# Lecture 4: Convex Optimization

## I. Introduction

- **Initial Remarks: Motivation and Overview**

**Why do we study optimization**

Optimization is about making **good** (may not be optimal) decisions or choices in a rigorous way, often subject to constraints. Applications appear everywhere in science and business, including:

| | |
|---|---|
| Managing a share portfolio | Designing electronic circuit layouts |
| Scheduling public transport | Choosing worker shift patterns |
| Fitting a model to measured data | Shaping aerodynamic components |
| Optimizing a supply chain | Recovering images from raw MRI data |

For this course, it is used to do: Linear Control design and Trajectory design for dynamic systems

**Describing an Optimization Problem**

The general format of the problem:

$$\min_{x} \ f(x)$$

$$\text{subj. to. } x \in \mathcal{X} \subseteq \operatorname{dom}(f)$$

The problem has several ingredients:

The vector $x$ collects the **decision variables**

The set $\operatorname{dom}(f)$ is the **domain** of the decision variables

The set $\mathcal{X} \subseteq \operatorname{dom}(f)$ is the **constraint set**, and describes the **feasible** decisions.

The **objective function** $f\colon \operatorname{dom}(f) \to \mathbb{R}$ assigns a cost $f(x)$ to each decision $x$

The problem can be more compactly written as:

$$\min_{x \in \mathcal{X} \subseteq \operatorname{dom}(f)} f(x)$$

We call this **nonlinear mathematical program** or **nonlinear program (NLP)**.

A more common problem format:

$$\min_{x \in \operatorname{dom}(f)} f(x)$$

$$\begin{aligned} \text{subj. to. } & g_i(x) \leq 0 \quad i = 1, \dots, m \\ & h_i(x) = 0 \quad i = 1, \dots, p \end{aligned}$$

Defined by the following problem data:

The **objective function** $f\colon \operatorname{dom}(f) \to \mathbb{R}$

**Domain** $\operatorname{dom}(f)$ of the objective function, from which the decision variable $x = (x_1; x_2 \dots; x_n)$ must be chosen.

Optional **inequality constraint functions** $g_i\colon \mathbb{R}^n \to \mathbb{R}$ for $i = 1, \dots, m$

Optional **equality constraint functions** $h_i: \mathbb{R}^n \to \mathbb{R}$ for $i = 1, \ldots, p$

Solving the optimization problem means computing the **least possible cost (optimal value)** $f^*$ and an associated **optimal solution** (or called **minimizer**) $x^*$:

$$f(x^*) = f^*$$

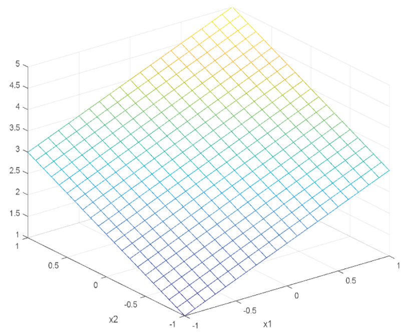
Note:

1. Minimization is used. Any maximization problem can be written this way by a change of sign.
2. In unconstrained optimization $\operatorname{dom}(f) = \mathbb{R}$

**Examples of Unconstrained Optimization**

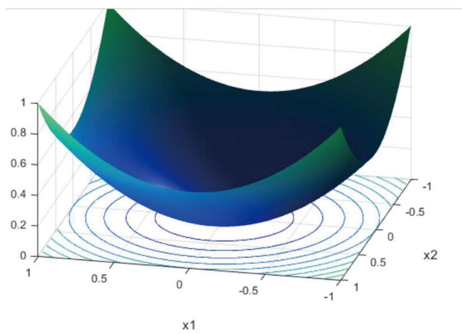
Min of a linear function

$$c^{\mathrm{T}} x + k$$



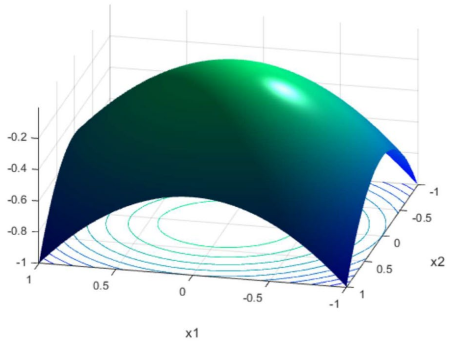

Min of a positive definite quadratic function

$$x^{\mathrm{T}} H x + c^{\mathrm{T}} x + k$$



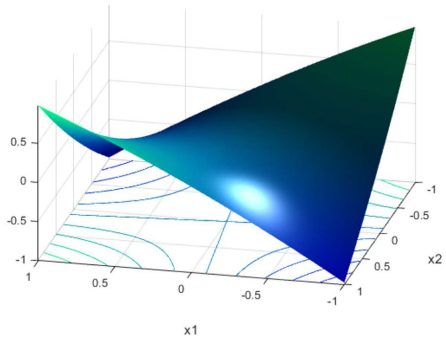

Min of a positive negative quadratic function

$$x^{\mathrm{T}} H x + c^{\mathrm{T}} x + k$$

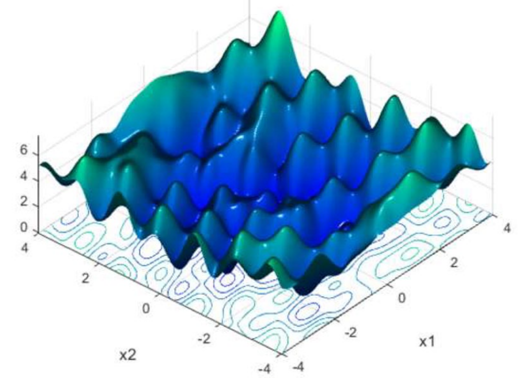



Min of a positive definite quadratic function

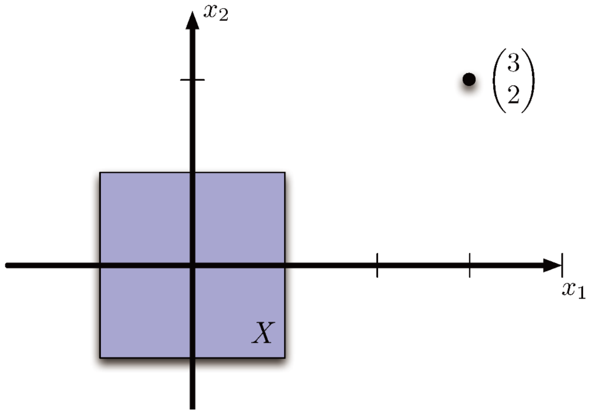
$$x^{\mathrm{T}} H x + c^{\mathrm{T}} x + k$$




Min of **Wolfram** function $f(x) = \left\| \begin{matrix} x_1 - \sin(2x_1 + 3x_2) - \cos(3x_1 - 5x_2) \\ x_2 - \sin(x_1 - 2x_2) + \cos(x_1 + 3x_2) \end{matrix} \right\|$



**Examples of Constrained Optimization**




Problem : In $\mathbb{R}^2$, find the point in the unit box $\mathcal{X}$ closest to the point $(x_1;x_2)=(3;2)$

In standard format, can be written as:

$$\min_{(x_1,x_2)\in\mathbb{R}^2} \ (x_1-3)^2+(x_2-2)^2$$

$$\begin{aligned} \text{subj. to.} \quad & x_1 \le 1 \\ & -x_1 \le 1 \\ & x_2 \le 1 \\ & -x_2 \le 1 \end{aligned}$$

**Properties of Optimization Problems**

Consider the **Nonlinear Program** (NLP):

$$J^* = \min_{x\in\mathcal{X}} f(x)$$

**Notations:**

If $J^* = -\infty$, then the problem is **unbounded below**.

If the set $\mathcal{X}$ is empty, then the problem is **infeasible** and we set $J^* = +\infty$.

If $\mathcal{X}=\mathbb{R}^n$, the problem is **unconstrained**,

There might be more than one solution. The set of solutions is:

$$\operatorname*{argmin}_{x\in\mathcal{X}} f(x) = \{x\in\mathcal{X} \mid f(x)=J^*\}$$

**Terminology:**

**Feasible point:** $x\in\operatorname{dom}(f)$ satisfying the inequality and equality constraints. i.e. $g_i(x)\le 0$ for $i=1,\dots,m$, $h_i(x)=0$ for $i=1,\dots,p$.

**Strictly feasible point:** Feasible $x\in\operatorname{dom}(f)$ satisfying the inequality constraints strictly, i.e. $g_i(x)<0$ for $i=1,\dots,m$.

**Optimal value:** The lowest possible objective value, $f(x^*)$, denoted by $f^*$ (or $J^*$ or $p^*$), also called the **infimum** sometimes (slightly different from the minimum, see the discussions below).

**Local optimality:** $x$ is locally optimal if there exists an $R>0$ such that $z=x$ is optimal for:

$$\min_{z\in\operatorname{dom}(f)} f(z)$$

$$\begin{aligned} \text{subj. to.} \quad & g_i(z)\le 0 \quad i=1,\dots,m \\ & h_i(z)=0 \quad i=1,\dots,p \\ & \|z-x\|_2 \le R \end{aligned}$$

**Optimal solution:** Any **feasible** $x^*\in\operatorname{dom}(f)$ such that $f(x^*)\le f(x)$ **for all the feasible** $x\in\operatorname{dom}(f)$, also called the **global optimal solution**

**Local optimum:** a point $x^*_{\text{local}}$ that is optimal within a neighborhood $\|x-x^*_{\text{local}}\|\le R$

**Discussion: What might go wrong?**

It is possible that **no minimizer will exist**

1. If the constraints are inconsistent, then the problem is **infeasible**.

$$\min_{x \in \mathbb{R}} x^2$$
$$\text{subj. to. } x \leq -1$$
$$x \geq 1$$

2. It might be possible to make $f(x)$ arbitrarily negative without violating any of the constraints. Then the problem is referred to as **unbounded.**

$$\min_{x \in \mathbb{R}} x$$
$$\text{subj. to. } x \leq 0$$

3. The value $J^*$ might be finite, but there is no $x$ that achieves it.

$$\min_{x \in \mathbb{R}} e^{-x}$$
$$\text{subj. to. } x \geq 0$$

The optimal value $J^* = 0$, exists, but there are no optimal solutions and $J^*$ is called **infimum.**

Note: Always remember that there might be more than one minimizer, or none at all.

**Active, Inactive and Redundant Constraints**

Consider the standard problem:

$$\min_{x \in \mathrm{dom}(f)} f(x)$$
$$\text{subj. to. } g_i(x) \leq 0 \quad i = 1, \ldots, m$$
$$h_i(x) = 0 \quad i = 1, \ldots, p$$

The $i$-th inequality constraint $g_i(x)$ is **active** at $\bar{x}$ if $g_i(\bar{x}) = 0$. Otherwise, it is **inactive**. Equality constraints are always active.

A **redundant** constraint does not change the feasible set. This implies that removing a redundant constraint does not change the solution.

Redundant constraint example:

$$\min_{x \in \mathbb{R}} f(x)$$
$$\text{subj. to. } x \leq 1$$
$$x \leq 2$$

Note: In practice, it is hard to judge the redundant constraints, but if it is identified, it could be helpful since it could be eliminated iteratively and thus help reduce the scale of the problem.

**Implicit and Explicit Constraints**

The equality and inequality constraints $g_i(x), i = 1, \ldots, m$ and $h_i(x)$, $i = 1, \ldots, p$ are referred to as the **explicit constraints** of the optimization problem. However, the **domains** of the objective function $f$ and constraint functions also define an **implicit constraint** on $x$:

$$x \in \operatorname{dom}(f) \bigcap_{i=1}^{m} \operatorname{dom}(g_i) \bigcap_{i=1}^{p} \operatorname{dom}(h_i)$$

If a problem has $m=0$ and $p=0$, it is referred to as an **unconstrained problem**, although the limited domain of the objective function may still represent an implicit constraint. For example:

$$\min_x \; f(x) = \sum_{i=1}^{k} \log(a_i^{\mathrm{T}} x - b_i)$$

is unconstrained but still has the implicit constraint that $a_i^{\mathrm{T}} x > b_i$ for $i=1,\ldots,k$. In other words, the constraint set $x \in \mathcal{X} = \{x \in \mathbb{R} \mid a_i^{\mathrm{T}} x > b_i, i=1\ldots,k\}$ is implied by the domain of $f$.

Note: We have two definitions for unconstraint problems. Narrowly, it means $x \in \operatorname{dom}(f) = \mathcal{X} = \mathbb{R}^n$, Generally, it means $x \in \operatorname{dom}(f)$ without explicit constraints, i.e. only the implicit constraints.

**Feasibility Constraints**

The "constraint satisfiability" problem

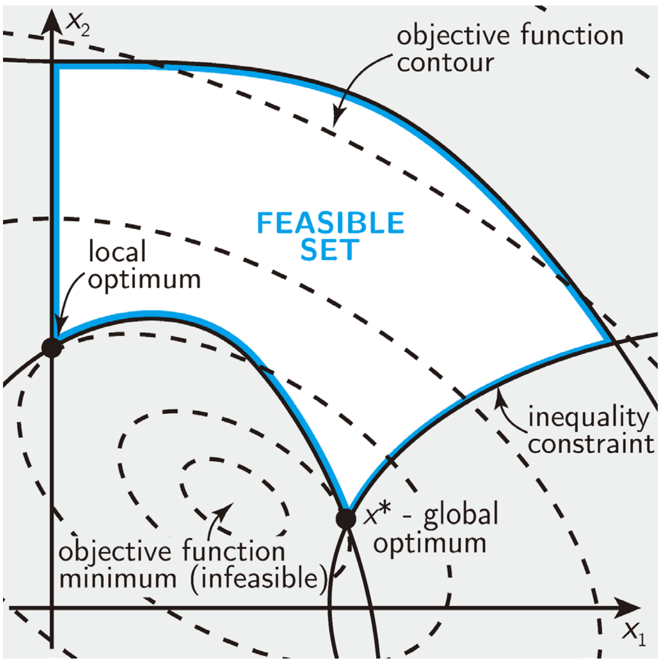
$$\begin{aligned} &\underset{x \in \operatorname{dom}(f)}{\text{find}} \quad x \\ &\text{subj. to. } g_i(x) \le 0 \quad i=1,\ldots,m \\ &\qquad\qquad h_i(x) = 0 \quad i=1,\ldots,p \end{aligned}$$

is a special case of the general optimization problem in which the objective is not dependent on $x$:

$$\begin{aligned} &\underset{x \in \operatorname{dom}(f)}{\text{find}} \quad 0 \\ &\text{subj. to. } g_i(x) \le 0 \quad i=1,\ldots,m \\ &\qquad\qquad h_i(x) = 0 \quad i=1,\ldots,p \end{aligned}$$

$J^* = 0$ if the constraints are feasible (consistent). Every feasible $x$ is optimal. $J^* = \infty$ otherwise.

**Geometry of an Optimization Problem**



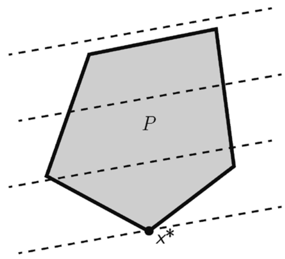
- **Common Types of Optimization Problems**

**Easier Problems**

**Linear Program (LP):**

Linear cost and constraint functions; feasible set is a polyhedron.

$$\min_x \; c^{\mathrm{T}} x$$
$$\text{subj. to. } Gx \le h$$
$$Ax = b$$

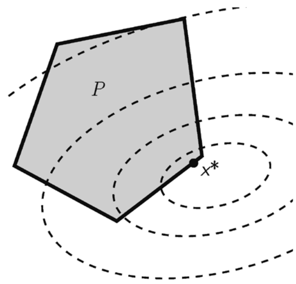



Linear optimization on a polytope.

**Convex Quadratic Program (QP):**

Quadratic cost and linear constraint functions; feasible set is a polyhedron. Convex if $P \succeq 0$

$$\min_x \; \frac{1}{2} x^{\mathrm{T}} P x + q^{\mathrm{T}} x$$
$$\text{subj. to. } Gx \le h$$
$$Ax = b$$



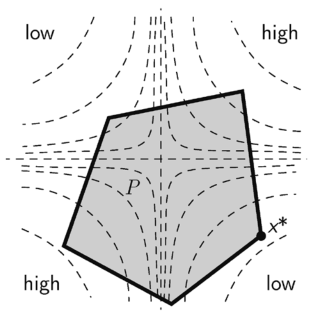

Convex quadratic optimization on a polytope.

**Harder problems**

**Nonconvex Quadratic Program (QP):**

Quadratic cost and linear constraint functions; feasible set is a polyhedron. $P \not\succeq 0$ (indefinite)

$$\min_x \; \frac{1}{2} x^{\mathrm{T}} P x + q^{\mathrm{T}} x$$
$$\text{subj. to. } Gx \le h$$
$$Ax = b$$

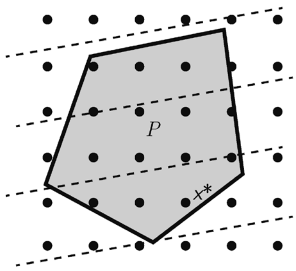



Nonconvex quadratic optimization on a polytope. Contours represent a saddle-shaped objective function.

**Mixed Integer Linear Program (MILP):**

Linear program with Linear cost plus binary or integer constraints.

$$\min_x \; c^{\mathrm{T}} x$$
$$\text{subj. to. } Gx \le h$$
$$Ax = b$$
$$x \in \{0,1\}^n \text{ or } x \in \mathbb{Z}^n$$




Linear optimization with integer constraints (dots).

**Software Tools for Optimization**

Huge variety of software tools for solving LPs and QPs (and other standard types):

Examples: MATLAB (linprog/quadprog), CPLEX, Gurobi, GLPK, XPRESS, qpOASES, OOQP, FORCES, SDPT3, Sedumi, MOSEK, IPOPT

General purposes modeling tools allow easy switching between solvers:

Examples: CVX, Yalmip, GAMS, AMPL

## II. Convex Sets

- **Definitions and Examples**

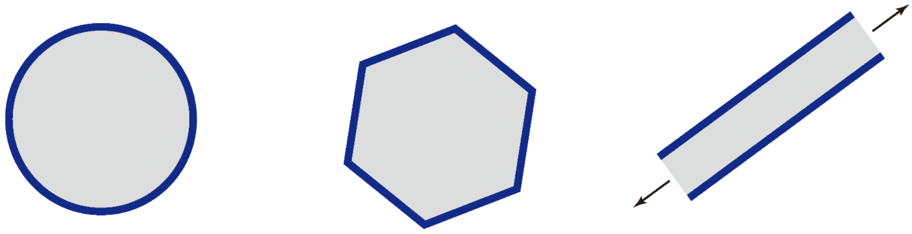
**Definition (Convex Set):**

A set $\mathcal{X}$ is convex **if and only if** for any pair of points $x$ and $y$ in $\mathcal{X}$, any **convex combination** $\mathrm{co}(x,y)$ lies in $\mathcal{X}$, i.e.:

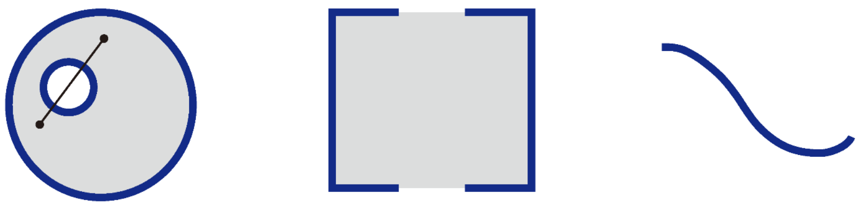
$$\mathcal{X} \text{ is convex} \quad \Leftrightarrow \quad \mathrm{co}(x,y) = \lambda x + (1-\lambda) y \in \mathcal{X},\ \forall \lambda \in [0,1],\ \forall x,y \in \mathcal{X}$$

Interpretation: All line segments starting and ending in $\mathcal{X}$ stay within $\mathcal{X}$

Convex set examples:



Non-Convex set examples:

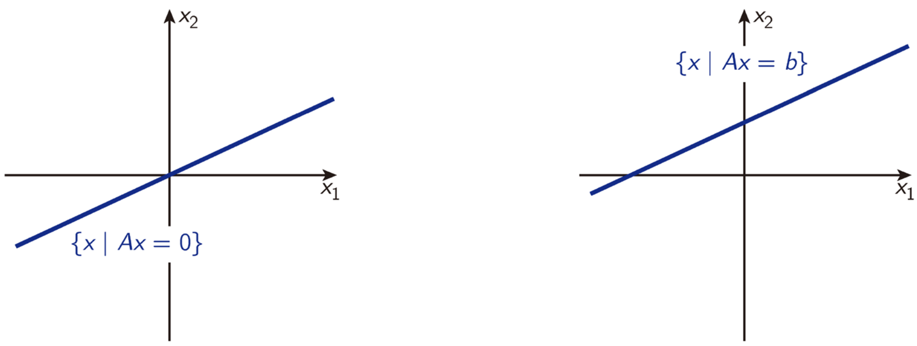


**Definition (Affine sets and Subspaces):**

An affine set is a **convex set** defined by $\mathcal{X} = \{x \in \mathbb{R}^n \mid Ax = b\}$.

A subspace is an affine set with $b = 0$

This definition encompasses lines, planes and individual points, for example:




A 1D subspace in $\mathbb{R}^2$ An affine space in $\mathbb{R}^2$

Note: Verify the convexity of the affine set by definition. For all $x, y \in \mathcal{X}$, for all $\lambda \in [0,1]$ we have:

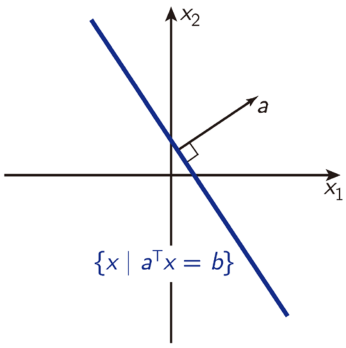
$$A(\lambda x + (1-\lambda) y) = \lambda A x + (1-\lambda) A y = \lambda b + (1 - \lambda b) = b$$

**Definition (Hyperplanes and halfspaces):**

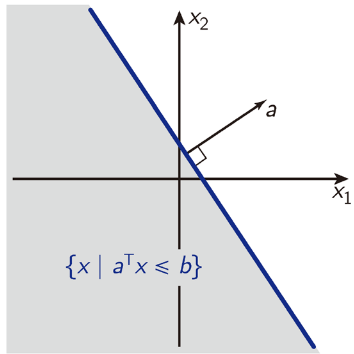
A hyperplane is defined by $\{x \in \mathbb{R}^n | \ a^{\mathrm{T}}x = b\}$ for $a \neq 0$, where $a \in \mathbb{R}^n$ is the normal vector to the hyperplane.

A halfspace is everything on one side of a hyperplane $\{x \in \mathbb{R}^n | \ a^{\mathrm{T}}x \leq b\}$ for $a \neq 0$. It can either be **open (strict inequality)** or **closed (non-strict inequality)**.

Hyperplanes and halfspaces **are always convex**. Examples:




A hyperplane




A closed halfspace

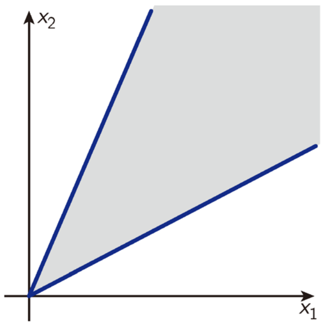
For $n = 2$ hyperplanes define lines. For $n = 3$, hyperplanes define planes. Compared to affine sets, which could define a line or a plane in $\mathbb{R}^3$.

Note: Actually, this means affine sets are "broader" than hyperplanes. Note that $a \in \mathbb{R}^n$ in $a^{\mathrm{T}}x = b$ (hyperplane) can be regarded **as one row of the equations** represented by $Ax = b$ (affine sets)

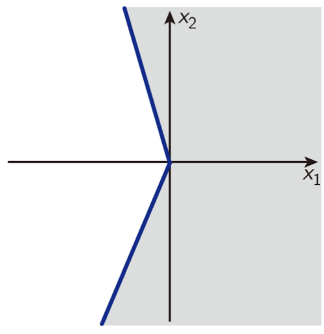
**Definition (Cone):**

A set $\mathcal{X}$ is a cone if for all $x \in \mathcal{X}$, and for all $\theta > 0$, $\theta x \in \mathcal{X}$. If the cone contains $x = 0$, then it is **pointed.** A conic combination, $\mathrm{cone}(x_1, x_2)$, is any point that can be expressed as $\theta_1 x_1 + \theta_2 x_2$, for some $\theta_1, \theta_2 \geq 0$.

A cone is **not necessarily convex.** Examples:




A convex cone




A non-convex cone

**Definition (Polyhedra and polytopes):**

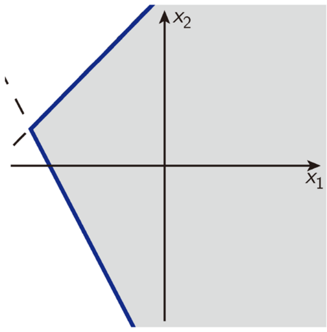
A polyhedron is the intersection of a **finite number of closed halfspaces**:

$$\begin{aligned}\mathcal{X} &= \{x | \ a_1^{\mathrm{T}}x \leq b_1, \ a_2^{\mathrm{T}}x \leq b_2, \dots, \ a_m^{\mathrm{T}}x \leq b_m\} \\ &= \{x | \ Ax \leq b\}\end{aligned}$$

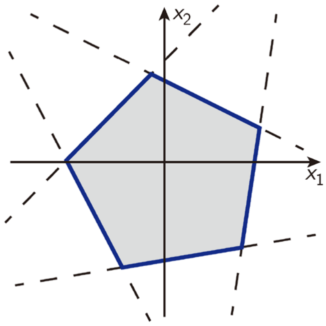
Where $A = [a_1 \ \ a_2 \ \ \cdots \ \ a_m]^{\mathrm{T}}$ and $b = [b_1 \ \ b_2 \ \ \cdots \ \ b_m]^{\mathrm{T}}$

A polytope is a **bounded polyhedron.**

Polyhedra and polytopes **are always convex.** Examples:




An (unbounded) polyhedron




A polytope

**Definition (Vector norm):**

A norm is any function $f: \mathbb{R}^n \to \mathbb{R}$ satisfying the following conditions:

$$f(x) \geq 0 \text{ and } f(x) = 0 \Rightarrow x = 0$$

$$f(tx) = |t| f(x) \text{ for scalar } t$$

$$f(x+y) \leq f(x) + f(y) \text{ for all } x, y, \in \mathbb{R}$$

A norm is denoted $\|x\|_{(\cdot)}$, where the subscript denotes the type of norm. The notation $\|x\|$ refers to any arbitrary norm.

**Definition ($\ell_p$ norms):**

The $\ell_p$ norm on $\mathbb{R}^n$ is denoted $\|x\|_p$ and is defined for any $p \geq 1$ by:

$$\|x\|_p = \left[\sum_{i=1}^{n} |x_i|^p\right]^{\frac{1}{p}}$$

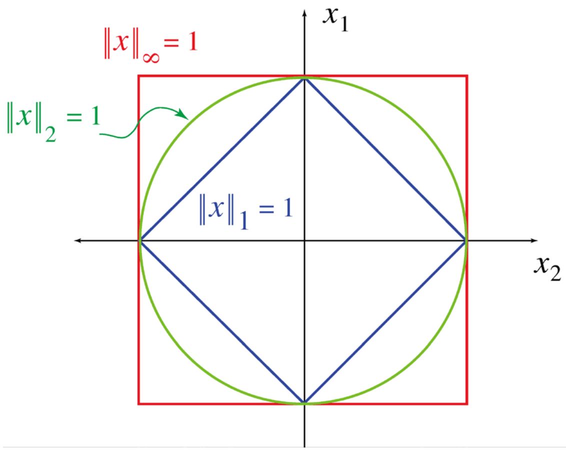
By far the most common $\ell_p$ norms are:

$p = 2$ (Euclidean norm): $\|x\|_2 = \sqrt{\sum_i x_i^2}$

$p = 1$ (Sum of absolute values): $\|x\|_1 = \sum_i |x_i|$

$p = \infty$ (Largest absolute value): $\|x\|_1 = \max_i |x_i|$

The norm ball, defined by $\{x \mid \|x - x_c\| \leq r\}$ where $x_c$ is the center of the ball and $r \geq 0$ is the radius, **is always convex** for any norm.




Note: in control practice, we use norms to represent errors. We use 2-norm most, sometimes with 1-norm, infinity norm is rarely used since 2-norm or 1-norm can mostly be converted into QP or LP.

**Definition (Ellipsoid):**

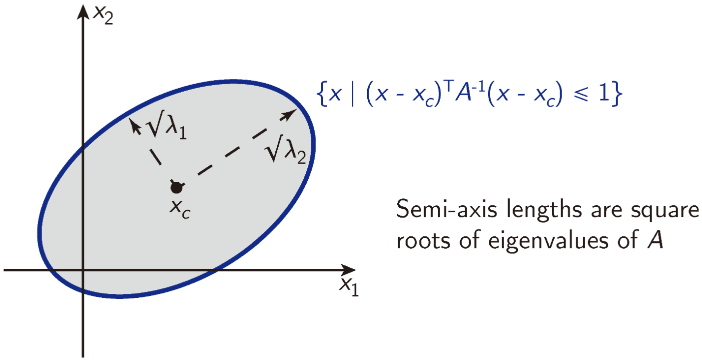
An ellipsoid is a set defined as:

$$\{x \mid (x - x_c)^{\mathrm{T}} A^{-1} (x - x_c) \leq 1\}$$

where $x_c$ is the center of the ellipsoid, and $A \succ 0$ (i.e. $A$ is positive definite).

The Euclidean ball $B(x_c, r)$ is a special case of the ellipsoid, for which $A = r^2 I$ so that $B(x_c, r) = \{x \mid \|x - x_c\|_2 \leq r\}$

Note: another common expression is to use $A$ instead of $A^{-1}$. In that case, the semi-axis lengths shown on the right would be $\lambda_1^{-1/2}$ and $\lambda_2^{-1/2}$, respectively




- **Set Operations**

**Intersection** $\mathcal{X} \cap \mathcal{Y}$

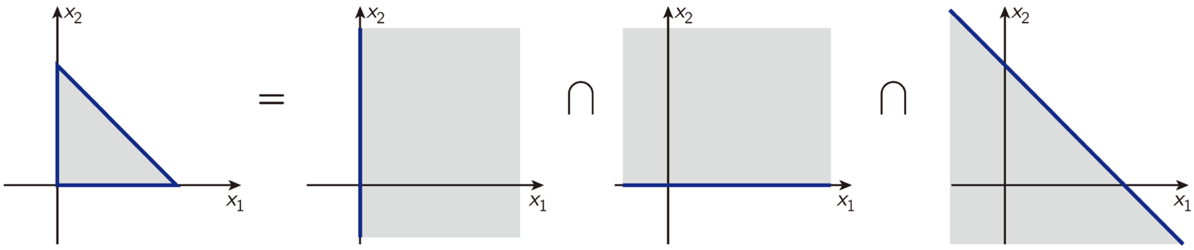
**Theorem (Intersection of convex sets):**

The intersection of two or more convex sets is itself convex.

**Proof (for two sets):**

Consider any two points $a$ and $b$ which **both lie in both of two convex sets** $\mathcal{X}$ and $\mathcal{Y}$. For any $\lambda \in [0, 1]$, $\lambda a + (1-\lambda) b$ is in both $\mathcal{X}$ and $\mathcal{Y}$ (by definition of convex sets). Therefore, we have $\lambda a + (1-\lambda) b \in \mathcal{X} \cap \mathcal{Y}, \forall \lambda \in [0, 1]$. This satisfies the definition of convexity for set $\mathcal{X} \cap \mathcal{Y}$.

Many sets can be written as the intersection of convex elements, and are therefore easily shown to be convex. **Any convex set can be written as a (possibly infinite) intersection of halfspaces.** Examples:




**Convex Hull** $\operatorname{conv}(\mathcal{X})$

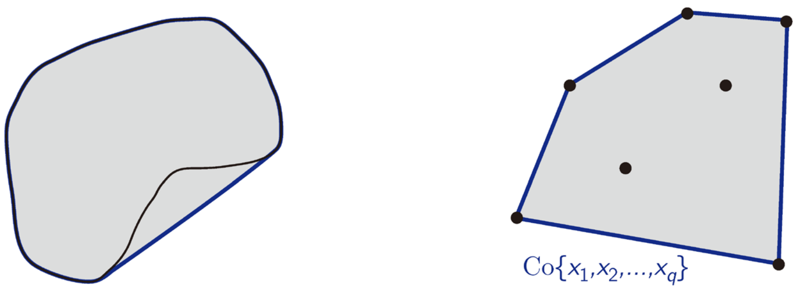
**Definition (Convex Hull):**

The convex hull of a set $\mathcal{X}$ is the set of all convex combinations of points in $\mathcal{X}$:

$$\operatorname{conv}(\mathcal{X}) = \{x \mid x = \lambda a + (1-\lambda) b,\ \lambda \in [0, 1],\ a, b \in \mathcal{X}\}$$

It is the smallest convex set that contains $\mathcal{X}$, i.e. for all convex sets $\mathcal{Y} \supseteq \mathcal{X}$, $\operatorname{conv}(\mathcal{X}) \subseteq \mathcal{Y}$

Examples (black is the set $\mathcal{X}$, blue is the corresponding convex hull):



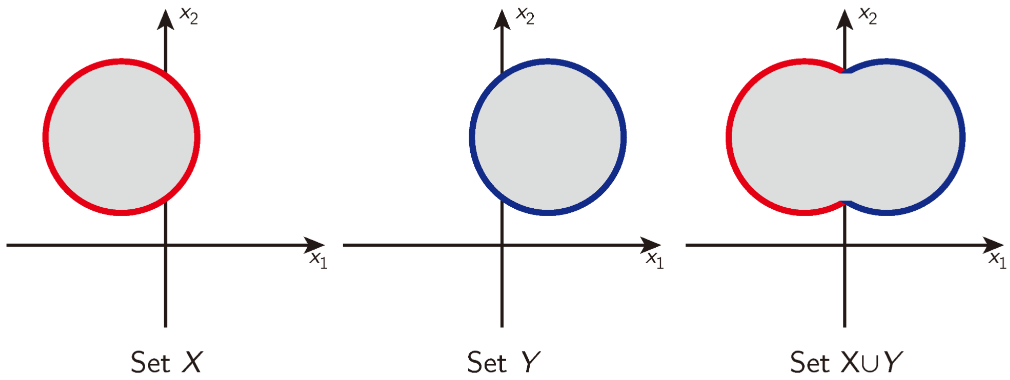
For a set $\mathcal{X} = \{x_1, x_2, \ldots, x_q\}$ with $q$ points, the convex hull can be written:

$$\text{conv}(\mathcal{X}) = \left\{ \lambda_1 x_1 + \lambda_2 x_2 + \ldots + \lambda_q x_q \;\middle|\; \lambda_i \geq 0,\ i = 1, 2, \ldots, q,\ \sum_{i=1}^{q} \lambda_i = 1 \right\}$$

Note: recall that previously we define polytopes through the halfspaces (i.e. through its edges). Using the convex hull, we can define the polytopes through its vertices.

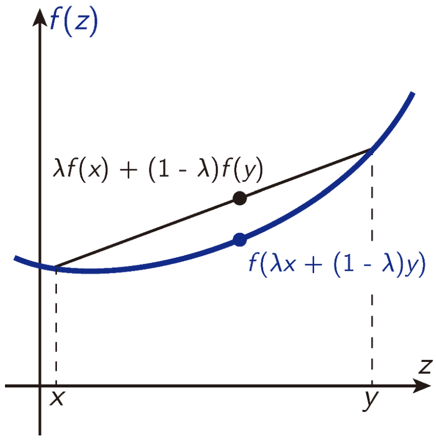
**Union** $\mathcal{X} \cup \mathcal{Y}$

Note that the **union of two sets is not convex in general**, regardless of whether the original sets were convex! Examples:




## III. Convex Functions

- **Definitions and Theorems**




**Definition (Convex Function):**

A function $f$: $\text{dom}(f) \to \mathbb{R}$ is convex if and only if its **domain** $\text{dom}(f)$ **is convex** and

$$f(\lambda x + (1-\lambda) y) \leq \lambda f(x) + (1-\lambda) f(y)$$

$\forall \lambda \in (0,1)$, $\forall x, y \in \text{dom}(f)$ **is satisfied**. The function $f$ is strictly convex if the above inequality is strict.

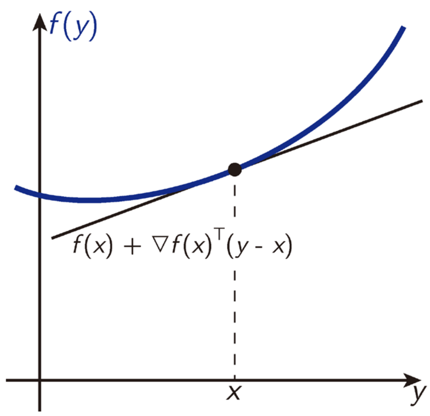
$f$ is concave if $-f$ is convex.

**Theorem (1st-order condition for convexity):**

A differentiable function $f$: $\text{dom}(f) \to \mathbb{R}$ with a convex domain is convex if and only if:

$$f(y) \geq f(x) + \nabla f(x)^{\mathrm{T}} (y - x) \qquad \forall x, y \in \text{dom}(f)$$

i.e. a first-order approximator of $f$ around any point $x$ is a **global underestimator** of $f$.




Note: recall that the gradient of $f$ is given by:

$$\nabla f(x) = \left[ \frac{\partial f(x)}{\partial x_1} \quad \frac{\partial f(x)}{\partial x_2} \quad \cdots \quad \frac{\partial f(x)}{\partial x_n} \right]^{\mathrm{T}}$$

**Theorem (2nd-order condition for convexity):**

A twice-differentiable function $f\colon \mathrm{dom}(f) \to \mathbb{R}$ is convex if and only if its domain is convex and:

$$\nabla^2 f(x) \succeq 0 \quad \forall x \in \mathrm{dom}(f)$$

where the Hessian $\nabla^2 f(x)$ is defined by:

$$\nabla^2 f(x)_{ij} = \frac{\partial^2 f(x)}{\partial x_i \partial x_j}$$

If $\mathrm{dom}(f)$ is convex and $\nabla^2 f(x) \succ 0 \quad \forall x \in \mathrm{dom}(f)$, then $f$ is strictly convex.

**Definition (Epigraph of a Function):**

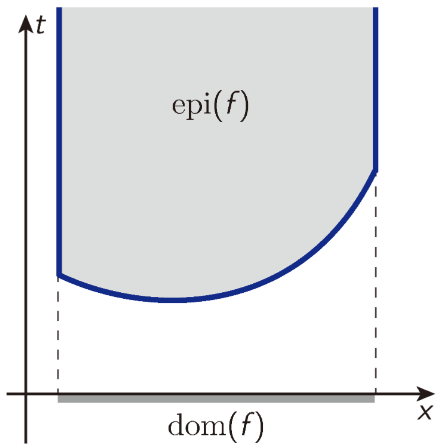
The epigraph of a function $f\colon \mathrm{dom}(f) \to \mathbb{R}$ is the set:

$$\mathrm{epi}(f) = \left\{ \begin{bmatrix} x \\ t \end{bmatrix} x \in \mathrm{dom}(f) \ f(x) \le t \right\} \subseteq \mathrm{dom}(f) \times \mathbb{R}$$

It is one dimension higher than the domain of $f$ (because we stack $x$ and $t$ together).

**A function is convex if and only if its epigraph is a convex set.**

Note: using epigraph, we connect the definition of convex sets and convex function together.




**Definition (Level and Sublevel Sets):**

The level set $L_\alpha$ of a function $f$ for value $\alpha$ is the set of all $x \in \mathrm{dom}(f)$ for which $f(x) = \alpha$:

$$L_\alpha = \{x \mid x \in \mathrm{dom}(f),\ f(x) = \alpha\}$$

For $f\colon \mathbb{R}^2 \to \mathbb{R}$ these are **contour lines** of constant "height"

The sublevel set $C_\alpha$ of a function $f$ for value $\alpha$ is defined by:

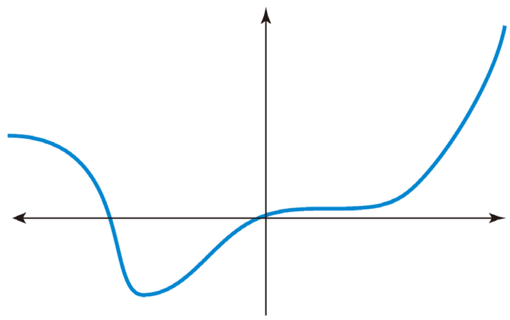
$$C_\alpha = \{x \mid x \in \mathrm{dom}(f),\ f(x) \le \alpha\}$$

Function $f$ is convex $\Rightarrow$ sublevel sets of $f$ are convex for all $\alpha$, but not $\Leftarrow$ (only necessity)!

**Definition (Quasi-convex Functions):**

Function $f$ is quasi-convex if and only if $\mathrm{dom}(f)$ is convex and all sublevel sets of $f$ are convex.

Note: figure on the right is a good example for why the above condition is only a necessary condition and why we need to define quasi-convex function.



A Quasiconvex function.

**Theorem (Restriction to a Line):**

A function $f\colon \mathbb{R}^n \to \mathbb{R}$ is convex **if and only if** its evaluation along any line in its domain is convex. In other words, parameterizing the "distance along the line" by $t$:

$$g(t) = f(x + tv), \quad \mathrm{dom}(g) = \{t \mid x + tv \in \mathrm{dom}(f)\}$$

Function $f$ is convex if and only $g$ is convex in $t$ for all $x \in \mathrm{dom}(f)$, for all $v \in \mathbb{R}^n$. This means **that convexity of $f$ can be tested by checking the functions of one variable.**

**Example: testing convexity of the function of matrices**

$f: \mathbb{S}^n \to \mathbb{R}$ with $f(X) = \log(\det(X))$, and $\operatorname{dom}(f) = \mathbb{S}^n_{++}$. This function is concave. To test this, define $g(t) = f(X + tV)$, and check the concavity of $g(t)$ for arbitrary $X$ and $V$.

Note: Here, the set $\mathbb{S}^n$ denotes the symmetric $n \times n$ matrices, $\mathbb{S}^n_{+}$ denotes the symmetric positive semidefinite matrices, and $\mathbb{S}^n_{++}$ denotes the symmetric positive definite matrices.

**Definition (Extended-value Extension):**

A function $f$ that is not defined everywhere can be extended to include points outside its domain by defining the **extended-value** function $\tilde{f}(x)$:

$$\tilde{f}(x) = f(x) \text{ for } x \in \operatorname{dom}(f), \ \tilde{f}(x) = +\infty \text{ for } x \notin \operatorname{dom}(f)$$

This often simplifies notation and does not change the epigraph of the function, i.e. $\operatorname{epi}\left(\tilde{f}\right) = \operatorname{epi}(f)$. Also, $\tilde{f}$ is convex if and only if $\operatorname{dom}(f)$ is convex and $f$ is convex.

- **Examples and Common Used Results**

**Examples of Convex Functions $\mathbb{R} \to \mathbb{R}$:**

The following functions are convex (on domain $\mathbb{R}$ unless otherwise stated):

Affine: $ax + b$ for any $a, b \in \mathbb{R}$

Exponential: $e^{ax}$ for any $a \in \mathbb{R}$

Powers: $x^\alpha$ on domain $\mathbb{R}_{++}$ for $\alpha \geq 1$ or $\alpha \leq 0$

Powers of absolute value: $|x|^p$ for $p \geq 1$

The following functions are concave (on domain $\mathbb{R}$ unless otherwise stated):

Affine: $ax + b$ for any $a, b \in \mathbb{R}$

Logarithm: $\log x$ on domain $\mathbb{R}_{++}$

Powers: $x^\alpha$ on domain $\mathbb{R}_{++}$ for $0 \leq \alpha \leq 1$

Entropy: $-x \log x$ on domain $\mathbb{R}_{++}$

**Examples of Convex Functions $\mathbb{R}^n \to \mathbb{R}$:**

1. **Affine functions** on $\mathbb{R}^n$ are **both convex and concave:**

   On $\mathbb{R}^n$, for some $a \in \mathbb{R}^n$ and $b \in \mathbb{R}$:

$$f(x) = a^{\mathrm{T}} x + b$$

2. **Vector Norms** on $\mathbb{R}^n$ are all convex:

   On $\mathbb{R}^n$, $\ell_p$ norms have the form, for $p \geq 1$,

$$\|x\|_p = \left( \sum_{i=1}^{n} |x|^p \right)^{\frac{1}{p}}, \text{ with } \|x\|_\infty = \max_i |x_i|$$

**Examples of Convex Functions $\mathbb{R}^{m \times n} \to \mathbb{R}$:**

1. **Affine functions** on $\mathbb{R}^{m \times n}$ are **both convex and concave:**

   On $\mathbb{R}^{m \times n}$, for some $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}$:

$$f(X) = \mathrm{tr}(A^{\mathrm{T}} X) + b = \sum_{i=1}^{m} \sum_{j=1}^{n} A_{ij} X_{ij} + b$$

1. **Matrix Norms** on $\mathbb{R}^{m \times n}$ are all convex:

   On $\mathbb{R}^{m \times n}$, the spectral, or **maximum singular value norm** is:

$$\| X \|_2 = \sigma_{\max}(X) = [\lambda_{\max}(X^{\mathrm{T}} X)]^{\frac{1}{2}}$$

- **Properties and Important Theorems**

Certain operations preserve the convexity of functions and we listed below

**Theorem (Non-negative Weighted Sum):**

If $f$ is a convex function, then $\alpha f$ is convex for $\alpha \geq 0$. For several convex functions $g_i$, $\sum_i \alpha_i g_i$ is convex if all $\alpha_i \geq 0$.

**Theorem (Composition with Affine Function):**

If $f$ is a convex function, then $f(Ax+b)$ is convex.

**Example:** $\|Ax - b\|$ is convex for any norm.

**Theorem (Pointwise Maximum):**

If $f_1, \ldots, f_m$ are convex functions, then $f(x) = \max\{f_1(x), \ldots, f_m(x)\}$ is convex.

**Example:** Piecewise linear functions $\max_{i=1,\ldots,m}\{a_i^{\mathrm{T}} x + b\}$ are convex.

**Theorem: Pointwise supremum:**

If $f(x,y)$ is convex in $x$ for every $y \in \mathcal{Y}$, then $g(x) = \sup_{y \in \mathcal{Y}} f(x,y)$ is convex.

**Example:**

1. Support function of set $\mathcal{A}$:

$$S_{\mathcal{A}}(x) = \sup_{y \in \mathcal{A}} y^{\mathrm{T}} x$$

2. Distance to farthest point in a (possibly non-convex) set $\mathcal{B}$:

$$f(x) = \sup_{y \in \mathcal{B}} \|x - y\|$$

3. Maximum eigenvalue of a matrix $X \in \mathbb{S}^n$

$$\lambda_{\max}(X) = \sup_{\|y\|_2 \leq 1} y^{\mathrm{T}} X y$$

**Theorem (Parametric Minimization, sometimes also called Partial Minimization):**

If $f(x,y)$ is convex in $(x,y)$ and set $\mathcal{C}$ is convex, then:

$$g(x) = \min_{y \in \mathcal{C}} f(x,y)$$

is convex

**Example:**

$\mathrm{dist}(x, \mathcal{S}) = \inf_{y \in \mathcal{S}} \|x - y\|$ is convex if $\mathcal{S}$ is convex

Note: this one is important and has wide applications in control since we have the value function defined as $V(x) = \min_{u} J(x.u)$ where $x$ is the states and $u$ is the control input

**Theorem (Composition with Scalar Functions):**

For $g: \mathbb{R}^n \to \mathbb{R}$ and $h$: $\mathbb{R} \to \mathbb{R}$, $f(x) = h(g(x))$ is convex if:

$g$ is convex, $h$ is convex, $\tilde{h}$ (extended value function) is non-decreasing.

$g$ is concave, $h$ is concave, $\tilde{h}$ (extended value function) is non-increasing.

**Example:** $\exp g(x)$ for convex $g$, $1/g(x)$ for concave positive $g$

**Theorem (Composition with Vector Functions):**

For $g$: $\mathbb{R}^n \to \mathbb{R}^k$ and $h$: $\mathbb{R}^k \to \mathbb{R}$, $f(x) = h(g_1(x), g_2(x), \ldots, g_k(x))$ is convex if:

Each $g_i$ is convex, $h$ is convex, $\tilde{h}$ is non-decreasing in each argument.

Each $g_i$ is concave, $h$ is concave, $\tilde{h}$ is non-increasing in each argument.

**Example:**

**1.** $\log \sum_{i=1}^{k} \exp g_i(x)$ is convex if all $g_i$ are convex

**2.** $\sum_{i=1}^{k} \log g_i(x)$ is concave for concave positive $g_i$

## IV. Convex Optimization Problems

- **Standard Form Convex Optimization Problem**

A standard form **convex optimization problem:**

$$\begin{aligned} \min_{x \in \operatorname{dom}(f)} \quad & f(x) \\ \text{subj. to.} \quad & g_i(x) \leq 0 \quad i = 1, \ldots, m \\ & a_i^{\mathrm{T}} x = b_i \quad i = 1, \ldots, p \end{aligned}$$

This problem is convex if:

1. The domain $\operatorname{dom}(f)$ is a convex set.
2. The objective function $f$ is a convex function.
3. The inequality constraint functions $g_i$ are all convex.
4. The equality constraint functions $h_i(x) = a_i^{\mathrm{T}} x$ **are all affine.** (if equality is curved, obviously it is not convex since the mid-point of the connecting line is not on the curve)

The affine constraints are typically gathered into matrix form:

$$\begin{aligned} \min_{x \in \operatorname{dom}(f)} \quad & f(x) \\ \text{subj. to.} \quad & g_i(x) \leq 0 \quad i = 1, \ldots, m \\ & Ax = b \qquad A = \mathbb{R}^{p \times m} \end{aligned}$$

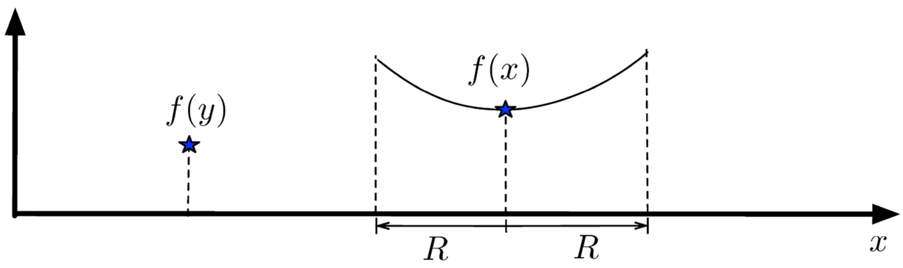
**Theorem (Local and Global Optimality for Convex Problems):**

For a convex optimization problem, **every locally optimal solution is globally optimal.**

**Proof (by contradiction):**

Assume that $x$ is locally optimal, but not globally optimal. Therefore, there is some other point $y$ such that $f(y) < f(x)$. However, $x$ is locally optimal, implies that there is some $R > 0$ such that:

$$\|z - x\|_2 \leq R \Rightarrow f(x) \leq f(z)$$




From the figure above, we can see that the problem can not be convex, thus we have contradiction.

- **Optimality Criterion for Differentiable Objective Function** $f$

**Theorem (Optimality Criterion for Convex Differentiable** $f$**):**

For a convex problem with a differentiable objective function $f$, $x$ is optimal **if and only if it is feasible, and:**

$$\nabla f(x)^{\mathrm{T}}(y-x) \geq 0, \text{ for all feasible } y$$

**Theorem: (Descent Direction)**

If there exists a vector $\mathbf{d}$ such that $\nabla f(\bar{x})'\mathbf{d} < 0$, then there exists a $\delta > 0$ such that $f(\bar{x}+\lambda\mathbf{d}) < f(\bar{x})$ for all $\lambda \in (0,\delta)$.

The vector $\mathbf{d}$ is called descent direction.

The direction of **steepest descent** $\mathbf{d}_s$ at $\bar{x}$ is defined as the normalized direction where $\nabla f(\bar{x})'\mathbf{d}_s$ is minimized.

The direction of $\mathbf{d}_s$ of steepest descent is:

$$\mathbf{d}_s = -\frac{\nabla f(\bar{x})}{\|\nabla f(\bar{x})\|}$$

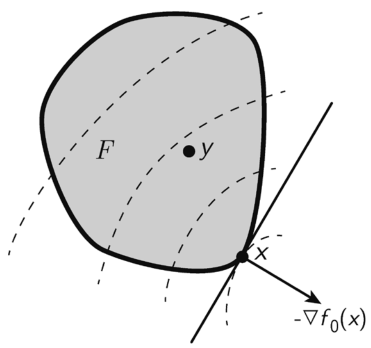



Illustration of the optimality condition for a feasible set $F \subset \mathbb{R}^2$.

Note:

1. For points **strictly inside the feasible set**, the optimality condition is $\nabla f(x) = 0$ (the familiar sufficient first-order optimality condition)
2. The expression above (optimality theorem) states that the gradient may be non-zero, as long as all other feasible points are not downhill from the optimum, so it is a more generalized expression that also holds **for points on the boundary**.
3. The expression from intuition should be $-\nabla f(x)^{\mathrm{T}}(y-x) \leq 0$, note that the $\nabla f(x)$ is the direction of objective ascent (hence $-\nabla f(x)$ is descent) $(y-x)$ is the vector pointing to possible feasible $y$ from the point $x$ which we want to judge is optimal or not. The expression then means all the feasible places $y$ we could go from the current point $x$ would lead to the ascent of objective value (thus, it is the optimal point).

**Unconstrained Problem:**

$$\min_{x \in \mathrm{dom}(f)} f(x)$$

Optimal condition: $x$ is optimal if and only if $x \in \mathrm{dom}(f), \nabla f(x) = 0$

**Equality Constrained Problem:**

$$\min_{x \in \operatorname{dom}(f)} f(x)$$
$$\text{subj. to. } Ax = b$$

Optimal condition: $x$ is optimal if and only if $x \in \operatorname{dom}(f),\ Ax = b,\ \exists \nu : \nabla f(x) + A^{\mathrm{T}} \nu = 0$

**Minimization over Non-negative Orthant:**

$$\min_{x \in \operatorname{dom}(f)} f(x)$$
$$\text{subj. to. } x \geq 0$$

Optimal condition: $x$ is optimal if and only if:

$$x \in \operatorname{dom}(f),\ x \geq 0,\ \nabla f(x)_i \geq 0, x_i = 0;\ \nabla f(x)_i = 0, x_i > 0$$

**Theorem (Simple Optimality Properties for Nonconvex Differentiable $f$):**

**Necessary condition:**

If $x^*$ is a local minimizer, then $\nabla f(x^*) = 0$

**Sufficient condition:**

Suppose that $f$ is twice differentiable at $x^*$. If $\nabla f(x^*) = 0$ and the Hessian of $f(x)$ at $x^*$ is positive definite, then $x^*$ is a local minimizer.

**Necessary and Sufficient condition:**

Suppose that $f(x) = x^{\mathrm{T}} H x + c^{\mathrm{T}} x + k$. If $H$ is positive definite, then $x^*$ is the global minimizer if and only if $\nabla f(x^*) = 0$.

Note:

Proofs are available in Ch.4 of Bazaraa, Sherali, and Shetty. Nonlinear Programming - Theory and Algorithms. John Wiley & Sons, Inc. New York, second edition, 1993.

**Applications: Least Squares**

Least Square problem:

$$\min_{x \in \mathbb{R}^n} \| Ax - b \|_2^2$$

Analytical solution $x^* = A^{\dagger} b$, if $A$ full column rank, where $A^{\dagger} = (A^{\mathrm{T}} A)^{-1} A^{\mathrm{T}}$ is often called the pseudo-inverse.

**Proof:**

$$\min_x \| Ax - b \|_2^2 = \min_x\ x^{\mathrm{T}} (A^{\mathrm{T}} A) x - x^{\mathrm{T}} (2 A^{\mathrm{T}} b) + b^{\mathrm{T}} b$$

If $A$ is full column rank than $A^{\mathrm{T}} A$ is positive definite. From the above theorem, we can know that at the optimal point $x^*$, we should have:

$$\nabla f(x^*) = 0 \Rightarrow (2 A^{\mathrm{T}} A) x^* = (2 A^{\mathrm{T}} b) \Rightarrow x^* = (A^{\mathrm{T}} A)^{-1} A^{\mathrm{T}} b$$

- **Equivalent Optimization Problems**

Two problems are **(informally) called equivalent** if the solution to one can be (easily) **inferred from** the solution to the other, and vice versa.

**Technique 1: Introducing equality constraints for separation:**

$$\min_{x} \ f(A_0 x + b_0)$$
$$\text{subj. to. } g_i(A_i x + b_i) \leq 0 \quad i = 1, \ldots, m$$

is equivalent to:

$$\min_{x} \ f(y_0)$$
$$\text{subj. to. } g_i(y_i) \leq 0 \quad i = 1, \ldots, m$$
$$A_i x + b_i = y_i \ \ i = 0, 1, \ldots m$$

**Technique 2: Introducing slack variable for linear inequalities:**

$$\min_{x} \ f(x)$$
$$\text{subj. to. } A_i x \leq b_i \quad i = 1, \ldots, m$$

is equivalent to:

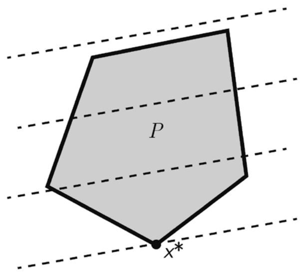
$$\min_{x, s_i} \ f(x)$$
$$\text{subj. to. } A_i x + s_i = b_i \quad i = 1, \ldots, m$$
$$s_i \geq 0 \quad i = 1, \ldots, m$$

- **Linear Program (LP)**

**General Linear Program**

Standard format:

$$\min_{x} \ c^{\mathrm{T}} x + d$$
$$\text{subj. to. } Gx \leq h$$
$$Ax = b$$



Linear optimization on a polytope.

The feasible set is a **polyhedron**. The constant component $d$ can be left out because it has no effect on the optimal solution.

An alternative format:

$$\min_{x} \ c^{\mathrm{T}} x + d$$
$$\text{subj. to. } Ax = b$$
$$x \geq 0$$

This is converted from introducing extra variables in the standard format. In this format, all components of $x$ are non-negative.

Note:

1. Many problems can be rewritten (with some effort!) into LPs
2. A huge variety of solution methods and software are available.

3. The feasible set can be either a polyhedron or a polytope. There is no difference in the case of LP. The objective function is linear so geometrically the level sets (contours) are straight lines.

**Examples for LP: Piecewise Affine Minimization**

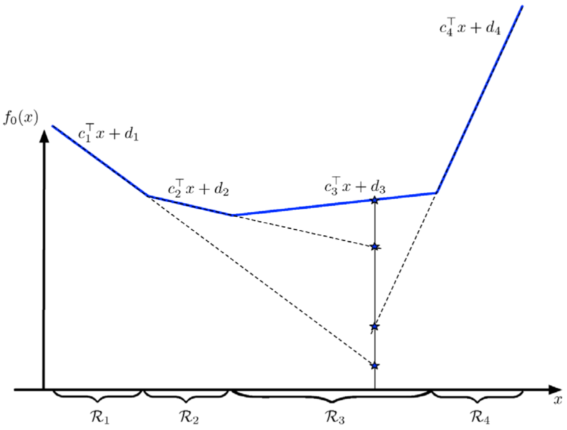
$$\min_x \left[ \max_{i=1,\dots,m} \{c_i^{\mathrm{T}} x + d_i\} \right]$$

$$\text{subj. to. } Gx \le h$$

The function is affine on each region $\mathcal{R}_i$, and from the figure on the right, we can see that the objective function $\max\{c_i^{\mathrm{T}} x + d_i\}$ is convex.

Any convex and piecewise affine function can be written this way.

This problem can be reformulated as an LP!




A piecewise affine (PWA) convex function.

Note that we can introduce a new variable $t$, representing the maximum that each affine segment should be less than, then the objective is to minimize the $t$ value, which can be written as:

$$\min_{x,t} \ t$$

$$\text{subj. to. } c_i^{\mathrm{T}} x + d_i \le t \quad i = 1, \dots, m$$

$$Gx \le h$$

Note: we can also understand this equivalent format from the **epigraph form** perspective.

**Examples for LP: $\ell_\infty$ (Chebyshev Minimization)**

Constrained $\ell_\infty$ (Chebyshev) minimization is the optimization as follows:

$$\min_{x \in \mathbb{R}^n} \|x\|_\infty \quad \Leftrightarrow \quad \min_{x \in \mathbb{R}^n} [\max\{x_1, \dots, x_n, -x_1, \dots, -x_n\}]$$

$$\text{subj. to. } Fx \le g \qquad\qquad \text{subj. to. } Fx \le g$$

This problem is exactly the same as the above problem in the special case that $c_i = \pm 1$ and $d = 0$, so we can still use the same technique to transform the problem as:

$$\begin{array}{llll} & \min_{x,t} \ t & & \min_{x,t} \ t \\ \text{subj. to.} & x_i \le t \quad i = 1, \dots, m & \Rightarrow \ \text{subj. to.} & x \le \mathbf{1}t \\ & -x_i \le t \quad i = 1, \dots, m & & -x \le \mathbf{1}t \\ & Fx \le g & & Fx \le g \end{array}$$

where $\mathbf{1}$ indicates a vector of ones. The constraint $-\mathbf{1}t \le x \le \mathbf{1}t$ bounds the absolute value of every element of $x$ with **a single** common scalar variable $t$.

**Examples for LP: Constrained $\ell_1$ Minimization**

$$\min_{x \in \mathbb{R}^n} \|Ax - b\|_1 \quad \Leftrightarrow \quad \min_{x \in \mathbb{R}^n} \left[ \sum_{i=1}^m \max\{(Ax-b)_i, -(Ax-b)_i\} \right]$$

$$\text{subj. to. } Fx \le g \qquad\qquad \text{subj. to. } Fx \le g$$

Still, using the same technique, this problem can be converted to piecewise affine minimization. The only difference is here we need to introduce multiple $t_i$ (or, a vector $t$):

$$\begin{aligned} \min_{x\in\mathbb{R}^n, t\in\mathbb{R}^n} \quad & t_1+\ldots+t_m \\ \text{subj. to.} \quad & (Ax-b)_i \le t_i \quad i=1,\ldots,m \\ & -(Ax-b)_i \le t_i \quad i=1,\ldots,m \\ & Fx \le g \end{aligned} \Rightarrow \begin{aligned} \min_{x\in\mathbb{R}^n, t\in\mathbb{R}^n} \quad & \mathbf{1}^{\mathrm{T}} t \\ \text{subj. to.} \quad & Ax-b \le t \\ & -(Ax-b) \le t \\ & Fx \le g \end{aligned}$$

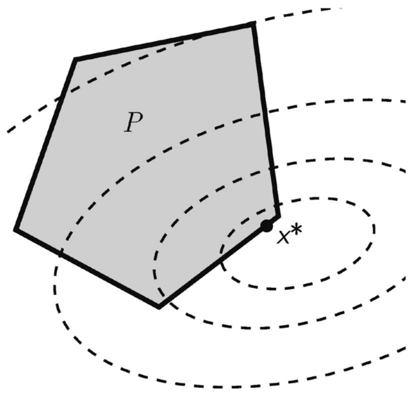
where $\mathbf{1}$ indicates a vector of ones. The constraint $-t \le (Ax-b) \le t$ bounds the absolute value of every element of $(Ax-b)$ with **each component of the vector variable** $t$.

- **Quadratic Program (QP)**

**General Convex Quadratic Program**

Standard format: Quadratic cost function with $P \in \mathbb{S}_+^n$, affine constraint functions

$$\begin{aligned} \min_x \quad & \frac{1}{2} x^{\mathrm{T}} P x + q^{\mathrm{T}} x + r \\ \text{subj. to.} \quad & Gx \le h \\ & Ax = b \end{aligned}$$




Similar to LP, the constant component $r$ can be left out because it has no effect on the optimal solution.

Note:

1. Maximization problems with a concave objective function, i.e. $-P \in \mathbb{S}_+^n$ are also quadratic programs
2. The feasible set is polyhedron and the objective contours are ellipsoids

**Examples for QP: Least Square**

The least square problem:

$$\min_x \ \|Ax-b\|_2^2$$

Without constraints, the analytical solution is: $x = A^{\dagger} b$ where $A^{\dagger}$ is the pseudo-inverse

When Extra linear constraints $l \le x \le u$ are added, it can still be solved but would no longer have an analytical solution.

**Examples for QP: Linear Program with a Random Cost:**

We have a random cost function vector $c$ with mean $\bar{c}$ and covariance $\Sigma$. Hence the linear objective $c^{\mathrm{T}} x$ is a random variable with mean $\bar{c}^{\mathrm{T}} x$ and variance $\gamma x^{\mathrm{T}} \Sigma x$. We wish to penalize the expected cost plus a "risk premium" on the variance.

$$\begin{aligned} \min_x \quad & \mathbb{E}[c^{\mathrm{T}} x] + \gamma \mathrm{var}(c^{\mathrm{T}} x) = \bar{c}^{\mathrm{T}} x + \gamma x^{\mathrm{T}} \Sigma x \\ \text{subj. to.} \quad & Gx \le h \\ & Ax = b \end{aligned}$$

Note: Intuition for setting the objective: Large $\gamma$ means large risk aversion, we prefer a small variance to the lowest expected cost.

**Examples for QP: Tikhonov Regularization**

Least squares with the **extra penalty for nonzero terms (1-norm term, also called the regularizer)**:

$$\min_{x \in \mathbb{R}} \|Ax - b\|_2^2 + \gamma \|x\|_1$$

Equivalent to:

$$\min_{t, x \in \mathbb{R}} \|Ax - b\|_2^2 + \gamma \mathbf{1}^{\mathrm{T}} t$$
$$\text{subj. to. } -t \leq x \leq t$$

Note:

1. We have converted an unconstrained problem into a larger constrained one to get it **into standard QP form.**
2. To ensure convexity, we need $\gamma \geq 0$
3. Intuition for the 1-norm term: Often when finding the optimal solution for least square QP in high dimensional space, the resulting solutions $x^*$ would be small, or many of its components would be small **but not zero**. If we can drive the small but not zero terms into zero, we can get the sparse solution, and sparsity can be useful in practice. Therefore, we need to penalize the non-zero terms. Also note that for the region near zero, 1-norm penalty is much more effective than 2-norm (gradient/ derivative for linear terms would maintain constant while for quadratic terms would be very small). That is why Tikhonov regularization is introduced.

## V. Duality

- **The Lagrangian Function and Dual Function**

Recall our standard (**possibly non-convex**) optimization problem $(P)$:

$$\min_{x \in \mathrm{dom}(f)} f(x)$$
$$\text{subj. to. } g_i(x) \leq 0 \quad i = 1, \ldots, m$$
$$h_i(x) = 0 \quad i = 1, \ldots, p$$

with (primal) decision variable $x$, domain $\mathrm{dom}(f)$ and optimal value $p^*(J^*)$

Define **Lagrangian function** $L \colon \mathrm{dom}(f) \times \mathbb{R}^m \times \mathbb{R}^p \rightarrow \mathbb{R}$**:**

$$L(x, \lambda, \nu) = f(x) + \sum_{i=1}^{m} \lambda_i g_i(x) + \sum_{i=1}^{p} \nu_i h_i(x)$$

Where $\lambda_i$ is **inequality Lagrange multiplier** for $g_i(x) \leq 0$, $\nu_i$ is **equality Lagrange multiplier** for $h_i(x) = 0$. The Lagrangian is a **weighted sum** of the objective and constraint functions.

Our optimization problem is now turned into **minimizing the Lagrangian function (original objective + penalty for constraints).**

Note: Why is the Lagrangian function in the form of the weighted sum of objectives and constraints?

A simple and naïve intuition for the Lagrangian function is that we want to tune the penalty on the constraints to more or less satisfy the constraints.

Following the idea above, we define the dual function $d$: $\mathbb{R}^m \times \mathbb{R}^p$:

$$
\begin{aligned}
d(\lambda,\nu) &= \inf_{x\in \mathrm{dom}(f)} L(x,\lambda,\nu) \\
&= \inf_{x\in \mathrm{dom}(f)} \left[ f(x) + \sum_{i=1}^{m} \lambda_i g_i(x) + \sum_{i=1}^{p} \nu_i h_i(x) \right]
\end{aligned}
$$

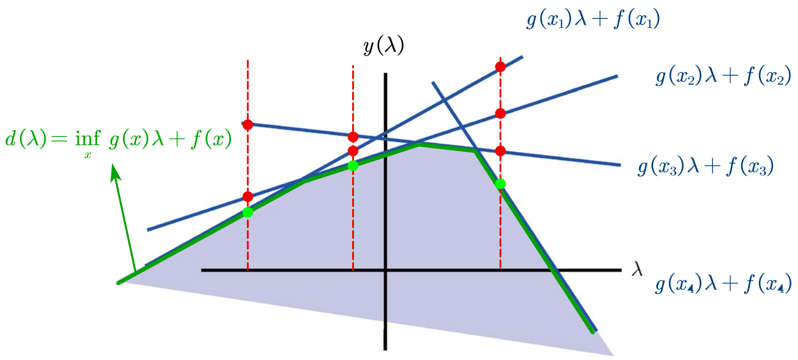
Important results:

1. The dual function $d(\lambda,\nu)$ is **always a concave function** because it is the pointwise infimum of affine functions (see illustration of simple cases below)
2. dual function generates lower bounds for $p^*$, i.e.:
$$d(\lambda,\nu) \leq p^*,\ \forall(\lambda \geq 0, \nu \in \mathbb{R}^p)$$
3. $d(\lambda,\nu)$ might be $-\infty$, i.e.:
$$\mathrm{dom}(d) = \{\lambda,\nu \mid d(\lambda,\nu) > -\infty\}$$

Note:

1. Illustration of dual function convexity

   consider a simple case where:
   $$x^* = \mathrm{argmin}\ f(x)$$
   $$\text{subj. to. } g(x) \leq 0$$
   Then the dual function is hence:
   $$
   \begin{aligned}
   d(\lambda) &= \inf_x L(x,\lambda) \\
   &= \inf_x f(x) + \lambda g(x)
   \end{aligned}
   $$




   The dual function $f(x) + \lambda g(x)$ is affine in $\lambda$, where $f(x)$ is the intercept term and $g(x)$ is the slope, and hence can be represented as straight lines in the figure above. To get the final dual function $d(\lambda)$, according to the definition, we need to find the infimum over $x$ when and treat $\lambda$ as fixed. Then in the figure it is a vertical line crossing across all the possible $x_i$ in $\mathrm{dom}(f)$ (here we assume there are 4 $x_i$ in the domain for simple illustration, there could be infinitely many for continuous decision variable space), and we need to select the lowest point (green point). For example, for the line on the rightmost, we choose $x_4$ for that $\lambda$ while for the leftmost, we choose $x_1$. If we keep doing so for all the $\lambda$, finally the green points would generate a concave function (green line), which is the dual function.
2. Proof (informal) for lower bound: still use the simple case above, we have
$$d(\lambda) = \inf_x L(x,\lambda) = \inf_x f(x) + \lambda g(x) \leq f(x^*) + \lambda g(x^*) \xRightarrow[g(x^*)\leq 0]{\lambda > 0} \leq p^*$$

   The first inequality holds because the dual function is the infimum over $x$ for a certain $\lambda$, the second inequality holds because $f(x^*)=p^*$ and $\lambda g(x^*)\leq 0$

3. Always remember when constructing dual function, we minimize over $x$ to get a function of $\lambda,\nu$.

**Examples for Dual function: Least Norm Solution to a linear system**

The corresponding optimization function $(P)$ is:

$$\min_{x\in\mathbb{R}^n} x^{\mathrm{T}}x$$

$$\text{subj. to. } Ax=b$$

From the formulation, we can construct Lagrangian function $L(x,v)=x^{\mathrm{T}}x+\nu^{\mathrm{T}}(Ax-b)$

The dual function is constructed by minimizing the Lagrangian, set the gradient of Lagrangian to zero:

$$\nabla_x L(x,\nu)=2x+A^{\mathrm{T}}\nu=0\Rightarrow x=-\frac{1}{2}A^{\mathrm{T}}\nu$$

Substituting this $x$ back into Lagrangian to get the dual function:

$$d(\nu)=-\frac{1}{4}\nu^{\mathrm{T}}AA^{\mathrm{T}}\nu-b^{\mathrm{T}}\nu$$

It's easy to check that $d(\nu)$ is concave for $\nu$, and according to the properties, we know that it generates the lower bound for every $\nu$, i.e. $d(\nu)\leq p^*,\ \forall\nu$.

- **The Dual Problem**

Using dual function, we get the lower bound for our optimal optimization objective $p^*$. The problem is how should we make the lower bound get the closest to the real value?

$\rightarrow$ Since the dual function is concave, it is intuitive to do maximization of dual function.

The dual problem $(D)$ is defined as:

$$\max_{\lambda,\nu} D(\lambda,\nu)$$

$$\text{subj. to. } \lambda\geq 0$$

Important facts:

1. Dual problem $(D)$ is **concave** even if the original problem $(P)$ is not
2. Dual problem $(D)$ has optimal value $d^*\leq p^*$
3. The point $(\lambda,\nu)$ is **dual feasible** if $\lambda\geq 0$ and $(\lambda,\nu)\in\mathrm{dom}(d)$
4. We can often impose the constraints $(\lambda,\nu)\in\mathrm{dom}(d)$ explicitly in dual problem $(D)$.

**Examples for dual problem: Dual of a Linear Program (LP)**

The original problem $(P)$ is:

$$\min_{x\in\mathbb{R}^n} c^{\mathrm{T}}x$$

$$\text{subj. to. } Ax=b$$

$$Cx\leq e$$

The dual function is:

$$d(\lambda.\nu)=\min_{x\in\mathbb{R}^n}\ [c^{\mathrm{T}}x+\nu^{\mathrm{T}}(Ax-b)+\lambda^{\mathrm{T}}(Cx-e)]$$

$$= \min_{x \in \mathbb{R}^n} \left[ (A^T\nu + C^T\lambda + c^T)x - b^T\nu - e^T\lambda \right]$$

$$= \begin{cases} -b^T\nu - e^T\lambda & \text{if } A^T\nu + C^T\lambda + c^T = 0 \\ -\infty & \text{otherwise} \end{cases}$$

Lower bound property: $-b^T\nu - e^T\lambda \le p^*$ whenever $\lambda \ge 0$ and $A^T\nu + C^T\lambda + c^T = 0$. Thus, the dual problem $(D)$ is:

$$\min_{\lambda, \nu} \ -b^T\nu - e^T\lambda$$

$$\text{subj. to. } A^T\nu + C^T\lambda + c^T = 0$$

$$\lambda \ge 0$$

Note:

1. We can see that the dual of a linear program is also a linear program.
2. Why do we emphasize $-b^T\nu - e^T\lambda \le p^*$ whenever $\lambda \ge 0$ and $A^T\nu + C^T\lambda + c^T = 0$ for constructing the dual problem? It is because we are maximizing for dual problem, and for cases that $A^T\nu + C^T\lambda + c^T \ne 0$, the lower bound is already $-\infty$, we focus on the **bounded and better lower bound**. Also, note that since the lower bound $-b^T\nu - e^T\lambda \le p^*$ only holds when $A^T\nu + C^T\lambda + c^T = 0$, so $A^T\nu + C^T\lambda + c^T = 0$ **should be written as a constraint in the dual problem.**
3. When is the dual problem easier to solve than the original problem? First, they are both LP, so we should compare the problem dimension. Note that when transforming $(P)$ to $(D)$. The variable to optimized over in transferred from $x$ to $(\lambda, \nu)$. Thus, we need to compare the dimension of the original decision variable and the number of constraints (i.e. the dimension of $\lambda$ and $\nu$)

**Examples for Dual problem: Dual of Norm Minimization with Equality Constraint**

The original problem $(P)$ is:

$$\min_{x \in \mathbb{R}^n} \ \|x\|_2$$

$$\text{subj. to. } Ax = b$$

The dual function is:

$$d(\nu) = \min_{x \in \mathbb{R}^n} \left[ \|x\| + \nu^T(Ax - b) \right] = \min_{x \in \mathbb{R}^n} \left[ \|x\| - (A^T\nu)^T x + b^T\nu \right]$$

$$= \begin{cases} b^T\nu & \text{if } \|A^T\nu\|_2 \le 1 \\ -\infty & \text{otherwise} \end{cases}$$

Lower bound property: $b^T\nu \le p^*$ whenever $\|A^T\nu\|_2 \le 1$.

The dual problem $(D)$ is:

$$\min_{\nu} \ b^T\nu$$

$$\text{subj. to. } \|A^T\nu\|_2 \le 1$$

**Examples for Dual problem: Dual of a Quadratic Program with Inequality Constraint**

The original problem $(P)$ is (with $Q \succ 0$):

$$\min_{x \in \mathbb{R}^n} \frac{1}{2} x^{\mathrm{T}} Q x + c^{\mathrm{T}} x$$

$$\text{subj. to. } Cx \leq e$$

The dual function is:

$$\begin{aligned} d(\lambda) &= \min_{x \in \mathbb{R}^n} \left[ \frac{1}{2} x^{\mathrm{T}} Q x + c^{\mathrm{T}} x + \lambda^{\mathrm{T}} (Cx - e) \right] \\ &= \min_{x \in \mathbb{R}^n} \left[ \frac{1}{2} x^{\mathrm{T}} Q x + (c^{\mathrm{T}} + C^{\mathrm{T}} \lambda)^{\mathrm{T}} x - e^{\mathrm{T}} \lambda \right] \end{aligned}$$

This is an unconstrained convex minimization, recall the optimality condition for unconstrained ($\nabla_x = 0$), the optimal $x$ satisfies $x = -Q^{-1}(c + C^{\mathrm{T}} \lambda)$ and hence:

$$d(\lambda) = -\frac{1}{2}(c + C^{\mathrm{T}} \lambda)^{\mathrm{T}} Q^{-1} (c + C^{\mathrm{T}} \lambda) - e^{\mathrm{T}} \lambda$$

Lower bound property: $-\frac{1}{2}(c + C^{\mathrm{T}} \lambda)^{\mathrm{T}} Q^{-1} (c + C^{\mathrm{T}} \lambda) - e^{\mathrm{T}} \lambda \leq p^*$ whenever $\lambda \geq 0$

The dual problem $(D)$ is:

$$\min_{\lambda} \frac{1}{2} \lambda^{\mathrm{T}} C Q^{-1} C^{\mathrm{T}} \lambda + (C Q^{-1} c + e)^{\mathrm{T}} \lambda + \frac{1}{2} c^{\mathrm{T}} Q^{-1} c$$

$$\text{subj. to. } \lambda \geq 0$$

Note: The dual of QP is another QP.

**Examples for Dual problem: Dual of Mixed-Integer Linear Program (MILP)**

The original problem $(P)$ is:

$$\min_{x \in \mathcal{X}} c^{\mathrm{T}} x$$

$$\text{subj. to. } Ax \leq b$$

$$\mathcal{X} = \{-1, 1\}^n$$

The dual function is:

$$\begin{aligned} d(\lambda) &= \min_{x_i \in \{-1, 1\}} [c^{\mathrm{T}} x + \lambda^{\mathrm{T}} (Ax - b)] \\ &= -\|A^{\mathrm{T}} \lambda + c\|_1 - b^{\mathrm{T}} \lambda \end{aligned}$$

Lower bound property: $-\|A^{\mathrm{T}} \lambda + c\|_1 - b^{\mathrm{T}} \lambda \leq p^*$ whenever $\lambda \geq 0$

The dual problem $(D)$ is:

$$\max_{\lambda} -\|A^{\mathrm{T}} \lambda + c\|_1 - b^{\mathrm{T}} \lambda$$

$$\text{subj. to. } \lambda \geq 0$$

Note:

1. The dual of a mixed-integer LP is a LP (without integers).
2. The minimization of $c^{\mathrm{T}} x + \lambda^{\mathrm{T}} A x$ for $x_i \in \{-1, 1\}$ results into $-\|A^{\mathrm{T}} \lambda + c\|_1$, which can be

easily checked by expanding and writing out the expression.

**Summary for Solving Optimization from Dual Procedure:**

1. Constrained optimization $(P) \to$ Unconstrained optimization (through Lagrangian function)
2. Solve the Unconstrained optimization over $x \to$ dual function $d$ (function about $\lambda, \nu$)
3. Dual function gives lower bound for $p^* \to$ maximizing dual function: the dual problem $(D)$
4. Solve the dual problem to get $d^*$: we hope to get as close to the optimal objective $p^*$ as possible.

Note:

In optimization problem, we care about 2 things: The first is the optimal objective $p^*$, the second is the optimal solution $x^*$ that leads to $p^*$. In dual procedure, we are getting the lower bound for $p^*$ regardless of whether $x$ satisfies the constraints. i.e., from dual procedure, we **don't necessarily get the feasible solution**.

- **Weak and Strong Duality, Duality Gap**

**Definition (Weak Duality, Strong Duality, Duality Gap):**

Weak duality is: $d^* \le p^*$ and it always holds

Strong duality is: $d^* = p^*$ and it generally does not hold, even for convex problems.

The duality gap is defined as: $p^* - d^*$

The certificate of optimality: $d \le d^* \le p^* \le p$

i.e. The dual and primal (the original problem) costs bound the optimal dual and primal costs.

**A Geometric Interpretation of Duality Gap:**

For simplicity, consider the problem only with the inequality constraints:

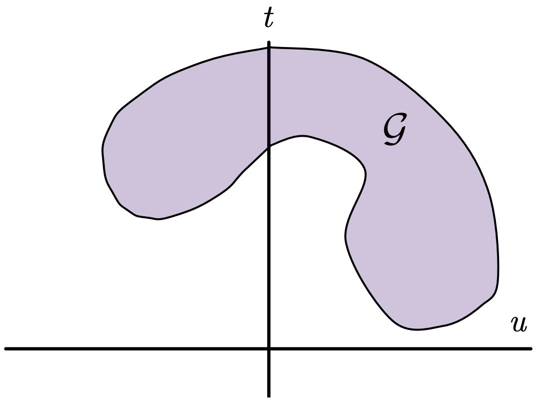
$$\min_{x \in \mathcal{X}} f(x)$$

$$\text{s.t. } g(x) \le 0$$

In order to consider it geometrically, we decide to visualize and map the problem to the objective-constraint value pair space. The set $\mathcal{G}$ in the following plot is given as:

$$\mathcal{G} = \{(u,t) \mid t = f(x),\ u = g(x),\ x \in \mathcal{X}\}$$

In some sense, $\mathcal{G}$ **can be regarded as a representation of mapping the domain** $\mathcal{X}$ **in the original space to** $t - u$ **space we want to discuss.**



1. Consider Primal (Original) Problem

The primal (original) problem $P$ is:

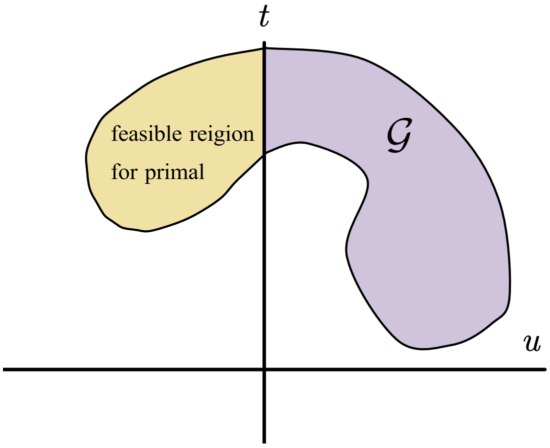
$$(P): \quad p^* = \min_{x \in \mathcal{X}} f(x)$$

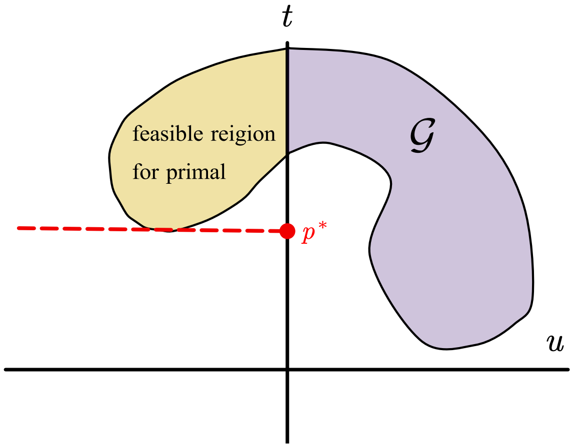
$$\text{s.t. } g(x) \leq 0$$

In the $t-u$ space, it should be written as:

$$p^* = \min\{t, (u,t) \in \mathcal{G}, \ u \leq 0\}$$

The primal problem is just the completely constrained problem and the feasible region in the original space is the decision variable **domain $\mathcal{X}$ plus the constraints. Therefore the feasible region in $t-u$ space is $\mathcal{G}$ plus** $u \leq 0$, i.e. we only search in the yellow region below, and the optimal solution can easily be found.




2. Consider Dual Procedure (Min-Max Game)

The dual procedure contains two stages: A. finding the dual function. B. maximizing the dual function. The process of maximizing dual function (stage B) is also called the dual problem $D$.

A. finding the dual function

The dual function is found through minimizing the Lagrange function. The Lagrange function transfers a constrained problem into an unconstrained one with the Lagrange multiplier.

The process of minimizing the unconstrained Lagrange function is originally written as:

$$d(\lambda) = \min_{x \in \mathcal{X}} f(x) + \lambda g(x)$$

In the $t-u$ space, it should be written as:

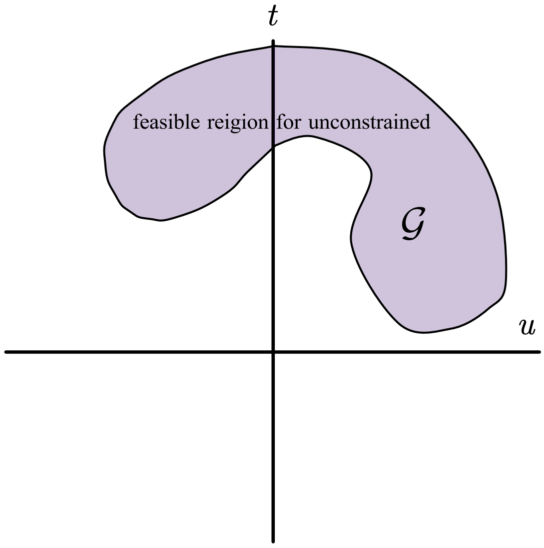
$$d(\lambda) = \min_{(u,t) \in \mathcal{G}} (t + \lambda u)$$

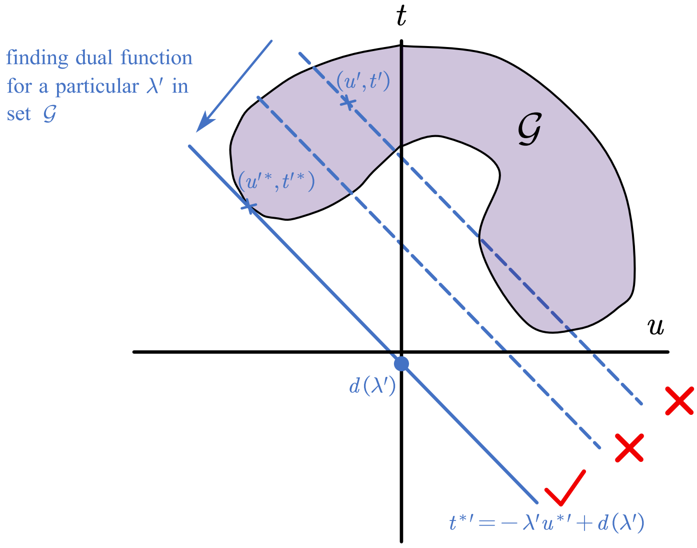
Note that the feasible region for the unconstrained problem in the original space is just the decision variable **domain $\mathcal{X}$ and hence the feasible region in $t-u$ space is the whole $\mathcal{G}$, not the yellow colored subset in the primal problem.**

The dual function is constructed by searching the $\mathcal{G}$ **for a particular** $\lambda$. To be more specific, a single $\lambda$ is given (fixed) and we try to find the $(t^*, u^*)$ and the corresponding $d$. **If for all the possible**

$\lambda \geq 0$**, its corresponding** $d$ **are found, we actually have the relationship between** $\lambda$ **and** $d$**, which is exactly the dual function** $d = d(\lambda)$. Also, note that finding $(t^*, u^*)$ in $\mathcal{G}$ is equivalent to finding $x^*$ in $\mathcal{X}$.

In the $(t, u)$ space, $d(\lambda) = t + \lambda u$ is a straight line, and the slope is $-\lambda$. For example, in the following figure, use the blue line to represent a particular $\lambda'$, then find the dual function means **selecting a line that has the lowest intercept from a series of parallel lines**.



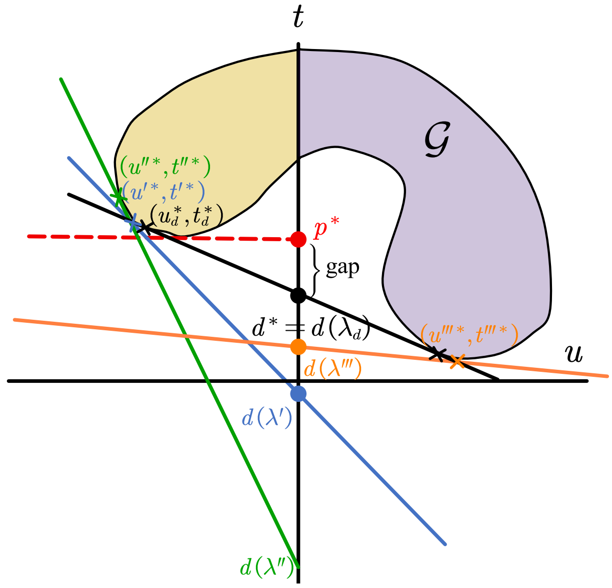




B. maximizing the dual function

After constructing the dual function, we have the correspondence between $\lambda$ and $d(\lambda)$ (also, $(u^*, t^*)$) for each of the $\lambda \geq 0$ (hence slope $-\lambda \leq 0$). In other words, for each $\lambda$, we have a corresponding straight line (and thus where the limit case is). Then we solve the dual problem:

$$(D):\ d^* = \max_{\lambda \geq 0} d(\lambda)$$

This procedure is equivalent to picking out the slope (and the corresponding straight line) that would maximize the intercept, in the figure we listed 4 possible values straight lines: line$'$, line$''$, line$'''$ and optimal line, subscripted $d$ (actually, there are infinitely many candidate lines).



Note that due to the **concave inward (hence non-convex) shape** of set $\mathcal{G}$, the best we can do is the optimal black line **as it is tangent to the boundary of $\mathcal{G}$ at two points**. We can never reach the optimal solution (horizontal red line), and the duality gap appears.

**Theorem (Slater Condition):**

Consider an optimization with $f$ and all $g_i$ **convex**:

$$\min f(x)$$
$$\text{subj. to. } g_i(x) \leq 0 \quad i=1,\ldots,m$$
$$Ax = b \quad A \in \mathbb{R}^{p \times n}$$

If there is **at least one strictly feasible point**, i.e.

$$\{x \mid Ax = b,\ g_i(x) < 0,\ \forall i \in \{1,\ldots,m\}\} \neq \varnothing$$

Then strong duality holds, i.e. $p^* = d^*$

There are several extended results:

1. Stronger version: Only $g_i(x)$ must be strictly satisfiable
2. Strong duality holds for LPs if at least one of the problems (primal or dual) is feasible.

Note: Other **constraint qualification conditions** exist to check strong duality in convex problems. Slater condition is only one of them.

- **Optimality Condition**

In the section above, we know that transferring the primal problem to the dual problem is good but not ideal because there is the duality gap and we give the sufficient condition (slater condition) for strong duality. The next thing we want to do is to derive **the necessary condition for strong duality, and thus conclude the necessary and sufficient condition for optimality**.

**Necessary condition derivation (proof):**

Assume we have strong duality ($d^* = p^*$). Let $x^*$ and $(\lambda^*, \nu^*)$ be primal and dual solution. Then from the definition of the dual function:

$$f(x^*) = p^* = d^* = d(\lambda^*, \nu^*) = \min_x L(x, \lambda^*, \nu^*)$$
$$= \min_x \left\{ f(x) + \sum_{i=1}^{m} \lambda_i^* g_i(x) + \sum_{i=1}^{p} \nu_i^* h_i(x) \right\}$$
$$\leq f(x^*) + \underbrace{\sum_{i=1}^{m} \lambda_i^* g_i(x)}_{\leq 0} + \underbrace{\sum_{i=1}^{p} \nu_i^* h_i(x)}_{=0} \leq f(x^*)$$

The equality holds if and only if $\sum_{i=1}^{m} \lambda_i^* g_i(x^*) = 0$. In other words, we want the inequality ($\leq 0$) term to be equality ($= 0$). Thus we need $\lambda_i^* = 0$ for every $g_i(x^*) < 0$ and $g_i(x^*) = 0$ for every $\lambda > 0$, which is called **complementary slackness**. And we have point $x^*$ minimizes $L(x, \lambda^*, \nu^*)$, which is called **stationary condition**.

All the above derivations can be concluded as the **KKT condition**.

**Theorem (Karush-Kuhn-Tucker Conditions (Necessity)):**

Assume that all $g_i$ and $h_i$ are differentiable. If $d^*=p^*$ (i.e. strong duality holds), then the optimal solutions $x^*,(\lambda^*,\nu^*)$ satisfy the Karush-Kuhn-Tucker Conditions (KKT condition):

1. Primal Feasibility

$$g_i(x^*)\leq 0 \quad i=1,2,\ldots,m$$
$$h_i(x^*)=0 \quad i=1,2,\ldots,p$$

2. Dual Feasibility

$$\lambda^*\geq 0$$

3. Complementary Slackness

$$\lambda_i^* g_i(x^*)=0 \quad i=1,2,\ldots,m$$

4. Stationarity

$$\nabla_x L(x^*,\lambda^*,\nu^*)=\nabla_x f(x^*)+\sum_{i=1}^{m}\lambda_i^*\nabla_x g_i(x^*)+\sum_{i=1}^{p}\nu_i^*\nabla_x h_i(x^*)=0$$

Note: This statement (necessity) actually assumes nothing a priori about the convexity of our problem.

**Theorem (Karush-Kuhn-Tucker Conditions for Convex Problems (Sufficiency)):**

For a convex optimization problem , if $(x^*,\lambda^*,\nu^*)$ satisfy the KKT condition, then:

$$p^*=f(x^*)=L(x^*,\lambda^*,\nu^*) \quad \text{(due to complementary slackness)}$$
$$d^*=d(\lambda^*,\nu^*)=L(x^*,\lambda^*,\nu^*) \quad \text{(due to convexity and stationarity)}$$

Hence $p^*=d^*$, i.e. the duality gap is 0. $x^*$ and $(\lambda^*,\nu^*)$ are primal and dual optimal solutions.

Note: for sufficiency, we require the convexity of the problem.

**Theorem (Karush-Kuhn-Tucker Conditions for Convex Problems (Necessity and Sufficiency)):**

For a convex optimization problem and If Slater's condition holds (i.e. strong duality holds):

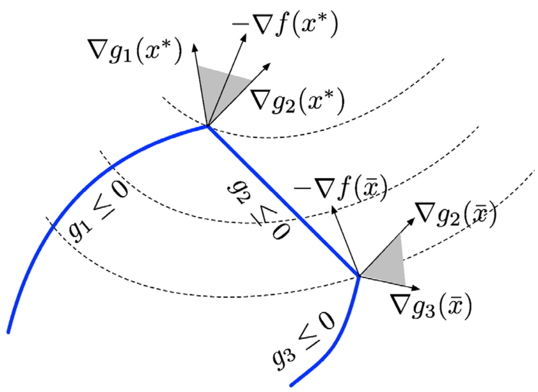
$$x^* \text{ and } (\lambda^*,\nu^*) \text{ are primal and dual optimal solutions}$$
$$\Longleftrightarrow x^* \text{ and } (\lambda^*,\nu^*) \text{ satisfy the KKT conditions.}$$

Note: In summary, zero duality gap (thus optimality) and KKT condition has the following relationship:

$$\text{KKT} \xRightarrow{\text{convexity}} \text{ZDP} \quad \text{KKT} \xLeftarrow{\text{strong duality}} \text{ZDP} \quad \text{KKT} \xLeftrightarrow{\text{slater condition}} \text{ZDP}$$

Here we assume Slater condition already implies the convexity of the problem.

**A Geometric Interpretation of KKT Conditions**




Assume inequality constraints only

Rewrite the stationary condition as:

$$-\nabla_x f(x)=\sum_{i=1}^{m}\lambda_i^*\nabla_x g_i(x)$$

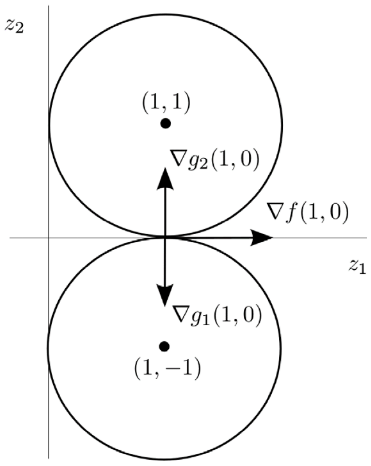
The geometric meaning of the above formula means the direction of the steepest descent is in a convex cone spanned by constraints gradients $\nabla g_i$

Note: In other words, the convex cone spanned by constraints gradients is the region where the would lead to infeasibility, so if the steepest descent is in that region, it means that the objective cannot be lower because we encounter a dead end, thus implies optimality.

**Example of Duality Gap and Slater Condition in Convex Problems**

Consider a specific example of the problem as follows:

$$\begin{aligned} &\min_{z_1, z_2} \; z_1 \\ \text{subj. to. } &(z_1 - 1)^2 + (z_2 - 1)^2 \leq 1 \\ &(z_1 - 1)^2 + (z_2 + 1)^2 \leq 1 \end{aligned}$$




From the plot, there is only a single feasible point $z = (1, 0)$. And the KKT condition does not hold for any $\lambda_i \geq 0$ because in this problem, the gradient of the objective function is not in the positive cone spanned by the gradients of the constraint (the cone is a line lie aligned with the gradient of the constraint and is normal to the gradient of the objective function). We know if strong duality holds, the point must satisfy the KKT condition, hence the strong duality does not hold for this case.

Note: This example verifies what we mentioned in the previous section: **even for convex problems, strong duality would not necessarily hold!** And it also shows what the Slater condition is doing: ensuring the feasible region has **a relative interior point.**

**Example of KKT condition: QP**

Consider a (convex) quadratic program $(P)$ with $Q \succeq 0$:

$$\begin{aligned} &\min_{x \in \mathbb{R}^n} \; \frac{1}{2} x^T Q x + c^T x \\ &\text{subj. to. } Ax = b \\ &\qquad\qquad\quad x \geq 0 \end{aligned}$$

The Lagrangian is:

$$L(x, \lambda, \nu) = \frac{1}{2} x^T Q x + c^T x + \nu^T (Ax - b) - \lambda^T x$$

The KKT conditions are:

| | |
|---|---|
| $\nabla_x L(x, \lambda, \nu) = Qx + A^T \nu - \lambda + c = 0$ | [stationary] |
| $Ax = b$ | [primal feasibility] |
| $x \geq 0$ | [primal feasibility] |
| $\lambda \geq 0$ | [dual feasibility] |
| $x_i \lambda_i = 0 \quad i = 1, \ldots n$ | [complementarity] |

The final three conditions are often written together as $0 \leq x \perp \lambda \geq 0$

**Example of KKT Condition: Implicit vs. Explicit Constraints**

It is sometimes helpful to keep some or all of the constraints in $\mathcal{X}$, e.g. In the box-constrained LP:

Method (Formulation) 1:

$$\min_{x \in \mathbb{R}^n} c^{\mathrm{T}} x$$

$$\text{subj. to. } Ax = b$$

$$-\mathbf{1} \le x \le \mathbf{1}$$

Dual function:

$$d\left(\overline{\lambda}, \underline{\lambda}, \nu\right) = \min_{x \in \mathbb{R}^n} \left[c^{\mathrm{T}} x + (Ax - b)^{\mathrm{T}} \nu + (-\mathbf{1} - x)^{\mathrm{T}} \underline{\lambda} + (-\mathbf{1} + x)^{\mathrm{T}} \overline{\lambda}\right]$$

$$= \begin{cases} -b^{\mathrm{T}} \nu - \mathbf{1}^{\mathrm{T}} \overline{\lambda} - \mathbf{1}^{\mathrm{T}} \underline{\lambda} & \text{if } c + A^{\mathrm{T}} \nu - \underline{\lambda} + \overline{\lambda} = 0 \\ -\infty & \text{otherwise} \end{cases}$$

Dual Problem $(D)$:

$$\max_{\nu, \overline{\lambda}, \underline{\lambda}} \; -b^{\mathrm{T}} \nu - \mathbf{1}^{\mathrm{T}} \overline{\lambda} - \mathbf{1}^{\mathrm{T}} \underline{\lambda}$$

$$\text{subj. to. } c + A^{\mathrm{T}} \nu - \underline{\lambda} + \overline{\lambda} = 0$$

$$\overline{\lambda} \ge 0, \; \underline{\lambda} \ge 0$$

Method (Formulation) 2:

$$\min_{\|x\|_\infty \le 1} \; c^{\mathrm{T}} x$$

$$\text{subj. to. } Ax = b$$

Dual function:

$$d(\nu) = \min_{\|x\|_\infty \le 1} \; [c^{\mathrm{T}} x + (Ax - b)^{\mathrm{T}} \nu] = -b^{\mathrm{T}} \nu - \|A^{\mathrm{T}} \nu + c\|_1$$

Dual Problem $(D)$: no constraints! A lot simpler and the dimension is lower.

$$\max_{\nu} \; -b^{\mathrm{T}} \nu - \|A^{\mathrm{T}} \nu + c\|_1$$

- **Sensitivity Analysis**

**Perturbed Optimization Problem**

A general optimization problem and its dual is:

| | |
|---|---|
| $\min_x \; f(x)$ | $\max_{\nu,\lambda} \; d(\nu, \lambda)$ |
| $\text{subj. to. } g_i(x) \le 0 \;\; i = 1, \ldots, m$ | $\text{subj. to. } \lambda \ge 0$ |
| $h_i(x) = 0 \;\; i = 1, \ldots, p$ | |

A **perturbed** optimization problem and its dual:

| | |
|---|---|
| $\min_x \; f(x)$ | $\max_{\nu,\lambda} \; d(\nu, \lambda) - u^{\mathrm{T}} \lambda - v^{\mathrm{T}} \nu$ |
| $\text{subj. to. } g_i(x) \le u_i \;\; i = 1, \ldots, m$ | $\text{subj. to. } \lambda \ge 0$ |
| $h_i(x) = v_i \;\; i = 1, \ldots, p$ | |

$x$ is the primal decision variable, $(\lambda, \nu)$ are the dual decision variables

$u$ and $v$ are parameters representing perturbations to the constraints. $p^*(u,v)$ is the optimal value as a function of $(u,v)$

Assume strong duality for the unperturbed problem with $(\nu^*, \lambda^*)$ dual optimal. Then we have:

$$p^*(u,v) \geq d^*(\nu^*, \lambda^*) - u^{\mathrm{T}}\lambda^* - v^{\mathrm{T}}\nu^* = p^*(0,0) - u^{\mathrm{T}}\lambda^* - v^{\mathrm{T}}\nu^*$$

The first inequality is due to the weak duality of the perturbed problem. The second equality is due to the strong duality of the original problem.

Note: Sensitivity is about analyzing **how the cost would change if the constraints are relaxed.** From the above formula we can see that the $\lambda^*$ and $\nu^*$ gives the lower bound of how much the objective changes with perturbation/relaxation $u, v$.

**Global/Local Sensitivity Analysis**

According to the formula above, we can have the following results (Global):

$\lambda_i^*$ large and $u_i < 0 \Rightarrow p^*(u,v)$ increase greatly

$\lambda_i^*$ small and $u_i > 0 \Rightarrow p^*(u,v)$ does not decrease much

$$\left.\begin{array}{l} \nu^* \text{ large and positive and } v_i < 0 \\ \nu^* \text{ large and negative and } v_i < 0 \end{array}\right\} \Rightarrow p^*(u,v) \text{ does not decrease much}$$

$$\left.\begin{array}{l} \nu^* \text{ small and positive and } v_i > 0 \\ \nu^* \text{ small and negative and } v_i > 0 \end{array}\right\} \Rightarrow p^*(u,v) \text{ does not decrease much}$$

Note: Results are not symmetrical. It is because we only have a lower bound on $p^*(u,v)$

If in addition $p^*(u,v)$ is differentiable at $(0,0)$, then (Local)

$$\lambda_i^* = \frac{\partial p^*(0,0)}{\partial u_i}, \ \nu_i^* = -\frac{\partial p^*(0,0)}{\partial v_i}$$

$\lambda_i^*$ is sensitivity of $p^*$ relative to $i$ th inequality.

$\nu_i^*$ is sensitivity of $p^*$ relative to $i$ th equality.