NONNEGATIVE RANK FACTORIZATION VIA RANK REDUCTION

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

Abstract. Any given nonnegative matrix $A \in \mathbb{R}^{m\times n}$ can be expressed as the product $A = UV$ for some nonnegative matrices $U \in \mathbb{R}^{m\times k}$ and $V \in \mathbb{R}^{k\times n}$ with $k \leq \min\{m, n\}$. The smallest k that makes this factorization possible is called the nonnegative rank of A. Computing the exact nonnegative rank and the corresponding factorization are known to be NP-hard. Even if the nonnegative rank is known a priori, no simple numerical procedure exists that can calculate the nonnegative factorization. This paper is the first to describe a heuristic approach to tackle this difficult problem. Based on the Wedderburn rank reduction formula, the idea is to recurrently extract, whenever possible, a rank-one nonnegative portion from the previous matrix, starting with A, while keeping the residual nonnegative and lowering its rank by one. With a slight modification for symmetry, the method can equally be applied to another important class of completely positive matrices. Numerical testing seems to suggest that the proposed algorithm, though still lacking in rigorous error analysis, might serve as an initial numerical 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. Let $\mathbb{R}_+$ stand for the half-array of nonnegative real numbers. Given any matrix $A \in \mathbb{R}_+^{m\times n}$, it is always possible to express A as the sum of a series of nonnegative rank one matrices. Among the many possible series representations of A by nonnegative rank one matrices, the number of terms that render the shortest nonnegative rank one series representation is attained is called the *nonnegative rank* of the matrix A. For convenience, we denote the nonnegative rank of A by $\text{rank}_+(A)$. It is known that the nonnegative rank has upper and lower bounds such as [14]

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

To determine the exact nonnegative rank for a matrix, however, is known to be NP-hard [40].

The series representation can be rewritten as the product $A = UV$ which obviously is in a nonnegative factorization form of A. Such a complete factorization, with its product equal to A, should be distinguished from what is known as the nonnegative matrix factorization (NMF) which has attracted much attention in the literature [3, 12, 15, 18, 20, 26, 33, 34, 37]. The notion of NMF is a low rank approximation formulated as the minimization problem

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

where $p < \min\{m, n\}$ is a preselected integer. Many specially developed numerical techniques are available for NMF, including the multiplicative update algorithm [30, 31], the gradient methods [9, 21], and alternating least square approaches [6, 7, 27, 29, 33]. These NMF techniques, mostly utilizing the notion of mathematical programming, cannot guarantee the required equality in a complete factorization. The reason is that, the objective function in (1.1) being non-convex, the factors U and V acquired by almost all NMF techniques even with $p = \text{rank}_+(A)$ are local minimizers and their products typically will not equal to A.

*Department of Applied Mathematics, Dalian University of Technology, Dalian, Liaoning, 116024. (dongbodlut@gmail.com). This work is partially supported by Chinese Scholarship Council and DLUT Grants 3004-842321 and 3004-893327.

†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.

In this paper, our focus is on a special subclass of nonnegative matrices,

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

We are interested in procuring a nonnegative factorization for $A \in \mathfrak{R}(m,n)$ which is both complete in the sense that $A = UV$ and minimal in the sense that $U \in \mathbb{R}_+^{m\times \operatorname{rank}(A)}$ and $V \in \mathbb{R}_+^{\operatorname{rank}(A)\times n}$. Such a factorization is called a *nonnegative rank factorization* (NRF) of A.

Every nonnegative matrix has a nonnegative factorization, but not every nonnegative matrix has an NRF [14]. Quite a few nonnegative matrices without NRF have been constructed in [25]. The simplest example is the 4×4 matrix

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

with $\operatorname{rank}(\mathscr{C}) = 3$ and $\operatorname{rank}_+(\mathscr{C}) = 4$.

Given a matrix $A \in \mathbb{R}_+^{m\times n}$, the question then arises about how to determine if A is in $\mathfrak{R}(m,n)$. A necessary and sufficient condition qualifying whether a nonnegative matrix has an NRF is given in [39], but the algebraic manipulations involved in checking out this condition for a given matrix A are onerous. Other known sufficient conditions for the existence of an NRF are for more restrictive subclasses of matrices such as the so called weakly monotone nonnegative matrices [24], λ-monotone [23], or matrices with nonnegative 1-inverse [8]. Our first contribution in this paper, which we think is new in theory, is to show that matrices in $\mathfrak{R}(m,n)$ are generic in the sense that those matrices whose rank is strictly less than its nonnegative rank form a set of measure zero. Even if the NRF is known to exist, the next question is how to actually compute the two factors in the NRF. The main thrust in this paper is to propose a general procedure that computes the NRF numerically, if it ever exists. We hasten to point out that an error analysis seems to be the norm in proposing any new algorithm, including our NRF computation. Nevertheless, the existence theory of the NRF is already scarcely understood in the literature, not to mention its perturbation analysis which is far beyond current research endeavor. Our goal at present is simply to develop a heuristic algorithm that, albeit its crudeness, might help as a tool to grasp more understanding of the difficult subject on nonnegative factorization. To the best of our knowledge, our numerical procedure is the first algorithm ever proposed to detect whether an NRF exists and to compute the NRF, if it does exist.

This paper is organized as follows. We begin in Section 2 by sketching a geometric meaning of the nonnegative rank by which we relate the problem of NRF to the classical Sylvester's problem. We then develop an interesting theoretical result on existence via the probability standpoint. In Section 3 we briefly review the critical Wedderburn formula for the purpose of exploiting its rank reduction property. To satisfy the nonnegativity constraints in the rank reduction process, we transform in Section 4 the problem of NRF to a sequence of maximin problems which can then be solved by any available optimization techniques. With a slight modification for symmetry, we apply in Section 5 our algorithm to the so called completely positive matrices. Again, we think our approach is new in this area. In Section 6, we raise an interesting question about the maximal nonnegative rank splitting of a nonnegative matrix when an NRF does not exist. Some interesting examples are included throughout the paper to enlighten the working of our algorithm.

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

$$\sigma(A) := \operatorname{diag}\left\{ \|\mathbf{a}_1\|_1, \ldots, \|\mathbf{a}_n\|_1 \right\}, \tag{2.1}$$

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

$$\vartheta(A) := A\sigma(A)^{-1}. \tag{2.2}$$

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

$$\mathcal{D}_m := \left\{ \mathbf{a} \in \mathbb{R}_+^m | \mathbf{1}_m^\top \mathbf{a} = 1 \right\}, \tag{2.3}$$

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

Suppose a given nonnegative matrix A can be factorized as $A = UV$, where $U \in \mathbb{R}_+^{m \times p}$ 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

$$A = \vartheta(A)\sigma(A) = UV = \vartheta(U)\vartheta(V)\sigma(V). \tag{2.4}$$

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

$$\vartheta(A) = \vartheta(U)\vartheta(V), \tag{2.5}$$
$$\sigma(A) = \sigma(V). \tag{2.6}$$

In particular, if $p = \text{rank}_+(A)$, then we see that $\text{rank}_+(\vartheta(A)) = p$, and vice versa. The expression (2.5) means that the columns in the pullback $\vartheta(A)$ are convex combinations of columns of $\vartheta(U)$. We thus obtain an interesting geometry interpretation of nonnegative rank.

LEMMA 2.1. *The nonnegative rank* $\text{rank}_+(A)$ *stands for the minimal number of vertices on* $\mathcal{D}_m$ *so that the resulting convex polytope encloses all columns of the pullback* $\vartheta(A)$.

A nonnegative matrix A has an NRF means that the minimal convex polytope enclosing $\vartheta(A)$ has exactly $\text{rank}(A)$ 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 $\mathbf{a} = [a_1, a_2, a_3, a_4]^\top \in \mathcal{D}_4$ satisfies $a_1 + a_2 + a_3 + a_4 = 1$, it suffices to represent the 4-dimensional vector $\mathbf{a}$ by the vector $[a_1, a_2, a_3]^\top$ 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 2.1. Note that the four points A_1, A_2, A_3, A_4 are coplanar

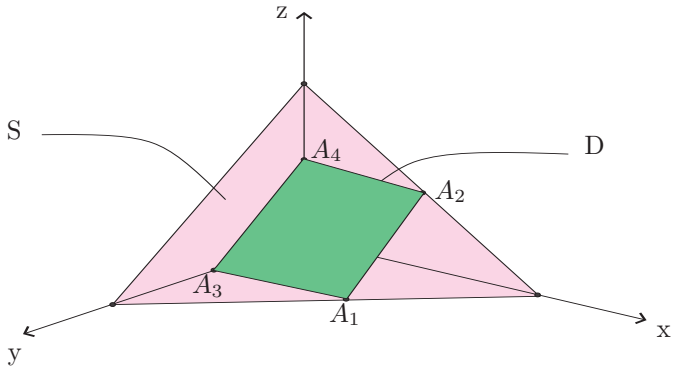

FIG. 2.1. *A geometric representation of the matrix* $\vartheta(\mathscr{C})$.

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 $\text{rank}_+(\mathscr{C}) = 4$. The point to make is that the columns of $\mathscr{C}$ are at very strategic 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?

To answer $R2R_+$, it suffices to consider matrices in the set

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

It is clear that $A \in \mathfrak{E}$ only if the four points obtained from columns by deleting the last row of A are in a plane that intersects the unit tetrahedron. There are two mutually exclusive cases. Firstly, the cross-section of the plane in the unit tetrahedron is a triangle. Naturally this triangle encloses the four points and the matrix has nonnegative rank 3. What then is the probability of a randomly selected plane in the unit tetrahedron to have a triangular cross-section? Secondly, the cross-section is a quadrilateral. What then is the probability that the four points representing A 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 [28]. Different definitions of randomness will lead to different answers. The second question appears to be a generalization of the well known Sylvester's four-point problem which asks for the probability, denoted by $p(4, K)$, of four random, independent, and uniform points from a compact set K such that none of them lies in the triangle formed by the other three [1]. The conditional probability of the event that $\operatorname{rank}_+(A) = 3$, given $A \in \mathfrak{E}$ is in a quadrilateral K, therefore is greater than or equal to $1 - p(4, K)$. However, it is a well known story that $p(4, K)$ itself has a number of different answers, prompting Sylvester to exclaim that [38], "This problem does not admit of a determinate solution!" At present, we do not have a definitive answer to the seemingly simple problem $R2R_+$.

The flipped side question of $R2R_+$ is also interesting.

($\mathbf{R_+2R}$) : 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. The following result shows that matrices which have an NRF are generic. It is based on this insight that we are able to generate many of our test problems in the subsequent discussion. We think this result is important and new.

THEOREM 2.2. *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* $\operatorname{rank}(A) = k$, *given $A \in R_+(k)$, is one.*

Proof. Without loss of generality, we may assume $A = \vartheta(A)$. The fact $\operatorname{rank}_+(A) = k$ means that k is the minimum number of vertices for a convex hull on D_m to enclose columns of A. On the other hand, the subspace orthogonal to columns of A is of dimension $m - \operatorname{rank}(A)$ while the column sum for A is always 1. Together, columns of A satisfy $m - \operatorname{rank}(A) + 1$ independent linear equations. Thus these vertices should reside on an affine subspace of dimension $\operatorname{rank}(A) - 1$. The probability of k distinct points in $\mathbb{R}^m$ 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 $\operatorname{rank}(A) = k$. □

It will be convenient to introduce two basic terms concerning a rank-one matrix. Firstly, any nonnegative matrix whose subtraction from a given nonnegative matrix remains nonnegative is called a *nonnegative component* (NC) of that given matrix. In [32], Levin described an algorithm for computing the "maximum" rank-one NC of a given nonnegative matrix. The trouble is that the residual after a rank-one NC subtraction might increase or maintain the rank as the original matrix. If the process is to be repeated, we might end up with an infinite series of nonnegative rank one matrices, which defies the minimal length of what an NRF desires. Levin's algorithm cannot be applied to compute the NRF of a given matrix in $\mathfrak{R}(m, n)$.

Secondly, any rank-one NC of a nonnegative matrix A such that the rank of the residual is one less than $\operatorname{rank}(A)$ is called a *nonnegative element* (NE) of A. The fundamental difference between an NC and an NE is significant. Unlike Levin's method, we are not interested in a maximum NC which typically will not reduce the rank. Rather, we gradually distribute A over a sequence of NEs each of which is an NC, but will also reduce the rank by one. Since the rank is reduced by one in

each step, the NRF of a matrix in $\mathfrak{R}(m,n)$ should be found by at most $\operatorname{rank}(A)$ many iterations in exact arithmetic. The principal tool used in our search for an NE is the Wedderburn rank reduction formula [10, 41] which will be reviewed in the next section.

Example 1. The matrix $\mathscr{C}$ defined in (1.3) has many NCs, but has no NE at all. Indeed, 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 residual matrix A_2 after its removal from A_1 would be nonnegative and of rank 2. But then by [14, 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.

3. Wedderburn rank reduction formula. The Wedderburn rank reduction formula appeared as a modest statement in Wedderburn's 1934 book [41] and as an exercise in Householder's 1964 book [19]. In the review paper [10], however, it was pointed out that almost all known matrix factorizations can be achieved by this seemingly straightforward expression. An interesting time line chronicling the appearance of the rank reduction results in the numerical linear algebra as well as the applied statistics and psychometrics literature can be found in [22, Figure 2.1]. For completion, we state two main theorems below.

The first formula concerns a necessary and sufficient condition for rank subtraction by a rank-one matrix. This result is of fundamental importance in our discussion.

THEOREM 3.1. *Let* $\mathbf{u} \in \mathbb{R}^m$ *and* $\mathbf{v} \in \mathbb{R}^n$. *Then the matrix*

$$B := A - \sigma^{-1}\mathbf{u}\mathbf{v}^\top \tag{3.1}$$

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*

$$\mathbf{u} = A\mathbf{x}, \quad \mathbf{v} = A^\top \mathbf{y}, \quad \sigma = \mathbf{y}^\top A\mathbf{x}. \tag{3.2}$$

Indeed, simultaneous multiple rank reduction is possible. Cline and Funderlic proved that a generalized formula holds in the following sense [13].

THEOREM 3.2. *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*

$$U = AX, \quad V = A^\top Y, \quad \text{and} \quad R = Y^\top AX. \tag{3.3}$$

The formula (3.1) provides a mechanism to break down a matrix into a sum of rank-one matrices. The basic idea is that, starting with $A_1 = A$, we define a sequence $\{A_k\}$ of matrices by defining

$$A_{k+1} := A_k - (\mathbf{y}_k^\top A_k \mathbf{x}_k)^{-1} A_k \mathbf{x}_k \mathbf{y}_k^\top A_k. \tag{3.4}$$

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 process can be continued so long as $A_k \neq 0$. Since $\operatorname{rank}(A_k)$ is reduced by one at each step, the sequence $\{A_k\}$ must be finite. In this way, the matrix A is decomposed as a finite series of rank one matrices. A detailed discussion in this regard can be found in [10].

For our application, we want to break a nonnegative matrix down by taking away one NE a time. For a rank-one matrix to be an NE, it must assume the Wedderburn form $(\mathbf{y}_k^\top A_k \mathbf{x}_k)^{-1} A_k \mathbf{x}_k \mathbf{y}_k^\top A_k$ which is nonnegative for some $\mathbf{x}_k \in \mathbb{R}^n$ and $\mathbf{y}_k \in \mathbb{R}^m$. It also needs to ensure $A_{k+1} \geq 0$. When these nonnegativity constraints are satisfied, the rank reduction mechanism kicks in. If we can repeat the process until the rank is reduced to zero (however, see Example 2), then an NRF is thus found. Note that if A_k is nonnegative, then nonnegative rank-one matrices in the Wedderburn form are easy to

find. If the process ever terminates prematurely, it is only because it is not possible to keep the residual A_{k+1} nonnegative by *any* nonnegative rank-one reduction. There are two probable causes for this to happen. One is that the matrix A is not a matrix in $\mathfrak{R}(m,n)$ to begin with. This would be a welcome conclusion which leads to the notion of maximal nonnegative rank splitting of A. The other is that some bad starting points have branched A_k into a "dead end" through the iteration. A restart might remedy the problem. In either cause, the point is that when our method ceases to iterate before the rank is diminished to zero, care must be taken to conclude whether A has an NRF.

4. NRF by Wedderburn formula. In this section, we give specifics about the application of the Wedderburn rank reduction formula to the NRF. We first characterize the nonnegativity constraints by which the NRF problem is cast as a maximin problem. Then we consider the prospect of multi-dimensional reduction.

Trivially, 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 scalar $\mathbf{y}_k^\top A_k \mathbf{x}_k$ in the Wedderburn formula (3.4) must be either positive or negative, we can assume without loss of generality that $A_k\mathbf{x}_k \geq 0$, $\mathbf{y}_k^\top A_k \geq 0$, and $\mathbf{y}_k^\top A_k \mathbf{x}_k > 0$. We assume further after scaling that $\mathbf{y}_k^\top A_k \mathbf{x}_k = 1$. These conditions constitute our "*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.

Intrinsic in the NRF is the nature of "additivity" by nonnegative rank-one matrices. We cannot immediately regard a nonnegative $A_k\mathbf{x}_k\mathbf{y}_k^\top A_k$ as an NE of A_k because its subtraction might over deduct the matrix A_k to have negative entries. To provide a safety guard against this over-subtraction, we consider the following optimization problem:

$$\begin{aligned} \max_{\mathbf{x}_k\in\mathbb{R}^n,\mathbf{y}_k\in\mathbb{R}^m} \quad & \min\left[A_k - A_k\mathbf{x}_k\mathbf{y}_k^\top A_k\right], \\ \text{subject to} \quad & A_k\mathbf{x}_k \geq 0, \\ & \mathbf{y}_k^\top A_k \geq 0, \\ & \mathbf{y}_k^\top A_k\mathbf{x}_k = 1, \end{aligned} \tag{4.1}$$

where the minimum is taken over all entries of the matrix. Our rationale is twofold. First, by the nonnegativity constraints, we have $A_k - A_k\mathbf{x}_k\mathbf{y}_k^\top A_k \leq A_k$. Thus, the maximizer of $\min\left[A_k - A_k\mathbf{x}_k\mathbf{y}_k^\top A\right]$ always exists. Second, a nonnegative objective value means $A_{k+1} \geq 0$. In this case, the rank-one matrix $A_k\mathbf{x}_k\mathbf{y}_k^\top A_k$ is a feasible NE and we can move on to find the next NE for A_{k+1}. We summarize our NRF computation in Algorithm 1.

Because (4.1) is a nonlinear problem, most maximin algorithms are able to find only a local solution. However, a local solution with nonnegative objective value is all we need to validate an NE. A negative objective value, on the other hand, indicates that a rank one NE has not been found yet. In this event, we may try a different starting point for the maximin algorithm with the hope that maybe another local solution can be found. Such a strategy has been adopted twice in Algorithm 1. Line (1.21) is to restart the optimization solver with the current A_k and Line (1.23) is to restart the entire NRF process with A_0.

Example 2. Consider the 4×5 matrix $A = [\mathscr{C}; \mathbf{c}]$, where $\mathbf{c}$ is a randomly chosen nonnegative vector from $\mathbb{R}^4_+$ such that this matrix A is of full row rank. Therefore $\operatorname{rank}_+(A) = \operatorname{rank}(A)$. Trivially, splitting A by its rows is automatically an NRF. On the other hand, the 4×5 matrix $\Delta := [0_4, \mathbf{c}]$ with 0_4 standing for the 4×4 zero matrix is an NE because it is nonnegative, has rank 1, and leaves behind a nonnegative matrix $A - \Delta = [\mathscr{C}; \mathbf{0}]$ which is of rank 3. But the resulting matrix $[\mathscr{C}; \mathbf{0}]$ does not have an NRF as we have already argued in Example 1. The matrix Δ is expressible in the Wedderburn form $\Delta = A\widehat{\mathbf{x}}\widehat{\mathbf{y}}^\top A$ with $\widehat{\mathbf{x}} = [0,0,0,0,1]^\top$ and $\widehat{\mathbf{y}} = [\alpha, -\alpha, -\alpha, \alpha]^\top$, where $\alpha = 1/(c_1 - c_2 - c_3 + c_4)$. It can be checked that $(\widehat{\mathbf{x}}, \widehat{\mathbf{y}})$ is a local maximizer to (4.1). Had the maximin solver converged to this maximizer, then Δ would have been the first NE extracted by our

Algorithm 1: $[U, V, p, \text{Iflag}] = \text{NRF}(A, \epsilon, Gmax, Lmax)$

Input:

A	$=$	matrix in $\mathbb{R}_+^{m\times n}$ to be factorized,
ϵ	$=$	threshold for machine zero,
$Gmax, Lmax$	$=$	maximal allowable numbers for retries,

Output:

p	$=$	an integer, is the numerical $\text{rank}_+(A)$ if Iflag $= 0$,
U	$\in$	$\mathbb{R}_+^{m\times p}$,
V	$\in$	$\mathbb{R}_+^{p\times n}$,
Iflag	$=$	$\begin{cases} 0, & \text{An NRF is found with } \|A - UV\|_F < \epsilon, \\ 1, & \text{Failed to find NRF.} \end{cases}$

1.1 **begin**
1.2 $Gstart \leftarrow 0$;
1.3 $B \leftarrow A$;
`initialization`
1.4 $Gstart \leftarrow Gstart + 1$;
1.5 $Lstart \leftarrow 0$;
1.6 $U \leftarrow []$;
1.7 $V \leftarrow []$;
1.8 $p \leftarrow 0$;
1.9 **if** $\|B\|_F \geq \epsilon$ **then**
1.10 **if** $Gstart \leq Gmax$ **then**
1.11 $\mathbf{x}, \mathbf{y} \leftarrow$ feasible random starting points;
1.12 $[\mathbf{x}, \mathbf{y}, \text{ObjValue}] \leftarrow$ Solve (4.1) with respect to B by available optimization routines;
1.13 **if** ObjValue ≥ 0 **then**
1.14 $p \leftarrow p + 1$;
1.15 $U \leftarrow [U, B\mathbf{x}]$;
1.16 $V \leftarrow [V; \mathbf{y}^\top B]$;
1.17 $B \leftarrow B - B\mathbf{x}\mathbf{y}^\top B$;
1.18 **else**
1.19 **if** $Lstart \leq Lmax$ **then**
1.20 $Lstart \leftarrow Lstart + 1$;
1.21 Go to line 1.11;
1.22 **else**
1.23 Go to line 1.3;
1.24 **end**
1.25 **end**
1.26 **else**
1.27 Report that an NE is not found after retries;
1.28 Iflag $= 1$;
1.29 return;
1.30 **end**
1.31 **else**
1.32 Iflag $= 0$;
1.33 **end**
1.34 **end**

algorithm. But then our iteration would get stuck because there is no more NE contained in $\mathscr{C}$. It is for situation like this that we suggest in Algorithm 1 by Line (1.23) to restart the process entirely.

The above example demonstrates a pathological phenomenon that a "bad" NE can causes a break-down for Algorithm 1, even if the NRF does exists. With the aid of the built-in restart mechanism, we rarely see this happening for generic matrices in $\mathfrak{R}(m,n)$, especially for matrices with no zero entries. Our extensive numerical experiments seem to evidence that Algorithm 1, apart from the fact that the accuracy of the numerical results depends on the stopping criteria set out in the underlying optimization solver, is generally robust.

For computation, the rank-one reduction procedure as is proposed in Algorithm 1 is sufficient. For theoretical consideration, Theorem 3.2 suggests that it is possible to reduce multiple ranks of the matrix A_k simultaneously in one step. A suitable change of variable can diminish the role of the matrix R^{-1} in Theorem 3.2 and gives rise to a natural generalization of (4.1) in the form

$$\begin{aligned} \max_{X_k\in\mathbb{R}^{m\times r}, Y_k\in\mathbb{R}^{n\times r}} \quad & \min\left[A_k - A_k X_k Y_k^\top A_k\right], \\ \text{subject to} \quad & A_k X_k \geq 0, \\ & Y_k^\top A_k \geq 0, \\ & Y_k^\top A_k X_k = I_{r\times r}, \end{aligned} \tag{4.2}$$

where $I_{r\times r}$ is the identity matrix of rank r. The next theorem observes an interesting connection that if the rank of the matrix A_k can be reduced by r in one step via (4.2), then the same reduction can be achieved by rank-one matrices in r steps via (4.1).

THEOREM 4.1. *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 3.2, we know there are matrices $X\in\mathbb{R}^{n\times r}$ and $Y\in\mathbb{R}^{m\times r}$ satisfying

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

and making the matrix $B := A - AXY^\top A$ nonnegative with $\operatorname{rank}(B) = \operatorname{rank}(A) - r$. 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 now show that each pair $(\mathbf{x}_k,\mathbf{y}_k)$, $k=1,\ldots,r$, has the desirable effect of rank-one reduction on A_k defined successively by (3.4), starting with $A_1 = A$, and that $A_{r+1} = B$.

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

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

By the mathematical induction, we conclude that the very same matrix B can be achieved by r rank-one reductions via the sequence of vectors $\{(\mathbf{x}_k, \mathbf{y}_k)\}_{k=1}^r$. □

Theorem 4.1 avers an important message in that no matrix can have a nonnegative rank-r reduction by a rank-r matrix without first having a nonnegative rank-one reduction by a rank-one matrix. For rank-one reduction, there are $n + m$ variables involved in the problem (4.1) and we have to solve the problem r times (if no restart). For rank-r reduction, there are $r(n+m)$ variables involved in the problem (4.2), but we just need to solve the problem once. As r grows larger, the complexity involved in the problem (4.2) even by the same optimization solver would grow far more rapidly than the total complexity of r applications to the problem (4.1). Nonnegative rank-one reduction is preferable for the computation of NRF.

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]. Similar to the NRF problem, there are two open questions about CP matrices. First, determine whether a given nonnegative semi-definite matrix is CP. Second, if a matrix is CP, determine its cp-rank [4].

If A is CP, then obviously $\text{rank}_+(A) \le \text{rank}_{cp}(A)$. There is also an upper bound estimate [2, 17],

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

provided $\text{rank}(A) > 1$. Some sufficient conditions under which $\text{rank}_{cp}(A) = \text{rank}(A)$ can be found in [35]. If A is generated by a Soules matrix, for example, then $\text{rank}_{cp}(A) = \text{rank}(A)$ [36]. Other than these results, we find very few discussions about how to compute the CP factorization. The symmetric form demanded in (5.1) seems to make the CP problem for a nonnegative seme-definite matrix more stringent than the NRF problem for a general nonnegative matrix. Our contribution to this subject is that our algorithm can determine heuristically whether $\text{rank}_{cp}(A) = \text{rank}(A)$ and, if affirmative, we can compute the factor B numerically.

Our idea is a slight modification of the Wedderburn formula. In order to maintain the symmetry, we only need to require $\mathbf{x} = \mathbf{y}$ in the rank reduction formula [11]. Starting with $A_1 = A$, the optimization problem in (4.1) is reformulated as follows:

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

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. The discussion above for the NRF problem can be carried over to the CP problem until such a symmetric NE can be found no more. As the rank is reduced

by one in each step, the process must terminate in at most $\text{rank}(A)$ steps in exact arithmetic. If $\text{rank}_{cp}(A) = \text{rank}(A)$, in particular, then we have now a numerical procedure to find its nonnegative factor $B = (A_1\mathbf{x}_1, \ldots, A_{\text{rank}(A)}\mathbf{x}_{\text{rank}(A)})$. Our approach perhaps represents only a modest advance toward the CP problem, but we are not aware of any other way to compute the CP in the literature.

6. Maximal nonnegative rank splitting. Due to its rank subtractivity property, our process cannot extract more than $\text{rank}(A)$ many rank-one NEs. Our approach cannot handle factorization for the cases $\text{rank}_+(A) > \text{rank}(A)$ or $\text{rank}_{cp}(A) > \text{rank}(A)$. Nevertheless, we still can use our algorithm to address another interesting notion of the so called maximal nonnegative rank splitting (MNRS) defined below.

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

$$A = B + C, \tag{6.1}$$

where both B and C are nonnegative matrices satisfying

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

and $\text{rank}(B)$ 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 has to be terminated (after many retries), say, at A_k, if we trust that A_k has no more NE, then $B := A - A_k$ and $C := A_k$ form an MNRS for A.

We mention two examples to demonstrate the notion of MNRS.

Example 3. The matrix $\mathscr{C}$ defined in (1.3) has zero B component in its MNRS.

Example 4. Consider the 8×8 nonnegative matrix $A := [W, H]$, where $W := \begin{bmatrix} \mathscr{C} \\ 0_4 \end{bmatrix} \in \mathbb{R}^{8\times 4}$ and $H \in \mathbb{R}^{8\times 4}$ is made of the product $H = H_1H_2$ with

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

It is easy to observe that $\text{rank}(H) = \text{rank}_+(H) = 3$ and, hence, $\text{rank}(A) = 6$. Does A have an NRF, or can we retrieve an MNRS of A?

Apply Algorithm 1 to the matrix A. The farthest we can go is, after three iterations, a splitting $A = B + C$ with

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

and

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

where for the ease of running text we have displayed all numerals in 5 digits only. Since this result is obtained from the Wedderburn rank reduction process, we naturally are guaranteed by theory that $\mathrm{rank}(B) = \mathrm{rank}_+(B) = 3$ and $\mathrm{rank}(C) = \mathrm{rank}(A) - \mathrm{rank}(B) = 3$. Observe, however, that the special matrix $\mathscr{C}$ is still embedded at the upper left corner of C, forcing the fact that $\mathrm{rank}_+(C) = 4$ and that C has no more NE. This splitting is an MNRS of A.

7. Numerical experiments. A general nonnegative matrix may not have an NRF. In this case, we have already exemplified in Section 6 how the MNRS can be calculated. In this section, we demonstrate the working of our algorithm by computing the NRF of a few nontrivial matrices.

To ensure the existence of an NRF, we rely on Theorem 2.2 to generate test data. For illustration purpose, we employ the MATLAB routine FMINIMAX as our optimization solver in all our computation. The threshold for machine zero in Algorithm 1 is set at $\epsilon = 10^{-10}$.

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

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

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

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

and

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

where, for convenience, we have transcribed all digits of the computed result.

Example 6. To demonstrate the NRF of a CP matrix, we randomly generate a nonnegative matrix W

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

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

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

Example 7. For a general matrix R with negative entries, it cannot be guaranteed that the product $RR^\top$ has an NRF. If R is a Soules matrix, however, then it is known that 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 elements [36].

The matrix

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

has negative entries, but is a Soules matrix [16]. With this R and with

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

we define

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

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

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

8. Conclusion. The notion of NMF has many important applications, but NMF usually only computes an approximation for a given nonnegative matrix. Detecting the nonnegative rank and computing a complete nonnegative factorization for a general nonnegative matrix are a very challenging problem both in theory and in practice. No existing NMF algorithms can guarantee to find a complete nonnegative factorization of a nonnegative matrix.

When $\text{rank}_+(A) = \text{rank}(A)$, we say that the matrix A has an NRF. The main thrust of this paper is to present a numerical method to assess whether a given nonnegative matrix does have an NRF. Our method is still heuristic, but empirical results seem to strongly suggest that our algorithm can handle the situation reasonable well for generic nonnegative matrices. Our method can also be used to study the notion of MNRS and works in a similar way for CP matrices.

The crux of our algorithm is the employment of the Wedderburn rank reduction formula. At each step, we look for an NE that not only reduces the rank by one, but also maintains the nonnegativity for the residual. To search for an NE, we need to resolve a maximin problem numerically by existing optimization solvers.

Described in this paper is only a numerical procedure that might serve as a possible computational tool for the NRF problem. The method is admittedly quite crude, but it is perhaps the first algorithm ever proposed in the literature which is able to explore the NP-hard nonnegative factorization numerically with considerable success. No backward stability of our algorithm has been analyzed. Indeed, perturbation analysis for NRF in general has not been studied in the literature at all. The following questions remain to be further investigated.

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 Algorithm 1 so that UV is a numerical NRF of A. Is UV the exact NRF of some perturbed nonnegative matrix $A + E$?

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 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.

[4] A. Berman and U. G. Rothblum, *A note on the computation of the CP-rank*, Linear Algebra Appl., 419 (2006), pp. 1–7.

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

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

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

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

[9] M. T. Chu, F. Diele, R. J. 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).

[10] 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.

[11] 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).

[12] 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.

[13] 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.

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

[15] 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.

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

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

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

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

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

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

[22] L. Hubert, J. Meulman, and W. Heiser, *Two purposes for matrix factorization: a historical appraisal*, SIAM Rev., 42 (2000), pp. 68–82 (electronic).

[23] 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.
[24] M. W. Jeter and W. C. Pye, *A note on nonnegative rank factorizations*, Linear Algebra Appl., 38 (1981), pp. 171–173.
[25] ———, *Some nonnegative matrices without nonnegative rank factorizations*, Indust. Math., 32 (1982), pp. 37–41.
[26] 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.
[27] H. Kim and H. Park, *Nonnegative matrix factorization based on alternating nonnegativity constrained least squares and active set method*, SIAM J. Matrix Anal. Appl., 30 (2008), pp. 713–730.
[28] D. A. Klain and G.-C. Rota, *Introduction to geometric probability*, Lezioni Lincee. [Lincei Lectures], Cambridge University Press, Cambridge, 1997.
[29] 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.
[30] D. D. Lee and H. S. Seung, *Learning the parts of objects by non-negative matrix factorization*, Nature, 401 (1999), pp. 788–791.
[31] ———, *Algorithms for non-negative matrix factorization*, in In Advances in Neural Information Processing Systems, vol. 13, 2001, pp. 556–562.
[32] B. Levin, *On calculating maximum rank one underapproximations for positive arrays*, tech. report, Division of Biostatistics, Columbia University, New York, 1985.
[33] 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.
[34] P. Paatero and U. Tapper, *Positive matrix factorization: A non-negative factor model with optimal utilization of error*, Environmetrics, 5 (1994), pp. 111–126.
[35] N. Shaked-Monderer, *Minimal cp-matrices*, ELA, 8 (2001), pp. 140–157.
[36] ———, *A note on the cp-rank of matrices generated by a soules matrix*, ELA, 12 (2004), pp. 2–5.
[37] S. Sra and I. S. Dhillon, *Nonnegative matrix approximation: Algorithms and applications*, tech. report, Deptartment of Computer Sciences, University of Texas at Austin, 2006.
[38] J. J. Sylvester, *On a special class of questions on the theory of probabilities*, Birmingham British Assoc. Rept., (1865), pp. 8–9.
[39] 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.
[40] s. A. Vavasis, *On the complexity of nonnegative matrix factorization*, 2007. Available at `http://www.citebase.org/abstract?id=oai:arXiv.org:0708.4149`.
[41] J. H. M. Wedderburn, *Lectures on matrices*, Dover Publications Inc., New York, 1964.