

RECOVERING SINUSOIDS FROM NOISY DATA USING BAYESIAN INFERENCE WITH SIMULATED ANNEALING

Dursun Üstündağ and Mehmet Cevri

The Department of Mathematics, Faculty of Science and Arts
Marmara University, 34722, Kadıköy, Istanbul, Turkey
dustundag@marmara.edu.tr, mcevri@gmail.com

Abstract- In this paper, we studied Bayesian analysis proposed by Bretthorst[6] for a general signal model equation and combined it with a simulated annealing (SA) algorithm to obtain a global maximum of a posterior probability density function (PDF) for frequencies. Thus, this analysis offers different approach to finding parameter values through a directed, but random, search of the parameter space. For this purpose, we developed a Mathematica code of this Bayesian approach together with SA and used it for recovering sinusoids from noisy data. Simulations results support its effectiveness.

Key Words- Bayesian Statistical Inference Simulated Annealing, Parameter Estimations, Power Spectral Density, Cramér-Rao lower bound.

1. INTRODUCTION

Let the vector $\boldsymbol{\theta}$ contain the parameters to be estimated from the (measurements) vector \mathbf{D} , which is the output of the physical system that one wants to be modeled. In many experiments, the recorded data $\mathbf{D} = \{d_1, \dots, d_N\}$ are sampled from an unknown model function $f(t; \boldsymbol{\theta})$ at discrete times $(t_1, \dots, t_N)^T$. Often one assumes a signal model equation as follows:

$$d_i = f(t_i; \boldsymbol{\theta}) + e(t_i), \quad (1)$$

where $e(t_i)$ represents the noise, assumed to be drawn independently from a zero mean Gaussian probability distribution with a standard deviation of σ . Different models correspond to different choices of signal model function $f(t; \boldsymbol{\theta})$. According to Bretthorst [6], the most general signal model may be given in the following form:

$$f(t; \boldsymbol{\theta}) = \sum_{j=1}^m B_j G_j(t; \{\omega\}), \quad (2)$$

where $G_j(t; \{\omega\})$ is a function of $\{\omega_1, \dots, \omega_\rho\}$ that we label collectively $\{\omega\}$ and B_j represents the amplitude corresponding to the j th model function $G_j(t; \{\omega\})$. The goals of data analysis are usually to use the observed data \mathbf{D} to infer the values of parameters $\boldsymbol{\theta} = \{\omega_1, \dots, \omega_\rho; B_1, \dots, B_{2\rho}\} (m = 2\rho)$.

In this paper, we address the problem of estimation of sinusoids in white Gaussian noise within a Bayesian framework. This is of great interest in many fields of science, including seismology, nuclear magnetic resonance and radar. Under an assumption of a known number of sinusoids, several algorithms have already been used in parameter estimation literature, such as least-square fitting [9], discrete Fourier transform [1], and periodogram [2]. The discrete Fourier transform has been a very powerful tool in Bayesian spectral analysis since Cooley and Tukey introduced the fast

Fourier transform (FFT) technique in 1965, followed by the rapid development of computers. After Jaynes [3] derived a periodogram directly from the principles of Bayesian inference, researchers in different branches of science have given much attention to the relationship between Bayesian inference and parameter estimation and they have done excellent works in this area for last fifteen years [4-7].

We consider here a Bayesian approach, proposed by Bretthorst, in which the posterior PDF for $\{\omega\}$ has to be maximized. Unfortunately, conventional algorithms [9] based on a gradient direction fail to converge for this maximization problem. Even when they converge, there is no assurance that they have found a global, rather than a local maximum. This is because a log of the posterior PDF is so sharply peaked and highly nonlinear function of $\{\omega\}$. Bretthorst pointed it out and used a pattern search algorithm described by Hook-Jeeves [10] to overcome this problem but, this approach does not converge unless starting point is much closer to the optimum $\{\omega\}$. Our contribution is therefore to develop a computer program using *Mathematica* [8] that combines this Bayesian approach with a very different optimization algorithm called SA [11]. It explores the entire surface of the posterior PDF for the frequencies and tries to optimize it while moving both uphill and downhill in order to escape from local maxima. Furthermore, it is largely independent of the starting values, often a critical input in conventional algorithms. Finally, we use it for estimating parameters of the sinusoids corrupted by a random noise.

2. BAYESIAN PARAMETER ESTIMATION

Let us now reconsider the problem given in Equation (1) within the Bayesian framework. If $\{\{\omega\}, \{B\}, \sigma^2\}$ is a set of parameters of interest, then their joint PDF given the observed data \mathbf{D} is proportional to

$$P(\{\omega\}, \{B\}, \sigma^2 | \mathbf{D}, \mathbf{I}) \propto P(\{\omega\}, \{B\}, \sigma^2 | \mathbf{I}) P(\mathbf{D} | \{\omega\}, \{B\}, \sigma^2, \mathbf{I}). \quad (3)$$

The quantity $P(\{\omega\}, \{B\}, \sigma^2 | \mathbf{I})$ is called the prior PDF; it represents prior knowledge of the parameters $\{\{\omega\}, \{B\}, \sigma^2\}$ given the information \mathbf{I} . The sampling PDF $P(\mathbf{D} | \{\omega\}, \{B\}, \sigma^2, \mathbf{I})$ is the likelihood of \mathbf{D} . A key component in Bayes theorem is the likelihood function $P(\mathbf{D} | \{\omega\}, \{B\}, \sigma^2, \mathbf{I})$ that is proportional to the PDF for the noise. If σ is known, then the likelihood function takes on the form:

$$P(\mathbf{D} | \{\omega\}, \{B\}, \sigma^2, \mathbf{I}) = (2\pi\sigma^2)^{-\frac{N}{2}} \exp\left(-\frac{NQ}{2\sigma^2}\right). \quad (4)$$

The exponent Q is defined as follows

$$Q = \overline{\mathbf{D}^2} - \frac{2}{N} \sum_{j=1}^m \sum_{i=1}^N d_i B_j G_j(t_i; \{\omega\}) + \frac{1}{N} \sum_{j=1}^m \sum_{k=1}^m \Omega_{jk} B_j B_k, \quad (5)$$

where $\overline{\mathbf{D}^2} = \frac{1}{N} \sum_{i=1}^N d_i^2$ and $\Omega_{jl} = \sum_{i=1}^N G_j(t_i; \{\omega\}) \otimes G_l(t_i; \{\omega\})$ ($j, l = 1, \dots, m$). In order to obtain the PDF for $\{\omega\}$, we integrate out the nuisance parameters $\{B\}$ and σ dependence,

$$P(\{\omega\} | \mathbf{D}, I) = \int P(\{\omega\}, \{B\}, \sigma | \mathbf{D}, I) d\{B\} d\sigma. \quad (6)$$

To do it analytically, it is necessary to make the matrix Ω_{jk} to be diagonal, effectively by introducing new orthogonal model functions,

$$H_j(t; \{\omega\}) = \frac{1}{\sqrt{\lambda_j}} \sum_{k=1}^m a_{jk} G_k(t; \{\omega\}), \quad (j = 1, \dots, m). \quad (7)$$

This diagonalization process gives a new expression for the signal model function,

$$f(t; \boldsymbol{\theta}) = \sum_{k=1}^m A_k H_k(t; \{\omega\}), \quad (8)$$

where the new amplitudes A_k 's are related to the old amplitudes B_j 's by

$$A_k = \sqrt{\lambda_k} \sum_{j=1}^m B_j u_{kj}, \quad (k = 1, \dots, m) \quad (9)$$

and where u_{kj} represents the j th component of the k th normalized eigenvector of Ω_{jk} , with λ_j as the corresponding eigenvalue. Substituting these expressions into Equation (5) and define

$$h_j = \sum_{i=1}^N d_i H_j(t_i, \{\omega\}), \quad (j = 1, \dots, m), \quad (10)$$

to be the projection of the data onto the orthogonal model functions $H_j(t, \{\omega\})$, we can then proceed to perform m Gaussian integrations over A_j . If σ is known, the joint PDF of the $\{\omega\}$ parameters, conditional on the data is given by

$$P(\{\omega\} | \mathbf{D}, I) \propto \exp\left(-\frac{m\overline{\mathbf{h}^2}}{2\sigma^2}\right), \quad (11)$$

where $\overline{\mathbf{h}^2} = \frac{1}{m} \sum_{j=1}^m h_j^2$. If σ is not known, by using the Jeffreys prior [15] and integrating

Equation (6) over σ parameter we obtain

$$P(\{\omega\} | \mathbf{D}, I) \propto \left(1 - \frac{m\overline{\mathbf{h}^2}}{N\overline{\mathbf{D}^2}}\right)^{\frac{m-N}{2}}. \quad (12)$$

This has the form of the Student t -distribution. It is desirable to compute the variances associated with those parameters. To do this, we can expand the function $\overline{\mathbf{h}^2}$ in a Taylor series at the point $\{\hat{\omega}\}$ which maximizes the PDF in Equation (11) or (12), such that

$$P(\{\omega\} | \mathbf{D}, I) \propto \exp\left(-\sum_{j=1}^{\rho} \sum_{k=1}^{\rho} \frac{b_{jk}}{2\langle\sigma^2\rangle} (\hat{\omega}_j - \omega_j)(\hat{\omega}_k - \omega_k)\right) \quad (13)$$

with $b_{jk} = -\frac{m}{2} \frac{\partial^2 \mathbf{h}^2}{\partial \omega_j \partial \omega_k} \Big|_{\substack{\omega_j = \hat{\omega}_j \\ \omega_k = \hat{\omega}_k}}$. For an arbitrary signal model the matrix b_{jk} cannot be calculated analytically; however, it can be evaluated numerically. This can be done by first changing to the orthogonal variables and performing the ρ Gaussian integrals. Let v_j and u_{kj} represent the j th eigenvalue and eigenvectors of the matrix b_{jk} , respectively. Then the new orthogonal variables are given by

$$s_j = \sqrt{v_j} \sum_{k=1}^{\rho} r_k u_{kj}, \quad r_j = \sum_{k=1}^{\rho} \frac{s_k u_{jk}}{\sqrt{v_k}}. \quad (14)$$

From these, we get the estimated frequency w_j in the form:

$$\langle \omega_j \rangle = \hat{\omega}_j \pm \sqrt{\langle \sigma^2 \rangle \sum_{l=1}^{\rho} \frac{u_{lj}^2}{v_l}}. \quad (15)$$

One can also show that $\langle A_j \rangle = h_j$. By using Equation (9) we get the estimated amplitude B_k in the form:

$$\langle B_k \rangle = \sum_{j=1}^m \frac{1}{\sqrt{\lambda_j}} h_j a_{jk} \pm \sqrt{\sigma^2 \sum_{j=1}^m \frac{1}{\lambda_j} a_{jk} a_{jk}} \quad (16)$$

3. CONTINUOUS GLOBAL OPTIMIZATION ALGORITHM

Bayesian parameter estimation turns into the global optimization problem [12] which is a task to find the best possible solution for the problem in Equations (11) and (12):

$$\max_{\{\omega\} \in \Xi^{\rho}} \{P(\{\omega\} | \mathbf{D}, I)\}, \quad (17)$$

where $\Xi = \{0, 2\pi\}$. Although there is numerous other approaches [13, 16] to solve this maximization problem, the SA algorithm, suggested by Corana et. al. [11] and modified by Goffe et. al. [12] is chosen here because it is generally applicable and easy to implement. This begins with initial guesses of ω^0 and Γ_0 (called temperature). Each step of this algorithm replaces the current frequency by a random nearby frequency. In other words, the next candidate point ω^k is generated by varying one component $j \in \{1, \dots, \rho\}$ of the current iteration $\omega^{(k-1)}$ at a time:

$$\omega_j^k = \omega_j^{k-1} + \delta\omega, \quad (18)$$

where $\delta\omega = \sqrt{\Upsilon(\omega)}N(0, 1)$ is recalculated until $\omega_j^{\min} < \omega_j < \omega_j^{\max}$ for upper and lower limit ω_j^{\max} and ω_j^{\min} , respectively (ordering $0 < \omega_1 < \dots < \omega_{\rho} < 2\pi$). The step size $\delta\omega$ is calculated from the square root of the Cramer-Rao lower bound (CRLB) $\Upsilon(\omega)$ [13] that

is a lower limit to the variance of the measurement of ω_j , so this generates a natural scale size of the search space around its estimated value. It is expected that better solutions lie close to solutions that are already good and so normally distributed step size are used. Thus, the point ω^k is accepted as the k th iteration point if $P(\omega^k | \mathbf{D}, I) \geq P(\omega^{(k-1)} | \mathbf{D}, I)$. Otherwise, it is accepted or rejected according to the Metropolis-criterion [14]:

$$\exp\left(\frac{P(\omega^{(k-1)} | \mathbf{D}, I) - P(\omega^k | \mathbf{D}, I)}{\Gamma_{k-1}}\right) > p, \quad (19)$$

where p is a uniformly distributed random number from $[0,1]$. This continues until all ρ components have been altered and thus ρ new points have successively accepted or rejected according to the Metropolis criterion. After this process is repeated n_s times, the whole cycle is then repeated n_Γ times, after which the temperature is decreased by a factor $\Gamma_k = \alpha_\Gamma \Gamma_0$ (called annealing or cooling schedule). Termination of the algorithm occurs when average function value of sequences of points after each $n_s n_\Gamma$ step cycle reaches a stable state:

$$\left|P_k(\omega | \mathbf{D}, I) - P_{k-l}(\omega | \mathbf{D}, I)\right| \leq \varepsilon \quad (l = 1, \dots, 4) \text{ and } \left|P_k(\omega | \mathbf{D}, I) - P_{opt}(\omega | \mathbf{D}, I)\right| \leq \varepsilon, \quad (20)$$

where ε is a small positive number defined by user, $P_{opt}(\omega | \mathbf{D}, I)$ is the value of the PDF at the best point ω_k and l indicates the last four successive iteration values of the PDF that are being stored.

4. COMPUTER SIMULATED EXAMPLES

To verify the performance of the proposed algorithm, we generated a data vector from multiple frequency sinusoids. Here t_i runs over the symmetric time interval $-T$ to T in $(2T+1) = 512$ integer steps and $e_i \sim N(0, \sigma)$. We first consider a signal model with two closed harmonic frequencies:

$$d_i = 0.5403 \cos(0.3t_i) - 0.8415 \sin(0.3t_i) - 0.4161 \cos(0.31t_i) - 0.9093 \sin(0.31t_i) + e_i, \quad (i = 1, \dots, 512) \quad (21)$$

Table 1. Computer simulations for two closed harmonic frequency signal model

		Bayesian Parameter estimation	
		Parameters	Estimated Values
σ : Unknown N : 512 m : 6 SNR: 1.02936 $\hat{\sigma}$: 0.99639	w_1		0.300082± 0.0009384
	w_2		0.310574± 0.0010515
	B_1		0.520649± 0.0632635
	B_2		-0.869805± 0.0631361
	B_3		-0.381937± 0.0634942
	B_4		-0.855687± 0.0636233

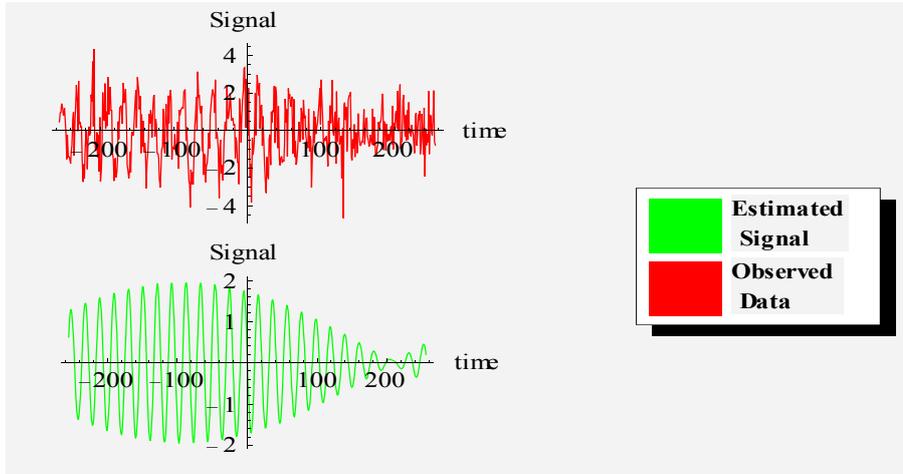


Figure 1. Recovering signals from noisy data produced from two closed harmonic Frequency signal model.

We carried out the Bayesian analysis of the simulated data, assuming that we know the mathematical form of the signal model but, not the value of the parameters. As an initial estimate of the frequencies ω^0 for the maximization procedure, it is possible to take random choices from the interval $(0, 2\pi)$. However, it is better to start with the locations of the peaks chosen automatically from the Fourier power spectral density graph by using a computer code written in Mathematica. After reasonable values of parameters that control the simulated annealing routine are chosen as $n_s = 20$, $n_T = \max(100, 5\rho)$ and $\varepsilon = 0.01$, the algorithm starts at some high temperature $\Gamma_0 = 100$ and generates the sequence of points until a sort of equilibrium is approached; that is a sequence of points $\omega^0, \omega^1, \omega^2, \dots$ whose average value of $P(\omega | \mathbf{D}, I)$ reaches a stable value as iteration proceeds. During this phase the step size is naturally adjusted. The best point ω reached so far is recorded. After thermal equilibration, Γ_0 is reduced by a factor $\alpha_\Gamma = 0.85$ and a new sequence is made starting from this best point ω , until thermal equilibrium is reached again and so on. Therefore it proceeds toward better maxima even in the presence of many local maxima. Consequently, the process is stop at a temperature low enough that no more useful improvement can be expected, according to a stopping criterion in Equation (20). Once the frequencies are estimated, we carried on calculating the estimated amplitudes associated with the errors. However, an evaluation of the posterior PDF at a given point ω cannot be made analytically and also requires a numerical calculation of projections on to orthonormal model functions, related to eigen-decomposition of $\Omega(\omega)$. It is therefore expected that its evaluation at ω requires larger consumption of time as the length of ω increases.

The computer program we developed was run on the workstation in two cases where the standard deviation of the noise is known or not. The computer simulations illustrated in Table 1 show the results when σ is known but they are almost similar with those obtained in the case σ is unknown. Estimated parameter values, quoted as

(value) \pm (standard deviation), indicate that all values of the parameters within the calculated accuracy are clearly recovered. The estimated value of signal to noise ratio (SNR) and the standard deviation of the noise are also shown in Table 1. On the other hand, Figure 1 shows the power of the method for recovering the signal from the noisy data. In general, we consider a multiple harmonic frequency signal model which is also used by Bretthorst [6]:

$$d_i = \cos(0.1 t_i + 1) + 2 \cos(0.15 t_i + 2) + 5 \cos(0.3 t_i + 3) + 2 \cos(0.31 t_i + 4) + 3 \cos(t_i + 5) + e_i \quad (i = 1, \dots, 512) \quad (22)$$

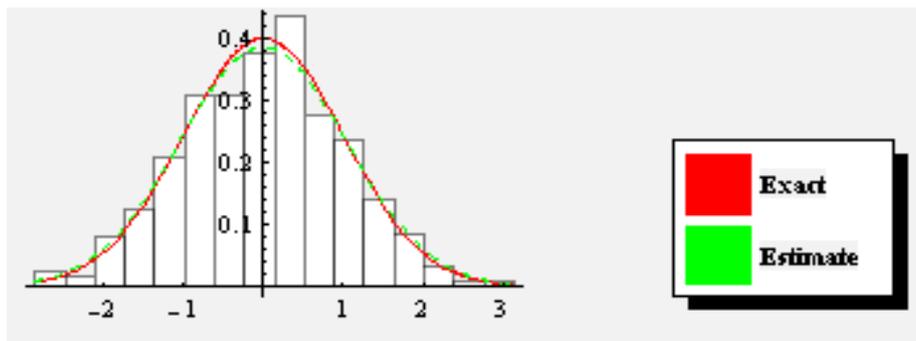
The best estimates of parameters are tabulated in Table 2. Once again, all the frequencies have been well resolved, even the third and fourth frequencies, which are very close not to be separated by the Fourier power spectral density shown in Figure 3. Actually with the Fourier spectral density when the separation of two frequencies is less than the Nyquist step, defined as $2\pi / N$, two frequencies are indistinguishable. In this example these two frequencies are separated by 0.01, which is less than the Nyquist step size. Therefore, there is no way by using the Fourier power spectral density that one can resolve these closed frequencies less than the Nyquist step, however Bayesian power spectral density shown in Figure 3 gives us very good results with high accuracy. The results we obtained so far are also consistent with that of Bretthorst [6].

Moreover, we initially assumed that the noise in data were drawn from the Gaussian density with the mean $\mu = 0$ and the standard deviation σ . Figure 2 shows the exact and estimate 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 data is also much closer to the true PDF. The histogram is known as a nonparametric estimator of the PDF because it does not depend on specified parameters.

We generated 64 data samples from a single frequency signal model and added it to the variety of noise levels. After 50 independent trials, the mean square errors (MSE) were calculated and their logarithmic values were plotted with respect to SNR that varies between zero and 20 dB. It can be seen from Figure 4 that the proposed estimator has threshold about 3 dB of SNR and follows nicely the CRLB after this value. As expected, larger SNR ratio gives smaller MSE.

Table 2. Computer simulations for a multiple harmonic frequency model¹

Parameters		True values	Estimated values	Bretthorst's results
Frequencies	ω_1	0.10	0.1006 ± 0.0004	0.0998 ± 0.0001
	ω_2	0.15	0.1495 ± 0.0002	0.1498 ± 0.0002
	ω_3	0.30	0.3001 ± 0.0001	0.3001 ± 0.0002
	ω_4	0.31	0.3099 ± 0.0004	0.3120 ± 0.0001
	ω_5	1.00	1.0000 ± 0.0001	0.9999 ± 0.0001
Amplitudes	a_1	1.0	0.99 ± 0.08	0.99 ∓ 0.08
	a_2	2.0	1.96 ± 0.08	2.08 ± 0.08
	a_3	5.0	5.10 ± 0.09	4.97 ± 0.08
	a_4	2.0	1.81 ± 0.09	1.95 ± 0.08
	a_5	3.0	2.95 ± 0.08	2.92 ± 0.08

**Figure 2.** Comparison of exact and estimate PDF of noise in data ($\sigma = 1$).

5. CONCLUSION

In this work we have partially developed a Bayesian approach with a simulated annealing and applied it to spectral analysis and parameter estimation problems. Overall results show that it provides rational approach for estimating, in an optimal way, values of parameters of sinusoids corrupted by random noise. Both frequency and amplitudes can be recovered from the experimental data and the prior information with high accuracy, especially the frequency which is the most important parameter in spectral analysis. Although it requires a large consumption of CPU time, it is competitive when compared to the multiple runs often used with conventional algorithms to test different starting values. For a sufficiently high SNR the MSE will attain Cramer-Rao lower bound so that it justifies the accuracy of the frequency estimation. Moreover,

¹ In order to compare the results with Bretthorst's in this example we converted $a_i = \sqrt{B_i^2 + B_{i+\rho}^2}$, ($i = 1, \dots, \rho$).

Mathematica is a powerful system for doing mathematics on the computer and it has grown to become today an unparalleled platform for all forms of computations.

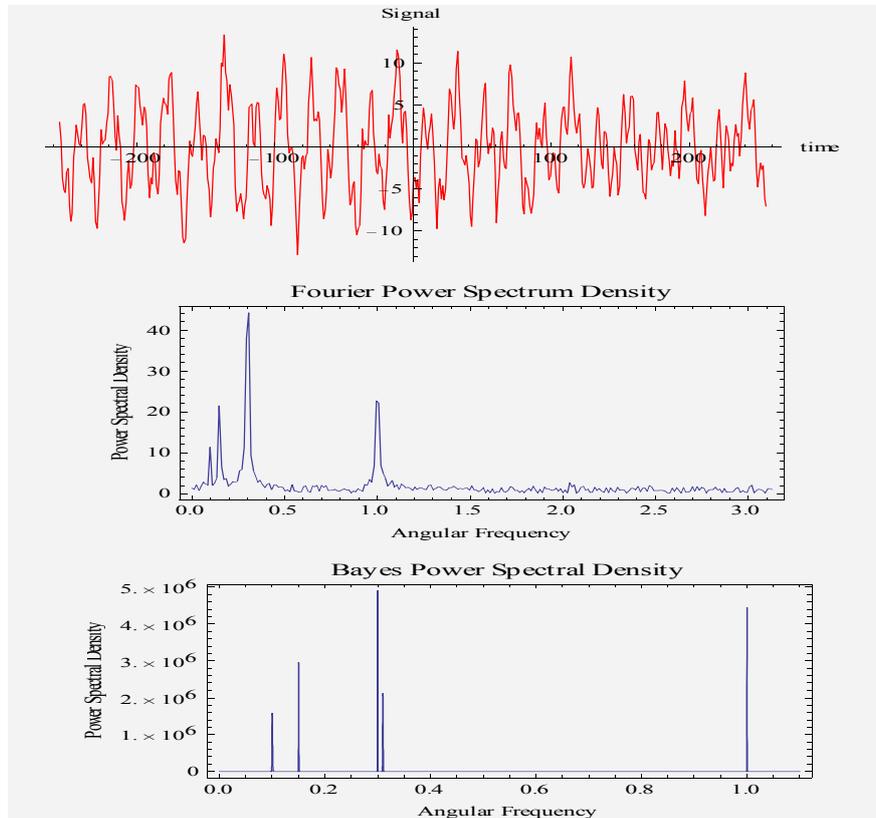


Figure 3. Spectral analysis of multiple frequency signal models.

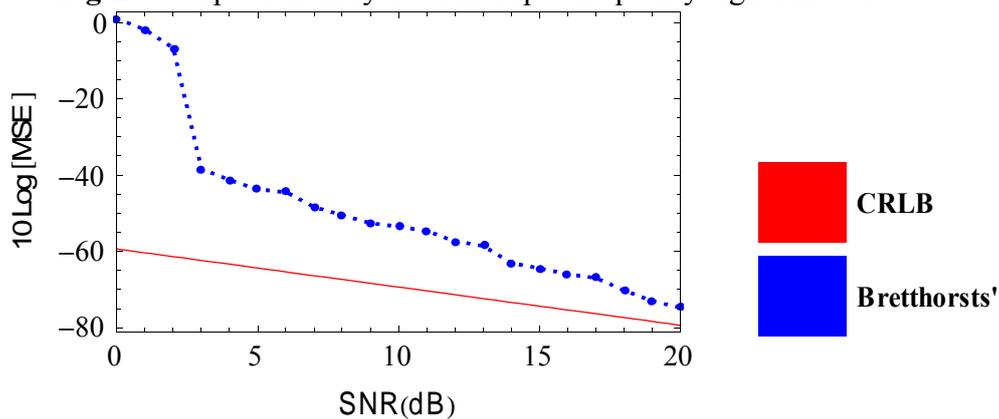


Figure 4. Mean square frequency errors versus SNR ($\omega = 0.3, N = 64$).

Acknowledgments- This work is a part of the projects, whose names are "Bayesian Model Selection and Parameter Estimation" with a number FEN - DKR - 200407 - 0082

and "Bayesian Spectrum Estimation of Harmonic Signals with a number FEN-BGS-060907-0191, supported by the University of Marmara, Istanbul, Turkey.

7. REFERENCES

1. J.W. Cooley and J.W. Tukey, An algorithm for the machine calculation of complex Fourier series, *Mathematics of Computation* **19**, 297-301, 1965.
2. A. Schuster, The periodogram and its optical analogy, *Proceedings of the Royal Society of London* **77**, 136-140, 1905.
3. E.T. Jaynes, Bayesian spectrum and chirp analysis, *Maximum Entropy and Bayesian Spectral Analysis and Estimation Problems*, 1-37, 1987.
4. D. Üstündağ, N.M. Queen, G.K. Skinner, J.E. Bowcock, Two new methods and comparison with the Cambridge MaxEnt Package, *Maximum Entropy and Bayesian Methods*, Kluwer Academic Publishers, Netherlands, 295-301, 1991.
5. D. Üstündağ and M. Cevri, Estimating of sinusoids from noisy data using Bayesian inference with simulated annealing, *Wseas Transactions on Signal Processing*, 432-441, 2008.
6. G.L. Bretthorst, *Bayesian Spectrum Analysis and Parameter Estimation*, Lecture Notes in Statistics, Springer-Verlag, New York, 1997.
7. H.L. Harney, *Bayesian Inference: Parameter Estimation and Decisions*, Springer-Verlag, Berlin Heidelberg, 2003.
8. P. Weillin, R. Gayllord and S. Kamin, *An Introduction to Programming with Mathematica*, Cambridge University Press, 2005.
9. W.H. Press, B.P. Flannery, S.A. Teukolshy and W.T. Vetterling, *Numerical Recipes in C: The Art of Computing*, Cambridge University Press, 1995.
10. T.R. Hooke and T.A. Jeeves, Direct Search Solution of Numerical and Statistical Problems, *Journal of Association of Computer Machinery*, 212-229, 1962.
11. A. Corana, M. Marchesi, C. Martini and S. Ridella, Minimizing multimodal functions of continuous variables with the simulated annealing algorithm, *ACM Transactions on Mathematical Software* **13**, 262-80, 1987.
12. W.L. Goffe, G.D. Ferrier and J. Rogers, Global optimization of statistical functions with simulated annealing, *Journal of Econometrics* **60**, 65-100, 1994.
13. J. Ireland, Simulated annealing and Bayesian posterior distribution analysis applied to spectral emission line fitting, *Journal of Solar Physics* **243**, 237-252, 2007.
14. N. Metropolis, A. Rosenbluth, M. Rosenbluth, A.M. Teller and E. Teller, Equation of states calculations by fast computing machines, *Journal of Chemical Physics* **21**, 1087-1092, 1953.
15. H. Jeffreys, *Theory of Probability*, Oxford University Press, London, 1961.
16. D. Stefankovic, S. Vempala and E. Vigoda, Adaptive simulated annealing: A near – optimal connection between sampling and counting. *Journal of the ACM (JACM)* **56**, 2009.