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

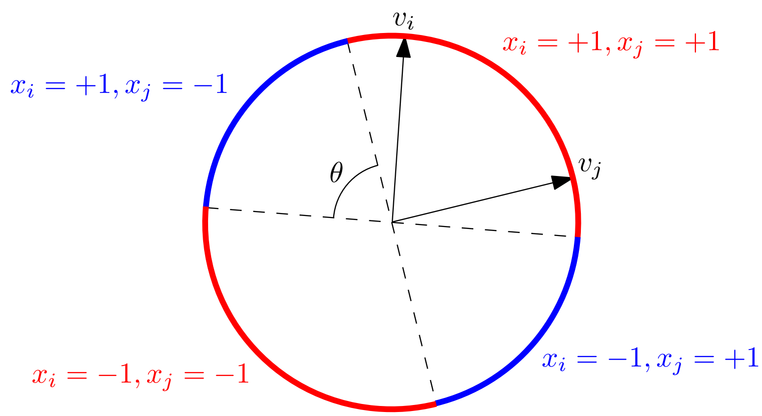

Figure 1: Computation of $\mathbb{E}[x_i x_j]$ for x defined in (3). Let $\theta = \arccos(\langle v_i, v_j \rangle)$ be the angle between v_i and v_j. The probability of having $x_i x_j = -1$ is $2\theta/2\pi$ and the probability of having $x_i x_j = +1$ is $(2\pi - 2\theta)/2\pi$.

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:

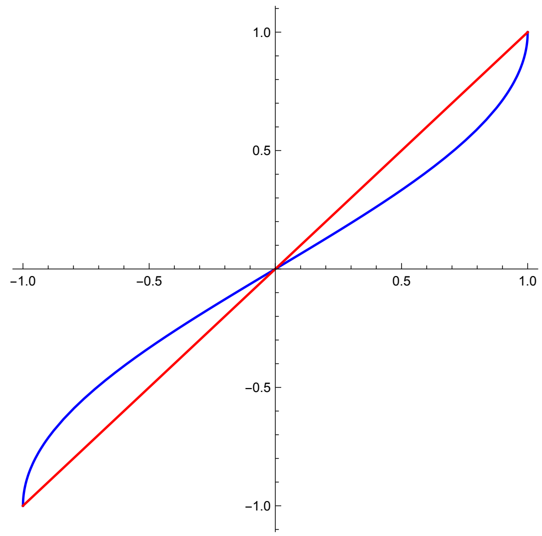

Figure 2: Plot of $f(t)$ given by (5).

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

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

□

Exercise 9.1 (Diagonally dominant matrices)**.** *Recall that a $n\times n$ matrix M is called diagonally dominant if $M_{ii} \geq \sum_{j\neq i} |M_{ij}|$ for all $i = 1, \ldots, n$.*

1. *Show that the matrix L_G is diagonally dominant.*
2. *Let $K \subset \mathbf{S}^n$ be the cone of symmetric diagonally dominant matrices. Show that the extreme rays of K are spanned by the matrices*

$$e_i e_i^T \quad (i = 1, \ldots, n) \qquad \text{and} \qquad (e_i \pm e_j)(e_i \pm e_j)^T \quad (1 \leq i < j \leq n).$$

   where $e_i \in \mathbb{R}^n$ is the vector with 1 in the i'th component and 0 elsewhere.

3. *Show that Theorem 9.1 still holds if in (MC) and (SDP) we replace L_G by any symmetric diagonally dominant matrix.*

Exercise 9.2 (Implementation)**.** *Implement the randomised rounding scheme of the proof of Theorem 9.1: choose a graph G, solve (SDP) (e.g., using CVX - `http://www.cvxr.com/cvx/`), sample points from (3) and plot the distribution of $x^T L_G x$.*

References

[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