Journal of Magnetic Resonance **145,** 108–114 (2000)
doi:10.1006/jmre.2000.2077, available online at http://www.idealibrary.com on IDEAL®

Exact ML Estimation of Spectroscopic Parameters[1]

Petre Stoica and Tomas Sundin

Department of Systems and Control, Uppsala University, P.O. Box 27, SE-75103 Uppsala, Sweden

E-mail: ps@syscon.uu.se, tsu@syscon.uu.se

Received November 17, 1999; revised February 29, 2000

In a paper on spectroscopic imaging published in this journal Spielman *et al.* (*J. Magn. Reson.* 79, 66–77 (1988)) made the important point that a priori information about the compounds present can and should be incorporated into the estimation of spectroscopic signal parameters. They proposed using the maximum likelihood (ML) approach for parameter estimation, but failed to incorporate properly the full a priori information that was assumed to be available. Consequently they ended up with a spectroscopic imaging method that is only a suboptimal approximation of the ML method. In this paper we derive the exact ML method, present a computationally efficient implementation of it (which is much faster than the direct implementation suggested by Spielman *et al.* for their suboptimal method), and illustrate numerically the performance gain that can be achieved over the method of Spielman *et al.*

Key Words: **magnetic resonance spectroscopy (MRS); magnetic resonance spectroscopic imaging (MRSI); maximum-likelihood estimation; global optimization.**

1. INTRODUCTION AND PROBLEM STATEMENT

The spectroscopic signal obtained by any of a number of localization methods applied to each voxel of interest can be reasonably well modeled as a sum of exponentially damped sinusoids in noise (see, e.g., (*1*)):

$$x(t) = \sum_{k=1}^{n} \beta_k e^{-\alpha_k t} e^{i(\omega_k t + \varphi_k)} + \epsilon(t) \quad t = 1, 2, \ldots, N, \quad [1]$$

where the sampling interval has been absorbed in $\{\alpha_k\}$ and $\{\omega_k\}$ for notational convenience. In Eq. [1], $\{\alpha_k\}$ are the damping coefficients, $\{\beta_k\}$ are the amplitudes, $\{\omega_k\}$ are the frequencies, $\{\varphi_k\}$ are the initial phases, and $\{\epsilon(t)\}$ denotes a noise sequence. Equation [1] constitutes a model commonly used in magnetic resonance (MR) spectroscopy, and the estimation of its unknown parameters is a well-studied problem for which a large number of solutions have appeared in both the MR literature and the signal processing literature (see, e.g., (*3*) and the many references therein). However, as pointed out in (*1*), in virtually all spectroscopic imaging applications the compounds being imaged and their MR spectroscopic structure are known. More exactly, with reference to (*1*), the number of components n is known. Owing to the main field inhomogeneity, the frequencies and initial phases of the spectral lines present are not exactly known; however, the frequency and phase differences are known. In other words we can write

$$\omega_k = \omega - \Delta\omega_k \quad k = 1, \ldots, n \quad [2]$$

and

$$\varphi_k = \varphi - \Delta\varphi_k \quad k = 1, \ldots, n, \quad [3]$$

where ω and φ are unknown but $\{\Delta\omega_k\}$ and $\{\Delta\varphi_k\}$ are known. Finally, the decaying coefficients $\{\alpha_k\}$ are also usually known with quite good accuracy. Inserting Eq. [2] and Eq. [3] into the spectroscopic signal equation, Eq. [1], we obtain

$$x(t) = \sum_{k=1}^{n} \Phi_{t,k} \beta_k e^{i(\omega t + \varphi)} + \epsilon(t) \quad t = 1, \ldots, N, \quad [4]$$

where

$$\Phi_{t,k} = e^{-(\alpha_k + i\Delta\omega_k)t - i\Delta\varphi_k} \quad [5]$$

are given for $t = 1, \ldots, N$ and $k = 1, \ldots, n$.

The basic problem considered in (*1*) was the estimation of $\{\beta_k\}$ (the parameters of major interest) and (ω, φ) (the nuisance parameters) in Eq. [4] from N noisy observations $\{x(t)\}_{t=1}^{N}$. To review the parameter estimation method in (*1*), let

$$a = [\beta_1 e^{i(\varphi - \Delta\varphi_1)} \ldots \beta_n e^{i(\varphi - \Delta\varphi_n)}]^T \quad [6]$$

$$x = [x(1) \ldots x(N)]^T \quad [7]$$

[1] This work was partly supported by the Senior Individual Grant Program of the Swedish Foundation for Strategic Research (SSF).

1090-7807/00 $35.00

and let Ψ be the $N \times n$ matrix with the (t, k)-element given by

$$\Psi_{t,k} = e^{[-\alpha_k + i(\omega - \Delta\omega_k)]t}. \quad [8]$$

Under the assumption that the noise $\{\epsilon(t)\}$ is white and circularly Gaussian distributed, the maximum likelihood (ML) estimation of the unknown parameters reduces to the minimization with respect to (w.r.t.) ω, φ, and $\{\beta_k\}$ of the function (see, e.g., (*1*, *3*)):

$$\|x - \Psi a\|^2, \quad [9]$$

where $\|\cdot\|$ denotes the Euclidean vector norm. *Ignoring the fact that the phase differences* $\{\Delta\varphi_k\}$ *in Eq.* [6] *are known,* (*1*) minimized Eq. [9] w.r.t. a to obtain

$$\hat{a} = (\Psi^*\Psi)^{-1}\Psi^* x. \quad [10]$$

The corresponding minimum value of Eq. [9] is

$$\|(I - \Psi(\Psi^*\Psi)^{-1}\Psi^*)x\|^2. \quad [11]$$

Consequently, ω was estimated in (*1*) by minimizing the function above,

$$\min_{\omega} \|(I - \Psi(\Psi^*\Psi)^{-1}\Psi^*)x\|^2. \quad [12]$$

The resulting parameter estimates are not MLE owing to the failure to incorporate the information that $\{\Delta\varphi_k\}$ are known. The larger n the less parsimonious is the model of $\{x(t)\}$ based on a and ω, and hence the less accurate are the parameter estimates derived from this model. For convenience's sake in what follows we will refer to the spectroscopic parameter estimator in Eq. [10] and Eq. [12] as the Spielman method.

The main goal of the present paper is to derive *the exact ML estimates of* $\{\beta_k\}$, ω, *and* φ. A second goal is to describe *a computationally efficient algorithm* for computing the exact MLE. In (*1*) the 1-D search problem in Eq. [12] was apparently solved by computing the projection matrix $\Psi(\Psi^*\Psi)^{-1}\Psi^*$ for each value of ω tested. However, this may be computationally quite demanding (and is in fact unnecessary). Here we will show that the main computational step of the exact MLE consists also of a 1-D search problem, for the solution of which we will develop a computationally efficient FFT-based algorithm. The exact MLE proposed in this paper will be shown to be not only more accurate than the suboptimal method in (*1*) but also computationally faster (when implemented as described below).

2. EXACT MLE

Let Φ be the $(N \times n)$ matrix with the (t, k)-element given by $\Phi_{t,k}$ in Eq. [5], and let

$$\beta = [\beta_1 \ldots \beta_n]^T \quad [13]$$

$$D = \begin{pmatrix} e^{i\omega} & & 0 \\ & \ddots & \\ 0 & & e^{iN\omega} \end{pmatrix}. \quad [14]$$

Using this notation we can rewrite the ML criterion in Eq. [9] as

$$\|x - D\Phi\beta e^{i\varphi}\|^2 = \|D^*x - \Phi\beta e^{i\varphi}\|^2. \quad [15]$$

Hereafter, the symbol * denotes the conjugate transpose. We remark on the fact that if the noise $\{\epsilon(t)\}$ in Eq. [4] is not white and circular Gaussian then the method based on the minimization of Eq. [15] is no longer MLE. However, even for colored non-Gaussian noise the parameter estimates obtained by minimizing Eq. [15] turn out to be quite accurate for a wide range of scenarios (see, e.g., (*2*)). Whenever the noise in Eq. [4] is either nonwhite or non-Gaussian the method that minimizes Eq. [15] to obtain parameter estimates is called *nonlinear least squares.*

In the following we minimize the (negative) log-likelihood function (LF) in Eq. [15] w.r.t. β, φ, and ω in a step-by-step manner.

Optimization of LF w.r.t. β

Since β is a real-valued vector, we have

$$\begin{aligned} \|D^*x - (\Phi e^{i\varphi})\beta\|^2 &= x^*x + \beta^T\Phi^*\Phi\beta - \beta^T e^{-i\varphi}\Phi^*D^*x \\ &\quad - x^*D\Phi e^{i\varphi}\beta \\ &= x^*x + \beta^T \mathrm{Re}(\Phi^*\Phi)\beta \\ &\quad - 2\beta^T \mathrm{Re}(e^{-i\varphi}\Phi^*D^*x), \end{aligned} \quad [16]$$

where $\mathrm{Re}(\cdot)$ stands for the real part of the quantity between the parentheses. The minimization of Eq. [16] w.r.t. β is straightforward. For any fixed ω and φ, the minimizer is given by

$$\hat{\beta} = \Gamma^{-1}\mathrm{Re}(\Phi^*D^*xe^{-i\varphi}), \quad [17]$$

where

$$\Gamma = \mathrm{Re}(\Phi^*\Phi). \quad [18]$$

It can be readily verified (using some simple property of Vandermonde matrices) that the inverse in Eq. [17] exists if and only if

$$\alpha_k + i\Delta\omega_k \neq \alpha_p + i\Delta\omega_p \text{ for } k \neq p, \quad [19]$$

which is evidently a weak requirement.

Inserting $\hat{\beta}$ into Eq. [16] yields the following function that is to be minimized w.r.t. φ and ω:

$$\hat{\beta}^T\Gamma\hat{\beta} - 2\hat{\beta}^T\mathrm{Re}(\Phi^*D^*xe^{-i\varphi}) = -[\mathrm{Re}(\Phi^*D^*xe^{-i\varphi})]^T\Gamma^{-1}[\mathrm{Re}(\Phi^*D^*xe^{-i\varphi})]. \quad [20]$$

Let $\Gamma^{1/2}$ denote a square root of the positive definite matrix Γ (i.e., $\Gamma = \Gamma^{1/2}\Gamma^{T/2}$), and let

$$\gamma(\omega) = \Gamma^{-1/2}\Phi^*D^*x. \quad [21]$$

With this notation the problem of determining the MLE of φ and ω reduces to the *maximization* of the function

$$f = \|\mathrm{Re}[\gamma(\omega)e^{-i\varphi}]\|^2. \quad [22]$$

Optimization of LF w.r.t. φ

Omitting the dependence of γ on ω, for notational convenience, we can rewrite f in Eq. [22] as

$$4f = (\gamma^Te^{-i\varphi} + \gamma^*e^{i\varphi})(\gamma e^{-i\varphi} + \bar{\gamma}e^{i\varphi}) = (\gamma^T\gamma)e^{-i2\varphi} + \gamma^*\bar{\gamma}e^{i2\varphi} + \gamma^T\bar{\gamma} + \gamma^*\gamma, \quad [23]$$

where the overbar denotes the complex conjugate. It follows that

$$2f = \gamma^*\gamma + \mathrm{Re}[(\gamma^T\gamma)e^{-i2\varphi}] = \gamma^*\gamma + |\gamma^T\gamma|\cos[\arg(\gamma^T\gamma) - 2\varphi], \quad [24]$$

where, for any complex variable ν, we write $\nu = |\nu|e^{i\arg(\nu)}$. The maximization of Eq. [24] w.r.t. φ is immediate. The maximizer is

$$\hat{\varphi} = \frac{1}{2}\arg(\gamma^T\gamma) \quad [25]$$

and the function left for maximization w.r.t. ω is given by

$$g(\omega) = \gamma^*(\omega)\gamma(\omega) + |\gamma^T(\omega)\gamma(\omega)|. \quad [26]$$

Optimization of LF w.r.t. ω

Estimation of ω by

$$\hat{\omega} = \arg\max_{\omega} g(\omega) \quad [27]$$

is a 1-D search problem. In the rest of this section we present an FFT-based method for solving Eq. [27] in a computationally efficient manner. Let $\mu_p(k)$ denote the (p, k)-element of the $n \times N$ matrix $\Gamma^{-1/2}\Phi^*$, that is,

$$\mu_p(k) = [\Gamma^{-1/2}\Phi^*]_{p,k} \quad p = 1, \ldots, n; \quad k = 1, \ldots, N. \quad [28]$$

The pth element of $\gamma(\omega)$ in Eq. [21] can then be written as

$$\gamma_p(\omega) = \sum_{k=1}^{N} \mu_p(k)\,x(k)e^{-i\omega k} \quad p = 1, \ldots, n. \quad [29]$$

Hence $\{\gamma_p(\omega)\}$, for $\omega \in (-\pi, \pi]$, can be efficiently evaluated by using *an FFT algorithm possibly with zero padding.* This will lead to a fast method for picking the peak of $g(\omega)$ in Eq. [26], which will be similar to the periodogram method in spectral analysis (*3*). We note that zero padding will be needed whenever $2\pi/N$ is deemed to be larger than (some fraction of) the expected estimation error in $\hat{\omega}$.

In Summary

1. We use an FFT-based method to obtain the estimate $\hat{\omega}$ in Eq. [27], assuming that no good initial estimate of ω is known; if a reliable estimate of ω is available then we replace the global search for the maximizer in Eq. [27] by a computationally more efficient local search (see the following section). This is the main computational step of the exact MLE algorithm.

2. Next we estimate φ as in Eq. [25]. The computational burden of this step is negligible.

3. Finally we obtain estimates of the parameters of major interest, $\{\beta_k\}$, using Eq. [17]. Note that in view of Eq. [21],

$$\hat{\beta} = \Gamma^{-T/2}\mathrm{Re}[\gamma(\hat{\omega})e^{-i\hat{\varphi}}], \quad [30]$$

where both $\Gamma^{-T/2}$ and $\gamma(\hat{\omega})$ are available from Step 1. Hence the computational burden of Step 3 is also quite modest.

In the sequel we assume, as stated before, that $\{\alpha_k\}$ are exactly known. What happens when this is not the case is an interesting question. More exactly, the question concerns the sensitivity of the ML estimates of $\{\beta_k\}$ to errors in $\{\alpha_k\}$ as well as possible means of reducing this sensitivity (if significant). While a study of these aspects is beyond the scope of the present paper (for clarity and conciseness reasons), we note briefly that relatively significant errors in $\{\alpha_k\}$ can be handled in a multistep manner as follows.

1. Using the latest available values of $\{\alpha_k\}$, obtain estimates of $\{\omega, \varphi, \beta_k\}$ in Eq. [1] by means of the ML algorithm described above.

2. Using the latest available estimates of $\{\omega, \varphi, \beta_k\}$ in the ML criterion in Eq. [15] and viewing the so-obtained criterion

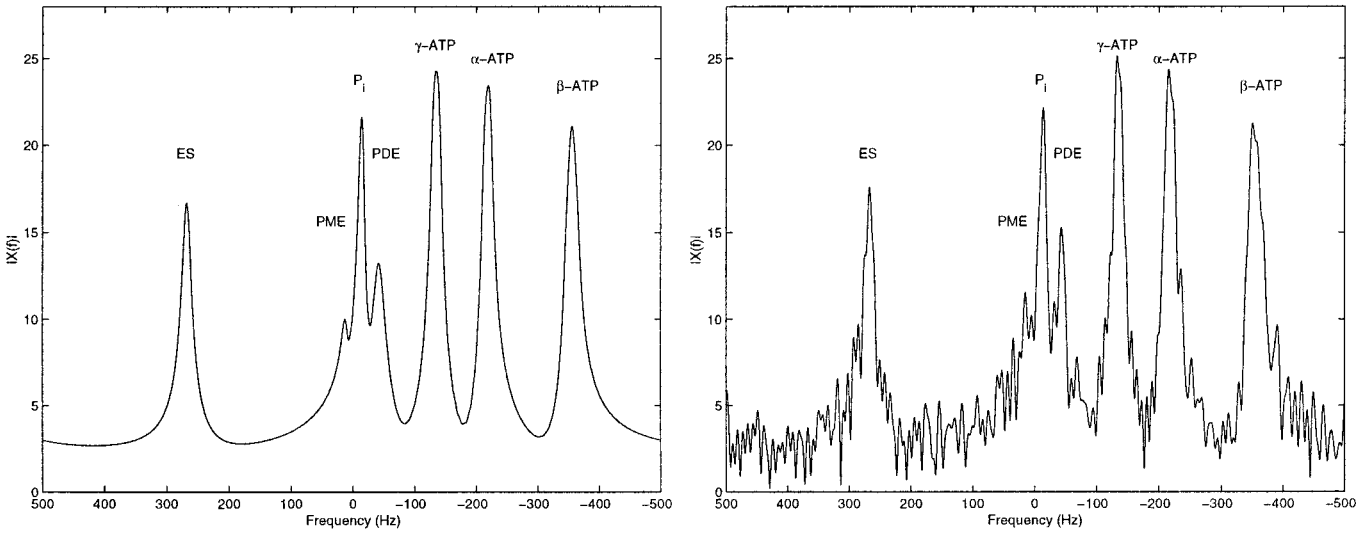

FIG. 1. Magnitude spectra (FFT) of simulated ^{31}P data. The peaks correspond to external standard (ES), phosphomonoesters (PME), iorganic phosphate (P_i), phosphodiesters (PDE), and the γ, α, and β phosphorus of adenosine triphosphate (ATP). Left: Magnitude spectrum of noise free data. Right: Magnitude spectrum of noisy data. $N = 128$ and SNR -12.5 dB for PME.

as a function of $\{\alpha_k\}$, perform *one-step* with a Newton–Raphson (or scoring) algorithm initialized in the most recent estimate of $\{\alpha_k\}$.

By iterating the above two steps only a few times (such as once or twice) we can mitigate the degradation of the ML estimates of $\{\omega, \varphi, \beta_k\}$ owing to errors in $\{\alpha_k\}$, unless the latter are rather significant (in which case a larger number of iterations of steps 1 and 2 may be required). Since the computational complexity of step 2 is comparatively low, the computational burden of the iterative algorithm outlined above is approximatively equal to that of the MLE of $\{\omega, \varphi, \beta_k\}$ times the number of iterations performed.

3. NUMERICAL EXAMPLE

In this section a numerical example is used to compare the two methods with respect to both their statistical accuracy and their computational complexity. Experimental signals contain many different sources of error which makes it difficult to distinguish between the actual estimation errors and the errors related to experimental imperfections. Therefore the following evaluation is performed using simulated data.

The signal considered consists of 11 peaks satisfying the model given in Eq. [1]. On the left-hand side of Fig. 1 the magnitude spectrum (obtained by the FFT) of the noise-free data is displayed (for $N = 128$). The data approximately correspond to a ^{31}P spectrum with the following metabolite peaks: external standard (ES), phosphomonoesters (PME), inorganic phosphate (P_i), phosphodiesters (PDE), and the γ (doublet), α (doublet), and β (triplet) phosphorus of adenosine triphosphate (ATP). On the right-hand side of Fig. 1 the magnitude spectrum of a noisy realization is displayed (again for $N = 128$). The added complex, white circular Gaussian noise sequence has a variance chosen as to obtain a signal-to-noise ratio (SNR) of -12.5 dB for the smallest peak (PME). Note that the closely spaced peaks of the γ and α-ATP doublets and the β-ATP triplet can not be resolved by the FFT.

The SNR is defined as the ratio of the power of the weakest component in Eq. [1] (with $N < \infty$) to the noise power:

$$\frac{1}{N}\sum_{t=0}^{N-1} \beta_1^2 e^{-2\alpha_1 t}/\sigma^2 = \frac{\beta_1^2}{\sigma^2}\frac{1 + e^{-2\alpha_1} + \cdots + e^{-2\alpha_1(N-1)}}{N} \triangleq SNR, \qquad [31]$$

where σ^2 denotes the noise variance, and we assume that the weakest component in Eq. [1] corresponds to $k = 1$ (for the sake of discussion). For $\alpha_1 = 0$, Eq. [31] becomes the usual ratio of powers, β^2/σ^2. For $\alpha_1 > 0$, Eq. [31] typically decreases as N increases, as might be expected. Note that N may be chosen to match the decaying time of the most slowly decaying component in Eq. [1] and that this component may not be the weakest. Hence the SNR definition in Eq. [31] may indicate a less favorable scenario than what we really have at hand, and perhaps a more informative definition of the SNR would correspond to summing Eq. [31] over all components in Eq. [1] (for $k = 1, \ldots, n$).

The frequency estimate $\hat{\omega}$ is obtained by solving the 1-dimensional optimization problem in Eq. [12] or Eq. [27] for the two methods, respectively. The optimization is performed with

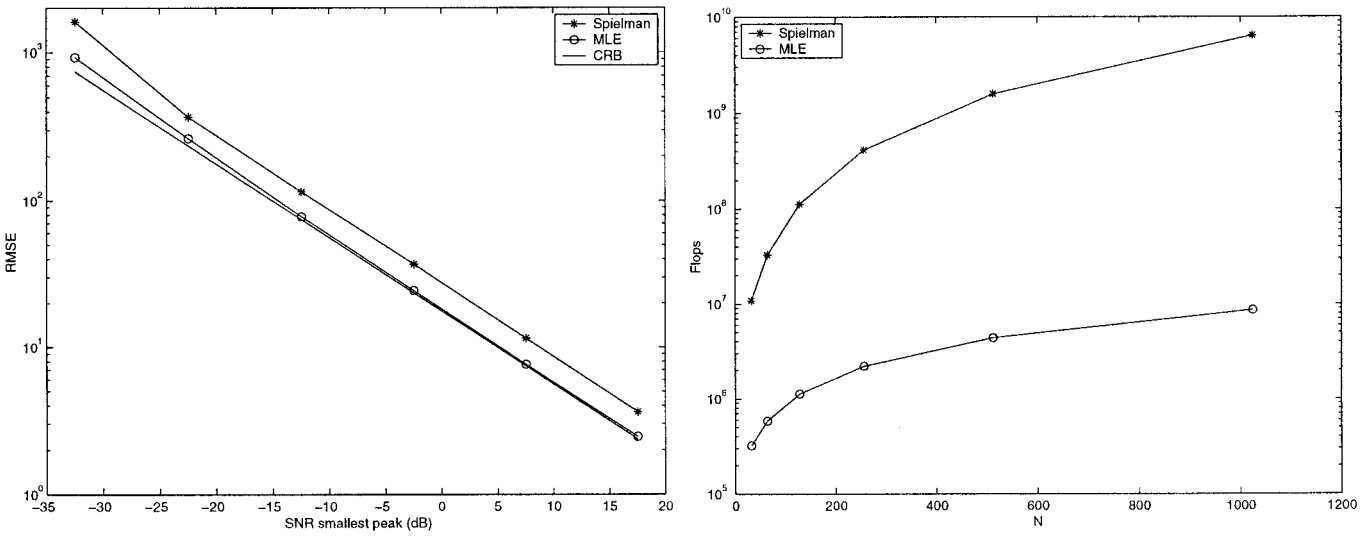

FIG. 2. Left: Sum of the estimated amplitude RMSE for the Spielman method and MLE as a function of SNR. Results were obtained via a local search from 1000 simulation runs using 128 data points. Right: Number of floating point operations (flops) associated with a local search for the Spielman method and MLE as a function of the number of data points. SNR $= -12.5$ dB for the smallest peak (PME).

either a *local* optimization algorithm or a *global* optimization algorithm. The local optimization algorithms have the advantage of fast convergence and small computational complexity. However, a drawback is that these local approaches are dependent on a good starting value to converge to the desired global optimum. In spectroscopic imaging applications such starting values are normally available from spectra in neighboring voxels. Therefore, local optimization algorithms such as gradient-based or simplex-based methods are the ones most used in practice. On the other hand global optimization algorithms are also of interest in those cases in which good initial estimates are not available. Note that the optimization is only w.r.t. one parameter, ω. This allows us to use a "brute force" grid search over the possible frequency interval to find the global optimum. Assuming that the grid is chosen fine enough, this ensures convergence to the global optimum without requiring any starting value. The applicability of this approach is dependent on the amount of computations required to evaluate the LF for each frequency. The desirable properties of the global approach can motivate its use even if the computational load is (slightly) higher than that of the local optimization algorithms.

First the accuracy of the estimates is evaluated using a local optimization algorithm. The optimization is performed using a Nelder–Mead simplex algorithm. To obtain a fair comparison between the methods the exact frequency is provided as a starting value. The empirical accuracy of the estimators is compared to the Cramér–Rao lower bound (CRB) which sets a lower bound on the accuracy of any (asymptotically) unbiased estimator (see, e.g., (*2, 3*)). On the left-hand side of Fig. 2 the sums of the experimental root mean square errors (RMSE) of the 11 amplitude estimates are compared to the corresponding CRB for different noise levels. The number of data points in each set is 128 and the number of simulation runs is 1000. It can be seen that the MLE is very close to the best possible performance while the results for the Spielman method are worse. On the right-hand side of Fig. 2 a comparison of the computational complexity as a function of the number of data points, N, is displayed. The figure shows that the local maximization of the criterion in our exact MLE method requires approximately 100 times less floating point operations (flops) than that of the loss function in the Spielman method.

In Fig. 3 experimental results for the individual peaks obtained from 1000 simulation runs are displayed. The number of data points is 128 and the SNR is -12.5 dB for the smallest peak (PME). To the left the mean Spielman estimates $\pm$ one standard deviation are shown along with the true amplitudes. To the right the corresponding results for the MLE are shown. The Spielman method leads to a visibly larger variance of the estimates in particular for the closely spaced γ- and α-ATP doublets and the β-ATP triplet (the estimated amplitudes of these components are also somewhat biased; see the figure). The reason is that for these difficult cases the correct incorporation of the assumption of known initial phase differences is crucial.

Next we turn our attention to evaluating the methods in combination with a global optimization algorithm. On the left-hand side of Fig. 4 the numbers of flops associated with a grid search are displayed for the two methods, respectively. It is evident that the number of flops for the Spielman method is much larger than for the MLE method. Note that the efficient manner in which the evaluation can be performed for the MLE makes the difference in computations between using the local and global optimization algorithms quite small. For the MLE it is therefore realistic to use the global grid search in practice,

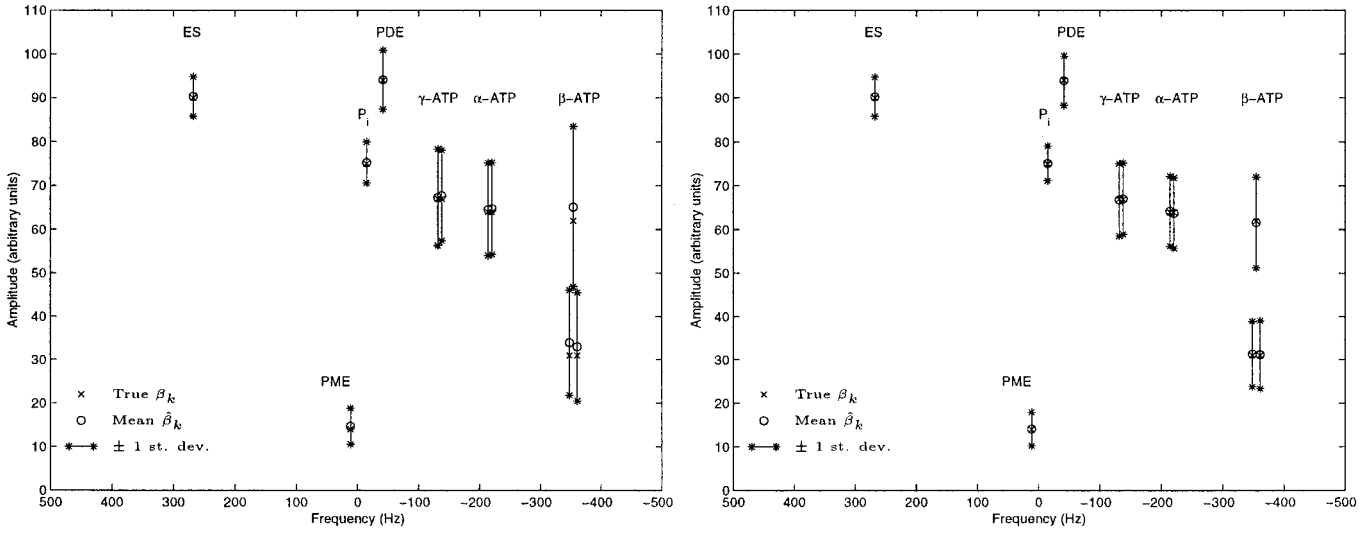

FIG. 3. Individual amplitude estimates. Means and standard deviations were obtained from 1000 simulation runs using 128 data points and SNR $= -12.5$ dB for the smallest peak (PME). Left: Spielman method. Right: MLE.

whereas this is not the case for the Spielman method. On the right-hand side of Fig. 4 simulation results showing the accuracy of the MLE using the global grid search are displayed. The experimental data are gathered from 1000 simulation runs and the number of data points is 128. It can be seen that the estimates are very good as long as the grid is chosen fine enough. For high SNR the finest grid of those considered is necessary to achieve results that are close to the CRB.

Finally, we note that the global search algorithm could also be implemented as follows: first perform a rough grid search by using an FFT with little or no zero padding to find the region where the optimum point lies; then perform a fine grid search in that region using a chirp or a zoom FFT (see, e.g., (*4*)) to accurately locate the optimum. Such an implementation would be computationally more efficient than the global fine-grid search based on an FFT with a lot of zero padding, but it may terminate in a local optimum owing to the fact that the initial rough global search may miss the global optimum.

4. CONCLUDING REMARKS

In summary the proposed MLE was seen to have significantly better accuracy than the Spielman suboptimal method. The numerical results supported the theoretical

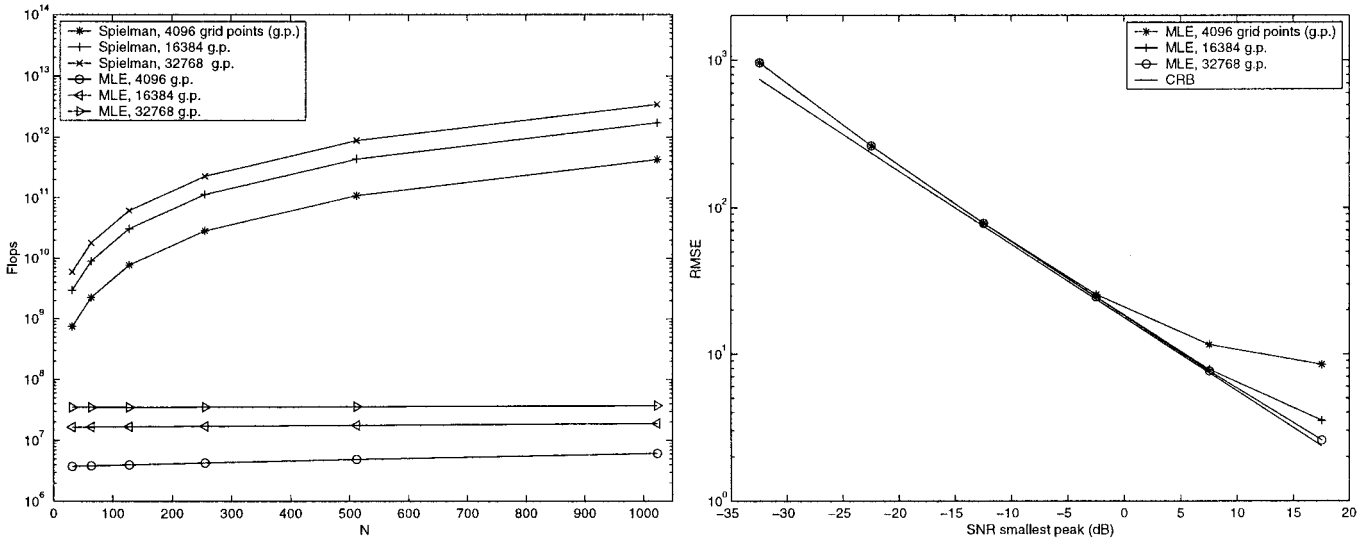

FIG. 4. Left: Number of floating point operations (flops) corresponding to using a grid search for the Spielman method and MLE as a function of the number of data points. Right: Sum of estimated amplitude RMSE for MLE as a function of SNR. Results were obtained from 1000 simulation runs using 128 data points.

prediction that our MLE has an accuracy close to the CRB. Furthermore, the efficient implementation of the MLE made its computational complexity several orders of magnitude less than that corresponding to the Spielman method. The choice of a local or global optimization algorithm is dependent on the availability of good starting values. However, the highly efficient evaluation of the LF for the MLE method using the FFT makes it possible to benefit from the desirable properties of the global search with little or no increase in the computational load.

REFERENCES

1. D. Spielman, P. Webb, and A. Macovski, A statistical framework for *in vivo* spectroscopic imaging. *J. Magn. Reson.* **79,** 66–77 (1988).
2. P. Stoica, A. Jakobsson, and J. Li, Cisoid parameter estimation in the colored noise case: Asymptotic Cramér–Rao bound, maximum likelihood and nonlinear least-squares. *IEEE Trans. Signal Processing* **45,** 2048–2059 (1997).
3. P. Stoica and R. Moses, "Introduction to Spectral Analysis," Prentice Hall, Upper Saddle River, NJ (1997).
4. B. Porat, "A Course in Digital Signal Processing," Wiley, New York (1997).