This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the IEEE Globecom 2011 proceedings.

A Maximum Likelihood Time Delay Estimator Using Importance Sampling

Ahmed Masmoudi, Faouzi Bellili, Sofiène Affes, and Alex Stéphenne

INRS-EMT, 800, de la Gauchetière Ouest, Bureau 6900, Montreal, Qc, H5A 1K6, Canada
Emails: {masmoudi, bellili, affes}@emt.inrs.ca, stephenne@ieee.org

***Abstract*—In this paper, we present a new time delay estimator for multipath environments using the importance sampling (IS) method. The new technique allows finding the maximum of the compressed likelihood function in an efficient manner. The main idea consists in generating realizations of a random variable distributed according to a function that approximates the actual compressed likelihood function and then computing the mean of the plausible realizations. We avoid eigen-decomposition operation that is widely used in the conventional high-resolution methods. We show through computer simulations that the new algorithm provides accurate estimates for closely spaced unknown time delays. Moreover, the method does not suffer from lack of convergence and initialization problems that arise with other iterative implementations of the maximum likelihood estimator.**

***Index Terms*—Timing recovery, Monte-Carlo methods, optimization methods, maximum likelihood (ML) estimation, high-resolution methods.**

I. Introduction

Time delay estimation is an important problem with applications in many areas such as radar, underwater acoustics and wireless communication system, etc. Traditionally, time delay estimation is accomplished by matched filter or correlation techniques. The performance of such procedures is inversely proportional to the signal bandwidth and they cannot separate paths with closely spaced delays [2]. Nevertheless, the separation of closely spaced delays has been studied intensively and many high-resolution algorithms have been proposed to accomplish this task. In this context, it is well known that the maximum likelihood (ML) estimator is asymptotically (in the sense of large number of snapshots) efficient and achieves the well known Cramer-Rao lower bound (CRLB) especially for moderate and high SNR values. Nonetheless, many existing procedures of finding the ML estimates can be daunting because of their numerical complexity and computational burden. In fact, the straightforward implementation of the ML criterion is based on a multidimensional grid search whose complexity increases with the number of parameters. Some other proposed methods tried to solve the ML problem in an iterative manner [3]. On the other hand, suboptimal techniques based on the signal subspace, such as MUSIC and ESPIRIT, which have been successfully applied to the estimation of other parameters such as directions of arrival, were also translated to the time delay estimation problem. Indeed, a direct application of the MUSIC algorithm was earlier proposed in [4] with not much performance gain with respect to the correlation-based methods. An analogy between time delay and frequency estimation problems was also found by Hou and Wu in [5] where the problem of time delay estimation has been formulated in the frequency domain. Later on, high-resolution methods for frequency finding were successfully applied to time delay estimation [6, 7]. However, the performance of all these estimators is severely affected by the bandwidth of the transmitted signal. Moreover, all these methods demand the computation of the autocorrelation matrix from large number of received data. Motivated by these facts, an alternative approach is proposed in this paper where grid search, iterative and eigen-decomposition operation are avoided. We implement the ML estimator using the concept of importance sampling (IS) which has been shown as a powerful tool to perform ML estimation of multiple unknown parameters. IS-based ML estimators were already derived and successfully implemented for the estimation of other crucial parameters such as the direction of arrival (DOA) [8] or more recently for the joint DOA-Doppler frequency finding [9]. While these results were primary established in the context of antenna arrays, we consider the IS-based ML time delay estimation in completely unknown multipath environments. It will be shown that the new method allows accurate estimation of the time delays even at low SNR values.

The remainder of this paper is organized as follows. In section II, we present the system model and derive the compressed likelihood function of the received signal in active systems. In section III, the global maximization method that will be used is described. Then we introduce in section IV the IS technique which will be used in section V for the estimation of the delays. The newly proposed algorithm is applied to passive systems in section VI and simulation results are discussed in section VII. Finally, some concluding remarks are drawn out in section VIII.

II. Signal Model and Compressed Likelihood Function

In a multipath environment, the received signal is a superposition of multiple delayed replicas of the known transmitted waveform. It is usually modeled as follows:

$$y(t) = \sum_{i=1}^{P} \alpha_i x(t-\tau_i) + w(t), \qquad (1)$$

Work supported by a Canada Research Chair in Wireless Communications and a Discovery Accelerator Supplement from NSERC.

where P is the number of multipath components, $x(t)$ is the *a priori* known transmitted waveform, $w(t)$ is an additive noise and $\boldsymbol{\alpha} = [\alpha_1,\ \alpha_2, \ldots,\ \alpha_P]^T$ are the unknown path amplitudes resulting from scatter characteristics and propagation fading through the medium. In addition, $\boldsymbol{\tau} = [\tau_1,\ \tau_2, \ldots,\ \tau_P]^T$ are the unknown time delays to be estimated. At the receiver side, the received signal is passed through an ideal lowpass filter of bandwidth $F_s/2$, where F_s is the sampling frequency sampled at a rate $T_s = 1/F_s$. The sampled signal can be written as:

$$y(nT_s) = \sum_{i=1}^{P} \alpha_i x(nT_s - \tau_i) + w(nT_s),\ n = 0,\ 1, \ldots,\ N-1, \tag{2}$$

where N represents the total number of available samples.

In the context of array signal processing, it was recently shown that the IS-based ML estimator can separate closely spaced DOAs in [8]. In the sequel, we apply the IS principle to the ML time delay estimation. However, this method is more suitable for the typical class of general linear models (GLM) having the following form:

$$\boldsymbol{y} = \boldsymbol{\Phi}(\boldsymbol{\tau})\boldsymbol{s} + \boldsymbol{w} \tag{3}$$

where $\boldsymbol{y}$ is the noise corrupted observation vector which depends on the unknown parameters $\boldsymbol{s}$ and $\boldsymbol{\tau}$ linearly and nonlinearly, respectively. The original input-output relationship in (2) does not allow a direct formulation which is analogous to (3). Therefore, the received samples are transformed into the frequency domain where the model could be expressed in a matrix form. In fact, taking the discrete Fourrier transform of (2), we obtain:

$$Y(k) \;=\; \sum_{i=1}^{P} \alpha_i X(k) \exp\{-\frac{j2\pi k \tau_i}{N}\} + W(k), \tag{4}$$

where $\{Y(k)\}_{k=0}^{N-1}$, $\{X(k)\}_{k=0}^{N-1}$ and $\{W(k)\}_{k=0}^{N-1}$ are the discrete Fourier transforms (DFTs) of $\{y(nT_s)\}_{n=0}^{N-1}$, $\{x(nT_s)\}_{n=0}^{N-1}$ and $\{w(nT_s)\}_{n=0}^{N-1}$, respectively. To bring the problem model into the form of (3), we denote $\boldsymbol{Y} = [Y(0),\ Y(1), \ldots,\ Y(N-1)]^T$ and we formulate a compact representation of (4) as follows:

$$\boldsymbol{Y} = \boldsymbol{\Phi}_a(\boldsymbol{\tau})\boldsymbol{\alpha} + \boldsymbol{W}, \tag{5}$$

in which the matrix $\boldsymbol{\Phi}_a(\boldsymbol{\tau})$ depends only on the unknown delays gathered in the vector $\boldsymbol{\tau}$ and is given by:

$$\boldsymbol{\Phi}_a(\boldsymbol{\tau}) = [\boldsymbol{\phi}_a(\tau_1),\ \boldsymbol{\phi}_a(\tau_2), \ldots,\ \boldsymbol{\phi}_a(\tau_P)], \tag{6}$$

with the columns $\{\boldsymbol{\phi}_a(\tau_i)\}_{i=1}^{P}$ being defined as:

$$\begin{aligned}\boldsymbol{\phi}_a(\tau_i) \;=\; & [S(0),\ S(1)e^{-\frac{j2\pi\tau_i}{N}},\ S(2)e^{-\frac{j2\pi 2\tau_i}{N}} \ldots, \\ & S(N-1)e^{-\frac{j2\pi(N-1)\tau_i}{N}}]^T,\end{aligned} \tag{7}$$

and $\boldsymbol{S} = [S(0),\ S(1), \ldots,\ S(N-1)]^T$ and $\boldsymbol{W} = [W(0),\ W(1), \ldots,\ W(N-1)]^T$ are the $N \times 1$ vectors containing the DFT coefficients of samples corresponding to the known transmitted pulse and the additive noise components, respectively.

In this work, the Gaussianity of the noise is not needed, since the noise samples are obtained from any continuous time process. But since for each k the Fourier coefficient $W(k)$ is simply a weighted sum of these samples, then according to the central limit theorem, $W(k)$ can be modeled as a Gaussian random variable for a sufficiently large number of received samples. Consequently, following the same arguments of [10], the likelihood function of the active model (5) is given by:

$$\begin{aligned}\Lambda(\boldsymbol{\tau},\boldsymbol{\alpha}) \propto p(\boldsymbol{Y};\boldsymbol{\tau},\boldsymbol{\alpha}) = \frac{1}{\pi^M \sigma_W^{2M}} \exp\Bigg\{ -\frac{1}{\sigma_W^2} \times \\ (\boldsymbol{Y} - \boldsymbol{\Phi}_a(\boldsymbol{\tau})\boldsymbol{\alpha})^H (\boldsymbol{Y} - \boldsymbol{\Phi}_a(\boldsymbol{\tau})\boldsymbol{\alpha}) \Bigg\},\end{aligned} \tag{8}$$

where $p(\boldsymbol{Y};\boldsymbol{\tau},\boldsymbol{\alpha})$ is the probability density function (pdf) of $\boldsymbol{Y}$ parameterized by $\boldsymbol{\tau}$ and $\boldsymbol{\alpha}$ and σ_W is the variance of $W(k)$. Traditionally, the ML solution $\widehat{\boldsymbol{\tau}}_{ML}$ is obtained by maximizing the likelihood function in (8) with respect to $\boldsymbol{\tau}$. However, $\Lambda(\boldsymbol{\tau},\boldsymbol{\alpha})$ depends on $\boldsymbol{\tau}$ and $\boldsymbol{\alpha}$ and their joint estimation is computationally intensive. Therefore, it is of interest to obtain a likelihood function that depends only on $\boldsymbol{\tau}$. To that end, we consider the nuisance parameter, $\boldsymbol{\alpha}$, as deterministic but unknown and substitute, in (8), $\boldsymbol{\alpha}$ by the solution $\widehat{\boldsymbol{\alpha}}(\boldsymbol{\tau})$ which maximizes the log-likelihood function $L(\boldsymbol{\tau},\boldsymbol{\alpha}) = \log\{\Lambda(\boldsymbol{\tau},\boldsymbol{\alpha})\}$ for a given $\boldsymbol{\tau}$. Actually, it can be shown that $\widehat{\boldsymbol{\alpha}}(\boldsymbol{\tau})$ is given by:

$$\widehat{\boldsymbol{\alpha}}(\boldsymbol{\tau}) = (\boldsymbol{\Phi}_a(\boldsymbol{\tau})^H \boldsymbol{\Phi}_a(\boldsymbol{\tau}))^{-1} \boldsymbol{\Phi}_a(\boldsymbol{\tau})^H \boldsymbol{Y}. \tag{9}$$

Substituting $\boldsymbol{\alpha}$ by $\widehat{\boldsymbol{\alpha}}(\boldsymbol{\tau})$ in (8) and omitting the terms that do not depend on $\boldsymbol{\tau}$, we get the so-called compressed likelihood function of the system as follows:

$$L_c(\boldsymbol{\tau}) = \frac{1}{\sigma_W^2} \boldsymbol{Y}^H \boldsymbol{\Phi}_a(\boldsymbol{\tau}) \Big(\boldsymbol{\Phi}_a(\boldsymbol{\tau})^H \boldsymbol{\Phi}_a(\boldsymbol{\tau}) \Big)^{-1} \boldsymbol{\Phi}_a(\boldsymbol{\tau})^H \boldsymbol{Y}. \tag{10}$$

III. Global Maximization of the Compressed Likelihood Function

Our goal is to maximize $L_c(\boldsymbol{\tau})$ over the unknown parameter $\boldsymbol{\tau}$. The direct implementation of this optimization problem requires a multidimensional grid search whose complexity increases exponentially with the number of parameters to be estimated. Many alternative approaches have been proposed to solve multidimensional optimization problems with relatively good performance. But, most of these alternatives are iterative and therefore require good initialization, otherwise they may not converge to the true maximum.

In this context, Pincus proposed an alternative approach that guarantees obtaining the global maximum of a multidimensional function. According to this theorem in [11], the global multidimensional maximum $\widehat{\boldsymbol{\tau}} = [\widehat{\tau}_1,\ \widehat{\tau}_2, \ldots,\ \widehat{\tau}_P]^T$ of the cost function $L_c(\boldsymbol{\tau})$ is given by:

$$\widehat{\tau}_i = \lim_{\rho \to \infty} \frac{\int_J \cdots \int_J \tau_i \exp\{\rho L_c(\boldsymbol{\tau})\}\, d\boldsymbol{\tau}}{\int_J \cdots \int_J \exp\{\rho L_c(\boldsymbol{u})\}\, d\boldsymbol{u}}, \tag{11}$$

where J is the interval in which the delays are confined. Defining the pseudo-pdf $L'_{c,\rho_0}(\boldsymbol{\tau})$, for some large value of ρ_0, as:

$$L'_{c,\rho_0}(\boldsymbol{\tau}) = \frac{\exp\{\rho_0 L_c(\boldsymbol{\tau})\}}{\int_J \cdots \int_J \exp\{\rho_0 L_c(\boldsymbol{u})\}\, d\boldsymbol{u}}, \tag{12}$$

the optimal value of τ_i is:

$$\widehat{\tau}_i = \int_J \cdots \int_J \tau_i L'_{c,\rho_0}(\boldsymbol{\tau}) d\boldsymbol{\tau}, \ i = 1,\ 2,\ \ldots,\ P. \tag{13}$$

In fact, as ρ_0 tends to infinity, the function $L'_{c,\rho_0}(\boldsymbol{\tau})$ becomes a P-dimensional Dirac function centered at the location of the maximum of $L_c(\boldsymbol{\tau})$. $L'_{c,\rho_0}(\boldsymbol{\tau})$ is designated as a pseudo-pdf since it has all the properties of a pdf although $\boldsymbol{\tau}$ is not really a random variable. Yet, the formulation given in (13) requires the evaluation of a multidimensional integral (a direct integration remains always difficult if not impossible). Alternatively, $\widehat{\tau}_i$ can be interpreted as the expected value of τ_i, the i^{th} element of the vector $\boldsymbol{\tau}$ that is distributed according the pseudo-pdf $L'_{c,\rho_0}(\boldsymbol{\tau})$. This kind of expectation can be efficiently evaluated using Monte-Carlo methods as follows:

$$\widehat{\boldsymbol{\tau}} = \frac{1}{R} \sum_{k=1}^{R} \boldsymbol{\tau}_k, \tag{14}$$

where $\{\boldsymbol{\tau}_k\}_{k=1}^{R}$ are R vector realizations of $\boldsymbol{\tau}$ generated according to $L'_{c,\rho_0}(\boldsymbol{\tau})$. In this way, the integration in (13) is substituted by a simple samples average. Clearly, as the number of generated values R increases, the variance of the sample mean becomes smaller. But one should find a way to easily generate realizations according to $L_{c,\rho_0}(\boldsymbol{\tau})$. Unfortunately, the introduced pseudo-pdf $L'_{c,\rho_0}(\boldsymbol{\tau})$ is non-linear with respect to the parameter of interest, $\boldsymbol{\tau}$, and is not suitable for easy generation. To overcome this problem, we consider a simpler pseudo-pdf and resort to the concept of IS to remove the bias of the estimates.

IV. The Importance Sampling Technique

It has been shown that the IS method is a powerful tool to compute multidimensional integrals of type (13) in an efficient manner [12]. The main idea is based on the following observation:

$$\int_J \cdots \int_J f(\boldsymbol{\tau}) L'_{c,\rho_0}(\boldsymbol{\tau}) d\boldsymbol{\tau} = \int_J \cdots \int_J f(\boldsymbol{\tau}) \frac{L'_{c,\rho_0}(\boldsymbol{\tau})}{g'(\boldsymbol{\tau})} g'(\boldsymbol{\tau}) d\boldsymbol{\tau}, \tag{15}$$

where $g'(\boldsymbol{\tau})$ is another pseudo-pdf, called normalized importance function (IF) and $f(.)$ is a given function. Then, the left-hand side of (15) is simply the mean of $f(\boldsymbol{\tau})\frac{L'_{c,\rho_0}(\boldsymbol{\tau})}{g'(\boldsymbol{\tau})}$ if $\boldsymbol{\tau}$ is generated according to $g'(\boldsymbol{\tau})$. Therefore, it is of interest to choose $g'(\boldsymbol{\tau})$ to be a simple function of $\boldsymbol{\tau}$ that should also remain similar to the original pseudo-pdf $L'_{c,\rho_0}(\boldsymbol{\tau})$. Then, we use Monte-Carlo methods to numerically compute the multidimensional integrals in (15) as follows:

$$\int_J \cdots \int_J f(\boldsymbol{\tau}) \frac{L'_{c,\rho_0}(\boldsymbol{\tau})}{g'(\boldsymbol{\tau})} g'(\boldsymbol{\tau}) d\boldsymbol{\tau} = \frac{1}{R} \sum_{k=1}^{R} f(\boldsymbol{\tau}_k) \frac{L'_{c,\rho_0}(\boldsymbol{\tau}_k)}{g'(\boldsymbol{\tau}_k)}, \tag{16}$$

where $\boldsymbol{\tau}_k$ is now the k^{th} realization of $\boldsymbol{\tau}$ according to $g'(.)$. Clearly, the estimation performance depends on the choice of R and $g'(.)$. While large values of R improve the estimation accuracy, they also result in higher computational complexity. In practice, the value of R depends on the choice of $g'(.)$. Typically, the shapes of the pseudo-pdfs $L'_{c,\rho_0}(.)$ and $g'(.)$ should be similar to reduce the variance of the estimates given by (16) [8]. On the other hand, the normalized IF, $g'(.)$, has to be simple enough so that realizations according to $g'(.)$ can be easily generated. Clearly, between choosing a function similar to $L'_{c,\rho_0}(.)$ and a simple one, some tradeoffs must be considered and next we discuss the appropriate selection of this IF.

First, it can be seen from (10) that the term $\left(\boldsymbol{\Phi}_a(\boldsymbol{\tau})^H \boldsymbol{\Phi}_a(\boldsymbol{\tau})\right)^{-1}$ makes the actual compressed likelihood function, $L_c(.)$, or equivalently $L'_{c,\rho_0}(.)$, non linear with respect to $\boldsymbol{\tau}$. Therefore, it would be of great interest if one can avoid the matrix inversion operation. In fact, the diagonal elements are given by:

$$[(\boldsymbol{\Phi}_a(\boldsymbol{\tau})^H \boldsymbol{\Phi}_a(\boldsymbol{\tau}))]_{l;l} = \sum_{k=0}^{N-1} |S(k)|^2, \ l = 1,\ 2, \ldots,\ P, \tag{17}$$

whereas the off-diagonal elements $\boldsymbol{\Phi}_a(\boldsymbol{\tau})^H \boldsymbol{\Phi}_a(\boldsymbol{\tau})$ are:

$$[(\boldsymbol{\Phi}_a(\boldsymbol{\tau})^H \boldsymbol{\Phi}_a(\boldsymbol{\tau}))]_{m;n} = \sum_{k=0}^{N-1} |S(k)|^2 \exp\left\{\frac{j2\pi k(\tau_m - \tau_n)}{N}\right\}, \quad m,\ n = 1,\ 2, \ldots,\ P. \tag{18}$$

We can verify, for $\tau_m \neq \tau_n$, that:

$$\left|[(\boldsymbol{\Phi}_a(\boldsymbol{\tau})^H \boldsymbol{\Phi}_a(\boldsymbol{\tau}))]_{m;n}\right| < [(\boldsymbol{\Phi}_a(\boldsymbol{\tau})^H \boldsymbol{\Phi}_a(\boldsymbol{\tau}))]_{m;m}. \tag{19}$$

Therefore, the diagonal elements of $\boldsymbol{\Phi}_a(\boldsymbol{\tau})^H \boldsymbol{\Phi}_a(\boldsymbol{\tau})$ are dominant compared to its off-diagonals ones and Fig. 1 confirms this result with high probability. So, a good approximation of $\left(\boldsymbol{\Phi}_a(\boldsymbol{\tau})^H \boldsymbol{\Phi}_a(\boldsymbol{\tau})\right)^{-1}$ is the diagonal matrix $(\sum_{k=1}^{N} |S(k)|^2)^{-1} \boldsymbol{I}_p$, where $\boldsymbol{I}_p$ is the $p \times p$ identity matrix. Thus, we define the IF $g_{\rho_1}(.)$ in the active case as:

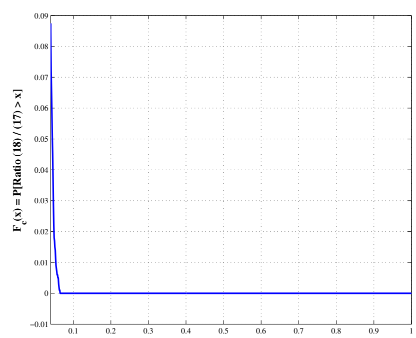

Fig. 1. Complementary cumulative distribution function of the ratio (18)/(17).

$$\begin{aligned} g_{\rho_1}(\boldsymbol{\tau}) &= \exp\left\{\frac{\rho_1}{\sigma_W^2 \sum_{k=1}^{N} |S(k)|^2} \boldsymbol{Y}^H \boldsymbol{\Phi}_a(\boldsymbol{\tau}) \boldsymbol{\Phi}_a(\boldsymbol{\tau})^H \boldsymbol{Y}\right\} \\ &= \prod_{i=1}^{P} \exp\left\{\frac{\rho_1}{\sigma_W^2 \sum_{k=1}^{N} |S(k)|^2} I(\tau_i)\right\}, \end{aligned} \tag{20}$$

where ρ_1 is another constant different from ρ_0 and $I(\tau_i)$ is the periodogram of the data in the frequency domain evaluated at the delay τ_i as follows:

$$I(\tau_i) = \Big|\sum_{k=1}^{N} S(k)^* Y(k) \exp\left\{\frac{j2\pi(k-1)\tau_i}{N}\right\}\Big|^2. \tag{21}$$

A further discussion on the appropriate choice of ρ_0 and ρ_1 can be found at the end of this section.

Actually, this choice for the IF in (20) presents many advantages. First, we notice that the joint contribution of the different delays in $g_{\rho_1}(.)$ is split into the product of their individual contributions as seen from (20). Hence, the generation of a realization of the vector $\boldsymbol{\tau}$ reduces to the generation of P independent scalar realizations (i.e., one realization for each entry of $\boldsymbol{\tau}$) using $\overline{g}_{\rho_1}(\tau_i)$, the normalized version of $\exp\left\{\frac{\rho_1}{\sigma_W^2\sum_{k=1}^{N}|S(k)|^2}I(\tau_i)\right\}$ $\left(g'_{\rho_1}(\boldsymbol{\tau}) = \prod_{i=1}^{P}\overline{g}_{\rho_1}(\tau_i)\right)$. Moreover, the multiplicative terms $S(k),\ k = 1,\ 2, \ldots,\ N$ act as weighting factors and they attenuate the contribution of the low-SNR frequencies in the computation of $I(\tau_i)$. This property may widen the application of the proposed algorithm to narrowband signals. Note also that this property improves considerably the performance of the estimator when compared to other approaches where the received signal is divided, in the frequency domain, by the known transmitted signal [7]. In fact, this operation usually results in some harmful effects by amplifying the noise for the low-energy frequencies. Indeed, while these techniques exhibit good performance for signals with large bandwidth and flat spectra, their performance depredates dramatically for narrowband signals. Finally, we rewrite the normalized IF as follows:

$$g'_{\rho'_1}(\boldsymbol{\tau}) = \frac{\prod_{i=1}^{P}\exp\left\{\rho'_1 I(\tau_i)\right\}}{\int_J\cdots\int_J\prod_{i=1}^{P}\exp\left\{\rho'_1 I(u_i)\right\}du_i}, \quad (22)$$

with

$$\rho'_1 = \frac{\rho_1}{\sigma_W^2\sum_{k=1}^{N}|S(k)|^2}. \quad (23)$$

Recall that the parameters ρ'_1, as well as ρ_0 are of great importance and impact seriously the new estimator performance. In fact, as already mentioned, $g'_{\rho'_1}(\boldsymbol{\tau})$ can be split into the product of P elementary IFs corresponding to each τ_i and hence we use $\overline{g}_{\rho'_1}(\tau)$ P times to generate the P elements of $\boldsymbol{\tau}$. In theory, $\overline{g}_{\rho'_1}(\tau)$ exhibits P lobes centered at the locations of the different delays and at each run, a realization is generated from the vicinity of one of the P lobes. However, in practice other secondary lobes appear and affect the generated values. To circumvent this problem, we may increase ρ'_1 indefinitely so that the unwanted lobes disappear. But it turn very large values of ρ'_1 may also destroy some useful lobes and therefore useful realizations may not be generated. This is why ρ'_1 should be chosen judiciously. Actually, the optimal value of ρ'_1 is the highest one for which $\overline{g}_{\rho'_1}(.)$ has at least P lobes[1]. Moreover, a good choice of ρ'_1 may allow us to reduce the number R of required realizations since it reduces the probability of generating undesired realizations. However, as previously mentioned, the global IF $g'_{\rho'_1}(\boldsymbol{\tau})$ is built upon an approximation of the actual compressed likelihood function and this results in a biased estimate of the delays. This bias is effectively avoided by the weighting factor $L'_{c,\rho_0}(\boldsymbol{\tau})/g'_{\rho'_1}(\boldsymbol{\tau})$ in (15). Therefore, we have to minimize the contribution of $g'_{\rho'_1}(\boldsymbol{\tau})$ in the considered weighting factor by making ρ_0 larger than ρ'_1.

[1]In theory, $\overline{g}_{\rho'_1}(.)$ has exactly P lobes. But the noisy observations make other lobes appear.

V. Estimation of the Time Delays

Following the developments disclosed above, the estimator of the time delays using the IS method can be written as follows:

$$\hat{\tau}_i = \frac{1}{R}\sum_{k=1}^{R}\boldsymbol{\tau}_k(i)\frac{L'_{c,\rho_0}(\boldsymbol{\tau}_k)}{g'_{\rho'_1}(\boldsymbol{\tau}_k)}. \quad (24)$$

It is noted that the interval in which the delays are confined depends on the nature of the system. Typically, in network communications, the time delays are confined within the symbol duration, whereas for radar or sonar transmissions, the actual delays introduced by the channel may exceed the symbol duration, but it can be assumed that they do not exceed L, where L is some positive real number. To unify both cases, we suppose that τ_i, for $i = 1,\ 2, \ldots,\ P$ are contained in $[0, L]$. Therefore, it is more convenient to use the circular mean since the parameters are bounded from below and above.

To introduce the concept of circular mean, define a circular random variable X taking values in the finite interval $[0, 1]$ with pdf $G(X)$. The circular mean of X is defined as:

$$E_c\{X\} = \frac{1}{2\pi}\angle\int_0^1\exp\{j2\pi x\}G(x)dx \quad (25)$$

where the operator $\angle(.)$ returns the angle of its complex argument. Suppose that we have R realizations of X, $x_1, \ldots,\ x_R$, then the circular mean in (25) is:

$$E_c\{X\} = \frac{1}{2\pi}\angle\frac{1}{R}\sum_{r=1}^{R}\exp\{j2\pi x_r\}. \quad (26)$$

Now we return to our estimation problem. As already mentioned, we look for the delays $\{\tau_i\}_{i=1}^{P}$ in $[0, L]$. To apply the circular mean, the delays are normalized by L to transpose them to the interval $[0, 1]$. Hence, the IS estimate of τ_i using (24) and (26) is given by:

$$\widehat{\tau}_i = \frac{1}{2\pi L}\angle\frac{1}{R}\sum_{k=1}^{R}F(\boldsymbol{\tau}_k)\exp\left\{j2\pi\frac{\boldsymbol{\tau}_k(i)}{L}\right\}, \quad (27)$$

where $F(\boldsymbol{\tau}_k)$ is the weighting factor defined by:

$$F(\boldsymbol{\tau}_k) = \frac{L'_{c,\rho_0}(\boldsymbol{\tau}_k)}{g'_{\rho'_1}(\boldsymbol{\tau}_k)}. \quad (28)$$

Notice also that we only need the angle of the complex number in the right-hand side of (27). Therefore, the two strictly positive constants $\int L'_{c,\rho_0}(\boldsymbol{x})d\boldsymbol{x}$ and $\int g'_{\rho'_1}(\boldsymbol{x})d\boldsymbol{x}$ used to normalize $L'_{c,\rho_0}(.)$ and $g'_{\rho'_1}(\boldsymbol{x})$, respectively, can be dropped. Moreover, another challenge arises from the computational overflow that may occur when evaluating the exponential terms in both the numerator and the denominator of $F(.)$. To circumvent this problem, we replace $F(.)$ by $F'(.)$ as recently

This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the IEEE Globecom 2011 proceedings.

done in [8][2]:

$$F'(\boldsymbol{\tau}_k) = \exp\left\{\rho_0 L_c(\boldsymbol{\tau}_k) - \rho_1' \sum_{i=1}^{P} I_i(\boldsymbol{\tau}_k(i)) - \max_{1\le l\le R}\left(\rho_0 L_c(\boldsymbol{\tau}_l) - \rho_1' \sum_{i=1}^{P} I_i(\boldsymbol{\tau}_l(i))\right)\right\}. \quad (29)$$

VI. Time Delays Estimation in Passive Systems

Under the assumption of an unknown transmitted signal, only the time difference of arrival (TDOA) can be estimated by using multiple received signals at spatially separated destinations [1]. One received signal at any sensor is designated as a reference and the other replica are compared to. In the sequel, we assume for simplicity the presence of two such sensors. The received signal on each one is represented as follows:

$$y_1(t) = \sum_{i=1}^{P_1} \alpha_{1;i} x(t-\tau_{1;i}) + w_1(t), \quad (30)$$

$$y_2(t) = \sum_{i=1}^{P_2} \alpha_{2;i} x(t-\tau_{2;i}) + w_2(t), \quad (31)$$

where $\{\tau_{n;i}\}_{i=1}^{P_n}$ and $\{\alpha_{n;i}\}_{i=1}^{P_n}$ are the delays and the complex amplitudes of the received signal on the n^{th} sensor and $\{P_n\}_{n=1}^{2}$ is the known number of multipath. To simplify, suppose that $y_1(t)$ has only one signal component ($P_1 = 1$). Then, as in (4), we express the sampled signals (30) and (31) in the frequency domain:

$$Y_1(k) = \alpha_{1;1} X(k) \exp\left\{-\frac{j2\pi k\tau_{1;1}}{N}\right\} + W_1(k), \quad (32)$$

$$Y_2(k) = \sum_{i=1}^{P_2} \alpha_{2;i} X(k) \exp\left\{-\frac{j2\pi k\tau_{2;i}}{N}\right\} + W_2(k), \quad k = 0,\ 1, \ldots,\ N-1, \quad (33)$$

where $\{Y_1(k)\}_{k=0}^{N}$, $\{Y_2(k)\}_{k=0}^{N}$, $\{W_1(k)\}_{k=0}^{N}$ and $\{W_2(k)\}_{k=0}^{N}$ are N samples of the Fourier transform of samples of $y_1(t)$, $y_2(t)$, $w_1(t)$ and $w_2(t)$. As mentioned above, the TDOA will be estimated by taking the received signal in the first sensor as a reference. This amounts to the estimation of the delay differences $\Delta_\tau^{(i)} = \tau_{2;i} - \tau_{1;1}$ for $i = 1,\ 3, \ldots,\ P_2$. Therefore, we manipulate (33) to highlight the desired parameter:

$$Y_2(k) = \sum_{i=1}^{P_2} \beta_i Y_1(k) \exp\left\{-\frac{j2\pi k\Delta_\tau^{(i)}}{N}\right\} + W_p(k), \quad (34)$$

in which

$$\beta_i = \frac{\alpha_{2;i}}{\alpha_{1;1}},\ i = 1,\ 2, \ldots,\ P_2, \quad (35)$$

$$W_p(k) = W_2(k) - \sum_{i=1}^{P_2} \beta_i W_1(k) \exp\left\{-\frac{j2\pi k\Delta_\tau^{(i)}}{N}\right\}. \quad (36)$$

[2]Another obvious solution to avoid occasional overflows is to use smaller values for the parameters ρ_0 and ρ_1', but this will affect the estimation performance.

As observed from the system models in (4) and (34), the formulations are essentially similar. The major difference is in the reference signal ($\boldsymbol{S}$ for the active case and $\boldsymbol{Y}_1$ for the passive case). Gathering all the frequency samples, we obtain the following matrix representation:

$$\boldsymbol{Y}_2 = [Y_2(0),\ Y_2(1), \ldots,\ Y_2(N-1)]^T = \boldsymbol{\Phi}_p(\boldsymbol{\Delta}_\tau)\boldsymbol{\beta} + \boldsymbol{W}_p, \quad (37)$$

where the matrix $\boldsymbol{\Phi}_p(\boldsymbol{\Delta}_\tau)$ is function of the TDOAs defined as:

$$\boldsymbol{\Phi}_p(\boldsymbol{\Delta}_\tau) = \left[\boldsymbol{\phi}_p(\Delta_\tau^{(1)}),\ \boldsymbol{\phi}_p(\Delta_\tau^{(2)}), \ldots,\ \boldsymbol{\phi}_p(\Delta_\tau^{(P_2)})\right], \quad (38)$$

$$\boldsymbol{\phi}_p(\Delta_\tau^{(i)}) = \left[Y_1(0),\ Y_1(1)\exp\left\{-\frac{j2\pi\Delta_\tau^{(i)}}{N}\right\}, \ldots, Y_1(N-1)\exp\left\{-\frac{j2\pi(N-1)\Delta_\tau^{(i)}}{N}\right\}\right]^T, i = 1, \ldots,\ P_2, \quad (39)$$

$$\boldsymbol{\beta} = \left[\frac{\alpha_{2;1}}{\alpha_{1;1}},\ \frac{\alpha_{2;2}}{\alpha_{1;1}}, \ldots,\ \frac{\alpha_{2;P_2}}{\alpha_{1;1}}\right]^T, \quad (40)$$

and

$$\boldsymbol{\Delta}_\tau = \left[\Delta_\tau^{(1)},\ \Delta_\tau^{(2)}, \ldots,\ \Delta_\tau^{(P_2)}\right]^T, \quad (41)$$

is the vector of the TDOAs of interest. Therefore, it turns out that the proposed algorithm developed primarily for active systems remains valid for passive systems as well, by substituting $\boldsymbol{\phi}_a(\tau)$ by $\boldsymbol{\phi}_p(\Delta_\tau)$ and the reference signal $x(t)$ by $y_1(t)$.

VII. Simulation Results

In this section, we present some numerical results to assess the performance of the new time delays estimator. For all simulations, the transmitted pulse is a chirp signal which is widely used in radar and sonar applications. We take $N = 70$ snapshots from the received waveform. We consider 3 propagation paths with delays $[3,\ 6,\ 8]$, normalized by the sampling period. To properly assess the performance of our new IS-based approach, we compare it to the *expectation maximization* (EM) ML algorithm and namely the MUSIC-type method proposed in [1] as an example of non-iterative algorithm. In Fig. 2, we plot the performance of the three methods as a function of the SNR for the three delays (paths). We see that the MUSIC-type technique approaches the CRLB as far as the SNR increases especially for the first path. However, the new IS-based estimator outperforms it over the entire SNR range. This is hardly surprising since the proposed method is an implementation of the ML criterion. Moreover, the MUSIC-type technique fails to estimate the delays of the second and third paths from a very noisy received signal while the newly proposed method exhibits almost the same performance for all the unknown delays. In addition, the IS-based estimator and the EM ML estimator perform almost the same when the later has good initial delays estimates. However, for less accurate initialization, the performance of the EM algorithm deteriorates considerably over the entire SNR range.

Moreover, as it was seen, all the treatments were performed in the frequency domain. Therefore, the signal bandwidth plays an important role in the estimation procedure for both methods. Therefore, in Fig. 3 we assess its impact on the performance of the two estimators and indeed we see that the latter degrades dramatically for small signal bandwidths. Yet, the IS-based estimator still outperforms the Music-type method over the entire considered bandwidth range. Note that the EM ML is less sensitive to bandwidth variations. While

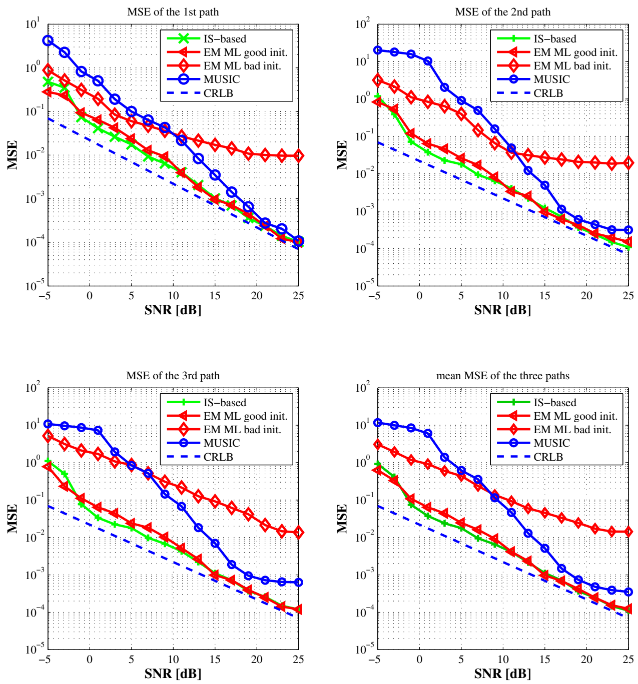

Fig. 2. Estimation performance of the IS-based, EM ML and the MUSIC-type algorithms vs. SNR.

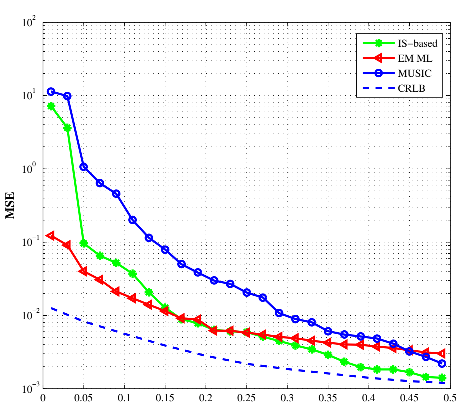

Fig. 3. Estimation performance of the IS-based, EM ML and the MUSIC-type algorithms vs. signal bandwidth at SNR = 10 dB.

the proposed methods is developed under the assumption of constant paths amplitudes, we verify through simulations that it outperforms MUSIC-type methods over time varying channels with relatively low Doppler. Nonetheless, the performance degrades considerably as the Doppler factor increases. In fact, the proposed likelihood function could be considered as a sub-optimal solution for slowly time varying channels since it do not take account mobility. This aspect is currently under investigation to be disclosed in the near future.

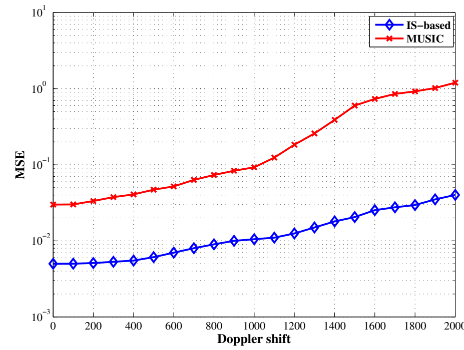

Fig. 4. Estimation performance of the IS-based and the MUSIC-type algorithms vs. Doppler shift at SNR = 10 dB.

VIII. Conclusion

In this paper, we proposed a new implementation of ML-based multiple time delays estimation. We avoided eigen-decomposition and grid search widely used in classical subspace-based and iterative methods. While these traditional methods perform well even for closely separated delays, at high SNRs values, only the proposed IS-based technique produces accurate estimates at low SNR. Moreover, the convergence to the global maximum of the likelihood function is guaranteed by our new estimator, unlike the previous iterative ML methods whose performance depends on the initialization values.

References

[1] F.-X. Ge, D. Shen, Y. Peng and V. O. K. Li, "Super-resolution time delay estimation in multipath environments," *IEEE Trans. Circuits Syst.,* vol. 54, no. 9, pp. 1977-1986, Sep. 2007.

[2] A. H. Quazi, "An overview on the time delay estimation in active and passive systems for target localization," *IEEE Trans. Acoust., Speech, Sign. Process.,* vol. 29, no. 3, pp. 527-533, Jun. 1981.

[3] P.J. Hahn, V.J. Mathews and T.D. Tran, "Adaptive realization of a maximum likelihood time delay estimator" *Conf. on Acoust., Speech, Sign. Process.,* vol. 6, pp. 3121-3124, May 1996.

[4] A. M. Bruckstein, T. J. Shan and T. Kailath, "The resolution of overlapping echoes," *IEEE Trans. Acoust., Speech, Sign. Process.,* vol. 33, no. 6, pp. 1357-1367, Dec. 1985.

[5] Z. Q. Hou and Z. D. Wu, "A new method for high resolution estimation of time delay," *in Proc. ICASSP,* pp. 420-423, May 1982

[6] M. A. Hasan, M. R. Azimi-Sadjadi and G. J. Dobeck, "Separation of multiple time delays using new spectral estimation schemes," *IEEE Trans. Sign. Process.,* vol. 46, no. 6, pp. 1580-1590, Jun. 1998.

[7] M. Pallas and G. Jourdain, "Active high resolution time delay estimation for large BT signals," *IEEE Trans. Sign. Process.,* vol. 39, no. 4, pp. 781-787, Apr. 1991.

[8] S. Saha and S. Kay, "An exact maximum likelihood narrowband direction of arrival estimator," *IEEE Trans. Sign. Process.*, vol. 56, pp. 1082-1092, Oct. 2008.

[9] H. Wang and S. Kay, "A maximum likelihood angle-doppler estimator using importance sampling," *IEEE Trans. Aerospace and Electro. Syst.*, vol. 46, Apr. 2010.

[10] L. Ng and Y. Bar-Shalom, "Multisensor multitarget time delay vector estimation," *IEEE Trans. Acoust., Speech, Sign. Process.,* vol. 34, pp. 669-677, Aug. 1986.

[11] M. Pincus, "A closed form solution for certain programming problems," *Oper. Research*, pp. 690-694, 1962.

[12] W. B. Bishop and P. M. Djuric, "Model order selection of damped sinusoids in noise by predictive densities," *IEEE Trans. Sign. Process.,* vol. 44, pp. 611-619, Mar. 1996.