Computing Symmetric Nonnegative Rank Factorizations

V. Kalofolias, E. Gallopoulos

Department of Computer Engineering and Informatics, University of Patras, Greece

Abstract

An algorithm is described for the nonnegative rank factorization (NRF) of some completely positive (CP) matrices whose rank is equal to their CP-rank. The algorithm can compute the symmetric NRF of any nonnegative symmetric rank-r matrix that contains a diagonal principal submatrix of that rank and size with leading cost $O(rm^2)$ operations in the dense case. The algorithm is based on geometric considerations and is easy to implement. The implications for matrix graphs are also discussed.

Keywords: nonnegative rank factorization, nonnegative matrix factorization, symmetric, rotation, isometry, principal submatrix, pivoted Cholesky factorization, symmetric positive semidefinite, arrowhead matrix, extreme ray, maximum independent set, rank reduction

1. Introduction

The problem of nonnegative rank factorization (NRF) is posed as follows [2, 3, 35]: Given[1] $A \in \mathbb{R}_+^{m\times n}$ of rank r find nonnegative factors $W \in \mathbb{R}_+^{m\times k}$ of full column rank and $H \in \mathbb{R}_+^{k\times n}$ of full row rank such that $k = r$ and $A = WH$. The problem is immediately related to that of nonnegative matrix factorization (NMF) that has been attracting a great deal of attention because of its interest in many applications; cf.[12]. Unlike it, however, NRF does not always have a solution, in the sense that the nonnegative rank of A, that is the minimum value k such that the above factorization exists (termed nonnegative rank, denoted by $\mathrm{rank}_+(A)$) can be larger than r. In fact, it is well known that $\mathrm{rank}(A) \le \mathrm{rank}_+(A) \le \min(m, n)$ [13].

In this paper we are concerned with symmetric NRF. That is, in the above factorization, we seek nonnegative W such that $A = WW^\top$. Matrices admitting such a factorization for some $k \ge \mathrm{rank}(A)$ are called completely positive (CP) (see e.g. [32]) and the minimum k for which such a factorization exists is called the CP-rank of A, sometimes denoted $\mathrm{rank}_{\mathrm{CP}}(A)$. CP matrices arise in applications (e.g. [6, 14, 24, 36]) and have even been used as test matrices [31]. Whether a given matrix is CP and computing its CP-rank are open research problems; cf. [4, 5].

For the special case of rank-2, there exist algorithms to compute the NRF [1, 13]. We show here how to extend these to handle the symmetric case. Under some conditions, the questions of

Email addresses: `kalofolb@ceid.upatras.gr` (V. Kalofolias), `stratis@ceid.upatras.gr` (E. Gallopoulos)

[1] Subscript "+" signifies the nonnegative section of the algebraic object.

existence and construction of the symmetric NRF can also be answered for matrices of higher rank. For example, CP matrices must have nonnegative eigenvalues. Diagonally dominant nonnegative matrices are always CP, and an algorithm to compute their symmetric NRF was described by Kaykobad [28]. That method, however, does not reveal the CP-rank since the dimension k of the computed factor $W \in \mathbb{R}_+^{m\times k}$ can turn out larger than the smallest possible.

In this paper we are concerned with symmetric matrices of rank r that satisfy a specific property. Because this will be used repeatedly throughout our discussion, for ease of reference we give it a special name:

> **Property $\Delta_{\max}$:** A matrix $A \in \mathbb{R}^{n\times n}$ will be said to have "property $\Delta_{\max}$" or simply "to be $\Delta_{\max}$" if it contains a diagonal principal submatrix of rank equal to the rank of A.

Clearly, the diagonal principal submatrix must contain strictly positive elements along its diagonal. As will be made clear, nonnegative symmetric matrices that are $\Delta_{\max}$ always possess a symmetric NRF. We develop the theory necessary to design an algorithm to construct the symmetric NRF for $\Delta_{\max}$ matrices. Like several other contributions (e.g. [24, 26, 29]), the proposed algorithm is based on a geometric intepretation of the problem. The contribution could be of value in applications where CP matrices arise. In this discussion we do not concern ourselves with implementation details and the effects of finite precision.

The outline of this paper is as follows. We review previous NRF efforts in Section 2. In Section 3 we recall an important property of symmetric positive semidefinite matrices that is then used to bind all possible solutions of the symmetric NRF. In Section 4 this property is utilized to solve the problem for CP matrices of rank 2. The remainder of this paper deals with the properties and NRF of symmetric nonnegative matrices that are $\Delta_{\max}$. Specifically, Section 5 provides properties of the sought factors and uniqueness results. It also shows the implications of $\Delta_{\max}$ for the associated matrix graph. Section 6 describes the algorithm, called IREVA, for the aforementioned types of matrices. A numerical example and illustration of IREVA are shown in Section 7.

Recall the following terminology: If K is a convex cone, and for any $x \in K$ such that $x = y + z$ where $y, z \in K$ implies that both y and z are nonnegative scalar multiples of x, then x is termed "extreme vector" of K and the ray it generates "extreme ray" of K [5]. Also, a nonnegative square matrix is characterized as "monomial" if it has exactly one nonzero in each row and column and thus can be expressed as product of a diagonal matrix and a permutation matrix [2]. Matrices that are nonnegative and positive semidefinite are termed "doubly nonnegative" [5]. Note that CP matrices are doubly nonnegative but the opposite does not hold in general, except for matrices of order at most 4 [7]. Finally, a matrix X is called a "1-inverse" of matrix A if $AXA = A$.

2. Related work

In the literature, there are no practical algorithms explicitly addressing the computation of symmetric NRF, so we first describe aspects and algorithms for the general (i.e. nonsymmetric) case that will be shown to be useful to the symmetric case for a certain class of matrices; cf. Theorem 4. It was proven by Vavasis that computing the NRF, if it exists, is NP-hard [37][2]. It

[2]The problem there is termed "exact NMF".

was shown by Berman and Plemmons in [2, Thm. 4] that any matrix $A \in \mathbb{R}_+^{m\times n}$ of rank r that contains an $r \times r$ monomial submatrix has an NRF, the matrix has a nonnegative 1-inverse and each of the terms in the NRF contains a monomial submatrix of rank r. The construction of an NRF for matrices that have a nonnegative 1-inverse was outlined by Campbell and Poole [8]. This relies on repeated subtractions of suitable scaled rows from each other and possible exchanges. The algorithm terminates when the matrix is brought into the form

$$P_t P_{t-1} \dots P_1 A = \begin{bmatrix} H \\ \mathbf{0} \end{bmatrix}$$

Factor W is readily obtained from the leading r columns of the product $P_1^{-1} P_2^{-1} \dots P_t^{-1}$. The algorithm is described at a high level and is concerned with the construction of the NRF for the nonsymmetric problem. The stable efficient implementation and its use for symmetric NRF remain interesting open problems.

Gray and Wilson in [24] used a geometric approach to show the existence of NRF for symmetric matrices up to size 4. Counterexamples were also provided to show the failure of this approach for matrices of larger size.

An optimization-based approach for NRF was proposed by Dong, Lin and Chu [15]. The algorithm relies on generating a nonnegative rank-1 factor such that when subtracted from A the algorithm is reduced in rank while remaining nonnegative. This is related to "underapproximation" [30] that was explored in a series of interesting papers by Gillis and collaborators to obtain the NMF [22, 21]. The algorithm uses the Wedderburn formula ([38]) for rank reduction that we also apply in the algorithm proposed herein. In [15], at each downdating step, a minimax problem is solved and after r steps and in the general case that the nonnegative rank is equal to the rank, the method generally returns an NRF. If a nonnegative remainder arises towards the end of the computation, some recomputation has to be applied. An advantage of the algorithm is that no prior constraints are imposed (though the method will fail when the nonnegative rank is higher than the rank) the price being that there are heuristic steps so there is no guarantee that the method will always work. The method can easily be modified to accommodate symmetric inputs and to produce a symmetric factorization. One difficulty is that the minimax problem solved at each step becomes a computational bottleneck for large matrices.

The Extreme Vector Algorithm (EVA), proposed by Klingenberg et al. in [29] considers further restrictions on the inputs in order to address the lack of uniqueness and subsequent ill-posedness inherent in some nonnegative factorizations. As presented, however, EVA is designed for the NRF of full rank matrices instead of the relatively small or moderate rank for which our algorithm is best suited.

It is worth noting that related ideas appeared in the past, e.g. [24], for small matrices and in proving the existence of NRF. More recently, the idea of rotations, that is central in the algorithm proposed here, appeared in the literature of factor analysis; cf. [33] and references therein, especially [26]. That context, however, is quite different. The challenge there is to compute (approximate) nonnegative matrix factorizations, and because symmetry is not an issue, the term "rotation" is used to characterize any invertible linear transformation that achieves positivity of the factors, i.e. compute T (not necessarily orthogonal) such that $A = \hat{W}TT^{-1}\hat{H}$, and $\hat{W}T, T^{-1}\hat{H}$ are nonnegative for given $\hat{W}, \hat{H}$ of arbitrary sign.

In [9], Catral et al. show that when the matrix is symmetric, the best reduced rank approximation of A need not be symmetric, nor does $W = H^\top$ hold necessarily. The authors also discuss conditions under which either of the above can hold. Since these results concern the approximate NMF, they are complementary to our discussion here.

Shaked-Monderer recently proved in [34, Thm. 4] a closely related result, namely that rank-r symmetric matrices that are doubly nonnegative and satisfy property $\Delta_{\max}$ are CP with CP-rank equal to r. It was also shown therein how to obtain the factor W for matrices of this type in which the rank-r principal submatrix is in the leading position. We will discuss these findings in Sections 5 and 6.

3. Relating symmetric factorizations

As is well known (cf. [16], [29] and references therein) the NRF factors might not be unique, causing a form of ill-conditioning. A similar lack of uniqueness can occur when A is CP and one seeks a symmetric NRF. In fact (see e.g. [27]), uniqueness is also lacking in the factorization of (mixed-sign) symmetric positive semidefinite matrices: However, all solutions are related via an orthogonal transformation. The following result is mentioned in [24]. For the sake of completeness we prove it in the Appendix.

Theorem 1. *Let $A \in \mathbb{R}^{m\times m}$ be symmetric positive semidefinite of rank r and let $W \in \mathbb{R}^{m\times r}$ and $H \in \mathbb{R}^{m\times r}$ be such that $A = WW^\top = HH^\top$. Then there exists an orthogonal matrix $R \in \mathbb{R}^{r\times r}$ s.t. $W = HR$.*

The lack of uniqueness and the fact that any two symmetric factorizations (i.e. the factors) are related by means of an orthogonal transformation, turn Theorem 1 into a useful starting point in the construction of symmetric NRF's. First note that when the matrix is symmetric positive semidefinite and nonnegative (all these being necessary conditions for the matrix to be CP), a straightforward symmetric rank factorization is directly computable as $V = Q\sqrt{E}$, where $Q \in \mathbb{R}^{m\times r}$ is the matrix of eigenvectors and $E \in \mathbb{R}_+^{r\times r}$ the diagonal matrix of (nonnegative, because of semidefiniteness) eigenvalues of A. Because of orthogonality, the distance between any two points in $\mathbb{R}^r$ is preserved when transformed by R, so this acts as an isometry. In case that the determinant of R is 1, then this will actually be a rotation. Since by a simple column interchange of R the matrix remains orthogonal and the determinant changes sign, in the sequel we can assume w.l.o.g. that R is a rotation. Interpreting the rows of W, say w_i, and H, say h_i, as points in $\mathbb{R}^r$, it also follows from $A = WW^\top = HH^\top$ that

$$h_i^T h_i = w_i^T w_i = \alpha_{i,i} \tag{1}$$

and thus both h_i and w_i have equal lengths (2-norm).

Similarly, for any two pairs of rows at corresponding positions of H, say h_i, h_j, and W, say w_i, w_j, it holds that $h_i^T h_j = w_i^T w_j = \alpha_{i,j}$, so from this and the fact shown earlier that corresponding row vectors of W and H are of equal length, it follows that the angle between vectors h_i and h_j is equal with the angle between w_i and w_j. Equivalently, for any given nonnegative matrix that has an NRF $A = WW^\top$, the lengths of and angles between rows of W are determined by the values of A.

It follows that the symmetric NRF of A can be reduced to computing an orthogonal matrix $R \in \mathbb{R}^{r\times r}$ such that $VR \geq 0$, where $V \in \mathbb{R}^{m\times r}$ is an arbitrary (mixed sign) solution to the symmetric factorization. The existence of such a matrix R is equivalent to matrix A being CP with CP-rank equal to its rank.

4. Symmetric NRF: The case of rank-2 matrices

Any symmetric $A \in \mathbb{R}_+^{m\times m}$ of rank-2 that is positive semidefinite is CP with $\text{rank}_{\text{CP}} = 2$ (cf. [5, Theorem 2.1]). Then, finding a solution of the symmetric NRF is relatively simple. In this section we propose an algorithm that is applicable to any symmetric, rank-2, matrix that is doubly nonnegative.

The algorithm proceeds as follows: In the first step, a mixed-sign factor $V \in \mathbb{R}^{m\times 2}$ s.t. $A = VV^\top$ is constructed to "seed" the second step, in which the rows of V are rotated into the positive orthant.

One way to implement this approach is to construct the factor V so that its first column is nonnegative. Then all rows of V lie in the nonnegative right-half plane. Moreover, the angle between any two of these vectors would be less than $\pi/2$ because of the nonnegativity of A. Therefore, there exists a rotation that can move all rows of V in the nonnegative quadrant. There remains to construct the sought V and an appropriate rotation. Regarding the former, two possible choices are the following: One is to construct $V = Q\sqrt{E}$, as before. Specifically, E is the 2×2 diagonal matrix containing the nonzero eigenvalues of A. Since E is the matrix of eigenvalues of A that are nonnegative and the first column of Q is the Perron vector of A that can be selected to be nonnegative, so will also be the entire first column of V. We next show that an alternative seeding method to the use of eigenvectors is the Cholesky decomposition. The following result is well known; cf. [27]:

Theorem 2. *Let the rank-r matrix $A \in \mathbb{R}^{m\times m}$ be symmetric positive semidefinite. Then there exists at least one lower triangular L with nonnegative diagonal elements such that $A = LL^\top$.*

The next corollary follows immediately and guarantees a Perron-like property, specifically that the first vector from the Cholesky decomposition of nonnegative matrices is nonnegative.

Corollary 3. *If A is doubly nonnegative then at least one of the possible L factors above has nonnegative first column.*

It remains to construct an appropriate rotation. We first note that the two row vectors in V that form the largest angle with the positive right axis actually are extreme vectors. These are readily found so the final step is to compute an angle θ such that rotating the extreme vectors by this angle, say counterclockwise, will bring them within the first quadrant. Clearly, this rotation will also move therein all remaining points. Note that θ is not necessarily unique, unless the angle between the two extreme vectors is maximal with $\phi = \pi/2$. In that case, after rotation, the two extreme vectors will align themselves, one with the x, the other with the y axis. Otherwise, θ will not be unique and can be chosen as follows: Letting ϕ_1 (resp. ϕ_2) be the angle between the x axis and the extreme vector closest to the negative (resp. positive) y axis, then it is enough to rotate by an angle $\theta \in [-\phi_1, \frac{\pi}{2} - \phi_2]$. Note also that whenever A contains zero rows, so will V;

such rows correspond to the axes' origin, are not affected by rotation and are discarded prior to the identification of extreme vectors. Algorithm 1 implements the scheme just described with ϕ selected so as to be in the middle of the interval of possible angles of rotation.

Algorithm 1 Symmetric NRF of rank-2 matrices

Input: CP matrix $A \in \mathbb{R}_+^{m\times m}$ of rank 2
Output: Factor $W \in \mathbb{R}_+^{m\times 2}$ s.t. $A = WW^\top$

1: Compute eigenvalue decomposition $A = QEQ^\top$, $Q \in \mathbb{R}^{m\times 2}$, $E \in \mathbb{R}_+^{2\times 2}$ with $Q_{:,1} \geq 0$
2: $V \leftarrow Q\sqrt{E}$
3: $\zeta \leftarrow$ the set of nonzero rows of V
4: $t_\theta \leftarrow V_{\zeta,2}./V_{\zeta,1}$
5: $\theta_{\min} \leftarrow -\arctan(\min(t_\theta))$
6: $\theta_{\max} \leftarrow \pi/2 - \arctan(\max(t_\theta))$
7: $\theta \leftarrow (\theta_{\min} + \theta_{\max})/2$
8: $R \leftarrow \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix}$
9: $W \leftarrow VR$;

5. Symmetric NRF: Uniqueness and graph intepretation

Any investigation of matrix factorizations (rank or otherwise) has to address the topic of uniqueness. Its absence is a source of difficulty (e.g. see [16, 25, 29] and references therein) but can also be used as a feature to help in NRF algorithm design, as will be done here.

One case where the factorization is unique is when the matrix $W \in \mathbb{R}_+^{m\times r}$ is sparse and contains row vectors (points) that "fill" the facets of the nonnegative orthant $\mathbb{R}_+^r$. Then, the smallest cone that contains the data and is inside the nonnegative orthant is the orthant itself. In that case, the extreme vectors have to lie along the orthant axes, that is to be positive multiples of $e_1, e_2, \ldots, e_r$.

In general, if a matrix $W \in \mathbb{R}^{m\times r}$ contains r rows that are multiples of $e_i, i \in \{1, 2, \ldots, r\}$, then it must contain an $r \times r$ monomial submatrix. As mentioned in Section 2 from [2, Thm. 4], it also holds that a rank-r matrix $A \in \mathbb{R}_+^{m\times n}$ contains a monomial submatrix of rank r if and only if it has a nonnegative rank factorization $A = WH$, where W and H contain a monomial submatrix of rank r each. A symmetric counterpart of this property is proven in Corollary 5.

Theorem 4. *Let $A \in \mathbb{R}_+^{m\times m}$ be rank-r symmetric and satisfy property $\Delta_{\max}$. Then A is also CP with $\operatorname{rank}_{\mathrm{CP}}(A) = r$ and has a unique symmetric NRF $A = VV^\top$, up to column permutations of V. Moreover, for any NRF $A = WH$, the columns of W and rows of H can always be rescaled with diagonal matrices D_W, D_H so that $WD_W = H^\top D_H = V$.*

Proof. First note that if A is as above then there exists a symmetric permutation to make diagonal its upper left $r \times r$ block. Without loss of generality, let the diagonal principal submatrix be the leading $r \times r$ block at the upper left corner of A. Then from the discussion preceding the statement of the theorem it follows that there is always an NRF, say $A = WH$, where W and H each contain monomial submatrices of rank r. Then, if W_1 is the uppermost $r \times r$ submatrix of W and H_1 the leftmost $r \times r$ submatrix of H, it follows that $W_1 H_1 = D$ is diagonal and hence monomial, therefore W_1 and H_1 are also monomial. We can thus write $W_1 = D_{W_1} P_W$ and $H_1 = P_H D_{H_1}$, where D_{W_1} and D_{H_1} are diagonal nonnegative and P_W, P_H permutation matrices. Hence $D_{W_1} P_W P_H D_{H_1} = D$ thus $P_W P_H = D_{W_1}^{-1} D D_{H_1}^{-1}$. Since the only diagonal matrix that is a product of two permutation matrices is the identity matrix, it follows that $P_W P_H = I$ and so by orthogonality $P_W = P_H^\top$ and thus $D_{W_1} D_{H_1} = D$.

Since all rows (columns) of W (H) may be written as conical combinations of the first r rows (columns) of W (H),

$$W = C_W W_1 = C_W D_{W_1} P_W, \ \ H = H_1 C_H = P_H D_{H_1} C_H,$$

where $C_W \in \mathbb{R}_+^{m \times r}$ and $C_H \in \mathbb{R}_+^{r \times m}$. Note that the uppermost $r \times r$ submatrix of C_W and the leftmost $r \times r$ submatrix of C_H are the identity matrix. It also holds that $A = C_W D C_H$.

Due to the symmetry of A, it holds that

$$\begin{aligned} W_1 H = (W H_1)^\top \quad &\Leftrightarrow \quad W_1 P_H D_{H_1} C_H = (C_W D_{W_1} P_W H_1)^\top \\ W_1 P_H D_{H_1} C_H = H_1^\top P_W^\top D_{W_1}^\top C_W^\top \quad &\Leftrightarrow \quad D_{W_1} P_W P_H D_{H_1} C_H = (P_H D_{H_1})^\top P_W^\top D_{W_1}^\top C_W^\top \\ D_{W_1} D_{H_1} C_H = D_{H_1}^\top D_{W_1}^\top C_W^\top \quad &\Leftrightarrow \quad D C_H = D^\top C_W^\top \Leftrightarrow C_H = C_W^\top. \end{aligned}$$

Letting $D_W = P_W^\top D_{W_1}^{-1} D^{1/2} P_W$ and $D_H = P_H D^{1/2} D_{H_1}^{-1} P_H^\top$, it follows that $D_W D_H = I$ and $W D_W = (D_H H)^\top = V$. Thus V is a solution of the symmetric NRF of A. The extreme rays that define the cone containing all row vectors of V are on the nonnegative axes of $\mathbb{R}^r$, so the cone is unique. So has to be the solution V, up to a permutation of columns. □

The above theorem is key as it characterizes the matrices for which the algorithm proposed in this paper is applicable. Specifically, it guarantees that if a symmetric rank-r nonnegative matrix satisfies property $\Delta_{\max}$, then it is CP and has a unique symmetric NRF. Moreover, any nonsymmetric NRF algorithm can be used to obtain a solution to this symmetric problem.

Corollary 5. *A symmetric matrix $A \in \mathbb{R}_+^{m \times m}$ of rank r satisfies property $\Delta_{\max}$ if and only if it has a symmetric NRF, say $A = WW^\top$, where W contains a monomial $r \times r$ submatrix of full rank.*

We next note that property $\Delta_{\max}$ is equivalent to the existence of an independent set of size r in the graph of the matrix [17]. Moreover, this is a maximum independent set, since the existence of an independent set of size larger than r, say $r + 1$, would imply the existence of a principal diagonal submatrix of size $r + 1$ so the rank of the matrix would be at least $r + 1$, thus violating property $\Delta_{\max}$. It is worth noting that for arbitrary graphs, the problem of finding the maximum independent set is NP-complete [20].

These matrices can be brought by symmetric permutation into what could be termed "thin shaft broad arrowhead" form, that is

$$A = \begin{bmatrix} D & A_{12} \\ A_{21} & A_{22} \end{bmatrix}, \tag{2}$$

where $D \in \mathbb{R}^{r \times r}$ is diagonal, $A_{22} \in \mathbb{R}^{(n-r)\times(n-r)}$, and A_{22}, A_{12}, A_{21} not trivially 0. Such matrices become standard arrowhead matrices when $r = n - 1$. Moreover, $A_{21} = A_{12}^\top$ because of summetry.

Furthermore, because $\text{rank}(A) = \text{rank}(D) = r$, there exists a matrix M such that $M[D, B] = [B^\top, A_{22}]$, where for convenience we set $B := A_{12}$ and thus $M = B^\top D^{-1}$ and $A_{22} = B^\top D^{-1} B$. Note that setting

$$W^\top = [D^{1/2}, B^\top D^{-1/2}], \tag{3}$$

it readily follows that A is doubly nonnegative as well as CP, corroborating the findings of Theorem 4. See also [34, Thm. 4] for a closely related result.

We assume next w.l.o.g. that the matrix has been scaled so that D is the $r \times r$ identity matrix and thus $A_{22} = B^\top B$. This latter factorization and expression (2) for A reveal the type of weighted undirected graphs associated with such matrices. Specifically, they show that the graphs in question can be partitioned into two (hitherto unknown) classes, say $\mathcal{T}$ and $\hat{\mathcal{T}}$, the former containing the r nodes of the maximum independent set and the latter all remaining ones. Then, the elements of each column of B correspond to the weights of the edges between nodes from $\mathcal{T}$ to nodes in $\hat{\mathcal{T}}$. Thus, for any two nodes in $\hat{\mathcal{T}}$, the weight of the edge between them is equal to the inner product of the weights of the edges linking each of these two nodes with each of the r nodes in $\mathcal{T}$. Finally, there is a self loop for each node in $\mathcal{T}$.

6. Identification and rotation of extreme vectors algorithm

In this section, we describe an algorithm for computing the symmetric NRF of matrices that satisfy the criteria of Theorem 4, that is they are positive symmetric of rank-r and have property $\Delta_{\max}$. From Theorem 4 it follows that the symmetric NRF of such matrices is unique. The algorithm consists of two major phases. In the first phase, the algorithm computes some matrix $V \in \mathbb{R}^{m \times r}$ such that $A = VV^\top$ with the first column of V strictly positive. This then becomes the seed for the next phase, in which an orthogonal matrix $R \in \mathbb{R}^{r \times r}$ is computed to bring the rows of V into the nonnegative orthant of $\mathbb{R}^r$. The existence of such a matrix is assured by Theorem 1.

6.1. Computing V: A rank reduction framework

The eigenvalue decomposition and Cholesky factorization are candidate methods, as shown in Section 4 for the case of rank-2, however the requirement for strictly positive first column requires some extra care. Specifically, when the matrix is reducible, it must first be brought by symmetric permutations into normal form, in which each diagonal block is irreducible [19]; algorithms to that effect exist, e.g. [18, Ch. 6]. Because of symmetry, the normal form will be block diagonal. Each diagonal block can then be factorized so that the first column of each factor is positive by irreducibility and ready for the rotation step described in Section 6.3.

The rank reduction method of Wedderburn offers a more general framework. The following is corollary to the Wedderburn rank reduction formula ([10, 38]):

Corollary 6. *Let matrix $A \in \mathbb{R}^{m \times m}$ be symmetric and $x \in \mathbb{R}^n$ such that $\omega = x^\top A x \neq 0$. Then matrix $B = A - \frac{1}{\omega} A x x^\top A$ has rank exactly one less then the rank of A.*

Therefore, when A is also nonnegative with nonzero rows, one could select any (strictly) positive $x \in \mathbb{R}^m$ and set $V_{:,1} = \frac{1}{\sqrt{\omega}} A x$. The following "Rank Reduction Theorem" is essential:

Theorem 7. *[11, Th. 3.1] Suppose that A is symmetric positive semidefinite, H symmetric and* $\mathrm{rank}(A - H) = \mathrm{rank}(A) - \mathrm{rank}(H)$. *Then $A - H$ is positive semidefinite.*

Here $H := \frac{1}{\omega} A x x^\top A$ and because of Corollary 6 the conditions of Theorem 7 hold, thus B remains positive semidefinite and the procedure can be repeated using B and selecting another x as before to reduce its rank, until a zero remainder is reached after $r = \mathrm{rank}(A)$ steps; at that point there will be r vectors that can be used to populate the columns of V so that $A = VV^\top$. Since the signs of the vectors of V other than the first do not matter, to generate them after the first step, one can use any convenient method suitable for symmetric semidefinite matrices, e.g. symmetric eigensolver, any decomposition suitable for symmetric positive semidefinite matrices, e.g. pivoted Cholesky [31], or directly the Wedderburn formula.

6.2. Identifying extreme vectors

The next phase of the algorithm consists of computing and applying the orthogonal transformation to bring the rows of V into the nonnegative orthant $\mathbb{R}^r_+$. To prepare for this the algorithm must identify the extreme vectors of the cone of rows of V. One way to identify the extreme rays in matrix V would be by comparing angles between pairs of row-vectors. This method is being used in the algorithm (EVA) proposed in [29]. In [13] Cohen and Rothblum used a different approach which motivated our method. Instead of the matrix multiplication necessary for the angular computations, they first scaled the columns of A to make it stochastic and then identified as extreme vectors the columns with minimum and maximum value in the first element. As described in [13], however, the method appeared to be applicable only for rank-2 matrices. We extend this strategy to arbitrary rank, utilizing the fact that the first column of V is positive. The basis for the algorithm is the next theorem.

Theorem 8. *Let $\mathcal{V}$ be a set of vectors in $\mathbb{R}^k$ with nonnegative first component and unit length under some p-norm, where $1 < p < \infty$. If all elements of $\mathcal{V}$ are contained in a conical hull defined by $r \leq k$ linearly independent extreme vectors also contained in $\mathcal{V}$ and at most one of the vectors in $\mathcal{V}$ has zero first component, then the vector with the smallest first component is an extreme vector.*

Proof. Denote by $v^{(1)}, \ldots, v^{(r)}$ the extreme vectors of $\mathcal{V}$ in order of ascending first coordinate, that is $0 \leq {v_1}^{(1)} \leq \ldots \leq {v_1}^{(r)}$. Let also u be an arbitrary vector of $\mathcal{V}$. This must be a conical combination of extreme vectors, therefore $u = \beta_1 v^{(1)} + \ldots + \beta_r v^{(r)}$, $\beta_i \geq 0$. From Minkowski's inequality, for $1 < p < \infty$, if vectors x, y are in l_p (that is a strictly convex space), then $\|x+y\|_p \leq \|x\|_p + \|y\|_p$ with equality holding if and only if $x = 0$ or $y = 0$ or $x = \gamma y$ for some $\gamma > 0$. This can be generalized, so that

$$\|u\|_p \;=\; \|\beta_1 v^{(1)} + \ldots + \beta_r v^{(r)}\|_p \;\leq\; \beta_1 \|v^{(1)}\|_p + \ldots + \beta_r \|v^{(r)}\|_p, \tag{4}$$

where we know that $v^{(1)}, v^{(2)}, \ldots, v^{(r)}$ are nonzero and linearly independent. Thus, equality holds if and only if at most one of $\beta_1, \ldots, \beta_r$ is strictly positive. All vectors $u, v^{(i)}$ have unit length, so from inequality (4)

$$\beta_1 + \ldots + \beta_r \geq 1. \tag{5}$$

Equality holds when exactly one of the β_i's is equal to 1, otherwise the inequality is strict. Let now u be the vector of $\mathcal{V}$ with smallest first component, u_1. Then

$$u_1 = \beta_1 v_1^{(1)} + \ldots + \beta_r v_1^{(r)} \geq (\beta_1 + \ldots + \beta_r) v_1^{(1)} \geq v_1^{(1)} \geq u_1,$$

Therefore $u_1 = v_1^{(1)}$ and thus

$$\beta_1 v_1^{(1)} + \ldots + \beta_r v_1^{(r)} = v_1^{(1)}. \tag{6}$$

If $v_1^{(1)} = 0$, from the statement of the theorem, this will be the only one that is 0, that is

$$0 < v_1^{(2)} \leq v_1^{(3)} \ldots \leq v_1^{(r)}.$$

From relation (6) it follows that

$$0 = \beta_2 v_1^{(2)} + \ldots + \beta_r v_1^{(r)} \Leftrightarrow \beta_2 = \beta_3 = \ldots = \beta_r = 0$$

and so $\beta_1 = 1$ and $u \equiv v^{(1)}$. If $v_1^{(1)} > 0$, let $v_1^{(1)} = v_1^{(2)} = \ldots = v_1^{(j)} < v_1^{(j+1)} \leq v_1^{(j+2)} \leq \ldots \leq v_1^{(r)}$, $r \leq k$, so from relations (5) and (6) follows that

$$1 = \beta_1 + \beta_2 + \ldots + \beta_j + \beta_{j+1} \frac{v_1^{(j+1)}}{v_1^{(1)}} + \ldots + \beta_r \frac{v_1^{(r)}}{v_1^{(1)}} \geq \beta_1 + \beta_2 + \ldots + \beta_r \geq 1,$$

so equalities must hold, therefore

$$\beta_{j+1} = \beta_{j+2} = \ldots = \beta_r = 0. \tag{7}$$

As the sum of the β_i's is 1, exactly one of them will be 1 while the remaining ones will be 0. From Eq. (7) we conclude that the latter $r - j$ coefficients are 0, so for some $t \leq j$ it holds that $\beta_t = 1$. Furthermore $\beta_i = 0$ for $i \neq t$ and thus $u \equiv v^{(t)}$. $\square$

Note that if two distinct extreme vectors of $\mathcal{V}$ have zero first component, any vector that is generated from their conical combination will also have zero first component, preventing the identification of the extreme vector using the above approach.

In the sequel we assume that the set of vectors in $\mathcal{V}$ are m rows[3] of a matrix $\widetilde{V} \in \mathbb{R}^{m \times k}$ that results after normalizing the rows of V. Then, as long as the first column of $\widetilde{V}$ is positive, the first extreme vector is readily found. Assuming w.l.o.g. that $v^{(1)}$ has been tagged as the first extreme vector, to find the second one, first all rows of $\widetilde{V}$ that are not perpendicular to $v^{(1)}$ are eliminated from consideration. The rows remaining are linearly independent to $v^{(1)}$ and form a cone that has dimensionality reduced by one. The assumption that $\widetilde{V}$ contains another $r - 1$ rows that are orthogonal to $v^{(1)}$ simplifies things further, since the next extreme vector will be one of them. Theorem 8 again applies, so the vector amongst those that remain that has the smallest first element is tagged as the second extreme vector, say $v^{(2)}$, and so on, until all r extreme vectors have been found.

[3]Hence the choice of superscript for each $v^{(j)}$.

6.3. Rotation

What remains is to rotate the extreme vectors (rows) of V into the nonnegative orthant $\mathbb{R}^r_+$. Since A satisfies the conditions of Theorem 4, its symmetric NRF is unique up to column permutations. Having computed a factorization $A = VV^\top$, it follows from Theorem 1 that there exists orthogonal $R \in \mathbb{R}^{r\times r}$ such that $W = VR$. To compute R, we utilize the properties of the extreme vectors of the cone containing all rows of V. They are pairwise perpendicular, as will be the extreme vectors of W after rotation, due to the orthogonality of R. After rotation, the extreme vectors would lie on each of the distinct nonnegative axes of $\mathbb{R}^r$, their lengths remaining unchanged.

Assume now that the extreme vectors of V are the rows of its uppermost square submatrix V_1. After rotation, they are transformed into extreme vectors of W and are positioned in the uppermost submatrix W_1 of W. Both matrices have rows that are pairwise perpendicular, so

$$V_1 V_1^\top = W_1 W_1^\top = D^2,$$

where D is a diagonal matrix containing the lengths of the extreme vectors. As the extreme vectors of W lie on the nonnegative axes of $\mathbb{R}^r$, W_1 has to be a monomial matrix, from Corollary 5, thus $W_1 = DP$, where P is a permutation matrix assigning each extreme vector of W_1 to a nonnegative axis of $\mathbb{R}^r$. Since the sequence of vectors is not important, we assume that P is the identity matrix, leaving W_1 diagonal, hence

$$W_1 = D = \sqrt{V_1 V_1^\top}$$

This assumption is equivalent to expecting to find a rotation matrix that brings the first row of V_1 on the nonnegative axis along e_1, the second row on the axis along e_2, and so on. Thus

$$V_1 R = W_1 = D^2 D^{-1} = V_1 V_1^\top D^{-1}$$

and thus $R = V_1^\top D^{-1}$. It follows that to compute R, it is enough to normalize the rows of V_1 to unit length and then to transpose it. The above procedure could also be done in the (easier) case that we were initially given V rather than A.

It is worth noting[4] that once the extreme vectors have been identified, as indicated in the previous subsection, we can readily compute W from expression (3). This form was shown in [34, Thm. 4]. That way of computing W is more direct and would be useful in practice though it is not as revealing of the geometrical features of our approach.

The entire method, using rotations, is listed as Algorithm 2, called IREVA (Identification & Rotation of Extreme Vectors Algorithm). The most expensive step is that of the computation of the r columns of V, its exact cost depending on the method chosen to implement it. If the matrix is rank-r and the first step is implemented with the Wedderburn formula followed by Cholesky for semidefinite matrices (e.g. [31]) for the remaining columns, then the cost is $O(m^2 r)$. The remaining steps cost much less, so this value would be a fair estimate of the total cost. If the matrix is reducible and one wants to apply the "eigenvalue decomposition" method (line 2) in the

[4] We thank the reviewer who suggested this.

first phase, some preprocessing is necessary to bring the matrix into block diagonal normal form (cf. Section 6.1) and then the algorithm must be slightly modified to accommodate it. It is also straightforward to modify the algorithm to accept as input any matrix and perform internally the tests for symmetry and positive semidefiniteness and rank detection.

Algorithm 2 IREVA

Input: Symmetric positive semidefinite matrix $A \in \mathbb{R}_+^{m\times m}$ of rank-r,
Method $\in$ {“eigenvalue decomposition”, “Wedderburn rank reduction”}
Output: Factor $W \in \mathbb{R}_+^{m\times r}$ s.t. $A = WW^\top$ or warning message
1: **if** Method = ”eigenvalue decomposition” **then**
2: Compute eigenvalue decomposition $A = QEQ^\top$, $Q \in \mathbb{R}^{m\times r}$, $E \in \mathbb{R}_+^{r\times r}$ with $Q_{:,1} \geq 0$
3: $V \leftarrow Q\sqrt{E}$
4: **else if** Method = ”Wedderburn rank reduction” **then**
5: Pick a positive vector $x \in \mathbb{R}^n$
6: $\omega \leftarrow x^\top Ax \neq 0$
7: $V_{:,1} \leftarrow \frac{1}{\sqrt{\omega}}Ax$
8: $B \leftarrow A - V_{:,1}V_{:,1}^\top$ //rank(B) = $r - 1$
9: Compute semidefinite Cholesky decomposition $PBP^\top = LL^\top$,
$P \in \mathbb{R}^{m\times m}$ permutation matrix, $L \in \mathbb{R}^{m\times(r-1)}$ lower triangular
10: $V_{:,2:r} \leftarrow P^\top L$
11: **end if**
12: $\mathcal{Z} \leftarrow \{j : j \in \{1, 2, \ldots, m\}, V_{j,:} \neq \mathbf{0}_{1\times r}\}$
13: **for all** $j \in \mathcal{Z}$ **do**
14: $\widetilde{V}_{j,:} = V_{j,:}/\|V_{j,:}\|_2$ //Normalization
15: **end for**
16: **for** $i = 1$ to r **do** //Identify extreme vectors
17: $\xi_i \leftarrow \arg\min_j \widetilde{V}_{j,1},\ j \in \mathcal{Z}$ //Index of row with minimum first component
18: $v^{(i)} \leftarrow \widetilde{V}_{\xi_i,:}^\top$ //i-th extreme ray
19: $c \leftarrow \widetilde{V}_{\mathcal{Z},:}v^{(i)}$ //Cosines of angles between rows of $\widetilde{V}$ and $v^{(i)}$
20: $\mathcal{Z} \leftarrow \{j : j \in \mathcal{Z}, c_j = 0\}$ //Keep rows orthogonal to $v^{(i)}$
21: **if** $|\mathcal{Z}| < r - i$ **then**
22: **return** “no full rank $r \times r$ diagonal principal submatrix exists”
23: **end if**
24: **end for**
25: $R \leftarrow \left[v^{(1)} \mid v^{(2)} \mid \ldots \mid v^{(r)}\right]$
26: $W \leftarrow VR$

To compute W directly using formula (3), it suffices to replace the last two lines of the algorithm with

$$25{:}W \leftarrow A(:, \xi)\, A(\xi, \xi)^{-1/2}.$$

Moreover, if at any iteration, i, of the loop 16 the cardinality of the index set $\mathcal{Z}$ is less than $r - i$, then it follows that there are not enough extreme vectors and hence A fails to be $\Delta_{\max}$ and the sought factorization cannot be computed with IREVA. This is monitored in line 21.

In view of the interpretation of property $\Delta_{\max}$ in terms of the associated matrix graph at the end of Section 5, if IREVA terminates successfully, vector ξ would contain the indices of its maximum independent set. This welcome side effect must be tempered, however, since it depends on the edge weights satisfying the constraints described earlier. In particular, even if a matrix is $\Delta_{\max}$, it will not necessarily remain so if all its nonzero elements are turned into ones (in fact, its rank would most likely increase) therefore the corresponding graph will not be amenable to our analysis.

7. Examples

We seek the factorization of symmetric matrix $A \in \mathbb{R}_+^{6\times 6}$ of rank-3 that contains a diagonal principal submatrix of equal rank,

$$A = \begin{bmatrix} 13 & 15 & 12 & 0 & 10 & 14 \\ 15 & 25 & 0 & 0 & 0 & 20 \\ 12 & 0 & 37 & 4 & 30 & 9 \\ 0 & 0 & 4 & 16 & 0 & 12 \\ 10 & 0 & 30 & 0 & 25 & 5 \\ 14 & 20 & 9 & 12 & 5 & 26 \end{bmatrix}.$$

In the first step, matrix $V \in \mathbb{R}^{6\times 3}$ is computed so that $A = VV^\top$ and having a positive first column. Vector $x = [1, 1, \dots, 1]^\top$ is selected and used in the Wedderburn rank reduction formula to compute $V_{:,1} \leftarrow Ax/\sqrt{\omega} > 0$, where $\omega = x^\top Ax > 0$. In the sequel, pivoted Cholesky is applied on the semidefinite matrix $A - Axx^\top A/\omega$. This provides the next two columns of V. Next, the rows of V are normalized to obtain $\widetilde{V}$, that contains the projections of the rows of V on the unit sphere. It is

$$V = \begin{bmatrix} 3.1841 & 1.3700 & -0.9923 \\ 2.9851 & 4.0111 & 0 \\ 4.5772 & -3.4064 & -2.1086 \\ 1.5921 & -1.1848 & 3.4730 \\ 3.4826 & -2.5918 & -2.4807 \\ 4.2787 & 1.8019 & 2.1086 \end{bmatrix}, \quad \widetilde{V} = \begin{bmatrix} 0.8831 & 0.3800 & -0.2752 \\ 0.5970 & 0.8022 & 0 \\ 0.7525 & -0.5600 & -0.3467 \\ 0.3980 & -0.2962 & 0.8682 \\ 0.6965 & -0.5184 & -0.4961 \\ 0.8391 & 0.3534 & 0.4135 \end{bmatrix}$$

As per Theorem 8, the row index of the first extreme vector $v^{(1)}$ is the position of the smallest element in the first column of $\widetilde{V}$. For the next step only row vectors perpendicular to $v^{(1)}$ need to be considered. These are readily identified by the values in $\widetilde{V}v^{(1)}$, while the remaining rows are discarded.

$$v^{(1)} = \begin{bmatrix} 0.3980 \\ -0.2962 \\ 0.8682 \end{bmatrix}, \quad \widetilde{V}v^{(1)} = \begin{bmatrix} 0 \\ 0 \\ 0.1644 \\ 1 \\ 0 \\ 0.5883 \end{bmatrix}, \quad \widetilde{V} \leftarrow \begin{bmatrix} 0.8831 & 0.3800 & -0.2752 \\ 0.5970 & 0.8022 & 0 \\ 0.6965 & -0.5184 & -0.4961 \end{bmatrix}$$

The updated $\widetilde{V}$ has fewer rows. The one with smallest first coordinate is identified, and this is the second extreme vector $v^{(2)}$. Following the same procedure, only rows of $\widetilde{V}$ that are perpendicular to $v^{(2)}$ are to be kept for the next step.

$$v^{(2)} = \begin{bmatrix} 0.5970 \\ 0.8022 \\ 0 \end{bmatrix}, \quad \widetilde{V} v^{(2)} = \begin{bmatrix} 0.8321 \\ 1 \\ 0 \end{bmatrix}, \quad \widetilde{V} \leftarrow \begin{bmatrix} 0.6965 & -0.5184 & -0.4961 \end{bmatrix}$$

Any row of $\widetilde{V}$ in the last step should be perpendicular to all extreme rays already found. In this example there is only one left, that corresponds to $v^{(3)}$.

$$v^{(3)} = \begin{bmatrix} 0.6965 \\ -0.5184 \\ -0.4961 \end{bmatrix}, \quad R \leftarrow \begin{bmatrix} v^{(1)} \mid v^{(2)} \mid v^{(3)} \end{bmatrix} = \begin{bmatrix} 0.3980 & 0.5970 & 0.6965 \\ -0.2962 & 0.8022 & -0.5184 \\ 0.8682 & 0 & -0.4961 \end{bmatrix}$$

Finally, V is multiplied with the orthogonal matrix R containing all normalized extreme vectors and the result is $W \geq 0$ such that $A = WW^\top$ holds.

$$W \leftarrow V\,R = \begin{bmatrix} 0 & 3 & 2 \\ 0 & 5 & 0 \\ 1 & 0 & 6 \\ 4 & 0 & 0 \\ 0 & 0 & 5 \\ 3 & 4 & 1 \end{bmatrix} \geq 0$$

The above step can be replaced by the direct computation of W from A immediately after the indices of the extreme vectors among the rows of $\widetilde{V}$ have been identified, as described in [34, Thm. 4] and explained in Section 6.3 above. Since $\xi = [4, 2, 5]$, it follows that

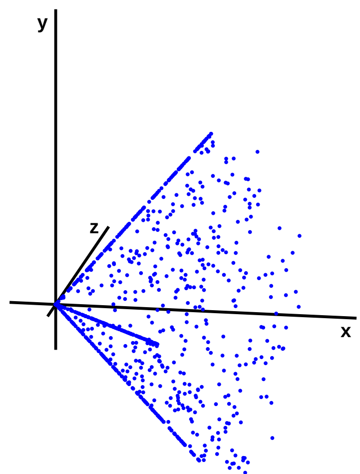
$$W \leftarrow A(:, \xi)\, A(\xi, \xi)^{-1/2} = \begin{bmatrix} 0 & 15 & 10 \\ 0 & 25 & 0 \\ 4 & 0 & 30 \\ 16 & 0 & 0 \\ 0 & 0 & 25 \\ 12 & 20 & 5 \end{bmatrix} \begin{bmatrix} 1/4 & 0 & 0 \\ 0 & 1/5 & 0 \\ 0 & 0 & 1/5 \end{bmatrix}$$

Figure 1 depicts the steps and geometrical features of IREVA for a CP matrix of rank-3 satisfying the conditions of Theorem 4.

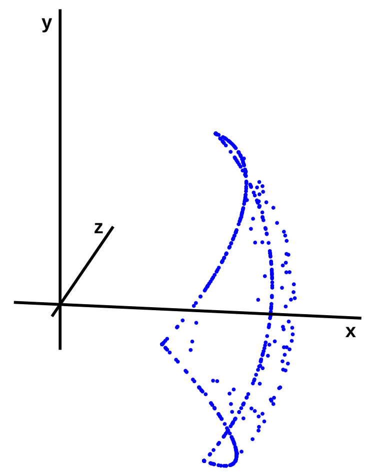
We finally note that even though the effects of finite precision were not the subject of this paper, in numerical experiments with many random matrices that satisfied the postulated conditions (Corollary 5) the computed factorization was accurate to machine precision. A more detailed analysis, however, remains to be done.

Acknowledgments

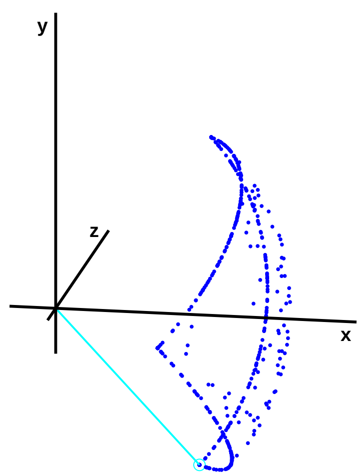
This paper is dedicated to Professor Hans Schneider. We thank Professor Moody Chu for sharing the MATLAB code from [15] and Professor Bob Plemmons for sending us a copy of [21].



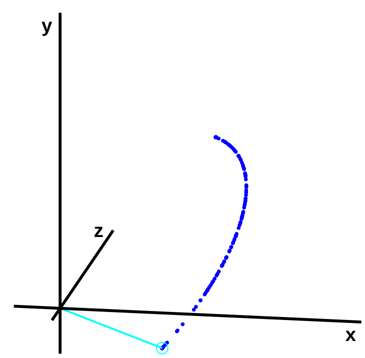
(a) Rows of V in $\mathbb{R}^3$ before rotation.



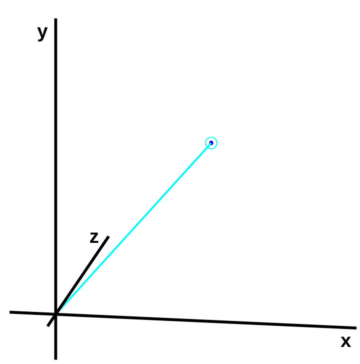
(b) Projection on the unit sphere. (Rows of $\widetilde{V}$)



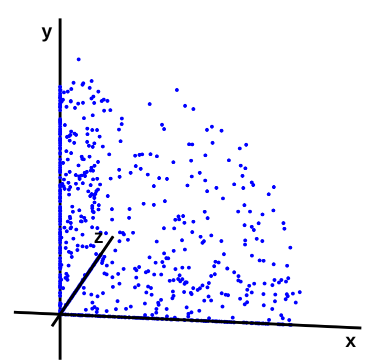
(c) Identification of the first extreme vector $v^{(1)}$



(d) Identification of the second extreme vector $v^{(2)}$



(e) Identification of the third extreme vector $v^{(3)}$



(f) Rotated data (rows of $W = VR$) falls in the nonnegative octant $\mathbb{R}^3_+$

Figure 1: Depiction of steps of Algorithm 2 for a rank-3 CP matrix.

This work was based on a talk presented at the ALA2010 Conference in Novi Sad by the second author who would like to thank Professor Lilja Cvetkovic and other organizers for their invitation. We are also grateful to the reviewers and the editor for their critical reading of the manuscript and their suggestions which helped improve various aspects of the paper and IREVA as well as for bringing reference [34] to our attention. The second author would also like to express his gratitude to INRIA Rennes and the Université de Rennes I for hosting him while this work was completed and Bernard Philippe for helpful discussions on aspects of this paper.

Appendix A. Proof of Theorem 1

Proof. Let $W = U_W\Sigma_W V_W^\top$ and $H = U_H\Sigma_H V_H^\top$ be the thin SVD's of the two factors. Thus V_W and V_H are orthogonal and U_W and U_H have orthonormal columns. Therefore $U_W\Sigma_W^2U_W^\top = U_H\Sigma_H^2U_H^\top$ and if we set $X = U_W\Sigma_W U_W^\top$ it will be the unique symmetric positive semidefinite square root of A, which is also symmetric positive semidefinite (cf. [23, Section 4.2.10]). From uniqueness, it also follows that $X = U_H\Sigma_H U_H^\top$. Let now $R = V_H U_H^\top U_W V_W^\top$. Then $HR = U_H\Sigma_H V_H^\top V_H U_H^\top U_W V_W^\top$ and after some algebra it follows that $HR = W$. Moreover, R is orthogonal. To prove this it suffices to show that $U_H^\top U_W$ is orthogonal: From the above $U_H\Sigma_H U_H^\top U_W = U_W\Sigma_W$ hence $U_H^\top U_W = \Sigma_H^{-1}U_H^\top U_W\Sigma_W$. So

$$
\begin{aligned}
(U_H^\top U_W)(U_H^\top U_W)^\top &= \Sigma_H^{-1}U_H^\top U_W\Sigma_W(\Sigma_H^{-1}U_H^\top U_W\Sigma_W)^\top = \Sigma_H^{-1}U_H^\top U_W\Sigma_W\Sigma_W U_W^\top U_H\Sigma_H^{-1} \\
&= \Sigma_H^{-1}U_H^\top(U_H\Sigma_H^2U_H^\top)U_H\Sigma_H^{-1} = I,
\end{aligned}
$$

thus $U_H^\top U_W$ is orthogonal and so R is also orthogonal. □

References

[1] S. Ambikkumar and S. W. Drury. A reverse Hadamard inequality. *Linear Algebra and its Applications*, 247:83–95, November 1996.

[2] A. Berman and R.J. Plemmons. Inverses of nonnegative matrices. *Linear Algebra and Multilinear Algebra*, 2:161–172, 1974.

[3] A. Berman and R.J. Plemmons. *Nonnegative Matrices in the Mathematical Sciences*. SIAM, Philadelphia, 1994.

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

[5] A. Berman and N. Shaked-Monderer. *Completely Positive Matrices*. World Scientific, 2003.

[6] S. Burer. On the copositive representation of binary and continuous nonconvex quadratic programs. *Mathematical Programming*, 120:479–495, 2009. 10.1007/s10107-008-0223-z.

[7] S. Burer, K.M. Anstreicher, and M. Doer. The difference between 5x5 doubly nonnegative and completely positive matrices. *Linear Algebra and its Applications*, 431(9):1539 – 1552, 2009.

[8] S.L Campbell and G.D Poole. Computing nonnegative rank factorizations. *Linear Algebra and its Applications*, 35:175–182, 1981.

[9] M. Catral, L. Han, M. Neumann, and R. Plemmons. On reduced rank nonnegative matrix factorization for symmetric nonnegative matrices. *Linear Algebra Appl.*, 393(1):107–126, 2004.

[10] M.T. Chu, R.E. Funderlic, and G.H. Golub. A rank-one reduction formula and its applications to matrix factorizations. *SIAM Review*, 37(4):512–530, 1995.

[11] M.T. Chu, R.E. Funderlic, and G.H. Golub. Rank modifications of semi-definite matrices with applications to secant updates. *SIAM J. Matrix Anal. Appl.*, 20(2):428-436, 1998.

[12] A. Cichocki, R. Zdunek, A.-H. Phan, and S. Amari. *Nonnegative matrix and tensor factorizations: Applications to exploratory multi-way data analysis*. John Wiley, 2009.

[13] J.E. Cohen and U.G. Rothblum. Nonnegative ranks, decompositions, and factorizations of nonnegative matrices. *Lin. Alg. Appl.*, 190:149–168, 1993.

[14] C. Ding, X. He, and H.D. Simon. On the equivalence of nonnegative matrix factorization and spectral clustering. In *Proc. SIAM Int'l Conf. Data Mining (SDM'05)*, pp. 606–610, 2005.

[15] B. Dong, M.M. Lin, and M.T. Chu. Nonnegative rank factorization via rank reduction. manuscript available from http://www4.ncsu.edu/ mtchu/Research/Papers/NRF04d.pdf.

[16] D.L. Donoho and V. Stodden. When does non-negative matrix factorization give a correct decomposition into parts? In Sebastian Thrun, Lawrence K. Saul, and Bernhard Schölkopf, editors, *NIPS*. MIT Press, 2003.

[17] I.S Duff and B. Ucar. Combinatorial problems in solving linear systems. In U Naumann, O. Schenk, H.D. Simon, and S. Toledo, editors, *Combinatorial Scientific Computing*, Dagstuhl, Germany, 2009. Dagstuhl Seminar Proceedings, 09061, Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik, Germany.

[18] I.S. Duff, A. M. Erisman, and J. K. Reid. *Direct Methods for Sparse Matrices*. Clarendon Press, Oxford, 1989.

[19] F.R. Gantmacher. *The Theory of Matrices*, volume 2. Chelsea Pub. Co., New York, N.Y., 2nd edition, 1959.

[20] M.R. Garey and D.S. Johnson. *Computers and Intractability: A Guide to the Theory of NP-Completeness*. W.H. Freeman, New York, N.Y., 1979.

[21] N. Gillis and R.J. Plemmons. Dimensionality reduction, classification, and spectral mixture analysis using nonnegative underapproximation. In S. Shen and P. Lewis, editors, *Proc. SPIE Conf. on Defense, Security and Sensing: Vol. 7695: Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XVI*, 2010.

[22] N. Gillis and F. Glineur. Using underapproximations for sparse nonnegative matrix factorization. *Pattern Recognition*, 43(4):11, 2010.

[23] G.H. Golub and C.F. Van Loan. *Matrix Computations*. The Johns Hopkins University Press, Baltimore, 3d edition, 1996.

[24] L. J. Gray and D. G. Wilson. Nonnegative factorization of positive semidefinite nonnegative matrices. *Linear Algebra and its Applications*, 31:119 – 127, 1980.

[25] R.C. Henry. Current factor analysis receptor models are ill-posed. *Atmospheric Environment (1967)*, 21(8):1815 – 1820, 1987.

[26] R.C. Henry. Receptor model applied to patterns in space (RMAPS) part I - model description. *J. Air Waste Managem. Assoc.*, 47:216–219, 1997.

[27] N.J. Higham. Analysis of the Cholesky decomposition of a semi-definite matrix. In M. G. Cox and S. J. Hammarling, editors, *Reliable Numerical Computation*, pages 161–185, 1990.

[28] M. Kaykobad. On nonnegative factorization of matrices. *Linear Algebra and its Applications*, 96:27–33, 1987.

[29] B. Klingenberg, J. Curry, and A. Dougherty. Non-negative matrix factorization : Ill-posedness and a geometric algorithm. *Pattern Recognition*, 42:918–928, 2009.

[30] B. Levin. On calculating maximum rank one underapproximations for positive arrays, http://biostats.bepress.com/columbiabiostat/papers/art10, Columbia University Biostatistics Technical Report Series. Working Paper 10. 1985.

[31] C. Lucas. LAPACK-style codes for level 2 and 3 pivoted Cholesky factorizations. Numerical Analysis Report No. 442 and LAPACK Working Note 161, Manchester Centre for Computational Mathematics, 2004.

[32] T.L. Markham. Factorizations of completely positive matrices. *Proc. Camb. Phil. Soc.*, 69:53–58, 1971.

[33] P. Paatero, P.K. Hopke, X.-H. Song, and Z. Ramadan. Understanding and controlling rotations in factor analytic models. *Chemometrics and Intelligent Laboratory Systems*, 60(1-2):253 – 264, 2002.

[34] N. Shaked-Monderer. A note on upper bounds on the cp-rank. *Linear Algebra and its Applications*, 431(12):2407–2413, 2009.

[35] L.B. Thomas. Problem 73-14, Rank factorization of nonnegative matrices. *SIAM Rev.*, 16(3):393–394, July 1974.

[36] A. Vandendorpe, N.-D. Ho, S. Vanduffel, and P.V. Dooren. On the parameterization of the CreditRisk+ model for estimating credit portfolio risk. *Insurance: Mathematics and Economics*, 42:736–745, 2008.

[37] S.A. Vavasis. On the complexity of nonnegative matrix factorization. *SIAM J. Optim.*, 20:1364–1377, 2009.

[38] J.H.M. Wedderburn. *Lectures on Matrices*. Colloqium Publications, Amer. Math. Soc., New York, 1934.