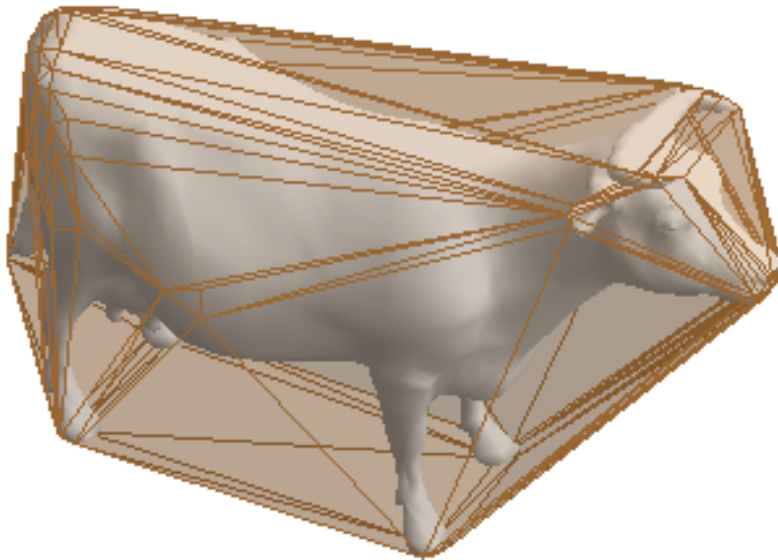


Alexandre St-Aubin

MATH 563 Convex Optimization

MARCH 13, 2024



Taught by Courtney Paquette

Contents

1	Chapter 1: Preliminaries	3
1.1	Euclidean Setting	3
1.2	Euclidean Matrix Spaces	3
1.3	Notions of Topology	4
1.4	Differentiation	4
1.5	Extended Arithmetic and Lower Semicontinuity	5
1.6	Optimization Problems	6
2	Chapter 2: Convex Sets	6
2.1	Definitions	6
2.2	Projections on Convex Sets	7
2.3	Separation Theorem	7
3	Chapter 3: Convex Functions	8
3.1	Definitions	8
3.2	Stronger Notions of Convexity	9
3.3	Functional Operations Preserving Convexity	10
4	Minimization and Convexity	10
4.1	General Existence Result	10
4.2	Convex Optimization	11
5	Infimal Convolution of Convex Functions	12
5.1	Moreau Envelopes and Proximal Mappings	13
6	Conjugacy of Convex Functions	14
6.1	Convex Hull of a Function	14
6.2	Fenchel Conjugate	15
7	Differential Theory	18
7.1	Convex Subdifferential	18
7.2	Directional Derivative of Convex Functions	20
7.3	Continuity Properties	24
7.4	Differentiable Convex Functions	24
7.5	Second Order Characterization	25
7.6	Subdifferential Calculus	25

1 Chapter 1: Preliminaries

1.1 Euclidean Setting

We denote by \mathbb{E} a n -dimensional Euclidian space equipped with a scalar product,

$$\langle \cdot, \cdot \rangle : \mathbb{E} \times \mathbb{E} \rightarrow \mathbb{R}$$

Definition 1.1 (Linear operators). If \mathbb{E}_1 and \mathbb{E}_2 are n and m dimensional Euclidian spaces, denote by $\mathcal{L}(\mathbb{E}_1, \mathbb{E}_2)$ the set of all linear operators from \mathbb{E}_1 to \mathbb{E}_2 .

Remark 1.2. For $L \in \mathcal{L}(\mathbb{E}_1, \mathbb{E}_2)$, there exists a unique mapping $L^* \in \mathcal{L}(\mathbb{E}_1, \mathbb{E}_2)$ such that

$$\langle Lx, y \rangle_{\mathbb{E}_2} = \langle x, L^*y \rangle_{\mathbb{E}_1} \quad \forall x \in \mathbb{E}_1, y \in \mathbb{E}_2$$

L^* is called the **Adjoint Mapping** of L .

Definition 1.3 (Orthogonal complement). For $\emptyset \neq S \in \mathbb{E}$, its orthogonal complement is

$$S^\perp = \{x \in \mathbb{E} \mid \langle s, x \rangle = 0 \ \forall s \in S\}$$

1.2 Euclidean Matrix Spaces

Matrix space of all real $m \times n$ matrices equipped with the standard scalar product

$$\langle A, B \rangle = \text{tr}(A^T B)$$

Definition 1.4 (Orthogonal Group). Denote the orthogonal group by

$$O(n) := \{A \in \mathbb{R}^{n \times n} \mid A^T A = I_n\}$$

Definition 1.5 (Symmetric Matrices). Let the space of symmetric matrices be defined as

$$\mathbb{S}^n := \{A \in \mathbb{R}^{n \times n} \mid A^T = A\}$$

Theorem 1.6 (Spectral Theorem). Let $A \in \mathbb{S}^n$, then there exists a $U \in O(n)$ such that $A = U^T \text{diag}(\underbrace{\lambda_1, \lambda_2, \dots, \lambda_n}_{\text{eigenvalues of } A})U$ and

$$\lambda_1, \lambda_2, \dots, \lambda_n \in \mathbb{R}.$$

Example 1.7 (Subsets of \mathbb{S}^n).

1. Positive semidefinite matrices,

$$\mathbb{S}_+^n := \{A \in \mathbb{S}^n \mid x^T A x \geq 0 \ \forall x \in \mathbb{R}^n\};$$

2. Positive definite matrices,

$$\mathbb{S}_{++}^n := \{A \in \mathbb{S}^n \mid x^T A x > 0 \forall x \in \mathbb{R}^n \setminus \{0\}\};$$

3. Negative semidefinite matrices,

$$\mathbb{S}_-^n := \{A \in \mathbb{S}^n \mid x^T A x \leq 0 \forall x \in \mathbb{R}^n\};$$

4. Negative definite matrices,

$$\mathbb{S}_{--}^n := \{A \in \mathbb{S}^n \mid x^T A x < 0 \forall x \in \mathbb{R}^n \setminus \{0\}\}.$$

1.3 Notions of Topology

Definition 1.8 (Affine set). We say that a set C is *affine* if for all $x, y \in C$ and every $\lambda \in \mathbb{R}$, $x + \lambda(y - x) \in C$.

Definition 1.9 (Linear set). We say that a set $U \subseteq \mathbb{E}$ is *linear* if for every pair of distinct points $x, y \in U$, the line \overleftrightarrow{xy} is contained in U ¹.

¹ Namely, $\forall x, y \in Q$ and $\lambda, \mu \in \mathbb{R}$, $\lambda x + \mu y \in Q$

Definition 1.10 (Affine Hull). The *Affine Hull* of a set $Q \subseteq \mathbb{E}$, denoted $\text{aff } Q$, is the intersection of all affine sets that contain Q .

Definition 1.11 (Relative Interior). The *Relative interior* of a set $Q \subseteq \mathbb{E}$, denoted $\text{ri } Q$, is the interior of Q relative to $\text{aff } Q$, i.e.

The **Relative Boundary** of Q is the boundary of $\text{ri } Q$, $\partial(\text{ri } Q)$

$$\text{ri } Q = \{x \in Q \mid \exists \varepsilon > 0 \text{ s.t. } B_\varepsilon(x) \cap \text{aff } Q \subseteq Q\}$$

1.4 Differentiation

Definition 1.12 (Differentiable). Let $\Omega \subset \mathbb{E}_1$ be open, a function $f : \Omega \rightarrow \mathbb{E}_2$ is *Differentiable* at $x \in \Omega$ if $\exists L_x \in \mathcal{L}(\mathbb{E}_1, \mathbb{E}_2)$ s.t.

$$\lim_{h \rightarrow 0, h \neq 0} \frac{f(x+h) - f(x) - L_x(h)}{\|h\|} = 0$$

If L_x exists, we call it the derivative of f at x .

Remark 1.13. If f is differentiable at x , $f'(x) \in \mathcal{L}(\mathbb{E}_1, \mathbb{E}_2)$ is uniquely determined through the limit

$$f'(x)(h) = \lim_{t \downarrow 0} \frac{f(x+th) - f(x)}{t} \quad \forall h \in \mathbb{E}$$

Remark 1.14. There exists a unique vector $\nabla f(x) \in \mathbb{E}$ such that $f'(x)(h) = \langle \nabla f(x), h \rangle \forall h \in \mathbb{E}$. We call $f(x)$ the **Gradient vector of f at x** .

Definition 1.15 (Hessian). We say that f is twice differentiable if f' is itself differentiable. This yields a mapping

$$x \in \Omega \rightarrow (f')'(x) = f'' \in \mathcal{L}(\mathbb{E}, \mathcal{L}(\mathbb{E}, \mathbb{R})) = \mathcal{L}^2(\mathbb{E})$$

We obtain a bilinear form

$$(h, d) \in \mathbb{E} \times \mathbb{E} \mapsto [f''(x)(h)](d) = f''(x)[h, d]$$

We can represent the bilinear form $f''(x)$ by a linear operator $\nabla^2 f(x) \in \mathcal{L}(\mathbb{E}, \mathbb{E})$ in the sense that

$$f''(x)[h, d] = \langle \nabla^2 f(x)h, d \rangle \quad \forall h, d \in \mathbb{E}$$

$\nabla^2 f(x)$ is called the **Hessian** of f at x .

Proposition 1.16. If $f : \Omega \rightarrow \mathbb{R}$ is twice differentiable at $x \in \Omega$, it admits a second order **Taylor Approximation** at x ,

$$f(x+h) = f(x) + \langle \nabla f(x), h \rangle + \frac{1}{2} \langle \nabla^2 f(x)h, h \rangle + O\|h\|^2$$

According to the Professor, this will be used a lot.

1.5 Extended Arithmetic and Lower Semicontinuity

Definition 1.17 (Domain). The *Domain* of a function $f : \mathbb{E} \rightarrow \mathbb{R}$ is given by

$$\text{dom } f = \{x \mid f(x) < \infty\}$$

We call f **proper** if $\text{dom } f \neq \emptyset$ and $f(x) > -\infty \forall x \in \mathbb{E}$.

Definition 1.18 (Epigraph). The *epigraph* of $f : \mathbb{E} \rightarrow \mathbb{R}$ is given by

$$\text{epi } f := \{(x, \alpha) \in \mathbb{E} \times \mathbb{R} \mid f(x) \leq \alpha\}$$

The *strict epigraph* is similar:

$$\text{epi}_{<} f := \{(x, \alpha) \in \mathbb{E} \times \mathbb{R} \mid f(x) < \alpha\}$$

Definition 1.19 (Level Sets). The *level sets* of $f : \mathbb{E} \rightarrow \mathbb{R}$ are

$$\text{lev}_{\leq \alpha} f = \{x \in \mathbb{E} \mid f(x) \leq \alpha\}$$

We call f **level bounded** if $\text{lev}_{\leq \alpha} f$ is bounded for any α .

Definition 1.20 (Indicator Function). The indicator function for a set $S \subset \mathbb{E}$, $\delta_S : \mathbb{E} \rightarrow \mathbb{R} \cup \{\infty\}$ is given by

$$\delta_S(x) = \begin{cases} 0, & \text{if } x \in S \\ \infty, & \text{if } x \notin S \end{cases}$$

Definition 1.21 (Lower semicontinuity). Let $f : \mathbb{E} \rightarrow \mathbb{R}$, we call f *Lower semicontinuous* (lsc) at \bar{x} if

$$\liminf_{x \rightarrow \bar{x}} f(x) \geq f(\bar{x})$$

Remark 1.22. f is lsc at \bar{x} if and only if \nexists a sequence $x_k \rightarrow \bar{x}$ such that $\lim_{k \rightarrow \infty} x_k < \bar{x}$.

Definition 1.23 (Lower semicontinuous closure). When a function is not lsc, we can make it. Denote the *Lower semicontinuous closure* of f by $\text{cl}(f) : \mathbb{E} \rightarrow \mathbb{R}$,

$$(\text{cl}(f))(\bar{x}) = \liminf_{x \rightarrow \bar{x}} f(x)$$

Remark 1.24.

$$\text{cl}(\text{epi } f) = \text{epi}(\text{cl } f)$$

Proposition 1.25. Let $f : \mathbb{E} \rightarrow \mathbb{R}$, T.F.A.E.

- (i) f is lsc on \mathbb{E} ;
- (ii) $\text{epi } f$ is closed;
- (iii) $\text{lev}_{\leq \alpha} f$ is closed for every $\alpha \in \mathbb{R}$.

1.6 Optimization Problems

Since we can always turn max into min problems by setting

$$\sup_C f = - \inf_C (-f),$$

we will always work on minimization problems. The function f is called the **Objective function** and C the **Feasible Set**. If $C = \mathbb{E}$, the minimization problem is called **Unconstrained**, otherwise constrained.

Definition 1.26 (Global Minimizers). A *Global minimizer* of C is the solution to the optimization problem $\min f(x)$, $x \in C$. It is denoted

$$\text{argmin}_C f = \{x \in C \mid f(x) = \inf_{\bar{x} \in C} f(\bar{x})\}$$

Remark 1.27. Using indicator functions, we can turn constrained problems into Unconstrained:

$$\begin{aligned} \inf_{x \in C} f(x) &\iff \inf_{x \in \mathbb{E}} f(x) + \delta_C \\ \min_{x \in C} f(x) &\iff \min_{x \in \mathbb{E}} f(x) + \delta_C \\ \text{argmin}_{x \in C} f(x) &\iff \text{argmin}_{x \in \mathbb{E}} f(x) + \delta_C \end{aligned}$$

Proposition 1.28.

$$\text{argmin}_{\mathbb{E}} f \neq \emptyset \implies \inf_{\mathbb{E}} f \in \mathbb{R}$$

Theorem 1.29 (Existence of Minima). Let $f : \mathbb{E} \rightarrow \mathbb{R} \cup \{+\infty\}$ be lsc, proper, and level bounded. Then

$$\text{argmin}_{\mathbb{E}} f \neq \emptyset \quad \text{and} \quad \inf_{\mathbb{E}} f \in \mathbb{R}$$

2 Chapter 2: Convex Sets

2.1 Definitions

Definition 2.1 (Convex Sets). A set $C \subseteq \mathbb{E}$ is convex if it contains all its line segments, that is

$$\lambda x + (1 - \lambda)y \in C, \quad \forall x, y \in C, \lambda \in [0, 1]$$

Definition 2.2 (Convex Combination). A vector of the form

$$\sum_{i=1}^r \lambda_i x_i, \quad \sum_{i=1}^r \lambda_i = 1, \quad \lambda_i \geq 0$$

is called a convex combination of the points $x_1, \dots, x_r \in \mathbb{E}$. It is easily seen that a set $C \subseteq \mathbb{R}^n$ is convex if and only if it contains all convex combinations of its elements.

2.2 Projections on Convex Sets

For a set $S \subseteq \mathbb{E}$ and a given point $x \in \mathbb{E}$, we want to assign to x the closest point in S .

Definition 2.3 (Projection on a Set). Let $S \subseteq \mathbb{E}$ be non empty and $x \in \mathbb{E}$, then the projection of x on S is

$$P_S(x) := \operatorname{argmin}_{y \in S} \|x - y\|$$

Lemma 2.4. Let $x \in \mathbb{E}$ and $S \subseteq \mathbb{E}$, then the following hold:

- (i) If S is closed, then $P_S(x) \neq \emptyset$;
- (ii) If S is convex, then $P_S(x)$ has at most one element.

Corollary 2.4.1 (Projection on closed, convex set). Let $C \subseteq \mathbb{E}$ be non empty, closed, and convex, then $P_C : \mathbb{E} \rightarrow C$ with $P_C(x) = x \iff x \in C$.

Theorem 2.5 (Projection Theorem). Let $C \subseteq \mathbb{E}$ be non empty, closed, and convex, and let $x \in \mathbb{E}$. Then $\bar{v} = P_C(x)$ if and only if

$$\bar{v} \in C \text{ and } \langle \bar{v} - x, v - \bar{v} \rangle \geq 0 \quad \forall v \in C$$

2.3 Separation Theorem

Theorem 2.6 (Separation Theorem). Let $C \subseteq \mathbb{E}$ be non empty, closed, and convex and let $x \notin C$. Then there exists an $s \in \mathbb{E} \setminus \{0\}$ with

$$\langle s, x \rangle > \sup_{v \in C} \langle s, v \rangle$$

Remark 2.7. Under the assumptions of Theorem 2.6, the following hold:

- (a) s can be substituted for $-s$ and thus $\langle s, x \rangle < \inf_{v \in C} \langle s, v \rangle$;
- (b) By positive homogeneity, we can assume wlog that $\|s\| = 1$;
- (c) A reformulation of the statement of the separation theorem is that for $C \subseteq \mathbb{E}$ non empty, closed, and convex and $x \notin C$, there exists an $s \in \mathbb{E} \setminus \{0\}$ and $\beta \in \mathbb{R}$ such that

$$\langle s, x \rangle > \beta \geq \langle s, v \rangle, \quad v \in C$$

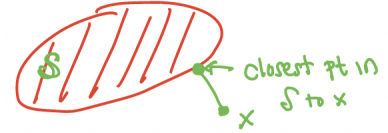


Figure 1: A projection

We say that $P_S(x)$ is a subset of S because it could contain more than one point (unless S is convex).

3 Chapter 3: Convex Functions

3.1 Definitions

Definition 3.1 (Convex Function). A function $f : \mathbb{E} \rightarrow \mathbb{R}$ is said to be convex if $\text{epi } f$ is a convex set.²

² Equivalently, f is convex if $\text{epi}_< f$ is a convex set, as proven in A2.

Proposition 3.2 (Hessian of a Convex Function). A twice differentiable function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is convex, if and only if the Hessian $\nabla^2 f(x)$ is positive semi-definite for all $x \in \mathbb{R}^n$.

*Proof.*³ Assume first that f is convex and let $x \in \mathbb{R}^n$. Define moreover the function $g : \mathbb{R}^n \rightarrow \mathbb{R}$ setting

$$g(y) := f(y) - \nabla f(x)^T(y - x).$$

Since the mapping $y \mapsto -\nabla f(x)^T(y - x)$ is affine, it follows that g is convex. Moreover

$$\nabla g(y) = \nabla f(y) - \nabla f(x)$$

and

$$\nabla^2 g(y) = \nabla^2 f(y)$$

for all $y \in \mathbb{R}^n$. In particular, $\nabla g(x) = 0$. Thus Corollary 6 implies that x is a global minimiser of g . Now the second order necessary condition for a minimiser implies that $\nabla^2 g(x)$ is positive semi-definite. Since $\nabla^2 g(x) = \nabla^2 f(x)$ and x was arbitrary, this proves that the Hessian of f is positive semi-definite for all $x \in \mathbb{R}^n$.

Now assume that the Hessian $\nabla^2 f(x)$ of f is positive semi-definite for all $x \in \mathbb{R}^n$. Let moreover $x, y \in \mathbb{R}^n$. Then Taylor's theorem implies that

$$f(y) = f(x) + \nabla f(x)^T(y - x) + \frac{1}{2}(y - x)^T \nabla^2 f(x + t(y - x))(y - x)$$

for some $0 \leq t \leq 1$. Since $\nabla^2 f$ is everywhere positive semi-definite, the quadratic term in this equation is always non-negative. Thus we can estimate

$$f(y) \geq f(x) + \nabla f(x)^T(y - x).$$

And so, f is convex. □

Remark 3.3.

- (i) Convex functions have convex level sets.
- (ii) The domain of convex functions is convex.

Definition 3.4 (Proper Convex Functions). A convex function is proper if $\text{dom } f \neq \emptyset$ and $f(x) > -\infty \forall x \in \mathbb{E}$.

³ Markus Grasmair. *BASIC PROPERTIES OF CONVEX FUNCTIONS*. Norwegian University of Science and Technology. URL https://wiki.math.ntnu.no/_media/tma4180/2016v/note2.pdf

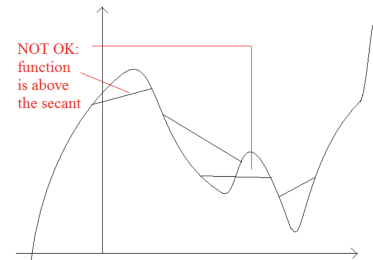


Figure 2: A nonconvex function does not lie below all of its secants.

Proposition 3.5 (Characterizing Convexity). A function $f : \mathbb{E} \rightarrow \mathbb{R} \cup \{+\infty\}$ is convex if and only if $\forall x, y \in \mathbb{E}$, we have ⁴

$$f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y) \quad \forall \lambda \in [0, 1] \quad (1)$$

Proposition 3.6 (Jensen's Inequality). A function $f : \mathbb{E} \rightarrow \mathbb{R} \cup \{\infty\}$ is convex \iff

$$f\left(\sum_{i=1}^p \lambda_i x_i\right) \leq \sum_{i=1}^p \lambda_i f(x_i), \quad x_i \in \mathbb{E}, \lambda_i \in \Delta_p$$

Where $\Delta_p := \{\lambda_i \in \mathbb{R}^p \mid \lambda_i \geq 0, \sum \lambda_i = 1\}$.

Definition 3.7 (Convexity on a Set). Let $f : \mathbb{E} \rightarrow \mathbb{R} \cup \{\infty\}$, and $C \subset \text{dom}(f)$ a convex, non-empty set. Then we call f convex on C if $\forall x, y \in C$, (1) holds.

Corollary 3.7.1. Let $f : \mathbb{E} \rightarrow \mathbb{R} \cup \{\infty\}$, then T.F.A.E.

- (i) f is convex;
- (ii) f is convex on its domain.

Remark 3.8 (Notation). Since we are mainly concerned with **proper**, **convex**, and sometimes **lower semicontinuous** functions, we shall use the following notation to denote such functions:

$$\Gamma = \Gamma(\mathbb{E}) := \{f : \mathbb{E} \rightarrow \mathbb{R} \cup \{\infty\} \mid f \text{ is proper \& convex} \}$$

$$\Gamma_0 = \Gamma_0(\mathbb{E}) := \{f : \mathbb{E} \rightarrow \mathbb{R} \cup \{\infty\} \mid f \text{ is proper, convex, \& lsc} \}$$

3.2 Stronger Notions of Convexity

Definition 3.9 (Strict Convexity). Let $f \in \Gamma$ and $C \subset \text{dom } f$ a convex set, then f is said to be **strictly convex** on C if

$$f(\lambda x + (1 - \lambda)y) < \lambda f(x) + (1 - \lambda)f(y),$$

for any $\lambda \in (0, 1)$, $x, y \in C$ s.t. $x \neq y$. For $C = \text{dom } f$, we simply say that f is strictly convex.

Definition 3.10 (σ -Strong Convexity). Let $f \in \Gamma$ and $C \subset \text{dom } f$ a convex set, then f is said to be **σ -strongly convex** on C if

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

for any $\lambda \in (0, 1)$, $x, y \in C$ s.t. $x \neq y$. For $C = \text{dom } f$, we simply say that f is strongly convex.

Remark 3.11. Strongly convex \implies Strictly convex, but the converse is not true.

⁴Stephen D. Boyles. *Convex Examples*. Spring 2015
Jensen's Inequality is a generalization of Proposition 3.5.

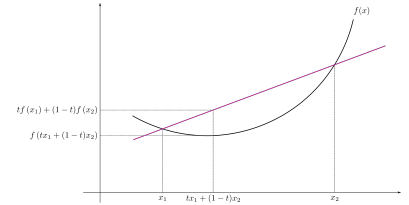


Figure 3: A convex function lies below all its secants.

The scalar $\sigma > 0$ is called the **modulus of strong convexity** of f on C

Proposition 3.12 (Characterizing Strong Convexity). Let $f \in \Gamma$ and $C \subset \text{dom } f$, then f is strongly convex if and only if $f(x) - \frac{\sigma}{2}\|x\|^2$ is convex on C .

Proof. See assignment 2. □

3.3 Functional Operations Preserving Convexity

Proposition 3.13 (Positive combinations of convex functions). For $p \in \mathbb{N}$, let f_i be convex (and lsc), $\alpha_i \geq 0$ for $i = 1, 2, \dots, p$, then

$$f = \sum_{i=1}^p \alpha_i f_i \quad \text{is convex (and lsc).}$$

If in addition $\bigcap_{i=1}^p \text{dom } f_i \neq \emptyset$, then f is also proper (i.e. $\text{dom } f \neq \emptyset$).

Proposition 3.14 (Pointwise supremum of convex functions). For an arbitrary index set \mathcal{I} , let f_i be convex (lsc) for all $i \in \mathcal{I}$. Then, the function

$$f(x) = \sup_{i \in \mathcal{I}} f_i(x) \quad \forall x \in \mathbb{E}$$

is convex (lsc).

Definition 3.15 (Affine Functions). A function $F : \mathbb{E}_1 \rightarrow \mathbb{E}_2$ is called *Affine* if

$$F(\alpha x + (1 - \alpha)y) = \alpha F(x) + (1 - \alpha)F(y) \quad \forall x, y \in \mathbb{E}_1, \alpha \in \mathbb{R}$$

Remark 3.16.

1. Affine functions are convex and continuous.
2. Linear functions are affine, and an affine function F is linear if and only if $F(0) = 0$.
3. Affine images and preimages of convex sets are convex.

Proposition 3.17 (Precomposition with an affine mapping). Let $H : \mathbb{E}_1 \rightarrow \mathbb{E}_2$ be affine and $g : \mathbb{E}_2 \rightarrow \mathbb{R} \cup \{\infty\}$ be convex (and lsc), then the function $f = g \circ H$ is convex (and lsc).

4 Minimization and Convexity

4.1 General Existence Result

Definition 4.1 (Coercivity and Supercoercivity). Let $f : \mathbb{E} \rightarrow \overline{\mathbb{R}}$, then f is called

(i) **Coercive** if

$$\lim_{\|x\| \rightarrow \infty} f(x) = +\infty$$

(ii) **Supercoercive** if

$$\lim_{\|x\| \rightarrow \infty} \frac{f(x)}{\|x\|} = +\infty$$

Proposition 4.2. A function is coercive if and only if it is level bounded.

Proof. See assignment 2. □

Proposition 4.3 (Existence of Minimizers). Let $f : \mathbb{E} \rightarrow \mathbb{R} \cup \{\infty\}$ be lsc and let $C \subset \mathbb{E}$ be closed such that $\text{dom } f \cap C \neq \emptyset$. If one of the following holds,

(i) f is coercive;

(ii) C is bounded;

then f has a minimizer over C .

Important corollary.

Corollary 4.3.1 (Existence of Minimizers II). Let $f, g : \mathbb{R} \rightarrow \mathbb{R} \cup \{\infty\}$ be lsc such that $\text{dom } f \cap \text{dom } g \neq \emptyset$. If f is coercive and g is bounded from below, then $f + g$ is coercive and has a minimizer over \mathbb{E} .

4.2 Convex Optimization

In the case of convex minimization, there is no difference between local and global solutions.

Proposition 4.4. Let $f \in \Gamma$, then every local minimizer over \mathbb{E} is a global minimizer.

Corollary 4.4.1 (Minimizers in Convex Optimization). Let $f \in \Gamma$ and $C \subset \mathbb{E}$ be non-empty and convex. Then every local minimizer of f over C is a global minimizer of f over C .

Proposition 4.5. Let $f \in \Gamma$, then $\text{argmin } f$ is a convex set.

Corollary 4.5.1. Let $f \in \Gamma$ and $C \in \mathbb{E}$ convex. Then $\text{argmin}_C f$ is convex.

Proposition 4.6 (Uniqueness of Minimizers). Let $f \in \Gamma$ be strictly convex, then f has at most one minimizer.

Proof. Assignment 2. □

Proposition 4.7 (Minimizing the sum of Convex Functions). Let $f, g \in \Gamma_0$ such that $\text{dom } f \cap \text{dom } g \neq \emptyset$. If one of the following holds,

(i) f is supercoercive;

(ii) f is coercive and g is bounded from below;

then $f + g$ is coercive and has a minimizer over \mathbb{E} . If f or g are strictly convex, then $f + g$ has a unique minimizer.

Theorem 4.8 (Parametric Minimization). Let $h : \mathbb{E}_1 \times \mathbb{E}_2 \rightarrow \mathbb{R} \cup \{\infty\}$ be convex. Then the **Optimal value function**

$$\varphi : \mathbb{E}_1 \rightarrow \overline{\mathbb{R}}, \quad \varphi(x) = \inf_{y \in \mathbb{E}_2} h(x, y),$$

is convex. Moreover, the set-valued mapping

$$x \mapsto \operatorname{argmin}_{y \in \mathbb{E}_2} h(x, y) \subset \mathbb{E}_2,$$

is convex-valued.

5 Infimal Convolution of Convex Functions

Definition 5.1 (Infimal Convolution). Let $f, g : \mathbb{E} \rightarrow \mathbb{R} \cup \{\infty\}$, then the function

$$f \# g : \mathbb{E} \rightarrow \overline{\mathbb{R}}, \quad (f \# g)(x) = \inf_{u \in \mathbb{E}} \{f(u) + g(x - u)\}$$

is called the **Infimal Convolution** of f and g .

Lemma 5.2. Let $f, g : \mathbb{E} \rightarrow \mathbb{R} \cup \{\infty\}$, then the following hold:

- (i) $\operatorname{dom}(f \# g) = \operatorname{dom} f + \operatorname{dom} g$;
- (ii) $f \# g = g \# f$;
- (iii) $(f \# g)(x) \leq f(u) + g(x - u) \quad \forall u \in \mathbb{E}$.

We call the infimal convolution $f \# g$ **exact** at $x \in \mathbb{E}$ if $\operatorname{argmin}_{u \in \mathbb{E}} \{f(u) + g(x - u)\} \neq \emptyset$. We simply say it is **exact** if it is exact everywhere.

Proposition 5.3 (Infimal convolution of convex functions). Let $f, g : \mathbb{E} \rightarrow \mathbb{R} \cup \{\infty\}$ be convex, then $f \# g$ is convex.

Proof. Assignment 3. □

Definition 5.4 (Distance function). Let $C \subset \mathbb{E}$, then the distance function $d_C = \delta_C \# \|\cdot\|$ is called the **Distance Function** to the set C . It holds that

$$d_C(x) = \inf_{u \in C} \|x - u\|$$

Corollary 5.4.1. If $C \subset \mathbb{E}$ is closed and convex, we have

$$d_C(x) = \|x - P_C(x)\|$$

Moreover, the distance function of a closed set is convex by the above proposition.

Proof. Follows from lemma 2.4. □

Proposition 5.5 (Affine Minorization Principle). For $f \in \Gamma_0$, there exists an *Affine minorant*, namely, $\exists a \in \mathbb{E}, \beta \in \mathbb{R}$ s.t.

$$f(x) \geq \langle a, x \rangle + \beta \quad \forall x \in \mathbb{E}$$

Theorem 5.6 (Infimal Convolution of Γ_0). Let $f, g \in \Gamma_0$ and suppose that one of the following holds:

- (i) f is *supercoercive*;
- (ii) f is *coercive and bounded from below*;

then $f \# g \in \Gamma_0$ and it is exact, i.e. $\forall x \in \mathbb{E}, \operatorname{argmin}_{u \in \mathbb{E}} \{f(u) + g(x - u)\} \neq \emptyset$.

5.1 Moreau Envelopes and Proximal Mappings

The Moreau envelope of a proper lower semi-continuous convex function $f \in \Gamma_0$, denoted $e_\lambda f$ is a smoothed version of f . It was proposed by Jean-Jacques Moreau in 1965. The Moreau envelope has important applications in mathematical optimization: minimizing over the Moreau envelope of f and minimizing over f are equivalent problems in the sense that the set of minimizers of f and $e_\lambda f$ are the same. However, first-order optimization algorithms can be directly applied to $e_\lambda f$, since f may be non-differentiable while $e_\lambda f$ is always continuously differentiable (given $f \in \Gamma_0$). Indeed, many proximal gradient methods can be interpreted as a gradient descent method over $e_\lambda f$.

In this section, if a function S maps points in a euclidian space \mathbb{E}_1 to subsets of another euclidian space \mathbb{E}_2 , we denote it by $S : \mathbb{E}_1 \rightrightarrows \mathbb{E}_2$.

Definition 5.7 (Moreau Envelope). Let $f : \mathbb{E} \rightarrow \overline{\mathbb{R}}$, the **Moreau envelope** of f to the parameter $\lambda > 0$ is the function $e_\lambda f : \mathbb{E} \rightarrow \overline{\mathbb{R}}$ defined by

$$e_\lambda f(x) = \inf_{u \in \mathbb{E}} \left\{ f(u) + \frac{1}{2\lambda} \|x - u\|^2 \right\}$$

Remark 5.8. Notice that

$$e_\lambda f = f \# \left(\frac{1}{2\lambda} \|\cdot\|^2 \right)$$

Definition 5.9 (Proximal Mapping). Let $f : \mathbb{E} \rightarrow \overline{\mathbb{R}}$, the **Proximal mapping** to the parameter $\lambda > 0$ of f is the (possibly set-valued mapping) $P_\lambda f : \mathbb{E} \rightrightarrows \mathbb{E}$ defined by

$$P_\lambda f(x) = \operatorname{argmin}_{u \in \mathbb{E}} \left\{ f(u) + \frac{1}{2\lambda} \|x - u\|^2 \right\}$$

Proposition 5.10. Let $f \in \Gamma_0$ and $\lambda > 0$, then $e_\lambda f \in \Gamma_0$ is finite-valued and $P_\lambda f$ is single valued.

Remark 5.11. The Moreau envelope is a smoothing operation, namely, the Moreau envelope of a closed, proper, convex function is differentiable.

Proposition 5.12. The proximal operator of f is related to the gradient of the Moreau envelope of f by the following identity

$$\nabla e_\lambda f = \frac{1}{\lambda}(x - P_\lambda f(x))$$

Example 5.13 (Moreau Envelope of an indicator function). Let $C \subset \mathbb{E}$ be nonempty, closed, and convex. Then

$$e_\lambda \delta_C(x) = \inf_{u \in C} \frac{1}{2\lambda} \|x - u\|^2 = \frac{1}{2\lambda} d_C^2(x),$$

i.e. $e_\lambda \delta_C$ is the distance function and $P_\lambda \delta_C = P_C$ is the projection onto C .

Remark 5.14. For $f \in \Gamma_0$, $\lambda > 0$ and $x \in \mathbb{E}$, we have

$$e_\lambda f(x) = f(P_\lambda f(x)) + \frac{1}{2\lambda} \|x - P_\lambda f(x)\|^2 \leq f(u) + \frac{1}{2\lambda} \|x - u\|^2 \quad \forall u \in \mathbb{E},$$

in particular, the Moreau envelope is a *Minorant* for its input function.

Proposition 5.15. Let $f \in \Gamma_0$ and let $x, p \in \mathbb{E}$. Then $p = P_1 f(x) \iff \langle y - p, x - p \rangle + f(p) \leq f(y) \quad \forall y \in \mathbb{E}$.

Proposition 5.16 (Firm nonexpansiveness of prox-operator). Let $f \in \Gamma_0$, then

$$\|P_1(x) - P_1(y)\|^2 \leq \langle x - y, P_1 f(x) - P_1 f(y) \rangle \quad \forall x, y \in \mathbb{E}.$$

In particular, $\|P_1(x) - P_1(y)\| \leq \|x - y\| \quad \forall x, y \in \mathbb{E}$. That is, $P_1 f$ is globally 1-Lipschitz continuous.

Remark 5.17. The above result extends to any $\lambda > 0$ through the identity

$$P_\lambda f = P_1(\lambda f)$$

6 Conjugacy of Convex Functions

6.1 Convex Hull of a Function

Main Result: Every proper, closed, convex function is the pointwise supremum of its affine minorants.

Theorem 6.1 (Envelope Representation of Γ_0). Let $f \in \Gamma_0$, then f is the pointwise supremum of all affine functions minorizing it, i.e.

$$f(x) = \sup\{h(x) \mid h \leq f, h \text{ affine}\}$$

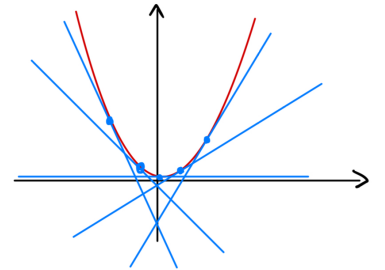


Figure 4: Pointwise affine minorants.

Definition 6.2 (Convex Hull of a Function). Let $f : \mathbb{E} \rightarrow \overline{\mathbb{R}}$, then the pointwise supremum of all convex functions minorizing f , i.e.

$$(\text{conv } f)(x) = \sup\{h(x) \mid h \leq f, h \text{ convex}\},$$

is called the **convex hull** of f . Moreover, we define the **closed convex hull** of f to be

$$(\overline{\text{conv}})(x) = \sup\{h(x) \mid h \leq f, h \text{ lsc and convex}\}$$

Remark 6.3. Since both convexity and lower semicontinuity are preserved under pointwise suprema, we find that $\text{conv } f$ is the largest convex function that minorizes f and $\overline{\text{conv}} f$ is the largest lsc and convex function that minorizes f .

Remark 6.4. (i) $\overline{\text{conv}} f \leq \text{conv } f \leq f$ and $\overline{\text{conv}} f = \text{cl}(\text{conv } f)$;

(ii) We can also define the convex hull of $C \subseteq \mathbb{E}$ as

$$\text{conv } C = \bigcap_{C \subset S, S \text{ convex}} S$$

(iii) Every affine function $\mathbb{E} \rightarrow \mathbb{R}$ is convex and lsc, hence the functions f , $\text{cl } f$, $\text{conv } f$, $\overline{\text{conv}} f$ have the same affine minorants.

Corollary 6.4.1 (Affine envelope representation of $\overline{\text{conv}}$). Let $f : \mathbb{E} \rightarrow \overline{\mathbb{R}}$ be proper and minorized by an affine function. Then $\overline{\text{conv}} f$ is the pointwise supremum of all affine functions minorizing f , i.e.

$$(\overline{\text{conv}} f)(x) = \sup\{h(x) \mid h \leq f, h \text{ affine}\}$$

6.2 Fenchel Conjugate

Definition 6.5 (Fenchel Conjugate). Let $f : \mathbb{E} \rightarrow \overline{\mathbb{R}}$, then the **Fenchel Conjugate** (or simply conjugate) of f is the function $f^* : \mathbb{E} \rightarrow \overline{\mathbb{R}}$ defined by

$$f^*(y) = \sup_{x \in \mathbb{E}} \{\langle x, y \rangle - f(x)\}$$

The function $f^{**} = (f^*)^*$ is called the Fenchel biconjugate of f .

Proposition 6.6 (Fenchel-Young Inequality). Let $f : \mathbb{E} \rightarrow \overline{\mathbb{R}}$, then

$$f(x) + f^*(y) \geq \langle x, y \rangle \quad \forall x, y \in \mathbb{E}$$

Remark 6.7. $f \leq g \implies f^* \geq g^*$ and in particular $f \geq f^{**}$.

Motivation for the Fenchel Conjugate. Let $F : \mathbb{E} \rightarrow \overline{\mathbb{R}}$, we notice that

$$\text{epi } f^* = \{(y, \beta) \mid \langle x, y \rangle - f(x) \leq \beta \quad \forall x \in \mathbb{E}\}$$

This means the conjugate of f is the function whose epigraph is the set of all (y, β) defining affine functions $x \mapsto \langle y, x \rangle - \beta$ that minorizes f . In view of Corollary 6.4.1, if f is proper and has an affine minorant, the pointwise supremum of these affine mappings is the closed, convex hull of f , i.e. through its epigraph, f^* encodes info about f and its closed convex hull. Since

$$f^*(y) = \sup\{\langle x, y \rangle - f(x)\} = \sup_{(x, \alpha) \in \text{epi } f} \{\langle y, x \rangle - \alpha\} \quad \forall y \in \mathbb{E},$$

we also have

$$\text{epi } f^* = \{(y, \beta) \mid \langle x, y \rangle - \alpha \leq \beta, (x, \alpha) \in \text{epi } f\}$$

Theorem 6.8 (Fenchel-Moreau Theorem). Let $f : \mathbb{E} \rightarrow \overline{\mathbb{R}}$ be proper and have an affine minorant, then the following hold:

- (i) f^*, f^{**} are closed, proper, and convex;
- (ii) $f^{**} = \overline{\text{conv}} f$;
- (iii) $f^* = (\text{conv } f)^* = (\text{cl } f)^* = (\overline{\text{conv}} f)^*$.

Remark 6.9. $f \geq f^{**}$ for any function $f : \mathbb{E} \rightarrow \overline{\mathbb{R}}$ that has an affine minorant, and it holds that $f^{**} = f$ if and only if f is closed and convex.

Proposition 6.10 (Conjugacy of infimal convolution). Let $f, g : \mathbb{E} \rightarrow \overline{\mathbb{R}}$, then

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

Proof. (HWK)

$$\begin{aligned} (f \# g)^*(x) &= \sup_{v \in \mathbb{E}} \{\langle v, x \rangle - (f \# g)(v)\} \\ &= \sup_{v \in \mathbb{E}} \{\langle v, x \rangle - \inf_{u \in \mathbb{E}} \{f(u) + g(v - u)\}\} \\ &= \sup_{v \in \mathbb{E}} \{\langle v, x \rangle + \sup_{u \in \mathbb{E}} \{-f(u) - g(v - u)\}\} \\ &= \sup_{v, u \in \mathbb{E}} \{\langle v, x \rangle + -f(u) - g(v - u)\} \\ &= \sup_{v, u \in \mathbb{E}} \{\langle v - u, x \rangle + \langle u, x \rangle - f(u) - g(v - u)\} \\ &= \sup_{v, u \in \mathbb{E}} \{(\langle v - u, x \rangle - g(v - u)) + (\langle u, x \rangle - f(u))\} \\ &= \sup_{v \in \mathbb{E}} \{\langle v, x \rangle - f(v)\} + \sup_{v \in \mathbb{E}} \{\langle v, x \rangle - g(v)\} \\ &= f^*(x) + g^*(x) \end{aligned}$$

□

Corollary 6.10.1 (Conjugacy correspondence). The Fenchel transform induces a 1-1 correspondence on Γ_0 . For $f, g \in \Gamma_0$, f is the conjugate to g if and only if f is the conjugate to f .

Proposition 6.11 (Elementary cases of Conjugacy). Let $f : \mathbb{E} \rightarrow \overline{\mathbb{R}}$, then the following hold:

1. $(f - \langle a, \cdot \rangle)^* = f^*((\cdot) + a)$ for $a \in \mathbb{E}$;
2. $(f + \gamma)^* = f^* - \gamma$ for all $\gamma \in \mathbb{R}$;
3. $(\lambda f)^* = \lambda f^*\left(\frac{(\cdot)}{\lambda}\right)$ for all $\lambda > 0$;
4. Let $g, h : \mathbb{E} \rightarrow \overline{\mathbb{R}}$ be proper, closed, convex functions, and let $A : \mathbb{E} \rightarrow Y$ be a linear map. Define $F(x, y) = h(Ax + y) + g(x)$. Then,

$$F^*(x, y) = g^*(x - A^*y) + h^*(y),$$

where A^* is the adjoint map.

Proof.

1.

$$\begin{aligned} f^*(\cdot + a) &= \sup_{x \in \mathbb{E}} \{ \langle x, \cdot + a \rangle - f(x) \} \\ &= \sup_{x \in \mathbb{E}} \{ \langle x, \cdot \rangle + \langle x, a \rangle - f(x) \} \\ &= \sup_{x \in \mathbb{E}} \{ \langle x, \cdot \rangle - (f - \langle a, \cdot \rangle)(x) \} \\ &= (f - \langle a, \cdot \rangle)^* \end{aligned}$$

2.

$$\begin{aligned} (f + \gamma)^* &= \sup_{x \in \mathbb{E}} \{ \langle x, \cdot \rangle - (f + \gamma)(x) \} \\ &= \sup_{x \in \mathbb{E}} \{ \langle x, \cdot \rangle - f(x) - \gamma \} \\ &= \sup_{x \in \mathbb{E}} \{ \langle x, \cdot \rangle - f(x) \} - \gamma \\ &= f^* - \gamma \end{aligned}$$

3.

$$\begin{aligned} (\lambda f)^* &= \sup_{x \in \mathbb{E}} \{ \langle x, \cdot \rangle - (\lambda f)(x) \} \\ &= \sup_{x \in \mathbb{E}} \left\{ \lambda \left(\frac{\langle x, \cdot \rangle}{\lambda} - f(x) \right) \right\} \\ &= \lambda \sup_{x \in \mathbb{E}} \left\{ \frac{1}{\lambda} \langle x, \cdot \rangle - f(x) \right\} \\ &= \lambda \sup_{x \in \mathbb{E}} \left\{ \left\langle x, \frac{(\cdot)}{\lambda} \right\rangle - f(x) \right\} \\ &= \lambda f^*\left(\frac{(\cdot)}{\lambda}\right) \end{aligned}$$

□

7 Differential Theory

We will study differentiability of convex functions and, in particular, develop tools where differentiability is lacking.

7.1 Convex Subdifferential

Some operations on convex functions destroy differentiability but preserve convexity - such as the max operation. In these situations, subgradients offer a method of generalizing gradients for optimizing convex functions that are not necessarily differentiable (where gradient descent does not work). For motivational purposes, we consider the following result on smooth, convex functions.

Proposition 7.1. Let $C \subset \mathbb{E}$ be open and convex and let $f : \mathbb{E} \rightarrow \mathbb{R}$ be continuous and differentiable on C . Then f is convex on C if and only if

$$f(x) \geq f(\bar{x}) + \langle \nabla f(\bar{x}), x - \bar{x} \rangle \quad \forall x, \bar{x} \in C$$

Proof. Assignment 3. □

Definition 7.2 (Convex Subdifferential). Let $f : \mathbb{E} \rightarrow \overline{\mathbb{R}}$ and $\bar{x} \in \mathbb{E}$, then $g \in \mathbb{E}$ is called a **subgradient** of f at \bar{x} if

$$f(x) \geq f(\bar{x}) + \langle g, x - \bar{x} \rangle \quad \forall x \in \mathbb{E}.$$

The set

$$\partial f(\bar{x}) = \{v \in \mathbb{E} \mid f(x) \geq f(\bar{x}) + \langle v, x - \bar{x} \rangle \quad \forall x \in \mathbb{E}\}$$

of all subgradients is called the convex **Subdifferential** of f at \bar{x} .

Remark 7.3. Note that we did not restrict ourselves to convex functions in the above definition but the class Γ is where the subdifferential is most meaningful.

Remark 7.4. If a function is differentiable at \bar{x} , then the *subdifferential* $\partial f(\bar{x})$ contains only one subgradient (∇f).

Example 7.5 (Subdifferential of Indicator Function). Let $C \subset \mathbb{E}$ be convex and $\bar{x} \in C$, then

$$\begin{aligned} g \in \partial \delta_C(\bar{x}) &\iff \delta_C(x) \geq \delta_C(\bar{x}) + \langle g, x - \bar{x} \rangle \quad \forall x \in \mathbb{E} \\ &\iff 0 \geq \langle g, x - \bar{x} \rangle \quad \forall x \in C \end{aligned}$$

that is,

$$\partial \delta_C(\bar{x}) = \{v \in \mathbb{E} \mid \langle v, x - \bar{x} \rangle \leq 0 \quad \forall x \in C\}$$

The latter set is called the **Normal cone** of C at \bar{x} .

The subdifferential $\partial f(\bar{x})$ of $f : \mathbb{E} \rightarrow \overline{\mathbb{R}}$ at $\bar{x} \in \mathbb{E}$ contains exactly the slopes of all affine minorants of f that coincide with f at \bar{x} .

Definition 7.6 (Normal Cone). If $C \subseteq \mathbb{E}$ is a set, then the **Normal cone** of C at \bar{x} is given by

$$N_C(\bar{y}) = \{v \in \mathbb{E} \mid \langle y - \bar{y}, v \rangle \leq 0 \ \forall y \in C\}$$

Proposition 7.7 (Properties of Subdifferentials). Let $f : \mathbb{E} \rightarrow \bar{\mathbb{R}}$, then the following hold,

- (i) $\partial f(\bar{x})$ is closed and convex for $\bar{x} \in \text{dom } f$
- (ii) If f is proper, then $\partial f(\bar{x}) = \emptyset \ \forall \bar{x} \notin \text{dom } f$
- (iii) $0 \in \partial f(\bar{x}) \iff \bar{x} \in \text{argmin } f$
- (iv) If f is convex, then

$$\partial f(\bar{x}) = \{v \in \mathbb{E} \mid (v, -1) \in N_{\text{epi } f}(\bar{x}, f(\bar{x}))\}$$

Proof.

- (i) We have

$$(\partial f)(\bar{x}) \cap \bigcap_{x \in \mathbb{E}} \{v \in \mathbb{E} \mid f(x) \geq f(\bar{x}) + \langle v, x - \bar{x} \rangle\},$$

since the intersection of convex, closed sets is also convex and closed.

- (ii) By proper, there exists an x s.t. $f(x) < \infty$ and since $\bar{x} \notin \text{dom}(f)$, we must have $f(\bar{x}) = \infty$. Using the definition of subgradient, we see that no g satisfies the equation $f(x) \geq \langle g, x - \bar{x} \rangle + f(\bar{x})$, thus $\partial f(\bar{x}) = \emptyset$.

- (iii) By definition, we have $0 \in \partial f(\bar{x}) \iff f(x) \geq f(\bar{x}) \ \forall x \in \mathbb{E}$.

Indeed,

(\implies) $0 \in \partial f(\bar{x}) \implies f(x) \geq f(\bar{x})$ by $f(x) \geq \langle 0, x - \bar{x} \rangle + f(\bar{x})$, for all x .

(\impliedby) $f(x) \geq f(\bar{x}) \ \forall x \implies f(x) \geq \langle 0, x - \bar{x} \rangle + f(\bar{x}) \ \forall x \implies$ by definition of $f(x) \geq \langle g, x - \bar{x} \rangle + f(\bar{x})$, $0 \in (\partial f)(\bar{x})$.

- (iv)

$$\begin{aligned} v \in (\partial f)(\bar{x}) &\iff f(x) \geq f(\bar{x}) + \langle v, x - \bar{x} \rangle \ \forall x \in \text{dom } f \\ &\iff \alpha \geq f(\bar{x}) + \langle v, x - \bar{x} \rangle \ \forall (x, \alpha) \in \text{epi } f \\ &\iff 0 \geq (f(\bar{x}) - \alpha) + \langle v, x - \bar{x} \rangle = \left\langle \begin{pmatrix} v \\ -1 \end{pmatrix}, \begin{pmatrix} x - \bar{x} \\ \alpha - f(\bar{x}) \end{pmatrix} \right\rangle \ \forall (x, \alpha) \in \text{epi } f \\ &\iff (v, -1) \in N_{\text{epi } f}(\bar{x}, f(\bar{x})) \end{aligned}$$

□

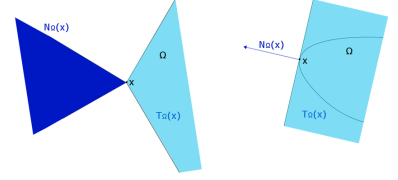


Figure 5: Normal cones.

Recall that $N_{\text{epi } f}$ denotes the normal cone of the epigraph of f . We note that if a function f is differentiable at x , then the normal cone at x contains only one vector.

Theorem 7.8 (Subdifferential and Conjugate Functions). Let $f : \mathbb{E} \rightarrow \bar{\mathbb{R}}$, then T.F.A.E.

- (i) $y \in \partial f(x)$
- (ii) $x \in \operatorname{argmax}_z \{\langle z, y \rangle - f(z)\}$
- (iii) $f(x) + f^*(y) = \langle x, y \rangle$ (equality in Fenchel Young if and only if $y \in \partial f(x)$)

If $f \in \Gamma_0$, these are equivalent to

- (iv) $x \in (\partial f^*)(y)$
- (v) $y \in \operatorname{argmax}_w \{\langle x, w \rangle - f^*(w)\}$

Proof.

$$\begin{aligned}
 y \in (\partial f)(x) &\iff f(z) \geq \langle x, z \rangle + \langle y, z - x \rangle \quad \forall z \\
 &\iff \langle y, x \rangle - f(x) \geq \langle y, z \rangle - f(z) \quad \forall z \\
 &\iff \langle y, x \rangle - f(x) \geq \sup_z \{\langle y, z \rangle - f(z)\} \\
 &\iff \langle y, x \rangle - f(x) \geq f^*(y)
 \end{aligned}$$

And, because Fenchel Young gives the reverse inequality, it must be true that we had equalities instead of inequalities the whole time. For (iv) and (v), apply the same reasoning to f^* and use that $f^{**} = f$ if $f \in \Gamma_0$. \square

Remark 7.9. (i) and (iv) $\iff (\partial f^*) = (\partial f)^{-1}$.

7.2 Directional Derivative of Convex Functions

Definition 7.10 (Directional Derivative). let $f : \mathbb{E} \rightarrow \bar{\mathbb{R}}$ be proper. For $x \in \operatorname{dom} f$, we say that f is directionally differentiable at x in the direction $d \in \mathbb{E}$ if

$$\lim_{t \downarrow 0} \frac{f(x + td) - f(x)}{t} \text{ exists}$$

In this case, we call

$$f'(x; d) = \lim_{t \downarrow 0} \frac{f(x + td) - f(x)}{t}$$

the directional derivative of f at x in direction d .

Remark 7.11. If f is differentiable at $x \in \operatorname{int}(\operatorname{dom} f)$, then

$$f'(x; d) = \langle \nabla f(x), d \rangle \quad \forall d$$

Example 7.12. Let $f : \mathbb{E} \rightarrow \mathbb{R}$, $f(x) = \|x\|$, then

$$f'(x; d) = \begin{cases} \left\langle \frac{x}{\|x\|}, d \right\rangle, & \text{if } x \neq 0 \\ \|d\|, & \text{otherwise.} \end{cases}$$

for $x \neq 0$, f is differentiable, so compute ∇f . For $x = 0$,

$$\lim_{t \rightarrow 0} \frac{\|0 + td\| - \|0\|}{t} = \lim_{t \rightarrow 0} \|d\| = \|d\|$$

Next, we show that

- (i) $f'(x; d)$ exists;
- (ii) positively homogeneous, finite, and convex in the interior of the domain.

if $f \in \Gamma$. Let $h : \mathbb{E} \rightarrow \mathbb{R} \cup \{\infty\}$, we call h positively homogeneous if

$$h(\lambda x) = \lambda h(x) \quad \forall x \in \mathbb{E}, \lambda \geq 0$$

HWK: show h is convex and homogeneous if and only if

$$h(\lambda x + \mu y) \leq \lambda h(x) + \mu h(y) \quad \forall x, y \in \mathbb{E}, \mu, \lambda \geq 0$$

This property is called **sublinearity**.

Remark 7.13. Observe that if h is positive homogeneous, then $h(0) = 0$, and h satisfies

$$-h(x) = h(-x) \quad \forall x \in \mathbb{E}$$

since $h(0) = 0 = h(-x + x) \stackrel{\text{sublinearity}}{\leq} h(x) + h(-x)$

Proposition 7.14 (Directional Derivative of Convex Functions). Let $f \in \Gamma$, $x \in \text{dom } f$ and $d \in \mathbb{E}$ direction. Define

$$t \in \mathbb{R} \setminus \{0\} \quad q(t) = \frac{f(x + td) - f(x)}{t}$$

then, the following hold,

- (i) $q(-t) \leq q(-s) \leq q(s) \leq q(t)$;
- (ii) We have $f'(x; d) = \inf_{t > 0} q(t)$. In particular, $f'(x; d)$ exists at least in the extended sense.
- (iii) $\text{dom}(f'(x; \cdot)) = \mathbb{R}_+(\text{dom } f - x) := \{\lambda y \mid \lambda \in \mathbb{R}_+, y \in \text{dom}(f - x)\}$
- (iv) If $x \in \text{int}(\text{dom } f)$, the function $f'(x; \cdot)$ is finite valued, positively homogeneous, and convex (also sublinear).

Proof.

(i) HWK. (use convexity, choose the right λ)

(ii) Since the infimum always exists, $\inf_{t \downarrow 0} q(t)$ exists. By (i), $q(t)$ decreases as $t \downarrow 0$, so

$$\inf_{t > 0} q(t) = \lim_{t \downarrow 0} q(t) = f'(x; d)$$

(iii) $d \in \text{dom } f'(x; \cdot)$,

$$\begin{aligned} &\iff \exists t > 0 \quad \frac{f(x + td) - f(x)}{t} < \infty \\ &\iff \exists t > 0 \quad f(x + td) - f(x) \leq \infty \\ &\iff \exists t > 0 \quad x + td \in \text{dom } f \\ &\iff d \in \mathbb{R}_+(\text{dom } f - x) \end{aligned}$$

(iv) Take $0 < s < t$ and observe

$$q(-t) \leq q(-s) \leq q(s) \leq q(t)$$

If $x \in \text{int}(\text{dom } f)$, we can choose t sufficiently small so that $x \pm td \in \text{int}(\text{dom } f)$. Thus, $q(t)$ and $q(-t)$ are finite. Hence,

$$-\infty < q(-t) \leq q(-s) \leq f'(x; d) = \lim_{s \downarrow 0} q(s) \leq q(t) < \infty$$

And, since d was arbitrary, $f'(x; d)$ is finite valued. Check that $f'(x; d)$ is positively homogeneous. For convexity, let $t > 0, d, h \in \mathbb{E}$ and $\lambda \in [0, 1]$. By convexity of f ,

$$\begin{aligned} \frac{f(x + t(\lambda d + (1 - \lambda)h)) - f(x)}{t} &= \frac{f(\lambda(x + td) + (1 - \lambda)(x + th)) - f(x)}{t} \\ &\leq \frac{\lambda f(x + td) + (1 - \lambda)f(x + th) - (\lambda + (1 - \lambda))f(x)}{t} \\ &= \frac{\lambda(f(x + td) - f(x)) + (1 - \lambda)(f(x + th) - f(x))}{t} \end{aligned}$$

Now, take $t \downarrow 0$ to get directional derivatives.

□

Definition 7.15. For sublinear functions $h : \mathbb{E} \rightarrow \mathbb{R} \cup \{\infty\}$, we define the linearity space,

$$\text{lin } h = \{x \in \mathbb{E} \mid h(-x) = -h(x)\}$$

Lemma 7.16. Let $h : \mathbb{E} \rightarrow \mathbb{R} \cup \{\infty\}$ be sublinear, then the following hold,

(i) The linearity space $\text{lin } h$ of h is the subspace on which h is linear.

(ii) For $\bar{x} \in \text{int}(\text{dom } h)$ and $q = h'(\bar{x}; \cdot)$, we have

$$(a) \quad q(\lambda \bar{x}) = \lambda h(\bar{x}) \quad \forall \lambda \in \mathbb{R}$$

$$(b) \quad q \leq h$$

$$(c) \quad \text{lin } q > \text{lin } h + \text{span}\{\bar{x}\}$$

Proof. HWK. □

Theorem 7.17. Let $f \in \Gamma$, then the following hold,

(i) For all $\bar{x} \in \mathbb{E}$,

$$\partial f(\bar{x}) = \{s \mid \langle s, d \rangle \leq f'(\bar{x}; d)\} \quad \forall d$$

(ii) For all $\bar{x} \in \text{int}(\text{dom } f)$, we have

$$f'(\bar{x}; d) = \max_{s \in \partial f(\bar{x})} \langle s, d \rangle$$

In particular, $(\partial f)(\bar{x})$ is non empty.

Proof. (i) is in HWK.

(ii) Let $d \in \mathbb{E} = \mathbb{R}^n$, by part (i), we need to show that $\exists s \in \partial f(\bar{x})$ such that $\langle s, d \rangle = f'(\bar{x}; d)$. Now let $\{b_1, \dots, b_n\}$ be a basis of \mathbb{E} with $b_1 = d$ if $d \neq 0$. If $d = 0$, then any basis of \mathbb{R}^n . We define p_0, p_1, \dots, p_n by

$$p_0 = f'(\bar{x}; \cdot), \quad p_k = p'_{k-1}(b_k; \cdot) \quad \forall k = 1, \dots, n$$

By proposition 7.14, p_k is finite and sublinear.

$$\text{lin } p_k > \text{lin } p_{k-1} + \text{span}\{b_k\} \quad k = 1, \dots, n$$

Thus,

$$\text{lin } p_n = \mathbb{E},$$

i.e. by lemma 7.16 part (a), p_n is linear on \mathbb{E} . Therefore, there exists an s such that $p_n(\cdot) = \langle s, \cdot \rangle$ (Riesz representation Theorem). Part (ii) of lemma 7.16, (b) implies

$$\langle s, \cdot \rangle = p_n \leq p_{n-1} \leq \dots \leq p_0 = f'(\bar{x}; \cdot),$$

Hence,

$$\langle s, x - \bar{x} \rangle = p_n(x - \bar{x}) \leq p_0(x - \bar{x}) = f'(\bar{x}, x - \bar{x}) \stackrel{7.14}{\leq} f(\bar{x} + (x - \bar{x})) - f(\bar{x}) = f(x) - f(\bar{x})$$

for any x , thus $s \in \partial f(\bar{x})$. If $d = 0$, then $f'(\bar{x}; 0) = \langle s, 0 \rangle$. If $d \neq 0$, i.e.

$b_1 = d$, then by lemma 7.16 (ii)(a), and definition of p_1 , we get

$$\begin{aligned} p_n(d) &\leq p_0(d) = p_0(b_1) = -p_1(-b_1) = -p_1(-d) \\ &\leq -p_n(-d) = p_n(d) \\ \implies p_n(d) &= p_0(d) \\ \implies \langle s, d \rangle &= f'(\bar{x}; d) \end{aligned}$$

□

7.3 Continuity Properties

Main result: The Subdifferential operator $\partial f : \mathbb{E} \rightrightarrows \mathbb{E}$ of a closed, proper and convex function has a closed graph

$$\text{gph } \partial f := \{(x, y) \in \mathbb{E} \times \mathbb{E} \mid y \in \partial f(x)\},$$

which is referred to as the **outer semi-continuity** of f .

Corollary 7.17.1. Let $f \in \Gamma_0$ and suppose $\{x_n\} \rightarrow x$, and $\{y_n \in \partial f(x_n)\} \rightarrow y$. Then, $y \in \partial f(x)$, i.e. $\text{gph } \partial f \in \mathbb{E} \times \mathbb{E}$ is closed.

Theorem 7.18. A function $f \in \Gamma$ is locally **Lipschitz** at every $x \in \text{int}(\text{dom } f)$.

Proposition 7.19. Let $F : \mathbb{E}_1 \rightarrow \mathbb{E}_2$ be locally **Lipschitz** at every point of the compact set C , i.e., for every $\bar{x} \in C$ there exists $L = L(\bar{x})$, $\varepsilon = \varepsilon(\bar{x}) > 0$ such that

$$\|F(x) - F(y)\| \leq L\|x - y\|, \quad x, y \in B_\varepsilon(\bar{x}).$$

Then, F is globally Lipschitz on C , i.e., there exists $K > 0$ such that

$$\|F(x) - F(y)\| \leq K\|x - y\|, \quad x, y \in C.$$

Definition 7.20 (Domain of Continuity). The **Domain of Continuity** of a function $f : \mathbb{E} \rightarrow \overline{\mathbb{R}}$ is defined as,

$$\text{cont}\{x \in \mathbb{E} \mid f(x) \in \mathbb{R} \text{ and } f \text{ is cts at } x\}$$

Corollary 7.20.1 (Domain of Continuity of Convex Functions). Let $f \in \Gamma$, then we have

$$\text{cont } f = \text{int}(\text{dom } f)$$

Proof. The forward inclusion holds for any function (HWK). The reverse follows from Theorem 7.18. \square

Theorem 7.21 (Boundedness Properties of ∂f). For $f \in \Gamma$, the following hold,

- (i) Let $X \subset \text{int}(\text{dom } f)$ be non empty, open, and convex. Then f is *Lipschitz* continuous with modulus $L > 0$ on X if and only if $\|v\| \leq L \forall x \in X \text{ and } v \in \partial f(x)$.
- (ii) ∂f maps bounded sets which are compactly contained in $\text{int dom } f$ to bounded sets.

Corollary 7.21.1. Let $f \in \Gamma$ and $\bar{x} \in \text{int}(\text{dom } f)$, then $\partial f(\bar{x})$ is non empty, compact, and convex.

7.4 Differentiable Convex Functions

Main Result: A convex function is differentiable at a point in the interior of its domain if and only if its subdifferential is a singleton (i.e. the gradient).

Theorem 7.22. Let $f \in \Gamma$, and $x \in \text{int}(\text{dom } f)$, then $\partial f(x)$ is a singleton if and only if f is differentiable at x . In this case, $\partial f(x) = \{\nabla f(x)\}$

Theorem 7.23. Let $f \in \Gamma$, and $x \in \text{int}(\text{dom } f)$, then f is continuously differentiable on $\text{int}(\text{dom } f)$ if and only if $\partial f(x)$ is a singleton for every $x \in \text{int}(\text{dom } f)$.

Corollary 7.23.1. Let $f \in \Gamma$, then f is differentiable if and only if f is continuously differentiable.

7.5 Second Order Characterization

Theorem 7.24. Let $f : \mathbb{E} \rightarrow \overline{\mathbb{R}}$ be twice differentiable on the open convex set $\Omega \in \text{int}(\text{dom } f)$. Then the following hold

- (i) f is convex on Ω if and only if $\nabla^2 f(x)$ is *positive semi-definite* for all $x \in \Omega$.
- (ii) If $\nabla^2 f(x)$ is *positive definite* for all $x \in \Omega$, then f is strictly convex on Ω .
- (iii) f is σ -strongly convex on Ω if and only if for all $x \in \Omega$, the smallest eigenvalue of $\nabla^2 f$ is bounded by σ from below. Namely, $\nabla^2 f \succeq \sigma I$.

Corollary 7.24.1. Let $f : \mathbb{R}^n \rightarrow \mathbb{R}$ be defined by

$$f(x) = \frac{1}{2}x^T A x + b^T x + \gamma,$$

with $A \in \mathbb{S}^n$, $b \in \mathbb{R}^n$, $\gamma \in \mathbb{R}$. Then,

- (i) f is convex if and only if $A \succeq 0$.
- (ii) f is strongly convex $\iff f$ is strictly convex $\iff A \succ 0$.

7.6 Subdifferential Calculus

Theorem 7.25. For all $i \in I = \{1, 2, \dots, m\}$, let

$$f_i \in \Gamma, x \in \bigcap_{i \in I} \text{int}(\text{dom } f),$$

and set

$$f = \max_{i \in I} f_i, \text{ and } I(x) = \{i \in I \mid f_i(x) = f(x)\}.$$

Then,

$$\partial f = \text{conv} \bigcup_{i \in I(x)} \partial f_i(x)$$

Corollary 7.25.1. For all $i \in I = \{1, 2, \dots, m\}$, let

$$f_i \in \Gamma \text{ be differentiable at } x \in \bigcap_{i \in I} \text{int}(\text{dom } f),$$

and set

$$f = \max_{i \in I} f_i, \text{ and } I(x) = \{i \in I \mid f_i(x) = f(x)\}.$$

Then,

$$\partial f = \text{conv}\{\nabla f_i(x) \mid i \in I(x)\}$$

Our goal is to establish a subdifferential rule for convex functions of the form $f + g \circ L$, with $f \in \Gamma(\mathbb{E}_1)$, $g \in \Gamma(\mathbb{E}_2)$, $L \in \mathcal{L}(\mathbb{E}_1, \mathbb{E}_2)$.

Theorem 7.26 (Fenchel-Rockafellar Duality). Let $f \in \Gamma(\mathbb{E}_1)$, $g \in \Gamma(\mathbb{E}_2)$, $L \in \mathcal{L}(\mathbb{E}_1, \mathbb{E}_2)$, then the following hold,

Weak Duality. We have

$$p := \inf_{x \in \mathbb{E}_1} \{f(x) + g(Lx)\} \geq \sup_{y \in \mathbb{E}_2} \{-f^*(L^*(y)) - g^*(-y)\} := d \quad (2)$$

Strong Duality. If f and g are convex and the qualification condition

$$0 \in \text{int}(\text{dom } g - L(\text{dom } f)) \quad (3)$$

is satisfied, then equality holds in 2, and the supremum is attained if it is finite.

Primal-Dual Recovery. If $f \in \Gamma_0(\mathbb{E}_1)$ and $g \in \Gamma_0(\mathbb{E}_2)$, the following are equivalent for $\bar{x} \in \mathbb{E}_1$, $\bar{y} \in \mathbb{E}_2$,

- (a) $p = d$, $\bar{x} \in \text{argmin}\{f(x) + g(L(x))\}$, $\bar{y} \in \text{argmax}\{-f^*(L^*(y)) - g^*(-y)\}$
- (b) $L^*(\bar{y}) \in \partial f(\bar{x})$, $-\bar{y} \in \partial g(L(\bar{x}))$
- (c) $\bar{x} \in \partial f^*(L^*(\bar{y}))$, $L(\bar{y}) \in \partial g^*(-\bar{y})$

Strong duality and Primal-dual recovery say that under some conditions, optimizing the primal is the same as solving the dual and recovering the solution.

Remark 7.27. The qualification condition (3) is satisfied under the stronger condition

$$L(\text{dom } f) \cap \text{dom } g \neq \emptyset$$

Remark 7.28. We call $\inf_{x \in \mathbb{E}_1} \{f(x) + g(Lx)\}$ the **Primal Problem**, and $\sup_{y \in \mathbb{E}_2} \{-f^*(L^*(y)) - g^*(-y)\}$ the **Dual Problem**.

Example 7.29. Consider the primal problem,

$$\min_x \langle c, x \rangle + g(b - L(x)) + f(x), \quad b \in \mathbb{E}_2, \quad c \in \mathbb{E}_1$$

then by replacing f with $f + \langle c, \cdot \rangle$ and g with $g(-\cdot)$, we may put the problem in the form stated in 7.26. Thus, the dual is

$$\max_y \langle b, y \rangle f^*(L^*(y) - c) - g^*(y)$$

and the qualification condition becomes $b \in L(\text{int dom } f) + \text{int dom } g$.

Theorem 7.30 (Generalized Sum Rule). Let $f : \mathbb{E}_1 \rightarrow \overline{\mathbb{R}}, g : \mathbb{E}_2 \rightarrow \overline{\mathbb{R}}$, and $L \in \mathcal{L}(\mathbb{E}_1, \mathbb{E}_2)$, then the following hold

- (i) $\partial(f + g \circ L)(\bar{x}) \supset \partial f(\bar{x}) + L^*[\partial g(L(\bar{x}))]$, $\bar{x} \in \mathbb{E}_1$
- (ii) if $f \in \Gamma(\mathbb{E}_1)$, $g \in \Gamma(\mathbb{E}_2)$, and condition (3) is satisfied, then equality holds in (i).

Theorem 7.31 (Subdifferential Sum Rule). Let $f, g \in \Gamma$, then

$$\partial(f + g)(x) \supset \partial f(x) + \partial g(x) \quad \forall x \in \mathbb{E}$$

Under the condition $0 \in \text{int}(\text{dom } f - \text{dom } g)$, equality holds.

Proposition 7.32 (Subdifferential Chain Rule). Let $g \in \Gamma(\mathbb{E}_2)$, $L \in \mathcal{L}(\mathbb{E}_1, \mathbb{E}_2)$, then

$$\partial(g \circ L) \supset L^*(\partial g) \circ L$$

Under the condition $0 \in \text{int}(\text{dom } f - \text{dom } g)$, equality holds.

Theorem 7.33 (Subdifferential of Infimal Convolution). Let $f, g \in \Gamma_0$ and $x \in \text{dom}(f \# g)$, then

$$\partial(f \# g)(\bar{x}) = \partial f(\bar{u}) \cap \partial g(\bar{x} - \bar{u}) \quad \forall \bar{u} \in \text{argmin}_u \{f(u) + g(\bar{x} - u)\}$$

Corollary 7.33.1 (Derivative of Moreau Envelope). Let $f \in \Gamma_0$ and $\lambda > 0$, then $e_\lambda f$ is continuously differentiable with $\frac{1}{\lambda}$ -Lipschitz gradient given by

$$\nabla e_\lambda f = \frac{1}{\lambda}(I - P_\lambda f)$$

Theorem 7.34 (Moreau Decomposition). For any proper, convex, closed function $f : \mathbb{E} \rightarrow \overline{\mathbb{R}}$, the following equality holds

$$\text{prox}_f(x) + \text{prox}_{f^*}(x) = x \quad \forall x \in \mathbb{E}$$

Proposition 7.35. For any proper, convex, closed function $f : \mathbb{E} \rightarrow \overline{\mathbb{R}}$ and $g : \mathbb{E} \rightarrow \mathbb{R}$ continuously differentiable, then

$$\partial(f + g)(x) = \partial f(x) + \nabla g(x) \quad \forall x \in \mathbb{E}$$

References

Stephen D. Boyles. *Convex Examples*. Spring 2015.

Markus Grasmair. *BASIC PROPERTIES OF CONVEX FUNCTIONS*.

Norwegian University of Science and Technology. URL https://wiki.math.ntnu.no/_media/tma4180/2016v/note2.pdf.