Lectures on Semidefinite Programming

Etienne de Klerk

[†]Delft University of Technology, The Netherlands

Topics for this talk

- Convex conic optimization and semidefinite programming (SDP);
- The conic duality theorem;
- Positive semidefinite matrices: a short review;
- Special cases/examples of SDP;
- Primal–dual interior point algorithms.

Conic Linear Optimization

A conic linear optimization problem (conic LP) takes the form:

$$\min \left\{ c^T x \mid Ax = b,\ x \in \mathcal{K} \right\},$$

where the data is $c \in \mathbb{R}^n$, $b \in \mathbb{R}^n$ and $\mathcal{K} \subset \mathbb{R}^n$ is a convex cone.
Some choices for $\mathcal{K}$:

- nonnegative orthant in $\mathbb{R}^n$ — linear programming (LP);
- the cone of $m \times m$ positive semidefinite matrices such that $n = \binom{m+1}{2}$ — semidefinite programming (SDP).

Importance of SDP

- Contains important convex optimization problems as special cases: LP, QP, etc.
- Important applications: Global and combinatorial optimization, control, etc.
- Can be solved efficiently by interior point algorithms.

More on convex cones

A subset $\mathcal{K}$ of $\mathbb{R}^m$ is a *cone* if

$$a \in \mathcal{K}, \;\; \lambda \geq 0 \;\Rightarrow\; \lambda a \in \mathcal{K},$$

and the cone $\mathcal{K}$ is a *convex cone* if moreover

$$a, a' \in \mathcal{K} \Rightarrow a + a' \in \mathcal{K}.$$

We impose three more conditions on $\mathcal{K}$ to obtain duality results similar to those for LP, namely:

- $\mathcal{K}$ must be *pointed* (may not contain a ray);
- $\mathcal{K}$ must be closed;
- $\mathcal{K}$ must be solid (have a nonempty interior).

Dual cones

The *dual cone* of a convex cone $\mathcal{K}$ is:

$$\mathcal{K}_* = \left\{ \lambda \in \mathbb{R}^m \ : \ \lambda^T a \geq 0 \ \ \forall a \in \mathcal{K} \right\}.$$

Theorem: Let $\mathcal{K} \subset \mathbb{R}^m$ be a nonempty cone.

(i) The set $\mathcal{K}_*$ is a closed convex cone.
(ii) If $\mathcal{K}$ is solid then $\mathcal{K}_*$ is pointed.
(iii) If $\mathcal{K}$ is a closed convex pointed cone, then $\mathcal{K}_*$ is solid.
(iv) If $\mathcal{K}$ is a closed convex cone, then so is $\mathcal{K}_*$, and the cone dual to $\mathcal{K}_*$ is $\mathcal{K}$ itself.

Corollary: If $\mathcal{K} \subset \mathbb{R}^m$ is a closed, pointed, solid, convex cone then so is $\mathcal{K}_*$, and vice versa.

Dual problem

Consider the conic optimization problem:

$$(COP): \quad \inf \left\{ c^T x \;\middle|\; Ax = b, \; x \in \mathcal{K} \right\},$$

where A has full rank. Feasible set $\mathcal{P}$, optimal set $\mathcal{P}^*$ (possibly empty).
Associated *dual problem*

$$(COD): \quad \sup \left\{ b^T y \;\middle|\; A^T y + s = c, \; s \in \mathcal{K}_* \right\}.$$

Feasible set $\mathcal{D}$, optimal set $\mathcal{D}^*$ (possibly empty).

Weak duality theorem

Let $x \in \mathcal{P}$ and $(y, s) \in \mathcal{D}$. Then

$$\begin{aligned} c^T x - b^T y &= (A^T y + s)^T x - b^T y \\ &= (Ax)^T y + s^T x - b^T y \\ &= s^T x \geq 0, \end{aligned}$$

because $x \in \mathcal{K}$ and $s \in \mathcal{K}_*$.

In words, the *duality gap* is *nonnegative* at feasible primal-dual solutions.

Strong duality theorem

Assume (COP) and (COD) are *strictly feasible*:

$$\mathcal{P} \cap \text{int}(\mathcal{K}) \neq \emptyset \text{ and } \mathcal{D} \cap \text{int}(\mathcal{K}_*) \neq \emptyset.$$

Then both problems are *solvable*, i.e:

$$\mathcal{P}^* \neq \emptyset \text{ and } \mathcal{D}^* \neq \emptyset,$$

and at optimal solutions $x^* \in \mathcal{P}^*$ and $(y^*, s^*) \in \mathcal{D}^*$ one has

$$c^T x^* - b^T y^* = {x^*}^T s^* = 0 \quad \text{(Duality gap zero.)}$$

We call (x^*, s^*) *complementary solutions*.

The trace operator

The trace of an $n \times n$ matrix A is

$$\mathrm{Tr}(A) = \sum_{i=1}^{n} a_{ii}.$$

Properties:

Let $A \in \mathbb{R}^{n \times n}$ and $B \in \mathbb{R}^{n \times n}$:

- $\mathrm{Tr}(A) = \sum_{i=1}^{n} \lambda_i(A)$;
- $\mathrm{Tr}(A) = \mathrm{Tr}\left(A^T\right)$;
- $\mathrm{Tr}(AB) = \mathrm{Tr}(BA)$
- $\mathrm{Tr}(AB^T) = \sum_{i,j=1}^{n} a_{ij} b_{ij}$.

An inner product

We denote the cone of real symmetric $n \times n$ matrices by $\mathcal{S}_n$. Let $A, B \in \mathcal{S}_n$. The inner product

$$\langle A, B \rangle := \mathrm{Tr}\,(AB) = \sum_{i,j} a_{ij} b_{ij}.$$

The inner product induces the *Frobenius* (Euclidean) norm:

$$\|A\|^2 := \langle A, A \rangle = \mathrm{Tr}\left(AA^T\right) = \sum_{i,j=1}^{n} a_{ij}^2,$$

which is sub-multiplicative: $\|AB\| \leq \|A\| \|B\|$.

Positive semidefinite matrices

Let $X \in \mathcal{S}_n$. The following are equivalent:

- $X \in \mathcal{S}_n^+$ or $X \succeq 0$ (X is PSD);
- $z^T X z \geq 0 \quad \forall z \in \mathbb{R}^n$;
- $\lambda_{\min}(X) \geq 0$;
- All principal minors of X are nonnegative;
- $X = LL^T$ for some $L \in \mathbb{R}^{n \times n}$.

A nonsingular matrix $X \succeq 0$ is called positive definite ($X \succ 0$).

P.s.d. matrices (ctd.)

Any $A \in \mathcal{S}_n$ has a *spectral decomposition*:

$$A = \sum_{i=1}^{n} \lambda_i q_i q_i^T := Q \Lambda Q^T$$

where

$$A q_i = \lambda_i q_i, \quad Q = [q_1 \ \ldots \ q_n], \quad Q^T Q = I,$$

and Λ is a diagonal matrix with $\Lambda_{ii} = \lambda_i$.
If $A \in \mathcal{S}_n^+$ then $\lambda_i \geq 0$. Square root factorization:

$$A^{\frac{1}{2}} := \sum_{i=1}^{n} \sqrt{\lambda_i} q_i q_i^T.$$

P.s.d. matrices (ctd.)

Theorem (Fejer): The cone $\mathcal{S}_n^+$ is self-dual with respect to the Euclidean inner product.
Proof: Let $A \in \mathcal{S}_n^+$ and $B \in \mathcal{S}_n^+$; then

$$\begin{aligned} \langle A, B \rangle &= \mathbf{Tr}\left(A^{\frac{1}{2}} A^{\frac{1}{2}} B^{\frac{1}{2}} B^{\frac{1}{2}} \right) \\ &= \mathbf{Tr}\left(A^{\frac{1}{2}} B^{\frac{1}{2}} B^{\frac{1}{2}} A^{\frac{1}{2}} \right) \\ &= \left\| A^{\frac{1}{2}} B^{\frac{1}{2}} \right\|^2 \geq 0. \end{aligned}$$

Conversely, if $A \in \mathcal{S}_n$ and $\langle A, B \rangle \geq 0$ for all $B \in \mathcal{S}_n^+$, then let $x \in \mathbb{R}^n$ be given and set $B = xx^T \in \mathcal{S}_n^+$. The result follows (why?)

SDP: The standard form

Let $\mathcal{K} = \mathcal{S}_n^+$. We get the conic optimization problems

$$(P): \quad p^* = \inf_{X \in \mathcal{S}_n^+} \operatorname{Tr}(CX)$$

$$\text{subject to } \operatorname{Tr}(A_i X) = b_i, \quad \forall\, i.$$

Dual problem:

$$(D): \quad d^* = \sup_{y \in \mathbb{R}^m, S \in \mathcal{S}_n^+} b^T y$$

$$\text{subject to } \sum_{i=1}^{m} y_i A_i + S = C.$$

Weak Duality in SDP

We can prove the weak conic duality directly:
Feasible $X \in \mathcal{P}$ and $y, S \in \mathcal{D}$ satisfy:

$$\begin{aligned} & \mathrm{Tr}(CX) - b^T y \\ = \; & \mathrm{Tr}\left(\left(S + \sum_{i=1}^m y_i A_i\right) X\right) - \sum_{i=1}^m y_i \mathrm{Tr}(A_i X) \\ = \; & \mathrm{Tr}(SX) \geq 0 \text{ (since } X, S \succeq 0) \end{aligned}$$

(Duality gap always nonnegative.)

Strong Duality in SDP

From the strong conic duality theorem we have:
If there exist *strictly* feasible $X \succ 0$ and $S \succ 0$, then there exist *complementary solutions* $X^* \in \mathcal{P}$ and $S^* \in \mathcal{D}$, i.e.

$$\mathrm{Tr}(X^* S^*) = 0 \Leftrightarrow X^* S^* = 0.$$

Bad duality example

$$(D): \quad \sup_{y \in \mathbb{R}^2} y_2$$

subject to

$$S = -y_1 \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} - y_2 \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} + \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix} \succeq 0.$$

(D) is *not solvable* but $\sup_{y \in \mathcal{D}} y_2 = 1$ (why?).

Bad duality example (ctd)

The dual problem of the example is:

$$(P): \quad \min_X \mathrm{Tr}\left(\begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix} X\right)$$

$$\text{subject to } X = \begin{bmatrix} 0 & x_{12} \\ x_{12} & 1 \end{bmatrix} \succeq 0.$$

Note that $X \succeq 0$ implies $x_{12} = 0$ so that

$$X^* = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \succeq 0,$$

is the unique optimal solution with optimal value 1.

Optimality conditions

From the strong duality theorem we get *sufficient* optimality conditions for (P) and (D):

$$\begin{aligned} \operatorname{Tr}(A_i X) &= b_i, \qquad\qquad i = 1, \ldots, m \\ \sum_{i=1}^m y_i A_i + S &= C \\ XS &= 0 \\ X \succeq 0, &\quad y \in \mathbb{R}^m, S \succeq 0 \end{aligned}$$

NB: Also *necessary* if (P) and (D) are *strictly feasible*. If (X, S) are complementary solutions $XS = 0$, and $X + S \succ 0$ then we call (X, S) a *strictly complementary solution pair*.

Special cases of SDP

Tools and examples

- The Schur complement theorem: a useful tool;
- Convex quadratic problems;
- Eigenvalue and matrix norm optimization;
- Logarithmic Chebychev approximation;
- Minimization of univariate polynomials;
- The Lovász theta function.

Schur complement theorem

Let

$$M = \begin{bmatrix} A & B \\ B^T & C \end{bmatrix}$$

where $A \succ 0$ and $C \in \mathcal{S}_n$. The matrix

$$C - B^T A^{-1} B$$

is called the *Schur complement of A in M*. The following are equivalent:

- M is positive (semi)definite;
- $C - B^T A^{-1} B$ is positive (semi)definite.

Convex quadratic problems

$$\min \left\{ c^T x \ : \ f_i(x) \le 0, \ i = 1, \ldots, m \right\},$$

where

$$f_i(x) = (B_i x + b_i)^T (B_i x + b_i) - c_i^T x - d_i, \quad \forall i.$$

Omitting the index i each constraint has the form

$$\|Bx + b\|^2 \le c^T x + d.$$

Via Schur complement theorem equivalent to:

$$\begin{bmatrix} I & Bx + b \\ (Bx + b)^T & c^T x + d \end{bmatrix} \succeq 0.$$

Smallest eigenvalue problem

Consider $M \in \mathcal{S}_n$ with eigenvalues

$$\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n.$$

One trivially has

$$(D) \qquad \lambda_1 = \max\{\lambda \ : \ \lambda I \preceq M\}.$$

Corresponding dual SDP problem is

$$(P) \qquad \min\{\mathrm{Tr} MX \ : \ \mathrm{Tr} X = 1 \ \ X \succeq 0\}.$$

Both problems are strictly feasible: In (D), take $\lambda < \lambda_1$ and in (P): take $X = \frac{1}{n} I$.

Eigenvalue optimization

Notation: $\lambda_{\max}(A)$ denotes the *largest eigenvalue* of $A \in \mathcal{S}_n$. Consider

$$\min_y \lambda_{\max}(A(y))$$

$$A(y) := A_0 + y_1 A_1 + \cdots + y_m A_m,$$

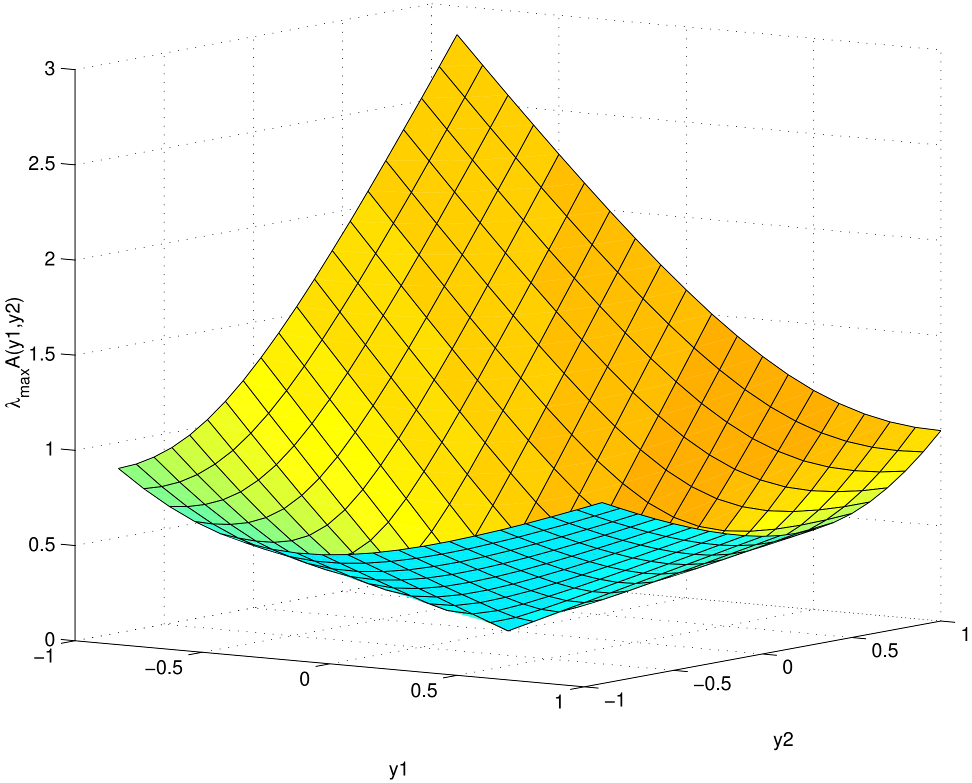
for given $A_i \in \mathcal{S}_n$ $(i = 0, \ldots, m)$. This can be formulated as an SDP:

$$\min \{t \ : \ tI - A(y) \succeq 0\}$$

NB: the function $f(y) = \lambda_{\max}(A(y))$ is convex but not *differentiable*.

Eig. optimization: example

$$\min_{y_1, y_2} \left\{ \lambda_{\max} \left(y_1 \begin{bmatrix} 1 & -1 \\ -1 & 0 \end{bmatrix} + y_2 \begin{bmatrix} 0 & 1 \\ 1 & -1 \end{bmatrix} \right) \right\}$$



Optimal solution $y_1^* = y_2^* = 0$.

Chebychev approximation

We wish to solve $Ax = b$ approximately, where $A = [a_1 \cdots a_m]^T \in \mathbb{R}^{n \times m}$. *Chebychev approximation*:

$$\min_x \|Ax - b\|_\infty := \min_x \max_i |a_i^T x - b_i|.$$

LP reformulation:

$$\min \left\{ t \ : \ -t \leq a_i^T x - b_i \leq t \ \ \forall i \right\}.$$

Logarithmic Chebychev approximation:

$$\min_x \max_i \left| \ln(a_i^T x) - \ln(b_i) \right|.$$

Chebychev approximation (ctd.)

$$\min_x \max_i \left| \ln(a_i^T x) - \ln(b_i) \right|$$

is equivalent to:

$$\min \left\{ t \ : \ 1/t \leq a_i^T x / b_i \leq t \ \ \forall i \right\}.$$

SDP formulation:

$$\min \left\{ t \ : \ \begin{bmatrix} t - a_i^T x/b_i & 0 & 0 \\ 0 & a_i^T x/b_i & 1 \\ 0 & 1 & t \end{bmatrix} \succeq 0 \ \ \forall i \right\}.$$

Minimization of polynomials

Let $p : \mathbb{R} \mapsto \mathbb{R}$ be a *univariate polynomial*.
Theorem:
$p(x) \geq 0 \; \forall x \in \mathbb{R}$ iff

$$p = \sum_i p_i^2$$

for some polynomials p_i.
We call p a *sum of squares* (SOS) in this case.

Minimization of polynomials (ctd.)

Now we have

$$\min_{x\in\mathbb{R}} p(x) = \max_{t,x}\{t \ : \ p(x)-t\geq 0 \ \forall x\in\mathbb{R}\}$$

$$= \max_{t,x}\left\{t \ : \ p(x)-t=\sum_i p_i(x)^2\right\}$$

for some p_i's.
SDP can be used to determine if a polynomial is an SOS (*Gram matrix method*).

The Gram matrix method

Theorem: A polynomial $p : \mathbb{R}^n \mapsto \mathbb{R}$ of degree $2m$ is an SOS iff

$$p(x) = \tilde{x}^T M \tilde{x}, \text{ for some } M \succeq 0, \tag{3}$$

where $\tilde{x} = [1\ x_1\ x_2\ \ldots\ x_n\ x_1^2\ x_1x_2\ \ldots\ x_n^m]^T$ is a vector of all possible monomials of degree at most m.

- Dimension of $\tilde{x}$ is $\binom{n+m}{m}$: polynomial in n if m is fixed.
- The right-hand-side in (1) is linear in the entries of $M \Rightarrow$ (1) is a linear matrix inequality (LMI) (semidefinite feasibility problem).

Example (Parrilo)

Is $P(x) := 2x_1^4 + 2x_1^3x_2 - x_1^2x_2^2 + 5x_2^4$ a sum of squares? For all $\lambda \in \mathbb{R}$ we have

$$P(x) = \begin{bmatrix} x_1^2 \\ x_2^2 \\ x_1x_2 \end{bmatrix}^T \begin{bmatrix} 2 & -\lambda & 1 \\ -\lambda & 5 & 0 \\ 1 & 0 & -1+2\lambda \end{bmatrix} \begin{bmatrix} x_1^2 \\ x_2^2 \\ x_1x_2 \end{bmatrix}.$$

If we call the 3×3 matrix in the last expression $M(\lambda)$, then $M(\lambda)$ defines an *affine space* parametrized by λ. **SDP problem:** is there a λ such that $M(\lambda) \succeq 0$ (positive semidefinite)?

Example (ctd.)

for $\lambda = 3$, $M(\lambda)$ is positive semidefinite, and

$$M(3) = L^T L, \quad L = \frac{1}{\sqrt{2}} \begin{bmatrix} 2 & -3 & 1 \\ 0 & 1 & 3 \end{bmatrix},$$

and consequently

$$P(x) = \tilde{x}^T M(3) \tilde{x} = \tilde{x}^T L^T L \tilde{x} = \|L\tilde{x}\|^2,$$

where $\tilde{x} = [x_1^2 \ \ x_2^2 \ \ x_1 x_2]^T$.
Thus P can be written as a sum of squares.

Nesterov's aproach

We have obtained:

$$\min_{x \in \mathbb{R}} p(x) = \max_{t,x} \left\{ t \;:\; p(x) - t = \sum_i p_i(x)^2 \right\}$$
$$= \max_{t,x} \{ t \;:\; p(x) - t = \tilde{x}^T M \tilde{x} \}$$

for some $M \succeq 0$, where $\tilde{x}^T = [1 \; x \; x^2 \ldots x^{\frac{1}{2}\deg(p)}]$.
Let $p(x) = \sum_\alpha a_\alpha x^\alpha$. Then the optimization problem becomes: maximize t such that

$$a_0 - t = M_{00}, \quad a_\alpha = \sum_{i+j=\alpha} M_{ij}, \quad M \succeq 0.$$

Example

$$p(x) := x^2 - 2x = (x-1)^2 - 1.$$

Equivalent problem: $\max t$ such that

$$x^2 - 2x - t = \begin{bmatrix} 1 \\ x \end{bmatrix}^T \begin{bmatrix} M_{00} & M_{01} \\ M_{10} & M_{11} \end{bmatrix} \begin{bmatrix} 1 \\ x \end{bmatrix}, \qquad (6)$$

for some $M \succeq 0$.
From (4):

$$M_{00} = -t, \quad M_{01} = M_{10} = -1, \quad M_{11} = 1.$$

Example (ctd.)

We therefore get

$$\min_{x \in \mathbb{R}} p(x) = \max_{t,M} t$$

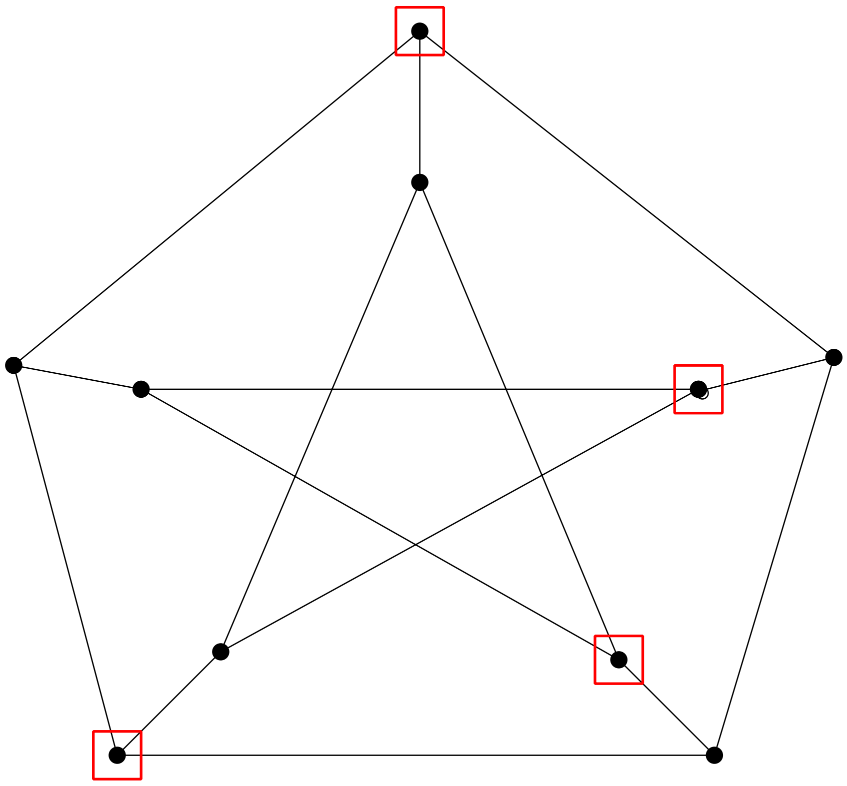
such that

$$M = \begin{bmatrix} -t & -1 \\ -1 & 1 \end{bmatrix} \succeq 0.$$

Note that the optimal value is -1, as it should be.

Co-cliques

A *co-clique* of $G = (V, E)$ is a subset $V' \subset V$ such that the *induced subgraph* on V' has no edges.



The *co-clique number* $\alpha(G)$ is the cardinality of the largest co-clique of G.

Vertex colourings

- An assignment of colours to the vertices V of G such that endpoints of each $e \in E$ are assigned different colours.
- Chromatic number $\gamma(G)$: smallest number of colours needed to colour V;
- It is NP hard to compute $\gamma(G)$ (or $\alpha(G)$), or even to give a non-trivial polynomial time approximation.

Lovász ϑ-function

A graph $G = (V, E)$ is given. Define:

$$\vartheta(G) := \max \operatorname{Tr}\left(ee^T X\right) = e^T X e$$

subject to

$$\begin{aligned} X_{ij} = 0,\ \{i, j\} &\in E \ (i \neq j) \\ \operatorname{Tr}(X) = 1 & \\ X \in \mathcal{S}_n^+, & \end{aligned}$$

where e denotes the all-one vector.

Lovász ‘sandwich theorem’

Let $\alpha(G)$ denote the independence number of G and $\gamma(\bar{G})$ the chromatic number of $\bar{G}$.

Lovász’s sandwich theorem

$$\alpha(G) \leq \vartheta(G) \leq \gamma(\bar{G}).$$

First equality is easy. Second inequality via strong duality theorem.
Implication: we can compute $\alpha(G)$ and $\gamma(\bar{G})$ in pol. time for *perfect graphs*.