NONNEGATIVE RANK FACTORIZATION VIA RANK REDUCTION

BO DONG*, MATTHEW M. LIN†, AND MOODY T. CHU‡

Abstract. Every nonnegative matrix can be expressed as the sum of nonnegative rank one matrices. Nonnegative rank factorization concerns finding the shortest rendition among all possible series representations. Such a complete representation is a special case of nonnegative matrix factorizations which play important roles in many scientific applications. Locating the exact nonnegative rank and computing the corresponding factorization are challenging both theoretically and computationally. Only a few abstract results, which are somewhat too conceptual for numerical implementation, have been developed for the existence of a nonnegative rank factorization. Available nonnegative matrix factorization techniques, which mostly utilize the notion of approximation, cannot guarantee a complete factorization. Employing the Wedderburn rank reduction formula, this paper presents a numerical procedure to detect whether a nonnegative rank factorization exists and, in the event that such a factorization does not exist, it is able to compute the maximum nonnegative rank splitting. With a slight modification for symmetry, the method can equally be applied to completely positive matrices. Numerical testing seems to suggest that the proposed method is a reasonable means for analyzing the nonnegative rank factorization.

Key words. nonnegative matrix, nonnegative matrix factorization, nonnegative rank, nonnegative rank factorization, Wedderburn rank reduction formula, completely positive matrix, cp-rank

AMS subject classifications. 15A23, 41A36, 52B11, 90C47, 65F30

1. Introduction.

Mathematical models with nonnegative data are abounding in science and engineering. The notion of nonnegative matrix factorization (NMF) plays a major role in a wide range of important applications, including text mining [17, 47], cheminformatics [37, 38], image articulation [34], bioinformatics [10], micro-array analysis [19], air pollution research [37], and spectral data analysis [6]. The basic idea behind the NMF of a given nonnegative matrix $A \in \mathbb{R}^{m\times n}$ is the typical low rank approximation which is often formulated as a minimization problem

$$\min_{U\in\mathbb{R}_+^{m\times k}, V\in\mathbb{R}_+^{k\times n}} \|A - UV\|_F, \tag{1.1}$$

where k is a specified integer, $\mathbb{R}_+$ stand for the set of nonnegative real numbers, and U, V are considered as low rank "factors" of A. It has been over a decade since the NMF was first utilized in application [38]. Numerous NMF techniques have been proposed in the literature, including the often cited multiplicative update algorithm by Lee and Seung [34, 35], the gradient methods [11, 26], and alternating least square methods [7, 8, 31, 33, 37]. The list is hardly exhaustive, see additional discussions in [14, 18, 23, 25, 30, 37, 38, 41], and the many references contained therein.

The potential impact of the NMF on scientific advancement might be as great as the singular value decomposition (SVD) in that it strives to accomplish what the SVD does with the additional nonnegativity constraints. In contrast to the SVD, the NMF, if successively realized, can improve interpretability and visualization of large scale data while maintaining the physical feasibility more closely. Unfortunately, most algorithms at present are data specific, the rate of convergence is slow, and the limit point is often far from being optimal. The commonly utilized Lee-Seung algorithm, for example, can break down and its limit point, when slowly reached, sometimes is not even stationary. Another major obstacle to a successful application of the NMF is that the low rank k must be given a prior, which is difficult to come by in practice. Many open questions remain to be resolved for the NMF.

*Department of Applied Mathematics, Dalian University of Technology, Dalian, Liaoning, 116024. (dongbodlut@gmail.com). This work is partially supported by Chinese Scholarship Council.

†Department of Mathematics, North Carolina State University, Raleigh, NC 27695-8205. (mlin@ncsu.edu). This research was supported in part by the National Science Foundation under grants DMS-0505880 and DMS-0732299.

‡Department of Mathematics, North Carolina State University, Raleigh, NC 27695-8205. (chu@math.ncsu.edu). This research was supported in part by the National Science Foundation under grants DMS-0505880 and DMS-0732299 and NIH Roadmap for Medical Research grant 1 P20 HG003900-01.

The purpose of this paper is to investigate a more fundamental issue to the NMF — the so called nonnegative rank and the associated nonnegative rank factorization (NRF). To motivate the idea, we adopt the notion by Cohen and Rothblum [16]. Observe that any given nonnegative matrix $A \in \mathbb{R}^{m\times n}$ can be expressed as the sum of a series of nonnegative rank one matrices. In particular, if $A = UV$, where $U = [\mathbf{u}_1, \ldots, \mathbf{u}_k]$ and $V = [\mathbf{v}_1^\top, \ldots, \mathbf{v}_k^\top]^\top$ with nonnegative vectors $\mathbf{u}_i \in \mathbb{R}^m$ and $\mathbf{v}_i \in \mathbb{R}^n$, then $A = \sum_{i=1}^k \mathbf{u}_i\mathbf{v}_i^\top$. Among the many possible series representations of A by nonnegative rank one matrices, the smallest integer k for which such a nonnegative rank one series representation is attained is called the *nonnegative rank* of the nonnegative matrix A. For convenience, we denote the nonnegative rank by $\mathrm{rank}_+(A)$. The bounds

$$\mathrm{rank}(A) \leq \mathrm{rank}_+(A) \leq \min\{m, n\}$$

for the nonnegative rank can readily be proved.

A careful distinction about the NRF needs to be made here. According to Cohen and Rothblum [16], the nonnegative decomposition $A = UV$ where U and V have precisely $\mathrm{rank}_+(A)$ columns and rows, respectively, is called a *nonnegative factorization* of A. As such, *every* nonnegative matrix has a "shortest" nonnegative factorization. On the other hand, the more traditional definition of nonnegative rank factorization refers to only this subclass of nonnegative matrices,

$$\mathfrak{R}(m, n) := \left\{A \in \mathbb{R}^{m\times n} \mid A \geq 0,\ \ \mathrm{rank}(A) = \mathrm{rank}_+(A)\right\}. \tag{1.2}$$

One of our contributions in this paper is to show that this class of matrices is generic in the sense that those matrices whose rank is strictly less than its nonnegative rank from a set of measure zero. For this reason, by an NRF, we refer to matrices in $\mathfrak{R}(m, n)$ only.

An NMF is not necessarily an NRF in that the latter demands a complete factorization whereas the former is only an approximation. Furthermore, not every nonnegative matrix has an NRF. Interesting examples of nonnegative matrices without NRF are constructed specifically in [29], the simplest being that

$$\mathscr{C} = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{pmatrix}, \tag{1.3}$$

with $\mathrm{rank}(\mathscr{C}) = 3$ and $\mathrm{rank}_+(\mathscr{C}) = 4$, despite the fact that this kind of matrices forms a measure zero set. A necessary and sufficient condition qualifying the existence of the NRF of a nonnegative matrix is given by Thomas [43], which would be onerous to check algebraically. Jeter and Pye characterize an existence condition of NRF for weakly monotone nonnegative matrices [28]. Jain and Tynan recently present another set of conditions of NRF for λ-monotone matrices which arise from the notion of nonnegative least-square solutions to linear systems [27].

Even if the NRF is known to exist in theory, to actually compute the factors in the decomposition is not an easy task by all means. It might be so speculated that a "good" NMF algorithm should be able to compute the NRF of a given nonnegative matrix, once the nonnegative rank is known. Indeed, this should be the starting point to assess the capability of any NMF algorithms, that is, to demand that the objective value in (1.1) be zero if k is precisely the nonnegative rank. It turns out that most existing NMF algorithms, designed to find local solutions only, cannot guarantee to compute the NRF even if k is properly specified. Campbell and Poole have suggested an idea similar to the Gaussian elimination by carefully selected nonnegative multipliers, but the matrix A must have nonnegative 1-inverse. For most nonnegative matrices, the selection process cannot be generalized [9]. Levin's idea of computing the "maximum" rank one approximation to a positive matrix while keeping the residual nonnegative seems appealing [36]. For the convenience of later discussions, any nonnegative matrix whose subtraction from a given nonnegative matrix remains nonnegative is

called a *nonnegative component* (NC) of that given matrix. Be noted that the residual of a rank one NC may have the same rank as the original matrix, which is precisely the shortcoming in Levin's approach. If the process is to be repeated, we may end up with an infinite series of nonnegative rank one matrices, defying the minimal length of what an NRF desires. So far as we know, little is done in the literature about computing the NRF of a given matrix in $\mathfrak{R}(m,n)$.

Our main contribution in this paper is to offer a numerical way to compute the NRF of a given nonnegative matrix in $\mathfrak{R}(m,n)$. Given the indication in [4, 44] that the determination of the nonnegative rank of a nonnegative matrix is still a subject of research and the recent discovery in [45] that the "exact NMF", which is our NRF, is NP-hard, we think that our algorithm, being capable to compute the NRF if exists, should be an important contribution to the field. We think our approach is innovative and might help to advance one extra step toward the understanding of the important subject of nonnegative factorizations.

The principal tool in our algorithm is the Wedderburn rank reduction formula. Similar to Levin's approach, our idea is to search for a rank one NC of a given matrix A while keeping the residual nonnegative. Different from Levin's approach, we insist that if this rank one NC is subtracted from the original matrix A, the rank of the residual must be reduced by one. For convenience, we refer to such a rank one reduction NC as a *nonnegative element* (NE) of A. The existence of an NE is determined by a maximin optimization. We repeat the procedure by taking away one element a time. Since the rank is reduced by one in each step, the iteration must be terminated in at most $\mathrm{rank}(A)$ steps. If at any step it is absolutely determined an NE cannot be found, then we know that $\mathrm{rank}_+(A) > \mathrm{rank}(A)$, that is, the matrix A is not in $\mathfrak{R}(m,n)$ and has no NRF.

Our paper is organized as follows. In Section 2 we briefly review the critical Wedderburn rank reduction formula. For our application, we have to introduce additional constraints that both the factors and the residual be kept nonnegative. To satisfy the nonnegativity constraints in the rank reduction process, we transform in Section 3 the problem of NRF to a maximin problem which can be solved recursively by numerical techniques. In Section 4, we raise an interesting question of maximal nonnegative rank splitting of a nonnegative matrix when an NRF does not exist. With a slight modification for symmetry, we discuss in Section 5 that our algorithm can be applied to the class of completely positive matrices and offers a numerical means to determine whether the the so called cp-rank of a completely positive matrix agrees with its rank. Again, we think this approach is new in the literature. In Section 6, we sketch a geometric meaning of the nonnegative rank by which we relate the problem of NRF to the classical Sylvester's problem and an interesting issue of probability is addressed. Finally, some enlightening empirical results by our algorithm are reported in Section 7.

2. Wedderburn rank reduction formula. It appeared as a simple statement in Wedderburn's 1934 book [46] and as an exercise in Household's 1964 book [24] that the rank one matrix $\sigma^{-1}A\mathbf{x}\mathbf{y}^\top A$, when subtracted from A, would result in a matrix of rank one less than A if $\sigma = \mathbf{y}^\top A\mathbf{x}$. Later, the converse was also proved true [20]. Together, the following is what we refer to as the *Wedderburn rank one reduction formula*.

THEOREM 2.1. *Let* $\mathbf{u} \in \mathbb{R}^m$ *and* $\mathbf{v} \in \mathbb{R}^n$. *Then the matrix*

$$B := A - \sigma^{-1}\mathbf{u}\mathbf{v}^\top \tag{2.1}$$

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

$$\mathbf{u} = A\mathbf{x}, \quad \mathbf{v} = A^\top\mathbf{y}, \quad \sigma = \mathbf{y}^\top A\mathbf{x}. \tag{2.2}$$

Indeed, simultaneous multiple rank reduction is possible. Cline and Funderlic proved that a generalized formula holds in the following sense [15].

THEOREM 2.2. *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*

$$U = AX, \quad V = A^\top Y, \quad \text{and} \quad R = Y^\top AX. \tag{2.3}$$

This seemingly straightforward formula (2.1) has profound consequences in that it provides a unified mechanism for constructing all known matrices factorizations. A detailed discussion in this regard can be found in the review paper [12]. The basic idea is that, starting with $A_1 = A$, then so long as $A_k \neq 0$, we can repeatedly apply the Wedderburn formula to generate a sequence $\{A_k\}$ of matrices by defining

$$A_{k+1} := A_k - (\mathbf{y}_k^\top A\mathbf{x}_k)^{-1} A_k\mathbf{x}_k\mathbf{y}_k^\top A_k. \tag{2.4}$$

for 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 rank of A_k being reduced by one at each step without any other additional condition, the sequence $\{A_k\}$ must be finite. In this way, the matrix A is decomposed as a finite series of rank one matrices. Different selections of $\mathbf{x}_k$ and $\mathbf{y}_k$ lead to different types of decompositions.

For our application, we want to break down a nonnegative matrix by taking away one piece of its NEs a time. The question is how "large" a rank one matrix can be removed. If too small a matrix is subtracted, then the process may be perpetuated and we will have more than $\operatorname{rank}_+(A)$ many terms in the series. If too large a matrix is subtracted, then the residual may become negative, which may have no more NE. Our remedy is to use the Wedderburn formula to control the "size" of the rank one matrix. More specifically, we search for $\mathbf{x}_k \in \mathbb{R}^n$ and $\mathbf{y}_k \in \mathbb{R}^m$ such that the rank one matrix $(\mathbf{y}_k^\top A\mathbf{x}_k)^{-1} A_k\mathbf{x}_k\mathbf{y}_k^\top A_k$ is nonnegative. This can be accomplished by requiring factors $A_k\mathbf{x}_k \geq 0$, $\mathbf{y}_k^\top A_k \geq 0$, and $\mathbf{y}_k^\top A\mathbf{x}_k > 0$, due to the homogeneity in the expression. Such a task is easy to do if A_k is nonnegative. To keep this process alive, we thus further require that the residual A_{k+1} remains nonnegative. Once these necessary nonnegativity constraints are satisfied, the rank reduction formula (2.4) kicks in. We repeat the process until the rank is reduced to zero and an NRF is achieved.

If $\operatorname{rank}_+(A) \neq \operatorname{rank}(A)$, then it is possible that the process will be terminated prematurely. This happens only in the case that, while the Wedderburn formula guarantees the rank reduction of A_k, the residual A_{k+1} cannot stay nonnegative by *any* nonnegative rank one reduction. This would be an indication that the matrix A does not have an NRF.

3. NRF by Wedderburn formula. In this section, we give specifics about the application of the Wedderburn rank reduction formula to the NRF. We begin with rank one reduction which will be used in computation and generalize the concept to multiple rank reduction which is of theoretical interest in its own right.

We first rephrase the notion of nonnegativity constraints. It is easy to see that the rank one matrix $\mathbf{u}\mathbf{v}^\top$ is nonnegative if and only if all entries of $\mathbf{u}$ and $\mathbf{v}$ are of one sign. Since the scale $\mathbf{y}_k^\top A_k\mathbf{x}_k$ in the Wedderburn formula (2.4) must be either positive or negative, we can assume without loss of generality nonnegative factors $A_k\mathbf{x}_k$ and $\mathbf{y}_k^\top A_k$ in our setting. We assume further after re-scaling that $\mathbf{y}_k^\top A_k\mathbf{x}_k = 1$ henceforth. These constitute the "*nonnegativity constraints*". If these constraints are satisfied, the subtraction of the nonnegative rank one matrix $A_k\mathbf{x}_k\mathbf{y}_k^\top A_k$ from A_k will reduce the rank by one automatically.

The nature of the NRF of a nonnegative matrix is the "additivity" of nonnegative rank one matrices. We cannot immediately regard $A_k\mathbf{x}_k\mathbf{y}_k^\top A_k$ as an NE of A because it might over deduct the matrix A to negative entries. To prevent this from happening, we consider the following optimization

problem:

$$\begin{aligned} \max_{\mathbf{x}_k\in\mathbb{R}^n,\mathbf{y}_k\in\mathbf{R}^m} \quad & \min\left(A_k - A_k\mathbf{x}_k\mathbf{y}_k^\top A\right), \\ \text{subject to} \quad & A_k\mathbf{x}_k \geq 0, \\ & \mathbf{y}_k^\top A_k \geq 0, \\ & \mathbf{y}_k^\top A_k\mathbf{x} = 1, \end{aligned} \tag{3.1}$$

where the minimum is taken over all entries of the matrix. We consider $A_k\mathbf{x}_k\mathbf{y}_k^\top A_k$ to be a feasible NE of A_k only if the objective value is nonnegative. In this case, we say that A_k has a *nonnegative rank one reduction*. We may proceed to define A_{k+1} by (2.4) and repeat the procedure. If the nonnegative matrix $A_1 = A$ has an NRF, that is, $\text{rank}_+(A) = \text{rank}(A)$, then in exactly $\text{rank}(A)$ steps we should get the NRF of the matrix A.

Be aware that by our nonnegativity constraints, $A_k - A_k\mathbf{x}_k\mathbf{y}_k^\top A_k \leq A_k$. Thus, the maximizer of $\min\left(A_k - A_k\mathbf{x}_k\mathbf{y}_k^\top A\right)$ always exists. There are readily available routines for solving (3.1). For example, the MATLAB routine FMINIMAX implements a sequential quadratic programming method and seems capable of handling our problem reasonably well. The only issue that will concern us is a negative objective value, indicating that a rank one NE has not been found yet. We summarize our algorithms as follows:

ALGORITHM 3.1. (NRF by rank one reduction)

Given a nonnegative matrix $A \in \mathbb{R}^{m\times n}$ and a threshold of machine zero ϵ, set $k = 1$ and $A_1 = A$.

Step 1. If $\|A_k\|_F \geq \epsilon$, go to Step 2. Otherwise, retrieve the following information and stop.
1. $\text{rank}(A) = \text{rank}_+(A) = k - 1$.
2. *The NRF of A is approximately given by the summation $A \approx \sum_{i=1}^{k-1} A_i\overline{\mathbf{x}}_i\overline{\mathbf{y}}_i^\top A_i$ with an error less than ϵ.*

Step 2. Randomly select feasible initial value $(\mathbf{x}_k^{(0)}, \mathbf{y}_k^{(0)})$ satisfying the nonnegativity constraints.

Step 3. Solve the maximin problem (3.1).

Step 4. If the objective value at the local maximizer $(\overline{\mathbf{x}}_k, \overline{\mathbf{y}}_k)$ is negative, go to Step 5. Otherwise, do update as follows and go to Step 1.
1. *Define $A_{k+1} := A_k - A_k\overline{\mathbf{x}}_k\overline{\mathbf{y}}_k^\top A_k$.*
2. *Set $k = k + 1$.*

Step 5. Carefully evaluate the situation of whether a change of the initial value in Step 2 can lead to a new nonnegative objective value in Step 3. That is, restart Steps 2 and 3 multiple times. If it is decided within reasonable trials that no initial value can result in nonnegative objective values, report with caution that the matrix A does not has an NRF and stop.

We must stress that most maximin algorithms, including the routine FMINIMAX in MATLAB, only find a local solution for the optimization problem (3.1). For our NRF application, however, a local solution is sufficient so long as the objective value at the local maximizer is nonnegative. Unlike Levin's method, we are not interested in a "maximal" NC which typically will not reduce the rank. Rather, we gradually distribute A over a sequence of nonnegative rank one matrices by ensuring that each of which will reduce the rank by one. In the event that a local solution returns a negative objective value, we suggest in Step 5 to reprocess the matrix by different initial values. From our extensive numerical experiments, however, we want to convey our experience that Algorithm 3.1 is generally robust and Step 5 rarely happens if the matrix A is in $\mathfrak{R}(m, n)$. Some testing results are presented in Section 7.

For computation, the rank one reduction procedure as is proposed in Algorithm 3.1 is sufficient. Nonetheless, Theorem 2.2 suggests that we should be able to reduce multiple ranks of the matrix A_k simultaneously in one step. Would this multiple rank reduction procedure be computationally more efficient? For NRF applications, suppose we want to reduce the rank by r, then we would initially

consider one of the following two optimization problems:

$$\begin{aligned} \max_{X_k\in\mathbb{R}^{m\times r},Y_k\in\mathbb{R}^{n\times r}} \quad & \min\left(A_k - A_kX_k(Y_k^\top A_kX_k)^{-1}Y_k^\top A_k\right), \\ \text{subject to} \quad & A_kX_k(Y_k^\top A_kX_k)^{-1} \geq 0, \\ & Y_k^\top A_k \geq 0, \end{aligned}$$

or

$$\begin{aligned} \max_{X_k\in\mathbb{R}^{m\times r},Y_k\in\mathbb{R}^{n\times r}} \quad & \min\left(A_k - A_kX_k(Y_k^\top A_kX_k)^{-1}Y_k^\top A_k\right), \\ \text{subject to} \quad & A_kX_k \geq 0, \\ & (Y_k^\top A_kX_k)^{-1}Y_k^\top A_k \geq 0. \end{aligned}$$

Certainly, it is not easy to invert the matrix $Y_k^\top A_kX_k$ while demanding either $A_kX_k(Y_k^\top A_kX_k)^{-1}$ or $(Y_k^\top A_kX_k)^{-1}Y_k^\top A_k$ be nonnegative. This inconvenience can easily be averted by absorbing the inverse matrix $(Y_k^\top A_kX_k)^{-1}$ into one of the variable matrices. For example, we can rewrite $A_kX_k(Y_k^\top A_kX_k)^{-1}$ as $A_k\left(X_k(Y_k^\top A_kX_k)^{-1}\right)$ and recognize $X_k(Y_k^\top A_kX_k)^{-1}$ as the new variable matrix X_k. In this way, the two optimization problems are equivalent to the following formulation:

$$\begin{aligned} \max_{X_k\in\mathbb{R}^{m\times r},Y_k\in\mathbb{R}^{n\times r}} \quad & \min\left(A_k - A_kX_kY_k^\top A_k\right), \\ \text{subject to} \quad & A_kX_k \geq 0, \\ & Y_k^\top A_k \geq 0, \\ & Y_k^\top A_kX_k = I_{r\times r}, \end{aligned} \tag{3.2}$$

where $I_{r\times r}$ is the identity matrix of rank r.

Some remarks are due on the computational complexity. For rank one reduction, there are $n+m$ variables involved in the problem (3.1). For rank r reduction, there are $r(n+m)$ variables involved in the problem (3.2). When r is large, the complexity for the problem (3.2) even by the same routine would grow far more rapidly than the total complexity of r applications to the problem (3.1). On the other hand, from the next theorem, we can see that if the rank of the matrix A_k can be reduced by r in one step, then we can achieve the same reduction by rank one matrices in r steps.

THEOREM 3.1. *If a nonnegative matrix has a nonnegative rank r reduction, then it must have r nonnegative rank one reductions.*

Proof. Suppose the nonnegative matrix A has a nonnegative rank r reduction. By Theorem 2.2, we know there are matrices $X \in \mathbb{R}^{n\times r}$ and $Y \in \mathbb{R}^{m\times r}$ such that the matrix

$$AX \geq 0, \quad Y^\top A \geq 0, \quad Y^\top AX = I_{r\times r}$$

and the nonnegative matrix $B := A - AXY^\top A$ has rank r less than the rank of the matrix A. Denote the columns of X and Y as

$$X := (\mathbf{x}_1, \ldots, \mathbf{x}_r), \quad Y := (\mathbf{y}_1, \ldots, \mathbf{y}_r).$$

then we have

$$A\mathbf{x}_i \geq 0, \quad \mathbf{y}_i^\top A \geq 0, \quad \mathbf{y}_i^\top A\mathbf{x}_j = \begin{cases} 1, & \text{if } i = j, \\ 0, & \text{if } i \neq j. \end{cases}$$

We need to show that the pair $(\mathbf{x}_k, \mathbf{y}_k)$ has the desirable effect of rank one reduction on A_k.

Beginning with $A_1 = A$, it is obvious that for $1 \leq i, j \leq r$, we have

$$A_1\mathbf{x}_i = A\mathbf{x}_i \geq 0, \quad \mathbf{y}_i^\top A_1 = \mathbf{y}_i^\top A \geq 0, \quad \mathbf{y}_i^\top A_1\mathbf{x}_j = \mathbf{y}_i^\top A\mathbf{x}_j = \begin{cases} 1, & \text{if } i = j, \\ 0, & \text{if } i \neq j. \end{cases}$$

Assume the statement that

$$A_k\mathbf{x}_i \geq 0, \quad \mathbf{y}_i^\top A_k \geq 0, \quad \mathbf{y}_i^\top A_k x_j = \begin{cases} 1, & \text{if } i = j, \\ 0, & \text{if } i \neq j, \end{cases}$$

is true for all $k \leq i, j \leq r$. Recall that $A_{k+1} = A_k - A_k\mathbf{x}_k\mathbf{y}_k^\top A_k$. It follows that for all $k+1 \leq i, j \leq r$, we have

$$\begin{aligned} A_{k+1}\mathbf{x}_i &= (A_k - A_k\mathbf{x}_k\mathbf{y}_k^\top A_k)\mathbf{x}_i = A_k x_i \geq 0, \\ \mathbf{y}_i^\top A_{k+1} &= \mathbf{y}_i^\top (A_k - A_k\mathbf{x}_k\mathbf{y}_k^\top A_k) = \mathbf{y}_i^\top A_k \geq 0, \\ \mathbf{y}_i^\top A_{k+1}\mathbf{x}_j &= \mathbf{y}_i^\top (A_k - A_k\mathbf{x}_k\mathbf{y}_k^\top A_k)\mathbf{x}_j = \mathbf{y}_i^\top A_k\mathbf{x}_j = \begin{cases} 1, & \text{if } i = j, \\ 0, & \text{if } i \neq j. \end{cases} \end{aligned}$$

By the mathematical induction, we have shown that the same matrix B can be achieved by r rank one reductions via the sequence of vectors $(\mathbf{x}_k, \mathbf{y}_k)$. □

Theorem 3.1 asserts an important consequence in that no matrix can have a nonnegative rank r reduction without first having a nonnegative rank one reduction. In other words, if Algorithm 3.1 fails to secure an NE for a matrix A_k, then A_k does not have any other rank reduction NC except for itself. This make-or-break feature enables us to determine conclusively whether a matrix has an NRF or not.

4. Maximal nonnegative rank splitting. Our algorithm gives rise to an interesting application when dealing with a general nonnegative matrix. It is natural to ask the following question about the maximal nonnegative rank splitting (MNRS) of an arbitrary nonnegative matrix.

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

$$A = B + C, \tag{4.1}$$

where both B and C are nonnegative matrices satisfying

$$\begin{aligned} \operatorname{rank}(B) &= \operatorname{rank}_+(B), \\ \operatorname{rank}(A) &= \operatorname{rank}(B) + \operatorname{rank}(C), \end{aligned}$$

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

If $A \in \mathfrak{R}(m, n)$, then trivially $B = A$ and $C = 0$. Consider the case that $A \notin \mathfrak{R}(m, n)$. We know that A has no NRF. However, it is plausible that A still has a few NEs. By repeatedly applying our method until it is terminated, say, at A_k, that is, A_k has no more NE, we obtain the byproduct that $B = A - A_k$ and $C = A_k$ form an MNRS for A.

It would be instructive to demonstrate the notion of MNRS by examples.

Example 1: The matrix $\mathscr{C}$ defined in (1.3) has many NCs, but our algorithm claims that it has no NE at all. The following analysis assures that this conclusion is true.

Recall that $\operatorname{rank}(\mathscr{C}) = 3$ and $\operatorname{rank}_+(\mathscr{C}) = 4$. If there were an NE for $A_1 = \mathscr{C}$, then the reduced matrix A_2 would be nonnegative and of rank 2. But then by [16, Theorem 4.1], the matrix A_2 would automatically have nonnegative rank 2, implying the matrix $\mathscr{C}$ would have nonnegative rank 3. This is a contradiction. In other words, the matrix $\mathscr{C}$ has zero B component in its MNRS.

Example 2. Consider the 8×8 nonnegative matrix $A := (W, H)$, where $W := (\mathscr{C}, 0_4)^\top \in \mathbf{R}^{8\times 4}$ with 0_4 standing for the zero matrix of size 4×4 and $H := H_1H_2 \in \mathbb{R}^{8\times 4}$ with

$$H_1 := \begin{pmatrix} 0.2917 & 0.3109 & 0.2026 \\ 0.4665 & 0.2558 & 0.9396 \\ 0.9439 & 0.1048 & 0.2107 \\ 0.0943 & 0.2903 & 0.9670 \\ 0.0119 & 0.4985 & 0.6356 \\ 0.3723 & 0.8205 & 0.4252 \\ 0.3542 & 0.3074 & 0.2262 \\ 0.0820 & 0.7715 & 0.9325 \end{pmatrix}, \quad H_2 := \begin{pmatrix} 0.7426 & 0.2143 & 0.0907 & 0.1922 \\ 0.5133 & 0.8007 & 0.8121 & 0.0639 \\ 0.5417 & 0.6280 & 0.0968 & 0.4969 \end{pmatrix}.$$

Note that $\text{rank}(H) = \text{rank}_+(H) = 3$. Clearly, $\text{rank}(A) = 6$. Can we tell whether A has an NRF, or can we retrieve an MNRS of A?

Upon applying Algorithm 3.1 to A, we obtain an MNRS $A = B + C$ with

$$B = \begin{pmatrix} 0 & 0 & 0 & 0 & 0.17825 & 0.36892 & 0.29782 & 0.08302 \\ 0 & 0 & 0 & 0 & 0.56400 & 0.47683 & 0.34027 & 0.18382 \\ 0 & 0 & 0 & 0 & 0.38505 & 0.33387 & 0.19112 & 0.16500 \\ 0 & 0 & 0 & 0 & 0.14403 & 0.42702 & 0.33791 & 0.09387 \\ 0 & 0 & 0 & 0 & 0.60902 & 0.80086 & 0.46744 & 0.34997 \\ 0 & 0 & 0 & 0 & 0.92796 & 1.00378 & 0.74126 & 0.33527 \\ 0 & 0 & 0 & 0 & 0.54335 & 0.46409 & 0.30366 & 0.20012 \\ 0 & 0 & 0 & 0 & 0.96204 & 1.22092 & 0.72424 & 0.52842 \end{pmatrix}$$

and

$$C = \begin{pmatrix} 1 & 1 & 0 & 0 & 0.30770 & 0.06976 & 0.00073 & 0.09358 \\ 1 & 0 & 1 & 0 & 0.42271 & 0.41803 & 0.00073 & 0.38908 \\ 0 & 1 & 0 & 1 & 0.48382 & 0.08464 & 0 & 0.12781 \\ 0 & 0 & 1 & 1 & 0.59883 & 0.43291 & 0 & 0.42331 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix},$$

where $\text{rank}(B) = \text{rank}_+(B) = 3$ and, due to the Wedderburn rank reduction process, $\text{rank}(C) = \text{rank}(A) - \text{rank}(B) = 3$. However, with $\mathscr{C}$ embedded at the upper left corner, the matrix C has nonnegative rank $\text{rank}_+(C) = 4$ and has no more NE.

5. Completely positive matrices. A nonnegative matrix $A \in \mathbb{R}^{n\times n}$ is said to be *completely positive* (CP) if and only if A can be factorized as

$$A = BB^\top, \tag{5.1}$$

where B is nonnegative [5]. The matrix B is not necessarily square. The smallest number of columns of B satisfying the factorization (5.1) is called the *cp-rank* of the matrix A, denoted by $\text{rank}_{cp}(A)$. General properties and some applications of CP matrices can be found in the reference [5]. So far as we know, there are two open questions about CP matrices. First, determine whether a given nonnegative semi-definite matrix is CP. Secondly, if a matrix is CP, determine its cp-rank [3].

If A is CP, then obviously $\text{rank}_+(A) \le \text{rank}_{cp}(A)$. Some sufficient conditions under which $\text{rank}_{cp}(A) = \text{rank}(A)$ can be found in [39]. If A is generated by a Soules matrix, for example, then $\text{rank}_{cp}(A) = \text{rank}(A)$ [40]. On the other hand, it has been proved that if A is CP, then we have the upper bound estimate of the cp-rank [22]

$$\text{rank}_{cp}(A) \le \frac{\text{rank}(A)(\text{rank}(A)+1)}{2},$$

which can be slightly improved to [2]

$$\text{rank}_{cp}(A) \le \frac{\text{rank}(A)(\text{rank}(A)+1)}{2} - 1,$$

provided $\text{rank}(A) > 1$. Other than these results, we find very few discussions about how to compute the CP factorization. Judging from the symmetric form in (5.1), the CP problem for a nonnegative seme-definite matrix seems more stringent than the NRF problem for a general nonnegative matrix. Our contribution to this subject is that we are able to determine whether $\text{rank}_{cp}(A) = \text{rank}(A)$ and, if affirmative, we can compute the factor B.

Our idea is to continue applying the Wedderburn formula with the requirement that the NE must be symmetric. In order to maintain this symmetry, it must be such that $\mathbf{x} = \mathbf{y}$ in the rank

reduction formula [13]. Starting with $A_1 = A$, the optimization problem in (3.1) is reformulated as follows:

$$\begin{aligned} \max_{\mathbf{x}_k \in \mathbb{R}^n} \quad & \min \left(A_k - A_k \mathbf{x}_k \mathbf{x}_k^\top A_k \right) \\ \text{subject to} \quad & A_k \mathbf{x}_k \geq 0 \\ & \mathbf{x}_k^\top A_k \mathbf{x}_k = 1 \end{aligned} \tag{5.2}$$

with $A_{k+1} := A_k - A_k \mathbf{x}_k \mathbf{x}_k^\top A_k$. If the objective value at the local maximizer is nonnegative, then the matrix A_k has a symmetric NE. Since the reduced matrix A_{k+1} remains symmetric and nonnegative, the rank reduction process can be repeated until such a symmetric NE can be found no more. As the rank is reduced by one in each step, the process must be terminated in at most $\operatorname{rank}(A)$ steps. If $\operatorname{rank}_{cp}(A) = \operatorname{rank}(A)$, by applying the rank reduction process $\operatorname{rank}(A)$ times, we can find the nonnegative factor $B = (A_1 \mathbf{x}_1, \ldots, A_{\operatorname{rank}(A)} \mathbf{x}_{\operatorname{rank}(A)})$. If $\operatorname{rank}_{cp}(A) \neq \operatorname{rank}(A)$, the rank reduction process will be terminated earlier. In this case, we certainly can address the issue of the maximal symmetric nonnegative rank splitting (MSNRS) in a similar way as the MNRS in Section 4, which we shall skip the details here.

6. The geometric meaning of nonnegative rank. We briefly outline a geometric meaning of the nonnegative rank, relate it to the classical Sylvester's problem, and draw an interesting conclusion on the conditional probability of the NRF in this section.

Let the columns of a given nonnegative matrix $A \in \mathbb{R}^{m \times n}$ be denoted by $A = [\mathbf{a}_1, \ldots, \mathbf{a}_n] \in \mathbb{R}^{m \times n}$. Define the *scaling factor* $\sigma(A)$ by

$$\sigma(A) := \operatorname{diag} \{ \|\mathbf{a}_1\|_1, \ldots, \|\mathbf{a}_n\|_1 \}, \tag{6.1}$$

where $\|\cdot\|_1$ stands for the 1-norm of a vector, and the *pullback map* $\vartheta(A)$ by

$$\vartheta(A) := A \sigma(A)^{-1}. \tag{6.2}$$

Each column of $\vartheta(A)$ can be regarded as a point on the $(m-1)$-dimensional *probability simplex* $\mathcal{D}_m$ defined by

$$\mathcal{D}_m := \left\{ \mathbf{a} \in \mathbb{R}^m | \mathbf{a} \geq 0, \mathbf{1}_m^\top \mathbf{a} = 1 \right\}, \tag{6.3}$$

where $\mathbf{1}_m = [1, \ldots, 1]^\top$ stands for the vector of all 1's in $\mathbb{R}^m$.

Suppose a given A can be factorized as the product of two nonnegative matrices $A = UV$, with $U \in \mathbb{R}^{m \times p}$ and $V \in \mathbb{R}^{p \times n}$. Because $UV = (UD)(D^{-1}V)$ for any invertible nonnegative matrix $D \in \mathbb{R}^{p \times p}$, we may assume without loss of generality that U is already a pullback so that $\sigma(U) = I_n$. We can write

$$A = \vartheta(A)\sigma(A) = UV = \vartheta(U)\vartheta(V)\sigma(V). \tag{6.4}$$

Note that the product $\vartheta(U)\vartheta(V)$ itself is on the simplex $\mathcal{D}_m$. It follows that

$$\vartheta(A) = \vartheta(U)\vartheta(V), \tag{6.5}$$

$$\sigma(A) = \sigma(V). \tag{6.6}$$

In particular, if $p = \operatorname{rank}_+(A)$, then we see that $\operatorname{rank}_+(\vartheta(A)) = p$, and vice versa. The expression (6.5) enjoys an interesting geometry interpretation. It means that the columns in the pullback $\vartheta(A)$ are convex combinations of columns of $\vartheta(U)$. The nonnegative rank $\operatorname{rank}_+(\vartheta(U))$ stands for the minimal number of vertices on $\mathcal{D}_m$ so that the resulting convex polytope encloses all columns of $\vartheta(A)$. A nonnegative matrix A has an NRF means that the minimal convex polytope enclosing $\vartheta(A)$ has exactly $\operatorname{rank}(A)$ many vertices.

Using this notion, it is insightful to explain by geometry why the matrix $\mathscr{C}$ in (1.3) does not have an NRF. Since each point $\mathbf{a} \in \mathcal{D}_4$ satisfies $a_1 + a_2 + a_3 + a_4 = 1$, it suffices to represent the 4-dimensional vector $\mathbf{a}$ by the vector $[a_1, a_2, a_3]$ of its first three entries. In this way, the probability simplex $\mathcal{D}_4$ can easily be visualized via the unit tetrahedron S in the first octant of $\mathbb{R}^3$. Specifically, columns of $\vartheta(\mathscr{C})$ can be interpreted as points A_1, A_2, A_3, A_4 depicted in Figure 6.1. Note that the

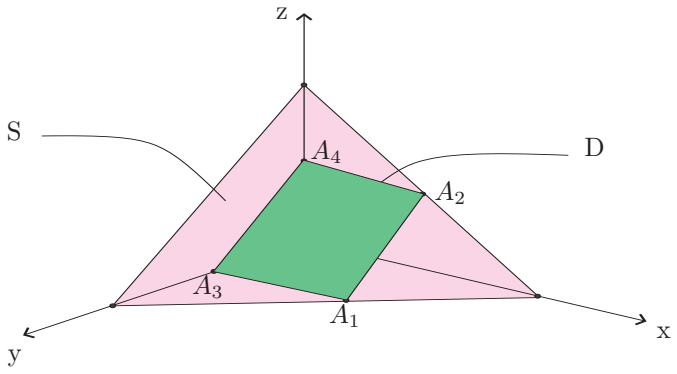


FIG. 6.1. *A geometric representation of the matrix $\vartheta(\mathscr{C})$.*

four points A_1, A_2, A_3, A_4 are coplanar because $\text{rank}(\vartheta(\mathscr{C})) = 3$. On the other hand, the convex hull D of these four points sits on four separate "ridges" of the tetrahedron, which cannot be enclosed by any three-vertex convex set in the tetrahedron. The minimum number of vertices for a convex set in the unit tetrahedron to cover D is four, hence $\text{rank}_+(A) = 4$.

Apparently, the columns of $\mathscr{C}$ occupy very specific positions in the unit tetrahedron. An interesting question to ask is how often this can happen.

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

For a matrix to be in the set

$$\mathfrak{E} := \left\{ A \in R_+^{4\times 4} \mid \text{rank}(A) = 3,\ A = \vartheta(A) \right\},$$

it is necessary that the four columns formed by deleting its last row are in a plane that intersects the unit tetrahedron. We consider two mutually exclusive cases. First, if the cross-section of the plane in the unit tetrahedron is a triangle, then this triangle naturally encloses the four points and the matrix has nonnegative rank 3. What is then the probability of a randomly selected plane in the unit tetrahedron having a triangular cross-section? Secondly, if the cross-section is a quadrilateral, then what is the probability that the four columns of A, identified as four points in the unit tetrahedron, are enclosed in a triangle within the quadrilateral?

There is no easy answer to the above questions. In particular, the first question touches upon the basis of geometric probability where we must reconcile first how the plane cuts through the unit tetrahedron [32]. Different definitions of randomness will lead to different answers. Likewise, the second question appears to be a generalization of the well known Sylvester's four-point problem which has a number of different answers, prompting Sylvester to conclude "This problem does not admit of a determinate solution" [42]. The Sylvester's question asks the probability $p(4, K)$ of four random, independent, and uniform points from a compact set K being such that none of them lies in the triangle formed by the other three [1]. The conditional probability of the event that $\text{rank}_+(A) = 3$, given $A \in \mathfrak{E}$ is in a quadrilateral K, therefore is greater than or equal to $1 - p(4, K)$. The answer to the problem $R2R_+$ is the sum of the probabilities of the two mutually exclusive cross-section cases. Still we do not have a definitive answer to the problem $R2R_+$ at present.

On the other hand, we may also ask the following question which is the other side of the coin:

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

It turns out we have an easy answer even for the general cases.

THEOREM 6.1. *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* $\operatorname{rank}(A) = k$ *with A restricted to $R_+(k)$ is one.*

Proof. Without loss of generality, we may assume $A = \vartheta(A)$. The fact that $\operatorname{rank}_+(A) = k$ means k is the minimum number of vertices for a convex hull on D_m to enclose columns of A. On the other hand, using the fact that columns of A must satisfy $m - \operatorname{rank}(A) + 1$ linear equations, these vertices should reside on an affine subspace of dimension $\operatorname{rank}(A) - 1$. The probability of k distinct points in $\mathbf{R}^m$ to be in an affine subspace of dimension strictly less than $k - 1$ is zero. Hence if $\operatorname{rank}_+(A) = k$, then with probability one we have $\operatorname{rank}(A) = k$. □

7. Numerical experiments. In general, a nonnegative matrix may not have an NRF, toward which we have already exemplified in Section 4 how the MNRS can be calculated. In this section, we demonstrate the working of our algorithm by computing the NRF of several nontrivial matrices. To ensure the existence of an NRF, we rely on Theorem 6.1 to generate test data.

Example 3. We randomly generate two nonnegative matrices $W \in \mathbb{R}^{5\times 3}$ and $H \in \mathbb{R}^{3\times 5}$

$$W = \begin{pmatrix} 0.9708 & 0.2140 & 0.4120 \\ 0.9901 & 0.6435 & 0.7446 \\ 0.7889 & 0.3200 & 0.2679 \\ 0.4387 & 0.9601 & 0.4399 \\ 0.4983 & 0.7266 & 0.9334 \end{pmatrix}, \quad H = \begin{pmatrix} 0.6833 & 0.2071 & 0.4514 \\ 0.2126 & 0.6072 & 0.0439 \\ 0.8392 & 0.6299 & 0.0272 \\ 0.6288 & 0.3705 & 0.3127 \\ 0.1338 & 0.5751 & 0.0129 \end{pmatrix}^\top,$$

and define $A = WH$. Obviously, the matrix A has an NRF by construction. Applying our method, we find a new NRF of the matrix $A = UV$ with

$$U = \begin{pmatrix} 0.02556251462152 & 0.00563995049828 & 0.05020141897100 \\ 0.02170073447446 & 0.00833461311952 & 0.10441556394589 \\ 0.01809150541560 & 0.01018126666230 & 0.05049608516928 \\ 0 & 0.01848758106428 & 0.11124821245836 \\ 0.00844625518446 & 0 & 0.11504734974715 \end{pmatrix}$$

and

$$V = \begin{pmatrix} 22.65196335171072 & 0 & 6.26680882859984 \\ 2.42501879200814 & 7.93355352839419 & 4.93382971188032 \\ 21.86596055059507 & 15.79385557484764 & 6.22840681607381 \\ 18.88971723342108 & 4.21210726740658 & 6.21364289968766 \\ 0 & 7.37487007847031 & 4.31631898597720 \end{pmatrix}^\top,$$

where, for convenience, we report only 14 decimals and entries with magnitude less than 10^{-10} are regarded as numerical zero.

Example 4. To demonstrate the NRF of a CP matrix, we randomly generate a nonnegative matrix W

$$W = \begin{pmatrix} 0.3840 & 0.0158 & 0.6315 & 0.3533 \\ 0.6831 & 0.0164 & 0.7176 & 0.1536 \\ 0.0928 & 0.1901 & 0.6927 & 0.6756 \\ 0.0353 & 0.5869 & 0.0841 & 0.6992 \\ 0.6124 & 0.0576 & 0.4544 & 0.7275 \\ 0.6085 & 0.3676 & 0.4418 & 0.4784 \end{pmatrix},$$

and define $A = WW^\top$. Note that by construction, A is CP. Our algorithm shows that the matrix A has a nonnegative decomposition $A = BB^\top$ with

$$B = \begin{pmatrix} 0.58354630329629 & 0.35203758768455 & 0.44908808470853 & 0.07198556063105 \\ 0.58153426004785 & 0.21962336143449 & 0.78677873223851 & 0 \\ 0.67910036260539 & 0.70358647639519 & 0.14987548760103 & 0.04842113133372 \\ 0 & 0.91741547294584 & 0 & 0 \\ 0.45303579486590 & 0.65652500722113 & 0.57125648786549 & 0.38921284477200 \\ 0.28134324531544 & 0.66368746544619 & 0.63911349062806 & 0.03680601340055 \end{pmatrix}.$$

Example 5. For a general matrix R with negative entries, it is not clear that the product $RR^\top$, even if it is nonnegative, has an NRF. If R is a Soules matrix, nonetheless, then it is known the product $RDR^\top$ is nonnegative and $\text{rank}_{cp}(RDR^\top) = \text{rank}(RDR^\top)$ for every nonnegative diagonal matrix D with nonincreasing diagonal element [40].

The matrix

$$R = \begin{pmatrix} 0.1348 & 0.1231 & 0.1952 & 0.3586 & 0.8944 \\ 0.2697 & 0.2462 & 0.3904 & 0.7171 & -0.4472 \\ 0.4045 & 0.3693 & 0.5855 & -0.5976 & 0 \\ 0.5394 & 0.4924 & -0.6831 & 0 & 0 \\ 0.6742 & -0.7385 & 0 & 0 & 0 \end{pmatrix}$$

is a Soules matrix [21]. With this R and with

$$D = \text{diag}([0.7, 0.5, 0.4, 0, 0]),$$

we define

$$A = RDR^\top = \begin{pmatrix} 0.035537749 & 0.071084934 & 0.106614875 & 0.027868556 & 0.018162837 \\ 0.071084934 & 0.142188747 & 0.213258065 & 0.055774870 & 0.036372868 \\ 0.106614875 & 0.213258065 & 0.319849520 & 0.083670750 & 0.054535705 \\ 0.027868556 & 0.055774870 & 0.083670750 & 0.511545776 & 0.072745736 \\ 0.018162837 & 0.036372868 & 0.054535705 & 0.072745736 & 0.590873073 \end{pmatrix}.$$

Because of the two zeros in D, we know $\text{rank}(A) = 3$. By this construction, A is a nontrivial, rank deficient, positive semi-definite matrix which, by theory, is also a CP matrix. Applying our method, we find $\text{rank}_{cp}(A) = 3$ as is expected and obtain a decomposition $A = BB^\top$ with

$$B = \begin{pmatrix} 0.18851458564260 & 0 & 0 \\ 0.37707922576755 & 0.00005556024891 & 0.00003751809670 \\ 0.56555239286434 & 0.00006079293882 & 0.00008502728571 \\ 0.14783235952195 & 0.07671246735655 & 0.69556205102811 \\ 0.09634711785325 & 0.76262068283226 & 0 \end{pmatrix}.$$

8. Conclusion. Nonnegative matrix factorization has many important applications. Nonnegative rank factorization is fundamental to nonnegative matrix factorization in that the former seeks a completely factorization whereas the latter seeks only a partial approximation. Detecting the nonnegative rank and computing the nonnegative rank factorization of a nonnegative matrix are theoretical challenging in many aspects. In our experience, no existing nonnegative matrix factorization algorithms can guarantee to find the nonnegative rank factorization of a nonnegative matrix.

In this paper, we present a numerical method to determine whether the nonnegative rank of a nonnegative matrix agrees with its rank. In the case that a nonnegative rank factorization does exist, our algorithm is able to compute this decomposition; otherwise, our algorithm is able to find its maximum nonnegative component where such a factorization exists. Our method also works for a completely positive matrix, provided its cp-rank and rank are the same.

The crux of our algorithm is the employment of the Wedderburn rank reduction formula. At each step, we look for a nonnegative element that not only can reduce the rank, but also maintain the nonnegativity for the residual. The nature of rank reduction forces the process to terminate in finitely many steps. In order to realize the multiple constraints of nonnegativity, we transform the problem of nonnegative rank factorization to the problem of an inequality system which is further transformed to a maximin problem. Plenty optimization techniques are available for solving the maximin problem numerically. Extensive numerical experiments seem to evidence that our approach works reasonably well.

REFERENCES

[1] I. Bárány, *Sylvester's question: the probability that n points are in convex position*, Ann. Probab., 27 (1999), pp. 2020–2034.

[2] F. Barioli and A. Berman, *The maximal cp-rank of rank k completely positive matrices*, Linear Algebra Appl., 363 (2003), pp. 17–33. Special issue on nonnegative matrices, M-matrices and their generalizations (Oberwolfach, 2000).

[3] A. Berman, *Nonnegative matrices—old problems, new applications*, in Positive systems, vol. 341 of Lecture Notes in Control and Inform. Sci., Springer, Berlin, 2006, pp. 1–9.

[4] A. Berman and R. J. Plemmons, *Nonnegative matrices in the mathematical sciences*, vol. 9 of Classics in Applied Mathematics, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1994. Revised reprint of the 1979 original.

[5] A. Berman and N. Shaked-Monderer, *Completely positive matrices*, World Scientific Publishing Co. Inc., River Edge, NJ, 2003.

[6] M. W. Berry, M. Browne, A. N. Langville, P. V. Pauca, and R. J. Plemmons, *Algorithms and applications for approximate nonnegative matrix factorization*, Computational Statistics and Data Analysis, 52 (2007), pp. 155–173.

[7] D. P. Bertsekas, *Nonlinear programming (2nd)*, Belmont, MA: Athena Scientific., 1999.

[8] R. Bro and S. D. Jong, *A fast non-negativity-constrained least squares algorithm*, Journal of Chemometrics, 11 (1997), pp. 393–401.

[9] S. L. Campbell and G. D. Poole, *Computing nonnegative rank factorizations*, Linear Algebra Appl., 35 (1981), pp. 175–182.

[10] Z. Chen, A. Cichocki, and T. M. Rutkowski, *Constrained non-negative matrix factorization method for eeg analysis in early detection of alzheimers disease*, in In IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP2006, Toulouse, France, 2006.

[11] M. Chu, F. Diele, R. Plemmons, and S. Ragni, *Optimality, computation and interpretation of nonnegative matrix factorizations*, Available online at http://www4.ncsu.edu/ mtchu/Research/Papers/nnmf.ps, (2005).

[12] M. T. Chu, R. E. Funderlic, and G. H. Golub, *A rank-one reduction formula and its applications to matrix factorizations*, SIAM Rev., 37 (1995), pp. 512–530.

[13] M. T. Chu, R. E. Funderlic, and G. H. Golub, *Rank modifications of semidefinite matrices associated with a secant update formula*, SIAM J. Matrix Anal. Appl., 20 (1999), pp. 428–436 (electronic).

[14] M. T. Chu and M. M. Lin, *Low-dimensional polytope approximation and its applications to nonnegative matrix factorization*, SIAM J. Sci. Comput., 30 (2008), pp. 1131–1155.

[15] R. E. Cline and R. E. Funderlic, *The rank of a difference of matrices and associated generalized inverses*, Linear Algebra Appl., 24 (1979), pp. 185–215.

[16] J. E. Cohen and U. G. Rothblum, *Nonnegative ranks, decompositions, and factorizations of nonnegative matrices*, Linear Algebra Appl., 190 (1993), pp. 149–168.

[17] C. Ding, X. He, and H. Simon, *On the equivalence of nonnegative matrix factorization and spectral clustering*, in In: Proceedings of the Fifth SIAM International Conference on Data Mining, Newport Beach, CA, 2005.

[18] D. Donoho and V. Stodden, *When does nonnegative matrix factorization give a correct decomposition into parts?*, in Proc. 17th Ann. Conf. Neural Information Processing Systems, NIPS, Stanford University, Stanford, CA, 2003, 2003.

[19] M. Dugas, S. Merk, S. Breit, and P. Dirschedl, *Mdclust-exploratory microarray analysis by multidimensional clustering*, Bioinformatics, 20 (2004), pp. 931–936.

[20] E. Egerváry, *On rank-diminishing operators and their applications to the solution of linear equations*, Z. Angew. Math. Phys., 11 (1960), pp. 376–386.

[21] L. Elsner, R. Nabben, and M. Neumann, *Orthogonal bases that lead to symmetric nonnegative matrices*, Linear Algebra Appl., 271 (1998), pp. 323–343.

[22] J. Hannah and T. J. Laffey, *Nonnegative factorization of completely positive matrices*, Linear Algebra Appl., 55 (1983), pp. 1–9.

[23] P. K. Hopke, *Receptor Modeling for Air Quality Management*, Elsevier, Amsterdam, 1991.

[24] A. S. Householder, *The theory of matrices in numerical analysis*, Dover Publications Inc., New York, 1975. Reprint of 1964 edition.

[25] P. O. Hoyer, *Nonnegative sparse coding*, in Proc. IEEE Workshop Neural Networks for Signal Processing, Martigny, 2002.

[26] ———, *Non-negative matrix factorization with sparseness constraints*, The Journal of Machine Learning Research, 5 (2004), pp. 1457–1469.

[27] S. K. Jain and J. Tynan, *Nonnegative rank factorization of a nonnegative matrix A with* $A^{\dagger}A \geq 0$, Linear Multilinear Algebra, 51 (2003), pp. 83–95.

[28] M. W. Jeter and W. C. Pye, *A note on nonnegative rank factorizations*, Linear Algebra Appl., 38 (1981), pp. 171–173.

[29] ———, *Some nonnegative matrices without nonnegative rank factorizations*, Indust. Math., 32 (1982), pp. 37–41.

[30] T. Kawamoto, K. Hotta, T. Mishima, J. Fujiki, M. Tanaka, and T. Kurita, *Estimation of single tones from chord sounds using non-negative matrix factorization*, Neural Network World, 3 (2000), pp. 429–436.

[31] H. Kim and H. Park, *Non-negative matrix factorization based on alternating non-negativity constrained least squares and activity set method*, Technical report GT-CSE-07-01, George Institute of Technology, 2007; SIAM J. Matrix Anal. Appl., to appear.
[32] D. A. Klain and G.-C. Rota, *Introduction to geometric probability*, Lezioni Lincee. [Lincei Lectures], Cambridge University Press, Cambridge, 1997.
[33] C. L. Lawson and R. J. Hanson, *Solving least squares problems*, vol. 15 of Classics in Applied Mathematics, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1995. Revised reprint of the 1974 original.
[34] D. D. Lee and H. S. Seung, *Learning the parts of objects by non-negative matrix factorization*, Nature, 401 (1999), p. 788C791.
[35] ———, *Algorithms for non-negative matrix factorization*, in In Advances in Neural Information Processing Systems, vol. 13, 2001, pp. 556–562.
[36] B. Levin, *On calculating maximum rank one underapproximations for positive arrays*, Columbia University Biostatistics Technical Report Series. Working Paper 10. http://biostats.bepress.com/columbiabiostat/papers/art10, (1985).
[37] P. Paatero, *The multilinear engine–a table-driven, least squares program for solving multilinear problems, including the n-way parallel factor analysis model.*, J. of Computational and Graphical Statistics, 8 (1999), pp. 854–888.
[38] P. Paatero and U. Tapper, *Positive matrix factorization: A non-negative factor model with optimal utilization of error*, Environmetrics, 5 (1994), pp. 111–126.
[39] N. Shaked-Monderer, *Minimal cp-matrices*, ELA, 8 (2001), pp. 140–157.
[40] ———, *A note on the cp-rank of matrices generated by a soules matrix*, ELA, 12 (2004), pp. 2–5.
[41] S. Sra and I. S. Dhillon, *Nonnegative matrix approximation: Algorithms and applications*, Technical report, Deptartment of Computer Sciences, University of Texas at Austin Austin, TX 78712, USA (2006). Available at http://www.cs.utexas.edu/ftp/pub/techreports/tr06-27.pdf.
[42] J. J. Sylvester, *On a special class of questions on the theory of probabilities*, Birmingham British Assoc. Rept., (1865), pp. 8–9.
[43] L. B. Thomas, *Solution to problem 73-14: Rank factorization of nonnegative matrices by a. berman and r. j. plemmons*, SIAM Review, 16 (1974), pp. 393–394.
[44] J. M. van den Hof, *Realization of positive linear systems*, Linear Algebra Appl., 256 (1997), pp. 287–308.
[45] S. A. Vavasis, *On the complexity of nonnegative matrix factorization*, (2007). Available at http://www.citebase.org/abstract?id=oai:arXiv.org:0708.4149.
[46] J. H. M. Wedderburn, *Lectures on matrices*, Dover Publications Inc., New York, 1964.
[47] W. Xu, X. Liu, and Y. Gong, *Document-clustering based on non-negative matrix factorization*, in In: Proceedings of SIGIR03, Toronto, CA, July 28–August 1 2003, pp. 267–273.