Decision Methods and Models

Master's Degree in Computer Science

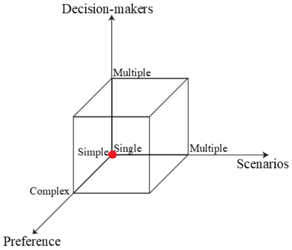
Roberto Cordone
DI - Università degli Studi di Milano

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

Mathematical Programming

We assume

- a preference relation Π with a known consistent utility function $u(f)$
- 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



The decision problem reduces to classical optimisation

$$\max u(f(x))$$
$$x \in X$$

We discuss a solving technique that is

- very general
- complex and inefficient

Basic assumptions

In mathematics, the most common form is

$$\min f(x)$$
$$x \in X$$

where $f(x)$ replaces $-u(f(x))$ *(It is not the original f!)*

We also assume regularity for the objective and the feasible region:

1. $f(x) \in C^1(X)$
2. $X = \{x \in \mathbb{R}^n : g_j(x) \le 0, j = 1, \ldots, m\}$ with $g_j(x) \in C^1(X)$

These are very general assumptions as

$$\max_{x \in X} f(x) \quad \Leftrightarrow \quad \min_{x \in X} -f(x)$$
$$g_j(x) \le a \quad \Leftrightarrow \quad g_j(x) - a \le 0$$
$$g_j(x) \ge a \quad \Leftrightarrow \quad a - g_j(x) \le 0$$
$$h_i(x) = 0 \quad \Leftrightarrow \quad \begin{cases} h_i(x) \le 0 \\ -h_i(x) \le 0 \end{cases}$$
$$x \in \mathbb{Z}^n \quad \Leftrightarrow \quad \sin(\pi x) = 0$$

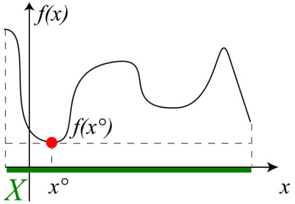
(computationally useless!)

Global and local optimum points

Given a set $X \subseteq \mathbb{R}^n$ and a function $f : X \to \mathbb{R}$

- global optimum point is a point $x^\circ \in X$ such that

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



Global and local optimum points

Given a set $X \subseteq \mathbb{R}^n$ and a function $f : X \to \mathbb{R}$

- global optimum point is a point $x^\circ \in X$ such that

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

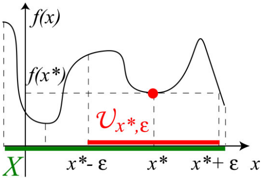
- local optimum point is a point $x^* \in X$ such that

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

$\mathcal{U}_{x^*,\epsilon} = \{x \in \mathbb{R}^n : \|x - x^*\| < \epsilon\}$ is a neighbourhood of x^* of radius ϵ

$\|x - x^*\|$ is the norm of vector $x - x^*$
(distance between x and $\tilde{x}$)

$$\|x -^*\| = \sqrt{\sum_{i=1}^{n} (x_i - x_i^*)^2}$$



All global optimum points are also local optimum points: $X^\circ \subseteq X^*$

The general process

Instead of X°, we pursue necessary conditions for local optimality

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

Then, we enumerate X^{KKT} exhaustively to find X°

The Karush-Kuhn-Tucker (*KKT*) conditions identify candidate points

1. solve the conditions to build the set of candidate points X^{KKT}
2. scan one by one the points in X^{KKT} comparing their values
3. the best ones yield X°

We hope that X^{KKT} is finite or $f(x)$ easy to optimise in it

The basic tool will be linear approximation in small neighbourhoods

This is why we will get false positives

Taylor's (first-order) series expansion

Any regular function can be locally approximated in $\tilde{x}$ by its tangent line

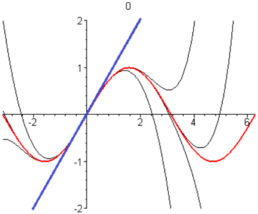
What happens to $f(x)$ moving a bit out of $\tilde{x}$?

If $f : \mathbb{R} \to \mathbb{R}$ and $f \in C^1(\mathcal{U}_{\tilde{x},\epsilon})$, then

$$f(x) = f(\tilde{x}) + f'(\tilde{x})(x - \tilde{x}) + R_1(|x - \tilde{x}|)$$

with $\lim_{x \to \tilde{x}} \frac{R_1(|x - \tilde{x}|)}{|x - \tilde{x}|} = 0$

Additional terms with higher exponents improve the approximation



We will not consider them

Taylor's (first-order) series expansion

For functions of many variables, the first-order expansion becomes

$$f(x) = f(\tilde{x}) + (\nabla f(\tilde{x}))^T (x - \tilde{x}) + R_1(\|x - \tilde{x}\|)$$

where

$$\lim_{x \to \tilde{x}} \frac{R_1(\|x - \tilde{x}\|)}{\|x - \tilde{x}\|} = 0$$

and $\nabla f(x)$ is the gradient vector

$$\nabla f(x) = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \dots \\ \frac{\partial f}{\partial x_n} \end{bmatrix}$$

It is the direction of quickest increase for $f(\cdot)$

The main difference is that $\mathbb{R}^n$ offers many ways to move away from $\tilde{x}$

Infinite straight lines and many more curves!

Given a point $\tilde{x} \in \mathbb{R}^n$, an arc in $\tilde{x}$ is a parametric curve $\xi : \mathbb{R}^+ \to \mathbb{R}^n$,
that is $\xi(\alpha) = \begin{bmatrix} \xi_1(\alpha) \\ \dots \\ \xi_n(\alpha) \end{bmatrix}$, such that $\xi(0) = \tilde{x}$ and $\xi_1(\alpha) \in C^1(\mathbb{R}^+)$

An arc $\xi(\alpha)$ is feasible for a given region $X \subseteq \mathbb{R}^n$
when the curve remains in X for small α

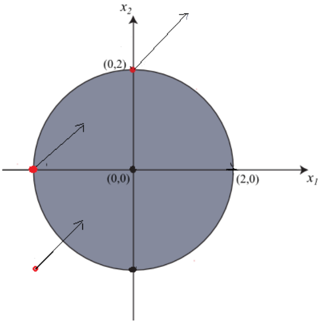
$$\exists \bar{\alpha}_f > 0 : \xi(\alpha) \in X \quad \forall \alpha \in [0; \bar{\alpha}_f)$$

An arc $\xi(\alpha)$ is improving for a given function $f : X \to \mathbb{R}$
when f is strictly better in $\xi(\alpha)$ than in $\tilde{x}$ for all small positive α

$$\exists \bar{\alpha}_i > 0 : f(\xi(\alpha)) < f(\tilde{x}) \quad \forall \alpha \in (0; \bar{\alpha}_i)$$

Example

Let $X = \left\{ x \in \mathbb{R}^2 : x_1^2 + x_2^2 \geq 4 \right\}$ and $f(x) = x_1^2 + x_2^2$

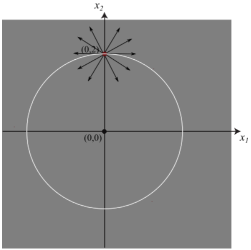


The rectilinear arc $\xi = \tilde{x} + \alpha \, [1 \ 1]^T = [\tilde{x}_1 + \alpha \ \tilde{x}_2 + \alpha]^T$ is

- feasible (with $\alpha \leq 2 - \sqrt{2}$) and improving in $\tilde{x} = (-2, -2)$
- feasible and nonimproving in $\tilde{x} = (0, 2)$
- nonfeasible and improving (with $\alpha \leq 1$) in $\tilde{x} = (-2, 0)$

Why not restricting to lines?

Nonlinear equalities imply that no feasible rectilinear arc exists



Example: $X = \left\{ x \in \mathbb{R}^2 : x_1^2 + x_2^2 = 4 \right\}$

For any constant vector d, the points of line $\xi(\alpha) = \tilde{x} + \alpha d$, are unfeasible

$$(\tilde{x}_1 + \alpha d_1)^2 + (\tilde{x}_2 + \alpha d_2)^2 = 4 \quad \forall \alpha \in [0; \bar{\alpha}_f)$$

implies

$$\cancel{\tilde{x}_1^2 + \tilde{x}_2^2} + \alpha^2 \left(d_1^2 + d_2^2 \right) + 2\alpha \left(d_1 \tilde{x}_1 + d_2 \tilde{x}_2 \right) = \cancel{4} \quad \forall \alpha \in [0; \bar{\alpha}_f)$$

that is

$$\alpha \left(d_1^2 + d_2^2 \right) + 2 \left(d_1 \tilde{x}_1 + d_2 \tilde{x}_2 \right) = 0 \quad \forall \alpha \in [0; \bar{\alpha}_f)$$

which is impossible

Lines are not enough for our purpose

A necessary local optimality condition

Theorem:
If $\tilde{x} \in X \subseteq \mathbb{R}^n$, $f(\cdot) \in C^1(X)$ and $\xi(\alpha)$ is an arc in $\tilde{x}$, feasible for X and improving for $f(\cdot)$, then $\tilde{x}$ is not locally optimal for $f(\cdot)$ in X.

By assumption, for suitable values $\bar{\alpha}_f > 0$ and $\bar{\alpha}_i > 0$, we have

- $\xi(\alpha)$ feasible: $\xi(\alpha) \in X$ for all $\alpha \in [0, \bar{\alpha}_f)$
- $\xi(\alpha)$ improving: $f(\xi(\alpha)) < f(\tilde{x})$ for all $\alpha \in (0, \bar{\alpha}_i)$

Since $\xi(\alpha)$ is a continuous arc

$$\lim_{\alpha \to 0} \xi(\alpha) = \tilde{x} \quad \Leftrightarrow \quad \forall \epsilon > 0, \exists \bar{\alpha}_\epsilon : \|\xi(\alpha) - \tilde{x}\| < \epsilon, \ \forall \alpha \in (0, \bar{\alpha}_\epsilon)$$

that is, $\xi(\alpha) \in \mathcal{U}_{\tilde{x},\epsilon}, \forall \alpha \in (0, \bar{\alpha}_\epsilon)$

Now $\alpha = \frac{1}{2} \min(\bar{\alpha}_f, \bar{\alpha}_i, \bar{\alpha}_\epsilon)$ satisfies all three conditions

- $\alpha < \bar{\alpha}_f \Rightarrow \xi(\alpha) \in X$
- $\alpha < \bar{\alpha}_i \Rightarrow f(\xi(\alpha)) < f(\tilde{x})$
- $\alpha < \bar{\alpha}_\epsilon \Rightarrow \xi(\alpha) \in \mathcal{U}_{x^*,\epsilon}$

but this contradicts local optimality

$$f(x) \geq f(\tilde{x}) \ \text{ for all } x \in \mathcal{U}_{\tilde{x},\epsilon} \cap X$$

A filtering approach

This suggests a possible approach to find candidate points:
remove from X all the points that are provably nonoptimal

$X^{KKT} := X$;
For each $x \in X^{KKT}$ *(continuous set for* x*)*
    For each arc $\xi(\alpha)$ in x feasible for X *(continuous set for* ξ*, interval for* α*)*
        If $\xi(\alpha)$ is improving in x for $f(\cdot)$ *(interval for* α*)*
            then $X^{KKT} := X^{KKT} \setminus \{x\}$
Return X^{KKT}

This is obviously not an algorithm: it loops on continuous sets!

Then, replace the loops with more efficient analitic conditions, that will be all based on first-order approximations

Tangent direction

Given an arc $\xi(\alpha)$ in $\tilde{x}$, its tangent direction is

$$p_\xi = \begin{bmatrix} \xi_1'(0) \\ \dots \\ \xi_n'(0) \end{bmatrix}$$

Straight lines $\xi(\alpha) = \tilde{x} + \alpha d$ have tangent direction d

In fact, arcs generalise directions

Example: The arc in $\tilde{x} = (2, 0)$

$$\xi(\alpha) = \begin{bmatrix} 2\cos\alpha \\ 2\sin\alpha \end{bmatrix}$$

describes the circumference with centre in the origin and radius 2

Its tangent direction is

$$p_\xi = \begin{bmatrix} -2\sin 0 \\ 2\cos 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 2 \end{bmatrix}$$

Theorem

If x^* is locally optimal in X for $f(\cdot)$ and $\xi(\alpha)$ is a feasible arc in x^* for X, then

$$\nabla f(\tilde{x})^T p_\xi \geq 0$$

In a locally optimal point, feasible arcs keep close to the gradient of f (angle $\leq 90°$), so that the objective cannot feasibly improve

Since the arc is feasible, $\xi(\alpha) \in X$ for small α

Since the arc is regular and x^* locally optimal, $f(\xi(\alpha)) \geq f(x^*)$ for small α

Apply Taylor's expansion to $f(\xi(\alpha))$ in $\alpha = 0$

$$\cancel{f(\xi(0))} + \alpha \left.\frac{df}{d\alpha}\right|_{\alpha=0} + R_1(\xi(\alpha) - \xi(0)) \geq \cancel{f(x^*)} \Rightarrow$$

$$\Rightarrow \nabla f(x^*)^T p_\xi + \frac{R_1(\xi(\alpha) - \xi(0))}{\alpha} \geq 0$$

As α converges to 0, the inequality is preserved

$$\lim_{\alpha \to 0} \left(\nabla f(x^*)^T p_\xi + \frac{R_1(\xi(\alpha) - \xi(0))}{\|\xi(\alpha) - \xi(0)\|} \frac{\|\xi(\alpha) - \xi(0)\|}{\alpha} \right) \geq 0 \Rightarrow \nabla f(x^*)^T p_\xi \geq 0$$

Example

$$\begin{aligned} \min f(x) &= x_2 \\ g_1(x) &= x_1^2 + x_2^2 \leq 4 \end{aligned}$$

with $\nabla f^T = [0\ 1]$

- Arc $\xi(\alpha) = \tilde{x} + \alpha [1\ \ -1]^T$ is improving in $\tilde{x} = (-2, 0)$: therefore, $\tilde{x}$ is not locally optimal

$$\nabla f(-2,0)^T p_\xi = [0\ 1] \cdot [1\ \ -1]^T = -1 < 0$$

- arc $\xi(\alpha) = \tilde{x} + \alpha [1\ 1]^T$ is nonimproving in $\tilde{x} = (0, -2)$: $\tilde{x}$ could be locally optimal (it remains candidate until disproval)

$$\nabla f(0,-2)^T p_\xi = [0\ 1] \cdot [1\ 1]^T = 1 \geq 0$$

A filtering approach

For a feasible arc $\xi(\alpha)$

- x^* locally optimal $\Rightarrow \nabla f(\tilde{x})^T p_\xi \geq 0$,
- conversely, $\nabla f(\tilde{x})^T p_\xi < 0 \Rightarrow x^*$ not locally optimal

This yields a sufficient condition to remove points

$X^{KKT} := X$;
For each $x \in X^{KKT}$ (*continuous set for* x)
  For each arc $\xi(\alpha)$ in x feasible for X (*continuous set for* ξ, *interval for* α)
    If $\xi(\alpha)$ is improving in x for $f(\cdot)$ (*interval for* α)
      then $X^{KKT} := X^{KKT} \setminus \{x\}$
Return X^{KKT}

can be simplified (possibly missing some removals) to

$X^{KKT} := X$;
For each $x \in X^{KKT}$ (*continuous set for* x)
  For each arc $\xi(\alpha)$ in x feasible for X (*continuous set for* ξ, *interval for* α)
    If $\nabla f(\tilde{x})^T p_\xi < 0$ ($\xi(\alpha)$ *is improving in* x *for* $f(\cdot)$)
      then $X^{KKT} := X^{KKT} \setminus \{x\}$
Return X^{KKT}

Then, we try and do the same for feasibility

Characterisation of the feasible arcs

Given the analytic description of the feasible region

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

we approximate each function $g_j(\cdot)$ with Taylor's expansion

However, feasibility differs from improvement in two regards

- it involves many inequalities, instead of a single objective
- it requires weak conditions, instead of a strict one

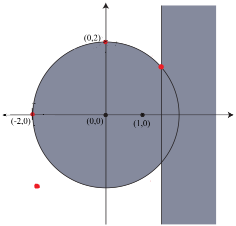
Given point $\tilde{x}$, we partition the constraints into two classes

1. the active constraints ($J_a(\tilde{x})$) are exactly satisfied: $g_j(\tilde{x}) = 0$
2. the nonactive constraints are largely satisfied: $g_j(\tilde{x}) < 0$

Example

$$J_a(x) = \{j \in \{1, \ldots, m\} : g_j(x) = 0\}$$

$$\begin{aligned} \min f(x) &= (x_1 - 1)^2 + x_2^2 \\ g_1(x) &= -x_1^2 - x_2^2 + 4 \leq 0 \\ g_2(x) &= x_1 - 3/2 \leq 0 \end{aligned}$$



The active constraints in various points are:

- for $x = (-2, -2)$, no active constraint: $J_a(-2, -2) = \emptyset$
- for $x = (-2, 0)$, one active constraint: $J_a(-2, 0) = \{1\}$
- for $x = (3/2, \sqrt{7}/2)$, two active constraints: $J_a(3/2, \sqrt{7}/2) = \{1, 2\}$

Characterisation of the feasible arcs

Theorem

If $\xi(\alpha)$ is a feasible arc in $\tilde{x}$ for X, then $\nabla g_j(\tilde{x})^T p_\xi \leq 0$ for all $j \in J_a(\tilde{x})$

Feasible arcs keep far away from the gradients of all active constraints g_j (angle $\geq 90°$), so that such constraints cannot be violated

If $\xi(\alpha)$ is a feasible arc in $\tilde{x}$ for X, there exists $\bar{\alpha}_f > 0$ such that

$$g_j(\xi(\alpha)) \leq 0 \quad \text{for all } \alpha \in [0; \bar{\alpha}_f) \text{ and for } j = 1, \dots, m$$

which implies

$$g_j(\xi(\alpha)) = g_j(\xi(0)) + \left.\frac{dg_j}{d\alpha}\right|_{\alpha=0} \alpha + R_1(\xi(\alpha) - \xi(0)) =$$
$$= g_j(\tilde{x}) + \alpha \nabla g_j(\tilde{x})^T p_\xi + R_1(\xi(\alpha) - \xi(0)) \leq 0$$

For small α, the inequality is guaranteed for all nonactive constraints, because $g_j(\tilde{x}) < 0$ dominates the other terms

Characterisation of the feasible arcs

For the active 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 by α and computing the limit as α converges to 0:

$$\lim_{\alpha \to 0} \left[\nabla g_j(\tilde{x})^T p_\xi + \frac{R_1(\xi(\alpha) - \xi(0))}{\alpha} \right] = \nabla g_j(\tilde{x})^T p_\xi \leq 0$$

□

Special case: equality constraints $h_i(x) = 0$ are always active and can be treated as pairs of active inequalities: $h_i(x) \leq 0$ and $-h_i(x) \leq 0$

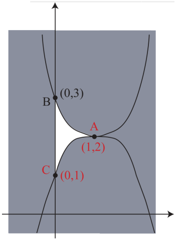
$$\begin{cases} \nabla h_i(\tilde{x})^T p_\xi & \leq 0 \\ -\nabla h_i(\tilde{x})^T p_\xi & \leq 0 \end{cases} \Rightarrow \nabla h_i(\tilde{x})^T p_\xi = 0$$

But the analytic condition is only necessary for feasibility, not sufficient!

Example

For any feasible arc $\xi(\alpha)$, vector p_ξ satisfies the conditions above,
but a vector p that satisfies them is not always tangent to a feasible arc

$$
\begin{aligned}
\min f(x) &= x_2 \\
g_1(x) &= (x_1 - 1)^3 + (x_2 - 2) \leq 0 \\
g_2(x) &= (x_1 - 1)^3 - (x_2 - 2) \leq 0 \\
g_3(x) &= -x_1 \leq 0
\end{aligned}
$$



Since $g_1(A) = g_2(A) = 0$ and $g_3(A) = -1$, $J_a(A) = \{1, 2\}$

$\nabla g_1(x) = \left[3(x_1 - 1)^2 \ \ 1\right]$, $\nabla g_2(x) = \left[3(x_1 - 1)^2 \ \ -1\right]$

Vector $p = [1 \ 0]^T$ satisfies the conditions:

$$
\begin{cases}
\nabla g_1(A)^T p = [0 \ 1] \begin{bmatrix} 1 \\ 0 \end{bmatrix} \leq 0 \\
\nabla g_2(A)^T p = [0 \ -1] \begin{bmatrix} 1 \\ 0 \end{bmatrix} \leq 0
\end{cases}
$$

but all arcs ξ with tangent vector $p_\xi = [1 \ 0]^T$ are unfeasible

Why? The linear approximation!

Regular points

Luckily, the problem concerns only some degenerate points

A point is regular when it satisfies the constraint qualification condition:
the gradients of all active constraints are linearly independent

Theorem
If $\tilde{x}$ is a regular point, then $\nabla g_j(\tilde{x})^T p \leq 0$ for all $j \in J_a(\tilde{x})$ if and only if there exists a arc $\xi(\alpha)$ in $\tilde{x}$ feasible for X with tangent direction $p_\xi = p$

The necessary conditions for feasibility are also sufficient in regular points

Problem: if equalities are turned into pairs of inequalities,
do all points become nonregular?

No, the equality guarantees the existence of a feasible arc lying on it

A filtering approach

Given the previous results

- the analytic conditions can be used to check feasibility in all regular points
- nonregular points must be explicitly tested: they are candidates by default

$X^{KKT} := X;$
For each $x \in X^{KKT}$ (*continuous set for* x)
  For each arc $\xi(\alpha)$ in x feasible for X (*continuous set for* ξ, *interval for* α)
    If $\nabla f(\tilde{x})^T p_\xi < 0$ ($\xi(\alpha)$ *is improving in* x *for* $f(\cdot)$)
      then $X^{KKT} := X^{KKT} \setminus \{x\}$
Return X^{KKT}

can be simplified (possibly missing some removals) to

$X^{KKT} := X \setminus \text{NonRegular}(g, X);$
For each $x \in X^{KKT}$ (*continuous set for* x)
  For each $p \in \mathbb{R}^n : \nabla g_j(x)^T p \leq 0, \forall j \in J_a(x)$ (*arc* $\xi(\alpha)$ *in* x *feasible for* X)
    If $\nabla f(\tilde{x})^T p_\xi < 0$ ($\xi(\alpha)$ *is improving in* x *for* $f(\cdot)$)
      then $X^{KKT} := X^{KKT} \setminus \{x\}$
$X^{KKT} := X^{KKT} \cup \text{NonRegular}(g, X);$
Return X^{KKT}

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 half-plane $C_{\text{impr}}(x)$: the set of improving vectors (scalar products < 0 with the objective gradient)

The first is close, the second open!

If a regular point is locally optimal,
then its feasible cone and improving half-space do not intersect

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

$$\begin{aligned} \min f(x) &= (x_1 - 1)^2 + x_2^2 \\ g_1(x) &= -x_1^2 - x_2^2 + 4 \leq 0 \\ g_2(x) &= x_1 - 3/2 \leq 0 \end{aligned}$$