Independent Component Analysis for Time-dependent Stochastic Processes

Aapo Hyvärinen
Helsinki University of Technology
Laboratory of Computer and Information Science
Rakentajanaukio 2 C, FIN-02150 Espoo, Finland
Email: aapo.hyvarinen@hut.fi

Abstract

The problem of linearly decomposing stochastic processes into 'independent' component processes is addressed. In contrast to ordinary independent component analysis, the time structure of the components is taken into account. It is shown that the data model of independent component analysis is identifiable if the innovation processes of the latent components (source signals) are independent. The results show the utility of performing independent component analysis on the innovation process instead of the original data.

1 Introduction

Independent component analysis (ICA) [7] is a statistical technique whose main applications are blind source separation, blind deconvolution, and feature extraction. In the simplest form of ICA [2], one observes m scalar random variables $x_1, x_2, ..., x_m$ which are assumed to be linear combinations of n unknown independent components, denoted by $s_1, s_2, ..., s_n$. The independent components s_i are assumed to be mutually *statistically independent*, and zero-mean. Arranging the observed variables x_j into a vector $\mathbf{x} = (x_1, x_2, ..., x_m)^T$ and the component variables s_i into a vector $\mathbf{s}$, the linear relationship can be expressed as

$$\mathbf{x} = \mathbf{A}\mathbf{s} \tag{1}$$

Here, $\mathbf{A}$ is an unknown $m \times n$ matrix of full column rank, called the mixing matrix. The basic problem of ICA is then to estimate both the mixing matrix $\mathbf{A}$ and the realizations of the independent components s_i *using only observations of the mixtures* x_j.

In this basic framework of ICA, the independent components are considered as random variables, with no time structure. Therefore, the assumption of independence is crucial for the identifiability of the model [2]. In this paper, we try to relax this assumption. We consider time-dependent stochastic processes instead of random variables, and utilize the notion of innovations. The innovation process of a stochastic process is roughly the new information fed to the process at a given time point. We show that it is enough for the identifiability of a time-dependent version of (1) that the components of the innovation process are independent. This is a more general condition than the independence of the components themselves. The mixing matrix $\mathbf{A}$ can then be estimated by applying ordinary ICA on the innovation process. We also argue that applying ICA on the innovation process makes the estimation more accurate in many cases.

2 Innovation process

The following developments are based on the concept of an innovation process of a stochastic process. Given a stochastic process $\mathbf{s}(t)$, we define its innovation process $\tilde{\mathbf{s}}(t)$ as the error of the best prediction (i.e. conditional expectation) of $\mathbf{s}(t)$, given its past:

$$\tilde{\mathbf{s}}(t) = \mathbf{s}(t) - \mathbf{E}(\mathbf{s}(t)|t, \mathbf{s}(t-1), \mathbf{s}(t-2), ...) \tag{2}$$

The expression 'innovation' describes the fact that $\tilde{\mathbf{s}}(t)$ contains all the new information about the process that can be obtained at time t. The t as given information in the expectation means that nonstationarities may occur. The innovation process is uniquely defined by (2). Our definition of innovation is a generalization of the conventional definition, see e.g. [4].

Estimation of the innovation process can be performed by approximating the conditional expectation, i.e. the best prediction of $\mathbf{s}(t)$ given its past (in the least mean-square sense). This is basically a regression problem that can be approximated in many cases by ordinary linear autoregressive models; in the very simplest case, a reasonable approximation of the innovation process may be given by the difference process $\Delta\mathbf{s}(t) = \mathbf{s}(t) - \mathbf{s}(t-1)$. In general, the nonlinear prediction may be approximated, e.g. by multi-layer perceptrons or radial basis functions.

3 ICA Estimation using innovations

Let us consider how the concept of innovation process can be used in the framework of estimation of the ICA data model. Consider a version of the ICA data model in (1) where the observed data is a stochastic process $\mathbf{x}(t)$ that is represented as a linear combination of component processes:

$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t). \tag{3}$$

The concept of innovations can be utilized in the estimation of the data model (3) due to the following lemma:

Lemma 1 *If $\mathbf{x}(t)$ and $\mathbf{s}(t)$ follow the instantaneous mixing model (3), then the innovation processes follow the ICA model (3) as well:*

$$\tilde{\mathbf{x}}(t) = \mathbf{A}\tilde{\mathbf{s}}(t). \tag{4}$$

To prove the lemma, it is enough to multiply both sides of (2) by $\mathbf{A}$ and use the linearity of the expectation operator, obtaining

$$\mathbf{A}\tilde{\mathbf{s}}(t) = \mathbf{x}(t) - \mathbf{E}(\mathbf{x}(t)|t, \mathbf{s}(t-1), \mathbf{s}(t-2), ...). \tag{5}$$

Since the information contained in $(\mathbf{s}(t-1), \mathbf{s}(t-2), ...)$ equals the information contained in $(\mathbf{x}(t-1), \mathbf{x}(t-2), ...)$ due to the invertibility of $\mathbf{A}$, this shows that $\mathbf{A}\tilde{\mathbf{s}}(t)$ is the innovation of $\mathbf{x}(t)$.

The lemma implies that it is enough for the model (3) to be identifiable that the innovation process fulfill the identifibility conditions usually required of the random vector $\mathbf{s}(t)$. In particular, *it is enough that* $\tilde{\mathbf{s}}(t)$ *has independent components* and is stationary as well as ergodic [2, 1]. This is a generalization of the ordinary identifiability conditions since the independence of the $s_i(t)$ implies the independence of the innovation processes $\tilde{s}_i(t)$. (This is due to the fact that independence of the $s_i(t)$ implies that the conditional expectation in (2) can be separated into a component-wise function.)

4 Why use innovations?

The relevant question here is: Why it would be better to apply ICA on the innovation instead of the original data. We claim that the following are valid reasons to use innovations:

1. The innovations are usually more independent from each other than the original processes. This is due to two factors. First, as mentioned above, the independence of the original processes implies the independence of the innovations, but not vice versa. Second, the innovations may correspond to physically independent processes that are mixed in an autoregressive process (inside the system) to give the source signals s_i that are no longer independent.

2. The innovations are usually more nongaussian than the original processes. Consider, for example, a simple 1-D autoregressive process:

   $$s(t) = as(t-1) + \tilde{s}(t). \tag{6}$$

   If the system is invertible, we have:

   $$s(t) = \sum_{\tau \geq 0} a^{\tau} \tilde{s}(t - \tau). \tag{7}$$

   This means that $s(t)$ is a sum of the innovation variables. A well-known result in ICA say that sums of nongaussian variables tend to be 'more gaussian' than the original variables. Thus the innovation process has a distribution that is more nongaussian than the distribution of s.

The accuracy of the estimation of the ICA model increases with increasing independence and nongaussianity of the components s_i [5]. In consequence, using innovations is likely to lead to much better estimates of the mixing matrix.

Thus these results show that to estimate the ICA data model when the independent components (source signals) have time-dependencies (i.e. spectral structure), it may be essential to *preprocess* the data before the application of conventional ICA algorithms. The preprocessing required is to extract the innovation process $\tilde{\mathbf{x}}(t)$ of the raw data, to be used in the conventional ICA method.

5 Simulations

Separating images of human faces has been considered a difficult case for ICA, since the faces, considered as 1-D signals of pixel gray-scale values, are not independent [8]. In this section we show that they can be nevertheless separated if ICA is applied on their innovation processes, which seem to be independent enough.

Four photos of faces are depicted in Fig 1. To show how dependent these signals are, we computed the correlation matrix:

$$\begin{bmatrix} 1.0000 & 0.2154 & 0.3988 & 0.4839 \\ 0.2154 & 1.0000 & 0.2594 & 0.1534 \\ 0.3988 & 0.2594 & 1.0000 & 0.6307 \\ 0.4839 & 0.1534 & 0.6307 & 1.0000 \end{bmatrix} \tag{8}$$

which shows that the source signals are rather strongly correlated. We then mixed these signals using a random mixing matrix. The mixed faces are shown in Fig. 2. First, ordinary ICA was applied to separation of the faces, using the algorithm in [6, 5]. The results are depicted in Fig. 3. Clearly, ordinary ICA failed to separate the faces. (The signs of the estimated independent components, which cannot be estimated by the ICA model, have been chosen to best approximate the original images.)

Next, we applied ICA on the innovation processes of the faces. To estimate (an approximation of) the innovation process of $\mathbf{x}$, we fitted a simple linear first-order autoregressive model on the data. Subtracting the obtained prediction from the data, we obtained a first approximation of the innovation process, which was essentially equal to the difference process. This trivial approximation is not good enough, however, because the signal is also quite nonstationary. When scanning the image row by row, significant nonstationarities are encountered in the beginning of each row. This gives large spikes in the obtained approximation of the innovation process. To make the approximation better, we discarded outliers in the first estimation of the innovation process. This means that we effectively used a predictor that takes into account the nonstationarity of the signal, obtaining a better approximation of the innovation process.

Thus we estimated the mixing matrix using the estimated innovation process. Separating the original images using the obtained estimate of the mixing matrix, we obtained the faces given in Fig. 4. Clearly, using innovation processes enabled accurate estimation of the mixing matrix.

Figure 1: The original faces used in the experiments.

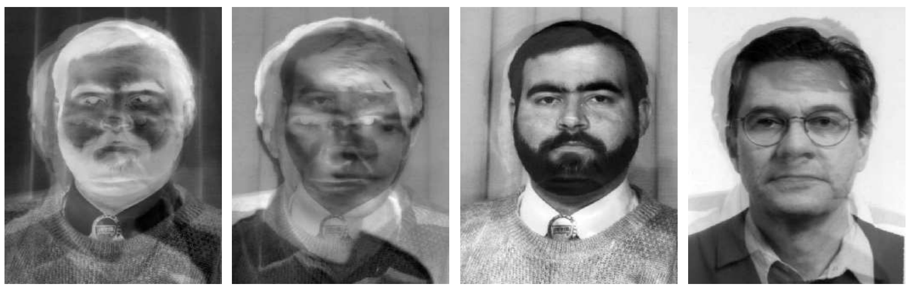

Figure 2: Linear mixtures of the faces.

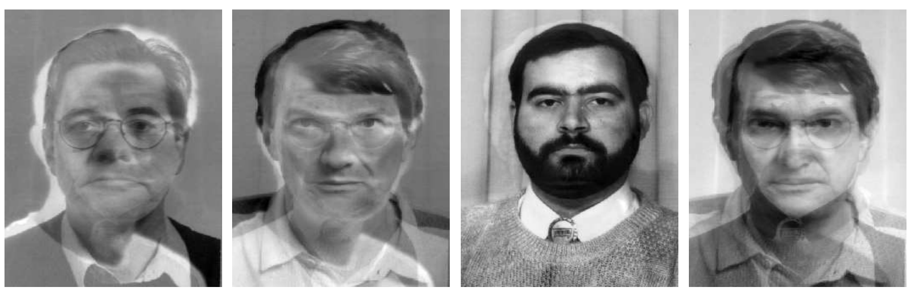

Figure 3: Original faces estimated by ordinary ICA.

Figure 4: Original faces estimated by ICA applied on innovation processes.

6 Conclusion

We considered independent component analysis of time-dependent stochastic processes. Using the concept of innovation processes, we generalized the identifiability conditions of the ordinary ICA model, showing that the independence of the components of the innovation process is sufficient for the identifiability of the model. We argued that using innovations enables a more accurate estimation of the ICA data model in many cases. In practice, this amounts to preprocessing the data before application of conventional ICA algorithms.

Our results are related to work on blind separation of convolutive mixtures [3, 9], but our framework is in some respects more general because it allows for nonlinearities in the conditional expectation, and does not require the signals to be stationary.[1] The framework of innovation processes is also closely related to methods using algorithmic complexity [8].

References

[1] J.-F. Cardoso and B. Hvam Laheld. Equivariant adaptive source separation. *IEEE Trans. on Signal Processing*, 44(12):3017–3030, 1996.

[2] P. Comon. Independent component analysis – a new concept? *Signal Processing*, 36:287–314, 1994.

[3] N. Delfosse and P. Loubaton. Adaptive blind separation of convolutive mixtures. In *Proc. ICASSP'96*, pages 2940–2943, 1996.

[4] S. Haykin. *Adaptive Filter Theory.* Prentice-Hall International, 3rd edition, 1996.

[5] A. Hyvärinen. Independent component analysis by minimization of mutual information. Technical Report A46, Helsinki University of Technology, Laboratory of Computer and Information Science, 1997.

[6] A. Hyvärinen and E. Oja. A fast fixed-point algorithm for independent component analysis. *Neural Computation*, 9(7):1483–1492, 1997.

[7] C. Jutten and J. Herault. Blind separation of sources, part I: An adaptive algorithm based on neuromimetic architecture. *Signal Processing*, 24:1–10, 1991.

[8] P. Pajunen. Blind source separation using algorithmic information theory. *Neurocomputing*, 1998. To appear.

[9] H.-L. Nguyen Thi and C. Jutten. Blind source separation for convolutive mixtures. *Signal Processing*, 45:209–229, 1995.

[1] Our framework is not a strict generalization of deconvolutive ICA, however, because in some cases where the convolutive ICA model holds, the innovation process as defined in (2) may not have independent components.