ELSEVIER

Signal Processing 51 (1996) 105–120

SIGNAL PROCESSING

An iterative MMSE procedure for parameter estimation of damped sinusoidal signals[1]

Hsiang-Tsun Li, Petar M. Djurić*

Department of Electrical Engineering, State University of New York at Stony Brook, Stony Brook, NY 11794-2350, USA

Received 4 May 1995; revised 4 December 1995

Abstract

We propose an iterative minimum mean square error (MMSE) approach to parameter estimation of multiple exponentially damped sinusoids embedded in white Gaussian noise. The approach splits the observed signal into its constituent components and proceeds to estimate the parameters of the highest energy signal component first, followed by the next highest energy signal component, and so on. This scheme reduces the computational complexity of the overall procedure by approximating the multidimensional integrals with several two-dimensional integrals. For the evaluation of these integrals, we introduce an adaptive Gaussian procedure. We present computer simulation results that show the performance of our approach and compare it with the performance of the alternating projection and the expectation-maximization schemes.

Zusammenfassung

In dieser Arbeit wird ein iterativer, auf dem Prinzip des minimalen mittleren quadratischen Fehlers (MMSE) beruhender Ansatz für die Parameterschätzung im Fall mehrfacher exponentiell gedämpfter Sinussignale in weißem Gaußschem Rauschen vorgeschlagen. Das beobachtete Signal wird zunächst in seine Komponenten zerlegt; danach werden die Parameter der Signalkomponente mit der größten Energie geschätzt, dann jene der Signalkomponente mit der zweitgrößten Energie usw. Diese Vorgangsweise reduziert den Rechenaufwand, indem mehrdimensionale Integrale durch mehrere zweidimensionale Integrale angenähert werden. Für die Auswertung dieser Integrale wird eine adaptive Gauß-Methode vorgeschlagen. Computersimulationen dokumentieren die Leistungsfähigkeit unseres Ansatzes und vergleichen sie mit der Leistungsfähigkeit der Methode alternierender Projektionen und der Expectation-Maximization-Methode.

Résumé

Nous proposons une approche itérative par l'erreur quadratique moyenne minimum (EQMM) de l'estimation de paramètres de sinusoides multiples amorties exponentiellement et noyées dans du bruit blanc gaussien. L'approche sépare le signal observé en ses différentes composantes et estime ensuite les paramètres de la composante du signal ayant la plus haute énergie, puis celle ayant la plus haute énergie parmi les restantes, et ainsi de suite. Cette méthode réduit la complexité de calcul de la procédure totale en approximant les intégrales multidimensionnelles par plusieurs intégrales

*Corresponding author. Tel.: 516 632 8423; fax: 516 632 8494; e-mail: djuric@sbee.sunysb.edu.

[1] This work has been supported by the National Science Foundation under Award No. MIP-9506743.

PII S0165-1684(96)00035-7

bidimensionnelles. Pour évaluer ces intégrales, nous introduisons une procédure adaptative gaussienne. Nous présentons des résultats de simulations informatiques qui montrent les performances de notre approche et les comparons avec les performances des méthodes de projection alternée et de maximisation de l'espérance.

Keywords: Parameter estimation; Damped sinusoids; Minimum mean square estimation; Bayes' theory; Adaptive Gaussian quadrature

1. Introduction

In many engineering and scientific problems the observed measurements are modeled as exponentially damped sinusoids distorted by additive noise. The important parameters of these models are the sinusoidal frequencies and the damping factors because they reveal significant information related to the phenomenon under investigation. A difficult but interesting problem has always been the estimation of these parameters. There have been a variety of approaches for estimation, most of them revolving around the maximum likelihood (ML) principle or linear prediction. As this is a highly nonlinear estimation problem, most of the methods in the literature maximize the likelihood function iteratively by employing variations of the expectation-maximization (EM) [2] or alternating projections (AP) schemes [12].

At high signal-to-noise ratios (SNRs) the ML-based methods perform equally well. However, these methods require a search over the frequency-damping factor plane to locate the global maximum or minimum, and their accuracy depends on the used grid resolution. Moreover, most of the iterative algorithms are sensitive to the choice of initial estimates, particularly when there are closely spaced signal components in the frequency domain, and the SNR is low. Recently, some efficient techniques based on linear prediction were proposed [5, 10], but they require higher SNRs to achieve the Cramer–Rao lower bounds (CRLBs).

In contrast to the above approaches, we focus on a procedure that yields MMSE estimates. It is well known that MMSE estimation requires multidimensional integrations. Motivated by the orthogonality of sinusoids, we reduce the multidimensional integrations into a set of two-dimensional integrations and propose an iterative approach similar in philosophy to the EM and AP. Instead of estimating all of the parameters simultaneously, we estimate the parameters of one signal at a time while considering the remaining parameters 'known' and equal to their current estimates. We also propose an efficient numerical approach, adaptive Gaussian quadrature (AGQ), for performing the numerical integrations. The computational load of our approach is low, and its performance is excellent. Similarly to the EM and AP, it is also sensitive to initializations. To desensitize the procedure, we have paid special attention to its initialization. The initial set of estimates are found by a two-step scheme that is based on notch periodograms. First the nonlinear parameters of the strongest signal component are estimated, followed by its annihilation and parameter estimation of the second strongest signal component, and so on. For initialization of the weakest component, we apply multimodal functions.

The paper is organized as follows. In Section 2, we state the signal model and formulate the problem. The parameter initialization is discussed in Section 3, and the MMSE procedure in Section 4. In Section 5, we provide simulation results and comparisons with the AP and EM methods. For the actual implementation of the estimator, we introduce an AGQ technique, which is outlined in Appendix A. Finally, in the last section we make some concluding remarks.

2. Problem formulation

A vector of measurements, $\mathbf{y}$, is observed whose components are samples of q complex sinusoids corrupted by additive white Gaussian noise, i.e.,

$$y[n] = \sum_{i=1}^{q} a_i \exp(\alpha_i n + \mathrm{j}2\pi f_i n) + w[n], \quad \alpha_i \leqslant 0, \tag{1}$$

where $n = 0, 1, 2, \ldots, N-1$; a_i, α_i and f_i, $i = 1, 2, \ldots, q$, are the complex amplitude, damping factor and frequency of the ith sinusoid, and $\mathrm{j} = \sqrt{-1}$. The random samples $w[n]$ are complex, independent and identically distributed. Moreover, the real and imaginary components of $w[n]$ are identically normally distributed with zero mean and unknown variance $\sigma^2/2$. The number of sinusoids q is assumed known.

From (1), the data model can be rewritten in a vector–matrix form according to

$$\boldsymbol{y} = \boldsymbol{H}(\boldsymbol{f}, \boldsymbol{\alpha})\boldsymbol{a} + \boldsymbol{w}, \tag{2}$$

where $\boldsymbol{H}$ is an $N \times q$ matrix whose columns span the signal space, $\boldsymbol{a}$ is a vector of complex amplitudes, and $\boldsymbol{w}$ is a noise vector, where $\boldsymbol{w} \sim \mathscr{CN}(\mathbf{0}, \sigma^2 \boldsymbol{I})$. The matrix $\boldsymbol{H}$ is defined by

$$\boldsymbol{H}(\boldsymbol{f}, \boldsymbol{\alpha}) = [\boldsymbol{d}(f_1, \alpha_1)\ \boldsymbol{d}(f_2, \alpha_2)\ \ldots\ \boldsymbol{d}(f_q, \alpha_q)],$$

where for $i = 1, 2, \ldots, q$

$$\boldsymbol{d}(f_i, \alpha_i)^{\mathrm{T}} = [1\ \ \exp(\alpha_i + \mathrm{j}2\pi f_i)\ \ \ldots\ \ \exp((N-1)\alpha_i + \mathrm{j}(N-1)2\pi f_i)], \quad \alpha_i \leqslant 0.$$

(In the remaining part of the paper, for notational convenience, whenever there is no ambiguity, $d(f, \alpha)$ and $\boldsymbol{H}(\boldsymbol{f}, \boldsymbol{\alpha})$ will be denoted as $\boldsymbol{d}$ and $\boldsymbol{H}$, respectively.) Given the observations $\boldsymbol{y}$ and the number of sinusoids q, the objective is to estimate the nonlinear parameters f_i and α_i, $i = 1, 2, \ldots, q$, of the signals. Note that if $\boldsymbol{\alpha}$ is a zero vector in (2), this becomes an undamped sinusoidal model, and it can be solved by the proposed algorithm as a special case.

3. Parameter initialization

To start the iterative algorithm, we have to find the initial approximate values of α_i and f_i for $i = 1, 2, \ldots, q$. First we estimate the initial values of α and f of the strongest sinusoid using the periodogram. We find the frequency estimate first, and, subsequently, the estimate of the damping factor without searching the whole f–α plane. Once they are obtained, we proceed with the second strongest sinusoid, and so on. This method of estimating the frequency is similar to the alternating notch-periodogram algorithm (ANPA) initialization for undamped sinusoids in [6, 7], except that our model is more general since it includes damping factors.

Before we describe the procedure for finding the initial estimates of the damping factors and frequencies, we define the notch periodogram of the data $\boldsymbol{y}$ at f with a notch set at $(\boldsymbol{f}, \boldsymbol{\alpha})$ by using the following result of orthogonal subspace decomposition [4]:

$$P_{\mathrm{n}}(f; \boldsymbol{f}, \boldsymbol{\alpha}) = \boldsymbol{y}^{\mathrm{H}}(\boldsymbol{P}(f, 0; \boldsymbol{f}, \boldsymbol{\alpha}) - \boldsymbol{P}(\boldsymbol{f}, \boldsymbol{\alpha}))\boldsymbol{y}, \tag{3}$$

where $\boldsymbol{P}(\boldsymbol{f}, \boldsymbol{\alpha})$ is a projection operator defined by $\boldsymbol{P}(\boldsymbol{f}, \boldsymbol{\alpha}) = \boldsymbol{H}(\boldsymbol{H}^{\mathrm{H}}\boldsymbol{H})^{-1}\boldsymbol{H}^{\mathrm{H}}$, with the superscript H denoting conjugate transposition.

We exploit (3) in obtaining the initial estimates of the unknown parameters. The frequency estimate of the first sinusoid is found from the periodogram of the data. To find the initial estimate of the first damping factor, we proceed as follows. Let $\boldsymbol{s}_1$ denote the first signal component given by

$$\boldsymbol{s}_1 = a_1 \boldsymbol{d}(f_1, \alpha_1).$$

Then we can show that

$$\boldsymbol{s}_1^{\mathrm{H}}(\boldsymbol{P}(f_1, 0; f_1, \alpha_1) - \boldsymbol{P}(f_1, \alpha_1))\boldsymbol{s}_1 = 0, \tag{4}$$

which implies that α can be estimated by minimizing the following function with respect to α:

$$P_{\mathrm{n}}(\hat{f}_1; \hat{f}_1, \alpha) = \boldsymbol{s}_1^{\mathrm{H}}(\boldsymbol{P}(\hat{f}_1, 0; \hat{f}_1, \alpha) - \boldsymbol{P}(\hat{f}_1, \alpha))\boldsymbol{s}_1, \tag{5}$$

where $P_{\mathrm{n}}(\hat{f}_1; \hat{f}_1, \alpha)$ is the notch periodogram of $\boldsymbol{s}_1$ at the estimated frequency $\hat{f}_1$ with a notch at $[\hat{f}_1, \alpha]$. The minimization of (5) yields a different result from the maximum projection method which searches for the α that corresponds to the minimum value of the orthogonal projection operator $\|\boldsymbol{P}^{\perp}(\hat{f}_1, \alpha)\boldsymbol{s}_1\|^2 = \|(\boldsymbol{I} - \boldsymbol{P}(\hat{f}_1, \alpha))\boldsymbol{s}_1\|^2$. If the observed data represent a single component, the difference between these two methods is insignificant.

The advantage of our method becomes distinct when the data $\boldsymbol{y}$ are composed of multiple damped sinusoids. Suppose for example that $\boldsymbol{y}$ consists of two signals, $\boldsymbol{s}_1$ and $\boldsymbol{s}_2$, and that $\boldsymbol{s}_2$ is the weaker one. Then the projection of $\boldsymbol{s}_2$ onto the space spanned by $\boldsymbol{P}^{\perp}(f_1, \alpha_1)$ is larger than the projection onto the notch periodogram, or

$$\begin{aligned}\|\boldsymbol{P}^{\perp}(f_1,\alpha_1)\boldsymbol{s}_2\|^2 &= \boldsymbol{s}_2^{\mathrm{H}}\boldsymbol{s}_2 - \boldsymbol{s}_2^{\mathrm{H}}\boldsymbol{P}(f_1,\alpha_1)\boldsymbol{s}_2 \\ &\geqslant \boldsymbol{s}_2^{\mathrm{H}}\boldsymbol{P}(f_1,0;\, f_1,\alpha_1)\boldsymbol{s}_2 \\ &\quad - \boldsymbol{s}_2^{\mathrm{H}}\boldsymbol{P}(f_1,\alpha_1)\boldsymbol{s}_2 \\ &= P_{\mathrm{n}}(f_1;\, f_1,\alpha_1), \end{aligned} \tag{6}$$

where $P_{\mathrm{n}}(f_1; f_1, \alpha_1)$ is the notch periodogram of the vector $\boldsymbol{s}_2$ at f_1 with respect to the notch set $[f_1, \alpha_1]$. Thus, compared with the periodogram, the notch periodogram can reduce the interference from other components in the parameter estimation of individual components. Later we will present the initialization performance and compare it with the performance of the classical maximum projection approach.

In summary, the initialization is implemented according to the following scheme. To begin, we search for the frequency $\hat{f}_1^{(0)}$ of the strongest component in the periodogram, then estimate $\hat{\alpha}_1^{(0)}$ that corresponds to the minimum of the notch periodogram (5) at $\hat{f}_1^{(0)}$ with respect to the notch set $[\hat{f}_1^{(0)}, \alpha]$. Fixing the initial estimate at $[\hat{f}_1^{(0)}, \hat{\alpha}_1^{(0)}]$, the initial parameter estimates of the second sinusoid are obtained by the notch periodogram with a notch at $[\hat{f}_1^{(0)}, \hat{\alpha}_1^{(0)}]$. Continuing in this fashion, at the ith iteration we find the initial estimates from

$$\begin{aligned}\hat{f}_i^{(0)} &= \arg\max_{f} P_{\mathrm{n}}(f_i;\, \hat{\boldsymbol{\theta}}_{i-1}), \\ \hat{\alpha}_i^{(0)} &= \arg\min_{\alpha} P_{\mathrm{n}}(\hat{f}_i^{(0)};\, \hat{\boldsymbol{\theta}}_{i-1}, \hat{f}_i^{(0)}, \alpha),\end{aligned} \tag{7}$$

where $P_{\mathrm{n}}(f; \hat{\boldsymbol{\theta}}_i)$ is the notch periodogram of the data vector $\boldsymbol{y}$ at f with respect to the notch set $\hat{\boldsymbol{\theta}}_i$, $\hat{\boldsymbol{\theta}}_i = [\hat{f}_1^{(0)}, \hat{\alpha}_1^{(0)}, \ldots, \hat{f}_l^{(0)}, \hat{\alpha}_l^{(0)}, \ldots, \hat{f}_i^{(0)}, \hat{\alpha}_i^{(0)}]$ with i notches, and $\boldsymbol{\theta}_0 = \emptyset$ (empty set). The subscript denotes the signal component number, and the superscript the iteration number.

Since the initialization algorithm represents a search along the frequency axis followed by a search along the damping factor axis instead of the whole f–α 2D plane, the estimated parameters do not necessarily correspond to the maximum of $\boldsymbol{y}^{\mathrm{H}}\boldsymbol{P}(\boldsymbol{f}, \boldsymbol{\alpha})\boldsymbol{y}$. In particular, for low SNRs or large damping factors, this initialization algorithm may yield estimates that result from local maxima in the f–α plane. Here we propose a scheme that alleviates this problem and improves the overall performance of the algorithm. It is based on the concept of integration of multimodal functions from [11]. It uses M_q initialization points, $(\hat{\boldsymbol{f}}_q^{(0)}, \hat{\boldsymbol{\alpha}}_q^{(0)})$, in the f–α plane for the weakest signal, where the lth initialization point is given by

$$\begin{aligned}\hat{f}_{(q,l)}^{(0)} &= f_{q+l-1}, \\ \hat{\alpha}_{(q,l)}^{(0)} &= \arg\min_{\alpha} P_{\mathrm{n}}(f_{q+l-1};\, \hat{\boldsymbol{\theta}}_{q-1}, f_{q+l-1}, \alpha),\end{aligned} \tag{8}$$

with f_{q+l-1} corresponding to the $(q+l-1)$th strongest peak of $P_{\mathrm{n}}(f; \hat{\boldsymbol{\theta}}_{q-1})$, and $l = 1, 2, \ldots, M_q$.

4. MMSE estimator

In this section we propose an iterative method similar in philosophy to the EM and AP approaches to obtain the MMSE of the frequencies and damping factors. To avoid integration of high multidimensional integrals over the space of unknown parameters, we 'decouple' the integrals by assuming that all the parameters except the parameters of one signal are known and equal to their current estimates. This will allow usage of two-dimensional integrals only.

The marginalized likelihood function can be expressed as

$$p(\boldsymbol{y}\,|\,\boldsymbol{f}, \boldsymbol{\alpha}) = \int_{\{\sigma\}} \int_{\{\boldsymbol{a}\}} p(\boldsymbol{y}\,|\,\boldsymbol{a}, \sigma, \boldsymbol{f}, \boldsymbol{\alpha})\, p(\boldsymbol{a}, \sigma\,|\,\boldsymbol{f}, \boldsymbol{\alpha})\, \mathrm{d}\boldsymbol{a}\, \mathrm{d}\sigma, \tag{9}$$

where $\{\sigma\}$ and $\{\boldsymbol{a}\}$ are the parameter spaces of σ and $\boldsymbol{a}$, respectively. For the prior probability density function (p.d.f) $p(\boldsymbol{a}, \sigma\,|\,\boldsymbol{f}, \boldsymbol{\alpha})$, we assume that it is noninformative and defined by Jeffrey's invariance principle [1]. Based on this principle, we have

$$p(\boldsymbol{a}, \sigma\,|\,\boldsymbol{f}, \boldsymbol{\alpha}) \propto \frac{|\boldsymbol{H}^{\mathrm{H}}(\boldsymbol{f}, \boldsymbol{\alpha})\boldsymbol{H}(\boldsymbol{f}, \boldsymbol{\alpha})|}{\sigma}, \tag{10}$$

where $\boldsymbol{H}(\boldsymbol{f}, \boldsymbol{\alpha})$ is given by (2). After inserting (10) into (9) and integrating out analytically the amplitudes and the noise variance σ^2, we obtain [9]

$$p(\boldsymbol{y}\,|\,\boldsymbol{f}, \boldsymbol{\alpha}) = \frac{c}{(\boldsymbol{y}^{\mathrm{H}}\boldsymbol{P}^{\perp}(\boldsymbol{f}, \boldsymbol{\alpha})\boldsymbol{y})^{N-m}}, \tag{11}$$

where c is the normalizing constant, and

$$\boldsymbol{P}^{\perp}(\boldsymbol{f},\boldsymbol{\alpha}) = \boldsymbol{I} - \boldsymbol{H}(\boldsymbol{f},\boldsymbol{\alpha})(\boldsymbol{H}^{\mathrm{H}}(\boldsymbol{f},\boldsymbol{\alpha})\boldsymbol{H}(\boldsymbol{f},\boldsymbol{\alpha}))^{-1}\boldsymbol{H}^{\mathrm{H}}(\boldsymbol{f},\boldsymbol{\alpha}). \tag{12}$$

By Bayes's theorem, the joint p.d.f. of the frequencies and damping factors is given by

$$p(\boldsymbol{f},\boldsymbol{\alpha}|\boldsymbol{y}) = \frac{p(\boldsymbol{y}|\boldsymbol{f},\boldsymbol{\alpha})p(\boldsymbol{f},\boldsymbol{\alpha})}{\int_{\{f\}}\int_{\{\alpha\}} p(\boldsymbol{y}|\boldsymbol{f},\boldsymbol{\alpha})p(\boldsymbol{f},\boldsymbol{\alpha})\,\mathrm{d}\boldsymbol{\alpha}\,\mathrm{d}\boldsymbol{f}}, \tag{13}$$

with $p(\boldsymbol{f},\boldsymbol{\alpha})$ being the prior p.d.f. of $(\boldsymbol{f},\boldsymbol{\alpha})$ for which we choose

$$p(\boldsymbol{f},\boldsymbol{\alpha}) \propto \begin{cases} C \neq 0, & f_i \neq f_k,\ \forall i \neq k,\ f_i \in \mathscr{F},\ \alpha_i \in \mathscr{A}, \\ 0, & \text{otherwise,} \end{cases}$$

because $\boldsymbol{f}$ and $\boldsymbol{\alpha}$ (by assumption) have finite supports. So, the MMSE of $(\boldsymbol{f},\boldsymbol{\alpha})$ are obtained by

$$\hat{\boldsymbol{f}} = \int_{\{f\}}\int_{\{\alpha\}} \boldsymbol{f}\,p(\boldsymbol{f},\boldsymbol{\alpha}|\boldsymbol{y})\,\mathrm{d}\boldsymbol{\alpha}\,\mathrm{d}\boldsymbol{f} \tag{14}$$

and

$$\hat{\boldsymbol{\alpha}} = \int_{\{f\}}\int_{\{\alpha\}} \boldsymbol{\alpha}\,p(\boldsymbol{f},\boldsymbol{\alpha}|\boldsymbol{y})\,\mathrm{d}\boldsymbol{\alpha}\,\mathrm{d}\boldsymbol{f}\ . \tag{15}$$

The integrals in (14) and (15) are $2q$-dimensional integrals, and, as such, would require reliance on numerical techniques for high-dimensional integration. This is computationally expensive and often precludes the use of the MMSE estimator in practical applications. Since the a posteriori p.d.f.'s of the parameters of each sinusoid, $p(f_i,\alpha_i|\boldsymbol{y})$ for $i=1,2,\ldots,q$, are highly concentrated functions around (f_i,α_i) for $i=1,2,\ldots,q$, respectively, when $\boldsymbol{y}$ is observed, and due to the orthogonality of sinusoids, imprecise knowledge of all the frequencies and damping factors except f_i and α_i, has little effect on $p(f_i,\alpha_i|\boldsymbol{y})$.

To be more specific, let k denote the current iteration, and $\hat{f}_i^{(k)}$ and $\hat{\alpha}_i^{(k)}$ the current estimates of f_i and α_i, respectively, $i=1,2,\ldots,q$. Then, if we approximate the posterior density $p(\boldsymbol{f},\boldsymbol{\alpha}|\boldsymbol{y})$ by

$$p(\boldsymbol{f},\boldsymbol{\alpha}|\boldsymbol{y}) \simeq p(f_i,\alpha_i|\boldsymbol{y},\hat{\boldsymbol{f}}_{(-i)}^{(k)},\hat{\boldsymbol{\alpha}}_{(-i)}^{(k)}) \times \prod_{\substack{l=1\\ l\neq i}}^{q} \delta(f_l - \hat{f}_l^{(k)},\alpha_l - \hat{\alpha}_l^{(k)}), \tag{16}$$

our $2q$-dimensional integral would reduce to q individual two-dimensional integrals. In (16), $\hat{\boldsymbol{f}}_{(-i)}^{(k)}$ and $\hat{\boldsymbol{\alpha}}_{(-i)}^{(k)}$ denote the estimates at the kth iteration of all the frequencies and damping factors except the ones of the ith sinusoid. The two-dimensional integrals are MMSE estimators of the form

$$\hat{f}_i^{(k)} = \int_{\{f_i\}}\int_{\{\alpha_i\}} f_i\,p(f_i,\alpha_i|\boldsymbol{y},\hat{\boldsymbol{f}}_{(-i)}^{(k)},\hat{\boldsymbol{\alpha}}_{(-i)}^{(k)})\,\mathrm{d}\alpha_i\,\mathrm{d}f_i \tag{17}$$

and

$$\hat{\alpha}_i^{(k)} = \int_{\{f_i\}}\int_{\{\alpha_i\}} \alpha_i\,p(f_i,\alpha_i|\boldsymbol{y},\hat{\boldsymbol{f}}_{(-i)}^{(k)},\hat{\boldsymbol{\alpha}}_{(-i)}^{(k)})\,\mathrm{d}\alpha_i\,\mathrm{d}f_i. \tag{18}$$

Thus, if $\hat{\boldsymbol{f}}_{(-i)}^{(k)}$ and $\hat{\boldsymbol{\alpha}}_{(-i)}^{(k)}$ approach $\boldsymbol{f}_{(-i)}$ and $\boldsymbol{\alpha}_{(-i)}$, then $E[f_i|\hat{\boldsymbol{f}}_{(-i)}^{(k)},\boldsymbol{y}]$ and $E[\alpha_i|\hat{\boldsymbol{\alpha}}_{(-i)}^{(k)},\boldsymbol{y}]$ approach $E[f_i|\boldsymbol{y}]$ and $E[\alpha_i|\boldsymbol{y}]$, respectively. Provided $\hat{\boldsymbol{f}}_{(-i)}^{(k)}$ and $\hat{\boldsymbol{\alpha}}_{(-i)}^{(k)}$ converge to the MMSE estimates $\hat{\boldsymbol{f}}_{(-i)\mathrm{MMSE}}$ and $\hat{\boldsymbol{\alpha}}_{(-i)\mathrm{MMSE}}$, we get $\hat{\boldsymbol{f}}_{\mathrm{MMSE}}$ and $\hat{\boldsymbol{\alpha}}_{\mathrm{MMSE}}$ iteratively rather than by performing a $2q$-dimensional integration. After substituting the density function (13) into the MMSE estimators, we have

$$\hat{f}_i^{(k)} = \frac{\int_{\{f_i\}}\int_{\{\alpha_i\}} f_i\,p(f_i,\alpha_i,\hat{\boldsymbol{f}}_{(-i)}^{(k)},\hat{\boldsymbol{\alpha}}_{(-i)}^{(k)}|\boldsymbol{y})\,\mathrm{d}\alpha_i\,\mathrm{d}f_i}{\int_{\{f_i\}}\int_{\{\alpha_i\}} p(f_i,\alpha_i,\hat{\boldsymbol{f}}_{(-i)}^{(k)},\hat{\boldsymbol{\alpha}}_{(-i)}^{(k)}|\boldsymbol{y})\,\mathrm{d}\alpha_i\,\mathrm{d}f_i} = \frac{\int\int \frac{f_i}{(\boldsymbol{y}^{\mathrm{H}}\boldsymbol{P}^{\perp}(\boldsymbol{f},\boldsymbol{\alpha})\boldsymbol{y})^{N-q}}\,\mathrm{d}\alpha_i\,\mathrm{d}f_i}{\int\int \frac{1}{(\boldsymbol{y}^{\mathrm{H}}\boldsymbol{P}^{\perp}(\boldsymbol{f},\boldsymbol{\alpha})\boldsymbol{y})^{N-q}}\,\mathrm{d}\alpha_i\,\mathrm{d}f_i} \tag{19}$$

and

$$\hat{\alpha}_i^{(k)} = \frac{\int_{\{f_i\}}\int_{\{\alpha_i\}} \alpha_i\,p(f_i,\alpha_i,\hat{\boldsymbol{f}}_{(-i)}^{(k)},\hat{\boldsymbol{\alpha}}_{(-i)}^{(k)}|\boldsymbol{y})\,\mathrm{d}\alpha_i\,\mathrm{d}f_i}{\int_{\{f_i\}}\int_{\{\alpha_i\}} p(f_i,\alpha_i,\hat{\boldsymbol{f}}_{(-i)}^{(k)},\hat{\boldsymbol{\alpha}}_{(-i)}^{(k)}|\boldsymbol{y})\,\mathrm{d}\alpha_i\,\mathrm{d}f_i} = \frac{\int\int \frac{\alpha_i}{(\boldsymbol{y}^{\mathrm{H}}\boldsymbol{P}^{\perp}(\boldsymbol{f},\boldsymbol{\alpha})\boldsymbol{y})^{N-q}}\,\mathrm{d}\alpha_i\,\mathrm{d}f_i}{\int\int \frac{1}{(\boldsymbol{y}^{\mathrm{H}}\boldsymbol{P}^{\perp}(\boldsymbol{f},\boldsymbol{\alpha})\boldsymbol{y})^{N-q}}\,\mathrm{d}\alpha_i\,\mathrm{d}f_i}. \tag{20}$$

In (19) and (20) the derivation of the matrix $\boldsymbol{P}^{\perp}(\boldsymbol{f},\boldsymbol{\alpha})$, which is a matrix function of $(\boldsymbol{f},\boldsymbol{\alpha})$, consumes much time. Fortunately, there is only one element (f_i,α_i) in the $(\boldsymbol{f},\boldsymbol{\alpha})$ vector that changes values, which allows us to simplify the calculation by using (3), or

$$\boldsymbol{P}^{\perp}(f_i,\alpha_i,\boldsymbol{f}_{(-i)},\boldsymbol{\alpha}_{(-i)}) = \boldsymbol{P}^{\perp}(\boldsymbol{f}_{(-i)},\boldsymbol{\alpha}_{(-i)}) + \boldsymbol{P}_v^{\perp}(f_i,\alpha_i,\boldsymbol{f}_{(-i)},\boldsymbol{\alpha}_{(-i)}) - \boldsymbol{I}, \tag{21}$$

where

$$v = P^{\perp}(f_{(-i)}, \alpha_{(-i)}) d(f_i, \alpha_i)$$

and

$$P_v^{\perp}(f_i, \alpha_i, f_{(-i)}, \alpha_{(-i)}) = I - v(v^H v)^{-1} v^H.$$

It is clear that $P^{\perp}(f_{(-i)}, \alpha_{(-i)})$ in (19) and (20) is a constant, and that there is only one inversion of a scalar in $P_a^{\perp}(f_i, \alpha_i, f_{(-i)}, \alpha_{(-i)})$ instead of the original $q \times q$ matrix inversion. This improves significantly the computational efficiency of our algorithm.

To enhance the computational efficiency even more, for the integration we propose an adaptive Gaussian quadrature technique (AGQ). This technique is based on dividing the support of the integrand into subintervals for which we use GQ formulae corresponding to different number of terms. The details of the technique are given in Appendix A.

5. Simulation results

In this section, we demonstrate the performance of our algorithm by some simulation results. We considered four examples where the data represent two or three exponentially damped sinusoids in additive complex Gaussian noise, as defined in Section 2. The length of the observed data was $N = 25$, and the amplitudes of the sinusoids were $a_i = 1$. The data in the first experiment were generated by

$$y[n] = \exp(-0.2n + j2\pi 0.42n) + \exp(-0.1n + j2\pi 0.52n) + w[n].$$

In the second experiment, the damping factors were $\alpha_1 = -0.4$ and $\alpha_2 = -0.07$, and we used

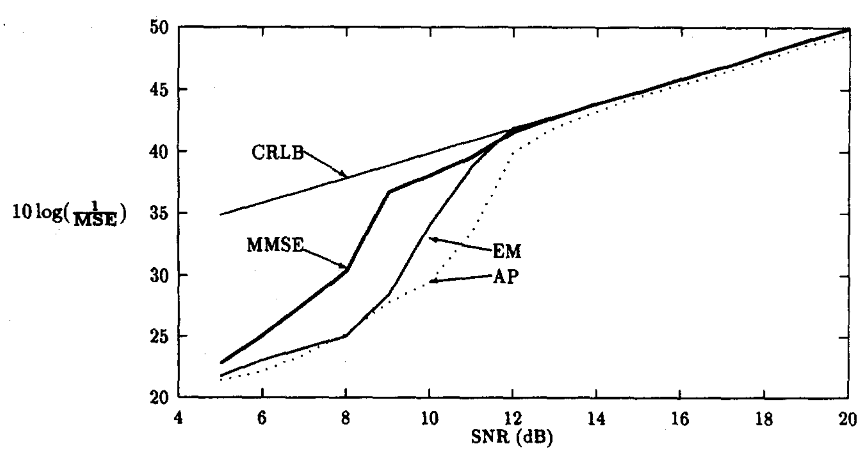
$$y[n] = \exp(-0.4n + j2\pi 0.42n) + \exp(-0.07n + j2\pi 0.52n) + w[n].$$

In the third experiment, the damping factors were the same as in the first one, but the frequencies were $f_1 = 0.46$ and $f_2 = 0.5$ to simulate a scenario of two damped sinusoids closely spaced in the frequency domain. Thus, $y[n]$ was obtained from

$$y[n] = \exp(-0.2n + j2\pi 0.46n) + \exp(-0.1n + j2\pi 0.5n) + w[n].$$

Finally, in the last experiment the data represent three damped sinusoids and were generated according to

$$y[n] = \exp(-0.2n + j2\pi 0.46n) + \exp(-0.1n + j2\pi 0.5n) + \exp(-0.1n + j2\pi 0.68n) + w[n].$$



(a) $f_1 = 0.42$

Fig. 1. Performance comparison of the MMSE, EM and AP for $q = 2$, (a) $f_1 = 0.42$ (experiment 1).

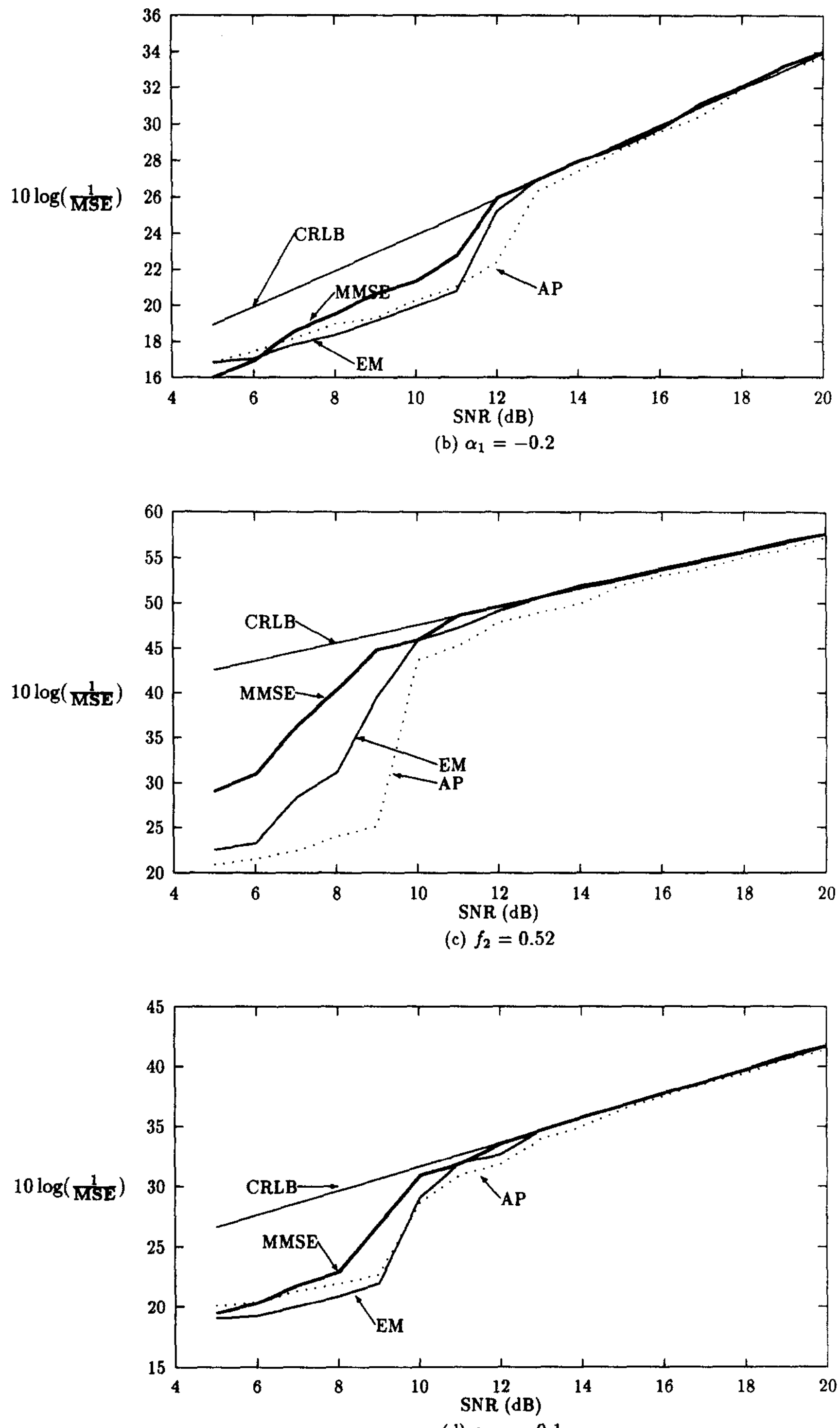


Fig. 1. (b) $\alpha_1 = -0.2$, (c) $f_2 = 0.52$, (d) $\alpha_2 = -0.1$.

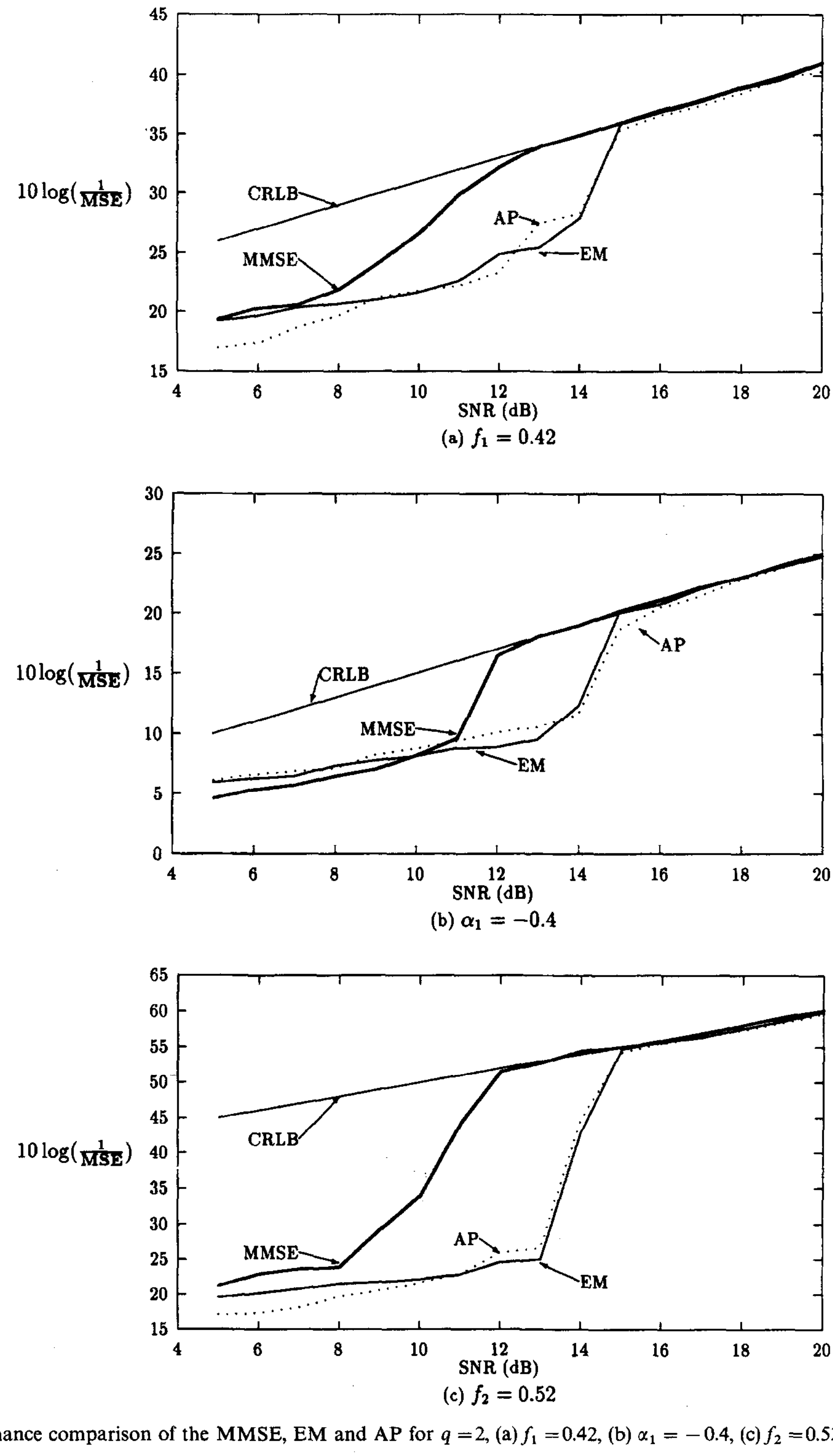


Fig. 2. Performance comparison of the MMSE, EM and AP for $q = 2$, (a) $f_1 = 0.42$, (b) $\alpha_1 = -0.4$, (c) $f_2 = 0.52$ (experiment 2).

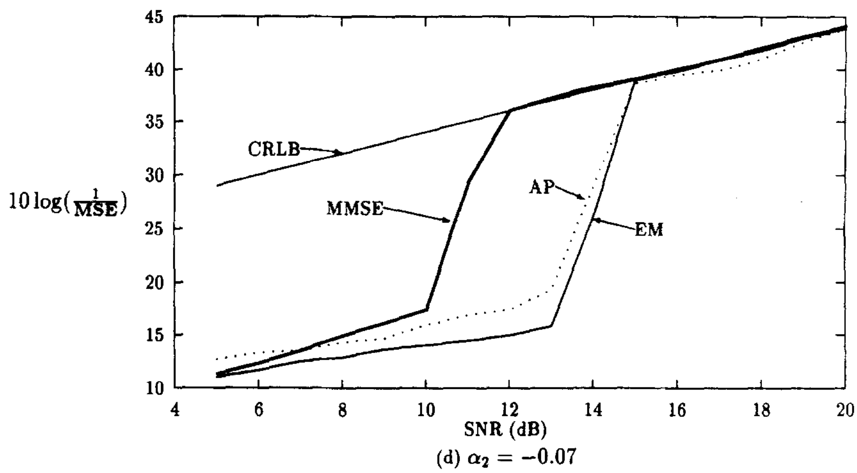


Fig. 2. (d) $\alpha_2 = -0.07$.

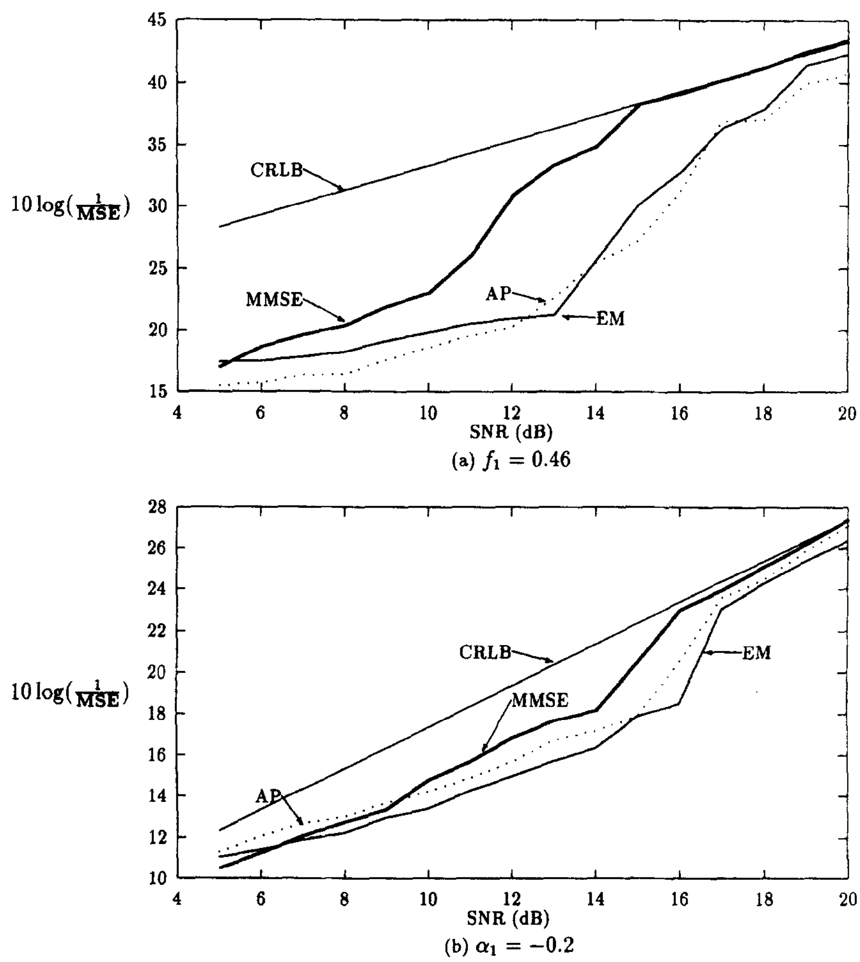
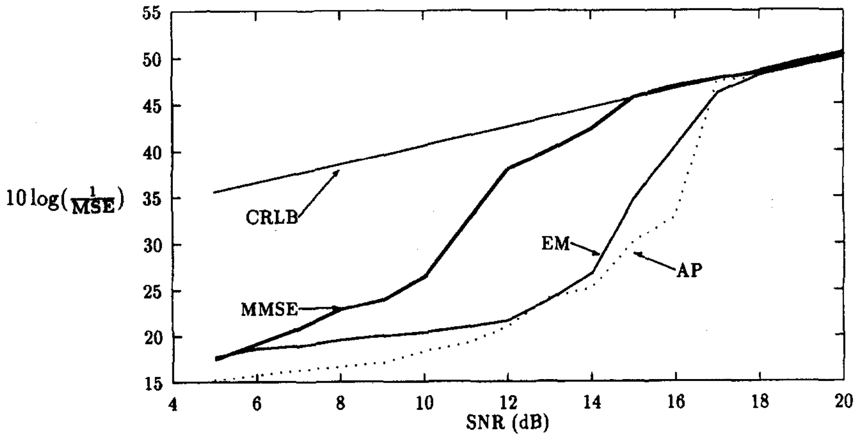
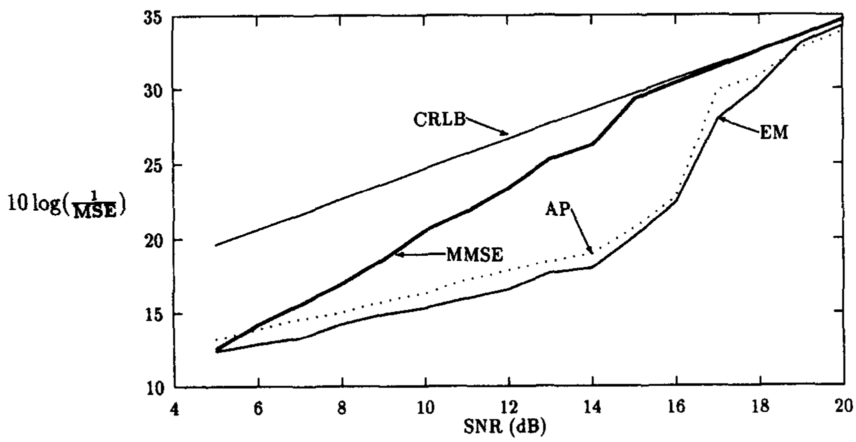


Fig. 3. Performance comparison of the MMSE, EM and AP for $q = 2$, (a) $f_1 = 0.46$, (b) $\alpha_1 = -0.2$ (experiment 3).

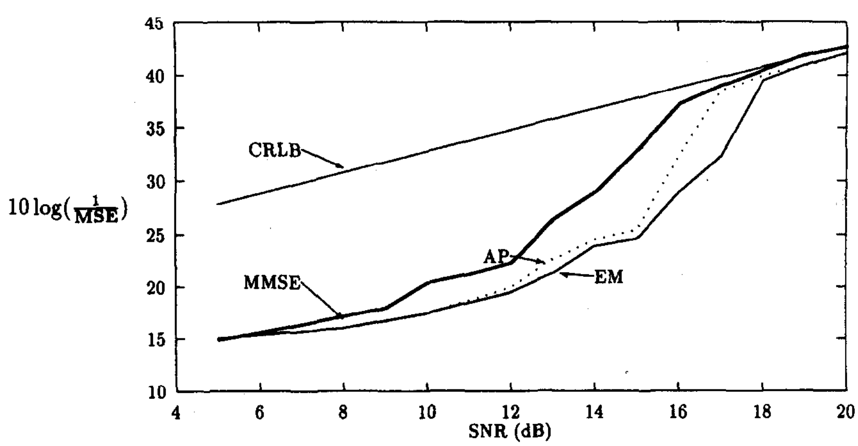


(c) $f_2 = 0.5$



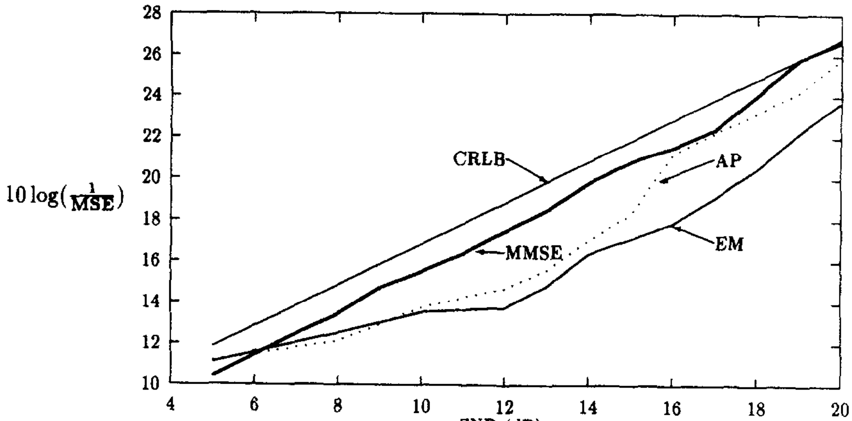
(d) $\alpha_2 = -0.1$

Fig. 3. (c) f_2 =0.5, (d) $\alpha_2 = -0.1$.

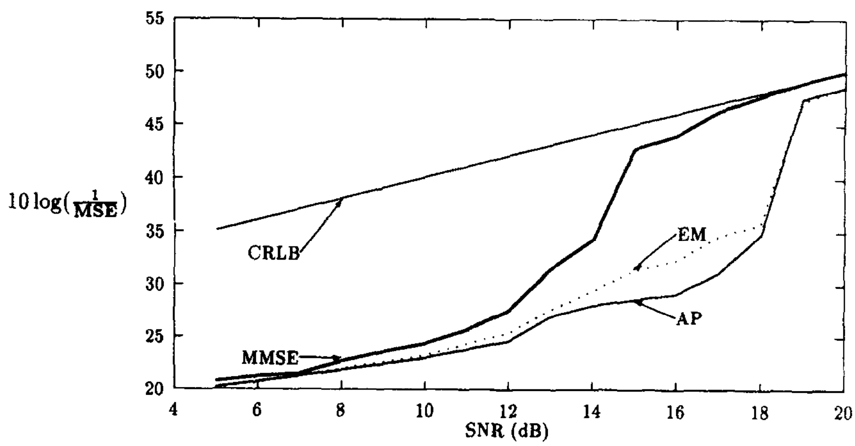


(a) $f_1 = 0.46$

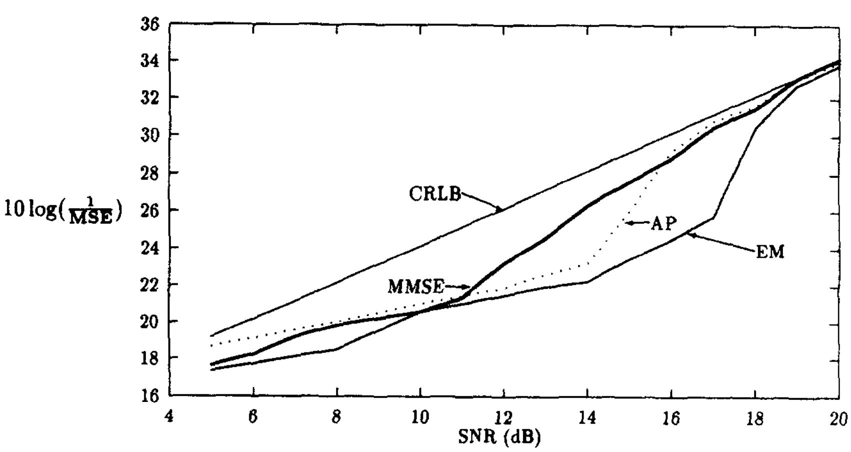
Fig. 4. Performance comparison of the MMSE, EM and AP for q =3, (a) f_1 =0.46 (experiment 4).



(b) $\alpha_1 = -0.2$



(c) $f_2 = 0.5$



(d) $\alpha_2 = -0.1$

Fig. 4. (b) $\alpha_1 = -0.2$, (c) $f_2 = 0.5$, (d) $\alpha_2 = -0.1$.

In each experiment the SNR, which is defined by

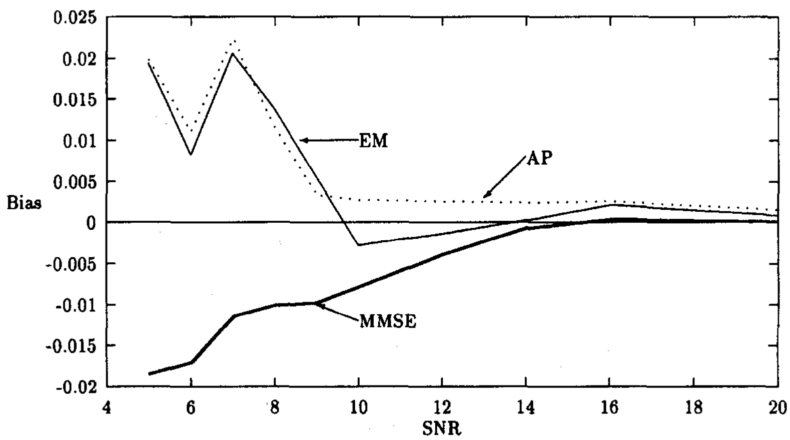
$$\mathrm{SNR} = 10 \log_{10} \frac{|a|^2}{\sigma^2},$$

was varied from 5 to 20 dB in steps of 1 dB. For each SNR we had 500 trials, and for each experiment the initial estimates of the weakest component were obtained by using $M_q = 2$. Note that for the EM and AP schemes, we chose the same initial estimates, except for using multimodal functions for the weakest component.

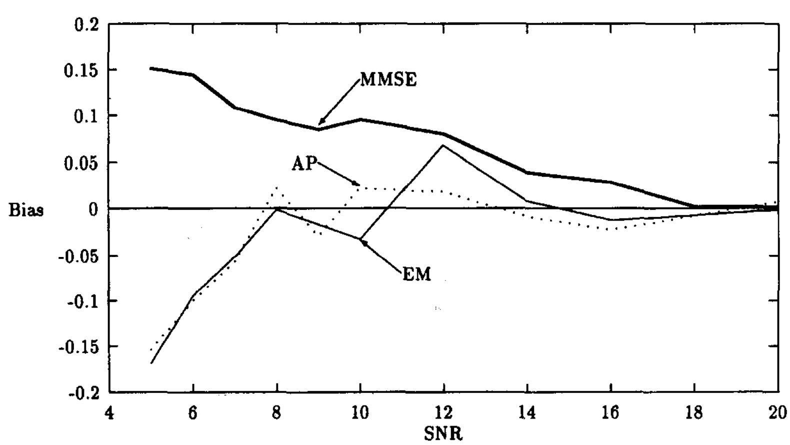
The results of the first experiment are shown in Figs. 1(a)–(d). In Figs. 1(a) and (b), we display the estimation performance for the frequency and damping factor of the first sinusoid, and in Figs. 1(c) and (d), the performance for the same parameters of the second sinusoid. In each figure there are four curves, the CRLBs and the MSEs of the MMSE, EM and AP methods calculated from the 500 trials. It is clear that the MMSE estimator outperforms the AP and the EM schemes.

The results of the second experiment are displayed in Figs. 2(a)–(d), where the same notation is used. Again, the MMSE procedure outperforms the AP and the EM, but this time with even larger margin.

In Figs. 3(a)–(d) we see the results of the third experiment. The MMSE continued to yield better results than the AP and EM.



(a) $f_1 = 0.42$



(b) $\alpha_1 = -0.4$

Fig. 5. Bias performance comparison of the MMSE, EM and AP for $q = 2$, (a) $f_1 = 0.42$, (b) $\alpha_1 = -0.4$ (experiment 2).

The results of the fourth experiment are shown in Figs. 4(a)–(d). We have only shown the performance for the parameters of the first two sinusoids since the performance for the third sinusoid is similar. When we compare these results with the ones from the previous experiment, we realize that the performance of the MMSE has slightly deteriorated. But so has the performance of the AP and EM.

Finally, in Figs. 5(a) and (b), we have graphed the bias of the frequency and damping factor estimates for the first sinusoid in the second experiment of the MMSE, EM and AP methods as a function of the SNR. It is obvious that the three methods are fairly biased for low SNRs.

In general, the simulation results show that for low SNRs and large damping factors, the MMSE algorithm will significantly outperform the AP and the EM methods. This is so because the shape of the likelihood function is not a sharp peak in the f–α plane, but a multitude of smaller peaks. This entails that the iterative methods become sensitive to initialization, which in our method is alleviated by applying the concept of integration of multimodal functions. The other advantage of the MMSE is in the computational load. The MMSE converges more rapidly than the EM and AP methods. In Table 1, we compare the CPU times of the three estimators needed in the above experiments. The numbers indicate the normalized CPU time required by each estimator to obtain the final estimates after the same initialization procedure. Note that these numbers correspond to the total times needed to complete the estimation of 500 independent trials for each SNR in the range 15–20 dB, with a step of 1 dB.

Table 1
The entries denote the normalized CPU times of various estimators after the same initialization

Estimator\EXP	EXP 1	EXP 2	EXP 3	EXP 4
MMSE	1.0	1.0	1.0	1.0
EM	6.83	5.01	5.60	5.98
AP	6.28	4.54	4.42	6.88

We also present results that show the difference in performance between our initialization minimum notch periodogram (MNP) scheme and the initialization by the classical maximum projection (MP) method. The results are displayed in Fig. 6 by the curves that show the mean square error initial estimates of $\hat{\alpha}_2^{(0)}$ obtained by the two methods. Obviously, the MNP yields much better results than the MP scheme. It is interesting to observe that the MP estimates do not improve with the increase of SNR, whereas the MNP estimates do improve. All these results are consistent with our analysis, which predicted that the notch periodogram should be a better procedure. The MNP has a better ability to reduce interference from other components in the parameter estimation of individual components than the periodogram.

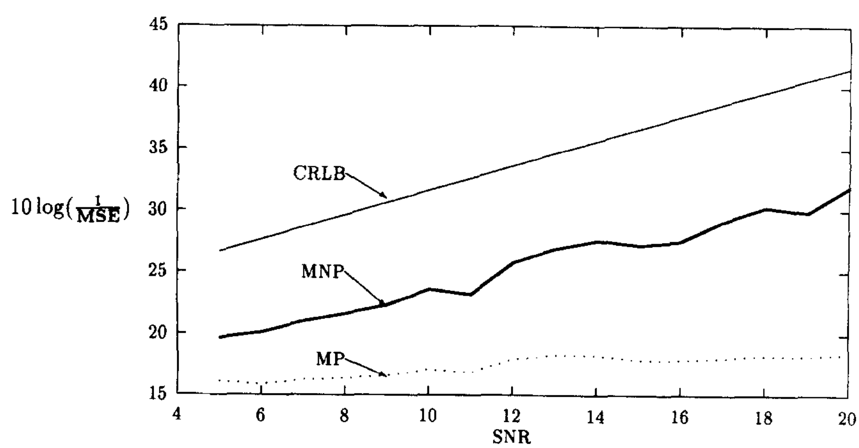


Fig. 6. Initialization performance comparison of the MNP and MP for $\alpha_2 = -0.1$ (experiment 1).

6. Conclusions

We proposed an efficient MMSE algorithm for parameter estimation of exponentially damped sinusoids. The algorithm is based on an iterative scheme, similar to the AP and EM methods. We also introduced a numerical technique, AGQ, to improve the computational efficiency of our algorithm. The experimental results showed agreement with our analysis as well as excellent performance, particularly for sinusoids with close frequencies and/or large damping factors.

Appendix A. Adaptive Gaussian quadrature

An important task in Bayesian analysis is the integration of various functions. Very often the integration cannot be carried out analytically, and one has to resort to numerical integrations. There are many approaches for numerical integration, and for some of them we provide in Table 2 their convergence rates to the exact values of the integrals for one- and d-dimensional integrations [8]. The entries in the table show the uncertainties of the results obtained by the various methods as a function of the number of evaluations b.

Since the integrations in (19) and (20) are two-dimensional and the Gaussian quadrature (GQ) has the best convergence rate for low-dimensional integration, we propose an adaptive numerical technique based on the concept of GQ [3], which we call AGQ. This method uses different number of evaluation points for various subintervals. The goal is to compute the expected value of $\boldsymbol{x}$, that is,

$$\boldsymbol{\mu} = \int_{\{x_1\}} \int_{\{x_2\}} \cdots \int_{\{x_l\}} \boldsymbol{x} p(\boldsymbol{x}), \tag{A.1}$$

Table 2
The entries denote the uncertainty of results as a function of number of evaluations b. In the last row, k denotes the number of terms in Gaussian quadrature

Uncertainty as a function of number of evaluations b	In one dimension	In d dimensions
Monte Carlo	$b^{-1/2}$	$b^{-1/2}$
Trapezoidal rule	b^{-2}	$b^{-2/d}$
Simpson's rule	b^{-4}	$b^{-4/d}$
Gaussian quadrature	b^{-2k+1}	$b^{-(2k-1)/d}$

where $\boldsymbol{x}^{\mathrm{T}} = [x_1, x_2, \ldots, x_l]$. We consider one-dimensional integrals first, where we use 10 function evaluations to calculate each one-dimensional integral. The number of evaluations can be different and the algorithm modified easily. We assume that the initial estimates of the first and second moments of x are $\hat{\mu}^{(0)}$ and $\hat{m}^{(0)}$. Since the p.d.f. $p(x)$ can be improper, to obtain the first estimate of the expected value, we use

$$\hat{\mu}^{(1)} = \frac{\int_{\{x\}} x p(x)\,\mathrm{d}x}{\int_{\{x\}} p(x)\,\mathrm{d}x}, \tag{A.2}$$

where $\{x\}$ is the support of x defined by $\{x\} = \{x\colon -4h + \hat{\mu}^{(0)} \leqslant x \leqslant 4h + \hat{\mu}^{(0)}\}$ and h is a constant defined later. In line with the concept of adaptive integration, we divide the interval $[-4h + \hat{\mu}^{(0)}, 4h + \hat{\mu}^{(0)}]$ into 3 suitable subintervals and use an appropriate number of terms of the GQ in each subinterval. We propose to partition $\{x\}$ into 3 subintervals according to $\{x_a\} = [-4h + \hat{\mu}^{(0)}, -h + \hat{\mu}^{(0)}]$, $\{x_b\} = [-h + \hat{\mu}^{(0)}, h + \hat{\mu}^{(0)}]$ and $\{x_c\} = [h + \hat{\mu}^{(0)}, 4h + \hat{\mu}^{(0)}]$. Moreover, for the middle set we use a GQ formula with four terms and for the remaining two, three terms. Thus, there are three, four and three evaluation points in the ranges $\{x_a\}$, $\{x_b\}$ and $\{x_c\}$, respectively. Note that the above choice depends on the prior information of the posterior distribution of the parameters but is still somewhat arbitrary.

Next we show how to choose the value of h. As an importance function we adopt $g(x)$ displayed in Fig. 7. The standard deviation of a random variable X, whose p.d.f. is $g(x)$, is $2.08h$. This implies that

$$2.08h = \sqrt{m - \mu^2},$$

where μ and m are the first and second moments of X, respectively. Hence, we could estimate h using the initial estimates $\hat{m}^{(0)}$ and $\hat{\mu}^{(0)}$ from

$$\hat{h}^{(1)} = 0.48\sqrt{\hat{m}^{(0)} - \hat{\mu}^{(0)2}}. \tag{A.3}$$

Now, we start the procedure by estimating $\hat{m}^{(1)}$ and $\hat{\mu}^{(1)}$ followed by the evaluation of $\hat{h}^{(2)}$ according to (A.3). The integral in (A.2) becomes

$$\hat{\mu}^{(1)} = \frac{\int_{\{x_a\}} x p(x)\,\mathrm{d}x + \int_{\{x_b\}} x p(x)\,\mathrm{d}x + \int_{\{x_c\}} x p(x)\,\mathrm{d}x}{\int_{\{x_a\}} p(x)\,\mathrm{d}x + \int_{\{x_b\}} p(x)\,\mathrm{d}x + \int_{\{x_c\}} p(x)\,\mathrm{d}x}, \tag{A.4}$$

where $\int_{\{x_a\}} xp(x)\,dx$ and $\int_{\{x_c\}} xp(x)\,dx$ are computed by a 3-term formula, and $\int_{\{x_b\}} xp(x)\,dx$ by a 4-term Gaussian formula. Similarly, for estimation of m, we use

$$\hat{m}^{(1)} = \frac{\int_{\{x_a\}} x^2 p(x)\,dx + \int_{\{x_b\}} x^2 p(x)\,dx + \int_{\{x_c\}} x^2 p(x)\,dx}{\int_{\{x_a\}} p(x)\,dx + \int_{\{x_b\}} p(x)\,dx + \int_{\{x_c\}} p(x)\,dx}. \quad (A.5)$$

Now we repeat the procedure, but we exploit $\hat{\mu}^{(1)}$ and $\hat{m}^{(1)}$ as initial estimates. In general, the kth estimates of $\hat{\mu}^{(k)}$ and $\hat{m}^{(k)}$ are evaluated by (A.4) and (A.5), where $\hat{h}^{(k)} = 0.48\sqrt{\hat{m}^{(k-1)} - \hat{\mu}^{(k-1)2}}$.

Next we consider the two-dimensional integration with the AGQ technique, where we have two random variables X_1 and X_2 whose joint p.d.f. is $p(x_1, x_2)$. Let the initial estimates of the first two moments be $[\hat{\mu}_1^{(0)}, \hat{\mu}_2^{(0)}]$ and $[\hat{m}_1^{(0)}, \hat{m}_2^{(0)}]$. Then, the new estimate of μ_1 is given by

$$\hat{\mu}_1^{(1)} = \frac{\int_{\{x_1\}} x_1 \int_{\{x_2\}} p(x_1, x_2)\,dx_2\,dx_1}{\int_{\{x_1\}} \int_{\{x_2\}} p(x_1, x_2)\,dx_2\,dx_1}, \quad (A.6)$$

where $\{x_i\}$ is the support of x_i defined by $\{x_i\} = \{x_i\colon\ -4\hat{h}_i^{(1)} + \hat{\mu}_i^{(0)} \leqslant x_i \leqslant 4\hat{h}_i^{(1)} + \hat{\mu}_i^{(0)}\}$, $\hat{h}_i^{(1)} = 0.48 \times \sqrt{\hat{m}_i^{(0)} - \hat{\mu}_i^{(0)}}$, $i = 1, 2$. For each $\{x_i\}$, $i = 1, 2$, we partition the support into 3 subintervals $\{x_{ia}\}$, $\{x_{ib}\}$ and $\{x_{ic}\}$ in the same way as in the one-dimensional case. First for each of the 10 evaluation points of x_1, we use 10 function evaluations to solve the second integral,

$$p(x_1) = \int_{\{x_2\}} p(x_1, x_2)\,dx_2,$$

and obtain the marginal p.d.f. $p(x_1)$. Thus, (A.6) is simplified to computing a set of one-dimensional integrals as (A.2), which all together involves 10^2 evaluation points. The remaining moments, $\hat{\mu}_2^{(1)}$, $\hat{m}_1^{(1)}$ and $\hat{m}_2^{(1)}$, are estimated similarly. These methods can be easily extended to l-dimensional integrals, where the number of evaluation points will be 10^l.

For the integration of multimodal functions for the weakest component in (8), (A.4) is modified to [11]

$$\hat{\mu}^{(1)} = \frac{\int_{\{x_{(1)}\}} xp(x)\,dx + \int_{\{x_{(2)}\}} xp(x)\,dx + \cdots + \int_{\{x_{(M)}\}} xp(x)\,dx}{\int_{\{x_{(1)}\}} p(x)\,dx + \int_{\{x_{(2)}\}} p(x)\,dx + \cdots + \int_{\{x_{(M)}\}} p(x)\,dx}, \quad (A.7)$$

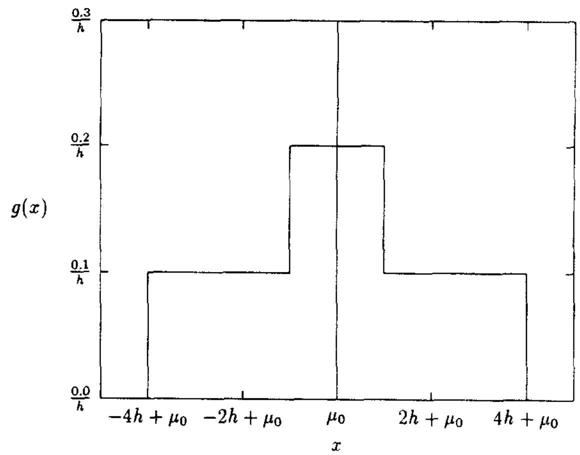


Fig. 7. The importance function $g(x)$ of the AGQ method.

where the interval of the lth mode $x_{(l)}$ is given by $\{x_{(l)}\} = \{x\colon\ -4h + \hat{\mu}_{(l)}^{(0)} \leqslant x \leqslant 4h + \hat{\mu}_{(l)}^{(0)}\}$, and $\hat{\mu}_{(l)}^{(0)}$ is the initial estimate of the lth mode, $l = 1, 2, \ldots, M_q$. In summary, in estimating $\hat{f}_q^{(1)}$ and $\hat{\alpha}_q^{(1)}$, the integral range includes the highest M_q peaks in the f–α plane of $P_n(f, \alpha; \hat{\theta}_{q-1})$. Therefore, we need $10^2 M_q$ function evaluations to estimate $\hat{f}_q^{(1)}$ and $\hat{\alpha}_q^{(1)}$, and 10^2 function evaluations in the remaining iterations, as well as 10^2 function evaluations for each of the remaining parameters in any iteration. The simulation results show that this modification extends the range of optimal performance by an SNR of 2–5 dB.

References

[1] G.E.P. Box and G.C. Tiao, *Bayesian Interference in Statistical Analysis*, Addison-Wesley, Reading, MA, 1973, Chapter 1, pp. 25–60.

[2] M. Feder and E. Weinstein, "Parameter estimation of superimposed signals using the EM algorithm", *IEEE Trans. Acoust. Speech Signal Process.*, Vol. 36, No. 4, April 1988, pp. 477–489.

[3] C.F. Gerald and P.O. Wheatley, *Applied Numerical Analysis*, Addison-Wesley, Reading, MA, 3rd Edition, 1984, Chapter 4, pp. 245–265.

[4] M.L. Honig and D.G. Messerschmitt, *Adaptive Filters Structure, Algorithms, and Applications*, Kluwer, Dordrecht, 1984, Chapter 6, pp. 145–160.

[5] Y. Hua and T.K. Sarkar, "Matrix pencil method for estimating parameters of exponentially damped/undamped sinusoids in noise", *IEEE Trans. Acoust. Speech Signal Process.*, Vol. 38, No. 5, May 1990, pp. 814–824.
[6] J.-K. Hwang and Y.-C. Chen, "A combined detection–estimation algorithm for the harmonic-retrieval problem", *Signal Processing*, Vol. 30, No. 2, January 1993, pp. 177–197.
[7] J.K. Hwang and Y.C. Chen, "Superresolution frequency estimation by alternating notch periodogram", *IEEE Trans. Signal Processing*, Vol. 41, No. 2, February 1993, pp. 727–741.
[8] F. James, "Monte Carlo theory and practice", *Rep. Phys.*, Vol. 43, 1980, pp. 1146–1189.
[9] D.E. Johnston and P.M. Djurić, "Bayesian detection and MMSE frequency estimation of sinusoidal signal via adaptive importance sampling", *Proc. Internat. Symposium Circuits and Systems*, London, England, Vol. 2, 1994, pp. 417–420.
[10] R. Kumaresan and D.W. Tufts, "Estimating the parameters of exponentially damped sinusoids and pole-zero modeling in noise", *IEEE Trans. Acoust. Speech Signal Process.*, Vol. 30, No. 6, December 1982, pp. 833–840.
[11] M.S. Oh and J.O. Berger, "Integration of multimodal functions by Monte Carlo importance sampling", *J. Amer. Statist. Assoc.*, Vol. 88, No. 422, June 1993, pp. 450–456.
[12] I. Ziskind and M. Wax, "Maximum likelihood localization of multiple source by alternating projection", *IEEE Trans. Acoust. Speech Signal Process.*, Vol. 36, No. 10, October 1988, pp. 1553–1560.