Independent Component Analysis: Algorithms and Applications

Aapo Hyvärinen and Erkki Oja

Presented by
Joshua Lewis and Deborah Goshorn
CSE291F 4/24/07

Plan

- Motivation
- ICA in a Nutshell
  - Caveats
- History
- Statistical Independence
- Nonlinear Decorrelation
- Central Limit Theorem
- Nongaussianity and Independence
- Kurtosis
- Deborah

Motivation (Blind Source Separation)

- *Imagine two signals (s_1 and s_2) measured by two recording devices (x_1 and x_2) at time t, where x_1 and x_2 are determined by some linear combination of s_1 and s_2:*
- $x_1(t) = a_{11}s_1(t) + a_{12}s_2(t)$ $\qquad$ $x_2(t) = a_{21}s_1(t) + a_{22}s_2(t)$
- In matrix-vector form: **x = As**
- **How does one recover the signal s from the mixture x?**

ICA in a Nutshell

- Given **$x = As$, if A is known then $s = A^{-1}x$**
- **If A is unknown, the problem is harder**
- **In most situations one can estimate $A^{-1} = W$ simply by assuming that *s_1 and s_2 are statistically independent at every time step t***
- ***Applications: EEG, image feature extraction, cocktail-party problem***

Caveats

- The variances of the independent components cannot be determined
  - With both **s and A unknown,** any scalar multiplier of a source s_i *can be canceled by dividing the appropriate column of* ***A by that same scalar***
- ***The order of the independent components cannot be determined***
  - ***Again with both s and A unknown, we can change the order of the terms in our original equation and call any signal the "first" one***

History of ICA

- Originally conceived of by H***érault, Jutten and Ans in the early 1980s to recover the position and velocity of a moving joint from sensory signals measuring muscle contraction***
- ***Largely ignored outside of France until the mid-1990s***
- ***Now several efficient and general ICA algorithms (such as FastICA) exist, and the technique has been applied to many problem domains***

Example

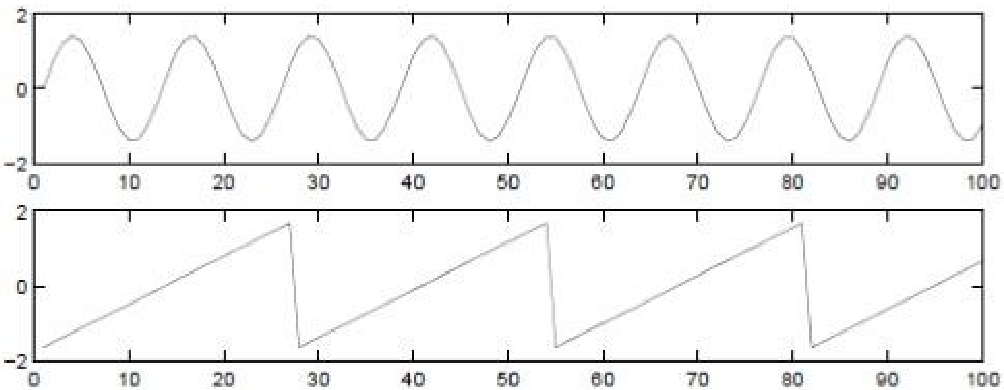

Figure 1: The original signals.

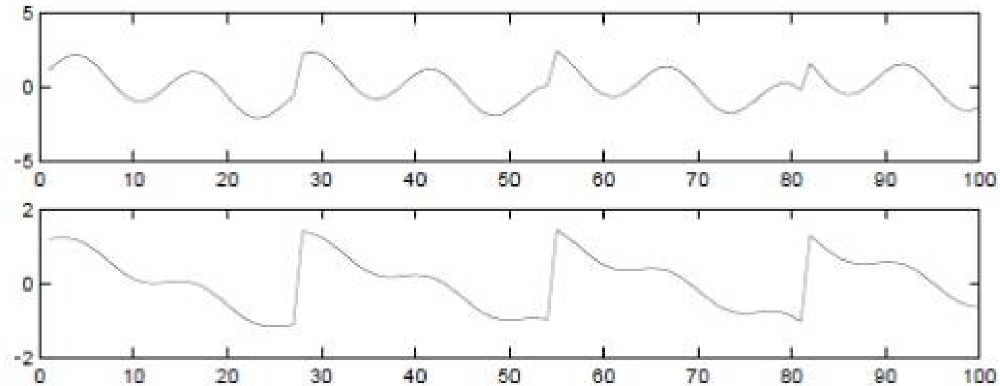

Figure 2: The observed mixtures of the source signals in Fig. 1.

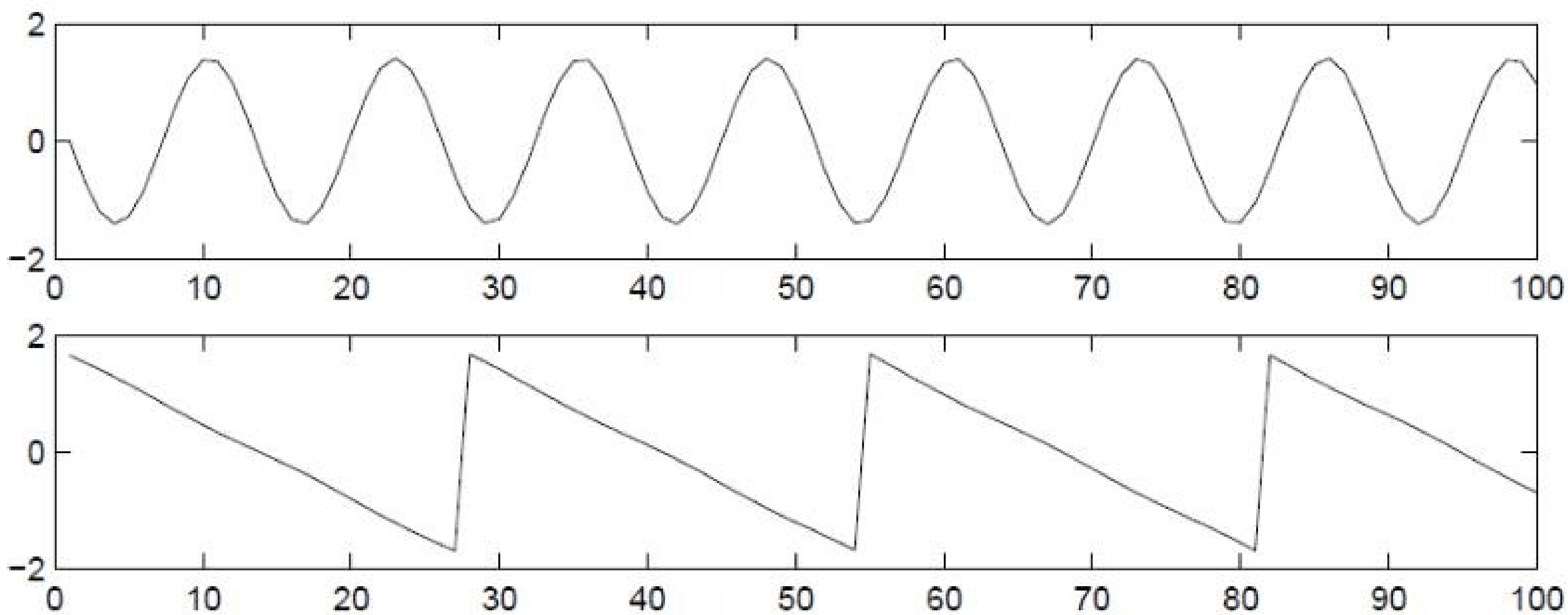

Figure 3: The estimates of the original source signals, estimated using only the observed signals in Fig. 2. The original signals were very accurately estimated, up to multiplicative signs.

Statistical Independence

- ***Consider two scalar random variables y_1 and y_2***
- ***y_1 and y_2 are statistically independent if information about y_1 does not give any information about y_2, in other words:***
- **Given** ***$p(y_1, y_2)$ as the joint pdf and***
- ***$p_1(y_1) = \int p(y_1,y_2)dy_2$ as the marginal pdf of y_1 (similarly for y_2)***
- ***$p(y_1, y_2) = p_1(y_1)p_2(y_2)$***

Nonlinear Decorrelation

- If *y_1 and y_2 are independent, transformations of y_1 and y_2 are uncorrelated:*

$$E\{h_1(y_1)h_2(y_2)\} = \int\int h_1(y_1)h_2(y_2)p(y_1,y_2)dy_1dy_2$$
$$= \int\int h_1(y_1)p_1(y_1)h_2(y_2)p_2(y_2)dy_1dy_2 = \int h_1(y_1)p_1(y_1)dy_1 \int h_2(y_2)p_2(y_2)dy_2$$
$$= E\{h_1(y_1)\}E\{h_2(y_2)\}.$$

- Find a candidate matrix **W** such that both the two estimated components, and some nonlinear transformation of those components are uncorrelated

Central Limit Theorem

- The distribution of a sum of independent random variables has a distribution that is closer to gaussian than the distribution of the original random variables

Why Nongaussian Is Independent

- Our estimate of an independent component $y = w^T x$, ***where w is to be determined***
- ***Define*** $z = A^T w$
- ***Now by definition*** $y = w^T x = w^T A s = z^T s$
- ***By central limit theorem,*** $z^T s$ ***is more gaussian than any component*** s_i, ***and when*** $z^T s$ ***is least gaussian, it must equal an element of s***
- ***Since*** $w^T x = z^T s$, ***choose w to maximize the nongaussianity of*** $w^T x$
- ***How does one estimate nongaussianity?***

Kurtosis

- The kurtosis (or fourth-order cumulant) of *y is defined by the following equation*

$$\text{kurt}(y) = E\{y^4\} - 3(E\{y^2\})^2$$

- If *y* is gaussian, the forth moment $E\{y^4\} = 3(E\{y^2\})^2$ and thus the kurtosis of *y* is 0
- If one maximizes the kurtosis of a random variable, one also maximizes the nongaussianity of that variable