Sequential EM Learning for Subspace Analysis

Seungjin Choi *

Department of Computer Science
POSTECH
San 31 Hyoja-dong, Nam-gu, Pohang 790-784, Korea

Abstract

Subspace analysis is one of popular multivariate data analysis methods, which has been widely used in pattern recognition. Typically data space belongs to very high dimension, but only a few principal components need to be extracted. In this paper, we present a fast sequential algorithm which behaves like expectation maximization (EM), for subspace analysis or tracking. In addition we also present a slight modification of the subspace algorithm by employing a rectifier, that is quite useful in handling nonnegative data (for example, images), which leads to *rectified subspace analysis*. The useful behavior of our proposed algorithms are confirmed through numerical experimental results with toy data and dynamic PET images.

Key words: Expectation maximization, PET images, Principal component analysis, Sequential learning, Subspace analysis

1 Introduction

The task of pattern recognition is to classify the data in an unlabelled set (test set), given a set of data labelled with its class (training set). Statistical pattern recognition relies on the information representation which is closely related to the statistical feature extraction. For example, linear data representation decomposes a given set of data into a linear sum of basis vectors. Feature vectors are obtained by projecting the data onto the subspace spanned by the basis vectors. Popular methods are factor analysis and principal component analysis (PCA). The eigenface method (Sirovich and Kirby, 1987; Turk

* Corresponding author. Tel.: +82-54-279-2259; Fax: +82-54-279-2299
Email: seungjin@postech.ac.kr (S. Choi)

and Pentland, 1991) might be one exemplary subspace analysis technique in pattern recognition.

In computer vision and pattern recognition, one often encounters into a set of huge dimensional data and wants to extract a small number of features which is able to represent the data as well as possible. The singular value decomposition (SVD) is a numerically robust method which calculates the eigenvectors of the data covariance matrix, however, it is computationally expensive, especially for the case of data with high dimension. For adaptive computation of eigenvectors, a variety of PCA neural networks have been developed (Diamantaras and Kung, 1996), most of which are gradient-based learning algorithms, so their convergence is very slow.

Recently probabilistic model-based methods for subspace analysis have been proposed. These include probabilistic PCA (PPCA) (Tipping and Bishop, 1999b), EM-PCA (Roweis, 1998), mixtures of factor analyzers (Ghahramani and Hinton, 1997), and mixtures of probabilistic principal component analyzers (Tipping and Bishop, 1999a). All these algorithms employ the EM learning which is an iterative maximum likelihood estimation method in the presence of hidden variables. PPCA and EM-PCA are batch algorithms, thus, when a new data arrives, whole calculation should be carried out again. In order to overcome this drawback, We present a sequential EM learning algorithm for subspace analysis.

In some applications such as image processing and text data processing, all data points are nonnegative. Imposing nonnegativity constraints in both basis vectors and encoding variables (factors) leads to a nonsubtractive combination of nonnegative basis vectors. Parts-based representation was shown to emerge from this nonsubtractive combination of basis vectors (Lee and Seung, 1999). Along this line, the nonnegative matrix factorization (NMF) was introduced (Lee and Seung, 1999) and drew extensive attraction in a variety of applications. It was demonstrated that the NMF gave parts-based representation (Lee and Seung, 1999) and was useful in dynamic positron emission tomography (PET) image analysis (Lee et al., 2001). On the other hand, Oja's subspace algorithm with nonnegativity constraints was shown be useful for the identification of face orientation (Lai and Fyfe, 2000). Motivated from these results (Lee and Seung, 1999; Lai and Fyfe, 2000), we present a slight modification of the sequential EM subspace algorithm, which leads to *rectified subspace analysis*. Experimental results show that our proposed algorithm is very useful for dynamic PET image analysis.

The rest of this paper is organized as follows. Probabilistic PCA is briefly reviewed in Sec. 2. The sequential EM algorithm for subspace analysis is proposed in Sec. 3 and its modification, the rectified subspace analysis, is presented in Sec. 4. Numerical experimental results are shown in Sec. 5. Con-

clusions are drawn in Sec. 6.

2 Probabilistic PCA

This section reviews PPCA (Tipping and Bishop, 1999b) for the case of isotropic Gaussian noise model and zero noise limit. In the limit of zero noise, PPCA is known as EM-PCA (Roweis, 1998).

2.1 Linear Generative Model

The linear generative model assumes that the set of m-dimensional observed vectors $\{\boldsymbol{x}_t\}$ is generated from a corresponding set of latent variables $\{\boldsymbol{s}_t\}$ by

$$\boldsymbol{x}_t = \boldsymbol{A}\boldsymbol{s}_t + \boldsymbol{v}_t, \tag{1}$$

where $\boldsymbol{s}_t \in \mathbb{R}^n$ $(n \leq m)$ and $\boldsymbol{v}_t \in \mathbb{R}^m$ is Gaussian noise vector that is assumed to be statistically independent of $\boldsymbol{s}_t$.

In the standard factor analysis, latent variables $\boldsymbol{s}$ are assumed to have a unit isotropic Gaussian distribution, i.e., $\boldsymbol{s} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$. The noise model is Gaussian, i.e., $\boldsymbol{v} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma})$ with $\boldsymbol{\Sigma}$ being a diagonal matrix. Given this formulation, the model for $\boldsymbol{x}$ is also Gaussian, $\boldsymbol{x} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{C})$ where $\boldsymbol{C} = \boldsymbol{\Sigma} + \boldsymbol{A}\boldsymbol{A}^T$. Due to the diagonal structure of $\boldsymbol{\Sigma}$, the observed variables $\boldsymbol{x}$ are conditionally independent given the values of the latent variables $\boldsymbol{s}$. Thus the reduced-dimensional distribution $\boldsymbol{s}$ is intended to model the dependencies between the observed variables. This is in contrast to PCA which treats the inter-variable dependencies and the independent noise identically. Factor analysis seeks for a factor loading matrix which best model the covariance structure of the observation data. In general, the columns of the factor loading matrix do not correspond to the principal subspace of the data. Maximum likelihood solution to factor analysis can be found in (Rubin and Thayer, 1982).

2.2 Isotropic Gaussian Noise

In general, factor loadings $\boldsymbol{A}$ differ from the principal axes due to the diagonal noise model $\boldsymbol{\Sigma}$. Principal components emerge when independent noise variables have common variance σ^2, i.e., noise is isotropic Gaussian. Recently Tipping and Bishop (Tipping and Bishop, 1999b) showed that under an isotropic noise structure, the maximum likelihood estimator $\boldsymbol{A}_{ML}$ spans a principal subspace

(which consists of scaled and rotated principal eigenvectors of the sample covariance matrix $\boldsymbol{R}$) even when the covariance model is approximate.

We assume an isotropic noise model, $\boldsymbol{v} \sim \mathcal{N}(\boldsymbol{0}, \sigma^2 \boldsymbol{I})$, which implies a probability distribution over the data space for a given $\boldsymbol{s}$ given by

$$p(\boldsymbol{x}|\boldsymbol{s}) = \frac{1}{(2\pi\sigma^2)^{\frac{m}{2}}} \exp\left\{-\frac{1}{2\sigma^2} \|\boldsymbol{x} - \boldsymbol{A}\boldsymbol{s}\|^2\right\}. \tag{2}$$

A Gaussian prior over the latent variables is used, i.e.,

$$p(\boldsymbol{s}) = \frac{1}{(2\pi)^{\frac{n}{2}}} \exp\left\{-\frac{1}{2} \|\boldsymbol{s}\|^2\right\}. \tag{3}$$

The marginal distribution of the data has the form

$$\begin{aligned} p(\boldsymbol{x}) &= \int p(\boldsymbol{x}|\boldsymbol{s}) p(\boldsymbol{s}) d\boldsymbol{s} \\ &= \frac{1}{(2\pi)^{\frac{m}{2}} |\boldsymbol{C}|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2} \boldsymbol{s}^T \boldsymbol{C}^{-1} \boldsymbol{s}\right\}, \end{aligned} \tag{4}$$

where the model covariance matrix is

$$\boldsymbol{C} = \boldsymbol{A}\boldsymbol{A}^T + \sigma^2 \boldsymbol{I}. \tag{5}$$

Then the log-likelihood of observing the data under this model is

$$\begin{aligned} \mathcal{L} &= \sum_{t=1}^{N} \log p(\boldsymbol{x}_t|\theta) \\ &= -\frac{Nm}{2} \log(2\pi) - \frac{N}{2} \log|\boldsymbol{C}| - \frac{N}{2} \mathrm{tr}\left\{\boldsymbol{C}^{-1}\boldsymbol{R}\right\}, \end{aligned} \tag{6}$$

where $\boldsymbol{R}$ is the sample covariance matrix given by

$$\boldsymbol{R} = \frac{1}{N} \sum_{t=1}^{N} \boldsymbol{x}_t \boldsymbol{x}_t^T. \tag{7}$$

It was shown in (Tipping and Bishop, 1999b) that with $\boldsymbol{C}$ given by (5), the only non-zero stationary points of $\frac{\partial \mathcal{L}}{\partial \boldsymbol{A}}$ occur for

$$\boldsymbol{A} = \boldsymbol{U}_n \left(\boldsymbol{\Lambda}_n - \sigma^2 \boldsymbol{I}\right)^{\frac{1}{2}} \boldsymbol{Q}, \tag{8}$$

where the n column vectors in $\boldsymbol{U}_n$ are eigenvectors of the sample covariance matrix $\boldsymbol{R}$, with corresponding eigenvalues in the diagonal matrix $\boldsymbol{\Lambda}_n$, and $\boldsymbol{Q}$ is an arbitrary $n \times n$ orthogonal rotation matrix.

Following Rubin and Thayer (Rubin and Thayer, 1982), Tipping and Bishop derived an EM algorithm for maximizing the log-likelihood (6). Here we briefly review the EM algorithm for probabilistic PCA. In the framework of EM, the latent variables $\{\boldsymbol{s}_t\}$ are treated as *missing data*. The complete-data log-likelihood $\mathcal{L}_c$ is given by

$$\begin{aligned}\mathcal{L}_c &= \sum_{t=1}^{N} \log p(\boldsymbol{x}_t, \boldsymbol{s}_t|\theta) \\ &= \sum_{t=1}^{N} \left[\text{const} - \frac{m}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\left\|\boldsymbol{x}_t - \boldsymbol{A}\boldsymbol{s}_t\right\|^2 - \frac{1}{2}\boldsymbol{s}_t^T\boldsymbol{s}_t\right]. \end{aligned} \tag{9}$$

In E-step, the expected complete-data log-likelihood, $\langle\mathcal{L}_c\rangle$ is calculated with the expectation being taken over the the posterior over the hidden variables. In M-step, the parameters are re-estimated in such a way that the expected complete-data log-likelihood is maximized. The probabilistic PCA algorithm is summarized below. See (Tipping and Bishop, 1999b) for more details.

Algorithm Outline: PPCA

E-step Compute sufficient statistics

$$\langle\boldsymbol{s}_t\rangle = \boldsymbol{M}^{-1}\boldsymbol{A}^T\boldsymbol{x}_t, \tag{10}$$

$$\left\langle\boldsymbol{s}_t\boldsymbol{s}_t^T\right\rangle = \sigma^2\boldsymbol{M}^{-1} + \langle\boldsymbol{s}_t\rangle\langle\boldsymbol{s}_t^T\rangle, \tag{11}$$

where $\boldsymbol{M} = \sigma^2\boldsymbol{I} + \boldsymbol{A}^T\boldsymbol{A}$.

M-step Re-estimate the parameters $\boldsymbol{A}$ and σ^2 by

$$\widehat{\boldsymbol{A}} = \left(\sum_{t=1}^{N}\boldsymbol{x}_t\left\langle\boldsymbol{s}_t^T\right\rangle\right)\left(\sum_{t=1}^{N}\left\langle\boldsymbol{s}_t\boldsymbol{s}_t^T\right\rangle\right)^{-1}, \tag{12}$$

$$\widehat{\sigma^2} = \frac{1}{m}\left\{\text{tr}\left[\boldsymbol{R} - \boldsymbol{R}\widehat{\boldsymbol{A}}\boldsymbol{M}^{-1}\widehat{\boldsymbol{A}}^T\right]\right\}. \tag{13}$$

2.3 Zero Noise Limit

Probabilistic PCA algorithm described in previous section is able to find scaled and rotated principal eigenvectors of the sample covariance matrix of the observed variables. It was derived in the framework of factor analysis with isotropic Gaussian noise model. PCA is a limiting case of the linear Gaussian

model (Roweis and Ghahramani, 1999) (which factor analysis is based on) as the covariance of the noise $\boldsymbol{v}$ becomes infinitesimally small and equal in all directions. Hence a simple EM algorithm for PCA can be obtained by taking the zero noise limit into account. In the zero noise limit ($\sigma^2 \to 0$), the likelihood of a data point $\boldsymbol{x}$ is dominated solely by the squared distance between it and its reconstruction $\boldsymbol{As}$. In such a case, the posterior collapses to a single point and the covariance becomes zero, i.e.,

$$\begin{aligned} p(\boldsymbol{s}|\boldsymbol{x}) &= \mathcal{N}\left(\left(\boldsymbol{A}^T\boldsymbol{A}\right)^{-1}\boldsymbol{A}^T\boldsymbol{x}, 0\right) \\ &= \delta\left(\boldsymbol{s} - \left(\boldsymbol{A}^T\boldsymbol{A}\right)^{-1}\boldsymbol{A}^T\boldsymbol{x}\right). \end{aligned} \tag{14}$$

Now inference reduces to simple least squares (LS) projection, which leads to a simple EM algorithm (Roweis, 1998) that is summarized below.

Algorithm Outline: PPCA (zero noise limit)

E-step Inference is carried out by LS projection,

$$\boldsymbol{S} = \left(\boldsymbol{A}^T\boldsymbol{A}\right)^{-1}\boldsymbol{A}^T\boldsymbol{X}, \tag{15}$$

where

$$\begin{aligned} \boldsymbol{S} &= [\boldsymbol{s}_1, \ldots, \boldsymbol{s}_N], \\ \boldsymbol{X} &= [\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N]. \end{aligned} \tag{16}$$

M-step Re-estimate the matrix $\boldsymbol{A}$ by

$$\widehat{\boldsymbol{A}} = \boldsymbol{X}\boldsymbol{S}^T\left(\boldsymbol{S}\boldsymbol{S}^T\right)^{-1}. \tag{17}$$

3 Sequential EM for Subspace Analysis

3.1 Separable Least Squares

The PPCA algorithm for the case of zero noise limit can be also derived in the framework of separable LS method. Moreover we can employ the sequential LS (also known as recursive LS) in order to develop an on-line algorithm which learns the principal subspace of the observed variables in a sequential manner.

As pointed out in (Roweis, 1998), in the case of zero noise limit, the likelihood of a data point $\boldsymbol{x}$ is dominated solely by the squared distance between it

and its reconstruction $\boldsymbol{As}$. In such a case, ML estimation of both $\boldsymbol{A}$ and $\boldsymbol{s}$ becomes a separable LS minimization problem. The LS estimates, $\boldsymbol{A}$ and $\boldsymbol{S}$ are computed by

$$\widehat{\boldsymbol{A}}, \widehat{\boldsymbol{S}} = \arg\min_{\boldsymbol{A},\boldsymbol{S}} \|\boldsymbol{X} - \boldsymbol{AS}\|_F^2 , \tag{18}$$

where $\|\cdot\|_F$ denotes the Frobenius norm. The separable LS minimization is carried out in two steps. First we minimize (18) with respect to $\boldsymbol{A}$ with $\boldsymbol{S}$ being fixed. It leads to

$$\widehat{\boldsymbol{A}} = \boldsymbol{X}\boldsymbol{S}^T \left(\boldsymbol{S}\boldsymbol{S}^T\right)^{-1} \tag{19}$$

which corresponds to the M-step in PPCA (for the case of zero noise limit).

The estimate $\widehat{\boldsymbol{A}}$ is substituted back into (18), then we obtain a new criterion which is a function of $\boldsymbol{S}$ only, i.e.,

$$\widehat{\boldsymbol{S}} = \arg\min_{\boldsymbol{S}} \left\|\boldsymbol{X}\boldsymbol{P}_S^{\perp}\right\|_F^2 , \tag{20}$$

where $\boldsymbol{P}_S^{\perp}$ is the orthogonal projection matrix given by

$$\boldsymbol{P}_S^{\perp} = \boldsymbol{I} - \boldsymbol{S}^T \left(\boldsymbol{S}\boldsymbol{S}^T\right)^{-1} \boldsymbol{S}. \tag{21}$$

The minimization of (20) is achieved when

$$\boldsymbol{S} = \left(\widehat{\boldsymbol{A}}^T\widehat{\boldsymbol{A}}\right)^{-1} \widehat{\boldsymbol{A}}^T \boldsymbol{X}, \tag{22}$$

which corresponds to the E-step.

3.2 Sequential LS

For sequential estimation of $\boldsymbol{A}$ and $\boldsymbol{s}$, we consider the weighted LS minimization problem where the objective function is given by

$$\mathcal{E} = \sum_{k=1}^{t} \beta^{t-k} \|\boldsymbol{x}_k - \boldsymbol{A}\boldsymbol{s}_k\|^2 , \tag{23}$$

where $0 < \beta \leq 1$ is the forgetting factor.

Our objective is to compute $\boldsymbol{A}_t$ and $\boldsymbol{s}_t$, assuming a good estimate of $\boldsymbol{s}_{t-1}$ (or equivalently $\boldsymbol{A}_{t-1}$) is available. The exponential weighting is used to de-emphasize old data in a time-varying environment. Set the derivative of $\mathcal{E}$ with respect to $\boldsymbol{A}$ to be zero, then we have

$$\boldsymbol{A}_t = \boldsymbol{R}_{xs,t} \left[\boldsymbol{R}_{ss,t}\right]^{-1}, \tag{24}$$

where

$$\begin{aligned} \boldsymbol{R}_{xs,t} &= \sum_{k=1}^{t} \beta^{t-k} \boldsymbol{x}_k \boldsymbol{s}_k^T, \\ \boldsymbol{R}_{ss,t} &= \sum_{k=1}^{t} \beta^{t-k} \boldsymbol{s}_k \boldsymbol{s}_k^T. \end{aligned} \tag{25}$$

Define $\boldsymbol{P}_t = \boldsymbol{R}_{ss,t}^{-1}$ and apply the matrix inversion lemma. Then we have a recursion equation for updating $\boldsymbol{P}_t$

$$\boldsymbol{P}_t = \frac{1}{\beta} \left\{ \boldsymbol{P}_{t-1} - \frac{\boldsymbol{P}_{t-1} \boldsymbol{s}_t \boldsymbol{s}_t^T \boldsymbol{P}_{t-1}}{\beta + \boldsymbol{s}_t^T \boldsymbol{P}_{t-1} \boldsymbol{s}_t} \right\}. \tag{26}$$

Using the recursion (26), the updating rule for $\boldsymbol{A}$ is given by

$$\begin{aligned} \boldsymbol{A}_t &= \boldsymbol{R}_{xs,t} \boldsymbol{P}_t \\ &= \boldsymbol{A}_{t-1} + \left[\boldsymbol{x}_t - \boldsymbol{A}_{t-1} \boldsymbol{s}_t\right] \frac{\boldsymbol{s}_t^T \boldsymbol{P}_{t-1}}{\beta + \boldsymbol{s}_t^T \boldsymbol{P}_{t-1} \boldsymbol{s}_t}. \end{aligned} \tag{27}$$

Algorithm Outline: Sequential EM

E-step Estimate $\boldsymbol{s}_t$ by the LS projection

$$\boldsymbol{s}_t = \left(\boldsymbol{A}_{t-1}^T \boldsymbol{A}_{t-1}\right)^{-1} \boldsymbol{A}_{t-1}^T \boldsymbol{x}_t. \tag{28}$$

M-step Estimate $\boldsymbol{A}_t$ by

$$\boldsymbol{A}_t = \boldsymbol{A}_{t-1} + \boldsymbol{\epsilon}_t \frac{\boldsymbol{s}_t^T \boldsymbol{P}_{t-1}}{\beta + \boldsymbol{s}_t^T \boldsymbol{P}_{t-1} \boldsymbol{s}_t}, \tag{29}$$

where

$$\boldsymbol{\epsilon}_t = \boldsymbol{x}_t - \boldsymbol{A}_{t-1} \boldsymbol{s}_t, \tag{30}$$

Remark: Our sequential EM subspace algorithm is reminiscent of PAST algorithm (Yang, 1995). However our algorithm is based on linear generative model, whereas the PAST algorithm considers the recognition model (which is an inverse of generative model). In addition, our algorithm does not find an orthogonal principal direction.

4 Rectified Subspace Analysis

Subspace analysis with nonnegativity constraints seeks for LS estimates, $\boldsymbol{A}$ and $\boldsymbol{S}$ which are computed from (18) with all elements of $\boldsymbol{A}$ and $\boldsymbol{S}$ being nonnegative. In order to find a solution which satisfies a nonnegativity constraint, NMF introduced a multiplicative updating rule. [1] On the other hand, Charles and Fyfe (Charles and Fyfe, 1998) used a simple rectification method with factor analysis network.

As in (Charles and Fyfe, 1998), we also use a rectifier defined by

$$[x]^+ = \begin{cases} x \text{ if } x \geq 0 \\ 0 \text{ if } x < 0. \end{cases} \tag{31}$$

The rectifier is operated in an elementwise fashion.

We call a subspace with nonnegativity constraint (through rectification) as *rectified subspace*. We incorporate the rectification into our sequential EM subspace algorithm described in previous section.

Algorithm Outline: Rectified Sequential EM

E-step Estimate $\boldsymbol{s}_t$ by the LS projection

$$\boldsymbol{s}_t = \left[\left(\boldsymbol{A}_{t-1}^T \boldsymbol{A}_{t-1} \right)^{-1} \boldsymbol{A}_{t-1}^T \boldsymbol{x}_t \right]^+. \tag{32}$$

M-step Estimate $\boldsymbol{A}_t$ by

$$\boldsymbol{A}_t = \left[\boldsymbol{A}_{t-1} + \boldsymbol{\epsilon}_t \frac{\boldsymbol{s}_t^T \boldsymbol{P}_{t-1}}{\beta + \boldsymbol{s}_t^T \boldsymbol{P}_{t-1} \boldsymbol{s}_t} \right]^+, \tag{33}$$

where

$$\boldsymbol{\epsilon}_t = \boldsymbol{x}_t - \boldsymbol{A}_{t-1} \boldsymbol{s}_t, \tag{34}$$

[1] They also considered a likelihood function under Poisson noise. See (Lee and Seung, 2001) for details.

5 Numerical Examples

5.1 Example 1

A simple numerical example is given here to confirm the fast convergence and high performance of the proposed sequential EM subspace algorithm. The algorithm is compared with the Oja's subspace rule (Oja, 1989). The 3-dimensional observation vector $\boldsymbol{x}$ was generated with its covariance matrix given by

$$\begin{bmatrix} 1.391 & 0.173 & -0.536 \\ 0.173 & 0.032 & -0.078 \\ -0.536 & -0.078 & 2.584 \end{bmatrix}.$$

As a performance measure, we use the subspace error (SE) which is defined by

$$\text{SE} = \frac{1}{\sqrt{n}} \left\| \boldsymbol{P}_A^{\perp} \boldsymbol{U}_n \boldsymbol{U}_n^T \right\|, \tag{35}$$

where $\boldsymbol{P}_A^{\perp}$ is the orthogonal projection matrix onto the signal subspace, i.e.,

$$\boldsymbol{P}_A^{\perp} = \boldsymbol{I} - \boldsymbol{A} \left(\boldsymbol{A}^T \boldsymbol{A} \right)^{-1} \boldsymbol{A}^T, \tag{36}$$

which is decided by the estimated values of $\boldsymbol{A}$. $\boldsymbol{U}_n \boldsymbol{U}_n^T$ is the projection matrix onto the signal subspace that is computed from the SVD of the covariance matrix of the input data. When the estimated $\boldsymbol{A}$ spanns the true signal subspace, $\boldsymbol{P}_A^{\perp}$ should orthogonal to $\boldsymbol{U}_n \boldsymbol{U}_n^T$. Hence the SE defined above is able to serve as a performance measure.

The matrices $\boldsymbol{A}_0$ (for the sequential EM) and $\boldsymbol{W}_0$ (for Oja's subspace rule) were initialized as a random matrix whose elements are drawn from uniformly distributed random variables over [0,1]. The learning rate in the Oja's subspace rule was $\eta_t = .01$. In Fig. 1, one can observe that the sequential EM algorithm converges to a solution much faster than the Oja's subspace rule and even after convergence, the former shows slight better performance than the latter.

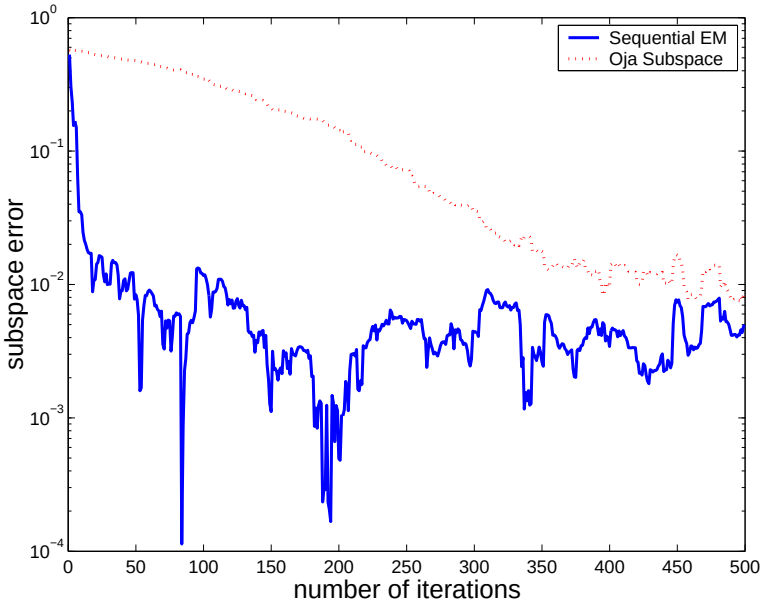


Fig. 1. Convergence comparison for sequential EM algorithm and Oja's subspace rule in terms of the subspace error.

5.2 Example 2

In this example, we tested the rectified subspace algorithm which is a slight modification of the sequential EM subsapce algorithm, with the simple bar data that was proposed by Földiák (Földiák, 1992). The simple bar data consists of an 8×8 square grid containing a random mixture of horizontal and vertical bars. The bars are chosen randomly such that each of the 16 possible bars appears with a probability of $\frac{1}{8}$ independent of each other (see Fig. 2). Several horizontal or vertical bars appear with some of them being overlapped. Hence, the conventional PCA fails to identify basis bars which constitute the data (see Fig. 3). On the other hand, the rectified subspace analysis method inherently allows nonsubtractive combination of nonnegative basis vectors and successfully identify basis bars (see Fig. 4). In the rectified subspace algorithm, $\beta = 0.99$ was used.

5.3 Example 3: PET Images

We performed $H_2^{15}O$ PET scans on seven dogs at rest and after pharmacological stress using Adnenosine or Dipyridamole. All the scans were acquired with an ECAT EXACT 47 scanner (Simens-CTI, Knoxville, USA) which has an intrinsic resolution of 5.2 mm FWHM (full width at half maximum) and images 47 contiguous planes with thickness of 3.4 mm simultaneously for a longitudinal field of view of 16.2 cm. Before $H_2^{15}O$ administration, transmis-

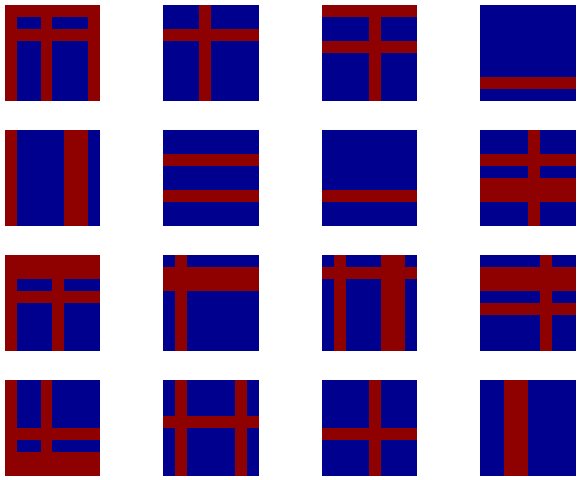


Fig. 2. Simple bar data.

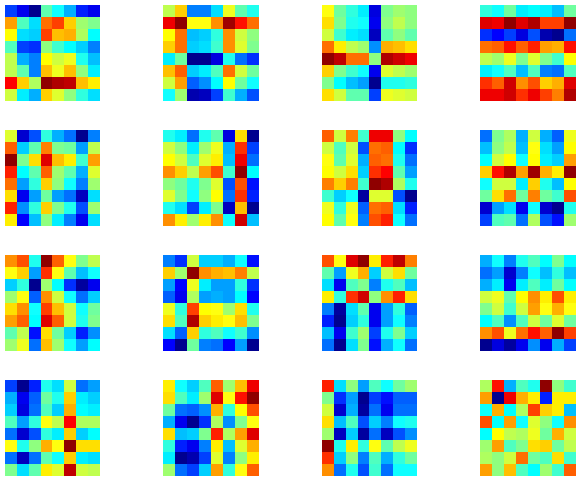


Fig. 3. Basis images learned by PCA.

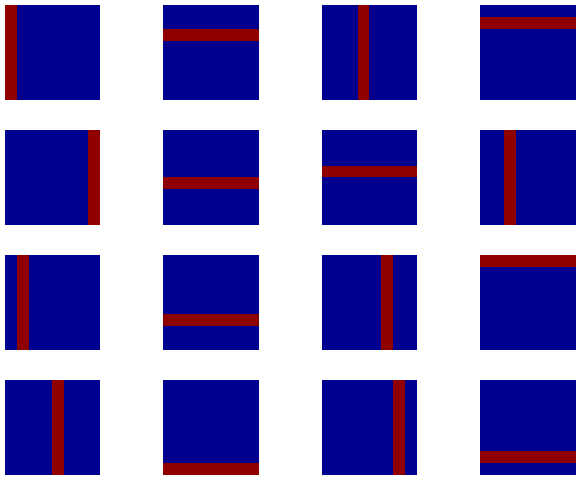


Fig. 4. Basis images learned by the rectified subspace analysis.

sion scanning was performed using three Ge-68 rod sources for attenuation correction. Dynamic emission scans (5 sec×12, 10 sec×9, 30 sec×3) were initiated simultaneously with the injection of 555-740 MBq $H_2^{15}O$. Transaxial images were reconstructed by means of a filtered back-projection algorithms as $128 \times 128 \times 47$ matrices with a size of $2.1 \times 2.1 \times 3.4$ mm

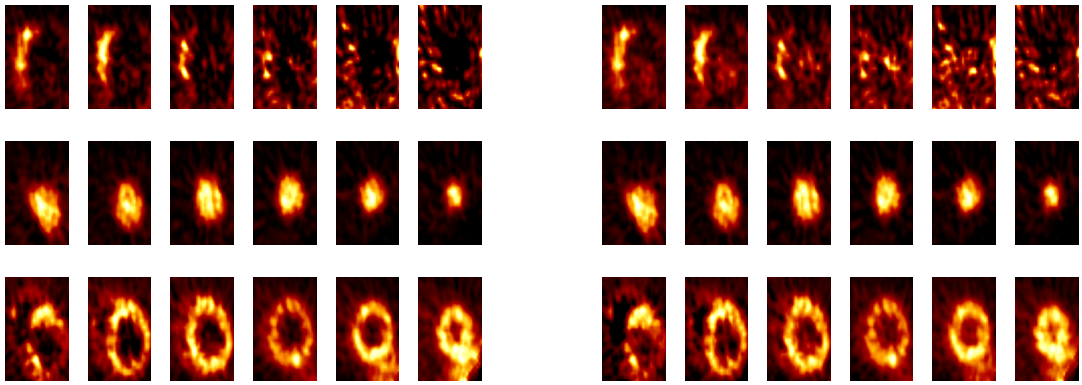
The initial eighteen frames (two minutes) of PET images were used for analysis. The dynamic PET images were re-oriented to short axis and were resampled to produce 1-cm-thick slices in order to increase the signal to noise ratio. Only the cardiac regions were then masked to remove extra cardiac components and to reduce the quantity of data and hence the burden of computation. The resulting masked images with dimension of $32 \times 32 \times 6 \times 18$ (pixel $\times$ pixel $\times$ plane $\times$ frame) were reformulated to 18×6144 (frame $\times$ pixel) data matrix $\boldsymbol{X}$.

In the rectified subspace analysis algorithm, we used $\beta = .99$. In most of simulations, all the parameters ($\boldsymbol{A}$ and $\boldsymbol{S}$ converged within 500 iterations. For NMF, all the data points were re-used 100 times to ensure the convergence.

Each row of the matrix $\boldsymbol{S}$ corresponds to basis image which represent cardiac component. Fig. 5 shows the basis images that we obtained using the rectified subspace algorithm and the NMF. For both cases, three cardiac components (right ventricle, left ventricle, myocardium) were successfully extracted. Each column vector of the matrix $\boldsymbol{A}$ represent the time activity curve (TAC) which is useful to calculate blood flow estimation (Lee et al., 2000). Fig. 6 shows the TAC for the rectified subspace algorithm and the NMF. These TACs showed reasonable shapes that we were familiar with (two peaks at each of right ventricle and left ventricle, more dispersion in left ventricle and mycardium). With the assumption of proper number of factors, both the rectified subspace analysis algorithm and NMF showed good results. The benefit of the rectified subspace analysis algorithm is a fast sequential algorithm (due to the sequential LS implementation), so will be useful for handling with high dimensional data matrix.

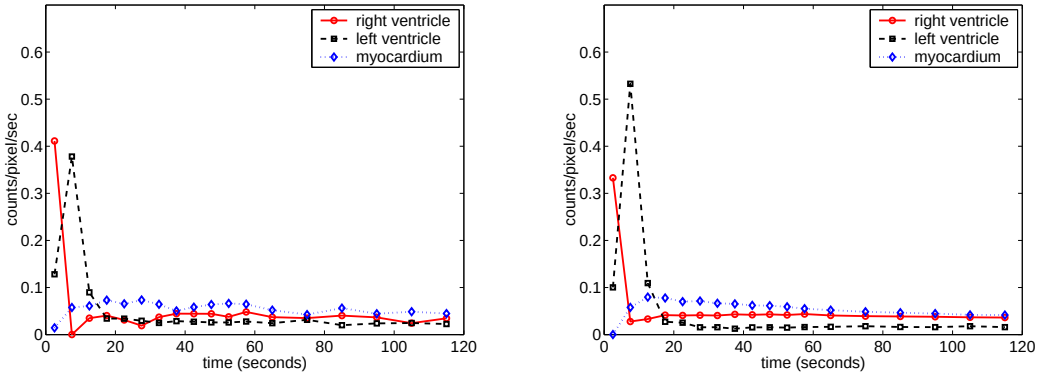
6 Conclusion

We have presented a sequential EM algorithm for subspace analysis which is able to compute the principal subspace of pre-specified dimension sequentially. Motivated by PPCA (Tipping and Bishop, 1999b) and EM-PCA (Roweis, 1998), in the framework of probabilistic model-based PCA, we employed the sequential LS method to develop the sequential EM subspace algorithm. The algorithm was shown to achieve the convergence much faster, compared to the Oja's subspace algorithm which is based on the gradient descent method. We



(a) (b)

Fig. 5. Basis images computed by: (a) the rectified subspace analysis algorithm; (b) NMF (LS method).



(a) (b)

Fig. 6. Time activity curves obtained by: (a) the rectified subspace analysis algorithm; (b) NMF (LS method).

also presented a slight modification of the sequential EM subspace algorithm, by employing a rectification operation, which leaded to the rectified subspace analysis algorithm. Its useful behavior was demonstrated through simple bar data and dynamic PET images. The number of feasible basis vectors in the rectified subspace algorithm, could be determined by a method in (Charles and Fyfe, 1998) where the difference between the variance of each observed variable and the squared Euclidean norm of each row vector of $\widehat{\boldsymbol{A}}$, was considered such that this difference is nonnegative.

7 Acknowledgment

The author would like to thank Prof. D. S. Lee and Dr. J. S. Lee in the Medical School of Seoul National University, for providing PET image data and their

useful discussion. This work was supported by KOSEF 2000-2-20500-009-5 and by Korea Ministry of Science and Technology under Brain Science and Engineering Research Program and by BK 21 in POSTECH.

References

Charles, D., Fyfe, C., 1998. Modelling multiple cause structure using rectification constraints. Network: Computation in Neural Systems 9, 167–182.

Diamantaras, K. I., Kung, S. Y., 1996. Principal Component Neural Networks: Theory and Applications. John Wiely & Sons, INC.

Földiák, P., 1992. Models of sensory coding. Ph.D. thesis, University of Cambridge.

Ghahramani, Z., Hinton, G. E., 1997. The EM algorithm for mixtures of factor analyzers. Tech. Rep. CRG-TR-96-1, Department of Computer Science, University of Toronto.

Lai, P. L., Fyfe, C., 2000. Simultaneous identification of face and orientation. Neural Processing Letters 12 (1), 33–40.

Lee, D. D., Seung, H. S., 1999. Learning the parts of objects by non-negative matrix factorization. Nature 401, 788–791.

Lee, D. D., Seung, H. S., 2001. Algorithms for non-negative matrix factorization. In: Advances in Neural Information Processing Systems. Vol. 13. MIT Press.

Lee, D. S., Lee, J. S., Ahn, J. Y., 2000. Application of independent component analysis to dynamic $H_2^{15}O$ positron emission tomography. In: Prof. ICONIP. pp. 1221–1226.

Lee, J. S., Lee, D. D., Choi, S., Park, K. S., Lee, D. S., 2001. Nonnegative matrix factorization of dynamic images in nuclear medicine. In: IEEE Medical Imaging Conference.

Oja, E., 1989. Neural networks, principal component analysis, and subspaces. International Journal of Neural Systems 1, 61–68.

Roweis, S. T., 1998. EM algorithms for PCA and SPCA. In: Advances in Neural Information Processing Systems. Vol. 10. MIT press, pp. 626–632.

Roweis, S. T., Ghahramani, Z., 1999. A unifying review of linear Gaussian models. Neural Computation 11 (2), 305–345.

Rubin, D., Thayer, D., 1982. EM algorithms for ML factor analysis. Psychometrika 47, 69–76.

Sirovich, L., Kirby, M., 1987. Low-dimensional procedure for the charaterization of human faces. Journal of the Optical Society of America A 4 (3), 519–524.

Tipping, M. E., Bishop, C. M., 1999a. Mixtures of probabilistic principal component analyzers. Neural Computation 11 (2), 443–482.

Tipping, M. E., Bishop, C. M., 1999b. Probabilistic principal component analysis. Journal of the Royal Statistical Society B 61 (3), 611–622.

Turk, M., Pentland, A., 1991. Eigenfaces for recognition. Journal of Cognitive Neuroscience 3 (1), 71–86.

Yang, B., 1995. Projection approximation subsapce tracking. IEEE Trans. Signal Processing 43 (1), 95–107.