MATH563: Honours Convex Optimization

Matthew He

March 21, 2026

This note is adopted from the course material of MATH563 Honours Convex Optimization (Winter 2026) by Prof. Courtney Paquette.
It covers all theoretical content taught in the course and required for the exam.

From a high-level perspective, the course focuses on understanding convexity from both geometric and linear-algebraic perspectives. A central lesson is developing the language and intuition required to work in abstract (potentially high-dimensional) inner product spaces, particularly Euclidean spaces. The course introduces key tools in convex optimization, including proximal operators, subdifferentials, Fenchel conjugates, and duality.

Prerequisites: For the honours version of the course, proficiency in metric space analysis (MATH255) is required. For non-honours version, familiarity with matrix calculus, abstract linear algebra, and nonlinear optimization will be beneficial.
Although, registration does not enforce any of these prerequisites.

1 *Preliminaries*

Definition 1 (Affine set). *A set $C \subseteq \mathbb{E}$ is called an affine set if for all $x, y \in C$ and $\lambda, \mu \in \mathbb{R}$, we have*

$$\lambda x + \mu y \in C. \tag{1}$$

Definition 2 (Affine hull). *The affine hull of a set $Q \subseteq \mathbb{E}$ denoteded by $aff(Q)$, is the intersection of all affine sets that contain Q.*

Affine hull is the smallest affine set that contains Q.

Definition 3 (Relative interior and boundary). *The relative interior of a set $Q \subseteq \mathbb{E}$, denoted by $ri(Q)$, is the interior of Q relative to its affine hull, i.e.*

$$ri(Q) = \{x \in Q : \exists \delta > 0, B(x,\delta) \cap aff(Q) \subseteq Q\}. \tag{2}$$

The relative boundary of Q is defined as $rb(Q) = cl(Q) \setminus ri(Q)$.

Relative interior is weaker than interior since it's easier to be satisfied, as $B_\varepsilon \cap \mathrm{aff} Q \subset B_\varepsilon$. We have $\mathrm{int}(Q) \subseteq \mathrm{ri}(Q)$ and $\mathrm{rb}(Q) \subseteq \mathrm{bd}(Q)$.

Definition 4 (Epigraph). *The epigraph of a function $f : \mathbb{E} \to \overline{\mathbb{R}}$ is the set*

$$epi(f) = \{(x, \alpha) \in \mathbb{E} \times \mathbb{R} : f(x) \leq \alpha\}. \tag{3}$$

We review the set operations that will be frequently used in the following sections:

Definition 5 (Minkowski sum). *Minkowski sum of sets: For $A, B \subseteq \mathbb{E}$, we define the Minkowski sum of A and B by*

$$A + B = \{a + b : a \in A, b \in B\}. \tag{4}$$

Proposition 1 (Minkowski sum of subspaces). *For subspaces $V, W \subseteq \mathbb{E}$, we have $V + W = span(V \cup W)$.*

Example 1 (***Common set algebra***)

1. *Translation:* $A + \{x\} = \{a + x : a \in A\}$
   It implies that $s \in (A + x) \iff s - x \in A$.
2. *Dilation:* $B(0, r_1) + B(0, r_2) = B(0, r_1 + r_2)$
3. *Smoothing;* $A + B(0, r)$ *has rounded corners.*

Definition 6 (Cone operator). *For $A \subseteq \mathbb{E}$, the cone operator is defined such that* $\mathbb{R}_+(A) = \{\lambda a : \lambda \geq 0, a \in A\}$.

Remark 1. *For the cone operator, we have the identity that*

$$x \in \mathbb{R}_+(A) \iff \exists \lambda > 0 texts.t. \frac{x}{\lambda} \in A \quad (5)$$

$$x \in \mathbb{R}_+(A) \setminus \{0\} \iff \frac{x}{\|x\|} \in \mathbb{R}_+(A) \quad (6)$$

Definition 7 (Polar Cone). *Let $C \subseteq \mathbb{E}$ be a cone. The **polar cone** of C, denoted C°, is defined as:*

$$C^\circ = \{s \in \mathbb{E} : \langle s, v \rangle \leq 0, \forall v \in C\}. \quad (7)$$

Definition 8 (Normal Cone). *Let $C \subseteq \mathbb{E}$ be a convex set and $x \in C$. The **normal cone** to C at x, denoted $N_C(x)$, is defined as:*

$$N_C(x) = \{s \in \mathbb{E} : \langle s, v - x \rangle \leq 0, \forall v \in C\}. \quad (8)$$

1.1 Lower semicontinuity

Definition 9 (Lower limit). *For $f : \mathbb{E} \to \mathbb{R}$, and $\bar{x} \in \mathbb{E}$, we define*

$$\liminf_{x \to \bar{x}} f(x) = \sup_{\delta > 0} \inf_{x \in B(\bar{x}, \delta)} f(x) \quad (9)$$

$$= \inf\{\alpha \in \overline{\mathbb{R}} : \exists x^k \to x \text{ s.t. } f(x^k) \to \alpha\}. \quad (10)$$

as the lower limit of f at $\bar{x}$.

Of course there are a lot more equivalent definition and characterization of liminf.

Definition 10 (Continuity notions). *Let $f : \mathbb{E} \to \overline{\mathbb{R}}$ and $\bar{x} \in \mathbb{E}$, then f is said to be lower semicontinuous at $\bar{x}$ if $f(\bar{x}) \leq \liminf_{x \to \bar{x}} f(x)$.*

Alternatively, f is lsc at $\bar{x}$ iff there does not exists a sequence $\{x^k\} \to \bar{x}$ such that $f(x^k) < f(\bar{x})$.

Definition 11 (Closure of function). *For $f : \mathbb{E} \to \overline{\mathbb{R}}$, the closure of f is the function $clf : \mathbb{E} \to \overline{\mathbb{R}}$ defined by*

$$clf(\bar{x}) = \liminf_{x \to \bar{x}} f(x). \quad (11)$$

As $\liminf_{x \to \bar{x}} f(x) \leq f(\bar{x})$, we always have $\text{cl}f \leq f$. We also have $\text{epi}(\text{cl}f) = \text{cl}(\text{epi}f)$. That said, epigraph and closure commute with each other.

Proposition 2 (Characterization of lsc)**.** *Let $f : \mathbb{E} \to \overline{\mathbb{R}}$. Then TFAE:*

1. *f is lsc on $\mathbb{E}$*
2. *epi(f) is closed*
3. *lev$_{\leq \alpha} f$ is closed for all $\alpha \in \mathbb{R}$*

Corollary 1 (lsc of indicator function)**.** *Let $C \subseteq \mathbb{E}$, its indicator function δ_C is proper, lsc iff C is nonempty and closed.*

Remark 2 (Closedness of a positive combination)**.** *For $p \in \mathbb{N}$, let $f_1, \ldots, f_p$ be lsc and $\alpha_i \geq 0$ for $i = 1, \ldots, p$. Then $f = \sum_{i=1}^{p} \alpha_i f_i$ is lsc.*

1.2 *Optimization problems*

We always have

$$\inf_C f = -\sup_C(-f), \ \sup_C f = -\inf_C(-f). \tag{12}$$

Definition 12 (Minimizer)**.** *Let $f : \mathbb{E} \to \mathbb{R} \cup \{\infty\}$ and $C \subseteq \mathbb{E}$. Then $\bar{x} \in C \cap dom(f)$ is called*

1. *local minimizer of f over C if*

$$\exists \varepsilon > 0 : f(\bar{x}) \leq f(x), \forall x \in C \cap B(\bar{x}, \varepsilon). \tag{13}$$

2. *global minimizer of f over C if*

$$f(\bar{x}) \leq f(x), \forall x \in C. \tag{14}$$

When minimizing a proper function $f : \mathbb{E} \to \mathbb{R} \cup \{\infty\}$, we have:

$$\arg\min_{\mathbb{E}} f \neq \emptyset \implies \inf_{\mathbb{E}} f \in \mathbb{R}. \tag{15}$$

but not the reverse.

Definition 13 (Coercivity and supercoercivity)**.** *Let $f : \mathbb{E} \to \overline{\mathbb{R}}$. Then f is called*

- *coercive if $\lim_{\|x\| \to \infty} f(x) = \infty$, which is sometimes called 0-coercive.*
- *supercoercive if $\lim_{\|x\| \to \infty} \frac{f(x)}{\|x\|} = \infty$, which is sometimes called 1-coercive.*

Lemma 1 (Level-boundedness and coercivity)**.** *A function $f : \mathbb{E} \to \overline{\mathbb{R}}$ is coercive iff f is level-bounded.*

Coercivity gives control of the function value at infinity, and hence guarantees the existence of the minimum. But coercivity alone does not guarantee the existence and uniqueness of the minimizer. We also need the function to be lsc to control the continuity, and thus the attainment of the minimum.

Theorem 1 (Existence of minima)**.** *Let $f : \mathbb{E} \to \mathbb{R} \cup \{\infty\}$ be proper, lsc, and level bounded. Then* $\arg\min f \neq \emptyset$ *and* $\min f = \inf f \in \mathbb{R}$.

We extend this to cover the problem of minimizing f over a closed set C. The closedness is required to give control on attainment of minimum at the boundary.

Corollary 2 (Existence of minimizer)**.** *Let $f : \mathbb{E} \to \mathbb{R} \cup \{\infty\}$ be **lower semicontinuous** and let $C \subseteq \mathbb{E}$ be closed such that $dom(f) \cap C \neq \emptyset$ and suppose one of the following holds:*

1. *f is coercive*
2. *C is bounded*

Then f has a minimizer on C.

Properness of f comes from definition that it cannot take value $-\infty$.

Proof: Consider the function $g = f + \delta_C$. Then it holds that $\text{lev}_{\leq \alpha} g = \text{lev}_{\leq \alpha} f \cap C$ for all $\alpha \in \mathbb{R}$.

Hence, under either assumption, g is closed and bounded level sets. Hence, it is lsc and level-bounded, Then the results follows from Theorem 1. □

An important corollary applies to the sum of functions.

Corollary 3 (Existence of minimizers II)**.** *Let $f, g : \mathbb{E} \to \mathbb{R} \cup \{\infty\}$ be lsc such that $dom(f) \cap dom(g) \neq \emptyset$. If f is coercive and g is bounded from below, then $f + g$ is coercive and has a minimizer over $\mathbb{E}$.*

2 Convex Sets

Definition 14 (Convex sets). *A set $C \subseteq \mathbb{E}$ is called convex if for all $x, y \in C$ and $\lambda \in [0,1]$, we have*

$$\lambda x + (1-\lambda)y \in C. \tag{16}$$

Example of convex sets includes: subspace of $\mathbb{E}$, Minkowski sum of convex sets, hyperplanes, affine sets, intersection of convex sets, linear image and preimage of convex sets, half spaces, interval (closed, open, half-open).

Remark 3 (Convexity preserving operations). *The following operations preserve convexity:*

- *(Intersection) Let I be an arbitrary index set (possibly uncountable) and $C_i \subseteq \mathbb{E}$ be convex for all $i \in I$, then $\bigcap_{i\in I} C_i$ is convex.*
- *(Linear image and preimage) Let $F \in \mathcal{L}(\mathbb{E}_1, \mathbb{E}_2)$ and $C \subseteq \mathbb{E}_1$, $D \subseteq \mathbb{E}_2$ be convex, then*

$$F(C) = \{Fx : x \in C\}, F^{-1}(D) = \{x : Fx \in D\} \tag{17}$$

  are convex.
- *(Minkowski sum) This implies translation of a convex set is convex. In particular, translation of subspace, which is called affine set, is convex.*

2.1 Projection on convex sets

Definition 15 (Projection on a set). *let $S \subset \mathbb{E}$ be nonempty and $x \in \mathbb{E}$, then we define the projection of x on S by*

$$P_S(x) = \arg\min_{y\in S} \|x - y\| \subset S \tag{18}$$

Note that there is no change if we substitute $\frac{1}{2}\|x - y\|^2$ for $\|x - y\|$ in the above definition.

Now we give sufficient conditions for this subset to be nonempty and unique.

Lemma 2. *Let $x \in \mathbb{E}$ and $S \subseteq \mathbb{E}$ be nonempty. Then the following hold:*

1. *If S is closed, then $P_S(x) \neq \emptyset$.*
2. *If S is convex, then $P_S(x)$ has at most one element.*

Corollary 4 (Projection on closed, convex set is unique). *Let $C \subset \mathbb{E}$ be nonempty, closed, and convex. Then $P_C(x)$ is a mapping from $\mathbb{E}$ to C with $x = P_C(x)$ iff $x \in C$. That said, projection onto a closed convex set is unique.*

The following theorem gives an important characterization of the projection on a closed convex set in terms of variational inequality.

Theorem 2 (Projection Theorem). *Let $C \subseteq \mathbb{E}$ be nonempty, closed, and convex, and let $x \in \mathbb{E}$. Then $\bar{v} = P_C(x)$ iff*

$$\langle x - \bar{v}, v - \bar{v}\rangle \leq 0, \quad \forall v \in C. \tag{19}$$

2.2 Separation Theorem

Definition 16 (Support Function). *For a nonempty set $C \subseteq \mathbb{E}$, the **support function** of C is the function $\sigma_C : \mathbb{E} \to \mathbb{R} \cup \{+\infty\}$ defined by*

$$\sigma_C(s) = \sup_{v\in C} \langle s, v\rangle. \tag{20}$$

Remark 4. *Support function probes the boundary of the set C in the direction of s. In particular, it's the largest shift of a hyperplane with normal vector s that still touches the set C.*

For a convex set C, the support function is the distance from the origin to the supporting hyperplane of C with normal vector s.

Definition 17 (Supporting Hyperplane)**.** *A hyperplane* $H = \{z \in \mathbb{E} : \langle s, z\rangle = \alpha\}$ *with* $s \neq 0$ *is said to* ***support*** *a set* C *at a point* $x \in cl(C)$ *if:*

1. $\langle s, v\rangle \leq \alpha$ *for all* $v \in C$ *(Set is in one half-space), and*

2. $\langle s, x\rangle = \alpha$ *(The plane actually touches the set at* x*).*

Note that these conditions combined imply that $\alpha = \sigma_C(s)$*.*

That said, support function returns the distance from the origin to the supporting hyperplane of C with normal vector s.

Remark 5 (Supporting hyperplane of convex set)**.** *Let* $C \subset \mathbb{E}$ *be a convex set and* x *in the boundary of* C*. Then there exists a non-zero vector* $s \in \mathbb{E} \setminus \{0\}$ *such that*

$$\langle s, x\rangle = \sigma_C(s) = \sup_{v \in C} \langle s, v\rangle. \tag{21}$$

The set $H = \{z \in \mathbb{E} : \langle s, z\rangle = \langle s, x\rangle\}$ *is a* ***supporting hyperplane*** *to* C *at* x*.*

Theorem 3 (Separation Theorem)**.** *Let* $C \subset \mathbb{E}$ *be nonempty, closed, and convex, and let* $x \notin C$*, then* ***there exists*** $s \in \mathbb{E} \setminus \{0\}$ *with*

$$\langle s, x\rangle > \sup_{v \in C} \langle s, v\rangle. \tag{22}$$

Substituting s *with* $-s$*, we also obtain*

$$\langle s, x\rangle < \inf_{v \in C} \langle s, v\rangle. \tag{23}$$

To see the separation, define $\gamma = \frac{1}{2}(\langle s, x\rangle + \sup_{y \in C}\langle s, y\rangle)$. Then, $x \in \{z : \langle s, z\rangle > \gamma\}$ and $C \subseteq \{z : \langle s, z\rangle < \gamma\}$, i.e., x and C lie in two distinct half-spaces defined by the hyperplane

$$H = \{z : \langle s, z\rangle = \gamma\}. \tag{25}$$

Remark 6. *The separation theorem implies that for* $C \subset \mathbb{E}$ *nonempty, closed, and convex and* $x \notin C$*, there exists* $s \in \mathbb{E} \setminus \{0\}$ *and* $\beta \in \mathbb{R}$ *such that*

$$\langle s, x\rangle > \beta > \sup_{v \in C} \langle s, v\rangle. \tag{24}$$

Moreover, this inequality can be flipped entirely by substituting s *with* $-s$*. So we can choose whichever direction that's more convenient.*

By positive homogeneity, we can assume $\|s\| = 1$ in the above theorem.

3 Convex Functions

3.1 Definition and characterization

Definition 18 (Convex functions). *A function $f : \mathbb{E} \to \overline{\mathbb{R}}$ is called convex if epi(f) is a convex set.*

We can substitute the the epigraph with strict epigraph.

Convex function also have convex level sets.

Examples of convex functions: scalar-valued linear functions $\mathcal{L}(\mathbb{E}_1, \mathbb{R})$ is convex; indicator of function of a convex set, any norm on $\mathbb{E}$, the scalar functions $x \mapsto x^2$, $x \mapsto \exp(x)$.

Proposition 3 (Domain of a convex function). *The domain of a convex function is convex.*

Proof:Using the projection linear mapping $L : (x, \alpha) \in \mathbb{E} \times \mathbb{R} \mapsto x \in \mathbb{E}$, we have $\text{dom}(f) = L(\text{epi}(f))$. We know that linear map preserve convexity.

Remark 7. *This implies convex function have* ***connected domain****.*

We further characterize the convex level set property:

Remark 8 (Quasiconvex functions). *A function is called quasiconvex if all its sublevel sets are convex. We have the following hold:*

1. *Every convex function is quasiconvex*
2. *$f : \mathbb{E} \to \mathbb{R} \cup \{\infty\}$ is quasiconvex then $\arg\min f$ is a convex set.*

Proposition 4 (Characterizing convexity). *A function $f : \mathbb{E} \to \mathbb{R} \cup \{\infty\}$ is convex iff for all $x, y \in \mathbb{E}$ and $\lambda \in (0, 1)$, we have*

$$f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y). \tag{26}$$

We can extend the above characterization.

Definition 19 (p-dimensional simplex). *The p-dimensional simplex is the set*

$$\Delta_p = \{\lambda \in \mathbb{R}^p : \lambda_i \geq 0, \sum_{i=1}^{p} \lambda_i = 1\}. \tag{27}$$

Remark 9. *Let $f : \mathbb{E} \to \mathbb{R} \cup \{\infty\}$, then TFAE:*

1. *f is convex*
2. *f is convex on its domain*

Convexity on the domain is equivalent to convexity on the whole space because the inequality trivially holds when x or y is outside the domain.

Corollary 5 (Jensen's Inequality). *Let $f : \mathbb{E} \to \mathbb{R} \cup \{\infty\}$ be a convex function, and let $x_1, \ldots, x_n \in$ dom(f) with $\lambda_1, \ldots, \lambda_n \geq 0$ such that $\sum_{i=1}^{n} \lambda_i = 1$. Then*

$$f\left(\sum_{i=1}^{n} \lambda_i x_i\right) \leq \sum_{i=1}^{n} \lambda_i f(x_i). \tag{28}$$

Definition 20 (Convexity on a set). *A function $f : \mathbb{E} \to \overline{\mathbb{R}}$ is convex on a nonempty convex set $C \subseteq$ dom(f) if the inequality (26) holds for every $x, y \in C$.*

We mainly interest in proper convex (sometimes lsc) functions $f : \mathbb{E} \to \mathbb{R} \cup \{\infty\}$. We take the following notation:

- $\Gamma = \Gamma(\mathbb{E}) = \{f : \mathbb{E} \to \mathbb{R} \cup \{\infty\} : f \text{ is proper and convex}\}$
- $\Gamma_0 = \Gamma_0(\mathbb{E}) = \{f : \mathbb{E} \to \mathbb{R} \cup \{\infty\} : f \text{ is proper, convex, and lsc}\}$

Now we introduce the important results about affine minorization of convex functions, extending from the separation theorem of convex sets (Theorem 3).

Theorem 4 (Affine minorization principle). *For $f \in \Gamma_0$, there exists an affine minorant, i.e. there exists $a \in \mathbb{E}$ and $\beta \in \mathbb{R}$ such that*

$$f(x) \geq \langle a, x \rangle + \beta, \forall x \in \mathbb{E}. \tag{29}$$

Lastly, we introduce stronger notions of convexity.

Definition 21 (Strict/strong convexity). *Let $f \in \Gamma$ and $C \subset dom(f)$ a convex set, then f is said to be*

- *strictly convex if for all $x, y \in \mathbb{E}$ with $x \neq y$ and all $\lambda \in (0,1)$, we have*

$$f(\lambda x + (1-\lambda)y) < \lambda f(x) + (1-\lambda) f(y). \tag{30}$$

- *strongly convex with parameter $m > 0$ if for all $x, y \in \mathbb{E}$ and all $\lambda \in (0,1)$, we have*

$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda) f(y) - \frac{m}{2}\lambda(1-\lambda)\|x-y\|^2. \tag{31}$$

  where m is called the modulus of strong convexity on C

Note that strong convexity implies strict convexity, but not the reverse.

Proposition 5 (Characterization of strong convexity). *Let $f \in \Gamma$ and $C \subset dom(f)$. Then f is strongly convex with parameter $m > 0$ on C iff the function $x \mapsto f(x) - \frac{m}{2}\|x\|^2$ is convex on C.*

3.2 *Functional operations preserving convexity*

Proposition 6 (Positive combinations of convex functions). *For $p \in \mathbb{N}$, let $f_1, \ldots, f_p$ be convex (and lsc) and $\alpha_i \geq 0$ for $i = 1, \ldots, p$. Then*

$$f = \sum_{i=1}^{p} \alpha_i f_i \tag{32}$$

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

Proposition 7 (Pointwise supremum of convex functions). *For an aribitrary index set I, let f_i be convex (lsc) for all $i \in I$. Then the function*

$$f(x) = \sup_{i \in I} f_i(x) \ \forall x \in \mathbb{E} \tag{33}$$

is convex (lsc).

The index set does not have to be finite.

Proposition 8 (Parametric infimum)**.** *Let $f : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ be convex. Then*

$$g(x) = \inf_{y \in \mathbb{R}^n} f(x, y) \tag{34}$$

is convex.

Proposition 9 (Infimum convolution)**.** *Let $f, g : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ be convex. Then the function (studied later in Definition 23)*

$$h(x) = (f \# g)(x) = \inf_{y \in \mathbb{E}} f(y) + g(x - y) \tag{35}$$

is convex.

Definition 22 (Affine functions)**.** *A function $f : \mathbb{E}_1 \to \mathbb{E}_2$ is called affine if*

$$f(\lambda x + (1-\lambda)y) = \lambda f(x) + (1-\lambda)f(y), \forall x, y \in \mathbb{E}_1, \lambda \in [0, 1]. \tag{36}$$

Every affine function can have a unique representation $f = L + v$ where $L \in \mathcal{L}(\mathbb{E}_1, \mathbb{E}_2)$ is linear and $v \in \mathbb{E}_2$.

If the coefficients need not sum to 1 then it's linear. Linear function is affine and affine function is linear if $f(0) = 0$.

Proposition 10 (Pre-composition of with an affine mapping)**.** *Let $H : \mathbb{E}_1 \to \mathbb{E}_2$ be affine and $g : \mathbb{E}_2 \to \mathbb{R} \cup \{\infty\}$ (lsc and) convex. Then the function $f = g \circ H$ is (lsc and) convex.*

This is not true in general, as the composition of two convex functions may not be convex.

4 Minimization and convexity

In an minimization problem:

$$\min_{x\in C} f(x) \iff \min_{x\in\mathbb{E}} f(x)+\delta_C(x). \tag{37}$$

If f is convex and C is convex, then we call this problem a convex minimization/optimization problem.

4.1 Convex minimization

Under convexity, local minimizer implies global minimizer. Thus we can extent our results from general optimization problem in Corollary 2.

Proposition 11. *Let $f \in \Gamma$, then every local minimizer (over $\mathbb{E}$) is a global minimizer.*

Corollary 6 (Minimizers in convex optimization)**.** *Let $f \in \Gamma$ and $C \subseteq \mathbb{E}$ be nonempty and convex. Then every local minimizer of f over C is a global minimizer over C.*

Proposition 12. *Let $f \in \Gamma$, then $\arg\min f$ is a convex set.*

This is because $\mathrm{lev}_{\leq\alpha} f$ is convex for all $\alpha \in \mathbb{R}$.

Corollary 7. *Let $f \in \Gamma$ and $C \subseteq \mathbb{E}$ be convex. Then $\arg\min_C f$ is a convex set.*

Proposition 13 (Uniqueness of minimizer)**.** *Let $f \in \Gamma$ be strictly convex, then $\arg\min f$ has at most one element.*

Corollary 8 (Minimizing the sum of convex functions)**.** *Let $f, g \in \Gamma_0$ such that $dom(f) \cap dom(g) \neq \emptyset$. Suppose that one of the following holds:*

1. *f is super coercive*
2. *f is coercive and g is bounded from below*

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

Theorem 5 (Parametric minimization)**.** *Let $h : \mathbb{E}_1 \times \mathbb{E}_2 \to \mathbb{R} \cup \{\infty\}$ be jointly convex. Then the optimal value function*

$$\varphi : \mathbb{E}_1 \to \overline{\mathbb{R}},\ \varphi(x) = \inf_{y\in\mathbb{E}_2} h(x,y) \tag{38}$$

is convex. Moreover, the set-valued mapping

$$S : \mathbb{E}_1 \rightrightarrows \mathbb{E}_2,\ S(x) = \arg\min_{y\in\mathbb{E}_2} h(x,y) \tag{39}$$

is convex valued.

The second assertion follows from $y \mapsto h(x, y)$ is convex for all $x \in \mathbb{E}_1$.

5 Infimal Convolution of Convex Functions

5.1 The operation of infimal convolution

Definition 23 (Infimal convolution). *Let $f, g : \mathbb{E} \to \mathbb{R} \cup \{\infty\}$. The infimal convolution of f and g is the function*

$$(f\#g) : \mathbb{E} \to \overline{\mathbb{R}} \ (f\#g)(x) = \inf_{u \in \mathbb{E}} \{f(u) + g(x-u)\} \tag{40}$$

for all $x \in \mathbb{E}$. We call the infimal convolution $f\#g$ exact at x if

$$\arg\min_{u \in \mathbb{E}} \{f(u) + g(x-u)\} \neq \emptyset. \tag{41}$$

We call $f\#g$ exact if it is exact at every $x \in dom(f\#g)$.

We have the representation

$$(f\#g)(x) = \inf_{u,v \in \mathbb{E}} \{f(u) + g(v) : u + v = x\}. \tag{42}$$

Lemma 3. *Let $f, g : \mathbb{E} \to \mathbb{R} \cup \{\infty\}$. Then the following hold:*

1. *$dom(f\#g) = dom(f) + dom(g)$*
2. *$f\#g = g\#f$*

For Lemma 3 (1), an example is $g = \delta_{\{x_0\}}$, in this case, $(f\#g)(x) = f(x - x_0)$ since $g(x - x_0) = 0$ and $g(x-u) = \infty$ for all $u \neq x - x_0$. Then $\mathrm{dom}(f\#g) = \mathrm{dom}(f) + \{x_0\}$ is a translation.

Moreover, we have the trivial inequality $(f\#g) \leq f(u) + g(x-u)$ for all $u \in \mathbb{E}$.

Convexity is preserved under infimal convolution.

Proposition 14 (Infimal convolution of convex functions). *Let $f, g : \mathbb{E} \to \mathbb{R} \cup \{\infty\}$ be convex, then $f\#g$ is convex.*

However, lsc is not preserved under infimal convolution.

Theorem 6 (Infimal convolution of Γ_0). *Let $f, g \in \Gamma_0$ and suppose that one of the following conditions hold:*

1. *f is supercoercive*
2. *f is coercive and g is bounded from below*

Then, $f\#g \in \Gamma_0$ and the infimal convolution is exact.

5.2 Moreau envelopes and proximal mappings

Definition 24 (Moreau envelop and proximal mapping). *Let $f : \mathbb{E} \to \overline{\mathbb{R}}$. The Moreau envelope of f (to the parameter $\lambda > 0$) is the function $e_\lambda f : \mathbb{E} \to \overline{\mathbb{R}}$ defined by*

$$e_\lambda f(x) = \inf_{u \in \mathbb{E}} \{f(u) + \frac{1}{2\lambda}\|x - u\|^2\}. \quad (43)$$

The (possibly set-valued) mapping $P_\lambda f : \mathbb{E} \rightrightarrows \mathbb{E}$

$$P_\lambda f(x) = \arg\min_{u \in \mathbb{E}} \{f(u) + \frac{1}{2\lambda}\|x - u\|^2\} \quad (44)$$

is called the proximal mapping or prox-operator to the parameter λ of f.

$e_\lambda(\alpha f) = \alpha e_{\lambda/\alpha} f$ for all $\alpha, \lambda > 0$.

Proposition 15 (Coefficients in Moreau envelope and proximal mapping). *We have:*

$$e_\lambda f(x) = \frac{1}{\lambda} e_1(\lambda f)(x), \quad P_\lambda f(x) = P_1(\lambda f)(x). \quad (45)$$

Proposition 16. *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 and nonempty.*

Hence, when $f \in \Gamma_0$, we have:

$$e_\lambda f(x) = f(P_\lambda f(x)) + \frac{1}{2\lambda}\|x - P_\lambda f(x)\|^2 \quad (46)$$

$$\leq f(u) + \frac{1}{2\lambda}\|x - u\|^2 \ \forall u \in \mathbb{E}. \quad (47)$$

So the Moreau envelope is a minorant of f.

Proposition 17. *Let $f \in \Gamma_0$ and $x, p \in \mathbb{E}$. Then $p = P_1 f(x)$ iff $\langle y - p, x - p\rangle + f(p) \leq f(y)$ for all $y \in \mathbb{E}$. Namely,*

$$\langle y - p, x - p\rangle \leq f(y) - f(p), \forall y \in \mathbb{E}. \quad (48)$$

Remark 10. *Proposition 17 in fact gives:*

$$x - p \in \partial f(p). \quad (49)$$

for $p = P_1 f(x)$ and $f \in \Gamma_0$.

Proposition 18 (Firm nonexpansiveness of the proximal mapping). *Let $f \in \Gamma_0$ and $x, y \in \mathbb{E}$. Then*

$$\|P_1 f(x) - P_1 f(y)\|^2 \leq \langle P_1 f(x) - P_1 f(y), x - y\rangle. \quad (50)$$

By CSI, this can be extent to show that $P_1(x)$ is globally Lipschitz with constant 1,

$$\|P_1 f(x) - P_1 f(y)\| \leq \|x - y\|. \quad (51)$$

Compared to nonexpansiveness, firm nonexpansiveness is a stronger property because it rules out the possibility of rotation.

For $T : \mathbb{E} \to \mathbb{E}$ satisfies firm nonexpansiveness, we can substitute $I - T$ instead of T in the inequality.

Firm nonexpansiveness also implies globally 1-Lipschitz, and hence continuous.

6 Conjugate of Convex Functions

6.1 Convex hull of a function

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

Theorem 7 (Envelop representation of Γ_0). *Let $f \in \Gamma_0$. Then f is the pointwise supremum of all affine functions minorizing it*

$$f(x) = \sup\{\langle a, x\rangle + \beta : a \in \mathbb{E}, \beta \in \mathbb{R}, f(y) \geq \langle a, y\rangle + \beta, \forall y \in \mathbb{E}\}. \tag{52}$$

or simply $f(x) = \sup\{h(x) : h \leq f,\ h \text{ affine }\}$.

Definition 25 (Convex hull of a function). *Let $f : \mathbb{E} \to \overline{\mathbb{R}}$. Then the pointwise supremum of all convex functions minorizing f, i.e.*

$$conv(f)(x) = \sup\{g(x) : g \leq f,\ g \text{ convex }\} \tag{53}$$

is called the convex hull of f. Moreover, we define the closed convex hull of f by

$$\overline{conv}(f)(x) = \sup\{g(x) : g \leq f,\ g \text{ convex and lsc }\}. \tag{54}$$

We have $\overline{\text{conv}}(f) = \text{cl}(\text{conv}(f))$.

Since both convexity and lsc are preserved under pointwise supremum, we find that the convex hull is the largest convex function that minorizes f and the closed convex hull is the largest convex and lsc function that minorizes f.

Moreover we have the following:

Remark 11. *for a set $C \subset \mathbb{E}$, defining the convex hull of C as*

$$conv(C) = \bigcap\{C \subset S, S \text{ is convex}\}$$

and the closed convex hull of C as

$$\overline{conv}(C) = \bigcap\{C \subset S, S \text{ is closed and convex}\}$$

then we have the following hold:

$$\overline{conv}(C) = cl(conv(C)). \tag{55}$$

Corollary 9 (Affine envelope representation of closed convex hull). *Let $f : \mathbb{E} \to \overline{\mathbb{R}}$ be proper and minorized by an affine function. Then the closed convex hull of f is the pointwise supremum of all affine functions minorizing f, i.e.*

$$\overline{conv}(f)(x) = \sup\{\langle a, x\rangle + \beta : a \in \mathbb{E}, \beta \in \mathbb{R}, f(y) \geq \langle a, y\rangle + \beta, \forall y \in \mathbb{E}\}. \tag{56}$$

Proof of this just need to show that the closed convex hull is proper. □

6.2 Fenchel Conjugate

Definition 26 (Fenchel conjugate). *Let $f : \mathbb{E} \to \overline{\mathbb{R}}$. The Fenchel conjugate of f is the function $f^* : \mathbb{E} \to \overline{\mathbb{R}}$ defined by*

$$f^*(y) = \sup_{x \in \mathbb{E}}\{\langle x, y\rangle - f(x)\}. \tag{57}$$

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

When we fix a slope y, the function $y \mapsto f^*(y)$ tells what is the highest we can push the affine minorant with slope y without exceeding f.

Corollary 10 (Fenchel-Young inequality). *Let $f : \mathbb{E} \to \overline{\mathbb{R}}$. Then for all $x \in \mathbb{E}$ and $y \in \mathbb{E}$,*

$$f(x) + f^*(y) \geq \langle x, y \rangle. \tag{58}$$

We have $f \leq g \implies f^* \geq g^*$ and $f^{**} \leq f$.

Theorem 8 (Fenchel-Moreau Theorem). *Let $f : \mathbb{E} \to \overline{\mathbb{R}}$ be proper and have an affine minorant. Then the following hold:*

1. f^*, f^{**} *are closed, proper, and convex*
2. $f^{**} = \overline{conv}(f)$
3. $f^* = (conv(f))^* = (\overline{conv}(f))^*$

(3) means they ahve the same affine minorantnt.

Proposition 19 (Elementary cases of conjugacy). *Let $f : \mathbb{E} \to \overline{\mathbb{R}}$. Then the following hold:*

1. $(f - \langle a, \cdot \rangle)^* = f^*(\cdot + a)$
2. $(f + \gamma)^* = f^* - \gamma$ *for all* $\gamma \in \mathbb{R}$
3. $(\lambda f)^* = \lambda f^*(\cdot/\lambda)$ *for all* $\lambda > 0$
4. *Let $g : \mathbb{E} \to \overline{R}$, $h : \mathbb{E} \to \overline{R}$ proper, closed, and convex functions, and $A \in \mathcal{L}(\mathbb{E}, Y)$. Define $F(x, y) = h(Ax + y) + g(x)$, then*

$$F^*(x, y) = h^*(y) + g^*(-A^*y + x). \tag{59}$$

Proposition 20 (Conjugate of infimal convolution). *Let $f, g : \mathbb{E} \to \overline{\mathbb{R}}$. Then*

$$(f \# g)^* = f^* + g^*. \tag{60}$$

Moreover, suppose that $f, g \in \Gamma_0$ and the qualification condition (e.g $dom(f) \cap dom(g) \neq \emptyset$) holds, then

$$(f + g)^* = f^* \# g^*. \tag{61}$$

(Proving the second assertion requires Fenchel duality type argument in the next Chapter.)

7 Differential Theory

7.1 Subdifferential

Definition 27 (Subdifferential). *Let $f : \mathbb{E} \mapsto \overline{\mathbb{R}}$, and $\bar{x} \in \mathbb{E}$. Then $g \in \mathbb{E}$ is call a subgradient of f at $\bar{x}$ if*

$$f(x) \geq f(\bar{x}) + \langle g, x - \bar{x} \rangle, \quad \forall x \in \mathbb{E}. \tag{62}$$

The set of all subgradients of f at $\bar{x}$:

$$\partial f(\bar{x}) = \{g \in \mathbb{E} : f(x) \geq f(\bar{x}) + \langle g, x - \bar{x} \rangle, \forall x \in \mathbb{E}\} \tag{63}$$

is called the subdifferential of f at $\bar{x}$.

Subdifferential connects back with the affine minorization principle (Theorem 4), in particular, it contains exactly the slopes of all affine minorants of f at $\bar{x}$.

That said, $s \in \partial f(x)$ iff the normal vector $(s, -1)$ gives a supporting hyperplane to the epigraph of f at $(x, f(x))$:

$$\langle \begin{bmatrix} s \\ -1 \end{bmatrix}, \begin{bmatrix} \bar{x} \\ f(\bar{x}) \end{bmatrix} \rangle \geq \langle \begin{bmatrix} s \\ -1 \end{bmatrix}, \begin{bmatrix} x \\ \alpha \end{bmatrix} \rangle, \forall (x, \alpha) \in \text{epi}(f). \tag{64}$$

Unpacking the inner product, we have

$$\langle s, \bar{x} \rangle - f(\bar{x}) \geq \langle s, x \rangle - \alpha \iff \alpha \geq f(\bar{x}) + \langle s, x - \bar{x} \rangle \tag{66}$$

for all $(x, \alpha) \in \text{epi}(f)$.

Remark 12. *Notice in inequality* (66)*, the LHS less than $-f^*(s)$. So we have*

$$-f^*(s) \geq \langle s, x \rangle - \alpha \tag{65}$$

for all $(x, \alpha) \in epi(f)$.

In particular, we can plugin $(x, \alpha) = (x, f(x)) \in \text{epi}(f)$, then we recover the subgradient inequality

$$f(x) \geq f(\bar{x}) + \langle s, x - \bar{x} \rangle.$$

Proposition 21 (Elementary properties of the subdifferential). *Let $f : \mathbb{E} \mapsto \overline{\mathbb{R}}$. Then the following hold:*

1. *$\partial f(\bar{x})$ is closed and convex for all $\bar{x} \in dom(f)$.*
2. *If f is proper, then $\partial f(\bar{x}) = \emptyset$ for all $\bar{x} \notin dom(f)$.*
3. *$0 \in \partial f(\bar{x})$ iff $\bar{x} \in \arg\min f$.*
4. *If f is convex, then $\partial f(\bar{x}) = \{v \in \mathbb{E} : (v, -1) \in N_{epi(f)}(\bar{x}, f(\bar{x}))\}$.*

Theorem 9 (Subdifferential and conjugate function). *Let $f : \mathbb{E} \mapsto \overline{\mathbb{R}}$. Then the following are equivalent:*

1. $y \in \partial f(x)$.
2. $x \in \arg\max_z \{\langle z, y\rangle - f(z)\}$
3. $f(x) + f^*(y) = \langle x, y\rangle$ ***[Fenchel-Young equality]***

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

4. $x \in \partial f^*(y)$
5. $y \in \arg\max_z \{\langle x, z\rangle - f^*(z)\}$

Note that (1) and (4) are equivalent to ∂f^* is an inverse of ∂f.

Proof: Notice that

$$y \in \partial f(x) \tag{67}$$
$$\iff f(z) \geq f(x) + \langle y, z - x\rangle, \forall z \in \mathbb{E} \tag{68}$$
$$\iff \langle z, y\rangle - f(z) \leq \langle x, y\rangle - f(x), \forall z \in \mathbb{E} \tag{69}$$
$$\iff \langle x, y\rangle - f(x) \geq \sup_z\{\langle z, y\rangle - f(z)\} = f^*(y) \tag{70}$$

Combine with Fenchel-Young inequality $f(x) + f^*(y) \geq \langle x, y\rangle$, we have

$$\iff f(x) + f^*(y) = \langle x, y\rangle \tag{71}$$

Apply the same argument to f^* and using the fact that $f^{**} = f$ for $f \in \Gamma_0$, we complete the proof. □

When $f \in \Gamma_0$ and differentiable at x, then

$$\nabla f^*(x) = (\nabla f(x))^{-1} \tag{72}$$

where the RHS is the inverse function of ∇f.

7.2 Directional Derivative of a convex function

Definition 28 (Directional derivative). *Let $f : \mathbb{E} \mapsto \overline{\mathbb{R}}$ be proper, and $\bar{x} \in dom(f)$. We say that f is directionally differentiable at $\bar{x}$ in the direction $d \in \mathbb{E}$ if the limit*

$$f'(\bar{x}; d) = \lim_{t \downarrow 0} \frac{f(\bar{x} + td) - f(\bar{x})}{t}. \tag{73}$$

exists in an extended real-valued sense (so it can be $+\infty$). In this case, we call $f'(\bar{x}; d)$ the directional derivative of f at $\bar{x}$ in the direction d.

Remark: if f is differentiable, then $f'(\bar{x}; d) = \langle \nabla f(\bar{x}), d\rangle$.

Definition 29 (Positive Homogeneity). *We call a function $f : \mathbb{E} \mapsto \mathbb{R} \cup \{+\infty\}$ positively homogeneous if $f(\lambda x) = \lambda f(x)$ for all $x \in \mathbb{E}$ and $\lambda \geq 0$.*

Proposition 22. *A function $f : \mathbb{E} \mapsto \mathbb{R} \cup \{+\infty\}$ is convex and positively homogeneous if and only if*

$$h(\lambda x + \mu y) \leq \lambda h(x) + \mu h(y), \quad \forall x, y \in \mathbb{E}, \lambda, \mu \geq 0. \tag{74}$$

a property called sublinearity. Sublinearity is also equivalent to subadditivity $h(x + y) \leq h(x) + h(y)$ and positive homogeneity.

That said, convexity and positive homogeneity give subadditivity:

$$f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda) f(y), \quad \forall x, y \in \mathbb{E}, \lambda \in [0, 1]. \tag{75}$$

Set $\lambda = \frac{1}{2}$, we have

$$f\left(\frac{x+y}{2}\right) \leq \frac{f(x)+f(y)}{2} \implies f(x+y) \leq f(x)+f(y). \quad (76)$$

□

Observe that any positive homogeneous function has $h(0) = 0$ and any sublinear function satisfies $-h(x) \leq h(0) + h(-x) \leq h(x)$.

Proposition 23 (Directional derivative of a convex function). *Let $f \in \Gamma$, $x \in dom(f)$, and $d \in \mathbb{E}$. Define:*

$$t \in \mathbb{R} \setminus \{0\} \mapsto q(t) = \frac{f(x+td) - f(x)}{t}. \quad (77)$$

then the following hold:

1. *$q(-t) \leq q(-s) \leq q(s) \leq q(t)$ for all $0 < s < t$.*
2. *We have $f'(x;d) = \inf_{t>0} q(t)$. In particular, $f'(x;d)$ always exists in an extended real-valued sense.*
3. *$dom(f') = \mathbb{R}_+(dom(f) - x)$.*
4. *If $x \in int(dom(f))$, then $f'(x;\cdot)$ is finite-valued, positively homogeneous, and convex, (so sublinear).*

For notation: $\mathbb{R}_+(S) := \{\alpha s \mid s \in S, \alpha \geq 0\}$. $d \in \mathbb{R}_+(S) \iff \exists \alpha \geq 0, s \in S : d = \alpha s$.

Proof: (2): Since infima always exists in $\overline{\mathbb{R}}$, $\inf_{t>0} q(t)$ exists. By (1), $q(t)$ monotonically decreasing as $t \downarrow 0$, so we have

$$\inf_{t>0} q(t) = \lim_{t\downarrow 0} q(t) = f'(x;d). \quad (78)$$

(3): Follows from (2) we have:

$$\begin{aligned} d \in \text{dom}(f') &\iff \exists t > 0 : \frac{f(x+td) - f(x)}{t} < \infty. && (79)\\ &\iff \exists t > 0 : f(x+td) - f(x) < \infty. && (80)\\ &\iff \exists t > 0 : x + td \in \text{dom}(f), \text{ since } f(x) < \infty \text{ for } x \in \text{dom}(f). && (81)\\ &\iff \exists t > 0 : td \in \text{dom}(f) - x. && (82)\\ &\iff d \in \mathbb{R}_+(\text{dom}(f) - x). && (83) \end{aligned}$$

(4): Take $0 < s < t$. If $x \in \text{int}(\text{dom}(f))$, we can choose $t > 0$ sufficiently small such that both $x \pm td \in \text{int}(\text{dom}(f))$. Thus $q(t)$ and $q(-t)$ are finite. Hence $+\infty > q(t) \geq q(s) \downarrow f'(x;d) \geq q(-t) > -\infty$. As $d \in \mathbb{E}$ is arbitrary, $f'(x;\cdot)$ is finite-valued.

$f'(x,\cdot)$ is also positively homogeneous as:

$$f'(x;\alpha d) = \lim_{t\downarrow 0} \frac{f(x+t\alpha d)-f(x)}{t} \tag{84}$$

$$= \lim_{t\downarrow 0} \alpha \frac{f(x+t\alpha d)-f(x)}{t\alpha} \tag{85}$$

$$= \alpha f'(x;d). \tag{86}$$

Convexity of $f'(x;\cdot)$: for $t>0, d, h \in \mathbb{E}$ and $\lambda \in (0,1)$, by convexity of f we have

$$f(x+t(\lambda d+(1-\lambda)h)) - f(x) = f(\lambda(x+td)+(1-\lambda)(x+th)) - f(x) \tag{87}$$

$$\le \lambda(f(x+td)-f(x)) + (1-\lambda)(f(x+th)-f(x)). \tag{88}$$

which implies

$$f'(x;\lambda d+(1-\lambda)h) = \lim_{t\downarrow 0} \frac{f(x+t(\lambda d+(1-\lambda)h))-f(x)}{t} \tag{89}$$

$$\le \lim_{t\downarrow 0} \lambda \frac{f(x+td)-f(x)}{t} + (1-\lambda)\frac{f(x+th)-f(x)}{t} \tag{90}$$

$$= \lambda f'(x;d) + (1-\lambda) f'(x;h). \tag{91}$$

□

Definition 30 (Linearity space). *For a sublinear function $h : \mathbb{E} \mapsto \mathbb{R} \cup \{+\infty\}$, we call the set*

$$lin(h) = \{d \in \mathbb{E} : -h(d) = h(-d)\} \tag{92}$$

the linearity space of h.

Lemma 4 (Linearity space of a sublinear function). *Let $h : \mathbb{E} \mapsto \mathbb{R} \cup \{+\infty\}$ be sublinear. Then the following holds:*

1. *The linearity space $lin(h)$ is a subspace on which h is linear.*
2. *For $\bar{x} \in int(dom(h))$ and $q = h'(\bar{x},\cdot)$, we have:*
   (a) $q(\lambda \bar{x}) = \lambda h(\bar{x})$
   (b) $q \le h$
   (c) $linq \supset linh + span\{\bar{x}\}$

Theorem 10 (Subdifferential vs. directional derivative). *Let $f \in \Gamma$, then the following holds:*

1. *For all $\bar{x} \in \mathbb{E}$, we have $\partial f(\bar{x}) = \{s \in \mathbb{E} : \langle s, d\rangle \leq f'(\bar{x}, d), d \in \mathbb{E}\}$.*
2. *For all $\bar{x} \in int(dom(f))$, we have*

$$f'(\bar{x}; d) = \sup_{s \in \partial f(\bar{x})} \langle s, d\rangle, \quad \forall d \in \mathbb{E}. \tag{93}$$

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

Remark 13. *That said, for a proper convex function, the directional derivative is the support function (Definition 16) of its subdifferential.*

7.3 Continuity properties

The subdifferential operator of a closed, proper, and convex function has a closed graph.

Corollary 11 (Outer semicontinuity of convex subdifferential). *Let $f \in \Gamma_0$ and suppose $\{x_k\} \to x$ and $\{y_k \in \partial f(x_k)\} \to y$. Then $y \in \partial f(x)$, i.e. the graph of ∂f is closed.*

Outer semicontinuity provides control on the subdifferential. For example, the indicator function of a closed set is outer semicontinuous but not continuous.

Proof: The proof relies on that Fenchel-Young equality holds under a limit, we have $f(x_k) + f^*(y_k) = \langle x_k, y_k\rangle$ for all k. On the other hand, by lsc of f and f^*, we have

$$f(x) + f^*(y) \leq \liminf_{k\to\infty} f(x_k) + \liminf_{k\to\infty} f^*(y_k) \leq \liminf_{k\to\infty}(f(x_k) + f^*(y_k)) = \liminf_{k\to\infty}\langle x_k, y_k\rangle = \langle x, y\rangle. \tag{94}$$

$f(x) + f^*(y) \leq \langle x, y\rangle$. Together with Fenchel-Young inequality $f(x) + f^*(y) \geq \langle x, y\rangle$, we have $f(x) + f^*(y) = \langle x, y\rangle$. This implies $y \in \partial f(x)$. □

A proper convex function is locally Lipschitz continuous on the interior of its domain.

Definition 31 (Locally Lipschitz continuous). *A function $f : \mathbb{E} \mapsto \mathbb{R} \cup \{+\infty\}$ is locally Lipschitz continuous at $\bar{x} \in dom(f)$ if $\exists L_{\bar{x}}, \delta_{\bar{x}} > 0$ such that*

$$|f(x) - f(y)| \leq L_{\bar{x}}\|x - y\|, \quad \forall x, y \in B(\bar{x}, \delta_{\bar{x}}). \tag{95}$$

Remark: every function that is locally Lipschitz at a point is also continuous at that point.

Theorem 11 (Local Lipschitz continuity of a proper convex function). *Let $f \in \Gamma$ and $\bar{x} \in int(dom(f))$. Then f is locally Lipschitz continuous at $\bar{x}$.*

Remark 14. *This implies that the jump discontinuity of a proper convex function can only happen at the boundary of its domain.*

Moreover, if a function is locally Lipschitz at every point of a compact set, then it is globally Lipschitz on that set.

Definition 32 (Domain of continuity)**.** *For a function $f : \mathbb{E} \mapsto \mathbb{R} \cup \{+\infty\}$, we call the set*

$$\mathrm{cont}(f) = \{x \in \mathbb{E} : f \text{ is continuous at } x\} \tag{96}$$

the domain of continuity of f.

Corollary 12. *Let $f \in \Gamma$. Then $\mathrm{int}(\mathrm{dom}(f)) = \mathrm{cont}(f)$.*

Remark 15. *$\mathrm{cont}(f) \subset \mathrm{int}(\mathrm{dom}(f))$ always holds for all extended-real valued functions.*

Theorem 12 (Boundedness properties of the subdifferential)**.** *For $f \in \Gamma$, the following holds:*

1. *Let $X \subset \mathrm{int}(\mathrm{dom}(f))$ be nonempty, open, and convex. Then f is Lipschitz continuous with modulus $L > 0$ on X if and only if $\|v\| \leq L$ for all $v \in \partial f(x)$ and $x \in X$.*
2. *∂f maps bounded sets which are compactly contained in $\mathrm{int}(\mathrm{dom}(f))$ to bounded sets.*

Corollary 13. *Let $f \in \Gamma$ and $\bar{x} \in \mathrm{int}(\mathrm{dom}(f))$. Then $\partial f(\bar{x})$ is nonempty, compact, and convex.*

Proof: By elementary properties of the subdifferential, $\partial f(\bar{x})$ is closed and convex since $\bar{x} \in \mathrm{int}(\mathrm{dom}(f))$. By the boundedness properties of the subdifferential, since $\{\bar{x}\}$ is compactly contained in $\mathrm{int}(\mathrm{dom}(f))$, $\partial f(\bar{x})$ is bounded. Hence $\partial f(\bar{x})$ is compact. □

7.4 *Differentiable convex function*

Differentiability of a convex function is defined on an open set.

Theorem 13. *Let $f \in \Gamma$ and $\bar{x} \in \mathrm{int}(\mathrm{dom}(f))$. Then $\partial f(\bar{x})$ is a singleton iff f is differentiable at $\bar{x}$. In this case, $\partial f(\bar{x}) = \{\nabla f(\bar{x})\}$.*

Differentiability on an open set implies continuous differentiability on that set.

Theorem 14. *Let $f \in \Gamma$ and $\bar{x} \in \mathrm{int}(\mathrm{dom}(f))$. Then f is continuously differentiable at $\bar{x}$ iff ∂f is a singleton for all $x \in \mathrm{int}(\mathrm{dom}(f))$ and the mapping $x \mapsto \partial f(x)$ is continuous at $\bar{x}$.*

Corollary 14. *Let $f : \mathbb{E} \to \mathbb{R} \cup \{+\infty\}$ be convex. Then f if differentiable iff f is continuously differentiable.*

Second-order characterizations

Theorem 15. *Let $f : \mathbb{E} \to \mathbb{R} \cup \{+\infty\}$ be twice differentiable on the open convex set $\Omega \subset int(dom(f))$. Then the following hold:*

1. *f is convex on Ω iff $\nabla^2 f(x) \succeq 0$ for all $x \in \Omega$.*
2. *If $\nabla^2 f(x)$ is positive definite for all $x \in \Omega$, then f is strictly convex on Ω.*
3. *f is σ strongly convex on Ω iff for all $x \in \Omega$, the smallest eigenvalue of $\nabla^2 f(x)$ is bounded below by σ, i.e. $\nabla^2 f(x) \succeq \sigma I$.*

Corollary 15. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be defined by*

$$f(x) = \frac{1}{2} x^T A x + b^T x + c, \tag{97}$$

with $A \in \mathbb{R}^{n \times n}$ symmetric, $b \in \mathbb{R}^n$, and $c \in \mathbb{R}$. Then f is convex iff $A \succeq 0$, and f is σ-strongly convex iff f is strictly convex iff A is positive definite.

7.5 *Subdifferential calculus*

Proposition 24 (Scaling rule). *For $f : \mathbb{E} \mapsto \mathbb{R} \cup \{+\infty\}$ and $\lambda > 0$, we have*

$$\partial(\lambda f)(x) = \lambda \partial f(x). \tag{98}$$

Theorem 16. *For all $i \in I = 1, \ldots, m$, let $f_i \in \Gamma$ and $x \in \bigcap_{i=1}^m int(dom(f_i))$. Set $f = \max_{i \in I} f_i$ and $I(x) = \{i \in I : f_i(x) = f(x)\}$. Then*

$$\partial f(x) = conv\left(\bigcup_{i \in I(x)} \partial f_i(x)\right). \tag{99}$$

Corollary 16. *For all $i \in I = 1, \ldots, m$, let $f_i \in \Gamma$ be differentiable at $x \in \bigcap_{i=1}^m int(dom(f_i))$. Set $f = \max_{i \in I} f_i$ and $I(x) = \{i \in I : f_i(x) = f(x)\}$. Then*

$$\partial f(x) = conv\left(\bigcup_{i \in I(x)} \{\nabla f_i(x)\}\right). \tag{100}$$

Theorem 17 (Fenchel-Rockafellar duality). *Let $f : \mathbb{E}_1 \to \mathbb{R} \cup \{+\infty\}$ and $g : \mathbb{E}_2 \to \mathbb{R} \cup \{+\infty\}$ and $L \in \mathcal{L}(\mathbb{E}_1, \mathbb{E}_2)$. Then the followings hold:*

1. *(Weak duality) We have*

$$p := \inf_{x \in \mathbb{E}_1} \{f(x) + g(L(x))\} \geq \sup_{y \in \mathbb{E}_2} \{-f^*(L^*(y)) - g^*(-y)\} =: d. \tag{101}$$

2. *(Strong duality) If f, g are convex and the qualification condition $0 \in int(dom(g) - L(dom(f)))$ holds, then equality holdes, i.e, $p = d$, and the supremum is attained in finite.*

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

   (a) $p = d$, $\bar{x} \in \arg\min\{f(x) + g(L(x))\}$, *and* $\bar{y} \in \arg\max\{-f^*(L^*(y)) - g^*(-y)\}$.

   (b) $L^*(\bar{y}) \in \partial f(\bar{x})$, *and* $-\bar{y} \in \partial g(L(\bar{x}))$.

   (c) $\bar{x} \in \partial f^*(L^*(\bar{y}))$, *and* $L(\bar{x}) \in \partial g^*(-\bar{y})$.

We note that the qualification condition is satisfied under a stronger condition $\text{int}(\text{dom}(g)) \cap L(\text{dom}(f)) \neq \emptyset$, which is often easier to verify in practice.

We call $\inf_{x \in \mathbb{E}_1} \{f(x) + g(L(x))\}$ the primal problem, and $\sup_{y \in \mathbb{E}_2} \{-f^*(L^*(y)) - g^*(-y)\}$ the dual problem. Part (2), (3) of the theorem say under the qualification condition, solving the primal problem is equivalent to solving the dual problem, and we can get the $\arg\min, \arg\max$ from (3).

Fenchel duality gives us the sum rule for subdifferential:

Theorem 18 (Generalized sum rule). *For $f : \mathbb{E} \to \mathbb{R} \cup \{+\infty\}$ and $g : \mathbb{E} \to \mathbb{R} \cup \{+\infty\}$, and $L \in \mathcal{L}(\mathbb{E}, \mathbb{E})$, the following hold:*

1. $\partial(f + g \circ L)(x) \supset \partial f(x) + L^*(\partial g(L(x)))$ *for all* $x \in \mathbb{E}$.

2. *If $f \in \Gamma(\mathbb{E})$, $g \in \Gamma(\mathbb{E})$, and the qualification condition $0 \in int(dom(g) - L(dom(f)))$ holds, the equality holds, i.e. $\partial(f + g \circ L)(x) = \partial f(x) + L^*(\partial g(L(x)))$ for all $x \in \mathbb{E}$.*

Corollary 17 (Subdifferential sum rule). *For $f, g \in \Gamma$, then*

$$\partial(f + g)(x) \supset \partial f(x) + \partial g(x), \quad \forall x \in \mathbb{E}. \tag{102}$$

If the qualification condition holds, equality holds, i.e. $\partial(f + g)(x) = \partial f(x) + \partial g(x)$ for all $x \in \mathbb{E}$.

Corollary 18 (Subdifferential chain rule)**.** *Let* $g \in \Gamma(\mathbb{E}_2)$ *and* $L \in \mathcal{L}(\mathbb{E}_1, \mathbb{E}_2)$*. Then* $\partial(g \circ L)(x) \supset L^*(\partial g(L(x)))$*. Under the qualification condition, the equality holds, i.e.* $\partial(g \circ L)(x) = L^*(\partial g(L(x)))$ *for all* $x \in \mathbb{E}_1$*.*

Theorem 19 (Subdifferential of infimal convolution)**.** *Let* $f, g \in \Gamma_0$ *and* $x dom(f\#g)$*. Then*

$$\partial(f\#g)(x) = \partial f(\bar{u}) \cap \partial g(x - \bar{u}), \tag{103}$$

for all $\bar{u} \in \arg\min_u\{f(u) + g(x-u)\}$*.*

Corollary 19 (Derivative of Mereau envelop)**.** *Let* $f \in \Gamma_0$ *and* $\lambda > 0$*. Then* $e_\lambda f$ *is continuously differentiable with* $\frac{1}{\lambda}$ *Lipschitz gradient given by*

$$\nabla e_\lambda f(x) = \frac{1}{\lambda}(I - prox_{\lambda f})(x). \tag{104}$$

Theorem 20 (Moreau decomposition)**.** *Let* $f \in \Gamma_0$ *and* $\lambda > 0$*. Then for all* $x \in \mathbb{E}$*, we have*

$$x = prox_{\lambda f}(x) + \lambda prox_{\frac{1}{\lambda} f^*}\left(\frac{x}{\lambda}\right). \tag{105}$$

Proposition 25. *For any proper, convex, closed function* $f : \mathbb{E} \to \overline{\mathbb{R}}$ *and* $g : \mathbb{E} \to \mathbb{R}$ *is continuously differentiable, then*

$$\partial(f+g)(x) = \partial f(x) + \nabla g(x), \quad \forall x \in \mathbb{E}. \tag{106}$$

Corollary 20 (Primal-dual optimality condition)**.** *Consider the problems*

P. $\min_x h(Lx) + g(x)$

D. $\max_y -g^*(-L^*(y)) - h^*(y)$

where $g : \mathbb{E} \to \overline{\mathbb{R}}$ *and* $h : Y \to \overline{\mathbb{R}}$ *are proper, closed, and convex functions, and* $L \in \mathcal{L}(\mathbb{E}, Y)$*.*

Suppose the optimal value of the primal and dual problem are equal, as implied by the qualification condition. Then, x *is the minimizer of the primal problem and* y *is the maximizer of the dual problem iff*

$$\begin{bmatrix} 0 \\ 0 \end{bmatrix} \in \begin{bmatrix} 0 & L^* \\ -L & 0 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \partial g(x) \times \partial h^*(y). \tag{107}$$

8 Important results on common functions

Here we see how the theories are put together.

8.1 Logorithmic determinant function

Proposition 26 (Logorithmic determinant). *We equip the linear space $\mathbb{S}^n$ of $n \times n$ symmetric matrices with the inner product $\langle X, Y \rangle = tr(XY)$ and the induced norm $\|X\|_F = \sqrt{tr(X^2)}$. Consider the function $f : \mathbb{S}^n \to \overline{\mathbb{R}}$ defined by:*

$$f(X) = \begin{cases} -\log\det(X), & X \in \mathbb{S}^n_{++} \\ +\infty, & \textit{otherwise} \end{cases} \tag{108}$$

We have that

$$\nabla f(X) = -X^{-1}, \quad \forall X \in \mathbb{S}^n_{++}. \tag{109}$$

$$\nabla^2 f(X)[V] = X^{-1}VX^{-1}, \quad \forall X \in \mathbb{S}^n_{++}, V \in \mathbb{S}^n. \tag{110}$$

$$\langle \nabla^2 f(X)[V], V \rangle = tr(X^{-1}VX^{-1}V) \tag{111}$$

$$= \|X^{-\frac{1}{2}}VX^{-\frac{1}{2}}\|_F^2 \geq 0, \quad \forall X \in \mathbb{S}^n_{++}, V \in \mathbb{S}^n. \tag{112}$$

namely, the logarithmic determinant is convex on $\mathbb{S}^n_{++}$.

Determinant of a square matrix is the product of its eigenvalues. For a triangular or diagonal matrix, the determinant is the product of its diagonal entries.

8.2 Indicator, distance, and norm

Distance and indicator

Proposition 27 (Distance function obtained from Moreau envelope). *Let $C \subset \mathbb{E}$ be nonempty, closed, and convex. Then the function $d_C = \delta_C \# \|\cdot\|$ is called the distance function to the set C, and is given by*

$$d_C(x) = \inf_{u \in C} \|x - u\| \tag{113}$$

If C is closed and convex, the projection operator $\text{proj}_C(x) = \arg\min_{u \in C} \|x - u\|$ is nonempty and single-valued, it follows that we have

$$d_C(x) = \|x - \text{proj}_C(x)\|. \tag{114}$$

Moreover, by Proposition 14, d_C is convex as the infimal convolution of two convex functions. It is in fact globally Lipschitz continuous with modulus 1.

Proposition 28 (Subdifferential of indicator function). *Let $C \subset \mathbb{E}$ be nonempty, closed, and convex. Then for all $x \in \mathbb{E}$, we have*

$$\partial\delta_C(x) = N_C(x) = \begin{cases} \{0\}, & x \in int(C) \\ N_C(x), & x \in bd(C) \\ \emptyset, & x \notin C \end{cases} \tag{115}$$

Norm

Proposition 29 (Conjugacy for norm)**.** *Let $\|\cdot\|$ be any norm on $\mathbb{E}$ and $\|\cdot\|_*$ be its dual norm. Then we have $\|\cdot\|^* = \delta_{\|\cdot\|_* \leq 1}$, i.e. the conjugate of a norm is the indicator function of the unit ball of its dual norm.*

Proposition 30. *For $f(x) = \frac{1}{2}\|x\|^2$, we have $f^* = f$.*

References