Chapter 2
Elements of Convex Analysis

(COSS 2018 Reading Material)

Kazuo Murota (July 14, 2018)

Minimal technical elements from convex analysis are given in this section. For comprehensive account, the reader is referred to books on convex analysis [1, 2, 3, 5, 6, 7, 8, 9, 10].

2.1 Convex Sets

For two vectors $a = (a_1, a_2, \ldots, a_n), b = (b_1, b_2, \ldots, b_n) \in (\mathbb{R} \cup \{-\infty, +\infty\})^n$ we define *closed interval* $[a, b]$ and *open interval* (a, b) as

$$[a,b] = [a,b]_{\mathbb{R}} = \{x \in \mathbb{R}^n \mid a_i \le x_i \le b_i \ (i = 1, 2, \ldots, n)\}, \tag{2.1}$$

$$(a,b) = (a,b)_{\mathbb{R}} = \{x \in \mathbb{R}^n \mid a_i < x_i < b_i \ (i = 1, 2, \ldots, n)\}, \tag{2.2}$$

where, if $a_i = -\infty$, for example, $a_i \le x_i$ is to be understood as $-\infty < x_i$.

A set $S \subseteq \mathbb{R}^n$ is called *convex* if it satisfies the condition

$$x, y \in S, \ \ 0 \le \lambda \le 1 \implies \lambda x + (1 - \lambda) y \in S, \tag{2.3}$$

where an empty set is a convex set. A *convex polyhedron* is a convex set S described by a finite number of linear inequalities as

$$S = \{x \in \mathbb{R}^n \mid \sum_{j=1}^{n} a_{ij} x_j \le b_i \ (i = 1, 2, \ldots, m)\}, \tag{2.4}$$

where $a_{ij} \in \mathbb{R}$ and $b_i \in \mathbb{R}$ $(i = 1, 2, \ldots, m; j = 1, 2, \ldots, n)$.

For a finite number of points $x^1, x^2, \ldots, x^m$ in a set S, a point represented as

$$\lambda_1 x^1 + \lambda_2 x^2 + \cdots + \lambda_m x^m \tag{2.5}$$

with nonnegative coefficients λ_i $(1 \leq i \leq m)$ having unit sum $(\sum_{i=1}^{m} \lambda_i = 1)$ is called a *convex combination* of those points. If S is convex, any convex combination of points in S belongs to S, and the converse is also true. Therefore, S is convex if and only if $S = \overline{S}$, where $\overline{S}$ denotes the set of all possible convex combinations of a finite number of points of S.

The intersection of any (finite or infinite) number of convex sets is a convex set. For any set S, the intersection of all the convex sets containing S is the smallest convex set containing S, which is called the *convex hull* of S and denoted as $\mathrm{conv}(S)$. The convex hull of S coincides with the set of all convex combinations of points in S. That is, we have $\mathrm{conv}(S) = \overline{S}$. The convex hull of a set S is not necessarily closed (in the topological sense). The smallest closed convex set containing S is called the *closed convex hull* of S. For a finite set S, the convex hull is always closed.

The *affine hull* of a set S is defined to be the smallest affine set (a translation of a linear space) containing S, and is denoted by $\mathrm{aff}\, S$. The *relative interior* of S, denoted as $\mathrm{ri}\, S$, is the set of points $x \in S$ such that $\{y \in \mathbb{R}^n \mid \|y - x\| < \varepsilon\} \cap \mathrm{aff}\, S$ is contained in S for some $\varepsilon > 0$. In other words, the relative interior of S is the set of the interior points of S with respect to the topology induced from $\mathrm{aff}\, S$.

For two sets S and T, the set

$$S + T = \{x + y \mid x \in S, y \in T\} \tag{2.6}$$

is called the *Minkowski sum* of S and T. If S and T are convex, the Minkowski sum $S + T$ is a convex set.

A set S is a *cone* if it satisfies

$$x \in S, \quad \lambda > 0 \implies \lambda x \in S. \tag{2.7}$$

A cone that is convex is called a *convex cone*. In other words, a set S is a convex cone if and only if it satisfies the condition

$$x, y \in S, \quad \lambda, \mu > 0 \implies \lambda x + \mu y \in S. \tag{2.8}$$

2.2 Convex Functions

For a function $f : \mathbb{R}^n \to \mathbb{R} \cup \{-\infty, +\infty\}$ we define

$$\mathrm{dom}\, f = \mathrm{dom}_{\mathbb{R}} f = \{x \in \mathbb{R}^n \mid -\infty < f(x) < +\infty\}, \tag{2.9}$$

which is called the *effective domain* of f.

A function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ is said to be *convex* if it satisfies

$$\lambda f(x) + (1 - \lambda) f(y) \geq f(\lambda x + (1 - \lambda) y) \qquad (x, y \in \mathbb{R}^n; 0 \leq \lambda \leq 1). \tag{2.10}$$

Note that $-\infty$ is excluded from the possible function values of a convex function, and that the inequality (2.10) is satisfied, by convention, if both sides are equal to $+\infty$. A convex function having a nonempty effective domain is called a *proper convex function*. A function is *strictly convex* if it satisfies (2.10) with strict inequalities, i.e., if

$$\lambda f(x) + (1-\lambda) f(y) > f(\lambda x + (1-\lambda) y) \qquad (x, y \in \mathrm{dom} f; 0 < \lambda < 1). \quad (2.11)$$

A function $g : \mathbb{R}^n \to \underline{\mathbb{R}}$ is *concave* if $-g$ is convex, that is, if

$$\lambda g(x) + (1-\lambda) g(y) \le g(\lambda x + (1-\lambda) y) \qquad (x, y \in \mathbb{R}^n; 0 \le \lambda \le 1). \quad (2.12)$$

The *epigraph* of a function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$, denoted as $\mathrm{epi}\, f$, is the set of points in $\mathbb{R}^n \times \mathbb{R}$ lying above the graph of $\alpha = f(x)$. Namely,

$$\mathrm{epi}\, f = \{(x, \alpha) \in \mathbb{R}^{n+1} \mid \alpha \ge f(x)\}. \quad (2.13)$$

Then we have

$$f \text{ is a convex function} \iff \mathrm{epi}\, f \text{ is a convex set.} \quad (2.14)$$

A function f is said to be *closed convex* if $\mathrm{epi}\, f$ is a closed convex set in $\mathbb{R}^{n+1}$.

The *indicator function* of a set $S \subseteq \mathbb{R}^n$ is a function $\delta_S : \mathbb{R}^n \to \{0, +\infty\}$ defined by

$$\delta_S(x) = \begin{cases} 0 & (x \in S), \\ +\infty & (x \notin S). \end{cases} \quad (2.15)$$

Then we have

$$S \text{ is a convex set} \iff \delta_S \text{ is a convex function.} \quad (2.16)$$

For a family of convex functions $\{f_k \mid k \in K\}$, indexed by K, the pointwise maximum of those functions, $f(x) = \sup\{f_k(x) \mid k \in K\}$, is again a convex function, where the index set K here may possibly be infinite. In particular, the maximum of a finite or infinite number of affine functions

$$f(x) = \sup\{\alpha_k + \langle p_k, x\rangle \mid k \in K\} \quad (2.17)$$

is a convex function, where $\alpha_k \in \mathbb{R}$ and $p_k \in \mathbb{R}^n$ for $k \in K$ and

$$\langle p, x\rangle = \sum_{i=1}^{n} p_i x_i \quad (2.18)$$

denotes the *inner product* of $p = (p_1, p_2, \ldots, p_n)$ and $x = (x_1, x_2, \ldots, x_n)$.

A function defined on $\mathbb{R}^n$ is said to be *polyhedral convex* if its epigraph is a convex polyhedron in $\mathbb{R}^{n+1}$. A polyhedral convex function is exactly such a function that can be represented as the maximum of a finite number of affine functions (i.e., (2.17) with finite K) on an effective domain represented as (2.4).

For two functions $f,g:\mathbb{R}^n\to\overline{\mathbb{R}}$, their *sum* is the function $f+g:\mathbb{R}^n\to\overline{\mathbb{R}}$ defined naturally by

$$(f+g)(x)=f(x)+g(x)\qquad(x\in\mathbb{R}^n),\tag{2.19}$$

and their *infimal convolution* is the function $f\,\Box\,g:\mathbb{R}^n\to\mathbb{R}\cup\{-\infty,+\infty\}$ defined by

$$(f\,\Box\,g)(x)=\inf\{f(y)+g(z)\mid x=y+z,\ \ y,z\in\mathbb{R}^n\}\qquad(x\in\mathbb{R}^n).\tag{2.20}$$

The sum of two convex functions is convex, and the infimal convolution of two convex functions is convex if it does not take the value of $-\infty$. If f and g are the indicator functions of sets S and T, then $f+g$ and $f\,\Box\,g$ are the indicator functions of the intersection $S\cap T$ and the Minkowski sum $S+T$, respectively.

For a function f and a vector p, we denote by $f[-p]$ the function defined by

$$f[-p](x)=f(x)-\langle p,x\rangle\qquad(x\in\mathbb{R}^n).\tag{2.21}$$

This is convex for a convex function f.

2.3 Minimization and Subgradients

The most appealing property of a convex function is that local minimality is equivalent to global minimality. A point (or vector) x is said to be a (*global*) *minimizer* of f if the inequality

$$f(x)\le f(y)\tag{2.22}$$

holds for every y. A point x is a *local minimizer* if the inequality (2.22) holds for every y in some neighborhood of x. Obviously, global minimality implies local minimality. The converse is not true in general, but it is true for convex functions.

Theorem 2.1. *For a convex function, local minimality implies global minimality.*

Proof. Let x be a local minimizer of a convex function f. Then we have $f(z)\ge f(x)$ for all z in some neighborhood U of x. For any y, we can choose $\lambda<1$ sufficiently close to 1 such that $z=\lambda x+(1-\lambda)y$ belongs to U. Then it follows from (2.10) that

$$\lambda f(x)+(1-\lambda)f(y)\ge f(\lambda x+(1-\lambda)y)=f(z)\ge f(x).$$

This implies $f(y)\ge f(x)$. □

The set of the minimizers of f is denoted as

$$\operatorname{argmin} f=\operatorname{argmin}_{\mathbb{R}} f=\{x\in\mathbb{R}^n\mid f(x)\le f(y)\ (\forall y\in\mathbb{R}^n)\}.\tag{2.23}$$

This is a convex set if f is convex.

The *subdifferential* of a function f at a point $x\in\operatorname{dom} f$ is defined to be the set

$$\partial f(x)=\{p\in\mathbb{R}^n\mid f(y)-f(x)\ge\langle p,y-x\rangle\ (\forall y\in\mathbb{R}^n)\}.\tag{2.24}$$

Note that $p \in \partial f(x)$ if and only if $x \in \operatorname{argmin} f[-p]$; in particular, $\mathbf{0} \in \partial f(x)$ if and only if $x \in \operatorname{argmin} f$. For a convex function f, the set $\partial f(x)$ is nonempty if x is in the relative interior of $\operatorname{dom} f$. An element of $\partial f(x)$ is called a *subgradient* of f at x. If f is convex and differentiable at x, the subdifferential $\partial f(x)$ consists of a single element, which is the *gradient* of f at x:

$$\nabla f = \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \ldots, \frac{\partial f}{\partial x_n} \right). \tag{2.25}$$

The *directional derivative* of a function f at a point $x \in \operatorname{dom} f$ in a direction $d \in \mathbb{R}^n$ is defined by

$$f'(x;d) = \lim_{\alpha \downarrow 0} \frac{f(x+\alpha d) - f(x)}{\alpha} \tag{2.26}$$

when this limit (finite or infinite) exists, where $\alpha \downarrow 0$ means that α tends to 0 from the positive side ($\alpha > 0$). For a convex function f, the limit exists for all d, and $f'(x;d)$ is a convex function in d. For a polyhedral convex function f, there exists $\varepsilon > 0$, independent of $x \in \operatorname{dom} f$, such that

$$f'(x;d) = f(x+d) - f(x) \qquad (\|d\| \le \varepsilon). \tag{2.27}$$

2.4 Conjugacy

As Fig. 2.1 (a,b) shows, a convex function $f(x)$ can be recovered from tangent lines as the upper envelope of all tangent lines with different slopes. Let α be the vertical intercept of the tangent line with slope p. Since α is dependent on slope p, we denote $\alpha = -f^\bullet(p)$. By considering the minimum distance between the graph of $y = f(x)$ and the line $y = px$, we see that the minimum of $f(x) - px$ over all x is equal to $\alpha = -f^\bullet(p)$; cf., Fig. 2.1(c). That is,

$$f^\bullet(p) = \sup\{px - f(x) \mid x \in \mathbb{R}\} \qquad (p \in \mathbb{R}). \tag{2.28}$$

This function $f^\bullet(p)$ should be equivalent to the original function $f(x)$ in some appropriate sense.

For a function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ with $\operatorname{dom} f \neq \emptyset$, the *convex conjugate* (or simply *conjugate*) of f is a function $f^\bullet : \mathbb{R}^n \to \overline{\mathbb{R}}$ defined by

$$f^\bullet(p) = \sup\{\langle p,x\rangle - f(x) \mid x \in \mathbb{R}^n\} \qquad (p \in \mathbb{R}^n), \tag{2.29}$$

which is indeed a convex function since it is the maximum of (infinitely many) affine functions in p indexed by x. The function $f^\bullet$ is also called the (convex) *Legendre–Fenchel transform* of f, and the mapping $f \mapsto f^\bullet$ is referred to as the (convex) *Legendre–Fenchel transformation*. Similarly, the *concave conjugate* of a function $g : \mathbb{R}^n \to \underline{\mathbb{R}}$ with $\operatorname{dom} g \neq \emptyset$ is a concave function $g^\circ : \mathbb{R}^n \to \underline{\mathbb{R}}$ defined by

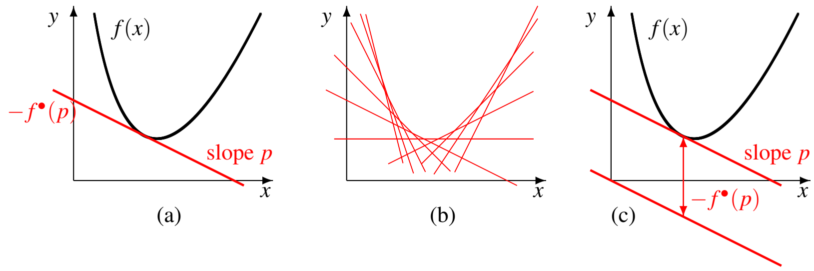


Fig. 2.1 Tangent lines of a convex function

$$g^\circ(p) = \inf\{\langle p,x\rangle - g(x) \mid x \in \mathbb{R}^n\} \qquad (p \in \mathbb{R}^n). \tag{2.30}$$

Note that $g^\circ(p) = -(-g)^\bullet(-p)$.

For a function f, the conjugate function of the conjugate of f, i.e., $(f^\bullet)^\bullet$, is called the *biconjugate* of f and denoted as $f^{\bullet\bullet}$. The biconjugate of f is the largest closed convex function that is dominated pointwise by f.

Theorem 2.2. *The Legendre–Fenchel transform $f^\bullet$ in* (2.29) *is a closed proper convex function for any function f with* $\mathrm{dom}\, f \neq \emptyset$*, and $f^{\bullet\bullet} = f$ for a closed proper convex function f.* ∎

This theorem shows that the Legendre–Fenchel transformation $f \mapsto f^\bullet$ gives a symmetric (or involutive) one-to-one correspondence in the class of all closed proper convex functions.

For a set $S \subseteq \mathbb{R}^n$, the conjugate $\delta_S{}^\bullet$ of its indicator function δ_S is expressed as

$$\delta_S{}^\bullet(p) = \sup\{\langle p,x\rangle \mid x \in S\} \qquad (p \in \mathbb{R}^n), \tag{2.31}$$

which is called the *support function* of S. The biconjugate $\delta_S{}^{\bullet\bullet}$ of the indicator function δ_S of a set S is the indicator function of the closed convex hull of S.

By Theorem 2.2 and the definition (2.24) we obtain the relationship

$$\boxed{\begin{array}{rcl} p \in \partial f(x) & \Longleftrightarrow & x \in \mathrm{argmin}\, f[-p] \\ & & \Updownarrow \\ & & f(x) + f^\bullet(p) = \langle p,x\rangle \\ & & \Updownarrow \\ x \in \partial f^\bullet(p) & \Longleftrightarrow & p \in \mathrm{argmin}\, f^\bullet[-x] \end{array}} \tag{2.32}$$

for a closed proper convex function f and vectors $x, p \in \mathbb{R}^n$. For a closed convex function f and a point x in the relative interior of $\mathrm{dom}\, f$, the support function of the subdifferential $\partial f(x)$ coincides with the directional derivative $f'(x;d)$ as a function in d, i.e.,

$$(\delta_{\partial f(x)})^\bullet(d) = f'(x;d) \qquad (d \in \mathbb{R}^n). \tag{2.33}$$

The addition (2.19) and the infimal convolution (2.20) are conjugate operations with respect to the Legendre–Fenchel transformation. For proper convex functions f and g we have

$$(f \Box g)^{\bullet} = f^{\bullet} + g^{\bullet}, \tag{2.34}$$

$$(f + g)^{\bullet} = f^{\bullet} \Box g^{\bullet}, \tag{2.35}$$

where the latter is true under the assumption that $\mathrm{ri}\,(\mathrm{dom}\, f) \cap \mathrm{ri}\,(\mathrm{dom}\, g) \neq \emptyset$.

Example 2.1. The conjugate of a quadratic function $f(x) = \frac{1}{2}x^{\top}Ax$ defined by a positive definite symmetric matrix A can be computed as follows. The maximizer x on the right-hand side of (2.29) is determined from $p = \nabla f(x) = Ax$ as $x = A^{-1}p$. Then

$$f^{\bullet}(p) = p^{\top}x - \frac{1}{2}x^{\top}Ax = \frac{1}{2}p^{\top}A^{-1}p.$$

Since $\nabla f(x) = Ax$ and $\nabla f^{\bullet}(p) = A^{-1}p$, we indeed have the equivalence "$p \in \partial f(x) \iff x \in \partial f^{\bullet}(p)$" in (2.32). ∎

2.5 Duality

The separation theorem for functions asserts that, if a convex function pointwise dominates a concave function, then there exists an affine function that lies between the convex function and the concave function; see Fig. 2.2 (a).

Theorem 2.3 (Separation for convex functions). *Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be a proper convex function and $g : \mathbb{R}^n \to \underline{\mathbb{R}}$ a proper concave function, and assume that* (a1) *or* (a2) *below is satisfied:*

(a1) $\mathrm{ri}\,(\mathrm{dom}\, f) \cap \mathrm{ri}\,(\mathrm{dom}\, g) \neq \emptyset$,

(a2) *f and g are polyhedral, and* $\mathrm{dom}\, f \cap \mathrm{dom}\, g \neq \emptyset$.

If $f(x) \geq g(x)$ ($\forall x \in \mathbb{R}^n$), there exist $\alpha^* \in \mathbb{R}$ and $p^* \in \mathbb{R}^n$ such that

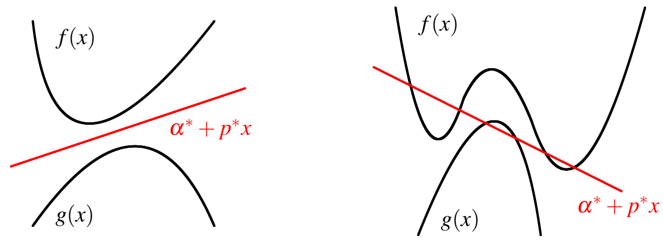
$$f(x) \geq \alpha^* + \langle p^*, x \rangle \geq g(x) \qquad (\forall x \in \mathbb{R}^n). \tag{2.36}$$

∎

Note that the convexity assumption is critical in Theorem 2.3. In Fig. 2.2 (b), we have $f(x) \geq g(x)$ for all x, but there exists no affine function $\alpha^* + p^*x$ that separates $f(x)$ and $g(x)$.

The *Fenchel duality* is a min-max relation between a pair of convex function f and concave function g and their conjugate functions $f^{\bullet}$ and g°.

Theorem 2.4 (Fenchel duality). *Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be a proper convex function and $g : \mathbb{R}^n \to \underline{\mathbb{R}}$ a proper concave function, and assume that at least one of the following four conditions* (a1)∼(b2) *below is satisfied:*



(a) Convex-concave pair (b) Nonconvex-concave pair

Fig. 2.2 Separation theorem

(a1) $\mathrm{ri}\,(\mathrm{dom}\, f) \cap \mathrm{ri}\,(\mathrm{dom}\, g) \neq \emptyset$,
(a2) *f and g are polyhedral, and* $\mathrm{dom}\, f \cap \mathrm{dom}\, g \neq \emptyset$,
(b1) *f and g are closed*[1], *and* $\mathrm{ri}\,(\mathrm{dom}\, f^{\bullet}) \cap \mathrm{ri}\,(\mathrm{dom}\, g^{\circ}) \neq \emptyset$,
(b2) *f and g are polyhedral, and* $\mathrm{dom}\, f^{\bullet} \cap \mathrm{dom}\, g^{\circ} \neq \emptyset$.
Then it holds that

$$\inf\{f(x) - g(x) \mid x \in \mathbb{R}^n\} = \sup\{g^{\circ}(p) - f^{\bullet}(p) \mid p \in \mathbb{R}^n\}. \tag{2.37}$$

Moreover, if this common value is finite, the supremum is attained by some $p \in$ $\mathrm{dom}\, f^{\bullet} \cap \mathrm{dom}\, g^{\circ}$ *under the assumption of* (a1) *or* (a2), *and the infimum is attained by some $x \in \mathrm{dom}\, f \cap \mathrm{dom}\, g$ under the assumption of* (b1) *or* (b2). ∎

If the supremum in (2.37) is attained by $p = p^*$, we have

$$\mathrm{argmin}\,(f - g) = \mathrm{argmin}\, f[-p^*] \cap \mathrm{argmax}\, g[-p^*]. \tag{2.38}$$

Remark 2.1. Theorem 2.4 above is formulated for a pair of convex and concave functions. In some cases it is convenient to reformulate it in terms of two convex functions. For convex functions f and g, the min-max formula (2.37) is rewritten as

$$\inf\{f(x) + g(x) \mid x \in \mathbb{R}^n\} = \sup\{-f^{\bullet}(p) - g^{\bullet}(-p) \mid p \in \mathbb{R}^n\}. \tag{2.39}$$

∎

References

1. Barvinok, A.: A Course in Convexity. American Mathematical Society, Providence, R.I. (2002)
2. Borwein, J.M., Lewis, A.S.: Convex Analysis and Nonlinear Optimization: Theory and Examples. Springer, Berlin (2000)
3. Boyd, S., Vandenberghe, L.: Convex Optimization. Cambridge University Press, Cambridge (2004)

[1] By this we mean that f and $-g$ are closed convex functions.

4. Ekeland, I., Témam, R.: Convex Analysis and Variational Problems. North-Holland, Amsterdam (1976); Society for Industrial and Applied Mathematics, Philadelphia (1999)
5. Hiriart-Urruty, J.-B., Lemaréchal, C.: Convex Analysis and Minimization Algorithms I, II. Springer, Berlin (1993)
6. Hiriart-Urruty, J.-B., Lemaréchal, C.: Fundamentals of Convex Analysis. Springer, Berlin (2001)
7. Krantz, S.G.: Convex Analysis. CRC Press, Boca Raton (2015)
8. Rockafellar, R.T.: Convex Analysis. Princeton University Press, Princeton (1970)
9. Rockafellar, R.T.: Conjugate Duality and Optimization. SIAM Regional Conference Series in Applied Mathematics, vol. 16, Society for Industrial and Applied Mathematics, Philadelphia (1974)
10. Rockafellar, R.T., Wets, R.J.-B.: Variational Analysis. Springer, Berlin (1998)