Activity Index Variance as an Indicator of the Number of Signal Sources

ROK ISTENIC, DAMJAN ZAZULA
System Software Laboratory, Institute of Computer Science
Faculty of Electrical Engineering and Computer Science
University of Maribor
Smetanova ulica 17, SI-2000 Maribor
SLOVENIA
rok.istenic@uni-mb.si http://storm.uni-mb.si

Abstract: - In this paper we introduce a novel technique that can be used as an indicator of the number of active signal sources in convolutive signal mixtures. The technique is designed so that the number of sources is estimated using only recorded signals and some marginal information, such as possible minimum and maximum triggering frequencies of sources, but no information on mixing matrix, other parameters of signal sources, etc. Our research is based on the convolution kernel compensation method (CKC), which is a blind source separation method. First, a correlation matrix of the recorded signals is estimated. Next, a measure of the global activity of the signal sources, called activity index, is introduced. The exact analytical model of the activity index variance was derived for the purpose of the estimation of the number of signal sources. Using the analytical model, the number of active signal sources can be estimated if some a priori marginal information is available. We evaluated these marginal parameter values in extensive simulations of compound signals. The number of sources, their lengths, signal-to-noise ratio, source triggerings, and the number of measurements were randomly combined in preselected ranges. By using the established marginal parameter values and increasing extension factors, the model of the activity index variance was deployed to estimate the number of signal sources. The estimation results using synthetic signal mixtures are very promising.

Key-Words: - Compound signals, estimation of the number of sources, correlation matrix, convolutive signal mixture, variance model, convolution kernel compensation.

1 Introduction

In all engineering disciplines signal measurements usually appear in complex forms, they are actually compounds of many sources that contribute to the recorded signal. Therefore, the terms signal mixtures or compound signals are commonly used to describe such signals [7], [9]. Once the signal is recorded, we are interested in its components. We want to estimate how many sources produced the mixture and how the individual sources look like [12]. The procedure of source estimation starts already with signal acquisition, where different recording techniques can support the final goal. Nowadays, sensor arrays are becoming quite common for several concurrent measurements, because more spatial information is gained this way [5], [6].

A well-known everyday example of compound signals is a meeting room with more people talking simultaneously. If the number of speakers and noise are limited, the human brain is capable of identifying the individual speakers, but the procedure is not so straightforward for the computer [14], [15].

The problem of estimating the number of signal sources is a common problem in all fields dealing with complex signals, such as surface electromyograms (sEMG) [7], [20], electroencephalograms (EEG) [11], radar and sonar signals [1], image processing [19], speech recognition systems [21], etc.

Various techniques have been proposed for estimating the number of signal sources in multichannel time-series, one of the first being eigenvalue-based method [1] that applies the fact that the number of dominant eigenvalues of signal correlation matrix corresponds to the number of signal sources. The main problem of eigenvalue based methods is how to set a threshold between dominant and non dominant eigenvalues. The information theoretic criteria, such as Akaike information criterion (AIC) and minimum description length (MDL) are used for optimal threshold setting, because these techniques do not require any subjective threshold settings; the number of signal sources is estimated by minimizing the AIC or the MDL criteria.

Many improvements of the original eigenvalue-based method have been introduced. The eigenvalue gradient method (EGM) [2] computes gradients of sorted eigenvalues of auto-correlation matrix to estimate the number of signal sources. The EGM performance is comparable to that of the MDL at high signal-to-noise ratio (SNR), but at low SNR the EGM is better than MDL.

Another extension of the original method uses Gerschgorin disks theorem [3]. By introducing the unitary transformation of covariance matrix, the Gerschgorin radii of the eigenvalues are used to determine the number of signal sources. It has been reported that Gerschgorin disk estimator provide more accurate estimation of the number of sources than MDL in both, simulated and real signal mixtures.

Gu et al. [4] deployed eigenvectors instead of eigenvalues and peak-to-average power ratio based frequency estimation algorithm to estimate the number of sources. Authors reported that their method is superior to others, especially at low SNR.

Other important approaches for the estimation of the number of signal sources include statistical approach [5], [16], independent component analysis (ICA) [9], overcomplete ICA [9], [18], wavelet packets [21], frequency-domain blind source separation (BSS) [20], inverse filters [17].

The common drawback of the proposed methods is their high computational complexity, because they require several multi-dimensional searches. Fishler and Poor [6], motivated by real life problems, implemented the low-complexity MDL-based method by eliminating the need for multi-dimensional searches. Their estimator became a favourable choice for practical application, because of its robustness and speed.

The vast majority of the researches tackled problems with overdetermined systems (where number of sources is smaller than number of measurements) [1]-[6], but there have also been some experiments with underdetermined systems (also known as overcomplete representations), where the number of sources is greater than the number of measurements [8], [9], [13]. The overview and comparison of some overcomplete ICA algorithms can be found in [18]. Dealing with overcomplete representations, the estimation of the number of sources is not excellent, but reduction in the number of needed measurements is achieved.

Olsson and Hansen [8] introduced a probabilistic blind source separation that is based solely on the time-varying second order statistics of the sources. Their algorithm employs a Gaussian linear model for the mixture and expectation maximization (EM) algorithm. The method performance was tested on speech signals. It was shown that the activity of the two speech sources in a single measurement can be detected.

Lee et al. [9] presented technique for the blind source separation of more sources than measurements for the speech signals. They used framework for learning overcomplete representations and reported that with two measurements they were able to extract up to four mixed speech signals.

Our paper is structured into six sections. The next section describes the methods used, namely the model of convolutive signal mixtures, convolutive kernel compensation (CKC), the activity index and analytical model of the activity index variance. Section 3 is devoted to the simulation studies and simulation results. Section 4 reveals how the model of activity index variance can be used to estimate the number of signal sources, and the advantages and weaknesses of the approach. Next section discusses the derived analytical model of the activity index variance and simulation results. Section 6 concludes the paper and presents some possibilities for future research.

2 Methods

Our main motivation for the research was to model and analyse the relationship between activity index and the number of signal sources in convolutive signal mixtures. The modelled relationship can be used to support compound signal decomposition, where it is helpful to know the number of active signal sources prior to the decomposition. To use an estimator before the decomposition, it has to be fast and reliable; therefore we based it on the CKC activity index, which has all desired properties [7].

2.1 Model of the signal mixtures

We consider convolutive multiple-input multiple-output (MIMO) system as in [7], [13]. Suppose that N source signals in the form of pulse trains $s_1(n),...,s_N(n)$ are convolved by impulse responses of $N\times M$ system channels and observed by M sensors, producing M measurements: $x_1(n),..., x_M(n)$. Taking into account the linearity of the modelled MIMO system, the i-th measurement $x_i(n)$ can be written as a sum of contributions from all N sources:

$$x_i(n) = \sum_{j=1}^{N}\sum_{l=0}^{L-1} h_{ij}(l)\cdot s_j(n-l), \quad i = 1,\ldots,M \qquad (1)$$

where $h_{ij}(l)$ denotes the impulse response of the *j*-th source as detected by the *i*-th sensor, and $s_j(n)$ stands for the train of δ pulses of the *j*-th source. Suppose all the system-channel impulse responses of length *L* samples.

Assuming *N* inputs and *M* outputs, the vector notations are as follows: $\mathbf{x}(n) = [x_1(n),....,x_M(n)]^{\mathrm{T}}$ for the vector of *M* measurements and $\mathbf{s}(n) = [s_1(n),....,s_N(n)]^{\mathrm{T}}$ for the vector of *N* sources. Using the vector notations, Eq. (1) can be transformed into a matrix form:

$$\mathbf{x}(n) = \mathbf{Hs}(n), \qquad (2)$$

where **H** stands for the so called mixing matrix of size $M\times NL$, which contains the system-channel impulse responses $h_{ij}(l)$.

When the influence of noise is considered we get:

$$\mathbf{y}(n) = \mathbf{x}(n) + \boldsymbol{\omega}(n), \qquad (3)$$

where $\mathbf{y}(n) = [y_1(n),....,y_M(n)]^{\mathrm{T}}$ denotes the transposed vector of *M* noisy measurements, and $\boldsymbol{\omega}(n) = [\omega_1(n),....,\omega_M(n)]^{\mathrm{T}}$ stands for the noise vector.

2.2 Activity index

In this section the global indicator of source activity, known also as activity index, is studied. Activity index was first introduced as a part of CKC algorithm [7]. Preliminary studies have shown that activity index indeed is a function of the number of signal sources [22]; therefore its analytical model is studied in details in this paper. Activity index is derived for the convolutive signal mixtures as presented in the Section 2.1. The only compulsory information to compute activity index are signal measurements (signal mixtures).

Having *M* measurements, and assuming that the number of measurements *M* is greater than the number sources *N* (the system in Eq. (2) is overdetermined), a positive integer *K* exists that satisfies: $M(K+1) > N(L+K)$. *K* is known as extension factor and stands for the number of shifted replicas of original measurements (see Fig. 1). Extended vectors of measurements are designated by the bars, for example $\bar{\mathbf{x}}$ or $\bar{\mathbf{y}}$.

Activity index is obtained by multiplying the inverse of correlation matrix with extended vectors of measurements on both sides:

$$I_{\mathrm{A}}(n) = \bar{\mathbf{y}}^{\mathrm{T}}(n)\cdot \mathbf{R}_{\mathrm{y}}^{-1}\cdot \bar{\mathbf{y}}(n). \qquad (4)$$

The correlation matrix of extended noisy measurements is computed as $\mathbf{R}_{\mathrm{y}} = \overline{\mathbf{y}\mathbf{y}}^{\mathrm{T}}$, its dimensions are M (K+1) × M (K+1), where M (K+1) stands for the number of extended measurements.

Assuming that there is no noise, the activity index could be thought of as an indicator of a global source activity as it differs from zero only at the time instants *n*, where at least one source is active.

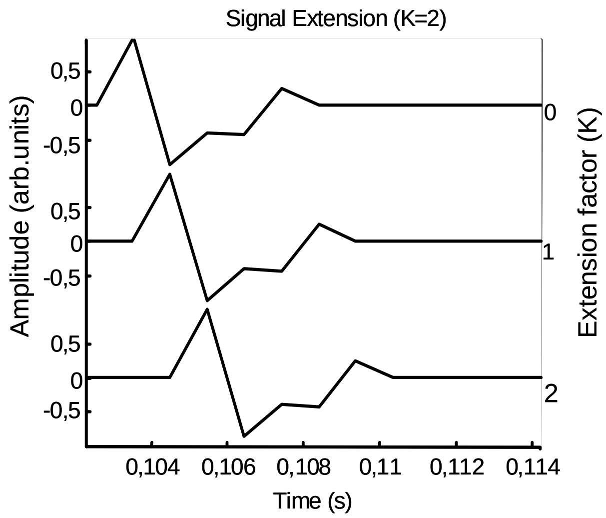

Fig. 1: Extension of a single synthetic measurement. The original measurement (*K*=0, upper row) is extended by 2 shifted replicas (*K*=1 and *K*=2). *X*-axis represents time, left *y*-axis the amplitude and right *y*-axis the extension factor.

Activity index can be understood as a superimposition of more signal sources and an additional component which stands for additive Gaussian white noise (Fig. 2). The variance of such superimposition is computed with four main terms: variances of each individual source, covariances between the sources, variance of the noise and covariances between the noise and all the sources.

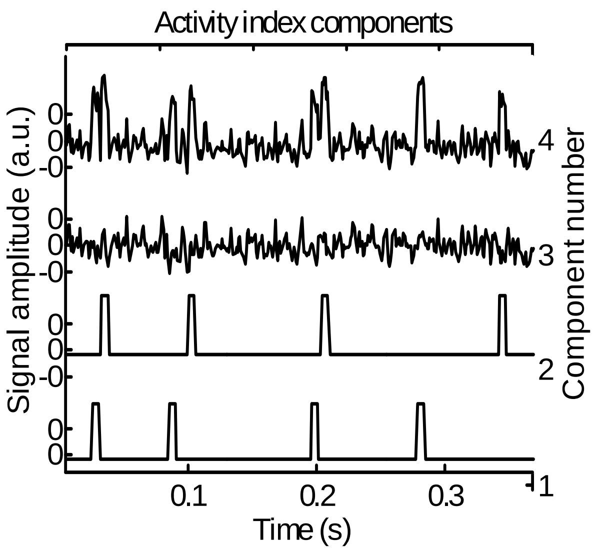

Fig. 2: Graphical representation of activity index components. The *x*-axis represents time, left *y*-axis amplitude and right *y*-axis component number. Components 1 and 2 represent signal sources, component 3 is white Gaussian noise, while component 4 depicts complete activity index, obtained as superimposition of components 1, 2 and 3.

2.3 Model of activity index variance

In the sequel, all four contributions within the activity index variance, var(I_A), will be modelled separately. The same notation will be used throughout the derivations: N as the number of sources, M the number of observed signals, L the length of system-channel impulse responses, K extension factor, D length of signals, E_i the number of firings of the i-th source, $\omega(n)$ the n-th sample of noise, and η_{ij} the number of superimposed samples in sources i and j.

The sum of variances of individual sources can then be modelled as follows:

$$\frac{-N(L+K)^2}{D(D-1)} + \frac{L+K}{D-1}\sum_{i=1}^{N}\frac{1}{E_i}. \quad (5)$$

The sum of covariances between the sources yields:

$$\frac{-N(N-1)(L+K)^2}{D(D-1)} + \frac{2}{D-1}\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}\eta_{ij}\frac{1}{E_i E_j}. \quad (6)$$

Variance of the noise contribution can be computed as:

$$-\frac{(M-NL+K(M-N))^2}{D(D-1)} + \frac{\sum_{n=0}^{D-1}\omega^2(n)}{D-1}. \quad (7)$$

Covariances between all the sources and noise contribution sum up into:

$$-2\frac{N(L+K)(M-NL+K(M-N))}{D(D-1)} + \frac{2}{D-1}\sum_{i=1}^{N}\sum_{n\in K_i}\frac{\omega(n)}{E_i}. \quad (8)$$

The total activity index variance is obtained by merging the models from expressions (5), (6), (7), and (8). If we sum up all the first (left) terms in expressions (5), (6), (7), and (8) the following simple result emerges:

$$\frac{-M^2(L+K)^2}{D(D-1)}. \quad (9)$$

The right-hand side terms in expressions (5), (6), (7), and (8) contain sums over one or two indices. The values of these sums depend on unknown parameters, such as samples of noise, number of firings of each individual source, number of superimposed samples of sources pairs, etc. In the final effort, the modelled relationships should be identified and the parameters estimated. So, the number of unknowns must be kept as low as possible. With this in mind, we tried to simplify the whole expression for the activity index variance in the following way:

$$\begin{aligned}\mathrm{var}(I_A) &= \frac{-M^2(L+K)^2}{D(D-1)} + \frac{N(L+K)}{D-1}c_0 \\ &+ \frac{N(N-1)(L+K)^2}{D-1}c_1 + \frac{M^2(K+1)^2}{D-1}c_2 \\ &+ \frac{2N(L+K)(M-NL+K(M-N))}{D-1}c_3,\end{aligned} \quad (10)$$

where only four new parameters c_0-c_3 were introduced and defined in Eqs. (11) to (14). These parameters are expected variable, but as we are going to see from our simulations, at least two of them are practically independent of the number of signal sources and their lengths. This fact may considerably mitigate the estimations based on the derived analytical models.

As we see from the following relationships:

$$c_0 = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{E_i}, \qquad (11)$$

$$c_1 = \frac{2\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}\eta_{ij}\frac{1}{E_iE_j}}{N(N-1)(L+K)^2}, \qquad (12)$$

$$c_2 = \frac{\sum_{n=0}^{D-1}\omega^2(n)}{M^2(K+1)^2}, \qquad (13)$$

$$c_3 = \frac{\sum_{i=1}^{N}\sum_{n\in K_i}\frac{\omega(n)}{E_i}}{N(L+K)(M-NL+K(M-N))}, \qquad (14)$$

c_0 is independent of extension factor K, while all others are functions of K. Their behaviour is interpreted by simulation studies on synthetic compound signals, as explained in the sequel.

3 Simulation study

Eq. (10) that stands for the model of the activity index variance involves two controllable parameters: number of extensions (K) and number of observed signals (M). In real cases, it can be expected the number of observations is given, at least after the signals have been recorded. So, K remains as the only controllable parameter. Therefore, the final estimation of the unknowns from Eq. (10) can be based on the computation of the activity index variance of real signals versus different extension factors.

3.1 Synthetic signals

For synthetic signal generation process, our custom generator was used, which generates convolutive signal mixtures in four steps. In the first step the firing patterns (trains of σ pulses) for all sources are generated, with their triggering moments according to a mean distance of 250 ms between consecutive triggerings and deviation of this distance of 50 ms. Sampling frequency was set to 1024 Hz.

Next, the system-channel impulse responses for all sources and all measurements are generated (see data model in Section 2.1 for the details). We resorted to random system-channel impulse responses to show that our method is capable of estimating the number of signal sources in many various applications, not only sEMG for example. The random system-channel impulse responses for a single active signal source and a matrix of 12 × 5 electrodes are depicted in Fig. 3. The Matlab random generator 'rand' was used and random numbers were generated on the interval [-1, 1].

In the third step the firing patterns of sources are convolved with the corresponding system impulse responses, and noiseless measurements are obtained. Finally, zero-mean white Gaussian noise of different SNRs (from 20 dB down to 0 dB) is added to the signals.

For each new simulation run, new system impulse responses, triggering instants, and white Gaussian noise are randomly generated.

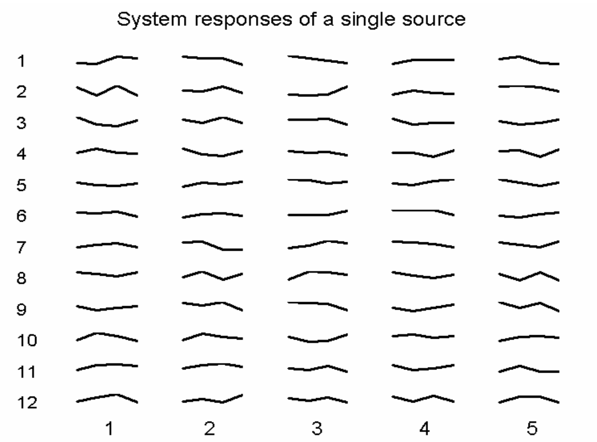

Fig. 3: Random shapes of generated system-channel impulse responses for 12 × 5 measurements and a single source. All system responses are of length L samples and their amplitude ranges between -1 and 1.

3.2 Simulation results

The behaviour of parameters c_0, c_1, c_2, and c_3 was observed through a series of simulations with synthetic compound signals. The signal generator has already been described in the previous section.

In our first experiment, the ranges of parameters c_i were estimated so that all 5 parameters (L, N, M, K, and SNR) were changed in the intervals that are most likely to happen in the reality: $N, L \in [1, 5, 10, 15, 20]$, $M \in [10, 50, 100, 200]$, $K \in [1,...,5]$, and SNR $\in [20, 10, 5, 1, 0]$ dB. This resulted in the following estimated ranges of parameters c_i that are reported in Tables 1- 4.

Our aim was to study, how the ranges of parameters c_i change with the number of measurements (M). Therefore, for one table, M was fixed, while other four parameters (L, N, K, and SNR) changed as reported before. The range minimum of c_i was obtained as minimal value that parameter c_i occupied at all possible combinations of L, N, K, and SNR, while the range maximum of c_i was obtained as maximal value of c_i for all possible combinations of L, N, K, and SNR.

Table 1: Ranges of parameters c_i at M = 10.

Parameter	Range minimum	Range maximum
c_0	9.09 e^{-2}	1.07 e^{-1}
c_1	4.00 e^{-4}	2.31 e^{-3}
c_2	4.17 e^{-5}	1.96 e^{-2}
c_3	-5.40 e^{-2}	3.03 e^{-2}

Table 2: Ranges of parameters c_i at M = 50.

Parameter	Range minimum	Range maximum
c_0	9.39 e^{-2}	1.06 e^{-1}
c_1	3.09 e^{-4}	4.55 e^{-3}
c_2	6.06 e^{-5}	9.56 e^{-4}
c_3	-1.61 e^{-3}	9.33 e^{-4}

Table 3: Ranges of parameters c_i at M = 100.

Parameter	Range minimum	Range maximum
c_0	9.09 e^{-2}	1.11 e^{-1}
c_1	2.27 e^{-4}	1.91 e^{-3}
c_2	3.69 e^{-4}	9.65 e^{-4}
c_3	5.38 e^{-4}	9.05 e^{-4}

Table 4: Ranges of parameters c_i at M = 200.

Parameter	Range minimum	Range maximum
c_0	9.09 e^{-2}	1.03 e^{-1}
c_1	5.56 e^{-4}	1.98 e^{-3}
c_2	5.89 e^{-4}	7.07 e^{-4}
c_3	7.34 e^{-4}	7.68 e^{-4}

Inspecting Tables 1- 4 carefully, one can see that ranges of parameters c_i are stabilizing with the increasing number of measurements. Having enough measurements (at least 100), the dynamics of parameters c_i is very limited, so that they can be represented only by close borders of their ranges.

Comparing the variability of individual parameters c_i at 200 measurements (Table 4), it is obvious that c_3 is the most stable one, followed by c_2, c_1, and c_0 in the increasing order of variability. Differences between the range minimum and maximum at 200 measurements are: 0.34e^{-5} for c_3, 1.18e^{-4} for c_2, 1.4e^{-3} for c_1, and 1.21e^{-2} for c_0.

In our second experiment functional relationships between parameters c_i, extension factor and the number of sources were studied. Measurements were generated in the same way as described in the first experiment, with the only exception that the number of measurements was fixed at 100. The results are depicted in Fig. 4. This experiment confirmed that c_0 is not a function of K, while c_1, c_2 and c_3 are. It is also evident that the parameters c_i converge to certain stable values as K increases. The experiment also revealed functional relationships between parameters c_i and the number of sources. Parameters c_1 and c_2 are decreasing with the increasing number of sources. The exception is c_1 that equals zero when only one source is active, because the covariance between sources does not exist at that special occasion. Parameters c_0 and c_3 do not change with the number of sources uniformly.

In the third experiment functional relationships between parameters c_i, SNR, and the number of sources were studied. Fig. 5 shows that parameters c_0, c_1, and c_2 are totally independent of SNR, which is a very desirable property. Parameter c_3 is a function of SNR (c_3 is decreasing with decreasing SNR), but it does not vary in a large content.

Besides all described functional relationships, the simulations also show that parameters c_i do not vary significantly by the changing number of sources. This means that the number of sources and the length of system-channel impulse responses may be searched for independently of the parameters c_i. Or in other words, as these parameters do not depend much on the extension factor, either, they can be treated constant and introduced in solving Eq. (10) for N and L with their marginal values evaluated a priori by extensive simulations and reported in Tables 1 - 4. However, special care must be taken, as using such an approximation of parameters c_i can decrease the estimation accuracy and hence deteriorate our proposed approach.

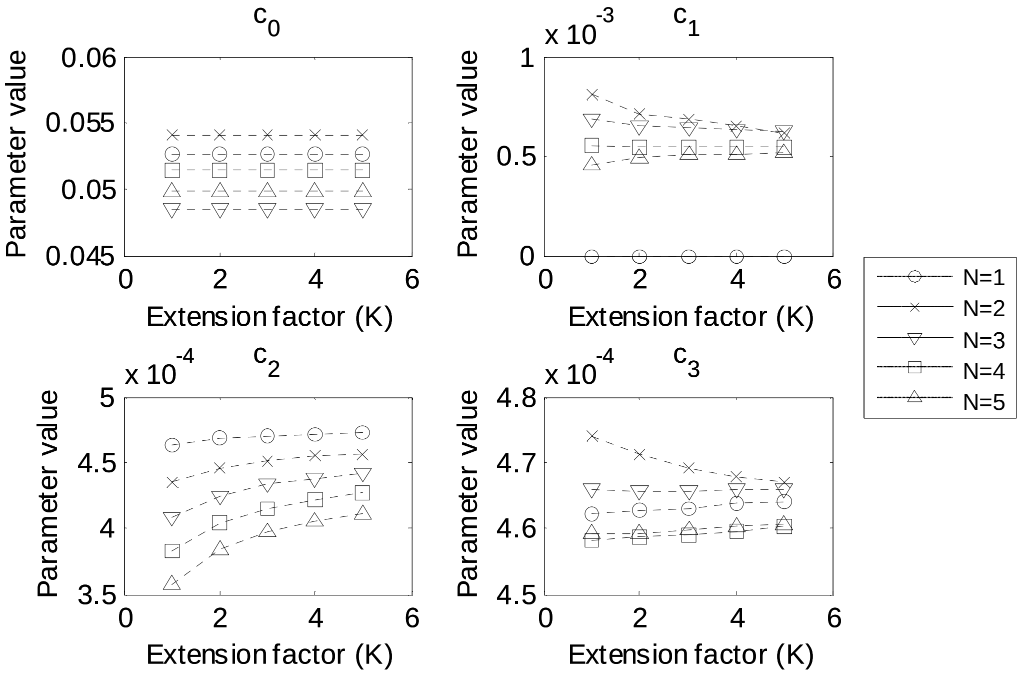

Fig. 4: The functional relationships between parameters c_i, extension factor (K) and the number of sources (N). The x-axis represents extension factor, while y-axis reports the values of parameters c_i. Each curve stands for a different number of sources, as it is reported in the figure legend. Fixed parameters were: L = 5, M = 100 and SNR = 20 dB.

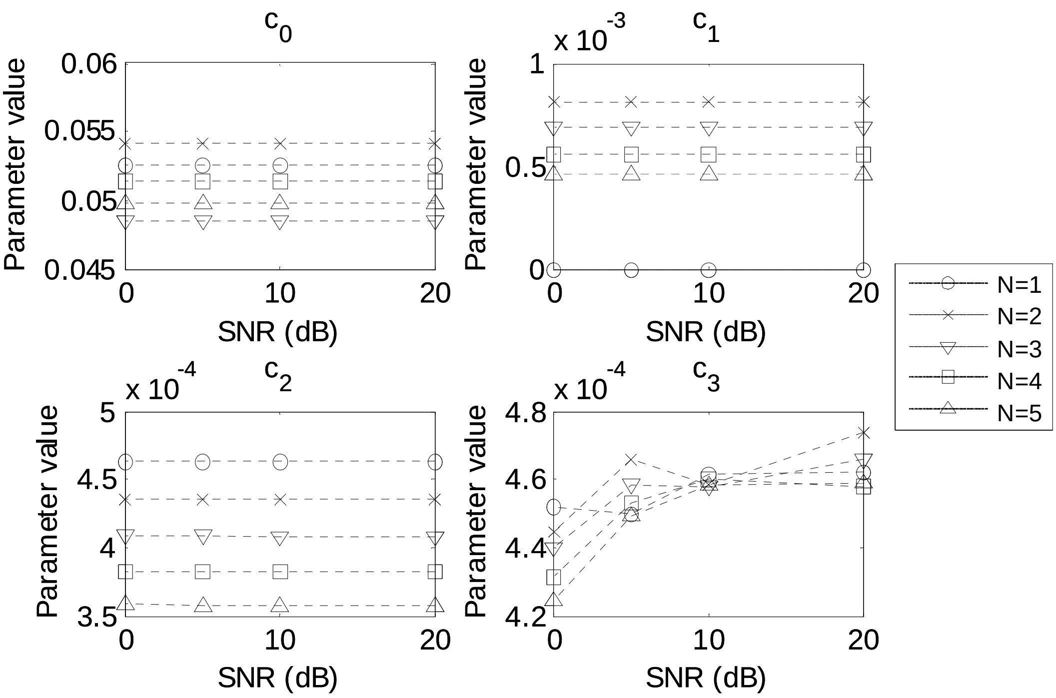

Fig. 5: The functional relationships between parameters c_i, SNR, and the number of sources. The x-axis represents SNR, while y-axis reports c_i values. Each curve stands for a different number of sources, as it is reported in the figure legend. Fixed parameters were: L = 5, M = 100 and K = 5.

4 Estimation of the number of sources

In previous sections we introduced the estimation of the number of sources, proposed our novel approach, derived analytical model and evaluate derived model with simulation studies. In this section we are going to demonstrate how our model can be used for the estimation of the number of sources. Because the approach is still in development stage, the estimation will be accomplished on synthetic signals only.

Our proposed model in Eq. (10) holds if only the correct values of parameters c_i are known. As it is depicted in Fig. 4, the main problem is that parameters c_i are not constant versus *K*. If they were, Eq. (10) would depend on 6 unknowns (*L*, *N*, c_0, c_1, c_2, c_3). But the parameters c_i depend on extension factor, which means that only by observing Eq. (10) at different values of *K* unknowns cannot be eliminated. However, our first experiment shows these parameters vary in rather limited ranges and this can help when the number of sources is estimated.

First we simulated the model, so that parameters c_i were computed exactly and inserted into Eq. (10). Such model has only two unknowns, *L* and *N*, so activity index has to be computed only at two different extension factor values, *K*. In our experiment we selected $K = 0, 1$; other parameters were set as follows: $L = 5$, $M = 100$, SNR = 20 dB. The estimation results for both *N* and *L* are reported in Table 5 .

Table 5: Estimation results for *N* and *L* if parameters c_i are known: the first column reports the actual number of sources *N*, the second the *N* estimate, and the third the *L* estimate. The actual length of system responses was 5 samples ($L = 5$).

N	*N* estimate	*L* estimate
1	1.02	4.89
2	2.02	4.94
3	3.02	4.96
4	4.02	4.97
5	5.02	4.98

Finally we attempted the estimation without knowing the exact values of parameters c_i. Instead of their exact values we used only minima and maxima of their ranges, as they are reported in Table 3 for the case with 100 measurements. Having four parameters c_i, each of them can have two possible values (the range minimum and maximum), which produces 16 possible combinations of c_i values. Having approximated parameters c_i with their minima or maxima, only two unknowns, i.e. *N* and *L*, are left, so that only two equations are needed again. Two different extension factors were chosen, $K = 0$, 1, and two quadratic equations with two unknowns, *N* and *L*, comprised the system of equations. Having two equations, we searched for their intersection (see Fig. 6, where the difference of two quadratic equations is depicted). The main problem is that surfaces, represented by quadratic equations are close to parallel, so their intersection area is large and consequently a lot of different solutions are possible. Another problem is that equations are quadratic and there are 16 possible combinations of the c_i values, therefore more solutions are possible and we have to decide, which one is the closest to the correct answer. We decided that the estimations have to meet the following requirements: they must be nonnegative and closest possible to the integer values.

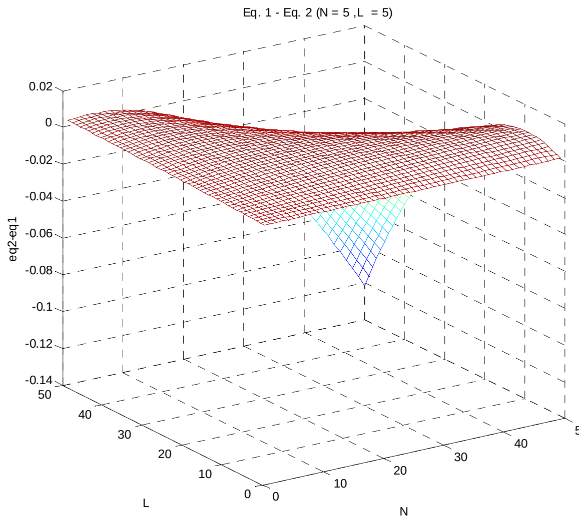

Fig. 6: The surface represents the difference of two quadratic equations that are used for the estimation of the number of sources and the length of the system responses. The area close to the zero is an equivalent for intersection. The large area that is close to zero and has moderate

5 Discussion

The proposed model, introduced in Eq. (10) holds if only the correct values of parameters c_i are known. The main problem is that parameters c_i are not constant versus *K*. If they were, Eq. (10) would depend on only on 6 unknowns. By constructing a system of equations based on the model and its derivatives, the unknowns could be recovered. But the parameters c_i depend on extension factor, which means that only by observing Eq. (10) at different values of K unknowns cannot be eliminated. However, our first experiment shows these parameters vary in rather limited ranges. This is important, because the derived model can be implemented in a search for the number of signal sources.

Some most important findings of our research are the following. The activity index level is independent of the SNR of additive Gaussian white noise, but it is proportional to the extension factor K. This is very interesting, because our method has therefore desirable noise characteristics. Moreover, from our analysis it became very probable that when the system becomes underdetermined, the number of recoverable sources, i.e. decomposed by the CKC method, converges to the correlation matrix dimensions, i.e. $M(K+1)$.

6 Conclusion

The main goal of our research was to develop an efficient and robust method for the estimation of the number of active signal sources in compound multi-channel signals. The research was based on activity index, as we found that the activity index variance is a function of the number of signal sources. The exact analytical model of the activity index variance was derived for the purpose of the estimation of the number of signal sources. Using the analytical model, the number of active signal sources can be estimated if some a priori marginal information is available, such as possible minimum and maximum triggering frequencies of sources.

Our technique can be applied in various applications, such as pre-processing of multi-channel sEMG prior to decomposition, in speech recognition systems, in estimation of direction of arrival in wireless signals, in biomedical engineering (EEG), various blind source separation algorithms, etc.

This research inspired many possible future extensions of our work. To evaluate our method, comparison to other known methods for the estimation of the number of sources on the same datasets will be carried out. All the experiments so far were conducted on synthetic signals. At the beginning stages of the research this is obvious, but for real applications the method must also be evaluated by real signals.

Acknowledgement

This work was supported by Slovenian Research Agency (Contract No. 27245 and Programme Funding No. 0796-010).

References:

[1] M. Wax and T. Kailath, Detection of Signals by Information Theoretic Criteria, *IEEE Transactions on Acoustics, Speech and Signal Processing*, Vol. 33, No. 2, 1985, pp. 387-392.

[2] J. Lou, Z. Zhang, Using Eigenvalue Grads Method to Estimate the Number of Signal Sources, in *Proceedings of 5th IEEE International Conference on Signal Processing*, 2000, pp. 223-225.

[3] H.T. Wu, J.F. Yang, and F.K. Chen, Source Number Estimator using Gerschgorin Disks, in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP),* 1994, pp. 261-264.

[4] J.-F. Gu, P. Wei and H.-M. Tai, Detection of the Number of Sources at Low Signal-to-Noise Ratio, *IET Signal Processing*, Vol. 1, No. 1, 2007, pp. 2-8.

[5] E. Fishler and H. Messer, Order Statistics Approach for Determining the Number of Sources Using an Array of Sensors, *IEEE Signal Processing Letters*, Vol. 6, No. 7, 1999, pp. 179-182.

[6] E. Fishler and H.V. Poor, Estimation of the Number of Sources in Unbalanced Arrays via Information Theoretic Criteria, *IEEE Transactions on Signal Processing*, Vol. 53, No. 9, 2005, pp. 3543-3553.

[7] A. Holobar and D. Zazula, Multichannel Blind Source Separation Using Convolution Kernel Compensation, *IEEE Transactions on Signal Processing*, Vol. 55, No. 9, 2007, pp. 4487-4496.

[8] R.K. Olsson and L.K. Hansen, Estimating the Number of Sources in a Noisy Convolutive Mixture Using BIC, in *Proceedings of 5th International Conference on Independent Component Analysis and Blind Signal Separation*, Berlin, 2004, pp. 618-625.

[9] T. Lee, M. S. Lewicki, M. Girolami and T. J. Sejnowski, Blind Source Separation of More Sources Than Mixtures Using Overcomplete Representations, *IEEE Signal Processing Letters*, Vol. 6, No. 4, 1999, pp. 87-90.

[10] H. Sawada, R. Mukai, S. Araki and S. Makino, Estimating the Number of Sources Using Independent Component Analysis, *Acoustical Science & Technology*, Vol. 26, No. 5, 2005, pp. 450-452.

[11] X. Bai and B. He, Estimation of Number of Independent Brain Electric Sources From the Scalp EEGs, *IEEE Transactions on Biomedical Engineering*, Vol. 53, No. 10, 2006, pp. 1883-1892.

[12] Q. Lv and X.-D. Zhang, A Unified Method for Blind Separation of Sparse Sources With Unknown Source Number, *IEEE Signal Processing Letters*, Vol. 13, No. 1, 2006, pp. 49-51.

[13] Y. Li, J. Wang, and A. Cichocki, Blind Source Extraction from Convolutive Mixtures in Ill-Conditioned Multi-Input Multi-Output Channels, *IEEE Transactions on Circuits and Systems*, Vol. 51, No. 9, 2004, pp. 1814-1822.

[14] K. Aczel, I. Vajk, Polyphonic music separation based on the Simplified Energy Splitter, *WSEAS Transactions on Signal Processing*, Vol. 4, No. 4, 2008, pp. 201-210.

[15] B. Gao, W.L. Woo, and S.S. Dlay, Single Channel Audio Source Separation, *WSEAS Transactions on Signal Processing*, Vol. 4, No. 4, 2008, pp. 173-182.

[16] A. Al-Qaisi, W.L. Woo, and S.S. Dlay, Novel Statistical Approach to Blind Recovery of Earth Signal and Source Wavelet using Independent Component Analysis, *WSEAS Transactions on Signal Processing*, Vol. 4, No. 4, 2008, pp. 231-240.

[17] V. Shtrauss, Nonlinear Extension of Inverse Filters for Decomposition of Monotonic Multi-Component Signals, *WSEAS Transactions on Signal Processing*, Vol. 7, No. 8, 2008, pp. 442-451.

[18] M. Borschbach and I. Hahn, Separation Capability of Overcomplete ICA Approaches, in *Proceedings of the 6th WSEAS International Conference on Signal Processing*, Dallas, Texas, USA, 2007, pp. 146-150.

[19] A. Khaparde, M. Madhavilatha, M.B.L. Manasa, P.A. Babu, S.P. Kumar, FastICA Algorithm for the Separation of Mixed Images, *WSEAS Transactions on Signal Processing*, Vol. 4, No. 5, 2008, pp. 271-278.

[20] M. Milanesi, N. Vanello, V. Positano, M.F. Santarelli, D.D. Rossi, L. Landini, Comparative Evaluation of Decomposition Algorithms based on Frequency Domain Blind Source Separation of Biomedical Signals, in *Proceedings of 7th WSEAS Int. Conf. on Mathematical Methods and Computational Techniques In Electrical Engineering*, Sofia, 2005, pp. 324-329.

[21] B. Mozaffary, M.A. Tinati, A. Aghagolzadeh, A. Erfanian, An Adaptive Algorithm for Speech Source Separation in Overcomplete Cases Using Wavelet Packets, in *Proceedings of the 5th WSEAS International Conference on Signal Processing*, Istanbul, Turkey, 2006, pp. 140-144.

[22] R. Istenic, D. Zazula, Analytical Model of the CKC-Based Activity Index Variance, in *Proceedings of the 10th WSEAS International Conference on Mathematical Methods, Computational Techniques and Intelligent Systems,* Corfu, Greece, 2008, pp. 246-249.