Applications of Semidefinite Programming

Etienne de Klerk

University of Waterloo

Tutorial at the Workshop Large-scale NLP and SDP

Waterloo, Canada, May 12, 2004

Topics

- The Schur complement theorem: a useful tool;
- Convex quadratic optimization and approximation;
- Eigenvalue and matrix norm optimization;
- Logarithmic Chebychev approximation;
- Representation of nonnegative polynomials;
- The Lovász theta function, max-k-cut, and Shannon capacity of a graph;

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.

Schur complement theorem

Proof: We have

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

with $A \succ 0$. Set $D = -A^{-1}B$. Now:

$$\begin{bmatrix} I & 0 \\ D^T & I \end{bmatrix} \begin{bmatrix} A & B \\ B^T & C \end{bmatrix} \begin{bmatrix} I & D \\ 0 & I \end{bmatrix} = \begin{bmatrix} A & 0 \\ 0 & C - B^T A^{-1} B \end{bmatrix}.$$

Convex quadratic problems

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

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

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

Via Schur complement theorem equivalent to:

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

Least squares approximation

We want to obtain the best convex quadratic least squares approximation $g(x) = x^T Qx + c^T x + d$ to a convex function $f : \mathbb{R}^n \to \mathbb{R}$ in a set of points

$$\mathcal{Z} := \{z_1, z_2, \cdots, z_N\}.$$

This is obtained by solving

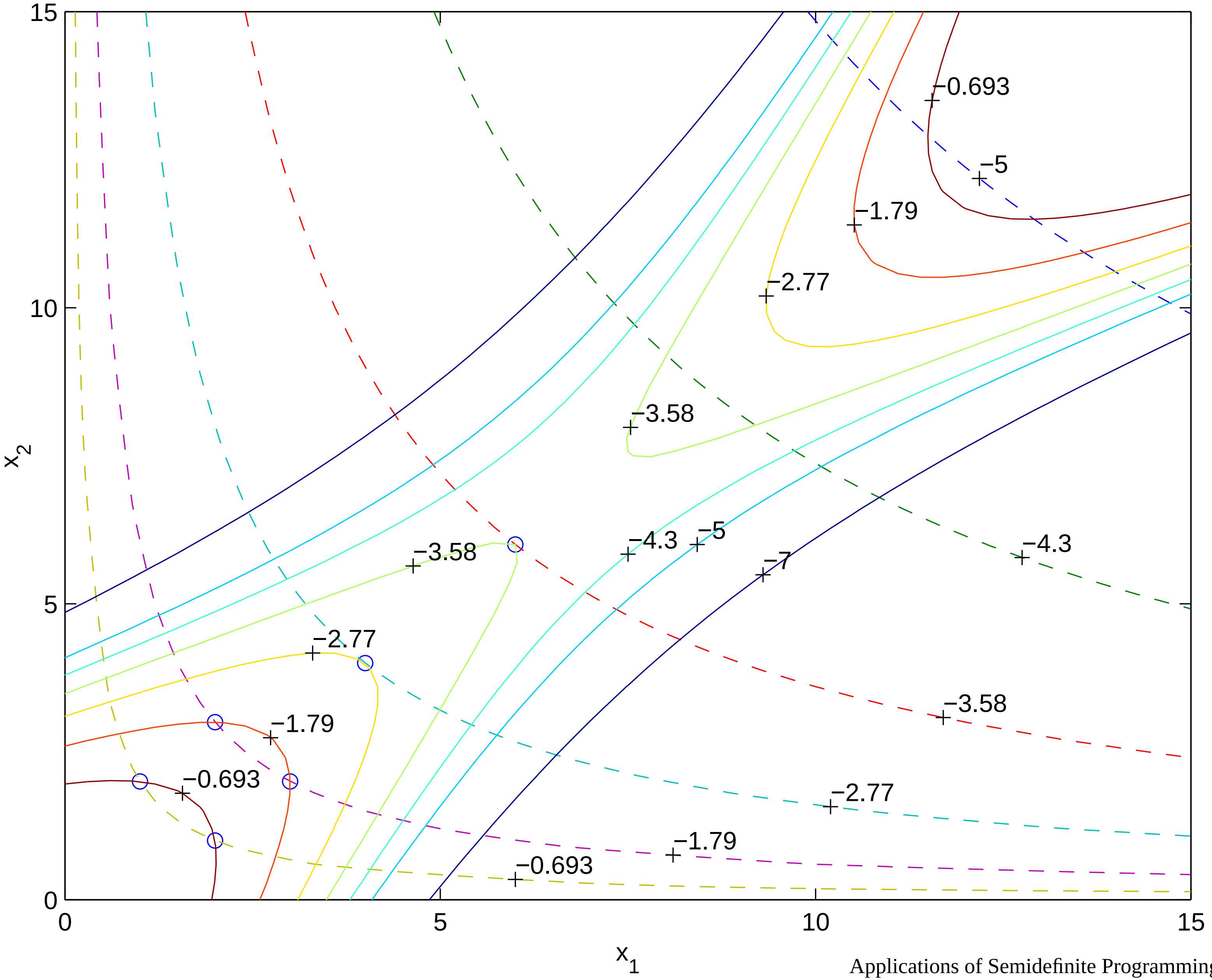
$$\min_{Q \succeq 0, c \in \mathbb{R}^n, d \in \mathbb{R}} \sum_{z \in \mathcal{Z}} \left(f(z) - (z^T Qz + c^T z + d) \right)^2 .$$

The objective is a convex quadratic function in the unknowns $Q \succeq 0$, $c \in \mathbb{R}^n$ and $d \in \mathbb{R}$, i.e. the problem has an SDP reformulation.

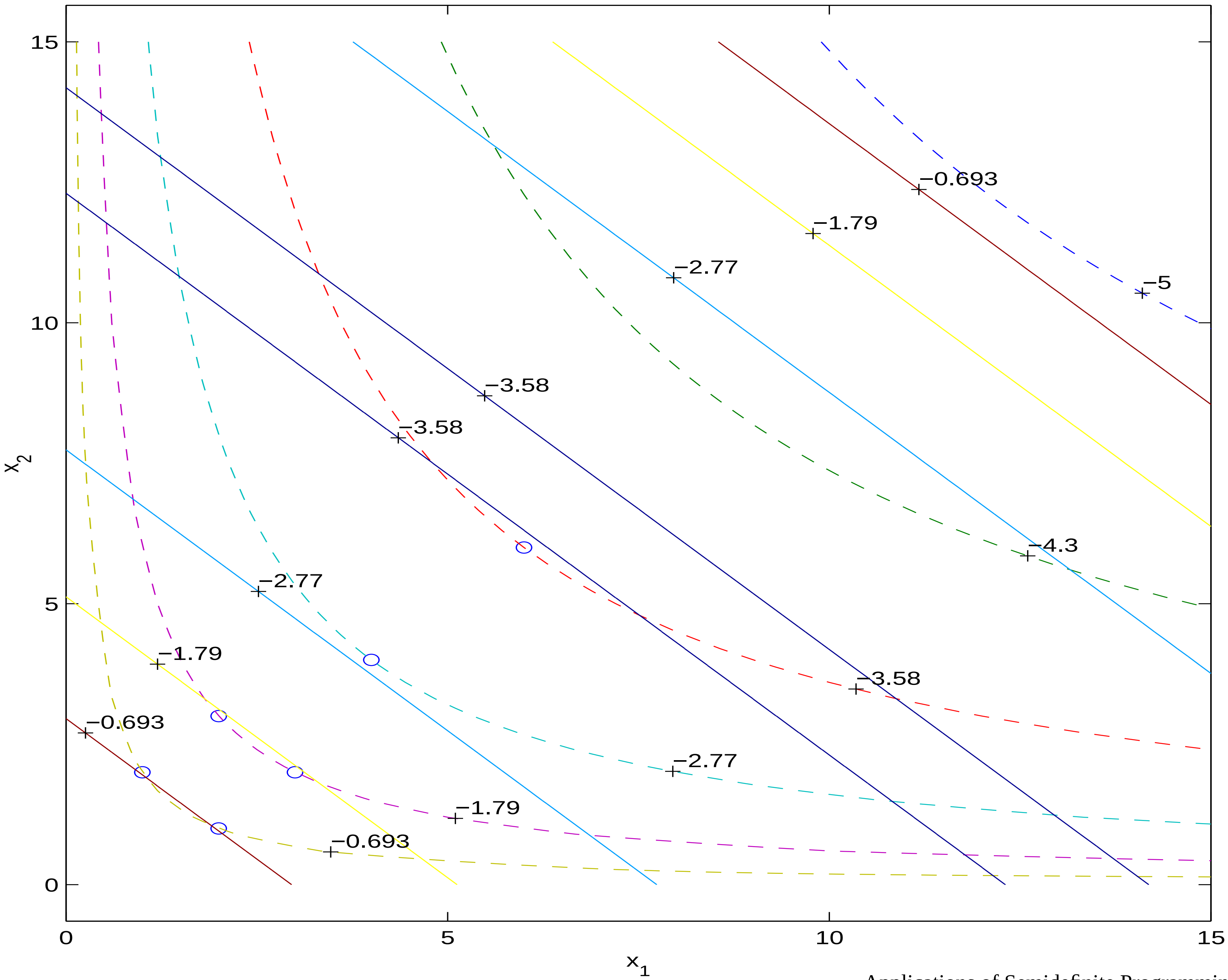
Example: $f(x) = -\ln x_1 x_2$

- Contours of f will be denoted by dashed lines.
- The points $\mathcal{Z} := \{z_1, z_2, \cdots, z_N\}$ will be denoted by circles.
- The quadratic least squares approximations in the points of $\mathcal{Z}$ will have solid contours.

Least squares approximation



Convex LS approximation



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

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

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

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

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

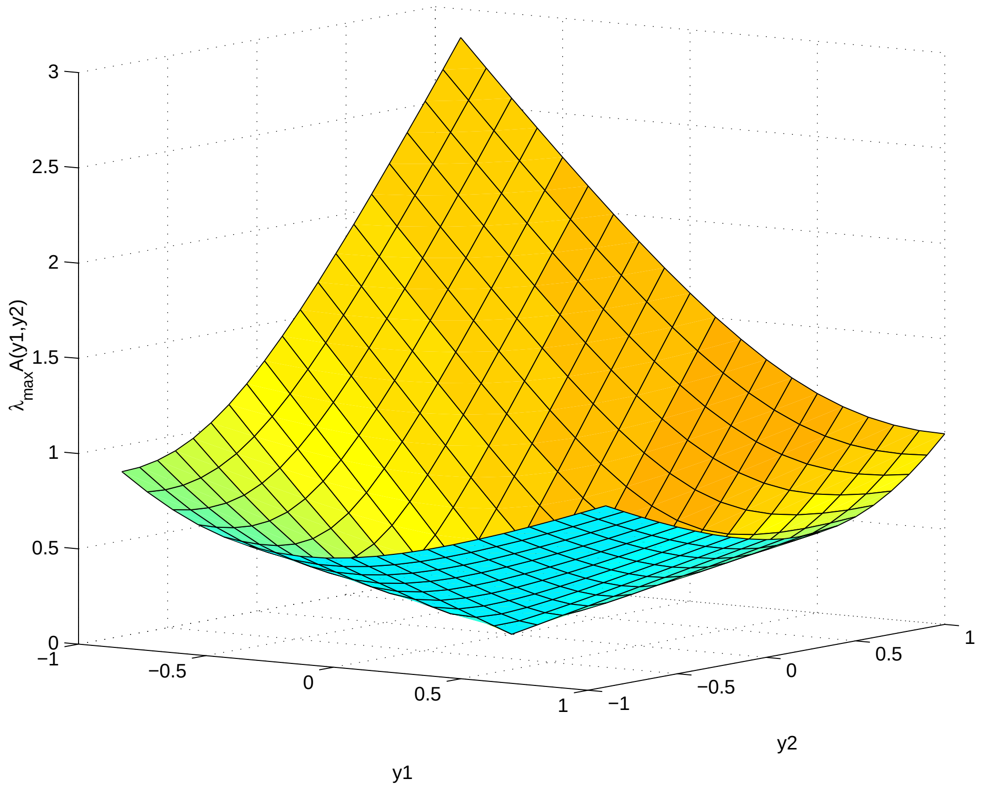
for given $A_i \in \mathcal{S}_n$ ($i = 0, \ldots, m$). This can be formulated as an SDP:

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

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:

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

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:

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

Nonnegative 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

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

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

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

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

- If p is homogeneous we only need the monomials of degree exactly m.
- Dimension of $\tilde{x}$ is $\binom{n+m}{m}$: polynomial in n if m is fixed.
- The right-hand-side in $(*)$ is linear in the entries of $M \Rightarrow (*)$ is a linear matrix inequality (LMI).

Example (Parrilo)

Is $P(x) := 2x_1^4 + 2x_1^3x_2 - x_1^2x_2^2 + 5x_2^4$ a sum of squares? YES, because

$$P(x) = \begin{bmatrix} x_1^2 \\ x_2^2 \\ x_1x_2 \end{bmatrix}^T \begin{bmatrix} 2 & -3 & 1 \\ -3 & 5 & 0 \\ 1 & 0 & 5 \end{bmatrix} \begin{bmatrix} x_1^2 \\ x_2^2 \\ x_1x_2 \end{bmatrix}.$$

The 3×3 matrix (say M) in the last expression is positive semidefinite.

Example (ctd.)

Since M is positive semidefinite, it has a Choleski factorization:

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

and consequently, using $\tilde{x} = [x_1^2 \;\; x_2^2 \;\; x_1 x_2]^T$,

$$\begin{aligned} P(x) &= \tilde{x}^T M \tilde{x} = \tilde{x}^T L^T L \tilde{x} = \|L\tilde{x}\|^2 \\ &= \frac{1}{2}\left(2x_1^2 - 3x_2^2 + x_1 x_2\right)^2 + \frac{1}{2}\left(x_2^2 + 3x_1 x_2\right)^2 . \end{aligned}$$

Minimization of polynomials

Note that:

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

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

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

for some $M \succeq 0$.

Equating the LHS and RHS coefficients:

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

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.

Nonnegative polynomials II

Artin's theorem (Hilbert's 17th problem):
Let $p : \mathbb{R}^n \mapsto \mathbb{R}$ be a *multivariate polynomial*.

Then $p(x) \geq 0 \; \forall x \in \mathbb{R}$ iff

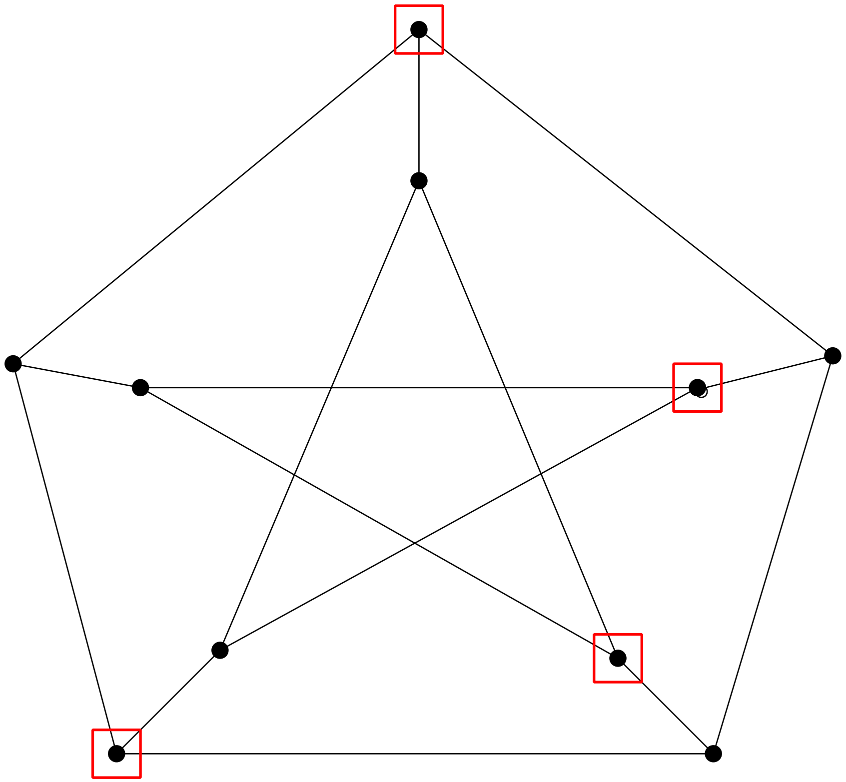
$$p \sum_j q_j^2 = \sum_i p_i^2$$

for some polynomials p_i and q_j.

Implication: one can obtain a *certificate* of nonnegativity of p via semidefinite programming.

Co-cliques

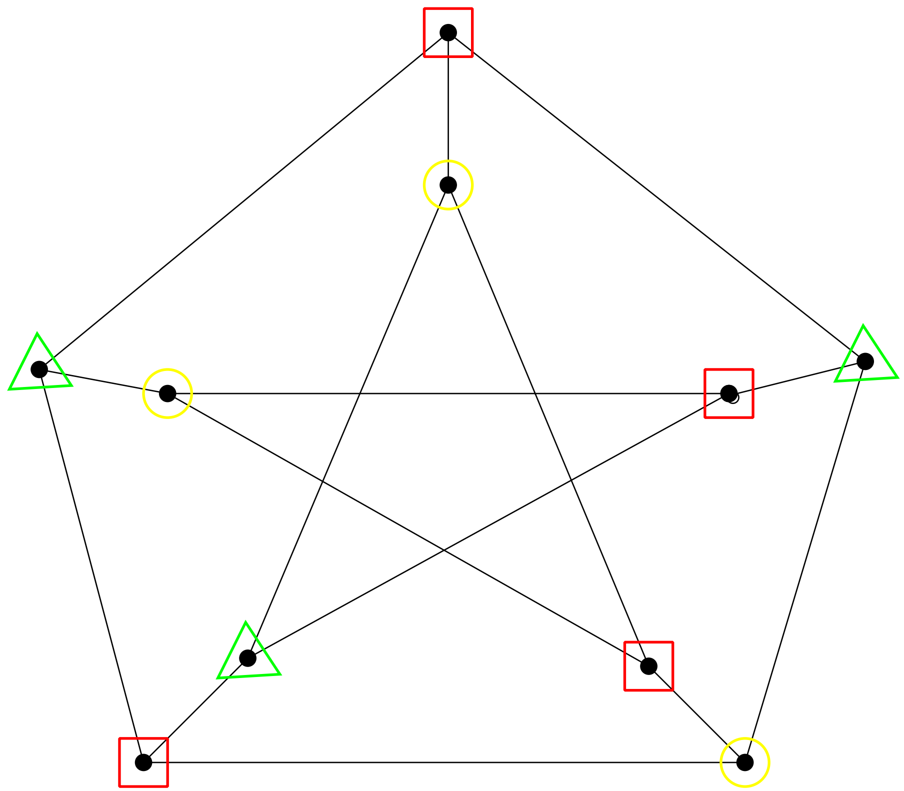
A *co-clique* of $G = (V, E)$ is a subset $V' \subset V$ such that the *induced subgraph* on V' has no edges.



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

Vertex colourings

A legal (proper) vertex colouring is an assignment of colours to the vertices V of G such that endpoints of each $e \in E$ are assigned different colours.



MAX−3−CUT of the Petersen graph

Vertex colourings (ctd.)

- 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.
- If $\bar{G}$ denotes the complementary graph of G, then obviously $\alpha(G) \leq \gamma(\bar{G})$.

Lovász ϑ-function

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

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

subject to

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

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

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

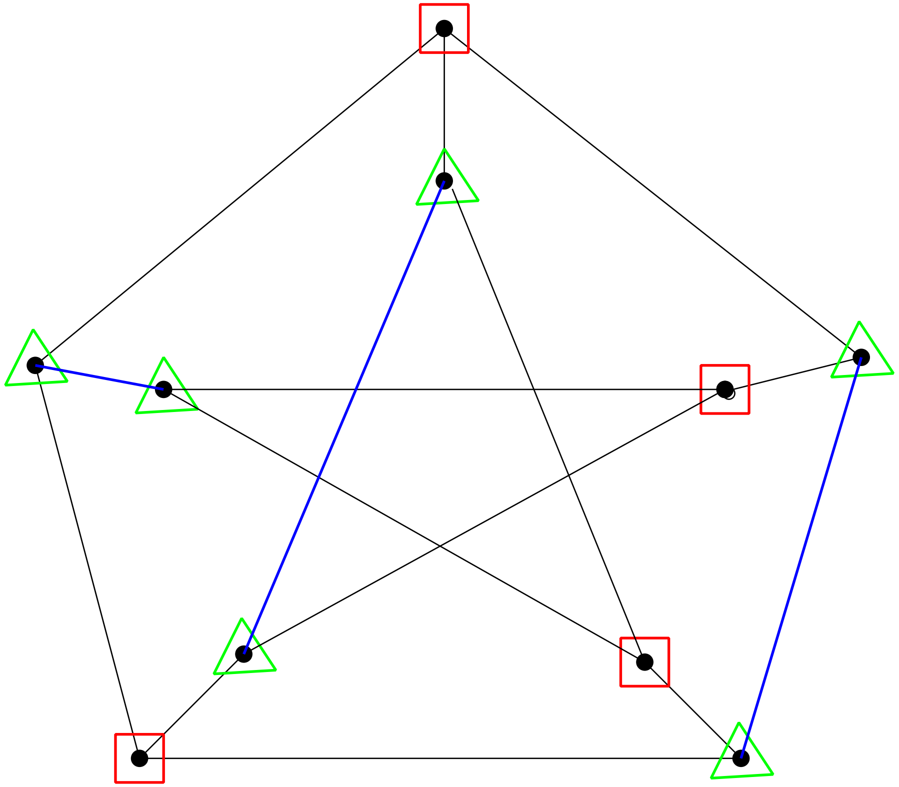
First equality is easy. Second inequality via strong duality theorem.

Example: For the pentagon, $\vartheta(G) = \vartheta(\bar{G}) = \sqrt{5}$, and

$$2 \equiv \alpha(G) \leq \vartheta(G) \leq \gamma(\bar{G}) \equiv 3.$$

Max-k-cut

A maximum k-cut is a vertex colouring using k colours such that the number of edges with endpoints of different colours is maximal.



2–CUT of the Petersen graph

Max-k-cut and $\vartheta(G)$

Let a graph $G = (V, E)$ and an integer $k > 2$ be given, and let |MAX-k-CUT| denote the cardinality of the maximum k cut.
One has

$$|\text{MAX-}k\text{-CUT}| \le \frac{k-1}{k}|E|\left(\frac{\vartheta(\bar{G})}{\vartheta(\bar{G})-1}\right).$$

Example: For the pentagon, $\vartheta(G) = \vartheta(\bar{G}) = \sqrt{5}$, and

$$4 = |\text{MAX-2-CUT}| \le \frac{1}{2}5\left(\frac{\sqrt{5}}{\sqrt{5}-1}\right) \approx 4.5225.$$

Data transmission problem

- We consider the problem of transmitting data via a communication channel. The data is coded as words consisting of the letters of an alphabet.
- During transmission, it may happen that any letter is changed to an ‘adjacent’ letter.
- We associate a set of vertices V with the letters of the alphabet, and join two vertices by an edge if the two corresponding letters are adjacent.
- What is the largest possible dictionary of r-letter words with the property that one word in the dictionary cannot be changed to another word in the same dictionary during transmission?

Strong graph product

The strong product $G_1 * G_2$ of graphs $G_1 = (V_1, E_1)$ and $G_2 = (V_2, E_2)$ is defined as the graph with vertex set $V = V_1 \times V_2$ and edge set:

$$E := \{((\bar{v}_i, v_j), (\bar{v}_k, v_l)) \mid [(\bar{v}_i, \bar{v}_k) \in E_1 \text{ or } i = k] \\ \text{and } [(v_j, v_l) \in E_2 \text{ or } j = l]\} \, .$$

NB: if $S_1 \subset V_1$ and $S_2 \subset V_2$ are stable sets of G_1 and G_2 respectively, then $S_1 \times S_2$ is a stable set of $G_1 \times G_2$. Thus

$$\alpha(G)^r \leq \alpha\Bigg(\underbrace{G * \ldots * G}_{r \text{ times}}\Bigg) := \alpha\left(G^r\right).$$

Shannon capacity

- Consider two r-letter words

$$(l_1, \ldots, l_r) \text{ and } (\hat{l}_1, \ldots, \hat{l}_r).$$

- They correspond to the endpoints of an edge in G^r if and only if for each $i = 1, \ldots, r$, either $l_i = \hat{l}_i$, or the letters l_i and $\hat{l}_i$ are adjacent.
- Therefore, the maximal number of words in the dictionary is $\alpha\left(G^r\right)$.

Shannon capacity

Theorem (Lovász):
Let two graphs $G_1 = (V_1, E_1)$ and $G_2 = (V_2, E_2)$ be given. Then

$$\vartheta\left(G_1 * G_2\right) = \vartheta(G_1)\vartheta(G_2).$$

Consequence:

$$\alpha\left(G^r\right) \leq \vartheta\left(G^r\right) = \left(\vartheta(G)\right)^r.$$

In words, $\left(\vartheta(G)\right)^r$ is an upper bound on the size of the dictionary. Moreover

$$\Theta(G) := \lim_{r\to\infty} \alpha\left(G^r\right)^{\frac{1}{r}} \leq \vartheta(G).$$

Shannon capacity

- The quantity $\Theta(G) := \lim_{r\to\infty} \alpha\left(G^r\right)^{\frac{1}{r}}$ is called the *Shannon capacity* of G.
- It is not known if the Shannon capacity can be computed by any algorithm.
- **Example:** if G is the pentagon then $\Theta(G) \leq \vartheta(G) = \sqrt{5}$. In fact, one can show that $\Theta(G) = \sqrt{5}$.
- The Shannon capacity of the 7-cycle (heptagon) is not known.

More info

Christoph Helmberg's SDP page with links to papers and software downloads:

http://www-user.tu-chemnitz.de/∼helmberg/semidef.html

Excellent introduction to SDP: L. Vandenberghe and S. Boyd. Semidefinite programming. *SIAM Review* 38, 49–95, 1996.

Today's lecture was largely based on: E. de Klerk. Aspects of Semidefinite Programming: Interior Point Algorithms and Selected Applications. Kluwer Academic Publishers, 2002.

Solving optimization problems via internet (NEOS server):

http://www-neos.mcs.anl.gov/