Scribe: Patrick Steele

1 Applications of semidefinite programming to combinatorial optimization

We consider the Max Cut problem and the Lovasz's theta function (discussed in the next lecture).

1.1 Max Cut

Consider an undirected graph $(G = (N, E)$, $N = \{1, \ldots, n\}$, with edge weights $w : E \to \mathbb{R}_+$. We desire a cut $\delta(S) = \{ij \in E | i \in S, j \notin S\}$ for some $S \subseteq N$ of maximum weight, where the weight of a cut is defined as

$$w(\delta(S)) = \sum_{e \in \delta(S)} w(e).$$

We can assume this graph is complete by setting $w(e) = 0$ for all edges e that must be added to complete the graph. This problem is known to be NP-hard.

1.2 Quadratic programming formulation

Let x_i be a variable, $i \in \{1, \ldots, n\}$, where $x_i = 1$ if $i \in S$, and $x_i = -1$ if $i \notin S$. Then, since $1 = x_i^2$,

$$w(\delta(S)) = \frac{1}{4} \sum_i \sum_{j \neq i} w_((i,j))(1 - x_i x_j).$$

We define the Laplacian of the graph $\mathcal{L} = \mathcal{L}(G)$ as having entries

$$l_{ij} = \begin{cases} -\frac{1}{4} w_{ij}, & i \neq j \\ \frac{1}{4} \sum_{k \neq i} w_{ik}, & i = j; \end{cases}$$

observe that $\mathcal{L} \in \mathbb{M}^n$. Then we can define a (nonconvex) quadratic program

$$\begin{aligned} \max \quad & \sum_{i,j} l_{ij} x_i x_j \\ \text{s.t.} \quad & x_i^2 = 1, \quad i = 1, \ldots, n. \end{aligned} \tag{QP}$$

(QP) is identical to the Max Cut problem, and so is also NP-hard.

We now consider three ways of relaxing (QP) to a semidefinite program.

1.3 Express (QP) as a linear function of $\mathbf{x}\mathbf{x}^\mathrm{T}$

We first write (QP) as

$$\begin{aligned} \max \quad & \mathcal{L} \cdot (\mathbf{x}\mathbf{x}^\mathrm{T}) \\ \text{s.t.} \quad & \operatorname{diag}(\mathbf{x}\mathbf{x}^\mathrm{T}) = \mathbf{1} \end{aligned}$$

and then re-write $\mathbf{x}\mathbf{x}^{\mathrm{T}}$ as $\mathbf{X}$, yielding

$$\begin{aligned} \max \quad & \mathcal{L} \cdot \mathbf{X} \\ \text{s.t.} \quad & \operatorname{diag}(\mathbf{X}) = e \\ & \mathbf{X} \succeq 0 \\ & \mathbf{rank}(\mathbf{X}) = 1. \end{aligned}$$

The rank-one requirement is to ensure that $\mathbf{X}$ can be expressed as $\mathbf{X} = \mathbf{x}\mathbf{x}^{\mathrm{T}}$:

Prop. $\left\{ \boldsymbol{x}\boldsymbol{x}^T \in \mathbb{M}^n | x_i = \pm 1 \forall i \right\} = \{ \boldsymbol{X} \in \mathbb{M}^n | \boldsymbol{X} \succeq 0, \;\; \operatorname{diag}(\boldsymbol{X}) = \boldsymbol{1}, \;\; \mathbf{rank}(\boldsymbol{X}) = 1 \}.$

Proof. The $\subseteq$ relation is trivial; we now show the $\supseteq$ relation. Suppose that $\mathbf{X} \in \mathbb{M}^n$ with $\operatorname{diag}(\mathbf{X}) = e$, $\mathbf{X} \succ 0$, and $\mathbf{rank}(\mathbf{X}) = 1$. All columns of $\mathbf{X}$ are multiples of the same vector, say $\mathbf{u} \neq 0$, since $\mathbf{X}$ is rank one; likewise, all rows are multiples of $\mathbf{v}^{\mathrm{T}}$. Since $\mathbf{X}$ is symmetric, by considering any nonzero column we can take $\mathbf{v} = \mathbf{u}$. So $\mathbf{X} = \alpha \mathbf{u}\mathbf{u}^{\mathrm{T}}$ for some nonzero $\alpha > 0$, since $\mathbf{X} \succeq 0$. Let $\mathbf{x} = \sqrt{\alpha}\mathbf{u}$ and we have that $\mathbf{X} = \mathbf{x}\mathbf{x}^{\mathrm{T}}$. Finally, since $\operatorname{diag}(\mathbf{X}) = \operatorname{diag}(\mathbf{x}\mathbf{x}^{\mathrm{T}}) = \mathbf{e}$, $x_i = \pm 1$ for all i. □

Observe that (QP) is equivalent to

$$\begin{aligned} \max \quad & \mathcal{L} \cdot \mathbf{X} \\ \text{s.t.} \quad & \operatorname{diag}(\mathbf{X}) = \mathbf{e} \\ & \mathbf{X} \succeq 0 \\ & \mathbf{rank}(\mathbf{X}) = 1, \end{aligned}$$

and so the min-rank SDP problem is NP-hard. With this in mind, we relax the problem by eliminating the rank constraint, yielding

$$\begin{aligned} \max \quad & \mathcal{L} \cdot \mathbf{X} \\ \text{s.t.} \quad & \operatorname{diag}(\mathbf{X}) = \mathbf{e} \\ & \mathbf{X} \succeq 0. \end{aligned} \tag{P}$$

1.4 Increase the dimension of the variables

The second relaxation technique we consider is increasing the dimension of each variable. We replace the variable $x_i = \pm 1$ (a unit vector in $\mathbb{R}^1$) with a unit vector $\mathbf{v}_i \in \mathbb{R}^n$. This yields the program

$$\begin{aligned} \max \quad & \sum_{i,j} l_{ij} \mathbf{v}_i^{\mathrm{T}} \mathbf{v}_j \\ \text{s.t.} \quad & \mathbf{v}_i^{\mathrm{T}} \mathbf{v}_i = 1, \; i = 1, \ldots, n, \\ & \mathbf{v}_i \in \mathbb{R}^n. \end{aligned}$$

Let $\mathbf{V} = [\mathbf{v}_1, \ldots, \mathbf{v}_n]$, and replace $\mathbf{V}^{\mathrm{T}}\mathbf{V}$ with $\mathbf{X} \in \mathbb{M}^n$; this yields (P), as every $\mathbf{X}$ can be so factored.

1.5 The double dual of (QP)

The next technique we consider involves taking the dual of the dual of (QP). The standard Lagrangian dual of (QP) is

$$\min_{\mathbf{y}\in\mathbb{R}^n} \max_{\mathbf{x}\in\mathbb{R}^n} \left(\mathbf{e}^{\mathrm{T}}\mathbf{y} - \mathbf{x}^{\mathrm{T}}\left(\mathrm{Diag}(\mathbf{y}) - \mathcal{L}\right)\mathbf{x}\right).$$

The inner maximization does not exist if $\mathrm{Diag}(\mathbf{y}) - \mathcal{L}$ is not psd, yielding the implied constraint $\mathrm{Diag}(\mathbf{y}) \succeq \mathcal{L}$; if it is psd, then the maximum is attained by choosing $\mathbf{x} = 0$, yielding

$$\begin{array}{rl} \min & \mathbf{e}^{\mathrm{T}}\mathbf{y} \\ \text{s.t.} & \mathrm{Diag}(\mathbf{y}) - \mathcal{L} \succeq 0 \\ & \mathbf{y} \in \mathbb{R}^n, \end{array} \tag{D}$$

first noted by C. Delorme and S. Poljak in 1993. The dual of this sdp is (P).

1.6 Goemans-Williamson '94

Theorem 1. *Let $v(\cdot)$ be the optimal value of a program. Then $v(QP) \geq .878 \cdot v(P)$.*

Proof. We show a randomized algorithm that, in expectation, yields a cut with value at least .878 of the optimal value. Let $\mathbf{X}$ be given that is optimal to (P), and factor $\mathbf{X}$ into $\mathbf{V}^{\mathrm{T}}\mathbf{V}$, $\mathbf{V} = [\mathbf{v}_1, \ldots, \mathbf{v}_n]$. We wish to map each vector into a one-dimensional subspace. Let $\mathbf{r} \in \mathbb{R}^n$ be a random vector drawn uniformly over the unit sphere in $\mathbb{R}^n$. Choose $x_i = \mathbf{sign}(\mathbf{v}_i^T\mathbf{r})$, and let S be the cut defined by $\{x_i\}$; Figure 1 is an example of this rounding scheme.

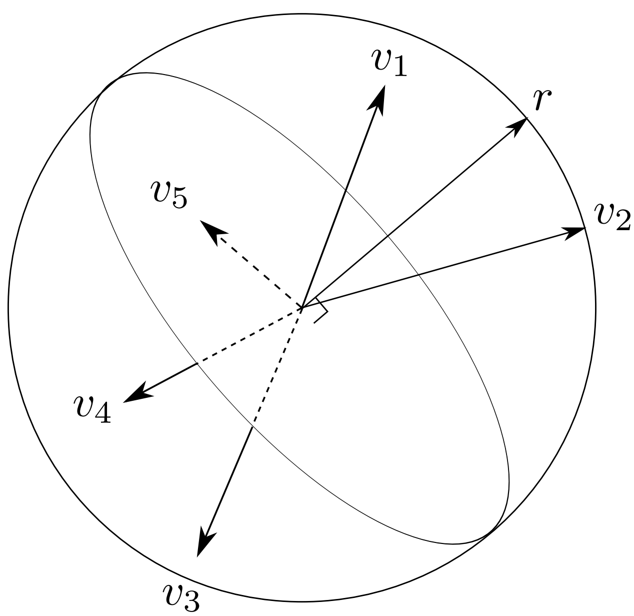

Figure 1: An example in $\mathbb{R}^3$ of the rounding scheme used. The vector $\mathbf{r}$ is chosen uniformly at random from the unit sphere in $\mathbb{R}^n$. The vectors $\mathbf{v}_1$ and $\mathbf{v}_2$ are rounded to 1, while $\mathbf{v}_3$, $\mathbf{v}_4$, and $\mathbf{v}_5$ are rounded to -1. We define the cut S as containing all nodes whose variables are rounded to 1.

We now show that $\mathrm{E}\left[w(\delta(S))\right] \geq .878 \cdot v(P)$. Observe that

$$\mathrm{E}\left[w(\delta(S))\right] = \sum_{(i,j)\in S} w_{ij} p_{ij},$$

where p_{ij} is the probability that ij is in the cut. For a given i, j,

$$p_{ij} = \Pr\left[\mathbf{r}^{\mathrm{T}}\mathbf{v}_i \geq 0 > \mathbf{r}^{\mathrm{T}}\mathbf{v}_j\right] + \Pr\left[\mathbf{r}^{\mathrm{T}}\mathbf{v}_j \geq 0 > \mathbf{r}^{\mathrm{T}}\mathbf{v}_i\right].$$

Consider the plane that $\mathbf{v}_i$ and $\mathbf{v}_j$ lie in, along with the normalized projection of $\mathbf{r}$ in this plane; see figure 2. For these two vectors to be separated by the randomized rounding, the vector $\mathbf{r}$ must lie in the region indicated, which has size $2\arccos(\mathbf{v}_i^{\mathrm{T}}\mathbf{v}_j)$; thus,

$$p_{ij} = \frac{1}{\pi}\arccos(\mathbf{v}_i^T\mathbf{v}_j).$$

By simple calculus, we can see that

$$\frac{1}{\pi}\arccos(\mathbf{v}_i^T\mathbf{v}_j) \geq .878 \cdot \frac{1}{2}(1 - \mathbf{v}_i^{\mathrm{T}}\mathbf{v}_j);$$

then

$$\mathrm{E}\left[w(\delta(S))\right] = \sum_{i<j} w_{ij}p_{ij} = \sum_{i<j} w_{ij}\frac{1}{\pi}\arccos(\mathbf{v}_i^{\mathrm{T}}\mathbf{v}_j) \geq .878 \cdot \frac{1}{2}\sum_{i<j} w_{ij}(1 - \mathbf{v}_i^{\mathrm{T}}\mathbf{v}_j).$$

Recall that

$$v(\mathrm{P}) = \mathcal{L}\cdot\mathbf{X} = \sum_{i,j} l_{ij}\mathbf{v}_i^{\mathrm{T}}\mathbf{v}_j = \frac{1}{2}\sum_{i<j} w_{ij}(1 - \mathbf{v}_i^{\mathrm{T}}\mathbf{v}_j),$$

and so $\mathrm{E}\left[w(\delta(s))\right] \geq .878 \cdot v(\mathrm{P})$, as required. In addition, we have a .878-approximate randomized algorithm.

□

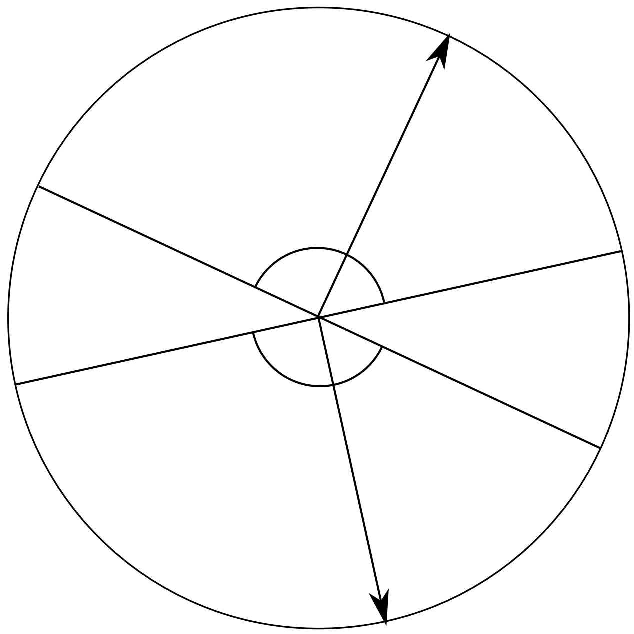

Figure 2: The plane that any two vectors lie in. For these two vectors to be cut by the randomized rounding, it must be that $\mathbf{r}$ lies in the region indicated. If the two vectors are $\mathbf{v}_1$ and $\mathbf{v}_2$, the total size of this region in radians is $2\arccos(\mathbf{v}_1^T\mathbf{v}_2)$.