3 Convex Conic Optimization

We saw from week 1 a number of examples of optimization problems with vectors or symmetric matrices as variables, and with constraints that are either linear or involve the 2-norm or positive (semi)definiteness. These problems belong to the class of *convex conic optimization* problems, which we now study in more depth.

3.1 Spaces

It is clear that we need to work with both the space of n-dimensional real column vectors $\mathbb{R}^n$ and the space of $n \times n$ real symmetric matrices, which we denote by $\mathbb{S}^n$, i.e.,

$$\mathbb{S}^n = \{X \in \mathbb{R}^{n\times n} \mid X = X^T\}.$$

For $X \in \mathbb{S}^n$, its trace is $\text{tr}[X] = \sum_{i=1}^n x_{ii} = \sum_i \lambda_i(X)$. More generally, let $\mathbb{H}$ be a finite-dimensional real linear space[4] equipped with an inner product $\langle x, y\rangle$[5] and norm $\|x\| = \sqrt{\langle x, x\rangle}$. This is a "Hilbert space." If this feels too abstract, then just consider the two cases of interest:

- $\mathbb{H} = \mathbb{R}^n,\ \langle x, y\rangle = x^T y = \sum_{i=1}^n x_i y_i,\ \|x\| = \|x\|_2,$
- $\mathbb{H} = \mathbb{S}^n,\ \langle X, Y\rangle = \text{tr}[X^T Y] = \sum_{i,j=1}^n x_{ij} y_{ij},\ \|X\| = \|X\|_F$ (so-called "Frobenius-norm").

We can identify each $X = [x_{ij}] \in \mathbb{S}^n$ 1-to-1 with the column vector $(x_{ij})_{i\le j} \in \mathbb{R}^{\frac{n(n+1)}{2}}$, and identify $\langle X, Y\rangle$ on $\mathbb{S}^n$ with the inner product $\langle (x_{ij})_{i\le j}, (y_{ij})_{i\le j}\rangle = \sum_i x_{ii}y_{ii} + 2\sum_{i<j} x_{ij}y_{ij}$ on $\mathbb{R}^{\frac{n(n+1)}{2}}$. However, we won't working with $\mathbb{R}^{\frac{n(n+1)}{2}}$ since operations such as matrix-matrix multiplication or matrix inversion become too complicated and unintuitive.

3.2 Convex sets and cones

Consider a set $K \subseteq \mathbb{H}$. We say K is *convex* if

$$\alpha x + (1-\alpha) y \in K \quad \forall x, y \in K,\ \forall 0 \le \alpha \le 1.$$

(Here we use lower case letters for the elements of $\mathbb{H}$ in general. We will switch to upper case letters in the special case where $\mathbb{H}$ is a space of matrices.) We say K is a *cone* if

$$\alpha x \in K \quad \forall x \in K,\ \forall \alpha \ge 0.$$

We denote the open ball by $\mathbb{B}(x, r) = \{y \in \mathbb{H} \mid \|y - x\| < r\}$, for $x \in \mathbb{H}$, $r > 0$. The interior of K is

$$\text{int}K = \{x \in K \mid \mathbb{B}(x, r) \subseteq K \text{ for some } r > 0\}.$$

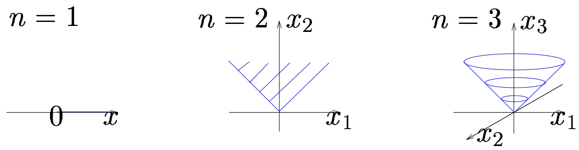
[4] so $x, y \in \mathbb{H}$ implies $x + y \in \mathbb{H}$ and $\alpha x \in \mathbb{H}\ \forall \alpha \in \mathbb{R}$.

[5] so $\langle x, y\rangle = \langle y, x\rangle$, $\langle x, x\rangle > 0$ for $x \neq 0$, and $\langle x, \cdot\rangle$, $\langle \cdot, y\rangle$ are linear.

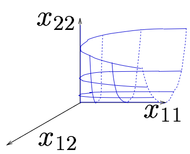
Then K is open if and only if (abbreviated as "iff" or "$\Longleftrightarrow$") $K = \text{int}K$. K is closed iff $\mathbb{H} \setminus K$ is open or, equivalently, the limit of any convergent sequence in K also lies in K, i.e., $x_k \in K$, $k = 1, 2, \ldots$, and $\lim_{k\to\infty} x_k = x$ (i.e., $\lim_{k\to\infty} \|x_k - x\| = 0$) implies $x \in K$.

Three exampes of closed convex cones are:

- $K_1 = \{x \in \mathbb{R}^n \mid x \geq 0\} = [0, \infty)^n$, "nonnegative orthant"
- $K_2 = \left\{x \in \mathbb{R}^n \,\middle|\, \sqrt{x_1^2 + \cdots + x_{n-1}^2} \leq x_n\right\}$, "second-order cone (SOC)"



- $K_3 = \{X \in \mathbb{S}^n \mid X \succeq 0\}$. "semidefinite cone"

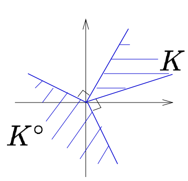


$n = 2$

$$\begin{pmatrix} x_{11} & x_{12} \\ x_{12} & x_{22} \end{pmatrix} \succeq 0 \Longleftrightarrow x_{11} \geq 0,\ x_{22} \geq 0,\ x_{11}x_{22} \geq x_{12}^2$$

For example, K_3 is closed since $X_k \in K_3$ iff $0 \leq u^T X_k u$ for all $u \in \mathbb{R}^n$, so that $\lim_{k\to\infty} X_k = X$ implies $0 \leq \lim_{k\to\infty} u^T X_k u = u^T X u$ and hence $X \in K_3$. K_3 is convex since, for $X, Y \in K_3$, we have $0 \leq u^T X u$ and $0 \leq u^T Y u$ for all $u \in \mathbb{R}^n$, so that $0 \leq \alpha\ u^T X u + (1-\alpha)\ u^T Y u = u^T(\alpha X + (1-\alpha)Y)u$ for all $0 \leq \alpha \leq 1$, and hence $\alpha X + (1-\alpha)Y$ (which is symmetric since X, Y are) is in K_3. K_3 is a cone by a similar reasoning. It's not hard to see that

$$\begin{aligned} \text{int}K_1 &= (0, \infty)^n, \\ \text{int}K_2 &= \left\{x \in \mathbb{R}^n \,\middle|\, \sqrt{x_1^2 + \cdots + x_{n-1}^2} < x_n\right\}, \\ \text{int}K_3 &= \{X \in \mathbb{S}^n \mid X \succ 0\}. \end{aligned}$$



For a cone $K \subseteq \mathbb{H}$, its *polar* is

$$K^\circ = \{y \mid \langle x, y\rangle \leq 0\ \forall x \in K\}.$$

We say that K is *self-dual* if $C^\circ = -C = \{-x \mid x \in C\}$. It's not hard to verify that K_1, K_2, K_3 are self-dual. In fact, these cones are very special. They are symmetric cones, with associated Jordan algebra (see the 1994 book *Analysis on Symmetric Cones* by Faraut and Korányi). This connection is useful in developing efficient primal-dual interior-point methods for solving the application problems we saw in the first week.

3.3 Convex functions

Consider a function $f : \mathbb{H} \to \mathbb{R} \cup \{\infty\}$. We say f is *proper* if its effective domain

$$\mathrm{dom} f = \{x \in \mathbb{H} \mid f(x) < \infty\}$$

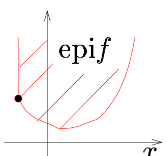
is nonempty. We say f is *closed* (equivalently, *lower semicontinuous*) if its epigraph

$$\mathrm{epi} f = \{(x, \alpha) \in \mathbb{H} \times \mathbb{R} \mid f(x) \le \alpha\}$$

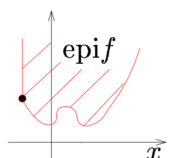
is a closed set. (This is needed for the attainment of minimum in minimization problems.) We say f is *convex* if $\mathrm{epi} f$ is a convex set or, equivalently,

$$f(\alpha x + (1-\alpha) y) \le \alpha f(x) + (1-\alpha) f(y) \quad \forall x, y \in \mathbb{H}, \ \forall 0 \le \alpha \le 1. \tag{14}$$

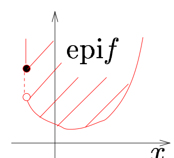
For example, the 2-norm $\|\cdot\|_2$ on $\mathbb{R}^{n-1}$ is a proper closed convex function and its epigraph is K_2. More generally, the norm $\|\cdot\|$ on $\mathbb{H}$ is proper, closed, convex. If $\mathrm{dom} f$ is closed and f is continuous on $\mathrm{dom} f$, then f is closed. However, a convex function f can be closed without being continuous on $\mathrm{dom} f$, so the notion of a closed function is weaker than continuity.



f is closed, convex



f is not convex



f is not closed

It's not hard to check, using basic properties of log function, that the following three functions are proper and closed.

$$f_1(x) = \begin{cases} -\sum_{i=1}^n \log x_i & \text{if } x \in \mathrm{int} K_1 \\ \infty & \text{else} \end{cases}$$

$$f_2(x) = \begin{cases} -\log\left(x_n^2 - (x_1^2 + \cdots + x_{n-1}^2)\right) & \text{if } x \in \mathrm{int} K_2 \\ \infty & \text{else} \end{cases}$$

$$f_3(X) = \begin{cases} -\log \det X & \text{if } X \in \mathrm{int} K_3 \\ \infty & \text{else} \end{cases}$$

Notice that $\mathrm{dom} f_j = \mathrm{int} K_j$, $j = 1, 2, 3$. We will see that they are also convex. In fact, they are very special functions associated with K_1, K_2, K_3 and will be used in interior-point methods.

Another important connection between convex function and convex set is the following: If $f : \mathbb{H} \to \mathbb{R} \cup \{\infty\}$ is proper, closed, convex, then the lower level set

$$\{x \in \mathbb{H} \mid f(x) \le \alpha\}$$

is closed convex (possibly empty) for any $\alpha \in \mathbb{R}$. Thus the unit-ball $\{x \in \mathbb{H} \mid \|x\| \le 1\}$ is closed and convex.

It's usually easy to check that a function $f : \mathbb{H} \to \mathbb{R} \cup \{\infty\}$ is proper and closed. How do we check that f is convex? There is a limited calculus for convex functions that allows complex convex functions to be built from "elementary" convex functions:

- Let $f : \mathbb{H} \to \mathbb{R} \cup \{\infty\}$ be closed, convex. For any linear mapping A from another real linear space $\mathbb{H}'$ to $\mathbb{H}$, the composite function $x' \mapsto f(Ax')$ is convex.
- Let $f : \mathbb{H} \to \mathbb{R} \cup \{\infty\}$ be closed, convex. For any function $\psi : \mathbb{R} \to \mathbb{R} \cup \{\infty\}$ that is closed, convex, increasing, the composite function $x \mapsto \psi(f(x))$ is convex.
- Let $f_1, f_2 : \mathbb{H} \to \mathbb{R} \cup \{\infty\}$ be closed, convex. The pointwise-sum function $x \mapsto f_1(x) + f_2(x)$ and the pointwise-maximum function $x \mapsto \max\{f_1(x), f_2(x)\}$ are also closed, convex.

The above properties are easily proved using the definition (14) of convex functions.

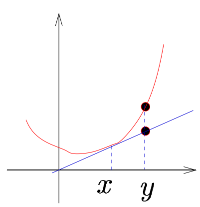
If f is twice continuously differentiable, then the convexity of f is characterized by the positive semidefiniteness of its Hessian. We say $f : \mathbb{H} \to \mathbb{R} \cup \{\infty\}$ is *differentiable* at $x \in \text{int}(\text{dom}f)$ if there exists an element of $\mathbb{H}$, which depends on x and is usually denoted $\nabla f(x)$, that satisfies

$$f(x+d) - f(x) = \langle \nabla f(x), d \rangle + o(\|d\|) \quad \forall d \in \mathbb{B}(0, r),$$

where $\lim_{\alpha \to 0^+} \frac{o(\alpha)}{\alpha} = 0$ and $r > 0$ is sufficiently small. It is not hard to show that if $\text{dom}f$ is open, convex, and f is differentiable on $\text{dom}f$, then f is convex if and only if

$$f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle \quad \forall x, y \in \text{dom}f.^{6} \tag{15}$$

(Geometrically, the above inequality says the graph of f lies above its tangent plane at $(x, f(x))$ for all $x \in \text{dom}f$.) This result is very useful when analyzing the convergence or complexity of gradient methods.



We say $f : \mathbb{H} \to \mathbb{R} \cup \{\infty\}$ is *twice differentiable* at $x \in \text{int}(\text{dom}f)$ if f is differentiable on an open set containing x and there exists a linear mapping from $\mathbb{H}$ to $\mathbb{H}$, which depends on x and is often denoted $\nabla^2 f(x)$, that satisfies

$$\nabla f(x+d) - \nabla f(x) = \nabla^2 f(x) d + o(\|d\|) \quad \forall d \in \mathbb{B}(0, r), \tag{16}$$

where $\lim_{\alpha \to 0^+} \frac{o(\alpha)}{\alpha} = 0$ and $r > 0$ is sufficiently small. Since $\mathbb{H}$ is finite-dimensional, $\nabla^2 f(x)$ may be represented as a matrix. It can be shown that if $\text{dom}f$ is open, convex, and f is twice continuously differentiable on $\text{dom}f$, then $\nabla^2 f(x)$ is self-adjoint for all $x \in \text{dom}f$ and f is convex if and only if

$$\nabla^2 f(x) \succeq 0 \quad \forall x \in \text{dom}f. \tag{17}$$

[6] Why? If f is convex, then (14) implies $f(y + \alpha(x-y)) - f(y) \leq \alpha(f(x) - f(y))\ \forall\ 0 < \alpha \leq 1$, so taking limit as $\alpha \to 0$ yields (15) (with "x" and "y" switched). Conversely, if (15) holds, then for any $x, y \in \text{dom}f$ and $0 \leq \alpha \leq 1$, we have $\alpha x + (1-\alpha)y \in \text{dom}f$ so that

$$\begin{aligned} f(y) &\geq f(\alpha x + (1-\alpha)y) + \langle \nabla f(\alpha x + (1-\alpha)y), \alpha(y-x) \rangle, \\ f(x) &\geq f(\alpha x + (1-\alpha)y) + \langle \nabla f(\alpha x + (1-\alpha)y), (1-\alpha)(x-y) \rangle. \end{aligned}$$

Multiplying the two inequalities by $1-\alpha$ and α, respectively, and summing yields (14).

Here, a linear mapping $L : \mathbb{H} \to \mathbb{H}$ being self-adjoint means $\langle Lu, v\rangle = \langle u, Lv\rangle$ for all $u, v \in \mathbb{H}$, and $L \succeq 0$ means $\langle u, Lu\rangle \geq 0$ for all $u \in \mathbb{H}$. Also, f being twice continuously differentiable at x means $\lim_{y\to x} \nabla^2 f(y) = \nabla^2 f(x)$, i.e., $\lim_{y\to x} |||\nabla^2 f(y) - \nabla^2 f(x)||| = 0$ with operator-norm $|||L||| = \max_{\|u\|=1} \|Lu\|$. In addition, $\nabla^2 f(x) \succ 0$ for all $x \in \text{dom}f$ implies f is strictly convex, i.e.,

$$f(\alpha x + (1-\alpha)y) < \alpha f(x) + (1-\alpha) f(y) \quad \forall\, x \neq y \in \text{dom}f,\ 0 < \alpha < 1.$$

Strict convexity is useful for showing uniqueness of the optimal solution.

When $\mathbb{H} = \mathbb{R}^n$, it can be seen that

$$\nabla f(x) = \left(\frac{\partial f}{\partial x_i}(x)\right)_{i=1}^n, \qquad \nabla^2 f(x) = \left(\frac{\partial^2 f}{\partial x_i \partial x_j}(x)\right)_{i,j=1}^n$$

In particular, we have that

$$\nabla f_1(x) = -\left(\frac{1}{x_i}\right)_{i=1}^n, \qquad \nabla^2 f_1(x) = \text{diag}\left(\frac{1}{x_i^2}\right)_{i=1}^n \qquad \forall x \in (0,\infty)^n,$$

and similarly for f_2. Since a diagonal matrix with positive diagonals is positive definite (the diagonals are its eigenvalues), we have $\nabla^2 f_1(x) \succ 0$ for all $x \in \text{dom}f_1$. Since $x_i \mapsto \frac{1}{x_i^2}$ is continuous on $(0,\infty)$, it follows that $\nabla^2 f_1$ is continuous on $\text{dom}f_1$. Thus f_1 is strictly convex. (This is not the only way to prove f_1 is strictly, though the gradient and Hessian formulas will be useful later in studying interior-point methods.) It can be similarly argued that f_2 is strictly convex. For f_3, the above approach does not work. Fortunately, we know from matrix analysis (see, e.g., the 1997 book *Matrix Analysis* by Bhatia) that

$$\nabla f_3(X) = -X^{-1}, \qquad \nabla^2 f_3(X)D = X^{-1}DX^{-1} \ \ \forall D \in \mathbb{S}^n, \qquad \forall X \succ 0.$$

(The formula for $\nabla^2 f_3(X)D$ can be derived from that for $\nabla f_3(X)$ using the definition (16).) Thus

$$\langle D, \nabla^2 f_3(X)D\rangle = \text{tr}[DX^{-1}DX^{-1}] = \text{tr}[X^{-\frac{1}{2}}DX^{-\frac{1}{2}}X^{-\frac{1}{2}}DX^{-\frac{1}{2}}] = \|X^{-\frac{1}{2}}DX^{-\frac{1}{2}}\|_F^2 \geq 0,$$

where the second equality we use the fact that $X = X^{\frac{1}{2}}X^{\frac{1}{2}}$ and $\text{tr}[AB] = \text{tr}[BA]$ for $A, B \in \mathbb{R}^{n\times n}$. Thus $\nabla^2 f_3(X) \succ 0$ for all $X \succ 0$. (It can be argued similarly that $\nabla^2 f_3(X)$ is self-adjoint.) It's also not hard to see that $\nabla^2 f_3$ is continuous on $\text{dom}f_3$. Thus f_3 is strictly convex.

3.4 Primal problem

We can now state the conic optimization problem. The "primal" form of the problem looks much like LP:

$$\begin{array}{ll} \max_x & \langle c, x\rangle \\ \text{s.t.} & \langle a_i, x\rangle = b_i, \quad i = 1, \ldots, m \\ & x \in K, \end{array} \tag{18}$$

where $a_i \in \mathbb{H}$, $b_i \in \mathbb{R}$, $c \in \mathbb{H}$, and $K \subseteq \mathbb{H}$ is a nonempty closed convex cone. Three special cases of particular interest are linear program (LP), second-order cone program (SOCP), and semidefinite program (SDP):

- $K_1 = [0, \infty)^n$, "LP"
- $K_2 = \mathrm{SOC}^{n_1} \times \cdots \times \mathrm{SOC}^{n_N}$, with $\mathrm{SOC}^n = \left\{ x \in \mathbb{R}^n \,\middle|\, \sqrt{x_1^2 + \cdots + x_{n-1}^2} \leq x_n \right\}$, "SOCP"
- $K_3 = \{X \in \mathbb{S}^n \mid X \succeq 0\}$. "SDP"

These three cases include (most) of the application problems we saw in the first week (will say more later). Here we continue to follow the convention of using upper case letters to denote the elements of $\mathbb{H}$ when $\mathbb{H} = \mathbb{S}^n$.

Since $[0, \infty) = \mathrm{SOC}^1$, we have

$$[0, \infty)^n = \mathrm{SOC}^1 \times \cdots \times \mathrm{SOC}^n,$$

so LP is a special case of SOCP. Moreover, it can be seen that

$$x \in \mathrm{SOC}^n \qquad \Longleftrightarrow \qquad \begin{pmatrix} x_n & & & x_1 \\ & \ddots & & \vdots \\ & & x_n & x_{n-1} \\ x_1 & \cdots & x_{n-1} & x_n \end{pmatrix} \succeq 0,^{7}$$

from which it's not hard to see that SOCP is a special case of SDP (i.e., any SOCP can be reformulated as an SDP). However, LP and SOCP have special structures that can be exploited to find solutions more efficiently than for a general SDP. (For example, LP can be solved by the simplex method, whereas SOCP and SDP cannot. LP can be solved in polynomial time under the Turing machine model of computation whereas this remains open for SOCP and SDP. LP enjoys stronger duality properies than SOCP and SDP, as we will see.) Thus, for reasons both theoretical and algorithmic, the three problems are often treated separately.

---

[7] Why? By a property of Schur complement, for any $A \succ 0$, $b \in \mathbb{R}^{n-1}$, and $c \in \mathbb{R}$, we have $\begin{pmatrix} A & b \\ b^T & c \end{pmatrix} \succeq 0$ if and only if $c - b^T A^{-1} b \geq 0$. Thus if $x_n > 0$, applying this property with $A = x_n I$, $b^T = \begin{pmatrix} x_1 & \cdots & x_{n-1} \end{pmatrix}$, $c = x_n$ and simplifying shows that the two sides are equal. If $x_n = 0$, then it is not hard to see that we must have $x_1 = \cdots = x_{n-1} = 0$ on both sides.