NON-MATRIX FACTORIZATION FOR BLIND IMAGE SEPARATION

Ichsan Mardani

Jurusan Teknik Informatika, Universitas Pembangunan Nasional Veteran Jakarta
Jl. RS. Fatmawati No 1 Pondok Labu Jakarta Selatan
Email: Mardani300@gmail.com

Abstract

Hyperspectral unmixing is a process to identify the constituent materials and estimate the corresponding fractions from the mixture, nonnegative matrix factions (NMF) is suitable as a candidate for the linear spectral mixture mode, has been applied to the unmixing hyperspectral data. Unfortunately, the local minima is cause by the nonconvexity of the objective function makes the solution nonunique, thus only the nonnegativity constraint is not sufficient enough to lead to a well define problems. Therefore, two inherent characteristic of hyperspectal data, piecewise smoothness (both temporal and spatial) of spectral data and sparseness of abundance fraction of every material, are introduce to the NMF. The adaptive potential function from discontinuity adaptive Markov random field model is used to describe the smoothness constraint while preserving discontinuities is spectral data. At the same time two NMF algorithms, non smooth NMS and NMF with sparseness constraint, are used to quantify the degree of sparseness of material abundances. Experiment using the synthetic and real data demonstrate the proposed algorithms provides an effective unsupervised technique for hyperspectial unmixing.

Keywords Hyperspectral unmixing, nonnegative matrix factorization

INTRODUCTION

Nonnegative Matrix Factorization (NMF) (Lee and Seung, 1999; Pattero and Tapper, 1994) has attracted many attentions for the past decade as a dimension reduction method in machine learning and data mining. NMF are considered as one of the highest dimensional data where each element has a nonnegative value, and provide a lower rank approximation that formed by factors whose elements are also nonnegative.

Due to the nonnegativity, the factors of lower rank approximation given a natural interpretation: for each object is explained by an additive linear it combines of intrinsic ‘parts’ of the data (Lee and Seung, 1999). Numerous successes were reported in application areas including text mining (Pauca et all, 2004), text clustering (xu et all, 2003), computer vision (Li et all, 2001), and cancer class discovery (Brunett et all, 2004; Kim and Park, 2007).

NMF can be traced back to 1970's (Notes from G. Golub) and it studies extensively by Paatero (Pattero and Tapper, 1994). Suggested that NMF factors contain coherent parts of the original images. They confirm that the difference between NMF and vector quantization (which is essentially the K-means clustering). However, later experiments (Hoyer,2004; Pattero and Tapper, 1994) do not support the coherent part interpretation of NMF. Moreover, most applications make use of the clustering aspect of NMF, which is de-emphasized by Lee and Seung (Lee and Seung, 1999). A recent theoretical analysis (Ding et all, 2005) shows the equivalence between NMF and K-means / spectral clustering. In these days, automatic organization of documents becomes crucial since the number of documents to be handled increases rapidly. Document clustering is an important task in organizing documents automatically, which simplifies many subsequent tasks such as retrieval, summarization, and recommendation. Document is represented as an unordered collection of words, which leads to a term-document matrix for a set of documents to be processed.

Term-document of matrix is nothing but co-occurrence table which is a simple case of dyadic data. Dyadic data refers to a domain with two finite sets of objects in which observations are made for dyads, i.e., pairs with one element from either set (Hofmann, Puzicha, & Jordan, 1999). Matrix factorization-based methods have established as a powerful techniques in dyadic data analysis where a fundamental problem, for example, is to perform document clustering or co-clustering words and documents given a term document matrix. Nonnegative matrix factorization (NMF) (Lee & Seung, 1999, 2001) was successfully applied to a task of document clustering (Shahnaz et all, 2006; Xu, Liu, & Gong, 2003), where a term-document matrix is taint into a product of two factor matrices, one of them is corresponds to a cluster canters (prototypes) and the other one which is associated with cluster indicator vectors. Orthogonal NMF, where an orthogonally constraint is imposed on a factor matrix in the decomposition, was shown to provide more clear interpretation on a link between clustering and matrix decomposition (Ding, Li, Peng, & Park, 2006).

New Extended algorithms for Non-negative Matrix Factorization (NMF). The proposed of the algorithms are to characterized by improving the efficiency and convergence rate, it can also be applied for various distributions of data and additive noise. Information theory and information geometry play an important roles in the derivation of new algorithms. Several loss or functions are used in information theory which allow us to obtain generalized forms of multiplicative NMF learning adaptive algorithms. Flexible and relaxed are also forms of the NMF algorithms to raise convergence speed and impose an additional constraint of sparsity.

The scope of these results is vast since discussed generalized divergence functions include a large number of useful loss functions such as the Amari α–divergence, Relative entropy, Bose-Einstein divergence, Jensen-Shannon divergence, J-divergence, Arithmetic-Geometric (AG) Taneja divergence, etc. Applied the developed algorithms successfully to Blind (or semi blind) Source Separation (BSS) where sources may be generally statistically dependent, however are subject to additional constraints such as nonnegativity and sparsity. Moreover, we applied a novel multilayer NMF strategy which improves performance of the most proposed algorithms (Cichocki et all, 2006).

LITERATURE

Non-negative matrix factorization can be revered to a paper of Paatero and Tapper in 1994 (Pattero and Tapper, 1994). The objective were to perform factor analysis that is based on environmental data, the problem involves of finding a small number of root causes that can explain large set of measurements. Every factor is a positive combination of several elementary variables. In a real condition, a factor is present (in some cases it has a certain positive effect) or it is absent (in some cases it has no effect). Therefore, it often makes sense to constrain the factors and their influences to be non-negative.

The problem can be posed formally. Assuming that the columns of A are the measurements, the columns of V are the factors and the rows of H are the influences of each factor. Use W to denote the weight associated to each part, which indicates the level of confidence in that measurement. Paatero and Tapper advocate optimizing the functional

Paatero and Tapper originally proposed using a constrained alternating least squares algorithm (ALS) to solve the problem. This method fixes V and solves the optimization with responds to H. after that it reverses the roles of the variables and repeats the process ad infinitum. The algorithm is initialized with different random matrices in an effort to obtain a global optimum. De Leeuw argues that ALS algorithms converge since they steadily decrease the objective function (Leeuw, 1994), but this convergence is not in the usual mathematical sense.

Futher report of Paatero, introduced by Lee and Seung the concept of NNMF in a 1997 paper on unsupervised learning (Lee and Seung, 1997). They begin by viewing the following encoding problem. Presume that the columns of V are fixed feature vectors and that a is an input vector to be encoded. The goal is to minimize the reconstruction

$$\min_{h} \quad \|\alpha - Vh\|_2^2$$

(2.1)

Different learning techniques can be obtained from several constraints on the vector h. PCA corresponds to an unconstrained minimization, while Vector Quantization (VQ) requires that h equal one of the canonical basis vectors. Lee and Seung suggst that two techniques compromise among PCA and VQ. The first, convex coding requires the whole of h to be nonnegative numbers which count to one. So the encoded vector is the best approximation to the input from the convex hull of the signalize vectors. The second, conic coding requires only that the entries of h be nonnegative. Then the encoded vector is the best approximation to the input from the cone generated by the feature vectors.

Future work of Lee and Seung consider how to find the best set of feature vectors for their new coding strategies. This leads them to the matrix approximation problem

$$\min_{V \geq 0, H \geq 0} \|A - VH\|_F^2$$

(2.2)

In this case columns of A contain training examples and V has far less columns than A. The two convex and conic coding, require V and H to be nonnegative. In addition, for convex coding, it forces the column sums of V and the row sums of H to equal one.

To solve their minimization problems, they suggest an alternating projected gradient method. Namely is to fix V; that perform a step of gradient descent with regard to H; next step is to zero all the negative components of H. Invert the roles of the variables and repeat.

To solve the convex coding problem, they also identify a penalty function into the minimization to preserve the row and column sums. The algorithms are executed a couple of times with a random starting points in an effort to come accross a global optimum.

Using these algorithms, they found that convex coding discovers locally linear models of the data, at the same time conic coding discovers features in the data. Their paper did not provide any proof of convergence, nor did it consider other types of algorithms which might apply.

The 1999 article in Nature by Daniel Lee and Sebastian Seung (Lee and Seung, 1999) started a flurry of research into the new Nonnegative Matrix Factorization. Hundreds of papers have cited Lee and Seung, but prior to its publication several lesser known papers by Pentti Paatero (Paatero and Tapper, 1994; Paatero, 1997, 1999) actually deserve more credit for the factorization's his- torical development. Though Lee and Seung cite Paatero's 1997 paper on his so-called positive matrix factorization in their Nature article, Paatero's work is rarely cited by subsequent authors. This is partially due to Paatero's unfortunate phrasing of positive matrix factorization, which is misleading as Paatero's algorithms create a nonnegative matrix factorization. Moreover, Paatero actually published his initial factorization algorithms years earlier in (Paatero and Tapper, 1994).

Since the introduction of the NMF problem by Lee and Seung, a great deal of published and unpublished work has been devoted to the analysis, extension, and application of NMF algorithms in science, engineering and medicine. The NMF problem has been cast into alternate formulations by various authors. (Lee and Seung, 2001) provided an information theoretic formulation based on the Kullback-Leibler divergence of A from WH that, in turn, lead to various related approaches. For example, (Cichocki et al., 2006) have proposed cost functions based on Csisz´ar's ϕ-divergence. (Wang et al., 2004) propose a formulation that enforces constraints based on Fisher linear discriminant analysis for improved determination of spatially localized features. (Guillamet et al., 2001) have suggested the use of a diagonal weight matrix Q in a new factorization model, $AQ \approx WHQ$, in an attempt to compensate for feature redundancy in the columns of W. This problem can also be alleviated using column stochastic constraints on H (Pauca et al., 2006).

Other approaches that propose alternative cost function formulations include but are not limitedto (Hamza and Brady, 2006; Dhillon and Sra, 2005). A theoretical analysis of nonnegative matrix factorization of symmetric matrices can be found in (Catral et al., 2004). Various alternative minimization strategies for the solution of (8) have also been proposed in an effort to speed up convergence of the standard NMF iterative algorithm of Lee and Seung. (Lin, 2005b) has recently proposed the use of a projected gradient bound-constrained optimization method that is computationally competitive and appears to have better convergence properties than the standard (multiplicative update rule) approach.

Use of certain auxiliary constraints in (8) may however break down the bound-constrained optimization assumption, limiting the applicability of projected gradient methods. (Gonzalez and Zhang, 2005) proposed accelerating the standard approach based on an interior-point gradient method. (Zdunek and Cichocki, 2006) proposed a quasi-Newton optimization approach for updating W and H where negative values are replaced with small $\varrho > 0$ to enforce nonnegativity, at the expense of a significant increase in computation time per iteration.

Further studies related to convergence of the standard NMF algorithm can be found in (Chu et al., 2004; Lin, 2005a; Salakhutdinov et al., 2003) among others. In the standard NMF algorithm W and H are initialized with random non- negative values, before the iteration starts. Various efforts have focused on alternate approaches for initializing or seeding the algorithm in order to speed up or otherwise influence convergence to a desired solution. (Wild et al., 2003) and (Wild, 2002), for example, employed a spherical k-means clustering approach to initialize W. (Boutsidis and Gallopoulos, 2005) use an SVD-based initialization and show anecdotical examples of speed up in the reduction of the cost function. Effective initialization remains, however, an open problem that deserves further attention.

Recently, various authors have proposed extending the NMF problem formulation to include additional auxiliary constraints on W and/or H. For example smoothness constraints have been used to regularize the computation of spectral features in remote sensing data (Piper et al., 2004; Pauca et al., 2005).(Chen and Cichocki, 2005) employed temporal smoothness and spatial correlation constraints to improve the analysis of EEG data for early detection of Alzheimer's disease. (Hoyer, 2002, 2004) employed sparsity constraints on either W or H to improve local rather than global representation of data.

The extension of NMF to include such auxiliary constraints is problem dependent and often reflects the need to compensate for the presence of noise or other data degradations in A.

2.2. Related work

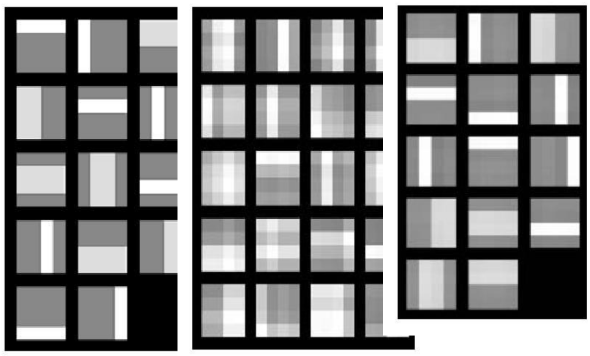
NMF problem is given a nonnegative n x m matrix V , find nonnegative n x r and r x m matrix factors W and H such that the difference measure between V and WH is the minimum according to some cost function, that is

$$V \approx WH \qquad (2.3)$$

NMF is a method to obtain a representation of data using nonnegative constraints. These constraints lead to a part-based representation because they allow only additive, not subtractive, combinations of the original data. For the th column of (9), that is vi = WHi, , where vi and hi are the ith column of V and H the ith observation is a nonnegative linear combination of the columns of W=(Wi,W2,…… ,Wr). while the combinatorial coefficients are the elements of hi Therefore, the W columns of , that is, W=(Wi,W2,…… ,Wr) , can be viewed as the basis of the data V when V is optimally estimated by its factors.

The linear bar problem (Földiák, 1990) is a blind separation of bars from their combinations. 8 nonnegative feature images (sources) sized including 4 vertical and 4 horizontal thin bar images, shown in Figure 1(a) are randomly mixtured to form 1000 observation images, the first 20 shown in Figure 1(b). The solution obtained from ICA and NMF with are shown in Figures 1(c) and 1(d), respectively, indicating that NMF can fulfill the task very well compared with ICA . However, when we extended this bar problem into the one which is composed of two types of bars, thin one and thick one, NMF failed to estimate the original sources. For example, fourteen source images sized with four thin vertical bars, four thin horizontal bars, three wide vertical bars, and three wide horizontal bars, shown in Figure 2(a), are nonnegative and evidently statistically dependent.

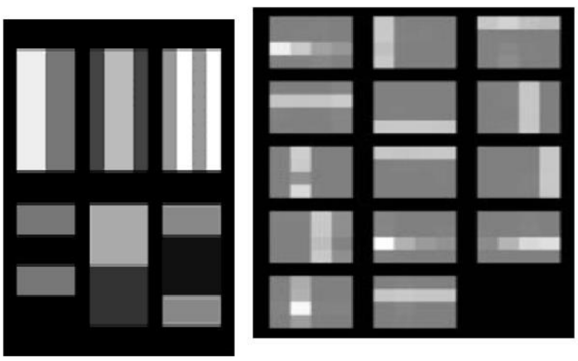
These source images were randomly mixed with mixing matrix of elements arbitrarily chosen in (0, 1) to form 1000 mixed images, the first 20 shown in Figure 2(b). The PE-NMF with parameter and was performed on these mixed images for . The resultant images, which are shown in Figure 2(c), indicate that the sources were recovered successfully with the proposed PE-NMF.



(a) (b) (c)

Figure II.1 Extended bar problem solution obtained from PE-NMF: (a) source images, (b) mixed images, (c) recovered images from PE-NMF. (Zhang et all, 2008)

For comparison, many times we tried using ICA and NMF on this problem for avoiding obtaining local minimum solutions, but always failed to recover the original sources. Shown in Figures 2(a) and 2(b) are the examples of the recovered images with these two approaches. Notice that both the ones recovered from ICA and NMF are very far from the original sources, and even the number of sources estimated from the ICA is only 6, rather than 14. It is noticeable that the recovered images from the PE-NMF with some other parameter such as and are comparable to the ones shown in Figure 2(c), indicating that the proposed method is not very sensitive to the parameter selection for this example.



(a) (b)

Figure II.2 Recovered images from (a) ICA, and (b) NMF for the extended bar problem (Zhang et all, 2008).

Blind Separation (BIS)

Blind source separation (BSS) is a very active topic recently in signal processing and neural network fields (Hyvärinen et all, 2001; Hoyer and Hyvärinen, 2000). It is an approach to recover the sources from their combinations (observations) without any understanding of how the sources are mixed. For a linear model, the observations are linear combinations of sources, that is, , where is an matrix indicating source signals each in -dimensional space, is an matrix showing observations in dimensional space, and is an mixing matrix. Therefore, BSS problem is a matrix factorization, that is, to factorize observation matrix into mixing matrix and source matrix .

Independent component analysis (ICA) has been found very effective in BSS for the cases where the sources are statistically independent. In fact, it factorizes the observation matrix into mixing matrix and source matrix by searching the most nongaussianity directions in the scatter plot of observations, and has a very good estimation performance of the recovered sources when the sources are statistically independent. This is based on the Central Limit Theorem, that is, the distribution of a sum (observations) of independent random variables (sources) tends toward a Gaussian distribution under certain conditions.

This induces the two serious constraints of ICA to the application of BSS:

a) The sources should be statistically independent to each other.
b) The sources should not follow Gaussian distribution.

The performance of the recovered sources with ICA approach depends on the satisfactory of these two constraints, and decreases very rapidly when either of them is not satisfied. However in real world, there are many applications of blind source separation where the observations are nonnegative linear combinations of nonnegative sources, and the sources are statistically dependent to some extent.

This is the model referred to as nonnegative linear model (NNLM), that is, with elements in both and nonnegative, and the rows in (the sources) may be statistically dependent to some extent. One of the applications of this model is gene expression profiles, where each of the profiles, which is only in nonnegative values, represents a composite of more than one distinct but partially dependent sources (Zhang et all, 2003), the profiles from normal tissue and from cancer tissue. What needs to be developed is an algorithm to recover dependent sources from the composite observations.

To recognize that BSS for NNLM is a nonnegative matrix factorization, that is, to factorize into nonnegative and nonnegative , where nonnegative matrix factorization (NMF) technique is applicable. Several approaches have been developed on applying NMF-based technique for BSS of NNLM. For example, we proposed a method for decomposition of molecular signatures based on BSS of nonnegative dependent sources with direct usage of standard NMF (Zhang et all, 2003).

Amari proposed a new algorithm for nonnegative matrix factorization in applications to blind source separation (Amari et all, 2006) by adding two suitable regularizations or penalty terms in the original objective function of the NMF to increase sparseness and/or smoothness of the estimated components. In addition, multilayer NMF was proposed by Cichocki and Zdunek for blind source separation (Cichocki and Zdunek, 2006), and nonsmooth nonnegative matrix factorization was proposed aiming at finding localized, part-based representations of nonnegative multivariate data items (Carazo et all, 2006).

Some other researches include the work of Zdunek and Cichocki, who proposed to take advantage of the second-order terms of a cost function to overcome the disadvantages of gradient (multiplicative) algorithms for NMF for tackling the slow convergence problem of the standard NMF learning algorithms (Cichocki and Zdunek, 2007)

The work by Ivica Kopriva and his colleagues, who proposed a single-frame blind image deconvolution approach with nonnegative sparse matrix factorization for blind image deconvolution (Borjanović, 2006); and the work by Liu and Zheng who proposed nonnegative matrix factorization-based methods for object recognition (Liu and Zheng, 2004).

The NMF to pattern expression NMF (PE-NMF) from the view point that the basis vector is desired to be the one which can express the data most efficiently. Its successful application to blind source separation of extended bar problem, nonnegative signal recovery problem, and heterogeneity correction problem for real gene microarray data indicates that it is of great potential in blind separation of dependent sources for NNLM model.

The loss function for the PE-NMF proposed here is a special case of that proposed in (Amari et all, 2006), and here not only the learning algorithm for the proposed PE-NMF approach is provided, but also the convergence of the learning algorithm is proved by introducing some auxiliary function. For speeding up the learning procedure, a technique based on independent component analysis (ICA) is proposed, and has been verified to be effective for the learning algorithm to converge to desired solutions.

Initialization of the Algorithm

It seems that there are two main reasons for NMF to converge to undesired solutions. One is that the basis of a space may not be unique theoretically, and therefore separate runs of NMF may lead to different results. Another reason may come from the algorithm itself, that the loss function sometimes gets stock into local minimum during its iteration. By revisiting the loss function of the proposed PE-NMF, it is seen that similar to NMF, the above PE-NMF still sometimes gets stock into local minimum during its iteration, and/or the number of iterations required for obtaining desired solutions is very large.

For the sake of these, an ICA-based technique was proposed for initializing source matrix instead of setting it to be a nonnegative matrix at random, ICA perform on the observation signals, and set the absolute of the independent components obtained from ICA to be the initialization of the source matrix. In fact, there are reasons that the resultant independent components obtained from ICA are generally not the original sources.

One reason is the nonnegativity of the original sources but centering pre-processes of the ICA makes each independent component both positive and negative in its elements: the means of each independent component is zero. Another reason is possibly dependent or partially independent original sources which does not follow the independence requirement of sources in the ICA study. Hence, the resultant independent components from ICA could not be considered as the recovery of the original sources. Even so, they still provide clues of the original sources: they can be considered as very rough estimations of the original sources. From this perspective, and by noticing that the initialization of the source matrix should be nonnegative, we set the absolute of the independent components obtained from ICA as the initialization of the source matrix for the proposed PE-NMF algorithm (Zhang, 2008).

The proposed PE-NMF algorithms have been extensively tested for many difficult benchmarks for signals and images with various statistical distributions. Three examples will be given in the following context for demonstrating the effectiveness of the proposed method compared with standard NMF method and/or ICA method. In ICA approach here, we decenteralize the recovered signals/images/microarrays for its nonnegativity property for compensating the centering preprocessing of the ICA approach. The NMF algorithm is simply the one proposed in (Lee and Seung, 1999) and the ICA algorithm is simply the FastICA algorithm generally used in many applications. The examples include blind source separation of extended bar problem, mixed signals, and real microarray gene expression data in which heterogeneity effect occurs.

1.3.2. Independent Component Analysis (ICA)

The proposed PE-NMF algorithms have been extensively tested for many difficult benchmarks for signals and images with various statistical distributions. Three examples will be given in the following context for demonstrating the effectiveness of the proposed method compared with standard NMF method and/or ICA method. In ICA approach, decenteralize the recovered signals/images/microarrays for its nonnegativity property for compensating the centering preprocessing of the ICA approach.

1.1 Extended Lee-Seung Algorithms and Fixed Point Algorithms.

Although the standard NMF (without any auxiliary constraints) provides sparseness of its component, it may achieve some control of this sparsity as well as smoothness of components by imposing additional constraints in addition to non-negativity constraints. In fact, it may incorporate smoothness or sparsity constraints in several ways (Hoyer, 2004). One of the simple approach is to implement in each iteration step a nonlinear projection which can increase the sparseness and/or smoothness of estimated components. An alternative approach is to add to the loss function suitable regularization or penalty terms. Consider the following constrained optimization problem:

$$D_F^{(\alpha)}(A,X) = \frac{1}{2}\|Y-AX\|_F^2 + \alpha_A J_A(A) + \alpha_X J_X(X) \qquad (2.4)$$

where αA and $\alpha X \geq 0$ are nonnegative regularization parameters and terms $JX(\boldsymbol{X})$ and $JA(\boldsymbol{A})$ are used to enforce a certain application-dependent characteristics of the solution. As a special practical case have $JX(\boldsymbol{X})$ = _jk $fX(xjk)$, where $f(\cdot)$ are suitably chosen functions which are the measures of smoothness or sparsity. In order to achieve sparse representation usually choose $f(xjk) = |xjk|$ or simply $f(xjk) = xjk$, or alternatively $f(xjk) = xjk \ln(xjk)$ with constraints $xjk \geq 0$. Similar regularization terms can be also implemented for the matrix $\boldsymbol{A}$. Note that treat both matrices $\boldsymbol{A}$ and $\boldsymbol{X}$ in a symmetric way. Applying the standard gradient descent approach,

$$a_{ij} \leftarrow a_{ij} - \eta_{ij} \frac{\partial D_F^{(\alpha)}(A,X)}{\partial a_{ij}},$$

$$x_{jk} \leftarrow x_{jk} - \eta_{jk} \frac{\partial D_F^{(\alpha)}(A,X)}{\partial x_{jk}}, \qquad (2.5)$$

where ηij and ηjk are positive learning rates. The gradient components can be expressed in a compact matrix form as: (Lee and Seung, 1999)

$$\frac{\partial D_F^{(\alpha)}(A,X)}{\partial a_{ij}} = [\,-\mathbf{YX}^{\mathbf{T}} + \mathbf{AXX}^{\mathbf{T}}\,]_{ij} + \alpha_a \frac{\partial J_A(A)}{\partial a_{ij}}, \qquad (2.6)$$

$$\frac{\partial D_F^{(\alpha)}(A,X)}{\partial a_{ij}} = [\,-\mathbf{A}^{\mathbf{T}}\mathbf{Y} + \mathbf{A}^{\mathbf{T}}\mathbf{AX}]_{ij} + \alpha_a \frac{\partial J_A(A)}{\partial a_{ij}}. \qquad (2.7)$$

Here is the Lee and Seung approach to choose specific learning rates

$$\eta_{ij} = \frac{a_{ij}}{[AXX^T]_{ij}}, \qquad \eta_{jk} = \frac{x_{jk}}{[A^T AX]_{jk}}, \qquad (2.8)$$

that leads to a generalized robust multiplicative update rules:

$$a_{ij} \leftarrow a_{ij} \frac{[[YX^T]_{ij} - \alpha_A \varphi_A(a_{ij})]_\varepsilon}{[AXX^T]_{ij} + \varepsilon}, \qquad (2.9)$$

$$x_{jk} \leftarrow x_{jk} \frac{[[A^T Y]_{jk} - \alpha_X \varphi_X(x_{jk})]_\varepsilon}{[A^T AX]_{jk} + \varepsilon}, \qquad (2.10)$$

where the nonlinear operator is defined as $[x]\varepsilon = \max\{\varepsilon, x\}$ with a small positive ε and the functions $\phi A(aij)$ and $\phi X(xjk)$ are defined as

$$\varphi_A(a_{ij}) = \frac{\partial J_A(A)}{\partial a_{ij}}, \qquad \varphi_X(X_{jk}) = \frac{\partial J_X(X)}{\partial x_{jk}}. \qquad (2.11)$$

Typically, $\varepsilon = 10-16$ is introduced in order to ensure non-negativity constraints and avoid possible division by zero. The above Lee-Seung algorithm can be considred as an extension of the well-known ISRA (Image Space Reconstruction Algorithm) algorithm. The above algorithm reduces to the standard Lee-Seung algorithm for $\alpha A = \alpha X = 0$. In the special case, by using the $l1$-norm regularization terms $f(\boldsymbol{x}) = \_\boldsymbol{x}\_1$ for both matrices $\boldsymbol{X}$ and $\boldsymbol{A}$ the above multiplicative learning rules can be simplified as follows:

$$a_{ij} \leftarrow a_{ij}\frac{[[YX^T]_{ij} - \alpha_A]_\varepsilon}{[AXX^T]_{ij} + \varepsilon}, \quad x_{jk} \leftarrow x_{jk}\frac{[[A^TY]_{jk} - \propto_X \varphi_X(x_{jk})]_\varepsilon}{[A^TAX]_{jk} + \varepsilon}, \quad (2.12)$$

with normalization in each iteration as follows *aij* ← *aij*/_*m i*=1 *aij* . Such normalizationis necessary to provide desired sparseness. Algorithm (11) providesa sparse representation of the estimated matrices and the sparseness measureincreases with increasing values of regularization coefficients, typically αX =0.01 ~0.5.It is worth to note that we can derive as alternative to the Lee-Seung algorithm(11) a Fixed Point NMF algorithm by equalizing the gradient components of (5)-(6) (for $l1$-norm regularization terms) to zero (Lee and Seung, 1997):

$$\nabla_X D_F^{(\alpha)}(Y||AX) = A^TAX - A^TY + \propto_X = 0, \quad (2.13)$$

$$\nabla_X D_F^{(\alpha)}(Y||AX) = AXX^T - YX^T + \propto_A = 0. \quad (2.14)$$

These equations suggest the following fixed point updates rules:

$$X \leftarrow \max\{\mathscr{E}, [(A^TA) + (A^TY - \propto_X)]\} = [(A^TA) + (A^TY - \propto_X)]_\varepsilon, \quad (2.15)$$

$$A \leftarrow \max\{\mathscr{E}, [(YX^T - \propto_A)(XX^T)^+]\} = [(YX^T - \propto_A)(XX^T)^+]_\varepsilon, \quad (2.16)$$

where $[\boldsymbol{A}]+$ means Moore-Penrose pseudo-inverse and *max* function is component wise. The above algorithm can be considered as nonlinear projected Alternating Least Squares (ALS) or nonlinear extension of EM-PCA algorithm.

RELATED WORK

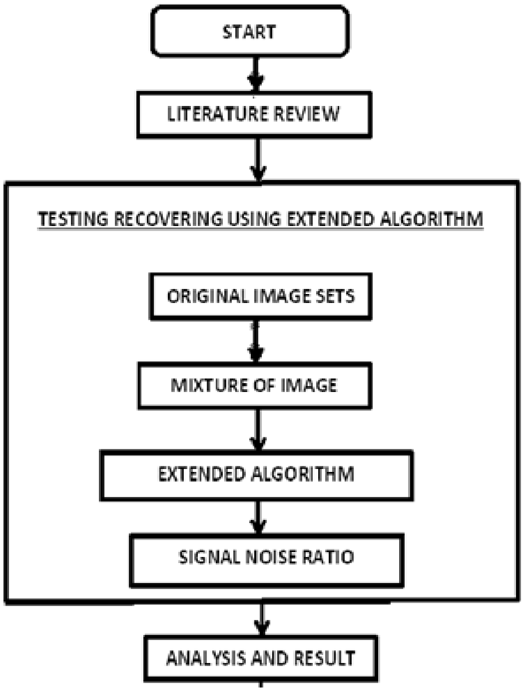
The proposed PE-NMF algorithms have been extensively tested for many difficult benchmarks for signals and images with various statistical distributions. Three examples will be given in the following context for demonstrating the effectiveness of the proposed method compared with standard NMF method and/or ICA method. In ICA approach, deceneralize the recovered signals/images/microarrays for its nonnegativity property for compensating the centering preprocessing of the ICA approach.

The NMF algorithm is simply the one proposed in (Lee and Seung, 1999) and the ICA algorithm is simply the FastICA algorithm generally used in many applications. The examples include blind source separation of extended bar problem, mixed signals, and real microarray gene expression data in which heterogeneity effect occurs.

Lee and Seung work also shows a magnificent result on comparing the NMF algorithms using the BSS technique, it works have shown that, the observations are linear combinations of sources, that is, , where is an matrix indicating source signals each in -dimensional space, is an matrix showing observations in dimensional space, and is an mixing matrix. Therefore, BSS problem is a matrix factorization, that is, to factorize observation matrix into mixing matrix and source matrix .

RESERCH FRAMEWORK

The methodology consists of six main steps of literature review until thesis writing. The proposed of nonnegative matrix factorization for blind image separation. Using the Nonnegative matrix factorization for Blind Image Seperatinon the original image will transform into a set of mixture of image, in this case it can reduce the noise from the original image (genererated image) into a new set of image by using the extended algorithm. The general workflow of the methodology is illustrated.



References

Brunet, J., Tamayo, P., Golub, T, and Mesirov, J. (2004). *Metagenes and molecular pattern discovery using matrix factorization*. Proceedings of the National Academy of Sciences, 101(12):4164–4169, 2004.

De Leeuw, J. (1994). *Information Systems and Data Analysis*. Springer, 1994, ch. Block-relaxation algorithms in statistics.

Dhillon, I. S., Mallela, S., and Modha, D. S. (2003). *Information-theoretic co-clustering*. In Proceedings of the ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD) (pp. 89–98).

Ding, C., He, X., and Simon, H. D. (2005). *On the equivalence of nonnegative matrix factorization and spectral clustering*. Proc. SIAM Data Mining Conf, 2005.

Field, D. J. (1994). *What is the goal of sensory coding? Neural Computation*, 6:559–601, 1994.

Foldi´ak, P., and Young, M. P. (1995). *Sparse coding in the primate cortex. In Michael A. Arbib, editor, The Handbook of Brain Theory and Neural Networks, pages 895–898*. The MIT Press, Cambridge, Massachusetts, 1995.

Hofmann, T., Puzicha, J., and Jordan, M. I. (1999). *Learning from dyadic data. Advances in neural information* processing systems NIPS (Vol. 11). MIT Press.

Hoyer, P. O. (2004). *Non-negative matrix factorization with sparseness constraints*. J. Machine Learning Research, 5:1457–1469, 2004.

Hoyer, P. O. (2002). *Non-negative sparse coding. In Neural Networks for Signal Processing XII (Proc. IEEE Workshop on Neural Networks for Signal Processing),* pages 557–565, Martigny, Switzerland, 2002.

Keshava, N., and Mustard, J. F. (2002). *"Spectral unmixing," IEEE Signal Process*. Mag., vol. 19, no. 1, pp. 44–57, Jan. 2002.

Kim, H. and Park, H. (2007). *Sparse non-negative matrix factorizations via alternating non-negativity-constrained least squares for microarray data analysis*. Bioinformatics, 23(12):1495–1502, 2007.

Landgrebe, D. (1997). *"Hyperspectral image data analysis," IEEE Signal Process*. Mag., vol. 19, no. 1, pp. 17–9.The MIT Press, 1997, pp. 515–521. [Online]. Available: citeseer.nj.nec.com/lee97unsupervised.html.

Lee, D. D. and Seung, H. S. (1997). *"Unsupervised learning by convex and conic coding," in Advances in Neural Information Processing Systems*, M. C. Mozer, M. I. Jordan, and T. Petsche, Eds., vol. 9.The MIT Press, 1997, pp. 515–521. [Online]. Available: citeseer.nj.nec.com/lee97unsupervised.html.

Lee, D. D. and Seung, H. S. (1999). *Learning the parts of objects by non-negative matrix factorization.* Nature, 40 (6755):788– 791, 1999.

Lee, D. D. and Seung, H. S. (2001). *Algorithms for non-negative matrix factorization*. In Advances in Neural Information Processing 13 (Proc. NIPS*2000). MIT Press, 2001.

Li, S. Z., Hou, X., Zhang, H., and Cheng, Q. (2001). *Learning spatially localized, parts-based representation*. In CVPR '01: Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, volume 1, 2001.

Liu, W., Zheng N., and Lu, X. (2003). *Non-negative matrix factorization for visual coding*. In Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP'2003), 2003.

Manolakis, D. and Shaw, G. (200) 2. *"Detection algorithms for hyperspectral imaging applications,"* IEEE Signal Process. Mag., vol. 19, no. 1, pp. 29–43, Jan. 2002. 28, Jan. 2002.

Manolakis, D. and Shaw, G. (2002). *"Detection algorithms for hyperspectral imaging applications*," IEEE Signal Process. Mag., vol. 19, no. 1, pp. 29–43, Jan. 2002.

Nascimento, J. M. P. and Dias, J. M. B. (2005). *"Does independent component analysis play a role in unmixing hyperspectral data?"* IEEE Trans. Geosci. Remote Sens., vol. 43, no. 1, pp. 175–187, Jan. 2005.

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

Parra, L., Spence, C., Sajda, P., Ziehe, A., and Müller, K. R. (1999). *"Unmixing hyperspectral data,"* in Proc. Adv. Neural Inf. Process. Syst., 1999, pp. 942–948.

Pauca, V. P., Shahnaz, F., Berry, W. M., and Plemmons, R. J. (2004). *Text mining using non-negative matrix factorizations.* In Proceedings of the 2004 SIAM International Conference on Data Mining, 2004.

Shahnaz, F., Berry, M., Pauca, P., & Plemmons, R. (2006). *Document clustering using nonnegative matrix factorization*. Information Processing and Management, 42, 373–386.

Shaw, G. A. and Burke, H. H. K. (2003). *"Spectral imaging for remote sensing,"* Linc. Lab. J., vol. 14, no. 1, pp. 3–28, 2003