On Reduced Rank Nonnegative Matrix Factorization for Symmetric Nonnegative Matrices

M. Catral
Department of Mathematics
University of Connecticut
Storrs, Connecticut 06269–3009

Lixing Han *
Department of Mathematics
University of Michigan–Flint
Flint, Michigan 48502

Michael Neumann†
Department of Mathematics
University of Connecticut
Storrs, Connecticut 06269–3009

R. J. Plemmons ‡
Departments of Computer Science and Mathematics
Wake Forest University
Winston–Salem, North Carolina 27109.

Abstract

Let $V \in \mathbb{R}^{m,n}$ be a nonnegative matrix. The *nonnegative matrix factorization* (NNMF) problem consists of finding nonnegative matrix factors $W \in \mathbb{R}^{m,r}$ and $H \in R^{r,n}$ such that $V \approx WH$. Lee and Seung proposed two algorithms which find nonnegative W and H such that $\|V - WH\|_F$ is minimized. After examining the case in which $r = 1$ about which a complete characterization of the solution is possible, we consider the case in which $m = n$ and V is symmetric. We focus on questions concerning when the best approximate factorization results in the product WH **being symmetric** and on cases in which the best approximation **cannot be a symmetric matrix**. Finally, we show that the class of positive semidefinite symmetric nonnegative matrices V generated via a Soules basis admit for every $1 \leq r \leq \text{rank}(V)$, a nonnegative factorization WH which coincides with the best approximation in the Frobenius norm to V in $\mathbb{R}^{n,n}$ of rank not exceeding r.

An example of applications in which NNMF factorizations for nonnegative symmetric matrices V arise is video and other media summarization technology where V is obtained from a distance matrix. We further mention that a vehicle for our results here is the Khun–Tucker conditions.

*Research was supported by a Rackham Fellowship of the University of Michigan.
†This author's research was supported in part by NSF grant No. DMS0201333.
‡His research was supported in part by the Air Force Office of Scientific Research under grant F49620–02–1–0107, and by the Army Research Office under grant DAAD19–00–1–0540.

1 Introduction

Given a nonnegative matrix V, the problem we consider here is to find nonnegative matrix factors W and H such that

$$V \approx WH. \tag{1.1}$$

This is the so–called *nonnegative matrix factorization* (NNMF) *problem* which was recently proposed by Paatero and Tapper [13, 14] and Lee and Seung [8]. Lee's and Seung's idea was to use the NNMF techniques to find a set of basis functions to represent image data where the basis functions enable the identification and classification of the intrinsic "parts" that make up the object being imaged by multiple observations. They showed that NNMF facilitates the analysis and classification of data from image or sensor articulation databases made up of images showing a composite object in many articulations, poses, or observation views. Given an initial database expressed as an $m \times n$ matrix V, where each column of V is an m–dimensional nonnegative vector of the original database (n vectors), the standard NNMF problem is to find two new reduced–dimensional matrices W and H, in order to approximate the original matrix V by the product WH. The dimensions of matrices W and H are $m \times r$ and $r \times n$, respectively. Usually, it is desirable that the number of columns in the new (basis) matrix W is chosen so that $r << m$. The choice of r is generally application–dependent and may also depend upon the characteristics of the particular database within the application.

In [9, 10], Lee and Seung suggested two useful algorithms which are easy to implement and have been successfully employed in many areas of applications, such as image processing, text information retrieval, and machine learning (see, for instance, [9, 11, 15]). We comment that the NNMF problem occurs in the context of symmetric nonnegative matrices V, for example, in video and other media summarization technology where V is the so–called (*distance*) *similarity matrix*, which is symmetric, see Cooper and Foote [4].

One of the two algorithms of Lee and Seung is based on the following formulation of the NNMF problem:

Suppose that $V \in \mathbb{R}^{m,n}$ is nonnegative. Find W and H which solve

$$\min \ f(W, H) \quad \text{subject to} \quad W \geq 0 \text{ and } H \geq 0, \tag{1.2}$$

where

$$f(W,H) \ := \ \frac{1}{2}\ \|V - WH\|_F^2 \tag{1.3}$$

and where $W \in \mathrm{I\!R}^{\mathrm{m,r}}$ *and* $H \in \mathrm{I\!R}^{\mathrm{r,n}}$ *and with* $r \leq \min\{m,n\}$.

In this paper we shall refer to any pair W and H which minimizes the objective function f over the feasible region $\{(W,H) \mid W \geq 0, H \geq 0\}$ as a *global minimizer* of Problem (1.2). Alternatively, we shall call any such pair a *solution* to the problem (1.2). We notice that if (W,H) is a solution, then so is $(\alpha W, H/\alpha)$ for any $\alpha > 0$. Thus the solutions to (1.2) are not unique.

When the nonnegativity constraints are not imposed, the best approximation to a given matrix by a lower rank matrix has been studied extensively in the literature. A well known result due to Eckart and Young is given in the following lemma, which plays an important role in some of our analysis. We shall assume, without loss of generality, that $m \geq n$.

LEMMA 1.1 (Eckart-Young Theorem, see [7]) *Let* $V \in \mathrm{I\!R}^{\mathrm{m}\times\mathrm{n}}$ *have a singular value decomposition*

$$V \ = \ P\Sigma Q^T, \quad \Sigma \ = \ \mathrm{diag}(\sigma_1, \sigma_2, \cdots, \sigma_n) \in \mathbb{R}^{m,n},$$

where $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n \geq 0$ *are the singular values of* V *and* $P \in R^{m,m}$ *and* $Q \in \mathbb{R}^{n,n}$ *are orthogonal matrices. Then for* $1 \leq r \leq n$, *the matrix*

$$B_r \ = \ P\ \mathrm{diag}(\sigma_1, \ldots, \sigma_r, \underbrace{0, \ldots, 0}_{n-r})\ Q^T, \tag{1.4}$$

is a global minimizer of the optimization problem

$$\min \ \left\{ \|V - B\|_F^2 \ \mid \ B \in \mathrm{I\!R}^{\mathrm{m,n}}\ ,\ \mathrm{rank}(B) \leq r \right\} \tag{1.5}$$

with the corresponding minimum value $\sum_{i=r+1}^{n} \sigma_i^2$. *Moreover, if* $\sigma_r > \sigma_{r+1}$, *then* B_r *is the unique global minimizer.*

In this paper, we will study some theoretical properties of the NNMF problem when the nonnegative V is also symmetric. In this case, of course, $m = n$ and the following comments are in order: (1) When rank(V)=1, then of course V has an exact factorization with $WH = WW^T$, where $W \geq 0$ and $W \in \mathbb{R}^{n,1}$. (2) In the case of rank$(V) = 2$, then Ambikkumar and Drury [1, Proposition] show that V continues to have an exact factorization $V = WH$, with $W, H \geq 0$ and with $W \in \mathbb{R}^{n,2}$ and $H \in \mathbb{R}^{2,n}$. However, as under the assumptions, V need not be positive semidefinite as shown by taking:

$$V = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix},$$

we see that even when an exact factorization of V exists, H need not equal W^T. (3) For nonnegative positive semidefinite matrices V, the problem of the minimal r for which $V = WW^T$, $W \geq 0$ and $W \in \mathbb{R}^{n,r}$ has been much studied in the literature. This is the so called *cp–rank* problem and the reader is referred to the recent book by Berman and Shaked–Monderer [3] on completely positive matrices and the references within.

In view of the above, we raise in this paper the following questions:

Question 1. *Is there a solution* (W, H) *to* (1.2) *such that* $H = W^T$*, where* $W \in \mathbb{R}^{n,r}$ *and* $H \in \mathbb{R}^{r,n}$?

Question 2. *Is there a solution* (W, H) *to* (1.2) *such that the product* WH *is symmetric* ?

We have already mentioned that the solutions to the NNMF problem are not unique. So we now raise the additional question of:

Question 3. *When do all solutions to* (1.2) *yield a unique product* WH ?

Conditions resulting in uniqueness in the special case of equality, $V = WH$, have been recently studied by Donoho and Stodden [5], using cone theoretic techniques (See also Chapter 1 in Berman and Plemmons [2]).

We shall see that when $r = 1$, the answers to the first two questions are in the affirmative and we are also able to answer Question 3. However, when $r > 1$, then even the answer to Question 2 can be negative, meaning that any solution that best approximates

V in the sense of (1.2) does **not** satisfy $WH = (WH)^T$.

The paper is organized as follows. In Section 2 we use the Kuhn–Tucker conditions to characterize solutions to (1.2). We shall show that generally when these conditions hold, it is always true that $\|WH\|_2 \leq \|V\|_2$ for a solution (W, H) to (1.2). In Section 3 we consider the case $r = 1$. In Section 4 we study the case when V is symmetric and $r > 1$. Finally, in Section 5 we answer Questions 2 and 3 for the class of symmetric nonnegative matrices V which are generated by orthogonal bases which are known as *Soules bases*, see Soules [16] and Elsner, Nabben, and Neumann [6]. We show that if a symmetric nonnegative matrix $V \in \mathbb{R}^{n,n}$, with distinct eigenvalues $\lambda_1 > \ldots > \lambda_n \geq 0$, is generated by a Soules basis, then for each $1 \leq r \leq n$, the product WH generated from a global minimizer for Problem (1.2) coincides with the global minimizer for V furnished via the Eckart–Young Theorem given in Lemma 1.1.

2 Initial Analysis Using the Kuhn–Tucker Conditions

In order to understand better the properties of solutions to (1.2), we shall appeal to the usual Kuhn–Tucker conditions, see Nocedal and Wright [12].

Assume then that (W, H) is a solution to (1.2). To begin with, the gradient of f with respect to W is given by

$$\nabla_W f \;=\; -VH^T + WHH^T,$$

while the gradient of f with respect to H is given by

$$\nabla_H f \;=\; -W^T V + W^T WH.$$

Now according to the Kuhn–Tucker conditions for constrained optimization, there exist Lagrange multipliers $\nu \in \mathbb{R}^{\mathrm{m,r}}$ and $\mu \in \mathbb{R}^{\mathrm{r,n}}$ such that

$$-VH^T + WHH^T \;=\; \nu, \tag{2.1}$$

$$-W^T V + W^T WH \;=\; \mu, \tag{2.2}$$

and

$$\nu_{i,j} W_{i,j} \;=\; 0,\; \mu_{st} H_{st} \;=\; 0,\; W_{i,j} \;\geq\; 0,\; H_{st} \;\geq\; 0,\; \nu_{i,j} \;\geq\; 0,\; \mu_{st} \geq 0, \tag{2.3}$$

for all $1 \le i \le m$, $1 \le j \le r$, $1 \le s \le r$, and $1 \le t \le n$.

We note that if $\nu = 0$ and $\mu = 0$, a situation which holds true, for example, when $W > 0$ and $H > 0$, then the Kuhn–Tucker conditions reduce to:

$$WHH^T \;=\; VH^T \quad \text{and} \quad W^TWH \;=\; W^TV. \tag{2.4}$$

If we further assume that W and H have full rank, then (2.4) can be rewritten as

$$W \;=\; VH^T(HH^T)^{-1} \quad \text{and} \quad H \;=\; (W^TW)^{-1}W^TV. \tag{2.5}$$

This shows that when $\nu = 0$ and $\mu = 0$ and W and H are of full rank, then we further have that

$$W \;=\; VH^{\dagger} \quad \text{and} \quad H \;=\; W^{\dagger}V, \tag{2.6}$$

in which case we see that

$$WH \;=\; VH^{\dagger}H \quad \text{and} \quad WH \;=\; WW^{\dagger}V \tag{2.7}$$

so that, in particular,

$$\mathcal{R}(WH) \;\subseteq\; \mathcal{R}(V) \quad \text{and} \quad \mathcal{N}(V) \;\subseteq\; \mathcal{N}(WH), \tag{2.8}$$

where in displays (2.6)–(2.8) we have used the notations $(\cdot)^{\dagger}$, $\mathcal{R}(\cdot)$, and $\mathcal{N}(\cdot)$ to denote, respectively, the Moore–Penrose generalized inverse, the range, and nullspace of a matrix. Moreover, either equalities in (2.7) have the implication that

$$\|WH\|_2 \;\le\; \|V\|_2. \tag{2.9}$$

In the next section, we shall show that the Lagrange multipliers ν and μ are always zero when $r = 1$ which enables us to answer the three questions we posed in the introduction in the rank one case.

3 The Rank One Case

The rank one NNMF problem, i.e., $r = 1$ in (1.2)–(1.3) can be written as

$$\min f(x, y) \;=\; \min\; \frac{1}{2}\left\|V - xy^T\right\|_F^2, \quad \text{subject to} \;\; x \ge 0 \text{ and } y \ge 0, \tag{3.1}$$

with $x \in \mathbb{R}^{m}$ and $y \in \mathbb{R}^{n}$, respectively. As we shall see in this section, when $r = 1$, a rather full analysis is possible for the NNMF problem.

We begin with the following theorem:

THEOREM 3.1 *Let (x, y) be a solution to (3.1). Then*

$$(y^T y)x - Vy \;=\; 0 \quad \text{and} \quad (x^T x)y - V^T x \;=\; 0. \tag{3.2}$$

Proof: We divide the proof into two cases, depending on $\sigma_1 = 0$ or $\sigma_1 > 0$, where σ_1 is the leading singular value of V.

Case 1: $\sigma_1 = 0$. In this case, $V = 0$ and therefore the minimum value of problem (3.1) is 0. If (x, y) is a solution to (3.1), then either $x = 0$ or $y = 0$, since otherwise $f(x, y) > 0$, which means that (x, y) can not be a solution. But $x = 0$ or $y = 0$ implies that (3.2) must hold.

Case 2: $\sigma_1 > 0$. For such a V, by the singular value decomposition there exist non-negative nonzero vectors a and b such that $\|a\|_2 = 1$, $\|b\|_2 = 1$, and

$$Vb \;=\; \sigma_1 a \quad \text{and} \quad V^T a \;=\; \sigma_1 b. \tag{3.3}$$

a and b are usually called the right and the left singular vectors of V. By the Eckart–Young Theorem, the pair $x^* = \sigma_1 a$ and $y^* = b$ is a global minimizer of the unconstrained optimization problem

$$\min f(x, y) \;=\; \min \; \frac{1}{2} \left\| V - xy^T \right\|_F^2 . \tag{3.4}$$

Notice that $x^* \geq 0$ and $y^* \geq 0$. Hence (x^*, y^*) must be a global minimizer of Problem (3.1) and, moreover, the global minimum value of Problem (3.1) is the same as the global minimum value of Problem (3.4). Therefore, any solution (x, y) to (3.1) (that is, a global minimizer of (3.1)) must also be a global minimizer of (3.4). This implies $\nu = 0$ and $\mu = 0$ by invoking the optimality conditions for the unconstrained problem (3.4), that is,

$$(y^T y)x - Vy \;=\; 0 \quad \text{and} \quad (x^T x)y - V^T x \;=\; 0.$$

□

COROLLARY 3.2 *If the two largest singular values of* V *satisfy that*

$$\sigma_1 > \sigma_2, \tag{3.5}$$

then the solutions to (3.1) *yield the unique product*

$$\sigma_1 ab^T,$$

where a *and* b *are as in* (3.3).

Proof: Since $\sigma_1 > \sigma_2$, by the Eckart–Young theorem we know that

$$B_1 = \sigma_1 ab^T$$

is the unique minimizer of (1.5). This implies that any solution (x, y) to (3.1) yields the same product $xy^T = B_1$. □

We are now ready to answer the questions we raised in Section 1 for the rank one case. Suppose that $V \in \mathbb{R}^{n,n}$ is both a symmetric and a nonnegative matrix. Then if $\sigma_1 = 0$, the answers to those questions are obvious. We consider the case when $\sigma_1 > 0$. Now since V is symmetric we have that $a = b$ where a and b are as in the proof of Theorem (3.1). The pair $W = x = \sqrt{\sigma_1}a$ and $H = y^T = \sqrt{\sigma_1}a^T$ is a solution to the rank one NNMF problem, which satisfies that $H = W^T$ and thus implies that the answer to Question 1 is in the affirmative. It follows then that the answer to Question 2 is also affirmative in this case.

To answer Question 3, let σ_1 and σ_2 be the two leading singular values of a symmetric and nonnegative matrix V. Then if $\sigma_1 > \sigma_2$, which is equivalent to the Perron eigenvalue of V being strictly greater than the absolute of all other eigenvalues of V, the solutions to (3.1) yield the unique product $\sigma_1 aa^T$, where a is as in (3.3). On the other hand, if $\sigma_1 = \sigma_2$, then we can show that the solutions to (3.1) do not yield a unique product. We illustrate this in the following example.

Consider the following symmetric and nonnegative matrix

$$V = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}. \tag{3.6}$$

For this matrix, both the pair $W_1 = x_1 = [0, 1]^T$ and $H_1 = y_1^T = [0, 1]$ and the pair $W_2 = x_2 = [0.5, 0.5]^T$ and $H_2 = y_2^T = [0.5, 0.5]$ are solutions to Problem (3.1). However, the second pair gives a symmetric product while the first does not. A quick check shows that $\sigma_1 = \sigma_2 = 1$ for this matrix V.

4 When V Is Symmetric

In this section, we study Questions 1 and 2 when $r > 1$. We begin with two examples.

EXAMPLE 1 Consider the matrix

$$\tilde{V} = \begin{bmatrix} 0.1000 & 1.100 & 0.1000 & 0.1000 \\ 1.100 & 0.1000 & 1.100 & 0.1000 \\ 0.1000 & 1.100 & 0.1000 & 1.100 \\ 0.1000 & 0.1000 & 1.100 & 0.1000 \end{bmatrix}.$$

We now let

$$V = \tilde{V} - \tau I = \begin{bmatrix} 0.09098 & 1.100 & 0.1000 & 0.1000 \\ 1.100 & 0.09098 & 1.100 & 0.1000 \\ 0.1000 & 1.100 & 0.09098 & 1.100 \\ 0.1000 & 0.1000 & 1.100 & 0.09098 \end{bmatrix}.$$

Here

$$\tau = \frac{\sqrt{5} - 2.2}{4}.$$

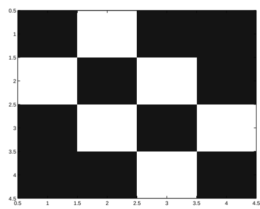
Applying the first Lee–Seung algorithm in [10] (based on the Euclidean distance) for computing the NNMF approximation to V, starting with a random positive matrix $W_0 \in \mathbb{R}^{4,3}$ normalized so that the sum of each of its columns is 1 and taking a random positive matrix $H_0 \in \mathbb{R}^{3,4}$, we obtain that

$$V \approx WH = \begin{bmatrix} 0.49817 & 0.031714 & 0.0038670 \\ 0.037361 & 0.022824 & 0.60536 \\ 0.42221 & 0.92820 & 0.0023015 \\ 0.042255 & 0.017261 & 0.38847 \end{bmatrix} \begin{bmatrix} 0.13900 & 2.1962 & 0.20404 & 0.12911 \\ 0.040412 & 0.18608 & 0.00048985 & 1.1261 \\ 1.3502 & 0.0084755 & 2.0975 & 0.13091 \end{bmatrix}$$

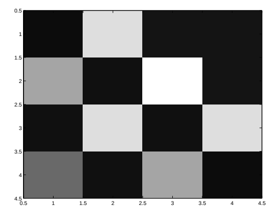
$$= \begin{bmatrix} 0.075748 & 1.1000 & 0.10977 & \boxed{0.10054} \\ 0.82346 & 0.091429 & 1.2774 & 0.10977 \\ 0.099304 & 1.1000 & 0.091429 & 1.1000 \\ \boxed{0.53108} & 0.099304 & 0.82346 & 0.075748 \end{bmatrix} \neq (WH)^T,$$

with $\|V - WH\|_F^2 = 0.3709$.

We note that WH is not a symmetric matrix, while V is symmetric. We can illustrate this example using gray–scale paneling:



True panel



NNMF approximation panel

A consequence of Theorem 4.3 will allow us to prove that, indeed, V in this example **does not** possess a solution (W, H) for the NNMF problem (1.2) which gives a symmetric approximation WH to V.

EXAMPLE 2 Let V be the following positive symmetric matrix:

$$V = \begin{bmatrix} \frac{1}{6} & \frac{2}{3} & \frac{1}{6} \\ \frac{2}{3} & \frac{1}{6} & \frac{1}{6} \\ \frac{1}{6} & \frac{1}{6} & \frac{2}{3} \end{bmatrix}.$$

It can be checked that the following pair (W, H), where

$$W = \begin{bmatrix} \frac{5}{4\sqrt{3}} & \frac{1}{2\sqrt{3}} \\ \frac{5}{4\sqrt{3}} & \frac{1}{2\sqrt{3}} \\ \frac{1}{2\sqrt{3}} & \frac{2}{\sqrt{3}} \end{bmatrix} \quad \text{and} \quad H = \begin{bmatrix} \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & 0 \\ 0 & 0 & \frac{1}{\sqrt{3}} \end{bmatrix},$$

is a solution to the NNMF problem (1.2) with $r = 2$ and that the product WH is symmetric.

In the following theorem we give sufficient conditions for a nonnegative symmetric matrix V to always have a symmetric approximate nonnegative factorization.

THEOREM 4.1 *Let $V = QDQ^T$ be a nonnegative symmetric matrix of rank t, $t \leq n$, where $Q \in \mathbb{R}^{n,n}$ is orthogonal and $D = \mathrm{diag}(d_1, \ldots, d_t, 0, \ldots, 0)$, with $d_i \neq -d_j$, for $i, j \in \{1, \ldots, t\}$, $i \neq j$. Let (W, H) be a solution to the NNMF problem* (1.2) *with Lagrange multipliers $\mu = 0$ and $\nu = 0$. Then the product WH is (also) symmetric.*

Proof: Let $X = Q^TWHQ$. Then $(WH)^T = WH$ if and only if $X^T = X$. We have that

$$WH \;=\; QXQ^T. \tag{4.1}$$

Since $\mu = 0$ and $\nu = 0$, equations (2.4) are satisfied and we see that

$$WH(WH)^T = V(WH)^T, \quad (WH)^TWH = (WH)^TV. \tag{4.2}$$

Substituting these relations in (4.1) we obtain that

$$XX^T \;=\; DX^T \;=\; XD, \quad X^TX \;=\; X^TD \;=\; DX. \tag{4.3}$$

Partition X and D as follows:

$$X \;=\; \begin{bmatrix} X_t & Y \\ Z & W \end{bmatrix} \tag{4.4}$$

and

$$D \;=\; \begin{bmatrix} D_t & 0 \\ 0 & 0 \end{bmatrix}, \tag{4.5}$$

where $X_t \in \mathbb{R}^{t,t}$, $W \in \mathbb{R}^{n-t,n-t}$, and $D_t =$diag$(d_1, \ldots, d_t)$.

On using the fact that $DX^T = XD$ and $X^TD = DX$ from (4.3), we have that

$$X_tD_t \;=\; D_tX_t^T, \;\; D_tZ^T \;=\; 0, \;\; ZD_t \;=\; 0,$$

and

$$X_t^TD_t \;=\; D_tX_t, \;\; D_tY \;=\; 0, \;\; Y^TD_t \;=\; 0.$$

Since rank$(D_t) = t$, we deduce that $Y = 0$ and $Z = 0$. Combining the equalities $X_tD_t = D_tX_t^T$ and $X_t^TD_t = D_tX_t$ now yields that

$$D_t(X_t^T - X_t) \;=\; (X_t - X_t^T)D_t.$$

Let $A = X_t^T - X_t$. Then the above becomes $D_tA = -AD_t$.

On computing the entries of D_tA and $-AD_t$ componentwise, we have for $i, j = 1, \ldots, t$,

$$d_ia_{i,j} \;=\; -a_{i,j}d_j.$$

Since $d_i \neq -d_j$, then we must have $a_{i,j} = 0$. That is, $A = X_t^T - X_t = 0$ and $X_t^T = X_t$. Thus, X is of the form

$$X = \begin{bmatrix} X_t & 0 \\ 0 & W \end{bmatrix}, \quad \text{with } X_t^T = X_t.$$

Finally, using $XX^T = DX^T$ and equating the lower right blocks, we get that $WW^T = 0$. This implies that $W = 0$ and $X^T = X$. □

Theorem 4.1 yields the following corollary:

COROLLARY 4.2 *If V is a nonnegative positive semidefinite matrix, then for any solution (W, H) to (1.2) with $\mu = 0$ and $\nu = 0$, WH is symmetric.*

We note that the (distance) similarity matrices $V = S$ studied in Cooper and Foote [4] in the context of video summarization, are in fact symmetric positive semidefinite. Thus Corollary 4.2 shows that the factorizations WH they obtain for such V are symmetric whenever the Lagrange multipliers are zero.

In our next result, we give necessary and sufficient conditions for the factorization WH to be symmetric under the assumption that the nonzero eigenvalues of V are mutually distinct.

THEOREM 4.3 *Let $V = QDQ^T$ be a nonnegative symmetric matrix of rank t, $t \leq n$, where $Q \in \mathbb{R}^{n,n}$ is orthogonal and $D = \operatorname{diag}(d_1, \ldots, d_t, 0, \ldots, 0)$, with $d_i \neq d_j$, for $i, j \in \{1, \ldots, t\}$, $i \neq j$. Let (W, H) be a solution to the NNMF problem (1.2) satisfying $\mu = 0$ and $\nu = 0$. Assume rank$(WH) = k \leq t$. Then WH is symmetric if and only if it is of the form $WH = QXQ^T$, where $X = \operatorname{diag}(x_1, \ldots, x_t, 0, \cdots, 0)$ with $x_i = 0$ for all but k i's and $x_i = d_i$ for $i \in S, |S| = k$.*

Proof. The "if part" is trivial. For the "if part", we let $X = Q^T WHQ$. Then since $(WH)^T = WH$, we have $X^T = X$.

Since $\mu = 0$ and $\nu = 0$, then the equations (2.4) are satisfied and we can derive

$$XX^T = DX^T = XD, \quad X^TX = X^TD = DX, \tag{4.6}$$

as in the proof of Theorem 1. Partition X and D as in (4.4) and (4.5) and use equations (4.6) to conclude that X is of the form

$$X = \begin{bmatrix} X_t & 0 \\ 0 & 0 \end{bmatrix}.$$

Using equations (4.6) and invoking the symmetry of X, we get that

$$X_t^2 = D_t X_t = X_t D_t.$$

Computing the entries of $D_t X_t$ and $X_t D_t$ componentwise, we have that for $i, j = 1, \dots, t$,

$$d_i x_{i,j} = x_{i,j} d_j.$$

Thus, since $d_i \neq d_j$ for all $i \neq j$, then we must have that $x_{i,j} = 0$ for all $i \neq j$. That is, $X_t = \mathrm{diag}(x_1, \dots, x_t)$. Now, using $X_t^2 = D_t X_t$, we have for $i = 1, \dots, t$,

$$x_i^2 = d_i x_i.$$

Since rank(X)=rank(WH) $= k$, then exactly k of these $x_i's$ are not zero. For these $x_i's$, we must have $x_i = d_i$. The proof is done. □

REMARK 4.4 (**Example 1 Revisited**) Let us return to the matrix V in Example 1. The eigenvalues of V are given by

$$\lambda_1 = 2 - \tau, \ \lambda_2 = \frac{\sqrt{5}-1}{2} - \tau, \ \lambda_3 = -\lambda_2, \ \lambda_4 = -\frac{\sqrt{5}+1}{2} - \tau.$$

The relation $\lambda_3 = -\lambda_2$ disqualifies V from availing of Theorem 4.1 which guarantees (under the assumption that the Lagrange multipliers ν and μ are both zero) a symmetric WH for every solution (W, H) to (1.2). In fact, we can show using Theorem 4.3, that for any solution (W, H) to (1.2) with $r = 3$, we have that the product WH is not symmetric.

To see this, let us first write $V = QDQ^T$, where $Q \in \mathbb{R}^{4,4}$ is orthogonal and $D = \mathrm{diag}(\lambda_1, \lambda_2, \lambda_3, \lambda_4)$. Let $\sigma_1, \sigma_2, \sigma_3$ and σ_4 be the singular values of V, arranged in decreasing order.

According to the Eckart–Young Theorem, the global minimum value of the unconstrained problem (i.e., if the nonnegativity constraints in problem (1.2) are dropped) is $\frac{1}{2}\sigma_4^2$. Running the NNMF algorithm gives that

$$\min\left\{\|V - WH\|_F^2 : W \in \mathbb{R}^{4,3}, H \in \mathbb{R}^{3,4}, W \geq 0,\ H \geq 0\right\} \ = \ \sigma_4^2 = 0.3709.$$

This implies that $\frac{1}{2}\sigma_4^2$ is also the global minimum of the NNMF problem (1.2) with $r = 3$. Thus, any solution (W, H) to the NNMF problem (1.2) with $r = 3$ also solves the unconstrained problem. Hence, we have that $\nu = 0$ and $\mu = 0$ respectively in equations (2.1) and (2.2).

Continuing, note that the eigenvalues of V are distinct and assume that WH is symmetric for some solution (W, H). Then on appealing to Theorem 4.3, we can deduce that WH must be of the form QXQ^T, where $X = \text{diag}(x_1, x_2, x_3, x_4)$, with $x_i = \lambda_i$ for $i \in S, |S| = 3$. We next run through the four possibilities QXQ^T to be :

$$Q\text{diag}(\lambda_1, \lambda_2, \lambda_3, 0)Q^T \ = \begin{bmatrix} 0.3158 & 0.7362 & 0.4638 & -0.1249 \\ 0.7362 & 0.6797 & 0.5113 & 0.4638 \\ 0.4638 & 0.5113 & 0.6797 & 0.7362 \\ -0.1249 & 0.4638 & 0.7362 & 0.3158 \end{bmatrix},$$

$$Q\text{diag}(\lambda_1, \lambda_2, 0, \lambda_4)Q^T \ = \begin{bmatrix} 0.3018 & 0.9595 & -0.0405 & 0.3108 \\ 0.9595 & 0.1847 & 1.1937 & -0.0405 \\ -0.0405 & 1.1937 & 0.1847 & 0.9595 \\ 0.3108 & -0.0405 & 0.9595 & 0.3018 \end{bmatrix},$$

$$Q\text{diag}(\lambda_1, 0, \lambda_3, , \lambda_4)Q^T \ = \begin{bmatrix} -0.1294 & 0.9638 & 0.2362 & 0.3203 \\ 0.9638 & 0.0068 & 1.1842 & 0.2362 \\ 0.2362 & 1.1842 & 0.0068 & 0.9638 \\ 0.3203 & 0.2362 & 0.9638 & -0.1294 \end{bmatrix},$$

and

$$Q\text{diag}(0, \lambda_2, \lambda_3, \lambda_4)Q^T = \begin{bmatrix} 0.2153 & 0.6405 & -0.3595 & -0.2063 \\ 0.6405 & -0.5982 & 0.4108 & -0.3595 \\ -0.3595 & 0.4108 & -0.5982 & 0.6405 \\ -0.2063 & -0.3595 & 0.6405 & -0.2153 \end{bmatrix}.$$

As none of these are nonnegative, we conclude that for V in this example there is no pair (W, H) solving (1.2) such that WH is symmetric.

5 When V Is Generated Via A Soules Basis

In this section we consider Questions 2 and 3 that we raised in Section 1 for a class of symmetric matrices that are generated via a Soules basis. A Soules basis is defined as follows.

DEFINITION 5.1 (See [16, 6]) Let $R \in \mathbb{R}^{n,n}$ be an orthogonal matrix with columns $R_1, R_2, \ldots, R_n$. The set $\{R_1, \ldots, R_n\}$ is called a *Soules basis* and R is called a *Soules matrix* if R_1 is a positive vector and if for every diagonal matrix

$$D = \text{diag}(d_1, \ldots, d_n), \quad \text{with } d_1 \geq d_2 \geq \ldots \geq d_n \geq 0, \tag{5.1}$$

the matrix RDR^T is nonnegative.

In their papers, Soules [16] and Elsner, Nabben, and Neumann [6] show how to construct a Soules matrix R using $R_1 = x$, where $x \in \mathbb{R}^n$ is a given **positive vector** with $\|x\|_2 = 1$. According to Definition 5.1, if R is a Soules matrix, then $V = RDR^T$ is a symmetric nonnegative matrix, where D is as in (5.1). We shall say that such a matrix V is "*generated via a Soules basis*".

Let R be a Soules matrix and $V = RDR^T \in \mathbb{R}^{n,n}$, where D is as in (5.1). Then for each $1 \leq r \leq n$, the matrix

$$B_r = R \text{ diag}(d_1, \ldots, d_r, \underbrace{0, \ldots, 0}_{n-r}) R^T,$$

is a global minimizer of the optimization problem (1.5). Since each of these B_r's is nonnegative, V has nonnegative global minimizers in the Eckart–Young sense for $r = 1, \ldots, n$.

We now study the NNMF problem (1.2) for matrices that are generated via Soules bases. We have the following theorem, which answers Questions 2 and 3 in Section 1.

THEOREM 5.2 *Let $R \in \mathrm{IR}^{\mathrm{n,n}}$ be a Soules matrix. Let $V = RDR^T$, where D is as in* (5.1)*. Then for any integer $r \in [1, n]$, there is a solution $(\hat{W}, \hat{H})$ to* (1.2) *such that*

$$\hat{W}\hat{H} \;=\; R\mathrm{diag}([d_1, \ldots, d_r, 0, \ldots, 0])R^T, \tag{5.2}$$

and

$$\left\|V - \hat{W}\hat{H}\right\|_F^2 \;=\; \sum_{i=r+1}^{n} d_i^2. \tag{5.3}$$

Moreover, if $d_r > d_{r+1}$, then the product $\hat{W}\hat{H}$ is unique.

Proof: We consider the case that the Soules matrix R is generated in the following way: Let $R_1 = x$. Define

$$x^{(j)} \;=\; x_j e_j \in \mathbb{R}^n,$$

for $j = 1, 2, \cdots, n$, where e_j is the j–th coordinate vector. For $r \in [2, n]$, introduce the vectors

$$y^{(r,r)} \;=\; (x_1, \ldots, x_{n-r+1})^T \in \mathrm{IR}^{\mathrm{n-r+1}}$$

and

$$x^{(r,j)} \;=\; (x_1, \ldots, x_{n-j+1}, 0, \cdots, 0)^T \in \mathrm{IR}^{\mathrm{n}},$$

for $j = 2, \ldots, r$. Then the i–th column of R is defined as

$$R_i \;=\; \frac{1}{\sqrt{\|x^{(r,i)}\|_2^2 + x_{n-i+2}^2}} \left\{ \frac{x_{n-i+2}}{\|x^{(r,i)}\|_2} x^{(r,i)} - \frac{\|x^{(r,i)}\|_2}{x_{n-i+2}} x^{(n-i+2)} \right\}, \tag{5.4}$$

for $i = 2, 3, \ldots, r$. (For a more general approach of constructing R, see Theorem 2.2 of Elsner, Nabben, and Neumann [6].)

According to the Eckart–Young theorem, the matrix $\hat{B} \;=\; R\,\mathrm{diag}([d_1, \ldots, d_r, 0, \ldots, 0])R^T$ is a global minimizer of (1.5) for $k = r$, with the corresponding minimum value

$$\|V - \hat{B}\|_F^2 \;=\; \sum_{i=r+1}^{n} d_i^2.$$

By direct computation, $\hat{B}$ has the following form $\hat{B} = CAC^T$, where

$$C := \begin{bmatrix} y^{(r,r)} & 0 & \cdots & 0 \\ 0 & x_{n-r+2} & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & x_n \end{bmatrix} \quad \text{and} \quad A := \begin{bmatrix} \alpha_{11} & \alpha_{12} & \cdots & \alpha_{1r} \\ \alpha_{21} & \alpha_{22} & \cdots & \alpha_{2r} \\ \vdots & \vdots & \cdots & \vdots \\ \alpha_{r1} & \alpha_{r2} & \cdots & \alpha_{rr} \end{bmatrix},$$

and where A is an $r \times r$ symmetric matrix whose entries are specified as follows:

$$\begin{aligned} \alpha_{1,1} &= d_1 + d_2 \frac{1}{\|x^{(r,2)}\|_2^2 + x_n^2} \frac{x_n^2}{\|x^{(r,2)}\|_2^2} + \ldots + d_{r-1} \frac{1}{\|x^{(r,r-1)}\|_2^2 + x_{n-r+3}^2} \frac{x_{n-r+3}^2}{\|x^{(r,r-1)}\|_2^2} \\ &+ d_r \frac{1}{\|x^{(r,r)}\|_2^2 + x_{n-r+2}^2} \frac{x_{n-r+2}^2}{\|x^{(r,r)}\|_2^2} \\ &\geq 0, \end{aligned}$$

$$\begin{aligned} \alpha_{i,1} &= d_1 \frac{x_n^2}{\|x^{(r,2)}\|_2^2} + \ldots + d_{r+1-i} \frac{1}{\|x^{(r,r-i+1)}\|_2^2 + x_{n+i-r+1}^2} \frac{x_{n+i-r+1}^2}{\|x^{(r,r-i+1)}\|_2^2} \\ &- d_{r+2-i} \frac{1}{\|x^{(r,r-i+2)}\|_2^2 + x_{n+i-r}^2}, \quad i = r, r-1, \ldots, 2, \end{aligned}$$

$$\begin{aligned} \alpha_{j,j} &= d_1 + d_2 \frac{1}{\|x^{(r,2)}\|_2^2 + x_n^2} \frac{x_n^2}{\|x^{(r,2)}\|_2^2} + \ldots + d_{r+1-j} \frac{1}{\|x^{(r,r-j+1)}\|_2^2 + x_{n-r+j+1}^2} \frac{x_{n-r+j+1}^2}{\|x^{(r,r-j+1)}\|_2^2} \\ &+ d_{r+2-j} \frac{1}{\|x^{(r,r-j+2)}\|_2^2 + x_{n-r+j}^2} \frac{\|x^{(r,r-j+2)}\|_2^2}{x_{n-r+j}^2} \\ &\geq 0, \quad j = r, r-1, \ldots, 2, \end{aligned}$$

$$\begin{aligned} \alpha_{i,j} &= d_1 + d_2 \frac{1}{\|x^{(r,2)}\|_2^2 + x_n^2} \frac{x_n^2}{\|x^{(r,2)}\|_2^2} + \ldots + d_{r+1-i} \frac{1}{\|x^{(r,r-i+1)}\|_2^2 + x_{n+i-r+1}^2} \frac{x_{n+i-r+1}^2}{\|x^{(r,r-i+1)}\|_2^2} \\ &- d_{r+2-i} \frac{1}{\|x^{(r,r-i+2)}\|_2^2 + x_{n+i-r}^2}, \quad i = r, r-1, \ldots, j+1. \end{aligned}$$

To prove the theorem it is sufficient to prove that the matrix A is nonnegative, since we can then choose $\hat{W} = CA$ and $\hat{H} = C^T$. Let us verify the nonnegativity of $\alpha_{i,j}$, keeping in mind that d_i's are decreasing.

$$\begin{aligned} \alpha_{i,j} &= d_1 + d_2 \frac{1}{\|x^{(r,2)}\|_2^2 + x_n^2} \frac{x_n^2}{\|x^{(r,2)}\|_2^2} + \ldots + d_{r+1-i} \frac{1}{\|x^{(r,r-i+1)}\|_2^2 + x_{n+i-r+1}^2} \frac{x_{n+i-r+1}^2}{\|x^{(r,r-i+1)}\|_2^2} \\ &- d_{r+2-i} \frac{1}{\|x^{(r,r-i+2)}\|_2^2 + x_{n+i-r}^2} \end{aligned}$$

$$
\begin{aligned}
&\geq\quad d_1 + d_2\ \frac{1}{\|x^{(r,2)}\|_2^2 + x_n^2}\frac{x_n^2}{\|x^{(r,2)}\|_2^2} + \ldots - d_{r+2-i}\ \frac{1}{\|x^{(r,r-i+1)}\|_2^2 + x_{n+i-r+1}^2} \\
&\geq\quad \ldots \\
&\geq\quad d_1 - d_{r+2-i}\ \frac{1}{\|x^{(r,2)}\|_2^2 + x_n^2} \\
&=\quad d_1 - d_{r+2-i} \\
&\geq\quad 0.
\end{aligned}
$$

Similarly, we can show that $\alpha_{i,1}$ is nonnegative. The Eckart–Young theorem also guarantees the uniqueness of the product $\hat{W}\hat{H}$ when $d_r > d_{r+1}$. □

REMARK 5.3 In the above proof of nonnegativity of $\alpha_{i,j}$, we used the inequalities

$$
\begin{aligned}
& d_{r+1-i}\frac{1}{\|x^{(r,r-i+1)}\|_2^2 + x_{n+i-r+1}^2}\frac{x_{n+i-r+1}^2}{\|x^{(r,r-i+1)}\|_2^2} - d_{r+2-i}\frac{1}{\|x^{(r,r-i+2)}\|_2^2 + x_{n+i-r}^2} \\
\geq\quad & d_{r+2-i}\ \frac{1}{\|x^{(r,r-i+1)}\|_2^2 + x_{n+i-r+1}^2}\frac{x_{n+i-r+1}^2}{\|x^{(r,r-i+1)}\|_2^2} - d_{r+2-i}\ \frac{1}{\|x^{(r,r-i+2)}\|_2^2 + x_{n+i-r}^2} \\
=\quad & -d_{r+2-i}\ \frac{1}{\|x^{(r,r-i+1)}\|_2^2 + x_{n+i-r+1}^2}
\end{aligned}
$$

which hold since $d_{r+1-i} \ \geq\ d_{r+2-i}$ and since

$$
\|x^{(r,r-i+1)}\|_2^2 \;=\; \|x^{(r,r-i+2)}\|_2^2 + x_{n+i-r}^2.
$$

As an example, when $r = 2$ and the Soules matrix R is generated via (5.4) we have that

$$
\hat{B} \;=\; \begin{bmatrix} y^{(2,2)} & 0 \\ 0 & x_n \end{bmatrix}\begin{bmatrix} \alpha_{1,1} & \alpha_{1,2} \\ \alpha_{2,1} & \alpha_{2,2} \end{bmatrix}\begin{bmatrix} y^{(2,2)} & 0 \\ 0 & x_n \end{bmatrix}^T,
$$

where

$$
\alpha_{1,1} \;=\; d_1 + d_2 x_n^2/\|y^{(2,2)}\|_2^2;\quad \alpha_{2,2} \;=\; d_1 + d_2\|y^{(2,2)}\|_2^2/x_n^2;\quad \alpha_{2,1} \;=\; d_1 - d_2.
$$

Now one choice of $\hat{W}$ and $\hat{H}$ is:

$$
\hat{W} \;=\; \begin{bmatrix} y^{(2,2)} & 0 \\ 0 & x_n \end{bmatrix}\begin{bmatrix} \alpha_{1,1} & \alpha_{1,2} \\ \alpha_{2,1} & \alpha_{2,2} \end{bmatrix}
$$

and

$$\hat{H} = \begin{bmatrix} y^{(2,2)} & 0 \\ 0 & x_n \end{bmatrix}^T.$$

Of course other choices are possible by decomposing the matrix

$$\begin{bmatrix} \alpha_{1,1} & \alpha_{1,2} \\ \alpha_{2,1} & \alpha_{2,2} \end{bmatrix}$$

into a product of two nonnegative matrices.

We note that the product $\hat{W}\hat{H}$ in (5.2) is symmetric. Therefore the answer to Question 2 in Section 1 is in the affirmative if V is generated via a Soules basis. Moreover, Theorem 5.2 gives the condition under which the product $\hat{W}\hat{H}$ is unique, which answers Question 3. We also observe that if $V \in \mathbb{R}^{n,n}$ is generated by a Soules basis with rank $t = \text{rank}(V) \leq n$, then according to Theorem 5.2, V has an *exact rank t* nonnegative factorization, i.e., there exist nonnegative $W \in \mathbb{R}^{n,t}$ and $H \in \mathbb{R}^{t,n}$ such that $V = WH$.

We comment that a symmetric nonnegative matrix $V \in \mathbb{R}^{n,n}$ can have, for each $r = 1, \ldots, n$, nonnegative global minimizers in the Eckart–Young sense without V being generated via a Soules basis. As an example let $V = Q\Lambda Q^T$, where

$$Q = \begin{bmatrix} 1/2 & 1/\sqrt{2} & 0 & 1/2 \\ 1/2 & -1/\sqrt{2} & 0 & 1/2 \\ 1/2 & 0 & 1/\sqrt{2} & -1/2 \\ 1/2 & 0 & -1/\sqrt{2} & -1/2 \end{bmatrix},$$

and

$$\Lambda = \begin{bmatrix} 6 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix},$$

in which case

$$V = \begin{bmatrix} 3.25 & 0.25 & 1.25 & 1.25 \\ 0.25 & 3.25 & 1.25 & 1.25 \\ 1.25 & 1.25 & 2.75 & 0.75 \\ 1.25 & 1.25 & 0.75 & 2.75 \end{bmatrix}.$$

One can verify that the columns of Q do not form a Soules basis, but that for $r = 1, \ldots, 4$, the Young–Eckart global minimizers $B_1, \ldots, B_4$ for V satisfying (1.4)–(1.5) are all nonnegative . For instance, B_3 is given by:

$$B_3 = Q\mathrm{diag}(6,3,2,0)Q^T = \begin{bmatrix} 3.0 & 0.0 & 1.5 & 1.5 \\ 0.0 & 3.0 & 1.5 & 1.5 \\ 1.5 & 1.5 & 2.5 & 0.5 \\ 1.5 & 1.5 & 0.5 & 2.5 \end{bmatrix}, \tag{5.5}$$

Continuing, letting $r = 3$ and applying the Lee and Seung algorithm to V yields the approximation

$$W_3H_3 = \begin{bmatrix} 2.865 & 0.04646 & 1.606 & 1.590 \\ 0.04646 & 3.074 & 1.450 & 1.432 \\ 1.606 & 1.450 & 2.421 & 0.6504 \\ 1.590 & 1.432 & 0.6504 & 2.449 \end{bmatrix}$$

with $\|V - W_3H_3\|_F = 1.0539$. Moreover, taking $r = 3$ and applying the Lee and Seung algorithm to B_3 yields

$$\tilde{W}_3\tilde{H}_3 = \begin{bmatrix} 2.900 & 0.05204 & 1.594 & 1.588 \\ 0.05204 & 3.039 & 1.463 & 1.452 \\ 1.594 & 1.463 & 2.413 & 0.6715 \\ 1.588 & 1.452 & 0.6715 & 2.424 \end{bmatrix}$$

with $\|B_3 - \tilde{W}_3\tilde{H}_3\|_F = 0.3597$. These numerical results suggest very strongly that for $r = 3$, the global Eckart–Young minimizer for V differs from the global minimizer to V in the sense of (1.2) furnished by the NNMF algorithm. Moreover, it suggests also that B_3 does not possess an exact nonnegative factorization.

We close the paper by raising the question of characterizing symmetric nonnegative matrices $V \in \mathbb{R}^{n,n}$ which, for each $r = 1, \ldots, \mathrm{rank}(V)$, have the property that their global Eckart–Young minimizers B_r have an exact nonnegative factorization W_rH_r, where $W_r \in \mathbb{R}^{n,r}$ and $H_r \in \mathbb{R}^{r,n}$.

References

[1] S. Ambikkumar and S. W. Drury, A reverse Hadamard inequality, *Lin. Alg. Appl.*, 247:83–95, 1996.

[2] A. Berman and R. Plemmons, *Non-Negative Matrices in the Mathematical Sciences*, SIAM Press Classics Series, Philadelphia, 1994.

[3] A. Berman and N. Shaked–Monderer, Completely Positive Matrices, *World Scientific Publishing Co.,* New–Jersey, 2003.

[4] M. Cooper and J. Foote, Summarizing Video via Similarity Matrix Decomposition, *Proc. IEEE Workshop on Multimedia Signal Processing,* 2002.

[5] D. Donoho and V. Stodden, When does Non-Negative Matrix Factorization Give a Correct Decomposition into Parts?, preprint, Department of Statistics, Stanford University, 2003.

[6] L. Elsner, R. Nabben, and M. Neumann, Orthogonal bases which lead to symmetric nonnegative bases, *Lin. Alg. Appl.,* 271:323–343, 1998.

[7] N. J. Higham, Matrix Nearest Problems and Applications, In M. Gover and S. Barnett, editors, *Applications of Matrix Theory,* pp.1–27, Oxford University Press, 1989.

[8] D. D. Lee and H. S. Seung, Unsupervised learning by convex and conic coding, *Advances in Neural Information Processing Systems* 9: 515–521, 1997.

[9] D. D. Lee and H. S. Seung, Learning the parts of the objects by non-negative matrix factorization, *Nature,* 401 (1999), 788–791.

[10] D. D. Lee and H. S. Seung, Algorithms for non-negative matrix factorization, *Advances in Neural Information Processing Systems* 13, 2001.

[11] S. Li, X. Hou and H. Zhang, Learning spatially localized, parts-based representation, In *Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition,* December 2001.

[12] J. Nocedal and S. Wright, *Numerical Optimization,* Springer, 1999.

[13] P. Paatero, Least squares formulation of robust non-negative factor analysis, *Chemometr. Intell. Lab.,* 37: 23–35, 1997.

[14] P. Paatero and U. Tapper, Positive matrix factorization: a non-negative factor model with optimal utilization of error estimates of data values, *Environmetrics,* 5: 111–126, 1994.

[15] L. Saul and D. Lee, Multiplicative updates for classification by mixture models, In *Advances in Neural Information Processing Systems* 14, 2002.

[16] G. W. Soules, Constructing Symmetric Nonnegative Matrices. *Lin. Multilin. Alg.*, 13:241–251, 1983.