LIKELIHOOD-BASED SELECTION OF FILTERING PARAMETERS

Joaquín Míguez †, *Mónica F. Bugallo* ‡
† Departamento de Electrónica e Sistemas, Universidade da Coruña.
Facultade de Informática, Campus de Elviña s/n, 15071 A Coruña, Spain.
Fax: +34 981 167160, e-mail: jmiguez@udc.es
‡ Dept. of Electrical and Computer Engineering, SUNY at Stony Brook
Stony Brook, NY 11794, USA.
Fax: +1 631 6328494, e-mail: monica@ece.sunysb.edu

ABSTRACT

Many important problems in signal processing can be reduced to the selection of the parameters in a filtering structure. In this paper, we introduce a general selection criterion that relies on the ability to characterize the *desired* signal to be obtained at the filter output in terms of its probability density function (pdf). Using this statistical reference, the filter parameters are chosen in order to maximize the likelihood of the filtered signal under the desired probability distribution. We study the feasibility and asymptotic properties of this approach and present an illustrative simulation example, where the Space Alternating Generalized Expectation-maximization (SAGE) algorithm is used in the numerical implementation of the proposed method.

1 INTRODUCTION

Many important problems in signal processing can be reduced to the adequate selection of the parameters in a filtering structure so that it accomplishes some prescribed task, such as deconvolution, system identification or pattern recognition, to mention just a few examples. The criteria proposed in the literature to solve such problems are diverse. One possible classification, according to the type of *reference* employed to assess the quality of the filtered signal, yields three categories: temporal reference, structural reference and statistical reference criteria. The methods in the first group choose the filter coefficients to make the output signal equal to, or highly correlated with, an *a priori* known reference signal. A classical example is the Minimum Mean Square Error (MMSE) criterion and its adaptive implementations via the well-known Least Mean Squares (LMS) and Recursive Least Squares (RLS) algorithms [7]. In the second group, we find those criteria that exploit the structure of the input signal, as subspace methods [10] do. Finally, criteria that rely on the statistical properties of the signal alone fall within the third category. This includes Godard's, or constant modulus, criterion [6] and other techniques aimed at Independent Component Analysis (ICA) [3]. Though appealing due to their very mild requirements (only statistical independence of the signals to be recovered), the lack of a more informative reference limits the practical use of ICA algorithms, which usually require the availability of huge observation records and are subject to local convergence problems.

In this paper, we introduce a general selection criterion that relies on the ability to characterize the *desired* signal at the filter output in terms of a *target* probability density function (pdf) which, in the situations of interest, depends on the input signal and the filter parameters. Using this statistical reference, much more informative than statistical independence, the filter coefficients are chosen in order to maximize the likelihood of the output signal under the target probability distribution. Similar techniques have been proposed in order to solve specific problems in digital communications, namely channel equalization and beamforming [8, 9] and multiuser interference suppression [1], but a more general study of the criterion, including its use in higher-dimensional signal processing problems, has not been tackled yet, to the best of our knowledge.

In the next section, we present a mathematical formulation of the filtering problem that covers several applications of interest in signal processing. Next, in section 3, we study the feasibility, asymptotic properties and limitations of the proposed approach. An illustrative computer simulation example, involving a numerical implementation of the method via the Space-Alternating Generalized Expectation-maximization (SAGE) algorithm [5], is presented in section 4. Finally, brief concluding remarks are made in section 5.

2 PROBLEM STATEMENT

Let us consider a stochastic dynamical system of the form

$$\begin{aligned} \mathbf{s}_t &\sim p\left[\mathbf{s}_t|\mathbf{s}_{0:t-1}\right] \\ \mathbf{x}_t &\sim p\left[\mathbf{x}_t|\mathbf{s}_{0:t}\right] \end{aligned}$$

where $t \in \{0\}\cup\mathbb{N}$ denotes discrete time, $\mathbf{s}_t$ is the $N\times 1$ signal of interest at time t (usually called the system *state*), $\mathbf{s}_{i:j} := \{\mathbf{s}_l, i \leq l \leq j\}$ denotes a set of vectors and $\mathbf{x}_t$ is the $L \times 1$ vector of observations at time t. We use an argument-wise notation for probability functions where $p[\mathbf{z}]$ is the true pdf[1] of the random vector $\mathbf{z}$. Also, $p\left[\mathbf{z}|\tilde{\mathbf{z}}\right]$ denotes the *conditional* density of $\mathbf{z}$ given known realizations $\tilde{\mathbf{z}}$.

We are interested in estimating the unobserved sequence $\mathbf{s}_{0:T}$ from the observations $\mathbf{x}_{0:T}$ using a filtering structure that implements some desired function $\varphi(\cdot,\cdot)$. Hence, we define the estimate of $\mathbf{s}_t$ as the $N \times 1$ vector of the form

$$\mathbf{y}_t := \varphi\left(\mathbf{x}_{t-\tau:t+\tau}, \mathbf{W}\right),$$

where the choice of τ is usually made based on computational or time complexity motivations, and our aim is to select the parameter matrix $\mathbf{W}$ in order to make $\mathbf{y}_{0:T}$ a *good estimate* of the signal of interest, $\mathbf{s}_{0:T}$.

In the sequel, we will leave the choice of τ implicit, for a simpler notation, and let

$$\mathbf{y}_{0:T} := \varphi\left(\mathbf{x}_{0:T}, \mathbf{W}\right)$$

[1] Or, alternatively, the true probability mass function (pmf), if $\mathbf{z}$ takes values in a discrete set.

denote the whole filtered signal.

3 SELECTION CRITERION

Let us define

$$p_{\mathbf{W}}[\mathbf{y}_{0:T}] := p[\varphi(\mathbf{x}_{0:T}, \mathbf{W})] \quad (1)$$

as the joint pdf of the filtered signal for a particular value of the filter parameters[2]. We assume that the transformation $\varphi(\cdot, \mathbf{W})$ has the necessary regularity conditions to guarantee the existence of $p_{\mathbf{W}}[\cdot]$. Since a tractable and general analytical expression for the latter pdf is hard to find even for the simplest cases, we only assume the availability of a *statistical reference*, meaning that the probability distribution of the optimal (in some sense) estimates is known and denoted as $p_o[\mathbf{y}_{0:T}]$. We write $p_o[\cdot]$ without explicit reference to the parameters in order to remark that this desired pdf is known even if the parameter matrix that makes the estimates follow that distribution *is not known or does not exist.*

We propose to compute the filter parameters solving an *inverse* model selection problem. Given an observation record, $\mathbf{z}_{0:T}$, and a set of probability models, $\mathcal{M}$, the aim in model selection is to determine the element of $\mathcal{M}$ that yields the highest likelihood of the data [2], i.e., model $M_o \in \mathcal{M}$ is selected if, and only if,

$$\log\left(\frac{p[\mathbf{z}_{0:T}|M_o]}{p\left[\mathbf{z}_{0:T}|\tilde{M}\right]}\right) \geq 0, \quad \forall \tilde{M} \in \mathcal{M}, \quad \tilde{M} \neq M_o,$$

where $\log(\cdot)$ denotes the natural logarithm. In our problem, the data record is $\mathbf{y}_{0:T}$, the probability models are induced by different choices of the parameter matrix, $\mathbf{W}$, and there is one extra model given by $p_o[\cdot]$, which has been defined independently of $\mathbf{W}$. Hence, if $\mathbf{W}$ is a complex-valued matrix with dimensions $R \times C$, we define a (possibly infinite) model set as $\mathcal{M} := \mathcal{W} \cup M_o$, where $\mathcal{W} \subseteq \mathbb{C}^{R\times C}$ and M_o denotes the target probability distribution. The objective is to find the parameter matrix $\mathbf{W}_o \in \mathcal{W}$ (if it exists) such that:

$$\begin{array}{ll} \text{if} & \mathbf{y}_{0:T} = \varphi(\mathbf{x}_{0:T}, \mathbf{W}_o) \\ \text{then} & \log\left(\dfrac{p_o[\mathbf{y}_{0:T}]}{p_{\mathbf{W}}[\mathbf{y}_{0:T}]}\right) \geq 0, \quad \forall \mathbf{W} \in \mathcal{W} \end{array}$$

Given the available knowledge, the best choice is the parameter matrix that maximizes the likelihood of the data for the target model, i.e., we select the parameter matrix as

$$\begin{aligned} \psi(\mathbf{W}) &\propto \log(p_o[\varphi(\mathbf{x}_{0:T}, \mathbf{W})]) \\ \hat{\mathbf{W}}_T &= \arg \max_{\mathbf{W}} \{\psi(\mathbf{W})\}. \end{aligned} \quad (2)$$

3.1 Properties

When (2) is considered as an estimation method, the resulting estimate, $\hat{\mathbf{W}}_T$, complies with the results proved and dicussed in this section.

Property 1 $\hat{\mathbf{W}}_T$*, as computed in (2), is a Maximum Likelihood (ML) estimate of* $\mathbf{W}_o$ *if there exists a function*

$$p_{x;\mathbf{W}}[\mathbf{x}_{0:T}] \propto p_o[\varphi(\mathbf{x}_{0:T}, \mathbf{W})] \quad (3)$$

that is a pdf for the process $\mathbf{x}_{0:T}$*, with parameters* $\mathbf{W}$*.*

[2]Function $p_{\mathbf{W}}[\cdot]$ *is not* argument-wise, but it is completely defined by the parameter matrix $\mathbf{W}$ instead. Hence, on writing $p_{\mathbf{W}}[\mathbf{z}]$ we refer to the same function even if $\mathbf{z}$ is not distributed according to the pdf $p[\varphi(\mathbf{x}_{0:T}, \mathbf{W})]$. We assume this notational convention whenever a subindex is used in a pdf.

Proof 1 *Function* $p_{x,\mathbf{W}}[\cdot]$*, depends on the matrix of parameters* $\mathbf{W}$*. Hence, if it is a pdf,*

$$\psi_x(\mathbf{W}) \propto \log(p_{x;\mathbf{W}}[\mathbf{x}_{0:T}])$$

is the likelihood of $\mathbf{W}$ *for the available observations,* $\mathbf{x}_{0:T}$*, and it is apparent, according to definition (3), that* $\hat{\mathbf{W}}_T$ *is the argument that maximizes* $\psi_x(\mathbf{W})$*.* □

Remark 1 *Analogously,* $\hat{\mathbf{W}}_T$ *is a Maximum A Posteriori (MAP) estimator of* $\mathbf{W}_o$ *if a conditional pdf of the form* $p_w[\mathbf{W}|\mathbf{x}_{0:T}] \propto p_o\left[\varphi\left(\mathbf{x}_{0:T}, \hat{\mathbf{W}}\right)\right]$ *can be defined for* $\mathbf{W}$*.*

As a result of Property 1, if the pdf $p_{x;\mathbf{W}}[\cdot]$ exists, $\hat{\mathbf{W}}_T$ verifies all properties of ML estimators. In particular, it is asymptotically unbiased and efficient (if efficient estimation is possible).

The next two results yield asymptotic properties of $\hat{\mathbf{W}}_T$, and their proofs require the following conditions to hold:

C1. The target pdf is feasible using the filter structure, i.e., there exists $\mathbf{W}_o \in \mathcal{W}$ such that $p_o[\cdot] = p_{\mathbf{W}_o}[\cdot]$ according to (1).

C2. The output random process $\mathbf{y}_t = \varphi\left(\mathbf{x}_{t-\tau:t+\tau}, \hat{\mathbf{W}}_T\right)$, $t = 0, 1, 2, \ldots$, is ergodic in the mean. This requires process $\mathbf{x}_t$ to be ergodic itself and mild regularity conditions for $\varphi(\cdot,\cdot)$.

C3. The output random variables (rv's), $\mathbf{y}_{0:T}$, are statistically independent.

As a consequence, we obtain:

Property 2 *If no further condition is imposed on the target pdf,* $\hat{\mathbf{W}}_T$ *is a biased estimator of* $\mathbf{W}_o$*.*

Proof 2 *Using condition C3 (independence), the log-likelihood of the filtered signal,* $\mathbf{y}_{0:T} = \varphi(\mathbf{x}_{0:T}, \mathbf{W})$*, can be written as*

$$\log(p_o[\mathbf{y}_{0:T}]) = \sum_{t=0}^{T} \log(p_o[\mathbf{y}_t]). \quad (4)$$

Equality (4), together with condition C2 (ergodicity), yields

$$\begin{aligned} \lim_{T\to\infty} \frac{1}{T} \log(p_o[\mathbf{y}_{0:T}]) &= \lim_{T\to\infty} \frac{1}{T} \sum_{t=0}^{T} \log(p_o[\mathbf{y}_t]) \\ &= E_{p_{\mathbf{W}}}[\log(p_o)] \end{aligned} \quad (5)$$

where $E_{p_{\mathbf{W}}}$ *denotes statistical expectation with respect to the random process with pdf* $p_{\hat{\mathbf{W}}_T}$ *and we omit the argument of the pdf in* $\log(p_o)$ *for notational simplicity. Let us define the asymptotic estimator* $\hat{\mathbf{W}}_\infty := \lim_{T\to\infty} \hat{\mathbf{W}}_T$*. Assuming the this limit exists, we can combine equations (5) and (2) to obtain*

$$\hat{\mathbf{W}}_\infty = \arg \max_{\mathbf{W}} \{E_{p_{\mathbf{W}}}[\log(p_o)]\}. \quad (6)$$

In order to incorporate the latter result, we consider the Kullback-Leibler Distance (KLD) [4] between the densities $p_{\hat{\mathbf{W}}_\infty}[\cdot]$ *and* $p_o[\cdot]$*, which yields*

$$KLD(p_{\hat{\mathbf{W}}_\infty}||p_o) = E_{p_{\hat{\mathbf{W}}_\infty}}\left[\log\left(\frac{p_{\hat{\mathbf{W}}_\infty}}{p_o}\right)\right] \geq 0 \quad (7)$$

with equality if, and only if, $p_{\hat{\mathbf{W}}_\infty} = p_o$ which, due to condition C1 (feasibility), is equivalent to $\hat{\mathbf{W}}_\infty = \mathbf{W}_o$. It is useful to rewrite (7) as

$$-E_{p_{\hat{\mathbf{W}}_\infty}}\left[\log(p_o)\right] \geq \mathcal{H}_{\hat{\mathbf{W}}_\infty} \tag{8}$$

where $\mathcal{H}_{\hat{\mathbf{W}}_\infty} = -E_{p_{\hat{\mathbf{W}}_\infty}} \log\left(p_{\hat{\mathbf{W}}_\infty}\right)$ is the differential entropy associated to $p_{\hat{\mathbf{W}}_\infty}[\cdot]$ [4]. It is apparent that equality cannot be guaranteed in (8). Since $\hat{\mathbf{W}}_\infty$ minimizes the left-hand side in (8), it is simple to devise an example where $p_{\hat{\mathbf{W}}_\infty}[\cdot]$ presents the same modes as $p_o[\cdot]$ but a lesser differential entropy, hence

$$\mathcal{H}_{\mathbf{W}_o} > -E_{p_{\hat{\mathbf{W}}_\infty}}\left[\log(p_o)\right] > \mathcal{H}_{\hat{\mathbf{W}}_\infty},$$

and, as a consequence, KLD $\left(p_{\hat{\mathbf{W}}_\infty} \| p_o\right) > 0$ and $\hat{\mathbf{W}}_\infty \neq \mathbf{W}_o$. □

Property 3 *If the target probability model, M_o, is the one with the minimum differential entropy over the set of probability models $\mathcal{M}$, then $\hat{\mathbf{W}}_T \to \mathbf{W}_o$ as $T \to \infty$.*

Proof 3 *When the desired pdf yields the minimum feasible differential entropy, equation (8) can be augmented as*

$$-E_{p_{\hat{\mathbf{W}}_\infty}}\left[\log(p_o)\right] \geq \mathcal{H}_{\hat{\mathbf{W}}_\infty} \geq \mathcal{H}_{\mathbf{W}_o}$$

and, by virtue of property C1, the solution to (6) is $\hat{\mathbf{W}}_\infty = \mathbf{W}_o$. □

Remark 2 *Properties 2 and 3 also hold, without need to resort to condition C3, when the estimator is originally defined as*

$$\hat{\mathbf{W}}_T := \arg \max_{\mathbf{W}} \left\{ \tilde{\psi}(\mathbf{W}) := \sum_{t=0}^{T} \log\left(p_o\left[\mathbf{y}_t\right]\right) \right\}, \tag{9}$$

since $\lim_{T\to\infty} \tilde{\psi}(\mathbf{W}) = E_{p_{\mathbf{W}}}[p_o]$ due to condition C2 alone. If we abide by definition (9), however, condition C3 must be claimed to prove Property 1.

The most important consequence of properties 2 and 3 is not quantitative but qualitative. They indicate that the proposed method searches the filter parameters, $\hat{\mathbf{W}}_T$, such that $p_{\hat{\mathbf{W}}_T}[\cdot]$ presents (a) the same modes as $p_o[\cdot]$ and (b) the minimum entropy, i.e., the minimum variance around these modes. Notice that many signal processing problems for which this method is of interest are, ultimately, classification problems[3] where we aim at ascribing $\mathbf{y}_t$ to a particular mode of $p_o[\cdot]$ and, therefore, the entropy minimization feature of the method can be, indeed, an advantage.

3.2 Global vs. Local Solutions

Properties P1 and P2 have been shown to hold for the global solution in problem (2). However, the cost function $\psi(\cdot)$ is multimodal for many systems and computing its global maximum can be difficult in practice. Although we will not address this issue here due to lack of space, it is possible to study the local maxima in $\psi(\cdot)$ both analytically and graphically in some simple cases [9].

[3]Thus the case is in digital communication applications, where symbols must be detected, or in pattern recognition tasks, where the reconstructed patterns must be subsequently classified.

4 SIMULATION EXAMPLE

Let us consider the simple dynamical system

$$\begin{aligned} \mathbf{s}_t &\sim \mathcal{U}(\mathcal{S}) \\ \mathbf{x}_t &= \mathbf{A}\mathbf{s}_t + \mathbf{g}_t \end{aligned} \tag{10}$$

where both $\mathbf{s}_t$ and $\mathbf{x}_t$ are 2×1 random vectors (i.e., $N = L = 2$); $\mathbf{A}$ is an unknown 2×2 distortion matrix; $\mathbf{g}_t \sim \mathcal{N}\left(\mathbf{0}, \sigma^2\mathbf{Q}\right)$ with unknown σ^2 and known $\mathbf{Q}$, and $\mathcal{U}(\mathcal{S})$ denotes the uniform probability distribution over the known set of three patterns $\mathcal{S} = \left\{\mathbf{s}^{(1)}, \mathbf{s}^{(2)}, \mathbf{s}^{(3)}\right\}$.

The goal is to estimate the unobserved state signal $\mathbf{s}_{0:T}$ from the observations $\mathbf{x}_{0:T}$. Since (10) is linear and the noise and state processes are white, we simply set $\tau = 0$ and employ a 2×2 matrix filter,

$$\mathbf{y}_t = \varphi\left(\mathbf{x}_{t-\tau:t+\tau}, \mathbf{W}\right) = \mathbf{W}^\dagger \mathbf{x}_t, \quad t = 0, 1, .., T. \tag{11}$$

in order to obtain statistically independent estimates, $\mathbf{y}_{0:T}$ († denotes matrix transposition). The target pdf is mixture Gaussian with modes at the patterns in $\mathcal{S}$, i.e.,

$$p_o\left[\mathbf{y}_t\right] \propto \left|\mathbf{\Sigma}_f\right|^{-\frac{1}{2}} \sum_{k=1}^{3} \mathrm{e}^{-\frac{1}{2}\left(\mathbf{y}_t - \mathbf{s}^{(k)}\right)^\dagger \mathbf{\Sigma}_f^{-1}\left(\mathbf{y}_t - \mathbf{s}^{(k)}\right)},$$

where $\mathbf{\Sigma}_f$ is the filtered noise covariance matrix and $|\cdot|$ denotes the determinant of a square matrix. Since it is difficult to choose an adequate value of $\mathbf{\Sigma}_f$ without knowledge of $\mathbf{A}$, we assume $\mathbf{\Sigma}_f = \sigma_f^2\mathbf{I}$ and estimate $\left(\mathbf{W}_o, \sigma_f^2\right)$ jointly. The optimization problem to be solved is

$$\begin{aligned} \psi\left(\mathbf{W}, \sigma_f^2\right) &= -2(T+1)\log\left(\sigma_f^2\right) + \\ &\quad + \sum_{t=0}^{T} \log\left(\sum_{k=1}^{3} \mathrm{e}^{-\frac{\left(\mathbf{y}_t - \mathbf{s}^{(k)}\right)^\dagger\left(\mathbf{y}_t - \mathbf{s}^{(k)}\right)}{2\sigma_f^2}}\right) \\ \left(\hat{\mathbf{W}}_T, \hat{\sigma}_{f,T}^2\right) &= \arg\max_{\mathbf{W}} \left\{\psi\left(\mathbf{W}, \sigma_f^2\right)\right\}. \end{aligned} \tag{12}$$

Being (12) a likelihood maximization problem, it is possible to solve it by applying the SAGE algorithm [5], that yields the iterative updating rules

$$\hat{\mathbf{W}}_T^{(i+1)} = \left(\sum_{j=0}^{T} \mathbf{x}_t \mathbf{x}_t^\dagger\right)^{-1} \left(\sum_{t=0}^{T} \mathbf{x}_t \mathrm{E}_{i,i}\left[\mathbf{s}_t^\dagger\right]\right) \tag{13}$$

$$\hat{\sigma}_{f,T}^{2(i+1)} = \frac{\sum_{t=0}^{T} \mathrm{E}_{i+1,i}\left[\left(\mathbf{y}_t^{(i+1)} - \mathbf{s}_t\right)^\dagger \left(\mathbf{y}_t^{(i+1)} - \mathbf{s}_t\right)\right]}{N(T+1)} \tag{14}$$

where, for some function $\phi(\cdot)$ of the state vector $\mathbf{s}_t$, $\mathrm{E}_{i,j}\left[\phi(\mathbf{s}_t)\right]$ denotes the *a posteriori* expected value of $\phi(\mathbf{s}_t)$ under the desired probability model,

$$\mathrm{E}_{i,j}\left[\phi(\mathbf{s}_t)\right] = \sum_{k=1}^{3} \phi\left(\mathbf{s}^{(k)}\right) p_o\left[\mathbf{s}^{(k)} | \mathbf{y}_t^{(i)}, \hat{\sigma}_{f,T}^{2(j)}\right] \tag{15}$$

and $\mathbf{y}^{(i)} = \left(\hat{\mathbf{W}}_T^{(i)}\right)^\dagger \mathbf{x}_t$. The posterior pmf in (15) can be analytically obtained using the Bayes theorem.

Figure 1 plots the normalized histogram of $\mathbf{y}_{0:T} = \varphi\left(\mathbf{x}_{0:T}, \hat{\mathbf{W}}_T\right)$ with system parameters

$$\mathcal{S} = \left\{ \begin{bmatrix} 5.84 \\ 7.18 \end{bmatrix}, \begin{bmatrix} -7.28 \\ -8.04 \end{bmatrix}, \begin{bmatrix} 8.46 \\ -7.48 \end{bmatrix} \right\},$$

$$\mathbf{Q} = \begin{bmatrix} 0.393 & 0.008 \\ 0.008 & 0.566 \end{bmatrix} \text{ and } \mathbf{A} = \begin{bmatrix} 1 & -0.8 \\ 0.3 & -1 \end{bmatrix}.$$

The Signal-to-Noise Ratio (SNR), defined as

$$\mathrm{SNR} = 10 \log_{10} \left(\frac{\mathrm{trace}\left(\mathbf{A}\mathrm{E}_{p_{\mathbf{s}_t}}\left[\mathbf{s}_t\mathbf{s}_t^\dagger\right]\mathbf{A}^\dagger\right)}{\sigma^2 \mathrm{trace}\left(\mathbf{Q}\right)} \right),$$

is set to 15 dB. The modes corresponding to the patterns in $\mathcal{S}$ are clearly observed.

Finally, figure 2 plots the normalized Mean Square Error (MSE), defined as

$$\mathrm{MSE} = \frac{\sum_{t=0}^{T} \mathrm{trace}\left((\mathbf{s}_t - \mathbf{y}_t)(\mathbf{s}_t - \mathbf{y}_t)^\dagger \right)}{\sum_{t=0}^{T} \mathrm{trace}\left(\mathbf{s}_t\mathbf{s}_t^\dagger\right)},$$

for several values of the SNR. The system parameters are:

$$\mathcal{S} = \left\{ \begin{bmatrix} 7.06 \\ -7.88 \end{bmatrix}, \begin{bmatrix} -9.65 \\ 4.72 \end{bmatrix}, \begin{bmatrix} 9.23 \\ -6.18 \end{bmatrix} \right\},$$

and $\mathbf{A}$ and $\mathbf{Q}$ like in the previous figure. We observe that the proposed algorithm attains the same performance as the Wiener filter, that requires knowledge of $\mathbf{A}$ and σ^2 and is optimal in terms of MSE for the class of linear matrix filters.

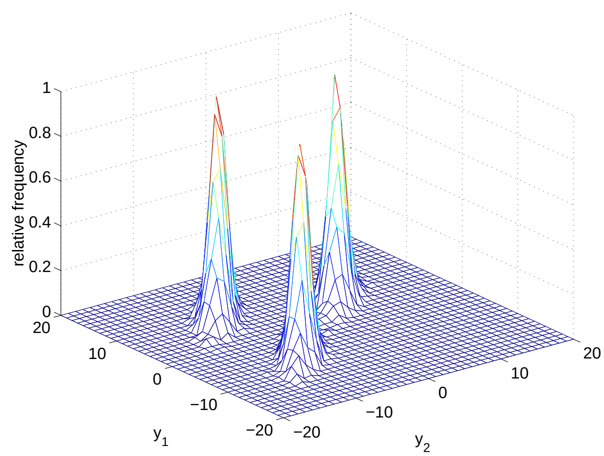

Figure 1: Normalized histogram of the filtered signal.

5 CONCLUSIONS

We have introduced a general criterion for selecting the parameters of arbitrary filtering structures. The method requires a characterization of the desired signal in terms of its pdf and consists of fitting the filter parameters in order to maximize the likelihood of the output signal under this target pdf. The statistical properties of the fitted parameters have been analyzed and the performance of the method has been illustrated via computer simulations.

6 ACKNOWLEDGEMENTS

This work has been supported by Xunta de Galicia, grant number PGIDT00PXI10504PR (†), and the National Science Foundation, award no. CCR-0082607 (‡).

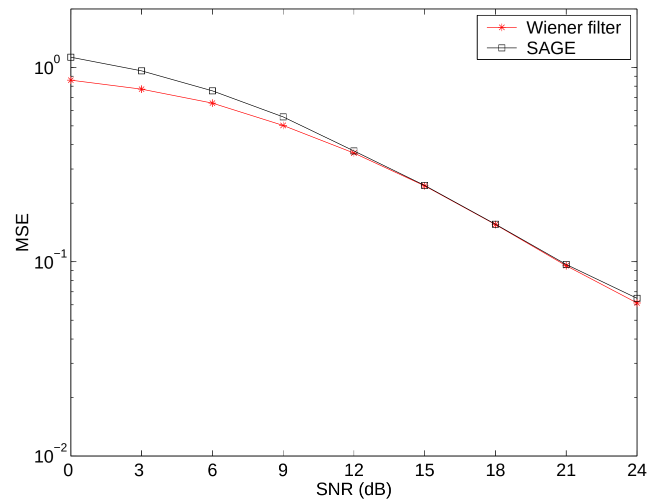

Figure 2: MSE for several values of the SNR.

References

[1] M. F. Bugallo, J. Míguez, and L. Castedo. A maximum likelihood approach to blind multiuser interference cancellation. *IEEE Trans. Signal Processing*, 49(6):1228–1239, June 2001.

[2] B. P. Carlin and S. Chib. Bayesian model choice via markov chain monte carlo methods. *J. of the Royal Stat. Society. Ser. B*, 57(3):473–484, 1995.

[3] P. Comon. Independent component analysis: A new concept. *IEEE Trans. Signal Processing*, 36:287–314, Abril 1994.

[4] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. Wiley-Interscience, New York (USA), 1991.

[5] J. A. Fessler and A. O. Hero. Space-alternating generalized expectation-maximization algorithm. *IEEE Trans. Signal Processing*, 42(10):2664–2677, October 1994.

[6] D. N. Godard. Self recovering equalization and carrier tracking in two-dimensional data communication systems. *IEEE Trans. Communications*, COMM-28:1867–1875, November 1980.

[7] S. Haykin. *Adaptive Filter Theory, 3rd Edition*. Prentice Hall, Information and System Sciences Series, 1996.

[8] S. J. Nowlan and G. E. Hinton. A soft decision-directed LMS algorithm for blind equalization. *IEEE Trans. Communications*, 41(2):275–279, February 1993.

[9] J. Sala-Álvarez and G. Vázquez-Grau. Statistical reference criteria for adaptive signal processing in digital communications. *IEEE Trans. Signal Processing*, 45(1):14–31, January 1997.

[10] P. Stoica and A. Nehorai. Performance comparison of subspace rotation and music methods for direction estimation. *IEEE Trans. on Signal Processing*, 39(2):357–361, February 1991.