30 Nonlinear PCA Networks and Optimization Criteria

Juha Karhunen, Erkki Oja, Petteri Pajunen, Liuyue Wang, and Jyrki Joutsensalo

In this work, which was our earlier principal research topic in 1994 and before that, we have developed several different relatively simple nonlinear and robust generalizations of neural PCA methods.
A common rigorous approach to these developments is to derive new unsupervised neural learning algorithms by considering generalizations of the optimization criteria leading to the standard PCA solution. There exist several different optimization problems which lead to a standard PCA solution. These include:

1. Maximization of linearly transformed variances $\mathrm{E}\{[\mathbf{w}(i)^T\mathbf{x}]^2\}$ or outputs of a linear network under orthonormality constraints ($\mathbf{W}^T\mathbf{W} = \mathbf{I}$). Here $\mathbf{x}$ is the input (data) vector, $\mathbf{w}(i)$ is the weight vector of i-th neuron, and $\mathbf{W} = \mathbf{w}(1), \dots, \mathbf{w}(M)$ is the weight matrix of a PCA network.

2. Minimization of the mean-square representation error $\mathrm{E}\{\| \mathbf{x} - \hat{\mathbf{x}} \|^2\}$, when the input data $\mathbf{x}$ are approximated using a lower dimensional linear subspace $\hat{\mathbf{x}} = \mathbf{W}\mathbf{W}^T\mathbf{x}$.

3. Uncorrelatedness of outputs $\mathbf{w}(i)^T\mathbf{x}$ of different neurons after an orthonormal transform ($\mathbf{W}^T\mathbf{W} = \mathbf{I}$).

4. Minimization of representation entropy.

In [2,3], we have derived a number of robust and nonlinear PCA learning algorithms from these generalized criteria for both symmetric and hierarchic network structures, and shown their relationships to existing neural PCA algorithms. In particular, generalization of the first variance maximization criterion leads for symmetric orthonormality constraint to the so-called Robust PCA rule:

$$\mathbf{W}_{k+1} = \mathbf{W}_k + \mu_k[\mathbf{I} - \mathbf{W}_k\mathbf{W}_k^T]\mathbf{x}_k\mathbf{g}(\mathbf{x}_k^T\mathbf{W}_k). \tag{81}$$

Here and later on the nonlinear odd function $g(t)$ is applied separately to each component of its argument vector. The index k denotes iteration or sample number, and μ_k is the learning parameter at iteration k. This rule has been shown to be useful in clustering, projection pursuit, and robust PCA. It is often useful to preprocess the data vectors $\mathbf{x}_k$ by whitening (sphering) them. After this, the learning rule (81) responds directly to higher-order statistics in the data.
Similarly, generalization of the second optimization problem, minimization of the mean-square representation error, leads to so-called Nonlinear PCA rule:

$$\mathbf{W}_{k+1} = \mathbf{W}_k + \mu_k[\mathbf{x}_k - \mathbf{W}_k\mathbf{g}(\mathbf{y}_k)]\mathbf{g}(\mathbf{y}_k^T), \tag{82}$$

where the output vector $\mathbf{y}_k = \mathbf{W}_k^T\mathbf{x}_k$. We have shown in several papers, summarized in [4], that with prewhitening the learning algorithm (82) can be successfully applied to blind separation of certain type source signals. The blind separation problem is discussed in several other sections of this report. The nonlinear PCA rule provides an especially simple neural solution to this difficult problem. This has been analyzed rigorously in [4,10].

The learning rules (81) and (82) were proposed on intuitive grounds already in [7]. Later on, their relationship to optimization problems were made rigorous in the theoretical papers [2,3]. We have also developed a number of other algorithms using this optimization based approach to nonlinear PCA; see [2,3,6]. In particular, the so-called bigradient algorithm developed and analyzed in [6] provides a versatile tool. In various forms, it can be applied both to robust PCA problems, making the results insensitive to outliers in the data and inpulsive noise, as well as to blind source separation.

We have also developed fast converging approximative least-squares algorithms [5] for minimizing so-called nonlinear PCA criterion given by

$$J(\mathbf{W}) = \| \mathbf{x} - \mathbf{W}\mathbf{g}(\mathbf{x}^T\mathbf{W}) \|^2 \tag{83}$$

These least-squares algorithms can again be applied to blind separation of sources after prewhitening of the input data [5]. - The same criterion (83) is used as a starting point in deriving the Nonlinear PCA rule (82), too.

Recently, we have derived new results on the nonlinear PCA criterion (83) in blind source separation and related problems [8,9]. The criterion can be expressed for prewhitened data in a simple form. This allows an easy comparison with other criteria used in blind signal processing and independent component analysis, including cumulants, Bussgang criteria, and information theoretic contrast functions. The results show the close connection of the nonlinear PCA learning rule (82) with certain well-known other algorithms used for blind source separation, and help in the optimal choice of the nonlinearity [8,9].

Still other theoretical results include stability considerations of these algorithms. In [1], a rigorous stability condition has been derived for PCA subspace rule, and the stability of the robust algorithm (81) is shown to be better if the the nonlinear function $g(t)$ grows less than linearly.

References

[1] J. Karhunen. Stability of Oja's PCA subspace rule. *Neural Computation*, 6:739–747, 1994.

[2] J. Karhunen and J. Joutsensalo. Representation and separation of signals using nonlinear PCA type learning. *Neural Networks*, 7(1):113–127, 1994.

[3] J. Karhunen and J. Joutsensalo. Generalizations of principal component analysis, optimization problems, and neural networks. *Neural Networks*, 8(4):549–562, 1995.

[4] J. Karhunen, E. Oja, L. Wang, R. Vigario, and J. Joutsensalo. A class of neural networks for independent component analysis. *IEEE Trans. on Neural Networks*, 8:486–514, 1997.

[5] P. Pajunen and J. Karhunen. Least-squares methods for blind source separation based on nonlinear PCA. *Int. J. of Neural Systems*, 8(5-6):601–612, 1997.

[6] L. Wang and J. Karhunen. A unified neural bigradient algorithm for robust PCA and MCA. *Int. J. of Neural Systems*, 7(1):53–67, 1996.

[7] E. Oja, H. Ogawa, and J. Wangviwattana. Learning in nonlinear constrained Hebbian networks. In T. Kohonen et al. (Eds.), *Artificial Neural Networks* (Proc. ICANN'91, Espoo, Finland, June 1991). North-Holland, Amsterdam, 1991, pp. 385-390.

[8] J. Karhunen, P. Pajunen, and E. Oja.. The nonlinear PCA criterion in blind source separation: relations with other approaches. *Neurocomputing*, 22:5–20, 1998.

[9] E. Oja. Nonlinear PCA criterion and maximum likelihood in independent component analysis. In *Proc. First Int. Workshop on Independent Component Analysis and Signal Separation (ICA'99).* Aussois, France, January 1999, pp. 143-148.

[10] E. Oja. The nonlinear PCA learning rule in independent component analysis.. *Neurocomputing*, 17:25–45, 1997.