A COMPLEXITY MINIMIZATION APPROACH FOR ESTIMATING FADING GAUSSIAN CHANNEL IN CDMA COMMUNICATIONS

Răzvan Cristescu[1], *Jyrki Joutsensalo*[2], *Juha Karhunen*[1] *and Erkki Oja*[1]

[1]Neural Networks Research Centre, Helsinki University of Technology
P.O.Box 5400, FIN-02015 HUT, Espoo, Finland
[2]Signal Processing Laboratory, Tampere University of Technology
P.O.Box 553, FIN-33101 Tampere, Finland

razvan@mail.cis.hut.fi, jyrkij@cs.tut.fi, Juha.Karhunen@hut.fi, Erkki.Oja@hut.fi

ABSTRACT

A complexity minimization approach has been recently introduced as a generalization of standard Independent Component Analysis. There also time correlations of the source signals are taken into account in addition to the independence assumption. Generally, minimization of the complexity measure can be computationally involved, but if the sources are assumed to be Gaussian and time correlated, minimization becomes clearly simpler. In this paper, this simplified method is applied to a timely communications problem, which is estimation of the fading channel coefficients of the desired user in a code-division multiple access (CDMA) system. Simulations with downlink data, propagated through a Rayleigh fading channel, show noticeable performance gains compared with blind minimum mean-square error channel estimation, which is currently a standard method for solving this problem.

1. INDEPENDENT COMPONENT ANALYSIS

Independent component analysis (ICA) is an unsupervised statistical technique where the goal is to represent a set of random variables as a linear transformation of statistically independent component variables. Recently, ICA and the closely related blind source separation (BSS) problem have attracted a lot interest both in statistical signal processing and neural network communities. Several good algorithms utilizing higher-order statistics either directly or indirectly via suitable nonlinearities are now available for solving the basic linear ICA/BSS problem [1, 5, 4, 8, 12], and new applications of these techniques are emerging. The main reason for the success is that the independence assumption is both powerful and realistic in many practical situations, making it possible to find meaningful source signals or independent components from the data to be analyzed in a completely blind manner.

The basic data model used in ICA is as follows. Denote by $\mathbf{x}(t) = [x_1(t), \ldots, x_n(t)]^T$ the n-dimensional t:th data vector. It is assumed that the components of the vectors $\mathbf{x}(t)$ are stationary and zero mean, and have some unknown non-Gaussian distributions. In ICA, the linear expansion

$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t) + \mathbf{n}(t) = \sum_{i=1}^{m} s_i(t)\mathbf{a}_i + \mathbf{n}(t) \qquad (1)$$

is fitted to the data vectors $\mathbf{x}(t)$. Here the vector $\mathbf{s}(t) = [s_1(t), \ldots, s_m(t)]^T$ contains the m independent components or source signals $s_i(t)$ for the data vector $\mathbf{x}(t)$. $\mathbf{A} = [\mathbf{a}_1, \ldots, \mathbf{a}_m]$ is a constant full-rank $n \times m$ matrix, called the mixing matrix. The vectors $\mathbf{a}_i$, $i = 1, \ldots, m$, are the basis vectors of ICA [5, 8, 12]. The n-vector $\mathbf{n}(t)$ denotes possible additive noise. In (1), the number of independent components m is usually assumed to be at most equal to the number of mixtures n, and often $m = n$.

The independent components or source signals are found by determining an $m \times n$ inverse mapping (separating matrix) $\mathbf{W}$ so that the m-vector

$$\mathbf{y}(t) = \mathbf{W}\mathbf{x}(t) \qquad (2)$$

becomes an estimate $\mathbf{y}(t) = \hat{\mathbf{s}}(t)$ of the source (independent component) vector $\mathbf{s}(t)$. To this end, many algorithms have been proposed; for reviews and more information, see [1, 5, 4, 8, 12].

A drawback of standard ICA and BSS approaches is that they are based solely on the assumption that the underlying source signals are statistically independent. Possible time information in the source signals is completely neglected. However, in practice the source signals are usually not some non-Gaussian temporally uncorrelated signals, but their subsequent samples can typically have significant time correlations. Certain alternative blind source separation algorithms are based on utilizing these time correlations. Then second-order statistics are sufficient for solving the BSS problem [2, 10]. On the other hand, these methods do not utilize the spatial independence information, and fail if the sources are not time correlated.

Ideally, both the spatial and temporal information in the data should be exploited for getting optimal results. Some such approaches have been introduced quite recently [11, 13]. In particular, the complexity minimization approach introduced in [13] is quite interesting, because it is a true generalization of standard ICA which tries to exploit optimally all the information in the data vectors. It is shown in [13] that if the source signals are temporally white (uncorrelated), the complexity minimization approach reduces to standard ICA.

In the following, we first consider in more detail our CDMA application and its associated signal model, and then the complexity minimization approach and its use in the CDMA channel estimation problem.

2. CDMA TRANSMISSION AND SIGNAL MODEL

Code Division Multiple Access (CDMA) [16] is a data transmission technique for the third generation of mobile phones, based on spread spectrum methods. In Direct-Sequence CDMA, the narrowband signal of each user is spreaded in frequency before actual transmission via a common medium. The spreading is performed by a wideband code sequence, which uniquely identifies each user. In the reception of such a system, the final objective is to estimate the transmitted symbols. However, both code timing and channel estimation are often prerequisite tasks.

Detection of the desired user's symbols is in CDMA systems more complicated than in the simpler TDMA and FDMA systems used previously in mobile communications. This is because the spreading code sequences of different users are typically non-orthogonal, and because several users are transmitting their symbols at the same time. However, CDMA systems offer several advantages over more traditional techniques. Their capacity is larger, and it degrades gradually with increasing number of simultaneous users who can be asynchronous [16]. Therefore, CDMA has been chosen as the transmission technique for new mobile phone systems.

In the following, we briefly describe the signal model studied in this work. This type of models and the formation of the data in them are discussed in detail in [16, 9]. Our signal model is a multipath downlink model with fading channel. The data has the form:

$$r(t) = \sum_{m=1}^{N} \sum_{k=1}^{K} b_{km} \sum_{l=1}^{L} a_{lm} s_k(t - mT - d_l) + n(t) \quad (3)$$

where a_{lm} is the fading factor of the lth path corresponding to the mth symbol, b_{km} is kth user's mth symbol, $s_k(\cdot)$ is kth user's chip (spreading code) sequence, $s_k(t) \in \{-1, +1\}$, $t \in [0, T)$, $s_k(t) = 0$, $t \notin [0, T)$, d_l is the lth path's delay, which is assumed to be constant during the observation time, and $n(t)$ is the noise. The length of the chip sequence is C and N is the number of bits in the observation interval. The system is supposed to have K simultaneous users and L independent transmission paths.

We collect C-length vectors from subsequent discretized equispaced data samples $r[n]$:

$$\mathbf{r}_m = [\ r[mC]\ r[mC+1] \ldots r[(m+1)C-1]\ ]^T \quad (4)$$

Then we may write [3]

$$\mathbf{r}_m = \sum_{k=1}^{K} \sum_{l=1}^{L} [a_{l,m-1} b_{k,m-1} \underline{\mathbf{g}}_{kl} + a_{l,m} b_{k,m} \overline{\mathbf{g}}_{kl}] + \mathbf{n}_m \quad (5)$$

where $\mathbf{n}_m$ denotes the noise vector, and the 'early' and 'late' parts of the code vectors are

$$\underline{\mathbf{g}}_{kl} = [\ s_k[C - d_l + 1] \ldots s_k[C]\ 0 \ldots 0\ ]^T \quad (6)$$

$$\overline{\mathbf{g}}_{kl} = [\ 0 \ldots 0\ s_k[1] \ldots s_k[C - d_l]\ ]^T \quad (7)$$

Here d_l is a discretized index, $d_l \in \{0, \ldots, (C-1)/2\}$. The matrix $\mathbf{R} = [\mathbf{r}_1 \ldots \mathbf{r}_N]$ can be represented in compact form as

$$\mathbf{R} = \mathbf{GF} + \mathbf{N} \quad (8)$$

where $C \times 2KL$ matrix $\mathbf{G}$ contains the basis vectors

$$\mathbf{G} = [\ \underline{\mathbf{g}}_{11}, \overline{\mathbf{g}}_{11}, \ldots, \underline{\mathbf{g}}_{KL}, \overline{\mathbf{g}}_{KL}\ ] \quad (9)$$

and $2KL \times N$ matrix $\mathbf{F} = [\mathbf{f}_1 \ldots \mathbf{f}_N]$ contains the symbols and fading terms

$$\mathbf{f}_m = [a_{1,m-1} b_{1,m-1}, a_{1m} b_{1m}, \ldots, a_{L,m-1} b_{K,m-1}, a_{Lm} b_{Km}]^T \quad (10)$$

Denote by $y_i(m)$, $i = 1, \ldots, 2KL$, the independent sources $a_{1,m-1}b_{1,m-1}, \ldots, a_{Lm}b_{Km}$. Here every $2L$ sources corresponds to each user.

The noisy linear ICA/BSS model (1) can be expressed in the matrix form

$$\mathbf{X} = \mathbf{AS} + \mathbf{N} \quad (11)$$

where $\mathbf{X}$ is the data matrix having as its columns the data vectors $\mathbf{x}(1), \mathbf{x}(2), \ldots$, and $\mathbf{S}$ and $\mathbf{N}$ are similarly compiled source and noise matrices whose columns consist of the source and noise vectors $\mathbf{s}(t)$ and $\mathbf{n}(t)$, respectively. Comparing the CDMA signal model (8) with (11) shows that it has the same form as the noisy linear ICA model, so that $\mathbf{F}$ is now the matrix of source signals, $\mathbf{R}$ is the observed data matrix, and $\mathbf{G}$ is the unknown mixing matrix.

For estimating the desired user's parameters and symbols, several techniques are available [9, 16]. Matched filter [15, 16] is the simplest estimator, but it is inadequate because of the nonorthogonal chip sequences used in CDMA transmission, suffering greatly from the performance degradation caused by interfering other users (near-far problem). Maximum likelihood method [15, 16] would be optimal, but it has a very high computational load. To remedy this problem while preserving acceptable performance, subspace approaches have been proposed for example in [3]. However, they are sensitive to noise, and fail when the signal subspace dimension exceeds the processing gain. This easily occurs even with moderate system load due to multipath propagation. Some semi-blind methods proposed for the problem are discussed in the recent review [9] and in [16].

It should be noted that the CDMA estimation problem is not completely blind, because there are some prior information available. In particular, the transmitted symbols are binary (more generally from a finite alphabet), and the spreading code (chip sequence) is known. On the other hand, multipath propagation, possibly fading channels, and time delays make separation of the desired user's symbols a very challenging estimation problem which is more complicated than the standard ICA/BSS problem.

3. MINIMIZATION OF COMPLEXITY

In this paper, we propose for estimating the fading channel coefficients of the desired user in the above CDMA signal model a new method based on complexity minimization. Generally, the method proposed originally in [13] for minimizing the (Kolmogorov) complexity measure can be computationally rather demanding. However, if the source signals are assumed to be Gaussian and non-white with significant time correlations, the minimization task becomes clearly simpler. This approach is applied in the following.

We thus assume that the communication channel is Gaussian, for example Rayleigh fading [16, 15]. In this paper, we also assume that a training sequence or a preamble is available for the desired user, although this might not be the case in practice. The possibility of relaxing this assumption is discussed. On these conditions, only the desired user's contribution in the sampled data is time correlated, which is then utilized. The method has the advantage of performing code timing estimation only implicitly, and hence it does not degrade the accuracy of channel estimation.

One method developed recently for separating the unknown source signals is based on minimization of the mutual information [7, 14] of the separated signals $\mathbf{f}_m = [y_1(m) \ldots y_{2KL}(m)]^T$:

$$\mathrm{MI}(\mathbf{f}) = \sum_i H(y_i) + \log \mid \det \mathbf{G} \mid \qquad (12)$$

where $H(y_i)$ is the entropy of y_i. But one possible interpretation of entropy is that it represents the optimum averaged codelength of a random variable. Hence the mutual information can be expressed by using algorithmic complexity as [14]

$$\mathrm{MI}(\mathbf{f}) = \sum_i K(y_i) + \log \mid \det \mathbf{G} \mid \qquad (13)$$

where $K(\cdot)$ is the per-symbol Kolmogorov complexity, given by the number of bits needed to describe y_i. By using the prior information about the signals, the coding costs can be explicitly approximated. For instance, if the signals are Gaussian, then independence becomes equivalent to uncorrelatedness. Then the Kolmogorov complexity can be replaced by the per-symbol differential entropy, which in this case depends on second-order statistics only.

Our assumption is that the transmission channel exhibits Rayleigh type fading. In this case, we can formulate the prior information by considering that the probability distributions of the mutually independent $y_i(m)$ source signals have zero-mean Gaussian distributions. Suppose we want to estimate the channel coefficients of the transmission paths, by sending a given length constant $b_{1m} = 1$ symbol sequence to the desired user. We consider the signals $y_i(m)$, $i = 1, \ldots, 2L$, with i representing the indexes of the $2L$ sources corresponding to first user. Then $y_i(m)$ will actually represent the channel coefficients of all the first user's paths. Since we assume that the channel is Rayleigh fading, then these signals are Gaussian and time correlated. In this case, blind separation of the sources can be achieved by using only second-order statistics. So, we can express the Kolmogorov complexity by coding these signals using principal component analysis [14].

4. CHANNEL ESTIMATION

Let $\mathbf{y}_i(m) = [y_i(m), \ldots, y_i(m-D+1)]$ denote the vector consisting of D last samples of every such source signal $y_i(m)$, $i = 1, \ldots, 2L$. Here D is the number of delayed terms, showing what is the range of time correlations taken into account when estimating the current symbol. The information contained in any of these sources can be approximated by the codelength needed for representing the D principal components, which have variances given by the eigenvalues of the time correlation matrix $\mathbf{C}_i = \mathrm{E}[\mathbf{y}_i(m)\mathbf{y}_i^T(m)]$ [14]. Since we assume that the transmission paths are mutually independent, the overall entropy of the source is given by summing up the entropies of the principal components. Using the result that the entropy of a Gaussian random variable is given by the logarithm of the variance, we get for the entropy of each source signal

$$H(y_i) = \frac{1}{2L}\sum_k \log \sigma_k^2 = \frac{1}{2L}\log\det \mathbf{C}_i \tag{14}$$

Inserting this into the cost function (12) yields

$$\mathrm{MI}(\mathbf{f}) = \sum_i \frac{1}{2L}\log\det \mathbf{C}_i - \log \mid \det \mathbf{W} \mid \tag{15}$$

where $\mathbf{W} = \mathbf{G^{-1}}$ is the separating matrix.

The separating matrix $\mathbf{W}$ can be estimated by using a gradient descent approach for minimizing the cost function (15), leading to the update rule [14]

$$\Delta \mathbf{W} = -\mu \frac{\partial \log \mathrm{MI}(\mathbf{f})}{\partial \mathbf{W}} + \alpha \Delta \mathbf{W} \tag{16}$$

where μ is the learning rate and α is the momentum term [6] that can be introduced to avoid getting trapped into a local minimum corresponding to a secondary path.

Let $\mathbf{w}_i^T$ denote the ith row vector of the separating matrix $\mathbf{W}$. Since only the correlation matrix $\mathbf{C}_i$ of the ith source depends on $\mathbf{w}_i$, we can express the gradient of the cost function by computing the partial derivatives

$$\frac{\partial \log\det \mathbf{C}_i}{\partial w_{ik}}$$

with respect to the scalar elements of the vector $\mathbf{w}_i^T = [w_{i1}, \ldots, w_{iC}]$. For these partial derivatives, one can derive the formula [14]

$$\frac{\partial \log\det \mathbf{C}_i}{\partial w_{ik}} = 2 \,\mathrm{trace}\left(\mathbf{C}_i^{-1}\mathrm{E}\left[\mathbf{y}_i^T \frac{\partial \mathbf{y}_i}{\partial w_{ik}}\right]\right) \tag{17}$$

Since $y_i(m) = \mathbf{w}_i^T\mathbf{r}_m$, we get

$$\frac{\partial y_i}{\partial w_{ik}} = [r_{k,m}, \ldots, r_{k,m-L+1}] \tag{18}$$

where $r_{k,i}$ is the element (i, j) of the observation matrix $\mathbf{R}$ defined earlier using the formulas (4) and (8).

What is left to do now is to compute the gradient update part due to the mapping information. It can be written [14]

$$\log\det \mathbf{W} = \sum_{i=1}^{C} \log \|(\mathbf{I} - \mathbf{P}_i)\mathbf{w}_i\| \tag{19}$$

where $\mathbf{P}_i = \mathbf{W}_i(\mathbf{W}_i^T\mathbf{W}_i)^{-1}\mathbf{W}_i^T$ is a projection matrix. Now the cost function can be separated, and the different independent components can be found one by one, by taking into account the previously estimated components, contained in the subspace $\mathbf{W}_i = [\mathbf{w}_1, \ldots, \mathbf{w}_{i-1}]$.

Since our principal interest lies in the transmission path having the largest power, corresponding usually to the desired user, it is sufficient to estimate the first such independent component. In this case, the projection matrix $\mathbf{P}_1$ becomes a zero matrix. Then the overall gradient (16) for the first row $\mathbf{w}_1^T$ of the separating matrix can be written

$$\frac{\partial \log(\mathrm{MI}(\mathbf{f}))}{\partial \mathbf{w}_1^T} = \frac{1}{D}\mathrm{trace}\left(\mathbf{C}_1^{-1}\mathrm{E}\left[\mathbf{y}_1^T \frac{\partial \mathbf{y}_1}{\partial \mathbf{w}_1^T}\right]\right) - \frac{\mathbf{w}_1^T}{\|\mathbf{w}_i^T\|} \tag{20}$$

We may consider the special case when only the two last samples are considered, so that the the delay $D = 2$. First, second-order correlations are removed from the data $\mathbf{R}$ by whitening. This can be done easily in terms of standard principal components analysis as explained in [5, 8, 12]. After this preprocessing step, the subsequent separating matrix will be orthogonal, and thus the second term in Eq. (15) disappears, yielding the cost function

$$\mathrm{MI}(\mathbf{f}) \sim \sum \log\det \mathbf{C}_k \tag{21}$$

with the 2×2 autocorrelation matrices given by

$$\mathbf{C}_k = \begin{bmatrix} 1 & \mathrm{E}[y_k(m)y_k(m-1)] \\ \mathrm{E}[y_k(m)y_k(m-1)] & 1 \end{bmatrix} \tag{22}$$

In this case, the separating vectors $\mathbf{w}_i^T$ can be found by maximizing sequentially $\mathrm{E}[y_i(m)y_i(m-1) + y_i(m-1)y_i(m)]$, which is the first-order correlation coefficient of y_i. It follows that the function to be maximized is

$$J(\mathbf{w}) = \mathbf{w}^T\mathrm{E}[\mathbf{r}_m\mathbf{r}_{m-1}^T + \mathbf{r}_{m-1}\mathbf{r}_m^T]\mathbf{w} \tag{23}$$

So the separating vector $\mathbf{w}_1^T$ corresponding to the most important path is given by the principal eigenvector of the matrix in Eq. (23). We have used symmetric expression for the correlation coefficients in order to avoid unsymmetry when the observed data set is finite. Finally, we separate the desired channel coefficients by computing the quantity

$$a_{11} = \mathbf{w}^T\mathbf{R} \tag{24}$$

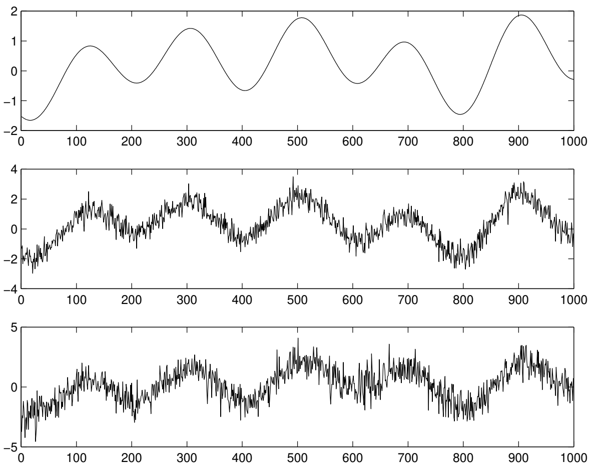

Figure 1: The original fading process (top), its estimate given by our method (middle), and estimate given by the blind MMSE method (bottom). The signal-to-noise ratio was 10 dB.

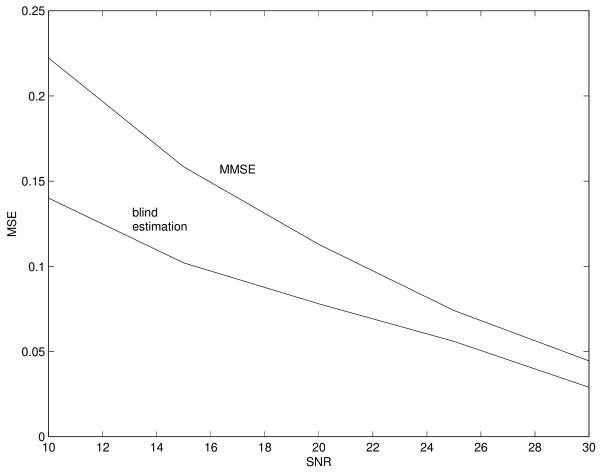

Figure 2: The mean-square errors of the MMSE method and our method as the function of the signal-to-noise ratio. The number of users was $K = 6$.

5. SIMULATIONS

We have compared the method described and derived above to a well-performing standard method used in multiuser detection, namely the Minimum Mean-Square Error (MMSE) estimator [17, 9]. In the MMSE method, the desired signal is estimated (up to scaling) using the formula

$$a_{\text{MMSE}} = \mathbf{g}_1^T \mathbf{V} \mathbf{\Lambda}^{-1} \mathbf{V}^T \mathbf{R} \tag{25}$$

where $\mathbf{\Lambda}$ and $\mathbf{V}$ are the matrices containing the eigenvalues and the respective eigenvectors of the data correlation matrix $\mathbf{R}\mathbf{R}^T/N$, and $\mathbf{g}_1$ is the code corresponding to the desired user. If the pilot signal $\mathbf{g}_1$ consists of ones, then a_{MMSE} provides estimates of the channel coefficients.

The algorithms were tested in a simulation using length $C =$ 31 quasi-orthogonal gold codes [15]. The number of users was $K = 6$, and the number of transmission paths was $L = 3$. The powers of the channel paths were -5, -5, and 0 dB respectively for every user, and the signal-to-noise ratio (SNR) varied from 30 dB to 10 dB with respect to the main path. Only the real part of the data was used. The observation interval was $N = 1000$.

We compared our algorithm with the blind MMSE method, where the pilot signal corresponding to the first user consisted of ones. The fading coefficients corresponding to the strongest path were estimated using both the methods. Figure 1 shows the original fading process and the estimated ones, giving an idea of the achieved accuracy. The figure shows that our method provides somewhat more accurate estimates than the MMSE method, though the estimated fading process is noisy. Figure 2 presents numerical values of the average mean-squared-error (MSE) as a function of SNR. The method introduced in this paper performs clearly better than the MMSE method especially at lower signal-to-noise ratios. The convergence of the gradient approach took place in this case in $10 - 15$ iterations for the learning parameters $\mu = 1$ and $\alpha = 0.5$.

6. CONCLUSIONS

In this paper we have introduced a new method for fading channel estimation based on minimization of generalized mutual information for statistically independent Gaussian signals that are temporally correlated. The proposed algorithm performs better than a widely used blind MMSE method. In its current form, the proposed method needs training symbols for providing a temporally correlated structure for the desired user's signal. If the channel varies rapidly during the training phase, the method is not able to estimate the channel as the data modulation is on. This is because the temporal correlatedness of the desired signal is lost. A future research topic is to overcome this problem.

Acknowledgment

This work has been supported by the Teletronics program of the Academy of Finland.

7. REFERENCES

[1] S.-I. Amari and A. Cichocki. Adaptive blind sig-

nal processing – neural network approaches. *Proceedings of the IEEE*, 86(10):2026–2048, October 1998.

[2] A. Belouchrani, K. Abed Meraim, J.-F. Cardoso, and E. Moulines. A blind source separation technique based on second order statistics. *IEEE Trans. on Signal Processing*, 45(2):434–444, 1997.

[3] S. Bensley and B. Aazhang. Subspace-based channel estimation for code division multiple access communication systems. *IEEE Trans. on Communications*, 44(8):1009–1020, August 1996.

[4] A. Nandi (Ed.). *Blind Estimation Using Higher-Order Statistics*. Kluwer, 1999.

[5] M. Girolami. *Self-Organising Neural Networks - Independent Component Analysis and Blind Source Separation*. Springer-Verlag, 1999.

[6] S. Haykin. *Neural Networks - A Comprehensive Foundation, 2nd ed.* Prentice-Hall, Englewood Cliffs, NJ, 1998.

[7] A. Hyvärinen. Independent component analysis by minimization of mutual information. Technical Report A46, Helsinki Univ. of Technology, Lab. of Computer and Information Science, 1997.

[8] T.-W. Lee. *Independent Component Analysis – Theory and Applications*. Kluwer, 1998.

[9] U. Madhow. Blind adaptive interference suppression for direct-sequence CDMA. *Proceedings of the IEEE*, 86(10):2049–2069, October 1998.

[10] L. Molgedey and H. Schuster. Separation of a mixture of independent signals using time delayed correlations. *Physical Review Letters*, 72:3634–3636, 1994.

[11] K.-R. Müller, P. Philips, and A. Ziehe. $JADE_{TD}$: Combining higher-order statistics and temporal information for blind source separation (with noise). In *Proc. Int. Workshop on Independent Component Analysis and Signal Separation (ICA'99)*, pages 87–92, Aussois, France, January 1999.

[12] E. Oja, J. Karhunen, A. Hyvärinen, R. Vigario, and J. Hurri. Neural independent component analysis - approaches and applications. In *Brain-Like Computing and Intelligent Information Systems*, pages 167–188. Springer-Verlag, Singapore, 1997.

[13] P. Pajunen. Blind source separation using algorithmic information theory. *Neurocomputing*, 22:35–48, 1998.

[14] P. Pajunen. Blind source separation of natural signals based on approximate complexity minimization. In *Proc. First Int. Conf. on Independent Component Analysis and Blind Source Separation (ICA'99)*, pages 267–270, Aussois, France, January 1999.

[15] J. Proakis. *Digital Communications, 3rd edition*. McGraw-Hill, 1995.

[16] S. Verdu. *Multiuser Detection*. Cambridge Univ. Press, Cambridge, United Kingdom, 1998.

[17] X. Wang and H.V.Poor. Blind multiuser detection: A subspace approach. *IEEE Trans. Information Theory*, 44(2):667–690, March 1998.