Dependent component analysis (DCA)

Ivica Kopriva

Ruđer Bošković Institute

e-mail:
  ikopriva@irb.hr
  ikopriva@gmail.com
Course outline

- Mathematical preliminaries - what is independent component analysis (ICA)? (lecture 1)
- ICA for linear static (memoryless) problems (lecture 2)
- ICA for linear dynamic (convolutive) problems (lecture 3)
- Dependent component analysis (DCA) (lecture 4)
- Underdetermined blind source separation (uBSS), sparse component analysis (SCA) and nonnegative matrix factorization (NMF) (lecture 5)
- Nonlinear blind source separation (lecture 6)
Outline

• Blind source separation (BSS) and independent component analysis (ICA)

• Preprocessing transforms for BSS – enhancing statistical independence

• Examples
  – separation of the images of human faces (project problem)
  – blind image deconvolution (project problem)
  – blind speech deconvolution (project problem)
**BSS and ICA**

A theory for multichannel blind signal recovery requiring minimum of a *priori* information.

**Problem:**
\[ x = As \]

**Goal:** find \( s \) based on \( x \) only (\( A \) is unknown).

By means of independent component analysis (ICA) algorithms \( W \) can be found such that:

\[ y \cong s = Wx \rightarrow y \cong P \Lambda s \]

When does ICA work!? 

• source signals $s_i(t)$ must be statistically independent.

$$p(s) = \prod_{i=1}^{N} p_i(s_i)$$

• source signals $s_i(t)$, except one, must be non-Gaussian.

$$C_n(s_i) \neq 0 \quad n > 2$$

• mixing matrix $A$ must be nonsingular.

$$W \cong A^{-1}$$

Understanding origin i.e. interpretation of the mixing matrix is sometimes of crucial importance.
Preprocessing transforms for BSS – enhancing statistical independence

• We want to find a linear operator $T$ with the property that $T(s_m)$ and $T(s_n)$ are more independent than $s_m$ and $s_n \forall m, n$.

• Then, $W \cong A^{-1}$ is learnt by applying ICA on $T(x) = AT(s)$.

• The challenge is how to find appropriate linear operator $T$?
Preprocessing transforms for BSS – enhancing statistical independence

• Sub-band decomposition ICA (SDICA): wideband source signals are dependent, but there exists some sub-bands where they are less dependent.\textsuperscript{48,49-52}

• Innovations-based approach.\textsuperscript{53}

• Adaptive prefiltering of measured signals.\textsuperscript{54,55}


\textsuperscript{50}T. Tanaka, A. Cichocki, Subband decomposition independent component analysis and new performance criteria, Proc. ICASSP, 200


\textsuperscript{52}I. Kopriva, D. Sersic, Robust blind separation of statistically dependent sources using dual tree wavelets, ICIP 2007.


Enhancing statistical independence: innovations-based approach

• Arguments for using innovations (prediction errors) are that they are more independent from each other and more non-Gaussian than original processes → essentially important for the success of the ICA algorithms.

\[ \tilde{s}_m(t) = s_m(t) - E\left[ s_m(t) \mid s_m(t-1), s_m(t-2), \ldots \right] \]

\[ \tilde{x}(t) = A\tilde{s}(t) \]

• The approach is computationally efficient. Linear time invariant prediction error filter is efficiently estimated by means of Levinson algorithm.

• This approach fails if observed data contain interferer that is non-i.i.d. process (such signals are non-predictable).
Enhancing statistical independence: adaptive prefiltering-based approach\textsuperscript{5657}
Enhancing statistical independence: adaptive prefiltering-based approach\textsuperscript{56,57}

- Measured signals are preprocessed by an adaptive filter with adaptation based on the minimum of mutual information (MI) among the restored sources.

- In the first version of the method, [57], prefilter and static ICA are implemented simultaneously. It has been found in [53] that learning is not stable.

- In the improved version of the method, [57], prefilter is adapted on the subset of source signals, that are assumed to be available. ICA is then applied on filtered observed data. Assumption about availability of the subset of source signals is somewhat unrealistic in the BSS scenario.
Enhancing statistical independence: SDICA approach\textsuperscript{51-54}

• In SDICA approach the operator $T$ represents prefilter applied to all observed signals.

• By assumption introduced by Cichocki in [51], the wideband source signals are dependent, but some of their subcomponents are independent.

$$s(t) = \sum_{l=1}^{L} s_l(t)$$

• The challenge is how to find a subband index $1 \leq k \leq L$, such that $s_k$ contains least dependent subcomponents?
Enhancing statistical independence: SDICA approach\textsuperscript{51-54}

• In [51][52] a filter bank has been used to implement operator $T$. An assumption has been made that \textit{at least two sub-bands with statistically independent components exist}. As a criteria for sub-band localization was vicinity of the global separation matrix $W_m W_n^{-1}$ to the diagonal matrix. A special case of this approach is use of high pass filter instead of filter bank.

• In [53][54] we have respectively proposed the use of real and complex discrete wavelet transform to implement multiscale sub-band decomposition.
Enhancing statistical independence: SDICA approach\textsuperscript{53,54}

• In order to locate sub-band with least dependent components we have used small cumulant based approximation to measure the mutual information\textsuperscript{58,53} between the components of the measured signals in the corresponding nodes of the wavelet trees.

\[
\hat{I}_k^j \left( x_{k1}^j, x_{k2}^j, \ldots, x_{kn}^j \right) \approx \frac{1}{4} \sum_{0 \leq n < l \leq N} \text{cum}^2 (x_{kn}^j, x_{kl}^j) + \frac{1}{12} \sum_{0 \leq n < l \leq N} \left( \text{cum}^2 (x_{kn}^j, x_{kn}^j, x_{kl}^j) + \text{cum}^2 (x_{kn}^j, x_{kl}^j, x_{kl}^j) \right) \\
+ \frac{1}{48} \sum_{0 \leq n < l \leq L} \left( \text{cum}^2 (x_{kn}^j, x_{kn}^j, x_{kn}^j, x_{kl}^j) + \text{cum}^2 (x_{kn}^j, x_{kn}^j, x_{kl}^j, x_{kl}^j) + \text{cum}^2 (x_{kn}^j, x_{kl}^j, x_{kl}^j, x_{kl}^j) \right)
\]

where \( j \) represents scale index and \( k \) represents sub-band index at the appropriate scale.

\textsuperscript{58} J.F. Cardoso, Dependence, correlation and gaussianity in independent component analysis, J. Mach. Learn. Res. 4 (2003) 1177-1203
Estimation of the Mutual information

MI approximation as a function of the sample size for two Laplacian distributed processes. ‘x’ denotes entropy-based approximation while ‘o’ denotes cumulant-based approximation.
Estimation of the Mutual information

MI approximation as a function of dependence factor. Two uniformly distributed processes were used as independent processes. Normally distributed process was added as a dependent process with the scale factor that influenced dependence level. ‘x’ denotes entropy-based approximation while ‘o’ denotes cumulant based approximation.
Computation time for two approximations of the MI as a function of the sample size. Two Laplacian distributed processes were used as independent processes. ‘x’ denotes entropy-based approximation while ‘o’ denotes cumulant-based approximation.
Enhancing statistical independence: SDICA approach\textsuperscript{53,54}

Mutiscale analysis SDICA

\[
k^* = \arg \min_k I\left(x_1^k, \ldots, x_N^k\right)
\]

\[
W \cong ICA\left(x_k\right)
\]

\[
s \cong Wx
\]
Enhancing statistical independence: SDICA approach\textsuperscript{53,54}

• In [53] it has been proven that sub-band index $k$ with the least dependent components of the observed signals corresponds with the sub-band index with least dependent components of the source signals.

• The cumulant based approximation is demonstrated in [53] to be consistent estimator of the mutual information as well as to be computationally more efficient than original entropy based formulation. Estimation of density function, especially multidimensional density function, is computationally very demanding.

\[
\hat{I}_H(y) = \sum_{n=1}^{N} \hat{H}(y_n) - \hat{H}(y) \\
\hat{H}(y_n) = -\frac{1}{T} \sum_{t=1}^{T} \log \hat{p}_{y_n}(y_n(t)) \\
\hat{H}(y) = -\frac{1}{T} \sum_{t=1}^{T} \log \hat{p}_y(y(t))
\]
Enhancing statistical independence: SDICA approach\textsuperscript{53,54}

- In [53] it has been chosen between shift invariant non-decimated wavelet packets filter bank and decimated filter banks to trade off between separation quality and computational efficiency.

- Dual tree complex wavelet transform (DTCWT)\textsuperscript{59,60} has been invented to provide nearly shift-invariance property (good quality in signal representation) with substantially reduced aliasing due to the fact that complex-valued wavelet and scaling functions are approximately analytic (and consequently have single sided spectrum). Thus aliasing problems caused by downsampling operations are much smaller than with real DWT.

- In a case of 1D signals DTCWT is 2 x redundant (i.e. two decimated wavelet trees). In case of \(d\)-dimensional signals it is \(2^d\) x redundant.

Separation of images of human faces

• We have tested wavelet packets approach to blind separation of statistically dependent sources on separation of the images of human faces. They are known to be highly dependent i.e. people look quite similarly.

• We have added background Gaussian noise as an wide-band interferer to all source images with an average SNR \( \approx 30 \text{dB} \).
A) Source images

B) Observed images

C) Direct application of the ICA

D) Innovations based approach

E) Dual tree WT approach
Single-frame multi-channel blind image deconvolution

• In recent work in [61] it has been demonstrated to be possible to formulate single frame blind image deconvolution as a multichannel BSS problem solved by DCA algorithms.

• Benefit: no *a priori* knowledge about the size and origin of the degradation kernel is required.

---

The single frame multi-channel presentation was first proposed by Umeyama [62]. The key insight is the Taylor series expansion of $f(x+s,y+t)$ around $f(x,y)$.

$$f(x+s, y+t) = f(x, y) + sf_x(x,y) + tf_y(x,y) + ...$$

Then the degraded image is obtained as

$$g(x, y) = a_1 f(x, y) + a_2 f_x(x, y) + a_3 f_y(x, y) + ...$$

where

$$a_1 = \sum_{s=-K}^{K} \sum_{t=-K}^{K} h(s,t) \quad a_2 = \sum_{s=-K}^{K} \sum_{t=-K}^{K} sh(s,t) \quad a_3 = \sum_{s=-K}^{K} \sum_{t=-K}^{K} th(s,t)$$

The PSF coefficients are absorbed into mixing coefficients. No a priori knowledge about the nature of the blurring process or size of the blurring kernel is required.

A multi-channel representation is obtained by applying a bank of 2-D Gabor filters to the degraded image \( g(x,y) \).

A single frame multi-channel model is obtained as:

\[
G = \begin{bmatrix}
    g_1^T \\
    g_2^T \\
    \vdots \\
    g_L^T
\end{bmatrix} \approx \begin{bmatrix}
    a_1 & a_2 & a_3 \\
    a_{11} & a_{12} & a_{13} \\
    \vdots & \vdots & \vdots \\
    a_{L1} & a_{L2} & a_{L3}
\end{bmatrix} \begin{bmatrix}
    f_1^T \\
    f_x^T \\
    f_y^T
\end{bmatrix} = AF
\]
Single-frame multi-channel blind image deconvolution

A Gabor filter bank of 7x7 filters with two spatial frequencies and four orientations.
Single-frame multi-channel blind image deconvolution

A problem with the ICA approach arises if the images $f$, $f_x$, $f_y$, etc. are not statistically independent.

In Ref. [51] we have used DWT based SDICA algorithm to enhance statistical independence among $T(f)$, $T(f_x)$, $T(f_y)$, etc.
Single-frame multi-channel blind image deconvolution

Picture acquired by digital camera in manually de-focused mode.

Output of the Gabor filter bank:
Single-frame multi-channel blind image deconvolution

Picture acquired by digital camera in manually de-focused mode.
Blind speech deconvolution

Single channel recording:

\[ x(t) = \sum_{\tau=0}^{T} h(\tau) s(t - \tau) \]

The original signal \( s(t-\tau) \) can be approximated by Taylor series expansion around \( s(t) \) giving:

\[
 s(t - \tau) = \sum_{n=0}^{N} \frac{(-\tau)^n}{n!} \frac{d^n s(t)}{dt^n} + \text{H.O.T.}
\]

this gives:

\[
 x(t) \approx \sum_{n=0}^{N} a_{1(n+1)} \frac{d^n s(t)}{dt^n}
\]

where \( a_{11} = \sum_{\tau=0}^{T} h(\tau) \) \( a_{12} = -\sum_{\tau=0}^{T} \tau h(\tau) \) \( a_{13} = \sum_{\tau=0}^{T} \left( \tau^2 / 2 \right) h(\tau) \)

\(^{63}\) I. Kopriva, Blind Signal Deconvolution as an Instantaneous Blind Separation of Statistically Dependent Sources, LNCS 4666 (2007) 504-511.
Blind speech deconvolution

Non-i.i.d. signal (female speech) has been passed through band-limited (low-pass) channel.

2\textsuperscript{nd} order Butterworth filter has been used to model the channel.

1D DWT was used to convert single channel to “multi-channel” representation.

Innovations were used to enhance statistical independence among the time derivatives of the source signal. Prediction-error filter of the 10\textsuperscript{th} order has been used to implement innovations.
Blind speech deconvolution

Source

Filtered source signal

Recovered source signal - Taylor series expansion

Normalized correlation coefficients between source-filtered and restored signals: 0.71774 and 0.88658.