# ITERATIVE PROJECTION APPROXIMATION ALGORITHMS FOR PCA

*Seungjin Choi* [§], *Jong-Hoon Ahn* [†], *Andrzej Cichocki* [‡]

[§] Department of Computer Science, POSTECH, Korea
*seungjin@postech.ac.kr*
[†] Department of Physics, POSTECH, Korea
*jonghoon@postech.ac.kr*
[‡] Advanced Brain Signal Processing Lab, BSI, RIKEN, Japan
*cia@brain.riken.jp*

**ABSTRACT**

In this paper we introduce a new error measure, *integrated reconstruction error* (IRE), the minimization of which leads to principal eigenvectors (without rotational ambiguity) of the data covariance matrix. Then we present iterative algorithms for the IRE minimization, through the projection approximation. The proposed algorithm is referred to as COnstrained Projection Approximation (COPA) algorithm and its limiting case is called COPAL. We also discuss regularized algorithms, referred to as R-COPA and R-COPAL. Numerical experiments demonstrate that these algorithms successfully find exact principal eigenvectors of the data covariance matrix.

## 1. INTRODUCTION

Principal component analysis (PCA) or principal subspace analysis (PSA) [1] is a fundamental multivariate data analysis method which is encountered into a variety of areas in neural networks, signal processing, and machine learning. A variety of adaptive (on-line) algorithms for PCA or PSA can be found in neural networks literature [2, 3, 4, 5, 6]. See also [7] and references therein.

The power iteration is a classical method for estimating the largest eigenvector of a symmetric matrix. The subspace iteration is a direct extension of the power iteration, computing subspace spanned by principal eigenvectors of a symmetric matrix. The natural power method is an exemplary instance of the subspace iteration, where the invariant subspace spanned by the $n$ largest eigenvectors of the data covariance matrix, is determined [8]. The natural power iteration provides a general framework for several well-known subspace algorithms, including Oja's subspace rule [2], PAST [9], and OPAST [10].

A common derivation of PSA, is terms of a linear (orthogonal) projection $\boldsymbol{W} = [\boldsymbol{w}_1 \cdots \boldsymbol{w}_n] \in \mathbb{R}^{m \times n}$ such that given a centered data matrix $\boldsymbol{X} = [\boldsymbol{x}(1) \cdots \boldsymbol{x}(N)] \in \mathbb{R}^{m \times N}$, the reconstruction error $\|\boldsymbol{X} - \boldsymbol{W}\boldsymbol{W}^\top \boldsymbol{X}\|_F^2$ is minimized, where $\|\cdot\|_F$ denotes the Frobenius norm (Euclidean norm). It is well known that the reconstruction error is blind to an arbitrary rotation. Thus, the minimization of the reconstruction error leads to $\boldsymbol{W} = \boldsymbol{U}_1 \boldsymbol{Q}$ where $\boldsymbol{Q} \in \mathbb{R}^{n \times n}$ is an arbitrary orthogonal matrix and the eigendecomposition of the covariance matrix $\boldsymbol{C} = \boldsymbol{X}\boldsymbol{X}^\top$ is given by

$$\boldsymbol{C} = \begin{bmatrix} \boldsymbol{U}_1 & \boldsymbol{U}_2 \end{bmatrix} \begin{bmatrix} \boldsymbol{\Lambda}_1 & 0 \\ 0 & \boldsymbol{\Lambda}_2 \end{bmatrix} \begin{bmatrix} \boldsymbol{U}_1 & \boldsymbol{U}_2 \end{bmatrix}^\top, \qquad (1)$$

where $\boldsymbol{U}_1 \in \mathbb{R}^{m \times n}$ contains $n$ largest eigenvectors, the rest of eigenvectors are in $\boldsymbol{U}_2 \in \mathbb{R}^{m \times (m-n)}$, and associated eigenvalues are in $\boldsymbol{\Lambda}_1,\ \boldsymbol{\Lambda}_2$ with $\lambda_1 > \lambda_2 > \cdots > \lambda_m$.

The natural power iteration-based methods and probabilistic methods such as PPCA [11] and EM-PCA [12], find principal subspace. In this paper we introduce the *integrated reconstruction error* (IRE) and show that its minimization leads to exact principal eigenvectors (without rotational ambiguity). We present iterative algorithms, referred to as COnstrained Projection Approximation (COPA) algorithm and its limiting case is called COPAL, which iteratively minimize the IRE, through projection approximation. We also present their regularized version, which is useful for a rank-deficient case.

## 2. INTEGRATED RECONSTRUCTION ERROR

It was shown in [9] that the reconstruction error $\mathcal{J}_{RE} = \|\boldsymbol{X} - \boldsymbol{W}\boldsymbol{W}^\top \boldsymbol{X}\|_F^2$ attains the global minimum if and only if $\boldsymbol{W} = \boldsymbol{U}_1 \boldsymbol{Q}$. Now we introduce the IRE that is summarized below.

**Definition 1 (IRE)** *The integrated reconstruction error, $\mathcal{J}_{IRE}$, is defined as a linear combination of $n$ partial reconstruction errors (PRE), $\mathcal{J}_i = \left\|\boldsymbol{X} - \boldsymbol{W}\boldsymbol{E}_i\boldsymbol{W}^\top \boldsymbol{X}\right\|_F^2$, i.e.,*

$$\begin{aligned} \mathcal{J}_{IRE}(\boldsymbol{W}) &= \sum_{i=1}^{n} \alpha_i \mathcal{J}_i \\ &= \sum_{i=1}^{n} \alpha_i \left\|\boldsymbol{X} - \boldsymbol{W}\boldsymbol{E}_i\boldsymbol{W}^\top \boldsymbol{X}\right\|^2, \end{aligned} \qquad (2)$$

*where $\alpha_i > 0$ and $\boldsymbol{E}_i \in \mathbb{R}^{n \times n}$ is a diagonal matrix, defined by*

$$[\boldsymbol{E}_i]_{jj} = \begin{cases} 1 & \text{for } j = 1, \ldots, i, \\ 0 & \text{for } j = i+1, \ldots, n. \end{cases}$$

**Theorem 1** *The IRE is minimized if and only if $\boldsymbol{W} = \boldsymbol{U}_1$.*

The last term in IRE, $\mathcal{J}_n$, is the standard reconstruction error. It was shown in [9] that $\boldsymbol{W}$ is a stationary point of $\mathcal{J}_n$ if and only if $\boldsymbol{W} = \boldsymbol{U}_1\boldsymbol{Q}$ (hence $\boldsymbol{W}^\top\boldsymbol{W} = \boldsymbol{I}$ is satisfied). All stationary points of $\mathcal{J}_n$ are saddle points, except when $\boldsymbol{U}_1$ contains the $n$ dominant eigenvectors of $\boldsymbol{C}$. In that case, $\mathcal{J}_n$ attains the global minimum.

PREs $\mathcal{J}_i$ are of the form

$$\mathcal{J}_i = \left\|\boldsymbol{X} - \left(\boldsymbol{w}_1\boldsymbol{w}_1^\top + \cdots + \boldsymbol{w}_i\boldsymbol{w}_i^\top\right)\boldsymbol{X}\right\|_F^2,$$

where $\boldsymbol{w}_i$ represents the $i$th column vector of $\boldsymbol{W}$. The $i$-dimensional principal subspace is completely included in the $(i+1)$-dimensional principal subspace. Hence,the IRE is minimized when each PRE

$\mathcal{J}_i$ is minimized. Minimizing each $\mathcal{J}_i$ is reminiscent of the deflation method where the eigenvectors of $\boldsymbol{C}$ is extracted one by one. Thus, it is expected that the minimization of IRE leads to principal eigenvectors of $\boldsymbol{C}$. The rigorous proof is left out due to the space limitation. A major difference between the deflation method and our method is that the former extracts principal components one by one and the latter find principal components simultaneously.

## 3. ITERATIVE ALGORITHMS

### 3.1. COPA

The projection approximation [9] assumes that the difference between $\boldsymbol{W}_{(k+1)}^\top \boldsymbol{X}$ and $\boldsymbol{W}_{(k)}^\top \boldsymbol{X}$ is small, which leads us to consider the following objective function

$$\mathcal{J}_{IRE}(\boldsymbol{W}) = \frac{1}{2}\sum_{i=1}^{n} \alpha_i \left\| \boldsymbol{X} - \boldsymbol{W}_{(k+1)}\boldsymbol{E}_i\boldsymbol{Y}_{(k)} \right\|^2, \tag{3}$$

where $\boldsymbol{Y}_{(k)} = \boldsymbol{W}_{(k)}^\top \boldsymbol{X}$.

The gradient of (3) with respect to $\boldsymbol{W}_{(k+1)}$ is given by

$$\frac{\partial \mathcal{J}_{IRE}}{\partial \boldsymbol{W}_{(k+1)}} = -\boldsymbol{X}\boldsymbol{Y}_{(k)}^\top \boldsymbol{\Sigma} + \boldsymbol{W}_{(k+1)}\left[\left(\boldsymbol{Y}_{(k)}\boldsymbol{Y}_{(k)}^\top\right)\odot \boldsymbol{\Gamma}\right], \tag{4}$$

where

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sum_{i=1}^n \alpha_i & 0 & 0 & \cdots & 0 \\ 0 & \sum_{i=2}^n \alpha_i & 0 & \cdots & 0 \\ \vdots & & & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \alpha_n \end{bmatrix},$$

$$\boldsymbol{\Gamma} = \begin{bmatrix} \sum_{i=1}^n \alpha_i & \sum_{i=2}^n \alpha_i & \sum_{i=3}^n \alpha_i & \cdots & \alpha_n \\ \sum_{i=2}^n \alpha_i & \sum_{i=2}^n \alpha_i & \sum_{i=3}^n \alpha_i & \cdots & \alpha_n \\ \sum_{i=3}^n \alpha_i & \sum_{i=3}^n \alpha_i & \sum_{i=3}^n \alpha_i & \cdots & \alpha_n \\ \vdots & & & \ddots & \vdots \\ \alpha_n & \alpha_n & \alpha_n & \cdots & \alpha_n \end{bmatrix},$$

and $\odot$ is the Hadamard product (element-wise product). With these definitions, it follows from $\frac{\partial \mathcal{J}_{IRE}}{\partial \boldsymbol{W}_{(k+1)}} = 0$ that

$$\begin{aligned} \boldsymbol{W}_{(k+1)} &= \left[\boldsymbol{X}\boldsymbol{Y}_{(k)}^\top\right]\boldsymbol{\Sigma}\left[\left(\boldsymbol{Y}_{(k)}\boldsymbol{Y}_{(k)}^\top\right)\odot\boldsymbol{\Gamma}\right]^{-1} \\ &= \left[\boldsymbol{X}\boldsymbol{Y}_{(k)}^\top\right]\left[\left(\boldsymbol{Y}_{(k)}\boldsymbol{Y}_{(k)}^\top\right)\odot\left(\boldsymbol{\Gamma}\boldsymbol{\Sigma}^{-1}\right)\right]^{-1} \\ &= \boldsymbol{X}\boldsymbol{Y}_{(k)}^\top\left[\mathsf{U}\left(\boldsymbol{Y}_{(k)}\boldsymbol{Y}_{(k)}^\top\right)\right]^{-1}, \end{aligned} \tag{5}$$

where $\mathsf{U}(\boldsymbol{Y})$ is an element-wise operator, whose arguments $Y_{ij}$ are transformed by

$$\mathsf{U}(Y_{ij}) = \begin{cases} Y_{ij}\frac{\sum_{l=i}^n \alpha_l}{\sum_{l=j}^n \alpha_l} & \text{if } i > j, \\ Y_{ij} & \text{if } i \le j. \end{cases} \tag{6}$$

The operator $\mathsf{U}(\boldsymbol{Y})$ results from the structure of $\boldsymbol{\Gamma}\boldsymbol{\Sigma}^{-1}$ given by

$$\boldsymbol{\Gamma}\boldsymbol{\Sigma}^{-1} = \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ \frac{\sum_{i=2}^n \alpha_i}{\sum_{i=1}^n \alpha_i} & 1 & 1 & \cdots & 1 \\ \frac{\sum_{i=3}^n \alpha_i}{\sum_{i=1}^n \alpha_i} & \frac{\sum_{i=3}^n \alpha_i}{\sum_{i=2}^n \alpha_i} & 1 & \cdots & 1 \\ \vdots & \vdots & \vdots & & \vdots \\ \frac{\alpha_n}{\sum_{i=1}^n \alpha_i} & \frac{\alpha_n}{\sum_{i=2}^n \alpha_i} & \frac{\alpha_n}{\sum_{i=3}^n \alpha_i} & \cdots & 1 \end{bmatrix}.$$

Replacing $\boldsymbol{Y}_{(k)}$ by $\boldsymbol{W}_{(k)}^\top \boldsymbol{X}$, leads to the updating rule for COPA:

$$\boldsymbol{W}_{(k+1)} = \boldsymbol{C}\boldsymbol{W}_{(k)}\left[\mathsf{U}\left(\boldsymbol{W}_{(k)}^\top \boldsymbol{C}\boldsymbol{W}_{(k)}\right)\right]^{-1}. \tag{7}$$

### 3.2. COPAL

We consider the limiting case where $\frac{\alpha_{i+1}}{\alpha_i} \to 0$ for $i = 1, \ldots, n-1$, that is, weighting $\alpha_i$'s are rapidly diminishing as $i$ increases. In such a case, the operator $\mathsf{U}(\cdot)$ becomes the conventional upper-triangularization operator $\mathsf{U}_T$ which is given by

$$\mathsf{U}_T(Y_{ij}) = \begin{cases} 0 & \text{if } i > j, \\ Y_{ij} & \text{if } i \le j. \end{cases} \tag{8}$$

This leads to the COPAL algorithm

$$\boldsymbol{W}_{(k+1)} = \boldsymbol{C}\boldsymbol{W}_{(k)}\left[\mathsf{U}_T\left(\boldsymbol{W}_{(k)}^\top \boldsymbol{C}\boldsymbol{W}_{(k)}\right)\right]^{-1}. \tag{9}$$

**Theorem 2** *The fixed point $\boldsymbol{W}$ of the COPAL (9) satisfies $\boldsymbol{W} = \boldsymbol{U}_1\boldsymbol{\Upsilon}$ (after each column vector of $\boldsymbol{W}$ is normalized), where $\boldsymbol{\Upsilon}$ is a diagonal matrix with its diagonal entries being 1 or -1, provided that the $n$th and $(n+1)$th eigenvalues of $\boldsymbol{C}$ are distinct and the initial weight matrix $\boldsymbol{W}_{(0)}$ meets a mild condition, saying that there exists a nonsingular matrix $\boldsymbol{L} \in \mathbb{R}^{(m-n)\times n}$ such that $\boldsymbol{U}_2^\top \boldsymbol{W}_{(0)} = \boldsymbol{L}\boldsymbol{U}_1^\top \boldsymbol{W}_{(0)}$ for a randomly chosen $\boldsymbol{W}_{(0)}$.*

*Proof.* It can be proved in a similar way to the method in [13].

### 3.3. R-COPA and R-COPAL

We consider the rank-deficient case where $n$ is greater than the rank of the data covariance matrix. For numerical stability, we add a regularizer term in (3). The modified objective function is given by

$$\begin{aligned} \widetilde{\mathcal{J}}_{IRE}(\boldsymbol{W}) &= \frac{1}{2}\sum_{i=1}^{n} \alpha_i \left\| \boldsymbol{X} - \boldsymbol{W}_{(k+1)}\boldsymbol{E}_i\boldsymbol{Y}_{(k)} \right\|^2 \\ &\quad + \frac{1}{2}\beta \,\mathrm{tr}\{\boldsymbol{W}_{(k+1)}\boldsymbol{W}_{(k+1)}^\top\}, \end{aligned} \tag{10}$$

where $\beta$ is the Lagrangian multiplier.

The minimization of (10) leads to the updating rule that is of form

$$\boldsymbol{W}_{(k+1)} = \boldsymbol{C}\boldsymbol{W}_{(k)}\left[\mathsf{U}\left(\boldsymbol{W}_{(k)}^\top \boldsymbol{C}\boldsymbol{W}_{(k)}\right) + \beta\boldsymbol{\Sigma}^{-1}\right]^{-1}. \tag{11}$$

This is referred to as R-COPA.

We also consider the following objective function with a weighted regularization term,

$$\begin{aligned} \bar{\mathcal{J}}_{IRE}(\boldsymbol{W}) &= \frac{1}{2}\sum_{i=1}^{n} \alpha_i \left\| \boldsymbol{X} - \boldsymbol{W}_{(k+1)}\boldsymbol{E}_i\boldsymbol{Y}_{(k)} \right\|^2 \\ &\quad + \frac{1}{2}\beta \,\mathrm{tr}\{\boldsymbol{W}_{(k+1)}\boldsymbol{\Sigma}\boldsymbol{W}_{(k+1)}^\top\}. \end{aligned} \tag{12}$$

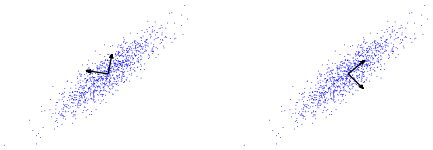
Incorporating with the limiting case, $\frac{\alpha_{i+1}}{\alpha_i} \to 0$, the minimization of (12) leads to

$$\boldsymbol{W}_{(k+1)} = \boldsymbol{C}\boldsymbol{W}_{(k)}\left[\mathsf{U}_T\left(\boldsymbol{W}_{(k)}^\top \boldsymbol{C}\boldsymbol{W}_{(k)}\right) + \beta\boldsymbol{I}\right]^{-1}. \tag{13}$$

Algorithms are summarized in Table 1, where the constrained natural power iteration [13] is a variation of the natural power iteration [8], incorporating with the upper-triangularization operator $\mathsf{U}_T$.

**Table 1**. The outline of updating rules and characteristics of algorithms.

| Algorithm | Updating rule | Type |
|---|---|---|
| Batch PAST [9] | $\boldsymbol{W}_{(k+1)} = \boldsymbol{C}\boldsymbol{W}_{(k)} \left[\boldsymbol{W}_{(k)}^{\top}\boldsymbol{C}\boldsymbol{W}_{(k)}\right]^{-1}$ | PSA |
| Batch Natural Power [8] | $\boldsymbol{W}_{(k+1)} = \boldsymbol{C}\boldsymbol{W}_{(k)} \left[\boldsymbol{W}_{(k)}^{\top}\boldsymbol{C}^2\boldsymbol{W}_{(k)}\right]^{-\frac{1}{2}}$ | PSA |
| Constrained Natural Power [13] | $\boldsymbol{W}_{(k+1)} = \boldsymbol{C}\boldsymbol{W}_{(k)} \left[\mathsf{U}_T\left(\boldsymbol{W}_{(k)}^{\top}\boldsymbol{C}^2\boldsymbol{W}_{(k)}\right)\right]^{-\frac{1}{2}}$ | PCA |
| COPA | $\boldsymbol{W}_{(k+1)} = \boldsymbol{C}\boldsymbol{W}_{(k)} \left[\mathsf{U}\left(\boldsymbol{W}_{(k)}^{\top}\boldsymbol{C}\boldsymbol{W}_{(k)}\right)\right]^{-1}$ | PCA |
| R-COPA | $\boldsymbol{W}_{(k+1)} = \boldsymbol{C}\boldsymbol{W}_{(k)} \left[\mathsf{U}\left(\boldsymbol{W}_{(k)}^{\top}\boldsymbol{C}\boldsymbol{W}_{(k)}\right) + \beta\boldsymbol{\Sigma}^{-1}\right]^{-1}$ | Regularized PCA |
| COPAL | $\boldsymbol{W}_{(k+1)} = \boldsymbol{C}\boldsymbol{W}_{(k)} \left[\mathsf{U}_T\left(\boldsymbol{W}_{(k)}^{\top}\boldsymbol{C}\boldsymbol{W}_{(k)}\right)\right]^{-1}$ | PCA |
| R-COPAL | $\boldsymbol{W}_{(k+1)} = \boldsymbol{C}\boldsymbol{W}_{(k)} \left[\mathsf{U}_T\left(\boldsymbol{W}_{(k)}^{\top}\boldsymbol{C}\boldsymbol{W}_{(k)}\right) + \beta\boldsymbol{I}\right]^{-1}$ | Regularized PCA |



(a) (b)

**Fig. 1**. Principal directions computed: (a) by the natural power (or the batch PAST); (b) COPA (or COPAL). The natural power method finds rotated principal directions, whereas our algorithms (COPA and COPAL) estimate exact principal directions of the two-dimensional data.

## 4. NUMERICAL EXPERIMENTS

Numerical examples are provided, in order to verify that the weight matrix $\boldsymbol{W}$ in COPA and COPAL converges to the true eigenvectors of the data covariance matrix $\boldsymbol{C}$. The first experiment was carried out with 2-dimensional vector sequences of length 1000. Fig. 1 shows the data scatter plots and principal directions computed by the natural power method and by our algorithms (COPA and COPAL). One can see that principal directions estimated by the natural power method are rotated eigenvectors of the data covariance matrix (i.e., principal subspace). On the other hand, COPA or COPAL finds exact principal directions (see Fig. 1 (b)).

The second experiment involves the useful behavior of our algorithms for high-dimensional data, showing that even for the case of high-dimensional data, our algorithms successfully estimate exact first few principal directions of data. To this end, we generated 5000 5-dimensional Gaussian vectors (with zero mean and unit variance) and applied a linear transform to construct the data matrix $\boldsymbol{X} \in \mathbb{R}^{1000\times 5000}$. The rank of the covariance matrix $\boldsymbol{C}$ is 5. COPA and COPAL algorithms in (7) and (9) were applied to find 3 principal eigenvectors from this data matrix. For the case of COPA, we used $\alpha_1 = 1,\ \alpha_2 = 0.1,\ \alpha_3 = 0.01$. Results are shown in Fig. 2 where both COPA and COPAL algorithms show similar behavior.

Next experiment is regarding the regularized algorithms. The 10-dimensional data vectors were generated by a linear transform of 3-dimensional Gaussian vectors. COPAL and R-COPAL were applied with $n = 5$ (see Fig. 3). In this case, COPAL showed severe numerical instability during iterations, although first three eigenvectors were successfully extracted. Frequently, COPAL became unstable. On the other hand, R-COPAL successfully extracted first three eigenvectors and found last two eigenvectors that are orthogonal to 4th and 5th eigenvectors computed by SVD. For R-COPAL,

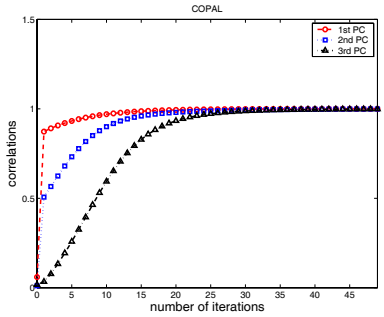
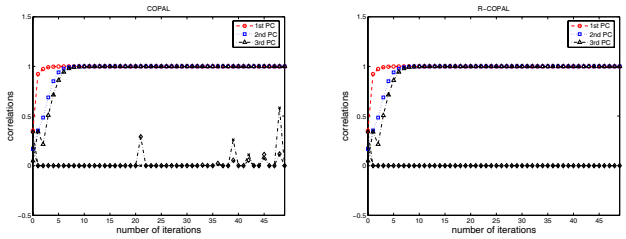



**Fig. 2**. Evolution of weight vectors in COPAL, is shown. Correlations represent the absolute value of the inner product between a weight vector and a true eigenvector (computed by SVD).

$\beta = 0.01$ was used in our simulation.




(a) (b)

**Fig. 3**. Evolution of weight vectors in COPAL and R-COPAL, is shown. Correlations represent the absolute value of the inner product between a weight vector and a true eigenvector (computed by SVD). The last two eigenvectors estimated by COPAL (a), are oscillating, which causes a numerical instability. In contrast, R-COPAL is numerically stable.

As a real world example, COPAL was applied to USPS handwritten digit data, in order to determine eigen-digits (see Fig. 4. Each image is the size of $16 \times 16$, which leads to a 256-dimensional vector. First 100 principal components were estimated by COPAL as well as SVD and batch version of PAST with deflation. Although the deflation method determines eigenvectors without rotational ambiguity, however, error accumulation is propagated as $n$ increases (see Fig. 4 (d)).

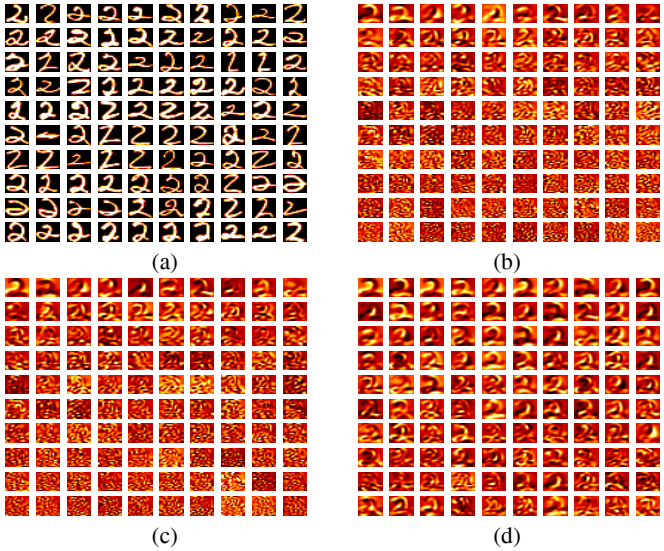



**Fig. 4**. USPS hand-written digit data, '2', is shown in (a). The rest are corresponding principal components estimated by: (b) SVD; (c) COPAL; (d) batch PAST with deflation. The eigen-digits estimated by COPAL is exactly same as ones found by SVD. On the other hand, first 10-20 eigen-digits computed by the deflation method are same as true eigen-digits, but eigen-digits are deteriorated as $n$ increases.

## 5. CONCLUSION

We have introduced a new error measure, IRE, for PCA, whose minimization produced the exact eigenvectors (without rotational ambiguity) of the data covariance matrix. We have presented two algorithms, COPA and COPAL which iteratively minimized the IRE, through the projection approximation, to determine the principal directions of a set of observed data. We have also presented regularized PCA algorithms, R-COPA and R-COPAL and have shown their useful behavior for the case where $n$ is greater than the rank of the data covariance matrix. The validity of our proposed algorithms was demonstrated through numerical experiments. The integrated error was also investigated in [14, 15] in the context of a generative model, leading to EM algorithms where principal directions were estimated through alternating two steps (E and M steps). In contrast, proposed algorithms in this paper, COPA and COPAL, need not go through two steps, which is a major advantage over EM type algorithms.

**Acknowledgments:** This work was supported by KISTEP International Cooperative Research Program, ITEP Brain Neuroinformatics Program, and Korea MIC under ITRC support program supervised by the IITA (IITA-2005-C1090-0501-0018).

## 6. REFERENCES

[1] I. T. Jolliffe, *Principal Component Analysis*, Spinger-Verlag, 2 edition, 2002.

[2] E. Oja, "Neural networks, principal component analysis, and subspaces," *International Journal of Neural Systems*, vol. 1, pp. 61–68, 1989.

[3] P. Baldi and K. Hornik, "Neural networks for principal component analysis: Learning from examples without local minima," *Neural Networks*, vol. 2, pp. 53–58, 1989.

[4] T. D. Sanger, "Optimal unsupervised learning in a single-layer linear feedforward neural network," *Neural Networks*, vol. 2, no. 6, pp. 459–473, 1989.

[5] R. W. Brockett, "Dynamical systems that sort lists, diagonalize matrices, and solve linear programming problems," *Linear Algebra and Applications*, vol. 146, pp. 79–91, 1991.

[6] A. Cichocki and R. Unbehauen, "Neural networks for computing eigenvalues and eigenvectors," *Biological Cybernetics*, vol. 68, pp. 155–164, 1992.

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

[8] Y. Hua, Y. Xiang, T. Chen, K. Abed-Meraim, and Y. Miao, "A new look at the power method for fast subspace tracking," *Digital Signal Processing*, vol. 9, pp. 297–314, 1999.

[9] B. Yang, "Projection approximation subsapce tracking," *IEEE Trans. Signal Processing*, vol. 43, no. 1, pp. 95–107, 1995.

[10] K. Abed-Meraim, A. Chkeif, and Y. Hua, "Fast orthonormal PAST algorithm," *IEEE Signal Processing Letters*, vol. 7, no. 3, pp. 60–62, 2000.

[11] M. E. Tipping and C. M. Bishop, "Probabilistic principal component analysis," *Journal of the Royal Statistical Society B*, vol. 61, no. 3, pp. 611–622, 1999.

[12] S. T. Roweis, "EM algorithms for PCA and SPCA," in *Advances in Neural Information Processing Systems*. 1998, vol. 10, pp. 626–632, MIT press.

[13] S. Choi, "On variations of power iteration," in *Proc. Int'l Conf. Artificial Neural Networks*, Warsaw, Poland, 2005, vol. 2, pp. 145–150, Springer.

[14] J. -H. Ahn and J. -H. Oh, "A constrained EM algorithm for principal component analysis," *Neural Computation*, vol. 15, no. 1, pp. 57–65, 2003.

[15] J. -H. Ahn, S. Choi, and J. -H. Oh, "A new way of PCA: Integrated-squared-error and EM algorithms," in *Proc. IEEE Int'l Conf. Acoustics, Speech, and Signal Processing*, Montreal, Canada, 2004.