Any typos should be emailed to a_a_a@princeton.edu.

In this lecture, we see some of the most well-known classes of convex optimization problems and some of their applications. These include:

- Linear Programming (LP)
- (Convex) Quadratic Programming (QP)
- (Convex) Quadratically Constrained Quadratic Programming (QCQP)
- Second Order Cone Programming (SOCP)
- Semidefinite Programming (SDP)

1 Linear Programming

Definition 1. *A linear program (LP) is the problem of optimizing a linear function over a polyhedron:*

$$\begin{aligned} &\min c^T x \\ &\text{s.t. } a_i^T x \leq b_i, \ i = 1, \ldots, m, \end{aligned}$$

or written more compactly as

$$\begin{aligned} &\min c^T x \\ &\text{s.t. } Ax \leq b, \end{aligned}$$

for some $A \in \mathbb{R}^{m \times n}, b \in \mathbb{R}^m$.

We'll be very brief on our discussion of LPs since this is the central topic of ORF 522. It suffices to say that LPs probably still take the top spot in terms of ubiquity of applications. Here are a few examples:

- A variety of problems in production planning and scheduling

- Exact formulation of several important combinatorial optimization problems (e.g., min-cut, shortest path, bipartite matching)
- Relaxations for all 0/1 combinatorial programs
- Subroutines of branch-and-bound algorithms for integer programming
- Relaxations for cardinality constrained (compressed sensing type) optimization problems
- Computing Nash equilibria in zero-sum games
- ...

2 Quadratic Programming

Definition 2. *A quadratic program (QP) is an optimization problem with a quadratic objective and linear constraints*

$$\min_x x^T Q x + q^T x + c$$
$$s.t.\ \ Ax \leq b.$$

Here, we have $Q \in S^{n\times n}$, $q \in \mathbb{R}^n,\ c \in \mathbb{R},\ A \in \mathbb{R}^{m\times n},\ b \in \mathbb{R}^m$.

The difficulty of this problem changes drastically depending on whether Q is positive semidefinite (psd) or not. When Q is psd, we call this *convex quadratic programming*, although under some conventions quadratic programming refers to the convex case by definition (and the nonconvex case would be called nonconvex QP).
In our previous lecture, we already saw a popular application of QP in *maximum-margin support vector machines.* Here we see another famous application in the field of finance, which won its developer the Nobel Prize in economics. Let's not forget that the basic least sqaures problem is also another instance of QP, possibly the simplest one.

2.1 The Markowitz minimum variance portfolio

We would like to invest our money in n assets over a fixed period. The return r_i of each asset is a random variable; we only assume to know its first and second order moments. Denote this random return by

$$r_i = \frac{P_{i,end} - P_{i,beg}}{P_{i,beg}}$$

where $P_{i,beg}$ and $P_{i,end}$ are the prices of the asset at the beginning and end of the period. Let $r \in \mathbb{R}^n$ be the random vector of all returns, which we assume has known mean $\mu \in \mathbb{R}^n$ and covariance $\Sigma \in S^{n \times n}$. If we decide to invest a portion x_i of our money in asset i, then the expected return of our portfolio would be

$$E[x^T r] = x^T \mu,$$

and its variance

$$\begin{aligned} E[(x^T r - x^T \mu)^2] &= E[(x^T (r - \mu))^2] = E[x^T (r - \mu)(r - \mu)^T x] \\ &= x^T E[(r - \mu)(r - \mu)^T] x \\ &= x^T \Sigma x. \end{aligned}$$

In practice, μ and Σ can be estimated from past data and be replaced with their empirical versions.

The minimum variance portfolio optimization problem seeks to find a portfolio that meets a given desired level of return $r_{\min}$, and has the lowest variance (or risk) possible:

$$\begin{aligned} &\min_x \; x^T \Sigma x \\ &\text{s.t. } \; x^T \mu \geq r_{\min} \\ &x \geq 0, \; \sum x_i = 1. \end{aligned}$$

In some variants of the problem the constraint $x_i \geq 0$ is removed on some of the variables ("*shorting*" is allowed). In either case, this problem is a convex QP (why?).

3 Quadratically Constrained Quadratic Programming

Definition 3. *A* quadratically constrained quadratic program (QCQP) *is an optimization problem with a quadratic objective and quadratic constraints:*

$$\begin{aligned} &\min_x \; x^T Q x + q^T x + c \\ &s.t. \; x^T Q_i x + q_i^T x + c_i \leq 0, \; i = 1, \ldots, m. \end{aligned}$$

Here, we have $Q_i, Q \in S^{n \times n}$, $q, q_i \in \mathbb{R}^n$, $c, c_i \in \mathbb{R}^n$.

Just like QP, the difficulty of the problem changes drastically depending on whether the matrices Q_i and Q are psd or not. In the case where $Q, Q_1, \ldots, Q_m$ are all psd, we refer to this problem as convex QCQP.

Notice that $QP \subseteq QCQP$ (take $Q_i = 0$).

A variant of the Markowitz portfolio problem described above gives a simple example of a QCQP.

3.1 A variant of the Markowitz portfolio theory problem

Once again, we would like to invest our money in n assets over a fixed period, with $r \in \mathbb{R}^n$ denoting the random vector of all returns, with mean $\mu \in \mathbb{R}^n$ and covariance matrix $\Sigma \in S^{n\times n}$. In our previous example, we wanted to find the minimum risk (or minimum variance) portfolio at a given level of return $r_{\min}$. It can also be interesting to consider the problem of finding the maximum return portfolio that meets a given level of risk $\sigma_{\max}$:

$$
\begin{aligned}
&\max_x \; x^T\mu \\
&\text{s.t. } x^T\Sigma x \leq \sigma_{\max} \\
&\sum x_i = 1 \\
&x \geq 0.
\end{aligned}
$$

This is a convex QCQP.

4 Second Order Cone Programming

Definition 4. *A* second order cone program (SOCP) *is an optimization problem of the form:*

$$
\begin{aligned}
&\min_x f^T x \\
&||A_i x + b_i||_2 \leq c_i^T x + d_i, \; i = 1, \ldots, m,
\end{aligned}
\tag{1}
$$

where $A_i \in \mathbb{R}^{k_i \times n}$, $b_i \in \mathbb{R}^{k_i}$, $c_i \in \mathbb{R}^n$ *and* $d_i \in \mathbb{R}$.

The terminology of "SOCP" comes from its connection to *the second order cone* (also called the Lorentz cone or the ice-cream cone).

Definition 5 (Second order cone)**.** *The second order cone in dimension $n+1$ is the set*

$$\mathcal{L}^{n+1} = \{(x,t) | \ ||x||_2 \leq t\}.$$

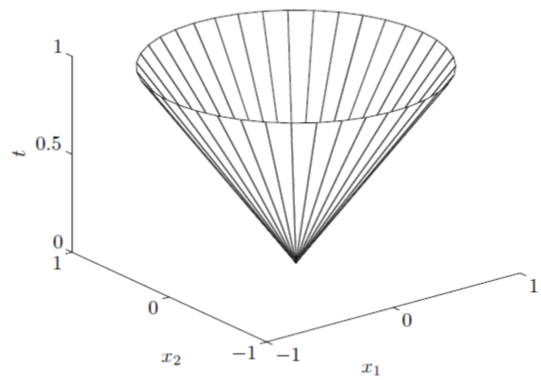

Figure 1: Boundary of the second order cone in $\mathbb{R}^3$.
Image credit: [1]

Notice then that (1) is equivalent to

$$\begin{aligned} &\min_x f^T x \\ &(A_i x + b_i, c_i^T x + d_i) \in \mathcal{L}^{n+1}, \ i = 1, \ldots, m. \end{aligned}$$

- If we take $A_i = 0$, we recover LPs.
- We also have (convex) QCQP $\subseteq$ SOCP (can you prove this?).

4.1 LASSO with Block Sparsity [2]

As an application of SOCP, we study a variant of the LASSO problem we saw earlier on. Consider the problem

$$\min_\alpha ||A\alpha - y||_2,$$

where $\alpha = \begin{pmatrix} \alpha_1 & \ldots & \alpha_p \end{pmatrix}^T \in \mathbb{R}^n$, $\alpha_i \in \mathbb{R}^{n_i}$, and $\sum_i n_i = n$.

- Similar to LASSO, we would like to obtain sparse solutions. However, in this new problem, we want to take into consideration the location of the zeros. To be more precise, we would like to set as many blocks α_i to zero as we can. If there is one or more nonzero element in a given block, then it does not matter to us how many elements in that block are nonzero.

- Naturally, the $||.||_1$ penalty of LASSO will not do the right thing here as it attempts to return a sparse solution without taking into consideration the block structure of our problem.

- Instead, we propose the penalty function

$$\left|\left|\begin{pmatrix} ||\alpha_1||_2 \\ \vdots \\ ||\alpha_p||_2 \end{pmatrix}\right|\right|_1 = \sum_{i=1}^{p} ||\alpha_i||_2.$$

  This will set many of the terms $||\alpha_i||_2$ to zero, which will force all elements of that particular block to be set to zero.

- The overall problem then becomes

$$\min_{\alpha} ||A\alpha - y||_2 + \gamma \sum_{i=1}^{p} ||\alpha_i||_2$$

  where $\gamma > 0$ is a given constant.

- The problem can be rewritten in SOCP form:

$$\begin{aligned} &\min_{\alpha,z,t_i} z + \gamma \sum_{i=1}^{p} t_i \\ &||A\alpha - y||_2 \leq z \\ &||\alpha_i||_2 \leq t_i,\ i = 1, \ldots, p. \end{aligned}$$

Let us mention a regression scenario where block sparsity can be relevant.

Example: Consider a standard regression scenario where you have m data points in $\mathbb{R}^n$ and want to fit a function f to this data to minimize the sum of the squares of deviations. You conjecture that f belongs to one of three subclasses of functions: polynomials, exponentials, and trigonometric functions. For example, f is of the form

$$f(x) = \beta_1 x + \ldots + \beta_5 x^5 + \beta_6 e^x + \ldots + \beta_{10} e^{5x} + \beta_{11} \cos(x) + \beta_{12} \sin(y) + \ldots + \beta_{20} \cos(5x) + \beta_{21} \sin(5x).$$

The problem of finding which subclass of functions is most important to the regression is a LASSO problem with block sparsity. Our blocks in this case would be $\alpha_1 = [\beta_1, \ldots, \beta_5]^T$, $\alpha_2 = [\beta_6, \ldots, \beta_{10}]^T$ and $\alpha_3 = [\beta_{11}, \ldots, \beta_{21}]^T$.

5 Semidefinite programming (SDP)

Semidefinite programming is the broadest class of convex optimization problems we consider in this class. As such, we will study this problem class in much more depth.

5.1 Definition and basic properties

5.1.1 Definition

Definition 6. *A semidefinite program is an optimization problem of the form*

$$\begin{aligned} &\min_{X \in S^{n \times n}} \operatorname{Tr}(CX) \\ &s.t. \ \operatorname{Tr}(A_i X) = b_i, i = 1, \ldots, m, \\ &X \succeq 0, \end{aligned}$$

where the input data is $C \in S^{n \times n}, A_i \in S^{n \times n}, i = 1, \ldots, m, b_i \in \mathbb{R}, i = 1, \ldots, m.$

Notation:

- $S^{n \times n}$ denotes the set of $n \times n$ real symmetric matrices.
- Tr denotes the trace of a matrix; i.e., the sum of its diagonal elements (which also equals the sum of its eigenvalues).

A semidefinite program is an optimization problem over the space of symmetric matrices. It has two types of constraints:

- Affine constraints in the entries of the decision matrix X.
- A constraint forcing some matrix to be positive semidefinite.

The trace notation is used as a convenient way of expressing affine constraints in the entries of our unknown matrix. If A and X are symmetric, we have

$$\operatorname{Tr}(AX) = \sum_{i,j} A_{ij} X_{ij}.$$

Since X is symmetric, we can assume without loss of generality that A is symmetric as we have $\mathrm{Tr}(AX) = \mathrm{Tr}((\frac{A+A^T}{2})X)$ (why?). In some other texts, this assumption is not made and instead you would see the expression $\mathrm{Tr}(A^T X)$, which is the standard inner product between two matrices A and X.

We should also comment that the SDP presented above is appearing in the so-called *standard form*. Many SDPs that we encounter in practice do not appear in this form. What defines an SDP is really a constraints that requires a matrix to be positive semidefinite, with the entries of this matrix being affine expressions of decision variables.

Another common form of a semidefinite constraint is the following:

$$A_0 + x_1 A_1 + \ldots, + x_n A_n \succeq 0.$$

This is called a *linear matrix inequality* (LMI). The decision variables here are the scalars $x_1, \ldots, x_n$ and the symmetric matrices $A_1, \ldots, A_n$ are given as input. Can you write this constraint in standard form?

5.1.2 Why SDP?

The reasons will become more clear throughout this and future lectures, but here is a summary:

- SDP is a very natural generalization of LP, but the expressive power of SDPs is much richer than LPs.
- While broader than LP, SDP is still a convex optimization problem (in the geometric sense).
- We can solve SDPs efficiently (in polynomial time to arbitrary accuracy). This is typically done by interior point methods, although other types of algorithms are also available.
- When faced with a nonconvex optimization problem, SDPs typically produce much stronger bounds/relaxations than LPs do.
- Just like LP, SDP has a beautiful and well-established theory. Much of it mirrors the theory of LP.

5.1.3 Characterizations of positive semidefinite matrices (reminder)

When dealing with SDPs, it is useful to recall the different characterizations of psd matrices:

$X \succeq 0$

$\Leftrightarrow y^T X y \geq 0,\ \forall y \in \mathbb{R}^n$

$\Leftrightarrow$ All eigenvalues of X are ≥ 0

$\Leftrightarrow$ Sylvester's critierion holds: all $2^n - 1$ principal minors of X are nonnegative (see Lecture 2)

$\Leftrightarrow X = MM^T,$ for some $n \times k$ matrix M. This is called a Cholesky factorization.

Remark: $A \succeq B$ means $A - B \succeq 0$.

Proof of the Cholesky factorization:

($\Rightarrow$) Since X is symmetric, there exists an orthogonal matrix U such that

$$X = U^T D U,$$

where

$$D = diag(\lambda_1, \ldots, \lambda_n),$$

and $\lambda_i, i = 1, \ldots, n$, are the eigenvalues of X. Since eigenvalues of a psd matrix are nonnegative, we can define

$$\sqrt{D} = diag(\sqrt{\lambda_1}, \ldots, \sqrt{\lambda_n})$$

and take

$$M = U^T \sqrt{D} U.$$

($\Leftarrow$) This follows by noticing that $x^T X x = x^T M M^T x = ||M^T x||_2^2 \geq 0$. $\square$

5.1.4 A toy SDP example and the CVX syntax

Consider the SDP

$$\begin{aligned} &\min x + y \\ &\text{s.t.} \begin{pmatrix} x & 1 \\ 1 & y \end{pmatrix} \succeq 0 \\ &x + y \leq 3. \end{aligned} \tag{2}$$

You would code this in CVX as follows:

```
cvx_begin
variables x y
minimize (x+y)
[x 1;1 y]==semidefinite(2);
x+y<=3;
cvx_end
```

Exercise: Write this SDP in standard form.

Note: All SDPs can be written in standard form but this transformation is often not needed from the user (most solvers do it automatically if they need to work with the standard form).

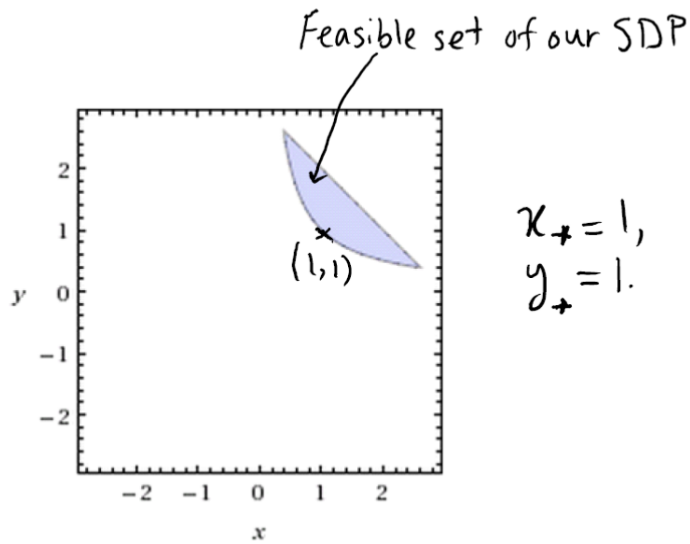

Figure 2: The feasible set of the SDP in (2).

5.1.5 Feasible set of SDPs

The feasible set of an SDP is called a *spectrahedron*. Every polyhedron is a spectrahedron (this is because every LP can be written as an SDP as we'll show shortly), but spectrahedra are far richer geometric objects than polyhedra (this is the reason why SDP is in general more powerful than LP). Examples of spectrahedra that are not polyhedra are given in Figure 2 and Figure 3.

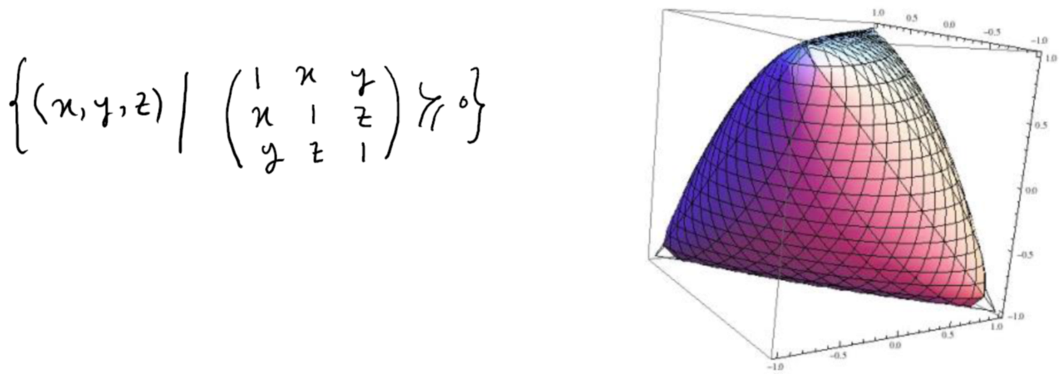

Figure 3: The so-called "elliptope."

Spectrahedra are always *convex* sets:

- Positive semidefinite $n \times n$ matrices form a convex set (why?).
- Affine constraints define a convex set.
- Intersection of convex sets is convex.

When we say an SDP is a convex optimization problem, we mean this is in the geometric sense:

- The objective is an affine function of the entries of the matrix.
- The feasible set is a convex set.
- However, the feasible set is not written in the explicit functional form "convex functions ≤ 0, affine function $= 0$".

To get a functional form, one can write an SDP as an infinite LP:

- Replace $X \succeq 0$ with linear constraints $y_i^T X y_i \geq 0$, for all $y \in \mathbb{R}^n$.
- We can reduce this to be a countable infinity by only taking $y \in \mathbb{Z}^n$ (why?).

Alternatively, we can write an SDP as a nonlinear program by replacing $X \succeq 0$ with $2^n - 1$ minor inequalities coming from Sylvester's critierion. However, treating the matrix constraint $X \succeq 0$ directly is often the right thing to do.

5.1.6 Attainment of optimal solutions

Unlike LPs, the minimum of an SDP may not always be attained. Here is a simple example:

$$\begin{aligned} &\min x_2 \\ &\text{s.t.} \begin{pmatrix} x_1 & 1 \\ 1 & x_2 \end{pmatrix} \succeq 0. \end{aligned} \tag{3}$$

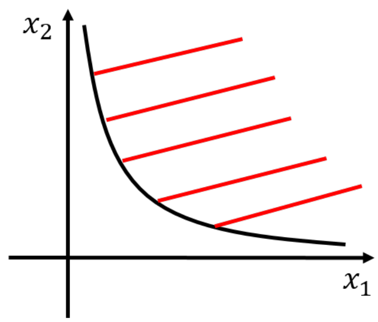

Figure 4: The feasible set of the SDP in (3).

5.2 Special cases of SDP: LP and SOCP

5.2.1 LP as a special case of SDP

Consider an LP

$$\begin{aligned} &\min_{x \in \mathbb{R}^n} c^T x \\ &\text{s.t. } a_i^T x = b_i, i = 1, \ldots, m, \\ &x \geq 0. \end{aligned}$$

For a vector v, let $\text{diag}(v)$ denote the diagonal matrix with v on its diagonal. Then, we can write our LP as the following SDP (why?):

$$\begin{aligned} &\min_X \ \text{Tr}(\text{diag}(c)X) \\ &\text{s.t. } \text{Tr}(\text{diag}(a_i)X) = b_i, \ i = 1, \ldots, m, \\ &\quad X \succeq 0. \end{aligned}$$

- So LP is really a special case of SDP where all matrices are diagonal — positive semidefiniteness for a diagonal matrix simply means nonnegativity of its diagonal elements.

- Like we mentioned already, geometry of SDP is far more complex than LP.
- For example, unlike polyhedra, spectahedra may have an infinite number of extreme points[1]. An example is the elliptope in Figure 3. Here's another simple example:

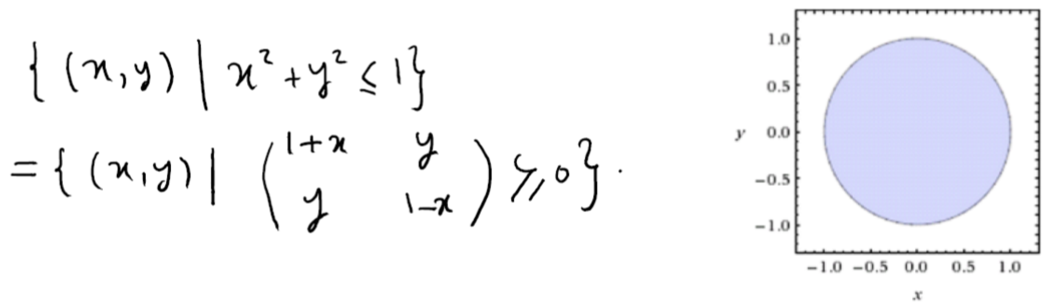

Figure 5: An example of a spectrahedron with an infinite number of extreme points

- This is the fundamental reason why SDP is not naturally amenable to "simplex type" algorithms.
- On the contrary, interior points for LP very naturally extend to SDP.

5.2.2 SOCP as a special case of SDP

To prove that SOCP is a special case of SDP, we first prove the following lemma that introduces the very useful notion of *Schur complements.*

Definition 7 (Schur complement)**.** *Given a symmetric block matrix* $X = \begin{bmatrix} A & B \\ B^T & C \end{bmatrix}$, *with* $det(A) \neq 0$, *the matrix* $S := C - B^T A^{-1} B$ *is called the Schur complement of* A *in* X*.*

Lemma 1. *Consider a block matrix* $X = \begin{pmatrix} A & B \\ B^T & C \end{pmatrix}$ *and let* $S := C - B^T A^{-1} B$*. If* $A \succ 0$*, then*

$$X \succeq 0 \Leftrightarrow S \succeq 0.$$

Proof: Let $f_v^* := \min_u f(u, v)$, where $f(u, v) := u^T A u + 2v^T B^T u + v^T C v$. Suppose $A \succ 0$, which implies that f is strictly convex in u. We can find the unique global solution of f over

[1]Recall that a point x is an extreme point of a convex set P if it cannot be written as a strict convex combination of two other points in P; i.e., $\nexists y, z \in P$ such that $x = \lambda y + (1 - \lambda) z$, for some $\lambda \in (0, 1)$.

u as follows:

$$\frac{\partial f}{\partial u} = 2Au + 2Bv = 0 \Rightarrow u = -A^{-1}Bv.$$

Hence, we obtain

$$\begin{aligned} f_v^* &= v^T B^T A^{-1} B v - 2 v^T B^T A^{-1} B v + v^T C v \\ &= v^T (C - B^T A^{-1} B) v \\ &= v^T S v. \end{aligned}$$

($\Rightarrow$) If $S \not\succeq 0$, then

$$\exists v \text{ s.t. } v^T S v < 0 \Rightarrow f_v^* < 0.$$

Picking

$$z = \begin{pmatrix} -A^{-1}Bv \\ v \end{pmatrix},$$

we obtain $z^T X z < 0$.

($\Leftarrow$) Take any $\begin{pmatrix} u \\ v \end{pmatrix}$. Then

$$\begin{pmatrix} u \\ v \end{pmatrix}^T X \begin{pmatrix} u \\ v \end{pmatrix} \geq f_v^* = v^T S v \geq 0.$$

□

Now let us use Schur complemets to show that SOCP is a special case of SDP.

Recall the general form of an SOCP:

$$\begin{aligned} &\min_x f^T x \\ &||A_i x + b_i||_2 \leq c_i^T x + d_i, \ i = 1, \ldots, m. \end{aligned}$$

We can assume $c_i^T x + d_i > 0$ (if not, one can argue separately and easily (why?)). Now we can write the constraint

$$||A_i x + b_i||_2 \leq c_i^T x + d_i$$

as

$$\begin{pmatrix} (c_i^T x + d)I & A_i x + b_i \\ (A_i x + b_i)^T & c_i^T x + d_i \end{pmatrix} \succeq 0. \tag{4}$$

Indeed, using Lemma 1,

$$
\begin{aligned}
(4) \Leftrightarrow & (c_i^T x + d_i) - (A_i x + b_i)^T \frac{1}{c_i^T x + d_i} (A_i x + b_i) \geq 0 \\
\Leftrightarrow & (c_i^T x_i + b_i)^2 \geq ||A_i x + b_i||_2^2 \\
\Leftrightarrow & c_i^T x_i + b_i \geq ||A_i x + b_i||_2
\end{aligned}
$$

as both terms are positive. □

5.3 Duality for SDP

Every SDP has a dual, which itself is an SDP. The primal and dual SDPs bound the optimal values of one another.

5.3.1 Deriving the dual of an SDP

Consider the primal SDP:

$$
\begin{aligned}
& \min_{X \in S^{n \times n}} \operatorname{Tr}(CX) \\
& \text{s.t. } \operatorname{Tr}(A_i X) = b_i, i = 1, \ldots, m, \\
& X \succeq 0,
\end{aligned}
$$

and denote its optimal value by p^*. To derive the dual, we define the Lagrangian function

$$
L(X, \lambda, \mu) = \operatorname{Tr}(CX) + \sum_i \lambda_i (b_i - \operatorname{Tr}(A_i X)) - \operatorname{Tr}(X\mu),
$$

and the dual function

$$
D(\lambda, \mu) = \min_X L(X, \lambda, \mu).
$$

The dual problem is then given by

$$
\begin{aligned}
& \max_{\lambda, \mu} D(\lambda, \mu) \\
& \text{s.t. } \mu \succeq 0.
\end{aligned}
\tag{5}
$$

Let us explain why the dual problem is defined this way.

Lemma 2. *For any λ, $\mu \succeq 0$ we have*

$$
D(\lambda, \mu) \leq p^*.
$$

Proof: We first prove a basic linear algebra fact, namely, if $A \succeq 0$ and $B \succeq 0$ then $\text{Tr}(AB) \geq 0$. Indeed, as $A \succeq 0$ and $B \succeq 0$, $\exists M, N$ such that $A = MM^T$ and $B = NN^T$ using the Cholesky decomposition. Then

$$\text{Tr}(AB) = \text{Tr}(MM^T NN^T) = \text{Tr}(N^T MM^T N) = ||M^T N||_F^2 \geq 0.$$

Now, let X^* be a primal optimal solution[2]. Then in view of the fact that $b_i - \text{Tr}(A_i X^*) = 0$, and $X^*, \mu \succeq 0$, we have

$$L(X^*, \lambda, \mu) = \text{Tr}(CX^*) - \text{Tr}(X^*\mu) = p^* - \text{Tr}(X^*\mu) \leq p^*,$$

where the last inequality follows from the claim we just proved above. Hence, we see that

$$D(\lambda, \mu) = \min_X L(X, \lambda, \mu) \leq p^*. \quad \square$$

So the motivation behind the dual problem (5) is to find the *largest lower bound* on p^*. Let us now write out the dual problem. Notice that

$$D(\lambda, \mu) = \min_X L(X, \lambda, \mu) = \begin{cases} \lambda^T b & \text{if } C - \sum_i \lambda_i A_i - \mu = 0 \\ -\infty & \text{otherwise} \end{cases}$$

(why?). In view of the fact that $\mu \succeq 0$, the condition $C - \sum_i \lambda_i A_i - \mu = 0$ can be rewritten as $C \succeq \sum_i \lambda_i A_i$. Hence, we can write out the dual SDP as follows:

$$\begin{aligned} &\max_{\lambda \in \mathbb{R}^m} \; b^T \lambda \\ &\text{s.t.} \; \sum_{i=1}^m \lambda_i A_i \preceq C. \end{aligned}$$

It is interesting to contrast this with the primal/dual LP pair in standard form:

$$\begin{aligned} (P) \quad & \min_x \; c^T x \\ & Ax = b \\ & x \geq 0, \end{aligned}$$

$$\begin{aligned} (D) \quad & \max_y \; b^T y \\ & A^T y \leq c. \end{aligned}$$

[2]We saw already that an SDP may not always achieve its optimal solution. We leave it to the reader to "fix" this proof for the case where the optimal solution is not achieved (hint: introduce an "ϵ").

5.3.2 Weak duality

Theorem 1 (Weak duality)**.** *Let X feasible be any feasible solution to the primal SDP (P) and let λ be any feasible solution to the dual SDP (D), then, we have*

$$Tr(CX) \geq b^T\lambda.$$

The proof was already done in Lemma 2.

5.3.3 Strong duality

Recall the strong duality theorem for LPs:

Theorem 2 (Strong duality for LP - reminder)**.** *Consider the primal-dual LP pair (P) and (D) given above. If (P) has a finite optimal value, then so does (D) and the two values match.*

Interestingly, this theorem does not hold for SDPs:

- It can happen that the primal optimal solution is achieved but the dual optimal solution is not (can you think of an example?).
- It can also happen that the primal and the dual both achieve their optima but the duality gap is nonzero (i.e., $d^* < p^*$) (can you think of an example?).

Fortunately, under mild additional assumptions, we can still achieve a strong duality result for semidefinite programming. One version of this theorem is stated below.

Theorem 3. *If the primal and dual SDPs are both strictly feasible (i.e., if there exists a solution that makes the matrix which needs to be positive semidefinite, positive definite), then both problems achieve their optimal value and $Tr(CX^*) = b^T\lambda^*$ (i.e., the optimal values match).*

6 Conclusion

In this lecture, we saw a hierarchy of convex optimization problems:

$$LP \subseteq \text{ (convex) } QP \subseteq \text{ (convex) } QCQP \subseteq SOCP \subseteq SDP.$$

In the upcoming lectures, we will dig deeper into SDP duality theory, as well as some applications of SDP. We will also cover the more general framework of *conic programming (CP)* which encompasses all problems classes studied in this lecture.

Notes

Further reading for this lecture can include Chapter 4 of [1] and Chapter 2 of [3].

References

[1] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, `http://stanford.edu/~boyd/cvxbook/`, 2004.

[2] C.G. Calafiore and L. El Ghaoui. *Optimization Models*. Cambridge University Press, 2014.

[3] M. Laurent and F. Vallentin. *Semidefinite Optimization*. 2012. Available at `http://www.mi.uni-koeln.de/opt/wp-content/uploads/2015/10/laurent_vallentin_sdo_2012_05.pdf`.