ROBUST PCA NEURAL NETWORKS FOR RANDOM NOISE REDUCTION OF THE DATA

Stanislaw Osowski, Andrzej Majkowski and Andrzej Cichocki[1]

Warsaw University of Technology, Warsaw, POLAND, sto@iem.pw.edu.pl

[1] also FRP RIKEN, Lab of Artificial Brain Systems, Tokyo, JAPAN, cia@hare.riken.go.jp

ABSTRACT

The paper presents principal component analysis (PCA) approach to the reduction of noise contaminating the data. The PCA performs the role of lossy compression and decompression. The compression/decompression provides the means of coding the data and then recovering it with some losses, dependent on the realized compression ratio. In this process some part of information contained in the data is lost. When the loss tolerance is equal to the noise strength, the noise and the loss tolerance are augmented and the decompressed signal is deprived of noise. This way of noise filtering has been checked on the examples of 1-dimensional and 2-dimensional data and the results of numerical experiments have been included in the paper.

1. INTRODUCTION

The elimination of the noise is an important subject since practical digital images are often degraded in some manner to some extend and need to be restored to improve their quality. The objective of removing the noise is to obtain the recovered image in such a way that it resembles the original (noiseless) image as closely as possible.

There are many different filtering algorithms for noise removal, following from the Wiener or Kalman filter theory [10, 11]. However to get good results of filtering using these methods we have to know in advance the spectral properties of the noise free data and the noise itself.

The paper will present the random noise reduction technique which does not need to know in advance the spectral properties of the data. It is based on the application of the compression and decompression (reconstruction) of the noisy data. The lossy compression/decompression technique provides the means of coding the data and then recovering it with some losses, dependent on the realized compression ratio. Some part of information contained in the data is then lost. This principle will be applied in the paper to reduce the random noise, distorting the data. When the loss tolerance is equal to the noise strength, the noise and the loss tolerance are augmented and the decompressed signal is deprived of noise. In this way the compression/decompression technique provides the filtering of the data.

This method of noise elimination has been checked on the examples of 1-dimensional and 2-dimensional data and some chosen results of numerical experiments are included in the paper. The important advantage of the approch is its universality. The neural network trained on the example of one data set can be applied to noise removal of the other sets of data of similar distortion. Thanks to the generalization ability of the neural network the quality of such operation is reasonably good.

2. PRINCIPLE OF THE RANDOM NOISE FILTERING USING COMPRESSION AND DECOMPRESSION

Let $\mathbf{f}_n$ and $\mathbf{r}_n$ denote the sample sequences composed of n independent samples of the image function f and random variable r, respectively. These sequences can be mathematically presented as

$$\mathbf{f} = [f(0),\ f(1),\ \cdots,\ f(n-1)]^T$$
$$\mathbf{r} = [r(0),\ r(1),\ \cdots,\ r(n-1)]^T$$

Let $\hat{\mathbf{f}}_n = \mathbf{f}_n + \mathbf{r}_n$ denotes the sequence of data $\mathbf{f}_n$ corrupted with noise, characterized by the vector $\mathbf{r}_n$. Let us introduce the notion of strength of the noise, understood here as the norm $\| \mathbf{r}_n \|$, where different kinds of norm may be applied in this definition, e.g., euclidean, L_∞ or L_1 norms. To estimate the noise strength we have followed the heuristic method of Natarajan [5], where many runs of compression/decompression algorithms of various compression losses, measured as Peak - Signal -to- Noise - Ratio (PSNR) have been tried. After all runs one has to plot the compression ratio K_r versus PSNR values to obtain the rate - distortion characteristic of the signal. If we create the plot for the noisy signal we notice that for the allowable losses higher than the noise strength, the plot follows the rate - distortion characteristic for the noise free signal. On the other hand at the allowable losses less than noise strength, the plot follows the rate - distortion characteristic of the noise itself. It means that the noise dominates in this region. At the point of PSNR corresponding to the strength of the noise the plot of the noisy signal shows the "knee point", that is a point at which the slope of the curve changes rapidly. The precise determination of the knee point can be obtained by drawing the second derivative characteristic. The point of PSNR at which the second derivative attains its maximum is the measure of the noise strength.

The elimination of noise is achieved in the algorithm through the lossy compression and then decompression of the noisy signal $\hat{\mathbf{f}}$. At high compression ratio K_r the noise introduced by coding/decoding plus additional random noise contaminating the data is relatively high. At small compression ratio K_r both signal and noise are passing throught the filter almost unchanged and no effect of filtering is observed. To obtain good results of noise elimination we have to find some compromise point, at which the attenuation of noise is high enough and coding/decoding error is on the acceptable level. This is the breaking point of the rate - distortion characteristic, corresponding to the PSNR value equal to the strength value of the noise, corrupting the data. Knowing the value of the noise strength it is enough to adjust the compression ratio K_r corresponding to this point.

3. NEURAL PCA TECHNIQUE OF COMPRESSION

The practical solution of filtering the random noise has been obtained by the authors through the use of PCA neural network [1, 2, 4], applied here as the coding/decoding means. The PCA is the statistical method defining the linear transformation $\mathbf{y} = \mathbf{W}\mathbf{x}$, transforming the stationary stochastic data $\mathbf{x} \in R^N$ into the vector $\mathbf{y} \in R^K$ using the matrix $\mathbf{W} \in R^{K\times N}$ at $K \ll N$ in such a way that the output space $\mathbf{y}$ of the reduced dimension preserves the most important information of the input space $\mathbf{x}$. Denote by $\mathbf{R}_{xx}$ the expected value of the correlation matrix $\mathbf{R}_{xx} = E[\mathbf{x}\mathbf{x}^T]$ of the input vectors $\mathbf{x}$ and by λ_i the ith eigenvalue of this matrix. The eigenvector $\mathbf{w}_i$ corresponding to λ_i fulfills the relation

$$\mathbf{R}_{xx}\mathbf{w}_i = \lambda_i \mathbf{w}_i$$

The eigenvectors corresponding to all eigenvalues are orthogonal to each other. If we arrange the eigenvalues in the decresing order and define the matrix $\mathbf{W}$ in the form

$$\mathbf{W} = [\mathbf{w}_1 \;\; \mathbf{w}_2 \;\; \cdots \;\; \mathbf{w}_K]$$

where each column vector $\mathbf{w}_i = [w_{i1} \;\; w_{i2} \;\; \cdots \;\; w_{iN}]^T$, then the aim of PCA transformation is the iterative determination of the principal eigenvectors $\mathbf{w}_i$ (i = 1, 2, ..., K) of the correlation matrix in such a way that the expected value $E\left(\| \mathbf{w}_i^T\mathbf{x} \|^2\right)$ is maximized. The determination of the principal vectors is performed only once. The reproduction of the data, $\hat{\mathbf{x}}$, may be then done to different accuracy taking into account different number of principal vectors forming the matrix $\mathbf{W}$. This can be achieved by using relation $\hat{\mathbf{x}} = \mathbf{W}^T\mathbf{y}$. The higher the number of principal vectors the better accuracy of reproduction, higher value of PSNR and lower compression ratio. So generaly we may summarize, that the idea of PCA is to convey the most information about a set of a data given a limited number of linear descriptors. High dimensional data is projected onto a smaller number of dimensions, adjusted in a way to enable the recovering of the data to some limited accuracy.

There are different methods of estimation of principal components [1, 2, 3, 4, 8]. In practice we have implemented the extended or modified Oja's learning rule [4, 2], according to which the results are obtained iteratively, where k-th iteration for eigenvectors is described by

$$\mathbf{w}_i(k+1) = \mathbf{w}_i(k) + \eta_i(k) y_i(k) \Psi[\mathbf{e}_i(k)], \quad (1)$$

In this relation the index i means the ith neuron and k – the iteration. The variable y_i is the output signal of the neuron, η_i – the learning rate and Ψ – the activation function. The recursive relations describing the variables are given in the form

$$\mathbf{e}_i \triangleq \mathbf{e}_{i-1} - \mathbf{w}_i y_i$$

$$y_i \triangleq \mathbf{w}_i^T \mathbf{e}_{i-1}$$

$$\mathbf{e}_0(k) \triangleq \mathbf{x}(k)$$

for i = 1, 2, ..., K. The vector $\Psi(\mathbf{e})$ represents the vector of suitable activation functions

$$\Psi(\mathbf{e}_i) = [\Psi_i(e_{i1}), \Psi_i(e_{i2}), ..., \Psi_i(e_{in})]^T$$

The most often used forms of activation functions are $tanh$ $\Psi_i(e_{ij}) = tanh(e_{ij}/\beta)$ or the polynomial $\Psi_i(e_{ij}) = e_{ij}^{2p+1}$ for robust nonlinear PCA and linear form $\Psi_i(e_{ij}) = e_{ij}$ for the standard linear PCA (which is optimal only for Gaussian distribution of noise).

The choice of activation function depends on character of the noise, i.e., its statistical distribution. In the special case $\Psi(e) = e$ the learning rule (1) is equivalent to the well known Sanger's algorithm [1,2] but it is much more numerically stable since the error accumulation is not as significant as in the Sanger rule.

The presented above learning rule is local assuming that the principal vectors are updated sequentially one after one, starting from the first (biggest one). The local character of the algorithm is a great advantage, since all calculations are done without inverting the whole system of linear equations and at the time of calculation only small part of information needs to be stored.

In the learning algorithm the key role plays the learning rate $\eta_i(k) \geq 0$. If the learning rate is too large, the algorithm is numerically unstable. Otherwise if it is fixed or exponentially decreasing parameter, the convergence speed of the algorithm may be very slow. Applying RLS (Recursive Least Squares) technique, it can be shown that the learning rate can be updated as follows

$$\eta_i(k+1) = [\frac{\lambda}{\eta_i(k)} + [y_i(k)]^2]^{-1} \quad (2)$$

with

$$\eta_i(0) = [\sigma^2(\mathbf{e}_{i-1})]^{-1} \quad (3)$$

where $\sigma^2(\mathbf{e}_{i-1})$ is the variance of the input signals), and with $0.9 < \lambda \leq 1$ representing the forgetting factor. This form of compression/decompression has been applied in practical solution of the noise filtering problem of the data.

4. RESULTS OF NUMERICAL EXPERIMENTS

The experiments checking the filtering ability of the proposed approach, have been carried out using both 1-D (the curve) and 2-D input data (the image). The data have been splitted into equal frames, forming the 1-D vectors. The neural network forming PCA has been trained using learning data corrupted with the random noise of either uniform or normal distribution of certain strength. Different levels of noise have been checked and the results of filtering have been assessed as good. Fig. 1 presents the results of filtering the 1-D data corrupted by the random noise of normal

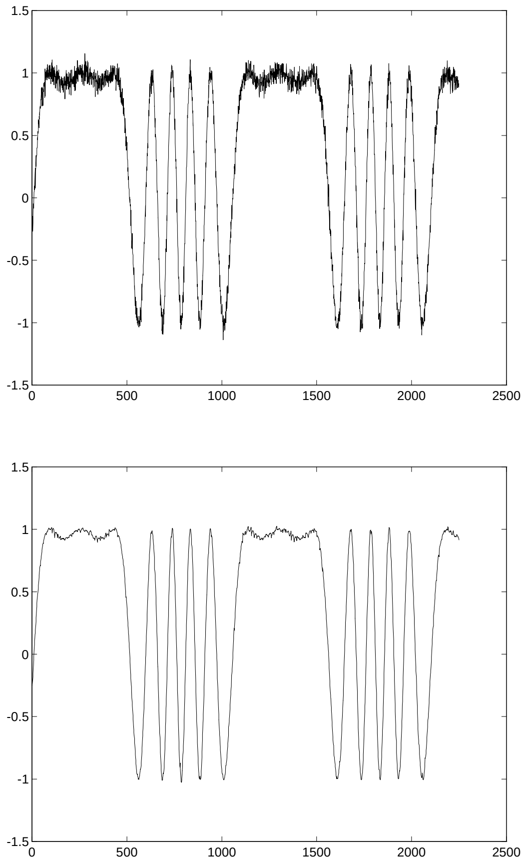

Figure 1. *The results of PCA application to the elimination of the noise from the learned image; upper - the curve corrupted with the noise, lower - the regenerated curve*

distribution of the SNR value equal 20dB (Fig. 1a). The original curve was described by the relation

$$f(n) = sin(300n + 6cos(60n))$$

It was plotted for the time t changing from $t = 0$ to $t = 0.22$ with the resolution of 0.0001. The length of the frame assumed in the experiments was equal 15. The neural network has been trained using modified Oja learning rule (1) with the recursively adjusted learning coefficients given by (2). After training the rate-distortion curve of the network has been plotted and on the basis of this the optimal compression ratio, corresponding to 3 principal components have been chosen. Fig. 1b presents the filtered curve, that is the curve after the compression and decompression using PCA with 3 principal components. As it is seen from the results the quality of noise elimination is reasonably good. The noise has been greatly supressed and the recovered curve only slightly differs from the original one. Fig. 2 illustrates the noise removal ability of the neural network. Fig. 2a presents the difference of the original and the distorted curves while Fig. 2b shows the difference between the original and recovered curves. After the noise removal the standard deviation of the error data has been reduced from the value of 0.0481 of the original noise to 0.0144 of the recovered data.

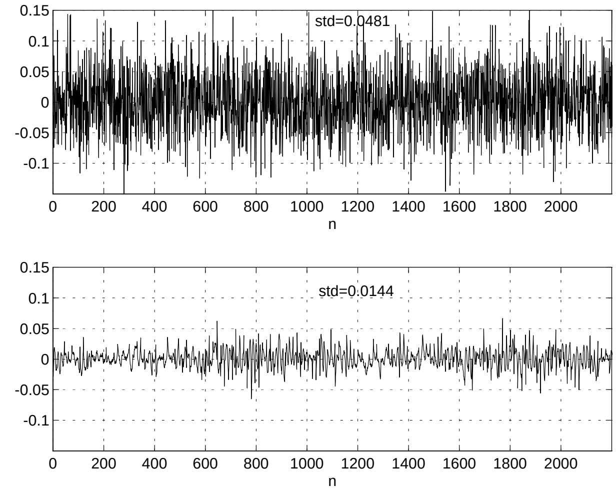

Figure 2. *The error of reproduction of the distorted 1-D data; the upper curve - the state on the input to the network, the lower curve - the final state after noise reduction*

Fig. 3a presents the 2-D image of the forest corrupted in 10% by the additive random noise of uniform distribution. Adjusting the proper compression ratio K_r, equal in the experiments $K_r \approx 6$, corresponding to 10 principal components, we got the results in the form of filtered image presented in Fig. 3b. Fig. 4 presents the results of generalization of the learned neural network. The network trained on the data of Fig. 3 has been used to eliminate noise corrupting the image of Lena of Fig. 4a. The resulting filtered image is shown in Fig. 4b. As it is seen the quality of the regenerated picture is good. It should be stressed that the presented filtered image has not been taking part in learning process, it is the result of the generalization ability of the trained neural network.

5. CONCLUSIONS

The paper has proposed the PCA based approach to the elimination of the noise, corrupting the data. According to the authors the main achievements of the work are:

- new coding/decoding strategy based on robust PCA neural network in application to noise removal
- the applicability of method to the image filtering without additional knowledge of the statistics of the distorting random noise
- the implementation of the developed technique to the reduction of random noise from the 1-D signals or 2-D images.

This method has been succesfully checked on many examples of 1-D and 2-D data proving its good performance on the images not taking part in learning process, for which the generalization ability of the neural network has been exploited. This is the important advantage of the approach, since it makes the method universal, applicable without retraining of the network to different data sets. The results of numerical experiments have confirmed the ability of the

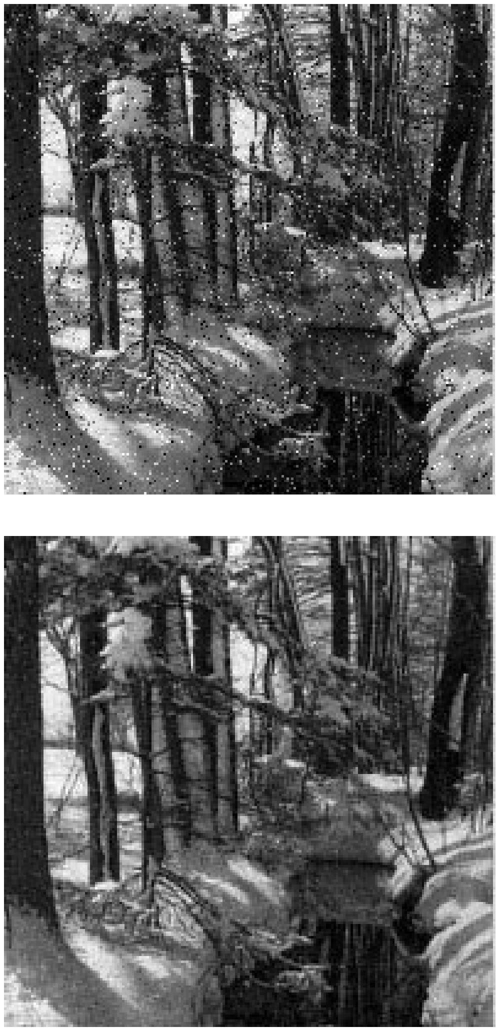

Figure 3. *The ability of the PCA neural network to random noise removal from the image used in learning: upper - picture corrupted with the noise, lower - the regenerated picture*

Figure 4. *The generalization ability of the PCA neural network - the results of elimination of the noise from the unlearned image; upper - picture corrupted with the noise, lower - the regenerated picture*

proposed procedure for removal the random noise from the corrupted data of different strength of corruption.

REFERENCES

[1] S. Haykin, Neural networks, a comprehensive foundation, Macmillan, 1994.

[2] A. Cichocki, R. Unbehauen, Neural networks for optimization and signal processing, Wiley, 1994 (second edition).

[3] G. Golub, C. Van Loan, Matrix computations, Academic Press, 1991, NY

[4] E. Oja, L. Wang , Image compression by MLP and PCA neural networks, SCIA-93, Tromso, 1993, pp. 1317 - 1324

[5] B. Natarajan, Filtering random noise from deterministic signals via data compression, IEEE Trans. SP, 1995, vol. 43, pp. 2595 - 2605.

[6] Y. Ephraim and H. Van Trees, A signal subspace approach for speech enhancement, IEEE Tr. SAP, vol.3, 251-266, 1995.

[7] S.H. Jensen, C. Hansen and S. Hansen, Reduction of broad-band noise in speech by truncated QVSD, IEEE Tr. SAP, vol.3, 439-448, 1995.

[8] G. Xu, H. Zha, G. Golub and T. Kailath, Fast algorithms for updating signal subspaces, IEEE Trans on CAS, vol.41, 1994, pp.537-549.

[9] N. H. Young, A. H. Lai, Performance evaluation of a feature-preserving filtering algorithm for removing additive random noise in digital images, Optical Engineering, 1996, vol. 35, pp. 1871 - 1885

[10] R. Gonzales, R. E. Wood, Digital image processing, Addison Wesley, Massachusetts, 1992

[11] S. A. Tretter, Introduction to discrete-time signal processing, Wiley, N. Y., 1976