Available online at www.sciencedirect.com
SCIENCE DIRECT®
NEUROCOMPUTING
Neurocomputing 69 (2006) 917–921
www.elsevier.com/locate/neucom

Letters

Gaussian moments for noisy complexity pursuit

Zhenwei Shi*, Changshui Zhang

State Key Laboratory of Intelligent Technology and Systems, Department of Automation, Tsinghua University, Beijing 100084, PR China

Received 12 August 2005; received in revised form 27 September 2005; accepted 31 October 2005
Available online 4 January 2006
Communicated by R.W. Newcomb

Abstract

Complexity pursuit is an extension of projection pursuit to time series data and the method is closely related to blind separation of time-dependent source signals and independent component analysis (ICA). In this paper, we consider the estimation of the data model of ICA when Gaussian noise is present and the independent components are time dependent. We derive a simple algorithm combining Gaussian moments and complexity pursuit for noisy ICA. Validity and performance of the described approaches are demonstrated by computer simulations.

Keywords: Independent component analysis; Blind source separation; Complexity pursuit; Projection pursuit; Time series

1. Introduction

Recently, blind source separation by independent component analysis (ICA) has received attention because of its potential signal processing applications. The model of ICA consists of mixing independent random variables, usually linearly [1–3,6,10,12,13]. In many applications, however, what is mixed is not random variables but time signals, or time series. Complexity pursuit, whose goal is to find projections of time series that have interesting structure, is introduced by Hyvärinen [9]. It is an extension of projection pursuit [5] to time series data having interesting structure, defined using criteria related to Kolmogoroff complexity [11] or coding length. Time series which have the lowest coding complexity are considered the most interesting. This idea is probably connected to information-processing principles used in the brain. Shi et al. [14] derived a simple approximation of Kolmogoroff complexity that takes into account both the non-Gaussianity and the autocorrelations of the time series. And they developed a fixed-point algorithm for complexity pursuit. The method is closely related to blind separation of time-dependent source signals and ICA.

In this paper, we consider the estimation of the data model of ICA when Gaussian noise is present and the independent components (ICs) are time dependent. Hyvärinen [7,8] introduced a modification of the FastICA algorithm [6] using the Gaussian moments for the problem of estimating the data model of ICA in the presence of Gaussian noise when the ICs are random variables. Motivated by these methods, we combine Gaussian moments to complexity pursuit and derive a simple algorithm for noisy ICA when the ICs are time dependent.

2. Complexity pursuit and Gaussian moments

In this section, we consider the estimation of the ICA model when the ICs are time signals. The model is expressed by the following noisy linear model:

$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t) + \mathbf{n}(t), \tag{1}$$

where $\mathbf{x}(t) = (x_1(t), \ldots, x_M(t))^{\mathrm{T}} \in \mathbb{R}^M$ is a sensor signal, $\mathbf{A} \in \mathbb{R}^{M\times M}$ is an unknown mixing matrix, $\mathbf{s}(t) = (s_1(t), \ldots, s_M(t))^{\mathrm{T}} \in \mathbb{R}^M$ is the source signal and $\mathbf{n}(t)$ is the noise which is modeled as Gaussian with zero mean and

*Corresponding author. Tel.: +86 10 627 96 872; fax: +86 10 627 86 911.
E-mail addresses: szw1977@yahoo.com.cn (Z. Shi), zcs@mail.tsinghua.edu.cn (C. Zhang).

0925-2312/$ - see front matter
doi:10.1016/j.neucom.2005.10.002

covariance matrix $\mathbf{\Sigma}$ (i.e., $\mathbf{\Sigma} = E\{\mathbf{n}(t)\mathbf{n}(t)^{\mathrm{T}}\}$), where $t = 1, \ldots, T$ and t is the time index. The source signals are assumed independent such that $p(\mathbf{s}(t)) = \prod_{i=1}^{M} p(s_i(t))$ where p denotes the probability density function. For simplicity, we assume that the noise covariance matrix $\mathbf{\Sigma}$ is known and the source signals have zero mean and unit variance.

2.1. Quasi-whitening

A useful preprocessing strategy in ICA is to first whiten the observed variables. However, the effect of noise must be considered in the preliminary whitening of the data. If the noise covariance matrix is known, the ordinary whitening should be replaced by the 'quasi-whitening' operation $\tilde{\mathbf{x}}(t) = (\mathbf{C} - \mathbf{\Sigma})^{-1/2}\mathbf{x}(t)$, where $\mathbf{C} = E\{\mathbf{x}(t)\mathbf{x}(t)^{\mathrm{T}}\}$ is the covariance matrix of the observed noisy data. The quasi-whitened data $\tilde{\mathbf{x}}$ follows a noisy ICA model as well, i.e., $\tilde{\mathbf{x}}(t) = \mathbf{B}\mathbf{s}(t) + \tilde{\mathbf{n}}(t)$, with an orthogonal mixing matrix $\mathbf{B}$ (from $E\{\mathbf{x}(t)\mathbf{x}(t)^{\mathrm{T}}\} = \mathbf{C} = \mathbf{A}\mathbf{A}^{\mathrm{T}} + \mathbf{\Sigma}$ and $\mathbf{B} = (\mathbf{C} - \mathbf{\Sigma})^{-1/2}\mathbf{A}$), and the following noise covariance matrix [7,8]:

$$\tilde{\mathbf{\Sigma}} = E\{\tilde{\mathbf{n}}(t)\tilde{\mathbf{n}}(t)^{\mathrm{T}}\} = (\mathbf{C} - \mathbf{\Sigma})^{-1/2}\mathbf{\Sigma}(\mathbf{C} - \mathbf{\Sigma})^{-1/2}. \tag{2}$$

2.2. Predictive coding and approximation of complexity

Kolmogoroff complexity is based on the interpretation of coding length as structure [9]. Most natural signals have redundancy; parts of the signal can be efficiently predicted from other parts. Such a signal can be coded, or compressed, so that the code length is shorter than the original code length. We could thus measure the amount of structure of the signal by the amount of compression that is possible in coding the signal. For signals of fixed length, the structure could be measured by the length of the shortest possible code for the signal. This is a nonrigorous definition of Kolmogoroff complexity [9] (for a more rigorous definition of the concept, see the paper [11]).

Denote the noise-free data by $\mathbf{y}(t) = \mathbf{B}\mathbf{s}(t)$, the basic idea in the complexity pursuit is to find projections $\mathbf{w}^{\mathrm{T}}\mathbf{y}(t)$ such that the Kolmogoroff complexity of the projection is minimized.

First, we derive an approximation of the Kolmogoroff complexity of a scalar signal $z(t)(t = 1, \ldots, T)$ [14]. We consider predictive coding of a scalar signal. The value $z(t)$ is predicted from the preceding values by some function to be specified:

$$\hat{z}(t) = f(z(t-1), \ldots, z(1)). \tag{3}$$

To code the actual value $z(t)$, the residual,

$$\delta z(t) = z(t) - \hat{z}(t), \tag{4}$$

is coded by a scalar quantization method. According to the basic principles of information theory [4], the length of this code is asymptotically approximated by the sum of the entropies H of the residuals. The coding complexity can be approximated by [9,14]:

$$\hat{K}(z) = TH(\delta z), \tag{5}$$

where δz denotes a random variable with the marginal distribution of the residual.

To use the approximation in Eq. (5) in practice, we need to fix the structure of the predictor f and find an approximation of the entropy of δz. We use a computationally simple predictor structure, given by a linear autoregressive model:

$$\hat{z}(t) = \sum_{\tau > 0} \alpha_\tau z(t - \tau). \tag{6}$$

To approximate the entropy of δz, we assume that we know a good approximation of the (negative) logarithm of the probability density of the residual, denoted by G. Then we obtain the approximation [14]

$$H(\delta z) \approx E\{G(\delta z)\}. \tag{7}$$

In ICA, if the signals to be reconstructed satisfy certain properties, an exact form of the contrast function is not required in order to achieve the desired estimation results [9,14]. We may therefore optimistically assume that the exact form of the function G is not very important here either, as long as it is qualitatively similar enough to the (negative) logarithm of the probability density of the residual [9,14].

To find the 'most interesting' directions $\mathbf{w}$, use the above approximation of complexity for $z(t) = \mathbf{w}^{\mathrm{T}}\mathbf{y}(t)$. Thus, we can express the approximation of complexity as a contrast function of $\mathbf{w}$ only [14]:

$$\min_{\|\mathbf{w}\|^2=1} \hat{K}(\mathbf{w}^{\mathrm{T}}\mathbf{y}(t)) = E\left\{ G\left(\mathbf{w}^{\mathrm{T}} \left(\mathbf{y}(t) - \sum_{\tau>0} \alpha_\tau \mathbf{y}(t-\tau) \right) \right) \right\}. \tag{8}$$

Using the contrast function, we can derive the original gradient descent complexity pursuit algorithm.

2.3. Gaussian moments

The approach above could be used for noisy complexity pursuit as well (i.e., ICs are time signals in the noisy model), if only we had measures of the Kolmogoroff complexity which are immune to Gaussian noise, or at least, whose values for the original data can be easily estimated from noisy observations. The main point is to show that for certain choices of G, it is simple to estimate the values of $\hat{K}$ from noisy observation.

Denote by

$$\varphi_c(x) = \frac{1}{c}\varphi\left(\frac{x}{c}\right) = \frac{1}{\sqrt{2\pi}c}\exp\left(-\frac{x^2}{2c^2}\right), \tag{9}$$

the Gaussian density function of variance c^2, and by $\varphi_c^{(l)}(x)$ the lth ($l>0$) derivative of $\varphi_c(x)$. Denote further by $\varphi_c^{(-l)}(x)$ the lth integral function of $\varphi_c(x)$, obtained by $\varphi_c^{(-l)}(x) = \int_0^x \varphi_c^{(-l+1)}(\xi)\,\mathrm{d}\xi$, where we define $\varphi_c^0(x) = \varphi_c(x)$. Then we have the following theorem [7,8]:

Theorem 1. *Let v be any non-Gaussian random variable, and denote by n an independent Gaussian noise variable of*