# Master-MIRI
# Topics on Optimization and Machine Learning (TOML)

**José M. Barceló Ordinas**
**Departament d'Arquitectura de Computadors**
**(UPC)**

- **Basic Terminology**

An **optimization problem (non-linear)** is expressed in its **standard form** as:

$$\begin{array}{ll} \textbf{minimize} & f_0(x) \\ \textbf{subject to} & f_i(x) \leq 0 \quad i=1,\ldots,m \\ & h_i(x) = 0 \quad i=1,\ldots,p \end{array}$$

where,

| | |
|---|---|
| $x \in R^n$ | optimization variable |
| $f_0: R^n \longrightarrow R$ | objective function |
| $f_i: R^n \longrightarrow R$ | inequality constraint functions |
| $h_0: R^n \longrightarrow R$ | equality constraint functions |

If **m=p=0** then the problem is called **unconstrained**

## Basic Terminology

The set of points at which the objective function and all constraint functions are defined is called **domain D**:

$$D = \mathrm{dom}\, f_0 \cap_{i=1\ldots m} \mathrm{dom}\, f_i \cap_{i=1\ldots p} \mathrm{dom}\, h_i$$

a point $x \in D$ is **feasible** if it satisfies the constraints,

$$f_i(x) \leq 0 \quad i=1,\ldots,m$$
$$h_i(x) = 0 \quad i=1,\ldots,p$$

Otherwise is called **unfeasible**

The set of all feasible points is called **feasible set**

The **optimal value p*** is defined as

$$\mathbf{p^* = \inf\{ f_0(x) \mid f_i(x) \leq 0,\ i=1,\ldots,m;\ h_i(x)=0,\ i=1,\ldots,p \}} < \infty$$

The problem is **unfeasible** if $p^* = \infty$

The problem is **unbounded below** if there are feasible points $x_k$ with $f(x_k) \rightarrow -\infty$ with $k \rightarrow \infty$ and then $p^* = -\infty$

## Optimal and locally optimal points

We say $x^*$ is an **optimal point** if **$x^*$ is feasible and $f_0(x^*)=p^*$**

The optimal set is then:

$$X_{opt}=\{x \mid f_i(x)\leq 0,\ i=1,\ldots m;\ h_i(x)=0,\ i=1,\ldots p;\ f_0(x)=p^*\}$$

A feasible point x with $f_0(x) \leq p^*+\varepsilon$ ($\varepsilon>0$) is called a **ε-suboptimal point** and the set of all ε-suboptimal points is called the ε-suboptimal set.

A point x is **locally optimal** if there is an R>0 such that

$$f_0(x) = \inf \{f_0(z) \mid f_i(z)\leq 0,\ i=1,\ldots m;\ h_i(z)=0,\ i=1,\ldots p;\ ||z-x||_2 \leq R\}$$

If x is feasible and $f_i(x)=0$ then the i-th inequality $f_i(x)\leq 0$ is **active**

If x is feasible and $f_i(x)<0$ then the i-th inequality $f_i(x)\leq 0$ is **inactive**

The equality constraints are always active

If m=p=0 then the problem is **unconstraint**

## Feasibility problems

- In order to find whether a point x is feasible we can solve the following optimization problem

$$
\begin{array}{ll}
\textbf{minimize} & \mathbf{0} \\
\textbf{subject to} & \mathbf{f_i(x) \leq 0 \quad i=1,\ldots,m} \\
 & \mathbf{h_i(x) = 0 \quad i=1,\ldots,p}
\end{array}
$$

that it has optimal solution $p^*=0$ if x is a feasible point and has solution $p=\infty$ if there is no any feasible point

# Conversion to the standard form

- Rearrange the inequality by subtracting any non-zero righthand side.
  - For example $g_i(x)=q_i(x) \rightarrow h_i(x)=g_i(x)-q_i(x)=0$
  - For example $g_i(x)>0 \rightarrow -g_i(x)\leq 0$
  - For example:

$$\begin{array}{ll} \textbf{minimize} & f_0(x) \\ \textbf{subject to} & l_i \leq x_i \leq u_i \quad i=1,\ldots,n \end{array}$$

is equivalent to:

$$\begin{array}{ll} \textbf{minimize} & f_0(x) \\ \textbf{subject to} & l_i - x_i \leq 0 \quad i=1,\ldots,n \\ & x_i - u_i \leq 0 \quad i=1,\ldots,n \end{array}$$

  - For example, **maximize $f_0(x)$** is equivalent to **minimize $-f_0(x)$**

# Equivalent problems

Two problems are **equivalent** if from a solution of one problem a solution of the other is readily found.

- **Examples:**
  - Transformation of the objective and constraint functions
  - Change of variables $x=g(z)$ if g is one-to-one (biyective)
  - Slack variables: $f_i(x)\leq 0 \rightarrow f_i(x)+s_i=0$ with $s_i \geq 0$ for $i=1,...,m$
  - Eliminate an equality constraint. Imagine that $h_i(x)=x-g(z)=0$ for each $i=1,...,p$. Then, we change variables $x=g(z)$ and eliminate $h_i(x)$.
  - Introduce equality constraints, e.g. $f_o(A_o x+b) \rightarrow f_o(y_o)$ and $y_o=A_o x+b$

- **Convex Optimization Problems (COP)**

Are those ones that

$$\begin{array}{lll} \textbf{minimize} & f_0(x) & \\ \textbf{subject to} & f_i(x) \leq 0 & i=1,\dots,m \\ & a^T_i x = b_i & i=1,\dots,p \end{array}$$

where $\mathbf{f_0}$, $\mathbf{f_1}$,…, $\mathbf{f_m}$ are convex functions. In other words, with respect a classical optimization problem (slide 2), the requirements are:

i. The objective function $\mathbf{f_0}$ must be convex
ii. The inequality constraint functions $\mathbf{f_i}$ (i=1,...,m) must be convex
iii. The equality constraint functions $\mathbf{h_i = a^T_i x - b_i}$ (i=1,...,p) must be affine
iv. Moreover, **the feasible set** of a convex optimization problem **is convex: t**he set $D = \cap_{i=0 \dots m} \text{dom } f_i$ is convex

If $\mathbf{f_0}$ is quasi-convex, then the problem is quasi-convex

- **Concave Optimization Problems**

Are those ones that

$$\begin{array}{lll} \textbf{maximize} & f_0(x) & \\ \textbf{subject to} & f_i(x) \leq 0 & i=1,\dots,m \\ & a^T_i x = b_i & i=1,\dots,p \end{array}$$

where $\mathbf{f_0}$ is concave and $\mathbf{f_1,\dots, f_m}$ are convex functions.

The problem is solved minimizing $-\mathbf{f_0}$

If $\mathbf{f_0}$ is quasi-concave, then the problem is quasi-concave

# An optimality criterion for differentiable $f_0$

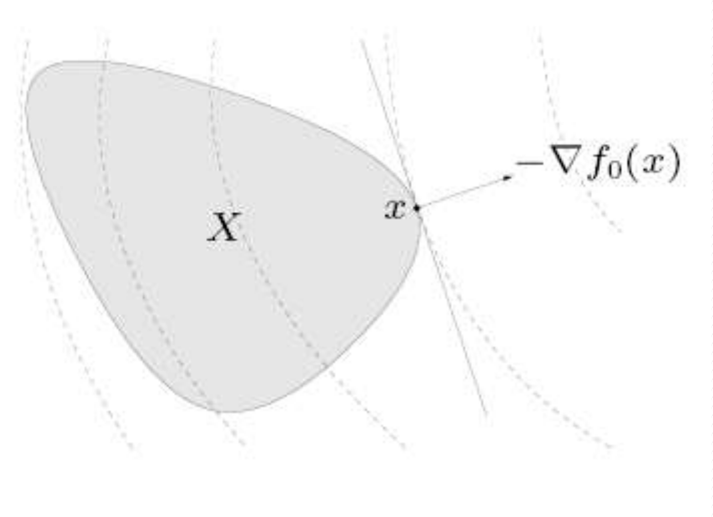
- Let the objective function $\mathbf{f_0}$ be differentiable (in a COP), so that for all $x,y \in \text{dom } f_0$

$$f_0(y) \geq f_0(x) + \nabla f_0(x)^T(y-x)$$

And let X denote the feasible set

$$X=\{x \mid f_i(x) \leq 0,\ i=1,\ldots m;\ h_i(x)=0,\ i=1,\ldots p\}$$

Then, x is **optimal** iif $x \in X$ and $\nabla f_0(x)^T(y-x) \geq 0$.
Geometrically: $\nabla f_0(x) \neq 0$ means that $-\nabla f_0(x)$ defines a supporting hyperplane to the feasible set at x



- **Unconstraint COP**
  - Are those ones in which m=p=0 → **minimize $f_0(x)$**
  - Then, the optimality condition $\nabla f_0(x)^T(y-x) \geq 0$ reduces to

$$\nabla f_0(x)= 0$$

  - To proof this, think in the following: since $f_0(x)$ is differentiable, its domain is open and all y close to x are feasible.

Let this $y=x-t\nabla f_0(x)$ with $t\in R$, for t small and positive:

$$\nabla f_0(x)^T(y-x) = -t\,||\nabla f_0(x)||_2 \geq 0 \rightarrow ||\nabla f_0(x)||_2=0 \rightarrow \nabla f_0(x)= 0$$

  - **Example:** Unconstrained quadratic optimization: $f_0(x)=(1/2)x^TPx+q^Tx+r$ → $\nabla f_0(x)=Px+q=0$. Then,
    - If $q\notin Rank(P)$ → there is no solution → $f_0$ is unbounded below
    - If P>0 then there is a unique solution $x^*=-P^{-1}q$
    - If P is singular and $q\in Rank(P)$ then the set of optimal points are $X_{opt}= -P^{-\dagger}q$ + Null(P) where Null(P) is the Nullspace of P (x such that Px=0), where $P^{-\dagger}=(P^T P)^{-1} P^T$ (pseudo-inverse)

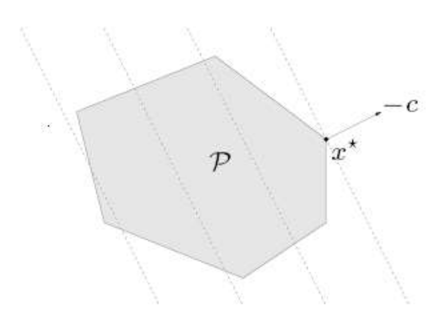
## Some COP: Linear Optimization (LP) problems

- LP: When the objective and constraint are all affine

| | |
|---|---|
| **minimize** | $c^T x+d$ |
| **subject to** | $Gx \leq h$ |
| | $Ax=b$ |

with $G \in R^{m \times n}$ and $A \in R^{p \times n}$, LP is convex in its different forms, e.g. standard form

| | |
|---|---|
| **minimize** | $c^T x$ |
| **subject to** | $Ax=b$ |
| | $x \geq 0$ |




The LP in its convex form can be easily converted in the classical standard form using the variables $x=x^+-x^-$,

| | |
|---|---|
| **minimize** | $c^T x^+ - c^T x^- + d$ |
| **subject to** | $Gx^+ - Gx^- + s = h$ |
| | $Ax^+ - Ax^- = b$ |
| | $s \geq 0,\ x^+ \geq 0,\ x^- \geq 0$ |

## Some COP: Linear Fractional problems

- **LFP**: When the objective are linear fractional functions

$$\begin{array}{ll} \textbf{minimize} & (c^Tx+d)/(e^Tx+f) \\ \textbf{subject to} & Gx \leq h \\ & Ax=b \end{array}$$

with dom f={x|$e^Tx+f>0$} can be transformed to a LP,

$$\begin{array}{ll} \textbf{minimize} & c^Ty+dz \\ \textbf{subject to} & Gy-hz \leq 0 \\ & Ay-bz=0 \\ & e^Ty+fz=1 \\ & z \geq 0 \end{array}$$

with:

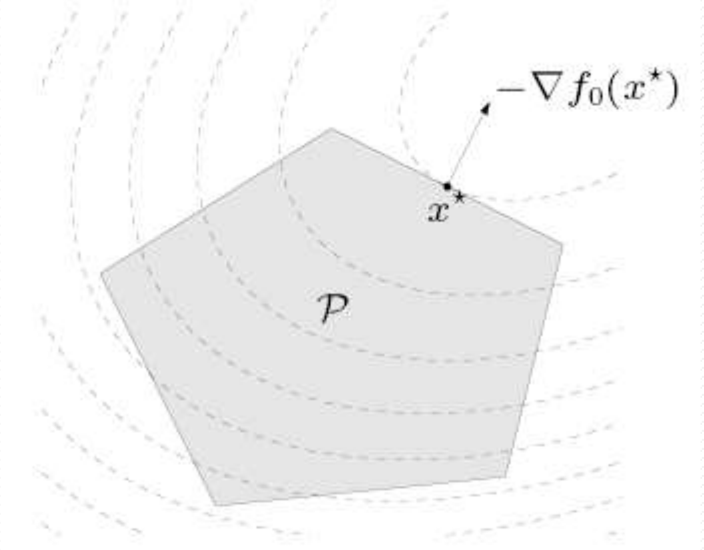
$$y=x/(e^Tx+f)$$
$$z=1/(e^Tx+f)$$

## Some COP: Quadratic (QP) optimization problems

- **QP**: When the objective function is quadratic

**minimize** $(1/2)x^TPx+q^Tx+r$
**subject to** $Gx \leq h$
$Ax=b$

with $P \in S^n_+$, $G \in R^{mxn}$ and $A \in R^{pxn}$,




- **QCQP (quadratically constrained QP)**: When the objective and constraint inequalities are quadratic

**minimize** $(1/2)x^TPx+q^Tx+r$
**subject to** $(1/2)x^TP_ix+q^Tx_i+r \leq 0$ for $i=1,\ldots,m$
$Ax=b$

with $P_i \in S^n_+$ (i=0,...,m), $G \in R^{mxn}$ and $A \in R^{pxn}$

- **Some COP: Quadratic (QP) optimization problems**
  - **Some examples:**
    - **Least squares and regression:** $||Ax-b||^2_2=x^TA^TAx-2b^TAx+b^Tb$ is a unconstrained QP
    - **Distance between polyhedra**: $\text{dist}(P_1,P_2)=\inf\{||x_1-x_2||_2 \mid x_1\in P_1, x_2\in P_2\}$ with $P_i=\{x \mid A_ix\leq b_i\}$ (i=1,2) is equivalent to the QP

      **minimize** $||x_1-x_2||^2_2$

      **subject to** $A_1x\leq b_1,\ A_2x\leq b_2$
    - **Variance problems:** remember that Var $f=E(f^2)-E(f)^2$, then

      **minimize** $E(f^2)-E(f)^2=\Sigma_{i=1..n} f_i^2p_i - (\Sigma_{i=1..n} f_ip_i)^2$

      **subject to** $p\geq 0,\ 1^Tp=1,$

## Some COP: Geometric (GP) optimization problems

- The function f: $R^n \longrightarrow R$ with dom $f=R^n_{++}$ defined as

$$f(x) = d\, x_1^{a1} x_2^{a2} \ldots x_n^{an},$$

with $d>0$ and $a_i \in R$ is called a **monomial function**.

- The sum of monomial functions is called a **posynomial function**

$$f(x) = \Sigma_{i=1..K} d_i\, x_1^{a1i} x_2^{a2i} \ldots x_n^{ani}, \text{ and } d_i>0$$

A **GP** is a problem such that

$$\begin{array}{ll} \textbf{minimize} & f_0(x) \\ \textbf{subject to} & f_i(x) \le 1 \quad i=1,\ldots,m \\ & h_i(x) = 1 \quad i=1,\ldots,p \end{array}$$

where $\mathbf{f_0}, \ldots, \mathbf{f_m}$ are posynomials and $\mathbf{h_1}, \ldots, \mathbf{h_p}$ are mononomials

- **Some COP: Geometric (GP) optimization problems**
  - **Example:**

Let us assume the following problem:

$$\begin{array}{ll} \textbf{maximize} & x/y \\ \textbf{subject to} & 2 \leq x \leq 3 \\ & x^2+3y/z \leq (y)^{1/2} \\ & x/y=z^2 \end{array}$$

Then, it can be transformed to:

$$\begin{array}{ll} \textbf{minimize} & x^{-1}y \\ \textbf{subject to} & 2x^{-1} \leq 1 \\ & (1/3)x \leq 1 \\ & x^2y^{-1/2}+3y^{1/2}z^{-1} \leq 1 \\ & xy^{-1}z^{-2}=1 \end{array}$$

- **Some COP: Geometric (GP) optimization problems**

  - A **GP is not a COP**, but it can easily be transformed to a COP

Let us remember that $y_i=\log(x_i)$ and then $x_i=\exp(y_i)$, then taking into account that $x_i^a=\exp(a \log x_i)=\exp(ay_i)$

$$\mathbf{f(x) = d\, x_1^{a1} x_2^{a2} \ldots x_n^{an} =}$$
$$\mathbf{= \exp(b) \exp(a_1 y_1) \ldots \exp(a_n y_n) =}$$
$$\mathbf{= \exp(a^T y+b)}$$

with $b=\log(d)$. In case of having a posynomial:

$$\mathbf{f(x) = \Sigma_{k=1..K}\, d_k\, x_1^{a1k} x_2^{a2k} \ldots x_n^{ank} =}$$
$$\mathbf{= \Sigma_{k=1..K} \exp(a_k^T y+b_k),}$$

with $y^T=[y_1,\ldots, y_n]$, $y_i=\log(x_i)$, $a^T_k=[a_{1k},\ldots, a_{nk}]$ and $b_k=\log(d_k)$

- **Some COP: Geometric (GP) optimization problems**
  - Thus the GP can be expressed as:

| | | |
|---|---|---|
| **minimize** | $\Sigma_{k=1..K0} \exp(a_{ok}^T y+b_{0k})$, | |
| **subject to** | $\Sigma_{k=1..Ki} \exp(a_{jk}^T y+b_{jk}) \leq 1$ | **j=1,...,m** |
| | $\exp(g_j^T y+h_j) = 1$ | **j=1,...,p** |

with $y_i=\log(x_i)$ for i=1,...,n

  - Now, we take logarithms:

| | | |
|---|---|---|
| **minimize** | $f_0(y)= \log(\Sigma_{k=1..K0} \exp(a_{ok}^T y+b_{0k}))$ | |
| **subject to** | $f_i(y)=\log(\Sigma_{k=1..Ki} \exp(a_{jk}^T y+b_{jk})) \leq 0$ | **j=1,...,m** |
| | $h_i(y)=g_i^T y+h_i = 0$ | **i=1,...,p** |

That is called the **GP in convex form**

The log-sum-exp function **log ($\Sigma_{i=1..n}$ exp($x_k$))** is convex in x

## Some COP: Geometric (GP) optimization problems

- For example

| | | |
|---|---|---|
| minimize | $x_1^{-1}x_2$ | $a_{01}=[-1,1,0]$ |
| subject to | $2x_1^{-1} \leq 1$ | $a_{11}=[-1,0,0]$ |
| | $(1/3)x_1 \leq 1$ | $a_{21}=[1,0,0]$ |
| | $x_1^2x_2^{-1/2}+3x_2^{1/2}x_3^{-1}\leq 1$ | $a_{31}=[2,-½,0]$, $a_{32}=[0, ½,-1]$ |
| | $x_1x_2^{-1}x_3^{-2}=1$ | $a_{41}=[1,-1,-2]$, |

would be expressed in its convex form as:

| | |
|---|---|
| minimize | $\log(\exp(-y_1+y_2)) = -y_1+y_2$ |
| subject to | $-y_1+ \log 2 \leq 0$ |
| | $y_1 - \log 3 \leq 0$ |
| | $\log( \exp(2y_1-½y_2)+\exp(½y_2-y_3+\log 3) ) \leq 0$ |
| | $y_1 - y_2 - 2y_3 = 0$ |

- **Some general applications:**
  - Regression, Least-squares estimation, residuals, ....
  - Maximum-Likelihood, Bayesian estimation, ....
  - Estimation and detection: hypothesis testing
  - Experiment design
  - Geometric Problems: euclidean distance problems, minimum distances to a point
  - Classification (pattern recognition and classification problems)