Decision Methods and Models

Master's Degree in Computer Science

Roberto Cordone

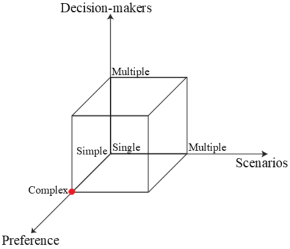
DI - Università degli Studi di Milano

Schedule:	Thursday 16.30 - 18.30 in Aula Magna (CS department)
	Friday 12.30 - 14.30 in classroom 301
Office hours:	on appointment
E-mail:	roberto.cordone@unimi.it
Web page:	https://homes.di.unimi.it/cordone/courses/2024-mmd/2024-mmd.html
Ariel site:	https://myariel.unimi.it/course/view.php?id=4467

Multiple criteria optimisation

We assume

- a preference relation Π that is no longer a weak order
- a certain environment: $|\Omega| = 1 \Rightarrow f(x, \bar{\omega})$ reduces to $f(x)$
- a single decision-maker: $|D| = 1 \Rightarrow \Pi_d$ reduces to Π



If Π is a preorder, we at least have nondominated solutions X°
but possibly incomparable with each other

Paretian preference

Vilfredo Pareto made the first historical proposal of a complex preference

$$f \preceq f' \Leftrightarrow f_l \leq f'_l \text{ for each } l \in P$$

meaning that each indicator is a cost

This is easily extended to the case in which the indicators are

- all benefits
- mixed costs and benefits

It is a very “poor” preference (including few impact pairs)

In pratice, it is often a starting point: a simplified restriction from which the real (richer) preference is built interacting with the decision-maker

Properties

Theorem

The Paretian preference is a partial order

Indeed, it is

- reflexive, because $f_l = f_l, \forall f \in F \Rightarrow f_l \leq f_l, \forall l \in P \Rightarrow f \preceq f, \forall f \in F$
- transitive, because $f \preceq f'$ and $f' \preceq f''$ imply $f_l \leq f'_l$ and $f'_l \leq f''_l, \forall l \in P$, but this clearly implies $f_l \leq f''_l, \forall l \in P$, and therefore $f \preceq f'', \forall f, f', f'' \in F$
- antisymmetric, because $f \preceq f'$ and $f' \preceq f$ imply $f_l \leq f'_l$ and $f'_l \leq f_l, \forall l \in P$, but this clearly implies $f_l = f'_l$ for each $l \in P$, and therefore $f = f', \forall f, f' \in F$
- not complete, because counterexamples exist:

$$\begin{bmatrix} 0 \\ 1 \end{bmatrix} \not\preceq \begin{bmatrix} 1 \\ 0 \end{bmatrix} \text{ and } \begin{bmatrix} 1 \\ 0 \end{bmatrix} \not\preceq \begin{bmatrix} 0 \\ 1 \end{bmatrix} \Rightarrow \begin{bmatrix} 0 \\ 1 \end{bmatrix} \bowtie \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

□

Therefore, incomparable impacts are possible

Domination

We can reformulate the concept of domination

The dominated solution set $X \setminus X^\circ$ includes the alternatives x for which

$$\exists x' \in X, \exists \bar{l} \in P : \begin{cases} f_l(x') \leq f_l(x) \text{ for all } l \in P \\ f_{\bar{l}}(x') < f_{\bar{l}}(x) \end{cases}$$

The nondominated solution set X°, known as Paretian region, includes the alternatives x for which

$$\forall x' \in X, \exists l' \in P : f_{l'}(x) < f_{l'}(x') \text{ or } f(x) = f(x')$$

Unless for infinite chains, X° is nonempty
and contains reciprocally incomparable solutions
(or indifferent solutions that have exactly the same impact)

We aim to enumerate X° and then ask the decision-maker

Paretian region enumeration

We will survey some methods to enumerate the Paretian region

Method	Generality	Practical limits	Result provided
definition	finite	slow for CO	exact
inverse transformation	$p = 2$	human intervention	exact
KKT conditions	regular	solving a system	overestimate
weigthed sum	general	parametric problem (inefficient sampling)	underestimate (bad for large p)
ϵ-constraint	general	parametric problem usually NP-hard (inefficient sampling)	overestimate (refined by repetition)

Applying the definition

For finite problems, X° can be found by pairwise comparisons, of course taking $\Theta\left(p\,|X|^2\right)$ time in the worst case

This can be huge in combinatorial problems, where $|X| \in \Theta(2^n)$

Example (choosing how to make a trip)

X	Time	Cost
Train	5.5	100
Car	4.0	150
Airplane	1.0	300
Coach	5.0	180
Taxi	4.0	400

The definition implies

- $4.0 < 5.0$ and $150 < 180 \Rightarrow$ Car $\prec$ Coach
- $4.0 \leq 4.0$ and $150 < 400 \Rightarrow$ Car $\prec$ Taxi
  (*equal times: the preference is strict, but we call it weak dominance*)
- $1.0 < 4.0$ and $300 < 400 \Rightarrow$ Airplane $\prec$ Taxi

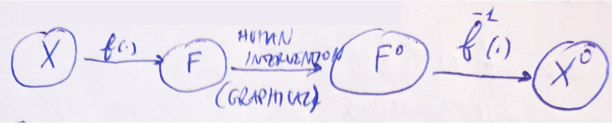
Inverse transformation method

This is a graphical method in the impact space $\mathbb{R}^p$

- it requires human intervention
- it can be applied only when the impact consists of two indicators

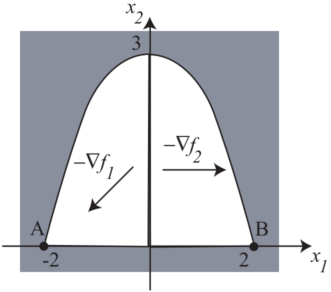
It consists of the following steps

1. compute the inverse $x(f)$ of the impact function $f(x)$
   (*if it is not unique, compute all inverse functions*)
2. build the image of X in F through the impact function $f(x)$
   by drawing the constraints $g_j(x(f)) \leq 0$
3. find graphically the nondominated impact subset $F^\circ \subseteq F$, exploiting the property that these impacts have an empty lower left quadrant
4. find a parametric form $f^\circ(\alpha)$ to describe F°
5. transform F° into X° through the inverse function $x(f^\circ(\alpha))$



Inverse transformation method: example

$$\begin{aligned} \min f_1 &= x_1 + x_2 \\ \min f_2 &= -x_1 \\ g_1 &= 3x_1^2 + 4x_2 - 12 \leq 0 \\ g_2 &= -x_2 \leq 0 \end{aligned}$$



1. compute the inverse $x(f)$ of the impact function $f(x)$

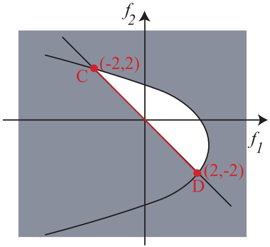
$$\begin{cases} f_1 = x_1 + x_2 \\ f_2 = -x_1 \end{cases} \Leftrightarrow \begin{cases} x_1 = -f_2 \\ x_2 = f_1 - x_1 = f_1 + f_2 \end{cases}$$

Inverse transformation method: example

2. build the image of X in F by drawing the constraints $g_j(x(f)) \leq 0$

$$\begin{cases} g1(x) = 3f_2^2 + 4(f_1 + f_2) - 12 \leq 0 \\ g2(x) = -f_1 - f_2 \leq 0 \end{cases} \Leftrightarrow \begin{cases} f_1 \leq -3/4f_2^2 - f_2 + 3 \\ f_1 + f_2 \geq 0 \end{cases}$$

3. find the nondominated impacts F° with an empty lower left quadrant



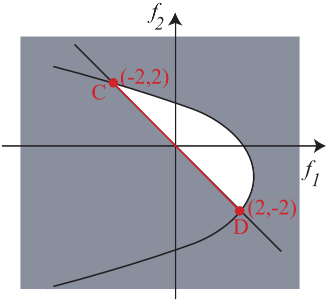
F° is segment $\overline{CD}$ with $C = (-2, 2)$ and $D = (2, -2)$

This is where human intervention and $p = 2$ are required

Inverse transformation method: example

❹ describe F° in parametric form $f^\circ(\alpha)$

$$\begin{cases} f_1^\circ = \alpha \\ f_2^\circ = -\alpha \end{cases} \quad \text{with } \alpha \in [-2, 2]$$



Inverse transformation method: example

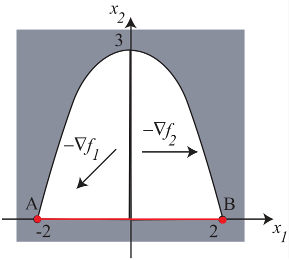
❺ transform F° into X° through the inverse function $x\left(f^\circ(\alpha)\right)$

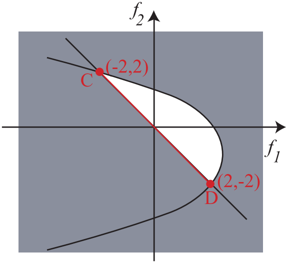
$$\begin{cases} x_1^\circ = x_1\left(f^\circ(\alpha)\right) = -f_2^\circ(\alpha) \\ x_2^\circ = x_2\left(f^\circ(\alpha)\right) = f_1^\circ(\alpha) + f_2^\circ(\alpha) \end{cases} \Rightarrow \begin{cases} x_1^\circ = \alpha \\ x_2^\circ = \alpha - \alpha = 0 \end{cases}$$

with $\alpha \in [-2, 2]$

This is segment $\overline{AB}$ with $A = (-2, 0)$ and $B = (2, 0)$

Notice that X° has $p - 1$ dimensions





Both f_1 and f_2 improve downwards; they conflict rightwards or leftwards

KKT conditions

The KKT conditions can be extended to Pareto preference by repeating the derivation with minor changes

A locally Paretian point is a feasible solution $x \in X$ not dominated by any other feasible solution in a suitably small neighbourhood

$$\exists \mathcal{U}_{\epsilon,x} : x' \not\prec x \text{ for all } x' \in \mathcal{U}_{\epsilon,x} \cap X$$

Let X^* denote the set of all locally Paretian points

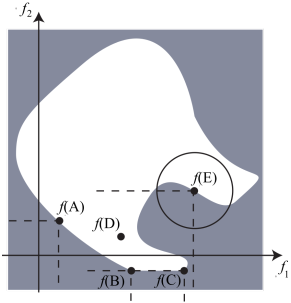
Theorem
If $X = \{x \in R^n : g_j(x) \leq 0 \text{ for } j = 1, \dots, m\}$ with $f_l(\cdot), g_j(\cdot) \in C^1(X)$, then

$$X^{KKT} \supseteq X^* \supseteq X^\circ$$

The KKT conditions are necessary for local paretianity

Global, local and weak Paretian points

First, generalise the concepts of globally and locally optimal point



In the picture (*notice that it shows the impacts, not the solutions!*)

- D is dominated (*there are impacts below and on the left of* $f(D)$)
- A and B are globally Paretian (*the lower left quadrant is empty*)
- E is locally Paretian (*the quadrant is empty in a neighbourhood*)
- D is weakly Paretian (*one indicator value is identical*)

Nonimprovement conditions

Theorem

If $\xi(\alpha)$ is a feasible arc in x and $\nabla f_l(x)^T p < 0$ for all $l \in P$, then x is not a locally Paretian point

It is a sufficient condition to prove that $\xi(\alpha)$ is an improving arc for $f(\cdot)$ in x, and therefore to filter x out of the candidate set X^{KKT}

Notice that improving all indicators is not necessary for strict dominance: Weakly Paretian points are dominated, but some indicators can't improve

Therefore, weakly Paretian points satisfy the KKT conditions

Feasibility conditions

Feasibility is treated in the same way as in Mathematical Programming

- nonregular points are candidate
- a regular point x is feasible if and only if

$$\nabla g_j(x)^T p \leq 0 \text{ for all } j \in J_a(x)$$

First geometric interpretation

Denote by

- feasible cone $C_{\text{feas}}(x)$ the set of vectors tangent to feasible arcs (scalar products ≤ 0 with all active constraint gradients)
- improving cone $C_{\text{impr}}(x)$: the set of improving vectors (scalar products < 0 with **all** objective gradient)

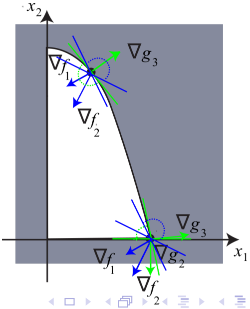
The first is close, the second open!

If a regular point is locally Paretian, then its feasible cone and improving cone do not intersect

$$x \in X^* \Rightarrow C_{\text{feas}}(x) \cap C_{\text{impr}}(x) = \emptyset$$

$$\begin{aligned} \min f_1 &= -2x_1 - x_2 \\ \min f_2 &= -x_1 - 2x_2 \\ g_1 &= -x_1 \leq 0 \\ g_2 &= -x_2 \leq 0 \\ g_3 &= x_1^2 + x_2 - 4 \leq 0 \end{aligned}$$

Check $A = (2, 0)$ and $B = (1/2, 15/4)$



Separating the improving and the feasible cone

In order to separate two cones, we need a separating hyperplane, that has two opposite orthogonal vectors:

- vector γ is on the opposite side of all $p \in C_{\text{feas}}$

$$p^T \gamma \leq 0 \text{ for all } p : p^T \nabla g_j(x) \leq 0 \text{ for all } j \in J_a(x)$$

- vector $-\gamma$ is on the opposite side of all $p \in C_{\text{impr}}$

$$p^T(-\gamma) \leq 0 \text{ for all } p : \nabla f_l(x)^T (-p) < 0 \text{ for all } l \in P$$

and now we can apply Farkas' lemma to both expressions, obtaining that

- vector γ falls in the cone of the gradients of the active constraints

$$\exists \mu_j \geq 0 : \gamma = \sum_{j \in J_a(x)} \mu_j \nabla g_j(x)$$

- vector $-\gamma$ falls in the cone of the gradients of the objectives

$$\exists w_l \geq 0 : -\gamma = \sum_{l \in P} w_l \nabla f_l(x)$$

and finally sum the two expressions

KKT conditions

$$\exists \mu_j \geq 0, w_l \geq 0 : \sum_{l \in P} w_l \nabla f_l(x) + \sum_{j \in J_a(x)} \mu_j \nabla g_j(x) = 0$$

Once again, we introduce the complementarity conditions

$$\sum_{l=1}^{p} w_l \nabla f(x) + \sum_{j=1}^{m} \mu_j \nabla g_j(x) = 0$$

$$\begin{aligned} \mu_j g_j(x) &= 0 \qquad j = 1, \dots, m \\ g_j(x) &\leq 0 \qquad j = 1, \dots, m \\ w_l &\geq 0 \qquad l = 1, \dots, p \\ \mu_j &\geq 0 \qquad j = 1, \dots, m \end{aligned}$$

Notice that for any solution (x, w, μ), also $(x, \alpha w, \alpha \mu)$ is a solution, that provides the same candidate point x

Then, introduce a normalisation condition

Normalisation condition

We could set $\sum_{l=1}^{p} w_l + \sum_{j=1}^{m} \mu_j = 1$, but we know that x is regular:

- if $w_l = 0$ for all $l \in P$, then $\sum_{j=1}^{m} \mu_j \nabla g_j(x) = \sum_{j \in J_a(x)} \mu_j \nabla g_j(x) = 0$,
- the active constraints should have linearly dependent gradients
- but in a regular point this is impossible!

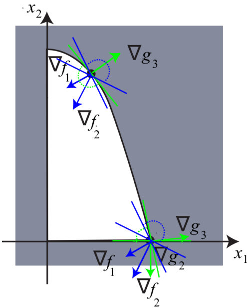
The w_l multipliers cannot be all equal to zero

Consequently, we add the normalisation condition

$$\sum_{l=1}^{p} w_l = 1$$

Example

$$\min f_1 = -2x_1 - x_2$$
$$\min f_2 = -x_1 - 2x_2$$
$$g_1 = -x_1 \leq 0$$
$$g_2 = -x_2 \leq 0$$
$$g_3 = x_1^2 + x_2 - 4 \leq 0$$



Setting $w_1 = w$ and $w_2 = 1 - w$, the KKT conditions become

$$\begin{aligned}
-2w - 1(1-w) - \mu_1 + 2x_1\mu_3 &= 0 \\
-w - 2(1-w) - \mu_2 + \mu_3 &= 0 \\
-\mu_1 x_1 &= 0 \\
-\mu_2 x_2 &= 0 \\
-\mu_3(x_1^2 + x_2 - 4) &= 0 \\
0 \leq w &\leq 1 \\
\mu_j &\geq 0 \\
g_j(x) &\leq 0
\end{aligned}$$

Example

The first equation becomes

$$2x_1\mu_3 = w + 1 + \mu_1 \Rightarrow \mu_3 > 0 \text{ and } x_1 > 0$$

which implies $\mu_1 = 0$ and $x_2 = 4 - x_1^2$

The equations reduce to

$$\begin{cases} x_1 = \dfrac{w+1}{2}\mu_3 \\ \mu_2 = w - 2 + \mu_3 \end{cases}$$

Example

Now we can split the problem with respect to the condition $\mu_2 x_2 = 0$

- in problem $\mu_2 > 0$, we have that $x_2 = 0 \Rightarrow x_1 = 2 \Rightarrow \mu_3 = \dfrac{w+1}{4}$, but this implies $\mu_2 = \dfrac{5w-7}{4} < 0$, unfeasible
- in problem $\mu_2 = 0$, we have that $\mu_3 = 2 - w$, which implies

$$\begin{cases} x_1 = \dfrac{w+1}{2(2-w)} \\ x_2 = 4 - x_1^2 \end{cases} \quad \text{with } w \in [0,1]$$

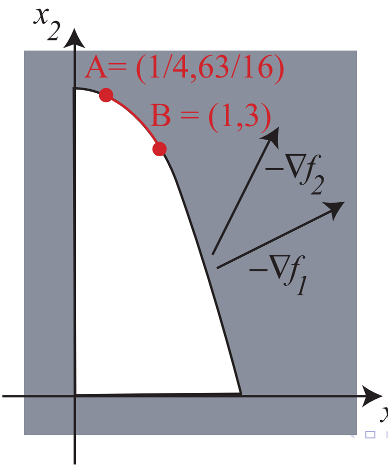
This is an arc of the parabola $x_2 = 4 - x_1^2$:

- it starts in $A = (1/4, 63/16)$ for $w = 0$
- it ends in $B = (1, 3)$ for $w = 1$
- since x_1 increases with w, the values of $w \in [0,1]$ correspond to the intermediate points

Example

This is an arc of the parabola $x_2 = 4 - x_1^2$:

- it starts in $A = (1/4, 63/16)$ for $w = 0$
- it ends in $B = (1, 3)$ for $w = 1$
- since x_1 increases with w, the values of $w \in [0, 1]$ correspond to the intermediate points



Equation balance

The KKT condition system has $n + p + m$ variables

- n variables for vector x
- p variables for vector w
- m variables for vector μ

and $n + 1 + m$ equalities

- n equations for the KKT conditions
- 1 equation for the normalisation condition
- m equations for the complementarity conditions

In general, it will have ∞^{p-1} solutions describing a hypersurface