

Lecture 11: SDP Relaxation and Randomized Methods for 0-1 Quadratic Program

(3 units)

Outline

- ▶ SDP problem
- ▶ Binary quadratic program and maximum cut problem
- ▶ SDP relaxation via Lagrangian dual
- ▶ SDP relaxation via lifting and rank relaxation
- ▶ Goemans and Williamson's bound and randomized scheme
- ▶ Nesterov's SDP bound

SDP Problem

- ▶ Example of Linear program and semidefinite program (SDP):

$$\begin{array}{ll} (LP) & \min 2x_1 + x_2 + x_3 \\ & \text{s.t. } x_1 + x_2 + x_3 = 1 \\ & \quad (x_1, x_2, x_3) \geq 0. \end{array}$$

$$\begin{array}{ll} (SDP) & \min 2x_1 + x_2 + x_3 \\ & \text{s.t. } x_1 + x_2 + x_3 = 1 \\ & \quad \begin{pmatrix} x_1 & x_2 \\ x_2 & x_3 \end{pmatrix} \succeq 0. \end{array}$$

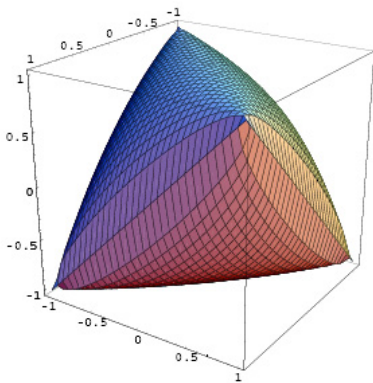


Figure: Set of 3×3 positive semidefinite matrices with unit diagonal

General form of SDP problem

- ▶ General form of SDP problem:

$$\begin{aligned} (SDP) \quad & \min C \bullet X \\ & \text{s.t. } A_i \bullet X = b_i, \quad i = 1, \dots, m, \\ & \quad X \succeq 0, \end{aligned}$$

where C, A_i are given $n \times n$ symmetric and b_i s are given scalars, and

$$A \bullet X = \sum_{i=1}^n \sum_{j=1}^n a_{ij} x_{ij} = \text{Tr}(A^T x).$$

- ▶ The dual of (SDP) is

$$\begin{aligned} (SDD) \quad & \max b^T y \\ & \text{s.t. } \sum_{i=1}^m y_i A_i \preceq C, \end{aligned}$$

where $y \in \mathbb{R}^m$.

- Or equivalently

$$\begin{aligned} (SDD) \quad & \min b^T y \\ & \text{s.t.} \quad \sum_{i=1}^m y_i A_i + S = C, \\ & \quad S \succeq 0. \end{aligned}$$

- **Strong duality:** $v(SDP) = v(SDD)$ if (SDP) or (SDD) is strictly feasible.
- The **SDP interior-point algorithm** finds an ϵ -approximate solution where solution time is linear in $\log(1/\epsilon)$ and polynomial in m and n .
- SDP solvers and software based on MatLab: **SeDuMi**, **CVX** ($n \leq 1000$).

Binary quadratic optimization and max-cut problem

- ▶ Binary quadratic optimization:

$$\begin{aligned} \min x^T Q x \\ \text{s.t. } x \in \{-1, 1\}^n. \end{aligned}$$

This is a well known **NP-hard** problem. It is NP-hard even if the matrix Q is positive definite, since $x^T Q x = x^T (Q + \text{diag}(\lambda)) x - e^T \lambda$.

- ▶ The Boolean constraints can be expressed using quadratic equations:

$$x_i^2 - 1 = 0 \Leftrightarrow x_i \in \{-1, 1\}.$$

- ▶ An equivalent problem:

$$\begin{aligned} \min x^T Q x \\ \text{s.t. } x_i^2 = 1, \quad i = 1, \dots, n. \end{aligned}$$

- ▶ Example. The **maximum cut** (MAXCUT) problem is to find a partition of the nodes of a graph $G = (V, E)$ into two disjoint sets V_1 and V_2 ($V_1 \cap V_2 = \emptyset$, $V_1 \cup V_2 = V$) in such a way to maximize the weights of edges that have one endpoint in V_1 and the other in V_2 . Let w_{ij} be the weight corresponding to the (i, j) edge, and is zero if the nodes i and j are not connected.
- ▶ Define:

$$y_i = -1 \Leftrightarrow i \in V_1, \quad y_i = 1 \Leftrightarrow i \in V_2.$$

The weight of the cut defined by $y \in \{-1, 1\}^n$ is:

$$\frac{1}{4} \sum_{i,j=1}^n w_{ij}(1 - y_i y_j)$$

$$(i \in V_1, j \in V_2 \Rightarrow w_{ij}(1 - y_i y_j) = 2w_{ij}, \\ i, j \in V_1 \text{ or } V_2 \Rightarrow w_{ij}(1 - y_i y_j) = 0).$$

- ▶ MAXCUT problem can be written as:

$$\begin{aligned} \max \quad & \frac{1}{4} \sum_{i,j} w_{ij} (1 - y_i y_j) \\ \text{s.t.} \quad & y_i \in \{-1, 1\}, \quad i = 1, \dots, n, \end{aligned}$$

- ▶ The maximum cut problem is clearly equivalent to:

$$\min \sum_{i,j} w_{ij} y_i y_j, \quad \text{s.t. } y_i \in \{-1, 1\}, \quad i = 1, \dots, n.$$

SDP relaxation via Lagrangian dual

- ▶ **Lagrangian duality** A general approach to obtain lower bounds on the value of general (non)convex minimization problems is to use Lagrangian duality. The original Boolean minimization problem can be written as:

$$(P) \quad \min x^T Q x \\ \text{s.t. } x_i^2 = 1, \quad i = 1, \dots, n.$$

- ▶ The Lagrangian function can be written as:

$$L(x, \lambda) = x^T Q x - \sum_{i=1} \lambda_i (x_i^2 - 1) = x^T (Q - \Lambda) x + \text{Tr}(\Lambda).$$

- ▶ The Lagrangian relaxation of (P) is

$$d(\lambda) = \min_{x \in \mathbb{R}^n} L(x, \lambda) = x^T (Q - \Lambda) x + \text{Tr}(\Lambda).$$

The dual of (P) is

$$\max d(\lambda) \\ \text{s.t. } \lambda \in \mathbb{R}^n.$$

- ▶ $d(\lambda) > -\infty \Leftrightarrow$
 - (i) $Q - \lambda \succeq 0$;
 - (ii) $\exists \bar{x} \in \mathbb{R}^n$ such that $(Q - \Lambda)\bar{x} = 0$.
- ▶ **Proof.** $d(\lambda) > -\infty \Rightarrow Q - \lambda \succeq 0$. If, otherwise, there exists $\tilde{x} \neq 0$ such that $\tilde{x}^T (Q - \Lambda)\tilde{x} < 0$, then

$$L(t\tilde{x}, \mu) = \frac{1}{2}t^2\tilde{x}^T(Q - \Lambda)\tilde{x} + Tr(\Lambda) \rightarrow -\infty, \quad t \rightarrow +\infty.$$

$Q - \Lambda \succeq 0$ and the KKT necessary condition $\Rightarrow (Q - \Lambda)\bar{x} = 0$.

- ▶ Conversely, if conditions (i)-(ii) hold, then $L(x, \lambda)$ is a convex function and \bar{x} satisfies the KKT sufficient condition. Thus,
 $d(\lambda) = L(\bar{x}, \lambda) = \text{Tr}(\Lambda)$
- ▶ $d(\lambda) > -\infty \Rightarrow d(\lambda) = \text{Tr}(\Lambda)$.
- ▶ The dual problem can be written as an SDP problem:

$$\begin{aligned} \max \quad & \text{Tr}(\Lambda) \\ \text{s.t.} \quad & Q \succeq \Lambda, \\ & \Lambda \text{ diagonal,} \end{aligned}$$

or

$$\begin{aligned} \max \quad & e^T \lambda \\ \text{s.t.} \quad & Q - \text{diag}(\lambda) \succeq 0. \end{aligned}$$

Lifting and rank relaxation

- ▶ Let $X = xx^T$ for $x \in \{-1, 1\}^n$. Then $X \succeq 0$, $X_{ii} = 1$, and X has rank one. Conversely, any matrix X with $X \succeq 0$, $X_{ii} = 1$, $\text{rank} X = 1$, then $X = xx^T$ for some $x \in \{-1, 1\}^n$.
- ▶ $x^T Q x = \text{Tr}(x^T Q x) = \text{Tr}(Q x x^T) = \text{Tr}(Q X)$.
- ▶ As a consequence, the original problem can be exactly rewritten as:

$$\begin{aligned} \min \quad & \text{Tr}(QX) \\ \text{s.t.} \quad & X_{ii} = 1, \quad i = 1, \dots, n \\ & X \succeq 0, \quad \text{rank}(X) = 1. \end{aligned}$$

- ▶ For any $X \succeq 0$ and $\text{diag}(X) = e$,

$$\text{rank}(X) = 1 \Leftrightarrow \|X\| = n.$$

- **Proof.** By definition

$$\|X\|^2 = \text{Tr}(X^2) = \sum_{i=1}^n \lambda_i(X)^2 = \|\lambda(X)\|^2.$$

For any $x \in \mathbb{R}^n$,

$$\|x\| = \sqrt{\sum_{i=1}^n x_i^2} \leq \sum_{i=1}^n |x_i| = \|x\|_1$$

with equality if and only if there is at most one **nonzero** x_i . So

$$\|X\| \leq \sum_{i=1}^n |\lambda_i(X)| = n$$

with equality if and only $\text{rank}(X) = 1$.

- Dropping the rank one constraint results in a **SDP** relaxation

$$\begin{aligned} \min \quad & \text{Tr}(QX) \\ \text{s.t.} \quad & X_{ii} = 1, \quad i = 1, \dots, n \\ & X \succeq 0. \end{aligned}$$

- ▶ It SDP dual is:

$$\begin{aligned} \max \quad & e^T \lambda \\ \text{s.t.} \quad & Q - \text{diag}(\lambda) \succeq 0. \end{aligned}$$

- ▶ A useful interpretation is in terms of a nonlinear **lifting** to a higher dimensional space. Indeed, rather than solving the original problem in terms of the n -dimensional vector x , we are instead solving for the $n \times n$ **matrix** X , effectively converting the problem from R^n to S^n , which has dimension $\binom{n+1}{2}$.
- ▶ If we find an optimal solution X of the SDP that has **rank one** \Rightarrow the original problem is solved.
- ▶ In general, it is **not the case** that the optimal solution of the SDP relaxation will be rank one. However, it is possible to use **rounding schemes** to obtain nearby **rank one** solutions. Furthermore, in some cases, it is possible to do so while obtaining some approximation guarantees on the quality of the rounded solutions.

Goemans and Williamson's bound

- ▶ **Basic questions:**
 - ▶ **Approximation guarantees:** is it possible to prove general properties on the quality of the bounds obtained by SDP?
 - ▶ **Feasible solutions:** can we (somehow) use the SDP relaxations to provide not just bounds, but actual **feasible points** with good (or optimal) values of the objective?
- ▶ In their celebrated MAXCUT paper (JACM, 1995), Goemans and Williamson developed the following randomized method for finding a **good feasible cut** from the solution of the SDP:
 - ▶ **Factorize** X as $X = V^T V$, where $V = [v_1, \dots, v_n] \in \mathbb{R}^{r \times n}$, where r is the rank of X .
 - ▶ Then $X_{ij} = v_i^T v_j$, and since $X_{ii} = 1$ this factorization gives n vectors v_i on the unit sphere in \mathbb{R}^r .
 - ▶ Now, choose a **random hyperplane** (passing through the origin) in \mathbb{R}^r , and assign to each variable x_i either $+1$ or -1 , depending on which side of the hyperplane the point v_i lies.

- ▶ It turns out that this procedure gives a solution that, on average, is quite close to the value of the SDP bound. The random hyperplane can be characterized by its normal vector p , which is chosen to be uniformly distributed on the unit sphere.
- ▶ The rounded solution is given by $x_i = \text{sign}(p^T v_i)$. The **expected value** of this solution in $x^T W x$ can then be written as:

$$\begin{aligned} E_p[x^T W x] &= \sum_{i,j} w_{ij} E_p[x_i x_j] \\ &= \sum_{i,j} w_{ij} E_p[\text{sign}(p^T v_i) \cdot \text{sign}(p^T v_j)]. \end{aligned}$$

- Consider the plane spanned by v_i and v_j , and let θ_{ij} be the angle between these two vectors. Then,

$$E_p[\text{sign}(p^T v_i) \cdot \text{sign}(p^T v_j)] = P_1 \times 1 + P_2 \times (-1) = 1 - \frac{2\theta_{ij}}{\pi}$$

P_1 =probability that both points are on the same side of the hyperplane= $1 - \frac{\theta_{ij}}{\pi}$.

P_2 =probability that they are on different sides= $\frac{\theta_{ij}}{\pi}$.

(if $\theta_{ij} = \pi \Rightarrow v_i$ and v_j must be in different sides of any plane.)

- Thus

$$\begin{aligned} E_p[x^T W x] &= \sum_{i,j} w_{ij} \left(1 - \frac{2\theta_{ij}}{\pi} \right) \\ &= \sum_{i,j} w_{ij} \left(1 - \frac{2}{\pi} \arccos(v_i^T v_j) \right) \\ &= \frac{2}{\pi} \sum_{i,j} w_{ij} \arcsin X_{ij} \end{aligned}$$

- ▶ Notice that the expression is of course well defined, since if $X \succeq 0$ and has unit diagonal $\Rightarrow |X_{ij}| \leq 1$.
- ▶ The objective of **maximum cut** is $\frac{1}{4} \sum_{i,j} w_{ij}(1 - y_i y_j)$. So the **expected value** of the cut is then:

$$\begin{aligned} c_{\text{sdp-expected}} &= \frac{1}{4} \sum_{i,j} \left(1 - \frac{2}{\pi} w_{ij} \arcsin X_{ij} \right) \\ &= \frac{1}{4} \cdot \frac{2}{\pi} \sum_{i,j} w_{ij} \arccos X_{ij}. \end{aligned}$$

- ▶ On the other hand, the solution of the SDP gives an upper bound on the cut capacity equal to:

$$c_{\text{sdp-upperbound}} = \frac{1}{4} \sum_{i,j} w_{ij}(1 - X_{ij}).$$

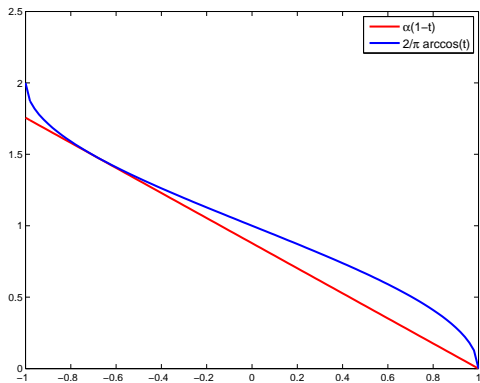
- Consider the problem of finding a constant α such that

$$\alpha(1 - t) \leq \frac{2}{\pi} \arccos(t), \quad \forall t \in [-1, 1].$$

This is

$$\alpha = \min_{t \in [-1, 1]} \frac{2 \arccos(t)}{\pi (1 - t)} = \min_{\theta \in [0, \pi]} \frac{2}{\pi} \frac{\theta}{1 - \cos \theta}.$$

- It can be shown that $0.87856 < \alpha < 0.87857$.



► Thus

$$c_{\text{sdp-upperbound}} \leq \frac{1}{4} \cdot \frac{1}{\alpha} \sum_{i,j} w_{ij} \arccos(X_{ij}) = \frac{1}{\alpha} c_{\text{sdp-expected}}.$$

► So far we have the following inequalities:

$$c_{\text{sdp-upperbound}} \leq \frac{1}{\alpha} c_{\text{sdp-expected}}$$

$$c_{\text{sdp-expected}} \leq c_{\text{max}}$$

$$c_{\text{max}} \leq c_{\text{sdp-upperbound}}$$

► Therefore

$$0.878 \cdot c_{\text{sdp-upperbound}} \leq c_{\text{max}}$$

(approximation ratio for SDP bound)

$$0.878 \cdot c_{\text{max}} \leq c_{\text{sdp-expected}}$$

(approximation ratio for feasible solution)

Nesterov's SDP bound

- ▶ In the MAXCUT problem, we are in fact maximizing the homogeneous quadratic form (omitting $\frac{1}{4}$):

$$x^T A x = \sum_{i,j=1}^n w_{ij}(1 - x_i x_j) = \sum_{i=1}^n \left(\sum_{j=1}^n w_{ij} \right) x_i^2 - \sum_{i,j=1}^n w_{ij} x_i x_j$$

over $\{-1, 1\}^n$.

- ▶ Special properties of A :
 - ▶ $A \succeq 0$ (why?);
 - ▶ $A_{ij} \leq 0$ for all $i \neq j$;
 - ▶ $\sum_{j=1}^n A_{ij} = 0$.
- ▶ What happens if A is a **general positive semidefinite matrix**?
- ▶ Let $A \succeq 0$, consider the problem:

$$(P) \quad \max_{x \in \{-1, 1\}^n} x^T A x.$$

- The SDP relaxation of (P) is

$$\begin{aligned} (SDP) \quad & \max A \bullet X \\ & \text{s.t. } \text{diag}(X) = e, \\ & X \succeq 0. \end{aligned}$$

$$\frac{v(SDP)}{v(P)} \geq 1.$$

- **Theorem** (Nesterov (1998)). Let $A \succeq 0$. Then

$$\boxed{v(P) \leq v(SDP) \leq \frac{\pi}{2} v(P)}$$

- ▶ Let X be the optimal solution to SDP. Let $X = V^T V$, where $V = (v_1, \dots, v_n)$ with $v_i \in \mathbb{R}^r$. Let ξ be a Gaussian random vector with **zero mean** and **covariance matrix** X . Let $\zeta = \text{sign}(\xi)$. Then

$$\begin{aligned} v_{SDP-e} = E(\zeta^T A \zeta) &= \frac{2}{\pi} \sum_{i,j=1}^n A_{ij} \arcsin(X_{ij}) \\ &= \frac{2}{\pi} \text{Tr}(A \arcsin[X]) \\ &= \frac{2}{\pi} \langle A, \arcsin[X] \rangle, \end{aligned}$$

where $|X_{ij}| \leq 1$ since $X \succeq 0$ and $X_{ii} = 1$.

- ▶ **Schur product of matrices**: Let $A = (a_{ij})$, $B = (b_{ij})$. Then $A \circ B = [a_{ij} b_{ij}]$ is called Schur product.

- **Schur Product Theorem:** Let $A, B \in \mathcal{S}_n^+$. Then $A \circ B \in \mathcal{S}_n^+$.
In particular $A \succeq 0 \Rightarrow A^{\circ k} \succeq 0$.
- **Proof.** For $v \in \mathbb{R}^n$, we have $(A \circ B)v = \text{Diag}(A \cdot \text{diag}(v)B)$.
So

$$\begin{aligned}
 v^T (A \circ B) v &= v^T \text{Diag}(A \cdot \text{diag}(v)B) \\
 &= \text{Tr}(\text{diag}(v)A \cdot \text{diag}(v)B) \\
 &= \langle \text{diag}(v)A \cdot \text{diag}(v), B \rangle \geq 0.
 \end{aligned}$$

$(A, B \in \mathcal{S}_n^+ \Rightarrow \langle A, B \rangle \geq 0, \text{ why? Prove it.})$

- Taylor expansion. For $t \in [-1, 1]$, we have

$$\arcsin(t) = t + \frac{1}{2} \frac{t^3}{3} + \frac{1 \cdot 3}{2 \cdot 4} \frac{t^5}{5} + \dots$$

- Since $|X_{ij}| \leq 1$, we have

$$\arcsin[X] = X + \frac{1}{2} \frac{X^{\circ 3}}{3} + \frac{1 \cdot 3}{2 \cdot 4} \frac{X^{\circ 5}}{5} + \dots$$

Since $X \succeq 0 \Rightarrow X^{\circ k} \succeq 0$, we get

$$\arcsin[X] \succeq X.$$

Again, since $X \succeq 0$, we obtain

$$\langle A, \arcsin[X] \rangle \geq \langle A, X \rangle.$$

- Thus,

$$v_{SDP-e} = \frac{2}{\pi} \langle A, \arcsin[X] \rangle \geq \langle A, X \rangle = \frac{2}{\pi} v(SDP).$$

Since $v_{SDP-e} \leq v(P) \leq v(SDP)$, we finally get

$$\frac{2}{\pi} v(P) \leq \frac{2}{\pi} v(SDP) \leq v_{SDP-e} \leq v(P) \leq v(SDP).$$

► Thus

$$v(P) \geq \frac{2}{\pi} v(SDP).$$

$$v(SDP - e) \geq \frac{2}{\pi} v(P).$$

$$\frac{2}{\pi} \approx 0.6366.$$