Adaptive MCA-Matched Filter Algorithms for Binary Detection

*Francisco Messina† and Bruno Cernuschi-Frías†‡

†Facultad de Ingeniería, Universidad de Buenos Aires, Argentina
fmessina@fi.uba.ar
‡IAM-CONICET, Buenos Aires, Argentina
bcf@ieee.org

Abstract. In this work, we present a method for signal-to-noise ratio maximization using a linear filter based on minor component analysis of the noise covariance matrix. As we will see, the greatest benefits are obtained when both filter and signal design are treated as a single problem. This general problem is then related to the minimization of the probability of error of a digital communication. In particular, the classical binary detection problem is considered when nonstationary and (possibly) nonwhite additive Gaussian noise is present. Two algorithms are given to solve the problem at hand with cuadratic and linear computational complexity with respect to the dimension of the problem.

Key words: Minor Component Analysis, Matched Filter, Optimal Signal Design, Binary Detection, Adaptive Algorithms.

1 Introduction

The matched filter is a very well known linear technique for signal-to-noise ratio (SNR) maximization of a signal embedded in additive noise [1]. It may also be combined with optimal signal design, in which case the problem is reduced to an eigenvalue problem. In particular, its solution is the eigenvector associated with the minimum eigenvalue of the noise covariance matrix. Typically, this matrix is not known in practice. A direct approach may involve then a spectral decomposition of the sample covariance matrix, but in many applications this is computationally intractable. This is specially so when the underlying noise process is nonstationary.

Minor component analysis (MCA) is a linear statistical technique that may be used to obtain the direction of minimum variance of a random vector. It is related to principal component analysis (PCA), but in the latter the maximum variance direction would be obtained instead [2]. In particular, when MCA is applied to noise, it can be used to solve both the matched filter and optimal

* This work was partially supported by the University of Buenos Aires, CONICET and ANPCyT.

signal design problem. Thus, iterative algorithms for MCA may be applied to the problem of SNR adaptive maximization.

The mentioned approach can be taken for reliable communication in a digital channel, since the probability of symbol error is basically a function of the SNR when the noise is Gaussian [3]. In this paper, we will exploit this connection in the simple binary detection problem with nonstationary additive Gaussian noise.

This paper is divided as follows. Section 2 presents a review of the matched filter and the optimal signal design problems with a signal-space point of view, culminating in an eigenvalue problem. Then, in section 3 we show explicitly the connection of this problem with the binary detection problem. To solve this adaptively, we present two algorithms in section 4 with different computation requirements. A convergence analysis of both algorithms is undertaken for the stationary case in section 5. Computer simulations are then presented for a stationary and nonstationary problem in section 6. The conclusions of the work are finally given in section 7.

2 Matched Filter and Optimal Signal Problem

The classical matched filter concept was introduced by North in 1963 [4], who considered a continous-time problem. Here, we take the well-known signal-space approach of digital communications [3], which allow us to pose the problem in terms of vectors and matrices (we are ignoring the noise that can not be represented in the signal space). In fact, let $\mathbf{s} \in \mathbb{R}^N$ be a deterministic signal that conveys information and is perturbed by stochastic additive noise $\mathbf{n}$ with zero mean and covariance matrix $\boldsymbol{C}_{\mathbf{n}}$. Note that we make no assumptions on the noise distribution in this section: it may be non-Gaussian. The observed signal is then given by

$$\mathbf{x} = \mathbf{s} + \mathbf{n}. \tag{1}$$

Our first objective is to design a filter $\mathbf{h}$ that yields, at the appropiate observation time, the maximum value for the output SNR [1] (the signal-to-noise ratio is defined as the ratio of the power of the signal component to the average power of the noise component). The response of the filter $\mathbf{h}$ to the signal $\mathbf{x}$ at this time may be written as

$$y = \mathbf{h}^T\mathbf{x} = \mathbf{h}^T\mathbf{s} + \mathbf{h}^T\mathbf{n} = y_{\mathbf{s}} + y_{\mathbf{n}}, \tag{2}$$

where $y_{\mathbf{s}}$ and $y_{\mathbf{n}}$ are the signal and noise components of the filter output y, respectively. Observe that $y_{\mathbf{n}}$ is a random variable whereas $y_{\mathbf{s}}$ is a deterministic function of $\mathbf{h}$. The output SNR is simply

$$SNR_o = \frac{|\mathbf{h}^T\mathbf{s}|^2}{\mathbf{h}^T\boldsymbol{C}_{\mathbf{n}}\mathbf{h}}. \tag{3}$$

Now, if the matrix $\boldsymbol{C}_{\mathbf{n}}$ is positive definite (if this is not the case, we can always find a direction for $\mathbf{h}$ such that the variance of $y_{\mathbf{n}}$ is zero, yielding an infinite SNR at the output of the filter), we can rewrite this equation as

$$SNR_o = \frac{|\mathbf{h}^T \boldsymbol{C}_\mathbf{n} \boldsymbol{C}_\mathbf{n}^{-1}\mathbf{s}|^2}{\mathbf{h}^T \boldsymbol{C}_\mathbf{n}\mathbf{h}} = \frac{\left|\langle \mathbf{h}, \boldsymbol{C}_\mathbf{n}^{-1}\mathbf{s}\rangle\right|^2}{\langle \mathbf{h}, \mathbf{h}\rangle}, \tag{4}$$

where we have defined the inner product $\langle \mathbf{x}, \mathbf{y}\rangle = \mathbf{x}^T \boldsymbol{C}_\mathbf{n}\mathbf{y}$ between two vectors $\mathbf{x}$ and $\mathbf{y}$ in $\mathbb{R}^N$. Applying the well known Cauchy-Schwarz inequality $|\langle \mathbf{x}, \mathbf{y}\rangle|^2 \leq \langle \mathbf{x}, \mathbf{x}\rangle \langle \mathbf{y}, \mathbf{y}\rangle$ to the equation (4), we get

$$SNR_o = \frac{\left|\langle \mathbf{h}, \boldsymbol{C}_\mathbf{n}^{-1}\mathbf{s}\rangle\right|^2}{\langle \mathbf{h}, \mathbf{h}\rangle} \leq \langle \boldsymbol{C}_\mathbf{n}^{-1}\mathbf{s}, \boldsymbol{C}_\mathbf{n}^{-1}\mathbf{s}\rangle = \mathbf{s}^T \boldsymbol{C}_\mathbf{n}^{-1}\mathbf{s}, \tag{5}$$

since $\boldsymbol{C}_\mathbf{n}^{-1}$ is a symmetric matrix. The maximum value of the output SNR is attained if and only if the filter is collinear with $\boldsymbol{C}_\mathbf{n}^{-1}\mathbf{s}$. Thus, the optima filters are

$$\mathbf{h}^* = \alpha \boldsymbol{C}_\mathbf{n}^{-1}\mathbf{s}, \tag{6}$$

where $\alpha \in \mathbb{R}$ is an arbitrary nonzero constant.

We can still go one step further and ask for the signal that maximizes the output SNR, while using the matched filter. This is a problem of optimal signal design. It is obvious from equation (5) that the problem must be constrained in some way, because the output SNR is directly proportional to the signal energy $\mathcal{E}_\mathbf{s} = \mathbf{s}^T\mathbf{s}$. To this purpose, we formulate the following constrained optimization problem:

$$\begin{aligned} \mathbf{s}^* &= \underset{\mathbf{s}\in\mathbb{R}^N}{\arg\max}\ \mathbf{s}^T \boldsymbol{C}_\mathbf{n}^{-1}\mathbf{s} \\ &\text{s.t. } \mathbf{s}^T\mathbf{s} = \mathcal{E}_\mathbf{s}. \end{aligned} \tag{7}$$

This problem is equivalent to the Rayleigh qoutient of the inverse of the noise covariance matrix, given by

$$R(\mathbf{s}) = \frac{\mathbf{s}^T \boldsymbol{C}_\mathbf{n}^{-1}\mathbf{s}}{\mathbf{s}^T\mathbf{s}}. \tag{8}$$

In fact, a well-known property of R is that its maximum value is given by the maximum eigenvalue of $\boldsymbol{C}_\mathbf{n}^{-1}$, and is obtained when $\mathbf{s}$ equals the corresponding eigenvector [5]. This is equivalent to finding the mininum eigenvalue of $\boldsymbol{C}_\mathbf{n}$ and its corresponding eigenvector, which is intuitively appealing: the best signal is in the direction of minimum noise power. Note also that when the optimum signal is used, the matched filter will be in the same direction. Then, the matched filter problem (at the receiver) and the optimum signal design (at the transmitter) are the same.

Finally, the maximum possible output SNR, obtained by setting $\mathbf{s}$ and $\mathbf{h}$ equal to the eigenvector of $\boldsymbol{C}_\mathbf{n}$ corresponding to its minimum eigenvalue $\lambda_{\min}$, is

$$SNR_o^* = \frac{\mathcal{E}_{\mathbf{s}}}{\lambda_{\min}}. \tag{9}$$

This result is of sum importance in digital communications, where the performance measure (i.e., the probability of symbol or bit error) depends critically in the effective signal-to-noise ratio at the receiver. We now present a review of the classical binary detection problem to show this connection.

3 Binary Detection Problem

Let $\mathcal{S} = \{-\mathbf{s}, +\mathbf{s}\}$ be the symbol constellation of a BPSK modulation scheme [3], where $\mathbf{s} \in \mathbb{R}^N$ is a signal to be determined. Suppose that the symbols $-\mathbf{s}$ and $\mathbf{s}$ represent the bits 0 and 1, respectively. Now, when one symbol is transmitted over a communication channel that introduces additive zero-mean Gaussian noise (not necessarily white), the received signal responds to either

$$\mathbf{x} = -\mathbf{s} + \mathbf{n}, \tag{10}$$

or

$$\mathbf{x} = \mathbf{s} + \mathbf{n}, \tag{11}$$

where $\mathbf{n} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{C}_{\mathbf{n}})$.

We define $\mathcal{H}_i$ as the hypothesis of the bit i being transmitted, with $i \in \{0, 1\}$. The optimal decision rule for minimizing the (symbol or bit) probability of error is given by the *maximum a posteriori* (MAP) rule [6], which can be compactly stated as

$$\frac{\mathbb{P}(\mathcal{H}_1|\mathbf{x})}{\mathbb{P}(\mathcal{H}_0|\mathbf{x})} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} 1. \tag{12}$$

By Bayes' theorem, this can be rewritten as

$$\Lambda(\mathbf{x}) = \frac{p(\mathbf{x}|\mathcal{H}_1)}{p(\mathbf{x}|\mathcal{H}_0)} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \frac{\mathbb{P}(H_0)}{\mathbb{P}(H_1)} = \gamma, \tag{13}$$

which is the simple likelihood ratio test (LRT), being $\Lambda(\mathbf{x})$ the likelihood ratio and γ a decision treshold. A more convenient form for the decision rule is given by the log-likelihood ratio test (LLRT):

$$\log \Lambda(\mathbf{x}) \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \log \gamma. \tag{14}$$

In this case, the log-likelihood takes the simple form

$$\log \Lambda(\mathbf{x}) = 2\, \mathbf{s}^T \boldsymbol{C}_{\mathbf{n}}^{-1} \mathbf{x} = 2\, \mathbf{h}^T \mathbf{x}, \tag{15}$$

where $\mathbf{h}$ is recognized as a matched filter. Finally, the decision rule can be written as

$$\mathbf{h}^T\mathbf{x} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \frac{1}{2}\log\gamma. \tag{16}$$

The probability of error, that gives the performance of the optimal detector, is given by

$$\mathbb{P}_e = \mathbb{P}\left(\mathbf{h}^T\mathbf{x} > \delta|\mathcal{H}_0\right)\mathbb{P}(\mathcal{H}_0) + \mathbb{P}\left(\mathbf{h}^T\mathbf{x} < \delta|\mathcal{H}_1\right)\mathbb{P}(\mathcal{H}_1), \tag{17}$$

with $\delta = \frac{1}{2}\log\gamma$. Letting $\rho = \mathbf{s}^T\boldsymbol{C}_\mathbf{n}^{-1}\mathbf{s}$, we observe that $\mathbf{h}^T\mathbf{x}|\mathcal{H}_0 \sim \mathcal{N}(-\rho,\rho)$ and $\mathbf{h}^T\mathbf{x}|\mathcal{H}_1 \sim \mathcal{N}(\rho,\rho)$. Therefore, we have

$$\mathbb{P}_e = \mathbb{P}(\mathcal{H}_0)Q\left(\frac{\delta+\rho}{\sqrt{\rho}}\right) + \mathbb{P}(\mathcal{H}_1)Q\left(\frac{\rho-\delta}{\sqrt{\rho}}\right), \tag{18}$$

where $Q(x) = \int_x^{+\infty}\frac{1}{\sqrt{2\pi}}\exp(-u^2/2)du$ is known as the Q-function. A fundamental property of the Q-function is that it is a decreasing function of its argument. Thus, the equation (18) shows that, in order to minimize the probabiliy of error, we must maximize ρ by choosing $\mathbf{s}$ appropiately. Since the energy of $\mathbf{s}$ is typically fixed by power contraints of the system, we must consider only the direction in the optimization. This is the optimal signal design problem already encountered in section 2, which answer is

$$\mathbf{s}^* = \sqrt{\mathcal{E}_\mathbf{s}}\mathbf{v}_{\min} \tag{19}$$

where $\mathbf{v}_{\min}$ is the eigenvector of the noise covariance matrix associated with its minimum eigenvalue $\lambda_{\min}$. Then, the optimum performance is

$$\mathbb{P}_e^* = \mathbb{P}(\mathcal{H}_0)Q\left(\frac{\delta+\rho^*}{\sqrt{\rho^*}}\right) + \mathbb{P}(\mathcal{H}_1)Q\left(\frac{\rho^*-\delta}{\sqrt{\rho^*}}\right). \tag{20}$$

with $\rho^* = \frac{\mathcal{E}_\mathbf{s}}{\lambda_{\min}}$.

In the following section, we consider the nonstationary scenario and present algorithms for the adaptive maximization of the effective signal-to-noise ratio in two different situations that determine the computation capabilities of the system.

4 Adaptive Algorithms

The results developed in the previous sections can be translated to a nonstationary scenario in a straightforward manner. Specifically, if the noise covariance matrix $\boldsymbol{C}_\mathbf{s}$ depends on a time index k, so will its eigenvalues and eigenvectors. This implies that the optimal signal, the matched filter and the performance will vary with k. The problem arises when this statistical variation is unknown, and must be "tracked" in some way.

There is a fundamental hypothesis that we will make in order to develop useful algorithms. Namely, we will consider a full-duplex symmetric channel in the communication, so that both transceivers can "observe" the noise statistic changes and use this information to udpate both the signal $\mathbf{s}$ and the filter $\mathbf{h}$. This means also that the algorithm must be used by the two communication parties so that both of them share the same "language".

Many MCA algorithms can be derived from a gradient descent approach applied to the Rayleigh quotient of the noise covariance matrix [7]:

$$R(\mathbf{s}) = \frac{\mathbf{s}^T \boldsymbol{C}_{\mathbf{n}} \mathbf{s}}{\mathbf{s}^T \mathbf{s}}. \tag{21}$$

Following this idea, we obtain the following update rule for the signal (and the matched filter):

$$\mathbf{s}(k+1) = \mathbf{s}(k) - \eta(k) \left. \frac{\partial R(\mathbf{s}, \boldsymbol{C}_{\mathbf{n}})}{\partial \mathbf{s}} \right|_{\mathbf{s}=\mathbf{s}(k), \boldsymbol{C}_{\mathbf{n}}=\boldsymbol{C}_{\mathbf{n}}(k)}, \tag{22}$$

where $\eta(k) > 0$ is the learning rate of the algorithm at time k. From the definition of equation (21), the partial derivative can be easily evaluated, yielding

$$\mathbf{s}(k+1) = \mathbf{s}(k) - \eta(k) \left\{ \frac{[\boldsymbol{C}_{\mathbf{n}}(k)\mathbf{s}(k)]\, \mathbf{s}^T(k)\mathbf{s}(k) - \left[\mathbf{s}^T(k)\boldsymbol{C}_{\mathbf{n}}(k)\mathbf{s}(k)\right] \mathbf{s}(k)}{\left[\mathbf{s}^T(k)\mathbf{s}(k)\right]^2} \right\}. \tag{23}$$

Note that a factor of 2 has been absorbed by $\eta(k)$. The quantity $\mathbf{s}^T(k)\mathbf{s}(k)$ is equal to the signal energy and does not depend on k. This will be proven later with the converge analysis in section 5 for the stationary case, but in general it can also be considered as a consequence of doing and explicit normalization after each iteration step. Thus, the update rule simplifies to

$$\mathbf{s}(k+1) = \mathbf{s}(k) - \eta(k) \left[\boldsymbol{C}_{\mathbf{n}}(k)\mathbf{s}(k) - \frac{\mathbf{s}^T(k)\boldsymbol{C}_{\mathbf{n}}(k)\mathbf{s}(k)}{\mathbf{s}^T(k)\mathbf{s}(k)} \mathbf{s}(k) \right]. \tag{24}$$

Again, all constants are absorbed by $\eta(k)$. In the following algorithms, $\boldsymbol{C}_{\mathbf{n}}(k)$ will be replaced by the actual estimation of the noise covariance matrix.

Before proceeding further, we shall distinguish two different cases corresponding to "low" and "high" data rates. The interpretation of the adjectives low and high shall be in terms of the computational capabilities of the system and will become precise in the subsequent discussion of the corresponding adaptive algorithms. Consider first the low data rate scenario, in which the covariance matrix may be estimated recursively as

$$\hat{\boldsymbol{C}}_{\mathbf{n}}(k) = (1-\alpha)\, \hat{\boldsymbol{C}}_{\mathbf{n}}(k-1) + \alpha\, \mathbf{n}(k)\mathbf{n}^T(k), \tag{25}$$

where α is some constant parameter between 0 and 1 that provides the algorithm's adaptivity. We now present the complete algorithm.

Algorithm 1 *MCA Detection for Low Data Rate.*

1. Let $\hat{\boldsymbol{C}}_{\mathbf{n}}(0)$ be the initial estimate or guess of the noise covariance matrix. Set $\mathbf{s}(0)$ and $\mathbf{h}(0)$ equal to its eigenvector associated with its minimum eigenvalue and let $\sqrt{\mathcal{E}_s}$ be the norm of $\mathbf{s}(0)$.

2. Set $k := 0$ and choose both the learning rate $\eta(k)$ and the paramater α.
3. For the received signal $\mathbf{x}(k)$, detect the symbol according to equation (16) using $\mathbf{h}(k)$. This produces a symbol estimation, say $\hat{\mathbf{u}}(k)$. Now, obtain the noise vector as $\mathbf{n}(k) = \mathbf{x}(k) - \hat{\mathbf{u}}(k)$ and perform the following computations:

$$\hat{\boldsymbol{C}}_{\mathbf{n}}(k) = (1-\alpha)\,\hat{\boldsymbol{C}}_{\mathbf{n}}(k-1) + \alpha\,\mathbf{n}(k)\mathbf{n}^T(k), \tag{26}$$

$$\mathbf{s}(k+1) = \mathbf{s}(k) - \eta(k)\left[\hat{\boldsymbol{C}}_{\mathbf{n}}(k)\mathbf{s}(k) - \frac{\mathbf{s}^T(k)\hat{\boldsymbol{C}}_{\mathbf{n}}(k)\mathbf{s}(k)}{\mathbf{s}^T(k)\mathbf{s}(k)}\mathbf{s}(k)\right], \tag{27}$$

$$\mathbf{s}(k+1) = \sqrt{\mathcal{E}_s}\frac{\mathbf{s}(k+1)}{\sqrt{\mathbf{s}^T(k+1)\mathbf{s}(k+1)}} \tag{28}$$

$$\mathbf{h}(k+1) = \mathbf{s}(k+1). \tag{29}$$

4. If a new symbols is received, set $k := k+1$ and go to step 3.

Note that we have added a normalization step to ensure that the signal energy remains constant over time. This is necessary even if theoretically the norm of $\mathbf{s}(k)$ (see section 5) because of numerical errors. The algorithm's computational complexity per time symbol is $O(N^2)$, where N is the dimension of the underlying signal space. This may be prohibitive in high data rate scenarios. In order to reduce the complexity, a simple instantaneous estimation of the noise covariance matrix may be used. That simplification produces the second algorithm.

Algorithm 2 *MCA Detection for High Data Rate.*

1. Set $\mathbf{s}(0)$ and $\mathbf{h}(0)$ equal to a random vector pointing with equal probability in any direction and let $\sqrt{\mathcal{E}_s}$ be the norm of $\mathbf{s}(0)$.
2. Set $k := 0$ and choose the learning rate $\eta(k)$.
3. For the received signal $\mathbf{x}(k)$, detect the symbol according to equation (16) using $\mathbf{h}(k)$. This produces a symbol estimation, say $\hat{\mathbf{u}}(k)$. Now, obtain the noise vector as $\mathbf{n}(k) = \mathbf{x}(k) - \hat{\mathbf{u}}(k)$ and perform the following computations:

$$y(k) = \mathbf{n}^T(k)\mathbf{s}(k), \tag{30}$$

$$\mathbf{s}(k+1) = \mathbf{s}(k) - \eta(k)\left[y(k)\mathbf{n}(k) - \frac{y^2(k)}{\mathbf{s}^T(k)\mathbf{s}(k)}\mathbf{s}(k)\right], \tag{31}$$

$$\mathbf{s}(k+1) = \sqrt{\mathcal{E}_s}\frac{\mathbf{s}(k+1)}{\sqrt{\mathbf{s}^T(k+1)\mathbf{s}(k+1)}} \tag{32}$$

$$\mathbf{h}(k+1) = \mathbf{s}(k+1). \tag{33}$$

4. If a new symbols is received, set $k := k+1$ and go to step 3.

The simple algorithm just described possess linear complexity (i.e, it is $O(N)$). We now present a brief convergence analysis of these algorithms for the case where the noise is a stationary random process.

5 Convergence Analysis

The presented algorithms are formally classsified as stochastic approximation algorithms [8]. Its convergence analysis is typically based on the averaged ordinary differential equation (ODE) associated with its update rule, obtained by simply taking the expected value and letting the time index be a continuous variable.

For the first algorithm, we shall study the update rule for the noise covariance matrix estimation. This can be done directly taking expectation on both sides of equation (26):

$$\boldsymbol{A}(k) = (1-\alpha)\,\boldsymbol{A}(k-1) + \alpha\,\boldsymbol{C}_{\mathbf{n}} \tag{34}$$

where $\boldsymbol{A}(k) = \mathbb{E}[\hat{\boldsymbol{C}}_{\mathbf{n}}(k)]$. This is a very simple matrix linear difference equation, which can be solved component by component to obtain that $\boldsymbol{A}(k) \to \boldsymbol{C}_{\mathbf{n}}$ for all value of $\alpha \in (0,1)$. Therefore, we can present a proof of the convergence of both algorithms based on the following ODE:

$$\frac{d\mathbf{s}(t)}{dt} = -\eta(t)\left[\boldsymbol{C}_{\mathbf{n}}\mathbf{s}(t) - \frac{\mathbf{s}^T(t)\boldsymbol{C}_{\mathbf{n}}\mathbf{s}(t)}{\mathbf{s}^T(t)\mathbf{s}(t)}\mathbf{s}(t)\right] \tag{35}$$

As mentioned in section 4, it can easily be shown that the norm of $\mathbf{s}(t)$ is a constant over time [7]:

$$\frac{d\|\mathbf{s}(t)\|^2}{dt} = 2\mathbf{s}^T(t)\frac{d\mathbf{s}(t)}{dt} = -2\eta(t)\left[\mathbf{s}^T(t)\boldsymbol{C}_{\mathbf{n}}\mathbf{s}(t) - \frac{\mathbf{s}^T(t)\boldsymbol{C}_{\mathbf{n}}\mathbf{s}(t)}{\mathbf{s}^T(t)\mathbf{s}(t)}\mathbf{s}^T(t)\mathbf{s}(t)\right] = 0 \tag{36}$$

Thus, $\|\mathbf{s}(t)\| = \|\mathbf{s}(0)\|$ and we can take it equal to 1 without loss of generality. To analyze the convergence, we write $\mathbf{s}(t)$ in terms of the eigenvectors $\mathbf{v}_i$ of the noise covariance matrix as (this eigenvectors span all $\mathbb{R}^N$ if $\boldsymbol{C}_{\mathbf{n}}$ is positive definite)

$$\mathbf{s}(t) = \sum_{i=1}^{N} \alpha_i(t)\mathbf{v}_i \tag{37}$$

where $\alpha_i(t)$ is the projection of $\mathbf{s}(t)$ onto $\mathbf{v}_i$. Using equation (37) in the ODE (35), we get the following differential equations for the projections:

$$\frac{d\alpha_i(t)}{dt} = \alpha_i(t)\left(\lambda_i - \frac{\sum_{j=1}^{N}\alpha_j^2(t)\lambda_j}{\sum_{j=1}^{N}\alpha_j^2(t)}\right) \tag{38}$$

If now we define the variables

$$\beta_i(t) = \frac{\alpha_i(t)}{\alpha_N(t)} \tag{39}$$

for $i = 1, \ldots, N$, we can show that

$$\frac{d\beta_i(t)}{dt} = -\beta_i(t)(\lambda_i - \lambda_N) \quad (40)$$

Thus, assuming that all eigenvalues are different and ordered as $\lambda_1 > \ldots > \lambda_N$, we have that $\beta_i(t) \to 0$ for all $i = 1, \ldots, N-1$, which implies that $\alpha_i(t) \to 0$ since the norm of $\mathbf{s}(t)$ is constant over time. Then, it must be that $\alpha_N(t) \to \pm 1$, that is, only the minor component remains in the limit. This completes the convergence proof.

6 Simulations

We performed two relatively simple experiments to show the performance of the proposed algorithms. The first one consists of a channel with stationary additive Gaussian noise, which covariance matrix is given by

$$\boldsymbol{C}_\mathbf{n} = \sigma^2 \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \quad (41)$$

where σ^2 is the variance of the noise components and ρ is the correlation between them which we will consider to be positive (analogous results are obtained when $\rho < 0$) so that $0 < \rho < 1$. The eigendecomposition of this matrix yields the factorization

$$\boldsymbol{C}_\mathbf{n} = \sigma^2 \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} 1+\rho & 0 \\ 0 & 1-\rho \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \quad (42)$$

which shows that the minimum noise power is given by $\sigma^2(1-\rho)$ and is obtained in the direction $[1, -1]^T$. In contrast, the maximum noise power is $\sigma^2(1+\rho)$ and is obtained in the direction $[1, 1]^T$. Thus, for a highly correlated noise, the benefits of our approach will be more remarkable. Moreover, the convergence speed of the gradient based algorithms also depends on ρ, since a greater (absolute) value of this parameter implies a more steepest form of the Rayleigh quotient. We have decided to choose a moderate value of $\rho = 1/2$.

The initial estimate for the covariance matrix was taken as

$$\hat{\boldsymbol{C}}_\mathbf{n}(0) = \sigma^2 \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad (43)$$

which means that the prior belief is that the channel is AWGN. The initialization of the signal vector $\mathbf{s}$ (and filter $\mathbf{h}$) was then randomly chosen with equal probability in all directions (note that in this case both algorithms do the same initialization procedure). Values used for the other parameters are $\mathcal{E}_s = 1, \sigma^2 = 10^{-3}, \alpha = 0.01, \mathbb{P}(\mathcal{H}_0) = \mathbb{P}(\mathcal{H}_1) = 0.5$.

In figure 1, we show the evolution of the standard inner product between the minor component direction and $\mathbf{s}(k)$ (which is simply the cosine of the angle between them) for different values of η while using algorithm 1. These graphics show that the convergence of this algorithm to the desired solution is practically insensitive to the value of the learning rate parameter. Nevertheless, the speed of convergence is strongly influenced by it (note that different number of iterations are shown in each graph for ease of visualization). Of course, it is possible to let the learning rate sequence take a great value in the first iterations and then reduce it to a small constante value to combine the advantages of the two choices. In the stationary case, we can even let $\eta(k)$ decay to zero, forcing the estimations to freeze after some criterion of convergence is met.

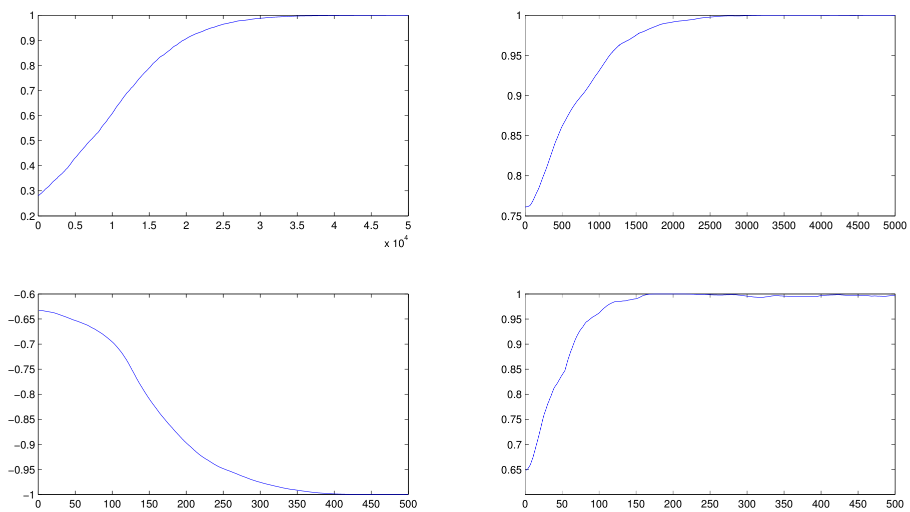

Fig. 1. Cosine of angle between the minor eigenvector of the noise covariance matrix and $\mathbf{s}(k)$ versus k for various values of the learning rate parameter η using algorithm 1. (a) $\eta = 10^{-1}$, (b) $\eta = 10^{0}$, (c) $\eta = 10^{1}$, (d) $\eta = 10^{2}$.

Using the second algorithm with the same learning rates, we obtained figure 2. Similar convergence conditions are observed, but in this case larger values for η produce larger oscillations in both the transient behaviour and in the "final value".

We now proceed to the second experiment in which the noise covariance matrix is going to be time-varying and given by

$$\boldsymbol{C}_{\mathbf{n}}(k) = \sigma^2 \begin{bmatrix} 1 & \rho(k) \\ \rho(k) & 1 \end{bmatrix} \tag{44}$$

As we have seen, the eigenvector associated with the minimum eigenvalue of this matrix is always either in the direction $[1, 1]^T$ (for $\rho(k) < 0$) or $[-1, 1]^T$ (for

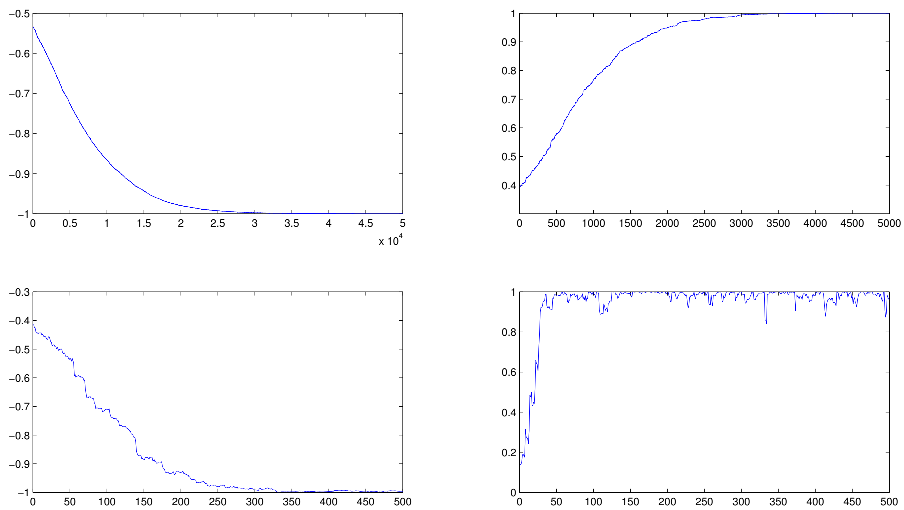

Fig. 2. Cosine of angle between the minor eigenvector of the noise covariance matrix and $\mathbf{s}(k)$ versus k for various values of the learning rate parameter η using algorithm 2. (a) $\eta = 10^{-1}$, (b) $\eta = 10^{0}$, (c) $\eta = 10^{1}$, $\eta = 10^{2}$.

$\rho(k) < 0$). If $\rho(k) = 0$ the noise is white at time k and any direction has equal variance. The secuence of correlation coefficients is set as

$$\rho(k) = \sin(\omega k) \tag{45}$$

where ω gives the frequency of this sequence and thus can be considered as the degree of nonstationarity . We let $\omega = \frac{2\pi}{L}$, being L the number of iterations performed in the experiment. The results of both algorithms using the same parameters as before (but with different learning rates) are shown in figure 3. It is seen that the first algorithm learns much faster the abrupt change that occurs in the minor component direction.

7 Conclusions

We have given a novel approach for signal-to-noise ratio maximization of a signal corrupted by additive noise and processed by a linear filter that is based on minor component analysis. The underlying idea here is to exploit the noise correlation. Then, we have shown how this relates to the probability of error of a digital communication, making special emphasis on the binary detection problem where this idea can be directly applied. This problem was then posed in terms of the Rayleigh quotient and two simple algorithms using gradient descent were developed with different computation requirements.

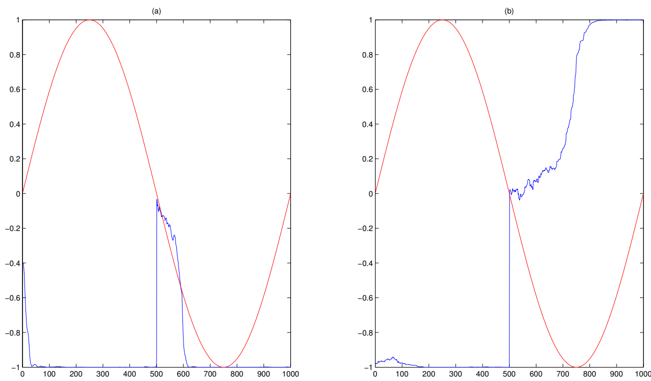

Fig. 3. Cosine of angle between the minor eigenvector of the noise covariance matrix and $\mathbf{s}(k)$ versus k for nonstationary Gaussian noise (in blue) and $\rho(k)$ (in red). (a) Algorithm 1 with $\eta = 10^3$, (b) Algorithm 2 with $\eta = 10$.

The results obtained for the stationary and nonstationary case are promising. Several extensions of this basic ideas are also possible. For example, this scheme could be used to adaptively select the best available channel for a single communication or to improve more complex constellation designs. Also, other algorithms for MCA could be used in order to gain robustness, convergence speed or to avoid the problem of the learning rate choice in practice [7].

References

1. Picinbono, B.: Random Signals and Systems. Prentice Hall (1993)
2. Oja, E.: Principal components, minor components and linear neural networks. Neural Networks **5** (1992) 927–935
3. Proakis, J., Salehi, M.: Digital Communications. 5th edn. McGraw-Hill (2007)
4. North, D.: An analysis of the factors which determine signal/noise discrimination in pulsed-carrier systems. Proceedings of the IEEE **51**(7) (1963) 1016–1027
5. Horn, R., Johnson, C.: Matrix Analysis. Cambridge University Press (1985)
6. Kay, S.: Fundamentals of Statistical Signal Processing, Volume II: Detection Theory. Prentice Hall (1998)
7. Cichocki, A., Amari, S.: Adaptive Blind Signal and Image Processing: Learning Algorithms and Applications. Communications technology. Wiley (2002)
8. Kushner, H.: Stochastic Approximation Methods for constrained and unconstrained systems. Springer (1978)