Independent Component Analysis
Term paper (Tutorial) Advanced Digital Signal Processing

Pin-Chun Hsu(B03901023), Prof. Jian-Jiun Ding

I. ABSTRACT

This tutorial presents an introduction to independent component analysis (ICA). Unlike principal component analysis, which is based on the assumptions of uncorrelatedness and normality, ICA is rooted in the assumption of statistical independence. Foundations and basic knowledge necessary to understand the technique are provided hereafter. Also included is a short tutorial illustrating the implementation of two ICA algorithms (FastICA and InfoMax) with the use of the Mathematica software.

II. INTRODUCTION

Imagine that you are in a room where two people are speaking simultaneously. You have two microphones, which you hold in different locations. The microphones give you two recorded time signals, which we could denote by $x_1(t)$ and $x_2(t)$, with x_1 and x_2 the amplitudes, and t the time index. Each of these recorded signals is a weighted sum of the speech signals emitted by the two speakers, which we denote by $s_1(t)$ and $s_2(t)$. We could express this as a linear equation:

$$x_1(t) = a_{11}s_1 + a_{12}s_2 \tag{1}$$

$$x_2(t) = a_{21}s_1 + a_{22}s_2 \tag{2}$$

where a_{11},a_{12},a_{21}, and a_{22} are some parameters that depend on the distances of the microphones from the speakers. It would be very useful if you could now estimate the two original speech signals $s_1(t)$ and $s_2(t)$, using only the recorded signals $x_1(t)$ and $x_2(t)$. This is called the cocktail-party problem. For the time being, we omit any time delays or other extra factors from our simplified mixing model. As an illustration, consider the waveforms in Fig. 1 and Fig. 2. These are, of course, not realistic speech signals, but suffice for this illustration. The original speech signals could look something like those in Fig. 1 and the mixed signals could look like those in Fig. 2. The problem is to recover the data in Fig. 1 using only the data in Fig. 2. Actually, if we knew the parameters a_{ij}, we could solve the linear equation in (1) by classical methods. The point is, however, that if you dont know the a_{ij}, the problem is considerably more difficult. One approach to solving this problem would be to use some information on the statistical properties of the signals $s_i t$ to estimate the aii. Actually, and perhaps surprisingly, it turns out that it is enough to assume that $s_1(t)$ and $s_2(t)$, at each time instant t, are statistically independent. This is not an unrealistic assumption in many cases, and it need not be exactly true in practice. The recently developed technique of Independent Component Analysis, or ICA, can be used to estimate the ai j based on the information of their independence, which allows us to separate the two original source signals $s_1(t)$ and $s_2(t)$ from their mixtures $x_1(t)$ and $x_2(t)$. Fig. 3 gives the two signals estimated by the ICA method. As can be seen, these are very close to the original source signals (their signs are reversed, but this has no significance.) Independent component analysis was originally developed to deal with problems that are closely related to the cocktail-party problem. Since the recent increase of interest in ICA, it has become clear that this principle has a lot of other interesting applications as well. Consider, for example, electrical recordings of brain activity as given by an electroencephalogram (EEG). The EEG data consists of recordings of electrical potentials in many different locations on the scalp. These potentials are presumably generated by mixing some underlying components of brain activity. This situation is quite similar to the cocktail-party problem: we would like to find the original components of brain activity, but we can only observe mixtures of the components. ICA can reveal interesting information on brain activity by giving access

to its independent components. Another, very different application of ICA is on feature extraction. A fundamental problem in digital signal processing is to find suitable representations for image, audio or other kind of data for tasks like compression and denoising. Data representations are often based on (discrete) linear transformations. Standard linear transforma- tions widely used in image processing are the Fourier, Haar, cosine transforms etc. Each of them has its own favorable properties (Gonzales and Wintz, 1987). It would be most useful to estimate the linear transformation from the data itself, in which case the transform could be ideally adapted to the kind of data that is being processed. Figure 4 shows the basis functions obtained by ICA from patches of natural images. Each image window in the set of training images would be a superposition of these windows so that the coefficient in the superposition are independent. Feature extraction by ICA will be explained in more detail later on. All of the applications described above can actually be formulated in a unified mathematical framework, that of ICA. This is a very general-purpose method of signal processing and data analysis.

III. DEFINITION

To rigorously define ICA (Jutten and Herault, 1991; Comon, 1994), we can use a statistical latent variables model. Assume that we observe n linear mixtures $x_1, ..., x_n$ of n independent components, for all j,

$$x_j = a_{j1}s_1 + a_{j2} + s_2 + ... + a_{jn}s_n \tag{3}$$

We have now dropped the time index t; in the ICA model, we assume that each mixture x_j as well as each independent component s_k is a random variable, instead of a proper time signal. The observed values $x_j(t)$, e.g., the microphone signals in the cocktail party problem, are then a sample of this random variable. Without loss of generality, we can assume that both the mixture variables and the independent components have zero mean: If this is not true, then the observable variables x_i can always be centered by subtracting the sample mean, which makes the model zero-mean. It is convenient to use vector-matrix notation instead of the sums like in the previous equation. Let us denote by x the random vector whose elements are the mixtures $x_1, ..., x_n$, and likewise by s the random vector with elements $s_1, ..., s_n$. Let us denote by $\mathbf{A}$ the matrix with elements a_{ij}. Generally, bold lower case letters indicate vectors and bold upper-case letters denote matrices. All vectors are understood as column vectors; thus x_T , or the transpose of $\mathbf{x}$, is a row vector. Using this vector-matrix notation, the above mixing model is written as

$$\mathbf{x} = \mathbf{A}\mathbf{s}. \tag{4}$$

Sometimes we need the columns of matrix $\mathbf{A}$; denoting them by $\mathbf{a}_j$ the model can also be written as

$$\mathbf{x} = \sum_{i=1}^{n} \mathbf{a}_i s_i \tag{5}$$

The statistical model in Eq. 4 is called **independent component analysis, or ICA model**. The ICA model is a generative model, which means that it describes how the observed data are generated by a process of mixing the components s_i. The independent components are latent variables, meaning that they cannot be directly observed. Also the mixing matrix is assumed to be unknown. All we observe is the random vector $\mathbf{x}$, and we must estimate both $\mathbf{A}$ and $\mathbf{s}$ using it. This must be done under as general assumptions as possible. The starting point for ICA is the very simple assumption that the components si are statistically independent. Statistical independence will be rigorously defined in Section 3. It will be seen below that we must also assume that the independent component must have nongaussian distributions. However, in the basic model we do not assume these distributions known (if they are known, the problem is considerably simplified.) For simplicity, we are also assuming that the unknown mixing matrix is square, but this assumption can be sometimes relaxed, as explained in Section 4.5. Then, after estimating the matrix $\mathbf{A}$, we can compute its inverse, say $\mathbf{W}$, and obtain the independent component simply by:

$$\mathbf{s} = \mathbf{W}\mathbf{x} \tag{6}$$

ICA is very closely related to the method called blind source separation (BSS) or blind signal separation. A source means here an original signal, i.e. independent component, like the speaker in a cocktail party problem. Blind means that we no very little, if anything, on the mixing matrix, and make little assumptions on the source signals. ICA is one method, perhaps the most widely used, for performing blind source separation. In many applications, it would be more realistic to assume that there is some noise in the measurements (see e.g. (Hyvarinen, 1998a; Hyvarinen, 1999c)), which would mean adding a noise term in the model. For simplicity, we omit any noise terms, since the estimation of the noise-free model is difficult enough in itself, and seems to be sufficient for many applications.

IV. CONCLUSION

V. REFERENCES