Applications of cone programming: a tutorial

Etienne de Klerk

Tilburg University, The Netherlands

Outline

- Background on cone programming;
- Robust linear and quadratic programming;
- The Lovász ϑ-function;
- The S-lemma;
- Eigenvalue optimization.

Convex sets

Definition (Convex combination of points)

Let two points $x, y \in \mathbb{R}^n$ and $0 \leq \lambda \leq 1$ be given.
Then the point

$$z = \lambda x + (1 - \lambda) y$$

is a *convex combination* of the two points x and y.

NB: the line connecting x and y is simply

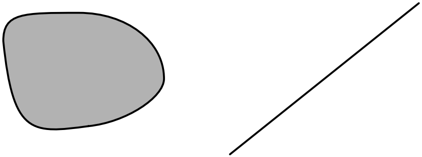
$$\{z \in \mathbb{R}^n \; : \; z = \lambda x + (1 - \lambda) y, \; 0 \leq \lambda \leq 1\}.$$

Definition (Convex set)

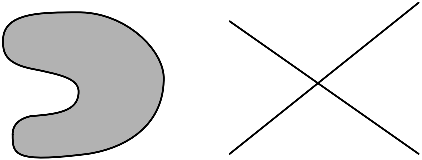
The set $\mathcal{C} \subset \mathbb{R}^n$ is called convex, if all convex combinations of any two points $x, y \in \mathcal{C}$ are again in $\mathcal{C}$.

Examples in $\mathbb{R}^2$

Convex sets:

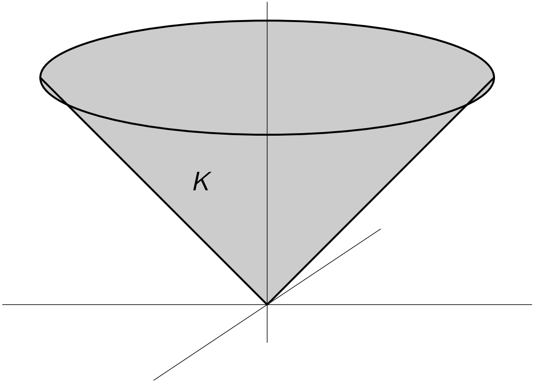


Non-convex sets:



Cones

The set $K \subset \mathbb{R}^n$ is a *convex cone* if it is a convex set and for all $x \in K$ and $\lambda > 0$ one has $\lambda x \in K$.



A convex cone is called *pointed* if it does not contain any subspace except the origin.

Conic optimization problem

Data:

- A convex cone $K \subset \mathbb{R}^n$;
- A linear operator $A : \mathbb{R}^n \to \mathbb{R}^m$;
- Vectors $c \in \mathbb{R}^n$ and $b \in \mathbb{R}^m$, and an inner product $\langle \cdot, \cdot \rangle$ on $\mathbb{R}^n$.

Conic optimization problem

$$\inf_{x \in K} \{\langle c, x \rangle \ : \ Ax = b\} .$$

One may solve conic optimization problems in polynomial time to any fixed accuracy using interior point methods, for certain convex cones ...

Choices for K

We consider the conic optimization problem for three choices of the cone K (or Cartesian products of cones of this type):

- Linear Programming (LP): K is the nonnegative orthant in $\mathbb{R}^n$:

$$\mathbb{R}^n_+ := \{x \in \mathbb{R}^n \; : \; x_i \geq 0 \; (i = 1, \ldots, n)\},$$

- Second order cone programming (SOCP): K is the second order (Lorentz) cone:

$$\left\{ \begin{bmatrix} x \\ t \end{bmatrix} \; : \; x \in \mathbb{R}^n, \; t \in \mathbb{R}, \; t \geq \|x\| \right\}.$$

- Semidefinite programming (SDP): K is the cone of symmetric positive semidefinite matrices.

The trace operator

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

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

Properties:

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

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

An inner product

We denote the cone of real symmetric $n \times n$ matrices by $\mathbb{S}^{n\times n}$. Let $A, B \in \mathbb{S}^{n\times n}$. Define

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

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

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

which is sub-multiplicative:

$$\|AB\| \le \|A\|\|B\|.$$

Positive semidefinite matrices

Theorem (Properties of p.s.d. matrices)

Let $X \in \mathbb{S}^{n\times n}$. The following are equivalent:

- $X \in \mathbb{S}^{n\times n}_{+}$ *or* $X \succeq 0$ *(X is p.s.d.);*
- $z^T X z \geq 0 \quad \forall z \in \mathbb{R}^n$*;*
- $\lambda_{\min}(X) \geq 0$*;*
- *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 ($X \succ 0$ or $X \in \mathbb{S}^{n\times n}_{++}$).

Linear regression

Recall the second order cone:

$$K := \left\{ \begin{bmatrix} x \\ t \end{bmatrix} \; : \; x \in \mathbb{R}^n, \; t \in \mathbb{R}, \; t \geq \|x\| \right\}.$$

The least squares solution of a linear system of equations $Ax = b$ is the solution of

$$\min_x \|Ax - b\| = \min\{t \; : \; t \geq \|Ax - b\|\}.$$

This is a second order cone programming problem (why?)

Robust LP

We consider an LP problem with ‘uncertain’ data.

Robust LP Problem

$$\min c^T x$$

subject to

$$a_i^T x \le b_i \ (i = 1, \ldots, m) \quad \forall a_i \in \mathcal{E}_i \ (i = 1, \ldots, m),$$

where the $\mathcal{E}_i$ are given ellipsoids:

$$\mathcal{E}_i = \{\bar{a}_i + P_i u \ : \ \|u\| \le 1\},$$

with P_i symmetric positive semidefinite.

Robust LP: SOCP formulation

We had

$$\mathcal{E}_i := \{\bar{a}_i + P_i u \ : \ \|u\| \leq 1\}.$$

Notice that

$$a_i^T x \leq b_i \ \forall a_i \in \mathcal{E}_i \iff \bar{a}_i^T x + \|P_i x\| \leq b_i \text{ (why?)}$$

Robust LP Problem: SOCP reformulation

$$\min c^T x$$

subject to

$$\bar{a}_i^T x + \|P_i x\| \leq b_i \ \ (i = 1, \ldots, m).$$

Note that this is indeed an SOCP problem.

Robust QP

We consider a quadratic programming (QP) problem with 'uncertain' objective.

Robust QP Problem

$$\min_{x} \max_{P \in \mathcal{E}} \; x^T P x + 2q^T x + r$$

subject to linear constraints, where $\mathcal{E}$ is the given ellipsoid:

$$\mathcal{E} = \{P_0 + \sum_{i=1}^{m} P_i u_i \; : \; \|u\| \leq 1\},$$

with the P_i's given symmetric positive semidefinite matrices.

Robust QP: SOCP formulation

Since

$$\mathcal{E} = \{P_0 + \sum_{i=1}^{m} P_i u_i \ : \ \|u\| \leq 1\},$$

one has

$$\max_{P \in \mathcal{E}} \ x^T P x = x^T P_0 x + \max_{\|u\| \leq 1} \sum_{i=1}^{m} x^T P_i x u_i.$$

Note that (Cauchy-Schwartz):

$$\max_{\|u\| \leq 1} \sum_{i=1}^{m} x^T P_i x u_i = \left(\sum_{i=1}^{m} (x^T P_i x)^2 \right)^{1/2}.$$

Robust QP: SOCP formulation (ctd.)

We obtained the reformulation:

$$\min_x x^T P_0 x + \left(\sum_{i=1}^m (x^T P_i x)^2\right)^{1/2} + 2q^T x + r.$$

This is the same as (why?)

$$\min\left\{t + v + 2q^T x + r \ : \ \|u\| \le t,\ x^T P_0 x \le v,\ x^T P_i x \le u_i \ (i = 1, \ldots, m)\right\}.$$

SOCP reformulation:

$$\min\left\{t + v + 2q^T x + r \ : \ \|u\| \le t, \left\| \begin{array}{c} 2P_i^{1/2} x \\ u_i - 1 \end{array} \right\| \le u_i + 1,\ u_i \ge 0 \ (i = 1, \ldots, m), \right.$$

$$\left. \left\| \begin{array}{c} 2P_0^{1/2} x \\ v - 1 \end{array} \right\| \le v + 1,\ v \ge 0. \right\}$$

Robust QP example: portfolio optimization

- You divide your capital over n possible investments (like stocks or bonds);
- x_i is the *fraction of your investment capital* you put in investment i;
- r_i is the expected return on investment i;
- $x^T Vx$ is the risk (or volatility) associated with your portfolio x;
- $r^T x$ is the expected return on your portfolio (prescribed to be at least α).

Markowitz QP model

$$\min_x \left\{ x^T Vx \ : \ \sum_{i=1}^{n} x_i = 1, \ r^T x \geq \alpha, \ x \geq 0 \right\}.$$

The covariance matrix V and vector r are calculated from historical performances of the n investments over a fixed period.

r and V are uncertain $\Rightarrow$ robust QP $\Rightarrow$ SOCP.

Lovász ϑ-function

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

Lovász ϑ-function

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

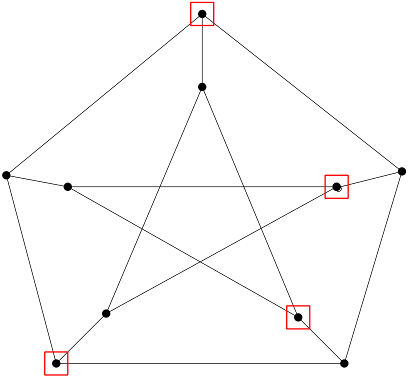
subject to

$$\begin{array}{rcl} X_{ij} & = & 0,\ \{i,j\} \in E\ (i \neq j) \\ \operatorname{trace}(X) & = & 1 \\ X & & \text{p.s,d.,} \end{array}$$

where e denotes the all-one vector.

Co-cliques

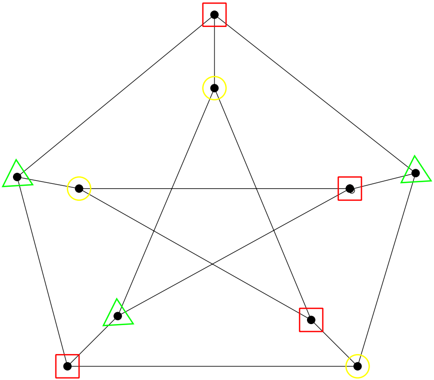
A *co-clique* of $G = (V, E)$ is a subset $V' \subset V$ such that the *induced subgraph* on V' has no edges.



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

Lovász 'sandwich theorem'

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

Theorem (Lovász's sandwich theorem)

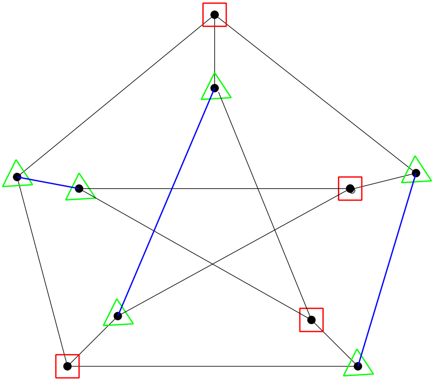
$$\alpha(G) \leq \vartheta(G) \leq \gamma(\bar{G}).$$

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 $|\text{MAX-}k\text{-CUT}|$ denote the cardinality of the maximum k cut.
One has

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

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

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

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

Definition

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:

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

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 $G_1 * G_2$. Thus

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

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 r-letter words in the dictionary is $\alpha(G^r)$.

Shannon capacity

Theorem (Lovász (1979))

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:

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

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

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

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 $\Theta(G) \le \vartheta(G) = \sqrt{5}$. In fact, one can show that $\Theta(G) = \sqrt{5}$.
- The Shannon capacity of the 7-cycle (heptagon) is not known.

S-lemma (Yakubovich, 1971)

When is one quadratic inequality implied by another quadratic inequality?

Let A_0 and A_1 be symmetric matrices and assume that $(x^0)^T A_1 x^0 > 0$ for some vector x^0.

Theorem (S-lemma)

The implication

$$x^T A_1 x \geq 0 \Rightarrow x^T A_0 x \geq 0$$

is valid if and only if

$$A_0 - \tau A_1 \ \textit{p.s.d.}$$

for some $\tau \geq 0$.

The question: "is $A_0 - \tau A_1$ p.s.d. for some τ?" can be answered using SDP.

Eigenvalue optimization

Notation: $\lambda_{\max}(A)$ denotes the *largest eigenvalue* of $A \in \mathbb{S}^{n\times n}$. Consider

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

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

for given $A_i \in \mathbb{S}^{n\times n}$ $(i = 0, \ldots, m)$.

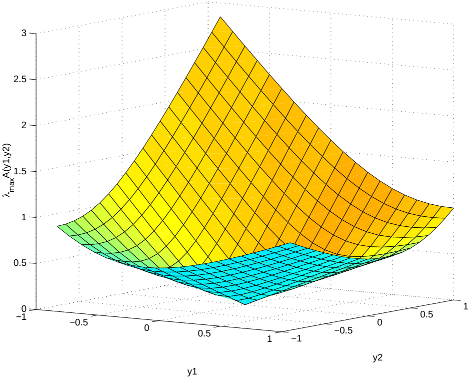
SDP reformulation

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

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

Eigenvalue 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 $y_1^* = y_2^* = 0$.

References and info

- The SOCP examples are from the online paper:
  M. Lobo, L. Vandenberghe, S. Boyd, H. Lebret, Applications of second-order cone programming. Linear Algebra and its Applications, 1998.
- The SDP examples are partly from the online paper:
  L. Vandenberghe, S. Boyd, Semidefinite programming. SIAM Review, 1996.
- The remaining SDP examples are described in:
  E. de Klerk. Aspects of Semidefinite Programming: Interior Point Algorithms and Selected Applications. Kluwer Academic Publishers, 2002.