

Convex Optimization

Yunwei Ren

Contents

2	Convex Sets	2
2.1	Definition of convexity	2
2.2	Examples	2
2.3	Operations that preserve convexity	5
2.4	Separation theorems and supporting hyperplanes	5
3	Convex Functions	7
3.1	Definition of convexity	7
3.2	Examples	8
4	Convex Optimization Problems	9
4.1	Basic terminology and optimality conditions	9

2 Convex Sets

2.1 Definition of convexity

1.

Proof. For $k = 2$, $\theta_1 x_1 + \theta_2 x_2 \in C$ holds by definition. We argue by induction on k and assume that the inclusion holds for $k < m$. When $k = m$, denoting $\sum_{i=1}^{m-1} \theta_i$ by s ,

$$\sum_{i=1}^m \theta_i x_i = s \sum_{i=1}^{m-1} \frac{\theta_i x_i}{s} + \theta_m x_m.$$

Since $\sum_{i=1}^{m-1} \theta_i / s = 1$, by the induction hypothesis, $\sum_{i=1}^{m-1} \theta_i x_i / s \in C$. Meanwhile, as $s + \theta_m = 1$, $\sum_{i=1}^m \theta_i x_i \in C$, completing the proof. \square

2.

Proof. Clear that the intersection of two convex sets is still convex. Hence, the intersection of $C \subset \mathbb{R}^n$ and any line is convex as long as C is convex.

Now we suppose that the intersection of C and any line is convex. For any $x_1, x_2 \in C$, $C_l = C \cap \{\theta x_1 + (1 - \theta)x_2 : \theta \in \mathbb{R}\}$ is convex and therefore $\theta x_1 + (1 - \theta)x_2 \in C_l \subset C$ for every $0 \leq \theta \leq 1$. Thus, C is convex.

The above argument, *mutatis mutandis*, gives the second result. \square

3.

Proof. For every $\theta \in [0, 1]$, the process of bisecting the interval implies there exists a series $\langle \delta_n \rangle$ whose sum is θ . Hence, for every $a, b \in C$, $x_n = a + (b - a) \sum_{n=1}^{\infty} \delta_n$ converges to $a + \theta(b - a)$. Meanwhile, the midpoint convexity implies $x_n \in C$ for every n . And since C is closed, $a + \theta(b - a) \in C$. Thus, C is convex. \square

4.

Proof. Let D be the intersection of all convex sets containing C . If $x \in C$, then it is a convex combination of some points in C . Hence, for every convex set containing C , it contains x . Therefore, $\mathbf{conv} C \subset D$. For the converse, since $\mathbf{conv} C$ itself is a convex set containing C , $D \subset \mathbf{conv} C$. Thus, $\mathbf{conv} C = D$. \square

2.2 Examples

5.

Solution. $|b_2 - b_1| / \|a\|_2$. \square

7.

Proof. $\|x - a\|_2 \leq \|x - b\|_2$ iff $\langle x - a, x - a \rangle \leq \langle x - b, x - b \rangle$ iff $2\langle x, b - a \rangle \leq \langle b, b \rangle - \langle a, a \rangle$. Namely, $2(b - a)^T x \leq \|b\|_2^2 - \|a\|_2^2$. \square

2.8

Proof.

(a) It is trivial when a_1 and a_2 are linearly dependent, so we assume that a_1 and a_2 are linearly independent. We first tackle the problem for orthonormal a_1 and a_2 and then reduce the general situation to it.

Suppose that a_1 and a_2 are orthonormal. Let $S_0 = \text{span}(a_1, a_2)$ and (b_1, \dots, b_{n-2}) a basis of S_0^\perp . Then

$$x \in S_0 \iff \begin{bmatrix} b_1^T \\ \vdots \\ b_{n-2}^T \end{bmatrix} x = Bx = 0.$$

For $y = y_1 a_1 + y_2 a_2 \in S_0$, $y_1 \leq 1$ iff $a_1^T y \leq 1$ as (a_1, a_2) is an orthonormal basis of S_0 . Hence,

$$-1 \leq y_1, y_2 \leq 1 \iff \begin{bmatrix} a_1^T \\ a_2^T \\ -a_1^T \\ -a_2^T \end{bmatrix} y = Ay \preceq \mathbf{1}.$$

Thus, for orthonormal a_1 and a_2 , $S = \{x : Bx = 0, Ax \preceq \mathbf{1}\}$, a polyhedron.

Now we only assume the linear independence of a_1 and a_2 . We know that there exists some invertible n -by- n matrix¹ R such that $[\tilde{a}_1, \tilde{a}_2] = R[a_1, a_2]$ and \tilde{a}_1 and \tilde{a}_2 are orthonormal. Denoting the set described in the problem with respect to u_1 and u_2 by $S(u_1, u_2)$, $x \in S(a_1, a_2)$ iff $Rx \in S(\tilde{a}_1, \tilde{a}_2)$ iff $Rx \in \{x : \tilde{B}x = 0, \tilde{A}x \preceq \mathbf{1}\}$ where the meaning of \tilde{A} and \tilde{B} are described in the previous passage. Hence,

$$S(a_1, a_2) = \{x : \tilde{B}Rx = 0, \tilde{A}Rx \preceq \mathbf{1}\}.$$

(b) Yes, and the provided form has already satisfied the requirement.

(c) No. Note that $\langle x, y \rangle_2 \leq 1$ for all y with 2-norm 1 implies

$$\|x\|_2 = \langle x, x/\|x\| \rangle_2 \leq 1.$$

And by the Cauchy-Schwarz inequality, for every $\|x\| \leq 1$, $\langle x, y \rangle_2$ holds for every $\|y\|_2 = 1$. Hence, S is the intersection of the unit ball and $\{x : x \succeq 0\}$, which is not a polyhedron.

(d) Yes. Let $\tilde{S} = \{x \in \mathbb{R}^n : x \succeq 0, \|x\|_\infty \leq 1\}$, which is clearly a polyhedron since when $x \succeq 0$, $\|x\|_\infty \leq 1$ is equivalent to $[e_1, \dots, e_n]x \preceq \mathbf{1}$ where e_i is the i -th vector in the standard basis of \mathbb{R}^n .

Now we show that $S = \tilde{S}$. Suppose that $x \succeq 0$. If $\langle x, y \rangle_2 \leq 1$ for all y with 1-norm 1, then $x_i = \langle x, e_i \rangle_2 \leq 1$. Namely, $\|x\|_\infty \leq 1$. Meanwhile, if $\|x\|_\infty \leq 1$,

$$\langle x, y \rangle \leq \sum_{i=1}^n x_i |y_i| \leq 1$$

as it is just the weighted average of x_1, \dots, x_n . Hence, $S = \tilde{S}$, completing the proof. \square

¹We can use QR factorization to construct the matrix explicitly

2.9

Proof.

(a) By the definition,

$$\begin{aligned}
x \in V &\Leftrightarrow \|x - x_0\|_2^2 - \|x - x_i\|_2^2 \leq 0 \\
&\Leftrightarrow 2\langle x, x_i - x_0 \rangle \leq \langle x_i, x_i \rangle - \langle x_0, x_0 \rangle \quad \text{for } i = 1, \dots, K \\
&\Leftrightarrow 2 \begin{bmatrix} \langle x, x_1 - x_0 \rangle \\ \vdots \\ \langle x, x_K - x_0 \rangle \end{bmatrix} \preceq \begin{bmatrix} \|x_1\|_2^2 - \|x_0\|_2^2 \\ \vdots \\ \|x_K\|_2^2 - \|x_0\|_2^2 \end{bmatrix} \\
&\Leftrightarrow 2 \begin{bmatrix} (x_1 - x_0)^T \\ \vdots \\ (x_K - x_0)^T \end{bmatrix} x \preceq \begin{bmatrix} \|x_1\|_2^2 - \|x_0\|_2^2 \\ \vdots \\ \|x_K\|_2^2 - \|x_0\|_2^2 \end{bmatrix}
\end{aligned}$$

Hence, V is a polyhedron. Intuitively, the border of a Voronoi set are the lines with the same distances to x_0 and x_i .

(b) Suppose that $P = \{x : \alpha_k^T x \leq b_k, k = 1, \dots, K\}$. Let x_0 be any point of P and we construct the other points by reflection. For each k , let \tilde{x}_k be any point of $\{x : \alpha_k^T x = b_k\}$, $U_k = I - 2\alpha_k\alpha_k^T/\|\alpha_k\|_2^2$, the Householder matrix, and

$$R_k(x) = U_k(x - \tilde{x}_k) + \tilde{x}_k = x + 2\frac{\alpha_k}{\|\alpha_k\|_2^2}(b_k - \alpha_k^T x).$$

It is easy to verified that P is the Voronoi region of x_0 with respect to $R_1(x_0), \dots, R_K(x_0)$. \square

10.

Proof.

(a) Suppose $x_1, x_2 \in C$ and $\theta \in (0, 1)$. Let $x = \theta x_1 + (1 - \theta)x_2$. Since A is symmetric, $x_2^T A x_1 = x_1^T A x_2$. Thus,

$$\begin{aligned}
f(x) &= x^T A x + b^T x + c \\
&= \theta^2 x_1^T A x_1 + 2\theta(1 - \theta)x_1^T A x_2 + (1 - \theta)^2 x_2^T A x_2 \\
&\quad + \theta b^T x_1 + (1 - \theta)b^T x_2 + \theta c + (1 - \theta)c.
\end{aligned}$$

Note that

$$\begin{aligned}
\theta^2 x_1^T A x_1 + \theta b_1^T x_1 + \theta c &= \theta(x_1^T A x_1 + b_1^T x_1 + c) - \theta(1 - \theta)x_1^T A x_1 \\
&\leq -\theta(1 - \theta)x_1^T A x_1
\end{aligned}$$

and we can get a similar inequality for x_2 . Hence,

$$\begin{aligned}
f(x) &\leq -\theta(1 - \theta)(x_1^T A x_1 - 2x_1^T A x_2 + x_2^T A x_2) \\
&= -\theta(1 - \theta)(x_1 - x_2)^T A (x_1 - x_2) \leq 0
\end{aligned}$$

as $A \succeq 0$. Hence, C is convex.

(b) Put $H = \{x : g^T x + h = 0\}$, $B = A + \lambda g g^T$ and

$$C_B = \{x \in \mathbb{R}^n : x^T B x + b^T x + c - \lambda h^2 \leq 0\}.$$

By (a), C_B is convex and so does $C_B \cap H$. Suppose $x \in H$, then $x^T B x = x^T A x + \lambda h^2$. Therefore, $C_B \cap H = C$. Thus, C is convex. \square

2.3 Operations that preserve convexity

16.

Proof. For every $(a, b_1 + b_2), (c, d_1 + d_2) \in S$ and $0 \leq \theta \leq 1$, let

$$z_\theta = \theta(a, b_1 + b_2) + (1 - \theta)(c, d_1 + d_2) = (x, y_1 + y_2)$$

where

$$x = \theta a + (1 - \theta)c, \quad y_i = \theta b_i + (1 - \theta)d_i \quad \text{for } i = 1, 2.$$

Since S_i is convex and $(a, b_i), (c, d_i) \in S_i$,

$$(x, y_i) = \theta(a, b_i) + (1 - \theta)(c, d_i) \in S_i.$$

Hence, S is convex. □

18.

Proof. Let $\theta : \mathbb{R}^n \rightarrow \mathbb{R}^{n+1}$ be defined by $x \mapsto (x, 1)$ and $P : \mathbb{R}^{n+1} \rightarrow \mathbb{R}^n$ the perspective function. It can be verified that $f = P \circ Q \circ \theta$. Now we show that $g = P \circ Q^{-1} \circ \theta$ is the inverse of f . Clear that $P \circ \theta = I$, the identity map on \mathbb{R}^n . Hence,

$$f \circ g = P \circ Q \circ \theta \circ P \circ Q^{-1} \circ \theta = I.$$

Similarly, $g \circ f = I$. Thus, f is invertible and $g = f^{-1}$. □

2.4 Separation theorems and supporting hyperplanes

20.

Proof. Let $N = A$ and x_0 be such that $Ax_0 = b$. We prove the hint first. Suppose for all $x \in N$, $\langle x_0 + x, c \rangle = d$. Hence, $\langle x_0, c \rangle + \langle x, c \rangle = d$, which implies $\langle x, c \rangle = 0$ and

$$\langle x_0, c \rangle = d. \tag{1}$$

Since $\langle x, c \rangle = 0$ for all $x \in N$, $N = \text{null } A \subset \{c\}^\perp$ and therefore, $\text{range } A^T \supset \{c\}$. Thus, there exists a λ such that $A^T \lambda = c$. Substituting this into (1) yields

$$d = \langle x_0, A^T \lambda \rangle = \langle Ax_0, \lambda \rangle = b^T \lambda.$$

And the proof of the converse is straightforward.

Now we show the proposition. First we suppose such an x does not exist. Namely, $D = x_0 + N$ and \mathbb{R}_{++}^n are disjoint. Since D is an affine set and \mathbb{R}_{++}^n is convex and open, by the converse separating theorem, there exists some nonzero $c \in \mathbb{R}^n$ and scalar d such that $c^T y \leq d$ for all $y \in D$ and $c^T y \geq d$ for all $y \in C$. Since the image of an affine set under a linear mapping is still an affine set, $c^T y \leq d$ for all $y \in D$ implies $c^T y = d$ for all $y \in D$. Then, by our previous result, there exists a λ such that $c = A^T \lambda$ and $d = b^T \lambda$. Since $c \neq 0$, $A^T \lambda \neq 0$. Meanwhile, from $c^T y \geq d$ for all $y \in C$ we conclude $y \succeq 0$, otherwise we may choose $y \in C$ which is a large positive number on the position where the component of y is negative and zero elsewhere to lead to a contradiction. Thus, $A^T \lambda \succeq 0$. Finally, with the same approach, we conclude that $d \leq 0$ and therefore $b^T \lambda \leq 0$.

For the converse, our discussion shows that the existence of such a λ implies a separating hyperplane of C and D . Since C is open, it does not intersect with the separating hyperplane. Hence, there is no x satisfying $x \succ 0$ and $Ax = b$, completing the proof. □

22. TODO

23.

Proof. $A = \{(x, y) \in \mathbb{R}^2 : y \leq 0\}$ and $B = \{(x, y) \in \mathbb{R}^2 : x > 0, y \geq 1/x\}$. □

25.

Proof. Since $P_{\text{inner}} = \mathbf{conv}\{x_1, \dots, x_K\}$ is the smallest convex set that contains $\{x_1, \dots, x_K\}$, $\{x_1, \dots, x_k\} \subset C$ as C is closed and C is convex, $P_{\text{inner}} \subset C$. Meanwhile, it follows from the definition that $C \subset P_{\text{outer}}$. □

26.

Proof. If $C = D$, then clear that $S_C = S_D$. For the converse, we argue by contradiction. Assume the existence of some $x_0 \in C$ such that $x_0 \notin D$. Since D is closed and convex, there exists a hyperplane strictly separate x_0 and D , that is, there exists some nonzero $a \in \mathbb{R}^n$ and $b \in \mathbb{R}$ such that $a^T x < b$ for all $x \in D$ and $a^T x_0 > b$. Then by the definition of the support function,

$$S_C(a) \geq a^T x_0 > b > a^T x, \quad \text{for all } x \in D.$$

Hence, $S_C(a) > \sup_{x \in D} a^T x = S_D(a)$. Contradiction. Thus, $C \subset D$. Interchanging the roles of C and D yields $C \supset D$. Therefore, $C = D$. □

27. TODO

3 Convex Functions

3.1 Definition of convexity

1.

Proof.

(a) Clear that $\frac{b-x}{b-a}, \frac{x-a}{b-a} \geq 0$ and the sum of them is 1 for all $x \in [a, b]$. Hence, by the definition of convexity,

$$f(x) = f\left(\frac{b-x}{b-a}a + \frac{x-a}{b-a}b\right) \leq \frac{b-x}{b-a}f(a) + \frac{x-a}{b-a}f(b).$$

(b) By (a),

$$\frac{f(x) - f(a)}{x - a} \leq \frac{\frac{b-x}{b-a}f(a) + \frac{x-a}{b-a}f(b) - f(a)}{x - a} = \frac{f(b) - f(a)}{b - a}.$$

And a similar argument gives the second inequality.

(c) Just let x approach a and b respectively and we get these two inequalities.

(d) By (c), for every $a < b \in \text{dom } f$,

$$f'(b) - f'(a) \geq \frac{f(b) - f(a)}{b - a} - f'(a) \geq 0.$$

Let $a \rightarrow b-$ and we get $f''(b-) \geq 0$. Since f is twice differentiable, this implies $f''(b) \geq 0$. This argument, *mutatis mutandis*, yields $f''(a) \geq 0$. \square

3. There is another proof which shows the concavity by showing the convexity of **hypo** g . But I think there exists some faults related to the domain of f in that proof.

Proof. We show that g is concave. For every $y_1, y_2 \in (f(a), f(b))$, suppose $y_1 = f(x_1)$ and $y_2 = f(x_2)$. Since f is convex,

$$\frac{y_1 + y_2}{2} = \frac{f(x_1) + f(x_2)}{2} \geq f\left(\frac{x_1 + x_2}{2}\right).$$

Since f is increasing, so is g . Hence,

$$g\left(\frac{y_1 + y_2}{2}\right) \geq g\left(f\left(\frac{x_1 + x_2}{2}\right)\right) = \frac{x_1 + x_2}{2} = \frac{1}{2}g(y_1) + \frac{1}{2}g(y_2).$$

Thus, g is concave. \square

5. Running average of a convex function

Proof. Put $t = sx$, then

$$F(x) = \frac{1}{x} \int_0^1 f(sx) d(sx) = \int_0^1 f(sx) ds.$$

It can be verified that for fixed s , $f(sx)$ is convex in x . Hence, for every $\lambda \in (0, 1)$, $a, b \in \text{dom } F$,

$$F(\lambda a + (1 - \lambda)b) \leq \int_0^1 \{\lambda f(sa) + (1 - \lambda)f(sb)\} ds = \lambda F(a) + (1 - \lambda)F(b).$$

Thus, F is convex. \square

8. Second-order conditions for convexity

Proof. First we suppose the convexity of f . Then by definition, $\mathbf{dom} f$ is convex. \square

13.

Proof. Define $f(x) = \sum_{i=1}^n x_i \log x_i$. Some computation yields $D_{kl}(u, v) = f(u) - f(v) - \nabla f(v)^T(u - v)$. The inequality and the equality condition follows immediately from the fact that f is strictly convex. \square

3.2 Examples

16.

Solution.

(a) Convex. For every $x \in \mathbb{R}$, $f''(x) = e^x > 0$.

(b) Quasiconcave. For every $(x_1, x_2)^T \in \mathbb{R}_{++}^2$, $\nabla^2 f = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$, which is neither positive semidefinite nor negative semidefinite. Hence, f is not convex or concave. Its superlevel sets S_α , however, are convex as

$$\frac{(x_1 + x_2)(y_1 + y_2)}{4} \geq \sqrt{x_1 x_2 y_1 y_2} \geq \alpha$$

as long as $(x_1, x_2), (y_1, y_2) \in S_\alpha$.

(c) Convex. For every $(x_1, x_2) \in \mathbb{R}_{++}^2$,

$$\nabla^2 f(x_1, x_2) = \frac{1}{x_1 x_2} \begin{bmatrix} \frac{2}{x_1^2} & \frac{1}{x_1 x_2} \\ \frac{1}{x_1 x_2} & \frac{2}{x_2^2} \end{bmatrix}.$$

Since both $2/x_1^3 x_2$ and $\det(\nabla^2 f)$ are positive, $\nabla^2 f$ is positive definite. Thus, f is convex.

(d) Quasilinear. For every $(x_1, x_2) \in \mathbb{R}_{++}^2$,

$$\nabla^2 f(x_1, x_2) = \begin{bmatrix} 0 & -1/x_2^2 \\ -1/x_2^2 & 2x_1/x_2^3 \end{bmatrix},$$

which is neither positive nor negative semidefinite since $(x \pm \sqrt{x_1^2 + x_2^2})/x_2^3$, the eigenvalues of $\nabla^2 f$, always have different signs. However, since its sublevel sets $S_\alpha = \{(x_1, x_2) \in \mathbb{R}_{++}^2 : x_1/x_2 \leq \alpha\} = \{(x_1, x_2) \in \mathbb{R}_{++}^2 : [1, -\alpha][x_1, x_2]^T \leq 0\}$, which is convex, f is quasiconvex. Similarly, f is quasiconcave. Thus, f is quasilinear.

(e) Convex. For every $(x_1, x_2) \in \mathbb{R} \times \mathbb{R}_{++}$,

$$\nabla^2 f(x_1, x_2) = \begin{bmatrix} 1/x_2 & -2x_1/x_2^2 \\ -2x_1/x_2^2 & 2x_1^2/x_2^3 \end{bmatrix},$$

which is positive semidefinite since both $1/x_2$ and $\det(\nabla^2 f)$ are nonnegative.

(f) Concave. For every $(x_1, x_2) \in \mathbb{R}_{++}^2$,

$$\nabla^2 f(x_1, x_2) = \alpha(\alpha - 1)x_1^\alpha x_2^{1-\alpha} \begin{bmatrix} x_1^{-2} & -(x_1 x_2)^{-1} \\ -(x_1 x_2)^{-1} & x_2^{-2} \end{bmatrix},$$

which is negative definite since both $\alpha(\alpha - 1)x_1^\alpha x_2^{1-\alpha} x_1^{-2}$ and $\det(\nabla^2 f)$ are negative. \square

4 Convex Optimization Problems

4.1 Basic terminology and optimality conditions

1.

Solution.

- (a) f_0 attains its optimal $3/5$ at $(2/5, 1/5)$.
- (b) It is unbounded below.
- (c) Optimal value = 0; Optimal set = $\{(0, x_2) : x_2 \geq 1\}$.
- (d) f_0 attains its optimal $1/3$ at $(1/3, 1/3)$.

□