Signal Processing ∎ (∎∎∎∎) ∎∎∎–∎∎∎

Contents lists available at ScienceDirect

Signal Processing

journal homepage: www.elsevier.com/locate/sigpro

Smooth nonnegative matrix and tensor factorizations for robust multi-way data analysis

Tatsuya Yokota [a,1], Rafal Zdunek [b], Andrzej Cichocki [a,d], Yukihiko Yamashita [c]

[a] RIKEN Brain Science Institute, Japan
[b] Wroclaw University of Technology, Poland
[c] Tokyo Institute of Technology, Japan
[d] Systems Research Institute, Polish Academy of Science, Poland

ARTICLE INFO

Article history:
Received 26 August 2014
Received in revised form
2 February 2015
Accepted 3 February 2015

Keywords:
Nonnegative matrix factorization (NMF)
Nonnegative tensor factorization (NTF)
Smooth component analysis (SmCA)
Blind source separation (BSS)
Multi-way data analysis
Gaussian radial basis function (GRBF)

ABSTRACT

In this paper, we discuss new efficient algorithms for nonnegative matrix factorization (NMF) with smoothness constraints imposed on nonnegative components or factors. Such constraints allow us to alleviate certain ambiguity problems, which facilitates better physical interpretation or meaning. In our approach, various basis functions are exploited to flexibly and efficiently represent the smooth nonnegative components. For noisy input data, the proposed algorithms are more robust than the existing smooth and sparse NMF algorithms. Moreover, we extend the proposed approach to the smooth nonnegative Tucker decomposition and smooth nonnegative canonical polyadic decomposition (also called smooth nonnegative tensor factorization). Finally, we conduct extensive experiments on synthetic and real-world multi-way array data to demonstrate the advantages of the proposed algorithms.

1. Introduction

Nonnegative matrix/tensor factorization (NMF/NTF) plays an important role in feature extraction, classification, blind source separation (BSS), denoising, completion of missing values, and clustering of nonnegative signals [4,8,9,11–13,18,22,25,30,37,40].

The standard NMF model is given by

$$\boldsymbol{Y} \cong \boldsymbol{AX} \in \mathbb{R}_{+}^{I \times J}, \tag{1}$$

where $\boldsymbol{A} = [\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_R] \in \mathbb{R}_{+}^{I \times R}$, $\boldsymbol{X} \in \mathbb{R}_{+}^{R \times J}$, and $\boldsymbol{Y} = [\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_J]$ is an input matrix consisting of J observed signals. The goal of NMF is to compute $\boldsymbol{A}$ and $\boldsymbol{X}$ from the matrix $\boldsymbol{Y}$ for a given parameter R. For example, we consider $R \le J \ll I$ in BSS problems [5]. In this case, we want to find R latent source signals from J mixed observations. NMF gives $\boldsymbol{A}$ as an estimator of the latent source signals. In the case of extracting parts of facial images [22], I and J denote the number of pixels in an image and the number of images, respectively. NMF then represents each facial image as a linear combination of R nonnegative parts. In the case of clustering tasks [37], $\boldsymbol{A}$ is the set of cluster centroids, and $\boldsymbol{X}$ represents the weight parameters of the clusters.

The standard criterion for NMF based on the Euclidean distance is given by

$$\min \|\boldsymbol{Y} - \boldsymbol{AX}\|_F^2, \quad \text{s.t. } \boldsymbol{A} \ge 0, \; \boldsymbol{X} \ge 0, \tag{2}$$

where $\|\cdot\|_F$ stands for the Frobenius norm. Obviously, it minimizes the Euclidean distance between the observed signals $\boldsymbol{Y}$ and the model $\boldsymbol{AX}$, imposing nonnegativity constraints onto latent source and mixing matrices. When there is no constraint, this decomposition model has an unlimited number of solutions, and it cannot provide any meaningful decomposition. However, nonnegativity constraints narrow down the set of the solutions to these which have some meaning for the latent components and the mixing systems.

E-mail address: yokota@brain.riken.jp (T. Yokota).
[1] Tel.: +81 48 462 1111x7173; fax: +81 48 467 9694.

http://dx.doi.org/10.1016/j.sigpro.2015.02.003
0165-1684/© 2015 Published by Elsevier B.V.

Such constraints imposed to both factors $\boldsymbol{A}$ and $\boldsymbol{X}$ can be justified by a non-subtracting mixing system of nonnegative signals, which can reduce redundancy of the solution.

Moreover, non-subtracting mixing systems are used for modeling light, sound, electromagnetic spectra, probability density functions, and image/video brightness. The observed signals, mixing matrices, and latent components are nonnegative, and NMF represents them using the linear superposition system (1). Hence NMF is a useful tool to analyze such kinds of data. The nonnegativity constraint plays an important role in physical interpretation of the decomposition and extraction of non-redundant signals from physically mixed observations. Basically, luminance signals, spectral signals, textual data, and financial data should be nonnegative, and their latent components are often preferred to be nonnegative for meaningful interpretation of the feature vectors.

In general, NMF/NTF is not unique. Thus, for many types of data, we need to impose some additional constraints to relax the problem of non-uniqueness and obtain physically meaningful components. To date, most researchers have imposed sparsity constraints [14,16,18]. In this paper, we investigate another fundamental constraint: smoothness. The smoothness means that the differences between neighboring values are small in some domain. For example, harmonic signals are smooth in the time domain and sparse in the frequency domain. Considering the nonnegative signals, natural image signals are smooth in the spatial domain, probability density functions may be often smooth in some domain, and spectral intensity of optical waves is smooth in the wavelength domain. The latent components of the above signals may be smooth and nonnegative, and the mixing matrices could be assumed as nonnegative based on non-subtracting physical mixing systems. For this reason, NMF/NTF with smoothness constraints is very useful to analyze such kinds of data. When we assume the non-smooth noise is included in observations, this version of NMF, which is referred to as the smooth NMF, may reduce the effects of the noise on the estimators of nonnegative and smooth latent components. In other words, it should be robust to non-smooth noise. In fact, smooth NMF is useful for analyzing temporally or spatially smooth signals (e.g., natural image data, brain waves, and financial data) [4,12,13,38–40].

Many smooth NMF methods can generally be separated into two groups. In the first one, a smoothness constraint term is added to the NMF criterion. For example, Chen et al. [4] proposed the addition of a temporal smoothness constraint and a spatial decorrelation constraint into the Frobenius norm, and into the Kullback–Leibler (KL) divergence-based NMF for electroencephalography (EEG) analysis. Zdunek and Cichocki [39,40] added a Gibbs regularization term for smooth NMF. Drakakis et al. [12] incorporated a sparseness constraint into the mixing matrix, and a smoothness constraint was added to the feature matrix in the Frobenius norm and KL divergence-based NMF for the analysis of financial data. Essid and Fevotte [13] applied the KL divergence-based smooth NMF for audio-visual document structuring, and Dong and Li [11] reported the application of smooth NMF using Laplacian regularization for incomplete matrix factorization.

In the second group, the feature vectors are approximated by a linear combination of several smooth basis vectors. This approach was first proposed by Zdunek [38], where Gaussian radial basis functions (GRBFs) were used with a single standard deviation parameter. This GRBF-NMF method provides effective performance for robust data analysis with respect to noise. However, the original algorithm was relatively slow, because it employed quadratic programming (QP) optimization and the active-set algorithm. The computational cost of QP optimization increases exponentially for large-scale problems. Thus, the original GRBF-NMF algorithm is not practical for large-scale data.

Another problem is that research into smooth nonnegative 'tensor' factorization is not sufficiently well progressed, despite many promising potential applications exist. One reason for this is that most existing algorithms for smooth NMF are quite complex and have a very high computational cost. In this paper, we address the following objectives: to simplify the GRBF-NMF method and develop a new practical algorithm (i.e., reduce the computational cost); to extend the method to the nonnegative Tucker and canonical polyadic (CP) decompositions with additional smoothness constraints. For this purpose, we modify the original problem and propose a new fast algorithm based on the hierarchical alternating least-squares (HALS) method, [6,8], which is a fast and stable algorithm for general NMF/NTF. Furthermore, we propose two extensions for GRBF-NMF. The first uses more flexible basis functions that consist of Gaussian functions with multiple standard deviation parameters. The second one involves two-dimensional Gaussian functions for processing image data. We call this extension GRBF-NMF-2D basis.

For the second objective, we propose two algorithms for smooth nonnegative Tucker decomposition (NTD) and smooth nonnegative CP decomposition (NCPD). These are extensions of our HALS-based GRBF-NMF algorithm. We call these extensions GRBF-NTD and GRBF-NCPD. Furthermore, the NTF methods are extended to the '2D basis' case. Note that we can select the target modes on which to impose the smoothness constraint. For example, for a 3D tensor with the temporally smooth domain (the first mode), the spatially smooth domain (the second mode), and the trial non-smooth domain (the third mode), the smoothness constraint can be applied to only the first and the second modes.

The remainder of this paper is organized as follows. Section 2 introduces the original GRBF-NMF algorithm for a smooth representation. In Section 3, we propose a novel fast algorithm for GRBF-NMF, and discuss its extensions. Section 4 explores the tensor versions of our approach based on the Tucker and CP models. In Section 5, we investigate the performance and applications of our new HALS-based GRBF-NMF/NTF algorithms, and compare them with some state-of-the-art methods. In Section 6, we discuss several aspects of our work, including potential applications and open problems. Finally, we give our conclusions in Section 7.

2. Smooth nonnegative matrix factorization with function approximation

In this section, we review the basic smooth NMF model using the function approximation proposed by Zdunek [38]. According to this method, a feature vector $\boldsymbol{a}_r$ is

Please cite this article as: T. Yokota, et al., Smooth nonnegative matrix and tensor factorizations for robust multi-way data analysis, Signal Processing (2015), http://dx.doi.org/10.1016/j.sigpro.2015.02.003