The Facial Structure of Convex Programs

Gábor Pataki

Dept. of IE/OR
Columbia University

The “Geometry” of Convex and Linear Programs

Convex program : minimizing a convex function subject to convex constraints.

What do we mean by the “geometry” of a convex program ?

- Characterization of solution set; uniqueness of solution.
- Same for the dual (if there is an explicit one).
- If we replace the objective by its linearization at the optimum, do we get an equivalent problem ?
- etc.

If the convex program is an LP, these questions can be studied through describing the **facial structure** of the feasible set.

There are 3 fundamental notions:

- Faces, extreme points (basic solutions).
- Nondegeneracy.
- Strict complementarity.

Very clear cut connections. E. g.

- x is nondegenerate $\Rightarrow$ dual optimal face is a singleton; $\Leftrightarrow$ dual solution is unique.
- If the dual solution is unique, then any (SC) primal solution must be nondegenerate.

Can we do the same for general convex programs ?

No comprehensive study so far. Some literature on the geometry of convex programs:

(1) Anderson and Nash : LP's in infinite-dimensional spaces.

(2) Faces of feasible sets of SDP's: Ramana, '94; P. '94.

(3) Nondegeneracy in SDP: Shapiro,Fan '94, Alizadeh, Haeberly, Overton '95.

(4) Nondegeneracy in nonlinear programs : Robinson.

(5) Nondegeneracy in cone programs: Shapiro '96.

(6) Characterization of solution sets of convex programs: Mangasarian '91; Burke and Ferris '92.

(7) Weak sharp minima in LP's, QP's: Ferris, Burke '91.

(8) Minimum principle sufficiency in convex programs: Ferris and Mangasarian '92.

- (1) is too general (even the dimension of the space can be infinite). Most of the others only work for specific problems. No treatment of basic solutions.
- Goal: to develop a unifying theory that subsumes, and generalizes many known results on the "geometry" of convex programs. (Started with SDP...)

Why study the facial structure ?

- We should not assume e.g. differentiability. But all closed convex sets have faces $\longrightarrow$ a good approach to describe the local structure of the feasible set.
- Everything we derive should be an easily recognizable generalization of the LP case.

Basic idea

The feasible set of every convex program is the *intersection* of *simple* convex sets.

E.g. the feasible set of an LP is

$$\{\, x \mid x \geq 0 \,\} \quad \cap \quad \{\, x \mid Ax = b \,\}$$

two sets with trivial geometry.

We will characterize the geometry of the intersection using the geometry of the simple sets.

Plan of talk

- Faces of general convex sets.
- The Main Tool: the FIT Theorem.
- The facial structure of cone-constrained linear programs.
- Diverse applications : eigenvalue-optimization; poly-time solvability of small quadratic programs; (partial) sensitivity analysis in cone programs; graph embedding.
- The facial structure of general convex programs.

Definition:

- If C is a convex set, then $F \subseteq C$ is a *face* of C, if F is convex, and $x, y \in C$, $\frac{1}{2}(x+y) \in F$ implies $x, y \in F$.
- A face consisting of only one element is called an *extreme point*.

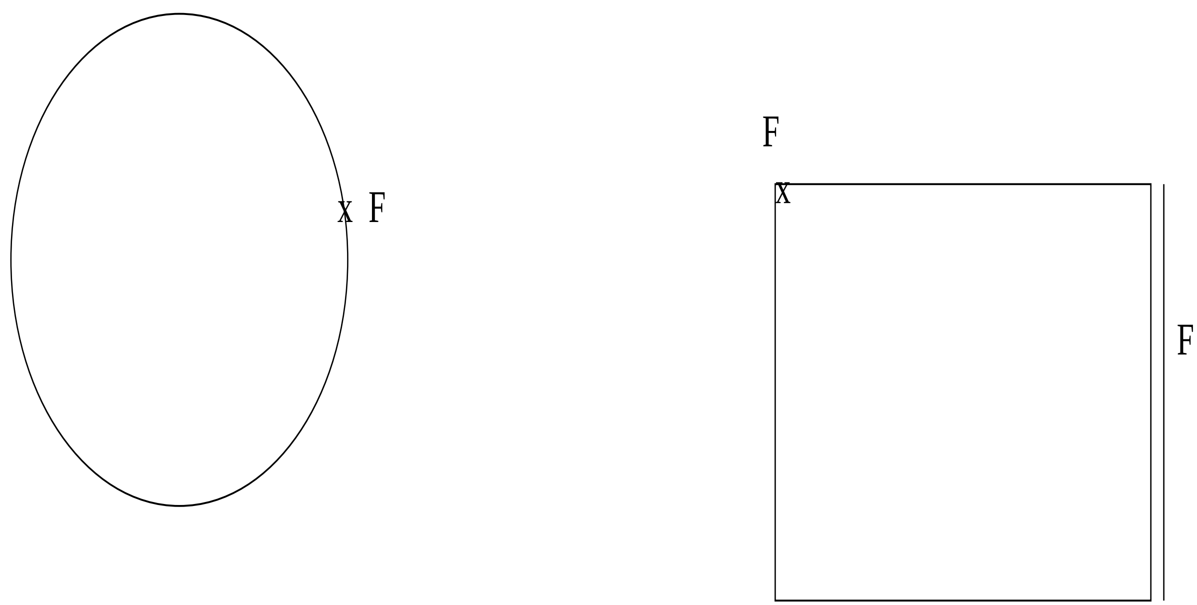


Figure 1: Faces of convex sets

The Main Tool: the FIT Theorem

(Faces of Intersection Theorem)

(by Bonnesen-Fenchel; Dubins; Klee).

Suppose that $C_1,\ C_2$ are closed, convex sets. Then

- F is a face of $C_1 \cap C_2 \;\Leftrightarrow\; F = F_1 \cap F_2$ for some F_i faces of C_i

$\Leftarrow$: easy.

$\Rightarrow$: F_1 and F_2 can be chosen as the *minimal* faces of C_1 and C_2 that contain F. In this case

$$\operatorname{aff} F \quad = \quad \operatorname{aff} F_1 \cap \operatorname{aff} F_2$$

(Example: $C_1 = \{\, x \mid Ax = b \,\}$, $C_2 = \{\, x \mid x \geq 0 \,\}$.)

A simple, important, (and somewhat forgotten) result.

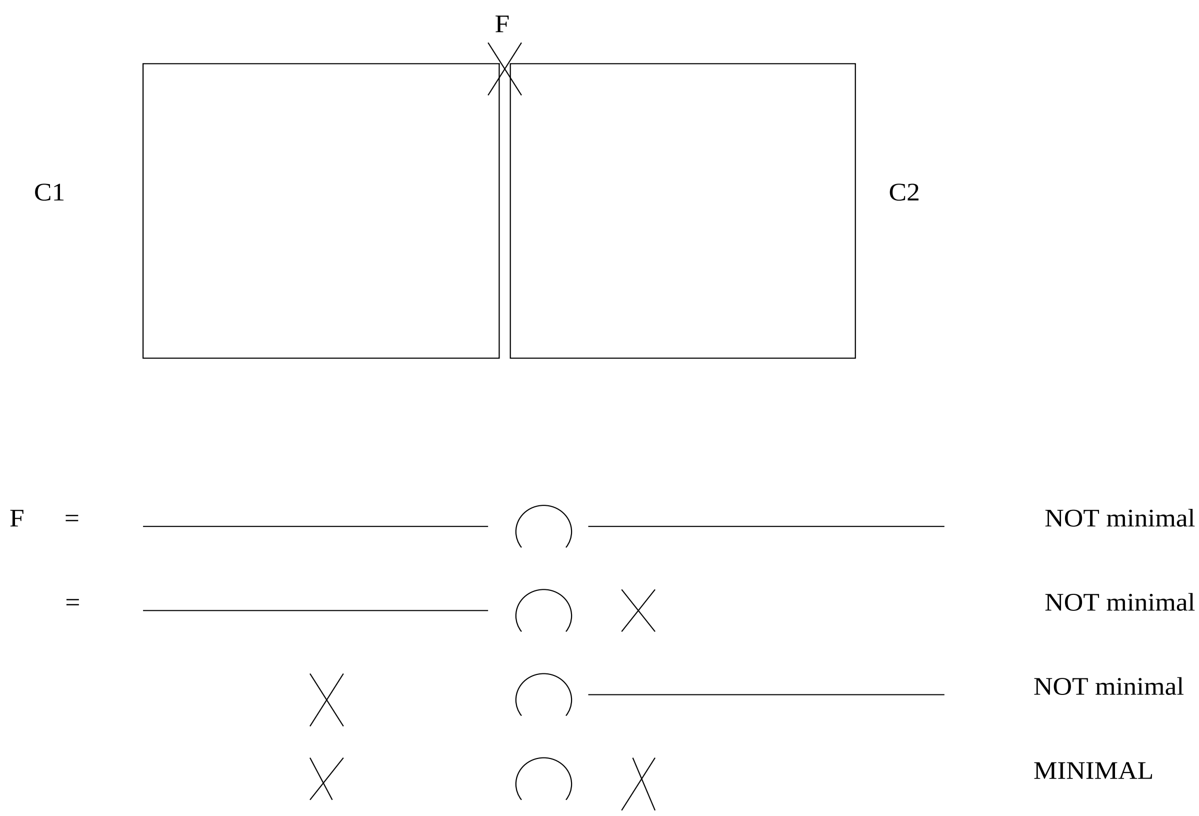


Figure 2:

The Facial Structure of Cone Programs

$$
\begin{array}{llllllll}
 & Min & c^T x & & Max & b^T y & \\
(P) & s.t. & x \in K & & s.t. & z \in K^* & (D) \\
 & & Ax = b & & & A^T y + z = c &
\end{array}
$$

where K is a closed convex cone in R^k,

$$K^* = \{z \mid zx \geq 0 \; \forall x \in K\} \text{ the } \textit{polar} \text{ of } K$$

Interesting choices of K

- $\mathcal{R}^k_+ \; \rightarrow$ LP
- Second-order (SO) cone, $K_2 = \{(t, x) \in \mathcal{R}^{1+d} \mid t \geq \| x \|\}$
- Positive semidefinite matrices $\rightarrow$ SDP

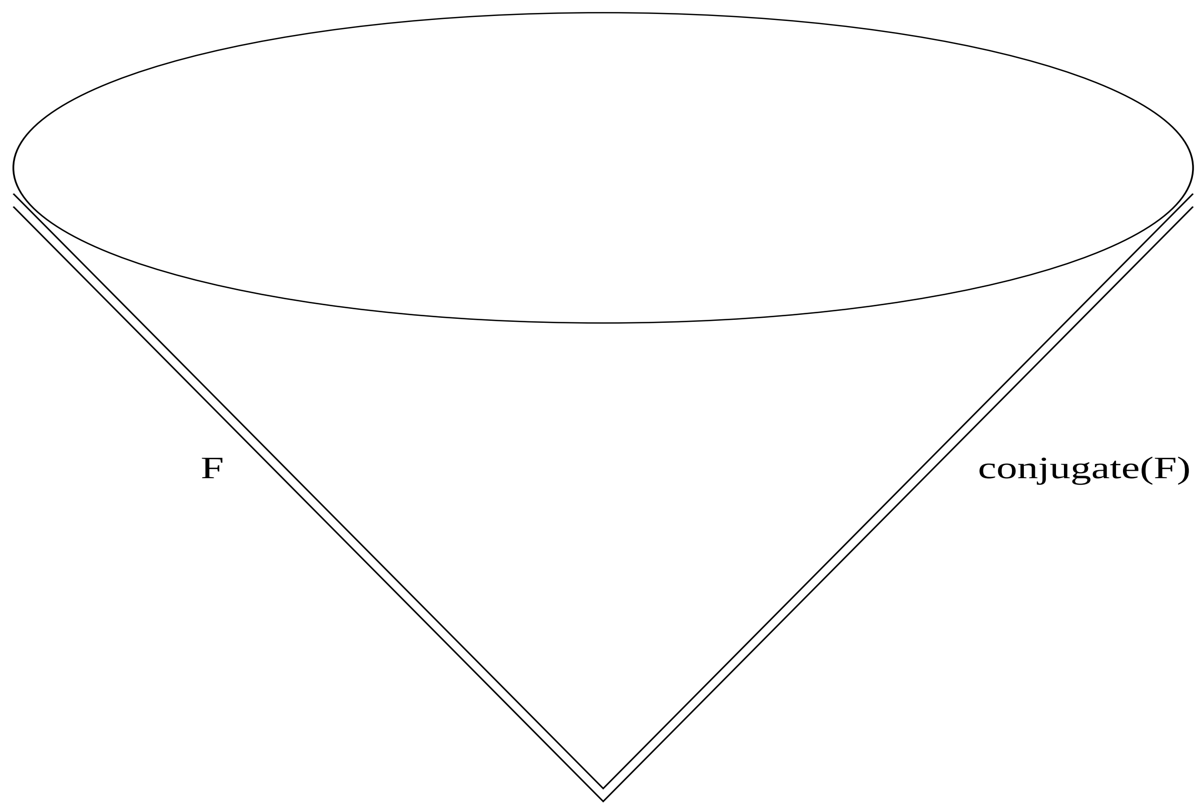


Figure 3: A second order cone

1. Basic solutions in cone programs

Definition: An extreme point of the feasible set of a cone program is called a *basic solution.*

Theorem:

- Suppose x feasible for (P), F the min. face of K that contains x. Then

$$x \text{ is basic} \quad \Leftrightarrow \quad \mathcal{N}(A) \cap \text{lin } F = \{0\}$$

Proof:

$$\begin{array}{lr} x \text{ is basic} & \Leftrightarrow \\ \text{Min. face of feasible set that contains } x \text{ is a singleton} & \Leftrightarrow \\ \text{its affine hull } \{\, x \mid Ax = b,\, x \in \text{lin } F \,\} \text{ is a singleton} & \Leftrightarrow \\ \mathcal{N}(A) \cap \text{lin } F = \{0\} & \end{array}$$

Moreover, if

$$\mathcal{N}(A) \cap \text{lin } F = \{0\}$$

fails, we can find a $\triangle x \neq 0$ in it, and solving

$$\max\{\, t \,:\, x \pm t\triangle x \in F \,\}$$

takes us to a lower-dimensional face of K (need to take care of precision).

Therefore, we can get to a basic solution in finitely many steps.

Characterization of dual basic solutions: analogous ($\mathcal{R}^m \times K^*$ is a cone also).

Special cases

Faces of the interesting cones

$\mathcal{R}^k_+$	$\{x \mid x = (\oplus, \ldots, \oplus, 0, \ldots, 0)\}$
SO cone	$\{\lambda(\parallel x^0 \parallel, x^0) \mid \lambda \geq 0 \}$ for some $x^0 \in \mathcal{R}^d$
Psd cone	$\{X \mid X = \begin{pmatrix} \oplus & 0 \\ 0 & 0 \end{pmatrix}\}$

or the orthogonal rotation of such a set $V(\bullet)V^T$

(Barker and Carlson $'75$)

LP

$$
\begin{array}{rcccccccccc}
 & & 1 & \dots & r & & & & & \\
x: & (& + & \dots & + & | & 0 & \dots & 0 &) \\
\text{lin } F: & (& \times & \dots & \times & | & 0 & \dots & 0 &) \\
A: & (& & B & & | & & N & &)
\end{array}
$$

Corollary:

- x basic $\Leftrightarrow$ columns of B are independent.

SDP

$$X \quad : \quad \begin{pmatrix} \overbrace{+}^{r} & 0 \\ 0 & 0 \end{pmatrix}$$

$$\text{lin } F \quad : \quad \begin{pmatrix} \times & 0 \\ 0 & 0 \end{pmatrix}$$

$$A_i \quad : \quad \begin{pmatrix} (A_i)_{11} & (A_i)_{12} \\ (A_i)_{21} & (A_i)_{22} \end{pmatrix}$$

($A_i \bullet VXV^T = V^T A_i V \bullet X \to$ rescaling.)

Corollary: X basic $\Leftrightarrow \{(A_1)_{11}, \ldots, (A_m)_{11}\}$ span the space of r by r symmetric matrices.

2. Nondegeneracy in cone programs

Definition: F face of K. The set

$$F^{\triangle} = \{z \in K^* \mid z^T x = 0 \; \forall \; x \in F\}$$

is called *the complementary (conjugate) face* of F. (Nonneg. orthant: flip the position of zeros)

Fact:

$$F^{\triangle\triangle} = F$$

for all faces, if K is facially exposed.

Definition: Suppose x is feasible for (P), F is the minimal face of K that contains x. We say that x is *nondegenerate*, if

$$\mathcal{R}(A^T) \cap \text{lin } F^{\triangle} = \{0\}$$

(recall: basic, if $\mathcal{N}(A) \cap \text{lin } F = \{0\}$)

Example: LP

$$
\begin{array}{rccccccccc}
 & & 1 & \ldots & s & & & & & \\
x: & (& + & \ldots & + & | & 0 & \ldots & 0 &) \\
\text{lin } F^{\triangle}: & (& 0 & \ldots & 0 & | & \times & \ldots & \times &) \\
A: & (& & B & & | & & N & &)
\end{array}
$$

Corollary:

- x nondegenerate $\Leftrightarrow$ rows of B are independent.

The duality gap for x and (y, z) is always $x^T z$.

Fact: S.t. x is a nondegenerate primal optimal solution. *Any* dual optimal solution (y, z) must satisfy

$$\begin{array}{ll} A^T y + z = c,\ z \in K^*, \quad z^T x = 0 & \Rightarrow \\ A^T y + z = c,\ z \in F^{\triangle} & \Rightarrow \end{array}$$

it must be basic $\Rightarrow$ dual optimal solution is unique.

Nondegeneracy of dual solution: analogous.

Examples of complementary faces

$$\begin{array}{lll}
\mathcal{R}^k_+ & \{(\oplus,\ldots,\oplus,0,\ldots,0)\} & \{(0,\ldots,0,\oplus,\ldots,\oplus\} \\
\text{SO cone} & \{\lambda(\| x^0 \|, x^0) \mid \lambda \geq 0\ \} & \{\lambda(\| x^0 \|, -x^0) \mid \lambda \geq 0\ \} \\
\text{Psd cone} & \{\begin{pmatrix} \oplus & 0 \\ 0 & 0 \end{pmatrix}\} & \{\begin{pmatrix} 0 & 0 \\ 0 & \oplus \end{pmatrix}\}
\end{array}$$

So, in these cases, it is easy to work out what nondegeneracy means.

3. Strict complementarity in cone programs

Definition: Let x and (y, z) be complementary primal and dual solutions. We say that they are *strictly complementary* if

$$\text{(SC)} \quad x \in \operatorname{ri} F \text{ and } z \in \operatorname{ri} F^{\triangle}$$

for a face F of K.

(LP : total number of nonzeros $= n$; SDP: total rank $= n$.)

4. Analogy of the bound on the number of nonzeros in LP

Suppose that x is feasible for (P), F is the min. face of K that contains x. Then x is basic $\Leftrightarrow$

$$\{\, x \mid Ax = b,\, x \in \mathrm{lin}\ F \,\} \text{ is a singleton}$$

Corollary: x, and F are as above. If x is basic, then

$$\dim F \quad \leq \quad m$$

(LP: $\dim F$ = number of nonzeros in x)

A sharper version: (For LP : Tijssen and Sierksma, Math. Progr. '98) Let d = dimension of dual solution set. Then

$$\dim F \quad \leq \quad m - d$$

with equality holding in LP.

Proof outline The independent dual solutions create dependence in the rows of the system

$$Ax = b, \; [(\text{lin } F)^{\perp}]x = 0$$

$\Longrightarrow$ this system must have more rows.

SDP

Corollary: Let d be the dimension of the set of dual optimal solutions, X a basic optimal solution of the primal SDP. Let r be the rank of X. Then

$$t(r) \quad \leq \quad m - d$$

where $t(r) = r(r+1)/2$ is the r^{th} *triangular number.*

(Existence of such a solution (without d): independently Barvinok, '95).

What fits into this framework

1. Eigenvalue-clustering in eigenvalue-optimization

$f_k(X) =$ sum of the k largest eigenvalues of the symmetric matrix X. **Fact:**

(1) $\exists f'(X) \Leftrightarrow \lambda_k(X) > \lambda_{k+1}(X)$.

(2) If (1) fails, then the subdifferential has dimension

t(multiplicity of $\lambda_k(X)$).

Consider

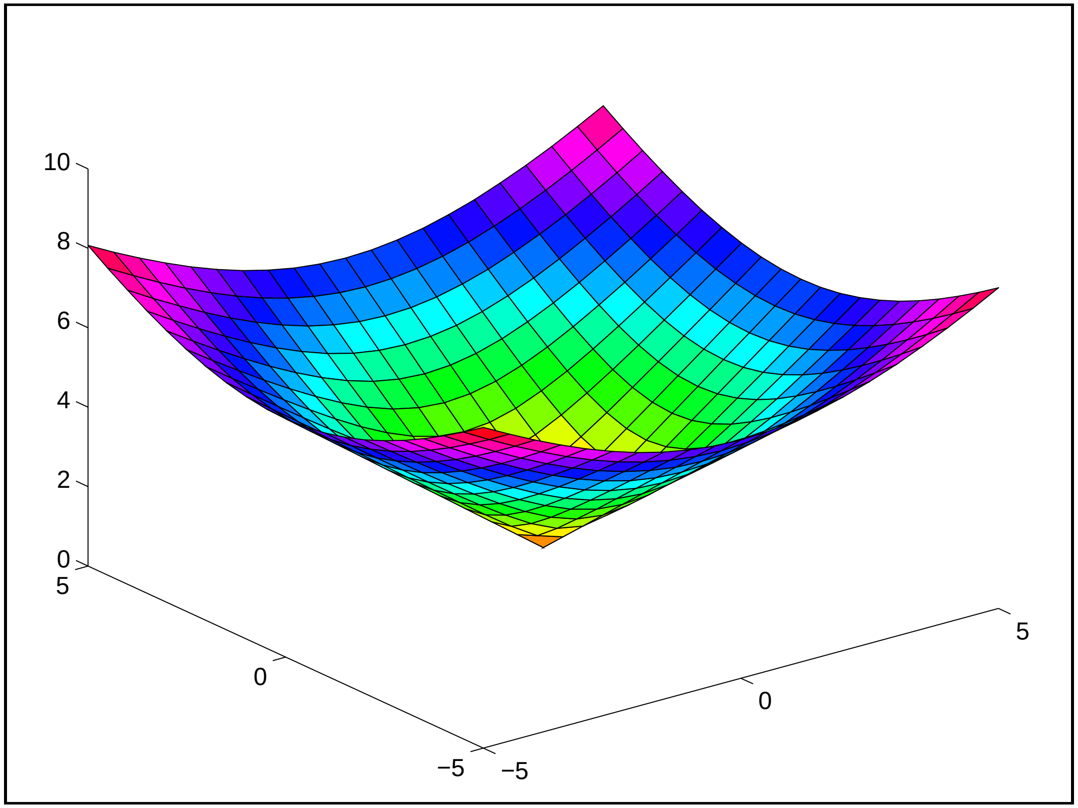
$$\begin{aligned} Min \quad & f_k(X) & \\ s.t. \quad & \mathcal{A}X & = b \end{aligned} \tag{1}$$

Observation : at *optimal solutions* frequently f_k is nondifferentiable $\longrightarrow$ a "model problem" of nonsmooth optimization.

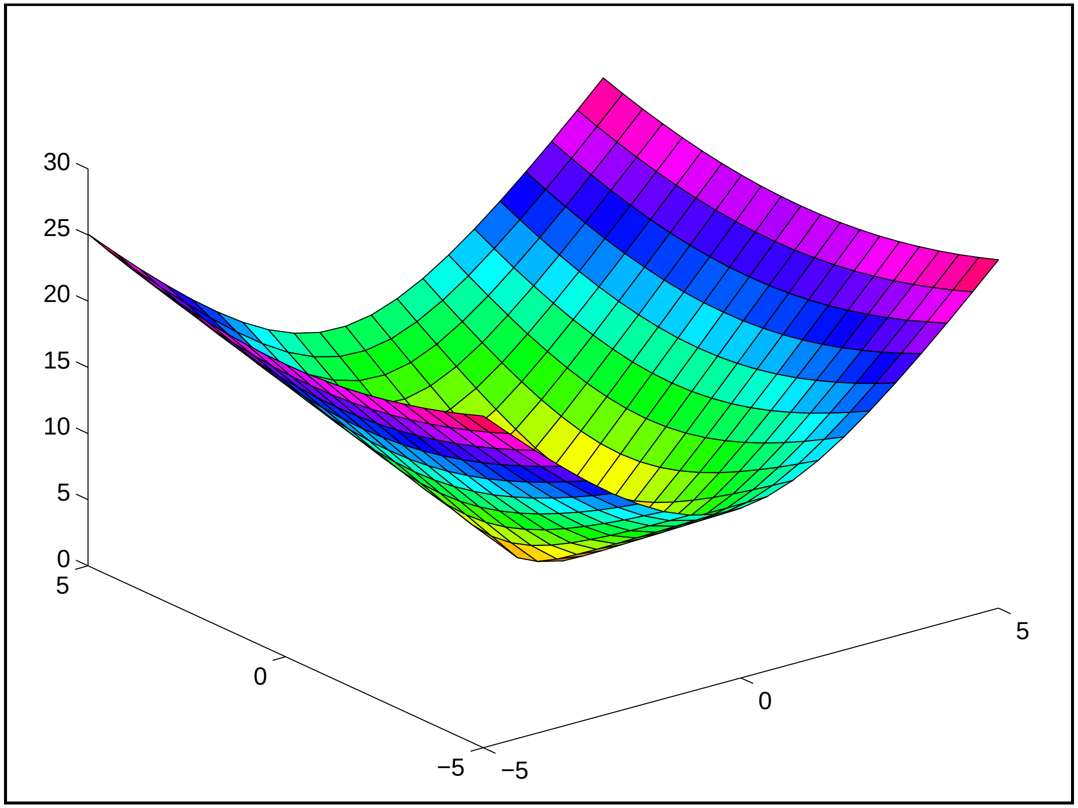
In fact, much of the machinery of NSO was developed to deal with nonsmoothness in (1).

The graph of $\lambda_{\max}(X)$ (parametrizing the feasible X matrices)

$$(1)\ X \quad = \quad \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + x_1 \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} + x_2 \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$



(2) The constraint system is randomly generated.



The clustering has been observed since the seventies without giving sound theoretical explanation: Cullum, Donath and Wolfe ('75); Fletcher; Overton; Shapiro; ... (≥ 20 references)

Theorem: (P, '95) At an extreme point X^* of the solution set of (1)

$$\lambda_k(X^*) = \lambda_{k+1}(X^*)$$

must hold, if the degrees of freedom (= t(n) - # of constraints) is at least $k(n-k)$. Moreover, there is a lower bound on the multiplicity of $\lambda_k(X^*)$ that increases with the the degrees of freedom (analogy in LP : few constraints $\Rightarrow$ few nonzeros in a basic solution).

Outline of proof

Problem (1) can be formulated with extra variables $(z \in \mathcal{R}, V \succeq 0, W \succeq 0)$ as an SDP (Alizadeh; N and N)

X is opt. with eigenvalues $\lambda_1 \geq \ldots \lambda_n \Rightarrow$ the opt. (z^*, V^*, W^*) are

$$\lambda_{k+1} \;\leq\; z^* \;\leq\; \lambda_k \tag{2}$$

$$\begin{aligned} \lambda(V^*) &= (\lambda_1 - z^*, \ldots, \lambda_k - z^*, \quad 0, \quad \ldots, \quad 0 \quad)^T \\ \lambda(W^*) &= (0, \quad \ldots, \quad 0, \quad z^* - \lambda_{k+1}, \ldots, z^* - \lambda_n)^T \end{aligned} \tag{3}$$

X is an extreme point of the solution set $\Rightarrow (z^*, V^*, W^*, X)$ is in a face of dim $\leq 1 \Rightarrow$ ub on rank V^* + rank $W^* \Rightarrow$

$$\lambda_k(X^*) = \lambda_{k+1}(X^*)$$

and lower bound on the multiplicity of $\lambda_k(X^*)$.

2. (Partial) Sensitivity Analysis

Suppose we have a pair of optimal solutions to (P) and (D), called x and (y, z). Now we change the objective from c to $c + t\triangle c$. How big can t be so that x remains optimal ? Denote by t^* the largest t.

(LP: well-known; SDP: Goldfarb and Scheinberg '97)

A simple common generalization, and extension.

Suppose that the primal and dual solutions are unique, and (SC) holds. Let the primal face be F, the dual face $F^{\triangle}$.

Then x is optimal, as long as

$$\begin{aligned} z(t) &\in F^{\triangle} \\ A^T y(t) + z(t) &= c + t\triangle c \end{aligned} \tag{4}$$

is feasible (since the duality gap is $x^T z(t)$).

Write

$$A^T \triangle y + \triangle z \quad = \quad \triangle c$$

with some $\triangle z \in \text{lin } \; F^{\triangle}$ (if it is impossible, then $t^* = 0$).

But the solution to (4) is unique $\Rightarrow$ it must be $(y(0) + t\triangle y, z(0) + t\triangle z)$.

Corollary:

$$t^* \quad = \quad \max\{\, t \,|\, z(0) + t\triangle z \in F^{\triangle} \,\}$$

LP: ratio-test; SDP: computing max. eigenvalue; SO-cone programming: quadratic linesearch.

3. Poly-time solvability of small nonconvex quadratic programs

$$
\begin{aligned}
Min \quad & x^T Q x + 2q^T x \\
s.t. \quad & x^T A_i x + 2b_i^T x + c_i \ \leq 0 \ (i = 1, \ldots, m)
\end{aligned}
\tag{5}
$$

where Q and A_i are not necessarily positive semidefinite $\longrightarrow$ a possibly nonconvex problem.

Equivalent formulation:

$$
\begin{aligned}
Min \quad & Q' \bullet \begin{pmatrix} x_0 \\ x \end{pmatrix} \begin{pmatrix} x_0 \\ x \end{pmatrix}^T \\
s.t. \quad & x_0^2 = 1 \\
& A_i' \bullet \begin{pmatrix} x_0 \\ x \end{pmatrix} \begin{pmatrix} x_0 \\ x \end{pmatrix}^T \ \leq \ c_i' \ (i = 1, \ldots, m)
\end{aligned}
$$

This can be relaxed to

$$
\begin{array}{ll}
Min & Q' \bullet X \\
s.t. & X \succeq 0 \\
 & X_{00} = 1 \\
 & A_i' \bullet X \;\leq\; c_i' \; (i = 1, \ldots, m)
\end{array}
\tag{6}
$$

Suppose that X is a basic optimal solution to (6), the rank of X is r and there are d nontight inequalities. Then

$$t(r) + d \quad \leq \quad m + 1$$

Corollary: If $m = 1$, then there is a rank 1 optimal solution $\Rightarrow$ the relaxation is exact. Also, this solution can be found in polynomial time from a possibly nonbasic solution.

Therefore for $m = 1$ the original problem is solvable in polynomial time (if computations are done exactly : Wolkowicz; Ye; more careful analyisis : Vavasis and Zipfel)

The same is true, if $m = 2$, and there are no linear terms (apparently new).

An extension to general convex programs

Any convex program can be written as

$$\min \{ f_1(x) + \ldots + f_m(x) \}$$

where the f_i's are "elementary" convex functions.

E.g. let $m = 3$,

$$\begin{aligned} f_1(x) &= cx \\ f_2(x) &= \delta\,(\,x \,|\, x \in K\,) \\ f_3(x) &= \delta\,(\,x \,|\, Ax = b\,) \end{aligned}$$

(δ is the *indicator function* of the corresponding convex set).

Denote the set of optimal solutions by S, and suppose that

$$f_i(x) \quad = \quad \alpha_i \quad \text{if}\, x \in S \tag{7}$$

Let

$$C_i \quad = \quad \{\, x \mid f_i(x) \leq \alpha_i \,\}$$

Then

$$S \quad = \quad C_1 \cap \ldots \cap C_m$$

$\longrightarrow$ characterization of the faces of S with the help of the faces of the C_i's.

Nondegeneracy: with the help of the Fenchel-dual.

Special case:

$$\begin{array}{ll} Min & f(x) \\ s.t. & g_i(x) \leq 0 \quad (i = 1, \ldots, m) \end{array} \tag{8}$$

where f and the g_i's are differentiable. Then a solution x is

- nondegenerate in the "facial structure" framework $\Leftrightarrow$ the vectors $\nabla g_{i_1}(x), \ldots, \nabla g_{i_p}(x)$ corresp. to the tight cosntraints are linearly independent.
- strictly complementary with the corresp. dual solution $\Leftrightarrow$ $\nabla f(x)$ is a strict positive combination of these vectors.

Related work

- Nondegeneracy, etc. is a generic property in cone programs. (Shapiro, AHO: for SDP, using differential geometry). In the general framework it is even easier.
- The nonsmoothness of *any* function of eigenvalues can be "predicted" from the case, when it is restricted to *diagonal* matrices (with A. Lewis).

Conclusion

- A theory to describe the "geometry" of general convex programs. Subsumes many known, and provides many new results.
- Facial structure: well-known tool in LP, (surprisingly) also works well in this general context.
- Applications : General results on basic (etc) solutions + structure of a specific problem = better understanding of the problem: "Convex combinatorics".