Convex Optimization: Reading Notes 2

GKxx

May 6, 2022

1 Convexity-preserving functions

Definition 1.1 (Affine function). *An* affine function $f : \mathbb{R}^n \to \mathbb{R}^m$ *is of the form*

$$f(x) = Ax + b,$$

with $A \in \mathbb{R}^{m \times n}$ *and* $b \in \mathbb{R}^n$.

Proposition 1.2. *The image of a convex set under an affine function is convex.*

Proof. Suppose $S \subseteq \mathbb{R}^n$ is convex and $f(x) = Ax + b$ is an affine function $f : \mathbb{R}^n \to \mathbb{R}^m$. For any $f(x), f(y) \in f(C)$ and any $\theta \in [0, 1]$, the convex combination is $\theta f(x) + (1 - \theta) f(y) = A(\theta x + (1 - \theta) y) + b = f(\theta x + (1 - \theta) y)$, while $\theta x + (1 - \theta) y \in C$ due to the convexity of C. Therefore, $\theta f(x) + (1 - \theta) f(y) \in f(C)$. □

Proposition 1.3. *The inverse image of a convex set under an affine function is convex.*

Proof. Suppose C is a convex set and $f(x) = Ax + b$ is an affine function. We will show that

$$f^{-1}(C) = \{x \in \mathbb{R}^n \mid f(x) \in C\}$$

is convex. For any $x, y \in \mathbb{R}^n$ with $f(x), f(y) \in C$ and any $\theta \in [0, 1]$, we have that

$$f(\theta x + (1 - \theta) y) = A(\theta x + (1 - \theta y) + b = \theta(Ax + b) + (1 - \theta)(Ay + b)) = \theta f(x) + (1 - \theta) f(y).$$

This is in the set C because C is convex and both $f(x)$ and $f(y)$ are in C. □

Example 1.4. *The hyperbolic cone*

$$H = \left\{ x \in \mathbb{R}^n \mid x^T P x \leqslant \left(c^T x\right)^2, c^T x \geqslant 0 \right\},$$

where $P \in \mathbb{S}_+^n$ *and* $c \in \mathbb{R}^n$, *is convex.*

Proof. H is the inverse image of the second-order cone

$$K = \left\{ (z, t) \mid z^T z \leqslant t^2, t \geqslant 0 \right\}$$

under the affine function $f : \mathbb{R}^n \to \mathbb{R}^{n+1}$ given by

$$f(x) = \begin{bmatrix} P^{1/2} \\ c^T \end{bmatrix} x.$$

□

Definition 1.5 (Perspective function)**.** *A* perspective function $P : \mathbb{R}^{n+1} \to \mathbb{R}^n$ *is of the form*

$$P(x, t) = x/t,$$

where $x \in \mathbb{R}^n$ *and* $t > 0$*. It can also be written as*

$$P(x) = \frac{1}{x_{n+1}} \begin{bmatrix} x_1 & \cdots & x_n \end{bmatrix}^T.$$

Proposition 1.6. *The image of a convex set under a perspective function is convex.*

Proof. Suppose $C \subseteq \mathbb{R}^n \times \mathbb{R}_{++}$ is a convex set and P is a perspective function. For every $x = (\tilde{x_n}, x_{n+1}), y = (\tilde{y_n}, y_{n+1}) \in C$ and every $\theta \in [0, 1]$, we have that

$$\begin{aligned} P(\theta x + (1-\theta)y) &= \frac{\theta \tilde{x_n} + (1-\theta)\tilde{y_n}}{\theta x_{n+1} + (1-\theta)y_{n+1}} \\ &= \frac{\tilde{x_n}}{x_{n+1}} \cdot \frac{\theta x_{n+1}}{\theta x_{n+1} + (1-\theta)y_{n+1}} + \frac{\tilde{y_n}}{y_{n+1}} \cdot \frac{(1-\theta)y_{n+1}}{\theta x_{n+1} + (1-\theta)y_{n+1}} \\ &= \mu P(x) + (1-\mu)P(y), \end{aligned}$$

where

$$\mu = \frac{\theta x_{n+1}}{\theta x_{n+1} + (1-\theta)y_{n+1}} \in [0, 1].$$

It is obvious that μ is monotonic with respect to θ, so the image of the line segment $[x, y]$ under P is $[P(x), P(yy)]$. Due to the convexity of C, we have $[x, y] \subseteq C$ and therefore $[P(x), P(y)] \subseteq P(C)$. Then for every $\theta \in [0, 1]$ we have $\theta P(x) + (1-\theta)P(y) \in [P(x), P(y)] \subseteq P(C)$, so $P(C)$ is convex. □

Proposition 1.7. *The inverse image of a convex set under a perspective function is convex.*

Proof. Suppose $C \subseteq \mathbb{R}^n$ is a convex set and P is a perspective function. Take any $x, y \in P^{-1}(C)$ so that $P(x), P(y) \in C$. For any $\theta \in [0, 1]$, let $x_0 = \theta x + (1-\theta)y \in [x, y]$ be the convex combination induced by θ, x, y. We have shown that $P([x, y]) = [P(x), P(y)]$, so

$$P(x_0) \in P([x, y]) = [P(x), P(y)].$$

Since C is convex and $P(x), P(y) \in C$, the line segment $[P(x), P(y)]$ is a subset of C. Hence $P(x_0) \in C \Rightarrow x_0 \in P^{-1}(C)$. □

Definition 1.8 (Linear-fractional function)**.** *A* linear-fractional function $f : \mathbb{R}^n \to \mathbb{R}^m$ *is of the form*

$$f(x) = \frac{Ax + b}{c^T x + d}, \quad \textbf{dom}\, f = \left\{ x \mid c^T x + d > 0 \right\}.$$

It is the composition of an affine function g *and the perspective function* P*, where*

$$g(x) = \begin{bmatrix} A \\ c^T \end{bmatrix} x + \begin{bmatrix} b \\ d \end{bmatrix}.$$

The following two propositions are easy to see by viewing a linear-fractional function as the composition of an affine function and a preserving function.

Proposition 1.9. *The image of a convex set under a linear-fractional function is convex.*

Proposition 1.10. *The inverse image of a convex set a linear-fractional function is convex.*

2 Generalized inequalities

Definition 2.1 (Proper cone)**.** *A cone* $K \subseteq \mathbb{R}^n$ *is said to be a* proper cone *if*

- K *is convex.*
- K *is closed (contains its boundary).*
- K *is solid (has nonempty interior).*
- K *is pointed (contains no line, or equivalently,* $\pm x \in K \Rightarrow x = 0$*).*

Definition 2.2 (Generalized inequalities)**.** *The generalized inequalities on* $\mathbb{R}^n$ *can be defined by a proper cone* $K \subseteq \mathbb{R}^n$ *as:*

$$x \preceq_K y \iff y - x \in K.$$

$$x \prec_K y \iff y - x \in \textbf{int}\, K.$$

Example 2.3. *The componentwise inequality is the generalized inequality defined by* $K = \mathbb{R}^n_+$*, which is the nonnegative orthant.*

Example 2.4. *The generalized inequality defined by the positive semi-definite cone* $K = \mathbb{S}^n_+$ *is*

$$X \preceq_{\mathbb{S}^n_+} Y \iff Y - X \in \mathbb{S}^n_+.$$

Proposition 2.5. *The generalized inequality* $\preceq_K$ *defined by a proper cone* K *is an ordering relation.*

Proof.

1. Reflectivity: For any $x \in \mathbb{R}^n$, $x \preceq_K x$ because $x - x = 0 \in K$.
2. Transitivity: If $x \preceq_K y$ and $y \preceq_K z$, then $z - x = (z - y) + (y - x) \in K$ because K is a convex cone.
3. Antisymmetry: If $x \preceq_K y$ and $y \preceq_K x$, then $\pm(x - y) \in K \Rightarrow x - y = 0 \Rightarrow x = y$.

□

Proposition 2.6. *The generalized inequality* $\preceq_K$ *defined by a proper cone* K *is preserved under addition, nonnegative scaling and limits.*

Proof.

- If $x \preceq_K y$ and $u \preceq_K v$, then $(y + v) - (x + u) = (y - x) + (v - u) \in K$ because K is a convex cone.
- If $x \preceq_K y$, then $\alpha x \preceq_K \alpha y$ for any scalar $\alpha \geqslant 0$ because $\alpha y - \alpha x = \alpha(y - x) \in K$.
- We first show that if $x_n \in K$ for every $n \in \mathbb{N}$ and $\lim_{n\to\infty} x_n$ exists, then $x = \lim_{n\to\infty} x_n \in K$. Assume that $x \notin K$. Since K is closed, there exists $u \in K$ such that $\|x - u\|_2 = \min_{v\in K} \|x - v\|_2$. Take $\varepsilon = \frac{1}{2}\|x - u\|_2$, then there exists $N \in \mathbb{N}$ such that for every $n > N$ we have $\|x_n - x\|_2 < \frac{1}{2}\|x - u\|_2$. This shows that $\|x - x_n\|_2 < \|x - v\|_2$ for every $v \in K$, so $x_n \notin K$ for $n > N$, a contradiction.

  Then we are done by definition of generalized inequality and noting that $\lim_{n\to\infty} y_n - \lim_{n\to\infty} x_n = \lim_{n\to\infty}(y_n - x_n)$.

□

Definition 2.7 (Minimum element). $x \in S$ *is the* minimum element *of* S *with respect to* $\preceq_K$ *if* $x \preceq_K y$ *holds for every* $y \in S$.

Definition 2.8 (Minimal element). $x \in S$ *is the* minimal element *of* S *with respect to* $\preceq_K$ *if for every* $y \in S$ *with* $y \preceq_K x$*, we have* $y = x$.

Remark 2.9. *The minimum element of a set with respect to a generalized inequality is unique, and it is also a minimal element.*

3 Separating and supporting hyperplanes

Theorem 3.1 (Separating hyperplane). *Suppose* C *and* D *are nonempty disjoint convex sets. Then there exist* $a \neq 0$ *and* b *such that* $a^T x \geqslant b$ *for all* $x \in C$ *and* $a^T x \leqslant b$ *for all* $x \in D$*. The hyperplane* $\{x \mid a^T x = b\}$ *is called the* separating hyperplane *for the sets* C *and* D.

Proposition 3.2. *For nonempty disjoint convex sets* C *and* D*, the set* $F = \{x - y \mid x \in C, y \in D\}$ *is a convex set that does not contain* 0*. There exist* $a \neq 0$ *such that* $a^T x \geqslant 0$ *holds for every* $x \in F$*. This is equivalent to the separating hyperplane theorem.*

Theorem 3.3 (Supporting hyperplane). *For any nonempty convex set* C *and any* $x_0 \in \mathbf{bd}\, C$*, there exists a supporting hyperplane* $\{x \mid a^T x = a^T x_0\}$*, for some* $a \neq 0$*, to* C *at the point* x_0.

Proof. If the interior of C is nonempty, then the result follows immediately from applying the separating hyperplane theorem to $\{x_0\}$ and $\mathbf{int}\, C$. If the interior of C is empty, then C must be contained in an affine set with dimension strictly less than n. Then any hyperplane containing this affine set contains C, which is a trivial supporting hyperplane. □

Now we prove the separating hyperplane theorem expressed in 3.2.

Lemma 3.4. *The closure of a convex set is convex.*

Proof. Suppose C is a convex set. We will show that for any $x, y \in \mathbf{cl}\, C$ and $\theta \in [0, 1]$, $x_0 = \theta x + (1 - \theta) y \in \mathbf{cl}\, C$, or equivalently any neighborhood of x_0 intersects C. Let O be an open neighborhood of x_0. Consider the function $g(u, v) = \theta u + (1 - \theta) v$. Since $g(x, y) = x_0 \in O$ and that $g(\cdot)$ is continuous, there exist open sets U and V such that $g(U, V) \subseteq O$ and that $x \in U, y \in V$. Since $x, y \in \mathbf{cl}\, C$, we can take $x_0 \in U \cap C$ and $y_0 \in V \cap C$ so that $g(x_0, y_0) \in C$. Since $g(x_0, y_0) \in O$, we can see that $O \cap C \neq \varnothing$, so $\mathbf{cl}\, C$ is convex. □

Proof of Proposition 3.2. First, we prove the separating hyperplane theorem for the case where the closure of the convex set F does not contain 0. We have shown in Lemma 3.4 that the closure $\mathbf{cl}\, F$ is convex. Now take $a = \arg\min_{u \in \mathbf{cl}\, F} \|u\|_2 \neq 0$. Assume that there exists $x \in \mathbf{cl}\, F$ such that $x^T a \leqslant 0$. Consider $f(\theta) = \|\theta a + (1 - \theta) x\|_2^2$. We have that

$$\begin{aligned} f'(\theta) &= \frac{d}{d\theta} (\theta a + (1 - \theta) x)^T (\theta a + (1 - \theta) x) \\ &= 2\theta \|a\|_2^2 + 2(1 - \theta) \|x\|_2^2 + 2 (1 - 2\theta)\, a^T x. \end{aligned}$$

When $\theta = 1$ we have $f'(1) = 2a^T a - 2a^T x > 0$. Therefore, there exists $\xi \in [0,1]$ such that $\|\xi a + (1-\xi)x\|_2 < \|a\|_2$, while $\xi a + (1-\xi)x \in \mathbf{cl}\, F$ due to the convexity of $\mathbf{cl}\, F$. This contradicts the fact that $a = \arg\min_{u \in \mathbf{cl}\, F} \|u\|_2$. From this we have in fact proved the strict separating hyperplane theorem for this case.

Now we consider the case where $\mathbf{cl}\, F$ contains 0. Suppose the affine dimension of F is m. Take the maximum set of linearly independent vectors of F, which is $\{v_1, \cdots, v_m\}$. Let $w = -v_1 - \cdots - v_m$. We claim that $\forall \alpha > 0$ the point αw is not in $\mathbf{cl}\, F$. Assume that there exists $\alpha > 0$ such that $\alpha w \in \mathbf{cl}\, F$. Take a sequence $\left\{w^{(n)}\right\}$ in C which converges to αw. Let $w^{(n)} = \lambda_1^{(n)} v_1 + \cdots + \lambda_m^{(n) v_m}$ for some coefficients $\lambda_1^{(n)}, \cdots, \lambda_m^{(n)}$. Since $\alpha > 0$ and that $\left\{w^{(n)}\right\}$ converges to αw, there exists $n_0 \in \mathbb{N}$ such that $\lambda_1^{(n_0)}, \cdots, \lambda_m^{(n_0)} < 0$. Then $w^{(n_0)} = \lambda_1^{(n_0)} v_1 + \cdots + \lambda_m^{(n_0)} v_m$, which implies that

$$0 = \frac{1}{1 - \sum_{i=1}^{m} \lambda_i^{(n_0)}} \left(w^{(n_0)} - \lambda_1^{(n_0)} - \cdots - \lambda_m^{(n_0)} \right).$$

The right-hand side of the equation above is an convex combination of $w^{(n_0)}, v_1, \cdots, v_m$. Since $w^{(n_0)}, v_1, \cdots, v_m$ are all in F, we obtain $0 \in F$.

Take $\alpha_n = 1/n$ and $F_n = F - \alpha_n w = \{x - \alpha_n w \mid x \in F\}$. Then F_n is a convex set whose closure does not contain 0. As what we have shown, there exist ξ_n such that $\forall x \in \mathbf{cl}\, F_n$, $\xi_n^T x > 0$. Without loss of generosity, we can assume that $\{\xi_n\}$ is bounded, so that there exists a convergent subsequence $\{\xi_{n_k}\}$ which converges to ξ for some $\xi \neq 0$, according to the Bolzano-Weierstrass' Theorem. Therefore, for every $x \in F$ we have

$$\xi^T x = \left(\lim_{k \to \infty} \xi_{n_k} \right)^T x = \lim_{k \to \infty} \xi_{n_k}^T \left(x - \alpha_{n_k} w \right) \geqslant 0.$$

$\square$