Mathematical Foundations of Data Sciences

Gabriel Peyré
CNRS & DMA, École Normale Supérieure

CHAPTER 18

Convex Analysis

September 9, 2026

Convexity turns local optimality conditions into certificates of global solutions, even when objectives have corners or encode constraints. This provides a common language for understanding regularization and designing optimization methods. We develop subgradients and normal cones, study convex conjugates and their relation to smoothness, and derive dual problems and optimality conditions.

We write $f(x)$ for the objective and x for the variable to be optimized. The main references are [2, 1].

18.1 Basics of Convex Analysis

We consider minimization problems of the form

$$\min_{x\in\mathcal{H}} f(x) \tag{18.1}$$

over the finite-dimensional Hilbert space $\mathcal{H} \stackrel{\text{def.}}{=} \mathbb{R}^N$, with the canonical inner product $\langle\cdot,\cdot\rangle$. Many of these results extend to infinite-dimensional Hilbert spaces, with additional hypotheses where compactness is used.

The objective $f : \mathcal{H} \longrightarrow \bar{\mathbb{R}} \stackrel{\text{def.}}{=} \mathbb{R}\cup\{+\infty\}$ is convex. Allowing f to take the value $+\infty$ encodes constraints directly in the objective; its effective domain is

$$\mathrm{dom}(f) \stackrel{\text{def.}}{=} \{x\,;\, f(x) < +\infty\}.$$

For a set $C \subset \mathcal{H}$, define the indicator function

$$\iota_C(x) \stackrel{\text{def.}}{=} \begin{cases} 0 & \text{if} \quad x\in C, \\ +\infty & \text{otherwise.}\end{cases}$$

18.1.1 Convex Sets and Functions

A set $\Omega \subset \mathcal{H}$ is convex if

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

A function is convex if

$$\forall\,(x,y,t)\in\mathcal{H}^2\times[0,1], \quad f((1-t)x+ty)\leqslant(1-t)f(x)+tf(y) \tag{18.2}$$

Equivalently, its epigraph $\{(x,r)\in\mathcal{H}\times\mathbb{R}\,;\, r\geqslant f(x)\}$ is convex. The inequality is interpreted in the extended real line $\bar{\mathbb{R}}$, with the convention $0\cdot(+\infty)=0$ at the endpoints $t=0,1$. The function f is strictly convex if the inequality in (18.2) is strict whenever $x,y\in\mathrm{dom}(f)$ are distinct and $0<t<1$. A set Ω is convex if and only if ι_Ω is a convex function.

Throughout the chapter, convex functions f are assumed proper, meaning $\mathrm{dom}(f)\neq\emptyset$, and lower semicontinuous (lsc), meaning that for every $x\in\mathcal{H}$,

$$\liminf_{y\to x} f(y)\geqslant f(x).$$

For a convex function, lower semicontinuity is equivalent to the epigraph $\mathrm{epi}(f)$ being closed. We denote by $\Gamma_0(\mathcal{H})$ the set of proper convex lsc functions.

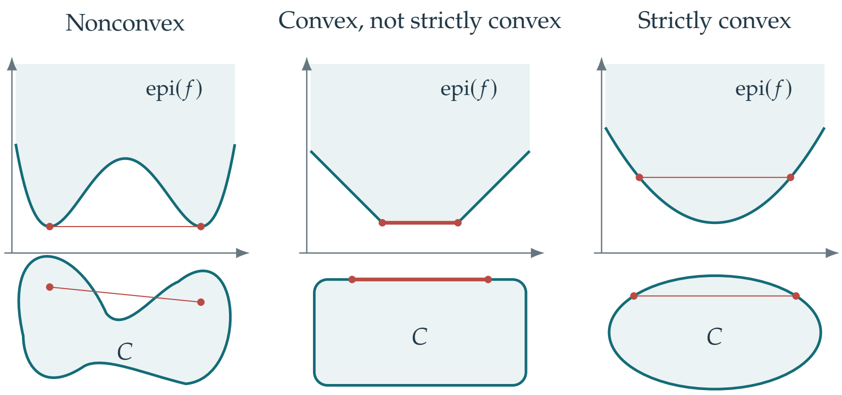


Figure 18.1. Convexity and strict convexity for functions and sets.

18.1.2 First-Order Conditions

Existence of minimizers. We first address existence. Lower semicontinuity and coercivity provide a useful sufficient condition.

Proposition 18.1. *If f is proper, lower semicontinuous, and coercive (i.e. $f(x) \to +\infty$ as $\|x\| \to +\infty$), then there exists a minimizer $x^\star$ of f.*

Proof. Choose $x_0 \in \mathrm{dom}(f)$. The sublevel set $K = \{x : f(x) \leqslant f(x_0)\}$ is nonempty, closed by lower semicontinuity, and bounded by coercivity. It is therefore compact in finite dimensions. A lower semicontinuous function attains its infimum on K, at some $x^\star$. Since $f(x) > f(x_0) \geqslant f(x^\star)$ outside K, this point is a global minimizer. □

This compactness argument is known as the direct method in the calculus of variations. For f in $\Gamma_0(\mathcal{H})$, the minimizer set $\operatorname{argmin} f$ is closed and convex, and every local minimizer is global. Strict convexity of f implies that there is at most one minimizer.

Subdifferential. The subdifferential at x of f is

$$\partial f(x) \stackrel{\text{def.}}{=} \{u \in \mathcal{H}^* \,;\, \forall y \in \mathcal{H}, f(y) \geqslant f(x) + \langle u,\, y - x\rangle\}, \qquad x \in \mathrm{dom}(f).$$

We set $\partial f(x) = \emptyset$ outside $\mathrm{dom}(f)$. Here $\mathcal{H}^* = \mathbb{R}^N$ denotes the space of dual vectors. The inner product identifies the dual space with $\mathcal{H}$, but distinguishing primal and dual variables remains useful. The duality pairing itself is canonical; the identification of a linear functional with a vector depends on the inner product. The subdifferential $\partial f(x)$ consists of the slopes u of supporting affine functions $f(x) + \langle u,\, z - x\rangle$ that lie below f.

At an interior point of $\mathrm{dom}(f)$, the function f is differentiable exactly when its subdifferential consists of the gradient alone:

$$\partial f(x) = \{\nabla f(x)\}.$$

Multiple elements of $\partial f(x)$ indicate distinct supporting slopes of f at x.

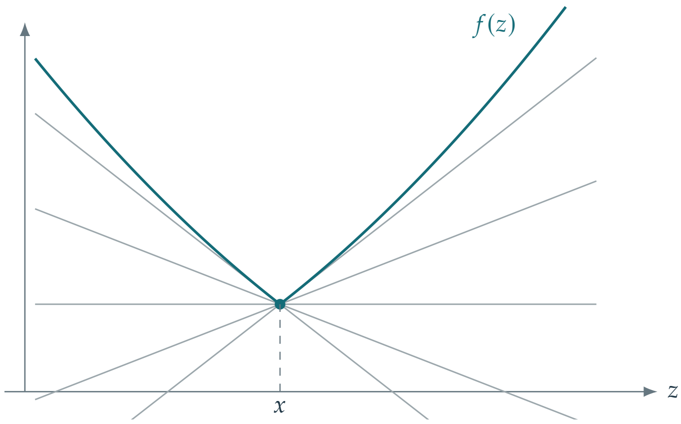
The subdifferential can also be empty at a boundary point of the effective domain. For example, $f(x) = -\sqrt{1 - x^2}$ on $[-1, 1]$, extended by $+\infty$ outside this interval, has no subgradient at $x = \pm 1$.

Since $\partial f(x) \subset \mathcal{H}^*$ is an intersection of half-spaces, it is a closed convex set. Thus $\partial f : \mathcal{H} \mapsto 2^{\mathcal{H}^*}$ is a set-valued operator, also written $\partial f : \mathcal{H} \hookrightarrow \mathcal{H}^*$.

Remark 18.2 (Maximally monotone operator). The subdifferential ∂f is monotone: setting $U = \partial f$ gives

$$\forall (u, v) \in U(x) \times U(y), \quad \langle y - x,\, v - u\rangle \geqslant 0.$$

For single-valued maps on the real line, monotonicity means being nondecreasing. Subdifferentials can also be shown to be maximally monotone, meaning that their graphs are not properly contained in the graph of another monotone operator. Conversely, a monotone map need not be a subdifferential; an example is $(x, y) \mapsto (-y, x)$. Many results extend from subdifferentials to arbitrary maximally monotone operators. We restrict attention to subdifferentials here.



$u \in \partial f(x)$

$f(z) \geq f(x) + u(z - x)$

for every z.

Several affine functions support the graph at this kink.

Figure 18.2. The subdifferential

For the absolute value $f(x) = |x|$, write $\partial f(x) = \partial|\cdot|(x)$. Its subdifferential is

$$\partial|\cdot|(x) = \begin{cases} \{-1\} & \text{if} \quad x < 0, \\ \{1\} & \text{if} \quad x > 0, \\ [-1, 1] & \text{if} \quad x = 0. \end{cases}$$

First-order conditions. The subdifferential gives a necessary and sufficient optimality condition.

Proposition 18.3. *$x^\star$ is a minimizer of f if and only if $0 \in \partial f(x^\star)$.*

Proof. One has

$$x^\star \in \operatorname{argmin} f \quad \Leftrightarrow \quad \big(\forall y, f(x^\star) \leqslant f(y) + \langle 0,\, x^\star - y\rangle\big) \quad \Leftrightarrow \quad 0 \in \partial f(x^\star).$$

□

Subdifferential calculus. Calculus rules simplify subdifferential computations. For a separable function $f(x_1, \ldots, x_K) = \sum_{k=1}^K f_k(x_k)$, the subdifferential is the Cartesian product

$$\partial f(x_1, \ldots, x_K) = \partial f_1(x_1) \times \ldots \times \partial f_K(x_K).$$

Applying this rule to the ℓ^1 norm $\|x\|_1 = \sum_{k=1}^N |x_k|$ gives

$$\partial\|\cdot\|_1(x) = \prod_{k=1}^N \partial|\cdot|(x_k)$$

a hyperrectangle. With $I = \operatorname{supp}(x)$, the condition $u \in \partial\|\cdot\|_1(x)$ is equivalent to

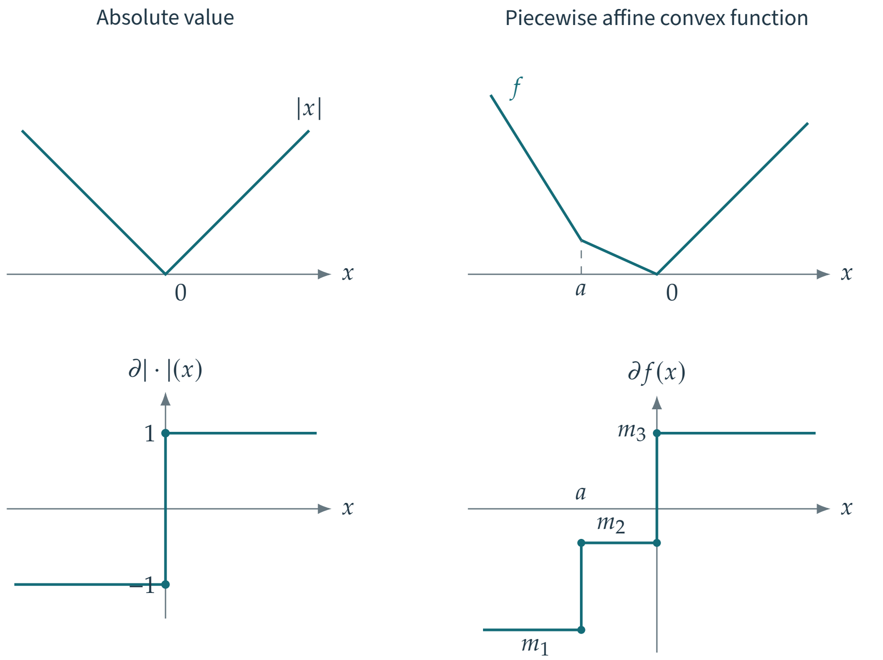
$$u_I = \operatorname{sign}(x_I) \quad \text{and} \quad \|u_{I^c}\|_\infty \leqslant 1.$$

A sum rule requires a domain qualification. If one function is finite and continuous at a point in the domain of the other, then for every $x \in \operatorname{dom}(f) \cap \operatorname{dom}(g)$,

$$\partial(f + g)(x) = \partial f(x) \oplus \partial g(x) = \{u + v \;;\; (u, v) \in \partial f(x) \times \partial g(x)\}$$

where $\oplus$ denotes the Minkowski sum. In particular, if f is differentiable at x, then

$$\partial(f + g)(x) = \nabla f(x) + \partial g(x) = \{\nabla f(x) + v \;;\; v \in \partial g(x)\}.$$



At each corner, the subdifferential contains every slope between the two one-sided slopes.

Figure 18.3. Subdifferential of the absolute value and a piecewise affine convex function.

Positive scaling gives

$$\forall\,\lambda > 0, \quad \partial(\lambda f)(x) = \lambda(\partial f(x)).$$

General compositions need not preserve convexity, so no unrestricted chain rule is available. Composition with a linear map does preserve convexity. For $A \in \mathbb{R}^{P\times N}$ and $f \in \Gamma_0(\mathbb{R}^P)$, assume $\mathrm{Im}(A) \cap \mathrm{relint}(\mathrm{dom}(f)) \neq \emptyset$. Then $f \circ A \in \Gamma_0(\mathbb{R}^N)$ and

$$\partial(f \circ A)(x) = A^*(\partial f)(Ax) \overset{\text{def.}}{=} \{A^*u \;;\; u \in \partial f(Ax)\}\,.$$

Remark 18.4 (Application to the Lasso). For $F(x) = \frac{1}{2}\|Ax - y\|^2 + \lambda\|x\|_1$ with $\lambda > 0$, the sum rule gives

$$0 \in A^*(Ax - y) + \lambda\partial\|{\cdot}\|_1(x).$$

Writing $r = y - Ax$, optimality is equivalent to $(A^*r)_i = \lambda\,\mathrm{sign}(x_i)$ on $\mathrm{supp}(x)$ and $|(A^*r)_i| \leqslant \lambda$ elsewhere. These conditions are useful both for analysis and for checking a numerical solution.

Normal cone. For a nonempty closed convex set C, the subdifferential of its indicator is the normal cone:

$$\forall\, x \in C, \quad \partial\iota_C(x) = \mathcal{N}_C(x) \overset{\text{def.}}{=} \{v \;;\; \forall\, z \in C, \langle z - x,\, v\rangle \leqslant 0\}\,.$$

Outside the set, $x \notin C$ implies $\partial\iota_C(x) = \emptyset$. For an affine space $C = a + \mathcal{V}$ with direction space $\mathcal{V} \subset \mathcal{H}$, the normal cone at each of its points is $\mathcal{N}_C(x) = \mathcal{V}^\perp$, the orthogonal complement of $\mathcal{V}$. More generally, at an interior point $x \in \mathrm{int}(C)$ of C, we have $\mathcal{N}_C(x) = \{0\}$. At a boundary point, the normal cone describes the supporting directions of C at x.

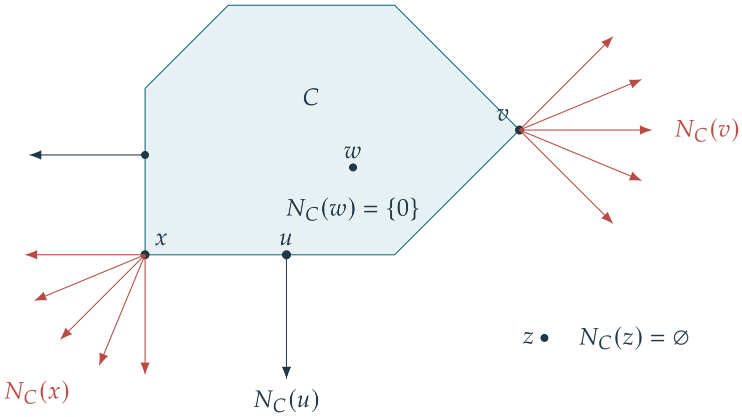
The normal cone expresses first-order optimality for the constrained problem

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

If f is finite and continuous at a point of C, the sum rule gives the optimality condition

$$0 \in \partial f(x) + \partial\iota_C(x) \quad \Leftrightarrow \quad \exists\xi \in \partial f(x), -\xi \in \mathcal{N}_C(x) \quad \Leftrightarrow \quad \partial f(x) \cap (-\mathcal{N}_C(x)) \neq \emptyset.$$

If f is differentiable, it reads $-\nabla f(x) \in \mathcal{N}_C(x)$.



$$N_C(x) = \{v : \langle z - x, v\rangle \le 0 \text{ for all } z \in C\}, \quad x \in C$$

Figure 18.4. Normal cones

18.2 Legendre–Fenchel Transform

The Legendre–Fenchel transform provides a dual representation of a convex function. It makes many Lagrange-duality calculations systematic by expressing inner minimizations through conjugates.

18.2.1 Legendre Transform

For $f \in \Gamma_0(\mathcal{H})$, we define its Legendre–Fenchel transform, or convex conjugate, by

$$f^*(u) \overset{\text{def.}}{=} \sup_x \langle x, u\rangle - f(x). \tag{18.3}$$

As a supremum of continuous affine functions, f^* is convex and lower semicontinuous; in fact, $f^\star \in \Gamma_0(\mathcal{H}^*)$. The following biconjugacy theorem recovers the original function.

Theorem 18.5. *One has*

$$\forall f \in \Gamma_0(\mathcal{H}), \quad (f^*)^* = f.$$

Even when f is not convex, f^* is convex. If f admits an affine minorant, f^{**} is its lower semicontinuous convex envelope: the largest lower semicontinuous convex function below f.

The subdifferentials of f and f^* satisfy the following inverse relation.

Proposition 18.6. *We have $\partial f^* = (\partial f)^{-1}$, where the inverse of a set-valued map is defined in (19.15). The Fenchel–Young inequality states that*

$$\forall (x, y), \quad \langle x, y\rangle \leqslant f(x) + f^*(y).$$

Equality is characterized by

$$\langle x, y\rangle = f(x) + f^*(y) \quad \Leftrightarrow \quad x \in \partial f^*(y) \quad \Leftrightarrow \quad y \in \partial f(x).$$

Proposition 18.7. *For $1 \leqslant p, q \leqslant \infty$ with $1/p + 1/q = 1$,*

$$(\iota_{\|\cdot\|_p \leqslant 1})^* = \|\cdot\|_q \quad \text{and} \quad (\|\cdot\|_q)^* = \iota_{\|\cdot\|_p \leqslant 1}$$

Here are useful examples and transformation rules.

Proposition 18.8. *If A is symmetric positive definite and $f(x) = \frac{1}{2}\langle Ax, x\rangle - \langle b, x\rangle$, then*

$$f^*(u) = \frac{1}{2}\langle A^{-1}(u + b), u + b\rangle.$$

In particular, $f = \|\cdot\|^2/2$ satisfies $f^* = f$. For $\lambda > 0$,

$$[f(\cdot - z)]^* = f^* + \langle z, \cdot\rangle, \qquad [f + \langle z, \cdot\rangle]^* = f^*(\cdot - z), \qquad (\lambda f)^* = \lambda f^*(\cdot/\lambda).$$

Proof. The maximizer in $f^*(u) = \sup_x\{\langle u + b, x\rangle - \frac{1}{2}\langle Ax, x\rangle\}$ satisfies $Ax = u + b$. Substitution of $x = A^{-1}(u + b)$ gives the formula. The translation and scaling identities follow directly by changing variables in (18.3). □

18.2.2 Legendre Transform and Smoothness

The Legendre–Fenchel transform exchanges smoothness and strong convexity. It inverts subdifferentials; when f is twice continuously differentiable with positive definite Hessian and $y = \nabla f(x)$, the inverse function theorem yields

$$\partial^2 f^*(y) = (\partial^2 f(x))^{-1}.$$

The next statement gives a version of this correspondence that does not require second derivatives.

Proposition 18.9. *For $f \in \Gamma_0(\mathcal{H})$ and $L > 0$,*

$$f \text{ is differentiable on } \mathcal{H} \text{ with L-Lipschitz gradient} \iff f^* \text{ is } 1/L\text{-strongly convex}.$$

Here strong convexity means that $f^* - \|\cdot\|^2/(2L)$ is convex, including when f^* is extended-valued.

This suggests a useful smoothing operation. The infimal convolution is

$$(f \otimes g)(x) \overset{\text{def.}}{=} \inf_{y+y'=x} \{f(y) + g(y')\}.$$

For proper convex functions, whenever this infimum defines a proper function, it is convex and satisfies

$$(f \otimes g)^* = f^* + g^*.$$

Moreover, $(f+g)^* = f^* \otimes g^*$ under the qualification $\mathrm{relint}(\mathrm{dom}(f)) \cap \mathrm{relint}(\mathrm{dom}(g)) \neq \emptyset$; without a qualification, a lower semicontinuous closure can be necessary.

For $\mu > 0$, the Moreau–Yosida regularization of $f \in \Gamma_0(\mathcal{H})$ is

$$f_\mu \overset{\text{def.}}{=} f \otimes \left(\frac{1}{2\mu}\|\cdot\|^2\right) = \left(f^* + \frac{\mu}{2}\|\cdot\|^2\right)^*. \tag{18.4}$$

The function inside the conjugate is μ-strongly convex, so f_μ has a $1/\mu$-Lipschitz gradient.

For the absolute value, the Moreau–Yosida regularization is

$$(|\cdot|_\mu)(x) = \begin{cases} \frac{1}{2\mu}x^2 & \text{if} \quad |x| \leqslant \mu, \\ |x| - \frac{\mu}{2} & \text{if} \quad |x| > \mu. \end{cases}$$

Compare this with the earlier regularization $\sqrt{x^2+\mu^2}$, which lies above the absolute value.

18.3 Convex Duality

A dual problem depends on the chosen primal formulation. A change of variables or a different representation of the constraints can produce a different dual problem.

18.3.1 Lagrange Duality: Affine Constraints

Begin with affine equality constraints:

$$\min_{Ax=y} f(x)$$

where $A \in \mathbb{R}^{P\times p}$, $y \in \mathbb{R}^P$, and $f : \mathbb{R}^p \to \mathbb{R}$ is convex and finite everywhere, hence continuous.

Assume that the affine constraint is feasible. Introducing $u \in \mathbb{R}^P$ gives the Lagrangian $\mathcal{L}(x,u) = f(x) + \langle Ax - y,\, u\rangle$. Minimization over x yields

$$F(u) = -f^*(-A^*u) - \langle y,\, u\rangle.$$

Because f is finite and continuous everywhere, feasibility is a sufficient qualification for strong duality:

$$\inf_{Ax=y} f(x) = \sup_{u\in\mathbb{R}^P} \{-f^*(-A^*u) - \langle y,\, u\rangle\}.$$

A primal-dual optimal pair satisfies $Ax^\star = y$ and $-A^*u^\star \in \partial f(x^\star)$.

18.3.2 Lagrange Duality: General Case

We consider a minimization of the form

$$p^\star = \inf_{x \in \mathbb{R}^N} \{f(x)\,;\; Ax = y \quad \text{and} \quad g(x) \leqslant 0\} \tag{18.5}$$

for a continuous convex function $f : \mathcal{H} \to \mathbb{R}$, a matrix $A \in \mathbb{R}^{P \times N}$ and a function $g : \mathcal{H} \to \mathbb{R}^Q$ such that each coordinate $g_i : \mathcal{H} \to \mathbb{R}$ is continuous and convex. In the standard formulation of a convex program, equality constraints are affine. Problems with convex inequality constraints can be written in the form (18.5). The representation is not unique, and different formulations lead to different dual problems.

For simplicity, f is assumed finite and continuous on its full domain $\mathrm{dom}(f) = \mathbb{R}^N$. The theory extends to more general $\mathrm{dom}(f)$, with additional domain qualifications. Here all constraints are expressed through $Ax = y$ and $g(x) \leqslant 0$.

Conic duality extends this construction from nonnegativity $x \geqslant 0$ to positive semidefiniteness $X \succeq 0$ of a matrix X, and more generally to membership in a convex cone.

We use the following fact

$$\sup_{u \in \mathbb{R}^P} \langle r,\, u \rangle = \begin{cases} 0 & \text{if} \quad r = 0, \\ +\infty & \text{if} \quad r \neq 0, \end{cases} \qquad \text{and} \qquad \sup_{v \in \mathbb{R}^Q_+} \langle s,\, v \rangle = \begin{cases} 0 & \text{if} \quad s \leqslant 0, \\ +\infty & \text{otherwise}, \end{cases}$$

to encode the constraints $r = Ax - y = 0$ and $s = g(x) \leqslant 0$.

Define the Lagrangian by

$$\mathcal{L}(x, u, v) \stackrel{\text{def.}}{=} f(x) + \langle Ax - y,\, u \rangle + \langle g(x),\, v \rangle.$$

The primal value in (18.5) is therefore

$$p^\star = \inf_x \sup_{u \in \mathbb{R}^P, v \in \mathbb{R}^Q_+} \mathcal{L}(x, u, v).$$

The dual objective is $F(u, v) \stackrel{\text{def.}}{=} \inf_x \mathcal{L}(x, u, v)$. Exchanging the order of optimization gives the dual value

$$d^\star = \sup_{(u,v) \in \mathbb{R}^P \times \mathbb{R}^Q_+} F(u, v). \tag{18.6}$$

The dual objective F is concave because it is an infimum of affine functions. The dual problem therefore maximizes a concave objective.

Weak duality states that every feasible dual value is a lower bound on the primal optimum.

Proposition 18.10. *One always has, for all* $(u, v) \in \mathbb{R}^P \times \mathbb{R}^Q_+$,

$$F(u, v) \leqslant p^\star \quad \Longrightarrow \quad d^\star \leqslant p^\star.$$

Proof. Since $g(x) \leqslant 0$ and $v \geqslant 0$, one has $\langle g(x),\, v \rangle \leqslant 0$, and since $Ax = y$, one has $\langle Ax - y,\, u \rangle = 0$, so that

$$F(u, v) \leqslant \mathcal{L}(x, u, v) \leqslant f(x) \quad \text{for every feasible } x.$$

Taking the infimum over feasible x gives $F(u, v) \leqslant p^\star$, and taking the supremum over feasible multipliers yields $d^\star \leqslant p^\star$. □

The next theorem gives a constraint qualification that guarantees equality of the primal and dual values.

Theorem 18.11. *If*

$$\exists x_0 \in \mathbb{R}^N, \quad Ax_0 = y \quad \text{and} \quad g(x_0) < 0, \tag{18.7}$$

then $p^\star = d^\star$. *If* $p^\star$ *is finite, the dual supremum is attained. Furthermore,* $x^\star$ *and* $(u^\star, v^\star)$ *solve* (18.5) *and* (18.6), *respectively, if and only if*

$$Ax^\star = y, \quad g(x^\star) \leqslant 0, \quad v^\star \geqslant 0 \tag{18.8}$$

$$0 \in \partial f(x^\star) + A^* u^\star + \sum_i v_i^\star \partial g_i(x^\star) \tag{18.9}$$

$$\forall\, i, \quad v_i^\star g_i(x^\star) = 0 \tag{18.10}$$

The existence of x_0 is Slater's constraint qualification; the strict inequality $g(x_0) < 0$ holds componentwise. Other, weaker qualifications are possible.

Condition (18.8) states primal and dual feasibility. Condition (18.9) expresses optimality of $\mathcal{L}(x, u, v)$ with respect to x. Together with feasibility, condition (18.10) expresses maximality of $\mathcal{L}(x^\star, u, v)$ over the admissible multipliers (u, v). Together these are the Karush–Kuhn–Tucker (KKT) conditions. Under the stated qualification, they are necessary and sufficient for primal-dual optimality.

Complementary slackness, $v_i^\star g_i(x^\star) = 0$, relates a multiplier to its constraint: if $g_i(x^\star) < 0$, the constraint is inactive and $v_i^\star = 0$; if $v_i^\star > 0$, the constraint must be active, $g_i(x^\star) = 0$.

For an affine inequality $g_i(x) = \langle x,\, h_i\rangle - c_i \leqslant 0$, strict feasibility $g_i(x_0) < 0$ may be replaced by ordinary feasibility $\langle x_0,\, h_i\rangle \leqslant c_i$ in the qualification.

If $\mathrm{dom}(f)$ is a proper subset of $\mathbb{R}^N$, the objective itself encodes constraints in addition to those written with $\leqslant$. The qualification then also requires $x_0 \in \mathrm{relint}(\mathrm{dom}(f))$.

Theorem 18.11 extends the Lagrange-multiplier conditions to convex problems with inequalities. Inequality constraints introduce nonnegative multipliers v. Convexity ensures that the KKT conditions are sufficient, and the stated constraint qualification ensures their necessity.

As an example, consider projection onto a nonempty affine space $\{x : Ax = y\}$:

$$p^\star = \min_{Ax=y} \frac{1}{2}\|x - z\|^2 = \max_u F(u), \qquad F(u) \overset{\text{def.}}{=} \min_x \left\{ \frac{1}{2}\|x - z\|^2 + \langle Ax - y,\, u\rangle \right\}.$$

The minimizing x satisfies $x - z + A^* u = 0$, hence $x = z - A^* u$. Therefore

$$F(u) = -\frac{1}{2}\|A^* u\|^2 + \langle u,\, Az - y\rangle.$$

An optimal multiplier solves $AA^* u^\star = Az - y$. Since $y \in \mathrm{Im}(A)$, a solution is $u^\star = (AA^*)^+(Az - y)$, where $^+$ denotes the Moore–Penrose inverse. If A has full row rank, $(AA^*)^+ = (AA^*)^{-1}$. The unique projection is

$$x^\star = \mathrm{Proj}_{A\cdot=y}(z) = z - A^+(Az - y) = (\mathrm{Id} - A^+ A)z + A^+ y. \tag{18.11}$$

A further example, the Lasso, is treated in Section 11.1.5.

18.3.3 Fenchel–Rockafellar Duality

Conjugates often give explicit expressions for the Lagrange dual of an objective f. Fenchel–Rockafellar duality is a particularly useful example.

We consider the following structured minimization problem

$$p^\star = \inf_x f(x) + g(Ax). \tag{18.12}$$

Introducing an auxiliary variable gives

$$\inf_{y=Ax} f(x) + g(y).$$

We can then form the primal-dual problem

$$\inf_{(x,y)} \sup_u f(x) + g(y) + \langle Ax - y,\, u\rangle. \tag{18.13}$$

Under the domain qualification in Theorem 18.12, one can exchange the infimum and supremum to obtain the dual problem

$$d^\star = \sup_u \inf_{(x,y)} f(x) + g(y) + \langle Ax - y,\, u\rangle \tag{18.14}$$

$$= \sup_u \left\{ \left(\inf_x \langle x,\, A^* u\rangle + f(x) \right) + \left(\inf_y \, -\langle y,\, u\rangle + g(y) \right) \right\}. \tag{18.15}$$

This gives the Fenchel–Rockafellar dual problem. The following theorem states a sufficient domain qualification for strong duality.

Theorem 18.12 (Fenchel–Rockafellar). *Let $f \in \Gamma_0(\mathbb{R}^N)$, $g \in \Gamma_0(\mathbb{R}^P)$, and $A \in \mathbb{R}^{P\times N}$. If*

$$0 \in \operatorname{relint}(\operatorname{dom}(g)) - A \operatorname{relint}(\operatorname{dom}(f)) \tag{18.16}$$

then strong duality holds

$$\inf_x f(x) + g(Ax) = \inf_x \sup_u \mathcal{L}(x,u) = \sup_u \inf_x \mathcal{L}(x,u) = \sup_u \; -f^*(-A^*u) - g^*(u) \tag{18.17}$$

$$\textit{where} \quad \mathcal{L}(x,u) \overset{\text{def.}}{=} f(x) + \langle Ax,\, u\rangle - g^*(u).$$

A pair $(x^\star, u^\star)$ is primal-dual optimal if and only if

$$-A^*u^\star \in \partial f(x^\star) \quad \textit{and} \quad Ax^\star \in \partial g^*(u^\star). \tag{18.18}$$

Condition (18.16) is the constraint qualification ensuring that one can exchange the infimum and supremum in (18.17). It is the relative-interior qualification corresponding to the equality-constrained reformulation (18.14). The relations (18.18) express first-order optimality in the primal variable x and dual variable u for the saddle function $\mathcal{L}$. They can be collected in matrix notation as

$$0 \in \begin{pmatrix} \partial f & A^* \\ -A & \partial g^* \end{pmatrix} \begin{pmatrix} x^\star \\ u^\star \end{pmatrix}.$$

References

[1] Stephen Boyd and Lieven Vandenberghe. *Convex optimization*. Cambridge university press, 2004.

[2] Philippe G Ciarlet. *Introduction à l'analyse numérique matricielle et à l'optimisation*. Masson, 1982.