Maximum Likelihood Parameter Estimation of Superimposed Chirps Using Monte Carlo Importance Sampling

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

***Abstract*—We address the problem of parameter estimation of superimposed chirp signals in noise. The approach used here is a computationally modest implementation of a maximum likelihood (ML) technique. The ML technique for estimating the complex amplitudes, chirping rates, and frequencies reduces to a separable optimization problem where the chirping rates and frequencies are determined by maximizing a compressed likelihood function that is a function of only the chirping rates and frequencies. Since the compressed likelihood function is multidimensional, its maximization via a grid search is impractical. We propose a noniterative maximization of the compressed likelihood function using importance sampling. Simulation results are presented for a scenario involving closely spaced parameters for the individual signals.**

I. Introduction

CHIRP signals are encountered in many different engineering applications including radar, active sonar, and passive sonar systems. The problem of parameter estimation of chirp signals has received a great deal of attention [3], [6], [7]. These approaches have been proven to be effective in the sense that they achieve the Cramér–Rao lower bound (CRLB). However, most of these approaches are designed for a single chirp signal. Parameter estimation of superimposed chirp signals is a difficult signal processing problem. The need for determining the parameters of superimposed chirp signals arises in passive sensor array systems, where it has been shown in [13] that the problem of range and direction-of-arrival estimation for moderately far, broadside targets reduces to that of estimating the parameters of sums of chirp signals. Liang and Arun [12] have also addressed an iterative maximum likelihood (ML) approach to this problem. Rank-reduction techniques were used to get good initial parameter estimates, which were then used in an ML procedure to obtain the final estimates. Although the approach has been shown to achieve good results at high SNRs, there is no guarantee that the global optimum will be achieved.

Our aim in this paper is to develop a noniterative computationally modest implementation of a ML estimator for the chirp signal parameters. To develop the estimator, we first show that the data model involves estimation of linear and nonlinear parameters of a partial general linear model [1]. The complex amplitudes form the linear parameter vector, and the chirp rates and frequencies form the nonlinear parameter vector. The parameter estimation problem becomes decoupled, where the nonlinear parameter vector needs to be estimated first by maximizing a compressed likelihood function involving only the chirp rates and frequencies as unknown parameters. The complex amplitude estimates are then obtained from the estimates of chirp rates and frequencies. In this paper, we focus on estimation of chirp rates and frequencies only. The straightforward implementation of the maximization of the compressed likelihood function involves a grid search that is impractical and whose computational complexity increases with the number of signals. To carry out this maximization noniteratively, we use a global optimization theorem proposed in [2]. This optimization algorithm has been used for estimation of frequencies of multiple sinusoids in noise [4]. This algorithm has also been used for the design of sensor locations and shading weights of a sparse linear array [5]. To efficiently implement the optimization, we use Monte Carlo importance sampling [8]. It is observed that the technique produces good estimates for the unknown parameters, even in cases where the individual parameters are closely spaced. The method achieves the CRLB for moderate and high SNRs. Furthermore, the computational burden is quite modest.

Manuscript received January 31, 2001; revised October 4, 2001. The associate editor coordinating the review of this paper and approving it for publication was Dr. Petar M. Djurić.

The authors are with the Department of Electrical and Computer Engineering, University of Rhode Island, Kingston, RI 02881 USA (e-mail: sup@ele.uri.edu).

Publisher Item Identifier S 1053-587X(02)00565-2.

II. Problem Definition

A sequence $x[n]$, $n = 0, \ldots, N-1$ is observed having the following parametric representation:

$$x[n] = \sum_{i=1}^{p} A_i \exp\left[j2\pi\left(f_i n + \frac{m_i}{2} n^2\right)\right] + w[n] \quad (1)$$

where the parameters, chirp rate m_i ($0 \le m_i \le 2$), frequency f_i ($0 \le f_i \le 1$), and the complex amplitudes A_i for $i = 1, \ldots p$ are unknown. The noise $w[n]$, $n = 0, \ldots, N-1$ is a segment of a zero mean complex white Gaussian random process. The aim is to obtain ML estimates of the chirp rate m_i and frequency f_i for $i = 1, \ldots, p$ from $x[n]$ for $n = 0, \ldots, N-1$.

A. ML Estimation

The data described by (1) can be expressed in matrix form as

$$\mathbf{x} = \mathbf{H}(\boldsymbol{\alpha}, \boldsymbol{\beta})\boldsymbol{\theta} + \mathbf{w} \quad (2)$$

where $\mathbf{x}$ is a $N \times 1$ vector given by $\mathbf{x} = [x[0] \ \cdots \ x[N-1]]^T$, $\mathbf{w}$ is a $N \times 1$ noise vector given by $\mathbf{w} = [w[0] \ \cdots \ w[N-1]]^T$, and

$\boldsymbol{\theta} = [\theta_1 \ \cdots \ \theta_p]^T$. The $N \times p$ matrix $\mathbf{H}(\boldsymbol{\alpha}, \boldsymbol{\beta})$ can be expressed as

$$\mathbf{H}(\boldsymbol{\alpha}, \boldsymbol{\beta}) = [\mathbf{e}(\alpha_1, \beta_1) \ \cdots \ \mathbf{e}(\alpha_p, \beta_p)] \tag{3}$$

where the $N \times 1$ vector $\mathbf{e}(\alpha_i, \beta_i)$ is given by

$$\mathbf{e}(\alpha_i, \beta_i) = \begin{bmatrix} \exp\left(j2\pi\left(f_i(0) + \frac{m_i}{2}(0)^2\right)\right) \\ \exp\left(j2\pi\left(f_i(1) + \frac{m_i}{2}(1)^2\right)\right) \\ \vdots \\ \exp\left(j2\pi\left(f_i(N-1) + \frac{m_i}{2}(N-1)^2\right)\right) \end{bmatrix} \tag{4}$$

where

$$\boldsymbol{\alpha} = [f_1 \ \cdots \ f_p] \tag{5}$$

and

$$\boldsymbol{\beta} = [m_1 \ \cdots \ m_p]. \tag{6}$$

Since the noise is assumed to be additive white Gaussian noise with variance σ^2, the probability density function (pdf) of the data vector $\mathbf{x}$ in (2) parameterized by $\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\theta}$ is given by $p(\mathbf{x}; \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\theta})$, which is equal to

$$\frac{1}{\pi^N \sigma^{2N}} \exp\left[-\frac{1}{\sigma^2}(\mathbf{x} - \mathbf{H}(\boldsymbol{\alpha}, \boldsymbol{\beta})\boldsymbol{\theta})^H(\mathbf{x} - \mathbf{H}(\boldsymbol{\alpha}, \boldsymbol{\beta})\boldsymbol{\theta})\right]. \tag{7}$$

Hence, the likelihood function of the data $L(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\theta})$ is given by

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

The joint ML estimate of $\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\theta}$ is obtained by maximizing $L(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\theta})$. From (7) and (8), this joint maximization is equivalent to

$$\min_{\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\theta}} (\mathbf{x} - \mathbf{H}(\boldsymbol{\alpha}, \boldsymbol{\beta})\boldsymbol{\theta})^H(\mathbf{x} - \mathbf{H}(\boldsymbol{\alpha}, \boldsymbol{\beta})\boldsymbol{\theta}).$$

The parameter vectors $\boldsymbol{\alpha}, \boldsymbol{\beta}$ that appear in the matrix $\mathbf{H}$ are nonlinearly related to $\mathbf{x}$, whereas the parameter vector $\boldsymbol{\theta}$ is linearly related to $\mathbf{x}$. It is known that for such kinds of joint parameter estimation problems, the parameter estimation procedure is decoupled [1], where estimation of the unknown nonlinear parameters is done first, and the estimated nonlinear parameters are inserted in the matrix $\mathbf{H}(\boldsymbol{\alpha}, \boldsymbol{\beta})$ to obtain the linear parameters. The estimates of the two nonlinear parameters are obtained as [1]

$$\left[\hat{\boldsymbol{\alpha}}_{mle}, \hat{\boldsymbol{\beta}}_{mle}\right] = \max_{\boldsymbol{\alpha}, \boldsymbol{\beta}} \left[\mathbf{x}^H\left(\mathbf{H}(\boldsymbol{\alpha}, \boldsymbol{\beta})\left[\mathbf{H}^H(\boldsymbol{\alpha}, \boldsymbol{\beta})\mathbf{H}(\boldsymbol{\alpha}, \boldsymbol{\beta})\right]^{-1}\mathbf{H}^H(\boldsymbol{\alpha}, \boldsymbol{\beta})\right)\mathbf{x}\right]. \tag{9}$$

The function in the right-hand side of (9) is called the compressed likelihood function $L_c(\boldsymbol{\alpha}, \boldsymbol{\beta})$. It can be observed from (9) that obtaining $[\hat{\boldsymbol{\alpha}}_{mle}, \hat{\boldsymbol{\beta}}_{mle}]$ will require a multidimensional grid search over the two parameter vectors. It is because of the lack of closed-form solution that the proposed approaches for these kinds of problems have been iterative. Pincus [2] showed that for a function of several variables having many local maxima, it is possible to have a closed-form expression for the values of the variables yielding the global maximum. Motivated by the result of [2], we develop a noniterative estimator for $[\boldsymbol{\alpha}\ \boldsymbol{\beta}]$.

III. Global Optimization Theorem

The theorem proposed by Pincus [2] is used to obtain the maximum/minimum of a multidimensional function having unique global maximum/minimum.

We apply this theorem to obtain the estimates of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ that maximize the compressed likelihood function $L_c(\boldsymbol{\alpha}, \boldsymbol{\beta})$. Based on the theorem [2], the estimates $\hat{\boldsymbol{\alpha}}$ and $\hat{\boldsymbol{\beta}}$ are given by

$$[\hat{\boldsymbol{\alpha}}]_i = \lim_{\rho \to \infty} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} [\boldsymbol{\alpha}]_i \overline{L_c'}(\boldsymbol{\alpha}, \boldsymbol{\beta})\, d\boldsymbol{\alpha}\, d\boldsymbol{\beta} \tag{10}$$

and

$$\left[\hat{\boldsymbol{\beta}}\right]_i = \lim_{\rho \to \infty} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} [\boldsymbol{\beta}]_i \overline{L_c'}(\boldsymbol{\alpha}, \boldsymbol{\beta})\, d\boldsymbol{\alpha}\, d\boldsymbol{\beta} \tag{11}$$

where $\overline{L_c'}(\boldsymbol{\alpha}, \boldsymbol{\beta})$ is the normalized compressed likelihood function

$$\overline{L_c'}(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \frac{\exp(\rho L_c(\boldsymbol{\alpha}, \boldsymbol{\beta}))}{\displaystyle\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \exp(\rho L_c(\boldsymbol{\alpha}, \boldsymbol{\beta}))\, d\boldsymbol{\alpha}\, d\boldsymbol{\beta}} \tag{12}$$

and $[\hat{\boldsymbol{\alpha}}]_i$ and $[\hat{\boldsymbol{\beta}}]_i$ are the ith components of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$. Although the theorem states that the global optimum is attained for the limiting case $\rho \to \infty$, it does not mean that the limit has to be always very large. In fact, this limit is problem specific and can be some finite number for a specific problem. If the global optimum is attained for some value of finite ρ, then it will be attained for all values of ρ above that finite value. It can be observed from (10) and (11) that the theorem provides a closed-form expression for obtaining the parameters that maximize the function, but its evaluation requires computation of a multidimensional integral. However, it can be noted that the integrations involved in (10) and (11) are closely related to integrations involved in probability theory required to compute expected values of random variables. This is because the normalized function $\overline{L_c'}(\boldsymbol{\alpha}, \boldsymbol{\beta})$ is positive and has all the properties of a joint PDF. This is because in (9), the matrix $[\mathbf{H}^H(\boldsymbol{\alpha}, \boldsymbol{\beta})\mathbf{H}(\boldsymbol{\alpha}, \boldsymbol{\beta})]$ satisfies the properties of a positive definite matrix. Thus, $L_c(\boldsymbol{\alpha}, \boldsymbol{\beta})$, and hence, $\overline{L_c'}(\boldsymbol{\alpha}, \boldsymbol{\beta})$ is positive. However, the parameter vectors $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are not random. Thus, the normalized function is termed a pseudo-PDF. Using this concept, the Monte Carlo techniques can be used to replace the multidimensional integrations in (10) and (11). The simplest Monte Carlo approach would require generation of random vectors $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ distributed according to the joint PDF $\overline{L_c'}(\boldsymbol{\alpha}, \boldsymbol{\beta})$. However, $\overline{L_c'}(\boldsymbol{\alpha}, \boldsymbol{\beta})$ is a highly nonlinear function of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$. As a result, direct generation of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ realizations is not easy, and one needs to resort to *other* Monte Carlo techniques [9] that generate samples according to some simpler PDF and use those samples to estimate the means. Importance sampling belongs to this class of Monte Carlo techniques and has been proven to

be a highly effective tool in evaluation of integrals in Bayesian theory [8]. We thus use importance sampling described in the next section to efficiently evaluate the estimates of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$.

IV. Importance Sampling

The importance sampling approach [9], [11] is based on the observation that integrals of the type $\int h(\mathbf{x})\overline{L}'(\mathbf{x})\,\mathbf{dx}$ can be expressed as

$$\int h(\mathbf{x})\overline{L}'(\mathbf{x})\,\mathbf{dx} = \int h(\mathbf{x})\,\frac{\overline{L}'(\mathbf{x})}{\overline{g}(\mathbf{x})}\,\overline{g}(\mathbf{x})\,d\mathbf{x} \tag{13}$$

where $\overline{g}(\mathbf{x})$ is assumed to possess all the properties of a PDF. $\overline{g}(\mathbf{x})$ is chosen in such a way that $\overline{g}(\mathbf{x}) > 0$ whenever $\overline{L}'(\mathbf{x}) > 0$. Then, the right-hand side of (13) can be expressed as the expected value of $h(\mathbf{x})(\overline{L}'(\mathbf{x})/\overline{g}(\mathbf{x}))$, with respect to the pseudo-PDF $\overline{g}(\mathbf{x})$. The function $\overline{g}(\mathbf{x})$ is called the normalized importance function. Unlike $\overline{L}'(\mathbf{x})$, which, in general, is a nonlinear function of $\mathbf{x}$, $\overline{g}(\mathbf{x})$ can be chosen to be some simple function of $\mathbf{x}$ so that realizations of $\mathbf{x}$ can be easily generated. Then, the value of the integral in (13) can be found by the Monte Carlo approximation

$$\frac{1}{R}\sum_{k=1}^{R} h(\mathbf{x}_k)\,\frac{\overline{L}'(\mathbf{x}_k)}{\overline{g}(\mathbf{x}_k)} \tag{14}$$

where $\mathbf{x}_k$ is the kth realization of the vector $\mathbf{x}$ distributed according to the pseudo-PDF $\overline{g}(\mathbf{x})$. The value of R needed for a good approximation depends on the choice of $\overline{g}$. Typically, $\overline{g}(\mathbf{x})$ should be chosen similar to $\overline{L}(\mathbf{x})$ as this reduces the variance of the estimate given by (14). However, another important point to keep in mind when choosing $\overline{g}(\mathbf{x})$ is that it should be simple enough so that $\mathbf{x} \sim \overline{g}(\mathbf{x})$ can be easily generated [8].

The ideas expressed by (13) and (14) can be applied to the estimation of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ once the importance function for this problem is defined. In particular, if the normalized importance function is $\overline{g}(\boldsymbol{\alpha}, \boldsymbol{\beta})$, then the estimates of the coordinates of the vector $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ computed using this importance function are expressed as

$$[\hat{\boldsymbol{\alpha}}]_i = \frac{1}{R}\sum_{k=1}^{R} [\boldsymbol{\alpha}_k]_i\,\frac{\overline{L_c'}(\boldsymbol{\alpha}_k, \boldsymbol{\beta}_k)}{\overline{g}(\boldsymbol{\alpha}_k, \boldsymbol{\beta}_k)} \tag{15}$$

and

$$\left[\hat{\boldsymbol{\beta}}\right]_i = \frac{1}{R}\sum_{k=1}^{R} [\boldsymbol{\beta}_k]_i\,\frac{\overline{L_c'}(\boldsymbol{\alpha}_k, \boldsymbol{\beta}_k)}{\overline{g}(\boldsymbol{\alpha}_k, \boldsymbol{\beta}_k)} \tag{16}$$

where $\boldsymbol{\alpha}_k$ and $\boldsymbol{\beta}_k$ are the kth realizations of the vectors $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ distributed according to the importance function $\overline{g}(\boldsymbol{\alpha}, \boldsymbol{\beta})$.

A. Choice of Importance Function

The normalized importance function $\overline{g}(\boldsymbol{\alpha}, \boldsymbol{\beta})$ needs to be chosen so that the samples $\boldsymbol{\alpha}_k$ and $\boldsymbol{\beta}_k$ can be easily generated. Furthermore, $\overline{g}(\boldsymbol{\alpha}, \boldsymbol{\beta})$ should be a close approximation to $\overline{L_c'}(\boldsymbol{\alpha}, \boldsymbol{\beta})$. Since

$$L_c(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \mathbf{x}^H\mathbf{H}(\boldsymbol{\alpha}, \boldsymbol{\beta})\left[\mathbf{H}^H(\boldsymbol{\alpha}, \boldsymbol{\beta})\mathbf{H}(\boldsymbol{\alpha}, \boldsymbol{\beta})\right]^{-1}\mathbf{H}^H(\boldsymbol{\alpha}, \boldsymbol{\beta})\mathbf{x}$$

it is obvious that $L_c'(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \exp(\rho L_c(\boldsymbol{\alpha}, \boldsymbol{\beta}))$ is a complicated function of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$. However, if we force the matrix $\mathbf{H}^H(\boldsymbol{\alpha}, \boldsymbol{\beta})\mathbf{H}(\boldsymbol{\alpha}, \boldsymbol{\beta})$ to be an identity matrix, then the function $L_c(\boldsymbol{\alpha}, \boldsymbol{\beta})$ will become separable in $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$. This is the main idea behind choosing the importance function. Thus, the importance function is chosen by forcing the $p \times p$ matrix $\mathbf{H}^H(\boldsymbol{\alpha}, \boldsymbol{\beta})\mathbf{H}(\boldsymbol{\alpha}, \boldsymbol{\beta})$ to be $N\mathbf{I}_p$, where $\mathbf{I}_p$ is a $p \times p$ identity matrix. To make the importance function similar to the function $L_c'(\boldsymbol{\alpha}, \boldsymbol{\beta})$, we choose the importance function as

$$g(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \exp\left(\rho_1\mathbf{x}^H\mathbf{H}(\boldsymbol{\alpha}, \boldsymbol{\beta})\,\frac{\mathbf{I}}{N}\,\mathbf{H}^H(\boldsymbol{\alpha}, \boldsymbol{\beta})\mathbf{x}\right) \tag{17}$$

and its normalized version as $\overline{g}(\boldsymbol{\alpha}, \boldsymbol{\beta})$

$$\overline{g}(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \frac{g(\boldsymbol{\alpha}, \boldsymbol{\beta})}{\displaystyle\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} g(\boldsymbol{\alpha}, \boldsymbol{\beta})\,d\boldsymbol{\alpha}\,d\boldsymbol{\beta}}. \tag{18}$$

Since $\mathbf{H}(\boldsymbol{\alpha}, \boldsymbol{\beta}) = [\mathbf{e}(\alpha_1, \beta_1)\ \cdots\ \mathbf{e}(\alpha_p, \beta_p)]$, where $\mathbf{e}(\alpha_p, \beta_p)$ has been defined in (4), we can write $g(\boldsymbol{\alpha}, \boldsymbol{\beta})$ as

$$g(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \exp\left[\rho_1\sum_{i=1}^{p}\frac{1}{N}\left|\sum_{n=0}^{N-1} x[n]\exp\left(-j2\pi\left(f_i n + \frac{m_i}{2}n^2\right)\right)\right|^2\right] \tag{19}$$

or

$$g(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \prod_{i=1}^{p}\exp[\rho_1 I(\alpha_i, \beta_i)] \tag{20}$$

where

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

As a result of this choice of $g(\boldsymbol{\alpha}, \boldsymbol{\beta})$ in (17), the importance function now becomes separable in $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ and can be expressed as

$$\overline{g}(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \prod_{i=1}^{p}\overline{g}(\alpha_i, \beta_i) \tag{21}$$

where

$$\overline{g}(\alpha_i, \beta_i) = \frac{\exp[\rho_1 I(\alpha_i, \beta_i)]}{\displaystyle\iint\exp[\rho_1 I(\alpha_i, \beta_i)]\,d\alpha_i\,d\beta_i}.$$

Note that this normalized importance function is a reasonable approximation to the target function $\overline{L_c'}(\boldsymbol{\alpha}, \boldsymbol{\beta})$. This is because the matrix $\mathbf{H}(\boldsymbol{\alpha}, \boldsymbol{\beta})$ has its columns parameterized by (α, β), and for well separated (α, β) pairs parameterizing the columns, the matrix $[\mathbf{H}^H(\boldsymbol{\alpha}, \boldsymbol{\beta})\mathbf{H}(\boldsymbol{\alpha}, \boldsymbol{\beta})]$ will be close to an identity matrix at large number of points in the multidimensional plane. This occurs because of the properties of complex exponentials that exhibit approximate orthogonality for well-separated parameters. This was the philosophy behind the choice of this importance function. This enables generation of p independent samples of (α_i, β_i) distributed according to the joint PDF $\overline{g}(\alpha_i, \beta_i)$ with the condition that no two of (α_i, β_i) are the

same. The p realizations of (α_i, β_i) for $i = 1, \ldots, p$ need to be distinct. This assumption is necessary for this problem as otherwise, the matrix $\mathbf{H}^H(\boldsymbol{\alpha}, \boldsymbol{\beta})\mathbf{H}(\boldsymbol{\alpha}, \boldsymbol{\beta})$ will become singular, and hence, the signal parameters are nonidentifiable. The allowable ranges of α_i and β_i during generation are determined based on the identifiability conditions of the problem. The allowable range of α_i and β_i ensuring identifiability is $0 < \alpha_i < 1$ and $0 < \beta_i < 1$. This is derived in the Appendix.

The variables α_i and β_i are now generated jointly using the following three steps.

1) Evaluate the two-dimensional (2-D) joint PDF $\overline{g}(\alpha, \beta)$ at $M \times M$ discrete set of points on a rectangular grid, and obtain the marginal PDF $\overline{g}(\alpha)$ as

$$\overline{g}(\alpha_l) = \sum_{i=1}^{M} \overline{g}(\alpha_l, \beta_i)\delta\beta_i$$

for $l = 1, \ldots M$. α_l and β_i are the lth frequency and ith chirp rate points, respectively, on the rectangular grid. $\delta\beta_i$ refers to the spacing between any two successive chirp rate parameters sampled on the rectangular grid. As there are M such points, $\delta\beta_i = 1/M$.

From the marginal PDF $\overline{g}(\alpha_l)$, obtain the cumulative distribution function $G(\alpha)$ as

$$G(\alpha) = \int_0^{\alpha} \overline{g}(x)\, dx$$

by approximating the integral as a sum.

2) From the marginal PDF $\overline{g}(\alpha_l)$ so obtained in Step 1, obtain the conditional PDF $\overline{g}(\beta|\alpha)$ as

$$\overline{g}(\beta_k|\alpha_l) = \frac{\overline{g}(\beta_k, \alpha_l)}{\overline{g}(\alpha_l)}.$$

Evaluate the conditional cumulative distribution function of the conditional PDF $\overline{g}(\beta|\alpha)$ as

$$G(\beta_k|\alpha_l) = \int_0^{\beta_k} \overline{g}(x|\alpha_l)\, dx$$

for all $k, l = 1, \ldots, M$.

3) Generate $u_1 \sim U[0, 1]$, $u_2 \sim U[0, 1]$, and obtain $\alpha = G^{-1}(u_1)$ and $\beta = G^{-1}(u_2|\alpha)$. Repeat this step p times to obtain a realization of the vector $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, each of which has dimension $p \times 1$.

4) Repeat Step 3 R times to obtain R realizations of the vector $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$. Since, in Step 3, the pairs (α, β) are generated independently p times, any two of them may turn out to be the same, which is in violation of the identifiability conditions of the problem. To avoid such a possibility, once α_1 is generated using $\alpha_1 = G^{-1}(u_1)$, α_2 is generated by the same procedure of first generating $u_1 \sim U[0, 1]$, except that now, u_1 is compared with $G(S)$, where the set S refers to all frequency points on the grid except α_1. This can be written in set notation as

$$S = \left\{0, \frac{1}{M}, \ldots, 1 - \frac{1}{M}\right\} - \{\alpha_1\}. \tag{22}$$

Thus, α_2 is obtained as $\alpha_2 = G^{-1}(u_1)$ with allowable values of α_2 being in the set S given by (22). Continuing in this manner, α_p is generated in the same way as α_2, except that now, the set S becomes

$$S = \left\{0, \frac{1}{M}, \ldots, 1 - \frac{1}{M}\right\} - \{\alpha_1, \ldots, \alpha_{p-1}\}. \tag{23}$$

Thus, while carrying out the generation of u_1, u_2 pth time in Step 3, α_p is obtained as

$$\alpha_p = G^{-1}(u_1)$$

with allowable values of α_p being the set S given by (23).

The β_i for each α_i is obtained using $\beta_i = G^{-1}(u_2|\alpha_i)$. This eliminates the possibility of any two of (α_i, β_i) being the same for $i = 1$ to p.

By generation of the pair (α_i, β_i) using the above Steps 1–3, it may appear that the vector no longer is generated from the pseudo PDF $\overline{g}(\boldsymbol{\alpha}, \boldsymbol{\beta})$ because we have forced a condition of distinct (α_i, β_i). This would be true from a theoretical viewpoint. However, by a choice of M sufficiently large, the generation of the pair $\overline{g}(\boldsymbol{\alpha}, \boldsymbol{\beta})$ will be closely distributed according to the desired pseudo-PDF. This can be understood as follows: If the pair (α_i, β_i) is the ith generated pair, then (α_j, β_j) is the jth generated pair. If $\alpha_j \neq \alpha_i$, which occurs with high probability if M is large, then the generation will not involve any conditioning. If α_j happens to equal α_i, then this conditional generation will prevent β_j from being equal to expected β_i. However, then β_j will equal $\beta_i +$ or $-\,(1/M)$. Again, for a large value of chosen M, this generation will be closely distributed according to the desired PDF $\overline{g}(\boldsymbol{\alpha}, \boldsymbol{\beta})$. Thus, a large value of M prevents the estimator from being biased.

Now, these realizations could be used in (15) and (16) to obtain the estimates that are essentially the linear means of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$. However, we do not use (15) and (16). Rather, we make further use of the periodicity properties of α_i and β_i in reducing the computations. Since α_i and β_i, are periodic with period 1 and 2, respectively, they posses the properties of a circular random variable. We thus compute the circular means [10], [14] and obtain the angle of the circular means to compute $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$. The key idea in defining a circular mean is to average position vectors [10]. The circular mean also alleviates the bias [14]. The expressions for the estimates based on the circular mean definition are given by

$$[\hat{\boldsymbol{\alpha}}]_i = \frac{1}{2\pi}\angle\frac{1}{R}\sum_{k=1}^{R}\exp(j2\pi[\boldsymbol{\alpha}_k]_i)\,\frac{\overline{L_c'}(\boldsymbol{\alpha}_k, \boldsymbol{\beta}_k)}{\overline{g}(\boldsymbol{\alpha}_k, \boldsymbol{\beta}_k)} \tag{24}$$

and

$$\left[\hat{\boldsymbol{\beta}}\right]_i = \frac{2}{2\pi}\angle\frac{1}{R}\sum_{k=1}^{R}\exp\left(j2\,\frac{\pi}{2}\,[\boldsymbol{\beta}_k]_i\right)\frac{\overline{L_c'}(\boldsymbol{\alpha}_k, \boldsymbol{\beta}_k)}{\overline{g}(\boldsymbol{\alpha}_k, \boldsymbol{\beta}_k)}. \tag{25}$$

Since the angle operation is invariant to multiplication by a constant, (24) and (25) reduce to

$$[\hat{\boldsymbol{\alpha}}]_i = \frac{1}{2\pi}\angle\frac{1}{R}\sum_{k=1}^{R}\exp(j2\pi[\boldsymbol{\alpha}_k]_i)\,\frac{L_c'(\boldsymbol{\alpha}_k, \boldsymbol{\beta}_k)}{\overline{g}(\boldsymbol{\alpha}_k, \boldsymbol{\beta}_k)} \tag{26}$$

and

$$\left[\hat{\boldsymbol{\beta}}\right]_i = \frac{2}{2\pi}\angle\frac{1}{R}\sum_{k=1}^{R}\exp\left(j2\,\frac{\pi}{2}\,[\boldsymbol{\beta}_k]_i\right)\frac{L_c'(\boldsymbol{\alpha}_k, \boldsymbol{\beta}_k)}{g(\boldsymbol{\alpha}_k, \boldsymbol{\beta}_k)} \tag{27}$$

respectively. It should be noted that by using (26) and (27), the normalizing constants are no longer required as a result of the $\angle$ operator, thus reducing the computational burden considerably. Since $L_c'(\boldsymbol{\alpha}_k, \boldsymbol{\beta}_k) = \exp(\rho L_c(\boldsymbol{\alpha}_k, \boldsymbol{\beta}_k))$, from (20), (26) and (27) are simplified to

$$[\hat{\boldsymbol{\alpha}}]_i = \frac{1}{2\pi}\angle\frac{1}{R}\sum_{k=1}^{R}\exp(j2\pi[\boldsymbol{\alpha}_k]_i)\,\frac{\exp(\rho L_c(\boldsymbol{\alpha}_k, \boldsymbol{\beta}_k))}{\exp(\rho_1 I(\boldsymbol{\alpha}_k, \boldsymbol{\beta}_k))} \tag{28}$$

and

$$\left[\hat{\boldsymbol{\beta}}\right]_i = \frac{2}{2\pi}\angle\frac{1}{R}\sum_{k=1}^{R}\exp\left(j2\,\frac{\pi}{2}\,[\boldsymbol{\beta}_k]_i\right)\frac{\exp(\rho L_c(\boldsymbol{\alpha}_k, \boldsymbol{\beta}_k))}{\exp(\rho_1 I(\boldsymbol{\alpha}_k, \boldsymbol{\beta}_k))}. \tag{29}$$

Note that since (28) and (29) both involve computation of arguments of exponentials that may be very large, computational difficulties may result. Thus, instead of evaluating the numerator $\exp(\rho L_c(\boldsymbol{\alpha}_k, \boldsymbol{\beta}_k))$ and the denominator $\exp(\rho_1 I(\boldsymbol{\alpha}_k, \boldsymbol{\beta}_k)$ for each k individually, the computation is actually carried out using

$$[\hat{\boldsymbol{\alpha}}]_i = \frac{1}{2\pi}\angle\frac{1}{R}\sum_{k=1}^{R}\exp(j2\pi[\boldsymbol{\alpha}_k]_i) \cdot \exp(\rho L_c(\boldsymbol{\alpha}_k, \boldsymbol{\beta}_k) - \rho_1 I(\boldsymbol{\alpha}_k, \boldsymbol{\beta}_k)) \tag{30}$$

and

$$\left[\hat{\boldsymbol{\beta}}\right]_i = \frac{2}{2\pi}\angle\frac{1}{R}\sum_{k=1}^{R}\exp\left(j2\,\frac{\pi}{2}\,[\boldsymbol{\beta}_k]_i\right) \cdot \exp(\rho L_c(\boldsymbol{\alpha}_k, \boldsymbol{\beta}_k) - \rho_1 I(\boldsymbol{\alpha}_k, \boldsymbol{\beta}_k)). \tag{31}$$

In (30) and (31), first, the difference argument $(\rho L_c(\boldsymbol{\alpha}_k, \boldsymbol{\beta}_k) - \rho_1 I(\boldsymbol{\alpha}_k, \boldsymbol{\beta}_k))$ is evaluated for each k, and then, the exponential of the difference argument is computed. Use of (30) and (31) eliminates the computation difficulties that arise due to large arguments in the exponentials.

V. Simulation Results

We present an example of estimation of parameters of two equipower closely spaced chirp signals for which $A_1 = 1$, $f_1 = 0.3$, $m_1 = 0.001$, $A_2 = 1$, $f_2 = 0.32$, and $m_2 = 0.002$. The data record length is 50. The value of ρ_1 is chosen as 0.4, and ρ is chosen equal to 4. It has been observed that the estimates do not change by increasing ρ further. It should be noted that the choice of ρ and ρ_1 are highly problem specific. In theory, ρ should be as large as possible. The value of ρ_1 should be chosen not too high but also not too low. This is because the choice of too high a ρ_1 may result in domination of one signal compared with the other, whereas too low a ρ_1 may result in contribution of noise of strength comparable with the actual signals. However, the choice of ρ_1 is not a very sensitive issue. A judicious choice helps in reducing the number of importance sampling realizations R. If a poor choice is made, then the number of importance sampling realizations R may have to be larger for identical performance. Realizations of the frequency and chirp rates using the three steps discussed in Section IV-A were generated with the spacing between each sample of frequency and chirp rate being 0.0005. The number of realizations R needed to obtain estimates for all the simulations was 3500.

First, we show that the technique is able to resolve the two closely spaced signals even for moderate and low SNRs. In Fig. 1, we plot the estimates of (f_1, m_1) and (f_2, m_2). The x-axis refers to frequency, whereas the y-axis refers to chirp rate. The true values are shown by circles. The estimate for 100 realizations is plotted. It can be observed that the technique is always able to resolve the signals and that the estimates lie very near the actual signal parameters. The SNR for this example was chosen as 5 dB. The CRLB is attained for this SNR.

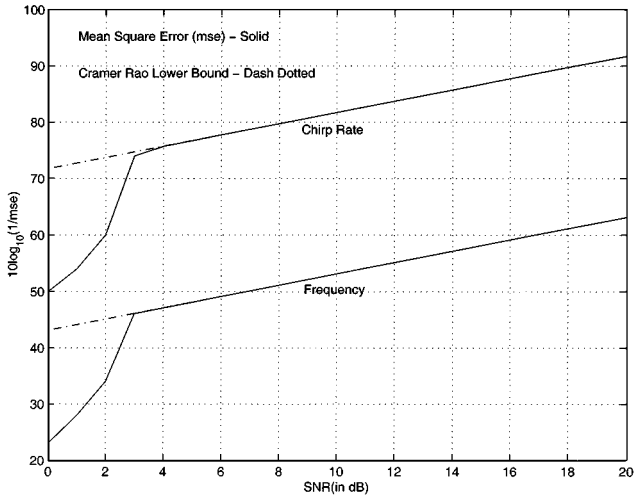

Fig. 1. Plot of true parameters and estimates for 100 realizations.

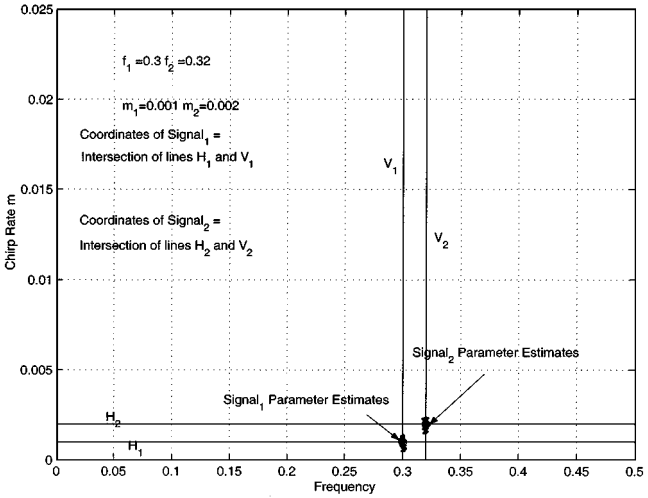

Fig. 2. Plot of mean square error versus SNR.

Next, we perform a Monte Carlo simulation for different SNRs for the same example to determine the threshold SNR at which the CRLB is no longer attained. There are 100 Monte Carlo estimations performed for each SNR from 1 to 20 dB, and the mean square error is computed for each. In Fig. 2, we plot $10\log_{10}$ (mean square error) versus the SNR. A comparison is made with the CRLB at each estimation. It can be observed that below 3 dB SNR, the CRLB is no longer attained.

VI. Conclusions

We have developed a noniterative technique to estimate the parameters of superimposed chirp signals. The important contributions of the paper are the application of the global optimization theorem, the use of importance sampling in efficiently im-

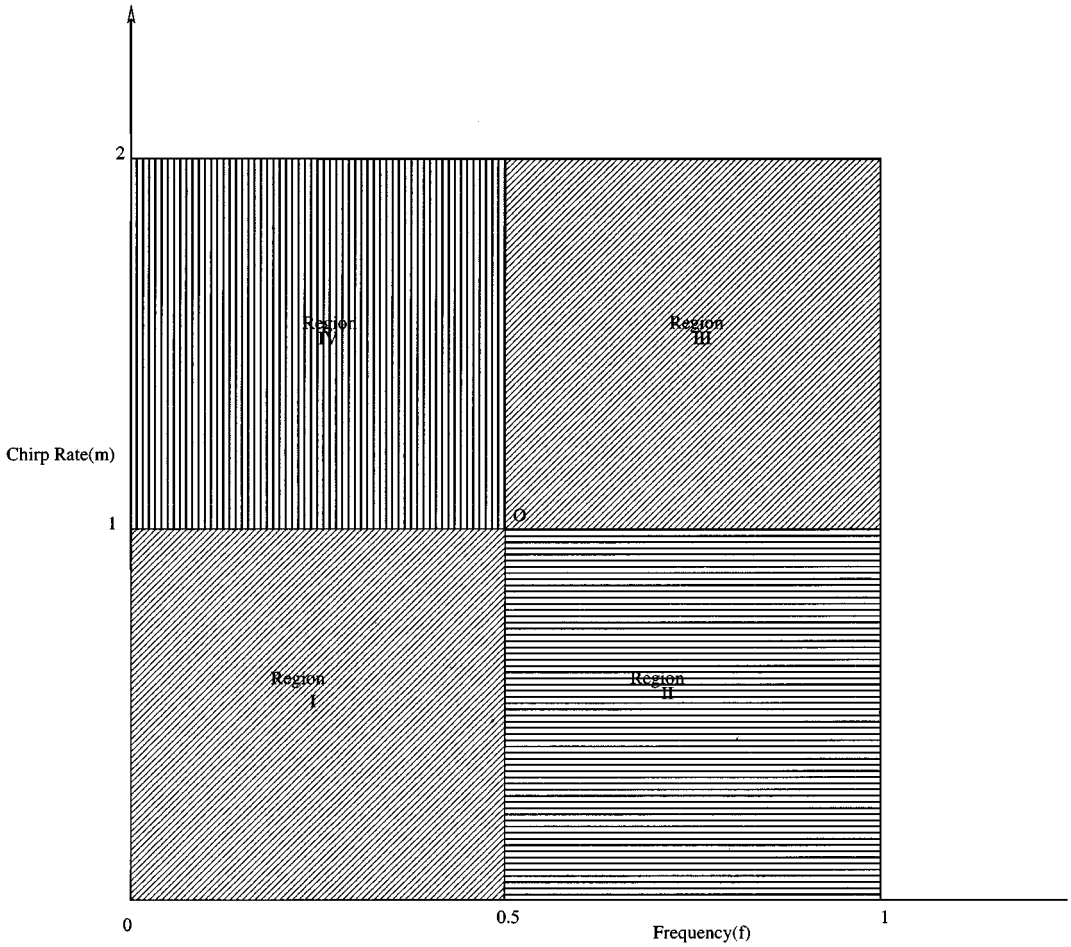

Fig. 3. Region of interest for frequency and chirp rate obtained from identifiability conditions.

plementing it, and the use of circular mean concept in reducing the computations. As a result of this, the technique does not suffer from the problem of convergence to a local maxima as is the case with iterative techniques. The threshold SNR for which the technique yields optimal estimates has also been determined. The main computation load of the algorithm is in the generation of the samples of frequency f and chirp rate m jointly. This involved sampling over a grid of size 2000 $\times$ 2000. Apart from this step the approach is not computationally intensive.

Using a direct ML approach, the computational load increases exponentially with an increase in the number of chirp signals. This is not the case using the proposed technique. If the number of chirp signals is three, then the only additional step is in the generation of the third pair (α_i, β_i). This step is computationally modest. Apart from this, the total number of realizations R is expected to increase for optimal results but not exponentially. This has been observed for a similar problem of multiple sinusoidal frequency estimation [4].

Appendix
Derivation of Region of Interest of Chirp Rate and Frequency

Since the parameterized chirp signal $s_{f,m}[n]$ has the form given by

$$s_{f,m}[n] = \exp\left(j2\pi\left(fn + \frac{m}{2}n^2 + \phi\right)\right) \tag{32}$$

the frequency parameter f is periodic with period 1, and the chirp rate m is periodic with period 2, but since the frequency and chirp rate appear in the argument of an exponential together, the range $0 \le f \le 1$ and $0 \le m \le 2$ may not be the correct allowable range that will permit identifiability. To determine the allowable range that leads to identifiability, we proceed as follows. Let $s_{f,m}[n]$ and $s_{f+f',m+m'}[n]$ be the two parametric signals, with $0 < f' < 1$ and $0 < m' < 2$. Now, it is required to determine the values of (f', m') for which the two parametric signals $s_{f,m}[n]$ and $s_{f+f',m+m'}[n]$ are the same. For

$$s_{f,m}[n] = s_{f+f',m+m'}[n]$$

or

$$\exp\left(j\left(2\pi\left(fn + \frac{m}{2}n^2\right) + \phi\right)\right) = \exp\left(j\left(2\pi\left[(f+f')n + (m+m')\frac{n^2}{2}\right] + \phi\right)\right)$$

it is required that

$$\exp\left(j2\pi\left(f'n + \frac{m'}{2}n^2\right)\right) = 1$$

or

$$f'n + \frac{m'}{2}n^2 = \text{integer for } 0 \le n \le N-1. \tag{33}$$

Now, for $n = 1$

$$f'n + \frac{m'}{2}n^2 = f' + \frac{m'}{2} = \frac{2f' + m'}{2} = \text{integer}. \quad (34)$$

Since $0 \le f' \le 1$ and $0 \le m' \le 2$

$$0 < \frac{2f' + m'}{2} < 2. \quad (35)$$

From (34) and (35)

$$\frac{2f' + m'}{2} = 1 \quad (36)$$

or

$$2f' + m' = 2. \quad (37)$$

For f' and m' satisfying (37), $f'n + (m'/2)n^2$ should be an integer for other values of n also. Thus, for $n = 2$

$$f'n + \frac{m'}{2}n^2 = f'2 + \frac{m'}{2}2^2 = 2f' + 2m' = \text{integer}. \quad (38)$$

From (37) and (38), we have

$$2f' + 2m' = 2 - m' + 2m' = 2 + m' = \text{integer} \quad (39)$$

but $2 < 2 + m' < 4$, Thus, from (39)

$$2 + m' = 3. \quad (40)$$

Thus, $m' = 1$, and $f' = 0.5$ is the only possibility. Now, we show that this choice ensures that $s_{f,m}[n] = s_{f+f',m+m'}[n]$. For $m' = 1$ and $f' = 0.5$

$$\begin{aligned} &s_{f+f',m+m'}[n] \\ &\quad = \exp\left(j2\pi\left(fn + \frac{m}{2}n^2\right)\right)\exp\left(j2\pi\left(0.5n + \frac{1}{2}n^2\right)\right) \\ &\quad = \exp\left(j2\pi\left(fn + \frac{m}{2}n^2\right)\right)\exp\left(j2\pi\,\frac{n(n+1)}{2}\right). \end{aligned}$$

Since $\exp(j2\pi(n(n+1)/2)) = 1$

$$s_{f,m}[n] = s_{f+f',m+m'}[n].$$

As a result of this, the identifiability conditions of the problem are determined by referring to Fig. 3. The value of $s_{f,m}[n]$ in regions I and III are symmetric with respect to the point O. The regions II and IV do not exhibit any form of symmetry and, thus, need to be examined individually. Thus, the regions that need to be examined for the frequency and chirp rate jointly are regions I, II, and IV. However, due to constraints imposed by narrowband chirp signal assumption, region IV may be excluded. This is because the chirp rate parameter is very small and is in the range of 10^{-3} to 10^{-4}. This can be understood by considering the digital chirp signal to have been obtained by sampling a narrowband analog chirp signal at a rate higher than Nyquist rate. The equivalent digital rate reduces to a very low quantity in the order of 10^{-3}. Thus, the regions of interest become regions I and II.

REFERENCES

[1] S. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Englewood Cliffs, NJ: Prentice-Hall, 1993.

[2] M. Pincus, "A closed form solution for certain programming problems," *Oper. Res.*, pp. 690–694, 1962.

[3] P. M. Djuric and S. M. Kay, "Parameter estimation of chirp signals," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 38, pp. 228–2126, Dec. 1990.

[4] S. Kay and S. Saha, "Mean likelihood frequency estimation," *IEEE Trans. Signal Processing*, vol. 48, pp. 1937–1946, July 2000.

[5] ——, "Design of sparse linear arrays by Monte Carlo importance sampling," in *Proc. Oceans MTS/IEEE*, Providence, RI, Sept. 2000.

[6] T. J. Abatzoglou, "Fast maximum likelihood joint estimation of frequency and frequency rate," *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES-13, pp. 708–715, Nov. 1986.

[7] C. S. Ramalingam, "Nonstationary signal analysis," Ph.D. dissertation, Univ. Rhode Island, Kingston, 1995.

[8] L. Stewart, "Bayesian analysis using Monte Carlo integration: A powerful methodology for handling some difficult problems," *Amer. Statist.*, pp. 195–200, 1983.

[9] A. F. M. Smith and A. Gelfand, "Bayesian statistics without tears: A sampling-resampling framework," *Amer. Statist. Assoc.*, 1992.

[10] K. V. Mardia, *Statistics of Directional Data*. London, U.K., New York: Academic, 1972.

[11] J. M. Bernardo and A. F. M. Smith, *Bayesian Theory*. New York: Wiley, 1994.

[12] R. M. Liang and K. S. Arun, "Parameter estimation for superimposed chirp signals," in *Proc. Int. Conf. Acoust., Speech, Signal Process.*, 1992, pp. 273–276.

[13] K. S. Arun and R. M. Liang, "A closed form solution for instantaneous amplitude and frequency estimation: Performance bounds and applications to source localization," *Proc. SPIE*, July 1991.

[14] B. C. Lovell, P. J. Kootsookos, and R. C. Williamson, "The circular nature of discrete-time frequency estimates," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, May 1991, pp. 3369–3372.

Supratim Saha received the B.Tech. (hons.) degree in instrumentation engineering from the Indian Institute of Technology, Kharagpur, India, in 1993, the M.S. degree from Indian Institute of Science, Bangalore, in 1996, and the Ph.D. degree from University of Rhode Island, Kingston, in 2001, all in electrical engineering.

His interests are in probability theory and Monte Carlo methods and their applications in signal processing and data communications.

Steven M. 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), and *Fundamentals of Statistical Signal Processing, Vol. I: Estimation Theory* (Englewood Cliffs, NJ: Prentice-Hall, 1993), *Fundamentals of Statistical Signal 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 and on several IEEE Oceans committees. He is currently a distinguished lecturer for the IEEE Signal Processing Society.