

Multichannel Blind Source Separation Using Convolution Kernel Compensation

Aleš Holobar, *Member, IEEE*, and Damjan Zazula, *Senior Member, IEEE*

Abstract—This paper studies a novel decomposition technique, suitable for blind separation of linear mixtures of signals comprising finite-length symbols. The observed symbols are first modeled as channel responses in a multiple-input–multiple-output (MIMO) model, while the channel inputs are conceptually considered sparse positive pulse trains carrying the information about the symbol arising times. Our decomposition approach compensates channel responses and aims at reconstructing the input pulse trains directly. The algorithm is derived first for the overdetermined noiseless MIMO case. A generalized scheme is then provided for the underdetermined mixtures in noisy environments. Although blind, the proposed technique approaches Bayesian optimal linear minimum mean square error estimator and is, hence, significantly noise resistant. The results of simulation tests prove it can be applied to considerably underdetermined convolutive mixtures and even to the mixtures of moderately correlated input pulse trains, with their cross-correlation up to 10% of its maximum possible value.

Index Terms—Blind source separation (BSS), convolution, convolution kernel compensation (CKC), multiple-input–multiple-output–(MIMO) systems, sparse signals.

I. INTRODUCTION

BLIND source separation (BSS) is becoming an increasingly important tool. Over the last decade, it has been successfully applied to the areas of radar, audio processing, telecommunications, separation of seismic signals, image processing, and to the analysis of biomedical data [1, pp. 391–448].

Recently, separation of sparse time series has gained a lot of attention [2], [3]. Assuming the source signals have a sparse representation on a given basis, the proposed methods utilize maximum likelihood (ML) estimators in order to iteratively learn both the mixing matrix and the source signals out of the observed data. They provide reasonably good results, also for the underdetermined mixtures. However, by artificially dividing signals into short blocks and by employing the basis functions of popular coding transformations (such as discrete cosine transform or wavelet transform) they completely ignore the time localizations of the underlying signal structures.

Manuscript received May 25, 2006; revised December 15, 2006. This work was supported in part by the Slovenian Ministry of Higher Education, Science and Technology, by the Lagrange Project of CRT Foundation, Italy, and by the Marie Curie Intra-European Fellowship within the 6th European Community Framework Programme. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Simon J. Godsill.

A. Holobar was with the Faculty of Electrical Engineering and Computer Science, University of Maribor, 2000 Maribor, Slovenia. He is now with the Politecnico di Torino, Torino, Italy (e-mail: ales.holobar@uni-mb.si).

D. Zazula is with the Faculty of Electrical Engineering and Computer Science, University of Maribor, Slovenia (e-mail: zazula@uni-mb.si).

Digital Object Identifier 10.1109/TSP.2007.896108

A more general single-channel generative model, in which the signal is described as a linear combination of shift-invariant basis functions was proposed by Lewicki and Sejnowski [4] and Olshausen [5]. Their methods inherently capture the best temporal positions of the predefined basis functions and preserve the information about the temporal structure of the signals. However, they focus on predefined basis functions only. The idea of shift-invariant basis functions was further extended by Blumenshath *et al.* [6], Wersing *et al.* [7], and Jost *et al.* [8], who proposed an iterative learning of fundamental signal structures considered as signal-specific basis functions. But they constrained their search to the set of uncorrelated basis functions.

In this paper, a different approach to sparse identification of shift-invariant signal components is presented as we address the multichannel linear mixtures of finite-length symbols (correlated or not). As explained in the next section, such observations can always be modeled as convolutive mixtures of sparse pulse trains, which carry information about the arising times of the detected symbols, and the symbols themselves. Instead of directly estimating the symbols, i.e., convolution kernels, we rather focus on the properties of triggering sparse trains. More precisely, we combine their spatial and temporal statistics with the information about their overlapping probability in order to blindly reconstruct their pulse sequences. The shapes of the observed symbols are lost during the decomposition, but can always be recovered by a phase-locked averaging of observations [9]. Throughout this manuscript, we do not assume any prior probability density function of the modeled pulse trains. We do, however, assume these pulse trains (at most) weakly correlated and sufficiently sparse, so that the subsequent repetitions of the same symbol are unlikely to overlap.

The assumed decomposition background is not completely new. The derived estimator shares almost identical functional form with the computationally attractive linear minimum mean square error (LMMSE) estimator, which is Bayesian optimal for linear mixing systems [10, pp. 379–418]. However, the LMMSE estimator supposes the first two moments of the source signals, i.e., their mean and their cross-correlation with the observations, are known in advance. A supervised way to overcome this problem was already proposed for interference suppression in the direct sequence (DS) code division multiple access (CDMA) receivers [11]. It starts by a transmission of predefined sequences of test signals that comprise known source symbols. These test signals are then used on the receiver side to determine the requested source statistics. Also more sophisticated estimation techniques were developed, such as the ML estimator described in [12, pp. 233–257]. But the ML approach suffers from high computational load and requires at least partial knowledge

of the mixing matrix [11]. Our approach upgrades the aforementioned decomposition techniques by iteratively improving the unknown source filter.

The paper is organized in six sections. In Section II, the assumed convolutive data model is derived in the form required by our decomposition approach, which is revealed in Sections III and IV. Section V presents numerical results obtained by the decomposition of synthetic mixtures, while a preliminary application of this method to the real surface electromyograms and electrocardiograms have been described in [13] and [14], respectively. Section VI concludes the paper.

Throughout this paper, boldface uppercase letters denote matrices, boldface lowercase letters denote vectors, while italics denote scalars. Discrete time series are denoted by the subscripted boldface lowercase letters, e.g., $\mathbf{x}_i = \{x_i(n); n = 0, 1, \dots\}$, where $x_i(n)$ denotes a single (the n th) sample. The vector of samples taken from M time series at the n th time instant is denoted by $\mathbf{x}(n) = [x_1(n), \dots, x_M(n)]^T$. The superscript $(\cdot)^T$ stands for transpose, while $^{-1}$ and $\#$ denote the matrix inverse and Moore-Penrose pseudoinverse, respectively.

II. DATA MODEL AND DECOMPOSITION BACKGROUND

Assume M different discrete-time observations $\mathbf{x}_i = \{x_i(n); n = 0, 1, 2, \dots\}; i = 1, \dots, M$ given, each comprising mixtures of up to N different L samples long symbols

$$x_i(n) = \sum_{j=1}^N \sum_{l=0}^{L-1} a_{ij}(l)t_j(n-l); \quad i = 1, \dots, M \quad (1)$$

where $\mathbf{a}_{ij} = \{a_{ij}(l); l = 0, 1, \dots, L-1\}$ stands for the j th symbol, as appearing in the i th observation, $\mathbf{t}_j = \{t_j(n); n = 0, 1, 2, \dots\}$ with $t_j(n) = \sum_k \delta(n - T_j(k))$ is the pulse train whose pulse at time $T_j(k)$ indicates the k th repetition of the j th symbol, and $\delta(\cdot)$ stands for the Dirac impulse. We additionally assume, the subsequent repetitions of the symbol are at least few samples apart, while $\forall j : E[|T_j(k) - T_j(k+1)|] > L$, where $E[\cdot]$ is mathematical expectation. When noisy observations are considered, (1) extends to

$$y_i(n) = x_i(n) + \omega_i(n) \quad (2)$$

where the additive noise $\omega_i = \{\omega_i(n); n = 0, 1, 2, \dots\}$ is commonly modeled as a stationary, temporally and spatially white zero-mean Gaussian random process.

The aforementioned multiple-input-multiple-output (MIMO) data model corresponds to many real world situations and can be applied whenever the observations can be individually interpreted as linear mixtures of separate signal components. A large area of application, which has been under intense investigation, is most certainly the field of biomedical signals. In the

case of electromyograms, for instance, each symbol \mathbf{a}_{ij} corresponds to an action potential of the j th motor unit as detected by the i th uptake electrode [15]. Similar interpretation can be found in the case of electrocardiogram, electroneurogram, and even in the case of electroencephalogram [16]. Another very popular field of application are digital communication systems. For instance, when employing DS-CDMA coding technique, the symbols \mathbf{a}_{ij} correspond to transmitted information bits modified by the users' spreading codes and by the transfer channel responses [12, pp. 849–861]. Further applications include speech recognition, audio separation, stereo image processing, etc.

A. Extension to the Matrix Form

The convolutive relationship described in (1) can always be expressed in the matrix form [17]. First, the vector of samples related to the n th time instant in M observations $\mathbf{y}(n) = [y_1(n), \dots, y_M(n)]^T$ is extended by $K-1$ delayed repetitions of each observation to comprise blocks of K samples for each observation

$$\bar{\mathbf{y}}(n) = [y_1(n), y_1(n-1), \dots, y_1(n-K+1), \dots, y_M(n), \dots, y_M(n-K+1)]^T. \quad (3)$$

Extending the noise vector $\boldsymbol{\omega}(n) = [\omega_1(n), \dots, \omega_M(n)]^T$ in the same way, (1) evolves to

$$\bar{\mathbf{y}}(n) = \mathbf{A}\bar{\mathbf{t}}(n) + \bar{\boldsymbol{\omega}}(n) \quad (4)$$

where the vector

$$\bar{\mathbf{t}}(n) = [t_1(n), t_1(n-1), \dots, t_1(n-L-K+2), \dots, t_N(n), \dots, t_N(n-L-K+2)]^T \quad (5)$$

is an extended form of vector $\mathbf{t}(n) = [t_1(n), \dots, t_N(n)]^T$, which contains a block of $(L+K-1)$ consecutive samples of each \mathbf{t}_j , and \mathbf{A} stands for the $KM \times N(L+K-1)$ mixing matrix

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_{11} & \dots & \mathbf{A}_{1N} \\ \vdots & \ddots & \vdots \\ \mathbf{A}_{M1} & \dots & \mathbf{A}_{MN} \end{bmatrix} \quad (6)$$

which contains the detected symbols \mathbf{a}_{ij} enclosed in $K \times (L+K-1)$ convolution kernels, shown in (7) at the bottom of the page. In the sequel, the j th element of vector $\bar{\mathbf{t}}(n)$ will be denoted by $\bar{t}_j(n)$, while the sequence $\bar{\mathbf{t}}_j = \{\bar{t}_j(n); n = 0, 1, \dots\}$ will be referred to as the j th extended pulse train.

Assuming the detected symbols are generated in random time instants, the pulse trains can be modeled as uncorrelated random pulse sequences. Moreover, when sampling frequency is high

$$\mathbf{A}_{ij} = \begin{bmatrix} a_{ij}(0) & a_{ij}(1) & \dots & a_{ij}(L-1) & 0 & \dots \\ 0 & a_{ij}(0) & \dots & a_{ij}(L-2) & a_{ij}(L-1) & \dots \\ \vdots & & \ddots & \ddots & \ddots & \ddots \\ 0 & \dots & 0 & a_{ij}(0) & \dots & a_{ij}(L-1) \end{bmatrix}. \quad (7)$$

enough (with respect to the symbol rate), the pulse trains become highly sparse. The decomposition approach described in this paper builds on these assumptions. In particular, the information contained in the mixing matrix \mathbf{A} is ignored as we try to compensate the convolution kernels \mathbf{A}_{ij} and focus strictly on the properties of the pulse trains $\bar{\mathbf{t}}_j$.

To make it more comprehensible, the description of the proposed convolution kernel compensation (CKC) approach is divided into two sections. Section III reveals the decomposition under ideal circumstances, i.e., when the assumed MIMO system is overdetermined and noise-free. The underdetermined case and the influence of noise are then discussed in Section IV.

III. CONVOLUTION KERNEL COMPENSATION IN A NOISE-FREE OVERDETERMINED CASE

Suppose the number of symbols N is smaller than the number of observations M , and that the extended pulse trains $\bar{\mathbf{t}}_j$ are weakly correlated, i.e., they have small, but significant number of overlapping pulses. In addition, assume the observations \mathbf{x}_i are ergodic and denote by $\mathbf{C}_{\bar{\mathbf{x}}\bar{\mathbf{x}}} = E[\bar{\mathbf{x}}(n)\bar{\mathbf{x}}^T(n)]$ the correlation matrix of extended observations. Finally, suppose the extension factor K is large enough to guarantee $KM \geq N(K + L - 1)$, while the matrix \mathbf{A} is of full column rank. Then, by calculating the square of Mahalanobis distance for vector $\bar{\mathbf{x}}(n)$, the convolution kernels \mathbf{A}_{ij} are compensated yielding a so-called activity index

$$\begin{aligned} \gamma(n) &= \bar{\mathbf{x}}^T(n) \mathbf{C}_{\bar{\mathbf{x}}\bar{\mathbf{x}}}^{-1} \bar{\mathbf{x}}(n) \\ &= \bar{\mathbf{t}}^T(n) \mathbf{A}^T (\mathbf{A}^T)^{-1} \mathbf{C}_{\bar{\mathbf{t}}\bar{\mathbf{t}}}^{-1} \mathbf{A}^{-1} \mathbf{A} \bar{\mathbf{t}}(n) \\ &= \bar{\mathbf{t}}^T(n) \mathbf{C}_{\bar{\mathbf{t}}\bar{\mathbf{t}}}^{-1} \bar{\mathbf{t}}(n) \end{aligned} \quad (8)$$

where $\mathbf{C}_{\bar{\mathbf{t}}\bar{\mathbf{t}}}$ stands for correlation matrix of $\bar{\mathbf{t}}(n)$. The activity index $\boldsymbol{\gamma} = \{\gamma(n); n = 0, 1, \dots\}$ can be thought of as an indicator of global pulse train activity. Being always positive, it differs from zero only at those time instants n_i where at least one extended pulse train is active, i.e., $\gamma(n_k) > 0 \Leftrightarrow \exists j; \bar{t}_j(n_k) = 1$.

Suppose there are $g_0 > 1$ trains active in the observed time instant n_0 and denote them by the set of indices $G_{n_0} = \{j_{n_0,1}, \dots, j_{n_0,g_0}\}$, i.e., $j \in G_{n_0} \Leftrightarrow \bar{t}_j(n_0) = 1$. By using a premultiplying vector $\bar{\mathbf{x}}^T(n_0)$ instead of $\bar{\mathbf{x}}^T(n)$ in (8), we obtain the following linear combination of extended pulse trains:

$$\begin{aligned} \nu_{n_0}(n) &= \bar{\mathbf{x}}^T(n_0) \mathbf{C}_{\bar{\mathbf{x}}\bar{\mathbf{x}}}^{-1} \bar{\mathbf{x}}(n) = \bar{\mathbf{t}}^T(n_0) \mathbf{C}_{\bar{\mathbf{t}}\bar{\mathbf{t}}}^{-1} \bar{\mathbf{t}}(n) \\ &= \sum_{k \in G_{n_0}} \sum_{j=1}^{N(L+K-1)} c_{jk}^{\text{inv}} \bar{t}_j(n) \end{aligned} \quad (9)$$

where c_{jk}^{inv} stands for the (j, k) th element of matrix $\mathbf{C}_{\bar{\mathbf{t}}\bar{\mathbf{t}}}^{-1}$.

Now, assume the correlation matrix $\mathbf{C}_{\bar{\mathbf{t}}\bar{\mathbf{t}}}$ is diagonally dominant. As proven in Appendix A, its inverse has a superior diagonal, while all the off-diagonal elements are much smaller than the diagonal ones. This implies the pulse sequence $\mathbf{v}_{n_0} = \{\nu_{n_0}(n); n = 0, 1, 2, \dots\}$ has strong contributions only from the pulse trains that are contained in G_{n_0} , while the contributions from all other trains are much smaller in amplitude. As a result

$$\nu_{n_0}(n) \approx \sum_{j \in G_{n_0}} c_{jj}^{\text{inv}} \bar{t}_j(n). \quad (10)$$

According to (10), the entire train $\bar{\mathbf{t}}_j$ can be reconstructed, providing we have found a time instant n_0 with a contribution from that train only. However, finding such a time instant is not a trivial task. Moreover, the probability of finding nonoverlapping pulses decreases with the number of extended pulse trains $N(K + L - 1)$. Hence, more formal procedure for separation of superimposed trains (10) is needed.

A. Separation of Superimposed Pulse Trains

Suppose all possible overlaps of pulses from the j th and the k th extended pulse train are independent, equally probable random events and denote their probability by $p_{j,k} = P(\bar{t}_j(n) = 1 | \bar{t}_k(n) = 1)$. Define $p = \max_{j \neq k} (p_{j,k})$ and note that p is, at least in the case of weakly correlated pulse trains, relatively close to zero ($p \ll 1$). Use (10) to reconstruct the pulse sequence $\mathbf{v}_{n_0} = \{\nu_{n_0}(n); n = 0, 1, 2, \dots\}$, randomly select a pulse in it and denote by n_1 the time of its occurrence. Now, compute \mathbf{v}_{n_1} and denote by $G_{n_1} = \{j_{n_1,1}, \dots, j_{n_1,g_1}\}$ the set of indices of all pulse trains with pulses at n_1 , i.e., $j \in G_{n_1} \Leftrightarrow \bar{t}_j(n_1) = 1$. Because the instant n_1 was chosen according to \mathbf{v}_{n_0} , there is at least one train active in both instants n_0 and n_1 , i.e., $\exists j; \bar{t}_j(n_0) = 1 \wedge \bar{t}_j(n_1) = 1$. Generate a sequence of element-wise products $\mathbf{h}_{n_0, n_1} = \{\nu_{n_0}(n) \cdot \nu_{n_1}(n); n = 0, 1, 2, \dots\}$, randomly select $R - 2$ pulses in it and denote by $n_r; r = 2, \dots, R - 1$ their arising times. For each $n_r; r = 2, \dots, R - 1$ there are two complementary explanations.

Assumption 1: At least in one train from $G_{n_0} \cap G_{n_1}$, a pulse appears at n_r .

Assumption 2: There is no train in $G_{n_0} \cap G_{n_1}$ with a pulse at n_r , but at least two different trains have pulses at n_r . The first is contained in set difference $G_{n_0} - G_{n_1}$, the second in $G_{n_1} - G_{n_0}$.

For each instant $n_r; r = 2, \dots, R - 1$ generate a new sequences \mathbf{v}_{n_r} , denote by $G_{n_r} = \{j_{n_r,1}, \dots, j_{n_r,g_r}\}$ the set of pulse trains active in n_r , and observe the number of pulses in element-wise product

$$\mathbf{h}_{n_0, \dots, n_{R-1}} = \left\{ \prod_{r=0}^{R-1} \nu_{n_r}(n); n = 0, 1, 2, \dots \right\}. \quad (11)$$

The probability that g trains are active in all R selected time instants can be estimated as

$$P \left(\text{card} \left(\bigcap_{r=0}^{R-1} G_{n_r} \right) = g \right) \leq p^{(R-1)(g-1)} \quad (12)$$

where $\text{card}(G_{n_r})$ denotes the cardinal number of G_{n_r} . For the sake of simplicity, let $R = 4$ and assume that the time instant n_2 fulfils Assumption 1, while n_3 fulfils Assumption 2. Then the intersection $G_{n_0} \cap G_{n_1} \cap G_{n_2}$ contains at least one pulse train, while Assumption 2 guarantees that $G_{n_0} \cap G_{n_1} \cap G_{n_3}$ is an empty set. Using inequality (12), we quickly realize the probability of having more than two pulse trains in any intersection $G_{n_i} \cap G_{n_j}; i \neq j$ becomes negligible when $p \ll 1$. For the same reason, the intersections $G_{n_0} \cap G_{n_2} \cap G_{n_3}$ and $G_{n_1} \cap G_{n_2} \cap G_{n_3}$ are, with high probability, empty sets. As a result, the total number of pulses in the product (11) can be estimated as (Appendix B, Case 1)

$$v(p, F) \approx F p^2 N(K + L - 1) \quad (13)$$

where, for the sake of simplicity, we assumed the total number of pulses for all pulse trains equals F .

Now, assume that both selected time instants n_2 and n_3 fulfil Assumption 1, i.e., the intersection $G_{n_0} \cap G_{n_1} \cap G_{n_2} \cap G_{n_3}$ comprises at least one pulse train, while (12) guarantees the probability of having more than one train in this intersection is negligible. Following the assumptions from the case above, the number of pulses in the product (11) can be estimated as (Appendix B, Case 2)

$$v(p, F) \approx F(1 + 3p). \quad (14)$$

According to (13) and (14), the superimposed pulse trains can be separated by observing the number of pulses in the element-wise product (11), providing that $p^2 N(K + L - 1) \ll 1$. When this is not the case, the number of components R in (11) can be increased. Generally speaking, the higher the probability p , the higher the optimal value for R . However, by increasing the value of R the time complexity of the CKC method also increases. In all our experiments, $R = 4$ proved to be a good compromise. The exact procedure for separation of superimposed trains is described in Section IV.

IV. CONVOLUTION KERNEL COMPENSATION IN UNDERDETERMINED CASE

Now, assume the extended pulse trains are uncorrelated and $\mathbf{C}_{\bar{\mathbf{t}}\bar{\mathbf{t}}}$ equals the identity (note that this assumption is only made to simplify the theoretical derivations in this section). If the number of symbols N is greater than the number of observations M , the $KM \times N(L + K - 1)$ mixing matrix \mathbf{A} becomes rectangular and has more columns than rows. As shown in Appendix C, its influence cannot be completely compensated by (9). What remains is an orthogonal projector \mathbf{B} :

$$\nu_{n_0}(n) = \bar{\mathbf{t}}^T(n_0)\mathbf{B}\bar{\mathbf{t}}(n) \approx \sum_{j \in G_{n_0}} b_{jj}\bar{t}_j(n) \quad (15)$$

where b_{ij} denotes the (i, j) th element of matrix \mathbf{B} . We found experimentally that when \mathbf{A} is close to a square matrix, \mathbf{B} is close to the identity matrix. By increasing the ratio $N(L + K - 1)/KM$, \mathbf{B} gradually loses its diagonal form and the impact of the mixing matrix in (15) increases. Nevertheless, keeping the ratio $N(L + K - 1)/KM$ below 2, at least a part of the diagonal in \mathbf{B} remains dominant (Appendix C).

The pulse trains corresponding to the dominant diagonal elements of the matrix \mathbf{B} still comply with the theory from the overdetermined case, and the procedure derived in Section III can readily be applied to the underdetermined MIMO systems. This is further confirmed by comparing the LMMSE estimator to our approach. Namely, employing probabilistic separation of superimposed pulse trains (Section III), the triggering times of a single, say the j th, pulse train can be grouped together into a common subset of time instants $\Psi_j = \{n_k; \bar{t}_j(n_k) = 1\}$. Assuming the mixing process is ergodic, the observation vectors $\bar{\mathbf{y}}(n)$ can then be averaged over all time instants n_k from the set Ψ_j . The obtained mean vector yields the cross-correlation between the j th pulse train and all the observations

$$\mathbf{c}_{\bar{\mathbf{y}}\bar{\mathbf{t}}_j} = \lim_{\text{card}(\Psi_j) \rightarrow \infty} \frac{1}{\text{card}(\Psi_j)} \sum_{n_k \in \Psi_j} \bar{\mathbf{y}}(n_k). \quad (16)$$

By inserting (16) into (9), we obtain the LMMSE estimator of the j th pulse train [10, p. 382]

$$\hat{\bar{t}}_j(n) = \mathbf{c}_{\bar{\mathbf{y}}\bar{\mathbf{t}}_j}^T \mathbf{C}_{\bar{\mathbf{y}}\bar{\mathbf{y}}}^{-1} \bar{\mathbf{y}}(n). \quad (17)$$

A. Noise Reduction

With respect to (16), the influence of noise $\bar{\mathbf{w}}(n)$ in (9) can be expressed as:

$$\begin{aligned} o_{\bar{\mathbf{w}}, \Psi_j}(n) &= \frac{\sum_{k \in \Psi_j} \bar{\mathbf{w}}^T(k)}{\text{card}(\Psi_j)} (\mathbf{A}^T)^{\#} \mathbf{C}_{\bar{\mathbf{t}}\bar{\mathbf{t}}}^{-1} \bar{\mathbf{t}}(n) + \frac{\sum_{k \in \Psi_j} \bar{\mathbf{w}}^T(k)}{\text{card}(\Psi_j)} \\ &\cdot (\mathbf{A}^T)^{\#} \mathbf{C}_{\bar{\mathbf{t}}\bar{\mathbf{t}}}^{-1} \mathbf{A}^{\#} \bar{\mathbf{w}}(n) + \frac{\sum_{k \in \Psi_j} \bar{\mathbf{t}}^T(k)}{\text{card}(\Psi_j)} \mathbf{C}_{\bar{\mathbf{t}}\bar{\mathbf{t}}}^{-1} \mathbf{A}^{\#} \bar{\mathbf{w}}(n). \end{aligned} \quad (18)$$

While the first two right-hand side factors in (18) converge to zero when the vector $\mathbf{c}_{\bar{\mathbf{y}}\bar{\mathbf{t}}_j}$ is averaged over a large enough set Ψ_j , this averaging hardly changes the rightmost factor. Its impact can be reduced by truncating the eigenvalues of matrix $\mathbf{C}_{\bar{\mathbf{y}}\bar{\mathbf{y}}}$. Namely, the influence of noise projected to the space of input pulse trains can be estimated as [18, pp. 411–417]

$$\frac{1}{\kappa} \frac{\|\bar{\mathbf{w}}(n)\|}{\|\bar{\mathbf{x}}(n)\|} \leq \frac{\|\bar{\mathbf{w}}_{\bar{\mathbf{t}}}(n)\|}{\|\bar{\mathbf{t}}(n)\|} \leq \kappa \frac{\|\bar{\mathbf{w}}(n)\|}{\|\bar{\mathbf{x}}(n)\|} \quad (19)$$

where κ stands for the condition number of matrix \mathbf{A} , and $\bar{\mathbf{w}}_{\bar{\mathbf{t}}}(n) = \mathbf{A}^{\#} \bar{\mathbf{w}}(n)$ is the noise projection to the space of extended pulse trains $\bar{\mathbf{t}}_j$. By setting the smallest singular values of \mathbf{A} to zero, we improve its condition number κ and increase the robustness of our decomposition [18, p. 418].

Now, assume the matrix $\mathbf{C}_{\bar{\mathbf{t}}\bar{\mathbf{t}}}$ equals the identity matrix. Then the conditional number of \mathbf{A} can be controlled by truncating the eigenvalues of $\mathbf{C}_{\bar{\mathbf{y}}\bar{\mathbf{y}}} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^T$, i.e., by setting the Q smallest eigenvalues to zero

$$\hat{\mathbf{\Lambda}} = \mathbf{\Lambda} \cdot \begin{bmatrix} \mathbf{I}_{KM-Q, KM-Q} & \mathbf{0}_{KM-Q, Q} \\ \mathbf{0}_{Q, KM-Q} & \mathbf{0}_{Q, Q} \end{bmatrix} \quad (20)$$

where $\mathbf{\Lambda}$ stands for the diagonal matrix with eigenvalues of $\mathbf{C}_{\bar{\mathbf{y}}\bar{\mathbf{y}}}$, sorted in descending order, \mathbf{U} is a matrix of corresponding eigenvectors, and $\mathbf{0}_{Q, K}$ denotes the $Q \times K$ matrix with all elements equal to zero. Afterwards, a new correlation matrix $\hat{\mathbf{C}}_{\bar{\mathbf{y}}\bar{\mathbf{y}}} = \mathbf{U}\hat{\mathbf{\Lambda}}\mathbf{U}^T$ is constructed and used in (17) in place of $\mathbf{C}_{\bar{\mathbf{y}}\bar{\mathbf{y}}}$. The optimal degree of eigenvalue truncation depends on the number of symbols, N , and the signal-to-noise ratio (SNR), and will be further clarified in Section V.

The final CKC decomposition procedure is described in Fig. 1. The noise variance σ^2 in step 1 can be estimated by observing the smallest eigenvalues of the matrix $\mathbf{C}_{\bar{\mathbf{y}}\bar{\mathbf{y}}}$ [1, p. 129], whereas the threshold J in step 5 can be computed as a product of the observed signal length and the lowest expected symbol rate.

V. SIMULATION RESULTS

The proposed decomposition algorithm was tested on three different sets of synthetic signals. The first experiment evaluated the influence of number of observed symbols, the second experiment studied the influence of pulse overlapping probability p , while in the third experiment the decomposition of

- 1: Using (8) compute γ . Estimate noise variance σ^2 and determine noise threshold $d_{\sigma^2} = \sigma^2 \left\| \hat{\mathbf{C}}_{\mathbf{y}\mathbf{y}}^{-1} \right\|_1$, where $\|\cdot\|_1$ denotes matrix 1-norm. Set $\forall n_i; \gamma(n_i) < d_{\sigma^2} \Rightarrow \gamma(n_i) = 0$.
- 2: Set $n_0 = \underset{\text{arg}}{\text{median}}(\gamma)$ and use (9) to calculate \mathbf{v}_{n_0} . Randomly choose one of its pulses exceeding the noise threshold d_{σ^2} , denote by n_1 its time of occurrence and reconstruct \mathbf{v}_{n_1} .
- 3: Collect all n_r for which $v_{n_0}(n_r) \cdot v_{n_1}(n_r) > d_{\sigma^2}$ into a set Υ_{n_0, n_1} .
- 4: For every $n_r \in \Upsilon_{n_0, n_1}$ reconstruct \mathbf{v}_{n_r} .
- 5: For all possible combinations $(n_3, n_4, \dots, n_{R-1})$, where $n_3, \dots, n_{R-1} \in \Upsilon_{n_0, n_1}$ determine the number of pulses in the product (11). Any combination $(n_3, n_4, \dots, n_{R-1})$ contributing the number of pulses exceeding a predefined threshold, J , is added to a common set Ψ_j .
- 6: Use Eqs. (16) and (17) to estimate the entire j -th pulse train.
- 7: Compare the obtained train to all previous reconstructed pulse sequences and classify it either as a new pulse sequence or a previously detected one.
- 8: Set $\forall n_k \in \Psi_j \Rightarrow \gamma(n_k) = 0$.
- 9: Repeat steps 2 to 9 until all the values of the activity index γ_i are zero.

Fig. 1. Pseudocode of the proposed CKC decomposition approach.

ill-conditioned linear mixtures was investigated. In all three experiments, three different performance measures were observed: the number of reconstructed pulse trains, sensitivity of decomposition algorithm, i.e., the percentage of accurately identified pulses per pulse train, and false alarms, i.e., the percentage of false pulses per reconstructed pulse train. The proposed decomposition approach was additionally compared to LMMSE estimator. Recall that the LMMSE estimator supposes the cross-correlation vector $\mathbf{c}_{\mathbf{y}\mathbf{t}_j}$ in (17) known in advanced. This means the comparison of the decomposition results obtained by LMMSE and our CKC is feasible just in simulated cases, while in real situations only the separation by CKC approach is implementable.

A. Experiment 1: The Influence of Number of Symbols

The first experiment evaluated the CKC performance in dependence of the number of symbols N . Fifteen simulation runs were performed, with N set equal to 10, 20, and 30 (five runs per each N). In each run, random input pulse trains, $t_j(n) = \sum_{k=1}^{200} \delta(n - k \cdot 50 + T_j(k))$, were generated with the mean interpulse interval (IPI) set equal to 50 samples and the values $T_j(k); k = 1, 2, \dots, 200$, uniformly distributed over the interval $[-\tau, \tau]$. In this way, the probability of overlapped pulses, p , was estimated according to $p \leq (2\tau)^{-1}$. Parameter τ was fixed to 20 samples, yielding pulse overlapping probability of $p \leq 0.025$. The length of simulated pulse trains was set equal to 10 000 samples. Random zero-mean symbols \mathbf{a}_{ij} of length of $L = 10$ samples were generated and convolved with the simulated pulse

trains to produce the observed signals (1). Note that the repetitions of the same symbol in each observation did not interfere. The number of observations M was set equal to 25. Finally, five realizations of noise per each simulation run and each SNR were simulated, resulting in 375 different mixtures. These mixtures were then extended by the empirically selected factor $K = 10$ for $N = 10$ and $K = 20$ for $N = 20$ and $N = 30$. As a result, the number of extended pulse trains increased to 190 for $N = 10$, 580 for $N = 20$, and 870 for $N = 30$, while the number of extended observations was fixed at 250 for $N = 10$, and at 500 for $N = 20$ and $N = 30$, respectively.

Each mixture was decomposed three times, with the degree of eigenvalue truncation T_λ set equal to 0%, 20%, and 40%, respectively. Here $T_\lambda = 0\%$ means no truncation at all, while $T_\lambda = 40\%$ indicates that, according to (20), 40% of the smallest eigenvalues of $\mathbf{C}_{\mathbf{y}\mathbf{y}}$ were set equal to zero. The results, averaged over all simulation runs, are reported in Figs. 2–4, respectively.

In the case of $N = 10$, almost all pulse trains were identified by both estimation techniques (CKC and LMMSE) down to SNR of 0 dB. For $N = 20$ and SNR down to 5 dB, the average percentage of pulse trains identified by the CKC method dropped to 65%, and further decreased to approx. 55% when dealing with 30 symbols ($N = 30$). This agrees perfectly with the theoretical expectations presented in Appendix C. Namely, both cases correspond to the underdetermined convolutive mixtures, with the ratio $N(L+K-1)/KM$ set at 1.17 and 1.74, respectively. This is also reflected in the performance of LMMSE estimator, which reconstructed only approx. 5% of input trains

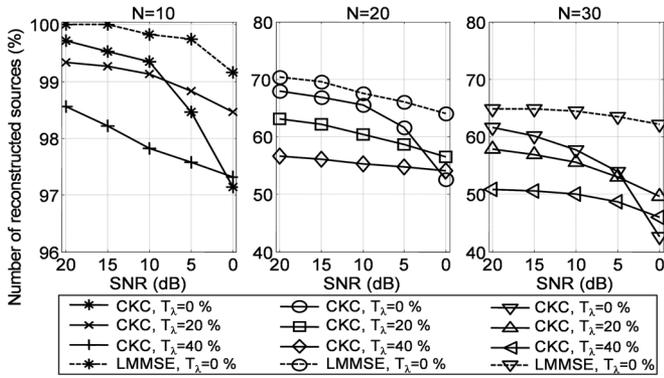


Fig. 2. Average number of recognised pulse trains depending on the number of symbols, N , the level of eigenvalue truncation (T_λ) and SNR. The results are normalized with respect to the number of symbols N . Note the different scale on the ordinate.

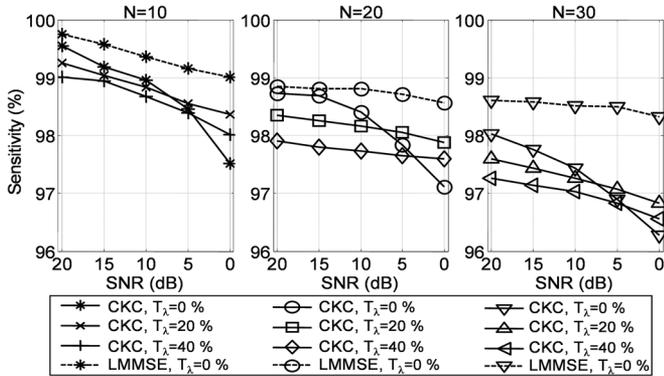


Fig. 3. The average number of accurately identified pulses per reconstructed pulse train (sensitivity) depending on the number of symbols, N , the level of eigenvalue truncation (T_λ) and SNR.

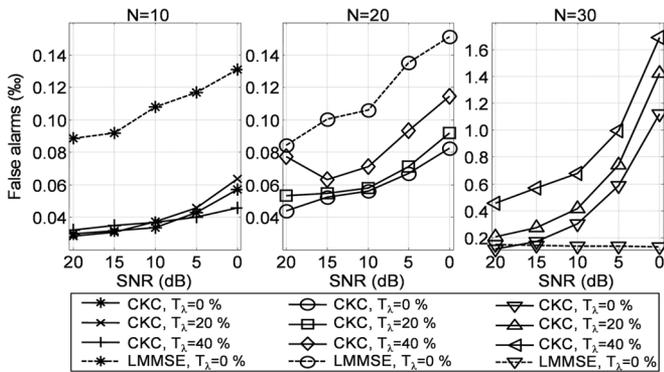


Fig. 4. The average percentage of incorrectly identified pulses per reconstructed pulse train (false alarms) depending on the number of symbols N , the level of eigenvalue truncation (T_λ), and SNR. Note the different scale on the ordinate.

more than the CKC method. However, even with a highly under-determined system, the eigenvalue truncation proved to be beneficial, increasing the number of reconstructed trains at SNR = 0 dB by more than 5%, on average. The optimal degree of truncation depends strongly on the level of noise. For example, for $N = 20$ the optimal value of T_λ yielded 0% at SNR = 15 dB and increased to 20% at SNR = 0 dB.

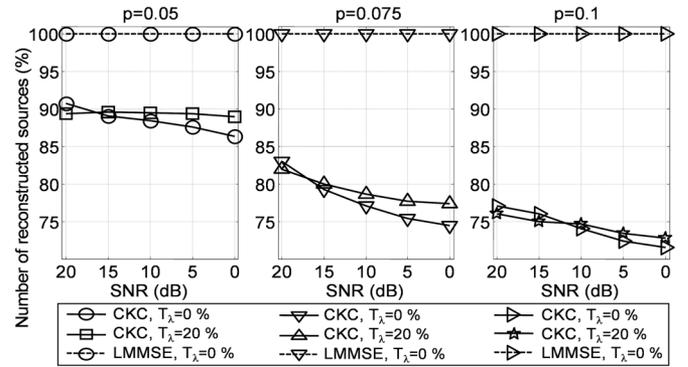


Fig. 5. The average number of reconstructed pulse trains for different values of overlapping probability p versus SNR. The number of symbols was set equal to $N = 10$.

The CKC method proved to be highly robust. On average, more than 97% of reconstructed pulses were accurately recognized down to the SNR of 5 dB, while there was hardly any misplaced pulse (Fig. 4). This is consistent with the theoretical derivation in (16), where it was shown that the accuracy of the CKC method matches the accuracy of the LMMSE estimator. In the case of $N = 20$ and $N = 30$, however, a slight decrease in CKC performance was noticed at SNR of 0 dB. The results in Fig. 3 also demonstrate a significant positive correlation between the sensitivity of CKC method and the degree of eigenvalue truncation. Separating the mixtures with $N = 20$ symbols at SNR of 0 dB using the 0% of eigenvalue truncation, for example, the sensitivity of CKC method dropped to 97%. Increasing the degree of truncation to 20%, the sensitivity index increased back to 98%. At the same time, there was a slight increase of false alarms, indicating a possible negative correlation between the specificity of the CKC algorithm and the eigenvalue truncation.

B. Experiment 2: The Influence of Level of Correlation Among Pulse Trains

The second experiment studied the influence of pulse overlapping probability p . Both the number of symbols, N , and the extension factor, K , were fixed at 10. The number of observations M was set equal to 25. Following the simulation protocol from the first experiment, 10 000 samples long random pulse trains were generated. The mean IPI was set at 50 samples, while the IPI variability was uniformly distributed over the interval $[-\tau, \tau]$. In this experiment, parameter τ was set at 10, 6.6, and 5, yielding pulse overlapping probability of $p \leq 0.05$, $p \leq 0.075$ and $p \leq 0.1$, respectively. Finally, random zero-mean symbols $\mathbf{a}_{i,j}$ of length $L = 10$ samples were convolved with the simulated pulse trains to produce the observed signals (1). The observations were additionally corrupted by additive zero-mean Gaussian noise (five realisations of noise per each SNR). The results, averaged over five simulation runs per each p , are depicted in Figs. 5–7.

As expected, the performance of the CKC method drops significantly with the overlapping probability p . Each time p is increased by 0.025, the percentage of reconstructed pulse trains drops by approx. 8%. At $p = 0.1$, only 75% of all the pulse

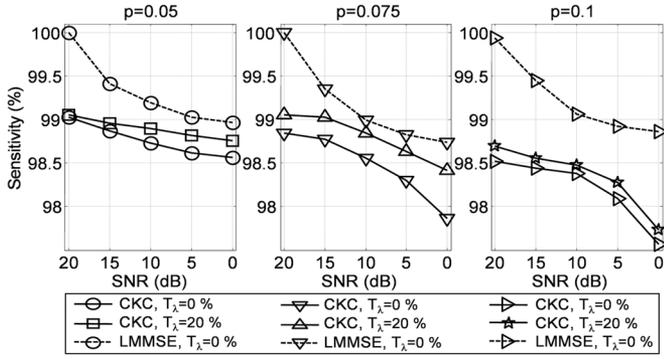


Fig. 6. The average number of accurately identified pulses per reconstructed pulse train (sensitivity) for different values of pulse overlapping probability p versus SNR. The number of symbols N was set equal to 10.

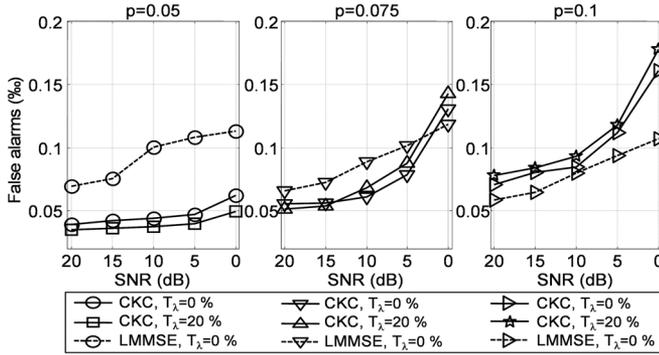


Fig. 7. False alarm index for different values of pulse overlapping probability, p , versus SNR. The number of symbols, N , was set equal to 10.

trains were reconstructed (with SNR of 10 dB and $T_\lambda = 20\%$). Also the sensitivity decreased, while the false alarms increased significantly (from 0.004% at $p = 0.025$ to 0.009% at $p = 0.1$ for SNR of 10 dB). There are several possible explanations of this phenomenon. The first, and most probable one, relies on the probabilistic separation of the superimposed pulse trains (Section III). By increasing the probability p the beneficial averaging effect in (16) decreases. As a result, the impact of both noise and superimposed pulses is increased. This also agrees with the observed increase of the CKC sensitivity to noise. In addition, high overlapping probability p causes the correlation matrix \mathbf{C}_{ff} (and, hence, its inverse) to become significantly non-diagonal, which additionally increases the devastating impact of the superimposed pulses in the reconstructed pulse trains. At the same time, no significant correlation between the probability p and the LMMSE estimation was noticed (except maybe in false alarm index). This was expected because the LMMSE estimator utilizes prior knowledge on the correlation between the observations and input pulse trains.

C. Experiment 3: Decomposition of Ill-Conditioned Mixtures

The final experiment was conducted on a simulated surface electromyograms (SEMG), based on a planar volume conductor model [19]. The thickness of the skin, fat and muscle layer was set equal to 1, 3, and 10 mm, respectively. Imitating the anatomy of skeletal muscles, synchronously active muscle fibers were first grouped into so called motor units (MUs) [20]. MUs were randomly scattered over the detection volume with their size

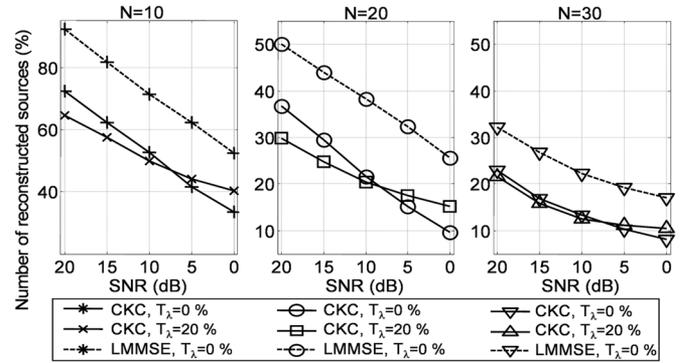


Fig. 8. Average number of recognised MUs depending on the number of simulated MUs, N , the level of eigenvalue truncation, T_λ , and SNR. The results are normalized with respect to the number of simulated MUs, N . Note the different scale on the ordinate.

varying from 25 to 2500 fibers. Average semifiber length was set equal to 70 mm. The mean muscle fiber conduction velocity was 4 ± 0.3 m/s. When activated, each MU transmitted detectable electric potentials. These potentials were additionally low-pass filtered to simulate the effect of biological tissues that separate the MU from the pick-up electrodes. Biopotentials, which correspond to the symbols \mathbf{a}_{ij} in our data model, were detected at the surface of the skin. Because of different MUs depths in the muscle tissue, the biopotentials differed significantly in amplitude, yielding the power ratio of up to 10 dB. The average length of all detected biopotentials \mathbf{a}_{ij} was 12 ms. The number of pick-up electrodes, M , was fixed at 60 (a 2-D electrode grid of 13×5 electrodes with interelectrode distance of 5 mm was simulated, while the signals were detected in longitudinal single differential configuration). Based on a MU recruitment model [21], three different muscle contraction levels were simulated corresponding to $N = 10$, $N = 20$, and $N = 30$ simultaneously active MUs. For each MU, a random sequence of innervation pulses was generated. The average motor unit discharge rate was set equal to 15.0 ± 4.0 pulses per second, while the average IPI variability equalled 20% of the mean IPI. Finally, SEMG observations were sampled at 1024 Hz. Before the decomposition took place, these observations were additionally corrupted by additive zero-mean Gaussian noise (five realizations of noise per each SNR) and extended by the factor $K = 10$ for $N = 10$ and $K = 20$ for $N = 20$ and $N = 30$. This resulted in a condition number of about 10^7 for the mixing matrix \mathbf{A} . The decomposition results, averaged over 15 simulation runs (5 runs per each N), are depicted in Figs. 8–10.

Both the CKC and LMMSE estimator exhibit a significant drop in the number of reconstructed pulse trains (when compared to the results of the first experiment). At $N = 10$ and SNR = 10 dB only 70% of simulated MUs were identified by the LMMSE estimator. Under the same conditions the CKC method reconstructed only a half of simulated MUs (compare this to almost 100% reconstruction in Experiment 1). This is not unexpected, as the decomposition of surface electromyograms is well known to be strongly ill-conditioned. In addition, due to large differences among the powers of different MUs, smaller MUs are often missed and considered a background noise. In spite of these facts, the reconstruction of pulse trains still proved

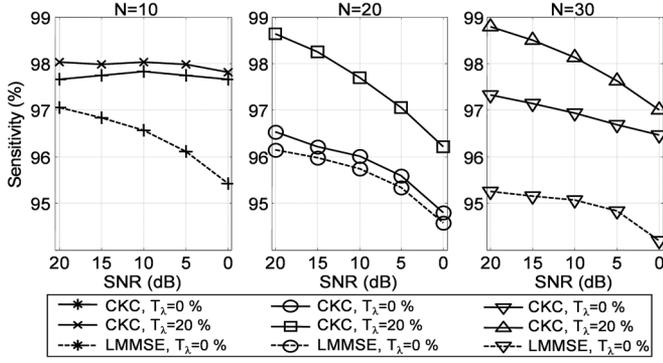


Fig. 9. The average number of accurately identified pulses per reconstructed MU (sensitivity) depending on the number of simulated MUs, N , the level of eigenvalue truncation T_λ , and SNR.

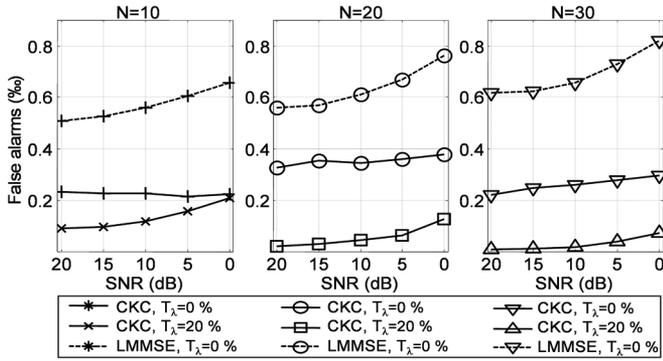


Fig. 10. The average percentage of incorrectly identified firings per reconstructed MU (false alarms) depending on the number of simulated MUs, N , the level of eigenvalue truncation T_λ , and SNR.

to be highly accurate. Slight decrease of the CKC sensitivity was noticed (compared to the results of Experiment 1). At the same time, a slight increase of false alarms was observed (for both the CKC and LMMSE estimator), but this had no significant influence on decomposition performance.

VI. CONCLUSION

The proposed CKC decomposition method can be applied to a variety of linear mixtures whenever the observations can be interpreted as compound signals comprising finite-length symbols. By compensating the shapes of the detected symbols, it efficiently combines all the assumed statistical properties of their arising times and operates directly in the space of sparse pulse trains. The proposed approach is significantly noise resistant, while successfully resolving also the underdetermined convolutive mixtures with a relatively large number of input trains.

In this paper, three different experimental settings were tested, evaluating the performance of CKC in the case of underdetermined and ill-conditioned mixtures (Experiments 1 and 3), and in the case of weakly correlated pulse trains (Experiment 2). Simulation results proved the accuracy of the CKC perfectly matches the accuracy of the LMMSE estimator. But, contrary to the LMMSE estimator, the CKC method exhibits a significant negative correlation between the number of reconstructed pulse trains and the pulse overlap probability p . There are several

possible explanations of this phenomenon, the most probable one including the induced non-diagonality of the correlation matrix $\mathbf{C}_{\mathbf{tt}}$ (see the discussion of Experiment 2). Invariance of the LMMSE estimator to the pulse overlapping originates from the necessity of a prior knowledge of the trains' cross-correlation with the observations. The drop of performance in the case of correlated pulse trains is the price we have to pay for reconstructing the pulse trains blindly.

Both the CKC and LMMSE estimator exhibit a significant performance drop when decomposing ill-conditioned and underdetermined mixtures. Typically, only the strongest observed symbols (in the sense of signal energy) are reconstructed, while all the others are treated as a background noise. In the case of biomedical signals, this is not a serious problem as the number of reconstructed input trains is already limited by the detection volume of the pick-up electrodes [13]. In the case of communication signals, on the other hand, the ratio between the number of observations and the number of users can be improved by increasing the sampling frequency. This also decreases the overlapping probability p and, hence, improves the performance of the proposed approach. The number of reconstructed pulse trains could be increased also by subtracting the identified symbol observations from the observed mixtures [22]. However, in the case of noise, this idea proves difficult to implement as it is very hard to obtain a perfect cancellation of the identified symbols.

Finally, the proposed decomposition approach reconstructs only the arising times and shapes of the detected symbols. In many cases (e.g., when decomposing biomedical signals), this ends the decomposition, as we are only interested in the detected form of symbols. When processing the communication signals, however, an additional single-input multiple-output (SIMO) decomposition step is needed in order to compensate the effect of the transfer channels and estimate the original source symbols out of their observations.

APPENDIX A

INVERSE OF DIAGONALLY DOMINANT MATRIX

Suppose the pulse train correlation matrix $\mathbf{C}_{\mathbf{tt}}$ is diagonally dominant, i.e., $\forall i, c_{ii} > \sum_{j \neq i} |c_{ij}|$ [18, p. 184]. If the pulse trains are weakly correlated we also have

$$\forall i \forall j, \quad i \neq j : 0 < |c_{ij}| c_{ii}^{-1} < \varepsilon; \quad \varepsilon \ll 1. \quad (\text{A.1})$$

The matrix $\mathbf{C}_{\mathbf{tt}}$ can be written as a sum of its diagonal and non-diagonal part $\mathbf{C}_{\mathbf{tt}} = \mathbf{D}_{\mathbf{tt}} + \mathbf{N}_{\mathbf{tt}}$, where $\mathbf{N}_{\mathbf{tt}} = \mathbf{C}_{\mathbf{tt}} - \mathbf{D}_{\mathbf{tt}}$ and $\mathbf{D}_{\mathbf{tt}} = \text{diag}([c_{1,1}, \dots, c_{N(K+L-1), N(K+L-1)}])$. Then, using the Neumann series, the inverse of $\mathbf{C}_{\mathbf{tt}}$ can be expressed as [18, pp. 126]:

$$\begin{aligned} \mathbf{C}_{\mathbf{tt}}^{-1} &= (\mathbf{D}_{\mathbf{tt}} + \mathbf{N}_{\mathbf{tt}})^{-1} \\ &= \left(\mathbf{D}_{\mathbf{tt}} \left(\mathbf{I} + \mathbf{D}_{\mathbf{tt}}^{-1} \mathbf{N}_{\mathbf{tt}} \right) \right)^{-1} \\ &= \mathbf{D}_{\mathbf{tt}}^{-1} + \left(\sum_{k=1}^{\infty} \left[-\mathbf{D}_{\mathbf{tt}}^{-1} \mathbf{N}_{\mathbf{tt}} \right]^k \right) \mathbf{D}_{\mathbf{tt}}^{-1} \end{aligned} \quad (\text{A.2})$$

providing the absolute value of each eigenvalue λ of the matrix $\mathbf{D}_{\mathbf{tt}}^{-1} \mathbf{N}_{\mathbf{tt}}$ is smaller than 1. According to Gerschgorin's the-

orem, the eigenvalues of $\mathbf{D}_{\text{tt}}^{-1}\mathbf{N}_{\text{tt}}$ are contained in a union of the circles defined by [18, p. 498]

$$|\lambda| \leq \sum_{\substack{j=1 \\ j \neq i}}^{N(K+L-1)} \left| \frac{c_{ij}}{c_{ii}} \right| = \frac{1}{|c_{ii}|} \sum_{\substack{j=1 \\ j \neq i}}^{N(K+L-1)} |c_{ij}| < 1. \quad (\text{A.3})$$

Therefore, the first-order approximation of $\mathbf{C}_{\text{tt}}^{-1}$ yields $\hat{\mathbf{C}}_{\text{tt}}^{-1} \approx \mathbf{D}_{\text{tt}}^{-1} - \mathbf{D}_{\text{tt}}^{-1}\mathbf{N}_{\text{tt}}\mathbf{D}_{\text{tt}}^{-1}$, which, when $\sum_j |c_{ij}| < c_{ii}$ and (A.1) are taken into account, proves the inverse of \mathbf{C}_{tt} has a superior diagonal. According to extensive numerical simulations, similar conclusions also apply to the matrices which are not strictly diagonally dominant but still fulfill the condition in (A.1).

APPENDIX B

AVERAGE NUMBER OF PULSES IN THE ELEMENT-WISE PRODUCT OF PULSE TRAINS

Although the number of components R in (11) can be considered an arbitrary value, we only focus on two cases ($R = 2$ and $R = 4$). The derivations for other cases follow those presented here and are left to the interested reader. For the sake of simplicity, we also assume the number of pulses in each train equals F .

First, observe the pulses in the element-wise product $\mathbf{h}_{n_0, n_1} = \{\nu_{n_0}(n) \cdot \nu_{n_1}(n); n = 0, 1, 2, \dots\}$, $n_0 \neq n_1$. This product will certainly contain the pulses of all the trains from $G_{n_0} \cap G_{n_1}$. Their average number of pulses can be estimated as $F \cdot \vartheta(g_{01}, 0)$ where $g_{01} = \text{card}(G_{n_0} \cap G_{n_1})$, while $\vartheta(g_0, g_1)$ with $g_i = \text{card}(G_{n_i})$; $i = 0, 1$, denotes the correction factor introduced by the fact that the pulses of the trains in $G_{n_0} \cap G_{n_1}$ mutually overlap

$$\vartheta(g_0, g_1) = \left(g_0 - \sum_{k=2}^{g_0+g_1} (-1)^k \binom{g_0+g_1}{k} p^{k-1} \right) \quad (\text{B.1})$$

where $\binom{m}{r}$ denotes the number of combinations of m elements, taken r elements at a time. The product \mathbf{h}_{n_0, n_1} will also contain all those pulses of any pulse train from the set difference $G_{n_0} - G_{n_1}$ which randomly overlap with the pulses of any pulse train from $G_{n_1} - G_{n_0}$. Their number can be estimated as

$$F \cdot p \cdot \vartheta(g_0 - g_{01}, g_{01}) \cdot \vartheta(g_1 - g_{01}, g_{01}). \quad (\text{B.2})$$

Hence, the average number of pulses in the product \mathbf{h}_{n_0, n_1} yields

$$v(p, F) \leq F \cdot [\vartheta(g_{01}, 0) + p \cdot \vartheta(g_0 - g_{01}, g_{01}) \cdot \vartheta(g_1 - g_{01}, g_{01})]. \quad (\text{B.3})$$

We can follow the same route in the case of $R = 4$ components in (11). First, define the following mutually disjunctive sets:

$$\begin{aligned} M_{0,1,2,3} &= G_{n_0} \cap G_{n_1} \cap G_{n_2} \cap G_{n_3} \\ M_{i,j,k} &= G_{n_i} \cap G_{n_j} \cap G_{n_k} - M_{0,1,2,3} \\ M_{i,j} &= \left(G_{n_i} \cap G_{n_j} - \bigcup_k M_{i,j,k} \right) - M_{0,1,2,3} \\ M_i &= \left(\left(G_{n_i} - \bigcup_k M_{i,k} \right) - \bigcup_{j,k} M_{i,j,k} \right) - M_{0,1,2,3} \end{aligned} \quad (\text{B.4})$$

with $i, j, k \in \{0, 1, 2, 3\}$ and $i \neq j, i \neq k, j \neq k$. Using denotation $m_i = \text{card}(M_i)$, the average number of pulses in (11) can be estimated as:

$$\begin{aligned} v(p, F) &\leq F \vartheta(m_{0123}, 0) \\ &\quad + pF[\vartheta(m_{012}, m_{0123})\vartheta(m_3, g_3 - m_3) \\ &\quad + \vartheta(m_{013}, m_{0123})\vartheta(m_2, g_2 - m_2) \\ &\quad + \vartheta(m_{023}, m_{0123})\vartheta(m_1, g_1 - m_1) \\ &\quad + \vartheta(m_{123}, m_{0123})\vartheta(m_0, g_0 - m_0) \\ &\quad + \vartheta(m_{01}, g_{012} + m_{013})\vartheta(m_{23}, g_{023} + m_{123}) \\ &\quad + \vartheta(m_{02}, g_{012} - m_{023})\vartheta(m_{13}, g_{013} + m_{123}) \\ &\quad + \vartheta(m_{03}, g_{013} - m_{023})\vartheta(m_{12}, g_{012} + m_{123})] \end{aligned} \quad (\text{B.5})$$

where, supposing $p \ll 1$, all the factors multiplied by p to the second or higher powers were neglected. According to (B.5), the average number of pulses depends mainly on the number of trains in the sets $G_{n_0}, G_{n_1}, G_{n_2}$ and G_{n_3} and their distribution in the corresponding M sets. In the sequel, we are going to study just the cases that prove the (13) and (14), respectively.

Case 1: Suppose the set $M_{0,1,2}$ contains a single pulse train, while sets $M_{0,1,3}, M_{0,2,3}, M_{1,2,3}$ and $M_{0,1,2,3}$ are empty. Further assume that $m_{ij}, i \neq j$, are all equal to or less than 1. Then (B.5) yields

$$v(p, F) \approx F \cdot p[\vartheta(1, 0)\vartheta(m_3, g_3 - m_3) + 3\vartheta(1, 1)\vartheta(1, 0)]. \quad (\text{B.6})$$

Use $m_3 = g_3 - m_{03} - m_{13} - m_{23} - m_{013} - m_{023} - g_{123}$ and recall that, on average, the number of pulse trains in the set G_{n_3} can be estimated as $g_3 = N(K+L-1)p$, where N stands for the number of symbols, K is extension factor and L is the symbol length. Then, (B.6) simplifies to $v(p, F) \approx Fp^2N(K+L-1)$.

Case 2: Suppose the set $M_{0,1,2,3}$ contains a single pulse train, while sets $M_{0,1,2}, M_{0,1,3}, M_{0,2,3}$ and $M_{1,2,3}$ are empty. Further assume that $m_{ij}, i \neq j$, are all equal to or less than 1. Then (B.5) simplifies to

$$v(p, F) \approx F \cdot [\vartheta(1, 0) + 3p\vartheta(1, 1)\vartheta(1, 0)] \approx F(1 + 3p). \quad (\text{B.7})$$

APPENDIX C

ELIMINATION OF UNDERDETERMINED MIXING MATRIX

Define $\mathbf{B} = \mathbf{A}^T(\mathbf{A}\mathbf{A}^T)\#\mathbf{A}$ where \mathbf{A} is an arbitrary mixing matrix of size $KM \times N(L+K-1)$, with $KM < N(L+K-1)$. By definition, \mathbf{B} is orthogonal projector and positive semidefinite [18, pp. 434]. Denoting by \mathbf{U} the matrix of the right singular vectors of \mathbf{A} , the (i, j) th element of matrix \mathbf{B} can be calculated as

$$b_{ij} = \sum_{k=1}^{KM} u_{ik}u_{jk} \quad (\text{C.1})$$

where u_{ij} denotes the (i, j) th element of the matrix \mathbf{U} . Considering the orthogonality of matrix \mathbf{U} and denoting by $\mathbf{u}(i)$ its i th column, three different cases can emerge.

1. The energy of $\mathbf{u}(i)$ is concentrated in the first KM elements. This is the ideal case as (C.1) guarantees the diag-

onal element b_{ii} will be close to 1, while all off-diagonal elements b_{ij} will be close to 0.

2. The energy of $\mathbf{u}(i)$ is distributed among all elements. This is the most probable case and implies the value of the diagonal element b_{ii} decreases with the ratio $N(L+K-1)/KM$, while the values of off-diagonal elements b_{ij} simultaneously increase.
3. The majority of $\mathbf{u}(i)$'s energy is concentrated in the last $N(L+K-1)/KM$ elements. This is the worst case scenario as (C.1) shows the diagonal element b_{ii} will be close to zero.

The number of columns $\mathbf{u}(i)$ being concomitant with case 3, is limited. Namely, supposing all the columns of the unitary matrix \mathbf{U} correspond to case 3, it would necessarily imply that \mathbf{U} is similar to the projection matrix mapping $N(L+K-1)$ -dimensional space onto $(N(L+K-1) - KM)$ -dimensional subspace. These arguments further support the following empirical observation:

Observation 1: For most \mathbf{A} , supposing the ratio $N(L+K-1)/KM$ small, the matrix \mathbf{B} will have at least a part of its diagonal elements superior to all off-diagonal elements.

The numerical simulations further reveal the dominant diagonal elements are at least several magnitudes higher than the corresponding off-diagonal elements, as long as the number of extended pulse trains, $N(L+K-1)$, does not exceed the number of extended observations, KM , by factor 2.

REFERENCES

- [1] A. Hyvarinen, J. Karhunen, and E. Oja, *Independent Component Analysis*. New York: Wiley, 2001.
- [2] M. Girolami, "A variational method for learning sparse and overcomplete representations," *Neural Comput.*, vol. 13, pp. 2517–2532, 2001.
- [3] M. Davies and N. Mitianoudis, "A simple mixture model for sparse overcomplete ICA," *Inst. Elect. Eng. Proc. Vision, Image and Signal Process.*, vol. 151, pp. 35–43, 2004.
- [4] M. S. Lewicki and T. J. Sejnowski, "Coding time-varying signals using sparse, shift invariant representations," in *Advances Neural Inf. Process. Syst.*, 1999, vol. 11, pp. 730–736.
- [5] B. A. Olshausen, R. P. N. Rao, B. A. Olshausen, and M. S. Lewicki, Eds., "Sparse codes and spikes," in *Probabilistic Models of the Brain*, 2002, pp. 257–272.
- [6] T. Blumensath and M. Davies, "Sparse and shift-invariant representations of music," *IEEE Trans. Audio, Speech Language Process.*, vol. 14, pp. 50–57, 2006.
- [7] H. Wersing, J. Eggert, and E. Körner, "Sparse coding with invariance constraints," in *Proc. ICANN 2003*, Istanbul, Turkey, 2003, pp. 285–392.
- [8] P. Jost, P. Vanderghenst, S. Lesage, and R. Groibonval, "MoTIF: An efficient algorithm for learning translation invariant dictionaries," in *Proc. ICASSP 2006*, 2006, vol. 5, pp. V-857–V-860.
- [9] C. Disselhorst-Klug, G. Rau, A. Schmeer, and J. Silny, "Non-invasive detection of the single motor unit action potential by averaging the spatial potential distribution triggered on a spatially filtered motor unit action potential," *J. Electromyogr. Kinesiol.*, vol. 9, pp. 67–72, 1999.
- [10] S. M. Key, *Fundamentals of statistical signal processing: Estimation theory*. Englewood Cliffs, NJ: Prentice-Hall Int., 1993.

- [11] U. Madhoo, "Blind adaptive interference suppression for direct-sequence CDMA," *Proc. IEEE*, vol. 86, pp. 2049–2069, 1998.
- [12] J. Proakis, *Digital Communications*, 3rd ed. New York: McGraw-Hill, 1995.
- [13] A. Holobar and D. Zazula, "Correlation-based approach to separation of surface electromyograms at low contraction forces," *Med. Biol. Eng. Comput.*, vol. 42, pp. 487–496, 2004.
- [14] D. Zazula and A. Holobar, "Blind source separation based on a single observation," in *Proc. BPC 2006*, Setúbal, Portugal, 2006, pp. 76–85.
- [15] R. Merletti, "Surface electromyography: Possibilities and limitations," *J. Rehab. Sci.*, vol. 7, pp. 25–34, 1994.
- [16] M. S. Lewicki, "A review of methods for spike sorting: The detection and classification of neural action potentials," *Network: Comput. Neural Syst.*, vol. 9, pp. 53–78, 1998.
- [17] B. Boashash, *Time-Frequency Signal Analysis and Processing*. Englewood Cliffs, NJ: Prentice-Hall PTR, 2001.
- [18] C. D. Meyer, *Matrix Analysis and Applied Linear Algebra*. Philadelphia, PA: SIAM, 2001.
- [19] D. Farina and R. Merletti, "A novel approach for precise simulation of the EMG signal detected by surface electrodes," *IEEE Trans. Biomed. Eng.*, vol. 48, pp. 637–646, 2001.
- [20] R. Merletti and P. A. Parker, *Electromyography: Physiology, Engineering and Noninvasive Applications*. Hoboken, NJ: Wiley, 2004.
- [21] A. J. Fuglevand, D. A. Winter, and A. E. Patla, "Models of recruitment and rate coding organization in motor unit pools," *J. Neurophysiol.*, vol. 70, pp. 2470–2488, 1993.
- [22] T. Ristaniemi and T. Huovinen, "Joint delay tracking and interference cancellation in DS-CDMA systems using successive ICA for oversaturated data," in *Proc. ICA2004*, Granada, Spain, 2004, pp. 1173–1180.



Aleš Holobar (M'98) received the B.S. and D.Sc. degrees in computer science from the University of Maribor, Slovenia, in 2000 and 2004, respectively.

From 2000 to 2006 he was a researcher at the Faculty of Electrical Engineering and Computer Science, University of Maribor. He is currently a Marie Curie Fellow at Politecnico di Torino, Torino, Italy. His research interests include virtual reality, conceptual learning, and signal processing, with current activities focused on blind source separation and biomedical signal processing.

Dr. Holobar is a member of ISEK, IAPR, and Slovenian Society of Pattern Recognition.



Damjan Zazula (M'87–SM'04) received the Dipl.Eng., Master's, and Doctor of Science degrees in electrical engineering from the University of Ljubljana, Slovenia, in 1974, 1978, and 1990, respectively.

After being involved in industrial R&D for 12 years, he joined the Faculty of Electrical Engineering and Computer Science, University of Maribor, Slovenia, in 1987. Currently, he holds a Full Professor position in computer science, and from 1998 to 2003, he was also appointed an Associate Dean of Research. He has spent several months as visiting professor with the ETH in Zurich, Switzerland, and Ecole Centrale de Nantes, France. His main research interests are compound signal decomposition, biomedical imaging, and virtual training tools.

Dr. Zazula is a member of the IEEE Signal Processing Society, EURASIP, IAPR, Slovenian Technical Society, Slovenian Society of Pattern Recognition, and Slovenian Society of Biomedical Engineering.