Energy predictability to blind source separation

Z. Shi and C. Zhang

A blind source separation (BSS) method based on the energy (square) predictability of original sources is proposed. The method exploits the nonstationarity of sources in the sense that the variance of each source signal can be assumed to change smoothly against time. In contrast to linear predictability, it is shown that nonlinear predictability can also be used for BSS. Simulations verify the efficient implementation of the proposed method, especially its robustness to the outliers.

Introduction: Blind source separation (BSS) [1, 2] is an increasingly popular data analysis technique which has received wide attention in various fields such as biomedical signal processing and analysis, data mining, speech and image processing. The task of BSS is to recover original sources from their mixtures using some statistical properties of original sources. In this Letter we consider linear, instantaneous, noiseless mixtures of the form $\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t)$, where $\mathbf{x}(t) = (x_1(t), \ldots, x_n(t))^T$ denotes the n-dimensional observation vector, $\mathbf{A}$ is the $n \times n$ unknown non-singular constant mixing matrix, and $\mathbf{s}(t) = (s_1(t), \ldots, s_n(t))^T$ is the n-dimensional vector of unknown zero-mean and unit-variance primary sources.

Several methods for BSS using the statistical properties of primary sources have been proposed, such as non-Gaussianity [1, 2], linear predictability or smoothness [1, 3], coding complexity [4, 5] and nonstationarity of variance [6], etc. In this Letter, we show that nonlinear predictability can also be a method for BSS. We assume that primary sources are mutually independent and sources are nonstationary in the sense that the energies (squares) of the signals are predictable (i.e. the variance of each independent source signal can be assumed to change smoothly against time). Comparison with a BSS algorithm by the nonstationarity of variance [6] is given in the 'Experimental results' Section.

Proposed algorithm: Assume that the measured sensor signals $\mathbf{x}$ have already been followed by an $n \times n$ whitening matrix $\mathbf{V}$ such that the components of $\tilde{\mathbf{x}}(t) = \mathbf{V}\mathbf{x}(t)$ are unit variance and uncorrelated. Furthermore, assume that we want to extract a source signal, for this purpose we design a single processing unit described as $\tilde{y}_i(t) = \mathbf{w}_i^T \tilde{\mathbf{x}}(t)$, where $\mathbf{w}_i = (w_{i1}, \ldots, w_{in})^T$ is the weight vector which corresponds to the estimate of one row of $(\mathbf{VA})^{-1}$, and $\tilde{y}_i(t)$ is the output signal which corresponds to the estimate of the source signal s_i.

We present the following constrained minimisation problem based on the generalised autocorrelation error function of the desired source; for simplicity we use just one predicting term:

$$\begin{aligned} \min_{\|\mathbf{w}_i\|=1} \Psi(\mathbf{w}_i, \alpha_i) &= E\{(G(\tilde{y}_i(t)) - \alpha_i G(\tilde{y}_i(t-\tau)))^2\} \\ &= E\{(G(\mathbf{w}_i^T \tilde{\mathbf{x}}(t)) - \alpha_i G(\mathbf{w}_i^T \tilde{\mathbf{x}}(t-\tau)))^2\} \end{aligned} \quad (1)$$

where τ is a specific time delay, α_i is a predicting parameter and G is a differentiable function which defines the linear or nonlinear autocorrelation of the desired source. Generally, we can choose $G(u) = u$ or $G(u) = u^2$. Assume that $G(u) = u$ is chosen, the objective function is just the mean squared error function used in the blind source extraction algorithm [3] based on the linear predictability. Using the linear predictability, one can perform BSS when sources have the linear temporal autocorrelations. However, we show that the energy (square) predictability can also be a BSS principle when sources have the square temporal autocorrelations.

To perform the optimisation in (1), we can use a simple gradient descent. The gradients of $\Psi(\mathbf{w}_i, \alpha_i)$ with respect to $\mathbf{w}_i$ and α_i are obtained as:

$$\begin{aligned} \frac{\partial \Psi(\mathbf{w}_i, \alpha_i)}{\partial \mathbf{w}_i} &= 2E\{(G(\tilde{y}_i(t)) - \alpha_i G(\tilde{y}_i(t-\tau)))(g(\tilde{y}_i(t))\tilde{\mathbf{x}}(t) \\ &\quad - \alpha_i g(\tilde{y}_i(t-\tau))\tilde{\mathbf{x}}(t-\tau))\} \end{aligned} \quad (2)$$

$$\frac{\partial \Psi(\mathbf{w}_i, \alpha_i)}{\partial \alpha_i} = -2E\{(G(\tilde{y}_i(t)) - \alpha_i G(\tilde{y}_i(t-\tau)))G(\tilde{y}_i(t-\tau))\} \quad (3)$$

where the function g is the derivative of G. Let $\partial \Psi(\mathbf{w}_i, \alpha_i)/\partial \alpha_i = 0$, we have

$$\alpha_i = \frac{E\{G(\tilde{y}_i(t))G(\tilde{y}_i(t-\tau))\}}{E\{G(\tilde{y}_i(t-\tau))G(\tilde{y}_i(t-\tau))\}} \quad (4)$$

Thus, the energy predictability BSS algorithm (EPBSS) is obtained as follows.

Algorithm outline: EPBSS (estimating one source)

1. Centre the data to make its mean zero and whiten the data to give $\tilde{\mathbf{x}}(t)$. Choose random initial values for $\mathbf{w}_i$ and α_i, and a suitable learning rate μ.
2. Update the weight vector by

$$\begin{aligned} \alpha_i(k+1) &= \frac{E\{G(\mathbf{w}_i(k)^T \tilde{\mathbf{x}}(t))G(\mathbf{w}_i(k)^T \tilde{\mathbf{x}}(t-\tau))\}}{E\{G(\mathbf{w}_i(k)^T \tilde{\mathbf{x}}(t-\tau))G(\mathbf{w}_i(k)^T \tilde{\mathbf{x}}(t-\tau))\}} \\ \mathbf{w}_i(k+1) &= \mathbf{w}_i(k) - \mu E\{(G(\mathbf{w}_i(k)^T \tilde{\mathbf{x}}(t)) - \alpha_i(k) \\ &\quad \times G(\mathbf{w}_i(k)^T \tilde{\mathbf{x}}(t-\tau)))(g(\mathbf{w}_i(k)^T \tilde{\mathbf{x}}(t))\tilde{\mathbf{x}}(t) \\ &\quad - \alpha_i(k) g(\mathbf{w}_i(k)^T \tilde{\mathbf{x}}(t-\tau))\tilde{\mathbf{x}}(t-\tau))\} \\ \mathbf{w}_i(k+1) &= \mathbf{w}_i(k+1)/\|\mathbf{w}_i(k+1)\| \end{aligned} \quad (5)$$

where $G(u) = u^2$, $g(u) = 2u$, and k means the iteration number.

3. If not converged, go back to step 2.

To estimate the separating matrix $\mathbf{W} = (\mathbf{w}_1, \ldots, \mathbf{w}_n)^T$, one can simply use a deflation scheme (one-by-one estimation) or the symmetric orthogonalisation [2].

Experimental results: We created ten artificial signals which had smoothly changing variances as follows (with Gaussian marginal distributions, zero linear autocorrelations and square temporal autocorrelations) [6]. First, we created ten signals using a first-order autoregressive model with constant variances of the innovations [4, 5], with 5000 time points. The signals were created with Gaussian innovations and had identical autoregressive coefficients (0.8). All these innovations had constant unit variance. Then, the signs of the signals were completely randomised by multiplying each signal by a binary i.i.d. signal that took the values ± 1 with equal probabilities. The source signals were mixed with 10×10 random matrices and the EPBSS algorithm using the symmetric orthogonalization was used to estimate the separating matrix (the learning rate $\mu = 0.1$ and $\tau = 1$). For comparison, we also ran the cumulant-based fixed-point approach using the nonstationarity of variance (FPNSV) ($\tau = 1$) [6]. To measure the accuracy of separation, we calculated the performance index

$$\mathbf{PI} = \frac{1}{n^2}\left\{\sum_{i=1}^{n}\left(\sum_{j=1}^{n}\frac{|p_{ij}|}{\max_k |p_{ik}|} - 1\right) + \sum_{j=1}^{n}\left(\sum_{i=1}^{n}\frac{|p_{ij}|}{\max_k |p_{kj}|} - 1\right)\right\} \quad (6)$$

where p_{ij} is the ijth element of $n \times n$ matrix $\mathbf{P} = \mathbf{WVA}$.

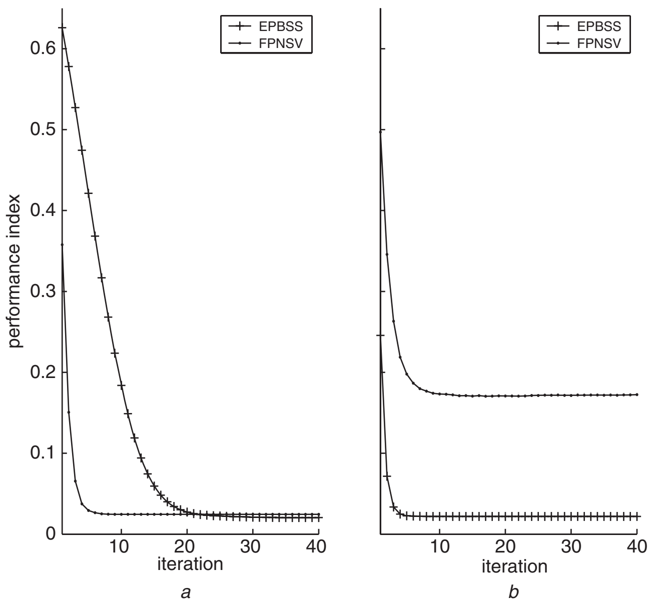

Fig. 1 *Average performance indexes*

a Average performance indexes over 100 independent runs for 10 sources
b Average performance indexes over 100 independent runs for 10 sources when outliers introduced

Fig. 1*a* shows the average performance indexes over 100 independent trials against iteration numbers. The EPBSS algorithm performed similarly to the cumulant-based fixed-point approach [6] in the sense of the separation accuracy. To investigate the robustness of algorithms, we randomly added 30 outliers the values of which were 10 in each source signal. Fig. 1*b* shows the average performance indexes over 100 independent trials against iteration numbers. Obviously, the EPBSS algorithm outperformed the FPNSV algorithm [6] when the outliers were introduced. In fact, the FPNSV algorithm was sensitive to the outliers and failed with the data in the case.

Conclusion: We propose a BSS algorithm based on a new simple principle: the energy predictability. When sources are nonstationary in the sense that the variances of the signals change smoothly, we have demonstrated the efficient implementation of the method and verified its robustness to the outliers.

Acknowledgments: The work was supported by the National Science Foundation of China (60475001, 10571018) and by the Chinese Postdoctoral Science Foundation (2005038075).

10 May 2006
Electronics Letters online no: 20061456
doi: 10.1049/el:20061456

Z. Shi and C. Zhang (*State Key Laboratory of Intelligent Technology and Systems, Department of Automation, Tsinghua University, Beijing 100084, People's Republic of China*)

E-mail: shizhenwei@mail.tsinghua.edu.cn

References

1 Cichocki, A., and Amari, S.: 'Adaptive blind signal and image processing: learning algorithms and applications' (Wiley, New York, 2002)
2 Hyvärinen, A., Karhunen, J., and Oja, E.: 'Independent component analysis' (Wiley, New York, 2001)
3 Barros, A.K., and Cichocki, A.: 'Extraction of specific signals with temporal structure', *Neural Comput.*, 2001, **13**, (9), pp. 1995–2003
4 Hyvärinen, A.: 'Complexity pursuit: separating interesting components from time-series', *Neural Comput.*, 2001, **13**, (4), pp. 883–898
5 Shi, Z., Tang, H., and Tang, Y.: 'A fast fixed-point algorithm for complexity pursuit', *Neurocomputing*, 2005, **64**, pp. 529–536
6 Hyvärinen, A.: 'Blind source separation by nonstationarity of variance: a cumulant-based approach', *IEEE Trans. Neural Netw.*, 2001, **12**, (6), pp. 1471–1474