Nonnegative Rank Factorization

A Heuristic Approach via Rank Reduction

(Work in Progress)

Moody T. Chu
(Joint work with Bo Dong and Matthew Lin)

North Carolina State University

March 8, 2012 @ National Cheng Kung University

一曲新詞酒一杯,去年天氣舊亭臺,
夕陽西下幾時回.
無可奈何花落去,似曾相識燕歸來,
小園香徑獨徘徊.

--- 宋.宴殊.浣溪沙

Take Home Question

Given a nonnegative matrix A, write

$$A = \sum_{i=1}^{k} \mathbf{u}_i \mathbf{k}_i^\top$$

where $\mathbf{u}_i, \mathbf{k}_i \geq 0$ and k is minimal.

Outline

Outline

Outline

Outline

Outline

Outline

Nonnegative Rank

- Given $A \in \mathbb{R}_+^{m \times n}$, write
$$A = UV.$$
  - $U \in \mathbb{R}_+^{m \times k}$ and $V \in \mathbb{R}_+^{k \times n}$ with $k \leq \min\{m, n\}$
  - Always possible.
- Interested in the *smallest* k rendering this factorization.
  - Denote by $\mathrm{rank}_+(A)$.
- A trivial fact —
$$\mathrm{rank}(A) \leq \mathrm{rank}_+(A) \leq \min\{m, n\}.$$
- A challenge —
  - Determining the exact nonnegative rank is NP-hard.

This Is Not ...

- $A = UV$ is a special nonnegative factorization of A.
- Should be unequivocally distinquished from what is known as the nonnegative matrix factorization (NMF).
  - Misnamed!
  - Is only a low rank approximation from

$$\min_{U \in \mathbb{R}_+^{m \times p}, V \in \mathbb{R}_+^{p \times n}} \|A - UV\|_F.$$

  - $p < \min\{m, n\}$ is preassigned.
  - Many numerical techniques.
  - Cannot guarantee the required equality in a complete factorization even with $p = \mathrm{rank}_+(A)$.

NMF Fails!

- Almost all NMF techniques adapt conventional mathematical programming schemes.
- The objective function in NMF is non-convex.
- The factors U and V retrieved by NMF techniques are typically local minimizers only.
- Cannot ensure equality to A.

Nonnegative Rank Factorization

$$\mathfrak{R}(m,n) := \left\{ A \in \mathbb{R}_+^{m\times n} \mid \mathrm{rank}(A) = \mathrm{rank}_+(A) \right\}.$$

- Interested in
  - Identifying if $A \in \mathfrak{R}(m,n)$.
  - Procuring a nonnegative factorization for $A \in \mathfrak{R}(m,n)$.
    - Is called a *nonnegative rank factorization* (NRF) of A.

An Example

- Not every nonnegative matrix has an NRF.
- The simplest non-NRF matrix:

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

  - $\mathrm{rank}(\mathscr{C}) = 3$.
  - $\mathrm{rank}_{+}(\mathscr{C}) = 4$.
- There is a necessary and sufficient condition qualifying an NRF matrix (Thomas'1974).
  - Not suitable for computation.

Probability Simplex

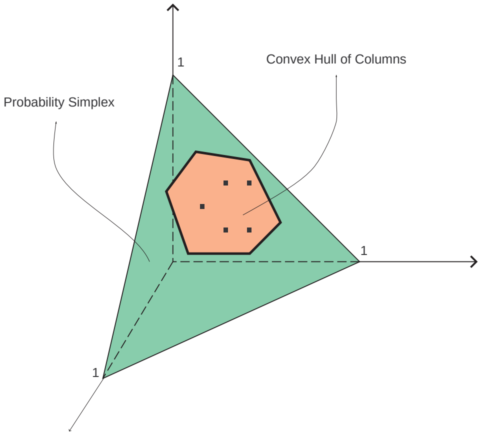
- Given $A \in \mathbb{R}_+^{m \times n}$, define

$$\begin{aligned} \sigma(A) &:= \operatorname{diag}\{\|\mathbf{a}_1\|_1, \ldots, \|\mathbf{a}_n\|_1\} \\ \vartheta(A) &:= A\sigma(A)^{-1}. \end{aligned}$$

- Columns of $\vartheta(A)$ are points on the probability simplex $\mathcal{D}_m$ in $\mathbb{R}_+^m$.

$$\mathcal{D}_m := \left\{\mathbf{a} \in \mathbb{R}_+^m | \mathbf{1}_m^\top \mathbf{a} = 1\right\},$$

Convex Hull of $\vartheta(A) \in \mathbb{R}_+^{3\times 11}$



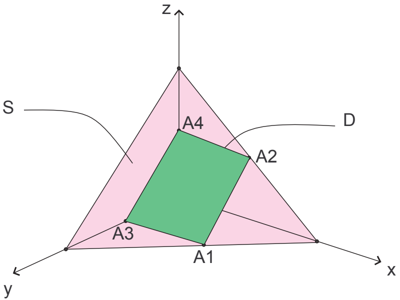
Minimal Convex Polytope

- $A = UV = (UD^{-1})(DV)$ with any invertible diagonal matrix D.
  - May assume $\sigma(U) = I_n$.
- Write
$$A = \vartheta(A)\sigma(A) = UV = \vartheta(U)\vartheta(V)\sigma(V).$$
  - It must be such that
$$\begin{aligned} \vartheta(A) &= \vartheta(U)\vartheta(V), \\ \sigma(A) &= \sigma(V). \end{aligned}$$
  - Columns of $\vartheta(A)$ are in the convex hull of $\vartheta(U)$.

Lemma

The nonnegative rank $\operatorname{rank}_+(A)$ stands for the minimal number of vertices on $\mathcal{D}_m$ so that the resulting convex polytope encloses all columns of the pullback $\vartheta(A)$.

Visualizing $\vartheta(\mathscr{C})$



- Suffices to represent the probability simplex $\mathcal{D}_4$ by the unit tetrahedron S in the first octant of $\mathbb{R}^3$.
- Columns of $\vartheta(\mathscr{C})$ can be interpreted as points A_1, A_2, A_3, A_4.
  - Coplanar because $\mathrm{rank}(\vartheta(\mathscr{C})) = 3$.
  - Four “edges" sitting on separate facets of the tetrahedron.
  - The minimum number of vertices for a convex set in the unit tetrahedron to cover D is four.

How Often Can This Happen?

Question

(**R2R**$_+$) : *Given an arbitrary nonnegative 4 by 4 matrix of rank 3, what is the probability that its nonnegative rank is 3?*

- No easy answer!

Question

Sylvester's Four-Point Problem *What is the probability of four random, independent, and uniform points from a compact set such that none of them lies in the triangle formed by the other three?*

- "This problem does not admit of a determinate solution!" (Sylvester'1865).

Flipped Side Question

Question

($\mathbf{R_{+}2R}$) : *Given an arbitrary nonnegative 4 by 4 matrix of nonnegative rank 3, what is the probability that its rank is 3?*

Theorem

Given $k < \min\{m, n\}$, let $R_{+}(k)$ denote the manifold of nonnegative matrices in $\mathbb{R}_{+}^{m \times n}$ with nonnegative rank k. Then the conditional probability of $\mathrm{rank}(A) = k$, given $A \in R_{+}(k)$, is one.

- Matrices which have an NRF are generic.

Almost Surely?

- If $A = UV$ with randomly generated $U \in \mathbb{R}_+^{m \times k}$ and $V \in \mathbb{R}_+^{k \times n}$,
  - *Almost surely* we have $\mathrm{rank}(A) = k$.
  - The converse is not true.
- Those matrices whose rank are not equal to the nonnegative rank form a measure zero set.
  - Not necessarily mean that the set is “unobservable", nor that nontrivial “exceptions" are difficult to come by.

Euclidean Distance Matrices

- Given n distinct points in $\mathbb{R}^r$:

$$E(\mathbf{q}_1, \ldots \mathbf{q}_n) := \left[\|\mathbf{q}_i - \mathbf{q}_j\|_2^2 \right].$$

  - Of rank $r + 2$, regardless of the number n of points.
  - If $r = 1$, then generically of nonnegative rank n (Chu'2010).
  - Certainly not have the generic phenomenon described above.
  - Form a large and characterizable set, but is too specific to have a nonzero measure.

Perturbation Theory

- Very little perturbation analysis for NRF in general has been studied in the literature.
- Some open questions:
  1. Given a nonnegative matrix A which has an NRF, under what condition will the perturbed nonnegative matrix $A + E$ still have an NRF?
  2. Given a nonnegative matrix A which has an NRF, let U and V be the nonnegative factors found by our (or any) numerical algorithm so that UV is a numerical NRF of A. Is UV the exact NRF of some perturbed nonnegative matrix $A + E$?

Local Rank Condition

Theorem

Given an $m \times n$ non-negative matrix with $\mathrm{rank}_+(A) = k$, then

1. *There exists a ball $B(A; \epsilon)$ such that $\mathrm{rank}_+(N) \geq k$ for all $N \in B(P; \epsilon)$.*
2. *For any $\epsilon > 0$, there exists $N \in B(A; \epsilon)$ such $\mathrm{rank}_+(N) = \mathrm{rank}_+(A)$ and $N \neq \lambda A$ for any λ.*

Chipping Away a Nonnegative Portion

- $B \geq 0$ is a *nonnegative component* (NC) of $A \geq 0$ iff $A - B \geq 0$.
  - Compute the “maximum" rank-one NC of A (Levin’1985).
  - The residual after an NC subtraction might have higher rank.
  - Might end up with an infinite series of NC matrices.
- $B \geq 0$ is a *nonnegative element* (NE) of $A \geq 0$ iff B is a rank-one NC and $\mathrm{rank}(A - B) = \mathrm{rank}(A) - 1$.
- The matrix $\mathscr{C}$ has many NCs, but has no NE at all.
- Need to gradually distribute A over a sequence of NEs.

A Major Difficulty

- A has an NRF, $\Longrightarrow A = \sum_{i=1}^{k} \mathbf{u}_i \mathbf{k}_i^\top$.
  - Each $\mathbf{u}_i \mathbf{k}_i^\top$ is an NE.
- Miss the particular sequence of NE's (not known to begin with)?
  - Any bad choices of NE's in the intermediate stages could cause the rank reduction process to break down.
    - Get stuck at a matrix that has no more NE at all.
- A major challenge:
  - Could not foresee which NE would be a “good" NE to continue on the rank reduction.
  - Finding the right NE's is precisely why the NRF problem is so challenging.
  - A weakness of our approach.
    - Still might be the first sensible way to crack the nut!

Wedderburn Rank Reduction Formula

Theorem

Let $\mathbf{u} \in \mathbb{R}^m$ *and* $\mathbf{v} \in \mathbb{R}^n$*. Then the matrix*

$$B := A - \sigma^{-1}\mathbf{u}\mathbf{v}^\top$$

satisfies the rank subtractivity $\operatorname{rank}(B) = \operatorname{rank}(A) - 1$ *if and only if there are vectors* $\mathbf{x} \in \mathbb{R}^n$ *and* $\mathbf{y} \in \mathbb{R}^m$ *such that*

$$\mathbf{u} = A\mathbf{x}, \quad \mathbf{v} = A^\top\mathbf{y}, \quad \sigma = \mathbf{y}^\top A\mathbf{x}.$$

Theorem

Suppose $U \in \mathbb{R}^{m\times k}$*,* $R \in \mathbb{R}^{k\times k}$*, and* $V \in \mathbb{R}^{n\times k}$*. Then*

$$\operatorname{rank}(A - UR^{-1}V^\top) = \operatorname{rank}(A) - \operatorname{rank}(UR^{-1}V^\top)$$

if and only if there exist $X \in \mathbb{R}^{n\times k}$ *and* $Y \in \mathbb{R}^{m\times k}$ *such that*

$$U = AX, \quad V = A^\top Y, \quad \text{and} \quad R = Y^\top AX.$$

Applications

- Provide a mechanism to break down a matrix into a sum of rank-one matrices.
  - Define a sequence $\{A_k\}$ of matrices by

$$A_{k+1} := A_k - (\mathbf{y}_k^\top A_k \mathbf{x}_k)^{-1} A_k \mathbf{x}_k \mathbf{y}_k^\top A_k.$$

  - Properly chosen vectors $\mathbf{x}_k \in \mathbb{R}^n$ and $\mathbf{y}_k \in \mathbb{R}^m$ satisfying $\mathbf{y}_k^\top A_k \mathbf{x}_k \neq 0$.
  - The process can be continued so long as $A_k \neq 0$.
  - The sequence $\{A_k\}$ must be finite.
- Almost all classical matrix decompositions can be found in this way (Chu, Funderlic, Golub'1995).

Wedderburn for NRF

- For NRF,
  - Break down a nonnegative matrix by taking away one NE a time.
  - An NE must assume the Wedderburn form

$$(\mathbf{y}_k^\top A_k \mathbf{x}_k)^{-1} A_k \mathbf{x}_k \mathbf{y}_k^\top A_k.$$

  - A_{k+1} must be nonnegative after the subtraction.
    - The most difficult part.
- Two probable causes for premature termination.
  - A is not a matrix in $\mathfrak{R}(m, n)$ to begin with.
    - A welcome conclusion.
    - Notion of maximal nonnegative rank splitting of A.
  - Bad starting points have branched A_k into a “dead end" , i.e., A_k has no NE.
    - A restart might remedy the problem.
- Only a heuristic way to find the (approximate) NRF.
- Need more analysis to conclude whether A has an NRF or not.

Maximin Problem

$$\max_{\mathbf{x}_k \in \mathbb{R}^n, \mathbf{y}_k \in \mathbb{R}^m} \quad \min \left[A_k - A_k \mathbf{x}_k \mathbf{y}_k^\top A_k \right],$$

$$\text{subject to} \quad \begin{array}{l} A_k \mathbf{x}_k \geq 0, \\ \mathbf{y}_k^\top A_k \geq 0, \\ \mathbf{y}_k^\top A_k \mathbf{x}_k = 1, \end{array}$$

- $A_k - A_k \mathbf{x}_k \mathbf{y}_k^\top A_k \leq A_k$.
  - maximizer of $\min \left[A_k - A_k \mathbf{x}_k \mathbf{y}_k^\top A \right]$ always exists.
- Nonnegative objective value $\Longrightarrow A_{k+1} \geq 0$.
  - A feasible NE is found.

A Pathological Example

- Consider $A = [\mathscr{C}; \mathbf{c}]$.
  - $\mathbf{c} \geq 0$ is random.
  - $\mathrm{rank}_+(A) = \mathrm{rank}(A) = 4$.
  - Splitting A by its rows is automatically an NRF.
- The matrix $\Delta := [0_4, \mathbf{c}]$ an NE.
  - Leave behind a nonnegative matrix $A - \Delta = [\mathscr{C}; \mathbf{0}]$ of rank 3.
  - $[\mathscr{C}; \mathbf{0}]$ does not have any NRF.
  - Iteration would get stuck.
- Restart $\Longrightarrow$

$$\begin{bmatrix} 1.0000 & 1.0000 & 0 & 0 & 0.0781 \\ 1.0000 & 0 & 1.0000 & 0 & 0.6690 \\ 0 & 1.0000 & 0 & 1.0000 & 0.5002 \\ 0 & 0 & 1.0000 & 1.0000 & 0.2180 \end{bmatrix} = \begin{bmatrix} 0.0000 & 0.0000 & 0.8624 & 0 \\ 0.0000 & 0.3862 & 0 & 0 \\ 3.4544 & 0 & 0 & 0 \\ 0.0000 & 0.0000 & 0.0000 & 1.7018 \end{bmatrix} \begin{bmatrix} 0 & 0.2895 & 0.0000 & 0.2895 & 0.1448 \\ 2.5895 & 0.0000 & 2.5895 & 0.0000 & 1.7325 \\ 1.1596 & 1.1596 & 0.0000 & 0.0000 & 0.0905 \\ 0 & 0 & 0.5876 & 0.5876 & 0.1281 \end{bmatrix}.$$

Multiple Rank Reduction

$$\max_{X_k \in \mathbb{R}^{m\times r}, Y_k \in \mathbb{R}^{n\times r}} \quad \min \left[A_k - A_k X_k Y_k^\top A_k \right],$$

$$\text{subject to} \quad \begin{array}{l} A_k X_k \geq 0, \\ Y_k^\top A_k \geq 0, \\ Y_k^\top A_k X_k = I_{r\times r}, \end{array}$$

Theorem

If a nonnegative matrix has a nonnegative rank-r reduction, then it must have r nonnegative rank-one reductions.

Completely Positive Matrices

- $A \in \mathbb{R}_+^{n \times n}$ is *completely positive* (CP) iff

$$A = BB^\top.$$

  - $B \geq 0$ is not necessarily square.
  - Interested in the smallest number of columns of B.
    - Denoted by $\mathrm{rank}_{cp}(A)$.
- Two questions (Berman'2003):
  - Determine whether a given nonnegative semi-definite matrix is CP.
  - Determine the cp-rank and compute its CP factorization.
    - More stringent than NRF.

Symmetric Wedderburn Formula

$$\max_{\mathbf{x}_k \in \mathbb{R}^n} \quad \min \left[A_k - A_k \mathbf{x}_k \mathbf{x}_k^\top A_k \right]$$

$$\text{subject to} \quad A_k \mathbf{x}_k \geq 0$$
$$\mathbf{x}_k^\top A_k \mathbf{x}_k = 1$$

Maximal Nonnegative Rank Splitting

Question

(MNRS): *Given a nonnegative matrix A, find a splitting*

$$A = B + C,$$

where both B and C are nonnegative matrices satisfying

$$\operatorname{rank}(B) = \operatorname{rank}_+(B),$$
$$\operatorname{rank}(A) = \operatorname{rank}(B) + \operatorname{rank}(C),$$

and $\operatorname{rank}(B)$ *is maximized.*

- If $A \in \mathfrak{R}(m, n)$, then trivially $B = A$ and $C = 0$.
- If $A \notin \mathfrak{R}(m, n)$,
  - A might still has a few NEs.
  - Retrieve all possible NEs of A.

Conclusion

- Detecting the nonnegative rank and computing the corresponding nonnegative factorization for a general nonnegative matrix are very challenging tasks both in theory and in practice.
- No existing algorithms can guarantee to find the NRF.
- Exploit the Wedderburn rank reduction formula to *downdate* a nonnegative matrix.
- Only a possible computational tool for the NRF problem.
- Need more perturbation analysis for NRF in general.