Quadratic Transformations: Convexity vs Nonconvexity

B. Polyak and P. Scherbakov

Weekly seminar, Lab 7
Institute for Control Science, Moscow
February 17, 2015

Problem

Have: Quadratic mapping $f\colon \mathbb{R}^n \to \mathbb{R}^m$ of the form

$$f(x) = \big(f_1(x), \ldots, f_m(x)\big)^\top, \qquad f_i(x) = (A_i x, x) + 2(b_i, x) + c_i, \quad i = 1, \ldots, m \le n$$

Consider

$$E = \{f(x)\colon\ x \in \mathbb{R}^n\}$$

or

$$F = \{f(x)\colon\ x \in \mathbb{R}^n,\ \|x\| \le 1\}$$

Is F convex or not?

Note: The A_is are not assumed to be sign-(semi)definite!

Applications

- General quadratic programming:

$$\min f_0(x)$$

$$\text{s.t.} \qquad f_i(x) \leq 0,\ i \in I, \quad f_i(x) = 0,\ i \in J$$

- Integer optimization

$$x_i = \{-1, +1\} \iff x_i^2 = 1$$

- Convex relaxation: When is it tight?

- S-theorem (absolute stability): When do the two quadratic inequalities imply the third one?

- Robustness analysis (so-called real μ)

- Power systems: Electric power is a quadratic function of the currents or voltages

- Quantum systems

Convexity vs Nonconvexity

- Simplest example:

$$\min(Ax, x) \quad \text{over} \quad \|x\| = 1$$

This problem is nonconvex! However the solution is explicit:

$$x^* = e_1,$$

where e_1 is the eigenvector associated with the minimal eigenvalue of A

- Titles of papers:

— *Hidden convexity in some nonconvex quadratically constrained quadratic programming* [Ben-Tal, Teboulle, 1996]

— *Permanently going back and forth between the "quadratic world" and the "convexity world" in optimization* [J.-B. Hiriart-Urruty, M. Tork, 2002]

- When the images of quadratic maps are convex?

Simple Illustrations

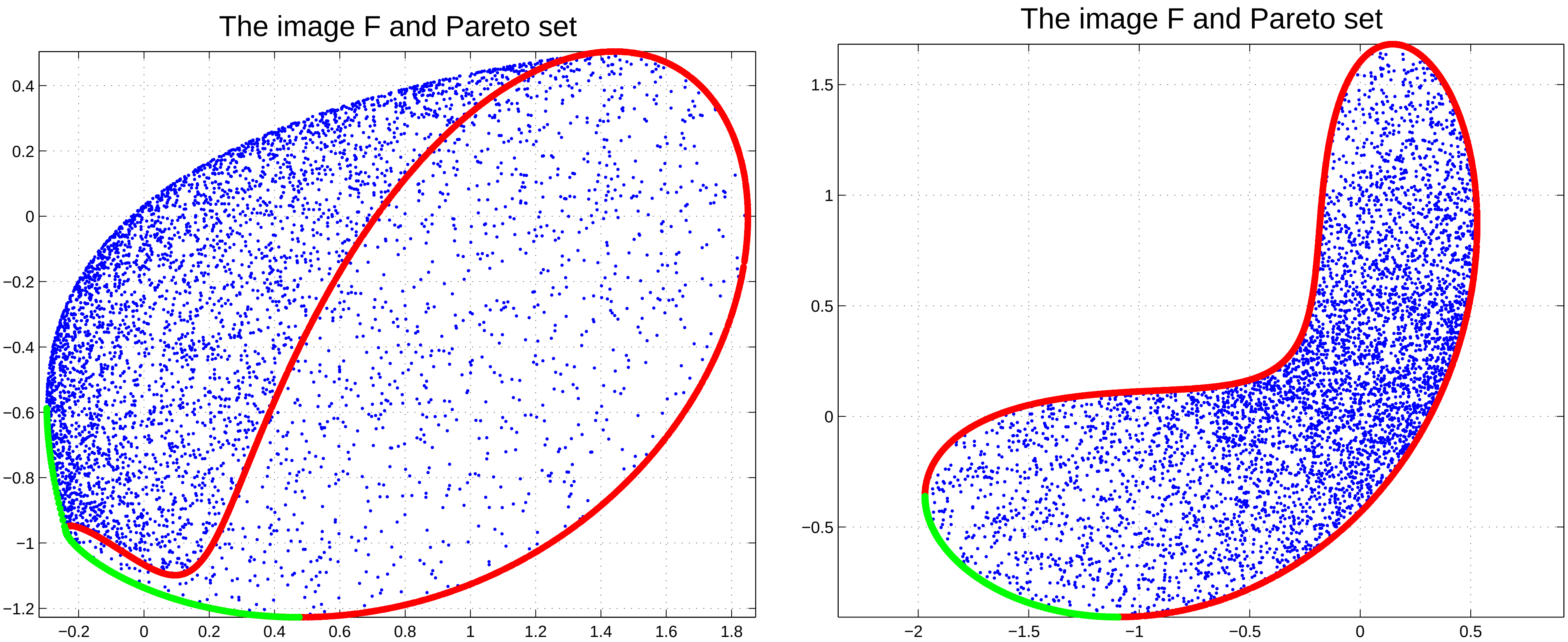


Figure 1: $n = m = 2$: Surface (red) and interior (blue)

Known Facts

- $m = 2,\ b_i = 0 \quad \Longrightarrow \quad E$ is convex [Dines, 1941]
- $m = 2,\ b_i = 0,\ n \geq 3 \quad \Longrightarrow \quad F$ is convex [Brickman, 1961]
- $m = 3,\ b_i = 0,\ n \geq 3,\ \sum c_i A_i \succ 0 \quad \Longrightarrow \quad E$ is convex [Polyak, 1998]
- m is arbitrary, $b_i = 0,\ A_i$ commute $\quad \Longrightarrow \quad F$ is convex [Fradkov, 1973]
- m is arbitrary, $b_i \neq 0 \quad \Longrightarrow \quad F_\varepsilon$ is convex for $\varepsilon > 0$ small enough [Polyak, 2001]

Convex Hull (i)

The idea of convex relaxations for quadratic problems goes back to [Shor, 1986];

see also [Nesterov 1998], [Zhang 2000], [Beck and Teboulle, 2005].

Recent survey:

Luo, Ma, So, Ye, Zhang, *Semidefinite relaxation of quadratic optimization problems,*
IEEE Sig. Proc. Magazine, 2010.

Two typical results:

Lemma 1. *For* $b_i = 0$ *have*

$$Conv(F) = \{\mathcal{A}(X): \; X \succcurlyeq 0, \; Tr X \le 1\},$$

where $X = X^\top \in \mathbb{R}^{n \times n}$, $\mathcal{A}(X) = \big(\langle A_1, X\rangle, \dots, \langle A_m, X\rangle\big)^\top$, *and* $\langle A, X\rangle = Tr AX$.

Convex Hull (ii)

Lemma 2. *In the general case* $(b_1 \neq 0)$ *have*

$$Conv(F) = \{\mathcal{H}(X) \colon\ X \succcurlyeq 0,\ TrX \leq 1,\ X_{n+1,n+1} = 1\}$$

where $X = X^\top \in \mathbb{R}^{(n+1)\times(n+1)}$, $\mathcal{H}(X) = \big(\langle H_1, X\rangle, \ldots, \langle H_m, X\rangle\big)^\top$,

and $H_i = \begin{bmatrix} A_i & b_i \\ b_i^T & c_i \end{bmatrix}$.

Idea of proof: $(A_i x, x) = \langle A_i, xx^\top\rangle = \langle A_i, X\rangle,\ X \succcurlyeq 0,\ \mathrm{rank}X = 1,\ \ TrX = \|x\|^2$.

For $z = (x;t) \in \mathbb{R}^{n+1}$ have $(H_i z, z) = (A_i x, x) + 2(b_i, x)t + c_i t^2 = f_i(x)$ if $t = 1$.

How Close are F and $Conv(F)$?

- the 0.87856 result for MaxCut [Goemans and Williamson, 1995]
- entropy distance [Barvinok, 2014]

However in general F and $Conv(F)$ can differ a lot.

Example:

$$n = m = 3, \quad f_1(x) = x_1x_2, \quad f_2(x) = x_1x_3, \quad f_3(x) = x_2x_3$$

FIGURE (four-dent image)

Nonconvexity Certificate N1

Assume

- $b_i = 0$
- *the LMI* $\sum c_i A_i \succcurlyeq 0$ *has no nonzero solution* c
- *there exists* $h \notin E$.

Then E *is nonconvex.*

Explanation: E is a cone; if it is convex it is either coincides with the whole space

or lies in a half-space

Example: $TrA_i = 0$ for all i as above

Similar result holds for $b_i \neq 0$

Convexity Certificate C1

Assume that for all $c \neq 0$, the minimal eigenvalue of the matrix $\sum c_i A_i$ is simple.

Then F is convex. If the eigenvalue is negative, then F is strictly convex.

Indeed, for $G = Conv(F)$ have

$$\min_{g \in G}(c, g) = \min_{X \succcurlyeq 0, TrX \leq 1} \langle A, X \rangle;$$

the unique minimizer is either $X = e_1 e_1^\top$ (if $\lambda_1 < 0$) or $X = 0$ (if $\lambda_1 \geq 0$).

Here, $\lambda_1 = \lambda_{\min}(A)$, $A = \sum c_i A_i$, and e_1 is the corresponding eigenvector.

Hence, any supporting hyperplane has a unique intersection point both with F and its convex hull.

Corollary

If the $A_i s$ *are matrices with positive entries, the "positive part" of* F *is convex.*

Extension [Jetakumar et al., 2009]). The matrix A is referred to as a *z-matrix* if $a_{ij} \leq 0$, $i \neq j$.

If all $A_i s$ *are z-matrices, the* Pareto set *in the multi-objective minimization problem with maps*

$$f_i(x) = (A_i x, x)$$

is convex.

Local Convexity

Theorem 1. *Let* $b_i = 0$ *and the eigensystem of the matrix* $A^c = \sum c_i A_i \not\succeq 0$ *be* $\lambda_1 < \lambda_2 \le \cdots \le \lambda_n,\ \lambda_1 < 0,$ *and* $e_1, \ldots, e_n$. *Denote* $x^c = e_1,\ f^c = f(x^c)$.

Then f^c *is the boundary point of* F*, and for all* d *such that*

$$\|d - c\|_\infty \le \varepsilon < \frac{\min\{-\lambda_1, \lambda_2 - \lambda_1\}}{2\kappa L};$$

f^d *is also a boundary point of* F.

Here, $L = \sum \|A_i\|$, $\kappa = \|T\|\|T^\top\|$, *where* T *is the matrix with columns* $e_i, T^{-1} = T^\top$, *and the matrix norm is* $\|M\| = \max_i \sum_j |m_{ij}|$, *while the vector* f^d *is defined similarly to* f^c.

In principle, this can provide us with a global convexity certificate: Sample points $c^1, \ldots, c^N$ (e.g., randomly on the unit sphere $\|c\| = 1$). If, for each c^i, Theorem 1 holds, and the corresponding small boxes cover the unit sphere, then F is convex.

Nonconvexity Certificate N2

Theorem 2. Let $m \geq 3$, $n \geq 3$, $b_i = 0$, and let for some $c_1, \ldots, c_m$, the matrix $A^c = \sum c_i A_i$ has eigenvalues $\lambda_1 = \lambda_2 < \lambda_3 \leq \cdots \leq \lambda_n$, $\lambda_1 < 0$,

and eigenvectors $e_1, \ldots, e_n$.

If $f^1 = f(e_1) \neq f^2 = f(e_2)$, $f^0 = \big((A_1 e_1, e_2), \ldots, (A_m e_1, e_2)\big)^\top \neq 0$,

then F is nonconvex.

Proof: $\operatorname{Arg}\min\limits_{f \in F}(c, f) = f^\alpha$, where $f^\alpha = f(x^\alpha)$; $x^\alpha = e_1\alpha_1 + e_2\alpha_2$; $\alpha_1^2 + \alpha_2^2 = 1$.

The set of such points f^α is a 2D ellipse, which is nondegenerate due to the assumptions $f^1 \neq f^2$, $f^0 \neq 0$.

Hence, the intersection of F and the supporting hyperplane $(c, f) = \lambda_1$ is nonconvex.

Example

$$n = m = 3, \quad A_1 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}, \quad A_2 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{pmatrix}, \quad A_3 = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

$$b_1 = b_2 = b_3 = 0$$

For $c = (0, -1, -1)^\top$ we have

$$\lambda_1 = \lambda_2 = -3, \;\; \lambda_3 = -2, \quad e_1 = (1, 0, 0)^\top, \quad e_2 = (0, 0, 1)^\top$$

$$f^1 = (0, 1, 2)^\top, \quad f^2 = (0, 2, 1)^\top, \quad f^0 = (1, 0, 0)^\top,$$

hence, Theorem 2 holds and F is nonconvex.

FIGURE (concave ice-cream cone)

How to Find Such c ?

Choose $a \neq 0$ and solve SDP in the scalar variables c, μ:

$$\max \mu$$

$$\text{s.t.} \qquad \sum c_i A_i \succcurlyeq \mu I$$

$$(c, a) = 1$$

It is likely that μ is the minimal eigenvalue of $\sum c_i A_i$ with multiplicity 2.

Then check other conditions of Theorem 2 for this c to assure the nonconvexity of F

Otherwise try another a.

Another Approach

Sample points on the boundary of $G = Conv(F) = \{\mathcal{A}(X)\colon\ X \succcurlyeq 0,\ TrX \leq 1\}$:

- If $\sum c_i A_i \succcurlyeq 0$ is infeasible, have $G = Conv(F) = \{\mathcal{A}(X)\colon\ X \succcurlyeq 0,\ TrX = 1\}$.

  Take $X^0 = (1/n)I$, pick a random matrix "direction" H, and compute

  $$X^1 = X^0 + \gamma_1\Big(H - \frac{1}{n}TrH \cdot I\Big), \quad \gamma_1 = \max\Big\{\gamma\colon\ X^0 + \gamma_1\Big(H - \frac{1}{n}TrH \cdot I\Big) \succcurlyeq 0\Big\}$$

  Then X^1 is a random point on the boundary of G.

  Compute the normal vector $c \neq 0$ to G at X^1 by solving the LMI

  $$\sum c_i A_i \succcurlyeq \mu I, \quad \Big\langle \sum c_i A_i - \mu I, X^1 \Big\rangle = 0$$

- If $\sum c_i A_i \succcurlyeq 0$ is feasible, set $X^0 = (1/2n)I$ and proceed similarly without scaling $TrX^1 = 1$.

If X^1 is not rank one, then it is likely to sit at the "flat" part of the boundary of G, and we are within the framework of Theorem 2.

Otherwise, keep on generating the points X^k similarly.

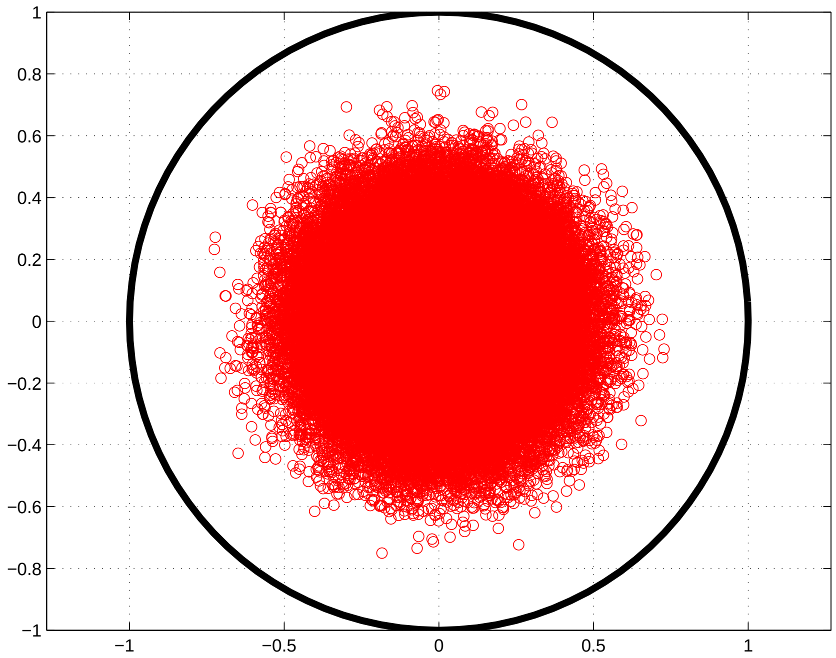
Different Problem Formulation

The goal is to approximate F or its boundary *point-wise*.

Brute force method: Monte Carlo. Sample points x^k in the unit ball $\|x\| \leq 1$

and calculate $f^k = f(x^k)$; adopt as the desired approximation.

OK for small n but very poor for n large.



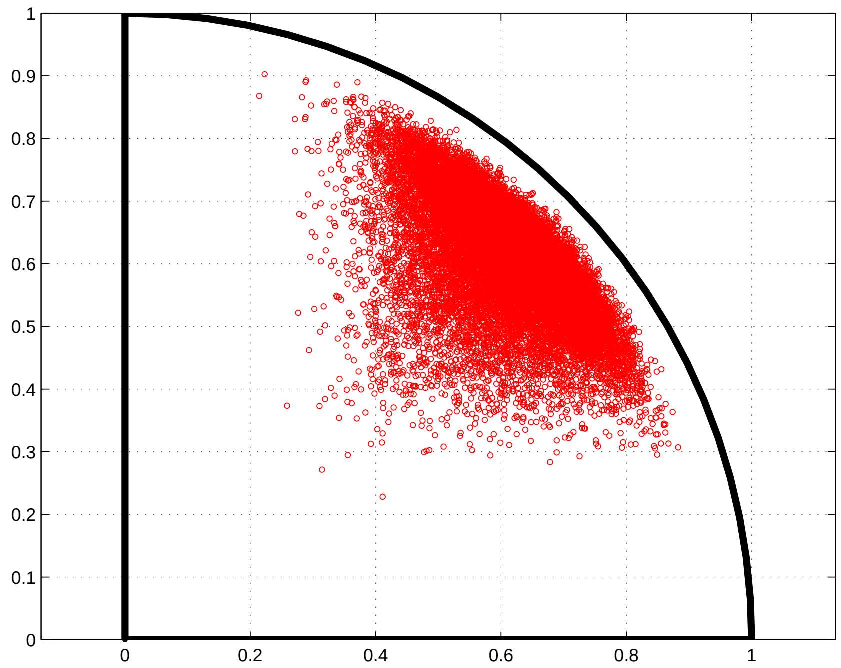


Figure 2: $n = 25$: Linear (left) and homogenous diagonal quadratic (right) mappings. True boundary and $N = 100,000$ MC samples

A Necessary Condition for a Boundary Point

Lemma 3 (The homogeneous case). If f^* is a boundary point of F, then there exists a $c \neq 0$ such that $f^* = f(e_i)$, e_i being an eigenvector of $A = \sum c_i A_i$ corresponding to a positive or zero eigenvalue of A.

Similar result holds for the nonhomogeneous case.

Hence, for m small, we can sample $c^k \in \mathbb{R}^m$, $\|c^k\| = 1$,

compute A^ks and their eigenvectors and

the points f^k.

However this is not efficient for $m > 2$.

2D image

Special case $m = 2$ is worth analyzing.

Of course only nonhomogenous maps are of interest (in view of Brickman theorem)

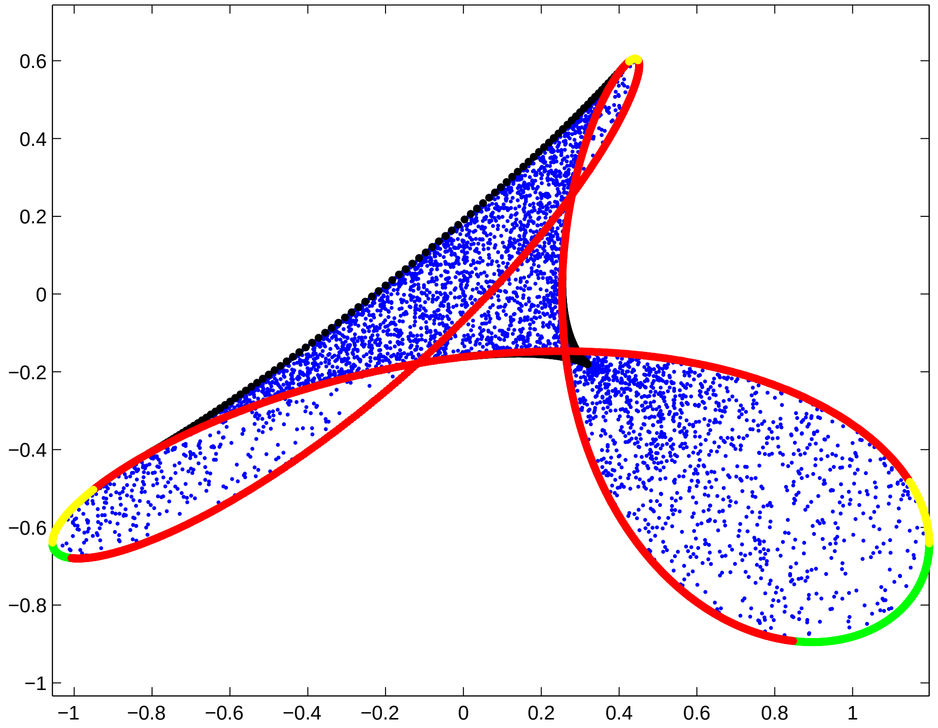


Figure 3: $n = 3$, $m = 2$

Future Work