

Performance analysis of Gibbs sampling for Bayesian extracting sinusoids

Mehmet Cevri

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

Dursun Üstündag

Department of Mathematics,
 Marmara University, Istanbul, Turkey
dustundag@marmara.edu.tr

Received: June 10, 2019. Revised: September 24, 2021. Accepted: October 14, 2021. Published: November 23, 2021.

Abstract—This paper involves problems of estimating parameters of sinusoids from white noisy data by using Gibbs sampling (GS) in a Bayesian framework. Modifications of its algorithm is tested on data generated from synthetic signals and its performance is compared with conventional estimators such as Maximum Likelihood (ML) and Discrete Fourier Transform (DFT) under a variety of signal to noise ratio (SNR) and different length of data sampling (N), regarding to Cramér-Rao lower bound (CRLB). All simulation results show its effectiveness in frequency and amplitude estimation of sinusoids.

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

I. INTRODUCTION

The sinusoidal frequency model embedded in noise is extensively important because of its wide applicability in many areas of science and engineering such as, modeling and manipulation of time-series from speech, audio to radar, seismology, nuclear magnetic resonance, communication problems and underwater acoustics[1].

We therefore address here a problem of estimating parameters of noisy sinusoids within a Bayesian inferential framework that provides a rigorous mathematical foundation for making inferences about them and a basis for quantifying uncertainties in their estimates. Under an assumption that a number of sinusoids is known a priori, several algorithms have already been applied to spectral analysis and parameter estimation problems, such as least-square fitting [2], maximum likelihood (ML)[3], discrete Fourier transform (DFT) [4,5], and a periodogram [6]. After Jayness' work [7], researchers in different fields of science have given much attention to the relationship between Bayesian inference and parameter estimation. Bretthorst and the others [8-16] have done excellent works in this area for the last sixteen years.

In this paper, we consider analysis of Gibbs sampling[11, 12] for recovering sinusoids from noisy data and compare its performance with classical estimators, regarding to Cramér-Rao lower bound (CRLB), that is a limit on the best possible performance achievable by an

unbiased estimator given a dataset[17]. For this purpose, a series of simulation studies with a variation in levels of noise and length of data sampling for a single sinusoid is set up.

II. HARMONIC SIGNAL MODEL

In many experiments, a discrete data set $\mathbf{D} = \{d_1, d_2, \dots, d_N\}^T$ denoted as an output of a physical system that we want to be modeled is sampled from an unknown function $y(t)$ at discrete times $\{t_1, \dots, t_N\}^T$:

$$\begin{aligned} d_i &= y(t_i) \\ &= f(t_i; \boldsymbol{\theta}) + e_i, \quad (i = 1, \dots, N), \end{aligned} \quad (1)$$

where $\boldsymbol{\theta}$ is a vector containing parameters that characterize behavior of physical system $f(t; \boldsymbol{\theta})$ and that are usually unknown. The term e_i is assumed to be drawn from a known random process. The choice of the model function $f(t; \boldsymbol{\theta})$ depends on the specific application, but we will consider here a superposition of k sinusoids:

$$f(t; \boldsymbol{\theta}) = \sum_{j=1}^k a_{c_j} \cos(t\omega_j) + a_{s_j} \sin(t\omega_j). \quad (2)$$

where $\{a_{c_j}, a_{s_j}\} \in \mathbb{R}^{2k}$ and $\omega_j \in (0, \pi)$ are amplitudes and angular frequencies, respectively. Hence, Equation (1) can be written in the matrix-vector form:

$$\mathbf{D} = \mathbf{G}\mathbf{a} + \mathbf{e}, \quad (3)$$

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

terms $\{a_{c_1}, a_{s_1}, \dots, a_{c_k}, a_{s_k}\}$. Then, the goal of data analysis is usually to infer $\theta = \{(a_{c_j}, a_{s_j}, \omega_j)\}_{j=1}^k$ from \mathbf{D} .

II. BAYESIAN DATA ANALYSIS

By using Bayes' rule [18,19,25], the context of the current problem can be expressed as follows:

$$p(\theta | \mathbf{D}, I) = \frac{p(\theta)p(\mathbf{D} | \theta, I)}{p(\mathbf{D})}, \quad (4)$$

where $p(\theta)$ is the prior probability density function (PDF) of the parameter vector θ that encapsulates our state of knowledge of the parameters before observing \mathbf{D} ; $p(\mathbf{D} | \theta, I)$ is called the likelihood function when considered as a function of θ , but it is known as the sampling distribution when considered as a function of \mathbf{D} . $p(\mathbf{D})$ is denoted as an evidence or the marginal likelihood and $p(\theta | \mathbf{D}, I)$ is the posterior PDF of the parameters θ of interest, which summarizes the last information about it:

$$p(\theta | \mathbf{D}, I) \propto p(\theta)p(\mathbf{D} | \theta, I). \quad (5)$$

To proceed further in the specification of the posterior PDF, we now need to assign functional forms for $p(\theta)$ and $p(\mathbf{D} | \theta, I)$. After computing $p(\theta | \mathbf{D}, I)$, the problem turns out to search a vector θ that satisfies

$$\hat{\theta} = \arg \max_{\theta \in \Theta} \{p(\theta | \mathbf{D}, I)\}, \quad (6)$$

where Θ is a parameter space.

IV. GIBBS SAMPLING

Let us first suppose that σ^2 is known and there is no any specific information about $\{\omega, \mathbf{a}_c, \mathbf{a}_s\}$ prior to the observation \mathbf{D} , then Equation (5) turns out to be the following form:

$$p(\omega, \mathbf{a}_c, \mathbf{a}_s | \mathbf{D}, \sigma^2, I) \propto p(\mathbf{D} | \omega, \mathbf{a}_c, \mathbf{a}_s, \sigma^2, I), \quad (7)$$

where $p(\omega, \mathbf{a}_c, \mathbf{a}_s) \propto \text{constant}$ as an uninformative uniform prior PDF for $\{\omega, \mathbf{a}_c, \mathbf{a}_s\}$. Secondly, in some cases the variance σ^2 is unknown so that the joint posterior PDF of $\{\omega, \mathbf{a}_c, \mathbf{a}_s\}$ and σ^2 becomes

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

Thus, by using Jeffreys prior $p(\sigma^2) = \frac{1}{\sigma^2}$ and integrating

this joint posterior PDF in (8) with respect to σ^2 , it turns into the Student's t distribution [11, 12]. Therefore, GS can give us an alternative way to find the marginal PDF for each parameter in Bayesian computation by dealing with it at each sampling step.

Suppose that a_{c_j} is the only unknown parameter among the others $\{\mathbf{a}_{c_{-j}}, \mathbf{a}_s, \omega\}$ where $\mathbf{a}_{c_{-j}} = \{a_{c_1}, \dots, a_{c_{j-1}}, a_{c_{j+1}}, \dots, a_{c_k}\}$.

Under the assumption of known distribution of the noise, the conditional PDF of a_{c_j} given that $\mathbf{a}_{c_{-j}}, \mathbf{a}_s, \omega, \mathbf{D}$ and σ^2 have already been known is a univariate Gaussian distribution:

$$p(a_{c_j} | \mathbf{a}_{c_{-j}}, \mathbf{a}_s, \omega, \mathbf{D}, \sigma^2) \propto N(\hat{a}_{c_j}, \sigma^2 (\mathbf{X}_{a_{c_j}}^T \mathbf{X}_{a_{c_j}})^{-1}), \quad (9)$$

where

$$\hat{a}_{c_j} = \frac{\hat{\mathbf{D}}^{(1)} \mathbf{X}_{a_{c_j}}}{\mathbf{X}_{a_{c_j}}^T \mathbf{X}_{a_{c_j}}}, \quad \mathbf{X}_{a_{c_j}} = \begin{bmatrix} \cos(\omega_j t_1) \\ \vdots \\ \cos(\omega_j t_N) \end{bmatrix} \quad (10)$$

and

$$\hat{\mathbf{D}}^{(1)} = \{d_1^{(1)}, d_2^{(1)}, \dots, d_N^{(1)}\} \quad (11)$$

whose components are defined by $d_i - \sum_{l=1}^k a_{c_l} \cos(\omega_l t_i) \delta_{lj} + a_{s_l} \sin(\omega_l t_i)$, ($i = 1, 2, 3, \dots, N$). The

$\delta_{lj} = \begin{cases} 1 & l \neq j \\ 0 & l = j \end{cases}$ helps to eliminate the contribution, which

comes from the cosine term of the j th sinusoid. When σ^2 is unknown, Equation (9) becomes a univariate Student's t distribution:

$$p(a_{c_j} | \mathbf{a}_{c_{-j}}, \mathbf{a}_s, \omega, \mathbf{D}, \sigma^2) \propto T(\hat{a}_{c_j}, s_{a_{c_j}} (\mathbf{X}_{a_{c_j}}^T \mathbf{X}_{a_{c_j}})^{-1}, N-1), \quad (12)$$

with

$$s_{a_{c_j}} = \frac{1}{N-1} (\hat{\mathbf{D}}^{(1)} - \hat{a}_{c_j} \mathbf{X}_{a_{c_j}})^T (\hat{\mathbf{D}}^{(1)} - \hat{a}_{c_j} \mathbf{X}_{a_{c_j}}) \quad (13)$$

When $\{\mathbf{a}_c, \mathbf{a}_{s_{-j}}, \omega\}$ is given, in a similar way, the conditional PDF of a_{s_j} given that $\mathbf{a}_c, \mathbf{a}_{s_{-j}}, \omega, \mathbf{D}$ and σ^2 have already been known is

$$p(a_{s_j} | \mathbf{a}_c, \mathbf{a}_{s_{-j}}, \omega, \mathbf{D}, \sigma^2) \propto N(\hat{a}_{s_j}, \sigma^2 (\mathbf{X}_{a_{s_j}}^T \mathbf{X}_{a_{s_j}})^{-1}), \quad (14)$$

where

$$\hat{\mathbf{a}}_{s_j} = \frac{\hat{\mathbf{D}}^{(2)} \mathbf{X}_{a_{s_j}}}{\mathbf{X}_{a_{s_j}}^T \mathbf{X}_{a_{s_j}}}, \quad \mathbf{X}_{a_{s_j}} = \begin{bmatrix} \sin(\omega_j t_1) \\ \vdots \\ \sin(\omega_j t_N) \end{bmatrix} \quad (15)$$

and

$$\hat{\mathbf{D}}^{(2)} = \{d_1^{(2)}, d_2^{(2)}, \dots, d_N^{(2)}\}, \quad (16)$$

whose components are defined by

$$\hat{d}_i^{(2)} = d_i - \sum_{l=1}^k a_{c_l} \cos(\omega_l t_i) + a_{s_l} \sin(\omega_l t_i) \delta_{ij}, \quad (i = 1, \dots, N).$$

When σ^2 is unknown, Equation (14) turns out to be

$$p(a_{s_j} | \mathbf{a}_{s_{-j}}, \mathbf{a}_c, \boldsymbol{\omega}, \mathbf{D}, \sigma^2) \propto \mathcal{T}(\hat{\mathbf{a}}_{s_j}, s_{a_{s_j}} (\mathbf{X}_{a_{s_j}}^T \mathbf{X}_{a_{s_j}})^{-1}, N-1) \quad (17)$$

with

$$s_{a_{s_j}} = \frac{1}{N-1} (\hat{\mathbf{D}}^{(2)} - \hat{\mathbf{a}}_{s_j} \mathbf{X}_{a_{s_j}})^T (\hat{\mathbf{D}}^{(2)} - \hat{\mathbf{a}}_{s_j} \mathbf{X}_{a_{s_j}}). \quad (18)$$

To be able to use the theory of GS for the nonlinear parameter $\boldsymbol{\omega}$, we need to introduce some reasonable approximations to linearize the nonlinear model function $f(t_i, \boldsymbol{\omega})$ with respect to $\boldsymbol{\omega}$ under the condition of the known amplitudes $\{\mathbf{a}_c, \mathbf{a}_s\}$. This can be done by expanding it around $\hat{\boldsymbol{\omega}}$ in a region where the posterior PDF is concentrated:

$$f(t_i, \hat{\boldsymbol{\omega}}) \cong \sum_{l=1}^k a_{c_l} \cos(\hat{\omega}_l t_i) + a_{s_l} \sin(\hat{\omega}_l t_i) + (-a_{c_j} t_i \sin(\hat{\omega}_j t_i) + a_{s_j} t_i \cos(\hat{\omega}_j t_i))(\omega_j - \hat{\omega}_j), \quad (19)$$

where $\hat{\omega}_j = \arg \min_{\hat{\boldsymbol{\omega}} \in \boldsymbol{\omega}} \sum_{i=1}^N (d_i - f(t_i, \boldsymbol{\omega}))^2$ and

$\hat{\boldsymbol{\omega}} = \{\omega_1, \dots, \omega_{j-1}, \hat{\omega}_j, \omega_{j+1}, \dots, \omega_k\}$. Thus, the conditional PDF of ω_j given that $\boldsymbol{\omega}_{-j}, \mathbf{a}_c, \mathbf{a}_s, \mathbf{D}$ and σ^2 have already been known is a univariate Gaussian distribution:

$$p(\omega_j | \boldsymbol{\omega}_{-j}, \mathbf{a}_c, \mathbf{a}_s, \mathbf{D}, \sigma^2) \propto \mathcal{N}(\hat{\omega}_j, \sigma^2 (\mathbf{X}_{\omega_j}^T \mathbf{X}_{\omega_j})^{-1}), \quad (20)$$

where

$$\mathbf{X}_{\omega_j} = \begin{bmatrix} -a_{c_j} t_1 \sin(\hat{\omega}_j t_1) + a_{s_j} t_1 \sin(\hat{\omega}_j t_1) \\ \vdots \\ -a_{c_j} t_N \sin(\hat{\omega}_j t_N) + a_{s_j} t_N \sin(\hat{\omega}_j t_N) \end{bmatrix}. \quad (21)$$

If σ^2 is unknown, Equation (20) becomes is a univariate Student's t distribution

$$p(\omega_j | \boldsymbol{\omega}_{-j}, \mathbf{a}_c, \mathbf{a}_s, \mathbf{D}, \sigma^2) \propto \mathcal{T}(\hat{\omega}_j, s_{\omega_j}^2 (\mathbf{X}_{\omega_j}^T \mathbf{X}_{\omega_j})^{-1}, N-1). \quad (22)$$

with

$$s_{\omega_j}^2 = \frac{1}{N-1} (\mathbf{D} - \hat{\mathbf{D}})^T (\mathbf{D} - \hat{\mathbf{D}}), \quad (23)$$

where $\hat{\mathbf{D}} = \{\hat{d}(t_1), \hat{d}(t_2), \dots, \hat{d}(t_N)\}$ whose components are

$$\hat{d}(t_i) = \sum_{l=1}^k a_{c_l} \cos(\hat{\omega}_l t_i) + a_{s_l} \sin(\hat{\omega}_l t_i).$$

A systematic form of GS algorithm [11,12,20] contains choosing initially arbitrary starting values $\{\mathbf{a}_c^{(0)}, \mathbf{a}_s^{(0)}, \boldsymbol{\omega}^{(0)}\}$ and drawing successively random samples from the full conditional distributions:

$$\begin{aligned} a_{c_j}^{(1)} &\sim p(a_{c_j} | \{a_{c_1}^{(1)}, \dots, a_{c_{j-1}}^{(1)}, a_{c_{j+1}}^{(0)}, \dots, a_{c_k}^{(0)}\}, \mathbf{a}_s^{(0)}, \boldsymbol{\omega}^{(0)}, \mathbf{D}) \\ a_{s_j}^{(1)} &\sim p(a_{s_j} | \mathbf{a}_{c_j}^{(1)}, \{a_{s_1}^{(1)}, \dots, a_{s_{j-1}}^{(1)}, a_{s_{j+1}}^{(0)}, \dots, a_{s_k}^{(0)}\}, \boldsymbol{\omega}^{(0)}, \mathbf{D}) \\ \omega_j^{(1)} &\sim p(\omega_j | \mathbf{a}_{c_j}^{(1)}, \mathbf{a}_{s_j}^{(1)}, \{\omega_1^{(1)}, \dots, \omega_{j-1}^{(1)}, \omega_{j+1}^{(0)}, \dots, \omega_k^{(0)}\}, \mathbf{D}), \quad (j = 1, \dots, k) \end{aligned} \quad (24)$$

At each iteration of the Gibbs sampler, we cycle through the set of conditional distributions and draw one sample from each. When a sample is drawn from one conditional distribution, the succeeding distributions are updated with the new value of that sample. This yields the following drawings at the K 'th iteration

$$\begin{aligned} a_{c_j}^{(K+1)} &\sim p(a_{c_j} | \{a_{c_1}^{(K+1)}, \dots, a_{c_{j-1}}^{(K+1)}, a_{c_{j+1}}^{(K)}, \dots, a_{c_k}^{(K)}\}, \mathbf{a}_s^{(K)}, \boldsymbol{\omega}_j^{(K)}, \mathbf{D}) \\ a_{s_j}^{(K+1)} &\sim p(a_{s_j} | \mathbf{a}_{c_j}^{(K+1)}, \{a_{s_1}^{(K+1)}, \dots, a_{s_{j-1}}^{(K+1)}, a_{s_{j+1}}^{(K)}, \dots, a_{s_k}^{(K)}\}, \boldsymbol{\omega}_j^{(K)}, \mathbf{D}) \\ \omega_j^{(K+1)} &\sim p(\omega_j | \mathbf{a}_{c_j}^{(K+1)}, \mathbf{a}_{s_j}^{(K+1)}, \{\omega_1^{(K+1)}, \dots, \omega_{j-1}^{(K+1)}, \omega_{j+1}^{(K)}, \dots, \omega_k^{(K)}\}, \mathbf{D}) \end{aligned} \quad (25)$$

We obtain $\{\mathbf{a}_c^{(K+1)}, \mathbf{a}_s^{(K+1)}, \boldsymbol{\omega}^{(K+1)}\}$. For a large enough K , $a_c^{(K+1)}$, $a_s^{(K+1)}$ and $\omega^{(K+1)}$ can be considered as random variables drawn from their posterior PDF distributions. Therefore we are able to generate samples of these posterior PDFs for each parameter. Using these samples, all of the estimates about the their corresponding can then be found, such as the most probable values for them, the mean value, the marginal variances with respect to the most probable value etc. When σ^2 is unknown, we do the same thing as above except that the random numbers are drawn from the Student's t distribution.

V. CRAMÉR-RAO LOWER BOUND

Estimators are often compared to the best possible performance given by CRLB [21-23] that provides a theoretical lower limit for their variances:

$$\text{Var}_{\hat{\boldsymbol{\theta}}}(\hat{\boldsymbol{\theta}}) \geq \text{CRLB}(\boldsymbol{\theta}) = \mathbf{J}^{-1}(\boldsymbol{\theta}), \quad (26)$$

where $\mathbf{J}(\boldsymbol{\theta})$ is called Fisher information matrix[21] and defined as an expectation of the second derivatives of the signal function with respect to $\boldsymbol{\theta}$ as follows:

$$\mathbf{J}(\boldsymbol{\theta}) = E \left[\frac{\partial^2 \ln P(\mathbf{D}|\boldsymbol{\theta}, I)}{\partial \boldsymbol{\theta}^2} \right] = \frac{1}{\sigma^2} \sum_{i=1}^N \frac{\partial f_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \left(\frac{\partial f_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right)^T. \quad (27)$$

For large N , $\mathbf{J}(\boldsymbol{\theta})$ is a diagonal matrix.

VI. COMPUTER SIMULATIONS

To demonstrate the proposed approach, let us consider a multiple harmonic frequency signal:

$$d_i = \cos(0.1 t_i + 1) + 2 \cos(0.15 t_i + 2) + 5 \cos(0.3 t_i + 3) + 2 \cos(0.31 t_i + 4) + 3 \cos(t_i + 5) + e_i \quad (28)$$

Here i runs in a symmetric time interval $-T$ and T ($2T+1=N$) and $e_i \sim N(0,1)$. We obtained noisy data samples ($N=512$), shown in Fig.1 and carried out Bayesian analysis.

All the methods, introduced in the paper were coded in *Mathematica* and run on a workstation in two cases where the standard deviation of noise is known or not. We need to give initial values for the parameters to start the iteration. Instead of choosing them randomly from a uniform distribution [9], we first performed a Fast Fourier Transformation (FFT) of the data and then chose the locations of the peaks in the power spectrum density, which is a squared magnitude of FFT, as an initial estimate for the frequencies. Once, initial frequencies were obtained, we carried on calculating the coefficients \mathbf{a}_c and \mathbf{a}_s , as initial values for the amplitudes, respectively.

In the case where the deviation of noise is unknown, the output of the computer simulation is illustrated in Table 1. The estimated parameter values are quoted as (*value*) \pm (*standard deviation*) and used to regenerate the given signal model, shown in Figure 1. It can be seen that five frequencies and their corresponding amplitudes are recovered very well.

The usual way the result from a spectral analysis is displayed is in the form of a power spectral density. In Fourier transform spectroscopy this is typically taken as the squared magnitude of the discrete Fourier Transform (DFT) of the data. Therefore, a comparison of Bayesian and Fourier spectral densities shown in Figure 1 indicate separation of frequencies. DFT spectral density shows only four peaks among five frequencies but, Bayesian spectral density indicates five frequencies with high accuracies. Moreover, we initially assumed that the values of the random noise in data were drawn from the Gaussian density. Figure 2 shows the exact and estimated PDF of the random

noise in data. 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 is also much closer to its true probability density. These results demonstrate how powerful Bayesian method is to separate noise from data.

Fig.3 shows only CPU time of different simulations for a variety of number of data samples and parameters and indicates that an increase in these numbers causes larger consumption of CPU time.

In order to evaluate the performance of GS, computer simulations have been performed and compared with the classical estimators such as ML and DFT, as well as CRLBs of ω and α , expressed in decibel (dB), which are a function of N and SNR:

$$\begin{aligned} CRLB(\omega) &\approx SNR + 10 \text{Log} \left(N^3 / 12 \right) \\ CRLB(\alpha) &\approx 10 \text{Log} (N) + 10 \text{Log} \left(1 / (2\sigma^2) \right) \end{aligned} \quad (29)$$

where $\alpha = \sqrt{a_{c1}^2 + a_{s1}^2}$. We fixed α to $\sqrt{2}$ and properly scaled $e(t_i)$ to obtain different SNRs, defined as $SNR = \frac{\alpha^2}{2\sigma^2}$. Unless stated otherwise, the angular

frequency is chosen as $\omega = 0.3\pi$ and $SNR=20\text{dB}$. We generated $N=100$ data samples from a single real tone frequency signal model in a variety of noise levels. After 50 independent trials under the same noise level, MSEs of the estimated frequency and amplitude were obtained for each method. Their logarithmic values were plotted with respect to SNR ratios, which vary from -10 dB to 30 dB and shown in Fig.4 and 6. They indicate the MSE performances for different estimators. The error curves in these figures were separated into two regions. The first one, on the left, shows that the estimator variances increase stronger than the CRLB and contain smaller threshold effects in Fig.6 than that of Fig.4. The second one, on the out of left indicates that the errors follow the CRLB and the curves close to it. In Fig.4, it can be seen that GS, ML and DFT estimators have threshold about -5 dB, -4 dB and -2 dB of the SNR, respectively and follow nicely with the CRLB after -1 dB. As expected, with increasing SNR, MSE values approaches to the CRLB but, with decreasing SNR, they get worse from it. This implies the higher the SNR, the lower CRLB. Moreover, all three estimators have same characters at high SNRs.

The above argument treats only with the case in which a size of data samples $N=100$ is used for the estimation. Therefore, one may ask how to vary accuracy of the estimation with N . To answer it, we set up an experiment in which the algorithms of three methods were run for 50 simulated data with different lengths. In this case Figure 5 and 7 show the MSE performances of three estimators with different data length which varies from $N=25$ to $N=300$

at $\omega = 0.3\pi$ under SNR=20 dB. They indicate that the larger data samples give the lower MSEs of frequencies than that of amplitudes. This implies that all three estimators are more effective for the frequency estimation, rather than amplitude estimation. On the other hand, an efficiency parameter [24], defined as

$$\eta_i = \frac{CRLB_i}{MSE_i} \times 100, \quad (30)$$

indicates the closeness of estimators to the CRLB. Table 2 and 3 contain the MSEs and the efficiency values η for the frequency and amplitude estimation obtained at the last states where SNR = 30dB and $N = 300$. It can be seen that the efficiency value for the GS at SNR = 30dB and for $N = 300$ is much closer to the CRLB than that of the other methods in both frequency and amplitude estimation. Thus, it is said that GS is more effective than the others for higher SNR and larger data sample.

CONCLUSIONS

In this work we have presented a numerical procedure, namely GS, based on Bayesian inference for estimating parameters of multiple sinusoids embedded in noise. Overall results show that Bayesian approach can not only give us the best estimates for the parameters but, it can also tell us uncertainties associated with their estimated values. Experiments with synthetic signals show that GS performs frequency estimations with a high-resolution, according to the CRLB. On the other hand, it requires a maximization of full conditional marginal probability density function of frequencies that can be difficult if SNR is low. Comparing with classical estimators such as ML and DFT, all three methods can give similar performance in higher SNRs and larger N . However ML and DFT require larger CPU times than GS estimator.

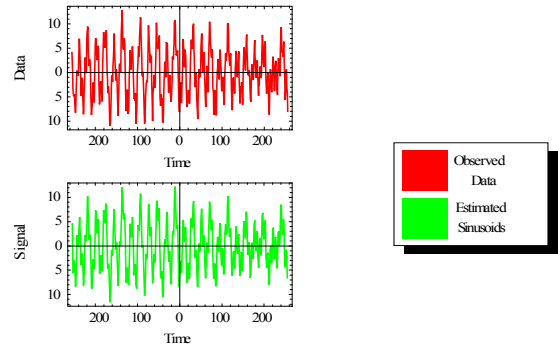
ACKNOWLEDGEMENT

This work has been supported by the Research Fund of Istanbul University with project numbers are UDP-33672 and YADOP-19681.

REFERENCES

- [1] Z.H.Michalopoulou and M. Picarelli, "Gibbs sampling for time-delay-and amplitude estimation in underwater acoustics", *J. Acoust. Soc. Am.*, vol. 117, pp. 799–808, 2005.
- [2] R. H. Swendsen and J.S. Wang, "Physical review of letters", vol. 57, pp.2607-2609, 1986.
- [3] R.J. Kenefic and A.H. Nuttall, "Maximum likelihood estimation of the parameters of tone using real discrete data", *IEEE J. Oceanic Eng.*, vol. 12 (1), pp. 279–280, 1987.
- [4] B.G. Quinn, "Estimating frequency by interpolation using Fourier coefficients", *IEEE Trans. Signal Process.* vol.42 (5), pp. 1264–1268, May 1994.
- [5] J.W.Cooley and J.W.Tukey, "An algorithm for the machine calculation of complex Fourier series", *Mathematics of computation*, vol.19, pp. 297-301, 1965.
- [6] N. Metropolis, A. Rosenbluth, M. Rosenbluth, A. Teller and E. Teller, "Equation of states calculations by fast computing machines", *Journal of chemical physics*, vol.21, pp. 1087-1092, 1953.
- [7] E.T Jaynes, "Bayesian Spectrum and Chirp Analysis", *In Proceedings of the Third Workshop on Maximum Entropy and Bayesian Methods*, Ed. C. Ray Smith and D. Reidel, Boston, pp. 1-37, 1987.
- [8] G.L. Bretthorst, "Bayesian Spectrum Analysis and Parameter Estimation", *Lecture Notes in Statistics*, Springer-Verlag Berlin Heidelberg New York, 1997.
- [9] P.Gregory, "Bayesian Logical Data Analysis for the Physical Science", Cambridge University Press, United Kingdom, 2005.
- [10] M. Cevri and D. Ustundag, "Bayesian recovery of sinusoids from noisy data with parallel tempering", *IET Signal Process.*, vol 6 (7), pp. 673–683, 2012.
- [11] L. Dou and R J W. Hodgson, "Bayesian inference and Gibbs sampling in spectral analysis and parameter estimation I", *Inverse Problem*, vol. 11, pp.1069-1085, 1995.
- [12] L. Dou, R J W. Hodgson, "Bayesian inference and Gibbs sampling in spectral analysis and parameter estimation II", *Inverse Problem*, vol. 11, pp.121-137, 1995.
- [13] C. Andreiu and A.Doucet, "Joint Bayesian model selection and estimation of noisy sinusoids via reversible jump MCMC", *IEEE Transactions on Signal Processing*, vol.47, pp.2667-2676, 1999.
- [14] D. Ustundag, and M. Cevri, *Simulated Annealing - Advances, Applications and Hybridizations*, In Tech, Croatia, ISBN: 978-953-51-0710-1, pp.67-90, 2012.
- [15] D. Üstündağ and M. Cevri, "Estimating parameters of sinusoids from noisy data using Bayesian inference with simulated annealing", *Wseas Transactions On Signal Processing*, vol.7, pp.432-441, 2008.
- [16] D. Üstündağ, M.Cevri, "Recovering Sinusoids from Noisy Data Using Bayesian Inference with Simulated Annealing", *Mathematical & Computational Applications*, Vol. 16(2), pp. 382-391, 2011.
- [17] D.C.Rife and R.R.Boorstyn, "Single-Tone Parameter Estimation from Discrete-Time Observations", *IEEE Transactions on Information Theory*, vol.20, pp.591-598, 1974.
- [18] J.M. Bernardo and A.F.M.Smith, "Bayesian Theory, Willey Series in Probability and Statistics", New York, 2000.
- [19] G.E.P. Box and C. Tiao, "Bayesian Inference in Statistical Analysis", New York, 1992.
- [20] P.Diaconis, K. Khare and L. S.Coste, "Gibbs Sampling, Exponential Families and Orthogonal Polynomials", *Statistical Science*, Vol. 23(2), pp. 151–178, 2008.
- [21] S.M. Kay, "Fundamentals of Statistical Signal Processing: Estimation Theory", Prentice-Hall, Englewood Cliffs, NJ, pp. 56–57, 1993.
- [22] S.M. Kay, "Accurate frequency estimation at low signal-to-noise ratio", *IEEE Trans. Acoust., Speech, Signal Processing*, ASSP-32, pp.540–547, 1984.

- [23] P. Händel, "Parameter estimation employing a dual-channel sine-wave model under a Gaussian assumption", *IEEE Transactions on Instrumentation and Measurement*, vol. 57(8), pp. 1661-1669, 2008.
- [24] B. Ristic, S. Arulampalam, N.Gordon, "Beyond the Kalman Filter Particle Filters for Tracking Applications", Artech House, London, 2004.
- [25] H.L.Harney, "Bayesian Inference: Parameter Estimation and Decisions", Springer-Verlag, Berlin Heidelberg, 2003.



TABLES AND FIGURES

Table 1. Computer simulations for five harmonic frequency signal model

Parameters		True values	Estimated values
Frequencies	ω_1	0.1	0.09979±0.0003
	ω_2	0.15	0.1498±0.0002
	ω_3	0.3	0.2999±0.00008
	ω_4	0.31	0.3095±0.0002
	ω_5	1.	1.000±0.0001
Amplitudes	a_{c_1}	0.540302	0.6542±0.0618
	a_{c_2}	-0.832294	-0.8582±0.0620
	a_{c_3}	-4.94996	-4.756±0.0624
	a_{c_4}	-1.30729	-1.4±0.06225
	a_{c_5}	0.850087	0.7496±0.06296
	a_{s_1}	-0.841471	-0.9683±0.06272
	a_{s_2}	-1.81859	-1.82±0.06311
	a_{s_3}	-0.7056	-0.8228±0.06247
	a_{s_4}	1.5136	1.443±0.06273
	a_{s_5}	2.87677	2.8227±0.06307

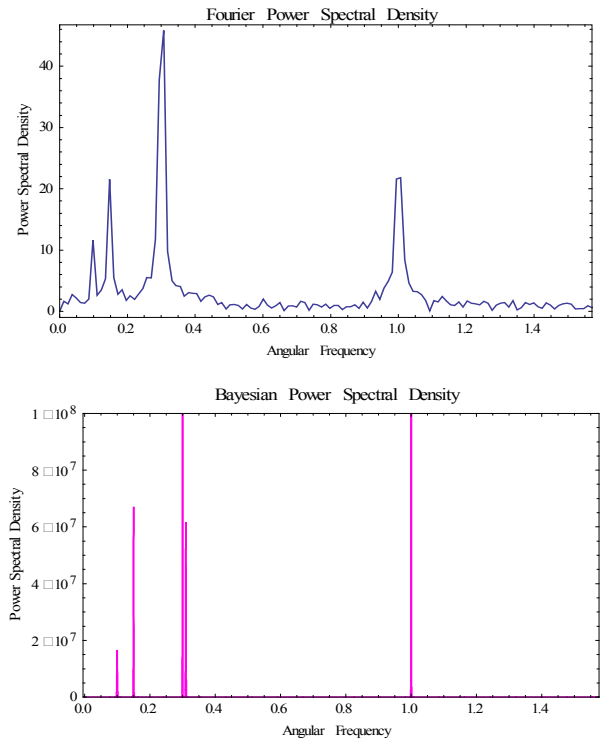


Fig. 1. Spectral analysis of multiple frequency model

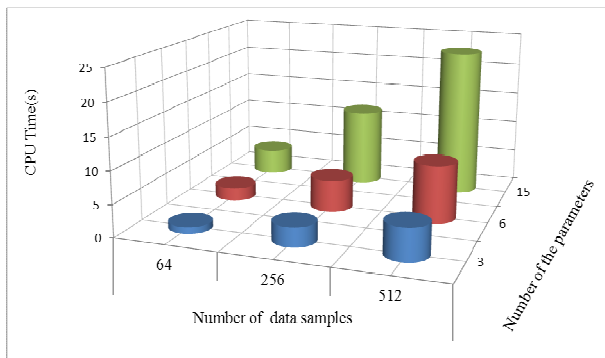


Fig.3 Different simulation times with respect to number of parameters and data samples

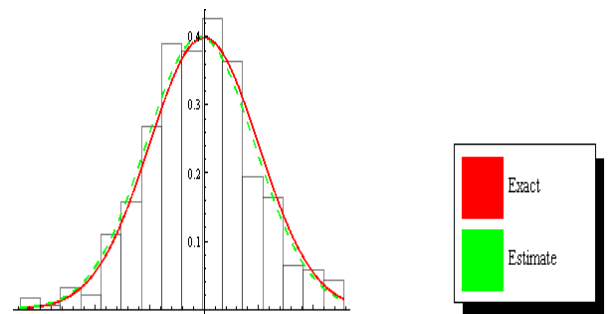


Fig.2 Comparison of exact and estimate probability densities of noise in data

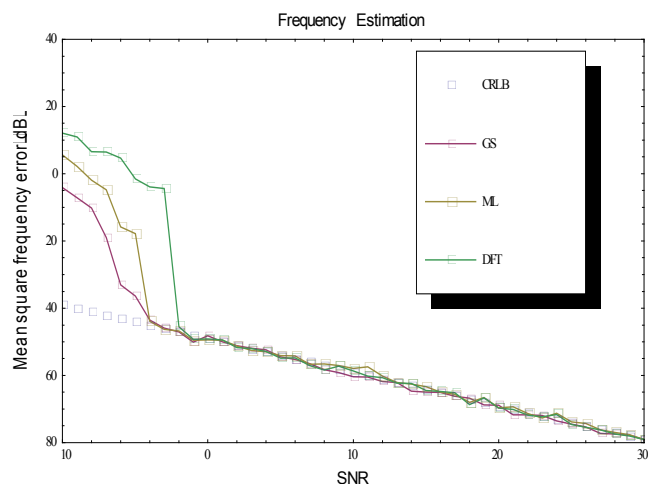


Fig. 4. Mean square frequency error versus SNR at $\omega = 0.3\pi$

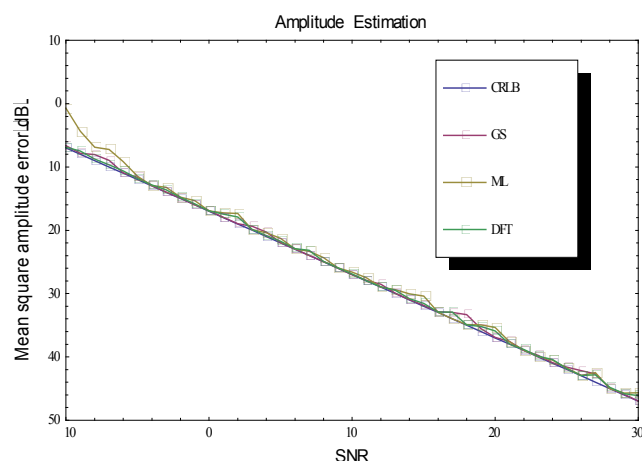


Fig.6. Mean square amplitude error versus SNR at $a = \sqrt{2}$

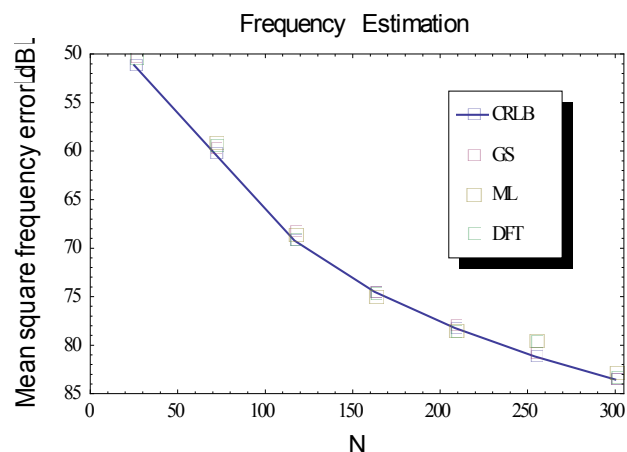


Fig. 5 Mean square frequency error versus N at $\omega = 0.3\pi$

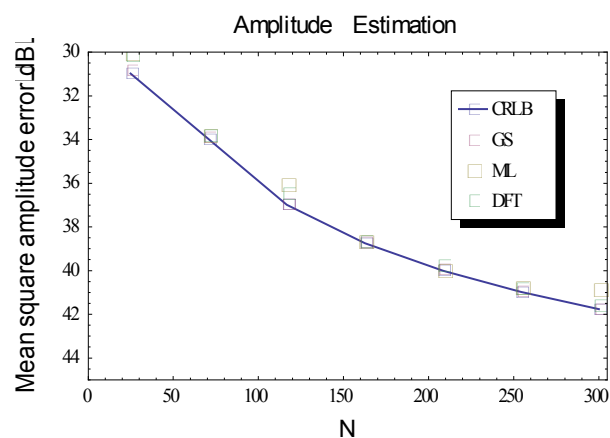


Fig.7. Mean square amplitude error versus N at $a = \sqrt{2}$

Table 2. Performance comparison of Bayesian methods for frequency estimation in single frequency signal model

Methods	SNR = 30dB		N = 300	
	MSE(dB)	Efficiency	MSE(dB)	Efficiency
GS	-79.175	100.041	-83.484	100.045
ML	-79.079	100.162	-82.850	100.81
DFT	-79.103	100.133	-83.401	100.144
CRLB	-79.208	100	-83.521	100

Table 3. Performance comparison of Bayesian methods for amplitude estimation in single frequency signal model

Methods	SNR = 30dB		N = 300	
	MSE(dB)	Efficiency	MSE(dB)	Efficiency
GS	-46.9800	100.021	-41.7415	100.046
ML	-45.6409	102.955	-40.8652	102.192
DFT	-46.0871	101.958	-40.8450	100.381
CRLB	-46.9897	100	-41.7609	100

**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