ECE285: Semidefinite and sum-of-squares optimization **Winter 2023**

Lecture 10: Applications of SDPs in combinatorial problems

Lecturer: Yang Zheng *Scribe: Yang Zheng*

Disclaimer: *These notes have not been subjected to the usual scrutiny reserved for formal publications. Any typos should be sent to* `zhengy@eng.ucsd.edu`.

Learning goals:

1. Binary quadratic optimization
2. The maximum cut problem
3. The independent set problem

1 Binary quadratic optimization

Binary quadratic optimization is a classical combinatorial optimization problem. We consider a problem with quadratic cost function and decision variables taking values ± 1, i.e., we aim to minimize an indefinite quadratic function over the vertices of n-dimensional hypercube. This is a problem of the form

$$\begin{aligned} \min_{x} \quad & x^{\mathsf{T}} Q x \\ \text{subject to} \quad & x_i \in \{-1, 1\}, i = 1, \ldots, n, \end{aligned} \tag{1}$$

where $Q \in \mathbb{S}^n$ is given. The binary constraints can be modeled using quadratic constraints, i.e.,

$$x_i \in \{-1, 1\} \quad \Longleftrightarrow \quad x_i^2 = 1.$$

These quadratic constraints define a finite set, but with an exponential number of elements. Many well-known problems are in the form of (1). We will mention the famous maximum cut problem and the stable set problem in this lecture.

The problem (1) is equivalent to

$$\begin{aligned} \min_{x} \quad & x^{\mathsf{T}} Q x \\ \text{subject to} \quad & x_i^2 = 1, i = 1, \ldots, n. \end{aligned} \tag{2}$$

It is known that this problem is NP-hard. This is true even if the objective function is strictly convex, i.e., Q is positive definite, since $x_i^2 = 1$ and we can add a large constant $d(x_1^2 + \ldots + x_n^2)$ to make Q positive definite.

1.1 Semidefinite relaxation

We denote the optimal value of (2) as p^* and an optimal solution as x^* such that $(x^*)^{\mathsf{T}} Q x^* = p^*$. Computing the exact solution for (2) is computationally hard. We are interested in computing accurate bounds on its optimal value. Upper bounds can be directly obtained from any feasible solutions: if $x_0 \in \mathbb{R}^n$ has binary values as ± 1, it always holds that $p^* \leq x_0^{\mathsf{T}} Q x_0$ (but this bound may be very loose).

To prove lower bounds, we need to use some sort of relaxation, i.e. we consider a minimization over a larger region (or you may consider the dual problem formulation in Lecture 7). One naive method is to relax the binary variables as $-1 \leq x_i \leq 1$, which naturally provides a lower bound. However, this bound is indeed often very loose.

Here, we introduce the following primal-dual pair of semidefinite programs to provide a lower bound:

$$\begin{aligned} \min_{X} \quad & \langle Q, X \rangle \\ \text{subject to} \quad & X_{ii} = 1, i = 1, \ldots, n \\ & X \in \mathbb{S}^n_+, \end{aligned} \tag{3a}$$

$$\begin{aligned} \max_{\Lambda} \quad & \text{trace}(\Lambda) \\ \text{subject to} \quad & Q - \Lambda \succeq 0 \\ & \Lambda \text{ diagonal.} \end{aligned} \tag{3b}$$

Several comments are:

1. It is easy to see that both (3a) and (3b) are strictly feasible. Therefore, strong duality holds and the optimal values are the same. We denote it as p^*_{sdp}.

2. SDPs (3a) and (3b) have matrix variables in $\mathbb{S}^n_+$, while the variable in (2) is a vector $x \in \mathbb{R}^n$. So, after the relaxation, the dimension of the optimization variable has been lifted.

3. For every feasible x in (2), the matrix $X = xx^\mathsf{T}$ is also feasible to (3a) with the same cost value:

$$\langle Q, X \rangle = \langle Q, xx^\mathsf{T} \rangle = x^\mathsf{T} Q x.$$

   Thus, solving (3a) returns a lower bound on $p^* \geq p^*_{\text{sdp}}$. Similarly, for very feasible solution $\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_n)$ in (3b), we have

$$x^\mathsf{T} Q x \geq x^\mathsf{T} \Lambda x = \sum_{i=1}^{n} \lambda_i x_i^2 = \text{trace}(\Lambda),$$

   where we used the fact that $x_i^2 = 1$, thus solving (3b) gives a lower bound as well.

4. In certain cases, the bound p^*_{sdp} from (3a) and (3b) is *provably* good. Well-known examples include when $-Q$ is diagonally dominant or positive semidefinite, or has a bipartite structure (see [3, Page 31, Chapter 2.2.2]). In these cases, it has been shown that there is at most a small constant factor between p^* and p^*_{sdp}. In the case of max cut, $-Q$ is diagonally dominant.

5. If the solution to the primal SDP (3a) has rank one, then we have obtained the global solution, i.e., $p^*_{\text{sdp}} = p^*$. Indeed, any rank one matrix $X \succeq 0$ with $X_{ii} = 1$ must be of the form $X = xx^\mathsf{T}$ with $x_i \in \{-1, 1\}, i = 1, \ldots, n$. If we add an additional constraint "rank(X) = 1" to (3a), the resulting rank-constrained SDP is equivalent to the original binary QP (2).

6. We can consider the matrix X in (3a) as a covariance matrix: suppose x is a random vector on the hypercube $\{-1, 1\}^n$ with mean $\mathbb{E}(x) = 0$, then its covariance matrix $\mathbb{E}(xx^\mathsf{T})$ satisfies $X \succeq 0$ and $X_{ii} = 1, i = 1, \ldots, n$. Furthermore, the cost in (3a) is just the expected cost

$$\mathbb{E}(x^\mathsf{T} Q x) = \mathbb{E}(\langle Q, xx^\mathsf{T} \rangle) = \langle Q, X \rangle.$$

Rounding: We know that the optimal value of (3a) and (3b) provides a lower bound. One natural question is how to generate a feasible solution to (2) that achieves a good cost value. Another question is whether we can quantify the quality of the bounds by (3a) and (3b).

Here, we describe a randomized rounding procedure introduced by Geomans and Williamson in the seminal work [5], which produces a binary vector x from the SDP solution $X \succeq 0$:

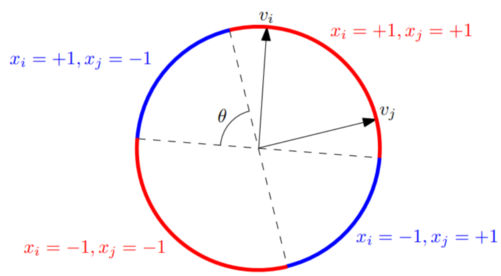

Figure 1: Computation of $\mathbb{E}(x_i x_j)$ for x defined in (4). Let $\theta = \arccos(\langle v_i, v_j\rangle)$ be the angle between v_i and v_j. The probability of having $x_i x_j = -1$ is $2\theta/2\pi$ and the probability of having $x_i x_j = +1$ is $(2\pi - 2\theta)/2\pi$. (Figure taken from [4, Lecture 8])

1. Factorize the solution $X = V^\top V$ where

$$V = \begin{bmatrix} v_1, v_2, \ldots, v_n \end{bmatrix} \in \mathbb{R}^{r \times n},$$

   where r is the rank of X. Since $X_{ij} = v_i^\top v_j$, and $X_{ii} = 1$. The vectors $v_i, i = 1, \ldots, n$ are on the unit sphere in $\mathbb{R}^r$.

2. Choose a uniformly distributed random hyperplane in $\mathbb{R}^r$, passing through the origin, and assign each variable x_i either $+1$ or -1, depending which side of hyperplane the point v_i lies in.

By a simple geometric argument, we can quantify the expected value of the cost function.

Lemma 10.1. *Given an $X \succeq 0$, we let*

$$x_i = \mathit{sign}(\langle v_i, z\rangle), i = 1, \ldots, n, \tag{4}$$

where $X = V^\top V$ and z is a standard random Gaussian vector. Then, we have

$$\mathbb{E}(x_i x_j) = 1 - \frac{2}{\pi}\arccos(X_{ij}).$$

Proof. We note the angle between v_i and v_j is θ, i.e.

$$\cos\theta = \frac{\langle v_i, v_j\rangle}{\|v_i\|\|v_j\|} = \langle v_i, v_j\rangle,$$

where we applied the fact that $\|v_i\| = \|v_j\| = 1$. From a geometric viewpoint (see Figure 1), it is not difficult to see that

- the probability of $x_i x_j = -1$ is $\frac{\theta}{\pi}$;
- the probability of $x_i x_j = 1$ is $1 - \frac{\theta}{\pi}$.

Therefore, we have

$$\begin{aligned}
\mathbb{E}(x_i x_j) &= -1 \times \frac{\theta}{\pi} + 1 \times \left(1 - \frac{\theta}{\pi}\right) \\
&= 1 - \frac{2\theta}{\pi} \\
&= 1 - \frac{2}{\pi}\arccos(X_{ij}).
\end{aligned}$$

□

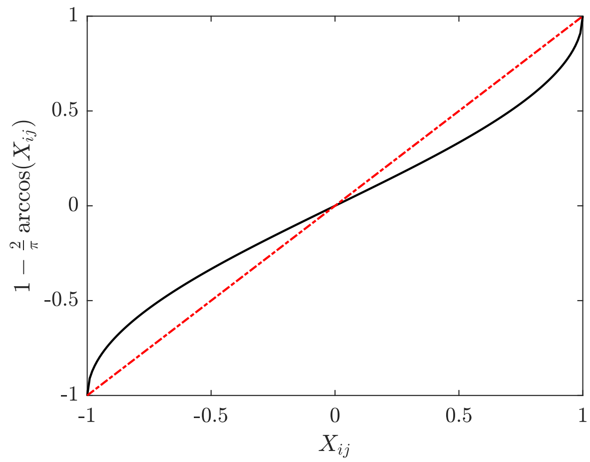

Figure 2: Plot of $\Sigma_{ij} = 1 - \frac{2}{\pi}\arccos(X_{ij})$.

Note that $\mathbb{E}(x_i x_j)$ and X_{ij} are actually close when $-1 < X_{ij} < 1$ (see Figure 2). Now, from the solution $X \in \mathbb{S}^n_+$ of (3a), we construct a random vector $x \in \{-1, 1\}^n$ in (4) that satisfies $\mathbb{E}(x) = 0$ with covariance $\Sigma = \mathbb{E}(xx^\mathsf{T})$ given by

$$\Sigma_{ij} = 1 - \frac{2}{\pi}\arccos(X_{ij}). \tag{5}$$

1.2 The maximum cut problem

We consider the famous maximum cut problem: Given an undirected weighted graph $\mathcal{G}(\mathcal{V}, \mathcal{E})$ with vertex set $\mathcal{V}$ and edge set $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$, and weights W, a cut is a partition of $\mathcal{V}$ into two disjoint sets $\mathcal{S}, \bar{\mathcal{S}}$ where $S \subset \mathcal{V}$ and $\bar{S} = \mathcal{V} \setminus \mathcal{S}$.

The value of a cut is the total weights of the edges connecting elements in $\mathcal{S}$ and elements in $\bar{\mathcal{S}}$:

$$\sum_{i \in \mathcal{S}, j \in \bar{\mathcal{S}}} w_{ij}.$$

The maximum cut problem is to find a cut with maximum value. We use a binary vector $x \in \{-1, +1\}^n$ (where $n = |\mathcal{V}|$) to denote a partition $\mathcal{S}, \bar{\mathcal{S}}$: $x_i = +1$ if $i \in \mathcal{S}$, otherwise $x_i = -1$. Then the value of a cut is

$$\frac{1}{2}\sum_{i,j \in \mathcal{V}} w_{ij}\frac{(x_i - x_j)^2}{4} = \frac{1}{2}\sum_{i,j \in \mathcal{V}} w_{ij}\frac{x_i^2 + x_j^2 - 2x_i x_j}{4} = \frac{1}{4}\sum_{i,j \in \mathcal{V}} w_{ij}(1 - x_i x_j) := \frac{1}{4}x^\mathsf{T} L x,$$

where the matrix $L = [l_{ij}] \in \mathbb{S}^n$ is defined as

$$l_{ij} = \begin{cases} \sum_{j \neq i} w_{ij} & \text{if } i = j \\ -w_{ij} & \text{otherwise.} \end{cases} \tag{6}$$

Note that this matrix L is diagonally dominant since the diagonal element $l_{ii} \geq \sum_{j \neq i} |l_{ij}|, i = 1, \ldots, n$.

The maximum cut problem can thus be written as

$$\begin{aligned} \max_{x} \quad & \frac{1}{4}x^\mathsf{T} L x \\ \text{subject to} \quad & x_i^2 = 1, i = 1, \ldots, n, \end{aligned} \tag{7}$$

which is in the form of (2) with $Q = -L$. The SDP relaxation of (7) is

$$\begin{aligned} \max_{x} \quad & \frac{1}{4}\langle L, X\rangle \\ \text{subject to} \quad & X_{ii} = 1, i = 1, \ldots, n, \\ & X \succeq 0. \end{aligned} \tag{8}$$

We have the famous result due to Geomans and Willianmson [5].

Theorem 10.1 (Geomans-Williamson [5])**.** *Let v^* be the optimal value of the maximum cut problem* (7) *and let p^*_{sdp} be the optimal value of its SDP relaxation* (8)*. Then*

$$\alpha \cdot p^*_{psd} \leq v^* \leq p^*_{sdp},$$

where $\alpha = \min_{t\in[-1,1)} \frac{1-f(t)}{1-t} \approx 0.878$ with $f(t) = 1 - \frac{2}{\pi}\arccos(t)$.

Proof. The part $v^* \leq p^*_{\text{sdp}}$ is directly from the relaxation. Given an optimal solution X of (8), we construct a random vector x as in (4). By definition, we have

$$v^* \geq \frac{1}{4}x^{\mathsf{T}}Lx \qquad \Rightarrow \qquad v^* \geq \mathbb{E}(\frac{1}{4}x^{\mathsf{T}}Lx) = \frac{1}{4}\langle L, \Sigma\rangle.$$

Then, we have

$$\langle L, \Sigma\rangle = \sum_{i,j} w_{ij}(1 - \Sigma_{ij}) = \sum_{i,j\in\mathcal{V}} w_{ij}(1 - f(X_{ij})) \geq \alpha \sum_{i,j\in\mathcal{V}} w_{ij}(1 - X_{ij}) = \alpha\langle L, X\rangle.$$

Therefore, we have proven

$$\alpha \cdot p^*_{\text{psd}} \leq v^* \leq p^*_{\text{sdp}}.$$

□

Remark 10.1. *In the maximum cut problem, the matrix $-Q = L$ is diagonally dominant; see* (6)*. When the matrix $-Q$ is positive semidefinite or has a bipartite structure, the optimal value of the SDP relaxation* (3a) *has provably quality as well; see [3, Page 31, Chapter 2.2.2].* □

2 The independent set problem

We now look at another combinatorial problem: the independent set problem. Given an undirected graph $\mathcal{G}(\mathcal{V}, \mathcal{E})$ with $|\mathcal{V}| = n$.

- An independent set (or stable set) of $\mathcal{G}$ is a subset $\mathcal{S} \subseteq \mathcal{V}$ such that no two vertices in $\mathcal{S}$ are connected by an edge, i.e., $(i, j) \notin \mathcal{E}, \forall i, j \in \mathcal{S}$.
- The size of the largest stable set of a graph, denoted by $\alpha(\mathcal{G})$, is called the stability number of the graph.
- The problem of testing if $\alpha(\mathcal{G})$ is larger than a given integer k is NP-hard.
- The maximum stable set problem is to find the largest stable set in a graph.

The stable set problem can be formulated as

$$\begin{aligned} \max_{x\in\mathbb{R}^n} \quad & \sum_{i=1}^{n} x_i \\ \text{subject to} \quad & x_i \in \{0, 1\}, i = 1, \ldots, n \\ & x_i x_j = 0, \ \ \forall (i, j) \in \mathcal{E}. \end{aligned} \tag{9}$$

The stable set $\mathcal{S}$ corresponds to the set of $x_i = 1$. The constraint $x_i x_j = 0$ ensures that $\mathcal{S}$ is a stable set, and the objective function counts the cardinality of $\mathcal{S}$. The constraint $x_i \in \{0, 1\}$ can be modeled by $x_i^2 = x_i$.

2.1 Semidefinite relaxation

We now introduce a semidefinite relaxation for (9), which provides an upper bound

$$\begin{aligned} \vartheta(\mathcal{G}) := \max_{x\in\mathbb{R}^n, X\in\mathbb{S}^n} \quad & \sum_{i=1}^n x_i \\ \text{subject to} \quad & X_{ii} = x_i, i = 1,\dots,n \\ & X_{ij} = 0,\ \forall (i,j)\in\mathcal{E} \\ & \begin{bmatrix} 1 & x^\top \\ x & X \end{bmatrix} \succeq 0. \end{aligned} \tag{10}$$

This relaxation was first proposed by Lovász in [6]. It is easy to see that (10) provides an upper bound $p^*_{\text{sdp}} \geq \alpha(\mathcal{G})$.

Theorem 10.2. *Let* $\alpha(\mathcal{G})$ *be the optimal value of* (9) *and* $\vartheta(\mathcal{G})$ *be the optimal value of* (10)*. Then, we have*

$$\alpha(\mathcal{G}) \leq \vartheta(\mathcal{G}).$$

Proof. This is observed from the fact that if x is feasible to (9), then the pair $x, X = xx^\top$ is feasible to (10) since

$$\begin{bmatrix} 1 & x^\top \\ x & X \end{bmatrix} = \begin{bmatrix} 1 \\ x \end{bmatrix} \begin{bmatrix} 1 \\ x \end{bmatrix}^\top \succeq 0.$$

□

Another question is to ask whether there is a constant $c > 0$ such that $c \cdot \vartheta(\mathcal{G}_n) \leq \alpha(\mathcal{G})$ for all graphs. Unfortunately, this is not true. The following result holds:

Theorem 10.3. *There exists a sequence of graphs* $\mathcal{G}_n$ *such that*

$$\frac{\vartheta(\mathcal{G}_n)}{\alpha(\mathcal{G}_n)} \geq \frac{\sqrt{n}}{3\log n} \to \infty, \ \text{as } n \to \infty.$$

The proof appears in [2], which we will not discuss in this lecture. You can also refer to [4, Lecture 9] for a proof.

Remark 10.2. *There is another similar semidefinite relaxation for the stable set problem* (9) *as follows*

$$\begin{aligned} \max_{X} \quad & \langle J, X\rangle \\ \textit{subject to} \quad & \textit{trace}(X) = 1 \\ & X_{ij} = 0, \quad (i,j)\in\mathcal{E}, \\ & X \succeq 0, \end{aligned} \tag{11}$$

where J *is a matrix with all entries equal to one. It is not difficult to verify that* (11) *is a relaxation of* (9)*. Indeed, we let* η *be a zero/one vector of length* n *where the* i*th element is* 1 *if and only if node* i *in* S*, and define*

$$x = \frac{1}{\sqrt{|S|}}\eta, \quad X = xx^\top,$$

where $|S|$ *denote the cardinality of set* S*. Then, we can verify that* X *is a feasible solution to* (11)*, i.e.*

$$X = xx^\top \succeq 0,$$

$$X_{ij} = \frac{1}{|S|}x_i x_j = 0, (i,j)\in\mathcal{E},$$

$$\text{trace}(X) = \text{trace}\left(\frac{1}{|S|}\eta\eta^\top\right) = \frac{1}{|S|}\text{trace}(\eta^\top\eta) = \frac{|S|}{|S|} = 1.$$

Also, we have the same cost value

$$\langle J, X\rangle = \left\langle 11^\intercal, \frac{1}{|S|}\eta\eta^\intercal \right\rangle = \frac{1}{|S|}\mathrm{trace}(\eta^\intercal 11^\intercal \eta) = \frac{1}{|S|}(1^\intercal \eta)^2 = |S|.$$

Notes

The preparation of this lecture was based on [1, Lecture 11] and [4, Lectures 7-9]. Further reading for this lecture can refer to [3, Chapter 2].

References

[1] Amir Ali Ahmadi. ORF523: Convex and Conic Optimization, Spring 2021.

[2] Aharon Ben-Tal and Arkadi Nemirovski. *Lectures on modern convex optimization: analysis, algorithms, and engineering applications*. SIAM, 2001.

[3] Grigoriy Blekherman, Pablo A Parrilo, and Rekha R Thomas. *Semidefinite optimization and convex algebraic geometry*. SIAM, 2012.

[4] Hamza Fawzi. Topics in Convex Optimisation, Michaelmas 2018.

[5] Michel X Goemans and David P Williamson. Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming. *Journal of the ACM (JACM)*, 42(6):1115–1145, 1995.

[6] László Lovász. On the shannon capacity of a graph. *IEEE Transactions on Information theory*, 25(1):1–7, 1979.