CPSC 536C: Algorithms for Convex Optimization
University of British Columbia

Fundamentals of Convexity

Akshay Ram
January 12, 2025

These notes are subject to change and may contain errors.

1 Preliminaries

- Notation: $\mathbb{R}$ for reals, $\mathbb{R}_+$ for non-negative reals, and $\mathbb{R}_{++}$ for positive reals. $[n] = \{1, ..., n\}$ for integer intervals. $e_1, ..., e_n \in \mathbb{R}^n$ for the standard basis. $\langle x, y\rangle = \sum_{i=1}^n x_i y_i$ for standard Euclidean inner product, and $\|x\|_2^2 = \langle x, x\rangle = \sum_{i=1}^n x_i^2$ for the Euclidean norm.

- Complexity Theory - $f \leq O(g)$ if there exists $n_0 \in \mathbb{N}, c \in \mathbb{R}_{++}$ such that for all $n \geq n_0$: $f(n) \leq cg(n)$. We will also use $f \lesssim g$ to denote the same thing.

- Topology - open vs closed

  Definition 1. *$S \subseteq \mathbb{R}^n$ is* open *if for every point $x \in S$, there is an $\varepsilon > 0$ such that the open ball $B^\circ(x, \varepsilon) := \{y \mid \|y - x\|_2 < \varepsilon\} \subseteq S$.*

  S is closed if the complement $\mathbb{R}^n/S$ is closed; equivalently, if S contains the limit point of every convergent sequence in S: $\{x_i \in S\} \implies \lim_i x_i \in S$.

  S is compact if it is closed and bounded.

- Differentiability and Taylor approximation

  Definition 2. *f is differentiable at x in the interior of the domain dom$(f)^\circ$ if all partial derivatives exist:*

  $$D_v f(x) := \lim_{t\to 0} \frac{f(x + tv) - f(x)}{t},$$

  and further the above limit is a linear function of v.

  It is k-times differentiable if all k-th order partial derivates exist, and k-times continuously *differentiable if furthermore the k-th derivative is continuous in a neighborhood of x.*

  Definition 3. *If f is differentiable at x, and an inner product $\langle \cdot, \cdot\rangle$ is given, then the* gradient *$\nabla f(x)$ is uniquely defined by*

  $$\forall v \in \mathbb{R}^n : \quad \langle \nabla f(x), v\rangle = \lim_{t\to 0} \frac{f(x + tv) - f(x)}{t} =: D_v f(x).$$

  For the standard inner product, this induces the more familiar definition

  $$(\nabla f(x))_i = \partial_{x_i} f(x).$$

Definition 4. *Similary, if f is twice-differentiable at x, the* Hessian $\nabla^2 f(x)$ *is uniquely defined by*

$$\forall u, v \in \mathbb{R}^n : \quad \langle u, \nabla^2 f(x) v \rangle = D_u D_v f(x).$$

For the standard inner product, this induces the more familiar definition

$$(\nabla^2 f(x))_{ij} = \partial_{x_i} \partial_{x_j} f(x).$$

- Linear and quadratic functions

  Definition 5. *An affine function $\ell : \mathbb{R}^n \to \mathbb{R}$ is of the form*

  $$\ell(x) = \langle a, x \rangle + b$$

  for $a \in \mathbb{R}^n, b \in \mathbb{R}$.

  A quadratic function $q : \mathbb{R}^n \to \mathbb{R}$ is of the form

  $$q(x) := \langle x, Ax \rangle + \langle b, x \rangle + c$$

  where $A \in \mathbb{R}^{n \times n}$ (symmetric matrix without loss of generality), $b \in \mathbb{R}^n, c \in \mathbb{R}$.

  Definition 6. *For once- and twice-differentiable functions $f : \mathbb{R}^n \to \mathbb{R}$, the linear and quadratic approximation at x are*

  $$\ell_x(y) := f(x) + \langle \nabla f(x), y - x \rangle;$$

  $$q_x(y) := f(x) + \langle \nabla f(x), y - x \rangle + \frac{1}{2} \langle (y - x), \nabla^2 f(x)(y - x) \rangle.$$

Remark 7. *By e.g. intermediate value theorem, the* remainder $f - \ell_x, f - q_x$ *are small in the neighborhood of x if f is appropriately differentiable at x.*

2 Introduction

2.1 Convex Sets

Definition 8. *A set $C \subseteq \mathbb{R}^d$ is* convex *if for all $x, y \in C$ and $\lambda \in [0, 1]$,*

$$\lambda x + (1 - \lambda) y \in C.$$

Definition 9. *For subset $S \subseteq \mathbb{R}^n$ we can define the* span, affine, *and* convex hull *in terms of linear combinations as*

$$\mathit{span}(S) := \{\sum_i^N a_i x_i \mid x_i \in S\};$$

$$\mathit{aff}(S) := \{\sum_i^N a_i x_i \mid x_i \in S, \sum_i^N a_i = 1\};$$

$$\mathit{conv}(S) := \{\sum_i^N a_i x_i \mid x_i \in S, a \geq 0, \sum_i^N a_i = 1\}.$$

Note $\mathrm{conv}(S) \subseteq \mathrm{aff}(S) \subseteq \mathrm{span}(S)$. Try to visualize these sets for small examples.

Fact 10. *The following operations preserve convexity of sets*

- *Scalar multiplication:* $K \to cK$*;*
- *Addition:* $K_1 + K_2$
- *Intersection:* $\cap_i K_i$
- *Affine transform:* $K \to AK + b$

2.2 Convex Functions

Definition 11. *A function* $f : \mathbb{R}^d \to \mathbb{R}$ *is* convex *if*

$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y)$$

for all $x, y \in \mathit{dom}(f)$ *and* $\lambda \in [0,1]$.

Here $\mathrm{dom}(f)$ is the domain of f, where it is well-defined. E.g. $\mathbb{R}_+$ for $\sqrt{x}$ or $\mathbb{R}_{++}$ for $\log(x)$. Function convexity depends on the domain, e.g. $\sqrt{|x|}$ is convex for dom $= \mathbb{R}_+$ but not for dom $= \mathbb{R}$. We can also use the convention that $f(x \notin \mathrm{dom}(f)) := +\infty$, which then implies $f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y)$ for all x, y, with the following arithmetic rules for infinity: $0 \cdot \infty = 0, \forall s \in \mathbb{R}_{++} : s \cdot \infty = \infty, \forall s \in \mathbb{R} : s + \infty = \infty$.

Exercise 1. *Verify that* $f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y)$ *for all* x, y *with the extended arithmetic rules given above iff* $\mathit{dom}(f)$ *is a convex set.*

For convex $K \subseteq \mathbb{R}^n$ *we define the* indicator function δ_K *and the* support function h_K *as*

$$\delta_K(x) := \begin{cases} 0 & x \in K \\ +\infty & \textit{otherwise} \end{cases}; \qquad h_K(y) := \sup_{x \in K} \langle y, x \rangle.$$

Verify that both of these are convex functions.

Fact 12. *The following operations preserve convexity of functions*

- *Non-negative scalar multiplication:* $f \to cf$*;*
- *Addition:* $f_1 + f_2$
- *Point-wise supremum:* $\sup_i f_i$
- *Restriction:* $t \to f((1-t)x + ty)$ *(or more generally a line or subspace).*
- *Affine transform:* $x \to f(Ax + b)$
- *Perspective:* $(x, t) \to tf(x/t)$ *for* $t > 0$.

Theorem 13. *If* f *is differentiable, then* f *is convex iff*

$$f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle \quad \textit{for all } x, y.$$

Proposition 14. *For convex* f*,* x *is a local minimum iff it is a global minimum.*

Proof: Clearly a global minimum is also a local minimum. For the converse, let $\varepsilon > 0$ such that $f(y) \geq f(x)$ for all $\|y - x\|_2 \leq \varepsilon$. Assume for contradiction x is not the global minimum, so $f(x_*) < f(x)$. Then letting $x_\delta := (1-\delta)x + \delta x_*$, by convexity we have

$$f(x_\delta) \leq (1-\delta)f(x) + \delta f(x_*) < f(x).$$

But for small enough δ this contradicts local minimality of x.

Definition 15. *For $f : \mathbb{R}^n \to \mathbb{R}$, let the* epigraph *be*

$$epi(f) := \{(x,t) \mid f(x) \leq t\}$$

i.e. the 'upwards closure' of the graph of f in $\mathbb{R}^{n+1}$.

Theorem 16. *$f : \mathbb{R}^n \to \mathbb{R}$ is a (closed) convex function iff the epigraph $epi(f) \subseteq \mathbb{R}^{n+1}$ is a (closed) convex set.*

Proof: First consider f convex, and we want to show $\text{epi}(f)$ is a convex set. Let $(x,t),(y,s) \in \text{epi}(f)$, which by definition means

$$f(x) \leq t, \qquad f(y) \leq s.$$

Now consider convex combination $z := (1-\lambda)x + \lambda y$ for $\lambda \in [0,1]$. Then by convexity

$$f(z) \leq (1-\lambda)f(x) + \lambda f(y) \leq (1-\lambda)t + \lambda s,$$

where the first step was by convexity of f, and in the second step we used that $(x,t),(y,s) \in \text{epi}(f)$. Rewriting this, we have shown

$$(z, (1-\lambda)t + \lambda s) = (1-\lambda)(x,t) + \lambda(y,s) \in \text{epi}(f),$$

which shows $\text{epi}(f)$ is a convex set.

For the converse, assume $\text{epi}(f)$ is closed and convex. Then for $(x, f(x)), (y, f(y)) \in \text{epi}(f)$ we have $(z,t) \in \text{epi}(f)$ for $z := (1-\lambda)x + \lambda y, t := (1-\lambda)f(x) + \lambda f(y)$, which implies

$$f((1-\lambda)x + \lambda y) = f(z) \leq t = (1-\lambda)f(x) + \lambda f(y),$$

verifying convexity of f.

Lemma 17. *For convex $f : \mathbb{R}^n \to \mathbb{R}$ and any $t \in \mathbb{R}$, the sub-level set*

$$L_t(f) := \{x \in \mathbb{R}^n \mid f(x) \leq t\}$$

is a convex set. Further, if $f(x) = t$ and f is differentiable at x, then the gradient $\nabla f(x)$ gives a supporting hyperplane *for the sub-level set L_t at x.*

Proof: The sub-level set is the projection of the intersection of the epigraph and a halfspace

$$\text{epi}(f) \cap \{(z,s) \mid s \leq t\},$$

and intersection and projection both preserve convexity, so L_t is convex.

3 Separation Theorems

3.1 Separating Hyperplane Theorem

Theorem 18. *For closed convex K and $p \in \mathbb{R}^n$ with $p \notin C$, there is a separating hyperplane $w \in \mathbb{R}^n$ satisfying*

$$\max_{x \in C} \langle w, x \rangle < b < \langle w, p \rangle.$$

Proof: The proof plan is simple: let $x_* := \arg\inf_{x \in K} \|x - p\|_2^2$, then we claim the direction $w := x_* - p$ gives a separating hyperplane.

To show this is well-defined, we reduce to the case of compact K. Take any $x_0 \in K$ and consider

$$L := K \cap \{y \in K \mid \|y - p\|_2^2 \leq \|x_0 - p\|_2^2\},$$

which is convex since K is convex and the second set is the sub-level set of convex function $y \to \|y - p\|_2^2$. Therefore we can assume K is compact and $x_* = \arg\inf_{x \in K} \|x - p\|_2^2$ is attained by the extreme value theorem. We can in fact show that the optimizer x_* is unique, as $\| \cdot \|_2^2$ is strictly convex, but we will not need this for the proof.

Now we claim that $x_* = \arg\min_{x \in K} \|x - p\|_2^2$ iff $\forall x \in K : \langle x - x_*, x_* - p \rangle \geq 0$. Indeed, we can rewrite the function difference

$$\|x - p\|_2^2 - \|x_* - p\|_2^2 = 2\langle x - x_*, x_* - p \rangle + \|x - x_*\|_2^2,$$

as can be verified directly. If the inner product term is ≥ 0 for all x, then since the norm term is always ≥ 0, we get $\|x - p\|_2^2 \geq \|x_* - p\|_2^2 \forall x \in K$. Conversely, if there exists $x \in K : \langle x - x_*, x_* - p \rangle < 0$, then consider $x_\varepsilon := (1 - \varepsilon)x_* + \varepsilon x$, which is in K by convexity, and note that

$$\|x_\varepsilon - p\|_2^2 - \|x_* - p\|_2^2 = 2\langle x_\varepsilon - x_*, x_* - p \rangle + \|x_\varepsilon - x_*\|_2^2 = \varepsilon \langle x - x_*, x_* - p \rangle + \varepsilon^2 \|x - x_*\|_2^2 < 0,$$

for small enough $\varepsilon > 0$, where we used $x_\varepsilon - x_* = \varepsilon(x - x_*)$, and the last step follows as the inner product term is negative and therefore dominates the quadratic term for small enough $\varepsilon > 0$.

Finally, we verify that $w := x_* - p$ gives a separating hyperplane:

$$\langle p - x_*, w \rangle = -\|p - x_*\|_2^2 < 0, \qquad \forall x \in K : \langle x - x_*, w \rangle = \langle x - x_*, x_* - p \rangle \geq 0$$

which by rearranging gives

$$\langle p, w \rangle < \langle x_*, w \rangle \leq \inf_{x \in K} \langle x, w \rangle.$$

We also get a version for two sets.

Theorem 19. *For closed convex $C, D \subseteq \mathbb{R}^n$, if they are disjoint $C \cap D = \emptyset$ and one of C or D is bounded, then there is a separating hyperplane $(w, b) \in \mathbb{R}^{n+1}$ satisfying*

$$\sup_{x \in C} \langle w, x \rangle < b < \inf_{x \in D} \langle w, x \rangle.$$

Proof: Consider $K = C - D$ and $p = 0$: there is a hyperplane separating C, D iff there is a hyperplane separating 0 from $C - D$. We need the condition that one of C or D is bounded in order for K to be closed.

Remark 20. *Both the assumptions (closed and bounded) are necessary for strict separation: open convex intervals* $(0,1),(1,2) \subseteq \mathbb{R}$ *are disjoint but cannot be strictly separated; also* $\{(x,y) \mid y \leq 0\}$ *and the epigraph of the function* $1/x$*, both in* $\mathbb{R}^2$ *are disjoint closed convex sets that 'meet at* ∞*' and so cannot be strictly separated.*

If we only require separation $\sup_{x\in C}\langle w,x\rangle \leq \inf_{x\in D}\langle w,x\rangle$*, then some of these technical conditions can be dropped, see [BV Section 2.5] for further discussion.*

Corollary 21. *For closed convex* $K \subseteq \mathbb{R}^n$*, recall the support function* $h_K(w) := \sup_{y\in K}\langle w,y\rangle$ *as defined in exercise 1. Show*

$$K = \{x \in \mathbb{R}^n \mid \forall w \in \mathbb{R}^n : \langle w,x\rangle \leq \sup_{y\in K}\langle w,y\rangle = h_K(w)\}.$$

The proof is left as an exercise, but this result is equivalent to the strong duality result proved in theorem 24.

3.2 Supporting Hyperplanes

Definition 22. *A supporting halfspace* H *of* $K \subseteq \mathbb{R}^n$ *at* $x \in \partial K$ *satisfies: (1)* $K \subseteq H$*; (2)* $x \in \partial H$*.*

Corollary 23. *For every closed convex* K *and* $x \in \partial K$*, there is a* supporting hyperplane $H \supseteq K$ *such that* $x \in \partial H$*.*

Proof: See Theorem 3.1.12 in [Nesterov] and page 51 of [Boyd, Vanderberghe].

Exercise 2 (Exercise 2.27 in BV)**.** $K \subseteq \mathbb{R}^n$ *closed with non-empty interior, then* K *is convex iff it has a* supporting *hyperplane at every point of its boundary.*

4 Strong Duality

Theorem 24. *Closed convex* $K \subseteq \mathbb{R}^n$ is *the intersection of all containing halfspaces*

$$K \equiv \cap_{\text{halfspace } H\supseteq K} H.$$

Proof: One direction is clear: for $x \in K, H \supseteq K \implies x \in H$, so $K \subseteq \cap_{H\supseteq K} H$. For the converse, consider $z \notin K$, then the separating hyperplane theorem 18 gives $w \in \mathbb{R}^n$ such that

$$\langle w,z\rangle > b > \sup_{x\in K}\langle w,x\rangle.$$

So we can consider halfspace $H := \{x \in \mathbb{R}^n \mid \langle w,x\rangle \leq b\}$, which contains K and does not contain z. Therefore $z \notin \cap_{H\supseteq K} H$.

We can also lift this to functions using the epigraph.

Theorem 25. *Closed convex* $f : \mathbb{R}^n \to \mathbb{R}$ *and proper* f *is the supremum of all affine minorants*

$$f(x) = \sup_{f\geq h} h(x).$$

Here proper means f *is never* $-\infty$ *and is not always* $+\infty$*, i.e.* $\inf f \notin \{\pm\infty\}$*.*

Proof: Recall $\text{epi}(f) \subseteq \mathbb{R}^n$ is a closed convex set by theorem 16. Therefore we can apply the previous duality theorem, so

$$\text{epi}(f) = \cap_{H \supseteq \text{epi}(f)} H,$$

where the intersection is over all halspaces containing the epigraph. Such a halfspace is defined in terms of normal and intercept

$$H = \{(y, s) \mid \langle (w, -b), (y, s) \leq c\}.$$

Since the epigraph is upwards closed, if H is a containing halfspace, then it must also be upwards closed, i.e. we claim that $b \geq 0$: indeed $(y, s) \in \text{epi}(f) \implies (y, s + t) \in \text{epi}(f) \forall t \geq 0$ so if $b < 0$ then H cannot contain the epigraph as

$$\langle (w, -b), (y, s + t) \rangle = \langle (w, -b), (y, s) \rangle - bt > c,$$

since $-tb$ can be made arbitrarily large.

If $b > 0$ then we claim this corresponds to an affine function. First note, if $b > 0$ then we can equivalently write

$$H = \{(y, s) \mid \langle (w, -b), (y, s) \leq c\} = \{(y, s) \mid \langle (w/b, -1), (y, s) \leq c/b\}.$$

Now we claim that this halfspace is exactly the epigraph of the affine function $h(y) := \langle \tilde{w}, y \rangle - \tilde{c}$ for $\tilde{w} := w/b, \tilde{c} := c/b$:

$$(y, s) \in H \iff \langle \tilde{w}, y \rangle - \tilde{c} \leq s \iff h(y) \leq s \iff (y, s) \in \text{epi}(h).$$

Further, $f \geq h$ iff $\text{epi}(f) \subseteq \text{epi}(h) = H$. So these are exactly the affine functions we want to consider in our sup and we call the H corresponding to $b > 0$ as affine halfspaces. Further recall that $\text{epi}(\sup_i h_i) \cap_i \text{epi}(h_i)$.

We want to show that $f = \sup_{f \geq h} h$, or equivalently $\text{epi}(f)$ is defined as the intersection of just the halfspaces with $b > 0$. So consider $(y, s) \notin \text{epi}(f)$ and let H be the separating halfspace corresponding to

$$\langle (w, -b), (y, s) \rangle > \sup_{(x,t) \in \text{epi}(f)} \langle (w, -b), (x, t) \rangle.$$

If $b > 0$ then we are done, so we must have $b = 0$. There must exist some containing affine halfspace H' defined by

$$\langle (w', -1), (y, s) \rangle > \sup_{(x,t) \in \text{epi}(f)} \langle (w', -1), (x, t) \rangle,$$

(this is the place we use that f is proper). Then let $w_\delta := (1 - \delta) w + \delta w', b' := (1 - \delta)$ and note that

$$\langle (w_\delta, -b_\delta), (y, s) \rangle > \sup_{(x,t) \in \text{epi}(f)} \langle (w_\delta, -b_\delta), (x, t) \rangle.$$

for small enough $\delta > 0$. But then this corresponds to an affine halfspace separating (y, s) and therefore corresponds to an affine function $f \geq h$.

The proof above reduces f to its epigraph and then applies strong duality for convex sets. It is valuable to consider what exactly the supporting hyperplanes of $\text{epi}(f)$ correspond to.

5 First Order Definition of Convexity

Recall that $f : \mathbb{R}^n \to \mathbb{R}$ is differentiable at $x \in \text{dom}(f)^\circ$ if the limit

$$\lim_{t\to 0} \frac{f(x+tv) - f(x)}{t}$$

exists, and further is a linear function of v. Given an inner product, the gradient is defined as

$$\forall v : \quad \langle \nabla f(x), v\rangle = \lim_{t\to 0} \frac{f(x+tv) - f(x)}{t}.$$

Proposition 26. *Let f be differentiable. Then f is convex iff*

$$\forall x, y : \quad f(y) \geq f(x) + \langle \nabla f(x), y - x\rangle.$$

We can apply this theory even in the non-differentiable setting.

Definition 27. *For $f : \mathbb{R}^n \to \mathbb{R}$ and $x \in dom(f)$, the set of* subgradients *$\partial f(x)$ are the elements $g \in \mathbb{R}^n$ satisfying*

$$\forall y \in \mathbb{R}^n : \quad f(y) \geq f(x) + \langle g, y - x\rangle.$$

Note that the set of subgradients is convex: since it is defined by the above inequalities, $g, h \in \partial f(x) \implies (1-\lambda)g + \lambda h \in \partial f(x)$.

The most important property of convexity, that local minima are global minima, can also be phrased in terms of first-order information:

Theorem 28. *For convex $f : \mathbb{R}^n \to \mathbb{R}$, $x = \arg\min_{y\in\mathbb{R}^n} f(y)$ is the global minimum iff $0 \in \partial f(x)$. For differentiable functions, this is equivalent to $\nabla f(x) = 0$.*

Proof: If x is the global minimum, then

$$\forall y \in \mathbb{R}^n : \quad f(y) \geq f(x) = f(x) + \langle 0, y - x\rangle,$$

i.e. $0 \in \partial f(x)$ verifies the definition of subgradient.

Conversely, if $0 \in \partial f(x)$ then we have

$$\forall y \in \mathbb{R}^n : \quad f(y) \geq f(x) + \langle 0, y - x\rangle = f(x),$$

i.e. x is the global minimum.

Proposition 29. *If f is convex, then for every $x \in dom(f)^\circ$, there is a subgradient $\partial f(x) \neq \emptyset$.*

Conversely, if $\partial f(x) \neq \emptyset$ for every $x \in dom(f)$, then f is convex.

Proof: Follows from exercise 2 applied to the epigraph.

6 GLS Oracle Model

Definition 30 (Algorithmic Oracle Model for Convex Optimization)**.** *MEM Input: $x \in \mathbb{R}^n$;*

Output: YES if $x \in K$; NO otherwise.

SEP Input: $x \in \mathbb{R}^n$;

Output: YES if $x \in K$; otherwise separating hyperplane $\langle h, \cdot\rangle$ such that $\sup_{z\in K}\langle h, z\rangle < \langle h, x\rangle$.

OPT *Input:* $c \in \mathbb{R}^n$;
Output: $\max_{x \in K} \langle c, x \rangle$

We can also consider *approximate* versions of all these oracles. For the formal definitions, see [GLS 2.1]. Many of the algorithms we study in this class can be thought of as reductions between

Think about what these are in the polar/dual context discussed below. Also what is the relation to function oracles?

7 Polar Sets and Cones

Definition 31. *For convex $K \subseteq \mathbb{R}^n$, the* polar set *is*

$$K^\circ := \{y \in \mathbb{R}^n \mid \forall x \in K : \langle y, x \rangle \leq 1\}.$$

Remark 32. *Relate this to the* support function *as described in exercise 1.*

Note that $0 \in K^\circ$ always; and note that K° is a convex set, even if K is not convex.

Theorem 33. *If $0 \in K$ then $K^{\circ\circ} = K$. More generally,*

$$K^{\circ\circ} = \mathit{conv}(K \cup \{0\}).$$