Convex Optimization

Yunwei Ren

Contents

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 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 its 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 $S_0 = \mathrm{span}(a_1, a_2)$ and $(b_1, \ldots, b_{n-2})$ a basis of $S_0^\perp$. Then

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

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,

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

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

Now we only assume the liner independence of a_1 and a_2. We know that there exists some invertible n-by-n matrix[1] 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 $Rx \in \{x : \, \tilde{B}x = 0, \, \tilde{A}x \preceq 1\}$ where the meaning of $\tilde{A}$ and $\tilde{B}$ are described in the previous passage. Hence,

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

(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 $[e_1, \ldots, e_n]x \preceq \mathbf{1}$ 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 $x_1, \ldots, x_n$. Hence, $S = \tilde{S}$, completing the proof. □

[1] We can use QR factorization to construct the matrix explicitly

2.9

Proof.
(a) By the definition,

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

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 $P = \{x : \alpha_k^T x \le b_k, k = 1, \dots, K\}$. 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 $R_1(x_0), \dots, R_K(x_0)$.
□

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

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

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

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

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

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

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

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.
□

2.3 Operations that preserve convexity

16.

Proof. For every $(a, b_1 + b_2), (c, d_1 + d_2) \in S$ and $0 \le \theta \le 1$, 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 $\theta : \mathbb{R}^n \to \mathbb{R}^{n+1}$ be defined by $x \mapsto (x, 1)$ and $P : \mathbb{R}^{n+1} \to \mathbb{R}^n$ 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 $x \in N$, $N = \operatorname{null} A \subset \{c\}^\perp$ and therefore, $\operatorname{range} A^T \supset \{c\}$. 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 $c^Ty \le d$ for all $y \in D$ and $c^Ty \ge d$ for all $y \in C$. Since the image of an affine set under a linear mapping is still an affine set, $c^Ty \le d$ for all $y \in D$ implies $c^Ty = d$ 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 $c \ne 0$, $A^T\lambda \ne 0$. Meanwhile, from $c^Ty \ge d$ 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 $d \le 0$ and therefore $b^T\lambda \le 0$.

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 $B = \{(x, y) \in \mathbb{R}^2 : x > 0, y \geq 1/x\}$. $\square$

25.

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

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, $C = D$. $\square$

27. TODO

2.5 Convex cones and generalized inequalities

31.

Solution.

(a) For $\lambda_1, \lambda_2 \in K^*$ and $\theta_1, \theta_2 > 0$, since

$$f = \langle \cdot, \theta_1 \lambda_1 + \theta_2 \lambda_2 \rangle = \theta_1 \langle \cdot, \lambda_1 \rangle + \theta_2 \langle \cdot, \lambda_2 \rangle,$$

f also maps K into $\mathbb{R}_+$. Namely, $\theta_1 \lambda_1 + \theta_2 \lambda_2 \in K^*$.

(b) If $f = \langle \cdot, \lambda \rangle$ maps K_2 into $\mathbb{R}_+$, then, as $K_1 \subset K_2$, it maps K_1 into $\mathbb{R}_+$. Thus, $K_2^* \subset K_1^*$.

(c) Suppose $(\lambda_n) \subset K^*$ be a sequence converging to $\lambda \in \mathbb{R}^n$. Then, by the continuity of the inner product, for every $x \in K$, $\langle x, \lambda \rangle = \lim_{n \to \infty} \langle x, \lambda_n \rangle \geq 0$. Hence, $\lambda \in K^*$. Namely, K^* is closed.

(d) If $y \in \mathbf{int}\, K^*$, then there exists some $\varepsilon > 0$ such that for all Δy with $\|\Delta y\| < \varepsilon$, $y + \Delta y \in K$, that is, $(y + \Delta y)^T x \geq 0$ for all $x \in \mathbf{cl}\, K$. For each x, put $\Delta y = -\varepsilon x / 2\|x\|$ and then we obtain $y^T x > 0$.

For the converse, suppose that $y \notin \mathbf{int}\, K^*$. Namely, for all $\varepsilon > 0$, there exists some Δy with $\|\Delta y\| < \varepsilon$ such that $(y + \Delta y)^T x_0 \leq 0$ for some $x_0 \in \mathbf{cl}\, K$. This time, put $\Delta y = \varepsilon x_0 / 2\|x_0\|$ and then we get $y^T x_0 \leq 0$.

(e) We argue by contradiction. Assume that there exists some nonzero $y \in K^*$ such that $-y \in K^*$. Then for every $x \in K$, $\langle x, \pm y \rangle \geq 0$, which yields $\langle x, y \rangle = 0$, i.e.,

$K \subset \{y\}^T$. Since $\dim\{y\}^T < n$, K can not have nonempty interior. Contradiction. Thus, K^* is pointed.

(f) For every $x \in \mathbf{cl}\, K$, $x^T y \geq 0$ for all $y \in K^*$. Hence, $x \in K^{**}$. Thus, $\mathbf{cl}\, K \subset K^{**}$. For the converse, note that $\mathbf{cl}\, K$, a closed convex cone, is fully determined by its supporting hyperplanes at the origin. Namely, if x satisfies $y^T x \geq 0$ for all $y \in (\mathbf{cl}\, K)^* = K^*$, then $x \in \mathbf{cl}\, K$. From this we conclude $K^{**} \subset \mathbf{cl}\, K$. Thus, $K^{**} = K$.

(g) We argue by contradiction. Assume that $\mathbf{int}\, K^*$ is empty. Then, by (d), if $y \in K^*$, then $y^T x = 0$ for all $x \in \mathbf{cl}\, K$. Namely, $K^* \subset (\mathbf{cl}\, K)^\perp$. Therefore, $(K^*)^\perp \supset \mathbf{cl}\, K = K^{**}$ where the equality comes from (f). Thus, for all $x \in K^**$, $-x^T y = x^T y = 0$ for all $y \in K^*$, which contradict the assumption that K is pointed. Thus, $\mathbf{int}\, K^* \neq \varnothing$. (This proof should be reviewed.) $\square$

32.

Solution. $\langle y, Ax \rangle \geq 0$ for all $x \succeq 0$ iff $\langle A^T y, x \rangle \geq 0$ for all $x \succeq 0$ iff $A^T y \succeq 0$. Hence, $K^* = \{y : A^T y \succeq 0\}$. $\square$

35.

Proof. Denote this set by $\mathcal{C}$. Note that $z^T X z = \mathbf{tr}(z z^T X)$. Hence, X is copositive iff $\langle z z^T, X \rangle \geq 0$ for all $z \succeq 0$. Namely,

$$\mathcal{C} = \bigcap_{z \succeq 0} \{X \in \mathbf{S}^n : \langle z z^T, X \rangle \geq 0\}, \tag{2}$$

the intersection of some half spaces. Hence, $\mathcal{C}$ is a closed convex cone. Since $\mathcal{C}$ contains the set of all positive semidefinite matrices, it is solid. Meanwhile, if $\pm X \in \mathcal{C}$, then $z^T X z = 0$ for all $z \succeq 0$. Hence, $X = 0$. Thus, $\mathcal{C}$ is a proper cone.

Note that $\mathcal{C}^*$ is just the collection of the inward normal vectors of supporting hyperplanes of $\mathcal{C}$ at the origin. By (2), $\mathcal{C}^* = \{z z^T : z \succeq 0\}$. $\square$

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 $a < b \in \mathbf{dom}\, f$,

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

Let $a \to b-$ 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

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

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

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

Thus, F is convex. $\square$

8. Second-order condition for convexity

Proof. First we prove the case $f : \mathbb{R} \to \mathbb{R}$. If f is convex, then $\mathbf{dom}\, f$ is convex by definition. Meanwhile, for every x and t, by the first-order condition,

$$\frac{f(x+t) - f(x) - f'(x)t}{t^2} \geq 0.$$

Let $t \to 0$ and we obtain $f''(x) \geq 0$. For the converse, $f''(x)$ implies that f' is monotonically increasing. Thus, by the mean-value theorem, there exists some c between x and y such that

$$f(y) - f(x) = f'(c)(y-x) \geq f'(x)(y-x),$$

Namely, f is convex.

Now we prove the general case. Recall that f is convex iff f is convex along all lines. For fixed $x, u \in \mathbb{R}^n$, define $g(t) = f(x+tu)$. By our previous result, g is convex iff

$$0 \leq g''(t) = u^T \nabla^2 f(x_0 + tu) u \quad \text{for all } t.$$

Namely, $\nabla^2 f(x) \succeq 0$ for all $x \in \mathbb{R}^n$. □

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. □

3.2 Examples

16.

Solution.

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

(b) Quasiconcave. For every $(x_1, x_2)^T \in \mathbb{R}^2_{++}$, $\nabla^2 f = \left[\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right]$, 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 $(x_1, x_2) \in \mathbb{R}^2_{++}$,

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

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 it sublevel sets $S_\alpha = \{(x_1, x_2) \in \mathbb{R}^2_{++} : x_1/x_2 \le \alpha\} = \{(x_1, x_2) \in \mathbb{R}^2_{++} : [1, -\alpha][x_1, x_2]^T \le 0\}$, 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 $(x_1, x_2) \in \mathbb{R}^2_{++}$,

$$\nabla^2 f(x_1, x_2) = \alpha(\alpha - 1)x_1^\alpha x_2^{1-\alpha} \begin{bmatrix} x_1^{-2} & -(x_1x_2)^{-1} \\ -(x_1x_2)^{-1} & x_2^{-2} \end{bmatrix},$$

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. $\square$

17.

Proof. Put $z_k = (x_1^k, \dots, x_n^k)$. Then the Hessian of f is

$$\nabla^2 f(x) = (1-p)(\mathbf{1}^T z_p)^{1/p-2}(z_{p-1}z_{p-1}^T - \mathbf{1}^T z_p \,\mathbf{diag}(z_{p-2})).$$

Put $K = (1-p)(\mathbf{1}^T z_p)^{1/p-2}$, a nonnegative constant. For every $v \in \mathbb{R}^n$,

$$\begin{aligned} v^T \nabla^2 f(x) v &= K v^T (z_{p-1}z_{p-1}^T - \mathbf{1}^T z_p \,\mathbf{diag}(z_{p-2}))v \\ &= K\left\{ \left(\sum_{i=1}^n v_i x_i^{p-1}\right)^2 - \left(\sum_{i=1}^n x_i^p\right)\left(\sum_{i=1}^n x_i^{p-2} v_i^2\right) \right\} \\ &\le 0, \end{aligned}$$

where the inequality comes from the Cauchy-Schwarz inequality $(a^T b)^2 \le (a^T a)(b^T b)$ with $a_i = x_i^{p/2}$ and $b_i = x_i^{p/2-1} v_i$. Thus, f is concave. $\square$

19. Nonnegative weighted sums and integrals

Proof.

(a) For each $k = 1, \dots, r$, let $f_k(x) = \sum_{i=1}^k x_{[i]}$, which is convex. Put

$$\beta_1 = \alpha_1 - \alpha_2, \quad \beta_2 = \alpha_2 - \alpha_3, \quad \dots \quad \beta_r = \alpha_r.$$

Since $\alpha_1 \ge \alpha_2 \ge \cdots \ge \alpha_r$, $\beta_i \ge 0$ for $i = 1, \dots, r$. Hence, $f = \beta_1 f_1 + \cdots + \beta_r f_r$, being a nonnegative weighted sum of convex functions, is convex.

(b) Note that $T(x, \omega)$ is linear in x for fixed ω. Hence, it can be verified via definition the convexity of $\mathbf{dom}\, f$ and $-\log T(x, \omega)$ is also convex in x. Hence, $f(x) = \int_0^{2\pi} \{-\log T(x, \omega)\} \mathrm{d}\omega$ is convex. $\square$

21.

Proof.

(a) By Prob.20(a), $\|A^{(i)}x - b^{(i)}\|$ is convex for each $i = 1, \dots, k$ and consequently f, the pointwise maximum of them, is convex.

(b) Let $E \subset \mathbb{R}^n$ be the collection of all vectors whose entries are ± 1 or 0. Then for each $c \in E$, $x \mapsto c^T x$ defines a convex function. Since $f(x) = \max_{c \in E} c^T x$, it is also convex. $\square$

22.(a)

Proof. Put $g(y) = \log(\sum_{i=1}^n e^{y_i})$ and $h(x) = Ax + b$ where $A = [a_1, \ldots, a_n]^T$ and $b = [b_1, \ldots, b_n]^T$. Then $j = g \circ h$ is convex on $\mathbb{R}^n$. Hence, $\mathbf{dom}\, f = \{x : j(x) < 1\}$ is convex. Meanwhile, $-j$ is concave, $-\log$ is convex and the extension of it to $\mathbb{R}$ is non-increasing. Therefore, $f(x) = -\log(-j(x))$ is convex. $\square$

3.3 Operations that preserve convexity

30. Convex hull or envelope of a function

Proof. Let h be any convex function such that $h(x) \le f(x)$ for all x. Then $\mathbf{epi}\, f \subset \mathbf{epi}\, h$. Since $\mathbf{conv}\,\mathbf{epi}\, f$ is the smallest convex set that contains $\mathbf{epi}\, f$ and $\mathbf{epi}\, h$ is convex as h is convex, $\mathbf{conv}\,\mathbf{epi}\, f \subset \mathbf{epi}\, h$. Namely, $(x,t) \in \mathbf{conv}\,\mathbf{epi}\, f$ implies $(x,t) \in \mathbf{epi}\, h$, that is, $h(x) \le t$. Take infimum on the both sides and we get $h(x) \le g(x)$. $\square$

31.

Proof.

(a) Note that $g(0) = 0$. Hence, if $t = 0$, $g(tx) = g(0) = 0 = tg(x)$. For $t > 0$, putting $\beta = \alpha/t$,

$$g(tx) = \inf_{\beta>0} \frac{f(\beta t x)}{\beta} = t \inf_{\alpha>0} \frac{f(\alpha x)}{\alpha} = tg(x).$$

(b) Let h be any homogenous underestimator of f. For every $\varepsilon > 0$, by definition, there is some β such that

$$g(x) + \varepsilon \ge \frac{f(\beta x)}{\beta} \ge \frac{h(\beta x)}{\beta} = h(x).$$

Since the choice of ε is arbitrary, this implies $g(x) \ge h(x)$.

(c) Consider the function $p : \mathbf{dom}\, f \times \mathbb{R}_{++}$, $(x, \alpha) \mapsto f(\alpha x)/\alpha$. Since $\mathbb{R}_{++}$ is convex and $g(x) = \inf_{\alpha>0} p(x, \alpha)$, g is convex as long as p is. Now we show the convexity of p. Note that

$$\begin{aligned} (x, \alpha, s) \in \mathbf{epi}\, p \quad &\iff \quad f(\alpha x)/\alpha \le s \\ &\iff \quad f(\alpha x) \le \alpha s \\ &\iff \quad (\alpha x, \alpha s) \in \mathbf{epi}\, f. \end{aligned}$$

As a consequence, p is convex as f is. $\square$

3.4 Conjugate functions

37. I assume that the space containing $\mathbf{dom}\, f$ is $\mathbb{S}^n$ so that $\mathbf{dom}\, f^* \subset \mathbb{S}^n$.

Proof. Define $g(X, Y) = \mathbf{tr}(YX) - f(X)$. First we show that for fixed $Y \notin -S_+^n$, g, as a function of X, is unbounded above. Since $Y \notin -S_+^n$, there exists some $\lambda_1 > 0$ and u with $\|u\| = 1$ such that $Yu = \lambda u$. Suppose $Y = S^{-1}\Lambda S$ where $\Lambda = \mathbf{diag}(\lambda_1, \ldots, \lambda_n)$. Put $X_k = \mathbf{diag}(k, 1, \ldots, 1)$. Then

$$g(X_k, Y) = \mathbf{tr}(\Lambda X_k) - \mathbf{tr}\,\mathbf{diag}(1/k, 1, \ldots, 1) = \lambda_1 k + \sum_{i=2}^n \lambda_i - \frac{1}{k} - n + 1 \to \infty$$

as $k \to \infty$. Hence, $\mathbf{dom}\, f^* \subset -\mathbb{S}^n_+$.

Then for $Y \in -\mathbb{S}^n_{++}$, $\nabla_X g(X, Y) = Y + X^{-2}$, which equals 0 at $X = (-Y)^{-1/2}$. Hence, $f^*(Y) = g((-Y)^{-1/2}, Y) = -2\,\mathbf{tr}(-Y)^{1/2}$. For $Y \in -\mathbb{S}^n_+$, there exists a sequence $(\varepsilon_k) \subset \mathbb{R}$ converges to 0 and $Y + \varepsilon_k I \in -\mathbb{S}^n_{++}$ for all k. Since $g(X, Y + \varepsilon_k I)$ is bounded above and $g(X, Y + \varepsilon_k I) \to g(X, Y)$ uniformly, $g(X, Y)$ is also bounded above. Hence, $\mathbf{dom}\, f^* = -\mathbb{S}^n_+$. Finally, by the continuity of f^*, which comes from the convexity, we conclude that $f^*(Y) = -2\,\mathbf{tr}(-Y)^{1/2}$ for all $y \in -\mathbb{S}^n_+$. □

38. Young's inequality I assume that f is continuous.

Proof. Since $\mathbf{dom}\, F = \mathbb{R}$ is closed and F is continuous, F is closed. Meanwhile, since f is increasing and $f \geq 0$, F is convex. Hence, $F = F^{++}$. Thus, it suffices to show that $G = F^*$.

Since f is continuous, F is differentiable. Hence,

Put $H(x, y) = yx - F(x)$. For fixed y,

$$H(x, y) = yx - \int_0^x f(a)\mathrm{d}a = \int_0^x \{y - f(a)\}\mathrm{d}a$$

attains its maximum at $x = g(y)$. Hence,

$$F^*(y) = H(g(y), y) = yg(y) - \int_0^{g(y)} f(a)\mathrm{d}a = G(y).$$

Thus, F and G are conjugates. Consequently, $xy \leq F(x) + G(y)$. □

40. Gradient and Hessian of conjugate function

Proof.

(a) The Legendre transformation yields

$$f^*(\bar{y}) = \bar{x}^T \nabla f(\bar{x}) - f(\bar{x}). \tag{3}$$

Differentiate (3) with respect to $\bar{x}$ yields

$$D_{\bar{x}} f^*(\bar{y}) = Df(\bar{x}) + \bar{x}^T \nabla^2 f(\bar{x}) - Df(\bar{x}) = \bar{x}^T \nabla^2 f(\bar{x}).$$

Meanwhile, the chain rule yields

$$D_{\bar{x}} f^*(\bar{y}) = D(f^* \circ \nabla f)(\bar{x}) = Df^*(\bar{y}) \nabla^2 f(\bar{x}).$$

These two equations gives

$$Df^*(\bar{y}) = \bar{x}^T \nabla^2 f(\bar{x}) (\nabla^2 f(\bar{x}))^{-1} = \bar{x}^T.$$

Namely, $\nabla f^*(\bar{y}) = \bar{x}$.

(b) Differentiate $\nabla f^*(\bar{y}) = \bar{x}$ with respect to $\bar{x}$ and we get $\nabla^2 f^*(\bar{y}) \nabla^2 f(\bar{x}) = I$. Thus, $\nabla^2 f^*(\bar{y}) = \nabla^2 f(\bar{x})^{-1}$. □

3.5 Quasiconvex functions

43.

Proof. Since f is quasiconvex iff its restriction to every line is, it suffices to prove the result for functions on $\mathbb{R}$. If $f : \mathbb{R} \to \mathbb{R}$ is quasiconvex, then, by definition, $\mathbf{dom}\, f$ is convex and for every $x, y \in \mathbf{dom}\, f$ with $f(x) \geq f(y)$ and $\theta \in (0, 1]$,

$$f(x + \theta(y - x)) \leq \max\{f(x), f(y)\} = f(x).$$

Thus,

$$f'(x)(y - x) = \lim_{\theta \to 0} \frac{f(x + \theta(y - x)) - f(x)}{\theta} \geq 0.$$

For the converse, we argue by contradiction. Assume that there exists some $x < y \in \mathbf{dom}\, f$ and $c \in (x, y)$ such that $f(c) > \max\{f(x), f(y)\}$. Define $D = \{z \in [x, y] : f(z) = f(c)\}$. Since D is bounded, $d = \inf D > -\infty$. By the continuity of f, $f(d) = f(c) > f(x)$. Namely, d is the leftmost point in $[x, y]$ the function value at which is $f(c)$. Similarly, we may find the rightmost point z in $[x, d]$ the function value at which is $f(x)$. Since $f(d) > f(z)$, there exists some $\xi \in (z, d)$ such that $f'(\xi) > 0$. Meanwhile, by our construction and the continuity of f, $f(\xi) \geq f(d) = f(x) \geq f(y)$. However, $f'(\xi)(y - \xi) > 0$. Contradiction. Thus, such a c does not exist and, consequently, f is quasiconvex. □

3.6 Log-concave and log-convex functions

47.

Proof. f is log-concave iff $\log f$ is concave iff for every $x, y \in \mathbf{dom}\, f$,

$$\log f(y) - \log f(x) \leq \frac{\nabla f(x)^T}{f(x)}(y - x) \quad \Longleftrightarrow \quad \frac{f(y)}{f(x)} \leq \exp\left(\frac{\nabla f(x)^T}{f(x)}(y - x)\right).$$

□

54.

Proof. (a) Some calculation yields

$$f'(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, \quad f''(x) = -\frac{x}{\sqrt{2\pi}} e^{-x^2/2}.$$

For $x \geq 0$, clear that the left hand side is nonpositive while the right hand side is nonnegative. Hence, $f''(x)f(x) \leq f'(x)^2$.

(b) It follows immediately from the AM–GM inequality.

(c) Take exponentials on the both side of the inequality in (b) and we get $e^{-t^2/2} \leq e^{x^2/2 - xt}$. Then integrating over t gives the other inequality.

(d) For $x \leq 0$, by (c),

$$-xe^{-x^2/2} \int_{-\infty}^{x} e^{-t^2/2} \mathrm{d}t \leq -x \int_{-\infty}^{x} e^{-xt} \mathrm{d}t = e^{-x^2}.$$

Thus, $f''(x)f(x) \leq f'(x)^2$. □

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 $(1/3, 1/3)$.

$\square$