Convex Optimization: Reading Notes 3

GKxx

September 5, 2022

1 Dual cones and generalized inequalities

Definition 1.1 (Dual cone)**.** *The* dual cone *of a cone* K *is defined as*

$$K^* = \left\{ y \mid \forall x \in K, y^T x \geqslant 0 \right\}.$$

Geometrically, $y \in K^*$ if and only if $-y$ is the normal of a hyperplane that supports K at the origin.

Proposition 1.2. *The dual cone of a linear subspace* $V \subseteq \mathbb{R}^n$ *is its orthogonal complement.*

Proof. Note that whenever $x \in V$, $-x$ is also contained in V, and by definition we can see $V^* = V^\perp$. □

Proposition 1.3. *The* nonnegative orthant $\mathbb{R}_+^n$ *is a self-dual cone, i.e.* $(\mathbb{R}_+^n)^* = \mathbb{R}_+^n$.

Proposition 1.4. *The* positive semidefinite cone *is self-dual.*

Proof. We first show that $(\mathbb{S}_+^n)^* \subseteq \mathbb{S}_+^n$. Suppose $Y \in (\mathbb{S}_+^n)^*$. Assume that $Y \notin \mathbb{S}_+^n$, then there exists $\xi \in \mathbb{R}^n$ such that $\xi^T Y \xi < 0$, which means $\mathbf{Tr}\left(\xi^T Y \xi\right) = \mathbf{Tr}\left(\xi\xi^T Y\right) < 0$, while $\xi\xi^T \in \mathbb{S}_+^n$, a contradiction.

Then suppose $Y \in \mathbb{S}_+^n$ and we show that $Y \in (\mathbb{S}_+^n)^*$. For every $X \in \mathbb{S}_+^n$, we write X as its eigenvalue decomposition

$$X = \sum_{i=1}^n \lambda_i u_i u_i^T, \quad \lambda_i \geqslant 0, u_i \in \mathbb{R}^n.$$

We have that

$$\mathbf{Tr}\,(XY) = \mathbf{Tr}\left(\sum_{i=1}^n \lambda_i u_i u_i^T Y\right) = \sum_{i=1}^n \lambda_i \,\mathbf{Tr}\left(u_i^T Y u_i\right) \geqslant 0.$$

Hence $(\mathbb{S}_+^n)^* = \mathbb{S}_+^n$. □

Definition 1.5 (Dual norm)**.** *For any norm* $\|\cdot\|$*, the* dual norm *is defined as*

$$\|u\|_* = \sup\left\{ u^T x \mid \|x\| \leqslant 1 \right\}.$$

Proposition 1.6. *The dual of a norm cone* $K = \left\{(x, t) \in \mathbb{R}^{n+1} \mid \|x\| \leqslant t\right\}$ *is the cone defined by the dual norm*

$$K^* = \left\{(y, u) \in \mathbb{R}^{n+1} \mid \|y\|_* \leqslant u\right\}.$$

Proof. We have to show that $\|y\|_* \leqslant u$ if and only if for every $x \in \mathbb{R}^n, t \geqslant 0$ with $\|x\| \leqslant t$, we have $y^T x + tu \geqslant 0$.

Suppose $\|y\|_* \leqslant u$, which gives by definition that $\sup\left\{y^T\xi \mid \|\xi\| \leqslant 1\right\} \leqslant u$, so $\forall \|\xi\| \leqslant 1, y^T\xi \leqslant u$. Therefore $\forall x \in \mathbb{R}^n, t \geqslant 0$ with $\|x\| = \|-x\| \leqslant t$, we have $y^T(-x) \leqslant tu \Rightarrow y^T x + tu \geqslant 0$.

Now suppose that $y^T x + tu \geqslant 0$ holds for every $x \in \mathbb{R}^n, t \geqslant 0$ with $\|x\| \leqslant t$. Assume that $\|y\|_* > u$, which implies that $\exists \xi \in \mathbb{R}^n$ with $\|\xi\| \leqslant 1$ such that $y^T\xi > u$. This gives that $y^T(-\xi) + 1 \cdot u < 0$ with $\|-\xi\| \leqslant 1$, a contradiction. □

Proposition 1.7. *The dual of the* ℓ_p*-norm is the* ℓ_q*-norm, where* $1/p + 1/q = 1, p, q > 0$.

Proof. We have to show that

$$\|u\|_q = \sup_x \left\{u^T x \mid \|x\|_p \leqslant 1\right\}$$

holds for every u and any $p, q > 0$ satisfying $1/p + 1/q = 1$. When $u = 0$ both sides are zero, so we only consider the case where $u \neq 0$. By Holder's inequality,

$$u^T x \leqslant \left\|u^T x\right\|_1 \leqslant \|u\|_q \|x\|_p \leqslant \|u\|_q,$$

so if suffices to find a vector x with $\|x\|_p \leqslant 1$ such that $\|u\|_q = u^T x$. Take

$$y = \left[\operatorname{sgn}(u_i) |u_i|^{q-1}\right]_{i=1}^{n},$$

so that

$$u^T y = \sum_{i=1}^{n} |u_i|^q = \|u\|_q^q,$$

and

$$\|y\|_p^p = \sum_{i=1}^{n} |u_i|^{p(q-1)} = \sum_{i=1}^{n} |u_i|^q = \|u\|_q^q.$$

Now let $x = y / \|u\|_q^{q-1}$, which satisfies

$$\|x\|_p = \frac{\|u\|_q^{q/p}}{\|u\|_q^{q-1}} = \|u\|_q^{q/p-q+1} = 1, \quad \text{and} \quad u^T x = \frac{\|u\|_q^q}{\|u\|_q^{q-1}} = \|u\|_q,$$

so we are done. □

Corollary 1.8. *The* ℓ_2*-norm is self-dual.*

The following will show step-by-step, that the dual cone of a proper cone is a proper cone.

Proposition 1.9. K^{**} *is the closure of a convex cone* K. *(Hence* $K^{**} = K$ *if* K *is closed.)*

Proof. We know that a nonzero vector $y \in K^*$ if and only if y is the normal vector of a homogeneous halfspace containing K. Since the closure of K is the intersection of all homogeneous halfspaces containing K, we have that

$$\mathbf{cl}\, K = \bigcap_{y \in K^*} \left\{x \mid y^T x \geqslant 0\right\} = \left\{x \mid y^T x \geqslant 0 \, \forall y \in K^*\right\} = K^{**}.$$

□

Proposition 1.10. *The dual of a cone is closed and convex.*

Proof. The dual of a cone K is defined as

$$K^* = \left\{y \mid y^T x \geqslant 0 \, \forall x \in K\right\} = \bigcap_{x \in K} \left\{y \mid y^T x \geqslant 0\right\},$$

which is the intersection of homogeneous halfspaces, and therefore closed and convex. □

Proposition 1.11. *If* K *has nonempty interior, then* K^* *is pointed.*

Proof. Assume that K^* is not pointed, i.e. contains a line $\theta v, v \neq 0, \theta \in \mathbb{R}$. Then both $v^T x \geqslant 0$ and $(-v)^T x \geqslant 0$ hold for every $x \in K$, so $v^T x = 0$ holds for every $x \in K$. Since K has nonempty interior, there must be a ball $B = \{x_c + u \mid \|u\|_2 \leqslant r\}$ contained in K, so $v^T x = 0$ holds for every $x \in B$ too. It follows that v must be zero, so K^* is pointed. □

Proposition 1.12. *If* $\mathbf{cl}\, K$ *is pointed, then* K^* *has nonempty interior.*

Proof. Assume that $\mathbf{int}\, K^* = \varnothing$. Since K^* is closed and convex, it must be contained in an affine space of lower dimension. Then the normal vector $v \neq 0$ to that affine space will be contained in $K^{**} = \mathbf{cl}\, K$, and so is $-v$. Hence $\mathbf{cl}\, K$ contains a line, a contradiction. □

Corollary 1.13. *The dual cone of a proper cone is a proper cone.*

2 Strong convexity

Definition 2.1 (Strongly convex function)**.** *A function* f *is said to be* μ-strongly convex *if*

$$g(x) = f(x) - \frac{\mu}{2} \|x\|_2^2$$

is convex, where $\mu > 0$ *is constant.*

Intuitively, strong convexity means that there exists a quadratic lower bound on the growth of the function. It is natural that strong convexity implies convexity and strict convexity. Moreover, we have the following equivalent condition.

Proposition 2.2. *A function* f *is* μ*-strongly convex if and only if for every* $\theta \in [0, 1]$*, we have*

$$f(\theta x + (1-\theta)y) \leqslant \theta f(x) + (1-\theta)f(y) - \frac{\theta(1-\theta)\mu}{2} \|x-y\|_2^2$$

for any x, y*.*

Proof. It follows directly from the convexity of $g(x) = f(x) - \mu \|x\|_2^2 /2$, which gives that $g(\theta x + (1-\theta)y) \leqslant \theta f(x) + (1-\theta)f(y)$ holds for every $x, y \in \mathbf{dom}\, f$ and $\theta \in [0, 1]$. □

Suppose further that $f(x)$ is μ-strongly convex differentiable function, we will have the following equivalent conditions.

Proposition 2.3. *A function* f *is* μ*-strongly convex if and only if*

$$f(y) \geqslant f(x) + \nabla f(x)^T(y-x) + \frac{\mu}{2} \|y-x\|_2^2$$

for every $x, y \in \mathbf{dom}\, f$*.*

Proof. It follows directly from the first-order condition of the convexity of $g(x) = f(x) - \mu \|x\|_2^2 /2$, which gives that

$$g(y) \geqslant g(x) + \nabla g(x)^T(y-x).$$

(Note that $\nabla \|x\|_2^2 = 2x$.) □

Remark 2.4. *From the convexity of a differentiable convex function* $g(x)$*, we have that*

$$g(y) \geqslant g(x) + \nabla g(x)^T(y-x),$$

so

$$\begin{aligned}(\nabla g(y) - \nabla g(x))^T (y-x) &= \nabla g(y)^T(y-x) - \nabla g(x)^T(y-x) \\ &\geqslant \nabla g(y)^T(y-x) - g(y) + g(x) \\ &\geqslant 0.\end{aligned}$$

Proposition 2.5. *A function* f *is* μ*-strongly convex if and only if*

$$(\nabla f(x) - \nabla f(y))^T (x-y) \geqslant \mu \|x-y\|_2^2$$

for every $x, y \in \mathbf{dom}\, f$*.*

Proof. It follows directly from Remark 2.4, noting that $\nabla \|x\|_2^2 = 2x$. □

Proposition 2.6. *Suppose* $f(x)$ *is differentiable and* μ*-strongly convex. Then for every* $x, y \in \mathbf{dom}\, f$*,*

$$\|\nabla f(x) - \nabla f(y)\|_2 \geqslant \mu \|x-y\|_2 .$$

Proof. By Cauchy-Swartz inequality,

$$\|\nabla f(x) - \nabla f(y)\|_2 \|x-y\|_2 \geqslant (\nabla f(x) - \nabla f(y))^T (x-y) \geqslant \mu \|x-y\|_2^2 .$$

Cancelling $\|x-y\|_2$ on both sides finishes the proof. □

Remark 2.7 (Lipschitz). *A function* $f : U \to V$ *is* Lipschitz continuous *with* Lipschitz constant L *if*

$$d_V(f(x) - f(y)) \leqslant L d_U(x - y)$$

for every $x, y \in U$. *Here* $d_U(\cdot)$ *and* $d_V(\cdot)$ *are metrics on* U *and* V *respectively.*

If a differentiable function f is μ-strongly convex with its gradient $\nabla f(x)$ L-Lipschitz continuous (suppose we use the ℓ_2-norm), then we would see that

$$\mu \|x - y\| \leqslant \|\nabla f(x) - \nabla f(y)\| \leqslant L \|x - y\|$$

holds for every $x, y \in \mathbf{dom}\, f$. These concepts are essential in analyzing the rate of convergence of some algorithms.

3 Operations preserving convexity

Proposition 3.1. *The maximum eigenvalue of symmetric matrices* $\lambda_{\max}(X)$ *is a convex function.*

Proof. By Rayleigh's Theorem $\lambda_{\max}(X) = \max_{\|y\|_2=1} y^T X y$, which is the maximum of convex functions, thus convex. $\square$

Theorem 3.2 (Rayleigh). *Let* $A \in \mathbb{C}^{n\times n}$ *be Hermitian with eigenvalues* $\lambda_1 \leqslant \cdots \leqslant \lambda_n$ *and corresponding unit eigenvectors* $x_1, \cdots, x_n$. *Let* $i_1, \cdots, i_k$ *be given integers with* $1 \leqslant i_1 < \cdots < i_k \leqslant n$. *Let* $S = \mathrm{span}\{x_{i_1}, \cdots, x_{i_k}\}$. *Then*

$$\lambda_{i_1} = \min_{\substack{x\in S\\ \|x\|_2=1}} x^* A x, \quad \lambda_{i_k} = \max_{\substack{x\in S\\ \|x\|_2=1}} x^* A x,$$

with minimum and maximum achieved when $x = x_{i_1}$ *and* $x = x_{i_k}$*, respectively. (Here* x^* *is the conjugate transpose of* x*.)*

Proof. For any $x \in S$ with $\|x\|_2 = 1$, there exist scalars $\alpha_1, \cdots, \alpha_k$ such that $x = \sum_{j=1}^k \alpha_j x_{i_j}$. Since x is a unit norm vector,

$$x^* x = \sum_{j=1}^k \alpha_j^2 x_{i_j}^* x_{i_j} = \sum_{j=1}^k \alpha_j^2 = 1.$$

Moreover, $Ax = \sum_{j=1}^k \alpha_j \lambda_{i_j} x_{i_j}$, so

$$x^* A x = \left(\sum_{j=1}^k \alpha_j x_{i_j}^*\right)\left(\sum_{j=1}^k \alpha_j \lambda_{i_j} x_{i_j}\right) = \sum_{j=1}^k \alpha_j^2 \lambda_{i_j},$$

which is a convex combination of $\lambda_{i_1}, \cdots, \lambda_{i_k}$, so it lies between λ_{i_1} and λ_{i_k}. When $x = x_{i_1}$ the minimum is achieved, and when $x = x_{i_k}$ the maximum is achieved. $\square$

With regard to eigenvalues there are some other interesting facts. The following can be viewed as generalized versions of the Rayleigh's theorem.

Theorem 3.3 (Courant-Fischer). *Let $A \in \mathbb{C}^{n\times n}$ be Hermitian with eigenvalues $\lambda_1 \leqslant \cdots \leqslant \lambda_n$. Then we have*

$$\lambda_i = \min_{\dim V = i} \max_{\substack{x\in V \\ \|x\|_2=1}} x^*Ax,$$

and its dual form

$$\lambda_i = \max_{\dim V = n-i+1} \min_{\substack{x\in V \\ \|x\|_2=1}} x^*Ax.$$

Proof. Let $x_1, \cdots, x_n$ be the eigenvectors associated with $\lambda_1, \cdots, \lambda_n$. For a given $i \in [n]$, let $S = \mathrm{span}\{x_i, \cdots, x_n\}$. For any subspace V with $\dim V = i$, since $\dim V + \dim S = n + 1 > n$, it follows that $V \cap S \neq 0$. Then according to the Rayleigh's Theorem we have

$$\max_{\substack{x\in V \\ \|x\|_2=1}} x^*Ax \geqslant \max_{\substack{x\in V\cap S \\ \|x\|_2=1}} x^*Ax \geqslant \min_{\substack{x\in V\cap S \\ \|x\|_2=1}} x^*Ax \geqslant \min_{\substack{x\in S \\ \|x\|_2=1}} x^*Ax = \lambda_i.$$

The equality is achieved when $V = \mathrm{span}\{x_1, \cdots, x_i\}$, so we have

$$\lambda_i = \min_{\dim V = i} \max_{\substack{x\in V \\ \|x\|_2=1}} x^*Ax.$$

The dual form can be proved by applying the primal form to $-A$, noting that the ordered eigenvalues are

$$\lambda_i(-A) = -\lambda_{n-i+1}(A).$$

□

Theorem 3.4. *Let $A \in \mathbb{C}^{n\times n}$ be Hermitian with eigenvalues $\lambda_1 \leqslant \cdots \leqslant \lambda_n$. Suppose that $1 \leqslant m \leqslant n$. Then*

$$\sum_{i=1}^{m} \lambda_i = \min_{\substack{V\in\mathbb{C}^{n\times m} \\ V^*V=I_m}} \mathbf{Tr}(V^*AV),$$

and

$$\sum_{i=1}^{m} \lambda_{i+n-m} = \max_{\substack{V\in\mathbb{C}^{n\times m} \\ V^*V=I_m}} \mathbf{Tr}(V^*AV).$$

The minimum and maximum are achieved for a matrix V whose columns are orthonormal eigenvectors associated with the m smallest or largest eigenvalues of A.

Moreover, it is not surprising that similar things can apply to singular values of arbitrary matrices. The following theorem could be derived from Courant-Fischer.

Theorem 3.5. *Let $A \in \mathbb{C}^{n\times m}$ with singular values $\sigma_1 \geqslant \cdots \geqslant \sigma_q$, where $q = \min\{n, m\}$. Then for every $i \in [q]$ we have*

$$\sigma_i = \max_{\dim V = i} \min_{0\neq x\in V} \frac{\|Ax\|_2}{\|x\|_2},$$

and

$$\sigma_i = \min_{\dim V = m-i+1} \max_{0\neq x\in V} \frac{\|Ax\|_2}{\|x\|_2}.$$

Theorem 3.6. *If* $f(x, y)$ *is convex and* C *is a convex set, then* $g(x) = \inf_{y \in C} f(x, y)$ *is convex. Here we take* $\mathbf{dom}\, g = \{x \mid (x, y) \in \mathbf{dom}\, f$ for some $y \in C\}$.

Proof. Take any $x_1, x_2 \in \mathbf{dom}\, g$. For any $\varepsilon > 0$, there exists $y_1, y_2 \in C$ such that $f(x_1, y_1) - \varepsilon \leqslant g(x_1), f(x_2, y_2) - \varepsilon \leqslant g(x_2)$. For any $\theta \in [0, 1]$, we have that

$$\begin{aligned} g(\theta x_1 + (1-\theta)x_2) &= \inf_{y \in C} f(\theta x_1 + (1-\theta)x_2, y) \\ &\leqslant f(\theta x_1 + (1-\theta)x_2, \theta y_1 + (1-\theta)y_2) \\ &\leqslant \theta f(x_1, y_1) + (1-\theta) f(x_2, y_2) \\ &\leqslant \theta g(x_1) + (1-\theta) g(x_2) + \varepsilon. \end{aligned}$$

Since the above inequality holds for every $\varepsilon > 0$, we can conclude that

$$g(\theta x_1 + (1-\theta)x_2) \leqslant \theta g(x_1) + (1-\theta) g(x_2),$$

so g is convex. □

Definition 3.7 (Perspective)**.** *The* perspective *of a function* $f : \mathbb{R}^n \to \mathbb{R}$ *is defined as*

$$g(x, t) = t f(x/t), \quad \mathbf{dom}\, g = \{(x, t) \mid x/t \in \mathbf{dom}\, f, t > 0\}.$$

Proposition 3.8. *The perspective of a convex function is convex.*

Proof. There are several ways to prove it. First, we prove it by showing the Jensen's inequality for every $(x, t), (y, s) \in \mathbf{dom}\, g, \theta \in [0, 1]$. We have that

$$\begin{aligned} & g(\theta x + (1-\theta)y, \theta t + (1-\theta)s) \\ &= (\theta t + (1-\theta)s)\, f\left(\frac{\theta x + (1-\theta)y}{\theta t + (1-\theta)s}\right) \\ &= (\theta t + (1-\theta)s)\, f\left(\frac{\theta t}{\theta t + (1-\theta)s} \cdot \frac{x}{t} + \frac{(1-\theta)s}{\theta t + (1-\theta)s} \cdot \frac{y}{s}\right) \\ &\leqslant (\theta t + (1-\theta)s)\left(\frac{\theta t}{\theta t + (1-\theta)s} f\left(\frac{x}{t}\right) + \frac{(1-\theta)s}{\theta t + (1-\theta)s} f\left(\frac{y}{s}\right)\right) \\ &= \theta g(x, t) + (1-\theta) g(y, s). \end{aligned}$$

Moreover, there is another important way to prove the convexity of g if we look at the epigraph of g. Note that for $t > 0$, $(x, t, s) \in \mathbf{epi}\, g$ if and only if $t f(x/t) \leqslant s$, which is equivalent to $(x, s)/t \in \mathbf{epi}\, f$. Therefore, $\mathbf{epi}\, g$ is the inverse image of $\mathbf{epi}\, f$ under the perspective function

$$P(u, v, w) = (u, v)/w.$$

So $\mathbf{epi}\, g$ is convex, and so is g. □

4 Conjugate function

Definition 4.1 (Convex conjugate). *The* convex conjugate *of a function* f *is*

$$f^*(y) = \sup_{x \in \mathbf{dom}\, f} \left(y^T x - f(x) \right).$$

It is also called the **Fenchel conjugate**, **Fenchel transformation**, or the **Legendre-Fenchel transformation**.

Remark 4.2. *The conjugate function of any function is convex.*

Proof. f^* is the supremum of affine functions, hence convex. □

Theorem 4.3 (Fenchel's inequality). *For a function* f *and its convex conjugate* f^**, we have that*

$$y^T x \leqslant f(x) + f^*(y).$$

Proof. It follows immediately from the definition that

$$f^*(y) \geqslant y^T x - f(x), \quad \forall x, y \in \mathbf{dom}\, f.$$

□

We are curious about the convex conjugate of the convex conjugate of a function. Below are some known facts, the proof of which is omitted here.

Definition 4.4 (Biconjugate). f^{**} *is called the* biconjugate *of a function* f.

Definition 4.5 (Lower semicontinuity). *A function* $f : X \to \bar{\mathbb{R}}$ *is called* lower semicontinuous *at a point* $x_0 \in X$ *if for every real* $y < f(x_0)$*, there exists a neighborhood* U *of* x_0 *such that* $f(x) > y$ *for all* $x \in U$.

Remark 4.6. f *is lower semicontinuous if and only if all sublevel sets are closed.*

Remark 4.7. f *is lower semicontinuous if and only if the epigraph of* f *is closed.*

Similarly we can define the concept of **upper semicontinuity**. A function is continuous if and only if it is both lower and upper semicontinuous.

Proposition 4.8. *The biconjugate* f^{**} *is the largest lower semi-continuous convex function with* $f^{**} \leqslant f$.

Theorem 4.9 (Fenchel-Moreau). $f = f^{**}$ *if and only if* f *is a lower semi-continuous and convex function.*

There is a very interesting fact about a function and its convex conjugate. Refer to [2] for a proof of this theorem.

Theorem 4.10.

1. *If* f *is closed and* μ*-strongly convex, then* f^* *has a* $1/\mu$*-Lipschitz continuous gradient.*

2. *If* f *is convex and has an* L*-Lipschitz continuous gradient, then* f^* *is* $1/L$*-strongly convex.*

5 Convex optimization

Theorem 5.1. *Any locally optimal point of a convex optimization problem is globally optimal.*

Proof. Suppose x is locally optimal, i.e. there exists $R > 0$ such that $f_0(y) \geqslant f_0(x)$ holds for every feasible y with $\|x - y\| \leqslant R$. Assume that there exists $x' \neq x$ such that $f_0(x') < f_0(x)$. Pick z on the line segment between x and x', with $0 < \|z - x\| < R$, so that z can be expressed in the form $z = \theta x + (1-\theta)x'$ for some $\theta \in [0, 1]$. Then by the convexity of f_0 we have

$$f_0(z) \leqslant \theta f_0(x) + (1-\theta) f_0(x') < \theta f_0(x) + (1-\theta) f_0(x) = f_0(x),$$

contradictory with the fact that $f_0(z) \geqslant f_0(x)$. □

References

[1] R. A. Horn and C. R. Johnson, *Matrix analysis*. Cambridge university press, 2012.

[2] X. Zhou, "On the fenchel duality between strong convexity and lipschitz continuous gradient," *arXiv preprint arXiv:1803.06573*, 2018.