Available online at www.sciencedirect.com
SCIENCE DIRECT®

Digital Signal Processing 15 (2005) 171–190

Digital Signal Processing

www.elsevier.com/locate/dsp

On the estimation of random unobserved signals by maximization of target likelihoods and its application to blind timing and phase recovery

Joaquín Míguez [a,*], Mónica F. Bugallo [b]

[a] *Departamento de Electrónica e Sistemas, Universidade da Coruña, Facultade de Informática, Campus de Elviña s/n, 15071 A Coruña, Spain*

[b] *Department of Electrical and Computer Engineering, Stony Brook University, Stony Brook, NY 11794, USA*

Available online 25 September 2004

Abstract

Many important problems in signal processing can be reduced to the adequate selection of the parameters of a (possibly nonlinear) filter in order to obtain an output signal that complies with some desired properties. In this work, we analyze a novel criterion for selecting filter parameters that relies on the ability to characterize the desired filter output in terms of a *target* probability density function (pdf). This target pdf can be handled as a likelihood function to be maximized, thus we refer to the new criterion as maximum target-likelihood (MTL). We present a very general signal model where the MTL criterion can be applied and derive necessary and sufficient conditions for asymptotic convergence of the method. The relationship and differences between MTL and standard maximum likelihood (ML), minimum Kullback–Leibler divergence (MKLD), and minimum entropy (ME) methods are explored. Finally, as an example, we apply the novel criterion to the problem of blind timing and phase recovery in a digital transmission system and show that the resulting algorithm is competitive with existing non-data-aided ML-based algorithms.

Keywords: Maximum likelihood estimation; Differential entropy; Kullback–Leibler divergence; Stochastic processes; Adaptive filtering; Synchronization; Blind timing recovery; Blind phase estimation; Blind detection

* Corresponding author. Present address: Departamento de Teoría de la Señal y Comunicaciones, Universidad Carlos III de Madrid, Spain.

E-mail addresses: jmiguez@udc.es (J. Míguez), monica@ece.sunysb.edu (M.F. Bugallo).

1051-2004/$ – see front matter
doi:10.1016/j.dsp.2004.08.005

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, e.g., deconvolution [1], system identification [2], or pattern recognition [3]. 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, spectral, and statistical reference criteria. The methods in the first group select the filter coefficients to make the output signal equal to, or highly correlated with, an a priori known signal. A typical example is the minimum mean square error (MMSE) criterion and its adaptive implementation via the well-known least mean squares (LMS) algorithm [4]. Filter design and optimization to match some prescribed spectral response is a thoroughly researched topic which includes classical techniques such as the minimax method (see, e.g., [5]). Finally, criteria that rely on the statistical properties of the signals fall within the third category, which includes some techniques aimed at independent component analysis (ICA) [6], e.g., the constant modulus criterion [7] or Cardoso and Laheld's equivariant algorithm [8]. 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 these purely-statistical ICA algorithms, which usually require the availability of huge observation records and are subject to local convergence problems. For some problems, better convergence properties can be provided by methods that also exploit the signal structure, as subspace methods do [9].

In this work, we analyze a novel criterion for parameter selection 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 practical situations, depends on the filter parameters and the distribution of the signal of interest. The target pdf serves as a statistical reference, much more informative than statistical independence alone, and the filter parameters can be chosen to maximize the likelihood of the output signal under the target probability model. For this reason, we refer to the proposed method as maximum target likelihood (MTL).

Similar techniques have been suggested in the past to deal with specific problems in digital communications, namely beamforming and channel equalization [10,11], as well as multiuser interference suppression [12,13]. These approaches, however, are concerned with simple schemes consisting of Gaussian-corrupted observations which are processed with a linear finite impulse response (FIR) filter. It is also worth mentioning a formally similar technique proposed in [14]. The latter method, however, is derived within the context of *conventional* maximum likelihood (ML) estimation by using adequate priors on nuisance variables which are integrated out.

Except for a first attempt by these authors in [15], a general study of the MTL criterion has not been tackled yet, to the best of our knowledge. In this work, we address a very general framework that comprehends dynamic systems in state-space form and arbitrary filtering functions, both linear and nonlinear. This broad setup allows to obtain both sufficient and necessary conditions for asymptotic convergence of MTL solutions. Compared to [15], we introduce stronger, and more detailed, theoretical results regarding convergence, together with a new application of the MTL methodology to the problem of blind timing and phase recovery in a digital receiver.

The remaining of the paper is organized as follows. Specific notation used through the article is introduced in Section 2. In Section 3, we present a mathematical formulation of the filtering problem that covers several applications of interest in signal processing. The proposed MTL criterion is formally stated as a method for *model fitting* in Section 4. In Section 5, an asymptotic convergence result that relates the MTL solution to the differential entropy (DE) of the filtered signal is proved. The connection of the MTL method with standard ML, minimum entropy (ME) and minimum Kullback–Leibler divergence (MKLD) techniques is investigated, and key differences are identified, in Section 6. Convergence in the presence of multiple target likelihood maxima is briefly commented upon in Section 7. The validity of the method is illustrated, in Section 8, by applying it to the problem of blind timing and phase recovery in a digital receiver. Finally, Section 9 is devoted to the conclusions.

2. Notation and definitions

We will abide by the following notation:

(1) *Vectors and matrices.* Vectors and matrices are denoted by lower-case bold-face and upper-case bold-face characters, respectively. The set of complex numbers is denoted as $\mathbb{C}$. Hence, $\mathbf{v} \in \mathbb{C}^N$ denotes an $N \times 1$ vector of complex elements.

(2) *Probability density functions.* Given random vectors $\mathbf{x}, \mathbf{y} \in \mathbb{C}^N$, $p(\mathbf{x}, \mathbf{y})$ is the true joint pdf of $\mathbf{x}$ and $\mathbf{y}$, $p(\mathbf{x}|\mathbf{y})$ is the true conditional pdf of $\mathbf{x}$ given $\mathbf{y}$ and $p(\mathbf{x})$ is the true marginal pdf of $\mathbf{x}$.

(3) *Probability models.* Probabilistic models are indices that completely identify a pdf. Let $\mathcal{M}$ be a set of models and let $M_1, M_2 \in \mathcal{M}$, $M_1 \neq M_2$. If a random vector $\mathbf{x}$ is distributed according to model M_1, written $\mathbf{x} \sim p(\mathbf{x}|M_1)$, then $p(\mathbf{x}|M_1) = p(\mathbf{x})$. If $\mathbf{x} \sim p(\mathbf{x}|M_2)$, then $p(\mathbf{x}|M_1) \neq p(\mathbf{x}) = p(\mathbf{x}|M_2)$. We use the model index to identify $p(\cdot|M)$, $M \in \mathcal{M}$, as a function that can be applied to any complex vector with the same dimension as $\mathbf{x}$.

(4) *Differential entropy.* The differential entropy (DE) [16] of a random vector, $\mathbf{x} \in \mathbb{C}^{N_x}$, $\mathbf{x} \sim p(\mathbf{x})$ is

$$H(\mathbf{x}) \triangleq -\mathrm{E}_{p(\mathbf{x})} \log p(\mathbf{x}),$$

where $\mathrm{E}_{p(\mathbf{x})}$ denotes statistical expectation with respect to (w.r.t.) the pdf in the subscript. The DE is a measure of the average information amount conveyed by the random source $\mathbf{x} \sim p(\mathbf{x})$.

(5) *Kullback–Leibler divergence.* The Kullback–Leibler divergence (KLD) between two pdf's, say $p(\mathbf{x}|M_1)$ and $p(\mathbf{x}|M_2)$, is defined as [16]

$$\mathrm{KLD}(M_1 \parallel M_2) \triangleq \mathrm{E}_{p(\mathbf{x}|M_1)} \log \frac{p(\mathbf{x}|M_1)}{p(\mathbf{x}|M_2)}.$$

The KLD measures the similarity between the two densities. It is guaranteed that $\mathrm{KLD}(M_1 \parallel M_2) \geqslant 0$, with equality if, and only if, $p(\mathbf{x}|M_1) = p(\mathbf{x}|M_2)$.

3. Problem statement

Let us consider the Markovian discrete-time random process

$$\mathbf{x}_n \sim p(\mathbf{x}_n|\mathbf{x}_{n-1}), \tag{1}$$

where $\mathbf{x}_n \in \mathbb{C}^{N_x}$ is a random signal of interest, with conditional pdf $p(\mathbf{x}_n|\mathbf{x}_{n-1})$. Notice that this model includes discrete processes, since probability mass functions (pmf's) can be represented in a continuous space as sums of Dirac deltas and then handled as pdf's. We are interested in estimating the sequence $\mathbf{x}_{0:n} = \{\mathbf{x}_0, \ldots, \mathbf{x}_n\}$, which cannot be observed directly. The signals we are able to collect through some type of sensor (or sensor array) during the nth observation period are $N_z \times 1$ complex vectors of the form

$$\mathbf{z}(t) = \theta_c\big(\mathbf{x}_n, \mathbf{u}(t), t\big), \quad t_{n-1} < t \leqslant t_n,$$

if they are continuous-time signals, or

$$\mathbf{z}_n = \theta_d(\mathbf{x}_n, \mathbf{u}_n),$$

if they are discrete-time processes. Functions θ_d and θ_c represent some unknown distortion of $\mathbf{x}_n$, as well as the effect of output and/or input transducers. Random processes $\mathbf{u}_n, \mathbf{u}(t) \in \mathbb{C}^{N_u}$ are used to model noise and other unknown nuisance signals.

The goal is to process the raw observations, either $\mathbf{z}(t)$ or $\mathbf{z}_n$, using some convenient filter that we formally describe by means of function

$$\varphi : \mathbb{C}^{N_z} \times \mathbb{C}^{N_w} \to \mathbb{C}^{N_x},$$

which takes the observations and a vector of adjustable parameters, $\mathbf{w} \in \mathcal{W} \subseteq \mathbb{C}^{N_w}$, as arguments in order to yield an estimate of $\mathbf{x}_n$. For the sake of generality, we assume a continuous-time observation and let the sampling of the collected signal be part of the processing whenever necessary (i.e., the sampling rate can be considered as one of the parameters in $\mathbf{w}$). Thus, the filter output corresponding to the nth observation period ($t_{n-1} < t \leqslant t_n$) can be written as

$$\mathbf{y}_n = \varphi_n\big(\mathbf{z}(t), \mathbf{w}\big) = \varphi_n\big(\theta_c\big(\mathbf{x}_n, \mathbf{u}(t), t\big), \mathbf{w}\big). \tag{2}$$

Finally, we drop the explicit reference to $\mathbf{x}_n, \mathbf{u}(t)$ and simply write

$$\mathbf{y}_n = \varphi_n(\mathbf{w}).$$

When considered together, Eqs. (1) and (2) yield a dynamic system where the signal of interest, $\mathbf{x}_n$, is the system state and $\mathbf{y}_n$ is an associated measurement. The aim is to adjust the filter parameters to ensure that the sequence of measurements is close to the sequence of states in some sense, i.e., we look for a choice of $\mathbf{w}$ such that the measurements, $\mathbf{y}_{0:n} = \varphi_{0:n}(\mathbf{w}) = \{\varphi_0(\mathbf{w}), \ldots, \varphi_n(\mathbf{w})\}$, can be used as estimates of $\mathbf{x}_{0:n}$. Figure 1 illustrates the system model.

4. Selection of filter parameters by maximization of a target likelihood

We assume that for each choice of filter parameters, $\mathbf{w} \in \mathcal{W}$, it is possible to define a probabilistic model identified by a real index $I(\mathbf{w}) \in \tilde{\mathcal{M}}$, where $\tilde{\mathcal{M}} \subseteq \mathbb{R}$ is the set of model

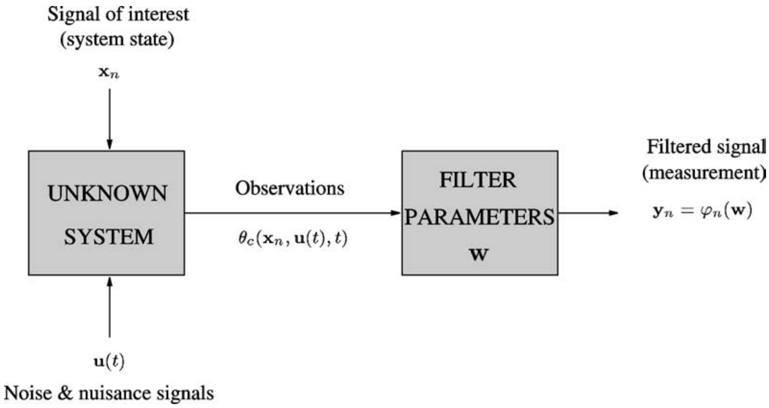

Fig. 1. Block diagram of the adopted system model.

indices. According to this notation, we can write

$$\mathbf{y}_n = \varphi_n(\mathbf{w}) \Rightarrow \mathbf{y}_n \sim p(\mathbf{y}_n) = p\big(\mathbf{y}_n | I(\mathbf{w})\big)$$

and

$$I(\mathbf{w}) \neq I(\mathbf{w}') \Leftrightarrow p\big(\mathbf{y}_n | I(\mathbf{w})\big) \neq p\big(\mathbf{y}_n | I(\mathbf{w}')\big).$$

In order to select the adequate value of the filter parameters, we assume there is an a priori known optimal probability distribution of the measurements. The optimal model is indexed as I_0 ($I_0 \notin \tilde{\mathcal{M}}$) and we hypothesize that if the filtered signal is distributed according to this model,

$$\mathbf{y}_n \sim p(\mathbf{y}_n | I_0),$$

then $\mathbf{y}_n$ is a good estimate of $\mathbf{x}_n$ in some statistical sense. Two important remarks are necessary at this point:

Remark 1. Model index I_0 is written without reference to a specific choice of filter parameters in order to emphasize that this optimal model is known even if the parameter vector that makes the measurements follow that distribution is not known, does not belong to $\mathcal{W}$ or simply does not exist.

Remark 2. It is important to realize that, for most problems of interest, even some of the simplest, it is impossible to find a tractable analytical expression for $p(\mathbf{y}_n | I(\mathbf{w}))$. Therefore, the straightforward strategy of solving equation $p(\mathbf{y}_n | I(\mathbf{w})) = p(\mathbf{y}_n | I_0)$ for $\mathbf{w}$ is not feasible in general.

In the sequel, the optimal model, I_0, is termed the *target model* and its associated pdf, $p(\mathbf{y}_n | I_0)$, the *target pdf*. Our aim is to compute a parameter vector $\hat{\mathbf{w}}$ such that $p(\mathbf{y}_n | I(\hat{\mathbf{w}}))$ is close to $p(\mathbf{y}_n | I_0)$ in some sense.

One appealing way to tackle this optimization problem is to pose it as one of model fitting. Let us define the complete model set as $\mathcal{M} = \tilde{\mathcal{M}} \bigcup \{I_0\}$. We propose to select the filter parameters so as to *fit* the target model, i.e., so as to make I_0 the most likely model in $\mathcal{M}$ according to the measurements. Mathematically, this means that, given a record of $n+1$ observations, $\hat{\mathbf{w}}$ can be chosen as

$$\hat{\mathbf{w}} = \arg\max_{\tilde{\mathbf{w}}} \left\{ \int_{\mathcal{W}} \sum_{k=0}^{n} \log \frac{p(\mathbf{y}_k|I_0)}{p(\mathbf{y}_k|I(\mathbf{w}))}\, d\mathbf{w} \right\} \quad \text{s.t. } \mathbf{y}_{0:n} = \varphi_{0:n}(\tilde{\mathbf{w}})$$
$$= \arg\max_{\tilde{\mathbf{w}}} \left\{ \int_{\mathcal{W}} \sum_{k=0}^{n} \log \frac{p(\varphi_k(\tilde{\mathbf{w}})|I_0)}{p(\mathbf{y}_k|I(\mathbf{w}))}\, d\mathbf{w} \right\}, \tag{3}$$

where s.t. stands for "subject to" and log denotes the natural logarithm. We have assumed that $p(\mathbf{y}_n|I(\mathbf{w}))$ does not have a tractable closed form for an arbitrary parameter vector $\mathbf{w} \in \mathcal{W}$ (see Remark 2 above). Thus, our knowledge is limited to the target pdf and the record of measurements and, as a consequence, we can do no better than increasing the value of the integral in (3) up to the maximization of the target density (which does not depend on the integration variable $\mathbf{w}$). Therefore, we substitute (3) by the simpler problem

$$\hat{\mathbf{w}} = \arg\max_{\tilde{\mathbf{w}}} \left\{ \sum_{k=0}^{n} \log p(\mathbf{y}_k|I_0) \right\} \quad \text{s.t. } \mathbf{y}_{0:n} = \varphi_{0:n}(\tilde{\mathbf{w}})$$
$$= \arg\max_{\mathbf{w}} l_n^0(\mathbf{w}), \tag{4}$$

where

$$l_n^0(\mathbf{w}) = \sum_{k=0}^{n} \log p\big(\varphi_k(\mathbf{w})|I_0\big) \tag{5}$$

is the target likelihood of $\mathbf{w}$ up to time n. Hence, the parameter vector obtained as the solution of problem (4) is termed the maximum target likelihood (MTL) solution of the filter parameters, and the resulting measurements, $\mathbf{y}_{0:n} = \varphi_{0:n}(\hat{\mathbf{w}})$, are the MTL estimates of the state sequence $\mathbf{x}_{0:n}$.

Clearly, when applying the MTL criterion in (4), we are disregarding the integral term $-\int_{\mathcal{W}} \sum_{k=0}^{n} \log p(\mathbf{y}_k|I(\mathbf{w}))\, d\mathbf{w}$, which appears in (3) and depends on $\tilde{\mathbf{w}}$ through $\mathbf{y}_k = \varphi_k(\tilde{\mathbf{w}})$. Therefore, we do not claim that the proposed criterion be optimal in any particular sense, or constitute a universal tool which can be successfully applied to every filtering problem. The relative merit of the MTL method should be found in making a sensible (yet possibly suboptimal) use of the available knowledge in order to achieve a good practical solution in those problems where a statistical reference in the form of a target pdf can be easily identified.

5. Asymptotic convergence

So far, we have developed the MTL criterion within a general framework and arrived at a relatively simple expression for the selection of the filter parameters, which reduces to the maximization of a known log-likelihood function. Although this approach has been proposed in the past to solve specific problems in communications [10–12], a general assessment of the convergence properties of the method has not been done yet. Indeed, the algorithms for channel equalization [11] and multiuser interference suppression [12] are supported by useful analytical results, but they are constrained to finite-alphabet uniform state signals, with linear observations and linear filtering. As a consequence of this lack of generality, the results reported in [11,12] cannot be directly applied to the MTL solution given by (4).

In this section, we state and prove a theorem that relates the MTL parameters, selected according to (4), to the DE of the measurements, $\mathbf{y}_n$. The theorem holds true asymptotically, as $n \to \infty$, and imposes few constraints on the observation and filtering functions

Section 5.1 is devoted to the theorem assumptions, while its statement and proof are developed in Section 5.2.

5.1. Assumptions and notation

The following assumptions are made in the remaining of this section:

(i) *Stationarity.* For a fixed $\mathbf{w}$, the measurement process is stationary up to the equality of the marginal densities, $p(\mathbf{y}_k|I(\mathbf{w})) = p(\mathbf{y}_l|I(\mathbf{w}))$, $\forall k, l$.

(ii) *Ergodicity.* For a fixed $\mathbf{w}$, the measurement process is ergodic in the mean, i.e.,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n} \nu(\mathbf{y}_k) = \mathrm{E}_{p(\mathbf{y}_n)} \nu(\mathbf{y}_n) < \infty,$$

with convergence in probability (i.p.), where ν is an arbitrary integrable function of $\mathbf{y}_n$ and $\mathrm{E}_{p(\mathbf{y}_n)}$ denotes mathematical expectation w.r.t. the pdf in the subscript.

(iii) *Existence.* The following summation converges i.p.

$$\lim_{n\to\infty} \frac{1}{n} l_n^0(\mathbf{w}) = l^0(\mathbf{w}).$$

(iv) *Gradient.* The gradient of the target likelihood, $\nabla_{\mathbf{w}} l_n^0(\mathbf{w})$, exists and it is an invertible function of $\mathbf{w}$ for
   (a) all parameter vectors belonging to the region of interest, i.e., $\forall \mathbf{w} \in \mathcal{W}$, and
   (b) all integers greater than some threshold, i.e., $\forall n > n_0$ ($n_0 < \infty$).

   Additionally, it is useful to introduce notation

$$\hat{\mathbf{w}}_n = \arg\max_{\mathbf{w}} l_n^0(\mathbf{w}) \tag{6}$$

   for the MTL parameter vector computed using the observations up to time n, and

$$\hat{\mathbf{w}}_\infty = \lim_{n\to\infty} \hat{\mathbf{w}}_n$$

   for the limit of (6) as the number of observations becomes arbitrarily large.

In general, we do not assume that an *exact* solution exists within $\mathcal{W}$ such that the target pdf $p(\mathbf{y}_n|I_0)$ can be attained. However, for notational convenience, we define the optimal parameter vector $\mathbf{w}_0$ that achieves $p(\mathbf{y}_n|I(\mathbf{w}_0)) = p(\mathbf{y}_n|I_0)$ although, possibly,

$\mathbf{w}_0 \in \mathbb{C}^{N_w} \setminus \mathcal{W}$. As a consequence, it is not at all guaranteed that $p(\mathbf{y}_n | I(\hat{\mathbf{w}}_\infty)) = p(\mathbf{y}_n | I_0)$. The properties of the asymptotic MTL solution are established in Section 5.2.

5.2. Asymptotic convergence and differential entropy

Let us write

$$H(\mathbf{w}_1) = -\mathrm{E}_{p(\mathbf{y}_n|I(\mathbf{w}_1))} \log p\big(\mathbf{y}_n | I(\mathbf{w}_1)\big),$$

for the DE of $\mathbf{y}_n \sim p(\mathbf{y}_n | I(\mathbf{w}_1))$, and

$$\mathrm{KLD}(\mathbf{w}_1 \parallel \mathbf{w}_2) = \mathrm{E}_{p(\mathbf{y}_n|I(\mathbf{w}_1))} \log \frac{p(\mathbf{y}_n|I(\mathbf{w}_1))}{p(\mathbf{y}_n|I(\mathbf{w}_2))},$$

for the Kullback–Leibler divergence (KLD) between $p(\mathbf{y}_n|I(\mathbf{w}_1))$ and $p(\mathbf{y}_n|I(\mathbf{w}_2))$. The feasible parameter vector that yields the closest-to-target pdf in the KLD sense plays a major role in the subsequent analysis and we denote it as

$$\mathbf{w}_* = \arg \min_{\mathbf{w} \in \mathcal{W}} \mathrm{KLD}(\mathbf{w} \parallel \mathbf{w}_0).$$

We have the following theorem.

Theorem 1. *Let $l_n^0(\mathbf{w})$ be a target likelihood in the region of interest, $\mathbf{w} \in \mathcal{W}$, which complies with assumptions* (i)–(iv)*. Necessary* (*a*) *and sufficient* (*b*) *conditions for asymptotic convergence can be established as shown below.*

(a) *If $\hat{\mathbf{w}}_\infty = \mathbf{w}_*$, then $H(\mathbf{w}_*) \leqslant H(\mathbf{w}) + \mathrm{KLD}(\mathbf{w} \parallel \mathbf{w}_0) - \mathrm{KLD}(\mathbf{w}_* \parallel \mathbf{w}_0),\ \forall \mathbf{w} \in \mathcal{W}$.*
(b) *If $H(\mathbf{w}_*) \leqslant H(\mathbf{w}) + \mathrm{KLD}(\mathbf{w} \parallel \mathbf{w}_*),\ \forall \mathbf{w} \in \mathcal{W}$, then $\hat{\mathbf{w}}_\infty = \mathbf{w}_*$.*

Proof (Outline). We proceed in three steps. First, we rely on assumptions (i)–(iv) to derive a convenient asymptotic version (as $n \to \infty$) of problem (6). Next, necessary condition (a) is easily derived from the definition of KLD. Finally, sufficient condition (b) is obtained by contradiction using the *Pythagorean theorem* of the KLD [16].

Step 1. *Asymptotic MTL solution.* We start with the definition of a modified target likelihood function of the form

$$\tilde{l}_n^0(\mathbf{w}) = \frac{1}{n} l_n^0(\mathbf{w}), \tag{7}$$

and the realization that

$$\hat{\mathbf{w}}_n = \arg \max_{\mathbf{w}} l_n^0(\mathbf{w}) = \arg \max_{\mathbf{w}} \tilde{l}_n^0(\mathbf{w}).$$

Substituting (5) into (7), we readily obtain an expression for the modified target likelihood,

$$\tilde{l}_n^0(\mathbf{w}) = \frac{1}{n} \sum_{k=0}^{n} \log p\big(\varphi_k(\mathbf{w}) | I_0\big) = \frac{1}{n} \sum_{k=0}^{n} \log p(\mathbf{y}_k | I_0), \tag{8}$$

subject to $\mathbf{y}_k = \varphi_k(\mathbf{w})$. According to the stationarity assumption (i), all densities in (8) are equal, hence we can apply the weak law of large numbers and the ergodicity assumption (ii) to arrive at

$$\lim_{n \to \infty} \tilde{l}_n^0(\mathbf{w}) = \mathrm{E}_{p(\mathbf{y}_k|I(\mathbf{w}))} \log p(\mathbf{y}_k | I_0) \quad \text{(i.p.)}. \tag{9}$$

To complete this first part of the proof, we recall the existence (iii) and gradient (iv) assumptions to obtain

$$\hat{\mathbf{w}}_\infty = \lim_{n\to\infty}\left\{\arg\max_{\mathbf{w}} \tilde{l}_n^0(\mathbf{w})\right\} = \arg\max_{\mathbf{w}}\left\{\lim_{n\to\infty}\tilde{l}_n^0(\mathbf{w})\right\}, \tag{10}$$

which, by substitution of (9) into (10), yields

$$\hat{\mathbf{w}}_\infty = \arg\max_{\mathbf{w}} \mathrm{E}_{p(\mathbf{y}_k|I(\mathbf{w}))} \log p(\mathbf{y}_k|I_0). \tag{11}$$

Step 2. *Necessary condition* (a). Given (11), $\hat{\mathbf{w}}_\infty = \mathbf{w}_*$ obviously yields

$$-\mathrm{E}_{p_*}\log p_0 \leqslant -\mathrm{E}_{p_\mathbf{w}}\log p_0, \quad \forall \mathbf{w} \in \mathcal{W},$$

where p_* and p_0 are shorthand for $p(\mathbf{y}_n|I(\mathbf{w}_*))$ and $p(\mathbf{y}_n|I(\mathbf{w}_0))$, respectively. Using the definitions of KLD and DE, the latter inequality can be successively transformed into

$$\mathrm{KLD}(\mathbf{w}_* \parallel \mathbf{w}_0) + H(\mathbf{w}_*) \leqslant \mathrm{KLD}(\mathbf{w} \parallel \mathbf{w}_0) + H(\mathbf{w}), \quad \forall \mathbf{w} \in \mathcal{W},$$

and

$$H(\mathbf{w}_*) \leqslant \mathrm{KLD}(\mathbf{w} \parallel \mathbf{w}_0) - \mathrm{KLD}(\mathbf{w}_* \parallel \mathbf{w}_0) + H(\mathbf{w}), \quad \forall \mathbf{w} \in \mathcal{W},$$

which is the necessary condition in part (a) of the theorem.

Step 3. *Sufficient condition* (b). We proceed by contradiction. Let us assume the following relationships hold jointly:

(i) $H(\mathbf{w}_*) \leqslant H(\mathbf{w}) + \mathrm{KLD}(\mathbf{w} \parallel \mathbf{w}_*)$, $\forall \mathbf{w} \in \mathcal{W}$,
(ii) $\hat{\mathbf{w}}_\infty \neq \mathbf{w}_*$.

From (ii) we obtain $\exists \mathbf{w}' \in \mathcal{W}$ such that

$$-\mathrm{E}_{p_{\mathbf{w}'}}\log p_0 < -\mathrm{E}_{p_*}\log p_0,$$

where $p_{\mathbf{w}'}$ is shorthand for $p(\mathbf{y}_n|I(\mathbf{w}'))$, which can be readily written in terms of KLD and DE as

$$\mathrm{KLD}(\mathbf{w}' \parallel \mathbf{w}_0) + H(\mathbf{w}') < \mathrm{KLD}(\mathbf{w}_* \parallel \mathbf{w}_0) + H(\mathbf{w}_*). \tag{12}$$

Using the *Pythagorean* inequality [16]

$$\mathrm{KLD}(\mathbf{w} \parallel \mathbf{w}_0) \geqslant \mathrm{KLD}(\mathbf{w} \parallel \mathbf{w}_*) + \mathrm{KLD}(\mathbf{w}_* \parallel \mathbf{w}_0), \quad \forall \mathbf{w} \in \mathcal{W},$$

in order to decompose $\mathrm{KLD}(\mathbf{w}' \parallel \mathbf{w}_0)$ in (12) yields

$$\mathrm{KLD}(\mathbf{w}' \parallel \mathbf{w}_*) + H(\mathbf{w}') < H(\mathbf{w}_*)$$

which is in contradiction with assumption (i) above. □

Corollary 1. *If* $\mathbf{w}_0 \in \mathcal{W}$ *then* $\hat{\mathbf{w}}_\infty = \mathbf{w}_0$ *if, and only if,*

$$H(\mathbf{w}_0) \leqslant H(\mathbf{w}) + \mathrm{KLD}(\mathbf{w} \parallel \mathbf{w}_0), \quad \forall \mathbf{w} \in \mathcal{W}.$$

Proof. Taking into account that $\mathbf{w}_0 \in \mathcal{W} \Rightarrow \mathbf{w}_0 = \mathbf{w}_* \Rightarrow \mathrm{KLD}(\mathbf{w}_* \parallel \mathbf{w}_0) = 0$, the proof follows trivially from Theorem 1. □

5.3. Discussion

The above theorem and corollary indicate that MTL solutions asymptotically converge to filter parameters which *may be not* the target parameters (or the closest feasible vector), but are close to them according to the KLD and yield a pdf with a smaller DE, if such a density exists. Note that this is not necessarily a drawback, since small entropy solutions are desirable in many applications, such as detection and classification.

One important case where convergence to a smaller-entropy pdf comes up is the problem of linear equalization of a dispersive channel in Gaussian noise. In [11], it is proved that the MTL linear equalizer is the filter that yields a measurement pdf with the same modes as the target one and minimum variance. For the symmetric mixture Gaussian target pdf's considered in [11], the minimization of the variance is equivalent to the minimization of the DE and, therefore, this result is consistent with the asymptotic convergence theorem in this paper.

The asymptotic convergence properties of the MTL method also provide an interesting ground for comparison with standard techniques that can be characterized in similar terms, namely ML, MKLD, and ME criteria. This is the aim of Section 6.

6. Connection with MKLD, ML, and ME techniques

6.1. MTL vs MKLD

There are two ways to build MKLD filter parameters,

$$\mathbf{w}_{\mathrm{MKLD}}^{(1)} = \arg\min_{\mathbf{w}\in\mathcal{W}} \mathrm{KLD}(\mathbf{w} \parallel \mathbf{w}_0) = \arg\min_{\mathbf{w}\in\mathcal{W}} \{\mathrm{E}_{p_{\mathbf{w}}} \log p_{\mathbf{w}} - \mathrm{E}_{p_{\mathbf{w}}} \log p_0\},$$

$$\mathbf{w}_{\mathrm{MKLD}}^{(2)} = \arg\min_{\mathbf{w}\in\mathcal{W}} \mathrm{KLD}(\mathbf{w}_0 \parallel \mathbf{w}) = \arg\min_{\mathbf{w}\in\mathcal{W}} \{-\mathrm{E}_{p_0} \log p_{\mathbf{w}}\}.$$

Obviously, if $\mathbf{w}_0 \in \mathcal{W}$ then $\mathbf{w}_{\mathrm{MKLD}}^{(1)} = \mathbf{w}_{\mathrm{KLD}}^{(2)} = \mathbf{w}_0$ but, otherwise, $\mathbf{w}_{\mathrm{MKLD}}^{(1)} \neq \mathbf{w}_{\mathrm{KLD}}^{(2)}$ in general because the KLD is not symmetric.

At any rate, Theorem 1 shows that the MTL solution may not necessarily converge to $\mathbf{w}_{\mathrm{MKLD}}^{(1)}$ (note that $\mathbf{w}_* = \mathbf{w}_{\mathrm{MKLD}}^{(1)}$), depending on the form of the set of feasible vectors, $\mathcal{W}$. Moreover, even when $\mathbf{w}_0 \in \mathcal{W}$, Corollary 1 states that if $\exists \mathbf{w}' \in \mathcal{W}$ such that $H(\mathbf{w}') < H(\mathbf{w}_0) - \mathrm{KLD}(\mathbf{w}' \parallel \mathbf{w}_0)$ then we have both $\hat{\mathbf{w}}_\infty \neq \mathbf{w}_{\mathrm{MKLD}}^{(1)}$ and $\hat{\mathbf{w}}_\infty \neq \mathbf{w}_{\mathrm{MKLD}}^{(2)}$.

6.2. MTL vs ML

We can similarly establish the difference between MTL and ML criteria. The ML principle can be used for the problem of filter parameter selection in the following way. Let $\mathbf{z}_{0:n}$ be a set $n+1$ independent observation vectors.[1] The ML filter parameter selection

[1] We consider the discrete-observation case and statistical independence for simplicity, but the argument can obviously be stated in more general terms.

constrained to the filtering function φ can be written as

$$\mathbf{w}_{\mathrm{ML}}^{(n)} = \arg\max_{\mathbf{w}\in\mathcal{W}} \left\{ \frac{1}{n} \sum_{k=0}^{n} \log p\big(\mathbf{z}_n | \mathbf{x}_n = \varphi_n(\mathbf{w})\big) \right\}.$$

Under mild ergodicity assumptions,

$$\mathbf{w}_{\mathrm{ML}} = \lim_{n\to\infty} \mathbf{w}_{\mathrm{ML}}^{(n)} = \arg\max_{\mathbf{w}\in\mathcal{W}} \big\{ \mathrm{E}_{p(\mathbf{z}_n)} \log p\big(\mathbf{z}_n | \mathbf{x}_n = \varphi_n(\mathbf{w})\big) \big\}$$

and simple manipulations yield

$$\mathbf{w}_{\mathrm{ML}} = \arg\min_{\mathbf{w}\in\mathcal{W}} \mathrm{KLD}\big(p(\mathbf{z}_n) \parallel p\big(\mathbf{z}_n | \mathbf{x}_n = \varphi_n(\mathbf{w})\big)\big),$$

i.e., ML estimation amounts to a search of the filter parameter vector that minimizes the KLD between the *true* pdf of the observations and the parameterized model pdf $p(\mathbf{z}_n|\mathbf{x}_n = \varphi_n(\mathbf{w}))$. If the latter model is not exact, i.e., $\nexists \mathbf{w} \in \mathcal{W}$ such that $p(\mathbf{z}_n|\mathbf{x}_n = \varphi_n(\mathbf{w})) = p(\mathbf{z}_n)$, then the model is said to be misspecified and $\mathbf{w}_{\mathrm{ML}}$ is termed a quasi-ML (QML) estimator [17].

It is straightforward to notice that the QML principle requires knowledge of system θ_d in order to build the probabilistic model of the observations. This at least involves the pdf of the noise and nuisance signals, $\mathbf{u}_n$, and the algebraic form of function θ_d. Such information is not necessarily required for MTL selection (we have assumed an unknown distortion of $\mathbf{x}_n$ all through our derivations).

6.3. MTL vs ME

When there is no model misspecification, i.e., $\exists \mathbf{w} \in \mathcal{W}$ such that $p(\mathbf{z}_n|\mathbf{x}_n = \varphi_n(\mathbf{w})) = p(\mathbf{z}_n)$, the ML solution can be written as

$$\begin{aligned}\mathbf{w}_{\mathrm{ML}} &= \arg\min_{\mathbf{w}\in\mathcal{W}} \big\{ -\mathrm{E}_{p(\mathbf{z}_n)} \log p\big(\mathbf{z}_n | \mathbf{x}_n = \varphi_n(\mathbf{w})\big) \big\} \\ &= \arg\min_{\mathbf{w}\in\mathcal{W}} H\big(\mathbf{z}_n | \mathbf{x}_n = \varphi_n(\mathbf{w})\big) = \mathbf{w}_{\mathrm{ME}},\end{aligned}$$

i.e., the ML principle yields a minimum entropy (ME) solution [17].

Besides the previous observation that finding $\mathbf{w}_{\mathrm{ME}}$ requires knowledge of the pdf of $\mathbf{z}_n$, it must be remarked that Theorem 1 and Corollary 1 show that the MTL method converges to *small*, but not minimal, DE solutions. The value of $\hat{\mathbf{w}}_\infty$ is the result of a trade off that involves the DE and the KLD between p_∞ and p_0.

7. Multiple maxima in the target likelihood

We have obtained asymptotic results for the global solution of problem (4) in some set $\mathcal{W}$. However, the target likelihood, $l_n^0(\mathbf{w})$, can be multimodal in many potential applications and computing its global maximum can become a difficult task in practice. Alternatively, the global maximum may not be unique, and the optimization algorithm should also select the desired maximum in some way.

Previous results in [13] suggest that, when possible, it is very effective to combine a limited temporal reference (i.e., a few known values within the state sequence) with the proposed statistical criterion in order to obtain a mixed scheme where optimization is easier. Alternatively, when a temporal reference is not available at all, it may be possible to set a structural constraint on the filter in order to guarantee convergence to the desired maximum. Such is the approach in [12], where the problem of blind multiuser interference suppression is investigated. Unfortunately, the latter strategy cannot be applied in all contexts, since it depends on the amount of side information available.

8. Blind timing and phase recovery in a digital receiver

In this section we study the application of the MTL criterion to blind (i.e., non-data aided) timing and phase recovery in a digital communication receiver. This is an important problem for which an optimal solution in closed-form cannot be derived. A comprehensive review of classical synchronization methods can be found in [18] and a more recent account of blind algorithms in [19].

8.1. Signal model

Let us consider a digital communication system where symbols from an arbitrary alphabet are transmitted over a frequency-flat channel. If data are transmitted in bursts of K symbols, the baseband-equivalent received signal has the form,

$$\tilde{r}(t) = ae^{j(\omega t+\theta)} \sum_{k=0}^{K-1} s_k \tilde{c}(t - kT_s + \tau) + \tilde{g}(t),$$

where t denotes continuous time, s_k is the kth transmitted symbol, $\tilde{c}(t)$ is a squared-root raised cosine pulse waveform, T_s is the symbol period, $0 \leqslant \tau < T_s$ is an unknown constant delay, ω and θ are the carrier frequency error and phase offset, respectively, and $\tilde{g}(t)$ is complex additive white Gaussian noise (AWGN). The receiver front-end consists of a matched filter that produces the signal

$$r(t) = ae^{j(\omega t+\theta)} \sum_{k=0}^{K-1} s_k c(t - kT_s + \tau) + g(t),$$

where

$$c(t) = \tilde{c}(t) * \tilde{c}^*(-t) = \frac{\sin(\pi t/T_s)}{\pi t/T_s} \frac{\cos(\pi\alpha t/T_s)}{1 - 4\alpha^2 t^2/T_s^2}$$

is a raised cosine waveform with roll-off factor $0 < \alpha \leqslant 1$ and $g(t) = \int_{-\infty}^{\infty} \tilde{g}(u) \times \tilde{c}^*(u - t)\,du$ is a Gaussian noise process with autocorrelation function $R_g(t) = N_0 c(t)$. If $r(t)$ is sampled at the symbol rate and the raised cosine waveform, $c(t)$, is time limited (which is always the case in practice), the discrete-time signal

$$r_k = r(kT) = ae^{j(\omega kT_s+\theta)} \sum_{n=1-L}^{L} s_{k+n} c(-nT_s + \tau) + g_k$$

is obtained, where $g_k = g(kT)$ is an AWGN process and L is the inter-symbol interference (ISI) span resulting from the limited time duration of $c(t)$. The latter equation can be expressed in a more compact form,

$$r_k = ae^{j(\nu k+\theta)}\mathbf{c}(\tau)^\top \mathbf{s}_k + g_k, \tag{13}$$

by defining:

- the normalized frequency error $\nu = \omega/T_s$,
- the $2L \times 1$ channel vector

$$\mathbf{c}(\tau) = \big[c\big((L-1)T+\tau\big), c\big((L-2)T+\tau\big), \ldots, c(-LT+\tau)\big]^\top,$$

- the symbol vector $\mathbf{s}_k = [s_{k-L+1}, s_{k-L+2}, \ldots, s_{k+L}]^\top$.

For clarity of presentation, we focus on the estimation of τ and θ alone, while assuming $a = 1$ and $\nu = 0$, and reduce the observations to the form

$$r_k = e^{j\theta}\mathbf{c}(\tau)^\top \mathbf{s}_k + g_k. \tag{14}$$

Notwithstanding, it is conceptually straightforward to extend the proposed MTL criterion to the general case where all parameters must be jointly estimated.

8.2. MTL blind timing and phase recovery

In order to obtain ISI free symbol estimates, the digital receiver should sample the received signal $r(t)$ at $t = kT_s - \tau$. However, τ is a priori unknown and the receiver samples the matched filter output at $t = kT_s - \tilde{\tau}$, where $\tilde{\tau}$ is an estimate of the true delay. The also unknown phase rotation, θ, is corrected using the estimate $\tilde{\theta}$. Hence, using the MTL terminology developed through the paper, we have the dynamic model

$$\mathbf{s}_k = \mathbf{A}\mathbf{s}_{k-1} + \mathbf{u}_k, \qquad \tilde{r}_k = e^{j(\theta-\tilde{\theta})}\mathbf{c}(\tau-\tilde{\tau})^\top \mathbf{s}_k + g_k,$$

where $\mathbf{s}_k$ is the system state, $\mathbf{A}$ is a $2L \times 2L$ shifting matrix with $A_{i,i+1} = 1$, $i = 1, \ldots, 2L-1$, and 0 otherwise, $\mathbf{u}_k = [0, \ldots, 0, s_{k+L}]^\top$ is the $2L \times 1$ system noise vector that contains the new symbol, and r_k is the discrete-time measurement. We assume s_{k+L} is a discrete uniform random variable in the alphabet $\mathcal{S}$, given by the modulation format. The signal we wish to estimate is the symbol sequence $s_{0:K-1}$ (which is equivalent to estimating the state sequence $\mathbf{s}_{0:K-1}$) and the kth measurement is obtained from the continuous-time received signal $r(t)$ by means of the transformation $\tilde{r}_k = \varphi(r(t), \tilde{\tau}, \tilde{\theta}) = e^{-j\tilde{\theta}} r(t = kT_s - \tilde{\tau})$. Therefore, our generic filter φ, in this case, consists of a symbol-rate sampler, which is fully parameterized by the timing epoch $\tilde{\tau}$, followed by a phase rotation.

If $\tilde{\tau}$ and $\tilde{\theta}$ are perfectly chosen in order to avoid ISI and correct the phase offset, $\tilde{\tau} = \tau$ and $\tilde{\theta} = \theta$, the channel vector reduces to $\mathbf{c}(0) = [0, \ldots, \underbrace{1}_{L}, \ldots, 0]^\top$ and the resulting pdf of the ISI-free measurements is the target pdf,

$$p_0(\tilde{r}_k) \propto \sum_{s \in \mathcal{S}} \exp\left\{-\frac{1}{N_0}|\tilde{r}_k - s|^2\right\},$$

where $p_0(\tilde{r}_k) = p(\tilde{r}_k)$ if, and only if, $\tilde{r}_k = \varphi(r(t), \tau, \theta) = e^{-j\theta} r(t = kT_s - \tau)$. The target likelihood at time n, therefore, is

$$l_n^0(\tilde{\tau}, \tilde{\theta}) = \sum_{k=0}^{n} \log p_0\bigl(e^{-j\tilde{\theta}} r(t = kT_s - \tilde{\tau})\bigr)$$

and the MTL solution at time n is

$$(\hat{\tau}_n, \hat{\theta}_n) = \arg\max_{\tilde{\tau},\tilde{\theta}} l_n^0(\tilde{\tau}, \tilde{\theta}) = \arg\max_{\tilde{\tau},\tilde{\theta}} \left\{ \sum_{k=0}^{n} \log \sum_{s\in\mathcal{S}} e^{-\frac{1}{N_0}|e^{-j\tilde{\theta}} r(t=kT_s-\tilde{\tau})-s|^2} \right\}. \tag{15}$$

8.3. Adaptive implementation using a gradient algorithm

Although other procedures can be explored (e.g., the expectation maximization algorithm as suggested in [12]), a simple way to adaptively solve (15) is by using a stochastic gradient algorithm of the form

$$\hat{\tau}_{n+1} = \hat{\tau}_n + \frac{\mu_\tau}{n} \frac{\partial l_n^0(\tilde{\tau}, \tilde{\theta})}{\partial \tilde{\tau}} \bigg|_{\tilde{\tau}=\hat{\tau}_n,\ \tilde{\theta}=\hat{\theta}_n}, \tag{16}$$

$$\hat{\theta}_{n+1} = \hat{\theta}_n + \frac{\mu_\theta}{n} \frac{\partial l_n^0(\tilde{\tau}, \tilde{\theta})}{\partial \tilde{\theta}} \bigg|_{\tilde{\tau}=\hat{\tau}_n,\ \tilde{\theta}=\hat{\theta}_n}, \tag{17}$$

where μ_τ/n and μ_θ/n are step-size parameters and the partial derivatives of the likelihood function are approximated using only the nth measurement, i.e.,

$$\frac{\mu_\tau}{n} \frac{\partial l_n^0(\tilde{\tau}, \tilde{\theta})}{\partial \tilde{\tau}} \bigg|_{\tilde{\tau}=\hat{\tau}_n,\ \tilde{\theta}=\hat{\theta}_n} \approx \mu_\tau \frac{\partial}{\partial \tilde{\tau}} \log p_0\bigl(e^{-j\tilde{\theta}} r(t = nT_s - \tilde{\tau})\bigr)\big|_{\tilde{\tau}=\hat{\tau}_n,\ \tilde{\theta}=\hat{\theta}_n}, \tag{18}$$

$$\frac{\mu_\theta}{n} \frac{\partial l_n^0(\tilde{\tau}, \tilde{\theta})}{\partial \tilde{\tau}} \big|_{\tilde{\tau}=\hat{\tau}_n,\ \tilde{\theta}=\hat{\theta}_n} \approx \mu_\theta \frac{\partial}{\partial \tilde{\theta}} \log p_0\bigl(e^{-j\tilde{\theta}} r(t = nT_s - \tilde{\tau})\bigr)\big|_{\tilde{\tau}=\hat{\tau}_n,\ \tilde{\theta}=\hat{\theta}_n}. \tag{19}$$

The right-hand sides of (18) and (19) can be easily calculated and yield

$$\frac{\partial}{\partial \tilde{\tau}} \log p_0(\tilde{r}_n) = -\frac{2\sum_{s\in\mathcal{S}} \exp\{-\frac{1}{N_0}|\epsilon_n|^2\} \mathfrak{Re}\{\epsilon_n^*\}}{N_0 \sum_{s\in\mathcal{S}} \exp\{-\frac{1}{N_0}|\epsilon_n|^2\}} \frac{dr(t)}{d\tilde{\tau}} \bigg|_{t=nT_s-\tilde{\tau}},$$

$$\frac{\partial}{\partial \tilde{\theta}} \log p_0(\tilde{r}_n) = \frac{2\sum_{s\in\mathcal{S}} \exp\{-\frac{1}{N_0}|\epsilon_n|^2\} \mathfrak{Im}\{e^{-j\tilde{\theta}} r(t) s^*\}}{N_0 \sum_{s\in\mathcal{S}} \exp\{-\frac{1}{N_0}|\epsilon_n|^2\}} \bigg|_{t=nT_s-\tilde{\tau}},$$

respectively, where we recall $\tilde{r}_n = e^{-j\tilde{\theta}} r(t = nT_s - \tilde{\tau})$, $\epsilon_n = r(t = nT_s - \tilde{\tau}) - e^{j\tilde{\theta}} s$ is an error signal, $\mathfrak{Re}\{\cdot\}$ and $\mathfrak{Im}\{\cdot\}$ denote the real and imaginary parts of a complex argument, respectively, and the derivative of the measurements, $dr(t = nT_s - \tilde{\tau})/d\tilde{\tau}$, can be obtained from a derivative matched filter, with impulse response $d\tilde{c}^*(-t)/dt$, in a way analogous to the approximate ML blind timing error detector (TED) in [18].

8.4. Computer simulations

We have carried out computer simulations to illustrate the performance of the MTL timing and phase recovery algorithm given by (16), (17), (18) and (19), and compare it with the ML TED described in [18, Chapter 7] for blind timing recovery and the ML-based clockless phase recovery algorithm of [18, Chapter 5]. The latter is a classical method derived under the assumption of low signal-to-noise ratio (SNR). It has similar requirements as (16), (18) (two filters are needed: one matched to the pulse waveform and another one matched to its derivative) and better performance than most practical techniques, e.g., the well-known early-late TED [18], which is a further simplification of the ML TED. Using the notation in this paper, the ML TED can be written as

$$\hat{\tau}_{n+1} = \tau_n + \mu \Re\mathfrak{e}\left\{ r^*(t) \frac{dr(t)}{dt} \right\}_{t=nT_s-\hat{\tau}_n}, \tag{20}$$

where μ is a step-size parameter. The phase recovery technique is a feedforward method based on the ML that can be written as

$$\hat{\theta}_n = \frac{1}{|\mathcal{S}|} \arg\left\{ \sum_{k=0}^{n} r^{|\mathcal{S}|}(t = kT_s - \hat{\tau}_k) \right\}, \tag{21}$$

where $|\mathcal{S}|$ is the symbol alphabet size. The ML TED (20) is independent of the phase rotation, but accurate convergence of (21) is highly dependent on the quality of the delay estimates.

For the simulations, we have considered a system with QPSK modulation (hence $s_k \in \{e^{j2\pi k/4}\}$, $k = 0, 1, 2, 3$) and raised-cosine pulses with roll-off factor $\alpha = 0.7$ and limited time duration of four periods, i.e., $L = 2$. Bits are transmitted in frames of size $K + 1 = 400$, the transmission rate is set to $R_b = 256$ Kb ps and, consequently, the symbol period is $T_s \simeq 3.9$ µs. We have chosen three representative values of the low-to-medium SNR region, namely SNR $= 2, 6$, and 10 dB s, and run 500 independent simulations for each value. In each simulation, we have compared the proposed MTL algorithm given by (16), (17), (18) and (19), and the approximate ML TED of (20) together with the feedforward phase estimator (21). The adaptive algorithms have been run with different choices of step-size parameters, as shown in Table 1 (note the exponentially decreasing adaptation steps

Table 1
Values of the step-size parameter for each algorithm and SNR value

	2 dB	6 dB	10 dB
MTL (1)	$\mu_\tau = 10^{-8} T_s$ $\mu_\theta = 0.990^n \times 10^{-1}$	$\mu_\tau = 6 \times 10^{-9} T_s$ $\mu_\theta = 0.985^n \times 3 \times 10^{-2}$	$\mu_\tau = 2 \times 10^{-9} T_s$ $\mu_\theta = 0.985^n \times 3 \times 10^{-2}$
MTL (2)	$\mu_\tau = 4 \times 10^{-8} T_s$ $\mu_\theta = 0.990^n \times 10^{-1}$	$\mu_\tau = 2.5 \times 10^{-8} T_s$ $\mu_\theta = 0.985^n \times 3 \times 10^{-2}$	$\mu_\tau = 7 \times 10^{-9} T_s$ $\mu_\theta = 0.980^n \times 3 \times 10^{-2}$
ML (1)	$\mu = 8 \times 10^{-3} T_s$	$\mu = 10^{-2} T_s$	$\mu = 10^{-2} T_s$
ML (2)	$\mu = 3 \times 10^{-2} T_s$	$\mu = 5 \times 10^{-2} T_s$	$\mu = 5 \times 10^{-2} T_s$

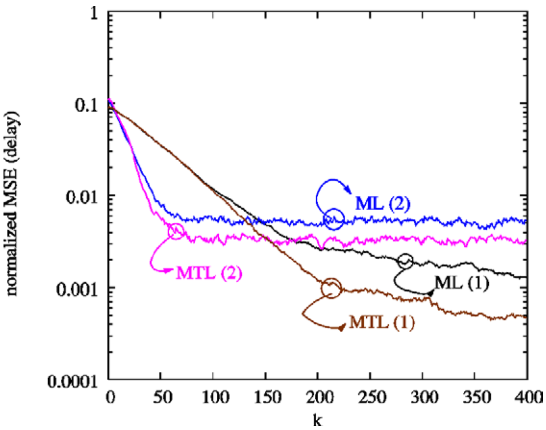

Fig. 2. Normalized MSE vs discrete time for timing recovery algorithms, SNR = 2 dB.

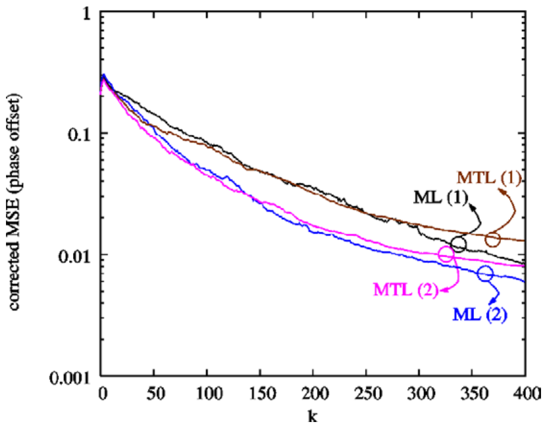

Fig. 3. Corrected MSE vs discrete time for phase recovery algorithms, SNR = 2 dB.

for phase estimation). As figures of merit, we have selected the normalized mean square error (MSE) with respect to τ, i.e.,

$$\mathrm{MSE}_n(\tau) = \frac{1}{500}\sum_{i=1}^{500}\frac{|\tau - \hat{\tau}_n|^2}{T_s},$$

and the corrected MSE

$$\mathrm{MSE}_n(\theta) = \min_{\alpha\in\{0,\pi/2,\pi,3\pi/2\}}\left\{|\theta - \hat{\theta}_n - \alpha|^2\right\}$$

with respect to θ, that accounts for the inherent ambiguity in blind phase recovery [18, Chapter 5].

Figure 2 shows the evolution of the normalized error, $\mathrm{MSE}_n(\tau)$, for the different timing recovery methods when the SNR is set to 2 dB. We observe that the MTL algorithms perform sensibly better, in the sense that they attain a lower MSE with the same convergence

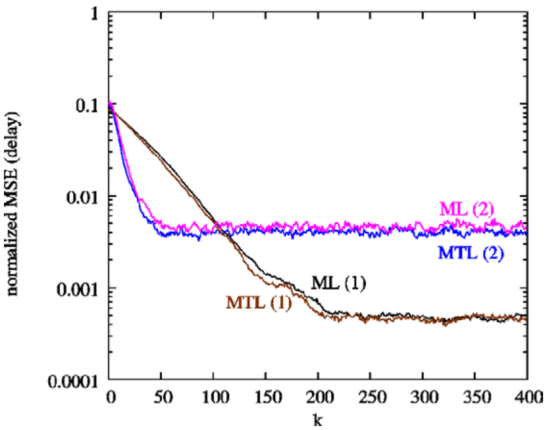

Fig. 4. Normalized MSE vs discrete time for timing recovery algorithms, SNR = 6 dB.

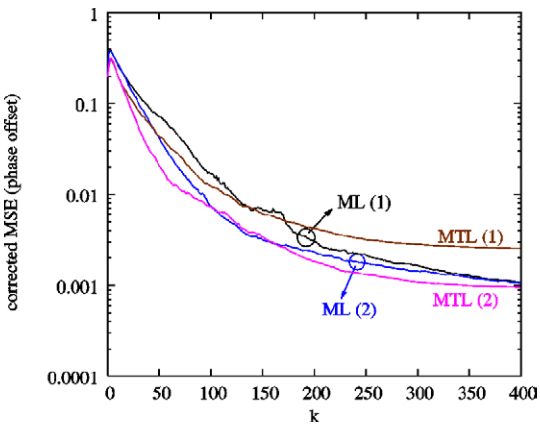

Fig. 5. Corrected MSE vs discrete time for phase recovery algorithms, SNR = 6 dB.

speed. This is true both for the ‘slow’ TEDs (labeled MTL (1) and ML (1)), which use a small step-size parameter, and the ‘fast’ TEDs (labeled MTL (2) and ML (2)) with larger adaptation steps.

The corrected phase error, $\mathrm{MSE}_n(\theta)$, for SNR = 2 dB is shown in Fig. 3. Although both MTL and ML-based techniques achieve similar results, it is the latter that attains a slightly lower MSE for this experiment. Remarkably, the best performance in terms of phase recovery is attained by the ‘fast’ algorithms, labeled MTL (2) and ML (2).[2] This means that phase offsets are better compensated for when timing is quickly adjusted, even if the steady value of $\mathrm{MSE}_n(\tau)$ is far from the lowest possible one.

[2] The ML phase estimator is the same for ML (1) and ML (2), but its performance depends on the convergence of the coupled ML TED.

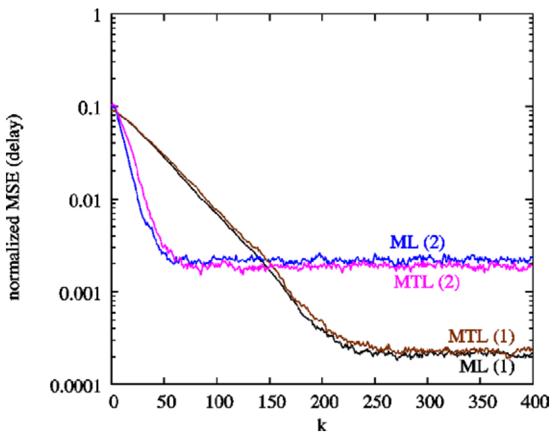

Fig. 6. Normalized MSE vs discrete time for timing recovery algorithms, SNR = 10 dB.

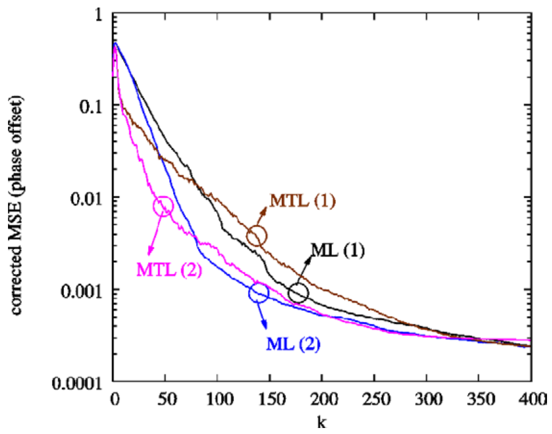

Fig. 7. Corrected MSE vs discrete time for phase recovery algorithms, SNR = 10 dB.

The performance of the MTL and ML synchronization algorithms becomes closer for higher SNR values. Figure 4 shows the normalized MSE attained by the TEDs when SNR = 6 dB. It is seen that both types of TEDs achieve a very similar performance, both for slow (ML (1) and MTL (1)) and fast (ML (2) and MTL (2)) adaptation. Although the steady state error of the MTL algorithms is slightly smaller than their ML counterparts, the difference is hardly significant.

Phase correction performance for SNR = 6 dB is illustrated in Fig. 5. Best phase recovery is attained by the fast-adaptation MTL algorithm, although the advantage with respect to the ML-based feedforward estimators is very small.

Finally, Figs. 6 and 7 plot the normalized delay MSE, $\mathrm{MSE}_n(\tau)$, and corrected phase MSE, $\mathrm{MSE}_n(\theta)$, respectively, for SNR = 10 dB. We observe that, as the SNR becomes higher, the performance of the algorithms becomes nearly equivalent, both in terms of timing recovery and phase correction.

8.5. Remarks

The adaptive algorithm for blind timing and phase recovery resulting from the MTL criterion turns out to be competitive with well-known ML techniques described in [18]. In particular, we have found that, compared to the conventional ML TED, the MTL algorithm has a superior convergence speed in the low SNR region, while both approaches become approximately equivalent (in terms of their normalized MSE curves) for higher SNR values. We have observed that phase recovery is highly dependent on the convergence properties of the timing estimation part of the algorithms. This is true both for the MTL techniques, where τ and θ are jointly estimated, and the ML-based approach, where the TED is phase-insensitive but phase recovery is very dependent on fast timing correction. In any case, phase correction attained by MTL and ML algorithms is equally satisfactory.

9. Conclusions

We have addressed the analysis of a novel criterion for the selection of filtering parameters that relies on the ability to characterize the filter output in terms of a *target* pdf. The latter density is then used as a likelihood function of the parameters, which can be selected as in a maximum likelihood problem. For this reason, the criterion has been termed maximum target likelihood (MTL).

The method has been described within a very general framework and an asymptotic convergence theorem that characterizes MTL solutions under few constraints has been stated and proved. Using this convergence result, the relationship and differences between the proposed approach and standard statistical (ML) and information theoretic (minimum KLD, minimum entropy) methodologies have also been explored. Finally, as an example, we have applied the MTL criterion to the problem of blind adaptive timing and phase recovery. The resulting algorithm has been shown to be competitive with existing maximum likelihood based algorithms, and we expect to successfully extend it in the future to tackle the generalized synchronization problem (joint timing, phase and frequency recovery) in more complex scenarios (e.g., multiple-input multiple-output channels).

Acknowledgments

J. Míguez acknowledges the support of the Ministry of Science and Technology of Spain and Xunta de Galicia (project TIC2001-0751-C04-01). M.F. Bugallo acknowledges the support of the National Science Foundation (award CCR-0082607).

References

[1] S. Haykin (Ed.), Blind Deconvolution, Prentice Hall International, Englewood Cliffs, NJ, 1994.

[2] L. Tong, S. Perreau, Multichannel blind identification: From subspace to maximum likelihood methods, Proc. IEEE 86 (10) (1998) 1951–1968.

[3] A.K. Jain, P.W. Duin, J. Mao, Statistical pattern recognition: A review, IEEE Trans. Pattern Anal. Machine Intell. 22 (1) (2000) 4–37.
[4] S. Haykin, Adaptive Filter Theory, third ed., Information and System Sciences Series, Prentice Hall International, Englewood Cliffs, NJ, 1996.
[5] A. Oppenheim, R. Schafer, Discrete-Time Signal Processing, Prentice Hall International, Englewood Cliffs, NJ, 1989.
[6] S. Roberts, R. Everson (Eds.), Independent Component Analysis. Principles and Practice, Cambridge Univ. Press, Cambridge, UK, 2001.
[7] D.N. Godard, Self recovering equalization and carrier tracking in two-dimensional data communication systems, IEEE Trans. Commun. COMM 28 (1980) 1867–1875.
[8] J.F. Cardoso, B.H. Laheld, Equivariant adaptive source separation, IEEE Trans. Signal Process. 4 (4) (1997) 112–114.
[9] P. Stoica, A. Nehorai, Performance comparison of subspace rotation and MUSIC methods for direction estimation, IEEE Trans. Signal Process. 39 (2) (1991) 357–361.
[10] S.J. Nowlan, G.E. Hinton, A soft decision-directed LMS algorithm for blind equalization, IEEE Trans. Commun. 41 (2) (1993) 275–279.
[11] J. Sala-Álvarez, G. Vázquez-Grau, Statistical reference criteria for adaptive signal processing in digital communications, IEEE Trans. Signal Process. 45 (1) (1997) 14–31.
[12] M.F. Bugallo, J. Míguez, L. Castedo, A maximum likelihood approach to blind multiuser interference cancellation, IEEE Trans. Signal Process. 49 (6) (2001) 1228–1239.
[13] M.F. Bugallo, J. Míguez, L. Castedo, Semiblind linear multiuser interference cancellation: A maximum likelihood approach, Signal Process. 81 (10) (2001) 2041–2057.
[14] N.S. Pannu, G.N. Murshudov, E.J. Dodson, R.J. Read, Incorporation of prior phase information strengthens maximum likelihood structure refinement, Acta Crystallogr. Biol. Crystallogr. 54 (2) (1998) 1285–1294.
[15] J. Míguez, M.F. Bugallo, Parametric estimation of hidden signals by likelihood maximization, in: Proceedings of 14th International Conference on Digital Signal Processing, vol. 1, 2002, pp. 143–146.
[16] T.M. Cover, J.A. Thomas, Elements of Information Theory, Wiley–Interscience, New York, 1991.
[17] H. White, Maximum likelihood estimation of misspecified models, Econometrica 50 (1) (1982) 1–26.
[18] U. Mengali, A. D'Andrea, Synchronization Techniques for Digital Receivers, Plenum, 1997.
[19] G. Vázquez, J. Riba, Non-data-aided digital synchronization, in: G.B. Giannakis, Y. Hua (Eds.), Signal Processing Advances in Wireless Communications, in: Trends in Single and Multi-User Systems, vol. II, Prentice Hall International, Englewood Cliffs, NJ, 2000, ch. 9.