Mean Likelihood Frequency Estimation

Steven Kay, *Fellow, IEEE,* and Supratim Saha

***Abstract*—Estimation of signals with nonlinear as well as linear parameters in noise is studied. Maximum likelihood estimation has been shown to perform the best among all the methods. In such problems, joint maximum likelihood estimation of the unknown parameters reduces to a separable optimization problem, where first, the nonlinear parameters are estimated via a grid search, and then, the nonlinear parameter estimates are used to estimate the linear parameters. We show that a grid search can be avoided by using the mean likelihood estimator for estimating the unknown nonlinear parameters and how its performance can be made equivalent to that of the maximum likelihood estimator (MLE). The mean likelihood estimator requires computation of a multidimensional integral. However, using the concepts of importance sampling, we obtain the mean likelihood estimate without using integration. The technique is computationally far less burdensome than the direct maximum likelihood method but performs just as well. Simulation examples for estimating frequencies of multiple sinusoids in noise are given. The general technique can be applied to a large class of nonlinear regression problems.**

***Index Terms*—Frequency estimation, Monte Carlo methods, optimization methods, sonar signal analysis.**

I. INTRODUCTION

MANY problems in statistical signal processing may be posed as ones that attempt to estimate signals with linear as well as nonlinear parameters in additive white Gaussian noise. These problems are referred to as nonlinear regression [13]. A common example is the estimation of frequencies of multiple sinusoids in noise. In this problem, the noise-corrupted data are linear with respect to the complex amplitudes of the sinusoids but nonlinear with respect to the frequencies. The problem of estimation of multiple time delays in a multipath environment is another example. The data in this case are linear with respect to the attenuations and nonlinear with respect to the time delays. The time delay estimation problem also reduces to the sinusoidal parameter estimation problem after transforming the data to the Fourier domain, where the attenuations take the role of amplitudes, and the time delays take the role of frequencies. The direction-of-arrival (DOA) estimation problem [17] in array processing also reduces to estimation of multiple frequencies, although in the spatial domain. As a result, the problem of sinusoidal frequency estimation has received a lot of attention in the signal processing community and, thus, will be our main focus in this paper. Our proposed approach, however, *can be applied to many other nonlinear regression problems as well*.

Manuscript received May 21, 1999; revised January 25, 2000. The associate editor coordinating the review of this paper and approving it for publication was Prof. Jian Li.

The authors are with the Department of Electrical and Computer Engineering, University of Rhode Island, Kingston, RI 02881 USA.

Publisher Item Identifier S 1053-587X(00)04940-0.

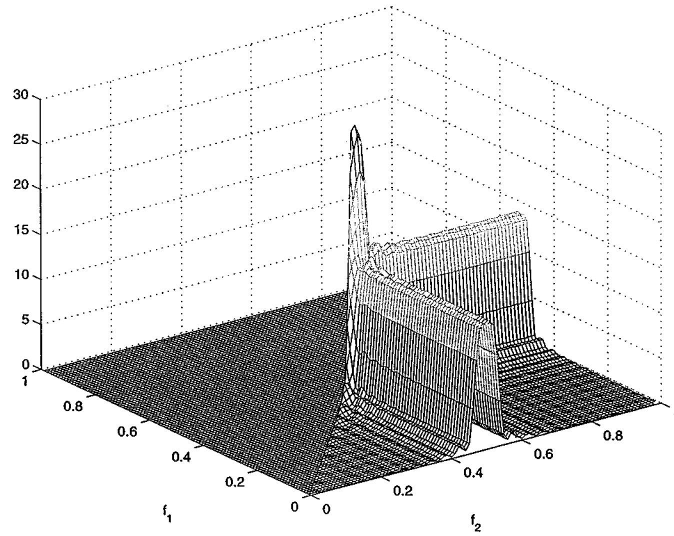


Fig. 1. Plot of $J(\mathbf{f})$ for two sinusoids (noise free case) and $f_2 > f_1$.

Although the maximum likelihood estimator (MLE) is not, in general, the optimal estimator for finite data records, the optimal minimum variance unbiased (MVU) estimator may be analytically difficult to obtain or may not exist at all [9]. Thus, the MLE, which is asymptotically optimal and has been shown to exhibit the best performance for finite data records [8], is the preferred estimator. The exact MLE requires a multidimensional grid search over the possible frequencies since the multidimensional likelihood function is a highly nonlinear function of the frequencies and has many local maxima, even in the absence of noise. The highly multimodal nature of the likelihood function can be seen in Fig. 1 for the noise-free case. The function, whose maximum location is the MLE of the frequencies, is plotted for two closely spaced sinusoids. With an increase in the number of sinusoids, the computational burden of the required grid search increases enormously. Because of the impracticality of the direct MLE, iterative approaches have been used, requiring a good initial guess of the frequencies. Furthermore, due to the highly multimodal nature of the likelihood function, there is no guarantee that an estimate obtained iteratively will be the global maximum. A typical iterative MLE approach is described by [16]. Other techniques that are not based on the MLE, such as subspace methods, have been shown to perform well only at higher SNR's [8]. The MLE, on the other hand, has been shown to perform well even at low signal-to-noise ratios (SNR's) [10]. *Our main goal, therefore, is to develop a noniterative estimator that approximates the MLE but is computationally efficient*.

In this paper, we propose an estimator whose performance is about the same as the MLE but can be implemented with a moderate amount of computation. The technique, called *mean likeli-*

hood estimation (MELE), is based on computation of the mean of the likelihood function [1]. It considers the *normalized likelihood function* as a probability density function. Even though the frequencies are not assumed to be random variables, an estimate can be obtained as the mean value of the normalized likelihood function. This is because the likelihood function, when properly normalized, possesses all the properties of a joint PDF in the unknown parameters. Since the frequencies are not random, we say that the normalized likelihood function is a *pseudo-PDF* of the unknown frequencies.

It should be emphasized that we are considering the frequencies to be unknown deterministic constants as opposed to realizations of random variables. The latter formalism is the basic assumption of the Bayesian philosophy. The problem could have been approached by assuming the frequencies are random and using a uniform prior (i.e., a noninformative prior) for the frequencies and a suitable noninformative prior for the in-phase and quadrature amplitude components [6]. This approach would then produce the minimum mean squared error estimator, where the mean square error criterion is defined in the Bayesian sense [2], i.e., when averaged over the prior PDF of the amplitude components and frequencies. Our approach, however, is a classical one in which we attempt to implement an estimator that approximates the MLE. No prior knowledge of the amplitude components and frequencies is assumed other than that required for identifiability. Note, however, that the approach described herein applies equally well under the Bayesian assumptions. In this case, the pseudo-PDF can be interpreted as a posterior PDF and the MELE as the posterior mean.

To implement the MELE requires a multidimensional integration, which at first appears impractical, but such types of integrals can be well approximated by Monte Carlo techniques [14], [15]. In particular, importance sampling has been shown to be a very powerful Monte Carlo technique, allowing multidimensional integrals to be evaluated efficiently. We will use the importance sampling approach to obtain the mean likelihood estimates. The proposed method will be shown via computer simulation to perform about the same as the MLE as previously mentioned. The MLE is recognized as the most accurate estimator for finite data records and, furthermore, can be shown to asymptotically achieve the Cramér-Rao lower bound (as the data record length becomes large and/or the signal-to-noise ratio becomes large [9]).

The paper is organized as follows. In Section II, we discuss the general nonlinear regression problem and the estimation of the nonlinear parameters using mean likelihood. In Section III, the use of importance sampling to efficiently obtain the mean likelihood estimate is described. In Section IV, the complete implementation details of the sinusoidal frequency estimator based on MELE and importance sampling is discussed. Section V contains some simulation results and also the details of the computational complexity of the proposed method. Finally, in Section VI, we give conclusions and future directions.

II. Nonlinear Regression

We are interested in the estimation of parameters of signals that have linear as well as nonlinear parameters in additive white Gaussian noise. The complex data model for these signals can be expressed in the form

$$\mathbf{x} = \mathbf{H}(\boldsymbol{\alpha})\boldsymbol{\theta} + \mathbf{w} \tag{1}$$

where

- $\mathbf{x}$ — $N \times 1$ noise-corrupted data vector;
- $\boldsymbol{\theta}$ — $q \times 1$ linear parameter vector;
- $\mathbf{H}(\boldsymbol{\alpha})$ — $N \times q$ matrix, which is dependent on the nonlinear signal parameters or the $p \times 1$ parameter vector $\boldsymbol{\alpha}$.

The data vector $\mathbf{x}$ is linearly related to the parameter vector $\boldsymbol{\theta}$ but depends nonlinearly on $\boldsymbol{\alpha}$. The noise vector $\mathbf{w}$ is assumed to consist of samples of complex white Gaussian noise and is of dimension $N \times 1$. Exponentials in noise, sinusoids in noise, FM signals in noise, etc. [9] can be modeled in the form given by (1). We now show how the sinusoidal parameter estimation problem can be expressed in the above form. Consider a signal consisting of p complex sinusoids embedded in complex white Gaussian noise. If $x[n]$ denotes the received data, then

$$x[n] = \sum_{i=1}^{p} A_i \exp[j(2\pi f_i n + \phi_i)] + w[n] \tag{2}$$

for $n = 0, \cdots, N-1$, and where $0 \le f_1 < f_2 < \cdots < f_p < 1$, and $w[n]$ is complex white Gaussian noise with variance σ^2. Note that we have assumed that the frequencies are ordered since without this assumption, they are not identifiable. That is to say, any reordering of the frequencies in (2) will produce the same data values and, hence, even in the absence of noise the frequencies cannot be uniquely determined. Now, the data model of (2) can be expressed in the form of (1), with $\boldsymbol{\theta}$ being the complex amplitudes or

$$\boldsymbol{\theta} = [A_1 \exp(j\phi_1) \quad A_2 \exp(j\phi_2) \quad \cdots \quad A_p \exp(j\phi_p)]^T \tag{3}$$

and $\boldsymbol{\alpha}$ being the frequencies so that

$$\mathbf{H}(\boldsymbol{\alpha}) = \mathbf{H}(\mathbf{f}) = [\mathbf{e}(f_1) \quad \mathbf{e}(f_2) \quad \cdots \quad \mathbf{e}(f_p)] \tag{4}$$

where

$$\begin{aligned}\mathbf{e}(f_i) = &[\exp(j2\pi f_i(0)) \quad \exp(j2\pi f_i(1)) \\ &\cdots \quad \exp(j2\pi f_i(N-1))]^T\end{aligned} \tag{5}$$

and $\boldsymbol{\alpha} = \mathbf{f} = [f_1 \cdots f_p]^T$. It is clear from (2) that the signal is nonlinear with respect to the frequencies f_i's and linear with respect to the complex amplitudes $A_i \exp(j\phi_i)$'s. For data models expressible in the form of (1), the joint MLE of $\boldsymbol{\theta}$ and $\boldsymbol{\alpha}$ reduces to a separable optimization problem so that $\hat{\boldsymbol{\theta}}_{\text{mle}}$ becomes a function of $\hat{\boldsymbol{\alpha}}_{\text{mle}}$, the latter being determined first.

The general likelihood function for the data model of (1) can be expressed as

$$\begin{aligned}L(\boldsymbol{\alpha}, \boldsymbol{\theta}) &\propto p(\mathbf{x}; \boldsymbol{\alpha}, \boldsymbol{\theta}) \\ &= \frac{1}{\pi^N \sigma^{2N}} \exp[-(1/\sigma^2)(\mathbf{x} - \mathbf{H}(\boldsymbol{\alpha})\boldsymbol{\theta})^H(\mathbf{x} - \mathbf{H}(\boldsymbol{\alpha})\boldsymbol{\theta})]\end{aligned} \tag{6}$$

where $p(\mathbf{x};\boldsymbol{\alpha},\boldsymbol{\theta})$ is the probability density function of $\mathbf{x}$ with parameters $\boldsymbol{\alpha}$ and $\boldsymbol{\theta}$. The MLE of $\boldsymbol{\theta}$ is given by [9]

$$\hat{\boldsymbol{\theta}} = (\mathbf{H}^H(\boldsymbol{\alpha})\mathbf{H}(\boldsymbol{\alpha}))^{-1}\mathbf{H}^H(\boldsymbol{\alpha})\mathbf{x} \tag{7}$$

when $\boldsymbol{\alpha}$ is replaced by its MLE. The compressed likelihood function is defined as

$$L(\boldsymbol{\alpha},\hat{\boldsymbol{\theta}}) = L'(\boldsymbol{\alpha}) \propto p(\mathbf{x};\boldsymbol{\alpha},\hat{\boldsymbol{\theta}}). \tag{8}$$

From (6) and (7), we have

$$\begin{aligned} p(\mathbf{x};\boldsymbol{\alpha},\hat{\boldsymbol{\theta}}) &= \frac{1}{\pi^N\sigma^{2N}} \\ &\quad\cdot \exp[-(1/\sigma^2)(\mathbf{x}-\mathbf{H}(\boldsymbol{\alpha})\hat{\boldsymbol{\theta}})^H(\mathbf{x}-\mathbf{H}(\boldsymbol{\alpha})\hat{\boldsymbol{\theta}})] \quad (9)\\ &= \frac{1}{\pi^N\sigma^{2N}}\exp[-(1/\sigma^2)\mathbf{x}^H \\ &\quad\cdot(\mathbf{I}-\mathbf{H}(\boldsymbol{\alpha})(\mathbf{H}^H(\boldsymbol{\alpha})\mathbf{H}(\boldsymbol{\alpha}))^{-1}\mathbf{H}^H(\boldsymbol{\alpha}))\mathbf{x}]. \quad (10)\end{aligned}$$

By omitting the terms not dependent on $\boldsymbol{\alpha}$, the compressed likelihood function becomes

$$L'(\boldsymbol{\alpha}) = \exp[(1/\sigma^2)\mathbf{x}^H\mathbf{H}(\boldsymbol{\alpha})(\mathbf{H}^H(\boldsymbol{\alpha})\mathbf{H}(\boldsymbol{\alpha}))^{-1}\mathbf{H}^H(\boldsymbol{\alpha})\mathbf{x}]. \tag{11}$$

The *normalized* compressed likelihood function is defined as

$$\overline{L}(\boldsymbol{\alpha}) = \frac{L'(\boldsymbol{\alpha})}{\int L'(\boldsymbol{\alpha})\mathbf{d}\boldsymbol{\alpha}}. \tag{12}$$

This is a function of the parameter vector $\boldsymbol{\alpha}$ and has all the properties of a PDF, although strictly speaking, it is not a PDF since $\boldsymbol{\alpha}$ is deterministic. We term $\overline{L}(\boldsymbol{\alpha})$ a pseudo-PDF in $\boldsymbol{\alpha}$. With this definition, we can define the *mean likelihood estimate* of the parameter vector $\boldsymbol{\alpha}$ as

$$\hat{\alpha}_{i,\text{mele}} = \int\cdots\int \alpha_i\overline{L}(\boldsymbol{\alpha})\,d\boldsymbol{\alpha} \qquad i=1,2,\cdots,p. \tag{13}$$

As can be seen from the above expression, the mean likelihood estimate $\hat{\alpha}_{i,\text{mele}}$ requires an evaluation of a p-dimensional integral (assuming $\boldsymbol{\alpha}$ is real), which is difficult to implement in practice. However, since $\overline{L}(\boldsymbol{\alpha})$ is a pseudo-PDF, we can interpret $\hat{\alpha}_{i,\text{mele}}$ as the expected value of α_i. It has been shown that for this type of problem, Monte Carlo approximation techniques can achieve good results without using direct integration [3]. A straightforward Monte Carlo approximation to the MELE, i.e., the mean of $\overline{L}(\boldsymbol{\alpha})$, is

$$\hat{\boldsymbol{\alpha}}_{\text{mele}} = \frac{1}{M}\sum_{k=1}^{M}\boldsymbol{\alpha}_k \tag{14}$$

where $\boldsymbol{\alpha}_k$ is the kth realization of the vector $\boldsymbol{\alpha}$ distributed according to $\overline{L}(\boldsymbol{\alpha})$ or $\boldsymbol{\alpha}\sim\overline{L}(\boldsymbol{\alpha})$. Computing $\hat{\boldsymbol{\alpha}}_{\text{mele}}$ by (14) requires the generation of $\boldsymbol{\alpha}\sim\overline{L}(\boldsymbol{\alpha})$. For the problem of interest in this paper, generation of the vector $\boldsymbol{\alpha}\sim\overline{L}(\boldsymbol{\alpha})$ is difficult, as $\overline{L}(\boldsymbol{\alpha})$ is a highly nonlinear function of $\boldsymbol{\alpha}$. Therefore, even though direct integration can be avoided by using (14), the generation of $\boldsymbol{\alpha}\sim\overline{L}(\boldsymbol{\alpha})$ may again demand integration. As a result, we do not use (14) to determine $\hat{\boldsymbol{\alpha}}_{\text{mele}}$. Rather, we use importance sampling [14], as described in the next section.

III. Importance Sampling

To compute a multidimensional integral of the type given in (13), importance sampling has been shown to be a powerful tool [4]. The approach is based on the observation that integrals such as $\int h(\boldsymbol{\alpha})\overline{L}(\boldsymbol{\alpha})\,\mathbf{d}\boldsymbol{\alpha}$ can be expressed as

$$\int h(\boldsymbol{\alpha})\overline{L}(\boldsymbol{\alpha})\,\mathbf{d}\boldsymbol{\alpha} = \int h(\boldsymbol{\alpha})\frac{\overline{L}(\boldsymbol{\alpha})}{\overline{g}(\boldsymbol{\alpha})}\overline{g}(\boldsymbol{\alpha})\,d\boldsymbol{\alpha} \tag{15}$$

where $\overline{g}(\boldsymbol{\alpha})$ is a pseudo-PDF, and we assume that $\overline{g}(\boldsymbol{\alpha}) > 0$. Note that we use the symbolism $\overline{g}(\boldsymbol{\alpha})$ to denote the normalized version of $g(\boldsymbol{\alpha})$, which will be used later. Then, the right-hand side of (15) can be interpreted as the expected value of $h(\boldsymbol{\alpha})(\overline{L}(\boldsymbol{\alpha})/\overline{g}(\boldsymbol{\alpha}))$, with respect to the pseudo-PDF $\overline{g}(\boldsymbol{\alpha})$. The function $\overline{g}(\boldsymbol{\alpha})$ is called the normalized importance function. Unlike $\overline{L}(\boldsymbol{\alpha})$, which, in general, is a complicated function of $\boldsymbol{\alpha}$, $\overline{g}(\boldsymbol{\alpha})$ can be chosen to be some simple function of $\boldsymbol{\alpha}$ so that realizations of $\boldsymbol{\alpha}$ can be easily generated. Then, the approximate value of the integral in (15) can be found by the Monte Carlo estimate

$$\frac{1}{M}\sum_{k=1}^{M}h(\boldsymbol{\alpha}_k)\frac{\overline{L}(\boldsymbol{\alpha}_k)}{\overline{g}(\boldsymbol{\alpha}_k)} \tag{16}$$

where $\boldsymbol{\alpha}_k$ is the kth realization of the vector $\boldsymbol{\alpha}$ generated according to the pseudo-PDF $\overline{g}(\boldsymbol{\alpha})$. The value of M needed for a good approximation depends on the choice of g. Typically, $\overline{g}(\boldsymbol{\alpha})$ should be chosen similar to $\overline{L}(\boldsymbol{\alpha})$, as this reduces the variance of the estimate given by (16). However, another important point to keep in mind when choosing $\overline{g}(\boldsymbol{\alpha})$ is that it should be simple enough so that $\boldsymbol{\alpha}\sim\overline{g}(\boldsymbol{\alpha})$ can be easily generated. We explain in the next section how to choose $\overline{g}(\boldsymbol{\alpha})$ for the sinusoidal parameter estimation problem described in Section II.

IV. Sinusoidal Frequency Estimation

For the sinusoidal parameter estimation problem, the compressed likelihood function was shown to be [see (11)]

$$L'(\mathbf{f}) = \exp\left[\frac{1}{\sigma^2}\mathbf{x}^H\mathbf{H}(\mathbf{H}^H\mathbf{H})^{-1}\mathbf{H}^H\mathbf{x}\right] \tag{17}$$

where $\mathbf{H}$ now depends on $\boldsymbol{\alpha}=\mathbf{f}$, per (4), and we have omitted the explicit dependence of $\mathbf{H}$ on $\mathbf{f}$. The MELE of the frequencies is computed using the importance sampling approach described previously. Due to the fact that the frequencies have the properties of a *circular* random variable [12], the mean likelihood estimate of f_i is obtained using the circular mean definition or

$$\hat{f}_{i,\text{mele}} = \frac{1}{2\pi}\angle\int\cdots\int\exp(j2\pi f_i)\overline{L}(\mathbf{f})\,\mathbf{df}.$$

This amounts to computing the angle of the mean likelihood estimate of $\exp(j2\pi f_i)$. Note that if the mean were to be evaluated directly as

$$\hat{f}_{i,\text{mele}} = \int\cdots\int f_i\overline{L}(\mathbf{f})\mathbf{df} \tag{18}$$

then the estimates obtained would be biased [7], especially at low SNR's and/or for short data records. The key idea in defining a circular mean is to average position vectors. Hence, if $\theta_1, \theta_2, \cdots, \theta_M$ are realizations of a random point $\exp(j\theta)$ on the circumference of a circle of unit radius, then the sample mean of the data is defined as [12]

$$\overline{\theta} = \angle \frac{1}{M} \sum_{k=1}^{M} \exp(j\theta_k). \tag{19}$$

The use of (19) alleviates the estimator bias. The difficulty of using the linear mean had also been remarked upon by [11]. Thus, $\hat{f}_{i,\text{mele}}$ is defined from (16) and (19) as

$$\hat{f}_{i,\text{mele}} = \frac{1}{2\pi} \angle \frac{1}{M} \sum_{k=1}^{M} \frac{\overline{L}(\mathbf{f}_k)}{\overline{g}(\mathbf{f}_k)} \exp(j2\pi[\mathbf{f}_k]_i) \tag{20}$$

for $i = 1, \cdots, p$, where $\mathbf{f}_k$ is the kth realization of the frequency vector. Note that since we need only find the angle of the complex quantity in (20), an equivalent estimator is

$$\hat{f}_{i,\text{mele}} = \frac{1}{2\pi} \angle \frac{1}{M} \sum_{k=1}^{M} \frac{L'(\mathbf{f}_k)}{g(\mathbf{f}_k)} \exp(j2\pi[\mathbf{f}_k]_i) \tag{21}$$

or finally

$$\hat{f}_{i,\text{mele}} = \frac{1}{2\pi} \angle \frac{1}{M} \sum_{k=1}^{M} w(\mathbf{f}_k) \exp(j2\pi[\mathbf{f}_k]_i) \tag{22}$$

where

$$w(\mathbf{f}) = \frac{L'(\mathbf{f})}{g(\mathbf{f})}.$$

This observation is quite important in that it simplifies the computation greatly. We no longer need to find the normalization constants $\int L'(\mathbf{f})\, d\mathbf{f}$ and $\int g(\mathbf{f})\, d\mathbf{f}$ in computing $\overline{L}(\mathbf{f})$ and $\overline{g}(\mathbf{f})$.

Having expressed the mean likelihood estimate of the frequencies in (22), we need to choose an appropriate importance function $g(\mathbf{f})$ that will allow $\mathbf{f}$ to be generated easily. From (17), we observe that if $(\mathbf{H}^H\mathbf{H})^{-1}$ is replaced by $\mathbf{I}/N$, where $\mathbf{I}$ is the $p \times p$ identity matrix, then $L'(\mathbf{f})$ becomes separable in the f_i's. Hence, with this choice, the joint pseudo-PDF can be written as a product of the marginals. Generation of realizations of $\mathbf{f}$ then reduces to the generation of p *independent* realizations of f. Note that such a choice is a good approximation for well-separated sinusoids. Thus, we let

$$g(\mathbf{f}) = \exp[(1/N\sigma^2)\mathbf{x}^H\mathbf{H}\mathbf{H}^H\mathbf{x}] = \exp\left[1/\sigma^2 \sum_{i=1}^{p} I(f_i)\right] \tag{23}$$

where $I(f_i)$ is the periodogram of the data evaluated at the frequency f_i and is given by

$$I(f_i) = \frac{1}{N}\left|\sum_{n=0}^{N-1} x[n]\exp(-j2\pi f_i n)\right|^2. \tag{24}$$

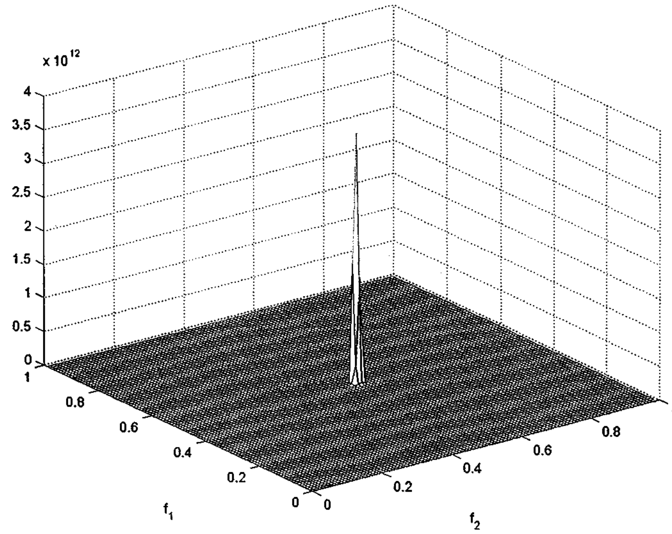


Fig. 2. Plot of modified $J(\mathbf{f})$ for two sinusoids (noise-free case).

Now, $g(\mathbf{f})$ can be expressed as (retaining g for the function of the scalar f)

$$g(\mathbf{f}) = \prod_{i=1}^{p} g(f_i) \tag{25}$$

where $g(f_i) = \exp(I(f_i)/\sigma^2)$.

Before proceeding further, recall that the main reason for using the MELE is to make the method perform similar to the MLE. The MLE is the location of the global maximum of the function $J(\mathbf{f}) = \mathbf{x}^H\mathbf{H}(\mathbf{H}^H\mathbf{H})^{-1}\mathbf{H}^H\mathbf{x}$. However, the function $J(\mathbf{f})$ has several local maxima, even in the absence of noise. This can be seen from Fig. 1, in which we have plotted the function $J(\mathbf{f})$ for the noise-free two-sinusoid case. Fig. 2 shows the plot of $\exp(\rho J(\mathbf{f}))$, where ρ is a factor used to make the function more peaked. In Fig. 2, we have set $\rho = 1$. The reason for making the function more peaked is that the global maximum will then have a relatively higher peak as compared with the local maxima points. If such a modified function is used, the mean likelihood estimate will be nearly equal to the maximum likelihood estimate. In fact, it can be shown that as $\rho \to \infty$, the mean likelihood estimate is equal to the maximum likelihood estimate, assuming that it is unique (see the proof in the Appendix). Therefore, instead of using the likelihood function of (17), we use the *modified* likelihood function defined as

$$L_{\text{mod}}(\mathbf{f}) = \exp[\rho\mathbf{x}^H\mathbf{H}(\mathbf{H}^H\mathbf{H})^{-1}\mathbf{H}^H\mathbf{x}] \tag{26}$$

and choose ρ for the best performance. Note that the only difference from the actual likelihood function is that instead of scaling by $1/\sigma^2$, which is dependent on the SNR and which is usually unknown, we use a scaling equal to ρ, which we choose independently of the SNR. In a similar way, we define the modified importance function as

$$g_{\text{mod}}(\mathbf{f}) = \exp[\rho\mathbf{x}^H\mathbf{H}\mathbf{H}^H\mathbf{x}/N] = \prod_{i=1}^{p} \exp(\rho I(f_i)) \tag{27}$$

so that we finally have from (26) and (27)

$$w(\mathbf{f}) = \frac{\exp[\rho \mathbf{x}^H \mathbf{H}(\mathbf{H}^H\mathbf{H})^{-1}\mathbf{H}^H\mathbf{x}]}{\prod_{i=1}^{p} \exp(\rho I(f_i))} \tag{28}$$

which is used in (22).

To obtain good estimation performance, it is important to choose ρ appropriately. Too small a value will result in a broad modified likelihood function, and the mean will not necessarily be the same as the location of the maximum. Too large a value will lead to numerical difficulties. A typical value and the one used for the two-sinusoid case described in Section V is $\rho = 1$. A further discussion of the choice of ρ is contained in that section.

A. Generation of $\mathbf{f}$

Due to separability of $g_{\text{mod}}(\mathbf{f})$ in the f_i's, as seen in (27), the frequencies can be considered independent. This makes generation of the f_i's quite simple. The only constraint on the f_i's is that there should be some minimum separation between any two of them. Such a constraint is necessary because in generating a frequency vector, two of the frequencies may turn out to be nearly the same. This violates the implicit assumption that the frequencies are distinct and, hence, identifiable. If this occurs, the matrix $\mathbf{H}^H\mathbf{H}$ will be singular, or alternatively, the covariance matrix of the complex sinusoidal amplitudes will not exist. The latter is, of course, just another indication that the frequencies are not identifiable. In the direct MLE, as well as our technique, the condition of distinct frequencies and, hence, a full-rank $\mathbf{H}$ is required in order to determine the frequencies. We have implicitly assumed this in ordering the frequencies of the model as $f_1 < f_2 < \cdots < f_p$.

Next, we show how to generate a realization of the vector $\mathbf{f}$. We first consider the case when $p = 2$. For f_1, we generate $u_1 \sim U[0,1]$ and then use $f_1 = G^{-1}(u_1)$, where $G(f)$ is the cumulative distribution function of f defined as [see (27)]

$$G(f) = \int_0^f \overline{g}_{\text{mod}}(u)\, du = \int_0^f \frac{\exp(\rho I(u))}{\int_0^1 \exp(\rho I(\xi))\, d\xi}\, du.$$

This is a standard method for generating a sample distributed according to a given PDF. However, due to the steep slope of $G(f)$, the direct method of finding the frequency sample by using $f_1 = G^{-1}(u_1)$ would require a fine search to obtain f_1 as $\arg\min_f |u_1 - G(f)|$. This would make the process of generating f_1 computationally intensive. Note, however, that the function $S(f) = |u_1 - G(f)|$ is unimodal because u_1 is fixed for a given realization, and $G(f)$ is a cumulative distribution function, which is an increasing function of f. This property of $S(f)$ allows us to use a golden search [18] to find the location of the minimum of $S(f)$. The golden search is known to converge after a small number of iterations and requires only one function evaluation per iteration.

Once f_1 is generated, f_2 is generated such that $|f_2 - f_1| > \delta$, where δ is the minimum allowable frequency separation. Generation of f_2 can be understood by referring to Fig. 3(a), where a sketch of $G(f)$—the cumulative distribution function of $\overline{g}_{\text{mod}}(f)$—is shown for two closely spaced sinusoids. As f_2 should satisfy the minimum frequency separation with respect to f_1, instead of generating f_2 in the same way as f_1, we first generate $u_2 \sim U[(0, G(f_1 - \delta)) \cup (G(f_1 + \delta), 1)]$. Then, from this u_2, we use the golden search to find f_2, which is given by $f_2 = \arg\min_f |u_2 - G(f)|$. This guarantees that $|f_2 - f_1| > \delta$.

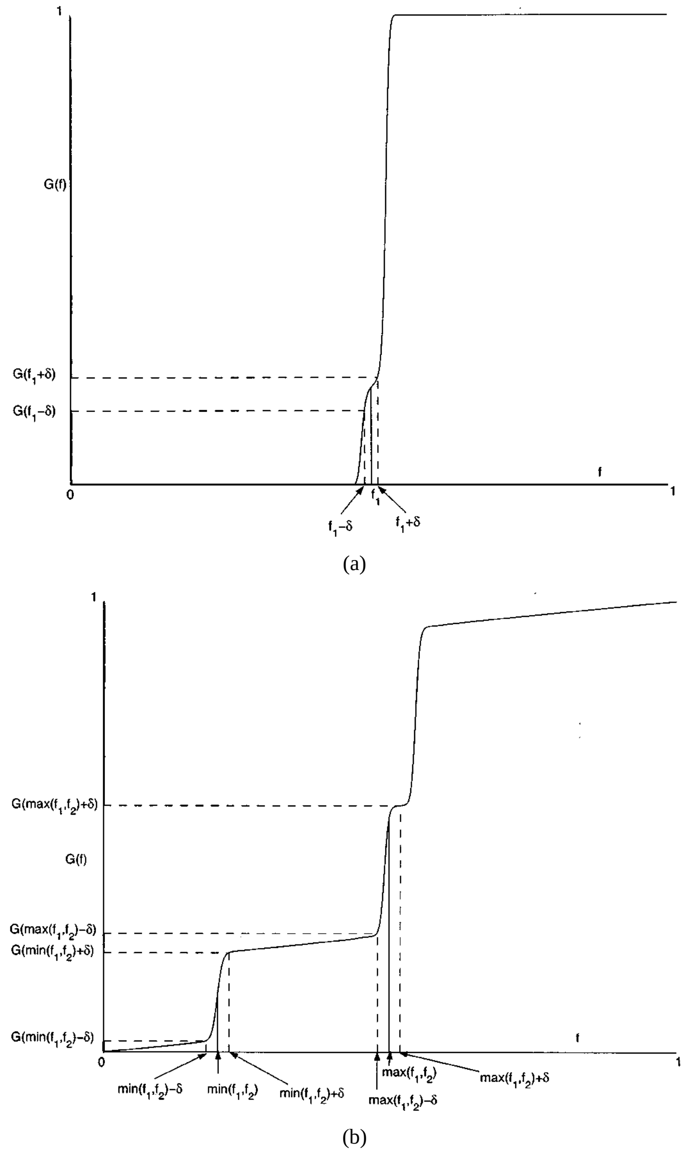


Fig. 3. (a) Generation of frequency f_2 for two sinusoids. (Generate $u_2 \sim U[(0, G(f_1-\delta)) \cup (G(f_1+\delta), 1)]$, where $f_1 = G^{-1}(u_1)$). (b) Generaton of frequency f_3 for three sinusoids. (Generate $u_3 \sim U[(0, G(\min(f_1,f_2)-\delta)) \cup (G(\min(f_1,f_2)+\delta), G(\max(f_1,f_2)-\delta)) \cup (G(\max(f_1,f_2)+\delta), 1)]$, after f_1 and f_2 have been generated with the condition $|f_1 - f_2| > \delta$).

When $p = 3$, we proceed as before to obtain the first two frequency samples. Then, the third frequency sample f_3 needs to be generated subject to the constraints $|f_3 - f_2| > \delta$ and $|f_3 - f_1| > \delta$. Now, we generate u_3, which is distributed uniformly in the union of three intervals. This satisfies the minimum frequency separation requirement with respect to f_1 and f_2. Fig. 3(b) illustrates this procedure. Then,

$u_3 \sim U[(0, G(\min(f_1, f_2) - \delta)) \cup (G(\min(f_1, f_2) + \delta), G(\max(f_1, f_2) - \delta)) \cup (G(\max(f_1, f_2) + \delta), 1)]$, if $\max(f_1, f_2) - \delta > \min(f_1, f_2) + \delta$, or $u_3 \sim U[(0, G(\min(f_1, f_2) - \delta)) \cup (0, G(\min(f_1, f_2) + \delta), 1)]$, if $\max(f_1, f_2) - \delta < \min(f_1, f_2) + \delta$. The latter will occur rarely if δ is very small but if it does, u_3 will be uniformly distributed in the union of two intervals instead of three. Then, find $f_3 = G^{-1}(u_3)$ by the golden search.

For more than three sinusoids, the process is repeated. The frequency realizations are grouped in ascending order to obtain the frequency vector sample $\mathbf{f}_k$. The process is repeated M times to yield M realizations of the frequency vector $\mathbf{f}_k$.

B. Algorithm Summary for $p = 3$

1) Find the periodogram $I(f)$ of the data $x[n]$, which is given by

$$I(f) = \frac{1}{N} \left| \sum_{n=0}^{N-1} x[n] \exp(-j2\pi f n) \right|^2.$$

2) Compute the normalized importance function given by

$$\bar{g}_{\text{mod}}(f) = \frac{\exp(\rho I(f))}{\displaystyle\int_0^1 \exp(\rho I(\xi))\, d\xi} \qquad 0 \le f < 1.$$

3) Determine the cumulative distribution function of $\bar{g}_{\text{mod}}(f)$ as

$$G(f) = \int_0^f \bar{g}_{\text{mod}}(u)\, du \qquad 0 \le f < 1.$$

4) Generate a realization of the frequency vector $\mathbf{f}$. To do so, first generate f_1 using $u_1 \sim U[0, 1]$ and a golden search, as described in Section IV-A, to find $f_1 = G^{-1}(u_1)$. Once f_1 is generated, put a band around it as $(f_1 - \delta, f_1 + \delta)$. Choose δ as 0.001 for good performance. More generally, δ should be chosen to be less than one tenth of the minimum frequency separation expected. Now, to guarantee that f_2 is separated from f_1 by at least δ, generate $u_2 \sim U[(0, G(f_1 - \delta)) \bigcup (G(f_1 + \delta), 1)]$, and again, use a golden search to obtain $f_2 = G^{-1}(u_2)$. In addition, put a band around f_2 as $(f_2 - \delta, f_2 + \delta)$, and generate u_3 uniformly distributed in the union of three intervals, as described in Section IV-A, and obtain f_3 from it. Now, define the frequency vector sample $\mathbf{f}_k$ by arranging f_1, f_2, f_3 in ascending order. Repeat the overall process to generate M realizations of $\mathbf{f}$ or $\mathbf{f}_k$ for $k = 1, 2, \cdots, M$.

5) Compute the mean likelihood estimate of the frequencies f_i for $i = 1, 2, 3$ using

$$\hat{f}_{i,\text{mele}} = \frac{1}{2\pi} \angle \frac{1}{M} \sum_{k=1}^{M} w(\mathbf{f}_k) \exp(j2\pi[\mathbf{f}_k]_i) \tag{29}$$

where

$$w(\mathbf{f}) = \frac{\exp[\rho \mathbf{x}^H \mathbf{H}(\mathbf{H}^H \mathbf{H})^{-1} \mathbf{H}^H \mathbf{x}]}{\displaystyle\prod_{i=1}^{3} \exp[\rho I(f_i)]} \tag{30}$$

and $\mathbf{H} = [\mathbf{e}(f_1)\mathbf{e}(f_2)\mathbf{e}(f_3)]^T$.

V. Computer Simulation Results

We consider the cases of two and three sinusoids. If the two sinusoids are widely spaced in frequency, the periodogram peak locations will indicate the frequencies correctly. However, if the two frequencies are closely spaced, then the two sinusoids interact with one another, and the periodogram peak locations are biased estimates of the frequencies. The following example considers the case of closely spaced sinusoids for which the periodogram peak locations are poor estimates of the frequencies.

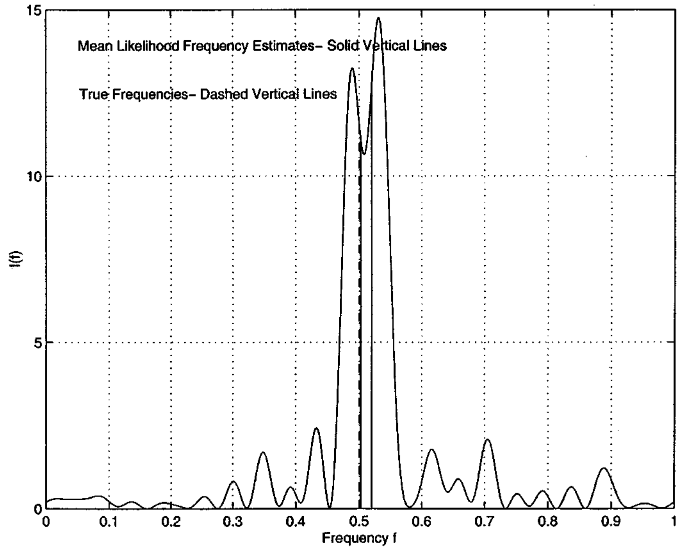
Example 1—Two Sinusoids: Two equiamplitude sinusoids in additive complex white Gaussian noise are considered. We have $N = 25$ data points of $x[n] = \exp(j2\pi f_1 n) + \exp(j(2\pi f_2 n + \pi/4)) + w[n]$, where $f_1 = 0.5$ and $f_2 = 0.52$. This example is a standard one and has been used extensively in [8]. The variance of $w[n]$ is chosen to result in a given SNR, which is defined as $10 \log_{10} 1/\sigma^2$ dB. As an illustration, a typical periodogram of the data is plotted in Fig. 4(a) for an SNR of 5 dB. It is clear from the plot that the peak locations of the periodogram fail to identify the true frequencies of the two sinusoids for this example. The MELE is able to accurately estimate the frequencies, as seen in the figure (as given by the solid lines, and note that at $f = 0.52$, the line indicating the true frequency and the one indicating the MELE coalesce). In Fig. 4(b), we have plotted the cumulative distribution function

$$G(f) = \int_0^f \frac{\exp(\rho I(u))}{\displaystyle\int_0^1 \exp(\rho I(\xi))\, d\xi}\, du$$

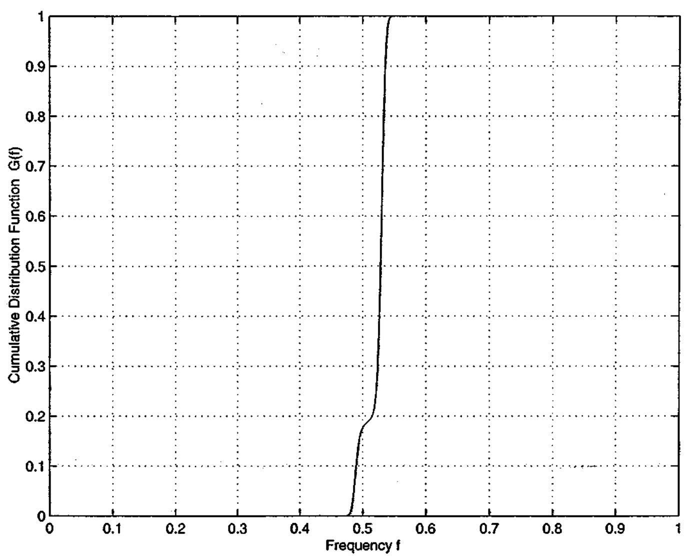
where ρ was chosen equal to 1 for this example. As can be observed, there is a very sharp transition from 0 to 1 in the vicinity of the actual frequencies (i.e., the region around 0.5). Using this cumulative distribution function, we generated $M = 2000$ realizations of the frequency vector using the technique described in Section IV-A. Then, the frequency estimate is obtained from the 2000 realizations of $\mathbf{f}$ using (29). The entire process is repeated for the same SNR with independent noise samples to obtain the estimates for different trials, and then, the mean square error for the given SNR is calculated. The number of Monte Carlo trials required was a maximum of 750, with more trials required for lower SNR's. In Fig. 5, we plot $10 \log_{10}$ (1/mean square error) versus the SNR. The mean square error was determined from the Monte Carlo trials as

$$\text{mean square error} = \frac{1}{M} \sum_{k=1}^{M} (\hat{f}_{i,\text{mele}}^{(k)} - f_i)^2$$

where $\hat{f}_{i,\text{mele}}^{(k)}$ is the estimate of the ith frequency from the kth Monte Carlo trial [see (29)]. The performance of the MELE is shown, as is the MLE. To obtain the MLE performance, a fine grid search was conducted of the function $J(\mathbf{f}) = \mathbf{x}^H \mathbf{H}(\mathbf{H}^H \mathbf{H})^{-1} \mathbf{H}^H \mathbf{x}$. From this plot, it can be concluded that the proposed method performs quite well in that it achieves the Cramér-Rao lower bound (CRLB) and does not have the drawback of poor performance at low SNR's, which is typical of the subspace-based methods [8]. In addition, it requires generation of only 2000 frequency realizations to achieve the CRLB. It should be noted that both the MLE and the proposed method have a threshold SNR of about 1 dB.



(a)



(b)

Fig. 4. (a) Periodogram of the data for two sinusoids in additive white Gaussian noise SNR $= 5$ dB. (b) Plot of the cumulative distribution function $G(f)$ for the two sinusoids case SNR $= 5$ dB.

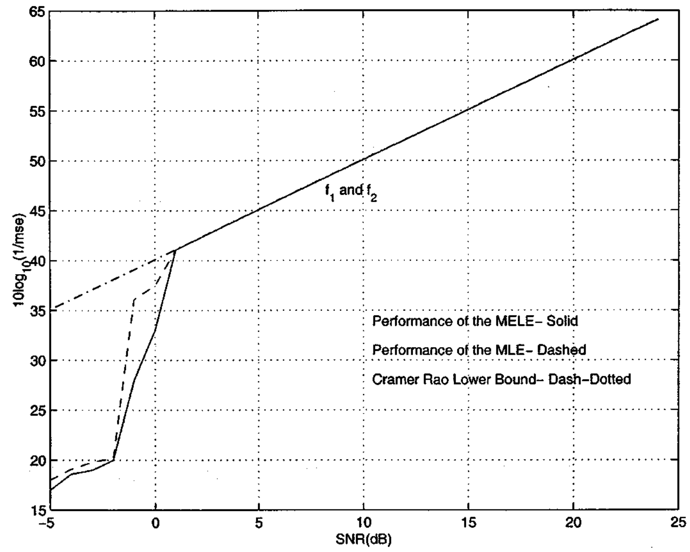


Fig. 5. Performance of the mean likelihood estimator for two sinusoids in additive white Gaussian noise.

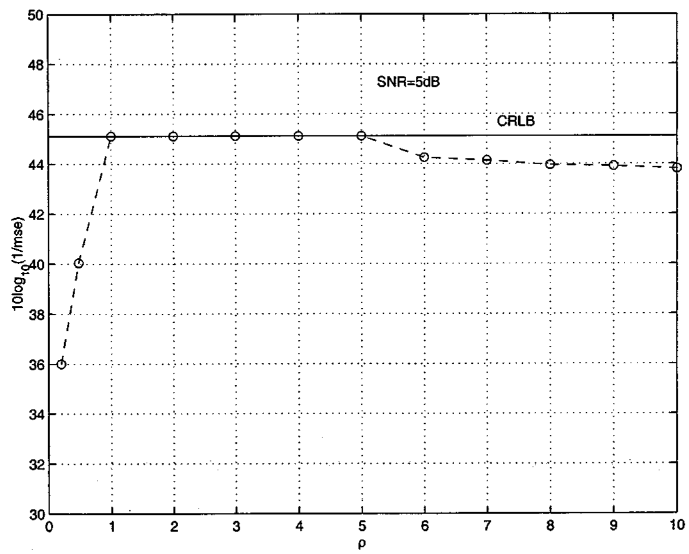
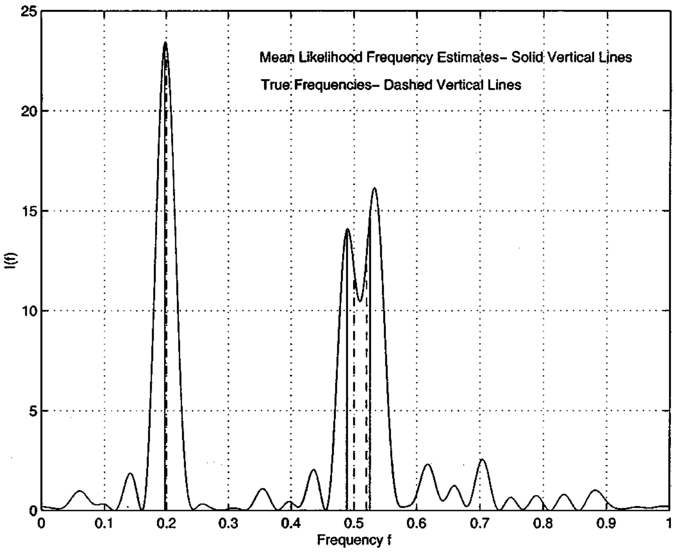


Fig. 6. Performance versus ρ for a given SNR for two sinusoids.

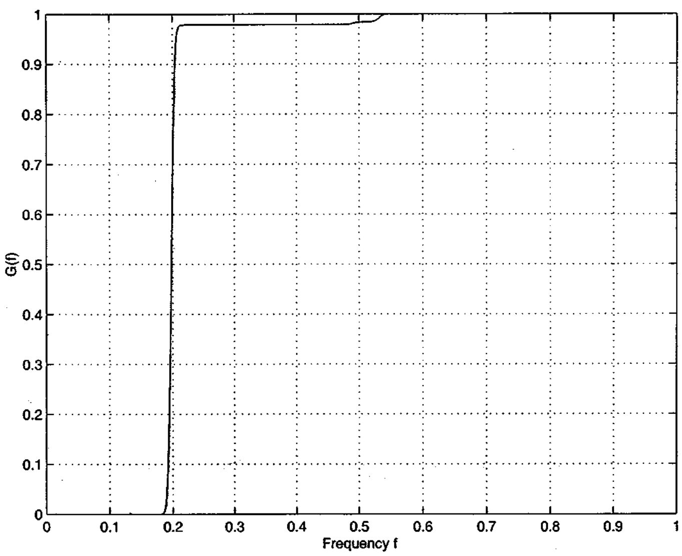
Hence, our method is about the same as the MLE in terms of performance but requires much less computation (see Section V-A).

As alluded to previously, it is important to choose an appropriate value for ρ. The effect of the value of ρ is shown in Fig. 6 for an SNR of 5 dB. The actual value of the mean square error for each ρ is shown as a circle. It is seen that for too small a value, the performance degrades. This is because the mean of the likelihood function is offset from the global maximum location due to the presence of local maxima. For too large a value, the performance also degrades, but this is due to numerical errors in computing the exponentials in (30). In theory, as $\rho \rightarrow \infty$, the MLE is obtained (see the Appendix). In practice, any value in the range shown can be used. Some experimentation is required to determine this range, which is somewhat dependent on data record length and SNR.

Example 2—Three Sinusoids: Now, we consider the case of three sinusoids with two closely spaced ones and the third one far from the other two. The sinusoidal parameters are $f_1 = 0.2, f_2 = 0.5, f_3 = 0.52, \phi_1 = 0, \phi_2 = 0, \phi_3 = \pi/4, A_1 = A_2 = A_3 = 1$. Fig. 7(a) and (b) shows the periodogram of the data and the cumulative distribution function, respectively. As in the first example, the periodogram peak locations are biased estimates. For comparison, the MELE is shown for this one realization of data. The cumulative distribution function exhibits transitions near 0.2, 0.5, and 0.52, i.e., at the true frequency locations. In Fig. 8, we plot $10\log_{10}$ (1/mean square error) versus SNR and show the Cramér-Rao lower bound on the same plot to benchmark the performance. We observe that the estimates attain the bound above an SNR of about 1 dB. In this example, 5000 realizations of the frequency vector were sufficient for the method to achieve the CRLB. For the best performance, ρ needs to be chosen appropriately , as previously discussed. For this example, we chose $\rho = 0.48$, which proved to be adequate. It has been observed that at high SNR's for the three sinusoid case, the performance becomes more sensitive to the choice of ρ, unlike



(a)



(b)

Fig. 7. (a) Periodogram of the data for three sinusoids in additive white Gaussian noise, SNR $=$ 5 dB. (b) Plot of cumulative distribution function $G(f)$ for the three-sinusoid case, SNR $= 5$ dB.

the case for two sinusoids. Thus, for the three-sinusoid case, it may be possible that there exists a better choice for the modified likelihood function for which the performance does not vary as much with the choice of ρ. This issue needs to be investigated further. As the MLE becomes too computationally intensive for more than two sinusoids, we did not carry out the performance evaluation via Monte Carlo runs for all SNR's. However, the threshold SNR for the MLE for the three sinusoid case was also found to be about 1 dB. According to theory, it should achieve the CRLB for all SNR's above the threshold. ◇

A. Analysis of Computation

In the MELE method, the two major sources of computations involved are generation of the realizations of the frequency vector and the evaluations of the function $w(\mathbf{f})$ for the realized frequency samples (see Section IV-B). In generating the frequency vector, we have used the golden search to reduce the computational burden involved. It required at most 24 evaluations of the cumulative distribution function per single frequency realization. The computation involved in these two steps is significantly less than that for the MLE, which for the two-sinusoid case requires at least $10^4/2$ evaluations of the likelihood function and at least $10^6/6$ evaluations for the three-sinusoid case. This is because for a minimum frequency separation of 0.02, a grid search for the MLE requires at least 100 grid points. Clearly, the computational burden of the MLE increases exponentially with the number of sinusoids, whereas for our technique, it does not. In particular, for the two simulation examples considered, the number of frequency realizations did not have to be increased exponentially to achieve the CRLB. The two-sinusoid case required 2000 frequency realizations, whereas the three-sinusoid case required 5000 frequency realizations for good performance. In terms of FLOP's, the MLE required more than a factor of 17:1 for the two-sinusoid case and a factor of 285:1 for the three-sinusoid case.

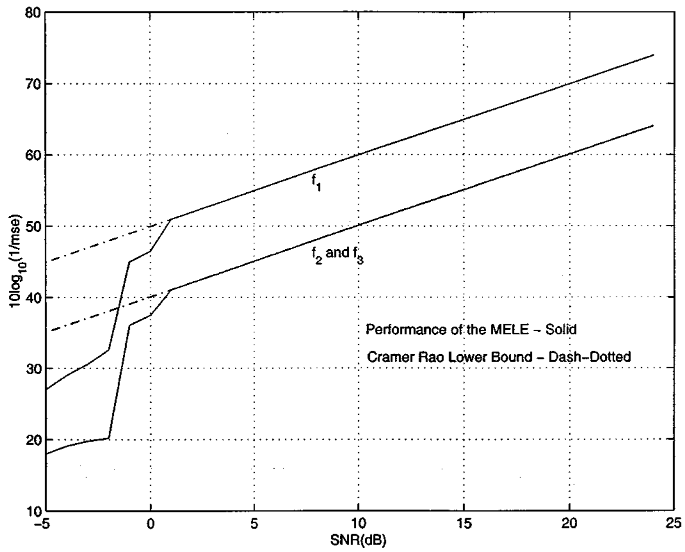


Fig. 8. Performance of the mean likelihood estimator for three sinusoids in additive white Gaussian noise.

VI. Other Problems of Interest

Although we have applied our approach to sinusoidal frequency estimation in this paper, the method is more generally applicable. Whenever the signal has linear as well as nonlinear parameters and is embedded in white Gaussian noise, the MELE approach can be applied. Some other examples follow.

1) Some extensions to the sinusoidal problem are the estimation of parameters of superimposed chirp signals in noise and the estimation of parameters for damped exponentials in noise. The first of these signals is described as

$$x[n] = \sum_{i=1}^{p} A_i \exp(j[2\pi(f_i n + \tfrac{1}{2} m_i n^2) + \phi_i]) + w[n].$$

Here, $x[n]$ is linear with respect to the complex amplitudes $A_i \exp(j\phi_i)$ and nonlinear with respect to the sweep rates m_i's and frequencies f_i's. The MLE of m_i, f_i, and A_i requires first the MLE of the nonlinear parameters, i.e., m_i's and f_i's, similar to the estimation of frequencies in the sinusoidal parameter estimation problem. The MLE of the chirp rates m_i's and the frequencies f_i's will, however, require a grid search. Instead of the impractical grid search, the importance sampling-based MELE can be implemented. Note that the sinusoidal parameter estimation problem is a special case of this in which the sweep rates m_i's are all zero. The second estimation problem concerns damped sinusoids in noise or

$$x[n] = \sum_{i=1}^{p} A_i r_i^n \exp[j(2\pi f_i n + \phi_i)] + w[n]$$

where each signal is a complex decaying exponential. Note that the sinusoidal parameter estimation problem is a special case of this in which the magnitude of each of the r_i's is unity.

2) Maximum likelihood estimation of multiple attenuations and multiple time delays in a multipath environment is another example. The data $x[n]$ in this case can be expressed as

$$x[n] = \sum_{i=1}^{p} a_i s(n - n_i) + w[n].$$

Here, the signal is linear with respect to the attenuations a_i but nonlinear with respect to the time delays n_i.

3) Finally, as mentioned in the introduction, the DOA estimation problem can be solved using the proposed approach. This extension is currently under investigation.

In summary, the MELE procedure can be used for a wide variety of signal processing problems of interest.

VII. Conclusions

Many important problems in statistical signal processing involve the estimation of nonlinear as well as linear parameters of signals in noise. In these cases, maximum likelihood estimation reduces to a separable optimization problem. By applying MELE, a reduction in estimator complexity has been achieved while retaining the good performance of the MLE. Furthermore, the method is not iterative in nature so that there is no question of convergence or accuracy required of an initial iterate. Although we have discussed mainly the sinusoidal frequency estimation problem, the technique can be applied to many other nonlinear regression problems. However, questions such as the best choice of the importance function in calculating the mean likelihood estimate will need to be investigated.

Appendix A
Asymptotic Equivalence of Mean Likelihood and Maximum Likelihood Frequency Estimators

Consider the basic MELE using (26) or

$$\begin{aligned}\hat{f}_i &= \frac{1}{2\pi}\angle \int_0^1 \cdots \int_0^1 L_{\mathrm{mod}}(\mathbf{f}) \exp(j2\pi f_i)\, d\mathbf{f} \\ &= \frac{1}{2\pi}\angle \int_0^1 \cdots \int_0^1 \exp(\rho \mathbf{x}^H \mathbf{H}(\mathbf{H}^H\mathbf{H})^{-1}\mathbf{H}^H\mathbf{x}) \\ &\quad \cdot \exp(j2\pi f_i)\, d\mathbf{f}\end{aligned}$$

for $i = 1, \cdots, p$. Let $J(\mathbf{f}) = \mathbf{x}^H \mathbf{H}(\mathbf{H}^H\mathbf{H})^{-1}\mathbf{H}^H\mathbf{x}$, and assume that $\mathbf{f}_0 = \arg\ \max J(\mathbf{f})$. Thus, by definition, $\mathbf{f}_0$ is the MLE and is assumed unique. Then

$$\hat{f}_i = \frac{1}{2\pi}\angle \int_0^1 \cdots \int_0^1 \exp(\rho J(\mathbf{f})) \exp(j2\pi f_i)\, d\mathbf{f}.$$

Since $J(\mathbf{f}_0)$ is real, we can rewrite this as

$$\hat{f}_i = \frac{1}{2\pi}\angle \frac{\int_0^1 \cdots \int_0^1 \exp(\rho J(\mathbf{f})) \exp(j2\pi f_i)\, d\mathbf{f}}{\exp(\rho J(\mathbf{f}_0))}$$

or

$$\begin{aligned}\hat{f}_i = \frac{1}{2\pi}\angle \int_0^1 \cdots \int_0^1 \exp[-\rho(J(\mathbf{f_0}) - J(\mathbf{f}))] \\ \cdot \exp(j2\pi f_i)\, d\mathbf{f}.\end{aligned}$$

Since $Q(\mathbf{f}) = J(\mathbf{f_0}) - J(\mathbf{f}) \geq 0$, $Q(\mathbf{f})$ has a minimum at $\mathbf{f} = \mathbf{f}_0$. In addition, note that the Hessian of $Q(\mathbf{f})$ is positive definite when evaluated at $\mathbf{f} = \mathbf{f}_0$. Thus, by Laplace's theorem, we have as $\rho \to \infty$ [5]

$$\begin{aligned}\hat{f}_i = \frac{1}{2\pi}\angle \Bigg\{ \left(\frac{2\pi}{\rho}\right)^{p/2} \det{}^{\frac{1}{2}} \left[\frac{\partial^2 Q(\mathbf{f})}{\partial \mathbf{f} \partial \mathbf{f}^T}\right]\Bigg|_{\mathbf{f_0}} \\ \cdot\ \exp(j2\pi[\mathbf{f}_0]_i) + \tilde{O}\left(\frac{1}{\rho^{p/2+1}}\right)\Bigg\}\end{aligned}$$

where $\partial^2 Q(\mathbf{f})/\partial\mathbf{f}\partial\mathbf{f}^T$ is the Hessian, and $\tilde{O}(\xi)$ denotes a complex number $\tilde{c}$ such that $|\tilde{c}| \leq K\xi$ for some constant K. Noting that the Hessian is real and positive definite and, thus, that the determinant is a positive real number, we have that as $\rho \to \infty$, $\hat{f}_i \to [\mathbf{f}_0]_i$, where the latter is the MLE.

Acknowledgment

The authors wish to thank P. Swaszek of the University of Rhode Island for useful discussions related to importance sampling.

References

[1] G. A. Barnard, "Statistical inference," *J. R. Stat. Soc.*, vol. B11, 1949.
[2] J. O. Berger, *Statistical Decision Theory and Bayesian Analysis*, New York: Springer, 1985.
[3] J. M. Bernardo and A. F. M. Smith, *Bayesian Theory*. New York: Wiley, 1994.
[4] W. B. Bishop and P. M. Djuric, "Model order selection of damped sinusoids in noise by predictive densities," *IEEE Trans. Signal Processing*, vol. 44, pp. 611–619, Mar. 1996.
[5] N. Bleistein and R. A. Handelsman, *Asymptotic Expansion of Integrals*. New York: Dover, 1986.
[6] G. L. Bretthorst, *Bayesian Spectrum Analysis and Parameter Estimation*. New York: Springer-Verlag, 1988.
[7] S. M. Kay, "Comments on frequency estimation by linear prediction," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-27, pp. 198–199, Apr. 1979.
[8] ——, *Modern Spectral Estimation: Theory and Application*. Englewood Cliffs, NJ: Prentice-Hall, 1988.
[9] ——, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Englewood Cliffs, NJ: Prentice-Hall, 1993.
[10] R. Kumaresan, "Estimating the parameters of exponentially damped or undamped sinusoidal signals in noise," Ph.D. dissertation, Univ. Rhode Island, Kingston, 1982.
[11] B. C. Lovell, P. J. Kootsookos, and R. C. Williamson, "The circular nature of discrete-time frequency estimates," in *Pro. IEEE Int. Conf. Acoust., Speech, Signal Process.*, May 1991, pp. 3369–3372.
[12] K. V. Mardia, *Statistics of Directional Data*. New York: Academic, 1972.

[13] G. A. F. Seber and C. J. Wild, *Nonlinear Regression*. New York: Wiley, 1989.
[14] A. F. M. Smith and A. Gelfand, *Bayesian statistics without tears: A sampling-resampling framework*. New York: Amer. Stat. Assoc., 1992.
[15] L. Stewart, "Bayesian analysis using Monte Carlo integration: A powerful methodology for handling some difficult problems," *Amer. Stat.*, pp. 195–200, 1983.
[16] P. Stoica, L. Moses, B. Friedlander, and T. Soderstrom, "Maximum likelihood estimation of parameters of multiple sinusoids from noisy measurements," *IEEE Trans. Acoust., Speech Signal Processing*, vol. 37, pp. 378–392, Mar. 1989.
[17] B. Van Veen and K. Buckley, "Beamforming: A versatile approach to spatial filtering," *IEEE ASSP Mag.*, Apr 1988.
[18] D. J. Wilde, *Globally Optimum Design*. Englewood Cliffs, NJ: Prentice-Hall, 1964.

Steven Kay (F'89) was born in Newark, NJ, on April 5, 1951. He received the B.E. degree from Stevens Institute of Technology, Hoboken, NJ, in 1972, the M.S. degree from Columbia University, New York, NY, in 1973, and the Ph.D. degree from Georgia Institute of Technology (Georgia Tech), Atlanta, in 1980, all in electrical engineering.

From 1972 to 1975, he was with Bell Laboratories, Holmdel, NJ, where he was involved with transmission planning for speech communications and simulation and subjective testing of speech processing algorithms. From 1975 to 1977, he attended Georgia Tech to study communication theory and digital signal processing. From 1977 to 1980, he was with the Submarine Signal Division, Portsmouth, RI, where he engaged in research on autoregressive spectral estimation and the design of sonar systems. He is presently Professor of electrical engineering at the University of Rhode Island, Kingston, and a consultant to industry and the Navy. He has written numerous papers and is a contributor to several edited books. He is the author of the textbooks *Modern Spectral Estimation* (Englewood Cliffs, NJ: Prentice-Hall, 1988), *Fundamentals of Statistical Signal Processing, Vol. I: Estimation Theory* (Englewood Cliffs, NJ: Prentice-Hall, 1993), and *Fundamentals of Statistical Signal Processing, Vol. II: Detection Theory* (Englewood Cliffs, NJ: Prentice-Hall, 1998). His current interests are spectrum analysis, detection and estimation theory, and statistical signal processing.

Dr. Kay is a member of Tau Beta Pi and Sigma Xi. He has served on the IEEE Acoustics, Speech, and Signal Processing Committee on Spectral Estimation and Modeling.

Supratim Saha received the B.Tech. (Hons.) degree from the Indian Institute of Technology, Kharagpur, in 1993 and the M.S. degree from the Indian Institute of Science, Bangalore, in 1996, both in electrical engineering. Since August 1997, he has been with the Department of Electrical Engineering, University of Rhode Island, Kingston, where he is pursuing the Ph.D degree.

From July 1993 to April 1994, he was with the Indian Aluminum Company, where he worked as an Instrumentation Engineer. From August 1996 to June 1997, he was with the Department of Electrical and Computer Engineering, Indian Institute of Science, Bangalore, where he worked as a Project Associate. His research interests are estimation, detection, and general Monte Carlo techniques.