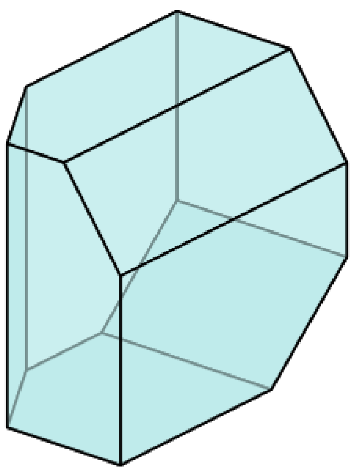
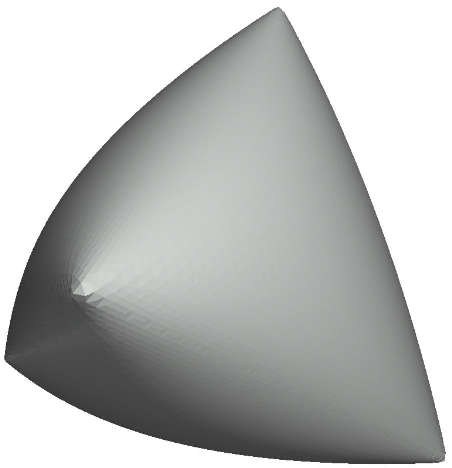
Linear and Semidefinite Optimization

Simone Naldi





Master ACSYON - Spring 2020
(code S8TQ158U)

Planning of the course

Online course (30h, 3 ECTS)

First week : March 23-27
Second week : March 30 - April 3
Third week : April 6-10
Fourth week : April 13-17
holidays : April 20-24
holidays : April 27-30
Fifth week : May 4-8 (Exam on thursday*)

** there could be modifications due to the covid-19 epidemic*

Online platform

The course material and all official communications are uploaded on the Moodle platform

Community

In particular, we are going to use the following online tools

- **Forum** – to post questions/answers, students are encouraged to use it
- **BBB (BigBlueBotton)** – for exam and some real-time classes
- **Devoir** – to send solutions to the exercise sheets

Grades

The final grade (over 20) is given by the formula

$$\frac{4}{10}TP + \frac{1}{10}P + \frac{5}{10}E$$

where

$TP =$ exercise sheets (due on the 3rd and the 5th weeks)
$P =$ class participation through Community
$E =$ written exam (5th week)

References for LP

Barvinok “A course in convexity”
Vol. 54. American Mathematical Soc., 2002.

Matousek-Gartner “Understanding and using Linear Programming” *Springer-Verlag, Berlin Heidelberg 2007*

Jansen “Theory of LP” *Vol. 6, Applied Optimization*

My slides available on Community or on my web page

References for SDP

Boyd-Vanderberghe "Semidefinite Programming"
SIAM Review – Vol. 38, No. 1, pp. 49-95, March 1996

De Klerk "Aspects of Semidefinite Programming: Interior Point Algorithms and Selected Applications" *Springer, 2006*

Anjos, Lasserre "Handbook on Semidefinite, Conic and Polynomial Optimization" *Springer Science & Business Media – Vol. 166, 2011*

My slides available on ☐ Community or on my ☐ web page

Outline of the course

1st week : Recaps and intro to LP

2nd week : Simplex and IPM for LP

3rd week : Preliminaries and intro to SDP

4th week : Algorithms for SDP

This course in a nutshell/1

Linear and Semidefinite Optimization (aka Linear and Semidefinite *Programming* or simply *LP, SDP*) are *convex conic* optimization problems. They model include several important classes of opt. problems (such as quadratic programs).

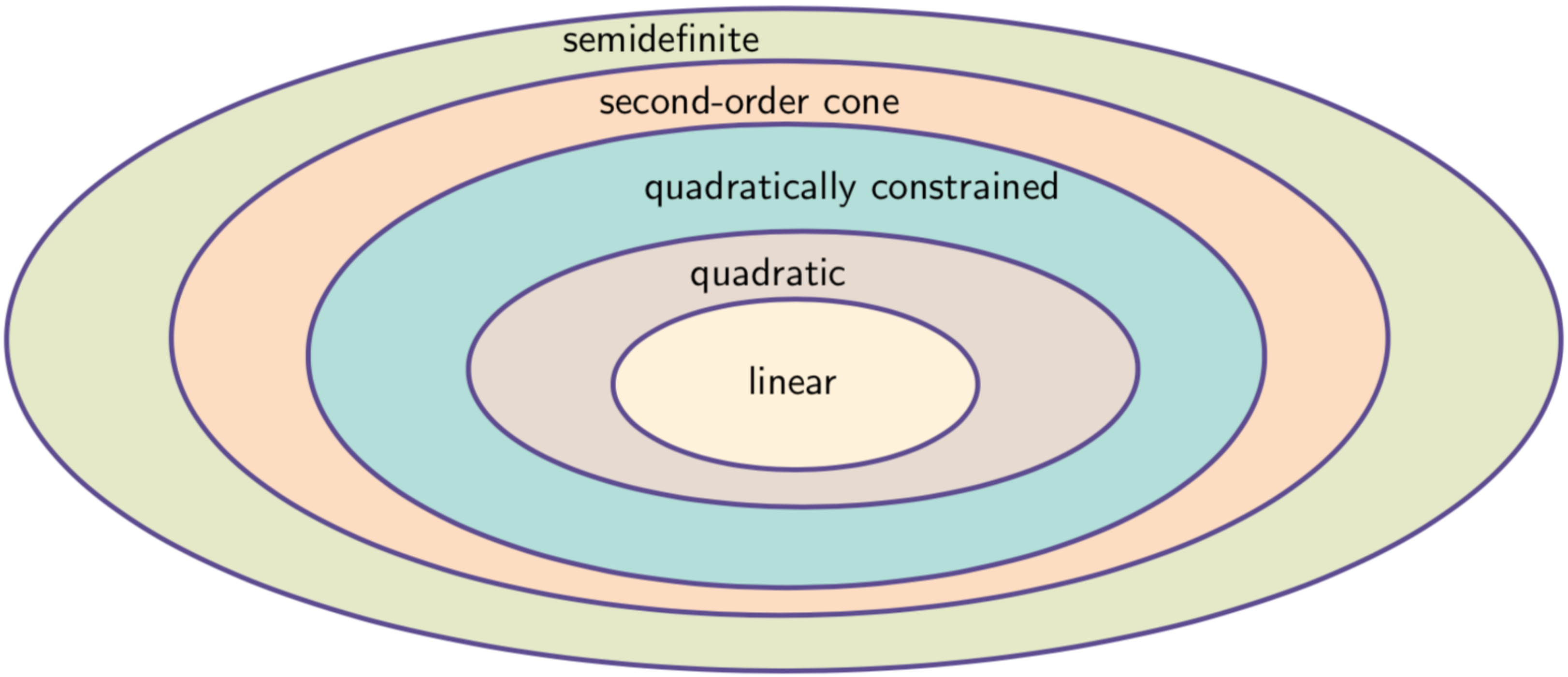
Several problems in mathematics can be cast *exactly* as LP or SDP. There are countless applications of such problems *e.g.* in: control theory, (combinatorial) optimization, quantum information . . .

This course in a nutshell/2

On the other hand, many classes of problems can be *relaxed* to (a sequence of) LP or SDP. For instance *polynomial optimization* (minimize a polynomial function over polynomial inequalities) or many *combinatorial optimization* problems (*e.g.* the MAX-CUT).

Barrier functions and IP methods are known for LP/SDP. Whereas LP can be solved in polynomial time, for SDP an approximate solution can be computed in polynomial time but, on the other hand, the complexity status of SDP in exact models (*e.g.* Turing) is still *unknown*.

A hierarchy of optimization problems



[courtesy of meboo.convexoptimization.com]

General notation

$\mathbb{Q}, \mathbb{R}, \mathbb{C}$: usual fields of rational, real and complex numbers.

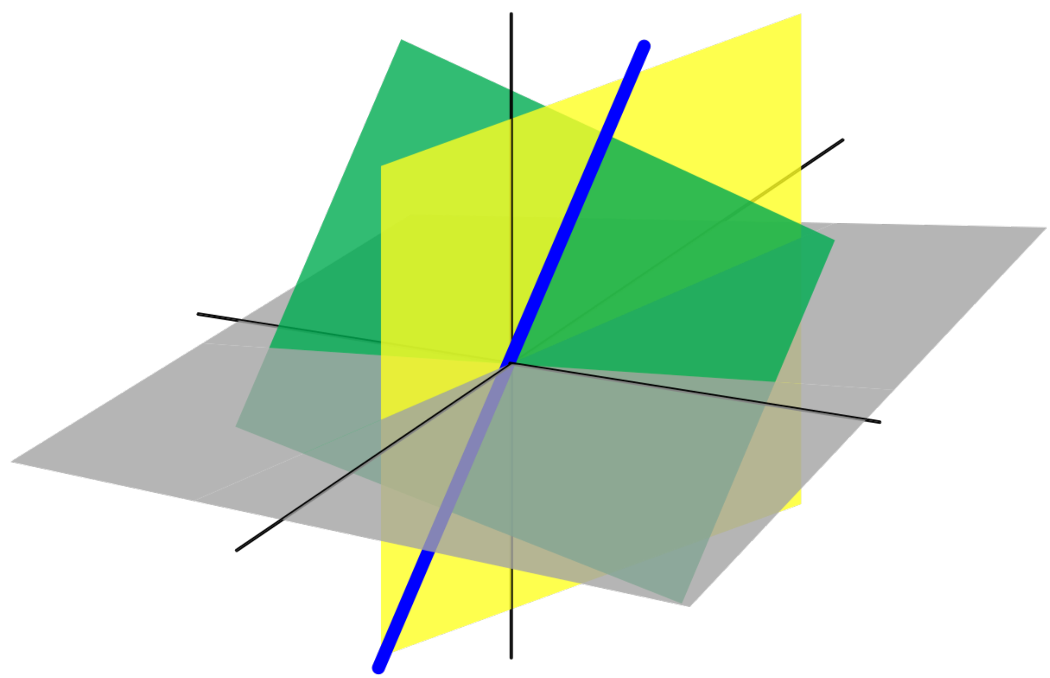
$\mathbb{R}_{\geq}, \mathbb{R}_{>}$: nonnegative/positive real numbers

$\mathcal{S}_d = \mathcal{S}_d(\mathbb{R})$: vector space of $d \times d$ real symmetric matrices

$\mathbf{K}$ is typically a (convex) cone (ex $\mathbf{K} = \mathbb{R}^n_{\geq}$: nonn. orthant)

PART I

Recaps on linear algebra and convex geometry and introduction to linear programming



Vector spaces

Basic objects of linear algebra are vector spaces V over a field $\mathbb{F}$ (for instance $\mathbb{F} = \mathbb{R}$ or $\mathbb{F} = \mathbb{C}$). A non-empty set V is a *vector space* if it satisfies the properties

$$\begin{aligned} &\alpha \in \mathbb{F}, v \in V \Rightarrow \alpha v \in V \\ &v_1, v_2 \in V \Rightarrow v_1 + v_2 \in V \end{aligned}$$

Elements $v_1, \ldots, v_n$ are linearly independent if

$$\lambda_1 v_1 + \cdots + \lambda_n v_n = 0 \quad \Rightarrow \quad \lambda_1 = \cdots = \lambda_n = 0$$

The supremum of all n for which there exist n linearly independent vectors in V is called the *dimension* of V (it can be $+\infty$).

Lines, planes, hyperplanes are examples of vector spaces.

Finite dimension

A real vector space V of dimension n is isomorphic to $\mathbb{R}^n$, but the isomorphism is non-canonical, that is, it depends on a choice of basis $B = \{v_1, \ldots, v_n\} \subset V$:

$$\begin{array}{rccc} \varphi : & V & \to & \mathbb{R}^n \\ & v_i & \mapsto & e_i \end{array}$$

and extended linearly $\varphi(\sum a_i v_i) := \sum a_i e_i$, where $e_1, \ldots, e_n$ is the "canonical basis" of $\mathbb{R}^n$ (e_i has i-th coordinate equal to 1 and 0 otherwise).

We will only consider real vector spaces, and often assume that a choice of a basis is done, hence we will consider that V is presented as $\mathbb{R}^n$ through some isomorphism φ.

Dual vector space

Let $V \approx \mathbb{R}^n$ be a fin.-dim real vector space. Its *dual vector space* is defined as the set of $\mathbb{R}$-linear forms defined on V, that is:

$$\boxed{V^\vee = \{\ell : V \to \mathbb{R} : \ell \text{ linear}\}}$$

$V^\vee$ is n-dimensional, hence (non-canonically) isomorphic to V.

Each element ℓ of $V^\vee$ can be associated to the null space

$$\ell^\perp = \{x \in V : \ell(x) = 0\} = \{x \in V : v_\ell^T x = 0\}$$

which is the hyperplane in V orthogonal to v_ℓ. We use this dual point of view to define objects in the primal vector space V (e.g. hyperplanes, subspaces, convex sets...)

Subspaces

Each *vector subspace* $U \subset V \approx \mathbb{R}^n$ has a *primal/image* definition and a *dual/kernel*) definition, as follows:

$$\text{Primal/image}: U = \{My : y \in V\} = span(M^{(1)}, \ldots, M^{(n)})$$

$$\text{Dual/kernel}: U = \{x : Nx = 0\} = N^{\perp}_{(1)} \cap \cdots \cap N^{\perp}_{(n)}$$

where $M, N \in \mathbb{R}^{n \times n}$ are matrices, that is linear maps $V \to V$. Compactly this means that $U = \text{im}(M) = \text{ker}(N)$.

Sets of type $p + U$ where $p \in V$ and $U \subset V$ vec subspace, are called *affine subspaces*.

Theorem of the alternative in linear algebra

This result was proved by Gauss (1809) and represents an example of "theorem of the alternative"

Theorem. Let $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^{m \times n}$. The system $Ax = b$ has a solution if and only if

$$(*) \qquad \boxed{y^T A = 0 \quad \Rightarrow \quad y^T b = 0}$$

proof. If $b = Ax$ for some x, then $y^T b = (y^T A)x$, hence (*) holds trivially. For the converse, let $U = \{Ax : x \in \mathbb{R}^n\} = \mathrm{im}(A) \subset \mathbb{R}^m$ be given in its primal representation. Then there exists $B \in \mathbb{R}^{n \times m}$ such that $U = \{y \in \mathbb{R}^m : By = 0\} = \ker(B)$. Thus $BA = 0$ (the null matrix). If (*) holds, then $Bb = 0$, hence $b \in \ker(B) = \mathrm{im}(A)$.

If $Ax = b$ is feasible, then there is a solution y to the alternative system : $y^T A = 0, y^T b \neq 0$, that "separates" b from $\mathrm{im}(A)$.

Metric spaces

One can define norms and distances on finite-dimensional vector spaces. On $V = \mathbb{R}^n$ the Euclidean inner product is given by

$$v, w \in \mathbb{R}^n \quad : \quad \langle v, w \rangle := \sum_i v_i w_i$$

and the corresponding Euclidean norms and distances

$$||v|| = \sqrt{\langle v, v \rangle} \qquad \text{and } d(v, w) = ||v - w|| \, .$$

Different norms can be defined over $\mathbb{R}^n$, all are asymptotically equivalent in the sense that for two norms $|| \cdot ||_1$ and $|| \cdot ||_2$ there are two constants c_1, c_2 s.t.

$$c_1 ||v||_1 \leq ||v||_2 \leq c_2 ||v||_2, \quad \text{for all } v \in \mathbb{R}^n.$$

Convexity

A set $C \subset \mathbb{R}^n$ is *convex* if

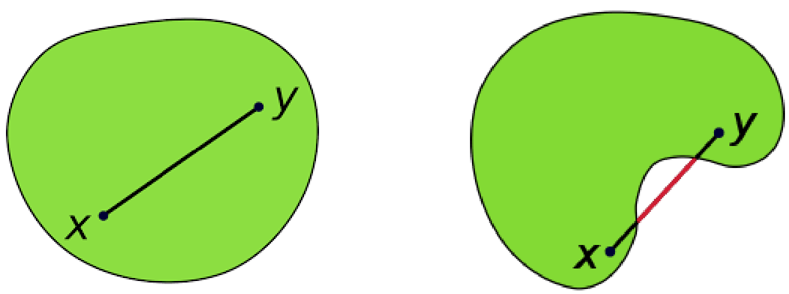
$$\forall\, x, y \in C,\ \forall t \in [0,1] \quad \text{one has} \quad tx + (1-t)y \in C$$

– closed under intersection, *not* closed under union

– closed under *Minkowski sum*: $C_1 + C_2 = \{x + y : x \in C_1, y \in C_2\}$

– $A \subset V$ convex, and $f : V \to W$ linear, then $f(A)$ is convex

– linear, affine subspaces, disks, segments, balls ... are convex



The smallest (for inclusion) convex set containing a set S is called the *convex hull* of S and denoted by $\mathrm{conv}(S)$.

Separation

In convex geometry there are several results that yield "separation properties" for pairs of (convex) sets. The main question is, given two (convex) sets S_1, S_2, can we "separate" them by providing a (elementary) function f that is positive on S_1 and negative on S_2 ?

For an affine hyperplane $H_0 = \{x \in \mathbb{R}^n : \ell(x) = a\}$, we denote by

$$H_0^+ = \{x \in \mathbb{R}^n : \ell(x) > a\} \qquad H_0^- = \{x \in \mathbb{R}^n : \ell(x) < a\}.$$

Isolation theorem. Let $S \subset \mathbb{R}^n$ be a non-empty convex set, and let $a \notin S$. Then there is an affine hyperplane $H_0 \subset \mathbb{R}^n$ which contains a ($H_0 = a + H$ where H is a hyperspace) and such that S belongs to H_0^+ or H_0^-.

Geometry of convex sets

We work with the Euclidean topology (a base is given by open balls). The interior of a convex set is convex, which comes as a corollary of the following

Theorem. Let $S \subset \mathbb{R}^n$ be convex, and $u \in Int(S)$. Then for all $v \in S$, and for all $\lambda \in [0, 1)$, one has $(1 - \lambda)u + \lambda v \in Int(S)$.

If a set S has empty interior, one can *a priori* find an affine space that contains S and such that the "relative" interior is non-empty.

Theorem. Let $S \subset \mathbb{R}^n$ be convex, and assume $Int(S) = \emptyset$. Then there is a proper affine space $L \subset \mathbb{R}^n$ such that $S \subset L$.

The *dimension* of a (convex) set S is the dimension of its affine hull, that is, the smallest affine space that contains it.

Cones

A set $\mathbf{K} \subset \mathbb{R}^n$ is a *(convex) cone* if

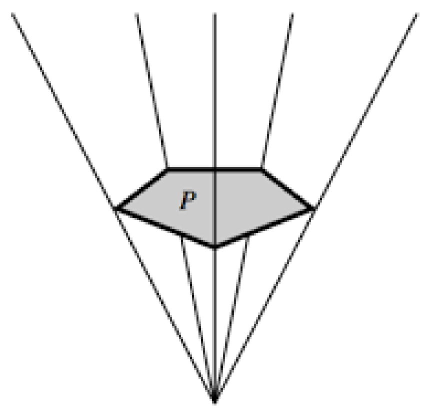
$$a, b \in \mathbf{K} \text{ and } \alpha, \beta \in \mathbb{R}_{\geq 0} \Rightarrow \alpha a + \beta b \in \mathbf{K}$$

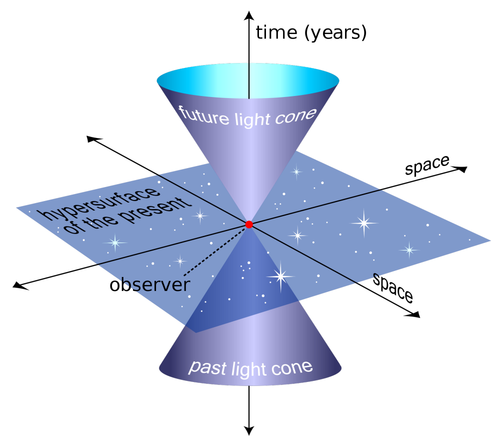
The smallest cone containing S is called its *conical hull* and denoted by cone(S). Interesting examples of cones :

the nonnegative orthant : $\mathbb{R}^n_{\geq} = \{x \in \mathbb{R}^n : x_i \geq 0, \forall\, i\}$

polyhedral cones : $\text{cone}(\{v_1, \ldots, v_s\})$

the second-order cone : $\mathscr{L}^n = \{(x,t) \in \mathbb{R}^{n+1} : \|x\| \leq t\}$



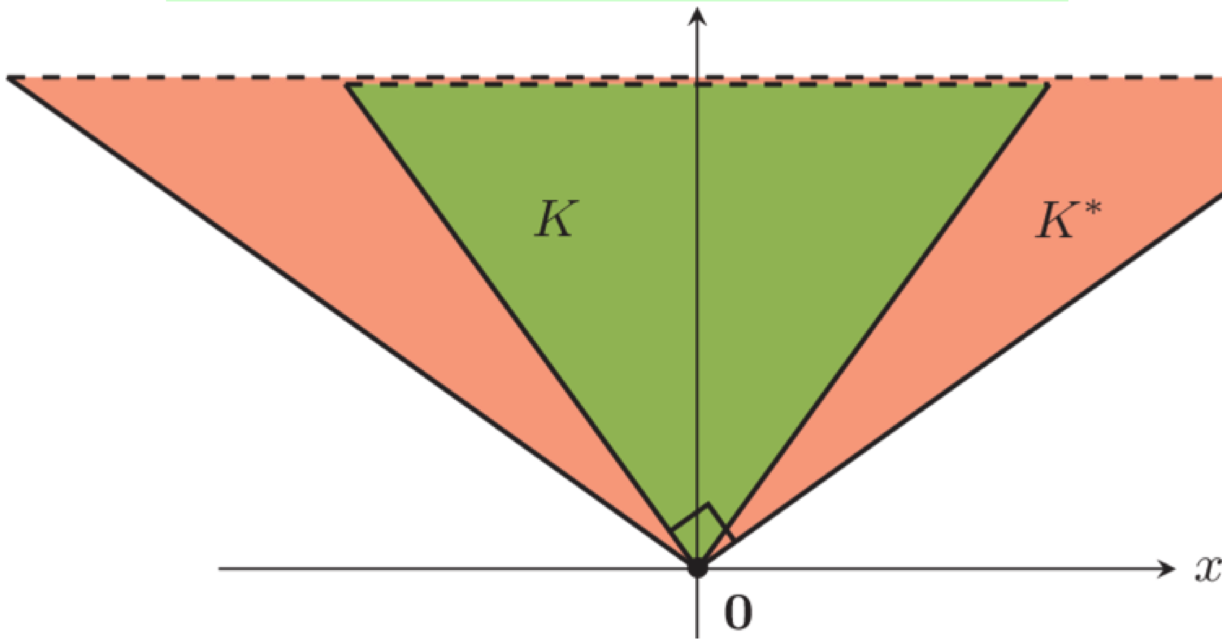


Dual cones

Let $\mathbf{K} \subset V$ be a convex cone in a $\mathbb{R}-$vector space V. Let $V^\vee$ be the *dual vector space* of V (space of $\mathbb{R}-$linear functionals on V).

The *dual cone* of $\mathbf{K}$ is the set of those $\mathbb{R}-$linear functionals that take nonnegative values on $\mathbf{K}$:

$$\mathbf{K}^* = \{\ell \in V^\vee : \ell(x) \geq 0, \ \forall x \in \mathbf{K}\} \subset V^\vee$$



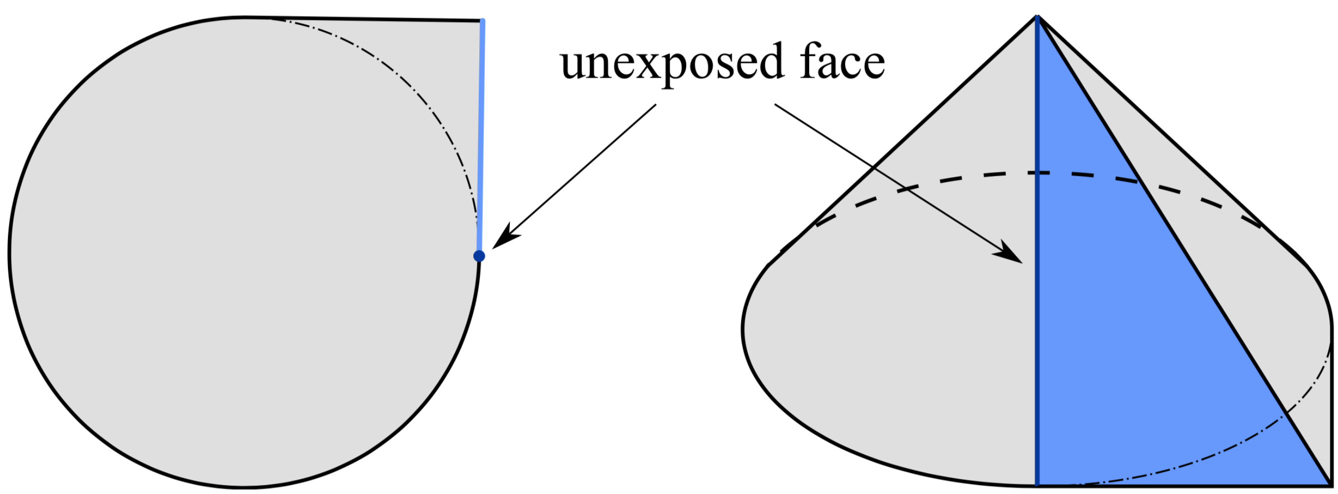
A cone $\mathbf{K}$ is *self-dual* whenever $\mathbf{K} = \mathbf{K}^*$ under isomorphism.

Faces

Let $S \subset V$ be a convex set. A convex subset $F \subset C$ is called a *face of* C if for every $x \in F$ and $y, z \in C$ one has

$$\text{if } x \in (y,z) := \{ty + (1-t)z : t \in (0,1)\}, \text{ then } y, z \in F.$$

A face is *exposed* if there exists an affine space $L \subset V$ such that $F = S \cap L$. For $C = \mathbb{R}^n_{\geq}$ (and polyhedral cones), $\mathcal{S}^+_d$ and $\mathscr{L}^n$ every face is exposed.



Extreme points

Let $C \subset \mathbb{R}^n$ be a convex set. A point $x \in C$ is called an *extreme point* if the following holds

$$\text{if } x \in (y,z) := \{ty + (1-t)z : t \in (0,1)\}, \text{ then } y = z = x.$$

Two crucial theorems :

Krein-Milman theorem. A compact convex set is the convex hull of the set of its extreme points.

Theorem (Minima of linear functions). Let $C \subset V$ be convex and $f : V \to \mathbb{R}$ be linear ($f \in V^\vee$). Let $c^* = \min_C f(x)$, and let $F_* = \{x \in C : f(x) = c^*\}$. Let $u \in F_*$ an extreme point of F_*. Then u is an extreme point of C.

Polyhedra

Polyhedra are the simplest example of convex sets after linear spaces. In a nutshell, a *polyhedron* $P \subset \mathbb{R}^n$ is a finite intersection of half spaces:

$$\boxed{P = \{x \in \mathbb{R}^n : Ax \leq b\}}$$

where $A \in \mathbb{R}^{m \times n}, b \in \mathbb{R}^m$, and "$Ax \leq b$" means that $b - Ax \in \mathbb{R}^m_{\geq}$. This is a *dual* definition because it uses linear inequalities on $\mathbb{R}^n$.

For $m = 1$ one gets hyperspaces (which are special cases of polyhedra). More generally one has intersections of hyperspaces.

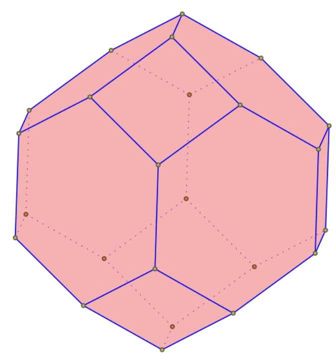
Polyhedra can be unbounded, but are convex and their faces are exposed. Very interesting combinatorial objects (cf. the book by Ziegler, out of scope of this course).

A polytope

If the polyhedron is *compact* (in which case it is also called a *polytope*) there is a *primal* definition through convex hulls:

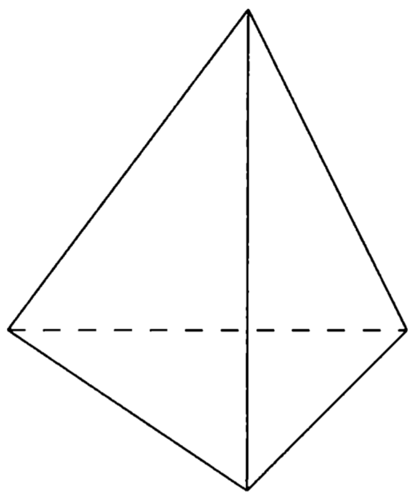
$$P = \mathrm{conv}(\{a_1, \ldots, a_k\})$$

for fixed $a_1, \ldots, a_k \in \mathbb{R}^n$.

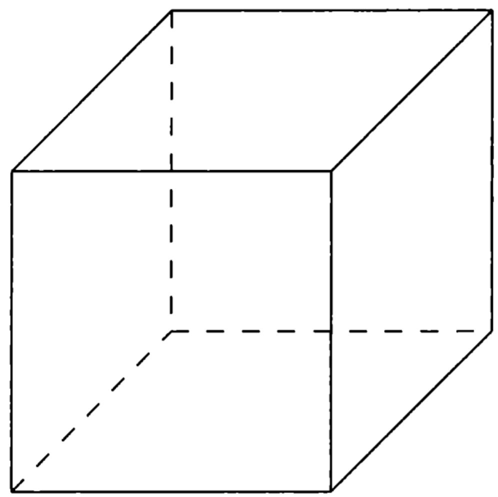


Representation theorem. A set $P \subset \mathbb{R}^n$ is a polyhedron if and only if $P = Q + C$ where $Q \subset \mathbb{R}^n$ is a polytope and $C \subset \mathbb{R}^n$ is a polyhedral cone.

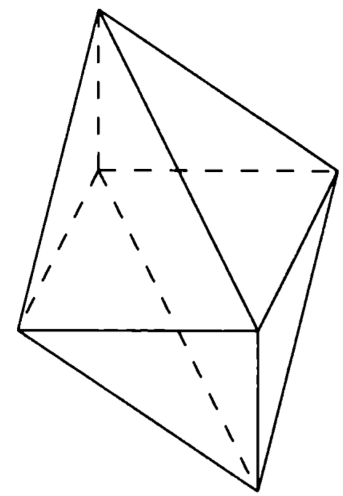
Example: Platonic solids



simplex



cube



octahedron (crosspolytope)

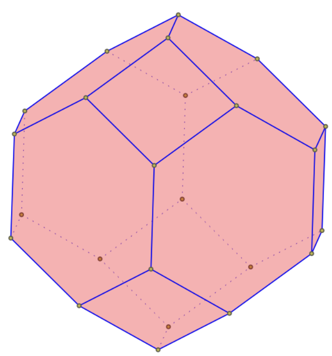
Vertices

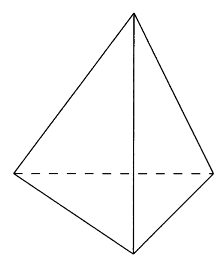
The extreme points of polyhedra are called *vertices*. They are characterized by the following theorem.

Theorem. Let $P = \{x \in \mathbb{R}^n : \langle a_i, x \rangle \leq b_i, i = 1, \ldots, m\}$ be a polyhedron. For $p \in P$, let

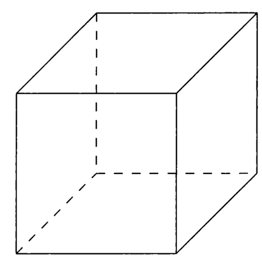
$$I(p) = \{i : \langle a_i, p \rangle = b_i\}$$

be the set of active constraints at p. Then p is a vertex if and only if $\{a_i : i \in I(p)\}$ spans $\mathbb{R}^n$.

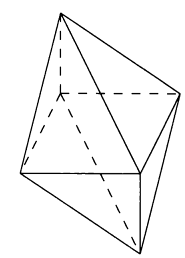




simplex



cube



octahedron (crosspolytope)

Slack variables/1

Every polyhedron can be seen as a section of the nonnegative orthant $\mathbb{R}^m_{\geq}$. This is done by adding what we call *slack variables*, that is, additional variables that allow to write the polyhedron P in a special form.

We have seen that

$$P = \{x \in \mathbb{R}^n : Ax \leq b\} = \{x \in \mathbb{R}^n : b - Ax \in \mathbb{R}^m_{\geq}\}.$$

Let $s = (s_1, \ldots, s_m)$ be new variables. Then one has

$$b - Ax \in \mathbb{R}^m_{\geq} \quad \Leftrightarrow \quad b - Ax = s \text{ and } s \in \mathbb{R}^m_{\geq}$$

Slack variables/2

Hence the original x variables are now free from conical constraints, but have to satisfy some linear equations, and the conical constraint is pushed to variables s.

Geometrically this corresponds to lifting P to a second polyhedron $P^* \subset \mathbb{R}^{n+m}$ so that P is the image of P^* through the projection on variables x:

$$P^* = \{(x, s) \in \mathbb{R}^n \times \mathbb{R}^m : b - Ax - s = 0, s \in \mathbb{R}^m_{\geq}\}$$

$$P = P^* \cap \mathbb{R}^n = \{x \in \mathbb{R}^n : \exists\, s \in \mathbb{R}^m : b - Ax - s = 0, s \in \mathbb{R}^m_{\geq}\}$$

Back to faces of polyhedra

For a system of inequalities $Ax \leq b$, we say that $A'x \leq b'$ is a subsystem whenever A' is obtaining from A by eliminating some rows, and the same for b' (corresponding rows).

Characterization of faces. Let P be a polyhedron defined by $Ax \leq b$, and $F \subset P$. Then F is a face of P if and only if one of the following (equivalent) properties hold :

– there exists a subsystem $A'x \leq b'$ of P such that

$$F = \{x \in P : A'x = b'\}$$

– there exists a vector $c \in \mathbb{R}^n$ for which F is the set of points in P where $\max_P c^T x$ is attained, that is

$$F = \arg\max_P c^T x$$

Conic optimization (settings) /1

Let $(V, V^\vee)$ be a dual pair with a duality pairing $\langle \cdot\,,\,\cdot \rangle_V$, namely a non-degenerate bilinear map

$$\langle \cdot\,,\,\cdot \rangle_V : V^\vee \times V \to \mathbb{R}$$

that formally acts as follows:

$$\langle \ell, v \rangle_V := \ell(v).$$

For $V \approx \mathbb{R}^n \approx V^\vee$ this is just the standard inner product $\sum_i \ell_i v_i$.

Let $\mathbf{K} \subset V$ be a convex cone in V, and let $\mathbf{K}^* \subset V^\vee$ be its dual cone, with respect to $\langle \cdot\,,\,\cdot \rangle_V$, meaning that

$$\boxed{\mathbf{K}^* = \{c \in V \,:\, \langle c, x \rangle_V \geq 0, \ \forall x \in \mathbf{K}\}}$$

Conic optimization (settings) /2

Let now W be a second real vector space.

Let $\mathscr{A}: V \to W$ be a linear map to a vector space W, and let $\langle\cdot,\cdot\rangle_W$ a duality pairing for $(W, W^\vee)$. Let $\mathscr{A}^*: W^\vee \to V^\vee$ be an adjoint of $\mathscr{A}$, that is an operator satisfying

$$\langle\mathscr{A}(x), y\rangle_W = \langle x, \mathscr{A}^*(y)\rangle_V \qquad \text{for all } x \in V,\ y \in W^\vee.$$

This is unique when $\dim_{\mathbb{R}}(V) < \infty$. For instance in LP, $\mathscr{A}(x) = Ax$ (matrix-vector multiplication) and $\mathscr{A}^*(y) = A^T y$.

In finite dimension, $\mathscr{A}$ is matrix-vector multiplication and $\mathscr{A}^*$ is its adjoint (transpose for real v.s.).

Primal conic program

The primal conic program in standard form reads as follows:

$$\begin{array}{rlll} p^* & := & \inf & \langle c, x\rangle_V \\ & & \text{s.t.} & \mathscr{A}(x) = b \\ & & & \boxed{x \in \mathbf{K}} \end{array}$$

Recall that $\mathscr{A}$ is linear hence $\mathscr{A}(x) = (\langle a_1, x\rangle_V, \ldots, \langle a_m, x\rangle_V)$.

In other words, CP is the problem of minimizing a linear function over affine sections of convex cones, in a given real v.s. V. Every such problem can be written in the form above for a cone $\mathbf{K}$.

Examples

Conic programming includes classical optimization problems :

– $\mathbf{K} = \mathbb{R}^n_{\geq}$ (non-negative orthant, linear programming)

– $\mathbf{K} = \mathscr{L}^n$ (Lorentz cone, second-order cone programming)

– $\mathbf{K} = \mathcal{S}^+_d$ (psd cone, semidefinite programming)

– Finite products : $\mathbf{K} = \bigtimes_i \mathbb{R}^{n_i}_{\geq} \times \bigtimes_j \mathscr{L}^{n_j} \times \bigtimes_\ell \mathcal{S}^+_{d_\ell}$

– The cone of nonnegative (multivariate) polynomials and the cone of sum-of-squares (multivariate) polynomials

Lagrangian function

Define the Lagrangian function (y is a Lagrange multiplier) :

$$
\begin{aligned}
\mathscr{L}(y) &:= \inf_{x\in\mathbf{K}} \langle c, x\rangle_V + \langle b - \mathscr{A}(x), y\rangle_W \\
&= \langle b, y\rangle_W + \inf_{x\in\mathbf{K}} \langle c, x\rangle_V - \langle \mathscr{A}(x), y\rangle_W \\
&= \langle b, y\rangle_W + \inf_{x\in\mathbf{K}} \langle c - \mathscr{A}^*(y), x\rangle_V
\end{aligned}
$$

Denote $s := c - \mathscr{A}^*(y)$, so that $\mathscr{L}(y) = \langle b, y\rangle_W + \inf_{x\in\mathbf{K}} \langle s, x\rangle_V$.
Remark that :

If $s \in \mathbf{K}^*$ then the above inf is equal to 0, whereas

if $s \notin \mathbf{K}^*$, then the above inf is equal to $-\infty$.

Lagrangian dual conic program

The Lagrangian dual problem asks to maximize the Lagrangian function over its domain :

$$\begin{array}{rcll} d^* & := & \text{sup} & \langle b, y \rangle_W \\ & & \text{s.t.} & c - \mathscr{A}^*(y) = s \\ & & & s \in \mathbf{K}^* \end{array}$$

The dual variables are (y, s) – we use s as a slack variable in order to have linear ($c - \mathscr{A}^*(y) = s$) and conical constraints ($s \in \mathbf{K}^*$) so that *the dual of a conic program is still a conic program*.

Feasibility and attainability

The following sets can be defined from primal input data :

$\mathcal{P} := \{x \in V : x \in \mathbf{K} \text{ and } \mathscr{A}(x) = b\}$ is the *primal feasible set*

$\mathcal{D} := \{y \in W^{\vee} : c - \mathscr{A}^*(y) \in \mathbf{K}^*\}$ is the *dual feasible set*

Let $L = \{x \in V : \mathscr{A}(x) = b\}$ (so that $\mathcal{P} = \mathbf{K} \cap L$) and suppose that $Int(\mathbf{K}) \neq \emptyset$. We say the primal conic program is *feasible* if $\mathbf{K} \cap L \neq \emptyset$ and in particular

strongly feasible if $Int(\mathbf{K}) \cap L \neq \emptyset$

weakly feasible if feasible but $Int(\mathbf{K}) \cap L = \emptyset$

We say it is *infeasible* if $\mathbf{K} \cap L = \emptyset$ and in particular

strongly infeasible if $d(\mathbf{K}, L) > 0$

weakly infeasible if infeasible but $d(\mathbf{K}, L) = 0$

We say the infimum is *attained* if $\exists \overline{x} \in \mathcal{P}$ s.t. $p^* = \langle c, \overline{x} \rangle_V$

Weak and strong duality

Suppose that $\mathcal{P} \neq \emptyset$ and $\mathcal{D} \neq \emptyset$ (primal and dual feasibility), and let $\overline{x} \in \mathcal{P}$ and $\overline{y} \in \mathcal{D}$. We know that $\overline{s} = c - \mathscr{A}^*(\overline{y}) \in \mathbf{K}^*$, hence $\langle \overline{s}, \overline{x} \rangle_V \geq 0$. Hence one gets

$$\langle c, \overline{x} \rangle_V = \langle \mathscr{A}^*(\overline{y}) + \overline{s}, \overline{x} \rangle_V = \langle \overline{y}, \mathscr{A}(\overline{x}) \rangle_W + \langle \overline{s}, \overline{x} \rangle_V = \langle \overline{y}, b \rangle_W + \langle \overline{s}, \overline{x} \rangle_V \geq \langle \overline{y}, b \rangle_W$$

which implies that the following *weak duality* always holds :

$$p^* = \inf_{x \in \mathcal{P}} \langle c, x \rangle_V \geq \sup_{y \in \mathcal{D}} \langle y, b \rangle_W = d^*$$

Conclusion : the dual optimal value d^* gives the best lower bound for the primal optimal value, and viceversa p^* gives the best upper bound for d^*.

We say that *strong duality* holds whenever $p^* = d^*$.

Example : absence of strong duality

In LP strong duality always holds, but this is false in general conic programs, as shown by this example.

Let $\mathbf{K} = \mathscr{L}^2 \times \mathbb{R}_{\geq}$. One has $\mathbf{K} = \mathbf{K}^*$, as product of self-dual cones. Consider the primal-dual conic program

$$p^* = \inf -x_1 \,:\, x_1 + x_4 = 1, x_2 + x_3 = 0,\ x \in \mathbf{K}$$
$$d^* = \sup y_1 \,:\, y_1 + s_1 = -1, y_2 + s_2 = y_2 + s_3 = y_1 + s_4 = 0,\ s \in \mathbf{K}^*$$

Remark that $\boxed{p^* = 0}$ since for $(x, y, z, w) \in \mathbf{K} :\ x_1^2 \leq x_3^2 - x_2^2 = 0$, and $(0, 0, 0, 1)$ is an optimal primal solution. The dual can be rewritten equivalently

$$d^* = \sup y_1 \,:\, -(1 + y_1, y_2, y_2, y_1) \in \mathbf{K}^*$$

and since $\mathbf{K} = \mathbf{K}^*$ one gets $(1 + y_1)^2 \leq y_2^2 - y_2^2 = 0$ hence

$$\boxed{d^* = -1}$$

Sufficient conditions for strong duality

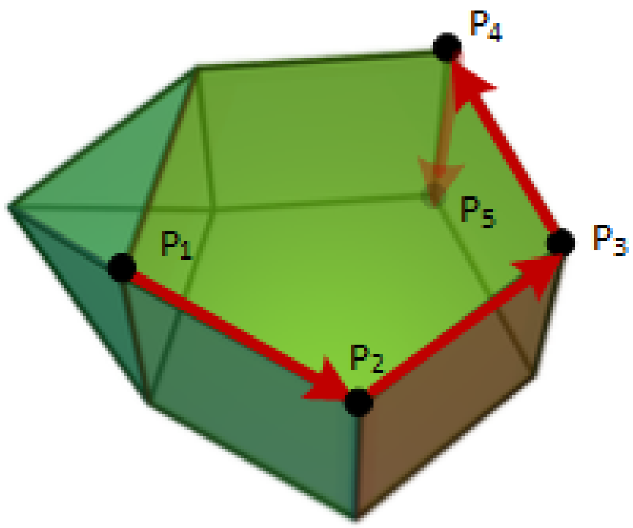
Theorem (Slater's condition)

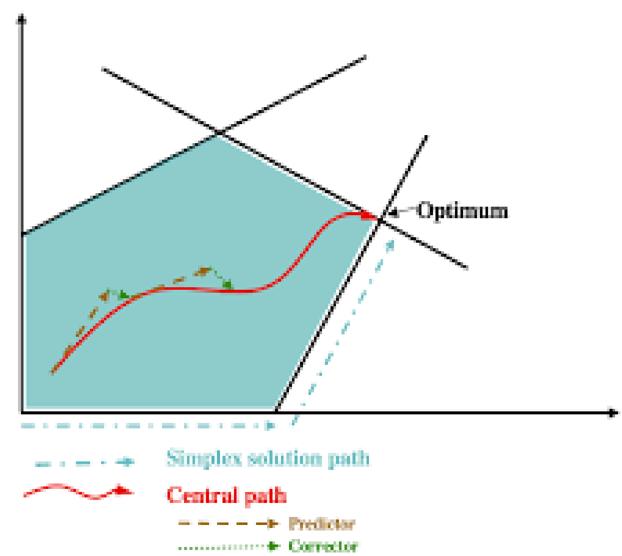
- If the primal program is strongly feasible and the dual is feasible, then $p^* = d^*$ and the dual supremum is attained.
- If the dual program is strongly feasible and the primal is feasible, then $p^* = d^*$ and the primal infimum is attained.

In the example above, the primal and dual feasible sets do not contain interior points of the primal and dual cones, hence Slater's condition does not hold and strong duality cannot be guaranteed (indeed, it fails).

PART II

Linear programming, simplex and interior-point methods





Goals of this part/week

- Give a precise description of LP, and of its duality

- Simplex method: construction

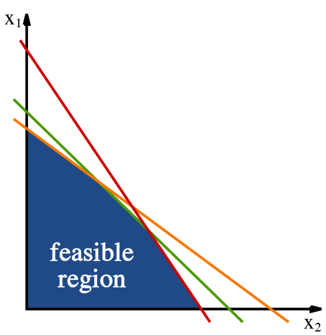
- Interior-point algorithms, barrier functions, central path...

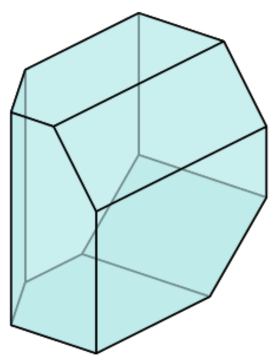
The primal linear program

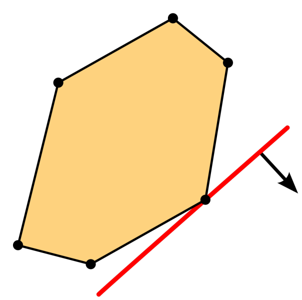
A *linear program* is the minimization of a linear function over linear inequalities in $\mathbb{R}^n$: for $c \in \mathbb{R}^n$, $b \in \mathbb{R}^m$, $A \in \mathbb{R}^{m\times n}$ solve

$$\begin{array}{rlll} p^* & := & \text{inf} & c^T x := c_1 x_1 + \cdots + c_n x_n \\ & & \text{s.t.} & Ax = b \quad \text{(or equiv. } Ax \le b\text{)} \\ & & & \boxed{x_i \ge 0 \quad i = 1, \ldots, n} \end{array}$$

Remark that A defines a linear map $\mathbb{R}^n \to \mathbb{R}^m$, and that the last constraint can be cast as $\boxed{x \in \mathbb{R}^n_{\ge 0}}$, the nonnegative orthant.







The dual linear program

The Lagrangian dual of a linear program (as for the general conic program) reads as follows:

$$\begin{array}{rcll} d^* & := & \text{sup} & b^T y := b_1 y_1 + \cdots + b_m y_m \\ & & \text{s.t.} & c - A^T y = s \quad \text{(slack vars)} \\ & & & \boxed{s_i \geq 0 \quad i = 1, \ldots, m} \end{array}$$

Now the linear map is the transpose $A^T : (\mathbb{R}^m)^\vee \to (\mathbb{R}^n)^\vee$, and the constraint on the slack variables can be cast as

$$\boxed{s \in (\mathbb{R}^n_{\geq 0})^* \approx \mathbb{R}^n_{\geq 0}.}$$

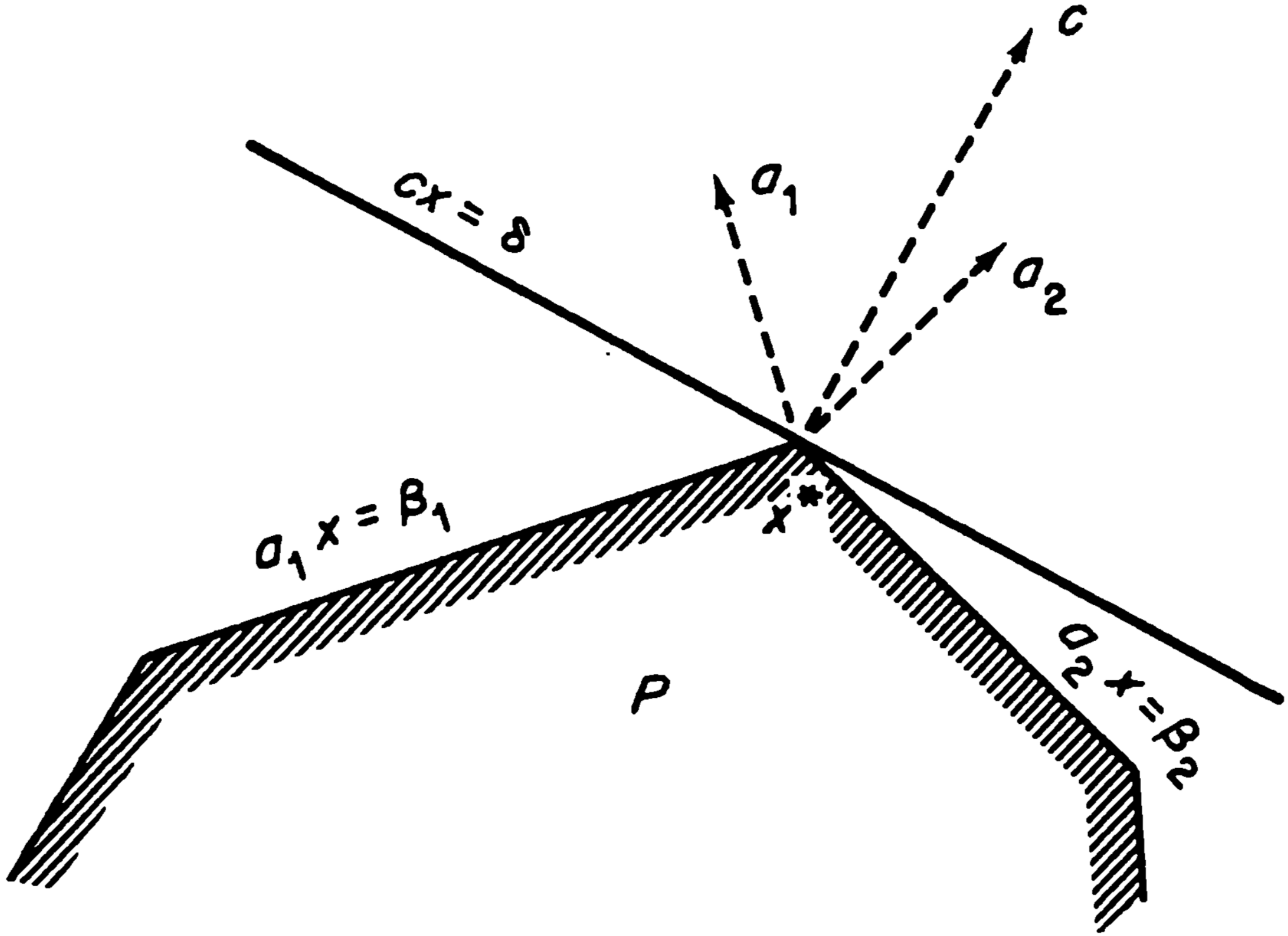


Image from Schrijver's book "Theory of LP and IP"

Farkas lemma

A powerful result, a theorem of the alternative for polyhedra:

either a system of linear inequalities is feasible, or a second one is feasible and certifies the infeasibility of the first one.

Theorem (Farkas, Minkowski). Let $b \in \mathbb{R}^m$ and $A \in \mathbb{R}^{m\times n}$. One of the two following systems is feasible

- There exists $x \in \mathbb{R}^n$ such that $Ax = b$ and $x \in \mathbb{R}^n_{\geq}$
- There exists $y \in \mathbb{R}^m$ such that $A^T y \in \mathbb{R}^n_{\geq}$ and $y^T b < 0$

Its natural generalization to CP (already in SDP) is false.
By the way there exist versions for the general conic program.

Absence of weak infeasibility

Theorem. An infeasible LP $Ax = b, x \geq 0$ is strongly infeasible.

proof. This essentially comes from Farkas lemma. Let $C = \text{cone}(A^{(1)}, \ldots, A^{(n)})$ is the (polyhedral) cone generated by the columns $A^{(1)}, \ldots, A^{(n)}$ of A. Then if the program $Ax = b, x \geq 0$ is infeasible, it means that $b \notin C$. By Farkas lemma, there is a vector y such that $A^T y \geq 0$ and $y^T b < 0$. The first condition means that $p^T y \geq 0$ for all $p \in C$ (since C is generated by the columns of A) and the second condition implies hence that b is strongly separated from C.

Strong duality in LP

As for every conic program, weak duality holds and implies that if the primal and dual programs are feasible, $p^* \geq d^*$.

It happens that LP is a "regular" conic program in the sense that **strong duality always holds** (with feasibility assumption).

Theorem (Strong duality). Let $c \in \mathbb{R}^n, b \in \mathbb{R}^m$ and $A \in \mathbb{R}^{m \times n}$. Then strong duality holds for the corresponding LP, that is

$$d^* = \sup\{b^T y : A^T y \leq c\} = \inf\{c^T x : x \geq 0, Ax = b\} = p^*$$

provided both sets are feasible.

Equivalent formulations of LP

One can represent linear programs in different settings, essentially varying equalities and inequalities, and membership to nonnegative orthant.

In particular the following systems are admissible formats for a (dual) linear program:

$$\sup b^T y \quad \text{subject to} \quad A^T y \leq c$$

$$\sup b^T y \quad \text{subject to} \quad A^T y \leq c, y \geq 0$$

$$\sup b^T y \quad \text{subject to} \quad A^T y = c, y \geq 0$$

$$\sup b^T y \quad \text{subject to} \quad A^T y \geq c$$

$$\sup b^T y \quad \text{subject to} \quad A^T y \geq c, y \geq 0$$

$$\sup b^T y \quad \text{subject to} \quad A^T y \geq c, y \geq 0$$

Complementary slackness

Important duality property for linear programs.

Recall that strong duality holds:

$$p^* = \inf_{Ax=b, x\geq 0} c^T x = \sup_{A^T y \leq c} b^T y = d^*.$$

The following are equivalent for two feasible points x_0, y_0:

(i) x_0 and y_0 are solutions to the primal-dual LP

(ii) $c^T x_0 = b^T y_0$

(iii) *complementary slackness* for each component of x_0, either it is 0 or the corresponding inequality in $A^T y \leq c$ is 0, that is

$$x_0^T (c - A^T y_0) = 0$$

Switching format of a LP

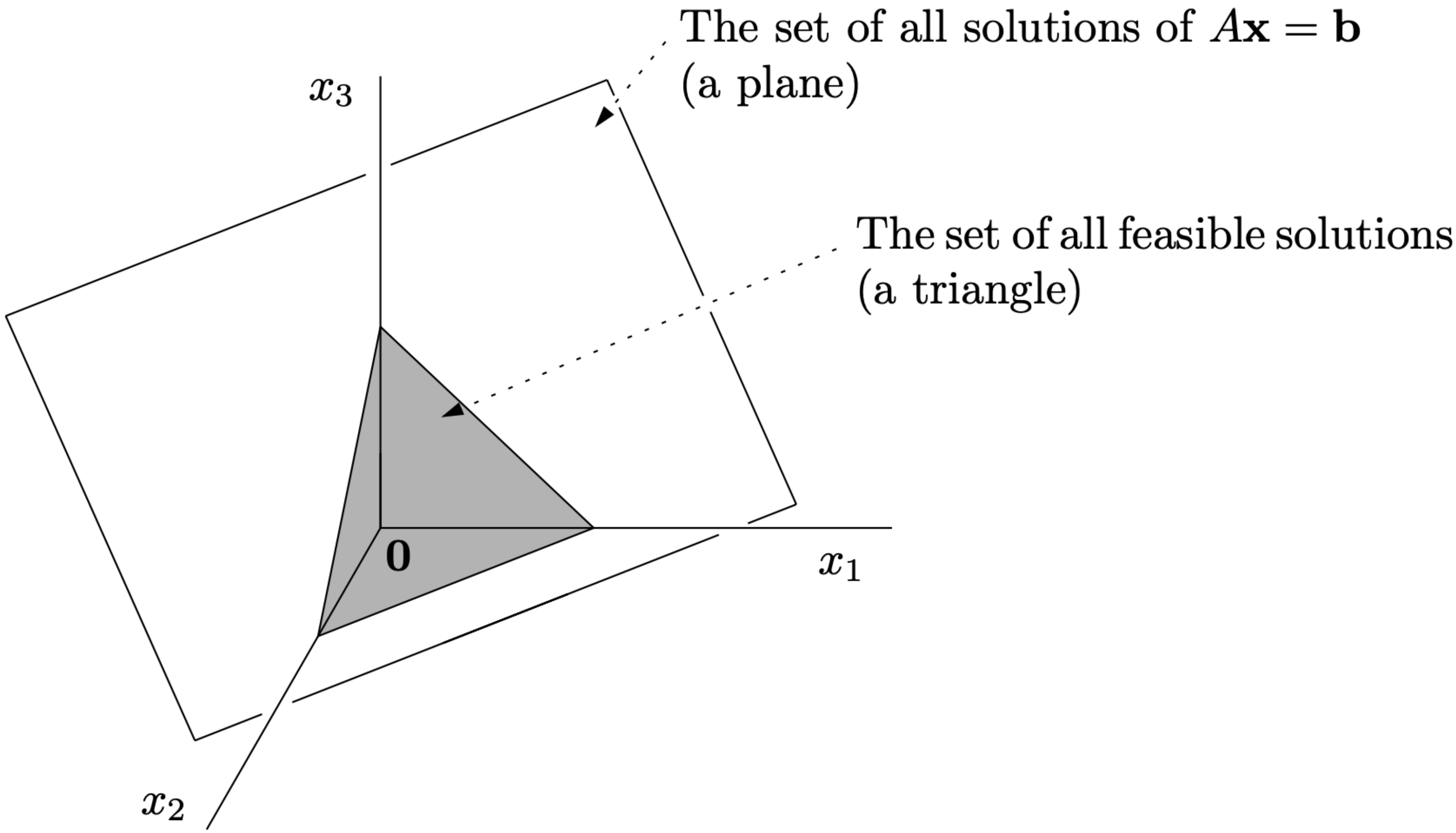
Often it can be useful to change format for a LP.

1. From inequalities to equalities:

$$
\begin{aligned}
a_1x_1 + a_2x_2 \le b \Leftrightarrow \quad & a_1(x_{11} - x_{12}) + a_2(x_{21} - x_{22}) \le b \\
& \text{with } x_{11}, x_{12}, x_{21}, x_{22} \ge 0 \\
\Leftrightarrow \quad & a_1x_1 + a_2x_2 \le b \\
\Leftrightarrow \quad & a_1(x_{11} - x_{12}) + a_2(x_{21} - x_{22}) + s = b \\
& \text{with } x_{11}, x_{12}, x_{21}, x_{22}, s \ge 0
\end{aligned}
$$

2. From equalities to inequalities:

$$
\begin{aligned}
Ax = b, x \ge 0 \Leftrightarrow \; & Ax \le b, -Ax \le -b, -x \ge 0 \\
\Leftrightarrow \; & [A \,|\, -A | -I]x \le [b | -b | 0].
\end{aligned}
$$



[from Matousek's book]

Geometry of polyhedra

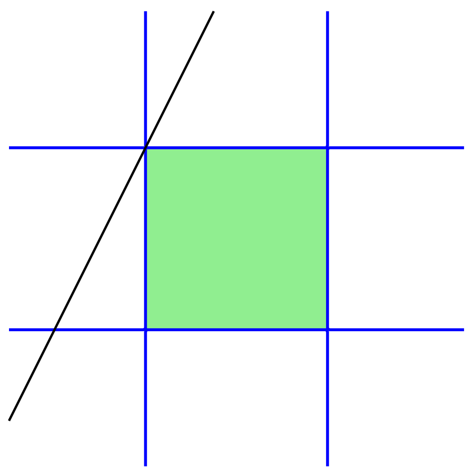
Recall these basic facts:

Every face of a polyhedron is the optimal face of some LP

Extreme pts of faces are extr. pts of the whole polyhedron

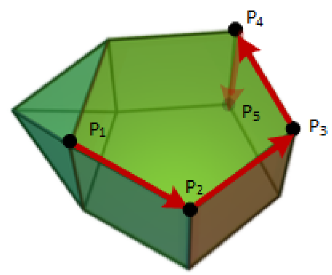
Extreme points of polyhedra are exactly the vertices

Vertices are defined by *maximal* subsystems $A'x \leq b'$ of $Ax \leq b$ (A' is an invertible submatrix of A).



The simplex method

We are going to give a simple description of the simplex method, which consists of drawing a *path* over the polyhedron, starting from a given vertex, and along the *edges* (1-dimensional faces):



Suppose our goal is to solve the (dual) linear program

$$d^* = \max\{b^T y : A^T y \leq c\}$$

where the feasible set is the polyhedron $P = \{y \in \mathbb{R}^m : A^T y \leq c\}$. Assume that a vertex $y_0 \in P$ is *known*.

The simplex method (continued)

1. Let $A_0^T y \leq c_0$ be a subsystem of $A^T y \leq c$, such that

$$A_0^T y_0 = c_0 \quad \text{and } A_0^T \text{ is non-singular}$$

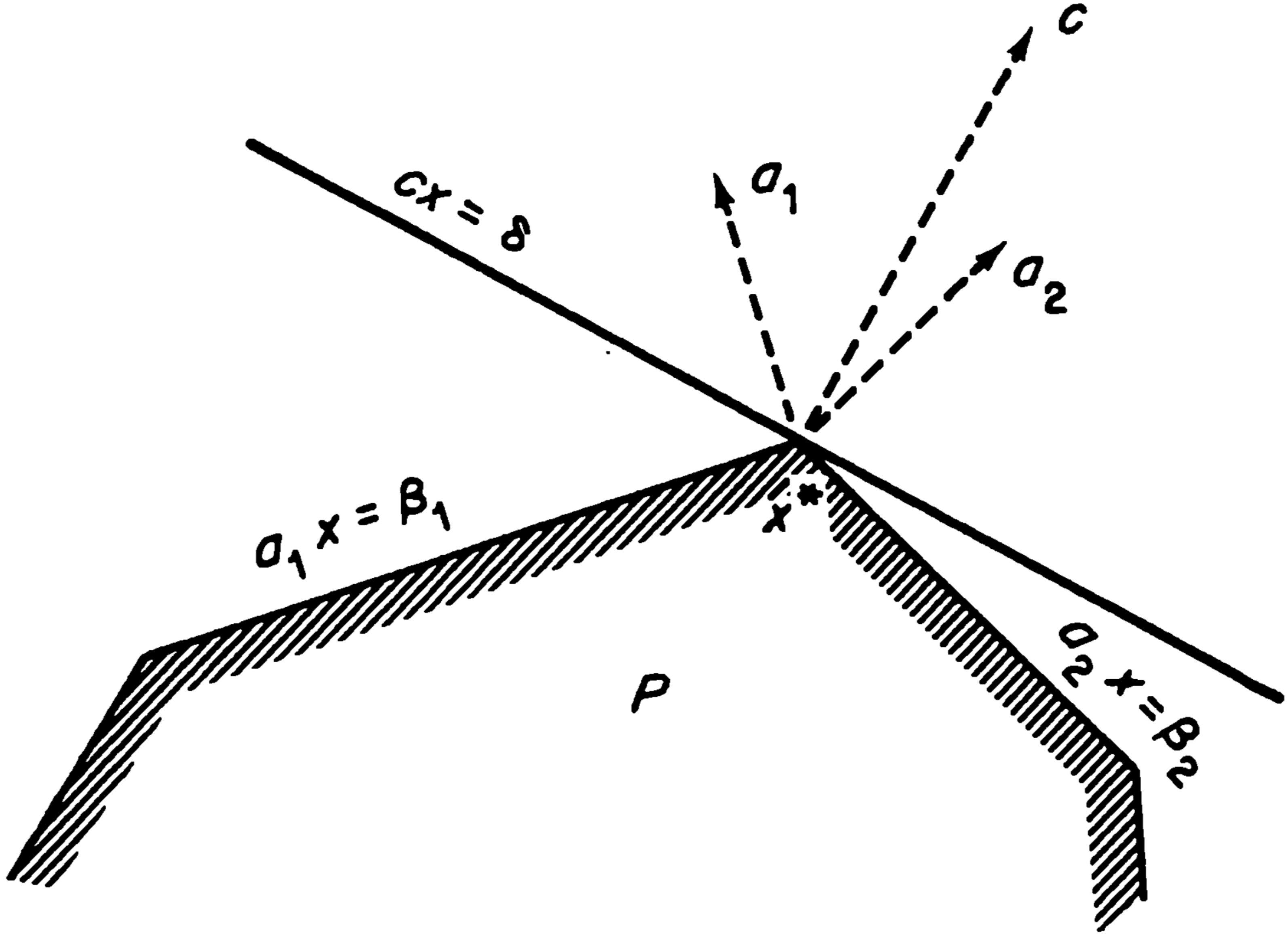
2. Compute $u \in \mathbb{R}^n$ such that $Au = b$ and u is 0 at components not corresponding to the rows selected by A_0^T. So one computes $(A_0^{-1})b$ and adds $0s$ on the rest of the components to get $u \in \mathbb{R}^n$.

3. If $u \geq 0$, we are **done**. Indeed, this means that y_0 is optimal:

$$d^* \geq b^T y_0 = u^T A^T y_0 = u^T c \geq \min_{Ax=b, x \geq 0} c^T x = p^* \geq d^* \qquad ☻$$

If u is not primal feasible (☹), there is some work to do.

Here comes the core part of the simplex method: we are going to trace a path on the boundary, along edges, and improve (increase) the objective value.


c
$cx = \delta$
a_1
a_2
x^*
$a_1 x = \beta_1$
$a_2 x = \beta_2$
P

The simplex method (continued)

4. Let $\bar{i}$ be the smallest component for which $u_{\bar{i}} < 0$. Let $x \in \mathbb{R}^m$ be such that $a^T x = 0$ for all rows a of A_0^T except the $\bar{i}$-th, that satisfies $a_{\bar{i}}^T x = -1$. [that is, x is the $\bar{i}$-th column of $-(A_0^T)^{-1}$]

Remark that the half-line $\{y_0 + \lambda x : \lambda \geq 0\}$ either traverses an edge of P, or it is outside P for all $\lambda \geq 0$. Moreover $b^T x = u^T A^T x = -u_{\bar{i}} > 0$.

5a. If $a^T x \leq 0$ for each row a of A^T, then

$$A^T(y_0 + \lambda x) = A^T y_0 + \lambda A^T x \leq A^T y_0 \leq c, \text{ for all } \lambda \geq 0$$

that is, $y_0 + \lambda x \in P$ for all λ. Thus $b^T(y_0 + \lambda x) = b^T y_0 - \lambda u_{\bar{i}} \mapsto +\infty$ for $\lambda \mapsto +\infty$, hence the max is unbounded. We are **done** ☻

The simplex method (continued)

5b. Otherwise, $a^T x > 0$ for some row a of A^T. Now we compute the largest λ (call it λ_0) such that $y_0 + \lambda x \in P$, which is

$$\lambda_0 = \min_{j, a_j^T x > 0} \frac{b_j - a_j^T y_0}{a_j^T x}.$$

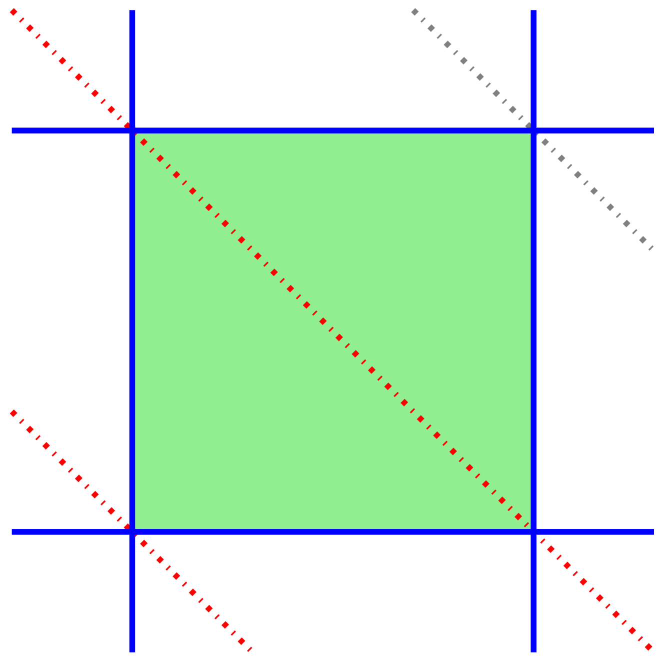
Let $\bar{j}$ be the smallest index j attaining the minimum above, and let $y_1 = y_0 + \lambda_0 x$ (the new vertex). Let A_1^T be the matrix obtained from A_0^T replacing $a_{\bar{i}}$ by $a_{\bar{j}}$.

Now start again the process (go to point 1) with A_1^T and y_1.

An example

Problem: we aim at maximizing the function $y_1 + y_2$ (that is, $b = (1, 1)$) over the following square (pictured below)

$$P = \mathrm{conv}((0,0),(1,0),(0,1),(1,1)) \subset \mathbb{R}^2$$



An example /2

The cube is defined by the dual inequalities: $A^T y \leq c$, where

$$A = \begin{bmatrix} -1 & 0 & 1 & 0 \\ 0 & -1 & 0 & 1 \end{bmatrix} \quad \text{and} \quad c^T = [0\ \ 0\ \ 1\ \ 1]$$

1. We start with the vertex $(0,0)$, defined by $A_0^T y = c_0$ with

$$A_0 = \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix} \quad \text{and} \quad c_0^T = [0\ \ 0]$$

2. Next we look for $u = (u_1, u_2, 0, 0) \in \mathbb{R}^4$ such that $Au = (1,1)$. This yields $u = (-1,-1,0,0)$. We expected this result, since $(0,0)$ is not optimal, and indeed u is not ≥ 0 (actually, it is the worst vertex, indeed all non-zero components are negative). Hence we go to Step 4.

An example /3

4. Here $\bar{i} = 1$. We get that $x \in \mathbb{R}^2$ must satisfy $x_2 = 0$ and $x_1 = 1$, hence $x = (1, 0)$: this is the direction where the simplex method is moving to, that is, it is going east !

5. We get that $y_0 + \lambda x = (\lambda, 0)$ and the maximum λ for which this is still in P is $\lambda_0 = 1$, yielding the new vertex $y_1 = (1, 0)$.

One then iterates, using this y_1 and its defining subsystem:

$$A_1^T = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}, c_1 = (0, 1)$$

Complexity of the simplex method

The simplex method, as many algorithms in mathematics, suffers of the following paradox:

> *Practice.* It works **WELL** in almost every instance (in practice, in a linear number of steps). By the way such efficiency depends on the *pivoting rule* (Step 4). [What we used is Bland's pivoting rule, which is not the most efficient, but prevents from cycling, that is, to visit the same vertex more times]
>
> *Theory.* It is **NOT** polynomial-time, at least for some choice of pivoting rule (see example of Klee-Minty, next slide). It is *average* polynomial-time, within a certain (natural) probabilistic setting, and through the pivoting rule called *Schatteneckenalgorithmus*.

Klee-Minty cube

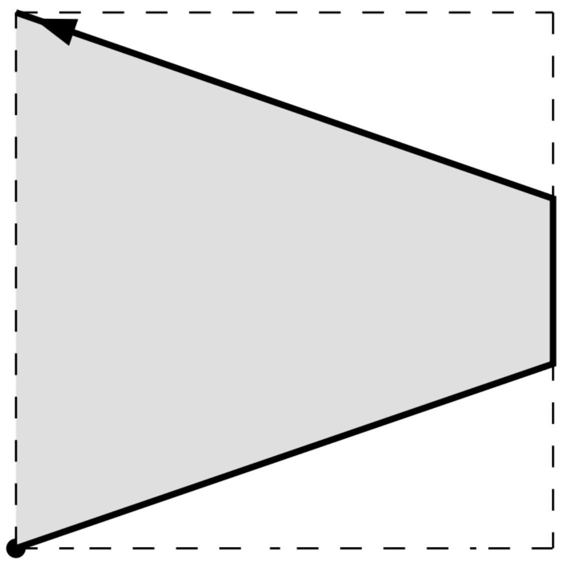
Example of a LP for which the simplex method reaches its worst behaviour (exponential number of visits to vertices before optimality, with Bland's pivoting rule).

For $n \in \mathbb{N}$, the simplex method applied to the following linear program performs exponentially many visits before optimality

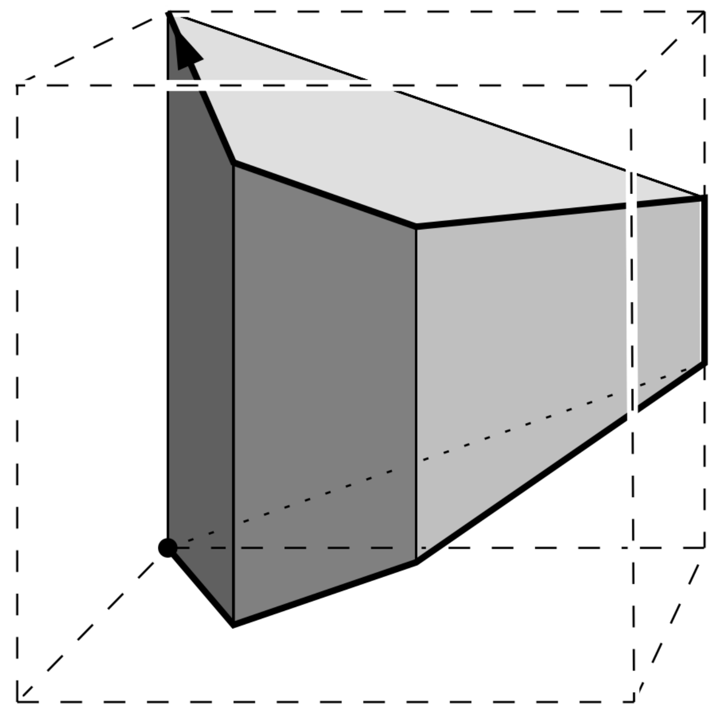
$$
\begin{array}{rcll}
d^* & := & \max & 2^{n-1}x_1 + 2^{n-2}x_2 + \cdots + 2x_{n-1} + x_n \\
& & \text{s.t.} & x_1 \leq 5 \\
& & & 4x_1 + x_2 \leq 25 \\
& & & 8x_1 + 4x_2 + x_3 \leq 125 \\
& & & \vdots \\
& & & 2^n x_1 + 2^{n-1}x_2 + \cdots + 4x_{n-1} + x_n \leq 5^n \\
& & & x_1, x_2, \ldots, x_n \geq 0
\end{array}
$$

[The simplex method (with Bland's pivoting rule) is not polynomial time]

The Klee-Minty cube (from Matousek)



$n = 2$



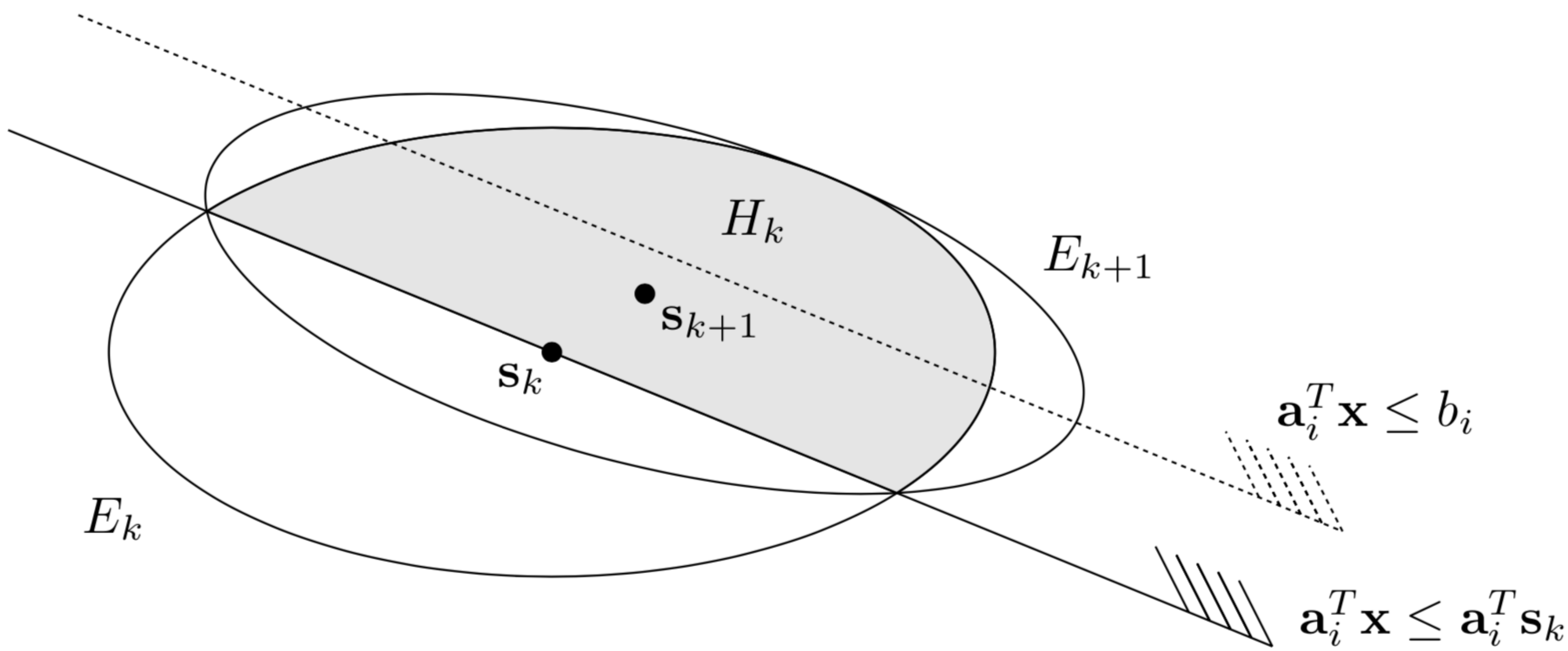
$n = 3$

Ellipsoid method

Invented by Shor, Nemirovski, Yudin (1970's).

First polynomial time for LP (Khachyian, 1979).

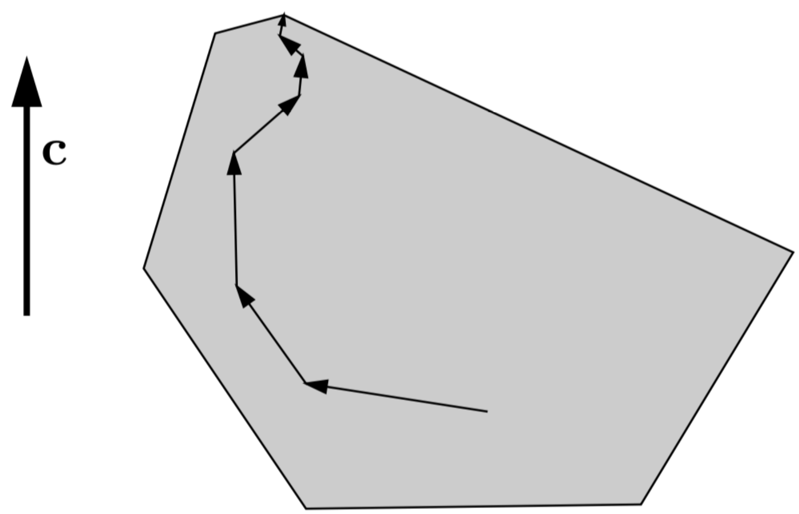
Not interesting in practice, since less efficient than simplex.



Interior point methods

We now introduce a third class of algorithms that can solve linear programming problems: *interior-point methods*.

Contrarily to ellipsoid (outer approximations) and simplex (boundary path), IPM work in the interior of the feasible set.



Developed since the 50's for nonlinear optimization.

Logarithmic barrier function for LP

We consider the LP $\min_P c^T x$ with $P = \{x \in \mathbb{R}^n : Ax \leq b\}$.

Remark that the *polynomial* $p(x) = \prod_i (b_i - a_i^T x)$ vanishes on the boundary of P, and is strictly positive on the interior.

The function $\log p(x)$ is called a *logarithmic barrier*, and

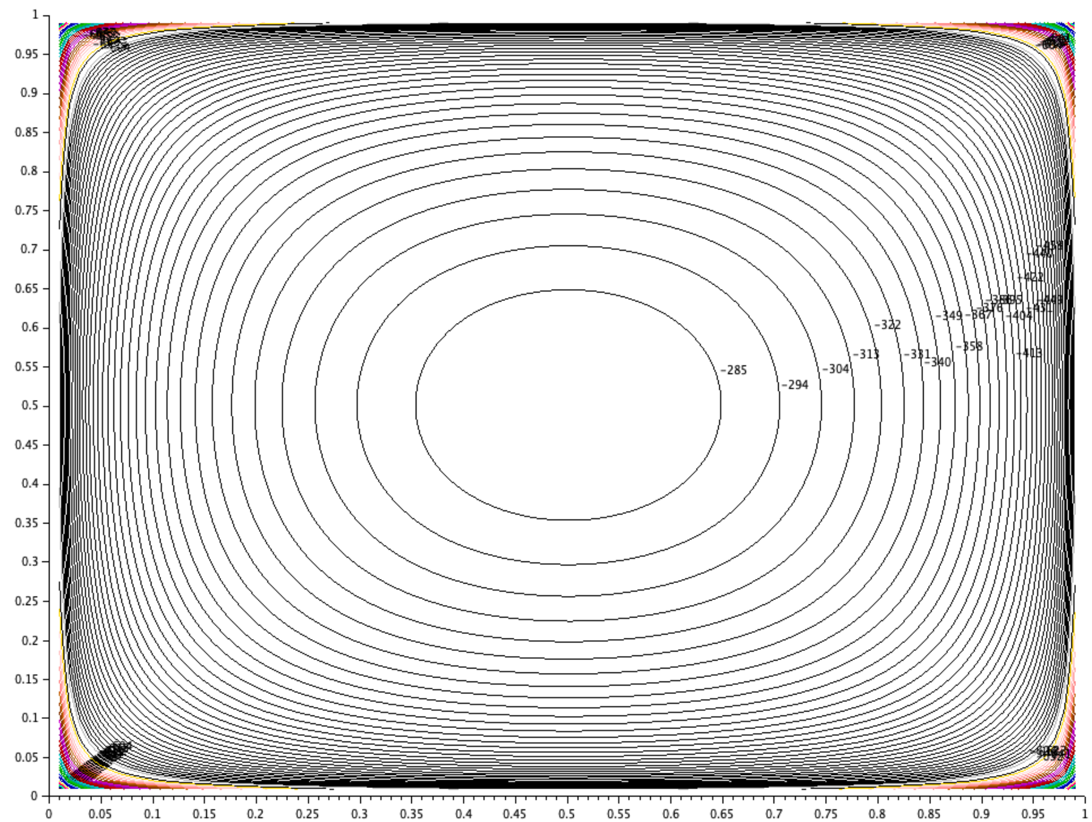
$$f_\mu(x) = c^T x + \mu \log p(x) = c^T x + \mu \sum_i \log (b_i - a_i^T x)$$

has the following properties:

- for $\mu > 0$, $f_\mu(x) \to -\infty$ whenever $x \to \partial P$ (boundary of P)
- $f_\mu(x)$ is concave in $\mathrm{int}(P)$ (the interior of P)
- for $\mu > 0$, P bounded, $f_\mu(x)$ has a unique max in $\mathrm{int}(P)$

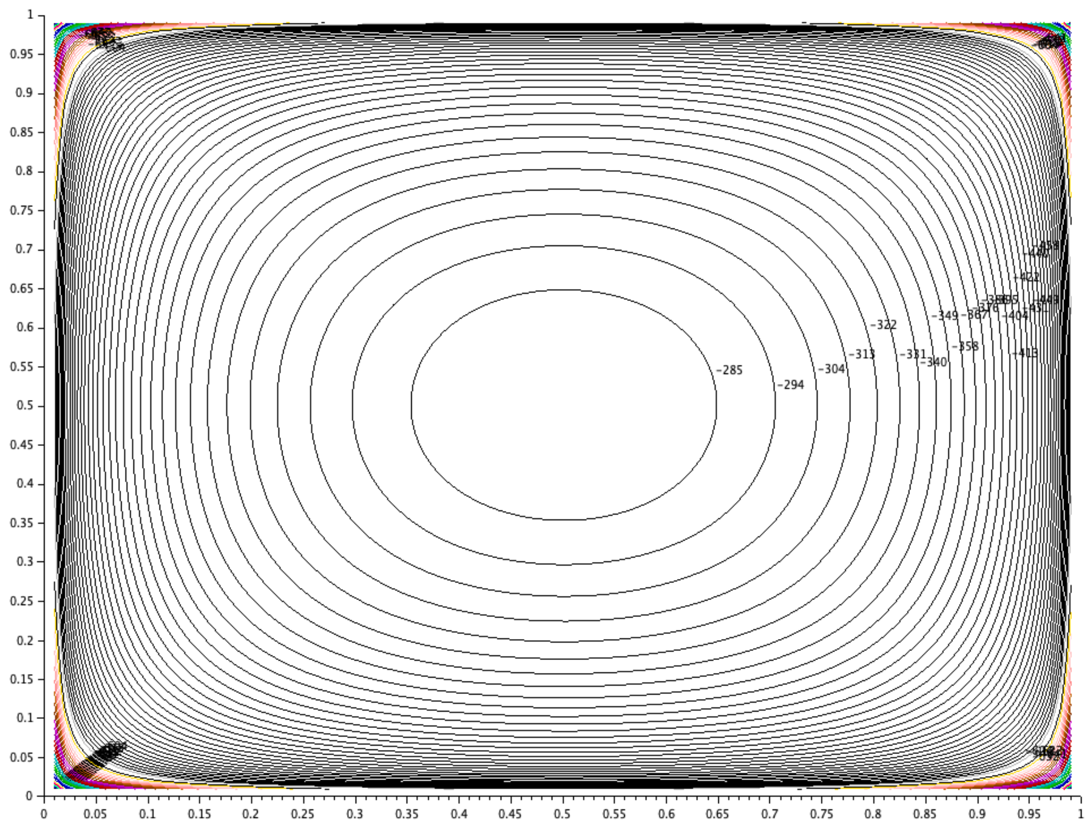
Logarithmic barrier for the unit square

$$c^T x = x_1 + x_2 \qquad \mu = 100$$



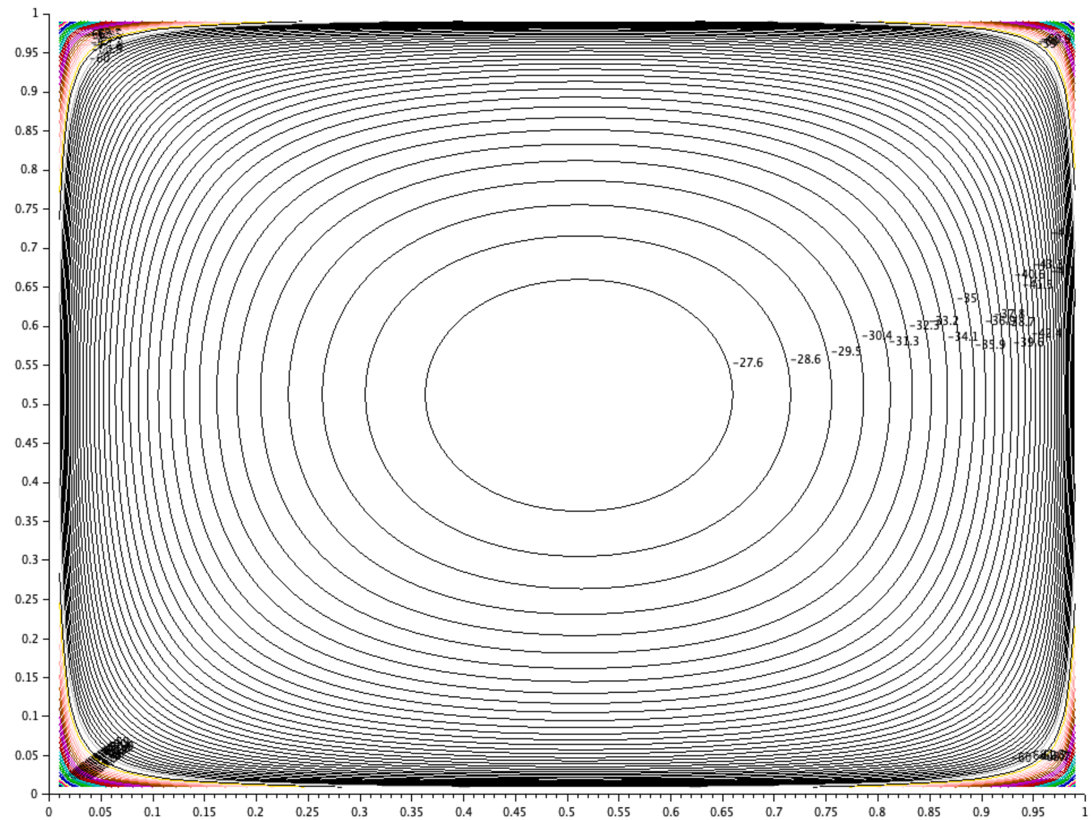
Logarithmic barrier for the unit square

$$c^T x = x_1 + x_2 \qquad \mu = 100$$



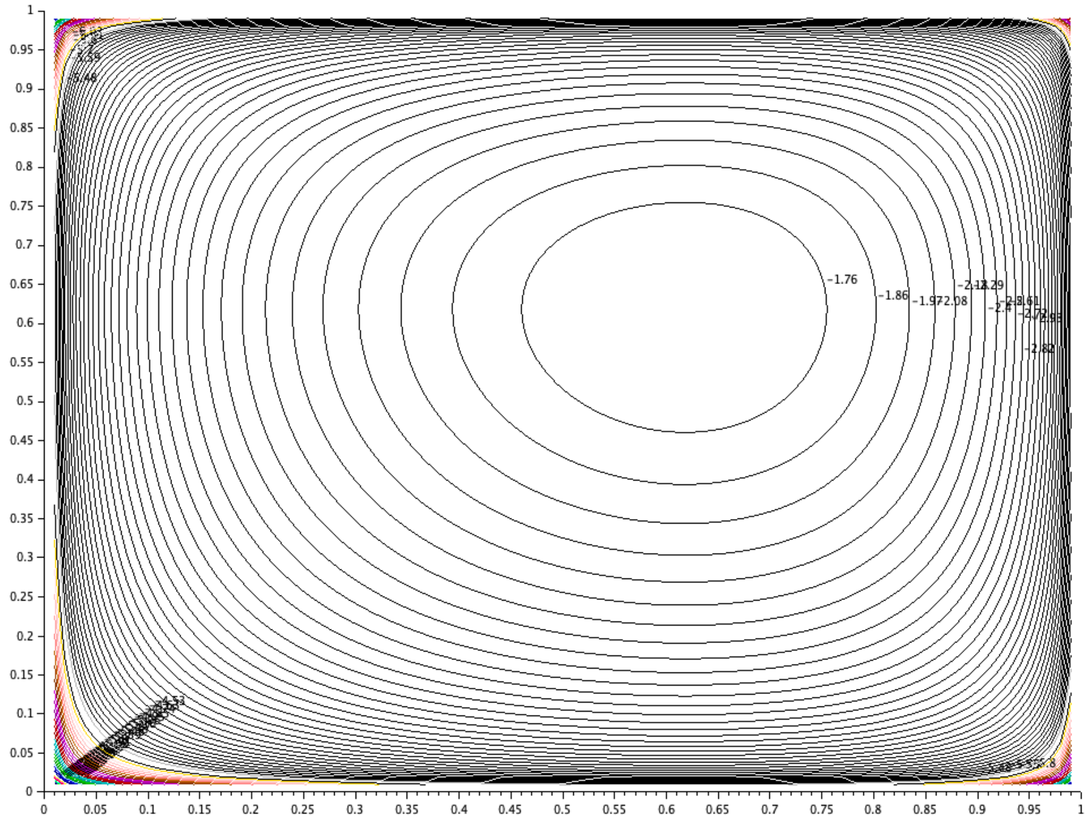
Logarithmic barrier for the unit square

$$c^T x = x_1 + x_2 \qquad \mu = 10$$



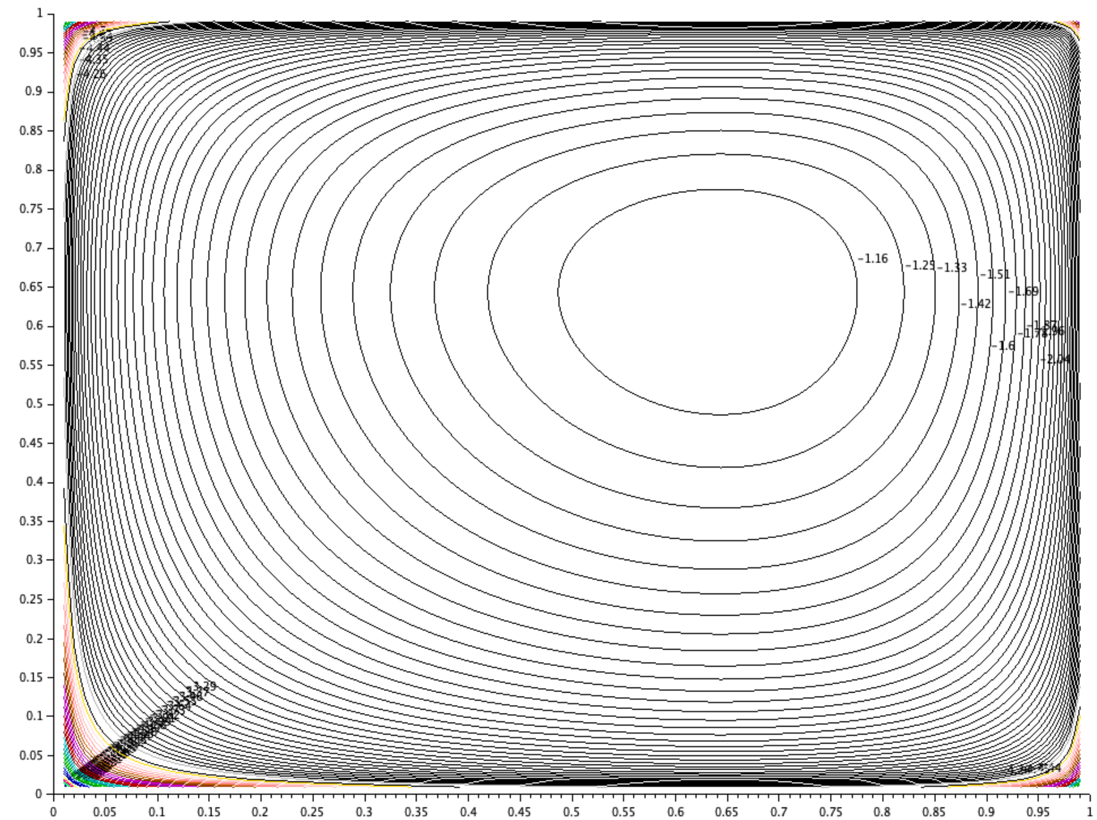
Logarithmic barrier for the unit square

$$c^T x = x_1 + x_2 \qquad \mu = 1$$



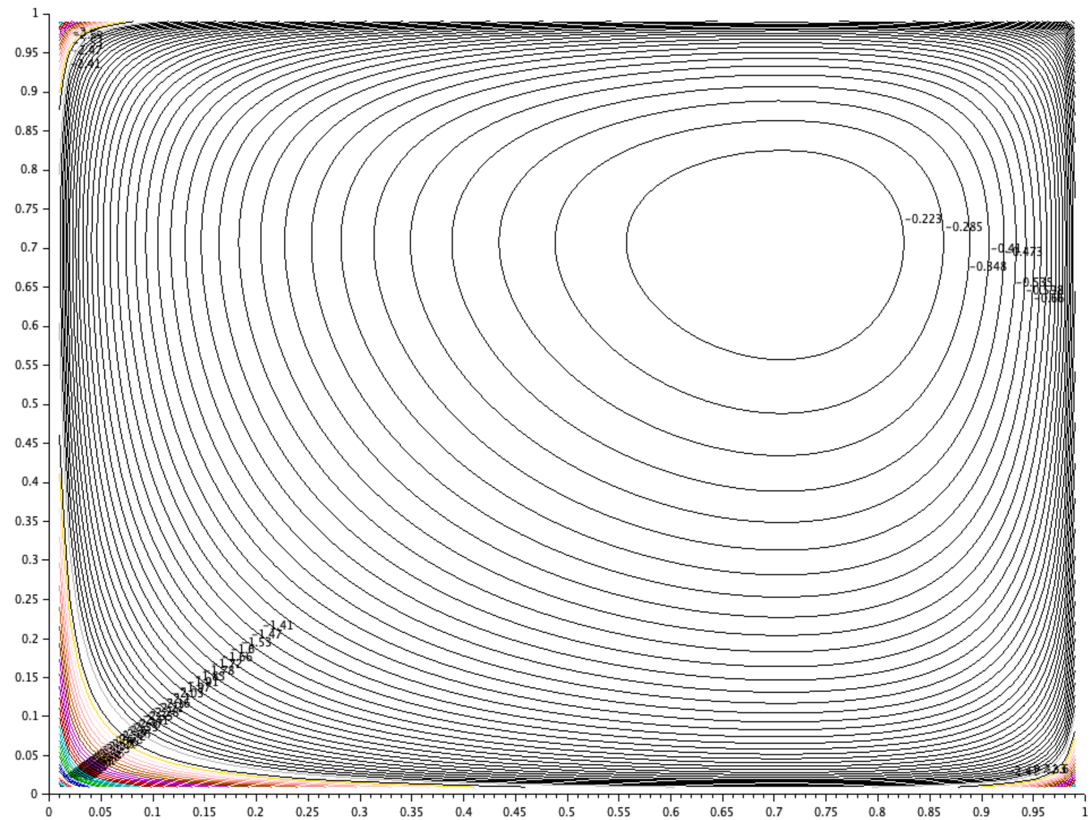
Logarithmic barrier for the unit square

$$c^T x = x_1 + x_2 \qquad \mu = 0.8$$



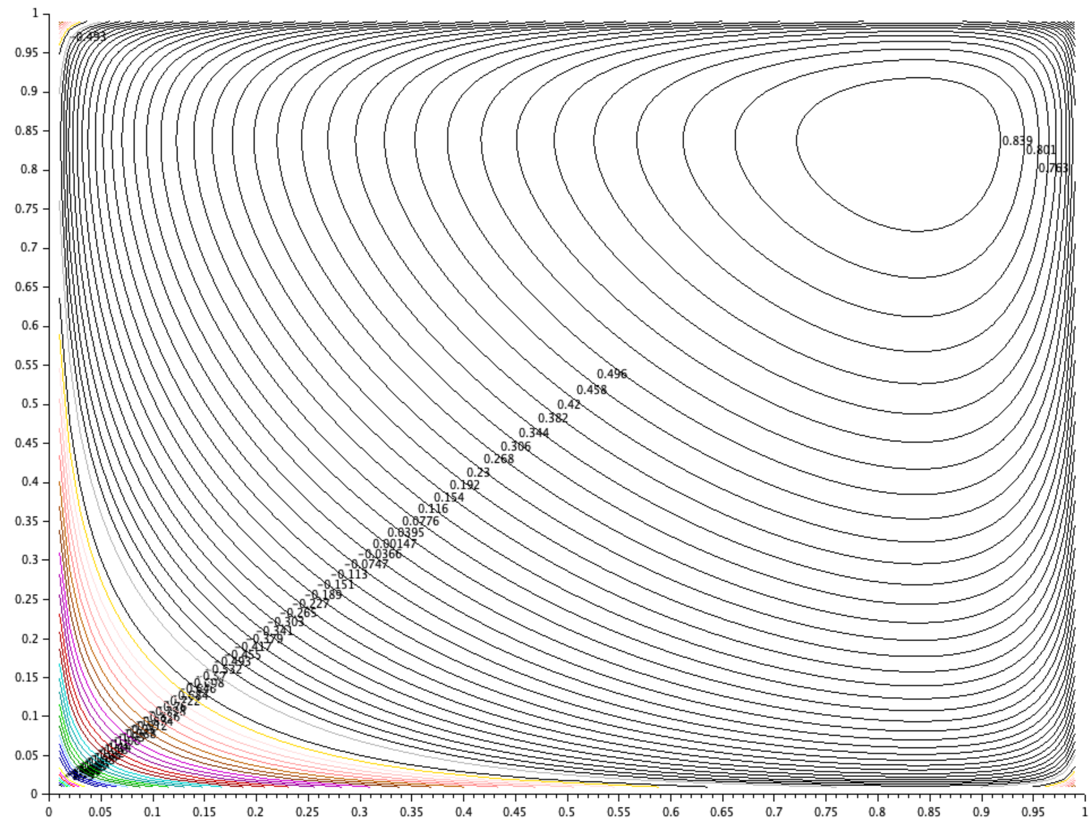
Logarithmic barrier for the unit square

$$c^T x = x_1 + x_2 \qquad \mu = 0.5$$



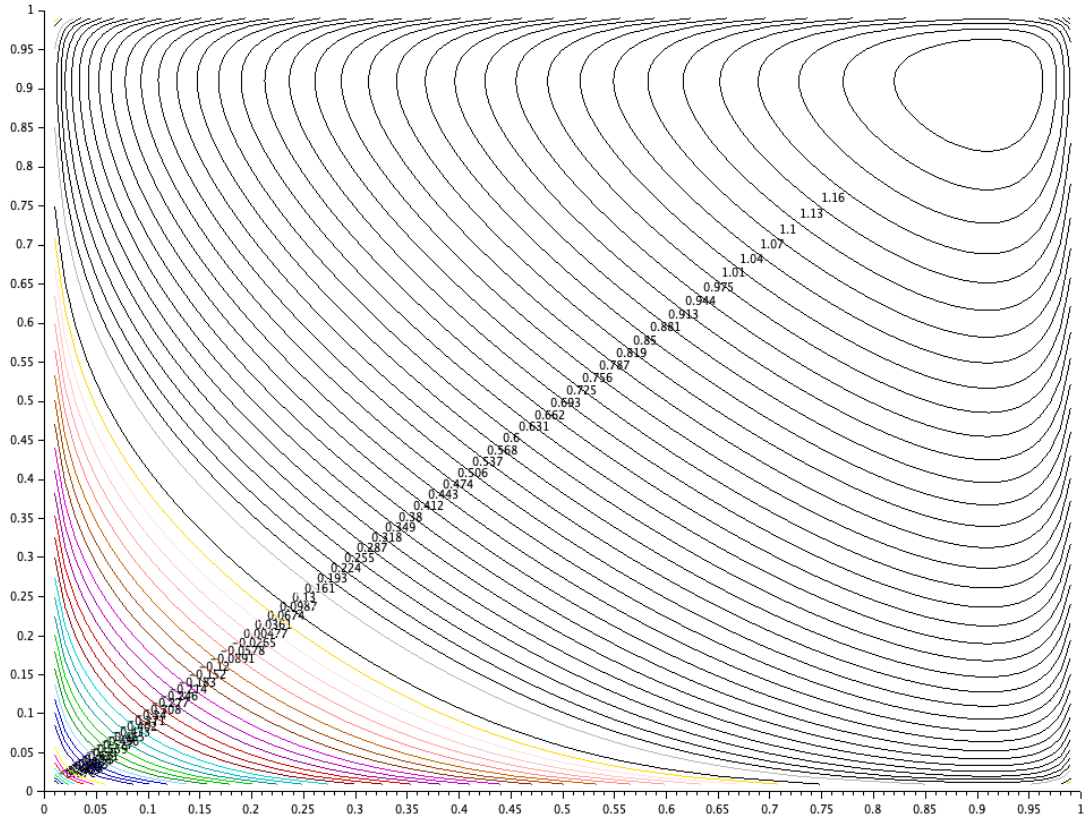
Logarithmic barrier for the unit square

$$c^T x = x_1 + x_2 \qquad \mu = 0.2$$



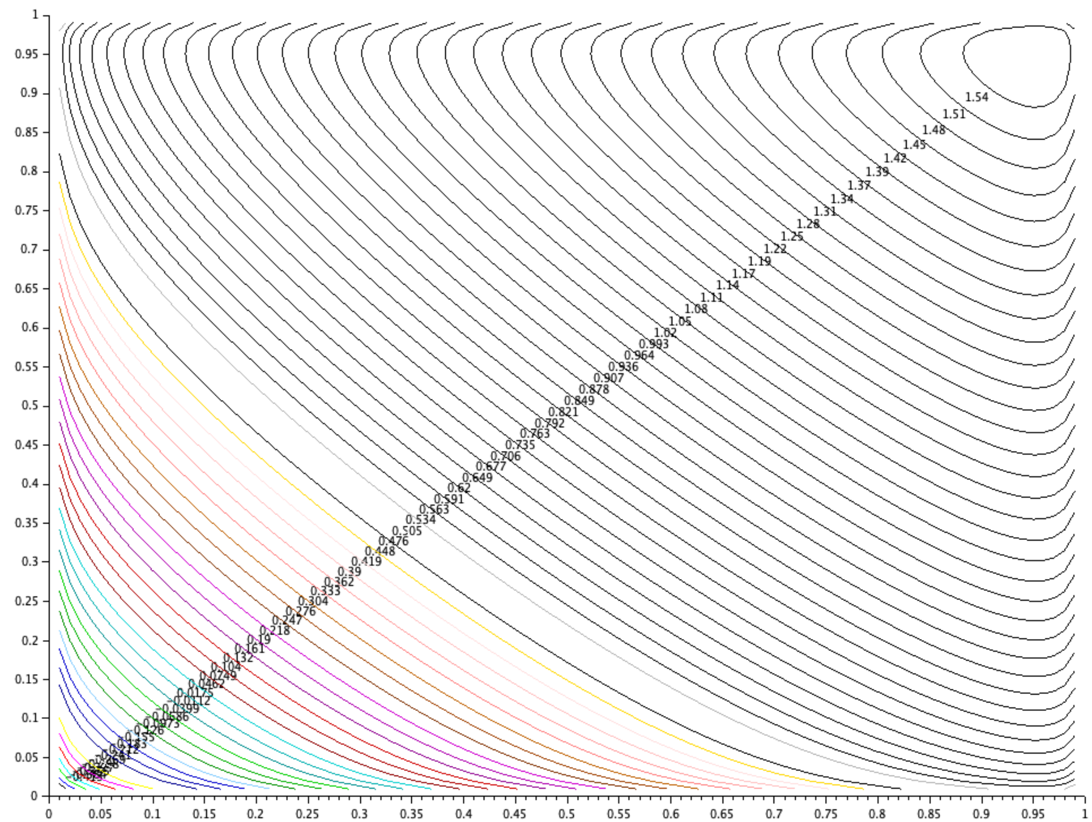
Logarithmic barrier for the unit square

$$c^T x = x_1 + x_2 \qquad \mu = 0.1$$



Logarithmic barrier for the unit square

$$c^T x = x_1 + x_2 \qquad \mu = 0.05$$



Central path for LP

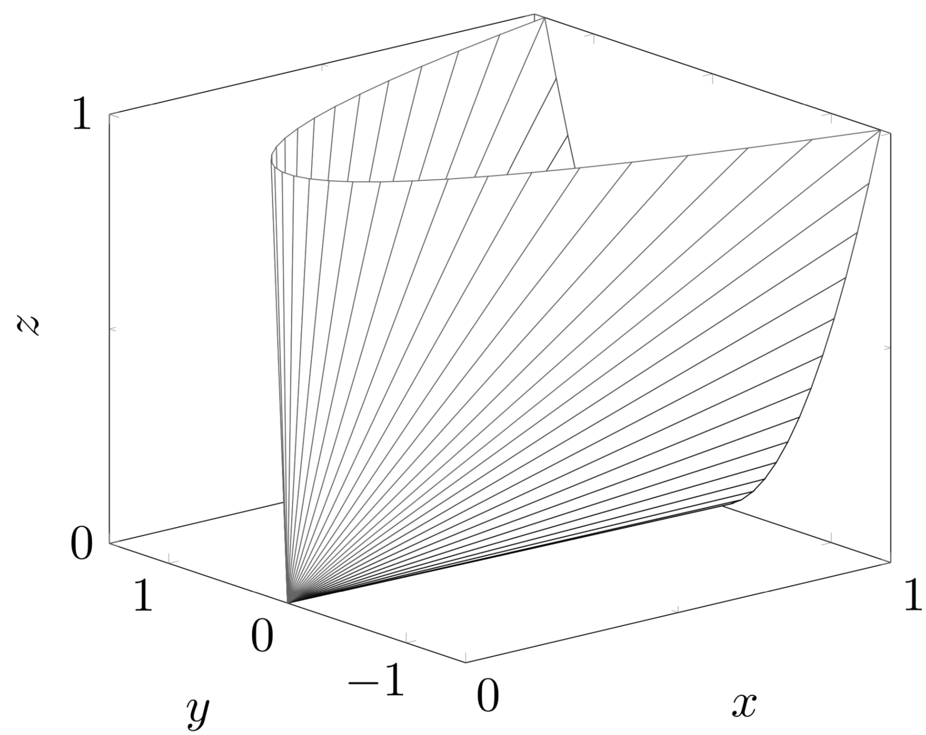
The curve $\{x^*(\mu) \in \mathbb{R}^n : \mu > 0\}$ parametrized by μ is called the *central path*. It is an *algebraic curve* (zero set of some polynomial equations).

The interior-point method (generally speaking, since there are many variants, here we stress on the *path-following* IPM) then consists of starting with a strictly feasible point $x^*(\mu)$ with $\mu \gg 0$, and follow the path.

Actually, what one does in practice is to numerically trace the path using Newton approximation or similar numerical methods.

PART III

The geometry of the psd cone and introduction to SDP



Positive semidefinite matrices

$X \in \mathcal{S}_d$ is *positive semidefinite (psd)*, resp. *definite (pd)*, if:

$v \mapsto v^T X v$ is globally nonnegative, (positive) $v \in \mathbb{R}^d (\setminus\{0\})$

The following conditions are equivalent (resp. blue ones) :

1. X is psd (resp. pd)
2. The eigenvalues of X are nonnegative (resp. positive)
3. $X = LL^T$ for some full-rank $L \in \mathbb{R}^{d\times r}$ (resp. $L \in \mathbb{R}^{d\times d}$)
4. The (leading) principal minors of X are ≥ 0 (resp. > 0)

The positive semidefiniteness induces the *Löwner order* on $\mathcal{S}_d$:

$X \succeq Y$ if and only if $X - Y$ is psd

Useful properties of (psd) symmetric matrices

- Spectral dec. for $X \in \mathcal{S}_d$: $X = Q^T \Lambda Q = \sum_i \lambda_i(X) q_i q_i^T$

- for $X \succeq 0$ and $v \in \mathbb{R}^n$: $v^T X v = 0$ if and only if $v \in \ker X$

- S invertible : $X \succeq 0\,(\succ 0)$ if and only if $S^T X S \succeq 0\,(\succ 0)$

- (Schur complement) Let $A \succ 0$, and $C \in \mathcal{S}_e$. Then

$$\begin{bmatrix} A & B \\ B^T & C \end{bmatrix} \succeq 0 \text{ (resp. } \succ 0\text{) iff } C - B^T A^{-1} B \succeq 0 \text{ (resp. } \succ 0\text{)}$$

A metric on $\mathcal{S}_d$

The *(Euclidean) inner product* on $\mathcal{S}_d$ is given by

$$X, Y \in \mathcal{S}_d : \langle X, Y \rangle_{\mathcal{S}_d} := \mathrm{Trace}(XY)$$

where $\mathrm{Trace}(XY) = \sum X_{ij} Y_{ij}$ is the usual trace function on XY.

Remarkable properties :

- $(\,\cdot\,,\,\cdot\,) \mapsto \langle\,\cdot\,,\,\cdot\,\rangle_{\mathcal{S}_d}$ is indeed an inner product on $\mathcal{S}_d$
- If $X \succeq 0, Y \succeq 0$ then $\langle X, Y \rangle_{\mathcal{S}_d} \geq 0$
- (*stronger*) : If $X \succeq 0$ then : $Y \succeq 0$ if and only if $\langle X, Y \rangle_{\mathcal{S}_d} \geq 0$
- If U is orthogonal then $\left\langle U^{-1} X U, U^{-1} Y U \right\rangle_{\mathcal{S}_d} = \langle X, Y \rangle_{\mathcal{S}_d}$
- Quadratic forms : $v \mapsto v^T X v = \left\langle X, v v^T \right\rangle_{\mathcal{S}_d}$

Semialgebraic sets

A *basic semialgebraic set* is a set $S \subset \mathbb{R}^n$ defined by (multivariate) polynomial inequalities:

$$S = \{x \in \mathbb{R}^n : f_1(x) \geq 0, \ldots, f_s(x) \geq 0\}$$

Examples: linear spaces, circles, algebraic curves, polyhedra . . .

Being psd is a semialgebraic condition on $\mathcal{S}_d$, that is $\mathcal{S}_d^+$ **is a (basic) semialgebraic set**. Ex ($d = 3$): $X = (x_{ij})_{i,j \leq 3}$ is psd iff

$\det X \geq 0$ (minors of order 3)

$$\begin{aligned} x_{11}x_{22} - x_{12}^2 &\geq 0 \\ x_{11}x_{33} - x_{13}^2 &\geq 0 \\ x_{22}x_{33} - x_{23}^2 &\geq 0 \end{aligned}$$

(minors of order 2)

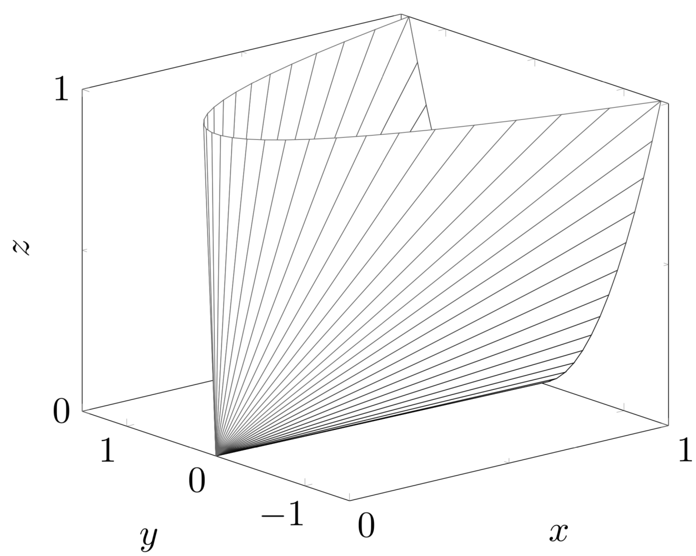
$x_{11} \geq 0,\ x_{22} \geq 0,\ x_{33} \geq 0$ (minors of order 1)

The psd cone

The set $\mathcal{S}_d^+ = \{X \in \mathcal{S}_d : X \succeq 0\}$, called the *psd cone*, satisfies the following properties:

- It is a convex cone, with non-empty (Euclidean) interior
- It is a basic semialgebraic set, it induces the Löwner order
- Every face of $\mathcal{S}_d^+$ is isomorphic to some $\mathcal{S}_e^+$, with $e \leq d$

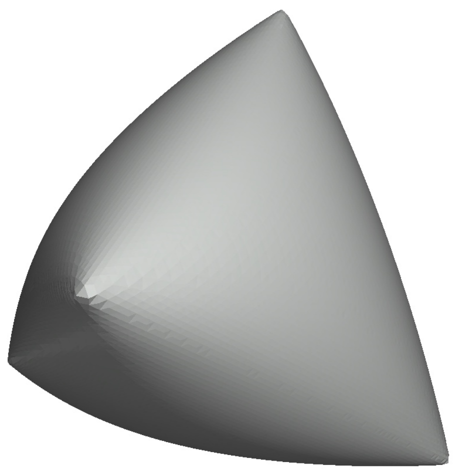
Here's the picture of $\mathcal{S}_2^+$ as a subset of $\mathbb{R}^3 \approx \mathcal{S}_2$:

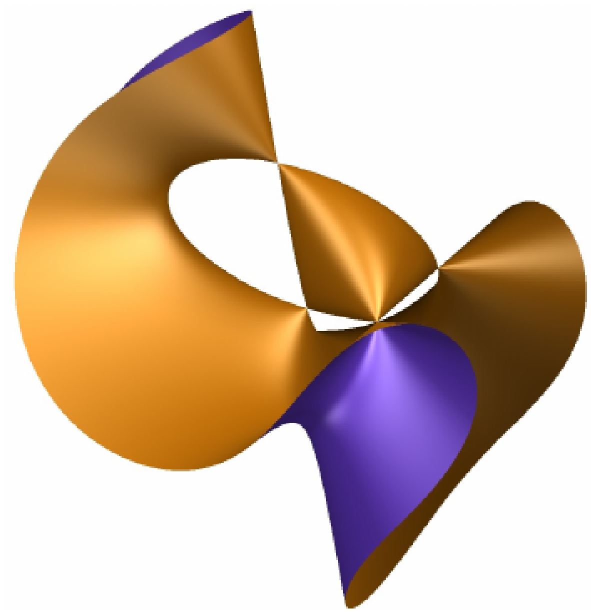


A 3-dim affine section of $\mathcal{S}_3^+$

$\mathbf{K} = \mathcal{S}_3^+$ has dimension 6 (full-dimensional in $\mathcal{S}_3 \approx \mathbb{R}^6$), and taking the following section one gets the figure below

$$E_3 = \left\{ \begin{bmatrix} x \\ y \\ z \end{bmatrix} \in \mathbb{R}^3 : A = \begin{bmatrix} 1 & x & y \\ x & 1 & z \\ y & z & 1 \end{bmatrix} \succeq 0 \right\} \overset{\approx}{\subset} \mathbf{K}$$





E_3 $\qquad$ $\{(x, y, z) \in \mathbb{R}^3 : \det(A) = 0\}$

Faces of the psd cone

We want to characterize the faces of the psd cone $\mathcal{S}_d^+$.

Remark that for all $M \in \mathcal{S}_d^+$, there is a unique minimal face of $\mathcal{S}_d^+$ containing M (take the intersection of all faces containing M), call it $\mathcal{F}_M$.

Theorem. For all $M \in \mathcal{S}_d^+$, then:

- M belongs to the relative interior of $\mathcal{F}_M$
- $\mathcal{F}_M = L_M \cap \mathcal{S}_d^+$, where $L_M = \{U \in \mathcal{S}_d : \ker(M) \subset \ker(U)\}$

The psd cone is self-dual /1

As for LP, the cone for SDP is self-dual, which will imply that one can exploit duality for designing efficient methods.

Theorem. $(\mathcal{S}_d^+)^* = \mathcal{S}_d^+$.

proof. First, let us identify $\mathcal{S}_d^\vee$ with $\mathcal{S}_d$, which is possible since $\mathcal{S}_d$ has finite dimension (what is this dimension?).

First, we claim that $\mathcal{S}_d^+$ is generated by rank-one psd matrices, that is $\mathcal{S}_d^+ = \text{cone}(\{vv^T : v \in \mathbb{R}^d\})$: indeed, this comes from the spectral decomposition $X = Q^T \Lambda Q = \sum_i \lambda_i(X) q_i q_i^T$, and by the fact that $\lambda_i(X) \geq 0$.

Recall that $\text{Trace}(ABC) = \text{Trace}(BCA)$. For all $X \in \mathcal{S}_d^+$ and $v \in \mathbb{R}^d$, one has

$$\left\langle X, vv^T \right\rangle_{\mathcal{S}_d} = \text{Trace}(Xvv^T) = \text{Trace}(v^T X v) = v^T X v \geq 0$$

The psd cone is self-dual /2

This implies that if $Y = \sum_j v_j v_j^T \in \mathcal{S}_d^+$, one has $\langle X, Y \rangle_{\mathcal{S}_d} = \sum_j \left\langle X, v_j v_j^T \right\rangle_{\mathcal{S}_d} \geq 0$, and hence that $\mathcal{S}_d^+ \subset (\mathcal{S}_d^+)^*$.

To prove that $(\mathcal{S}_d^+)^* \subset \mathcal{S}_d^+$, the idea is the same. Indeed, if $Y \in (\mathcal{S}_d^+)^*$, then by definition the scalar product $\langle X, Y \rangle_{\mathcal{S}_d}$ is nonnegative, for all $X \in \mathcal{S}_d^+$.

In particular, for every $v \in \mathbb{R}^d$

$$0 \leq \left\langle vv^T, Y \right\rangle_{\mathcal{S}_d} = v^T Y v$$

which proves that $Y \in \mathcal{S}_d^+$. We conclude the equality $(\mathcal{S}_d^+)^* = \mathcal{S}_d^+$.

LP as optimization over diagonal matrices

Consider the natural *diagonal embedding* of $\mathbb{R}^d$ into the space of symmetric matrices

$$\text{diag} : \mathbb{R}^d \to \mathcal{S}^d, \qquad \text{diag}(x_1, \ldots, x_d) = \begin{bmatrix} x_1 & & \\ & \ddots & \\ & & x_d \end{bmatrix}$$

Remark that $x \in \mathbb{R}^d_{\geq}$ if and only if $\text{diag}(x) \in \mathcal{S}_d^+$ and that the standard primal LP reads as follows

$$\begin{array}{lll} p^* := & \inf & \langle c, x \rangle \\ & \text{s.t.} & Ax = b \\ & & \boxed{\text{diag}(x) \in \mathcal{S}_d^+} \end{array}$$

How about optimizing over the whole psd cone, instead of just the diagonal embedding of the nonnegative orthant ?

This is the idea behind the definition of SDP.

The primal SDP ...

Let $A_1, \ldots, A_n \in \mathcal{S}_d$, and define the linear map

$$\begin{array}{rcl} \mathscr{A} : \mathcal{S}_d & \to & \mathbb{R}^n \\ X & \mapsto & \left(\langle A_1, X\rangle_{\mathcal{S}_d}, \ldots, \langle A_n, X\rangle_{\mathcal{S}_d}\right) \end{array}$$

where we stress on the index $\rangle_{\mathcal{S}_d}$ for the trace inner product on $\mathcal{S}_d$. The primal SDP in standard form reads as follows:

$$\begin{array}{rll} p^* := & \text{inf} & \langle C, X\rangle_{\mathcal{S}_d} \\ & \text{s.t.} & \mathscr{A}(X) = b \\ & & \boxed{X \succeq 0} \end{array}$$

In other words, SDP is the problem of minimizing a linear function on $\mathcal{S}_d$, subject to affine-linear constraints ($\mathscr{A}(X) = b$) and conical contraints given by the membership to the psd cone $\mathcal{S}_d^+$.

... and its Lagrangian dual

Remark that for the previous linear operator $\mathscr{A}$ one has

$$\langle \mathscr{A}(X), y\rangle_{\mathbb{R}^n} = \textstyle\sum_i y_i \langle A_i, X\rangle_{\mathcal{S}_d} = \langle X, \textstyle\sum_i y_i A_i\rangle_{\mathcal{S}_d} \qquad \forall\, X \in \mathcal{S}_d,\ y \in \mathbb{R}^n$$

which gives the (unique) adjoint map $\mathscr{A}^*(y) = \sum_i A_i y_i$.

The Lagrangian conic dual of a semidefinite program is hence the semidefinite program

$$\begin{array}{rcll} d^* & := & \text{sup} & \langle b, y\rangle_{\mathbb{R}^n} \\ & & \text{s.t.} & C - \sum_i A_i y_i = S \\ & & & S \succeq 0 \end{array}$$

As above, the symmetric matrix S plays the role of a slack matrix, allowing us to separate the affine and the semidefinite (dual) constraints.

Spectrahedra and Linear matrix inequalities (LMI)

Both the primal and the dual feasible sets of a semidefinite program are called *spectrahedra*, that is a set defined by a *linear matrix inequality* :

$$\mathscr{S}_M = \{x \in \mathbb{R}^s : M(x) := M_0 + M_1 x_1 + \cdots + M_s x_s \succeq 0\}$$

The name comes from the fact that a spectrahedron is a poly*hedron* in the space of eigenvalues, that is the *spectrum*, of a real symmetric matrix.

The matrix $M(x) = M_0 + M_1 x_1 + \cdots + M_s x_s$ is called a linear matrix, it defines the affine space

$$L = M_0 + \langle M_1, \ldots, M_s \rangle_{\mathbb{R}} \subset \mathcal{S}_d$$

Properties and examples

Let M be a linear matrix and $\mathcal{S}_M$ it associated spectrahedron:

$\mathcal{S}_M$ is a basic closed semialgebraic set

It is a convex set, as pre-image of $\mathcal{S}_d^+$ under $M : \mathbb{R}^s \to \mathcal{S}_d$

It is non-polyhedral in general, but its faces are exposed

Examples :

The $n-$dim disk: $\{x \in \mathbb{R}^n : \|x\| \leq 1\} = \mathcal{S}_M$ with $\boxed{M = \begin{bmatrix} I_n & x \\ x^T & 1 \end{bmatrix}}$

The SOS cone $\Sigma_{n,2d}$ is a (projection of a) spectrahedron, while its dual cone $\Sigma^*_{n,2d}$ is a spectrahedron.

Degenerate SDP /1

If A and B are two linear matrices, we denote by $A \oplus B$ the block-diagonal matrix with blocks A, B. Then it is easy to see* that $\mathcal{S}_{A \oplus B} = \mathcal{S}_A \cap \mathcal{S}_B$.

Using this fact, one can cook up examples of spectrahedra $\mathcal{S}_M$ which are "doubly exponentially far from the origin", that is, such that every point in $\mathcal{S}_M$ has exponential size with respect to input size:

$$M_n = \begin{bmatrix} 1 & 2 \\ 2 & x_1 \end{bmatrix} \oplus \begin{bmatrix} 1 & x_1 \\ x_1 & x_2 \end{bmatrix} \oplus \begin{bmatrix} 1 & x_2 \\ x_2 & x_3 \end{bmatrix} \oplus \cdots \oplus \begin{bmatrix} 1 & x_{n-1} \\ x_{n-1} & x_n \end{bmatrix}$$

Remark that $\mathcal{S}_{M_n}$ is **not** contained in a ball of radius 2^{2^n}.

---

*Check it by exercise.

Degenerate SDP /2

Irrational solutions : there are examples of spectrahedra $\mathcal{S}_M$ defined with rational matrices, such that every point in $\mathcal{S}_M$ is irrational, for instance*:

$$\left\{\sqrt{2}\right\} = \left\{x \in \mathbb{R} : \begin{bmatrix} 1 & x \\ x & 2 \end{bmatrix} \oplus \begin{bmatrix} x & 2 \\ 2 & 2x \end{bmatrix} \succeq 0\right\}$$

Recall that this is not the case for polyhedra (LP) : indeed, if a polyhedron $P \subset \mathbb{R}^n$ is defined over $\mathbb{Q}$ (with facet inequalities with rational coefficients), then every vertex of P is in $\mathbb{Q}^n$.

The same for a 2×2 LMI in the following SDP:

$$\sqrt{2} = \inf x \quad \text{s.t. } \begin{bmatrix} 1 & x \\ x & 2 \end{bmatrix} \succeq 0.$$

---

*Indeed, in this case remark that the first LMI $\begin{bmatrix} 1 & x \\ x & 2 \end{bmatrix} \succeq 0$ implies that $x^2 \leq 2$, and the second $\begin{bmatrix} x & 2 \\ 2 & 2x \end{bmatrix} \succeq 0$ imples that $x^2 \geq 2$.

Optimality

One has the following general result that ensure strong duality:

Theorem. Suppose that both the primal and the dual semidefinite programs are strongly feasible (existence of primal and dual feasible positive definite matrices). Then strong duality holds, there exists a pair of primal-dual optimal solutions $(\overline{X}, \overline{S})$ satisfying the following system

$$\overline{X} \succeq 0 \qquad \overline{S} \succeq 0 \qquad \mathscr{A}(X) = b \qquad \overline{X}\,\overline{S} = 0$$

Moreover, both $\overline{X}$ and $\overline{S}$ are singular matrices, their ranks satisfy the relation $\text{rank}(\overline{X}) + \text{rank}(\overline{S}) \leq d$ (weak complementarity), and for general data $\text{rank}(\overline{X}) + \text{rank}(\overline{S}) = d$ (strong complem.).

Pataki inequalities

As we have seen, the solution to a SDP is a singular (psd) matrix. Can we bound its rank?

Pataki inequalities. Under the assumption that strong complementarity holds for a pair of primal-dual semidefinite programs, that is rank($\overline{S}$) $= r$ and rank($\overline{X}$) $= d - r$ for some $0 \leq r \leq d$, then the following inequalities hold

$$\binom{d-r+1}{2} \leq n \quad \text{and} \quad \binom{r+1}{2} + n \leq \binom{d+1}{2}$$

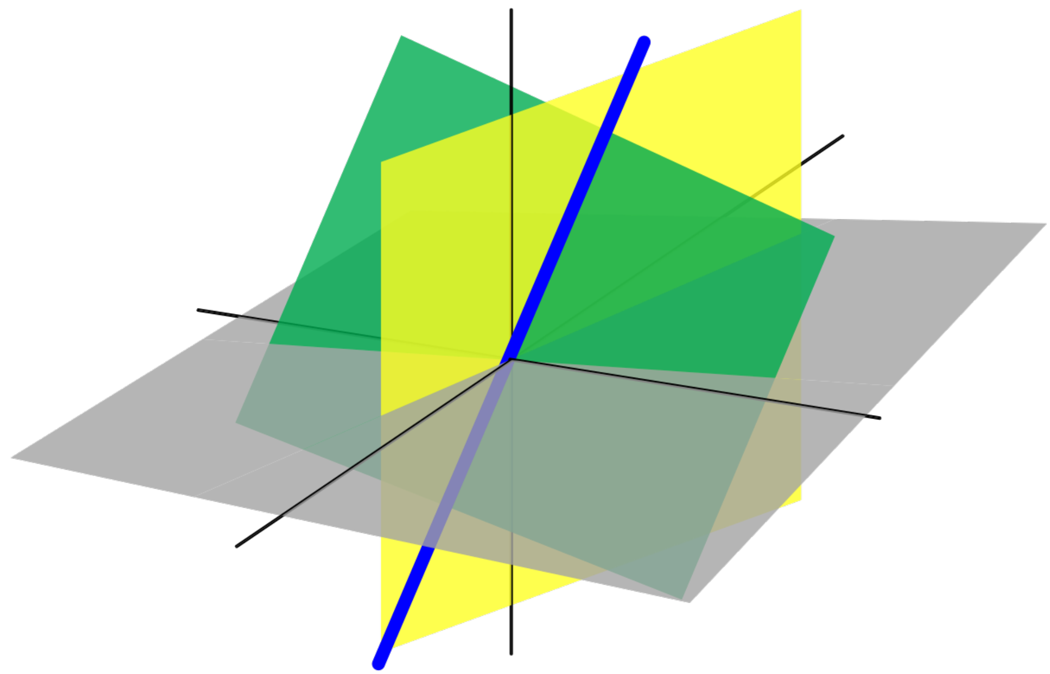
A statistical analysis of the optimal rank

d	3		4		5		6	
s	rank	percent	rank	percent	rank	percent	rank	percent
3	2	24.00%	3	35.34%	4	79.18%	5	82.78%
	1	76.00%	2	64.66%	3	20.82%	4	17.22%
4			3	23.22%	4	16.96%	5	37.42%
	1	100.00%	2	76.78%	3	83.04%	4	62.58%
5					4	5.90%	5	38.42%
	1	100.00%	2	100.00 %	3	94.10%	4	61.58%
6							5	1.32%
			2	67.24%	3	93.50%	4	93.36%
			1	32.76%	2	6.50%	3	5.32%
7			2	52.94%	3	82.64%	4	78.82%
			1	47.06%	2	17.36%	3	21.18%
8					3	34.64%	4	45.62%
			1	100.00%	2	65.36%	3	54.38%
9					3	7.60%	4	23.50%
			1	100.00%	2	92.40%	3	76.50%

courtesy of J. Nie, K. Ranestad, B. Sturmfels "The algebraic degree of Semidefinite Programming", Math. Prog. 122(2):379–405 (2010)

PART IV

Applications of SDP and interior-point method



On the power of semidefinite programming

SDP has become a central problem in math. programming:

- It can model a large class of problems, that is, many problems can be cast directly or indirectly (relaxations) as SDP.

- Applications exist in combinatorial optimization, control theory, mathematical engineering...

- One word about the "complexity": interior-point algorithms can solve SDP in polynomial time **in finite precision**. On the other hand, the complexity status of SDP is fairly unknown in Turing or other exact models of computation.

LP and SOCP are SDP

We have already seen that LP is a special case of SDP, indeed we can re-write it as

$$
\begin{array}{rcll}
p^* & := & \text{inf} & \langle C, X \rangle_{\mathcal{S}_d} \\
 & & \text{s.t.} & \langle A_i, X \rangle_{\mathcal{S}_d} = b_i, A_i = \text{diag}(a_i) \\
 & & & C = \text{diag}(c), X = \text{diag}(x) \\
 & & & X \succeq 0
\end{array}
$$

where a_i is the $i-$th row of the matrix A in the standard LP.

The same holds for the general second-order cone program (SOCP). Indeed the Lorentz cone $\mathscr{L}^n = \{(x,t) \in \mathbb{R}^{n+1} : \|x\| \leq t\}$ is a spectrahedron admitting the following LMI representation :

$$
\mathscr{L}^n = \left\{ (x,t) \in \mathbb{R}^{n+1} : \begin{bmatrix} tI_n & x \\ x^T & t \end{bmatrix} \succeq 0 \right\}
$$

hence the cone constraint $(x,t) \in \mathscr{L}^n$ becomes $\begin{bmatrix} tI_n & x \\ x^T & t \end{bmatrix} \in \mathcal{S}_{n+1}^+$.

QCQP is SDP

A (convex) quadratic programming problem of the form

$$\begin{array}{rcll} f^* & := & \inf & f_0(x) \\ & & \text{s.t.} & f_i(x) \leq 0,\ i = 1, \ldots, m \end{array}$$

where $f_i(x) = (A_i x + b)^T (A_i x + b) - c_i^T x - d_i$, can be cast directly as a semidefinite program :

$$\begin{array}{rcll} f^* & := & \inf & \theta \\ & & \text{s.t.} & \left[\begin{array}{cc} I & A_0 x + b_O \\ (A_0 x + b_O)^T & c_0^T x + d_0 + \theta \end{array} \right] \succeq 0 \\ & & & \left[\begin{array}{cc} I & A_i x + b_i \\ (A_i x + b_i)^T & c_i^T x + d_i \end{array} \right] \succeq 0,\ i = 1, \ldots, m \end{array}$$

By the way, from an algorithmic viewpoint, one prefers to cast a convex QCQP as a second-order cone program.

Eigenvalues optimization is SDP

Let A be a symmetric matrix and let $\lambda_1 \leq \cdots \leq \lambda_d$ be its eigenvalues. Recall the following well-known inequality in linear algebra

$$v^T A v \leq \lambda_d \|v\|^2 \qquad \text{for all } \ v \in \mathbb{R}^d$$

This implies for instance that $\lambda_d I - A$ is positive semidefinite, and more precisely the following semidefinite characterization:

$$\begin{array}{rcll} \lambda_d & := & \inf & \theta \\ & & \text{s.t.} & \theta I_d - A \succeq 0 \end{array}$$

Remark that one can also solve the same problem for a symmetric linear matrix $M(x) = M_0 + \sum_i x_i M_i$, in which case one minimizes the max.eig. in the affine space $L = M_0 + \langle M_1, \ldots, M_s \rangle_{\mathbb{R}}$.

Similarly the sum of the k largest eigenvalues of $M(x)$ is minimized by

$$\begin{array}{rcll} p^* & := & \inf & k\theta + \langle I_d, X \rangle_{\mathcal{S}_d} \\ & & \text{s.t.} & \theta I_d + X - M(x) \succeq 0 \\ & & & X \succeq 0 \end{array}$$

Spectral norm minimization is SDP

The *spectral norm* of a (possibly rectangular) real matrix M is the maximum singular value of M, that is the (square root of the) maximum eigenvalue of the real symmetric matrix M^TM.

Let $M(x) = M_0 + \sum_i x_i M_i$ be a (possibly rectangular) linear matrix describing the affine space $L = M_0 + \langle M_1, \ldots, M_s \rangle_{\mathbb{R}} \subset \mathbb{R}^{p \times q}$.

Then the minimum spectral norm of a matrix in the affine space L can be computed via the following semidefinite program

$$
\begin{array}{rcll}
p^* & := & \inf & \theta \\
 & & \text{s.t.} & \begin{bmatrix} \theta I_p & M(x) \\ M(x)^T & \theta I_q \end{bmatrix} \succeq 0
\end{array}
$$

of size $p + q$ in $s + 1$ variables.

Lyapunov stability is a LMI/SDP

The linear dynamical system $\dot{x} = Ax$ for a given real matrix A, is *(exponentially) stable* whenever the eigenvalues of A have negative real part.

This is equivalent to the existence of a *quadratic Lyapunov function* that is a state function $V(x) = x^T P x$ such that its matrix P satisfies

$$P \succ 0 \qquad A^T P + P A \succ 0$$

which is equivalent to the strong feasibility of a SDP.

More generally one can use LMI to model the design of *state-feedback controllers* minimizing the norm of input-output transfer functions.

A general relaxation scheme for quadratic programs

Consider (with $A_i \in \mathcal{S}_d$) the quadratic optimization problem

$$\begin{array}{rcll} f^* & := & \inf & x^T A_0 x + b_0^T x + c_0 \\ & & \text{s.t.} & x^T A_i x + b_i^T x + c_i \leq 0, \; i = 1, \ldots, m \end{array}$$

Remark that $x^T A_i x = \text{Trace}(x^T A_i x) = \text{Trace}(A_i x x^T) = \left\langle A_i, xx^T \right\rangle_{\mathcal{S}_d}$ and that xx^T is a *rank-one psd matrix*. One gets

$$\begin{array}{rcll} f^* & = & \inf & \langle A_0, X \rangle_{\mathcal{S}_d} + b_0^T x + c_0 \\ & & \text{s.t.} & \langle A_i, X \rangle_{\mathcal{S}_d} + b_i^T x + c_i \leq 0, \; i = 1, \ldots, m \\ & & & X = xx^T \end{array}$$

and hence that the original problem can be relaxed by dropping the rank constraint, obtaining

$$\begin{array}{rcll} f^* & \geq & \inf & \langle A_0, X \rangle_{\mathcal{S}_d} + b_0^T x + c_0 \\ & & \text{s.t.} & \langle A_i, X \rangle_{\mathcal{S}_d} + b_i^T x + c_i \leq 0, \; i = 1, \ldots, m \\ & & & \begin{bmatrix} X & x \\ x^T & 1 \end{bmatrix} \succeq 0 \end{array}$$

(Remark that the LMI constraint is equivalent to $X \succeq xx^T$).

Relaxation for boolean optimization

A *boolean* or *(0,1)-quadratic program* is the opt. problem

$$\begin{array}{rcll} f^* & := & \inf & x^T A x + b^T x \\ & & \text{s.t.} & x_i^2 = 1,\ i = 1, \ldots, d \end{array}$$

whose feasible set is the set of vertices of the hypercube in $\mathbb{R}^d$.

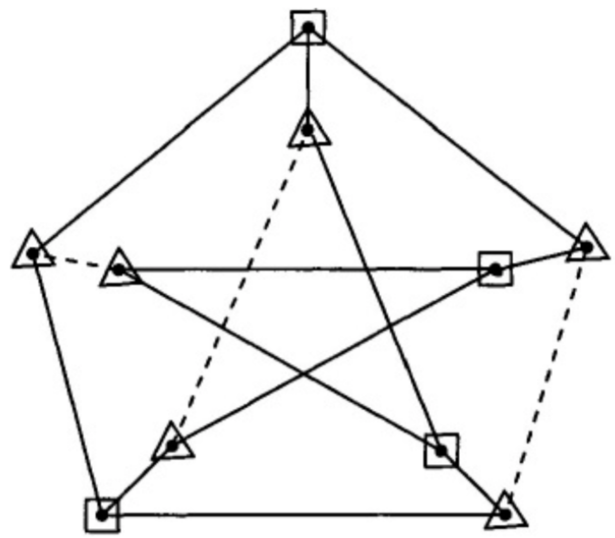
It can be relaxed as a semidefinite program by simply remarking that if x is feasible, then the diagonal elements of $X = xx^T$ are all equal to one (and, still, X has rank one and is psd).

The following is a semidefinite relaxation of the above boolean program

$$\begin{array}{rcll} f^* & \geq & \inf & \langle A, X \rangle_{\mathcal{S}_d} + b^T x \\ & & \text{s.t.} & X_{ii} = 1,\ i = 1, \ldots, d \\ & & & \begin{bmatrix} X & x \\ x^T & 1 \end{bmatrix} \succeq 0 \end{array}$$

MAXCUT and its semidefinite relaxation

Let $G = (V, E)$ be an *undirected graph* on a set of vertices $V = \{v_1, \ldots, v_n\}$, with edges E. A *cut* on G is a subset $S \subset V$. A *maximum cut* is a cut that maximises the number of crossing edges (edges linking an element of S and another outside it). The MAXCUT and its generalizations are NP-complete.



A maximum cut on the Petersen graph

[courtesy of de Klerk "Aspects of semidefinite programming"]

MAXCUT and its semidefinite relaxation

Say $x_i = 1$ if the $v_i \in S$, -1 otherwise. Then $1 - x_i x_j = 2$ if (v_i, v_j) is a crossing edge and $= 0$ otherwise and

$$\text{MAXCUT} := \operatorname{argsup}_{S \subset V} \sum_{(v_i,v_j)\in E} \frac{1}{2}\left(1 - x_i x_j\right), x_i = \pm 1$$

Its semidefinite relaxation reads as follows

$$\begin{array}{rcll} f^* & = & \text{sup} & \sum_{i,j} \frac{1}{2}\left(1 - S_{ij}\right) \\ & & \text{s.t.} & S_{ii} = 1,\ i = 1, \dots, d \\ & & & S \succeq 0 \end{array}$$

where the rank-constrained condition $S = xx^T$ has been relaxed as above to $S \succeq xx^T$. Goemans/Williamson proved that such a relaxation gives an 87% approximation of the original solution, in average.

Lovász theta-function /1

One of the most celebrated (as MAXCUT) examples in combinatorics is the computation of the Lovász θ-function (from Lászláo Lovász) of a graph $G = (V, E)$.

We denote by:

$\chi(G)$ the *chromatic number* of G, that is, the minimum number of colours needed to colour the vertices, no two adjacent of the same one.

$\omega(G)$ the *clique number* of G, that is, the size of a maximum clique (= subgraph whose vertices are 2-by-2 adjacent).

Lovász theta-function /2

Let $e = (1, 1, \ldots, 1) \in \mathbb{R}^n$ and let X be a generic symmetric matrix. The *Lovász θ-function* of an undirected graph $G = (V, E)$ is defined as the maximum value of the following SDP:

$$\begin{array}{rcll} \theta(G) & = & \text{sup} & \text{Trace}(ee^T X) = e^T X e \\ & & \text{s.t.} & x_{ij} = 0 \text{ for all } (i, j) \notin E \\ & & & \text{Trace}(X) = 1 \\ & & & X \succeq 0. \end{array}$$

It satisfies the following inequalities:

$$\omega(G) \leq \theta(\overline{G}) \leq \chi(G)$$

where G is the complement of G (graph with same vertices and complement edges). Hence it is used to lowerbound $\chi(G)$ and to upperbound $\omega(G)$ in polynomial time.

Semidefinite relaxation of polynomial optimization

Let $f \in \mathbb{R}[x_1, \ldots, x_n]$, and let the global optimization problem be

$$\begin{array}{rllcll} f^* & = & \inf & f(x) & = & \sup & \lambda \\ & & \text{s.t.} & x \in \mathbb{R}^n & & \text{s.t.} & f - \lambda \in \mathcal{P}(\mathbb{R}^n) \end{array}$$

where $\mathcal{P}(\mathbb{R}^n)$ is the cone of polynomials nonnegative on $\mathbb{R}^n$.

A classical relaxation of this problem is to consider the subcone of $\mathcal{P}(\mathbb{R}^n)$ consisting of *sums-of-squares polynomials*

$$\Sigma_{n,2d} = \left\{ f \in \mathbb{R}[x_1, \ldots, x_n] \; : \; f = \sum_i f_i^2 \right\}$$

which gives a lower bound. By the way such lower bound can be computed through SDP (whereas the computation of f^* is NP-hard in general) :

$$\begin{array}{rllcll} f^{sos} & = & \sup & \lambda & = & \sup & \lambda \\ & & \text{s.t.} & f - \lambda \in \Sigma_{n,2d} & & \text{s.t.} & f - \lambda = v^T X v \\ & & & & & & X \succeq 0 \end{array}$$

Interior-point methods in SDP

Interior-point methods for LP have been successfully extended to the case of SDP:

- Alizadeh's PhD thesis (1991)
- Nesterov and Nemirovski (1990-1994), developed the general theory of interior-point algorithms for conic programming

IPM for SDP are based on the *logarithmic barrier function*:

$$\boxed{F(X) = -\log\det X}$$

defined for non-singular $X \in \mathcal{S}_d$.

Central path

The central path can be defined in SDP as the set of solutions (X_μ, S_μ, y_μ), depending on a parameter $\mu > 0$, of the following system of equations and inequalities:

$$\begin{array}{rcl} \mathscr{A}(X) & = & b \\ \mathscr{A}^*(y) + S & = & C \\ XS & = & \mu\,\mathrm{Id}_m \\ X & \succeq & 0 \\ S & \succeq & 0. \end{array}$$

Theorem. If the primal and the dual semidefinite programs are strongly feasible, then the central path exists and it is an analytic curve.

Limit of the central path

Under the assumption of strict complementarity (there are solutions X^* and S^* with $X^*S^* = 0$ and having complementary rank) the limit of the central path for $\mu \to 0^+$ is a feasible solution.

Lemma. For $\overline{\mu} > 0$, the set $\{(X_\mu, S_\mu) : 0 < \mu \leq \overline{\mu}\}$ is contained in a compact subset of $\mathcal{P} \times \mathcal{D}$ (cart. product of primal and dual sets).

This implies that the limit of the central curve

$$(X_0, S_0) = \lim_{\mu \to 0^+} (X_\mu, S_\mu)$$

is a (primal-dual) feasible solution. Under some assumptions it coincides with the analytic center of the optimal face, in which case it can be computed by maximising the determinant over suitable sections of the primal or dual feasible set.

Complexity of SDP / finite precision

For the problem $d^* = \max\{b^T y : y \in \mathcal{S}_A\}$, where $A_i \in \mathcal{S}_d$, and $\mathcal{S}_A = \{y \in \mathbb{R}^n : A(y) := C - y_1 A_1 - \cdots - y_n A_n \succeq 0\}$, we denote by L the total input bit-size and

$$S(A, \epsilon) = \mathcal{S}_A + B(0, \epsilon) \quad \text{and} \quad S(A, -\epsilon) = \{y : B(y, \epsilon) \subset \mathcal{S}_A\}$$

- If $R > 0$ is known such that either $\mathcal{S}_A = \emptyset$ or $\mathcal{S}_A \cap B(0, R) \neq \emptyset$, then there is an algorithm that, for any $\epsilon > 0$ decides that $S(A, -\epsilon) = \emptyset$ or computes $y \in S(A, \epsilon)$ satisfying

  $$b^T z \leq b^t y + \epsilon, \quad \text{for all } z \in S(A, -\epsilon).$$

  Its **bit complexity** is polynomial in d, n, L and $\log(\frac{1}{\epsilon})$.

- There are algorithms that given y such that $A(y) \succ 0$, and R as above, compute z such that $A(z) \succ 0$ and $c^T z \geq d^* - \epsilon$. Its **arithmetic complexity** is polynomial in $d, n, L, \log(\frac{1}{\epsilon})$ and $\log(R)$. No good bounds for intermediate bit size.

Complexity of SDP / exact models

If one is interested in the theoretical complexity of computing the exact solution of a SDP, it is worth to say that it is an open problem generally speaking. More precisely:

- In the bit model (Turing), SDP or its feasibility is **not** known to be in NP.

- In general, it is known to be in NP if and only if it is in coNP.

- In the real numbers model (Blum, Shub, Smale) it is in NP and in coNP.