Conic Programming

Recall the linear programming (LP) problem that we have seen in the last lecture:

$$\min_{\mathbf{x}\in\mathbb{R}^n} \mathbf{c}^{\mathrm{T}}\mathbf{x} \quad \text{s.t.} \quad \mathbf{A}\mathbf{x}-\mathbf{b}\geq \mathbf{0}.$$

with $\mathbf{c}\in\mathbb{R}^n$, $\mathbf{A}\in\mathbb{R}^{m\times n}$ and $\mathbf{b}\in\mathbb{R}^m$. The constraints of this problem (interpreted as an element-wise inequality) can be written equivalently as

$$\mathbf{A}\mathbf{x}-\mathbf{b}\in\mathbb{R}^m_+ = \{\mathbf{y}\in\mathbb{R}^m : y_i \geq 0, \forall i\}.$$

The space $\mathbb{R}^m_+$ is usually called the *positive* (or, more accurately, non-negative) *orthant*; the constraints of LP can be interpreted as the intersection of the affine space $\mathbf{A}\mathbf{x}-\mathbf{b}$ with the positive orthant.

1 Convex cones

Geometrically, the orthant is a particular case of a construction called a convex cone:

Definition. *A set C is a* convex cone *iff $\alpha C+\beta C = C$ for any $\alpha,\beta>0$. The cone is said* pointed *if $\mathbf{0}\in C$ and* blunt otherwise*. The cone is said* salient *if for every $\mathbf{0}\neq\mathbf{x}\in C$, $-\mathbf{x}\notin C$, and* flat *otherwise.*

The way we defined $\mathbb{R}^m_+$ makes it a pointed salient convex cone. Other examples of cones we will see next fall into this category:

1. *Lorentz* (aka "ice cream" or "second-order") cone defined as

$$L^m = \{\mathbf{x}\in\mathbb{R}^m : x_m^2 \geq x_1^2+\cdots+x_{m-1}^2\} = \{\mathbf{x} : x_m^2 \geq \|(x_1,\dots,x_{m-1})\|_2\}.$$

2. *Positive semidefinite cone*

$$S^m_+ = \{\mathbf{A}\in\mathbb{R}^{m\times m} : \mathbf{A}\succeq 0\}.$$

Exercise 1. *Show that $\mathbb{R}^m_+$, L^m, and S^m_+ are pointed salient convex cones.*

The notion of a cone gives rise to a generalized form of *conic inequalities*: for a cone K, we will denote $\mathbf{a}\in K$ equivalently as $\mathbf{a}\geq_K \mathbf{0}$. Similarly, we will say $\mathbf{a}\geq_K\mathbf{b}$ implying $\mathbf{a}-\mathbf{b}\geq_K\mathbf{0}$ or, equivalently, $\mathbf{a}-\mathbf{b}\in K$. In the same way, $\mathbf{A}\mathbf{x}-\mathbf{b}\geq_K\mathbf{0}$ defines the intersection of an affine space with the cone K.

2 Conic programs

The notion of a conic inequality allows to generalize a linear program into a *conic program* (CP)

$$\min_{\mathbf{x}\in\mathbb{R}^n} \mathbf{c}^{\mathrm{T}}\mathbf{x} \quad \text{s.t.} \quad \mathbf{A}\mathbf{x}-\mathbf{b} \geq_K \mathbf{0}$$

by simply replacing the standard inequality (i.e., $\geq_{\mathbb{R}^m_+}$) with a conic inequality w.r.t. some convex cone K. Apart from linear programming, which is an obvious particular case of CP, the following two problems are of major importance:

1. *Second-order conic program* (SOCP) corresponding to $K = L^m$:

   $$\min_{\mathbf{x}\in\mathbb{R}^n} \mathbf{c}^{\mathrm{T}}\mathbf{x} \quad \text{s.t.} \quad \begin{pmatrix} \mathbf{A}\mathbf{x}-\mathbf{b} \\ \mathbf{d}^{\mathrm{T}}\mathbf{x}-\mathbf{e} \end{pmatrix} \geq_K \mathbf{0}$$

   which can be rewritten explicitly as

   $$\min_{\mathbf{x}\in\mathbb{R}^n} \mathbf{c}^{\mathrm{T}}\mathbf{x} \quad \text{s.t.} \quad \|\mathbf{A}\mathbf{x}-\mathbf{b}\|_2 \leq \mathbf{d}^{\mathrm{T}}\mathbf{x}-\mathbf{e}.$$

2. *Semidefinite program* (SDP) corresponding to $K = S^m_+$. In order to avoid high-order tensor notation, we will define a general linear operator $\mathcal{A} : \mathbb{R}^n \mapsto \mathbb{R}^{m\times m}$ as

   $$\mathcal{A}\mathbf{x} = \sum_{i=1}^{n} x_i \mathbf{A}_i,$$

   where $\mathbf{A}_1, \ldots, \mathbf{A}_n \in S^m_+$. In these terms, an SDP can be written as

   $$\min_{\mathbf{x}\in\mathbb{R}^n} \mathbf{c}^{\mathrm{T}}\mathbf{x} \quad \text{s.t.} \quad \mathcal{A}\mathbf{x} - \mathbf{B} \succeq 0.$$

Surprisingly, SDP generalizes both LP and SOCP! Any linear program can be written in the form

$$\min_{\mathbf{x}\in\mathbb{R}^n} \mathbf{c}^{\mathrm{T}}\mathbf{x} \quad \text{s.t.} \quad \mathbf{a}_i^{\mathrm{T}}\mathbf{x} - b_i \geq 0, i = 1, \ldots, m,$$

which can be written as an equivalent SDP

$$\min_{\mathbf{x}\in\mathbb{R}^n} \mathbf{c}^{\mathrm{T}}\mathbf{x} \quad \text{s.t.} \quad \begin{pmatrix} \mathbf{a}_1^{\mathrm{T}}\mathbf{x} - b_1 & & \\ & \ddots & \\ & & \mathbf{a}_m^{\mathrm{T}}\mathbf{x} - b_m \end{pmatrix} \succeq 0.$$

Note that a diagonal matrix is positive semi-definite iff all its diagonal elements are non-negative.

In order to reduce SOCP to SDP, we will need the following very useful result:

Theorem 1 (Schur complement)**.** *Let* $\mathbf{A}$ *be a symmetric block matrix*

$$\mathbf{A} = \begin{pmatrix} \mathbf{P} & \mathbf{Q}^{\mathrm{T}} \\ \mathbf{Q} & \mathbf{R} \end{pmatrix}$$

with $\mathbf{R} \succ 0$*. Then,* $\mathbf{A} \succ 0$ *iff* $\mathbf{P} - \mathbf{Q}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{Q} \succ 0$*.*

The theorem holds for a semi-definite $\mathbf{A}$ as well. To prove this result, observe that $\mathbf{A} \succeq 0$ iff for every $\mathbf{u}$,

$$\inf_{\mathbf{v}} (\mathbf{u}^{\mathrm{T}}, \mathbf{v}^{\mathrm{T}}) \begin{pmatrix} \mathbf{P} & \mathbf{Q}^{\mathrm{T}} \\ \mathbf{Q} & \mathbf{R} \end{pmatrix} \begin{pmatrix} \mathbf{u} \\ \mathbf{v} \end{pmatrix} = \inf_{\mathbf{v}} \mathbf{u}^{\mathrm{T}}\mathbf{P}\mathbf{u} + 2\mathbf{u}^{\mathrm{T}}\mathbf{Q}^{\mathrm{T}}\mathbf{v} + \mathbf{v}^{\mathrm{T}}\mathbf{R}\mathbf{v}.$$

The latter is an infimum of a quadratic function in $\mathbf{v}$, for which we can write the following necessary optimality condition (which is also sufficient, since $\mathbf{R} \succ 0$ and the function is convex):

$$2\mathbf{Q}\mathbf{u} + 2\mathbf{R}\mathbf{v} = 0,$$

from where

$$\mathbf{v} = -\mathbf{R}^{-1}\mathbf{Q}\mathbf{u}.$$

Substituting the solution into the quadratic function yields

$$\begin{aligned} \inf_{\mathbf{v}} (\mathbf{u}^{\mathrm{T}}\mathbf{v}^{\mathrm{T}})\mathbf{A} \begin{pmatrix} \mathbf{u} \\ \mathbf{v} \end{pmatrix} &= \mathbf{u}^{\mathrm{T}}\mathbf{P}\mathbf{u} - 2\mathbf{u}^{\mathrm{T}}\mathbf{Q}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{Q}\mathbf{u} + \mathbf{u}^{\mathrm{T}}\mathbf{Q}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{R}\mathbf{R}^{-1}\mathbf{Q}\mathbf{u} \\ &= \mathbf{u}^{\mathrm{T}}(\mathbf{P} - \mathbf{Q}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{Q})\mathbf{u} \end{aligned}$$

The latter quadratic form is non-negative iff $\mathbf{P} - \mathbf{Q}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{Q} \succeq 0$.

Armed with Schur's complement theorem, we can now return to the SOCP problem, which we will now write as

$$\min_{\mathbf{x} \in \mathbb{R}^n} \mathbf{c}^{\mathrm{T}}\mathbf{x} \quad \text{s.t.} \quad \|\mathbf{u}(\mathbf{x})\|_2 \le v(\mathbf{x}),$$

where $\mathbf{u}(\mathbf{x}) = \mathbf{A}\mathbf{x} - \mathbf{b}$ and $v(\mathbf{x}) = \mathbf{d}^{\mathrm{T}}\mathbf{x} - \mathbf{e}$. Note that v in the constraint upper-bounds a norm, so it cannot be negative. We rewrite the constraint as $\mathbf{u}^{\mathrm{T}}\mathbf{u} \le v^2$ or, dividing by v as

$$v - \frac{1}{v}\mathbf{u}^{\mathrm{T}}\mathbf{u} = v - \mathbf{u}^{\mathrm{T}} \begin{pmatrix} v & & \\ & \ddots & \\ & & v \end{pmatrix}^{-1} \mathbf{u} \ge 0.$$

Invoking the Schur's complement, we can equivalently write the constraint as

$$\begin{pmatrix} v(\mathbf{x}) & - & \mathbf{u}^{\mathrm{T}}(\mathbf{x}) & - \\ | & v(\mathbf{x}) & & \\ \mathbf{u}(\mathbf{x}) & & \ddots & \\ | & & & v(\mathbf{x}) \end{pmatrix} \succeq 0,$$

turning SOCP into an SDP.

3 Barrier (interior point) methods

One of the most successful ways of solving conic programs are barrier methods, also know as interior point methods. Recall that we defined barriers as a particular family of penalty functions that disallow infeasible solutions. For our purpose, a barrier representing a cone K is a function $\varphi : K \to \mathbb{R}$ such that

$$\lim_{K \ni \mathbf{x} \to \partial K} \varphi(\mathbf{x}) = \infty.$$

A barrier method consists of solving a sequence of unconstrained minimization problems with the barrier aggregate

$$\min_{\mathbf{x} \in \mathbb{R}^n} F_p(\mathbf{x}) = \min_{\mathbf{x} \in \mathbb{R}^n} \mathbf{c}^{\mathrm{T}} \mathbf{x} + \frac{1}{p} \varphi(\mathcal{A}\mathbf{x} - \mathbf{b}),$$

with increasing p. In the limit $p \to \infty$, unconstrained minimization of F_∞ is equivalent to the original conic program. As we have already mentioned, the barrier method has to be initialized with a feasible point, and it will always produce a strictly feasible solution (this is the reason for the name *interior point*).

The following barrier functions are frequently used for standard convex cones:

Cone	Barrier function
$\mathbb{R}_+^m$	$\varphi(\mathbf{x}) = -\sum_{i=1}^{m} \log x_i$
L^m	$\varphi(\mathbf{x}) = -\log \left(x_m^2 - \sum_{i=1}^{m-1} x_i^2 \right)$
S_+^m	$\varphi(\mathbf{X}) = -\log \det \mathbf{X}$

Exercise 2. *Derive the gradients of the above functions and show that they qualify as barriers.*
Hint: for the semidefinite cone barrier, show that $-\log \det \mathbf{X}$ *can be represented as* $-\mathrm{tr} \log \mathbf{X}$*, where* $\log \mathbf{X}$ *is understood as a function of a matrix.*

4 Dual cones

Definition. *The dual of a cone K is* $K_* = \{\mathbf{y} : \langle \mathbf{x}, \mathbf{y} \rangle \geq 0 \quad \forall \mathbf{x} \in K\}$

Exercise 3. *Prove that K_* is a cone.*

Geometrically, the dual cone consists of vectors forming acute angles with all the vectors of the cone K.

Exercise 4. *Prove that* $(K_*)_* = K$.

Definition. *A cone K is said to be* self-dual *iff* $K_* = K$.

Exercise 5. *Show that* $\mathbb{R}_+^m$, L^m *and* S_+^m *are self-dual.*

5 Dual conic programs

Before we show the dual form of a general convex conic program, we need another important notion from linear algebra.

Definition. *An* adjoint *of a linear operator $\mathcal{A} : U \to V$ is the linear operator $\mathcal{A}^* : V \to U$ such that for every $\mathbf{u} \in U$ and $\mathbf{v} \in V$,*

$$\langle \mathbf{u}, \mathcal{A}^* \mathbf{v} \rangle_U = \langle \mathcal{A}\mathbf{u}, \mathbf{v} \rangle_V.$$

In general, there is no relation between the adjoint operator and the inverse operator!

Exercise 6. *Show that the adjoint of a linear operator expressed by a real matrix $\mathbf{A}$ is the matrix $\mathbf{A}^{\mathrm{T}}$.*

Let us now consider the (primal) conic program

$$\min_{\mathbf{x} \in \mathbb{R}^n} \mathbf{c}^{\mathrm{T}} \mathbf{x} \quad \text{s.t.} \quad \mathcal{A}\mathbf{x} - \mathbf{b} \geq_K \mathbf{0}.$$

For every feasible $\mathbf{x}$, $\mathcal{A}\mathbf{x} - \mathbf{b} \in K$. Therefore, by definition of the dual cone, for every $\mathbf{y} \in K_*$, $\langle \mathcal{A}\mathbf{x} - \mathbf{b}, \mathbf{y} \rangle \geq 0$, from where

$$\langle \mathcal{A}\mathbf{x}, \mathbf{y} \rangle = \langle \mathbf{x}, \mathcal{A}^* \mathbf{y} \rangle \geq \langle \mathbf{b}, \mathbf{y} \rangle$$

In particular, for $\mathcal{A}^* \mathbf{y} = \mathbf{c}$, we have

$$\langle \mathbf{c}, \mathbf{x} \rangle \geq \langle \mathbf{b}, \mathbf{y} \rangle.$$

In other words, $\langle \mathbf{b}, \mathbf{y} \rangle$ is the lower bound on the primal objective. Maximizing the latter bound yields the dual problem:

$$\max_{\mathbf{y} \in K_*} \mathbf{b}^{\mathrm{T}} \mathbf{y} \quad \text{s.t.} \quad \mathcal{A}^* \mathbf{y} = \mathbf{c}.$$

Most practical conic program satisfy the conditions of the following strong duality theorem:

Theorem 2 (Strong duality)**.** *If either the primal or the dual problems are strictly feasible and bounded, then the other one is solvable and $\langle \mathbf{c}, \mathbf{x}^* \rangle = \langle \mathbf{b}, \mathbf{y}^* \rangle$.*

Let $\mathbf{x}^*$ and $\mathbf{y}^*$ the primal and the dual solutions, respectively. Substituting $\mathcal{A}^* \mathbf{y} = \mathbf{c}$ into the strong duality result yields

$$\begin{aligned} 0 &= \langle \mathbf{c}, \mathbf{x}^* \rangle - \langle \mathbf{b}, \mathbf{y}^* \rangle = \langle \mathcal{A}^* \mathbf{y}^*, \mathbf{x}^* \rangle - \langle \mathbf{b}, \mathbf{y}^* \rangle \\ &= \langle \mathbf{y}^*, \mathcal{A}\mathbf{x}^* \rangle - \langle \mathbf{y}^*, \mathbf{b} \rangle = \langle \mathbf{y}^*, \mathcal{A}\mathbf{x}^* - \mathbf{b} \rangle. \end{aligned}$$

Note that we have already seen this condition in the general nonlinear programming case, $\boldsymbol{\lambda}^{*\mathrm{T}} \mathbf{g}(\mathbf{x}^*) = 0$, under the name of complementary slackness. As before, the dual variable $\mathbf{y}$ plays the role of the Lagrange multipliers.

As an example, let us derive the dual problem of SDP, in which

$$\mathcal{A}\mathbf{x} = \sum_{i=1}^{m} x_i \mathbf{A}_i.$$

Since S_+^m is self-dual, the dual problem can be written as

$$\max_{\mathbf{Y} \succeq 0} \langle \mathbf{Y}, \mathbf{B} \rangle \quad \text{s.t.} \quad \mathcal{A}^* \mathbf{Y} = \mathbf{c}.$$

Here, the inner product in the objective is a matrix inner product $\langle \mathbf{Y}, \mathbf{B} \rangle = \mathrm{tr}(\mathbf{YB})$ (no transpose due to symmetry). The adjoint operator $\mathcal{A}^*$ assigns to each matrix $\mathbf{Y} \in \mathbb{R}^{m \times m}$ a vector in $\mathbb{R}^n$. In order to derive the adjoint, let fix some $\mathbf{x}$ and $\mathbf{Y}$ and use the definition

$$\begin{aligned} \langle \mathbf{x}, \mathcal{A}^* \mathbf{Y} \rangle &= \langle \mathcal{A}\mathbf{x}, \mathbf{Y} \rangle = \langle \sum_{i=1}^{m} x_i \mathbf{A}_i, \mathbf{Y} \rangle \\ &= \sum_{i=1}^{m} x_i \langle \mathbf{A}_i, \mathbf{Y} \rangle = \langle \mathbf{x}, \begin{pmatrix} \langle \mathbf{A}_1, \mathbf{Y} \rangle \\ \vdots \\ \langle \mathbf{A}_n, \mathbf{Y} \rangle \end{pmatrix} \rangle \end{aligned}$$

Since the latter equality holds for every $\mathbf{x}$ and $\mathbf{Y}$, we recognize in the vector of inner products $\langle \mathbf{A}_i, \mathbf{Y} \rangle$ the action of the adjoint operator on $\mathbf{Y}$.