We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,900
Open access books available

185,000
International authors and editors

200M
Downloads

Our authors are among the

154
Countries delivered to

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com



Chapter 4

Bayesian Recovery of Sinusoids with Simulated Annealing

Dursun Üstündag and Mehmet Cevri

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/50449

1. Introduction

The ultimate goal of collecting data is to gain meaningful information about a physical system. However, in many situations, the quantities that we would like to determine are different from the ones which we are able to have measured. If the data we measured depends on the quantities we want, it contains at least some information about them. Therefore, our general interest is to subtract this information from data.

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. The physical system is described by a vector function f in the form:

$$y(t) = f(t;\boldsymbol{\theta}), \tag{1}$$

where t represents time. In many experiments, the recorded data $\mathbf{D} = \{d_1, d_2, ..., d_N\}$ are sampled from an unknown function $y(t)$ together with errors $e(t)$ at discrete times $(t_1, t_2, ..., t_N)^T$:

$$d_i = y(t_i) + e(t_i), \quad (i = 1, ..., N). \tag{2}$$

The measurement errors $e(t)$ are generally assumed to be drawn independently from a zero mean Gaussian probability distribution with a standard deviation of σ. On the other hand, different signal models correspond to different choices of signal model function $f(t, \boldsymbol{\theta})$. In this chapter, we restrict our attention to the static[1] sinusoidal model given by

[1] Static refers to that the amplitudes of the sinusoids do not change with time.

INTECH
open science | open minds

© 2012 Üstündag and Cevri, licensee InTech. This is an open access chapter distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/3.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

$$\begin{aligned} f(t;\boldsymbol{\theta}) &= \sum_{j=1}^{\rho} B_j \cos(t\omega_j) + B_{j+\rho} \sin(t\omega_j) \\ &= \sum_{j=1}^{2\rho} G_j(t;\boldsymbol{\omega}) B_j \end{aligned} \tag{3}$$

where B_j's represent the amplitudes of the signal model and

$$G(t;\boldsymbol{\omega}) = \begin{bmatrix} \cos(\omega_1 t_1) & \dots & \cos(\omega_\rho t_1) & \sin(\omega_1 t_1) & \dots & \sin(\omega_\rho t_1) \\ \cos(\omega_2 t_2) & \dots & \cos(\omega_\rho t_2) & \sin(\omega_2 t_2) & \dots & \sin(\omega_\rho t_2) \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \cos(\omega_1 t_N) & \dots & \cos(\omega_\rho t_N) & \sin(\omega_1 t_N) & \dots & \sin(\omega_\rho t_N) \end{bmatrix}. \tag{4}$$

The goal of data analysis is usually to use the observed data $\mathbf{D}$, to infer the values of parameters $\boldsymbol{\theta} = \{\boldsymbol{\omega}, \mathbf{B}\}$. Besides estimating the values of the parameters, there are two additional important problems. The one is to obtain an indication of the uncertainties associated with each parameter, i.e. some measures of how far they are away from the true parameters. The other we will not consider here is to assess whether or not the model is appropriate for explaining the data.

Sinusoidal parameter estimation in additive white noise within a Bayesian framework has been an important problem in signal processing and still now is an active research area because of its wide variety of applications in multiple disciplines such as sonar, radar, digital communications and biomedical engineering. The purpose of this research is therefore to develop accurate and computationally efficient estimators for sinusoidal parameter, namely, amplitudes and frequencies. In above problem, one may or may not know the number of sinusoids. When it is unknown, it is called model selection (Andrieu and Doucet, 1999; Üstündag, 2011) and is not subject to this chapter. Under an assumption of a known number of sinusoids, several algorithms have already been used in the parameter estimation literature, such as least-square fitting (Press et al., 1995), discrete Fourier transform (Cooley & Tukey, 1964), and periodogram (Schuster, 1905). With least square fitting, the model is completely defined and the question remaining is to find the values of the parameters by minimizing the sum of squared residuals. 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. In 1987, Jaynes derived periodogram directly from the principles of Bayesian inference. After his work, 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 (Bretthorst, 1990; Üstündag et al., 1989, 1991; Harney, 2003; Gregory, 2005; Üstündag & Cevri, 2008, 2011; Üstündag, 2011).

In this chapter, we studied Bayesian recovery of sinusoids using estimation approach proposed by Bretthorst for a general signal model equation and combined it with a simulated annealing (SA) algorithm to obtain a global maximum of the posterior probability density function (PDF) $P\left(\boldsymbol{\omega}\middle|\mathbf{D},I\right)$ for frequencies . Unfortunately, conventional algorithms (Press et al., 1995) based on the gradient direction fail to converge for this problem. Even when they converge, there is no assurance that they have found the global, rather than a local maximum. This is because the logarithm of the PDF $P\left(\boldsymbol{\omega}\middle|\mathbf{D},I\right)$ is so sharply peaked and highly nonlinear function of frequencies. In this respect, a pattern search algorithm described by Hook-Jevees (Hooke & Jevees, 1962) to overcome this problem has already been used by some researchers in literature of estimation. However, we have found out that this approach does not converge unless the starting point is much closer to the optimum so that we have developed an algorithm in which this Bayesian approach is combined with a simulated annealing (SA) algorithm (Kirkpatrick, et al., 1983; Corana et al., 1987; Goffe et al., 1994; Ingber, 1996), which is a function optimization strategy based on an analogy with the creation of crystals from their melts. This explores the entire surface of the posterior PDF for the frequencies and tries to maximize it while moving both uphill and downhill steps, whose sizes are controlled by a parameter Γ that plays the role of the temperature in a physical system. By slowly lowering the temperature Γ towards zero according to a properly chosen schedule, one can show that the globally optimal solutions are approached asymptotically. Thus, it is largely independent of the starting values, often a critical input in conventional algorithms, and also offers different approach to finding parameter values of sinusoids through a directed, but random, search of the parameter space. In this context, an algorithm of this Bayesian approach is developed and coded in Mathematica programming language (Wellin at al., 2005) and also tested for recovering noisy sinusoids with multiple frequencies. Furthermore, simulation studies on synthetic data sets of a single sinusoid under a variety of signal to noise ratio (SNR) are made for a comparison of its performance with Cramér-Rao lower bound (CRLB), known as a lower limit on variance of any unbiased estimator. The simulations results support its effectiveness.

2. Bayesian parameter estimation

Let us now reconsider above problem within a Bayesian context (Bernardo & Smith, 2000; Bretthorst, 1988; Gregory, 2005; Harney, 2003; Jaynes, 2003; Ruanaidh & Fitzgerald, 1996; Üstündag & Cevri, 2008*).* As with all Bayesian calculations, the first step is to write down Bayes' theorem for the joint PDF for all of the unknown parameters $\{\boldsymbol{\omega},\mathbf{B},\sigma^2\}$:

$$P\left(\boldsymbol{\omega},\mathbf{B},\sigma^2\middle|\,\mathbf{D},\mathbf{I}\right)=\frac{1}{P\left(\mathbf{D}\middle|\,\mathbf{I}\right)}P\left(\boldsymbol{\omega},\mathbf{B},\sigma^2\middle|\,\mathbf{I}\right)P\left(\mathbf{D}\middle|\,\boldsymbol{\omega},\mathbf{B},\sigma^2,\mathbf{I}\right). \tag{5}$$

The quantity $P\left(\boldsymbol{\omega},\mathbf{B},\sigma^2\middle|\,\mathbf{I}\right)$ is called the prior PDF; it represents prior knowledge of the parameters $\{\boldsymbol{\omega},\mathbf{B},\sigma^2\}$ given the information $\mathbf{I}$. The sampling distribution $P\left(\mathbf{D}\middle|\,\boldsymbol{\omega},\mathbf{B},\sigma^2,\mathbf{I}\right)$ is the likelihood (Edwards, 1972) of the data $\mathbf{D}$, given the model parameters. The probability

function $P(\mathbf{D}\,|\,\mathbf{I})$ is the evidence, which is a constant for parameter estimation and is used here for normalizing the posterior PDF $P\left(\boldsymbol{\omega},\mathbf{B},\sigma^2\,|\,\mathbf{D},\mathbf{I}\right)$. Therefore, it can be dropped in Equation (5) so that the joint PDF for the unknown parameters turns out to be the following form:

$$P\left(\boldsymbol{\omega},\mathbf{B},\sigma^2\,|\,\mathbf{D},\mathbf{I}\right)\propto P\left(\boldsymbol{\omega},\mathbf{B},\sigma^2\,|\,\mathbf{I}\right)P\left(\mathbf{D}\,|\,\boldsymbol{\omega},\mathbf{B},\sigma^2,\mathbf{I}\right). \tag{6}$$

A key component in Bayes theorem is the likelihood function $P\left(\mathbf{D}\,|\,\boldsymbol{\omega},\mathbf{B},\sigma^2,\mathbf{I}\right)$ which is proportional to the probability density of the noise. If its standard deviation σ is assumed to be known, then the likelihood function takes on the following form:

$$P\left(\mathbf{D}\,|\,\boldsymbol{\omega},\mathbf{B},\sigma^2,\mathbf{I}\right)=(2\pi\sigma^2)^{\frac{-N}{2}}\,exp\left(-\frac{NQ}{2\sigma^2}\right), \tag{7}$$

where the exponent Q is defined as follows

$$Q=\sum_{i=1}^{N}\left(d_i-\sum_{j=1}^{2\rho}B_jG_j(t_i;\boldsymbol{\omega})\right)^2. \tag{8}$$

This is equivalent to

$$Q=\overline{\mathbf{D}^2}-\frac{2}{N}\sum_{j=1}^{2\rho}\sum_{i=1}^{N}d_iB_jG_j\left(t_i;\boldsymbol{\omega}\right)+\frac{1}{N}\sum_{j=1}^{2\rho}\sum_{k=1}^{2\rho}\Omega_{jk}B_jB_k, \tag{9}$$

where

$$\overline{\mathbf{D}^2}=\frac{1}{N}\sum_{i=1}^{N}d_i^{\,2} \tag{10}$$

and

$$\Omega_{jk}=\sum_{i=1}^{N}G_j\left(t_i;\boldsymbol{\omega}\right)\otimes G_k\left(t_i;\boldsymbol{\omega}\right),\quad (j,k=1,...,2\rho). \tag{11}$$

In order to obtain the posterior PDF for $\boldsymbol{\omega}$, Equation (6) can be integrated with respect to the nuisance parameters $\mathbf{B}$ under the knowledge of σ^2:

$$P\left(\boldsymbol{\omega},\sigma\,|\,\mathbf{D},\mathbf{I}\right)=\int P\left(\boldsymbol{\omega},\mathbf{B},\sigma\,|\,\mathbf{D},\mathbf{I}\right)d\mathbf{B}. \tag{12}$$

With the choices of an uniform prior PDF or independent Gaussians distributions with mean zero and known standard deviation for the amplitudes, the integral equations in (12) turn out to be a Gaussian integral which can be evaluated analytically (Bretthorst, 1988). To do this it is simply to convert the square matrix $\boldsymbol{\Omega}$ into a special type of matrix- a so called diagonal matrix- that shares the same fundamental properties of the underlying matrix. In other words, it is equivalent to transforming the underlying system of equations into a special set of coordinate axes. Therefore, this diagonalization process (Bretthorst, 1988) effectively introduces new orthonormal model functions,

$$H_j(t;\boldsymbol{\omega}) = \frac{1}{\sqrt{\lambda_j}} \sum_{k=1}^{2\rho} \upsilon_{jk} G_k(t;\boldsymbol{\omega}),\ (j = 1,...,2\rho), \tag{13}$$

and also gives a new expression for the signal model function in Equation (3):

$$f(t;\boldsymbol{\theta}) = \sum_{k=1}^{2\rho} A_k H_k(t;\boldsymbol{\omega}), \tag{14}$$

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

$$A_k = \sqrt{\lambda_k} \sum_{j=1}^{2\rho} B_j \upsilon_{kj}, (k = 1,...,2\rho) \tag{15}$$

where υ_{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 (9) and defining

$$h_j = \sum_{i=1}^{N} d_i H_j(t_i;\boldsymbol{\omega}), \quad (j = 1,2,...,2\rho) \tag{16}$$

to be the projection of the data onto the orthonormal model functions $H_j(t;\boldsymbol{\omega})$, we can then proceed to perform the 2ρ integration over A_j in Equation (12) to obtain

$$P(\mathbf{D}\,|\,\boldsymbol{\omega},\mathbf{B},\sigma,I) \propto \sigma^{-N+2\rho} exp\left(-\frac{N\overline{\mathbf{D}^2} - 2\rho\overline{\mathbf{h}^2}}{2\sigma^2}\right) \tag{17}$$

with

$$\overline{\mathbf{h}^2} = \frac{1}{2\rho}\sum_{\mathrm{j}=1}^{2\rho} \mathrm{h}_\mathrm{j}^2. \tag{18}$$

This represents the mean-square of the observed projections. If σ is known, the joint posterior probability density of the parameters $\boldsymbol{\omega}$, conditional on the data and our knowledge of σ is given by

$$P\left(\boldsymbol{\omega}\middle|\mathbf{D},\sigma,I\right) \propto exp\left(\frac{\rho\overline{\mathbf{h}^2}}{\sigma^2}\right). \tag{19}$$

If there is no prior information about noise, then σ is known as a nuisance parameter and must also be eliminated by integrating it out. Using Jeffreys prior (Jeffreys, 1961) $\frac{1}{\sigma}$ and integrating Equation (12) over σ we obtain

$$P\left(\boldsymbol{\omega}\middle|\mathbf{D},I\right)\propto\left(1-\frac{2\rho\overline{\mathbf{h}^2}}{N\overline{\mathbf{D}^2}}\right)^{\frac{2\rho-N}{2}}. \tag{20}$$

This has the form of the "Student *t*- distribution". As well as determining the values of the $\boldsymbol{\omega}$ parameters for which the posterior PDF for $\boldsymbol{\omega}$ is a maximum, it is also desirable to compute uncertainties associated with them. To do this, let us assume the case where σ is known and let $\hat{\boldsymbol{\omega}}$ represent the estimated values of $\boldsymbol{\omega}$. Following to Bretthorst's work (Bretthorst, 1988), we can expand the function $\mathbf{h}^2$ in a Taylor series at the point $\hat{\boldsymbol{\omega}}$, such that

$$P(\boldsymbol{\omega}\,|\,\mathbf{D},I)\propto exp\left(-\sum_{j=1}^{\rho}\sum_{k=1}^{\rho}\frac{b_{jk}}{2\sigma^2}\Delta_j\Delta_k\right) \tag{21}$$

with b_{jk} defined as $b_{jk}=-\rho\left.\frac{\partial^2\overline{\mathbf{h}^2}}{\partial\omega_j\partial\omega_k}\right|_{\substack{\omega_j=\hat{\omega}_j\\ \omega_k=\hat{\omega}_k}}$ and $\Delta_j=(\hat{\omega}_j-\omega_j)$ for a single frequency case. For an arbitrary model the matrix b_{jk} cannot be calculated analytically; however, it can be evaluated numerically. The calculations of the mean and standard deviations for $\boldsymbol{\omega}$ parameters require for evaluating Gaussian integrals by first changing to the orthogonal variables as was done above with the amplitudes. Let τ_j and u_{kj} represent the jth eigenvalue and eigenvector of the matrix b_{jk}, respectively. Then the new orthogonal variables are given by

$$s_j=\sqrt{\tau_j}\sum_{k=1}^{\rho}\Delta_k u_{kj},\quad \Delta_j=\sum_{k=1}^{\rho}\frac{s_k u_{jk}}{\sqrt{\tau_k}}. \tag{22}$$

By using these orthogonal variables to perform the ρ Gaussian integrals, the estimate variance for ω_k can be calculated:

$$\sigma_\omega^2=\left\langle\sigma^2\right\rangle\sum_{j=1}^{\rho}\frac{u_{jk}^2}{\tau_j}. \tag{23}$$

Therefore, the approximations for $\boldsymbol{\omega}$ can be implemented in the form:

$$\langle\omega_j\rangle=\hat{\omega}_j\pm\sigma_{\omega_j},\quad (j=1,2,3,...,\rho). \tag{24}$$

In agree with Bretthorst, the expectation values of the amplitudes are given by $\langle A_j\rangle=h_j$. From Equation (15) the expected values for the old amplitudes $\mathbf{B}$ becomes

$$\langle B_k\rangle=\sum_{j=1}^{2\rho}\frac{1}{\sqrt{\lambda_j}}h_j\upsilon_{jk}. \tag{25}$$

The uncertainty in A_j is $\pm\sigma$, so that the corresponding uncertainty in $\mathbf{B}$ is

$$\sigma_{\mathbf{B}} = \sigma \sum_{j=1}^{2\rho} \frac{1}{\sqrt{\lambda_j}} \upsilon_{jk} \upsilon_{jk} \quad (k = 1,2,3,...,2\rho) \tag{26}$$

3. Implementation of Simulated Annealing (SA) algorithm

Bayesian approach introduced by Bretthorst is briefly summarized in section 2 but, it is referred to Bretthorst's works (Bretthorst, 1988) for more detail information. Consequently, Bayesian parameter estimation turns into a global optimization problem which is a task to find the best possible solution $\boldsymbol{\omega}$ for Equations (19) or (20) satisfying

$$\max_{\boldsymbol{\omega} \in \Xi_\rho} \left\{ P\left(\boldsymbol{\omega} \middle| \mathbf{D}, I\right) \right\}, \tag{27}$$

in the solution space $\Xi_\rho = \{0, \pi\}$ .Because there is no variation at negative frequencies and the highest possible frequencies corresponds to wave that under goes a complete cycle in two unit intervals, so the lower limit on the range is 0 and all the variation is accounted for by frequencies less than π .

Over last few decades, researchers have developed many computational algorithms to address such type of global optimization problems (Metropolis et al., 1953; Jeffreys, 1961; Kirkpatrick, et al., 1983; Laarhoven & Aarts, 1987; Stefankovic et al., 2009). Although there are numerous algorithms which are suggested to achieve this goal, few of them are capable of locating it effectively. Therefore, we follow the SA algorithm, suggested by Corana (Corana et. al., 1987) and modified by Goffe (Goffe et al., 1994), which is a kind of probabilistic algorithm for finding the global optimum of Equation (27) although its various alternative versions have already been used in statistical applications. A brief review of the most work on many algorithms based on that of SA, together with areas of applications is provided by Ingber and Binder (Binder, 1986; Ingber, 1994).

The algorithm begins with an initial guess of the frequencies $\boldsymbol{\omega}^0$, a trial step-length vector $\mathbf{v}^0$ and a global parameter Γ_0 (called the initial temperature). Each step of the SA algorithm replaces the current frequency with randomly generated new frequency. In other words, the next candidate point $\boldsymbol{\omega}^{k+1}$ is generated by varying one component $j \in \{1,...,\rho\}$ of the current point $\boldsymbol{\omega}^k$ at a time:

$$\omega_j^{k+1} = \omega_j^k + \xi v_j^k, \tag{28}$$

where ξ is a uniformly distributed random number from the interval [-1,1] and $\mathbf{v}^k = \left\{ v_j^k \right\}$ is a step vector. The function value of $P\left(\boldsymbol{\omega}^{k+1} \middle| \mathbf{D}, I\right)$ is then computed. If

$$P\left(\boldsymbol{\omega}^{k+1} \middle| \mathbf{D}, I\right) \geq P\left(\boldsymbol{\omega}^{k} \middle| \mathbf{D}, I\right) \tag{29}$$

then the point $\boldsymbol{\omega}^{k+1}$ is accepted as the $(k+1)$th iteration point, it is replaced with $\boldsymbol{\omega}^{k}$ and algorithm moves uphill. If $P\left(\boldsymbol{\omega}^{k+1}\middle|\mathbf{D},I\right)$ is the largest posterior PDF, denoted as $P_{opt}\left(\boldsymbol{\omega}\middle|\mathbf{D},I\right)$ its value and $\boldsymbol{\omega}^{k+1}$ are recorded since this is the best current value of the optimum. This forces the system toward a state corresponding to a local maximum or possibly a global maximum. However, most large optimization problems, like the one given in Equation (27), have many local maxima and optimization algorithm is therefore often trapped in a local maximum. To get out of a local maximum, a decrease of the function value $P\left(\boldsymbol{\omega}^{k+1}\middle|\mathbf{D},I\right)$ is accepted with a certain probability. This is accomplished by the Metropolis-criterion (Metropolis et al., 1953) which is based on the changes of obtaining new state with the posterior PDF of frequencies value, defined as

$$\begin{aligned} p &= \frac{\exp\left(-\dfrac{P\left(\boldsymbol{\omega}^{k+1}\middle|\mathbf{D},I\right)}{\Gamma^{k}}\right)}{\exp\left(-\dfrac{P\left(\boldsymbol{\omega}^{k+1}\middle|\mathbf{D},I\right)}{\Gamma^{k}}\right)+\exp\left(-\dfrac{P\left(\boldsymbol{\omega}^{k}\middle|\mathbf{D},I\right)}{\Gamma^{k}}\right)}, \\ &\approx \exp\left(-\frac{\Delta P}{\Gamma^{k}}\right) \end{aligned} \tag{30}$$

where ΔP represents difference between the present and previous posterior PDF values of frequencies, e.g., $\Delta P = P\left(\boldsymbol{\omega}^{k+1}\middle|\mathbf{D},I\right) - P\left(\boldsymbol{\omega}^{k}\middle|\mathbf{D},I\right)$. Whenever

$$P\left(\boldsymbol{\omega}^{k+1}\middle|\mathbf{D},I\right) < P\left(\boldsymbol{\omega}^{k}\middle|\mathbf{D},I\right) \tag{31}$$

p is computed and compared to p', a uniformly distributed random number from the interval $(0,1)$. If $p > p'$, the new point $\boldsymbol{\omega}^{k+1}$ is accepted and replaced with $\boldsymbol{\omega}^{k}$ and the algorithm moves downhill, i.e. lower temperatures and larger differences in posterior PDF's values. 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 step vector $\mathbf{v}^{k}$ is adjusted by the following rule:

$$v_j^{k+1} = \begin{cases} v_j^{k}\left(1 + c_j \dfrac{\left(n_j - 0.6n_s\right)}{0.4n_s}\right) & if\ \dfrac{n_j}{n_s} > 0.6 \\ \dfrac{v_j^{k}}{\left(1 + c_j \dfrac{\left(0.4n_s - n_j\right)}{0.4n_s}\right)} & if\ \dfrac{n_j}{n_s} < 0.4 \\ v_j^{k} & \text{otherwise} \end{cases} \tag{32}$$

where n_j is the number of accepted moves along the direction j and the parameter c_j, which is initially defined by user, controls the step variation along the j th direction. The aim of these variations in a step length is to maintain average percentage of accepted moves at about one-half of the total number of moves.

An alternative step size of above SA algorithm is given by replacing Equation (28) with

$$\omega_j^{k+1} = \omega_j^k + \Gamma_k \sqrt{\gamma_j{}^k} N(0,1),\ (j = 1,2,...,\rho). \tag{33}$$

where $N(0,1)$ is a standard Normal distribution, Γ_k represents the temperature at the k th iteration and rescales the step size and ordering ω_j. On the other hand, $\gamma_j{}^k$ is the CRLB for the j th component of angular frequencies at the $(k+1)$ th iteration (Ireland, 2007) and provides a theoretical lower limit on how precisely parameter ω_j can be extracted from noisy measurements. In this respect, it is defined in the form:

$$\gamma_j = \mathbf{J}^{-1}(\omega_j) \tag{34}$$

where the Fisher information matrix $\mathbf{J}(\boldsymbol{\omega})$ (Kay, 1993), defined by

$$\mathbf{J}(\boldsymbol{\omega}) = E\left[\frac{\partial^2 \ln P(\mathbf{D}|\boldsymbol{\omega}, I)}{\partial \boldsymbol{\omega}^2}\right] = \frac{1}{\sigma^2}\sum_{j=1}^{N} \frac{\partial f_j(\boldsymbol{\omega})}{\partial \boldsymbol{\omega}}\left(\frac{\partial f_j(\boldsymbol{\omega})}{\partial \boldsymbol{\omega}}\right)^T, \tag{35}$$

is an expectation of the second derivatives of the signal function with respect to $\boldsymbol{\omega}$. Assuming that the matrix $\mathbf{J}(\boldsymbol{\theta})$ is diagonal for a large N so that its inversion is straightforward. In this case, the diagonal elements yield the lower bound (Stoica et al., 1989; Lagrange, 2005) for the variance of $\hat{\boldsymbol{\omega}}$ asymptotically and we can write,

$$\gamma_j = \mathrm{var}_{CRLB}(\hat{\omega}_j) \cong \frac{24\hat{\sigma}^2}{N^3(\hat{B}_j^2 + \hat{B}_{j+\rho}^2)}, \tag{36}$$

where $\hat{\sigma}^2$ represents the estimated variance of the noise and is described in Equation (41). This whole cycle is then repeated n_Γ times, after which the temperature is decreased by a factor $\alpha_\Gamma \in (0,1)$. This process is generally called annealing (or cooling) schedule which is the heart of the algorithm and effects the number of times the temperature is decreased. If a fast cooling takes place then the problem will be trapped in a local maximum. Therefore, there are various annealing schedules suggested by different researchers (Ingber, 1994; Stefankovic, et al., 2009) for lowering the temperature but we choose the following:

$$\Gamma_{k+1} = \alpha_\Gamma \frac{\Gamma_k}{\exp\{(\alpha_\Gamma - 1)k^{\frac{1}{\rho}}\}} \tag{37}$$

Because of being exponential rather than logarithmic, it is sometimes known as simulated quenching (SQ) (Vasan et al., 2009; Aguiare et al., 2012). In case of a well conditioned estimation problem like, say, frequency estimation problem in signal processing, it is clear that the convenience of SA algorithm, together with a need for some global search over local optima, makes a strong practical case for the use of SQ. Therefore, different parameters have different finite ranges and different annealing time- dependent sensitivities. Classical annealing algorithms have distributions that sample infinite ranges and there is no decision for considering differences in each parameter dimension; e.g. different sensitivities might be necessary to use different annealing schedules. This requires the development of a new PDF to embed in the desired features (Ingber, 1994) so that it leads to variants of SA that justifies exponential temperature annealing schedules.

Termination of the algorithm occurs when average function value of the sequences of the points after each $n_s \times n_\Gamma$ step cycle reaches a stable state:

$$\begin{aligned} &\left| P_k\left(\boldsymbol{\omega}\middle|\mathbf{D},I\right) - P_{k-l}\left(\boldsymbol{\omega}\middle|\mathbf{D},I\right)\right| \le \varepsilon \ \ (l=1,...,4) \\ &\left| P_k\left(\boldsymbol{\omega}\middle|\mathbf{D},I\right) - P_{opt}\left(\boldsymbol{\omega}\middle|\mathbf{D},I\right)\right| \le \varepsilon \end{aligned} \tag{38}$$

where ε is a small positive number defined by user and l indicates the last four successive iteration values of the posterior PDF of the frequencies that are being stored. Further details of the algorithm initialization are problem-dependent and are given in Section 5.

4. Power spectral density

Before we discuss the computer simulated examples, there is something we need to say about how to display the results. The usual way the result from a spectral analysis is displayed is in the form of a power spectral density that shows the strength of the variation (energy) as a function of frequency. In Fourier transform spectroscopy this is typically taken as the squared magnitude of the discrete Fourier transform of the data. In order to display our results in the form of a power spectral density (Bretthorst, 1988; Gregory, 2005), it is necessary to give an attention to its definition that shows how much power is contained in a unit frequency. According to Bretthorst (Bretthorst, 1988) the Bayesian power spectral density is defined as the expected value of the power of the signals over the joint posterior PDF:

$$S(\boldsymbol{\omega}) = \frac{N}{2}\int \sum_{j=1}^{\rho}\left(B_j^2 + B_{j+\rho}^2\right) P\left(\boldsymbol{\omega},\mathbf{B}\middle|D,\sigma^2,I\right) dB_1 ... dB_{2\rho}. \tag{39}$$

Performing integrals analytically over $B_1,...,B_{2\rho}$ by using these orthonormal model functions defined in section 2, the power spectral density can therefore be approximated as

$$S(\boldsymbol{\omega}) = 2\left(\sigma^2 + \sum_{j=1}^{2\rho} h_j^2(\hat{\boldsymbol{\omega}})\right) \sum_{k=1}^{\rho} \sqrt{\frac{b_{kk}}{2\pi\sigma^2}} \exp\left(-\frac{b_{kk}(\hat{\omega}_k - \omega)^2}{2\sigma^2}\right). \tag{40}$$

This function stresses information about the total energy carried by the signal and about the accuracy of each line. In the next section, we will present some numerical examples how well this technique works.

5. Computer simulated examples

To verify the performance of the algorithm, we generated a simulated data vector according to one, two and with five sinusoids. Here t_i runs over the symmetric time interval $-T$ to T in $(2T+1)=512$ integer steps and the components of the noise e_i are generated from the zero mean Gaussian distribution with a known deviation σ, initially and added to the simulated data. However, one of the interests in an experiment is also to estimate noise power σ^2 so that it is assumed to be unknown. In agreement with Bretthorst, this is given in the following form:

$$\langle \sigma^2 \rangle = \frac{\left(N\overline{\mathbf{D}^2} - 2\rho\overline{\mathbf{h}^2} \right)}{N-2\rho-2}\left(1 \pm \sqrt{\frac{2}{N-2\rho-4}} \right). \tag{41}$$

Clearly, it is seen that the accuracy of the estimate depends on how long we sample and the signal-to-noise ratio (SNR), defined as the ratio of the root mean square of the signal amplitude to the noise σ^2. In addition, one may also get the following expression of SNR:

$$\text{SNR} = \sqrt{\frac{2\rho}{N}\left(1 + \frac{\overline{\mathbf{h}^2}}{\sigma^2} \right)}. \tag{42}$$

When the standard deviation of the noise is unknown, an empirical SNR is obtained by replacing σ^2 in Equation (42) with the estimated noise variance in (41).

In our first example, we generate the data set from the following equation:

$$d_i = 0.001 + 0.5403\cos\left(0.3t_i\right) - 0.8415\sin\left(0.3t_i\right) + e_i \quad (i = 1,...,512). \tag{43}$$

We then carried out the Bayesian analysis of the simulated data, assuming that we know the mathematical form of the model but not the value of the parameters. We first gave starting values to the list of frequencies to begin a multidimensional search for finding a global maximum of the posterior PDF of the frequencies $\boldsymbol{\omega}$ given in Equations (19) or (20). As an initial estimate of the frequencies $\boldsymbol{\omega}^0$ for the maximization procedure, it is possible to take random choices from the interval $(0,\pi)$. 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.

In agreement with Corana (Corana, et al., 1987), reasonable values of the parameters that control the SA algorithm are chosen as $n_s = 20$, $n_\Gamma = \max(100, 5\rho)$ and $c_j = 2, (j = 1,...,\rho)$. Then the global optimization algorithm starts at some high temperature $\Gamma_0 = 100$. Thus the

sequence of points is generated until a sort of equilibrium is approached; that is a sequence of points $\boldsymbol{\omega}^0, \boldsymbol{\omega}^1, \boldsymbol{\omega}^2, \ldots$ whose average value of $P\left(\boldsymbol{\omega}\middle|\mathbf{D}, I\right)$ reaches a stable value as iteration proceeds. During this phase the step vector $\mathbf{v}$ is periodically adjusted by the rule defined in Equation (32). The best point $\boldsymbol{\omega}$ reached so far is recorded. After thermal equilibration, the temperature Γ is reduced by using the annealing scheduled in Equation (37) with a factor $\alpha_\Gamma = 0.85$ and a new sequence is made starting from this best point $\boldsymbol{\omega}$, until thermal equilibrium is reached again and so on. The SA algorithm first builds up a rough view of the surface by moving large step lengths. As the temperature Γ falls and the step decreases, it is slowly focuses on the most promising area. Therefore it proceeds toward better maximum even in the presence of many local maxima by adjusting the step vector that can allow the algorithm to shape the space within it may move, to that within which ought to be as defined by the PDF $P\left(\boldsymbol{\omega}\middle|\mathbf{D}, I\right)$. 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 (38).

	Bayesian Parameter estimation	
	Parameters	Estimated Values
σ : Known N : 512 3ρ : 3 SNR : 0.725464 $\hat{\sigma}$: 1	ω_1	0.2998 ± 0.0005
	B_1	0.6035 ± 0.0589
	B_2	-0.8174±0.0626

Table 1. Computer simulations for a single harmonic frequency model

Once the frequencies are estimated, we then carried on calculating the amplitudes and parameter errors approximately using Equations (25), (23) and (26), respectively. However, an evaluation of the posterior PDF at a given point $\boldsymbol{\omega}$ cannot be made analytically. It requires a numerical calculation of projections onto orthonormal model functions, related to Eigen-decomposition of the $\left(2\rho \times 2\rho\right)$ dimensional matrix $\Omega\left(\boldsymbol{\omega}\right)$. Therefore, the proposed algorithm was coded in *Mathematica* programming language (Wellin, P., et al., 2005), that provides a much flexible and efficient computer programming environment. Furthermore, it also contains a large collection of built-in functions so that it results much shorter computer codes than those written in C or FORTRAN programming languages.

The computer program was run on the workstation with four processors, which of each has got Intel Core 2 Quad Central Processing Unit (CPU), in two cases where the standard deviation of noise is known or not. The output of the computer simulations when $\sigma = 1$ is illustrated in Table 1. The results when σ is unknown are almost similar with that of Table 1. Parameter values are quoted as (*value*) ± (*standard deviation*). It can be seen that a single frequency and amplitudes are recovered very well. The estimated value of SNR and the standard deviation of the noise are also shown in Table 1.

In our second example, we consider a signal model with two close harmonic frequencies:

$$d_i = 0.5403\cos\left(0.3t_i\right) - 0.8415\sin\left(0.3t_i\right) \\ -0.4161\cos\left(0.31t_i\right) - 0.9093\sin\left(0.31t_i\right), \quad (i = 1,...,512). \tag{44}$$

In a similar way, we produced the same size data corrupted by the zero mean Gaussian noise with $\sigma = 1$. We run our *Mathematica* code again in the case where the deviation of noise is unknown. The results, shown in Table 2, indicate that all values of the parameters within the calculated accuracy are clearly recovered.

	Bayesian Parameter estimation	
	Parameters	Estimated Values
σ : Unknown N : 512 3ρ : 6 *SNR* : 1.02936 $\hat{\sigma}$: 0.99639	ω_1	0.3000±0.0009
	ω_2	0.3105±0.0010
	B_1	0.5206±0.0632
	B_2	-0.8698±0.0631
	B_3	-0.3819±0.0634
	B_4	-0.8556±0.0636

Table 2. Computer simulations for two closed harmonic frequency model

In general, we consider a multiple harmonic frequency model signal:

$$d_i = \cos\left(0.1\ t_i + 1\right) + 2\cos\left(0.15t_i + 2\right) + 5\cos\left(0.3t_i + 3\right) \\ +2\cos\left(0.31t_i + 4\right) + 3\cos\left(t_i + 5\right) + e_i \quad (i = 1,...,512). \tag{45}$$

The best estimates of parameters are tabulated in Table 3. Once again, all the frequencies have been well resolved, even the third and fourth frequencies which are too closed 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 $2\pi / N$, the two frequencies are indistinguishable. This is simply because there are no sample points in between the two frequencies in the frequency domain. If $\left|\omega_3 - \omega_4\right| > 2\pi / N$ theoretically the two frequencies can then be distinguished. If $\left|\omega_3 - \omega_4\right|$ is not large enough, the resolution will be very poor. Therefore, it is hard to tell where the two frequencies are located. This is just the inherent problem of the discrete Fourier power spectral density. In this example two frequencies are separated by 0.01, which is less than the Nyquist step size. There is no way by using Fourier power spectral density that one can resolve the closed frequencies less than the Nyquist step. However, Bayesian power spectral density shown in Figure 3 gives us very good results with high accuracy. Finally, we constructed the signal model in Equation (3), whose parameters, amplitudes and frequencies, are randomly

chosen from uniform distribution in the intervals $[-5,5]$ and $[0,\pi]$, respectively and used it to generate data samples of $N=512$ by adding a zero mean Gaussian random noise with $\sigma=1$. The proposal algorithm was rerun for recovering sinusoids from it and the results are tabulated in Table (4). It can be seen that frequencies are specially recovered with high accuracies. Ten frequencies signal model are shown in Fig.5.

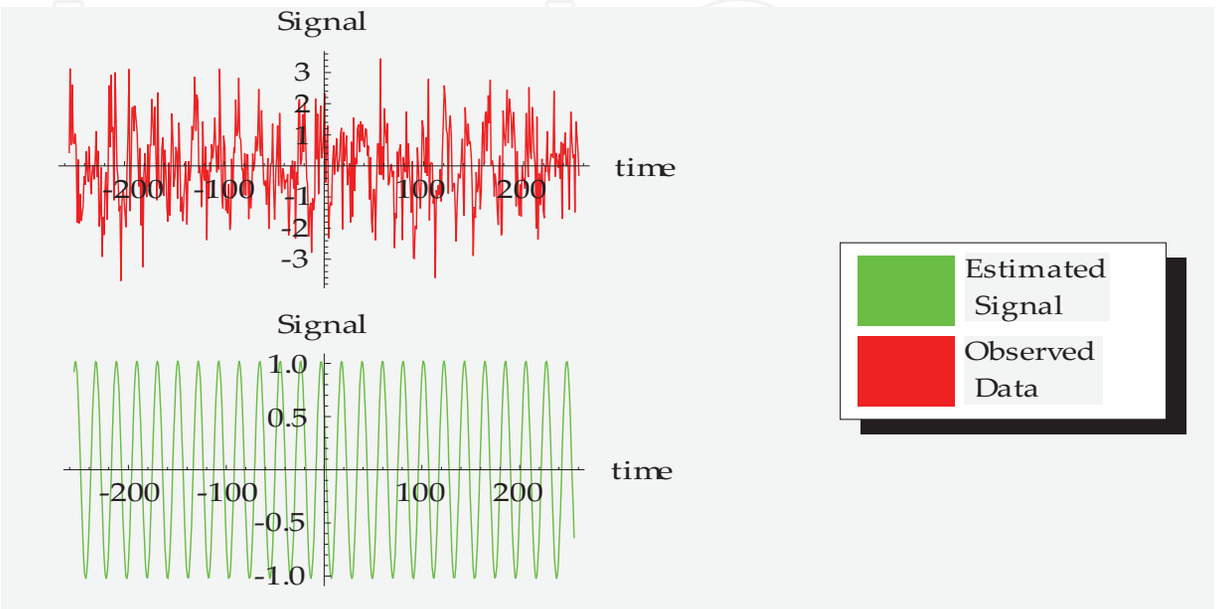


Figure 1. Recovering signal from noisy data produced from a single harmonic frequency signal model.

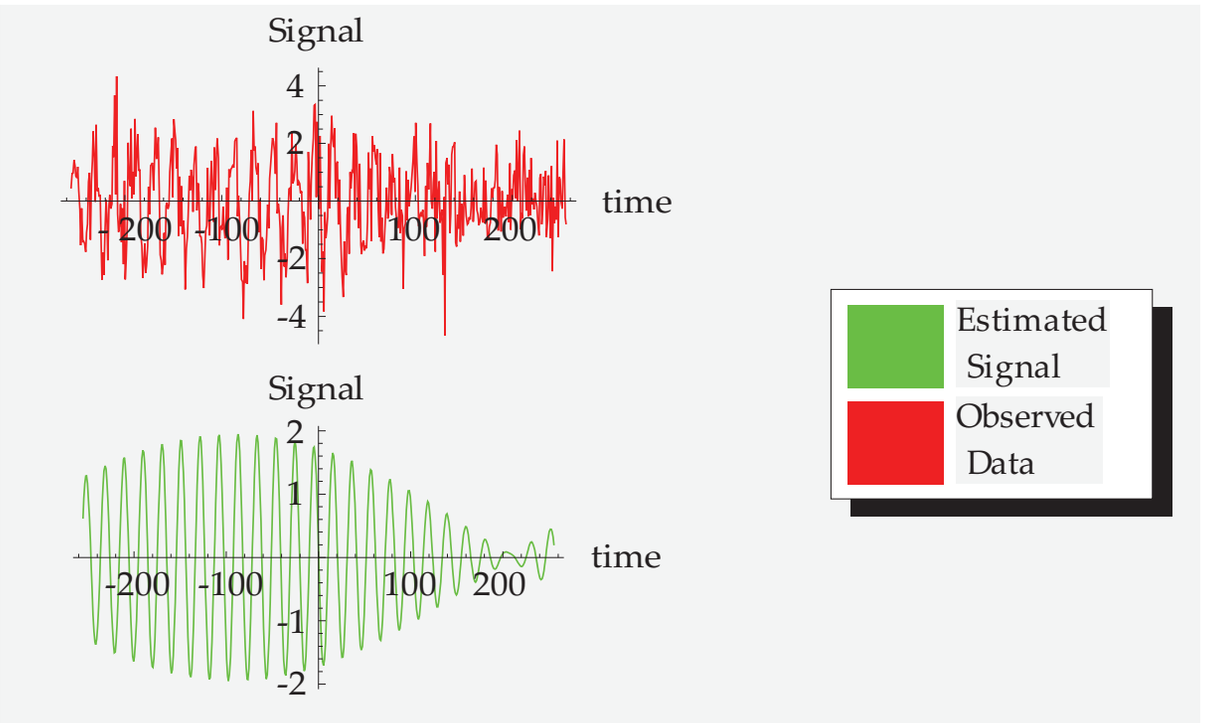


Figure 2. Recovering signals from noisy data produced from two closed harmonic frequency signal model.

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	1.00	0.9905 ± 0.08	0.99 ∓ 0.08
	a_2	2.00	1.9600 ± 0.08	2.08 ± 0.08
	a_3	5.00	5.1058 ± 0.09	4.97 ± 0.08
	a_4	2.00	1.8199 ± 0.09	1.95 ± 0.08
	a_5	3.00	2.9556 ± 0.08	2.92 ± 0.08

Table 3. Computer simulations for a multiple harmonic frequency model[2]

On the other hand, modifications of this algorithm (Üstündag & Cevri, 2008; 2011) have already been made by generating the next candidate in Equation (33) from normal distribution with a mean of the current estimation whose standard deviation is a square root of the CRLB (Ireland, 2007) given in Equation (36), which is a lower limit to the variance of the measurement of the frequencies, so this generates a natural scale size of the search space around their estimated values. It is expected that better solutions lie close the ones that are already good and so normally distributed step size is used. Consequently, the results we obtained are comparable with or higher than those obtained previously. In addition, all the results discussed so far are also consistent with those of Bretthorst (Bretthorst, 1988) and also demonstrate the advantages of the Bayesian approach together with SA algorithm. Moreover it appears to be very reliable, in the sense that it always converged to neighborhood of the global maximum. The size of this neighborhood can be reduced by altering the control parameters of the SA algorithm, but this can be expensive in terms of CPU consumption. Moreover, we initially assumed that the values of the random noise in data were drawn from the Gaussian density with the mean $\mu = 0$ and the standard deviation σ. Figure 4 shows the exact and estimate probability densities of the random noise in data. It is seen that the estimated (dotted) probability density is closer to the true (solid) probability density and the histogram of the data is also much closer to the true probability density. The histogram is known as a nonparametric estimator of the probability density because it does not depend on specified parameters.

Computer simulations had been carried out to compare the performance of the method with the CRLB. To do this, we generated 64 data samples from a single real tone frequency signal

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

model $\left(\omega = 0.3,\ B_1 = B_2 = 1\right)$ and added it to the variety of noise levels. After 50 independent trials under different SNR ratios, the mean square errors (MSE) for the estimated frequencies were obtained and their logarithmic values were plotted with respect to SNR ratios that vary between zero and 20 dB (deci Bell). It can be seen from Figure 6 that the proposed estimator has threshold about 3 dB of the SNR and follows nicely the CRLB after this value. As expected, larger SNR gives smaller MSE. However, many of existing methods in signal processing literature have a MSE that is close to the CRLB when the SNR is more than 20 dB and they usually perform poorly when the SNR is decreased.

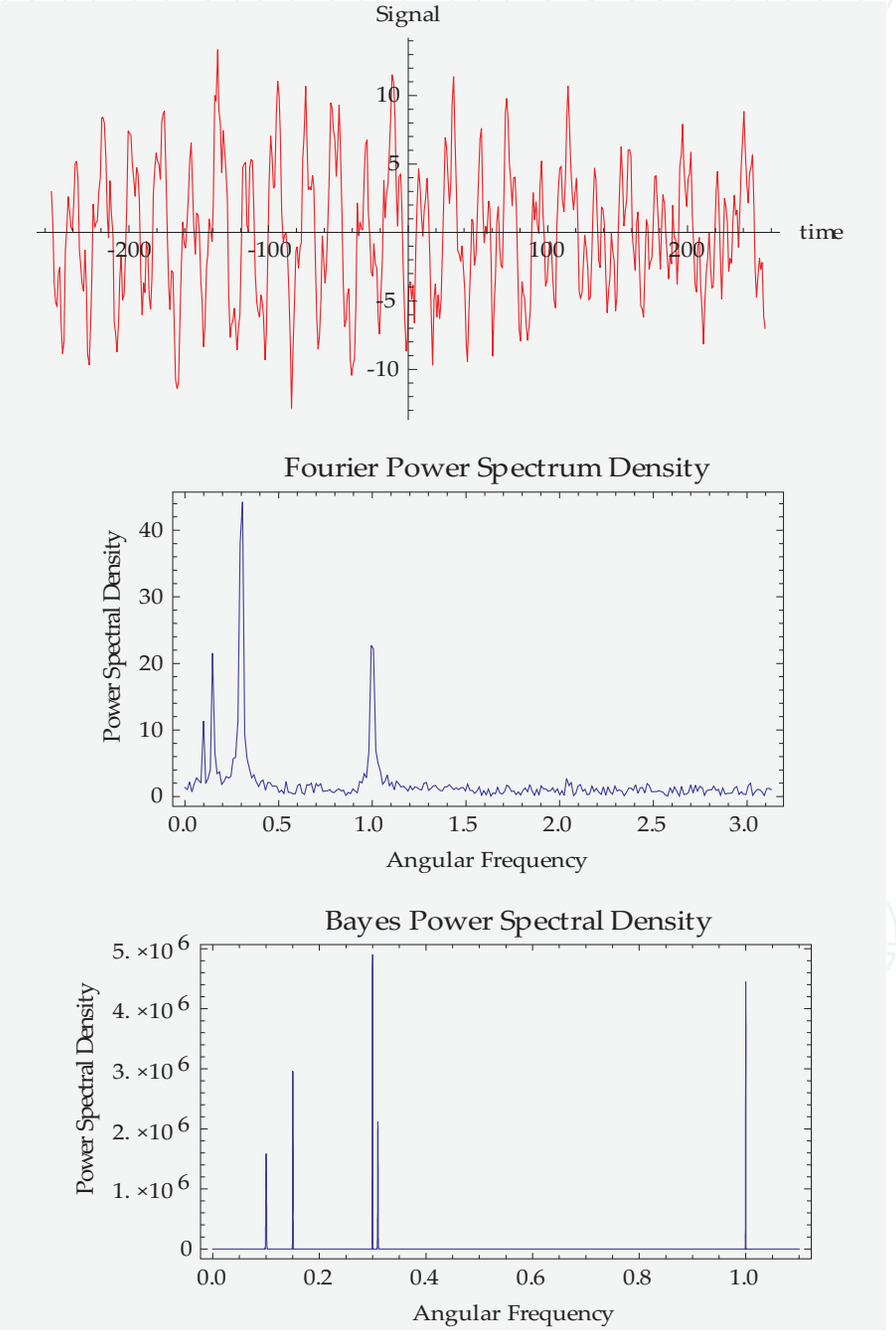


Figure 3. Spectral analysis of multiple frequency signal model

True Frequencies									
ω_1	ω_2	ω_3	ω_4	ω_5	ω_6	ω_7	ω_8	ω_9	ω_{10}
0.36824	0.4109	0.4841	0.4961	0.8146	0.9715	1.0167	1.5978	3.0807	3.1150
True Amplitudes									
B_1	B_2	B_3	B_4	B_5	B_6	B_7	B_8	B_9	B_{10}
-1.838	-2.156	-3.096	2.919	2.765	2.519	-0.185	4.994	-3.862	1.371
B_{11}	B_{12}	B_{13}	B_{14}	B_{15}	B_{16}	B_{17}	B_{18}	B_{19}	B_{20}
4.634	-0.691	-4.947	-4.441	-2.984	-1.409	4.559	1.0755	-3.004	0.206
Estimated Frequencies									
ω_1	ω_2	ω_3	ω_4	ω_5	ω_6	ω_7	ω_8	ω_9	ω_{10}
0.3683 ±0.00002	0.4107 ±0.00003	0.4842 ±0.00002	0.4961 ±0.00004	0.8145 ±0.00003	0.9715 ±0.00002	1.0166 ±0.00002	1.5979 ±0.00002	3.0807 ±0.00002	3.1153 ±0.00007
Estimated Amplitudes									
B_1	B_2	B_3	B_4	B_5	B_6	B_7	B_8	B_9	B_{10}
-1.8644 ±0.06	-2.1858 ±0.06	-3.1385 ±0.06	2.7857 ±0.06	2.7536 ±0.06	2.5492 ±0.06	-0.2409 ±0.06	4.9877 ±0.06	-3.8841 ±0.06	1.2779 ±0.06
B_{11}	B_{12}	B_{13}	B_{14}	B_{15}	B_{16}	B_{17}	B_{18}	B_{19}	B_{20}
4.6103 ±0.06	-0.6783 ±0.06	-5.0206 ±0.06	-4.3572 ±0.06	-3.0533 ±0.06	-1.3511 ±0.06	4.6550 ±0.06	1.1273 ±0.06	-2.9807 ±0.06	0.2479 ±0.06

Table 4. Computer simulations for ten harmonic frequency signal model

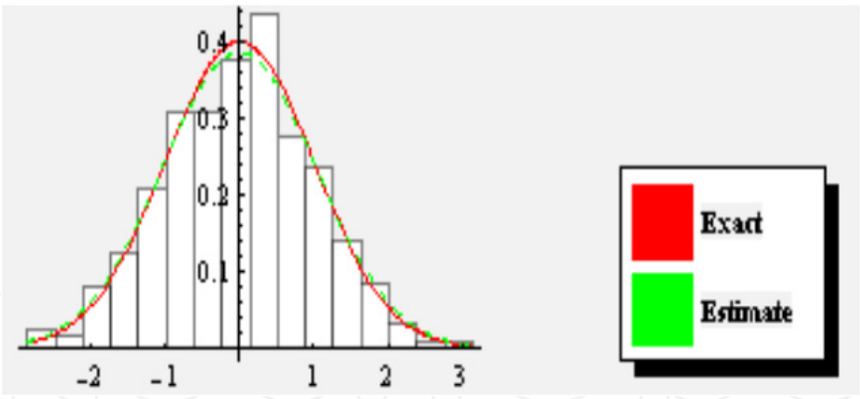


Figure 4. Comparison of exact and estimate probability densities of noise in data $(\sigma = 1)$.

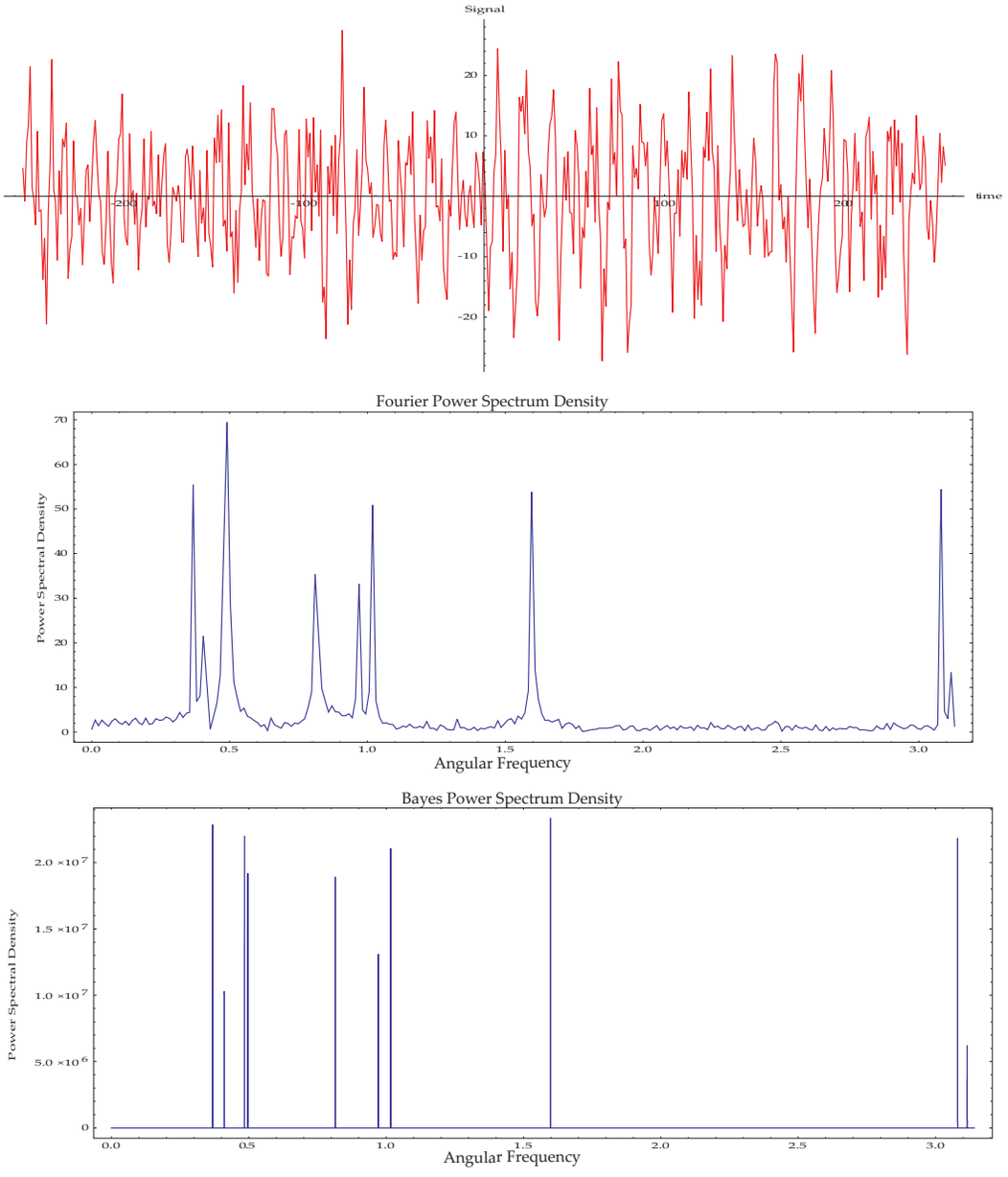


Figure 5. Spectral analysis of ten frequency signal model.

The computational complexity of the algorithm is dependent upon a few parameters such as annealing schedule, data samples and parameters. Under using same annealing schedule, Fig. 7 shows only CPU time of different simulations in a variety of number of data samples. It can be clearly seen that an increase in these numbers causes larger consumption of CPU time. With fixed size of components set and specifically annealing schedule of SA algorithm, the overall execution time of the cooling and decision is almost constant, but the runtime of the first two stages (move and evaluate) mostly depends on complicated design constraints and objective functions. Because the move and the evaluation process in the SA algorithm play an important role in CPU resource usage, improving the calculation ability for these stages will be the most feasible approach for an optimizing SA so that parallel computing is one of best approaches for this goal.

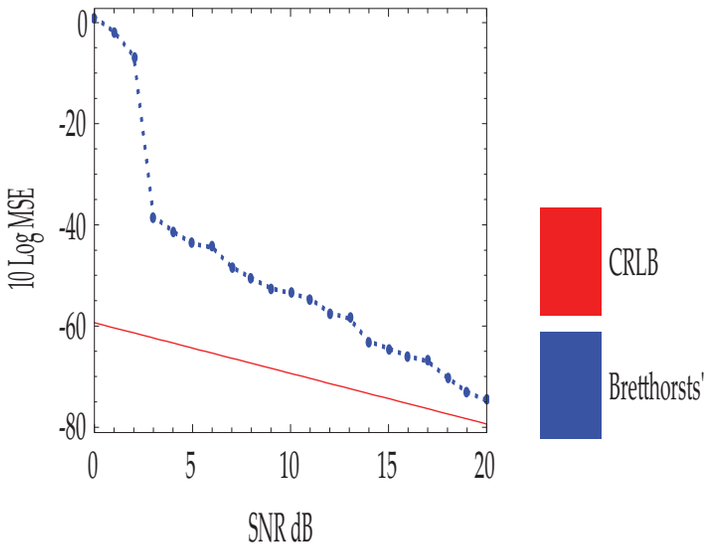


Figure 6. The calculated MSE of the proposed method compared with CRLB versus different SNR with a white noise.

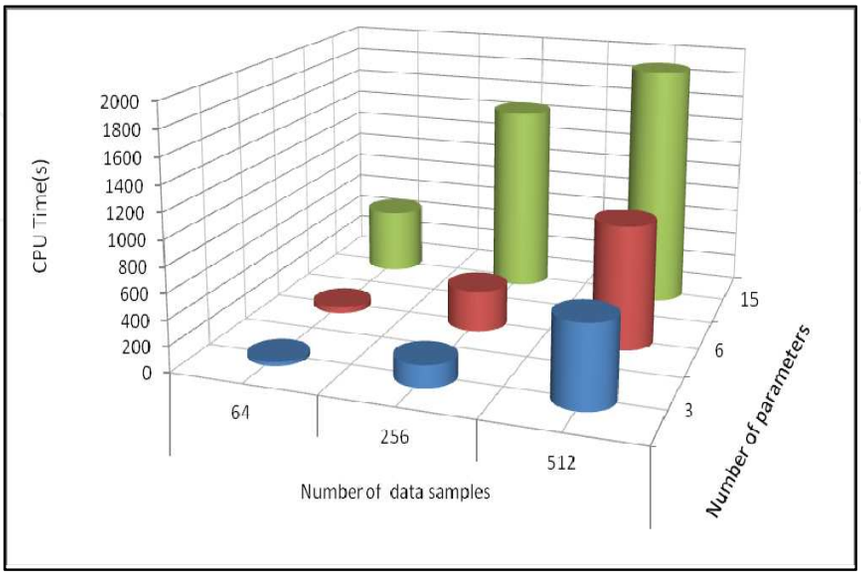


Figure 7. CPU times versus with number of parameters and data samples.

6. Conclusion

In this work we have partially developed a Bayesian approach combined with SA algorithm and applied it to spectral analysis and parameter algorithm for estimating parameters of sinusoids from noisy data. Overall, results presented here show that it provides rational approach for estimating the parameters, namely the frequencies and the 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. A significant advantage of this approach comes from the very large posterior probabilities, which are sharply peaked in the neighborhood of the best fit. This helps us to simplify the problem of choosing starting values for the iteration and it provides a rational approach for estimating, in an optimal way, values of parameters by performing a random search. On the other hand, for sufficiently high SNR, MSEs of frequencies will attain CRLB so that it justifies the accuracy of the frequency estimation. Although the SA algorithm spends large consumption of CPU time, it is competitive when compared to the multiple runs often used with conventional algorithms to test different starting values. As expected, parallel implementation of SA algorithm reduces CPU resource usage.

Data analysis given in this chapter has also been applied to more complicated models and conditions, such as signals decay, periodic but non-harmonic signals, signals with non-stationary, etc.,. In general, we have also not addressed the problem of model selection which is the part of spectral analysis. In this case, one has enough prior information in a given experiment to select the best model among a finite set of model functions so that Bayesian inference helps us to accomplish it. Therefore, it will deserve further investigations.

Nomenclature

$\mathbf{B}$	: Nuisance parameters
$\{B_j, B_{j+\rho}\}$	: Amplitudes of j th sinusoidal
$P(.)$	: Marginal probability density function
$P(.\|.)$	: Conditional probability density function
b_{jk}	: A matrix defined as $-\rho \frac{\partial^2 \overline{\mathbf{h}^2}}{\partial \omega_j \partial \omega_k}\Big\|_{\substack{\omega_j=\hat{\omega}_j \\ \omega_k=\hat{\omega}_k}}$
c_j	: A parameter which controls the step variation along the j th direction
$\mathbf{D}$	: A set of observed data
$e(t)$	: Measurement errors
ε	: A small positive number defined by user
$E[.]$	: Expectation of a function or random variable
$G(.;.)$	: Model functions that contains sinus and cosines terms
$\mathbf{h}$	: Projections of data onto orthonormal model functions

$H(.;.)$: Orthonormal model functions

$\overline{\mathbf{h}^2}$: Mean of square of projections

$\mathbf{I}$: Prior information

$\mathbf{J}(.)$: Fisher information matrix

$\mathbf{J}^{-1}(.)$: Inverse of Fisher information matrix

ρ : Number of frequencies.

N : Number of data samples

$N(0,1)$ A standard Normal distribution

n_j : The number of accepted moves along the direction j .

k : Number of iterations

Q : Least square function

$S(.)$: Bayesian power spectral density

s^2 : Sampling variance

$(t_1, t_2, ..., t_N)^T$: Discrete time set

σ_ω : Standard deviation of angular frequency

σ_B : Standard deviation of amplitude

Ξ_ρ : Solution space of angular frequencies

τ_j : jth Eigenvalue of the matrix b_{jk}

u_{kj} : jth Eigenvector of the matrix b_{jk}

α_Γ : A factor between 0 and 1.

$\boldsymbol{\theta}$: Parameters vector

λ_j : j th component of normalized Eigenvalues of Ω_{jk}

υ_{kj} : j th component of k th normalized Eigenvector of Ω_{jk}

$\mathbf{v}^k$: Step-length vector

$\mathbf{v}^0$: Trial step-length vector

ξ : Uniformly distributed random number from the interval $[-1,1]$

σ^2 : Variance of noise

$\langle \sigma^2 \rangle$: Expected value of noise variance

$\gamma_j^{(k)}$: CRLB for j th component of $\boldsymbol{\omega}^k$

$\boldsymbol{\omega}$: Vector of angular frequency

$\hat{\boldsymbol{\omega}}$: Estimated angular frequency vector

$\boldsymbol{\omega}^0$: An initial guess vector for frequencies

Γ_0 : The initial temperature

$\boldsymbol{\Omega}$: Diagonal matrix

$|.|$: Absolute value

CRLB : Cramér-Rao lower bound
PDF : Probability Density Function
SA : Simulated Annealing
SNR : Signal to Noise Ratio
SQ : Simulated Quenching
MSE : Mean Squared Errors
dB : deci Bell

Author details

Dursun Üstündag
Marmara University, Faculty of Science and Letters, Department of Mathematics, Turkey

Mehmet Cevri
Istanbul University, Faculty of Science, Department of Mathematics, Turkey

Acknowledgement

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 Marmara University, Istanbul, Turkey.

7. References

Aguiare, H., Junior, O., Ingber, L., Petraglia, A., Petraglia, M.R & Machado, M.A.S. (2012). Stochastic Global Optimization and Its Applications with Fuzzy Adaptive Simulated Annealing. Springer Heidelberg New York Dorddrecht London, pp.36-57.

Barbedo, J.G.A. & Lopes, A. (2009). Estimating Frequency, Amplitude and Phase of two Sinusoids with Very Close Frequencies, *International Journal of Signal processing*, Vol.5, No.2, pp. 138-145.

Bernardo, J.M. & Smith, A.F.M. (2000). Bayesian *Theory, Wiley*, New York.

Binder, K. (1986). *Monte Carlo Methods in Statistical Physics*, 2nd Edition, Springer-Verlag.

Bretthorst, G.L. (1988). *Bayesian Spectrum Analysis and Parameter Estimation, Lecture Notes in Statistics,* Springer-Verlag Berlin Heidelberg, New York.

Bretthorst, G. L. (1990). An Introduction to Parameter Estimation Using Bayesian Probability Theory, *Maximum Entropy and Bayesian Methods*, Kluwer Academic Publishers, pp. 53-79.

Brooks S.P. & Morgan B.J.T. (1995). Optimization using simulated annealing, *Journal of Royal Society* (The Statistician) 44, pp. 241-257.

Chan K.W. & So, H. C. (2004). Accurate frequency estimation for real harmonic sinusoids, *IEEE Signal Processing Letters* 11, pp.609-612.

Cooley, J.W. & Tukey, J.W. (1964). An algorithm for the machine calculation of complex Fourier series, Mathematics of Computation *19, pp.* 297-301.

Corana, A., Marchesi, M., Martini, C. and Ridella, S. (1987). Minimizing multimodal functions of continuous variables with the simulated annealing algorithm, *ACM Transactions on Mathematical Software* 13, pp. 262-280.

Edwards, A.W.F. (1972). *Likelihood*, Cambridge University Press.

Gregory, P.C. (2005). Bayesian Logical Data Analysis for the Physical Science, Cambridge University Press.

Goffe, W.L., Ferier, G.D. & Rogers, J. (1994). Global optimization of statistical functions with simulated annealing, Journal of Econometrics *60, pp.* 65-100.

Harney, H.L. (2003). Bayesian Inference: Parameter Estimation and Decisions, Springer-Verlag Berlin Heidelberg.

Hooke, T.R. & Jevees, T.A. (1962). Direct search solution of numerical and statistical problems, *Journal of Association of Computer Machinery* 5, pp. 212-229.

Ingber, L. (1994). Simulated Annealing: Practice versus theory, *Mathematical computer Modeling* 18, pp. 29-57.

Ingber, L. (1996). Adaptive Simulated Annealing ASA: Lessons learned, *Control and Cybernetic 25*, pp. 33-54.

Ireland, J. (2007). Simulated annealing and Bayesian posterior distribution analysis applied to spectral emission line fitting, *Journal of Solar Physics* 243, pp. 237-252.

Jaynes, E. T. (1987). Bayesian Spectrum and Chirp Analysis, in Maximum Entropy and Bayesian Spectral Analysis and Estimation Problems, C. R. Smith and G. J. Erickson (eds.), D. Reidel, Dordrecht, p. 1-37.

Jaynes, E.T. (2003). *Probability Theory: The Logic of Science*, Cambridge University Press.

Jeffreys, H. (1961*). Theory of Probability*, Oxford University Press, London.

Kay, S. M. (1993). *Fundamentals of statistical signal processing: Estimation theory*, Prentice Hall, V.1.

Kirkpatrick, S., Gelatt, C. D., & Vecchi, M.P. (1983). Optimization by Simulated Annealing, *Journal of Science* 220, pp. 671-680.

Laarhoven, P.V. & Aarts, E. (1987). *Simulated annealing: Theory and Applications*, D. Reidel Publishing Company.

Locatelli, M. (2000). Simulated annealing algorithms for Continuous global optimization: Convergence Conditions, *Journal of Optimization Theory and Applications* 106, pp. 121-133.

Lagrange, M., Marchand, S. & Rault, J-Rr. (2005). Improving sinusoidal frequency estimation using a trigonometric approach, *Proceedings of the Digital Audio Effects Conference*, Spain,pp.110-115.

Metropolis, N., Rosenbluth, A., Rosenbluth, M., Teller, A. M. & Teller, E. (1953). Equation of states calculations by fast computing machines, *Journal of Chemical Physics* 21, pp. 1087-1092.

McWhorter T. & Scharf, L. L. (1993). Cramer-Rao Bound for deterministic modal analysis, *IEEE Transactions on Signal Processing* 41, pp. 1847-1866.

Nishiyama, K. (1997). A nonlinear filter for estimating a sinusoidal signal and its parameters in white noise: On the case of a single sinusoid, *IEEE transactions on signal processing* 45, pp. 970-981.

Norton, J.P. (1986). *An introduction to identification,* Academic Press, 1986.

Press, W.H., Flannery, B.P., Teukolshy, S.A. & Vetterling, W.T. (1995). Numerical recipes in C: The art of computing, Cambridge University Press.

Ruanaidh, J.J.K.Q. & Fitzgerald, W.J. (1996). *Numerical Bayesian Methods Applied to Signal Processing*, Springer-Verlag, New York.

Stoica, P. & Nehorai, A.(1989). MUSIC, Maximum Likelihood, and Cramer-Rao Bound, *IEEE Transactions on Acoustic Speech, and Signal Processing* 37, pp.720-741.

Schuster, A. (1905). The Periodogram and its optical analogy, Proceedings of the Royal Society of London, *Series A 77, pp.* 136-140.

Stefankovic, D., Vempala, S. & Vigoda, E. (2009). Adaptive simulated annealing: A near – optimal connection between sampling and counting. *Journal of the ACM* 56, doi:10.1145/1516512.1516520

Tsoulos, G. & Lagaris, I.E. (2006). GenAnneal: Genetically modified Simulated Annealing, *Computer Physics Communications* 174, pp. 846-85.

Üstündag, D., Queen & Bowcock, J.E. (1989). A method for retrieving images from noisy, incomplete data, *Proceeding of the Fifth Alvey Vision Conference*, University of Reading, pp. 239–245.

Üstündag, D., Queen, N.M., Skinner, G.K. & Bowcock, J.E. (1991). Two new methods for retrieving an image from noisy, incomplete data and comparison with the Cambridge MaxEnt package, *10. International Workshop on Maximum Entropy and Bayesian Methods*, MaxEnt 90, pp. 295-301.

Üstündağ, D. & Cevri, M. (2008). Estimating parameters of sinusoids from noisy data using Bayesian inference with simulated annealing, *WSEAS Transactions on Signal Processing* 7, pp. 432–441.

Üstündağ, D. & Cevri, M. (2011). Recovering sinusoids from noisy data using Bayesian inference with simulated annealing, *Mathematical and Computational Applications 16,* pp. 382-391.

Üstündağ, D. (2011). Recovering sinusoids from data using Bayesian inference with RJMCMC, *Proceeding of Seventh International Computation on Natural Computation (ICNN)*, Shanghai, pp. 1850-1854.

Vasan, A. & Raju, K.S. (2009). Comparative analysis of Simulated annealing, Simulated quenching and genetic algorithms for optimal reservoir operation, *Applied Soft Computing* 9, pp.274-281.

Wellin, P., Gayllord, R. & Kamin, S. (2005). An *Introduction to programming with Mathematica*, Cambridge University Press.