points of the Paretian region, the cone of the feasible directions does not intersect the cone of the improving directions. Moreover, the cone of the gradients of the active constraints intersects the cone of the antigradients of the indicators.

In the following example, the KKT-conditions determine a set that is strictly larger than the Paretian region. This confirms that they are only necessary conditions.

Example 42 *Consider the following problem:*

$$
\begin{aligned}
&\min f_1 = -x_2 \\
&\min f_2 = -x_1 - x_2 \\
&g_1(x) = x_1 - 1 \leq 0 \\
&g_2(x) = -x_1 \leq 0 \\
&g_3(x) = -x_1^2 + x_2 - 1 \leq 0
\end{aligned}
$$

Figure 6.8 represents the feasible region of the problem. The arrow indicate the antigradients of the indicators, that is the directions of quickest improvement.

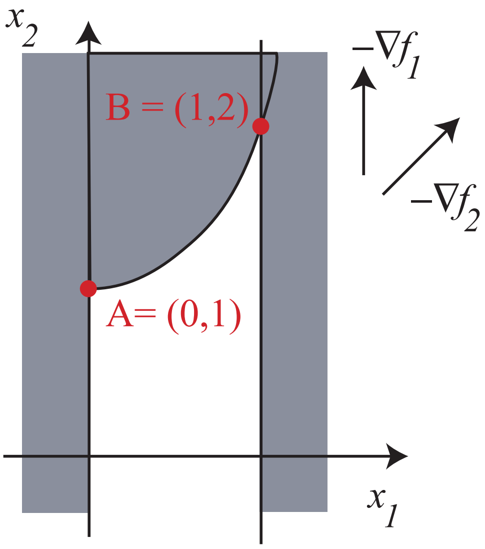


Figure 6.8: Identification of the Paretian region: point A respects the KKT-conditions, and is therefore candidate, even though it is not even locally Paretian

The gradients of the objective functions are:

$$\nabla f_1(x) = \begin{bmatrix} 0 \\ -1 \end{bmatrix} \quad \nabla f_2(x) = \begin{bmatrix} -1 \\ -1 \end{bmatrix}$$

while the gradients of the constraints are

$$\nabla g_1(x) = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \quad \nabla g_2(x) = \begin{bmatrix} -1 \\ 0 \end{bmatrix} \quad \nabla g_3(x) = \begin{bmatrix} -2x_1 \\ 1 \end{bmatrix}$$

As in the previous example, we take into account the normalization of the mul-

tipliers w_l. The overall system to solve is:

$$\begin{aligned}
w - 1 + \mu_1 - \mu_2 - 2\mu_3 x_1 &= 0\\
-w + w - 1 + \mu_3 &= 0\\
\mu_1 g_1(x) = \mu_1(x_1 - 1) &= 0\\
\mu_2 g_2(x) = -\mu_2 x_1 &= 0\\
\mu_3 g_3(x) = -\mu_3\left(-x_1^2 + x_2 - 1\right) &= 0\\
0 \leq w &\leq 1\\
\mu_1 &\geq 0\\
\mu_2 &\geq 0\\
\mu_3 &\geq 0\\
g_1(x) = x_1 - 1 &\leq 0\\
g_2(x) = -x_1 &\leq 0\\
g_3(x) = -x_1^2 + x_2 - 1 &\leq 0
\end{aligned}$$

The second condition implies that $\mu_3 = 1$, so that constraint $g_3(x) = 0$ is active: the Paretian region lies on the parabola.

$$\begin{array}{rcl}
w - 1 + \mu_1 - \mu_2 - 2x_1 & = & 0\\
\mu_1(x_1 - 1) & = & 0\\
\mu_2 x_1 & = & 0\\
x_2 & = & x_1^2 + 1\\
0 & \leq & w \leq 1\\
\mu_1 & \geq & 0\\
\mu_2 & \geq & 0\\
0 & \leq & x_1 \leq 1
\end{array}$$

Let us split the problem P_0. Given the form of the conditions, we split P_0 into three subproblems: P_1 ($x_1 = 0$ and $\mu_1 = 0$), P_2 ($0 < x_1 < 1$ and $\mu_1 = \mu_2 = 0$) and P_3 ($x_1 = 1$ and $\mu_2 = 0$).

P_1 ($x_1 = 0$ **and** $\mu_1 = 0$) *The first constraints implies that $\mu_2 = w - 1$. Since $\mu_2 \geq 0$, the only feasible value is $w = 1$, from which $\mu_2 = 0$. Point $A = (0, 1)$ is candidate, with $\mu = (0, 0, 1)$ and $w = 1$. Yet, A is not even locally Paretian, because moving rightwards f_1 keeps constant and f_2 improves (and moving slightly upwards, it is even possible to improve f_1). In fact, this is consistent with the KKT-conditions, given that they only require that at least one of the indicators do not improve when approximated by its first-order Taylor expression. Just to be sure, we must consider as a candidate a point in which even a single indicator keeps constant at the first order while all the other ones improve.*

P_2 ($0 < x_1 < 1$ **and** $\mu_2 = 0$) *The first constraint implies that $\mu_1 = \mu_2 = 0$, and therefore $x_1 = (w - 1)/2$. However, the constraint $0 < x_1 < 1$ would require $1 < w < 3$, that is unfeasible. Hence, this subproblem does not provide any candidate point.*

P_3 ($x_1 = 1$ **and** $\mu_2 = 0$) *The first constraints implies that $\mu_1 = 3 - w$, the fourth one implies that $x_2 = 2$. Then, Point $B = (1, 2)$ is candidate, with $\mu = (3 - w, 0, 1)$*

for each $w \in [0;1]$. This is a degenerate case, in which infinite solutions of the KKT-conditions correspond to a single candidate point.

In summary, the problem has two candidate points; only B is actually Paretian.

6.6 The weighted-sum method

The weighted-sum method consists in building a linear combination of the indicators and minimising it. The following theorem proves that using combinations with strictly positive coefficients always yields a Paretian solution. It is a very simple and frequently used method, but we shall see that it has several disadvantages.

Theorem 19 *Let $z(x) = \sum_{l=1}^{p} w_l f_l(x)$ be a convex combination of the indicators, with coefficients $w_l > 0$ for $l \in \{1, \dots, p\}$ and $\sum_{l=1}^{p} w_l = 1$. If x° is a globally optimal point for problem*

$$\begin{array}{rcl} \min z(x) & = & \displaystyle\sum_{l=1}^{p} w_l f_l(x) \\ x & \in & X \end{array}$$

then x° is a globally Paretian point with respect to the impact $f(x)$.

Proof. The proof is by contradiction. If x° were not a Paretian point, by Definition 22 there should exist another point $x' \in X$ such that $x' \prec x^\circ$, that is $f_l(x') \leq f_l(x^\circ)$ for each $l \in \{1, \dots, p\}$ and an index $\bar{l}$ such that $f_{\bar{l}}(x') < f_{\bar{l}}(x^\circ)$. But this would imply

$$\begin{cases} w_l f_l(x') & \leq w_l f_l(x^\circ) \ \text{ for all } l \in \{1, \dots, p\} \setminus \{\bar{l}\} \\ w_{\bar{l}} f_{\bar{l}}(x') & < w_{\bar{l}} f_{\bar{l}}(x^\circ) \end{cases}$$

Notice that the second property holds only if $w_{\bar{l}}$ is strictly positive. Summing the p inequalities, one obtains

$$\sum_{l=1}^{p} w_l f_l(x') < \sum_{l=1}^{p} w_l f_l(x^\circ) \Rightarrow z(x') < z(x^\circ)$$

and therefore x° cannot be a globally optimal point for $z(x)$. From the contradiction, the thesis follows. ■

This result appears very similar to the KKT-conditions, but it is on the contrary quite different:

1. the hypothesis of the theorem requires that solution x° be a globally optimal point, not simply to satisfy the KKT-conditions;

2. the coefficients w_l used in the combination must be strictly positive, and not simply nonnegative.

For these two reasons, the assumptions of the weighted-sum method are much stricter than the KKT-conditions. This makes them sufficient to guarantee that a point satisfying them be Paretian, instead of only necessary. Consequently, the resulting set X^{ws} is a subset of X°, instead of a superset.

Remark 6 *Should Theorem 19 only require x° to be a globally optimal point, without imposing to the weights w_l to be strictly positive, there could exist a point x' with values $f_l(x') = f_l(x^\circ)$ for the indicators with positive weight and with values $f_l(x') < f_l(x^\circ)$ for the indicators with zero weight; point x° would still be a globally optimal point for $z(x)$, but it would be dominated by x', and therefore the thesis would no longer hold.*

Example 43 *Let us consider once again the problem of Example 41, represented in Figure 6.7. We apply the weighted-sum method combining the two indicators f_1 e f_2 with coefficients $w_l > 0$. The objective function*

$$z(x) = w_1 f_1 + w_2 f_2$$

has an antigradient which is equal to the combination of the antigradients ∇f_1 and ∇f_2 with the same coefficients w_1 e w_2.

As the coefficients vary, this vector always falls inside the cone identified by the two combined antigradients. This suggests how to solve the problem graphically: one obtains a single globally optimal point, that, as the coefficients vary, moves along the arc AB of the parabola in Figure 6.7. The method does not provide the whole Paretian region, that would be the whole arc, but only the arc deprived of its two extreme points. In fact, these points, that correspond to $w = [1\ 0]^T$ and $w = [0\ 1]^T$ cannot be obtained using strictly positive coefficients.

Example 44 *Let us consider once again the problem of Example 41. Applying the weighted-sum method, we obtain for all values of the coefficients w_l point B. This is correct, because that point is the only Paretian one. Then, in this case the subset identified by the method coincides with the whole Paretian region.*

In order to discuss the strengths and the limitations of the weighted-sum method, we need some definitions.

Definition 26 *Given a Paretian solution $x^\circ \in X^\circ$, we denote as* support *of x° the set $\mathrm{Supp}(x^\circ) \subset \mathrm{R}^p$ of all vectors w such that $w_l > 0$ for all $l \in \{1, \ldots, p\}$, $\sum_{l=1}^{p} w_l = 1$ and x° is a globally optimal point for the problem $\min\limits_{x \in X} \sum_{l=1}^{p} w_l f_l(x)$.*

Definition 27 *We denote as* supported solution *a Paretian solution $x^\circ \in X^\circ$ whose support is nonempty:* $\mathrm{Supp}(x^\circ) \neq \emptyset$.

Example 45 *Consider the minimum spanning tree problem on the complete graph K_3 (that is trivially a triangle with three vertices and three edges) with the two cost functions reported in Table 6.3.*

f_1	1	2	3
1	-	1	3
2	1	-	6
3	3	6	-

f_2	1	2	3
1	-	13	10
2	13	-	8
3	10	8	-

Table 6.3: Two cost functions defined on the edges of graph K_3

The KKT-conditions are useless in this case, because the problem is discrete, and therefore all solutions are locally Paretian. The inverse transformation method is usually impossible to apply in combinatorial problem, because X consists of an exponential number of isolated points in the decision variable space, and corresponds to *an exponential set of isolated impacts in the indicator space $\mathbb{R}^2$. Finding the image of the Paretian region in this space requires to survey all the possible impacts. In this specific case, however, the task is easy, because the problem has only $n = 3$ decision variables, associated to the three edges of the graph: $x_{12}, x_{13}, x_{23} \in \{0, 1\}$, and the feasible region includes only three solutions: $X = \{(1,1,0), (1,0,1), (0,1,1)\}$. These three solutions correspond to three different impacts, represented in Figure 6.9. All three impacts have an empty lower left quadrant, and therefore all three feasible solutions are globally Paretian.*

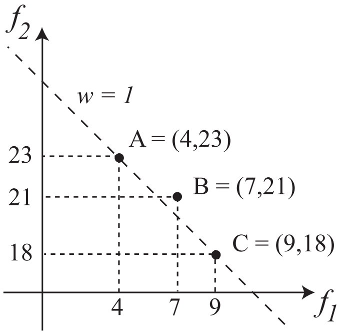


Figure 6.9: All three feasible solutions are globally Paretian, but B, that is the tree $\{(1,2), (2,3)\}$, which is not optimal for any value of w

Let us apply the weighted-sum method, combining the objectives f_1 and f_2 with the weights w and $(1-w)$, where $w \in (0,1)$, and solving the minimum spanning tree problem for each value of w. The costs of the edges are

- $f(1,2) = w \cdot 1 + (1-w) \cdot 13 = 13 - 12w$
- $f(1,3) = w \cdot 3 + (1-w) \cdot 10 = 10 - 7w$
- $f(2,3) = w \cdot 6 + (1-w) \cdot 8 = 8 - 2w$

Let us compute the optimal solution with Prim's algorithm, starting from vertex 1 and applying it parametrically, that is considering all possible values of parameter $w \in (0,1)$ and gradually subdividing all the cases and subcases that this implies[5]*. The minimum cost edge incident in vertex 1 can be either $(1,2)$ or $(1,3)$ according to the value of w, and in particular to the sign of the difference $f(1,2) - f(1,3) = (13 - 12w) - (10 - 7w) = 3 - 5w$:*

1. *if $w \leq 3/5$, edge $(1,3)$ is cheaper, and becomes the first edge of the minimum spanning tree. At this point, we must find the minimum cost edge in the cut that separates vertices 1 and 3 from vertex 2, that is the minimum between $f(1,2)$ and $f(3,2)$:*

$$f(1,2) - f(3,2) = (13 - 12w) - (8 - 2w) = 5 - 10w$$

   The comparison generates two subcases:

   (a) *if $w \leq 1/2$, edge $(3,2)$ is cheaper, so that the minimum spanning tree is $\{(1,3), (3,2)\}$, which costs $18 - 9w$*

   (b) *if $1/2 \leq w \leq 3/5$, edge $(1,2)$ is cheaper, so that the minimum spanning tree is $\{(1,3), (1,2)\}$, which costs $23 - 19w$*

[5] Of course, any other correct algorithm for the minimum spanning tree problem could be applied, for example Kruskal's algorithm.

2. *if $w \geq 3/5$, edge $(1,2)$ is cheaper, and becomes the first edge of the minimum spanning tree. At this point, we must find the minimum cost edge in the cut that separates vertices 1 and 2 from vertex 3, that is the minimum between $f(1,3) = 10 - 7w$ and $f(2,3) = 8 - 2w$:*

$$f(1,3) - f(2,3) = (10 - 7w) - (8 - 2w) = 2 - 5w$$

   Since this expression is always negative for $w \geq 3/5$, the minimum spanning tree is $\{(1,2),(1,3)\}$, which costs $23 - 19w$.

In summary, only two Paretian points have been found.

However, the convex combination of the two objectives yields a value of the combined objective function equal to $21 - 14w$ in the third solution. This value is always larger than the value of the other two Paretian solutions. Hence, the third solution cannot be determined with the weighted-sum method: it is nonsupported.

6.6.1 Advantages and disadvantages

The weighted-sum method has some strong advantages:

1. it can be applied to any problem, even the ones for which the KKT-conditions are useless (e. g., the discrete problems);
2. it is very intuitive, as often the decision-makers find it natural to combine the indicators by summing them, after choosing units of measure for which their numerical values are not too different;
3. it only requires to build an auxiliary objective for the problem, keeping every other feature, and in particular the feasible region; quite often, a valid algorithm for the optimization of the single indicators is also valid for the optimization of the auxiliary objective.

At the same time, however, it has strong disadvantages:

1. it requires to consider all possible values of the weight vector w, which form a continuous infinite set; this requires a parametric solution, that is in general nontrivial, and for large-size problems can become intractable;
2. it requires to find all globally optimal solutions for each value of w (one is not enough, contrary to Mathematical Programming);
3. it does not provide the whole Paretian region X°, but only the supported solutions, which could even be a very restricted underestimate.

Theorem 20 *In a combinatorial problem, the number of nonsupported solutions can grow exponentially with respect to the number n of the decision variables; correspondingly, the weighted-sum method can provide a fraction of the Paretian region gradually converging to zero.*

Example 46 [*] *An example illustrating Theorem 20 can be rather easily built for the biobjective shortest path problem. Examples for other combinatorial problems with any number of indicators can be built applying similar ideas. Consider the graph reported in Figure[6] composed by $2k+3$ nodes: an origin node s, an intermediate node*

[*] This example provides advanced concepts, that are not part of the course's syllabus.

[6] THE FIGURE WITH THE GRAPH IS NOT YET AVAILABLE.

i_0, a destination node t and k pairs of nodes i_r and i'_r for $r = 1, \ldots, k$. Three arcs go from node s to nodes i_k (with cost $f(s, i_k) = \left[2^k + 1\ 0\right]^T$), i'_k (with cost $f(s, i'_k) = \left[0\ 2^k + 2\right]^T$) and i_0 (with cost $f(s, i_k) = [1\ 1]^T$). Two arcs exit from each of the nodes i_r with $r = 0, \ldots, k-1$ and from each of the nodes i'_r with $r = 1, \ldots, k-1$, and go to the nodes of the following pair, i_{r+1} and i'_{r+1}, with cost $f(i_r, i_{r+1}) = f(i'_r, i_{r+1}) = [2^r\ 0]^T$ and $f\left(i_r, i'_{r+1}\right) = f\left(i'_r, i'_{r+1}\right) = [0\ 2^r]^T$. Finally, a single arc connects nodes i_k and i'_k to node t, with cost $f(i_k, t) = f(i'_k, t) = [2\ 2]^T$. By construction, there are 2^k+2 different paths from node s to node t: the path (s, i_k, t), whose cost is $\left[2^k + 2\ 0\right]^T$, the path (s, i'_k, t), whose cost is $\left[0\ 2^k + 2\right]^T$ and 2^k paths go from s to i_0 crossing the sequence of the node pairs (i_r, i'_r) visiting exacty one node of each pair. By construction, each of these paths costs $\left[h + 2\ 2^k + 1 - h\right]^T$, with $h = 0, \ldots, 2^k - 1$, so that in these paths f_1 assumes all integer values from 2 to $2^k + 1$, while f_2 correspondingly assumes the complementary value, and their sum is always $f_1 + f_2 = 2^k + 1$. If all these impacts are drawn in plane $f_1 f_2$, one realizes that the image of the family of paths passing by i_0 consists of the integer points of a segment of the line $f_1 + f_2 = 2^k + 1$, whereas the other two paths represent extreme points lying on the axes f_1 and f_2. All solutions are Paretian. On the other hand, for every convex combination of weights, the optimal solution of the single-objective auxiliary problem is one of the two extreme paths; all other solutions are non supported. Their number grows exponentially with k, that is with the number of decision variables of the problem, that is equal to the number of arcs, $4k+5$. The fraction of nonsupported solutions is 2^k over $2^k + 2$, and obviously converges to the whole solution set as the number of variables increases.

In order to avoid the need for a parametric algorithm, a possible approach is to *sample* the set of weights $w \in (0;1)^p$ and to list the optimal solutions found for each considered weight. This approach simplifies the computation, because the algorithm used is the same that optimises the single indicators, but it introduces two new difficulties:

1. the result is not even guaranteed to include all supported solutions, so that further Paretian solutions are lost;
2. several different sampled values of the weights could produce the same Paretian solution, so that the method can be computationally inefficient, even when compared to the parametric one.

Relations between the weighted-sum method and *MAUT*

There are strong similarities between the weighted-sum method to determine the Paretian region and the *MAUT*, but the context is completely different:

1. the weighted-sum method does not require complete preferences, nor mutual preferential independence or the corresponding trade-off condition, but only that the attributes be costs or benefits;
2. the weighted-sum method does not require to determine a specific weight vector w, but it scans all possible vectors with strictly positive components;
3. the weighted-sum method does not provide an "optimal" solution, but a subset of the Paretian region;
4. the weighted-sum method directly combines the attributes, without filtering them through utility functions suitably built so as to reflect the relative utilities associated to different values of an indicator; consequently, it cannot access

the nonsupported solutions; on the contrary, suitable utility functions (that is, suitably shaped indifference curves) allow the overall utility function to admit such solutions as optimal.

Example 47 *Consider the minimum spanning tree problem discussed in Example 45 of Section 6.6, and assume that the two attributes, which both represent costs, are associated to the following utility functions*

$$u_1(f_1) = \sqrt{9 - f_1} \qquad u_2(f_2) = \sqrt{23 - f_3}$$

The problem admits only three feasible solutions, corresponding to impacts $f_A = (4, 23)$, $f_B = (7, 21)$ and $f_C = (9, 18)$.

FIGURA DA AGGIUNGERE, CON LE CURVE DI INDIFFERENZA NELLO SPAZIO DEGLI INDICATORI E NELLO SPAZIO DELLE UTILITA'

All three solutions are Paretian, but the weighted sum method is able to identify only A and C, because B is an unsupported solution, nested inside a concavity of the impact set, and is nonoptimal for any linear combination of the two costs. By contrast, nonlinear indifference curves could imply that $f(B)$ is actually preferable to $f(A)$ and $f(C)$. Correspondingly, a nonlinear utility functions, such as the one above mentioned, allows to bend the indifference curves in the utility component space, making it explicit that B is also Paretian, and that it can be optimal for suitable utility functions. In fact,

$$u(f) = u_1(f_1) + u_2(f_2) \Rightarrow \begin{cases} u(f(A)) = \sqrt{5} \\ u(f(B)) = 2\sqrt{2} \\ u(f(C)) = \sqrt{5} \end{cases}$$

so that $u(f(B)) > u(f(A)) = u(f(C))$, and $f(B) \prec f(A) \sim f(C)$.

In principle, any Paretian solution admits a utility function for which it is optimal.

6.6.2 The weighted-sum method in Linear Programming*

Since the globally optimal points of Linear Programming problems are the vertices of the feasible region (or the faces identified by optimal vertices), and since the weighted-sum method provides solutions that are optimal for some combination of the objectives with weights w_l, the Paretian solution thus obtained will always fall on the vertices or, possibly, on the faces. It can be proved that in this case the method provides the whole Paretian region, and that therefore it is only necessary to determine the Paretian vertices. To this purpose, there exist *ad hoc* methods, which we do not treat. Let us see an example, solved with an explicit parametric method (analysis of cases and subcases).

Example 48 *Consider the folowing problem with linear indicators and constraints:*

$$\begin{aligned} \min f_1 &= -x_1 + 3x_2 \\ \min f_2 &= 4x_1 - x_2 \\ -x_1 + x_2 &\leq 7/2 \\ x_1 + x_2 &\leq 11/2 \\ 2x_1 + x_2 &\leq 9 \\ x_1 &\leq 4 \\ x_1, x_2 &\geq 0 \end{aligned}$$

*This section provides advanced concepts, that are not part of the course's syllabus.

for which Figure 6.10 reports the feasible region and the improving directions of the objective functions.

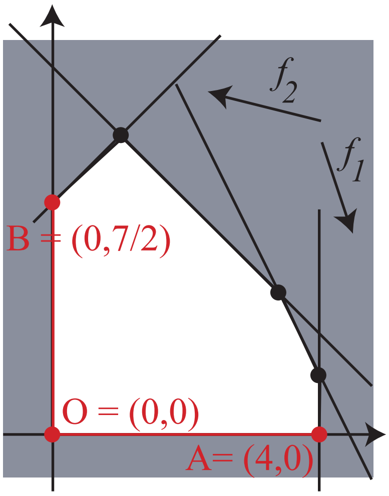


Figure 6.10: In Linear Programming problems, the Paretian region coincides with the part of frontier of the feasible regione that is identified by the vertices which are globally optimal for suitable values of the weights

The weighted-sum method combines the objectives with a coefficient vector $w \in (0;1)^p$*, obtaining a single objective function* $z(x) = (4-5w)\,x_1 + (4w-1)\,x_2$*. Then, it solves to optimality the resulting problem, for example with the simplex algorithm applied parametrically. Luckily, the problem is already in a feasible canonical basis form, with the following* tableau*:*

0	$4-5w$	$4w-1$	*0*	*0*	*0*	*0*
7/2	*-1*	*1*	*1*	*0*	*0*	*0*
11/2	*1*	*1*	*0*	*1*	*0*	*0*
9	*2*	*1*	*0*	*0*	*1*	*0*
4	*1*	*0*	*0*	*0*	*0*	*1*

Case $0 < w < 1/4$ *The reduced cost of the second column is negative, so that it is necessary to perform a* pivot *operation on element* $a_{12} = 1$*, letting variable* x_2 *replace variable* x_3 *in the basis. The resulting* tableau *is*

$7/2\,(1-4w)$	$3-w$	*0*	$1-4w$	*0*	*0*	*0*
7/2	*-1*	*1*	*1*	*0*	*0*	*0*
2	*2*	*0*	*-1*	*1*	*0*	*0*
11/2	*3*	*0*	*-1*	*0*	*1*	*0*
4	*1*	*0*	*0*	*0*	*0*	*1*

where all reduced costs are positive. The optimal solution is $x^\circ = (0, 7/2)$ *and its value is* $z^* = 7/2\,(4w-1)$*.*

Case $w = 1/4$ *This limit case corresponds to the situation in which the canonical form with respect to basis* x_3, x_4, x_5, x_6 *is optimal, but variable* x_2 *can enter the basis without worsening the current solution, so that the basis* x_2, x_4, x_5, x_6 *is also optimal. In fact, all solutions that are convex combination of* $(0,0)$ *and* $(0,7/2)$ *are optimal. It is an infinite set of Paretian points, forming a whole segment. Notice* *that the gradient of the combined objective function is perfectly perpendicular to constraint* $x_1 = 0$. *In fact, for* $w = 1/4$, *it is* $f(x) = 15/4\,x_1$.

Case $1/4 < w < 4/5$ *The current canonical form corresponds to an optimal basis. The optimal solution is* $x^\circ = (0,0)$, *whose value is zero.*

Case $w = 4/5$ *This limit case corresponds to the situation in which the canonical form with respect to basis* x_3, x_4, x_5, x_6 *is optimal, but variable* x_1 *can enter the basis without worsening the current solution, so that the basis* x_1, x_3, x_4, x_5 *is also optimal. In fact, all solutions that are convex combinations of* $(0,0)$ *and* $(4,0)$ *are optimal. It is an infinite set of Paretian points, forming a whole segment. Notice that the gradient of the combined objective function is perfectly perpendicular to constraint* $x_2 = 0$. *In fact, for* $w = 4/5$, *it is* $f(x) = 12/5\,x_2$.

Case $4/5 < w < 1$ *The reduced cost of the first column is negative, and therefore it is necessary to perform a* pivot *operation on element* $a_{41} = 1$, *letting variable* x_1 *replace variable* x_6 *in the basis.*

$4(5w-4)$	0	$4w-1$	0	0	0	$5w-4$
15/2	0	1	1	0	0	1
3/2	0	1	0	1	0	-1
1	0	1	0	0	1	-2
4	1	0	0	0	0	1

Luckily, the optimality conditions are now satisfied (since $w > 4/5$, *all reduced costs are nonnegative). The optimal solution is* $x^\circ = (4,0)$ *and its value is* $z^* = 4(4-5w)$.

6.6.3 Sensitivity analysis

.

Sensitivity analysis is in general the study of how the solution of a problem changes as one or more numerical parameters are modified. In the weighted-sum method, sensitivity analysis consists in partitioning the region $\{w \in \mathbb{R}^p : w_l > 0, \sum_{l=1}^{p} w_l = 1\}$ into the supports of the single solutions, that is into the subsets corresponding to which each supported solution is optimal for the single-objective problem $\sum_{l=1}^{p} w_l f_l$.

If the problem is finite, the cardinality of these subsets is finite. If, in addition, it has only two objectives, the set of all possible weight vectors w is the line segment identified by points $(0,1)$ and $(1,0)$ in plane $w_1 w_2$. In that case, it is possible to plot the profile of the objective function for each given solution depending on one of the two weights (the other weight is determined by the normalization condition). This study allows to easily determine the support of each solution, that is a closed interval of values of the remaining weight.

Example 49 *Consider once again Example 45, which concerned a minimum spanning tree problem. Figure 6.11 shows the behaviour, as weight* w *varies between* 0 *and* 1, *of the combined objective* $f_w(x) = w f_1(x) + (1-w) f_2(x)$ *for each of the three solutions* $x \in X$. *We remind that in this case the problem has two weights, but the normalization condition reduces them to a single independent one. For each fixed* x, *then, the objective* $f_w(x)$ *is a simple function of a single variable* w. *Moreover, such a function is linear by construction. Therefore, in order to plot it, it is enough to compute its values in two points, such as* $w = 0$ *and* $w = 1$.

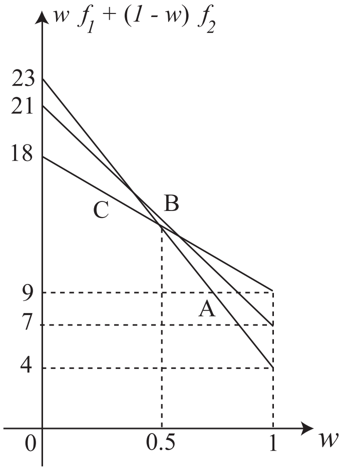


Figure 6.11: The support of solution C is $(0, 1/2]$; the support of solution A is $[1/2, 1)$; solution B is globally Paretian, but has an empty support.

The figure shows that the three linear profiles evolve so that for small values of w the Paretian solution identified is C and for large values it is A. Solution B is never identified, even though it is Paretian, because its support is empty.

6.7 The ϵ-constraint method

The ϵ-constraint method consists in replacing all indicators but one with constraints that require the solution to respect a quality threshold, and in solving the resulting auxiliary problem. The following theorem proves that all Paretian solutions are optimal points for a problem of this family, provided that the thresholds are suitably determined. It is a frequently used method, with an intuitive interpretation, given that it corresponds to keep a single indicator and to replace the search for an optimal value of the other ones with the search for a satisfying value.

Theorem 21 *If x° is a globally Paretian point for indicators $f_l(x)$ in X, then for any index $l^* \in \{1, \ldots, p\}$, chosen* ad libitum, *x° is a globally optimal point for problem:*

$$\begin{array}{rcll} \min f_{l^*}(x) & & & \\ f_l(x) & \leq & \epsilon_l & l \in \{1, \ldots, p\} \setminus \{l^*\} \\ x & \in & X & \end{array}$$

where $\epsilon_l = f_l(x^\circ)$ for all $l \in \{1, \ldots, p\} \setminus \{l^*\}$.

Proof. By contradiction, assume that x° is not a globally optimal point for this problem. Then, there exists a point $x' \in X$ such that $f_l(x') \leq \epsilon_l$ for all $l \neq l^*$ and $f_{l^*}(x') < f_{l^*}(x^\circ)$. Since $\epsilon_l = f_l(x^\circ)$, such a point dominates x° according to the definition of Paretian dominance. Since this contradicts the hypothesis, x° must be a globally optimal point. ■

As in the weighted-sum method, a globally optimal point for an auxiliary problem is taken into account. The idea is to solve the auxiliary problem parametrically for all values of the coefficients ϵ_l. This time, however, the condition appears as the thesis of the theorem, and is therefore a necessary condition, not a sufficient one. Each solution, then, is a candidate point, instead of a certainly Paretian point; computing all globally optimal points for all values of the parameters, the candidate set provides a superset of the Paretian region. A candidate point x° could be non Paretian for the original problem. This happens when there exists another globally optimal point $x^{\circ\prime}$, in which f_{l^*} has the same value and at least one of the other indicators has a better value. Notice that each indicator l^* provides in general a different superset, and it is possible to intersects them to get a more precise estimate of the Paretian region[7]. Anyway, the method generates only weakly Paretian points.

The coefficients ϵ_l are also denoted as *standards* (and the method is also called the *standards method*) because the technique corresponds to focusing on a single objective, neglecting the other ones, but setting for each of them a performance threshold. Since there are $p-1$ standards, reciprocally independent, the set obtained letting them vary parametrically is a $(p-1)$-dimension parametric hypersurface, exactly as in the weighted-sum method[8].

Example 50 *A nuclear waste dump must be located at a distance not larger than* 2 *from the plant that produces it, that is situated in point* $(1, 0)$*). Moreover, the dump must keep north of the plant and out of a park whose border runs along line* $x_1 + x_2 = 3$*. The park lies in the north-east side of the line. While respecting such constraints, the dump must be as far as possible from a town situated at the origin of the axes, and at the same time as far as possible from the park.*

$$\begin{aligned} \max f_1(x) &= x_1^2 + x_2^2 \\ \min f_2(x) &= x_1 + x_2 \\ g_1(x) &= (x_1-1)^2 + x_2^2 \leq 4 \\ g_2(x) &= x_1 + x_2 \leq 3 \\ g_3(x) &= -x_2 \leq 0 \end{aligned}$$

Figure 6.12 represents the feasible region of the problem. The first objective aims to keep the dump far away from the origin, so that its improving direction varies from point to point and is oriented radially. The second objective, on the contrary, improves moving towards the lower left.

We can set up the ϵ*-constraint method by turning either indicator* f_1 *or indicator* f_2 *into a constraints. In general, it is preferable to operate on linear objectives, because the resulting constraints are easier to deal with, especially in a graphical resolution process. Let us proceed, therefore, with respect to* f_2[9]*.*

$$\begin{aligned} \max f_1(x) &= x_1^2 + x_2^2 \\ g_1(x) &= (x_1-1)^2 + x_2^2 \leq 4 \\ g_2(x) &= x_1 + x_2 \leq 3 \\ g_3(x) &= -x_2 \leq 0 \\ g_4(x) &= x_1 + x_2 \leq \epsilon \end{aligned}$$

Any solution method is acceptable. For example, one could apply the KKT-conditions, but their solution is quite hard, so that we resort to a graphical approach.

[7] My feeling is that the intersection provides exactly the Paretian region, but as long as I do not provide a proof, I prefer to avoid stating this property.

[8] In general, it is a larger surface, but it has the same number of dimensions.

[9] In this case, selecting f_1 would probably be viable: the additional constraint would be $x_1^2 + x_2^2 \geq \epsilon_1$, which would force the solution of the auxiliary problem to stay outside a circle centred in the origin and having a suitable radius.

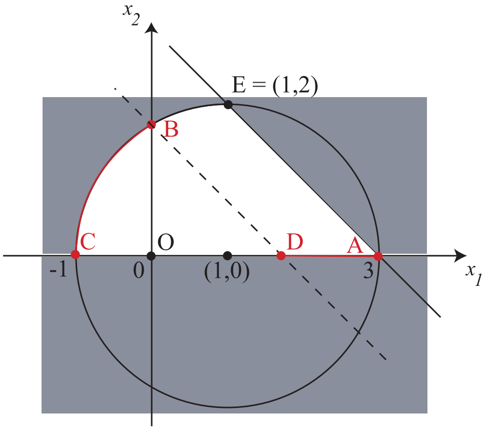


Figure 6.12: Feasible region and Paretian region

It is clear that the parametric constraint $x_1 + x_2 \leq \epsilon$ corresponds to a half-space gradually restricting the feasible region as ϵ decreases. Moreover, the separating line is parallel to the border of the park. In detail:

- *for $\epsilon \geq 3$, the ϵ-constraint does not remove any feasible solution. The globally optimal point is the farthest one from the origin: it is not difficult to see that such point is $A = (3, 0)$.*
- *for $\sqrt{3} < \epsilon < 3$, the ϵ-constraint becomes tighter than the park constraint ($g_2(x) \leq 0$): the globally optimal point lies on axis x_1, and it is point $(\epsilon, 0)$, which gradually moves from A to $D = \left(\sqrt{3}, 0\right)$.*
- *for $\epsilon = \sqrt{3}$, the intersections of the ϵ-constraint, respectively with the circumference $g_1(x) = 0$ (point $B = \left(0, \sqrt{3}\right)$) and with axis x_1 (point $D = \left(\sqrt{3}, 0\right)$) have the same distance from the origin, and are both globally optimal points.*
- *for $-1 \leq \epsilon \leq \sqrt{3}$, the intersection of the ϵ-constraint with the circumference $g_1(x) = 0$ becomes farther from the origin than the intersection with the axis x_1; the globally optimal point lies therefore on the circumference and gradually moves from $B = \left(0, \sqrt{3}\right)$ to $C = (-1, 0)$.*
- *for $\epsilon < -1$, there are no longer feasible solutions.*

As a consequence, the ϵ-constraint method returns as a candidate Paretian set the union of segment AD and of circumference arc BC. In fact, that is exactly the Paretian set.

Notice that the points of segment between $(1, 0)$ and D are locally Paretian, given that the improving directions for both indicators are the directions oriented towards the lower right more downwards than rightwards), and are all unfeasible.

Other points would satisfy KKT-conditions, even though they are not even locally Paretian. The points of the bisector of the first quadrant are not locally Paretian (moving orthogonally with respect to the bisector gets farther away from the origin while keeping the same distance from the park), but they are candidate anyway, because the two indicators have opposite gradients, and therefore $w_1 \nabla f_1 + w_2 \nabla f_2 = 0$ (the gradients of the constraints play no role, because no constraint is active in these points).

Example 51 *Consider the same problem of the previous exercise, but reverse the sign of indicator f_2 (for example, line $g_2(x) = 0$ might represent a state border which cannot be trespassed, but to which one wants to get as close as possible so as to dump abroad the negative effects of the waste.*

$$\begin{aligned} \max f_1(x) &= x_1^2 + x_2^2 \\ \max f_2(x) &= x_1 + x_2 \\ g_1(x) &= (x_1 - 1)^2 + x_2^2 \leq 4 \\ g_2(x) &= -x_1 - x_2 \leq -3 \\ g_3(x) &= -x_2 \leq 0 \end{aligned}$$

Rather intuitively, this problem has a degenerate Paretian region, that consists only of point $A = (3,0)$, that is at the same time the farthest one from the town and the closest one to the border.

Let us solve the problem turning indicator f_1 (with the opposite sign, as it must be maximised) into a constraint.

$$\begin{aligned} \min f_2(x) &= -x_1 - x_2 \\ g_1(x) &= (x_1 - 1)^2 + x_2^2 \leq 4 \\ g_2(x) &= -x_1 - x_2 \leq -3 \\ g_3(x) &= -x_2 \leq 0 \\ g_5(x) &= -x_1^2 - x_2^2 \leq \epsilon \end{aligned}$$

The ϵ-constraint corresponds to requiring a minimum distance from the origin, that is to excluding the points of a circle centred in the origin:

- *for $\epsilon \geq -9/2$, all points of segment AE, with $A = (3,0)$ and $E = (1,2)$ are feasible for the threshold and optimal with respect to the remaining objective; therefore, they are all candidate.*
- *for $-5 \leq \epsilon < -9/2$, circumference $g_5(x) = 0$ removes some of those points: the only globally optimal points are those on the segment between A and $\left(\left(3 + \sqrt{-2\epsilon - 9}\right)/2, \left(3 - \sqrt{-2\epsilon - 9}\right)/2\right)$ and those on the segment between E and $\left(\left(3 - \sqrt{-2\epsilon - 9}\right)/2, \left(3 + \sqrt{-2\epsilon - 9}\right)/2\right)$.*
- *for $-9 \leq \epsilon \leq -5$, the globally optimal points are those on the segment between A and $\left(\left(3 + \sqrt{-2\epsilon - p}\right)/2, \left(3 - \sqrt{-2\epsilon - p}\right)/2\right)$.*
- *for $\epsilon < -9$, there is no feasible solution.*

Therefore, the ϵ-constraint method returns the whole segment AE as a candidate, though only point A is Paretian. It is true that, using indicator f_1 as an objective and imposing a threshold on f_2 would provide the correct region.

6.7.1 Advantages and disadvantages

The ϵ-constraint method has some strong advantages:

1. it can be applied to any problem, even the ones for which the KKT-conditions are useless (e. g., the discrete problems);
2. it is very intuitive, as often the decision-makers find it natural to focus on a single indicator and to translate their preferences into minimum satisfaction

thresholds (see again in Figure 1.2 the iterative decision process according to Simon, which terminates when the decision-makers are satisfied with the result);

3. it provides an overestimate of the Paretian region that is often quite tight and that can be easily refined applying again the method with a different choice of the main indicator.

At the same time, however, it has strong disadvantages:

1. it requires to consider all possible values of the thresholds ϵ, which form a continuous infinite set; this requires a parametric solution, that is in general nontrivial, and for large-size problems can become intractable;
2. it requires to find all globally optimal solutions for each value of ϵ (one is not enough, contrary to Mathematical Programming);
3. in combinatorial problems, while changing the objective function usually does not affect the computational complexity of the problem (the polynomial ones remain polynomial), introducing additional constraints nearly always increases the complexity (for example, a polyomial problem turns into an $\mathcal{NP}$-complete one).

Relations between the ϵ-constraint method and the lexicographic preference with aspiration levels

Notice the affinity between the ϵ-constraint method to determine the Paretian region and the lexicographic preference with aspiration levels: the aspiration levels correspond to a special choice of the thresholds used in the ϵ-constraint method. The main difference is that the thresholds ϵ_l assume all possible values, instead of being fixed once for all at the beginning. It must be admitted, however, that the decision-maker might also make some experiment, adapting the choice of the aspiration levels based on the results they produce. This makes the method even closer to a variant of the ϵ-constraint method based on sampling. The relation between the two methods suggests that the generated solution be, if not Paretian, at least candidate; in fact, optimizing the indicators step-by-step guarantees that the generated solution is actually Paretian.

Remark 7 *The lexicographic order method with aspiration levels generates by construction a Paretian solution, for any chosen order and any value of the aspiration levels that produce a nonempty restricted feasible region.*

Proof. The property can be proved by contradiction, assuming the generation of a dominated solution, which would respect the same aspiration levels of the dominating one and at the same time would be worse for one or more indicators, thus contradicting the selection process. ■

6.7.2 The ϵ-constraint method in Linear Programming[*]

As for the weighted-sum method, Linear Programming is an easier special case, in which *ad hoc* methods identify exactly the Paretian region. This requires some adaptation, because the basic method still returns an overestimate (contrary to the weighted-sum method). The adaptations, however, are rather straighforward.

[*] This section presents advanced material; it is not part of the exam programme.

We do not treat the *ad hoc* methods, but simply show an example of an explicit parametric method, analysing cases and subcases.

Example 52 *Let us solve again, this time with the ϵ-constraint, the Linear Programming problem considered in Example 48.*

$$\begin{array}{rcl} \min f_1 & = & -x_1 + 3x_2 \\ \min f_2 & = & 4x_1 - x_2 \\ -x_1 + x_2 & \leq & 7/2 \\ x_1 + x_2 & \leq & 11/2 \\ 2x_1 + x_2 & \leq & 9 \\ x_1 & \leq & 4 \\ x_1, x_2 & \geq & 0 \end{array}$$

Figure 6.13 reports the feasible region of the problem and the improving directions of the objective functions.

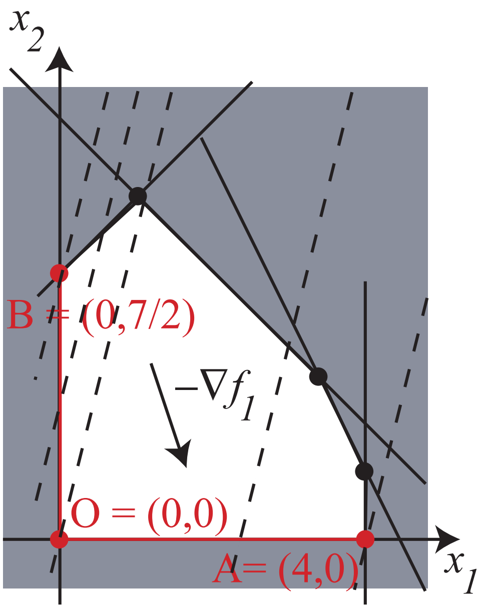


Figure 6.13: Feasible region and improving directions in a linear problem

Let us translate objective f_2 into an inequality:

$$\begin{array}{rcl} \min f_1 & = & -x_1 + 3x_2 \\ -x_1 + x_2 & \leq & 7/2 \\ x_1 + x_2 & \leq & 11/2 \\ 2x_1 + x_2 & \leq & 9 \\ x_1 & \leq & 4 \\ 4x_1 - x_2 & \leq & \epsilon \\ x_1, x_2 & \geq & 0 \end{array}$$

Luckily, we are already in a feasible canonical basis form. The corresponding tableau *is:*

0	-1	3	0	0	0	0	0
7/2	-1	1	1	0	0	0	0
11/2	1	1	0	1	0	0	0
9	2	1	0	0	1	0	0
4	1	0	0	0	0	1	0
ϵ	4	-1	0	0	0	0	1

First of all, let us make all reduced costs nonnegative: this requires to pivot *either on* $a_{41} = 4$ *or on* $a_{51} = 1$*, respectively for* $4 < \epsilon/4$ *and* $4 \geq \epsilon/4$*.*

Case $\epsilon > 16$ *A simplex iteration on* $a_{41} = 4$ *returns the following* tableau.

4	0	3	0	0	0	1	0
15/2	0	1	1	0	0	1	0
3/2	0	1	0	1	0	-1	0
1	0	1	0	0	1	-1	0
4	1	0	0	0	0	1	0
ϵ-16	0	-1	0	0	0	-1	1

which corresponds to a global optimum in $x = (4, 0)$*.*

Case $\epsilon \leq 16$ *A* pivot *operation on* a_{51} *returns the following* tableau.

ϵ/4	0	11/4	0	0	0	0	1/4
7/2+ϵ/4	0	3/4	1	0	0	0	1/4
11/2-ϵ/4	0	5/4	0	1	0	0	-1/4
9-ϵ/2	0	3/2	0	0	1	0	-1/2
4-ϵ/4	0	1/4	0	0	0	1	-1/4
ϵ/4	1	-1/4	0	0	0	0	1/4

If the right-hand-sides are all nonnegative, we have reached the optimum. Otherwise, the dual simplex method must be applied. The former situation holds foor $0 \leq \epsilon \leq 16$*, so that there are two subcases.*

Subbase $0 \leq \epsilon \leq 16$ *We reached the optimum with solution* $x^\circ(\epsilon) = (\epsilon/4, 0)$*, which means that the Paretian region runs along the* x_1 *axis from* A *to* O*.*

Subcase $\epsilon < 0$ *A dual simplex iteration must be performed on* $a_{52} = -1/4$*, yielding:*

3ϵ	11	0	0	0	0	0	3
7/2+ϵ	3	0	1	0	0	0	1
11/2+ϵ	5	0	0	1	0	0	1
9+ϵ	6	0	0	0	1	0	1
4	1	0	0	0	0	1	0
-ϵ	-4	1	0	0	0	0	-1

Once again, we have two subcases: if $-7/2 \leq \epsilon < 0$*, we reached the optimum in* $x = (0, -\epsilon)$*, which means that the Paretian region runs along the* x_2 *axis from* O *to* B*.*

If, on the contrary, $\epsilon < -7/2$, we should make a further step, but there is no feasible (nonnegative) pivot *element. This corresponds to the situation in which the dual problem is unlimited, which implies that the primal problem is unfeasible. In fact, the first constraint has only nonnegative coefficients and a strictly negative right-hand side, so that there is no way to satisfy it. This corresponds to the case in which no solution respects the quality threshold.*

In conclusion, the Paretian region consists of the two line segments AO and OB, as already determined with the weighted-sum method.

6.8 Exercises[*] [†]

The following exercises are often solved only with one of the methods described in the chapter, but they can be solved also with other ones. Of course, different methods can provide different approximations of the Paretian region. In particular, the finite problems can be solved with all methods, but the KKT-conditions return the whole feasible region. The infinite two-dimensional problems can be solved with all methods, except for the definition.

Exercise 1

Consider the following biobjective problem:

$$\begin{aligned} \min f_1 &= \frac{1}{4}\left(x_1-4\right)^2+\frac{1}{4}x_2^2 \\ \min f_2 &= 2-x_2 \\ 2x_1+x_2 &\leq 4 \\ x_1, x_2 &\geq 0 \end{aligned}$$

Draw the feasible region in plane (x_1, x_2) and the impact region in plane (f_1, f_2).

Determine the Paretian region with the ϵ-constraint method.

Solution

The problem can be solved graphically

The Paretian region X^* is the segment of line $2x_1+x_2=4$ included between points $A=(2,0)$ and $B=(0,4)$. Its image F^* is the arc of parabola ... included between points $A'=(1,2)$ and $B'=(8,-2)$.

Exercise 2

Consider the following biobjective problem:

$$\begin{aligned} \max f_1 = x_1 - x_2 \\ \max f_2 = x_2 \\ x_1 + x_2 \leq 3 \\ 0 \leq x_1 \leq 2 \\ x_2 \geq 0 \end{aligned}$$

Draw the feasible region in plane (x_1, x_2) and the impact region in plane (f_1, f_2).

Determine the Paretian region with the ϵ-constraint method.

Solution

The Paretian region X^* is the poly-line (spezzata poligonale) ABC, with $A=(2,0)$, $B=(2,1)$ and $C=(0,3)$. Its image F^* is the poly-line (spezzata poligonale) $A'B'C'$ with $A'=(2,0)$, $B'=(1,1)$ and $C'=(-3,3)$.

[*] The solutions of these exercises have not yet been revised: error reports are welcome.

[†] I owe several of these exercises to exam texts of professor Alberto Colorni.

Exercise 3

Consider the following biobjective problem:

$$\begin{aligned}&\min f_1 = x'Ax + b'x\\&\min f_2 = c'x\\&x \in X = \left\{x \in \mathbb{R}^2 : x_1 \geq 0, x_2 \geq 0\right\}\end{aligned}$$

con

$$A = \begin{bmatrix} 1 & 0 \\ 0 & 4 \end{bmatrix} \quad b = \begin{bmatrix} -4 \\ -8 \end{bmatrix} \quad c = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

Draw the feasible region in plane (x_1, x_2) and the impact region in plane (f_1, f_2).

Determine the Paretian region with the ϵ-constraint method.

Solution

The Paretian region X^* is the poly-line (spezzata poligonale) ABC, with $A = (0, 0)$, $B = (0, 1/2)$ and $C = (2, 1)$.

Exercise 4

Consider the following problem:

$$\begin{aligned}\max f_1 &= x_1 + 3x_2\\\max f_2 &= -3x_1 - 2x_2\\2x_1 + x_2 &\leq 32\\x_1 + x_2 &\leq 20\\x_1 + 5x_2 &\leq 72\\x_1, x_2 &\geq 0\end{aligned}$$

Determine the Paretian region with the weighted-sum method and with the ϵ-constraint method.

Draw the Paretian region in the decision variable space and its image in the impact space.

Solution

The Paretian region X^* is the poly-line (spezzata poligonale) ABC, with $A = (0, 0)$, $B = (0, 72/5)$ and $C = (7, 13)$. Its image F^* is the poly-line (spezzata poligonale) $A'B'C'$ with $A' = (0, 0)$, $B' = (216/5, -144/5)$ and $C' = (46, -47)$.

Exercise 5

Determine the Paretian region of the following problem:

$$\begin{aligned}\max f_1 &= -x_1 - x_2\\\max f_2 &= x_1\\3x_1^2 + 4x_2 &\leq 12\\x_2 &\geq 0\end{aligned}$$

Draw the Paretian region in the decision variable space and its image in the impact space.

Solution[*]

The Paretian region X^* is the line segment AB, with $A = (-2, 0)$ and $B = (0, 2)$. Its image F^* is the line segment $A'B'$ with $A' = (?, ?)$ and $B' = (?, ?)$.

Exercise 6

A fire station must be located as close as possible to a dangerous point in which the cartesian axes are centred. A river flows approximately along a straight line from east to west three kilometres north of that point. The river could be used as a source of water for the station, building a dedicated aqueduct, which should be as short as possible in order to limit its cost. For the sake of a quick intervention, the station cannot be located north of the river, nor south-west of a large road represented by line $2x_1 + x_2 = 4$.

Determine the Paretian region in the space of the decision variables and its image in the space of the indicators.

Solution

In synthesis, the problem imposes two constraints on the location of the service and two choice criteria: the closeness to the dangerous points and the closeness to the river. The Paretian region X^* is segment AB, with $A = (1/2, 3)$ and $B = (8/5, 4/5)$.

Exercise 7

Consider the following biobjective problem:

$$\begin{aligned} \max f_1 = 9x_1^2 + 4x_2^2 - 18x_1 - 16x_2 \\ \max f_2 = -x_1 \\ 3x_1 + x_2 \leq 6 \\ 3x_1 + 2x_2 \leq 9 \\ x_2 \geq 0 \end{aligned}$$

Draw the Paretian region in the decision variable space and its image in the indicator space.

Solution

The Paretian region X^* is the poly-line (spezzata poligonale) ABC, with $A = (1, 2)$, $B = (4/3, 2)$ and $B = (2, 0)$. Its image F^* is the curve line (spezzata curva) $A'B'C'$ with $A' = (-25, -1)$, $B' = (-24, -4/3)$ and $C' = (0, -2)$.

Exercise 8

The following table represents the performance of five alternatives with respect to four indicators (all of them to be maximised), in una scala di valori tra 0 e 100.

[*] This problem is equivalent to the one solved in Example 37 with the inverse transformation method.

	a_1	a_2	a_3	a_4	a_5
f_1	100	70	60	40	20
f_2	60	45	40	100	80
f_3	60	25	20	80	100
f_4	20	100	90	50	40

Identify the dominated alternatives (if any exists) and specify by which other alternatives they are dominated.

Solution

Alternative a_3 is dominated by alternative a_2.

Exercise 9

Consider the following biobjective problem:

$$\begin{aligned} \max f_1(x) &= -x_1 + 2x_2 \\ \max f_2(x) &= 2x_1 - x_2 \\ x_1 &\leq 4 \\ x_2 &\leq 4 \\ x_1 + x_2 &\leq 7 \\ -x_1 + x_2 &\leq 3 \\ x_1 - x_2 &\leq 3 \\ x_1 &\geq 0 \\ x_2 &\geq 0 \end{aligned}$$

Draw the Paretian region in the decision variable space and its image in the indicator space.

Solution

The Paretian region X^* is the polyline (spezzata poligonale) $FEDC$, with $C = (1, 4)$, $D = (3, 4)$, $E = (4, 3)$ and $F = (4, 1)$.

Exercise 10

Consider the following biobjective problem:

$$\begin{aligned} \max f_1(x) &= x_1 - 3x_2 \\ \max f_2(x) &= -4x_1 + x_2 \\ -2x_1 + 2x_2 &\leq 7 \\ 2x_1 + 2x_2 &\leq 11 \\ x_1 &\leq 4 \\ x_1, x_2 &\geq 0 \end{aligned}$$

Draw the Paretian region in the decision variable space and its image in the indicator space.

Solution

The Paretian region X^* is the poly-line (spezzata poligonale) AOB, with $A = (0, 7/2)$, $O = (0, 0)$ and $B = (4, 0)$.

Exercise 11

Consider the following biobjective problem:

$$\begin{aligned} \min f_1(x) &= x_1^2 + x_2^2 \\ \max f_2(x) &= x_2 \\ x_2 &\leq 10 \end{aligned}$$

Draw the Paretian region in the decision variable space and its image in the indicator space.

Solution

The Paretian region X^* is segment OA, with $O = (0, 0)$ and $A = (0, 10)$.

Exercise 12

A club is looking for a new site: there are four alternatives (A, B, C and D), besides alternative 0 (staying in the present site). The choice among the five alternatives is definitive and will be based on three factors: cost, accessibility and prestige. The following table indicates the benefits associated to each alternative and factor on a scale between 0 and 100.

Indicatori	A	B	C	D	0
Cost	90	90	90	1	100
Accessibility	12	13	10	100	37
Prestige	30	1	5	100	10

Identify the dominated alternatives and state by which other alternatives they are dominated.

Solution

Alternative C is dominated by alternative A, while alternative B is dominated by alternative 0.

Exercise 13

Consider the following biobjective problem:

$$\begin{aligned} \min f_1(x) &= x^2 - 4x \\ \min f_2(x) &= -x^2 \\ 0 \leq x &\leq 3 \end{aligned}$$

Draw the Paretian region in the decision variable space and its image in the indicator space.

Solution

The Paretian region X^* is the interval $x \in [2;3]$. Its image F^* is the arc of the parabola ... included between points $A' = (-4,-4)$ and $B' = (-3,-9)$.

Exercise 14

Consider the following biobjective problem:

$$\begin{aligned}
&\min f_1(x) = x_1^2 + 4x_2^2 - 2x_1 - 16x_2 \\
&\min f_2(x) = -5x_1 - x_2 \\
&\quad x_1^2 + x_2^2 \leq 8 \\
&\qquad\quad x_2 \geq 0
\end{aligned}$$

Give an analytical description of the feasible cone $\mathcal{D}_f$ and of the improving cone $\mathcal{D}_i$ in point $P = (2,2)$. Based on the two cones, can point P be Paretian? Why?

Solution

The feasible cone and the improving cone are:

$$\mathcal{D}_f = \left\{d \in \mathbb{R}^2 : d_1 + d_2 \leq 0\right\} \qquad \mathcal{D}_i = \left\{d \in \mathbb{R}^2 : d_1 \leq 0, 5d_1 + d_2 geq 0\right\}$$

The two cones have no common direction. Then, P satisfies the necessary KKT-conditions for local paretianity: it can be Paretian, even if it is not guaranteed to be.

.

Exercise 16

Given the following decision problem with three alternatives and four attributes (benefits):

Attributi	A	B	C
u_1	0	100	80
u_2	100	83	0
u_3	70	20	100
u_4	40	100	20

Identify the dominated alternatives and state by which other alternatives they are dominated.

Solution

There are no dominated alternatives.

Exercise 17

Consider the following biobjective problem:

$$\begin{aligned}
&\min f_1(x) = x_1^2 + x^2 - 2x_1 \\
&\min f_2(x) = -x_2 \\
&x_1^2 + 4x^2 \leq 8 \\
&x_1 - 2x_2 \geq 0
\end{aligned}$$

Draw the Paretian region in the decision variable space and its image in the indicator space.

Solution

The Paretian region X^* is the poly-line (spezzata poligonale) ABC, with $A = (1, 0)$, $B = (1, 1/2)$ and $C = (2, 1)$. Its image F^* is the curve line (spezzata curva) $A'B'C'$, with $A' = (-1, 0)$, $B' = (-3/4, -1/2)$ and $C' = (1, -1)$.

Chapter 7

Weak rationality methods

7.1 Partially inconsistent decision-makers

Applying the machinery of Multiple Attribute Utility Theory to the additive case, if each passage is performed with due precision, the result consists of p normalized utility functions $\tilde{u}_l$ and p weights w_l that tune their contribution to the overall utility function. The marginal rates of substitution between the normalized utility components are uniform and equal to the ratios of the weights ($\tilde{\lambda}_{lm} = w_l/w_m$), and the indifference curves in the utility space form families of parallel hyperplanes.

In pratice, the problem is that the decision-maker is usually unable to estimate correctly the rates of substition as required by the theory, because he/she considers indifferent also pairs of impacts between which there is on the contrary a slight preference. The result is that the computation of the weights w_l based on $p-1$ pairs of indifferent impacts is unreliable.

Example 53 *A decision-maker has been asked to indicate the rates of substitution between three normalized utility functions $\tilde{u}_1$, $\tilde{u}_2$ and $\tilde{u}_3$. The answers of the decision-maker generate the following matrix.*

$$\tilde{\Lambda} = \begin{array}{c|ccc|} & \tilde{u}_1 & \tilde{u}_2 & \tilde{u}_3 \\ \cline{2-4} \tilde{u}_1 & \mathit{1} & \mathit{2} & \mathit{4} \\ \tilde{u}_2 & \mathit{1/2} & \mathit{1} & \mathit{3} \\ \tilde{u}_3 & \mathit{1/4} & \mathit{1/3} & \mathit{1} \\ \cline{2-4} \end{array}$$

The matrix is positive and reciprocal, but not consistent, since $\tilde{\lambda}_{31} \neq \tilde{\lambda}_{32}\tilde{\lambda}_{21}$. Each column suggests a different weight vector

$$w^{(1)} = \begin{bmatrix} 0.571 \\ 0.286 \\ 0.143 \end{bmatrix} \qquad w^{(2)} = \begin{bmatrix} 0.6 \\ 0.3 \\ 0.1 \end{bmatrix} \qquad w^{(3)} = \begin{bmatrix} 0.5 \\ 0.375 \\ 0.125 \end{bmatrix}$$

The difference between the three vectors is relevant, and it could lead to choose different solutions.

In an inconsistent matrix, each spanning tree produces its own weight estimate, in general different from the one base on the other trees.

7.1.1 Reconstructing consistent matrices

If the attributes are numerous, in practice no human decision-maker is able to keep perfectly consistent. The weights derived from $p-1$ comparisons are therefore unreliable. One can, howerver, build the whole pairwise comparison matrix, evaluate the amount of the inconsistency and build a more reliable weight vector. Such a vector w allows, therefore, to reconstruct a consistent comparison matrix. If one wants to use the weight vector w for a decision, it is necessary that the comparison matrix generated from it be as close as possible to the measured one. In other words, one wants to solve the following minimisation problem:

$$\begin{aligned} \min_w \left\| W - \tilde{\Lambda} \right\| & \\ W_{lm} &= \frac{w_l}{w_m} \\ \sum_{l \in P} w_l &= 1 \\ w_l &\geq 0 \qquad l \in P \end{aligned}$$

where the unknown variables are the weights w_l, W denotes the matrix composed by the ratios w_l/w_m, $\tilde{\Lambda}$ the matrix denoted by the ratios estimated by the decision-maker and $\|W - \tilde{\Lambda}\|$ is the norm of matrix $W - \tilde{\Lambda}$.

Definition 28 *We denote as* norm *any function associating to a matrix a real value such that:*

- *it is nonnegative for any matrix:*

$$\|M\| \geq 0 \textit{ for each } M$$

- *it is zero if and only if the matrix is null:*

$$\|M\| = 0 \Leftrightarrow M = 0$$

- *if all elements of the matrix are multiplied by a constant factor, the norm is multiplied by the absolute value of such a constant:*

$$\|\alpha M\| = |\alpha| \; \|M\|$$

- *satisfies the* triangle inequality*:*

$$\|M_1 + M_2\| \leq \|M_1\| + \|M_2\|$$

There are infinitely many different definitions of norm that respect these properties; each one provides a different solution to the problem. The most common ones in the literature are:

$$\left\| W - \tilde{\Lambda} \right\|_1 = \sum_{l \in P} \sum_{m \in P} \left| \frac{w_l}{w_m} - \tilde{\lambda}_{lm} \right|$$

$$\left\| W - \tilde{\Lambda} \right\|_2 = \sqrt{\sum_{l \in P} \sum_{m \in P} \left| \frac{w_l}{w_m} - \tilde{\lambda}_{lm} \right|^2}$$

$$\left\| W - \tilde{\Lambda} \right\|_\infty = \max_{l \in P, m \in P} \left| \frac{w_l}{w_m} - \tilde{\lambda}_{lm} \right|$$

In particular, norm $\|W - \tilde{\Lambda}\|_2$ is the Euclidean norm, which gives rise to the classical *least-squares method.*

The basic idea of all these methods is to start from the information provided by the decision-maker and to modify it as little as possible in order to reach consistency. *The optimal value of the objective function can be assumed as a measure of the initial inconsistency.* In fact, it is zero if and only if the estimated matrix $\tilde{\Lambda}$ is consistent.

In order to keep into account the uncertainty of the decision-maker, it has also been proposed to replace the estimated values $\tilde{\lambda}_{lm}$ with intervals and to define the norm by combining the distances between each ratio w_l/w_m and the corresponding interval, instead of the corresponding value.

The eigenvalue method

A method frequently used in the literature was proposed by Saaty[1], the founder of the Analytic Hierarchy Process (see Section 7.2). Is is based on the eigenvalues of the estimated matrix $\tilde{\Lambda}$. Since this is positive, Perron-Frobenius' theorem guarantees that the dominant eigenvalue $\mu^{\max}$ (that is, the eigenvalue whose absolute value is maximum) is real and positive. If $\tilde{\Lambda}$ is also reciprocal (an easy condition to impose in practice), $\mu^{\max} \geq p$ and $\mu^{\max} = p$ if and only if the matrix is consistent.

The inconsistency of matrix $\tilde{\Lambda}$ can therefore be measured by

$$\gamma = \frac{\mu^{\max} - p}{p - 1}$$

that is the opposite of the average value of the other $p - 1$ eigenvalues, since their sum is equal to the trace, that is to p.

If the matrix is weakly inconsistent, $\mu^{\max} \approx p$ and the other eigenvalues are close to zero. The eigenvector $x^{\max}$ associated with $\mu^{\max}$, normalized, provides a vector of weights with which a consistent matrix can be built. Saaty also suggested to consider as acceptable an inconsistency not larger than 10% of the average inconsistency obtained from random positive and reciprocal matrices of the required size p. Such average inconsistency is a function of p whose empirical values have been tabulated.

This method has been severely criticised in recent times, because it has been proved that it can produce *dominated* matrices: it sometimes computes a matrix whose elements are farther away (one-by-one) from the estimated matrix than an alternative consistent matrix obtained with alternative methods. Such an alternative matrix is strictly better than the dominant eigenvector matrix with respect to any definition of norm.

Example 54 *Consider the pairwise comparison matrix:*

$$\tilde{\Lambda} = \begin{array}{c|ccc|} & u_1 & u_2 & u_3 \\ \hline u_1 & \mathit{1} & \mathit{2} & \mathit{4} \\ u_2 & \mathit{1/2} & \mathit{1} & \mathit{3} \\ u_3 & \mathit{1/4} & \mathit{1/3} & \mathit{1} \\ \hline \end{array}$$

whose eigenvalues are $\mu_1 = 3.108$*,* $\mu_2 = -0,0091 + 0,2348i$ *and* $\mu_3 = -0,0091 - 0,2348i$*. The inconsistency is* $\gamma = 0.054$*, rather limited, and the consistent matrix suggested by the method is:*

$$\tilde{\Lambda} = \begin{array}{c|ccc|} & \tilde{u}_1 & \tilde{u}_2 & \tilde{u}_3 \\ \hline \tilde{u}_1 & \mathit{1} & \mathit{1.747} & \mathit{4.579} \\ \tilde{u}_2 & \mathit{0.572} & \mathit{1} & \mathit{2.621} \\ \tilde{u}_3 & \mathit{0.218} & \mathit{0.381} & \mathit{1} \\ \hline \end{array}$$

[1] Thomas. L. Saaty (1926-2017), Iranian mathematician working in the United States.

Example 55 *Consider the pairwise comparison matrix:*

$$\tilde{\Lambda} = \begin{array}{c|ccc|} & \tilde{u}_1 & \tilde{u}_2 & \tilde{u}_3 \\ \hline \tilde{u}_1 & 1 & 2 & 1/2 \\ \tilde{u}_2 & 1/2 & 1 & 3 \\ \tilde{u}_3 & 2 & 1/3 & 1 \\ \hline \end{array}$$

whose eigenvalues are $\mu_1 = 3.7262$*,* $\mu_2 = -0.3631 + 1.6044i$ *and* $\mu_3 = -0.3631 - 1.6044i$*.*

The inconsistency is $\gamma = 0.3631$*, quite strong. Evidently, the first attribute is more important than the second, and the second is more important than the third, but the third is more important than the first one: not only consistency, but even transitivity is violated.*

The consistent matrix suggested by the eigenvector method is:

$$\tilde{\Lambda} = \begin{array}{c|ccc|} & u_1 & u_2 & u_3 \\ \hline u_1 & 1 & 0.874 & 1.145 \\ u_2 & 1.145 & 1 & 1.310 \\ u_3 & 0.874 & 0.763 & 1 \\ \hline \end{array}$$

but in this case it is doubtful whether it is a meaningful matrix.

7.1.2 Consistency at any cost?

The replacement of the given pairwise comparison matrix $\tilde{\Lambda}$ with a consistent matrix W in general distorts the information provided by the decision-maker. Part of the literature reports that the new matrix W can be more appreciated by the decision-maker than the original one, as if the operation were able to help him/her reach a hard to attain, but desired, consistency. Part of the literature, on the contrary, reports that the new matrix W can be less appreciated, as if the operation contradicted the preferences of the decision-maker under some important aspects.

So far, the only established conclusion is that it is better to perform this operation in an interactive way, instead of as a black-box. Possibly, it is also better to proceed by subsequent iterations, submitting the results to the decision-maker for a confirmation, and asking to modify himself/herself the indications provided in order to approach consistency.

7.2 Analytic Hierarchy Process

The *Analytic Hierarchy Process* (*AHP*) was introduce in 1980, based on the following criticisms made by Saaty to the classical methods to build utility functions:

1. the reconstruction of the single-variable normalized utility functions is subject to strong approximation errors;
2. the estimation of the weights is subject to strong approximation errors when the number of attributes p is large;
3. the various approximation errors combine in cascade.

The approximations in the normalized utility functions are particularly strong in the case of qualitative attributes, for which it is easy to sort the impact, but much harder to measure the relative strength of the preference between different impacts If the normalized utility functions are not correct, however, the subsequent operations of weight estimation and linear combination in order to evaluate the overall utility become meaningless. In fact, any Paretian alternative of a problem can be the winner if a suitable utility function is chosen. These practical difficulties risk to reduce the classical methods to pure "number-crunching", that is manipulation of numbers without any practical meaning and out of the control of the decision-maker.

The *AHP* aims to define in a simpler and more natural way the measure of preference between alternative impacts. Its characteristic elements are:

1. the evaluation of solution utility with respect to each attribute through pairwise comparisons, instead of direct absolute measures;
2. the use of qualitative scales, instead of quantitative ones;
3. the evaluation of the attribute weights through pairwise comparisons;
4. the hierarchical structuring of attributes;
5. the multiplicative recombination of the weights at different levels of the hierarchy.

Even if it does not properly make sense to speak of preferential independence and of the corresponding trade-off condition, the operations performed by the *AHP* implicitly assume that each indicator contribute in an additive way to the overall utility of each solution. In fact, the fundamental concepts of the *AHP* can be interpreted as an approximated and qualitative version (but a more robust one with respect to modelling errors) of the fundamental concepts of the classical theory.

7.2.1 Computation of the utilities with pairwise comparisons

Given an indicator f_l, the classical method tries to reconstruct the whole absolute profile of utility as a function $\tilde{u}_l(f_l)$. The criticism by Saaty is based on cognitive psychology results according to which it is very difficult for a human decision-maker to associate quantitative utility values to the values of the indicator. The *AHP*, therefore, focuses on pair of solutions and on the strength of the relative preference betewen the values of an indicator in the two solutions of the pair. A matrix $\Lambda_l = \{\lambda_{xy}^{(l)}\}$ is built, whose element $\lambda_{xy}^{(l)}$ is associated to a pair of alternatives $(x, y) \in X \times X$ and evaluates how much $f_l(x)$ is preferable to $f_l(y)$. Therefore, the process does not reconstruct the whole normalized utility function with respect to an indicator, but evaluates the strength ratio between given pairs of alternative values, that is something similar to a utility ratio $\tilde{u}_l(f_l(x)) / \tilde{u}_l(f_l(y))$.

Since it requires the explicit enumeration of all solution pairs, the *AHP can be applied only to finite problems* and with a small number of solutions, or to problems whose solution set has been preliminarly pruned reducing it to a small finite set.

7.2.2 Qualitative scales

The second basic idea of the *AHP* is to simplify the evaluation process by measuring the preference between two values $f_l(x)$ and $f_l(y)$ not with a quantitative scale, but a qualitative one, denoted as *Saaty's scale*:

1. $\lambda_{xy}^{(l)} = 1$: $f_l(x)$ e $f_l(y)$ are *indifferent*;
2. $\lambda_{xy}^{(l)} = 3$: $f_l(x)$ is *moderately* preferable to $f_l(y)$;

3. $\lambda_{xy}^{(l)} = 5$: $f_l(x)$ is *strongly* preferable to $f_l(y)$;

4. $\lambda_{xy}^{(l)} = 7$: $f_l(x)$ is *very strongly* preferable to $f_l(y)$;

5. $\lambda_{xy}^{(l)} = 9$: $f_l(x)$ is *absolutely* preferable to $f_l(y)$.

The weights 2, 4, 6 and 8 are used for intermediate evaluations. Symbol λ is used by analogy with the marginal rate of substitution, but it is clearly a much more empirical and approximated measure. The values of Saaty's scale are absolutely arbitrary, but derive from psychology studies, according to which it is inappropriate to ask a decision-maker to indicate exact quantitative values for the preference ratios, because the human beings are unable to discriminate more than $5 - 9$ levels in a sensible and consistent way.

Moreover, the use of a qualitative scale allows to compare heterogeneous quantities, or quantities which are not expressed in a quantitative way, translating verbal judgments into numerical values. The idea is to give for granted that the numerical values used are only a rough approximation of reality; knowing it, automatically raises an alert against trusting them too much, a risk that is easily taken with the normalized utilities estimated by the classical methods. Moreover, building a whole pairwise comparison matrix also allows to evaluate its inconsistency and to tune it *a posteriori* in order to make it consistent, instead of forcing a fictitious consistency since the beginning, which could introduce unamendable errors in the following phases.

The methods described in Section 7.1.1 allow to derive from the pairwise comparison matrix a consistent matrix, and from the latter a vector u_l that associates to each value considered for the indicator, that is to each feasible solution $x \in X$, a positive and normalized numerical value u_{xl}.

Definition 29 *We denote as* evaluation matrix *a matrix $U = \{u_{xl}\}$ containing the evaluation u_{xl} of each alternative $x \in X$ with respect to each indicator $l \in P$, obtained starting from the pairwise comparison matrix Λ_l.*

The values u_{xl} obtained in this way replace the values of the normalized utility functions $\tilde{u}_l(f_l(x))$ used in the classical theory. These values also fall in $[0, 1]$ and are larger for the preferred alternatives, but in general the worst and the best value of an indicator will not correspond to the extreme values 0 e 1.

7.2.3 Computation of the weights with pairwise comparisons

Also the attribute weights are built starting from pairwise comparison matrices filled with qualitative values drawn from Saaty's scale. The comparison between two attributes aims to evaluate the relative weight of the former with respect to the latter. With the methods described in Section 7.1.1 one can derive from the pairwise comparison matrix of the attributes a vector of normalized weights w_l, associated to the attributes of the problem. This vector, therefore, does not derive from the identification of pairs of indifferent impacts, but from the processing of judgments (provided by the decision-maker) on the relative weights of the attributes. The practical meaning of these weights, however, is the same: they are used to combine in an additive way the evaluations u_{xl}, so as to obtain the overall evaluation of $u_x = \sum_{l=1}^{p} u_{xl}$ for each alternative $x \in X$, based on which the choice will be made.

7.2.4 Hierarchical structuring of the attributes

Building a pairwise comparison matrix between all attributes of a problem can be impractical if the number of attributes is large. As anticipated in Chapter 2, however, one can collect the attributes of the problem into homogeneous categories, structured in a hierarchy:

1. the level of the leaves includes the elementary attributes;

2. at the upper levels summarising attributes appear, progressively more general;

3. the root corresponds to a sort of general objective.

Each attribute of the problem, in other words, is summarised into a more general objective and decomposed into more specialised objectives. For example, in an Environmental Impact Assessment (*EIA*) the general objective of satisfying the decision-maker is often decomposed into macro-objectives guaranteeing a positive impact in the environmental, economic and social sectors, respectively. The environmental impact, on its turn, can be decomposed into sectors concerning water, ground, air and noise; the sector "air" into the different kinds of pollutants, the latter into the different geographical areas, and so on. Figure 7.1 provides a (simplified) example concerning the tram-train in the town of Como, where environment, economy and society were completed by an additional sector conerning traffic.

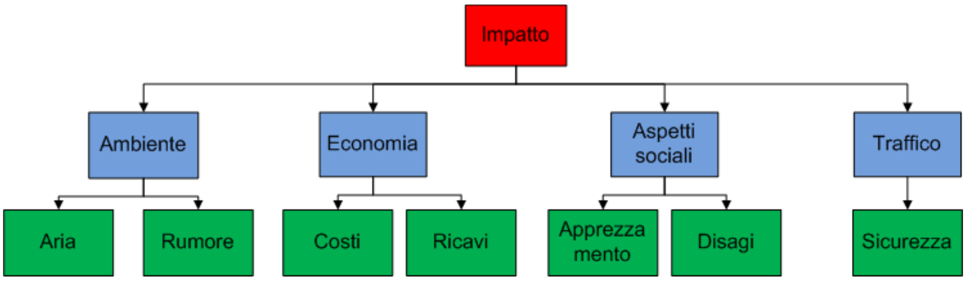


Figure 7.1: The hierarchical structure of the *EIA* problem for the tram-train of Como (strongly simplified version)

The advantages of structuring the attributes into a hierarchy are:

1. only homogeneous attributes are compared with each other (for example, the pollution by CO with the pollution by NO_x, and not with the increase in employment or the financial costs);

2. each subset of pairwise comparisons can be assigned to an *expert subdecision-maker* specialised in a specific field, so as to obtain more meaningful indications;

3. the total number of pairwise comparisons strongly decreases.

The fundamental disadvantage is that the pairwise comparisons at the level of the leaves consider concrete and well-defined attributes, whereas those at the upper levels consider abstract and general objectives, for which a quantitative measure does not even make sense. For example, at the level of the leaves the weight of the pollution by CO is compared with that of the pollution by NO_x, whereas at the upper levels the environmental impact of a project if compared with its social impact. The latter comparison, even though it is expressed with a numerical scale, is obviously qualitative. Typically, the upper levels are assigned to the political decision-makers, as experts of trade-offs between general themes.

7.2.5 Hierarchical recomposition

The use of a hierarchical structure slightly complicates the computation of the weights of the attributes. In fact, only the attributes that are children of the same node in the tree are compared with each other. This gives rise to weights that are normalized within eah group of children nodes, but are not comparable with those of the other groups. The tree structure, however, allows to progressively build the attribute weight vector scanning the tree level-by-level from the leaves up to the root.

At each level, one builds a pairwise comparison matrix between the children of the same node, and derives from it a vector of weights with the methods described above. The weights in each vector have sum equal to one, and describe the relative weight between each other. In order to compare such weights with those of the other nodes of the whole tree, it is enough to renormalize them so that their sum coincides with the weight of the father node, that is simply to multiply them by the weight of the father node. If the tree has several levels, it is enough to multiply the weight of each leaf attribute by those of all the nodes visited along the path that leads to the root of the tree. In this way, the sum of the weights of all nodes on a level is always equal to one; in particular, the sum of the weights w_l of the leaves is equal to one.

Once the weights w_l have been reconstructed, th evaluation matrix u_{xl} provides a value for each alternative and each elementary attribute. As in the classical theory, the overall utility $u(x)$ of each alternative x is obtained with a convex combination of the utilities with the weights. This corresponds to multiplying the evaluation matrix U by the weight vector w:

$$u_x = \sum_{l=1}^{p} u_{xl} w_l \quad \Leftrightarrow \quad u = U \cdot w$$

The utility values thus obtained are used to sort the alternatives and solve the decision problem.

Example 56 *Let us consider the purchase of a house. Let us suppose that there are three alternatives, named a_1, a_2 and a_3, and that the choice depends on four attributes (price, size, life quality in the area, purchase conditions). The hierarchy of the problem, therefore, includes a root, corresponding to the global objective, from four leaves are appended (the attributes).*

With respect to each of the four attributes, the pairwise comparison phase between the alternatives leads to the following matrices, and therefore to the following weight vectors (since there are few alternatives, we suppose to obtain consisten matrices, independent from the method used to build the utilities):

1. *price*

$$\Lambda_p = \begin{bmatrix} 1 & 1/9 & 1/8 \\ 9 & 1 & 9/8 \\ 8 & 8/9 & 1 \end{bmatrix} \Rightarrow u_p = \begin{bmatrix} 1/18 \\ 9/18 \\ 8/18 \end{bmatrix}$$

2. *size*

$$\Lambda_d = \begin{bmatrix} 1 & 9 & 9 \\ 1/9 & 1 & 1 \\ 1/9 & 1 & 1 \end{bmatrix} \Rightarrow u_d = \begin{bmatrix} 9/11 \\ 1/11 \\ 1/11 \end{bmatrix}$$

3. life quality in the area

$$\Lambda_q = \begin{bmatrix} 1 & 1/9 & 1/4 \\ 9 & 1 & 9/4 \\ 4 & 4/9 & 1 \end{bmatrix} \Rightarrow u_q = \begin{bmatrix} 1/14 \\ 9/14 \\ 4/14 \end{bmatrix}$$

4. purchase condition

$$\Lambda_c = \begin{bmatrix} 1 & 3 & 3/5 \\ 1/3 & 1 & 1/5 \\ 5/3 & 5 & 1 \end{bmatrix} \Rightarrow u_c = \begin{bmatrix} 3/9 \\ 1/9 \\ 5/9 \end{bmatrix}$$

Let us assume that the four attributes have the same weight according to the decision-maker.

$$\Lambda_w = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix} \Rightarrow w = \begin{bmatrix} 1/4 \\ 1/4 \\ 1/4 \\ 1/4 \end{bmatrix}$$

Since the tree has a single level, it is not necessary to recombine the weights. It is enough to reconstruct the evaluation matrix, using as columns the vectors which express the utility of the alternatives with respect to the attributes:

$$U = [u_p \,|\, u_d \,|\, u_q \,|\, u_c]$$

and performing the convex combination of the utilities with the weights, multiplying the evaluation matrix by the weight vector.

$$u = Uw = \begin{bmatrix} 1/18 & 9/11 & 1/14 & 3/9 \\ 9/18 & 1/11 & 9/14 & 1/9 \\ 8/18 & 1/11 & 4/14 & 5/9 \end{bmatrix} \cdot \begin{bmatrix} 1/4 \\ 1/4 \\ 1/4 \\ 1/4 \end{bmatrix} = \begin{bmatrix} 0.320 \\ 0.336 \\ 0.344 \end{bmatrix}$$

which leads to choose alternative a_3, even though the other two are not far away. If the coefficients assigned during the pairwise comparison were not very reliable, this suggests the opportunity to perform a sensitivity analysis to evaluate possible variations, as in Example 49 of Section 6.6.3.

If the tree has several levels, it is also completely equivalent to evaluate the alternatives in each node of the tree, that is with respect to the corresponding attribute: it is enough to go up the tree terminating the recombination in the node that corresponds to the attribute considered. For instance, if one wants to know the environmental utility of the alternatives, it is enough to go up the tree until the node Environment, and to renormalize the weights only for the environmental attributes, multiplying them only by the weights of the nodes visited along the path from the leaves to such a node, and finally to perform the convex combination of the environmental utilities with such weights.

7.2.6 Hybridisations with the utility theory

The hierarchical structuring of the attributes is an innovation introduced by the *AHP*, that has later been adapated to the body of the classical methods. It can in fact happen that the estimated normalized utility functions are reliable, but tha the indicators are too many to determine their weights by comparing them in a meaningful way. In this case, it is possible to accept the normalised utility values, but to estimate the weights with pairwise comparisons, qualitative scales and multiplicative recombination, as proposed by the Analytic Hierarchy Process.

Example 57 *Let us consider a very simplified problem concerning the case study of the tram-train in Como, introduced in Chapter 2. For the sake of simplicity, we consider a limited number of alternatives and attributes, but we keep a multilevel attribute tree. The chosen alternatives are:*

0. *Alternative zero: no intervention;*
1. *Railway shuttle;*
2. *Tram-train (having chosen one of the paths);*
3. *Ring tramway.*

			Alternatives			
Submatrix	Sector	Subsector	0	1	2	3
U_a^T	Environment	Air	0.0	0.1	0.6	0.8
		Noise	0.4	0.2	0.4	0.6
		Vibrations	0.4	0.2	0.4	0.6
		Landscape	1.0	1.0	0.8	0.5
		Territ. struct.	0.0	0.3	1.0	0.6
U_e^T	Economy	Costs	1.0	0.9	0.6	0.0
		Revenues	0.0	0.3	0.7	1.0
		Housing values	0.0	0.5	1.0	1.0
U_s^T	Society	Appreciation	0.0	0.2	0.7	0.9
		Discomfort	1.0	1.0	0.7	0.2
		Accessibility	0.0	0.1	0.7	0.9
		Employment	0.0	0.5	1.0	1.0
		Induced effects	0.0	0.3	0.6	1.0
U_t^T	Transportation	Security	0.0	0.2	0.6	1.0
		Congestion	0.0	0.4	0.9	1.0
		Interferences	1.0	0.8	0.6	0.2

Table 7.1: Evaluation matrix (transposed) for the (simplified) problem of the tram-train of Como

Once the impacts have been estimated for each alternative and the normalized utility functions have been built, we suppose to obtain the evaluation matrix U reported in Table 7.1. The matrix has been transposed for ease of representation, and it has been hierarchically divided into submatrices. Each column of the matrix (row in the table) provides the normalized utility of an indicator, with values included between 0 *(for the worst performance) and* 1 *(for the best one).*

We can therefore proceed building the weights of the attributes with pairwise comparisons and qualitative evaluations. In a first phase, four specialised decision-makers evaluate the relative importance of the subsectors collected in each sector. Let us assume that the pairwise comparison matrices obtained for each sector are the ones represented in Tables 7.2, 7.3, 7.4 and **??***.*

From each matrix, we derive the corresponding subsector weight vector. As the decision-makers are not completely consistent, each matrix must be made consisten, before deriving the weight vector. This has been done with the least-square method. The four resulting vectors are reported in Table 7.6 and added to the overall tree structure in Table 7.7

In a second phase, a political decision-maker evaluates the relative importance of the sectors, once again by pairwise comparisons. Table 7.8 reports the comparisons.

Λ_a	Air	Noise	Vibrations	Landscape	Territ. struct.
Air	1	1	1	5/4	1
Noise	1	1	1	3/4	1
Vibrations	1	1	1	3/4	3/4
Landscape	4/5	4/3	4/3	1	1
Territ. struct.	1	1	4/3	1	1

Table 7.2: Pairwise comparison matrix for the subsectors of sector Environment

Λ_e	Costs	Revenues	Housing values
Costs	1	5/4	5/4
Revenues	4/5	1	1
Housing values	4/5	1	1

Table 7.3: Pairwise comparison matrix for the subsectors of sector Economy

Λ_s	Apprec.	Discomfort	Access.	Empl.	Ind. eff.
Appreciation	1	1	1	5/4	1
Discomfort	1	1	5/4	3/4	1
Accessibility	1	4/5	1	3/2	1
Employment	4/5	4/3	2/3	1	3/4
Induced effects	1	1	1	4/3	1

Table 7.4: Pairwise comparison matrix for the subsectors of sector Society

Λ_t	Security	Congestion	Interferences
Security	1	3/4	1
Congestion	4/3	1	5/4
Interferences	1	4/5	1

Table 7.5: Pairwise comparison matrix for the subsectors of sector Transportation

Environment		Economy		Society		Transportation	
Subsect.	w_a	Subsect.	w_e	Subsect.	w_s	Subsect.	w_t
Air	0.209	Costs	0.384	Apprec.	0.209	Secur.	0.301
Noise	0.188	Revenues	0.308	Discomfort	0.197	Congest.	0.392
Vibrations	0.178	Housing values	0.308	Access.	0.207	Interf.	0.307
Landscape	0.214			Empl.	0.176		
Territ. Struct.	0.211			Ind. eff.	0.211		

Table 7.6: Weight vectors for the subsectors of the four sectors

Submatrix	Sector	Subsector		Alternatives			
				0	1	2	3
U_a^T	Environment	Air	0.209	0.0	0.1	0.6	0.8
		Noise	0.188	0.4	0.2	0.4	0.6
		Vibrations	0.178	0.4	0.2	0.4	0.6
		Landscape	0.214	1.0	1.0	0.8	0.5
		Territ. struct.	0.211	0.0	0.3	1.0	0.6
U_e^T	Economy	Costs	0.384	1.0	0.9	0.6	0.0
		Revenues	0.308	0.0	0.3	0.7	1.0
		Housing values	0.308	0.0	0.5	1.0	1.0
U_s^T	Society	Appreciation	0.209	0.0	0.2	0.7	0.9
		Discomfort	0.197	1.0	1.0	0.7	0.2
		Accessibility	0.207	0.0	0.1	0.7	0.9
		Employment	0.176	0.0	0.5	1.0	1.0
		Induced effects	0.211	0.0	0.3	0.6	1.0
U_t^T	Transportation	Security	0.301	0.0	0.2	0.6	1.0
		Congestion	0.392	0.0	0.4	0.9	1.0
		Interferences	0.307	1.0	0.8	0.6	0.2

Table 7.7: Evaluation matrix (transposed) for the (simplified) problem of the tram-train of Como, with the weights computed for the subsectors of the four sectors

	Environment	Economy	Society	Transportation
Environment	1	3/2	3/4	1
Economy	2/3	1	1/2	3/4
Society	4/3	2	1	1
Transportation	1	4/3	1	1

Table 7.8: Pairwise comparison matrix among the four sectors

This matrix yields the weights $w^{(1)}$ *of the sectors at level* 1 *in the tree (see Table 7.9), once again applying the least-square method to solve the inconsistencies of the decision-maker. Table 7.10 reports the overall tree structure with the weights at both levels.*

Sectors	$w^{(1)}$
Environment	0.252
Economy	0.173
Society	0.312
Transportation	0.263

Table 7.9: Weight vector of the four sectors

				Alternatives			
Submatrix	Sector	Subsector		0	1	2	3
		Air	0.209	0.0	0.1	0.6	0.8
		Noise	0.188	0.4	0.2	0.4	0.6
U_a^T	Environment	Vibrations	0.178	0.4	0.2	0.4	0.6
	0.252	Landscape	0.214	1.0	1.0	0.8	0.5
		Territ. struct.	0.211	0.0	0.3	1.0	0.6
		Costs	0.384	1.0	0.9	0.6	0.0
U_e^T	Economy	Revenues	0.308	0.0	0.3	0.7	1.0
	0.173	Housing values	0.308	0.0	0.5	1.0	1.0
		Appreciation	0.209	0.0	0.2	0.7	0.9
		Discomfort	0.197	1.0	1.0	0.7	0.2
U_s^T	Society	Accessibility	0.207	0.0	0.1	0.7	0.9
	0.312	Employment	0.176	0.0	0.5	1.0	1.0
		Induced effects	0.211	0.0	0.3	0.6	1.0
		Security	0.301	0.0	0.2	0.6	1.0
U_t^T	Transportation	Congestion	0.392	0.0	0.4	0.9	1.0
	0.263	Interferences	0.307	1.0	0.8	0.6	0.2

Table 7.10: Evaluation matrix (transposed) for the (simplified) problem of the tram-train of Como, with the weights computed for the subsectors of the four sectors

At this point, two equivalent ways are possible. The former consists in computing the weight of each subsector by multiplying its relative weight for the corresponding sector (Table 7.6) by the weight of the sector, and so on for every other weight visited along the path to the root of the attribute tree (in this case, there are only two levels). This provides an overall vector of $5 + 3 + 5 + 3 = 16$ *weights summing to one, represented in Table 7.11.*

With these weights the utility of Table 7.1 can be combined, simply multiplying the evaluation matrix by the overall vector. This produces the utilities of Table 7.13.

The second way allows also to evaluate the partial utilities of each alternative with respect to the single sectors. In order to do that one gradually scans the tree, multiplying the utility submatrices by the normalised weight subvectors level-by-level. In this way, one obtains the utilities of the alternatives normalised with respect to each sector. Table 7.12 reports the aggregated matrix $U^{(1)}$ *of the utilities for the first level of the tree+.*

From this matrix, the global utility for each alternative (referring to the root of the tree) can be obtained by multiplying the aggregated first-level matrix by the sector

		Alternatives			
Sectors	w	0	1	2	3
Air	0.053	0.0	0.1	0.6	0.8
Noise	0.047	0.4	0.2	0.4	0.6
Vibrations	0.045	0.4	0.2	0.4	0.6
Landscape	0.054	1.0	1.0	0.8	0.5
Territ. struct.	0.053	0.0	0.3	1.0	0.6
Costs	0.066	1.0	0.9	0.6	0.0
Revenues	0.053	0.0	0.3	0.7	1.0
Housing values	0.053	0.0	0.5	1.0	1.0
Appreciation	0.065	0.0	0.2	0.7	0.9
Discomfort	0.061	1.0	1.0	0.7	0.2
Accessibility	0.065	0.0	0.1	0.7	0.9
Employment	0.055	0.0	0.5	1.0	1.0
Induced effects	0.066	0.0	0.3	0.6	1.0
Security	0.079	0.0	0.2	0.6	1.0
Congestion	0.103	0.0	0.4	0.9	1.0
Interferences	0.081	1.0	0.8	0.6	0.2

Table 7.11: Evaluation matrix (transposed) for the (simplified) problem of the tram-train of Como, with the overall weight vector

		Alternatives			
Submatrix	Sector	0	1	2	3
U_a^T	Environment 0.252	0.360	0.371	0.654	0.620
U_e^T	Economy 0.173	0.385	0.593	0.755	0.616
U_s^T	Society 0.312	0.197	0.411	0.732	0.801
U_t^T	Transportation 0.263	0.307	0.463	0.718	0.754

Table 7.12: Evaluation matrix of the alternatives, aggregated with respect to the four sectors

weight vector, once again obtaining the values reported in Table 7.13.

	Alternatives			
Utilities	0	1	2	3
$u = U^{(1)} w^{(1)}$	0.299	0.446	0.713	0.711

Table 7.13: Overall utility vector

7.2.7 Rank reversal

The main defect of the Analytic Hierarchy Process, which has given rise to a lively debate, also involving philosophical aspects, is the phenomenon of *rank reversal*: *the order of the alternatives substantially depends on what alternatives are present.* The reason is that the evaluations u_{xl} are not absolute utility values, but derive from pairwise comparisons, and therefore depend on the values actually compared.

The same problem occurs in the soccer championship, which is actually an ordering system based on pairwise comparisons, even if the computation of the ordering function is performed in a different way. Let us suppose that teams A and B win over different subsets of other teams, and that A is slightly stronger than B; if a new team C is admitted, which is dominated by both, but plays slightly better in the matches against A than in those against B, this induces a redistribution of the weights that can lead to the final victory of B. Similar phenomena occur in electoral competitions, where the presence of a losing candidate can hamper in a different measure the stronger candidates, thus modifying the final result of the elections.

The problem also occurs in the case of perfectly consistent pairwise comparison matrices. In particular, there are examples in which even splitting an alternative into two perfectly identical alternatives, which yields a problem absolutely equivalent to the original one, modifies the final order. Of course, in such a situation, the two identical alternatives should simply end up in the same position without influencing the order of the other ones.

Example 58 *Let us consider a variation of Example 56, which concerned the purchase of a house. Without any modification with respect to the three alternatives, let us introduce a fourth alternative a_4. The new pairwise comparison matrices are identical to the original ones, except that they are framed with an additional row and column providing the comparisons which concern a_4.*

1. *price*

$$\Lambda_p = \begin{bmatrix} 1 & 1/9 & 1/8 & 1/4 \\ 9 & 1 & 9/8 & 9/4 \\ 8 & 8/9 & 1 & 2 \\ 4 & 4/9 & 1/2 & 1 \end{bmatrix} \Rightarrow u_p = \begin{bmatrix} 1/22 \\ 9/22 \\ 8/22 \\ 4/22 \end{bmatrix}$$

2. *size*

$$\Lambda_d = \begin{bmatrix} 1 & 9 & 9 & 9 \\ 1/9 & 1 & 1 & 1 \\ 1/9 & 1 & 1 & 1 \\ 1/9 & 1 & 1 & 1 \end{bmatrix} \Rightarrow u_d = \begin{bmatrix} 9/12 \\ 1/12 \\ 1/12 \\ 1/12 \end{bmatrix}$$

3. *life quality in the area*

$$\Lambda_q = \begin{bmatrix} 1 & 1/9 & 1/4 & 1/8 \\ 9 & 1 & 9/4 & 9/8 \\ 4 & 4/9 & 1 & 1/2 \\ 8 & 8/9 & 2 & 1 \end{bmatrix} \Rightarrow u_q = \begin{bmatrix} 1/22 \\ 9/22 \\ 4/22 \\ 8/22 \end{bmatrix}$$

4. purchase conditions

$$\Lambda_c = \begin{bmatrix} 1 & 1/3 & 3/5 & 3/5 \\ 1/3 & 1 & 1/5 & 1/5 \\ 5/3 & 5 & 1 & 1 \\ 5/3 & 5 & 1 & 1 \end{bmatrix} \Rightarrow u_c = \begin{bmatrix} 3/14 \\ 1/14 \\ 5/14 \\ 5/14 \end{bmatrix}$$

The new alternative subtracts from the other ones part of the utility with respect to each attribute. Its performance is rather scarse on the first two attribute, whereas it is the second-best for attribute "quality" and the first on a par with a_3 for the attribute "conditions". The new evaluation matrix and global utility vector are as follows:

$$u = Uw = \begin{bmatrix} 1/22 & 9/12 & 1/22 & 3/14 \\ 9/22 & 1/12 & 9/22 & 1/14 \\ 8/22 & 1/12 & 4/22 & 5/14 \\ 4/22 & 1/12 & 8/22 & 5/14 \end{bmatrix} \cdot \begin{bmatrix} 1/4 \\ 1/4 \\ 1/4 \\ 1/4 \end{bmatrix} = \begin{bmatrix} 0.264 \\ 0.243 \\ 0.246 \\ 0.246 \end{bmatrix}$$

which leads to choose alternative a_1. Alternative a_4, in fact, performing badly with respect to the second attribute (where a_1 is the first) and well with respect to the third attribute (where a_2 is the first) and the fourth (where a_3 is the first), in practice opens the way to a_1.

The *rank reversal* is unavoidable, unless by assuming an absolute scale of values.

7.2.8 Absolute scales, or *a priori* estimate method

In order to face the *rank reversal* problem, Saaty proposed a system based on pairwise comparisons not between single alternatives, but between predefined classes of alternatives.

For instance, when evaluating a group of candidates for a job which requires the knowledge of English and of computer software, the human resource department will not compare pairs of candidates with respect to their skills in these two aspects, but will establish standard skill levels (such as, "scarse", "sufficient", "fairly good", "good" and "very good") and will compare pairs of levels. Thus, a very good knowledge of English could be worth 7 times a fairly good knowledge, whereas a very good skill in computer software could be worth 3 times a fairly good knowledge, and so o. Then, the method will consist in assigning each alternative (that is, each candidate) to one of the classes.

The method has the additional advantage to reduce the number of comparisons required, as the size of the comparison matrix is given by the number of the classes, instead of the number of the alternatives. This also allows to apply the *AHP* to problems with a large number of alternatives. As the method allows to evaluate in advance the utility of each class, it can even be applied when not all alternatives are known since the beginning. For example, the candidates might show up in a long period of time, but should be immediately accepted or rejected, without knowing the following candidates. In such a situation, establishing that candidate A is 5 times better than candidate B is impossible if candidate B is still unknown and could be difficult even in the case of an already known candidate because a possible future candidate might make that ratio too small or too large.

The disadvantage of the method resides in the arbitrariety of the classes: moving the border of a class even by a small amount could move some alternatives from class to class, and consequently change their evaluation, possibly by a large amount.

Example 59 *Let us condider the hiring of a worker. There are four choice criteria: studies, curriculum, experience and proficiency in English.* A priori*, we do not know how many candidates will show up. Consequently, it is not possible to assign values to the pairwise comparisons until they have all been performed. We can, however establish, for each attribute, some predefined levels:*

- *studies (S)*
  1. *very good (vg)*
  2. *good (g)*
  3. *fairly good (fg)*
- *curriculum (C)*
  1. *excellent (e)*
  2. *very good (vg)*
  3. *good (g)*
- *experience (E)*
  1. *very relevant (vr)*
  2. *relevant (r)*
  3. *irrelevant (i)*
- *English proficiency (P)*
  1. *professional (p)*
  2. *basic (b)*

The fact that no level of studies below "fairly good" is considered could indicate that such levels are considered as unacceptable, so that the candidate is immediately discarded. In other words, such solutions are not part of the feasible region X and the associated impact is not part of its image F.

Now, before even interviewing the first candidate, we compare the levels by pairs, and obtain a vector of relative weights among them. For the studies:

$$\Lambda_S = \begin{array}{c|ccc} & vg & g & fg \\ \hline vg & 1 & 3 & 6 \\ g & 1/3 & 1 & 2 \\ fg & 1/6 & 1/2 & 1 \end{array} \Rightarrow u_S = \begin{bmatrix} 6/9 \\ 2/9 \\ 1/9 \end{bmatrix}$$

A very good level of studies is considered moderately preferable to a good one (weight 3 *on Saaty's scale) and somewhat intermediate between strongly and very strongly preferable to a fairly good one (weight* 6*).*

As for the curriculum:

$$\Lambda_C = \begin{array}{c|ccc} & e & vg & g \\ \hline e & 1 & 4 & 6 \\ vg & 1/4 & 1 & 3/2 \\ g & 1/6 & 2/3 & 1 \end{array} \Rightarrow u_C = \begin{bmatrix} 12/17 \\ 3/17 \\ 2/17 \end{bmatrix}$$

As for experience:

$$\Lambda_E = \begin{array}{c|ccc} & vr & r & i \\ \hline vr & 1 & 2 & 4 \\ r & 1/2 & 1 & 2 \\ i & 1/4 & 1/2 & 1 \end{array} \Rightarrow u_E = \begin{bmatrix} 4/7 \\ 2/7 \\ 1/7 \end{bmatrix}$$

As for the proficiency in English

$$\Lambda_P = \begin{array}{c|cc} & p & b \\ \hline p & 1 & 3 \\ b & 1/3 & 1 \end{array} \Rightarrow u_I = \begin{bmatrix} 3/4 \\ 1/4 \end{bmatrix}$$

Then, as usual, we will build a pairwise comparison matrix for the attributes, from which a weight vector will be derived.

$$\Lambda_w = \begin{array}{c|cccc} & S & C & E & P \\ \hline S & 1 & 1 & 3 & 4 \\ C & 1 & 1 & 3 & 4 \\ E & 1/3 & 1/3 & 1 & 4/3 \\ P & 1/4 & 1/4 & 3/4 & 1 \end{array} \Rightarrow w = \begin{bmatrix} 12/31 \\ 12/31 \\ 4/31 \\ 3/31 \end{bmatrix}$$

So, a candidate with very good studies, an excellent curriculum, irrelevant experience and a basic proficiency in English receives an evaluation equal to

$$u_S = 6/9 \quad u_C = 12/17 \quad u_E = 1/7 \quad u_P = 1/4$$

from which, as $U = [u_S\, u_C\, u_E\, u_P]$

$$Uw = \begin{bmatrix} 6/9 & 12/17 & 1/7 & 1/4 \end{bmatrix} \cdot \begin{bmatrix} 12/31 \\ 12/31 \\ 4/31 \\ 3/31 \end{bmatrix} = 0.574$$

On the other hand, a candidate with good studies, a very good curriculum, very relevant experience and a professional proficiency in English receives an evaluation equal to

$$u_S = 2/9 \quad u_C = 3/17 \quad u_E = 4/7 \quad u_P = 3/4$$

from which

$$Uw = \begin{bmatrix} 2/9 & 3/17 & 4/7 & 3/4 \end{bmatrix} \cdot \begin{bmatrix} 12/31 \\ 12/31 \\ 4/31 \\ 3/31 \end{bmatrix} = 0.301$$

and is therefore considered as less preferable. All this holds a priori*, even before the candidates start showing up.*

7.2.9 Rank reversal in utility theory

In the classical utility theory, the rank reversal phenomenon is impossible if the normalised utility functions are correctly built. However, the phenomenon can occur if the construction if incorrect. The most typical case is that in which a subsequent modelling phase introduces new alternatives in X, whose impact falls outside of the original region F, modifying the extreme values of some attributes, and therefore affecting their normalisation[2]. As a consequence, the other alternatives change their normalised utility with respect to such attributes, and their variations are

[2] New solution can appear during the modelling process due to the fact that such a process is not linear, but goes by subsequent loops. For example, in a big project, the solutions are detailed and modified in order to make them more agreeable as the analysis proceeds and determines their properties. Usually, the new solutions are not very far away from the original region F, so that the effect is negligible, but in general it should be taken into account.

not uniform. In particular, if the new solution is very good for an attribute, the best original solutions become much worse, whereas the worst original solutions are scarsely affected. It is possible that these variation modify the order of the solutions. In principle, the marginal rates of substitution should change to keep into account the new scaling, but this requires to tune them correctly.

Example 60 *A problem has two alternatives A and B, with impacts $f(A) = (100, 0)$ and $f(B) = (90, 100)$, that represent costs and have linear utility functions. The normalised utilities are $U_A = (1, 0)$ and $U_B = (0, 1)$. If the weights of the two components of the utility function are $w_1 = 0.6$ and $w_2 = 0.4$, the utilities of the two alternatives are:*

$$u(A) = 0.6 \cdot 1 + 0.4 \cdot 0 = 0.6 \qquad u(B) = 0.6 \cdot 0 + 0.4 \cdot 1 = 0.4$$

and A is the best one.

If a new alternative C is introduced, with impact $f(C) = (50, 0)$ and the utility u_1 is renormalised withouth updating the weights (and this is the mistake), U_A remains the same, whereas $U_B = (0.8, 1)$. Therefore, utility u_A remains 0.6 and utility u_B grows from 0.4 to $u_B = 22/25 = 0.88$, becoming the best one. In fact, the variation in the range of f_1 should have implied the retuning of the weights, as if the unit of measure of the two components of the utility had changed.

7.3 ELECTRE methods

The *ELECTRE methods* (*ELimination Et Choix Traduisant la REalité*[3]), developed by Roy[4] since the Sixties, start from a basic criticism of the assumption that the decision-maker be able to compare all pairs of impacts.

A typical example of this criticism is the *coffee paradox*, briefly presented in Section 3.5.5: two cups of coffee contain the same quantity of the same kind of coffee, but the former contains a given quantity f of sugar, whereas the latter contains a grain more ($f' = f + \epsilon$). For a sufficiently small value of ϵ, the two cups are obviously indifferent. As well, the latter cup is indifferent with respect to a third one, containing a quantity of sugar equal to $f'' = f + 2\epsilon$, which is indifferent with respect to a fourth cup, containing a quantity of sugar equal to $f + 3\epsilon$, and so on... By transitivity, two cups at the opposite extremes of the chain, and containing amounts of sugar equal to f and $2f$, respectively, must also be indifferent. But this is obviously false. The classical theory of utility avoids the paradox assuming that the decision-maker be able to discriminate each link of the chain, always indicating the better between the two cups.

Let us extend the concept to a multi-attribute framework, in order to introduce the solution proposed by the ELECTRE methods. Let us assume that the cup of coffee A contains a superior quality coffee with one grain more than the ideal quantity of sugar, whereas the cup of coffee B contains an average quality coffee with the ideal quantity of sugar. According to the approach of Pareto, the two cups would be incomparable. According to the classical multi-attribute utility theory, they can be compared by establishing a marginal rate of substitution between the attributes "coffee quality" and "sugar quantity". The ELECTRE methods follow yet another way. We talk of ELECTRE methods, with a plural, because several different methods have been proposed along the time. In the following, we present a summary of their main concepts, without referring explicitly to specific methods.

[3] Elimination and choice translating reality

[4] Bernard Roy (1934), French mathematician.

7.3.1 The outranking relation

The ELECTRE methods start from the Pareto preference relation which considers A preferable to B when it is not worse for all attributes, and enrich it by admitting that A could be preferable to B also when it is worse for some attributes, provided that the difference does not exceed a given threshold. This special kind of preference relation is usually denoted as *outranking relation* S (in French, *surclasser*). In the following, we assume that all indicator correspond to benefits for the decision-maker[5]. Moreover, we deal with finite problems, in which alternatives and impacts are in one-to-one correspondence, so that the distinction between the two concepts is blurred.

Definition 30 *Given two impacts $f, f' \in F \subseteq \mathbb{R}^p$, we say that f* outranks *f' ($f \preceq_S f'$), based on the thresholds $\epsilon_l \geq 0$ when impact f is not exceedingly worse than f' for all indicators $l \in P$ with respect to the thresholds:*

$$f \preceq_S f' \Leftrightarrow f_l \geq f'_l - \epsilon_l \text{ for all } l \in P$$

It is easy to see that, setting $\epsilon_l = 0$ for all $l \in P$, the outranking relation becomes the Paretian preference. Since usually $\epsilon_l > 0$, the outranking relation is richer, that is, it contains a larger set of pairs. The Paretian preference, therefore, implies the outranking relation, but not the opposite.

The outranking relation is clearly reflexive, but *in general it does not have the transitive property*, as shown in the following example.

Example 61 *Table 7.14 reports the benefits associated to four solutions A, B, C and D, with respect to four attributes f_l.*

	f_1	f_2	f_3	f_4
A	0.50	0.50	0.50	0.60
B	0.45	0.45	0.45	0.70
C	0.40	0.40	0.40	0.80
D	0.35	0.35	0.35	0.90

Table 7.14: Evaluation matrix (example)

Assuming that $\epsilon_l = 0.05$ ($l = 1, \ldots, 4$), the impacts of alternatives A and B differ less than the threshold ϵ_l for the first three attributes, while B is clearly better than A for the fourth one. So, B outranks A ($B \preceq_S A$), but not vice versa. For the same reason, C outranks B and D outranks C. It would seem to be possible to conclude that D outranks A, but this is false, because D is clearly worse than A with respect to three of the four criteria; indeed, D and A are incomparable.

$$D \preceq_S C \preceq_S B \preceq_S A \text{ but } D \not\preceq_S B, D \not\preceq_S A, C \not\preceq_S A$$

In summary, in this case the Paretian preference would be completely empty and would provide no useful information. On the contrary, the outranking relation, translated into a directed graph, would correspond to a path going from D to C, to B, to A. This does not mean that solution D is the best and solution A the worst: in fact, A is better than D for at least three indicators. While in the Paretian case the solutions with no ingoing arcs dominate the other ones, the outranking relation is harder to deal with.

[5] Notice that we have made the opposite assumption for the Paretian case.

7.3.2 Refinement of the outranking relation

The definition of outranking introduced above easily produces relations too weak and too similar to the Paretian one when the thresholds ϵ_l are very small, or relations too rich, in which all impacts outrank each other, when the thresholds are too large. In both cases, the relation gives little information to the decision-maker.

The ELECTRE methods, therefore, propose to combine the basic definition with other conditions: the final outranking relation will include the pairs of impact that verify all conditions. Each condition is a relaxation of the Paretian preference, so that the Paretian preference implies each of them. Consequently, the Paretian preference is always a sufficient condition for the outranking relation. The conditions allow to take into account additional remarks, for example expressing the relative importance of the indicators through suitable weights w_l. The decision-maker must indicate both the thresholds ϵ_l and the other parameters that rule the outranking conditions. The values attributed to these parameters influence the result of the decision process. The weights w_l are positive and have a sum equal to one, as in the multi-attribute utility theory and in the *AHP*, but they are not used as multpliers in a convex combination. On the contrary, they are used with an approach that recalls the elections: an impact is considered preferable to another when the indicators with respect to which it is better have larger weights.

For each pair of impacts (f, f'), it is possible define the following three values:

- the *sum of the weights for the attributes for which* f_l *is better than* f'_l

$$w^+_{ff'} = \sum_{l \in P : f_l > f'_l} w_l$$

- the *sum of the weights for the attributes for which* f_l *is indifferent to* f'_l

$$w^=_{ff'} = \sum_{l \in P : f_l = f'_l} w_l$$

- the *sum of the weights for the attributes for which* f_l *is worse than* f'_l

$$w^-_{ff'} = \sum_{l \in P : f_l < f'_l} w_l$$

Thanks to the normalisation assumption, the following property holds:

$$w^+_{ff'} + w^=_{ff'} + w^-_{ff'} = \sum_{l \in P} w_l = 1 \text{ for all } f, f' \in F$$

Outranking conditions

Given a set of conditions, the outranking relation is given by the pair of impacts that satisfy all of them, and is therefore obtained by intersecting the relations defined by each condition. Some of these conditions are the following:

1. the *satisfaction of comparability thresholds*: impact f is not much worse than impact f' for all attributes:

$$f \preceq_{S_\epsilon} f' \Leftrightarrow f_l \geq f'_l - \epsilon_l \text{ for all } l \in P$$

   where $\epsilon_l \geq 0$ for all $l \in P$. This is the basic condition introduced above. Notice that for $\epsilon_l = 0$ this relation coincides with the Paretian preference, whereas for $\epsilon_l > 0$ it grows to include other pairs of impacts, without losing any Paretian pair.

2. the *concordance condition*: a subset of attributes of sufficient weight agree that f is not worse than f':

   $$f \preceq_{S_c} f' \Leftrightarrow c_{ff'} = w^{+}_{ff'} + w^{=}_{ff'} \geq \alpha_c \text{ with } \alpha_c \in [0;1]$$

   For $\alpha_c = 1$, this yields the Paretian preference, since $w^{-}_{ff'} \leq 1 - \alpha_c = 0$. For smaller values of α_c (down to 0), the relation includes larger and larger subsets of pairs.

3. the *discordance condition*: no attribute rejects with exceeding strength the statement that f is better than f':

   $$f \preceq_{S_d} f' \Leftrightarrow d_{ff'} = \begin{cases} \dfrac{\max\limits_{l \in P} \left[\max \left(f'_l - f_l, 0 \right) \right]}{\max\limits_{l \in P} |f_l - f'_l|} & (\text{for } f \neq f') \\ 0 & (\text{for } f = f') \end{cases} \leq 1 - \alpha_d \text{ with } \alpha_d \in [0;1]$$

   The numerator measures the maximum difference between the two impacts with respect to the indicators for which f' is better, that is $f'_l > f_l$. The denominator measures the overall maximum difference between the two impacts. The ratio varies between 0, when f is never worse than f', and 1, when f is worse than f' for the indicator with the maximum difference. This index expresses the regret that one would feel rejecting f' in favour of f. For $\alpha_d = 1$, one obtains the Paretian preference[6], for other values the relation enlarges including other pairs of impacts. Since different attributes are combined in the index, it is important that they are normalised in advance, in order to make the comparison independent from the units of measure adopted.

The fundamental rationale of the ELECTRE methods consists in taking into account both the values of the indicators f_l and of their weights w_l, but dealing with them separately. This allows to avoid the linear combination process of weights and indicators, that makes sense only under rather strict assumptions, that is when the indicators are somehow turned into utility functions, which must be summable, adimensional (that is, normalised), etc... Moreover, this also allows to tune the relative influence of indicators and weights on the final result. If the values estimated for the weights appear more reliable, one can tighten the concordance threshold (or similar thresholds defined on the weights), so that the corresponding conditions remove from the outranking relation several pairs of impacts. On the other hand, if the values estimated for the indicators appear more reliable, one can tighten the discordance threshold (or similar thresholds defined on the indicators), so that the conditions on the indicators remove from the outranking relation several pairs of impacts. This allows to obtain outranking relations that depend more on the weights than on the indicator and vice versa. It should be noticed that, contrary to this rationale, some authors have proposed conditions in which weights and indicators are combined (in particular, there is a weighted definition of discordance, in which the differences between the values of the indicators are multiplied by the corresponding weights), but we do not go into details about these variants.

The values of the parameters strongly influences the final result. This requires, unless there is a strong consensus, a sensitivity analysis showing the appearance and disappearance of preference pairs as the parameters vary. If the analysis suggests that the results have a weak dependence on the parameter values, the choice is correct. For the concordance and discordance thresholds, in particular, it has been

[6] The textbooks usually report α_d on the right-hand-side of the inequality. Here, we have preferred to write $1-\alpha_d$ so that all conditions reduce to the Paretian preference when the threshold is equal to 1, as it was for the concordance condition.

proposed to set the thresholds in an adaptive fashion, assigning them the average values of the concordance and discordance coefficients with respect to all the pairs of impacts:

$$\alpha_c = \frac{\sum\limits_{f,f' \in F} c_{ff'}}{|F|\,(|F|-1)} \qquad \alpha_d = 1 - \frac{\sum\limits_{f,f' \in F} d_{ff'}}{|F|\,(|F|-1)}$$

These values have a good probability to remove a large, but not excessive, fraction of the potential pairs, thus yielding a final relation that is neither too rich nor too poor[7].

Example 62 *Let us consider a simple example with three solutions, evaluated on the basis of four indicators. Table 7.15 reports their performance under the form of adimensional benefits in an evaluation matrix $\{f_{xl}\}$ and the normalised weights of the indicators in a vector w.*

f_{xl}	f_1	f_2	f_3	f_4
a_1	1	0.7	0.6	0.8
a_2	0.8	0.5	1	1
a_3	0.4	1	0.6	0.6

w	0.3	0.4	0.2	0.1

Table 7.15: Evaluation matrix for an application of the ELECTRE methods

Let us suppose that the comparability thresholds be large ($\epsilon_l = 0.5$): all three solutions outrank each other, with the exception of the pair (a_3, a_1) (see the upper left graph in Figure 7.2).

Let us now consider the weights of the indicators and build the concordance matrix C, including all coefficients $c_{ff'}$:

C	a_1	a_2	a_3
a_1	*1*	*0.7*	*0.6*
a_2	*0.3*	*1*	*0.6*
a_3	*0.6*	*0.4*	*1*

Let us set a concordance threshold equal to $\alpha_c = 0.5$. The relation S_c includes all pairs with concordance $c_{ff'} \geq \alpha_c$, that is

$$a_1 \preceq_{S_c} a_2,\ a_2 \preceq_{S_c} a_3 \text{ and } a_3 \sim_{S_c} a_1$$

besides the obvious reflexive pairs $a_i \preceq_{S_c} a_i$ for all $a_i \in X$ (see the upper left graph in Figure 7.2).

Finally, from the indicator values we derive the discordance matrix D, which includes all coefficients $d_{ff'}$:

D	a_1	a_2	a_3
a_1	*0*	*1*	*0.5*
a_2	*0.5*	*0*	*1*
a_3	*1*	*0.8*	*0*

[7] But they make the threshold dependent on the considered impacts, therefore reproposing the insidious problem of *rank reversal*: a bad alternative can lower the thresholds, introducing new outranking pairs that can modify the result of the decision process.

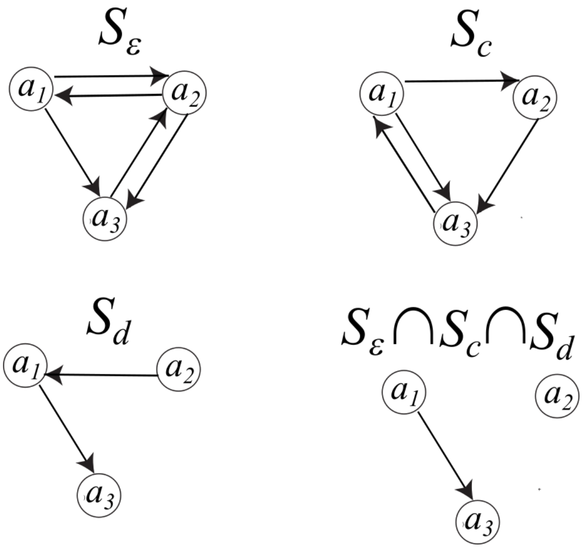


Figure 7.2: Combination of the outranking relations generated by the comparability thresholds (upper left graph), the concordance condition (upper right graph) and the discordance condition (lower left graph) in a resulting outranking relation (lower right graph)

Let us set a discordance threshold equal to $\alpha_d = 0.5$. *The relation* S_d *includes all pairs with discordance* $d_{ff'} \leq 1 - \alpha_d$, *that is*

$$a_1 \preceq_{S_d} a_3 \text{ and } a_2 \preceq_{S_d} a_1$$

besides the obvious reflexive pairs $a_i \preceq_{S_d} a_i$ *for all* $a_i \in X$ *(see the lower left graph in Figure 7.2).*

Intersecting the three relations, one obtains the final outranking relation

$$a_1 \preceq_S a_3$$

besides the reflexive pairs $a_i \preceq_{S_c} a_i$ *for all* $a_i \in X$ *(see the lower right graph in Figure 7.2).*

7.3.3 Kernel identification

Given a finite problem, the impacts in F can be described as a graph whose arcs represent the preference relation. The Paretian region corresponds in that graph to the subset of nodes with zero indegree. The outranking relation, by contrast, is not transitive, and this forbids to apply the same simple criterium to determine the subset of plausible alternatives for the decision process. In the ELECTRE methods, such a subset is obtained with a slightly more complex, iterative procedure.

Definition 31 *We denote as* kernel *the subset of alternatives obtained with the following procedure:*

1. *start with an empty kernel* ($K := \emptyset$)*;*
2. *add to the kernel the subset of all solutions with no ingoing arcs in the current graph* ($K := K \cup \{x \in X : \nexists x' \in X : x' \prec_S x\}$)*;*

3. *remove from the graph all solutions outranked by a kernel solution* ($X := X \setminus \{x \in X : \exists x' \in K : x' \prec_S x\}$)*;*

4. *if the reduced graph contains only the kernel, terminate; otherwise, go back to step 2.*

Example 63 *Given the graph in Figure 7.3, at the first step G, D and E enter the kernel. Then, B is removed from the graph (it is outranked by E), as well as C and F (outranked by D and E). Now, A enters the kernel. The algorithm terminates.*

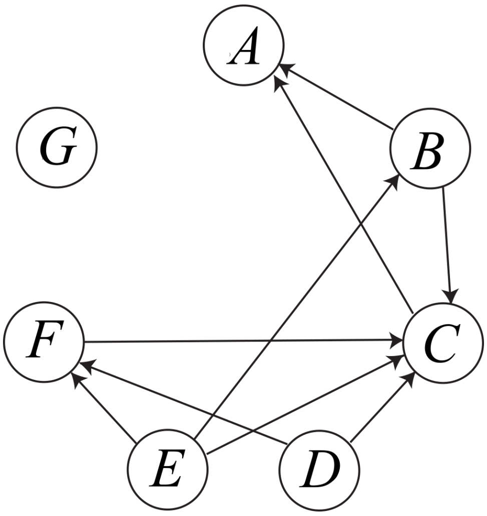


Figure 7.3: A possible outranking graph (self-loops not reported for the sake of simplicity)

The presence of A in the kernel could look strange, since it is outranked by B and C, that do not belong to the kernel. The reason is that being outranked is a weaker condition than being dominated. If the plausible alternatives are the ones in the kernel and none of them outranks A, this means that there is no strong reason to reject A.

The case of cyclic graphs

If the outranking graph contains circuits, that is directed outranking cycles, it is possible that the procedure does not terminate. For example, there could be pairs of alternatives that outrank each other, which correspond to circuits of two arcs in the graph. Sometimes, such circuits can be removed by asking the decision-maker to choose between the two alternatives, or to tune the compatibility thresholds and the weights, and therefore the graph. In general, however, the problem concerns longer circuits, that are harder to remove.

The following matrix evaluates three alternatives with respect to three indicators.

$f(x)$	f_1	f_2	f_3
A	3	2	1
B	2	1	3
C	1	3	2

If we set thresholds $\epsilon_l = 2$ for $l = 1, \dots, 3$, the three impacts outrank each other with respect to the comparability thresholds. If the three indicators have the same weights, the concordance condition can only leave all pairs unchanged or remove all of them from the relation, and the same occurs for the discordance condition. The symmetry of the problem imposes to keep or reject all pairs.

The literature has proposed different, more or less satisfactory, approaches to deal with the case of cyclic graphs. The simplest one is to merge each circuit into a supernode, dealing with the single nodes as if they were indifferent, that is including or excluding all of them from the kernel.

Outranking regions

It is useful to perform a sensitivity analysis with respect to the concordance and discordance thresholds, α_c and α_d, in order to determine for each solution the limit values that imply its being outranked by another solution[8]

Definition 32 *We denote as* outranking region *of a solution the set of all pairs* (α_c, α_d) *given which the solution is outranked only by itself.*

We remark that decreasing the thresholds increases the number of impact pairs in the outranking relation, and therefore makes it easier for each solution to be outranked. Every potential arc (f', f) entering a node f admits a minimum value for α_c under which the arc exists and above which it does not exist. The same holds for α_d. Consequently, the region of plane $\alpha_c\alpha_d$ in which an arc exists is a rectangle including the origin. The region in the plane in which a node is outranked, that is it has an ingoing arc, is the union of all such rectangles. The outranking region of a solution, therefore, is the complement of that union.

The stronger solutions are hard to outrank, and therefore have larger outranking areas, which get close to the origin, whereas the weaker solutions have smaller areas, limited to the region around the opposite corner $(1, 1)$. The outranking regions allow to sort the solutions with a reflexive and transitive, but not complete, relation, that is a partial order.

Example 64 *Let us consider again Example 62. The outranking area of alternative* a_1 *is determined by the two arcs* (a_2, a_1) *and* (a_3, a_1)*, that is by the two outranking relations* $a_2 \preceq_S a_1$ *and* $a_3 \preceq_S a_1$*. These arcs exist, respectively, for*

$$a_2 \preceq_S a_1 \Leftrightarrow \begin{cases} c_{21} = 0.3 \geq \alpha_c \\ d_{21} = 0.5 \leq 1 - \alpha_d \end{cases} \Leftrightarrow \begin{cases} \alpha_c \leq 0.3 \\ \alpha_d \leq 0.5 \end{cases}$$

and

$$a_3 \preceq_S a_1 \Leftrightarrow \begin{cases} c_{31} = 0.6 \geq \alpha_c \\ d_{31} = 1 \leq 1 - \alpha_d \end{cases} \Leftrightarrow \begin{cases} \alpha_c \leq 0.6 \\ \alpha_d \leq 0 \end{cases}$$

Figure 7.4 represents these two regions as grey rectangles (one of them is reduced to a segment of the horizontal axis). The white area complementary to these two rectangles is the outranking area of alternative a_1*.*

As for alternative a_2*, the arcs* (a_1, a_2) *and* (a_3, a_2) *correspond to the outranking relations* $a_1 \preceq_S a_2$ *and* $a_3 \preceq_S a_2$ *and to the rectangles*

$$\begin{cases} c_{12} = 0.7 \geq \alpha_c \\ d_{12} = 1 \leq 1 - \alpha_d \end{cases} \Leftrightarrow \begin{cases} \alpha_c \leq 0.7 \\ \alpha_d \leq 0 \end{cases}$$

[8] In principle, the analysis should refer only to the alternatives of the kernel, because only in that case the outranked alternative would be expelled from the kernel, but the textbooks are not clear under this respect. For the sake of simplicity, we consider a generic outranking.

and

$$\begin{cases} c_{32} = 0.4 \geq \alpha_c \\ d_{32} = 0.8 \leq 1 - \alpha_d \end{cases} \Leftrightarrow \begin{cases} \alpha_c \leq 0.4 \\ \alpha_d \leq 0.2 \end{cases}$$

and therefore the outanking area is the region depicted in Figure 7.5.

Finally, the outranking area for alternative a_3 is determined by the arcs (a_1, a_3) and (a_2, a_3), the outranking relations $a_1 \preceq_S a_3$ and $a_2 \preceq_S a_3$ and the rectangles

$$\begin{cases} c_{13} = 0.6 \geq \alpha_c \\ d_{13} = 0.5 \leq 1 - \alpha_d \end{cases} \Leftrightarrow \begin{cases} \alpha_c \leq 0.6 \\ \alpha_d \leq 0.5 \end{cases}$$

and

$$\begin{cases} c_{23} = 0.6 \geq \alpha_c \\ d_{23} = 1 \leq 1 - \alpha_d \end{cases} \Leftrightarrow \begin{cases} \alpha_c \leq 0.6 \\ \alpha_d \leq 0 \end{cases}$$

and reported in Figure 7.6.

The area of alternative a_1 is clearly larger than that of alternative a_3, whereas other relations are not strictly given.

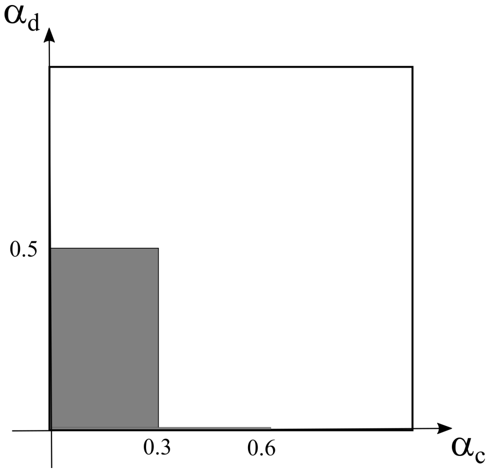


Figure 7.4: Outranking region of the alternative a_1

7.3.4 Creation of a weak ordering

Some ELECTRE methods include a final phase in which additional criteria are introduced to sort the solutions of the kernel. As in the Paretian case, this requires to perform a partial stretching.

Topological ordering

The nodes with no ingoing arcs correspond to nonoutranked solution; such solutions naturally stand out as the best ones. Similarly, the nodes with no outgoing arcs stand out as the worst ones[9].

Any topological ordering of the graph respects the outranking relation, meaning that it sets the outranking nodes in positions preceding the outranked ones. Notice that the topological order requires an acyclic graph, so that circuits must have been

[9] Of course, a node could fall under both categories.

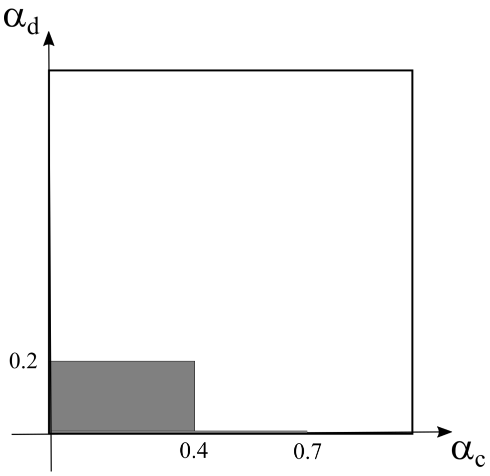


Figure 7.5: Outranking region of the alternative a_2

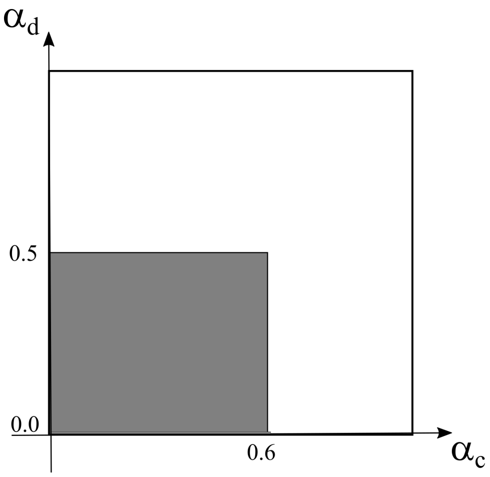


Figure 7.6: Outranking region of the alternative a_3

removed in a previous step, as already discussed in Section 7.3.3. Such a graph is partially ordered. The topological ordering introduces a stretching, because it selects one of the several total orders that are consistent with the original partial order. Notice that starting from the nodes with zero indegree and proceeding forward (*forward ordering*) produces a different result with respect to starting from the nodes with zero outdegree and proceeding backward (*backward ordering*). It has been proposed to determine both orderings, to assign to each node a ranking index in each of them (from $n = |X|$ down to 1 as in the Borda count), to make an average of the two indices and to use such an average to generate the final ordering. They are obviously highly empirical methods.

In the example of Figure 7.3, the forward ordering starts from G, D and E, that cannot be distinguished; the ordering goes on with B and F, then C and finally A. The backward ordering starts from A and G, then C, the pair B and F, and the pair D and E. The two orderings are similar, but (for example) the position of the isolated node G is completely different. If there are circuits with no ingoing arcs from the rest of the graph, all the nodes of the circuit are extracted together.

	A	B	C	D	E	F	G
Forward ordering	1	3.5	2	6	6	3.5	6
Backward ordering	1.5	4.5	3	6.5	6.5	4.5	1.5
Composition	1.25	4	2.5	6.25	6.25	4	3.75

The average of the ranking indices of the nodes in the two orderings suggests D and E as the best solutions, followed by B and F, G, C, ending with A.

Ordering with aggregated indices

This method builds for each impact a concordance and a discordance aggregated index, starting with those associated with the pairs of impacts:

- *concordance index*: it tries to describe the satisfaction associated to the choice of an impact; its values is large when the impact prevails on the other ones for many weighty attributes, while the other impacts prevail on it for few light attributes:
$$C_f = \sum_{g \in F} (c_{fg} - c_{gf}) \qquad f \in F$$
- *discordance index*: it decreases when the regret for a victory of the impact is small, while the regret for the defeat is large:
$$D_f = \sum_{g \in F} (d_{fg} - d_{gf}) \qquad f \in F$$

The *utopia point*, that is the impact that dominates (according to Pareto) all other impacts, and that is in general unfeasible, would have the maximum theoretical value of concordance: $C^* = n - 1$. The *distopia point*, that is the impact that is dominated by all other impacts, would on the contrary have the minimum theoretical value: $C^\dagger = 1 - n$.

It has been proposed to sort the impacts, and therefore the alternatives, by decreasing concordance indices C_f, or by increasing discordance indices D_f. As well, it has been proposed to build a biobjective programming problem with the two indices as objectives. The strength of the two indices resides in the fact that the former refers to the weights, while the latter refers to the attribute values.

Example 65 *In Example 62, the aggregate indices values are:*

$$C = \begin{bmatrix} 1 + 0.7 + 0.6 - 1 - 0.3 - 0.6 \\ 0.3 + 1 + 0.6 - 0.7 - 1 - 0.4 \\ 0.6 + 0.4 + 1 - 0.6 - 0.6 - 1 \end{bmatrix} = \begin{bmatrix} 0.4 \\ -0.2 \\ -0.2 \end{bmatrix}$$

and

$$D = \begin{bmatrix} 0 + 1 + 0.5 - 0 - 0.5 - 1 \\ 0.5 + 0 + 1 - 1 - 0 - 0.8 \\ 1 + 0.8 + 0 - 0.5 - 1 - 0 \end{bmatrix} = \begin{bmatrix} 0 \\ -0.3 \\ 0.3 \end{bmatrix}$$

Therefore, the ranking based on the aggregated concordance index (by decreasing values) puts a_1 before a_2 and a_3 in a tie. The ranking based on the aggregated discordance index (by increasing values) puts a_2 before a_1 and a_1 before a_3.

7.4 Exercises[*] [†] [‡]

Exercise 1

Consider the following decision problem with 5 alternatives and 3 criteria, characterized by the following impacts (benefits) and weights:

Attributes	a_1	a_2	a_3	a_4	a_5
f_1	30	0	60	90	120
f_2	5	6	4	2	7
f_3	18	26	17	20	16

	weights
w_1	0.1
w_2	0.6
w_2	0.3

Normalise the matrix for each criterium between the minimum ad maximum value, and compute the concordance and discordance matrices.

Build the outranking graph based on the three following conditions:

1. concordance with $\alpha_c = 0.75$, that is $w^+_{ff'} + w^=_{ff'} \geq 0.75$;
2. discordance with $\alpha_d = 0.25$, that is $\dfrac{\max_{l \in P: f_l < f'_l} |f_l - f'_l|}{\max_{l \in P} |f_l - f'_l|} \leq 0.75$

with arbitrarily large comparability thresholds ($\epsilon_l = +\infty$ for all $l \in P$).

Determine the kernel of the outranking graph.

Sort the alternatives based on the concordance index and on the discordance index.

Solution

The normalised evaluation matrix is:

Attributes	a_1	a_2	a_3	a_4	a_5
$\tilde{f}_1$	0.25	0.00	0.50	0.75	1.00
$\tilde{f}_2$	0.6	0.8	0.4	0.0	1.0
$\tilde{f}_3$	0.2	1.0	0.1	0.4	0.0

The outranking graph associated to the comparability thresholds is a complete graph of 5 nodes that correspond to the alternatives.

The concordance matrix is:

$$C = \begin{bmatrix} 1 & 0.1 & 0.9 & 0.6 & 0.3 \\ 0.9 & 1 & 0.9 & 0.9 & 0.3 \\ 0.1 & 0.1 & 1 & 0.6 & 0.3 \\ 0.4 & 0.1 & 0.4 & 1 & 0.3 \\ 0.7 & 0.7 & 0.7 & 0.7 & 1 \end{bmatrix}$$

The corresponding outranking graph includes 5 nodes corresponding to the alternatives, and the following arcs ($c_{ij} \geq \alpha_c = 0.75$): (a_1, a_3), (a_2, a_1), (a_2, a_3), (a_2, a_4), plus all self-loops (a_i, a_i) for $i = 1, \ldots, 5$.

[*] The solutions of these exercises have not yet been revised: error reports are welcome.

[†] I owe several of these exercises to exam texts of professor Alberto Colorni.

[‡] Welcome in the roaring Nineties. Even though I was a student in those years, I am astonished at the idea that exam exercises could take so many calculations.

The discordance matrix is:

$$D = \begin{bmatrix} 0 & 1 & 1 & 0.8333 & 1 \\ 0.3125 & 0 & 0.5555 & 0.9375 & 1 \\ 0.8 & 1 & 0 & 0.75 & 1 \\ 1 & 1 & 1 & 0 & 1 \\ 0.2666 & 1 & 0.1666 & 0.4 & 0 \end{bmatrix}$$

The corresponding outranking graph includes 5 nodes and the following arcs ($d_{ij} \leq 1-\alpha_d = 0.75$): (a_2, a_1), (a_2, a_3), (a_3, a_4), (a_5, a_1), (a_5, a_3), (a_5, a_4), plus all self-loops (a_i, a_i) for $i = 1, \dots, 5$.

The overall outranking graph includes 5 nodes and the arcs (a_2, a_1) and (a_2, a_3), plus all self-loops (a_i, a_i) for $i = 1, \dots, 5$.

Its kernel is therefore the subset $K = \{a_2, a_4, a_5)$.

The concordance indices of the alternatives are:

$$C_f = \sum_{f' \in F} c_{ff'} - \sum_{f' \in F} c_{f'f} = [-0.2 \; 2.0 \; -1.8 \; -1.6 \; 1.6]$$

which implies the ordering $a_2 \prec a_5 \prec a_1 \prec a_4 \prec a_3$.

The discordance indices of the alternatives are:

$$D_f = \sum_{f' \in F} d_{ff'} - \sum_{f' \in F} d_{f'f} = [1.4542 \; -1.1945 \; 0.8279 \; 1.0792 \; -2.1668]$$

which implies the ordering $a_5 \prec a_2 \prec a_3 \prec a_4 \prec a_1$.

Exercise 2

A community centre is looking for a new location: there are four alternatives (A, B, C and D) besides alternative 0 (staying in the present location). It has been decided that the choice among the five alternative should be definitive and that it shall be taken based on three factors: costs, accessibility and prestige of the location. The following table provides the utilities associated to each alternative and factor, on a scale from 0 to 100. It provides also a weight vector for the three factors.

Indicators	A	B	C	D	0
Costs	90	90	90	1	100
Accessibility	12	13	10	100	37
Prestige	30	1	5	100	10

	weights
w_1	1/3
w_2	1/3
w_3	1/3

Represent the problem with the *AHP*, expressing the pairwise comparison matrices between the alternatives with Saaty's scale (i. .e., when assigning the values, always choose the closest available one in Saaty's scale).

State whether any of the matrices is consistent, indicating the associated ordering vector.

Compute the concordance and the discordance matrices according to the ELECTRE methods.

Solution

Solution not available.

Exercise 3

Consider the following hierarchy in a multiple criteria decision problem:

- the decision-maker wants to optimise two main criteria f_1 and f_2, for which the following pairwise comparison matrix built with Saaty's scale is known:

	f_1	f_2
f_1	1	4
f_2	1/4	1

- criterium f_1 combines two subcriteria s_{11} and s_{12} with the following pairwise comparison matrix:

	s_{11}	s_{12}
s_{11}	1	1/9
s_{12}	9	1

- criterium f_2 combines two subcriteria s_{21} and s_{22} with the following pairwise comparison matrix:

	s_{21}	s_{22}
s_{21}	1	3
s_{22}	1/3	1

- there are three alternatives a_1, a_2 e a_3;
- the pairwise comparison matrix between the alternatives according to subcriterium s_{11} are:

s_{11}	a_1	a_2	a_3
a_1	1	3	6
a_2	1/3	1	2
a_3	1/6	1/2	1

- the pairwise comparison matrix between the alternatives according to subcriterium s_{12} are:

s_{12}	a_1	a_2	a_3
a_1	1	1/4	1/4
a_2	4	1	1
a_3	4	1	1

- the pairwise comparison matrix between the alternatives according to subcriterium s_{21} are:

s_{21}	a_1	a_2	a_3
a_1	1	1/4	2
a_2	4	1	8
a_3	1/2	1/8	1

- the pairwise comparison matrix between the alternatives according to subcriterium s_{22} are:

s_{22}	a_1	a_2	a_3
a_1	1	7	2
a_2	1/7	1	1/3
a_3	1/2	3	1

State whether all the matrices are consistent, motivating the answer.

If there are inconsistent matrices, turn them into consistent ones, modifying only the pair of values λ_{13} and λ_{31}.

Once the matrices are consistent, compute the corresponding weight vectors.

Reapply the hierarchy in order to obtain the final ordering of the alternatives, indicating the single passages.

Insert a new alternative so as to provoke a *rank reversal.*

Solution

All matrices are consistent, except for the pairwise comparison matrix with respect to subcriterium s_{22}.

That matrix can be made consistent by replacing it with:

s_{11}	a_1	a_2	a_3
a_1	1	7	7/3
a_2	1/7	1	1/3
a_3	3/7	3	1

The weight vectors are:

- for criteria f_1 and g_2: $[\ 0.8\ 0.2\ ]^T$;
- for subcriteria s_{11} and s_{12}: $[\ 0.1\ 0.9\ ]^T$;
- for subcriteria s_{21} and s_{22}: $[\ 0.75\ 0.25\ ]^T$;
- for the alternatives with respect to subcriterium s_{11}: $[\ 0.66\ 0.22\ 0.11\ ]^T$;
- for the alternatives with respect to subcriterium s_{12}: $[\ 0.11\ 0.44\ 0.44\ ]^T$;
- for the alternatives with respect to subcriterium s_{21}: $[\ 0.18\ 0.72\ 0.09\ ]^T$;
- for the alternatives with respect to subcriterium s_{22}: $[\ 0.63\ 0.09\ 0.27\ ]^T$.

This provides the following pseudoutilities.

- $u_{a_1} = 0.053 + 0.079 + 0.027 + 0.031 = 0.190$;
- $u_{a_2} = 0.018 + 0.317 + 0.108 + 0.005 = 0.448$;
- $u_{a_3} = 0.009 + 0.317 + 0.013 + 0.014 = 0.362$.

that imply the ordering: $a_2 \prec a_3 \prec a_1$.

In order to provoke a rank reversal, it is necessary to introduce an alternative that behaves “nearly” as the winning alternative a_2 in the four pairwise comparison matrices between the alternatives [10]

[10] My guess is that in this way the pseudoutilities of a_2 are approximately halved, thus favouring a_3, but I have not checked yet.

Exercise 4

Consider the following decision problem with three alternatives and four attributes (benefits), associated with the weights w_i $(i = 1, \ldots, 4)$:

Attributes	A	B	C
u_1	0	100	80
u_2	100	83	0
u_3	70	20	100
u_4	40	100	20

	weights
w_1	0.25
w_2	0.30
w_3	0.40
w_3	0.05

State whether there are outranking relations, with concordance threshold $\alpha_c = 0.70$ and discordance threshold $\alpha = 0.60$.

Solution

The concordance matrix of the problem is:

$$C = \begin{bmatrix} 1 & 0.7 & 0.35 \\ 0.3 & 1 & 0.6 \\ 0.65 & 0.4 & 1 \end{bmatrix}$$

The discordance matrix is:

$$D = \begin{bmatrix} 0 & 1 & 0.8 \\ 0.5 & 0 & 0.9638 \\ 1 & 1 & 0 \end{bmatrix}$$

There is no outranking, because the discordance values are all $> 1 - \alpha_d$. Therefore, the outranking graph only includes the self-loops.

If considered alone, the concordance matrix would imply the outranking relation $A \prec_{S_c} B$.

Exercise 5

Determine the kernel according to the ELECTRE methods for a problem which has seven alternatives a_i $(i = 1, \ldots, 7)$ and an outranking graph with arcs: (a_1, a_2), (a_1, a_4), (a_2, a_4), (a_2, a_7), (a_4, a_3), (a_4, a_6), (a_3, a_7).

Solution

The kernel is a set of alternatives that do not outrank each other and such that every other alternative is outranked by at least one kernel alternative.

Based on this definition, the procedure:

1. includes a_1 and a_5 in the kernel;
2. removes a_2 and a_4 from the kernel;
3. includes a_3 and a_6 in the kernel;
4. removes a_7 from the kernel

so that in the end $K = \{a_1, a_3, a_5, a_6\}$.

Part IV

Models with multiple scenarios

Chapter 8

Models of uncertainty

The *Programming in conditions of uncertainty* deals with decision problems in which the choice is among alternatives whose impact depends not only on the choice of the decision-maker, but also on external factors that cannot be exactly predicted. These problems are characterised by having:

- a single decision-maker;
- a single objective;
- an uncertain environment: the impact of the decision depends also on variables which are not under the control of the decision-maker and about whose value only partial information is available.

Such problems can be formulated as:

$$\begin{gathered} \min_{x \in X} f(x, \omega) \\ \omega \in \Omega \end{gathered}$$

where

- X is the set of the *alternatives* or *solutions* x;
- Ω is the set of the *scenarios* ω (in statistics, *sample space*);
- $f(x, \omega)$ is the *impact* of solution x and of scenario ω, and is a real number;
- lower impacts are preferable to higher impacts (or vice versa).

It is important to observe that the decision-maker can choose the alternative x, but not the scenario ω, and that the choice of x precedes the unravelling of the scenario ω. If x were chosen after ω is known, the problem would reduce to a Mathematical Programming problem parameterised in ω, that is to solving

$$\min_{x \in X} f(x, \omega)$$

for each single value of $\omega \in \Omega$. The solution would then be a *strategy* $x^*(\omega)$.

Some typical examples of decision problems in conditions of uncertainty are:

- finance: selection of a stock portfolio (uncertainty on the future value of the stocks);

- marketing: development of new products and advertising campaigns (uncertainty on the response of the customers);
- research and development: opportunity of performing researches (uncertainty on the amount of costs and revenues);
- contract competitions: definition of the price of an offer (uncertainty on the costs and on the entity of the offers of the competitors);
- oil drillings: choice of where and when to drill (uncertainty on the size of the oilfield).

The problems in conditions of uncertainty include two main classes, though other ones have been proposed:

1. decisions in conditions of *ignorance*: the only information on scenario ω is that it falls within Ω;
2. decisions in conditions of *risk*: *the probability* π_ω *of each scenario* $\omega \in \Omega$ *is known* (if Ω is a discrete set) *or the probability density* $\pi(\omega)$ *is known* (if Ω is a continuous set).

In all these classes, it is possible to define relations of dominance that allow to reduce the set of interesting solutions. In general, these relations will not lead to a single choice, as they typically are *partial orders*, that is reflexive and transitive for the usual basic rationality considerations, but incomplete).

8.1 Dominance relations

Definition 33 *We say that alternative x* strongly dominates *alternative x' when its impact is at least as good in all scenarios $\omega \in \Omega$*

$$x \preceq x' \Leftrightarrow f(x,\omega) \leq f(x',\omega) \ \textit{for all}\ \omega \in \Omega$$

There is a clear formal analogy here with the concept of Paretian preference (see Definition 21 in Section 6): the scenarios are treated as distinct attributes, and one would like to achieve an optimal performance with respect to each of them. The fundamental difference is that the scenarios can be infinitely many, possibly even a continuous infinity. The conclusion, however, is the same: the alternative that are strictly dominated (that is dominated, but not dominating) can be rejected *a priori*; rejecting them is the only rational action that can be performed if no further information is added to the problem.

Notice that the term “strong” here is not meant to exclude the opposite relation (as in “strong” preference relations): unfortunately, the word is used to distinguish this kind of dominance from the probabilistic dominance discussed in the following.

Definition 34 *In the finite case, the impact $f(x,\omega)$ can be represented with an* evaluation matrix U*, having the alternatives x on the rows and the scenarios ω on the columns.*

This matrix has the same name as the matrix that represents the multidimensional impacts as a function of the alternatives x in multiple attribute problems (see Section 3.3.1), with the scenarios taking the place of the attributes. The two matrices are not equivalent, but there are some formal similarities in the way to deal with them. The first one is that, in order to find the nondominated alternatives, one must perform pairwise comparisons on the rows of the matrix, exactly as in the Paretian case (see Section 6.3.1).

Example 66 *Let us consider an investment on a given time horizon. There are four alternatives:*

1. *stock fund*
2. *bond fund*
3. *treasuries*
4. *"toxic" bonds*

*and a single objective, the ROI (*return-on-investment*), that must be maximised. Therefore, the best values are the larger ones, contrary to what assumed in Definition 33: either one adapts the definition of strong dominance of one reverts the sign of the objective function. The ROI of each alternative, however, is unknown* a priori*, as it depends also on the behaviour of economy in the considered period, which we will model with three possible scenarios:*

1. *recession*
2. *moderate growth*
3. *strong growth*

ROI	Recession	Moderate growth	Strong growth
Stock fund	-25%	0%	35%
Bond fund	-10%	5%	15%
Treasuries	8%	8%	8%
"Toxic" bonds	-5%	6%	8%

Table 8.1: Evaluation matrix for an investment problem with four alternatives and three scenarios (the impact is the ROI)

Table 8.1 reports the evaluation matrix of the investment problem. The "toxic" bonds are dominated by the treasuries. In fact, the latter have a strictly better ROI in case of recession and moderate growth (8% *versus* -5% *and* 8% *versus* 6%*) and identical in case of strong growth (*8%*). Therefore, the "toxic" bonds can be rationally excluded from the choice, whereas all other alternatives are rationally defensible.*

8.1.1 Probabilistic dominance*

An alternative definition of dominance, that is weaker (*i. e.*, it is always respected when the strong dominance holds) is based on the probability function, and therefore can be used only in conditions of risk.

Definition 35 *We say that alternative x* probabilistically dominates *alternative x' when for any threshold $\bar{f}$ the probability that x have impacts not worse than the threshold is not inferior to that of x'.*

$$x \preceq x' \Leftrightarrow P\left[f\left(x,\omega\right) \leq \bar{f}\right] \geq P\left[f\left(x',\omega\right) \leq \bar{f}\right] \text{ \textit{per ogni} } \bar{f} \in \mathbb{R}$$

*This section provides advanced concepts, that are not part of the course's syllabus.

This definition is not trivial to apply, given that it requires to verify a property on two distributions of probability. It is often applied, however, in the comparison of algorithms or models, where it is very rare that strong dominance hold.

Definition 36 *Suppose that three bets on the throw of a die are available, indicated with $X = \{a_1, a2, a_3\}$. The corresponding gains are reported in Table 8.2, of course with respect to the six possible outcomes of the throw. Notice that, since they are gains, we have to modify suitably the definitions of strong and probabilistic dominance.*

	ω_1	ω_2	ω_3	ω_4	ω_5	ω_6
a_1	1	1	2	2	2	2
a_2	1	1	1	2	2	2
a_3	3	3	3	1	1	1

Table 8.2: Choice among three bets on the throw of a die

Alternative a_1 strongly dominates alternative a_2, because it yields at least as good a gain for any outcome of the throw. Alternative a_3 does not dominate strongly alternative a_2, because the latter is better in the scenarios from ω_4 to ω_6. It dominates the latter, however, probabilistically, because

- $P[B \geq 1] = P[C \geq 1] = 1$
- $P[B \geq 2] = P[C \geq 2] = 3/6$
- $P[B \geq 3] = 0 < P[C \geq 3] = 3/6$

By contrast, a_1 and a_3 have no probabilistic dominance since

- $P[A \geq 2] = 4/6 > P[C \geq 2] = 3/6$
- $P[C \geq 3] = 3/6 > P[A =\geq 3] = 0$

8.2 Models of uncertainty

There are two main ways to describe uncertain situations, that reflect into two main forms of the scenario set Ω:

1. *scenario description*: Ω is a finite set, in which the single scenarios are explicitly listed;
2. *interval description*: Ω is the Cartesian product of a finite number of intervals, that is each exogeneous variable ω_k varies in a given interval, independently from the values of the other variables.

These are not the only possible cases: Ω could assume much more sophisticated forms. However, they are two very frequent special cases, which it is interesting to discuss in more detail.

Example 67 *Consider a shortest path problem between two nodes of a directed graph. Let us assume that the uncertainty be modelled by two exogeneous variables ω_1 and ω_2, that provide the travel time on two arcs of the graph, which represent two streets in the network on which the traffic congestion is particularly uncertain.*

A scenario description could consist in listing the possible combinations of the values of the two variables. For example, Ω *could consist of three scenarios, represented by the pairs* $\omega^{(1)} = (2,1)$, $\omega^{(2)} = (5,1)$ *and* $\omega^{(3)} = (2,6)$, *that are reported as red squares in Figure 8.1. The three scenarios correspond to situations in which, respectively, there is low congestion in both streets, high congestion in the former and low in the latter, or low congestion in the former and high in the latter. This model excludes the possibility of having high congestion in both streets.*

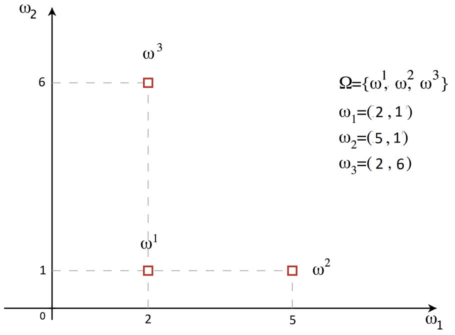


Figure 8.1: Scenario description of uncertainty in a shortest path problem

An interval description, on the contrary, could define a range of possible values for each of the two variables ω_1 *and* ω_2. *Figure 8.2 shows as a dashed region the scenario set* Ω *obtained as the Cartesian product of the intervals* $[2,5]$ *and* $[1,6]$. *This is a continuous set.*

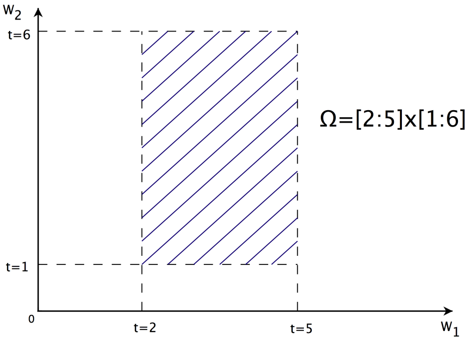


Figure 8.2: Interval description of uncertainty in a shortest path problem

The main difference between the two description is that often in the interval description the worst scenario is shared by all solutions (not always: it depends on the problem), whereas in the scenario description in general every solution has its own worst scenario. We shall see in Section 9.9 that this can make it easier to solve the interval models, with strong implications on the computational complexity.

8.3 Exercises*

Exercise 1

Solution

[*] The solutions of these exercises have not yet been revised: error reports are welcome.

Chapter 9

Decisions in conditions of ignorance

Full ignorance is the most problematic situation for a decision-maker, since the information on the exogenous part of the problem reduces to knowing that tha scenario falls within a given set. In order to optimise the objective, one should choose the best pair (x, ω), that is fix the best alternative and impose the best scenario. In the example of the financial investment, in order to maximise the return, it is necessary to invest in stocks and to guarantee that the economy grow strongly. However, this is out of the control of the decision-maker, who can fix only the x variables, and not the ω variables. This problem has no solution.

The approaches proposed in the literature consist in redefining the problem, replacing $f(x, \omega)$ with an auxiliary function $\phi_\Omega(x)$, that depends on x and on the overall set Ω, but not on the (unknown) value of ω, and then optimising the latter.

Definition 37 *We denote as* choice criterium *every definition of $\phi_\Omega(x)$ aimed to replace the impact $f(x, \omega)$.*

We now survey several choice criteria proposed in the literature. We shall see that each of these criteria has some advantage, but none satisfies all the properties that would be desirable for a rationally founded decision. The choice among them, therefore, only depends on the attitude of the decision-maker. We will apply each criterium to an example with four alternatives ($X = \{x^{(1)}, x^{(2)}, x^{(3)}, x^{(4)}\}$) and four scenarios ($\Omega = \{\omega^{(1)}, \omega^{(2)}, \omega^{(3)}, \omega^{(4)}\}$), whose evaluation matrix is reported in Table 9.1 (the values reported are costs).

$f(x, \omega)$	$\omega^{(1)}$	$\omega^{(2)}$	$\omega^{(3)}$	$\omega^{(4)}$
$x^{(1)}$	2	2	4	3
$x^{(2)}$	3	3	3	3
$x^{(3)}$	4	0	4	6
$x^{(4)}$	3	1	4	4

Table 9.1: Sample evaluation matrix: costs associated to the four alternatives and the four possible scenarios

9.1 Worst-case criterium

This criterium, also known as *pessimism criterium* or *Wald criterium*[1], consists in being pessimist and assuming that for each chosen solution x the future prepare the scenario implying the largest cost $\omega^{\dagger}(x) = \arg \max_{\omega \in \Omega} f(x, \omega)$. This assumption allows to compute the impact as a function only of the decision variable x (and of the overall set Ω), that is to reduce the problem to the minimisation of $\phi_{\text{worst}}(x) = f\left(x, \omega^{\dagger}(x)\right)$

$$\min_{x \in X} \phi_{\text{worst}}(x) = \min_{x \in X} \max_{\omega \in \Omega} f(x, \omega)$$

Table 9.2 reports for each alternative the value of the criterium and the scenario that produces it.

	$\omega^{(1)}$	$\omega^{(2)}$	$\omega^{(3)}$	$\omega^{(4)}$	$\phi_{\text{worst}}(x)$	$\omega^{\dagger}(x)$
$x^{(1)}$	2	2	4	3	4	$\omega^{(3)}$
$x^{(2)}$	3	3	3	3	3	$\omega^{(1)}$, $\omega^{(2)}$, $\omega^{(3)}$, $\omega^{(4)}$
$x^{(3)}$	4	0	4	6	6	$\omega^{(4)}$
$x^{(4)}$	3	1	4	4	4	$\omega^{(3)}$, $\omega^{(4)}$

Table 9.2: Application of the worst-case criterium to a sample problem with four alternatives and four scenarios (cost minimisation)

Now, it is enough to select the best alternative, that is the cheapest one. The worst-case criterium suggests to order the alternatives as follows: $x^{(2)} \prec x^{(1)} \sim x^{(4)} \prec x^{(3)}$.

9.2 Best-case criterium

The best-case criterium, also known as *optimism criterium*, is complementary to that of the worst-case: it consists in assuming that, for each solution x chosen by the decision-maker, the future answer by proposing the best scenario $\omega^{*}(x) = \arg \min_{\omega \in \Omega} f(x, \omega)$. This assumption allows to compute the impact as a function only of the decision variables x (and of the overall set Ω), reducing the problem to the minimisation of function $\phi_{\text{best}}(x) = f(x, \omega^{*}(x))$

$$\min_{x \in X} \phi_{\text{best}}(x) = \min_{x \in X} \min_{\omega \in \Omega} f(x, \omega)$$

Table 9.3 reports for each alternative the value of the criterium and the scenario that produces it. Then, one selects the best alternative, that is the one of minimum cost. The optimism criterium suggests to order the alternatives as follows: $x^{(3)} \prec x^{(4)} \prec x^{(1)} \prec x^{(2)}$.

9.3 Hurwicz criterium

The first two criteria look rather unbalanced in the evaluation of the scenarios. In order to guarantee a stronger balance, Hurwicz[2] proposed to build a choice criterium

[1] Abraham Wald (1902-1950), Hungarian Jew mathematician, fled to the United States after the annexion of Austria to Nazi Germany.

[2] Leonid Hurwicz (1917-2008), Polish Jew economist, who was stuck abroad by the Nazi invasion of Poland, and later took shelter in the United States.

	$\omega^{(1)}$	$\omega^{(2)}$	$\omega^{(3)}$	$\omega^{(4)}$	$\phi_{\text{best}}(x)$	$\omega^*(x)$
$x^{(1)}$	2	2	4	3	2	$\omega^{(1)}, \omega^{(2)}$
$x^{(2)}$	3	3	3	3	3	$\omega^{(1)}, \omega^{(2)}, \omega^{(3)}, \omega^{(4)}$
$x^{(3)}$	4	0	4	6	0	$\omega^{(2)}$
$x^{(4)}$	3	1	4	4	1	$\omega^{(2)}$

Table 9.3: Application of the best-case criterium to a sample problem with four alternatives and four scenarios (cost minimisation)

making a convex combination of the two:

$$\min_{x\in X} \phi_{\text{Hurwicz}}(x) = \min_{x\in X}\left[\alpha \max_{\omega\in\Omega} f(x,\omega) + (1-\alpha)\min_{\omega\in\Omega} f(x,\omega)\right]$$

where the *pessimism coefficient*[3] $\alpha \in [0;1]$ weighs the worst impact, while the complementary value $1-\alpha$ weighs the best one.

Such a coefficient allows to tune the weight of the two extreme scenarios in the choice: setting $\alpha = 1$ corresponds to the pessimism criterium, whereas setting $\alpha = 0$ corresponds to the optimism criterium.

Table 9.4 reports for each alternative the value of the criterium under the assumption that the pessimism coefficient be $\alpha = 0.6$, that is, that the decision-maker be slightly pessimistic. The resulting order is $x^{(4)} \prec x^{(2)} \prec x^{(1)} \prec x^{(3)}$.

	$\omega^{(1)}$	$\omega^{(2)}$	$\omega^{(3)}$	$\omega^{(4)}$	$\phi_{\text{Hurwicz}}(x)\ (\alpha = 0.6)$
$x^{(1)}$	2	2	4	3	$0.6 \cdot 4 + 0.4 \cdot 2 = 3.2$
$x^{(2)}$	3	3	3	3	$0.6 \cdot 3 + 0.4 \cdot 3 = 3$
$x^{(3)}$	4	0	4	6	$0.6 \cdot 6 + 0.4 \cdot 0 = 3.6$
$x^{(4)}$	3	1	4	4	$0.6 \cdot 4 + 0.4 \cdot 1 = 2.8$

Table 9.4: Application of Hurwicz criterium with $\alpha = 0.6$ to a sample problem with four alternatives and four scenarios (cost minimisation)

9.3.1 Tuning the pessimism coefficient

Since the solution depends on the value of the pessimism coefficient α, it is useful to discuss how it is possible to tune this parameter. A possible approach, similar to that used to determine the weight coefficients in the utility function for multi-attribute problems, is to find (possibly, to invent) a pair of reciprocally indifferent alternatives, to impose the equality of the corresponding values of the criterium $\phi_{\text{Hurwicz}}(x)$ and to solve the resulting equation in α. The simplest way to do that is to choose an alternative x, to ask the decision-maker to indicate its *certainty equivalent*, that is an alternative y, that in general is not a true alternative of the problem, with a uniform impact in all scenarios and that is overall indifferent to x.

In the example we are studying, let us choose the alternative $x^{(3)}$: its cost is 6 in the worst case and 0 in the best one. If the decision-maker states that an alternative y with a uniform cost equal to 4 would be indifferent to $x^{(3)}$, we can conclude that

[3] Of course, some authors define an optimism coefficient, that multiplies the best value of the impact, and use the complementary value as a multiplier of the worst impact. As well, if the impact is a benefit, instead of a cost, the maximum and minimum operations must be suitably modified.

$\phi_{\text{Hurwicz}}\left(x^{(3)}\right) = \phi_{\text{Hurwicz}}\left(y\right)$, that is

$$\alpha \max_{\omega\in\Omega} f(x^{(3)},\omega) + (1-\alpha) \min_{\omega\in\Omega} f(x^{(3)},\omega) = \alpha \max_{\omega\in\Omega} f(y,\omega) + (1-\alpha) \min_{\omega\in\Omega} f(y,\omega) \Rightarrow$$

$$\Rightarrow 6\alpha + 0\,(1-\alpha) = 4\alpha + 4\,(1-\alpha) \Rightarrow \alpha = \frac{2}{3}$$

It is not at all obvious that the decision-maker is able to indicate a certainty equivalent, given that we suppose not to have any information on the scenarios, besides the fact that they exist.

9.3.2 Sensitivity with respect to the pessimism coefficient

Given the uncertainty of the pessimism coefficient, it is in general adviceable to perform a sensitivity analysis showing whether the chosen solution would remain the same under small variations of its value. The analysis requires to build for each alternative $x \in X$ the function $\phi_{\text{Hurwicz}}\left(x\right)$, that depends on parameter $\alpha \in [0;1]$:

- $\phi_{\text{Hurwicz}}(x^{(1)}) = 4\alpha + 2(1-\alpha) = 4\alpha + 2 - 2\alpha = 2\alpha + 2$
- $\phi_{\text{Hurwicz}}(x^{(2)}) = 3\alpha + 3(1-\alpha) = 3\alpha + 3 - 3\alpha = 3$
- $\phi_{\text{Hurwicz}}(x^{(3)}) = 6\alpha + 0(1-\alpha) = 6\alpha$
- $\phi_{\text{Hurwicz}}(x^{(4)}) = 4\alpha + 1(1-\alpha) = 4\alpha + 1 - 1\alpha = 3\alpha + 1$

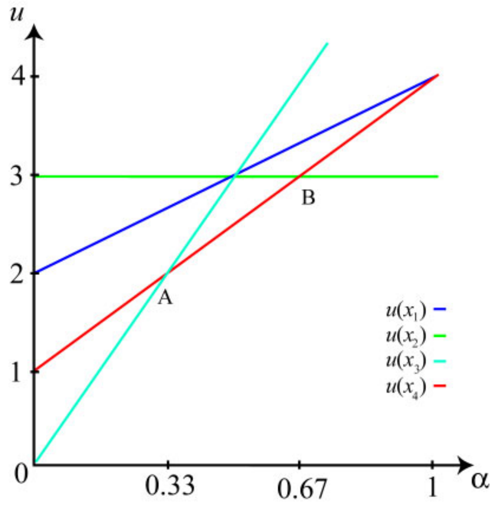


Figure 9.1: Sensitivity analysis with respect to the pessimism coefficient for a sample problem

Figure 9.1 reports the graph of the four functions. For every $\alpha \in [0;1]$, one must choose the solution, that is the function, that minimises the cost. We must therefore determine the values of α that identify point A, intersection of $\phi(x^{(3)})$ and $\phi(x^{(4)})$:

$$\phi(x^{(3)}) = \phi(x^{(4)}) \Rightarrow 6\alpha = 3\alpha + 1 \Rightarrow \alpha = \frac{1}{3} \Rightarrow A = (\frac{1}{3}, 2)$$

and point B, intersection of $\phi(x^{(2)})$ and $\phi(x^{(4)})$:

$$\phi(x^{(2)}) = \phi(x^{(4)}) \Rightarrow 3 = 3\alpha + 1 \Rightarrow \alpha = \frac{2}{3} \Rightarrow B = (\frac{2}{3}, 3)$$

Consequently, if $0 \leq \alpha \leq \frac{1}{3}$ the best alternative is $x^{(3)}$ (as with the optimism criterium); if $\frac{1}{3} \leq \alpha \leq \frac{2}{3}$ the best alternative is $x^{(4)}$; if $\frac{2}{3} \leq \alpha \leq 1$ the best alternative is $x^{(2)}$ (as with the pessimism criterium). The extreme values of α obviously confirm the results of the optimism and pessimism criteria. For intermediate values, however, Hurwicz criterium selects an alternative that the two previous criteria do not take into account.

The graph in Figure 9.1 also shows that alternative $x^{(1)}$ is always worse than alternative $x^{(4)}$, because the function associated to the former is above that associated to the latter. This does not mean that $x^{(1)}$ is dominated by $x^{(4)}$ (in fact, is is better in scenario $\omega^{(3)}$): the situation is similar to that of the Paretian, but unsupported, solutions, which were never optimal for any linear combination of the attributes (see Section 6.6). This is an intrinsic limitation of Hurwicz criterium: it is unable to generat all the nondominated solution by tuning its only parameter. The limitation is even worse than in the Paretian case, because the criterium only considers the two extreme scenarios, whereas the weighted-sum method combines all the attributes of the problem. On the other hand, the aim of Hurwicz criterium is not to generate the solutions, but only to order them.

9.4 Equiprobability criterium

The equiprobability criterium, also known as *Laplace criterium*, modifies Hurwicz criterium considering all the scenarios, instead of only the extreme ones. Without any information on the likelyhood of the scenarios, it combines the impacts applying the same weight to all of them:

$$\min_{x \in X} \phi_{\text{Laplace}}(x) = \min_{x \in X} \frac{\sum\limits_{\omega \in \Omega} f(x, \omega)}{|\Omega|}$$

This expression presumes that the scenario set is finite[4]. The Laplace criterium can be interpreted as the minimisation of the expected cost under the assumption that all scenarios have the same probability to occur (hence the first name of the criterium). We remind, however, that we are assuming to have no information on the probability of the scenarios.

Table 9.5 reports for each alternative the value of the criterium, that is the arithmetic mean of the impacts in the various scenarios. The resulting order is $x^{(1)} \prec x^{(2)} \sim x^{(4)} \prec x^{(3)}$.

	$\omega^{(1)}$	$\omega^{(2)}$	$\omega^{(3)}$	$\omega^{(4)}$	$\phi_{\text{Laplace}}(x)$
$x^{(1)}$	2	2	4	3	(2+2+4+3)/4 = 2.75
$x^{(2)}$	3	3	3	3	(3+3+3+3)/4 = 3.00
$x^{(3)}$	4	0	4	6	(4+0+4+6)/4 = 3.50
$x^{(4)}$	3	1	4	4	(3+1+4+4)/4 = 3.00

Table 9.5: Application of Laplace criterium to a sample problem with four alternatives and four scenarios (cost minimisation)

[4] A generalisation to infinite sample spaces is possible, but exceeds the limits of this course.

9.5 Regret criterium

This criterium, also known as *Savage criterium*[5], consists in evaluating the regret that the decision-maker would feel if the decision taken were wrong. The criterium works in subsequent phases:

1. determine the best alternative for each scenario $x^*(\omega)$
$$x^*(\omega) = \arg \min_{x \in X} f(x, \omega)$$
2. evaluate the *regret* $\rho(x, \omega)$ associated to each alternative x and each scenario ω, defined as the *loss incurred choosing x instead of the best alternative in scenario ω*:
$$\rho(x, \omega) = f(x, \omega) - f(x^*(\omega), \omega) = f(x, \omega) - \min_{x \in X} f(x, \omega)$$
3. assign to each solution the value of the maximum regret over all scenarios:
$$\phi_{\text{Savage}}(x) = \max_{\omega \in \Omega} \rho(x, \omega)$$
4. choose the solution with the smallest worst regret:
$$\min_{x \in X} \phi_{\text{Savage}}(x) = \min_{x \in X} \max_{\omega \in \Omega} \left[f(x, \omega) - \min_{x \in X} f(x, \omega) \right]$$

The main idea is to adopt a worst-case perspective, as in Wald criterium, but defining the worst case not as the one in which the cost is maximum, but as the one which maximises the difference between the cost incurred and the cost that could have been paid with a better choice. In other words, this criterium is relative and opportunistic, instead of absolute and conservative: it does not aim to lose as little as possible, but to keep as close as possible to the best alternative.

Table 9.6 reports the best alternative for each scenario and its cost.

$f(x,\omega)$	$\omega^{(1)}$	$\omega^{(2)}$	$\omega^{(3)}$	$\omega^{(4)}$
$x^{(1)}$	2	2	4	3
$x^{(2)}$	3	3	3	3
$x^{(3)}$	4	0	4	6
$x^{(4)}$	3	1	4	4
$x^*(\omega)$	$x^{(1)}$	$x^{(3)}$	$x^{(2)}$	$x^{(1)}$,$x^{(2)}$
$f(x^*(\omega), \omega)$	2	0	3	3

Table 9.6: Application of Savage criterium to a sample problem with four alternatives and four scenarios (cost minimisation): determination of the best alternative for each scenario

From this table, one can derive Table 9.7, that reports the regret associated to each alternative and scenario, that is the difference between the impact of each alernative and the best alternative for the same scenario. The last column of the table reports the worst-case regret for each alternative. Now we try to minimise that regret: the resulting order is $x^{(4)} \prec x^{(1)} \prec x^{(2)} \sim x^{(3)}$.

[5] Leonard Jimmie Savage, born Leonard Ogashevitz (1917-1971), American mathematician and statistician; his parents were Russian Jews fled from tsarist Russia.

$\rho(x,\omega)$	$\omega^{(1)}$	$\omega^{(2)}$	$\omega^{(3)}$	$\omega^{(4)}$	$\phi_{\text{Savage}}(x)$
$x^{(1)}$	0	2	1	0	$2\ (\omega^{(2)})$
$x^{(2)}$	1	3	0	0	$3\ (\omega^{(2)})$
$x^{(3)}$	2	0	1	3	$3\ (\omega^{(4)})$
$x^{(4)}$	1	1	1	1	$1\ (\omega^{(1)}, \omega^{(2)}, \omega^{(3)}, \omega^{(4)})$

Table 9.7: Application of Savage criterium to a sample problem with four alternatives and four scenarios (cost minimisation): computation of the regret for each alternative and scenario and determination of the maximum regret for each alternative

9.6 Surplus criterium

This criterium is complementary to the regret criterium: it considers in each scenario the surplus that the decision-maker obtains with respect to the worst alternative, and tries to maximise the surplus in the worst possible case. One proceeds therefore in a similar, but complementary, way:

1. determine the worst alternative for each scenario $x^{\dagger}(\omega)$

$$x^{\dagger}(\omega) = \arg\max_{x \in X} f(x,\omega)$$

2. evaluate the *surplus* $\sigma(x,\omega)$ associated to each alternative x and scenario ω, as the *gain obtained choosing x with respect to the worst alternative in scenario ω*:

$$\sigma(x,\omega) = f\left(x^{\dagger}(\omega),\omega\right) - f(x,\omega) = \max_{x \in X} f(x,\omega) - f(x,\omega)$$

3. assign to each solution the minimum surplus (worst case):

$$\phi_{\text{surplus}}(x) = \min_{\omega \in \Omega} \sigma(x,\omega)$$

4. choose the solution with the largest minimum surplus:

$$\max_{x \in X} \phi_{\text{surplus}}(x) = \max_{x \in X} \min_{\omega \in \Omega} \left[\max_{x \in X} f(x,\omega) - f(x,\omega)\right]$$

With respect to the other criteria considered above, and in particular the regret criterium, the main difference is that function σ does not measure a loss with respect to the best choice, but a “gain” with respect to the worst one (in the example, this gain is actually a smaller loss). Therefore, the worst scenario is that in which σ is minimum, whereas the worst scenario for the regret criterium was the one which maximised ρ. As the regret criterium, the surplus criterium does not use an absolute scale of cost, but a scale relative to a reference situation: in this case, one tries to keep as far as possible from the worst alternative.

Table 9.8 shows the worst alternative for each scenario and its cost.

Then, one computes how much the other alternatives save with respect to the worst alternative. Table 9.9 reports these values for each alternative and scenario. The last column of the table reports the surplus in the worst case for each alternative. All values are zero, which corresponds to the fact that each solution admits a scenario in which it is the worst one. Now, we should maximise this surplus, which implies that the four solutions are indifferent: $x^{(1)} \sim x^{(2)} \sim x^{(3)} \sim x^{(4)}$.

$f(x,\omega)$	$\omega^{(1)}$	$\omega^{(2)}$	$\omega^{(3)}$	$\omega^{(4)}$
$x^{\dagger}(\omega)$	$x^{(3)}$	$x^{(2)}$	$x^{(1)}$, $x^{(3)}$, $x^{(4)}$	$x^{(3)}$
$f(x^{\dagger}(\omega),\omega)$	4	3	4	6

Table 9.8: Application of the surplus criterium to a sample problem with four alternatives and four scenarios (cost minimisation): determination of the worst alternative for each scenario

$\sigma(x,\omega)$	$\omega^{(1)}$	$\omega^{(2)}$	$\omega^{(3)}$	$\omega^{(4)}$	$\phi_{\text{surplus}}(x)$
$x^{(1)}$	2	1	0	3	0 ($\omega^{(3)}$)
$x^{(2)}$	1	0	1	3	0 ($\omega^{(2)}$)
$x^{(3)}$	0	3	0	0	0 ($\omega^{(1)}$, $\omega^{(3)}$, $\omega^{(4)}$)
$x^{(4)}$	1	2	0	2	0 ($\omega^{(3)}$)

Table 9.9: Application of the surplus criterium to a sample problem with four alternatives and four scenarios (cost minimisation): computation of the surplus for each alternative and scenario and determination of the maximum surplus for each alternative

9.7 Examples

**** DA TRADURRE ****

Consideriamo nel seguito una serie di esempi di applicazione dei sei criteri prima descritti, allo scopo di illustrare piccole varianti, come l'uso di valori parametrici per definire una strategia di decisione al variare di un dato incerto oppure il trattamento dell'incertezza nel caso continuo.

9.7.1 Installazione di un dispositivo

Una casa automobilistica deve decidere se installare su un nuovo modello un dispositivo di sicurezza del costo di 2500 € oppure non installarlo. Se il dispositivo non viene installato e si verifica un incidente, la casa automobilistica dovrà procedere all'installazione e ripagare le spese dell'incidente. Se installa il dispositivo, invece, dovrà pagare solo il costo dell'installazione, perché non sarà considerata responsabile di eventuali incidenti.

Il modello del problema prevede quindi due alternative:

- $x^{(1)}$: installare il dispositivo di sicurezza;
- $x^{(2)}$: non installarlo.

e due scenari:

- $\omega^{(1)}$: non si verifica alcun incidente;
- $\omega^{(2)}$: si verifica almeno un incidente.

A rigore, lo scenario $\omega^{(2)}$ si dovrebbe decomporre in un'intera gamma di sottoscenari, corrispondenti a incidenti di diversa gravità e con diversi costi. Per semplicità, però, fonderemo tutti questi sottoscenari in uno solo, introducendo un parametro C_I che misura il costo dell'incidente.

La Tabella 9.10 riporta la matrice di valutazione del problema: se la casa automobilistica installa il dispositivo, paga in entrambi gli scenari solo il costo fisso di installazione di 2500 €; se non lo installa, nello scenario migliore non paga nulla, mentre nello scenario peggiore paga il dispositivo più il costo C_I dell'incidente.

$f(x,\omega)$	$\omega^{(1)}$	$\omega^{(2)}$
$x^{(1)}$	2 500	2 500
$x^{(2)}$	0	$2\,500 + C_I$

Table 9.10: Matrice di valutazione per un problema di scelta riguardo l'installazione di un dispositivo di sicurezza (minimizzazione di costi)

Ipotizziamo di essere in condizioni di completa ignoranza, cioè di non saper assegnare alcuna probabilità a $\omega^{(1)}$ e $\omega^{(2)}$.

Criterio del pessimismo La Tabella 9.11 riporta la valutazione del problema con il criterio del caso pessimo. Poiché $2\,500 + C_I > 2\,500$, per minimizzare il costo nel caso pessimo conviene adottare l'alternativa $x^{(1)}$, cioè installare il dispositivo di sicurezza.

$f(x,\omega)$	$\omega^{(1)}$	$\omega^{(2)}$	$\phi_{\text{worst}}(x)$
$x^{(1)}$	2 500	2 500	2 500
$x^{(2)}$	0	$2\,500 + C_I$	$2\,500 + C_I$

Table 9.11: Applicazione del criterio del caso pessimo al problema di installazione

Criterio dell'ottimismo La Tabella 9.12 riporta la valutazione del problema con il criterio del caso ottimo. Questo consiglia di adottare l'alternativa $x^{(2)}$, cioè di non installare il dispositivo di sicurezza. Infatti, se non capitano incidenti, il costo da sostenere è nullo.

$f(x,\omega)$	$\omega^{(1)}$	$\omega^{(2)}$	$\phi_{\text{best}}(x)$
$x^{(1)}$	2 500	2 500	2 500
$x^{(2)}$	0	$2\,500 + C_I$	0

Table 9.12: Applicazione del criterio del caso ottimo al problema di installazione

Criterio di Hurwicz La Tabella 9.13 riporta la valutazione del problema con il criterio di Hurwicz: poiché il decisore non ha specificato alcun valore per il coefficiente di pessimismo α, esso compare come parametro nel criterio.

$f(x,\omega)$	$\omega^{(1)}$	$\omega^{(2)}$	$\phi_{\text{Hurwicz}}(x)$
$x^{(1)}$	2 500	2 500	2 500
$x^{(2)}$	0	$2\,500 + C_I$	$(2500 + C_I)(1-\alpha)$

Table 9.13: Applicazione del criterio di Hurwicz con α parametrico al problema di installazione

Il costo dell'incidente C_I è un secondo parametro. La natura dei due parametri è molto diversa: α è in mano al decisore, mentre C_I è una variabile esogena generata dall'ambiente. Tuttavia, si possono trattare entrambe come variabili rispetto a cui condurre uno studio di sensitività.

La Figura 9.2 riporta il grafico delle due funzioni $\phi_{\text{Hurwicz}}\left(x^{(1)}\right)$ e $\phi_{\text{Hurwicz}}\left(x^{(2)}\right)$ rispetto ad α e C_I: il piano orizzontale blu rappresenta l'alternativa $x^{(1)}$, la superficie curva rossa l'alternativa $x^{(2)}$.

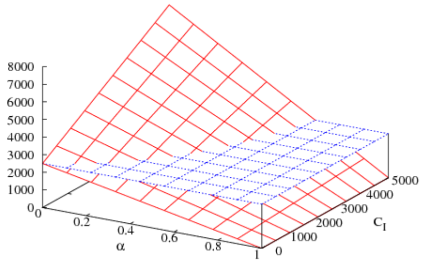


Figure 9.2: Studio di sensitività rispetto ad α e C_I per il problema di installazione: il piano orizzontale blu corrisponde all'installazione, la superficie curva rossa alla non installazione

Poiché $x^{(1)} \preceq x^{(2)}$ quando $\phi_{\text{Hurwicz}}\left(x^{(1)}\right) \leq \phi_{\text{Hurwicz}}\left(x^{(2)}\right)$, dalla figura si ricava che

$$x^{(1)} \preceq x^{(2)} \Leftrightarrow 2500 < (1-\alpha)(2500 + C_I) \Leftrightarrow C_I > \frac{\alpha}{1-\alpha} 2500$$

La Figura 9.3 mostra la proiezione sul piano αC_I del grafico in Figura 9.2: la zona colorata in blu contiene le coppie (α, C_I) in cui conviene adottare l'alternativa $x^{(1)}$, quella in rossof le coppie in cui conviene l'alternativa $x^{(2)}$; le due aree sono divise dalla curva $C_I = 2\,500\alpha/\left(1-\alpha\right)$, che tende all'infinito per $\alpha \rightarrow 0$ e poi va calando fino a zero.

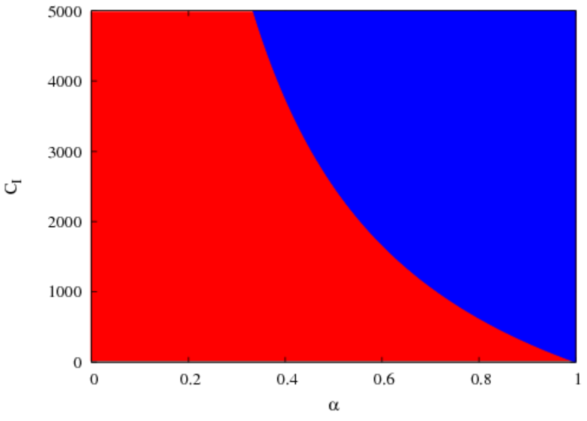


Figure 9.3: Studio di sensitività del problema di installazione rispetto ad α e C_I: supporto delle due alternative nello spazio dei parametri

In termini qualitativi:

- se il decisore ha un'alta propensione al rischio (valori bassi di α), gli conviene adottare la strategia $x^{(1)}$, cioè non installare il dispositivo, anche per valori alti (al limite, per qualsiasi valore) del costo di un eventuale incidente;

- se il decisore ha una bassa propensione al rischio (alti valori di α) gli conviene adottare la strategia $x^{(2)}$, cioè installare il dispositivo anche per stime basse del costo di un eventuale incidenti;
- il valore di soglia del costo C_I che divide le due soluzioni cresce più che linearmente al crescere della propensione al rischio.

Criterio di Laplace Si assume che la probabilità che accada un incidente sia uguale a quella che non accada alcun incidente. La Tabella 9.14 riporta i corrispondenti valori del criterio di Laplace. Se ne deduce che $x^{(1)} \preceq x^{(2)}$ quando $2\,500 \leq 1\,250 + C_I/2$, cioè $C_I \geq 2\,500$.

$f(x,\omega)$	$\omega^{(1)}$	$\omega^{(2)}$	$\phi_{\text{Laplace}}(x)$
$x^{(1)}$	2 500	2 500	2 500
$x^{(2)}$	0	2 500 + C_I	1 250+$C_I/2$

Table 9.14: Applicazione del criterio di Laplace al problema di installazione

Criterio del rammarico Il costo minimo e la corrispondente alternativa ottima per ciascuno dei due scenari sono riportati nell'ultima riga della Tabella 9.15.

$f(x,\omega)$	$\omega^{(1)}$	$\omega^{(2)}$
$x^{(1)}$	2 500	2 500
$x^{(2)}$	0	2 500 + C_I
$x^*(\omega)$	$x^{(2)}$	$x^{(1)}$
$f(x^*(\omega),\omega)$	0	2 500

Table 9.15: Applicazione del criterio del rammarico al problema di installazione: determinazione delle soluzioni ottime per ogni scenario

La funzione di rammarico $\rho(x,\omega)$ assume i valori riportati nella Tabella 9.16; l'ultima colonna riporta il rammarico nel caso pessimo per ogni soluzione.

$\rho(x,\omega)$	$\omega^{(1)}$	$\omega^{(2)}$	$\phi_{\text{regret}}(x)$
$x^{(1)}$	2 500	0	2 500
$x^{(2)}$	0	C_I	C_I

Table 9.16: Applicazione del criterio del rammarico al problema di installazione: calcolo della funzione rammarico e determinazione del caso pessimo

Anche in questo caso la soluzione non è univoca ma dipende dal costo dell'incidente. Poiché vogliamo minimizzare il rammarico nel caso peggiore, diremo che $x^{(1)} \preceq x^{(2)}$ quando $C_I \geq 2\,500$.

Criterio delle eccedenze L'alternative pessima e il relativo costo per ciascuno dei due scenari sono riportati nell'ultima riga della Tabella 9.17.

La funzione di eccedenza $\sigma(x,\omega)$ assume i valori riportati nella Tabella 9.18; l'ultima colonna riporta l'eccedenza nel caso pessimo per ogni soluzione. Si dovrebbe

$f(x,\omega)$	$\omega^{(1)}$	$\omega^{(2)}$
$x^{(1)}$	2 500	2 500
$x^{(2)}$	0	2 500 + C_I
$x^{\dagger}(\omega)$	$x^{(1)}$	$x^{(2)}$
$f\left(x^{\dagger}(\omega),\omega\right)$	2 500 + C_I	2 500

Table 9.17: Applicazione del criterio dell'eccedenza al problema di installazione: determinazione delle soluzioni pessime per ogni scenario

scegliere la soluzione che massimizza l'eccedenza nel caso pessimo, ma in questo caso le due soluzioni sono equivalenti da questo punto di vista: $x^{(1)} \sim x^{(2)}$.

$\sigma(x,\omega)$	$\omega^{(1)}$	$\omega^{(2)}$	$\phi_{\text{surplus}}(x)$
$x^{(1)}$	0	C_I	0
$x^{(2)}$	2500	0	0

Table 9.18: Applicazione del criterio dell'eccedenza al problema di installazione: calcolo della funzione eccedenza e determinazione del caso pessimo

9.7.2 Un investimento finanziario

Applichiamo ciascuna delle logiche su elencate all'esempio di investimento introdotto al principio del capitolo. In questo esempio gli impatti sono rendimenti, e vanno quindi massimizzati.

Criterio del caso pessimo Questo criterio dà luogo alla Tabella 9.19, secondo la quale si dovrebbe investire in titoli di stato.

		Crescita	Crescita		
$f(x,\omega)$	Recessione	moderata	forte	$\phi_{\text{worst}}(x)$	$\omega_{\text{worst}}(x))$
Azioni	-25%	0%	35%	-25%	Recessione
Obbligazioni	-10%	5%	15%	-10%	Recessione
Titoli di stato	8%	8%	8%	8%	Recessione

Table 9.19: Applicazione del criterio di Wald a un problema di investimento (massimizzazione di rendimenti)

Criterio del caso ottimo Questo criterio dà luogo alla Tabella 9.20, secondo la quale si dovrebbe investire in fondi azionari.

Criterio di Hurwicz Il criterio di Hurwicz con $\alpha = 0.6$, cioè con una visione bilanciata, ma leggermente più pessimista che ottimista, dà luogo alla Tabella 9.21, secondo la quale si dovrebbe investire in titoli di stato[6].

[6] Poiché l'impatto è un rendimento, il parametro di pessimismo $\alpha = 0.6$ va applicato ai valori peggiori, cioè ai più bassi.

$f(x,\omega)$	Recessione	Crescita moderata	Crescita forte	$\phi_{\text{worst}}(x)$	$\omega_{\text{worst}}(x))$
Azioni	-25%	0%	35%	35	Crescita forte
Obbligazioni	-10%	5%	15%	15	Crescita forte
Titoli di stato	8%	8%	8%	8	Crescita forte

Table 9.20: Applicazione del criterio dell'ottimismo a un problema di investimento (massimizzazione di rendimenti)

$f(x,\omega)$	Recessione	Crescita moderata	Crescita forte	$\phi_{\text{Hurwicz}}(x)$
Azioni	-25%	0%	35%	$-25\% \cdot 0.6 + 35\% \cdot 0.4 = -1\%$
Obbligazioni	-10%	5%	15%	$-10\% \cdot 0.6 + 15\% \cdot 0.4 = 0\%$
Titoli di stato	8%	8%	8%	$+8\% \cdot 0.6 + 8\% \cdot 0.4 = 8\%$

Table 9.21: Applicazione del criterio di Hurwicz con $\alpha = 0.6$ a un problema di investimento (massimizzazione di rendimenti)

Studio di sensitività rispetto al coefficiente di pessimismo Tornando sul criterio di Hurwicz, si può approfondire il discorso conducendo anche l'analisi di sensitività. Definiamo per ogni investimento possibile il rendimento, funzione del parametro α:

$$\begin{aligned} \phi_A(\alpha) &= -25\alpha + 35(1-\alpha) &&= 35 - 60\alpha \\ \phi_O(\alpha) &= -10\alpha + 15(1-\alpha) &&= 15 - 25\alpha \\ \phi_T(\alpha) &= 8\alpha + 8(1-\alpha) &&= 8 \end{aligned}$$

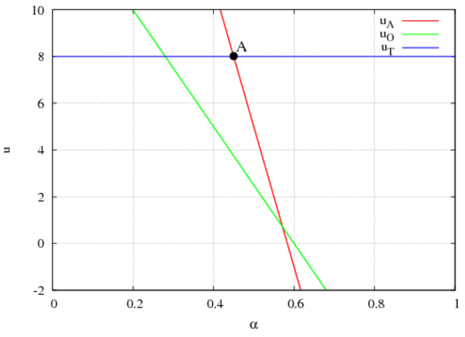


Figure 9.4: Studio di sensitività rispetto al coefficiente di pessimismo per il problema di investimento (massimizzazione di rendimenti)

La Figura 9.4 mostra i grafici delle tre funzioni. Per ogni $\alpha \in [0;1]$, bisogna scegliere la funzione con il valore massimo perché stiamo massimizzando un rendimento. Anche senza fare calcoli, si nota che le obbligazioni non sono mai convenienti, poiché la relativa retta è sempre superata da almeno una delle altre due. Questo non significa che esse siano una soluzione dominata: il fenomeno è analogo a quello delle soluzioni paretiane non supportate (la combinazione lineare non è sufficiente a far emergere soluzioni dominate dalle combinazioni convesse delle altre. Il fenomeno è poi ulteriormente aggravato dal fatto che il criterio di Hurwicz combina solo due degli scenari, mentre il metodo dei pesi combinava tutti gli indicatori. Per individuare l'investimento migliore in funzione di α, bisogna trovare il punto di intersezione A fra la retta delle azioni e quella dei titoli di stato: $\phi_A(\alpha) = \phi_T(\alpha) \Rightarrow 35 - 60\alpha = 8 \Rightarrow \alpha = 0.45$. Quindi, lo studio di sensitività conferma i risultati ottenuti con i metodi del pessimismo e dell'ottimismo: se si è ottimisti ($0 \leq \alpha \leq 0.45$), conviene investire in azioni; se si è pessimisti ($0.45 \leq \alpha \leq 1$), conviene investire in titoli di stato.

Criterio di Laplace Il criterio di Laplace dà luogo alla Tabella 9.22, che suggerisce, ancora una volta, di investire in titoli di stato.

$f(x, \omega)$	Recessione	Crescita moderata	Crescita forte	$\phi_{\text{Laplace}}(x)$
Azioni	-25%	0%	35%	$(-25\% + 0\% + 35\%)/3 = 3.33\%$
Obbligazioni	-10%	5%	15%	$(-10\% + 5\% + 15\%)/3 = 3.33\%$
Titoli di stato	8%	8%	8%	$(8\% + 8\% + 8\%)/3 = 8\%$

Table 9.22: Applicazione del criterio di Laplace a un problema di investimento (massimizzazione dei rendimenti)

Criterio del rammarico Il criterio del rammarico richiede anzi tutto di calcolare l'alternativa ottima per ogni scenario e il suo costo (vedi Tabella 9.23).

$f(x, \omega)$	Recessione	Crescita moderata	Crescita forte
Azioni	-25%	0%	35%
Obbligazioni	-10%	5%	15%
Titoli di stato	8%	8%	8%
$x^*(\omega)$	Titoli di stato	Titoli di stato	Azioni
$f(x^*(\omega), \omega)$	8%	8%	35%

Table 9.23: Applicazione del criterio di Savage a un problema di investimento (massimizzazione dei rendimenti): calcolo dell'alternativa ottima per ogni scenario

Quindi, richiede di valutare la funzione rammarico per ciascuna alternativa e scenario e trovarne il massimo per ciascuna alternativa (vedi Tabella 9.24). Investendo in azioni, la cosa peggiore che possa succedere è di andare in recessione, perdendo il 25% anziché guadagnare l'8%, con un rammarico del 33%. Investendo in obbligazioni, lo scenario peggiore è la crescita forte, che fa guadagnare il 15% anziché il 35%, con un rammarico del 20%; investendo in titoli di stato, lo scenario peggiore è la crescita forte, che fa guadagnare l'8% anziché il 35%, con un rammarico del 27%[7].

Infine, si può procedere alla scelta dell'alternativa che produce il rammarico minimo, cioè investire in obbligazioni.

Criterio delle eccedenze Il criterio delle eccedenze dà anzi tutto luogo alla Tabella 9.25. che riporta l'alternativa pessima per ogni scenario e il suo costo.

Quindi, il criterio genera la Tabella 9.26, che riporta l'eccedenza per ciascuna alternativa e scenario, e nell'ultima colonna l'eccedenza minima per ogni alternativa. Investendo in azioni, la cosa peggiore che possa succedere è che vi sia recessione o crescita moderata, nel qual caso l'eccedenza è nulla; investendo in obbligazioni, lo

[7] Tutte queste percentuali sono riferite all'ammontare dell'investimento iniziale.

$\rho\,(x,\omega)$	Recessione	Crescita moderata	Crescita forte	$\phi_{Savage}\,(x)$
Azioni	8%-(-25%) = 33%	8%-0% = 8%	35%-35% = 0%	33%
Obbligazioni	8%-(-10%) = 18%	8%-5% = 3%	35%-15% = 20%	20%
Titoli di stato	8%- 8% = 0%	8%-8% = 0%	35%- 8% = 27%	27%

Table 9.24: Applicazione del criterio di Savage a un problema di investimento (minimizzazione dei rendimenti): calcolo del rammarico per ogni alternativa e scenario

$f\,(x,\omega)$	Recessione	Crescita moderata	Crescita forte
Azioni	-25%	0%	35%
Obbligazioni	-10%	5%	15%
Titoli di stato	8%	8%	8%
$x^{\dagger}\,(\omega)$	Azioni	Azioni	Titoli di stato
$f(x^{\dagger}\,(\omega)\,,\omega)$	-25%	0%	8%

Table 9.25: Applicazione del criterio delle eccedenze a un problema di investimento (massimizzazione dei rendimenti): calcolo dell'alternativa pessima per ogni scenario

scenario pessimo è la crescita moderata, che fa guadagnare il 5% anziché lo 0%, con un'eccedenza del 5%; investendo in titoli di stato, lo scenario pessimo è la crescita forte, che fa guadagnare l'8%, con un'eccedenza nulla.

$\sigma(x,\omega)$	Recessione	Crescita moderata	Crescita forte	$\phi_{surplus}\,(x)$
Azioni	-25%-(-25%)= 0%	0%-0%= 0%	35%-8%= 27%	0%
Obbligazioni	-10%-(-25%)= 15%	5%-0%= 5%	15%-8%= 7%	5%
Titoli di stato	8%-(-25%)= 33%	8%-0%= 8%	8%-8%= 0%	0%

Table 9.26: Applicazione del criterio delle eccedenze a un problema di investimento (minimizzazione dei rendimenti): calcolo dell'eccedenza per ogni alternativa e scenario

Infine, si può procedere alla scelta dell'alternativa che produce l'eccedenza massima: investire in obbligazioni.

9.7.3 Un esempio continuo*

Consideriamo un problema nel quale la regione ammissibile X è un insieme continuo (un intervallo), mentre l'insieme degli scenari Ω rimane finito:

$$X = [10^{-9}, 3] \qquad \Omega = \{\omega^{(1)}, \omega^{(2)}, \omega^{(3)}\}$$

Poichè X è continuo non si può descrivere la funzione impatto $f(x,\omega)$ con una tabella. Si può però descriverla, come nella Tabella 9.27, specificando una funzione $f(x,\omega_i)$ per ogni $\omega_i \in \Omega$. Supponiamo che gli impatti rappresentino dei costi da minimizzare.

La Figura 9.5 mostra i grafici delle tre funzioni impatto. Poiché saranno utili in

*Questa è una sezione di approfondimento, che non fa parte del programma d'esame.

	$\omega^{(1)}$	$\omega^{(2)}$	$\omega^{(3)}$
$f(x,\omega)$	$\frac{1}{2}$	$\frac{1}{x}$	$-\frac{1}{2}x^2+2x$

Table 9.27: Valori dell'impatto $f\,(x,\omega)$ in ciascuno scenario $\omega_i \in \Omega$

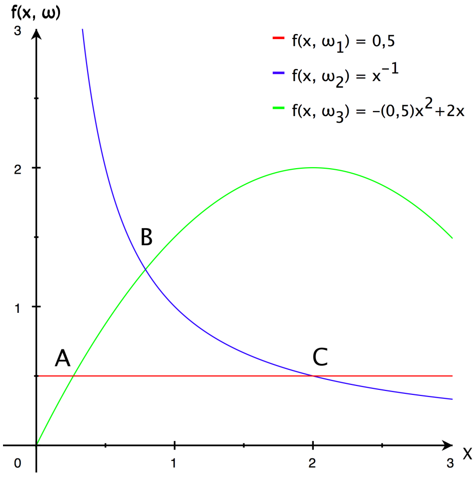


Figure 9.5: Andamento dell'impatto $f\,(x,\omega)$ in ciascuno scenario $\omega_i \in \Omega$

seguito, riportiamo anche le coordinate dei punti di intersezione delle tre curve:

$$A = \left(2-\sqrt{3}, \frac{1}{2}\right) \quad B = (0.79, 1.27) \quad C = \left(2, \frac{1}{2}\right)$$

Criterio del caso pessimo Per ogni $x \in X$ bisogna selezionare la funzione con valore massimo; della curva ottenuta bisogna selezionare il punto minimo. La Figura 9.6 mostra la soluzione suggerita dal criterio, che è $x^*_{\text{worst}} = x_B = 0.79$.

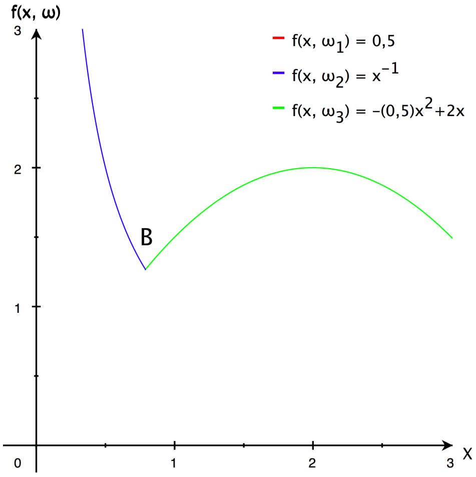


Figure 9.6: Applicazione del criterio del caso pessimo all'esempio continuo

Criterio del caso ottimo Per ogni $x \in X$ bisogna selezionare la funzione con valore minimo; della curva ottenuta bisogna selezionare il punto minimo. La Figura 9.7 mostra la soluzione suggerita dal criterio, che è in $x^*_{\text{best}} = 10^{-9}$.

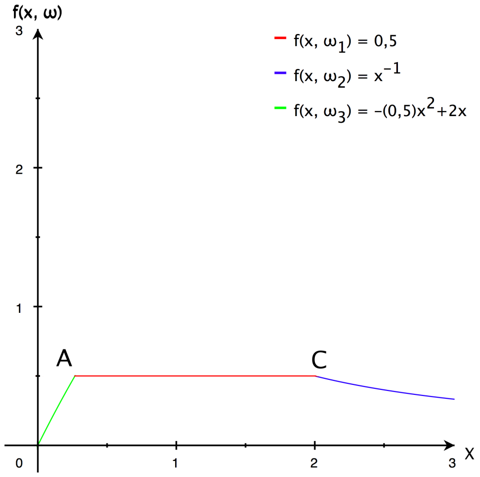


Figure 9.7: Applicazione del criterio del caso ottimo all'esempio continuo

Criterio di Hurwicz e studio di sensitività La funzione che esprime il criterio di Hurwicz è:

$$\phi_\Omega\,(x) = \alpha \min[f(x,\omega^{(1)}), f(x,\omega^{(2)}), f(x,\omega^{(3)})]+(1-\alpha) \max[f(x,\omega^{(1)}), f(x,\omega^{(2)}), f(x,\omega^{(3)})]s$$

La Figura 9.8 mostra l'applicazione del criterio di Hurwicz con diversi valori di α: le curve estreme rappresentano il criterio dell'ottimismo ($\alpha = 1$) e del pessimismo ($\alpha = 0$). Per ogni valore di α campionato viene segnata in neretto la soluzione del criterio, cioè il punto di minimo della curva. Per alcuni valori di α si trova una soluzione ($x^*_{\text{Hurwicz}} = 3$) diversa da quelle trovate con i criteri del pessimismo e dell'ottimismo.

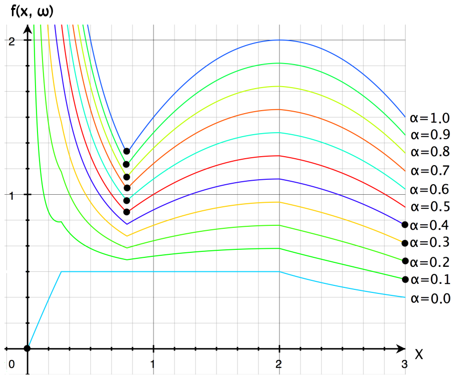


Figure 9.8: Applicazione del criterio di Hurwicz all'esempio continuo

La Tabella 9.28 riporta le soluzioni per ogni valore campionato di α.

α	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
$x^*_{\text{Hurwicz}}(\alpha)$	x_B	x_B	x_B	x_B	x_B	x_B	3	3	3	3	$\frac{1}{10^9}$

Table 9.28: Soluzioni ottime secondo il criterio di Hurwicz per ciascun valore campionato di α

Criterio di Laplace Il criterio consiste nel tracciare la curva media fra le tre date:

$$\phi_\Omega(x) = \frac{f(x,\omega^{(1)}) + f(x,\omega^{(2)}) + f(x,\omega^{(3)})}{3}$$

e calcolarne il minimo. La Figura 9.9 riporta l'andamento di tale curva in nero, evidenziandone il minimo, cioè $x^*_{\text{Laplace}} = 3$.

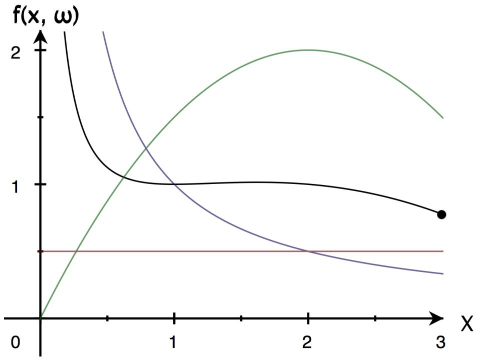


Figure 9.9: Applicazione del criterio di Laplace all'esempio continuo

Criterio del rammarico Per calcolare la funzione rammarico, occorre sottrarre a ciascuna funzione $f(x,\omega)$ il valore ottimo per lo scenario ω:

$$\rho(x,\omega) = f(x,\omega) - \min_{x\in X} f(x,\omega) \text{ per ogni } \omega \in \Omega$$

Le tre componenti della funzione rammarico sono riportate nella Tabella 9.29. La Figura 9.10 ne riporta i grafici.

	$\omega^{(1)}$	$\omega^{(2)}$	$\omega^{(3)}$
$f(x,\omega)$	$\frac{1}{2}$	$\frac{1}{x}$	$-\frac{1}{2}x^2 + 2x$
$f(x^*(\omega),\omega)$	$\frac{1}{2}$	$\frac{1}{3}$	10^{-18}
$\rho(x,\omega)$	0	$\frac{1}{x} - \frac{1}{3}$	$2x - \frac{1}{2}x^2$

Table 9.29: Valori della funzione rammarico per l'esempio continuo

Per concludere, bisogna individuare la curva del caso peggiore, cioè l'inviluppo superiore delle tre curve, ovvero i valori massimi per ogni $x \in X$. Nella figura, questa curva è ricalcata in nero. Il suo punto di ottimo è $x^*_{\text{Savage}} = x_D$.

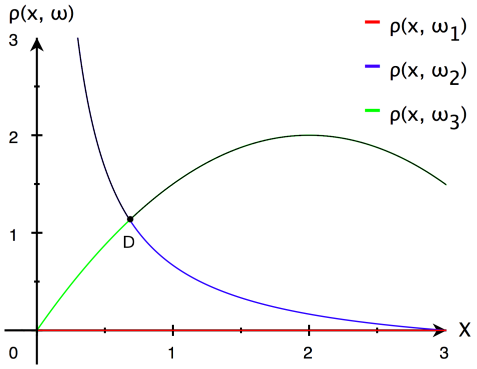


Figure 9.10: Andamento delle componenti della funzione rammarico per l'esempio continuo

Criterio delle eccedenze Per calcolare la funzione eccedenza, occorre sottrarre al valore pessimo per ciascuno scenario ω l'impatto $f\left(x, \omega\right)$:

$$\sigma(x, \omega) = \max_{x \in X} f(x, \omega) - f(x, \omega) \text{ per ogni } \omega \in \Omega$$

Le tre componenti della funzione eccedenza sono riportate nella Tabella 9.30. La Figura 9.11 ne riporta i grafici.

	$\omega^{(1)}$	$\omega^{(2)}$	$\omega^{(3)}$
$f\left(x^{\dagger}\left(\omega\right), \omega\right)$	$\frac{1}{2}$	10^9	2
$f(x, \omega)$	$\frac{1}{2}$	$\frac{1}{x}$	$-\frac{1}{2}x^2 + 2x$
$\sigma(x, \omega_i)$	0	$10^9 - \frac{1}{x}$	$\frac{1}{2}x^2 - 2x + 2$

Table 9.30: Valori della funzione eccedenza per l'esempio continuo

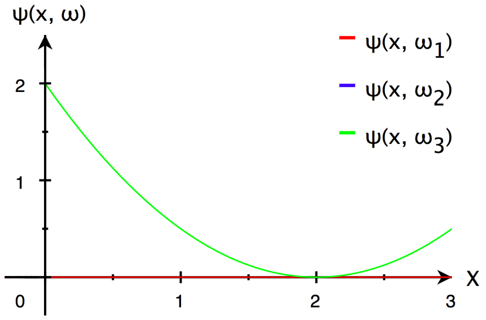


Figure 9.11: Andamento delle componenti della funzione eccedenza per l'esempio continuo

Per concludere, bisogna individuare la curva del caso peggiore, cioè l'inviluppo inferiore delle tre curve, ovvero i valori minimi per ogni $x \in X$. Questo inviluppo coincide con la componente $\sigma(x, \omega^{(1)}) = 0$. Tutte le soluzioni ammissibili $x \in X$ sono quindi indifferenti rispetto al criterio delle eccedenze.

9.8 Formal defects of the choice criteria

A satisfactory criterium should enjoy the following properties:

1. *weak ordering*: it should always induce a weak ordering on the alternatives;
2. *indipendence from the labelling*: the induced ordering should not change when the alternatives or the scenarios are permuted, or their names are changed[8];
3. *scale invariance*: the ordering should not change if the impact undergoes a linear transformation $f' = \alpha f + \beta$, with $\alpha > 0$ and any real value of β that is measuring the impact with a different scale (unit of measure and offset);
4. *strong dominance preservation*: if x dominates x', the criterium should prefer x to x':
$$x \preceq x' \Rightarrow \phi_\Omega(x) \leq \phi_\Omega(x')$$
5. *indipendence from irrelevant alternatives*: the ordering should not change if new alternatives are added or some alternatives are removed (*rank reversal*);
6. *indipendence from the duplication of scenarios*: the ordering should not change adding a new scenario identical to an given one;
7. *indipendence from the uniform variation of the impacts in a scenario*: the ordering should not change when the impacts associated to a scenario improve or worsen by the same amount for all alternatives.

Unfortunately, none of the choice criteria listed above respects all desirable properties. Even worse, no possible choice criterium can respect them. There is no rational basic reason to prefer a criterium to another one.

In particular, the criteria listed above enjoy the first four properties, but not all the other ones. For example, the first property automatically derives from the fact that the impact $f(x, \omega)$ ia turned into a function $\phi_\Omega(x)$ and the alternatives are ordered on the basis of the numerical values of that function (which necessarily implies a weak ordering). In the following, we consider in detail the last three properties, referring to the example with four alternatives and four scenarios discussed above.

9.8.1 Dependence from irrelevant alternatives

The criteria of pessimism, optimism, Hurwicz and Laplace enjoy this property. If a new alternative is added, in fact, it simply joins the ordering of the old alternatives, in a position determined by its score. The order of the old solutions does not change, because their score remains the same: the values of criterium $\phi_\Omega(\bar{x})$ for a specific alternative $\bar{x}$, in fact, depend only on the values of impact $f(\bar{x}, \omega)$ for that alternative. Table 9.31 reports the computation of these four criteria for the new problem. It is easy to realise that the values of the four original alternatives are identical to those reported in Tables 9.2, 9.3, 9.4 and 9.5, given that they are computed treating each row independently from the other ones.

The regret and the surplus criteria, on the contrary, violate this property, because the regret function $\rho(x, \omega)$ and the and the surplus function $\sigma(x, \omega)$ for each x also depend on the best alternative $x^*(\omega)$ and on the worst alternative $x^\dagger(\omega)$ in each scenario ω. Since the new alternative can be the best or the worst one in some

[8] Trivial examples of criteria that would violate this condition are: "choose the first alternative" and "choose the alternative that performs best in the first scenario".

	$\omega^{(1)}$	$\omega^{(2)}$	$\omega^{(3)}$	$\omega^{(4)}$	$\phi_{\text{worst}}(x)$	$\phi_{\text{best}}(x)$	$\phi_{\text{Hurwicz}}(x)$	$\phi_{\text{Laplace}}(x)$
$x^{(1)}$	2	2	4	3	4	2	3.2	2.75
$x^{(2)}$	3	3	3	3	3	3	3.0	3.00
$x^{(3)}$	4	0	4	6	6	0	3.6	3.50
$x^{(4)}$	3	1	4	4	4	1	2.8	3.00
$x^{(5)}$	6	4	0	4	6	0	3.6	3.50

Table 9.31: Application of the pessimism, optimism, Hurwicz and Laplace criteria to the sample problem after the introduction of a fifth alternative: the ordering of the first four ones does not change

scenarios, and not in other, the regret and the surplus of other alternative increase (they cannot decrease!) in these scenarios, whereas they remain the same in the other scenarios. Since the criterium considers the worst scenario with respect to the regret or the surplus, some values of the criteria ϕ_{Savage} and ϕ_{surplus} change, while others do not, yielding a potentially different ordering.

x	$\omega^{(1)}$	$\omega^{(2)}$	$\omega^{(3)}$	$\omega^{(4)}$
$x^{(1)}$	**2**	2	4	**3**
$x^{(2)}$	3	3	3	**3**
$x^{(3)}$	4	**0**	4	6
$x^{(4)}$	3	1	4	4
$x^{(5)}$	6	4	**0**	4

Table 9.32: Application of Savage criterium to the sample problem after the addition of a fifth alternative: the best alternative for scenario $\omega^{(3)}$ changes

Table 9.32 reports the costs of the problem extended with solution $x^{(5)}$; the best alternatives for each scenario are bolded. Comparing it with Table 9.6, we notice that $x^{(5)}$ has become the best alternative for scenario $\omega^{(3)}$, replacing $x^{(2)}$. Conseguently, the regret function changes (only in column $\omega^{(3)}$) and its maximum values row by row (that is the worst case for each alternative) change in an unpredictable way: some of them remain the same, others increase by the whole difference between the old and the new optimal value $f\left(x^*\left(\omega^{(3)}\right),\omega^{(3)}\right)$, others increase by intermediate values: Table 9.33 shows all this. The resulting order is $x^{(2)} \prec x^{(1)} \sim x^{(3)} \sim x^{(4)} \sim x^{(5)}$. Before adding alternative $x^{(5)}$, the ordering of the other four alternatives was $x^{(4)} \prec x^{(1)} \prec x^{(2)} \sim x^{(3)}$, that was different from the current one. For example, $x^{(1)}$ was preferable to $x^{(2)}$, whereas now the opposite occurs.

$\rho(x,\omega)$	$\omega^{(1)}$	$\omega^{(2)}$	$\omega^{(3)}$	$\omega^{(4)}$	$\phi_{\text{regret}}(x)$
$x^{(1)}$	0	2	4	0	4 ($\omega^{(3)}$)
$x^{(2)}$	1	3	3	0	3 ($\omega^{(2)}$, $\omega^{(3)}$)
$x^{(3)}$	2	0	4	3	4 ($\omega^{(3)}$)
$x^{(4)}$	1	1	4	1	4 ($\omega^{(3)}$)
$x^{(5)}$	4	1	0	1	**4** ($\omega^{(1)}$)

Table 9.33: Evaluation of the regret for the sample problem after the addition of a fifth alternative

As well, the surplus criterium violates this property. Table 9.34 reports the costs of the problem extended with solution $x^{(5)}$. The worst alternatives for each scenario are bolded. Let us notice that $x^{(5)}$ has become the worst alternative for scenario $\omega^{(1)}$, replacing $x^{(3)}$, and for scenario $\omega^{(2)}$, replacing $x^{(2)}$.

$f(x,\omega)$	$\omega^{(1)}$	$\omega^{(2)}$	$\omega^{(3)}$	$\omega^{(4)}$
$x^{(1)}$	2	2	**4**	3
$x^{(2)}$	3	3	3	3
$x^{(3)}$	4	0	**4**	**6**
$x^{(4)}$	3	1	**4**	4
$x^{(5)}$	**6**	**4**	0	4

Table 9.34: Application of the surplus criterium to the sample problem after the addition of a fifth alternative

Table 9.35 reports the new values of the surplus for each alternative and scenario and, in the last column, its value in the worst case for each alternative and the corresponding scenario. The first and second column of the matrix are changed, increasing by the difference between the new and the old worst value $f\left(x^{\dagger}(\omega_i),\omega_i\right)$ $(i=1,2)$. At first, the four alternatives were all indifferent, now alternative $x^{(2)}$ is better than the other ones.

$\sigma(x,\omega)$	$\omega^{(1)}$	$\omega^{(2)}$	$\omega^{(3)}$	$\omega^{(4)}$	$\phi_{\text{surplus}}(x)$
$x^{(1)}$	4	2	0	3	0 $(\omega^{(3)})$
$x^{(2)}$	3	1	1	3	1 $(\omega^{(2)},\ \omega^{(3)})$
$x^{(3)}$	2	4	0	0	0 $(\omega^{(3)},\ \omega^{(4)})$
$x^{(4)}$	3	3	0	2	0 $(\omega^{(3)})$
$x^{(5)}$	0	0	4	2	0 $(\omega^{(1)},\ \omega^{(2)})$

Table 9.35: Evaluation of the surplus for the sample problem after the addition of a fifth alternative

9.8.2 Dependence from the duplication of scenarios

Nearly all analysed criteria are immune to the duplication of scenarios, because the best and the worst scenarios, even if duplicated, remain the same, that is, the best and the worst values of the impact for each alternative do not change. If, however, one uses the equiprobability criterium, duplicating a scenario corresponds to (nearly) doubling its probability, and therefore its weight on the overall objective function. This modifies the values of the criterium $\phi_\Omega(x)$, and therefore can modify the ordering of the alternatives.

Table 9.36 reports an extension of the example in which scenario $\omega^{(2)}$ has been duplicated adding a new scenario ω_2' with identical impacts for each alternative. The ordering implied by Laplace criterium is now $x^{(1)} \sim x^{(4)} \preceq x^{(3)} \preceq x^{(2)}$, whereas it was $x^{(1)} \preceq x^{(2)} \sim x^{(4)} \preceq x^{(3)}$ before the duplication: alternative $x^{(3)}$ has improved, because its cost in scenario $\omega^{(2)}$ is particularly low, and therefore increasing the weight of such a scenario favours it with respect to the other solutions.

9.8.3 Dependence from uniform variations of a scenario

The pessimism and optimism criteria (and, consequently, the Hurwicz criterium) are affected by the uniform modification of all impacts in a given scenario. In fact,

	$\omega^{(1)}$	$\omega^{(2)}$	$\omega^{(3)}$	$\omega^{(4)}$	$\omega'^{(2)}$	ϕ_{Laplace}
$x^{(1)}$	2	2	4	3	2	2.6
$x^{(2)}$	3	3	3	3	3	3.0
$x^{(3)}$	4	0	4	6	0	2.8
$x^{(4)}$	3	1	4	4	1	2.6

Table 9.36: Application of Laplace criterium to a sample problem (cost minimisation) with four alternatives and five scenarios

such a modification can turn that scenario into the worst one, or the best one, for one of the alternatives, or (on the contrary) imply that the scenario is no longer the worst or the best one.

	$\omega^{(1)}$	$\omega^{(2)}$	$\omega^{(3)}$	$\omega^{(4)}$	ϕ_{worst}	ϕ_{best}	ϕ_{Hurwicz}	ϕ_{Laplace}
$x^{(1)}$	2	~~2~~ 6	4	3	~~4~~ 6	2	~~3.2~~ 4.4	~~2.75~~ 3.75
$x^{(2)}$	3	~~3~~ 7	3	3	~~3~~ 7	3	~~3.0~~ 5.4	~~3.00~~ 4.00
$x^{(3)}$	4	~~0~~ 4	4	6	6	~~0~~ 4	~~3.6~~ 5.2	~~3.50~~ 4.50
$x^{(4)}$	3	~~1~~ 5	4	4	~~4~~ 5	~~1~~ 3	~~2.8~~ 4.6	~~3.00~~ 4.00

Table 9.37: Application of the pessimism, optimism, Hurwicz and Laplace criteria to a sample problem with four alternatives and four scenarios (cost minimisation), in which all impacts associated to scenario $\omega^{(2)}$ increase by the same amount

Let us increase by a uniform value $\delta f = 4$ all the impacts associated to scenario $\omega^{(2)}$ for the example that we are taking into account. Table 9.37 reports the new value of the impact, as well as the values of the pessimis, optimism, Hurwicz (with $\alpha = 0.6$) and Laplace criteria. Some values are unchanged, others increase by $\delta f = 4$, others increase by values intermediate between 0 and δf. The orderings of the alternatives change as follows:

- according to the pessimism criterium, from $x^{(2)} \preceq x^{(1)} \sim x^{(4)} \preceq x^{(3)}$ to $x^{(4)} \preceq x^{(1)} \sim x^{(3)} \preceq x^{(2)}$;
- according to the optimism criterium, from $x^{(3)} \preceq x^{(4)} \preceq x^{(1)} \preceq x^{(2)}$ to $x^{(1)} \preceq x^{(2)} \preceq x^{(3)} \sim x^{(4)}$;
- according to Hurwicz criterium, from $x^{(4)} \preceq x^{(2)} \preceq x^{(1)} \preceq x^{(3)}$ to $x^{(1)} \preceq x^{(4)} \preceq x^{(3)} \preceq x^{(2)}$.

The ordering according to Laplace criterium, by contrast, does not change, because adding a constant δf to the impacts of a scenario corresponds to adding a uniform value $\delta f / |\Omega|$ for each $x \in X$ to the Laplace criterium $\phi_{\text{Laplace}}(x)$. In this case, the criterium increases by $\delta f / 4 = 1$ for all $x \in X$. Therefore, the ordering of the alternatives does not change.

Also the regret and the surplus criteria respect this property. In fact, both the regret function and the surplus function are defined as differences between impacts of different alternatives (the given one and the best one, or the worst one and the given one) in the same scenario. Since both impacts increase by the same constant amoung, the values of the regret and of the surplus remain the same. Let $\tilde{f}(x, \omega) = f(x, \bar{\omega}) + \delta$ be the new impact function. Then, the new regret function

$\tilde{\rho}(x,\omega)$ and the new surplus function $\tilde{\sigma}(x,\omega)$ coincide exactly with the old ones:

$$\begin{aligned}\tilde{\rho}(x,\omega) &= \tilde{f}(x,\omega) - \tilde{f}(x^*(\omega),\omega) = (f(x,\omega)+\delta) - (f(x^*(\omega),\omega)+\delta)\\ &= f(x,\omega) - f(x^*(\omega),\omega) = \rho(x,\omega)\\ \tilde{\sigma}(x,\omega) &= \tilde{f}\left(x^\dagger(\omega),\omega\right) - \tilde{f}(x,\omega) = \left(f(x^\dagger,\omega)+\delta\right) - (f(x(\omega),\omega)+\delta)\\ &= f(x^\dagger,\omega) - f(x(\omega),\omega) = \rho(x,\omega)\end{aligned}$$

9.9 Robust Programming*

**** DA TRADURRE ****

I concetti teorici presentati sinora nel caso finito si estendono quasi tutti banalmente ai casi in cui X e Ω sono infiniti (magari continui) oppure combinatorici, cioè finiti, ma di cardinalità assai elevata. Tuttavia, la soluzione pratica di tali problemi pone difficoltà molto serie: non è più possibile scorrere uno per uno gli scenari e le alternative applicando le definizioni. Spesso anche solo conoscere il valore di un criterio $\phi_\Omega(\cdot)$ in una prefissata alternativa $\bar{x}$ richiede di risolvere un problema di ottimizzazione, che può non essere banale. Ad esempio, il criterio del caso pessimo richiede di calcolare $\min_{\omega\in\Omega} f(x,\omega)$. Il criterio del rammarico richiede addirittura tre ottimizzazioni successive, di cui due parametriche:

1. calcolare la miglior soluzione per ogni scenario, $\max_{x\in X} f(x,\omega)$, che è un'ottimizzazione parametrica in $\omega \in \Omega$;
2. calcolare lo scenario con il rammarico massimo per ogni alternativa, $\max_{\omega\in\Omega} \rho(x,\omega)$, che è un'ottimizzazione parametrica in $x \in X$;
3. calcolare l'alternativa con il criterio minimo: $\min_{x\in X} \phi_{\text{Savage}}(x)$.

Recentemente, però, lo studio della Programmazione in Condizioni di Ignoranza per problemi di Ottimizzazione Combinatoria ha visto un grosso fermento. Trattandosi di un'altra comunità scientifica rispetto a quella dell'economia classica, il nome assegnato al filone di studi è diverso: si parla infatti di *Programmazione Robusta.*

9.9.1 Criteri adottati in Ottimizzazione Combinatoria

Lo studio della robustezza in Ottimizzazione Combinatoria si fonda, come quello della programmazione in condizioni di ignoranza, sul sostituire l'impatto $f(x,\omega)$ con funzioni ausiliarie che eliminino la dipendenza dalle variabili esogene ω. A causa del diverso ambito applicativo, le funzioni ausiliarie usate non sono esattamente le stesse. Alcune vengono a mancare, perché non avrebbero molto senso nello studio di problemi con applicazioni tecnologiche e con moltissime soluzioni. Ad esempio, tutti i criteri ottimisti, che considerano lo scenario ottimo per ogni alternativa o l'eccedenza rispetto all'alternativa pessima per ogni scenario, vengono considerati poco utili, perché i sistemi tecnologici richiedono garanzie forti di buon funzionamento, e quindi non ammettono l'affidarsi alla fortuna. Di conseguenza, non si ritrovano in Programmazione Robusta il criterio del caso ottimo, quello di Hurwicz e quello delle eccedenze. Il criterio di Laplace richiederebbe una media aritmetica su un numero esponenziale o su un insieme infinito di scenari, e quindi è computazionalmente impraticabile. I criteri residui si conservano, ma assumono

[*] This section provides advanced concepts, that are not part of the course's syllabus.

nomi diversi. Inoltre, sono stati proposti anche criteri del tutto nuovi. I più studiati sono:

1. la *robustezza assoluta*, che è il criterio del caso pessimo

$$\phi_{\mathrm{RA}}(x) = \max_{\omega \in \Omega} f(x, \omega)$$

2. la *deviazione robusta*, che è il criterio del rammarico

$$\phi_{\mathrm{DR}}(x) = \max_{\omega \in \Omega} \rho(x, \omega) = \max_{\omega \in \Omega} \left[f(x, \omega) - \min_{x \in X} f(x, \omega) \right]$$

3. la *robustezza relativa*, che valuta il rapporto fra rammarico e valore ottimo:

$$\phi_{\mathrm{RR}}(x) = \max_{\omega \in \Omega} \frac{\rho(x, \omega)}{\min\limits_{x \in X} f(x, \omega)} = \max_{\omega \in \Omega} \frac{f(x, \omega) - \min\limits_{x \in X} f(x, \omega)}{\min\limits_{x \in X} f(x, \omega)}$$

La robustezza assoluta esprime un atteggiamento molto conservativo: ottimizzarla significa difendersi dalle situazioni peggiori. La deviazione robusta e la robustezza relativa, invece, vedono l'incertezza anche come un'opportunità da sfruttare, con un approccio di tipo *benchmark*, che si confronta con il meglio che si poteva fare: la deviazione robusta misura quanto migliorerebbero in valore assoluto le prestazioni se si potesse eliminare l'incertezza, la robustezza relativa valuta tale miglioramento in rapporto al costo stesso, come se si trattasse di una perdita percentuale.

Example 68 *Per semplicità, confrontiamo le tre definizioni di robustezza su un esempio finito di piccole dimensioni, anziché combinatorico:* $X = \{x^{(1)}, x^{(2)}, x^{(3)}\}$ *e* $\Omega = \{\omega^{(1)}, \omega^{(2)}, \omega^{(3)}\}$*. La seguente tabella riporta i costi per ogni alternativa e scenario, nonché quelli delle alternative ottime per ogni scenario.*

	$\omega^{(1)}$	$\omega^{(2)}$	$\omega^{(3)}$
$x^{(1)}$	*8*	*10*	*9*
$x^{(2)}$	*2*	*14*	*7*
$x^{(3)}$	*12*	*12*	*1*
$f(x^*(\omega), \omega)$	*2*	*10*	*1*

Se ne ricava la robustezza assoluta $\phi_{\Omega}(x)$ *(in grassetto è evidenziato il valore della soluzione ottima).*

x	$\phi_{\mathrm{RA}}(x)$
$x^{(1)}$	***10***
$x^{(2)}$	*14*
$x^{(3)}$	*12*

Di seguito viene riportata la tabella del rammarico e i valori della deviazione robusta (in grassetto è evidenziata la soluzione ottima).

$\rho(x, \omega)$	$\omega^{(1)}$	$\omega^{(2)}$	$\omega^{(3)}$	$\phi_{\mathrm{DR}}(x)$
$x^{(1)}$	*6*	*0*	*8*	*8*
$x^{(2)}$	*0*	*4*	*6*	***6***
$x^{(3)}$	*10*	*2*	*0*	*10*

Per la robustezza relativa, il rammarico $\rho(x, \omega)$ viene rapportato al valore ottimo di ogni scenario. La tabella seguente riporta il rammarico "relativo" e i valori della robustezza relativa (la soluzione ottima è in grassetto).

	$\omega^{(1)}$	$\omega^{(2)}$	$\omega^{(3)}$	$\phi_{\text{RR}}(x)$
$x^{(1)}$	$6/2 = 3$	$0/10 = 0.0$	$8/1 = 8$	*8*
$x^{(2)}$	$0/2 = 0$	$4/10 = 0.4$	$6/1 = 6$	*6*
$x^{(3)}$	$10/2 = 5$	$2/10 = 0.2$	$0/1 = 0$	***5***

*Logiche diverse indicano soluzioni diverse: la robustezza assoluta indica $x^{(1)}$, la deviazione robusta $x^{(2)}$, la robustezza relativa $x^{(3)}$. Nel primo caso, importa il fatto di spendere il meno possibile (*10*); nel secondo caso, importa il fatto di spendere il meno possibile in più rispetto alla soluzione ottima (*6 *in più); nel terzo importa il fatto di spendere il meno possibile in più e in rapporto alla soluzione ottima ("solo"* 5 *volte di più, cioè il sestuplo).*

9.9.2 Risultati teorici di complessità

Come si è detto, nel caso combinatorico non è praticamente possibile applicare le definizioni per valutare i criteri di scelta. Questo non vuol dire che non esistano algoritmi più efficienti per ottimizzare tali criteri: vuol dire invece che occorre studiare volta per volta la complessità del problema. Si verifica che, partendo dallo stesso problema deterministico, a seconda del criterio scelto per trattare l'incertezza (caso pessimo, ottimo, ecc. . .) e della descrizione dell'incertezza scelta (a scenari o a intervalli, si veda la Sezione 8.2) si possono ottenere problemi di complessità molto diversa tra loro.

In linea generale, si osserva che:

1. la controparte robusta (ovvero incerta) di un problema è più complessa del problema originale (spesso è $\mathcal{NP}$-completa anziché polinomiale);

2. ottimizzare la deviazione robusta o la robustezza relativa è più complicato che ottimizzare la robustezza assoluta, dato che richiede di conoscere la soluzione ottima per ogni scenario;

3. la descrizione dell'incertezza a scenari dà luogo a problemi più complessi di quella a intervalli, dato che:

    - non tutte le combinazioni di valori ammissibili per le singole variabili esogene corrispondono a scenari ammissibili;
    - ogni soluzione può avere un diverso scenario pessimo, mentre nella descrizione a intervalli, spesso tutte le soluzioni condividono lo stesso scenario pessimo[9].

Nel seguito, discutiamo a mo' di esempio il problema del cammino minimo robusto, distinguendo la complessità e gli algoritmi risolutivi per i due approcci nella descrizione dell'incertezza (scenari e intervalli) e per le tre funzioni obiettivo (robustezza assoluta, deviazione robusta e robustezza relativa).

[9] Questo succede in particolare quando l'incertezza riguarda i coefficienti di una funzione di costo lineare: il caso pessimo, allora, è quello in cui tutti i coefficienti hanno il valore massimo.

Il problema del cammino minimo robusto: incertezza a scenari[*]

Consideriamo un grafo orientato $G = (N, A)$ con due nodi $s, t \in N$ e una funzione di costo $c : A \times \Omega \to \mathbb{N}$ definita sugli archi. Tale funzione di costo è incerta, cioè dipende non solo dall'arco, ma anche dallo scenario che si verificherà. Supponiamo di descrivere l'incertezza con scenari esplicitamente elencati. La versione deterministica del problema ammette notoriamente diversi algoritmi polinomiali.

Nel seguito mostreremo che la versione robusta del problema, con tutte e tre le funzioni obiettivo, è $\mathcal{NP}$-completa. La dimostrazione avviene riducendo un opportuno problema $\mathcal{NP}$-completo al problema del cammino minimo robusto, cioè mostrando che un algoritmo che risolvesse quest'ultimo risolverebbe necessariamente anche il primo.

Definition 38 *Dato un insieme N di oggetti e una funzione $c : S \to \mathbb{N}$ che associa un costo intero ad ogni oggetto, il* Partition Problem *richiede di determinare se sia possibile dividere N in due sottoinsiemi C e $S \setminus C$ che abbiano lo stesso costo totale:*

$$\exists C \subset S : \sum_{i \in C} c_i = \sum_{i \in S \setminus C} c_i \ ?$$

La Figura 9.12 riporta un insieme per cui non è possibile effettuare la partizione, mentre la Figura 9.13 riporta un esempio per cui la partizione è possibile.

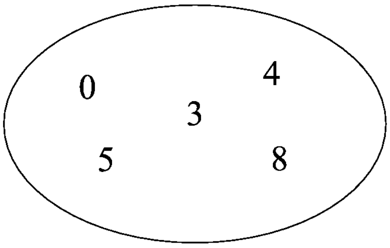


Figure 9.12: Insieme non partizionabile in due sottoinsiemi di ugual peso

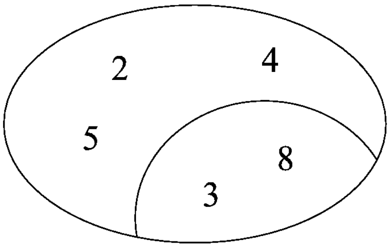


Figure 9.13: Insieme partizionabile in due sottoinsiemi di ugual peso

La Tabella 9.38 riporta tutti i possibili partizionamenti in due dell'insieme della Figura 9.13 e i costi dei due sottoinsiemi. La riga in grassetto mostra la partizione che soddisfa la proprietà richiesta, costituita dai due sottoinsiemi $\{3, 8\}$ e $\{2, 4, 5\}$.

Per ogni istanza del *Partition Problem*, si può costruire un'istanza del problema del cammino minimo robusto ad essa equivalente. Si consideri infatti un grafo $G = (N, A)$, in cui l'insieme dei nodi N contiene un nodo di partenza s, un nodo di

[*] Questa è una sezione di approfondimento, che non fa parte del programma d'esame.

Partizione	Costi
$\emptyset$ – $\{2,3,4,5,8\}$	0 – 22
$\{2\}$ – $\{3,4,5,8\}$	2 – 20
$\{3\}$ – $\{2,4,5,8\}$	3 – 19
$\{4\}$ – $\{2,3,5,8\}$	4 – 18
$\{5\}$ – $\{2,3,4,8\}$	5 – 17
$\{8\}$ – $\{2,3,4,5\}$	8 – 14
$\{2,3\}$ – $\{4,5,8\}$	5 – 17
$\{2,4\}$ – $\{3,5,8\}$	6 – 16
$\{2,5\}$ – $\{3,4,8\}$	7 – 15
$\{2,8\}$ – $\{3,4,5\}$	10 – 12
$\{3,4\}$ – $\{2,5,8\}$	7 – 15
$\{3,5\}$ – $\{2,4,8\}$	8 – 14
{3, 8} - {2, 4, 5}	**11-11**
$\{4,5\}$ – $\{2,3,8\}$	9 – 13
$\{4,8\}$ – $\{2,3,5\}$	12 – 10
$\{5,8\}$ – $\{2,3,4\}$	13 – 9

Table 9.38: Elenco delle partizioni in due sottoinsiemi dell'insieme di Figura 9.13

arrivo v_{n+1} e $n = |S|$ coppie di nodi v_i e v'_i con $i = 1, \ldots, n$, associate agli elementi dell'insieme S dato. Definiamo inoltre un insieme Ω con due scenari. Il nodo s è collegato al nodo v_1 da un arco il cui costo è 1 in entrambi gli scenari; ciascun nodo v_i $(i = 1, \ldots, n)$ è collegato al nodo v_{i+1} da un arco il cui costo è c_i nel primo scenario e 0 nel secondo e al nodo v'_i da un arco di costo 0 nel primo scenario e c_i nel secondo. Ogni nodo v'_i è collegato al nodo v_{i+1} da un arco di costo nullo. La Figura 9.14 mostra il grafo equivalente al problema di partizione della Figura 9.13.

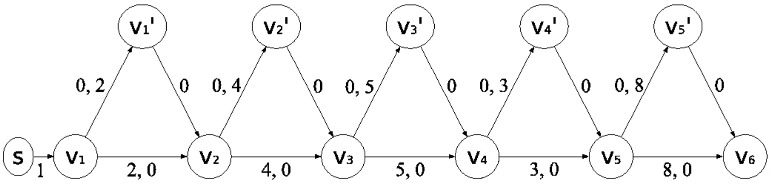


Figure 9.14: Grafo equivalente al problema partizione

Fissato uno scenario, il cammino minimo da s a v_{n+1} si ottiene in tempo $O(n)$, perché il grafo è aciclico e il numero di archi è proporzionale a quello dei nodi. È anche facile vedere che ogni cammino da s a v_{n+1} ha nel primo scenario un costo pari a 1 più la somma di alcuni c_i $(1 + \sum_{i \in C} c_i)$, mentre nel secondo ha un costo pari a 1 più la somma degli altri c_i $(1 + \sum_{i \in N \setminus C} c_i)$. Se si conoscesse lo scenario in anticipo, si potrebbe seguire sempre l'arco di costo nullo, e quindi ottenere un cammino di costo 1, ma il costo di tale cammino nell'altro scenario è pari a $1 + \sum_{i \in S} c_i$. Il numero di possibili soluzioni è 2^n; infatti ci sono n nodi v_i in ciascuno dei quali si può scegliere tra due strade.

Si può osservare che, in realtà, il numero di alternative interessanti è 2^{n-1}, dato che metà delle soluzioni sono speculari all'altra metà: ogni soluzione x_j $(j = 1, \ldots, 2^{n-1})$ ha una soluzione complementare x_{j_c} in cui i valori per i due scenari sono invertiti e in cui in ogni nodo v_i $(i = 1, \ldots, n)$ viene selezionato l'arco uscente che non è stato visitato in x_j. Questo fatto si nota meglio nel *Partition Problem* originale, in cui non avevamo fatto distinzione tra il sottoinsieme C e il sottoinsieme

$N \setminus C$ e, infatti, avevamo elencato $2^{(5-1)} = 16$ possibili partizioni.

La Tabella 9.39 riporta i $2^{(5-1)} = 16$ cammini interessanti, con i relativi costi nei due scenari. In particolare:

- $x^{(1)}$ è l'alternativa ottima per lo scenario $\omega^{(1)}$;
- $x^{(2)}$ è un'alternativa in cui il costo è lo stesso per i due scenari $\omega^{(1)}$ e $\omega^{(2)}$;
- altre quattordici alternative, $x^{(3)}, \ldots, x_{16}$.

All'alternativa $x^{(1)}$ ottima per lo scenario $\omega^{(1)}$ corrisponde l'alternativa speculare $x'^{(1)} = (s - v_1 - v_2 - v_3 - v_4 - v_5 - v_6)$, che costa 23 nel primo scenario e 1 nel secondo, e quindi è ottima per lo scenario $\omega^{(2)}$. All'alternativa $x^{(2)}$ corrisponde l'alternativa speculare $x'^{(1)} = (s - v_1 - v'_1 - v_2 - v'_2 - v_3 - v'_3 - v_4 - v_5 - v_6)$, che ha anch'essa costo 12 in entrambi gli scenari.

X	$\omega^{(1)}$	$\omega^{(2)}$	$\phi_{RA}(x)$
$x^{(1)} = (s - v_1 - v'_1 - v_2 - v'_2 - v_3 - v'_3 - v_4 - v'_4 - v_5 - v'_5 - v_6)$	1	23	23
$x^{(2)} = (s - v_1 - v_2 - v_3 - v_4 - v'_4 - v_5 - v'_5 - v_6)$	12	12	**12**
$x^{(3)} = (s - v_1 - v_2 - v'_2 - v_3 - v'_3 - v_4 - v'_4 - v_5 - v'_5 - v_6)$	3	21	21
$x^{(4)} = (s - v_1 - v'_1 - v_2 - v'_2 - v_3 - v'_3 - v_4 - v_5 - v'_5 - v_6)$	4	20	20
$x^{(5)} = (s - v_1 - v'_1 - v_2 - v_3 - v'_3 - v_4 - v'_4 - v_5 - v'_5 - v_6)$	5	19	19
$x_6 = (s - v_1 - v'_1 - v_2 - v'_2 - v_3 - v_4 - v'_4 - v_5 - v'_5 - v_6)$	6	18	18
$x_7 = (s - v_1 - v'_1 - v_2 - v'_2 - v_3 - v'_3 - v_4 - v'_4 - v_5 - v_6)$	9	15	15
$x_8 = (s - v_1 - v_2 - v'_2 - v_3 - v'_3 - v_4 - v_5 - v'_5 - v_6)$	6	18	18
$x_9 = (s - v_1 - v_2 - v_3 - v'_3 - v_4 - v'_4 - v_5 - v'_5 - v_6)$	7	17	17
$x_{10} = (s - v_1 - v_2 - v'_2 - v_3 - v_4 - v'_4 - v_5 - v'_5 - v_6)$	8	16	16
$x_{11} = (s - v_1 - v_2 - v'_2 - v_3 - v'_3 - v_4 - v'_4 - v_5 - v_6)$	11	13	13
$x_{12} = (s - v_1 - v'_1 - v_2 - v_3 - v'_3 - v_4 - v_5 - v'_5 - v_6)$	8	16	16
$x_{13} = (s - v_1 - v'_1 - v_2 - v'_2 - v_3 - v_4 - v_5 - v'_5 - v_6)$	9	15	15
$x_{14} = (s - v_1 - v'_1 - v_2 - v_3 - v_4 - v'_4 - v_5 - v'_5 - v_6)$	10	14	14
$x_{15} = (s - v_1 - v'_1 - v_2 - v_3 - v'_3 - v_4 - v'_4 - v_5 - v_6)$	13	11	13
$x_{16} = (s - v_1 - v'_1 - v_2 - v'_2 - v_3 - v_4 - v'_4 - v_5 - v_6)$	14	10	14
$\vdots$	$\vdots$	$\vdots$	

Table 9.39: Elenco delle soluzioni interessanti del problema di cammino minimo di Figura 9.14

Vediamo in dettaglio che relazione c'è fra un cammino nel grafo e una partizione dell'insieme: in ogni cammino i pesi diversi da zero che vengono selezionati nello scenario $\omega^{(1)}$ sono quelli che appartengono al primo sottoinsieme della partizione associata; quelli che vengono selezionati con lo scenario $\omega^{(2)}$, invece, sono quelli che appartengono al secondo sottoinsieme. Ad esempio la partizione $\{5, 8\} - \{2, 3, 4\}$ corrisponde al cammino x_{16}. Notiamo che ogni alternativa contiene in entrambi gli scenari l'arco (s, v_1), di peso 1, che non corrisponde ad alcun elemento nel problema di partizione. Tale arco serve solo per fare in modo che l'alternativa ottima abbia sempre costo maggiore di zero, così da evitare problemi nel denominatore della robustezza relativa. Questo spiega perché il costo di ogni cammino supera di un'unità il peso del sottoinsieme corrispondente nel problema della partizione.

Calcoliamo i tre criteri di robustezza per il problema di cammino minimo. Il valore $\phi_{RA}(x)$ della robustezza assoluta per ogni alternativa è riportato nell'ultima colonna della Tabella 9.39. L'alternativa migliore per la robustezza assoluta è $x^{(2)}$

f	$x^{(1)}$	$x^{(2)}$	$x^{(3)}$	$x^{(4)}$	$x^{(5)}$	x_6	x_7	x_8	x_8	x_{10}	x_{11}	x_{12}	x_{13}	x_{14}	x_{15}	x_{16}
$\omega^{(1)}$	0	11	2	3	4	5	8	5	6	7	10	7	8	9	12	13
$\omega^{(2)}$	22	11	20	19	18	17	14	17	16	15	12	15	14	13	10	9
ϕ_{DR} ϕ_{RR}	22	**11**	20	19	18	17	14	17	16	15	12	15	14	13	12	13

Table 9.40: Valori della deviazione robusta e della robustezza relativa per le soluzioni interessanti del problema di cammino minimo

(oppure l'alternativa complementare $x'^{(2)}$). Per costruzione, in questo specifico esempio la deviazione robusta e la robustezza relativa di ciascun cammino coincidono, poiché l'alternativa ottima costa 1 in ogni scenario, e quindi il denominatore presente nella robustezza relativa vale sempre 1. Possiamo, quindi, usare la funzione rammarico per minimizzare entrambi i criteri (vedi Tabella 9.40).

L'ultima riga della tabella riporta i valori della deviazione robusta e della robustezza relativa. L'alternativa che minimizza tale valore è $x^{(2)}$ (oppure la complementare $x'^{(2)}$). Quindi le tre funzioni obiettivo hanno lo stesso ottimo, che corrisponde a un cammino che distribuisce nel modo più equo i costi c_i. Un algoritmo polinomiale che ottimizzasse questo obiettivo sarebbe in grado di dire in tempo polinomiale se l'insieme dei c_i è divisibile in due parti uguali. Se nella soluzione ottima i due scenari non hanno ugual costo, allora l'insieme di partenza non è partizionabile in sottoinsiemi di ugual peso.

Il fatto che il *Partition Problem* sia $\mathcal{NP}$-completo, ma non fortemente $\mathcal{NP}$-completo lascia aperta la via a un algoritmo pseudopolinomiale. In effetti, esiste un algoritmo di programmazione dinamica che risolve il problema considerando nello spazio degli stati tutte le possibili durate dei cammini in ognuno dei due scenari. Se il numero di scenari è illimitato, però, il problema è fortemente $\mathcal{NP}$-completo, per riduzione dal 3-*Partition Problem.*

Il problema del cammino minimo robusto: incertezza a intervalli[10]

Analizziamo ora il caso in cui l'incertezza sul costo di percorrenza di un arco (i, j) viene descritta a intervalli, cioè $c_{ij} \in [l_{ij}; \phi_{ij}]$. La Figura 9.15 mostra un grafo di esempio.

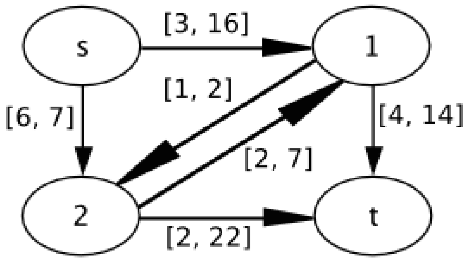


Figure 9.15: Problema del cammino minimo robusto, con incertezza descritta a intervalli

Ottimizzare la robustezza assoluta è un problema polinomiale, perché lo scenario pessimo è lo stesso per tutte le soluzioni ammissibili: si tratta dello scenario $\omega^{\dagger}$ in

[10] Questo è un argomento avanzato: non fa parte del programma d'esame. Vi si addentri chi è interessato a un approfondimento.

cui il costo di ciascun arco (i, j) è pari a ϕ_{ij}. Quindi basta applicare un classico algoritmo per il cammino minimo, assegnando a ogni arco il costo massimo ϕ_{ij}. La Tabella 9.41 riporta la robustezza assoluta di tutti i cammini semplici da s a t, evidenziando in grassetto l'alternativa prescelta.

x	$\phi_{RA}(x)$
$s-1-t$	$16+14=30$
$s-2-t$	$7+22=29$
$s-1-2-t$	$16+2+22=40$
$s-2-1-t$	$7+7+14=\mathbf{28}$

Table 9.41: Valori della robustezza assoluta per le soluzioni del problema di cammino minimo della Figura 9.15

Minimizzare la deviazione robusta è invece $\mathcal{NP}$-difficile, anche nel caso in cui il grafo sia orientato e aciclico[11]. La deviazione robusta di un cammino x in uno scenario ω è il rammarico, cioè la differenza tra il costo $f(x, \omega)$ di x in ω e il costo $f(x^*(\omega), \omega)$ del cammino minimo nello scenario ω. È stato dimostrato in letteratura[12] che la deviazione robusta è massima nello scenario in cui i costi degli archi sono:

- $c_{ij} = \phi_{ij}$ se $(i, j) \in x$,
- $c_{ij} = l_{ij}$ se $(i, j) \notin x$.

Questo vuol dire che, nonostante l'insieme dei possibili scenari sia infinito e continuo, occorre valutare solo quelli in cui gli archi assumono valori estremi di costo, che sono in numero finito, benché esponenziale. Anzi, gli scenari potenzialmente interessanti sono solo quelli in cui gli archi di un cammino da s a t assumono il costo minimo e tutti gli altri il costo massimo.

Una possibile formulazione del problema è:

$$\min \sum_{(i,j)\in A} \phi_{ij} x_{ij} - d_t \tag{9.1}$$

$$\sum_{(i,j)\in A} x_{ij} - \sum_{(j,k)\in A} x_{jk} = \begin{cases} -1 \text{ per } j = s \\ 1 \text{ per } j = t \\ 0 \text{ altrimenti} \end{cases} \tag{9.2}$$

$$d_s = 0 \tag{9.3}$$

$$d_j \leq d_i + l_{ij} + (\phi_{ij} - l_{ij}) x_{ij} \qquad (i, j) \in A \tag{9.4}$$

$$x_{ij} \in A \tag{9.5}$$

$$d_i \geq 0 \tag{9.6}$$

$$x_{ij} \in \{0, 1\} \qquad (i, j) \in A \tag{9.7}$$

dove le variabili binarie x_{ij} identificano gli archi del cammino robusto, mentre d_t è il costo del cammino descritto dalle x_{ij} nello scenario pessimo. Ai vincoli classici imposti sulle variabili x_{ij} perché descrivano un cammino si aggiungono i nuovi vincoli, che impongono al costo effettivo di ogni arco (i, j), cioè alla differenza $d_j - d_i$, di essere pari a l_{ij} se l'arco non fa parte del cammino descritto dalle x e pari a ϕ_{ij} se ne fa parte.

[11] Averbakh I, Lebedev V Interval data minmax regret network optimization problems. *Discrete Applied Mathematics* 138: 289–301, 2004

[12] Karaşan OE, Pinar MÇ, Yaman H The robust shortest path problem with interval data. *Operations Research Letters* 29: 31–40, 2001.

Effettuiamo il calcolo sull'esempio precedente. Poiché ci sono quattro cammini, dobbiamo valutare quattro scenari. Il $\sum_{(i,j)\in A} \phi_{ij} x_{ij}$ ci viene fornito dalla tabella della robustezza assoluta.

Per il cammino $(s - 1 - t)$ lo scenario che dobbiamo considerare è ω_{s-1-t} che viene mostrato in Figura 9.16.

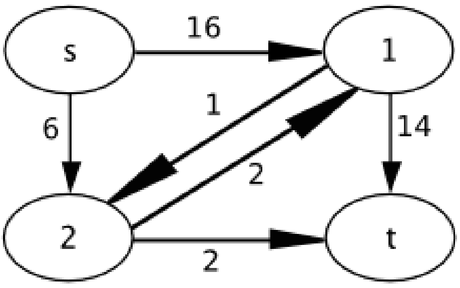


Figure 9.16: Scenario ω_{s-1-t}

Il cammino minimo per lo scenario ω_{s-1-t} è $(s - 2 - t)$ il cui costo è $6 + 2 = 8$. Quindi il valore della deviazione robusta per $s - 1 - t$ è $DR(s - 1 - t) = RA(s - 1 - t) - 8 = 30 - 8 = 22$. Per il cammino $(s - 2 - t)$ lo scenario che dobbiamo considerare è ω_{s-2-t} che viene mostrato in Figura 9.17.

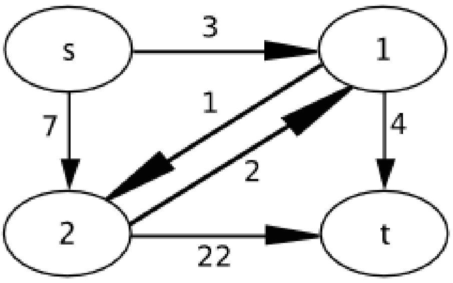


Figure 9.17: Scenario ω_{s-2-t}

Il cammino minimo per lo scenario ω_{s-2-t} è $(s - 1 - t)$ il cui costo è $3 + 4 = 7$. Quindi $DR(s-2-t) = RA(s-2-t)-7 = 29-7 = 22$. Per il cammino $(s-1-2-t)$ lo scenario che dobbiamo considerare è $\omega_{s-1-2-t}$ che viene mostrato in Figura 9.18.

Il cammino minimo per lo scenario $\omega_{s-1-2-t}$ è $(s-1-t)$ il cui costo è $16+4 = 20$. Quindi $DR(s - 1 - 2 - t) = RA(s - 1 - 2 - t) - 20 = 40 - 20 = 20$. Per il cammino $(s - 2 - 1 - t)$ lo scenario che dobbiamo considerare è $\omega_{s-2-1-t}$ che viene mostrato in Figura 9.19.

Il cammino minimo per lo scenario $\omega_{s-2-1-t}$ è $(s - 1 - 2 - t)$ il cui costo è $3 + 1 + 2 = 6$. Quindi $DR(s - 2 - 1 - t) = RA(s - 2 - 1 - t) - 6 = 28 - 6 = 22$. La Tabella 9.42 riassume i risultati: il cammino con deviazione robusta minima è $x^* = (s - 1 - 2 - t)$.

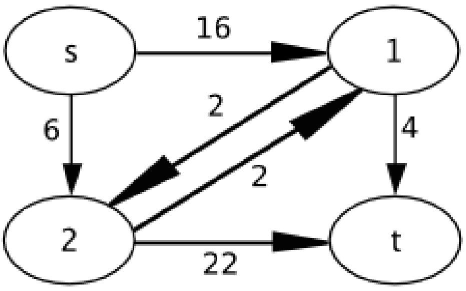


Figure 9.18: Scenario $\omega_{s-1-2-t}$

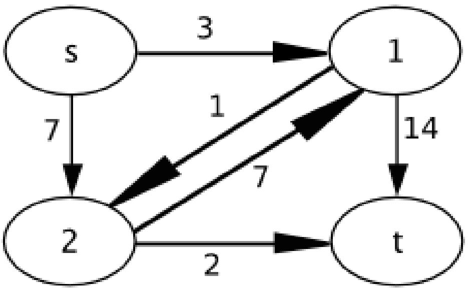


Figure 9.19: Scenario $\omega_{s-2-1-t}$

x	$\phi_{DR}(x)$
$(s-1-t)$	22
$(s-2-t)$	22
$(s-1-2-t)$	**20**
$(s-2-1-t)$	22

Table 9.42: Valutazione della deviazione robusta per il problema del cammino minimo con incertezza descritta a intervalli

Chapter 10

Programming in conditions of risk

With respect to the programming in conditions of ignorance, this model includes not only the possible scenario set Ω, but also a formalisation of their probability. For each given solution x, the associated impact $f(x, \omega)$ is a *random variable* depending on the scenario ω. We assume then to know:

- in the discrete case, a function π_ω assigning to each scenario a *probability* value
$$\pi : \Omega \rightarrow [0; 1] \text{ with } \sum_{\omega \in \Omega} \pi_\omega = 1$$
- in the continuous case, a function $\pi(\omega)$ assigning to each scenario a *probability density* value
$$\pi : \Omega \rightarrow \mathbb{R}^+ \text{ with } \int_\Omega \pi(\omega)\, d\omega = 1$$

For some brief recalls of probability theory, see Appendix C.

As in the programming in conditions of ignorance, the main approaches consist in reducing the problem to the optimisation of an auxiliary function $\phi(x)$ (to be minimised or maximised), which removes the dependency on the scenario ω. Thanks to the stronger assumptions, however, it is possible to design decision criteria which enjoy the properties expected from a rational approach.

10.1 Definitions of probability

All the numerical values and the functions that appear in decision problems are the result of descriptive models: the values of the impacts, the weights of the indicators, the coefficients used in the various choice criteria, and so on. Of course, this introduces the problem of the reliability of such values, and therefore of the choice that derives from them. This also holds for the probability functions, with the additional complication that the concept itself of probability is extremely problematic, and in fact can be defined in completely different ways. As in the decision problems all definitions of probability are commonly applied, it is appropriate to be aware of their nature and limits.

The *classical* definition indicates as probability the ratio between the number of elementary cases that form a scenario and the total number of possible elementary cases. This definition assumes that there exists a finite number of elementary cases with the same probability, and typically is applied to simple examples, such as games.

The *frequentist* definition indicates as probability of a scenario the limit to which its relative frequency tends as the number of observations, or experiments, increases. This definition requires an archive of the scenarios that one wants to model, assumes that such an archive be sufficiently large to provide a good approximation of the limit frequencies of the scenarios, and that the frequencies of past scenarios be the same that will occur in the future, after applying the decision. It is not very fit to describe scenarios for which past examples are missing or not very significant.

The *subjective* definition indicates as probability of a scenario the price an individual would consider fair to pay in order to receive 1 if the scenario occurs and 0 if it does not. This definition is clearly related to decision problems, as it is based on the idea of having a gain or not, and it is typically used in economics and finance. It provides values that vary from individual to individual.

The *axiomatic* definition indicates as probability a function that associates values to scenarios respecting suitable axioms. The theory that derives from it is perfectly consistent, but the definition does not indicate how to obtain the values in practical applications.

10.2 Expected value criterium

The first investigations on the decision problems in conditions of risk were done in the Seventeenth century by Pascal[1], and were based on the compact representation of the impact $f(x,\omega)$ of an alternative x in the various scenarios $\omega \in \Omega$ with its expected value, defined as:

$$\phi_{EV}(x) = E[f(x,\omega)] = \begin{cases} \sum\limits_{\omega\in\Omega} \pi_\omega f(x,\omega) & \text{in the discrete case} \\ \int_\Omega \pi(\omega) f(x,\omega)\, d\omega & \text{in the continuous case} \end{cases}$$

In the finite case, this corresponds to the product of the evaluation matrix times the probability vector:

$$\phi_{EV} = U \cdot \pi$$

Example 69 *Consider the following decision problem in conditions of risk: there are four alternatives and four scenarios. Table 10.1 reports the evaluation matrix with the impacts (that are costs) for each alternative and each scenario, but in addition it also reports the probability that each scenario occurs, that is vector π.*

The expected value criterium requires to compute the expected value for each alternative $x \in X$:

$$\phi_{EV} = U \cdot \pi \Rightarrow \begin{cases} \phi_{EV}\left(x^{(1)}\right) = 1 \cdot 0.20 + 3 \cdot 0.25 + 4 \cdot 0.50 + 6 \cdot 0.05 = 3.25 \\ \phi_{EV}\left(x^{(2)}\right) = 2 \cdot 0.20 + 2 \cdot 0.25 + 2 \cdot 0.50 + 4 \cdot 0.05 = 2.10 \\ \phi_{EV}\left(x^{(3)}\right) = 3 \cdot 0.20 + 2 \cdot 0.25 + 1 \cdot 0.50 + 9 \cdot 0.05 = 2.05 \\ \phi_{EV}\left(x^{(4)}\right) = 6 \cdot 0.20 + 6 \cdot 0.25 + 1 \cdot 0.50 + 3 \cdot 0.05 = 3.35 \end{cases}$$

This yields the ordering $x^{(3)} \prec x^{(2)} \prec x^{(1)} \prec x^{(4)}$, which leads to choose alternative $x^{(3)}$.

[1] Blaise Pascal (1623-1662), French mathematician, physicist and philosopher.

$f(x,\omega)$	$\omega^{(1)}$	$\omega^{(2)}$	$\omega^{(3)}$	$\omega^{(4)}$
$x^{(1)}$	1	3	4	6
$x^{(2)}$	2	2	2	4
$x^{(3)}$	3	2	1	9
$x^{(4)}$	6	6	1	3

	$\omega^{(1)}$	$\omega^{(2)}$	$\omega^{(3)}$	$\omega^{(4)}$
$\pi^T(\omega)$	0.20	0.25	0.50	0.05

Table 10.1: Evaluation matrix and probability vector for a problem of programming in conditions of risk with four alternatives and four scenarios

10.3 Sensitivity analysis in the probability space

The probabilities of the single scenarios are often obtained by sampling or estimated based on models, and therefore are not known with absolute precision. It is often useful, therefore, to evaluate the dependency of the solution suggested by a criterium from the probabilities of the scenarios, in order to understand whether a possible error can imply a wrong decision, and how large would the mistake be. This corresponds to identifying in the probability space the regions in which each alternative is optimal.

Definition 39 *We denote as* probability space *the region*

$$\mathscr{P}_\Omega = \left\{ \pi_\omega \in [0;1]^r : \sum_{\omega \in \Omega} \pi_\omega = 1, \pi_\omega \geq 0, \forall \omega \in \Omega \right\}$$

where $r = |\Omega|$.

Definition 40 *We denote as* probabilistic support *of a solution* $x \in X$ *for a given choice criterium the subset* Ω_x *of the probability space* $\mathscr{P}_\Omega$ *in which* x *is optimal according to the choice criterium adopted.*

If the alternative chosen based on the estimated probabilities π is optimal in a wide region around point π, one can feel rather safe. In the opposite case, one should take into account for further analysis the other solutions nearby, or the estimate of the probabilities should be improved. Estimating the size of the probabilistic support of each alternative can be particularly useful when the probabilities have been estimated in a very rough way, that is in the decisions that tend to be close the the programming in conditions of ignorance. Many financial decisions fall under this case, as the probabilities are often defined subjectively in such applications.

Example 70 *Consider the problem of Example 69: alternative* $x^{(3)}$*, suggested by the expected value criterium, has a value much similar to that of* $x^{(2)}$*, so close that one could doubt that approximations in the value of the impacts or of the probabilitiese might have determined this choice. It is therefore reasonable to evaluate the sensitivity of the result. For the sake of simplicity, we do not perform the analysis in the whole probability space, which would have three independent dimensions, but we only analyse the sensitivity with respect to one of the probabilities. The other ones cannot remain unchanged, because by definition the sum of all probabilities is equal to* 1*, but their relative size can be fixed, by keeping unchanged the fraction that each of them occupies of the complement of the investigated probability. This* *simplification could derive from a weaker trust of the decision-maker with respect to the value estimated for a scenario as compared to the other ones.*

Let us evaluate the sensitivity of the solution of a problem in conditions of risk with respect to the probability of a scenario $\bar{\omega}$. Setting $\pi(\bar{\omega}) = \alpha$, the complementary probability $1-\alpha$ will divide up among the other scenarios of $\Omega \setminus \{\bar{\omega}\}$, proportionally to their original probabilities. Let $\tilde{\pi}_\omega$ be the starting probabilities of the various scenarios: the new values of probability $\pi_\omega(\alpha)$ with $\omega \neq \bar{\omega}$ can be obtained multiplying the starting ones $\tilde{\pi}_\omega$ by a factor $k(\alpha)$ such that their sum is equal to $1-\alpha$:

$$\sum_{\omega \neq \bar{\omega}} \pi_\omega = \sum_{\omega \neq \bar{\omega}} k(\alpha)\, \tilde{\pi}_\omega = (1-\alpha) \Rightarrow k(\alpha) = \frac{1-\alpha}{\sum\limits_{\omega \neq \bar{\omega}} \tilde{\pi}_\omega} = \frac{1-\alpha}{1-\tilde{\pi}_{\bar{\omega}}}$$

Hence, the new probabilities are:

$$\pi_\omega(\alpha) = \begin{cases} \alpha & \textit{per } \omega = \bar{\omega} \\ \dfrac{1-\alpha}{1-\tilde{\pi}_{\bar{\omega}}} \tilde{\pi}_\omega & \textit{for } \omega \neq \bar{\omega} \end{cases}$$

In the considered example:

$$\begin{cases} \pi\left(\omega^{(1)}\right) = 0.20 \dfrac{(1-\alpha)}{1-0.50} = \dfrac{4}{10}(1-\alpha) \\ \pi\left(\omega^{(2)}\right) = 0.25 \dfrac{(1-\alpha)}{1-0.50} = \dfrac{5}{10}(1-\alpha) \\ \pi\left(\omega^{(3)}\right) = \alpha \\ \pi\left(\omega^{(4)}\right) = 0.05 \dfrac{(1-\alpha)}{1-0.50} = \dfrac{1}{10}(1-\alpha) \end{cases}$$

Now, we can apply the expected value criterium in a parametric way, and observe the performance of the four alternatives as α varies:

$$\begin{cases} \phi_\alpha\left(x^{(1)}\right) = 1 \cdot \dfrac{4}{10}(1-\alpha) + 3 \cdot \dfrac{5}{10}(1-\alpha) + 4 \cdot \alpha + 6 \cdot \dfrac{1}{10}(1-\alpha) = \dfrac{25}{10} + \dfrac{15}{10}\alpha \\ \phi_\alpha\left(x^{(2)}\right) = 2 \cdot \dfrac{4}{10}(1-\alpha) + 2 \cdot \dfrac{5}{10}(1-\alpha) + 2 \cdot \alpha + 4 \cdot \dfrac{1}{10}(1-\alpha) = \dfrac{22}{10} - \dfrac{2}{10}\alpha \\ \phi_\alpha\left(x^{(3)}\right) = 3 \cdot \dfrac{4}{10}(1-\alpha) + 2 \cdot \dfrac{5}{10}(1-\alpha) + 1 \cdot \alpha + 9 \cdot \dfrac{1}{10}(1-\alpha) = \dfrac{31}{10} - \dfrac{21}{10}\alpha \\ \phi_\alpha\left(x^{(4)}\right) = 6 \cdot \dfrac{4}{10}(1-\alpha) + 6 \cdot \dfrac{5}{10}(1-\alpha) + 1 \cdot \alpha + 3 \cdot \dfrac{1}{10}(1-\alpha) = \dfrac{57}{10} - \dfrac{47}{10}\alpha \end{cases}$$

In practice, in order to draw the profile of the choice criterium for the four alternatives, it is enough to notice that they are linear functions of α, and therefore it is enough to determine its value in two points. For $\alpha = 1$, it is enough to copy the value of the impact in the considered scenario: $f(x, \bar{\omega}) = [4\ 2\ 1\ 1]^T$. For $\alpha = 0$, on the contrary, it is necessary to compute the component of the choice criterium deriving from the other scenarios and to divide it by $1 - \tilde{\pi}_{\bar{\omega}}$: $[1.25\ 1.10\ 1.55\ 2.85]^T / 0.50$, that is $[2.5\ 2.2\ 3.1\ 5.7]^T$.

Figure 10.1 shows the behaviour of the criterium in the four alternatives: for low values of α, that is when scenario $\omega^{(3)}$ is unlikely, alternative $x^{(2)}$ is the best, whereas for large values of α, that is when scenario $\omega^{(3)}$ is likely, alternative $x^{(3)}$ is the best. The intersection between the two diagrams is identified by equation

$$u_\alpha\left(x^{(2)}\right) = u_\alpha\left(x^{(3)}\right) \Rightarrow \frac{22}{10} - \frac{2}{10}\alpha = \frac{31}{10} - \frac{21}{10}\alpha \Rightarrow \alpha = \frac{9}{19} \approx 0.474$$

This confirms that, if the probability estimate is uncertain, the choice of $x^{(3)}$ *is not necessarily the best one. Alternatives* $x^{(1)}$ *and* $x^{(4)}$*, on the contrary, lag behind* $x^{(2)}$ *and* $x^{(3)}$*, respectively, for every value of probability. This could suggest that they are strongly dominated, but this is not true: the two solutions are worse only as long as the probabilities of the scenarios* $\omega^{(1)}$*,* $\omega^{(2)}$ *and* $\omega^{(4)}$ *remain in the reciprocal ratios fixed at the beginning. There is no dominance, therefore.*

In fact, is is even possible for a nondominate alternative to have an empty probabilistic support, that is to be considered by the expected value criterium as inferior to other alternatives for any value of probability, exactly as a Paretian solution can have an empty support. The reason is also the same: the linear combinations of the impact values are not enough to determine all the nondominated solutions.

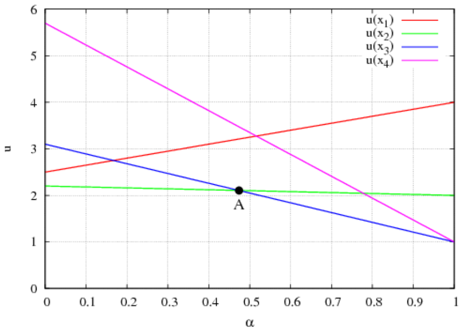


Figure 10.1: Sensitivity analysis of a cost minimisation problem with respect to probability $\pi_{\omega^{(3)}} = \alpha$

10.4 Formal defects of the expected value criterium

Since the Seventeenth century, the expected value criterium was criticised for the unrealistic consequences (sometimes even paradoxical) it leads to. Let us survey the main ones.

10.4.1 Inconsistency between expected value and actual preferences

According to the expected value criterium, all the combinations of impacts and probabilities producing the same result are reciprocally indifferent. In practice, this is often observed to be false: if one proposes to a decision-maker different combinations of impacts and probabilities with the same expected value, nearly always the decision-maker shows a preference, even if this changes depending on the decision-maker. In particular, this occurs for the so called *certainty equivalent*, that is a certain impact with a value equal to the expected one in every scenario.

Example 71 *Compare the following* 4 *alternatives:*

1. *throw a die and gain* 100 *Euros certainly;*

2. *throw a die and gain* 200 *Euros if the outcome is* 4*,* 5 *or* 6*, nothing otherwise;*

3. *throw a die and gain* 600 *Euros if the outcome is* 6*, nothing otherwise;*

4. *throw a die and gain* 200 *Euros if the outcome is* 2*,* 3*,* 4*,* 5 *or* 6*, lose* 400 *Euros if the outcome is* 1*;*

According to the expected value criterium, the four solutions are indifferent: the expected value of the impact, in fact, is always 100 *Euros. Yet, nearly no decision-maker would consider them as indifferent. The first solution is in this case the certainty equivalent.*

The previous example shows that the expected value does not correctly model the preference relation of the decision-maker from the descriptive point of view. It can be objected that this is an empirical remark on human behaviour, which concerns the descriptive models of human preference (how decision-makers act in practice), and not necessarily the prescriptive models (how decision-makers should act in order to be effective). Also from the prescriptive point of view, however, it does not seem desirable to consider the four situations as indifferent.

10.4.2 Infinite expected values and infinitesimal probabilities

A couple of famous thought experiments show inconsistencies between practical preferences and the expected value criterium when applied to situations in which some impacts are unlimitedly large and some probabilities are unlimitedly small.

Pascal's wager

A very famous application of the expected values criterium is *Pascal's wager*, which he considered as an argument to "prove" the existence of God, or better to support the sensibleness of faith (not its logical rationality). The well-known argument states as follows. Every individual must choose whether to believe or not: we can model this situation with two alternatives, $x^{(1)}$ and $x^{(2)}$. Moreover, God either exists or does not exist, that is, there are two scenarios, $\omega^{(1)}$ e $\omega^{(2)}$. Table 10.2 provides the evaluation matrix of the problem. If the individual believes and God exists, he gains paradise, which we can see as a very large benefit A (possibly, infinite). The three impacts associated to believing in a nonexistent God (and therefore losing time, pleasant occasions, but also living honestly and serenely, says Pascal), to not believing in an existing God (and therefore, perhaps, being punished) and not believing in a nonexistent God (probably no impact) are indicated, respectively, with b, c and d, which are all values (positive or negative) much smaller than A (in particular, somebody could affirm that c is a huge negative value, corresponding to an infinite punishment). Notice that, since they are benefits, the expected value criterium must be maximised, instead of minimised.

$f(x,\omega)$	$\exists$ God	$\nexists$ God
Believe	A	b
Not believe	c	d

Table 10.2: Evaluation matrix of the impacts (benefits) for Pascal's wager problem: $A \gg |b|$, $A \gg c$ and $A \gg |d|$

If we denote by α the (unknown) probability that God exist, the expected criterium generates the following utilities for the two alternatives:

$$\begin{cases} \phi\left(x^{(1)}\right) = A\alpha + b\,(1-\alpha) \\ \phi\left(x^{(2)}\right) = c\alpha + d\,(1-\alpha) \end{cases}$$

and the former exceeds the latter when $A\alpha + b(1-\alpha) > c\alpha + d(1-\alpha)$, that is $(A-b-c+d)\alpha > d-b$. If the reward A is very large (infinite), it is sensible to believe in God, even if the probability is very low; it is enough that:

$$\alpha > \alpha_{\min} = \frac{d-b}{A-b-c+d} = \frac{1}{1+\dfrac{A-c}{d-b}}$$

Since $\lim_{A \to +\infty} \alpha_{\min} = 0$ for any fixed value of b, c and d, the possibility of an infinite reward A suggests the sensibleness of faith even in the face of an extremely small probability α.

The argument immediately received objections, besides from the ethical seriousness point of view (that other authors have defended reformulating the wager in a different way), also from the point of view of mathematical validity. The idea of two alternatives was attacked (there are many ways to believe or not believe, and many faiths in which to believe or not), as well as the idea of applying the concept of probability to such a situation (it is not possible to make several random experiments and verify in the end whether God exists or not), etc... The point we are interested in, here, is however the use of the expected value criterium, that is the assumption that human beings actually care about the average gain of their actions.

Saint Petersburg's paradox

In 1730, Bernouilli[2] proposed the following gamble, in which the gambler always wins: flip a coin until it comes up tails; if that occurs after ω heads (with $\omega \in \mathbb{N}$), the gambler gains 2^ω Euros. What sum c would be reasonable to pay in order to take part to this game?

The problem includes two alternatives: to play or not play. The former offers zero gain. The second offers a gain equal to $-c + 2^\omega$, where ω is the number of heads preceding the first tails, which obviously depends on the scenario. The possible scenarios are infinitely many, one for each possible number of consecutive heads ($\Omega = \mathbb{N}$). The probability of the scenario in which ω heads are followed by a tail is $\pi_\omega = 1/2^{\omega+1}$. Hence, the expected value of the gain for the latter alternative is:

$$E[v] = -c + \sum_{\omega=0}^{+\infty} \frac{1}{2^{\omega+1}} 2^\omega = -c + \sum_{\omega=0}^{+\infty} \frac{1}{2} = +\infty$$

Therefore, playing is profitable for any possible cost c, that is, one should be willing to pay any sum of money in order to take part to the game. Everybody can see, however, that nobody would spend large sums on this game, because the games nearly always pays only a very small sum. A very powerful factor comes into play here, that is known as *risk propensity*, which depends on the amount of money at stake, on the wealth of the gambler, on his current mood, etc...

Bernouilli proposed a solution to the paradox, suggesting that the expected value should be applied not to the gain, but to its logarithm, based on the idea that every additional sum of money has for a human being a decreasing value as his wealth increases. This does not solve completely the paradox, because it is possible to design games whose rewards increase faster than in Saint Petersburg's paradox. Only a utility limited from above completely solves the issue.

[2] Daniel Bernouilli (1700-1782), Swiss mathematician and physicist, particularly interested in mechanics and fluidodynamics.

10.5 Stochastic utility theory

During the Thirties of the twentieth century, Von Neumann[3] and Morgenstern[4] introduced the stochastic utility theory, with the aim to overcome the limitations of the expected value criterium through an axiomatic approach. The basic idea is to assume that the decision-maker is able to establish a preference relation Π not only between pairs of deterministic impacts, but also between pairs of uncertain impacts, described as random variables. In the following, for the sake of simplicity, we will focus on the case of finite scenario sets.

Definition 41 *We denote as* finite simple lottery $\ell_{f,\pi}$ *a pair of functions* $(f(\omega), \pi(\omega))$, *where* $f(\omega) : \Omega \to F$ *is a random variable on a finite sample space* Ω, *while* $\pi(\omega) : \Omega \to [0;1]$ *is a probability function on* Ω.

The set of all finite simple lotteries on F and Ω is therefore $F^{|\Omega|} \times \mathcal{P}_\Omega$, as it contains all pairs of vectors such that the former has $|\Omega|$ components in F and the latter belongs to the probability space on Ω. In order to improve legibility[5], in the following we will rearrange the components of a lottery $\ell_{f,\pi} = \left([f(\omega_1) \dots f(\omega_r)]^T, [\pi(\omega_1) \dots \pi(\omega_r)]^T\right)$ to better underline the combination of probabilities and impacts, as follows:

$$\ell_{f,\pi} = \langle f(\omega_1), \pi(\omega_1)\rangle \oplus \dots \oplus \langle f(\omega_r), \pi(\omega_r)\rangle$$

Lotteries with a small number of possible scenarios have special names and notation.

Definition 42 *We denote as* degenerate lottery *a lottery* $\langle f, 1\rangle$ *(in short,* ℓ_f*), in which a scenario has probability equal to* 1 *and the others have zero probability. We denote as* binary lottery *a lottery* $\langle f, \alpha\rangle \oplus \langle f', 1-\alpha\rangle$ *(in short,* $\ell_{f,\alpha,f'}$*) in which two scenarios have probabilities summing to* 1 *and the others have zero probability.*

A single deterministic impact in F is clearly equivalent to a degenerate lottery.

We also introduce the concept of a lottery with multiple phases, in which each phase (except for the last one) is a lottery whose rewards are tickets to take part to the lottery associated with the following phase; only the final phase provides deterministic gains and losses. The left side of Figure 10.2 shows a tree representation of a compound lottery, where each leaf node corresponds to a deterministic impact, each internal node to a lottery and each arc to a scenario, with probabilities summing to 1 for the arcs going out of each node. In detail, the first phase of lottery ℓ has two possible outcomes, that correspond to lotteries ℓ_1 and ℓ_2, respectively with probabilities π_1 and π_2 (of course, $\pi_1 + \pi_2 = 1$). On its turn, lottery ℓ_1 has three possible outcomes, corresponding to impacts f_1 (with probability π_{11}), f_2 (with probability π_{12}) and f_3 (with probability $\pi_{13} = 1 - \pi_1 - \pi_2$), whereas lottery ℓ_2 has two possible outcomes, corresponding to impacts f_1 (with probability π_{21}) and f_3 (with probability $\pi_{23} = 1 - \pi_{21}$).

Definition 43 *We denote as* compound lottery *a lottery whose impacts are other lotteries (possibly degenerate).*

[3] John Von Neumann, or Jànos Neumann (1903-1957), Hungarian Jew, moved to the United States from Germany before the Nazists came to power. He was one of the fathers of game theory and of Computer Science.

[4] Oskar Morgenstern (1902-1977), antinazist Austrian economist emigrated to the United States after the annexion of Austria to Germany, one of the founders of game theory.

[5] At least, I hope: this notation is nonstandard, but the standard ones (to the best of my knowledge) are in my opinion either heavy or confuse.

Definition 44 *We denote as $L_{F,\Omega}$ the set of all possible lotteries, simple or compound, on F and Ω.*

Given a finite decision problem in conditions of risk, for any generic alternative $\bar{x} \in X$, the impact $f(\bar{x}, \omega)$ and the probability π respect the definition of finite lottery. Therefore, any $x \in X$ corresponds to a lottery $\ell(x) \in L_{F,\Omega}$ and a method allowing to compare lotteries also allows to compare alternatives.

10.5.1 Fundamental axioms of stochastic utility

The stochastic utility theory first defines the properties that a preference relation between lotteries $\Pi \subset 2^{L_{F,\Omega} \times L_{F,\Omega}}$ should respect in order to be rational. Then, it proves that only a well determined family of preference relations satisfies such properties, and that such relations can be represented by real-valued consistent functions. In this way, the selection of an alternative (lottery) among the feasible ones to the optimisation of the consistent function.

Definition 45 *A preference relation between lotteries $\Pi \subset 2^{L_{F,\Omega} \times L_{F,\Omega}}$ admits a* consistent stochastic utility function $u : L_{F,\Omega} \to \mathbb{R}$ *when, for every pair of lotteries ℓ and ℓ', the utility of the preferred one exceeds the utility of the other one:*

$$\ell \preceq \ell' \Leftrightarrow u(\ell) \geq u(\ell')$$

Since every alternative $x \in X$ of a decision problem in conditions of risk corresponds to a lottery $\ell(x)$, with stochastic utility $u(\ell(x))$, the decision problem reduces to finding an alternative x° whose corresponding lottery $\ell(x^\circ)$ have maximum stochastic utility with respect to all $x \in X$. In the following, therefore, we adopt the point of view of the economists, maximising benefits, instead of minimising costs.

The properties required from a rational preference relation between lotteries, or axioms of stochastic utility, are the following:

1. *weak ordering*: the preference relation between lotteries Π is reflexive, transitive and complete;
2. *monotony*: lotteries (both simple and compound) that assign larger probabilities to better impacts or lotteries are preferable:
$$\alpha \geq \beta \Leftrightarrow \langle \ell, \alpha \rangle \oplus \langle \ell', 1-\alpha \rangle \preceq \langle \ell, \beta \rangle \oplus \langle \ell', 1-\beta \rangle \text{ for all } \ell \preceq \ell'$$
3. *continuity*: given two lotteries ℓ and ℓ', for any intermediate impact between lotteries ℓ and ℓ' there exists a suitable probability value which allows to compose the two given lotteries into a lottery indifferent to the impact:
$$\ell \preceq f \preceq \ell' \Rightarrow \exists \alpha \in [0;1] : f \sim \langle \ell, \alpha \rangle \oplus \langle \ell', 1-\alpha \rangle$$
This means that modifying in a continuous way the probabilities of the lotteries, the preference also varies continuously, leaving no impact "uncovered".
4. *independence* (or *substitution*): the preference between two lotteries does not changing adding or removing the same lottery with the same positive probability:
$$f \preceq f' \Leftrightarrow \langle f, \alpha \rangle \oplus \langle f'', 1-\alpha \rangle \preceq \langle f', \alpha \rangle \oplus \langle f'', 1-\alpha \rangle \text{ for all } \alpha \in (0;1]$$

5. *reduction*: any compound lottery is indifferent to the simple lottery obtained listing the final impact of the compound lottery and assigning each of them the probability given by the classical composition rules:
    - the probability of each final impact is the product of the probabilities met along the sequence of phases of the lottery that lead to it, since they are conditional probabilities;
    - if a final impact appears in several final scenarios, generated by different branches of the compound lottery, the corresponding probabilities are summed, since they are associated to mutually exclusive events.

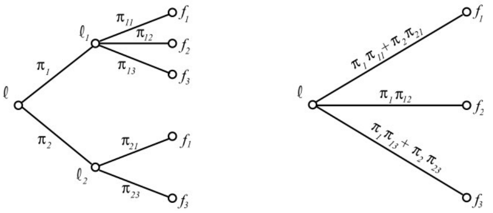


Figure 10.2: Tree representation of a compound lottery, and its replacement with an equivalent simple lottery

As already discussed, Figure 10.2 shows on the left the tree representation of a compound lottery. On the right side, the figure reports the star representation of the equivalent simple lottery, where every arc links the root directly to a leaf node. The leaf nodes are only three, because different outcomes of the compound lottery correspond to the same final impact. The probability on each arc is the sum, on all paths that in the original tree linked the root to each leaf node with a given impact, of the products of the probabilities on the arcs of each path. For example, in the original tree two paths linked the root to impact f_1: the first path went along an arc of probability π_1 and an arc of probability π_{11}, while the second path went along an arc of probability π_2 and an arc of probability π_{21}. Consequently, in the final star the root is linked to impact f_1 by an arc with probability $\pi_1\pi_{11} + \pi_2\pi_{21}$.

10.5.2 Von Neumann-Morgenstern stochastic utility theorem

The following theorem proves that, under the axioms discussed above, it is always possible to build a consistent stochastic utility function to compare lotteries. Moreover, it is possible to build infinitely many such functions, that coincide with one another up to a linear scaling, so that there is one and only one normalised stochastic utility function. Since the proof is constructive, it also shows how to build such a function.

Theorem 22 *Given a set of impacts F not all reciprocally indifferent, a sample space Ω and a preference relation Π between lotteries on F and Ω that respects the five axioms of Von Neumann and Morgenstern, there exists one and only one utility function $u(\ell) : L_{F,\Omega} \to [0,1]$ consistent with Π and normalised so as to have value 0 in the worst impact and 1 in the best impact.*

Proof. The first axiom guarantees that the preference relation Π is a weak order on the lotteries of $L_{F,\Omega}$, and therefore on the degenerate lotteries, that is on the impacts of F. Notice that this axiom is absolutely necessary to obtain a consistent utility function, since associating lotteries to real numbers and the preference relation to comparisons between real numbers automatically implies that the relation enjoys the properties of a weak order. It is not however sufficient: building the utility function requires a process that applies one by one also the other axioms.

The existence of a weak order on F guarantees that there exists at least one worst impact $f^\dagger$ and one best impact f° in F. In a decision problem, they are the extreme impacts with respect to all variables, both x and ω. Let $\ell^\dagger = \left\langle f^\dagger, 1 \right\rangle$ be the degenerate lottery that certainly returns the worst impact and $\ell^\circ = \langle f^\circ, 1 \rangle$ the degenerate lottery that certainly returns the best impact. If not all impacts are reciprocally indifferent, these two lotteries are certainly not indifferent: $f^\circ \prec f^\dagger$, that is $\ell^\circ \prec \ell^\dagger$. Now we can assign extreme conventional values to the utility of these two lotteries: zero utility to the worst degenerate lottery ($u(\ell^\dagger) = 0$) and unitary utility to the best degenerate lottery ($u(\ell^\circ) = 1$).

As a second step, we build the utility values for all degenerate lotteries, that is for all impacts in F, exploiting the continuity axiom. For every impact $f \in F$, there exists certainly a probability $\alpha_f \in [0;1]$ that produces a lottery between the two extreme impacts and is indifferent with respect to the degenerate lottery corresponding to f:

$$\exists \alpha_f \in [0;1] : \langle l^\circ, \alpha_f \rangle \oplus \left\langle l^\dagger, 1 - \alpha_f \right\rangle \sim f$$

The value of α_f is unique thanks to the monotony axiom: in fact, if two lotteries $\langle l^\circ, \alpha_f \rangle \oplus \left\langle l^\dagger, 1 - \alpha_f \right\rangle$ and $\langle l^\circ, \beta_f \rangle \oplus \left\langle l^\dagger, 1 - \beta_f \right\rangle$ were both equivalent to f, they would be reciprocally indifferent, that is, each would be preferable to the other one, and this would imply both $\alpha_f \leq \beta_f$ and $\beta_f \leq \alpha_f$. In particular, $\alpha_f = 1$ only for lotteries combining the best impacts and $\alpha_f = 0$ only for lotteries combining the worst impacts.

As a third step, we build the utility values for general lotteries. In order to do that, we exploit the substitution and reduction axioms. Every possible final impact $f \in F$ of a lottery can be seen as a degenerate lottery ℓ_f, which is equivalent to a binary lottery $\langle l^\circ, \alpha_f \rangle \oplus \left\langle l^\dagger, 1 - \alpha_f \right\rangle$ between the extreme impacts. Given a general lottery ℓ, each of its final impacts f can be replaced with the degenerate lottery or with the equivalent binary lottery obtaining a compound two-phase lottery. The first phase does no longer provide the single impacts, but tickets to take part to the second phase, in which only the extreme impacts $f^\dagger$ e f° are possible. The reduction axiom allows to combine the two phases into a simple lottery with the same final impacts (that is, the two extreme ones) and probabilities determined by the theorem of total probability. In particular, the probability of the best impact is equal to the sum over all scenarios $\omega \in \Omega$ of the product of the scenario probability π_ω times the probability α_f of the best impact in the binary lottery replacing f. Now, the given lottery ℓ corresponds to a simple lottery between the extreme impacts:

$$\ell \sim \left\langle f^\circ, \sum_{\omega \in \Omega} \pi_\omega \alpha_{f(\omega)} \right\rangle \oplus \left\langle f^\dagger, 1 - \sum_{\omega \in \Omega} \pi_\omega \alpha_{f(\omega)} \right\rangle$$

We define the utility of lottery ℓ as the probability of the best impact f°. Since $\alpha_f = u(\ell_f)$, the stochastic utility of a lottery coincides with the expected value of the utility of the impact:

$$u(\ell) = \sum_{\omega \in \Omega} \pi_\omega u(f(\omega)) = E[u(f)]$$

■

Getting back from lotteries to decision problems in conditions of risk, each alternative $x \in X$ corresponds to a lottery $\ell(x)$, formed by impact $f(x, \omega)$ and probability π_ω. Von Neumann-Morgenstern theorem allows to build a utility value $u(\ell(x))$, which can be seen directly as a utility function $u(x)$ defined on the alternatives:

$$u(x) = \sum_{\omega \in \Omega} \pi_\omega u(f(x,\omega)) = E[u(f(x,\omega))] \qquad x \in X$$

The expression of stochastic utility is very similar to that of the expected value criterium ($u(x) = \sum_{\omega \in \Omega} \pi_\omega f(x,\omega) = E[f(x,\omega)]$), but, instead of combining with the probabilities the impacts, it combines the utilities of the corresponding degenerate lotteries. In other words, this is an updated and revised version of the expected value criterium, in which each impact is not used directly, but is first filtered through the stochastic utility function.

Example 72 *Consider a simple decision problem with four possible impacts ($F = \{10, 20, 30, 50\}$), to be interpreted as benefits, and a given set of scenarios Ω. Let us build, following the steps of Von Neumann-Morgenstern theorem, a stochastic utility function that will allow to compare all possible finite lotteries $\ell = (f, \pi)$ on F and Ω.*

Since the impacts are benefits, the worst impact is the smallest one and the best impact is the largest: $f^\dagger = 10$ and $f^\circ = 50$. The degenerate lotteries $\ell^\dagger = \ell_{f^\dagger}$ e $\ell^\circ = \ell_{f^\circ}$ have, respectively, utility $u\left(\ell^\dagger\right) = 0$ and $u(\ell^\circ) = 1$. The continuity axiom guarantees that each intermediate impact $f^{(2)} = 20$ and $f^{(3)} = 30$ admits an equivalent binary lottery between the extreme impacts, $\left\langle f^\circ, \alpha_{f^{(2)}} \right\rangle \oplus \left\langle f^\dagger, 1 - \alpha_{f^{(2)}} \right)$ and $\left\langle f^\circ, \alpha_{f^{(3)}} \right\rangle \oplus \left\langle f^\dagger, 1 - \alpha_{f^{(3)}} \right)$. The decision-maker must indicate which probability α guarantees the indifference between the intermediate impact and this lottery. Let us assume that the decision-maker set $\alpha_{f^{(2)}} = 0.25$ and $\alpha_{f^{(3)}} = 0.60$. In other words, he considers indifferent to gain $f^{(2)} = 20$ certainly or having a probability equal to 0.25 to gain $f^\circ = 50$ and a complementary probability equal to 0.75 to gain $f^\dagger = 10$. As well, the decision-maker is indifferent between certainly gaining $f^{(3)} = 30$ and having a probability eual to 0.60 to gain $f^\circ = 50$ and a complementary probability equal to 0.40 to gain $f^\dagger = 10$. Then, $u\left(f^{(2)}\right) = 0.25$ and $u\left(f^{(3)}\right) = 0.60$.

Now, the utility of any lottery among the four impacts is automatically determined. For example, let us assume that scenarios $\omega^{(1)}$, $\omega^{(2)}$ and $\omega^{(3)}$ have probabilities $\pi_{\omega^{(1)}} = 0.25$, $\pi_{\omega^{(2)}} = 0.50$ and $\pi_{\omega^{(3)}} = 0.25$ and let us compare the alternatives x and x', which produce the following impacts: $f\left(x, \omega^{(1)}\right) = 10$, $f\left(x, \omega^{(2)}\right) = 20$, $f\left(x, \omega^{(3)}\right) = 50$ and $f\left(x', \omega^{(1)}\right) = 50$, $f\left(x', \omega^{(2)}\right) = 10$, $f\left(x', \omega^{(3)}\right) = 30$. Both alternatives are lotteries. Their stochastic utilities can be computed as the expected value of the utility, that is the convex combinations of the utilities associated to the single impacts with the probabilities of the corresponding scenarios:

$$u(x) = \sum_{\omega \in \Omega} \pi_\omega u(f(x,\omega)) = 0.25 \cdot 0 + 0.50 \cdot 0.25 + 0.25 \cdot 1 = 0.375$$

$$u(x') = \sum_{\omega \in \Omega} \pi_\omega u(f(x',\omega)) = 0.25 \cdot 1 + 0.50 \cdot 0 + 0.25 \cdot 0.60 = 0.40$$

Since $u(x') > u(x)$, the second alternative is preferable to the first one: $x' \prec x$.

This constructive process has a strong analogy with the construction of the multiple attribute utility function. We have, in fact, already remarked that there is a similarity in the definition of dominance: the scenarios are treated as if they were attributes. On the other hand, there are also strong differences. For example, a preference relation between deterministic impacts admits infinite consistent value functions, which are in general not additive with respect to the single attributes.

Only under strong additional assumptions, and with a complex modelling effort, it is possible to build a utility function that is a linear combination of terms referring to the single attributes. In the case of preference relations between lotteries, the five axioms guarantee that there is always a single normalised consistent utility function. Even relaxing the requirement of a normalised function, there are infinite consistent functions, but they are all connected by a simple rescaling $u'(\ell) = au(\ell) + b$ with $a > 0$. In fact, Von Neumann-Morgenstern can be proved replacing the conventional values 0 and 1 for the two extreme impacts with other two real values. Moreover, the coefficients of the linear combination are not marginal rates of substitution, to be estimated with interviews with the decision-maker, but probabilities, which can often be estimated based on the observed frequencies of the scenarios. The reason is that in this case we are combining homogeneous quantities (the impact is a single indicator, measured in different situations) instead of heterogeneous ones (the impact is a vector composed by different indicators).

10.6 Risk aversion and risk propensity

Every decision-maker has a characteristic stochastic utility function $u : L_{F,\Omega} \rightarrow [0, 1]$, that reflects his attitude towards uncertainty. The profile of such a function also indicates how much the indications of stochastic utility differ from those provided by the classical expected value criterium. In order to give a simple description of the stochastic utility function, we can focus on degenerate lotteries ℓ_f with $f \in F$, that is consider the restriction of function u from the lottery set $L_{F,\Omega}$ to the impact set F. This simplifies the task because it allows to draw a diagram with a single argument f, whereas, even considering only simple lotteries, there would be $2r - 1$ arguments: the r impacts $f(\omega)$ and the r probabilities π_ω, constrained by the sum equal to 1. On the other hand, as the theorem proves, the values of u in F univocally determine those in the whole of $L_{F,\Omega}$.

Definition 46 *We denote as* risk profile *the profile of the stochastic utility function on the degenerate lotteries ℓ_f as impact f varies in F.*

If we assume, as done so far, that impact f describes a benefit, the risk profile is a strictly increasing monotone function of f, starting from $u\left(\ell_{f^\dagger}\right) = 0$ and arriving at $u(\ell_{f^\circ}) = 1$. Consider two impacts f' and f'', with $f'' > f'$ (therefore, $f'' \prec f'$). Any intermediate impact $f \in [f', f'']$ can be obtained with a suitable convex combination of the two extreme impacts: $f_\alpha = \alpha f'' + (1 - \alpha) f'$, and it can also be seen as a degenerate lottery $\ell_{f_\alpha} = \langle f_\alpha, 1\rangle = \langle \alpha f'' + (1 - \alpha) f', 1\rangle$. Such an impact is also intermediate with respect to preference: $f'' \preceq f_\alpha \preceq f'$.

Consider the three following cases:

1. the risk profile $u(f)$ is a concave function of f in $[f', f'']$:

$$u(\ell_{f_\alpha}) = u(\langle \alpha f'' + (1 - \alpha) f', 1\rangle) \geq \alpha u(\ell_{f''}) + (1 - \alpha) u(\ell_{f'}) \text{ for all } \alpha \in [0; 1]$$

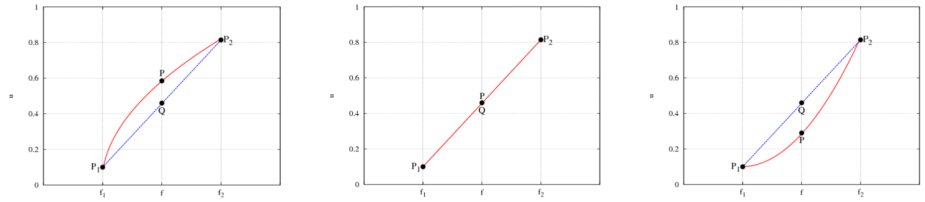
2. the risk profile $u(f)$ is a linear function of f in $[f', f'']$:

$$u(\ell_{f_\alpha}) = u(\langle \alpha f'' + (1 - \alpha) f', 1\rangle) = \alpha u(\ell_{f''}) + (1 - \alpha) u(\ell_{f'}) \text{ for all } \alpha \in [0; 1]$$

3. the risk profile $u(f)$ is a convex function of f in $[f', f'']$:

$$u(\ell_{f_\alpha}) = u(\langle \alpha f'' + (1 - \alpha) f', 1\rangle) \leq \alpha u(\ell_{f''}) + (1 - \alpha) u(\ell_{f'}) \text{ for all } \alpha \in [0; 1]$$

Figure 10.3 represents the three cases. In each of them, the extreme points of the risk profile are marked by points $P_1 = (f', u(\ell_{f'}))$ and $P_2 = (f'', u(\ell_{f''}))$, whereas the intermediate point $P = (f_\alpha, u(\ell_{f_\alpha}))$ indicates impact f and its stochastic utility. These three cases do not cover all possibilities, but they are model situations. Each of the three diagrams also reports a point Q with abscissa equal to f_α and ordinate equal to $\alpha u(f'', 1) + (1-\alpha) u(f', 1)$. This point lies on line segment $\overline{P_1 P_2}$. Let us investigate the meaning of this point.



(a) (b) (c)

Figure 10.3: Risk profiles of three decision-makers: risk-averse (a), risk-neutral (b) and risk-prone (c)

Instead of a deterministic impact intermediate between f' and f'', let us consider a lottery $\ell_\alpha = \langle f'', \alpha \rangle \oplus \langle f', 1-\alpha \rangle$ between these two impacts, assigning probability α to f''. The stochastic utility of lottery ℓ_α is:

$$u(\ell_\alpha) = \alpha u(f'') + (1-\alpha) u(f')$$

that is the ordinate of point Q, whereas the stochastic utility of f_α, that is of the degenerate lottery ℓ_{f_α}, is $u(\ell_{f_\alpha})$, that is the ordinate of point P. Therefore, in general $f_\alpha \nsim \ell_\alpha$.

Let us apply the classical expected value criterium to compare the two alternatives ℓ_{f_α} and ℓ_α. Since ℓ_{f_α} is degenerate, the value of the criterium is trivially $\phi_{\text{EV}}(\ell_{f_\alpha}) = 1 \cdot f_\alpha = f_\alpha$. On the other hand, the expected value criterium applied to lottery ℓ_α yields $\phi_{\text{EV}}(\ell_\alpha) = \alpha f'' + (1-\alpha) f' = f_\alpha$. Then, $f_\alpha \sim \ell_\alpha$ according to the expected value criterium.

Now the three profiles acquire an intuitive meaning. Building with the same coefficient α a deterministic impact f_α intermediate between f' and f'' and a lottery ℓ_α between the same extreme impacts, the expected value criterium considers them as indifferent, whereas:

1. if the risk profile $u(f)$ is a concave function, the decision-maker prefers the deterministic impact, and is therefore *risk-averse*;
2. if the risk profile $u(f)$ is a linear function, the decision-maker is indifferent, therefore *risk-neutral*;
3. if the risk profile $u(f)$ is a convex function, the decision-maker prefers the lottery, and is therefore *risk-prone*.

In general, a sufficiently regular risk profile will consist of intervals in each of which the profile assumes one of the three typical behaviours. For example, a decision-maker could consider pleasing to bet small sums, even if the expected value of the reward is negative, that is worse than not playing. The same decision-maker could refuse to bet large sums even when the expected value of the reward is positive. Conversely, another decision-maker could love risk only for large sums, and avoid it for small ones.

10.6.1 Certainty equivalent and risk premium

An equivalent description of the risk profile of a decision-maker is given by the inverse of the utility function, that is by the function that rebuilds an impact $f(u)$ for each possible value of the stochastic utility u. This inverse function certainly exists, given that $u(f)$ is strictly increasing, and its profile can be obtained by simply exchanging abscissa and ordinate in the diagram of the risk profile.

Definition 47 *Given a lottery ℓ, we denote as* certainty equivalent *the deterministic impact f_ℓ that is equivalent to the lottery, and* risk premium *the difference between the expected value of the lottery and its certainty equivalent.*

The risk premium measures the additional utility that the decision-maker requires in order to accept a lottery instead of its expected value. Therefore (see Figure 10.4):

- for risk-averse decision-makers, the certainty equivalent of a lottery is smaller than its expected value, and the risk premium is positive;
- for risk-neutral decision-makers, the certainty equivalent of a lottery is equal to its expected value, and the risk premium is zero;
- for risk-prone decision-makers, the certainty equivalent of a lottery is larger than its expected value, and the risk premium is negative.

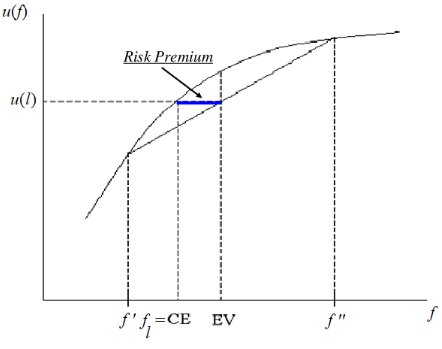


Figure 10.4: Certainty equivalent and risk premium

Example 73 *Consider a decision problem with interval* $[0; 1000]$ *as the impact set, and suppose that the risk profile of the decision-maker be* $u(\ell_f) = \sqrt{f/1000}$. *If the decision-maker should choose between a certain gain equal to* $\bar{f} = 250$ *Euros and a lottery* $\ell = (f, \pi)$ *with a gain of* 810 *Euros with probability* 0.1, *a gain of* 360 *Euros with probability* 0.5 *and a gain of* 160 *Euros with probability* 0.4, *which alternative should be chosen? What is the certainty equivalent of the lottery?*

The utility of the certain gain is $u(\bar{f}) = \sqrt{\frac{250}{1000}} = 0.5$; *the utility of the lottery is:*

$$
\begin{aligned}
u(\ell) &= \sum_{\omega\in\Omega} \pi_\omega u(f(\omega)) = \\
&= 0.1 \cdot u(810) + 0.5 \cdot u(360) + 0.4 \cdot u(160) = \\
&= 0.1 \cdot 0.9 + 0.5 \cdot 0.6 + 0.4 \cdot 0.4 = 0.55
\end{aligned}
$$

and therefore the lottery is preferable to the certain gain.

The certainty equivalent of the lottery is the impact f_ℓ that yields the same utility:

$$u(f_\ell) = \sqrt{\frac{f_\ell}{1000}} = 0.55 \Rightarrow f_\ell = 302.5$$

In order to compute the risk premium, it is necessary to evaluate the expected value of the gain from the lottery:

$$E[f] = \sum_{\omega\in\Omega} \pi_\omega f(\omega) = 0.1 \cdot 810 + 0.5 \cdot 360 + 0.4 \cdot 160 = 81 + 180 + 64 = 325$$

Therefore, the risk premium is $E[f] - f_\ell = 325 - 302.5 = 23.5$, positive. In fact, the decision-maker has a concave risk profile, and is therefore risk-averse: with respect to the lottery, the decision-maker prefers a certain gain lower than the expected value of the lottery.

10.7 Critiche alla teoria dell'utilità stocastica*

Una critica abbastanza immediata alla teoria dell'utilità stocastica è che essa, come la teoria dell'utilità a molti attributi, richiede un decisore estremamente attento, preciso e capace di soppesare le alternative propostegli, che non sono semplici impatti, ma lotterie ipotetiche. Crieri più semplici, come il bistrattato criterio del valore atteso, sono direttamente applicabili senza dover aggiungere ulteriori informazioni opinabili.

Un filone di critiche più fondamentale si indirizza invece contro gli stessi assiomi, e in particolare l'assioma di indipendenza. Questo richiede che la preferenza fra due lotterie non cambi se si aumenta o riduce in ugual misura la probabilità di impatti identici nelle due lotterie. Sembra un'ipotesi ragionevole, ma diversi esperimenti psicologici mostrano che non accade in pratica.

10.7.1 Paradosso di Allais

Questo paradosso, formulato da M. Allais[6] nel 1953, propone la scelta tra:

1. una lotteria con premi $f = [2\,500\ 2\,400\ 0]^T$ e probabilità $\pi = [0.33\ 0.66\ 0.01]^T$
2. una lotteria degenere con premio certo pari a $f_2 = 2\,400$.

I decisori più avversi al rischio preferiscono la seconda lotteria, cioè il premio certo.

Supponiamo ora di modificare entrambe le lotterie allo stesso modo, sostituendo in un sottoinsieme di scenari di probabilità 0.66 l'impatto f_2 con l'impatto f_3. Questo equivale a eliminare tale sottoinsieme di scenari e aggiungere un altro sottoinsieme di uguale probabilità con un impatto diverso. Secondo l'assioma di sostituzione, la preferenza fra le due lotterie non dovrebbe cambiare. Nella prima lotteria, l'impatto f_2 viene sostituito da f_3 in tutti gli scenari dove compariva, mentre nella seconda viene sostituito in una parte. Il risultato sono le due lotterie seguenti:

1. una lotteria con premi $f = [2\,500\ 0]^T$ e probabilità $\pi = [0.33\ 0.67]^T$
2. una lotteria con premi $f = [2\,400\ 0]^T$ e probabilità $\pi = [0.34\ 0.66]^T$

*Questa è una sezione di approfondimento, che non fa parte del programma d'esame.

[6] Maurice Allais (1911-2010), fisico ed economista, premio Nobel per l'Economia nel 1988.

Sperimentalmente, gli stessi decisori che preferivano la secondo lotteria ora scelgono in maggioranza la prima: la preferenza si è invertita.

Il fatto è che nella prima scelta l'impatto nullo, benché molto improbabile, impressiona i decisori e li spinge a cercare una certezza, mentre nel secondo test la probabilità di non vincere nulla è talmente alta da indurre i decisori a rischiare per guadagnare di più nel caso fortunato. In altre parole, 0.01 e 0 sembrano probabilità molto diverse, mentre 0.33 e 0.34 sembrano molto simili.

Il paradosso di Allais e altre considerazioni simili portarono D. Kahneman[7] e A. Tversky[8] a proporre la *teoria del prospetto*, nella quale

1. l'utilità degli impatti ha un andamento a forma di S asimmetrica, per cui il decisore soffre le perdite più di quanto goda dei guadagni di uguale entità;

2. la funzione valore combina linearmente i valori delle utilità non con le probabilità, ma con funzioni non lineari che sovrastimano le probabilità piccole e sottostimano quelle grandi.

Queste complicazioni consentono di descrivere meglio il comportamento dei decisori reali, ma riaprono il problema di quale ruolo abbia il modello decisionale: se debba solo descrivere lo stato di fatto, o anche prescrivere un approccio efficace alla decisione.

10.7.2 Il paradosso di Ellsberg

Questo paradosso, proposto da D. Ellsberg[9] nel 1961, riguarda invece un aspetto più fondamentale, cioè la definizione di probabilità usata nel modello. Il paradosso ipotizza che un'urna contenga tre biglie colorate, di cui una rossa, mentre le altre due possono essere nere o gialle, ma non si sa esattamente di quale colore siano. In un primo test si può scegliere tra due lotterie, relative all'estrazione di una biglia dall'urna:

1. nella prima lotteria, si vince 1 euro quando si estrae una biglia rossa;

2. nella seconda lotteria, si vince 1 euro quando si estrae una biglia nera.

La probabilità di vittoria per la prima lotteria è $1/3$, mentre quella per la seconda è $\alpha/3$, dove $\alpha \in \{0, 1, 2\}$ è il numero sconosciuto di biglie nere nell'urna. La maggior parte dei decisori sottoposti a questo test preferisce la prima lotteria. Secondo la teoria di Von Neumann e Morgenstern, questo significa che

$$u(\ell_1) > u(\ell_2) \Rightarrow \frac{1}{3}u(1) + \frac{2}{3}u(0) > \frac{\alpha}{3}u(1) + \left(1 - \frac{\alpha}{3}\right)u(0)$$

In un secondo test, gli stessi decisori, posti di fronte alla stessa urna, possono scegliere tra altre due lotterie:

1. nella prima lotteria, si vince 1 euro quando si estrae una biglia rossa o gialla;

2. nella seconda lotteria, si vince 1 euro quando si estrae una biglia nera o gialla.

[7] Daniel Kahneman (1934-), psicologo e premio Nobel per l'Economia nel 2002.

[8] Amos Tversky (1937-1996), psicologo.

[9] Daniel Ellsberg (1931-), economista e attivista per la libertà di stampa (dal Vietnam negli anni '70 a Wikileaks oggi).

La probabilità di vittoria è $1 - \alpha/3$ per la prima lotteria e $2/3$ per la seconda. La maggior parte dei decisori sottoposti al test preferisce la seconda lotteria. Secondo la teoria di Von Neumann e Morgenstern, questo significa che

$$u\left(\ell_2'\right) > u\left(\ell_1'\right) \Rightarrow \frac{2}{3}u\left(1\right) + \frac{1}{3}u\left(0\right) > \left(1 - \frac{\alpha}{3}\right)u\left(1\right) + \frac{\alpha}{3}u\left(0\right)$$

Se sommiamo membro a membro le due disugualianze, però, otteniamo che

$$u\left(1\right) + u\left(0\right) > u\left(1\right) + u\left(0\right) \Rightarrow 0 > 0$$

che è ovviamente un assurdo.

La differenza sostanziale sta nella costruzione della probabilità. Le lotterie vincenti in entrambi i test sono più semplici. In esse, il valore della probabilità è dato: in qualsiasi estrazione si vince con una biglia su tre, oppure con due biglie su tre. Le lotterie perdenti, invece, sono quelle in cui la probabilità di vincere non è sempre uguale, ma cambia probabilisticamente ad ogni estrazione, pur avendo un valore atteso identico a quello delle lotterie vincenti. In certe estrazioni, in effetti, il decisore può essere certo di perdere o certo di vincere. Questo rende il problema intermedio fra la programmazione in condizioni di rischio e quella in condizioni di ignoranza, e gli esseri umani tendono a preferire il rischio all'ignoranza, ma in questa situazione gli assiomi di Von Neumann e Morgenstern sono troppo stretti per essere verificati.

Chapter 11

Decision theory

Decision theory is not a distinct topic with respect to the ones discussed above, but an extension of the concepts previously exposed to more complex problems and models, both for the decisions in conditions of ignorance and for the decisions in conditions of risk. In particular, decision theory takes into account the fact that *decisions in an uncertain environment can be distributed along the time in various phases*, and that, therefore, *some decision variables could be fixed after the time in which the value of some exogenous variables is revealed.* This modifies the assumption made so far that first all decision variables should be fixed and only later the values of all exogenous variables become known.

11.1 Decision tree

The *decision tree* is a tool to solve finite decision problems: it is equivalent to the evaluation matrix, but more flexible. It introduces a hierarchical structure on the decision variables and on the exogenous variables.

The fundamental idea is that the decision process takes place in $t_{\max}$ subsequent phases, represented by a time index t increasing from 0 to $t_{\max} - 1$. In each phase t:

- first, the decision-maker fixes a subvector of decision variables $x^{(t)}$;
- then, the external world fixes a subvector of exogenous variables $\omega^{(t)}$.

The problem can be compared to a game, in which alternatively the decision-maker chooses a move $x^{(t)}$ and nature replies with a countermove $\omega^{(t)}$. When the decision-maker must choose $x^{(t)}$, he known not only the problem data, but also his own moves and those of nature in all previous phases, from 0 to $t-1$. Then, the problem does no longer consist in searching for a solution, that is a vector of real numbers $x_i^{(t)}$, but for a *strategy*, that is a vector of functions $x_i^{(t)}\left(\omega^{(0)}, \ldots, \omega^{(t-1)}\right)$, whose components depend in general from the values of the exogeneous variables known at the time of the decision.

The decision tree is composed by levels in strict chronological order, from level 0, which includes only the root of the tree, to level $2t_{\max}$, which includes only leaves. The levels are organised in pairs associated to the phases of the decision:

- the even-indexed levels $(2t)$, excluding the last one, correspond to the first half of phase t, in which the decision-maker makes choices, that is fixes the value of the decision variables $x_i^{(t)}$; the arcs going out of each node represent all possible values of such decision variables;

- the odd-indexed levels $(2t+1)$ correspond to the second half of phase t, in which part of the state of nature is revealed, that is, the value of the exogenous variables $\omega_j^{(t)}$ is fixed; the arcs going out of each node represent all possible values of such exogenous variables, that is mutually exclusive random events;

- the leaves of the tree (level $2t_{\max}$) describe situations in which decisions and states of nature are entirely fixed, that is the system configurations (x, ω).

The chronological order of the levels corresponds to the unravelling of decisions and external events, along a branch of the decision tree from the root to one of the leaves. The data of the problem are associated with the various nodes and arcs of the tree:

- each leaf (x, ω) is labelled with the corresponding impact $f(x, \omega)$;

- the arcs going out of the nodes in the odd levels can be labelled with the probabilities $\pi(\omega)$ of the random events they represent (if they are known); the sum of the probabilities for the arcs going out of each node is 1.

In the problems studied so far, the decision process underwent a single phase ($t_{\max} = 1$), so that the corresponding tree has three levels: a starting decision level, that is the root, a scenario disclosure level, that is the intermediate nodes, and a finale impact evaluation level, that is the leaves (see Figure 11.1). For these problems, the decision tree is perfectly equivalent to the evaluation matrix, plus the probability vector (if this is known).

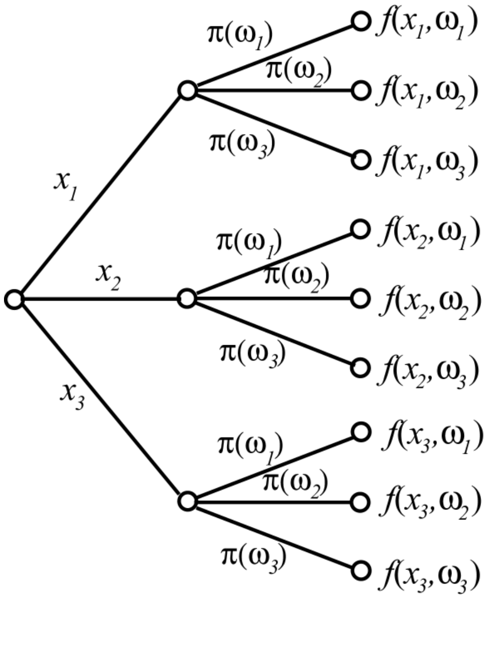


Figure 11.1: Two-level decision tree for a problem with a single decision phase, three alternatives and three scenarios

In order to solve the problem, one applies the *backward induction* method: make a backward visit of the tree, from the leaves to the root, climbing up the tree level by level and labelling each father node based on the labels of its children nodes according to the a predefined solution approach:

- if the father node corresponds to a decision, label it with the best among the labels of the children nodes, to indicate that the decision-maker always selects the alternative that produces the best result; this corresponds to fixing the values of the decision variables $x_i^{(t)}$ that represent the decision;

- if the father node corresponds to an external event, label it according to the criterium chosen to face uncertainty: worst-case, regret, equiprobability, expected value, stochastic utility, etc...

For example, if it has been decided to adopt the expected utility criterium, the label of the father node derives from the convex combination of the labels of the children nodes associated to all possible alternative events, combined with the corresponding probabilities. Adopting other choice criteria, the method builds different labels on the same tree, thus obtaining different solutions.

Example 74 *A company wants to introduce in the market a new product, choosing among three possible models (A, B, C). Three possible levels of demand have been forecast (Low, Medium and High) and the corresponding probabilities have been estimated. The models represent the feasible solutions $x \in X$, while the demand levels represent the possible scenarios $\omega \in \Omega$. Table 11.1 shows the estimated profits $f(x, \omega)$ for each configuration model/demand and the probabilities estimated for the possible scenarios. The company wants to maximise the total profit, applying the worst-case, Laplace and expected value criteria.*

$f(x,\omega)$	Demand level ω		
Model x	Low	Medium	High
A	200 000	350 000	600 000
B	250 000	350 000	540 000
C	300 000	375 000	490 000

	Low	Medium	High
Probability π	0.1	0.5	0.4

Table 11.1: Evaluation matrix and probability vector for the launch of a new product onto the market

The decision tree representing the problem is reported in Figure 11.2. Level 0 *corresponds to the decision x and the arcs going out of the root are associated with the models. Level* 1 *corresponds to the state of nature and the arcs going out of each node are associated with the demand levels, with the corresponding probabilities $\pi(\omega)$. Notice that they repeat identically in each of the three subtrees associated to the possible solutions. In fact, the probabilities do not depend on the chosen alternative. The leaves are associated with the values of the impact $f(x, \omega)$.*

Worst-case criterium *Starting from the leaves, we label each father node associated to a solution x_i computing the worst value $\min_{\omega \in \Omega} f(x_i, \omega)$ among the labels of the children nodes, that is among the possible scenarios in the father node. This is equivalent to computing the minimum of each row in the evaluation matrix, as already seen in Section 9.1*9.1*. The result is reported on the nodes of level* 1 *in Figure 11.3. At level* 0*, on the contrary, since it represents a decision, we choose the best alternative, that maximises the labels of the children nodes, that is model C. Figure 11.3 summarises the process.*

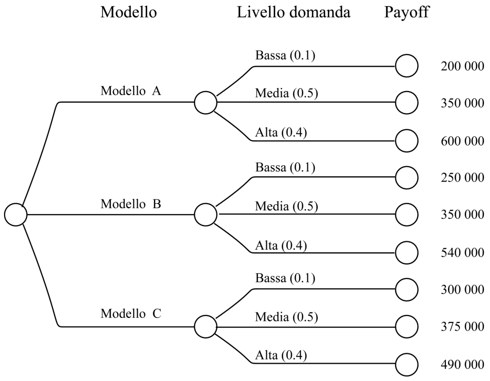


Figure 11.2: Decision tree for the launch of a new product onto the market

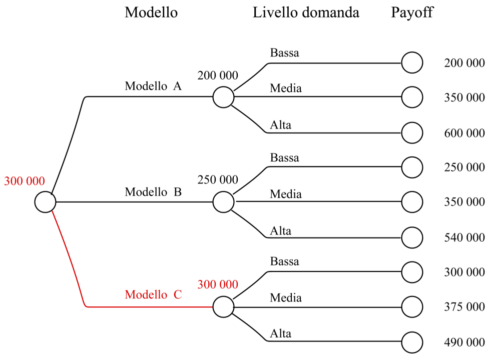


Figure 11.3: Decision tree for the launch of a new product onto the market and solution with the worst-case criterium

Laplace criterium *Laplace criterium requires the same backward induction process, but this time in the nodes at level* 1 *we apply a different labelling criterium: in each node we report the arithmetic mean of the labels of the children nodes. At level* 0*, on the contrary, since it represents a decision, we choose again the alternative that maximises the labels of the children nodes, that is once again model C, even if the labels are different. Figure 11.4 summarises the process.*

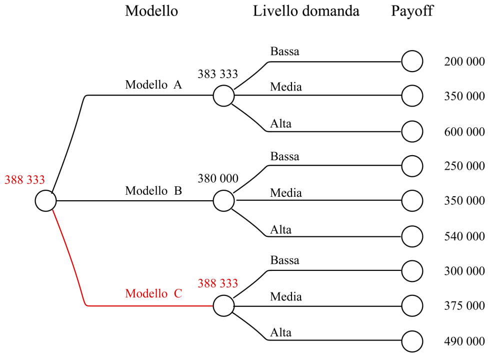


Figure 11.4: Decision tree for the launch of a new product onto the market and solution with the Laplace criterium

Expected value criterium *Starting from the leaves, we label each father node associated to a solution* x_i *computing the expected value* $E\left[f\left(x_i,\omega\right)\right]=\sum_{\omega\in\Omega}\pi_\omega f\left(x_i,\omega\right)$*. This is equivalent to computing the product of each row of the evaluation matrix by the probability vector. This time, therefore, we use the information provided by vector* π*. The result is reported on the nodes of level* 1 *in Figure 11.5. Now, we can go back from level* 1 *to the root, using the deterministic strategy, because the root corresponds to a decision: we choose the alternative that maximises the labels of the children nodes, that is model A. The associated label is copied into the root node, and provides the solution of the problem.*

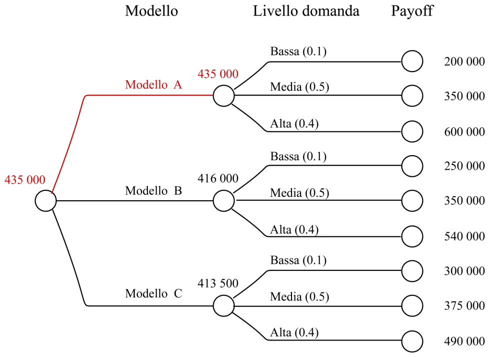


Figure 11.5: Decision tree for the launch of a new product onto the market and solution with the expected value criterium

Decision tree and evaluation matrix are two equivalent ways to represent the data of a decision problem in an uncertain environment. The solution process is the same: working on the rows of the matrix or the arcs of the tree, in fact, the same operations are performed. What is the advantage of the decision tree, then?

In the following, we present three situations in which the decision tree proves a better modelling tool, as it is more flexible than the matrix.

11.2 Scenarios conditioned by the decision

When the state of nature is influenced by the decision variables, the probabilities of the scenarios do no longer form a vector $\pi(\omega)$ of absolute values, but a matrix $\pi(\omega|x)$ of conditional values, depending on the chosen alternative. This modifies the process only slightly: in the decision tree, each node obtained with a decision has outgoing arcs associated to the scenarios opened by such a decision. It is necessary to associate with those arcs not the absolute probabilities of the scenarios, but the conditional probabilities related to the decision taken in the node.

Example 75 *Let us modify the previous example, concerning the launch of a new product onto the market. We assume that the three scenarios about the level of demand depend on the model launched, In other words, we assume that, since the possible models appeal to different targets (unknown, but estimated), the three scenarios denoted as low, medium and high demand have different probabilities to occur according to the chosen model: a more appealing model increases the probability of a high level of deman. Table 11.2 reports the conditional probabilities* $\pi(\omega_k|x_i)$.

	$\pi(\omega_k\|x_i)$		
x_i	Low	Medium	High
A	0.3	0.5	0.2
B	0.1	0.5	0.4
C	0.1	0.6	0.3

Table 11.2: Probability matrix of the states of nature, conditioned by the alternatives for the launch of a new product onto the market

Figure 11.6 reports the decision trees corresponding, respectively, to the original evaluation matrix and to the new conditional probabilities, and the resolution process with the expected value criterium. Compare the labels of the arcs with that in Figure 11.2: the arcs associated with the same scenario in different nodes no longer have the same probabilities: they are different. This time, in fact, model B wins, because it tends to be more appealing or the market, and in particular it has a stronger probability to stimulate a high demand.

11.3 Decisions distributed along multiple phases

When the decisions are distributed along time and some variables are fixed after that the state of nature is partially revealed, the decision tree has more than two levels. This problem cannot be represented in an elementary way on a two-dimension matrix. We shall see in Section , in the context of game theory, that it is possible to translate a tree representation into a matrix representation, by introducing remarkable complications.

Example 76 *Assume that, once a product has been launched, it is possible to decide whether to run also an advertising campaign to support the sales or not. In this problem, the decisions to be taken are two: which product to launch and whether to*

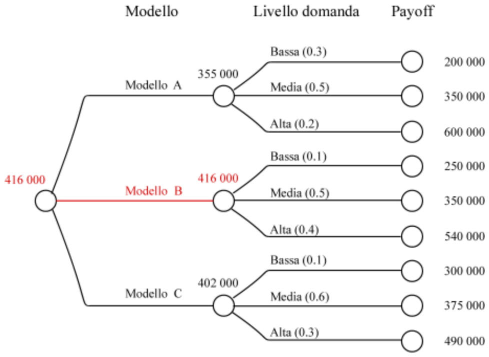


Figure 11.6: Decision tree for the launch of a new product onto the market in the case of conditional probabilities

run the campaign or not[1]*. The second decision must be taken after the launch of the product, when it has become known whether the demand is low, medium or high. On its turn, the campaign can have an uncertain outcome, so that we should introduce a further exogenous variable, and a further level in the decision tree. In order to limit the size of the example, and in order to show how deterministic phenomena can be modelled in this context, we assume the effect of the campaign to be certain, as reported in Table 11.3. The profit values reported keep into account on the one hand the increase in sales provoked by the campaign, on the other hand the cost of the campaign itself. The result, therefore, is sometimes better and sometimes worse than in Table 11.1.*

	Demand level ω		
Model x	Low	Medium	High
A	220 000	340 000	560 000
B	300 000	380 000	530 000
C	360 000	415 000	480 000

Table 11.3: Evaluation matrix for the launch of a new product onto the market after running an advertising campaign (to be compared with Table 11.1)

The assumption that the effect of the campaign be known a priori *reduces level* 3 *of the decision tree (scenarios implied by the campaign) to a set of single arcs, each going out of a node of level* 2 *and with associated probability equal to* 1. *If the campaign had different potential outcomes, each node would emit different arcs, with probabilities summing to* 1 *and potentially conditioned both by the decisions and the scenarios met along the path from the root to the father node.*

Figure 11.7 reports the data and the solving process for this extended problem. One proceeds, as usual, climbing the tree backward from the leaves to the root. Level 3 *is trivial, and only consists in copying the labels of the leaves in the corresponding father nodes. If some father nodes have several outgoing arcs, we should choose and apply a choice criterium in order to determine the label (for example, the expected value criterium). The labels at level* 2 *are obtained choosing the maximum value among the labels of the children nodes, given that the level represents a decision (whether to run or not the advertising campaign). The labels at level* 1

[1] Alternatively, it is possible to decide among different campaigns, more or less strong, and therefore more or less effective, but also more or less expensive.

is computed with the expected value criterium, and the label of the root is computed maximising. The result, marked in red in the figure, is not a simple solution, but a two-phase strategy: the former phase is determined univocally, whereas the second depends on the partial scenario decided by nature:

1. *launch model C;*
2. *measure the demand:*
   - *if it is low, run the advertising campaign;*
   - *if it is medium, run the advertising campaign;*
   - *if it is high, do not run the advertising campaign.*

The optimal strategy includes not only the choice of the model, but also the indication of how to react to the market demand, once it is revealed: in case of a high demand, do not run the campaign, because its cost is not compensated by the expected increase of the sales. The gain expected from applying this strategy is 439 500 *Euros.*

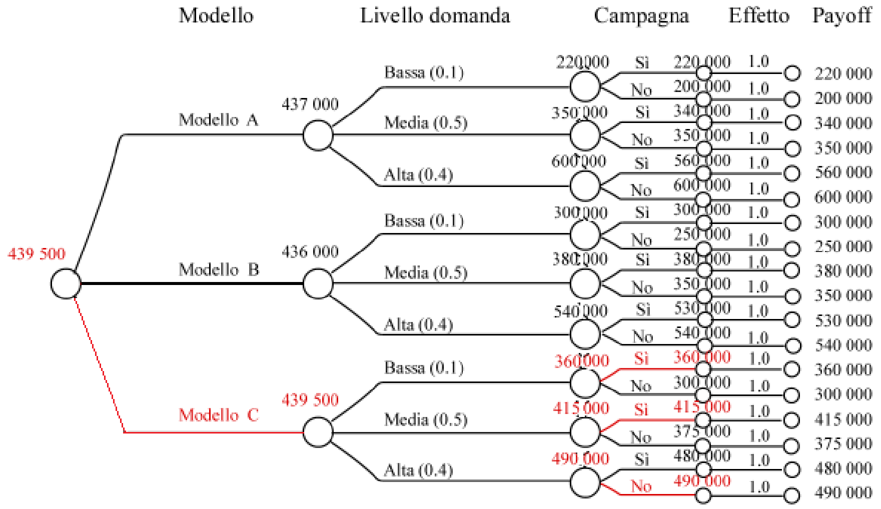


Figure 11.7: Decision tree for the launch of a new product onto the market in the case of a two-phase decision

11.4 Random experiments

Sometimes, the estimate of the probabilities of the possible scenarios can be refined through the execution of an "experiment". A typical example concerns meteorological phenomena. We can estimate the probability of rain in a given period based on historical data. However, if we check an instrument, such as a barometer, thermometer or hygrometer (or even several instruments at a time), we obtain measures that on their turn are correlated somehow with the probability of rain. The value of such measures is the outcome of a random experiment, and in general do not allow to know precisely the state of nature, that is to predict with certainty whether it will rain or not. However, knowing the precision of the instrument, that is the correlation between the outcomes of its measures and the possible scenarios, one can derive a better estimate of the probabilities of the scenarios.

The decision tree allows to represent this situation with an additional level representing uncertain scenarios, this time upstream of the basic decision. Since the decision-maker must also decide whether to perform the experiment or not, another level must be also added, referring to this decision, upstream of the level associated to the outcomes of the experiment. It must, in fact, be considered that any experiment has a cost, and therefore it is not at all obvious that it would be adviceable to perform it: this depends on the advantage gained from it in terms of the quality of the solution chosen. Figure 11.8 gives a compact representation of the decision tree, which only shows the levels of the tree, collapsing all the nodes of a level in a single one. The tree has the following structure:

- level 0 refers to the decision whether to perform the experiment or not, before considering the given problem: the decision variable $x' \in X'$ can assume only two values;
- level 1 refers to the random experiment: the exogenous variables $\omega' \in \Omega'$ indicate the outcome of the experiment, also called *signal*;
- level 2 refers to the solution $x \in X$ chosen for the given problem;
- level 3 refers to the scenarios $\omega \in \Omega$ of the given problem.

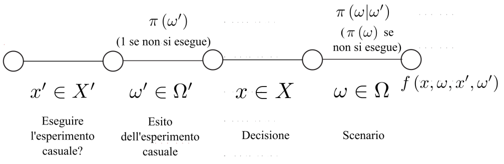


Figure 11.8: Decision tree when a random experiment is available: for the sake of simplicity, all nodes of the same level are collapsed in a single one

If one wants to apply the expected value criterium, or the expected utility criterium, one has to report on the arcs associated with the outcomes of the experiment, that is those going out of the nodes of level 1, the probabilities $\pi(\omega')$ of such outcomes. On the arcs associated with the scenarios, that is those going out of the nodes of level 3, one must report the conditional probabilities $\pi(\omega|\omega')$ of such scenarios with respect to the outcomes of the experiment. In general, the probability reported on each arc is conditioned by every event occuring on the previous arcs along the path starting at the root. In particular, the probabilities could be conditioned by the previous decisions, as in Section 11.2, that is on x' for the arcs going out of the nodes of level 1 and on x' and x for the arcs going out of the nodes of level 3. In this example, however, the decision have no influence on the scenarios. We also notice that the relation is not causal, but simply a statistical correlation: the probabilities $\pi(\omega|\omega')$ of the meteorological scenarios are conditioned by the outcomes of the instrument check. This does not mean that measuring the barometric pressure modifies the probability of rain. It means that such a measure provides further information with respect to the historical frequency, and such information modify our estimate of the probability of rain. More in general, the random experiment does not modify the state of nature, but deepens our knowledge changing the probabilities of the scenarios. On the other hand, the experiment does not give any absolute guarantee on the scenario; if it did, it would trivially reduce the scenarios to a single one, completely removing uncertainty, and reducing the whole problem to a Mathematical Programming model.

The subtree associated to the decision not to perform the experiment should skip the stochastic level associated to the outcomes of the experiment, directly linking level 0 (decision about the experiment) to level 2 (decision about the given problem). In order to keep the structure with alternate levels, we introduce also in this subtree a fictitious level 1, composed by a node that represents a degenerate experiment, with a single outgoing arc that represents a fictitious outcome of probability 1. The probabilities of the arcs on level 3 of this subtree, that is the probabilities of the states of nature, are not conditioned by the outcome of the experiment, which is not performed: they are the probabilities estimated *a priori* based on the historical data.

Finally, the leaves of the tree are, as usual, associated with the impacts of each final system configuration. The configurations are not the classical pairs (x, ω), but also include the decision variables and the exogenous variables concerning the experiment: they are therefore quadruplets (y, ω', x, ω). In fact, the impact includes also the cost of the experiments in the configurations that require to pay such a cost. The outcome of the experiment, in this example, has no influence on the final cost, but in general it could have.

Definition 48 *We denote as* information value *V the difference between the utility gained performing the experiment and the utility gained not performing it.*

The information value measures the improvement allowed by the experiment itself, and therefore the maximum cost that the decision-maker should be willing to pay in order to perform it. An experiment whose cost is larger than its information value should not be performed. Of course, this can be generalised to the situations in which there are several kinds of experiment: each of them has its own subtree, and its information value can be computed by comparison with the subtree associated to not performing any experiment.

11.4.1 Probability computation for the decision tree

In order to solve the problem with the usual backward induction method, it is required to associate the arcs of the stochastic levels with the corresponding probabilities. Each of them is the conditional probability with respect to the decisions and the outcomes associated to the arcs of the path leading from the root to the father node of the arc considered. In the case we are analysing, one must report on the tree:

1. at level 1 the total probabilities $\pi(\omega')$ of the outcomes of the experiment;
2. at level 3 the probabilities $\pi(\omega|\omega')$ of the states of nature conditioned by the outcomes of the experiment.

In order to provide an example, let us consider weather forecasts, in which the exogenous variables ω describe the future weather and exogeneous variables ω' describe the measure of an instrument. Since the measure precedes the disclosure of weather, the level of variables ω' precedes the level of variables ω. The decision tree requires on level 1 the probability $\pi(\omega')$ of each measure, and on level 3 the conditional probability $\pi(\omega|\omega')$ of each possible weather for each possible measure. Both these pieces of information are not available. However, they can be reconstructed from the available information. Now, let us consider the available information and how to reconstruct the missing one.

Besides the total probabilities of the scenarios ($\pi(\omega)$), we can assume to have information on the reliability of the instrument, that is on the conditional probability $\pi(\omega'|\omega)$ that the experiment have a given outcome ω' in a determined scenario ω. These probabilities derive from historical experience. In the case of weather forecasts, the conditional probabilities provide the frequency with which (for example) a barometer has indicated high pressure, given that the situation was destined to evolve towards nice weather, or towards bad weather. This is a historical information: we do not know whether the high pressure indicated by the barometer will truly guarantee a nice weather, but we know that in the past it has done it with a certain frequency. Figure 11.9 represents the situation: events A_i correspond to the possible weather conditions; event B to the observed fact that the barometer indicates high pressure: this reduces the possible cases, thus modifying the probability that the scenario will fall in each possible event; in other words, it modifies the probabilities of different weather conditions. The fundamental point is that we have to report on the tree $\pi(\omega')$ and $\pi(\omega|\omega')$, whereas we know $\pi(\omega)$ and $\pi(\omega'|\omega)$. In order to obtain them from the available data, it is enough to apply *Bayes' theorem*.

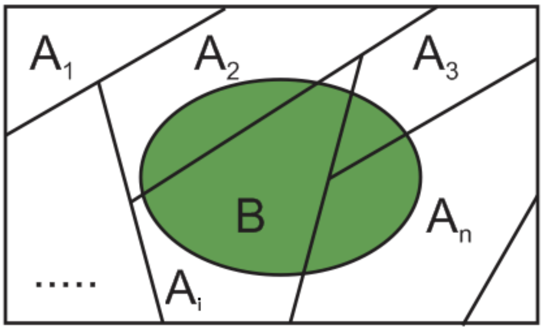


Figure 11.9: The conditional probabilities of the disjoint possible events A_i with respect to event B are different from the absolute probabilities: therefore, knowing that B has occurred gives a more precise probabilistic information on the occurrence of all events A_i

Theorem 23 Bayes' theorem*: Given a family of mutually exclusive events A_i and an event B:*

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{\sum_j P(B|A_j)P(A_j)} = \frac{P(B \cap A_i)}{P(B)}$$

In our case, the events A_i are the states of nature $\omega \in \Omega$, whereas event B is each of the outcomes ω' of the random experiment (the theorem is applied separately on each outcome). Therefore:

$$\pi(\omega|\omega') = \frac{\pi(\omega'|\omega)\,\pi(\omega)}{\sum\limits_{\omega\in\Omega} \pi(\omega'|\omega)\,\pi(\omega)}$$

Finally, the total probabilities of the outcomes of the experiment can be obtained starting from the conditional probabilities $\pi(\omega'|\omega)$ and from the probabilities of the scenarios $\pi(\omega)$, by summing their products:

$$\pi(\omega') = \sum_{\omega\in\Omega} \pi(\omega',\omega) = \sum_{\omega\in\Omega} \pi(\omega'|\omega)\,\pi(\omega)$$

Example 77 *A tourist is informed that, when the weather is bad, the rain starts falling in the morning and goes on for the whole day. The tourist office recommends to check the barometer before leaving for a trip, in order to decide what to wear. The barometer indicates nice weather, changeable variable or bad weather, while the weather can be nice and dry or bad and rainy. Moreover, the tourist can choose among three types of clothing: light, light with an umbrella, or warm, with umbrella, raincoat, hat and boots.*

Summarising, there are two possible scenarios ($|\Omega| = 2$):

- ω_1 = *nice weather*
- ω_2 = *bad weather*

The available alternatives are three ($|X| = 3$):

- x_1 = *light clothing*
- x_2 = *light clothing with umbrella*
- x_3 = *warm clothing*

Checking the barometer is a random experiments with three possible outcomes: ($|\Omega'| = 3$):

- ω'_1 = *fair*
- ω'_2 = *change*
- ω'_3 = *rain*

Table 11.4 reports the values of the "cost" $f(x,\omega)$ for every combination of decision (clothing) and scenario (weather). It is therefore a minimisation problem.

$f(x,\omega)$	ω_1	ω_2
x_1	0	5
x_2	1	3
x_3	3	2

Table 11.4: Cost function for the tourist problem

ω	ω_1	ω_2
$\pi(\omega)$	0.40	0.60

Table 11.5: Probability associated with the scenarios of the tourist problem

The historical data provide the a priori *probability $\pi(\omega)$ of each scenario (see Table 11.5) and the reliability of the barometer, that is the conditional probability $\pi(\omega'|\omega)$ of getting each measure from the barometer for each scenario (see Table 11.6).*

An "ideal" barometer would have a conditional probability equal to 1 for (ω_1, ω'_1) and (ω_2, ω'_3) and equal to 0 otherwise, but the barometer is far from being ideal. However, using it can improve the expected cost of the decision.

Figure 11.8 reports the structure of the decision tree for this problem:

1. *the arcs going out of the root (first level) represent the decision $x' \in \{0,1\}$ whether to perform the experiment or not;*

$\pi(\omega'\|\omega)$	ω_1	ω_2
ω'_1	0.60	0.20
ω'_2	0.25	0.30
ω'_3	0.15	0.50

Table 11.6: Probability of the outcomes of the random experiment, conditioned by the scenarios for the tourist problem

2. *the arcs of the second level represent the result ω' of the experiment and are labelled with their total probabilities;*
3. *the arcs of the third level represent the decision $x(\omega')$, made by the decision-maker based on the outcome of the experiment;*
4. *the arcs of the fourth level represent the states of nature ω, and are labelled with their conditional probabilities with respect to the outcomes of the experiment, and thefore determine the final cost, which is reported on the leaves.*

Applying Bayes' theorem, we proceed as follows:

1. *compute the conjoint probabilities $p(\omega, \omega')$ (Table 11.7), multiplying the conditional probabilities of the outcomes by the* a priori *probabilities of the states of nature;*

$\pi(\omega' \cap \omega)$	ω_1	ω_2
ω'_1	0.24	0.12
ω'_2	0.10	0.18
ω'_3	0.06	0.30

Table 11.7: Conjoint probabilities between the states of nature and the outcomes of the random experiment for the tourist problem

2. *derive the* a priori *probabilities of the outcomes of the experiment $\pi(\omega')$ (Table 11.8),8 summing over the state of nature, that is row by row;*

ω'	$\pi(\omega')$
ω'_1	0.36
ω'_2	0.28
ω'_3	0.36

Table 11.8: Probabilities of the outcomes of the experiment for the tourist problem

3. *derive from these the conditional probabilities of the states of nature with respect to the outcomes $\pi(\omega|\omega')$ (Table 11.9), dividing each conjoint probability by the* a priori *probability of the result.*

$\pi(\omega\|\omega')$	ω_1	ω_2
ω'_1	0.67	0.33
ω'_2	0.36	0.64
ω'_3	0.17	0.83

Table 11.9: Conditional probability of the states of nature with respect to the outcomes of the random experiment for the tourist problem

The last two families of values are used to label the arcs of the decision tree.

Now, we can climb up the tree solving the problem with the expected value criteriu: in the stochastic levels, we label the father node with the expected value of the labels of the children nodes; in the deterministic levels, we label the father node with the best (i. e., minimum) value of the labels of the children nodes. Figure 11.10 reports the decision tree, with the labels obtained from the leaves up to the root, under the assumption that the cost of the experiment be zero.

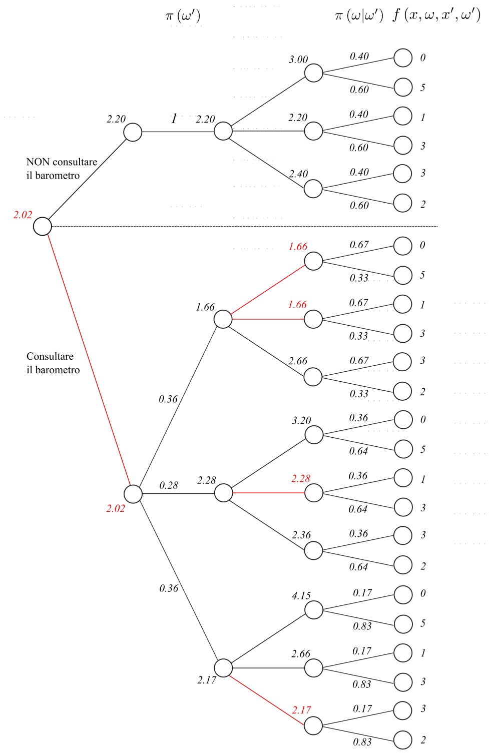


Figure 11.10: Decision tree for the tourist problem

The bolded branches indicated the choices of the decision-maker, which provide a strategy. There are actually two optimal strategies, reciprocally equivalent, which consist in setting:

$$x' = 1 \text{ (perform the experiment)} \qquad x = \begin{cases} x_1 & se \quad \omega' = \omega'_1 \\ x_2 & se \quad \omega' = \omega'_2 \\ x_3 & se \quad \omega' = \omega'_3 \end{cases}$$

or

$$x' = 1 \text{ (perform the experiment)} \qquad x = \begin{cases} x_2 & se \quad \omega' \in \{\omega'_1, \omega'_2\} \\ x_3 & se \quad \omega' = \omega'_3 \end{cases}$$

If the experiment were not performed, the best solution would be $x = x_2$. This implies that the information value is:

$$V = 2.2 - 2.02 = 0.18$$

If the experiment has a cost $C > 0$, the process must take it into account: the labels of the leaves of the subtree which derives from the choice of performing the experiment must be increased by C, so that the last three nodes of the third level have *a label increased by* C*, and the same occurs for the second node of the second level. The two nodes at level* 1 *have labels* 2.2 *(without the experiment) and* $2.02+C$ *(with the experiment). The label of the root should be the minimum of the two values. Therefore, if the experiment has a cost* $C > 0.18$*, it is not advantageous to perform it: in that case, the optimal strategy is not to perform the experiment, and to directly choose* x_2*.*

11.5 Exercises[*] [†] [‡]

Exercise 1

A decision problem admits three alternatives and two states of nature. The following table reports the benefits associate to each possible configuration.

u	ω_1	ω_2
x_1	60	20
x_2	160	-70
x_3	36	46

Indicate the alternative chosen applying the worst-case criterium, Hurwicz criterium (with a pessimism coefficient $\alpha = 0.4$), Savage criterium and Laplace criterium.

Suppose that the state of nature have probabilities given by vector $\pi = [\ 0.4\ 0.6\ ]$ and indicate the alternative chosen with the expected value criterium.

Finally, assume to have the opportunity of performing a random experiment, with two possible outcomes (y_1 e y_2), which provides information on the state of nature based on the conditional probabilities reported in the following table.

$p(y\|\omega)$	ω_1	ω_2
y_1	0.90	0.20
y_2	0.10	0.80

Indicate the best strategy according to the expected value criterium and the value of the information provided by the random experiment.

Solution

Next table report the values of the various criteria for the three alternatives.

X	$u_{\text{Wald}}(x)$	$u_{\text{Hurwicz}}(x)$	$u_{\text{Savage}}(x)$	$u_{\text{Laplace}}(x)$
x_1	20	44	100	40
x_2	-70	68	116	45
x_3	36	42	124	41

which implies that the best alternative is x_3 according to the worst-case criterium, x_2 according to Hurwicz criterium, x_1 according to the regret criterium and x_2 according to Laplace criterium.

Introducing the given probabilities, the expected value criterium provides: $E[f(x)] = [\ 36\ 22\ 42\ ]'$, so that the best alternative is x_3.

Performing the experiment, there are $3^2 = 9$ possible strategies, with the following performances.

[*] The solutions of these exercises have not yet been revised: error reports are welcome.

[†] I owe several of these exercises to exam texts of professor Alberto Colorni.

[‡] All the exercises on decision theory with random experiments can be solved also as exercises of programming in conditions of risk (just neglect the experiment) or exercises of programming in conditions of ignorance (just neglect the probabilities of the scenarios)

$f(x(y),\omega)$	ω_1	ω_2	u_{media}
(x_1, x_1)	60	20	36
(x_1, x_2)	70	-52	-3.2
(x_1, x_3)	57.6	40.8	47.5
(x_2, x_1)	150	2	61.2
(x_2, x_2)	160	-70	22
(x_2, x_3)	147.6	22.8	72.7
(x_3, x_1)	38.4	25.2	30.5
(x_3, x_2)	48.4	-46.8	-8.7
(x_3, x_3)	36	46	42

Therefore, the best strategy is the sixth one ((x_2, x_3)), with an expected value equal to 72.7. The value of the information provided by the random experiment is $V = 72.7 - 42 = 30.7$.

Exercise 2

Consider the following problem of decision theory. Given 10 boxes, let 8 of them contain 4 red tokens and 6 black tokens, while the other 2 boxes contain 9 red tokens and 1 black token. One of the boxes is chosen at random and given to the decision-maker, giving rise to two scenarios ω_1 and ω_2. The decision-maker can either guess which of the two kinds the box belongs to (alternative x_1, or x_2), or fold, that is abstain from guessing (alternative x_3). The following table reports the benefits $u(x, \omega)$.

u	ω_1	ω_2
x_1	40	-20
x_2	-5	100
x_3	0	0

Before deciding, the decision-maker can perform one of the following random experiments:

1. experiment e_1: pick up a token from the box, paying 8;
2. experiment e_2: pick up two tokens from the box, paying 12.

Of course, the decision-maker can also decide to make no experiment (e_0).

Formulate and solve with the decision tree the problem of maximising the expected value of the gain, given the following tables which provide the conjoint probabilities of the experiment outcomes and the states of nature for each of the two experiments.

e_1	ω_1	ω_2
R	0.32	0.18
N	0.48	0.02

e_2	ω_1	ω_2
RR	0.128	0.162
RN	0.384	0.036
NN	0.288	0.002

Solution

The decision tree has the following structure:

	e	y	x	ω_1	u
-	e_0	-	x_1	ω_1	40
29.5	28	28	28	ω_2	-20
		-	x_2	ω_1	5
		-	16	ω_2	100
		-	x_3	ω_1	0
		-	0	ω_2	0
	e_1	R	x_1	ω_1	40
	35.2-8=27.2	32.8	18.4	ω_2	-20
		-	x_2	ω_1	5
		-	32.8	ω_2	100
		-	x_3	ω_1	0
		-	0	ω_2	0
		N	x_1	ω_1	40
		37.6	37.6	ω_2	-20
		-	x_2	ω_1	5
		-	-0.8	ω_2	100
		-	x_3	ω_1	0
		-	0	ω_2	0
	e_2	RR	x_1	ω_1	40
	41.5-12=29.5	53.8	6.4	ω_2	-20
		-	x_2	ω_1	5
		-	53.8	ω_2	100
		-	x_3	ω_1	0
		-	0	ω_2	0
		RN	x_1	ω_1	40
		34.0	34.0	ω_2	-20
		-	x_2	ω_1	5
		-	5.5	ω_2	100
		-	x_3	ω_1	0
		-	0	ω_2	0
		NR	x_1	ω_1	40
		40.0	40	ω_2	-20
		-	x_2	ω_1	5
		-	-5	ω_2	100
		-	x_3	ω_1	0
		-	0	ω_2	0

from which it can be deduced that it is more profitable to perform experiment e_2 and then:

- if RR is picked up, choose x_2;
- otherwise, choose x_1.

The value of the information for e_1 is $V_1 = 7.2$, which is smaller than its cost, 8. By contrast, the value for e_2 is $V_2 = 13.5$, while its cost is 12.

Exercise 4

Consider the following problem with two possible alternatives (x_1 and x_2), three possible states of nature (ω_1, ω_2 and ω_3), a random experiment with four possible outcomes (y_1, y_2, y_3 and y_4) and the following benefits $u\,(x, \omega)$ and conjoint probabilities $\pi\,(y, \omega)$.

$f(x,\omega)$	ω_1	ω_2	ω_3
x_1	10	20	4
x_2	12	3	16

$\pi(y,\omega)$	ω_1	ω_2	ω_3
y_1	0.1	0.0	0.3
y_2	0.1	0.1	0.0
y_3	0.1	0.0	0.1
y_4	0.1	0.1	0.0

How many strategies are possible?

Solve the problem with the decision tree.

What is the value of the information provided by the experiment?

Solution

There are $2^4 = 16$ possible strategies.

In order to build the decision tree, it is necessary to compute the conditional probabilities $\pi(\omega|y)$, reported in the following table.

$\pi(\omega\|y)$	ω_1	ω_2	ω_3
y_1	0.25	0.00	0.75
y_2	0.50	0.50	0.00
y_3	0.50	0.00	0.50
y_4	0.50	0.50	0.00

The decision tree has the following structure:

	e	y	x	ω_1	u
-	e_0	-	x_1	ω_1	10
14.8	11.8	11.8	9.6	ω_2	20
		-	-	ω_3	4
		-	x_2	ω_1	12
		-	11.8	ω_2	3
		-	-	ω_3	16
	e_1	y_1	x_1	ω_1	10
	14.8	15.0	5.5	ω_2	20
		-	-	ω_3	4
		-	x_2	ω_1	12
		-	15.0	ω_2	3
		-	-	ω_3	16
		y_2	x_1	ω_1	10
		15.0	15.0	ω_2	20
		-	-	ω_3	4
		-	x_2	ω_1	12
		-	7.5	ω_2	3
		-	-	ω_3	16
		y_3	x_1	ω_1	10
		14.0	7.0	ω_2	20
		-	-	ω_3	4
		-	x_2	ω_1	12
		-	14.0	ω_2	3
		-	-	ω_3	16
		y_4	x_1	ω_1	10
		15.0	15.0	ω_2	20
		-	-	ω_3	4
		-	x_2	ω_1	12
		-	7.5	ω_2	3
		-	-	ω_3	16

which suggests to perform the experiment and then apply strategy $s_{11} = (x_2, x_1, x_2, x_1)$, that is choosing x_2 if the outcome of the experiment is y_1 or y_3 and choosing x_1 in the opposite case. The expected value of the benefit is 14.8.

Since the expected value when the experiment is not performed is 11.8, the value of the experiment is $V = 14.8 - 11.8 = 3.0$.

Exercise 5

A discrete decision problem in conditions of risk is given. There are two possible alternatives and two states of nature. Moreover, it is possible to perform either experiment e_1, with two possible outcomes, or experiment e_2, with three possible outcomes, besides not performing any experiment (e_0). The following tables represent, respectively, the benefits, the conjoint probabilities for the first experiment and for the second experiment.

$f(x,\omega)$	ω_1	ω_2
x_1	500	300
x_2	0	600

$\pi(y,\omega)$	ω_1	ω_2
y_1	0.2	0.3
y_2	0.1	0.4

$\pi(t,\omega)$	ω_1	ω_2
t_1	0.1	0.3
t_2	0.1	0.3
t_3	0.1	0.1

Solve the problem with the decision tree.

State how many strategies are possible, based on experiment e_1 or e_2, and list them.

Solution

The possible strategies are $2^2 = 4$ for e_1 and $2^3 = 8$ for e_2.

In order to build the decision tree, it is necessary to compute the conditional probabilities $\pi(\omega|y)$, reported in the following tables.

$\pi(\omega\|y)$	ω_1	ω_2
y_1	0.4	0.6
y_2	0.2	0.8

$\pi(\omega\|t)$	ω_1	ω_2
t_1	0.25	0.75
t_2	0.25	0.75
t_3	0.50	0.50

The decision tree has the following structure[2]:

	e	y	x	ω_1	u
-	e_0	-	x_1	ω_1	500
440	420	420	?	ω_2	300
		-	x_2	ω_1	0
		-	420	ω_2	600
	e_1	y_1	x_1	ω_1	500
	430	380	380	ω_2	300
		-	x_2	ω_1	0
		-	?	ω_2	600
		y_2	x_1	ω_1	500
		480	?	ω_2	300
		-	x_2	ω_1	0
		-	480	ω_2	600
	e_2	t_1	x_1	ω_1	500
	440	450	?	ω_2	300
		-	x_2	ω_1	0
		-	450	ω_2	600
		t_2	x_1	ω_1	500
		450	?	ω_2	300
		-	x_2	ω_1	0
		-	450	ω_2	600
		t_3	x_1	ω_1	500
		400	400	ω_2	300
		-	x_2	ω_1	0
		-	?	ω_2	600

It is profitable to perform experiment e_2, ammesso che il suo costo sia ≤ 10 e adottare la strategia $s_7 = (x_2, x_2, x_1)$, che opera la scelta x_2 se il risultato dell'esperimento è t_1 o t_2 e la scelta x_1 altrimenti.

Exercise 6

Consider the decision theory problem characterised by the benefits $f(x, \omega)$, the absolute probabilities $\pi(\omega)$ and the conditional probabilities $\pi(y|\omega)$ reported in the following tables.

[2] Some intermediate result is missing.

$f(x,\omega)$	ω_1	ω_2
x_1	100	10
x_2	50	30
x_2	25	90

	ω_1	ω_2
$\pi(\omega)$	0.5	0.5

$\pi(y\|\omega)$	ω_1	ω_2
y_1	0.8	0.4
y_2	0.2	0.6

Solve the problem using the expected value criterium, assuming to perform the random experiment with outcomes y_1 and y_2.

Solve the problem without performing the random experiment, once again with the expected value criterium.

Compute the value V of the information associated to the experiment, motivating the answer.

Solution

There are $3^2 = 9$ possible strategies. The best one, according to the expected value criterium, is $s_3 = (x_1, x_3)$, with an expected value of the benefit equal to 71.5.

Not performing the experiment, there are only three possible solutions, respectively with expected value of the benefit equal to $E[f(x_1)] = 55$, $E[f(x_2)] = 40$ and $E[f(x_3)] = 57.5$, so that the best one is x_3.

The value of the information provided by the experiment is $V = 71.5 - 57.5 = 14$, given that performing the experiment allows to increase the expected value of the benefit exactly by that amount.

Exercise 7

A decision problem in conditions of uncertainty has the following impact function, which represents benefits

$f(x,\omega)$	ω_1	ω_2
x_1	100	0
x_2	300	-200

and the following absolute probabilities for the scenarios: $\pi(\omega) = [\ 0.4\ 0.6\ ]'$.

The decision-maker can either directly make a choice, or make one of two experiments. The relation between the outcomes of the first experiment e_1 and the states of nature is given by the conjoint probabilities $\pi(y,\omega)$, the relation between the outcomes of the second experiment e_2 and the states of nature is given by the conjoint probabilities $\pi(t,\omega)$.

$\pi(y,\omega)$	ω_1	ω_2
y_1	0.30	0.20
y_2	0.10	0.40

$\pi(t,\omega)$	ω_1	ω_2
t_1	0.25	0.25
t_2	0.15	0.35

Solve the problem with the decision tree, deciding which experiment to perform and which strategy to adopt.

Compute the value of the information for both experiments.

Solution

The decision tree has the following structure:

	e	y	x	ω_1	u
-	e_0	-	x_1	ω_1	100
60	40	40	40	ω_2	0
		-	x_2	ω_1	300
		-	0	ω_2	-200
	e_1	y_1	x_1	ω_1	100
	60	100	60	ω_2	0
		-	x_2	ω_1	300
		-	100	ω_2	-200
		y_2	x_1	ω_1	100
		20	20	ω_2	0
		-	x_2	ω_1	300
		-	-100	ω_2	-200
	e_2	t_1	x_1	ω_1	100
	40	50	50	ω_2	0
		-	x_2	ω_1	300
		-	50	ω_2	-200
		t_2	x_1	ω_1	100
		30	30	ω_2	0
		-	x_2	ω_1	300
		-	-50	ω_2	-200

The optimal solution consists in performing experiment e_1 and adopting the following strategy: select x_2 if the outcome of the experiment is y_1, select x_1 if the outcome is x_1.

The value of the information for the two experiments is $V_1 = 60 - 40 = 20$ and $V_2 = 40 - 40 = 0$, respectively.

Exercise 8

A decision problem in conditions of uncertainty has an impact function $f(x, \omega)$ whose values correspond to benefits. It is possible to perform a random experiment with conjoint probabilities $\pi(x, \omega)$. The two functions are described by the following tables.

$f(x, \omega)$	ω_1	ω_2	ω_3
x_1	-5	40	10
x_2	50	0	-10

$\pi(y, \omega)$	ω_1	ω_2	ω_3
y_1	0.30	0.05	0.15
y_2	0.06	0.40	0.04

Solve the problem with a decision tree, using the expected value criterium.

Compute the value of the information provided by the experiment.

Solution

The absolute probabilities of the scenarios are $\pi(\omega) = [\ 0.36\ 0.45\ 0.19\ ]'$. The conditional probabilities with respect to the outcomes of the experiment are $\pi(\omega|y_1) = [\ 0.60\ 0.10\ 0.30\ ]'$ and $\pi(\omega|y_2) = [\ 0.12\ 0.80\ 0.08\ ]'$.

The decision tree has the following structure:

	e	y	x	ω_1	u
-	e_0	-	x_1	ω_1	-5
29.6	18.1	18.1	18.1	ω_2	40
				ω_3	10
		-	x_2	ω_1	50
		-	16.1	ω_2	0
				ω_3	-10
	e_1	y_1	x_1	ω_1	-5
	29.6	27.0	4.0	ω_2	40
				ω_3	10
		-	x_2	ω_1	50
		-	27.0	ω_2	0
				ω_3	-10
		y_2	x_1	ω_1	-5
		32.2	32.2	ω_2	40
				ω_3	10
		-	x_2	ω_1	50
		-	5.2	ω_2	0
				ω_3	-10

The optimal strategy consists in performing experiment e_1 and adopting the following strategy: select x_2 if the outcome of the experiment is y_1, select x_1 if the outcome is x_1.

The value of the information is $V = 29.6 - 18.1 = 11.5$.

Part V

Models with multiple decision-makers

Chapter 12

Game theory

The decision problems involving more than one decision-maker include two main extreme cases:

1. *game theory* studies the situations in which each decision-maker has his/her own decision variables, which are set independently from the other decision-makers;

2. *group decision-making* studies the situations in which the decision-makers share the same decision variables and must coordinate in order to set their value together.

Of course, there can be intermediate cases, in which some variables derive from compromises among decision-makers, while other variables are fixed autonomously, but the two extreme cases show the characteristic features that are worth investigating.

Game theory is modelled defining:

- a finite set of decision-makers $D = \{1, \dots, |D|\}$;

- a feasible region given by the Cartesian product of feasible regions X_d associated with the decision-makers:

$$X = X^{(1)} \times \dots \times X^{(|D|)} \Leftrightarrow x = \left[\ x^{(1)}\ \dots\ x^{(|D|)}\ \right]^T$$

  and, consequently, a decision variable vector x composed by $|D|$ subvectors $x^{(d)}$ associated with the single decision-makers and subject to constraints that involve a single subvector at a time;

- a perfectly deterministic environment, that is a single scenario:

$$\Omega = \{\omega\} \Leftrightarrow |\Omega| = 1$$

  so that it is not necessary to explicitly mention the scenario ω in the model;

- a vectorial impact function composed by d one-dimensional functions associated with the decision-makers:

$$f = \left[\ f^{(1)}\ \dots\ f^{(|D|)}\ \right]^T \text{ con } f^{(d)} : X^{(d)} \to F^{(d)} \subseteq \mathbb{R} \qquad d \in D$$

- a function Π defining for each decision-maker a preference relation Π_d, which trivially consists in preferring larger impacts:

$$\Pi_d = \left\{ \left(f^{(d)}, f'^{(d)} \right) \in F^{(d)} \times F^{(d)} : f^d \geq f'^{(d)} \right\}$$

  that is the impact components $f^{(d)}$ represent benefits, to be maximised (if it is convenient, they can become costs, to be minimised).

A game theory problem, therefore, can be summarised in the following notation:

$$\begin{aligned} \max f^{(d)} = f^{(d)}\left(x^{(1)}, \ldots, x^{(|D|)}\right) & \qquad d \in D \\ x^{(d)} \in X^{(d)} & \qquad d \in D \end{aligned}$$

The basic point in this formalisation, that distinguishes game theory from the simple combination of $|D|$ Mathematical Programming problems on independent variables and differen objective functions, is that the objective function $f^{(d)}$ of each decision-maker $d \in D$ not only depends on the variables $x^{(d)}$ controlled by the corresponding decision-maker, but also on those controlled by the other ones. This remarkably complicates the solution of the problem, so much so that it becomes dubious, in some cases, whether it makes sense to talk of a solution.

Game theory adopts a terminology partially different from the one used so far for decision problems:

- the decision-makers can also be called *players*;
- the impacts can also be called *payoffs*;
- the subvectors describing the part of solution set by each decision-maker can also be called *pure strategies*: a reason for this is that games can take place in several phases, and therefore some variables can depend on the values of other decision variables (of the same, or other, decision-makers); we will also discuss, however, another reason for this name, which is used also in single-phase games;
- the whole vector describing a solution, that is the combination of strategies adopted by all decision-makers can also be called a *strategy profile*;
- when a game takes place in several phases, each single value assigned to the variables fixed in a phase is called *move*.

Game theory has been extended to the case of multiple criteria and uncertain environment. We will neglect such extensions for the sake of simplicity, and also because they often consist in combinations of the concepts already introduced to face separately each of the three main complications of decision problems: the multiplicity of criteria, of scenarios and of decision-makers.

It is possible to classify game theory problems based on different points of view. The resulting classifications intersect one another, in general producing many cases and subcases. In particular, it is possible to distinguish:

1. *noncooperative games*, in which the decision-makers act in a completely independent way, based only on the data of the model;
2. *cooperative games*, which admit coalitions among decision-makers, constraining agreements, utility transfers among decision-makers or other similar complications which introduce new elements in the model, such as decision variables specifying whether to propose or not agreements, to accept them or not, how much utility to transfer among decision-makers in each situation, etc...).

From another point of view, it is possible to distinguish:

1. *complete information games*, in which each decision-maker knows the feasible solutions and the payoffs of all decision-makers;
2. *incomplete information games*, in which some decision-makers do not possess the whole information (for example, the players knows their payoffs, but not those of the other players).

The games in which the decision process takes place in several phases along time admit the distinction between *perfect information games*, in which each decision-maker knows the moves made in the past by all players, and those in which this does not occur. Complete and perfect information are two different situations: the former concerns the data, the latter the moves of the players. Notice that, even when the information is perfect, only past moves are known: those that are performed simultaneously by the other players, in the same phase, are still unknown.

In the following, we will only consider noncooperative, complete information games, with perfect information if the game involves several phases.

Game representations

There are two main representations of games:

1. the *extended form*, in which the game is represented as a tree;
2. the *strategic form*, in which the game is represented as a table.

This distinction closely resembles the two representations of decision problems in conditions of uncertainty: the decision tree and the evaluation matrix. In fact, as in that case, the distinction between the two representations is simply a matter of convenience: every game can be represented in both ways. However, much more than in the problems in conditions of uncertainty, each game tends to have a natural representation, whereas the alternative one appears to be forced and unconvenient. In the following, we discuss in more detail the two representations, providing examples of games in which one representation is more natural than the other.

12.1 Games in extended form

The game is described as a tree, whose structure reflects the structure of the game itself. Every level corresponds to the choice of a player, and the arcs going out of a node describe the possible moves in the configuration described by the node. The sequence of levels corresponds to the time sequence of the players' choices, and therefore depends on the rules of the game. The single nodes of the tree are denoted as *game positions* and each leaf describes one of the possible conclusions of the game, and is associated to the vector $f = \left[f^{(1)} \ \dots \ f^{(|D|)} \right]^T$ of the impacts determined for each player by the choices $x^{(d)}$ performed by all players along the path that links the root to the leaf itself.

The extended form games with sequential moves can be analysed with the same approach used for the decision trees. The only difference is that all levels are controlled by players (unless the game is stochastic), and therefore it is not necessary to model the behaviour of nature with some more or less debatable criterium (worst-case, expected value, etc...). At each level, the father node is deterministically labelled with the best among the labels of the children nodes, where "best" must be referred to the objective function $f^{(d)}$ of the player taking the choice.

Example: the tic-tac-toe game in extended form

A simple example of extended form game is tic-tac-toe. On a board with 3 rows and 3 columns, two players put, by turns, a cross or a nought. The winner is the player who first completes a sequence of three equal symbols, along a vertical, horizontal or diagonal line. If the board is full and contains no winning sequence, the game ends with a draw. Figure 12.1 shows part of the game tree for tic-tac-toe.

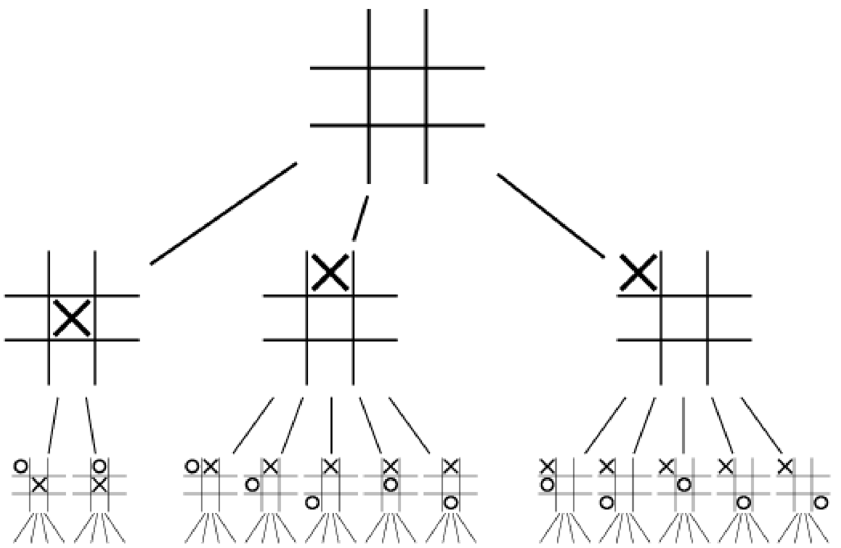


Figure 12.1: Game tree for tic-tac-toe

It is possible to analyse the game tree with the *backward induction* method, already presented in Section 11.1 and proposed by Von Neumann and Morgenstern. Climbing up the tree from the final game positions with the associated impacts (cross win, nought win or draw), one labels each node with the best label from the children nodes, where "best" refers to the impact for the player associated to the level of the father node. So, if in a level associated to the nought player the father node has at least one children node labelled with a nought win, it will be labelled in the same way; otherwise, it at least one children node is labelled with a draw, it will be labelled with a draw; finally, if all children nodes lead to a cross win, it will be labelled with a cross win. For the levels associated to the cross player, the complementary rules apply. In the case of tic-tac-toe, the result is that the root will be labelled with a draw, which means that tic-tac-toe, if played rationally, always ends up with a draw.

Example: the Marienbad game in extended form

Several rows of matches are given, each composed of n_i matches. At each turn, a player picks up any number of matches from a single row. The player picking up the last match loses the game. Figure 12.2 shows the game tree in the very simple case in which two rows are composed of $n_1 = 2$ and $n_2 = 2$ matches. The tree has been simplified considering that, if two nodes on the same level correspond to "symmetric" situations (that is situations in which the cardinalities of the rows are exchanged), the following of the game, and in particular the final result, is necessarily also the same, besides exchanging the moves:

- the root, besides arcs (A, B) and (A, C) should also emit other two arcs going to nodes "symmetric" with respect to the first two, in which one or two matches are picked up from the second row;
- node 5, besides arc (E, L) would also emit another arc, going to a node in which a single match is picked up from the second row.

Under this simplification, the game tree is not actually a tree, but a directed acyclic graph organised into levels. On the other hand, this is just a compact representation of a tree, that reduces the number of nodes, and consequently the time required to study and solve the game. The reduction is similar to that obtained by dynamic programming methods for optimisation, even if the situation is quite different.

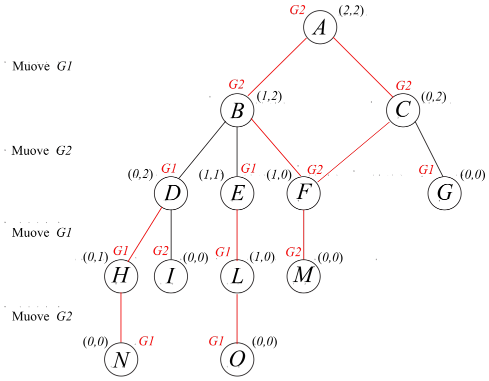


Figure 12.2: Game tree of the Marienbad game (compacted merging equivalent game positions) and solution with the backward induction method: every node has an alphabetic identifier and is labelled with the current game position (n_1, n_2); the arcs represent the possible moves; the arcs marked in red represent the optimal strategies for the two players (in the root node A, player $P1$ has two "optimal" strategies). If both players play rationally, $P2$ wins.

Figure 12.2 applies the backward induction method, showing that, if the players are rational, the second one wins. Player $P1$ has two optimal strategies. Both lead to lose, but they are the best possible one (they allow to win if $P2$ makes a mistake). Indicating the single moves with the arc names:

	Position			
Strategy	A	D	E	F
$x^{(1)}$	(A, B)	(D, H)	(E, L)	(F, M)
$x'^{(1)}$	(A, C)	(D, H)	(E, L)	(F, M)

Player $P2$ has an optimal strategy, that allows to win the game:

	Posizione			
Strategy	B	C	H	L
$x^{(2)}$	(B, F)	(C, F)	(H, N)	(L, O)

Example: the rock-paper-scissors game in extended form

Another example of extended form game is the rock-paper-scissors game, in which the players *simultaneously* choose a symbol among rock, paper and scissors. Figure 12.3 shows the game tree of this game. The basic difference with respect to the previous game is that the moves are simultaneous, so that the order between the two levels is arbitrary and the second player has an imperfect information, that is he/she does not know in which branch of the tree the current node is. In the figure, this is represented by the subset including nodes B, C and D. This complications forbids to solve the game with backward induction, because the strategy suggested by this method provides the move to perform as a function of the exact position in which the player is, indicating three different moves for the three nodes of set $\{B, C, D\}$. In order to solve the game in extended form, it is necessary to represent the incompleteness of information. This is done partitioning the game tree into subsets of nodes belonging to the same level. The player associated to that level known that he/she is in the subset, but not in which node. Consequently, the strategies must be expressed not as functions indicating a move for each given node, but as functions indicating a move for each given subset.

Definition 49 *We denote as* information set *each subset of nodes in a level of a game tree among which the player associated to that level is unable to discriminate.*

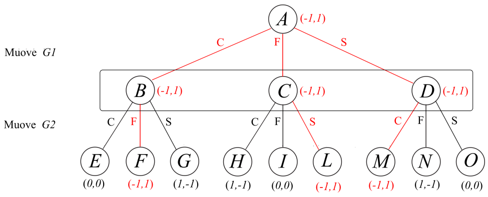


Figure 12.3: Game tree for rock-paper-scissors game: player $P2$ seems to be the winner, but in practice the player is unable to apply the winning strategy, because he/she does not know whether the current node is B, C or D; for the same reason, if the game level are exchanged, player $P1$ would deceivingly appear to be the winner.

12.2 Games in strategic form

Games in strategic form are represented by a $|D|$-dimension matrix reporting in each cell the payoffs gained by each decision-maker in the strategy profile (that is, the combination of strategies played by the single decision-makers) identified by the cell. If the player are two, they are usually denoted as *row player* and *column player*: the strategies of the former correspond to the rows of the matrix, the strategies of the latter to the columns.

Example: the rock-paper-scissors game in strategic form

Table 12.1 reports the rock-paper-scissors game in strategic form. There are two player, and each of them has three available pure strategies: rock, paper and scissors. Paper prevails on rock, rock on scissors and scissors on paper. The payoff is 1 in case of win, 0 in case of draw and -1 in case of loss. For example, if the first player chooses “paper” and the second “rock”, the payoff of the first player is 1 and that of the second is -1.

This representation makes it evident that there is no winning strategy, since no row or column contains only winning impacts for one of the two players.

	Paper	Scissors	Rock
Paper	$(0, 0)$	$(-1, 1)$	$(1, -1)$
Scissors	$(1, -1)$	$(0, 0)$	$(-1, 1)$
Rock	$(-1, 1)$	$(1, -1)$	$(0, 0)$

Table 12.1: Payoff table for the rock-paper-scissors game

Example: the Marienbad game in strategic form

In order to build the *payoff* table, it is necessary to clarify the meaning of the rows and columns of the matrix: they represent strategies, that is not single choices, but functions suggesting to the decision-maker which move to perform in every situation in which a decision is required, that is functions providing for each node associated to a decision-maker the arc going out of that node which must be selected. A strategy, therefore, can be defined in terms of subsets of arcs in the game tree.

Definition 50 *A* strategy *for player $d \in D$ corresponds to a consistent maximal subset of arcs in the game tree going out of nodes associated to the player, where:*

- consistent *means it includes no more than one arc going out of each node;*
- maximal *means that it includes at least one arc going out of each node (provided that the node has outgoing arcs).*

A strategy for a player, in fact, describes moves of the player, prescribes a single move to perform and provides an indication for each possible situation. For example, referring to the game tree of Marienbad game represented in Figure 12.2, arc (B, F) cannot belong to any strategy of player $P1$, because it describes a move of player $P2$. As well, arcs (A, B) and (A, C) are inconsistent, because they denote alternative moves starting from position A. Finally, subset $\{(A, B), (E, L)\}$ is not a complete strategy, because it does not specify what to do in node D.

Notice that this definition of strategy can be redundant, because the subsets also contain:

- forced choices (for example, arc (E, L) is the only outgoing arc for node E, so that it appears redundant to specify it explicitly in any strategy);
- impossible choices (for example, strategy $\{(A, C), (D, I), (E, L), (F, M)\}$ contains arcs (D, I) and (E, L), that it will never be possible to choose once move (A, C) is performed.

Now, we list the strategies for the restricted Marienbad game. Player $P1$ has 5 nodes associated to decisions: $\{A, D, E, F, G\}$. Nodes N and O, in fact, belong to a level associated to player $P1$, but has no outgoing arc: it is a final configuration, not associated to any decision. The possible strategies for player $P1$ are the maximal consistent subsets of arcs associated with $P1$, that is:

S1 $\{(A, B), (D, H), (E, L), (F, M)\}$,

S2 $\{(A, B), (D, I), (E, L), (F, M)\}$,

S3 $\{(A, C), (D, H), (E, L), (F, M)\}$,

S4 $\{(A, C), (D, I), (E, L), (F, M)\}$.

Remark that three arcs appear everywhere, since they describe forced choices. In the following, we will neglect them, for the sake of simplicity. On the other hand, the strategies $\{(A,C),(D,H)\}$ and $\{(A,C),(D,I)\}$ differ for an arc that represents an impossible choice. Therefore, the two strategies are in fact identical. We will merge them into a single strategy, denoted for the sake of simplicity as $\{(A,C)\}$. Player $P2$ takes decisions in nodes $\{B,C,H,L\}$, since nodes I and M represent final configurations. This implies the following strategies:

S1 $\{(B,D),(C,F),(H,N),(L,O)\}$,

S2 $\{(B,D),(C,G),(H,N),(L,O)\}$,

S3 $\{(B,E),(C,F),(H,N),(L,O)\}$,

S4 $\{(B,E),(C,G),(H,N),(L,O)\}$,

S5 $\{(B,F),(C,F),(H,N),(L,O)\}$,

S6 $\{(B,F),(C,G),(H,N),(L,O)\}$.

For the sake of simplicity, in the following we will neglect arcs (H,L) and (L,O), which describe forced choices and appear in all strategies. The payoff matrix is reported in Table 12.2. It is evident that the column player ($P2$) can always win with the strategy $\{(B,F),(C,F)\}$, whatever strategy is adopted by the row player ($P1fs$). The row player, on the contrary, has not this possibility: all strategies lead to win or lose based on the choices of the adversary.

	$\{(B,D),$ $(C,F)\}$	$\{(B,D),$ $(C,G)\}$	$\{(B,E),$ $(C,F)\}$	$\{(B,E),$ $(C,G)\}$	$\{(B,F),$ $(C,F)\}$	$\{(B,F),$ $(C,G)\}$
$\{(A,B),(D,H)\}$	(1,-1)	(1,-1)	(1,-1)	(1,-1)	(-1,1)	(-1,1)
$\{(A,B),(D,I)\}$	(-1,1)	(-1,1)	(1,-1)	(1,-1)	(-1,1)	(-1,1)
$\{(A,C)\}$	(-1,1)	(1,-1)	(-1,1)	(1,-1)	(-1,1)	(1,-1)

Table 12.2: Payoff matrix for the Marienbad game in strategic form

12.2.1 Dominance between strategies

Definition 51 *Given two strategies $x^{(d)}$ e $x'^{(d)}$ for player d, we say that $x^{(d)}$* dominates *$x'^{(d)}$ when*

$$f^{(d)}(x^{(1)},\ldots,x^{(d)},\ldots,x^{(|D|)}) \geq f^{(d)}(x^{(1)},\ldots,x'^{(d)},\ldots,x^{(|D|)})$$

for all $x^{(j)} \in X^{(j)}, j \in D \setminus \{d\}$.

This means that, for any possible behaviour of the other players $j \neq d$, strategy $x^{(d)}$ yields a better impact than strategy $x'^{(d)}$ for player d. Since the all players are assumed to be rational, that is to have a preference relation that maximises their own payoff, the dominated strategy will never be chosen.

Example 78 *Consider the following payoff matrix for a game with $|D| = 2$ players.*

	1	*2*	*3*
1	*(4,5)*	*(5,0)*	*(5,2)*
2	*(2,6)*	*(9,1)*	*(3,2)*
3	*(3,2)*	*(2,8)*	*(6,0)*

With respect to the column player, the third column is evidently a strategy dominated by the first one. In fact, it offers payoffs equal to 2, 2 *and* 0*, according to the strategy chosen by the row player and each of these payoffs is worse than those offered by the first column (*5, 6 *and* 2*). Both players know this, and therefore can remove the dominated column.*

	1	2
1	(4,5)	(5,0)
2	(2,6)	(9,1)
3	(3,2)	(2,8)

As well, with respect to the row player, the third row is dominated by the first one (payoffs equal to 3 *and* 2*, versus* 4 *and* 5*), per cui si può togliere. Notice that this dominance did not exist at the beginning, because the third column contained a gain equal to* 6 *versus* 5*. But the strategy must be removed anyway, because it is winning only under the assumption that the adversary play a dominated strategy, which we can exclude* a priori*. One could talk of a "hidden dominance".*

	1	2
1	(4,5)	(5,0)
2	(2,6)	(9,1)

Now, the second column is dominated by the first one (this dominance was previously hidden, too), and therefore can be removed: the column player can only rationally apply the first strategy.

	1
1	(4,5)
2	(2,6)

This means that the row player will apply the first strategy:

	1
1	(4,5)

The subsequent removal of dominated strategies is a method to simplify the game: it does not always yield a single strategy for both players.

Example: dominance in continuous games

Let us consider an example of game with continuous feasible regions: even if it is no longer possible to apply the matrix representation, the basic concepts still apply. Let us suppose that two individuals have different preferences on the temperature of a room: player 1 would prefer a temperature of 22 degrees, player 2 a temperature of 20 degrees. Assume that player 1 can tune the room's thermostat, setting it on a value between 16 and 26 degrees, while player 2 can operate a window, modifying its opening degree in a continuous way between the extreme states "closed window" and "wide open window". Of course, the physical model of the way in which the two control variables interact is complex, and involves many other data and quantities, among which some unknown values. Let us simplify the model, assuming that

$$T\left(x_1, x_2\right) = \left(x_1 - 5x_2\right)$$

where x_1 is the temperature set on the thermostat by player 1 and x_2 is the fraction of opening of the window set by player 2. Moreover, let us assume that for both players the discomfort be measured by the difference between the actual temperature and the desired one. The problem becomes:

$$\begin{aligned}\min_{x_1} f^{(1)} &= \left(T\left(x_1, x_2\right) - 22\right)^2 \\ \min_{x_2} f^{(2)} &= \left(T\left(x_1, x_2\right) - 20\right)^2 \\ x_1 &\in [16; 26] \\ x_2 &\in [0; 1]\end{aligned}$$

Let us compare two strategies for player 1: $x_1 = 22$ and $x_1 = \bar{x}_1 < 22$. In other words, we are wondering whether it would make sense for player 1 to set a temperature lower than 22 degrees. The intuitive answer is that, since the adversary can only decrease the temperature, it makes no sense for player 1 to set values lower than his/her optimum. In fact:

$$T\left(\bar{x}_1, x_2\right) < T\left(22, x_2\right) \leq 22 \Rightarrow f^{(1)}\left(\bar{x}_1, x_2\right) > f^{(1)}\left(22, x_2\right)$$

for every $x_2 \in X^{(2)} = [0; 1]$. Notice that the components of the impact here are costs, so that the definition of dominated strategy is exactly complementary the one given in Definition 51. Thus, strategy $\bar{x}_1$ is dominated for any value $\bar{x}_1 < 22$. This means that player 1 will never set the thermostat to values < 22 degreesm and that it is possible to reduce $X^{(1)}$ to $[22; 26]$.

Correspondingly, player 2 can use this information to exclude dominated strategies. We can compare strategy $x_2 = 0.4$ and strategy $x_2 = \bar{x}_2 < 0.4$. Since player 1 will always set the thermostat to values ≥ 22 degrees, player 2 has an advantage in always opening the window enough to let the temperature decrease by at least 2 degrees. In fact:

$$T\left(x_1, \bar{x}_2\right) > T\left(x_1, 0.4\right) \geq 20 \Rightarrow f^{(2)}\left(x_1, \bar{x}_2\right) > f^{(2)}\left(x_1, 0.4\right)$$

for every $x_1 \in X^{(1)} = [22; 26]$. Thus, the strategy $\bar{x}_2$ is dominated for any value $\bar{x}_2 < 0.4$. This means that player 2 will never open the window less than 0.4, and that it is possible to reduce $X^{(2)}$ to $[0.4; 1]$.

Now, we can go back to analyse the strategies of player 1, who knows that the window will never be kept close. Therefore, it is adviceable to set a temperature larger than the one which would be optimal in the case of closed window. With a derivation similar to the one made above, it can be proved that it does not make sense to set a temperature lower than 24 degrees. Going back to the second player, this implies that it is adviceable for him/her to open the window at least by a factor of 0.8. Consequently, player 1 should set the thermostat to the maximum temperature, that is 26 degrees. Now, the only possibility for player 2 is to open the window completely. Both players, therefore, have a single nondominated strategy:

$$x_1^* = 26 \qquad x_2^* = 1.0$$

This pair of strategies leads to a temperature equal to $T = 21$ gradi, which corresponds to a cost equal to $f^{(1)} = f^{(2)} = 1$ for both players. An interesting remark is that the same result could be obtained by setting the thermostat to 21 degrees and keeping the window close ($x_1^* = 21$ e $x_2^* = 0.0$). On the other hand, this solution would induce each of the two players to operate on his/her variable in order to improve the payoff, since the temperature is too low for the former and too high for the latter. The solution found above has the same defect but does not induce the players to change the decision, because each of them has already pushed his/her own variable to the extreme.

12.2.2 Equilibrium

Definition 52 *We denote a strategy profile* $\left(x^{*(1)},\dots,x^{*|D|}\right) \in X$ *as an* equilibrium point*, or* Nash equilibrium*, when*

$$f^{(d)}(x^{*(1)},\dots,x^{*(d)},\dots,x^{*(|D|)}) \geq f^{(d)}(x^{*(1)},\dots,x^{(d)},\dots,x^{*(|D|)})$$

for any $d \in D, x^{(d)} \in X^{(d)}$*.*

The concept of equilibrium has been investigated by Nash[1]. The intuitive meaning of the definition is that every player moving away from an equilibrium point ends up by damaging himself/herself, if all other players choose to keep their current strategies.

Notice the differences and the similiarities with the definition of dominant strategy: here a strategy of player d is compared with all the other ones (instead of a single one), but the combination of strategies of all other players is fixed (instead of considering all possible combinations of strategies). On the one hand, the definition is tighter, on the other hand it is looser. Since an equilibrium cannot be dominated, in the (rare) case in which dominance allows to remove all strategies but one, this is necessarily an equilibrium strategy. In general, however, the removal of dominated strategies will leave several residual strategies, and their combinations will not always be an equilibrium point.

Methods to determine Nash equilibria

A Nash equilibrium in a finite game in strategic form corresponds to a cell of the payoff matrix. In order to find the Nash equilibria, it is possible to use the *best response method*, which consists in:

- marking in each column the best payoff for the row player;
- marking in each row the best payoff for the column player

A cell whose payoffs are all marked is a Nash equilibrium.

Table 12.3 provides an example in which the only Nash equilibrium is cell $(3,1)$. Notice that this game has no dominated strategy. If each player assumes that the other one will not change strategy, the strategy profile $(3,1)$ is preserved, because it is the best in the whole row for the column payoff, and the best in the whole column for the row payoffs.

	1	2	3
1	$(4,\bar{3})$	$(5,1)$	$(\bar{6},2)$
2	$(2,1)$	$(8,\bar{4})$	$(3,3)$
3	$(\bar{5},\bar{9})$	$(\bar{9},6)$	$(2,8)$

Table 12.3: A payoff matrix with a Nash equilibrium in position $(3,1)$

An equivalent method to find Nash equilibria is the *arrow method*, in which one draws from each cell an arrow to the best cell in the column with respect to the row payoffs and an arrow to the best cell in the row with respect to the column payoffs. The Nash equilibria are the cells with only ingoing arrows.

A game can admit no Nash equilibrium, or several Nash equilibria.

[1] John Forbes Nash (1928-2015), American mathematician, winner of the Nobel Prize for Economy in 1994.

Example 79 *The rock-paper-scissors game represented in Table 12.4 with the corresponding markings has no equilibria.*

	Paper	Scissors	Rock
Paper	$(0,0)$	$(-1,\bar{1})$	$(\bar{1},-1)$
Scissors	$(\bar{1},-1)$	$(0,0)$	$(-1,\bar{1})$
Rock	$(-1,\bar{1})$	$(\bar{1},-1)$	$(0,0)$

Table 12.4: Payoff matrix for the rock-paper-scissors game with the markings: there is no equilibrium point, that is no cell completely marked

Table 12.5 reports an example of game with multiple equilibria: both cell $(1,1)$ *and cell* $(3,2)$ *are Nash equilibria.*

	1	2	3
1	$(\overline{15},\overline{11})$	$(4,1)$	$(\overline{6},5)$
2	$(10,\overline{4})$	$(7,2)$	$(3,1)$
3	$(3,6)$	$(\overline{12},\overline{8})$	$(5,3)$

Table 12.5: Payoff table with markings for a game with multiple Nash equilibria in positions $(1,1)$ e $(3,2)$

12.3 The worst-case strategy

How can a player choose the correct strategy when the backward induction method does not provide a single solution? The problem is similar to what happened in the programming in conditions of ignorance. The main difference is that each player does not face nature, but another rational player. In spite of that, it is not at all obvious which choice should be expected from the adversary, because the latter, on its turn, will analyse the situation and adapt his/her own decision to the one expected from the first player, who will think in the same way, and so on.

A simplifying approach drawn from the programming in conditions of ignorance is the worst-case criterium: each player determines, for each possible strategy, the strategy of the adversary that would generate the worst result for himself/herself. Then, the player chooses the strategy that, under this assumption, guarantees the best payoff. This approach in general does not provide the best performance, but only the safest performance: in general the outcome of the game will be better than the predicted one, given that perhaps the adversary will not choose exactly the most damaging strategy.

Definition 53 *We denote as* value of the game for player d *the maximum payoff that can be obtained by player d in the worst case, that is the best possible guarantee on the performance of d:*

$$u^{(d)} = \max_{x^{(d)} \in X^{(d)}} \min_{x^{(j)} \in X^{(j)}, \forall j \neq d} f^{(d)}\left(x^{(1)}, \ldots, x^{(|D|)}\right)$$

In detail, the definition evaluates for each strategy $x^{(d)}$ the worst possible result for player d with respect to the strategies adopted by the other players, that is the minimum payoff $f^{(d)}$ with respect to the strategies $x^{(j)}$ of all players $j \neq d$.

We will in particular discuss the case of the games with $|D| = 2$ players: player $d = 1$ is denoted as row player and player $d = 2$ is denoted as column player: we will use indices (r) and (c), instead of (1) and (2). The value of the game for the two players is defined as:

$$u^{(r)} = \max_{x^{(r)} \in X^{(r)}} \min_{x^{(c)} \in X^{(c)}} f^{(r)}\big(x^{(r)}, x^{(c)}\big)$$

$$u^{(c)} = \max_{x^{(c)} \in X^{(c)}} \min_{x^{(r)} \in X^{(r)}} f^{(c)}\big(x^{(r)}, x^{(c)}\big)$$

In general, these two values are completely unrelated, because they derive from different and independent functions. We shall see, however, that for some special games, they are strictly related.

Example 80 *Let us apply the worst-case strategy to the example of Table 12.3. The results are reported in Table 12.6: for every row, we evaluate the minimum of payoff $f^{(r)}$ for each row, and the minimum of payoff $f^{(c)}$ for each column.*

$(f^{(r)}, f^{(c)})$	1	2	3	$\min f^{(r)}$
1	(4,3)	(5,1)	(6,2)	4
2	(2,1)	(8,4)	(3,3)	2
3	(5,9)	(9,6)	(2,8)	2
$\min f^{(c)}$	1	1	2	

Table 12.6: Application of the worst-case criterium

Hence, $u^{(r)} = 4$ and $u^{(c)} = 2$: applying the worst-case criterium, the row player wins at least 4, whatever strategy the column player adopts; applying the worst-case criterium, the column player wins at least 2, whatever strategy the row player adopts. In particular, if both follow this criterium, the former gains 6 and the latter gains 2. In general, the gains could coincide with the values of the game or be larger. In the present case, one of the two players gains the minimum guaranteed payoff, the other one a value strictly higher.

The profile $x_{\text{worst}} = (1, 3)$ generated by the worst-case criterium is not an equlibrium point: if the game were repeated, each of the two players would have an incentive to choose a different strategy. Of course, following that incentive, the player would lose the guarantee about the worst possible performance. If the profile had been an equilibrium point, the two players would have no incentive to modify their choices.

Chapter 13

Zero-sum games

A fundamental aspect in games is the overall utility distributed to the players (i. e., the sum of the payoffs on the players) and the way in which it is distributed in the different strategy profiles. A particularly important case is provided by the zero-sum games.

Definition 54 *We denote as* zero-sum game *a game in which the overall utility of the game is zero for every strategy profile:*

$$\sum_{d \in D} f^{(d)}\left(x^{(1)}, \ldots, x^{(|D|)}\right) = 0 \textit{ for all } (x^{(1)}, \ldots, x^{(|D|)}) \in X$$

Notice that changing the scale of the game payoffs (unit of measure and offset), that is applying a linear transformation $f' = af + b$ with $a > 0$ and any b, modifies neither the dominances nor the equilibria of the game. All games with a uniform sum $\sum_{d \in D} f'(x) = b|D|$ must therefore be considered as zero-sum games.

In the zero-sum games with two players, *the win of one equals the loss of the other.* Consequently, the payoff matrix of the column player is opposite to that of the row player, and it is redundant to specify it. Table 13.1 considers the case of a two-player zero-sum game in which both players have two strategies: on the left side, the complete representation with the payoffs of both players; on the right side, the compact one, with the payoffs only of the row player. The game has only 4 indipendent impacts, instead of 8.

	1	2
1	(a,-a)	(c,-c)
2	(b,-b)	(d,-d)

	1	2
1	a	c
2	b	d

Table 13.1: A two-player zero-sum game, represented in a complete way on the left, in a compact way on the right

Let us briefly review how the fundamental concepts exposed in the previous chapter change, with respect to the strategic form, if one considers two-player zero-sum games.

13.1 Dominated strategies

Dominance is defined as usual for the row player: a row strategy dominates another one when the corresponding row has higher values, element by element. For the column player, on the contrary, the values reported in the table represent costs, so that a column strategy dominates another one when the corresponding column has lower values, element by element.

Example 81 *Figure 13.1 reports three sample games. The one with matrix F_1 has no dominated row strategies, whereas the first column dominates the second (therefore, one can remove the second column, and,* a posteriori, *conclude that the row player will certainly choose the second strategy). In matrix F_2, the third strategy of the row player dominates the other two, as it is characterised by larger gains; for the column player, the third strategy dominates the other two, as it is characterised by smaller losses. Finally, in matrix F_3 the first row is dominated, whereas there are no dominated columns; once the first row is removed, the third column is dominated by the second one; once also the third column is removed, the resulting matrix has no dominances.*

$$F_1 = \begin{bmatrix} -2 & 3 \\ -1 & 1 \end{bmatrix} \quad F_2 = \begin{bmatrix} 2 & 1 & 1 \\ -1 & 0 & -1 \\ 3 & 1 & 1 \end{bmatrix} \quad F_3 = \begin{bmatrix} -1 & 0 & -1 \\ 1 & 2 & 3 \\ 2 & -1 & -1 \end{bmatrix}$$

Figure 13.1: Three sample two-player zero-sum games

13.2 Equilibria

Definition 55 *In a two-person zero-sum game, the strategy profile $(x^{*(r)}, x^{*(c)})$ is an equilibrium point if and only if cell (r, c) is a* saddle point *of the payoff matrix: a point of maximum for column c and of minimum for row r:*

$$\begin{cases} f_{rc} \geq f_{ic} \text{ for all } i \in X^{(r)} \\ f_{rc} \leq f_{rj} \text{ for all } j \in X^{(c)} \end{cases}$$

Example 82 *Considering again the three examples of Figure 13.1, matrix F_1 admits the saddle point $(2, 1)$ with $f_{21} = -1$. Matrix F_2 has four saddle points, that are all the elements of value 1. Finally, matrix F_3 has no saddle point.*

13.3 Value of the game

In two-player zero-sum games, the worst-case assumption is more reasonable than in other games, because the impact on the two players are always exactly opposite, and therefore the worst-case for a player coincides with the best one for the other. It is natural to assume that the adversary will play so as to provoke the maximum damage, and that the worst-case criterium will be a useful guide.

The concept of value of the game for each of the two players, that is of best possible guarantee on the result, is defined as usual for the row player:

$$u^{(r)} = \max_{i \in X^{(r)}} \min_{j \in X^{(c)}} f_{ij}$$

whereas the definition for the column player is:

$$\max_{j \in X^{(c)}} \min_{i \in X^{(r)}} (-f_{ij}) = \max_{j \in X^{(c)}} \left(- \max_{i \in X^{(r)}} f_{ij} \right) = - \min_{j \in X^{(c)}} \max_{i \in X^{(r)}} f_{ij}$$

By convention, in the two-player zero-sum games, the sign in the definition of the value of the game is reversed for the column player:

$$u^{(c)} = \min_{j \in X^{(c)}} \max_{i \in X^{(r)}} f_{ij}$$

With this convention, the value of the game does no longer describe a minimum guaranteed gain, but a maximum guaranteed loss: the column player does not maximise the gain, but minimises the loss.

While in general the value of the game for a player has no relation with the value of the game for the other players, in the two-player zero-sum games the two values are strictly related by the following theorem, which admits a generalisation even more interesting, as we shall see.

Theorem 24 *Given a two-player zero-sum game, if $u^{(r)}$ is the value of the game for the row player and $u^{(c)}$ is the value of the game for the colum player:*

1. $u^{(r)} \leq u^{(c)}$;
2. $u^{(r)} = u^{(c)}$ *if and only if the game has at least a Nash equilibrium, that is the payoff matrix F has a saddle point.*

The theorem states that the minimum guaranteed gain for the row player cannot exceed the maximum guaranteed loss for the column player. The reason is rather intuitive, given that they are two limits, a lower and an upper bound, on the same (unknown) quantity that is transferred from one player to the other. In general, the two bounds do not coincide, because each player could have a result better than the guaranteed one: a larger gain, or a smaller loss. If F has no saddle point, the game does not admit equilibria: in any situation, at least one of the players has an incentive to change strategy, in order to improve the result (assuming that the adversary will keep the same strategy). On the other hand, if and only if F has a saddle point, the choice produced by the worst-case criterium are consistent and lead to the equilibrium that corresponds to the saddle point. In that case, the value of the game is the same for the two players, and this is denoted as *value of the game* without any further specification. It is the maximum utility obtainable by both players. This also means that, if there is more than one equilibrium, or saddle point, all of them have the same value.

$$F_1 = \begin{bmatrix} -2 & 3 \\ -1 & 1 \end{bmatrix} \quad F_2 = \begin{bmatrix} -2 & -3 \\ 0 & 3 \end{bmatrix} \quad F_3 = \begin{bmatrix} 2 & -3 \\ 0 & 2 \\ -5 & 10 \end{bmatrix}$$

Figure 13.2: Three sample two-person zero-sum games

Example 83 *Figure 13.2 reports other three samples of two-person zero-sum games. In the first game, the worst-case for the row player corresponds to the minimum payoff on each row, that is -2 and -1, and the best strategy is the second one, which implies $u^{(r)}(F_1) = -1$. For the column player, the worst-case corresponds to the maximum loss on each column, that is -1 and 3, and the best strategy is the first one, which implies $u^{(c)}(F_1) = -1$.*

In the second game, the worst-case for the row player corresponds to the minimum payoff on each row, that is -3 and 0, and the best strategy is the second, which implies $u^{(r)}(F_2) = 0$. For the column player, the worst-case corresponds to *the maximum loss on each column, that is* 0 *and* 3*, and the best strategy is the first, which implies* $u^{(c)}(F_2) = 0$.

Finally, in the third game, the worst-case for the row player corresponds to the minimum payoff on each row, that is -3, 0 *and* -5*, and the best strategy is the second, which implies* $u^{(r)}(F_3) = 0$. *For the column player, the worst-case corresponds to the maximum loss on each column, that is* 2 *and* 10*, and the best strategy is the first, which implies* $u^{(c)}(F_3) = 2$.

In the first two games, the strategy profiles suggested by the worst-case criterium are equilibria, whereas in the third one there is no equilibrium, and any given strategy profile gives to at least one player an incentive to change strategy.

13.4 Mixed strategies

If the game is played several times in sequence, it is licit for the players to change the strategy adopted in each round. In this situation, the decision variable for each player has no longer a finite feasible region, but a continuous set, corresponding to the frequency with which the player should choose each basic alternative. These combinations of frequencies are denoted as *mixed strategies*, which explains why the basic alternatives of the game are denoted as pure strategies.

This model also describes the situation in which the same game is played in parallel by two teams of players, and the players of each team can adopt different strategies. As well, the same model describes the situation in which two players play a single round, and each selects an alternative with a random extraction, after deciding the probabilities of the single alternatives available. In all these cases, the utility of the players becomes a random variable, and the purpose of the game for the row player is to maximise the expected value of the gain, while the purpose for the column player is to minimise the expected value of the loss.

Definition 56 *We denote as* mixed strategy *for a player a probability vector*

$$\xi = [\xi, \ldots, \xi_n]' \in \Xi = \left\{ \xi \in \mathbb{R}^n : \sum_{i=1}^{n} \xi_i = 1 \textit{ and } \xi_i \geq 0 \textit{ for all } i = 1, \ldots, n \right\}$$

where n *is the number or pure strategies available for the player.*

In a two-player game, the row player has a set $X^{(r)}$ of $n^{(r)} = \left|X^{(r)}\right|$ possibles pure strategies, the column player a set $X^{(c)}$ of $n^{(c)} = \left|X^{(c)}\right|$ possibles pure strategies. Combining them with suitable probabilities, they can obtain the mixed strategies $\xi^{(r)} \in \Xi^{(r)}$ e $\xi^{(c)} \in \Xi^{(c)}$. Determining the expected impact for the two players is easy: the expected gain for the former is equal to the expected loss for the latter and can be obtained by summing on each combination of outcomes $(i, j) \in \Xi^{(r)} \times \Xi^{(c)}$ the product of the gain f_{ij} times the probability $\xi_i^{(r)} \xi_j^{(c)}$ associated to the specific combination.

Definition 57 *Given a two-person zero-sum game* F *and two strategies* $\xi^{(r)}$ *and* $\xi^{(c)}$*, we denote as* expected value of the game *the expected value of the gain for the row player and of the loss for the column player.*

$$v(F) = E\left[f\left(\xi^{(r)}, \xi^{(c)} \right) \right] = \sum_{i=1}^{n^{(r)}} \sum_{j=1}^{n^{(c)}} \xi_i^{(r)} \xi_j^{(c)} f_{ij}$$

If the two players adopt the worst-case criterium, the row player aims to maximise the minimum of the expected value with respect to the strategies of the adversary, whereas the column player aims to minimise the maximum of the expected value with respect to the strategies of the adversary. Both operate on the respective probability vectors $\xi^{(r)}$ and $\xi^{(c)}$, instead of simply choosing one of the pure strategies.

Definition 58 *We denote as* expected value of the game for the row player *the optimal expected value of the gain in the worst case:*

$$v^{(r)}(F) = \max_{\xi^{(r)} \in \Xi^{(r)}} \min_{\xi^{(c)} \in \Xi^{(c)}} E\left[f\left(\xi^{(r)}, \xi^{(c)}\right)\right]$$

and as expected value of the game for the column player *the optimal expected value of the loss in the worst case:*

$$v^{(c)}(F) = \min_{\xi^{(c)} \in \Xi^{(c)}} \max_{\xi^{(r)} \in \Xi^{(r)}} E\left[f\left(\xi^{(r)}, \xi^{(c)}\right)\right]$$

Notice that these values do not coincide with the values $u^{(r)}$ e $u^{(c)}$ introduced in Definition 53 of Section 12.3, as they refer to pure strategies, whereas these refer to mixed strategies. Technically speaking, the problems which define these expected values are nothing else than continuous relaxations of the corresponding problems which define the value of the game with respect to pure strategies. In fact, a pure strategy is a special case of mixed strategy in which a probability is equal to 1 and the others are equal to zero. The objective function is the same, the constraints imposing nonnegative values and a sum equal to 1 are the same, but a pure strategy also respects an integrality constraints. The relation between u and v is not obvious, because in the second case both players have more choices: on the one hand, each player has more chances to improve his/her own result, on the other hand, also the chances to be damaged by the adversary are much larger.

13.5 Von Neumann's minimax theorem

For both players, however, the advantage of using mixed strategies prevails on the disadvantage of facing a mixed strategy. This depends on the following important remark, that allowed Von Neumann to prove a fundamental theorem on zero-sum games.

Proposition 2 *Whatever mixed strategy a player adopts, the worst-case corresponds to the application of a pure strategy by the adversary. Consequently, the expected value of the game for the row player can be expressed as:*

$$v^{(r)}(F) = \max_{\xi^{(r)} \in \Xi^{(r)}} \min_{j \in X^{(c)}} E\left[f_{\cdot j}\left(\xi^{(r)}\right)\right] \text{ with } E\left[f_{\cdot j}\left(\xi^{(r)}\right)\right] = \sum_{i=1}^{n^{(r)}} \xi_i^{(r)} f_{ij}$$

and the expected value of the game for the column player as:

$$v^{(c)}(F) = \min_{\xi^{(c)} \in \Xi^{(c)}} \max_{i \in X^{(r)}} E\left[f_{i \cdot}\left(\xi^{(c)}\right)\right] \text{ with } E\left[f_{i \cdot}\left(\xi^{(c)}\right)\right] = \sum_{j=1}^{n^{(c)}} \xi_j^{(c)} f_{ij}$$

Proof. Consider the case of the row player. For a fixed strategy $\xi^{(r)}$, the payoff in the worst case will be:

$$\phi(\xi^{(r)}) = \min_{\xi^{(c)} \in \Xi^{(c)}} \sum_{i=1}^{n^{(r)}} \sum_{j=1}^{n^{(c)}} \xi_i^{(r)} \xi_j^{(c)} f_{ij} = \min_{\xi^{(c)} \in \Xi^{(c)}} \sum_{j=1}^{n^{(c)}} \left(\sum_{i=1}^{n^{(r)}} \xi_i^{(r)} f_{ij} \right) \xi_j^{(c)}$$

where the term in round parenthesis is a vector of constant coefficients with index j, given that the probabilities $\xi_i^{(r)}$ that define the mixed strategy adopted by the row player are fixed. Also notice that each of those coefficients is the expected value of the corresponding column in the payoff matrix:

$$\sum_{i=1}^{n^{(r)}} \xi_i^{(r)} f_{ij} = E\left[f_{\cdot j}\left(\xi^{(r)}\right)\right]$$

that is the expected value of the payoff obtained by the row player against each of the pure strategies of the column player.

Now let us determine the worst case with respect to the strategy $\xi^{(r)}$ that we have fixed. In order to do that, one must solve the optimisation problem:

$$\phi\big(\xi^{(r)}\big) = \min_{\xi^{(c)} \in \Xi^{(c)}} \sum_{j=1}^{n^{(c)}} E\left[f_{\cdot j}\left(\xi^{(r)}\right)\right] \xi_j^{(c)}$$

$$\sum_{j=1}^{n^{(c)}} \xi_j^{(c)} = 1$$

$$\xi_j^{(c)} \geq 0$$

This problem consists in minimising a convex combination of $n^{(c)}$ numbers. Ths optimal solution of this problem is trivially the minimum of those numbers:

$$\xi_j^{\circ(c)}\big(\xi^{(r)}\big) = \begin{cases} 1 \text{ for } j = \arg \min\limits_{j \in X^{(c)}} E\left[f_{\cdot j}\left(\xi^{(r)}\right)\right] \\ 0 \text{ otherwise} \end{cases}$$

which consists in determining the worst impact and fixing to 1 the probability of the adversary strategy that produces it. The corresponding value is:

$$\phi\big(\xi^{(r)}\big) = \min_{j \in X^{(c)}} E\left[f_{\cdot j}\left(\xi^{(r)}\right)\right]$$

Now, the value of the game for the row player is obtained fixing the mixed strategy that maximises $\phi\big(\xi^{(r)}\big)$, which implies the thesis. With the same process, one obtains the value of the game for the column player. ■

This proposition shows that, assuming the prudential point of view of the worst-case criterium, one should give for granted that the adversary will behave in a deterministic way, selecting the pure strategy that is most damaging for the chosen mixed strategy, instead of a stochastic way, selecting a combination of strategies. In fact, if a pure strategy is more damaging than the other ones, even if the adversary could soften it combining it with the latter, in a worst-case perspective it would not make sense to do that.

The minimax theorem

Theorem 25 *For any two-person zero-sum finite game, the expected values of the game for the row and the column player are equal. Moreover, there exists a pair of mixed strategies* $\big(\xi^{*(r)}, \xi^{*(c)}\big) \in \Xi^{(r)} \times \Xi^{(c)}$ *such that*

$$v^{(r)} = v^{(c)} = E\left[f\left(\xi^{*(r)}, \xi^{*(c)}\right)\right]$$

These two strategies are optimal *for the players, that is*

$$v^{(r)} = \min_{\xi^{(c)} \in \Xi^{(c)}} E\left[f\left(\xi^{*(r)}, \xi^{(c)}\right)\right]$$
$$v^{(c)} = \max_{\xi^{(r)} \in \Xi^{(r)}} E\left[f\left(\xi^{(r)}, \xi^{*(c)}\right)\right]$$

Proof. We will prove this theorem exploiting the relation that links the theory of zero-sum games and the theory of duality in Linear Programming. First consider the point of view of the row player. This player defines the probability vector $\xi^{(r)}$ which defines his/her mixed strategy so as to maximise the expected utility in the worst-case. From a mathematical point of view, this problem can be represented as a maximisation problem in the decision variables $\xi_i^{(r)}$ ($i = 1, \ldots, n^{(r)}$):

$$\begin{aligned} &\max \phi\big(\xi^{(r)}\big) \\ &\sum_{i=1}^{n^{(r)}} \xi_i^{(r)} = 1 \\ &\qquad \xi_i^{(r)} \geq 0 \qquad i = 1, \ldots, n^{(c)} \end{aligned}$$

with $\phi\big(\xi^{(r)}\big) = \min\limits_{j \in X^{(c)}} \sum\limits_{i=1}^{n^{(r)}} f_{ij} \xi_i^{(r)}$.

This problem has linear constraints on variables $\xi_i^{(r)}$, but the objective function is nonlinear, because it includes the minimisation operator (min). However, it is possible to linearise the problem with a simple trick. First, add a new variable v that describes the utility in the worst case. Then, impose that v does not exceed none of the terms $\sum_{i=1}^{n^{(r)}} f_{ij} \xi_i^{(r)}$. In this way, v could be smaller than all those terms, but since the objective function tends to maximise v, in any optimal solution at least one of the inequalities will be an equality, that is, v will be equal to the minimum term.

In this way, one obtains an equivalent Linear Programming problem (both the constraints and the objective function are linear):

$$\begin{aligned} &\max_{v, \xi^{(r)}} \phi\big(v, \xi^{(r)}\big) = v \\ &\qquad\qquad \sum_{i=1}^{n^{(r)}} \xi_i^{(r)} = 1 \\ &\qquad v - \sum_{i=1}^{n^{(r)}} f_{ij} \xi_i^{(r)} \leq 0 \qquad j = 1, \ldots, n^{(c)} \\ &\qquad\qquad\quad \xi_i^{(r)} \geq 0 \qquad i = 1, \ldots, n^{(r)} \end{aligned}$$

The problem for the column player is similar, except that we must minimise a loss. Applying a similar trick to linearise the objective function, in order to remove a maximisation operator (max):

$$\begin{aligned}\min_{w,\xi^{(c)}} \psi\big(w,\xi^{(c)}\big) &= w \\ \sum_{j=1}^{n^{(c)}} \xi_j^{(c)} &= 1 \\ w - \sum_{j=1}^{n^{(c)}} f_{ij}\xi_j^{(c)} &\geq 0 \qquad i = 1,\ldots,n^{(r)} \\ \xi_j^{(c)} &\geq 0 \qquad j = 1,\ldots,n^{(c)}\end{aligned}$$

The two problems are reciprocally dual, as:

1. one is a maximisation problem and the other is a minimisation problem;
2. the objective coefficient vector of each problem of one problem is the right-hand-side vector of the other: both consist of a 1 accompanied by a series of 0;
3. the coefficient matrix of each problem is equal to the transpose of the matrix of the other problem: the first row consists of a 0 and a series of 1, the following ones by a 1 and the coefficients f_{ij} with a reversed sign; since one of the sums is performed on index i and the other on index j, the constraints in one of the two problems scan a column of matrix F and those in the other a row.

The theory of duality in Linear Programming guarantees that both problems have an optimal solution, and that the values of the two optima are the same, that is the maximum the row player can gain is equal to the minimum that column player can lose. This is exactly the statement of the thesis. ■

Von Neumann's minimax theorem has three main consequences:

1. the mixed strategies offer to both players better guarantees with respect to the pure strategies[1]. Trivially, since the pure strategies are special cases of mixed strategies in which one has probability 1 and the other probability 0, the mixed strategies cannot perform worse.
2. the optimal mixed strategies have the same value for the two players, and give rise to a (mixed) equilibrium, that is better than the values of the game for the pure strategies:
$$u^{(r)} \leq v^{(r)} = E\left[f\big(\xi^{\circ(r)},\xi^{\circ(c)}\big)\right] = v^{(c)} \leq u^{(c)}$$
3. if the game has a saddle point f_{rc}, then an optimal mixed strategy profile coincides with the pure strategy profile in which the players choose the row and the column associated to the saddle[2]:
$$\exists \text{ saddle point } x^\circ = \big(x^{\circ(r)},y^{\circ(r)}\big) \Rightarrow \begin{cases}\xi_i^{\circ(r)} \in \{0,1\} \text{ for } i \in X^{(r)} \\ \xi_j^{\circ(c)} \in \{0,1\} \text{ for } j \in X^{(c)} \\ u^{(r)} = v^{(r)} = v^{(c)} = u^{(c)}\end{cases}$$

[1] It is true that the guarantee provided by the pure strategies is deterministic, whereas that provided by the mixed strategies concerns the expected values.

[2] The optimal solution can be nonunique, as in general is nonunique the optimal solution of a Linear Programming problem. It can also happen that several equilibria in pure strategies exist. In that case, we have seeen that they all provide to the players the same *payoff*, that is equal to that gained from the mixed strategies obtained combining the pure strategies with probabilities summing to 1.

13.5.1 Computation of the equilibrium mixed strategy

If the game is particularly simple, that is each of the two players has only two strategies, the search for the equilibrium mixed strategies requires to solve optimisation problems with only two variables (the probabilities $\xi_1^{(r)}$ and $\xi_2^{(r)}$ for the row player and the probabilities $\xi_1^{(c)}$ and $\xi_2^{(c)}$ for the column player). Since the sum of such probabilities is 1, they are actually simple one-dimension problems, which can be solved graphically. In the following, we shall see some examples.

Example 84 *The "Odds and evens" game is a two-player zero-sum game in which each player has two strategies. The payoff matrix, reported in Table 13.2, does not contain equilibria.*

	O	E
O	1	-1
E	-1	1

Table 13.2: Payoff matrix for the "Odds and evens" game

We know that the worst case for any strategy is determined by a pure strategy of the adversary. Let us assign to the row player a generic mixed strategy $\xi^{(r)} = [\alpha\ (1-\alpha)]^T$, and let use evaluate the results of the two possible pure strategies of the adversary:

- $E\left[f\left(\xi^{(r)}, O\right)\right] = \alpha \cdot 1 + (1-\alpha)\cdot(-1) = 2\alpha - 1$;
- $E\left[f\left(\xi^{(r)}, E\right)\right] = \alpha \cdot (-1) + (1-\alpha)\,1 = 1 - 2\alpha$.

The problem for the row player is to maximise the worst of two results:

$$v^{(r)} = \max_{\alpha\in[0;1]} \min\left(2\alpha - 1, 1 - 2\alpha\right)$$

The left side of Figure 13.3 shows the graphic resolution of the problem for the row player: for each mixed strategy, that is for each value of α, we consider the two pure strategies of the column player, represented by two straight lines. The mixed strategies for the column players would fall inside the area delimited by the two lines, while function $\phi^{(r)}\left(\xi^{(r)}\right)$, which represents the worst case, corresponds to the lower envelope of this area, depicted with a continuous line. Now, the row player must decide his/her optimal mixed strategy $\xi^{\circ(r)}$, that is given by the maximum of the piecewise linear function $\phi^{(r)}\left(\xi^{(r)}\right)$. Therefore, the optimal strategy is $\xi^{\circ(r)} = [1/2\ 1/2]^T$, and the value of the game is zero. In fact, the best strategy for the game is to play odds half the time and even the other half; choosing one of the two strategies more often than the other exposes to the risk of being predicted and anticipated. The expected gain of the optimal strategy is zero.

The problem of the column player can be solved considering the generic mixed strategy $\xi^{(c)} = [\beta\ (1-\beta)]^T$ and evaluating its worst case. Once again, one can consider only the two pure strategies of the adversary:

- $E\left[f\left(O, \xi^{(c)}\right)\right] = \beta \cdot 1 + (1-\beta)\cdot(-1) = 2\beta - 1$;
- $E\left[f\left(E, \xi^{(c)}\right)\right] = \beta \cdot (-1) + (1-\beta)\cdot 1 = 1 - 2\beta$;

given that the mixed ones have intermediate results. Therefore, the column player must minimise the worst of the two results:

$$v^{(c)} = \min_{\beta\in[0;1]} \max\left(2\beta - 1, 1 - 2\beta\right)$$

The right side of Figure 13.3 shows the graphical resolution of the problem: the objective function is represented by the upper envelope of the region delimited by the two straight lines. The point of maximum is identified by the intersection of the two lines, which corresponds to the optimal strategy $\xi^{*(c)} = [1/2\ 1/2]^T$*, for which the value of the game is zero.*

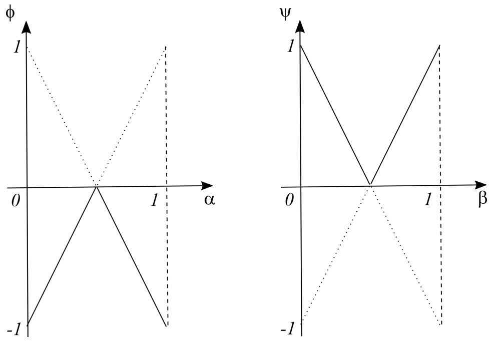


Table 13.3: Graphical resolution of the "Odds and evens" game for the row player (on the left) and the column player (on the right)

Example 85 *Consider the payoff matrix reported in Table 13.4.*

	1	2
1	2	-3
2	-1	1

Table 13.4: Payoff matrix for a zero-sum game

The mixed strategies for the row player can be represented as $\xi^{(r)} = [\alpha\ 1-\alpha]^T$ *con* $\alpha \in [0,1]$*. For every value of alpha, the worst case corresponds to the worst of the two pure strategies available for the column player represented on the left side of Figure 13.3 by the two lines:*

- $E\left[f\left(\xi^{(r)},1\right)\right] = \alpha \cdot 2 + (1-\alpha)\cdot(-1) = 3\alpha - 1$*;*
- $E\left[f\left(\xi^{(r)},2\right)\right] = \alpha \cdot (-3) + (1-\alpha)\,1 = 1 - 4\alpha$*.*

Now, for every mixed strategy $\xi^{(r)}$*, that is for every* α*, the worst case is represented by the lower envelope of the region delimited by the two lines, depicted in continuous lines. The row player must decide the mixed strategy that maximises this result:*

$$v^{(r)} = \max_{\alpha\in[0,1]} \min\left(3\alpha - 1, 1 - 4\alpha\right)$$

Such a strategy is identified by the intersection of the two lines, for which $\xi^{\circ(r)} = [2/7\ 5/7]^T$ *and the value of the game is* $v^{(r)} = -1/7$*. With the same process, on the right side of Figure 13.3 we compute that* $\xi^{\circ(c)} = [4/7\ 3/7]^T$ *and* $v^{(c)} = -1/7$*.*

13.5.2 Metodo del pivot*

C'è un altro modo per trasformare il problema in un programma lineare che rende più semplice i calcoli nel caso in cui il valore del gioco sia positivo. D'altra parte,

*Questa è una sezione di approfondimento, che non fa parte del programma d'esame.

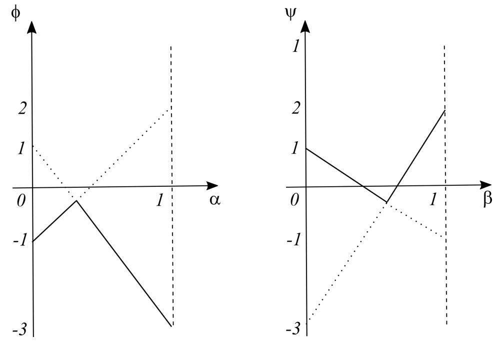


Figure 13.3: Graphical resolution of the game of Table 13.4 for the row player (on the left) and the column player (on the right)

se si aggiunge una costante a tutti i payoff della matrice, il gioco sostanzialmente non cambia, e si può certamente rendere positivo il suo valore. Ipotizziamo quindi che sia $v > 0$ e riformuliamo il problema del giocatore di riga:

$$\begin{aligned} &\max_{v,\xi^{(r)}} \phi\big(v,\xi^{(r)}\big) = v \\ &\sum_{i=1}^{n^{(r)}} \xi_i^{(r)} = 1 \\ &v - \sum_{i=1}^{n^{(r)}} f_{ij}\xi_i^{(r)} \leq 0 \qquad j = 1,\dots,n^{(c)} \\ &\xi_i^{(r)} \geq 0 \qquad i = 1,\dots,n^{(r)} \end{aligned}$$

Introduciamo le variabili ausiliarie $z_i = \xi_i^{(r)}/v$, da cui $\xi_i^{(r)} = vz_i$.

$$\begin{aligned} &\max_{v,z} \phi\big(v,z\big) = v \\ &\sum_{i=1}^{n^{(r)}} vz_i = 1 \\ &v - \sum_{i=1}^{n^{(r)}} f_{ij}vz_i \leq 0 \qquad j = 1,\dots,n^{(c)} \\ &z_i \geq 0 \qquad i = 1,\dots,n^{(r)} \end{aligned}$$

Il vincolo $\sum_{i=1}^{n^{(r)}} z_i = 1/v$ non è lineare, ma possiamo sfruttare il fatto che massimizzare v equivale a minimizzare $1/v$, per rimuovere completamente v dal problema minimizzando al suo posto $\sum_{i=1}^{n^{(r)}} z_i$. Negli altri vincoli, infatti, v compare come fattore moltiplicativo a entrambi i membri, e quindi si può semplificare:

$$\begin{aligned} &\max_{z} \phi\big(v,z\big) = \sum_{i=1}^{n^{(r)}} z_i \\ &\sum_{i=1}^{n^{(r)}} f_{ij}z_i \geq 1 \qquad j = 1,\dots,n^{(c)} \\ &z_i \geq 0 \qquad i = 1,\dots,n^{(r)} \end{aligned}$$

Risolto questo problema, la soluzione del gioco originale può essere facilmente trovata tenendo conto che $v = 1/\sum_{i=1}^{n^{(r)}} z_i^\circ$ e $\xi_i^{(r)} = vz_i^\circ$. Per il giocatore di colonna, si può applicare un procedimento del tutto analogo.

Per risolvere il problema, si può applicare un metodo equivalente all'algoritmo del simplesso[3]

Step 1 Se necessario, aggiungere una costante a tutti gli elementi della matrice di utilità affinchè il valore del gioco sia positivo (a questo punto, il valore del gioco originale è uguale a quello del nuovo gioco meno la costatnte aggiunta).

Step 2 Creare un *tableau* aumentando la matrice delle utilità (coefficienti delle variabili x nei vincoli) con una riga di -1 in alto (coefficienti della funzione obiettivo cambiati di segno) e una colonna di 1 all'estrema sinistra (termini noti dei vincoli). Nell'angolo in alto a sinistra porre 0 (termine noto della funzione obiettivo). Aggiungere una colonna a destra con le etichette da x_1a x_m (variabili fuori base) e una riga in basso con le etichette da y_1a y_n (variabili basiche). Sia $A = [a_{ij}]$ la nuova tabella.

0	-1	-1	$\cdots$	-1	
1	a_{11}	a_{12}	$\cdots$	a_{1n}	x_1
1	a_{21}	a_{22}	$\cdots$	a_{2n}	x_2
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
1	a_{m1}	a_{m2}	$\cdots$	a_{mn}	x_m
	y_1	y_2	$\cdots$	y_n	

Step 3 Selezionare un elemento a_{pq}, detto *pivot*, tale che

1. $q \in \{j : a_{0,j} < 0, 1 \le j \le n\}$
2. $p = \arg\min_{1 \le i \le m, a_{i,q} > 0} \left\{ \frac{a_{i,0}}{a_{i,q}} \right\}$.

Nota: il pivot esiste sempre perchè il problema è limitato sotto le condizioni espresse dal teorema minmax.

Step 4 Fare pivot nel seguente modo:

1. $a_{ij} := a_{ij} - a_{pj} \cdot a_{iq}/a_{pq} \quad i = 1, ..., m \; j = 1, ..., n \; i \neq p \; j \neq q$
2. nella riga del pivot, ad eccezione del pivot, $a_{pj} := \frac{a_{pj}}{a_{pq}} \quad j = 1, ..., n \; j \neq q$
3. nella colonna del pivot, ad eccezione del pivot, $a_{iq} := -\frac{a_{iq}}{a_{pq}} \quad i = 1, ..., m \; i \neq p$
4. $a_{pq} := \frac{1}{a_{pq}}$.

Step 5 Scambiare l'etichetta a destra nella riga del *pivot* (la variabile esce di base) con l'etichetta in basso nella colonna del *pivot* (la variabile entra in base), *swap* $A_{p,n+1}$ $A_{m+1,q}$.

[3] La parte seguente non è stata ancora rivista.

Step 6 Se rimane qualche elemento negativo nella riga zero, tornare allo *step 3*.

Step 7 Altrimenti, ricavare la soluzione nel seguente modo:

1. il valore del gioco v è il reciproco del numero nell'angolo in alto a sinistra (valore della funzione obiettivo all'ottimo)

2. la strategia ottima del giocatore riga è così costruita: le variabili x_i che si trovano nella colonna a destra ricevono probabilità 0 (variabili fuori base), mentre quelle sulla riga in basso una probabilità pari a $v \cdot a_{0,i}$ (variabili basiche)

3. la strategia ottima del giocatore colonna è così costruita (soluzioni duali del problema appena risolto): le variabili $y_j$che si trovano nella riga in basso ricevono probabilità 0, mentre quelle sulla colonna a destra una probabilità pari a $v \cdot a_{j,0}$.

Example 86 *Consideriamo il gioco descritto dalla matrice*

$$B = \begin{bmatrix} 2 & -1 & 6 \\ 0 & 1 & -1 \\ -2 & 2 & 1 \end{bmatrix}$$

Come si può notare non esiste un punto di sella e quindi nessuna strategia pura ottimale. Il valore del gioco è positivo? Forse. In ogni caso è più semplice sommare 2 ad ogni elemento e ottenere la matrice.

$$B' = \begin{bmatrix} 4 & 1 & 8 \\ 2 & 3 & 1 \\ 0 & 4 & 3 \end{bmatrix}$$

Il valore di questo gioco è almeno 1 dal momento che il giocatore riga si può garantire questa vincita scegliendo la riga 1 o 2. Questo completa lo step *1.*

Nello step *2, prepariamo un* tableau *per B' come il seguente:*

0	-1	-1	-1	
1	4	1	8	x_1
1	2	3	1	x_2
1	0	4	3	x_3
	y_1	y_2	y_3	

Nello step *3, dobbiamo scegliere il* pivot*. Poichè tutte e tre le colonne hanno nella prima riga un valore negativo, possiamo sceglierne una qualunque come colonna del pivot. Prendiamo la prima. La riga del pivot deve avere un numero positivo su questa colonna, così deve essere una delle prime due righe. Per decidere quale è, calcoliamo il rapporto tra il numero nella prima colonna e il candidato pivot. Per la prima riga è 1/4, per la seconda 1/2. Il più piccolo è il primo quindi scegliamo 4 come pivot.*

0	-1	-1	-1	
1	**4**	1	8	x_1
1	2	3	1	x_2
1	0	4	3	x_3
	y_1	y_2	y_3	

Nello step *4, dopo aver combinato linearmente le righe del tableau, otteniamo*

$\frac{1}{4}$	$\frac{1}{4}$	$-\frac{3}{4}$	1	
$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	2	x_1
$\frac{1}{2}$	$-\frac{1}{2}$	$\frac{5}{2}$	-3	x_2
1	0	4	3	x_3
	y_1	y_2	y_3	

Nello step *5, scambiamo le etichette della riga e della colonna del pivot, cioè* x_1 *e* y_1*.*

$\frac{1}{4}$	$\frac{1}{4}$	$-\frac{3}{4}$	1	
$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	2	y_1
$\frac{1}{2}$	$-\frac{1}{2}$	$\frac{5}{2}$	-3	x_2
1	0	4	3	x_3
	x_1	y_2	y_3	

Nello step *6, controlliamo che ci siano valori negativi nella prima riga. Poichè ci sono torniamo allo* step *3.*

Rifacendo tutti i passaggi precedenti otteniamo

0.4	0.1	0.3	0.1	
0.2	0.3	-0.1	2.3	y_1
0.2	-0.2	0.4	-1.2	y_2
0.2	0.8	-1.6	7.8	x_3
	x_1	x_2	y_3	

Nello step *7, possiamo leggere la soluzione del gioco con la matrice B':*

- *il valore del gioco è* 2.5.
- *la strategia ottima per il giocatore riga è* $\mathbf{p} = (0.25, 0.75, 0)$
- *la strategia ottima per il giocatore colonna è* $\mathbf{q} = (0.5, 0.5, 0)$

Il gioco con matrice B ha le stesse strategie ottime, ma valore $2.5 - 2 = 0.5$.

Chapter 14

Symmetric games

Another special case of particular importance in the distribution of utility among players is the case in which the players are perfectly interchangeable, that is, have the same available strategies and, if they exchange their strategies, they correspondingly exchange the same payoffs.

Definition 59 *We denote as* symmetric game *a game in which all players have the same number of pure strategies* $n = \left|X^{(d)}\right|$ *(for all* $d \in D$*) and, for every permutation* $p = \left(p_1, \ldots, p_{|D|}\right)$ *of the players, if the strategies adopted are permuted according to* p*, also the resulting payoffs are permuted according to* p*:*

$$f^{(d)}\left(x^{(1)}, \ldots, x^{(|D|)}\right) = f^{(p_d)}\left(x^{(p_1)}, \ldots, x^{(p_{|D|})}\right) \text{ for all } d \in D$$

Intuitively, this means that the results of the strategies do not depend on the players who apply them: exchanging the strategies among the players, the results are exchanged in the same way. In practice, this typically corresponds to the fat that each player has strategies somehow corresponding *one-by-one* to those of any other player, and that it is therefore reasonable to give them the same names and order. Therefore, the first strategy of each player corresponds in some sense to the first strategy of each other player, the second to the second, and so on. They do not necessarily represent the same physical operations, but they are operations producing the same results.

In the following, we consider the case of the two-person symmetric games. In this case, there exist only two possible permutations of the players, so that the definition is strongly simplified:

$$f_{ij}^{(r)} = f_{ji}^{(c)} \text{ for all } i, j = 1, \ldots, n$$

If the row player obtains the gain $f_{ij}^{(r)}$ applying strategy i while the column player applies strategy j, when the two players exchange their strategies and the row player applies j while the column the column player applies i, the column player will obtain exactly the same gain in the new situation.

Remark 8 *In a two-person symmetric game, the payoff matrix of the second player is the transpose of the payoff matrix of the first player.*

Contrary to the case of the zero-sum games, even if the information associated to the second player is redundant, usually one adopts the complete representation. This is because the missing information is not simply the opposite of the reported one, but should be retrieved in the symmetric cell of the matrix, with a nontrivial process. Table 14.1 reports a sample two-person symmetric game.

	1	2	3
1	(2,2)	(3,0)	(5,2)
2	(0,3)	(4,4)	(-1,8)
3	(2,5)	(8,-1)	(6,6)

Table 14.1: Sample two-person symmetric game: the payoff matrices of the two players are the transpose of each other

14.1 A taxonomy of two-person two-strategy symmetric games

An extremely simplified case, and yet a very interesting one in practice, is that of two-person games in which each player has only two available strategies. It is possible to classify such games based on two different methodologies, one of which leads to 4 classes, whereas the othe leads to 12 classes. The two methodologies are largely consistent, meaning that the latter adds a further specification to the former, splitting some of its 4 classes into subclasses. In the following, we will present the two classifications in abstract terms, and then we will discuss some games to which the literature has given appealing symbolic names, explaining their fundamental characteristics in terms of the two classifications.

14.1.1 Classification based on Nash equilibria

	1	2
1	(f_{11}, f_{11})	(f_{12}, f_{21})
2	(f_{21}, f_{12})	(f_{22}, f_{22})

Table 14.2: General two-person two-strategy symmetric game: the payoff matrices of the two players are the transpose of each other

Given the general payoff matrix in Table 14.2, and excluding the limit cases in which some payoff coincide, there are only four possible situations. For each, we evaluate whether it admits dominated strategies and we determine the equilibria, if any exist:

1. $f_{11} > f_{21}$ and $f_{12} > f_{22}$: the first strategy strictly dominates the second for both players; there is a single Nash equilibrium in point $(1, 1)$;

2. $f_{11} > f_{21}$ and $f_{12} < f_{22}$: there is no dominated strategy, and there are two Nash equilibria lying on the main diagonal in points $(1, 1)$ and $(2, 2)$;

3. $f_{11} < f_{21}$ and $f_{12} > f_{22}$: there is no dominated strategy, and there are two Nash equilibria lying on the secondary diagonal in point $(1, 2)$ and $(2, 1)$;

4. $f_{11} < f_{21}$ and $f_{12} < f_{22}$: the second strategy strictly dominates the first for both players; there is a single Nash equilibrium in point $(2, 2)$.

As it can be noticed, *two-person and two-strategy symmetric games have always at least one equilibrium.* This is a remarkable characteristic, given that in general this is not true for asymmetric games, nor for symmetric games with more than two strategies (for instance, the rock-scissors-paper game). We will see in the following, considering some relevant examples, what practical meaning these equilibria assume.

Class	Subclass	Order	Examples
1	a	$f_{11} > f_{12} > f_{21} > f_{22}$	Ideal marriage
	b	$f_{11} > f_{12} > f_{22} > f_{21}$	
	c	$f_{11} > f_{21} > f_{12} > f_{22}$	
	d	$f_{12} > f_{11} > f_{21} > f_{22}$	
	e	$f_{12} > f_{11} > f_{22} > f_{21}$	
2	a	$f_{11} > f_{21} > f_{22} > f_{12}$	Stag hunt
	b	$f_{11} > f_{22} > f_{12} > f_{21}$	Coordination (1)
	c	$f_{11} > f_{22} > f_{21} > f_{12}$	Coordination (2)
3	a	$f_{21} > f_{11} > f_{12} > f_{22}$	Chicken's game
	b	$f_{12} > f_{21} > f_{11} > f_{22}$	Battle of the sexes (1)
	c	$f_{21} > f_{12} > f_{11} > f_{22}$	Battle of the sexes (2)
4	a	$f_{21} > f_{11} > f_{22} > f_{12}$	Prisoner's dilemma

Table 14.3: List of the 12 possible two-player two-strategy symmetric games according to the classification based on the order of the payoffs and their relation with the classification based on the Nash equilibria

14.1.2 Classification based on the order of the payoffs

A more detailed way to classify two-person two-strategy symmetric games is based on the relative order of the four independent payoffs, f_{11}, f_{12}, f_{21} and f_{22}. In this classification, there are as many games as permutations of the four values, that is $4! = 24$. In practice, however, it is possible to reduce these classes to 12 by remarking that the names of strategies 1 and 2 are completely conventional, so that there is no loss of generality in the additional assumption that $f_{11} > f_{22}$. If the given game does not satisfy this condition, it is enough to exchange the names of the two strategies in order to introduce it.

Table 14.1.2 lists the 12 possible classes according to this taxonomy, specifying the order of the payoffs and the class of the first taxonomy to which each of them belongs, as well as some famous names of games described in the literature which fall inside each subclass.

The following sections present six important games, chosen among the 12 categories. In order to make the description more concrete, instead of using symbolic payoffs, we will attribute to the impacts the conventional values 0, 1, 2 and 3. For the same reason, we will attribute to the two strategies the conventional names of *cooperation* (C) for the former and *noncooperation* (NC) for the latter. This is consistent with the assumption that two cooperating strategies have a payoff f_{11} larger than the payoff f_{22} of two noncooperating strategies.

14.2 The ideal marriage

This game corresponds to the order

$$f_{11} > f_{21} > f_{12} > f_{22}$$

and is represented in Table 14.4.

This game describes the situation in which cooperation increases the gain of the two players, as in an ideal marriage. Going into details, the order among the possible situations is:

1. if both spouses/players cooperate (C, C), both gain the best possible result;

	C	NC
C	$(\bar{3},\bar{3})$	$(\bar{1},2)$
NC	$(2,\bar{1})$	$(0,0)$

Table 14.4: Payoff matrix for the ideal marriage: the markings on the best payoffs for each row and column indicate that point (C,C) is the only Nash equilibrium

2. if the first player does not cooperate, while the other one cooperates (NC,C), he/she gains the second best possible result;
3. if the first player cooperates, while the othe one does not cooperate (C,NC), he/she gains the third result;
4. if both players do not cooperate (NC,NC), both gain the worst result.

The strategy of not cooperating (NC) is dominated by the strategy of cooperating (C) for both players. This means that it is reasonable for both to cooperate, and in fact the game has a single Nash equilibrium, that corresponds to the mutual cooperation (in Table 14.4, both payoffs are marked).

Other four games belong to class 1 (games with a single cooperative equilibrium); they are rather similar to each other, but have not inspired fancy names to the literature, probably because the correspond to situations in which cooperating with a noncooperating adversary is more profitable than non cooperating with a cooperating adversary ($f_{12} > f_{21}$). It seems unlikely that such situations arise spontaneously, without an external intervention.

14.3 The stag hunt

This game corresponds to the order

$$f_{11} > f_{21} > f_{22} > f_{12}$$

and is represented in Table 14.5.

	C	NC
C	$(\bar{3},\bar{3})$	$(0,2)$
NC	$(2,0)$	$(\bar{1},\bar{1})$

Table 14.5: Payoff matrix for the stag hunt: the markings on the best payoffs for each row and column indicate that points (C,C) and (NC,NC) are both Nash equilibria

This game derives from a passage of the *Discourse on the Origin and Basis of Inequality Among Men* (1755) by Rousseau[1], according to which human societies derive from the temporary alliances among human beings for hunts to large animals, which a single individual could not catch. Assuming that two players take part to the hunt, the order of the possible situations is:

1. if both cooperate (C,C), it is quite likely that they succeed in catching a stag, which is the best possible result for both;

[1] Jean-Jacques Rousseau (1712-1778), Swiss philosopher, writer and musician.

2. if one of them does not cooperate, while the other cooperates, the noncooperating player can dedicate time to catch a minor prey (a hare) and keep a small hope to catch also the stag with the residual effort, thus obtaining the second best result;
3. if both players do not cooperate (NC, NC), both can catch a hare, but the probability that they also catch a stag becomes approximately zero, so that this is the third best possible result for both;
4. finally, if one of the two players cooperates, while the other does not, the cooperating player has only a very small probability to catch the stag, which is the worst possible result.

As Table 14.5 shows, neither strategy is dominated, and the game has two Nash equilibria, which correspond to mutual cooperation and mutual noncooperation. In fact, if one can be sure that the other player will cooperate, it is more profitable not to abandon the cooperative equilibrium, because it leads to catch the stag. On the contrary, if one can be sure that the other player will not cooperate, it is more profitable not to abandon the noncooperative equilibrium, in order to keep at least the capture of a hare. There are therefore two symmetric possibilities. Mutual cooperation is a desirable equilibrium, because it corresponds to higher gains for both players. Mutual noncooperation, however, is also an equilibrium, though less desirable for both, and it is conservative, meaning that it obeys the worst-case criterium. Therefore, both strategies are rational, and the choice between the will be taken based on other factors not modelled in the basic form of the game, such as the amount of trust in the other player.

14.4 Pure coordination games

These games correspond to the two orders:

$$f_{11} > f_{22} > f_{12} > f_{21} \quad \text{and} \quad f_{11} > f_{22} > f_{21} > f_{12}$$

and are represented in Table 14.6. Of course, the intermediate case in which $f_{12} = f_{21}$ falls in the same category, and is actually the most common representation of this class of games. These games describe the situations in which the best results are obtained when both players choose the same strategy, whereas the asymmetric strategies are damaging for both. In this sense, the labels *cooperation* and *noncooperation* are inadequate labels, and in fact are not used in the table.

	1	2
1	$(\bar{3},\bar{3})$	$(0,1)$
2	$(1,0)$	$(\bar{2},\bar{2})$

	1	2
1	$(\bar{3},\bar{3})$	$(1,0)$
2	$(0,1)$	$(\bar{2},\bar{2})$

Table 14.6: Payoff matrix for the pure coordination games: the markings on the best responses for each row and column show that points $(1,1)$ and $(2,2)$ are both Nash equilibria

A classical example is that of two drivers who meet each other, coming from opposite directions, along a narrow road, with no centreline dividing the two lanes. If both keep on the right side, or on the left side, they cross each other without problems. One of the two strategy profiles is slightly better (for example, if both drivers are used to drive on the right side, it will be more natural for both to keep on the right). By contrast, if one of the drivers keeps on his/her own right and the other on the left, the result is a head-on collision, that is a bad outcome for both. Also in this case, one could consider one of the two situations are slightly better than the other, but anyway worse than in the case of coordinated strategies.

The pure coordination games have no dominated strategies and have the same two Nash equilibria as the stag hunt. The difference is that the contrast between the more desirable equilibrium and the less desirable one is less pronounced: they are indeed less interesting situations.

14.5 The chicken race

This game corresponds to the order:

$$f_{21} > f_{11} > f_{12} > f_{22}$$

and is represented in Table 14.7.

	C	NC
C	$(2, 2)$	$(\bar{1}, \bar{3})$
NC	$(\bar{3}, \bar{1})$	$(0, 0)$

Table 14.7: Payoff matrix for the chicken race: the markings on the best response for each row and column indicated that the points (NC, C) and (C, NC) are both Nash equilibria

This game derives from the movie *Rebel Without a Cause* (1955): two youngsters challenge each other driving stolen cars towards a cliff; the first one jumping out of his car will prove a "chicken", that is a coward, and will lose the challenge. An alternative name for the same game is *hawks-and-doves*, a phrase coined by Russell[2] in his book *Common sense and nuclear warfare* (1959), to describe the situation of political tension between USA and URSS. In these challenges (military, political, economical, etc. . .), one of the two competitors decides to go the end hoping that the adversary will give up, risking a disaster for both.

The order of the possible situations is:

- if a player does not cooperate and the other does (NC, C), that is the former persists and the latter gives up, the noncooperating player wins the challenge, and obtains the best possible result;
- if both players cooperate (C, C), that is both give up, both get the second best possible result, that is an honourable tie;
- if a player cooperates, that is gives up, whereas the other does not, the cooperating one obtains the third possible result, that is to be publicly disgraced, but save one's life;
- if both players do not cooperate (NC, NC), that is they persist, both lose their lives, which is the worst possible result for both.

As Table 14.7 shows, neither strategy is dominated, and the game has two Nash equilibria, that correspond to complementary strategies, with one cooperating player and one noncooperating player. In fact, if one of the two player can be sure

[2] Bertrand Arthur William Russell (1872-1970), Welsh philosopher and mathematician.

that the other will cooperate, it is profitable not to cooperate. Conversely, if one can be sure that the other player will not cooperate, it is profitable to cooperate. The worst-case criterium suggests to both players to cooperate, but the resulting solution is not an equilibrium, and therefore tends not to preserve if the game is repeated along time. In spite of having two Nash equilibria, the game admits no rational strategy.

14.6 Battle of the sexes

This game corresponds to the orders:

$$f_{12} > f_{21} > f_{11} > f_{22} \quad \text{and} \quad f_{21} > f_{12} > f_{11} > f_{22}$$

and is represented in Table 14.8.

	1	2
1	$(1,1)$	$(\bar{3},\bar{2})$
2	$(\bar{2},\bar{3})$	$(0,0)$

	1	2
1	$(1,1)$	$(\bar{2},\bar{3})$
2	$(\bar{3},\bar{2})$	$(0,0)$

Table 14.8: Payoff matrix for the battle of the sexes: the markings of the best response for each row and column indicate that points $(1,2)$ and $(2,1)$ are both Nash equilibria

As in the pure coordination games, the labels "cooperation" and "noncooperation" are inadequate, and in fact are not used. The game can be exemplified as follows. Two lovers have decided to spend the day together. They animatedly discuss whether to go to a soccer match (as preferred by him) or a ballet (as preferred by her). The call is cut off before an agreement with no possibility to establish it again, so that each one must decide what to do independently. Both prefer to spend time together rather than staying alone, even enjoying the favourite entertainment. The order of the possible situations is the following:

- the best result is obtained attending the favourite entertainment together with the partner, that is in case $(1,2)$ for the first player and in case $(2,1)$ for the second;
- the second best result is obtained attending the less favourite entertainment together with the partner, that is in case $(2,1)$ for the first player and in case $(1,2)$ for the second;
- the third possible result is obtained attending separately one's favourite entertainment;
- the worst possible result, finally, is obtained attending separately one's less favourite entertainment.

This game has no dominated strategy, and has two Nash equilibria, corresponding to the asymmetric strategies, as in the chicken race. The difference is that the two equilibria are fundamentally positive for both players, instead of being very positive for one and very negative for the other.

14.7 The prisoner's dilemma

This game corresponds to the order:

$$f_{21} > f_{11} > f_{22} > f_{21}$$

and is represented in Table 14.9.

	C	NC
C	$(2,2)$	$(0,\bar{3})$
NC	$(\bar{3},0)$	$(\bar{1},\bar{1})$

Table 14.9: Payoff matrix for the prisoner's dilemma: the markings on the best response for each row and column indicate that point (NC, NC) is the only Nash equilibrium

This game derives from a 1950 paper by Flood[3] and Dresher[4] supported by the *Rand Corporation* (*Applications to a global nuclear strategy*). The name of the game was invented by Tucker[5], together with the following tale to explain it: two thieves are arrested and imprisoned in two separate cells on suspicion of a heinous crime, besides of a theft. The judge has sufficient evidence to prove that they are guilty of the minor crime, but would need a confession in order to sentence them to a severe punishment for the major crime. The judge, therefore, proposes to each of the two thieves, who are kept strictly isolated, a reduced sentence in exchange for testimony against the accomplice. The one who accepts will receive a light sentence, whereas the other one will receive the maximum sentence. If both confess, they will only receive a small reduction. If both deny the claim, they will be cleared of the major charge, and will be imprisoned only for the minor one.

The order of the possible situations is:

- if a player does not cooperate and the other one does (NC, C), that is if the former confesses and the second does not, the noncooperating one obtains the best possible result, that is the reduced sentence;
- if both players cooperate (C, C), that is they both deny the claim (the cooperation is between the two players, not with the judge, who is an external subject), both receive a light sentence, that is the second best possible result;
- if both players do not cooperate (NC, NC), that is both confess, they both receive a small reduction, that is the third best possible result for both;
- if a player cooperates and the other does not (C, NC), the one who cooperates, that is denies the claim, receives the maximum sentence, that is the worst possible result.

The cooperation strategy (C) is dominated by the noncooperation strategy (NC) for both players. This means that it is reasonable for both players to confess, and in fact the game has a single Nash equilibrium, that corresponds to mutual noncooperation (in the table, both payoffs are marked). Even though the solution is logically derived, it also raises a dilemma. In fact, noncooperation is the only nondominated

[3] Merrill Meeks Flood (1908-1991), American mathematician.

[4] Melvin Dresher (1911-1992), Polish-born American mathematician.

[5] Albert William Tucker (1905-1995), Canadian mathematician who also proposed the KKT-conditions.

strategy, therefore the only reasonable one. However, mutual cooperation would be more desirable for both players, since $f_{11} > f_{22}$).

The dilemma formalises a conflict between rationality and generlal interest, which looks typical of many practical situations. The prisoner's dilemma is probably the most investigated of all games.

14.8 Other games

We will say very little about general games (nonsymmetric and non zero-sum). We will only describe a rather curious asymmetric game which gives rise to paradoxical consequences. Such consequences are however supported not only by mathematical theory, but also by the empirical investigation of situations that respect the model assumption.

14.8.1 The pigsty game

This game, also known as *pigeon coop game* is not symmetric and not zero-sum. Table 14.10 reports its payoffs.

		Weak pig	
		Lever	Wait
Strong	Lever	$(4, 2)$	$(\bar{3}, \bar{3})$
pig	Wait	$(\bar{5}, 0)$	$(1, \bar{1})$

Table 14.10: Payoff matrix for the pigsty game: the markings on the best response for each row and column indicate that point (NC, NC) is the only Nash equillibrium

The game involves a stronger player (1) and a weaker one (2). A lever allows to obtain a prize (for instance, food). The prize, however, is accessible only from a position far away from the lever (the opposite side of the sty). Each of the two players can decide whether to operate the lever or to wait:

- if both operate on the lever, the stronger one reaches first the food, and gets most of it;
- if the stronger player operates the lever and the weaker one waits near the position where the food appears, both get a good portion (the weaker one gets it at the beginning, the stronger one in the end, shooing the weaker);
- if the weaker player operates the lever and the stronger one waits near the position where the food appears, the stronger players gets nearly all the food, using both the advantages of position and strength;
- if both players wait near the position where the food appears, the food does not appear and both have the worst result.

The stronger player (row player in Table 14.10) has no dominated strategies, whereas for the weaker player (column player) the strategy of operating the lever is dominated by that of waiting. Therefore, the weaker player waits. Then, it is reasonable for the stronger player to operate the lever. The paradoxical conclusion is that it is adviceable for the stronger player to service the weaker one, since it will anyway get the scraps, whereas the other player has no hope to get anything unless by waiting.

14.9 Finite games and mixed strategies

Nash extended the theory of Von Neumann and Morgenstern on mixed strategies from zero-sum games to any game with finite sets of players and strategies. Briefly, he obtained two remarkable results:

1. *every finite game admits at least an equilibrium in mixed strategies*;
2. it is not possible to use Linear Programming to find such equilibria: there are algorithms to find a single Nash equilibrium by solving a sequence of linear problems whose solution converges to an equilibrium, but the number of equilibria can be exponential with respect to the number of players and strategies, and the different equilibria can have different values (in zero-sum games all equilibria have the same value, and usually only one exists); at present, enumerating all Nash equilibria requires exponential-time implicit enumeration algorithms.

14.10 Exercises[*] [†] [‡]

Exercise 1

The following table reports the payoff matrix of a two-person zero-sum game, in which the row player A aims to maximise the benefit and the column player B to minimise the cost.

f	b_1	b_2
a_1	50	-20
a_2	-10	20
a_3	-50	10

Formulate and solve the problem of determining the best strategy (mixed, if necessary) for the row player A and the corresponding expected win.

Then, determine the best strategy (mixed, if necessary) for the column player B and the corresponding expected loss.

Assuming that player B follow a mixed strategy which selects at random b_1 with probability 0.75 and b_2 with probability 0.25 and that player A knows that, what is the minimum guaranteed win for player A?

Solution

Strategy a_3 is dominated by a_2, so that it will never be used. In order to build an optimal mixed strategy combining a_1 and a_2, it is necessary to determine the probability to select a_1 so as to maximise the expected result in the worst case:

$$\max\left[50\alpha - 10\left(1-\alpha\right), -20\alpha + 20\left(1-\alpha\right)\right] = \max\left[60\alpha - 10, -40\alpha + 20\right] \text{ con } \alpha \in [0;1]$$

The graphical resolution suggests to set $\alpha^* = 0.3$, which implies $f^* = 8$.

Correspondingly, the column player minimises the expected result in the worst case:

$$\min\left[50\beta - 20\left(1-\beta\right), -10\beta + 20\left(1-\beta\right)\right] = \min\left[70\beta - 20, -30\beta + 20\right] \text{ con } \beta \in [0;1]$$

which implies $\beta^* = 0.4$, confirming that $f^* = 8$.

If player B behaves mechanically, it is licit to consider him/her as a state of nature and to apply the expected value criterium. As the given probabilities imply that $E\left[f\left(a_1\right)\right] = 32.5$, $E\left[f\left(a_2\right)\right] = -2.5$ and $E\left[f\left(a_3\right)\right] = -35$, the best strategy is a_1.

Exercise 2

Given the finite two-person zero-sum game represented by the following payoff matrix (where the payoffs are the benefits of the row player), determine the best mixed strategy for the row player.

[*] The solutions of these exercises have not yet been revised: error reports are welcome.

[†] I owe several of these exercises to exam texts of professor Alberto Colorni.

[‡] Several exercises are much more complex than those required for the exam, or concern more general topics than those discussed in the notes. I have decided to keep them for the sake of completeness, marking them.

f	b_1	b_2	b_3	b_4
a_1	50	10	90	30
a_2	70	120	40	80

Then, determine the best mixed strategy for the column player[6].

Solution

The best strategy for the row player is $\xi^*_{a_1} = 0.6$ and $\xi^*_{a_2} = 0.4$, with $f^* = 60$. The best strategy for the column player is $\xi^*_{b_1} = \xi^*_{b_2} = 0$, and $\xi^*_{b_3} = \xi^*_{b_4} = 0.5$, also with $f^* = 60$.

Exercise 3

Consider the two-player zero-sum game defined by the following rules[7]:

- at the beginning, set $s = 0$;
- player A moves first, summing either 1 or 2 to s;
- player B moves second, summing either 1 or 2 to s;
- the players move iteratively, first A then B, until one of the two players sets $s \geq 5$: that player wins the game;
- if A wins, receives 100 (and B loses 100); if B wins, receives 300 (and A loses 300).

List the possible strategies for each of the two players.

Describe the game in extended form (as a tree) and in strategic form (as a matrix), computing the labels for each node of the tree and each cell of the matrix (strategy profile).

Based on the description in strategic form, state whether there exist dominated strategies and Nash equilibria.

Solution

Since at every round s increases at least by 1 and the game ends when $s \geq 5$, there are at most five moves (three rounds). The strategy of each player, therefore, is described by a pair of choices (all moves in the third round have the same impact). The payoff matrix is reported in the following table, which reports on the rows the strategies of player A, on the columns the strategies of player B and in the cells the gains of A (the losses of B are exactly opposite, given that it is a zero-sum game).

	$(1,1)$	$(1,2)$	$(2,1)$	$(2,2)$
$(1,1)$	100	-300	-300	-300
$(1,2)$	-300	-300	100	100
$(2,1)$	-300	-300	100	100
$(2,2)$	100	100	100	100

Strategy $(2,2)$ for player A dominates the other ones and guarantees to win 100. Strategy $(1,2)$ for player B dominates the other ones. There is therefore a Nash equilibrium in $((2,2),(1,2))$.

[6] This requires a reasoning on the duality between the two problems, that exceed the syllabus of the course.

[7] This exercise is more sophisticated than those proposed in the exam.

Exercise 4

Given the two-person finite game with the following payoff matrix, determine all dominated strategies and all Nash equilibria (if any exists).

f	C	D	E
A	(3,2)	(5,4)	(7,8)
B	(5,9)	(1,11)	(4,3)

Solution

Strategy C is dominated by strategy D. After removing it, strategy B becomes dominated on its turn. Then, strategy D becomes dominated and the only remaining strategy profile is (A, E), which is also necessarily a Nash equilibrium.

Exercise 5

Determine the possible dominated strategies and Nash equilibria in the finite two-person game with the following payoff matrix.

f	B_1	B_2	B_3
A_1	(0,1)	(-1,3)	(1,3)
A_2	(0,2)	(1,1)	(2,0)
A_3	(1,0)	(1,-1)	(3,0)

Solution

Strategy A_3 dominates the other two strategies for the row player. The column player at first has no dominated strategies, but after the removal of A_1 and A_2, strategy B_2 becomes dominated by the other ones. The two remaining strategy profiles, (A_3, B_1) and (A_3, B_3), are also Nash equilibria.

Exercise 6

Two opposed parties are standing for election: the Good and the Bad. Each party puts up one candidate per constituency. There are three constituencies: North, Centre and South. Both parties have a leader, who can run in at most two of the three constituencies. The following table provides an estimate, based on polls, of the millions of votes that the candidates (either the leader or another one) would get if they should run for each of the three constituencies.

	Good		Bad	
f	Leader	Other	Leader	Other
North	16	14	19	13
Centre	18	14	15	13
South	17	14	18	14

In each constituency, the candidate who gets more votes wins for his/her party 10 seats for each million votes more with respect to the adversary. The strategists of the two parties must decide where their leader should run, respecting the limit of at most two constituencies.

Describe the problem as a finite two-person zero-sum game, providing the payoff matrix for the Good.

State whether there exist dominated strategies and equilibrium points.

Solution

First of all, it is clear that it is profitable to put up the leader in two constituencies, because the leader always gets more votes than the alternative candidate. Therefore, both players have exactly three strategies, each one characterised by the constituency in which the leader does not run. This yields the following payoff matrix.

f	North	Centre	South
North	30	-10	10
Centre	10	-30	-10
South	20	-20	0

The strategy profile (Nord,Centro), with the leader of the Goods running at the Centre and South and the leader of the Bads running at the North and the South, is an equilibrium, with a gain equal to 10 for the Bads. The other rows and columns are all dominated.

Exercise 7

The OX game is played by tho players as follows: player A selects Even (E) or Odd (O), and announces this choice; then, player B chooses an integer number from 2 to 5, and announces the chosen number. If A selects E, he/she pays to B as many units as the number chosen by B; if A selects O, he/she gains the ssame amount from B. The game must be played in "sets" of four rounds, during which each player must select the available alternatives in a balanced way: A must select twice E and twice O; B must choose once each integer value from 2 to 5.

Describe the game tree for a single round.

List the strategies of both players for a single round.

Show an example of path on the tree of the whole "set" of four rounds, associating to the leaf the corresponding final impact.

Indicate the number of levels and leaves for a whole "set" of four rounds, specifying how many leaves are feasible[8].

Estimate whether the game offers to both players the same chances of victory on a set of four rounds[9].

Solution

The game tree has the following structure.

-	E	5	$(-5,5)$
		4	$(-4,4)$
		3	$(-3,3)$
		2	$(-2,2)$
	O	5	$(5,-5)$
		4	$(4,-4)$
		3	$(3,-3)$
		2	$(2,-2)$

[8] This question is more complex.

[9] This question is more complex.

In the single round, the first player has two strategies (E and O), the second one has four strategies for each move of the first one, that is 8. An example of strategy for the second player is to choose 5 against E and 2 against O.

An example of path along the whole game tree is $(P, 5, P, 4, D, 3, D, 2)$ which yields the payoffs $(-4, 4)$.

On a whole set, the tree has 8 levels: those associated to the first player have two outgoing arcs per node, those associated to the second player have a number of outgoing arcs per node progressively decreasing from 4 to 1. Therefore, the leaves are $2 \cdot 4 \cdot 2 \cdot 3 \cdot 2 \cdot 2 \cdot 2 \cdot 1 = 2^4 \cdot 4! = 384$. However, the balancing rule reduces the choices of the first player from 2^4 to $\binom{4}{2}$, so that the feasible leaves are only 144.

The game is unfair, because B can always win the single round, and consequently the whole set, by choosing the maximum remaining integer value when A has selected E and the minimum one when A has selected O.

Exercise 8[10]

Determine the Nash equilibria (if any exist) for the following continuous two-person zero-sum game:

$$\begin{cases} f_1(x_1, x_2) = \dfrac{1}{3}x_1^3 + x_1 x_2^2 - 2x_1 - \dfrac{2}{3}x_2^3 \\ f_2(x_1, x_2) = -f_1(x_1, x_2) \end{cases} \qquad \text{con } x_1 \in X^{(1)} = \mathbb{R} \text{ e } x_2 \in X^{(2)} = \mathbb{R}$$

Solution

An equilibrium point, if any exists, is a maximum point for the row player and a minimum point for the column player. In a continuous game, maxima and minima can be determined with the usual first-order conditions (in an unconstrained, one-dimension problem, these conditions reduce to setting the first derivative to zero):

$$\begin{cases} \dfrac{\partial f_1}{\partial x_1} = x_1^2 + x_2^2 - 2 = 0 \\ \dfrac{\partial f_1}{\partial x_2} = 2x_1 x_2 - 2x_2^2 = 0 \end{cases}$$

plus checking the sign of the second-order derivatives:

$$\begin{cases} \dfrac{\partial^2 f_1}{\partial x_1^2} = 2x_1 < 0 \\ \dfrac{\partial^2 f_1}{\partial x_2^2} = 2x_1 - 4x_2 > 0 \end{cases}$$

The solutions with $x_2 = 0$ are unfeasible ($2x_1 < 0$ e $2x_1 > 0$). Those with $x_2 \neq 0$ have $x_2 = x_1$, and therefore $x_1 = x_2 = \pm 1$. Since $x_1 - 2x_2 > 0$, the only equilibrium point is $(-1, -1)$.

Exercise 9

Consider the continuous two-person zero-sum game with benefit function $f(a, b) = -2a^2 + b^2 + 3ab - a - 2b$ for player A who decides the value of variable a. Determine whether $H = (1/4, 3/4)$ is a Nash equilibrium. If it is not, determine the equilibrium points, if any exists.

[10] This exercise is about continuous games. The notes have only hinted at such games. The basic property of Nash equilibria, of being maximum points for the row player and minimum points for the column player, is trivially extended.

Solution

Point H is not a Nash equilibrium, because it is not a maximum point for the row player and a minimum point for the column player. Point $K = (4/17, 11/17)$ is a Nash equilibrium.

Exercise 10[11]

A nation has three political parties, more or less of the same strength: the Left (L), the Centre (C) and the Right (R). Each party can decide between two possible political platforms: a conservative one (Cons.) and a reformist one (Ref.). The platforms are published simultaneously. The pair or triplet of parties which propose the same platform win the elections and form a majority for the government. Each of the three parties receives a benefit in terms of parliament seats if it joins the majority, whereas it pays a cost if is remains isolated at the opposition. The benefit for the parties with the winning platform is +30 for the reformist platform, +10 for the conservative one. The loss for the party at the opposition (if any) is −50 if its platform is reformist, −10 if it is conservative.

Model this problem as a game in strategic form. Since there are three players, instead of two, it is not possible to report the payoffs in the classical matrix, but it is possible to list on the rows the strategy profiles and on the columns the corresponding payoffs.

Determine whether there are equilibrium points, and list them.

Determine the strategy suggested by the worst-case criterium.

Solution

The payoff matrix is the following one.

x_S	x_C	x_D	$f_S(x_S, x_C, x_D)$	$f_C(x_S, x_C, x_D)$	$f_D(x_S, x_C, x_D)$
Cons.	Cons.	Cons.	10	10	10
Cons.	Cons.	Ref.	10	10	-50
Cons.	Ref.	Cons.	10	-50	10
Cons.	Ref.	Ref.	-10	30	30
Ref.	Cons.	Cons.	-50	10	10
Ref.	Cons.	Ref.	30	-10	30
Ref.	Ref.	Cons.	30	30	-10
Ref.	Ref.	Ref.	30	30	30

The equilibrium points are the homogeneous triplets of strategies (Cons.,Cons.,Cons.) and (Ref.,Ref.,Ref.). In fact, in these situations every player avoids changing strategy because it would end up at the opposition, and it would be penalised. On the other hand, in the mixed situations, the player at the opposition is encouraged to join the majority, but also the ones in a conservative majority are encouraged to change platform and join forces with the opposition.

The worst-case criterium suggests to adopt the conservative platform, because the worst case consists in ending up at the opposition, and in that case the conservative platform implies a smaller loss.

[11] This exercise considers a three-person game. Except for this, the concepts of equilibrium point and worst-case strategy are the same.

Exercise 11[12]

Waiting for the sheriff, the deputy sheriffs A and B play a simplified version of poker with a deck of three cards: J, Q and K. At the beginning of the game, each player puts one dollar in the pot. Then, they pick one card each, leaving the third card on the table. Each player only known his/her own card. A chooses whether to "raise", that is add another dollar to the pot, or "fold", that is to leave the game. B chooses whether to fold, to raise (adding a new dollar to the pot, only if A has not raised), or to "call" (only if A has raised). Finally, A can fold or call. In summary, the game can unfold according to the following scheme:

1. A folds:
   (a) B folds: the pot is divided between the two players;
   (b) B raises, that is adds a third dollar to the pot:
      i. A folds: B wins the pot, getting back his/her money, plus one dollar from A;
      ii. A calls: the player with the higher card wins the pot, getting back his/her money, plus one dollar from the other player;
2. A raises, that is adds a third dollar to the pot:
   (a) B folds: A wins the pot, getting back his/her money, plus one dollar from B;
   (b) B calls: the player with the higher card wins the pot, getting back his/her money, plus one dollar from the other player.

List the possible strategies of the two players.

Determine the evaluation matrix of the first player for any possible card deal.

Determine matrix of the expected values of the payoffs.

List the possible strategies for the first player, that is the possible moves depending on the card picked up (for example, the triplet $[P-P, P-V, A]$ means fold if the picked card is J or Q, raise if it is K).

Solution

The possible strategies for the first player are $P-P$, $P-V$ and A, those for the second player are P, A and V.

According to the cards picked up by the two players, the payoff matrix for the first one is the following (label '-' marks the unfeasible strategy profiles):

- if the card is J and for any card of the second player, or if the card is Q e and the card of the second player is K:

f	P	A	V
$P-P$	0	-1	-
$P-V$	0	-2	-
A	1	-	-2

[12] This exercise also covers advanced material, not required by the exam, as it includes conditions of uncertainty. The extension to this kind of games is based on concepts discussed in the corresponding chapters.

- if the card is Q and that of the second player is J:

f	P	A	V
$P-P$	0	-1	-
$P-V$	0	2	-
A	1	-	2

- if the card is K and for any card of the second player:

f	P	A	V
$P-P$	0	-1	-
$P-V$	0	2	-
A	1	-	0

The matrix of the expected values of the payoffs is:

f	P	A	V
$P-P$	0	-1	-
$P-V$	0	0	-
A	1	-	0

The first player has $3^3 = 27$ possible strategies, given that there are 3 possible moves for each of the 3 possible initial cards.

Exercise 12[13]

The "truth game" employs a box containing two tokens (a red one, R, and a black one, N). Player A extracts from the box a token without showing it to the adversary, and announces its colour, possibly lying. If he/she states that the token in black, B gains 1 and A loses 1. If he/she states that it is red, player B can either accept the announcement or reject it. If he/she accepts it, A wins 1 and B loses 1. Finally, if he/she rejects it, the token colour is checked: if it is red, A wins 2 and B loses 2; if it is black, B wins 2 and A loses 2. The probabilities to extract each token are balanced (50% each).

Model the game in strategic form, through a payoff matrix.

Compute graphically the best strategies for both players.

Discuss how the game and the best strategies change if the box contains a red token and two black ones.

Solution

The main complication of the exercise is in distinguishing what falls under the scenario and what falls under the strategy. The colour of the extracted token is the scenario. The strategies for player A are:

1. announce the true colour of the token;

[13] Also this exercise concerns a game in conditions of uncertainty. If someone feels that the exercise or the solution is unclear, it could be comforting to read the following anecdote: this exercise, in a more ambiguous form, was part of a mid-term exam I had to pass when I was a student. I misunderstood it completely, losing 6 points out of 30. The feeling of having been defrauded was so strong that I decided to reject the grade; in the following exam, my grade was 30 *cum laude.*

2. announce that the token is red, whatever is its colour.

It would be more natural to distinguish between making a true or a false announcement, but lying is forbidden in the case of a red ball, whereas a correct description of the game requires the strategies to be feasible in every scenario.

The strategies for player B are:

1. accept the announcement;

2. reject the announcement.

If the token is red, the payoff matrix is:

f	Accept	Reject
True	1	2
Red	1	2

If the token is black, the payoff matrix is:

f	Accept	Reject
True	-1	-1
Red	1	-2

The matrix of the expected values of the payoffs is therefore:

f	Accept	Reject
True	0	1/2
Red	1	0

The best strategy for player A is $(2/3, 1/3)$, the best strategy for player B is $(1/3, 2/3)$. The expected result is a gain of 1/3 for A.

The modified case produces the following table:

f	Accept	Reject
True	-1/3	0
Red	0	-2/3

with optimal strategies $(5/6, 1/6)$ for A and $(1/3, 2/3)$ for B. The expected result is a gain of $-1/9$ for A.

Chapter 15

Group decision-making

As in game theory, also in group decision-making the decision-maker set D is finite, but not reduced to a single element ($|D| > 1$). The fundamental difference is that the single decision-makers do not fix independently the values of the decision variables. The must agree on the alternative to choose coordinating somehow. Another typical situation is that in which a main decision-maker indicates the alternative, but in such a way to satisfy as much as possible the preferences of the other ones, besides his/her own.

A classical approach to the modelling of these problems is the construction, starting from the preference relations Π_d of the decision-makers $d \in D$, of a single preference relation $\Pi_D \subseteq F \times F$, denoted as *group preference*, which formulates the criteria based on which the common decision should be taken. Once the group preference relation has been constructed, the problem becomes equivalent to that of a single decision-maker, and all the techniques described in the previous chapters can be applied to it. The fundamental passage of the resolution process is therefore how to build a preference relation starting from $|D|$ given relations. The investigations on this problem mainly refer to political and social sciences, so that it can be expected that the notation, once again, changes. In fact, the decision-maker are often denoted as *individuals*, *citizens*, *agents*, *voters* or *judges*.

In the following, we will assume that:

- there is a single possible scenario ($|\Omega| = 1$), a condition introduced for the sake of simplicity, but also because in group decision-making each individual often has a rather precise opinion (be that right or wrong) about the impact expected from each alternative;

- the feasible region X is finite;

- the impact function $f(x)$ is invertible, that is each solution has a different impact, so that the preference relations can be thought as relations on X, instead of on F (it is enough to define $x \preceq x'$ if and only if $f(x) \preceq f(x')$);

- the preference relations of the single decision-makers are weak order relations, that is reflexive, transitive and complete.

The last condition is the most important one. As we known, it means that each decision-maker is required to respect rather strict conditions of rationality.

Definition 60 *Given a set of solutions X, we denote as $\mathcal{D}(X)$ the set of all weak orders on the elements of X.*

Since every preference relation on solutions is a set of pairs of solutions ($\Pi_d \subseteq X \times X$, or $\Pi_d \in 2^{X \times X}$), set $\mathcal{D}(X)$ is the collection of all sets that enjoy the reflexive, transitive and complete property, and therefore $\mathcal{D}(X) \subset 2^{2^{X \times X}}$.

Example 87 *Consider a set of three solutions, $X = \{a, b, c\}$. The set of all weak orders on X includes 13 different relations:*

- *the 6 total orders, that is the permutations of the three solutions:*

$$a \prec b \prec c \qquad a \prec c \prec b \qquad b \prec c \prec a$$

$$b \prec a \prec c \qquad c \prec a \prec b \qquad c \prec b \prec a$$

- *the 3 weak orders in which two solutions are reciprocally indifferent and preferable with respect to the third one:*

$$a \sim b \prec c \qquad a \sim c \prec b \qquad b \sim c \prec a$$

- *the 3 weak orders in which one solution is preferable with respect to the other two, which are reciprocally indifferent:*

$$a \prec b \sim c \qquad b \prec a \sim c \qquad c \prec a \sim b$$

- *the weak order in which the three solutions are indifferent with respect to each other:*

$$a \sim b \sim c$$

Each of these weak orders can be expressed (in a rather lengthy way) listing the pairs of solutions such that the former is weakly preferable with respect to the latter. For example, order $a \prec b \prec c$ can be formulated as $\{(a,a),(a,b),(a,c),(b,b),(b,c),(c,c)\}$, whereas order $a \sim b \sim c$ can be formulated as $\{(a,a),(a,b),(a,c),(b,a),(b,b),(b,c),(c,a),(c,b),(c,c)\}$.

15.1 Social welfare function

The problem we face is how to derive a group preference relation from a finite set of weak orders, one for each individual.

Definition 61 *Given a set of solutions X, we denote as* social welfare function *a function that associates to each $|D|$-uple of weak orders on X a "group" weak order on X:*

$$g : \mathcal{D}(X)^{|D|} \rightarrow \mathcal{D}(X)$$

It is possible to build a huge number of completely arbitrary functions that reduce several weak orders to a single one. The interesting aspect is to search for a function that respect as much as possible the individual weak orders. In order to do that, such a function must enjoy suitable properties. In the following, we will survey some proposals that have been put forth in time, and we investigate their limitations. Then, we will list the properties that appear to be necessary to characterise an acceptable social welfare function, and we will prove the fundamental result according to which there exists no function which satisfies all the necessary properties. Finally, we will discuss the criticisms that have been made to that result and some subsequent developments of the research on the topic.

15.2 Condorcet method

The *Condorcet method*[1], also known as *simple majority method* is based on the following definition:

$$x \preceq_D x' \Leftrightarrow |\{d \in D : x \preceq_d x'\}| \geq |\{d \in D : x' \preceq_d x\}|$$

that is, a solution is preferable to another one according to the group when the number of individuals who prefer the former to the latter exceeds the number of individuals who prefer the latter to the former.

Notice that the individuals according to whom the two solutions are indifferent appear on both sides of the inequality and therefore cancel each other: only the ones having a strict preference matter. Consequently, two solutions are indifferent when the same number of individuals prefer each of them to the other one (in particular when all individuals consider them as indifferent).

The Condorcet method provides a simple algorithm to compute a social welfare function. However, this function has a strong drawback, which is known in the literature as *Condorcet paradox*. The drawback is represented in Table 15.1: there are three individuals $D = \{1, 2, 3\}$ and three solutions $X = \{a, b, c\}$. Each column is associated to an individual and describes his/her preference relation on the solutions: row 1 reports the best solution, row 2 the second best one and row 3 the worst one. The three individual preference relations are therefore total orders, without any tie.

	Individuals		
Order	1	2	3
1	a	b	c
2	b	c	a
3	c	a	b

Table 15.1: Example of Condorcet paradox: based on Condorcet method, each of the three solutions is better than another one; the resulting preference relation is not transitive, and therefore is not a weak order.

Applying Condorcet method to the table, one obtains that:

- $a \prec_D b$ because a is preferable to b for two individuals (1 and 3), whereas the opposite holds only for one (2);
- $b \prec_D c$ because b is preferable to c for two individuals (1 and 2), whereas the opposite holds only for one (3);
- $c \prec_D a$ because c is preferable to a for two individuals (2 and 3), whereas the opposite holds only for one (1).

Moreover, each solution is trivially preferable to itself because this holds for all individuals. The resulting relation exhibits a circular strict preference:

$$a \prec b \prec c \prec a$$

and is nontransitive, since $a \prec b$ and $b \prec$, but $a \not\prec c$. Therefore, *the Condorcet method does not always allow to build a weak order.* This does not mean that it

[1] Marie Jean Antoine Nicolas de Caritat, Marquis of Condorcet (1743-1794), French mathematician, economist, philosopher and politician activist of the Gironde party during the French revolution, died as a suicide in prison during the Terror.

never works; it means that it can fail. And, in particular, it can be unable to provide a solution preferable to all other ones, that is the minimum requirement for a choice criterium.

History proposes several examples of parliamentary votes in which three versions of a law had the support of voter groups structured as in the Condorcet paradox. In some cases, the *impasse* was solved setting an arbitrary order of business, that is discussing first the conflict between solutions a and b (which led to the victory of solution b), and then the conflict between the winning solution b and the third solution c (which led to the victory of solution c). A different order of business would have led to a different choice. The experts in politics know that setting the agenda often grants a strong position in parliamentary proceedings.

15.3 Borda method

The *Borda method*[2] is based on the auxiliary definition of *Borda count*, that we have introduced in Section 4.3:

$$B_d(x) = |\{x' \in X : x \preceq_d x'\}|$$

that is, in a finite problem the value of a solution x for individual d can be measured by the number of solutions to which the individual prefers x. Once a value function is defined for each individual, aggregating them with a simple sum appear like a natural development, from an egalitarian and democratic point of view:

$$B_D(x) = \sum_{d \in D} B_d(x)$$

and it appears equally natural to base the group preference on the aggregated function, that is on the overall Borda count:

$$x \preceq_D x' \Leftrightarrow B_D(x) \geq B_D(x')$$

that is, a solution is preferable to another one according to the group when the overall Borda count on all individuals of the former exceeds that of the latter.

This definition introduces a consistent value function for the group and bases on it the preference. The resulting relation is necessarily a weak order. Therefore, contrary to the Condorcet method, the Borda method always allows to build a social welfare function.

Also the Borda method, however, has a drawback, that we have already met in another context, that is the dependence on irrelevent alternatives. In other words, the weak order that is created between two solutions does not depend only on the preferences of the individuals between the two solutions, but in general also by the preferences they have with respect to other solutions. This is bad, because in general the feasible region built during the modelling phase is at least partly arbitrary (this has been discussed in Section 2.1.2, outlining the iterative process to build the model, and in particular the feasible region; for instance, the mitigation measures for large public works are typically defined in subsequent phases of the study, after identifying and discussing the direct impacts of the project). If the final result of the decision depends on the alternatives taken into account, there is a margin for manipulating the process which is not desirable.

Example 88 *Consider the decision problem represented in Table 15.2, characterized by seven individuals ($D = \{1,2,3,4,5,6,7\}$) and four solutions ($X = \{a,b,c,d\}$.*

[2] Jean-Charles de Borda (1733-1799), French mathematician, physicist and admiral operating during the French revolution.

	Individuals						
Order	1	2	3	4	5	6	7
1	a	b	c	a	b	c	a
2	b	c	d	b	c	d	b
3	c	d	a	c	d	a	c
4	d	a	b	d	a	b	d

Table 15.2: A group decision problem with 7 individuals and 4 solutions; every column is associated to an individual and reports from row 1 to row 4 the solutions in (total) preference order.

Once again, the preference relations of all individuals are total orders: the preferred solution is associated to a Borda count equal to $B_d(x) = 4$, the second best solution to a value equal to 3, the third one to 2 and the worst one to 1. The overall count is reported in Table 15.3, from which one derives the group total order $c \prec b \prec a \prec d$.

X	$B_D(x)$
a	$4+1+2+4+1+2+4 = 18$
b	$3+4+1+3+4+1+3 = 19$
c	$2+3+4+2+3+4+2 = 20$
d	$1+2+3+1+2+3+1 = 13$

Table 15.3: Overall Borda count for the alternatives of the problem reported in Table 15.2.

If however solution d were removed, that is a rather uninteresting solution, given that no individual prefers it and overall its result is clearly inferior to that of the other ones, one would obtain the problem and the overall count reported, respectively in Tables 15.4 and 15.5.

	Individuals						
Order	1	2	3	4	5	6	7
1	a	b	c	a	b	c	a
2	b	c	a	b	c	a	b
3	c	a	b	c	a	b	c

Table 15.4: The problem of Table 15.2 reduced by removing solution d.

X	$B_D(x)$
a	$3+1+2+3+1+2+3 = 15$
b	$2+3+1+2+3+1+2 = 14$
c	$1+2+3+1+2+3+1 = 13$

Table 15.5: Overall Borda count for the alternatives of the problem reported in Table 15.4.

The overall count yields the total order $a \prec b \prec c$, which obviously does not include the removed alternative, but also reverts completely the order relation among the other three alternatives. The reason for this change is that the Borda count of each alternative, and therefore the preference relation between two alternatives, *depends on the position of all the other ones. In particular, in this case alternative d was always worse than c, often worse than b, but only sometimes worse than a; removing it reduces the count of d strongly, that of b significantly and that of a only weakly, reverting the preference relations among these three alternatives.*

The previous examples leads to conclude that also the Borda method can be unsatisfactory, given that is it open to manipulations in the phase of the generation of the alternatives.

15.4 Plurality system

The plurality system is the method traditionally used in elections: each individual summarises his/her own preference in the choice of a solution. The solution with the larger number of votes wins. Strictly speaking, given that we assume the individual preference relations to be weak orders, we summarise the preference of an individual in the choice of a subset of solutions (reciprocally indifferent) that are strictly preferable to all the other ones. In formal terms, we define:

$$V(x) = |\{d \in D : x \preceq_d y \text{ for all } y \in X\}|$$

and

$$x \preceq_D x' \Leftrightarrow V(x) \geq V(x')$$

This system is based on a consistent value function, as the Borda method, and therefore gives rise to a weak order. However, it also has the drawback to depend on irrelevant alternatives, even if slightly less than the Borda method, since the value function does not consider all the positions of every solution, but only the winning positions. In order to change the number of winning positions, it is necessary to choose with some care the alternatives to introduce or remove.

Example 89 *Consider the group decision problem with* 7 *individuals and* 4 *solutions represented in Table 15.6.*

	Individuals						
Order	1	2	3	4	5	6	7
1	a	a	b	b	d	d	d
2	b	b	a	a	a	b	c
3	c	c	c	c	b	c	a
4	d	d	d	d	c	a	b

Table 15.6: A decision problem with 7 individuals and 4 solutions.

Once again, the individual preference relations are all total orders. The plurality system leads to choose solution d, because it is the best one for three individuals, whereas a and b are the best for two individuals and c for none. Therefore, $d \prec a \sim b \prec c$.

However, if solution b should be rejected, reducing the problem to the one represented in Table 15.7, the method would suggest solution a, as it is preferred by four individuals, whereas solution d is preferred by three and solution c by none. Therefore, $a \prec d \prec c$.

Actually, the method has another, rather strong, drawback: a solution that is strongly supported by a minority prevails upon a solution that is not preferred by

	Individuals						
Order	1	2	3	4	5	6	7
1	a	a	a	a	d	d	d
2	c	c	c	c	a	c	c
3	d	d	d	d	c	a	a

Table 15.7: The problem of Table 15.2 reduced removing solution b.

anyone, but that is agreeable for everybody. This derives from the fact that function $V(x)$ only considers the first position in the individual weak orders, completely neglecting the subsequent positions.

Example 90 *Considering the problem with* 4 *solutions and* 7 *individuals represented in Table 15.8, one can notice that the winning solution d is very disagreeable for the individuals not supporting it, whereas solution a, which nobody supports as the best one, is the second best for everybody, and would therefore be probably a good compromise. Any method ignoring the solutions in positions following the first one cannot detect the occurrence of such situations.*

	Individuals						
Order	1	2	3	4	5	6	7
1	b	b	c	c	d	d	d
2	a	a	a	a	a	a	a
3	c	c	b	b	b	c	c
4	d	d	d	d	c	b	b

Table 15.8: Another group decision problem with 7 individuals and 4 solutions.

15.5 Lexicographic method

The lexicographic method loses a fundamental feature of the previous two methods, that is the symmetry with respect to the individuals. In fact, the method consists in selecting a permutation of the individuals, $p = \left(d_1, d_2, \ldots, d_{|D|}\right)$ and in defining the preference between two solutions as the preference of the first individual according to the permutation who is not indifferent between them. Two alternatives are therefore indifferenti if and only if all individuals consider them as indifferent.

$$x \prec_D x' \Leftrightarrow \begin{cases} x \prec_{d_1} x' \text{ or} \\ x \sim_{d_1} x' \text{ and } x \prec_{d_2} x' \text{ or} \\ \ldots \\ x \sim_{d_1} x', x \sim_{d_2} x', \ldots \text{ and } x \prec_{d_{|D|}} x' \end{cases}$$

$$x \sim_D x' \Leftrightarrow \quad x \sim_{d_1} x', x \sim_{d_2} x', \ldots \text{ and } x \sim_{d_{|D|}} x'$$

This method describes an extreme version of an absolute monarchy, in which the will of the monarch d_1 is the law (at least for the choices in set X: for other choices, the order could change) and, if the monarch is indifferent, the choice is delegated to another individual, who in case of indifference delegates it to a third one, and so on down to the last citizen of the kingdom, who can express a preference if and only if only all the other ones are indifferent.

15.6 The axiomatic approach

Now, after surveying a number of rather reasonable and natural method, all of them afflicted by some disadvantage, one could wonder whether there exists a method to aggregate individual preferences in a group preference while respecting some desirable basic properties. And one can assume that listing such properties be the first step towards the design of such a method. This is the axiomatic approach proposed for group decision-making by Arrow[3] in 1951, with an unexpected outcome, that is the proof that it is impossible to formulate a method enjoying all the properties considered as necessary to aggregate preferences in a democratically acceptable way. In order to describe Arrow's result in a compact way, it is useful to give a name to each possible set of preferences that a social welfare function is expected to aggregate.

Definition 62 *Given a finite set of alternatives X and a finite set D of individuals, we denote as* profile $\mathcal{P}(X)$ *any $|D|$-uple of weak orders on the alternatives associated to the individuals.*

The problem consists therefore in the identification of a social welfare function g that aggregate every possible profile $\mathcal{P}(X) \in \mathcal{D}(X)^{|D|}$ into a group preference relation $\Pi_D(X)$ that is a weak order:

$$\Pi_D(X) = g(\mathcal{P}(X)) \in D(X) \text{ for all } \mathcal{P} \in \mathcal{D}(X)^{|D|}$$

Given two profiles and two alternatives, it is interesting to remark whether one of them is more unbalanced than the other in favour of one of the two alternatives. A possible way to visualise the situation is to think of two different polls: an earlier one registers the initial preference profile of the inividuals with respect to the two alternatives; a later one, after an event convinces some supporters of the second alternative to change their minds, becoming indifferent or even supporting the first, registers the new profile, which is more favourable to the first alternative.

Definition 63 *Given two due alternatives $x, y \in X$ and two profiles $\mathcal{P}(X)$ e $\mathcal{P}'(X)$, we denote as $\prec_d$ and $\prec'_d$ the strict preferences of individual $d \in D$ in profiles $\mathcal{P}(X)$ and $\mathcal{P}'(X)$. Similarly, we denote as $\sim_d$ and $\sim'_d$ the indifference relation and with $\preceq_d$ and $\preceq'_d$ the weak preference in the two profiles. We say that* $\mathcal{P}'(X)$ promotes x over y more than $\mathcal{P}(X)$ *when*

- *every individual who strictly prefers x in $\mathcal{P}(X)$ also strictly prefers x in $\mathcal{P}'(X)$:*
$$x \prec_d y \Rightarrow x \prec'_d y$$
- *every individual who is indifferent between x and y in $\mathcal{P}(X)$ weakly prefers x in $\mathcal{P}'(X)$:*
$$x \sim_d y \Rightarrow x \preceq'_d y$$

Finally, we define the concept of dictatorship, in order to exclude it. A dictator is an individual whose strict preferences imply strict preferences for the whole group. In the lexicographics method, for example, individual d_1 is a dictator. The other methods considered, on the contrary, do not admit dictators, because there always exists a profile in which at least one of the strict preferences of each individual is violated by the group.

[3] Kenneth Joseph Arrow (1921 - 2017), American economist, won the Nobel Prize for Economy in 1972.

Definition 64 *Given a social welfare function g, a* dictator *according to such a function is an individual $d \in D$ such that for every profile in which $x \prec_d y$, the social welfare function imposes that $x \prec_D y$.*

There is also a limited concept of dictatorship: instead of a single individual whose strict preferences determine those of the group, it is possible to have a subset of individuals whose strict preference on a specific pair of alternatives, if unanimous, determines that of the group.

Definition 65 *Given a social welfare function g, a subset of individuals $V \subseteq D$ is* decisive *for a pair of alternatives (x, y) when, for every profile $\mathcal{P}(X)$ such that $x \prec_d y$ for all $d \in V$, the social welfare function imposes $x \prec_D y$.*

A dictator is of course a special case of a decisive set, consisting of a single individual, which is decisive for all pairs of alternatives.

Arrow's approach first introduces some axioms which define the properties required from a social welfare function $g : \mathcal{D}(X)^{|D|} \to \mathcal{D}(X)$ in order to be considered democratic. Then, it proves that, if all axioms but the last one are true, necessarily the last one is false. The original proof by Arrow considers seven axiom.

Theorem 26 *(Arrow's theorem) Let X be a finite set of alternatives, D a finite set of individuals and $\Pi(X) : D \to \mathcal{D}(X)$ a preference relation that associates to each individual a weak order on the set of the alternatives. The following seven statements are mutually incompatible:*

1. nontriviality *axiom: there are at least three alternatives and at least two individuals:*
$$|X| \geq 3 \qquad |D| \geq 2$$
2. universality *axiom: there exists a social welfare function g defined on all possible profiles, that is, g describes how to aggregate any set of weak orders associated to the individuals;*
3. weak order *axiom: the social welfare function g returns a weak order (that is a reflexive, transitive and complete preference relation) on X;*
4. independence from irrelevant alternatives *axiom: given a problem with a reduced set of alternatives $X' \subset X$, but the same individuals and preference relations restricted to X', the resulting weak order is the one obtained restricting $\Pi_D(X)$ to X';*
5. monotony *axiom: if a profile $\mathcal{P}'(X)$ promotes x over y more than $\mathcal{P}(X)$, if for all pairs of solutions which do not include x the two profiles are identical and if $x \prec_D y$ (that is, the group prefers x in profile $\mathcal{P}(X)$), then $x \prec'_D y$ (that is, the group prefers x also in $\mathcal{P}'(X)$);*
6. popular sovereignty*: every group preference can be obtained by suitable fixing the preferences of the individuals (the social welfare function is* surjective*);*
7. non-dictatorship *axiom: the social welfare function g does not admit any dictator.*

Proof. The proof goes through the following phases:

1. prove the existence of a decisive subset for a pair of solutions;

2. prove that such a subset can be reduced to a single individual;
3. prove that such an individual is decisive for every pair of solutions, and is therefore a dictator.

First, notice that the nontriviality axiom guarantees the existence of at least three alternatives. Then, we remark that the monotony and popular sovereignty axioms imply the following property of *unanimity* (or Pareto efficiency): if $x \prec_d y$ for all $d \in D$, then $x \prec_D y$, that is, the social welfare function imposes that unanimous strict preferences turn into group strict preferences. In fact, if there exists a profile for which the group prefers x to y and all the other profiles which promote x over y more than it also imply that the group prefers x to y, then the profile in which all individuals are unanimous in preferring x will force the group to prefer x. The opposite is not necessarily true.

Thanks to the unanimity axiom, according to every social welfare function each pair of alternatives (x, y) admits a decisive subset. In fact, if all individuals of D, unanimous, strictly prefer x to y, the whole group strictly prefers x to y. This implies that D is a decisive subset for any pair (x, y).

The following point is the most delicate of the proof: the search for a minimal decisive subset, that is a subset that cannot lose any member without losing the status of being a decisive subset. Let us assume that $V \subseteq D$ is a minimal decisive subset. This means that it is decisive for one or more pairs of solutions, and that removing any of its members it is no longer decisive for any pair of solutions. How many elements has V?

Let d be an individual of V and $W = V \setminus \{d\}$. Finally, let $D \setminus V$ be the set of individuals not belonging to V; this set could be empty. Let (x, y) be one of the pairs for which V is decisive. The property of being decisive holds for any preference profile: for any opinion of the individuals, if the members of V unanimously prefer x to y, then the group prefers x to y, even if all the other individuals have exactly the opposite preference. Then, let us consider the situation described in the following table, where each column reports the order of the preferences of a subset of individuals with respect to three alternatives, denoted as x, y and z (the nontriviality axiom is applied once again here). They are total orders, and therefore weak orders. Therefore, the social welfare function g can be applied to this preference profile, yielding a group preference (thanks to the universality axiom). What is the preference of the group?

$\{d\}$	$V \setminus \{d\}$	$D \setminus V$
x	z	y
y	x	z
z	y	x

The individuals of V are unanimous in strictly preferring x to y. Since V is a decisive subset for (x, y), this implies that the group has the same preference:

$$x \prec_V y \Rightarrow x \prec_D y$$

Now there are two possible cases: either V includes more than an individual, or it includes exactly one. By contradiction, assume the former case: $V \setminus \{d\}$ is not empty. Moreover, it is not decisive for any pair of solutions (otherwise, it would contradict the initial assumption that V is a minimal decisive set). In particular, it is not decisive for (z, y). Now, let us observe the preferences of the individuals for pair (z, y): all members of $V \setminus \{d\}$ are unanimous in strictly preferring z to y, all other individuals are unanimous in the opposite preference. What is the group preference between z and y? First of all, a preference exists because the weak order axiom guarantees that the group preference relation is complete. On the other hand, the preference cannot be strictly in favour of z, because it would agree with that of the members of $V \setminus \{d\}$, and this would still hold for any other profile promoting z over y; consequently, $V \setminus \{d\}$ would be decisive. Rejecting the strict preference implies that $y \preceq_D z$. However, since $x \prec_D y$ and the group preference is transitive (once again the weak order axiom):

$$x \prec_D y \text{ and } y \preceq_D z \Rightarrow x \prec_D z$$

But the only individual who strictly prefers x to z is d, which means that d is decisive for pair (x, z). This remarks contradict the assumption that V is a minimal decisive subset. The conclusion is that the only possible case is the latter, that is, that V includes a single individual. Consequently, there exists an individual d who is specifically a "dictator" for the pair of alternatives (x, y).

The second step of the proof shows that d is a dictator for all pairs of alternatives. Let us start from the pairs (x, w) with $w \in X \setminus \{x, y\}$. The following table shows a possible preference profile (notice that w could be the same as z: we use a different name to stress that it is *any* solution different from x and y).

$\{d\}$	$D \setminus \{d\}$
x	y
y	w
w	x

The table shows that the individuals are unanimous in strictly preferring y to w, and therefore $y \prec_D w$. On the other hand, since d is a specific dictator for pair (x, y), his/her strict preference for x turns into the preference of the whole group, even if nobody else shares it. Now, transitivity strikes again, implying that:

$$x \prec_D y \text{ and } y \prec_D w \Rightarrow x \prec_D w$$

By monotony any other profile that promotes x over w more than the considered one would yield the same result. But the only individual who strictly prefers x to w is d, who is therefore a dictator also for the new pair. Since w is absolutely general, d is a dictator for all pairs whose first member is alternative x.

Now, we prove that d is also a dictator for the pairs (z, w) with $z, w \in X \setminus \{x\}$. In order to do that, consider the following preference profile:

$\{d\}$	$D \setminus \{d\}$
z	w
x	z
w	x

The individuals are unanimous in strictly preferring z to x, and therefore $z \prec_D x$. On the other hand, since d is a specific dictator for all pairs (x, w), his/her strict preference for x turns into the group preference, even if nobody else shares it. Now, transitivity strikes again, implying that:

$$z \prec_D x \text{ and } x \prec_D w \Rightarrow z \prec_D w$$

not only in the considered profile, but also (by monotony) in those where other individuals strictly prefer z to w. Hence, d is a dictator for pair (z, w), which is completely general, except for excluding alternative x from both members.

Finally, we can prove that d is also a dictator for the pairs (z, x) with $z \in X \setminus \{x\}$. In order to do that, consider the following preference profile:

$\{d\}$	$D \setminus \{d\}$
z	w
w	x
x	z

The individuals are unanimous in strictly preferring w to x, and therefore $w \prec_D x$. On the other hand, since d is a specific dictator for all pairs (z, w), his/her strict preference for z turns into the group preference, even if nobody else shares it. Now, transitivity strikes again, implying that:

$$z \prec_D w \text{ and } w \prec_D x \Rightarrow z \prec_D x$$

but the only individual who strictly prefers z to x is d, who therefore (extending the conclusion by monotony to the profiles in which other individuals prefer z to x) is a dictator for pair (z, x), where z is completely general, except for excluding alternative x. Of course, the pairs (x, x) are indifferent for all individuals and for the group, due to the reflexivity of the preferences.

The conclusion is that d is a dictator, which violates the last statement of the theorem. ■

More recently, a proof of Arrow's theorem has been proposed, which is based on six axioms, weaker than the original ones and easier to state. The modern formulation includes:

- instead of the indipendence from the irrelevant alternatives, a *binary relevance* axiom: for any pair of alternatives (x, y), the group preference depends only on the preferences of the single individuals between x and y.

  It can be proved that this axiom is equivalent to the independence from irrelevant alternatives, that is, that adding or removing alternatives does not modify the preference between x and y; therefore, the substitution does not change anything;

- instead of the monotony and popular sovereignty axioms, the unanimity axiom, that is weaker.

Even if the modern formulation of the theorem considers a wider set of social welfare functions, it also fails finding a satisfactory one.

15.7 Criticisms to Arrow's axioms

Arrow's proof raised a lively debate for many years. In particular, the reviewers tried to undermine the proof suggesting that the axioms are not as natural and obvious as they look. The criticisms are partially justified in stating that Arrow's model does not correspond exactly to the real world, but cannot anyway destroy the fundamental core of the theorem, that is the fact that no preference aggregation rule is able to satisfy the requirements that it would be desirable to enjoy. Let us survey in further detail some criticisms.

15.7.1 Criticisms to the nontriviality axiom

The nontriviality axiom is rather difficult to criticise: in practice, a group always includes more than one individual and the alternatives are often more than two. However, the supporters of two-party political systems have often stressed the fact that such systems exactly correspond to the case left open by Arrow's proof. In fact, according to *Black's theorem*[4], if the feasible region includes only two alternatives x and y, all of the other axioms proposed by Arrow can be satisfied by a social welfare function that counts the individuals who strictly prefer each of the alternatives to the other one and assigns the group preference to the alternative with the largest number of supporters. The theorem is also known as the *median voter theorem* because it also proves that, if the strength of the preference of each individual for the two alternative is known, and the individuals are distributed along a continuous line according to such preference strength, the winning alternative is the one preferred by the median individual, that is by the individual who has half of the individuals on the left and half on the right.

This has led the supporters of two-party systems to consider it as the only implementable form of democracy, and to support the voting systems that favour the aggregation on two alternative fronts. This position has some strong points, but it can be objected that the reduction to two alternatives in problems in which the feasible solutions are more numerous requires on its turn decision processes for which theory does not offer any guarantee of reaching democratic results.

15.7.2 Criticisms to the universality axiom

More theoretical criticisms have been made to the universality axiom, which requires a social welfare function to be able to aggregate the preferences for any possible opinion of the individuals. Black's theorem come once again into play. In fact, it proves that, if the individual preferences respect some additional restrictive assumptions, plurality voting respects all the other axioms introduced by Arrow. This means that, if the individuals distribute their preferences in a suitable way instead of having any imaginable profile, it is possible to guarantee a democratic aggregation rule.

The special restriction assumed by Black requires that the alternative could be ordered in a sequence $X = (x_1, \dots, x_n)$ according to any arbitrary criterium (it is not required to follow any rule) and that (this is the fundamental assumption) the preference relation of each individual should be consistent with a value function that is unimodal with respect to such a sequence. In simple words, each individual must have a preferred alternative (or a set of consecutive preferred alternatives) and must consider the alternatives that follow the preferred one as progressively worse as one proceeds forward along the sequence; as well, the alternatives that precede the preferred one should be considered as progressively worse as one proceeds backward along the sequence. Figure 15.7.2 shows the result: the single individuals correspond to the unimodal lines (first increasing, then decreasing), that describe the strength of their preference for the alteratives (distributed along the horizontal axis). Arrow's axioms are all respected, except for the universality axiom, by the Condorcet method, which leads to the victory of the alternative preferred by the median individual, that is the individual whose maximum has half of the maxima of the other individuals on the left and half on the right.

This situation clearly recalls the periods in which the political life of a country is polarised along a linear sequence of positions between two extreme wings, classically denoted as "left" and "right". If the individuals have a favourite political position and consider the other ones progressively worse as they get farther away in one or in the other direction (even in an asymmetric way, provided that it is in a monotonous way), Condorcet's method allows to aggregate the preferences satisfying Arrow's axioms. If, on the contrary, some individuals do not respect this attitude (for instance, they prefer extreme positions, be them right-wing or left-wing, to moderate

[4] Duncan Black (1908-1991), Scottish economist.

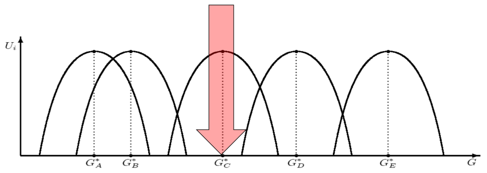


Figure 15.1: Five individuals express their preferences, which are all unimodal, for the alternatives distributed from left to right along the horizontal axis; Condorcet's method leads to select the alternative pointed by the arrow, which is the favourite one for the median individual, that is the individual who has two individuals on the left and two on the right.

ones), the process fails and Arrow's theorem is once again valid.

15.7.3 Criticisms to the axiom of binary relevance

Other criticisms have been raised against the binary relevance axiom, according to which the group preference between two alternatives must depend on the individual preferences only between those two alternatives, and not the other ones. It has already been remarked that this is equivalent to the independence from irrelevant alternatives, and it aims to avoid that the group preference could be manipulated removing or introducing alternatives. However, this assumption looks unnecessarily strong. Let us explain why with an example.

Example 91 *Two individuals meet at a coffee shop and decide to buy two identical drinks. In order to decide which, they express their preference by sorting the available drinks. The result is reported in next table.*

d_1	d_2
Coffee	*Tea*
Water	*Coffee*
Lemonade	*Water*
Cola	*Lemonade*
Tea	*Cola*

As in the other examples, for the sake of simplicity we have assumed that the preferences of the individuals form total orders. From the table, it is reasonable to deduce that coffee would be a good candidate for a common choice, given that it is the favourite drink for the first individual and the second favourite drink for the second individual. However, the binary relevance axiom requires to consider the group preference between the two drinks as perfectly identical to the one obtained removing the other drinks, as reported in the following table.

d_1	d_2
Coffee	*Tea*
Tea	*Coffee*

This situation, however, looks quite different: the two drinks contrast in an exactly balanced way, as a matter of fact forbidding a choice (each choice appears arbitrarily favourable to one of the two individuals).

The fundamental point is that the presence of the other drinks allows not only to describe the order of the individual preferences, but also to somehow measure *their strength. This can be a deceptive effect: possibly, the first individual considers all drinks approximately equivalent, even if ordered in the given way, whereas the second individual likes only tea, detesting all the other ones. In fact, in order to take a decision, one should know the strength of the preference of each individual with respect to each alternative. This closely recalls the composition of the values of multiple attributes in a single utility function, where the individuals correspond to the attributes. Remembering the complications required by this process, it can be easily understood how little it can be applied to the aggregation of the preferences of many individuals (potentially, hundreds of millions, for political choices in very populated nations) in a single national preference. Therefore, the criticism is valid, but the countermeasure is hard to apply in practice.*

15.7.4 Criticisms to the unanimity axiom

Even the unanimity axiom is open to criticism. It would seem obvious that, if all the individuals share the same preference, the group should replicate it faithfully. Yet, this contradicts some basic principles supported by several currents of political thought. When the catholic mention "unnegotiable values" and the liberals "fundamental human rights" they mean exactly that some preferences (in favour of life versus choice, or in favour of freedom of speech versus state censorship) cannot be modified even if all individuals believe that they should.

On the other hand, it is also clear that, if the foundation of these principles is not the preference of the individuals, the problem of establishing its ultimate source remains open. The answers that have been given to this problem is not taken into account in Arrow's model.

15.8 Impossibility theorem for voting systems[*]

There is a theorem that closely resembles Arrow's theorem, but concerns the voting systems, and specifically the allotment of seats to parties in a voting competition. It is well known that there are many different voting systems, from the purely proportional ones to the majoritarian ones. A voting system can be interpreted as an aggregation and simplification rule for the individual preferences on the existing parties. There preferences are simplified *a priori*, assuming that each individual chooses a single party. Starting from this simplified representation, we aim to obtain an aggregated description of such preferences assigning to each party a suitable number of seats in a parliament.

Definition 66 *Let D be a finite set of individuals and X a finite set of $n = |X|$ parties. The set of the* vote vectors $V = \left\{v \in \mathbb{N}^n : \sum_{x \in X} v_x = |D|\right\}$ *includes all vectors of n integer numbers summing to $|D|$.*

The vote vectors count the number of individuals who prefer each party.

Definition 67 *Given a set S of seats to assign in a parliament, we denote as* voting system *any function $s : V \to \mathbb{N}^n$ that turns any vote vector $v \in V$ in the number of seats assigned to each party, so that $\sum_{x \in X} s_x(v) = |S|$.*

This definition includes all possible voting systems: those with a single party which, for any vote vector, assign all seats to party x_1; those which assign all seats to the party getting the relative majority of the votes; those which assign each to

[*] This section provides advanced concepts, that are not part of the course's syllabus.

each party a number of seats proportional to the number of votes, with some more or less complex rounding mechanism (which is far from being a trivial task).

Given this definition, it makes sense to investigate which systems enjoy the properties that a theoretically ideal system should guarantee to ensure a good operation of a nation. Unfortunately, also this problem exhibits an impossibility result similar to that of Arrow.

Theorem 27 *If the set of parties has at least three elements, no voting system enjoys all of the following properties:*

1. *remuneration: the parties with no votes should get no seat*

$$s_x(v) = 0 \text{ per ogni } v \in V \text{ tale che } v_x = 0$$

2. *monotony: the parties with more votes should get at least the same number of seats (as the votes are many more than the seats, the seat distribution is necessarily more coarse-grained than the vote distribution)*

$$s_x(v) \geq s_y(v) \text{ for all } v \in V \text{ such that } v_x > v_y$$

3. *absolute majority: a party with the absolute majority of the votes should get the absolute majority of the seats, and vice versa*

$$v_x > \frac{|D|}{2} \Leftrightarrow s_x(v) > \frac{|S|}{2} \text{ per ogni } v \in V$$

4. *superadditivity: the merger of two parties should have at least as many seats as the two individual parties*

$$v = \begin{bmatrix} v_1 \\ \dots \\ v_x \\ \dots \\ v_y \\ \dots \\ v_n \end{bmatrix}, \quad v' = \begin{bmatrix} v_1 \\ \dots \\ v_x + v_y \\ \dots \\ 0 \\ \dots \\ v_n \end{bmatrix} \Rightarrow s_x(v) + s_y(v) \leq s_x(v')$$

The fourth property deserves some explanation: vector v' aggregates the votes of the two parties x and y in a single one (indicato come x per semplice comodità), mentre y rimane come semplice segnaposto nel vettore, senza voti. La proprietà di superadditività richiede che il nuovo partito abbia un numero di seggi non inferiore alla somma dei due partiti originari. Il partito fittizio non ha seggi, grazie alla proprietà di remunerazione. Questa proprietà ha lo scopo di evitare che un partito si frammenti in due per sfruttare l'incentivo offerto dal sistema elettorale, che consente di ottenere complessivamente più seggi scindendosi che aggregandosi.

Part VI

Descriptive models[*]

[*] This part provides advanced concepts, that are not part of the course syllabus.

Chapter 11

Modelli per i sistemi di trasporto

Ci occupiamo dei sistemi per il trasporto di persone, sia pubblici sia privati. Consideriamo *modelli descrittivi*, nei quali cioè si vuole *determinare il valore di alcune grandezze incognite a partire da quello di altre grandezze note*. Non consideriamo invece modelli decisionali, cioè non imponiamo noi il valore di alcuna grandezza in gioco. Il modello descrittivo verrà poi usato in supporto a modelli decisionali (aprire o chiudere strade, cambiare sensi unici, modificare l'ampiezza, istituire corsie riservate, modificare i cicli semaforici, introdurre rotonde, ecc...). Il modello che presentiamo è il più classico modello per i sistemi di trasporto. Viene detto *modello a quattro stadi* in quanto si compone di una catena di quattro sottomodelli, ognuno dei quali riceve come dati i risultati del precedente e fornisce come dati al successivo i propri risultati (Fig. 11.1). Questo ha due conseguenze immediate:

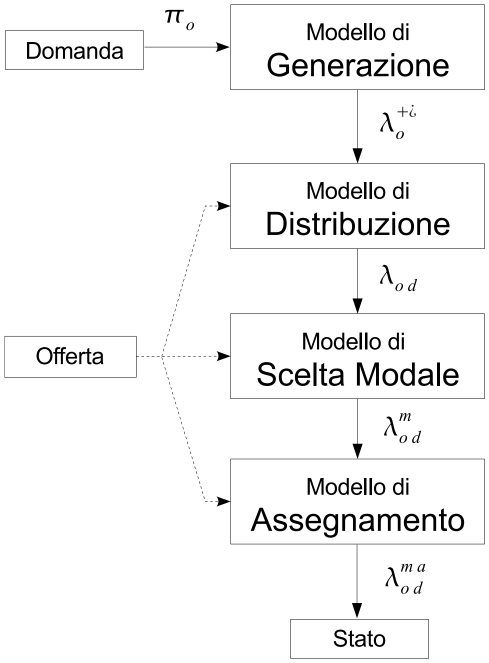


Figure 11.1: Sistemi di trasporto, modello a quattro stadi

1. gli errori di modellazione compiuti in uno stadio si ripercuotono sui successivi, accumulandosi
2. ciascun sottomodello può essere realizzato in vari modi, dando origine a parecchi diversi modelli complessivi

In particolare, alcuni dati possono essere ricavati da misurazioni anziché dai modelli precedenti. Questo è possibile se il modello di trasporto viene usato per descrivere situazioni identiche o simili a quella attuale, per cui i valori misurati sono significativi, oltre ad essere esenti da errori di modellazione. D'altra parte, se il modello va usato per descrivere situazioni molto diverse dall'attuale, oppure se i dati misurati sono incompleti, se sono campionari e poco significativi o se sono affetti da errori di misurazione, l'uso di modelli offre una possibilità di sostituirli o correggerli.

Ogni sottomodello del modello a quattro stadi ha parametri che devono essere determinati attraverso una fase di *calibrazione*, nella quale si individuano i valori che avvicinano il più possibile i valori previsti dal modello a quelli registrati. La fase di calibrazione comporta l'uso di modelli decisionali (si tratta di decidere il valore dei parametri) in supporto a un modello descrittivo, che poi verrà usato in supporto a modelli decisionali, come si è detto. Tali modelli possono essere modelli di stima ai minimi quadrati, modelli di massima verosimiglianza, ecc...

Il modello a quattro stadi parte dalla descrizione di

- *domanda di trasporto*: popolazione e sue caratteristiche socio-economiche; distribuzione spaziale e temporale, abitudini, servizi e attrattori di traffico presenti sul territorio
- *offerta di trasporto*: rete stradale (distanze, caratteristiche di ampiezza e forma, cicli semaforici, caratteristiche degli incroci, ecc...), reti di trasporto pubblico (linee, fermate, distanze, orari, frequenze).

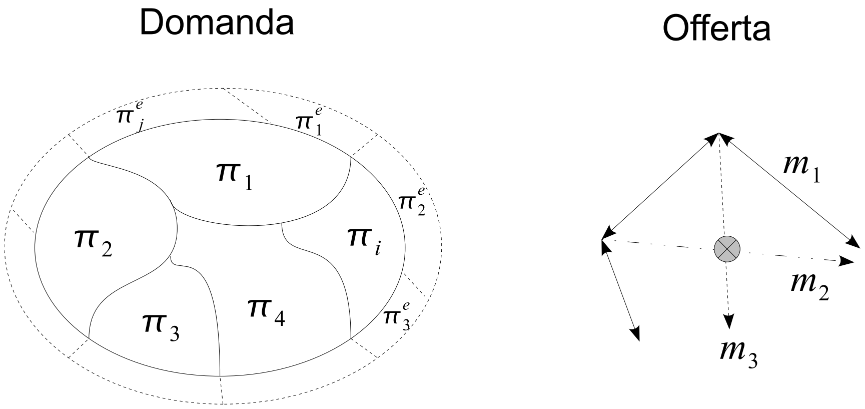


Figure 11.2: Sistemi di trasporto, modellazione di domanda ed offerta

Questi due elementi fondamentali sono spesso schematizzati come riportato in Fig. 11.2: per la domanda si utilizza un insieme di aree di origine π_i con una serie di aree periferiche π_j^e che esercitano un'influenza marginale sulla popolazione; per l'offerta si utilizza un grafo a più livelli dove ognuno rappresenta un differente mezzo di trasporto. In questa rappresentazione devono essere tenuti in considerazione sia i punti di giunzione fra diversi livelli (ad es. se è possibile scendere dal treno per prendere l'autobus), sia i vincoli derivanti dalla circolazione (semafori, sensi unici, ecc...).

In generale, l'offerta di trasporto deve rappresentare l'intera struttura viaria a disposizione nell'area considerata. É inoltre importante che la modellazione avvenga con un adeguato livello di dettaglio: questa rappresentazione verrà utilizzata in più livelli del modello.

Gli stadi componenti il modello sono i seguenti:

1. *generazione* o emissione della domanda (*chi fa il viaggio*?): definita una segmentazione della popolazione in fasce orarie, ed eventualmente in motivazioni di viaggio, categorie di viaggiatori, ecc... e una suddivisione spaziale della popolazione in aree dette *origini* (abbastanza grandi da essere gestibili, abbastanza piccole da rendere trascurabili gli spostamenti nella singola area), si procede a valutare il numero di spostamenti generati da ogni area
$$\lambda_o^+ = \phi_g\left(\pi_o, \ldots\right)$$
2. *distribuzione* (*da dove a dove si viaggia*?): per ogni origine, si valuta la distribuzione dei viaggi fra le aree di arrivo (*destinazioni*). La segmentazione della popolazione serve a tener conto del fatto che per fasce orarie, motivazioni e categorie diverse di popolazione la distribuzione sarà anche diversa.
$$\lambda_{od} = \phi_d\left(\lambda_o^+, \ldots\right)$$
3. *scelta modale* (*con che mezzo si viaggia*?): per ogni coppia origine-destinazione (O/D), si determina la distribuzione dei viaggiatori tra i diversi mezzi disponibili (privati e pubblici)
$$\lambda_{od}^m = \phi_m\left(\lambda_{od}, \ldots\right)$$
4. *assegnamento* (*lungo quale percorso si viaggia*?): per ogni coppia O/D e per ogni mezzo, si determina il percorso effettivamente seguito sulla rete di trasporto
$$\lambda_{od}^{mp} = \phi_p\left(\lambda_{od}^m, \ldots\right)$$

L'idea di fondo è esprimere il numero di spostamenti dalla zona o alla zona d con il mezzo m e lungo il percorso p (nella fascia oraria, per la motivazione e la categoria di persone scelte) come prodotto della popolazione presente nella zona o per la probabilità che ciascun individuo si sposti in d col mezzo m lungo il percorso p. Tale probabilità verrà espressa come prodotto di probabilità condizionate

$$\lambda_{od}^{mp} = \pi_o P\left(N = n\right) P\left(D = d|O = o\right) P\left(M = m|O = o, D = d\right) P\left(P = p|O = o, D = d, M = m\right)$$

Tramite questa catena di probabilità condizionate si tenta di modellare in modo trattabile il processo decisionale del singolo individuo. E' da notare come questa descrizione sia approssimata in quanto ipotizza che la catena sia orientata: la scelta della destinazione influenza la scelta del mezzo di trasporto, ma non viceversa. La stessa considerazione vale per tutte le altre scelte: ognuna influenza quelle a valle ma non è vero il contrario. Questa approssimazione rende quindi ogni livello di scelta indipendente dai successivi. L'ipotesi di fondo non descrive perfettamente la realtà ma, fra i possibili ordinamenti, la teoria ha scelto quello che riflette meglio le reali dinamiche di trasporto.

11.1 Modello di generazione

Si tratta di un modello molto semplice. Fissata la fascia oraria, ed eventualmente la motivazione e la categoria di viaggiatori di interesse, il modello stima il numero di spostamenti generati da una certa zona come prodotto della popolazione per la probabilità che un individuo della zona (e della categoria specificata) compia uno spostamento all'interno della fascia oraria data. Più precisamente, ammettendo che un individuo possa compiere più spostamenti, si impiegano le probabilità che l'individuo compia un dato numero n di spostamenti, moltiplicato per il numero stesso:

$$\lambda_o^+ = \pi_o \sum_{n=1}^{n_{\max}} nP\,(N = n)$$

Si noti che la popolazione non va intesa come popolazione residente, ma come popolazione presente nell'area durante la fascia oraria di interesse (magari a causa di spostamenti precedenti). Quindi, le zone che generano traffico nelle fasce orarie mattutine, diventano zone di destinazione di traffico generato dalle zone lavorative nelle fasce orarie serali, quelle in cui la popolazione rientra da una giornata di lavoro. Nel corso della giornata le zone possono quindi assumere sia il ruolo di origini di traffico, sia di destinazioni.

Si possono stimare direttamente le probabilità $P\,(N = n)$, oppure esprimerle come funzioni (in genere lineari) di variabili $x_g^{(l)}$ che descrivano le caratteristiche socio-economiche della zona di origine, dalla categoria di viaggiatori, dalle fascia oraria e dalla motivazione di viaggio.

$$P\,(N = n) = \sum_l \beta_g^{(l)} x^{(l)}$$

dove gli *attributi* $x^{(l)}$ sono grandezze socio-economiche come note come il tipo di zona, il reddito, la diffusione dell'automobile, ecc... mentre i coefficienti $\beta_g^{(l)}$ dovranno essere calibrati rispetto ad un campione di dati significativo.

11.2 Modello di distribuzione

A questo punto, si tratta di stimare quanta parte dei viaggi generati si diriga verso ciascuna area di destinazione. L'idea alla base del modello di distribuzione è che ogni individuo sceglierà la propria destinazione come la più conveniente per soddisfare il bisogno all'origine del viaggio. Se i criteri che costituiscono l'utilità associata a un viaggio fossero perfettamente uniformi per tutti i viaggiatori di una certa zona di origine, essi si dirigerebbero alla zona che meglio li soddisfa. Se tali criteri fossero perfettamente noti, il loro comportamento sarebbe perfettamente prevedibile. Ad esempio, se l'unico criterio fosse la distanza (da minimizzare), tutti i λ_o^+ viaggi generati dalla zona o sarebbero diretti alla più vicina zona in grado di soddisfare la motivazione del viaggio.

Poiché così non è, senza abbandonare del tutto l'idea di un comportamento razionale e ottimizzante, si adottano *modelli di utilità aleatoria*, nei quali l'utilità del singolo individuo viene descritta come parzialmente nota e parzialmente aleatoria. Cercheremo quindi la probabilità che ogni individuo scelga una certa destinazione e valuteremo il flusso diretto a quella destinazione come prodotto del numero di individui in viaggio per tale probabilità.

$$\lambda_{od} = \lambda_o^+ P\,(D = d | O = o)$$

Le ipotesi di fondo sono:

1. il generico utente ha un insieme finito A di alternative disponibili; l'insieme è lo stesso per tutti (da qui l'esigenza di segmentare categorie di utenti diversi, per esempio dotati o no di patente di guida);

2. ogni decisore associa a ciascuna alternativa $a \in A$ un'utilità percepita U_a e sceglie l'alternativa che massimizza U_a

3. l'utilità U_a dipende da una serie di caratteristiche misurabili (*attributi*) $x_{(i)a}$ propri dell'alternativa e del decisore; quindi, il decisore sceglie un'alternativa confrontando gli attributi di quell'alternativa con quelli delle altre;

4. l'utilità U_a non è nota con certezza all'osservatore esterno, e pertanto deve essere rappresentata con una variabile aleatoria, pari alla somma di una componente nota (costituita da quegli attributi che l'osservatore può misurare) e di una componente aleatoria o *residuo* (costituita da quegli attributi che l'osservatore non può misurare):

$$U_a = V_a + \epsilon_a \; a \in A$$

5. la dipendenza della componente deterministica V_a dagli attributi viene in genere modellata come lineare, con coefficienti uniformi per tutti gli individui. Gli errori compiuti con tale ipotesi vengono anch'essi attribuiti alla componente aleatoria.

$$V_a = \sum_i \beta_d^{(i)} x_a^{(i)} \; a \in A$$

dove $x_a^{(i)}$ sono gli attributi dell'alternativa a e i coefficienti $\beta_d^{(i)}$ esprimono il peso dell'attributo i sull'utilità dell'individuo.

A questo punto, la probabilità di scegliere l'alternativa $\bar{a}$ è

$$P\left(A = \bar{a}\right) = P\left(U_{\bar{a}} \geq U_a, \forall a \in A\right) = P\left(\epsilon_{\bar{a}} \geq \epsilon_a + V_a - V_{\bar{a}}, \forall a \in A\right)$$

Per valutare questa probabilità, occorre un'ipotesi sulla distribuzione stocastica dei residui. Il modello più diffuso in letteratura è il *modello Logit multinomiale.* Esso assume che *i residui seguano identicamente e indipendentemente una distribuzione di Gumbel a media nulla e varianza θ^2.*

$$F\left(x\right) = e^{-e^{-\frac{x}{\theta} - \Phi}}$$

dove Φ è la costante di Eulero.

Questa distribuzione è abbastanza simile alla distribuzione normale (quella tipica dei valori casuali), ma ha in più la proprietà di poter esprimere in forma chiusa la probabilità precedente.

$$P\left(A = \bar{a}\right) = \frac{e^{U_{\bar{a}}}}{\sum\limits_{a \in A} e^{U_a}}$$

che ha un andamento sigmoidale piuttosto verosimile (Fig. 11.3). Qualora tutte le alternative avessero la stessa utilità deterministica ($U_a = \bar{U}, \forall a \in A$), il modello suggerisce che gli individui si disperdano uniformemente fra le diverse destinazioni ($P(A = a) = 1/ \mid A \mid$).

Altre proprietà interessanti sono il fatto che le probabilità dipendono dalle differenze tra i valori delle utilità sistematiche, e non dal loro valore assoluto e quindi risultano indipendenti alla scelta dello zero per ciascun attributo. Inoltre, il modello risulta indipendente dalle alternative irrilevanti, cioè aggiungere o togliere alternative non altera l'ordinamento relativo fra quelle presenti: cambia il denominatore ma allo stesso modo per tutti i coefficienti, come se fossero moltiplicati per un identico

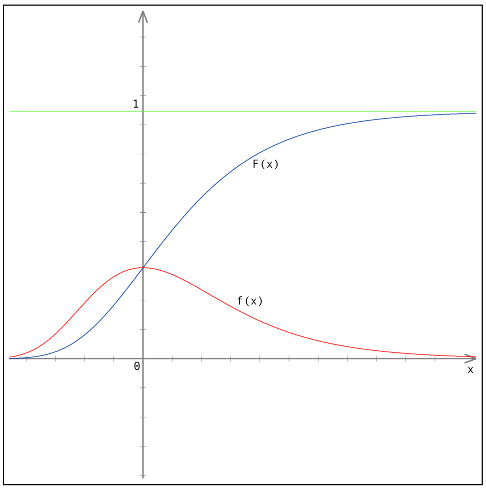


Figure 11.3: Distribuzione di Gumbel $F(x)$ e relativa funzione di densità di probabilità $f(x)$.

fattore. Infine, esiste l'influenza della varianza dei residui: al crescere di θ, le probabilità tendono ad avvicinarsi, cioè meno sono significativi i termini noti, meno si differenziano fra loro le scelte.

Il modello richiede una fase di calibrazione per valutare il parametro θ e i coefficienti $\beta_d^{(i)}$ dei diversi attributi.

Il modello Logit è ampiamente utilizzato anche nell'ambito del marketing, per descrivere il comportamento del consumatore posto di fronte ad un insieme di possibili prodotti o servizi da acquistare.

11.3 Modello di scelta modale

Stima quanta parte dei viaggi dall'origine o alla destinazione d utilizza il mezzo m. I modelli Logit multinomiali sono stati applicati per la prima volta a questa fase, estendendosi poi a quella di distribuzione, nella quale in precedenza si adottavano modelli diversi (ad esempio, il modello gravitazionale o quello a massima entropia).

Gli attributi tipici usati nel modello Logit per la scelta modale sono:

- tempo di viaggio
- costo monetario
- comodità
- tempo di avvicinamento al mezzo e di allontanamento dal mezzo
- tempo di attesa (nullo per il mezzo privato)

Trattandosi di criteri eterogenei, i coefficienti $\beta^{(i)}$ devono correlare fra loro le unità di misura, traducendo tutti gli attributi in termini di utilità adimensionale. Nel caso di costi (cioè escludendo solo la comodità), i coefficienti sono negativi.

Ricavare il peso $\beta_m^{(i)}$ assegnato ad ogni attributo $x_a^{(i)}$ dell'alternativa a richiede uno studio dedicato che può essere svolto secondo differenti modalità. Nel caso in cui si contatti un campione dell'utenza per rilevarne direttamente utilità e bisogni si parla di metodi delle *preferenze dichiarate* o *stated preferences.* Sotto questa dicitura viene raccolto un insieme di metodi di pesatura (o *pricing*) dove il bacino d'utenza viene interpellato direttamente tramite opportuni questionari. Le domande sono strutturate in modo tale da ricavare, per ognuna delle alternative presenti, un valore monetario: si può chiedere all'intervistato quanto sarebbe disposto a pagare per evitare un peggioramento del servizio o, in alternativa, quanto sconto esigerebbe sul prezzo del servizio per sentirsi in un certo modo "risarcito" del calo di qualità subito. Un'ulteriore possibilità prevede di chiedere all'intervistato di determinare un tradeoff accettabile fra alternative differenti, dal quale poi è possibile stimare il valore assegnato ad ognuna. La seconda categoria di metodi, definita delle *preferenze rilevate* (*revealed preferences*), richiede l'osservazione del comportamento dell'utenza a fronte di modifiche della qualità o del prezzo delle alternative. I metodi appartenenti a questa categoria sono considerati i più affidabili dalle discipline *econometriche* in quanto si osservano i comportamenti reali dell'utenza, evitando le potenziali imprecisioni dovute alla incongruenze di valutazione che possono sorgere nel decisore messo di fronte ad un questionario. Nonostante questo difetto, i metodi delle preferenze dichiarate sono, per ovvi motivi, decisamente meno costosi ed impegnativi e permettono inoltre di assegnare un valore di utilità anche ad alternative che non sono ancora state implementate.

11.4 Modello di scelta del percorso

Questo modello parte dalle matrici O/D associate a ciascun mezzo e arriva a determinare il numero di individui che seguono ogni percorso. Nel caso dei mezzi pubblici, il modello è generalmente banale, dato che il percorso disponibile è uno solo, oppure si riduce ad un'ulteriore applicazione del modello Logit, oppure a un'applicazione di modelli statistici basati sulla frequenza delle corse (l'ipotesi è che l'individuo prenda il mezzo che passa per primo fra quelli che possono servire la sua richiesta).

Nel caso del trasporto privato invece, a partire dalle matrici O/D il modello assegna il valore del flusso di veicoli su ogni arco della rete stradale. In questo caso, la scelta del percorso è tutt'altro che banale perché molti percorsi sono disponibili e soprattutto perché il tempo di viaggio su ogni percorso dipende pesantemente dal numero di veicoli che lo seguono. I due aspetti sono quindi profondamente correlati, influendo l'uno sull'altro: un percorso più breve attira più traffico e diventa quindi automaticamente più lungo.

Quando un insieme di veicoli percorre una strada, il tempo che impiega è inversamente proporzionale alla sua velocità. La velocità di un veicolo è una variabile aleatoria, che il veicolo tipicamente tende a conservare costante. Se sulla strada ci sono però altri veicoli, la cosa comporta sorpassi, e questo è possibile solo se l'intervallo fra il veicolo precedente e quello prima ancora lo consente. Quindi, il tempo di viaggio medio tende a crescere al crescere del flusso. Sulle autostrade, si considera approssimativamente lineare la relazione fra tempo e flusso.

$$t_a = \alpha_a + \beta_a \frac{x_a}{K_a} \tag{11.1}$$

dove α_a, β_a e K_a sono parametri dell'arco di strada a. Ovviamente esiste un limite superiore al flusso sopportabile dal tratto prima dell'imbottigliamento dei viaggiatori come mostrato in Fig. 11.4. Per comprendere il significato del parametro K_a, si consideri la dinamica del flusso di veicoli in viaggio su di un tratto autostradale. Se un conducente che sta percorrendo la corsia di destra incontra un veicolo che viaggia ad una velocità inferiore, cercherà di spostarsi sulla corsia a sinistra per superarlo.

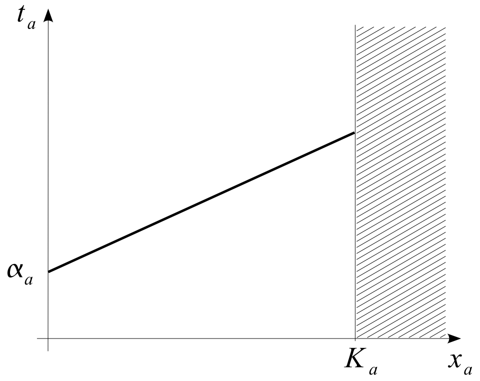


Figure 11.4: Tempo di percorrenza su arco autostradale

Allo stesso modo, se un conducente che sta percorrendo la corsia di sorpasso incontra un veicolo più lento, tenterà di indurlo a spostarsi sulla corsia di destra per mantenere la propria velocità. La capacità di spostarsi da una corsia all'altra è quindi una proprietà fondamentale per mantenere una velocità di crociera. Questa possibilità però non è sempre garantita. All'aumentare del traffico infatti, la facoltà di cambiare corsia diminuisce, poiché diventano più rari gli intervalli di spazio fra due veicoli entro i quali è possibile inserirsi con sicurezza, con conseguente diminuzione della velocità media del flusso veicolare. Quando il traffico è molto denso, i veicoli che si spostano sulla corsia di sorpasso si inseriscono in intervalli abbastanza stretti, costringendo chi li precede a decelerare, spesso bruscamente. Queste decelerate si propagano anche ai veicoli successivi, dando luogo a forti oscillazioni di velocità. In queste condizioni il flusso diventa instabile: si verificano rapide ed improvvise cadute di velocità, con possibile arresto della circolazione. Si definisce *capacità fisica* di una carreggiata autostradale quel valore del flusso veicolare al di là del quale il rischio della instabilità è considerato inaccettabile. Imponendo quindi che

$$x_a \leq K_a$$

la funzione lineare (11.1) diventa verosimile. Come si nota dalla Figura 11.4, l'andamento oltre K_a non è rappresentato in quanto non sarebbe significativo: oltre quel valore il flusso veicolare è in condizione di instabilità. Il valore di *capacità fisica* dipende dalla geometria della carreggiata, dalla composizione del traffico, dalle condizioni atmosferiche e dal comportamento dei conducenti.

Nelle strade urbane invece, anche quando non ci sono carreggiate separate, il sorpasso rimane difficoltoso o vietato. Questo vincolo imposto sulla circolazione urbana rende le strade di questo tipo simili ad una coda *fifo* in quanto tutti i veicoli che intraprendono il percorso devono necessariamente assestarsi alla velocità del più lento. Il tempo di percorrenza è descritto in modo semplificato dalla seguente relazione (vedi Fig. 11.5).

$$t_a = \overline{t_a}\left[1 + \alpha_a \left(\frac{x_a}{K_a}\right)^{\beta_a}\right] \tag{11.2}$$

dove $\overline{t_a}$ rappresenta il tempo richiesto per la percorrenza del tratto quando sgombro (detto *tempo a vuoto*), K_a la *capacità fisica* della strada espressa in numero di veicoli, α_a e β_a sono parametri dipendenti dalle caratteristiche del tratto. Come si nota dal grafico in Figura 11.5, l'andamento del tempo di percorrenza t_a aumenta polinomialmente in funzione del flusso a differenza di quanto si aveva nel caso autostradale dove incrementava linearmente il tempo di percorrenza del tratto fino al

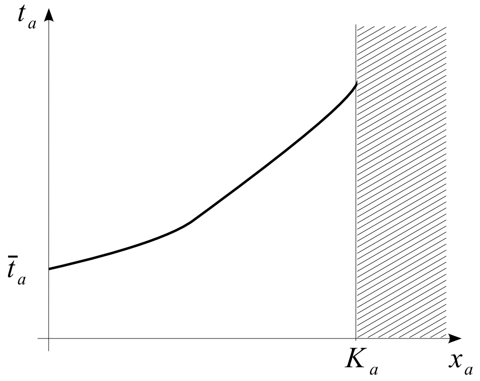


Figure 11.5: Tempo di percorrenza su arco stradale urbano

raggiungimento del limite K_a. Anche in questo caso, la validità della (11.2) è garantita solo dove $x_a \leq K_a$.

Principi di Wardrop Per descrivere il comportamento del traffico privato si utilizzano i *Principi di Wardrop.* Essi tengono conto delle cosiderazioni fatte fino a questo punto e tentano di modellare il processo decisionale del singolo una volta messo di fronte al problema della scelta del percorso da seguire.

Siano P_{od} l'insieme dei percorsi da o a d, λ_{od} il flusso complessivo dall'origine o alla destinazione d ed il coefficiente binario δ_{pa} l'indicatore di appartenenza del tratto stradale a ad uno dei percorsi p utilizzati ($\delta_{pa} = 1$ in caso positivo, $\delta_{pa} = 0$ altrimenti). Sia inoltre la variabile x_a il flusso sul tratto stradale a e λ^p il flusso presente sul singolo percorso p. I vincoli di conservazione diventano:

$$\sum_{p \in P_{od}} \lambda^p = \lambda_{od} \qquad \forall o, \, \forall d \tag{11.3}$$

$$x_a = \sum_p \lambda^p \delta_{pa} \qquad con \quad \delta_{pa} = \begin{cases} 1 & \text{se } a \in p \\ 0 & \text{se } a \notin p \end{cases} \tag{11.4}$$

$$\lambda^p \geq 0 \qquad \forall p \tag{11.5}$$

Primo principio di Wardrop *Il tempo di viaggio su ogni percorso usato è minore o uguale al tempo di viaggio che un singolo veicolo sperimenterebbe se usasse uno qualsiasi dei percorsi rimasti inutilizzati.* Ciò equivale a dire che ogni singolo utente cerca di minimizzare, in modo non cooperativo, il proprio tempo di viaggio. I flussi di traffico che soddisfano questo principio sono detti *di equilibrio.* Il punto di equilibrio è raggiunto quando nessun utente, tramite una decisione personale, può diminuire il proprio tempo di viaggio. Il primo principio di Wardrop afferma che:

$$T^*_{od} \leq T^p \qquad \forall p \in P_{od} \tag{11.6}$$

$$\lambda^p \left(T^p - T^*_{od}\right) \quad = 0 \quad \forall p \in P_{od} \tag{11.7}$$

In particolare, la (11.7) è detta *vincolo di complementarità* ed impone che, se il percorso p non è ottimo, il flusso su di esso è nullo e non passa alcun veicolo ($T^p > T^*_{od} \Rightarrow \lambda_p = 0$).

Secondo principio di Wardrop Si consideri ora il tempo totale T^p del percorso p dato dalla somma di tutti i tempi degli archi stradali a che lo compongono;

$$T^p = \sum_{a \in p} t_a(x_a)$$

Si consideri anche la seguente

$$T^*_{od} = \min_{p \in P_{od}} T^p$$

che rappresenta il tempo ottimo di percorrenza dall'origine o alla destinazione d. L'equazione (11.7) è la condizione necessaria di equilibrio ed assume lo stesso ruolo che le condizioni di Karush-Kuhn-Tucker ricoprono nell'ambito della Programmazione Matematica (vedere capitolo dedicato). Infatti, il secondo principio di Wardrop afferma che *all'equilibrio il tempo medio di viaggio è minimo*:

$$\varphi_x = \sum_{a \in A} \int_0^{x_a} t_a(x)dx \text{ è minimo}$$

Per dimostrare la relazione (11.7), si consideri la funzione φ si cerchi di minimizzarla nel rispetto dei vincoli (11.3), (11.4) e (11.5). Le condizioni di Karush-Kuhn-Tucker diventano:

$$l = \sum_{a \in A} \int_0^{x_a} t_a(x)dx + \sum_{od} \mu_{od} \left(\sum_{p \in P_{od}} \lambda^p - \lambda_{od} \right) + \sum_{a \in A} \mu_a \left(x_a - \sum_{p} \lambda^p \delta_{pa} \right) + \sum_{p} \mu_p \lambda^p$$

$$\frac{dl}{dx_a} = t_a(x_a) + \mu_a = 0 \quad \Longrightarrow \quad \mu_a = -t_a(x_a)$$

$$\frac{dl}{d\lambda^p} = \mu_{od} - \sum_{a \in A} \mu_a \delta_{pa} + \mu_p = 0 \qquad \Longrightarrow \qquad \sum_{a \in A} \delta_{pa} t_a(x_a) + \mu_{od_p} + \mu_p = 0$$

$$\mu_p = \mu_{od_p} - \sum_{a \in A} t_a(x_a)\delta_{pa} \geq 0 \quad \Longrightarrow \quad T^p - T^* \geq 0$$

$$x_a = \sum_{p} \lambda^p \delta_{pa}$$

$$\lambda_{od} = \sum_{p \in P} \lambda^p$$

$$\mu_p \lambda^p = 0$$

$$\mu_p \geq 0$$

dalle quali si ottiene la (11.7). Quindi si minimizza il tempo *medio* di viaggio nell'intervallo $(0; x_a)$ ma sulla rete *tutti* subiscono il tempo associato al flusso massimo x_a che è maggiore. Come si può notare quindi, la situazione di equilibrio che si configura non è ottimale per alcun viaggiatore.

Il paradosso di Braess Un esempio significativo delle conseguenze dei principi di Wardrop è costituito dalla seguente situazione. Si consideri una rete stradale, modellata nel grafo di Fig. 11.6, dove l'andamento del tempo di percorrenza dei

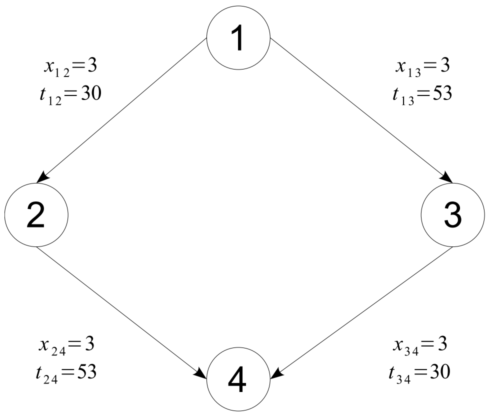


Figure 11.6: Paradosso di Braess, situazione di equilibrio iniziale

singoli tratti è descritto dalle seguenti equazioni

$$\begin{aligned} t_{12} &= 10x_{12} \\ t_{13} &= x_{13} + 50 \\ t_{24} &= 10x_{24} \\ t_{34} &= x_{34} + 50 \end{aligned}$$

e la rete soggetta ad una domanda di trasporto pari a $\lambda_{14} = 6$. Seguendo il primo principio di Wardrop, la situazione di ottimalità si raggiunge suddividendo equamente il flusso in uscita dal nodo 1, ottenendo così tempi di percorrenza equivalenti su entrambi i percorsi possibili:

$$\begin{aligned} p_1 &= \{1,2,4\} & t_1 &= 83 \\ p_2 &= \{1,3,4\} & t_2 &= 83 \end{aligned}$$

Il decisore si pone però il problema di diminuire il tempo di percorrenza, migliorando così la rete stradale sotto il suo controllo. Per raggiungere in modo semplice questo obiettivo egli decide di aggiungere un nuovo tratto molto veloce che collega i nodi 2 e 3: il tempo di percorrenza del nuovo passaggio è descritto dalla seguente

$$t_{23} = x_{23} + 10$$

e la situazione che si viene a creare è quella rappresentata in Fig. 11.7. Considerando i percorsi possibili nella situazione finale, si ha:

$$\begin{aligned} p_1 &= \{1,2,4\} & t_1 &= 92 \\ p_2 &= \{1,3,4\} & t_2 &= 92 \\ p_3 &= \{1,2,3,4\} & t_3 &= 92 \end{aligned}$$

Il primo principio di Wardrop può essere interpretato anche dal punto di vista singolare: ognuno tenta di minimizzare il tempo di percorrenza del proprio percorso in modo assolutamente egoistico, senza nessun tipo di coordinazione, portando ad una situazione finale che può essere peggiore per tutti. Il paradosso consiste proprio nell'aspetto di peggioramento per ognuno: se anche uno solo traesse beneficio dalla situazione finale, si tratterebbe di prepotenza di chi trova giovamento nei confronti di chi invece sarebbe costretto ad allungare il proprio viaggio. Si nota come nella situazione finale con la nuova strada ci sia una parte di viaggiatori (pari ad una

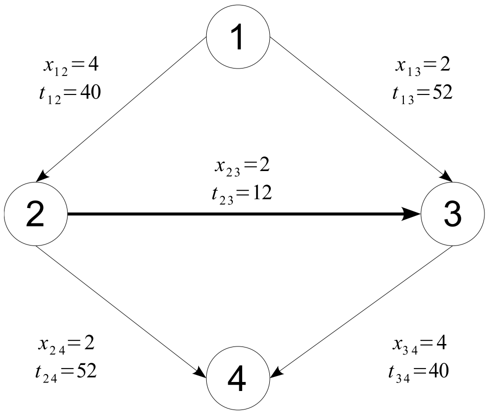


Figure 11.7: Paradosso di Braess, situazione di equilibrio finale

sola unità di flusso) che sceglie di cambiare percorso, passando dal nodo 2 per poter usufruire del nuovo tratto che si sa essere molto rapido. Questa decisione porta ad una perdita di flusso da parte delle strade più lente, che diventano leggermente più veloci, in favore di quelle più scorrevoli che diventano decisamente più lente. In questo caso, la situazione finale vede un deciso aumento del tempo di percorrenza di *tutti* i viaggiatori che da 83 passa a 92. La nuova situazione di equilibrio porta quindi a percorsi più lenti: aprire una nuova strada ha portato ad un netto peggioramento. Questi sono fenomeni che si osservano nella realtà e questo paradosso risulta significativo anche nell'ambito della Teoria dei Giochi: in questo caso tutti i giocatori (i viaggiatori) defezionano, tentati da una strada veloce, portando ad un'inevitabile punizione collettiva. L'aspetto di non-collaborazione del primo principio di Wardrop è così verificato; nel caso in cui tutti collaborassero, l'apertura di una nuova strada non modificherebbe i flussi in quanto tutti i viaggiatori la ignorerebbero (scegliendo di non defezionare ma di collaborare): basta che anche uno solo decida di farsi tentare per portare inevitabilmente alla defezione di massa.

Chapter 12

Modelli di teoria delle code

Chapter 13

Modelli di simulazione a eventi discreti

Chapter 14

Modelli dinamici

Appendix A

Richiami di Analisi Matematica

Queste note riassumono alcuni concetti di Analisi Matematica necessari per affrontare i fondamenti della Programmazione Matematica (Cap. 1).

A.1 Richiami sulle funzioni di una sola variabile

Una funzione di una sola variabile $u = f(x)$ è una legge che fa corrispondere a ogni valore x di un opportuno insieme S in $\mathbb{R}$ uno o più valori reali $u \in \mathbb{R}$.

$$u = f(x) \quad \text{oppure}$$

In generale, si intende che $f(x)$ assuma uno e un solo valore per ogni x in S.

Al variare di P in S, il punto $(P, f(P))$ descrive un luogo geometrico che si chiama *grafico* della funzione, ed è in generale una linea sul piano x, f. La Programmazione Matematica si occupa di determinare massimi e minimi di tale linea, con i relativi punti di massimo e minimo.

Teorema di Lagrange (teorema del valor medio) Se $f(x)$ è continua in un intervallo $[a; b]$ e derivabile in $(a; b)$, esiste un punto $\xi \in (a, b)$ tale che $f(b) - f(a) = (b - a) f'(\xi)$.

Questo teorema consente di esprimere il valore di una funzione nell'intorno di un punto x^*. Infatti, per ogni coppia di punti x^* e x in $[a; b]$ risulta

$$f(x) = f(x^*) + (x - x^*) f'(\xi)$$

dove ξ è un opportuno punto compreso fra x^* e x.

Formula di Taylor e formula di MacLaurin Se $f(x)$ è continua in un intervallo $[a; b]$ con le sue derivate sino all'ordine $n - 1$ ed esiste la sua derivata n-esima in $(a; b)$, per ogni coppia di punti x^* e x in $[a; b]$ risulta

$$f(x) = \sum_{k=0}^{n-1} \frac{f^{(k)}(x^*)}{k!} (x - x^*)^k + \frac{f^{(n)}(\xi)}{n!} (x - x^*)^n =$$

$$= f(x^*) + \frac{f'(x^*)}{1!} (x - x^*) + \ldots + \frac{f^{(n-1)}(x^*)}{(n-1)!} (x - x^*)^{n-1} + \frac{f^{(n)}(\xi)}{n!} (x - x^*)^n$$

dove $\frac{f^{(n)}(\xi)}{n!}(x-x^*)^n$ si chiama *resto nella forma di Lagrange.*

Se in x^* esiste anche la derivata n-esima, si può scrivere

$$f(x) = \sum_{k=0}^{n-1} \frac{f^{(k)}(x^*)}{k!}(x-x^*)^k + \frac{(x-x^*)^n}{n!}\left(f^{(n)}(x^*) + \alpha_n(x)\right) =$$

dove $\lim_{x\to x^*} \alpha_n(x) = 0$, oppure

$$f(x) = \sum_{k=0}^{n} \frac{f^{(k)}(x^*)}{k!}(x-x^*)^k + o_n(x)$$

dove $o_n(x) = \frac{\alpha_n(x)}{n!}(x-x^*)^n$ si chiama *resto nella forma di Peano* ed è *infinitesimo di ordine superiore a n* rispetto a $(x-x^*)$, cioè $\lim_{x\to x^*} \frac{o_n(x)}{(x-x^*)^n} = 0$.

Il vantaggio della forma di Peano consiste nel far dipendere tutti i coefficienti dalle derivate di f in x^*, lasciando un termine infinitesimo che si può spesso eliminare attraverso un passaggio al limite.

Per $x^* = 0$, la formula di Taylor dà luogo alla *formula di MacLaurin.*

$$f(x) = \sum_{k=0}^{n} \frac{f^{(k)}(0)}{k!}x^k + o_n(x) \quad \text{con} \lim_{x\to 0} \frac{o_n(x)}{x^n} = 0$$

A.2 Funzioni di più variabili

Una funzione di più variabili $u = f(P)$ è una legge che fa corrispondere a ogni punto P di un opportuno insieme S in $\mathbb{R}^n$ uno o più valori reali $u \in \mathbb{R}$.

$$u = f(P) \quad \text{oppure} \quad u = f(x_1, x_2, \ldots, x_n)$$

In generale, si intende che $f(P)$ assuma uno e un solo valore per ogni P in S.

Al variare di P in S, il punto $(P, f(P))$ descrive un luogo geometrico che si chiama *grafico* della funzione, ed è in generale una (iper)superficie a n dimensioni in uno spazio a $n+1$ dimensioni. La Programmazione Non Lineare si occupa di determinare massimi e minimi di tale superficie, con i relativi punti di massimo e minimo.

Limiti di funzioni a più variabili Il concetto di limite si estende banalmente da funzioni di una a funzioni di più variabili.

$$\lim_{P\to P^*} f(P) = l$$

significa che per ogni $\epsilon > 0$ esiste un valore $\delta_\epsilon > 0$ tale che $|f(P) - l| < \epsilon$ per ogni $P \in \mathcal{U}_{P^*,\delta_\epsilon}$, cioè per ogni P tale che $\|P - P^*\| < \delta_\epsilon$.

Geometricamente, ciò significa che per ogni intervallo di semiampiezza ϵ intorno a l si riesce a trovare una sfera di raggio δ_ϵ centrata in x^* tale che il valore di f in tutti i punti della sfera cade nell'intervallo.

La distanza $\|P - P^*\|$ va definita opportunamente, ed è in genere quella euclidea.

$$\|P - P^*\| = \sqrt{\left(x_1 - x_1^*\right)^2 + \left(x_2 - x_2^*\right)^2 + \dots}$$

Una funzione $f(P)$ è continua in P^* quando

$$\lim_{P \to P^*} f(P) = f(P^*)$$

Composizione di funzioni Supponiamo che $z = f(P) = f(x_1, \dots, x_n)$ sia definita in un sottoinsieme S del piano $\mathbb{R}^n$ e che le variabili x_i siano a loro volta funzioni $x_i = \phi_i(Q) = \phi_i(\xi_1, \dots, \xi_m)$ in un sottoinsieme Σ di $\mathbb{R}^m$. Allora z risulta essere funzione composta delle variabili ξ_j in Σ.

Si dimostra che, se ciascuna funzione ϕ_i ammette limite finito l_i per $Q \to Q_0$ e se f è continua in $P^* = (l_1, \dots, l_n)$, allora la funzione composta è continua in Q_0 e il limite è $f(\phi(Q_0), \psi(Q_0))$. Ad esempio, per $n = 2$:

$$\begin{cases} z = f(x_1, x_2) \\ x_1 = \phi_1(Q) \\ x_2 = \phi_2(Q) \\ \lim\limits_{Q \to Q_0} \phi_1(Q) = l_1 \\ \lim\limits_{Q \to Q_0} \phi_2(Q) = l_2 \\ f(P) \, \text{continua in } P^* = (l_1, l_2) \end{cases} \Rightarrow \begin{cases} z = f(\phi_1(Q), \phi_2(Q)) \text{ è continua in } Q_0 \\ \lim\limits_{Q \to Q_0} f(\phi_1(Q), \phi_2(Q)) = f(\phi_1(Q_0), \phi_2(Q_0)) \end{cases}$$

Derivate parziali Data una funzione di due variabili x e y, definita in un campo (insieme aperto e non vuoto), si attribuisca alla sola x un incremento $h \neq 0$, mantenendo invariata la y. Con h piccolo, il nuovo punto è ancora valido per la funzione. Si può definire il *rapporto incrementale parziale* e il suo limite, che è la *derivata parziale* rispetto a x. Analoghe sono le definizioni per la variabile y.

$$f_x = \frac{\partial f}{\partial x} = \lim_{h \to 0} \frac{f(x+h, y) - f(x, y)}{h}$$

Esempio: $z = e^{x/y}$ ha derivate parziali $z_x = \frac{1}{y} e^{x/y}$ e $z_y = -\frac{x}{y^2} e^{x/y}$

Le derivate parziali ammettono un'interpretazione geometrica: sono le tangenti trigonometriche degli angoli che le rette tangenti alla superficie $(P, f(P))$ in P e parallele ai piani xz e yx formano con gli assi x e y.

Esempio: $z = x^2 + 2y^2$ è un paraboloide con vertice nell'origine. Si vede che le tangenti vanno via via crescendo quando ci si allontana dall'origine e che crescono più in fretta rispetto alla y che rispetto alla x.

Una funzione può essere discontinua, ma derivabile parzialmente, al contrario di quanto avviene per le funzioni di una sola variabile. Basta infatti che siano contine le sue restrizioni agli assi.

Esempio:

$$f(x, y) = \begin{cases} \dfrac{xy}{x^2 + y^2} & \text{per } (x, y) \neq (0, 0) \\ 0 & \text{per } (x, y) = (0, 0) \end{cases}$$

in $(0, 0)$ ammette entrambe le derivate parziali, perché lungo gli assi è identicamente nulla. D'altra parte, non esiste il limite per $(x, y) \to (0, 0)$; infatti, se ci si avvicina lungo la bisettrice $y = x$, f ha limite $1/2$; se ci si avvicina lungo l'altra bisettrice $y = -x$ ha limite $-1/2$; se ci si avvicina lungo gli assi, ha limite 0.

Derivazione delle funzioni composte di una e più variabili Sia $f(x,y)$ definita in un campo A, con x e y entrambe funzioni di t in un intervallo $(a;b)$ e supponiamo che (x,y) resti in A mentre t varia in $(a;b)$. Allora, f è funzione composta di t nell'intervallo $(a;b)$.

Se $f \in C^1(A)$ e ϕ e ψ sono in $C^1(a,b)$, allora la funzione composta è $C^1(a,b)$ e la sua derivata è

$$f'(t) = f_x(\phi(t), \psi(t))\,\phi'(t) + f_y(\phi(t), \psi(t))\,\psi'(t)$$

Dimostrazione: La dimostrazione si basa sul teorema di Lagrange: per piccoli incrementi di t, x e y cambiano poco e i punti $(x+\delta x, y)$ e $(x+\delta x, y+\delta y)$ cadono in A. Quindi la variazione δz è la somma dell'incremento dovuto alla variazione di x e di quello dovuto alla variazione di y.

$$f(t+h) - f(t) = f(\phi(t+h), \psi(t+h)) - f(\phi(t), \psi(t)) =$$

$$= f(\phi(t+h), \psi(t+h)) - f(\phi(t+h), \psi(t)) + f(\phi(t+h), \psi(t)) - f(\phi(t), \psi(t))$$

$$= f_y(\phi(t+h), \eta)(\psi(t+h) - \psi(t)) + f_x(\xi, \psi(t))(\phi(t+h) - \phi(t))$$

dove $\eta \in (t, t+h)$ e $\xi \in (t, t+h)$. Quindi, tendendo al limite per $h \to 0$:

$$f'(t) = \lim_{h\to 0} \frac{f(t+h) - f(t)}{h} =$$

$$= \lim_{h\to 0}\left[f_y(\phi(t+h), \eta)\frac{\psi(t+h) - \psi(t)}{h} + f_x(\xi, \psi(t))\frac{\phi(t+h) - \phi(t)}{h}\right]$$

per la continuità di f_x e f_y e per la definizione di derivata, segue la tesi.

Analogamente, se le n variabili x_i sono funzioni di m variabili ξ_j

$$f_{\xi_j} = \frac{\partial f}{\partial \xi_j} = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}\frac{\partial x_i}{\partial \xi_j}$$

Basta ripetere il ragionamento, tenendo ogni volta fisse $m-1$ variabili.

Derivata direzionale Consideriamo una funzione f definita in un campo A, un punto $P^* \in A$ e un asse r, con coordinate misurate a partire da P^*. Se consideriamo i valori di f solo in corrispondenza ai punti dell'asse r, si possono definire un rapporto incrementale e una derivata rispetto a r. Questa è la tangente trigonometrica dell'angolo che la retta tangente alla superficie in P e parallela al piano rz forma con l'asse r.

Abbiamo quindi infinite possibili derivate: le derivate parziali sono quelle per cui r coincide con uno degli assi. Fortunatamente, per conoscere le infinite derivate direzionali, è sufficiente conoscere le n derivate parziali. Scrivendo x e y come funzioni di r con le coordinate polari:

$$x = x^* + r\cos\theta \text{ e } y = y^* + r\sin\theta$$

la funzione f ristretta ai punti dell'asse r si può valutare come funzione composta, e la derivata direzionale si può ottenere come sopra descritto, ottenendo

$$f_r = f_x\frac{dx}{dr} + f_y\frac{dy}{dr} = f_x\cos\theta + f_y\sin\theta$$

Gradient The vector whose components are the partial derivatives of a function with respect to the variables is denoted as *gradient* of the surface. It identifies a *vectorial field*, that is a function of several variables with vectorial values.

$$\nabla f = \begin{bmatrix} \dfrac{df}{dx_1} \\ \dots \\ \dfrac{df}{dx_n} \end{bmatrix}$$

L'asse r è univocamente individuato dal suo *versore* u_r (vettore di norma unitaria)

$$u_r = \begin{bmatrix} \cos\theta_1 \\ \dots \\ \cos\theta_n \end{bmatrix}$$

dove θ_i è l'angolo che r forma con l'asse x_i e i coseni si chiamano *coseni direttori.*

Quindi, la derivata direzionale lungo l'asse r è pari al prodotto scalare del vettore gradiente con il versore di r.

$$f_r = f_x \cos\theta + f_y \sin\theta = (\nabla f)^T \cdot u_r$$

e questo implica che *la derivata direzionale è massima quando viene calcolata nella direzione del gradiente* e nulla in direzione ortogonale.

Il gradiente indica la direzione di più rapido incremento della funzione. Dette *linee di livello* i luoghi dei punti in cui $f(P) = k$, cioè la funzione ha valore uniforme, *il gradiente è in ogni punto normale alle linee di livello.*

Piano tangente La relazione fra le derivate direzionali si traduce in una relazione geometrica fra le rette tangenti al grafico di f. Infatti, i punti della tangente lungo l'asse r in $(x^*, y^*, f(x^*, y^*))$ si possono scrivere come

$$z - f(x^*, y^*) = f_r(0)\, r = [f_x(x^*, y^*) \cos\theta + f_y(x^*, y^*) \sin\theta]\, r$$

ma poiché $r\cos\theta = x - x^*$ e $r\sin\theta = y - y^*$:

$$z - f(x^*, y^*) = f_x(x^*, y^*)(x - x^*) + f_y(x^*, y^*)(y - y^*)$$

che rappresenta un piano: *tutte le rette tangenti a una superficie regolare in un punto giacciono su un piano.* Questo piano si dice *piano tangente.*

Formula di Taylor per funzioni di più variabili Ora possiamo tirare le somme: dato un punto P^*, approssimiamo il valore di f in qualsiasi punto P nell'intorno di P^* semplicemente tracciando un asse da P^* a P, considerando la restrizione di f a tale asse ed approssimandola con la formula di Taylor. In essa compare la derivata direzionale, che però si può esprimere in funzione delle derivate parziali.

Se $f(x, y) \in C^n(A)$, si consideri un punto $P^* = (x^*, y^*)$ e un punto $P = (x, y)$ tale che il segmento P^*P cada interamente in A. I punti del segmento hanno coordinate $(x^* + t(x - x^*), y^* + t(y - y^*))$ con $t \in [0; 1]$. Lungo il segmento, f è funzione composta di t. La formula di MacLaurin fornisce il valore di f lungo il segmento (in particolare in $t = 1$):

$$f(1) = \sum_{k=0}^{n-1} \frac{f^{(k)}(0)}{k!} 1^k + \frac{f^{(n)}(\xi)}{n!} 1^n$$

con un opportuno $\xi \in (0;1)$. Arrestandosi al primo termine ($n = 1$), considerando $t = 1$ e riscrivendo nelle variabili x e y, si ottiene

$$f(x, y) = f(x^*, y^*) + [f_x(P(\xi))(x - x^*) + f_y(P(\xi))(y - y^*)]$$

dove $P(\xi) = (x^* + \xi(x - x^*), y^* + \xi(y - y^*))$.

Se le derivate prime sono continue, al resto in forma di Lagrange si può sostituire il resto in forma di Peano

$$f(x, y) = f(x^*, y^*) + [f_x(x^*, y^*)(x - x^*) + f_y(x^*, y^*)(y - y^*)] + o_1(x, y)$$

con $\lim\limits_{P \to P^*} \dfrac{o_1(P)}{\|P - P^*\|} = 0$.

A.3 Linee in forma parametrica

Consideriamo due funzioni reali di variabile reale t definite e continue in un *insieme base* S. Se le interpretiamo come coordinate cartesiane di un punto, il loro insieme definisce una *linea in forma parametrica.* Il parametro t stabilisce un orientamento sulla linea.

Se un punto $P(t) = (x(t), y(t))$ di una linea in forma parametrica deriva da diversi valori di t, cioè se la linea si ripiega su sé stessa, si parla di *punto multiplo.*

Esempio:

$$\begin{cases} x(t) = t^2 \\ y(t) = t^3 - t \end{cases} \quad \text{con } t \in [-2, 2]$$

Esempio:

$$\begin{cases} x(t) = t^2 \\ y(t) = t^2 \end{cases} \quad \text{con } t \in [-1, 1]$$

Una *linea semplice* è priva di punti multipli. Essa crea una corrispondenza biunivoca fra punti e valori di t.

Vettore tangente Si può estendere al vettore $P(t)$ il concetto di derivata, sempre basandolo sul rapporto incrementale: il risultato è un vettore (detto *vettore tangente* alla linea) le cui componenti sono le derivate delle singole componenti.

$$P'(t) = \begin{bmatrix} x'(t) \\ y'(t) \end{bmatrix}$$

Il piano normale al vettore tangente si dice *piano normale* alla linea.

Linee regolari Il concetto di linea in forma parametrica include un insieme di oggetti più ampio di quello suggerito dall'intuizione. Ad esempio, $x(t) = 1$ e $y(t) = 2$ è una linea in forma parametrica, pur riducendosi geometricamente a un solo punto. Ancora, la curva di Peano è una linea in forma parametrica, anche se geometricamente ricopre un intero quadrato.

Per evitare queste patologie, ci concentriamo sulle *linee regolari*:

1. *semplici*

2. aventi come *insieme base un intervallo* (a, b)
3. dotate di *vettore tangente continuo e non nullo* (cioè le singole derivate sono continue e non contemporaneamente nulle)

Esempio:

$$\begin{cases} x(t) = \sin t \\ y(t) = \cos t \end{cases} \quad \text{con } t \in [-\pi/2, \pi/2] \ \text{ è una linea regolare}$$

Esempio:

$$\begin{cases} x(t) = \sin t \\ y(t) = \cos t \end{cases} \quad \text{con } t \in [-\pi, \pi]$$

è un caso limite (*ciclo*) perché il punto $P(-\pi) = P(\pi) = (0, -1)$ è doppio.

Esempio:

$$\begin{cases} x(t) = t^2 \\ y(t) = t^3 \end{cases} \quad \text{con } t \in [-1, 1]$$

è semplice, ma non regolare: perché il vettore tangente $P'(t) = \left[2t \ 3t^2\right]^T$ si annulla in $(0, 0)$.

Rappresentazioni Una linea orientata ha infinite rappresentazioni parametriche: basta sostituire t con $t = \phi(\tau)$, dove ϕ è una funzione strettamente monotona in un intervallo. Per mantenere la regolarità, occorre e basta che ϕ sia continua e abbia derivata continua e mai nulla.

In particolare, la *rappresentazione cartesiana* impiega come parametro una delle coordinate. Una linea regolare ammette sempre una rappresentazione cartesiana locale, cioè valida nell'intorno di un punto $P^* = P(t_0)$: basta invertire una delle funzioni componenti, dato che non possono avere tutte derivata nulla. Però la rappresentazione cartesiana in generale copre solo un intorno di P^*.

Esempio (elica cilindrica):

$$\begin{cases} x(t) = R\cos t \\ y(t) = R\sin t \\ z(t) = \dfrac{p}{2\pi}t \end{cases} \quad \text{con } t \in \mathbb{R}$$

invertendo la terza relazione $z(t) = \dfrac{p}{2\pi}t \Leftrightarrow t(z) = \dfrac{2\pi}{p}z$, si traduce in

$$\begin{cases} x(t) = R\cos\dfrac{2\pi z}{p} \\ y(t) = R\sin\dfrac{2\pi z}{p} \end{cases} \quad \text{con } z \in \mathbb{R}$$

Altre rappresentazioni cartesiane si ottengono invertendo una delle prime due relazioni, ma esse valgono solo per un semigiro dell'elica, cioè in un intorno del punto scelto.

Appendix B

Linear algebra

B.1 Vector combinations

Definition 68 *Given a set of m vectors g_j in space $\mathbb{R}^n$, we denote as* linear combination *of such vectors any vector obtained multiplying them by real coefficients and summing the results:*

$$f = \sum_{j=1}^{m} \mu_j g_j$$

A linear combination is (see Figure B.1):

- *a* conic combination *when the coefficients μ_j are all nonnegative:*

$$\mu_j \geq 0 \textit{ for } j = 1, \ldots, m$$

- an affine combination *when the coefficients μ_j sum to 1:*

$$\sum_{j=1}^{m} \mu_j = 1$$

- *a* convex combination *when it is both conic and affine:*

$$\mu_j \geq 0 \textit{ for } j = 1, \ldots, m \textit{ and } \sum_{j=1}^{m} \mu_j = 1$$

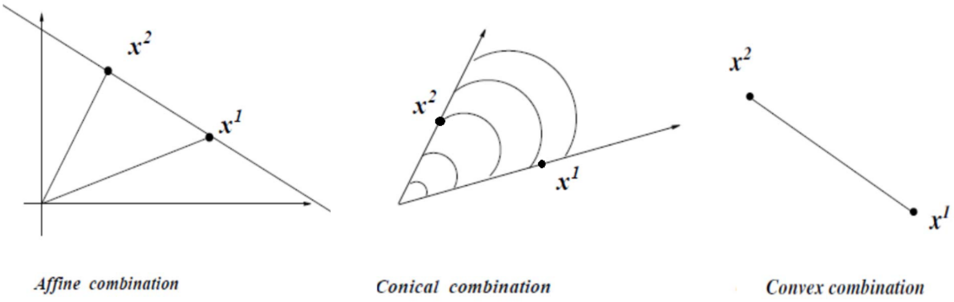


Figure B.1: Combinations of two vectors: conic, affine, convex

Appendix C

Richiami di calcolo delle probabilità

Seguiremo un approccio assiomatico, cioè definiremo le proprietà matematiche degli oggetti che ci occorrono, senza entrare nei (grossi) problemi filosofici di come dare loro un significato pratico. Vi sono approcci molto diversi alla soluzione di tali problemi.

1. È dato uno *spazio degli eventi* Ω, che contiene tutti i possibili *esiti* ω di un esperimento casuale; è anche data una collezione $\mathcal{A}$ di sottoinsiemi dello spazio Ω detti *eventi*. Altrove nel corso gli esiti sono definiti scenari.

2. Ad ogni evento $A \in \mathcal{A}$ è associato un numero reale non negativo $P(A)$, detto *probabilità* di A
$$P : \mathcal{A} \to \mathbb{R}^+$$

3. L'intero spazio è un evento (*evento certo*) di probabilità pari a 1
$$P(\Omega) = 1$$

4. Se due eventi A e B contengono esiti distinti, la probabilità della loro unione coincide con la somma delle loro probabilità
$$A \cap B = \emptyset \Rightarrow P(A \cup B) = P(A) + P(B)$$

5. Se A_n è una successione di eventi contenuti ciascuno nel precedente e al tendere di n all'infinito la loro intersezione tende all'insieme vuoto, allora $P(A_n)$ tende a zero:
$$A_n \to \emptyset \Rightarrow \lim_{n\to\infty} P(A_n) = 0$$

Dagli assiomi derivano le seguenti conseguenze o teoremi.

Teorema della probabilità totale La probabilità dell'unione di due o più eventi, cioè la probabilità che se ne verifichi almeno uno, è la somma delle probabilità dei singoli eventi meno la somma delle probabilità delle intersezioni due a due, più la somma delle probabilità delle intersezioni a tre a tre e così via. Ad esempio:

$$P(A\cup B\cup C) = P(A)+P(B)+P(C)-P(A\cap B)-P(A\cap C)-P(B\cap C)+P(A\cap B\cap C)$$

Se gli eventi sono a due a due incompabitibili, è la semplice somma delle probabilità.

Probabilità condizionata La *probabilità condizionata* di A dato B, indicata con $P(A|B)$ è la probabilità che l'evento A ha di verificarsi quando si sappia che B si è verificato:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

Teorema della probabilità composta La probabilità dell'intersezione di due o più eventi, ovvero la probabilità che essi si verifichino tutti è $P(A \cap B) = P(A|B)P(B)$.

Se la probabilità di A dato B, $P(A|B)$, è uguale a $P(A)$, i due eventi vengono definiti *indipendenti stocasticamente* (o probabilisticamente) e $P(A \cap B) = P(A)P(B)$.

Teorema di Bayes Dato un insieme finito o numerabile di eventi A_i, a due a due incompatibili, quando si sappia che si è verificato un evento E e che se E si verifica, allora si verifica necessariamente uno degli eventi di tale insieme (ed uno solo, dato che sono incompatibili), la probabilità a posteriori di A_i si può calcolare da quelle a priori degli eventi A_i e da quelle condizionate $P(E|A_i)$:

$$P(A_i|E) = \frac{P(E|A_i)P(A_i)}{\sum\limits_j P(E|A_j)P(A_j)}$$

In altre parole, se si conoscono sia le probabilità a priori delle diverse possibili "cause" di E (ma non si sa per effetto di quale di loro E si è verificato), sia le probabilità condizionate di E data ciascuna delle cause, è possibile calcolare la probabilità che E si sia verificato per effetto di una particolare causa.

Probabilità e densità di probabilità Se Ω è un insieme discreto, la funzione P definita su eventi induce una funzione probabilità $\pi_\omega : \Omega \to [0;1]$ definita sugli esiti, che è semplicemente la P sugli eventi costituiti da un singolo esito.

Se Ω è un insieme continuo, si definisce per ogni esito ω una *densità di probabilità* $\pi(\omega)$, come limite del rapporto fra la probabilità P e la misura Δ degli eventi che contengono ω

$$\pi(\omega) = \lim_{\|\Delta(A)\| \to 0} \frac{P(A)}{\|\Delta(A)\|}$$

dove $A \subseteq \Omega$ è un evento che va via via restringendosi, ma contiene sempre ω e $\|\Delta(A)\|$ è la sua misura.

Variabili aleatorie Una *variabile aleatoria* è una funzione che associa ad ogni esito ω un valore numerico. Per esempio, fissata una soluzione $\bar{x}$, l'impatto $f(\bar{x}, \omega)$ è una variabile aleatoria.

Funzione di distribuzione La probabilità che una variabile aleatoria abbia valore $\leq x$ è detta *funzione di distribuzione* $F(x)$: è una funzione monotona non decrescente di x.

$$F(x) = P(X \leq x)$$

Di conseguenza $F(x) \geq 0$ per ogni x e valgono i limiti

$$\lim_{x \to +\infty} F(x) = 1 \qquad \lim_{x \to -\infty} F(x) = 0$$

Inoltre

$$P(a < X \leq b) = F(b) - F(a)$$

Se X è una variabile casuale discreta, ossia ammette una collezione numerabile di possibili valori $x_1, \ldots, x_n, \ldots$, la distribuzione $F(x)$ è discontinua e presenta crescite brusche in corrispondenza ai valori possibili; l'entità di ogni salto coincide con la probabilità totale degli eventi in cui X ha quel valore.

$$F(x) = \sum_{x_i \leq x} f(x_i)$$

dove $f(x) = P(X = x)$ è detta funzione di densità discreta di X ed è, per ogni valore reale, la probabilità che la variabile X assuma esattamente quel valore.

Nel caso continuo, la funzione di densità $f(x)$ è il limite del rapporto fra la probabilità che X assuma valori in un intervallo contenente x e l'ampiezza di tale intervallo, e quindi indica su quali valori X tende a concentrarsi.

Di conseguenza, $f(x)$ è la derivata di $F(x)$ e, se X è una variabile casuale assolutamente continua

$$F(x) = \int_{-\infty}^{x} f(u)du$$

Valore atteso Il *valore atteso* $\mathbb{E}[X]$ di una variabile aleatoria X è definito

- per variabili discrete come

$$\mathbb{E}[X] = \sum_{i=1}^{+\infty} x_i f(x_i)$$

- per variabili assolutamente continue come

$$\mathbb{E}[X] = \int_{-\infty}^{+\infty} x f(x)dx$$