# C. Convex Optimization Problems

## 1. Basic notation and terminology.

An **optimization problem** is a problem of the form

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & x \in C \end{array} \qquad (1)$$

- $f : \mathbb{R}^n \to \mathbb{R}$ is called the **objective function**
- $x$ is called the **optimization variable**
- $C \subseteq \mathbb{R}^n$ is called the **constraint set** (or **feasible set**)

Typically the constraint set $C$ is given to us in functional form

$$C = \{ x \in \mathbb{R}^n : g_i(x) \le 0,\ i = 1, \dots, m,\quad h_j(x) = 0,\ j = 1, \dots, p \}$$

for some functions $g_i, h_j : \mathbb{R}^n \to \mathbb{R}$.

Equivalenty we write

$$\begin{array}{lll} \text{minimize} & f(x) & \\ \text{subject to} & g_i(x) \le 0, & i = 1, \dots, m \\ & h_j(x) = 0, & j = 1, \dots, p \end{array} \qquad (1)$$

- $g_i : \mathbb{R}^n \to \mathbb{R}$ are the **inequality constraint** functions
- $h_j : \mathbb{R}^n \to \mathbb{R}$ are the **equality constraint** functions.

**Definition (Local and global minima)** Consider the optimization problem (1). A point $\bar{x} \in \mathbb{R}^n$

- **feasible** if $\bar{x} \in C$, equivalently $g_i(\bar{x}) \le 0\ \forall i = 1, \dots, m$ and $h_j(\bar{x}) = 0\ \forall j = 1, \dots, p$.

  If $f, g_i, h_j$ are **extended-valued** functions (that is if $f, g_i, h_j : \mathbb{R}^n \to (-\infty, +\infty]$) a point $\bar{x}$ is called **feasible** if

  $$\bar{x} \in \underbrace{\operatorname{dom} f \cap \bigcap_{i=1}^{m} \operatorname{dom}(g_i) \cap \bigcap_{j=1}^{p} \operatorname{dom}(h_j)}_{\text{implicit constraints}} \cap \underset{\text{explicit constraint}}{C}$$

W.l.o.g. we will study **real-valued** functions

an optimal solution **may not exist** or **may not be unique**

- **global minimizer** or **(globally) optimal solution** if $\bar{x} \in C$ and $f(\bar{x}) \le f(x)\ \forall x \in C$. We denote $\bar{x} \in \underset{x \in C}{\operatorname{argmin}} f(x)$.

- **local minimizer** if $\bar{x} \in C$ and there exists $\delta > 0$ such that $f(\bar{x}) \le f(x)$, for all $x \in C$ with $\|x - \bar{x}\| \le \delta$.

  any norm would result to the same definition

- the **optimal value** $p^*$ of problem (1) is $p^* \triangleq \inf_{x \in C} f(x)$

  - $p^* = \inf_{x \in C} f(x) = +\infty \iff$ problem is **infeasible** $\iff C = \emptyset$

    there is no feasible $x$ (satisfying the constraints)

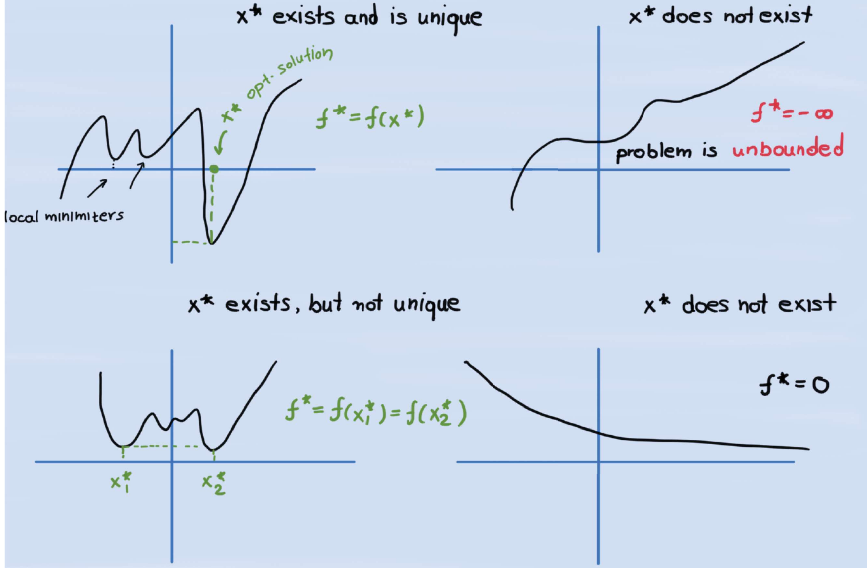
  - $p^* = \inf_{x \in C} f(x) = -\infty \iff$ problem is **unbounded below**

    $\iff \exists\ (x_n)_{n=1}^{\infty} \subset C$ s.t. $\lim_{n \to +\infty} f(x_n) = -\infty$

    sequence of feasible points

- $\exists$ optimal solution $x^*$ to (1) $\iff p^* = \min_{x \in C} f(x) = f(x^*)$

**Example:**




**Fact (Weierstrass Theorem)** If $f: \mathbb{R}^n \to \mathbb{R}$ is **continuous** and $C \subseteq \mathbb{R}^n$ is **compact**, then an optimal solution $x^*$ to $\min\{f(x) : x \in C\}$ is guaranteed to exist

**Pathologies of nonconvex problems:** The optimality conditions associated to general nonconvex optimization problems suffer from:

1. it is possible to be inconclusive about testing local optimality
2. they say absolutely nothing about global optimality

## 2. Convex Optimization Problems

**Why focusing on convex optimization problems?**

1. broadest class of problems we know how to solve efficiently.
2. enjoy nice geometric properties (e.g., local minima are global)
3. excellent software available (eg., CVX)
4. numerous important problems are convex

### Standard form convex optimization problem

$$\begin{aligned} &\text{minimize} && f(x) \\ &\text{subject to} && g_i(x) \le 0, \quad i = 1, \dots, m \\ & && Ax = b \end{aligned}$$

where $A = \begin{bmatrix} a_1^T \\ \vdots \\ a_m^T \end{bmatrix} \in \mathbb{R}^{m \times n}$ $\quad b = \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix} \in \mathbb{R}^m$

- the **objective** $f: \mathbb{R}^n \to \mathbb{R}$ is **convex**
- the **inequality constraint functions** $g_i: \mathbb{R}^n \to \mathbb{R}$ are **convex**
  therefore the sub-level sets $\{x : g_i(x) \le 0\}$ are convex
- the **equality constraint map** $\mathbb{R}^n \ni x \to Ax - b \in \mathbb{R}^m$ **affine**
  therefore the set $\{x \in \mathbb{R}^n : Ax = b\}$ is affine and thus convex
- the **feasible set** $C \triangleq \{x \in \mathbb{R}^n : g_i(x) \le 0, i = 1, \dots, m, \; Ax = b\}$

**Facts:** For a **standard form convex optimization program**

1. The **feasible set** is convex
2. The **set of optimal solutions** is convex
3. **local minimizers are global minimizers**

Note that although for **standard form** convex optimization programs the feasible set is convex, the converse is not true

**Example**

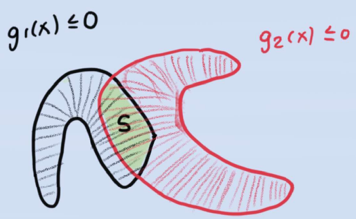
$C = \{x \in \mathbb{R} : x^3 \leq 0\}$ is convex but minimizing a convex function over $C$ is not a convex optimization program per our definition. ( $g(x) = x^3$ is not convex function over $\mathbb{R}^n$ )

However, the same set can be represented as $C = \{x \in \mathbb{R} : x \leq 0\}$ and then this would result to a convex optimization problem in the standard form

**We require this stronger definition :**

1. to **avoid** **complicated optimization problems** that can be formulated as optimization problems over a convex set

For example




The feasible set $S = \{x : g_1(x) \leq 0,\ g_2(x) \leq 0\}$ is convex however $g_1$ and $g_2$ are not even quasi-convex.

2. standard form convex optimization problems can be solved efficiently by convex optimization solvers (algorithms) eg cvx.

## 3. Existence, Uniqueness and Optimality Conditions

**Definition (strict and strong convexity)** A function $f : \mathbb{R}^n \to \mathbb{R}$ is

- **strictly convex** if for all $x \neq y$ and all $\theta \in (0,1)$
$$f(\theta x + (1-\theta)y) < \theta f(x) + (1-\theta) f(y)$$
- **strongly convex** if $\exists\ \sigma > 0$ : such that $x \to f(x) - \frac{\sigma}{2}\|x\|^2$ is convex

actually, we say that $f$ is **$\sigma$-strongly convex w.r.t $\|\cdot\|$**.

**Fact 1** $f : \mathbb{R}^n \to \mathbb{R}$ is **$\sigma$-strongly convex** if and only if for all $x, y \in \mathbb{R}^n$ and $\theta \in [0,1]$

$$f(\theta x + (1-\theta)y) \leq \theta f(x) + (1-\theta) f(y) - \frac{\sigma}{2}\theta(1-\theta)\|x-y\|^2$$

**Fact 2** strong convexity $\Rightarrow$ strict convexity $\Rightarrow$ convexity

But the converse of neither application is true.

**Fact 3 (strict convexity and 1st-order characterization)** A differentiable function $f : \mathbb{R}^n \to \mathbb{R}$ is **strictly convex** if and only if

$$f(y) > f(x) + \nabla f(x)^T (y-x) \quad \forall\ x, y \in \mathbb{R}^n \quad x \neq y$$

**[second order sufficient condition]** If $f : \mathbb{R}^n \to \mathbb{R}$ is twice continuously differentiable, $\nabla^2 f(x) \succ 0\ \forall x \Rightarrow f$ strictly convex

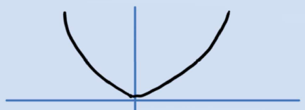
The converse is not true.

**FACT 4 (Characterization of differentiable strongly convex functions)**

differentiable
$f: \mathbb{R}^n \to \mathbb{R}$ strongly convex $\iff f(y) \geq f(x) + \nabla^T f(x)(y-x) + \frac{\sigma}{2}\|x-y\|^2 \quad \forall x, y$

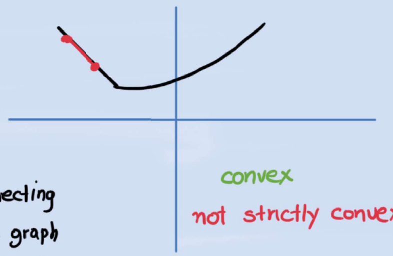
$f: \mathbb{R}^n \to \mathbb{R}$ strongly convex $\iff \nabla^2 f(x) \succeq mI \quad \forall x \in \mathbb{R}^n$
twice continuously differentiable

## Examples:

- 

  strictly convex

  the line segment connecting any two points of the graph of f lies **strictly** above the graph

  convex

  not strictly convex

- Any norm $x \to \|x\|$ is convex but not strictly convex

  indeed for $x \in \mathbb{R}^n$, $y = 2x$, $\theta = 1/2$
  $\| 1/2\, x + \frac{1}{2} y \| = \frac{3}{2} \|x\| = \frac{1}{2}\|x\| + \frac{1}{2}\|y\|$

- The square of 2-norm $x \to \|x\|_2^2$ is strictly convex

  indeed the Hessian of $f$ equals $\nabla^2 f(x) = 2I \succ 0$.

- The sum of a convex function and a strictly convex function is strictly convex (easy by using definitions)

---

- $f$ strictly convex $\not\Rightarrow \nabla^2 f(x) \succ 0$
  $f$ strictly convex $\not\Rightarrow$ strongly convex

counterexample: $f(x) = x^4$ is strictly convex but $f''(0) = 0$.
So $f''$ is not strictly positive. Since this is necessary for strong convexity, it is not strongly convex

indeed $f(x) = \varphi(g(x))$ where $\varphi(t) = t^2$, $t \geq 0$ (strictly increasing) and convex and $x \to g(x) = x^2$ strictly convex.

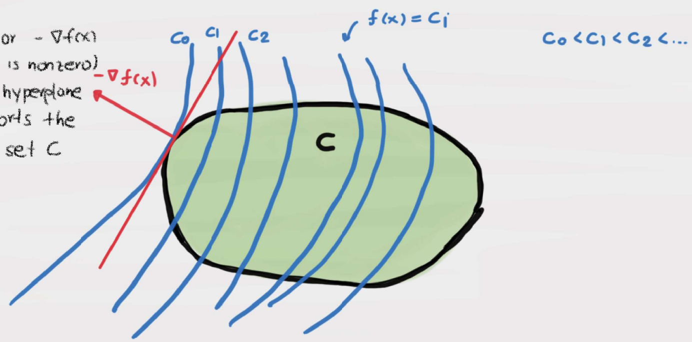
**Fact (Uniqueness of solutions)**

- strictly convex functions have at most one minimizer
- strongly convex functions have exactly one minimizer

**Fact (First order optimality conditions)** Consider an optimization problem: $\min\{f(x) : x \in C\}$ where $f: \mathbb{R}^n \to \mathbb{R}$ is convex and differentiable and $C \subseteq \mathbb{R}^n$ is convex. Then a point $x$ is a (global) minimizer if and only if $x \in C$ and

$$\nabla f(x)^T (y - x) \geq 0 \quad \forall y \in C$$

the vector $-\nabla f(x)$ (assuming is nonzero) defines a hyperplane that supports the feasible set $C$ at $x$.



If $C = \mathbb{R}^n$ (unconstrained case), the necessary and sufficient condition above reduces to $\nabla f(x) = 0$ (why?)

**Example (quadratic functions)** Let $f(x) = x^T A x + b^T x + c$ where $A$ is symmetric, then $f$ is convex $\Leftrightarrow A \succeq 0$.
Consider now the unconstrained optimization problem $\min_x f(x)$.
We have:

- If $A \not\succeq 0$ ($f$ not convex) $\Rightarrow$ problem is unbounded below
  proof: let $\bar{x}$ be eigenvector with a negative eigenvalue $\lambda$. Then, $A\bar{x} = \lambda \bar{x} \Rightarrow \bar{x}^T A \bar{x} = \lambda \|\bar{x}\|_2^2 < 0$.
  Moreover, $f(\alpha x) = \alpha^2 \bar{x}^T A \bar{x} + \alpha b^T \bar{x} + c$.
  So $f(\alpha \bar{x}) \to -\infty$ when $\alpha \to +\infty$

  it is also strongly convex.
- If $A \succ 0$ then $f$ is strictly convex. There is a unique optimal solution $x^* = -\frac{1}{2} A^{-1} b$ ($\nabla f(x^*) = 0$).

  The first order condition says that $x^*$ is optimal $\Leftrightarrow 2Ax^* + b = 0$
- If $A \succeq 0$ ($f$ convex). If $A$ is positive semidefinite but not positive definite, then the problem can be unbounded (e.g. take $A=0$, $b \neq 0$) or it can be bounded below and have infinitely many solutions (e.g. $f(x) = (x_1 - x_2)^2$.) One can distinguish between the two cases by testing whether $b$ lies in the range of $A$

**Example (projection onto a convex set)** Consider the projection onto a convex set

$a \in \mathbb{R}^n$ is fixed vector

$$\begin{aligned} &\text{minimize } \tfrac{1}{2}\|a - x\|_2^2 \\ &\text{subject to } x \in C \end{aligned}$$

According to the first order optimality condition,

$$\nabla f(x)^T (y - x) = (a - x)^T (y - x) \geq 0 \quad \forall x \in C.$$

normal cone

Thus $x$ is an optimal solution $\Leftrightarrow a - x \in N_C(x)$

**Example (equality-constrained minimization)** Consider the equality-constrained convex optimization problem

$$\begin{aligned} &\text{minimize } f(x) \\ &\text{subject to } Ax = b \end{aligned}$$

convex and differentiable

where $f: \mathbb{R}^n \to \mathbb{R}$ is a convex function and $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$.
A point $x \in \mathbb{R}^n$ is an optimal solution if and only if $Ax = b$ and $\exists \mu \in \mathbb{R}^m$ s.t. $\nabla f(x) = A^T \mu$.

proof: $x$ is optimal solution $\Leftrightarrow Ax = b$ and $\nabla f(x)^T \underbrace{(y - x)}_{= v} \geq 0 \ \forall y: Ay = b$.

$(Av = Ay - Ax = 0)$

Equivalently a feasible $x$ is optimal $\Leftrightarrow \nabla f(x)^T v \geq 0 \quad \forall v \in \mathrm{NULL}(A)$

vector space

Since $\mathrm{NULL}(A)$ is a vector space we have that $v \in \mathrm{NULL}(A) \Leftrightarrow -v \in \mathrm{NULL}(A)$. So for $v \in \mathrm{NULL}(A)$
$\nabla f(x)^T v \geq 0 \ \forall v \in \mathrm{NULL}(A) \Leftrightarrow \nabla f(x)^T v \leq 0 \ \forall v \in \mathrm{NULL}(A)$

All in all a feasible $x$ is optimal $\Leftrightarrow \nabla f(x)^T v = 0 \ \forall v \in \mathrm{NULL}(A)$
$\Leftrightarrow \nabla f(x)^T \in \mathrm{NULL}(A)^\perp \Leftrightarrow$
$\Leftrightarrow \nabla f(x)^T \in \mathrm{RANGE}(A^T) \Leftrightarrow$

$\mathrm{NULL}(A)^\perp = \mathrm{RANGE}(A^T)$

$\Leftrightarrow \exists \mu \in \mathbb{R}^m : \nabla f(x)^T = A^T \mu$.

**Example (Least squares)** Consider the problem

$$\text{minimize } \|Ax-b\|_2^2$$

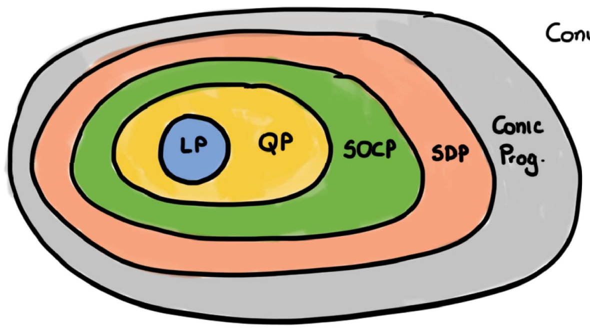
where $A \in \mathbb{R}^{m\times n}$ has **full column rank**

$$\nabla f(x) = 2A^TAx - 2A^Tb = 0 \Rightarrow A^TAx = A^Tb \Rightarrow$$
$$\Rightarrow x = (A^TA)^{-1}A^Tb$$

Note that the matrix $A^TA$ is indeed invertible.
because its nullspace is just the origin.
Indeed $A^TAx = 0 \Rightarrow x^TA^TAx = 0 \Rightarrow \|Ax\|^2 = 0 \Rightarrow Ax = 0 \Rightarrow x = 0$,
where the last implication follows by the linear independency of the columns of $A$.

Since $\nabla^2 f(x) = 2A^TA \succ 0$ (as $x^TA^TAx = \|Ax\|^2 \geq 0$ and $= 0 \Leftrightarrow x = 0$)
we conclude that $f$ is strictly (and strongly) convex.
Thus $x^* = (A^TA)^{-1}A^Tb$ is the unique global minimizer.

## Convex Optimization Problems




the formulations in slides are the dual ones (cf next lect.)

## Linear Conic Optimization Problem (Standard form)

$$\begin{aligned} \underset{x}{\text{minimize}} \quad & \langle x, c\rangle \\ \text{subject to} \quad & \mathcal{A}x = b \\ & x \succeq_K 0 \end{aligned}$$

where:

- $X, Y$ are two **finite-dimensional vector spaces**
- $K \subseteq X$ is a (closed) **convex cone** in $X$
- $\mathcal{A}: X \to Y$ is a **linear operator**
- $\langle\cdot,\cdot\rangle$ is an **inner product** in $X$
- $c \in X$, $b \in Y$

## Semidefinite Programming (Standard form)

$$\begin{aligned} \underset{X}{\text{minimize}} \quad & \text{trace}(X^TC) \\ \text{subject to} \quad & \text{trace}(A_i^TX) = b_i, \; i = 1,\dots,m \\ & X \succeq 0 \end{aligned}$$

the dual (next lecture) is described by LMIs (linear matrix inequalities)

- $X = S^n$ vector space of symmetric $n\times n$ matrices
- $Y = \mathbb{R}^m$
- $K = S^n_+$ convex cone of symmetric positive semi-definite $n\times n$ matrices
- $\langle A, B\rangle = \text{trace}(A^TB) = \sum_{i=1}^n \sum_{j=1}^n A_{ij}B_{ij}$, for $A, B \in S^n$
- $\mathcal{A}: S^n \to \mathbb{R}^m$ is a linear operator defined by some fixed $A_1, \dots, A_m \in S^n$ by:

$$\mathcal{A}(X) = \begin{pmatrix} \langle A_1, X \rangle \\ \vdots \\ \langle A_m, X \rangle \end{pmatrix}$$

- $C \in S^n$, $b \in \mathbb{R}^m$

## Second Order Cone Programming (Standard form)

$$\begin{aligned} \underset{x}{\text{minimize}} \quad & c^T x \\ \text{subject to} \quad & Ax = b \\ & x_1 \geq \sqrt{x_2^2 + \cdots + x_n^2} \end{aligned}$$

- $X = \mathbb{R}^n$, $Y = \mathbb{R}^m$
- $K = \{ (x_1, \dots, x_n) : \left( \sum_{i=2}^{n} x_i^2 \right)^{1/2} \leq x_1 \}$ is the second order cone in $\mathbb{R}^n$
- $\langle x, y \rangle = x^T y$
- $\mathcal{A} : \mathbb{R}^n \to \mathbb{R}^m$ is given by $\mathcal{A}x = Ax$ where $A \in \mathbb{R}^{m \times n}$
- $c \in \mathbb{R}^n$, $b \in \mathbb{R}^m$

$$= \left\{ x \in \mathbb{R}^n : \begin{pmatrix} x_1 & x_2 & \cdots & x_n \\ x_2 & x_1 & \ddots & 0 \\ \vdots & \vdots & & \vdots \\ x_n & 0 & \cdots & x_1 \end{pmatrix} \succeq 0 \right\}$$

## Linear Programming (standard form)

$$\begin{aligned} \underset{x}{\text{minimize}} \quad & c^T x \\ \text{subject to} \quad & Ax = b \\ & x \geq 0 \end{aligned}$$

- $X = \mathbb{R}^n$, $Y = \mathbb{R}^m$
- $K = \mathbb{R}^n_+$ nonnegative orthant
- $\langle x, y \rangle = x^T y$
- $\mathcal{A} : \mathbb{R}^n \to \mathbb{R}^m$, $\mathcal{A}x = Ax$, $A \in \mathbb{R}^{m \times n}$.