Journal of Computational Mathematics
Vol.30, No.1, 2012, 34–46.
http://www.global-sci.org/jcm
doi:10.4208/jcm.1110-m11si10

# AN EFFECTIVE INITIALIZATION FOR ORTHOGONAL NONNEGATIVE MATRIX FACTORIZATION*

Xuansheng Wang
*School of Mathematical Science, Xiamen University, Xiamen 361005, China*
*Email: wxs111111@163.com*
Xiaoyao Xie
*School of Mathematics and Computer Science, Guizhou Normal University, Guiyang 550001, China*
*Email: xyx@gznu.edu.cn*
Linzhang Lu
*School of Mathematics and Computer Science, Guizhou Normal University, Guiyang 550001, China*
*School of Mathematical Science, Xiamen University, Xiamen 361005, China*
*Email: lzlu@xmu.edu.cn*

**Abstract**

The orthogonal nonnegative matrix factorization (ONMF) has many applications in a variety of areas such as data mining, information processing and pattern recognition. In this paper, we propose a novel initialization method for the ONMF based on the Lanczos bidiagonalization and the nonnegative approximation of rank one matrix. Numerical experiments are given to show that our initialization strategy is effective and efficient.

*Mathematics subject classification:* 65F99.
*Key words:* Lanczos bidiagonalization, Orthogonal nonnegative matrix factorization, Low-rank approximation, Nonnegative approximation.

## 1. Introduction

Let $m$ and $n$ be two integers, denote by $\mathbb{R}_+^{m,n}$ the set of all $m \times n$ nonnegative matrices. The nonnegative matrix factorization (NMF) problem means that for given $A \in \mathbb{R}_+^{m,n}$ and $k \ll \min(m,n)$, finding $W \in \mathbb{R}_+^{m,k}$ and $H \in \mathbb{R}_+^{k,n}$ such that

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

That is, finding two nonnegative matrices of low rank $W$ and $H$, such that their product is an approximation of a given nonnegative matrix $A$ in some distance metrics (in this paper, the distance metric will be the Frobenius norm $\|.\|_F$) [14]. The NMF, or approximation of a nonnegative matrix, has become a useful tool in a large applications, such as, images processing, text mining and space situation alertness. Scientific literature and soft tools [4] on the subject and variants thereof are rapidly expending. The orthogonal nonnegative matrix factorization (ONMF), where an orthogonality constraint is imposed on a factor ($W$ or $H$) in the decomposition (1.1), was shown to provide a more clear interpretation on a link between clustering and matrix decomposition [5]. Multiplicative updates for the NMF with preserving orthogonality were recently proposed in [3]. Numerical experiments on face image data for an image representation task show that the ONMF algorithm preserves the orthogonality, while

* Received February 28, 2011 / Revised version received June 7, 2011 / Accepted June 20, 2011 / Published online January 9, 2012 /

the goodness-of-fit (GOF) is minimized. In [3], the GOF is compared with standard NMF. As this is not the point of this paper, we will not describe it in details.

To speed up the convergence of the NMF methods and the minimization of the objective function, most research papers to date for the NMF algorithms have discussed the need to investigate good initialization strategies [1]. However, few of them mentioned the initialization of the ONMF. Therefore, in this paper, we propose a novel initialization algorithm for the ONMF based on the Lanczos algorithm and nonnegative approximation of rank one matrices (see [2]). The proposed algorithm has some good features: it can be combined with all ONMF algorithms and allows a little randomization by free choice of the initial vectors in the Lanczos process. Moreover, our initialization can preserve some original information from given data. From our numerical experiments, it is seen that the initialization algorithm work effectively and efficiently.

The rest of this paper is organized as follows. Section 2 reviews the Lanczos bidiagonalization process to get a low-rank approximation of a nonnegative matrix. Section 3 presents and analyzes our algorithm. In section 4, we give some numerical experiments to demonstrate our algorithms. The last section provides some conclusion.

## 2. Lanczos Bidiagonalization

Since the ONMF is a constrained low-rank approximation problem of a matrix, we need to seek an initialization strategy among alternative low-rank factorizations. For such a problem, the following Eckart-Young theorem [12] is important.

**Theorem 2.1.** *Let $A \in \mathbb{R}^{m,n}$ have the singular values decomposition (SVD)*

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

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

$$A_r = Pdiag(\sigma_1, \cdots, \sigma_r, \underbrace{0, \cdots, 0}_{n-r})Q^T \tag{2.1}$$

*is a global minimize of the optimization problem*

$$\min \left\{ \|A - B\|_F^2 \ \ |B \in \mathbb{R}^{m,n}, rank(B) \leq r \right\} \tag{2.2}$$

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

It follows from Theorem 1 that once the SVD of a matrix $A$ is available, the best rank $r$ approximation $A_r$ of $A$ is easily computed. When $A$ is large, however, computing the SVD of $A$ can be costly. If we are only interested in some $A_r$ with $r \ll \min(m, n)$, the computation of the complete SVD of A is rather wasteful. It is therefore desirable to develop less expensive alternatives for computing a good approximation of $A_r$. In this section, we show that we can obtain a good low-rank nonnegative approximation of a nonnegative matrix $A$ directly from the Lanczos bidiagonalization process without computing the SVD of $A$.

In the following, we describe the Lanczos bidiagonalization process presented in [11, 20].