Positive semidefinite rank

Pablo A. Parrilo

Laboratory for Information and Decision Systems
Electrical Engineering and Computer Science
Massachusetts Institute of Technology

Based on joint work with **Hamza Fawzi** (MIT), **João Gouveia** (U. Coimbra),
James Saunderson (MIT), **Richard Robinson** and **Rekha Thomas** (U. Washington)

Cargese 2014

Question: representability of convex sets

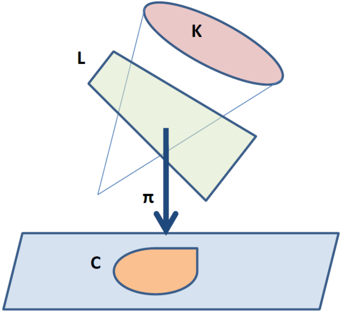
Existence and efficiency:

- When is a convex set representable by conic optimization?
- How to quantify the number of additional variables that are needed?

Given a convex set C, is it possible to represent it as

$$C = \pi(K \cap L)$$

where K is a cone, L is an affine subspace, and π is a linear map?



Factorizations

Given a matrix $M \in \mathbb{R}^{m \times n}$, can factorize it as $M = AB$, i.e.,

$$\mathbb{R}^n \xrightarrow{B} \mathbb{R}^k \xrightarrow{A} \mathbb{R}^m$$

Ideally, k is small (matrix M is low-rank), so we're factorizing through a "small subspace."

Why is this useful?

- Realization theory (e.g., factorization of a Hankel matrix)
- Principal component analysis (e.g., factorization of covariance of a Gaussian process)

And many others... Standard notion in linear algebra.

Factorizations

Given a matrix $M \in \mathbb{R}^{m \times n}$, can factorize it as $M = AB$, i.e.,

$$\mathbb{R}^n \xrightarrow{B} \mathbb{R}^k \xrightarrow{A} \mathbb{R}^m$$

Ideally, k is small (matrix M is low-rank), so we're factorizing through a "small subspace."

Why is this useful?

- Realization theory (e.g., factorization of a Hankel matrix)
- Principal component analysis (e.g., factorization of covariance of a Gaussian process)

And many others... Standard notion in linear algebra.

More Factorizations...

However, often we need further conditions on $M = AB$...

- Norm conditions on the factors A, B:
  - Want factors A, B to be "small" in some norm
  - Well-studied topic in Banach space theory, through the notion of *factorization norms*
  - For instance, the nuclear norm $\|M\|_\star := \min_{A,B\,:\,M=AB} \frac{1}{2}(\|A\|_F^2 + \|B\|_F^2)$
- Nonnegativity conditions:
  - Matrix M is (componentwise) nonnegative, and so must be the factors.
  - This is the *nonnegative factorization* problem.
  - Many applications, e.g., in probability (conditional independence) and machine learning (additive features).

More Factorizations...

However, often we need further conditions on $M = AB$...

- Norm conditions on the factors A, B:
  - Want factors A, B to be "small" in some norm
  - Well-studied topic in Banach space theory, through the notion of *factorization norms*
  - For instance, the nuclear norm $\|M\|_\star := \min_{A,B\,:\,M=AB} \frac{1}{2}(\|A\|_F^2 + \|B\|_F^2)$
- Nonnegativity conditions:
  - Matrix M is (componentwise) nonnegative, and so must be the factors.
  - This is the *nonnegative factorization* problem.
  - Many applications, e.g., in probability (conditional independence) and machine learning (additive features).

Nonnegative factorization and hidden variables

Let X, Y be discrete random variables, with joint distribution

$$\mathbf{P}[X = i, Y = j] = P_{ij}.$$

The nonnegative rank of P is the smallest support of a random variable Z, such that X and Y are *conditionally independent* given Z (i.e., $X - Z - Y$ is Markov):

$$\mathbf{P}[X = i, Y = j] = \sum_{s=1,\ldots,k} \mathbf{P}[Z = s] \cdot \mathbf{P}[X = i | Z = s] \cdot \mathbf{P}[Y = j | Z = s].$$

- Relations with information theory, "correlation generation," communication complexity, etc.

As we'll see, also fundamental in optimization ...

Nonnegative factorization and hidden variables

Let X, Y be discrete random variables, with joint distribution

$$\mathbf{P}[X = i, Y = j] = P_{ij}.$$

The nonnegative rank of P is the smallest support of a random variable Z, such that X and Y are *conditionally independent* given Z (i.e., $X - Z - Y$ is Markov):

$$\mathbf{P}[X = i, Y = j] = \sum_{s=1,\ldots,k} \mathbf{P}[Z = s] \cdot \mathbf{P}[X = i | Z = s] \cdot \mathbf{P}[Y = j | Z = s].$$

- Relations with information theory, "correlation generation," communication complexity, etc.

As we'll see, also fundamental in optimization ...

Conic factorizations

We're interested in a different class: *conic factorizations* [GPT11]

Let $M \in \mathbb{R}_+^{m\times n}$ be a nonnegative matrix, and $\mathcal{K}$ be a convex cone in $\mathbb{R}^k$. Then, we want $M = AB$, where

$$\mathbb{R}_+^n \xrightarrow{B} \mathcal{K} \xrightarrow{A} \mathbb{R}_+^m$$

- M maps the nonnegative orthant into the nonnegative orthant.
- For $\mathcal{K} = \mathbb{R}_+^k$, this is a standard nonnegative factorization.
- In general, factorize a linear map through a "small cone"

Important special case: $\mathcal{K}$ is the cone of psd matrices...

Conic factorizations

We're interested in a different class: *conic factorizations* [GPT11]

Let $M \in \mathbb{R}_+^{m \times n}$ be a nonnegative matrix, and $\mathcal{K}$ be a convex cone in $\mathbb{R}^k$. Then, we want $M = AB$, where

$$\mathbb{R}_+^n \xrightarrow{B} \mathcal{K} \xrightarrow{A} \mathbb{R}_+^m$$

- M maps the nonnegative orthant into the nonnegative orthant.
- For $\mathcal{K} = \mathbb{R}_+^k$, this is a standard nonnegative factorization.
- In general, factorize a linear map through a "small cone"

Important special case: $\mathcal{K}$ is the cone of psd matrices...

Conic factorizations

We're interested in a different class: *conic factorizations* [GPT11]

Let $M \in \mathbb{R}_+^{m \times n}$ be a nonnegative matrix, and $\mathcal{K}$ be a convex cone in $\mathbb{R}^k$. Then, we want $M = AB$, where

$$\mathbb{R}_+^n \xrightarrow{B} \mathcal{K} \xrightarrow{A} \mathbb{R}_+^m$$

- M maps the nonnegative orthant into the nonnegative orthant.
- For $\mathcal{K} = \mathbb{R}_+^k$, this is a standard nonnegative factorization.
- In general, factorize a linear map through a "small cone"

Important special case: $\mathcal{K}$ is the cone of psd matrices...

PSD rank of a nonnegative matrix

Let $M \in \mathbb{R}^{m \times n}$ be a nonnegative matrix.

Definition [GPT11]: The PSD rank of M, denoted rank_{psd}, is the smallest r for which there exists $r \times r$ PSD matrices $\{A_1, \ldots, A_m\}$ and $\{B_1, \ldots, B_n\}$ such that

$$M_{ij} = \text{trace}\, A_i B_j, \qquad i = 1, \ldots, m \quad j = 1, \ldots, n.$$

(The maps are then given by $x \mapsto \sum_i x_i A_i$, and $Y \mapsto \text{trace}\, Y B_j$.)

Natural definition, generalization of nonnegative rank.

PSD rank of a nonnegative matrix

Let $M \in \mathbb{R}^{m \times n}$ be a nonnegative matrix.

Definition [GPT11]: The PSD rank of M, denoted rank_{psd}, is the smallest r for which there exists $r \times r$ PSD matrices $\{A_1, \ldots, A_m\}$ and $\{B_1, \ldots, B_n\}$ such that

$$M_{ij} = \mathrm{trace}\, A_i B_j, \qquad i = 1, \ldots, m \quad j = 1, \ldots, n.$$

(The maps are then given by $x \mapsto \sum_i x_i A_i$, and $Y \mapsto \mathrm{trace}\, Y B_j$.)

Natural definition, generalization of nonnegative rank.

PSD rank of a nonnegative matrix

Let $M \in \mathbb{R}^{m \times n}$ be a nonnegative matrix.

Definition [GPT11]: The PSD rank of M, denoted rank_{psd}, is the smallest r for which there exists $r \times r$ PSD matrices $\{A_1, \dots, A_m\}$ and $\{B_1, \dots, B_n\}$ such that

$$M_{ij} = \operatorname{trace} A_i B_j, \qquad i = 1, \dots, m \quad j = 1, \dots, n.$$

(The maps are then given by $x \mapsto \sum_i x_i A_i$, and $Y \mapsto \operatorname{trace} Y B_j$.)

Natural definition, generalization of nonnegative rank.

Example (I)

$$M = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix}.$$

M admits a psd factorization of size 2:

$$A_1 = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \quad A_2 = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \quad A_3 = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}$$

$$B_1 = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \quad B_2 = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \quad B_3 = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}.$$

One can easily check that the matrices A_i and B_j are positive semidefinite, and that $M_{ij} = \langle A_i, B_j \rangle$. This factorization shows that $\text{rank}_{\text{psd}}(M) \leq 2$, and in fact $\text{rank}_{\text{psd}}(M) = 2$.

Example (I)

$$M = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix}.$$

M admits a psd factorization of size 2:

$$A_1 = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \quad A_2 = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \quad A_3 = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}$$

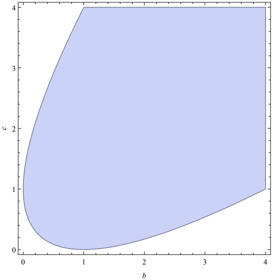
$$B_1 = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \quad B_2 = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \quad B_3 = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}.$$

One can easily check that the matrices A_i and B_j are positive semidefinite, and that $M_{ij} = \langle A_i, B_j \rangle$. This factorization shows that $\text{rank}_{\text{psd}}(M) \leq 2$, and in fact $\text{rank}_{\text{psd}}(M) = 2$.

Example (II)

Consider the matrix

$$M(a,b,c) = \begin{bmatrix} a & b & c \\ c & a & b \\ b & c & a \end{bmatrix}.$$



- Usual rank of $M(a,b,c)$ is 3, unless $a = b = c$ (then, rank is 1).
- One can show that

$$\text{rank}_{\text{psd}}\,(M(a,b,c)) \leq 2 \quad \Longleftrightarrow \quad a^2 + b^2 + c^2 \leq 2(ab + bc + ac).$$

Back to representability...

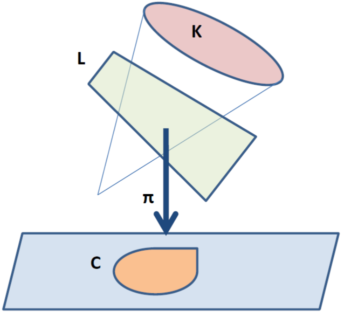
Existence and efficiency:

- When is a convex set representable by conic optimization?
- How to quantify the number of additional variables that are needed?

Given a convex set C, is it possible to represent it as

$$C = \pi(K \cap L)$$

where K is a cone, L is an affine subspace, and π is a linear map?



Question: representability of convex sets

"Complicated" objects are sometimes easily described as "projections" of "simpler" ones.

A general theme: algebraic varieties, unitaries/contractions, graphical models, . . .

Extended formulations

These representations are usually called *extended formulations*.
Particularly relevant in combinatorial optimization (e.g., TSP).

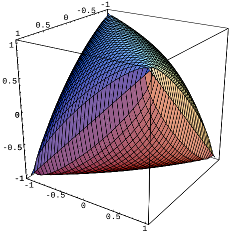
Seminal work by Yannakakis (1991). He gave a beautiful characterization (for LP) in terms of *nonnegative factorizations*, and used it to disprove the existence of "symmetric" LPs for the TSP polytope.
Nice recent survey by Conforti-Cornuéjols-Zambelli (2010).

Our goal: to understand this phenomenon for convex optimization (SDP), not just LP.

"Extended formulations" in semidefinite programming

Many convex sets can be modeled by SDP and LMIs. Among others:

- Sums of eigenvalues of symmetric matrices
- Convex envelope of univariate polynomials
- Multivariate polynomials that are sums of squares
- Unit ball of matrix operator and nuclear norms
- Geometric and harmonic means
- (Some) orbitopes – convex hulls of group orbits



How to produce extended formulations?

- Clever, non-obvious constructions
  - E.g., the KYP (Kalman-Yakubovich-Popov) lemma, LMI solution of interpolation problems (e.g., AAK, Ball-Gohberg-Rodman), ...
  - Work of Nesterov/Nemirovski, Boyd/Vandenberghe, Scherer, Gahinet/Apkarian, Ben-Tal/Nemirovski, Sanyal/Sottile/Sturmfels, etc.
- Systematic "lifting" techniques
  - Reformulation/linearization (Sherali-Adams, Lovasz-Schrijver)
  - Sum of squares (or moments), Positivstellensatz, (Lasserre, Putinar, P.)
  - Determinantal representations (Helton/Vinnikov, Nie)
  - Hyperbolic polynomials (Guler, Renegar)

Much research in this area. More recently, efforts towards understanding the general case (not just specific constructions).

Polytopes

What happens in the case of polytopes?

$$P = \{x \in \mathbb{R}^n : f_i^T x \leq 1\}$$

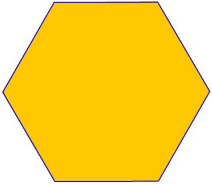
(WLOG, compact with $0 \in \operatorname{int} P$).

Polytopes have a finite number of facets f_i and vertices v_j.
Define a nonnegative matrix, called the *slack matrix* of the polytope:

$$[S_P]_{ij} = 1 - f_i^T v_j, \qquad i = 1, \ldots, |F| \quad j = 1, \ldots, |V|$$

Example: hexagon (I)

Consider a regular hexagon in the plane.



It has 6 vertices, and 6 facets. Its slack matrix has rank 3, and is

$$S_H = \begin{pmatrix} 0 & 0 & 1 & 2 & 2 & 1 \\ 1 & 0 & 0 & 1 & 2 & 2 \\ 2 & 1 & 0 & 0 & 1 & 2 \\ 2 & 2 & 1 & 0 & 0 & 1 \\ 1 & 2 & 2 & 1 & 0 & 0 \\ 0 & 1 & 2 & 2 & 1 & 0 \end{pmatrix}.$$

"Trivial" representation requires 6 facets. Can we do better?

Cone factorizations and representability

"Geometric" LP formulations exactly correspond to "algebraic" factorizations of the slack matrix.

For polytopes, this amounts to a *nonnegative factorization* of the slack matrix:

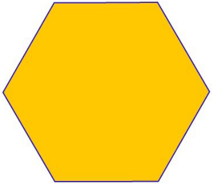
$$S_{ij} = \langle a_i, b_j \rangle, \qquad i = 1, \ldots, v, \qquad j = 1, \ldots, f$$

where a_i, b_i are nonnegative vectors.

Theorem (Yannakakis 1991): The minimal lifting dimension of a polytope is equal to the *nonnegative rank* of its slack matrix.

Example: hexagon (II)

Regular hexagon in the plane.



Its slack matrix is

$$S_H = \begin{pmatrix} 0 & 0 & 1 & 2 & 2 & 1 \\ 1 & 0 & 0 & 1 & 2 & 2 \\ 2 & 1 & 0 & 0 & 1 & 2 \\ 2 & 2 & 1 & 0 & 0 & 1 \\ 1 & 2 & 2 & 1 & 0 & 0 \\ 0 & 1 & 2 & 2 & 1 & 0 \end{pmatrix}.$$

Nonnegative rank is 5.

Beyond LPs and nonnegative factorizations

LPs are nice, but what about broader representability questions?

In [GPT11], a generalization of Yannakakis' theorem to the full convex case. General theme:

"Geometric" extended formulations exactly correspond to "algebraic" factorizations of a slack operator.

polytopes/LP	convex sets/convex cones
slack matrix	slack operators
vertices	extreme points of C
facets	extreme points of polar C°
nonnegative factorizations	conic factorizations
$S_{ij} = \langle a_i, b_j \rangle, \quad a_i \geq 0, b_j \geq 0$	$S_{ij} = \langle a_i, b_j \rangle, \quad a_i \in K, b_j \in K^*$

Polytopes, semidefinite programming, and factorizations

Even for polytopes, SDP representations can be interesting.

(Example: the *stable set* or *independent set* polytope of a graph. For *perfect graphs*, efficient SDP representations exist, but no known subexponential LP.)

Thm: ([GPT 11]) Positive semidefinite rank of slack matrix exactly characterizes the complexity of SDP-representability.

PSD factorizations of slack matrix $\iff$ SDP extended formulations

Polytopes, semidefinite programming, and factorizations

Even for polytopes, SDP representations can be interesting.

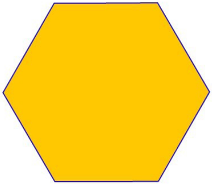
(Example: the *stable set* or *independent set* polytope of a graph. For *perfect graphs*, efficient SDP representations exist, but no known subexponential LP.)

Thm: ([GPT 11]) Positive semidefinite rank of slack matrix exactly characterizes the complexity of SDP-representability.

PSD factorizations of slack matrix $\iff$ SDP extended formulations

SDP representation of hexagon

A regular hexagon in the plane.



Projection onto (x, y) of a 5-dimensional spectrahedron:

$$\begin{bmatrix} 1 & x & y & t \\ x & (1+r)/2 & s/2 & r \\ y & s/2 & (1-r)/2 & -s \\ t & r & -s & 1 \end{bmatrix} \succeq 0$$

Representation has nice symmetry properties (equivariance).

Towards understanding psd rank

Generally difficult, since it's semialgebraic (inequalities matter), and symmetry group is "small".

- Basic properties
- Other interpretations (e.g., information-theoretic)
- Dependence on field and topology of factorizations
- Special cases and extensions

Basic inequalities

- For any nonnegative matrix M

$$\frac{1}{2}\sqrt{1+8\,\mathrm{rank}(M)}-\frac{1}{2}\leq \mathrm{rank}_{psd}(M)\leq \mathrm{rank}_{+}(M).$$

- Gap between $\mathrm{rank}_{+}(M)$ and $\mathrm{rank}_{psd}(M)$ can be arbitrarily large:

$$M_{ij}=(i-j)^2=\left\langle \begin{pmatrix} i^2 & -i \\ -i & 1 \end{pmatrix}, \begin{pmatrix} 1 & j \\ j & j^2 \end{pmatrix} \right\rangle$$

has $\mathrm{rank}_{psd}(M)=2$, but $\mathrm{rank}_{+}(M)=\Omega(\log n)$.

Arbitrarily large gaps between all pairs of ranks (rank, rank_{+} and $\mathrm{rank}_{\mathrm{psd}}$). For slack matrices of polytopes, arbitrarily large gaps between rank and rank_{+}, and rank and $\mathrm{rank}_{\mathrm{psd}}$.

Real and rational PSD rank can be different

If the matrix M has rational entries, sometimes it is natural to consider only factors A_i, B_i that are rational.

In general we have

$$\operatorname{rank}_{\text{psd}}(M) \leq \operatorname{rank}_{\text{psd}\,\mathbb{Q}}(M).$$

and inequality can be strict. Explicit examples (Fawzi-Gouveia-Robinson).

Same question for nonnegative rank is open (since Cohen-Rothblum 93).

Real and rational PSD rank can be different

If the matrix M has rational entries, sometimes it is natural to consider only factors A_i, B_i that are rational.

In general we have

$$\text{rank}_{\text{psd}}(M) \leq \text{rank}_{\text{psd}\,\mathbb{Q}}(M).$$

and inequality can be strict. Explicit examples (Fawzi-Gouveia-Robinson).

Same question for nonnegative rank is open (since Cohen-Rothblum 93).

Real and rational PSD rank can be different

If the matrix M has rational entries, sometimes it is natural to consider only factors A_i, B_i that are rational.

In general we have

$$\text{rank}_{\text{psd}}(M) \leq \text{rank}_{\text{psd}\,\mathbb{Q}}(M).$$

and inequality can be strict. Explicit examples (Fawzi-Gouveia-Robinson).

Same question for nonnegative rank is open (since Cohen-Rothblum 93).

Computing and bounding PSD rank

Computing PSD rank seems to be quite hard, both in theory and practice. Are there situations where it is easy?

- $k = 1$ easy, since $\text{rank}_{\text{psd}}(M) = 1$ if and only if $\text{rank}(M) = 1$.
- $k = 2$ also easy, since it is reducible to semidefinite programming (e.g., via S-lemma).

What about for fixed psd rank (even $k = 3$)? Analogue of Arora-Ge-Kannan-Moitra polynomiality result for nonnegative rank?

What about bounds? And before psd, what about nonnegative rank?

Computing and bounding PSD rank

Computing PSD rank seems to be quite hard, both in theory and practice. Are there situations where it is easy?

- $k = 1$ easy, since $\text{rank}_{\text{psd}}(M) = 1$ if and only if $\text{rank}(M) = 1$.
- $k = 2$ also easy, since it is reducible to semidefinite programming (e.g., via S-lemma).

What about for fixed psd rank (even $k = 3$)? Analogue of Arora-Ge-Kannan-Moitra polynomiality result for nonnegative rank?

What about bounds? And before psd, what about nonnegative rank?

Computing and bounding PSD rank

Computing PSD rank seems to be quite hard, both in theory and practice. Are there situations where it is easy?

- $k = 1$ easy, since $\text{rank}_{\text{psd}}(M) = 1$ if and only if $\text{rank}(M) = 1$.
- $k = 2$ also easy, since it is reducible to semidefinite programming (e.g., via S-lemma).

What about for fixed psd rank (even $k = 3$)? Analogue of Arora-Ge-Kannan-Moitra polynomiality result for nonnegative rank?

What about bounds? And before psd, what about nonnegative rank?

Bounding nonnegative rank

Want techniques to *lower bound* the nonnegative rank of a matrix.

In applications, these bounds may yield:

- Minimal size of latent variables
- Complexity lower bounds on extended representations

Many known bounds (e.g. rank, combinatorial, information-theoretic, etc.).

New "self-scaled bounds" via SOS (Fawzi-P., `arXiv:1404.3240`), that extend to other "product cone" ranks (e.g., NN tensor rank, CP-rank, etc). We describe these next...

Bounding nonnegative rank

Want techniques to *lower bound* the nonnegative rank of a matrix.

In applications, these bounds may yield:

- Minimal size of latent variables
- Complexity lower bounds on extended representations

Many known bounds (e.g. rank, combinatorial, information-theoretic, etc.).

New "self-scaled bounds" via SOS (Fawzi-P., `arXiv:1404.3240`), that extend to other "product cone" ranks (e.g., NN tensor rank, CP-rank, etc). We describe these next...

Lower bound

- Main observation: Assume
$$M = X_1 + \cdots + X_r \tag{1}$$
nonnegative factorization of M where $X_i \geq 0$ and rank-one.
Then $X_i \leq M$ (componentwise) for all $i = 1, \ldots, r$.

Lower bound

- Main observation: Assume

$$M = X_1 + \cdots + X_r \tag{1}$$

nonnegative factorization of M where $X_i \geq 0$ and rank-one.
Then $X_i \leq M$ (componentwise) for all $i = 1, \ldots, r$.

- Define

$$\mathcal{A}(M) = \left\{ X \in \mathbb{R}^{p \times q} \ : \ X \text{ rank-one and } 0 \leq X \leq M \right\}$$

Each X_i from Equation (1) belongs to $\mathcal{A}(M)$.

Lower bound

- Main observation: Assume

$$M = X_1 + \cdots + X_r \tag{1}$$

nonnegative factorization of M where $X_i \geq 0$ and rank-one.
Then $X_i \leq M$ (componentwise) for all $i = 1, \ldots, r$.

- Define

$$\mathcal{A}(M) = \left\{ X \in \mathbb{R}^{p \times q} \ : \ X \text{ rank-one and } 0 \leq X \leq M \right\}$$

Each X_i from Equation (1) belongs to $\mathcal{A}(M)$.

Proposition

Assume $L : \mathbb{R}^{p \times q} \to \mathbb{R}$ linear function such that $L(X) \leq 1$ for all $X \in \mathcal{A}(M)$. Then $L(M) \leq \operatorname{rank}_+(M)$.

Proof.

$$L(M) = L(X_1) + \cdots + L(X_r) \leq 1 + \cdots + 1 = r = \operatorname{rank}_+(M).$$

□

Lower bound

- Look for the linear function L which gives the best lower bound (call the resulting quantity $\tau(M)$):

$$\begin{aligned} \tau(M) := \quad & \max_{L} \quad L(M) \\ & \text{s.t.} \quad L : \mathbb{R}^{p \times q} \to \mathbb{R} \text{ linear} \\ & \qquad\;\; L \leq 1 \text{ on } \mathcal{A}(M) \end{aligned}$$

- From previous proposition, $\tau(M)$ satisfies:

$$\tau(M) \leq \text{rank}_+(M)$$

- Computing $\tau(M)$ is a convex optimization problem (but feasible set may be complicated to represent)

Duality

$$\tau(M) := \max_{L \text{ linear}} \quad L(M) \quad \text{s.t.} \quad L \leq 1 \text{ on } \mathcal{A}(M)$$

$$= \quad \min \quad t \quad \text{s.t.} \quad M \in t\,\text{conv}(\mathcal{A}(M))$$

Duality

$$\tau(M) := \max_{L \text{ linear}} \quad L(M) \quad \text{s.t.} \quad L \le 1 \text{ on } \mathcal{A}(M)$$

$$= \min \quad t \quad \text{s.t.} \quad M \in t\,\mathrm{conv}(\mathcal{A}(M))$$

- $\tau(M)$ is Minkowski gauge function of conv($\mathcal{A}(M)$), evaluated at M.
- "*Self-scaled*": the atoms $\mathcal{A}(M)$ *depend* on the matrix M

Example: diagonal matrices

$$M = \begin{bmatrix} a_1 & 0 \\ 0 & a_2 \end{bmatrix}$$

$$\mathcal{A}(M) = \left\{ X \in \mathbb{R}^{2\times 2} \ : \ \text{rank}\, X \leq 1 \text{ and } 0 \leq X \leq \begin{bmatrix} a_1 & 0 \\ 0 & a_2 \end{bmatrix} \right\} \tag{2}$$

Example: diagonal matrices

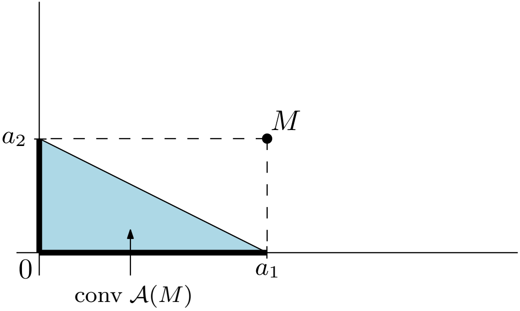
$$M = \begin{bmatrix} a_1 & 0 \\ 0 & a_2 \end{bmatrix}$$

$$\begin{aligned} \mathcal{A}(M) &= \left\{ X \in \mathbb{R}^{2\times 2} \ : \ \text{rank}\, X \leq 1 \text{ and } 0 \leq X \leq \begin{bmatrix} a_1 & 0 \\ 0 & a_2 \end{bmatrix} \right\} \\ &= \left\{ \begin{bmatrix} x & 0 \\ 0 & 0 \end{bmatrix} \text{ with } 0 \leq x \leq a_1 \right\} \cup \left\{ \begin{bmatrix} 0 & 0 \\ 0 & y \end{bmatrix} \text{ with } 0 \leq y \leq a_2 \right\}. \end{aligned} \tag{2}$$

Example: diagonal matrices

$$M = \begin{bmatrix} a_1 & 0 \\ 0 & a_2 \end{bmatrix}$$

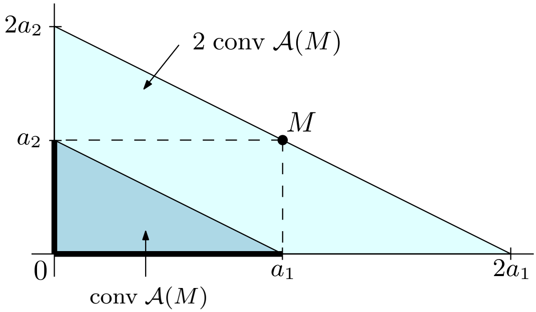
$$\begin{aligned} \mathcal{A}(M) &= \left\{ X \in \mathbb{R}^{2\times 2} \;:\; \text{rank}\, X \leq 1 \text{ and } 0 \leq X \leq \begin{bmatrix} a_1 & 0 \\ 0 & a_2 \end{bmatrix} \right\} \\ &= \left\{ \begin{bmatrix} x & 0 \\ 0 & 0 \end{bmatrix} \text{ with } 0 \leq x \leq a_1 \right\} \cup \left\{ \begin{bmatrix} 0 & 0 \\ 0 & y \end{bmatrix} \text{ with } 0 \leq y \leq a_2 \right\}. \end{aligned} \tag{2}$$



Example: diagonal matrices

$$M = \begin{bmatrix} a_1 & 0 \\ 0 & a_2 \end{bmatrix}$$

$$\begin{aligned} \mathcal{A}(M) &= \left\{ X \in \mathbb{R}^{2\times 2} \ : \ \text{rank}\, X \leq 1 \text{ and } 0 \leq X \leq \begin{bmatrix} a_1 & 0 \\ 0 & a_2 \end{bmatrix} \right\} \\ &= \left\{ \begin{bmatrix} x & 0 \\ 0 & 0 \end{bmatrix} \text{ with } 0 \leq x \leq a_1 \right\} \cup \left\{ \begin{bmatrix} 0 & 0 \\ 0 & y \end{bmatrix} \text{ with } 0 \leq y \leq a_2 \right\}. \end{aligned} \tag{2}$$



Computing the lower bound

$$\tau(M) = \max_{L \text{ linear}} \quad L(M) \quad \text{s.t.} \quad L \leq 1 \text{ on } \mathcal{A}(M)$$

- To compute $\tau(M)$, we need efficient description of

$$C = \Big\{ L \text{ linear} \quad : \quad L(X) \leq 1 \quad \forall X \in \mathcal{A}(M) \Big\}.$$

Computing the lower bound

$$\tau(M) = \max_{L \text{ linear}} \quad L(M) \quad \text{s.t.} \quad L \leq 1 \text{ on } \mathcal{A}(M)$$

- To compute $\tau(M)$, we need efficient description of

$$C = \left\{ L \text{ linear} \ : \ L(X) \leq 1 \ \ \forall X \in \mathcal{A}(M) \right\}.$$

- We construct a tractable SOS relaxation $C^{sos} \subseteq C$:

$$C^{sos} = \left\{ L \text{ linear} \ : \ \text{identity below holds for some } \alpha_{ij} \geq 0, \beta_{ijkl}, SOS(X) \right\}.$$

$$1 - L(X) = SOS(X) + \sum_{\substack{1 \leq i \leq p \\ 1 \leq j \leq q}} \alpha_{ij} X_{ij}(M_{ij} - X_{ij}) + \sum_{\substack{1 \leq i < k \leq p \\ 1 \leq j < l \leq q}} \beta_{ijkl}(X_{ij}X_{kl} - X_{il}X_{kj})$$

Computing the lower bound

$$\tau(M) = \max_{L \text{ linear}} \quad L(M) \quad \text{s.t.} \quad L \leq 1 \text{ on } \mathcal{A}(M)$$

- To compute $\tau(M)$, we need efficient description of

$$C = \Big\{ L \text{ linear} \ : \ L(X) \leq 1 \ \ \forall X \in \mathcal{A}(M) \Big\}.$$

- We construct a tractable SOS relaxation $C^{sos} \subseteq C$:

$$C^{sos} = \Big\{ L \text{ linear} \ : \ \text{identity below holds for some } \alpha_{ij} \geq 0, \beta_{ijkl}, SOS(X) \Big\}.$$

$$1 - L(X) = SOS(X) + \sum_{\substack{1 \leq i \leq p \\ 1 \leq j \leq q}} \alpha_{ij} X_{ij}(M_{ij} - X_{ij}) + \underbrace{\sum_{\substack{1 \leq i < k \leq p \\ 1 \leq j < l \leq q}} \beta_{ijkl}(X_{ij}X_{kl} - X_{il}X_{kj})}_{=0 \text{ if } X \text{ is rank-one}}$$

Computing the lower bound

$$\tau(M) = \max_{L \text{ linear}} \quad L(M) \quad \text{s.t.} \quad L \leq 1 \text{ on } \mathcal{A}(M)$$

- To compute $\tau(M)$, we need efficient description of

$$C = \Big\{ L \text{ linear} \; : \; L(X) \leq 1 \;\; \forall X \in \mathcal{A}(M) \Big\}.$$

- We construct a tractable SOS relaxation $C^{sos} \subseteq C$:

$$C^{sos} = \Big\{ L \text{ linear} \; : \; \text{identity below holds for some } \alpha_{ij} \geq 0, \beta_{ijkl}, SOS(X) \Big\}.$$

$$1 - L(X) = SOS(X) + \underbrace{\sum_{\substack{1 \leq i \leq p \\ 1 \leq j \leq q}} \alpha_{ij} X_{ij} (M_{ij} - X_{ij})}_{\geq 0 \text{ if } 0 \leq X \leq M} + \underbrace{\sum_{\substack{1 \leq i < k \leq p \\ 1 \leq j < l \leq q}} \beta_{ijkl} (X_{ij} X_{kl} - X_{il} X_{kj})}_{= 0 \text{ if } X \text{ is rank-one}}$$

Computing the lower bound

$$\tau(M) = \max_{L \text{ linear}} \quad L(M) \quad \text{s.t.} \quad L \leq 1 \text{ on } \mathcal{A}(M)$$

- To compute $\tau(M)$, we need efficient description of

$$C = \Big\{ L \text{ linear} \;\; : \;\; L(X) \leq 1 \quad \forall X \in \mathcal{A}(M) \Big\}.$$

- We construct a tractable SOS relaxation $C^{sos} \subseteq C$:

$$C^{sos} = \Big\{ L \text{ linear} \;\; : \;\; \text{identity below holds for some } \alpha_{ij} \geq 0, \beta_{ijkl}, SOS(X) \Big\}.$$

$$1 - L(X) = \underbrace{SOS(X)}_{\geq 0} + \underbrace{\sum_{\substack{1 \leq i \leq p \\ 1 \leq j \leq q}} \alpha_{ij} X_{ij}(M_{ij} - X_{ij})}_{\geq 0 \text{ if } 0 \leq X \leq M} + \underbrace{\sum_{\substack{1 \leq i < k \leq p \\ 1 \leq j < l \leq q}} \beta_{ijkl}(X_{ij}X_{kl} - X_{il}X_{kj})}_{= 0 \text{ if } X \text{ is rank-one}}$$

SOS relaxation

- Define:

$$\tau^{\text{sos}}(M) := \max_{L} \quad L(M)$$
$$\text{s.t.} \quad L \text{ linear and has SOS representation}$$

- Quantity $\tau^{\text{sos}}(M)$ can be computed using semidefinite programming. Satisfies:

$$\tau^{\text{sos}}(M) \leq \tau(M) \leq \text{rank}_+(M)$$

Structural properties of τ and τ^{sos}

- Invariant under scaling:

$$\tau(D_1 M D_2) = \tau(M)$$

  for any D_1, D_2 diagonal matrices with positive diagonal elements.

- Block-diagonal matrices: $\tau(\text{blockdiag}(M_1, M_2)) = \tau(M_1) + \tau(M_2)$
- Subadditivity: $\tau(M + N) \leq \tau(M) + \tau(N)$
- Product: $\tau(MN) \leq \min(\tau(M), \tau(N))$
- Monotonicity: If P submatrix of M then $\tau(P) \leq \tau(M)$.

Comparison with combinatorial bounds

- Combinatorial bounds on $\text{rank}_+(M)$ are bounds that only depend on sparsity pattern of M. Can be expressed in terms of the *rectangle graph* G_M of M:

$$\underbrace{\omega(G_M)}_{\text{fooling set bound}} \leq \overline{\vartheta}(G_M) \leq \chi_{\text{frac}}(G_M) \leq \underbrace{\chi(G_M)}_{\text{rect. cover number}} \leq \text{rank}_+(M)$$

- The quantities $\tau(M)$ and $\tau^{sos}(M)$ can be shown to be *non-combinatorial* counterparts of *fractional rectangle cover number* and of $\overline{\vartheta}(G_M)$:

Theorem

$$\tau(M) \geq \chi_{frac}(G_M) \qquad \tau^{sos}(M) \geq \overline{\vartheta}(G_M)$$

General "atomic cone ranks"

- General framework: K convex cone and V is some set

$$M = \sum_{i=1}^{r} X_i \quad \text{where} \quad X_i \in K \cap V$$

Define $\text{rank}_{K,V}(M)$ to be the size of the smallest such decomposition.

$$\mathcal{A}(M) = \{X \ : \ 0 \preceq_K X \preceq_K M\}$$

$$\tau(M) = \max L(M) \ : \ L \leq 1 \text{ on } \mathcal{A}(M).$$

Then $\tau(M) \leq \text{rank}_{K,V}(M)$.

General “atomic cone ranks”

- General framework: K convex cone and V is some set

$$M = \sum_{i=1}^{r} X_i \quad \text{where} \quad X_i \in K \cap V$$

Define $\text{rank}_{K,V}(M)$ to be the size of the smallest such decomposition.

$$\mathcal{A}(M) = \{X \ : \ 0 \preceq_K X \preceq_K M\}$$

$$\tau(M) = \max L(M) \ : \ L \leq 1 \text{ on } \mathcal{A}(M).$$

Then $\tau(M) \leq \text{rank}_{K,V}(M)$.

- Examples:
    - Completely positive matrices: $M = \sum_{i=1}^{r} u_i u_i^T \quad \text{where} \quad u_i \geq 0.$
    - Quadrature formulae:

$$\int_{\Omega} p(x)dx = \sum_{i=1}^{r} w_i p(x_i) \quad w_i \geq 0.$$

Lower bounding PSD rank?

Bounds on PSD rank are of high interest, since combinatorial methods (based on sparsity patterns) don't quite work.

But, a few unexpected difficulties...

- In the PSD case, the underlying norm is non-atomic, and the corresponding "obvious" inequalities do not hold...
- "Noncommutative" trace positivity, quite complicated structure...

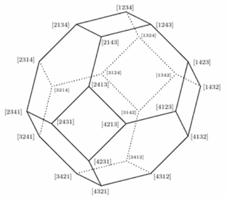
Nice links between rank_{psd} and quantum communication complexity, mirroring the situation between rank_+ and classical communication complexity (e.g., Fiorini *et al.* (2011), Jain *et al.* (2011), Zhang (2012)).

Orbitopes and equivariant lifts

Special class of convex bodies: regular orbitopes

$$C = \{\mathrm{conv}(g \cdot x_0) \;:\; g \in G\},$$

where G is a compact group.



Many important examples: hypercubes, hyperspheres, Grassmannians, Birkhoff polytope, parity polytope, cut polytope, etc..

(More about this in tomorrow's talk.)

Symmetric PSD factorizations

Given a symmetric $M \in \mathbb{R}^{n \times n}$, do there exist $A_i \succeq 0$ such that

$$M_{ij} = \langle A_i, A_j \rangle \qquad i, j = 1, \ldots, n.$$

Equivalently, is M the Gram matrix of a set of psd matrices?

- Dual to *trace positivity* of noncommutative polynomials (e.g., Klep, Burgdorf, etc.)
- Of interest in quantum information (e.g., Piovesan-Laurent)
- Many open questions; related to outstanding conjectures of Connes and Tsirelson

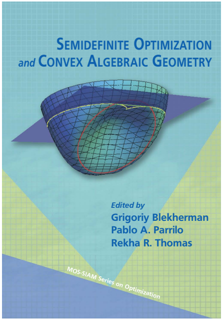
Many questions

Many open aspects of positive semidefinite rank:

- Computation of lower/upper bounds?
- Approximate factorizations?
- Topology of space of factorizations?
- For polytopes, separations between rank_+ and rank_{psd} for slack matrices?
- Are current constructions (e.g., SOS) far from optimal?

Summary

- Interesting, new class of factorization problems
- Interplay of algebraic and geometric aspects
- Many open questions, lots to do!



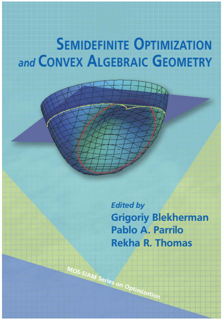
If you want to know more:

- H. Fawzi, J. Gouveia, P.A. Parrilo, R. Robinson, R. Thomas, Positive semidefinite rank, arXiv:1407.4095.
- J. Gouveia, P.A. Parrilo, R. Thomas, Lifts of convex sets and cone factorizations, *Mathematics of Operations Research*, 38:2, 2013. arXiv:1111.3164.
- SIAM book and NSF FRG project "SDP and convex algebraic geometry" website www.math.washington.edu/~thomas/frg/frg.html (Helton/P./Nie/Sturmfels/Thomas)

Thanks for your attention!

Summary

- Interesting, new class of factorization problems
- Interplay of algebraic and geometric aspects
- Many open questions, lots to do!



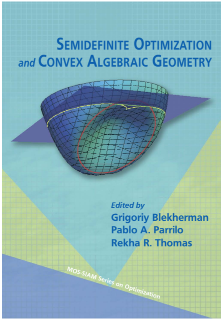
If you want to know more:

- H. Fawzi, J. Gouveia, P.A. Parrilo, R. Robinson, R. Thomas, Positive semidefinite rank, arXiv:1407.4095.
- J. Gouveia, P.A. Parrilo, R. Thomas, Lifts of convex sets and cone factorizations, *Mathematics of Operations Research*, 38:2, 2013. arXiv:1111.3164.
- SIAM book and NSF FRG project "SDP and convex algebraic geometry" website www.math.washington.edu/~thomas/frg/frg.html (Helton/P./Nie/Sturmfels/Thomas)

Thanks for your attention!

Summary

- Interesting, new class of factorization problems
- Interplay of algebraic and geometric aspects
- Many open questions, lots to do!

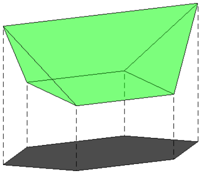


If you want to know more:

- H. Fawzi, J. Gouveia, P.A. Parrilo, R. Robinson, R. Thomas, Positive semidefinite rank, `arXiv:1407.4095`.
- J. Gouveia, P.A. Parrilo, R. Thomas, Lifts of convex sets and cone factorizations, *Mathematics of Operations Research*, 38:2, 2013. `arXiv:1111.3164`.
- SIAM book and NSF FRG project "SDP and convex algebraic geometry" website `www.math.washington.edu/~thomas/frg/frg.html` (Helton/P./Nie/Sturmfels/Thomas)

Thanks for your attention!

Example: hexagon (III)



A nonnegative factorization:

$$S_H = \begin{pmatrix} 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 2 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 2 & 1 & 0 \end{pmatrix} \begin{pmatrix} 0 & 0 & 0 & 1 & 2 & 1 \\ 1 & 2 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$