Chapter III
Convex Analysis

Skills to acquire

- **How to recognize convexity of sets, functions, and problems.**
- **How to characterize convexity of sets, functions, and problems.**

1 Convex sets

1.1 Basic notions

Let $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k \in \mathbb{R}^n$ and $C \subset \mathbb{R}^n$.

Definition 1 (Points combination)**.** A a *linear combination* of $(\boldsymbol{x}_i)_i$ is any sum

$$\boldsymbol{x} = \theta_1 \boldsymbol{x}_1 + \ldots + \theta_k \boldsymbol{x}_k.$$

where $\theta_i \in \mathbb{R}$. This combination is called *convex* if $0 \leqslant \theta_i \leqslant 1, \forall\, i$ and $\sum\limits_{i=1}^{k} \theta_i = 1$; *affine* if $\sum\limits_{i=1}^{k} \theta_i = 1$; and *conic* if $\theta_i \geqslant 0,\ \forall\, i$.

Definition 2. The set $\mathcal{C}$ is called *convex* if it contains all of its convex combinations:

$$\forall\, (x, y, t) \in \mathcal{C}^2 \times [0, 1], \quad (1-t)x + ty \in \mathcal{C}.$$

Similarly, $\mathcal{C}$ is called *affine subspace* if it contains all its affine combinations.

Example 1 (Convex sets)**.** Let $A \in \mathbb{R}^{m \times n}$, $\mathbf{a} \in \mathbb{R}^n$, $\mathbf{c} \in \mathbb{R}^m$, and $b \in \mathbb{R}$. The following sets are convex.

- Euclidean balls (for any norm).
- Positive Orthant: $\{\boldsymbol{x} \in \mathbb{R} : x_i \geqslant 0, \forall\, i\}$.
- Hyperplane: $\{\boldsymbol{x} \in \mathbb{R}^n : \mathbf{a}^T \boldsymbol{x} = b\}$, Halfspace: $\{\boldsymbol{x} \in \mathbb{R}^n : \mathbf{a}^T \boldsymbol{x} \leqslant b\}$.
- Affine space:$\{\boldsymbol{x} \in \mathbb{R}^n : A\boldsymbol{x} = \mathbf{c}\}$.

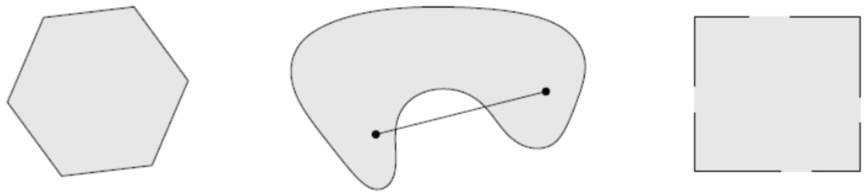

Fig. III.1: Examples of convex (first) and non-convex sets.

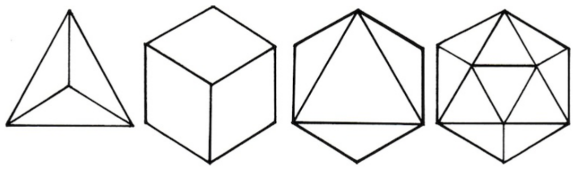

Fig. III.2: Polyhedra

- Polyhedron: $\{\boldsymbol{x} \in \mathbb{R}^n : A\boldsymbol{x} \preceq \mathbf{c}\}$, where inequality is component-wise.

- Simplex: $conv\{\boldsymbol{x}_0, \ldots, \boldsymbol{x}_k\}$, where these points are affinely independent[1]. It is a special case of polyhedron and a generalization of the notion of a triangle to arbitrary dimensions.

- Ellipsoid: for a symmetric $Q \succ 0$ (positive definite)

$$\left\{\boldsymbol{x} \in \mathbb{R}^n : (\boldsymbol{x} - \mathbf{c})^T Q(\boldsymbol{x} - \mathbf{c}) \leqslant r^2\right\}.$$

Definition 3. $\mathcal{C}$ is called a *cone* if:

$$x \in \mathcal{C} \implies tx \in \mathcal{C},\ \forall t \geqslant 0.$$

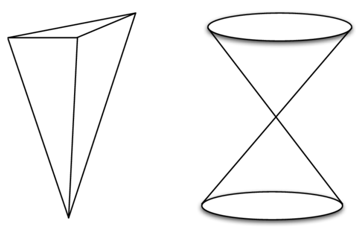

Fig. III.3: Examples of convex and non-convex cones

It is easy to show that $\mathcal{C}$ is a convex cone if:

$$x_1, x_2 \in \mathcal{C} \implies t_1 x_1 + t_2 x_2 \in \mathcal{C},\ \forall t_1, t_2 \geqslant 0.$$

Example 2 (Convex cones)**.** The following sets are convex cones.

- *Norm cone* (for any norm $\|\cdot\|$):

$$\{(\boldsymbol{x}, t) \in \mathbb{R}^n \times \mathbb{R} \ : \ \|\boldsymbol{x}\| \leqslant t\}.$$

For ℓ_2-norm, it is called second-order cone.

[1] $\boldsymbol{x}_0, \ldots, \boldsymbol{x}_k$ are affinely independent means $\boldsymbol{x}_1 - \boldsymbol{x}_0, \ldots, \boldsymbol{x}_k - \boldsymbol{x}_0$ are linear independent.

- *Normal cone* (given a closed convex $\mathcal{C}$ and point $\boldsymbol{x} \in \mathcal{C}$):

$$\mathcal{N}_{\mathcal{C}}(\boldsymbol{x}) = \{\mathbf{g} \in \mathbb{R}^n | \mathbf{g}^T(\mathbf{z} - \boldsymbol{x}) \leqslant 0, \ \forall \mathbf{z} \in \mathcal{C}\}.$$

  By convention, we let $\mathcal{N}_C(\boldsymbol{x}) := \emptyset$ when $\boldsymbol{x} \notin \mathcal{C}$.

  1. $\mathcal{C} = [0,1]$: $\mathcal{N}_{\mathcal{C}}(x) = \begin{cases} \mathbb{R}_- & \text{if } x = 0 \\ \mathbb{R}_+ & \text{if } x = 1 \\ \{0\} & \text{if } x \in]0,1[\\ \emptyset & \text{otherwise} \end{cases}$.
  2. $\mathcal{C} = \{\boldsymbol{x} \in \mathbb{R}^n \mid \|\boldsymbol{x}\| \leqslant 1\}$: $\mathcal{N}_{\mathcal{C}}(\boldsymbol{x}) = \begin{cases} \mathbb{R}_+ \boldsymbol{x}, \text{ if } & \|\boldsymbol{x}\| = 1 \\ \{0\}, \quad \text{if } & \|\boldsymbol{x}\| < 1 \\ \emptyset, & \text{otherwise} \end{cases}$.
  3. The normal cone of a triangle at some points is depicted in Figure III.4.

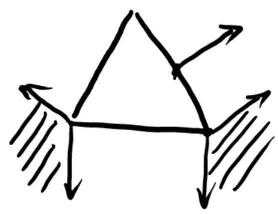

Fig. III.4: Some normal cones.

- *Positive semidefinite matrices*

The convexification of a nonconvex set X is achieved by defining the convex hull, which is the smallest convex set containing X. It is also the intersection of all convex sets containing X.

Definition 4. *Convex* (resp. *conic*, *affine*) hull of S, denoted by *conv*(S) (resp. where aff(S)), is the set of all convex (resp. conic, affine) combinations of elements of S.

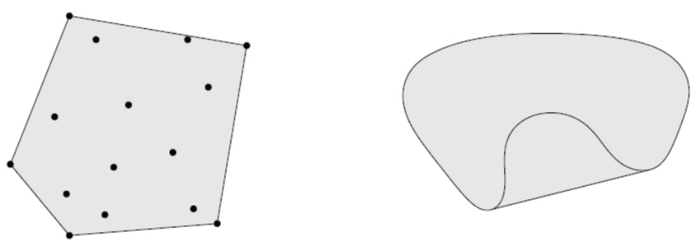

Fig. III.5: Examples of convex hulls

> **Remark:** A convex hull is always convex even when $\mathcal{C}$ is not convex.

We now consider some generic topological properties of convex sets.

Proposition 1. The closure and the interior of a convex set are convex.

The relative interior of a set S, denoted by $\mathrm{relint}(S)$, is a refinement of the concept of the interior, which is often more useful when dealing with low-dimensional sets placed in higher-dimensional spaces.

Definition 5. The relative interior of a set S (denoted $\mathrm{relint}(S)$) is defined as its interior within the affine hull of S, i.e.,

$$\mathrm{relint}(S) := \{x \in S : \exists \epsilon > 0, \mathbb{B}(x, \epsilon) \cap \mathrm{aff}(S) \subseteq S\}.$$

All metrics define the same $\mathrm{relint}(S)$.

Example 3. Consider the closed unit square $S := \left\{(x, y, 0) \in \mathbb{R}^3 \,\middle|\, 0 \leq x, y \leq 1\right\}$. We have $\mathrm{int}(S) = \emptyset$ but $\mathrm{relint}(S) = \left\{(x, y, 0) \in \mathbb{R}^3 \,\middle|\, 0 < x, y < 1\right\}$.

1.2 Key properties of convex sets

Theorem 1 (Separating hyperplane theorem)**.** If C, D are nonempty convex sets with $C \cap D = \emptyset$, then there exists $\mathbf{a} \in \mathbb{R}^n$, $b \in \mathbb{R}$ such that

$$C \subset \{\boldsymbol{x} : \mathbf{a}^T \boldsymbol{x} \leqslant b\} \;\text{ and }\; D \subset \{\boldsymbol{x} : \mathbf{a}^T \boldsymbol{x} \geqslant b\}.$$

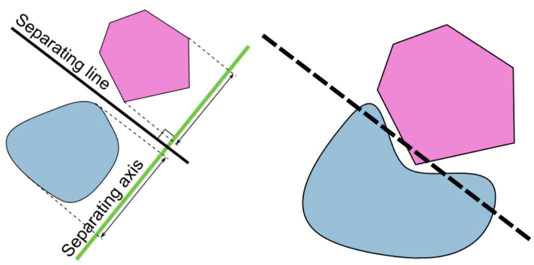

Fig. III.6: Separating hyperplane theorem

Theorem 2 (Supporting hyperplane theorem)**.**
If C is a nonempty convex set, and $\boldsymbol{x}_0 \in \partial(C)$, then there exists $\mathbf{a}$ such that

$$C \subset \{\boldsymbol{x} : \mathbf{a}^T \boldsymbol{x} \leqslant \mathbf{a}^T \boldsymbol{x}_0\}.$$

1.3 Operations preserving sets convexity

To prove the convexity of sets, one can use the definition. Nevertheless, it is usually easier and more effective to use some properties to prove convexity of sets that can be obtained from other sets, for which convexity is easier to establish.

Theorem 3 (Operations preserving sets convexity)**.** The following operations preserve convexity.

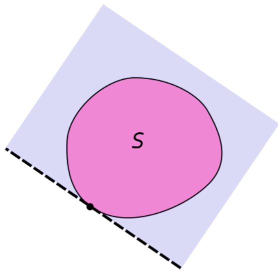

Fig. III.7: Supporting hyperplane

- The *intersection* of convex sets is convex.
- The vector sum $C_1 + C_2$ of two convex sets C_1 and C_2 is convex.
- *Scaling and translation*: if C is convex, then $\lambda C + \boldsymbol{a}$ is convex for any a and λ. Furthermore, if C is a convex set and λ_1, λ_2 are positive scalars,
$$(\lambda_1 + \lambda_2)C = \lambda_1 C + \lambda_2 C.$$
- *Affine images and preimages*: Let $f : \mathbb{R}^n \to \mathbb{R}^m$ with $f(\boldsymbol{x}) = A\boldsymbol{x} + \boldsymbol{b}$, $A \in \mathbb{R}^{n\times m}$ and $\boldsymbol{b} \in \mathbb{R}^m$. If $\mathcal{C}$ is convex then $f(\mathcal{C})$ and $f^{-1}(\mathcal{C})$ are convex.

Example 4. content...

2 Convex functions

Let $f : \mathbb{R}^n \to \mathbb{R}$.

2.1 Basic notions

Definition 6. The function f is called *convex* if $\mathrm{dom}(f)$ convex and

$$\forall\, \boldsymbol{x}, \boldsymbol{y} \in \mathrm{dom}(f),\ \forall t \in [0,1],\ \ f((1-t)\boldsymbol{x} + t\boldsymbol{y}) \leqslant (1-t)f(\boldsymbol{x}) + tf(\boldsymbol{y}) \qquad \text{(III.1)}$$

f is *strictly convex* if f is convex and if equality only holds for $t \in]0,1[$. If $-f$ is convex, f is called concave.

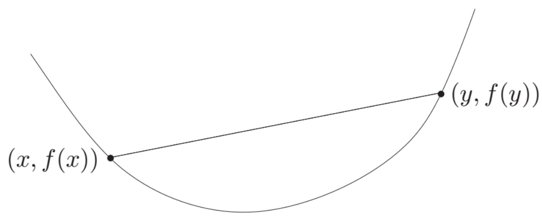

Fig. III.8: Convex function

Definition 7. The function f is called *strongly convex* with parameter $m > 0$ if $\mathrm{dom}(f)$ is convex and

$$\forall\,(\boldsymbol{x},\boldsymbol{y},t) \in \mathrm{dom}(f)^2 \times [0,1]:\ f(t\boldsymbol{x}+(1-t)\boldsymbol{y}) \leq tf(\boldsymbol{x})+(1-t)f(\boldsymbol{y})-\frac{\mu}{2}t(1-t)\|\boldsymbol{x}-\boldsymbol{y}\|^2.$$

f is quasi-convex if

$$\forall\,(\boldsymbol{x},\boldsymbol{y},t) \in \mathrm{dom}(f)^2 \times [0,1], \quad f((1-t)\boldsymbol{x}+t\boldsymbol{y}) \leqslant \max\{f(\boldsymbol{x}), f(\boldsymbol{y})\}$$

> **Remark:** It is often convenient to extend a convex function to all of $\mathbb{R}^n$ as the following:
> $$\tilde{f}(\boldsymbol{x}) = \begin{cases} f(\boldsymbol{x}), & \boldsymbol{x} \in \mathrm{dom}(f), \\ \infty, & \boldsymbol{x} \notin \mathrm{dom}(f). \end{cases}$$

> **Remark:** Note that strongly convex $\Longrightarrow$ strictly convex $\Longrightarrow$ convex. For example, function $f(x) = \frac{1}{x}$ is strictly convex but not strongly convex.

It is often easy to draw a function defined on $\mathbb{R}$ to check geometrically whether it is convex or not. Here are some basic examples.

Example 5. Functions on $\mathbb{R}$.

- Exponential functions e^{ax} are convex for any $a \in \mathbb{R}$.
- Even powers x^p (p is even) and powers of absolute vaue $|x|^p$ for $p \geqslant 1$ are convex.
- Power function x^a is convex for $a \geqslant 1$ or $a \leqslant 0$, and concave for $0 \leqslant a \leqslant 1$.
- Logarithmic function $\log x$ is concave over $\mathbb{R}_{++}$ and $x \log x$ is convex.

Example 6. For any norm $\|\cdot\|$, the following functions are convex:
$f(\boldsymbol{x}) = \|\boldsymbol{x}\|$, $f(\boldsymbol{x}) = \|\boldsymbol{x}\|^2$, and the least squares loss $\|\mathbf{y} - A\boldsymbol{x}\|_2^2$ for any matrix A.

Example 7. For a convex set C, the indicator function $\mathbb{1}_C(\boldsymbol{x})$ convex.

2.2 Characterizations of convex functions

The convexity of functions defined on $\mathbb{R}$ is easy to check geometrically. The following theorem links between convexity on $\mathbb{R}$ and convexity on $\mathbb{R}^n$.

Theorem 4. A function is convex iff its restriction to any line is convex.

The next theorem relates function convexity with sets convexity.

Theorem 5. A function f is convex if its epigraph is convex. Furthermore, if f is convex, then every level set is convex.

> **Remark:** The converse is not true. For example, $f(x) = \sqrt{|x|}$ is not a convex function but each of its sublevel sets are convex sets.

When the function is differentiable, convexity can be characterized using its derivatives.

Theorem 6 (First-order characterization)**.** Assume that f is differentiable and $\text{dom}(f)$ is convex. Then, f is convex if and only if f completely lies above each of its tangent hyperplanes, i.e.,

$$\forall \boldsymbol{x}, \boldsymbol{y} \in \text{dom}(f): \quad f(\mathbf{y}) \geqslant f(\boldsymbol{x}) + \nabla f(\boldsymbol{x})^T(\mathbf{y} - \boldsymbol{x}).$$

Furthermore, if f is convex then ∇f is a monotone mapping:

$$(\nabla f(\boldsymbol{x}) - \nabla f(\mathbf{y}))^T (\boldsymbol{x} - \mathbf{y}) \geq 0, \quad \forall \boldsymbol{x}, \mathbf{y} \in dom(f).$$

Theorem 7 (Second-order characterization)**.** If f is twice differentiable and $\text{dom}(f)$ is convex then

$$f \text{ is convex} \quad \Longleftrightarrow \quad \forall x \in \text{dom}(f): \; \nabla^2 f(x) \succeq 0.$$

Example 8. Functions on $\mathbb{R}^n$

- Any affine function $f(\boldsymbol{x}) = \mathbf{a}^T\boldsymbol{x} + b$ is both convex and concave.
- Quadratic function $\frac{1}{2}\boldsymbol{x}^T Q\boldsymbol{x} + \mathbf{b}^T\boldsymbol{x} + c$ is convex provided that $Q \succeq 0$.
- Affine function on $\mathbb{R}^{m\times n}$: $f(X) = \text{tr}(A^TX) + b = \sum_{i,j} A_{i,j}X_{i,j} + b$, is convex.

2.3 Operations preserving functions convexity

The following propositions gives useful techniques to establish function convexity from convexity of simpler functions.

Proposition 2. Operations preserving functions convexity.

- Any conic combination of convex functions is convex.
- Affine composition: if f is convex, then $g(\boldsymbol{x}) = f(A\boldsymbol{x} + \mathbf{b})$ is convex.
- Point-wise maximization: if $(f_i)_{i\in I}$ are convex then $f(\boldsymbol{x}) = \max_{i\in I} f_i(\boldsymbol{x})$ is convex (I here can be infinite).
- Partial minimization: if $g(\boldsymbol{x}, \mathbf{y})$ is convex in and C is convex, then $f(\boldsymbol{x}) = \min_{\boldsymbol{y}\in C} g(\boldsymbol{x}, \boldsymbol{y})$ is convex.

Example 9. The following functions are convex.

- $g(\boldsymbol{x}) = \max(\mathbf{a}_1^T\boldsymbol{x} + b_1, \cdots + \mathbf{a}_k^T x + b_k)$, e.g., $f(\boldsymbol{x}) = \max\{x_1, \ldots, x_n\}$.
- $d(x, C)$, where C is convex.

- For any set C (convex or not), the support function for is convex: $\mathbb{1}_C^*(\boldsymbol{x}) = \max_{\boldsymbol{y} \in C} \boldsymbol{x}^T \boldsymbol{y}$.

Proposition 3. Consider the *composition* $f(\boldsymbol{x}) = h(g(\boldsymbol{x}))$, where $h : \mathbb{R} \to \mathbb{R}$ and $g : \mathbb{R}^n \to \mathbb{R}$. If h is convex then f is convex if $\Big(h$ is nondecreasing and g is convex $\Big)$ or $\Big(h$ is nonincreasing and g is concave $\Big)$

> **Remark:** Trick to remember the rule: $f''(x) = h''(g(x))\, g'(x)^2 + h'(g(x))\, g''(x)$.

Example 10. The following functions are convex: $\exp f(x)$ with f convex and $-\log f(x)$ with f concave.

3 Convex optimization

A convex optimization problem is of the form

$$\begin{aligned} \underset{\boldsymbol{x} \in \mathbb{R}^n}{\text{minimize}} \quad & f(\boldsymbol{x}) \\ \text{subject to} \quad & g_i(\boldsymbol{x}) \leqslant 0,\ i = 1, \ldots, m, \\ & h_j(\boldsymbol{x}) = 0,\ j = 1, \ldots, \ell \end{aligned} \tag{III.2}$$

- $f : \mathbb{R}^n \longrightarrow \overline{\mathbb{R}}$ and $g_i : \mathbb{R}^n \longrightarrow \mathbb{R}$ are all convex,
- h_j are affine; the equality constraints can be written as $A\boldsymbol{x} = \mathbf{b}$.

For an optimization problem, every local minimizer is a global one.

Proposition 4. A local minimizer for a convex optimization is a global minimizer and the solution set X_{opt} is convex.

Example 11 (Basis pursuit). Given $\mathbf{y} \in \mathbb{R}^n$ and $X \in \mathbb{R}^{n \times p}$, where $p > n$, we seek the sparsest solution to the under-determined linear system $X\beta = \mathbf{y}$. A straightforward nonconvex formulation can be expressed as:

$$\begin{aligned} \underset{\beta}{\text{minimize}} \quad & \|\beta\|_0 \\ \text{subject to} \quad & X\beta = \boldsymbol{y} \end{aligned} \tag{PB0}$$

where $\|\beta\|_0$ corresponds to the total number of nonzero elements in a vector. It is actually not a norm. So, we approximate $\|\cdot\|_0$ by a ℓ_1 norm to obtain the *basis pursuit*:

$$\begin{aligned} \underset{\beta}{\text{minimize}} \quad & \|\beta\|_1 \\ \text{subject to} \quad & X\beta = \boldsymbol{y} \end{aligned} \tag{PB}$$

Example 12 (Lasso). Given $\boldsymbol{y} \in \mathbb{R}^n$ and $X \in \mathbb{R}^{n \times p}$, a lasso problem can be formulated as follows.

$$\begin{aligned} \underset{\beta \in \mathbb{R}^p}{\text{minimize}} \quad & \|\boldsymbol{y} - X\beta\|_2^2 \\ \text{subject to} \quad & \|\beta\|_1 \leq s \end{aligned}$$

Lasso consists in finding sparse approximate solution to the system $\boldsymbol{y} = X\beta$.

The optimization problem III.2 can be rewritten as

$$\begin{aligned} \underset{\boldsymbol{x}}{\text{minimize}} \quad & \tilde{f}(\boldsymbol{x}) \\ \text{subject to} \quad & \boldsymbol{x} \in \mathcal{C} \end{aligned} \tag{III.3}$$

where $\mathcal{C}$ represents the constraints set and $\tilde{f}(\boldsymbol{x}) = f(\boldsymbol{x}) + \mathbb{1}_{\mathcal{C}}(\boldsymbol{x})$, where $\mathbb{1}_{\mathcal{C}}(\boldsymbol{x}) = +\infty$ if $\boldsymbol{x} \notin \mathcal{C}$.

Theorem 8 (First order condition for optimality)**.** Given a differentiable function f and convex set $\mathcal{C}$, consider the problem (III.3). Then, a feasible point $\boldsymbol{x}$ is optimal if and only if:

$$\nabla f(\boldsymbol{x})^T(\mathbf{y} - \boldsymbol{x}) \geqslant 0, \quad \forall\, \mathbf{y} \in \mathcal{C}.$$

If $\mathcal{C} = \mathbb{R}^n$ (unconstrained optimization), then optimality condition reduces to familiar $\nabla f(\boldsymbol{x}) = 0$.

Example 13 (Equality-constrained minimization)**.** Consider the equality-constrained problem:

$$\begin{aligned} \underset{\boldsymbol{x}}{\text{minimize}} \quad & f(\boldsymbol{x}) \\ \text{subject to} \quad & A\boldsymbol{x} = \mathbf{b} \end{aligned}$$

with $A \in \mathbb{R}^{n\times p}$ and f differentiable. Using the first-order optimality condition, solution $\boldsymbol{x}$ satisfies $A\boldsymbol{x} = \mathbf{b}$ (so as to be a feasible point) and

$$\nabla f(\boldsymbol{x})^T(\mathbf{y} - \boldsymbol{x}) \geqslant 0, \quad \forall\, \mathbf{y} \text{ such that } A\mathbf{y} = \mathbf{b}.$$

This is equivalent to

$$\nabla f(\boldsymbol{x})^T\mathbf{v} = 0, \quad \forall\ \mathbf{v}\ \in\ \mathrm{Ker}(A).$$

On the other hand, $\ker(A) = \mathrm{Im}(A^T)^\perp$. Indeed,

$$\begin{aligned} \mathbf{v}\ \in \mathrm{Ker}(A) & \iff A\mathbf{v} = 0 \\ & \iff \forall \mathbf{w}\ \langle A\mathbf{v}, \mathbf{w}\rangle = 0 \\ & \iff \forall \mathbf{w}\ \langle \mathbf{v}, A^T\mathbf{w}\rangle = 0 \\ & \iff \mathbf{v} \in \mathrm{Im}(A^T)^\perp. \end{aligned}$$

So, we have

$$\langle \mathbf{v}, \nabla f(\boldsymbol{x})\rangle = 0, \quad \forall\ \mathbf{v} \in \mathrm{Im}(A^T)^\perp.$$

This implies that $\nabla f(\boldsymbol{x}) \in (\mathrm{Im}(A^T)^\perp)^\perp = \mathrm{Im}(A^T)$ and hence,

$$\exists \lambda \in \mathbb{R}^p : \nabla f(\boldsymbol{x}) = -\lambda^T A,$$

which are Lagrange multipliers.

Example 14 (Projection onto a convex set)**.** Consider a convex set $\mathcal{C}$ and the optimization problem:

$$\begin{array}{ll} \underset{\boldsymbol{x}}{\text{minimize}} & \|\mathbf{a} - x\|_2^2 \\ \text{subject to} & \boldsymbol{x} \in \mathcal{C} \end{array}$$

First-order optimality condition says that the solution $\boldsymbol{x}$ satisfies

$$\nabla f(\boldsymbol{x})^T(\mathbf{y} - \boldsymbol{x}) = 2(\boldsymbol{x} - \mathbf{a})^T(\mathbf{y} - \boldsymbol{x}) \geqslant 0, \quad \text{for all } \ \mathbf{y} \in \mathcal{C}$$

Equivalently, this says that

$$\mathbf{a} - \boldsymbol{x} \in \mathcal{N}_{\mathcal{C}}(\boldsymbol{x})$$

where $\mathcal{N}_{\mathcal{C}}$ is the normal cone.