New Lower Bounds on Nonnegative Rank using Conic Programming

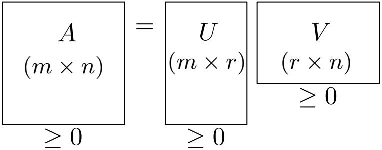
Hamza Fawzi
Joint work with Pablo Parrilo

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

ICCOPT, Lisbon, Portugal, July 2013

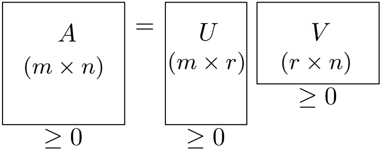
Nonnegative rank

- $A \in \mathbb{R}^{m \times n}_{\geq 0}$ *elementwise* nonnegative matrix. Nonnegative factorization:



Nonnegative rank

- $A \in \mathbb{R}^{m \times n}_{\geq 0}$ *elementwise* nonnegative matrix. Nonnegative factorization:



- Nonnegative rank of A is smallest r such that A has a nonnegative factorization of inner dimension r (denoted $\text{rank}_+(A)$).

$$\text{rank}_+(A) = \min\{r \in \mathbb{N} \ : \ \exists U \in \mathbb{R}^{m \times r}_{\geq 0}, V \in \mathbb{R}^{r \times n}_{\geq 0} \ \ A = UV\}$$

(Note: if we drop requirement $U, V \geq 0$, we get the usual rank)

Extended formulations of polytopes (1)

- Linear program:

$$(LP) \quad \min \; c^T x \; \text{ subject to } Gx \leq h$$

- Feasible set is polytope $P = \{x \mid Gx \leq h\}$.
  Number of inequalities (num. of rows of G) is number of facets of P.

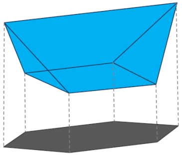
Extended formulations of polytopes (1)

- Linear program:

$$(LP) \quad \min \; c^T x \; \text{ subject to } Gx \leq h$$

- Feasible set is polytope $P = \{x \mid Gx \leq h\}$.
  Number of inequalities (num. of rows of G) is number of facets of P.

- Given polytope P, an extended formulation of P is another polytope Q (in higher-dimensional space) that projects onto Q.



Extended formulation of a 2D hexagon

- Sometimes Q can be much simpler to represent than P (has much fewer facets)
- Extension complexity of P = smallest f s.t. P has an extended formulation that has f facets

Extended formulations of polytopes (2)

- How does this relate to nonnegative rank?
- Yannakakis '91: Extension complexity of P is equal to

$$\mathrm{rank}_+(S(P))$$

  where $S(P)$ is *slack matrix* of P.
- Proof of Yannakakis' theorem is constructive: any nonnegative factorization of $S(P)$ yields an extended formulation, and vice-versa.

Extended formulations of polytopes (2)

- How does this relate to nonnegative rank?
- Yannakakis '91: Extension complexity of P is equal to

$$\text{rank}_+(S(P))$$

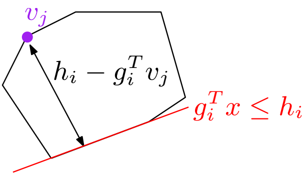
  where $S(P)$ is *slack matrix* of P.

- Proof of Yannakakis' theorem is constructive: any nonnegative factorization of $S(P)$ yields an extended formulation, and vice-versa.
- $S(P)$ is a nonnegative matrix of size #facets(P) $\times$ #vertices(P) defined by:

$$S(P)_{i,j} = h_i - g_i^T v_j$$

  where

  - $g_i^T x \leq h_i$ are the facet inequalities of P
  - v_j are the vertices of P



Other applications...

- Other applications of nonnegative rank in latent variable modeling/correlation generation as well as in communication complexity.
- Nonnegative matrix factorization is used in practice in different domains:
  - topic modeling (identifying a set of topics in documents)
  - hyperspectral unmixing
  - etc...
- Unlike the usual rank, nonnegative rank is hard to compute (Vavasis 2009, Arora et al. 2012)
- Objective: Use convex optimization techniques to obtain lower bound on nonnegative rank

Existing combinatorial lower bounds

- Nonnegative factorization:

$$A = \underbrace{u_1 v_1^T}_{\geq 0} + \cdots + \underbrace{u_r v_r^T}_{\geq 0}$$

- Nonzero entries of $u_k v_k^T$ define a rectangle: $R_k = \text{supp}(u_k) \times \text{supp}(v_k)$.
- Rectangles R_k cover the nonzero entries of A without “touching” the zero entries:

$$\text{supp}(A) \quad = \quad R_1 \cup R_2 \cup \cdots \cup R_k$$

- Hence we have:

$$\text{rank}_+(A) \geq rc(A)$$

where $rc(A)$ is minimal number of rectangles needed to cover supp(A) (rectangle covering number of A).

$$\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \cup \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$$

Nuclear norm

- $A \in \mathbb{R}^{m \times n}$ arbitrary matrix. The **nuclear norm** of A is the sum of the singular values of A:

$$\nu(A) = \sum_{i=1}^{\mathrm{rank}(A)} \sigma_i(A)$$

where $\sigma_i(A)$ singular values of A.

Nuclear norm

- $A \in \mathbb{R}^{m \times n}$ arbitrary matrix. The **nuclear norm** of A is the sum of the singular values of A:

$$\nu(A) = \sum_{i=1}^{\text{rank}(A)} \sigma_i(A)$$

  where $\sigma_i(A)$ singular values of A.

- Nuclear norm gives lower bound on rank(A):

$$\text{rank}(A) \geq \left(\frac{\nu(A)}{\|A\|_F} \right)^2$$

  where

$$\|A\|_F = \sqrt{\sum_{i,j} A_{i,j}^2} = \|\sigma\|_2$$

- Proof:

$$\nu(A) = \sum_{i=1}^{\text{rank}(A)} \sigma_i(A) \leq \sqrt{\text{rank}(A)}\|\sigma\|_2 = \sqrt{\text{rank}(A)}\|A\|_F$$

Nuclear norm

- Alternative definition of nuclear norm (without using singular values):

$$\nu(A) = \min \left\{ \sum_i \|A_i\|_F \ : \ A = \sum_i A_i \text{ where } A_i \text{ rank } 1 \right\}$$

Atomic norm where the *atoms* are rank-1 matrices with unit Frobenius norm.

Nuclear norm

- Alternative definition of nuclear norm (without using singular values):

$$\nu(A) = \min \left\{ \sum_i \|A_i\|_F \; : \; A = \sum_i A_i \text{ where } A_i \text{ rank } 1 \right\}$$

  Atomic norm where the *atoms* are rank-1 matrices with unit Frobenius norm.

- $\nu(A)$ is a semidefinite program:

$$\nu(A) = \min_{X,Y} \left\{ \frac{1}{2}(\mathrm{Tr}(X) + \mathrm{Tr}(Y)) \; : \; \begin{bmatrix} X & A \\ A^T & Y \end{bmatrix} \succeq 0 \right\}$$

“Nonnegative” nuclear norm

- For nonnegative factorizations, natural to define:

$$\nu_{+}(A) = \min \left\{ \sum_i \|A_i\|_F \;:\; A = \sum_i A_i \quad A_i \text{ rank } 1 \;\; \text{and } A_i \geq 0 \right\}$$

  Atomic norm where the *atoms* are nonnegative rank-1 matrices with unit Frobenius norm.

- Can we compute $\nu_{+}(A)$? Does $\nu_{+}(A)$ give a lower bound to $\text{rank}_{+}(A)$?

“Nonnegative” nuclear norm

- For nonnegative factorizations, natural to define:

$$\nu_+(A) = \min \left\{ \sum_i \|A_i\|_F \ : \ A = \sum_i A_i \quad A_i \text{ rank } 1 \ \text{ and } A_i \geq 0 \right\}$$

  Atomic norm where the *atoms* are nonnegative rank-1 matrices with unit Frobenius norm.

- Can we compute $\nu_+(A)$? Does $\nu_+(A)$ give a lower bound to $\text{rank}_+(A)$?

$$\nu_+(A) = \min_{X,Y} \frac{1}{2}(\text{Tr}(X) + \text{Tr}(Y))$$

$$\text{s.t. } \begin{bmatrix} X & A \\ A^T & Y \end{bmatrix} \text{ completely positive}$$

Lower bound on nonnegative rank

- *Does* $\nu_+(A)$ *give a lower bound to* $\text{rank}_+(A)$*?* Yes:

Theorem:

$$\text{rank}_+(A) \geq \left(\frac{\nu_+(A)}{\|A\|_F} \right)^2$$

Lower bound on nonnegative rank

- *Does* $\nu_+(A)$ *give a lower bound to* $\text{rank}_+(A)$*?* Yes:

  Theorem:

  $$\text{rank}_+(A) \geq \left(\frac{\nu_+(A)}{\|A\|_F}\right)^2$$

- Proof: Let $A = \sum_{i=1}^r u_i v_i^T$ nonnegative decomposition of A with $r = \text{rank}_+(A)$.

  Cauchy-Schwarz:

  $$\frac{\sum_{i=1}^r \|u_i\|_2 \|v_i\|_2}{\sqrt{\sum_{i=1}^r \|u_i\|_2^2 \|v_i\|_2^2}} \leq \sqrt{r} = \sqrt{\text{rank}_+(A)}$$

Lower bound on nonnegative rank

- *Does* $\nu_+(A)$ *give a lower bound to* $\text{rank}_+(A)$*?* Yes:

  Theorem:

  $$\text{rank}_+(A) \geq \left(\frac{\nu_+(A)}{\|A\|_F}\right)^2$$

- Proof: Let $A = \sum_{i=1}^{r} u_i v_i^T$ nonnegative decomposition of A with $r = \text{rank}_+(A)$.

  Cauchy-Schwarz:

  $$\frac{\sum_{i=1}^{r} \|u_i\|_2 \|v_i\|_2}{\sqrt{\sum_{i=1}^{r} \|u_i\|_2^2 \|v_i\|_2^2}} \leq \sqrt{r} = \sqrt{\text{rank}_+(A)}$$

  One then shows

  - $\nu_+(A) \leq \sum_{i=1}^{r} \|u_i\|_2 \|v_i\|_2$, by definition of $\nu_+(A)$
  - $\|A\|_F \geq \left(\sum_{i=1}^{r} \|u_i\|_2^2 \|v_i\|_2^2\right)^{1/2}$, using nonnegativity of u_i, v_i

Computing $\nu_+(A)$

$$\nu_+(A) = \min_{X,Y} \frac{1}{2}(\mathrm{Tr}(X) + \mathrm{Tr}(Y))$$

$$\text{s.t.} \begin{bmatrix} X & A \\ A^T & Y \end{bmatrix} \text{ completely positive}$$

- Matrix $M \in \mathbb{R}^{n \times n}$ is completely positive if $M = BB^T$ for some B nonnegative

Computing $\nu_+(A)$

$$\nu_+(A) = \min_{X,Y} \frac{1}{2}(\mathrm{Tr}(X) + \mathrm{Tr}(Y))$$

$$\text{s.t.} \begin{bmatrix} X & A \\ A^T & Y \end{bmatrix} \text{ completely positive}$$

- Matrix $M \in \mathbb{R}^{n \times n}$ is completely positive if $M = BB^T$ for some B nonnegative
- Closed convex cone $\subseteq \mathbf{S}_+^n \cap \mathbb{R}_{\geq 0}^{n \times n}$. Equality holds for $n \leq 4$. For $n > 4$ inclusion is strict.

Computing $\nu_+(A)$

$$\nu_+(A) = \min_{X,Y} \frac{1}{2}(\mathrm{Tr}(X) + \mathrm{Tr}(Y))$$
$$\text{s.t. } \begin{bmatrix} X & A \\ A^T & Y \end{bmatrix} \text{ completely positive}$$

- Matrix $M \in \mathbb{R}^{n \times n}$ is completely positive if $M = BB^T$ for some B nonnegative
- Closed convex cone $\subseteq \mathbf{S}_+^n \cap \mathbb{R}_{\geq 0}^{n \times n}$. Equality holds for $n \leq 4$. For $n > 4$ inclusion is strict.
- Dual is the cone of copositive matrices

$$M \text{ copositive} \quad \overset{\text{def}}{\iff} \quad \forall z \geq 0, \quad z^T M z \geq 0$$

Computing $\nu_+(A)$

Duality

$$\nu_+(A) = \min_{X,Y} \quad \frac{1}{2}(\mathrm{Tr}(X) + \mathrm{Tr}(Y))$$

$$\text{s.t.} \quad \begin{bmatrix} X & A \\ A^T & Y \end{bmatrix} \text{ completely positive}$$

Computing $\nu_+(A)$

Duality

$$\nu_+(A) = \min_{X,Y} \quad \frac{1}{2}(\mathrm{Tr}(X) + \mathrm{Tr}(Y))$$

$$\text{s.t.} \quad \begin{bmatrix} X & A \\ A^T & Y \end{bmatrix} \text{ completely positive}$$

$$= \max_{W} \quad \mathrm{Tr}(A^T W)$$

$$\text{s.t.} \quad \begin{bmatrix} I & -W \\ -W^T & I \end{bmatrix} \text{ copositive}$$

Computing $\nu_+(A)$

Duality

$$\nu_+(A) = \min_{X,Y} \quad \frac{1}{2}(\mathrm{Tr}(X) + \mathrm{Tr}(Y))$$
$$\text{s.t.} \quad \begin{bmatrix} X & A \\ A^T & Y \end{bmatrix} \text{ completely positive}$$

$$= \max_{W} \quad \mathrm{Tr}(A^T W)$$
$$\text{s.t.} \quad \begin{bmatrix} I & -W \\ -W^T & I \end{bmatrix} \text{ copositive}$$

Linear programs over completely-positive/copositive cones are NP-hard in general. Fortunately there are nice SDP relaxations...

SDP relaxations of the copositive cone

$$M \text{ copositive} \quad \stackrel{\text{def}}{\iff} \quad \forall z \geq 0, \quad z^T M z \geq 0$$

$$\iff \forall x, \quad p_M(x) := \sum_{i,j} M_{ij} x_i^2 x_j^2 \geq 0$$

SDP relaxations of the copositive cone

$$M \text{ copositive} \quad \overset{\text{def}}{\iff} \quad \forall z \geq 0, \quad z^T M z \geq 0$$

$$\iff \forall x, \quad p_M(x) := \sum_{i,j} M_{ij} x_i^2 x_j^2 \geq 0$$

Sums-of-squares relaxation [Parrilo'2000]

$$\mathcal{C}^{[k]} = \left\{ M \in \mathcal{S}^n \ : \ \left(\sum_{i=1}^{n} x_i^2 \right)^k \left(\sum_{i,j=1}^{n} M_{i,j} x_i^2 x_j^2 \right) \text{ is SOS} \right\}.$$

SDP relaxations of the copositive cone

$$M \text{ copositive} \quad \overset{\text{def}}{\Longleftrightarrow} \quad \forall z \geq 0, \quad z^T M z \geq 0$$

$$\Longleftrightarrow \forall x, \quad p_M(x) := \sum_{i,j} M_{ij} x_i^2 x_j^2 \geq 0$$

Sums-of-squares relaxation [Parrilo'2000]

$$\mathcal{C}^{[k]} = \left\{ M \in \mathcal{S}^n \ : \ \left(\sum_{i=1}^{n} x_i^2 \right)^k \left(\sum_{i,j=1}^{n} M_{i,j} x_i^2 x_j^2 \right) \text{ is SOS} \right\}.$$

$\mathcal{C}^{[k]}$ can be described using Linear Matrix Inequalities (semidefinite programming)

SDP relaxations of the copositive cone

$$M \text{ copositive} \overset{\text{def}}{\iff} \forall z \geq 0, \quad z^T M z \geq 0$$
$$\iff \forall x, \quad p_M(x) := \sum_{i,j} M_{ij} x_i^2 x_j^2 \geq 0$$

Sums-of-squares relaxation [Parrilo'2000]

$$\mathcal{C}^{[k]} = \left\{ M \in \mathcal{S}^n \ : \ \left(\sum_{i=1}^n x_i^2 \right)^k \left(\sum_{i,j=1}^n M_{i,j} x_i^2 x_j^2 \right) \text{ is SOS} \right\}.$$

$\mathcal{C}^{[k]}$ can be described using Linear Matrix Inequalities (semidefinite programming)

$$\mathcal{C}^{[0]} \subseteq \mathcal{C}^{[1]} \cdots \subseteq \mathcal{C}$$

SDP-based lower bounds on nonnegative rank

Define

$$\nu_+^{[k]}(A) = \max_{W} \quad \mathrm{Tr}(A^T W) \quad \text{s.t.} \begin{bmatrix} I & -W \\ -W^T & I \end{bmatrix} \in \mathcal{C}^{[k]}$$

Then

$$\nu(A) \;\le\; \nu_+^{[0]}(A) \;\le\; \nu_+^{[k]}(A) \;\le\; \nu_+(A) \;\le\; \sqrt{\mathrm{rank}_+(A)}\|A\|_F$$

Examples

- A 4×4 matrix:

$$A = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix} \qquad \text{rank}(A) = 3 \quad \text{rank}_+(A) = 4$$

Examples

- A 4×4 matrix:

$$A = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix} \qquad \text{rank}(A) = 3 \quad \text{rank}_+(A) = 4$$

$$\nu_+^{[0]}(A) = 4\sqrt{2} \Longrightarrow \left(\frac{\nu_+^{[0]}(A)}{\|A\|_F} \right)^2 = 4 = \text{rank}_+(A)$$

Examples

- A 4×4 matrix:

$$A = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix} \qquad \mathrm{rank}(A) = 3 \quad \mathrm{rank}_+(A) = 4$$

$$\nu_+^{[0]}(A) = 4\sqrt{2} \Longrightarrow \left(\frac{\nu_+^{[0]}(A)}{\|A\|_F} \right)^2 = 4 = \mathrm{rank}_+(A)$$

- Let $C_n = [0,1]^n$ be the hypercube in n dimensions and let $S(C_n) \in \mathbb{R}^{2n \times 2^n}$ be its slack matrix. Then

$$\mathrm{rank}_+(S(C_n)) = \left(\frac{\nu_+^{[0]}(S(C_n))}{\|S(C_n)\|_F} \right)^2 = 2n.$$

Examples

- Derangement matrix

$$D_n = J_n - I_n = \begin{bmatrix} 0 & 1 & \dots & 1 \\ 1 & 0 & & \vdots \\ \vdots & & \ddots & 1 \\ 1 & \dots & 1 & 0 \end{bmatrix}$$

- $\text{rank}_+(D_n) = \text{rank}(D_n) = n$
- However we can show that for all n,

$$\left(\frac{\nu_+(D_n)}{\|D_n\|_F} \right)^2 \leq 4$$

- D_n is "badly-conditioned": $\sigma_1(D_n) = n - 1, \sigma_2(D_n) = \cdots = \sigma_n(D_n) = 1$.

Summary

- $\nu_+(A)$ natural extension of nuclear norm $\nu(A)$ when dealing with nonnegative factorizations
- $\nu_+(A)$ allows to give lower bound on $\text{rank}_+(A)$.
- $\nu_+(A)$ can be approximated (from below) efficiently using semidefinite programming.

Summary

- $\nu_+(A)$ natural extension of nuclear norm $\nu(A)$ when dealing with nonnegative factorizations
- $\nu_+(A)$ allows to give lower bound on $\text{rank}_+(A)$.
- $\nu_+(A)$ can be approximated (from below) efficiently using semidefinite programming.
- Advantage over combinatorial lower bounds: applies to any nonnegative matrix (does not rely on presence of zeros).
- Value of bound depends on scaling of matrix. Can be poor when A is not well-conditioned.
- Current work in progress: new lower bound based on same atomic norm ideas but invariant under scaling

Summary

- $\nu_+(A)$ natural extension of nuclear norm $\nu(A)$ when dealing with nonnegative factorizations
- $\nu_+(A)$ allows to give lower bound on $\text{rank}_+(A)$.
- $\nu_+(A)$ can be approximated (from below) efficiently using semidefinite programming.
- Advantage over combinatorial lower bounds: applies to any nonnegative matrix (does not rely on presence of zeros).
- Value of bound depends on scaling of matrix. Can be poor when A is not well-conditioned.
- Current work in progress: new lower bound based on same atomic norm ideas but invariant under scaling

Thank you!

Manuscript: `http://arxiv.org/abs/1210.6970`

Latent variable modeling

- Let (X, Y) pair of random variables, and $P(x, y)$ their joint distribution. P is a nonnegative matrix with $\sum_{x,y} P(x, y) = 1$.
- Assume X and Y conditionally independent given some latent variable W. Then:

$$
\begin{aligned}
P(x, y) &= \sum_{w=1}^{|W|} \Pr[X = x, Y = y, W = w] \\
&= \sum_{w=1}^{|W|} \Pr[W = w]\ \Pr[X = x | W = w] \Pr[Y = y | W = w]
\end{aligned}
$$

Latent variable modeling

- Let (X, Y) pair of random variables, and $P(x, y)$ their joint distribution. P is a nonnegative matrix with $\sum_{x,y} P(x, y) = 1$.
- Assume X and Y conditionally independent given some latent variable W. Then:

$$\begin{aligned} P(x, y) &= \sum_{w=1}^{|W|} \Pr[X = x, Y = y, W = w] \\ &= \sum_{w=1}^{|W|} \Pr[W = w] \; \Pr[X = x | W = w] \Pr[Y = y | W = w] \end{aligned}$$

- This is a nonnegative factorization of P! Inner dimension of factorization is $|\mathrm{support}(W)|$.
- "Probabilistic" formulation of nonnegative rank:

$$\mathrm{rank}_+(P) = \min_{X - W - Y} |\mathrm{support}(W)|$$

  where $X - W - Y$ means X and Y are conditionally independent given W.
- Measure of correlation of X and Y. Related to Wyner's common information $C(X; Y)$