Self-scaled bounds for atomic ranks: applications to nonnegative rank and cp-rank

Hamza Fawzi
Joint work with Pablo Parrilo

Laboratory for Information and Decision Systems
Massachusetts Institute of Technology

May 21st 2014
SIAM Optimization Conference 2014
San Diego

Nonnegative rank

- $M \in \mathbb{R}^{p \times q}$ *nonnegative* matrix: $M_{ij} \geq 0 \;\; \forall i, j$

- Nonnegative factorization of M:

$$\underbrace{\boxed{\begin{matrix} M \\ (p \times q) \end{matrix}}}_{\geq 0} = \underbrace{\boxed{\begin{matrix} U \\ (p \times r) \end{matrix}}}_{\geq 0} \underbrace{\boxed{\begin{matrix} V \\ (r \times q) \end{matrix}}}_{\geq 0} = u_1 v_1^T + \ldots + u_r v_r^T$$

- Nonnegative rank of M, denoted $\operatorname{rank}_+(M)$, is smallest r such that M has a nonnegative decomposition with r terms.

- Quantity of interest in different areas (optimization, probability, etc.)

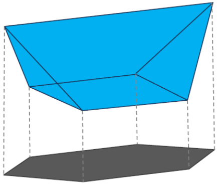
  This talk: new method to compute lower bounds on $\operatorname{rank}_+(M)$ using convex optimization

Applications of nonnegative rank

- **Statistical modeling:** $M_{ij} = \Pr[X = i, Y = j]$

$$\begin{array}{c}\text{Nonnegative}\\ \text{factorization of } M\end{array} \iff \begin{array}{c}\text{Finding “hidden” variable}\\ W \text{ such that } \underbrace{X - W - Y}_{\substack{\text{conditional}\\ \text{independence}}}\end{array}$$

- **Optimization:** extended formulations of polytopes



Yannakakis theorem (1991): $\mathrm{rank}_+(S_P) =$ the smallest number of linear inequalities needed to represent P

- Other applications in communication complexity, etc...
Unfortunately, nonnegative rank is **NP-hard** [Vav09] to compute (unlike standard rank).

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.

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

$\square$

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

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

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

Duality

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

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

- $\tau(M)$ is Minkowski gauge function of $\mathrm{conv}(\mathcal{A}(M))$, 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

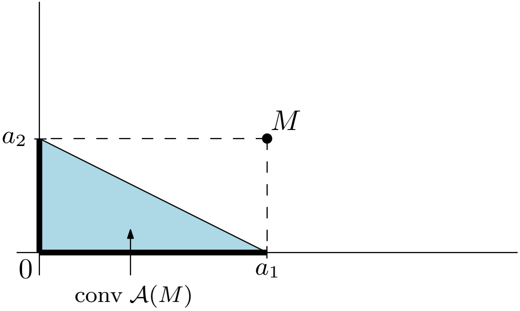
$$M = \begin{bmatrix} a_1 & 0 \\ 0 & a_2 \end{bmatrix}$$

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

Example: diagonal matrices

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

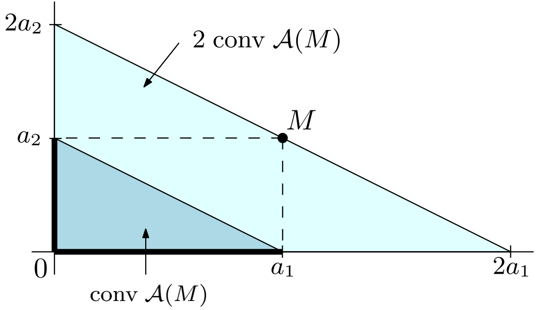
$$\begin{aligned} \mathcal{A}(M) &= \left\{ X \in \mathbb{R}^{2\times 2} \; : \; \text{rank}\, X \leq 1 \text{ and } 0 \leq X \leq \begin{bmatrix} a_1 & 0 \\ 0 & a_2 \end{bmatrix} \right\} \\ &= \left\{ \begin{bmatrix} x & 0 \\ 0 & 0 \end{bmatrix} \text{ with } 0 \leq x \leq a_1 \right\} \cup \left\{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



Example: diagonal matrices

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

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



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

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

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

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

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

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

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

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

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

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

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 $C^{sos} \subseteq C$:

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

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

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

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

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

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

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

SOS relaxation

- Define:

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

- 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

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

Comparison with "norm-based" bounds

One can show that $\tau(M)$ is at least as good as any norm-based bounds:

Theorem

Let $\mathcal{N} : \mathbb{R}_+^{m \times n} \to \mathbb{R}_+$ be any monotone positively homogeneous function. Let L be a linear function such that:

$$L(X) \leq 1 \quad \forall X \geq 0,\ \textit{rank-one}, \mathcal{N}(X) = 1.$$

Then for any $M \in \mathbb{R}_+^{m \times n}$,

$$\frac{L(M)}{\mathcal{N}(M)} \leq \tau(M) \leq \text{rank}_+(M).$$

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

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

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

Conclusion

- A new lower bound on $\text{rank}_+(M)$ using convex optimization and sum-of-squares techniques
- Improves on existing combinatorial and norm-based bounds and has appealing structural properties
- Technique applies to other "atomic rank functions" defined on convex cones.

Conclusion

- A new lower bound on $\text{rank}_+(M)$ using convex optimization and sum-of-squares techniques
- Improves on existing combinatorial and norm-based bounds and has appealing structural properties
- Technique applies to other "atomic rank functions" defined on convex cones.

Thank you!

`http://arxiv.org/abs/1404.3240`