

Classical & Bayesian Spectral and Tracking Analysis

Keywords: Fundamental frequency estimation

By

HARRIS K. GONDO

**THESIS FOR THE DEGREE OF THE MSc IN ELECTRICAL &
ELECTRONICS ENGINEERING**

COORDINATOR: SØREN WORRE - BRÜEL & KJÆR A/S

SUPERVISOR: OLE WINTHER - PROFESSOR at DTU

TECHNICAL UNIVERSITY OF DENMARK

DECEMBER 28, 2007

Preface

The thesis has been prepared at Informatics and Mathematical Modeling (IMM), at the Technical University of Denmark in the second semester of 2007. The project has been carried out in collaboration with Brüel and Kjør Sound & Vibration Measurement A/S. It is the thesis for the degree of Civil Engineering in Electrical & Electronic engineering.

This project is the result of my interest for intelligence signal processing which begins at DTU. Spectral analysis and Bayesian parameter estimation form the broad spectrum of the project. However, this is by no means an exhaustive overview of all frequency estimation methods developed. Also I cannot go through every single methods mentioned in the project in great details because of time limit and it is not my goal with writing this thesis. It is rather the study of some relevant and successful techniques defined as those used in autotracking.

The aim of this thesis is to make a survey and investigate the Bayesian probability for fundamental frequency estimation. The emphasis is on classical spectral estimation and Bayesian tracking analysis for both noisy stationary and nonstationary time series. Performance analyses through computer simulations are undertaken to emphasize the most important results in separate points. Illustration of estimates error sensitivity for the sake of estimator comparison and hyperparameter adjustment impact on the estimates is shown and evaluated.

Acknowledgements

I would like express my deepest gratitude to my supervisor, Ole Winther professor from Intelligence Signal Processing-Group in the department of Information Mathematic Modeling-IMM at the Technical University of Denmark, for his guidance through my thesis in collaboration with Brüel & Kjør vibration and sound Measurement A/S. His vast knowledge, patience and valuable advices helped me to accomplish this civil engineering thesis at DTU.

I am thankful to my late father, Gondo Gaston for his moral support and his faithful protection. I am very grateful to my coordinator Søren Worre at Brüel & Kjør and Thorkild Pedersen PhD for their support and helps.

I am please to my Mother Non Monné in Cote D'Ivoire (Ivory Coast) and my lovely son Jesse Frederic Gondo in Denmark for their unconditional support in prayers.

Finally, I would like to thank the greatest God who has given me the energy, the motivation and the health to keep going despite the multiple devastating challenges; and who has been hugely supportive throughout the 2 and 1/2 years it has taken to finish my studies at the Technical University of Denmark.

Abstract

Analysis of rotating machines for design purpose or fault diagnosis requires generally an estimation of parameters that characterizes the vibration and sound patterns. Spectral estimation methods based on classical techniques assume stationarity and high signal-to-noise ratio (SNR). The nonstationarity of vibration and acoustic data is accommodated by the commonly used windowing technique. This thesis explores the Bayesian fundamental frequency estimation theory and investigates both classical and Bayesian approaches to the problem of spectral analysis and slowly varying frequency tracking. We use Periodogram, MUSIC, linear Kalman filter and Bayesian techniques to jointly estimate and track the spectral components. The error sensitivity is shown and the performance for frequency estimation is compared. Such a comparison is based on stationary time series corrupted by additive white Gaussian noise (AWGN). Further, the effect of the prior hyperparameters adjustment is illustrated on the speed profile estimated. The most important results are shown through the experiments in computer simulation. The Bayesian estimator performs well regardless the nature of the signal. Moreover, it provides a reliable and new way of determining the running speed of rotating mechanical system. The marginalization property of the Bayesian can be used to remove DC component (if present) in the data and target the fundamental frequency of interest. That is, the Bayesian method can provide more accurate estimation than stochastic and classical methods when the hyperparameters are adjusted correctly. The reason for such performance status is detailed.

Contents

Preface.....	i
Abstract.....	ii
Introduction.....	vi
Problem formulation.....	vi
Problem analysis.....	vi
Solution strategies.....	vii
Requirement specifications.....	vii
Deadline.....	vii
1. Basic statistics.....	1
1.1 Autocorrelation.....	1
1.2 Fourier Transform.....	1
2. Basic Probability Theory.....	2
2.1 Descriptive statistics.....	2
2.2 Gaussian distribution.....	3
2.2.1 Introduction.....	3
2.2.2 Maximum Likelihood for Gaussian.....	5
2.2.3 Bayesian inference for Gaussian.....	7
2.3 random Walk.....	8
2.4 Conditional Probability.....	8
2.5 Markov Chain.....	9
3. Estimation Methods Pros & Cons.....	10
3.1 Pitch detection algorithms.....	10
3.1.1 Time domain.....	10
3.1.2 Frequency domain.....	10
3.1.3 Summary of frequency estimation algorithms.....	11
3.1.4 Cramer-Rao-Bound.....	12

4. Spectral analysis.....	13
4.1 Classical methods.....	13
4.1.1 Periodogram methodology.....	13
4.1.2 Pisarenko Harmonic Decomposition.....	14
4.1.3 MUSIC.....	15
4.1.4 Linear Kalman filter.....	16
5. Rotating Machines based on vibration and sound analysis.....	20
5.1 Introduction.....	20
5.2 Vibration analysis.....	21
5.2.1 Vibration and sound waveform descriptions.....	21
5.2.2 Spectrogram of the data.....	23
5.2.3 Data model.....	25
5.3 Robust Bayesian tracking analysis.....	28
5.3.1 Modeling the informative prior.....	29
5.3.2 Tracking location parameter.....	31
5.3.3 Procedure of fundamental frequency tracking using informative prior.....	31
6. Result for computer simulations.....	32
6.1 Spectral analysis simulations.....	32
6.1.1 Performance analysis using stationary signal.....	32
• Experiment 1: <i>Single harmonic frequency estimation</i>	32
• Experiment 2: <i>Two harmonic frequency estimation</i>	33
• Experiment 3: <i>Multi-stationary harmonic frequency estimation</i>	35
• Experiment 4: <i>Multiple nonstationary harmonic frequency estimation</i>	37
6.2 Classical and Bayesian estimator noise sensitivity.....	41
6.3 Stationary fundamental frequency tracking.....	44
6.4 Nonstationary frequency tracking.....	51
6.4.1 Bayesian tracking analysis using vibration signal.....	51
6.4.2 Hyperparameter effect.....	54
7. General Conclusion.....	60
Appendix.....	63
A. Review materiel for Bayesian Analysis for linear for regression model....	63
A.1 Bayesian parameter estimation.....	63
A.1.1 Linear model for regression.....	63
A.1.2 Maximum likelihood for regression.....	64
A.1.3 Evidence approximation.....	66
A.1.4 Case study: Inference for Normal mean with known variance.....	69
A.1.5 Vague Prior.....	73
A.1.6 Conjugate Priors.....	74

A.2 Stationary frequency estimation.....	76
A.2.1 Single harmonic frequency estimation.....	77
A.2.2 Model selection.....	82
A.3 Nonstationary frequency tracking.....	88
A.3.1 Likelihood method.....	88
A.3.2 Likelihood procedure.....	90
A.4 Robust Bayesian tracking supplement.....	91
B. Matlab code for robust Bayesian tracking	92
C. Figures	101
D. Reference list.....	104

Introduction

Frequency estimation and tracking is a topic which has been studied in several literatures. It is one of the important areas for application concerning radar, speed estimation of rotational systems among others. Its importance requires a research process that covers qualitative and quantitative approaches to data analysis. Such a study explores specific areas of data analysis that may be applicable and benefit to engineers and companies. Thus the thesis aims to support the development of the critical appraisal skill, thorough considering systematic reviews based on quantitative and qualitative data analysis of existing techniques of frequency estimation and tracking. The benefits of frequency estimation and tracking are many and well known in the medical sector, industries such as Brüel and Kjær Vibration and Sound Measurement A/S. The purpose of the current thesis responds to a new way to determine the running speed of a rotating machine, and can be used to assist the application designer to improve the comfort of automotive products. Frequency estimation entails however, the introduction of a parameter estimation problem in low SNR and nonstationary frequency tracking. The basic problem in frequency estimation is parameter estimation where we assign probabilities to represent what we actually know about the noise uncertainty. As such, we formulate our problem because we know the number of harmonics and constants made of the linear regression model involved in the observations through the spectrogram of the data. In Bayesian probability theory, when these are known the problem is one of parameter estimation. When the harmonic order or the presence of a constant is not known, the problem can refer to model selection. Both problems may be solved using Bayesian theorem and rules of probability theory. However, the parameter estimation and model selection problems have different solutions. In this thesis, we will address the parameter estimation problem through spectral analysis and nonstationary frequency tracking. The framework will be based on classical spectral analysis using synthetic signals plus Gaussian noise for one hand. And in the other hand, Bayesian nonstationary frequency tracking using both vibration and sound data will be investigated. Therefore we will examine the Bayesian technique applied for stationary frequency estimation. In addition, we will compare the Bayesian and the classical performance in a noisy environment to observe the effect of low SNR on the estimates and also the behaviour of the estimator. Further, we will extend such investigation to nonstationary frequency tracking of a real world signal. Moreover we will use both Brüel and Kjær technical signal processing software package called Pulse and Matlab simulation of a Bayesian method based on Thorikild Pedersen's algorithm.

Problem analysis

Learning is a reverse problem of generating a sample from a given model. In our work, we are given a model of the signal with unknown parameters. And then our task is to estimate a fundamental frequency parameter. We formulate the estimation of the fundamental frequency in a Bayesian perspective so that the uncertainty in our model is expressed through the posterior distribution over the parameter of interest. The posterior probability is specified by the likelihood function and the prior distributions. In such a formulation to parameter estimation, the major issue is the choice of the informative prior and the determination of the optimal hyperparameters. In fact, it has been shown in the literature that the incorporation of the prior distribution can yield satisfactory results. However, if the choice of an informative prior can be made more or less for convenience sake, the determination of the optimal hyperparameters associated remains an ill-posed problem.

Solution strategies

In our framework, we will adopt as mentioned above the Bayesian inference as an approach to statistics in which all forms of the unknown fundamental frequency uncertainty may be expressed in term of probability. Although, Bayesian algorithm may show some limit due to process time and high complexity, it offers more flexibility and yield accurate results. Despite the vast field of its study, we will concentrate on parameter estimation and some analyses based on theory and computer simulations to emphasize its performance against noise and its ability to track nonstationary frequency.

In order to achieve our goal, it appears necessary to organize our work in different chapters:

- Chapter 1. We review the basic statistic.
- Chapter 2. Basic probability theory is introduced.
- Chapter 3. Estimation method pros. & cons are tabulated to give an overview of some existing methods performance and comparison.
- Chapter 4. Spectral analysis emphasizes the performance of both classical and Bayesian methods.
- Chapter 5. Rotating machine based on vibration and sound analysis
- Chapter 6. The experiments results for computer simulations are provided.
- Chapter 7. General conclusion
- Further, the appendix follows with the A general survey of the Bayesian analysis for linear regression models to provide us the understanding of the background theory we need to carry out the thesis framework.

Requirement specifications

The experimental vibration, tacho and sound signals were provided by Brüel & Kjær A/S.
Literature: Bayesian analysis of rotating machines by Thorkild Fin Pedersen from IMM bookstore.
Software packages: PULSE Labshop from Bruel & Kjær, student version Matlab from DTU.

Deadline

27 December 2007

Chapter 1

Basic statistics

In this section, we describe the basic analytical and simple nontrivial spectral estimation methods which may be used in this thesis. The explanatory of these basic concepts obviously will help later as the fundament of frequency estimation to understand some advanced related theory we may use.

1.1 Autocorrelation function

It is frequently necessary to be able to quantify the degree of interdependence of one process upon another, or to establish the similarity between one set of data and another. In other words, the correlation between processes or data is sought. The mathematical description of such a tool is as follows

$$R_{xx}(k) = \frac{1}{N} \sum_{n=0}^{N-|k|-1} x(n)x(n+|k|). \quad \text{Eq1}$$

This autocorrelation is a deterministic descriptor of the waveform which may be best modeled by a random sequence. The use of $|k|$ in Eq1 makes $R_{xx}(\cdot)$ symmetric about $k = 0$.

1.2 Fourier Transform

The Discrete Fourier Transform (DFT) is a Fourier series representation where Fourier coefficients are the samples of the sequence. In other words, it provides the description of the signal $x(n)$ in the frequency domain, in the sense that $X(k)$ represents the amplitude and phase associated with the frequency component as defined bellow:

$$X(k) = \frac{1}{N} \sum_{n=0}^{N-1} x(n) e^{-j2\pi kn/N}. \quad \text{Eq2}$$

Chapter 2

Basic Probability Theory

In practice, data often contain some randomness or uncertainty. Statistics handle such data using methods of probability theory which concern the analysis of random phenomena. Before we make any informed decision, we use analysis methods based on the following.

2.1 Descriptive statistics

This forms the quantitative analysis of the data. We will use these to describe basic features of the data in study. Generally they provide summary of the data. In this project, we will use the following:

- *Mean as a measure of location*

$$\mu_x = \frac{1}{N} \sum_{n=1}^N x(n) \quad \text{Eq3}$$

- *Variance as a measure of the statistical variability*

$$\sigma_x^2 = \frac{1}{N} \sum_n (x(n) - \mu_x)^2 \quad \text{Eq4}$$

- *Skewness is a measure of asymmetry. It concerns the shape of the distribution. The coefficient of skewness may be positive (right tail), negative (left tail) or zero (symmetric).*

$$\text{skew}(x(n)) = \frac{\left(\frac{1}{N} \sum_{n=1}^N [x(n) - \mu]\right)^3}{\sigma^3} \quad \text{Eq4}$$

- Kurtosis¹ is a measure of the peakedness (sharpness of the spike) of a unimodal probability density function (pdf).

$$\text{kur}(x(n)) = \frac{\left(\frac{1}{N} \sum_{n=1}^N [x(n) - \mu]\right)^4}{\sigma^4} \quad \text{Eq5}$$

¹ See page 157 – Ledermann handbook of Applicable Mathematics – Volume 2- Probability – Emlyn Lloyd, 1980

2.2 Gaussian distribution

Probability theory provides a consistent framework for the quantification and manipulation of uncertainty. It forms one of the important keys in pattern recognition. Therefore it appears necessary to explore a specific probability distribution model and its properties. The popular Gaussian distribution will provide us the opportunity to discuss some statistical key concepts, such as mean, variance and Bayesian inference in the context of simple model before the proposed robust model. One role for the distribution is to model the probability distribution $p(x)$ of the random variable x from a given finite set x_1, \dots, x_N of observations. This problem is known as density estimation. For that purpose, we shall assume that the data points are all independent and identically distributed (iid). It should be emphasized that the problem of density estimation is fundamentally ill-posed, because there are infinitely many probability distributions that could have given rise to the observed data. Indeed any distribution that is nonzero can be a potential candidate. The issue of choosing a suitable model is related to the problem of model selection which is one of the central issues in pattern recognition. We will focus here on the Gaussian distribution for a simple mathematical tractability.

2.2.1 Introduction

We introduce one of the most important probability distribution for continuous variables called also normal distribution. For the case of single real-valued variable x , the Gaussian distribution is defined by

$$N(x | \mu, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{1/2}} \exp\left\{-\frac{1}{2\sigma^2}(x - \mu)^2\right\} \quad \text{Eq6}$$

which is governed by two parameters: μ called mean and σ^2 called the variance. The square root of the variance is called standard deviation $\sqrt{\sigma^2}$ and the reciprocal of the variance, written as $\beta = 1/\sigma^2$, is called precision. Figure 1 shows the plot of the Gaussian distribution.

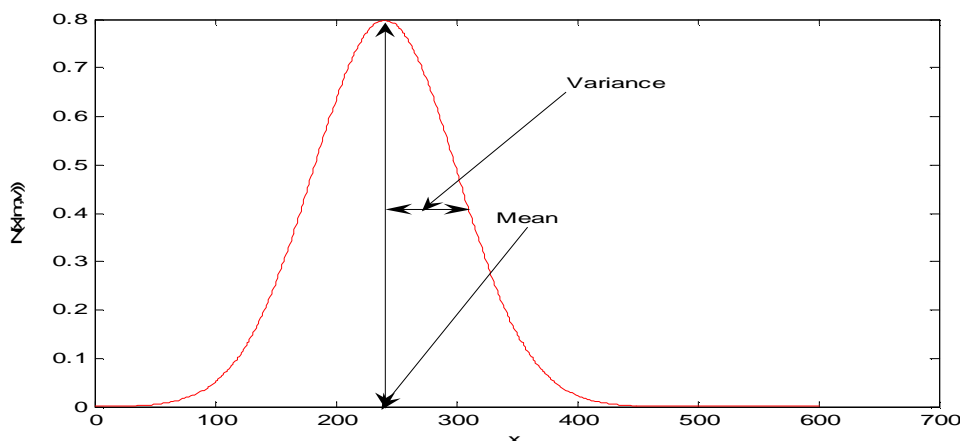


Figure 1: plot of univariate Gaussian showing the mean and the standard deviation.

This is a common used model due to its simple property. The Gaussian distribution can arise when we consider the sum of multiple random variables. The central limit theorem (due to Laplace) tells us that under certain condition, the sum of a set of random variable, which is also random variable, has a distribution that becomes increasingly Gaussian as number in term increases (Walker 1969). The Figure 2 shows the illustration of the central limit theorem.

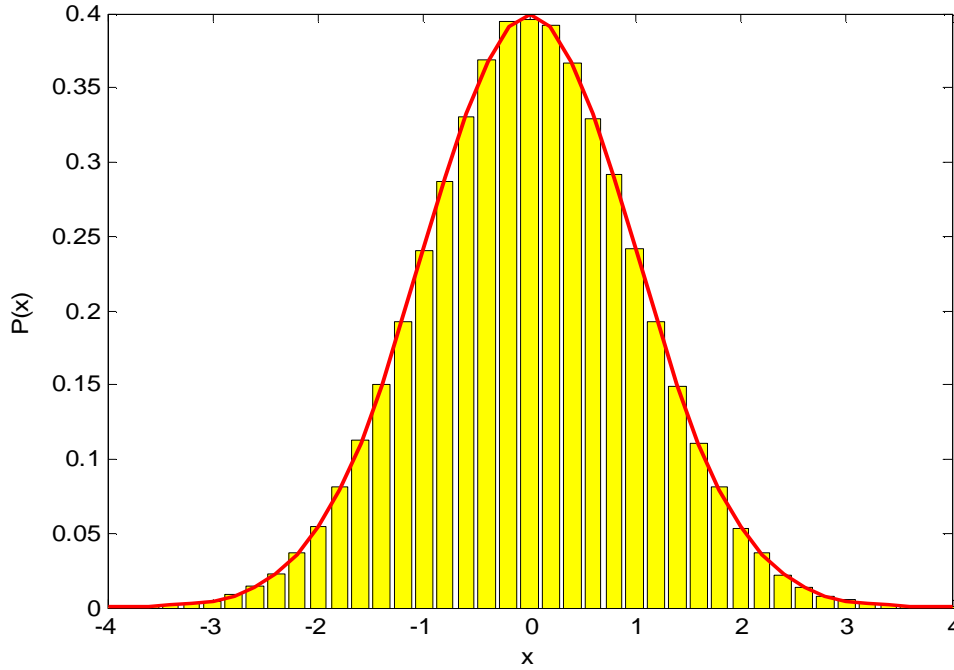


Figure 2: Central limit theorem simulated by an histogram forming a Gaussian distribution.

From Eq6, we see $N(x | \mu, \sigma^2) > 0$ and it is straightforward to show that the Gaussian is normalized, so that $\int_{-\infty}^{\infty} N(x | \mu, \sigma^2) dx = 1$.

Thus Eq6 satisfies the two requirements for a valid probability density. We can then find the expectations of function of x under the Gaussian distribution. The maximum of the Gaussian distribution is called mode, and it coincides with the mean. We are also interested in multivariate Gaussian distribution defined over D -dimensional vector of x of continuous variables, which is given by

$$N(\mathbf{X} | \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{D/2}} \frac{1}{|\boldsymbol{\Sigma}|^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{X} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{X} - \boldsymbol{\mu})\right\} \quad \text{Eq7}$$

Where the D -dimensional vector $\boldsymbol{\mu}$ is the mean, the $D \times D$ matrix $\boldsymbol{\Sigma}$ is the covariance, and $|\boldsymbol{\Sigma}|$ is the determinant of $\boldsymbol{\Sigma}$.

2.2.2 Maximum Likelihood for Gaussian

- *Univariate case:*

One common criterion of determining μ and σ^2 in such a distribution using an observed data set is to find the parameter values that maximize the likelihood function. In practice, it is more convenient to maximize the log of the likelihood function. Because the logarithm is a monotonically increasing function, maximizing the log of the function is equivalent to maximizing the function itself. Taking the log not only simplifies the subsequent mathematical analysis, but also avoids underflow of the numerical precision of the computer by using sum of log probabilities. From Eq6 and Eq7, the log likelihood is written in the form

$$\ln p(X | \mu, \sigma^2) = -\frac{1}{2\sigma^2} \sum_{i=1}^N (x_i - \mu)^2 - \frac{N}{2} \ln \sigma^2 - \frac{N}{2} \ln(2\pi) \quad \text{Eq8}$$

In practice, it is more convenient to consider the negative log of Eq7 to find the minimum of error sum which is equivalent to maximizing the likelihood since the negative log is a monotonically decreasing function. However, for the special case of the univariate normal density, we can find the maximum likelihood solution by analytic differentiation of Eq8 (same procedure applies for the multivariate case). We then obtain the maximum likelihood solution given by

$$\hat{\mu}_{ML} = \mu_x. \quad \text{Eq9}$$

This is a sample mean, i.e. the mean of the observed values. Similarly differentiating eq8 with respect to with regard to (wrt) σ^2 , we obtain the maximum likelihood solution for the variance in the form

$$\sigma_{ML}^2 = \sigma_x^2 \quad \text{Eq10}$$

which is the sample variance measure wrt the sample mean. In fact, it appears at this stage necessary to point out that the maximum likelihood approach underestimates the true variance of the distribution by factor $(N-1)/N$ and yields the correct mean value as follows (Pattern Recognition and Machine Learning – C. M. Bishop 2006).

$$E[\sigma_{ML}^2] = \left(\frac{N-1}{N}\right) \sigma_x^2. \quad \text{Eq11}$$

From Eq8 it follows that the following estimate for the variance parameter is unbiased

$$\hat{\sigma}^2 = \frac{N}{N-1} \sigma_{ML}^2 = \frac{1}{N-1} \sum_{n=1}^N (x_n - \mu_{ML})^2 \quad \text{Eq12}$$

We note that the bias problem due to the underestimation of the true variance becomes less significant as the number of N of data points increases. When N approaches infinite, the maximum likelihood solution for the variance equals the true variance of the distribution that generates the data. In the multivariate case, the maximum likelihood for the Gaussian yields the following parameter estimates:

- *Multivariate case:*

Let's consider a data set $\mathbf{X} = \{x_1, \dots, x_N\}^T$ in which the observations are also assumed to be drawn independently from a multivariate Gaussian distribution. We can estimate again the parameters of the distribution by maximum likelihood. The log likelihood function is given by

$$\ln p(\mathbf{X} | \boldsymbol{\mu}, \boldsymbol{\Sigma}) = -\frac{ND}{2} \ln(2\pi) - \frac{N}{2} \ln |\boldsymbol{\Sigma}| - \frac{1}{2} \sum_{n=1}^N (x_n - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (x_n - \boldsymbol{\mu}). \quad \text{Eq13}$$

Using the derivative of the log likelihood wrt $\boldsymbol{\mu}$ is given by

$$\frac{\partial}{\partial \boldsymbol{\mu}} \ln p(\mathbf{X} | \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \sum_{n=1}^N \boldsymbol{\Sigma}^{-1} (x_n - \boldsymbol{\mu}). \quad \text{Eq14}$$

And setting this derivative to zero, we obtain the solution for the maximum likelihood estimate of the mean. The maximization of the Eq13 wrt $\boldsymbol{\Sigma}$ is rather more involved. After some manipulations, the result is as expected and takes the form

$$\boldsymbol{\Sigma}_{ML} = \frac{1}{N} \sum_{n=1}^N (x_n - \boldsymbol{\mu}_{ML})(x_n - \boldsymbol{\mu}_{ML})^T \quad \text{Eq15}$$

But this is less than the true value. Hence it is biased. We can correct this biased by

$$\boldsymbol{\Sigma}_{ML} = \frac{1}{N-1} \sum_{n=1}^N (x_n - \boldsymbol{\mu}_{ML})(x_n - \boldsymbol{\mu}_{ML})^T \quad \text{Eq16}$$

NB: *we must note that all the $\boldsymbol{\mu} = \boldsymbol{\mu}_x$.*

2.2.3 Bayesian Inference for Gaussian

The maximum likelihood gives the point estimates for the parameter μ and Σ . Now we develop the Bayesian analysis by introducing the prior distributions over the parameters. We start by a simple example in which we suppose that the variance σ^2 is known and consider the task of inferring the mean μ given the data set of N observations. The likelihood function that is the probability of observed data given the mean is defined by

$$p(\mathbf{X} | \mu) = \prod_{n=1}^N p(x_n | \mu) = \frac{1}{(2\pi\sigma^2)^{N/2}} \exp\left\{-\frac{1}{2\sigma^2} \sum_{n=1}^N (x_n - \mu)^2\right\} \quad \text{Eq17}$$

We take the prior which has the same probability distribution over the mean parameter as the likelihood function to yield a posterior probability with the same Gaussian distribution. Hence conjugacy is obtained.

$$p(\mu) = N(\mu | \mu_0, \sigma_0^2) \quad \text{Eq18}$$

and the posterior probability distribution is given by

$$p(\mu | \mathbf{X}) \propto p(\mathbf{X} | \mu) p(\mu) \quad \text{Eq19}$$

After some manipulation involving completing the square in the exponent the posterior distribution is given by

$$p(\mu | \mathbf{X}) = N(\mu | \mu_N, \sigma_N^2) \quad \text{Eq20}$$

where

$$\mu_N = \frac{\sigma^2}{N\sigma_0^2 + \sigma^2} \mu_0 + \frac{N\sigma_0^2}{N\sigma_0^2 + \sigma^2} \mu_{ML} \quad \text{Eq21}$$

$$\frac{1}{\sigma_N^2} = \frac{1}{\sigma_0^2} + \frac{N}{\sigma^2} \quad \text{Eq22}$$

$$\sigma_N^2 = \frac{\sigma^2}{N\sigma_0^2 + \sigma^2} \sigma_0^2 \quad \text{Eq23}$$

It is worth to study to mean and the variance of posterior probability which are given by the compromise between the prior and the likelihood. We can notice that if the number of observed data points is zero, the posterior mean is equal the prior mean. For an infinitely number of N , the mean of the posterior distribution is given by the maximum likelihood solution. When we consider the variance, we see that there are expressed in terms of inverse variance, which is called precision. Furthermore, the precisions are additive, so that the precision of the posterior is given by the precision of the prior and one contribution of the data precision from each of the observed data points. As we increase the number of data points, the precision increases. If N is infinitely large, the posterior variance goes to zero and the posterior distribution becomes infinitely peaked around the maximum likelihood solution.

2.3 Random walk

A random process is the stochastic process. It is also a probabilistic description of a system developing or changing in time or space. Here we represent such a process by a point which moves at each trial either one 1, 2,...steps upward (with probability p_1, p_2, \dots) or 1, 2 steps downwards (with probabilities q_1, q_2, \dots). The unrestricted simple random walk process S_r is defined as follows:

$$S_{r+1} = S_r + X_{r+1} \quad \text{Eq24}$$

where $r = 0, 1, \dots, S_0 = k$ (a given constant) and X_r are mutually independent random variable with a distribution given by $P(X_r = 1) = p$, $P(X_r = -1) = 1 - p = q$. We are not going any further, because we will be using only this basic property as our ground to build the proposed tracking prior later in this project.

2.4 Conditional probability

In the deterministic world model which is adequate for greater part of the elementary science and technology, phenomena are either independent of one another, or completely determined one by another.

The Rules of probability theory

There only two basic rules for manipulating probabilities, the product and the sum rule; all other rule may be derived from them. If A , B and C stand for three arbitrary propositions then

$$P(A | B) = \frac{P(A \text{ and } B)}{P(B)} \quad \text{Eq25}$$

If A and B are independent $P(A \text{ and } B) = P(A)P(B)$

Thus Eq25 becomes $P(A | B) = P(A)$ and $P(B | A) = P(B)$

Sum rule $P(A, B) = P(A | B)P(B)$

Product rule $P(A) = \sum_Y P(A, B)$

According to Aristotelian logic, the proposition “ A and B ” is the same as “ B and A ” so the truth value of the propositions must be the same in the product rule. That is the probability of “ A and B given C ” must be equal the probability of B and A given C ”, this can be defined by

$$P(B, A | C) = P(B | C)P(A, B | C) \quad \text{Eq26}$$

Likewise

$$P(A, B | C) = P(A | C)P(B, A | C) \quad \text{Eq27}$$

These equations may be combined to obtain the following result

$$P(A | B, C) = \frac{P(A | C)P(B | A, C)}{P(B | C)} \quad \text{Eq28}$$

This is also Bayes theorem. It is named after Reverend Thomas Bayes, an 18th century mathematician who derived a special case of the theorem. This is a starting point of the all Bayesian calculations.

$$P(A | C) = \int dB P(A, B | C) \quad \text{Eq29}$$

This is a form that the sum rule uses to remove uninteresting or nuisance parameters (B in this example).

2.5 Markov chain

We now consider a more complex problem involving the chain or state of evolution of the frequency. From rotational system, the signal is the sum of harmonic related signal. The change of one will always affect the other. To express such effects in probabilistic model, we need to relax the independent identical distributed (iid) assumption of the observation. And then consider the Markov model to design the slowly change of the frequencies. This concept will lay the foundation for the tracking process of the slowly varying harmonically related frequency in nonstationary environment. First order Markov chain is defined to be a series of random variables $w^{(1)}, w^{(2)}, \dots, w^{(N)}$ such that the following conditional independence property holds for $m \in \{1, 2, \dots, N\}$

$$P(w^{(m)} | w^{(1)}, \dots, w^{(m-1)}) = P(w^{(m)} | w^{(m-1)}) \quad \text{Eq30}$$

Under this model the joint distribution of a sequence of m observation is given by

$$P(w^{(1)}, \dots, w^{(m)}) = P(w^{(1)}) \prod_{n=2}^m P(w^{(n)} | w^{(n-1)}) \quad \text{Eq31}$$

This model can be used to model distribution of the slowly changing frequency which is characterized by high correlation. We will see the full description later.

Chapter 3

Estimation Methods Pros. & Cons.

3.1 Pitch detection algorithms

There are two categories of pitch detection algorithms: time domain and frequency domain. In this section, we give the Pros & cons in both time and frequency domains, and then the summary of the frequency estimators will follow.

3.1.1 Time domain

- *Autocorrelation*
Pros. Relative impervious to noise.
Cons. Sensitive to sampling rate results in low resolution, expensive computation.
- *Zero crossings*
Pros. Simple, inexpensive
Cons. Inaccurate, poor with noisy signals or harmonics signals.
- *Maximum likelihood*
Pros. Accuracy is high.
Cons. Complex

3.1.2 Frequency domain

- *Harmonic Product Spectrum (HPS)*
Pros. Computationally inexpensive, reasonably resistant to noise, inputs dependent.
Cons. Low pitch may be tracked less accurately than high pitches.
- *Discrete Fourier transform (DFT)*
Pros. Powerful analysis tool for stationary and periodic signals.
Cons. Inefficient to noise

This section compares four pitch detection algorithms in real time pitch-detection application. The four algorithms are HPS, DFT, Maximum Likelihood and weighted Autocorrelation. They mentioned the issues on the discontinuity of the result, which depends on the frame size of the detection windows. An interesting point raised is the sensitivity of the algorithm to the type of input signal. However, it does not contain much about the real time issues, such as the window size and sampling frequency.

3.1.3 Summary of frequency estimation algorithms

In this section, we will give a brief summary and some results the frequency estimation algorithms have achieved. For that purpose, we shall categorize frequency estimation algorithms as follows:

- Block estimators, where the frequency estimate is obtained for fixed sample size T in $O(T \log T)$ or more floating point operations.
- Fast block estimators, where the sample size again is fixed, but the number of operations required is $O(T)$.
- On-line estimators, which allow recursively updated frequency estimates to be generated.

These last class of estimators is of particular interest, because they may be more amenable to extension to the frequency tracking problem than the block processing methods. The block processing methods may only be used for tracking when it is known that the instantaneous frequency does not change significantly over known time.

- *Block estimators:*
It has been found in the literature [1] that the most attractive of these estimators appears to be the estimator of Quin and Fernandes [1991], for several reasons. The estimator is unbiased, asymptotically efficient, requires fewer operations than the full maximum likelihood and is more robust to initial conditions than that algorithm.
- *Fast block estimators:*
Of the weights phase averaging estimators, that proposed by Lovell and Williamson [1992] has the best performance. The Kay [1989] estimator has similar performance for small noise levels, but its bias in the presence of unbounded, in particular Gaussian, noise is a problem.
- *On-Line estimators*
Because of the frequency tracking problem, the interest increases around on-line estimators. The Hannan-Huang estimator has been so modified (Hannan-Huang [1993]) and Nehorai and Porat frequency estimator only requires a suitable choice of system dynamics to be used as a frequency tracker.

From the table 1, we have found that only four estimators namely: Maximum Likelihood (ML), periodogram maximizer, Fernandes-Goodwin-de Souza and Quin-Fernandes achieves Cramer Rao Bound.

3.1.4 Cramer-Rao-Bound

The Cramer-Rao-Bound on the variance of an unbiased estimator of the frequency, $\hat{\omega}_0$ of a signal tone in noise is

$$\text{var}(\hat{\omega}_0) \geq \frac{12\sigma^2}{N(N^2 - 1)B^2} \quad \text{Eq32}$$

For the multi-harmonic frequency estimation problem, Barrett and McMahon [1987] have derived the analogous bound, which is

$$\text{var}(\hat{\omega}_0) \geq \frac{12\sigma^2}{N(N^2 - 1) \sum_{k=1}^p k^2 B_k^2} \quad \text{Eq33}$$

where σ^2 is the variance, N is the sample size and B is the amplitude of the signal.

Frequency estimators Summary			
Paradigm	Algorithms	Complexity	AACRB
ML	ML	$> O(T \log T)$	Yes
Approximate ML	Periodogram maximiser	$> O(T \log T)$	Yes
	DF-Periodogram M^2	$O(T \log T)$	No
Fourier coefficient	FTI 1	$O(T \log T)$	No
Fourier coefficient	FTI2	$O(T \log T)$	No
	GPIE	$O(T \log T)$	No
Signal	Minimum Variance	$O(T^3)$	No
Subspace	Barlett	$O(T^3)$	No
Noise	Pisarenko	$O(T)$	No
Subspace	MUSIC	$O(T^3)$	No
Phase	Lank-Reed-Pollon	$O(T)$	No
Weighted	Kay	$O(T)$	No
Averaging	Lovell	$O(T)$	Yes*
	Clarkson	$O(T)$	No
Filtering	Fernandes-Goodwin-de-Souza	$O(T)$	Yes
	Quin-Fernandes	$O(T)$	Yes
	Hannan-Huang	N/A	N/A
	Nehorai-Porat	N/A	N/A

Table 1: Summary of frequency estimators.

² M = Maximizer. * Further investigation of the asymptotic performance of this algorithm is need. N/A: not applicable to online estimators. Asymptotically Achieves Cramer-Rao-Bound (AACRB) ?.

Chapter 4

Spectral Analysis

4.1 Classical methods

This section deals with transformation of data from time domain to frequency domain. Spectral analysis is thus applied when the frequency property of the phenomena is investigated and when the time contains periodicities.

4.1.1 Periodogram methodology

We consider an important class of signal characterized as stationary random process.

$$C(\omega) = \frac{1}{N} \left| \sum_{i=1}^N d_i e^{-j\omega t_i} \right|^2 \quad \text{Eq34}$$

This is the so called periodogram. It was originally introduced by Schuster (1898) to detect and measure “hidden periodicities” in data. The problem of the periodogram is that the variance of the estimate $C(\omega)$ does not decay to zero as $N \rightarrow \infty$. That is, it does not converge to the true power density spectrum. This inconsistency can be seen when the estimates fluctuate more and more wildly from realization to realization. However the periodogram has the advantage of possible implementation using fast Fourier transform (FFT), but with the disadvantage in the case of short data lengths of limit frequency resolution. The deleterious effects of spectral leakage and smearing may be minimized by windowing the data by a suitable window function. It has been shown that averaging a number of N realizations can significantly improve the estimate of the spectrum accuracy. The accuracy of the spectra may be obtained in term of variance. The smaller variance of the power spectral density yields more accurate estimate. We can also decrease the variance by first dividing the whole data into k equal length section followed by zero padding, and smooth the estimate spectrum to removing randomness. The DFT consists of harmonic amplitude and phase components regularly spaced in frequency. The spacing of the spectral lines decreases with length of the sampled waveform. If a signal component falls between two adjacent harmonic frequency spectra then it cannot be properly represented (See computer simulations results). Its energy will be shared between neighbouring harmonics and the nearby spectral amplitude will be distorted. Windowing is very relevant also to smooth the estimate spectral component. However, when data is windowed the zero ends point can represent loss of information. To avoid such loss, we need to partition the data into overlap section, say 50% to 75% to include most of the feature. The resulting spectra are then averaged to obtain an estimate of the true spectrum.

4.1.2 Pisarenko Harmonic Decomposition

In this section, we will discuss the Pisarenko and introduce the MUSIC techniques of frequency estimation based on phase and autocovariance. We won't spend much time in Pisarenko for the reason that its asymptotic variance is proportional to T^{-1} . Any estimator which may be expressed as a nonlinear function of these sample autocovariances will inherit an asymptotic variance of the same order. The periodogram and the MLE approaches, on the other hand, yield estimators with asymptotic variances of order T^{-3} . It has been shown that the Pisarenko's technique which uses eigenvectors of autocovariance matrix, is consistent when the noise is white, but produces estimators which have variances of a higher order than those of the MLE.

Observation model is given by:

$$y(n) = x(n) + w(n) \quad \text{Eq35}$$

where $x(n) = \sum_{i=1}^p A_i e^{j(2\pi f_i n + \phi)}$, $w(n)$ is an additive white Gaussian noise with zero mean and variance

σ_w^2 ; and $Y' = [y(n), y(n-1), \dots, y(n-p)]$ is the observed data vector of dimension $(p+1)$ and

$W' = [w(n), w(n-1), \dots, w(n-p)]$ is the noise vector.

The autocorrelation function $\gamma_{yy}(m)$ is

$$\gamma_{yy}(m) = \gamma_{xx}(m) + \sigma_w^2 \delta(m) \quad m = 0, \pm 1, \dots, \pm(M-1) \quad \text{Eq36}$$

Hence the $M \times M$ autocorrelation matrix for $y(n)$ can be expressed as

$$\Gamma_{yy} = \Gamma_{xx} + \sigma_w^2 \mathbf{I} \quad \text{Eq37}$$

where Γ_{xx} is the autocorrelation matrix for the signal $x(n)$ and $\sigma_w^2 \mathbf{I}$ is the autocorrelation of the matrix of the noise. In fact, the signal matrix can be expressed as

$$\Gamma_{xx} = \sum_{i=1}^p P_i \mathbf{s}_i \mathbf{s}_i^H \quad \text{Eq38}$$

where $P_i = A_i^2$ the power of the i th sinusoid, H denotes the conjugate transpose and \mathbf{s}_i is a signal vector of dimension M defined as

$$\mathbf{s}_i = [1, e^{j2\pi f_i}, e^{j4\pi f_i}, \dots, e^{j2\pi(M-1)f_i}]^T \quad \text{Eq39}$$

Let us perform an eigen-decomposition of the matrix Γ_{yy} . Let the eigenvalues $\{\lambda_i\}$ be ordered in decreasing value with $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \dots \geq \lambda_M$ and let the corresponding eigenvectors be denoted as $\{\mathbf{v}_i, i = 1, 2, \dots, M\}$. We assume that the eigenvectors are normalized so that $\mathbf{v}_i^H \cdot \mathbf{v}_j = \delta_{ij}$.

The signal correlation matrix is

$$\Gamma_{xx} = \sum_{i=1}^p \lambda_i \mathbf{v}_i \mathbf{v}_i^H \quad \text{Eq40}$$

In the presence of the noise, the autocorrelation matrix can be represented by

$$\sigma_w^2 \mathbf{I} = \sigma_w^2 \sum_{i=1}^M \mathbf{v}_i \mathbf{v}_i^H \quad \text{Eq41}$$

After substitution of some of the equations above, we obtain

$$\Gamma_{yy} = \sum_{i=1}^p \lambda_i \mathbf{v}_i \mathbf{v}_i^H + \sum_{i=1}^M \sigma_w^2 \mathbf{v}_i \mathbf{v}_i^H = \sum_{i=1}^p (\lambda_i + \sigma_w^2) \mathbf{v}_i \mathbf{v}_i^H + \sum_{i=p+1}^M \sigma_w^2 \mathbf{v}_i \mathbf{v}_i^H \quad \text{Eq42}$$

This eigen-decomposition separates the eigenvectors in two sets. The set $\{\mathbf{v}_i, i = 1, 2, \dots, p\}$

which is principal eigenvector, span the signal subspace, while the set $\{\mathbf{v}_i, i = p+1, \dots, M\}$

which is orthogonal to the principal vector, are said to belong to the noise space. In this context we see that the Pisarenko method is based on an estimation of the frequencies by using the orthogonality property between the signal vector and vector in the noise subspace. The frequencies can be determined by solving for the zeros of the polynomial

$$V(z) = \sum_{k=0}^p v_{p+1}(k+1) z^{-k} \quad \text{Eq43}$$

all of which lie on the unit circle. The angles of the roots are $2\pi f_i, i = 1, 2, \dots, p$. When the number of sinusoids is unknown, the determination of p is difficult, especially if the signal level is not much higher than the noise level. The location of the peaks in the frequency estimation function is defined by

$$P_P(\hat{f}) = \frac{1}{\left| \mathbf{s}_i^H(\hat{f}) \mathbf{v}_{p+1} \right|^2} \quad \text{Eq44}$$

4.1.3 Multiple Signal Classification (MUSIC)

This method is also a noise subspace frequency estimator. The estimates of the sinusoidal frequencies are the peaks of the $P_M(f)$

$$P_M(\hat{f}) = \frac{1}{\sum_{k=p+1}^M \left| \mathbf{s}^H(\hat{f}) \mathbf{v}_k \right|^2} \quad \text{Eq45}$$

For further details see digital signal processing, principles, algorithms and application, John G. Proakis, Dimitris G. Monalakis (3edition – 1996-page 948) and The estimation and tracking of frequency (2001) – B.G. Quin & E.J. Hannan (p.143-179). The asymptotic variance is in the $O(T^{-1})$.

4.1.4 Linear Kalman Filter

Since 1960, Kalman filtering has been the subject of extensive research and application [1], particularly in the area as diverse as aerospace, demographic modelling, manufacturing, radio communication and other. The Kalman filter is an efficient recursive filter that estimates the states of a dynamic system from a series of incomplete and noisy measurements. Further it provides computational means to estimate the state of the process, in a way it minimizes the mean of the square error. It is applied when the dynamic and the observation equation are linear with additive Gaussian noise. The Kalman filters are based on linear dynamic system discretised in the time domain. They are modelled on Markov chain built on linear operator and perturbed by Gaussian noise. In order to use the Kalman filter to estimate the internal state of a process given a sequence of noisy observations, the process must be modelled by specifying the matrices A , H , Q , R and sometimes B for each time-step k as described below.

- *Kalman filter Model*

The Kalman filter model addresses the general problem of trying to estimate the state x_k of a discrete time controlled process that is governed by the linear stochastic difference equation

$$x_k = Ax_{k-1} + Bu_{k-1} + w_{k-1} \quad \text{Eq46}$$

Where

A is the state transition model which is applied with the previous state x_{k-1}

B is the control input model which is applied to the control vector u_k

w_k is the process noise assumed to be drawn from Gaussian zero mean multivariate normal distribution.

$$p(w) \sim N(0, Q)$$

- *The observation (measurement) model*

At time k an observation (or measurement) z_k of the true state x_k can be described by

$$z_k = Hx_k + v_k \quad \text{Eq47}$$

Where H is the observation model which maps the true state space into the observed space and v_k is the observation noise which is assumed to be independent (of each other), white, and with normal distribution $p(v) \sim N(0, R)$.

In practice, the process noise covariance Q and the measurement noise covariance R matrices might change with each time step or measurement, however here we assume they are constant. The $n \times n$ matrix A in the equation Eq46 relates the state at previous time step $k - 1$ to the state at the current step k in either the presence of the driving function or process noise. Note that in practice A must be change with each time step, but here we assume it is constant. The $n \times l$ matrix B relates the optional control input u to the state. The $m \times n$ matrix H in k^{th} measurement equation Eq47 relates the state to

measurement z_k . In practice H might also change with each time step or measurement, but here we assume it is constant.

- *Computation origins of the filter*

We define \hat{x}_k^- to be our a priori state estimate at step k given knowledge of the process prior to step k , and \hat{x}_k to be our a posteriori state estimate at step k given measurement z_k . We can define a priori and a posteriori estimate errors as

$$e_k^- \equiv \hat{x}_k - \hat{x}_k^- \quad \text{and} \quad e_k \equiv \hat{x}_k - \hat{x}_k^- \quad \text{Eq48}$$

The a priori estimate error covariance is then

$$P_k^- = E \left[e_k^- e_k^{-T} \right] \quad \text{Eq49}$$

and the a posteriori estimate error covariance is

$$P_k = E \left[e_k e_k^T \right]. \quad \text{Eq50}$$

In deriving the equations of Kalman filter, we start with a goal in finding an equation that compute the a posteriori state estimate \hat{x}_k as a linear combination of a priori estimate \hat{x}_{k-}^- and a weighted

difference between an actual measurement z_k and a prediction $H \hat{x}_{k-}^-$ as shown below.

$$\hat{x}_k = \hat{x}_{k-}^- + K (z_k - H \hat{x}_{k-}^-) \quad \text{Eq51}$$

The difference $(z_k - H \hat{x}_{k-}^-)$ in Eq6 is called measurement innovation, or residual. The residual reflects the discrepancy between the predicted measurement $H \hat{x}_{k-}^-$ and the actual measurement z_k . A residual of zero means that two are in complete agreement. The $n \times m$ matrix K in Eq51 is chosen to be the gain factor that or blending factor that minimizes the a posteriori error covariance Eq51. This minimization can be achieved by substituting Eq51 into Eq50 and performing the indicate expectations, then solving for K . For further details see [Maybeck 79; Brown 92; Jacobs 93]. K is given by

$$K_k = P_k^- H^T (H P_k^- H^T + R)^{-1} \quad \text{Eq52}$$

$$K_k = \frac{P_k^- H^T}{H P_k^- H^T + R} \quad \text{Eq53}$$

As we can see that when the measurement error covariance approaches zero, the gain weights the residual more heavily, $\lim_{R_k \rightarrow 0} K_k = H^{-1}$.

On the other hand, as the a priori estimate error covariance P_k^- approaches zero, the gain K weights the residual less heavily. $\lim_{P_k^- \rightarrow 0} K_k = 0$.

More sophisticated models of Kalman filter can be found for the purpose of fitting model based on nonlinear dynamic system. However, the Kalman filter is designed to make a good approximation fit when the system has a linear dynamic system.

- *Kalman filter algorithm*

This algorithm is based on recursive estimation by which the only estimate needed to compute the current state is the state from the previous time step and the current measurement. The state of the filter is represented by two variables:

\hat{x}_k^- the estimate of state at time k.

P_k , the error covariance matrix (a measure of the estimate accuracy of the state estimate).

Predict		
Predicted state	$\hat{x}_k = A \hat{x}_{k-1} + B u_k$	Eq54
Predicted estimate covariance	$P_k^- = A P_{k-1} A^T + Q$	Eq55
Update		
Innovation or measurement residual	$\hat{x}_k = z_k - H \hat{x}_k^-$	Eq56
Innovation (or residual) covariance	$S_k = H P_k^- H^T + R$	Eq57
Optimum Kalman gain	$K_k = P_k^- H^T S_k^{-1}$	Eq58
Update state estimate	$\hat{x}_k = \hat{x}_k^- + K_k \hat{x}_k^-$	Eq59
Update estimate covariance	$P_k = (I - K_k H) P_k^-$	Eq60

From the above equations, we can notice that the time update denoted by predict project the state and covariance estimates forward from time step $k - 1$ to step k . The first task during the measurement update is to compute the Kalman gain and the remaining follow as shown above.

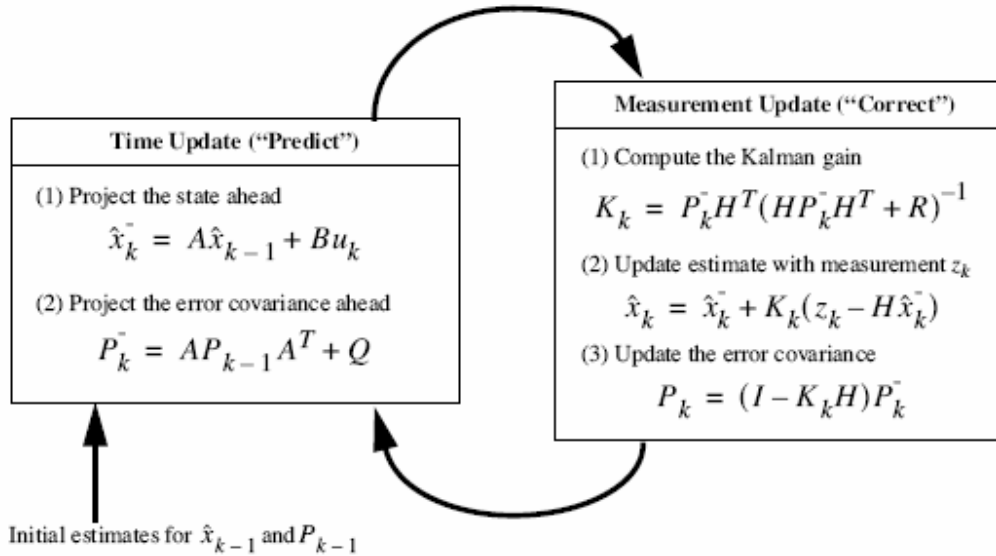


Figure 3: Complete picture of the linear Kalman filter operation

The Figure 3 shows a diagram to shortly give a compact explanatory of the Kalman filter operation as designed above. In closing we note that under conditions where Q and R are in fact constant, both estimation error covariance P_k and the Kalman gain K_k will stabilize quickly and then remain constant. In this case the parameter will be pre-computed off-line, or by determining the state value of P_k as describe in [Grewal 93]. In either case, we can obtain a good performance by tuning the filter parameters. The Kalman filter is a generalization of wiener filter. Unlike the Wiener filter, which is designed under the assumption that the signal and the noise are stationary, the Kalman filter has the ability to adapt itself to non-stationary environment. If the signal and noise are jointly Gaussian, the Kalman filter is optimal in a minimum MSE sense. If the signal and / or the noise are non-Gaussian, then the Kalman filter is the best linear estimator that minimizes MSE among all possible linear estimators. Moreover it is not convenient for online operation. It is also not shown to guarantee bounded error variance.

Chapter 5

Rotating Machine based on Vibration and Sound analysis

5.1 Introduction

In rotating machines, vibration and sound analysis can yield information based on a change of system vibration pattern which can be relevant to design engineer. Therefore these changes are analyzed and used as parameters to predict fault (in condition monitoring), improve the comfort and enhance the design quality in automotive products. A mechanical system which encompasses a car motor has been used for vibration and sound measurement. We are not going to specify the scientific condition of the data acquisition. We assume, that the data has obtained in the normal scientific condition and sampled with respect to Nyquist. In order to analysis data, it appears relevant to give a brief analysis of the rotating machines based on vibration analysis. The purpose is not to provide some technical analysis where the results can indicate a fault detection method, even though possible, for rotating machine based on a change of system vibration pattern or critical element in a system. We will thus give the waveform characteristic of the data, the statistical property of the data and the data model.

5.2 Vibration analysis

Machines are complex mechanical structures with articulated elements. The parts that are excited could oscillate; where joint to other coupled elements transmit such oscillations. The result is the complex frequency spectrum that characterizes the system. Each time a behaviour of component changes one of its mechanical characteristics because of wear or crack, a frequency component of the system will be affected. However, in an automotive, we are more concerned with tracking the auto to translate the rotational angular velocity into a speed profile. Therefore, for our study and the requirement specification purposes based on fundamental frequency tracking and performance analysis, we will focus on the technical analysis through the use of the Pulse software package to analyze the obtained data by using a statistical approach called Bayesian analysis. In such an analysis, the discrete time data is processed to track the fundamental frequency in a full band vibration and sound signals made of more than 11 harmonics associated with noise. The purpose of this work is to get acquainted with the utilization of PULSE software. Moreover, it is to find the suitable parameters that yield the optimal estimate or track the fundamental frequency of interest.

5.2.1 Vibration and Sound waveforms description

In previous chapters, we have been dealing with some concepts and some theories behind the Bayesian analysis which may be used to get acquainted with Bayesian probability theory. That is the derivation of the posterior probability distribution and the incorporation of the informative prior based on mathematic manipulation has yielded good theoretical results. Now we consider vibration and sound signals which are highly nonstationary but relevant for our study. The reason is the interests of Brüel & Kjær to investigate a new way to determine the running speed of a car engine for application design purpose. Thus we are concerned with fundamental frequency tracking one of the core tasks in automobile department at Brüel and Kjær vibration and sound measurement A/S. The goal is to use these measurements to track the trajectory of the fundamental frequency. Such a goal can be reached with a well formulated technique which can take into account the practical aspect of the nonstationary data model and the uncertainty of the stochastic parameter being estimated. In automotive jargon, it is baptised auto tracking. In order to achieve our goal, it appears necessary to organize the task as follows:

- *Waveform characteristic*

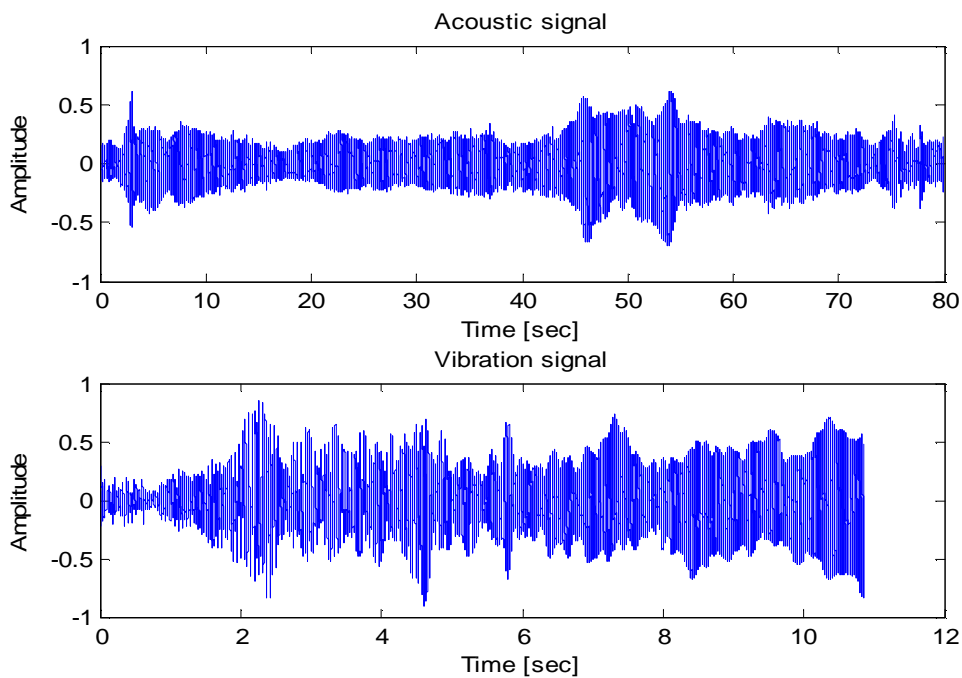


Figure 4: Acoustic (upper panel) and Vibration (lower panel) waveform signal. These show amplitude plot versus time. As we see, the amplitudes of these signals are characterised by an unpredictable fluctuation over time.

We section the signal for the purpose of having a closer look on the waveform characteristics. The Figure 5 shows both the tacho for each signal (lower panels) and the signal characteristics (upper panel).

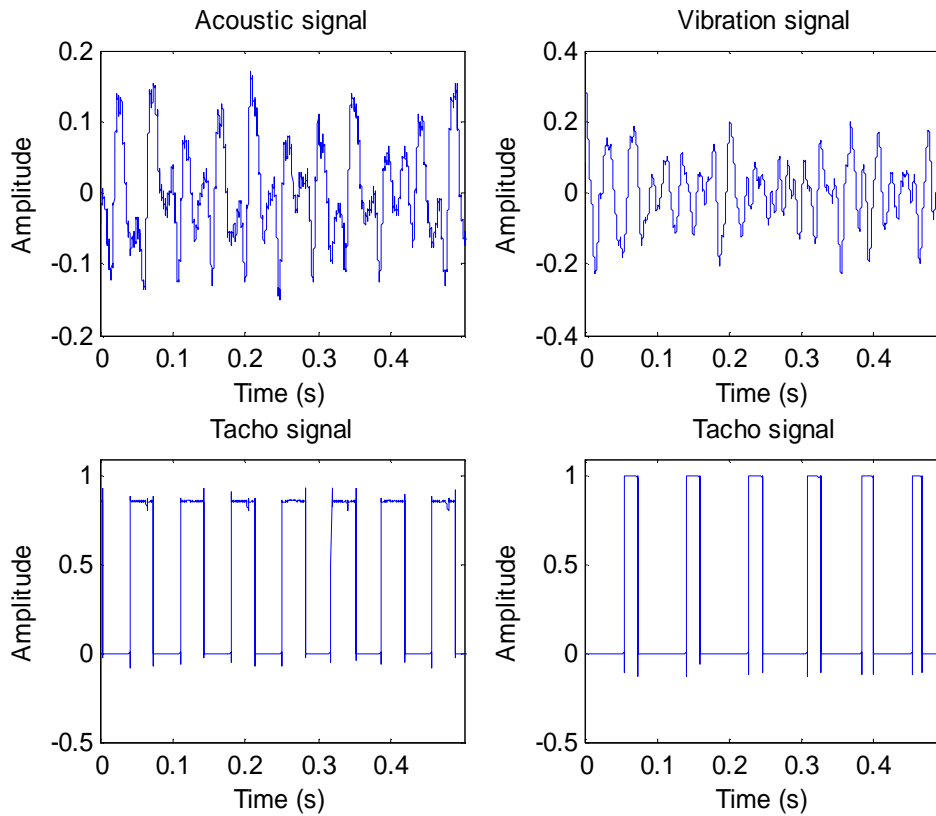


Figure 5: Acoustic and Vibration signals (upper panel) which are nonharmonic signals and their respective tacho signals dominated by several pulses to designate the periodicities.

5.2.2 Spectrogram of the data

Next, we consider the frequency content of these signals. The reason is simple. We are interested to track the fundamental frequency of the signal which requires knowledge about the number of harmonics in the signals. Thus we plot the spectrogram in Figure 6.

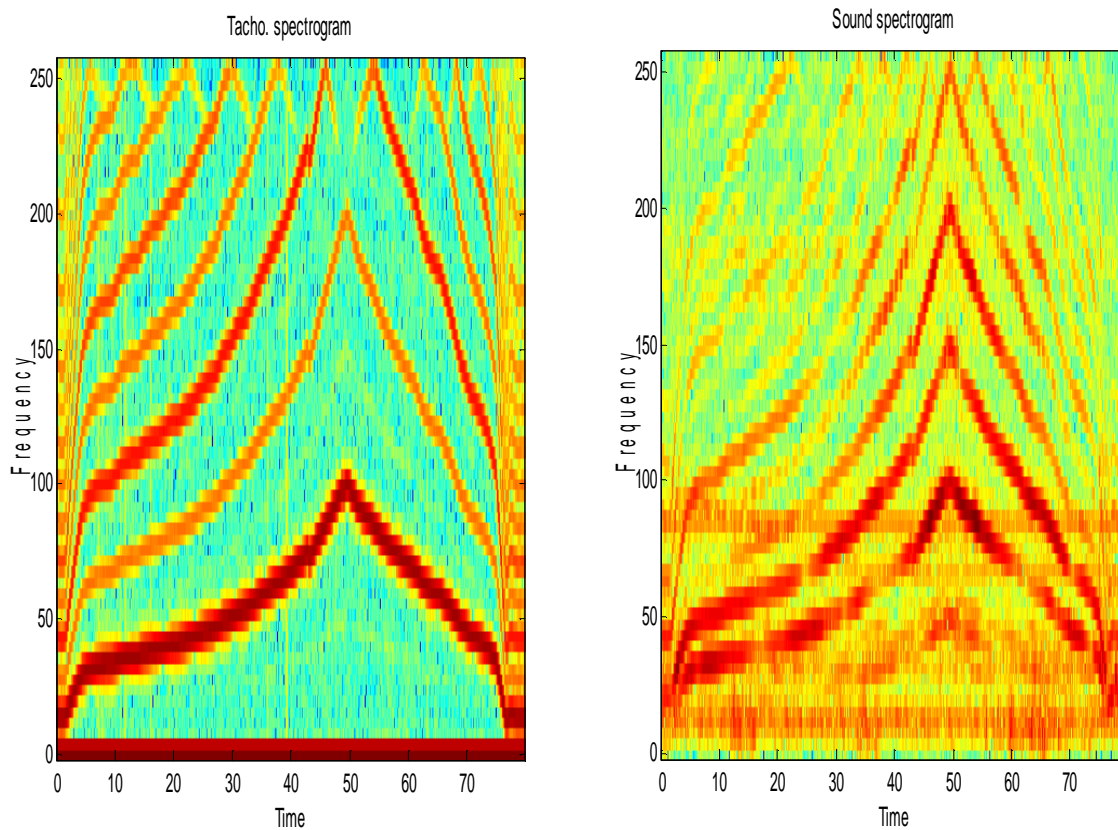


Figure 6: Spectrogram of both tacho (left) and the acoustic (right) harmonics order change.

As we can see Figure 6 presents two spectrograms of both the tacho (left) and the sound (right) signals. These spectrograms show a clear picture of amplitude of several harmonics components evolving in time. These start at the low frequency of 100 Hz with first order, then as the frequency increases the number of dominating order increases. Thus the amplitude of the related harmonics changes with the fundamental frequency. Moreover these spectrograms present the energy of the frequency contents (frequency spectrum) of windowed frames for the run up and run down when the frequency changes over time.

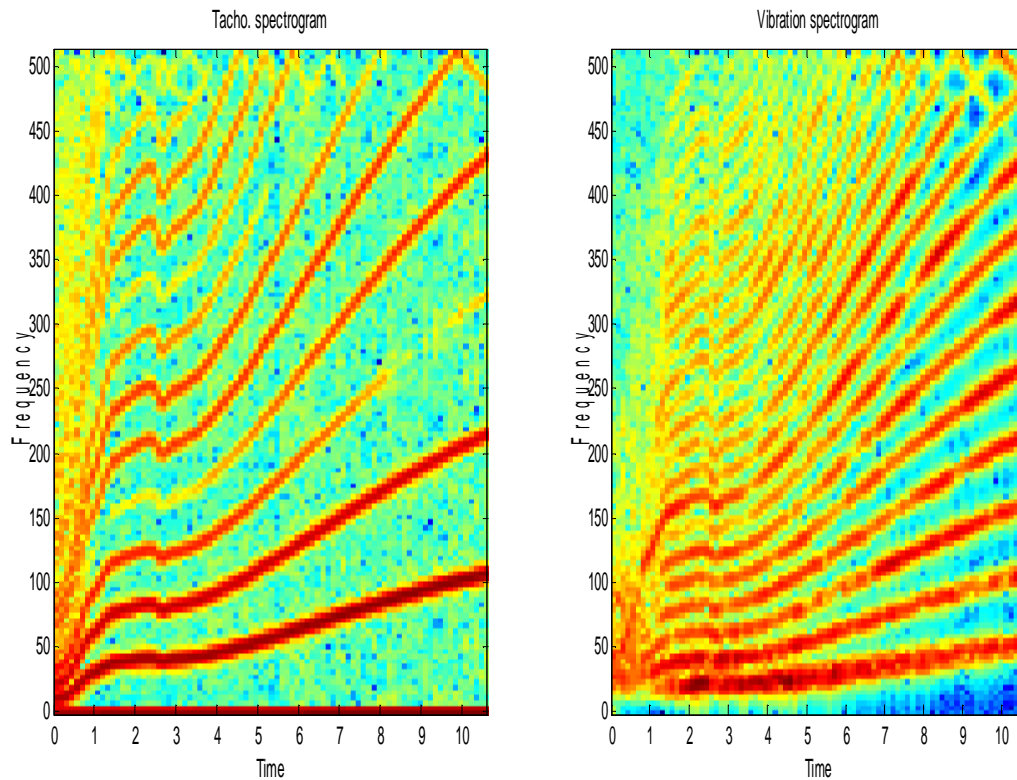


Figure 7: Spectrogram of both Tacho (left) and Vibration (right) harmonics order change.

These spectrograms show the amplitude of the harmonic components and the frequency content of these signals. As the harmonic order changes, the amplitude changes as well. This change means that the frequency is varying over time. Thus for nonstationary signals, such a tool based on time-frequency spectrum is suitable for that purpose. Furthermore, by inspection, we see that the fundamental frequency for the acoustic has its peak at 100 Hz. The vibration of the motor yields two pulses per rotation represented by the tacho. That is, when the tacho has peak at 100 Hz, the vibration is at 50 Hz (see the first harmonic on the Figure 7). These are the fundamental frequencies of both the acoustic and the vibration signals we are going to track. Moreover, these figures present strong harmonics, DC (in tacho spectrograms) and some aliasing (top of each figure) for run up movement. These data have been generated in unknown condition by Bruel & Kjaer and consist of three signals:

- *Tacho reference measured optically from a cam-shaft of the engine*
- *Vibration (called also acceleration) in the vertical direction of the engine block measured with accelerometer.*
- *Acoustic sound pressure measured with a microphone approximately 1 meter above a car engine.*

We assume that the data have sampled according to the Nyquist theorem.

5.2.3 Data model

We have seen that in the spectrogram, both the vibration and the sound signals encompass several harmonics and other artefacts due to aliasing. Because we are concerned more with the harmonic frequencies, we will reduce our model to the sum of the some harmonics. This will be given as follows:

$$x(t) = \sum (a_n \cos(\eta_n \Omega(t)) + b_n \sin(\eta_n \Omega(t))) \quad \text{Eq61}$$

where

$$\Omega(t) = \int_0^t w(\tau) d\tau$$

This is the representative signal, where the fundamental frequency will be estimated with respect to (wrt) the harmonic structure, frequency order and the amplitude of the orders. This model adds with noise will yield the regression model described above. In such a case, we may determine statistical properties of the data.

5.2.4 Descriptive statistics

In order to study the complete informative description of the waveform, statistical variability, the shape of the distribution and a quantitative analysis are setup.

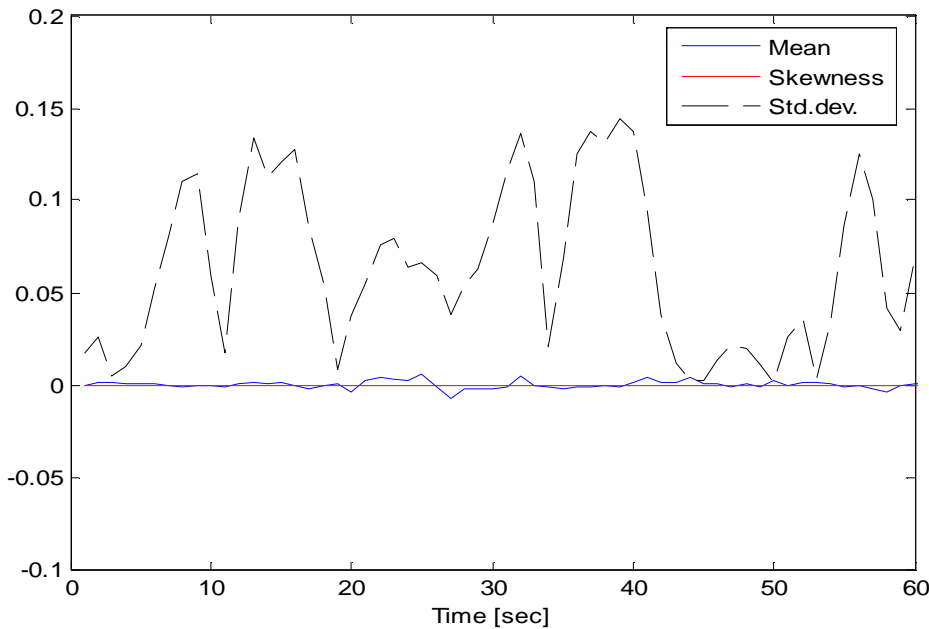


Figure 8: Graphical representation of the partial quantitative description.

Figure 8 depicts the descriptive statistics. It is clear that the signal is highly nonstationary due to the variability of the standard deviation. The skewness lying in on the x axis tells us that the distribution shape of our sampled data may be symmetrical. Hence we will use the Gaussian distribution.

- *Student's t-distribution*

The conjugate prior for the precision of a Gaussian is given by a Gamma distribution. If we have a univariate Gaussian $N(x | \mu, \tau^{-1})$ together with Gamma prior $Gam(\tau | a, b)$ and we integrate out the precision, we obtain the marginal distribution

$$p(x | \mu, a, b) = \int_0^\infty N(x | \mu, \tau^{-1}) Gam(\tau | a, b) d\tau \quad \text{Eq62}$$

After some manipulations, we obtain the student's t-distribution defined by

$$St(x | \mu, \lambda, \nu) = \frac{\Gamma(\nu/2 + 1/2)}{\Gamma(\nu/2)} \left(\frac{\lambda}{\pi\nu}\right)^{1/2} \left[1 + \frac{\lambda(x - \mu)^2}{\nu}\right]^{-(\nu+1)/2} \quad \text{Eq63}$$

Where λ is sometimes called the precision of the t-distribution even though it is not in general equal to the inverse of the variance. The variance is $\text{var}[X] = \frac{1}{\lambda} \frac{\nu}{\nu-2}$ and $\nu > 2$. The parameter ν is called the degree of freedom, and its effect can control the shape of the t-distribution. For the particular case of $\nu = 1$, the t-distribution reduces to the Cauchy distribution, while in the limit $\nu \rightarrow \infty$ the t-distribution $St(x | \mu, \lambda, \nu)$ becomes a Gaussian $N(x | \mu, \tau^{-1})$ with mean and μ precision λ .

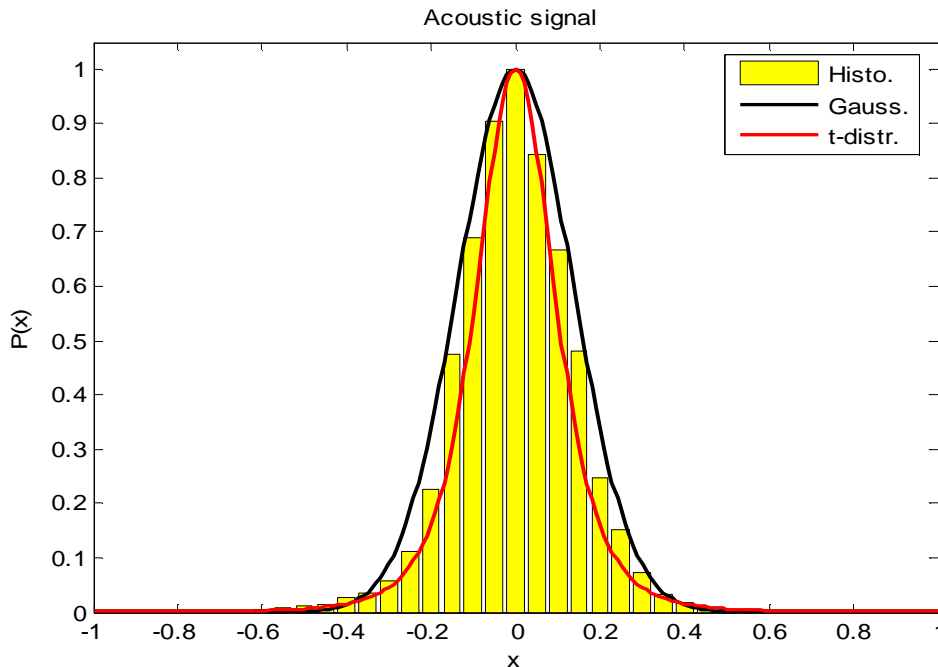


Figure 9a: Complete descriptive information of sound signal model.

In the Figures 9a and 9b, we consider 3 distributions namely the histogram, Gaussian and the Student's t-distributions representing the probability distribution of the acoustic data. The bars of the histogram in both right and left sides taper in the same way. These tapering sides are called tails (or snakes), and provide a visual shape of the distribution. Such a distribution contains both right longer tail (positive skew) and left longer tail (negative skew). The distribution is said to be skewed. From equation eq1, we notice that student's t-distribution is formed by adding up an infinite number of Gaussian distributions with same mean and different precisions. This is referred to an infinite mixture of Gaussians. That is, the distribution has in general longer tails than a Gaussian. This gives the t-distribution an important property called robustness, which means that it is much less sensitive than a Gaussian to the presence of outliers. Outliers can arise in practical applications either because the process that generates the data corresponds to a distribution having heavy tail or simply mislabelled data. Robustness is also an important property of the regression problem.

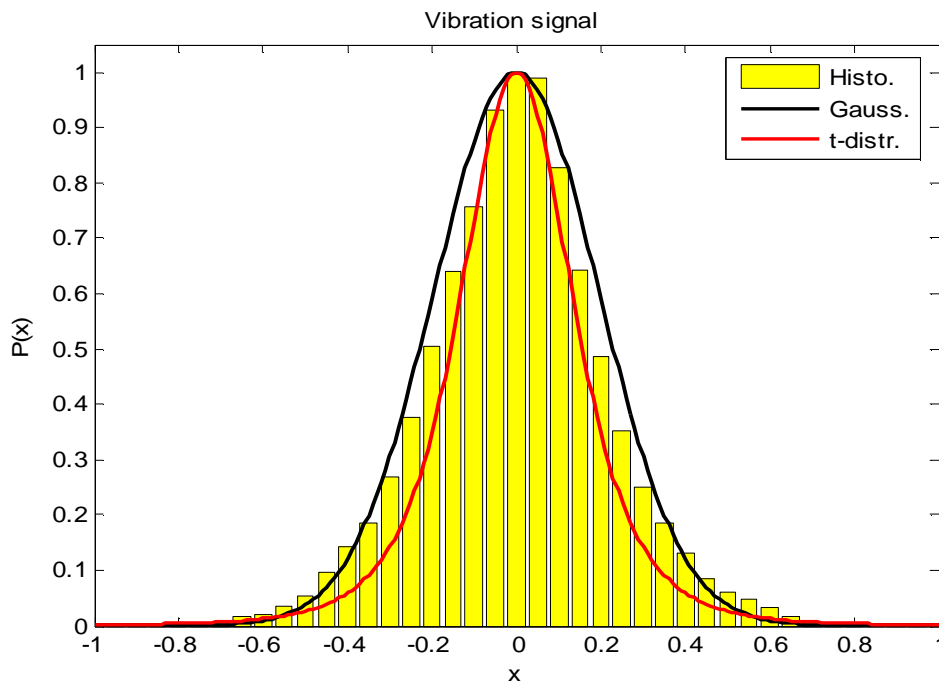


Figure 9b: Complete descriptive information of vibration signal model.

- *Interpretation*

Illustration of the robustness of Student's t-distribution compared to Gaussian. Histogram distribution of 401 data points from a maximum likelihood fit obtained from a t-distribution (red curve) and a Gaussian distribution (dark curve). The t-distribution contains the Gaussian as a special case. It gives almost the same solution as the Gaussian. The complete statistical description of the data confirms that the distribution of the acoustic and vibration data have a symmetric distribution shape. Further, the elongation of the histogram tails confirms the heavier tail. However, such tails happen to be short and die out faster. Therefore, we will consider the Gaussian distribution to be suitable probability density function (pdf) for short bandwidth. The variability observed by the variance indicates that the signal is highly nonstationary.

5.3 Robust Bayesian tracking algorithm

The previous chapters were concerned with the estimation of the stationary frequency. If the instantaneous frequency changes substantially under low SNR, there is not much that can be done to track the frequency as it changes. However, if the frequency is changing so slowly, then we could track the instantaneous frequency simply by estimating it independently over time blocks using the above mentioned method. In this section we are concerned with the estimation of a stochastic variable namely the unknown fundamental frequency from vibration and sound data sampled uniformly.

The data record is defined by

$$d^{(k)} = [d(t_{k\Delta N}), d(t_{\Delta N+1}), \dots, d(t_{\Delta N+M-1})]^T \quad \text{Eq64}$$

where M is the number of samples in each record. The records are offset from each other ΔN .

The parameter to track

$$\Theta_L \equiv \{w_F^{(i)} : 0 \leq i < L\} \quad \text{Eq65}$$

The observations are defined by

$$D_L \equiv \{d^{(i)} : 0 \leq i < L\} \quad \text{Eq66}$$

The likelihood function is defined by

$$p(d | \theta) = \frac{1}{(2\pi\sigma)^{N/2}} \exp\left\{-\frac{1}{2\sigma^2}(d^T d - f^T f)\right\} \quad \text{Eq67}$$

- *Parameter vector or matrix*

$$t_{N \times 1} = [t_0, \dots, t_{N-1}]^T$$

$$d_{N \times 1} = [d(t_0), \dots, d(t_{N-1})]^T$$

$$b_{2K+1 \times 1} = [A_0, \dots, A_K \ B_1, \dots, B_K]^T$$

$$f = G \hat{b}$$

$$\hat{b} = (G^T G)^{-1} G^T d$$

$$d = Gb + e$$

$$h_{K \times 1} = [\eta_1, \dots, \eta_K]^T$$

$$u_{N \times 1} = [1, \dots, 1]^T$$

$$G_{N \times 2K+1} = [u_{N \times 1} \quad \cos(w_F t_{N \times 1} h^T) \quad \sin(w_F t_{N \times 1} h^T)]$$

The value of Θ_L corresponding to the MAP estimate of $p(\Theta_L | D_L)$ is the optimum track. In order to calculate the posterior probability, we need to find the prior model.

5.3.1 Modeling the informative prior distribution

We know that the frequency changes slowly over time. The successive samples will not be too different, suggesting that there is high degree of correlation between the samples as mentioned earlier. Such effect leads us to consider a conditional probabilistic model namely Markov model. The reason is that we need to relax the identically independent distributed assumption of the observations; to captures the slowly change effects. Thus with such a model, all information about $w_F^{(i)}$ from the past observation is contained in the previous observation $w_F^{(i-1)}$.

$$p(w_F^{(k)} | w_F^{(k-1)}, \dots, w_F^{(k-P)})$$

If $w_F^{(i)}$ obeys Eq6 it is said to be P-order Markov process. The joint posterior becomes

$$\begin{aligned} p(\Theta_L | D_L) &= p(w_F^{(L-1)} | \Theta_{L-1}, D_L) p(\Theta_{L-1} | D_L) \\ &\propto p(d^{(L-1)} | w_F^{(L-1)}) p(w_F^{(L-1)} | \Theta_{L-1}) p(\Theta_{L-1} | D_{L-1}) \\ &\propto p(d^{(0)} | w_F^{(0)}) p(w_F^{(0)}) \prod_{i=1}^{P-1} p(d^{(i)} | w_F^{(i)}) p(w_F^{(i)} | \Theta_i) \\ &\quad \times \prod_{i=P}^{L-1} p(d^{(i)} | w_F^{(i)}) p(w_F^{(i)} | w_F^{(i-1)}, \dots, w_F^{(i-P)}) \end{aligned}$$

- *The posterior probability*

$$\begin{aligned} p(\Theta_L | D_L) &\propto p(d^{(0)} | w_F^{(0)}) p(w_F^{(0)}) \prod_{i=1}^{P-1} p(d^{(i)} | w_F^{(i)}) p(w_F^{(i)} | w_F^{(i-1)}, \dots, w_F^{(0)}) \\ &\quad \times \prod_{i=P}^{L-1} p(d^{(i)} | w_F^{(i)}) p(w_F^{(i)} | w_F^{(i-1)}, \dots, w_F^{(i-P)}) \end{aligned} \quad \text{Eq68}$$

As we can clearly see, the posterior probability is the resultant of Bayesian inference. Three scenarios may be designed to compute the sets of the posterior probabilities which can yield the fundamental frequency components.

5.3.2 Tracking location parameter

As we know that the unknown parameter varies slowly over records or time, in the case of Gaussian distribution, with the fundamental frequency corresponding to the mean, the regression function is linear. It may therefore be defined by

$$w_F^{(n-k)} = w_F^{(n)} + \alpha k + \eta(n-k) , \tag{Eq69}$$

Where η is $N(0, \sigma_T^2)$, $1 \leq k \leq P$ and α is the rate of change. If α is too big, the change will be too fast and the tracker may not perform well. The other parameters are important and these will be given in a simulation part.

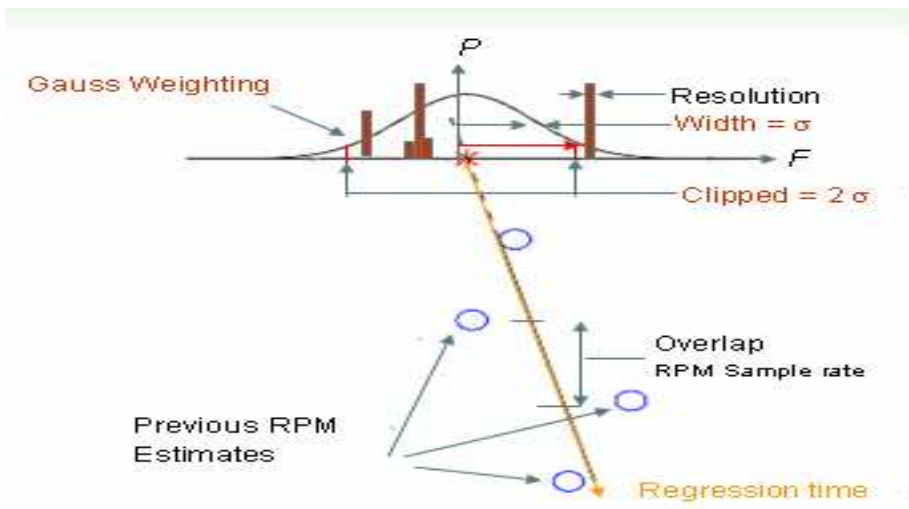


Figure 10: Tracking prior from linear regression.

The linear regression can be used to estimate the location parameter in the subsequent observation. In the Figure 10, the description of how to determine the prior probability distribution is shown by linear regression.

- *Determination of the prior mean*

We use a conjugate prior based on Gaussian distribution. The determination of its parameter follows Thorkild Pedersen procedure:

$$p(w_F^{(n)} | w_F^{(n-1)}, \dots, w_F^{(n-P)}) = N(\mu_T, \sigma_T^2) \tag{Eq70}$$

$$\mu_T \equiv \begin{cases} \frac{\sum_{k=1}^P (2(2P+1) - 6k) w_K^{(k-1)}}{P(P-1)} & , \text{ for } P > 1 \\ w_F^{(n-1)} & , \text{ for } P = 1 \end{cases} \tag{Eq71}$$

5.3.3 Procedure of fundamental frequency tracking using informative prior

- 1- Segmentation of the data set.
- 2- Overlap the each data segmented to each other if $\Delta N < M$.
- 3- Compute the posterior distribution $p(w_F^{(k)} | d^{(k)})$ of the fundamental frequency
- 4- Find maximum a posteriori (MAP): $p(\Theta_L | D_L)$
 - *Prior information not available wrt $w_F^{(i)}$ the MAP is defined by*

$$p(\Theta_L | D_L) = \prod_{i=0}^{L-1} p(w_F^{(i)} | d^{(i)})$$

It is defined by $p(w_F^{(i)} | d^{(i)})$ such that

$$\left\{ \arg \max_{w_F^{(i)}} p(w_F^{(i)} | d^{(i)}) : 0 \leq i < L \right\}$$

- *Prior knowledge available the MAP is defined as follows*

$$p(\Theta_L | D_L) \propto p(d^{(0)} | w_F^{(0)}) p(w_F^{(0)}) \prod_{i=1}^{P-1} p(d^{(i)} | w_F^{(i)}) p(w_F^{(i)} | w_F^{(i-1)}, \dots, w_F^{(0)}) \\ \times \prod_{i=P}^{L-1} p(d^{(i)} | w_F^{(i)}) p(w_F^{(i)} | w_F^{(i-1)}, \dots, w_F^{(i-P)})$$

5- Compute the posterior probability as follows

1. Initialization

$$p(w_F^{(0)}), \sigma_T^2$$

2. Compute

$$w_F^{-(0)} = \arg \max_{w_F^{(0)}} (p(d^{(0)} | w_F^{(0)}) p(w_F^{(0)}))$$

3. For $1 \leq k < P$:

$$w_F^{-(k)} = \arg \max_{w_F^{(k)}} (p(d^{(k)} | w_F^{(k)}) p(w_F^{(k)} | w_F^{(k-1)}, \dots, w_F^{(0)}))$$

4. For $P \leq k < L$:

$$w_F^{-(k)} = \arg \max_{w_F^{(k)}} (p(d^{(k)} | w_F^{(k)}) p(w_F^{(k)} | w_F^{(k-1)}, \dots, w_F^{(k-P)}))$$

Chapter 6

Results for Computer Simulations

6.1 Spectral Analysis simulation

6.1.1 Performance analysis using stationary signal

- Experiment 1: Single harmonic frequency estimation

This experiment is a simple frequency estimation based on a single harmonic in sine wave. We generate a 2501 periodic discrete time samples. We will mention that in these experiments, we assume that all data are uniformly sampled. We apply only the periodogram and the student t-distribution. The results are depicted in Figure 11.

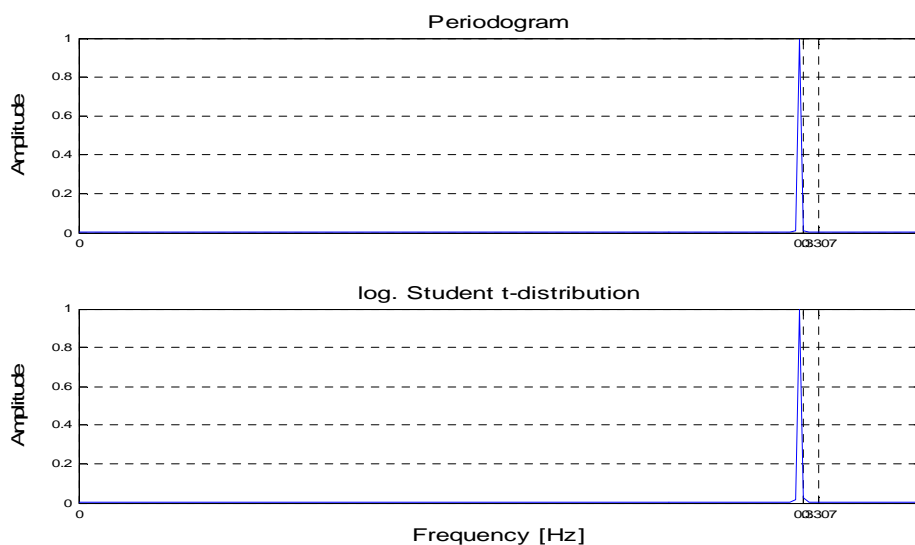


Figure11: Spectral estimate comparison

The figure shows evidence of one peak in each panel. The upper panel is the result of the periodogram resolving perfectly the single harmonic frequency. The second harmonic is also at the right position. Hence, these two estimators have successfully yielded the single harmonic in the signal.

- Experiment 2: Two harmonic frequencies estimation

In this experiment, we are interested in the power carried by each line; not in the total power carried by the signal. This can be a real issue as the two lines become closer and closer together so that power is shared between them. The figure 12 shows an example of such an issue. To illustrate this point, we generate a discrete time sine wave sampled uniformly. We use 2501 sampled data. We then estimate the frequencies. Figure 12 shows the spectral components of two closed harmonic frequencies. In the upper panel, the periodogram shows only one peak. This estimator has estimated a frequency which is the average of the two frequencies. In the lower panel of the figure, the student t- distribution shows two frequency peaks at wrong position. Thus the inclusion of the improper prior has enhanced the ability of the estimator in the lowest panel to emphasize the evidence of two harmonic frequencies.

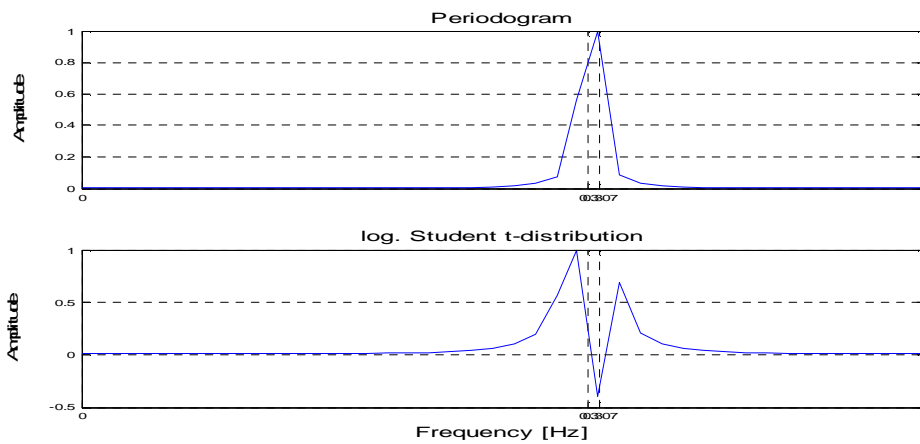


Figure 12: Power of the prior and spectral estimation.

Therefore, we note that prior, even uninformative can have a major effect on the conclusion we are able to draw from a given data set. This plot illustrates clearly some of the points we have been mentioning earlier even though the estimate of the student t-distribution may seem very conservative (see Figure 12). When we increase the data size, the Figure 13 shows evidence of two peaks for each of these estimators. The periodogram and the student t-distribution yield successfully the two frequency components as shown in both upper and lower panels respectively. As we may know from literature, it is not easy to retrieve too closed harmonics. At some limit, it may be even very difficult.

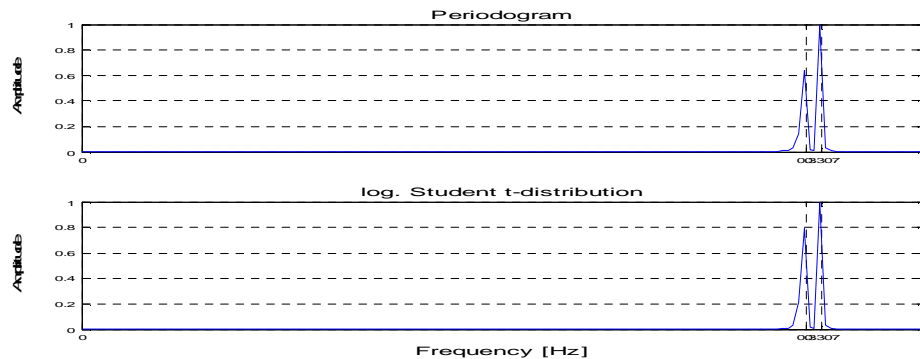


Figure 13: Spectral analysis of too closed harmonics.

However, we can solve such an issue by applying the likelihood method introduced in section 5.3.

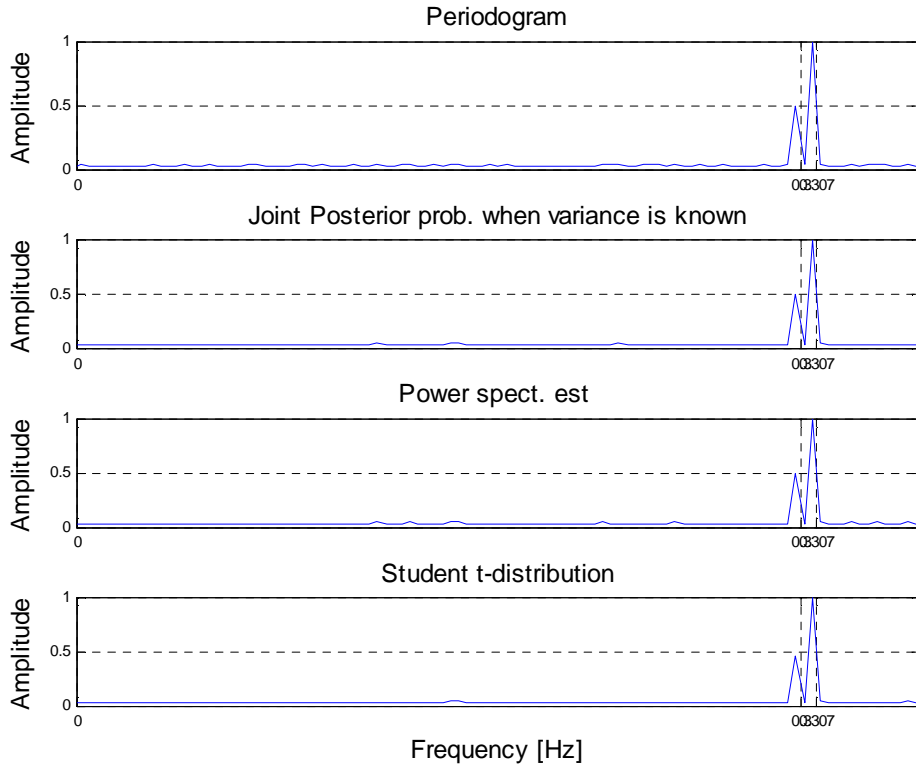


Figure 14: Spectral analysis showing spurious peaks caused by noise effect.

We now want to observe the effect of noise on these estimators. Therefore using the same signal with two harmonics frequencies, we increase the noise variance level beyond the fading limit say SNR set to -40dB . In such a lower SNR we apply these estimators to resolve the frequencies of interest. The noise effect has been attenuated by the ensemble averaging technique employed on the power spectral of each estimator. This is to reduce the variability of the power spectral estimates due to random noise effect. It results that the noise is filtered out. However, the increase of the noise at certain level has significant effect on the periodogram (see upper first panel). It presents spurious effects on its estimates which may be considered as frequencies components. On the other hand, the estimators based on the posterior probability distribution do not suffer from the same effects of the spurious components. The periodogram is not even a sufficient statistic in noisy environment because it becomes significantly affected by noise (see Figure 14). We have shown that the periodogram is very powerful to single tone signal. Despite the sample size of the data, student t-distribution can demonstrate the evidence of the exact number frequency present in the signal. This difference of resolving frequencies in a low SNR signal is due to the additional effect of the prior. Such a prior can help to enhance the ability of emphasizing the evidence of the frequency component in the signal. Moreover, the student t-distribution withstands the effect of the noise at certain level. Therefore without concluding, we may say that the marginal posterior probability remains the flexible estimator and yield good performance with the inclusion of the prior distribution.

- Experiment 3: Multi stationary harmonic frequency estimation

In this experiment, we generate a discrete time uniformly spaced sinusoid sample with a four low frequencies: $f_1=0.1$, $f_2=0.2$, $f_3=0.4$ and $f_4=0.6$. The sample size was 3001. The sampling frequency is 50 Hz. We apply the periodogram and the joint posterior probability distributions with and without knowing the variance, the results are depicted in the Figure 16. Matlab code used:

bayes_stationary_spect_ana.m. The Figure 15 shows the results which perform the multiple stationary frequency estimation with closed four closed harmonic related frequencies. The results are shown in the Figure 15.

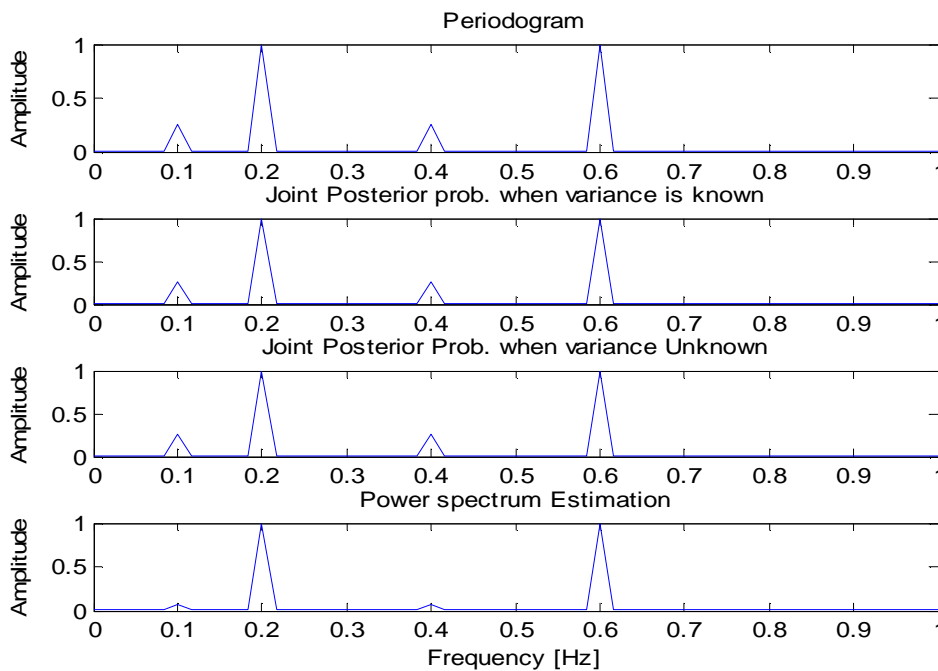


Figure 15: Spectrum of four related low frequencies using both periodogram, the joint posterior probability and the spectral estimator $\hat{p}(w)$ with and without variance being known. All the estimators show evidence of four peaks at the right position. Only the $\hat{p}(w)$ estimator shows a low amplitude of the estimates $2f_0$ and $4f_0$ where $f_0=0.1$.

We add noise a high noise level say SNR to -35.4 dB beyond the fading criterion. And then we apply these estimators. They successfully show four peaks at the right frequencies position. Although the success of these estimators the spectral estimator in the lowest panel appears to withstand the noise effect. The remaining ones, periodogram and the two joint posterior probability estimators yield both the right spectrum and also show evidence of spurious peaks. This is due to the low level of the SNR. The scenario is presented in Figure 16.

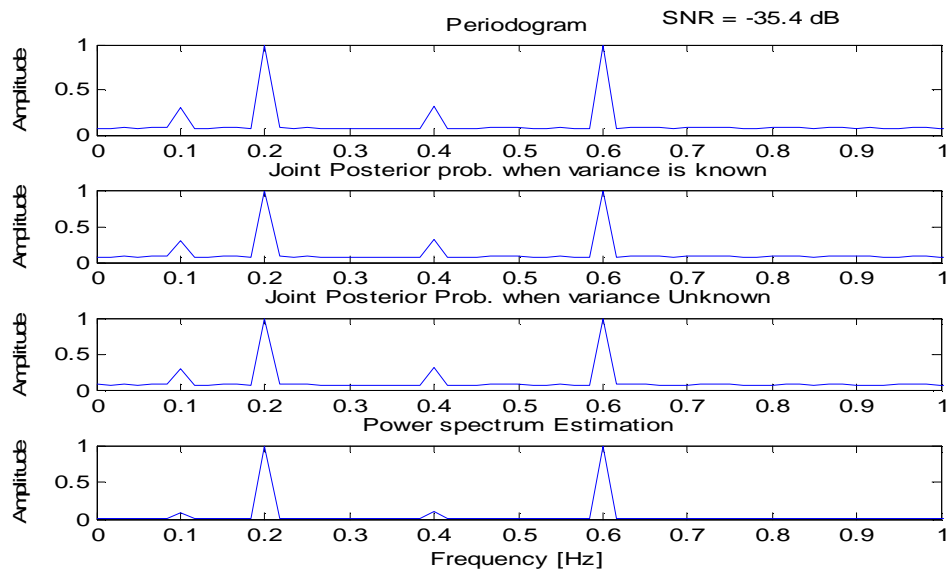


Figure 16: Ensemble average spectrum when SNR is set to -35.4 dB. The performance of these estimators shows evidence of four correct peaks.

Further, we examine the performance of these estimators when three of these four harmonic frequencies are too close. These estimate successfully the four harmonics as shown in Figure 17.

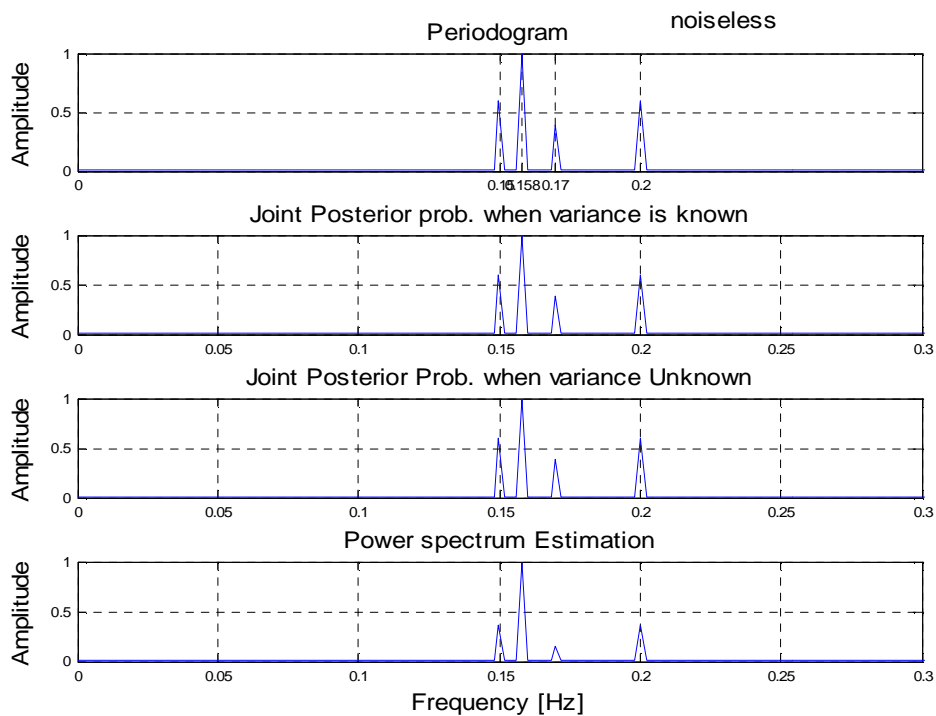


Figure 17: Ensemble average spectral estimation when the frequencies are clustered together.

- Experiment 4: Multiple nonstationary harmonic frequency estimation

In this experiment, we will investigate the capability of the periodogram and the student t-distribution to estimate the nonstationary frequencies from two uniformly sampled signals with two separate frequencies and decay factor. The signal is modeled as follows:

$$f_1(t) = [B_1 \cos(w_1 t) + B_2 \sin(w_1 t)]e^{(\alpha_1 t + \phi_1)} \text{ and } f_2(t) = [[B_3 \cos(w_2 t) + B_4 \sin(w_2 t)]e^{(\alpha_2 t + \phi_2)}]$$

Parameters used are:

$$B_1 = 1.5, B_2 = 4, B_3 = 2, B_4 = 3, w_1 = 0.3 \text{ rad/s}, w_2 = 0.5 \text{ rad/s}, \phi_1 = 0^\circ, \phi_2 = 90^\circ, f_s = 50 \text{ Hz}.$$

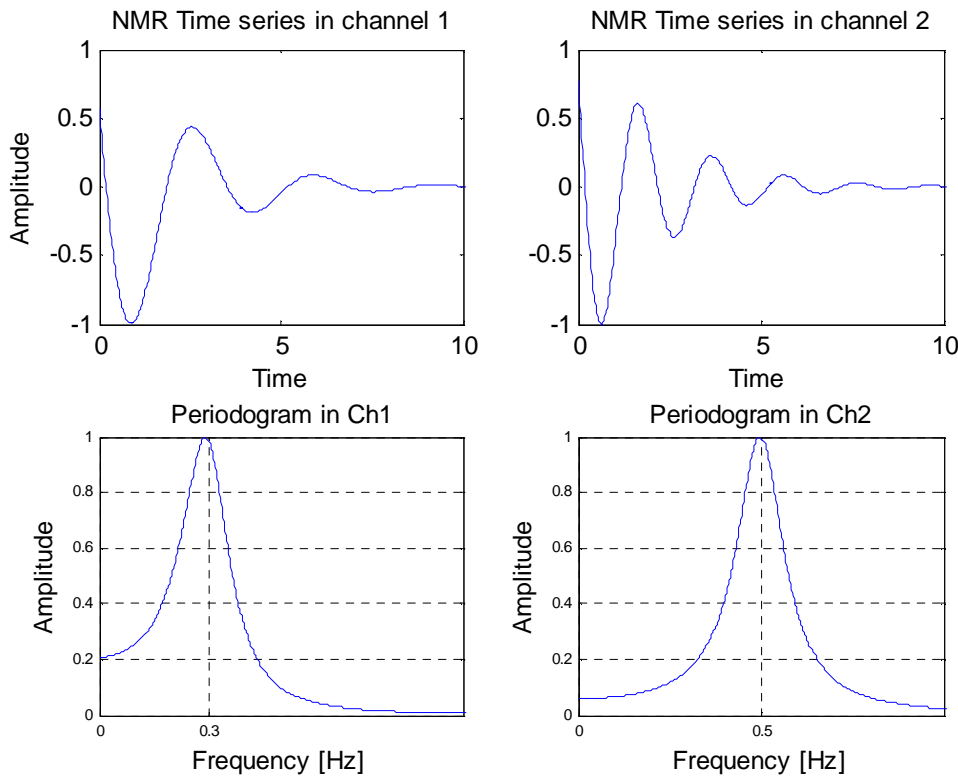


Figure 18: Performance of the spectral estimates for the periodogram.

Figure 18 shows the time series of the NMR free induction decay data from two different channels represented by channel 1 and channel 2 (upper panel). In the lowest panel, the estimates of the periodogram are shown for of each channel. We see only one peak for each signal or channel. This is reasonable because each signal contains only one frequency component. When these channels are added or combined, the estimation of both frequencies by periodogram fails. This is due to the incapacity of the periodogram to resolve nonstationary frequency (see upper panel in Figure 19).

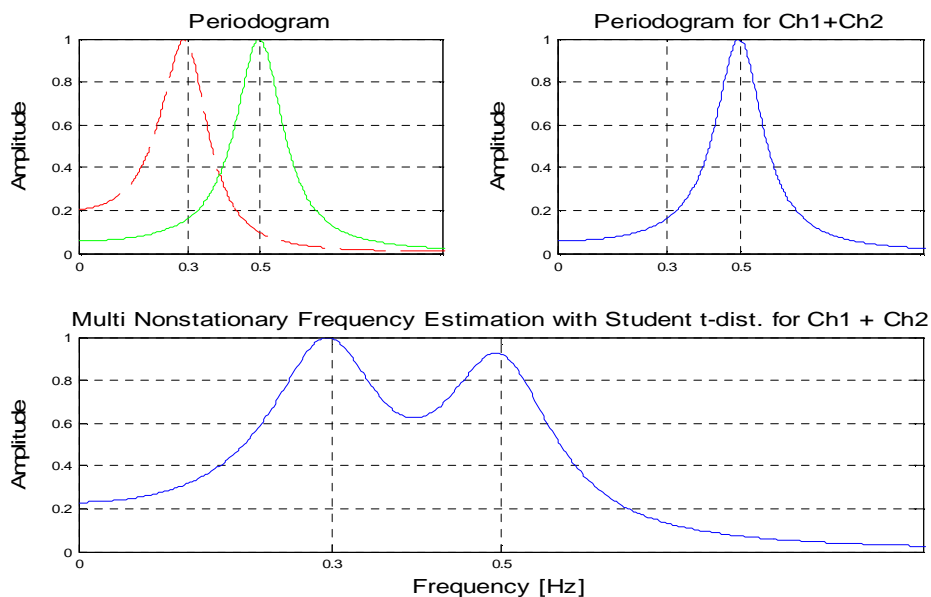


Figure 19: Performance of the spectral analysis by the periodogram and the student t-distribution.

In the lowest panel, we note the evidence of two peaks at the right position of the frequencies needed. This indicates that the frequencies of interested have been successfully estimated by the student t-distribution. This is also in harmony with the theories in many literatures that postulate that the student t-distribution outperforms the periodogram in certain conditions. We now add white Gaussian to our signal model (Figure 20); and then we apply both the periodogram and the student t-distribution to the unnormalized signals. The results of the experiment are exactly the same as in Figure 19.

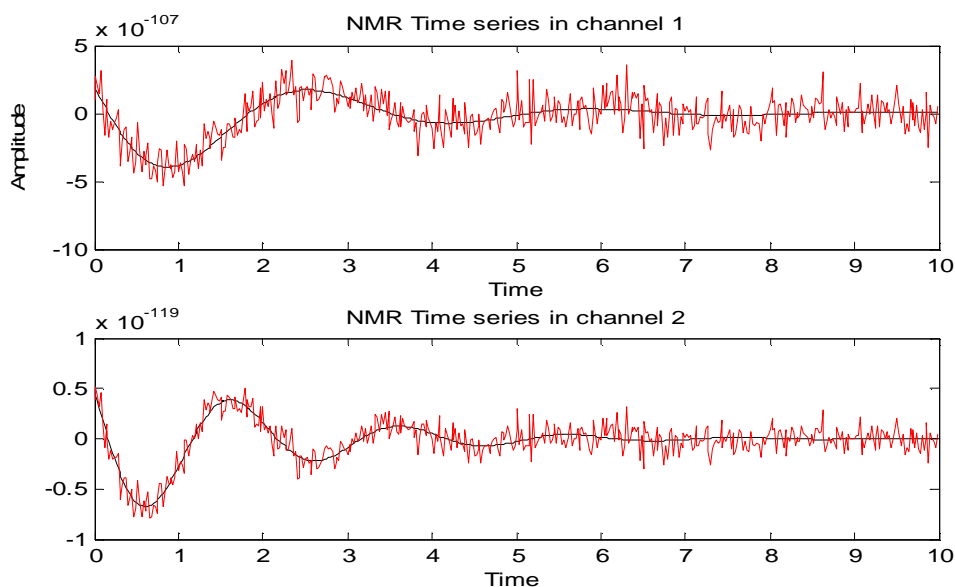


Figure 20: The true signals are corrupted (black) into white Gaussian noise (red).

Now, we would like to know the behaviour of such an estimator under different condition such as normalization and in a noisy environ. Therefore, we undertake a new experiment with the same signal and same parameters as before. We normalize the original signals $f_1(t)$ and $f_2(t)$, and then we apply these two estimators. The results are shown in Figures 21. We can clearly see the evidence of two peaks in the upper and lower panels. The periodogram and the student t-distribution have successfully estimated these two frequencies.

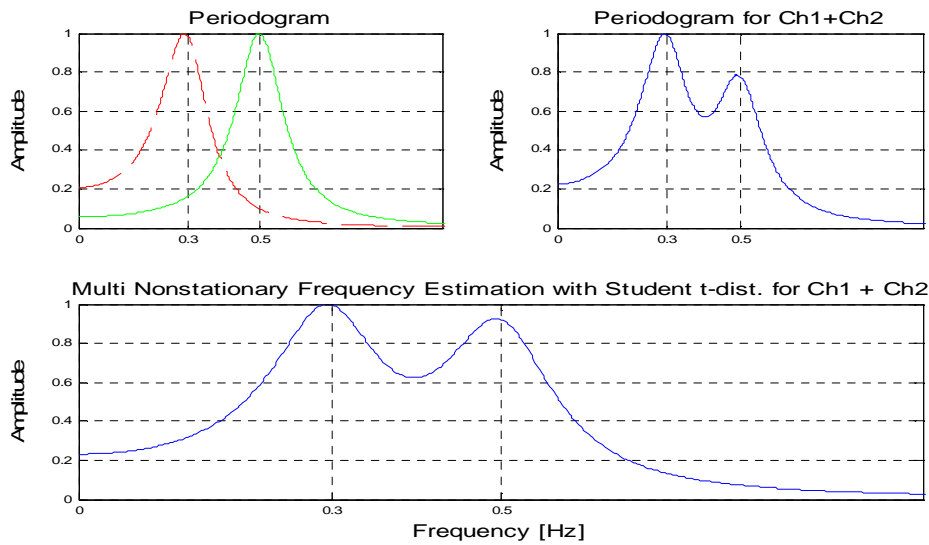


Figure 21: Frequency estimation under signal normalization condition.

At last, we normalized the signals and then add white Gaussian noise with variance set to 0.005. The results are shown in the Figure 21.

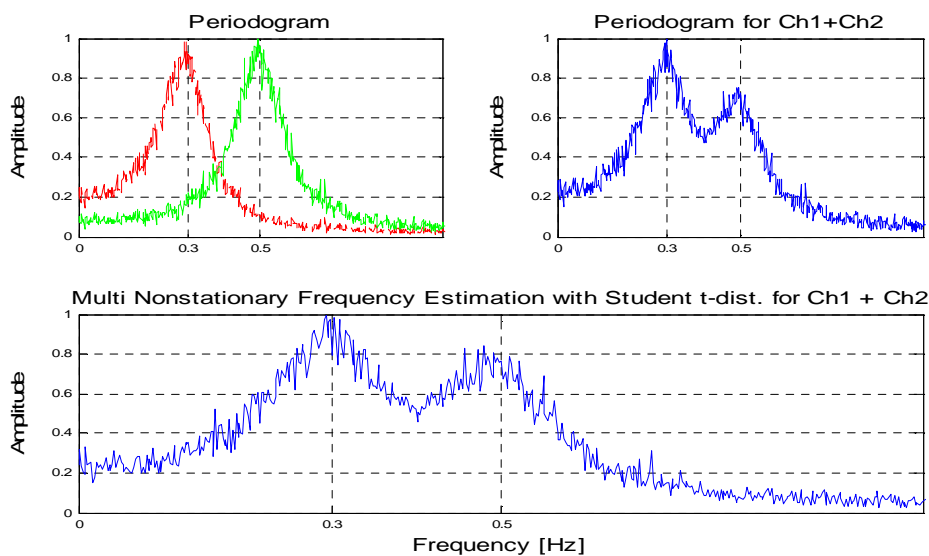


Figure 22: Performance comparison when variance is set to 0.005

It is not surprising to see that the periodogram achieves the same performance as the student t-distribution does. Because we can see the all these estimators yield the same result. This is simply because of the normalization effect on the signals and the axis.

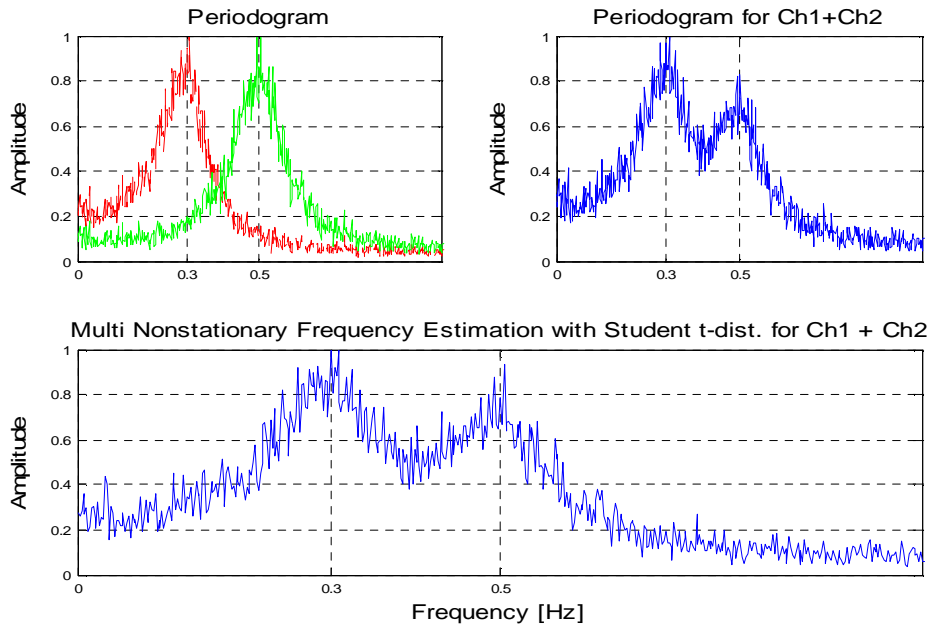


Figure 23: Performance of the periodogram vs Student t-distribution in noisy environment.0.01

The normalization [15] has the following effect on the signals of interest.

- Amplification
- Base line shift
- Stretch or concentration- a scale along the x-or y axis
- Phase shift
- Orientation – a rotation along the axis

Therefore the periodogram has a correct estimator appearance. The result is unsatisfactory although the periodogram has yielded the two correct peaks at the right frequency positions. The estimation is correct due to the effect of the normalization process which changes the signal.

Comment:

The student t-distribution works better on spectral estimation. Further, we have also seen the effect of the normalization, which amplifies noise and shifts the phase and the base line to give another signal. Thus the signals lose their intrinsic shape. In addition, we have shown when more one channel is present; the periodogram is not an appropriate estimator for indication of multi nonstationary frequencies. We have shown that the logarithm of the student t-distribution is a proper statistic estimator which can resolve all the peaks in these channels, while the periodogram fails to do so. We have also seen that prior distribution can have an impact on the estimate although it is vague. Thus the student t-distribution can be used as frequency estimator in a frequency modulator system.

6.2 Classical and Bayesian estimators' noise sensitivity

We have seen the simulated results of the Bayesian method and the periodogram in different context. Now we want to analyze the performance of the Bayesian technique compared to the classical methods. We use for such a purpose a synthetic data to estimate the spectral components of the signal under noiseless and noisy conditions. The difference here is that we focus more on the error sensitivity. Thus we implement the generation of the noisy sequence $y(t)$ and the computation of the frequency estimation. We would note that the spectral estimates of the methods applied here exhibit a significant variability. Therefore, it is necessary to average the noise over several realizations for the sake filtering the noise and stability. We use 10000 realizations in our current experiment. Figures 24 - 25 illustrate the results obtained by running the whole program (Matlab script: *method_sim_rev.m*). We assume that the signal being used in this experiment is uniformly sampled.

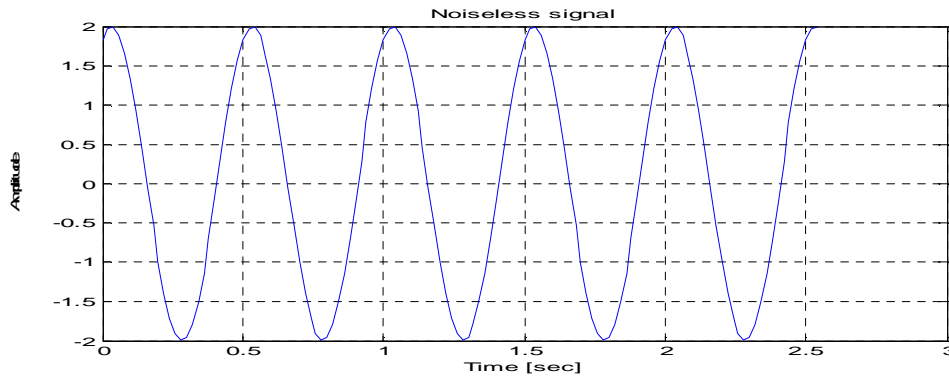


Figure 24: Noiseless sine wave

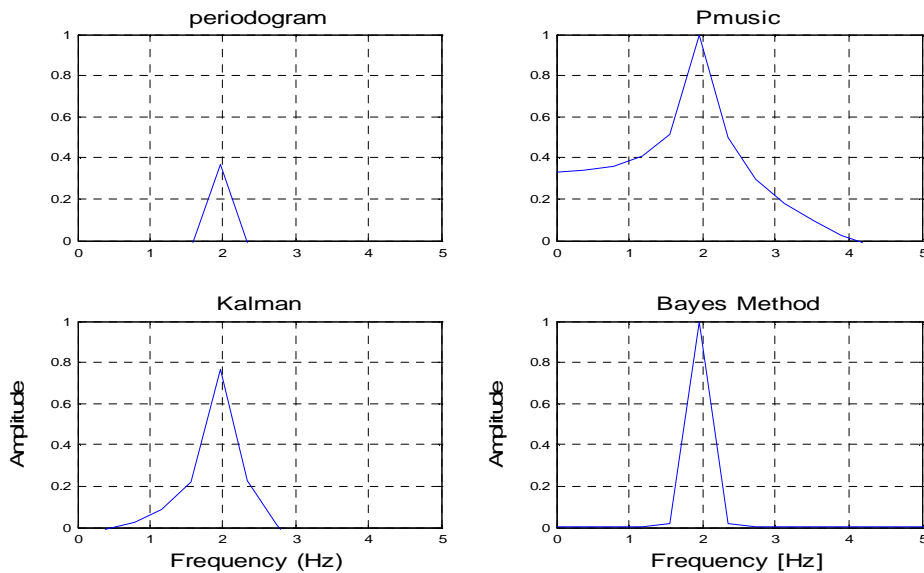


Figure 25: Noiseless spectral components

The Figures 24-25 show the sine wave and the successful spectral estimation of these four estimators except the MUSIC which shows a DC level. Now we add noise and then increase it to a certain level beyond the fading criterion. The results of such an experiment using the same signal are depicted in Figures 26 and 27. We see how the additive Gaussian noise corrupts the signal (see Figure 26). When we apply the same estimators as that of above, all these estimators yield a pronounced peak at the right frequency position.

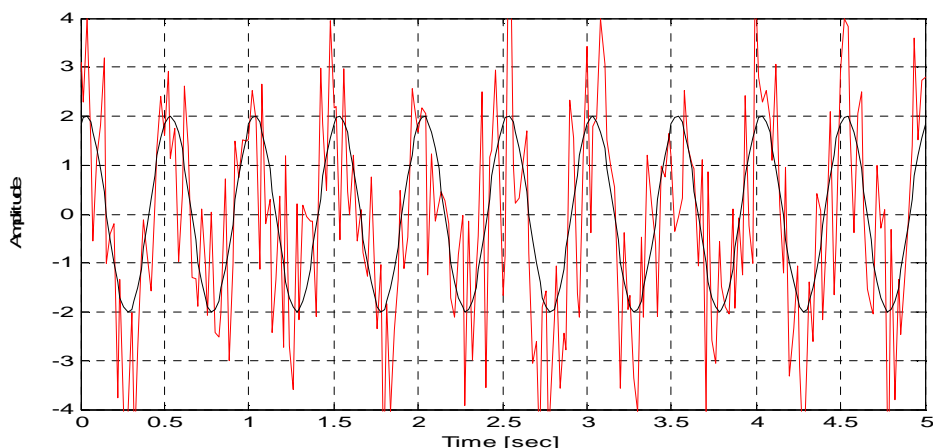


Figure 26: Single tone signal (dark) embedded in white Gaussian noise (red).

By inspection, we see that the Figure 27 shows four frequency spectral components from the periodogram, Music, linear Kalman filter and the Bayesian. However, Kalman and the periodogram introduce spurious peaks. This is a great sign of disturbance; whereas the Music and the Bayesian methods withstand such a noise level. This also demonstrates the power of eigenanalysis-based algorithm for Music and prior for Bayesian.

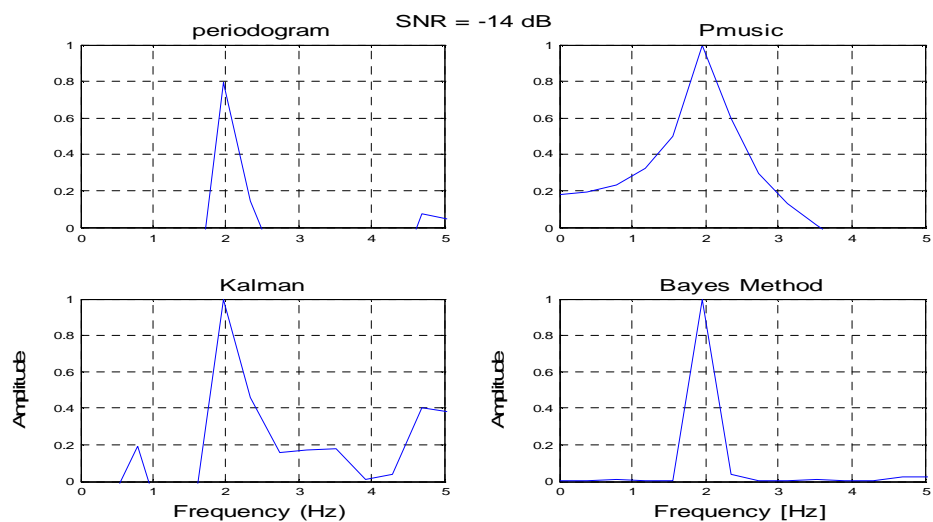


Figure 27: Single frequency spectral form classical and Bayesian estimators and noise effect.

We increase the noise level. At this point, all the estimates are affected. The result tells us that these estimators are significantly deteriorated by the noise as shown in the Figure 28.

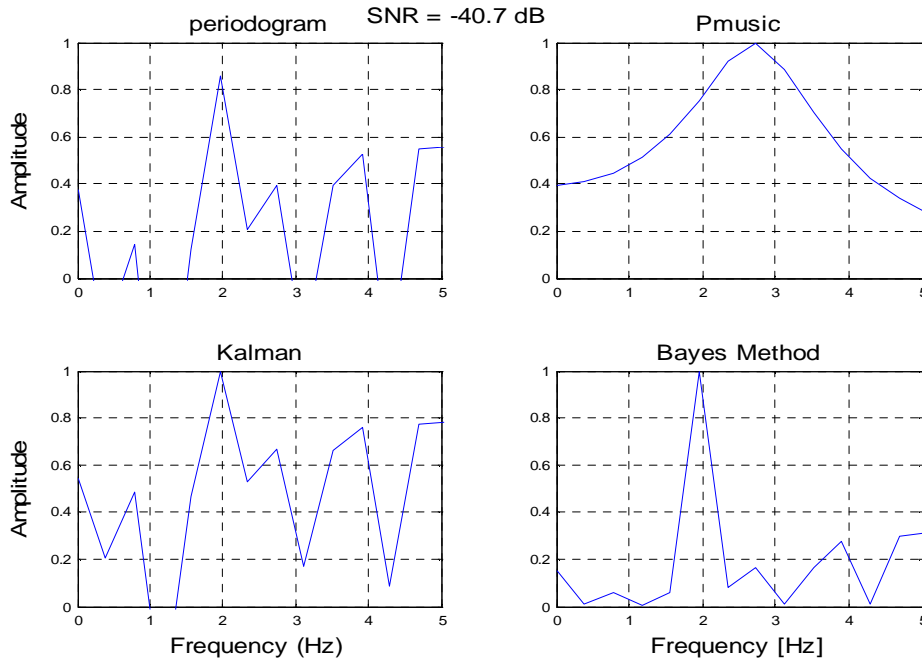


Figure 28: Noise effect on spectral estimators.

Figure 28 shows the effect of low SNR of the estimated spectrum by introducing several peaks. Although the periodogram and the Kalman filter keep the evidence of the peak, the noise has severely deteriorates the estimator by the presence of spurious effects. The Bayesian estimator shows a less impact to the noise due to its low spectral disturbance. The Music fails to estimate the frequency of interest. The key point in the Kalman filter theory is that the underlying state space model is accurate. When this assumption is violated, the performance of the filter can deteriorate appreciably. The filter sensitivity to modelling nonlinear error has led to the development of robust state space filters. Eventhough it is difficult to draw any conclusion, the results nevertheless demonstrate the power of the posterior probability including vague prior in resolving the frequency component in additive noise. Although the results were not satisfactory for the MUSIC, it has been stressed out in literature [18] that the MUSIC are good estimator for sinusoids and can be applied more generally to the estimation of the narrow band signals. Furthermore, the Bayesian technique used in this experiment remains a better estimator. However it must be reinforced by a more robust algorithm including an informative prior with adjustable hyperparameter to be a general purpose estimator. Whereas the linear Kalman assumption and adaptive capability need to be robust against noise.

6.3 Stationary Fundamental frequency tracking

In previous experiments, we study the performance of our estimators with fixed frequency. This analysis extends the ideas developed above under the condition in which the Bayesian algorithm with adjustable parameters is applied to track the frequency variation. We then use a sine wave with fundamental frequency which varies slowly over time. The slow motion of the frequency may be linear and nonlinear. The results of the experiment are shown in the figures below. Thus we consider the following signal and the parameter are listed below.

- Problem statement: linear fundamental frequency tracking

Signal model setup:

1. $ff(t) = F_0 + 0.1t$: Fundamental frequency with a low rate of change.

2. $x(t) = \sin(2\pi \int_0^t ff(t))$, Periodic signal

Parameters:

Record size: 125 samples - Overlap: 100 samples - F0: 5 Hz - Fs: 100 Hz

Signal duration: 60 seconds

P:1 number of regression order

Variance: $\frac{1}{4}$

K: [1] order of the harmonic

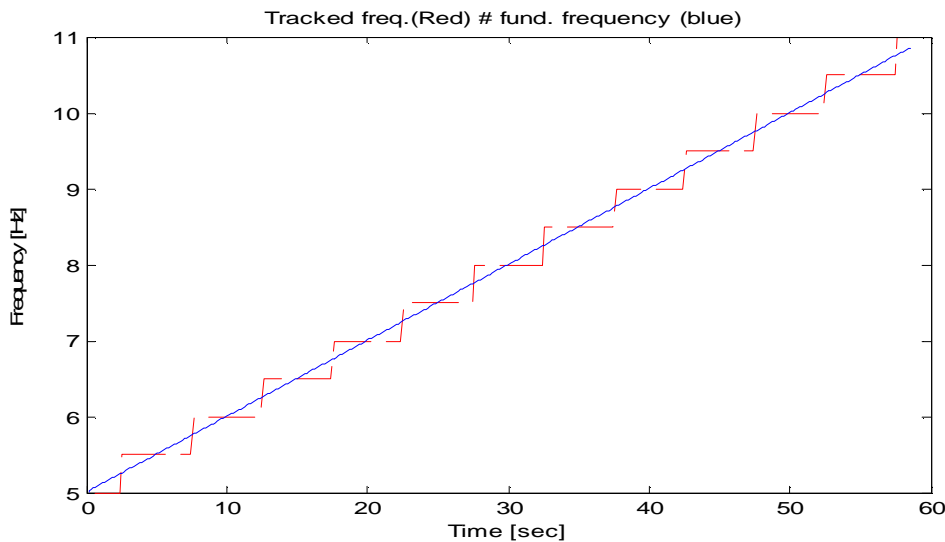


Figure 29a: Fitted linear fundamental frequency track when the signal is noiseless.

As the Bayesian procedure has been described earlier, we will only give interpretations of the results. Thus the Figures 29a and 29b, show the linear fundamental frequency (white line) versus the true

fundamental frequency (blue line). Eventhough the Figure 29b does provide more information; it shows a successful of segmentation and overlap of the data record and the tracked fundamental frequency trajectory followed by the tracker in Figure 29b. In Figure 29a, successful frequency tracking is depicted.

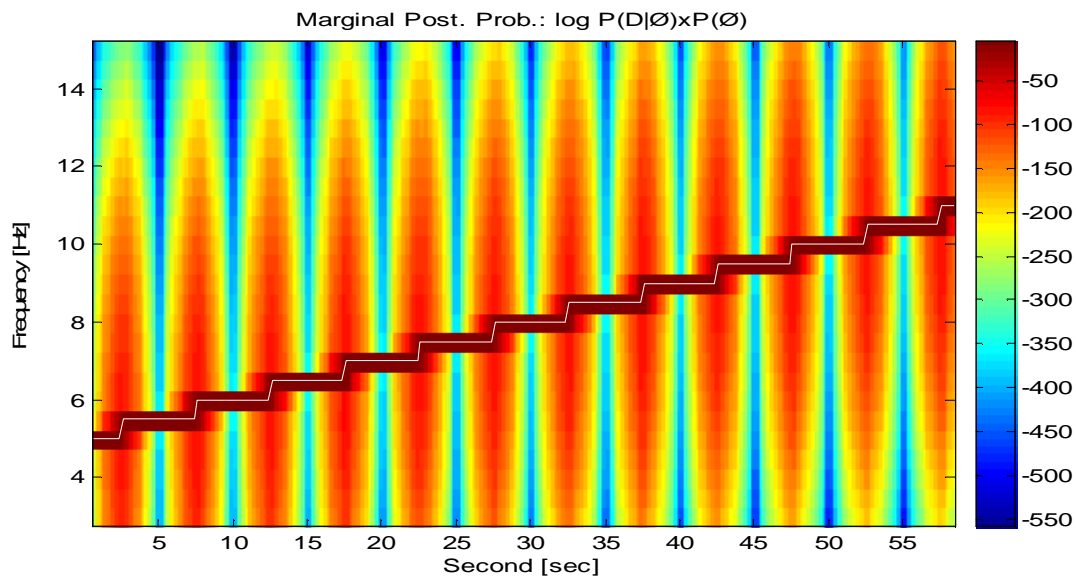


Figure 29b: Image of the linear frequency tracking process of the noiseless signal.

Further, we will now carry out the performance test by adding a white Gaussian noise to the signal and simulate the impact of the decreasing SNR on the Bayesian performance by means of the accuracy and error sensitivity. Let us consider by now that the signal to be tested is as follows: $y(t) = x(t) + n(t)$. A regression model with additive white Gaussian noise with variance set to 1. The result of such a test applying Bayesian is depicted in Figure 30a and 30b.



Figure 30a: Tracking (red) the true fundamental frequency (blue) when the noise variance is set to 1. Despite the noise, the model is fitted well. A negligible degradation is noted in the Figure 30a. At a noise level set to 3, the tracker cannot follow correctly the true fundamental frequency as shown in

Figure 30b. The increase of noise variance has created high uncertainty in the estimates such that it appears difficult to fit the model. This is shown in the Figure 30b, where the fitted curve (red) deviates to follow the trajectory of the detected track (blue).

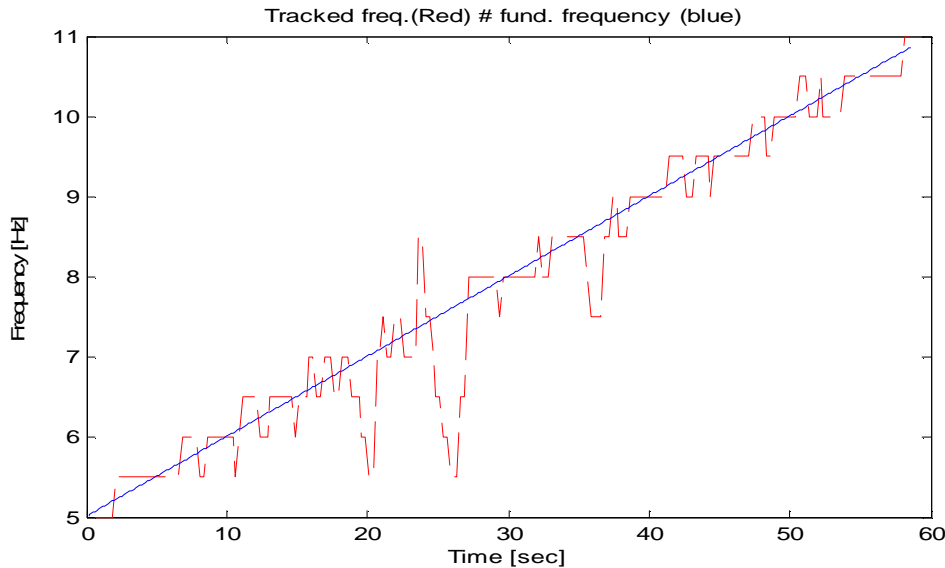


Figure 30b: High degradation of the tracker due to variance set to 3: model cannot be fitted well.

Comment:

In this experiment, we test the performance of our Bayesian algorithm including an informative prior linear time varying frequency signal³. We consider the signal evenly spaced for the first evaluation of the error sensitivity. In the absence of noise, the track and the fitted curves overlap. The model is fitted well. When increase the variance of the noise, the model does not fit well. The effect of noise deteriorates the performance of our Bayesian method. This effect of the noise is that it increases the uncertainty of the parameter to be estimated. Thus confusing the decision making process of the posterior probability by providing wrong and inaccurate estimate to adapt itself later to such a noise level. Moreover the algorithm can yield best result where the model can be fitted well. However under low SNR condition the algorithm fails to fit well the model. Therefore care should be taken to reduce the noise or improve the algorithm. Nevertheless it has shown that our algorithm can drastically deteriorate in low SNR.

³ NB: we must note that all the vertical axes are frequency axe in this experiment of section 7.1.3.

- Problem statement: nonlinear fundamental frequency

Signal model setup:

1. $ff(t) = F_0 + 2.5(1 - \cos(2\pi(0.1)t))$:a fundamental frequency with a low rate of change.

2. $x(t) = \sin(2\pi \int_0^t ff(t))$, periodic signal. Parameters are the same as the above.

When we consider the signal described above, and the search range sets from 5 to 10 Hz. The Bayesian algorithm fitting the model is shown in Figure 33a. The model is well fitted.

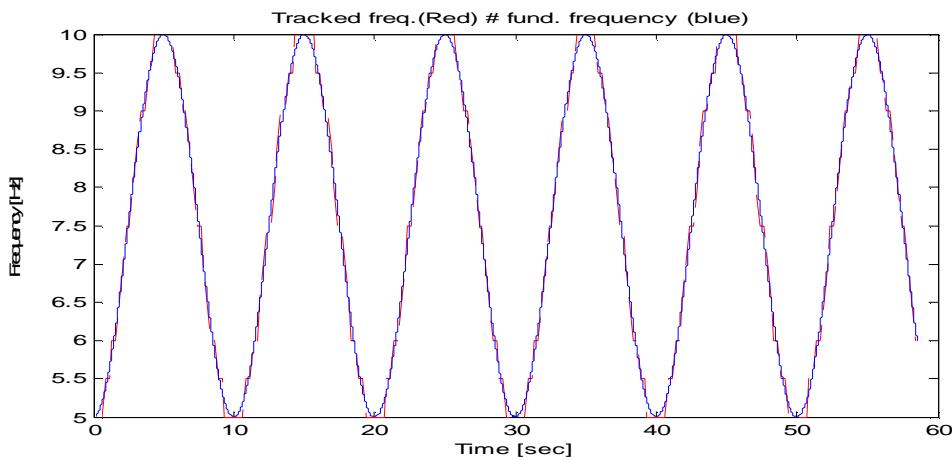


Figure 31a: Satisfactory model fitting

Thus the curve of the tracked fundamental frequency (red) and the true fundamental frequency (blue) overlap quite well (see Figure 31a). Figure 31b shows the posterior distribution of the fundamental frequency with the estimate tracked (white line).

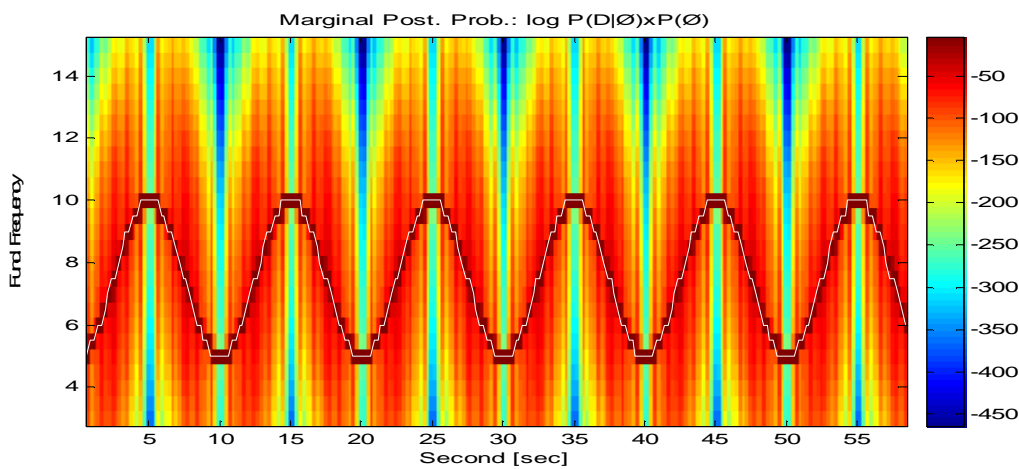


Figure 31b: Tracking fundamental frequency (white line is the fitted track) in log domain.

We add the same Gaussian noise to the signal. The results show that by increasing the noise level, our estimator becomes sensitive to noise.

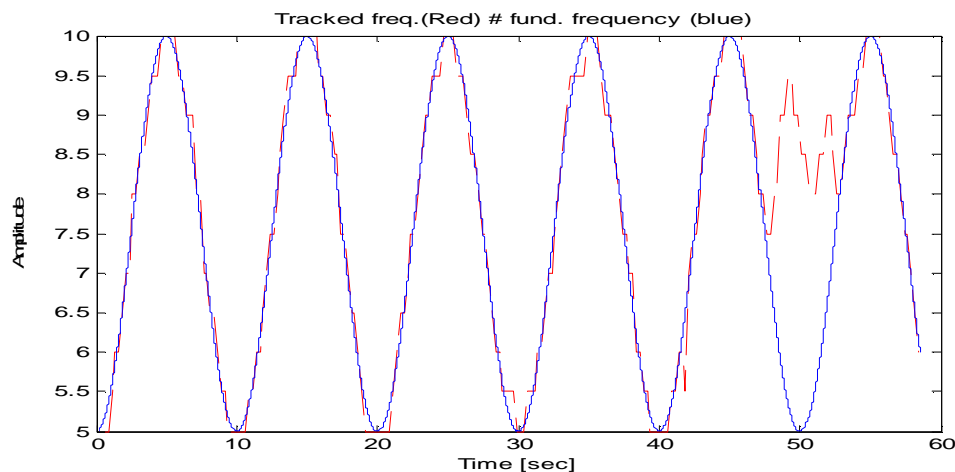


Figure 32a: Measurement (blue) and fitted frequency track (red) when the variance is 2.5.

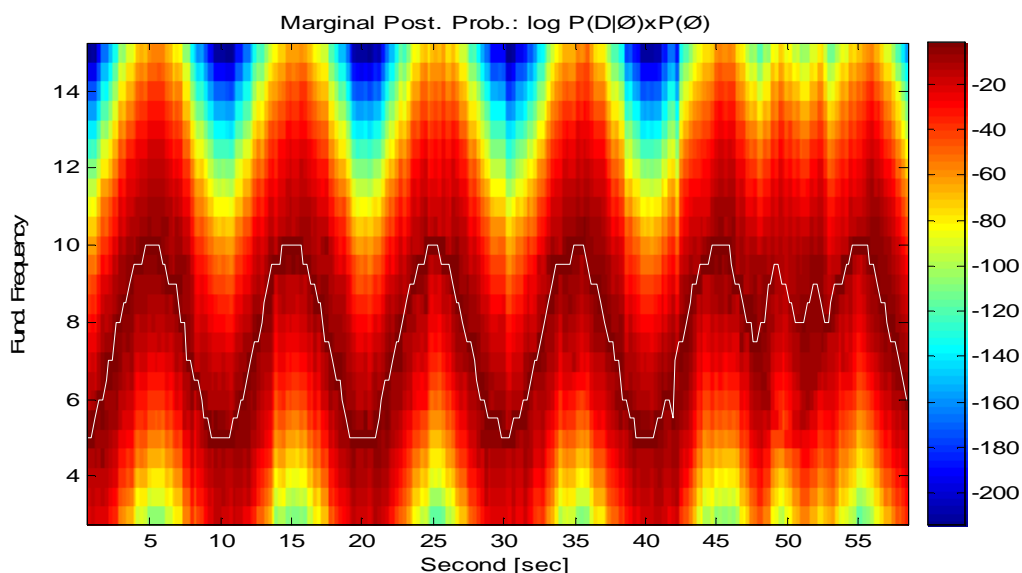


Figure 32b: The posterior of the underfitted fundamental frequency when variance is 2.5.

This implies that the effect the noise disturbs the estimator. Consequently the model cannot be fitted well. The reason is that the estimator cannot withstand such a noise effect. Thus the posterior probability decision yields wrong decision. Hence the estimator yields inaccurate fundamental frequency as shown in Figure 32a -32b.

Now we set the variance noise level to 3. Figure 33b shows the behaviour effect noise on the estimates. The Bayesian algorithm cannot fit the model. As shown the tracker capability deteriorates more and more. Hence the curve represented by the estimate frequencies deviates significantly from the true fundamental frequency trajectory. We have emphasized the performance analysis and the error sensitivity of the Bayesian algorithm when tracking the slowly change of the fundamental frequency. In order to validate the result of the experiment, we first test the signal without noise.

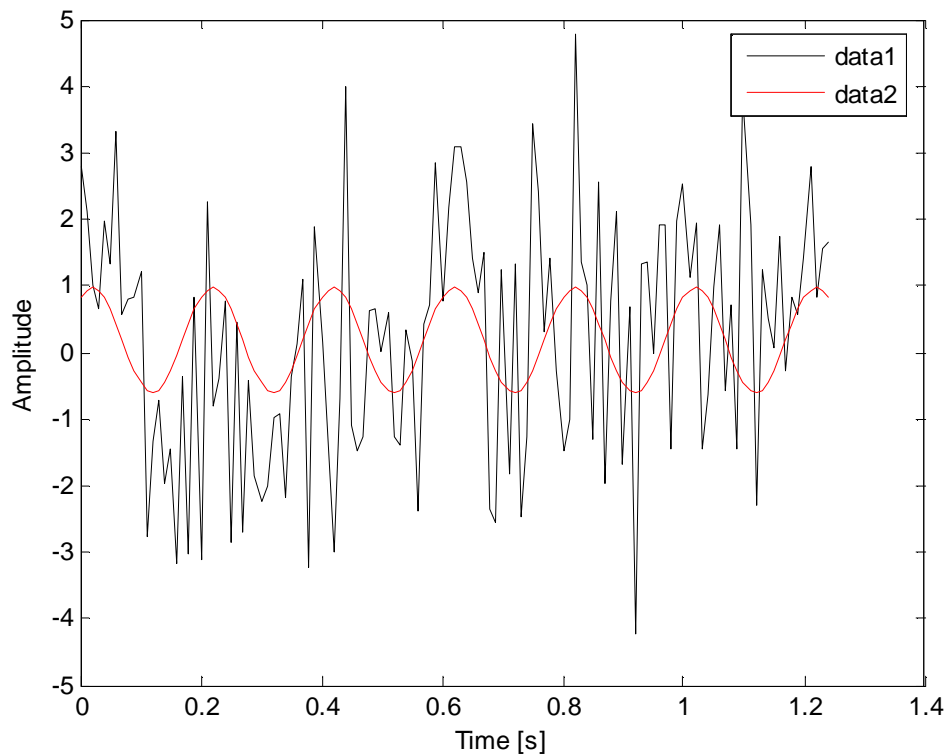


Figure 33a: Tracking fundamental frequency from a signal (data1) in noise (data2).

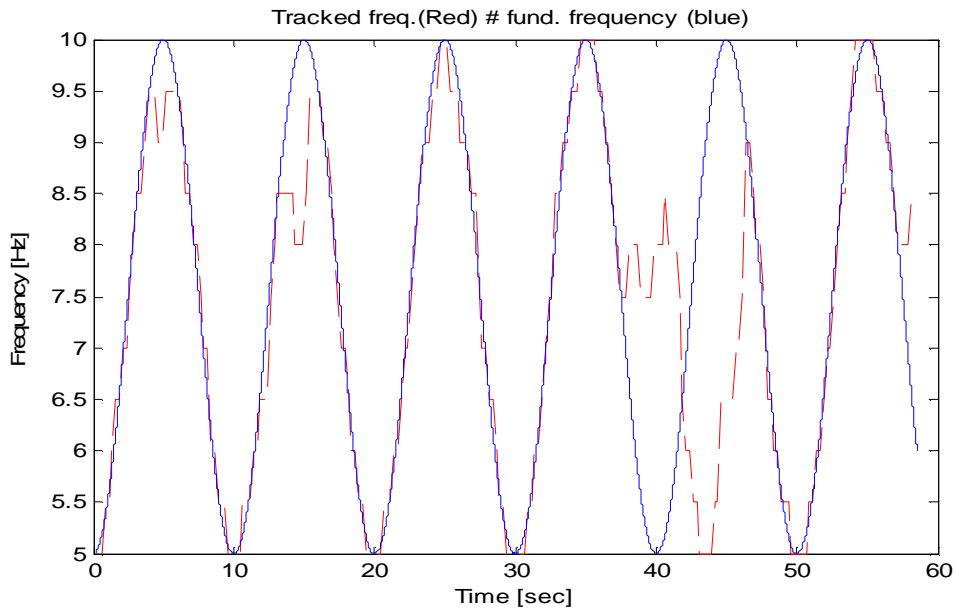


Figure 33b: Performance of the tracker when the noise level is set 3: model cannot be fitted.

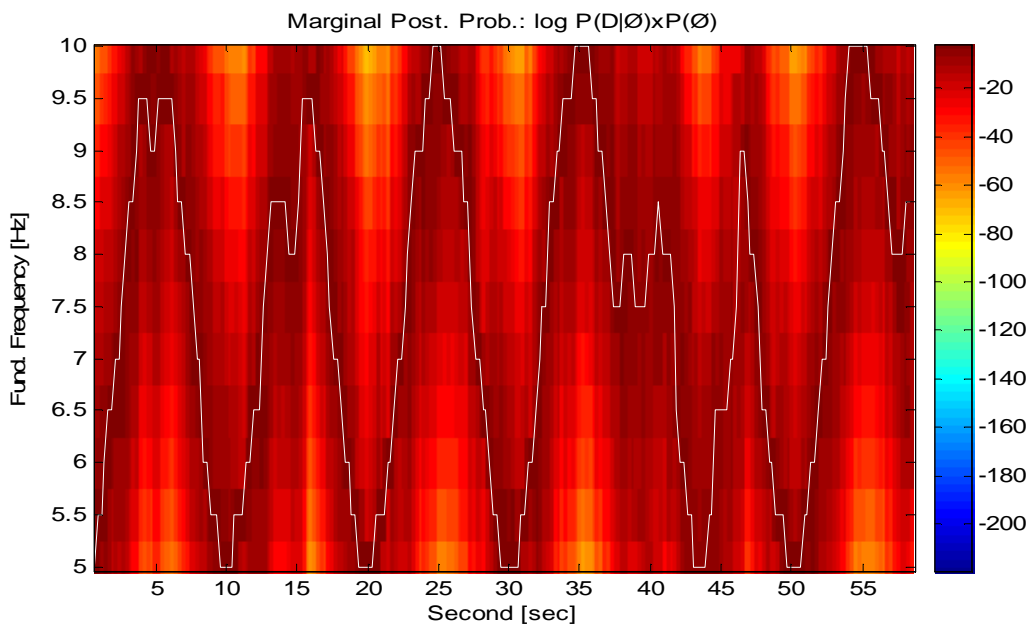


Figure 33c: Posterior probability of the tracked fundamental frequency (white line⁴) when noise level is set to 3.

The Bayesian algorithm can achieve good tracking performance of a stationary the fundamental frequency. In a very low SNR condition the algorithm can suffers from erroneous decision that yield inaccurate estimates. Thus it fails to fit the model.

⁴ White line in these figures is the representation of the tracked fundamental frequency.

6.4 Nonstationary frequency tracking

6.4.1 Bayesian Tracking analysis using vibration signal

This experiment is the results of applying robust Bayesian algorithm to the vibration signal. Note that the parameters are first selected and fixed except the variance. The reason is that we don't know the bound of the variance. Thus the choice of the variance can be time consuming when we need to optimize the accuracy of the estimate. In our case we use the tachometer as reference speed profile to compare the estimate speed profile based on the real data set. Before we go through it,

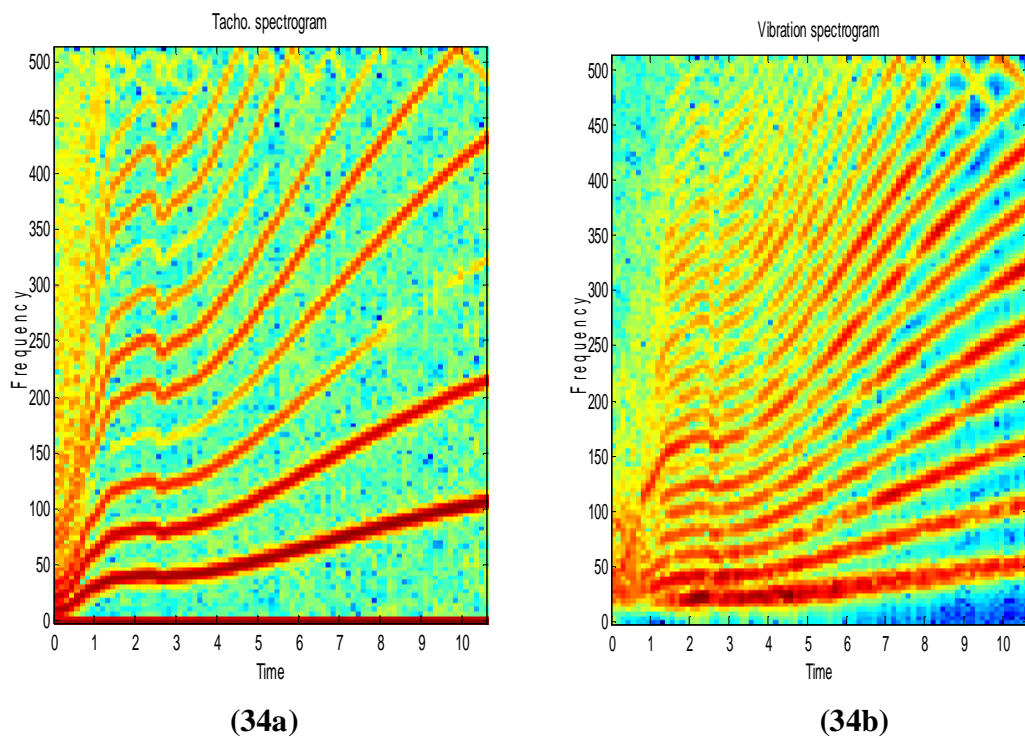


Figure 34: Spectrograms of the tachometer (34a) and the vibration (34b). The spectrogram is the energy in the time-frequency spectrum.

Figure 34 denotes the time-frequency spectrum consisting of several harmonics. These harmonics describe the frequency versus time run-up situation of a car engine. Inspection of Figure 34a gives starts frequency around 10 Hz. It then increases around 40 Hz linearly until 5 seconds at the end (100 Hz). This is the fundamental frequency of the vibration signal. Comparing the tachometer spectrogram with the vibration spectrogram indicates that the harmonic orders in the vibration spectrogram are

multiple of $\frac{1}{2}$ order. Thus we use the order model $K = [1, 1.5, 2]$. This means the first order; the 1.5th order and the 2nd order are selected to be the search region. The other parameters are variance = 0.6, the initial guessed frequency $f_0 = 10$ Hz, the order number $K = 1:2$, the frequency range is set to $[5:0.5:100]$; and the number of the previous record $P = 3$. We apply the Bayesian algorithm again. The

results from the Matlab code: *non_exp_demo.m*, are depicted in the following figures. Figure 35 denotes the effect of a tracking prior with normal distribution. In fact the normal distribution becomes a parabola in log domain. And then tends infinity when moving away from its mean value as shown in Figure 35 (upper panel). When we add the prior the result is shown in the lowest panel in Figure 35.

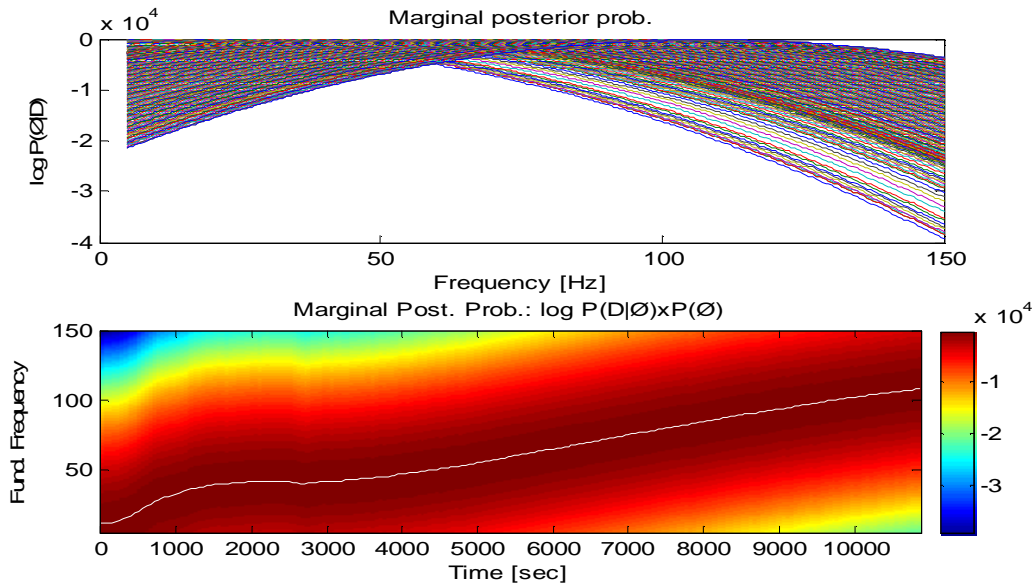


Figure 35: The parabola curve of the posterior probability of the records in log domain (upper panel). And the posterior of the fundamental frequency tracked (white line).

This is the image of the tracked fundamental in log domain. We will see later that this is a correct fundamental frequency estimate (white line) in Figure 48.

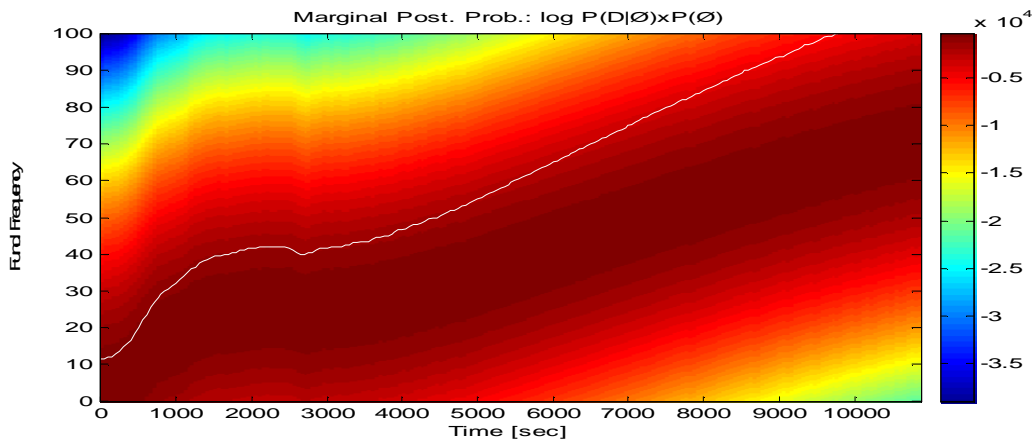


Figure 36: The posterior of the fundamental frequency trajectory (white line).

The Figures 36 describes the MAP results for the run up of all the records of the vibration signal. The algorithm has been able to handle the computation need for drawing inference about the fundamental frequency estimate (white line). We give an illustration in time domain comparison to show how

accurate the algorithm yields the model parameters of interest. Therefore we plot the noisy observations against the tacho (green pulses). As we can see, in the upper panel of the Figure 37, the pulses rise at the start of each vibration signal period by a close look.

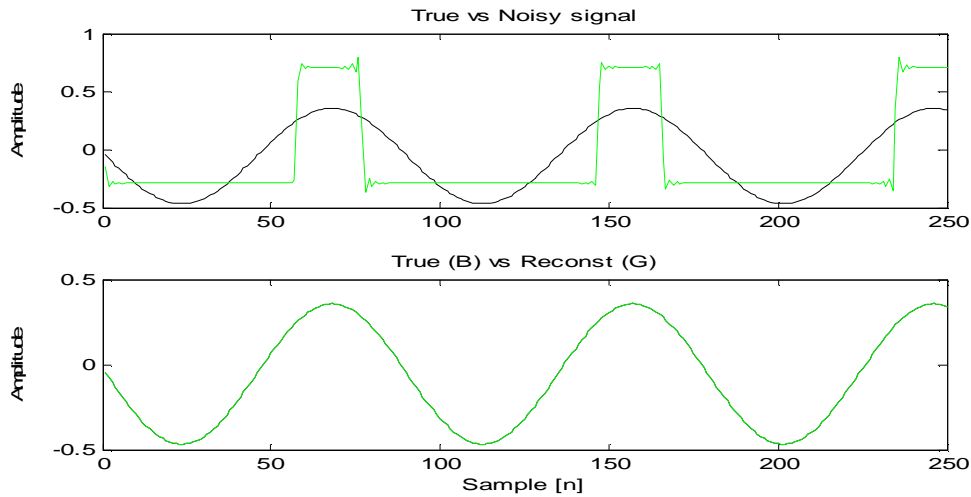


Figure 37: Signal comparison (lower panel) and period matching (upper panel)

Further we compare the true signal with the reconstructed signal. We see that these two signals match each other. This comparison can also tell us that the tracking has been successfully done. However the result is not perfect but satisfactory because the reference tacho speed profile (red in Figure 50) shows a strange discrepancy due may be to our algorithm (does not start at zero on the y-axis).

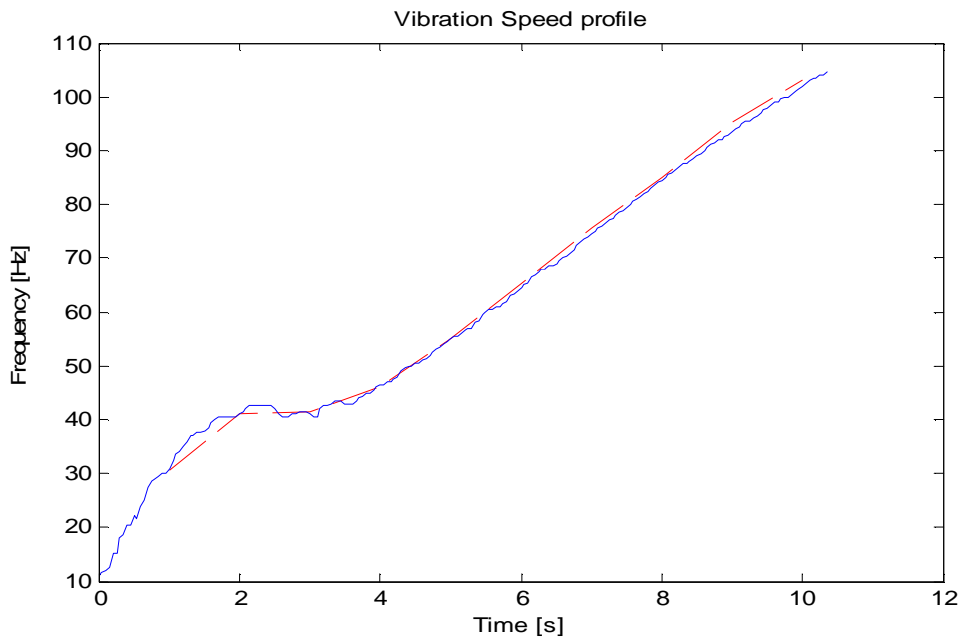


Figure 38: Speed profile from tacho (red) and vibration (blue) signals. The two speed profiles follow each other. This tells us that the tracking has been successful. The model is more or less fitted.

6.4.2 Hyperparameter effects

If tracking is shown to be successful in one hand, parameter adjustment has been creating instability in the shape of the estimate. One of the difficulties here has been to determine the optimal parameters. That is the parameter which can yield the “best estimate “. This is because there is no clear bound for the parameter. It is vague to consider that the parameter space is defined only from on zero to infinity. This makes the work time consuming. Because adjusting the parameter, specifically, it is referred to manipulate the shape of the prior (width by variance adjustment) and the parameter location (by the mean through the number of previous record P). However when the “true parameters” have been found, the algorithm can handle well the fundamental frequency tracking. The variance and the number of the previous record (used by the mean) are the governing parameters. Thus the prior shows its influence through these parameters. The wrong choice of these parameters yields inaccurate estimates. We will demonstrate this influence of these parameters below when we use the sound signal.

The simulation has the same scenario with the vibration one. The only is that we test the impact of the wrong adjustment on the estimate which has not done in the vibration side. The reason is that the sound signal represents both run up and coast down. Therefore doing the experiment on one will give a result for both at once. As before, we setup the parameters. And then we apply our new algorithm based on robust Bayesian method. The results are described in the Figures below.

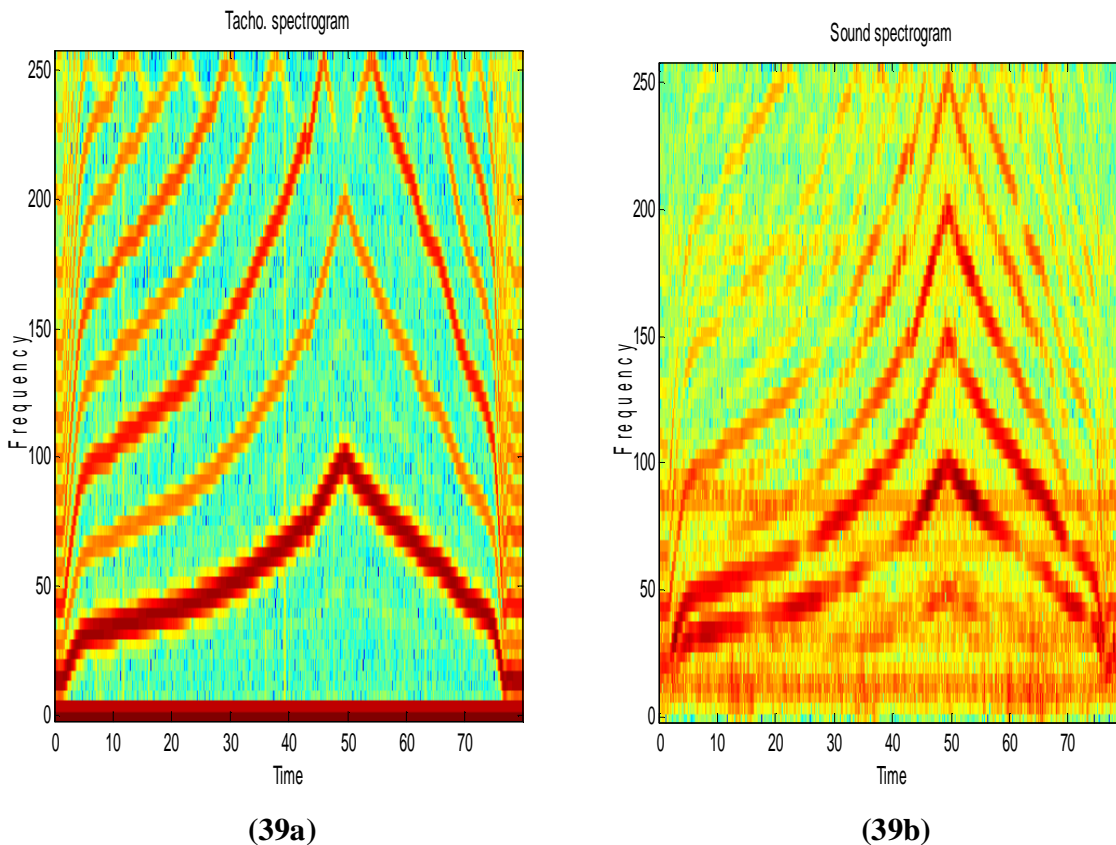


Figure: 39: Spectrogram of the tacho (39a) and sound (39b) signal.

Figure 39 shows the spectrogram of the sound and the tacho signal. A closer look at these spectrograms shows that in tacho spectrogram, the first harmonic starts around 20 Hz. It then increases to around 100 Hz where it stays for 2.4 sec, where after it decreases almost linearly to around 10 Hz until the end. When we compare the tacho spectrogram with the sound spectrogram we observe that the harmonics orders in the sound (acoustic signal) are multiple of the $\frac{1}{2}^{th}$ as the vibration one. In this way, we select the model order to be $K = [1, 1.5, 2]$.

Figure 40 describes the result of the marginal posterior probability distribution in log domain. We can see the fundamental frequency which has been tracked correctly (see Figure 41).

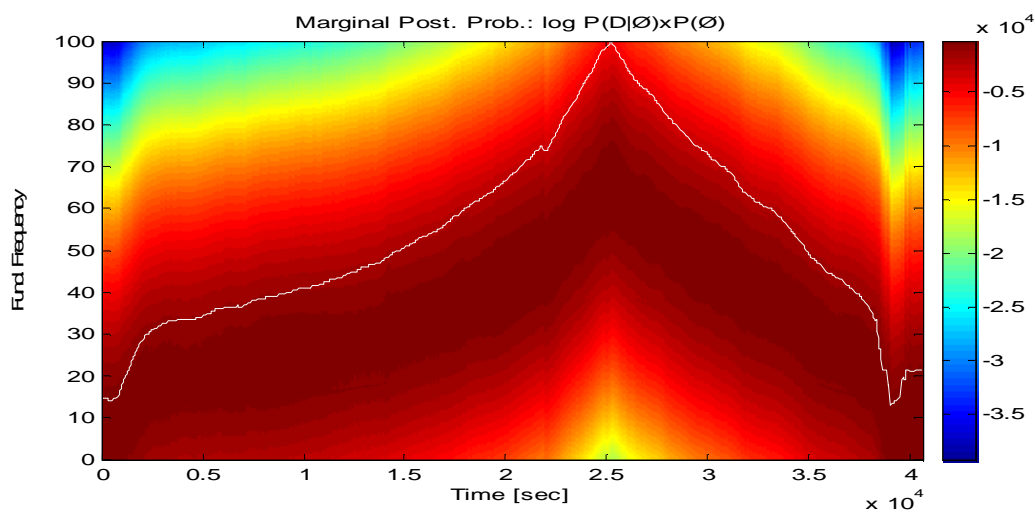


Figure 40: Tracking successfully with the prior the fundamental frequency estimated (white line).

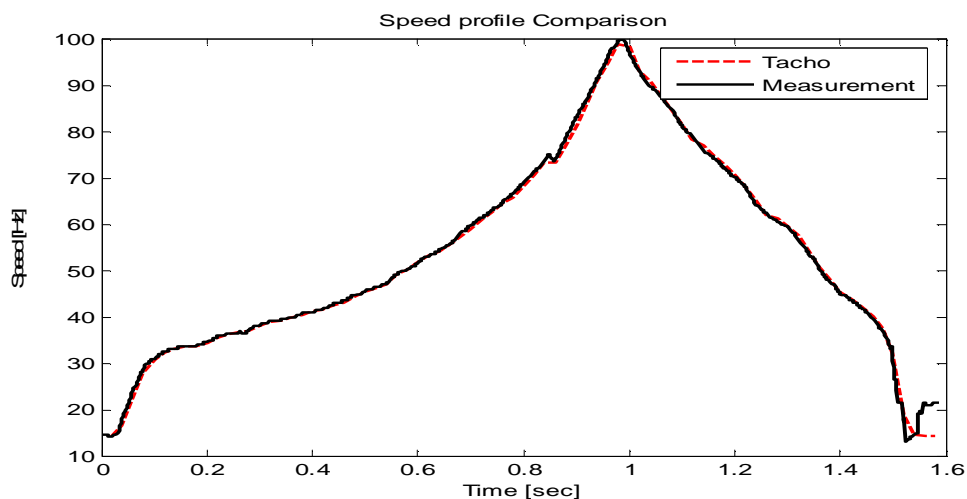


Figure 41: Speed profile estimated (measurement) overlapping the reference (tacho): model is fitted.

Figure 41 shows the true speed profile and its corresponding estimate determined by applying the robust Bayesian algorithm. The estimate speed profile (measurement) is virtually identical to the exact speed profile (tacho). The result tells us that the parameters fit well the data model. This is because the estimates speed profile is in good agreement with the true speed profile. These two speed profiles describe the run up and run down situation of a car engine. Hence we see that tracking has been achieved successfully. The algorithm has been well capable to track the precise fundamental frequency. However the task has not been so easy because of the adjustment of the parameters time consuming. Alternatively, we can also compare the true and the reconstructed signals. And then the error is computed. The results appear in Figures (42-43).

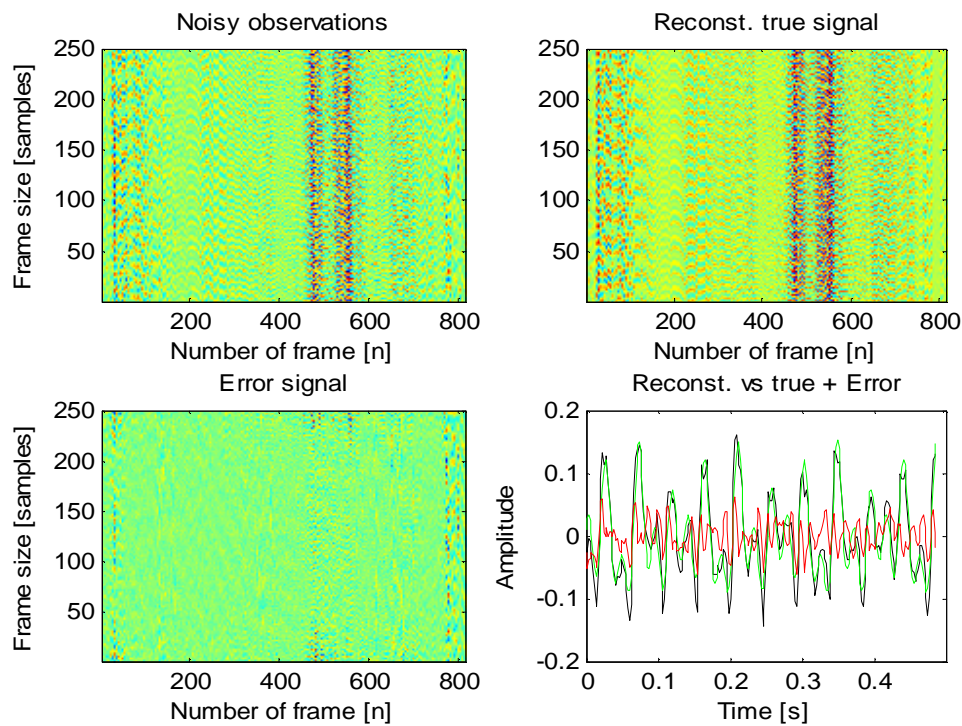


Figure 42: Image of the signals and the error.

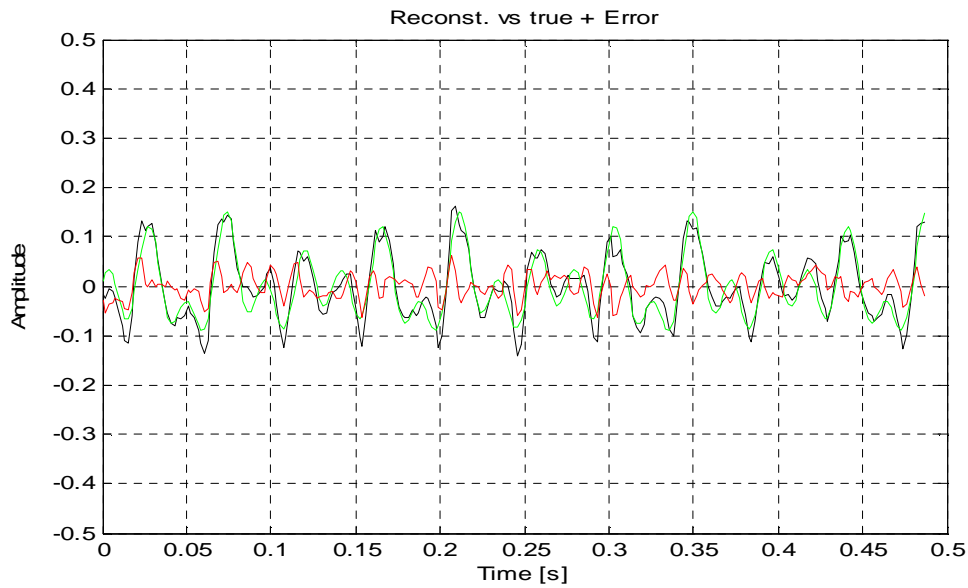


Figure 43: The reconstructed and true signal plus the reconstructed error.

Although the information from this comparison may not be objective, it gives quite good impression of the reliability and the robustness of the Bayesian algorithm by looking at Figures 43. The result was shown to be successful.

Now, we are concerned with the behaviour of the algorithm while adjusting the parameters of interest. We will be using the variance, the number of the record (includes in the mean) and may be the number of order to test their effect. The results when we did not adjust correctly the parameter of the Bayesian algorithm is shown in Figure 44.

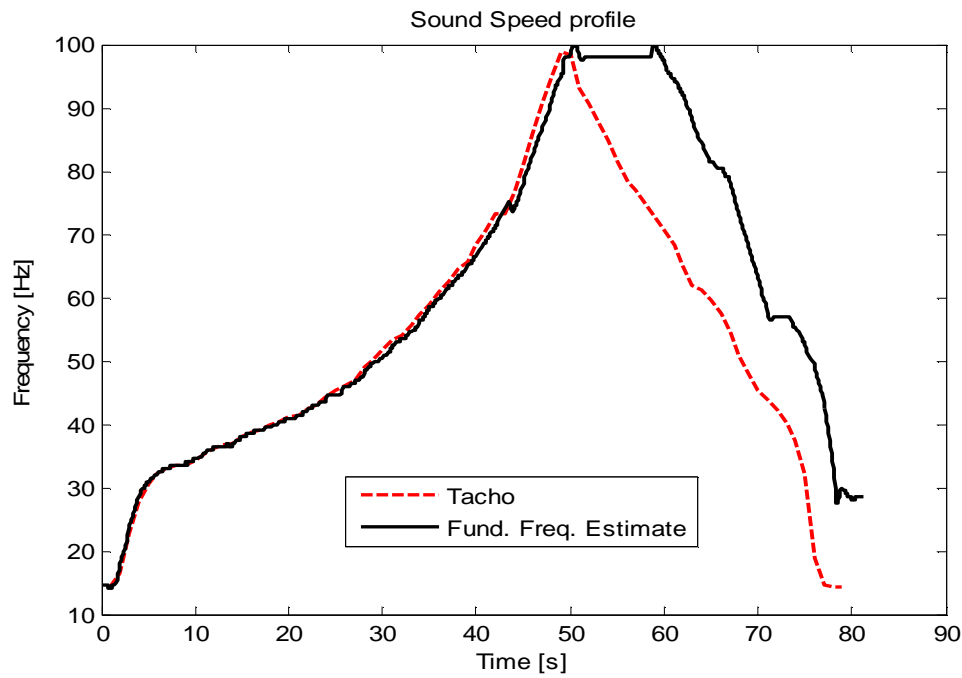


Figure 44: Speed profile being controlled by adjustable parameters. $K = [1.5 \ 2]$, $\text{var} = 1/4$, $P=3$. In this case the model is not fitted.

As we can see from Figure 44, when we change the order K parameter value, the algorithm tracks the run up and deviates to follow the run down. This tells us that the order parameter controls the search region of the fundamental frequency (see Figure 44). This is also true, because it is the order K which

allows tracking the right fundamental frequency. Hence the search region depends on the parameter K .

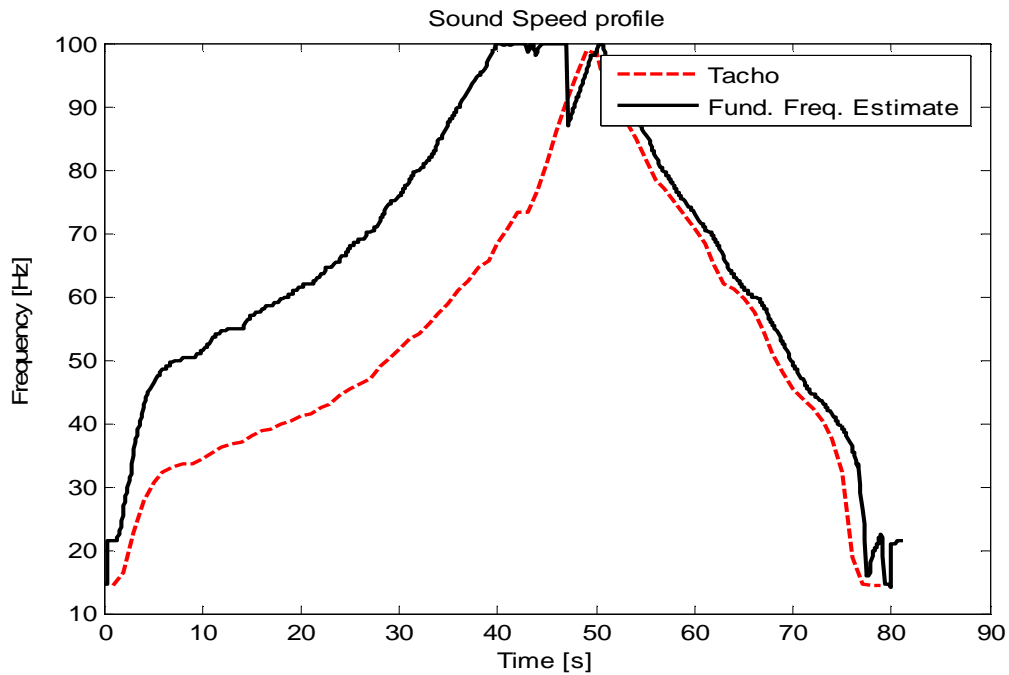


Figure 45: Speed profile being controlled by adjustable parameters. $K=[1.5 \ 2]$; $\text{var} = 0.3$, $P=3$. The model is not fitted because the parameters are optimized.

We now fix the other parameter and then change the variance value, shape of the speed profile changes as shown in Figure 45. The tracker cannot follow the run down properly. This change has a harmful impact on the performance of our robust algorithm. This is also expected because the variance controls the width of the prior distribution which is very important for the posterior probability to draw inference about parameters to be estimated. We have stated earlier that the prior probability distribution is a Gaussian bell-shaped curve. And the standard deviation (square root of the variance) controls the width of the prior distribution. Any change of variance value will imply changes in the prior shape. Consequently, the change in the prior shape will influence the posterior probability decision. The model won't be fitted well with such parameters. Furthermore the deterioration of the performance can result as shown in the Appendix C.

Chapter 7

General Conclusion

In this thesis we have investigated the classical spectral and Bayesian tracking analysis. The performance analysis of the overall estimators involved in this work is emphasized through the experiment simulations. The investigation and analysis works are described through xxx fundamental and complementary processes:

1. Basic statistics and probability theory
2. Estimation methods pros. And cons
3. Spectral analysis methodologies
 - Periodogram
 - MUSIC
 - Linear Kalman filter
 - Pisarenko
4. Bayesian analysis for linear regression models
 - Maximum likelihood for regression
 - Likelihood procedure for low SNR, too closed frequency and low frequency estimation
 - Vague and conjugate prior introduction
 - Bayesian parameter estimation-case study
 - Bayesian tracking analysis using vibration and acoustic signals
5. Performance analysis using stationary time series plus white Gaussian noise
 - Single harmonic frequency estimation
 - Two harmonic frequency estimation
 - Multi-stationary harmonic frequency estimation
 - Multiple nonstationary harmonic frequencies estimation
6. Comparison of low SNR effect on both classical and Bayesian estimates
7. Slowly time varying fundamental frequency tracking using noisy time series
8. Robust Bayesian tracking analysis and procedure proposal

We have established a relation between theory and engineering technical software application in a broad field of Classical spectral and Bayesian tracking analysis in rotating mechanical system. In order to understand and implement the statistical approach to the fundamental frequency tracking problem using vibration and acoustic data, we have simplified the random parameter estimation problem at stationary noisy time series level in accordance with my supervisor at DTU. We have given a survey of Bayesian analysis for linear regression models, provided a possibility of understanding the Bayesian parameter estimation technique, comparing the performance of both classical and Bayesian and analysing the error sensitivity and the effect of the hyperparameter on the estimates through computer simulations experiments. We have found that for single harmonic frequency estimation provided it is not too closed to zero, the periodogram performs well. Although the periodogram can estimate multi-

stationary harmonic frequencies in the presence of Gaussian noise, the log Student t-distribution yields better estimates. For two closed stationary harmonic frequencies with short data size, we have reported that the introduction of uninformative prior has an effect to emphasize the evidence of these frequencies although uncorrected.

We have given some basic methods and the summary of some previous estimators which are used in both off line and on-line frequency estimation to. By doing so, we have been able understand the strength and the accuracy in function of the Cramer-Rao-Bound (CRB) of these frequency estimators. From the summary it has been shown that only maximum likelihood, the periodogram, Fernandes-Goodwin-de-Souza and Quin-Fernandes asymptotically achieves Cramer-Rao-Bound. That is, these can be used to provide good estimates in the application of interest.

Bayesian parameter estimation technique for linear regression models has been investigated. It been derived that the posterior probability distribution is proportional to the product of the likelihood function and the prior. Our focus has been on how to determine the hyperparameters of the prior distribution in parameters estimation problem. It has been found that for optimal determination of these hyperparameters, we could use empirical Bayes, type 2- maximum likelihood, general maximum likelihood or evidence approximation. Further if the prior is flat, the evidence is obtained by maximizing the likelihood function. If we define conjugate (Gamma) prior distribution over the hyperparameters, then the marginalization over these hyperparameters can be performed analytically to give student t-distribution. Alternatively the expectation maximization (EM) algorithm provides practical evidence framework if the integral is no longer analytically tractable. It is relevant to mention that there other method which can be used such as Monte Carlo simulation or importance sampling (see section 6.4 in Bayesian Method, 2005). These estimators can yield good results at the expense of high complexity.

Time constraint for the sake of efficiency requires that simple algorithms are preferable and some trade-off between algorithms complexity, accuracy, delay and quality must be made to select the desired estimator scheme.

For the sake of accuracy, comparison and reliability in fundamental frequency estimation, we have considered to perform spectral analysis of classical and Bayesian methods. Therefore we have simulated six experiments using sinusoidal discrete time series added to white Gaussian noise. Since sinusoids plus additive white Gaussian noise describes well stationary signal, we have simulated single stationary harmonic frequency estimation, multi-stationary harmonic frequencies estimation and nonstationary harmonic frequency estimation. The results of these experiments have proved that although, the periodogram achieved a better performance when frequencies are separated, it introduces spurious peaks and deteriorates significantly as the SNR becomes small. The linear Kalman filter can yield good performance in high SNR. It is a best estimator when the signal and noise are non-Gaussian. The performance of Kalman filter is not optimal in the presence of Gaussian noise. It has also been found the MUSIC algorithm achieves good performance but it cannot ensure Cramer-Rao-Bound. All these classical estimators, despite these efforts to perform well sometimes, the posterior probability including prior knowledge outperforms all of these. This is due the power of the prior to yield correct. We have seen also that the prior has an impact on the posterior distribution. Therefore if the prior is

vague, the posterior results become conservative. However the estimates or results from the posterior probability distribution are corrected if the posterior distribution is based on informative prior. These experiments have been simulated successfully. In addition we have simulated the error sensitivity of the Bayesian method. It has been found that Bayesian method shows an undesirable effect and moreover, it yields bad performance. This behaviour is comprehensible because it is beyond the fading limit or the normal experimental limit. Furthermore, we have simulated the effect of the adjustable hyperparameters of the prior distribution on tracking the fundamental frequency. It has been shown that when these hyperparameters are not well adjusted, wrong estimates can be yielded out by the robust Bayesian scheme. If the hyperparameters are setup correctly, the Bayesian achieved successfully correct results. Although the robust Bayesian remains the reference in our case for tracking speed profile, it is sensitive to noise. It is very simple and provides good quality and high accuracy despite the noisy nonstationary signals of interest.

The main problem about the robust Bayesian algorithm implementation is the choice of the optimal hyperparameters to accurately create the reliability condition in tracking speed profile. We have found, through our simulations, a bound for the variance and the way of setting up the number of order to avoid a long time consuming. Hence we have found that the variance can be found setup between an interval of [0.1 0.6] and the number of order to track depending of the real application, we have in our case found that it may be assigned to [1 1.5 2], which means the 1 for the first order, 1.5 for slight shift of the first order frequency due may be to the nonlinear effect of the system. Therefore the region of tracking of the fundamental frequency in such a condition will take into account both first frequency, the slight shift first frequency and the second order of the harmonic which is designated by 2. We consider 1.5th order as the fundamental frequency and the frequency range is fixed and of course known.

We have found that these hyperparameter control the behaviour of the prior. Specially, the variance controls the width of the prior distribution. Moreover, the adjustment of the hyperparameter offers more flexibility to the Bayesian algorithm to adapt itself to any type of parameter estimation problem. We have little prior is available, the posterior estimates reduces to the maximum likelihood estimates.

The principle of least square or maximum likelihood provides no way to eliminate nuisance parameters, and thus oblige to seek a global maximum in a space of much high dimensionality, which requires an heavy computation burden. Having found that, they only provide the sampling distribution in a longer calculation which does not answer the question of interest. Thus they cannot assess the accuracy of the estimates.

We have also found that although the vibration tacho speed profile was successfully achieved, however its representation by tacho suffers from my code deficiency to yield a correct size of the speed profile. In other side, the Bayesian method achieves successfully the tracking process for both vibration and acoustics nonstationary signals.

The future works to improve the robust Bayesian method are:

- Robbin –Monro method to estimate the stochastic location parameter in nonstationary data.
- Improvement Bayesian algorithm using robust Kalman filtering or Particle filtering

Appendix

A Review materiel for Bayesian linear regression

Bayesian Analysis for Linear Regression Models

A.1 Bayesian parameter estimation

A.1.1 Linear model for regression

Linear regression model is a mathematic method to model the relationship between the dependent variables and independent variables. The general linear regression model

$$y(X, W) = \sum_{j=0}^{M-1} w_j \phi_j(x) = W^T \Phi(X) \tag{Eq72}$$

Where $\Phi = (\phi_1, \dots, \phi_M)^T$ and $W = (w_1, \dots, w_M)^T$

A simple model equation is represented in Figure 48. The figure shows the linear regression model (a straight line governed by $y = w_0 + w_1 x$) and data points.

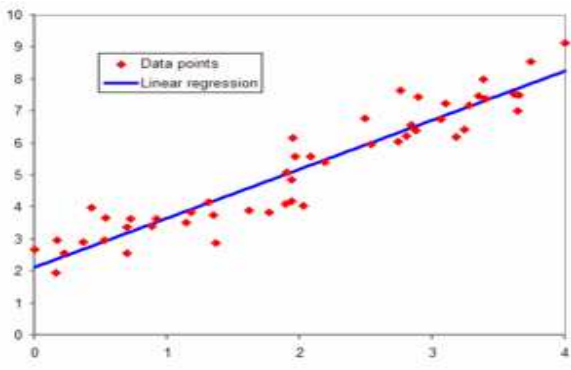


Figure 46: linear regression and data point plot of y versus input x.

Much of our discussion in this section will be applicable to situation in which the vector $\Phi(X)$ of basis functions is simply the identity $\Phi(X) = X$. Further, we will derive the maximum likelihood and Bayesian treatment of linear regression model and explain how to determine the hyperparameters of the prior distribution.

A.1.2 Maximum likelihood for regression

We have seen several times that the maximization of the likelihood function under conditional Gaussian noise distribution for linear model is equivalent to minimizing the sum square error function. Before we derive such an error function, let us re-establish the equation. This will be repeated even though there may similar formula above for the purpose of conformity between variables. We may assume that the target variable is defined by the deterministic function with a Gaussian noise as follows

$$t = y(X, W) + \varepsilon \quad \text{Eq73}$$

where ε is zero mean Gaussian random variable with precision (inverse variance) β . Thus the likelihood function is

$$p(t | X, W, \beta) = N(t | y(W, X), \beta^{-1}). \quad \text{Eq74}$$

Making the assumption that these data point are drawn independently from the distribution Eq63 we obtain the following likelihood expression

$$p(T | X, W, \beta) = \prod_{n=1}^N N(t_n | W^T \Phi(X_n), \beta^{-1}). \quad \text{Eq75}$$

Because in this supervised learning problem such as regression, we are not seeking to model the distribution of the input variables, therefore we drop the input variable form now to keep the notation uncluttered. $p(T | X, W, \beta) = p(T | W, \beta)$. Taking the logarithm of the likelihood function and making use of the standard form (1.46) for multivariate Gaussian, we have

$$\begin{aligned}\ln p(T | W, \beta) &= \sum_{n=1}^N \ln N(t_n | W^T \phi(x_n), \beta^{-1}) \\ &= \frac{N}{2} \ln \beta - \frac{N}{2} \ln(2\pi) - \beta E_D(w)\end{aligned}\tag{Eq76}$$

This is called the maximum likelihood. Where the sum of the square error function is defined by

$$E(w) = \frac{1}{2} \sum_{n=1}^N \{t_n - W^T \phi(x_n)\}^2\tag{Eq77}$$

$$W_{ML} = (\Phi^T \Phi)^{-1} \Phi^T t \quad \text{and} \quad \frac{1}{\beta_{ML}} = \frac{1}{N} \sum_{n=1}^N \{t_n - W_{ML}^T \phi(x_n)\}^2$$

In practice we are not interested in finding the value of w itself but rather making a prediction of t for new values of x . This requires that we integrate over the parameter w . This is called marginalization. We thus evaluate the predictive distribution defined by

$$p(t | T, \alpha, \beta) = \int p(t | W, \beta) p(W | T, \alpha, \beta) dW\tag{Eq78}$$

in which T is the vector of the target values from the training set. The result is as follows

$$p(t | X, T, \alpha, \beta) = N(t | m_N^T \phi(x), \sigma_N^2(x))\tag{Eq79}$$

where the predictive variance is given by

$$\sigma_N^T(x) = \frac{1}{\beta} + \phi(x)^T S_N \phi(x).\tag{Eq80}$$

The first term represents the noise of the data whereas the second term is the uncertainty associated with parameters w . The conditional distribution for $p(t | X, W, \beta)$ of the target variables is given in Eq5 without X and the posterior weight distribution is given by

$$p(W | T) = N(m_N, S_N)\tag{Eq81}$$

where

$$m_N = S_N (S_0^{-1} m_0 + \beta \Phi^T T)\tag{Eq82}$$

$$S_N^{-1} = S_0^{-1} + \beta \Phi^T \Phi\tag{Eq83}$$

If we consider a broader prior $S_0 = \alpha^{-1} I$ with $\alpha \rightarrow 0$, the mean m_N of the posterior distribution reduces to the maximum likelihood value given by ((3.15) in Pattern recognition for Machine Learning – C. M. Bishop, 2006). Similarly, if $N=0$, then the posterior reverts to the prior. Furthermore, if the data arrive sequentially, the posterior distribution at any stage acts as the prior distribution for the subsequent data point., such that new posterior distribution is again given by Eq70.

A.1.3 Evidence approximation

In Bayesian treatment of the linear basis function, although we can integrate over the hyperparameters w , the complete marginalization over these variables is analytically intractable. We will adopt here the popular approximation method of determining the hyperparameters. This can be achieved by maximizing the marginal likelihood function obtained by first integrating over the parameter w . This is known in the literature as *empirical Bayes* (Bernardo and Smith, 19994; Gelman et al., 2004), or *type 2 maximum likelihood* (Berger, 1985), or *generalized maximum likelihood* (Wahba, 1975), and in the machine learning literature is also called *evidence approximation* (Gull, 1999; MacKay, 1992a). If we now introduce hyperpriors over α and β , the predictive distribution is obtained by marginalization over w, α and β so that

$$p(t | T) = \iiint p(t | W, \beta) p(W | T, \alpha, \beta) p(\alpha, \beta | t) dW d\alpha d\beta \quad \text{Eq84}$$

where $p(t | W, \beta)$ is given by (3.8 – page 140) and $p(W | T, \alpha, \beta)$ is given by ((3.49) – page 153) in “Pattern recognition and Machine Learning, C. M. Bishop 2006”. From Bayesian theorem the posterior distribution for α and β is given by

$$p(\alpha, \beta | T) \propto p(T | \alpha, \beta) p(\alpha, \beta) \quad \text{Eq85}$$

If the prior is flat, then the values of α and β can be determine through the maximization of the marginal likelihood. If we define conjugate (Gamma) prior distribution over α and β , then the marginalization over theses hyperparameters in Eq73 can be performed analytically to give a student t-distribution over w . However, the integrand as a function of w has a strong skewed mode so that the Laplace approximation fails to capture the bulk of the probability mass, leading to poorer results than those obtained by maximizing the evidence (MacKay, 1999). In the evidence frame work, there are two approaches that e can take to the marginalization of log evidence. We can evaluate the evidence function analytically and then set its derivative to zero to obtain re-estimation equations for α and β . Alternatively, we use the expectation maximum (EM) algorithm. Thus we derive the marginal likelihood function by integrating over the weight parameters as follows

$$p(T | \alpha, \beta) = \int p(T | W, \beta) p(W | \alpha) dW \quad \text{Eq86}$$

From some manipulations we obtain,

$$p(T | \alpha, \beta) = \left(\frac{\beta}{2\pi} \right)^{N/2} \left(\frac{\alpha}{2\pi} \right)^{M/2} \int \exp\{-E(w)\} dw \quad \text{Eq87}$$

The integral over w can be evaluated as follows

$$\int \exp(-E(w)) dw = \exp\{-E(m_N)\} (2\pi)^{M/2} |A|^{1/2} \quad \text{Eq88}$$

Using equation Eq13 we can then write the log marginal likelihood in the form

$$\ln p(T | \alpha, \beta) = \frac{M}{2} \ln \alpha + \frac{N}{2} \beta - E(m_N) - \frac{1}{2} \ln |A| - \frac{N}{2} \ln(2\pi) \quad \text{Eq89}$$

which is the required expression for the evidence function. We first find α by using ((3.81*) and (3.82*) and (3.86*), page 167) in “*Pattern recognition and Machine Learning, C. M. Bishop 2006” and also adding the fact that A has eigenvalues $\alpha + \lambda_i$, we have

$$\frac{d}{d\alpha} \ln |A| = \frac{d}{d\alpha} \ln \prod_i (\lambda_i + \alpha) = \frac{d}{d\alpha} \sum_i \ln(\lambda_i + \alpha) = \sum_i \frac{1}{\lambda_i + \alpha} \quad \text{Eq90}$$

Thus rearranging the equation, we obtain

$$0 = \frac{M}{2\alpha} - \frac{1}{2} m_N^T m_N - \frac{1}{2} \sum_i \frac{1}{\lambda_i + \alpha} \quad \text{Eq91}$$

Rearranging we obtain

$$\alpha m_N^T m_N = M - \alpha \sum_i \frac{\lambda_i}{\lambda_i + \alpha} = \gamma \quad \text{Eq92}$$

Since there are M terms in the sum over I, when we multiply by 2α through some manipulations we obtain

$$\gamma = \sum_i \frac{\lambda_i}{\lambda_i + \alpha} \quad \text{Eq93}$$

From Eq80, we can derive the value of α that maximizes the marginal likelihood as follows

$$\alpha = \frac{\gamma}{m_N^T m_N} \quad \text{Eq94}$$

To find β , we maximize the log marginal likelihood with respect to β . To do this, we denote that the eigenvalues λ_i defined in ((3.87), page 168 *). Hence $d\lambda_i / d\beta = \lambda_i / \beta$ giving by

$$\frac{d}{d\alpha} \ln |A| = \frac{d}{d\beta} \sum_i \ln(\lambda_i + \alpha) = \frac{1}{\beta} \sum_i \frac{1}{\lambda_i + \alpha} = \frac{\gamma}{\beta} \quad \text{Eq95}$$

The stationary point of the marginal likelihood therefore satisfies

$$0 = \frac{N}{2\beta} - \frac{1}{2} \sum_{n=1}^N \{t_n - m_N^T \phi(x_n)\}^2 - \frac{\gamma}{2\beta} \quad \text{Eq96}$$

And rearranging we obtain

$$\frac{1}{\beta} = \frac{1}{N - \gamma} \sum_{n=1}^N \{t_n - m_N^T \phi(x_n)\}^2 \quad \text{Eq97}$$

Both α and β can be calculated by iterative procedure by choosing an initial values using (3.53*) and (3.95*⁵) respectively. For further information see the above the above mentioned book from “C. M. Bishop, 2006”. In the case that the number of the data points is large in relation to the number of the parameters, all the parameters will be well determined by the data because $\Phi^T \Phi$ from (3.83*) involves an implicit sum over data points, and re-estimation equations for α and β become

$$\alpha = \frac{M}{2E_W(m_N)} \quad \text{Eq98}$$

$$\beta = \frac{N}{2E_D(m_N)} \quad \text{Eq99}$$

Where E_W and E_D are defined by (3.25*) and (3.26*) respectively.

⁵* refers to Pattern Recognition and machine Learning – C.M Bishop

A.1.4 Case study: Inference for normal mean with known variance

Take $Y = \{y_1, \dots, y_N\}$ to denote a random sample from a normal distribution with unknown mean θ and known variance τ^2 . Then the likelihood function of θ , given the observation Y is

$$p(Y | \theta) \propto \exp\left\{-\frac{1}{2\tau^2} \sum_{i=1}^N (y_i - \theta)^2\right\} \quad (-\infty < \theta < \infty) \quad \text{Eq100}$$

The likelihood may be expressed more simply by noting that

$$\sum_{i=1}^N (y_i - \theta)^2 = s^2 + n(\bar{y} - \theta)^2 \quad \text{Eq101}$$

Where \bar{y} is the sample mean and $s^2 = \sum_{i=1}^N (y_i - \bar{y})^2$. Consequently, as function of θ ,

$$p(Y | \theta) \propto \exp\left\{-\frac{1}{2\tau^2} (\theta - \bar{y})^2\right\} \quad \text{Eq102}$$

The known situation will seldom arise in practice. However, for normal mean θ , we first consider a conjugate prior distribution, which is normal with mean μ and variance σ^2 . This may be justified by Boltzmann's maximum entropy theorem (Cercignami, 1988; Rosenkrantz, 1989). Suppose, we specify can only the mean μ and the variance σ^2 but nothing else about our prior distribution. Therefore we will choose the prior distribution $p(\theta)$ that maximize the entropy

$$\varphi(p) = -\int_0 p(\theta) \log p(\theta) d\theta, \quad \text{Eq103}$$

But subject to the mean μ and the variance σ^2 of the being equal to our specified values for μ and σ^2 . A straightforward calculation (Bellman, 1971, chapter 4) tells us that our optimal prior is normal $N(\mu, \sigma^2)$. This is the special case of Boltzmann's theorem (*Bayesian Methods, 2005, p122*):

The density $p(\theta)$ that maximizes $\varphi(p)$, subject to the constraints

$$E[t_i(\theta)] = t_i \quad (i = 1, \dots, q) \quad \text{Eq104}$$

takes the parameter exponential family form

$$p(\theta) \propto \exp\{\lambda_1 t_1(\theta) + \lambda_2 t_2(\theta) + \dots + \lambda_q t_q(\theta)\} \quad (\theta \in \Theta), \quad \text{Eq105}$$

where $\lambda_1, \lambda_2, \dots, \lambda_q$ can be determined, via the p-constraints, in terms of t_1, t_2, \dots, t_q .

Under the maximum entropy $N(\mu, \sigma^2)$ prior distribution, the hyperparameters μ can be specified as a prior estimate of θ , and σ^2 denotes the prior standard deviation. With

$$\sigma^2 = \frac{\tau^2}{k} \quad \text{Eq106}$$

The prior sample is

$$k = \frac{\tau^2}{\sigma^2} \quad \text{Eq107}$$

The signal-to-noise ratio is

$$k^{-1} = \frac{\sigma^2}{\tau^2} \quad \text{Eq108}$$

Then the posterior probability density of θ is

$$p(\theta | Y) \propto \exp\left\{-\frac{1}{2\sigma^2}(\theta - \mu)^2 - \frac{N}{2\tau^2}(\theta - \bar{y})^2\right\} \quad \text{Eq109}$$

The close form of the posterior probability density can be found by using the lemma below.

Lemma (Completing the square): For any constants A, B, a and b

$$A(\theta - a)^2 + B(\theta - b)^2 = (A + B)(\theta - \theta^*)^2 + (A^{-1} + B^{-1})^{-1}(a - b)^2 \quad \text{Eq110}$$

where

$$\theta^* = (A + B)^{-1}(Aa + Bb). \quad \text{Eq111}$$

NB: notice that we are not going to prove all these results. For more information about these see "Bayesian Methods, 2005, page 123". Thus, when we apply these results to the posterior density, we obtain

$$p(\theta | Y) \propto \exp\left\{-\frac{1}{2}\nu^{-1}(\theta - \theta^*)^2\right\} \quad \text{Eq112}$$

Where

$$\theta^* = \frac{N\tau^{-2}\bar{y} + \sigma^{-2}\mu}{N\tau^{-2} + \sigma^{-2}} = \frac{N\bar{y} + k}{N + k} \quad \text{Eq113}$$

And

$$\nu^{-1} = N\tau^{-2} + \sigma^{-2} = \tau^{-2}(N + k). \quad \text{Eq114}$$

In other words, θ is a posteriori normally $N(\theta^* | \nu)$ distributed. This is the maximum entropy distribution, given the posterior mean θ^* and variance ν . The expressions in Eq113 define the posterior mean, mode and median of θ , since these are identical for normal distribution. They all equal the weighted average

$$\theta^* = \rho \bar{y} + (1 - \rho)\mu \quad \text{Eq115}$$

where $\rho = \frac{N\tau^{-2}}{(N\tau^{-2} + \sigma^{-2})}$ describes the “reliability” of \bar{y} as an estimator of θ . Equation Eq112 tell

us that the posterior precision ν^{-1} equal the sum of the sampling precision and the prior precision. Therefore the posterior variance is, in this special case, less than both the prior variance σ^2 and the sampling variance $N^{-1}\tau^2$ of \bar{y} . This is not generally the case. Posterior probabilities can be calculated from

$$p(a < \theta < b | y) = \Phi\left(\frac{b - \theta^*}{\sqrt{\nu}}\right) - \Phi\left(\frac{a - \theta^*}{\sqrt{\nu}}\right), \quad \text{Eq116}$$

Under a general prior density $p(\theta)$ for normal mean θ , the posterior density of θ is

$$p(\theta | Y) \propto p(\theta) \exp\left\{-\frac{N}{2\tau^{-2}}(\theta - \bar{y})^2\right\} \quad \text{Eq117}$$

Note that the prior predictive density of the sample mean \bar{y} is

$$p(\bar{y}) = \int_{-\infty}^{\infty} p(\bar{y} | \theta) p(\theta) d\theta \quad \text{Eq118}$$

$$= (2\pi\tau^2)^{-\frac{1}{2}} \int_{-\infty}^{\infty} \exp\left\{-\frac{N}{2\tau^2}(\theta - \bar{y})^2\right\} p(\theta) d\theta. \quad \text{Eq119}$$

Then the first two of $\log p(\bar{y})$, with respect to (wrt), satisfy

$$\frac{\partial \log(p(\bar{y}))}{\partial \bar{y}} = -N\tau^{-2}\bar{y} + N\tau^{-2}E[(\theta | Y)] \quad \text{Eq120}$$

and

$$\frac{\partial^{-2} \log(p(\bar{y}))}{\partial \bar{y}^2} = -N\tau^{-2} - N^2\tau^{-4} \text{var}[(\theta | Y)] \quad \text{Eq121}$$

Therefore the general expressions for the posterior mean and variance of θ are

$$E[(\theta | Y)] = \bar{y} + N^{-1} \tau^2 \frac{\partial \log(p(\bar{y}))}{\partial \bar{y}} \quad \text{Eq122}$$

and

$$\text{var}[(\theta | Y)] = N^{-1} \tau^2 + N^2 \tau^4 \frac{\partial^2 \log(p(\bar{y}))}{\partial \bar{y}^2} \quad \text{Eq123}$$

These results relate to the regression in classical theory. Dawid (1973) and Leonard (1974) address the issue that the estimator in Eq114 based on the conjugate prior, can discredit the prior estimate as \bar{y} moves large away from μ . Thus they show in their analysis that the prior distribution with thicker tails yield possibly more desirable properties. There Dawid recommends a generalized t-prior density, taking the form

$$p(\theta) \propto \left[\nu \lambda + (\theta - \mu)^2 \right]^{-\frac{1}{2}(\nu+1)} \quad \text{Eq124}$$

In situations where the parameters space Θ is unbounded, Bayesian theory is faced with the problem that their estimates are generally quite sensitive to the thickness of the tails of the prior density. However, in practice it quite difficult to model the thickness of the tails based upon the prior information, for example how to determine the value of ν in Eq115. Therefore in practice, we refer to the entropy criterion Eq103 and a conjugate normal prior. Further we may point out that the posterior expectations of bounded function of unbounded parameters θ are not as sensitive to the tail behaviour of the prior density as the posterior mean of θ . Sensitivity issues are discussed by lavine (1991), and robust estimates of location are considered by Doksum and Lo (1990). We will consider the analytical procedure to estimate the parameter by using type2-maximu likelihood or empirical Bayes techniques. Before lounging into depth, we give a brief description of the improper prior and its relevance.

A.1.5 Vague prior

In a situation of complete prior ignorance (which may happen rarely) regarding the unknown parameter θ , we may consider vague prior information. The Bayesian paradigm cannot formally handle complete prior ignorance. In such a situation, we can use likelihood methods if a sampling model is available. However, Bayesian analysis can handle situations where prior information is fairly vague. For a clear explanation, let us consider a $N(\mu, \sigma^2)$ normal prior distribution where the normal mean μ is unknown and the variance σ^2 is known. A small value of σ^2 indicates the feeling that θ is quite likely to be close to μ . As σ^2 increases, the prior density becomes more and more dispersed around μ . Then the limit as $\sigma^2 \rightarrow \infty$, the prior $p(\theta) \rightarrow K$ for all θ , where the constant K is arbitrary and does not depend on θ , that is

$$p(\theta) \propto 1 \quad (-\infty < \theta < \infty) \quad \text{Eq125}$$

The limit is not a density, since it does not integrate to unity. The prior distribution of θ becomes improper. It represents a specific prior information that θ is equally likely to fall in the interest search interval. Under such a prior distribution, the posterior density for a normal mean θ reduces to a

$N(\bar{y}, N^{-1}\tau^2)$ density. We can therefore state that the posterior probability that θ lies in the 95%

confidence interval $(\bar{y} - 1.96\tau/N^{1/2}, \bar{y} + 1.96\tau/N^{1/2})$ is 0.95% (Bayesian method, 2005, p134).

However, under a wide range of regularity conditions, it is true that any $100(1-\epsilon)\%$ Bayesian region will give frequency coverage approaching $100(1-\epsilon)\%$ as N gets large and for any prior density $p(\theta)$ for θ .

Now we consider another way of choosing a vague prior distribution. The general Jeffrey's prior which yields excellent frequency properties (Bayesian methods, 2005). We will only set up the way to derive such a popular prior. Thus the Jeffreys' invariant prior (Berger, 1985, p.390) more generally can be defined by

$$p(\theta) \propto |F(\theta)|^{-1/2}, \quad \text{Eq126}$$

where

$$F(\theta) = -E_{y|\theta} \left[\frac{\partial^2 \log p(Y|\theta)}{\partial \theta^2} \right] \quad \text{Eq127}$$

denotes Fisher's information for θ . The choice of these prior distributions is situation dependent. That is, in some cases, both can yield good results. When using the improper distribution in prior

A.1.6 Conjugate priors

Given a distributional choice, the prior parameters are chosen to interject the least information possible. We will illustrate the conjugate prior distributions in different situations.

Case 1: variance known and we should infer the mean given the observation

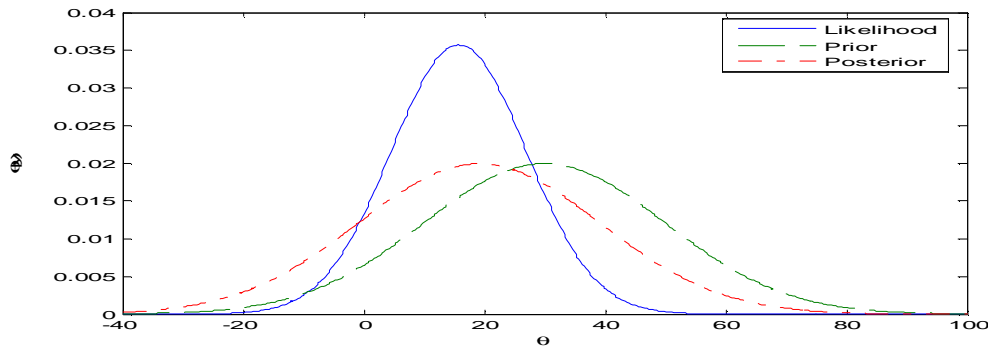


Figure 47: The posterior probability distribution formed by likelihood and the conjugate prior. As we can observe, the prior distribution shape increases with the variance value. Thereby, the posterior distribution shape also increases. The prior has a strong impact on the posterior probability distribution. This is due to first the standard deviation which controls the width of the prior and the sample size. It is stated in the literature that the larger the sample size, the less impact the prior has on the posterior probability distribution.

Case 2: Mean known and we wish to infer the variance itself

In such a situation, the conjugate prior used is the inverse Gamma. We simulate the inverse Gamma distribution using : $\alpha=[1 \ 6 \ 3 \ 4 \ 3 \ 3]$; $\beta=[1 \ .5 \ 1 \ 2 \ 1 \ .5]$ for the prior.

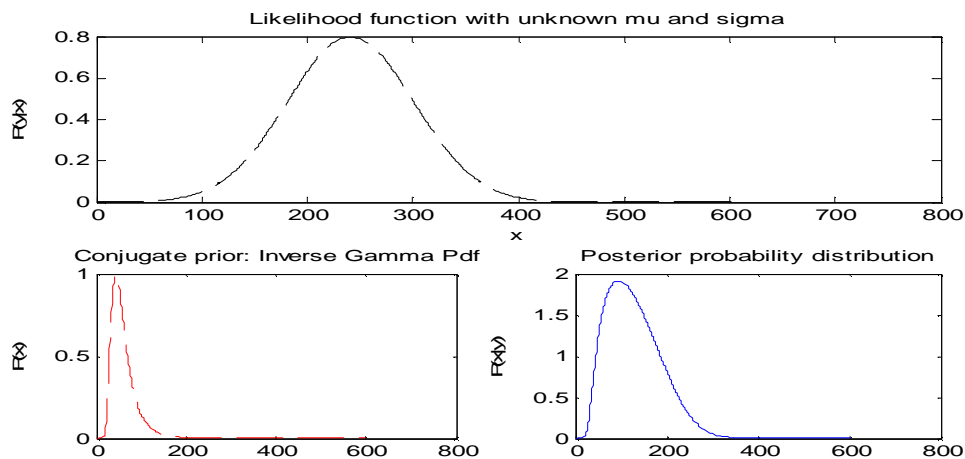


Figure 48: Bayesian theorem simulated using inverse Gamma conjugate prior. This inverse Gamma conjugate distribution is convenient in the situation, in which we suppose that the mean is known and we wish to infer the variance. Now suppose both mean and precision are unknown.

Case 3: Mean is known and we want to infer the precision λ .

It turns out to be most convenient to work with precision $\lambda = \frac{1}{\sigma^2}$. In this situation, the conjugate prior distribution corresponds to the Gamma distribution: $Gamma(x | a, b) = \frac{1}{\Gamma(a)} b^a x^{a-1} \exp(-bx)$, where $\Gamma(a) = (a-1)!$ if a is an integer.

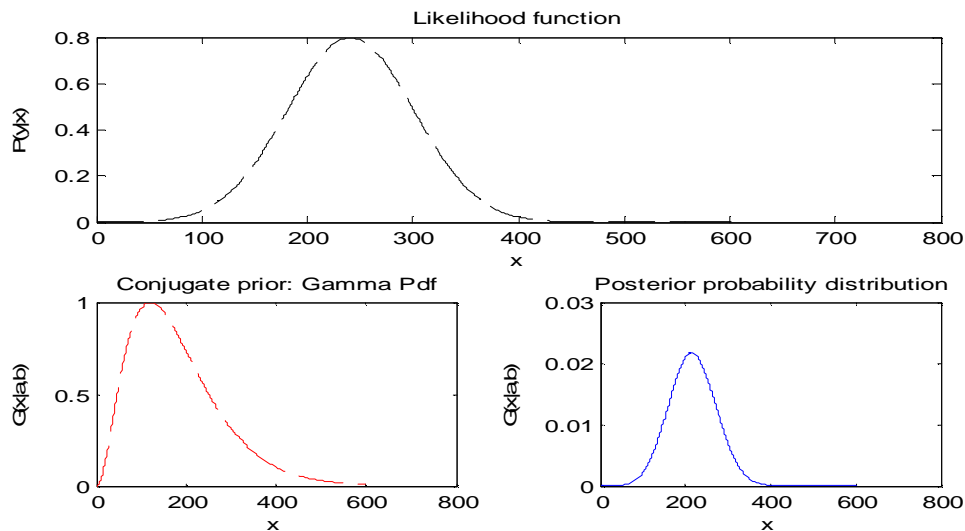


Figure 49: Gamma conjugate prior distribution simulated using $a = 3$ and $b = 1/2$.

Case 4: Mean and the precision are unknown. The conjugate is thus a normal Gamma distribution.

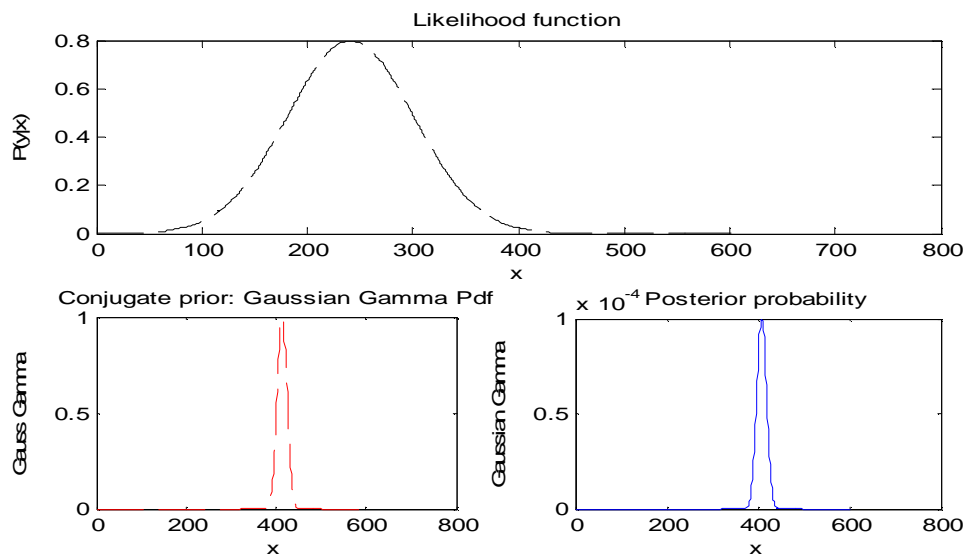


Figure 50: Simulation of the normal Gamma prior distribution

Case 5: Mean and the variance are unknown. The convenient conjugate prior distribution is the normal inverse Gamma distribution which is not plotted.

Comment

The illustration of the prior probability is in fact an experiment in which we demonstrate the desirable convenience and its important role played in the Bayesian analysis. We have seen that the shape of the prior, when the size of the data is small can contribute more in the decision making process of the posterior probability. Further, the flexibility in change of the shape can accurately yield best estimate with high frequency resolution, as we can see when the shape of the prior is very small. However, such a shape becomes insignificant in participating to the decision when the observation of interest is too large. In this experiment, we have not included the constant normalization factor and the log terms. We have only focus on the conjugate prior which is our focus in our framework. Moreover, the results show that the incorporation of the prior in the Bayesian analysis is can deal effective with the standard uncertainty associated with the best estimate and yield a supplement of information to estimate and track the parameter of interest. In addition, we have found that the selection of the conjugate prior distribution depends on the parameters mean and the variance. Such information is relevant for the posterior distribution and the accuracy of the estimate. The only issue we will outline is that if there are outliers in the data, the Gaussian model which has light tail cannot cope with it. We will need other distribution which present heavy tail such as student-t distribution or stable distribution to take into account such an outlier or trend if present in the observation data.

A.2 Stationary frequency estimation

Data modeled as a sum of sinusoids or exponentials arise in many areas of science namely nuclear magnetic resonance (NMR), functional magnetic resonances imagine (fMRI), auto tracking and more. However parameter estimation is a challenging problem. In this section we will present frequency parameter estimation theory in stationary case and nonstationary case as well. In the stationary case, we will introduce the signal harmonic case to followed by the multi-harmonic case to continue the analyses started by Larry Bretthorst in “Bayesian Spectrum analysis and parameter estimation” and Lars Kai Hansen, Finn Årup and Jan Larsen⁶ in their Bayesian framework about the “Exploring fMRI data for periodic signal components”. The use of such data are relevant because many studies are non-standard, and it is not always possible to provide a complete convincing analysis based upon pre-existing techniques. Therefore our study based on pre-existing algorithms to continue to develop the available understanding and apply these to specific knowledge. Thus the basic methods and two robust parameter estimation algorithms are presented. Several methods have been considered, trying to deal with such a problem by locating the maxima of an approximately periodic function. In this way, the least square method has been considered by Gauss [1] to estimate model parameters in noisy data. In this procedure, the problem is formulated in term of minimizing the sum of the discrepancies between the model and the data. Ideally, the problem will be formulated in such a way that only the frequency remains, but it is

⁶ All names mentioned above are professors in Intelligence Signal Processing (ISP) group from IMM at DTU.

not possible with direct least square, which require us to fit all the model parameters. The method of least squares may be difficult in practice even though it is well understood. Under Gaussian noise assumption, the least squares estimates are simply the parameter values that maximize the probability that we would obtain the data given the parameters.

The spectral method of dealing with this problem is based on the popular and powerful tool Fourier transform which is often used to estimate the frequency of the signal. The discrete Fourier transform (DFT) is a different method that can estimate the spectrum of the original discrete time series.

Even though such technique is well defined analysis tool, it does not work well when the signal to signal-to-noise ratio (SNR) of the data is small or when the data are nonstationary. Then it appears necessary to use the probability theory. The technique of the DFT has also been a problem when the signal is other than simple harmonic frequency. For example the chirped signal. The peak will spread out relative to a simple harmonic spectrum. This creates the noise to interfere with parameter estimation problem much more severely, and probability theory becomes essential.

In reaction against these difficulties encountered by DFT, Arthur Schuster [3] introduces the periodogram method of detecting a periodicity and estimating its frequency. The periodogram is based on averaging the square magnitude of the DFT and does yield useful frequency estimates under a wide range of conditions. Due to its statistical relevance in parameter estimation, Jaynes [4] establishes it as a “sufficient statistic” for inferences about single stationary frequency or discrete time sampled data set under Gaussian noise assumption. That is, the periodogram which summarises all the information in the data can be used to estimate the frequency under certain condition. We will investigate the basic methods, implement the probability theory behind the Bayesian analysis and combine the experimental and computational resources to the usefulness of the data. Further we will compare some classical and Bayesian spectral estimators through a Matlab simulation to analyze the performance and the error sensitivity of the Bayesian method.

A.2.1 Single harmonic estimation

We construct the likelihood model defined by $P(D | H, I)$ because it is dependence of the parameters which concerns us here. The time series we are $y(t)$ we are considering is postulated to contain a single stationary harmonic $f(t)$ plus noise $\varepsilon(t)$. The basic model is always we are recorded a discrete time data set $D = (d_1, \dots, d_N)$; sampled from $y(t)$ at discrete time $\{t_1, \dots, t_N\}$; with a model equation

$$d_i = y(t_i) = f(t_i) + e_i, \quad (1 \leq i \leq N). \quad \text{Eq128}$$

We will follow up the analysis of the Larry Bretthorst by introducing the prior probability for the amplitudes, which simplifies the calculation but has no effect on the final result. And also to discuss and introduce the calculation techniques without the complex model functions confusing the issues. The model is described as follows

$$f(t) = B_1 \cos(\omega t) + B_2 \sin(\omega t) \quad \text{Eq129}$$

which has three parameters (B_1, B_2, ω) that may be estimated from the observation data. There are several ways of estimating the parameters of interest. The problem to be solved is to compute the probability of the frequency ω conditional on the data and the prior information, this is abbreviated

as $P(w | D, I)$. But when we take the equation Eq4, there are four parameters $\{w, B_1, B_2, \sigma\}$. In this problem the two parameters B_1 and B_2 are referred to nuisance parameter, because the probability distribution that is to be calculated does not depend on these parameters. To perform this calculation we will apply the Bayes theorem to compute the joint probability of the all the parameters and them use the sum rule to eliminate the nuisance parameters.

Applying Bayesian theorem gives:

$$P(w, B_1, B_2 | D, \sigma, I) = \frac{P(D | w, B_1, B_2, \sigma, I)P(w, B_1, B_2, \sigma | I)}{P(D | \sigma, I)} \quad \text{Eq130}$$

which indicates that to compute the joint probability density, we must obtain three terms:

- $P(D | w, B_1, B_2, \sigma, I)$ is the likelihood function of the data given the parameters and the information I.
- $P(w, B_1, B_2, \sigma | I)$ is the prior probability distribution of the parameter given only the information.
- $P(D | \sigma, I)$ is the probability of the data given only the information I. It is called the normalization constant.

The sum rule can be applied to remove the dependence on the amplitudes:

$$P(w | D, \sigma, I) = \int_0^{\infty} \int_0^{\infty} \frac{P(D | w, B_1, B_2, \sigma, I)P(w, B_1, B_2, \sigma | I)}{P(D | I)} dB_1 dB_2 \quad \text{Eq131}$$

Assigning the likelihood function

This is equivalent at inserting the single stationary sinusoid frequency model in the expression of the noise $e = d - f$. changing f to indicate that it is the parameter that interest us, we obtain

$$P(D | w, B_1, B_2, \sigma, I) = (2\pi\sigma)^{-N/2} \exp\left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^N [d_i - B_1 \cos(\omega t_i) - B_2 \sin(\omega t_i)]^2 \right\} \quad \text{Eq132}$$

Assigning the prior probability

Assigning the prior probability is one of the most controversial area in Bayesian probability. Yet, to a Bayesian it is the most natural of things. The controversy arises when we try solve a problem in which we have a little prior information. If one has highly informative prior measurement, there is little discussion on how to assign the priors: the posterior probability derived in analyzing the previous

measurement can be used as the prior probability for the current measurement. But this delays the problem of how to assign probability of knowing little. In assigning the prior probability

$$P(\omega, B_1, B_2 | I) = P(B_1, B_2 | \omega, I)P(\omega | I) \quad \text{Eq133}$$

The prior probability of the frequency must be assigned completely independent of the amplitudes values. Here the only thing know about the frequency is that the data has been sampled uniformly, thus frequency values greater than the Nyquist frequency are aliased. So the frequency must be bounded between 0 and 2π . Using this bound and the normalization constraint in a maximum entropy calculation results in the assignment of

$$P(\omega | I) = \frac{1}{2\pi} \quad \text{Eq134}$$

as a prior probability of the frequency. Of course this is not the only prior probability that could be assigned. There is no contradiction in arriving at different prior probability assignments. The two different assignments correspond to being in different stage of knowledge, and different prior information result in different assignments. But this different assignment represent knowing little, effectively nothing, and regardless of what functional form one assigns to the prior; if the prior is slowly varying compared to the likelihood function, the prior will look like a constant over the range of the values where the likelihood is sharply peaked and its behaviour outside of this region will make little effectively no difference in the results. It is only when the width of the prior is comparable to the width of the likelihood function that it can have any significant effect.

Equation Eq9 becomes then

$$P(\omega, B_1, B_2 | I) = \frac{P(B_1, B_2 | \omega, I)}{2\pi} \quad \text{Eq135}$$

The probability of the amplitudes depends explicitly on the value of the frequency. In this calculation, it will be assumed that knowing the frequency tells us nothing about the amplitudes. This is not true in general, for example if the experiment is repeatable and a previous measurement is available, knowledge of the frequency will relevant about the value of amplitude. But if knowledge of the

frequency does not tell us anything about the amplitudes then $P(B_1, B_2 | \omega, I) = P(B_1, B_2 | I)$ and the joint prior probability of all the parameters may be written as

$$P(B_1, B_2, \omega | I) = \frac{P(B_1, B_2 | I)}{2\pi} \quad \text{Eq136}$$

In order to state what we know about the amplitudes, we suppose that we repeat this experiment a number of times. The signal is a stationary sinusoid. When the experiment is repeated, each of the amplitudes will take on both positive and negative values, (the phase will be different in each run of the data). Thus the average value of the amplitudes will be zero, but the mean square value will be nonzero. Applying the principle of maximum entropy will result in assigning a Gaussian prior probability to the amplitudes:

$$P(B_1, B_2 | \delta, I) = (2\pi\delta)^{-1} \exp\left\{-\frac{B_1^2 + B_2^2}{2\delta^2}\right\}, \quad \text{Eq137}$$

where δ^2 represents the uncertainty in the amplitude. If this prior probability is to represent little knowledge, then δ must be very large. But if δ is very large this prior probability is effectively a uniform prior probability over the range where the likelihood function is peaked. Due to the lack of information, we use uniform prior. This will yield conservative results. It is called improper prior.

This prior is needed to ensure that the total probability is one. So in parameter estimation problem, this prior is not relevant and can be dropped provided that the probability is normalized at the end of the calculation.

After manipulations involving elimination of nuisance parameter and removing constant, we obtain the likelihood function as defined below.

$$P(D | w, \sigma, I) \propto \sigma^{-N/2} \exp\left\{-\frac{N}{2\sigma^2} \left[\bar{d}^2 - \frac{2C(w)}{N} \right]\right\} \quad \text{Eq138}$$

The prior probability distribution is as follows

$$P(\sigma) = \frac{1}{\sigma}. \quad \text{Eq139}$$

This is Jeffreys prior, which can yield conservative result. Thus we obtain the posterior probability if the variance σ^2 is known by

$$P(w | D, \sigma, I) \propto \exp\left\{\frac{C(w)}{\sigma^2}\right\}. \quad \text{Eq140}$$

We see that the conditional posterior probability is related to the periodogram.

However, when the noise information is not available, the variance is unknown. To determine the posterior probability, we multiply the prior distribution and the likelihood function. Then we integrate out the variance parameter.

$$P(w | D, I) = \int_0^{\infty} P(D | w, \sigma, I) P(w | \sigma, I) d\sigma \quad \text{Eq141}$$

Thus we obtain the posterior probability called student t-distribution

$$P(w | D, I) \propto \left[1 - \frac{2C(w)}{N\bar{d}^2} \right]^{\frac{2-N}{2}} \quad \text{Eq142}^7$$

In our case it is the posterior probability density that a stationary harmonic frequency w is present in the data when no prior information about σ . These two posterior probabilities show why the discrete Fourier transform tends to peak at the location of a frequency when the data are noisy. Namely the discrete Fourier transform is directly related to the probability that a single harmonic frequency is present in the data, even when noise level is unknown. If the signal, being analysed, is a simple harmonic frequency plus noise, then the maximum of the periodogram will be the “best” estimate of the frequency that we can make in the absence of additional information about it. We now see the Fourier transform in a entirely new light: the highest peak in the discrete Fourier transform (DFT) is an optimal frequency estimator for data set which contains a single harmonic frequency in the present of Gaussian white noise.

- *Power Spectral density:*

We will express the result in probabilistic term to simplify the comparison between techniques, although there is no correspondence between a spectral density defined with reference to a stochastic model and one that pertains to a parameter estimation model.

$$\hat{p}(w) = \frac{N}{2} \int dB_1 dB_2 (B_1^2 + B_2^2) P(w, B_1, B_2 | D, \sigma, I) = 2[\sigma^2 + C(w)] P(w | D, \sigma, I)$$

Eq143a

This is the probabilistic meaning of the power spectral density (psd) defined by integrating the product of the total energy carried by the signal (not the noise) during our observation time by the joint posterior probability distribution for all the parameters. We now see that the peak of the periodogram is indicative of the total energy carried by the signal. One interesting thing this formula, is that the probability theory will handle those secondary maxima (side lobes) that occur in the periodogram by assigning them negligible weight. If the noise variance is known, Eq133a may be approximated by

$$\hat{p}(w) \approx \left[\sigma^2 + C(\hat{w}) \right] \left[\delta(\hat{w} - w) + \delta(\hat{w} + w) \right] \quad \text{Eq143b}$$

for most purposes. But for the term σ^2 , the peak of the periodogram is, in the model, nearly the total energy carried by the signal. These formulae can be useful in some context, we will show later in the

⁷ Where $\bar{d} = \frac{1}{N} \sum_{i=1}^N d_i^2$ and for two channels the student-t distribution is the product of two posterior distributions (Eq142).

simulation part of the project. For more information see computer simulation in section 6.1.1. Although these formulae can all be useful at certain level in stationary environment, the problem of estimating too closed harmonic frequencies remain. It has been notice that in Larry Bretthorst analysis that the student t-distribution Eq132 and Equation Eq133a are for most purposes. However, we may notice that these formulae result in yielding conservative results due to the incorporation of the improper prior in the Bayesian analysis. Therefore it appears necessary to describe a technique of parameter estimation of stationary signal which can yield satisfactory result while using proper informative prior. Such a technique is detailed below for multi-harmonic frequency estimation and also model comparison.

A.2.2 Model selection

As introduced above, in this section we will show the method of calculation for parameter estimation in stationary signal. Such a topic has been treated by several signal processing groups. Among those, Lars Kai, Finn Årup Niesen and Jan Larsen professors at Technical University of Denmark from ⁸ISP group in Informatic Mathematic Modeling department has provided a paper entitled “Exploring FMRI data for periodic signal components”. In these frameworks, the technique of parameter estimation has explored. In this analyze to perform the parameter estimation technique; we will introduce their method used in accordance with a linear regression, while the basis function is typically sinusoidal function. The technique of calculation to find an informative conjugate prior id described below. The signal will be modelled as a sum of multi-harmonic components plus noise. The most general form of the model is as follows:

$$f(t) = \sum_{j=1}^K b_j x_j(t) , \quad t = 1, 2, \dots, N \quad \text{Eq144}$$

Where N is the number of the data set, $x_j(t)$ is a set of periodic basic function such that and

$$x_{2j}(t) = \cos(j\omega_0 t) , \quad b_j \text{ is a } j \text{ linear amplitude parameter, } j=1, 2, \dots, K \text{ is the number}$$

$x_{2j+1}(t) = \sin(j\omega_0 t)$ of harmonics (model order) and ω_0 is the nonlinear fundamental frequency parameter. In matrix form, equation Eq144 may be written as

$$f = Xb \quad \text{Eq145}$$

However, more often than not, when the data is measured, it comes with noise, which can often be assumed to be additive. A model of the data might therefore be.

$$y(t) = f(t) + n(t) \quad \text{Eq146}$$

Where $n(t)$ is the additive white noise with zero mean and unknown variance σ^2 .

⁸ ISP means Intelligence Signal Processing

This difference equation forms a general linear regression model with a model of $K = 2j$ basis function and $2j$ dimension amplitudes vector b .

A fundamental problem we encounter is that the parameters $\{b, \omega_0, k, \sigma\}$ are unknown. In this way, the estimation of the fundamental frequency parameter will be posed in Bayesian term. That is we will develop a Bayesian paradigm that allows us to make inference about these parameters independently to the amplitude b and phase of harmonic independently of the noise variance σ^2 . We are only interested in estimating ω_0 the fundamental frequency and K the number of harmonics. This can be achieved by using nuisance parameters elimination technique.

- *Calculation technique for parameter estimations*

We will introduce the calculation method to introduce how to eliminate the other unknown parameter (nuisance parameters) which we don't need and also how to determine the hyperparameters of both the prior and the posterior probability. What we need here is only ω_0 and K the fundamental frequency and the harmonic order respectively. We can thus eliminate the unneeded parameters by explicit integration. Before we do so, it appears necessary to specify the Bayes theorem to clearly formulate our aim.

Problem statement 2: estimation of the posterior probability density $P(\omega, K | y)$

Solution strategy:

$$P(\omega, K | y) = \frac{P(y | \omega_0, K)P(\omega_0, K)}{P(y)} \quad \text{Eq147}$$

Where $P(y | \omega_0, K)$ is the likelihood function, $P(\omega_0, K)$ is the prior distribution and $P(y)$ is the normalization factor. For the fixed set the joint likelihood function; i.e the conditional probability density of the measurement given the parameters may be written as

$$P(y | \omega_0, b, K, \sigma^2) = (2\pi\sigma^2)^{-N/2} \exp\left(-\frac{1}{2\sigma^2}(y - bX)^2\right) \quad \text{Eq148}$$

Since we are only interested in estimating $\{\omega_0, K\}$ the fundamental frequency and the number of harmonic (model order) respectively, and the amplitude b and the noise variance σ^2 are unknown, we consider them as nuisance parameters to be eliminated. To do this, we use the prior distribution $P(b, \sigma^2)$ which quantifies the general knowledge we have on the domain and which potentially depends on the given basis set and model order.

We proceed to formulate explicitly the prior distribution by including some new hyper-parameter m, V denoted mean and variance of b and d, a designating mean and precision respectively for σ^2 .

$$P(b, \sigma^2) = P(b | \sigma^2)P(\sigma^2) = P(b | \sigma^2, m, V)P(\sigma^2 | d, a)$$

$$P(y | \omega_0, K) = \int d\sigma^2 \int db P(b, \sigma^2) P(y | \omega_0, b, K, \sigma^2)$$

$$P(y | \omega_0, K) = \int d\sigma^2 \int db P(b, \sigma^2) (2\pi\sigma^2)^{-T/2} \exp\left(-\frac{(y - bX)}{2\sigma^2}\right) \quad \text{Eq149}$$

The likelihood function has a normal distribution with the mean and variance unknown and assuming dependence on the likelihood function. Thus the convenient conjugate prior distribution to be chosen is the Normal Inverse Gamma (NIG), with four prior hyper-parameters $\{m, d, V, a\}$ and also four posterior hyper-parameters $\{m_p, d_p, V_p, a_p\}$ (see table of conjugate distribution in the Appendix). There are many different properties of conjugate distributions. To find the conjugate prior we consider the dependence of the likelihood function on the mean and the variance. In the following Figures, we show two examples for the sake of illustration of Gaussian and Inverse Gamma priors respectively.

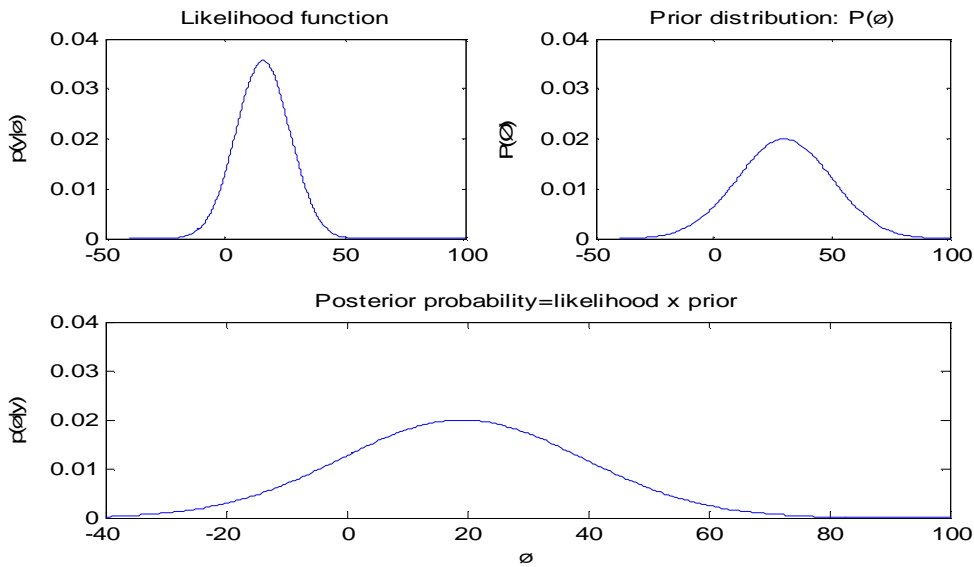


Figure 51: Illustration likelihood function which is normal with known variance and unknown mean, thus the prior is Gaussian distribution.

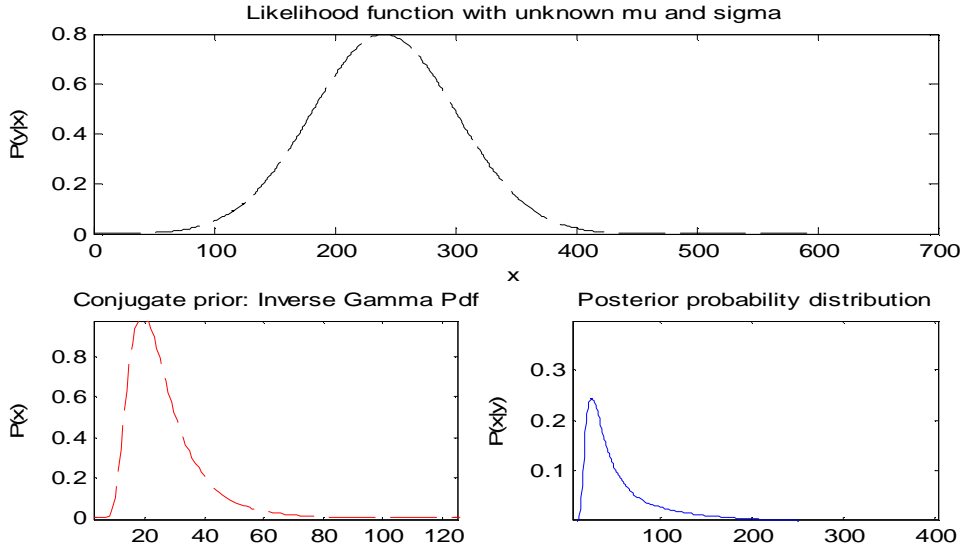


Figure 52: Continuous distribution with Normal likelihood function, Inverse Gamma conjugate prior

As we can notice in all these Figures, the priors have the same distribution as the posterior probability distribution as shown in the lowest panel in Figures 51-52. This is the key idea of the prior conjugacy.

More explanation about the prior distribution is comprehensively described in the literatures. The main idea is to choose the prior distribution such that the posterior probability density has the same form but with “updated ” i.e data dependent parameters. In our linear regression model presented above as a combination of systematic and Gaussian noise, the conjugate prior which can derive is the Normal Inverse Gamma or NIG(a, d, m, V).

$$P(b, \sigma^2 | a, d, K, m, V) = P(b | a, d, K, m, V)P(\sigma^2 | a, d) \quad \text{Eq150}$$

Where

$$P(b | a, d, K, m, V) = \int d\sigma^2 P(b, \sigma^2 | m, K) \quad \text{Eq151}$$

We obtain the marginal prior distribution over the amplitudes b as follows

$$P(b|a,d,K,m,V) = \frac{(a/2)^{-K/2} \Gamma((d+K)/2)}{(2\pi)^{K/2} |V|^{1/2} \Gamma(d/2)} \left\{ 1 + (b-m)^T (aV)^{-1} (b-m) \right\}^{-(d+K+2)/2}$$

Eq152

Where T means transpose. This is a multivariate t-distribution with mean m and covariance determined by $(\frac{a}{d-2}V)$, and centred at m with heavier tails than the normal distribution. The marginal prior distribution of the noise variance σ^2 is given by

$$P(\sigma^2|a,d) = \frac{(a/2)^{-d/2} (\sigma^2)^{-(d+2)/2}}{\Gamma(d/2)} \exp\left(-\frac{1}{2\sigma^2}a\right) \quad \text{Eq153}$$

The prior distribution over noise is the Inverse Gamma distribution of mean $\frac{a}{d-2}$ with $d > 2$. Hence the Normal Inverse Gamma distribution defined by equation Eq30 is explicitly expressed as

$$P(b, \sigma^2 | a, d, m, K, V) = \frac{(a/2)^{d/2} (\sigma^2)^{-(d+K+2)}}{(2\pi)^{K/2} |V|^{1/2} \Gamma(d/2)} \exp\left(-(b-m)^T (aV)^{-1} (b-m)\right) \quad \text{Eq154}$$

Now, we must give a parameters value, so that for long time series, their minimal influence on the result vanishes completely. Thus we set the prior mean to the observed signal variance

$\frac{a}{d-2} = \sigma_y^2 = \frac{y^T y}{N}$. The result of this calculation yield a small variance of the observation noise which is lower than that of the total observed variance. $d = 3$ is a small noise value for which the prior noise variance is finite, i.e hence the weak the prior. $m = 0$. The form of the prior covariance is determined by $V = \nu 1$, where 1 is a unit matrix. The parameter ν will be determined by

$$\nu = \frac{T}{Tr[XX^T]} \quad \text{where we let variance be equal to the variance of the measured signal.}$$

The prior covariance of the fitted signal \hat{y} is given by

$$\left\langle \hat{y} \hat{y}^T \right\rangle_{\text{Prior}} / N = Tr \left[XX^T \left\langle bb^T \right\rangle_{\text{Prior}} / N \right] = \frac{\nu a}{(d-2)} Tr[XX^T] / N \quad \text{Eq155}$$

Where Tr is a Matlab function called trace.m. This function calculates the sum of diagonal elements of an input matrix or the sum of the eigenvalues of an input. After integration and multiplication manipulations, we obtain the likelihood function

$$P(y|\omega_0, K) = \left(\frac{|V_P| a^d}{|V| (a_P)^{d_P} \pi^T} \right)^{1/2} \frac{\Gamma(d_P / 2)}{\Gamma(d / 2)} \quad \text{Eq156}$$

Where the parameters can be determined as follows:

$$V_P^{-1} = V^{-1} + X^T X \quad \text{Eq157}$$

$$m_P = V_P (V^{-1} m + X^T y) \quad \text{Eq158}$$

$$a_P = a + m^T V^{-1} m + y^T y - m_P^T V_P^{-1} m_P \quad \text{Eq159}$$

$$a_P = d + N \quad \text{Eq160}$$

When we use our specification, we obtain

$$V_P^{-1} = v \mathbf{1} + X^T X \quad \text{Eq161}$$

$$m_P = V_P X^T y \quad \text{Eq162}$$

$$a_P = (N+1) \sigma_y^2 - y^T X V_P X^T y \quad \text{Eq163}$$

$$d_P = 3 + N \quad \text{Eq164}$$

We can observe that the influence of the prior choice of a and d is weak for $N \gg 1$, because the prior contributions are of order one relative to N in equation Eq163 and Eq164 respectively. The probabilities of the complete set of hypotheses (parameterized by ω_0 and K) including the null – hypothesis are then given by

$$P(\omega_0, K|y) = \frac{P(y|\omega_0, K)}{P(y|0) + \sum_{\omega_0, K} P(y|\omega_0, K)} \quad \text{Eq165}$$

$$P(0|y) = \frac{P(y|0)}{P(y|0) + \sum_{\omega_0, K} P(y|\omega_0, K)} \quad \text{Eq166}$$

This algorithm has been designed based on Bayesian probability theory to detect periodic components in fMRI data. The requirement specification has involved the fundamental for Bayesian analysis using a specific linear regression model. It allows us to get also acquainted and understand the underlying calculation procedure for parameter estimation. And also it provides insight to master how to determine hyperparameter in in practical situation. Although, the simulation of this algorithm in Matlab did not work perfectly at the end, we have explored the technique and the capability of the algorithm to estimate the multi harmonic frequency and the order from the noisy signal. Thus the algorithm may be used with more flexibility to estimate the fundamental frequency and also detect the correct number of harmonics in periodic signal even though the fundamental frequency is beyond the Nyquist interval. Moreover, the result is useful for signal detection to localize the regions of periodicity. Such an technique can be useful in medical application, whereby may be used to localize region of highly affected by periodic physiological artefacts, such as cardiac pulsation.

A.3 Nonstationary frequency tracking

A.3.1 Likelihood method

The more important problem of frequency estimation is where the frequency is changing over time and frequency being grouped very close together. However, situation arises where frequencies are fixed. To cope with these problems, we treat the changing frequency as constant over intervals where barely changing and to estimate the frequency over each interval. The model considered here is defined by

$$y(t) = \mu + \sum_{j=1}^r (a_j \cos(w_j t) + b_j \sin(w_j t)) + n(t) \quad \text{Eq167}$$

The parameters to be estimated are thus μ and $a_j, b_j, w_j, \quad j = 1, 2, \dots, r$

We shall fist take r as known and later discuss its estimation. The observation noise $n(t)$ will be generated by a stationary process with zero mean and variance σ^2 .

Let

$$y_T = [y(1), y(2), \dots, y(T-1)]^T \quad \text{Eq168}$$

a column vector of T elements and let X_T be the matrix of whose t th row, $t = 0, 1, \dots, T-1$, is

$$\left[1 \quad \cos(w_1 t) \quad \sin(w_1 t) \dots \cos(w_r t) \quad \sin(w_r t) \right] \quad \text{Eq169}$$

The likelihood function is

$$P(y_T | X_T) = \frac{1}{(2\pi)^{T/2}} \frac{1}{|\Gamma_T|^{1/2}} \exp \left\{ -\frac{1}{2} (y_T - X_T B)' \Gamma_T^{-1} (y_T - X_T B) \right\} \quad \text{Eq170}$$

The log-likelihood is, for $n(t)$ Gaussian, $\log P(y_T | X_T)$

$$\log P(y_T | X_T) = \text{constant} - \frac{1}{2} \log |\Gamma_T| - \frac{1}{2} (y_T - X_T B)' \Gamma_T^{-1} (y_T - X_T B) \quad \text{Eq171}$$

We put Γ_T for the $T \times T$ matrix with $\gamma_n(t-s)$ in rows s , column t , where

$$\gamma(s) = E[(n(s)n(s+t))] - \mu^2 = E[n(s)n(s+t)] - E[n(0)]^2$$

Where $B' = \{\mu, a_1, b_1, \dots, a_r, b_r\}$. For the moment, we assume that Γ_T is known and maximise

Eq160 as though the w_j were known, depending only on w_j . Since

$$\hat{B}(w_1, \dots, w_r) = (X_T' \Gamma_T^{-1} X_T)^{-1} X_T' \Gamma_T^{-1} y_T, \quad \text{Eq172}$$

The reduced likelihood, i.e the likelihood with matrix is replaced by its estimator, is

$$\text{Constant} - \frac{1}{2} \log |\hat{\Gamma}_T| - \frac{1}{2} y_T' \hat{\Gamma}_T^{-1} y_T + \hat{Q}_T(w_1, \dots, w_r) \quad \text{Eq173}$$

where

$$\tilde{Q}_T(w_1, \dots, w_r) = y_T' \Gamma_T^{-1} X_T (X_T' \Gamma_T^{-1} X_T)^{-1} X_T' \Gamma_T^{-1} y_T, \quad \text{Eq174}$$

Which we called the “regression sum squares”. It is evident that must \tilde{Q}_T be maximized with respect to w_j in order to find the maximum likelihood estimator (MLE) of the w_j . After some manipulations and if the noise $n(t)$ is Gaussian white noise, eq174 becomes

$$\tilde{Q}(w_1, \dots, w_r) = \sigma^2 y_T' X_T (X_T' X_T)^{-1} X_T' y_T \quad \text{Eq175}$$

where the variance of the noise is σ^2 . Ignoring the variance, we obtain

$$\tilde{Q}(w_1, \dots, w_r) = y_T' X_T (X_T' X_T)^{-1} X_T' y_T \quad \text{Eq176}$$

After some manipulation eq176 becomes

$$\tilde{Q}(w_1, \dots, w_r) = T \bar{y}^{-2} + \sum_{k=1}^r C(w_k) \quad \text{Eq177}$$

The estimator may be obtained by choosing the locations of the r greatest maxima of $C(w)$, ignoring local maxima so close to others that may be assumed to be due to “sidelobes”. The estimator obtained by the method cited above (via maximization of Eq174, Eq176 and $C(w)$) will, under more general condition, have the same asymptotic properties. For more details see the “Estimation and Tracking of Frequency, 2001- by B.G Quin & E.J. Hannan”. It is somewhat more difficult to use Eq176 rather than a periodogram. But there is a good argument for using it. The reasons to use the Likelihood method can be described through some simulation later (see section 6.1.1).

A.3.2 Likelihood Procedure

Use $Q_T(\hat{w})$, with $r = 1$ or $C(w)$, computed from $y(t) - \bar{y}$, to obtain \hat{w}_1 as the maximizer value. Compute \hat{a}_1 and \hat{b}_1 by regressing $y(t) - \bar{y}$ on $\cos(\hat{w}_1 t)$ and $\sin(\hat{w}_1 t)$, and compute $Q_T^{(1)}(w)$ or $C^{(1)}(w)$ from the residuals $y(t) - \bar{y} - \hat{a}_1 \cos(\hat{w}_1 t) + \hat{b}_1 \sin(\hat{w}_1 t)$. Determine \hat{w}_2 as the maximising value of one or the other of these, and so on. Having found $\hat{w}_1, \dots, \hat{w}_r$, recompute \hat{a}_j and \hat{b}_j , $j = 1, \dots, r$ by computing

$$B(w_j; j = 1, \dots, r) = (X_T' X_T)^{-1} X_T' y_T$$

It is necessary we can perform a further iteration, by beginning from the residuals from the regression on $\cos(\hat{w}_j t)$, and $\sin(\hat{w}_j t)$, $j = 2, \dots, r$ to re-estimate \hat{w}_1 doing the same as above by omitting \hat{w}_2 from the regression procedure, and replacing \hat{w}_1 by new estimate of w_{01} xx, to get a new estimate of w_{02} and so on. For details see Bloomfield (1976).

- *Discussion*

Estimating a fundamental frequency depends also on which context and the application. If we are concerned with some frequencies near zero or are close together, then $Q_T(w_1, \dots, w_r)$ should be used to evaluate the maximum likelihood estimators of the frequency. In other words, the full likelihood procedure for the Gaussian white noise case should be used, not that of the periodogram $C(w)$. Of course, if it were known that the frequencies were separate from each other and far from zero except for pair of close frequencies, we might use $Q_T(w_1, w_2)$ for the pairs. Similarly, if we knew that there was only one frequency close to zero, we would use $Q_T(w)$ for that frequency. In either case of these cases, we would use the periodogram $C(w)$ for the remainder.

A.4 Robust Bayesian tracking supplement

- *Harmonic model*

$$d(t) = A_0 + \sum_{k=1}^K (A_n \cos(w_k t) + B_k \sin(w_k t)) + e(t), \quad e(t) \sim N(0, \sigma^2) \quad \text{Eq178}$$

where $w_k = \eta_k w_F$ and η_k is the known harmonic order then the parameters to be estimated are

$$\theta = \{A_0, A_1, \dots, A_k, B_1, \dots, B_k, w_F, \sigma^2\} \quad \text{Eq179}$$

$$d = Gb + e \text{ then } e = d - Gb; f = G(G^T G)^{-1} G^T d \quad \text{Eq180}$$

When estimating the fundamental frequency, we must consider the other parameter as nuisance. We need integrate them out of the joint posterior probability; we need to assign them to suitable priors that reflect the knowledge we have.

$$p(b | I_k) = k_b \quad \text{Eq181}$$

$$p(\sigma | I_k) = \frac{1}{\sigma} \quad \text{Eq182}$$

$$p(w_F | d, I_k) \iint \frac{p(d | w_F, \sigma, b, I_k) p(\sigma | I_k) p(b | I_k)}{p(d | I_k)} db d\sigma \quad \text{Eq183}$$

The evidence and the prior are constant.

$$\begin{aligned} p(w_F | d, I_k) &\propto \iint p(d | w_F, \sigma, d, I_k) \frac{1}{\sigma} db d\sigma \\ &\propto \iint (2\pi\sigma^2)^{-N/2} \exp\left\{-\frac{d^T d + b^T G^T G b - 2b^T G^T G d}{2\sigma^2}\right\} \frac{1}{\sigma} db d\sigma \end{aligned} \quad \text{Eq184}$$

After some analytical manipulations, we obtain

$$p(w_F | d, I_K) \propto \frac{\left(d^T d - d^T G(G^T G)^{-1} G^T d\right)^{\frac{-N+2K}{2}}}{\sqrt{\det(G^T G)}} \quad \text{Eq185}$$

$$p(w_F | d, I_K) \propto \frac{\left(d^T d - f^T f\right)^{\frac{-N+2K}{2}}}{\sqrt{\det(G^T G)}} \quad \text{Eq186}$$

B Matlab code

- *Main.m*

```
close all;
clear all;

%sound signal
[Acous_sig, Fs_a, Nbits]=wavread('Lyd og tacho signal');

% Vibration signal
[Vibro_sig, Fs_v, Nbits]=wavread('Vibration og tacho signal');

% the time series data
%y=Vibro_sig(:,1);
y=Acous_sig(:,1);

%parameters
%fs=Fs_v;    % sampling frequency
fs=Fs_a;    % sampling frequency

len=length(y); % length of the measurement
%n=(0:length(y)-1)/Fs_v; % time vector
n=(0:length(y)-1)/Fs_a; % time vector

% Reduce sample rate to 1 kHz i.e. important frequencies are
% below 500 Hz
nr=64;
fs=fs/nr;
y=resample(y,1,nr);
n=(0:length(y)-1)/fs;
%specgram(y, fs/4, fs)
% Inspection of spectrogram indicates the fundamental frequency is in the
% range of 10 to 100 Hz.
%nr=64;fs=fs_v/nr;specgram(resample(Vibro_sig(:,1),1,nr),fs/4,fs_)
%nr=64;fs=fs_v/nr;specgram(resample(Vibro_sig(:,2),1,nr),fs/4,fs_)

%aply bayes
reclsize=250; % Segmentation
overlap=200; % overlap parameter

% segment the signal
%% the function recordize100.m segments the signal and and overlap them
%% sequentially.
[X,Xt] = recordize100(y,reclsize,overlap);

% apply bayes to generate the log probability
Trec = Xt/fs; % time vector
Lbf=5;%3; % lower bound frequency(Lbf)
```

```

Ubf=100;%150;%50;%40; % Upper bound frequency(Ubf)
df=0.5;%0.25; % step size
Ff=(Lbf:df:Ubf)'; % frequency vector including the frequency band interval of the machine
 % (In this example we are from f0 til 3*f0=[min_f max_f]=[lbf ubf].

% number of the harmonics in the signals
%K = 1:0.5:3; % works well for Fund. Fred. (max top @ 50 Hz)%1:0.5:6; % K=[1 2.5 3]
K=[1 1.5 2]; % works well for the Fund. freq. (top @ 100 Hz)for acoustic
%K=1:2; % vibration
%K=[1.5 2];
vW.P = 2;%3;%3;%1; %3 % Number of regression records
vW.var = 1/4;%0.1%0.2;%/4;%0.1;%1/16;%1/4;%1/4;%1/8; % variance
%vW.var=0.60 ; % vibration
f=(0:len-1)/len*fs;

% determine the maximum likelihood using  $L_p=p(d|w)$ 
[Lp,Qf,F_] = bayes_w(X,Ff,K,n(1:size(X,1)))';

% track the fund. frequency using posterior probability  $z=p(w|d)$ 
% the conjugate prior is computed by using  $p(w|w,...w)$  in the
% following function btrack100.m
%f0 = 10; % start frequency
f0=15;
[Ftrack,z] = btrack100(lpnorm(Lp),Ff,f0,vW);

% The reconstruction is done by using process
[Xr] = bayes_r100(X,Ftrack,K,n(1:size(X,1)))';

t=n(1:250);

[Acous, Fs_a,Nbits]=wavread('Lyd og tacho signal');

% assign the pulse per revolution=one revolution in 1 pulse
PulsePerRevolution=1; % for acoustic signal
%PulsePerRevolution=2; % for vibration signal

%tacho=Vibro_sig(:,2);
%tacho=Vibro(:,2);
tacho=Acous(:,2);
% assign tacho signal
%tacho=vibro(1:end,2);
%tacho=acous(1:end,2);

% compute the trigger level
triglevel=min(tacho)+max(tacho)/2;

% compute sequence which exceeds level
levelExceedVector=tacho>=triglevel;

% find the corresponding values

%EdgeIndexVector=0;
EdgeIndexVector=zeros(1,length(levelExceedVector));

```

```

n1=1;

for n2=2:length(levelExceedVector)
    if levelExceedVector(n2)==1
        if levelExceedVector(n2-1)==0
            EdgeIndexVector(n1)=n2;
            n1=n1+1;
        end
    end
end
EdgeIndexVector = EdgeIndexVector(1:n1-1);

% convert the edage index vector in second
%EdgeIndexVectorInSecond=EdgeIndexVector/Fs_v;
EdgeIndexVectorInSecond=EdgeIndexVector/Fs_a;
%
% % plot
% figure(1)
% clf
% plot(tacho)
% hold on
% plot(levelExceedVector,'g')
% plot(EdgeIndexVector,1,'r*')
% hold off
% zoom on
% grid on
% title('Time signal')
% xlabel('Sample index')

% compute the delta time
deltatime=diff(EdgeIndexVectorInSecond);

ff=zeros(1,length(deltatime));
uf=zeros(1,length(deltatime));

for n=1:length(EdgeIndexVector)-1
    T(n)=(EdgeIndexVectorInSecond(n)+EdgeIndexVectorInSecond(n+1))/2; % midle point of the pulse
    %ff(n)=1/(2*(EdgeIndexVectorInSecond(n+1)-EdgeIndexVectorInSecond(n)+eps)); % for vibration signal
    ff(n)=1/(1*(EdgeIndexVectorInSecond(n+1)-EdgeIndexVectorInSecond(n)+eps)); % for acoustics signal
end

figure(1);
subplot(311); plot(Ff,lpnorm(Lp));xlabel('Freq [Hz]'); ylabel('log P');title('log P Records')
subplot(312); imagesc(Trec,Ff,lpnorm(Lp)), axis xy, title('log P')
xlabel('Time [s]'),ylabel('Freq [Hz]')

subplot(313);
plot(Trec,Ftrack);xlabel('Time [s]');ylabel('Freq [Hz]');title('Tracked Frequency')
subplot(312);
hold on;
plot(Trec,Ftrack,'w');
hold off;

```



```

figure(2)
%fs=Fs_v;nr=64;fs_=fs/nr;specgram(resample(Vibro_sig(:,2),1,nr),fs_/4,fs_)
fs=Fs_a;nr=64;fs_=fs/nr;specgram(resample(Acoustic_sig(:,2),1,nr),fs_/4,fs_)
title('Tacho. spectrogram')
figure(3)
%fs=65536;nr=64;fs_=fs/nr;specgram(resample(Vibro_sig(:,1),1,nr),fs_/4,fs_)
%title('Vibration spectrogram')
fs=32768;nr=64;fs_=fs/nr;specgram(resample(Acoustic_sig(:,1),1,nr),fs_/4,fs_)
title('Sound spectrogram')

% plot the speed profile
figure(4);
nn=(0:max(size(z))-1)/fs;
plot(nn(1:length(Ftrack)),Ftrack,xlabel('Time [s]'),ylabel('Frequency [Hz]'),title('Speed profile'))

figure(5);
subplot(211),plot(Ff,z),xlabel('Frequency [Hz]'),ylabel('log P( $\emptyset|D$ )'),title('Marginal posterior prob.')
subplot(212), imagesc(Xt,Ff,z),axis xy,xlabel('Frequency [Hz]'),colorbar,
ylabel('log P( $\emptyset|D$ )'), hold on;
plot(Xt,Ftrack,'w')
xlabel('Time [sec]'),ylabel('Fund. Frequency'),title('Marginal Post. Prob.: log P(D| $\emptyset$ )xP( $\emptyset$ )')

figure(6);
subplot(221),imagesc(X),axis xy,xlabel('Number of frame [n]'),
ylabel('Frame size [samples]'),title('Noisy observations')
subplot(222),imagesc(Xr),xlabel('Number of frame [n]'),axis xy,title('Reconst. true signal')
subplot(223),imagesc(X-Xr),axis xy,title('Error signal'),
xlabel('Number of frame [n]'),ylabel('Frame size [samples]')
subplot(224),plot(t,X(:,1),'k'),hold on,plot(t,Xr(:,1),'g'),plot(t,X(:,1)-Xr(:,1),'r'),
xlabel('Time [s]'),xlim([0 0.5]),ylabel('Amplitude'),title('Reconst. vs true + Error')

figure(7);
[M,N]=size(Xr); xr=reshape(Xr,1,M*N);
subplot(211),plot(xr(1:250),'k'),hold on,plot(y(1:250),'g'),ylabel('Amplitude'),title('True vs Noisy signal')
subplot(212),plot(xr(1:250),'k'),hold on,plot(xr(1:250),'g-'),xlabel('Sample [n]')
ylabel('Amplitude'), title('True (B) vs Reconst (G)')

figure(8);
%imagesc(Xt,Ff,z),axis xy,xlabel('Fundamental Frequency [Hz]'),colorbar,
% for FFmax = 100 Hz
%imagesc(Xt,[0 150],z),axis xy,xlabel('Fundamental Frequency
%[Hz]'),colorbar,
% for FFmax = 50 Hz
imagesc(Xt,[0 100],z),axis xy,xlabel('Fundamental Frequency [Hz]'),colorbar,

%imagesc(t1,[0 4],Pp
ylabel('log P( $\emptyset|D$ )'), hold on;
plot(Xt,Ftrack,'w')
xlabel('Time [sec]'),ylabel('Fund. Frequency'),title('Marginal Post. Prob.: log P(D| $\emptyset$ )xP( $\emptyset$ )')

```

```

figure(9)
%tau=1:length(tacho)/Fs_v;
tau=1:length(tacho)/Fs_a;

uf=interp1(T,ff,tau,'nearest');
plot(tau,uf,'r--','LineWidth',2),ylabel('Speed [Hz]'),title('Tacho speed profile')
%hold on,plot(Xt(1:end-10)/100,Ftrack(1:end-10),'b'),xlim([0 10]),xlabel('Time [s]'),ylabel('Frequency
[Hz]'),title('Vibration Speed profile')
hold on,plot(Xt(1:end-20)/500,Ftrack(1:end-20),'k','LineWidth',2),xlabel('Time [s]'),ylabel('Frequency
[Hz]'),title('Sound Speed profile')
,legend('Tacho','Fund. Freq. Estimate ')

```

```

figure(10)
plot(tau,uf,'r--','LineWidth',2),ylabel('Speed [Hz]'),title('Tacho speed profile')
%hold on,plot(Xt(1:end-10)/100,Ftrack(1:end-10),'b'),xlim([0 10]),xlabel('Time [s]'),ylabel('Frequency
[Hz]'),title('Vibration Speed profile')
hold on,plot(Xt(1:end)/500,Ftrack(1:end),'k','LineWidth',2),xlabel('Time [s]'),ylabel('Frequency [Hz]'),title('Sound
Speed profile')
,legend('Tacho','Fund. Freq. Estimate ')

```

- *m-files*

recordize100.m segments the signal and overlap the different records.

```

function [X,Xt] = recordize100(x,recsize,overlap)
% Synopsis:
% [X, Xt] = recordize(x,recsize,overlap)
% Input:
% x - data vector to segmentize
% recsize - size of segments
% overlap - number of samples each segment overlap, if overlap < 1
% it is taken as the percentage of the recordsize.
% Output:
% X [recsize,M] - segmented data
% Xt[recsize,1] - Indices in X of first value in records

```

```

if abs(overlap)<1,
    overlap = fix(overlap*recsize);
end

```

```

N = length(x);
i = (1:recsize)';
j = 0:recsize-overlap:N-recsize;
X = x(i*ones(1,length(j)) + ones(recsize,1)*j);
Xt=j+1;

```

Bayes_w.m determines the ML

```

function [Lp,Qf,F_] = bayes_w(D,Ff,K,t)
% Synopsis:
%
% Lp = bayes_w(D,Ff,K,t)

```

```

%
% Description:
%
% Does bayesian frequency estimation on the columns in D.
%
% Input:
%
% D [NxM] - Data matrix
% Ff [NFx1] - Frequencies for which to compute p(w|D), where w =
%           2*pi*Ff.
% K [NKx1] - Vector with the harmonic orders in signal.
% t [Nx1] - Time vector. Used to construct basis vectors,
%           G = [ cos(2*pi*Ff*t) sin(2*pi*Ff*t) ... sin(2*pi*Ff*K*t) ]
%
% Output:
%
% Lp [NFxM] - Log of probability
%
% TFP 2002/2/21
if ~exist('K'), K =1;end

[SzRec,NRec] = size(D);
NFreq      = length(Ff);
NK         = length(K);

Lp         = zeros(NFreq,NRec);
% Er       = zeros(NFreq,NRec);
Qf         = zeros(size(Ff));
F_         = zeros(NFreq,NRec);

d_ = sum(D.*D);      % sum all rows in the matrix
dof = SzRec-(2*NK+1);
% % Normalize data for unit energy
% dmin = min(d_)/dof;
% d_ = d_/dmin;
% D = D/sqrt(dmin);
Fs=1/(t(2)-t(1));    % sampling frequency
ndiv = floor(NFreq/20); % number of division
for i=1:20;fprintf('X');end;
fprintf('\r');

for i=1:NFreq
if ~mod(i,ndiv),
%disp(sprintf('Ff=%g',Ff(i)));
fprintf('.');
end

% Avoiding aliasing
% ix = find((Ff(i)*K/Fs<0.5).*(Ff(i)*K>min(Ff))); % Min & max
% ix = find(Ff(i)*K>min(Ff));% Min only
% ix = find(Ff(i)*K/Fs<0.5); % Max only
% ix = find(Ff(i)>0); % ALL
nk = length(ix);

```

```

W = 2*pi*t*Ff(i)*K(ix); % w=2*pi*f*t where f=Ff*K
G = exp(1i*W); %

% G = [ones(size(t)),reshape([real(G);imag(G)],SzRec,2*NK)];
G = [ones(size(t)),reshape([real(G);imag(G)],SzRec,2*nk)];
Q = G'*G; %

B = inv(Q)*G'*D; %
F = G*B; % estimate of the record

f_ = sum(F.*F); % energy of the reconstructed signal
detQ = det(Q);

F_(i,:) = f_;
Qf(i) = detQ;

% Equaiton from book:
lp = -0.5*size(G)*[1;-1]*log((d_-f_)+eps) - log(detQ+eps)/2; % the likelihood function P(w|d,Ik)-->eq(4.15)
% lp = -0.5*(size(G,1)-2*length(K)-1)*log((d_-f_)) -
% log(detQ)/2; % different scaling
% Equation modified to relative difference:
% lp = -0.5*dof*log((1-f_/d_)) - log(detQ)/2;
Lp(i,:) = lp;

% B_ = G'*D;
% F_ = G*B_;
% Er(i,:) = d_ - sum(F_.*F_);
end

btrack100.m tracks the fundamental frequency using prior
function [f0,z] = btrack100(Lpw,Ff,f0,vW)
%function z = btrack(Lpw,Ff,f0,vW)
% tracking prior from linear regression
%-----
% Lpw =likelihood probability density P(d|w)
% Ff = fundamental frequencies from c(w)=abs(fft(d)).2
% periodogram
% f0 = initialize fundament frequency
% z = estimates frequencis
%
%
z = Lpw;

P = 2; % number of record
u = 1; % mean value

if ~exist('vW'), vW = 2; end
if isstruct(vW),
P=vW.P;
vW = vW.var;
end

% compute tracking mean
k=1:P;

```

```

if P>1,
    %u=(2*(2*P+1)-6*(1:P'))/(P*(P-1));
    u=(2*(2*P+1)-6*k')/(P*(P-1));
    u=flipud(u); % revere the column in the up/down direction
end

if ~exist('f0')| f0<0,
    [zx,zi]=max(sum(z(:,1:8),2));
    f0 = Ff(zi);
end

f0 = repmat(f0(1),1,size(z,2));

for i=1:size(z,2),
    if i<P+1,
        f_ = [repmat(f0(1),1,P-i+1),f0(1:i-1)];%
    else
        f_ = f0(i-P:i-1);
    end
    % p2=log(exp(-1/(2*sigma)*(w-uT)))
    %p2 = -0.5/vW(i)*(Ff-f_*u).^2; % prior distribution
    p2 = -0.5/vW*(Ff-f_*u).^2; % prior distribution
    pw = z(:,i)+p2; % log (p(d(k)|w(k))*p(w(k)|w(k-1),...w(k-p)))
    [px,ix] = max(pw); % estimate of the frequency mean
    z(:,i) = pw; % fundamental frequency estimates
    f0(i) = Ff(ix); % frequency sampling points
end

bayes_r100.m reconstructs the true signal

function [f0,z] = btrack100(Lpw,Ff,f0,vW)
%function z = btrack(Lpw,Ff,f0,vW)
% tracking prior from linear regression
%-----
% Lpw =likelihood probability density P(d|w)
% Ff = fundamental frequencies from c(w)=abs(fft(d)).2
% periodogram
% f0 = initialize fundamnt frequency
% z = estimates frequencis
%
%
z = Lpw;

P = 2; % number of record
u = 1; % mean value

if ~exist('vW'), vW = 2; end
if isstruct(vW),
    P=vW.P;
    vW = vW.var;
end

% compute tracking mean

```

```

k=1:P;
if P>1,
    %u=(2*(2*P+1)-6*(1:P'))/(P*(P-1));
    u=(2*(2*P+1)-6*k')/(P*(P-1));
    u=flipud(u); % revere the column in the up/down direction
end

```

```

if ~exist('f0')| f0<0,
    [zx,zi]=max(sum(z(:,1:8),2));
    f0 = Ff(zi);
end

```

```

f0 = repmat(f0(1),1,size(z,2));

```

```

for i=1:size(z,2),
    if i<P+1,
        f_ = [repmat(f0(1),1,P-i+1),f0(1:i-1)];%
    else
        f_ = f0(i-P:i-1);
    end
    % p2=log(exp(-1/(2*sigma)*(w-uT)))
    %p2 = -0.5/vW(i)*(Ff-f_*u).^2; % prior distribution
    p2 = -0.5/vW*(Ff-f_*u).^2; % prior distribution
    pw = z(:,i)+p2; % log (p(d(k)|w(k))*p(w(k)|w(k-1),...w(k-p)))
    [px,ix] = max(pw); % estimate of the frequency mean
    z(:,i) = pw; % fundamental frequency estimates
    f0(i) = Ff(ix); % frequency sampling points
end

```

lpnorm.m normalizes the joint posterior probability

```

function [LPN,Lnorm] = lpnorm(LP)
% LPN = lpnorm(LP)
% Probability normalization of estimate og log p() for each column
% in LP
[M,N] = size(LP);
Lmax = max(LP);
LPN = LP - ones(M,1)*Lmax;
Lnorm = log(sum(exp(LPN)));
LPN = LPN - ones(M,1)*Lnorm;

```

- For the rest of the Matlab codes see CD ROM attached

C Figures

Performance deterioration due to wrong parameter setup to demonstrate one of the issues in prior parameters adjustment problems.

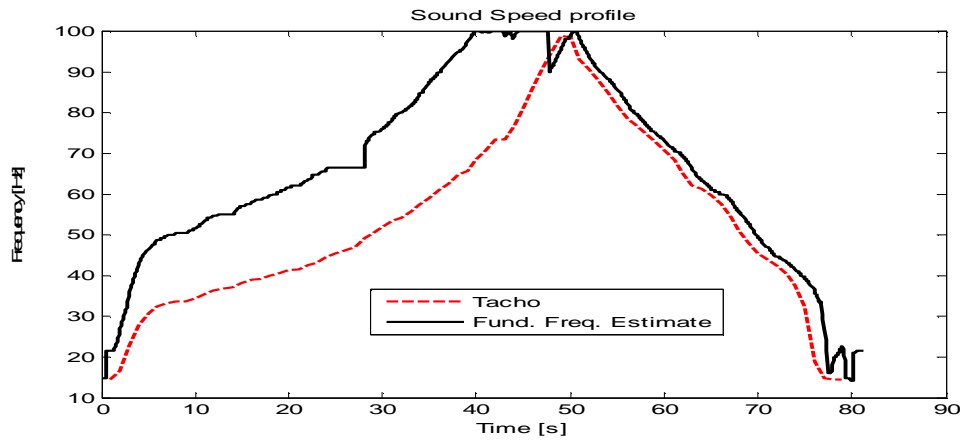


Figure 1c: $K=[1 \ 1.5 \ 2]$; $\text{var} = 1/4$; $P=3$

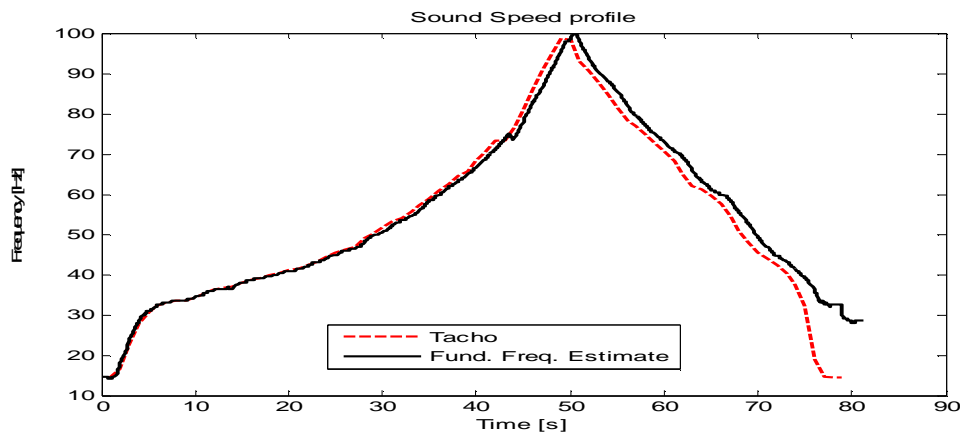


Figure 2c: $K=[1 \ 1.5 \ 2]$; $\text{var} = 0.1$, $P=3$

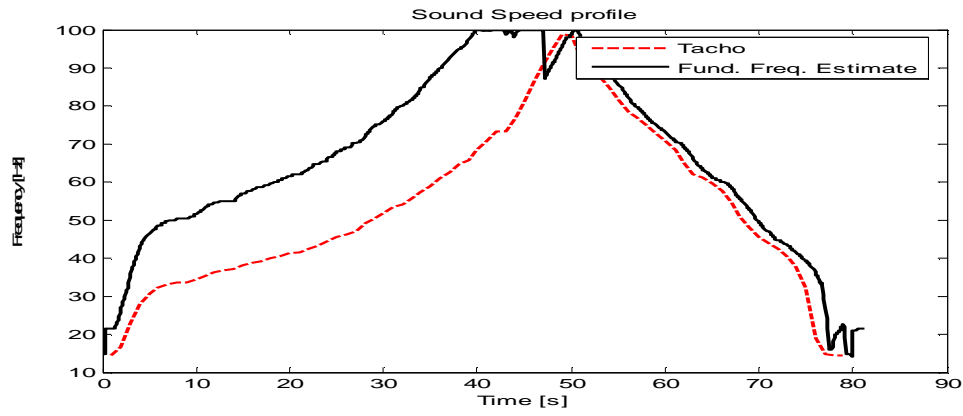


Figure 3c: $K=[1.5 \ 2]$; $\text{var} = 0.3$, $P=3$.

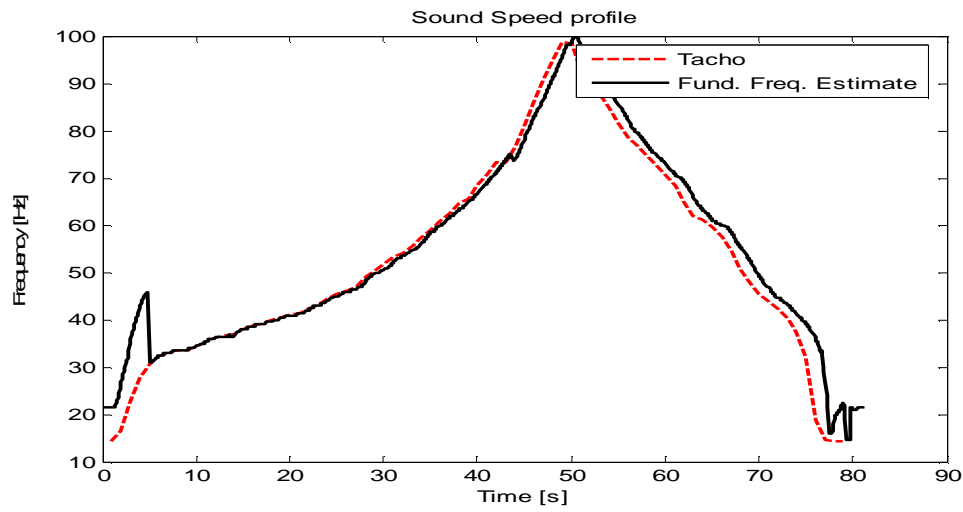


Figure 4c: $K=[1.5 \ 2]$; $\text{var} = 0.5$; $P = 3$

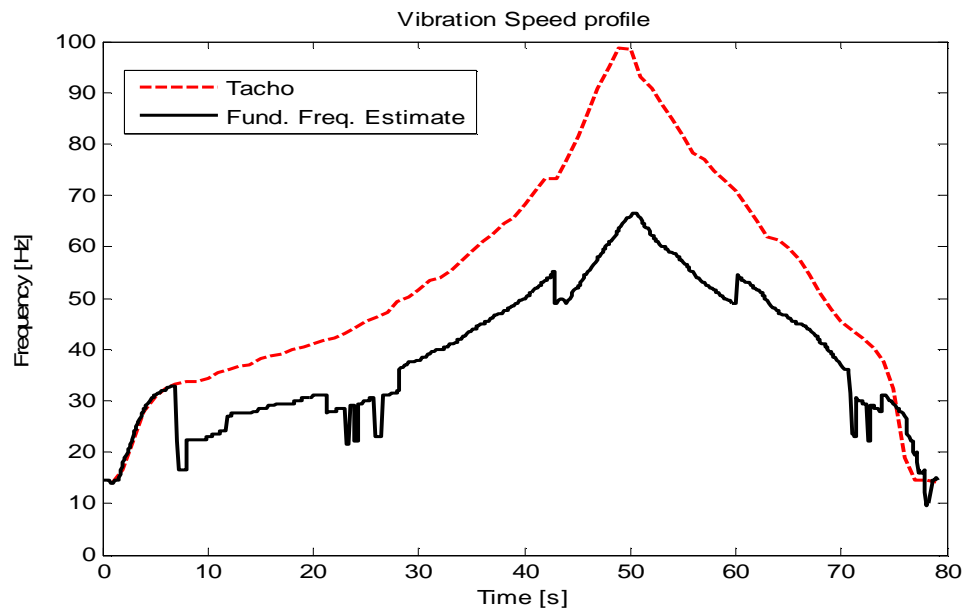


Figure 5c: $K=[1.5 \ 2]$; var = 0.6; P = 3;

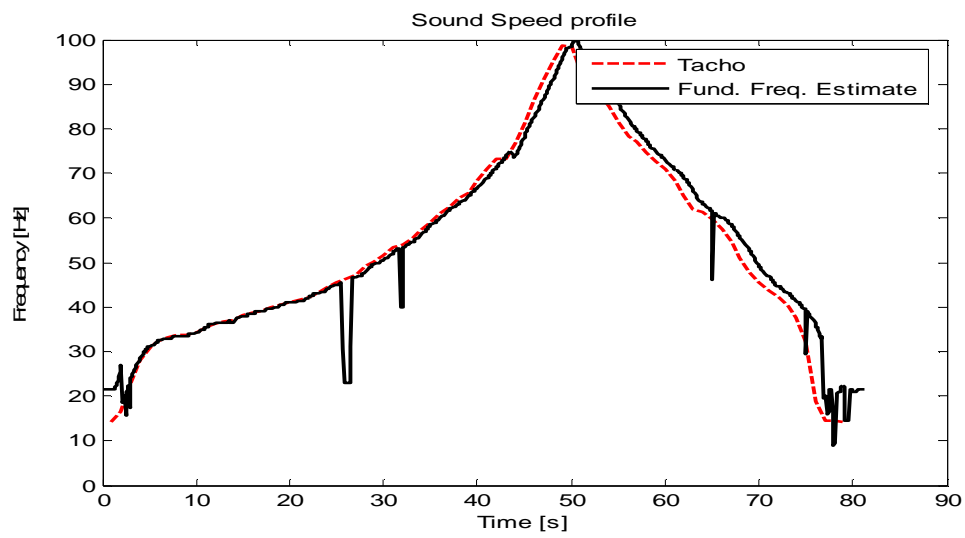


Figure 6c: $K=[1.5 \ 2]$; var = 5; P=3

D References list

- [1] James O. Berger, *Statistical Decision Theory and Bayesian analysis*, Second edition. Spriger-Verlag, 1985. ISBN 0-387-96098-8
- [2] Edwin T. Jaynes, "Prior Probabilities," *IEEE Transactions on the Systems Science and the Cybernetic*, SSC- α , 227-241, sept. 19668. Reprinted in Roger D. Rosenkrantz, Compiler, E.T. Jaynes: *Papers on Probability, Statistics and Statistical Physics*. Dordrecht, Holland: Reidel Publishing Company, pp. 116-130, 1983. ISBN: 90-277-1448-7
- [3] www.science.direct.com
- [4] Transfer learning by constructing informative priors, Rajat Raina, Andrew Y. Ng, Daphne Koller, Computer science department, standford University, Standford, CA 94305
- [5] <http://en.wikipedia.org>
- [6] Introduction to the Kalman Filter, Greg Welch and Gary Bishop
- [7] A. Herment G. demoment, P. Dumée, J-P. Guglielmi, and A. Delouche, "A new adaptive mean frequency estimator: Application to constant variance color flow mapping," *IEEE Trans. Ultrason ferroelectr. Freq. Contr.*, vol. 40, pp. 796-804, 1993
- [8] A review on the frequency Estimation and Tracking Problems – P.J. Kootsookos – CRC for Robust and Adaptive Systems – DSTO, Salisbury Site – February 21, 1999.
- [9] Legendre, P.S., (1812), *Theorie Analytique des Probabilités*, Paris, (2nd edition, 1814; 3rd edition, 1820)
- [10] The estimation and tracking of frequency – B.G.Quin, E.J. Hannan (Cambridge series in statistic and probabilistic mathematics)
- [11] Schuster, A., (1905), "The periodogram and its Optical analogy," *Proceedings of the Royal Society of London*, 77, pp.136
- [12] Jaynes. E.T (1987, "Bayesian spectrum and shirp Analsyis" in *Maximum Entropy and Bayesian Spectral Analysis and Estimation Problems*. C. Ray Smith, and G. J. Erickson, ed.. D. Reidel, Dordrecht-Holland, pp 1-37
- [13] Blackman, R.B., and J.W. Tukey, (1959), *The Measurement of the Power Spectra*, Dover Publications, Inc., New York.

- [14] Lecture notes in statistics – “Bayesian Spectrum Analysis and Parameter Estimation” G. Larry Bretthorst, (ebook).
- [15] <http://www.edw789.addr.com/norm.htm>
- [16] A. Herment G. demoment, P. Dumée, J-P. Guglielmi, and A. Delouche, ”A new adaptive mean frequency estimator: Application to constant variance color flow mapping,” IEEE Trans. Ultrason ferroelectr. Freq. Contr., vol. 40, pp. 796-804, 1993
- [17] Pattern Recognition and Machine Learning, Christopher M. bishop, 2006
- [18] Digital Signal processing, 3 edition, John Proakis, Dimitris G. Monalakis, 1996
- [19] Probability Volume 2, Emilyn Lloyd, 1980
- [20] The Estimation and Tracking of Frequency, B.G. Quin, E. J. Hannan, 2001
- [21] Bayesian Methods, Thomas Leonard, John S. J. Hsu, 2005
- [22] Technical view, N0 1. 1987, Vibration Monitoring, Brüel & Kjær
- [23] Technical Review N0 2, 1996, Nonstationary Signal Analysis Using wavelet Transform, Short-time Fourier Transform and Wigner-ville Distribution
- [24] Unsupervised Frequency Tracking Beyond the Nyquist Frequency using Markov Chains, Jean Francois, Giovannelli, Jerome Idier, Redha Boubertakh, and Alain Herment, IEEE Transactions on Signal Processing, Vol. 50, N0 12, Decembre 2002.
- [25] An Introduction To Parameter Estimation Using Bayesian Probability Theory, Larry Bretthorst, Washington University, Department of Chemistry, 1 Brooking Drive, St. louis, Missouri 63130
- [26] Bayesian Spectrum and Chirp Analysis, E. T. Jaynes, Waynes Crow Professor of Physics, Washington University, St. Louis MO 63130.
- [27] Wavelet Bayesian Block Shrinkage via Mixture of Normal –Inverse Gamma Priors, Daniela De Canditiis, Istituto Per Applicazioni della Matematica, CNR, Naples, Italy, Bravini Vidakovic, Georgia Institute of Technology, Atlanta, GA 30332-0205, USA.
- [28] Bayesian Analysis of Rotating Machines, A statistical approach to estimate and track the fundamental frequency, Thorkild Find Pedersen
- [29] Exploring FMRI data for Periodic signal components, Lars Kai Hansen, Finn Årup Nielsen and Jan Larsen, Informatics Mathematical Modeling, Technical University of Denmark B321, DK-2800 Lyngby, Denmark.

- [30] Time series Analysis, Henrik Madsen, 2000
- [31] Neural Networks Pattern Recognition, Christopher M. Bishop 1995
- [32] Introduction to the Kalman Filter, Greg Welch and Gary Bishop, TR 95-041, Department of Computer Sciences, University of North Carolina at Chapel Hill, Chapel Hill , C 27599-3175
- [33] YIN, a Fundamental Frequency Estimator for Speech and Music, Alain de Cheveigne, Hideki kawahara
- [34] Bayesian Spectrum estimation of unevenly sampled Nonstationary data, Yuan Qi, Thomas P. Minka, and Rosalind W. Picard, MIT Media Laboratory, Cambridge, MA, 02139, USA. Department of Statistics, Carnegie Mellon University, Pittsburgh, PA 15213, USA.
- [35] Bayesian Revolution in Spectral analysis, published in “Bayesian inference and Maximum Entropy methods in science and Engineering”, Paris 2000, ed. A. Mohammad-Djafari, American Institute of Physics proceedings, 568, p.557, 2001.
- [36] Leonard Janer, Modulated Gaussian Wavelet Transform based Speech Analyser (MGWTSA) Pitch Detection Algorithm (PDA). In Proceeding EUROSPEECH, volume 1, page 401-404, 1995.
- [37] Leonard Janer, Juan José Bonet, Eduardo Lleida –Solano, Pitch detection and Voiced / Unvoiced Decision algorithm based on Wavelet Transform.
- [38] A framework for state-space Estimation with Uncertain Models, Ali Sayed, Fellow, IEEE transaction automatic control, vol. 46, NO. 7, pp998-1013, July 2001.