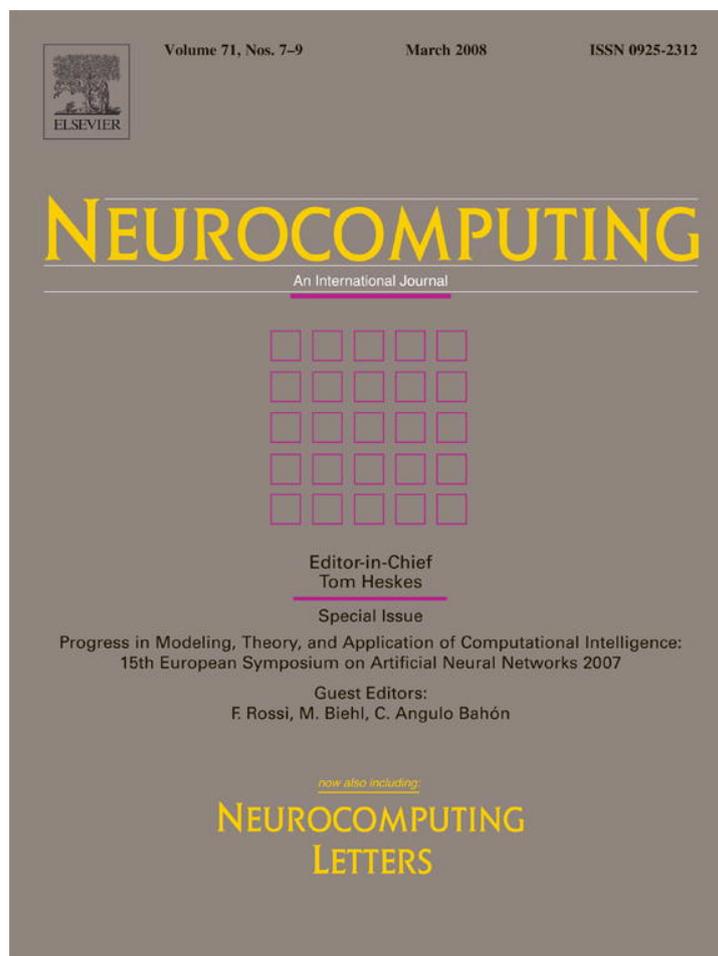


Provided for non-commercial research and education use.
Not for reproduction, distribution or commercial use.



This article was published in an Elsevier journal. The attached copy is furnished to the author for non-commercial research and education use, including for instruction at the author's institution, sharing with colleagues and providing to institution administration.

Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier's archiving and manuscript policies are encouraged to visit:

<http://www.elsevier.com/copyright>



Wavelet packets approach to blind separation of statistically dependent sources

Ivica Kopriva^{a,*}, Damir Seršić^b

^a*Rudjer Bošković Institute, Bijenička cesta 54, P.O. Box 180, 10002 Zagreb, Croatia*

^b*Department of Electronic Systems and Information Processing, Faculty of Electrical Engineering and Computing, University of Zagreb, Unska 3, 10000 Zagreb, Croatia*

Received 25 September 2006; received in revised form 2 April 2007; accepted 13 April 2007

Communicated by E.W. Lang

Available online 22 April 2007

Abstract

Sub-band decomposition independent component analysis (SDICA) assumes that wide-band source signals can be dependent but some of their sub-components are independent. Thus, it extends applicability of standard independent component analysis (ICA) through the relaxation of the independence assumption. In this paper, firstly, we introduce novel wavelet packets (WPs) based approach to SDICA obtaining adaptive sub-band decomposition of the wideband signals. Secondly, we introduce small cumulant based approximation of the mutual information (MI) as a criterion for the selection of the sub-band with the least-dependent components. Although MI is estimated for measured signals only, we have provided a proof that shows that index of the sub-band with least dependent components of the measured signals will correspond with the index of the sub-band with least dependent components of the sources. Unlike in the case of the competing methods, we demonstrate consistent performance in terms of accuracy and robustness as well as computational efficiency of WP SDICA algorithm.

© 2007 Elsevier B.V. All rights reserved.

Keywords: Sub-band decomposition; Independent component analysis; Wavelet packets; Mutual information

1. Introduction

Independent component analysis (ICA) is a statistical technique to extract non-Gaussian and statistically independent source signals given only the observed or measured data [12,23]. The problem is known as blind source separation (BSS) and is formally described as

$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t), \quad (1)$$

where $\mathbf{x} \in \mathbb{R}^N$ represents vector of measured signals, $\mathbf{A} \in \mathbb{R}^{N \times M}$ represents an unknown mixing matrix and $\mathbf{s} \in \mathbb{R}^M$ represents unknown vector of the source signals. In the subsequent derivation, we shall assume $M = N$. For $M > N$

the BSS problem is reduced on the squared problem by the dimensionality reduction technique, which is realized by the principal component analysis or singular value decomposition based transforms. The case $M < N$ leads to underdetermined BSS problem, which is not solvable under general ICA assumptions (non-Gaussianity and statistical independence of the source signals). Some additional *a priori* information about source signals, such as sparseness [20,29,30,44], must be known in order to solve the underdetermined BSS problem. Underdetermined case will not be treated in this paper.

Statistical independence assumption of the source signals is satisfied in many situations which leads to the successful application of the ICA in various fields functional magnetic resonance imaging signal processing [33], processing of radio signals from multiantenna based stations in wireless communication systems [36], multispectral astronomical and remotely sensed images [19,34], multiframe

*Corresponding author. Tel.: +44 385 1 4571 286;

fax: +44 385 1 4680 104.

E-mail addresses: ikopriva@gmail.com (I. Kopriva), damir.sersic@fer.hr (D. Seršić).

blind image deconvolution with non-stationary blurring process [10,27], speech separation and enhancement [2,28], etc.

However, in practice the independence property of the source signals does not always hold. The example is brain source signals observed by EEG [15]. One approach to solve this problem and relax statistical independence assumption is to assume that the wideband source signals are dependent, but there exist some sub-bands where they are independent. This assumption leads to the SDICA [13–15,38,40,41]. In [14,38], measured signals had to be passed through the filter bank in order to achieve sub-band decomposition. The problem remains how to select the sub-band with the least dependent components. Some *a priori* knowledge in a form of statistical measure such as kurtosis was used in [14] while in [38] it has been additionally assumed the existence of at least two sub-bands where sub-components of the source signals are independent. These assumptions may not necessarily hold in practical situations. In addition to that, in the later approach the problem still remains how to select the best sub-band.

A simplified version of the filter bank approach is to use the high pass (HP) filter only [14,15] in preprocessing the observed signals and apply standard ICA algorithm on filtered data in order to learn the mixing matrix. This is motivated by the fact that high pass filtered version is usually more independent than original signals. At the same time, this approach is computationally very efficient, which makes it attractive for BSS problems with statistically dependent sources. Nevertheless, we shall illustrate that due to its simplicity the method is not robust under certain interference conditions. The similar comments apply for innovation-based blind separation of statistically dependent sources [21,24]. The arguments for using innovations are that they are usually more independent from each other and more non-Gaussian than original processes [21]. In relation to [21] where innovations representation was proposed as a preprocessing method to increase accuracy of the linear instantaneous ICA algorithms, in [24] this was additionally extended to the post-nonlinear instantaneous ICA problems. In [24] it was assumed that source signals are statistically independent and temporally correlated. Usage of the linear filter to temporally decorrelate sources will make them more non-Gaussian and consequently increase accuracy of the ICA algorithms. We comment here that in the context of the BSS problem with statistically dependent sources, we use both properties of innovations: to be more independent and more non-Gaussian than original processes. Another reason to use innovations representation for discussed problem is its computational efficiency, because it is implemented by simple linear time invariant filter. However, as in the case of HP filter, we shall illustrate that innovations approach is also not robust under certain interference conditions.

Another approach has been formulated in [40,41] where measured signals were preprocessed by an adaptive filter

with adaptation based on the minimum of mutual information (MI) among the filter outputs. In the first version of the algorithm [40], prefilter adaptation and demixing matrix adaptation of instantaneous ICA stage are implemented simultaneously. In the second version of the algorithm [41], the prefilter is adapted alone on the subset of source signals, which are assumed to be available. ICA is then applied on filtered observed data. We find the assumption about the availability of the subset of the source signals unfair and somewhat unrealistic in the context of the blind processing scenario. Additional critique of this approach is problems associated with the adaptation of the prefilter. Standard score functions are needed in learning. It is known that *a priori* knowledge of at least statistical class to which source signals belong is required for their estimation. In order to obtain optimal score functions, some form of the kernel-based estimation of the unknown density functions is required [5,35], which is used in the implementation of the adaptive filter algorithm [40]. This is one part that adds to the computational complexity of this algorithm. Moreover, estimation of the joint score function is also required during filter adaptation phase. The joint score function depends on joint density function, which is known to be computationally very demanding to estimate, especially when the BSS problem is highly dimensional. The tradeoff has been found in [40] by minimizing pairwise MI. Anyway, this is another part that adds to the computational complexity of this algorithm. We shall demonstrate computational inefficiency of this algorithm as well as difficulties associated with the stability of the adaptive learning process in Section 4.

Alternative approach for separation of statistically dependent sources has been proposed in [7] where maximization of the non-Gaussianity measure equivalent to the minimization of the Shannon entropy is used. Performance of this approach strongly depends on the correlation level between the sources. In addition to that, there is also relatively high computational cost involved with the entropy-based approximation of MI.

We propose an approach to SDICA, which is based on multiresolution decomposition using wavelet packets (WPs) based iterative filter banks [31,39]. In order to enable the filter bank to adaptively select the sub-band with the least dependent sub-components of the source signals, we have introduced a criterion based on small cumulant approximation of MI. As it has been demonstrated in the Appendix A the cumulant based approximation is consistent estimator of the MI and computationally more efficient than entropy-based estimator. Although MI is estimated for measured signals only, we provide a proof at the end of Appendix A that shows that index of the sub-band with least dependent components of the measured signals will correspond with the index of the sub-band with least dependent components of the sources. Thus, our approach is adaptive as in [40,41] and computationally efficient as in [14,15,21,38]. Moreover, it is directly

extendable to time-frequency representation. Additionally, as demonstrated in Section 4, it exhibits consistent performance in terms of robustness and accuracy for various characters of interfering signals that act as dependent components.

The rest of the paper is organized as follows. We introduce the sub-band decomposition independent component analysis (SDICA) model in Section 2. The WP-based approach to SDICA is introduced in Section 3. Simulation results with comparative performance evaluation are presented in Section 4. Conclusion is given in Section 5.

2. SDICA model

In the BSS problem (1) to be solved by the ICA algorithms, it is assumed that sources $s_m(t)$, $m \in \{1, \dots, M\}$, are non-Gaussian and statistically independent. A powerful extension and generalization of this basic ICA model is SDICA. It assumes that wide-band source signals can be dependent, but some of their sub-components are independent. Thus, source signals can be represented as [14,15,38,40,41]:

$$s_m(t) = s_{m,1}(t) + s_{m,2}(t) + \dots + s_{m,L}(t), \quad (2)$$

where $s_{m,k}(t)$, $k = 1, \dots, L$, are narrow-band sub-components. The ICA problem is to find a separation or demixing matrix $\mathbf{W} \triangleq \mathbf{A}^{-1}$, which recovers independent components or source signals

$$\mathbf{y}(t) = \mathbf{W}\mathbf{x}(t), \quad (3)$$

where $\mathbf{y} \in \mathbb{R}^M$ and $\mathbf{y} \triangleq \mathbf{s}$. Like [40,41] we shall assume that for certain set of k , sub-components in (2) are least dependent or possibly independent. This significantly relaxes assumptions on the sub-band model (2) in relation to assumptions made in [14], where some *a priori* information about the sources is assumed. That contradicts to blindness assumption used to solve the BSS problem (1), or in [38], where existence of at least two sets of k sub-components is assumed. Under presented assumption the standard linear ICA algorithms can be applied to the selected set of k sub-components in order to learn demixing matrix \mathbf{W} :

$$\mathbf{y}_k(t) = \mathbf{W}\mathbf{x}_k(t). \quad (4)$$

The problems to be resolved are which preprocessing transform should be used to obtain sub-band representation of the original wideband BSS problem (1) and which criteria should be used to select the set k with least dependent sub-components.

3. Multiresolution SDICA

In order to obtain sub-band representation of the original wideband BSS problem (1), we can use any linear operator T_k which will extract a set k of sub-components

$$\mathbf{s}_k(t) = T_k[\mathbf{s}(t)], \quad (5)$$

where T_k can, for example, represent a linear time-invariant bandpass filter. Using (5) and sub-band representation of the sources (2), application of the operator T_k on the wideband BSS model (1) yields

$$\mathbf{x}_k(t) = T_k[\mathbf{A}\mathbf{s}(t)] = \mathbf{A}T_k[\mathbf{s}(t)] = \mathbf{A}\mathbf{s}_k(t). \quad (6)$$

For this purpose a fixed filter bank has been used in [14,38]. A high pass filtering can be seen as a special case of the filter bank approach. An adaptive pre-filter has been applied on x_n , $n \in \{1, \dots, N\}$ in [40,41] and trained by minimizing MI information among the filter outputs, thus eliminating dependent sub-components \mathbf{s}_k . In this paper we propose WP transform to be used for T_k in order to obtain sub-band representation of the wideband BSS problem (1). The main reason is existence of the WP transform in a form of iterative filter bank and multiresolution property of the WP transform which allows isolation of the fine details within each decomposition level and enable adaptive sub-band decomposition [31,39]. In the context of SDICA, it means that an independent sub-band that is arbitrarily narrow can be isolated, by progressing to the higher decomposition levels. Also, computationally efficient implementations of the WP transform exist for both two-dimension (2D) and 1D signals. This will be illustrated in the next section by applying WP-based SDICA approach to blind separation of human faces, as well as to 1D signals with a sine wave as interferer, i.e. dependent component. We further use properties of the WP transform in order to confirm existence of (6). That property was extensively exploited in various versions of the sparse ICA, where it has been found that either WP or short-time Fourier transform are very useful for obtaining new representation of data which is sparser than original formulation. As it has been shown, executing ICA in sparse domain produced more accurate solutions in solving linear instantaneous BSS problem and enabled to solve underdetermined (more sources than sensors) BSS problem [20,25,29,30,44]. In particular case of WP, we express each source signal (image) in terms of its decomposition coefficients:

$$s_{km}^j(t) = \sum_l c_{kml}^j \varphi_{jl}(t), \quad (7)$$

where j represents scale level, k represents sub-band index, m represents source index and l represents shift index. $\varphi_j(t)$ is the chosen wavelet also called atom or element of the representation space and c_{kml}^j are decomposition coefficients. In our implementation of the described SDICA algorithm, we have used shift-invariant 2D WP decomposition for separation of human faces and 1D WP for separation of 1D signals. If we choose the same representation space as for the source signals, we express each component of the observed data \mathbf{x} as

$$x_{kn}^j(t) = \sum_l f_{knl}^j \varphi_{jl}(t), \quad (8)$$

where n represents sensor index. Let vectors \mathbf{f}_l and \mathbf{c}_l be constructed from the l th coefficients of the mixtures and

sources, respectively. From (1) and (8) using the orthogonality property of the functions $\varphi_j(t)$ we obtain

$$\mathbf{f}_l = \mathbf{A}\mathbf{c}_l. \quad (9)$$

If additive noise is present this relation holds approximately, i.e. expanding the noise term in the same representation $n_{km}^j(t) = \sum_l e_{kml}^j \varphi_{jl}(t)$ would add up to (9) a vector \mathbf{e}_l of the decomposition coefficients of the noise [25]. Thus, when the noise is present Eq. (9) becomes $\mathbf{f}_l \approx \mathbf{A}\mathbf{c}_l$. From (1) and (9) we see the same relation between signals in the original domain and WP representation domain. Inserting (9) into (8) and using (7) we obtain:

$$\mathbf{x}_k^j(t) = \mathbf{A}\mathbf{s}_k^j(t), \quad (10)$$

as it was announced by (6) where no assumption of multiscale decomposition has been made. For each component x_n of the observed data vector \mathbf{x} , the WP transform creates a tree with the nodes that correspond to the sub-bands of the appropriate scale. In order to select the sub-band with least dependent components \mathbf{s}_k , we measure MI between the same nodes in the WP trees. For this purpose we use the small cumulant approximation of the Kullback–Leibler divergence, which is an exact measure of MI, obtained under weak correlation and weak non-Gaussianity assumptions [9]:

$$\begin{aligned} \hat{I}_k^j(x_{k1}^j, x_{k2}^j, \dots, x_{kN}^j) &\approx \frac{1}{4} \sum_{\substack{0 \leq n < l \leq N \\ n \neq l}} \text{cum}^2(x_{kn}^j, x_{kl}^j) \\ &+ \frac{1}{12} \sum_{\substack{0 \leq n < l \leq N \\ n \neq l}} \left(\text{cum}^2(x_{kn}^j, x_{kn}^j, x_{kl}^j) + \text{cum}^2(x_{kn}^j, x_{kl}^j, x_{kl}^j) \right) \\ &+ \frac{1}{48} \sum_{\substack{0 \leq n < l \leq L \\ n \neq l}} \left(\text{cum}^2(x_{kn}^j, x_{kn}^j, x_{kn}^j, x_{kl}^j) \right. \\ &\left. + \text{cum}^2(x_{kn}^j, x_{kn}^j, x_{kl}^j, x_{kl}^j) + \text{cum}^2(x_{kn}^j, x_{kl}^j, x_{kl}^j, x_{kl}^j) \right), \quad (11) \end{aligned}$$

where $\text{cum}()$ in (11) denotes second, third or fourth-order cross-cumulants [6,32]. Approximation of the joint MI by the sum of pair-wise MI is commonly used in the ICA community in order to simplify computational complexity of the linear instantaneous ICA algorithms [17]. We have demonstrated in Appendix A as follows: (i) that cumulant based approximation (11) of the MI is consistent as and computationally more efficient than entropy based approximation of MI [18]; (ii) although MI is estimated for measured signals only index of the sub-band with least dependent components of the measured signals will correspond with the index of the sub-band with least dependent components of the sources. However, because we use small cumulant-based approximation (11) of the MI, it means that sometimes instead of selecting a sub-band with least dependent sub-components of the source signals we shall only get close to this sub-band. By consistency, we assume that for mutually independent processes approximation (11) of the MI converges toward zero when the sample size grows toward infinity and that

its value is increased when dependence level between the processes is increased. Once the sub-band with the least dependent components is selected, we obtain either estimation of the inverse of the basis matrix \mathbf{W} or estimation of the basis matrix \mathbf{A} by applying standard ICA algorithms on the model (10). Reconstructed source signals \mathbf{y} are obtained by applying \mathbf{W} on the original data \mathbf{x} as it is given by (3). Alternatively, mixed signals can be reconstructed through the synthesis part of the WP transform, where sub-bands with a high level of MI are removed from the reconstruction. This is demonstrated in experiments 1 and 2 in the next section. We summarize multiscale SDICA BSS algorithm in the following four steps:

1. Perform multiscale WP decomposition of each component of the multivariate data \mathbf{x} . Wavelet tree will be associated to each component of \mathbf{x} (Eq. (6)–(10)).
2. Select sub-band with the least dependent components by estimating MI between the same nodes (sub-bands) in the wavelet trees (Eq. (10) and (11)).
3. Learn basis matrix \mathbf{A} or its inverse \mathbf{W} by executing standard ICA algorithm for linear instantaneous problem on the selected sub-band (Eq. (10)).
4. Obtain recovered sources \mathbf{y} by applying \mathbf{W} on data vector \mathbf{x} (Eq. (3)).

4. Simulation examples

In this section we examine performance of derived WP SDICA algorithm using the same examples as in [40]. The code for MATLAB implementation of the WP SDICA algorithm can be found and downloaded from [26]. In order to quantify separation quality, we also use the Amari's performance index P_{err} [1] that measures closeness of the global separation matrix $\mathbf{Q} = \mathbf{W}\mathbf{A}$ to the general permutation matrix $\mathbf{P} = \mathbf{\Lambda}\mathbf{I}$, where $\mathbf{\Lambda}$ represents diagonal matrix and \mathbf{I} represents identity matrix. P_{err} is calculated as

$$\begin{aligned} P_{\text{err}} &= \frac{1}{N(N-1)} \sum_{i=1}^N \left\{ \left(\sum_{j=1}^N \frac{|q_{ij}|}{\max_k |q_{ik}|} - 1 \right) \right. \\ &\left. + \left(\sum_{j=1}^N \frac{|q_{ji}|}{\max_k |q_{ki}|} - 1 \right) \right\}, \quad (12) \end{aligned}$$

where $q_{ij} = [\mathbf{Q}]_{ij}$ and $0 \leq P_{\text{err}} \leq 2$. P_{err} approaches zero when \mathbf{Q} approaches \mathbf{P} . We compare WP SDICA algorithm with adaptive prefiltering algorithms [40,41], innovations approach [21,24] and high pass prefiltering approach [14,15]. We have implemented the high pass prefiltering by means of 1D wavelet transform with the one decomposition level. It realizes coarse approximation and details by means of the two half band filters. All the experiments have been conducted in MATLAB[®] environment on a 3GHz PC machine with dual core microprocessor and 2GB of internal memory. In a case of all tested algorithms, we have used the FastICA algorithm in a symmetric mode

with *tanh* nonlinearity [22]. We would like to comment that choice of the nonlinearity for the FastICA algorithm is not critical as it would be for the ICA algorithm based on the classical score function [4,43]. We have also tested FastICA algorithm with *cubic* nonlinearity as well as JADE algorithm [8] that is known to be “nonlinearity free”. The obtained results were only slightly different than those obtained by FastICA algorithm with *tanh* nonlinearity. The reason why FastICA and any standard ICA algorithm will fail when applied directly on measured data is violation of statistical independence assumption.

4.1. Experiment 1: WP-based SDICA with artificially generated data

In this experiment, we have used the four independent sub-components: amplitude-modulated signal, a square wave, a high-frequency noise signal and a speech signal. Each signal has 10,000 samples. Each original signal s_i contains one of the above independent signals together with a sine wave with the same frequency, $\Omega = 0.41$ rad, but different phases for different sources, which represents dependent sub-component. Fig. 1 shows these signals as well as their magnitude spectra. Fig. 2 shows waveforms (left) and corresponding power spectrums (right) of the observed or mixed signals x_i . We have used the same mixing matrix as in [40]. Fig. 3 shows waveforms of the true dependent source signals (left) and waveforms of the estimated dependent source signals (right) reconstructed by direct application of the FastICA algorithm on the observed signals x_i . It is evident that separation quality is

poor. The value of the corresponding Amari’s error is $P_{err} = 0.4728$, which is significantly greater than zero. Evidently, the presence of the dependent sub-components prevented ICA algorithm from learning de-mixing matrix \mathbf{W} correctly. Fig. 4 shows equivalent results when described WP-based SDICA has been applied on observed data. We have used 1D shift-invariant WPs with five decomposition levels. Regarding the type of the wavelet our choice was symmlets with eight vanishing moments. The amount of the MI equals to $9.3059e-8$, where MI has been normalized with respect to the maximal value in the whole WP tree. In order to derive a truly unsupervised approach to blind separation of statistically dependent sources, we need a criterion for the selection of decomposition level. One way to automate algorithm completely is to monitor the rate of change of the MI, as decomposition level is increasing and stop when rate of change is small. For example, by carefully looking Table 1 we observed that change of MI in the part of the wavelet tree that spreads from the high pass part of the decomposition level 1 is insignificant in comparison with the MI at the high pass part of decomposition level 1. This issue is commented in more details at the end of the section. Table 1 shows normalized MI among the corresponding nodes of the wavelet trees obtained after WP-based decomposition of the observed data. The observed data were reconstructed using the synthesis part of the WP tree. Only the nodes with normalized $MI < 0.05$ have been retained, i.e. the nodes with higher level of MI were eliminated. In this way the dependent sub-component has been treated as interferer and eliminated from measurements before separation. We

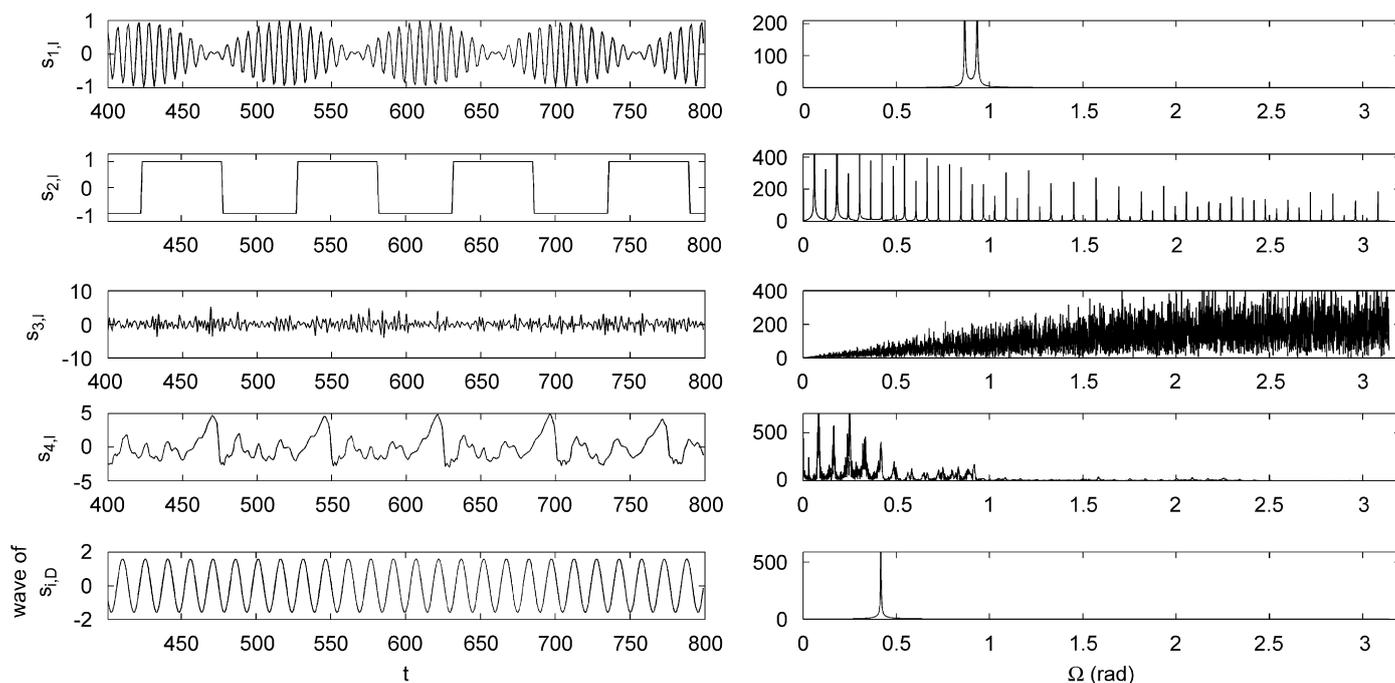


Fig. 1. The source independent sub-components (top four) and the waveform of dependent sub-component (the bottom one), as well as their magnitude spectra. Only 400 samples are plotted for illustration. The sources are $s_i = s_{i,I} + s_{i,D}$ and $s_{i,D}$ are sinusoid waves with the same frequency but different phases.

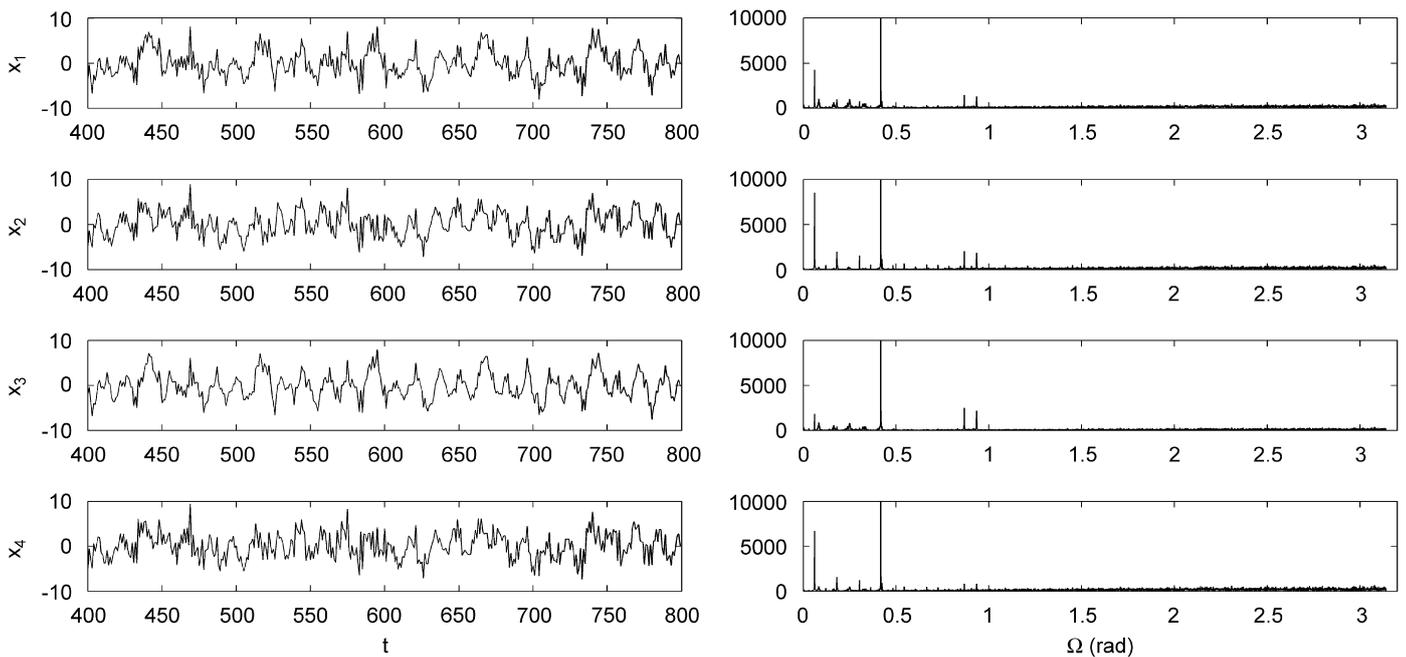


Fig. 2. The observed signals (left) and their magnitude spectra (right) in the experiment 1.

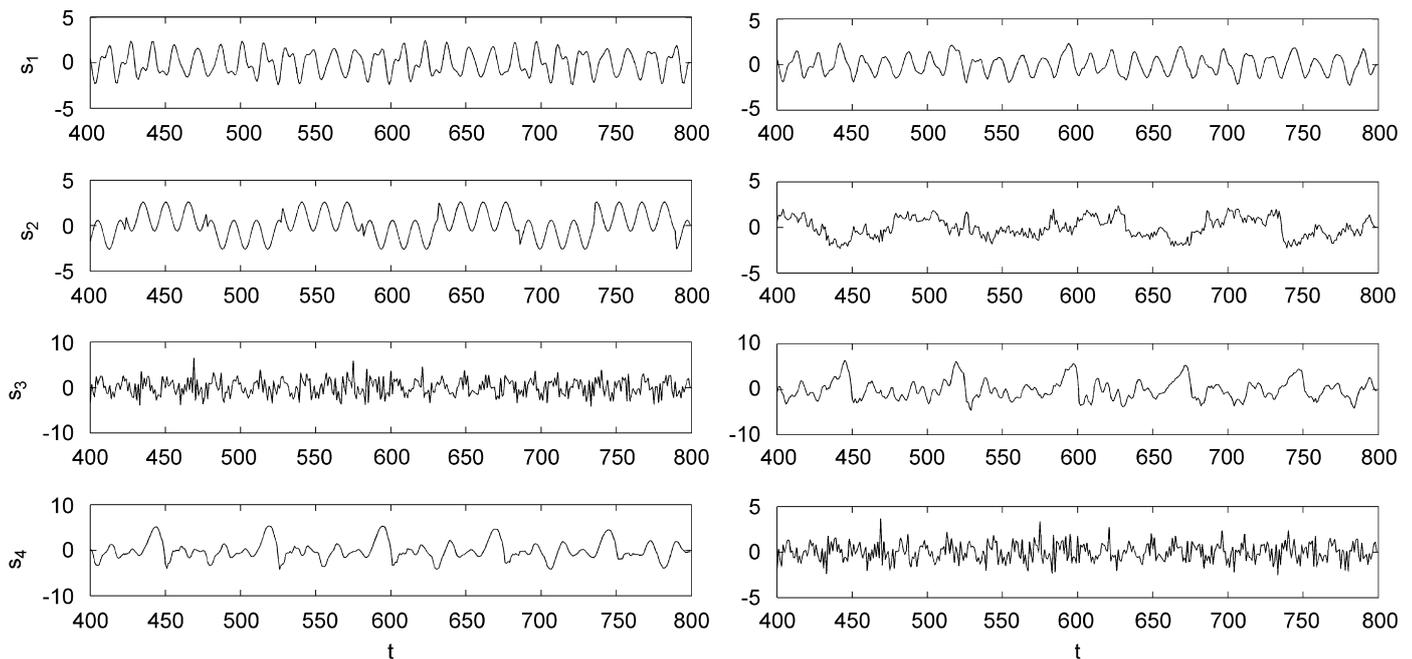


Fig. 3. The source independent sub-components with added dependent component (left) and reconstructed waveforms by direct application of FastICA method on the observed data (right).

then applied FastICA algorithm on the reconstructed observed data in order to learn de-mixing matrix \mathbf{W} . The Amari's error was $P_{\text{err}} = 0.0051$, that is 92.7 times less than the value when the FastICA algorithm has been applied directly on measured data. The whole process, analysis and synthesis part, took approximately 9.5 s.

The competing innovations approach has been tested with a 10th order of the AR model. The prediction filter

was very efficient in eliminating the dependent sine wave component. The FastICA algorithm has been applied on innovations to learn the de-mixing matrix \mathbf{W} . The Amari's error was $P_{\text{err}} = 0.0885$ that is 5.3 times less than the value when the FastICA algorithm has been applied directly on measured data. The whole process took approximately 0.3 s. When HP filter has been applied to solve described BSS problem value of the Amari's error was $P_{\text{err}} = 0.0823$,

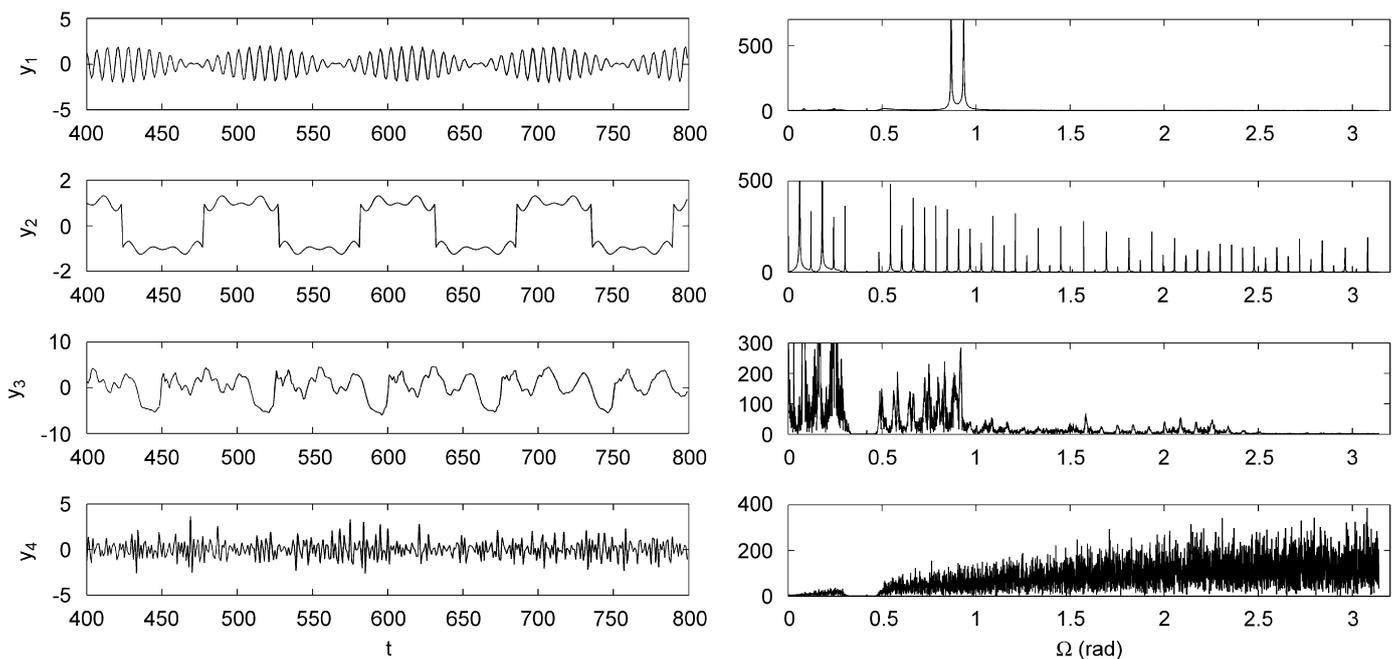


Fig. 4. Reconstructed source independent sub-components (left) and their magnitude spectra (right) obtained by WP-based SDICA. It can be seen from the magnitude spectra that dependent component with $\Omega \approx 0.41$ rad is removed from the reconstructed sources.

that is 5.7 times less than the value when the FastICA algorithm has been applied directly on measured data. The whole process took approximately 0.22 s. The HP filter used in the experiment has been obtained from the WPs. It is interesting to note that use of simple differentiator, i.e. FIR filter with coefficients $[1 \ -1]$, that works very well with face images gives very poor performance in the presented case of 1D signals. For such a filter value of Amari's error was $P_{\text{err}} = 0.5646$.

Thus, it is evident that both HP and innovations are reasonably accurate and yet computationally very efficient. On the other hand, WP approach offers an order of magnitude better accuracy with an increased computational time that is still of the order of a few seconds. However, if we change the frequency of the sine dependent components to $\Omega = 2.14$ rad, the WP approach achieves the value of Amari's error $P_{\text{err}} = 0.0069$, the innovations approach $P_{\text{err}} = 0.01271$ and the HP filtering approach $P_{\text{err}} = 0.4423$. The reason while HP filtering approach failed has been of course that interference is in a high pass part of the spectrum. Without *a priori* information about location of the interference HP approach fails. By this simple experiment we have demonstrated the meaning and importance of adaptability of the proposed WP SDICA approach.

The adaptive prefilter algorithm has been applied on mixed data with adaptive filter learning gain $\eta_h = 0.1$, demixing matrix learning gain $\eta_w = 0.15$ and filter order equal to 17. These values were suggested in [40]. We have used the code provided on the web-site [42]. The algorithm failed to remove sine dependent component. When applied

in *independence enhanced* mode [40], which assumes availability of the subset of source signals, the adaptive filter strongly suppressed sine dependent component and achieved value of the Amari's error was $P_{\text{err}} = 0.0046$. However, we find assumption about availability of the subset of source signals unfair in the context of blind scenario.

In [16] additive information cost function was proposed for the selection of the best basis of WPs. Table 1 shows the values of normalized MI for all nodes in the WP tree. It is important to note that MI is not an additive function that fits in the best basis scheme. Nevertheless, we can compare MI measure in the parent and child nodes and decide whether the further splitting was fruitful. If the child nodes do not differ significantly among each other in the MI, we can stop the further splitting in that direction.

4.2. Experiment 2: WP-based SDICA for separating images of human faces

This example has been chosen from the same reason as in [40]. Namely, it has been already found out that face images, represented as 1D signals, are dependent and hard to separate [21]. We demonstrate here the capability of the WP SDICA algorithm to successfully separate four images of human faces and compare its performance with adaptive prefiltering algorithm [40], innovations approach and HP prefiltering approach. We have used differentiator with the impulse response $h = [1 \ -1]$ to implement HP filter for which it is known to works well with the face images. Regarding the type of the wavelet, we have also used

symmlets with eight vanishing moments. The four original images are shown in Fig. 5. It has been already demonstrated that both innovations and HP prefiltering are successful in separating faces from their mixtures [21,40]. However, let us assume the existence of the external background Gaussian noise that is treated in BSS scenario as another source signal [12]. It obviously represents a

dependent sub-component. We have added it such that an average SNR with respect to source images was 29.1 dB. The noise corrupted source images are mixed with the 4×4 random mixing matrix. Mixed images are shown in Fig. 6(a). It can be observed that external background noise is not visible. Fig. 6(b) shows separated images obtained after applications of the FastICA algorithm on the mixed images. The quality of the separated images is evidently very poor, that is also confirmed by the value of the Amari's error $P_{err} = 0.7382$. If we now apply the WP SDICA algorithm, innovations approach and HP prefiltering approach to these noisy data, we respectively obtain the following values of the Amari's error $P_{err} = 0.0344, 0.2761$ and 0.1685 . Performance of WP SDICA algorithm has been improved 21.4 times, performance of the innovations approach has been improved only 2.7 times and performance of the HP prefiltering approach has been improved 4.38 times. Fig. 6(c) shows separated image by WP SDICA algorithm, while Fig. 6(d) shows separated images by innovations approach. Normalized mutual information among the corresponding nodes of the wavelet trees obtained after WP based decomposition of the observed data is shown in Table 2. The observed data were reconstructed using the synthesis part of the wavelet packet tree. Only the nodes with normalized mutual information less than 0.05 have been retained. The reason why innovations failed is that dependent white Gaussian noise term is not predictable. The HP prefiltering approach failed due to the wideband character of the external noise term, i.e. the HP filtering kept half of its power. The WP SDICA algorithm with two decomposition levels took approximately 91 s. The innovations approach took approximately 19 s while HP prefiltering approach took approximately 12 s. When WP transform was implemented in decimated version obtained Amari's error for two decomposition levels was $P_{err} = 0.04$, which is slightly worse than for non-decimated version while the execution time was equal to 14 s. This experiment is additional evidence about the robustness of the WP SDICA algorithm as well as its flexibility in achieving tradeoff between accuracy and computational efficiency.

The adaptive prefilter algorithm has been additionally tested on a case without external noise with the parameter values as proposed in [40] and using code provided by Zhang et al. [42]. The filter order was 11, filter learning

Table 1
Values of the normalized mutual information between the same nodes in the wavelet packet trees of the mixed signals in experiment 1

Level 0	Level 1	Level 2	Level 3	Level 4	Level 5
6.6338e-6	1.0165e-4	0.001568	0.001114	<i>0.004123</i>	0.03762
				0.005760	4.3069e-5
			0.003929	<i>9.8547e-8</i>	0.09283**
				0.06284	<i>3.7627e-4</i>
		<i>9.5130e-8</i>	1.1057e-7	1.1187e-7	1.0896e-7
				1.0864e-7	9.3059e-8*
			1.2969e-7	9.2024e-7	<i>1.1561e-7</i>
				1.0114e-7	1.1368e-7
					1.1205e-7
					1.0740e-7
					1.2148e-7
					1.6753e-6
					3.9823e-6
					1.0295e-7
					9.8636e-8
	1.1252e-7	<i>1.1725e-7</i>	1.3701e-7	3.3412e-7	7.5134e-7
					2.2103e-7
					1.1139e-7
					1.7694e-7
					2.6185e-7
					2.6591e-7
					1.2917e-7
					6.6695e-7
					1.2161e-7
					1.1270e-7
					1.2488e-7
					1.5636e-7
					1.2049e-7
					3.2190e-7
					1.2617e-7
					2.1985e-7
					1.4093e-7

Node that corresponds to the minimum MI is marked with an asterisk, while nodes that represent sub-bands with a high level of mutual information are marked with two asterisks. The italic nodes denote the sub-bands from which the partial reconstruction has been done.

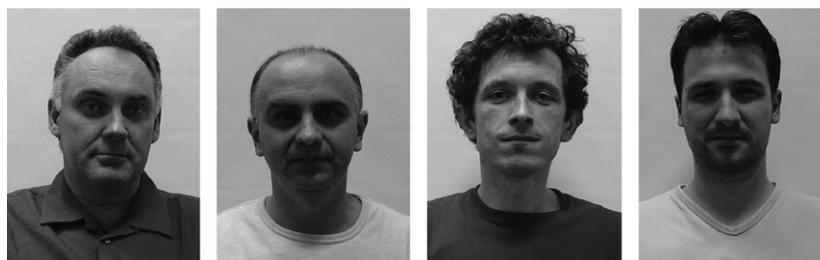


Fig. 5. Separating mixtures of images of human faces: original images.



Fig. 6. Separating mixtures of images of human faces with an external Gaussian noise source. Average SNR value was equal to 29.1 dB. (a) Mixed images. (b) Separation results obtained by FastICA algorithm. (c) Separation results obtained by WP SDICA algorithm. (d) Separation results obtained by innovations based algorithm.

gain 0.04 and de-mixing matrix learning gain in the instantaneous ICA stage 0.08. Initial value of de-mixing matrix has been obtained by the FastICA algorithm. The best result obtained after 265 iterations was $P_{\text{err}} = 0.0134$. However after 500 iterations algorithm stopped at similar value $P_{\text{err}} = 0.0124$, but the error has been oscillating during the adaptation process. The value of the Amari's error vs. iteration index is shown in Fig. 7(a), which illustrates oscillating character of the learning process. We have decreased the prefilter and de-mixing matrix learning gains 10 times to 0.004 and 0.008. Fig. 7(b) shows behavior of the Amari's error vs. iteration index for 5000 iterations. Minimum of the Amari's error $P_{\text{err}} = 0.0077$ has been

achieved after 4567 iterations. Evidently, the learning is more stable but small oscillations still remained. We found very difficult to determine when to stop the algorithm. The oscillating character of the learning process may represent serious difficulty in real world applications where Amari's error is impossible to be calculated and monitored. There are no such difficulties associated with the WP SDICA approach. We have also tested the [41] on the same example in a slightly modified mode, i.e. the prefilter has been applied on observed data only and ICA algorithm has been applied to filtered data in order to learn de-mixing matrix. Obtained results are equivalent to those already reported when filtering and instantaneous ICA are

executed together [40]. Regarding computational complexity of the [40] algorithm it took approximately 360 s for 500 iterations. For an accuracy of $P_{err} = 0.0077$ achieved after 4567 iterations it took 3240 s. The same accuracy has been achieved by WP SDICA algorithm in 91 s only.

5. Conclusion

We have formulated a new method for blind separation of statistically dependent sources. The method is based upon assumption that wideband sources are dependent, but there exists sub-band where sources are independent. Adaptive sub-band decomposition is realized through WP

transform implemented in a form of iterative filter bank. The sub-band with least dependent components is selected by measuring MI between the same nodes of the wavelet trees in the multiscale decomposition of the mixed signals. The computationally efficient small cumulant approximation of the MI has been used for this purpose. If reconstruction of dependent sub-components is desired, the de-mixing matrix, learnt by applying the standard ICA algorithm on selected sub-band mixed signals, is applied on original mixed signals. This is illustrated by successful separation of images of human faces from the mixtures. If reconstruction of dependent sub-components is not desired, the learnt de-mixing matrix is applied on a *cleaned* version of the mixed signals obtained through the synthesis part of the WP transform, where only sub-bands with normalized MI less than predefined threshold retained. Demixing matrix is learnt by applying ICA on reconstructed mixed signals with eliminated dependent sub-components. This is illustrated by successful separation of artificially generated 1D signals corrupted by a sine wave as dependent sub-component. In relation to the innovations algorithm proposed in [21] and high pass prefiltering algorithm proposed in [14,15], it has been demonstrated that WP SDICA algorithm exhibits consistent performance in terms of accuracy and robustness with respect to various interfering signals that act as dependent components, while the other two methods fail under certain circumstances. Yet, the WP SDICA algorithm keeps reasonably small computational complexity. Unlike the adaptive prefiltering algorithms proposed in [40,41], the WP SDICA algorithm does not have any stability problems associated with an adaptive learning process and is an order of magnitude computationally more efficient. Thus, we consider that WP SDICA algorithm may be quite useful in practical applications due to its robustness and computational efficiency.

Table 2
Values of the normalized mutual information between corresponding nodes in the wavelet packet trees of the mixed signals in experiment 2

Level 0	Level 1	Level 2
3.1104e-2	3.6502e-2	1.00000** 3.4570e-2 2.5430e-2*
	3.9920e-2	3.2171e-2 5.2337e-2** 3.8559e-2 2.9874e-2 3.3041e-2
	2.8660e-2	2.9893e-2 2.8736e-2 2.7758e-2 3.1539e-2
	3.0871e-2	2.8029e-2 2.9845e-2 3.0672e-2 3.1825e-2

For description of used notation see Table 1.

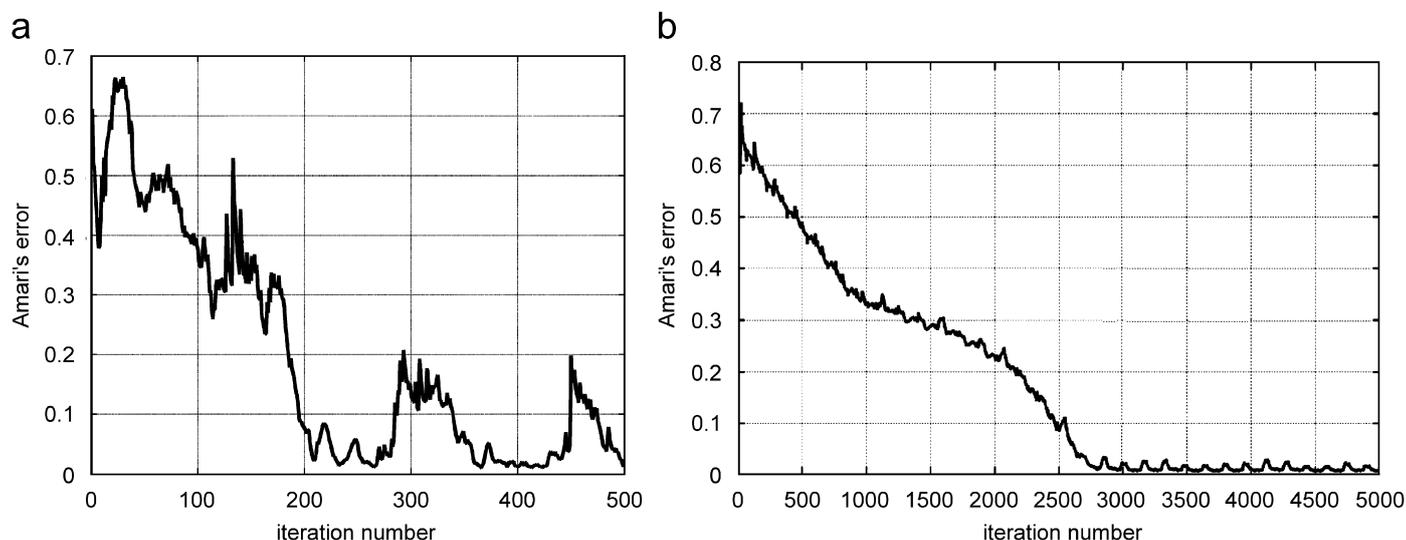


Fig. 7. Amari's error vs. iteration index for algorithm (Zhang and Chan, 2006a) in blind separation of human faces. (a) Prefilter learning gain 0.04, de-mixing matrix learning gain 0.08. (b) Prefilter learning gain 0.004, de-mixing matrix learning gain 0.008.

Acknowledgment

The authors wish to thank the anonymous reviewers for their constructive comments that certainly improved the quality of the paper. Part of this work was supported through grant 098-0982903-2558 funded by the Ministry of Science, Education and Sport, Republic of Croatia.

Appendix A. Approximation of the MI and its use in sub-band selection

Under weak correlation and weak non-Gaussian assumptions, it has been shown in [9], Eqs. (70), (66) and (68), that MI can be approximated via small cumulant approximation of the Kullback–Leibler divergence (Gram–Charlier expansion of non-Gaussian distributions around normal distribution) as

$$I(y_1, y_2, \dots, y_N) \approx \frac{1}{4} \sum_{\substack{1 \leq i < j \leq N \\ i \neq j}} c_{ij}^2 + \frac{1}{2} \sum_{r \geq 3} \frac{1}{r!} \sum_{i_1 i_2, \dots, i_r=1}^N c_{i_1 i_2, \dots, i_r}^2, \quad (\text{A.1})$$

where $c_{i_1 i_2, \dots, i_r}^2$ denotes square of the related r th order cross-cumulant and $i_1 i_2, \dots, i_r$ denotes partition of indices such that they are not all identical and c_{ij} denotes cross-correlation or second-order cross-cumulant. If we further assume that all the cumulants of the order higher than 4 (weak non-Gaussian assumptions) are small, then (A.1) is reduced to

$$I(y_1, y_2, \dots, y_N) \approx \frac{1}{4} \sum_{\substack{1 \leq i < j \leq N \\ i \neq j}} c_{ij}^2 + \frac{1}{12} \sum_{i_1 i_2 i_3=1}^N c_{i_1 i_2 i_3}^2 + \frac{1}{48} \sum_{i_1 i_2 i_3 i_4=1}^N c_{i_1 i_2 i_3 i_4}^2. \quad (\text{A.2})$$

We now apply another approximation commonly done in ICA community [17] and approximate joint independence $I(y_1, y_2, \dots, y_N)$ by the sum of pairwise independences and arrive to

$$I(y_1, y_2, \dots, y_N) \approx \sum_{\substack{1 \leq i < j \leq N \\ i \neq j}} I(y_i, y_j) \approx \frac{1}{4} \sum_{\substack{1 \leq i < j \leq N \\ i \neq j}} c_{ij}^2 + \frac{1}{12} \sum_{\substack{1 \leq i < j \leq N \\ i \neq j}} \sum_{n,m} \left(\text{cum}(\underbrace{y_i, \dots, y_i}_n, \underbrace{y_j, \dots, y_j}_m) \right)^2 + \frac{1}{48} \sum_{\substack{1 \leq i < j \leq N \\ i \neq j}} \sum_{p,q} \left(\text{cum}(\underbrace{y_i, \dots, y_i}_p, \underbrace{y_j, \dots, y_j}_q) \right)^2, \quad (\text{A.3})$$

where in (A.3) $n, m \in \{1, 2\}$, $p, q \in \{1, 2, 3\}$, $n+m=3$ and $p+q=4$. As it has been proven in [9] it applies for such a measure $I(y_1, y_2, \dots, y_N) \geq 0$. It is equal to zero when $\{y_n\}_{n=1}^N$ are mutually statistically independent or pairwise

statistically independent. This pairwise approximation of MI has been also used in [40], Eq. (19), as a cost function for filter adaptation. It follows that above approximation represents consistent measure of statistical (in)dependence. From the practical reasons we shall rewrite Eq. (A.3) in terms of explicit use of second-, third- and fourth-order cross-cumulants

$$\hat{I}_c(y_1, y_2, \dots, y_N) \approx \frac{1}{4} \sum_{\substack{1 \leq i < j \leq N \\ i \neq j}} \text{cum}^2(y_i, y_j) + \frac{1}{12} \sum_{\substack{1 \leq i < j \leq N \\ i \neq j}} \left(\text{cum}^2(y_i, y_i, y_j) + \text{cum}^2(y_j, y_j, y_i) \right) + \frac{1}{48} \sum_{\substack{1 \leq i < j \leq N \\ i \neq j}} \left(\text{cum}^2(y_i, y_i, y_i, y_j) + \text{cum}^2(y_i, y_i, y_j, y_j) + \text{cum}^2(y_j, y_j, y_j, y_i) \right), \quad (\text{A.4})$$

where subscript c denotes cumulant-based approximation. Alternative to this approximation is to employ the entropy-based approximation that follows from the definition of the MI [18]:

$$I(\mathbf{y}) = \sum_{n=1}^N H(y_n) - H(\mathbf{y}), \quad (\text{A.5})$$

where H denotes the entropy, $H(y_n) = -E[\log(p_{y_n}(y_n))]$ and p_{y_n} is the density function of y_n . It is necessary to approximate marginal entropies, joint entropy and related density functions:

$$\hat{I}_H(\mathbf{y}) = \sum_{n=1}^N \hat{H}(y_n) - \hat{H}(\mathbf{y}), \quad (\text{A.6})$$

where subscript H denotes entropy-based approximation. We follow approach exposed in [3]:

$$\hat{H}(y_n) = -\frac{1}{T} \sum_{t=1}^T \log \hat{p}_{y_n}(y_n(t)),$$

$$\hat{p}_{y_n}(y_n(t)) = \frac{1}{T} \sum_{m=1}^T K_h(y_n(t) - y_n(m)),$$

$$\hat{H}(\mathbf{y}) = -\frac{1}{T} \sum_{t=1}^T \log \hat{p}_{\mathbf{y}}(\mathbf{y}(t)),$$

$$\hat{p}_{\mathbf{y}}(\mathbf{y}(t)) = \frac{1}{T} \sum_{m=1}^T \prod_{n=1}^N K_h(y_n(t) - y_n(m)), \quad (\text{A.7})$$

where T denotes the sample size. In (A.7) K_h represents kernel. Various kernels might be used for density estimation [11,35,37]. The commonly used kernel is Gaussian kernel

$$K_h(y_n(t) - y_n(m)) = \exp\left(-\frac{(y_n(t) - y_n(m))^2}{h}\right), \quad (\text{A.8})$$

where h denotes the bin width of the kernel. Recommendation for choosing h is [37]:

$$h = 0.6 \frac{\hat{\sigma}}{\sqrt[5]{T}}, \quad (\text{A.9})$$

where $\hat{\sigma}$ denotes sample standard deviation of y_n . The joint density estimator in (A.7) is expressed using a product of 1D kernels. This simplifies computational complexity, but represents a restriction since joint density does not factorize except if marginal variables are independent. However, it has been proven in [3] that this choice improves the bias of the estimator of the MI that asymptotically converges toward zero with a speed $1/T$. Therefore, MI estimator (A.6) satisfies $\hat{I}_H(\mathbf{y}) \geq 0$ with the equality to zero only when variables of \mathbf{y} are statistically independent.

We now demonstrate consistency of the two MI estimators $\hat{I}_c(\mathbf{y})$ and $\hat{I}_H(\mathbf{y})$. Fig. A.1 shows MI as a function of the sample size for two Laplacian distributed processes, where ‘x’ denotes entropy-based approximation and ‘o’ denotes cumulant-based approximation. It is evident that both approximations converge toward zero as sample size increases. Fig. A.2 shows approximations of the MI for two partially dependent processes $s_1 = s_1'' + cs_d$ and $s_2 = s_2'' + cs_d$ where s_1'' and s_2'' were two zero mean uniformly distributed processes and dependent component s_d being normally distributed process with zero mean and unit variance. The scale factor c has been varied between 0.1 and 1 with the step equal to 0.1. It is evident from Fig. A.2 that both \hat{I}_c and \hat{I}_H are consistent, i.e. $\hat{I}_{c(H)}(s_1(c_1), s_2(c_1)) > \hat{I}_{c(H)}(s_1(c_2), s_2(c_2))$ for $c_1 > c_2$. In the case of dependent processes both approximations are less accurate, \hat{I}_H due to the already discussed reasons related to the use of 1D kernel in the joint density estimator (A.7) and \hat{I}_c due to the fact that only cumulants up to the order four were used in the approximation (A.4). Nevertheless, demonstrated consistency makes both measures suitable for detection of the sub-band with the least dependent components. Due to the huge difference in computational complexity, we have selected cumulant-based approximation \hat{I}_c for the use in sub-band detection criterion. For the sake of illustration, Fig. A.3 shows estimated computation time for MATLAB implementation of both approxima-

tions as a function of the sample size for the example analyzed in Fig. A.1. Note that small inconsistency with the cumulant-based measure is due to the poor resolution capability of the MATLAB functions *clock* and *etime*, on the milliseconds scale, that were used for measuring the computation time.

After demonstrating consistency of the MI estimators $\hat{I}_c(\mathbf{y})$ and $\hat{I}_H(\mathbf{y})$, we now want to prove that index of the sub-band with the statistically least dependent components of the measured signals, that is found at the minimum of the MI of the measured signals, corresponds with the index of the sub-band with the statistically least dependent components of the source signals. For that purpose, we rewrite sub-band decomposition representation, Eq. (6), of the wideband BSS model:

$$\mathbf{x}_k(t) = \mathbf{A}\mathbf{s}_k(t). \quad (\text{A.10})$$

Index of the sub-band with the statistically least dependent components of the measured signals k^* is obtained as

$$k^* = \arg \min_k I(\mathbf{x}_k). \quad (\text{A.11})$$

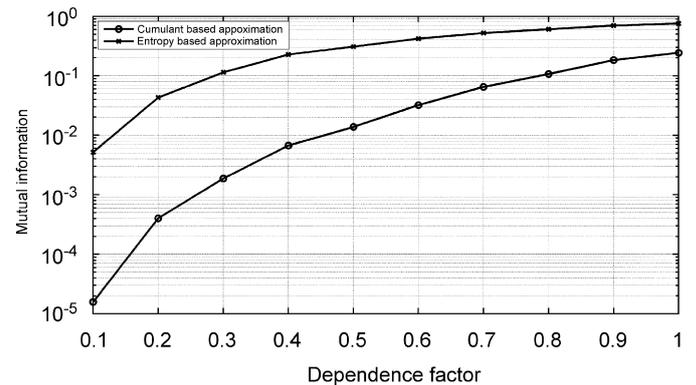


Fig. A2. MI approximation as a function of dependence factor. Two uniformly distributed processes were used as independent processes. Normally distributed process was added as a dependent process with the scale factor that influenced dependence level. ‘x’ denotes entropy-based approximation while ‘o’ denotes cumulant based approximation.

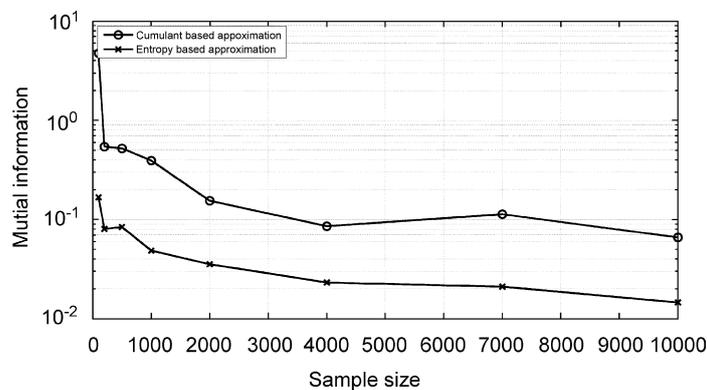


Fig. A1. MI approximation as a function of the sample size for two Laplacian distributed processes. ‘x’ denotes entropy-based approximation while ‘o’ denotes cumulant-based approximation.

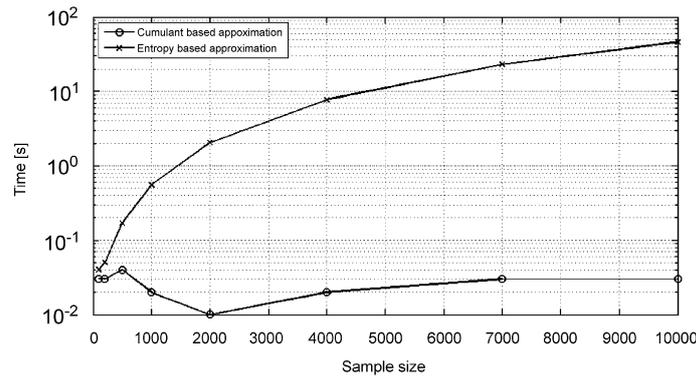


Fig. A3. Computation time for two approximations of the MI as a function of the sample size. Two Laplacian distributed processes were used as independent processes. 'x' denotes entropy-based approximation while 'o' denotes cumulant-based approximation.

Consequently, we want to prove

$$k^* = \arg \min_k I(\mathbf{x}_k) = \arg \min_k I(\mathbf{s}_k). \quad (\text{A.12})$$

Let us now express the MI in a form of Kullback divergence:

$$\begin{aligned} I(\mathbf{x}_k) &= D\left(p(\mathbf{x}_k) \parallel \prod_{n=1}^N p_n(x_{kn})\right) \\ &= \int p(\mathbf{x}_k) \log \frac{p(\mathbf{x}_k)}{\prod_{n=1}^N p_n(x_{kn})} \\ &= \sum_{n=1}^N H(x_{kn}) - H(\mathbf{x}_k), \end{aligned} \quad (\text{A.13})$$

where $H(x_{kn})$ and $H(\mathbf{x}_k)$ respectively represent marginal and joint differential entropy [18, pp. 224–233]. MI defined as Kullback divergence is a convex function with the property $I(\mathbf{x}_k) \geq 0$ and equality with zero only if x_{kn} are independent, i.e. $p(\mathbf{x}_k) = \prod_{n=1}^N p_n(x_{kn})$, see theorem 9.6.1 in [18]. Also from the corollary defined by Eq. (9.59) in [18], it follows:

$$H(\mathbf{x}_k) \leq \sum_{n=1}^N H(x_{kn}), \quad (\text{A.14})$$

with equality only if x_{kn} are independent. Because we are looking for the sub-band with least dependent components of the measured signals, sum of the marginal entropies in such a case will be closest to the joint entropy. Therefore, from (A.13) and (A.14) it follows:

$$k^* = \arg \min_k I(\mathbf{x}_k) = \arg \min_k \sum_{n=1}^N H(x_{kn}) - \arg \max_k H(\mathbf{x}_k). \quad (\text{A.15})$$

It follows from (A.14) and (A.15) that minimizing the MI is equivalent to the simultaneous minimization of the sum of marginal entropies and maximization of the joint entropy. Thus, maximization of joint entropy implies simultaneous minimization of the sum of marginal entropies driving MI toward zero. We now express joint entropy $H(\mathbf{x}_k)$ in terms of the joint entropy $H(\mathbf{s}_k)$ using the rule for the entropy of the linear transformation $\mathbf{x}_k = \mathbf{A}\mathbf{s}_k$

[18] (theorem 9.6.4, i.e. Eq. (9.67)):

$$H(\mathbf{x}_k) = H(\mathbf{s}_k) + \log |\det \mathbf{A}|. \quad (\text{A.16})$$

Taking account that second term in (A.16) is invariant with respect to the sub-band index k it follows:

$$\arg \max_k H(\mathbf{x}_k) = \arg \max_k H(\mathbf{s}_k). \quad (\text{A.17})$$

Due to already discussed reasons maximization of $H(\mathbf{s}_k)$ will simultaneously imply minimization of $\sum_{n=1}^N H(s_{kn})$. Hence

$$\begin{aligned} k^* &= \arg \min_k I(\mathbf{x}_k) \\ &= \arg \min_k \sum_{n=1}^N H(x_{kn}) - \arg \max_k H(\mathbf{x}_k) \\ &= \arg \min_k \sum_{n=1}^N H(s_{kn}) - \arg \max_k H(\mathbf{s}_k) \\ &= \arg \min_k I(\mathbf{s}_k), \end{aligned} \quad (\text{A.18})$$

which proves (A.12). However, because we use small cumulant-based approximation (A.4) of the MI (A.5), the equality (A.12) holds approximately. It means that sometimes instead of selecting a sub-band with least dependent sub-components of the source signals we shall only get close to it.

References

- [1] S. Amari, S.A. Cichocki, A.H.H. Young, A new learning algorithm for blind signal separation, MIT Press, Cambridge, MA, 1996, pp. 757–763.
- [2] S. Araki, S. Makino, R. Aichner, T. Nishikawa, H. Saruwatari, Subband-based blind separation for convolutive mixtures of speech, IEICE Trans Fundament E88-A (2005) 3593–3603.
- [3] S. Archard, D.T. Pham, Ch. Jutten, Criteria based on mutual information minimization for blind source separation in post nonlinear mixtures, Signal Process 85 (2005) 965–975.
- [4] A.J. Bell, T.J. Sejnowski, An information-maximization approach to blind separation and blind deconvolution, Neural Comput. 7 (1995) 1129–1159.
- [5] R. Boscolo, H. Pan, Independent component analysis based on nonparametric density estimation, IEEE Trans. Neural Networks 15 (2004) 55–65.

- [6] D.R. Brillinger, Time Series Data Analysis and Theory, McGraw-Hill, New York, 1981.
- [7] C.F. Caiafa, A.N. Proto, Separation of statistically dependent sources using L^2 -distance non-Gaussian measure, Signal Process 86 (2006) 3404–3420.
- [8] J.F. Cardoso, A. Soulomniac, Blind beamforming for non-Gaussian signals, Proc. IEE F 140 (1993) 362–370.
- [9] J.F. Cardoso, Dependence, correlation and gaussianity in independent component analysis, J. Mach. Learning Res, 4 (2003) 1177–1203.
- [10] J. Chen, X.Z. Wang, A new approach to near-infrared spectral data analysis using independent component analysis, J. Chem. Inform. Comput. Sci. 41 (2001) 992–1001.
- [11] A. Chen, Fast kernel density independent component analysis, Lecture Notes in Computer Science 3889 (2006) 24–31.
- [12] A. Cichocki, S. Amari, Adaptive blind signal and image processing, John Wiley, New York, 2002.
- [13] A. Cichocki, Blind source separation: new tools for extraction of source signals and denoising, Proceedings of the SPIE 5818, Orlando, FL, 2005, pp. 11–25.
- [14] A. Cichocki, P. Georgiev, Blind source separation algorithms with matrix constraints, IEICE Trans. Fund. Electron. Commun. Comput. Sci. E86-A (2003) 522–531.
- [15] A. Cichocki, General component analysis and blind source separation methods for analyzing multichannel brain signals, in: M.J. Wenger, C. Schuster (Eds.), Statistical and Process Models of Cognitive Aging, Notre Dame Series on Quantitative Methods, Erlbaum, Mahwah, NJ, 2007.
- [16] R.R. Coifman, M.V. Wickerhauser, Entropy-based algorithms for best basis selection, IEEE Trans. Inform. Theory 38 (1992) 713–718.
- [17] P. Comon, Independent component analysis, a new concept?, Signal Process 36 (1994) 287–314.
- [18] T. Cover, J. Thomas, Elements of Information Theory, Wiley, New York, 1991.
- [19] Q. Du, I. Kopriva, H. Szu, Independent component analysis for hyperspectral remote sensing, Opt Eng 45 (2006), 017008-1-13.
- [20] P. Georgiev, F. Theis, A. Cichocki, Sparse component analysis and blind source separation of underdetermined mixtures, IEEE Trans. Neural Networks 16 (2005) 992–996.
- [21] A. Hyvärinen, Independent component analysis for time-dependent stochastic processes. Proceedings of the International Conference on Artificial Neural Networks (ICANN'98), Skovde, Sweden, 1998, pp. 541–546.
- [22] A. Hyvärinen, Fast and robust fixed-point algorithms for independent component analysis, IEEE Trans. Neural Networks 10 (1999) 626–634.
- [23] A. Hyvärinen, J. Karhunen, E. Oja, Independent Component Analysis, Wiley Interscience, New York, 2001.
- [24] J. Karvanen, T. Tanaka, Temporal decorrelation as preprocessing for linear and post-nonlinear ICA, Lecture Notes Comput Sci 3195 (2004) 774–781.
- [25] P. Kisilev, M. Zibulevsky, Y.Y. Zeevi, Multiscale framework for blind separation of linearly mixed signals, J. Mach. Learning Res. 4 (2003) 1339–1363.
- [26] I. Kopriva, D. Seršić, MATLAB code for WP SDICA algorithm: <http://www.zesoi.fer.hr/~sdamir/icasb>.
- [27] I. Kopriva, Q. Du, H. Szu, W. Wasykiwskyj, Independent component analysis approach to image sharpening in the presence of atmospheric turbulence, Opt. Commun. 233 (2004) 7–14.
- [28] R.H. Lambert, C.L. Nikias, Blind Deconvolution of Multipath Mixtures, Chapter 9, in: S. Haykin (Ed.), Unsupervised Adaptive Filtering—Volume I Blind Source Separation, Wiley, New York, 2004.
- [29] Y. Li, A. Cichocki, S. Amari, Analysis of sparse representation and blind source separation, Neural Comput 16 (2004) 1193–1234.
- [30] Y. Li, S. Amari, A. Cichocki, D.W.C. Ho, S. Xie, Underdetermined blind source separation based on sparse representation, IEEE Trans. Signal Process. 54 (2006) 423–437.
- [31] S. Mallat, A Wavelet Tour of Signal Processing, Academic Press, New York, 1999.
- [32] P. McCullagh, Tensor Methods in Statistics, Chapman & Hall, London, 1995.
- [33] M. McKeown, S. Makeig, G. Brown, T.P. Jung, S. Kinderman, T.W. Lee, T.J. Sejnowski, Spatially independent activity patterns in functional magnetic resonance imaging data during the stroop color-naming task, Proc. Natl. Acad. Sci. 95 (1998) 803–810.
- [34] D. Nuzillard, A. Bijaoui, Blind source separation and analysis of multispectral astronomical images, Astron Astrophys Suppl Ser 147 (2000) 129–138.
- [35] J.C. Principe, D. Xu, J.W. Fisher, Information-theoretic learning. In: Unsupervised Adaptive Filtering—Volume I Blind Source Separation. Wiley, New York, 2000 (Chapter 7).
- [36] T. Ristaniemi, J. Joutsalo, Advanced ICA-based receivers for block fading DS-CDMA channels, Signal Process 85 (2002) 417–431.
- [37] B.W. Silverman, Density Estimation for Statistics and Data Analysis, Chapman & Hall, London, 1986.
- [38] T. Tanaka, A. Cichocki, Subband decomposition independent component analysis and new performance criteria, in: Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2004), Montreal, Canada, 2004, pp. 541–544.
- [39] M.V. Wickerhauser, Adapted Wavelet Analysis from Theory to Software, A.K. Peters, 1994.
- [40] K. Zhang, L.W. Chan, An adaptive method for subband decomposition ICA, Neural Comput 18 (2006) 191–223.
- [41] K. Zhang, L.W. Chan, Enhancement of source independence for blind source separation, Lecture Notes in Computer Science 3889 (2006) 731–738.
- [42] K. Zhang, L.W. Chan, MATLAB code for adaptive SDICA algorithm <http://appsrv.cse.cuhk.edu.hk/~kzhang/research.html>.
- [43] L. Zhang, A. Cichocki, S. Amari, Self-adaptive blind source separation based on activation function adaptation, IEEE Trans. Neural Networks 15 (2004) 233–244.
- [44] M. Zibulevsky, P. Kisilev, Y.Y. Zeevi, B. Pearlmutter, Blind source separation via multinode sparse representation, Advances in Neural Information Processing Systems, Morgan Kaufman, Los Altos, CA, 2002, pp. 185–191.



Ivica Kopriva received the B.S. degree in electrical engineering from Military Technical Faculty, Zagreb, Croatia in 1987, and M.S. and Ph.D. degrees in electrical engineering from the Faculty of Electrical Engineering and Computing, Zagreb, Croatia in 1990 and 1998, respectively. Currently, he is senior scientist at the Rudjer Bošković Institute, Zagreb, Croatia. His current research activities are related to the algorithms for blind signal and image processing and non-negative matrix factorization with applications in

unsupervised segmentation of multispectral and hyperspectral images. Prior to joining Rudjer Bošković Institute Dr Kopriva spent 4 years, 2001–2005, at the George Washington University, working on theory and applications of blind signal processing in imaging as well as on direction finding systems.



Damir Seršić received his diploma degree, as well as M.S. and Ph.D. in Electrical Engineering degrees from University of Zagreb, Faculty of Electrical Engineering and Computing in 1986, 1993 and 1999 respectively. Since 1987 he has been with the Department of Electronic Systems and Information Processing, Faculty of Electrical Engineering and Computing, University of Zagreb, Croatia, where he is currently assistant professor. His research interests are in the theory and applications of multiresolution analysis.