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 - big-O and poly time
- 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:*
  $$D_v f(x) := \lim_{t\to 0} \frac{f(x+tv) - f(x)}{t}.$$
  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, the* gradient *$\nabla f(x)$ is uniquely defined by*
  $$\forall v \in \mathbb{R}^n : \quad \langle \nabla f(x), v\rangle = D_v f(x).$$
  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 function $f : \mathbb{R}^n \to \mathbb{R}$, the linear and quadratic approximation at x, for once- and twice-differentiable functions, respectively, 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*

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

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

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

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: $\cap_i K_i$*

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 $x, y \in \mathbb{R}^d$ and $\lambda \in [0,1]$.

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).*

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 iff it is a stationary point.*

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

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

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

Theorem 16. *f is (closed) convex iff the epigraph epi(f) is a (closed) convex set.*

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.*

3 Strong Duality

3.1 Separating Hyperplane Theorem

Theorem 18. *For closed convex $C, D \subseteq \mathbb{R}^n$, if they are disjoint $C \cap D = \emptyset$ then there is a separating hyperplane $(w, b) \in \mathbb{R}^{n+1}$ satisfying*

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

Corollary 19. *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.

Theorem 20. *Closed convex $K \subseteq \mathbb{R}^n$* is *the intersection of all containing halfspaces*

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

Corollary 21. *Let $K \subseteq \mathbb{R}^n$ be a closed convex set. Then*

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