ADVANCED ALGORITHMS (IV)

CHIHAO ZHANG

1. Goemans and Williamson Rounding of MaxCut

We already meet the problem of MaxCut last time and has been convinced that the integrality gap of an LP relaxation of the problem can be $2-\varepsilon$ for any $\varepsilon > 0$.

MaxCut	
Input:	An undirected graph $G=(V,E)$.
Problem:	Compute a set $S \subseteq V$ that maximizes $\left\lvert E(S,\bar{S})\right\rvert$.

In fact, it is more natural to model the problem as a quadratic programming:

$$\begin{aligned} \max \quad & \frac{1}{2}\sum_{e=\{u,v\}\in E}(1-x_u x_v) \\ \text{s.t.} \quad & x_u \in \{-1,1\}, \quad \forall u \in V. \end{aligned}$$

If we view each number x_u as a one-dimonsional vector $x_u \in \mathbb{R}^1$, and relax it to a n-dimensional vector $\mathbf{w}_u \in \mathbb{R}^n$, the following vector proramming is obtained:

$$\begin{aligned} \max \quad & \frac{1}{2}\sum_{e=\{u,v\}\in E}\left(1-\mathbf{w}_u^T\mathbf{w}_v\right) \\ \text{s.t.} \quad & \mathbf{w}_u \in \mathbb{R}^n, \quad \forall u \in V; \\ & \mathbf{w}_u^T\mathbf{w}_u = 1, \quad \forall u \in V. \end{aligned}$$

We can therefore solve this vector program by transforming it to an equivalent SDP. Let $\left\{\mathbf{w}_u^*\right\}_{u\in V}$ be an optimal solution of the vector program, we now need to round it to a cut, or equivalent a partition $(S,\bar{S})$ of the vertex set V.

Intuitively, if the angle between two vectors $\mathbf{w}_u^*$ and $\mathbf{w}_v^*$ is large, we prefer to cut u and v. The Goemans and Williamson rounding randomly samples a hyperplane crossing the orgin, and therefore separates the vectors into two sets: those on one side of the hyperplane and those on the other side. Therefore, those pairs of vectors with large angle between them are more likely to be separated.

In order to implement this idea, we need to know how to uniformly sample a hyperplane crossing the origin. We achieve this by uniformly sample a point in the $n-1$-sphere $S^{n-1}=\{x\in\mathbb{R}^n : \|x\|=1\}$ to serve as the normal of the plane. We sample a vector $\mathbf{r}=(r_1,\dots,r_n)$ where each $r_i \sim \mathcal{N}(0,1)$ follows the Gaussian distribution with mean 0 and variance 1 independently.

Lemma 1. *$\frac{\mathbf{r}}{\|r\|}$ is a point in S^{n-1} uniformly at random.*

Proof. Consider the probability density function of $\mathbf{r}$. For every $\mathbf{r}=(r_1,\dots,r_n)$, since each r_i is an independent $\mathcal{N}(0,1)$, we have

$$p(r_1,\dots,r_n)=\prod_{i=1}^n \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{r_i^2}{2}\right)=(2\pi)^{-\frac{n}{2}}\exp\left(-\frac{\sum_{i=1}^n r_i^2}{2}\right)=(2\pi)^{-\frac{n}{2}}\exp\left(-\frac{\|\mathbf{r}\|^2}{2}\right).$$

The density only depends on the norm of $\mathbf{r}$, therefore, after normalization, $\frac{\mathbf{r}}{\|r\|}$ is uniform in S^{n-1}. □

The rounding algorithm is

GOEMANS-WILLIAMSON ROUNDING

1. Compute $\left\{\mathbf{w}_u^*\right\}_{u \in V}$.
2. Choose a point $\mathbf{r} = (r_1, \ldots, r_n) \in S^{n-1}$ uniformly at random.
3. Let $S \triangleq \left\{u \in V \;:\; \mathbf{r}^T \mathbf{w}_u^* \geq 0\right\}$.

Theorem 2. *Goemans-Williamson rounding is a randomized α^*-approximation of* MAXCUT *with $\alpha^* > 0.878$.*

Proof. We use $\theta_{u,v}$ denote the angle between vectors $\mathbf{w}_u^*$ and $\mathbf{w}_v^*$, namely $\theta_{u,v} = \arccos \mathbf{w}_\mathbf{u}^{*\,T} \mathbf{w}_v^*$. Since the separating hyperplane is uniformly chosen , for every edge $e = \{u, v\} \in E$, the probability that that u and v are separated (lie on two sides of the hyperplane respectively) is $\frac{\theta}{\pi}$. Therefore, if we let X denote the size of the cut obtained, we have

$$\mathbf{E}\left[X\right] = \sum_{\{u,v\} \in E} \mathbf{Pr}\left[u \text{ and } v \text{ are separated}\right] = \sum_{\{u,v\} \in E} \frac{\arccos \mathbf{w}_\mathbf{u}^{*\,T} \mathbf{w}_v^*}{\pi}. \tag{1}$$

If we let

$$\alpha^* = \min_{-1 \leq x \leq 1} \frac{2 \arccos x}{\pi(1-x)} > 0.878,$$

(1) becomes to

$$\mathbf{E}\left[X\right] \geq \frac{\alpha^*}{2} \sum_{\{u,v\} \in E} (1 - \mathbf{w}_\mathbf{u}^{*\,T} \mathbf{w}_v^*) = \alpha^* \cdot \mathbf{OPT}(VP) \geq \alpha^* \cdot \textsc{MaxCut}(G).$$

□

2. QUADRATIC PROGRAMMING

It is nature to ask whether one can apply Goemans-Williamson rounding to approximate any quadratic programs. In binary quadratic programs, each variable takes value -1 or $+1$:

$$\begin{aligned} \max \quad & \sum_{1 \leq i,j \leq n} a_{i,j} x_i x_j \qquad\qquad \text{(P1)} \\ \text{s.t.} \quad & x_i \in \{-1, +1\}, \quad i = 1, \ldots, n. \end{aligned}$$

In this lecture, we only consider the case that the optimal value of the quadrtic program is *nonnegative*, so that the approximation ratio is well-defined. Therefore, we assume that the coefficient matrix $A = (a_{i,j})_{1 \leq i,j \leq n} \geq 0$. See [WS11, Chapter 13] for a more general treatment.

We can directly apply the Goemans-Williamson relaxation and rounding to quadrtic programs. We relax (P1) to the following vector program:

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

Then apply the following rounding procedure

GOEMANS-WILLIAMSON FOR QUADRATIC PROGRAMMING

1. Compute $\left\{\mathbf{v}_i^*\right\}_{1 \leq i \leq n}$.
2. Pick a vector $\mathbf{r}$ in S^{n-1} uniformly at random.
3. $\hat{x}_i = 1$ if $\mathbf{v}_i^{*\,T} \mathbf{r} \geq 0$; $\hat{x}_i = -1$ otherwise.

For every $i, j \in [n]$, it holds that

$$\mathbf{E}\left[\hat{x}_i \hat{x}_j\right] = (1 - \mathbf{Pr}\left[\hat{x}_i \neq \hat{x}_j\right]) - \mathbf{Pr}\left[\hat{x}_i \neq \hat{x}_j\right] = 1 - 2\mathbf{Pr}\left[\hat{x}_i \neq \hat{x}_j\right].$$

It follows from our argument for MaxCut that the probability of $\hat{x}_i \neq \hat{x}_j$, or equivalently the probability of $\mathbf{v}_i^*$ and $\mathbf{v}_j^*$ being separated by the random hyperplane, is $\frac{\arccos {\mathbf{v}^*}_i^T \mathbf{v}_j^*}{\pi}$. Therefore, we have

$$\mathbf{E}\left[\hat{x}_i \hat{x}_j\right] = 1 - \frac{2}{\pi} \arccos {\mathbf{v}^*}_i^T \mathbf{v}_j^* = \frac{2}{\pi} \arcsin {\mathbf{v}_i^*}^T \mathbf{v}_j^*,$$

where in the last equality we used the identity $\arcsin(x) + \arccos(x) = \frac{\pi}{2}$.

In order to establish a bound for the approximation ratio, we would like to find some α^* so that

$$\sum_{1 \le i,j \le n} a_{i,j} \mathbf{E}\left[\hat{x}_i \hat{x}_j\right] = \sum_{1 \le i,j \le n} a_{i,j} \cdot \frac{2}{\pi} \arcsin {\mathbf{v}_i^*}^T \mathbf{v}_j^* \ge \alpha^* \cdot \sum_{1 \le i,j \le n} a_{i,j} {\mathbf{v}^*}^T \mathbf{v}_j^* \tag{2}$$

holds. Unlike the MaxCut case, it is not enough to simply prove

$$\frac{2}{\pi} \arcsin x \ge \alpha^* \cdot x$$

holds for every $-1 \le x \le 1$, since $a_{i,j}$ might be negative. We shall apply a global argument.

Let $x_{i,j}$ denote ${\mathbf{v}_i^*}^T \mathbf{v}_j$, then it follows from (2) that we want to establish

$$\sum_{1 \le i,j \le n} a_{i,j} \cdot \left(\frac{2}{\pi} \arcsin x_{i,j} - \alpha^* x_{i,j} \right) \ge 0$$

holds for some α^*.

We prove this for $\alpha^* = \frac{2}{\pi}$. Recall that $A = (a_{i,j})_{1 \le i,j \le n}$ is a positive semi-definite matrix and let $Z = (\arcsin x_{i,j} - x_{i,j})_{1 \le i,j \le n}$ be another $n \times n$ matrix. We show the *Frobenius inner product* of A and Z is nonnegative, namely

$$A \bullet Z \ge 0.$$

For every two $n \times n$ matrices $A = (a_{i,j})_{1 \le i,j \le n}$ and $B = (b_{i,j})_{1 \le i,j \le n}$, their *Hadamard product* $C = A \circ B$ is a matrix $C = (c_{i,j})_{1 \le i,j \le n}$ with $c_{i,j} = a_{i,j} \cdot b_{i,j}$. Therefore, if we can show $A \circ Z \ge 0$, then it follows that

$$A \bullet Z = \mathbf{1}^T A \circ Z \mathbf{1} \ge 0,$$

where $\mathbf{1}$ is the all-one vector.

We make use of the following theorem.

Theorem 3 (Schur product theorem). *If $A, B \ge 0$, then $A \circ B \ge 0$.*

Proof. We show for every $x \in \mathbb{R}^n$, $x^T A \circ B x \ge 0$. By definition,

$$x^T A \circ B x = \sum_{i,j} x_i a_{i,j} b_{i,j} x_j = (\operatorname{diag}(x) \cdot A) \bullet (B \cdot \operatorname{diag}(x)),$$

where $\operatorname{diag}(x)$ is a diagonal matrix whose ith entry on the diagonal is x_i. Recall that for a matrix $A = (a_{i,j})_{1 \le i,j \le n}$, its *trace*, denoted by $\operatorname{Tr}(A)$, is defined to be the sum of its diagonal entries, i.e.,

$$\operatorname{Tr}(A) \triangleq \sum_{i=1}^{n} a_{i,i}.$$

It is not hard to verify by definition that for any two matrices A and B, it holds that $A \bullet B = \operatorname{Tr}(AB^T)$. Therefore, we have

$$(\operatorname{diag}(x) \cdot A) \bullet (B \cdot \operatorname{diag}(x)) = \operatorname{Tr}\left(\operatorname{diag}(x) \cdot A \cdot \operatorname{diag}(x) \cdot B\right)$$

Since $A \ge 0$, let $\Lambda = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$ be the diagonal matrix consisting of its eigenvalues, the spectral decomposition theorem says that we can write

$$A = S \Lambda S^{-1}$$

for some matrix S. Let $\sqrt{A} \triangleq S\sqrt{\Lambda}S^{-1}$ where $\sqrt{\Lambda} \triangleq \operatorname{diag}(\sqrt{\lambda_1}, \ldots, \sqrt{\lambda_n})$, then obviously we have $A = \sqrt{A} \cdot \sqrt{A}$. Defining $\sqrt{B}$ similarly, we can continue to write

$$
\begin{aligned}
x^T A \circ Bx &= \operatorname{Tr}\left(\operatorname{diag}(x) \cdot \sqrt{A} \cdot \sqrt{A} \cdot \operatorname{diag}(x) \cdot \sqrt{B} \cdot \sqrt{B}\right) \\
&\overset{①}{=} \operatorname{Tr}\left(\sqrt{A} \cdot \operatorname{diag}(x) \cdot \sqrt{B} \cdot \sqrt{B} \cdot \operatorname{diag}(x) \cdot \sqrt{A}\right) \\
&= \operatorname{Tr}\left(\left(\sqrt{A} \cdot \operatorname{diag}(x) \cdot \sqrt{B}\right) \cdot \left(\sqrt{A} \cdot \operatorname{diag}(x) \cdot \sqrt{B}\right)^T\right) \\
&\overset{②}{\geq} 0,
\end{aligned}
$$

where ① is due to the *cyclic permutation invariance* property of the trace operator, and ② follows from the easy-verified identity $\operatorname{Tr}(MM^T) \geq 0$ for any M. □

Given Schur product theorem, we only need to verify that $Z \succeq 0$. The taylor series of the function $\arcsin x$ is

$$\arcsin x = x + \sum_{n \geq 1} \frac{\prod_{i=1}^{n+1}(2i-1)}{\prod_{i=1}^{n} 2i} \frac{x^{2n+1}}{2n+1}.$$

Therefore, if we let $X = (x_{i,j})_{1 \leq i,j \leq n}$, it holds that

$$Z = \sum \sum_{n \geq 1} \frac{\prod_{i=1}^{n+1}(2i-1)}{\prod_{i=1}^{n} 2i} \frac{X^{(2n+1)}}{2n+1},$$

where $X^{(k)} \triangleq \underbrace{X \circ X \circ \cdots \circ X}_{k \text{ consecutive } \circ}$. Then $Z \succeq 0$ follows from Schur product theorem and the fact that the sum of positive semi-definite matrices is positive semi-definite. So we can conclude with the following theorem.

Theorem 4. *Goemans-Williamson rounding is a randomized $\frac{2}{\pi}$-approximation of binary quadratic programming when the coefficient matrix is positive semi-definite.*

Remark. We can add arbitrary quadratic constriants to (P1). The approximation ratio of Goemans-Williamson is at least as good as the non-constraint case, since the same analysis applies (and we may have chance to improve it under some contraints).

3. Correlation Clustering

The last problem today is called "Correlation Clustering". Given a undirected graph $G = (V, E)$ in which each edge $e \in E$ has two nonnegative weights $w_e^+, w_e^- \geq 0$.The problem is to find a parition $\mathcal{S} = (S_1, \ldots, S_k)$ (clustering) of V such that

$$\sum_{e \in E^+(\mathcal{S})} w_e^+ + \sum_{e \in E^-(\mathcal{S})} w_e^-$$

is maximized, where $E^+(\mathcal{S})$ (resp. $E^-(\mathcal{S})$) consists of edges whose two ends are in the same (resp. different) S_i.

To relax the problem, we first formalize it into a vector program. Let $n = |V|$. For every $i \in [n]$, we use $\mathbf{e}_i$ to denote the i-th unit vector, i.e., the vector whose i-th entry is one and all other entries are zero. Then the correlation clustering problem is equivalent to

$$
\begin{aligned}
\max \quad & \sum_{\{u,v\} \in E} \left(w_{u,v}^+(\mathbf{x}_u^T \mathbf{x}_v) + w_{u,v}^-(1 - \mathbf{x}_u^T \mathbf{x}_v)\right) \\
\text{s.t.} \quad & \mathbf{x}_u \in \{e_1, \ldots, e_n\}, \quad \forall u \in V.
\end{aligned}
$$

We relax the domain of $\mathbf{x}_u$ to unit vectors in $\mathbb{R}^n$:

$$\begin{aligned} \max \quad & \sum_{\{u,v\}\in E} \left(w^+_{u,v}(\mathbf{x}_u^T\mathbf{x}_v) + w^-_{u,v}(1-\mathbf{x}_u^T\mathbf{x}_v) \right) \\ \text{s.t.} \quad & \mathbf{x}_v^T\mathbf{x}_v = 1, \quad \forall v \in V; \\ & \mathbf{x}_u^T\mathbf{x}_v \geq 0, \quad \forall u,v \in V; \\ & \mathbf{x}_u \in \mathbb{R}^n, \quad \forall u \in V. \end{aligned}$$

Note that we have additional constraints of the form $\mathbf{x}_u^T\mathbf{x}_v \geq 0$. We can safely add these constraints since they are obviously satisfied by $\mathbf{x}_u \in \{\mathbf{e}_i\}_{i\in[n]}$. They are useful in our analysis since we only need to argue in the domain where these constraints are satisfied.

We shall obtain a 0.75-approximation by using at most four clusters. The four clusters are formed by using two random hyperplanes crossing the origin, in the Goemans-Williamson manner.

Correlation Clustering

1. Compute $\{\mathbf{x}_u^*\}_{u\in V}$.
2. Pick two vectors $\mathbf{r}_1$ and $\mathbf{r}_2$ in S^{n-1} uniformly at random.
3. We let
   - $R_1 = \left\{u \in V \,:\, (\mathbf{x}_u^*)^T\mathbf{r}_1 \geq 0 \text{ and } (\mathbf{x}_u^*)^T\mathbf{r}_2 \geq 0\right\}$,
   - $R_2 = \left\{u \in V \,:\, (\mathbf{x}_u^*)^T\mathbf{r}_1 \geq 0 \text{ and } (\mathbf{x}_u^*)^T\mathbf{r}_2 < 0\right\}$,
   - $R_3 = \left\{u \in V \,:\, (\mathbf{x}_u^*)^T\mathbf{r}_1 < 0 \text{ and } (\mathbf{x}_u^*)^T\mathbf{r}_2 \geq 0\right\}$,
   - $R_4 = \left\{u \in V \,:\, (\mathbf{x}_u^*)^T\mathbf{r}_1 < 0 \text{ and } (\mathbf{x}_u^*)^T\mathbf{r}_2 < 0\right\}$.

Then for every $\{u,v\} \in E$, if we use $q_{u,v}$ to denote the probability that $\mathbf{x}_u^*$ and $\mathbf{x}_v^*$ are on the same side of both hyperplanes, the expected cost of our rounding algorithm is

$$\sum_{\{u,v\}\in E} w^+_{u,v} q_{u,v} + w^-_{u,v}(1-q_{u,v}).$$

Moreover,

$$q_{u,v} = \left(1 - \frac{\arccos {\mathbf{x}_u^*}^T\mathbf{x}_v^*}{\pi}\right)^2 .$$

So if we let

$$\alpha_1 = \min_{0\leq z\leq 1} \left(1-\frac{\arccos z}{\pi}\right)^2 \Big/ z\,, \quad \alpha_2 = \min_{0\leq z\leq 1} \left(1-\left(1-\frac{\arccos z}{\pi}\right)^2\right)\Big/ (1-z)\,,$$

and $\alpha^* = \min\{\alpha_1,\alpha_2\} = 0.75$, it holds that the expected cost of our rounding algorithm is at least

$$\alpha^* \cdot \sum_{\{u,v\}\in E} \left(w^+_{u,v}({\mathbf{x}_u^*}^T\mathbf{x}_v) + w^-_{u,v}(1-{\mathbf{x}_u^*}^T\mathbf{x}_v^*) \right) \geq \alpha^* \cdot \mathbf{OPT}.$$

4. Remark

The presentation of this lecture mainly follows [WS11, Chapter 6].

References

[WS11] David P Williamson and David B Shmoys. *The design of approximation algorithms*. Cambridge university press, 2011. 2, 5