

Comparison of Bayesian Methods for Recovering Sinusoids

Dursun Üstündağ

Department of Mathematics
 Faculty of Science and Arts,
 Marmara University, Istanbul, Turkey
dustundag@marmara.edu.tr

Mehmet Cevri

Department of Mathematics
 Faculty of Science,
 İstanbul University, Istanbul, Turkey
mcevri@gmail.com.tr

Abstract— In this paper, we study a problem of estimating parameters of sinusoids from noisy data within Bayesian inferential framework. In this context, three different computational schemes such as, Bretthorst’s integral method (BRETTHORST), Gibbs sampling (GIBBS) and parallel tempering (PT) are studied and modifications of their algorithms were tested on data generated from synthetic signals. In addition, our emphasis is given to a comparison of their performances with respect to Cramér-Rao lower bound (CRLB).

Keywords— Bayesian inference; parameter estimation; Gibbs sampling; simulated annealing; parallel tempering; Cramér-Rao lower bound; power spectral density.

I. INTRODUCTION

In many experiments, a discrete data set $\mathbf{D} = \{d_i\}_{i=1}^N$ denoted as an output of a superposition of m sinusoids is sampled at discrete times $\{t_i\}_{i=1}^N$:

$$d_i = \sum_{j=1}^m a_{c_j} \cos(\omega_j t_i) + a_{s_j} \sin(\omega_j t_i) + e(t_i), \quad (1)$$

Due to the applicability of the sinusoidal model, it has received a great interest in many fields of science [1]. Especially, the frequency parameter has been subject to extensive research since it enters the signal model in non-linear fashion. The term $e(t)$ represents a random process at time t , due to measurement error. Then Equation (1) can commonly be written in the matrix-vector form:

$$\mathbf{D} = \mathbf{G}\mathbf{a} + \mathbf{e}, \quad (2)$$

where \mathbf{D} is $(N \times 1)$ matrix of data points; \mathbf{e} is $(N \times 1)$ matrix of independent identically distributed Gaussian noise samples with variance σ^2 ; \mathbf{G} is $(N \times 2m)$ matrix whose each column is a basis function evaluated at each point of time series and \mathbf{a} is $(2m \times 1)$ matrix whose components are arranged in order of coefficients of cosine and sine terms.

The vector $\boldsymbol{\theta} = \{a_{c_j}, a_{s_j}, \omega_j\}_{j=1}^m, \sigma\}$ consists of parameters

of signals and noise. Then, the goal of data analysis is usually to infer it from \mathbf{D} .

We consider here three Bayesian approaches [1, 2, 4, 5] based on different numerical procedures for estimating parameters of noisy sinusoids. Although different researchers suggested different Bayesian approaches [6, 7, 9], there has been a little work about a comparison of their performances. Therefore, a series of simulation studies with variations in the signal to noise ratio (SNR) and the length of the data sampling (N) are set up and the performances of the methods are compared with the CRLB[1,11] which is defined to be a limit on the best possible performance achievable given a dataset.

II. BAYESIAN DATA ANALYSIS

To estimate the parameters of signals from \mathbf{D} , the methods based on Bayesian inference provide a mathematical foundation for making inferences about them. Therefore, the basic relationship quantifying inferences is given by Bayes’ rule [3, 12, 13]:

$$p(\boldsymbol{\theta} | \mathbf{D}, I) = \frac{p(\boldsymbol{\theta} | I) p(\mathbf{D} | \boldsymbol{\theta}, I)}{p(\mathbf{D} | I)}, \quad (3)$$

where I represents the prior information; $p(\boldsymbol{\theta} | I)$ is the prior PDF of the parameter vector $\boldsymbol{\theta}$ that encapsulates our state of knowledge of the parameters before the receipt of the measurements \mathbf{D} ; $p(\mathbf{D} | \boldsymbol{\theta}, I)$ is termed the likelihood function when considered as a function of $\boldsymbol{\theta}$, but is known as the sampling distribution when considered as a function of \mathbf{D} ; $p(\boldsymbol{\theta} | \mathbf{D}, I)$ is the posterior PDF of $\boldsymbol{\theta}$, that corresponds to the update of $p(\boldsymbol{\theta} | I)$ incorporating the knowledge gained about $\boldsymbol{\theta}$ after the receipt of the observations \mathbf{D} and $p(\mathbf{D} | I)$ is termed as an evidence which is a measure of the probability but, it is constant in parameter estimation. In that case, the posterior PDF of $\boldsymbol{\theta}$ given in Eq. (3) becomes

$$p(\boldsymbol{\theta} | \mathbf{D}, I) \propto p(\boldsymbol{\theta} | I) p(\mathbf{D} | \boldsymbol{\theta}, I). \quad (4)$$

To proceed further in the specification of $p(\boldsymbol{\theta}|\mathbf{D}, I)$, we now need to assign a functional form for $p(\boldsymbol{\theta}|I)$ and $p(\mathbf{D}|\boldsymbol{\theta}, I)$. Because different methods use different prior PDFs, we postpone discussing their assignments to the next sections but, in the case of independent measurements the assignment of a functional form for $p(\mathbf{D}|\boldsymbol{\theta}, I)$ becomes

$$p(\mathbf{D}|\boldsymbol{\theta}, I) = \left(2\pi\sigma^2\right)^{-\frac{N}{2}} \exp\left(-\frac{1}{2}\chi^2(\boldsymbol{\theta})\right), \quad (5)$$

where

$$\chi^2(\boldsymbol{\theta}) = \sum_{i=1}^N \left(\frac{d(t_i) - f(t_i, \boldsymbol{\theta})}{\sigma}\right)^2. \quad (6)$$

After making assignments to the prior and posterior PDF, the problem turns out to search $\boldsymbol{\theta}$ in a parameter space \mathfrak{S} :

$$\hat{\boldsymbol{\theta}} = \arg \max_{\boldsymbol{\theta} \in \mathfrak{S}} \{p(\boldsymbol{\theta}|\mathbf{D}, I)\}. \quad (7)$$

III. BAYESIAN METHODS

A. Bretthorst's Integral Method

Let us rewrite the joint posterior PDF of all parameters in Equation (4):

$$p(\boldsymbol{\omega}, \mathbf{a}_c, \mathbf{a}_s, \sigma^2 | \mathbf{D}, I) \propto p(\mathbf{D} | \boldsymbol{\omega}, \mathbf{a}_c, \mathbf{a}_s, \sigma^2, I) \times p(\boldsymbol{\omega}, \mathbf{a}_c, \mathbf{a}_s, \sigma^2 | I). \quad (8)$$

By use of product rule, the joint prior PDF becomes

$$p(\boldsymbol{\omega}, \mathbf{a}_c, \mathbf{a}_s, \sigma^2 | I) \propto p(\mathbf{a}_c, \mathbf{a}_s | I) p(\boldsymbol{\omega} | I) p(\sigma^2 | I). \quad (9)$$

Here $\boldsymbol{\omega}$ is drawn uniformly so that $p(\boldsymbol{\omega} | I) \propto \pi^{-m}$. On the other hand, one can suppose that $p(\mathbf{a}_c, \mathbf{a}_s | I) \propto \text{constant}$ and $p(\sigma^2 | I) \propto \text{constant}$. Thus Eq. (9) is reduced to

$$p(\boldsymbol{\omega}, \mathbf{a}_c, \mathbf{a}_s, \sigma^2 | I) \propto \pi^{-m}. \quad (10)$$

By using Equations (5) and (10) and dropping constant terms the posterior PDF in Eq. (4) becomes

$$p(\boldsymbol{\omega}, \mathbf{a}_c, \mathbf{a}_s | I) \propto \exp\left(-\frac{1}{2}\chi^2(\boldsymbol{\omega}, \mathbf{a}_c, \mathbf{a}_s)\right). \quad (11)$$

In order to obtain the marginal PDF of $\boldsymbol{\omega}$, we need to take the integration of Eq. (11) with respect to the amplitudes. If σ^2 is known, then the posterior PDF of $\boldsymbol{\omega}$ is given by

$$p(\boldsymbol{\omega} | \mathbf{D}, I) \propto \exp\left(-\frac{m\overline{h^2}}{2\sigma^2}\right), \quad (12)$$

where $h_l = \sum_{i=1}^N d_i H_l(t_i)$, ($l=1, \dots, 2m$) is a projection of data onto new orthogonal model functions $H_l(t)$:

$$H_j(t) = \left(\lambda_j\right)^{-1/2} \sum_{l=1}^{2m} \varphi_{jl} G_l(t, \boldsymbol{\omega}). \quad (13)$$

Here φ_{lj} represents the j th component of the l th normalized eigenvector of the matrix $\boldsymbol{\Omega} = \mathbf{G}^T \mathbf{G}$, with λ_l as the corresponding eigenvalue. If there is no prior information about the noise, by using Jeffreys' prior [14] $p(\sigma^2 | I) = \sigma^{-2}$ and integrating the expression in Eq. (11) with respect to σ^2 we obtain

$$p(\boldsymbol{\omega} | \mathbf{D}, I) \propto \left(1 - \frac{m\overline{h^2}}{N\mathbf{D}^2}\right)^{\frac{m-N}{2}}. \quad (14)$$

This is the form of the "Student's t distribution" with $(N - m)$ degree of freedom.

The approach summarized above requires analytical or numerical approximation of integrals which is not given here but, we refer to Bretthorst's work [6]. [Consequently, Bayesian parameter estimation problem turns into maximization of the posterior PDF given in Eqs (13) and (14). Unfortunately, conventional algorithms [15, 16] based on the gradient direction fail to converge. Even when they converge, there is no assurance that they have found a global, rather than a local maximum. This is because the log of the posterior PDF is so sharply peaked and highly nonlinear function of $\boldsymbol{\omega}$. To overcome this problem, Bretthorst used a pattern search algorithm described by Hook-Jeeves [17] but, we found out that this approach does not converge unless the starting point is much closer to the optimum $\boldsymbol{\omega}$. Therefore, we combined it with a simulated annealing (SA) algorithm [18, 19] to obtain a global maximum of the posterior PDF of $\boldsymbol{\omega}$. For detail information, we refer to our papers and book's chapter in [26, 27, 28, 32].

B. Gibbs Sampling

In order to avoid solving the difficult multivariate maximization problem in Section III.A, an alternative way proposed by Dou and Hodgson [7] combines Gibbs sampling (GS) with Bayesian inference. We extend this derivation for multiple frequency signals and summarize it below, but refer to their papers [7, 8] for detail information.

Assume that σ^2 is known and there is no any specific information about $\{\boldsymbol{\omega}, \mathbf{a}_c, \mathbf{a}_s\}$. Then Equation (8) turns out to be the following form:

$$p(\boldsymbol{\omega}, \mathbf{a}_c, \mathbf{a}_s | \mathbf{D}, \sigma^2, I) \propto p(\mathbf{D} | \boldsymbol{\omega}, \mathbf{a}_c, \mathbf{a}_s, \sigma^2, I), \quad (15)$$

because of $p(\boldsymbol{\omega}, \mathbf{a}_c, \mathbf{a}_s | I) \propto \text{constant}$. Suppose also that a_{c_j} is the only unknown parameter among $\{\mathbf{a}_{c_j}, \mathbf{a}_s, \boldsymbol{\omega}\}$, where $\mathbf{a}_{c_j} = \{a_{c_1}, \dots, a_{c_{j-1}}, a_{c_{j+1}}, \dots, a_{c_m}\}$. If the distribution of noise is known as a-priori, the conditional PDF of a_{c_j} is considered to be as a univariate normal distribution:

$$p(a_{c_j} | \mathbf{a}_{c_j}, \mathbf{a}_s, \boldsymbol{\omega}, \mathbf{D}, \sigma^2) \propto N(\hat{a}_{c_j}, \sigma^2 (\mathbf{X}_{c_j}^T \mathbf{X}_{c_j})^{-1}), \quad (16)$$

where

$$\hat{a}_{c_j} = \frac{\hat{\mathbf{D}}^{(1)} \mathbf{X}_{c_j}}{\mathbf{X}_{c_j}^T \mathbf{X}_{c_j}}, \quad \mathbf{X}_{c_j} = \begin{bmatrix} \cos(\omega_j t_1) \\ \vdots \\ \cos(\omega_j t_N) \end{bmatrix} \quad (17)$$

and $\hat{\mathbf{D}}^{(1)} = \{d_i^{(1)}\}_{i=1}^N$ whose components are defined by

$$\hat{d}_i^{(1)} = d_i - \sum_{l=1}^m a_{c_l} \cos(\omega_l t_i) \delta_{lj} + a_{s_l} \sin(\omega_l t_i), \quad (18)$$

where $\delta_{lj} = \begin{cases} 1 & l \neq j \\ 0 & l = j \end{cases}$ helps to eliminate the contribution of the cosine term of the j th sinusoid. However, in some cases where σ^2 is unknown the joint posterior PDF of $\{\boldsymbol{\omega}, \mathbf{a}_c, \mathbf{a}_s\}$ and σ^2 can be implemented in the form:

$$p(\boldsymbol{\omega}, \mathbf{a}_c, \mathbf{a}_s, \sigma^2 | \mathbf{D}, I) \propto p(\mathbf{D} | \boldsymbol{\omega}, \mathbf{a}_c, \mathbf{a}_s, \sigma^2, I) p(\sigma^2 | I) \quad (19)$$

We eliminate it by taking Jeffreys' prior [13] and integrating the expression in Eq. (16) with respect to σ^2 and obtain a univariate Student's t distribution:

$$p(a_{c_j} | \mathbf{a}_{c_j}, \mathbf{a}_s, \boldsymbol{\omega}, \mathbf{D}, \sigma^2) \propto T(\hat{a}_{c_j}, s_{a_{c_j}} (\mathbf{X}_{c_j}^T \mathbf{X}_{c_j})^{-1}, N-1), \quad (20)$$

with

$$s_{a_{c_j}} = \frac{1}{N-1} (\hat{\mathbf{D}}^{(1)} - \hat{a}_{c_j} \mathbf{X}_{c_j})^T (\hat{\mathbf{D}}^{(1)} - \hat{a}_{c_j} \mathbf{X}_{c_j}). \quad (21)$$

In a similar way, the conditional PDF of a_{s_j} is given in the form of Eqs (17) and (20) but, \hat{a}_{s_j} and $\mathbf{X}_{a_{s_j}}$ are obtained by replacing a_{s_j} and sine terms in Eqs (17) and (21) with a_{c_j} and cosine terms, respectively. In Eqs (17) and (21), $\hat{\mathbf{D}}^{(1)}$ term is calculated by eliminating the contribution of the sine term of the j th sinusoid instead of cosine term in Eq. (18).

To be able to use the theory of GS algorithm for $\boldsymbol{\omega}$, we need Taylor series expansion of $\chi^2(\boldsymbol{\omega})$ at $\hat{\boldsymbol{\omega}}$:

$$\chi^2(\boldsymbol{\omega}) \approx \chi^2(\hat{\boldsymbol{\omega}}) + \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m \frac{\partial^2 \chi^2(\boldsymbol{\omega})}{\partial \omega_i \partial \omega_j} \Big|_{\hat{\boldsymbol{\omega}}} (\omega_i - \hat{\omega}_i) (\omega_j - \hat{\omega}_j) + \dots, \quad (22)$$

where $\hat{\boldsymbol{\omega}} = \arg \min_{\boldsymbol{\omega} \in [0, \pi]^m} \chi^2(\boldsymbol{\omega})$. Then the conditional PDF of ω_j turns out to be in the form:

$$p(\omega_j | \boldsymbol{\omega}_{-j}, \mathbf{a}_c, \mathbf{a}_s, \mathbf{D}, \sigma^2) \propto \exp\left(-\frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m \frac{\partial^2 \chi^2(\boldsymbol{\omega})}{\partial \omega_i \partial \omega_j} \Big|_{\hat{\boldsymbol{\omega}}} (\omega_i - \hat{\omega}_i) (\omega_j - \hat{\omega}_j)\right) \propto N(\hat{\omega}_j, \sigma^2 (\mathbf{X}_{\omega_j}^T \mathbf{X}_{\omega_j})^{-1}) \quad (23)$$

where

$$\mathbf{X}_{\omega_j} = \begin{bmatrix} -a_{c_j} t_1 \sin(\hat{\omega}_j t_1) + a_{s_j} t_1 \sin(\hat{\omega}_j t_1) \\ \vdots \\ -a_{c_j} t_N \sin(\hat{\omega}_j t_N) + a_{s_j} t_N \sin(\hat{\omega}_j t_N) \end{bmatrix}. \quad (24)$$

If σ^2 is unknown, Equation (23) becomes

$$p(\omega_j | \boldsymbol{\omega}_{-j}, \mathbf{a}_c, \mathbf{a}_s, \mathbf{D}, \sigma^2) \propto T(\hat{\omega}_j, s_{\omega_j}^2 (\mathbf{X}_{\omega_j}^T \mathbf{X}_{\omega_j})^{-1}, N-1). \quad (25)$$

with

$$s_{\omega_j}^2 = \frac{1}{N-1} (\mathbf{D} - \hat{\mathbf{D}})^T (\mathbf{D} - \hat{\mathbf{D}}), \quad (26)$$

where the components of $\hat{\mathbf{D}}$ is defined by the use of the current estimated values of parameters. In numerical calculations, a systematic form of GS algorithm proceeds in the following manner. With arbitrary starting values $\{\mathbf{a}_{c,0}, \mathbf{a}_{s,0}, \boldsymbol{\omega}_0\}$ random drawings are chosen from the full conditional PDFs described above as follows:

$$\begin{aligned} a_{c_j,1} &\sim p(a_{c_j} | \{a_{c_1,1}, \dots, a_{c_{j-1},1}, a_{c_{j+1},0}, \dots, a_{c_m,0}\}, \mathbf{a}_{s,0}, \boldsymbol{\omega}_0, \mathbf{D}) \\ a_{s_j,1} &\sim p(a_{s_j} | \mathbf{a}_{c_j,1}, \{a_{s_1,1}, \dots, a_{s_{j-1},1}, a_{s_{j+1},0}, \dots, a_{s_m,0}\}, \boldsymbol{\omega}_0, \mathbf{D}) \\ \omega_{j,1} &\sim p(\omega_j | \mathbf{a}_{c_j,1}, \mathbf{a}_{s_j,1}, \{\omega_{1,1}, \dots, \omega_{j-1,1}, \omega_{j+1,0}, \dots, \omega_{m,0}\}, \mathbf{D}) \end{aligned} \quad (27)$$

After the first iteration, we get $\{\mathbf{a}_{c,1}, \mathbf{a}_{s,1}, \boldsymbol{\omega}_1\}$. Repeating this procedure K times, we obtain $\{\mathbf{a}_{c,K}, \mathbf{a}_{s,K}, \boldsymbol{\omega}_K\}$. In Bayesian context, for a large enough K the joint PDF can be replaced by the conditional PDF so that $a_{c_j,K}$, $a_{s_j,K}$ and $\omega_{j,K}$ become random variables. Then we draw M random samples of $\{\mathbf{a}_{c,K}^l\}_{l=1}^M$, $\{\mathbf{a}_{s,K}^l\}_{l=1}^M$ and $\{\boldsymbol{\omega}_K^l\}_{l=1}^M$ from their marginal PDFs, respectively and using them, we obtain the estimates about the corresponding parameters.

C. Parallel Tempering

The GIBBS (GS) overcomes some problems associated with BRETTHORST but, it faces with a problem of simulating frequencies from their conditional PDF in Eq. (23) and (25). Because their direct simulations are impossible, they are approximated by a Taylor expansion at $\hat{\omega}$ that can be difficult with a lower SNR to obtain. To overcome these problems and provide a flexibility of choice of priors, we implement PT method [9, 10] that is originated with Swendsen [20], extended by Geyer [21] and later developed and successfully used in a number of general optimization problems.

Suppose that a suitable prior PDF for ω is given by

$$p(\omega|I) = \begin{cases} \frac{1}{3(\pi)} + \frac{1}{3(\omega+c)\ln(\frac{\omega+c}{c})} + \frac{1}{3\sqrt{(2\pi\sigma_\omega^2)}} \exp\left(-\frac{(\omega-\mu_\omega)^2}{2\sigma_\omega^2}\right) & 0 \leq \omega \leq \pi \\ 0 & \text{otherwise} \end{cases} \quad (28)$$

where c is the mean of the standard deviation of the noise vector and the logarithmic term ensures that the prior PDF is normalized. In a similar way, the prior PDF for the amplitude a_{c_j} can also be taken as

$$p(a_{c_j}|I) = \begin{cases} \frac{1}{2(D_{\max}-D_{\min})} + \frac{1}{2\sqrt{2\pi\sigma_{a_j}^2}} \exp\left(-\frac{(a_{c_j}-\mu_{a_j})^2}{2\sigma_{a_j}^2}\right) & D_{\min} \leq a_{c_j} \leq D_{\max} \\ 0 & \text{otherwise} \end{cases}, \quad (29)$$

where $\mu_{a_c} = \frac{D_{\max} + D_{\min}}{2}$, and $\sigma_{a_s} = \frac{D_{\max} - D_{\min}}{3}$. A similar prior PDF is also assigned to a_{s_j} by replacing a_{c_j} with a_{s_j} in Eq. (29). Putting Eqs (28) and (29) into Eq. (5), we obtain the posterior PDF $p(\omega, \mathbf{a} | \mathbf{D}, I)$, denoted here as a tempered PDF $\pi(\omega, \mathbf{a} | \mathbf{D}, \beta, I)$:

$$\pi(\omega, \mathbf{a} | \mathbf{D}, \beta, I) = p(\omega, \mathbf{a} | I) e^{\beta \ln(p(\mathbf{D}|\omega, \mathbf{a}, I))} \quad (30)$$

Now, the problem turns out finding the parameter values that maximize Eq. (30) using the PT algorithm. It consists of two main updating steps. The first one is the state update of each chain in which there exists $n\beta$ multiple copies of Markov chain Monte Carlo (MCMC) simulations[20], which are run simultaneously in parallel each at different values of tempering parameter β . The second one is the swapping update between two neighboring chains at each n_s step.

MCMC algorithm generates desired samples $X_t = \{\omega, \mathbf{a}\}$ by constructing a kind of random walk in a model parameter space. Briefly, we pick a proposed value for X_{t+1} drawn from a proposal PDF $q(X_{t+1}|X_t)$, which is considered to be a multivariate Normal distribution with a mean equaled to current sample X_t and a deviation σ_X named as a step size which is taken to be a square root of the CRLB for estimated

parameters [21]. Let a random number u_1 drawn from a uniform distribution $U_{(0,1)}$. Then X_{t+1} is accepted as a new value for X_t satisfying

$$u_1 \leq \alpha(X_t, X_{t+1}), \quad (31)$$

with a probability:

$$\alpha(X_t, X_{t+1}) = \min\left\{1, \frac{p(X_{t+1} | \mathbf{D}, I)}{p(X_t | \mathbf{D}, I)}\right\}. \quad (32)$$

This process continues until a random number u_2 drawn from $U_{(0,1)}$ satisfies $u_2 \leq 1/n_s$. Then, at time t the simulation β_i in the state $X_{t,i}$ and the simulation β_{i+1} in the state $X_{t,i+1}$ can be interchanged if a random number u_3 drawn from $U_{(0,1)}$ satisfies

$$u_3 \leq \alpha(X_{t,i}, X_{t,i+1}), \quad (1 \leq i \leq n\beta - 1), \quad (33)$$

where

$$\alpha(X_{t,i}, X_{t,i+1}) = \min\left\{1, \frac{\pi(X_{t,i+1} | \mathbf{D}, \beta_i, I) \cdot \pi(X_{t,i} | \mathbf{D}, \beta_{i+1}, I)}{\pi(X_{t,i} | \mathbf{D}, \beta_i, I) \cdot \pi(X_{t,i+1} | \mathbf{D}, \beta_{i+1}, I)}\right\}. \quad (34)$$

This is called a probability of the swap acceptance. As expected, after an initial burn-in period this proposed method generates samples X_t with a PDF equal to the desired posterior PDF $p(X_t | \mathbf{D}, I)$. Finally, inferences about parameters are based on these samples drawn from the output corresponding to the lowest temperature chain ($\beta = 1$). For detail information about the algorithm, we refer to our recent papers in [30, 31].

IV. COMPUTER SIMULATIONS

To demonstrate the proposed approaches with examples which are used by previous researchers [6-9], we firstly have created $N = 512$ data samples according to a multiple harmonic frequency signal model:

$$d_i = \cos(0.1t_i + 1) + 2\cos(0.15t_i + 2) + 5\cos(0.3t_i + 3) + 2\cos(0.31t_i + 4) + 3\cos(t_i + 5) + e_i \quad (35)$$

Here i runs in a symmetric time interval from $-T$ to T ($N = 2T + 1$) and $e_i \sim N(0,1)$. Then, we carried out the Bayesian analysis of the simulated data.

All three methods, introduced here were coded in Mathematica, which contains a large collection of built-in functions so that it provides much shorter computer codes than those written in C or FORTRAN. They were run on a workstation with four processors which of each has got Intel Core 2 Quad CPU. For each method, initial values for the parameters are required to start iteration so that we first identified the locations of the peaks of the Fourier power spectral density with the greatest magnitude as an initial estimate for angular frequencies and then we carried on

calculating the coefficients \mathbf{a}_c and \mathbf{a}_s as initial values for amplitudes, respectively. In the case where the deviation of noise is unknown, the output of the computer simulation is illustrated in Table 1. It can be seen that the estimated five frequencies and their corresponding amplitudes for the model signal are recovered very well and quoted as (*value*) \pm (*standard deviation*).

The usual way the result from a spectral analysis is displayed is in the form of a power spectral density (PSD) [3, 6]. In order to do it, we focus on its definition that shows how much power is contained in a unit frequency. According to Bretthorst [6] the Bayesian PSD is defined as the expected value of the power of the signals over the joint posterior PDF. In Fourier transform spectroscopy this is typically taken as the squared magnitude of the Discrete Fourier Transform (DFT) of the data so that Bayesian and Fourier PSDs for the signal are shown in Figure 1. Fourier PSD in Figure 1 (b) indicates only four of five well separated frequencies but, Bayesian PSDs in Figure 1 (c), (d) and (e) show that all the five frequencies have been well separated while the height is indicative of the resolution. A comparison of them implies that frequencies obtained by using the proposed methods are separated very well although the separation of the sinusoids is less than the Rayleigh resolution [3, 6, 24]. These results demonstrate ability of resolving closely spaced frequencies using the methods based on Bayesian inference.

Moreover, we initially assumed that the values of the random noise in data were drawn from the Gaussian distribution. Figure 2 shows the exact and estimated PDF of the random noise in data using each of the proposed methods. It is seen that the estimated (dotted) PDF is closer to the true (solid) PDF and the histogram of the errors, which is known as nonparametric estimator of the PDF of the noise is also much closer to the true PDF.

When we look the Tables (1) and also PSDs in Figure (1), all three methods give similar results. Therefore, in order to evaluate their frequency estimation performances, computer simulations were performed and comparisons were also made with the CRLB for different data length and various SNR values. We first generated 64 data samples with different noise levels from a single real tone frequency signal model. After 50 independent trials under the same SNR, the mean square errors (MSE) of the estimated frequency were obtained by each of methods so that their logarithmic values were plotted as a function of the SNR, which varies from 0 to 20 dB and shown in Figure 3(a). It can be seen that BRETTHORST, GIBBS and PT estimators have thresholds about 3 dB, 4 dB and 5 dB of the SNR, respectively and they follow nicely with the CRLB after 5 dB. As expected, larger SNR gives smaller MSE. To see changes in performances we calculated an efficiency parameter η [34], defined as $\eta = (\text{CRLB} / \text{MSE}) \times 100$, that indicates closeness to the CRLB. Table 2 contains the MSEs and the efficiency parameter values η for the frequency estimation obtained at SNR = 20dB and $N = 64$. It is seen

that BRETTHORST's efficiency value is much closer to the CRLB than that of the others so that BRETTHORST is better estimator at SNR = 20dB than the others.

The above argument treats with only the case in which a data set of $N = 64$ is used for estimation. Therefore, one may ask how to estimation accuracy varies with N . To answer it, we secondly generate data sets with different lengths from a single real tone frequency signal model. After independent 20 trials under SNR = 10dB, the MSEs of the estimated frequency were obtained by each of methods so that their logarithmic values were plotted as a function of N , that vary between 64 and 676 and shown in Figure 3(b). From these results, larger N makes higher accuracy but, requires larger consumption of CPU time. It is also expected that lower SNR and less value for N deteriorate their performances. Table 2 also contains the MSEs and efficiency values η for the frequency estimation obtained in the case of $N = 676$. According to the CRLB all three methods have almost same performance under different data lengths. As a result, the performances of the proposed methods are close to optimal with a minimum variance, which is close to the predictions made by the CRLB.

CONCLUSIONS

In this paper, we have partly improved three Bayesian approaches using different numerical procedures for estimating parameters of multiple sinusoids embedded in Gaussian noise. Overall results show that Bayesian approach can not only give us the best estimates for the parameters but, also tell us uncertainties associated with their estimated values. Although they are more computationally intensive than the usual power spectrum methods, they are the best suited to those datasets where the values are noisy and aliasing that causes difficulties in the interpretation of the power spectrum. Moreover, the performances of the proposed algorithms for single frequency estimation are demonstrated via computer simulations at different data lengths and signal to noise ratios.

ACKNOWLEDGEMENTS

This work is some part of the projects, whose names are "Positron lifetimes spectrum analysis with Bayesian statistical inference" with a number FEN - A -060510-0135 and "Comparison of Bayesian methods for recovering sinusoids" with a number FEN-D-150513-0187 supported by Marmara University, Istanbul, Turkey.

REFERENCES

- [1]. Kay S.M.: "Fundamentals of Statistical Signal Processing: Estimation Theory", Prentice Hall, 1993.

- [2]. E.T. Jaynes: "Bayesian Spectrum and Chirp Analysis", *In Proceedings of the Third Workshop on Maximum Entropy and Bayesian Methods* (1983), Ed. C. Ray Smith and D. Reidel, Boston 1987, 1-37.
- [3]. E.T. Jaynes: "Probability Theory: The Logic of Science", Cambridge University Press, Cambridge, UK, 2003.
- [4]. D.S. Sivia, J. Skilling: "Data Analysis: A Bayesian Tutorial", Oxford University Press Inc., New York, 2006.
- [5]. G. D' Agostini: "Bayesian inference in processing experimental data: principles and basic applications", *Report on Processing Physics* 2003; **66**: 1383-1419.
- [6]. G.L. Bretthorst: "Bayesian Spectrum Analysis and Parameter Estimation", *Lecture Notes in Statistics*, Springer-Verlag Berlin Heidelberg New York, 1997.
- [7]. L. Dou and R. J. W. Hodgson: "Bayesian inference and Gibbs sampling in spectral analysis and parameter estimation I", *Inverse Problem* 1995; **11**: 1069-1085.
- [8]. L. Dou and R. J. W. Hodgson: "Bayesian inference and Gibbs sampling in spectral analysis and parameter estimation II", *Inverse Problem* 1995; **11**: 121-137.
- [9]. J. P. Gregory: "Bayesian Logical Data Analysis for the Physical Science", Cambridge University Press, United Kingdom, 2005.
- [10]. Gregory P. C.: "A Bayesian Kepler Periodogram Detects a Second Planet in HD 208487", *Mon. Not. R. Astron. Soc.* 2006; 1-14.
- [11]. Kay S.M.: "Accurate frequency estimation at low signal-to-noise ratio", *IEEE Transactions on Acoustic, Speech and Signal Processing* 1984; **ASSP-32**: 540-547.
- [12]. H.L. Harney: "Bayesian Inference: Parameter Estimation and Decisions", Springer-Verlag, Berlin Heidelberg, 2003
- [13]. M. Bernardo, A.F.M. Smith: "Bayesian Theory", Willey Series in Probability and Statistics, New York, 2000.
- [14]. H. Jeffreys: "Theory of Probability", Oxford University press, 1988.
- [15]. D. W. Marquardt: "An algorithm for least-squares estimation of nonlinear parameters", Vol. 11(1963) 431-441.
- [16]. W.H Press, B.P Flannery, S.A Teukolshy and W.T. Vetterling: "Numerical Recipes in C: The Art of Computing", Second Edition, Cambridge University Press, 1995.
- [17]. T.R. Hooke and T.A. Jevees.: "Direct search solution of numerical and statistical problems", *Journal of Association of Computer Machinery* 1962; **5**: 212-229.
- [18]. W.L. Goffe, G.D. Ferier and J. Rogers: "Global optimization of statistical functions with simulated annealing", *Journal of Econometrics* 1994; **60**: 65-100.
- [19]. A. Corana, M. Marchesi, C. Martini, S. Ridella: "Minimizing multimodal functions of continuous variables with the simulated annealing algorithm", *ACM Transactions on Mathematical Software* 1987; **13**: 262-280.
- [20]. R. H. Swendsen and J.S. Wang: "Replica Monte Carlo simulation of spin-glasses", *Physical Review of Letters* 1986; **57**: 2607.
- [21]. C. J. Geyer. "In computing science and statistics, *Proceedings of the 23rd Symposium on the Interface*", New York, 1991.
- [22]. N. Metropolis, A. Rosenbluth, M. Rosenblatt, A. Teller and E. Teller: "Equation of states calculations by fast computing machines", *Journal of Chemical Physics* 1953; **21**:1087-1092.
- [23]. J. Ireland: "Simulated annealing and Bayesian posterior distribution analysis applied to spectral emission line fitting", *Journal of Solar Physics* 2007; **243**: 237-252.
- [24]. A. Schuster: "The Periodogram and its Optical Analogy", *Proceedings of the Royal Society of London* 1905; **77**: 136.
- [25]. J.W. Cooley, J.W. Tukey: "An algorithm for the machine calculation of complex Fourier series", *Mathematics of Computation* 1965; **19**: 297-301.
- [26]. D. Üstündağ and M. Cevri: "Estimating parameters of sinusoids from noisy data using Bayesian inference with simulated annealing", *Wseas Transactions on Signal Processing* 2008; **7**: 432-441.
- [27]. D. Üstündağ and M. Cevri: "Bayesian Parameter Estimation of Sinusoids with Simulated Annealing", *8th Wseas International Conference On Signal Processing, Computational Geometry And Artificial Vision (Iscgav'08)*, Rhodes, Greece 2008; 106-112,
- [28]. D. Üstündağ and M. Cevri: "Recovering sinusoids from noisy data using Bayesian inference with simulated annealing" *Mathematical and Computational Applications* 2011; **16**: 382-391.
- [29]. L. Tierney: "Markov chains for exploring posterior distributions", *The Annals of Statistics* 1994; **22**: 1701-1728.
- [30]. M. Cevri and D. Üstündağ: "Bayesian Estimation of Sinusoidal Signals via Parallel Tempering", *Proceedings of the 3rd WSEAS international symposium on Wavelets theory and applications in applied mathematics, signal processing & modern science*, Istanbul, Turkey, 2009, 67-72.
- [31]. M. Cevri and D. Üstündağ: "Bayesian recovery of sinusoids from noisy data with parallel tempering", *IET Signal Processing* 2012; **6**: 673-683, 2012.
- [32]. D. Üstündağ and M. Cevri: "Bayesian recovery of sinusoids with simulated annealing" *Simulated Annealing - Advances, Applications and Hybridizations*, ISBN 978-953-51-0710-1, 2012.
- [33]. D. Üstündağ: "Recovering Sinusoids from Data Using Bayesian inference with RJMCMC", *Seventh International Conference on Natural Computation. Shanghai, China*, 2011; 1850-1854.
- [34]. B. Ristic, S. Arulampalam and N. Gordon: "Beyond the Kalman Filter: Particle Filters for Tracking Applications", Artech House, London, 2004.

Creative Commons Attribution License 4.0 (Attribution 4.0 International, CC BY 4.0)

This article is published under the terms of the Creative Commons Attribution License 4.0
https://creativecommons.org/licenses/by/4.0/deed.en_US

TABLES AND FIGURS

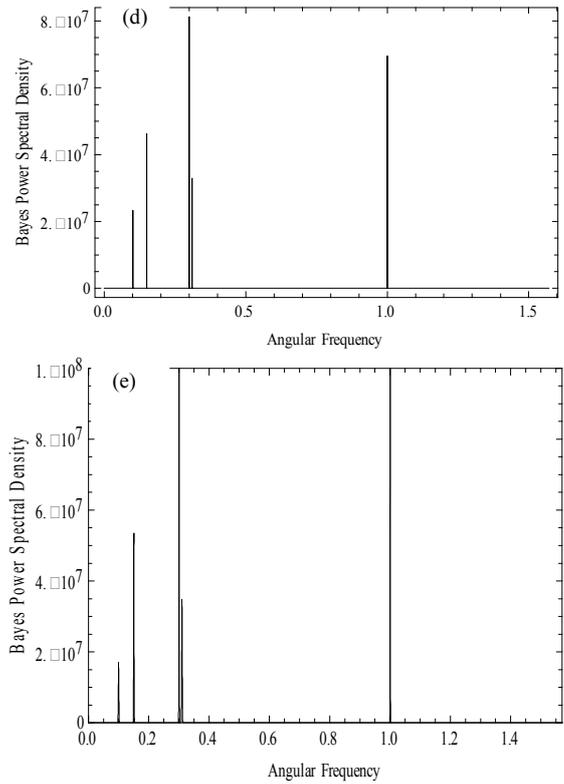
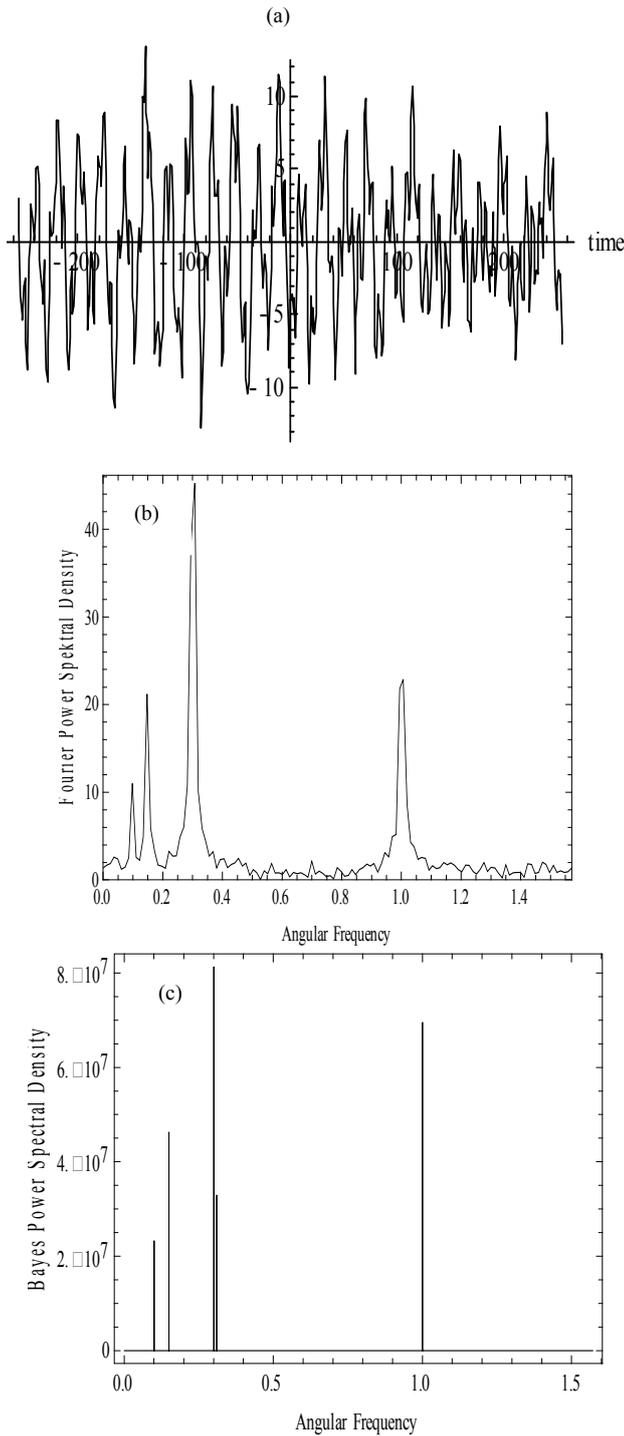


Figure 1 Spectral analysis of multiple frequencies signal model: (a) Observed data (b) Fourier power spectral density (c) Bretthorsts’ integral method (d) Gibbs sampling (e) Parallel Tempering.

Table 3. Performance comparison of Bayesian methods for single frequency estimation

PARAM	TRUE	BRETTHO	GIBBS	PT
ETERS	VALUES	RSTS		
ω_1	0.1000	0.1001 ± 0.0005	0.0989 ± 0.0004	0.1005 ±0.0004
ω_2	0.1500	0.1502 ± 0.0003	0.1499 ± 0.0002	0.1502 ±0.0002
ω_3	0.3000	0.2997 ± 0.0002	0.3001 ± 0.0001	0.2998 ±0.0001
ω_4	0.3100	0.3097 ± 0.0004	0.3102 ± 0.0002	0.3094 ±0.0003
ω_5	1.000	0.9999 ± 0.0002	1.0000 ± 0.0001	1.0001 ±0.0001
B_1	0.5403	0.5669 ± 0.0645	0.5740 ± 0.0657	0.5645 ±0.0573
B_2	-0.8414	-0.8922 ± 0.0646	-0.7589 ± 0.0662	-0.7664 ±0.0586
B_3	-0.8322	-0.8873 ± 0.0662	-0.8922 ± 0.0654	-0.7985 ±0.0604
B_4	-1.8185	-1.8215 ± 0.0661	-1.8140 ± 0.0662	-1.8304 ±0.066
B_5	-4.9499	-4.7931 ± 0.0649	-5.0060 ± 0.0658	-4.9055 ±0.0934
B_6	-0.7056	-0.7841 ± 0.0655	-0.7615 ± 0.0655	-0.7161 ±0.0915
B_7	-1.3072	-1.3671 ± 0.0653	-1.1790 ± 0.0651	-1.3965 ±0.1178
B_8	1.5136	1.4932 ±0.0628	1.4840 ± 0.0656	1.4921 ±0.0681
B_9	0.8500	0.8879 ± 0.0667	0.9373 ± 0.0655	0.8414 ±0.0651
B_{10}	2.8767	2.9218 ± 0.0649	2.9260 ± 0.0657	2.8529 ±0.0612

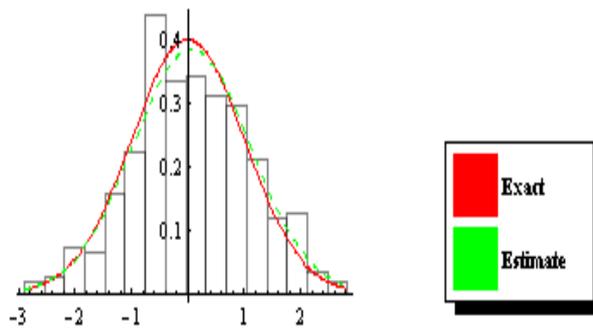


Figure 2. Comparison of exact and estimate PDFs of the noise in data

Table2. The best estimates of parameters for a multiple frequency sinusoidal signal model

Methods	SNR = 20dB		N = 672	
	MSE(dB)	Efficiency	MSE(dB)	Efficiency
GIBBS	-72.0742	110	-81.371	122
BRETTTHORST	-74.7158	106	-83.007	120
P T	-68.266	116	-85.010	117
CRLB	-79.3572	100	-99.357	100

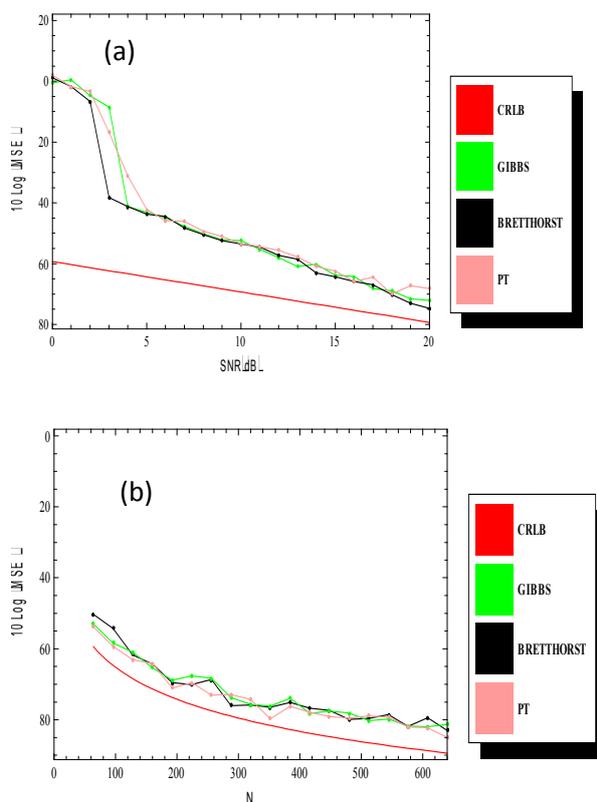


Figure 4. (a) Mean squared frequency errors for $\omega = 0.3$ versus SNR at $N = 64$. **(b)** Mean squared frequency errors for $\omega = 0.3$ versus N at SNR = 10dB.