Lecture 32 (Nov 28): Semidefinite Programming

Lecturer: Zachary Friggstad *Scribe: Zachary Friggstad*

For square matrices $\mathbf{A}, \mathbf{C}$ we let $\mathbf{A} \circ \mathbf{C} = \sum_{i,j} \mathbf{A}_{i,j} \cdot \mathbf{C}_{i,j}$. Recall that $\mathbf{A} \succeq 0$ means $\mathbf{A}$ is a symmetric, positive semidefinite matrix

In a **semidefinite program (SDP)**, we are given square matrices $\mathbf{A}^1, \ldots, \mathbf{A}^m \in \mathbb{R}^{n \times n}$ with associated values $b_1, \ldots, b_m \in \mathbb{R}$ plus an **objective function** matrix $\mathbf{C} \in \mathbb{R}^{n \times n}$. Finally, there is a matrix of variables $\mathbf{X}$ over $\mathbb{R}^{n \times n}$.

The goal is to optimize the following:

$$\begin{array}{rrcll} \text{maximize :} & \mathbf{C} \circ \mathbf{X} & & & \\ \text{subject to :} & \mathbf{A}^i \circ \mathbf{X} & = & b_i & \forall\, 1 \leq i \leq m \\ & \mathbf{X} & \succeq & 0 & \end{array}$$

The left side of the equality constraints and the objective function itself are linear functions over the variables in $\mathbf{X}$. So the only really new ingredient here over linear programming is the p.s.d. constraint $\mathbf{X} \succeq 0$.

We note this is a convex optimization problem.

Lemma 1 *The set $\mathbf{S}_+ = \{\mathbf{X} \in \mathbb{R}^{n \times n} : \mathbf{X} \succeq 0\}$ is closed and convex.*

Proof. For convexity, suppose $\mathbf{X}, \mathbf{Y} \succeq 0$. Let $0 \leq \lambda \leq 1$. For any $\mathbf{z} \in \mathbb{R}^n$ we have

$$\mathbf{z}^T \cdot (\lambda \mathbf{X} + (1-\lambda)\mathbf{Y}) \cdot \mathbf{z} = \lambda(\mathbf{z}^T \mathbf{X} \mathbf{z}) + (1-\lambda) \cdot (\mathbf{z}^T \mathbf{Y} \mathbf{z}) \geq 0.$$

So $\lambda \cdot \mathbf{X} + (1-\lambda) \cdot \mathbf{Y} \succeq 0$. Therefore $\mathbf{S}_+$ is convex.

To see it is closed, let $\mathbf{X}^1, \mathbf{X}^2, \ldots,$ be any sequence of p.s.d. matrices whose entries converge to $\overline{\mathbf{X}}$. For any $\mathbf{z} \in \mathbb{R}^n$ it is easy to see then that the sequence of values $\rho_i := \mathbf{z}^T \cdot \mathbf{X}^i \cdot \mathbf{z}$ also converges to $\overline{\rho} := \mathbf{z}^T \cdot \overline{\mathbf{X}} \cdot \mathbf{z}$. As $\rho_i \geq 0$ for each i, then $\overline{\rho} \geq 0$ as well. This holds for each $\mathbf{z}$ so $\overline{\mathbf{X}} \geq 0$. This shows $\mathbf{S}_+$ is closed. ■

32.1 Bad Examples

Here are some examples demonstrating how semidefinite programming does not enjoy many nice properties exhibited by linear programs.

Cannot Achieve the Optimum

The theory of linear programming is quite nice. If an LP is feasible but not unbounded, then one can achieve the optimum solution. This is not the case with semidefinite programming, as the following example shows. Let $\mathbf{X}$ be a 2×2 matrix.

Consider

$$\begin{array}{rrcl} \text{maximize :} & -\mathbf{X}_{1,1} & & \\ \text{subject to :} & \mathbf{X}_{1,2} & = & 1 \\ & \mathbf{X} & \succeq 0 & \end{array}$$

Any feasible solution is a positive semidefinite matrix of the form $\begin{pmatrix} \mathbf{X}_{1,1} & 1 \\ 1 & \mathbf{X}_{2,2} \end{pmatrix}$. This is p.s.d. if and only if $\mathbf{X}_{1,1}, \mathbf{X}_{2,2} \geq 0$ and $\mathbf{X}_{1,1} \cdot \mathbf{X}_{2,2} \geq 1$. Equivalently $\mathbf{X}_{1,1} \geq \frac{1}{\mathbf{X}_{2,2}}$. So there are feasible solutions whose value is arbitrarily close to 0 but this maximum is never obtained.

Only Irrational Optimal Solutions

Consider the following SDP where the constraint matrices, r.h.s. values, and objective function are all given by rational values. However, there is only one optimal solution and it is irrational.

$$\begin{array}{rrcl} \text{maximize :} & \mathbf{X}_{1,2} & & \\ \text{subject to :} & \mathbf{X}_{1,1} & = & 1 \\ & \mathbf{X}_{2,2} & = & 2 \\ & \mathbf{X} & \succeq & 0 \end{array}$$

Recall that a 2×2 matrix is p.s.d. if and only if both its trace (sum of diagonal entries) and determinant are nonnegative. If we write $\mathbf{X} = \begin{pmatrix} 1 & x \\ x & 2 \end{pmatrix}$, we see trace$(\mathbf{X}) = 1+2$ and so $\mathbf{X} \succeq 0$ if and only if $\det \mathbf{X} = 2 - x^2 \geq 0$. That is, $x \leq \sqrt{2}$. So the unique optimum solution has $x = \sqrt{2}$.

Doubly-Exponential Feasible Solutions

One can imagine that the irrationality or lack of convergence to a true optimum is not a big problem in practice as long as we can find the optimum solution up to some small error tolerance (perhaps even violating feasibility by only a small amount). But a more fundamental problem is that a semidefinite program may have all feasible solutions being doubly-exponential in value! That is, expressing an explicit fractional feasible solution (or even near-feasible solution) would itself require exponentially many bits!

For example, consider the following matrix $\mathbf{A} \in \mathbb{R}^{2n \times 2n}$. All entries of $\mathbf{A}$ are 0 except for the following. For each $1 \leq i \leq n$ let $\mathbf{A}^i$ denote the 2×2 submatrix of $\mathbf{A}$ with rows $2i, 2i+1$ and columns $2i, 2i+1$ of $\mathbf{A}$ (so a small square around the diagonal). Let $x_1, \ldots, x_n$ be variables that will compose part of $\mathbf{A}$. Then $\mathbf{A}$ has the following form

$$\mathbf{A}^1 = \begin{pmatrix} 1 & 2 \\ 2 & x_1 \end{pmatrix},$$

and for $2 \leq i \leq n$ we have

$$\mathbf{A}^i = \begin{pmatrix} 1 & x_{i-1} \\ x_{i-1} & x_i \end{pmatrix},$$

If we assert $\mathbf{A} \succeq 0$, we can view this as asserting $\mathbf{X} \succeq 0$ for a matrix of variables and then setting the individual entries as above using various linear equalities.

It is easy to prove by induction on n that $x_n \geq 2^{2^n}$ for any feasible solution n and that this can be achieved by setting $x_i = 2^{2^i}$. So the SDP that maximizes $-x_n$ has doubly-exponential value.

To see the above claim, we argue about the individual 2×2 blocks. That is, it is simple to show $\mathbf{A} \succeq 0$ if and only if $\mathbf{A}^i \succeq 0$ for each i. In turn, this implies $\det \mathbf{A}^i \geq 0$ for each i. That is, $1 \cdot x_i - x_{i-1}^2 \geq 0$ (where we use $x_0 = 2$). Inductively, this means $x_i \geq x_{i-1}^2 \geq (2^{2^{i-1}})^2 = 2^{2^i}$. One can also check easily that setting $x_i = 2^{2^i}$ indeed satisfies $\mathbf{A}^i \succeq 0$ for each i.

Thankfully, we can at least avoid this issue in most combinatorial settings as the intended entries of the matrix are constrained to lie in a bounded set (e.g. $\mathbf{X}$ is often, ideally, containing only 0/1 entries or $-1/+1$ entries in the intended integer optimum).

32.2 Solving SDPs

We can still solve semidefinite programs in the following sense. Our statement is slightly informal, see [GS12] for more proper statements about the solvability of semidefinite programs.

"Theorem"
Assume the set of all feasible solutions $\mathbf{X}$ to a semidefinite program have $\sum_{i,j} \mathbf{X}_{i,j}^2 \leq R^2$ and that the input matrices and values are rational numbers. For $\epsilon > 0$, in time that is polynomial in $\log R, \log \frac{1}{\epsilon}$ and the bit complexity of the input we can find a solution with value at least $OPT - \epsilon$.

In truth, it is slightly messier than this because $\mathbf{X}$ itself might be slightly infeasible. We ignore this matter and simply note that what we discuss here can be adapted to handle this issue.

The main thing that drives such an algorithm is that we have a separation oracle for $\mathbf{S}_+$, the set of p.s.d. matrices. That is, given $\mathbf{X} \in \mathbb{Q}^{n \times n}$ we can either correctly assert $\mathbf{X} \succeq 0$ or find $\mathbf{z} \in \mathbb{Q}^n$ such that $\mathbf{z}^T \cdot \mathbf{X} \cdot \mathbf{z} < 0$. Reinterpreting this in the latter case, if we set $\mathbf{Z} = \mathbf{z} \cdot \mathbf{z}^T$ (the "outer" product) then $\mathbf{X} \circ \mathbf{Z} = \mathbf{z}^T \cdot \mathbf{X} \cdot \mathbf{z} < 0$ so such $\mathbf{z}$ produces a hyperplane separating $\mathbf{X}$ from $\mathbf{S}_+$.

We formalize this, with some low-level work left to the reader.

Theorem 1 *Let $\mathbf{X} \in \mathbb{Q}^{n \times n}$ be symmetric. In time that is polynomial in the size of $\mathbf{X}$, we can either correctly assert $\mathbf{X} \succeq 0$ or find $\mathbf{z} \in \mathbb{Q}^n$ with $\mathbf{z}^T \cdot \mathbf{X} \cdot \mathbf{z} < 0$.*

Proof. We essentially step through the Cholesky decomposition algorithm which attempts to find vectors $\mathbf{v}_1, \ldots, \mathbf{v}_i$ such that $\mathbf{X}_{i,j} = \mathbf{v}_i \circ \mathbf{v}_j$. The Cholesky decomposition algorithm itself uses square roots to compute the $\mathbf{v}_i$ vectors, but one can observe that if the decomposition fails then we can in fact produce a rational vector certifying $\mathbf{X}$ is not p.s.d.

Write

$$\mathbf{X} = \begin{pmatrix} \alpha & \mathbf{q}^T \\ \mathbf{q} & \mathbf{X}' \end{pmatrix}$$

with $\alpha \in \mathbb{Q}, \mathbf{q} \in \mathbb{Q}^{n-1}, \mathbf{X}' = \mathbb{Q}^{(n-1)\times(n-1)}$.

- If $\alpha < 0$ then $\mathbf{X}$ is not p.s.d. and we output $\mathbf{z} = (1, 0, 0, \ldots, 0)^T$.
- If $\alpha = 0$ then one can show we must have $\mathbf{q} = \mathbf{0}$. Otherwise, setting $\mathbf{z}$ as
  - $\mathbf{z}_1 = -\frac{\mathbf{X}' \cdot \mathbf{q}}{2\mathbf{q}^T \cdot \mathbf{q}} - 1$
  - $\mathbf{z}_i = \mathbf{q}_{i-1}$ for $2 \leq i \leq n$

  has $\mathbf{z}^T \cdot \mathbf{X} \cdot \mathbf{z} < 0$.
- If $\alpha > 0$, we recursively check that $\mathbf{X}'' := \mathbf{X}' - \frac{1}{\alpha} \cdot \mathbf{q} \cdot \mathbf{q}^T \succeq 0$. If so, we can declare $\mathbf{X}$ is p.s.d., if not then we are also given (recursively) $\mathbf{z}' \in \mathbb{Q}^{n-1}$ with $\mathbf{z}'^T \cdot \mathbf{X}'' \cdot \mathbf{z}' < 0$. We then set $\mathbf{z}_1 = -\frac{1}{\alpha} \sum_{i=1}^{n-1} \mathbf{z}'_i \mathbf{q}_i$ and $\mathbf{z}_i = \mathbf{z}'_{i-1}$ for $2 \leq i \leq n$. One can check $\mathbf{z}^T \cdot \mathbf{X} \cdot \mathbf{z} < 0$.

A careful inspection reveals the bit complexity of the returned value $\mathbf{z}$ is polynomial in the bit complexity of $\mathbf{X}$. ■

32.3 Application: Approximating Maximum Cut

Let $\mathcal{G} = (V; E)$ be an undirected graph. The goal is to find $S \subseteq V$ such that $|\delta(S)|$ is as large as possible. Note this is a special case of symmetric submodular maximization, and we already saw a naive 2-approximation for this problem by letting S be a random subset of V.

Consider the following SDP where we view the matrix of variables as being indexed by vertices in V.

$$\begin{array}{rrcll} \text{maximize :} & \sum_{uv \in E} \frac{1-\mathbf{X}_{u,v}}{2} & & & \\ \text{subject to :} & \mathbf{X}_{v,v} & = & 1 & \text{for each } v \in V \\ & \mathbf{X} & \succeq & 0 & \end{array}$$

The objective function sums over each edge precisely once. Note that the objective function is affine, not just linear, but clearly we can maximize the linear function $\sum_{uv \in E} -\mathbf{X}_{u,v}$ to maximize the affine function.

Let $S^* \subseteq V$ be an optimum cut. By setting $\mathbf{X}_{u,v} = -1$ if $uv \in \delta(S^*)$ and $\mathbf{X}_{u,v} = 1$ if $uv \notin \delta(S^*)$ we get a feasible solution. That is, if we let $x_v = 1$ if $v \in S^*$ and $x_v = -1$ if $v \notin S^*$ then $\mathbf{X}_{u,v} = x_u \cdot x_v$, so we have described a Cholesky factorization of $\mathbf{X}$ meaning $\mathbf{X} \succeq 0$. For $uv \in E$, in this solution we have $\frac{1-\mathbf{X}_{u,v}}{2} = 1$ if $uv \in \delta(S^*)$ and $\frac{1-\mathbf{X}_{u,v}}{2} = 0$ if $uv \notin \delta(S^*)$. That is, the objective function of this given $\mathbf{X}$ value is exactly $|\delta(S^*)|$.

Therefore, if we let $OPT = |\delta(S^*)|$ denote the maximum cut value and OPT_{SDP} denote the optimum SDP value, then we have seen

Lemma 2 *$OPT \leq OPT_{SDP}$*

Conversely, let $\alpha := \frac{2}{\pi} \cdot \min_{0 \leq \theta \leq \pi} \frac{\theta}{1 - \cos\theta}$. Numeric evaluation shows $\alpha > 0.87856$.

Theorem 2 *$OPT \geq \alpha \cdot OPT_{SDP}$. Moreover, given a feasible solution $\mathbf{X} \in \mathbb{Q}^{n \times n}$ to the above SDP there is a randomized algorithm that finds a cut S with $\mathbf{E}[\delta(S)] \geq \alpha \cdot \sum_{uv \in E} \frac{1-\mathbf{X}_{u,v}}{2} - \epsilon$ in time that is polynomial in the size of $\mathbf{X}$ and $\log \frac{1}{\epsilon}$.*

Proof. Let $\mathbf{X}$ be feasible and let $\mathbf{x}_v \in \mathbb{R}^n$ for each $v \in V$ be vectors such that $\mathbf{x}_v^T \cdot \mathbf{x}_u = \mathbf{X}_{u,v}$. In practice, if $\mathbf{X} \in \mathbb{Q}^{n \times n}$ one can compute such vectors within arbitrary precision using the Cholesky decomposition algorithm where each square root is computed within arbitrary precision. The precise statements about the relationship with the desired ϵ are left up to the reader.

Cut the n-dimensional unit sphere in half along a random plane. In particular, sample a random $\eta \in \mathbb{R}^n$ where each entry $\eta_v \sim \mathcal{N}(0, 1)$ is an independent standard Gaussian. Set $S = \{v \in V : \eta^T \cdot \mathbf{x}_v > 0\}$. That is, set S to be all points lying on one side of the cut. Once again, we can effectively do this sampling up to an ϵ in practice (e.g. through some appropriate discretization of the density function of the normal distribution), but we will not focus on these details.

We will be done after showing the following.

Claim: For each $uv \in E$ we have $\mathbf{Pr}[e \in \delta(S)] \geq \alpha \cdot \frac{1-\mathbf{X}_{u,v}}{2}$.

We prove the claim using geometric arguments that can be formalized with some care. Consider some $uv \in E$ and consider the plane including points $\mathbf{x}_u, \mathbf{x}_v$ and the origin $\mathbf{0}$. See Figure 32.1 for an illustration.

By symmetry, the random hemispherical cut is cutting this plane along a random line through the origin. So $uv \in \delta(S)$ if and only if this random line cuts $\mathbf{x}_u$ and $\mathbf{x}_v$ on this plane.

Let θ_{uv} be the angle between $\mathbf{x}_u$ and $\mathbf{x}_v$. The random line separates u from v with probability exactly $\frac{\theta_{uv}}{\pi}$.

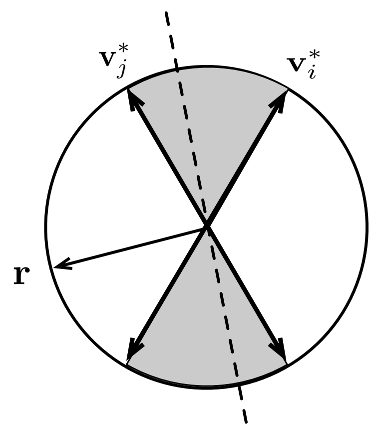

Figure 32.1: The vectors η, $\mathbf{x}_u$, and $\mathbf{x}_v$ from the text are shown as $\mathbf{r}$, $\mathbf{v}_i^*$, and $\mathbf{v}_j^*$; this image is borrowed from one of my previous courses. Note the two vectors separating $\mathbf{x}_u, \mathbf{x}_v$ are separated if and only if the dashed line (normal to η) cuts through the shades region, which happens precisely with probability θ_{uv}/π.

Recall $\mathbf{X}_{u,v} = \mathbf{x}_u \circ \mathbf{x}_v = ||\mathbf{x}_u|| \cdot ||\mathbf{x}_v|| \cdot \cos\theta_{uv} = \cos\theta_{uv}$. Thus, by definition of α we have

$$\mathbf{Pr}[uv \in \delta(S)] = \frac{\theta_{uv}}{\pi} \geq \alpha \cdot \frac{1 - \mathbf{X}_{u,v}}{2}.$$

■

References

GS12 B. Gartner and A. Schrijver, Approximation Algorithms and Semidefinite Programming, Springer, 2012.