9 Binary quadratic optimisation (continued)

Recall the maximum cut problem:

$$\begin{array}{ll} \text{maximise} & x^T L_G x \\ \text{subject to} & x_i \in \{-1,1\}, \; i = 1,\dots,n. \end{array} \tag{MC}$$

where L_G is the Laplacian quadratic form associated with a weighted graph G:

$$x^T L_G x = \frac{1}{2} \sum_{i,j \in V} w_{ij} (x_i - x_j)^2.$$

More generally for any matrix $Y \in \mathbf{S}^n$ we have

$$\operatorname{Tr}(L_G Y) = \frac{1}{2} \sum_{i,j \in V} w_{ij} (Y_{ii} + Y_{jj} - 2Y_{ij}). \tag{1}$$

We introduced the following semidefinite relaxation of (MC) in the last lecture.

$$\begin{array}{ll} \underset{X \in \mathbf{S}^n}{\text{maximise}} & \operatorname{Tr}(L_G X) \\ \text{subject to} & X \succeq 0 \\ & X_{ii} = 1 \; (i = 1,\dots,n) \end{array} \tag{SDP}$$

Let v^* be the optimal value of (MC) and p^*_{SDP} be the optimal value of (SDP). We already saw that $p^*_{SDP} \geq v^*$. In today's lecture we prove the following result due to Goemans and Williamson:

Theorem 9.1 (Goemans-Williamson, [GW95])**.** *Let v^* be the optimal value of (MC) and let p^*_{SDP} be the optimal value of (SDP). Then*

$$\alpha \cdot p^*_{SDP} \leq v^* \leq p^*_{SDP} \tag{2}$$

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

Proof. We have already proved the inequality (2) $v^* \leq p^*_{SDP}$ last lecture: if $x \in \{-1,1\}^n$ then letting $X = xx^T$ we see that X is feasible for the SDP (SDP) and $\operatorname{Tr}(L_G X) = x^T L_G x$.

The main part of the proof is to show the inequality $\alpha p^*_{SDP} \leq v^*$. For this we will use a technique called *randomised rounding*. Let X be a solution of (SDP). Since $X \succeq 0$ we can write $X = V^T V$ where $V \in \mathbb{R}^{r \times n}$, or in other words $X_{ij} = \langle v_i, v_j \rangle$ where $v_i \in \mathbb{R}^r$ and $r = \operatorname{rank}(X)$. Since $X_{ii} = 1$ we know that $\|v_i\| = 1$. We are now going to see a way to use the vectors $v_1, \dots, v_n$ to produce a random vector $x \in \{-1,1\}^n$ whose covariance matrix will be "close to" X. The random vector x is defined by:

$$x_i = \operatorname{sign}(\langle v_i, z \rangle), \quad i = 1,\dots,n. \tag{3}$$

where z is a standard Gaussian random vector in $\mathbb{R}^r$. It is not difficult to verify that $\mathbb{E}[x_i] = 0$. The next lemma computes the covariance matrix of x:

Lemma 1. *For the random variables $x_1, \dots, x_n$ defined in (3) we have $\mathbb{E}[x_i x_j] = 1 - \frac{2}{\pi} \arccos(X_{ij})$.*

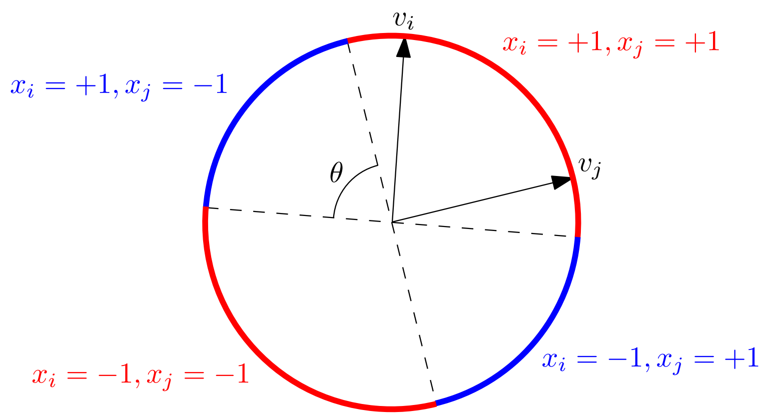

Figure 1: Computation of $\mathbb{E}[x_i x_j]$ for x defined in (3). Let $\theta = \arccos(\langle v_i, v_j \rangle)$ be the angle between v_i and v_j. The probability of having $x_i x_j = -1$ is $2\theta/2\pi$ and the probability of having $x_i x_j = +1$ is $(2\pi - 2\theta)/2\pi$.

Proof. The proof of this lemma is summarised in Figure 1. First note that the value of the pair $(\langle v_i, z\rangle, \langle v_j, z\rangle)$ only depends on the orthogonal projection of z on the subspace $\text{span}(v_i, v_j)$. Since z is standard Gaussian we know its orthogonal projection on $\text{span}(v_i, v_j)$ is distributed like a standard Gaussian vector on that two-dimensional subspace. In Figure 1 we represent vectors v_i and v_j in that subspace. Since the standard Gaussian distribution is rotation-invariant we see that the probability of having $x_i x_j = -1$ (blue region in the figure) is $2\theta/2\pi$ and the probability of having $x_i x_j = +1$ (red region in the figure) is $(2\pi - 2\theta)/2\pi$. Thus the expected value of $x_i x_j$ is given by:

$$\mathbb{E}[x_i x_j] = (-1) \cdot (2\theta/2\pi) + (+1) \cdot (1 - 2\theta/2\pi) = 1 - \frac{2}{\pi}\theta.$$

Since $\theta = \arccos(\langle v_i, v_j \rangle) = \arccos(X_{ij})$ we get the desired formula. □

To summarize: from the solution $X \in \mathbf{S}^n$ of (SDP), we constructed a random vector x in $\{-1, 1\}^n$ (defined in (3)) that satisfies $\mathbb{E}[x] = 0$ and whose covariance matrix $\Sigma = \mathbb{E}[xx^T]$ is given by

$$\Sigma_{ij} = f(X_{ij}) \tag{4}$$

where

$$f(t) = 1 - \frac{2}{\pi}\arccos(t). \tag{5}$$

Figure 2 shows the plot of $f(t)$. Qualitatively, we see that $f(t)$ is not too far from t and so the entries of Σ are not too far from the entries of X. Remember we know that

$$v^* \geq \mathbb{E}[x^T L_G x] = \text{Tr}(L_G \Sigma).$$

(The inequality $v^* \geq \mathbb{E}[x^T L_G x]$ simply comes by taking expectations in the inequality $v^* \geq x^T L_G x$ which holds with probability 1 by definition of v^*.) Now, it is reasonable to expect since Σ is not too far off from X, that we can relate $\text{Tr}(L_G \Sigma)$ to $\text{Tr}(L_G X) = p^*_{SDP}$. Indeed it is not very difficult to do this here. Define:

$$\alpha = \min_{t \in [-1,1)} \frac{1 - f(t)}{1 - t} \approx 0.878. \tag{6}$$

The constant α measures in some sense how much you have to tilt the line $y = t$ in Figure 2 so that it lies above the curve of f, while keeping the point $(t = 1, y = 1)$ fixed. Then we can show:

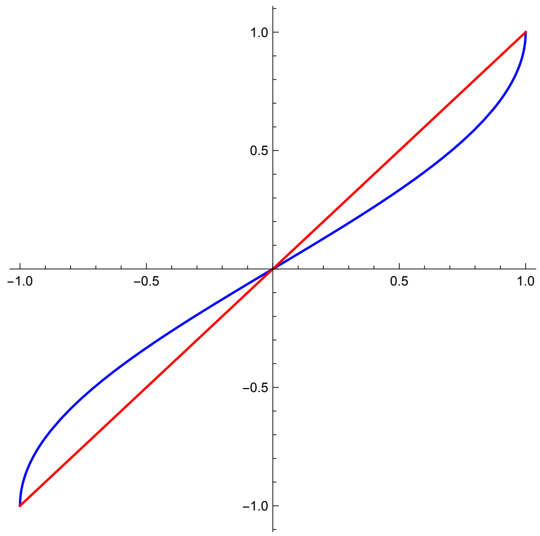

Figure 2: Plot of $f(t)$ given by (5).

Claim 9.1. *With* Σ *defined in* (4) *and* α *in* (6) *we have* $\mathrm{Tr}(L_G\Sigma) \geq \alpha\,\mathrm{Tr}(L_G X)$.

Proof. From the definition of L_G (see (1)) and since $\Sigma_{ii} = X_{ii} = 1$ for $i = 1, \ldots, n$ we have:

$$\mathrm{Tr}(L_G\Sigma) = \sum_{i,j\in V} w_{ij}(1-\Sigma_{ij}) = \sum_{i,j\in V} w_{ij}(1-f(X_{ij})) \overset{(*)}{\geq} \alpha \sum_{i,j\in V} w_{ij}(1-X_{ij}) = \alpha\,\mathrm{Tr}(L_G X)$$

where in $(*)$ we used that $w_{ij} \geq 0$. □

The proof of the theorem is now complete since we showed

$$v^* \geq \mathrm{Tr}(L_G\Sigma) \geq \alpha\,\mathrm{Tr}(L_G X) = \alpha p^*_{SDP}.$$

□

Stable set problem

We now look at another application of semidefinite optimisation to combinatorial optimisation, namely to the maximum stable set problem.

Stable set Let $G = (V, E)$ be an undirected graph. A *stable set* (also known as an *independent set*) in G is a subset $S \subseteq V$ such that no two vertices in S are connected by an edge, i.e., $i, j \in S \Rightarrow \{i, j\} \notin E$. The *maximum stable set problem* is the problem of finding the largest stable set in a graph. The stable set problem can be formulated as the following problem:

$$\begin{array}{lll} \underset{x\in\mathbb{R}^n, X\in\mathbf{S}^n}{\text{maximise}} & \displaystyle\sum_{i=1}^n x_i & \\ \text{subject to} & x_i^2 = x_i & \forall i \in V = \{1,\ldots,n\} \\ & x_i x_j = 0 & \forall ij \in E. \end{array} \tag{7}$$

The constraint $x_i^2 = x_i$ is equivalent to saying that $x_i \in \{0, 1\}$ and the stable set S corresponds to the set of i such that $x_i = 1$. Note that the constraint $x_i x_j = 0$ ensures that S is a stable set. The objective function $\sum_{i=1}^n x_i$ counts the cardinality of S. Solving the optimisation problem (7) is computationally hard in general.

Semidefinite relaxation We are now going to define a semidefinite relaxation for (7). This relaxation was first proposed by Lovász in [Lov79]. It allows us to get an upper bound on the solution (7) by solving a semidefinite program.

$$\begin{array}{lll} \underset{x\in\mathbb{R}^n, X\in\mathbf{S}^n}{\text{maximise}} & \sum_{i=1}^n x_i & \\ \text{subject to} & X_{ii} = x_i & i \in V \\ & X_{ij} = 0 & ij \in E \\ & \begin{bmatrix} 1 & x^T \\ x & X \end{bmatrix} \succeq 0 & \end{array} \tag{8}$$

Problem (8) can be solved efficiently using algorithms for semidefinite programming. The next theorem shows that (8) yields an upper bound on (7).

Theorem 9.2. *Let $\alpha(G)$ be the solution of (7) and $\vartheta(G)$ be the solution of (8). Then $\alpha(G) \leq \vartheta(G)$.*

Proof. It suffices to observe that if x is feasible for (7), then the pair $(x, X = xx^T)$ is feasible for (8) since

$$\begin{bmatrix} 1 & x^T \\ x & xx^T \end{bmatrix} = \begin{bmatrix} 1 \\ x \end{bmatrix} \begin{bmatrix} 1 \\ x \end{bmatrix}^T \succeq 0.$$

□

A natural question is to ask whether there is a constant $c > 0$ such that $c \cdot \vartheta(G) \leq \alpha(G)$ for all graphs G. Unfortunately this is not the case. Indeed one can show:

Theorem 9.3. *There exists a sequence of graphs (G_n) such that $\alpha(G_n)/\vartheta(G_n) \to 0$ as $n \to \infty$.*

Proof. See Exercise 9.1. □

Exercise 9.1 (Proof of Theorem 9.3, see [BTN01, p. 169])**.** *In this exercise we will prove Theorem 9.3. In fact we will show something more precise. We will prove that for any large enough n there is a graph G on n nodes such that $\alpha(G)/\vartheta(G) \leq O(\log(n)/\sqrt{n})$ as $n \to \infty$.*

1. *Show that the dual of (8) can be expressed as*

$$\begin{array}{lll} \text{min.} & Z_{00} & \\ \text{s.t.} & z_i = (1 + Z_{ii})/2 & \forall i \in V \\ & Z_{ij} = 0 & \forall \{i,j\} \in \overline{E} \\ & \begin{bmatrix} Z_{00} & z^T \\ z & Z \end{bmatrix} \succeq 0 & \end{array} \tag{9}$$

where $\overline{E} = \{\{i,j\} : i \neq j \text{ and } \{i,j\} \notin E\}$ is the complement of E [hint: you may need to use the fact that $\left[\begin{smallmatrix} A & B^T \\ B & C \end{smallmatrix}\right] \succeq 0 \iff \left[\begin{smallmatrix} A & -B^T \\ -B & C \end{smallmatrix}\right] \succeq 0$].

2. *Show that* (9) *can be simplified to:*

$$\begin{array}{lll} \textit{min.} & Z_{00} & \\ \textit{s.t.} & Z_{ii} = 1 & \forall i \in V \\ & Z_{ij} = 0 & ij \in \overline{E} \\ & \begin{bmatrix} Z_{00} & \mathbf{1}^T \\ \mathbf{1} & Z \end{bmatrix} \succeq 0 & \end{array} \tag{10}$$

where $\mathbf{1}$ *denotes the vector with all ones [hint: given* (z, Z) *feasible for* (9)*, consider* $\tilde{Z}_{ij} = Z_{ij}/(z_i z_j)$*].*

3. *Use Slater condition to verify that* (10) *and* (8) *have the same optimal values.*

4. *Show that for any graph* G *with* n *vertices we have* $\vartheta(G)\vartheta(\bar{G}) \geq n$ *where* $\bar{G} = (V, \bar{E})$ *[hint: use the minimisation formulation* (10) *of* $\vartheta(G)$ *to construct a feasible point for* (8) *applied to* $\bar{G}$*]. Deduce that for any graph* G *we have either* $\vartheta(G) \geq \sqrt{n}$ *or* $\vartheta(\bar{G}) \geq \sqrt{n}$*.*

5. *We are now going to assume the following fact: for any* n *large enough (i.e.,* $n \geq N_0$ *for some* N_0*) there is a graph* G *on* n *vertices such that* $\max(\alpha(G), \alpha(\bar{G})) \leq 3\log(n)$*. Using this fact together with question 4, prove the desired result.*

Note: One way to prove the existence of a graph such that $\max(\alpha(G), \alpha(\bar{G})) \leq 3\log(n)$ *is using the probabilistic method. If we let* G *be a random undirected graph on* $V = \{1, \ldots, n\}$ *where we draw an edge between a pair* $\{i, j\} \subset V$ *with probability 1/2 (independently of the other pairs) a well-known result states that* $\alpha(G)$ *concentrates around* $2\log(n)$*.*

References

[BTN01] Aharon Ben-Tal and Arkadi Nemirovski. *Lectures on modern convex optimization: analysis, algorithms, and engineering applications.* SIAM, 2001. 4

[GW95] Michel X. Goemans and David P. Williamson. Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming. *Journal of the ACM*, 42(6):1115–1145, 1995. 1

[Lov79] László Lovász. On the Shannon capacity of a graph. *IEEE Transactions on Information Theory*, 25(1):1–7, 1979. 4