A Coupled Helmholtz Machine for PCA

Seungjin Choi [1]

Department of Computer Science
Pohang University of Science and Technology
San 31 Hyoja-dong, Nam-gu
Pohang 790-784, Korea
seungjin@postech.ac.kr

August 3, 2006

Abstract

In this letter we present a coupled Helmholtz machine for principal component analysis (PCA), where sub-machines are related through sharing some latent variables and associated weights. Then, we present a wake-sleep PCA algorithm for training the coupled Helmholtz machine, showing that the algorithm iteratively determines principal eigenvectors of a data covariance matrix without any rotational ambiguity, in contrast to some existing methods that performs factor analysis or principal subspace analysis. The coupled Helmholtz machine provides a unified view of principal component analysis, including various existing algorithms as its special cases. The validity of the wake-sleep PCA algorithm is confirmed by numerical experiments.

Indexing terms: Dimensionality reduction, Helmholtz machines, PCA.

to appear in

Electronics Letters

[1] Please address correspondence to Prof. Seungjin Choi, Department of Computer Science, POSTECH, San 31 Hyoja-dong, Nam-gu, Pohang 790-784, Korea, Tel: +82-54-279-2259, Fax: +82-54-279-2299, Email: seungjin@postech.ac.kr

1 Introduction

Spectral decomposition of a symmetric matrix involves determining eigenvectors of the matrix, which plays an important role in various methods of machine learning and signal processing. For instance, principal component analysis (PCA) or kernel PCA requires the calculation of first few principal eigenvectors of a data covariance matrix or a kernel matrix, respectively. A variety of methods have been developed for PCA (see [4] and references therein). A common derivation of PCA, is illustrated in terms of a linear (orthogonal) projection $\boldsymbol{W} \in \mathbb{R}^{m \times n}$ such that given a centered data matrix $\boldsymbol{X} \in \mathbb{R}^{m \times N}$, the reconstruction error $\|\boldsymbol{X} - \boldsymbol{W}\boldsymbol{W}^\top \boldsymbol{X}\|_F^2$ is minimized, where $\|\cdot\|_F$ denotes the Frobenius norm. 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 $\boldsymbol{U}_1 \in \mathbb{R}^{m \times n}$ contains n principal eigenvectors.

The Helmholtz machine [3] is a statistical inference engine where a recognition model is used to infer a probability distribution over the underlying causes from the sensory input and a generative model is used to train the recognition model. The wake-sleep learning is a way of training the Helmholtz machine and the delta-rule wake-sleep learning was used for factor analysis [5]. In this letter, we present a coupled Helmholtz machine for PCA, where sub-Helmholtz machines are related through sharing some latent variables as well as associated weights. We develop a wake-sleep PCA algorithm for training the coupled Helmholtz machine, showing that the algorithm iteratively determines principal eigenvectors of a data covariance matrix without any rotational ambiguity, in contrast to some existing methods that performs factor analysis or principal subspace analysis. In addition, we show that the coupled Helmholtz machine includes [1, 2] as its special cases.

2 Proposed Method

2.1 Coupled Helmholtz Machine

We denote a centered data matrix by $\boldsymbol{X} = [X_{it}] \in \mathbb{R}^{m \times N}$ and the latent variable matrix by $\boldsymbol{Y} \in \mathbb{R}^{n \times N}$ ($n \leq m$ represents the intrinsic dimension).

The coupled Helmholtz machine that is proposed in this letter, is described by a set of generative models and recognition models, where a set of n generative models has the form

$$\boldsymbol{X} = \boldsymbol{A}\boldsymbol{E}_i\boldsymbol{Y}, \quad i = 1, \ldots, n, \tag{1}$$

where $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ is the generative weigh matrix and $\boldsymbol{E}_i \in \mathbb{R}^{n\times n}$ is a diagonal matrix, defined by

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

The recognition model infers latent variables by

$$\boldsymbol{Y} = \boldsymbol{W}^\top \boldsymbol{X}, \tag{2}$$

where $\boldsymbol{W} \in \mathbb{R}^{m\times n}$ is the recognition weight matrix.

The set of generative models shares some latent variables Y_{it} as well as associated generative weights A_{ij}, in such a way that the 2nd sub-model shares Y_{1t} with the 1st sub-model and the 3rd sub-model shares Y_{1t} and Y_{2t} with the 2nd sub-model as well as Y_{1t} with the 1st submodel, and so on.

2.2 Wake-Sleep PCA Algorithm

The objective function that we consider here, is the *integrated squared error* that has the form

$$\mathcal{J} = \sum_{i=1}^{n} \alpha_i \|\boldsymbol{X} - \boldsymbol{A}\boldsymbol{E}_i\boldsymbol{W}^\top\boldsymbol{X}\|_F^2, \tag{3}$$

where $\alpha_i > 0$ are positive coefficients.

We apply the alternating minimization to derive updating rules for $\boldsymbol{A}$ and $\boldsymbol{W}$ that iteratively minimize (3). In the sleep phase, we fix $\boldsymbol{A}$ and solve $\frac{\partial \mathcal{J}}{\partial \boldsymbol{W}} = 0$ for $\boldsymbol{W}$, leading to

$$\boldsymbol{W} = \boldsymbol{A}\left[\mathsf{U}\left(\boldsymbol{A}^\top\boldsymbol{A}\right)\right]^{-1}, \tag{4}$$

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

$$\mathsf{U}(Z_{ij}) = \begin{cases} Z_{ij}\frac{\sum_{l=i}^{n}\alpha_l}{\sum_{l=j}^{n}\alpha_l} & \text{if } i > j, \\ Z_{ij} & \text{if } i \le j. \end{cases} \tag{5}$$

Next, in the wake phase, we fix $\boldsymbol{W}$ and solve $\frac{\partial \mathcal{J}}{\partial \boldsymbol{A}} = 0$ for $\boldsymbol{A}$, leading to

$$\boldsymbol{A} = \boldsymbol{X}\boldsymbol{Y}^\top \left[\mathsf{U}\left(\boldsymbol{Y}\boldsymbol{Y}^\top\right)\right]^{-1}. \tag{6}$$

The updating algorithms in (4) and (6) are referred to as *wake-sleep PCA* (WS-PCA), that is summarized below. As in [1], we can also consider the limiting case where $\frac{\alpha_{i+1}}{\alpha_i} \to 0$ for $i = 1, \ldots, n-1$, that is, weighting α_i's are rapidly diminishing as i increases. In such a case, the operator $\mathsf{U}(\cdot)$ becomes the conventional upper-triangularization operator $\mathsf{U}_T(\cdot)$ where $\mathsf{U}_T(Z_{ij}) = 0$ for $i > j$ and $\mathsf{U}_T(Z_{ij}) = Z_{ij}$ for $i \leq j$. The resulting algorithm is referred to as WS-PCA (limiting case).

Algorithm Outline: WS-PCA

Sleep phase

$$\begin{aligned}\boldsymbol{W} &= \boldsymbol{A}\left[\mathsf{U}\left(\boldsymbol{A}^\top\boldsymbol{A}\right)\right]^{-1}, \\ \boldsymbol{Y} &= \boldsymbol{W}^\top\boldsymbol{X}.\end{aligned}$$

Wake phase

$$\boldsymbol{A} = \boldsymbol{X}\boldsymbol{Y}^\top \left[\mathsf{U}\left(\boldsymbol{Y}\boldsymbol{Y}^\top\right)\right]^{-1}.$$

---

3 Numerical Experiments

We provide a numerical example with $\boldsymbol{X} \in \mathbb{R}^{10\times 1000}$ (intrinsic dimension $n = 5$), in order to demonstrate that the WS-PCA indeed finds the exact eigenvectors of $\boldsymbol{X}\boldsymbol{X}^\top$ without rotational ambiguity. Fig 1 shows the convergence behavior of the WS-PCA algorithm (and its limiting case) with different choice of α_i. Regardless of values of α_i, generative weights (or recognition weights) converge to true eigenvectors. However, the convergence behavior of the WS-PCA algorithm is slightly different, especially according to the ratio $\frac{\alpha_{i+1}}{\alpha_i}$ for $i = 1, \ldots, n-1$ (see Fig. 1). The WS-PCA achieves the faster convergence, as the ratio, $\frac{\alpha_{i+1}}{\alpha_i}$ for $i = 1, \ldots, n-1$ decreases. In fact, the limiting case of WS-PCA (where U_T is used instead of U) shows the fastest convergence (see Fig. 1).

4 Conclusion

We have introduced a coupled Helmholtz machine where latent variables as well as associated weights are shared by a set of Helmholtz machines. We have presented a wake-sleep-like algorithm in the framework of the coupled Helmholtz machine, showing that the algorithm indeed determines the exact principal eigenvectors of a data covariance matrix without rotational ambiguity. The WS-PCA algorithm includes [1, 2] as its special case, each of which is the generative-only and the recognition-only counterpart, respectively. The WS-PCA algorithm is useful in applications where only first few principal eigenvectors are required to be computed from the high-dimensional data covariance matrix.

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

References

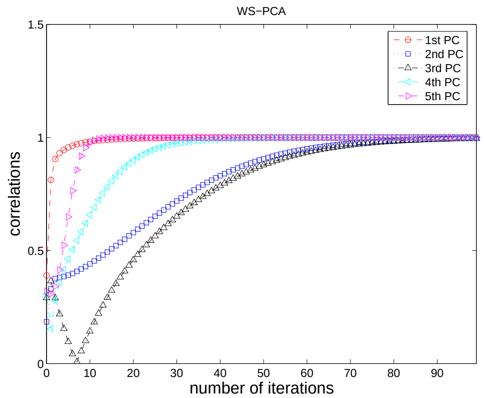
[1] J. -H. Ahn, S. Choi, and J. -H. Oh. A new way of PCA: Integrated-squared-error and EM algorithms. In *Proc. IEEE Int'l Conf. Acoustics, Speech, and Signal Processing*, Montreal, Canada, 2004.

[2] S. Choi. On variations of power iteration. In *Proc. Int'l Conf. Artificial Neural Networks*, volume 2, pages 145–150, Warsaw, Poland, 2005. Springer.

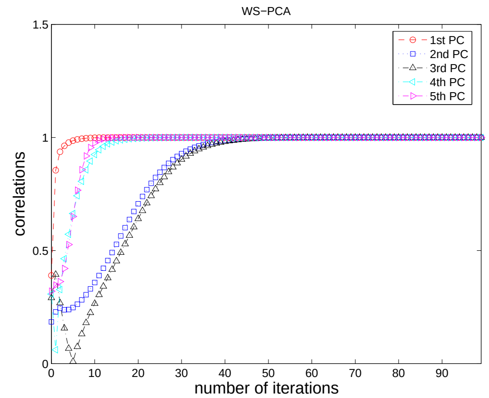
[3] P. Dayan, G. E. Hinton, R. M. Neal, and R. S. Zemel. The Helmholtz machine. *Neural Computation*, 7:1022–1037, 1995.

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

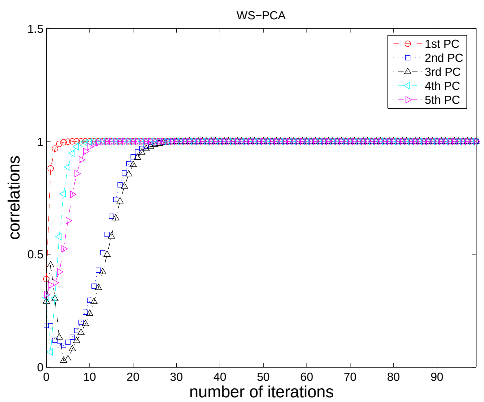
[5] R. M. Neal and P. Dayan. Factor analysis using delta-rule wake-sleep learning. *Neural Computation*, 9:1781–1803, 1997.



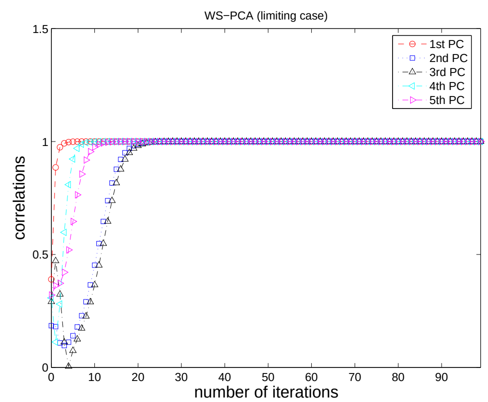
(a)



(b)



(c)



(d)

Figure 1: Evolution of generative weight vectors, is shown in terms of the absolute value of the inner produce between a weight vector and a true eigenvector (computed by SVD): (a) WS-PCA with $\frac{\alpha_{i+1}}{\alpha_i} = 1$ and $\alpha_1 = 1$; (b) WS-PCA with $\frac{\alpha_{i+1}}{\alpha_i} = 0.5$ and $\alpha_1 = 1$; (c) WS-PCA with $\frac{\alpha_{i+1}}{\alpha_i} = 0.1$ and $\alpha_1 = 1$; (d) WS-PCA (limiting case)