MATH 329, Lecture 14: Semidefinite programming

- **Linear programming** is: $\min_{x\in\mathbf{R}^n} c^\top x$ s.t. $Ax = b$ and $x \geq 0$. (Can reduce LPs to this std form.)
  This is: linear cost function, linear equality constraints, and specific inequalities.
- Abstract point of view: The inequalities express the fact that x belongs to a convex cone: $\mathbf{R}^n_+$.
- **Conic programming** generally: minimize linear cost under linear constraints ∩ convex cone.
- **Semidefinite programming** is conic programming in $\mathrm{Sym}(n) = \{X \in \mathbf{R}^{n\times n}: X = X^\top\}$ with cone
  $$\mathbf{K} = \{X \in \mathrm{Sym}(n) : X \succcurlyeq 0\} = \{X \in \mathrm{Sym}(n) : u^\top X u \geq 0 \text{ for all } u \in \mathbf{R}^n\}$$
  Note: $\mathbf{K}$ is clearly a cone. It is closed because $\mathbf{K} = \{X: \lambda_{\min}(X) \geq 0\}$ and $\lambda_{\min}$ is continuous.
  It is also convex since $X, Y \in \mathbf{K} \Rightarrow u^\top\big((1-t)X + tY\big)u \geq 0$ for all $t \in [0,1]$ and $u \in \mathbf{R}^n$.
- Let $\langle A, B\rangle = \mathrm{Tr}(A^\top B)$ be the usual inner product on $\mathrm{Sym}(n)$. An SDP in standard form is:
  $$\min_{X\in\mathrm{Sym}(n)} \langle C, X\rangle \quad \text{subject to } \mathcal{A}(X) = b \text{ and } X \succcurlyeq 0,$$
  where $\mathcal{A}: \mathrm{Sym}(n) \to \mathbf{R}^m$ is linear: $\mathcal{A}(X)_i = \langle A_i, X\rangle$ given $A_1, \ldots, A_m \in \mathrm{Sym}(n)$. This is convex.
- There are standard **algorithms** good for n up to ~10,000 (interior point method, …). Can use CVX.
- **Example 1**. An LP is an SDP:
  $\min_{X\in\mathrm{Sym}(n)} \langle \mathrm{diag}(c), X\rangle$ s.t. $\langle e_i e_j^\top + e_j e_i^\top, X\rangle = 0$ for $i \neq j$, $\langle \mathrm{diag}(a_i), X\rangle = b_i$ for $i = 1 \ldots m$, $X \succcurlyeq 0$.
- **Example 2**. $x_1^2 + x_2^2 \leq 1$ is the same as $\begin{bmatrix} 1 - x_1 & x_2 \\ x_2 & 1 + x_1 \end{bmatrix} \succcurlyeq 0$. Also $\begin{bmatrix} 1 & 0 & x_1 \\ 0 & 1 & x_2 \\ x_1 & x_2 & 1 \end{bmatrix} \succcurlyeq 0$.
  Can reason with Sylvester's criterion, but careful: for $\succcurlyeq$ (as opposed to $\succ$), must check *all* minors.
  More general: $\|Ax + b\|^2 \leq c^\top x + d$ holds iff $\begin{bmatrix} I & Ax + b \\ (Ax + b)^\top & c^\top x + d \end{bmatrix} \succcurlyeq 0$ (Schur complement).
  This is actually in "dual form" (see later), but (just as for LPs), there are standard tricks to express the above as an SDP in standard form. CVX knows these tricks and will apply them for you.
- A couple of facts from **linear algebra** we'll need in a moment. (This is really saying $\mathbf{K}^* = \mathbf{K}$.)
  - $\langle A, B\rangle \geq 0$ for all $A, B \in \mathbf{K}$. Proof: $B = \sum_i \lambda_i u_i u_i^\top$ so $\langle A, B\rangle = \sum_i \lambda_i u_i^\top A u_i \geq 0$.
    In particular: if $A \succcurlyeq 0$ then $\min_{B\in\mathbf{K}} \langle A, B\rangle = 0$.
  - On the other hand, if $A \not\succcurlyeq 0$, then there exists v such that $v^\top A v < 0$. Let $B = tvv^\top \succcurlyeq 0$ with $t \geq 0$ so $\langle A, B\rangle = tv^\top A v$; let $t \to \infty$ to see $\langle A, B\rangle \to -\infty$; hence $\min_{B\in\mathbf{K}} \langle A, B\rangle = -\infty$.
- **Duality**. The standard SDP is not quite of the form $h(x) = 0$ and $g(x) \leq 0$, but almost. By analogy, let's write a Lagrangian and derive a primal and a dual problem from there:
  $$L: \mathbf{K} \times \mathbf{R}^m \to \overline{\mathbf{R}}: (X, \mu) \mapsto L(X, \mu) = \langle C, X\rangle + \mu^\top\big(b - \mathcal{A}(X)\big)$$
  Note: we force X to live in its cone as part of the definition of L.
  1. $L_P(X) = \sup_{\mu\in\mathbf{R}^m} L(X, \mu) = \begin{cases} \langle C, X\rangle & \text{if } \mathcal{A}(X) = b \\ +\infty & \text{if } \mathcal{A}(X) \neq b \end{cases}$ so $\min_{X\in\mathbf{K}} L_P(X)$ is the original problem.
  2. $L_D(\mu) = \inf_{X\in\mathbf{K}} L(X, \mu) = \inf_{X\in\mathbf{K}} \langle C - \sum_i \mu_i A_i, X\rangle + \mu^\top b = \begin{cases} b^\top \mu & \text{if } C - \sum_i \mu_i A_i \succcurlyeq 0 \\ -\infty & \text{if } C - \sum_i \mu_i A_i \not\succcurlyeq 0 \end{cases}$
- **Dual problem**: $\max_{\mu\in\mathbf{R}^m} b^\top \mu$ subject to $\sum_i \mu_i A_i \preccurlyeq C$.
- **Notation**: $\langle \mu, \mathcal{A}(X)\rangle_{\mathbf{R}^m} = \mu^\top \mathcal{A}(X) = \sum_i \mu_i \langle A_i, X\rangle = \langle \sum_i \mu_i A_i, X\rangle$ so $\mathcal{A}^*(\mu) = \sum_i \mu_i A_i$ is adjoint.
- **Weak duality**: $L_D(\bar\mu) = \inf_{X\in\mathbf{K}} L(X, \bar\mu) \leq L(\bar X, \bar\mu) \leq \sup_{\mu\in\mathbf{R}^m} L(\bar X, \mu) = L_P(\bar X)$ for all $\bar X \in \mathbf{K}$, $\bar\mu \in \mathbf{R}^m$.
  Now (and not earlier) take sup over $\bar\mu$ and inf over $\bar X$ to conclude $d^* \leq p^*$.
- **Strong duality**: By analogy with convex programs: if the primal has a global minimizer and a "constraint qualification" holds there (we could extend definitions), then strong duality holds. Slater's CQ extends in a natural way: it holds globally iff there exists $X \succ 0$ such that $\mathcal{A}(X) = b$.

- **Dual of the dual**. $L(\mu, X) = b^\top \mu + \langle X, C - \mathcal{A}^*(\mu)\rangle = \left(b - \mathcal{A}(X)\right)^\top \mu + \langle C, X\rangle$ so $L_D(X) = \sup_{\mu \in \mathbf{R}^m} L(\mu, X) = \begin{cases} \langle C, X\rangle \text{ if } \mathcal{A}(X) = b \\ +\infty \quad \text{otherwhise} \end{cases}$, therefore the dual of the dual is $\min_{X \in \mathbf{K}} \langle C, X\rangle$ s.t. $\mathcal{A}(X) = b$.
- **MAX CUT**. NP-hard combinatorial problem. Given an undirected graph G with weights $w_{ij} \geq 0$, assign labels ± 1 to each node to maximize the sum of weights of edges connecting nodes with opposite labels. As an optimization problem:
$$\max_{x \in \mathbf{R}^n} \sum_{i,j} w_{ij} \frac{\left(x_i - x_j\right)^2}{4} \quad \text{s.t.} \quad x_i \in \{\pm 1\}, \quad i = 1 \ldots n.$$
- Develop cost using $x_i^2 = 1$: $\sum \frac{w_{ij}}{4}\left(x_i - x_j\right)^2 = \frac{1}{2}\sum w_{ij} - \frac{1}{2}\sum w_{ij} x_i x_j = \text{cst} - \frac{1}{2} x^\top W x$.
- Now again, but as a minimization problem with smooth cost and constraints:
$$\min_{x \in \mathbf{R}^n} x^\top W x \quad \text{s.t.} \quad 1 - x_i^2 = 0, i = 1 \ldots n.$$
- **Lagrange dual**. $L(x, \mu) = x^\top W x + \sum_i \mu_i \left(1 - x_i^2\right) = x^\top W x + 1^\top \mu - x^\top \text{diag}(\mu) x$.
Thus, $L_D(\mu) = \inf_{x \in \mathbf{R}^n} x^\top \left(W - \text{diag}(\mu)\right) x + 1^\top \mu = \begin{cases} 1^\top \mu \ \text{ if } \ W - \text{diag}(\mu) \succcurlyeq 0, \\ -\infty \quad \text{otherwise.} \end{cases}$
Dual problem is: $\max_{\mu \in \mathbf{R}^n} 1^\top \mu$ s.t. $\sum_i \mu_i e_i e_i^\top \preccurlyeq W$. This is an SDP in dual form.
Dual of the dual: $\min_{X \succcurlyeq 0} \langle W, X\rangle$ s.t. $\langle e_i e_i^\top, X\rangle = 1$ that is $\min_{X \succcurlyeq 0} \langle W, X\rangle$ s.t. $\text{diag}(X) = 1$.
- **Viewpoint 1**: Slater holds (because $I_n \succ 0$ and $\text{diag}(I_n) = 1$), and compact, hence we have **strong duality** for the two SDPs. Yet the first SDP is the dual of the original Max-Cut problem. Thus, the optimal value of the SDP (either of them) provides a bound on the optimal cut value.
- **Viewpoint 2**: The dual-of-the-dual SDP is a **relaxation** of the original problem. Indeed,
  If $x \in \mathbf{R}^n$ satisfies $x_i^2 = 1$ for all i, then $xx^\top \succcurlyeq 0$, $\text{diag}(xx^\top) = 1$ and $\text{rank}(xx^\top) = 1$.
  Conversely, if $X \succcurlyeq 0$, $\text{diag}(X) = 1$ and $\text{rank}(X) = 1$, then $X = xx^\top$ for x as above.
Also, $x^\top W x = \text{Tr}(x^\top W x) = \langle W, xx^\top\rangle$. Thus, the original problem is equivalent to
$$\min_{X \succcurlyeq 0} \langle W, X\rangle \quad \text{s.t.} \quad \text{diag}(X) = 1 \quad \text{and} \quad \text{rank}(X) = 1.$$
We see that the dual-of-the-dual is the same thing, only with the **rank constrained removed**. From that viewpoint, it is clear that the optimal value of the dual-of-the-dual provides a bound.
- How can we use this to (approximately) solve our original problem?
Solve the SDP (e.g., through CVX, to call a proper SDP solver). Two things can happen:
  Either X happens to have **rank = 1**. If so, great: we have solved the true problem! (Why?)
  Or X has **rank > 1**. If so, we could project: $X \approx uu^\top$ (dominant eigvec) and let $x = \text{sign}(u)$.
- **Goemans-Williamson**. Showed *random* rounding of X provides 87%-optimal cut in expectation.
- **Burer-Monteiro**. Instead of relaxing the rank from 1 to n, could relax the rank to p:
$$\min_{X \succcurlyeq 0} \langle W, X\rangle \quad \text{s.t.} \quad \text{diag}(X) = 1 \quad \text{and} \quad \text{rank}(X) \leq p$$
Observe: every $X \succcurlyeq 0$ with $\text{rank}(X) \leq p$ can be written as $X = YY^\top$ where $Y \in \mathbf{R}^{n \times p}$.
If also $\text{diag}(X) = 1$, then $\text{diag}(YY^\top) = 1$, which is equivalent to each row of Y having norm 1.
So:
$$\min_{Y \in \mathbf{R}^{n \times p}} \langle W, YY^\top\rangle \quad \text{s.t.} \quad \|y_i\|^2 = 1, i = 1 \ldots n.$$
For each $p < n$, you get a different non-convex relaxation. The non-convexity is well understood (and it's not too bad). Instead of labeling nodes ± 1, we place them on a sphere in $\mathbf{R}^p$.

- ~~(if time) **Polynomial optimization (multivariate) and sum-of-squares**. See end of these slides.~~