Optimization problems with eigenvalue constraints

*Masaru Ito (Nihon University, Japan)

Bruno F. Lourenço (Institute of Statistical Mathematics, Japan)

SIAM Conference on Optimization
July 1, 2023
Session MS147 “Recent Advancements in Conic Optimization - Part III of III”

Eigenvalue programming

Motivation: Consider the optimization problem

$$\text{minimize}_{x \in \mathcal{V}} \quad f(x) \quad \text{subject to} \quad \underbrace{x \in \mathcal{D}}_{\text{simple}}, \quad \underbrace{\lambda(x) \in \mathcal{C}}_{\text{eigenvalue constraints}}$$

or the feasibility problem

$$\text{Find} \quad x \quad \text{such that} \quad x \in \mathcal{D}, \quad \lambda(x) \in \mathcal{C}$$

- $f : \mathcal{V} \to \mathbb{R}$ is a smooth function defined on a real inner product space $\mathcal{V}$.
- $\lambda(x) = (\lambda_1(x), \ldots, \lambda_r(x)) \in \mathbb{R}^r$: "eigenvalues" of x
- $\mathcal{C}, \mathcal{D}$ are "simple" (e.g., polyhedral or projection is computable)
- The constraint $\lambda(x) \in \mathcal{C}$ can be nonconvex even if $\mathcal{C}$ is convex
- Applications: Low-rank matrix completion, Inverse eigenvalue problems

Target problem

Motivation: Consider the optimization problem

$$\text{minimize}_{x \in \mathcal{V}} \quad f(x) \quad \text{subject to} \quad x \in \mathcal{D}, \quad \lambda(x) \in \mathcal{C}$$

or the feasibility problem

$$\text{Find} \quad x \quad \text{such that} \quad x \in \mathcal{D}, \quad \lambda(x) \in \mathcal{C}$$

	$x \in \mathcal{D}$	$\lambda(x) \in \mathcal{C}$	Both
Optimization problem	"NLP"	This talk	Future interest
Feasibility problem	"Simple"	"Simple"	This talk

Agenda

- The projection onto the constraint $\{x : \lambda(x) \in \mathcal{C}\}$ can be done by projection onto $\mathcal{C} \cap \operatorname{ran} \lambda$ + "spectral decomposition" [Gowda 2019]
- Analyze the convergence of projected gradient method for $\min\{f(x) : \lambda(x) \in \mathcal{C}\}$
- Some numerical examples

Notion of eigenvalues

Examples of vector space $\mathcal{V}$:

- $\mathcal{V} = \mathcal{S}^n$, the vector space of real symmetric matrices.
  $\lambda(x) = (\lambda_1(x), \ldots, \lambda_n(x))$: eigenvalues of x as usual (here $\lambda_i \geq \lambda_{i+1}$).
- $\mathcal{V} = \mathbb{C}^{m\times n}$ or $\mathbb{R}^{m\times n}$
  $\lambda(x) = (\sigma_1(x), \ldots, \sigma_{\min(m,n)}(x))$: singular values of x (here $\sigma_i \geq \sigma_{i+1}$).
- $\mathcal{V}$: Euclidean Jordan algebra. For instance:
  - $\mathcal{V} = \mathbb{R}^n$, $\lambda(x) = x^{\downarrow}$ (rearrangement of x in descending order)
  - $\mathcal{V} = \mathcal{S}^n$, the same as above.
  - $\mathcal{V} = \mathbb{R}^{1+n}$: Euclidean Jordan algebra associated with Lorentz cone
    $\lambda(t,x) = (\lambda_+(t,x), \lambda_-(t,x)) = \frac{1}{\sqrt{2}}(t + \|x\|, t - \|x\|)$
- Direct product of the above spaces.

We will introduce the Fan-Theobald-von Neumann (FTvN) system [Gowda 2019] to unify these concepts.

Preliminary 1: Spectral decomposition

Examples of vector space $\mathcal{V}$:

- $\mathcal{V} = \mathcal{S}^n$:

$$x = U\,\mathrm{Diag}(\lambda(x))U^T = \sum_{i=1}^{n} \lambda_i(x) u_i u_i^T, \quad U = (u_1, \ldots, u_n) \text{ orthogonal}$$

- $\mathcal{V} = \mathbb{C}^{m\times n}$ or $\mathbb{R}^{m\times n}$

$$x = \sum_{i=1}^{\min(m,n)} \sigma_i(x) u_i v_i^*, \quad \{u_i\}, \{v_j\} \text{ orthonormal}$$

- $\mathcal{V} = \mathbb{R}^{1+n}$: Euclidean Jordan algebra associated with Lorentz cone:

$$(t,x) = \lambda_+(t,x)e_+ + \lambda_-(t,x)e_-, \quad e_\pm = \frac{1}{\sqrt{2}\|x\|}(\pm x, \|x\|)$$

Preliminary 2: Simultaneous diagonalization

For $\mathcal{V} = \mathcal{S}^n$,

- Trace inner product: $\langle x, y \rangle_{\mathcal{S}^n} := \mathrm{tr}(xy)$.
- Spectral decomposition $x = U\,\mathrm{Diag}(\lambda(x))U^T$, U orthogonal
- Fact 1: $\langle x, y \rangle_{\mathcal{S}^n} \leq \langle \lambda(x), \lambda(y) \rangle_{\mathbb{R}^n}$ holds.
- Fact 2: $\langle x, y \rangle_{\mathcal{S}^n} = \langle \lambda(x), \lambda(y) \rangle_{\mathbb{R}^n} \Leftrightarrow x, y$ are simultaneously diagonalizable:

$$\exists U \text{ orthogonal s.t. } x = U\,\mathrm{Diag}(\lambda(x))U^T,\ y = U\,\mathrm{Diag}(\lambda(y))U^T$$

Fact (It will be the definition of FTvN system)

For any $x, y \in \mathcal{S}^n$, one can construct $x' \in \mathcal{S}^n$ such that

$$\lambda(x) = \lambda(x') \quad \text{and} \quad \langle x', y \rangle_{\mathcal{S}^n} = \langle \lambda(x'), \lambda(y) \rangle_{\mathbb{R}^n}$$

Method:

1. Decompose $\quad y = U_y\,\mathrm{Diag}(\lambda(y))U_y^T$
2. Replace eigenvalues: $x' := U_y\,\mathrm{Diag}(\lambda(x))U_y^T$.

FTvN system

Definition: Fan-Theobald-von Neumann (FTvN) system

Let $\mathcal{V}$ be an inner product space and $\lambda : \mathcal{V} \to \mathbb{R}^r$.
The tuple $(\mathcal{V}, \mathbb{R}^r, \lambda)$ is called a FTvN system if

- $\|x\|_{\mathcal{V}} = \|\lambda(x)\|_{\mathbb{R}^r} \qquad (\|x\|_{\mathcal{V}} = \sqrt{\langle x, x \rangle_{\mathcal{V}}})$
- $\langle x, y \rangle_{\mathcal{V}} \leq \langle \lambda(x), \lambda(y) \rangle_{\mathbb{R}^r}$
- $\forall x, y \in \mathcal{V}$, $\exists x' \in \mathcal{V}$ such that $\lambda(x) = \lambda(x')$ and $\langle x', y \rangle_{\mathcal{S}^n} = \langle \lambda(x'), \lambda(y) \rangle_{\mathbb{R}^r}$

Target problem: Let $(\mathcal{V}, \mathbb{R}^r, \lambda)$ be an FTvN system.

$$\text{minimize}_{x \in \mathcal{V}} \quad f(x) \quad \text{subject to} \quad \lambda(x) \in \mathcal{C}$$

Useful fact: projection

Target problem: Let $(\mathcal{V}, \mathbb{R}^r, \lambda)$ be an FTvN system.

$$\text{minimize}_{x\in\mathcal{V}} \quad f(x) \quad \text{subject to} \quad \lambda(x) \in \mathcal{C}$$

Useful fact [Gowda 2019]: Projection of c onto $\{x : \lambda(x) \in \mathcal{C}\}$

$$(*) \quad \min_{x\in\mathcal{V}} \frac{1}{2}\|x - c\|_{\mathcal{V}}^2 \ \text{ s.t. } \ \underbrace{\lambda(x) \in \mathcal{C}}_{\text{possibly nonconvex}}$$

can be obtained using the projection

$$(**) \quad \min_{\xi\in\mathbb{R}^r} \frac{1}{2}\|\xi - \lambda(c)\|_{\mathbb{R}^r}^2 \text{ s.t. } \xi \in \mathcal{C} \cap \underbrace{\operatorname{ran}\lambda}_{\text{convex cone}}$$

Simple setting: $\operatorname{ran}\lambda = \mathbb{R}^r_{\downarrow}$

Method:

1. $\xi^* := \operatorname{proj}_{\mathcal{C}\cap\operatorname{ran}\lambda}(\lambda(c))$: solution to $(**)$
2. $x^* \in \mathcal{V}$ is such that $\lambda(x^*) = \xi^*, \langle c, x^*\rangle_{\mathcal{V}} = \langle \lambda(c), \lambda(x^*)\rangle_{\mathbb{R}^r}$
   (when $\mathcal{V} = \mathcal{S}^n$: $c = U_c \operatorname{Diag}(\lambda(c)) U_c^T$ and set $x^* = U_c \operatorname{Diag}(\xi^*) U_c^T$)

Projected gradient method

Target problem:

$$\text{minimize}_{x \in \mathcal{V}} \quad f(x) \quad \text{subject to} \quad \lambda(x) \in \mathcal{C}$$

We shall analyze the projected gradient method

$$x_{k+1} = \text{proj}_{\lambda^{-1}(\mathcal{C})}(x_k - \alpha_k \nabla f(x_k))$$

More explicit form:

Algorithm: Projected gradient method

1. $y_k \leftarrow x_k - \alpha_k \nabla f(x_k)$
2. $\xi_k^* \leftarrow$ solution to $(**)$ $\min_{\xi \in \mathbb{R}^r} \frac{1}{2}\|\xi - \lambda(y_k)\|_{\mathbb{R}^r}^2$ s.t. $\xi \in \mathcal{C} \cap \text{ran}\, \lambda$
3. $x_{k+1} \in \mathcal{V}$ is such that $\lambda(x_{k+1}) = \xi_k^*$ and $\langle y_k, x_{k+1} \rangle_{\mathcal{V}} = \langle \lambda(y_k), \lambda(x_{k+1}) \rangle_{\mathbb{R}^r}$
(when $\mathcal{V} = \mathcal{S}^n$: $y_k = U \,\text{Diag}(\lambda(y_k)) U^T$ and set $x_{k+1} = U \,\text{Diag}(\xi_k^*) U^T$)

Convergence results

- Target problem: $\min_{x \in \mathcal{V}} \; f(x) \;$ s.t. $\; \lambda(x) \in \mathcal{C}$
- Projected gradient: $x_{k+1} = \mathrm{proj}_{\lambda^{-1}(\mathcal{C})}(x_k - \alpha_k \nabla f(x_k))$

Assumptions for convergence results

- ∇f is L-Lipschitz continuous, $\alpha_k \in [\alpha_{\min}, \alpha_{\max}]$ where $\alpha_{\min}, \alpha_{\max} \in (0, 1/L)$.
- $\mathcal{C}$ is a semialgebraic set (a union of system polynomial inequalities)
- f, λ are semialgebraic functions (i.e. their graphs are semialgebraic sets)

λ is semialgebraic in many cases.

Theorem: Convergence results

Under the above assumptions, let $\{x_k\}$ be the iterates of projected gradient.

- Global convergence to stationary point: if $\{x_k\}$ is bounded, then it converges to x^* such that $0 \in \nabla f(x^*) + N_{\lambda^{-1}(\mathcal{C})}(x^*)$.
- Local convergence to optimal solution x^*: if $\|x_0 - x^*\|$ and $f(x_0) - f(x^*)$ are sufficiently close to zero, $\{x_k\}$ converge to x^*.

Proof is an application of [Attouch et al. 2013].

Application to feasibility problem

Consider the feasibility problem:

$$\text{Find } \; x \; \text{ such that } \; x \in \mathcal{D}, \quad \lambda(x) \in \mathcal{C},$$

where $\mathcal{C}, \mathcal{D}$ are convex, semialgebraic and $\mathrm{proj}_{\mathcal{C} \cap \mathrm{ran}\, \lambda}(\cdot), \mathrm{proj}_{\mathcal{D}}(\cdot)$ are computable.

- Reformulate as

$$\text{minimize } \; f(x) = \frac{1}{2} \mathrm{dist}(x, \mathcal{D})^2 \; \text{ s.t. } \; \lambda(x) \in \mathcal{C}$$

  and apply projected gradient method.
- The resulting iteration is a weighted alternating projection

$$y_k = (1 - \alpha_k) x_k + \alpha_k \mathrm{proj}_{\mathcal{D}}(x_k), \qquad x_{k+1} = \mathrm{proj}_{\lambda^{-1}(\mathcal{C})}(y_k)$$

- Convergence results follow from the previous slide.

Numerical example 1: Inverse eigenvalue problem

Inverse eigenvalue problem: Given linear map $\mathcal{A} : \mathbb{R}^d \to \mathcal{V}$, $b \in \mathcal{V}$ and $\lambda^* \in \mathbb{R}^r$,

$$\text{Find} \quad z \in \mathbb{R}^d \quad \text{such that} \quad \lambda(\mathcal{A}z + b) = \lambda^*$$

Reformulation with variable transformation $x := \mathcal{A}z + b$ is

$$\text{Find} \quad x \in \mathcal{V} \quad \text{such that} \quad x \in \mathcal{D} := \underbrace{\operatorname{ran} \mathcal{A} + b}_{d\text{-dim affine set}}, \quad \lambda(x) \in \mathcal{C} := \{\lambda^*\}.$$

- We examine traditional setting $\mathcal{V} = \mathcal{S}^n$ and more structured one

$$\mathcal{V} := \underbrace{\mathbb{R}^{1+n} \times \cdots \times \mathbb{R}^{1+n}}_{m \text{ times}} \times \mathcal{S}^n,$$

$$\lambda(z_1, \ldots, z_m, X) := (\lambda(z_1), \ldots, \lambda(z_m), \lambda(X)) \in \mathbb{R}^{2m+n}, \quad z_i \in \mathbb{R}^{1+n}, \ X \in \mathcal{S}^n$$

with $n = 10$, $m = 1, 3, 5$. $\mathbb{R}^{1+n}$ is the Jordan algebra associated with $\mathcal{L}_2^n$.

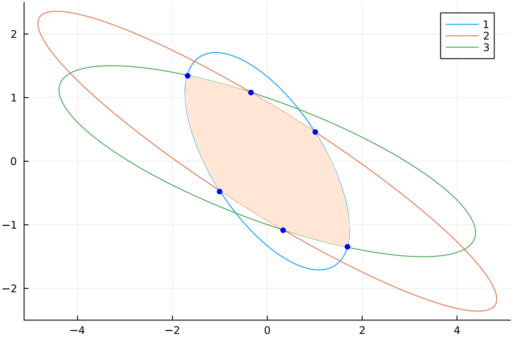
Numerical example 1: Inverse eigenvalue problem

- Algorithm: Projected gradient with stepsize 0.99
- x^* is a known feasible solution such that $\lambda(x^*)$ has multiplicity two.
- Initial point: $\|x_0 - x^*\|/\|x^*\| = r_k := 100/2^k$ for a known feasible point x^*. Rerun increasing k until finding ε-feasible point within 10000 iterations.
- $r_k \in [0, 100]$ estimates the relative distance of local convergence.
- Average over 10 random instances:

n	m	d	mean iter	mean of r_k
10	0	5	10.7	100
		16	21.7	100
		38	96.7	100
	1	6	9.2	100
		19	20.3	100
		46	99.6	100
	5	11	15.2	100
		33	51.8	100
		77	1087.5	71.3 (max 100, min 0.19)

Numerical example 2: Intersection of ellipsoids

Suppose that we have m-ellipsoids $\mathcal{C}_i$ on $\mathbb{R}^n$ $(1 \leq i \leq m)$.
Figure: $m = 3$ ellipsoids



- Problem: Find $x \in \mathbb{R}^n$ such that $x \in \mathcal{C}_1 \cap \cdots \cap \mathcal{C}_m$
  and x belong to the boundary of at least ℓ of these ellipsoids.
- $\ell = 0$: orange region
- $\ell = 1$: boundary of the orange reigon
- $\ell = 2$: blue points

Numerical example 2: rank constraint reformulation

- Let $\mathcal{C}_i = \{x : \|Q_i(x - c_i)\|_{\mathbb{R}^n} \leq 1\}$ for some $Q_i \in \mathcal{S}^n_{++}$ and center $c_i \in \mathbb{R}^n$
- Define the affine map

$$\mathcal{A}x + b = (\underbrace{1, Q_1(x - c_1)}_{\in \mathbb{R}^{1+n}}; \cdots ; \underbrace{1, Q_i(x - c_m)}_{\in \mathbb{R}^{1+n}}) \in \prod_{i=1}^{m} \mathbb{R}^{1+n}$$

- Define the eigenvalue

$$\lambda(z_1, \ldots, z_m) := (\lambda(z_1), \ldots, \lambda(z_m))^{\downarrow} \in \mathbb{R}^{2m}$$

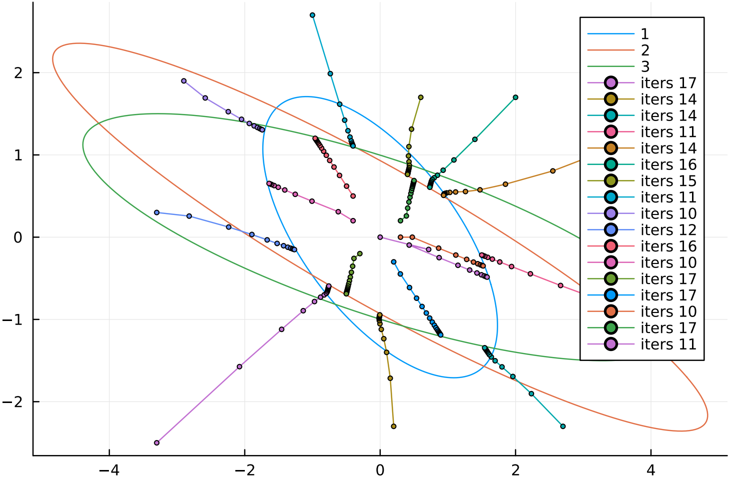
- Problem: Find $x \in \mathbb{R}^n$ such that $x \in \bigcap_{1 \leq i \leq \ell} \mathcal{C}_i$
  and x belong to the boundary of at least ℓ of these ellipsoids.
- Reformulation:
  Find $x \in \mathbb{R}^n$ such that $\lambda(\mathcal{A}x + b)$ is nonnegative and has at least ℓ-zeros.

Reformulation

Find $z \in \prod_{i=1}^{m} \mathbb{R}^{1+n}$ such that $z \in \mathcal{D} := \operatorname{ran} \mathcal{A} + b$ and
$\lambda(z) \in \mathcal{C} := \{\lambda \in (\mathbb{R}^{2m}_+)^{\downarrow} \mid \lambda_{n-\ell} = \cdots = \lambda_\ell = 0\}$

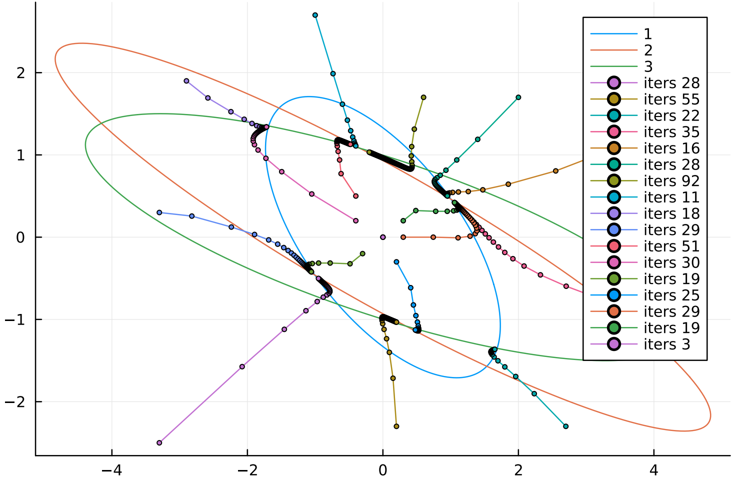
Numerical example 2: The case $\ell = 1$ for three ellipsoids

Projected gradient trajectory from various initial points:



Numerical example 2: The case $\ell = 2$ for three ellipsoids

Projected gradient trajectory from various initial points:



Summary

Motivation: $\quad \text{minimize}_{x\in\mathcal{V}} \;\; f(x) \;\; \text{subject to} \;\; x \in \mathcal{D}, \;\; \lambda(x) \in \mathcal{C}$

or

Find x such that $x \in \mathcal{D}, \;\; \lambda(x) \in \mathcal{C}$

	$x \in \mathcal{D}$	$\lambda(x) \in \mathcal{C}$	Both
Optimization problem	"NLP"	This talk	Future interest
Feasibility problem	"Simple"	"Simple"	This talk

- The projection onto the constraint $\{x : \lambda(x) \in \mathcal{C}\}$ can be done by $\text{proj}_{\mathcal{C}\cap \text{ran}\,\lambda}(\cdot)$ + "spectral decomposition" [Gowda 2019]
- Analyze the convergence of projected gradient method for $\min\{f(x) : \lambda(x) \in \mathcal{C}\}$

Future interests:

- Algorithm for more general setting with convergence results
- Potential application

Thank you for your attention!

References I

H. Attouch, J. Bolte, and B. F. Svaiter, Convergence of descent methods for semi-algebraic and tame problems: proximal algorithms, forward-backward splitting, and regularized Gauss-Seidel methods, Math. Program., 137:91–129, 2013

M. Baes. Convexity and differentiability properties of spectral functions and spectral mappings on Euclidean Jordan algebras. Linear Algebra and its Applications, 422:664–700, 2007.

J. Faraut and A. Korányi. Analysis on symmetric cones. Oxford mathematical monographs. Clarendon Press, Oxford, 1994.

M. S. Gowda, Optimizing certain combinations of spectral and linear/distance functions over spectral sets, arXiv:1902.06640v2, 2019

M.-F. R. Jacek Bochnak, Michel Coste. Real Algebraic Geometry. Springer-Verlag, Berlin, Heidel- berg, 1998.

A. S. Lewis and H. S. Sendov. Nonsmooth analysis of singular values. part I: Theory. Set-Valued Analysis, 13:213 241, 2005.