Université d'Ottawa - University of Ottawa

APPLICATIONS OF BAYESIAN INFERENCE METHODS TO TIME SERIES DATA ANALYSIS AND HYPERFINE PARAMETER EXTRACTIONS IN MÖSSBAUER SPECTROSCOPY

By

© Lixin Dou

A thesis submitted to
the School of Graduate Studies and Research
in partial fulfillment of the requirements
for the degree of Doctor of Philosophy

Department of Physics

University of Ottawa
Ottawa, Ontario
December 9, 1998

Abstract

The Bayesian statistical inference theory is studied and applied to two problems in applied physics: spectral analysis and parameter estimation in time series data and hyperfine parameter extraction in Mössbuaer spectroscopy.

The applications to spectral analysis and parameter estimation for both single- and multiple-frequency signals are presented in detail. Specifically, the marginal posterior probabilities for the amplitudes and frequencies of the signals are obtained by using Gibbs sampling without performing the integration, no matter whether the variance of the noise is known or unknown. The best estimates of the parameters can be inferred from these probabilities together with the corresponding variances. When the variance of the noise is unknown, an estimate about the variance of the noise can also be made. Comparisons of our results have been made with results using the Fast Fourier Transformation(FFT) method as well as Bretthorst's method.

The same numerical approach is applied to some complicated models and conditions, such as periodic but non-harmonic signals, signals with decay, and signals with chirp. Results demonstrate that even under these complicated conditions the Bayesian inference and Gibbs sampling can still give very accurate results with respect to the true result. Also through the use of the Bayesian inference methods it is possible to choose the most probable model based on known prior information of data, assuming a model space.

The Bayesian inference theory is applied to hyperfine parameter extraction in Mössbauer spectroscopy for the first time. The method is a free-form model extraction approach and gives full error analysis of hyperfine parameter distributions. Two applications to quadrupole splitting distribution analysis in Fe-57 Mössbauer spectroscopy are presented. One involves a single site of Fe^{3+} and the other involves two sites for Fe^{3+} and Fe^{2+}. In each case the method gives a unique solution to the distributions with arbitrary shape and is not sensitive to the elemental doublet parameters.

The Bayesian inference theory is also applied to the hyperfine field distribution extraction. Because of the complexity of the elemental lineshape, all the other extraction methods can only use the first order perturbation sextet as the lineshape function. We use Blaes' exact lineshape model to extract the hyperfine field distribution. This is possible because the Bayesian inference theory is a free-form model extraction method. By using Blaes'

lineshape function, different cases of orientations between the electric field gradient principle axis directions and the magnetic hyperfine field can be studied without making any approximations. As an example the ground state hyperfine field distribution of $Fe_{65}Ni_{35}$ Invar is extensively studied by using the method. Some very interesting features of the hyperfine field distribution are identified.

Acknowledgments

I would like to take this opportunity to express my sincere gratitude to my thesis supervisor Dr. R.J.W. Hodgson for his encouragement, advice, help and kindness throughout the course of this work. Special thanks to Dr. D.G. Rancourt for the wonderful collaboration, discussion and help during the course of this research. I would like to thank Dr. Benson, B. Chan and H. Lacasse for their help with my teaching works. I would also thank the support staff and the graduate students in the Department of physics who helped to make life, study and work more bearable and joyful during my study and research.

Dedication

To my dear father who passed away just two weeks after I submited this thesis. I miss you.

To my husband Ping for your great support and understanding.

To my lovely daughter Vivian, you have been bringing so much joy and love to my life.

I love you all !

Contents

List of Tables

List of Figures

Chapter 1. Introduction

In this thesis the powerful Bayesian inference theory is studied in detail and applied to two problems in applied physics: 1) spectral analysis and parameter estimation in discrete time signals; and 2) hyperfine parameter distribution extraction in Mössbauer spectroscopy.

The first chapter (the current one) presents introductions to all the chapters to give readers an idea of the thesis structure.

Chapter 2, Inference in Science[5], is a mini course on inference in science. The syllogisms of deductive and plausible reasoning are discussed in detail. The inference in science fits into the plausible reasoning syllogisms and is an iterative process. Two generic investigation models[1] in scientific reasoning are introduced. Discussions on the advantages of the Bayesian inference theory in plausible reasoning are followed.

Chapter 3, Bayesian Inference Theory[1], is a crash course on Bayesian inference theory. First a brief survey about probability is presented. This is aimed to define some vocabulary which will be a necessary background to the following chapters. Then the Bayes' theorem is introduced. Generic applications to both linear model and non-linear models in standard normal inference problems are shown. This builds up a solid mathematical foundation for chapter five and chapter six.

Chapter 4, Gibbs Sampling[6]-[11] - a numerical technique in applications of Bayesian inference method. As a numerical technique, Gibbs sampling is used to calculate marginal posteriors instead of using the integration which, sometimes , is

very difficult to do both analytically and numerically. Because the method is a version of the Monte Carlo method based on the Markov chains, a brief introduction to Markov chains is presented first. Then the procedures of Gibbs sampling technique are followed.

Chapter 5, Spectral Analysis and Parameter Estimation in Discrete Time Signals. In this chapter the Bayesian inference theory and the Gibbs sampling technique are applied to an old and well known problem - spectral analysis and parameter estimation in discrete time signals. First a historical review on the problem is given. Then the application to the signals with multiple frequencies and white noise are presented. This includes the numerical models for the problem, the numerical algorithm and procedure, the computer simulated examples and the comparison of the results to the results from other approaches to the same problems. In practice, however, there are always situations that the models and the real problem do not match to each other. So some applications to the problems which violate the conditions in the model are conducted to test the robustness of the method. Again the results are compared with the results from other approaches. The application mentioned above are based on known models. Also in most of the real situations, the model itself is unknown. The only thing which is known is that the possible model space. Besides finding the parameters, the most probable model for the problem from the model space has to be found. So the last section of this chapter is dedicated to the model selection. The numerical procedures and examples are presented.

Chapter 6, Hyperfine Parameter Distribution Extraction in Mössbauer Spectroscopy. In this chapter the problems involving hyperfine parameter distribution extraction in Mössbauer spectroscopy are studied extensively by using the Bayesian

inference theory and Gibbs sampling. In the first section a background introduction to Mössbauer spectroscopy and hyperfine interactions in Fe-57 (which is the main focus of the application) is presented in detail as well as a review of previous work in hyperfine parameter distribution extraction. Also the method called Voigt-based fitting is described in detail because this method is the most powerful fitting method for the extraction at present and the results for the same problems by using the Bayesian inference method are compared with the results by using the Voigt-based method. The numerical procedures and algorithms are described in the second section. What follows next is the application to both single generalized site and multiple generalized sites quadrupole splitting distributions(QSD) extraction which is the simpler case where only electric quadrupole interaction exists. When both electric quadrupole interactions and magnetic hyperfine interactions exist, things are more complicated because the elemental lineshape function involves more hyperfine parameters and is in a more complex format. The fitting methods can only deal with the approximated elemental lineshape function from the first order perturbation treatment because they use a fixed distribution-type model to do the extraction. What they try to extract are the model parameters. Because Bayesian inference theory is a model free method, any kinds of elemental lineshape function can be used. A detailed description onf the elemental lineshape models is given and followed by applications to some simulated spectra which demonstrate the robustness of the Bayesian inference theory in the hyperfine parameter extraction with known elemental lineshape functions. The last section is dedicated to a very complicated and well known problem in Mössbauer spectroscopy, the ground state hyperfine distribution of $Fe_{65}Ni_{35}$ Invar. Different elemental lineshape models are used to study the sensitivity of the hyperfine field distribution(HFD) at low fields

to the way how the electrical field gradient(EFG) is represented.

Chapter 2. Inference in Science

2.1 Deductive and Plausible Reasoning

In our everyday lives, we often make decisions, judgments, and propositions about our lives and the world we live in. In most of the cases there is not enough information in our hands when we make decisions; but still we must decide immediately what to do. For example, on a summer morning the sky is very dark and from last night's forecast we know there is an 80% chance of rain the next morning. What are we going to do, take an umbrella to work or not? Of course almost everybody would take an umbrella. But by what reasoning process do we arrive at this decision? A moment's thought makes it clear that our decision is not a logical deduction from the information we have. The information does not make 'it is going to rain' certain, however the information makes it extremely plausible. In fact we are hardly able to get through one waking hour without facing some situation where we do not have enough information to permit deductive reasoning. The syllogisms in the next two sections give clear definitions for deductive reasoning and plausible reasoning.

2.1.1 Deductive reasoning

Deductive reasoning can be analyzed ultimately into the repeated application of two strong syllogisms:

Syllogism 1:

If A is true, then B is true

A is true

Therefore, B is true

and its inverse:

Syllogism 2:

If A is true, then B is true

B is false

Therefore, A is false

For example, we know $x+y=3$. Let : "$A = x$ is 2" and "$B = y$ is 1". So if A is true, without any doubt we know that B is true; or in the opposite case where B is false (y is not equal to 1) then A is false (since $x+y=3$, y is not equal to 1, x can not be 2). These are the kinds of situations in which we have sufficient information to allow this kind of reasoning, i.e. deductive reasoning. But in almost all of the situations confronting us we do not have the right kind of information to do this. We have to fall back on some weaker syllogisms according to which the plausible reasoning can be analyzed.

2.1.2 Plausible reasoning

The first weaker syllogism is:

Syllogism 3:

If A is true, then B is true

B is true

Therefore, A becomes more plausible

In this situation, the evidence does not prove that A is true, but verification of one of its consequences does give us more confidence in A. This is the plausible reasoning

syllogism that we use to decide upon taking an umbrella to work since there is an 80% of chance of rain and the sky is dark which makes 'it is going to rain' very plausible.

Another weak syllogism, using the same major premise, is

Syllogism 4:

If A is true, then B is true

A is false

Therefore, B becomes less plausible

In this case, the evidence does not prove that B is false. However one of the possible reasons for its being true has been eliminated, and so we feel less confident about B.

The human brain, in doing plausible reasoning, not only decides whether something becomes more plausible or less plausible, but it evaluates the degree of plausibility in some way. The brain also makes use of old information (from experience, we know that the plausibility of rain very much depends on the darkness of those clouds) as well as the specific new data of the problem (the weather man predicted that it is going to rain). In deciding what to do we try to recall our past experiences, with clouds and rain, and what the weather man predicted last night. The reasoning of a scientist, by which he/she accepts or rejects his/her theories, consists almost entirely of the third and fourth kinds.

2.2 Reasoning in Science and the Bayesian Approach

2.2.1 Reasoning in science

Scientists conduct experiments in order to postulate and test deductions about the world in which we live. We learn quickly that the world is too complicated for us to analyze all at once. Rarely can a series of logical deductions be made which lead from data to theory. We can progress only if we dissect problems into little pieces and study them separately. Sometimes we can invent a mathematical model based on the available information (incomplete), which reproduces several features of one of these pieces, and whenever this happens we feel that progress has been made. As knowledge advances, more information is available, we are able to invent better and better models, which produce more and more features of the real world, and do so more and more accurately.

Scientific investigation is a controlled learning process in which various aspects of a problem are illuminated as the study proceeds. It can be thought of as a major iteration within which secondary iterations occur. The major iteration is that in which a tentative conjecture suggests an experiment, appropriate analysis of the data so generated leads to a modified conjecture, and this in turn leads to a new experiment, and so on. An idealization of this process is seen in Figure 2.1, involving an alternation between conjecture and experiment carried out via experimental design and data analysis.

Because we can never be sure that a postulated model is entirely appropriate, we must proceed in such a manner that inadequacies can be taken account of and their implications considered as we go along. To do this we must do more than merely postulate a model; we must build and test a tentative model at each stage of

the investigation. Only when the analyst and the investigator are satisfied that no important fact has been overlooked and that the model is adequate to the purpose, should it be used to further that major iteration. The iterative model building process taking place within the major iteration is depicted in Figure 2.2.

2.2.2 The Bayesian approach

The Bayesian inference theory is a statistical inference theory. The method of reasoning under the circumstances in which we do not have complete information is called inference. The Bayesian inference theory provides us a unique mathematical language to describe the essential features of this plausible reasoning. Among different systems of statistical inference, that derived from Bayes' theorem will be seen to have properties which make it particularly appropriate to its role in scientific investigation. In particular:

1. Precise assumptions introduced on the left in Fig. 2.2 lead, via a leak-proof route to consequent inferences on the right.

2. It follows that, given the model, Bayesian analysis automatically makes use of all the information from the data.

3. It further follows that inferences that are unacceptable must come from inappropriate assumptions and not from inadequacies of the inferential system. Thus all parts of the model, including the prior distribution, are exposed to appropriate criticism.

4. Because this system of inference may be readily applied to any probability model, much less attention need be given to the mathematical convenience of the models considered and more to their scientific merit.

5. Awkward problems encountered in sampling theory (another statistical inference method [1]), concerning choices of estimator and of confidence intervals, do not arise.

6. Bayesian inference provides a satisfactory way of explicitly introducing and keeping track of assumptions about prior knowledge or ignorance. (It should be recognized that some prior knowledge is employed in all inferential systems.)

These properties suggest that Bayes' theorem provides a system of statistical inference intrinsically suited to iterative model building, which is in turn an essential part of scientific investigation. The detailed introduction about the Bayesian inference theory will be given in chapter 3.

In the last decade, there have been an increasing number of successful applications of the Bayesian inference theory as more and more people have begun to realize that this theory has the intrinsic properties suited to the inference problems in scientific investigation. Most of the applications face the extreme challenge of the computational aspects (as we will see in chapter 5 and 6 of this thesis). It would not be possible to apply the theory broadly in science without modern high speed computers. Two major contributions to the use of the Bayesian inference theory are the development of the theory and the algorithms to reconstruct large images [3]in Cambridge, England and the work [4], at Washington University, St. Louis, U.S.A., of demonstrating the power of the Bayesian spectrum analysis and parameter estimation which will be quantitatively compared to the results of the study in this thesis.

For those readers who are interested in probability theory in scientific investigation and the Bayesian inference theory in statistical analysis, more detailed materials

are available in the book of Jaynes [4], which is still incomplete, and the book by Box and Tiao [1].

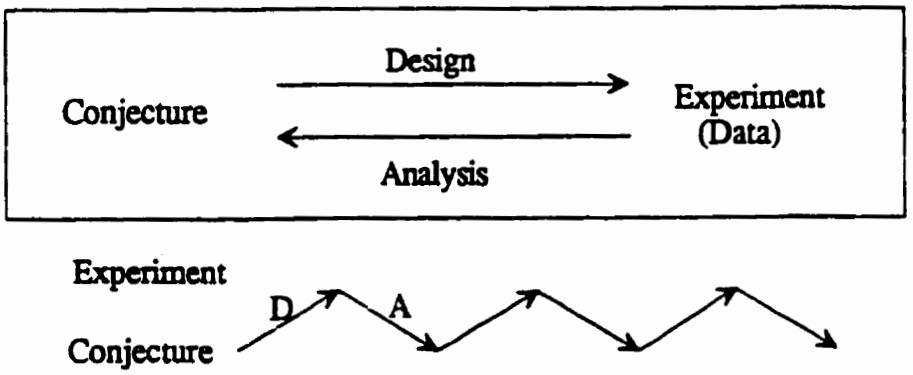

Figure 2.1: Iterative process of scientific investigation (the alternation between conjecture and experiment).

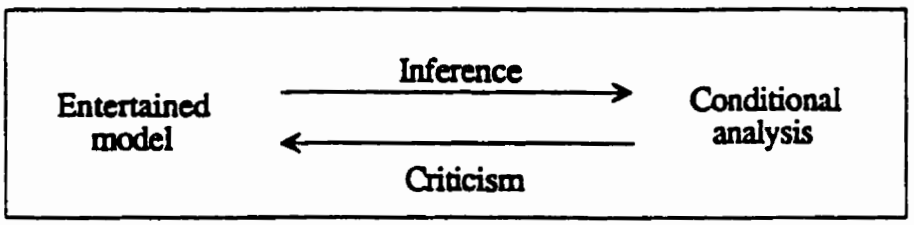

Figure 2.2: Statistical analysis of data as an iterative process of model building.

Chapter 3. Bayesian Inference Theory

In this chapter some background of probability theory and a brief introduction about Bayesian inference theory are presented. The intention of this chapter is to give general readers a neccessary probability and Bayesian inference theory backgroud to understand the rest of the chapters. For readers who are interested in the detail of these theory, please refer to References: [2] and [1]

3.1 Some Background of Probability Theory

3.1.1 Experiments and sample spaces

The word "experiment" is used here in a very general sense to describe virtually any process of which all possible outcomes can be specified in advance and of which the actual outcome will be one of those specified. The outcome of an experiment may be random or non random. For the purpose of this study, the interesting feature of an experiment is simply that its outcome is not definitely known by the experimenter before hand. The set of all possible outcomes of an experiment is called sample space S of the experiment. For a given experiment, the number of outcomes and the name of the outcomes in S may be chosen differently by different people. The basic requirements are that S must include all possible outcomes and that each outcome must be described with all essential detail. The sample space S of any experiment is conveniently regarded as a set of elements, each element being a possible outcome of the experiment.

3.1.2 Events and probabilities

An event refers to the result of a trial which refers to the occurrence of an experiment. The limiting cases of events are also considered to be events. Let Ω denote the total or certain event characterized by every possible outcome of a trial, and Φ the null or impossible event containing no outcomes. Both Ω and Φ are viewed as events. A complete or exhaustive set of events is a collection whose union is the total event Ω.

The notation of probability corresponds to the chance that some event of interest will occur. Probability can be called a measure applied to the events that can occur when an experiment is performed. Formally a probability function is a real-valued set function defined on the class of all subsets of the sample space S; the value that is associated with a subset A is denoted by $P(A)$. The value of a probability $P(A)$ ranges from 0 to 1 inclusively. The probability of the null event is 0 and that of the total event is 1; that is $P(\Phi) = 0$ and $P(\Omega) = 1$. Actual specification of the value of $P(A)$ must come from analytical considerations of the experiment performed and the mechanism behind it.

A set of events is independent if the events have no effect on each other. A set event is called dependent if it is not independent. In other words, at least one event affects at least one other event in the set. The joint probability of a collection of independent events is given by the product of their respective probabilities.

$$P(AB) = P(A)P(B) \tag{3.1}$$

where A and B are independent.

The joint probability of a collection of dependent events must take into account the conditional probability that characterize their dependence. The conditional

probability is the probability of the occurrence of an event A when additional information about the outcome of the experiment has been obtained from the occurrence of some other event B. Then

$$P(AB) = P(A \mid B)P(B) \tag{3.2}$$

where we call $P(A \mid B)$ the conditional probability of A occurring, given that B has occurred.

3.1.3 Random variables and probability density distribution(PDD) functions

A random variable X is a variable that might assume any one of several values in trials conducted under identical conditions. Each of the alternative values corresponds to a different outcome. Random variables may be classified broadly into two categories: a discrete random variable and a continuous random variable. A random variable X is discrete if its range forms a discrete (countable) set of real numbers. A random variable X is continuous if its range forms a continuous (uncountable) set of real numbers.

A random variable X is characterized completely by its PDD function which defines the relative likelihood of assuming one value over the others. The PDD function has a different meaning depending on whether the random variable is discrete or continuouse. For discrete distributions, the PDD is defined as the probability of observing a particular outcome. The PDD is denoted by $P(X = x_i)$, where x_i is a possible value that X takes. For the continuous case, the PDD function, $P(X)$ at a value x_i, is not the probability of observing the value x_i. The probability of observing any particular value is zero. To get a probability over an interval $a \leq X \leq b$,

one needs to integrate the PDD function over the interval as follows:

$$F(a \leq X \leq b) = \int_a^b P(X)dX \tag{3.3}$$

Here $F(a \leq X \leq b)$ is the probability of $a \leq X \leq b$.

For two continuouse random variables X and Y, we can define the joint distribution density function PDD

$$F(a \leq X \leq b, c \leq Y \leq d) = \int_a^b \int_c^d P(X,Y)dXdY \tag{3.4}$$

The marginal PDD of X can be obtained by integrating out Y i.e.

$$P(X) = \int_{-\infty}^{\infty} P(X,Y)dY \tag{3.5}$$

In the discrete case, summation signs replace integrals.

3.1.4 Moments of the distribution

We have seen that the probability density distribution can evaluate probability statements about random variable X. Frequently, problems are phrased that require the notion of the average or expected value of X, the measure of the volatility of X etc., not merely a statement of the probability that X will lie in a certain interval. The expectation of a random variable locates its center of mass. For discrete random variables, the expectation is

$$E(X) = \sum_{x=-\infty}^{\infty} x_i P(X = x_i) \tag{3.6}$$

For the continuous case,

$$E(X) = \int_{-\infty}^{\infty} XP(X)dX \tag{3.7}$$

A measure of the volatility of X is found in the expected value of its squared deviation from the center of mass. This is called the variance of X:

$$Var(X) \equiv E\{[X - E(X)]^2\} \tag{3.8}$$

Since $E(X)$ is the center of mass for variable X, the transformed quantity $X' \equiv X - E(X)$ is called the centralized form of X. In other word, the centralized version of a random variable is its deviation from the mean.

3.2 Bayesian Inference Theory

3.2.1 Bayes' theorem

Suppose that $Y = [y_1, \cdots \cdot, y_N]^T$ is a vector of N observations from an experiment whose probability distribution $P(Y \mid \Theta)$ depends on the values of L parameters $\Theta = [\theta_1, \cdots \cdot, \theta_L]^T$ and Θ itself has a probability distribution $P(\Theta)$. Here Θ could be either discrete or continuous. Bayesian theory then tells us:

$$P(Y \mid \Theta)P(\Theta) = P(Y, \Theta) = P(\Theta \mid Y)P(Y). \tag{3.9}$$

Given the observed data Y, the conditional distribution of Θ is

$$P(\Theta \mid Y) = \frac{P(Y \mid \Theta)P(\Theta)}{P(Y)}. \tag{3.10}$$

For a given Y, $P(Y \mid \Theta)$ in the above equation may be regarded not as a function of Y, but a function of Θ. When so regarded it is called the likelihood function of Θ for given Y and can be written as $L(\Theta \mid Y)$. $P(Y)$ can be regarded as a normalization constant, so that Bayes' formula can be written as

$$P(\Theta \mid Y) = \frac{L(\Theta \mid Y)P(\Theta)}{P(Y)} \tag{3.11}$$

where we call $P(\Theta \mid Y)$ the probability density distribution (PDD) posterior to Y. $P(\Theta)$ is the PDD for Θ prior to the data, and $L(\Theta \mid Y)$ is the likelihood function for Θ given Y. Here the likelihood function plays a very important role. It is the function through which the data Y modifies the prior knowledge of Θ. It can therefore be regarded as representing the information about Θ coming from the data.

3.2.2 Application to the linear model in standard Normal inference problems

We consider a linear model as follows:

$$Y = X\Theta + \mathcal{E}. \tag{3.12}$$

where Y is an $N \times 1$ vector of observations, X is an $N \times L$ matrix of known constants, Θ is an $L \times 1$ vector of unknown parameters, and $\mathcal{E}$ is an $N \times 1$ vector of random variables distributed as Normal $N(0, \sigma^2)$ (often referred to as statistical error). Explicitly, the model is

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1L} \\ x_{21} & x_{22} & \cdots & x_{2L} \\ \vdots & \vdots & \vdots & \vdots \\ x_{N1} & x_{N2} & \cdots & x_{NL} \end{bmatrix} \begin{bmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_L \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_1 \\ \vdots \\ \varepsilon_N \end{bmatrix} \tag{3.13}$$

and in particular, for the ith observation,

$$y_i = x_{i1}\theta_1 + x_{i2}\theta_2 + \cdots + x_{iL}\theta_L + \varepsilon_i. \tag{3.14}$$

Given the observations Y, the constant matrix X, and knowing ε_i is a normally distributed random error with zero mean and variance σ^2, we want to find the best estimates for Θ.

First, suppose σ^2 is known. According to Bayesian theory, the posterior PDD of Θ given Y and σ^2 is

$$P(\Theta \mid Y, \sigma^2) \propto L(\Theta \mid Y, \sigma^2) P(\Theta). \tag{3.15}$$

If we don't have any specific information about Θ prior to the observation data Y, we may employ $P(\Theta) \propto constant$ as a non informative reference prior for Θ [1], then

$$P(\Theta \mid Y, \sigma^2) \propto L(\Theta \mid Y, \sigma^2). \tag{3.16}$$

If we have the true model, the difference between the data Y and the model $X\Theta$ is just the noise. Then the likelihood, or the probability that we should obtain the data Y given X and Θ, is [1]

$$\begin{aligned} L(\Theta \mid Y, \sigma^2) &= P(\mathcal{E}) \equiv \frac{1}{(\sqrt{2\pi}\sigma)^N} exp\left[-\frac{1}{2\sigma^2}\mathcal{E}^T\mathcal{E}\right] \\ &= \frac{1}{(\sqrt{2\pi}\sigma)^N} exp\left[-\frac{1}{2\sigma^2}(Y - X\Theta)^T(Y - X\Theta)\right] \qquad (3.17) \\ &= \frac{1}{(\sqrt{2\pi}\sigma)^N} exp\left\{-\frac{1}{2\sigma^2}\left[\nu s^2 + (\Theta - \hat{\Theta})^T X^T X(\Theta - \hat{\Theta})\right]\right\} \end{aligned}$$

where

$$\begin{aligned} \hat{\Theta} &= (X^T X)^{-1} X^T Y, \\ \nu &= N - L, \\ s^2 &= \frac{1}{\nu}(Y - \hat{Y})^T(Y - \hat{Y}), \\ \hat{Y} &= X\hat{\Theta}. \end{aligned}$$

L is the rank of the matrix X. It then follows that:

(a). $\hat{\Theta}$ is a vector of statistics jointly sufficient for Θ if σ^2 is known.

(b). $\hat{\Theta}$ and s^2 are jointly sufficient for (Θ, σ^2) if σ^2 is unknown.

(c). $\hat{\Theta}$ has the multivariate Normal distribution $N_L[\Theta, \sigma^2(X^TX)^{-1}]$.

(d). νs^2 is distributed independently of $\hat{\Theta}$ as $\sigma^2\chi^2_\nu$.

Since σ^2 is known

$$P(\Theta \mid Y, \sigma^2) \propto L(\Theta \mid Y, \sigma^2) \propto exp\left[-\frac{1}{2\sigma^2}(\Theta - \hat{\Theta})^T X^T X(\Theta - \hat{\Theta})\right] \tag{3.18}$$

and

$$\int_R exp\left[-\frac{1}{2\sigma^2}(\Theta - \hat{\Theta})^T X^T X(\Theta - \hat{\Theta})\right] d\Theta = \frac{(\sqrt{2\pi}\sigma)^L}{\mid X^TX \mid^{1/2}}$$

where R is the region $-\infty < \theta_i < \infty$, $i = 1, 2, \cdots, L$. From this we have

$$P(\Theta \mid Y, \sigma^2) = \frac{\mid X^TX \mid^{1/2}}{(\sqrt{2\pi}\sigma)^L} exp\left[-\frac{1}{2\sigma^2}(\Theta - \hat{\Theta})^T X^T X(\Theta - \hat{\Theta})\right] \tag{3.19}$$

Therefore all the relevant inferences about Θ can now be made from the knowledge that the posterior distribution of Θ is the multivariate Normal distribution $N_L[\hat{\Theta}, \sigma^2(X^TX)^{-1}]$. $P(\Theta \mid Y, \sigma^2)$ is maximized at $\hat{\Theta}$, so $\hat{\Theta}$ is our best estimate for Θ with variance equal to $\sigma^2(X^TX)^{-1}$. This is just identical to least-squares fitting for Θ, but Bayesian inference gives a reasonable justification for the procedure.

In some cases the variance σ^2 is unknown, and information about σ^2 coming from the observations is used. In this case the joint PDD of Θ and σ^2 is

$$P(\Theta, \sigma^2 \mid Y) = L(\Theta, \sigma^2 \mid Y)P(\Theta, \sigma^2) \tag{3.20}$$

By using a non informative prior with Θ and $\log\sigma$ approximately independent and locally uniform [1], i.e.

$$P(\Theta, \sigma^2) = P(\Theta)P(\sigma^2) \propto \sigma^{-2} \tag{3.21}$$

we can then follow the same reasoning as when σ^2 is known. The likelihood in this case is given by

$$\begin{aligned} L(\Theta, \sigma^2 \mid Y) &= \frac{1}{(\sqrt{2\pi})\sigma)^N} exp\left[\frac{1}{2\sigma^2}(Y - X\Theta)^T(Y - X\Theta)\right] \\ &= \frac{1}{(\sqrt{2\pi}\sigma)^N} exp\left\{-\frac{1}{2\sigma^2}\left[\nu s^2 + (\Theta - \hat{\Theta})^T X^T X(\Theta - \hat{\Theta})\right]\right\} \end{aligned} \tag{3.22}$$

and thus

$$P(\Theta, \sigma^2 \mid Y) \propto \sigma^{-(N+2)} exp\left\{-\frac{1}{2\sigma^2}\left[\nu s^2 + (\Theta - \hat{\Theta})^T X^T X(\Theta - \hat{\Theta})\right]\right\}. \tag{3.23}$$

Integrating out σ^2 from the above equation and doing a normalization, the marginal posterior distribution function for Θ becomes

$$P(\Theta \mid Y) = \frac{\Gamma[(\nu + L)/2] \mid X^T X \mid^{1/2} s^{-L}}{[\Gamma(1/2)]^L \Gamma[\nu/2](\sqrt{\nu})^L}\left[1 + \frac{(\Theta - \hat{\Theta})^T X^T X(\Theta - \hat{\Theta})}{\nu s^2}\right]^{-(\nu+L)/2} \tag{3.24}$$

$$-\infty < \theta_i < \infty \quad i = 1, 2, \cdots, L$$

which is a multivariate (L-dimensional) Student-t distribution, often denoted as $t_L(\hat{\Theta}, s^2(X^T X)^{-1}, \nu)$. Once again $P(\Theta \mid Y)$ is maximized at $\hat{\Theta}$.

We can also obtain an estimate about the variance of the noise by finding the marginal PDD of σ^2. Integrate out the θ_i's from equation (3.25) and do a normalization. The marginal PDD of σ^2 is then

$$P(\sigma^2 \mid Y) = \frac{(\nu s^2)^{(N-L)/2}}{2^{(N-L)/2}\Gamma((N - L)/2)(\sigma^2)^{(N-L+2)/2}} exp(-\frac{\nu s^2}{2\sigma^2}). \tag{3.25}$$

Letting

$$\chi^2 = \frac{\nu s^2}{\sigma^2} \tag{3.26}$$

we find that

$$P(\chi^2 \mid Y) = \frac{1}{2^{(N-L)/2}\Gamma((N - L)/2)}(\chi^2)^{(N-L-2)/2} exp(-\frac{\chi^2}{2}) \tag{3.27}$$

which is a χ^2 distribution with $\nu = N - L$ degrees of freedom. So σ^2 is distributed a posteriori as $\nu s^2 \chi_\nu^{-2}$. The value of the point which gives the maximum marginal PDD of σ^2 is the best estimate for σ^2.

3.2.3 Linear approximation to non-linear models

Unfortunately, most real experimental data are from non-linear models as follows:

$$Y = G(\Xi, \Theta) + \mathcal{E} \tag{3.28}$$

where Y is an $N \times 1$ vector of observations, Ξ is a set of known constants $\{\xi_1, \ldots, \xi_N\}$, Θ is an $L \times 1$ vector of unknown parameters, and $\mathcal{E}$ is an $N \times 1$ vector of random variables distributed as Normal $N(0, \sigma^2)$. $G(\Xi, \Theta)$ is a non-linear function of Ξ and Θ. To be able to use the theory described in the previous section we have to introduce some reasonable approximations to linearize the non-linear model function. Box and Tiao [1] have shown a very simple method for doing this. In general the specific form of the posterior distributions which we shall obtain for the linear case will often provide reasonably close approximations even when the model functions $G(\Xi, \Theta)$ are nonlinear in Θ. This is because we need only require that the model functions be approximately linear in the region of the parameters space covered by most of the posterior distribution, say within the 95% highest posterior density (H.P.D.) region. For moderate N, this can happen with functions which are highly nonlinear in the parameters when considered over their whole range[1]. Then, in the region where the posterior probability mass is concentrated (the 95% H.P.D. region), we may expand the model functions around the mode $\hat{\Theta}$ ($\hat{\Theta}$ minimizes $\mid Y - G(\Xi, \Theta) \mid^2$), giving

$$y_i = G(\xi_i, \Theta) \cong G(\xi_i, \hat{\Theta}) + \sum_{j=1}^{L} x_{ij}(\theta_j - \hat{\theta}_j) \tag{3.29}$$

where

$$x_{ij} = \left.\frac{\partial G(\xi_i, \Theta)}{\partial \theta_j}\right|_{\theta_j = \hat{\theta}_j} \qquad i = 1, 2, \cdots, N \tag{3.30}$$

which is in the form of a linear model. Thus the posterior distributions found for linear models can now be used to provide a close approximation to the true distribution. This linear approximation will be applied to solve the problems presented in Chapter Five.

Chapter 4. Gibbs Sampling - A Numerical Technique in Applications of Bayesian Inference Methods

4.1 Introduction

In the previous chapter, applications of the Bayesian inference theory to the linear or non-linear model in standard Normal inference problems were presented. We ended up with a joint posterior probability distribution for the unknown parameters $\Theta = \{\theta_1, \theta_2, \ldots, \theta_L\}$ (see equation (3.21) and (3.27)). What we want is to determine the best estimate for each parameter as well as the uncertainty about the estimate. To do this, we have to find the marginal posterior probability distribution for each parameter. Generally for the parameter of interest we can get rid of nuisance parameters from the joint posterior using integration. However there are some technical difficulties in the calculations as we mentioned in the previous chapter. In the last few years there have been a number of advances in numerical and analytic approximation techniques for such calculations[6]-[11], but implementation of these approaches typically requires sophisticated numerical or analytic approximation expertise and possibly specialist software.

Fortunately, there are alternative ways for the calculation of marginal posteriors. These are the sampling-based approaches, which, by contrast, are essentially trivial to implement. The basic idea is to generate a sample of marginal distribution for the interested parameter. Then consequently all the information about the interested parameter can be drawn from the sample. Since we do not have the marginal

distribution function for the parameter this can not be done directly. What we can do is to use some special sampling algorithm to generate the sample from the joint or the conditional posterior probability distributions.

There are three well developed sampling algorithms: the data-augmentation algorithm described by Tanner and Wang[12] and detailed in Tanner's book "Tools for Statistical Inference"[13], the Gibbs sampler algorithm introduced by Geman and Geman[14], and the form of importance-sampling algorithm proposed by Rubin[15] [16]. Among them the Gibbs sampler has been widely taken up in the image-processing literature and in other large-scale models. There is a close relationship between the Gibbs sampler and the data augmentation algorithm which are both iterative Monte Carlo procedures. The importance-sampling algorithm involves a non iterative Monte Carlo integration approach to calculating marginal distributions. Since in this study only the Gibbs sampling will be employed, we will only give a detailed introduction about the Gibbs sampling. For those readers who are interested in the other two techniques, please refer to the references.

4.2 Gibbs Sampling

The Gibbs sampling is a version of the Monte Carlo method based on the Markov chains and is a computer-based technique which allows us to generate the marginal probability density distributions to any degree of accuracy from the conditional probability density distributions. In the procedure the objective is to generate a sample distribution $\pi(x)$ which can not be done directly. However, if we can construct a Markov chain in the same state space, having an equilibrium distribution $\pi(x)$, then by running the chain for a long time, simulated values of the chain can

be used as a basis for summarizing those features of $\pi(x)$ which are of interest. To better understand the Gibbs sampling technique, first a brief introduction about Markov chains is presented.

4.2.1 Markov chains

Suppose X is a random variable which takes only two possible values 0 or 1. $X = 1$ means a certain machine works, $X = 0$ means a certain machine dose not work. Adding the time factor, at time n, if $X_n = 0$ i.e. the machine is down, then suppose

$$P(X_{n+1} = 1 \mid X_n = 0) = p \tag{4.1}$$

and

$$P(X_{n+1} = 0 \mid X_n = 1) = q \tag{4.2}$$

we want to know the limiting probability that a machine is functioning, i.e.

$$\lim_{n \to \infty} P(X_n = 1) \tag{4.3}$$

. Write

$$\begin{aligned} P(X_{n+1} = 0) &= P(X_{n+1} = 0, X_n = 0) + P(X_{n+1} = 0, X_n = 1) \\ &= P(X_{n+1} = 0 \mid X_n = 0)P(X_n = 0) \\ &\quad + P(X_{n+1} = 0 \mid X_n = 1)P(X_n = 1) \\ &= (1-p)P(X_n = 0) + q[1 - P(X_n = 0)] \\ &= (1-p-q)P(X_n = 0) + q \\ &= (1-p-q)[(1-p-q)P(X_{n-1} = 0) + q] + q \\ &= (1-p-q)^2 P(X_{n-1} = 0) + q + q(1-p-q) \end{aligned} \tag{4.4}$$

This implies

$$P(X_n = 0) = (1-p-q)^n P(X_0 = 0) + q\sum_{j=0}^{n-1}(1-p-q)^j \tag{4.5}$$

$$= (1-p-q)^n P(X_0 = 0) + q\left[\frac{1-(1-p-q)^n}{p+q}\right]$$

where $P(X_0 = 0)$ and $P(X_0 = 1)$ are known as the initial probability distribution of X. Since $0 < p < 1$ and $0 < q < 1$, then $0 < p+q < 2$ or $\mid 1-(p+q) < 1 \mid$, it is obvious that

$$\lim_{n\to\infty} P(X_n = 0) = \frac{q}{p+q} \tag{4.6}$$

This implies that we may generate X_n's using the conditional probabilities. For large enough n, the value of X_n generated will be random variable from the marginal.

4.2.2 Gibbs Sampler

We now consider a particular form of Markov chain scheme to help us generate marginal probability distribution samples from the conditional probability distributions. Suppose for a set of random variables $\{x_1, x_2, \cdots, x_N\}$, we know the conditional densities $P(x_i \mid x_1, \cdots, x_{i-1}, x_{i+1}, \cdots, x_N)$, $i = 1, 2, , \cdots, N$. Gibbs sampling then proceeds in the following manner; first arbitrary starting values $x_{1,(0)}, x_{2,(0)}, \cdots, x_{N,(0)}$ are chosen; then random drawings are made from the conditional densities:

$$x_{1,(1)} \quad from \quad P(x_1 \mid x_{2,(0)}, \cdots, x_{i,(0)}, \cdots, x_{N,(0)})$$

$$\vdots$$

$$x_{i,(1)} \quad from \quad P(x_i \mid x_{1,(1)}, \cdots, x_{i-1,(1)}, x_{i+1,(0)}, \cdots, x_{N,(0)})$$

$$\vdots$$

$$x_{N,(1)} \quad from \quad P(x_N \mid x_{1,(1)}, \cdots, x_{i,(1)}, \cdots, x_{N-1,(1)}).$$

After one iteration, we obtain

$$\{x_{1,(1)}, \cdots, x_{i,(1)}, \cdots, x_{N,(1)}\}$$

and after K iterations, we obtain

$$\{x_{1,(K)}, \cdots, x_{i,(K)}, \cdots, x_{N,(K)}\}$$

It has been shown [17, 18] that for large enough K, $x_{i,(K)}$, $i = 1, 2, \cdots, N$, will be a random variable drawn from the marginal densities

$$P(x_i) = \int P(x_1, \cdots, x_i, \cdots, x_N) dx_1 \cdots dx_{i-1} dx_{i+1} \cdots dx_N.$$

$$i = 1, 2, \cdots, N$$

Thus in order to estimate the marginal densities of x_i, $P(x_i)$, $i = 1, 2, \cdots, N$, we need to generate several values of $x_{i,(K)}$

$$x^1_{1,(K)}, x^2_{1,(K)}, \cdots, x^{M_s}_{1,(K)};$$
$$x^1_{2,(K)}, x^2_{2,(K)}, \cdots, x^{M_s}_{2,(K)};$$
$$\vdots$$
$$x^1_{N,(K)}, x^2_{N,(K)}, \cdots, x^{M_s}_{N,(K)};$$

Then

$$X_i = \{x^1_{i,(K)}, x^2_{i,(K)}, \cdots, x^{M_s}_{i,(K)}\},$$

$$i = 1, 2, \cdots, N$$

are random variable samples of size M_s from the marginal densities $P(x_i)$, $i = 1, 2, \cdots, N$. The determination of K depends upon the particular application. This

will be discussed further in Chapter 5. By plotting the marginal histograms, we can draw information from them, such as the mean, the variance, as well as the position corresponding to the maximum marginal probability, etc.. Of course the larger the size of the sample the more accurate the information that we can draw from it.

Chapter 5. Spectral Analysis and Parameter Estimation in Discrete Time Signals

5.1 Introduction

5.1.1 The problem

In brief, spectral analysis and parameter estimation in discrete time series is the fitting of model functions, most commonly the sinusoid functions, to a data set in order to determine the parameters, such as frequencies and amplitudes (or phases). There are different ways to formulate the problem. We must be very clear what we already know and what we want to know. For example, if we know the model functions for the data are sinusoidal and how many periodicities there are, we might want to estimate their parameters; alternatively we might know there is a number of periodicities but we do not know exactly how many, we want to estimate that number; or even a more complicate case where there are a few probable model functions, we want to know which one is the most probable. The problem is old and well known. Comprehensive histories of the spectral analysis methods are given by Robinson[19] and Marple[20]. In what follows, a brief historical introduction of these methods is presented.

5.1.2 Historical review

The first significant advance in the spectral analysis and parameter estimation problem took place in the early nineteenth century, when two methods came into being:

the use of probability theory, and the use of the Fourier transform. It was Laplace[21] who formulated the probabilistic methods in some generality in the late 18th century. Then the method was applied by Legendre[22] and Gauss[23] who first used the method of least squares to estimate model parameters in noisy data. Some idealized model function is postulated to represent the signal and the criterion of minimizing the sum of the squares of the residuals is used to estimate the parameters in the model. This approach is usually too restrictive. It assumes more about the problem than is really known. Usually, we don't know which model functions will fit the data best, or how many parameters are required to fix the model functions.

The spectral method of dealing with the problem also occurred in the early 19th century. The Fourier transform is one of the most powerful tools. Since Cooley and Tukey [24] introduced the fast Fourier transform technique in 1965, followed by the rapid development of computers, the discrete Fourier transform has been widely used in spectral analysis. In a typical situation in spectral analysis, one is attempting to gain information about an underlying signal in a time series which has been sampled discretely over a finite time, and which is compounded by the addition of noise and possible distortion. In this situation, the Fourier transform consists of the signal transform plus a noise transform. As such, the relationship of the Fourier transform to the underlying time series is a nontrivial technical problem. One of the central questions then becomes: under what conditions is the discrete Fourier transform an optimal frequency estimator?

Bretthorst [4] has shown that the discrete Fourier transform is able to give optimal frequency estimates to a signal with noise if certain conditions are met:

1. The number of data values N is large.

2. There is no constant component in the data.

3. There is no evidence of a low frequency.

4. The data contain only one frequency.

5. The frequency must be stationary.

6. The noise is white.

If any of these conditions are violated, the discrete Fourier transform may give misleading or simply incorrect results in light of more realistic models. This is not because the discrete Fourier transform is wrong, but because it is attempting to answer the wrong question. In practice, as long as certain conditions are satisfied, the zero-padded discrete Fourier-transform power spectrum is the best estimator of the frequencies in the signal. Bayesian analysis gives added justification to this [26]. The discrete Fourier-transform power spectrum is essentially the logarithm of the posterior probability of a single stationary sinusoidal frequency, or the logarithm of the posterior probability of a series of well-separated sinusoids.

Schuster[25] introduced the periodogram near the beginning of this century as a method of detecting a periodicity and estimating its frequency. The periodogram is essentially the squared magnitude of the discrete Fourier transform of the data. It has all the drawbacks of the Fourier transform. Since Jaynes [27] derived the periodogram directly from the principles of probability theory and demonstrated it to be a "sufficient statistic" for inferences about a single stationary frequency in a time-sampled data set when a Gaussian probability distribution is assigned for the noise, much attention has been paid to the relationship between Bayesian inference, spectral analysis and parameter estimation. Following Jaynes' work, Bretthorst [4],

[28]-[32] has done some excellent work in this area and his approach has been applied to various NMR signals [26], [33]-[36]. (Since we will give quantitative comparisons between our results and Bretthorst's results, a detailed background introduction to Bretthorst's method will been given in next section.) However the calculations in his work are quite complicated and the method can only give the joint probabilities of all parameters. The only way to find the marginal probability for a specific parameter is to integrate out all the nuisance parameters. When more and more parameters are involved, there is usually no analytical result for the integral and the numerical calculation of the integral can be very complicated and difficult.

5.2 Stationary Sinusoid Signals with White Noise

5.2.1 Numerical model of the problem

The time series $y(t)$ we are considering is postulated to contain a signal $f(t)$ plus a noise $e(t)$. The basic model is : we have recorded a discrete data set $D = \{d_1, d_2, \ldots . d_N\}^T$ sampled from $y(t)$ at discrete time $\{t_1, t_2, \ldots, t_N\}^T$ with a model equation:

$$d_i = y(t_i) = f(t_i) + e(t_i) \qquad 1 \leq i \leq N \tag{5.1}$$

Here we call $f(t)$ the model function. Different models corresponds to different choices of the model function $f(t)$. In this section we suppose the model function is known(we will talk about model selection in 5.4) as a single or multiple stationary harmonic signal. The function model for this signal can be written as:

$$f(t_i) = \sum_{j=1}^{M} [B_{j1} cos(\omega_j t_i) + B_{j2} sin(\omega_j t_i)] (1 \leq i \leq N) \tag{5.2}$$

which has $3M$ parameters $\{B_{j1}, B_{j2}, \omega_j\}, j = 1, 2, \ldots, M$. The white noise means $e(t_i)$ $i = 1, 2, \ldots, N$ is independently normaly distributed, $e(t_i) \sim N(0, \sigma^2)$, i.e. the magnitude of the noise at each datum point has a Gaussian distribution with zero mean and variance σ^2. Now based on this model and the data set D with σ^2 known or unknown, we want to make the best estimates for all the $3M$ parameters using the Bayesian inference methods described in 3.2.

5.2.2 The likelihood and the prior

According to the Bayesian theorem, given the model function and the data D, the joint posterior probability of the parameters is:

$$P(\{B_{j1}\}, \{B_{j2}\}, \{\omega_j\} \mid D, \sigma^2) = \tag{5.3}$$
$$P(\{B_{j1}\}, \{B_{j2}\}, \{\omega_j\}) L(\{B_{j1}\}, \{B_{j2}\}, \{\omega_j\} \mid D, \sigma^2)$$

when σ^2 is known or

$$P(\{B_{j1}\}, \{B_{j2}\}, \{\omega_j\}, \sigma^2 \mid D) \tag{5.4}$$
$$= P(\{B_{j1}\}, \{B_{j2}\}, \{\omega_j\}, \sigma^2) L(\{B_{j1}\}, \{B_{j2}\}, \{\omega_j\}, \sigma^2 \mid D)$$

when σ^2 is unknown.

where

$$\{B_{j1}\} = \{B_{11}, B_{21}, \ldots, B_{M1}\}, \tag{5.5}$$

$$\{B_{j2}\} = \{B_{12}, B_{22}, \ldots, B_{M2}\},$$

$$\{\omega_j\} = \{\omega_1, \omega_2, \ldots, \omega_M\}$$

If we have the true model, the difference between the data d_i and the function model $f(t_i)$ is just the noise $e(t_i)$. Then the likelihood, or the probability, that we should obtain the data D given the model and the parameters is the probability of the noise. As we know

$$P(e(t_1), e(t_2), \ldots, e(t_N) \mid \sigma^2) \propto exp[-\frac{1}{\sqrt{2\pi\sigma^2}} \sum_{i=1}^{N} e^2(t_i)] \tag{5.6}$$

so

$$\begin{aligned} &L(\{B_{j1}\}, \{B_{j2}\}, \{\omega_j\} \mid D, \sigma^2) \\ &\propto exp\left\{-\tfrac{1}{2\sigma^2} \textstyle\sum_{i=1}^{N}[d_i - \sum_{j=1}^{M}(B_{j1}cos(\omega_j t_i) + B_{j2}sin(\omega_j t_i))]^2\right\} \end{aligned} \tag{5.7}$$

when σ^2 is known, and

$$\begin{aligned} &L(\{B_{j1}\}, \{B_{j2}\}, \{\omega_j\}, \sigma^2 \mid D) \\ &\propto \tfrac{1}{\sigma^N} exp\left\{-\tfrac{1}{2\sigma^2} \textstyle\sum_{i=1}^{N}[d_i - \sum_{j=1}^{M}(B_{j1}cos(\omega_j t_i) + B_{j2}sin(\omega_j t_i))]^2\right\} \end{aligned} \tag{5.8}$$

when σ^2 is unknown.

If we do not have any information about the parameters and σ^2(if σ^2 is unknown) before we analyse the data, we may use the non informative prior for $P(\{B_{j1}\}, \{B_{j2}\}, \{\omega_j\})$, so

$$P(\{B_{j1}\}, \{B_{j2}\}, \{\omega_j\}) \propto const \tag{5.9}$$

and

$$P(\{B_{j1}\},\{B_{j2}\},\{\omega_j\},\sigma^2) \tag{5.10}$$
$$= P(\{B_{j1}\},\{B_{j2}\},\{\omega_j\})P(\sigma^2) \propto \sigma^{-2}$$

Then the joint posterior probability distribution is

$$P(\{B_{j1}\},\{B_{j2}\},\{\omega_j\} \mid D,\sigma^2) \tag{5.11}$$
$$\propto exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{N}[d_i - \sum_{j=1}^{M}(B_{j1}cos(\omega_j t_i) + B_{j2}sin(\omega_j t_i))]^2\right\}$$

when σ^2 is known, and

$$P(\{B_{j1}\},\{B_{j2}\},\{\omega_j\},\sigma^2 \mid D) \tag{5.12}$$
$$\propto \sigma^{-(N+2)}exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{N}[d_i - \sum_{j=1}^{M}(B_{j1}cos(\omega_j t_i) + B_{j2}sin(\omega_j t_i)]^2\right\}$$

when σ^2 is unknown

Now the question is how to make the estimates about each parameter from the joint posterior? In the next section we will see how the Gibbs sampling and linear approximation help us to answer this question.

5.2.3 Why do we need Gibbs sampling?

If just Bayesian theory is applied to the model described above, we end up with the joint posterior probability distribution of $(B_{j1}, B_{j2}, \omega_j)$, $j = 1, 2, \cdots, M$ shown in equation(5.12) and (5.13). To find the marginal distributions, we have to integrate out all the nuisance parameters from the joint distribution. For example, to find $P(\omega_j \mid D, \sigma^2)$ we have to evaluate the integral:

$$P(\omega_j \mid D, \sigma^2) \propto \int exp\{-\frac{1}{2\sigma^2}\sum_{i=1}^{N}[d_i - \sum_{j=1}^{M}(B_{j1}cos(\omega_j t_i) + B_{j2}sin(\omega_j t_i))]^2\} d\{B_{j1}\}d\{B_{j2}\}d\omega_1 \cdots d\omega_{j-1}d\omega_{j+1} \cdots d\omega_M \quad (5.13)$$

where

$$d\{B_{j1}\} = dB_{11}dB_{21} \cdots dB_{M1} \quad (5.14)$$
$$d\{B_{j2}\} = dB_{12}dB_{22} \cdots dB_{M2}.$$

As we can see this integration is not easy to compute since the model functions are non-linear in the parameters. By using the linear approximation we have to find $(\{\hat{B}_{j1}\}, \{\hat{B}_{j2}\}, \{\hat{\omega}_j\})$ which minimizes the exponent in equation (5.13), then expand the model function around the minimum up to a linear order. The multivariate minimization, however, is not trivial and sometimes only the local minimum can be found. In order to avoid computing the integrals or the multivariate minimization, Gibbs sampling can give us an alternative way to find the marginal distribution for each parameter by dealing with just a single parameter at each sampling step. Since only the frequency is non-linear in our model function, we only need to do linear approximation for each frequency which only involves a univariate minimization. We will see that this method is very successful.

Before we go to the numerical procedures of our method, a brief introduction about Bretthorst's approach is presented in the next section since we quantitatively compare our results with his results in the numerical examples.

5.2.4 Bretthorst's approach

Starting with a general model function

$$f(t) = \sum_{j=1}^{m} B_j G_j(t, \Omega) \tag{5.15}$$

$G_j(t, \Omega)$ are functions of a parameter set $\Omega = \{\omega_1, \omega_2, \ldots, \omega_k\}$ which may represent frequencies, phase shifts, decay rates, etc.

If we substitute the model function, equation(5.15), into the expression for the likelihood,

$$L(\{B\}, \{\omega\}, \sigma) = \sigma^{-N} \exp\{-\frac{1}{2\sigma^2} \sum_{i=1}^{N} [d_i - f(t_i)]^2\} \tag{5.16}$$

it then takes the form,

$$L(\{B\}, \{\omega\}, \sigma) \propto \sigma^{-N} \exp\{-\frac{NQ}{2\sigma^2}\} \tag{5.17}$$

where

$$Q \equiv \bar{d^2} - \frac{2}{N} \sum_{j=1}^{m} \sum_{i=1}^{N} B_j d_i G_j(t_i) + \frac{1}{N} \sum_{j=1}^{m} \sum_{k=1}^{m} g_{jk} B_j B_k \tag{5.18}$$

$$g_{jk} = \sum_{i=1}^{N} G_j(t_i) G_k(t_i) \tag{5.19}$$

In the above, $\bar{d^2} = \frac{1}{N} \sum_{i=1}^{N} d_i^2$.

In order to carry out the integration over the nuisance parameters, it is necessary to make the matrix g_{jk} diagonal, effectively introducing new model functions $H_j(t_i)$ which are orthogonal [4]. This diagonalization process yields a new expression for the model function,

$$f(t) = \sum_{k=1}^{m} A_k H_k(t) \tag{5.20}$$

where the new amplitudes A_k are related to the B_k, by

$$A_k = \sqrt{\lambda_k} \sum_{j=1}^{m} B_j e_{kj} \tag{5.21}$$

and where e_{kj} represents the jth component of the kth normalized eigenvector of g_{jk}, with λ_l as the lth eigenvalue.

Substituting these expressions into the expression for the likelihood function, and defining

$$h_j = \sum_{i=1}^{N} d_i H_j(t_i) \tag{5.22}$$

to be the projection of the data onto the orthonormal model functions H_j, we can then proceed to perform the m integrations over the A_j to obtain,

$$L(\{\omega\},\sigma) \propto \sigma^{-N+m} \exp\left\{-\frac{N\bar{d^2} - m\bar{h^2}}{2\sigma^2}\right\} \tag{5.23}$$

with

$$\bar{h^2} = \frac{1}{m}\sum_{j=1}^{m} h_j^2 \tag{5.24}$$

representing the mean-square of the observed projections.

If σ^2 is known, and assuming that we have no prior information, the joint posterior probability function of the $\{\omega\}$ parameters, conditional on the data and our knowledge of σ is given by

$$P(\{\omega\}|D,I) \propto \exp\{\frac{m\bar{h^2}}{2\sigma^2}\} \tag{5.25}$$

If there is no prior information about the noise, then σ is a nuisance parameter and must also be eliminated by integrating it out. Using the Jeffreys prior [1] $1/\sigma$ and integrating over σ gives

$$P(\{\omega\}|D,I) \propto \left[1 - \frac{m\bar{h^2}}{N\bar{d^2}}\right]^{\frac{m-N}{2}} \tag{5.26}$$

This has the form of the "Student t-distribution".

As well as determining the values of the $\{\omega\}$ parameters for which the posterior probability is a maximum, it is also desirable to compute the variances associated with these parameters. If we assume the case where σ is known, and let $\{\hat{\omega}\}$ represent the values of $\{\omega\}$ for which the posterior is a maximum, then we can expand

the function $\bar{h^2}$ in a Taylor series, such that the posterior probability function is proportional to

$$\exp\Big\{-\sum_{j,k=1}^{r}\frac{b_{jk}}{2<\sigma^2>}\Delta_j\Delta_k\Big\}$$

with b_{jk} defined as

$$b_{jk}\equiv-\frac{m}{2}\frac{\partial^2\bar{h^2}}{\partial\omega_j\partial\omega_k} \tag{5.27}$$

and $\Delta_j\equiv\hat{\omega}_j-\omega_j$ for the single-frequency case.

In calculating the mean and standard deviations for the $\{\omega\}$ parameters, the Gaussian integrals are evaluated by first changing to the orthogonal variables as was done above with the amplitudes. The new variables are obtained from the eigenvalues and eigenvectors of b_{jk}. If we let u_{jk} represent the kth component of the jth eigenvector of b_{jk}, and let v_j represent the eigenvalue, then the orthogonal variables are given by,

$$s_j=\sqrt{v_j}\sum_{k=1}^{r}\Delta_k u_{kj}\qquad\Delta_j=\sum_{k=1}^{r}\frac{s_k u_{jk}}{\sqrt{v_k}} \tag{5.28}$$

From these, the variance γ_k^2 of the posterior distribution for ω_k is

$$\gamma_k^2\equiv<\sigma^2>\sum_{j=1}^{r}\frac{u_{jk}^2}{v_j} \tag{5.29}$$

Thus the estimated ω_j parameters are

$$(\omega_j)_{est}=\hat{\omega}_j\pm\gamma_j \tag{5.30}$$

where $\hat{\omega}_j$ is the location of the maximum of the probability distribution as a function of the $\{\omega\}$ parameter.

One can also show that the expectation values of the parameters $\{A\}$ are given by $<A_j>=h_j$, which when transformed to the parameters $\{B\}$ become

$$<B_k>=\sum_{j=1}^{m}\frac{h_j e_{jk}}{\sqrt{\lambda_j}} \tag{5.31}$$

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

$$\sigma^2 \sum_{j=1}^{m} \frac{e_{jk}e_{jk}}{\lambda_j} \tag{5.32}$$

5.2.5 Numerical procedures

Let's consider a single harmonic-frequency model described by,

$$d_i = B_1 cos(\omega t_i) + B_2 sin(\omega t_i) + \varepsilon_i \tag{5.33}$$

$$\varepsilon_i \sim N(0, \sigma^2) \quad i = 1, 2, \cdots, N. \tag{5.34}$$

Suppose B_2, and ω are given, leaving B_1 as the only unknown linear parameter. According to the theory in Section 3.2.2 , we know that when σ^2 is known $P(B_1 \mid B_2, \omega, D, \sigma^2)$ is a univariate Normal distribution $N_1(\hat{B}_1, \sigma^2(X_1^T X_1)^{-1})$, where

$$\hat{B}_1 = \frac{\sum_{i=1}^{N} y_i^{(1)} cos(\omega t_i)}{\sum_{i=1}^{N} cos^2(\omega t_i)}, \quad X_1 = \begin{bmatrix} cos(\omega t_1) \\ cos(\omega t_2) \\ \vdots \\ cos(\omega t_N) \end{bmatrix}, \tag{5.35}$$

$$y_i^{(1)} = d_i - B_2 sin(\omega t_i) \tag{5.36}$$

When σ^2 is unknown, $P(B_1 \mid B_2, \omega, D)$ is a univariate Student-t distribution $t_1(\hat{B}_1, s_1^2(X_1^T X_1)^{-1}, N-1)$ where

$$s_1^2 = \frac{1}{N-1}(Y^{(1)} - \hat{Y}^{(1)})^T (Y^{(1)} - \hat{Y}^{(1)}) \tag{5.37}$$

with

$$Y^{(1)} = [y_1^{(1)}, y_2^{(2)}, \cdots, y_N^{(1)}]^T \quad \hat{Y}^{(1)} = X_1 \hat{B}_1. \tag{5.38}$$

Similarly given B_1 and ω, $P(B_2 \mid B_1, \omega, D, \sigma^2)$ is a univariate Normal distribution $N_1(\hat{B}_2, \sigma^2(X_2^T X_2)^{-1})$, where

$$\hat{B}_2 = \frac{\sum_{i=1}^N y_i^{(2)} sin(\omega t_i)}{\sum_{i=1}^N sin^2(\omega t_i)}, \quad X_2 = \begin{bmatrix} sin(\omega t_1) \\ sin(\omega t_2) \\ \vdots \\ sin(\omega t_N) \end{bmatrix}, \tag{5.39}$$

$$y_i^{(2)} = d_i - B_1 cos(\omega t_i) \tag{5.40}$$

and whe σ^2 is unknown, $P(B_2 \mid B_1, \omega, D)$ is a univariate Student-t distribution $t_1(\hat{B}_2, s_2^2(X_2^T X_2)^{-1}, N-1)$ where

$$s_2^2 = \frac{1}{N-1}(Y^{(2)} - \hat{Y}^{(2)})^T (Y^{(2)} - \hat{Y}^{(2)}) \tag{5.41}$$

with

$$Y^{(2)} = [y_1^{(2)}, y_2^{(2)}, \cdots, y_N^{(2)}]^T, \quad \hat{Y}^{(2)} = X_2 \hat{B}_2. \tag{5.42}$$

When B_1 and B_2 are given, ω is a non-linear parameter. We then have to linearize the function with respect to ω. By using the approximation presented in Section 3.2.3, we have that

$$\begin{aligned} B_1 cos(\omega t_i) + B_2 sin(\omega t_i) \;&\cong\; B_1 cos(\hat{\omega} t_i) + B_2 sin(\hat{\omega} t_i) \\ &+\; [-B_1 t_i sin(\hat{\omega} t_i) + B_2 t_i cos(\hat{\omega} t_i)](\omega - \hat{\omega}) \end{aligned} \tag{5.43}$$

where $\hat{\omega}$ minimizes $\sum_{i=1}^{N}[d_i - B_1 cos(\omega t_i) - B_2 sin(\omega t_i)]^2$ with B_1 and B_2 being given. Then when σ^2 is known, $P(\omega \mid B_1, B_2, D, \sigma^2)$ is also a univariate Normal distribution $N_1(\hat{\omega}, \sigma^2(X_\omega^T X_\omega)^{-1})$, where

$$X_\omega = \begin{bmatrix} -B_1 t_1 sin(\hat{\omega} t_1) + B_2 t_1 cos(\hat{\omega} t_1) \\ -B_1 t_2 sin(\hat{\omega} t_2) + B_2 t_2 cos(\hat{\omega} t_2) \\ \vdots \\ -B_1 t_N sin(\hat{\omega} t_N) + B_2 t_N cos(\hat{\omega} t_N) \end{bmatrix}. \tag{5.44}$$

when σ^2 is unknown, $P(\omega \mid B_1, B_2, D)$ is a univariate Student-t distribution $t_1(\hat{\omega}, s_\omega^2(X_\omega^T X_\omega)^{-1}$, 1) with

$$s_\omega^2 = \frac{1}{(N-1)}(D - \hat{D})^T(D - \hat{D}) \tag{5.45}$$

and

$$\hat{D} = [\hat{f}(t_1), \hat{f}(t_2), \cdots, \hat{f}(t_N)]^T \tag{5.46}$$
$$\hat{f}(t_i) = B_1 cos(\hat{\omega} t_i) + B_2 sin(\hat{\omega} t_i)$$

In our numerical calculation when σ^2 is known we start with a set of initial guesses $(B_{1,(0)}, B_{2,(0)}, \omega_{(0)})$, and the Gibbs sampler generates random numbers as follows:

$$\begin{aligned} B_{1,(1)} \quad & from \quad P(B_1 \mid B_{2,(0)}, \omega_{(0)}, D, \sigma^2) \\ B_{2,(1)} \quad & from \quad P(B_2 \mid B_{1,(1)}, \omega_{(0)}, D, \sigma^2) \\ \omega_{(1)} \quad & from \quad P(\omega \mid B_{1,(1)}, B_{2,(1)}, D, \sigma^2). \end{aligned} \tag{5.47}$$

After one iteration, we get $(B_{1,(1)}, B_{2,(1)}, \omega_{(1)})$. After K iterations , we obtain $(B_{1,(K)}, B_{2,(K)}, \omega_{(K)})$. For a large enough K, $B_{1,(K)}$ can be considered as a random variable from $P(B_1 \mid D, \sigma^2)$; $B_{2,(K)}$ a random variable from $P(B_2 \mid D, \sigma^2)$; and $\omega_{(K)}$ a random variable from $P(\omega \mid D, \sigma^2)$. Repeat this M_s times we obtain three random number samples of the same size M_s, $(B^1_{1,(K)}, B^2_{1,(K)}, \cdots, B^{M_s}_{1,(K)})$ from the marginal distribution $P(B_1 \mid D, \sigma^2)$, $(B^1_{2,(K)}, B^2_{2,(K)}, \cdots, B^{M_s}_{2,(K)})$ from $P(B_2 \mid D, \sigma^2)$ and $(\omega^1_{(K)}, \omega^2_{(K)}, \cdots, \omega^{M_s}_{(K)})$ from $P(\omega \mid D, \sigma^2)$. From each sample, all of the estimates about the corresponding parameter can be found, such as the most probable value for the parameter, the mean value, the variances with respect to the most probable value and the mean, etc.. For multiple harmonic frequency signals, the procedures are analogous to the above.

When σ^2 is unknown, we do the same thing as above except that the random numbers are generated from the Student-t distribution. However in this case we must also obtain an estimate for the noise level or the noise variance. Since the marginal posterior probability of σ^2 is $\nu s^2 \chi_\nu^{-2}$, a first set of random numbers $\{x_i\}$ are generated from a chi-square random number generator. Then a set of random numbers $\{\sigma_i^2\}$ which distribute as $\nu s^2 \chi_\nu^{-2}$ can be calculated from $\{x_i\}$ as $\{\sigma_i^2\} = \nu s^2 \{x_i\}^{-1}$. Once again the best estimate for σ^2 can be made from this random numbers sample.

Before we discuss the computer simulated examples, there is something we need to say about the display of the results. The usual way to display the results from a spectral analysis is in the form of a power spectral density. As we know, in Fourier transform spectroscopy this is typically taken as the squared magnitude of the discrete Fourier transform of the data. Here, in order to display our results in the form of a power spectral density, we need to give some attention to the definition

of the power spectral density. The power spectral density shows how much power is contained in a unit frequency. Following Bretthorst's [4] definition, when σ^2 is known the power spectral density is related to the joint posterior probability in the single-frequency model as

$$\hat{p}(\omega) = \frac{N}{2}\int (B_1^2 + B_2^2) P(B_1, B_2, \omega \mid D, \sigma^2) dB_1 dB_2 \tag{5.48}$$

Performing the integrals over B_1 and B_2, we get

$$\hat{p}(\omega) \cong \left[2\sigma^2 + \frac{N}{2}(\hat{B}_1 + \hat{B}_2)\right] P(\omega \mid D, \sigma^2) \tag{5.49}$$

When σ^2 is unknown, we find

$$\hat{p}(\omega) \cong \left[\frac{2(N-2)}{N-4}s^2 + \frac{N}{2}(\hat{B}_1 + \hat{B}_2)\right] P(\omega \mid D) \tag{5.50}$$

where the following approximations have been used,

$$\sum_{i=1}^{N} cos^2(\omega t_i) = \frac{N}{2} + \frac{1}{2}\sum_{i=1}^{N} cos(2\omega t_i) \cong \frac{N}{2} \tag{5.51}$$

$$\sum_{i=1}^{N} sin^2(\omega t_i) = \frac{N}{2} - \frac{1}{2}\sum_{i=1}^{N} cos(2\omega t_i) \cong \frac{N}{2} \tag{5.52}$$

$$\sum_{i=1}^{N} cos^2(\omega t_i) \sum_{i=1}^{N} sin^2(\omega t_i) - \left[\sum_{i=1}^{N} cos(\omega t_i) sin(\omega t_i)\right]^2 \cong \sum_{i=1}^{N} cos^2(\omega t_i) \sum_{i=1}^{N} sin^2(\omega t_i) \tag{5.53}$$

Since

$$\sum_{i=1}^{N} cos(\omega t_i) sin(\omega t_i) = \frac{1}{2}\sum_{i=1}^{N} sin(2\omega t_i) \ll \frac{N}{2} \tag{5.54}$$

If there is more than one frequency in the signal, it won't be easy to do the integral in equation (5.48) and it will only give the joint power spectral density. For multiple-frequency signals only the line power spectral density was computed in Bretthorst's [4] work, which is the indication of the energy carried by each frequency line. In our calculation, since we have already found our best estimates for each frequency and the corresponding amplitudes, we can always use equation (5.49) or (5.50) to find the power spectral density for each frequency. For example, suppose there are two frequencies in the signal,

$$f(t_i) = B_{11}cos(\omega_1 t_i) + B_{12}sin(\omega_1 t_i) + B_{21}cos(\omega_2 t_i) + B_{22}sin(\omega_2 t_i) \tag{5.55}$$

Using Bayesian theory and Gibbs sampling, the marginal posterior probabilities, $P(\omega_1 \mid D, \sigma^2)$ and $P(\omega_1 \mid D, \sigma^2)$, are found and $(\hat{B}_{11}, \hat{B}_{12}, \hat{\omega}_1)$ and $(\hat{B}_{11}, \hat{B}_{12}, \hat{\omega}_1)$ are our best estimates for $(B_{11}, B_{12}, \omega_1)$ and $(B_{21}, B_{22}, \omega_2)$ from the corresponding posterior probabilities. The power spectral densities for ω_1 and ω_2 are

$$\hat{p}(\omega_1) \cong \left[2\sigma^2 + \frac{N}{2}(\hat{B}_{11} + \hat{B}_{12})\right] P(\omega_1 \mid D, \sigma^2) \tag{5.56}$$

$$\hat{p}(\omega_2) \cong \left[2\sigma^2 + \frac{N}{2}(\hat{B}_{21} + \hat{B}_{22})\right] P(\omega_2 \mid D, \sigma^2) \tag{5.57}$$

In the next section, we will present some numerical examples to show how well this technique works.

5.2.6 Computer simulated examples

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

$$\begin{aligned} d_i &= 0.001 + cos(0.3t_i + 1) + \varepsilon_i \\ &= 0.001 + 0.5403cos(0.3t_i) - 0.8415sin(0.3t_i) \end{aligned} \tag{5.58}$$

$$(\omega = 0.3,\ \ B_1 = 0.5403,\ \ B_2 = -0.8415)$$

Here t_i runs over the symmetric interval $-T$ to T in $(2T + 1) = 512$ integer steps and $\varepsilon_i \sim N(0, 1)$. To ensure that the data have zero mean value the average value of the data is computed and subtracted from each data point. Figure 5.1 is a plot of this computer simulated time series, and Figure 5.2 shows the posterior probability density for ω generated from samples of size 10^3 with 40 bins. The sequence length K in Gibbs sampling for this calculation is five. Different values for K have been tested and it turns out that five is a good choice for both accuracy and CPU time saving. The starting values for the parameters didn't affect the results significantly since after a few iteration steps the process is relaxed and stable. The only effect of the starting values may be on the K value. If the starting values are close to the true values, then the value of K does not have to be large. Further investigation about this problem will be conducted.

The best estimates, i.e. the values which maximize the marginal posterior probabilities, are tabulated in Table 5.1 together with the true values.

Table 5.1: The Best Estimates of Parameters For A Single Frequency Model

Parameters	True Values	Estimated Values	Estimated Variances
ω	0.3	0.29999428	0.2×10^{-6}
B_1	0.5403	0.546	0.004
B_2	-0.8415	-0.777	0.004

The estimated variances for B_1, B_2, ω are calculated from the random number samples $(B^1_{1,(K)}, B^2_{1,(K)}, \cdots, B^{M_s}_{1,(K)})$, $(B^1_{2,(K)}, B^2_{2,(K)}, \cdots, B^{M_s}_{2,(K)})$ and $(\omega^1_{(K)}, \omega^2_{(K)}, \cdots, \omega^{M_s}_{(K)})$ respectivly as follows:

$$\sigma^2_{B_1} = \frac{1}{10^3}\sum_{m_s=1}^{10^3}(B^{m_s}_{1,(K)} - \hat{B}_1)^2, \quad \sigma^2_{B_2} = \frac{1}{10^3}\sum_{m_s=1}^{10^3}(B^{m_s}_{2,(K)} - \hat{B}_2)^2, \quad \sigma^2_{\omega} = \frac{1}{10^3}\sum_{m_s=1}^{10^3}(\omega^{m_s}_{(K)} - \hat{\omega})^2. \tag{5.59}$$

As we can see, the computed values are very close to the true values, especially the frequency ω which is the most important parameter in spectral analysis and parameter estimation. Comparing our estimate for ω

$$\omega_{est} = 0.2999 \pm 0.4 \times 10^{-3} \tag{5.60}$$

with Bretthorst's estimate [4]

$$\omega_{est} = 0.2997 \pm 0.6 \times 10^{-3} \tag{5.61}$$

for the same example we can see the difference in the level of accuracy. Figure 5.3 is the power spectral density calculated by using equation (5.56). It is very sharply peaked at the true frequency. This example gives us an indication of the level of accuracy that can be obtained with this approach.

As our second example, we consider two close frequencies:

$$\begin{aligned} d_i &= cos(0.3t_i + 1) + cos(0.31t_i + 2)\varepsilon_i \qquad (5.62) \\ &= 0.5403cos(0.3t_i) - 0.8415sin(0.3t_i) - 0.4161cos(0.31t_i) - 0.9093sin(0.31t_i) \end{aligned}$$

The time interval and the size of the sample are the same as in the first example. The sequence length K is 10. Figure 5.4 shows the time series, and Table 5.2 contains the estimated values, variances and true values for (B_1, ϕ_1, ω_1) and (B_1, ϕ_1, ω_1).

Here again the computed values are all very close to the true values and the two frequency lines are clearly resolved. This is also clearly evident from the power spectral density, shown in Figure 5.5. In Figure 5.6 the Fourier transform power spectral density for this signal is shown. It can be seen that the FFT shows only a single peak in between the two frequencies. Actually with the Fourier transform when the separation of two frequencies is less than the Nyquist step $\mid \omega_1 - \omega_2 \mid < 2\pi/N$, the two frequencies are indistinguishable. This is simply because there are no sample points in between the two frequencies in the frequency domain. Even when $\mid \omega_1 - \omega_2 \mid > 2\pi/N$ theoretically the two frequencies can be distinguished. However, if $\mid \omega_1 - \omega_2 \mid$ is not large enough the resolution will be very poor. Figure 5.7 is the FFT of a two-frequency signal separated by $1.5 \times (2\pi/N)$. It is hard to tell where the two frequencies are located. This is just the inherent problem of the discrete Fourier transform. In this example the two frequencies are separated by 0.01, which is less than $2\pi/N = 2\pi/512 = 0.0123$. There is no way by using FFT that one can resolve the two frequencies, however Bayesian inference and Gibbs sampling give us very good results with high accuracy.

Table 5.2: The Best Estimates of Parameters For a Two Close Frequency Model

Parameters	True Values	Estimated Values	Estimated Variances
ω_1	0.3	0.2999633	0.2×10^{-6}
B_{11}	0.5403	0.520	0.003
B_{12}	-0.8415	-0.772	0.004
ω_2	0.31	0.3101024	0.2×10^{-6}
B_{21}	-0.4161	-0.401	0.003
B_{22}	-0.9093	-0.948	0.007

For the two-frequency model we have also investigated the dependence of the results on the sequence length K. Figure 5.8 shows how ω_1 fluctuates with increasing K. It is obvious that as K increases, there is a definite decrease in the scatter, with a tendency to a value very close to the exact value. The variance associated with this ω_1 value is also shown as a function of K in Figure 5.9. Near $K = 20$ there is a clear settling of the variance to a value near 7×10^{-7}. Plots for the coefficients B_{11}, B_{12}, etc show a similar tendency. We are thus satisfied that for these examples, a value of 20 for the sequence length is suitable.

Finally we consider a multiple harmonic-frequency model:

$$d_i = cos(0.1t_i + 1) + 2cos(0.15t_i + 2) + 5cos(0.3t_i + 3) \qquad (5.63)$$
$$+2cos(0.31t_i + 4) + 3cos(1t_i + 5) + \varepsilon_i$$

The time series is shown in Figure 5.10. The best estimates and the variances for all the parameters are tabulated in Table 5.3[1] with the true values and Bretthorst's results for the same example.

[1]In order to compare our results with Bretthorst's [4] in this example we converted $\{B_{j1}, B_{j2}, \omega_j\}$ to $\{B_j, \phi_j, \omega_j\}$ where $B_j = \sqrt{B_{j1}^2 + B_{j2}^2}$, $\phi_j = arctan(-B_{j2}/B_{j1})$, $j = 1, \cdots, 5$.

Table 5.3: The Best Estimates of Parameters For a Multiple Frequency Model

Parameters	True Values	Estimated Values	Bretthorst's Results
ω_1	0.1	0.09999359 $\pm$0.0004	0.0998 $\pm$0.0001
B_1	1	0.966 ± 0.06	0.9 ± 0.08
ϕ_1	1	0.987 ± 0.06	N/A
ω_2	0.15	0.15003654 $\pm$0.0002	0.1498 $\pm$0.0002
B_2	2	2.116 ± 0.06	2.08 ± 0.08
ϕ_2	2	2.019 ± 0.03	N/A
ω_3	0.3	0.3000216 $\pm$0.0001	0.3001 $\pm$0.0002
B_3	5	4.988 ± 0.07	4.97 ± 0.08
ϕ_3	3	3.0080 ± 0.024	N/A
ω_4	0.31	0.3100121 $\pm$0.0002	0.3102 $\pm$0.0001
B_4	2	2.018 ± 0.084	1.95 ± 0.08
ϕ_4	4	3.985 ± 0.03	N/A
ω_5	1	1.00003730 $\pm$0.0002	0.9999 $\pm$0.0001
B_5	3	3.069 ± 0.06	2.92 ± 0.08
ϕ_5	5	4.9790 ± 0.02	N/A

By comparing the estimated values in colum 3 and colum 4 with the true values in colum 2,it is obvious that our results are closer to the true values than those of Bretthorst's considering the estimated standard deviation from the both method are in the same order. Once again all the frequencies have been well resolved, even the third and the fourth frequencies which are too close to be separated by FFT. That is why in Figure 5.11, which is the FFT power spectral density, it only shows four peaks. The Bayesian spectral density plot for this example shows five sharp peaks. Once again the advantages of Bayesian inference and Gibbs sampling are

Table 5.4: The Best Estimates of Parameters For A Single Frequency Model with Unknown Noise Variance

Parameters	True Values	Estimated Values	Estimated Variances
ω	0.3	0.3000352	0.2×10^{-6}
B_1	0.5403	0.525	0.004
B_2	-0.8415	-0.779	0.004
σ^2	1	1.002	0.004

demonstrated by these results.

Let us now return to the first example. This time we suppose that the variance of the noise is unknown. We have repeated the same calculation except that the random numbers are generated from the Student-t distribution. All of the estimated values are tabulated in Table 5.4 together with the true values. As we can see, even when the noise level is unknown Bayesian inference and Gibbs sampling can still give very accurate results together with an estimate of the noise level.

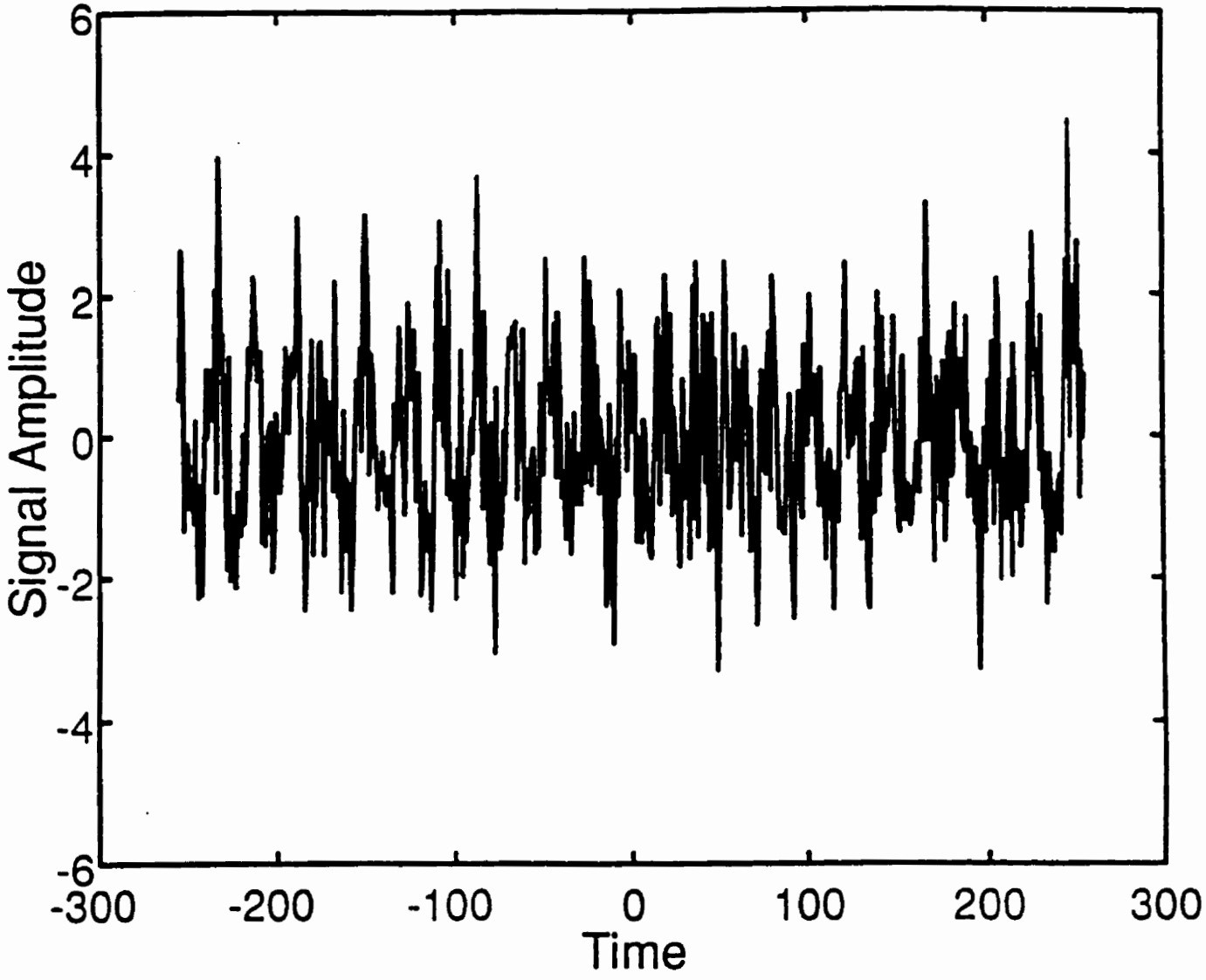

Figure 5.1: Time series of example one.

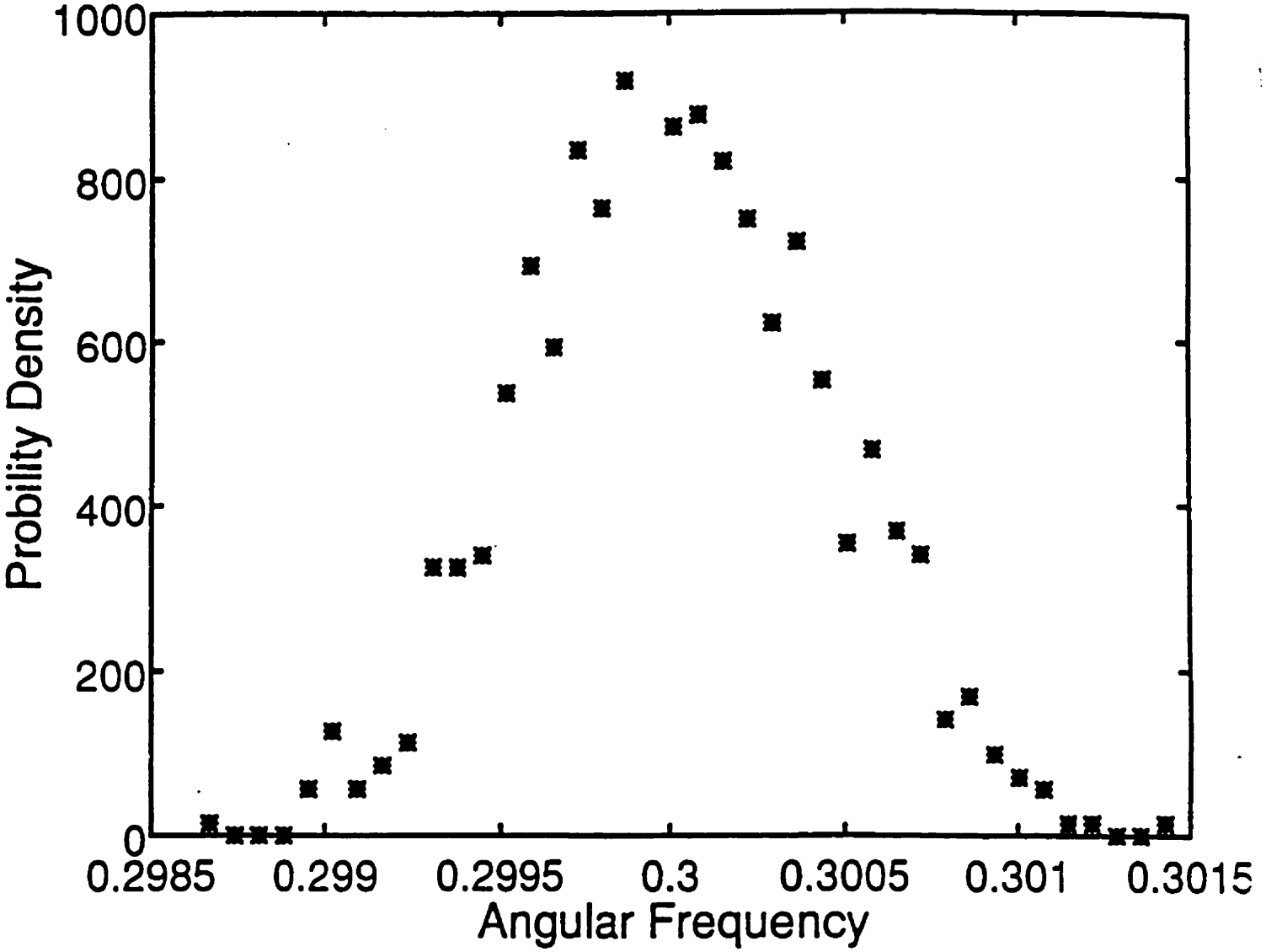

Figure 5.2: Posterior probability distribution for example one.

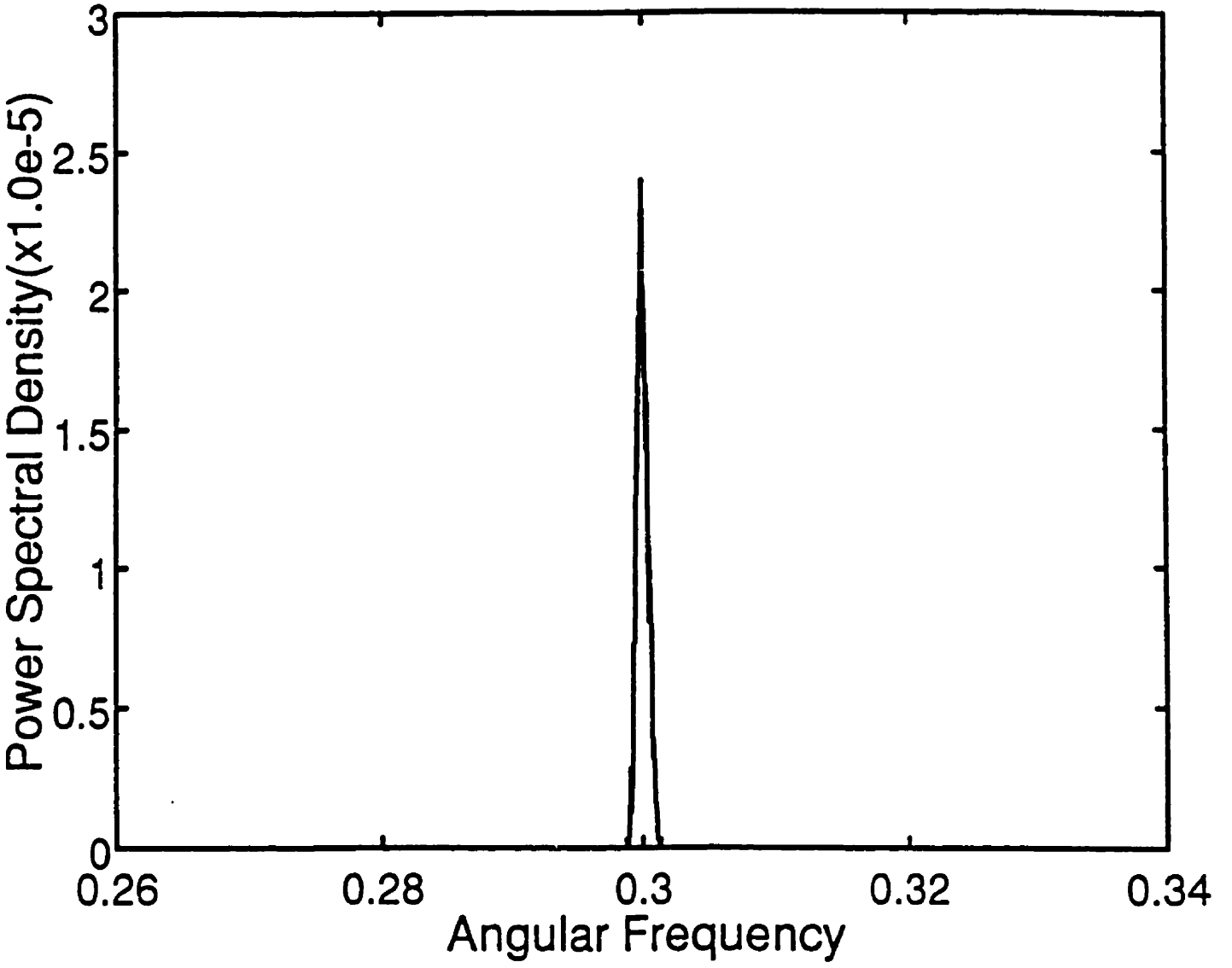

Figure 5.3: FFT Power spectral density for example one.

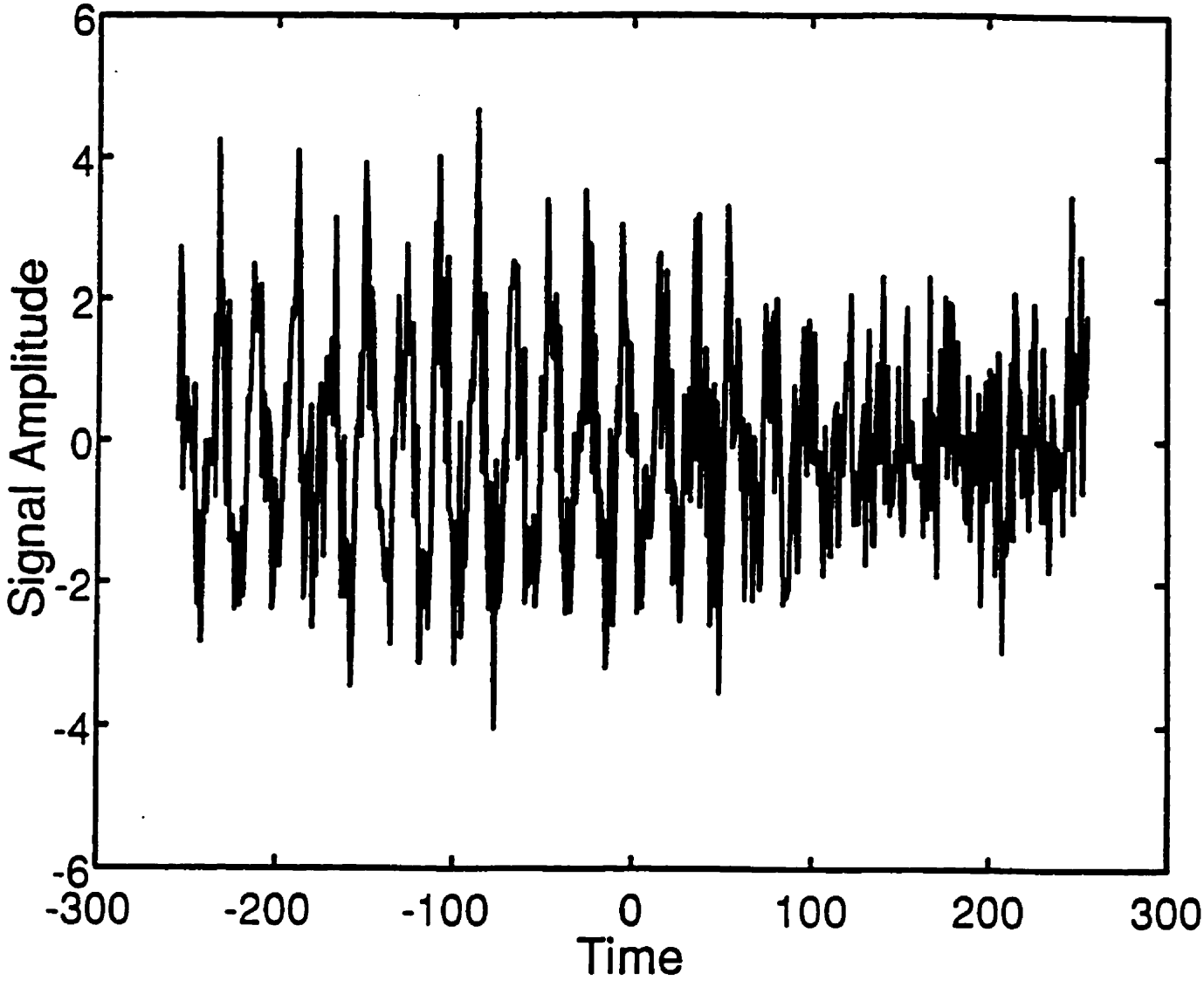

Figure 5.4: Time series of example two.

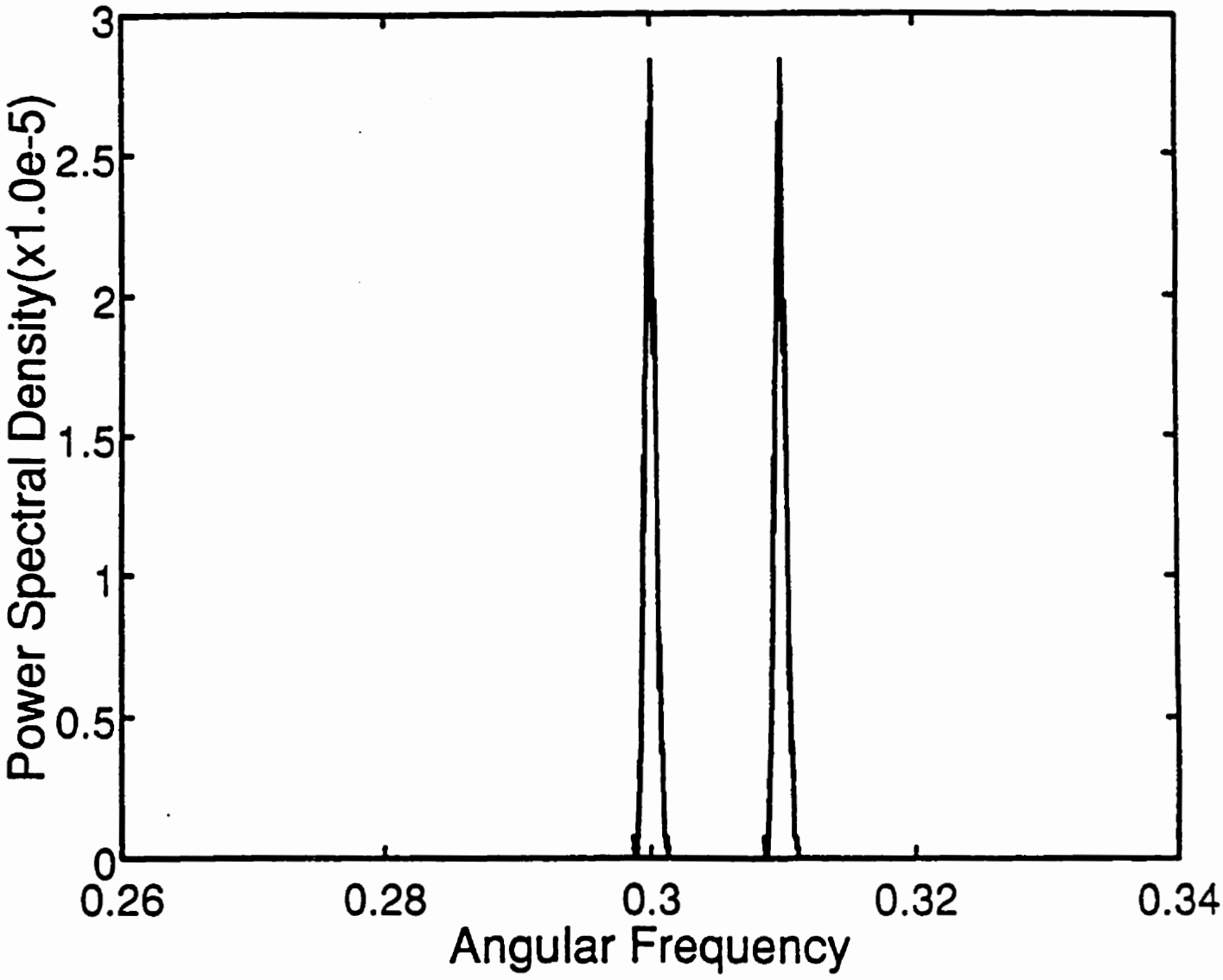

Figure 5.5: Posterior probability distribution for example two.

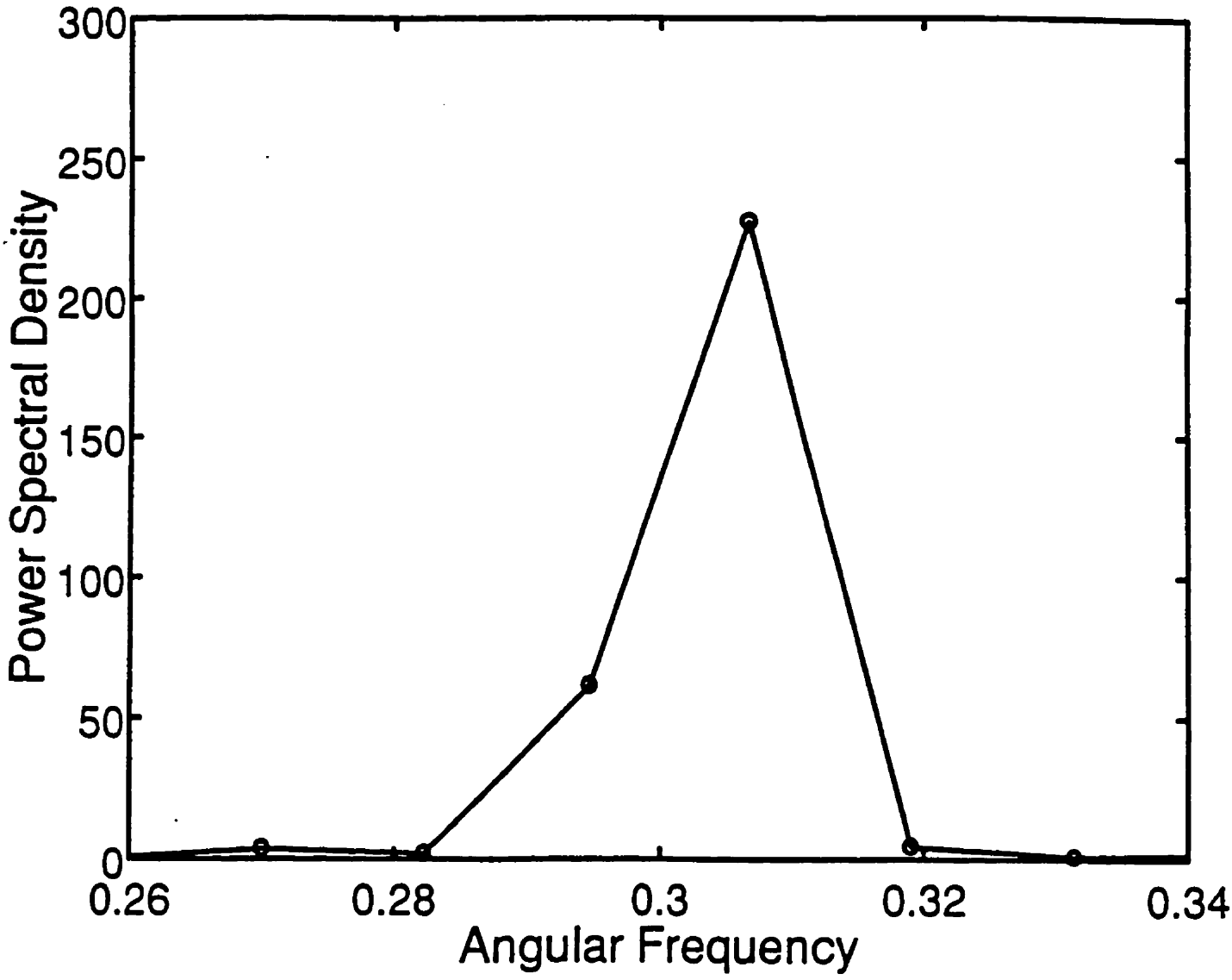

Figure 5.6: FFT Power spectral density for example two.

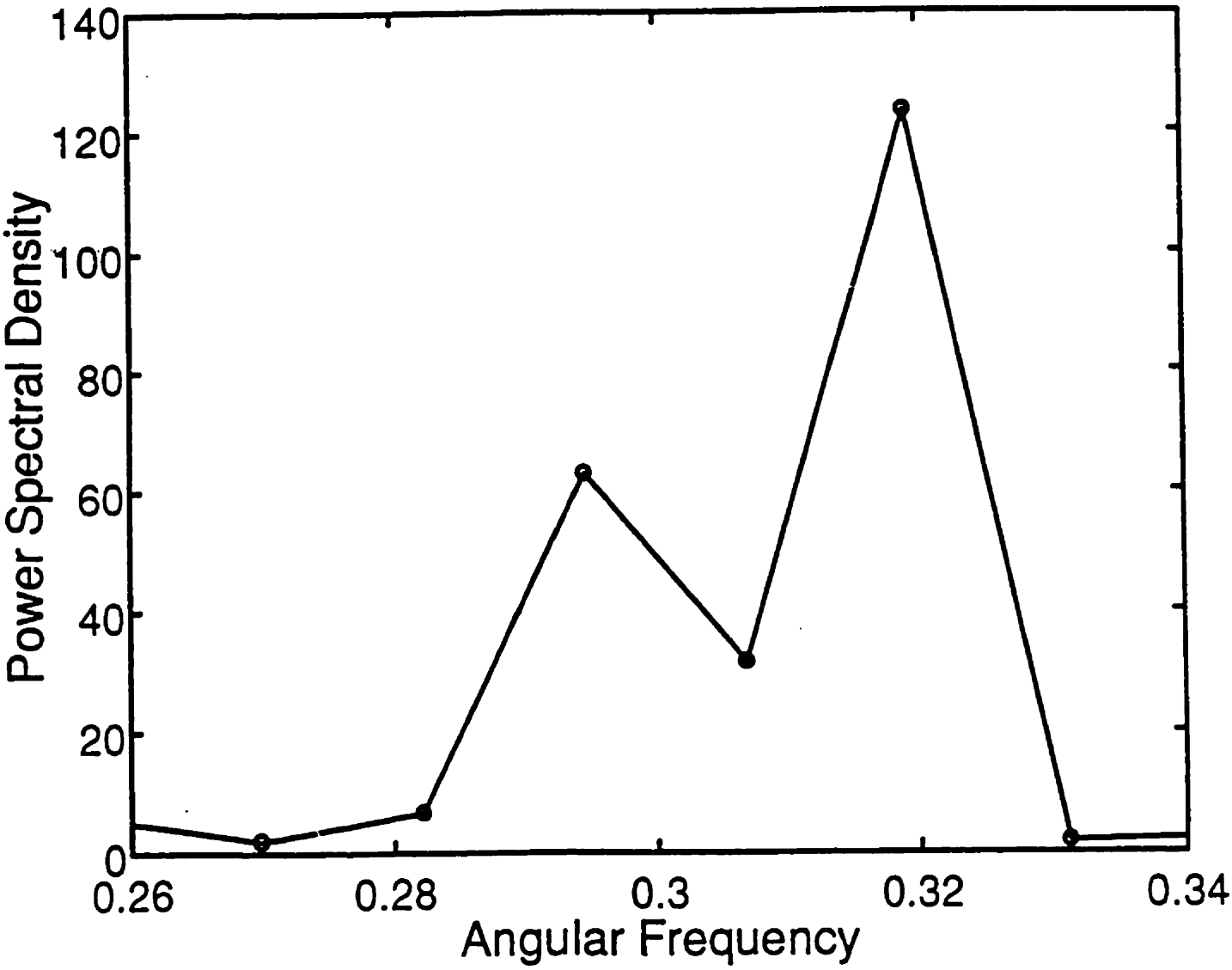

Figure 5.7: FFT Power spectral density for a two-frequency signal separated by $1.5 \times (2\pi/N)$.

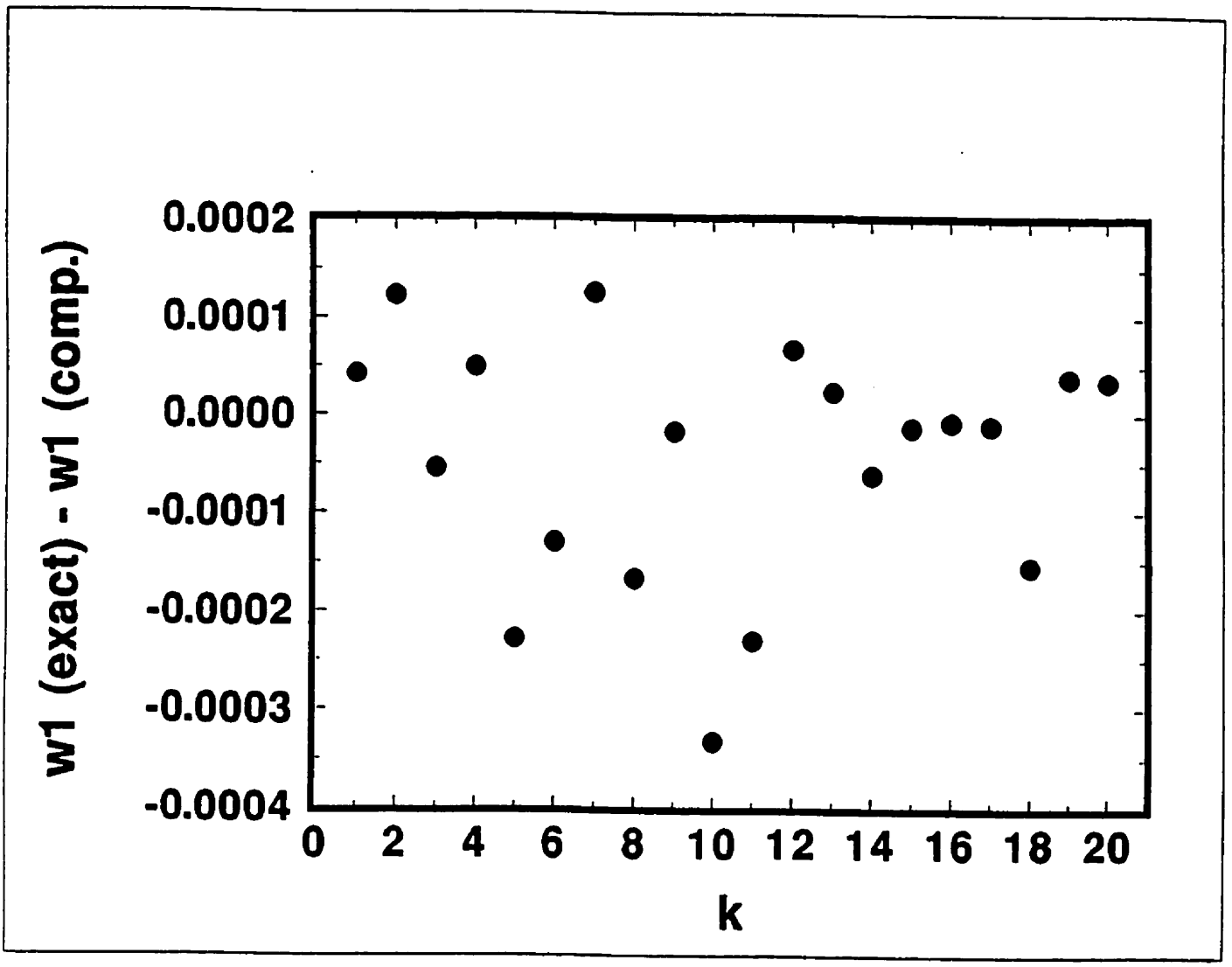

Fig. 3(a)

Figure 5.8: Dependency of ω_1 on K value.

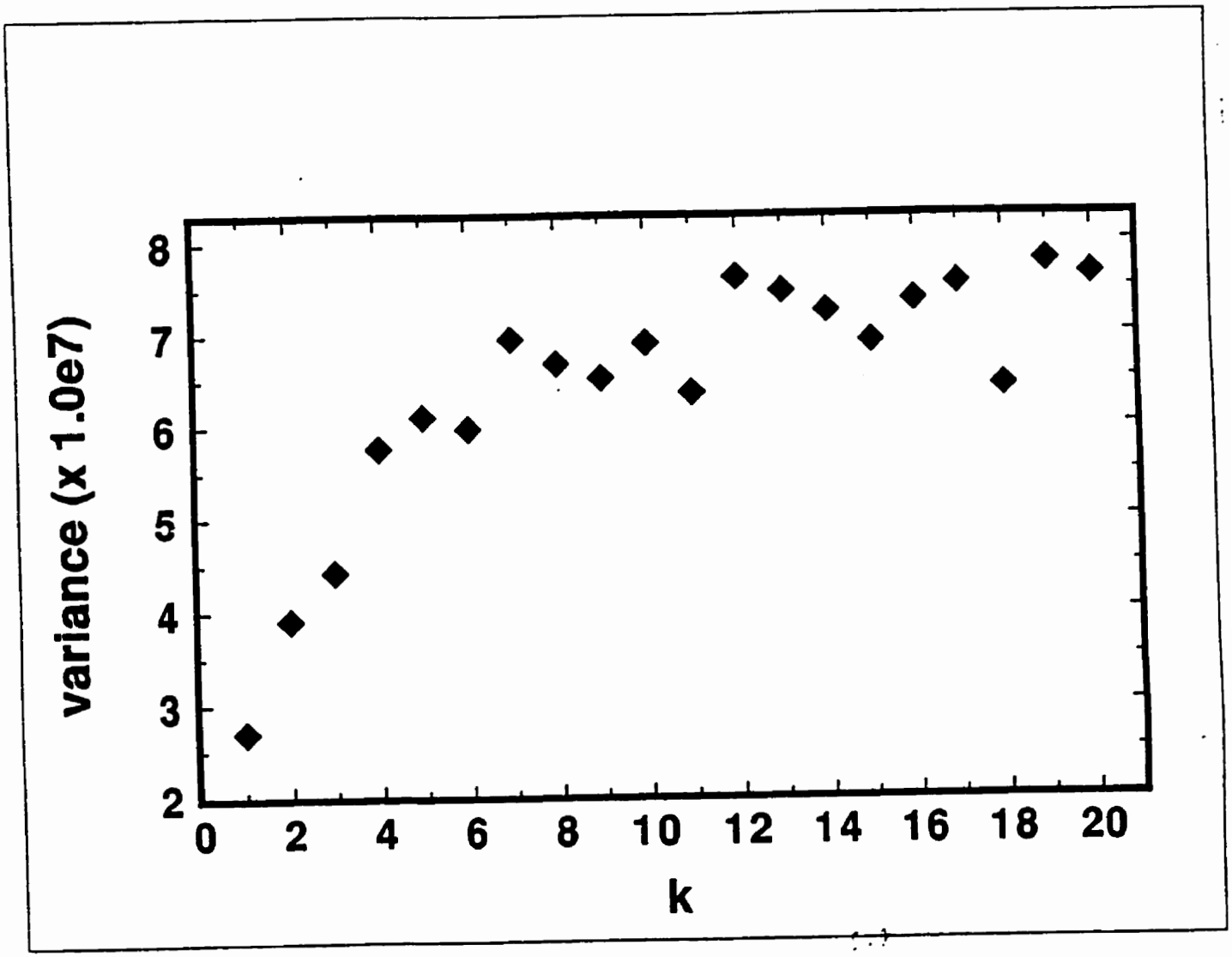

Figure 5.9: Dependency of variance of ω_1 on K value.

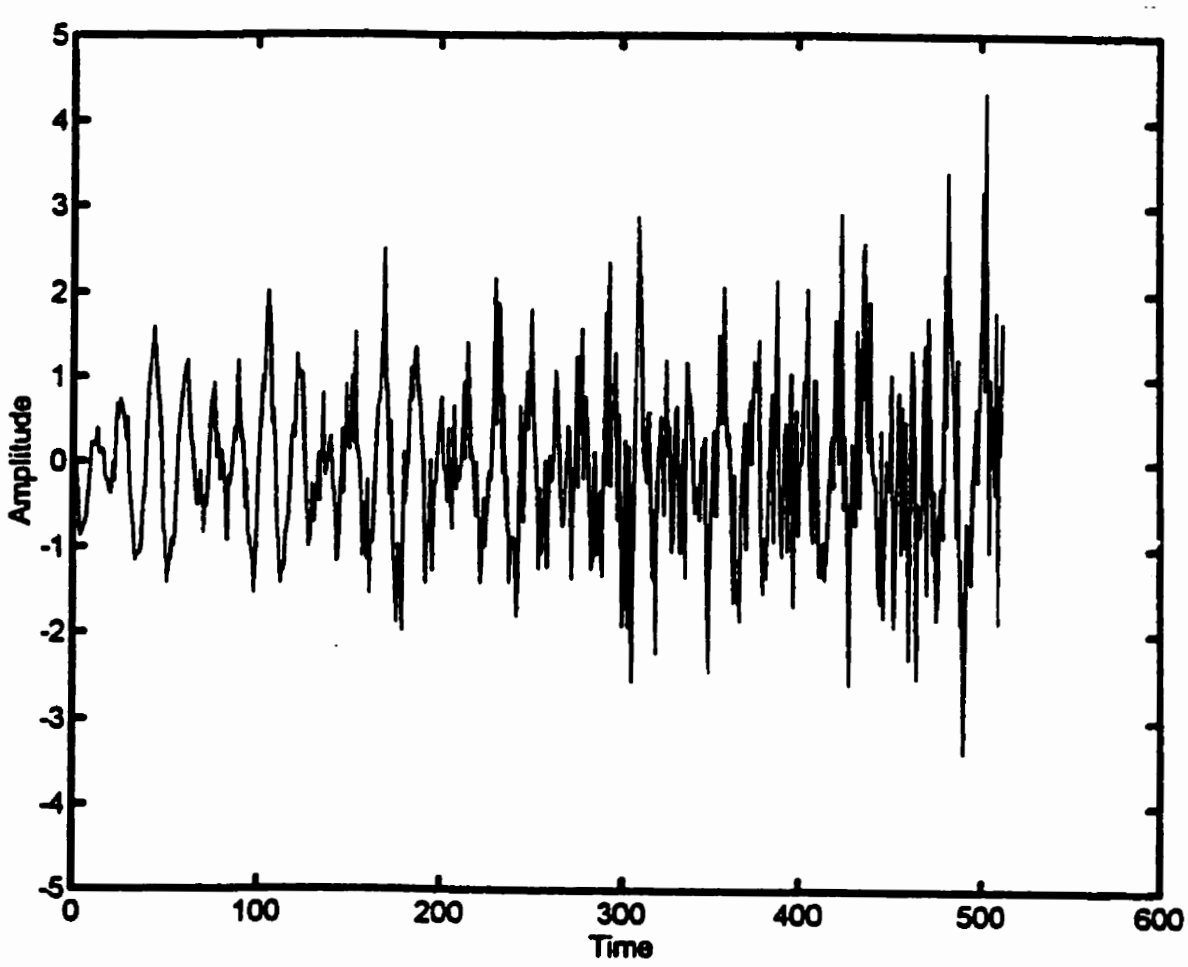

Figure 5.10: Time series of example three.

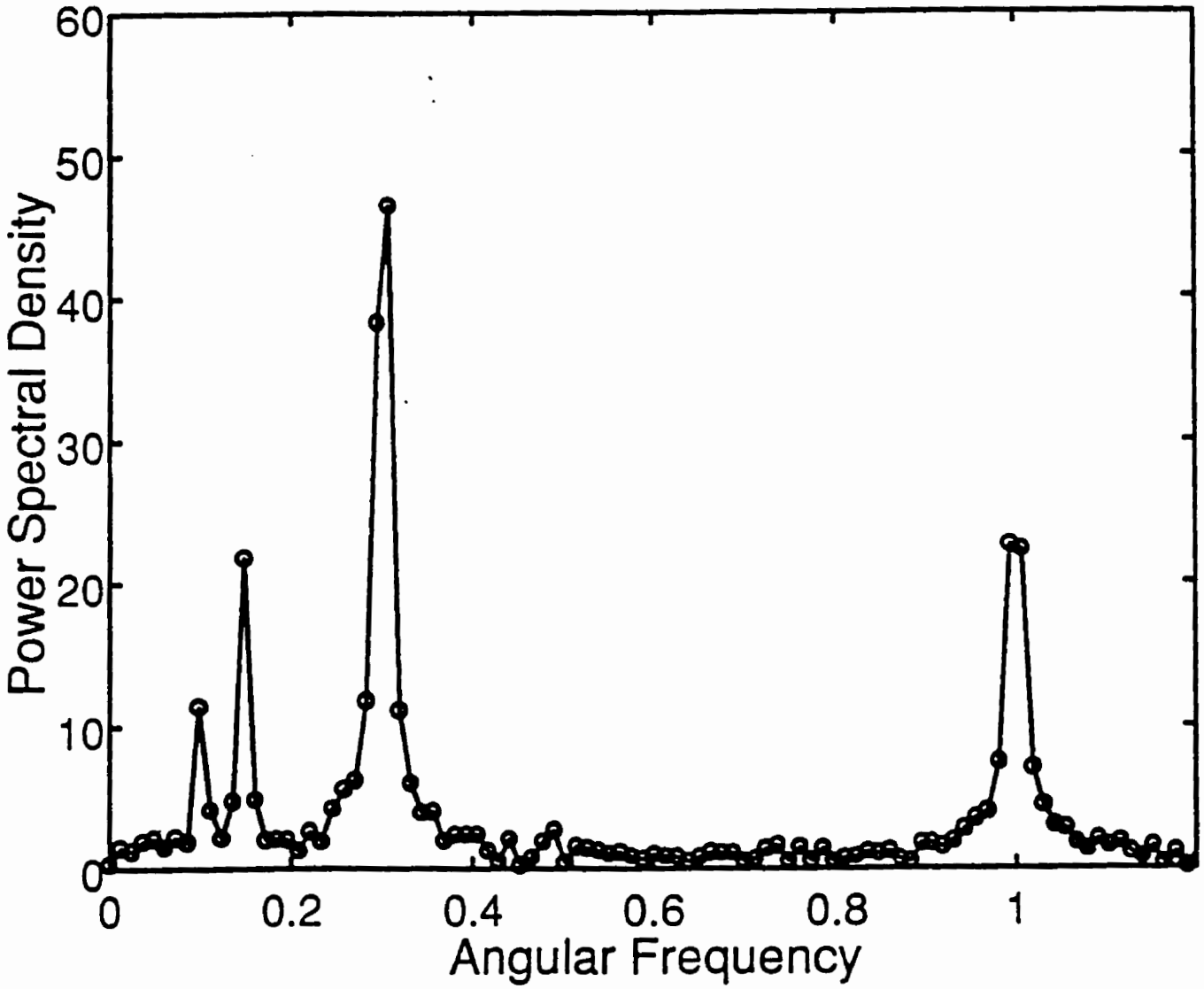

Figure 5.11: FFT Power spectral density for example three.

5.3 Violating the Conditions

In the previous chapter we applied the Gibbs sampling approach to signals which were harmonic, containing one or more frequencies. It was demonstrated that it was possible to determine the frequencies with a fair degree of accuracy, in the presence of white noise. We now proceed to examine a few situations where these simple assumptions are no longer true. Our objective is to see if we can still extract the key properties of the signal, and to also compare the use of approximate integral Bayesian methods with the Gibbs sampling approach.

5.3.1 Periodic but non harmonic signal

Initially we will look at extending the results of our earlier work by applying the harmonic model to a time series which is periodic but non harmonic. Three examples will be studied: a ramp function, a triangle function, and a Gaussian-based non-harmonic signal. Assuming we have no knowledge of the form of the underlying signal, we will assume a harmonic signal model function. In practise, when noise is added, it is often very difficult to determine the form of the underlying signal.

The ramp signal is built by repeatedly running a counter from 0 to 15, and then subtracting 7.5 from each point so that the mean value is zero. Gaussian white noise is then added, with a variance of one. For the triangle wave, a counter is run from 0 to 16, and then from 15 to 1. This cycle is repeated, and also adjusted to have a zero mean. Our third non-harmonic signal is constructed by first counting from $i = 1$ to $i = 9$ with the function $\exp[(i-5)^2/2]$, and then by counting from $i = 9$ to $i = 16$ with the function $-\exp[(i-13)^2/2]$. A section of this latter waveform is shown in Figure 5.12 without noise, and with noise added in Figure 5.13. In all

three cases we have employed $N = 1024$ points.

We have focused our attention on determining the frequency of the underlying signal. For the ramp and Gaussian signals, the exact frequency is $2\pi/16$, whereas for the triangle function it is $2\pi/32$. If we apply the FFT to each of these, we observe a number of alias frequencies. In the case of the triangle function, the amplitudes of these aliases are quite small in comparison with that at the fundamental frequency. The ramp function displays a series of alias frequencies, of decreasing amplitudes, at multiples of the fundamental, and the Gaussian function shows a large number of aliases, many with significant amplitudes.

We then proceed in the same manner as described in section 5.2. Our model is

$$\hat{d}_i = B_1 \cos(\omega i) + B_2 \sin(\omega i) + \varepsilon_i \qquad i = 1, 2, \ldots, 1024 \qquad (5.64)$$

and we want to find the best estimates for ω given the data set $D = \{d_1, d_2, \ldots, d_N\}$.

Applying the approximate Bayesian formalism, using equation (5.25), we are able to compute the posterior probability function for each of these three examples. The form of the plot of the posterior as a function of frequency is very similar to that of the FFT except that there is an enormous difference in the amplitudes of the peaks. For example, for the ramp function, in the FFT plot the ratio of amplitudes of the largest peak to the second largest is approximately 2:1, while in the plot of the posterior function of equation (5.25) the same ratio is roughly $10^{378} : 1$, an enormous difference. In the latter case this is a signal of the level of probability for each of the frequencies. In the case of the gaussian function, the ratio of the amplitude of the largest peak to the next largest drops to only $10^{11} : 1$.

Table 5.5 summarizes the results of the computations for all three signals. The estimated value of the frequency in each case is obtained from the location of the

Table 5.5: Estimated frequencies for non-harmonic signals

Signal:	Ramp	Triangle	Gaussian
Bretthorst	0.39265 $\pm 3 \times 10^{-5}$	0.19637 $\pm 2.3 \times 10^{-5}$	0.3922 $\pm 7 \times 10^{-4}$
Gibbs	0.39258 $\pm 4 \times 10^{-5}$	0.19626 $\pm 2 \times 10^{-5}$	0.3920 $\pm 5 \times 10^{-4}$
Exact	0.392699082	0.196349541	0.392699082

largest peak in the plot of the posterior probability. The associated variance is computed using two different procedures. For the approximate integral approach of Bretthorst, we have employed the method outlined above, wherein the probability function is expanded in a Taylor series about the maximum, and the coefficient of the quadratic term is used to compute the variance. In the Gibbs sampling case, we consider the values of the probability in the vicinity of the maximum (everything beyond is zero in comparison), normalize the total area to one, and compute the marginal variance from the data points. The uncertainty quoted in the table is the standard deviation. The frequency estimates given here are of the same order of magnitude as one can obtain if the true signal actually met the assumptions of the model.

5.3.2 A Signal with non-stationary and non-white noise

The theory developed earlier for spectral estimation is partially based on the assumption that the noise involved in the signal is both stationary and white. As a consequence the noise is distributed as a Normal distribution with zero mean and variance σ^2 (we suppose σ^2 is known). If instead the noise values are increasing with time and are also correlated, then the noise data are no longer Normally distributed

random numbers.

In this section we investigate the application of the theory which is based on stationary, white noise to signals which have non-stationary and correlated noise added to them.

The basic time series is generated from the series,

$$\hat{d}_i = 0.001 + B_1 \cos(0.3i) + B_2 \sin(0.3i) + \varepsilon_i \qquad i = 1, 2, \ldots, 1024 \quad (5.65)$$

Different types of noise signals are then added to this time series. A non-stationary noise signal is obtained by multiplying a standard white noise signal by $0.1 + 0.002t$. In addition we can introduce correlation by applying a four-point averaging filter. In our analysis, we have investigated these cases separately, as well as together.

The non-stationary noise signal quickly dominates the combined signals, as can be seen in Figure 5.14. Even so, application of the FFT to this signal gives a maximum very close to $\omega = 0.3$. Many other small peaks also occur, but the ratio of the amplitudes of the peak at 0.3 to the others is at least 5:1. If only correlations are introduced, then this ratio improves considerably, as the amplitude of the smaller peaks are much less.

In Table 5.6 to 5.8 we summarize and compare the calculations for the frequency estimation using the approximation method of Bretthorst, and the Gibbs sampling approach for three different situations: non-stationary noise; four-point averaging filter; and both together. As can be seen, the results are very similar. There is no difficulty in obtaining an accurate estimate of the frequency. It is clear from this that we are still able to obtain very good estimates of the signal frequency in all three cases. The amplitudes are much more difficult to deduce in this kind of a situation. This should not be surprising given the amount of noise in the signal.

Table 5.6: The Best Estimates of Parameters For A Signal with Non-Stationary Noise

Parameter	Integration	Gibbs	True values
ω	0.2998 $\pm 2 \times 10^{-4}$	0.3000 $\pm 1 \times 10^{-4}$	0.3
B_1	0.60 ± 0.04	0.65 ± 0.1	0.5403
B_2	-0.83 ± 0.04	-0.90 ± 0.08	-0.8415

Table 5.7: The Best Estimates of Parameters For A Signal with Correlated Noise

Parameter	Integration	Gibbs	True values
ω	0.2998 $\pm 7 \times 10^{-4}$	0.3002 $\pm 2 \times 10^{-4}$	0.3
B_1	0.51 ± 0.04	0.57 ± 0.1	0.5403
B_2	-0.86 ± 0.04	-0.80 ± 0.08	-0.8415

Table 5.8: The Best Estimates of Parameters For A Signal with Non-Stationary and Correlated Noise

Parameter	Integration	Gibbs	True values
ω	0.2998 $\pm 1 \times 10^{-4}$	0.3001 $\pm 2 \times 10^{-4}$	0.3
B_1	0.44 ± 0.04	0.56 ± 0.1	0.5403
B_2	-0.79 ± 0.04	-0.80 ± 0.08	-0.8415

5.3.3 A signal with Lorentzian decay

In experiments one often encounters signals which are not stationary, i.e. they decay as a function of time. A particular example is the free-induction decay time series associated with NMR (Nuclear Magnetic Resonance) spectroscopy. In our next example we apply the Gibbs sampling approach to actual NMR experimental data. The data used are of a free-induction decay shown in Figure 5.15. NMR time series consist of two data channels: the real channel and the imaginary channel, 90° out of phase with each other. It is well known [40] that the decay in NMR time series is Lorentzian decay, so that the model we use here is

$$d_i = \sum_{j=1}^{M} [B_{1j}\cos(\omega_j i) + B_{2j}\sin(\omega_j i)]e^{-\alpha_j i} + \varepsilon_i \qquad i = 1, \ldots, 512 \quad (5.66)$$

From a preliminary study of the data using FFT, we observe that there are two frequency lines in both data channels. Therefore we have analyzed the data using a two-frequency model, i.e. $M = 2$ in equation (5.66). With this assumption we have four parameters to estimate for each frequency line, B_{1j}, B_{2j}, ω_j, and α_j, $j = 1, 2$. Four of these, ω_j and α_j, $j = 1, 2$, are non-linear parameters.

In applying the integral approach, we follow essentially the same procedure as for the previous examples, employing equation (5.26) to compute the posterior probability. The difference now is that this probability is a function of the four variables, ω_j and α_j. Hence we locate the values of these variables which make this function a maximum. This is accomplished using standard quasi-Newton search procedures. Once these have been found, we can then estimate the values of B_{1j} and B_{2j} as before. The results for the real channel, including the standard deviations as computed using equations (5.27) and (5.29) are displayed in Table 5.9.

In order to apply the Gibbs method, the linear approximation described in Section 3.2.3 of Chapter 3 has been used to linearize the sinusoidal functions with respect to frequency, and the Lorentzian decay functions with respect to the decay rate. Following similar numerical procedures to those used in the two-frequency model studied in 5.2.6, the marginal posterior probability distributions for all the parameters have been computed. From these distributions the best estimates for the parameters can be made with corresponding variances. The results for the real channel are given in Table 5.9. These parameters have been used to generate the real NMR signal, shown in Figure 5.16. It can be seen in comparing Figure 5.15 with Figure 5.16 that the model signals and the experimental signals look essentially identical except that there is a certain amount of random noise in the latter. In Figure 5.17 the residual of the experimental signal and the model signal for the real channel is shown, which looks like white noise. In addition we calculated the FFT spectral density for the two residual signals (real and imaginary channels). The results for the real channel are shown in Figure 5.18. There appear to be some very small effects in the two residuals. However these effects are not the same in both channels. We think that these small effects may be due to an instrumental artifact.

Table 5.9: Best Parameter Estimates for Lorentzian Decay Signal

Parameter	Integration	Gibbs
α_1	0.0098 $\pm 7 \times 10^{-4}$	0.0109999 $\pm 2 \times 10^{-7}$
ω_1	1.9251 $\pm 5 \times 10^{-4}$	1.9300000 $\pm 1 \times 10^{-7}$
B_{11}	-17626 ± 410	-12480.4 ± 0.4
B_{21}	-13832 ± 410	18567.2 ± 0.5
α_2	0.0103 $\pm 8 \times 10^{-4}$	0.0100000 $\pm 1 \times 10^{-7}$
ω_2	1.8715 $\pm 6 \times 10^{-4}$	1.8700000 $\pm 1 \times 10^{-7}$
B_{12}	-16642 ± 520	-4336.9 ± 0.4
B_{22}	-13330 ± 500	18750.0 ± 0.4

5.3.4 A signal with chirp

Chirp signals are common in several areas of science and engineering (e.g., physics, sonar, radar, communications). In one application they are used to estimate the trajectories of moving objects with respect to fixed receivers. In long-haul optical communications systems, one of the primary parameters that determines performance is the frequency chirp. A number of schemes have been developed in order to estimate the chirp parameters from a signal [38].

Sometimes the phase shift in signals is instantaneous, e.g. the phase could be varying slowly with time. A signal with chirp is one example of these kinds of signals. In what follows we will demonstrate that the Bayesian inference and Gibbs sampling procedure can also deal with these kinds of signals. Here we will apply

Table 5.10: Best Parameter Estimates for A Signal with Chirp

Parameter	True Value	Integration	Gibbs
ω	0.3	0.2988 $\pm 6 \times 10^{-4}$	0.29979 $\pm 4 \times 10^{-4}$
α	0.01	0.0100044 $\pm 8 \times 10^{-7}$	0.01000063 $\pm 2 \times 10^{-6}$
B	1.0	0.888 ± 0.004	1.066 ± 0.04
ϕ	1.0	0.90 ± 0.1	1.0 ± 0.2

the Bayesian analysis to a real chirp signal. Figures 5.19 and 5.20 show plots of the signal generated by using the following formula

$$d_i = B\cos(\omega i + \alpha i^2 + \phi) + \varepsilon_i \tag{5.67}$$

where

$$B = 1,\ \omega = 0.3,\ \alpha = 0.01,\ \phi = 1$$

and ε_i represents a Normally distributed random noise with zero mean. Here again we have four parameters (B, ω, α, ϕ) to estimate. We have used the same linear approximation as was introduced in Chapter 3 to linearize the function with respect to ω and α respectively. In this situation, however, the model functions used when applying the method of Bretthorst are $\cos(\omega t + \alpha t^2)$ and $\sin(\omega t + \alpha t^2)$. In this approach we search for the values of ω and α which give the maximum of the posterior. In the Gibbs sampling approach the marginal posterior probabilities for each of the four parameters are obtained. The best estimates for the four parameters (B, ω, α, ϕ), are tabulated in Table 5.10 together with the true values.

Just for comparison, in Figure 5.21 we show the discrete Fourier transform of the chirp signal. Obviously nothing can be estimated from it. This is due to the

fact that there is more structure in this signal than in a simply periodic signal. Intrinsically FFT is not able to give the correct answer to this problem.

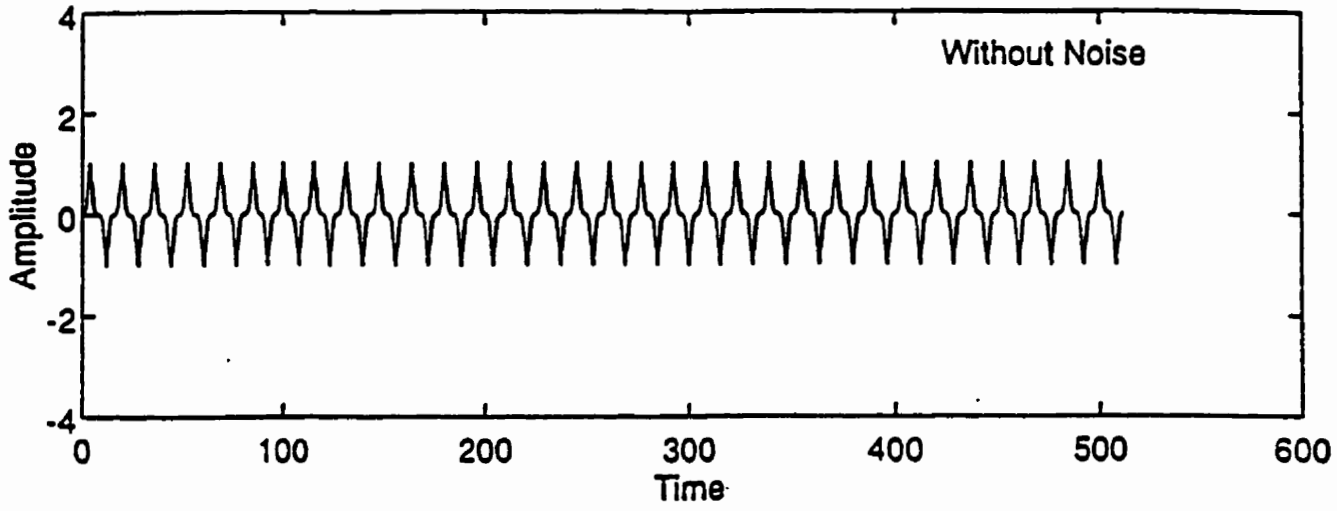

Figure 5.12: Time series of non-harmonic signal without noise.

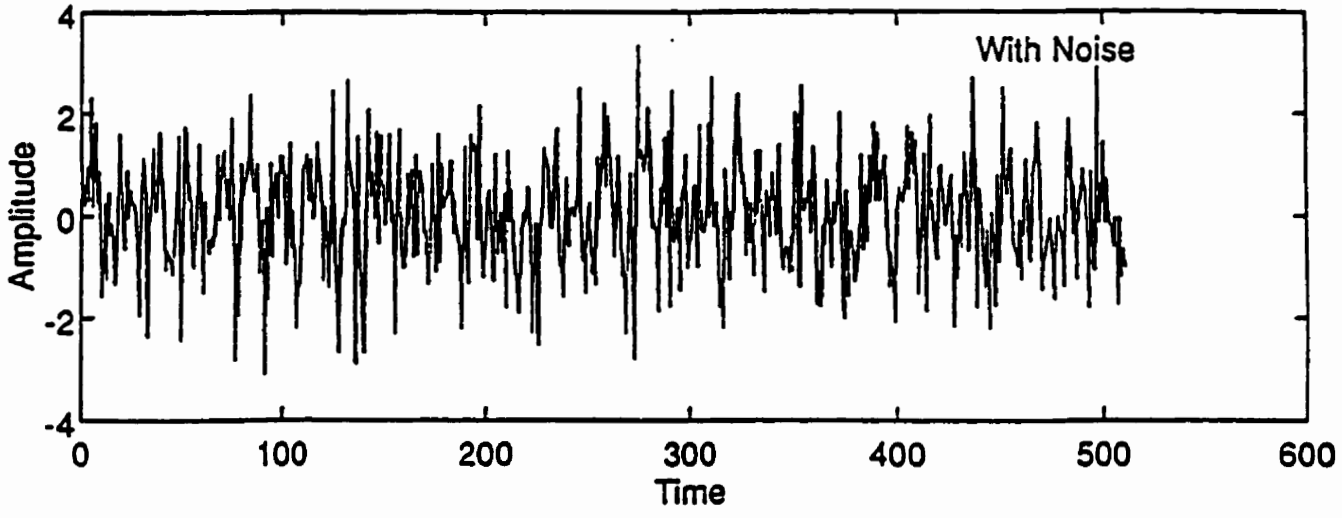

Figure 5.13: Time series of non-harmonic signal with noise.

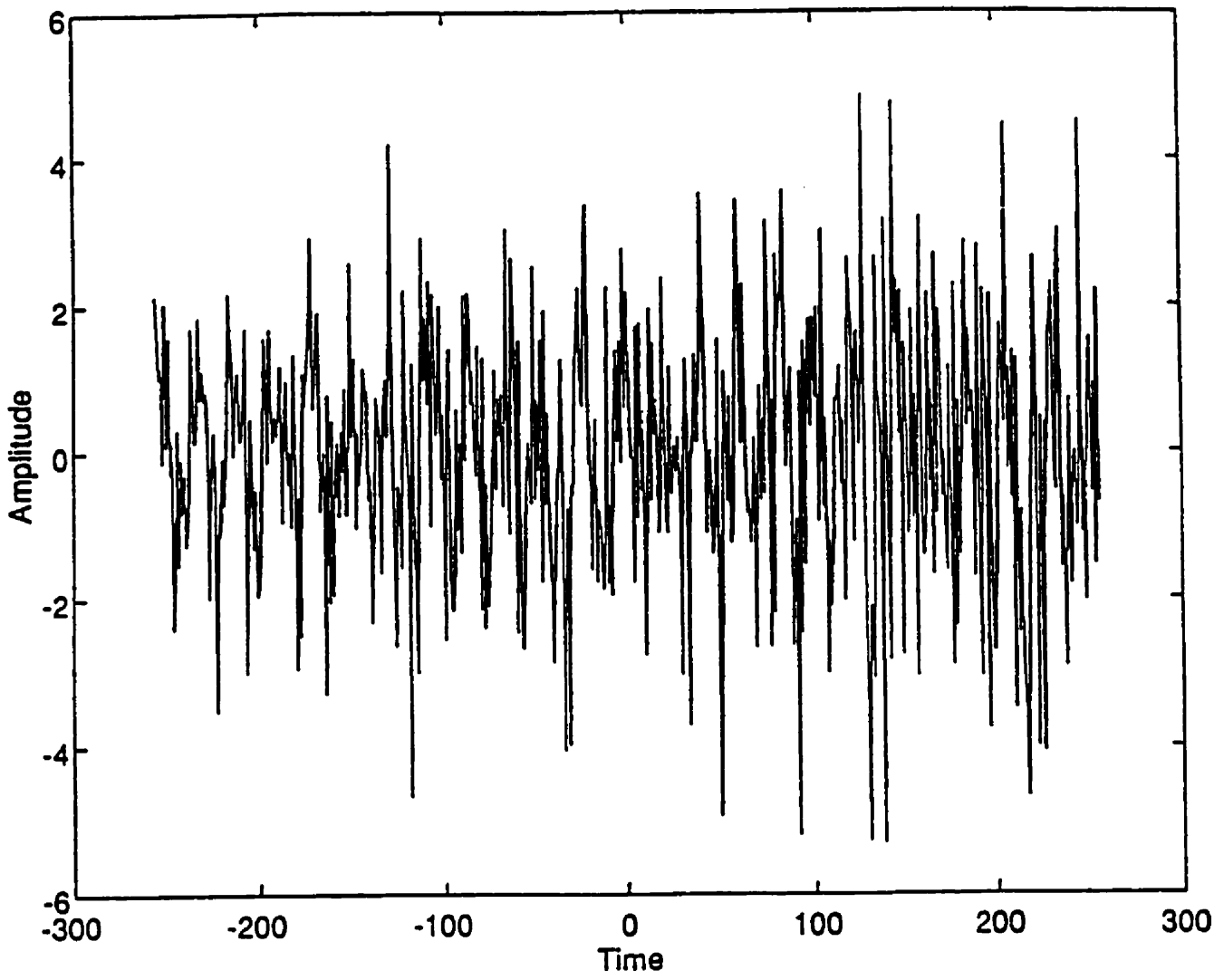

Figure 5.14: Time series of a signal with non-white and non-stationary noise.

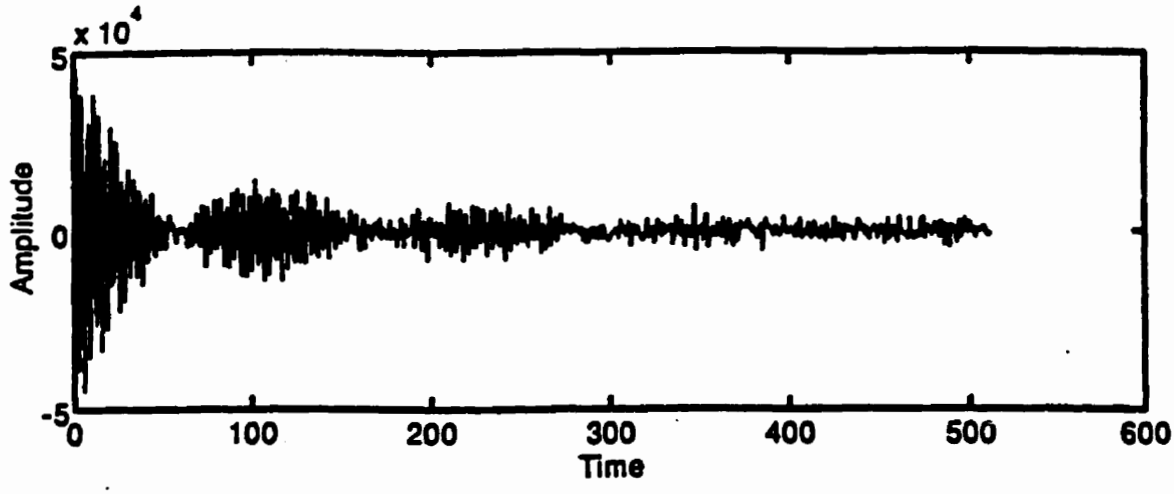

Figure 5.15: Real channel NMR data from a real experiment.

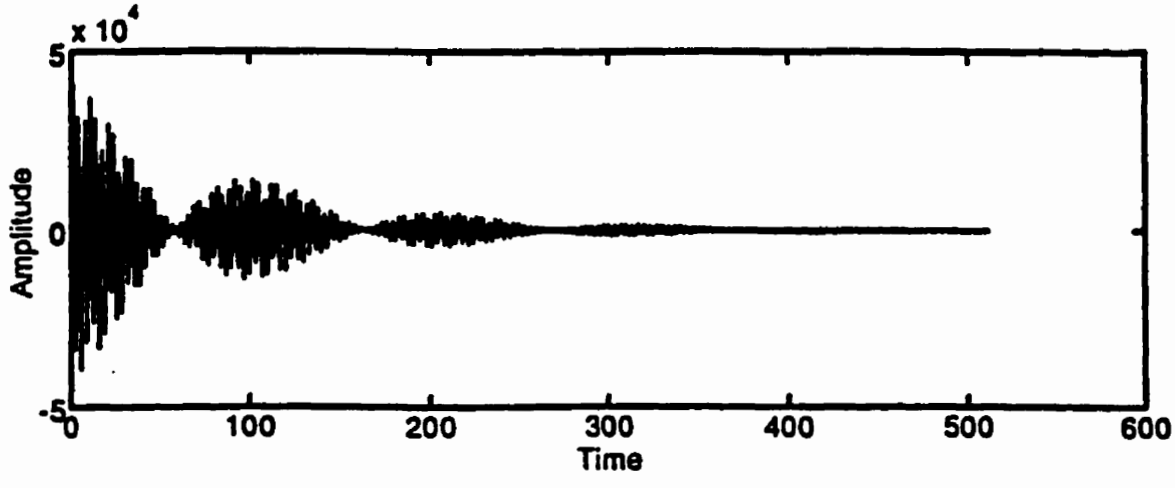

Figure 5.16: Real channel NMR data from a model with extracted parameters.

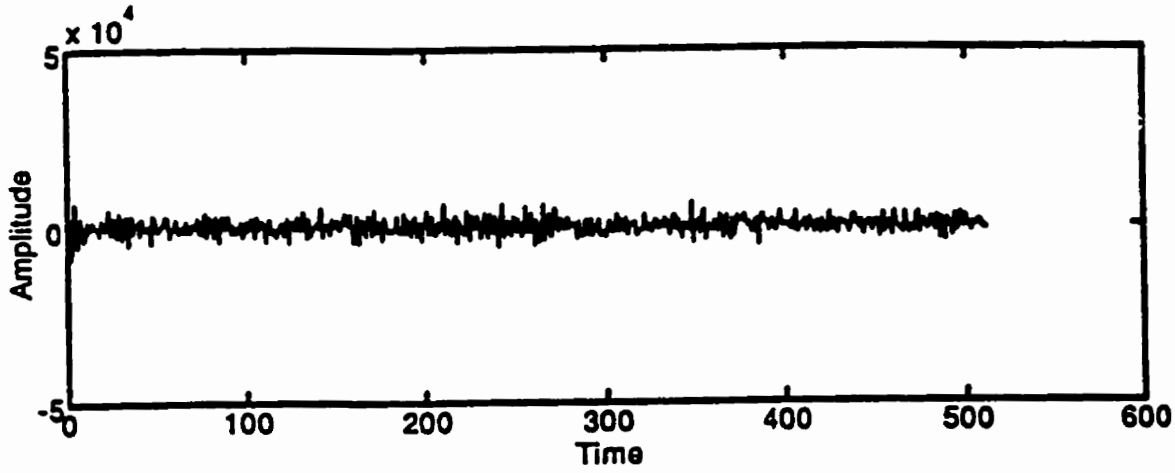

Figure 5.17: Residual.

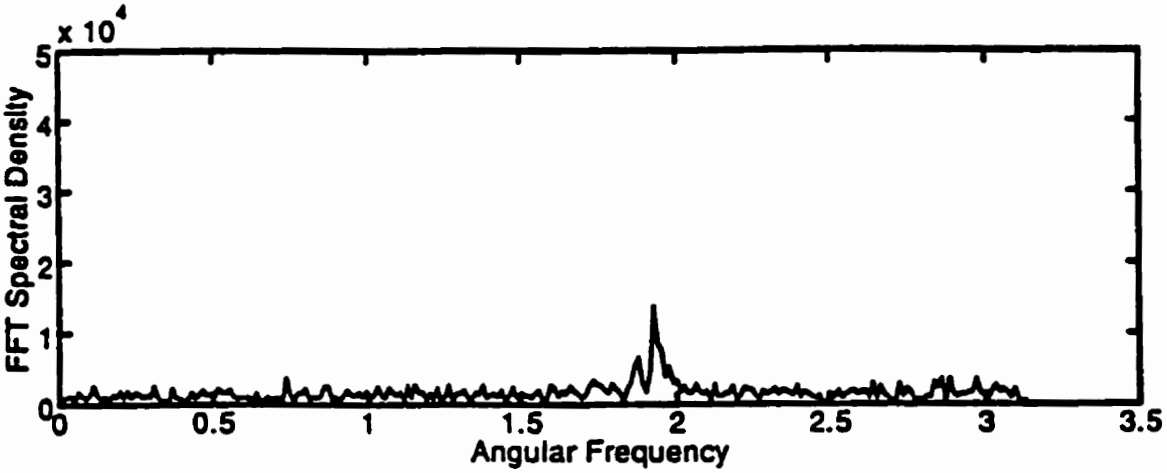

Figure 5.18: FFT spectrum of the residual.

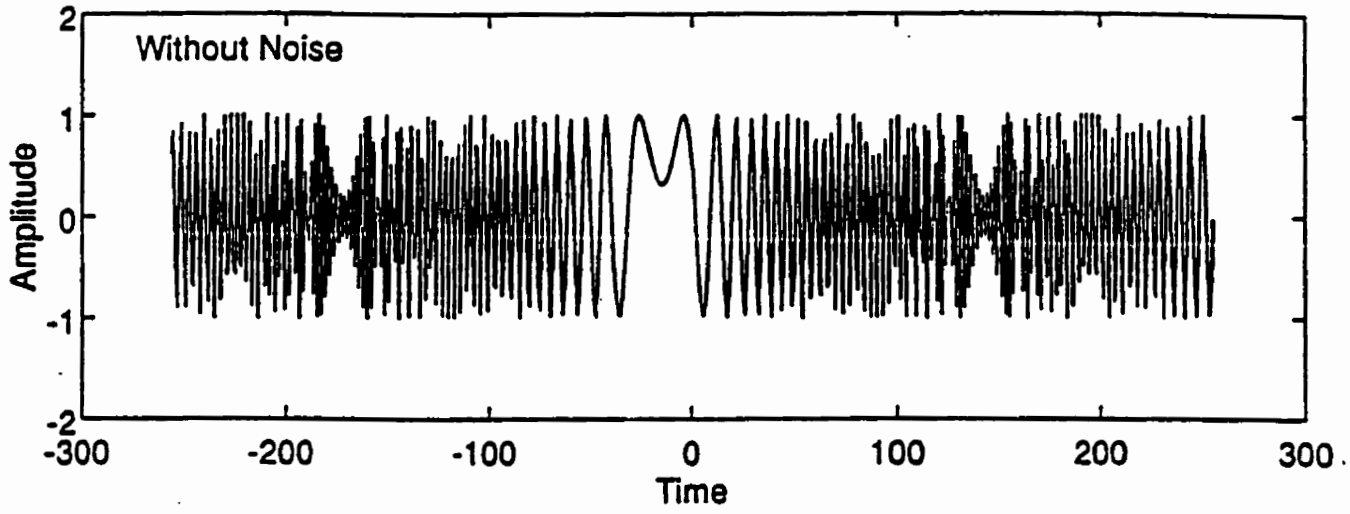

Figure 5.19: Time serise of a signal with chirp (no noise).

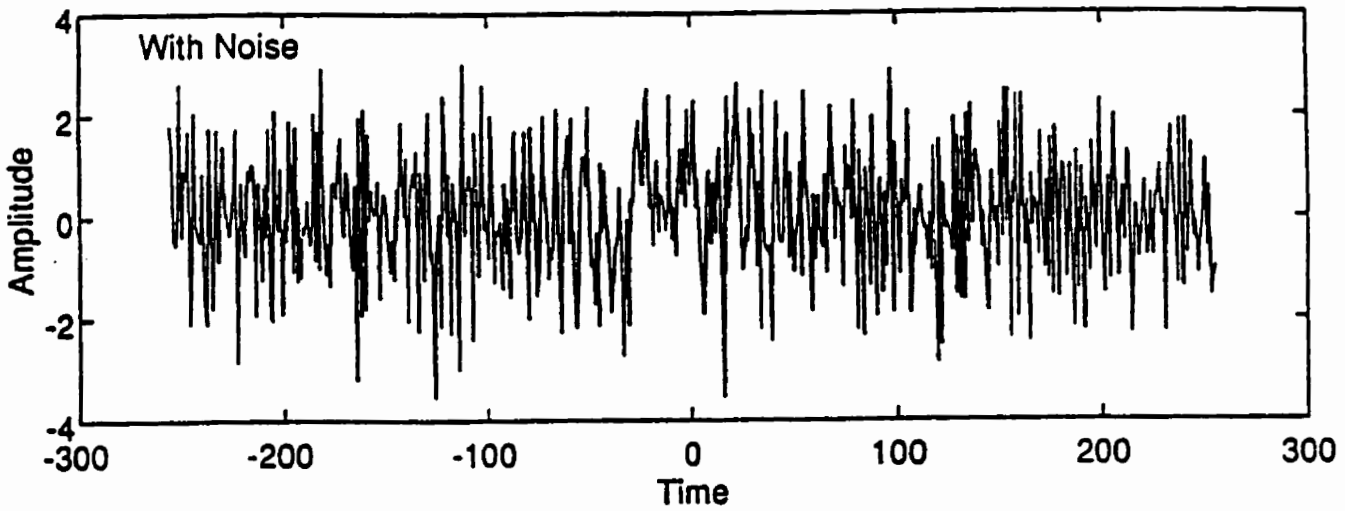

Figure 5.20: Time series of a signal with chirp (with noise).

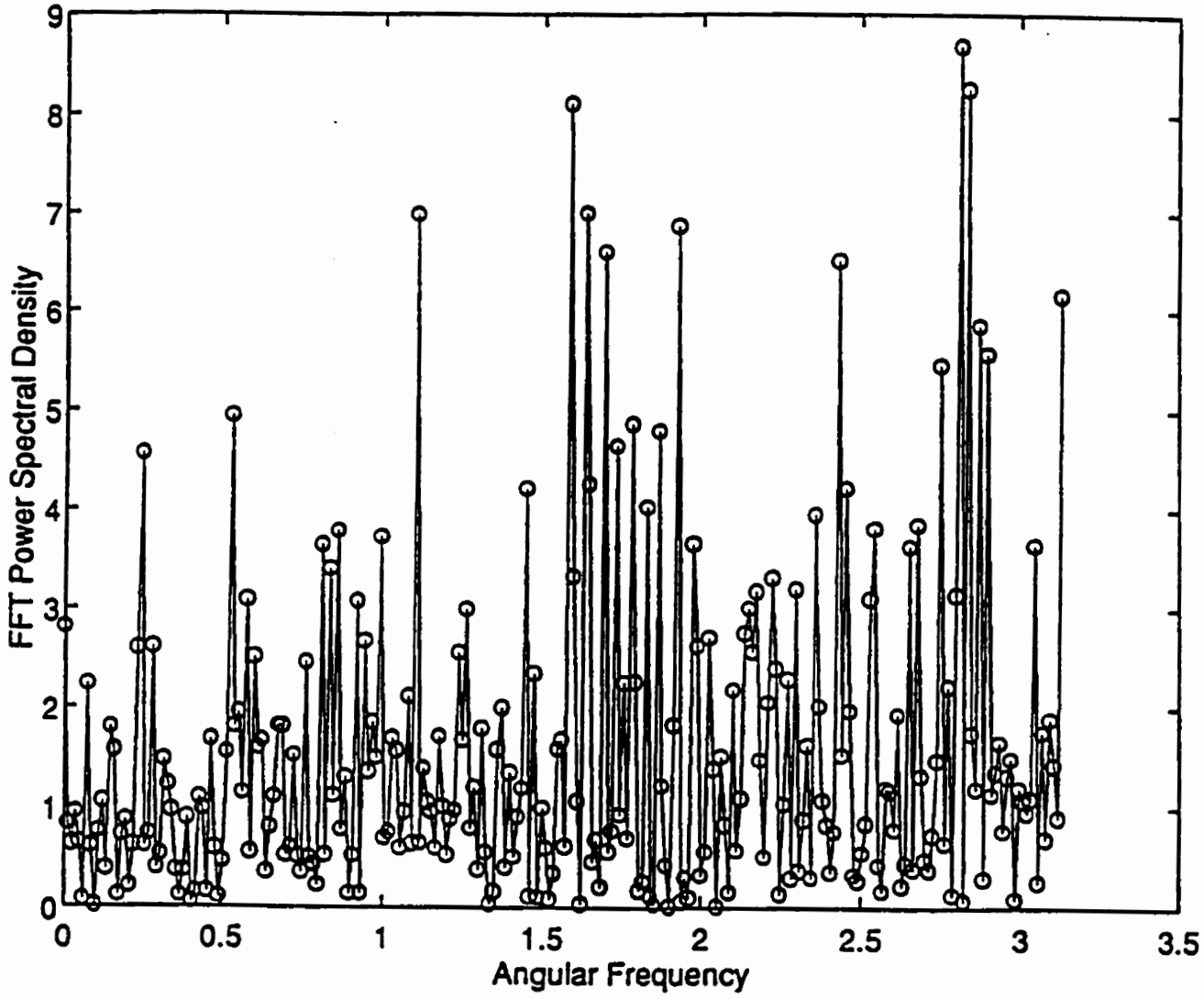

Figure 5.21: FFT spectrum.

5.4 Model Selection

5.4.1 Model space

When analyzing a spectrum from experimental data, it is not always known which model applies. We need a way to choose the best model from several possible models. For example, according to some prior information about the experiment, we may already know that the model functions are sinusoidal, and we may also know that the maximum number of frequency lines in the model is k. Hence there are k possible models, each with a different number of frequency lines. What we need to do in this situation is find out which one of the possible models is the best choice, and determine the corresponding parameters. Sometimes the situation may be even worse in that we don't even know what kind of function applies. We only know that there are several possible functions, any one of which could be the true function. In this situation there is one more step to carry out in comparison to the first case - that is choosing a best model function. In the following, a general procedure, which can deal with both of the above situations, is presented.

According to our knowledge about the experiment, all the possible models, $\{m_1, m_2, \ldots, m_k\}$, are enumerated, which is our model space. For each model there are some undetermined parameters. As a first step these parameters have to be found. The procedure for accomplishing this has already been described in 5.2. Now the question which can also be answered by Bayesian inference theory is "Given the model space and looking only within that model space, which model is most likely according to the information from all the data in conjunction with known prior information? Furthermore, how strongly is the most possible model supported relative to the others in the model space?" If the inferences we made are unacceptable,

are unacceptable, the problem must come from inappropriate assumptions about the model and the prior information and not from inadequacies of the inferential system, since given the model Bayesian analysis automatically makes use of all the information from the data. When this happens, at least we know that our model space is not a proper one, and other possible models need to be considered. In this way Bayesian theory is the kind of theory which either gives us the most possible answer or tells us that something is wrong with our assumptions. The latter case may then be regarded as prior information for a further investigation.

5.4.2 The best choice from a model space

In this section we will confine our attention to the model space which contains all the possible models $\{m_1, m_2, \ldots, m_k\}$ according to our knowledge about the problem. There is no reason to favor any one model over another. Therefore each of the models is equally possible. We may then assign a uniform a priori probability to the data D on the model space., i.e.:

$$P(m_1) = P(m_2) = \ldots = P(m_k) = const \tag{5.68}$$

From Bayesian theory the posterior probability of model m_i is

$$P(m_i \mid D) = \frac{P(m_i)P(D \mid m_i)}{P(D)} \tag{5.69}$$

where

$$P(D \mid m_i) = (1/\sqrt{2\pi})^N \sigma^{-N} \exp\{-\frac{1}{2\sigma^2}\sum_{j=1}^{N}[d_j - m_i(t_j)]^2\}$$

if the noise in the data is Normally distributed with zero mean and known variance σ^2.

Now what we need to do is to compare the posterior probabilities of the models in our model space. Because we don't know $P(D)$ which is independent of the model, we may calculate the relative posterior probabilities or posterior odds,

$$\frac{P(m_i \mid D)}{P(m_j \mid D)} = \frac{P(m_i)P(D \mid m_i)}{P(m_j)P(D \mid m_j)} = \frac{P(D \mid m_i)}{P(D \mid m_j)} \tag{5.70}$$

From these posterior odds the most probable model can be chosen and the confidence of choosing this model can be described by the posterior odds as well. To make all these ideas more clear, two numerical examples are presented in the next section.

5.4.3 Numerical examples

Consider a signal, Figure 5.22, which contains three close harmonic frequency lines generated as follows:

$$d_i = \cos(0.3i + 1) + 2\cos(0.31i + 2) + \cos(0.33i + 3) + \varepsilon_i \tag{5.71}$$

Here i runs over the symmetric interval $-T$ to T in integer steps, $(2T + 1) = 512$, and $\varepsilon_i \sim N(0, 1)$. The discrete Fourier transform of this signal, Figure 5.23, only shows one peak. From the plot of the signals itself, however, it is evident that there is more than one frequency involved. There may be one, two, three, or even four frequency lines, so in our model space we assume four models $\{m_1, m_2, m_3, m_4\}$, each of which consists of a sinusoidal function as the model function, but with a different number of frequency lines.

Now for each model, the best estimates of the parameters which fully fix the model are made by maximizing the marginal posterior probabilities for each parameter. We can then compute the posterior probability for each model, and then

Table 5.11: The Best Estimates of Parameters and Posterior Odds For a Signal with Three Close Frequencies Using Different Models

Model	Frequency	Amplitude	Phase	Sum-of-Squared Residuals and Posterior Odds
One Frequency	0.3101 ±0.0002	2.04 ±0.08	1.8 ±0.3	0.919×10^3 0.3×10^{-88}
Two Frequencies	0.3101 ±0.0002	2.16 ±0.09	1.9 ±0.2	0.717×10^3 0.3×10^{-44}
	0.3299 ±0.0004	0.87 ±0.08	2.79 ±0.07	
Three Frequencies	0.3000 ±0.0004	0.91 ±0.05	1.0 ±0.3	0.512×10^3 1
	0.3100 ±0.0002	2.0 ±0.1	2.0 ±0.2	
	0.3300 ±0.0004	0.95 ±0.08	2.90 ±0.08	
Four Frequencies	0.2995 ±0.0004	2.9 ±0.9	2.27 N/A	9.871×10^3 0
	0.3000 ±0.0004	3.3 ±0.8	2.4 ±0.8	
	0.3101 ±0.0002	2.0 ±0.1	2.0 ±0.2	
	0.329 ±0.008	0.9 ±0.3	2.9 ±0.2	

obtain the posterior odds relative to the model with three frequency lines. Finally the best choice for the model can be made from these posterior odds. In Table 5.11 the best estimates of the parameters for each model are tabulated together with the corresponding uncertainties, and the posterior odds for each model relative to the three-frequency model.

These results clearly show that the model with three frequency lines has the largest posterior odds. (Note that a posterior odds value of zero results when the

value is too small for the precision of the computer.)

In the next example, a single-frequency model with Lorentzian decay is considered.

$$d_i = (B\cos(\omega i + \phi))\exp^{-\alpha i} + \varepsilon_i \tag{5.72}$$

where

$$B = 1, \omega = 0.3, \phi = 1, \alpha = 0.01$$

In this case we will assume that we don't know which decay model applies. We will then assume that it is one of four possible models: a stationary signal $B\cos(\omega t+\phi)$, a Gaussian in time, $B\cos(\omega t+\phi)e^{-\alpha t^2}$, a Lorentzian in time, $B\cos(\omega t+\phi)(1+\alpha t^2)^{-1}$ and a Lorentzian in frequency, $B\cos(\omega t+\phi)e^{-\alpha t}$,

For each of the decay models, the associated parameters are estimated by using Bayesian inference with Gibbs sampling. Then the posterior odds for each model relative to the Lorentzian-in-frequency model are calculated. All of the results are tabulated in Table 5.12.

Obviously the Lorentzian-in-frequency decay model gives the largest posterior odds, so it is the best choice for the decay model, and the corresponding parameters are the best estimates for the amplitude, frequency, phase and decay rate.

Table 5.12: The Best Estimates of Parameters and Posterior Odds For a Signal with Unknown Decay Model

Decay Model	Stationary	Gaussian in Time	Lorentzian in Time	Lorentzian in Frequency
ω	0.3000 ± 0.0001	0.29 ± 0.01	0.301 ± 0.003	0.30000 $\pm 5 \times 10^{-5}$
B	2.5 ± 0.07	1.6 ± 0.2	2.5 ± 0.3	1.00 ± 0.02
ϕ	1.0 ± 0.1	0.5 ± 0.5	0.8 ± 0.3	1.02 ± 0.04
α	N/A	0.001	0.001	0.01
Sum-of Squared Residuals	3.246×10^3	4.769×10^3	4.682×10^3	0.517×10^3
Post. Odds	0	0	0	1

5.5 Conclusion

In this chapter we developed and applied Bayesian inference and Gibbs sampling to the spectral analysis and parameter estimation problem. It shows that this approach to spectral analysis is quite promising. Both frequency and amplitudes (then, of course, the phase shift) can be inferred from experimental data and the prior information with high accuracy, especially the frequency, which is the most important parameter in spectral analysis. We have also shown that two close frequencies which are too close to be distinguished by FFT can be clearly resolved using Bayesian inference and Gibbs sampling. Even when the noise level is unknown the estimates for all parameters, including the noise variance can be made accurately. The accuracy of our approach is either comparable with or higher than that employed by Bretthorst. All of the calculations are very straightforward. We also demonstrated the application to some very complicated models and conditions. We have shown that this approach can not only give us the best estimates for the parameters when

we know what the underlying model is, but it can also tell us which model is most probable in situations where we can specify a model space. The comparison of the true model and true parameters with the estimated model and parameters shows us the high accuracy and reliability of this approach.

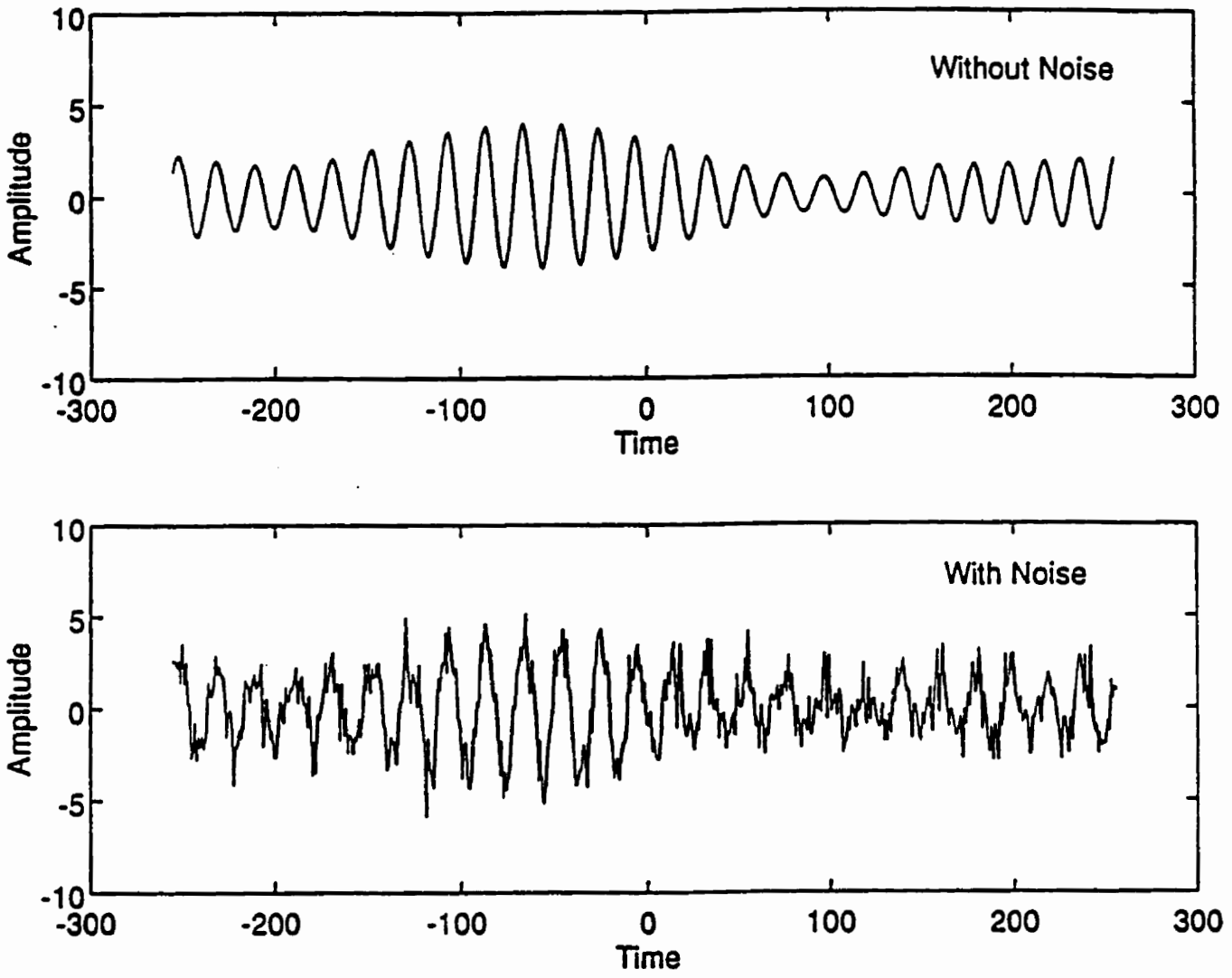

Figure 5.22: Time series of a signal with three close frequencies.

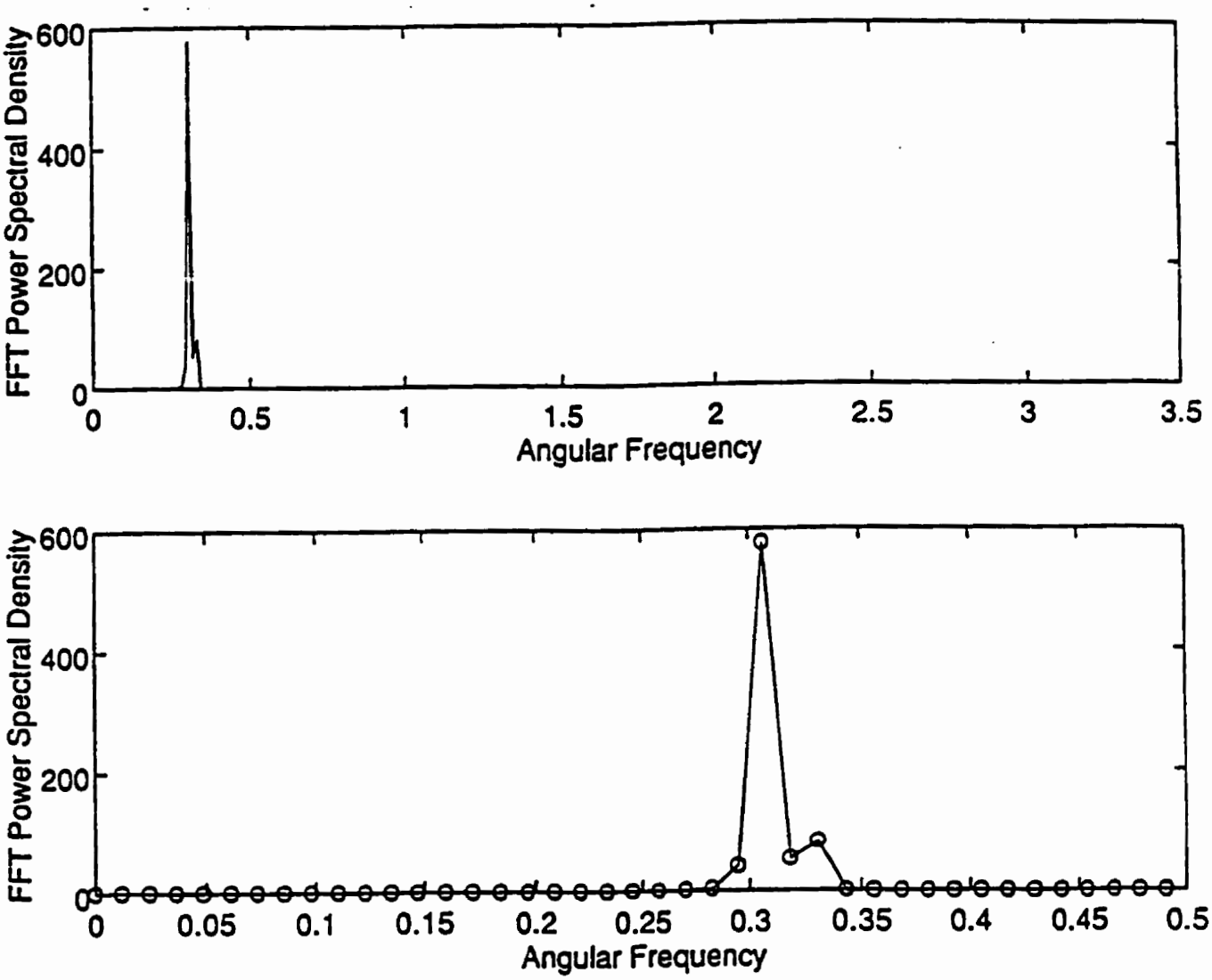

Figure 5.23: FFT power spectrum.

Chapter 6. Hyperfine Parameter and Field Distribution Extraction in Mössbauer Spectroscopy

6.1 Introduction

6.1.1 Mössbauer spectroscopy

Mössbauer spectroscopy is a nuclear spectroscopy technique that has an energy resolution sufficient to resolve the hyperfine structure of nuclear levels. This extraordinary technique is based on a physical phenomenon known as the Mössbauer effect, discovered by Rudolf Mössbauer in 1957. The effect allows recoilless emission and absorption of γ-rays by nuclei imbedded in solid substances. There are about 40 elements which have suitable Mössbauer transitions. Among them Fe-57 Mössbauer spectroscopy is the most commonly utilized and is the one used in this study. The advantage of Mössbauer spectroscopy is that the probe isotope reveals a spectrum which is sensitive to subtle variations in the local electronic and magnetic environment. This makes the technique very useful in materials science, mineralogy, chemistry, and condensed matter physics.

For the readers who are not familiar with the technique and interested in Mössbauer spectroscopy, there are many reviews and books which provide complete and excellent introductions to the technique. Among them Greenwood and Gibb's book "Mössbauer spectroscopy" [43] is most recommended.

6.1.2 Hyperfine interactions in Fe-57

The transition that produces the Fe-57 Mössbauer γ-ray is the decay of the $I = 3/2$ first nuclear excited state to the $I = 1/2$ ground state. There is a four fold degeneracy($(2I+1)$) in the excited state and a double fold degeneracy in the ground state. In an isolated nucleus the transition is a single line (see Figure 6.1) however there are three hyperfine interactions which can either affect the energy levels of the states or remove the degeneracies of the states partially or completely. They are:

1. Center shift.
2. Electric quadrupole splitting.
3. Magnetic hyperfine interaction.

The center shift is a electric monopole interaction which does not split the levels(i.e. it does not remove any of the degeneracy) but only affects their precise energy. The electric quadrupole interaction partially splits the degeneracies of nuclear staes in which nuclei have non-zero quadrapole moments. The magnetic hyperfine field interaction completely removes all the angular momentum(or spin) related degeneracies.

Center shift

The measured center shift(CS) or δ has two terms:

$$\delta = IS + SOD \tag{6.1}$$

where IS is the isomer shift and SOD is the second order Doppler shift. Both CS and IS are with respect to a standard reference shift such as the CS of metallic iron at room temperature ($RT = 22^oC$)

The isomer shift is caused by the electric monopole interaction between the nucleus and the surrounding electrons. The nucleus has a finite volume and this must be taken into account when considering nucleus-electron interactions because an s-electron wave function implies a non-zero electron charge density within the nuclear volume. The effect of this interaction is to shift the energy levels of the nucleus with respect to the bare nucleus without lifting the degeneracy of the two states (see Figure 6.2). This shift is written as:

$$IS = const_1\{\mid \psi_s(0)_A \mid^2\} + const_2 \tag{6.2}$$

where $\mid \psi_s(0)_A \mid^2$ is the total s-electron density at the nucleus in the absorber, and $const_2$ depends on the standard reference shift. $const_1$ depends only on nuclear properties of the ^{57}Fe probe.

In material samples where only the electric monopole interaction operates and where all probe nuclei are in identical electronic environment, the corresponding Mössbauer resonance is a single line called a singlet having a full width at half maximum(FWHM) equal to 2Γ, where the *Gamma* is the intrinsic linewidth of the nuclear transition.

The SOD shift is caused by the emitting or absorbing atom vibrating on its lattice site, which is a temperature dependent contribution. The contribution appears as a second order Doppler shift in the energy of the emitted photon. The fact that it is second order means that the shift depends on the square of the velocity of the atom. To calculate SOD a particular phonon model is required. In the Debye model, the SOD can be written as

$$SOD = \frac{< v^2 >}{2c} = -\frac{3k_B\theta_D}{2mc}[\frac{3}{8} + 3(\frac{T}{\theta_D})^2 \int_0^{\theta_D/r} \frac{x^3}{(e^x - 1)}dx] \tag{6.3}$$

where v is the nuclear vibration velocity, c is the speed of light, k_B is the Boltzman constant, Θ_D is the Debye temperature, and T is the temperature.

Electric quadrupole interaction

Any nucleus with a spin quantum number great than $I = 1/2$ has a non-spherical charge distribution which, if expanded as a series of multipoles, contains a quadrupole term. The magnitude of the charge deformation is described as the nuclear quadrupole moment Q. On the other hand, the electronic charge distribution surrounding the nucleus is usually not spherically symmetric, resulting in an EFG at the probe nucleus. The interaction between the nuclear quadrupole moment Q and the EFG is called the electric quadrupole interaction. The effect of this interaction, in addition to the electric monopole interaction, partially splits the $I = 3/2$ state into substates having $I_z = \pm 3/2$ and $I_z = \pm 1/2$. The centroid of the first excited state remains that determined by the electric monopole interaction alone. The splitting between the two sublevels is given (see Figure 6.3) by an amount

$$\Delta = \frac{1}{2} e^2 q Q (1 + \frac{1}{3}\eta^2)^{1/2} \tag{6.4}$$

where Q is the nuclear quadrupole moment, eq is the maximum value of the EFG , and η is the asymmetry parameter. We call Δ the quadrupole splitting.

The above situation results in the Mössbauer resonance consisting of two lines called doublet. If all the probe nuclei in the sample experience identical EFGs (and have identical CSs), then the corresponding absorption spectrum consists of two Lorentzian lines, each have FWHM of 2Γ and being separated by the *Delta*.

The relative depths (or intensities) of the two absorption lines in a doublet depend on the orientation of the largest EFG principle axis relative to the incident

γ-ray direction.

The EFG is a tensor. When a coordinate system is properly chosen the EFG tensor can be diagonalized. Then only two independent parameters are needed to specify the EFG completely, and the two which are usually chosen are

$$V_{zz} = \frac{\partial^2 V}{\partial z^2} \tag{6.5}$$

and

$$\eta = \frac{\mid V_{xx} - V_{yy} \mid}{V_{zz}} \tag{6.6}$$

where V is the electrostatic potential, and also $V_{zz} = eq$. Using the convention that $\mid V_{zz} \mid \geq \mid V_{yy} \mid \geq \mid V_{xx} \mid$ ensures that $0 \leq \eta \leq 1$.

Magnetic hyperfine interaction

The magnetic hyperfine interaction is the interaction between the nuclear magnetic dipole moment and the magnetic field that the nucleus experiences. This interaction splits the degeneracies of both the excited state and the ground state (see Figure 6.4). The energy shifts, in the absence of an electric quadrupole interaction and in addition to the monopole interaction, are given by

$$\Delta E_m = g_I \mu_N m_I H \tag{6.7}$$

where g_I is the level specific nuclear g-factor, μ_N is the nuclear magnetron, and H is the magnetic field experienced by the nucleus. The field can originate either within the atom itself, within the crystal via exchange interactions or as a result of placing the compound in an externally applied magnetic field. The Mössbauer resonance corresponds to eight possible transitions between the the two sublevels of the ground state and the four levels of the excited state. In this case because of $q = 0$ (and

in certain cases where $q \neq 0$), two of the eight possible transitions are forbidden by a quantum mechanical selection rule, the Mössbauer resonance consists of six lines called sextet. If all the probe nuclei in the sample experience indentical HFs (and have identical CSs), then the corresponding absorption spectrum consists of six Lorentzian lines, each having a FWHM of 2Γ and positions that depend on the magnitude of H. The relative intensities of the six lines depend on the orientation of the HF relative to the incident γ-ray direction.

In practice, the electric quadrupole interaction and magnetic hyperfine interaction could both exist at the same time. Then the energy levels have to be determined by the solution of the full hyperfine Hamiltonian. Since both the magnetic and quadrupole interactions are direction dependent, the general interaction is quite complex. However there are some special situations which will make the determination simpler. We will discuss these situations in the following sections gradually.

6.1.3 Mössbauer spectrum and static hyperfine parameter distributions

There are different ways to interpret the Mössbauer spectrum depending on the different circumstances. In this study, we exclusively interpret the Mössbauer spectrum in terms of static HFDs in that we assume that dynamic effects are negligible in all the cases studied. As we have seen in section 6.1.2, the probe nuclear energy levels depend on the three interactions which are characterized by the hyperfine parameters, such as δ, Δ, and H_{hf}. The following illustration presents a picture of how the Mössbauer spectrum is related to the hyperfine parameters.

Suppose in an absorber all the probe nuclei experience the same center shift interaction, the same electric quadrupole interaction, and no HF interaction. The corresponding Mössbauer spectrum is then just an elemental Lorentzian doublet

with the Heisenberg natural line width (see Figure 6.6(a)). If there are two families of probe nuclei with the same valence state (we will talk about probe nuclei with different valence states in section 6.4), but different local surroundings such as different bonding anions, local symmetry, local distortions, then the energy levels for nuclei in one family will be different from the energy loevels for the nuclei in the other family, i.e. the two δ's are different, as are the two quadrupole splittings. Hence the Mössbauer spectrum is a composition of two elemental Lorentzian doublets (see Figure 6.6(b)) with different δ and different Δ. However various solids of interest contain probe nuclei in wide ranges of different local environments which lead to effectively continuous distributions of static hyperfine parameters. The corresponding Mössbauer spectrum is a superposition of elemental Lorentzian doublets with δ and Δ continuously distributed (see Figure 6.6(c)). A simple mathematical model can be formulated to describe this:

$$M(v) = \int\int E(\delta, \Delta; v) P(\delta, \Delta) d\delta d\Delta \tag{6.8}$$

where $M(v)$ is the Mössbauer spectrum, $E(\delta, \Delta; v)$ is the elemental Lorentzian doublet, and $P(\delta. \Delta)$ is the distribution function for δ and Δ.

In general when all the interactions are present, we may have:

$$M(v) = \int\int E(p_1, p_2, \ldots, p_n; v) P(p_1, p_2, \ldots, p_n) d\delta d\Delta \tag{6.9}$$

assuming the absence of artifacts such as thickness effects. We call $E(p_1, p_2, \ldots, p_n; v)$ the elemental line shape which is a function of the hyperfine parameters, $p_1, p_2, \ldots, p_n$. $P(p_1, p_2, \ldots, p_n)$ is the distribution function for the parameters. The goal of this study is to extract meaningful hyperfine parameter distributions from spectra, based on the model equation (6.9). Before we get into the study, a brief review of previous

work in hyperfine parameter distribution extraction is given in the next subsection.

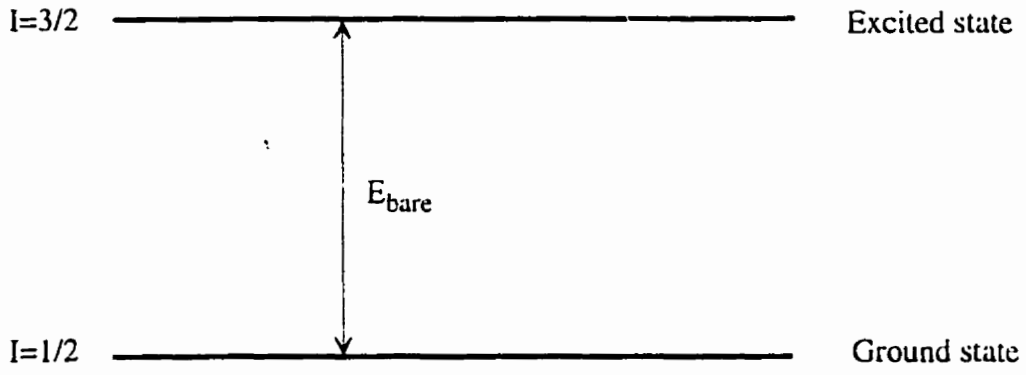

Figure 6.1: Energy level diagram of Fe-57: the standard isolated nucleus.

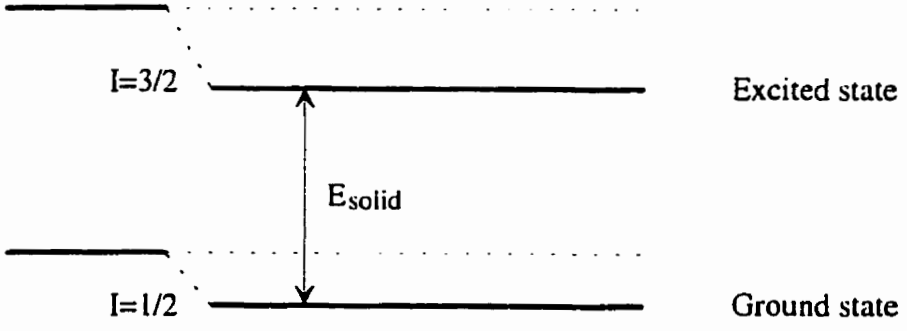

Figure 6.2: Energy level diagram of Fe-57: the effect of the isomer shifts which shifts both states.

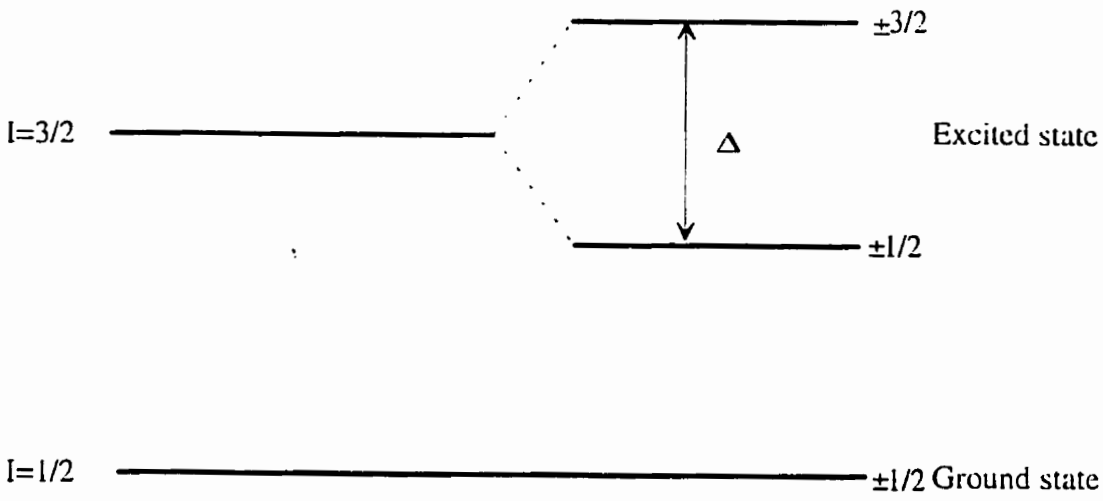

Figure 6.3: Energy level diagram of Fe-57: the electronic quadrupole interaction which splits the excited state into a doublet.

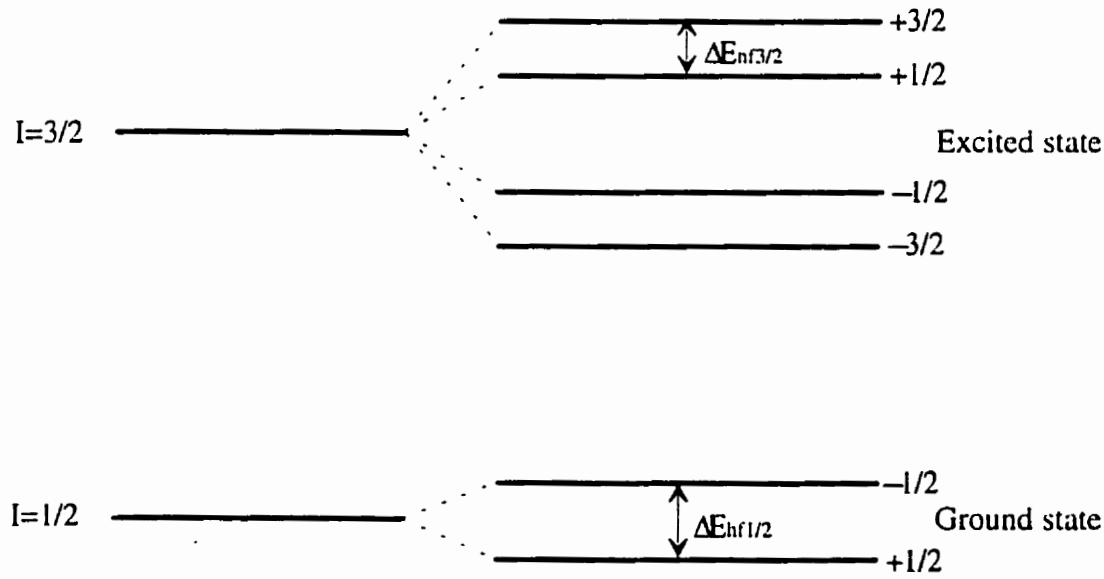

Figure 6.4: Energy level diagram of Fe-57: the effects of the magnetic hyperfine interaction (no quadrupole interaction) which removes the degeneracies of both states.

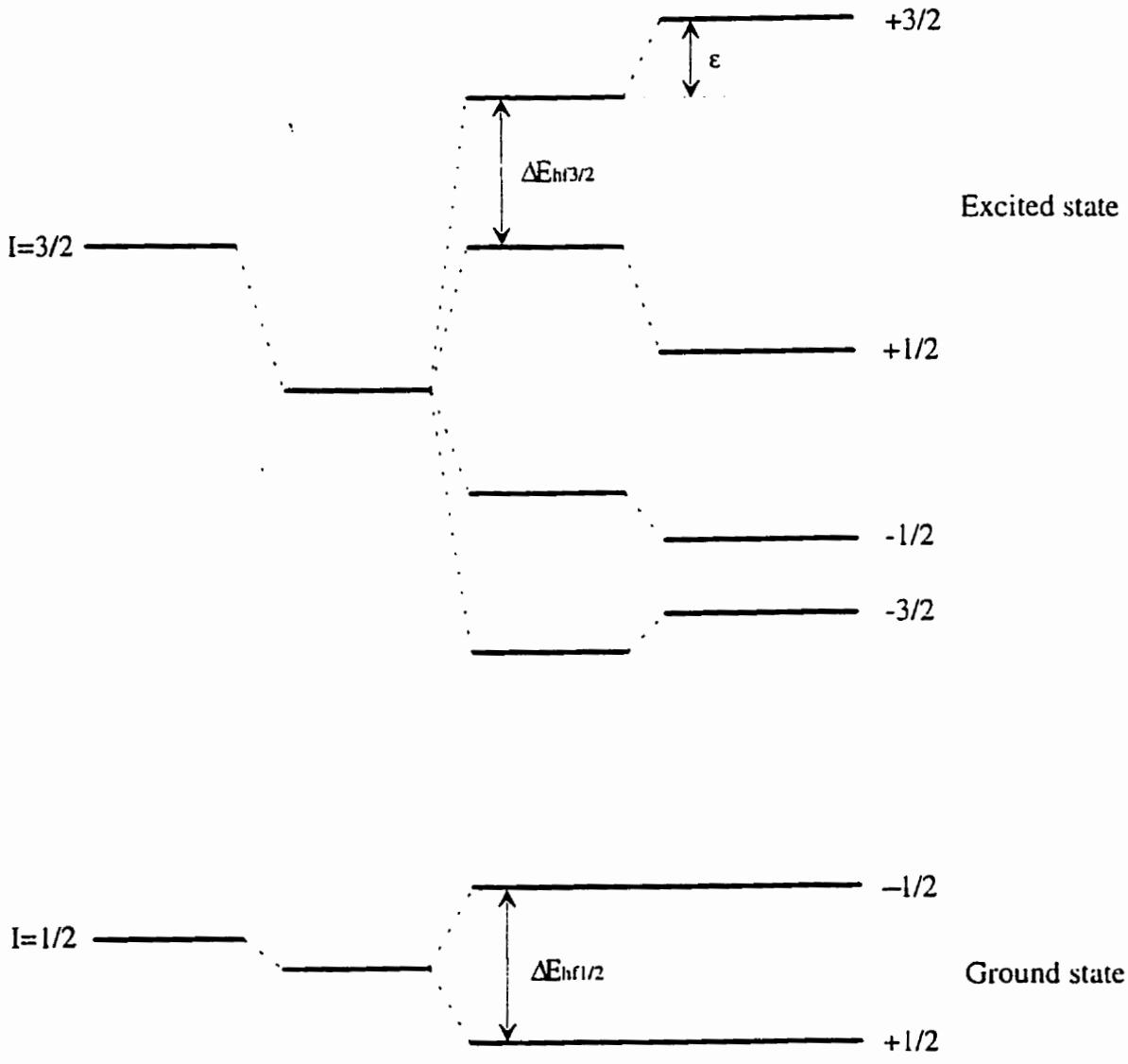

Figure 6.5: Energy level diagram of Fe-57: the effects of all hyperfine interactions when $e^2qQ/4 << g^{3/2}\mu_N H$.

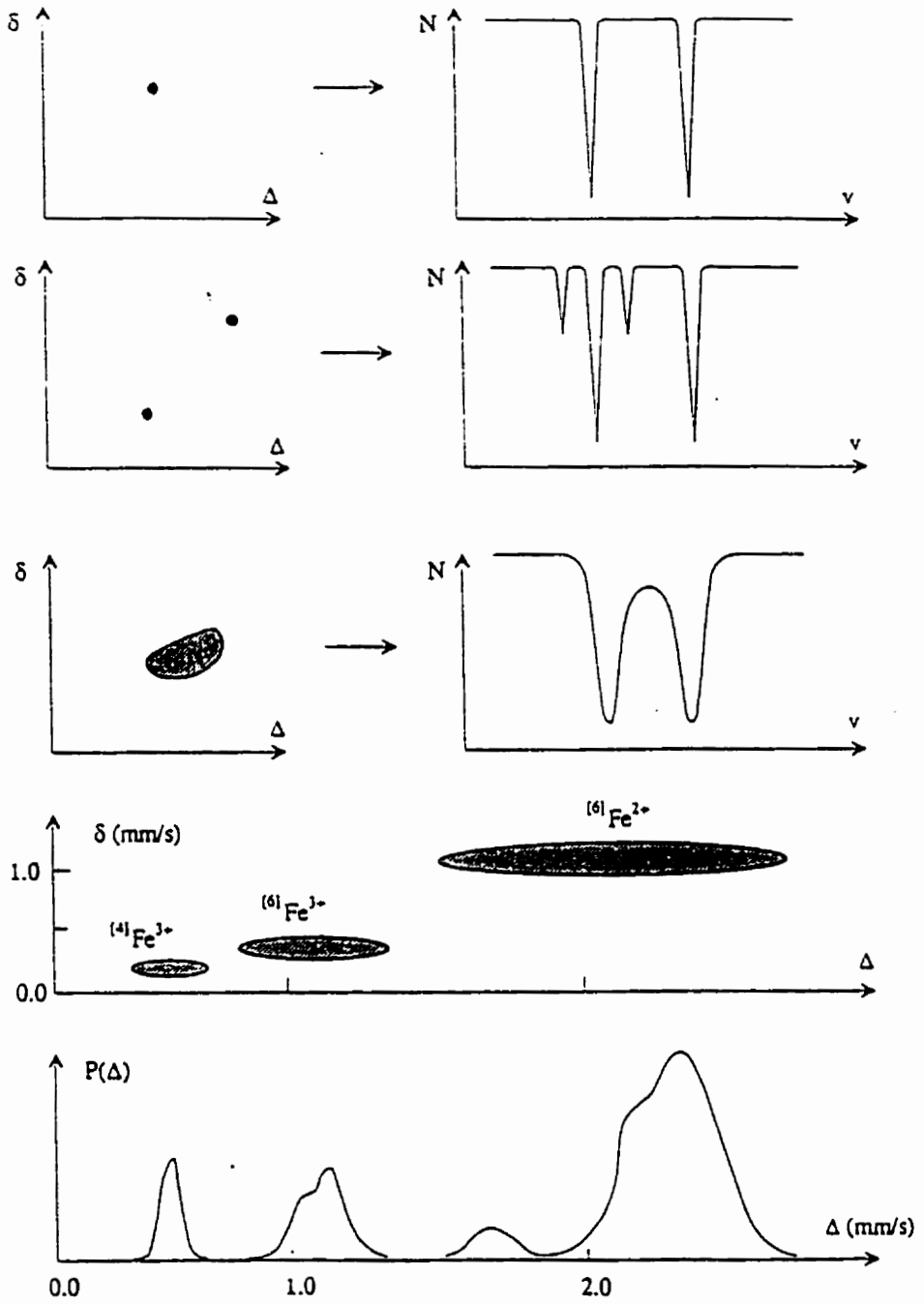

Figure 6.6: Relation between $\delta - \Delta$ distribution and Mössbauer spectrum.

6.1.4 Review of previous work in hyperfine parameter distribution extraction

The extraction of meaningful distributions of hyperfine parameters from measured Mössbauer spectra is a nontrivial task for which many methods have been proposed and developed[45]-[57] and about which several reviews have been written[58]-[64]. The different approaches can be classified as: i) simple analytical functions where the distribution is assumed to have a particular shape (especially Gaussian), ii) Fourier methods (e.g. Window's method), iii) step functions used either in conjunction with a physical model for the discrete hyperfine parameter values or as an approximation to a continuous distribution, iv) improved step function methods with smoothing, end point restrictions, positive value restrictions, and so on, v) the Voigt-based fitting method of Rancourt and Ping for arbitrary-shape continuous distributions, and vi) the combined partial deconvolution method of Ping and Rancourt. Here we distinguish fitting methods (i, iii, iv, and v) which use non-linear least-square fitting and direct extraction methods (ii and vi) that obtain the distribution in a single mathematical operation, usually involving matrix inversion that may be equivalent to linear least-square minimization.

All methods are plagued to various degrees by several problems such as: non unique solutions due to tradeoff effects, unphysical oscillations near rapidly varying features, unavoidable (and meaningless) negative components, excessive numbers of free parameters, overly large and nonlinear program structures leading, respectively, to slow and ambiguous convergence, difficulty in handling simultaneous distributions of more than one hyperfine parameter with correlation between the different distributions, difficulty in handling distributions containing both relatively sharp and broad features, etc.. Even the most powerful Voigt-based fitting (VBF) method

still has two main disadvantages: (1) the usual perturbation sextet must be used as the elemental subspectrum, and (2) as in all other methods, error analysis of the resulting HFD is not performed.

In the next sections, we will see for the first time that the Bayesian inference theory can solve these problems and permits several different elemental line shape models to be tested. Before we get into the next section, a brief introduction about the VBF method is presented since we will compare the Bayesian inference method with the VBF method. Also a preliminary study using the Voigt-based fitting method can give us some useful information about a spectrum.

6.1.5 Voigt-based fitting method

As described above, the VBF method is the most powerful fitting method at present. This method was devised by Rancourt and Ping for arbitrary-shape continuous distributions in 1991 [55]. It has been applied to spectra arising from different materials and under different conditions. It is applicable in all cases involving quadrupole splitting distributions or hyperfine field distributions, in the perturbation limit, or to coexisting quadrupole splitting distributions and hyperfine field distributions, when all couplings to non-distributed parameters are at most linear and as long as the true distributions are not so pathological that they can not be represented as sums of Gaussian components. The central physical assumption in the method is that the distribution can be expressed as a sum of Gaussians with different widths, positions, and relative areas

$$P(p) = \sum_{i=1}^{R} w_i G_i(p_{oi}, \sigma_{pi}; p) \tag{6.10}$$

where

$$G_i(p_{0i}, \sigma_{pi}; p) = \frac{1}{\sigma_{pi}\sqrt{2\pi}} exp\left[\frac{-(p - p_{oi})^2}{2\sigma_{pi}^2}\right] \tag{6.11}$$

and

$$\int P(p)dp = \sum_i^R w_i \int G_i dp = \sum_i^R w_i = 1 \tag{6.12}$$

For quadrupole splitting distributions , it is assumed that the center shift, δ, is linearly coupled to Δ, i.e.,

$$\delta = \delta_0 + \delta_1 \Delta \tag{6.13}$$

and, with hyperfine field distributions, one assumes that both the center shift,

$$\delta = \delta_0 + \delta_1 g_{3/2} \mu_N H \tag{6.14}$$

and the quadrupole shift,

$$\epsilon = \epsilon_0 + \epsilon_1 g_{3/2} \mu_N H \tag{6.15}$$

are linearly coupled to H. It follows that, given equation(6.10), the analytical expression for the total Mössbauer lineshape is

$$M(v) = \sum_{i=1}^R w_i \sum_k V_{ik}(A_k + B_k p_{oi}, \mid B_k \mid \sigma_{pi}, \Gamma, h_k; v) \tag{6.16}$$

where $V_{ik}(\alpha, \beta, \Gamma, h; v)$ is a Voigt function[109]. Here, the sum on k has two terms for quadrupole splitting distributions and six terms for hyperfine field distributions. The parameters A_k and B_k are functions of the coupling parameters, $\delta_0, \delta_1, \epsilon_0$ and ϵ_1, which can be treated as fitting parameters. In their numerical calculation a particularly efficient and accurate approximation to the Voigt function is used [109],[55], [72] and equation (6.16) is incorporated into a least squares minimization routine. For detailed applications to hyperfine field distributions see references [55] and [108]. For applications to quadrupole splitting distributions see references [73] and [74].

6.2 Numerical Procedures and Algorithms

In this section the numerical procedures and algorithms of the Bayesian inference method for the general case is introduced in detail. The concrete elemental lineshape and parameter distribution function could be different for different circumstances. The exact expressions will be introduced in the applications.

6.2.1 Numerical expression of the problem

When we study a specific problem, equation (6.9) can often be simplified to a single parameter distribution with coupling to other parameters

$$M(v) = \int E(p; v)P(p)dp \tag{6.17}$$

where $E(p; v)$ could be a doublet, with p representing the quadrupole splitting; or $E(p; v)$ could be a sextet, in which case p is the hyperfine field parameter when magnetic order is present. What we want to find from the spectrum is the distribution function of hyperfine parameters, given the elemental line shape $E(p; v)$. This is an example of an inverse problem involving a Fredholm equation of the first kind [44]. As we know there is always noise in any experimental data, so the experimental spectrum should be expressed as:

$$M(v) = \int E(p; v)P(p)dp + \mathcal{N}(v) \tag{6.18}$$

where $\mathcal{N}$ represents the noise. In numerical calculations, equation (6.18) has to be made discrete. Using M discrete velocity channels and N regularly spaced values of p, equation (6.18) can be written in matrix notation as

$$[M]_{M\times 1} = [E]_{M\times N}[P]_{N\times 1} + \mathcal{N}_{M\times 1}$$

or simply

$$M = EP + \mathcal{N}. \tag{6.19}$$

Here M is a vector with elements $M(v_i)$, $i = 1, \cdots, M$ and P is a vector with elements $P(p_j)$, $j = 1, \cdots, N$. E is an $M \times N$ matrix with elements $E(p_j; v_i)\Delta p$, where Δp is the hyperfine parameter spacing $\Delta p = p_{j+1} - p_j$. $\mathcal{N}$ is a vector with elements $\mathcal{N}(v_i)$, $i = 1, \cdots, M$. $\mathcal{N}(v_i)$ is distributed as a Normal distribution with zero mean and known variance σ_i^2, $i = 1, \cdots, M$, equal to the square root of the measured spectrum datum at the corresponding velocity channel [117].

For convenience we let

$$b = \begin{bmatrix} b_1 \\ \vdots \\ b_M \end{bmatrix} = \begin{bmatrix} M(v_1)/\sigma_1 \\ \vdots \\ M(v_M)/\sigma_M \end{bmatrix}; \tag{6.20}$$

$$A = \begin{bmatrix} a_{11} & \cdots & a_{1N} \\ \vdots & \vdots & \vdots \\ a_{M1} & \cdots & a_{MN} \end{bmatrix} = \begin{bmatrix} E(p_1; v_1)\Delta p/\sigma_1 & \cdots & E(p_N; v_1)\Delta p/\sigma_1 \\ \vdots & \vdots & \vdots \\ E(p_1; v_M)\Delta p/\sigma_M & \cdots & E(p_N; v_M)\Delta p/\sigma_M \end{bmatrix}; \tag{6.21}$$

$$X = \begin{bmatrix} x_1 \\ \vdots \\ x_N \end{bmatrix} = \begin{bmatrix} P(p_1) \\ \vdots \\ P(p_N) \end{bmatrix}; \tag{6.22}$$

$$\mathcal{E} = \begin{bmatrix} e_1 \\ \vdots \\ e_M \end{bmatrix} = \begin{bmatrix} \mathcal{N}(v_1)/\sigma_1 \\ \vdots \\ \mathcal{N}(v_M)/\sigma_M \end{bmatrix}, \tag{6.23}$$

and equation (6.19) becomes

$$b = AX + \mathcal{E} \tag{6.24}$$

where $\mathcal{E}$ is a vector and the probability distribution of $\mathcal{E}$ is a multivariate Normal distribution, $N_M(0, I)$, of dimension M where 0 is a vector with all M elements equal to zero and I is an $M \times M$ identity matrix.

Given equation (6.24), and given the experimental data b, what is the best estimate for X, and what is the uncertainty of this best estimate? Bayesian inference theory can answer this question from a statistical point of view. The probability distribution of X posteriori to b is

$$P(X \mid b) = \frac{P(b \mid X)P(X)}{P(b)} \tag{6.25}$$

where $P(b \mid X)$ is the likelihood function of X, and $P(X)$ is the prior probability distribution of X. Since $P(b)$ is just a constant for given b, the posterior probability distribution of X is simplified to

$$P(X \mid b) \propto P(b \mid X)P(X) \quad or \quad P(X \mid b) \propto L(X \mid b)P(X). \tag{6.26}$$

From a statistical point of view, the best estimate for X should be that $\hat{X}$ which maximizes the posterior probability distribution of X. It is now necessary to derive expressions for the likelihood function and the prior probability distribution of X.

6.2.2 The likelihood function

The likelihood function of X is that function which tells us how likely it is to obtain the data b if X is given. Since the elemental lineshape A is known, according to equation (6.24) the difference between the data b and AX is just the noise $\mathcal{E}$. Obviously, therefore, the required likelihood function is equal to the probability distribution of the noise $\mathcal{E}$. Since $P(\mathcal{E})$ is a multivariate Normal distribution, we

have

$$\begin{aligned} L(X \mid b) = P(\mathcal{E}) &\equiv \frac{1}{(\sqrt{2\pi})^M} exp[-\frac{1}{2}\mathcal{E}^T\mathcal{E}] \\ &= \frac{1}{(\sqrt{2\pi})^M} exp[-\frac{1}{2}(b-AX)^T(b-AX)]. \end{aligned} \tag{6.27}$$

If we don't have any specific information about the distribution of the hyperfine parameter before we have the data, we might choose a constant prior, i.e. $P(X) = const$ for X. In this case the posterior probability distribution of X is :

$$P(X \mid b) \propto exp[-\frac{1}{2}(b-AX)^T(b-AX)]. \tag{6.28}$$

Hence the best estimate for X is that $\hat{X}$ which maximizes the posterior or minimizes the exponent $\Phi(X) = (b-AX)^T(b-AX)$. This is just the maximum likelihood or least-square estimator

$$\hat{X} = (A^TA)^{-1}A^Tb. \tag{6.29}$$

Thus Bayesian inference theory gives us a justification to use the least-square estimator. However the least-square estimator equation (6.29) can only give reasonable solutions [66] when 1) the data has a sufficiently large signal/noise (S/N) ratio and, 2) the condition number of the matrix A is small. "*Small*" here means small enough to make the estimator insensitive to the particular noise level in the data. In practice most inverse problems do not satisfy this condition. This means that we cannot expect data to determine the solution accurately by itself. The prior information becomes crucial. Indeed, the limits of equation (6.29) (and its $N = M$ simplification: $\hat{X} = A^{-1}b$) have been tested extensively in the context of Mössbauer spectroscopy [56].

6.2.3 The prior probability distribution and error analysis

It is quite reasonable that from physics we may have some information about the distribution of the hyperfine parameter before we analyze the data. By imposing this prior information into our calculation, we hope to find a solution which is a good approximation to the desired unknown solution. The least-square estimator usually gives a solution with such violent oscillations that the real solution is submerged by these oscillations [56, 67]. The cause for these oscillations is not the least-square method, but the problem itself, i.e. the problem is ill-conditioned. On the other hand, from physics we know that the distribution of the hyperfine parameter will often be a reasonably smooth function. In this case, from various possible smoothness conditions, we assume the distribution function of the hyperfine parameter to have finite second derivatives, i.e. we assume

$$X^T L_2^T L_2 X = C_1 \tag{6.30}$$

where L_2 is the second derivative operator. With this constraint we can get rid of those solutions with high oscillations from the solution space since those solutions have infinite second derivatives. Another reason for choosing this constraint is that the prior probability distribution of X constructed from this constraint can be easily combined with the likelihood function to give a standard distribution for the posterior probability distribution of X. Consequently we can obtain analytic expressions for the most probable solution and the relevant errors.

Hesse and Rübartsch [47] used the same smoothness condition (6.30) to regularize their numerical calculation and obtained a similar solution to ours. Because they did not work within the framework of Bayesian inference theory, however, they had

no way to either: (i) determine the smoothness parameter that affects the solution, or (ii) evaluate the error in the resulting solution (i.e. distribution).

Now the question is how to formulate this constraint (6.30) into the prior probability distribution of X. Here the maximum entropy principle can be applied [68]. Note that we use the maximum entropy principle only to obtain the prior probability distribution given constraint (6.30), without using the spectral data in any way. Brand and LeCaër [53] have used maximum entropy in a totally different way and get a totally different result. They apply it directly to determining the distribution from the spectrum by maximizing the entropy of the distribution itself subject to the chi squared criterion.

Suppose there is a set of discrete functions $\{X_1, X_2, \cdots, X_K\}$ which are the K possible solutions for the distribution function of the hyperfine parameter, i.e. the possible solutions for equation (6.24). We can assign a set of probability values $\{P(X_1), P(X_2), \cdots, P(X_K)\}$ to these possible solutions, with normalization condition $\sum_{k=1}^{K} P(X_k) = 1$. According to maximum entropy theory, the Shannon measure of the entropy of $P(X_k)$

$$S = -\sum_{k=1}^{K} P(X_k) \log P(X_k) \tag{6.31}$$

should be maximized subject to

$$\sum_{k=1}^{K} P(X_k) = 1$$

and

$$\sum_{k=1}^{K} P(X_k)(X_k^T L_2^T L_2 X_k) = C_1.$$

This leads to a prior (see Appendix A) which is a Gaussian function of X:

$$P(X) \propto exp(-\frac{1}{2}\lambda^2 X^T L_2^T L_2 X) \tag{6.32}$$

where $\lambda^2/2$ is a Lagrangian constant. If we know the value of the constant C_1, λ can be determined from it. Unfortunately the value of C_1 is unknown. We need a way to determine λ. This will be discussed in Section 6.2.4.

From equations (6.26), (6.28) and (6.32), we obtain the posterior probability distribution of X as

$$P(X \mid b) \propto exp\left\{-\frac{1}{2}\left[(b - AX)^T(b - AX) + \lambda^2 X^T L_2^T L_2 X\right]\right\}. \tag{6.33}$$

Therefore the most probable solution to $b = AX + \mathcal{E}$ is that $\hat{X}$ which maximizes this posterior or minimizes the exponent

$$\Phi(X) = (b - AX)^T(b - AX) + \lambda^2 X^T L_2^T L_2 X.$$

By doing some simple calculations (see Appendix B), we obtain

$$\hat{X} = (A^T A + \lambda^2 L_2^T L_2)^{-1} A^T b. \tag{6.34}$$

This is just the Tikhonov regularization solution [69] to $b = AX + \mathcal{E}$ with L_2 as the regularization operator. Using $\hat{X}$, $\Phi(X)$ can be expressed as (see Appendix B)

$$\Phi(X) = (X - \hat{X})^T(A^T A + \lambda^2 L_2^T L_2)(X - \hat{X}) + C_2 \tag{6.35}$$

where $C_2 = b^T b - \hat{X}^T(A^T A + \lambda^2 L_2^T L_2)\hat{X}$, which is independent of X. Finally the normalized posterior probability distribution of X can be written as,

$$P(X \mid b) = \frac{\mid A^T A + \lambda^2 L_2^T L_2 \mid^{1/2}}{(\sqrt{2\pi})^N} exp\left[-\frac{1}{2}(X - \hat{X})^T(A^T A + \lambda^2 L_2^T L_2)(X - \hat{X})\right]. \tag{6.36}$$

This posterior tells us that X is distributed as a multivariate Normal distribution

$$X \propto N_N\left(\hat{X}, [A^T A + \lambda^2 L_2^T L_2]^{-1}\right).$$

Thus the marginal distribution of each element x_j, $j = 1, \cdots, N$ in X is distributed as $N(\hat{x}_j, c_{jj})$ with mean $\hat{x}_j$ and variance c_{jj} where c_{jj} is the jth diagonal element of $[A^T A + \lambda^2 L_2^T L_2]^{-1}$. This means that the true one-sigma standard deviation error in each point $\hat{x}_j$ of the extracted hyperfine parameter distribution is c_{jj}. In other words, given our extraction procedure (with particular choice of prior to which the solution is not sensitive) and given the particular values of the non-distributed hyperfine and spectral parameters, each point in the extracted distribution, for many spectral data sets from identically performed experiments, has a Gaussian distribution with standard deviation c_{jj} that is a direct consequence of the spectral noise. This represents a complete and precise error analysis of the distribution problem in Mössbauer spectroscopy.

6.2.4 Determination of λ

Before we can proceed to the numerical calculation, we have to find a way to choose a proper value for λ. The best solution $\hat{X}$ maximizes $\Phi(X)$. In $\Phi(X)$ there are two parts, one is $X^T L_2^T L_2 X$ from the prior, another is $(b - AX)^T(b - AX)$ from the likelihood or data. It has been shown [70] that a plot of $X^T L^T L X$, where L is a general regularization operator, including L_2, versus $(b - AX)^T(b - AX)$ with different $\lambda's$ is almost always shaped like an L-curve as in Figure 6.7. The vertical part of the L-curve corresponds to solutions where $X^T L^T L X$ is very sensitive to the changes in λ. The horizontal part corresponds to solutions where $(b-AX)^T(b-AX)$ is very sensitive to the changes in λ. The $\lambda's$ corresponding to the vertical part take less prior and more likelihood into the solution than those $\lambda's$ corresponding to the horizontal part. An optimal choice is that λ which corresponds to the corner of the L-curve. This λ gives a good compromise between the prior and the likelihood.

In numerical calculations, first by calculating $\hat{X}_\lambda$ for different λ values, the corresponding $\hat{X}_\lambda^T L_2^T L_2 \hat{X}_\lambda$ and $(b - A\hat{X}_\lambda)^T(b - A\hat{X}_\lambda)$ can be evaluated for each λ value. Then by plotting $log[\hat{X}_\lambda^T L_2^T L_2 \hat{X}_\lambda]$ versus $log[(b - A\hat{X}_\lambda)^T(b - A\hat{X}_\lambda)]$ and locating the corner, that point on the L-curve having maximum curvature, an optimal value for λ can be determined. This is an efficient procedure that does not present any particular difficulty.

6.2.5 Finding $\hat{X}$ using generalized singular value decomposition(SVD)

When we do numerical calculations, the direct use of the expression $\hat{X} = (A^T A + \lambda^2 L_2^T L_2)^{-1} A^T b$ to compute $\hat{X}$ is not recommended. Instead, a technique called generalized singular value decomposition (GSVD) is employed. It has been proven [71] that given

$$A \in R^{M \times N} \quad L_2 \in R^{H \times N} \quad M \geq N \geq H$$

with

$$rank(L_2) = H \quad and \quad rank\left(\begin{bmatrix} A \\ L \end{bmatrix}\right) = N$$

then there exist matrices

$$U \in R^{M \times N}, \quad V \in R^{H \times H}$$

with

$$U^T U = I_N, \quad V^T V = I_H$$

and a nonsingular

$$W \in R^{N \times N}$$

such that

$$A = UZW^{-1}, \quad L_2 = V[\Gamma, 0]W^{-1} \tag{6.37}$$

where

$$Z = diag(\zeta_1, \cdots, \zeta_H, 1, \cdots, 1) \in R^{N \times N} \tag{6.38}$$

$$\Gamma = diag(\gamma_1, \cdots, \gamma_H) \in R^{H \times H} \tag{6.39}$$

and such that

$$0 \leq \zeta_1 \leq \cdots \leq \zeta_H \leq 1;$$

$$1 \geq \gamma_1 \geq \cdots \geq \gamma_H > 0;$$

$$\zeta_i^2 + \gamma_i^2 = 1 \quad i = 1, \cdots, H.$$

The quantities $\alpha_i \equiv \zeta_i/\gamma_i,\ i = 1, \cdots, H$ are termed the generalized singular values of (A, L_2).

For convenience, the matrices U, Z and W are partitioned as follows:

$$U = [U_H, U_0]; \quad Z = \begin{bmatrix} Z_H & 0 \\ 0 & I_0 \end{bmatrix}; \quad W = [W_H, W_0] \tag{6.40}$$

where

$$U_H \in R^{M \times H}; \quad U_0 \in R^{M \times (N-H)};$$

$$Z_H \in R^{H \times H}; \quad I_0 \in R^{(N-H) \times (N-H)};$$

$$W_H \in R^{N \times H}; \quad W_0 \in R^{N \times (N-H)}.$$

The subscript "0" is a short-hand notation for $N - H$.

By doing some matrix calculations (see Appendix C), we obtain:

$$(A^T A + \lambda^2 L_2^T L_2)^{-1} A^T = W_H F Z_H^{-1} U_H^T + W_0 U_0^T \tag{6.41}$$

where

$$F = diag\left[\frac{(\zeta_1/\gamma_1)^2}{(\zeta_1/\gamma_1)^2+\lambda^2}, \cdots\cdots, \frac{(\zeta_H/\gamma_H)^2}{(\zeta_H/\gamma_H)^2+\lambda^2}\right]$$

so

$$\hat{X} = \left[W_H F Z_H^{-1} U_H^T + W_0 U_0^T\right] b. \tag{6.42}$$

6.3 Application to Quadrupole Splitting Distribution (QSD) Extraction

We now present an application of the Bayesian direct extraction method to a real spectrum that is understood to arise from a quadrupole splitting distribution (QSD).

6.3.1 A real Fe-57 Mössbauer spectrum of an Fe^{3+}-chlorite

The ^{57}Fe spectrum of our example was collected at room temperature ($22°C$) using a thin random-orientation powder absorber of an Fe-bearing natural chlorite that was non-destructively oxidized by heating in air. The spectrum was folded to obtain a flat background and a zero velocity corresponding to the center shift of α-Fe at room temperature. The absorber material is pure chlorite with the so-called modified chlorite structure as seen by X-ray diffraction analysis. It has only Fe^{3+} in both the modified hydroxide layer and the octahedral sites of the T-O-T (or 2:1) layer. No Fe is present in the tetrahedral sites [77].

The same spectrum has also been used to demonstrate the combined partial deconvolution direct extraction method of Ping and Rancourt [57]. The latter method has the relative disadvantages that: 1) ideal fits with Voigt lines are required in a preliminary step as a way of filtering out noise and to help with the deconvolution,

2) errors on the distribution of hyperfine parameters are not obtained, and 3) it cannot easily be generalized to treat multisite problems where several (i.e. more than one) independent distributions of hyperfine parameters occur simultaneously.

6.3.2 The elemental line shape and the extracted QSD

In this case of a QSD, the elemental line shape in equation (6.18) is an elemental doublet consisting of two Lorentzian lines:

$$E(\delta, \Delta; v) = \sum_{k=-1}^{1} L_k(\delta + k\Delta/2, \Gamma, h_k; v) \tag{6.43}$$

where

$$L_k(\delta + k\Delta/2, \Gamma, h_k; v) \equiv \frac{h_k\Gamma^2/4}{(v - \delta + k\Delta/2)^2 + \Gamma^2/4} \tag{6.44}$$

is a elemental Lorentzian with height h_k, full width at half maximum (FWHM) Γ, and centered on $\delta + k\Delta/2$. The two lines are separated by Δ which is given by the usual expression:

$$\Delta = \frac{1}{2}e^2qQ(1 + \frac{1}{3}\eta^2)^{1/2} \tag{6.45}$$

The relevant distributed hyperfine parameter is the QS(Δ). $P(\Delta)$ is the distribution function which we want to extract from the spectrum.

In constructing the matrix A, symmetric elemental doublets (having Lorentzian lines of equal heights and areas) were assumed and a δ-Δ coupling of the form $\delta = \delta_0 + \delta_1\Delta$ was used. When $\delta_1 \neq 0$, therefore, there is an associated distribution of center shifts having the same shape as $P(\Delta)$.

Given the spectrum (Figure 6.8), we chose to use $M = 512$, $N = 41$ and velocity and quadrupole splitting ranges of -4 to 4 mm/s and 0 to 4 mm/s, respectively. $P(\Delta)$ distributions were extracted using several different trial sets of δ_0 and δ_1: $\delta_0 = 0.37$ mm/s and $\delta_1 = 0$ as might be guessed from a preliminary inspection of the spectrum and values such as $\delta_0 = 0.355$ mm/s and $\delta_1 = -0.00723$ that are typically obtained when the spectrum is fitted using the VBF method of Rancourt and Ping [55]. In all cases the Lorentzian width was taken to be $\Gamma = 0.222$ mm/s[55]. The Lorentzian height was constrained to the value that gives the correct total spectral area obtained by background subtraction, and the extracted $P(\Delta)$ distributions were equal within error.

The extracted curve for $\delta_0 = 0.356$ mm/s and $\delta_1 = -0.00813$ is shown in Figure 6.9 with the calculated σ-error for each point obtained by the procedure described in Section 6.2. The corresponding recalculated spectrum ($M = EP$) is shown by the solid line in Figure 6.8 and gives a reduced chi squared value of 1.23.

6.4 Extraction of Two Overlapped QSDs Using Gibbs Sampling

As we have mentioned in 6.1.1, Mössbauer Spectroscopy is a very useful technique in materials science and mineralogy. In this section we present an example of extracting QSDs from the spectrum of an oxibiotite sample using both Bayesian inference theory and the Gibbs sampling technique.

Biotite is a kind of layer silicate. Layer silicates are present in virtually all kinds of rocks and are also the main minerals in clays and soils. Microscopic crystal

chemical information of layer silicates can make significant contributions to petrology, which is the study of rocks and their formation and transformation. Also our existence on the planet and much of the world's economy is based on soils and their properties. These properties emanate from key reactions such as intercalation, ion exchange, dehydroxylation, oxidation, etc.. To have a good understanding of these properties, we need to have the microscopic information of the materials. It is well understood that the Mössbauer spectrum from biotite arises from electric quadrupole splitting distributions. Therefore the success of the QSDs analysis will make a big contribution to the study of these materials.

6.4.1 The problem

The spectrum, which we are going to study, is from an oxibiotite (sample MOC 2661 which was supplied by the Mineral Sciences Division of the Canadian Museum of Nature). It was originally from the Silver Crater Mine, near Bancroft, Ontario. At this locality, biotite occurs as large crystals in a calcite-rich band associated with nepheline syenite [75]. The study of the oxidation reaction of the same sample was presented by Rancourt et al.[76] in 1993. All the details about the sample, such as the crystal structure, structural formula etc., can be found in that article. In this study we only concentrate on how the spectrum is related to the QSDs and on how to extract the QSDs from the spectrum.

It has been established[76] that in the virgin sample 10% of the iron is Fe^{3+} and the rest is Fe^{2+}, assuming equal Fe^{2+} and Fe^{3+} Mössbauer recoilless fraction at room temperature(RT). When the sample is heated in air, an oxidation reaction occurs. This means that some of the Fe^{2+} cations in the virgin sample will be oxidized to Fe^{3+} cations. As the sample is heated to higher temperatures, the site population

of Fe^{3+} increases.

The spectrum shown in Figure 6.10 was collected at room temperature after the sample was heated in air for 24 hours at 689 K. By looking at the spectrum, we immediately know that there are two overlapping doublets in the spectrum. The reason for this is that there are now both Fe^{2+} and Fe^{3+} sites in the sample. Fe^{2+} and Fe^{3+} cations have different valence states which make their quadruple splittings and center shifts very differently from one to the other.

As we have shown in 6.1.3, for those sites which have the same valence state cations but have different local environments, there is a quadrupole splitting distribution. Here we call a group of sites with the same valence state a generalized site. From each generalized site there is a contribution, called the subspectrum, to the total spectrum. Each subspectrum has its own QSD and corresponding set of other hyperfine parameters. The mathematical model can be written as:

$$\begin{aligned} M(v) &= \int E_{2+}(\delta_{2+}, \Delta_{2+}; v) P_{2+}(\Delta_{2+}) d\Delta_{2+} \\ &\quad + \int E_{3+}(\delta_{3+}, \Delta_{3+}; v) P_{3+}(\Delta_{3+}) d\Delta_{3+} + \mathcal{N}(v) \end{aligned} \tag{6.46}$$

where $E_{2+}(\delta_{2+}, \Delta_{2+}; v)$ and $P_{2+}(\Delta_{2+})$ are the elemental doublet and the QSD for the generalized site Fe^{2+}; $E_{3+}(\delta_{3+}, \Delta_{3+}; v)$ and $P_{3+}(\Delta_{3+})$ are the elemental doublet and QSD for the generalized site Fe^{3+}; $\mathcal{N}(v)$ is the noise. From the spectrum we know that the two subspectra are overlapped. This is because the corresponding two QSDs are overlapped. Extracting the two overlapped QSDs directly from the spectrum is not a trivial task. It has never been done before by a direct extraction method. In the following we will see that the Bayesian inference theory together with the Gibbs sampling technique can give us very encouraging results.

6.4.2 Numerical procedures

Starting with equation (6.46), we do the same discretization as we did in 6.2.1. We obtain the matrix formulation of the model:

$$[M]_{M\times 1} = [E_{2+}]_{M\times N}[P_{2+}]_{N\times 1} + [E_{3+}]_{M\times L}[P_{3+}]_{L\times 1} + [\mathcal{N}]_{M\times 1} \tag{6.47}$$

It is not necessary that $N = L$. But for the convenience of the calculation, we let $N = L$. In simplified notation we have:

$$M = E_{2+}P_{2+} + E_{3+}P_{3+} + \mathcal{N} \tag{6.48}$$

Now in equation (6.48) there are two unknown QSDs, P_{2+} and P_{3+}. Direct application of Bayesian inference theory cannot extract both distributions. However by using Gibbs sampling with the Bayesian inference theory we can easily solve the problem. Suppose from our experience we can give an initial guess, say P^0_{3+}, about P_{3+} before we analyze the data. If we let

$$M^0_{2+} = M - E_{3+}P^0_{3+} \tag{6.49}$$

then we have:

$$M^0_{2+} = E_{2+}P_{2+} + \mathcal{N} \tag{6.50}$$

This equation is exactly the same as equation (6.19). We can follow the same numerical procedure as in 6.2.1 to extract a P^1_{2+} from the posterior probability distribution of P_{2+}, $P(P_{2+} \mid M, P^0_{3+})$, with the condition that on $P_{3+} = P^0_{3+}$. Next we substitute P_{2+} with P^1_{2+} in equation (6.48), then let

$$M^1_{3+} = M - E_{2+}P^0_{2+} \tag{6.51}$$

We obtain:

$$M^1_{3+} = E_{3+}P_{3+} + \mathcal{N} \tag{6.52}$$

Similarly a P_{3+}^1 can be extracted from the posterior probability distribution of P_{3+}, $P(P_{3+} \mid M, P_{2+}^1)$, (conditionally on) $P_{2+} = P_{2+}^1$. After one iteration we have (P_{2+}^1, P_{3+}^1). After K iterations we obtain (P_{2+}^K, P_{3+}^K). For a large enough K, P_{2+}^K can be considered as a random vector from the marginal posterior probability distribution of P_{2+}, $P(P_{2+} \mid M)$, and P_{3+}^K can be considered as a random vector from the marginal posterior probability distribution of P_{3+}, $P(P_{3+} \mid M)$. If we repeat this M_s times, we obtain two random vector samples $(P_{2+}^{K,1}, P_{2+}^{K,2}, \ldots, P_{2+}^{K,M_s})$ and $(P_{3+}^{K,1}, P_{3+}^{K,2}, \ldots, P_{3+}^{K,M_s})$, which are from $P(P_{2+} \mid M)$ and $P(P_{3+} \mid M)$ respectively. Consequently all information about P_{2+} and P_{3+}, such as the best estimates of the QSDs, the variances of the estimates, etc., can be found from these samples.

6.4.3 Applications to the spectrum of the oxybiotite sample

We now show the application of the above algorithm to the spectrum collected at RT after heating the sample MOC 2661 in air for 24 hours at 689°C. In constructing the elemental line shapes E_{2+} and E_{3+}, symmetric elemental doublets (having Lorentzian lines of equal heights and areas) were assumed and $\delta - \Delta$ couplings of the form $\delta = \delta_0 + \delta_1 \Delta$ were used. Given the spectrum, we chose to use $M = 512$, $N = 41$ and velocity ranges from -4 mm/s to 4 mm/s. The quadrupole splittings range from 0 mm/s to 4 mm/s. $P_{2+}(\Delta_{2+})$ and $P_{3+}(\Delta_{3+})$ were extracted using a set of $(\delta_{0,2+}, \delta_{1,2+}, \delta_{0,3+}, \delta_{1,3+})$ which are from the Voigt-based fitting method. On the Gibbs sampling stage, different K values, from 2 to 10, were tested. We found that for K larger than six, there is no significant improvement in the extracted QSDs. The sample sizes are $M_s = 1000$. The two extracted QSDs are shown in Figure 6.11 and 6.12 and the corresponding recalculated spectrum (solid line) is shown in Figure 6.10. We can see that there is an overlapping area between $P_{2+}(\Delta_{2+})$ and

$P_{3+}(\Delta_{3+})$ and there are negative values for the QSDs in that area. This is because the numerical method has the freedom to allow tradeoff between the two QSDs in that area. This is inevitable. The things we could do here to prevent this are: to use more realistic models for the elemental line shapes, the coupling relationship between δ and Δ; and to impose more prior information, such as positive QSDs. This study is a preliminary study to see whether the Bayesian inference theory and the Gibbs sampling technique can tackle this problem which has never been solved before. Obviously it gives us very encouraging results. More detailed studies will be conducted in the future.

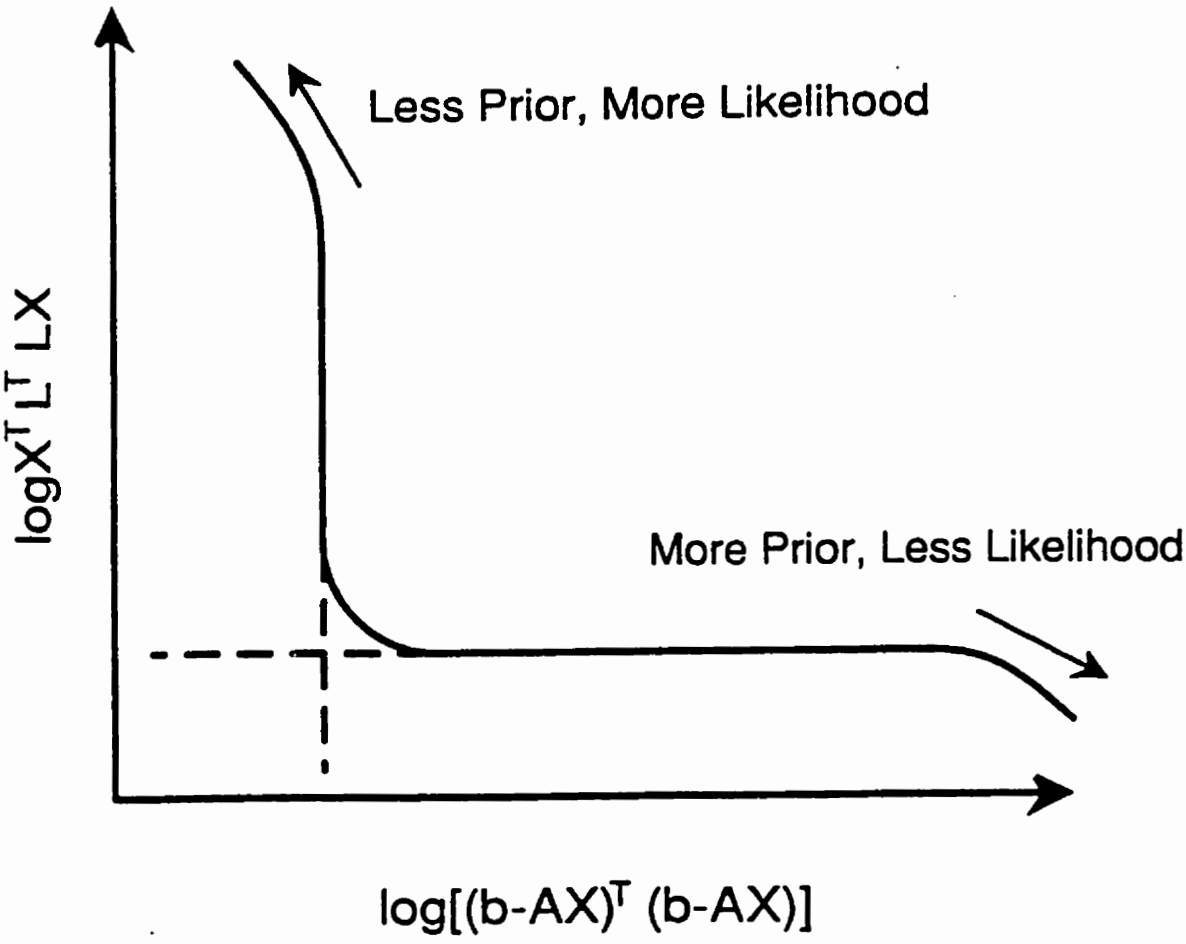

Figure 6.7: The generic form of the L-curve

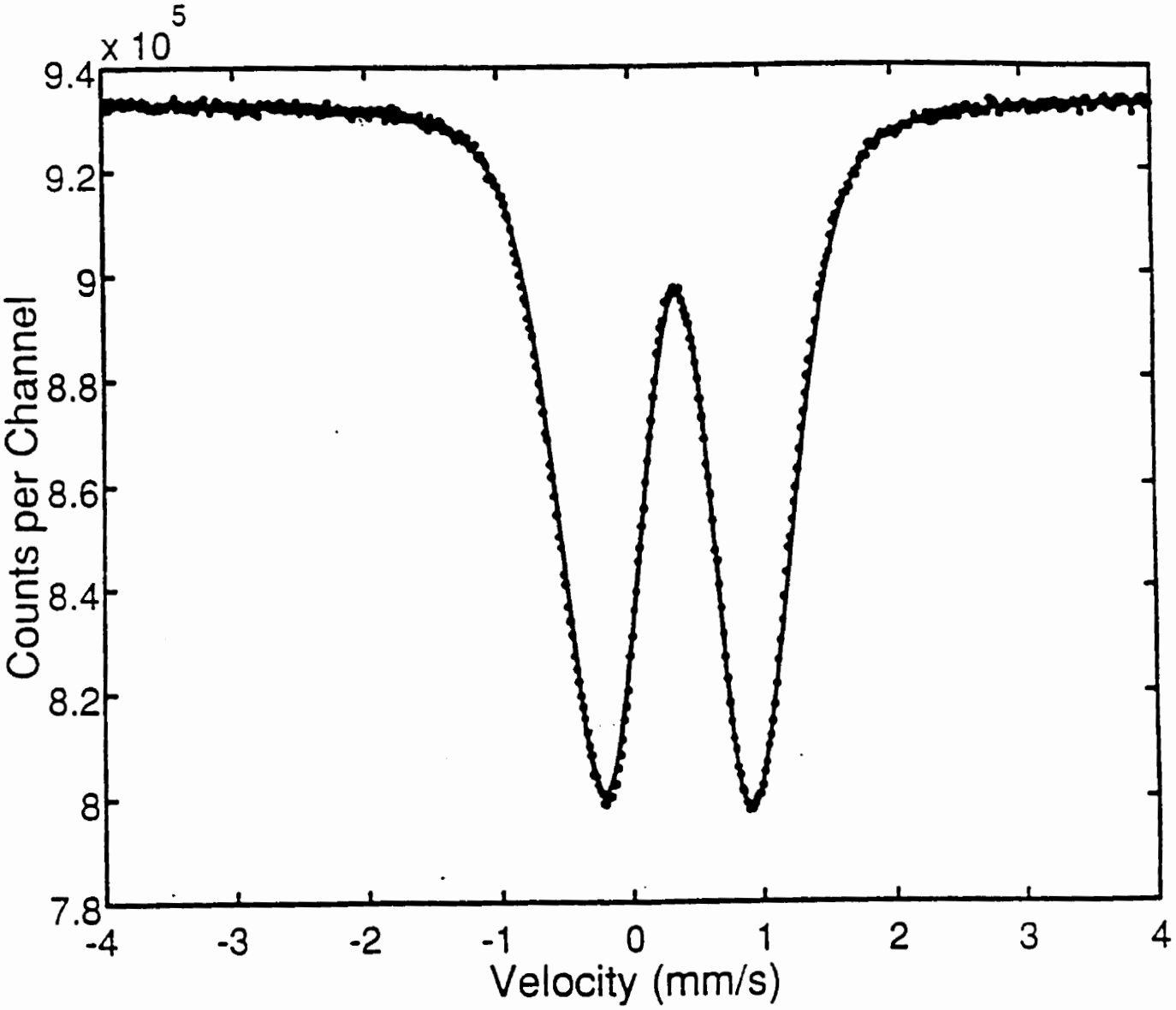

Figure 6.8: The room temperature Mössbauer spectrum (dots) of the Fe^{3+}-chlorite and the recalculated spectrum (solid line).

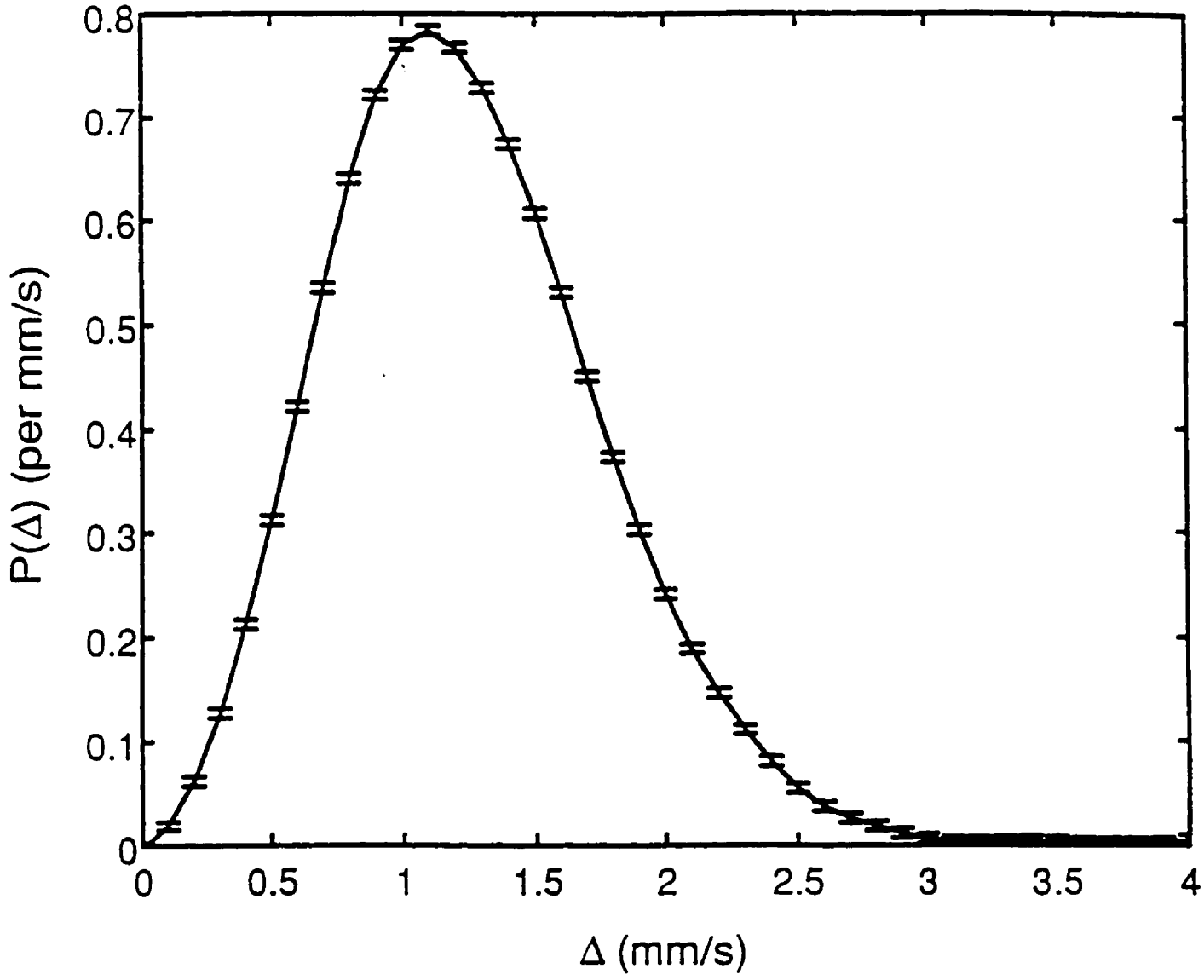

Figure 6.9: The extracted quadrupole splitting distribution of the Fe3+-chlorite with the standard deviation σ.

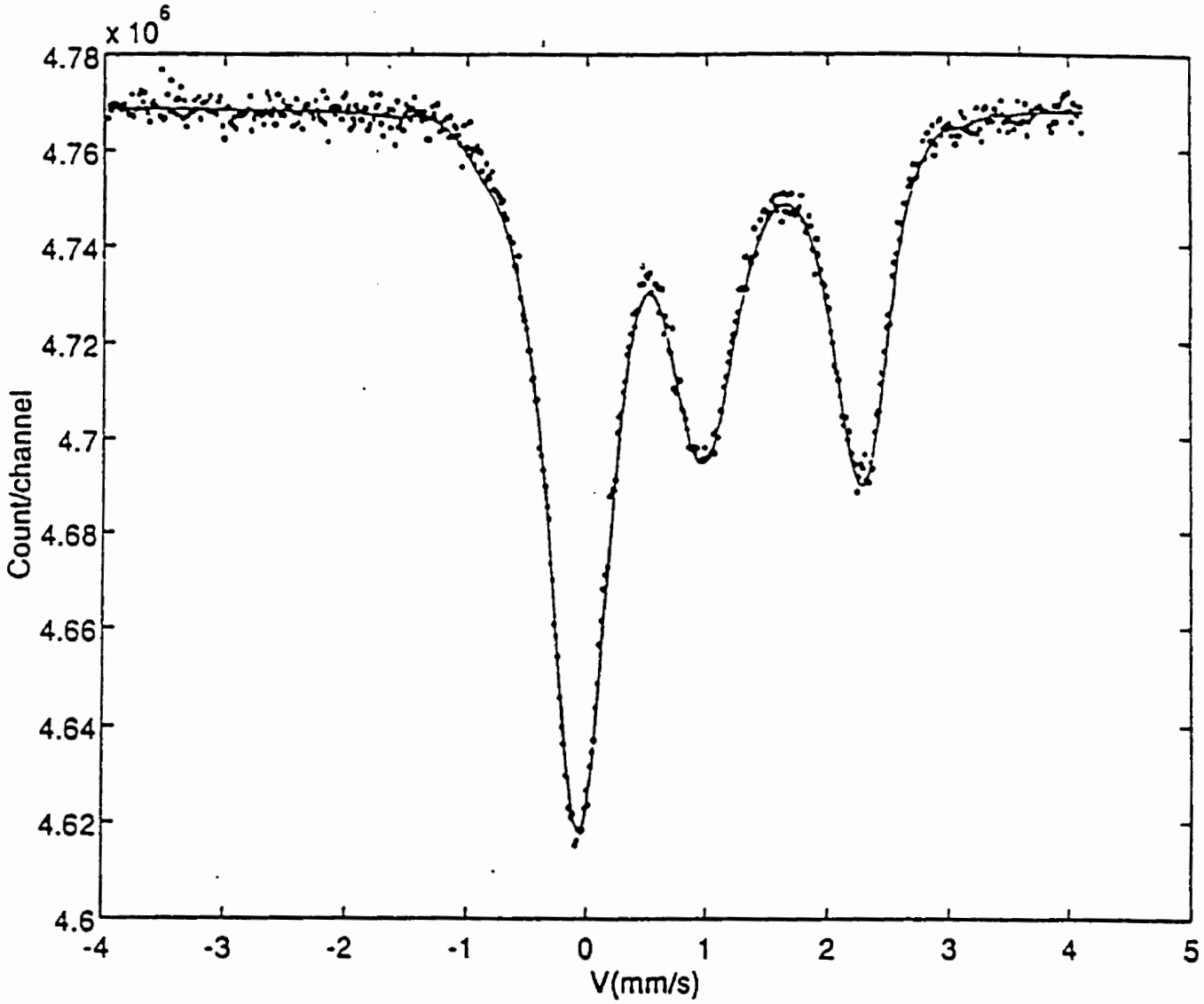

Figure 6.10: The room temperature spectrum (dots) and fitted spectrum of the oxibiotite heated in air for 24 hours at 689 K .

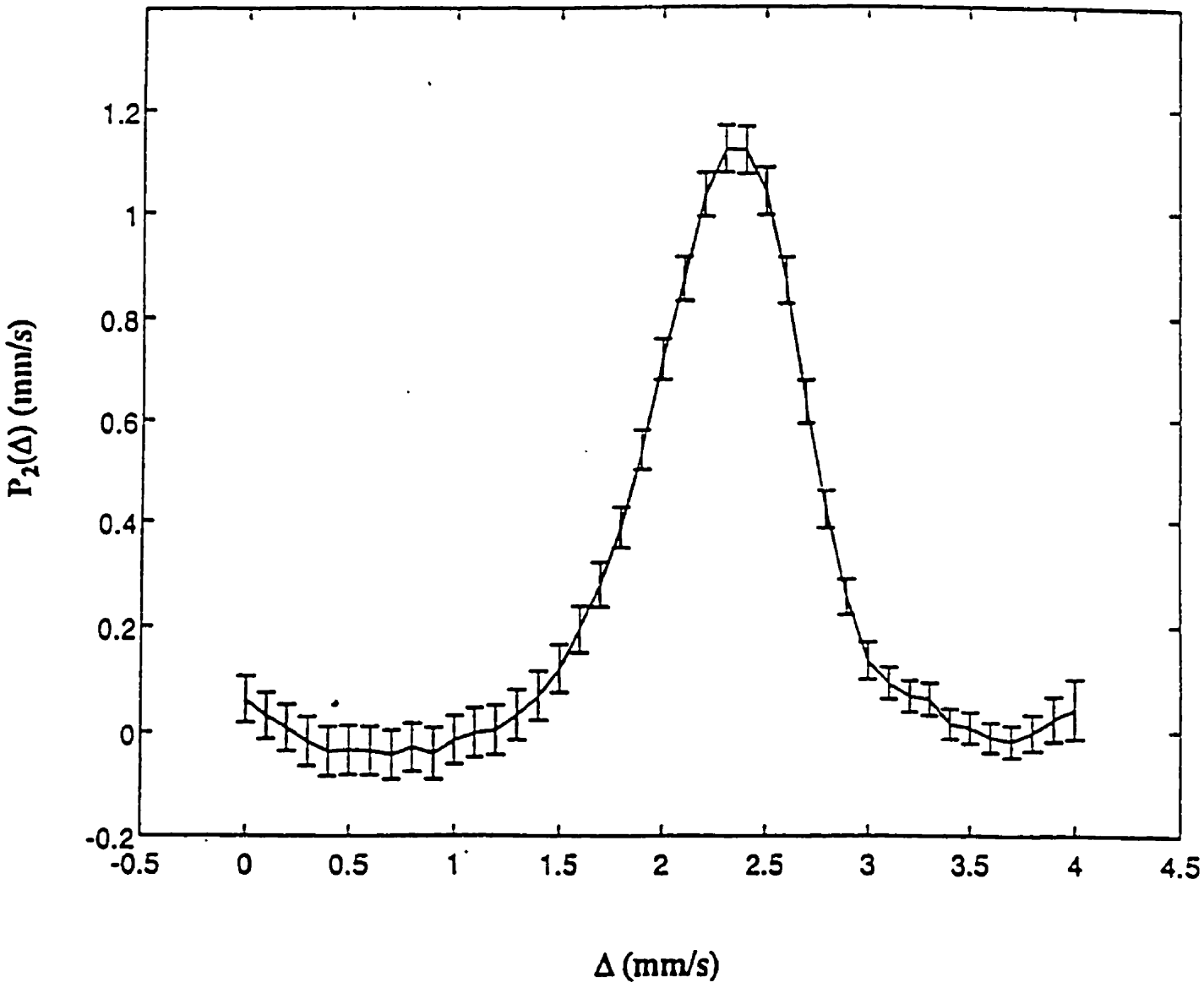

Figure 6.11: The extracted quatrupole splitting distribution of the Fe^{2+} site with the standard deviation σ.

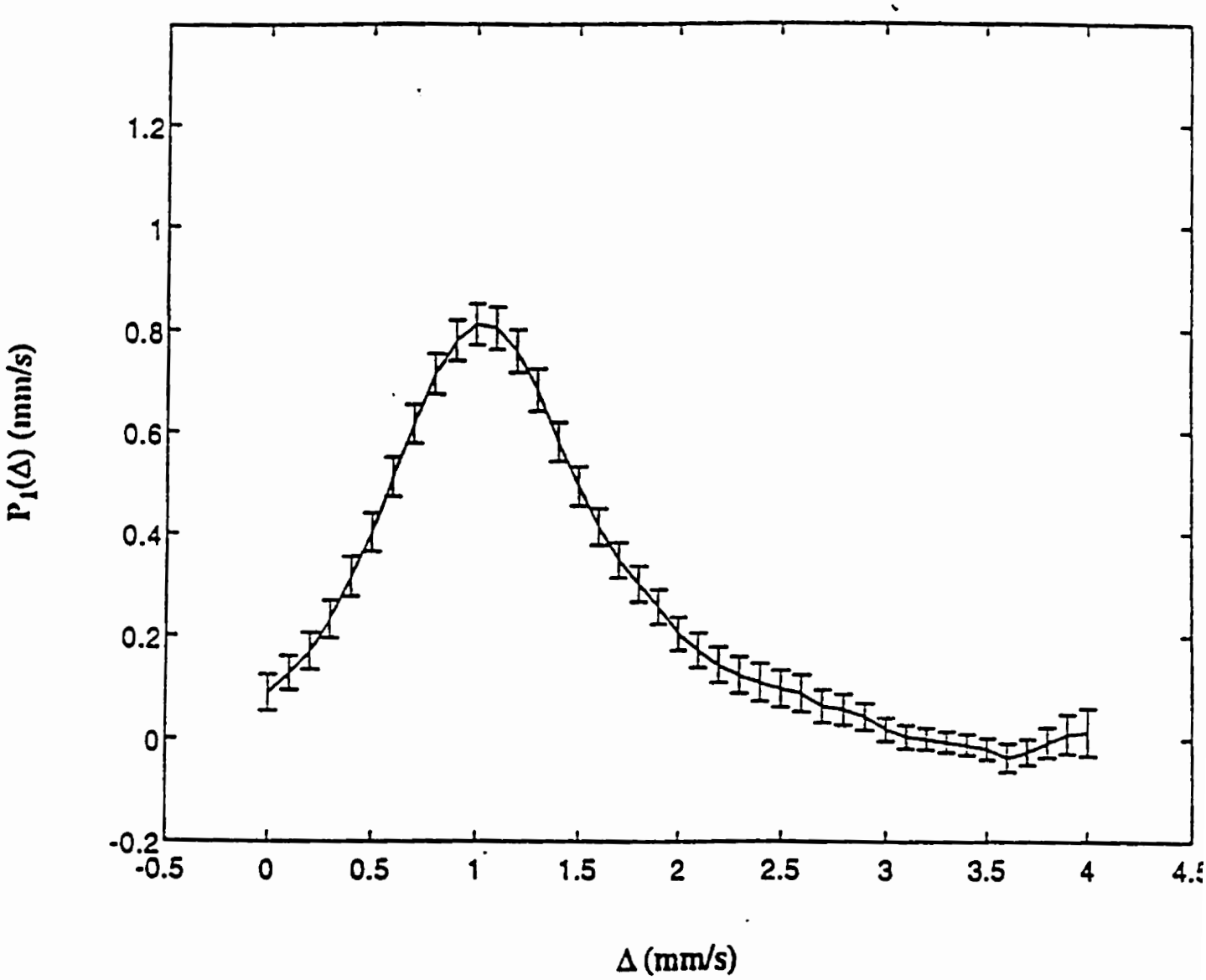

Figure 6.12: The extracted quadrupole splitting distribution of the Fe^{3+} site with the standard deviation σ .

6.5 Extraction of the Hyperfine Field Distributions

6.5.1 The problem

Hyperfine field distributions are probability density distributions, $P(H)$, of magnitudes of hyperfine fields. They occur in the Fe-57 Mössbauer spectra of magnetically ordered materials that have atomic and/or crystallographic disorder. In the absence of dynamic line shape effects and artifacts such as thickness effects, the corresponding Mössbauer spectrum can be written as

$$M(v) = \int E(H, \delta, \epsilon; v) P(H) + \mathcal{N}(v), \tag{6.53}$$

where $E(H; v)$ is the elemental line shape for one site in the absorber. It is a function of the hyperfine field, H, and other hyperfine parameters for that site (see below) or family of sites having the same H. $P(H)$ is the distribution function for the hyperfine field (or HFD). $\mathcal{N}(v)$ represents the experimental noise. In this case we want to extract the hyperfine field distribution $P(H)$ from the corresponding spectrum.

In the absorber, all three interactions, the center shift, the electric quadrupole interaction and the magnetic field interaction, are present. The elemental line shape $E(H, \delta, \epsilon; v)$ is much more complicated than the situations where either only CS and QS interactions are present or only CS and HF interaction are present. The choice of different elemental line shapes depends on how much information we have about the sample before we analyze the spectrum. Sometimes different models have to be tested to see which one can give the reasonable results. In the next section we will give a detailed introduction about the elemental line shapes in the situation wher all three interaction are present.

6.5.2 The elemental lineshapes

A. Perturbation sextet model

It is well known that when the electric field gradient (EFG) parameter $K = e^2qQ/2$ is assumed to be much smaller than the excited state Zeeman splitting $z = \mid g_{3/2}\mu_N H \mid$ ($\mid e^2qQ/2g_{3/2}\mu_N H < 0.8 \mid$ is a usable limit [118]), the first order EFG perturbation on the Zeeman split excited state level results in a shift

$$\epsilon = (e^2qQ/8)[2 - (3 - \eta cos2\phi)sin^2\theta] \tag{6.54}$$

where θ and ϕ are the polar coordinates of hyperfine field in the usual EFG coordinate system. The line positions of the six allowed transitions which make up the sextet (lines 1 to 6 in order of increasing energy) turn out to be:

$$\begin{aligned}
\omega_1 &= \delta - (Z+3)z/2 + \epsilon, \\
\omega_2 &= \delta - (Z+1)z/2 - \epsilon, \\
\omega_3 &= \delta - (Z-1)z/2 - \epsilon, \\
\omega_4 &= \delta + (Z-1)z/2 - \epsilon, \\
\omega_5 &= \delta + (Z+1)z/2 - \epsilon, \\
\omega_6 &= \delta + (Z+3)z/2 + \epsilon,
\end{aligned} \tag{6.55}$$

where Z is the ratio of nuclear g-factors $Z = g_{1/2}/g_{3/2} = 1.7509$ and δ is the center shift. Then the elemental sextet can be written as:

$$E(z, \delta, \epsilon; v) = \sum_{k=1}^{6} L_k(\omega_k, \Gamma, h_k; v), \tag{6.56}$$

where

$$L_k(\omega_k, \gamma, h_k; v) \equiv \frac{h_k\Gamma^2/4}{(v - \omega_k)^2 + \Gamma^2/4} \tag{6.57}$$

is a Lorentzian with height h_k, full width at half maximum (FWHM) Γ, and centered on ω_k.

B. Blaes' exact line shape model

In Blaes' exact line shape model [119], a super operator formalism is used to yield a closed form expression for the Mössbauer line shape in the presence of magnetic dipole and electric quadrupole hyperfine interactions for fixed orientations between the EFG principle axes, the magnetic field, and the γ-ray direction. Then besides the hyperfine field H, the center shift δ, and the EFG parameter $K = e^2qQ/2$, the lineshape is also a function of the polar and azimuthal angles of the γ-ray direction in the laboratory system and a function of the Euler angles α, β, and γ which transform the principle axes system x' y' z' of the EFG tensor into the laboratory system x y z where the magnetic field is parallel to z. (See Figure 6.13). The details of the calculations and the general expression for the lineshape can be found in Ref. [119]. Although the resulting expressions appear complicated, they allow some angular averages to be performed analytically.

A great degree of test and comparison with other methods of calculating the lineshape, such as the first order perturbation and solving explicitly the hyperfine Hamiltonian under certain conditions, has been made. Perfect agreements were obtained. This was to make sure that the expressions in Blaes' paper are correctly presented and that our further calculations for some special conditions, such as the angular averages, were correctly conducted. Such applications of the Blaes line shape expressions are given in the following subsections (C-E).

C. The exact lineshape for powder samples with symmetric EFG ($\eta = 0$)

This is the case where the orientation of the EFG is fixed and the γ-ray direction is randomly oriented with respect to the magnetic field. Thus, only the following angular average has to be performed:

$$E_1(H, \delta, Q, \beta; v) = \int_0^{2\pi} \int_0^{\pi} E(H, \delta, K, \eta, \alpha, \beta, \gamma, \phi, \theta; v) sin(\theta) d\theta d\phi \qquad (6.58)$$

which can be calculated analytically. This eliminates γ from the line shape expression E_1 [119]. Also, because of the rotational symmetry of the EFG ($\eta = 0$) with respect to its principle component, the nuclear Hamiltonian for the electric quadrupole interaction is independent of the angle α. Then the line shape E_1 only depends on the hyperfine parameters $\{H, \delta, K\}$ and the angle β representing the orientation between the magnetic field and the principal axis of the EFG.

D. The lineshape for powder samples with both EFG ($\eta = 0$) and magnetic field H independently and randomly oriented

This is the case when both the EFG with rotational symmetry and the magnetic field randomly orient with respect to the γ-ray direction. Based on the line shape equation for the case described in C, an additional average on the angle β

$$E_2(H, \delta, K; v) = \int_0^{\pi} E_1(H, \delta, K, \theta; v) sin(\beta) d\beta \qquad (6.59)$$

is needed. Once again this can be performed analytically and results in a closed expression for the line shape E_2 which is simply a function of the hyperfine parameters $\{H, \delta, K\}$.

E. Lineshapes with distributed magnitudes of EFG

We have considered two cases, as follows.

Firstly, when the magnitude of the EFG is not a single value but has an independent probability distribution, equation (6.53) should be written as

$$M(v) = \int\int E(H, \delta, K; v)P(H)P(K)dHdK + \mathcal{N}(v). \tag{6.60}$$

To extract the distribution $P(H)$ based on equation (6.60), we have to give a reasonable guess for $P(K)$. Then, instead of imposing a single value for K into equation (6.53), a known function for $P(K)$ is substituted into equation (6.60) and equation (6.60) is simplified as

$$M(v) = BG - S(v) = \int E_3(H, \delta; v)P(H)dH + \mathcal{N}(v), \tag{6.61}$$

where

$$E_3(H, \delta; v) = \int E(H, \delta, K; v)P(K)dK, \tag{6.62}$$

is the corresponding elemental line shape.

Secondly, we assume that K is linearly coupled with H. That is $K = k_0 + k_1 H$ with known $\{k_0, k_1\}$. Imposing $K = k_0 + k_1 H$ into the lineshape in equation (6.53), the corresponding line shape,

$$E_4(H, \delta; v) = E(H, \delta, K = k_0 + k_1 H; v) \tag{6.63}$$

is just a function of $\{H, \delta\}$.

6.5.3 Applications to simulated spectra

In real spectrum analysis, due to a lack of information about the sample, choosing a correct line shape model is not an easy task. Different models have to be used to see

which one can give more reasonable results. Having good results also depends on the reliability of the numerical method used to analyze the spectrum. Before we go to analyze the real spectra, applications of the Bayesian inference method to simulated spectra are conducted using the Blaes' line shape model. These applications will tell us whether the Bayesian inference method can give the reasonable results when the line shape model is exactly known.

In the following simulation, the line shape model described in subsection D is used, i.e. both the EFG with rotational symmetry and the magnetic field randomly orient independently with respect to the γ-ray direction. The line shape is a function of $(H, \delta, Q; v)$. Here, the assumed parameters for both the spectrum simulation and the extraction are zero center shift and $e^2qQ = 1.2$ mm/s. The only distributed parameter is the hyperfine field. The known Gaussian hyperfine distributions used to simulate the spectra, with simulated Gaussian noise, are full or truncated (make $H \geq 0$). The standard deviation of the Gaussian distributions is $|g_{3/2}|\mu_N\sigma_H$ =0.2 mm/s and the maximum positions of $|g_{3/2}|\mu_N H_{max}$ =2, 1, 0.5, and 0.2 mm/s. In constructing the matrix A, we choose to use $M = 512$, $N = 41$ and the velocity ranges from -8 mm/s to 8 mm/s. As we can see from Figure 6.14 to 6.17 in each case the Bayesian direct extraction method exactly recovers the correct hyperfine distribution and also gives the uncertainty of the distribution which aries from the spectral noise.

The above demonstration clearly shows us that when we have the right model, the Bayesian method has excellent reliability and can give us the uncertainty about the extraction which is related to the noise level in the spectra. Therefore we can say that in the real spectrum analysis when an unreasonable result comes out, it means the model used is not correct. In this case the Bayesian method sends us the

message that the model needs to be refined.

6.6 Ground state hyperfine field distribution of $Fe_{65}Ni_{35}$ Invar at 4.2 K

The purpose of this study is to use the Blaes' exact elemental lineshape model with BIF-DE method to extract the ground state hyperfine field distribution of $Fe_{65}Ni_{35}$ Invar. By testing different orientations of EFG and HF with respect to the incident *gamma*-ray direction, we try to find out that how the HFD is affected by those different assumptions, especially at the low field area. The results of this extensive testing will be compared with the results from the Monte Carlo simulations[127]-[128], a method which calculates HFDs using the spin structure and a phenomenological model relating thermal average local moments to resulting site-specific hyperfine fields. In the hope this comparison will help us understand the microscopic structure of Invar. Because this comparison involves detailed introduction of the magnetic structure model and the Monte Carlo simulation which are out of the scope of this thesis, it will be present somewhere else[129].

6.6.1 Introduction

In the physics literature related to fcc Fe-Ni alloys, the emphasis has clearly been on Invar where low temperature behavior has been given special attention in the hope that it might elucidate the higher temperature Invar (i.e. magneto-volume) behavior. Many low temperature anomalies have been reported and are reviewed in Ref. [79]. Many of these are non-intrinsic and arise from precipitates of either body centered cubic (BCC) phase (α-phase) or low-moment (LM) face centered cubic (FCC) phase (γ_{LM}-phase).

Invar itself is a high-moment (HM) FCC phase (γ_{HM}-phase) that is structurally metastable and is obtained by quenching from the melt. The true equilibrium structure at 65 at. % Fe is not the quenched state and its exact nature is still being elaborated [80]. The true structural ground state may be an epitaxial intergrowth of Fe-rich γ_{LM}-phase and chemically ordered FeNi phase (tetrataenite) [81]. The present paper is concerned with the magnetic ground state of Invar in its quenched (or metastable) structural state.

It is important to realize, however, that the quenched structural state of Invar is that of a random binary alloy on an FCC lattice only to a first approximation and that it has not been fully characterized. Many structural studies report partial Fe and Ni segregation on a length scale of $\sim$ 15 Å[82]. Such non-randomness must affect the magnetic properties and the ground state spin structure.

Different models for the magnetism of Invar have been proposed, notably: the two-γ-state model [83]-[85], weak itinerant ferromagnetism [86], and latent antiferromagnetism [87, 88]. Recently, also, it has been suggested that the theoretically predicted occurrence of LM phases in metallic Fe [89]-[99], chemically ordered Fe_3Ni [99]-[101], and Fe-rich random FCC Fe-Ni alloys [102, 103] is related to Invar behavior. In all cases, the magnetism is believed to cause Invar behavior via some form of magneto-volume or magneto-structural coupling.

Recently, it has been shown [104, 105] that a simple local moment model with three composition-independent pair-specific nearest neighbor (NN) magnetic exchange parameters (J_{FeFe}, J_{FeNi}, and J_{NiNi}) reproduces all of the main purely magnetic properties of the quenched FCC Fe-Ni alloys: the Curie point versus composition, the spontaneous saturation moment versus composition, the spontaneous

magnetization versus temperature at each composition, and the high field (paraprocess) susceptibility at $T = 0\ K$ versus composition.

Both cluster-method mean field theory (MFT) calculations [104, 106] and Ising approximation Monte Carlo (MC) simulations [105] were used. The MC simulations give superior agreement with experiment except for the paraprocess susceptibility where the Ising approximation of the MC simulations gives artificial steps, whereas the Heisenberg Hamiltonian of the MFT calculations gives realistic behavior with gradual spin rotation. The anomalies in all the magnetic properties are seen to arise from an antiferromagnetic NN Fe-Fe coupling ($J_{FeFe} < 0$) in these otherwise ferromagnetic γ_{HM}-phase alloys. Magnetic frustration on the FCC lattice occurs at the Fe-rich end. These calculations strongly support the earlier latent antiferromagnetism ideas.

It follows, based on the above agreement with the macroscopic magnetic properties, that Invar is well approximated as a local moment system with large stable moments (0.6 μ_B and 2.8 μ_B on Ni and Fe atoms, respectively) and NN-only Ising exchange interactions. In this study, we test the degree to which this approximation is valid by a detailed comparison with a microscopic property: the ground state ^{57}Fe hyperfine field distribution (HFD).

In order to accomplish this, we implement: (1) absorber thickness corrections of the Mössbauer spectrum from which the HFD is extracted, (2) our HFD extraction method based on Bayesian inference theory that gives error propagation down to the measured HFD and that is not limited to any particular elemental line shape model, and (3) a simple vector model[127]-[128], of the relationship between the spin structure and the HFD, that has been extensively tested at all compositions and that allows a simulated HFD to be uniquely generated from an equilibrium MC

spin structure.

6.6.2 Measurement and data treatment

Preparation and characterization of the splat quenched $Fe_{65}Ni_{35}$ sample is described elsewhere [107, 108]. A 55 μm thick splat quenched flake that showed no grain or magnetization orientation texture was used for the present investigation. The RT ^{57}Fe Mössbauer spectrum showed no trace of either α-phase or γ_{LM}-phase and the measured RT saturation magnetic moment, first-heating Curie point (measured on a separate piece), and RT lattice parameter were consistent with the known values for $Fe_{65}Ni_{35}$.

The main ^{57}Fe transmission Mössbauer spectrum was collected at $T = 4.2\ K$ in an exchange gas type liquid helium cryostat using a $\sim$ 20 mCi ^{57}Co rhodium-matrix single-line thin source on a velocity range of ± 7.5 mm/s with a constant acceleration drive. Data was acquired on 1024 channels and folded to give a flat background (BG) and a zero velocity position corresponding to the center shift (CS) of metallic α-Fe at RT in the 512-channel calibrated spectrum.

Absorber thickness effect corrections were performed by a new method [109, 110] that involves an analytic deconvolution of the transmission integral. The intrinsic absorber resonant cross section is extracted and used to reconstruct a thin-limit spectrum that is devoid of any thickness effects. An absorber recoilless fraction at $T = 4.2\ K$ of 0.9 was assumed.

6.6.3 Spectral analysis: obtaining the measured HFD

A. Overview of the problem

$Fe_{65}Ni_{35}$ Invar has been extensively studied by ^{57}Fe Mössbauer spectroscopy [108]. A recently reported low temperature anomaly [111] was not seen in other studies and was later found to be an experimental artifact arising from a cryogenic problem (personal communication with J. Hesse). All studies, therefore, show a behavior that saturates at low temperatures such that the $T = 4.2\ K$ spectrum is expected to be identical within error to the $T = 0\ K$ spectrum, if the latter could be measured.

It has been proposed that even the low temperature spectra could be interpreted in terms of dynamic effects rather than HFDs [82],[112]- [114]. Whereas such dynamic line shape effects do occur in broad neighborhoods below and above the Curie point, they only start to affect the low temperature spectral line shape above $T \approx 200\ K$ [108]. It is therefore generally accepted that the low temperature spectra can be analyzed and interpreted in terms of distributions of static hyperfine field magnitudes (i.e. HFDs). This is strongly supported by the present work where the extracted HFD, obtained by assuming a static HFD interpretation of the measured spectrum, is understood in terms of a simple physical model.

Although such a HFD analysis is conceptually simple, in practice it is very difficult for both mathematical and physical reasons. Several HFD analysis methods have been devised and are reviewed in Ref. [115]. In its domain of applicability, the most powerful of these is the Voigt-based fitting (VB-F) method [115]. However, it has two main disadvantages: (1) the usual perturbation sextet must be used as the elemental subspectrum, and (2) as in all previous methods, error analysis of the resulting HFD is not performed. The Bayesian inference theory direct extraction

(BIT-DE) method [116] that we apply here to HFDs for the first time solves both these problems and allows several different elemental line shape models to be tested, including exact line shapes, angle averaging, independent or coupled distributions of other hyperfine parameters, etc.

B. Application to the spectrum of $Fe_{65}Ni_{35}$ invar at 4.2 K

We now present the results from our study of the thickness corrected spectrum of $Fe_{65}Ni_{35}$ Invar at 4.2 K using the BIT-DE method. Based on equation (6.53), i.e. interpreting the spectrum in terms of distributions of static hyperfine field magnitudes, different elemental line shape models, such as those described in C-E, were tested.

In our study we began with the perturbation sextet model for the elemental line shape and using the VB-F method [115] to extract the HFD. By doing this we can have some initial ideas about the distribution and the parameters involved in the line shape (see below).

Figure 6.18 shows the fitted spectrum and corresponding HFD from the fitting with three Gaussian components for the HFD and allowing linear coupling to the HFD for both δ and ϵ ($\delta = \delta_0 + \delta_1 H$, $\epsilon = \epsilon_0 + \epsilon_1 H$). Since we used thickness corrected data, we let the sextet height ratios $\{h_1/h_3, h_2/h_3\}$ and FWHM Γ be fixed at their known theoretical values. The resulting fit parameters are given in Table 6.1.

From this initial study we can see that the extracted HFD has a maximum around 2.4 mm/s and a long tail in the low field area from 0.0 mm/s to 1.8 mm/s. This tells us that there may be some features in the low field area in the true HFD. Since VB-F is based on the assumption that the HFD is represented by sums of Gaussians with different positions, the details of the true HFD may not be extracted due to

Table 6.1: Fit parameters for $Fe_{65}Ni_{35}$ using VB-F with three Gaussian components for the HFD

Parameters	(units)	Values	Parameters	(units)	Values
$h_{3,1}$	(counts)	6620.7	Γ	(mm/s)	0.194
σ_{z1}	(mm/s)	0.46891	δ_0	(mm/s)	0.04716
z_{01}	(mm/s)	1.43485	δ_1		-0.00969
$h_{3,2}$	(counts)	6267.4	ϵ_0	(mm/s)	0.40516
σ_{z2}	(mm/s)	0.18409	ϵ_1		-0.17221
z_{02}	(mm/s)	2.17670	h_1/h_3		3
$h_{3,3}$	(counts)	29551.0	h_2/h_3		2
σ_{z3}	(mm/s)	0.13018	Reduced χ^2		4.306
z_{03}	(mm/s)	2.35370	$< z >_d$	(mm/s)	2.18422

pre-fixed distribution model limitations. Also, the error analysis of the resulting HFD cannot easily be performed using this method.

The latter means that for small features we do not know whether the features are caused by true structures in the material or by experimental errors. To further study the spectrum, we used BIT-DE with the perturbation sextet elemental line shape and allowed linear coupling for δ and ϵ. Different sets of (δ_0, δ_1) and (ϵ_0, ϵ_1), which can either be guessed from a preliminary inspection of the spectrum or obtained from Table 6.1., were used to extract the HFD.

From this study we found that when $\delta_1 = 0$, i.e. no coupling for δ, there is no significant effect on the HFD, but there is significant change on the HFD in low field when the coupling for ϵ is changed. This can be seen from Figure 6.19. Also from Figure 6.19 the error bars give us confidence that the features in the low field are not due to experimental error but to true structure in the material. However, at this stage we are still not sure that the true HFD in the low field is correctly

represented by the extracted HFD because the perturbation sextet model is a good approximation to the true elemental line shape only when $e^2qQ/2$ is much smaller than $g^*\mu_N H$.

To correctly extract the HFD in the low field range where $e^2qQ/2$ is no longer much smaller than $g^*\mu_N H$, exact line shape models must be employed. Here Blaes' exact line shape models described in C to E for different situations were used.

Figure 6.20 shows the extracted HFD using the line shape described in C with single value for the magnitude of the EFG and with angle $\beta = 0$. We also conducted one dimensional mapping for β from 0 to 90° (E_1 is a function of $cos^2\beta$). It turned out that there is no significant effect on the extracted HFD when β changes.

Next the elemental line shape with angle β averaged and single values for both K and δ was used. A two dimensional search of K and δ for the best chi-squared was conducted as well. Significant changes in the HFD in the low field range showed up with changing values of K. The extracted HFD with the best chi-squared is shown in Figure 6.21.

To further see how the EFG affects the extraction of the HFD, an independent Gaussian distribution and a linearly coupled distribution to H for K were imposed respectively. In both cases the angle β was averaged. The corresponding extracted HFDs are shown in Figure 6.22 and Figure 6.23.

By comparing the extracted HFDs in Figure 6.18 to Figure 6.23 we see that they are almost identical from $z = 3\ mm/s$ down to $z \cong 0.8\ mm/s$. There is almost no difference among the four extracted HFDs without imposing the coupled distribution for K. These clearly show us the robustness of the BIT-DE method and the sensitivity of the HFD in the low field to the way in which the EFG is represented.

We observed that all the extracted HFDs have negative values at low field area(see Figure 6.19 to 6.23). As a distribution function the negative value of $P(H)$ is unacceptable. From the discussion in section 2.2.2 of Chapter 2 we know that inferences that are unacceptable must be from inappropriat assumptions and not from inadequacies of the inferential system. All parts of the model, including the prior information, are exposed to appropriat criticism. The Bayesian inference theory is sending us the information that more realistic model and prior information need to be used to solve this problem.

6.7 Conclusion

Bayesian inference theory has been used to give an elegant solution to the inverse problem related to static hyperfine parameter distributions in Mössbauer spectroscopy. Whereas a constant prior (i.e. assuming no a priori knowledge of the distribution) leads to the usual (and unsuccessful) least-square estimator approach, application of the maximum entropy principle gives a prior that is a multivariate Gaussain function which leads to a well behaved extraction algorithm that incorporates full error propagation. The result is a fast and robust direct extraction method which, for a given assumed elemental line shape model, gives a unique distribution that is not sensitive to the spectral noise or to uncertainties in the values of the non-distributed hyperfine parameters.

The method can be generalized to the cases involving 1) multisite problems where several independent distributions of the same hyperfine parameter occur simultaneously, and 2) several independent or coupled hyperfine parameters that are distributed simultaneously in a single "site" model.

Finally, Bayesian inference theory is seen as a general framework which can accommodate: more complicated situations involving multiple distributions, more highly tuned prior functions resulting from more stringent physical constraints, and various error propagation schemes. This is an effective direct extraction procedure that uses the data directly without filtering and it represents the first time that error propagation to the distribution is performed in Mössbauer spectroscopy.

6.8 Appendix A. Deriving the prior probability distribution of X

We maximize the Shannon measure of the entropy of $P(X_k)$

$$S = -\sum_{k=1}^{K} P(X_k) \log P(X_k) \tag{6.64}$$

subject to

$$\sum_{k=1}^{K} P(X_k) = 1 \tag{6.65}$$

and

$$\sum_{k=1}^{K} P(X_k)(X_k^T L_2^T L_2 X_k) = C_1. \tag{6.66}$$

The Lagrangian is given by

$$\begin{aligned} L \equiv & -\sum_{k=1}^{K} P(X_k) log P(X_k) \\ & -(\lambda_0^2 - 1)\left[\sum_{k=1}^{K} P(X_k) - 1\right] - \frac{1}{2}\lambda^2\left[\sum_{k=1}^{K} P(X_k)(X_k^T L_2^T L_2 X_k) - C_1\right] \end{aligned} \tag{6.67}$$

where $(\lambda_0^2 - 1)$ and $\lambda^2/2$ are the two Lagrange multipliers corresponding to the two constraints. Note that $(\lambda_0^2 - 1)$ and $\lambda^2/2$ are used as the Lagrange multipliers instead of λ_0 and λ as a matter of convenience. Now

$$\frac{\partial L}{\partial P(X_k)} = 0 \implies -log P(X_k) - \lambda_0^2 - \frac{1}{2}\lambda^2 X_k^T L_2^T L_2 X_k = 0, \tag{6.68}$$

such that

$$P(X_k) = exp(-\lambda_0^2) exp(-\frac{1}{2}\lambda^2 X_k^T L_2^T L_2 X_k). \tag{6.69}$$

Substituting equation (6.69) into equation (6.65), we obtain

$$exp(\lambda_0^2) = \sum_{k=1}^{K} exp(-\frac{1}{2}\lambda^2 X_k^T L_2^T L_2 X_k) \tag{6.70}$$

which is just a normalization constant. So the prior probability distribution for X is

$$P(X) \propto exp(-\frac{1}{2}\lambda^2 X^T L_2^T L_2 X). \tag{6.71}$$

6.9 Appendix B. Deriving $\hat{X}$ by minimizing $\Phi(X)$

To minimize the exponent

$$\Phi(X) = (b - AX)^T(b - AX) + \lambda^2 X^T L_2^T L_2 X$$

we have to give an explicit expression for $\Phi(X)$. By using equations (6.20) - (6.22) where

$$L_2 = \begin{bmatrix} 1 & -2 & 1 & & & \\ & 1 & -2 & 1 & & \\ & & \ddots & \ddots & \ddots & \\ & & & 1 & -2 & 1 \end{bmatrix}_{(N-2)\times N} \tag{6.72}$$

$\Phi(X)$ can be written as

$$\Phi(X) = \sum_{i=1}^{M}(b_i - \sum_{j=1}^{N} a_{ij}x_j)^2 + \lambda^2 \sum_{j=2}^{N-1}(x_{j-1} - 2x_j + x_{j+1})^2. \tag{6.73}$$

Now, let

$$\frac{\partial \Phi(X)}{\partial x_k} = 0 \quad k = 1, \cdots, N, \tag{6.74}$$

that is,

$$\begin{cases} -\sum_{i=1}^{M} b_i a_{i1} + \sum_{i=1}^{M} \left(\sum_{j=1}^{N} a_{ij}x_j\right) a_{i1} + \lambda^2(x_1 - 2x_2 + x_3) = 0 \\ \\ -\sum_{i=1}^{M} b_i a_{i2} + \sum_{i=1}^{M} \left(\sum_{j=1}^{N} a_{ij}x_j\right) a_{i2} + \lambda^2(-2x_1 + 5x_2 - 4x_3 + x_4) = 0 \\ \\ -\sum_{i=1}^{M} b_i a_{ik} + \sum_{i=1}^{M} \left(\sum_{j=1}^{N} a_{ij}x_j\right) a_{ik} + \lambda^2(x_{k-2} - 4x_{k-1} + 6x_k - 4x_{k+1} + x_{k+2}) = 0 \\ \\ k = 3, \cdots, N-2 \\ \\ -\sum_{i=1}^{M} b_i a_{iN-1} + \sum_{i=1}^{M} \left(\sum_{j=1}^{N} a_{ij}x_j\right) a_{iN-1} + \lambda^2(x_{N-3} - 4x_{N-2} + 5x_{N-1} - 2x_N) = 0 \\ \\ -\sum_{i=1}^{M} b_i a_{iN} + \sum_{i=1}^{M} \left(\sum_{j=1}^{N} a_{ij}x_j\right) a_{iN} + \lambda^2(x_{N-2} - 2x_{N-1} + x_N) = 0. \end{cases} \tag{6.75}$$

This set of linear algebraic equations can be rewritten in matrix notation as

$$(A^T A + \lambda^2 L_2^T L_2)X = A^T b, \tag{6.76}$$

so the $\hat{X}$ which maximizes the posterior probability distribution $P(X \mid b)$ is

$$\hat{X} = (A^T A + \lambda^2 L_2^T L_2)^{-1} A^T b. \tag{6.77}$$

Now we can use $\hat{X}$ to rewrite $\Phi(X)$,

$$\begin{aligned} \Phi(X) &= (b - AX)^T(b - AX) + \lambda^2 X^T L_2^T L_2 X \\ \\ &= b^T b - X^T A^T b - b^T AX + X^T A^T AX + \lambda^2 X^T L_2^T L_2 X \end{aligned}$$

$$= X^T(A^TA + \lambda^2 L_2^T L_2)X - X^T A^T b - b^T AX + b^T b. \quad (6.78)$$

From equation (6.77), we have

$$A^T b = (A^TA + \lambda^2 L_2^T L_2)\hat{X} \quad (6.79)$$

and

$$bA^T = \hat{X}^T(A^TA + \lambda^2 L_2^T L_2). \quad (6.80)$$

Substituting equations (6.79) and (6.80) into equation (6.78), we obtain

$$\Phi(X) = (X - \hat{X})^T(A^TA + \lambda^2 L_2^T L_2)(X - \hat{X}) + b^T b - \hat{X}^T(A^TA + \lambda^2 L_2^T L_2)\hat{X}. \quad (6.81)$$

We also let

$$C_2 = b^T b - \hat{X}^T(A^TA + \lambda^2 L_2^T L_2)\hat{X} \quad (6.82)$$

which is independent of X. Finally $\Phi(X)$ is expressed as:

$$\Phi(X) = (X - \hat{X})^T(A^TA + \lambda^2 L_2^T L_2)(X - \hat{X}) + C_2. \quad (6.83)$$

6.10 Appendix C. Calculating $\hat{X}$ using GSVD

From equations (6.37) and (6.40) in Section 3.2, we have

$$\begin{aligned} A^TA + \lambda^2 L_2^T L_2 &= (W^{-1})^T Z^T U^T U Z W^{-1} + \lambda^2 (W^{-1})^T \begin{bmatrix} \Gamma^T \\ 0 \end{bmatrix} V^T V [\Gamma, 0] W^{-1} \\ &= (W^{-1})^T \left\{ \begin{bmatrix} Z_H^T Z_H & 0 \\ 0 & I_0 \end{bmatrix} + \lambda^2 \begin{bmatrix} \Gamma^T\Gamma & 0 \\ 0 & 0 \end{bmatrix} \right\} W^{-1} \\ &= (W^{-1})^T \left\{ diag \left[(\zeta_1^2 + \lambda^2\gamma_1^2), \cdots, (\zeta_H^2 + \lambda^2\gamma_H^2), 1, \cdots, 1 \right] \right\} W^{-1} \end{aligned} \tag{6.84}$$

such that

$$(A^TA + \lambda^2 L_2^T L_2)^{-1} = W \left\{ diag \left[\frac{1}{(\zeta_1^2 + \lambda^2\gamma_1^2)}, \cdots, \frac{1}{(\zeta_H^2 + \lambda^2\gamma_H^2)}, \cdots, 1, \cdots, 1 \right] \right\} W^T. \tag{6.85}$$

Now let

$$F = diag \left[\frac{(\zeta_1/\gamma_1)^2}{(\zeta_1/\gamma_1)^2 + \lambda^2}, \cdots\cdots, \frac{(\zeta_H/\gamma_H)^2}{(\zeta_H/\gamma_H)^2 + \lambda^2} \right]. \tag{6.86}$$

Then

$$(A^TA + \lambda^2 L_2^T L_2)^{-1} = W \begin{bmatrix} F & 0 \\ 0 & I_0 \end{bmatrix} \begin{bmatrix} Z_H^{-1} Z_H^{-1} & 0 \\ 0 & I_0 \end{bmatrix} W^T \tag{6.87}$$

and

$$(A^TA + \lambda^2 L_2^T L_2)^{-1}A^T = W \begin{bmatrix} F & 0 \\ 0 & I_0 \end{bmatrix} \begin{bmatrix} Z_H^{-1}Z_H^{-1} & 0 \\ 0 & I_0 \end{bmatrix} W^T(W^{-1})^T Z^T U^T$$

$$= [W_H, W_0] \begin{bmatrix} F & 0 \\ 0 & I_0 \end{bmatrix} \begin{bmatrix} Z_H^{-1} & 0 \\ 0 & I_0 \end{bmatrix} \begin{bmatrix} U_H^T \\ U_0^T \end{bmatrix}$$

$$= W_H F Z_H^{-1} U_H^T + W_0 U_0^T. \tag{6.88}$$

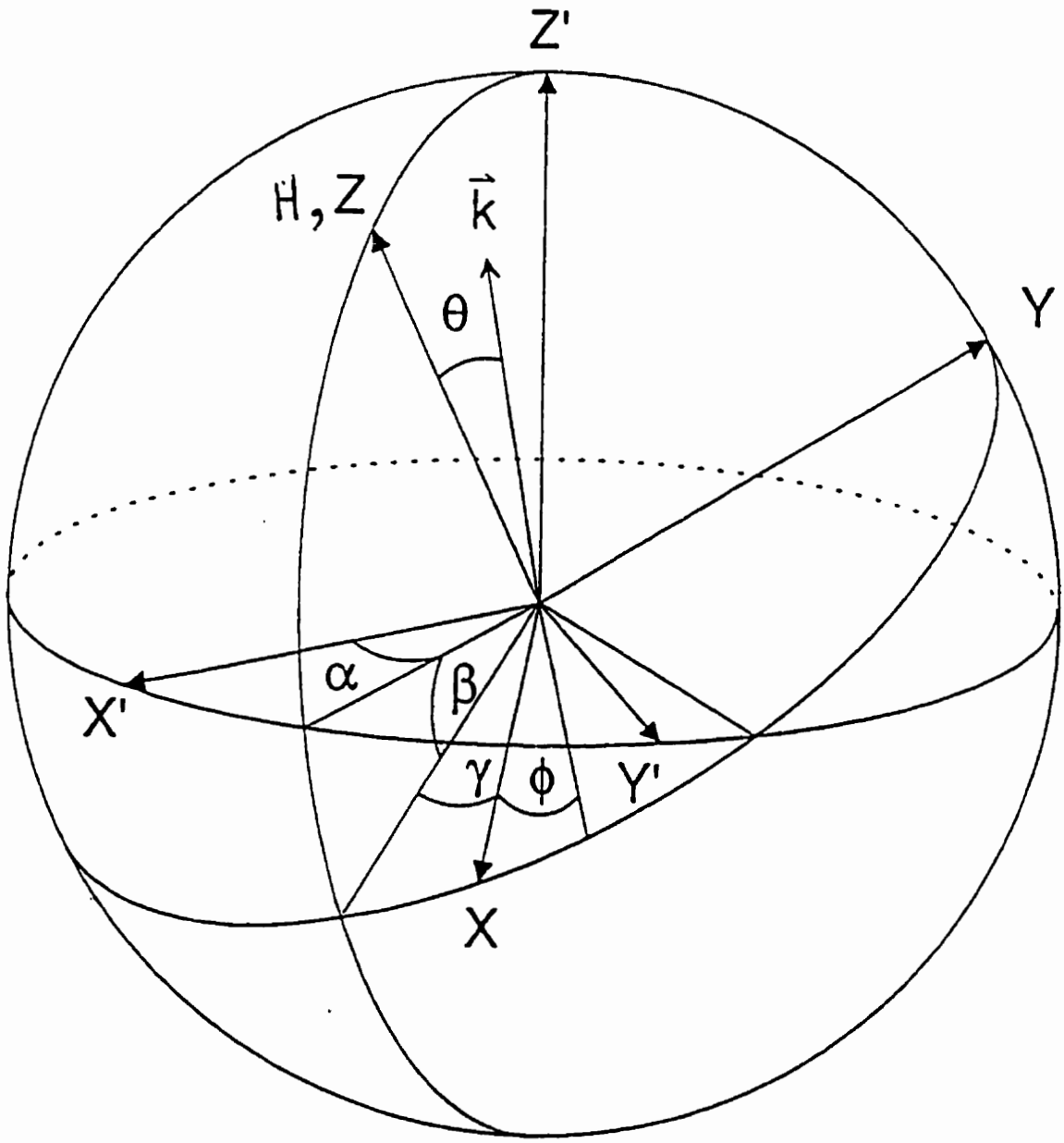

Figure 6.13: Illustration of the Euler angles α, β, and γ transforming the principle axes x', y', z' where the magnetic field H is parallel to Z. θ and ϕ represent the polar and azimuthal angle of the γ-ray direction in the laboratory system.

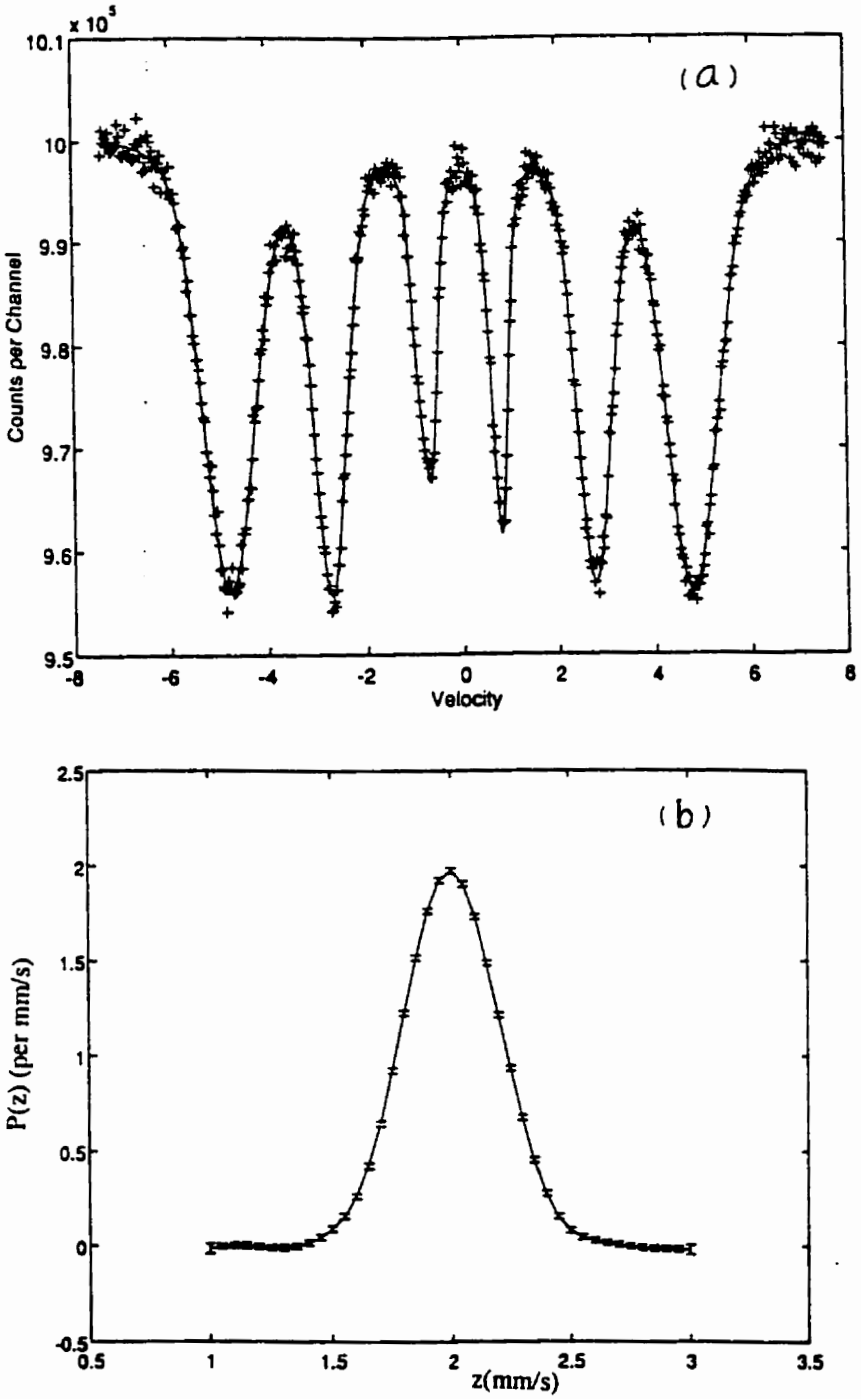

Figure 6.14: The extracted hyperfine field distribution from the simulated spectrum using Blaes' line shape model. (a) The simulated spectrum (solid line) and the recalculated spectrum (dots); (b) The exact distribution centered at $\mid g_{3/2}\mu_{NH} \mid = 2$ mm/s (solid line) and the extracted distribution with error bars.

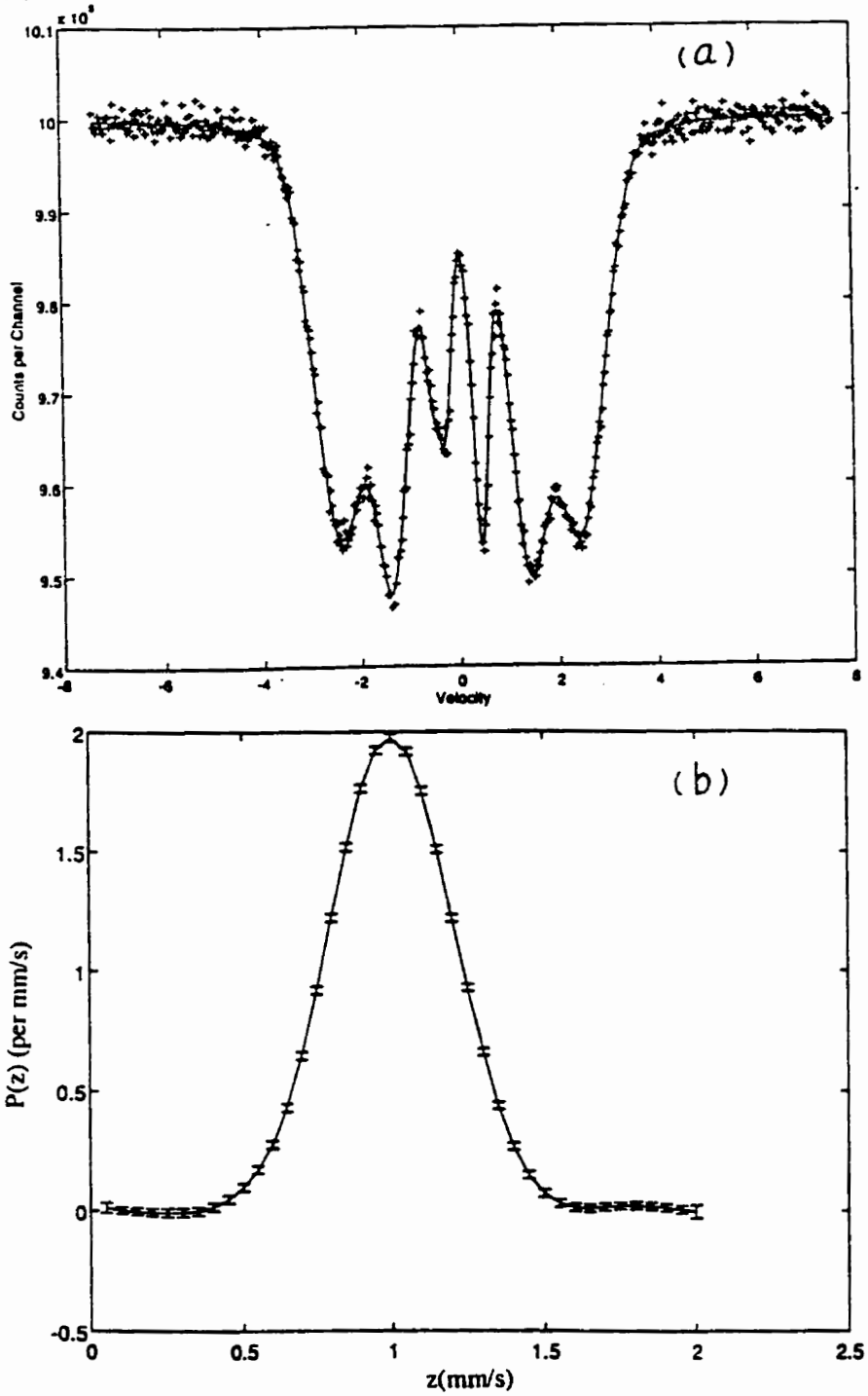

Figure 6.15: The extracted hyperfine field distribution from the simulated spectrum using Blaes' line shape model. (a) The simulated spectrum (solid line) and the recalculated spectrum (dots); (b) The exact distribution centered at $\mid g_{3/2}\mu_{NH} \mid = 1$ mm/s (solid line) and the extracted distribution with error bars.

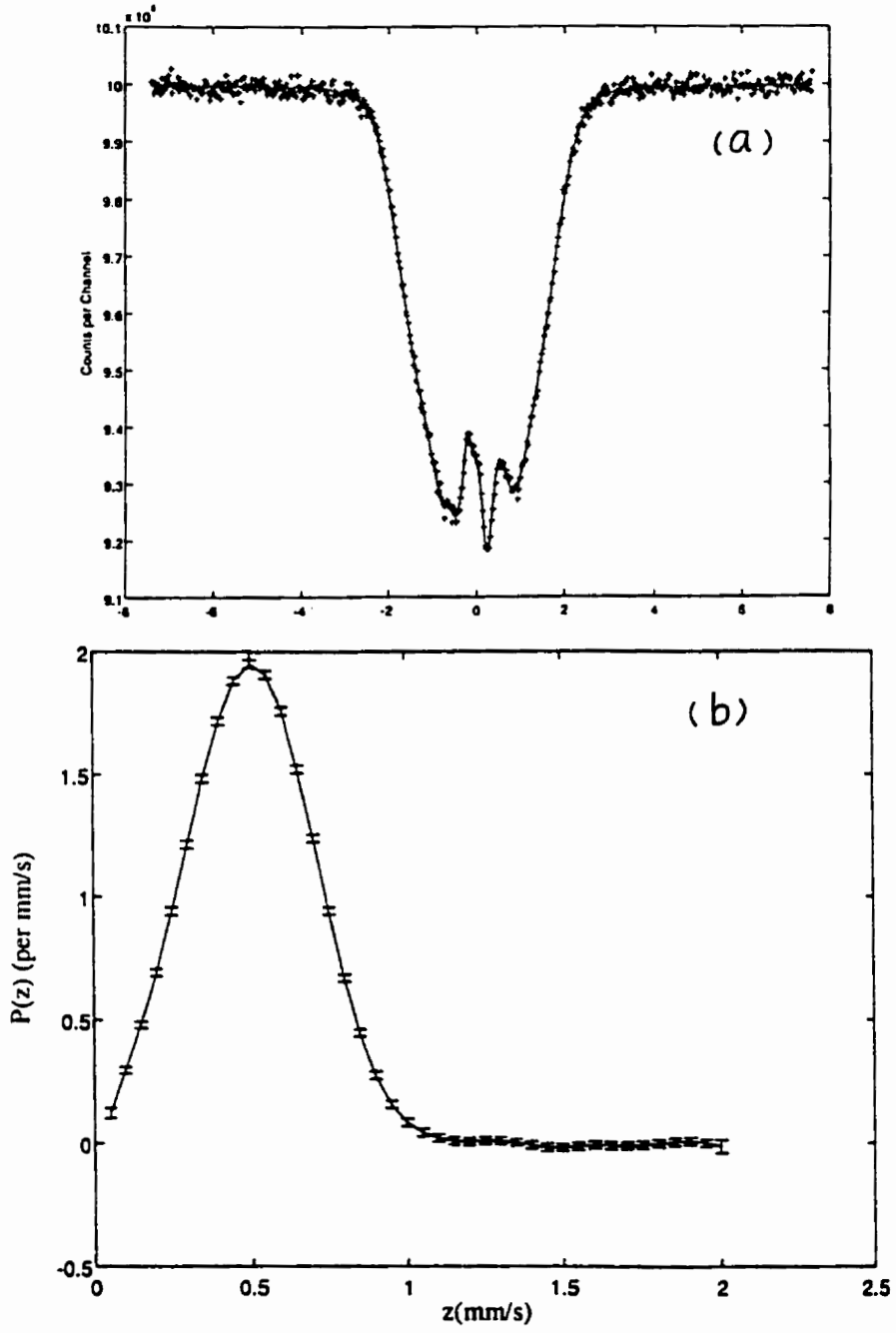

Figure 6.16: The extracted hyperfine field distribution from the simulated spectrum using Blaes' line shape model. (a) The simulated spectrum (solid line) and the recalculated spectrum (dots); (b) The exact distribution centered at $\mid g_{3/2}\mu_{NH} \mid = 0.5$ mm/s (solid line) and the extracted distribution with error bars.

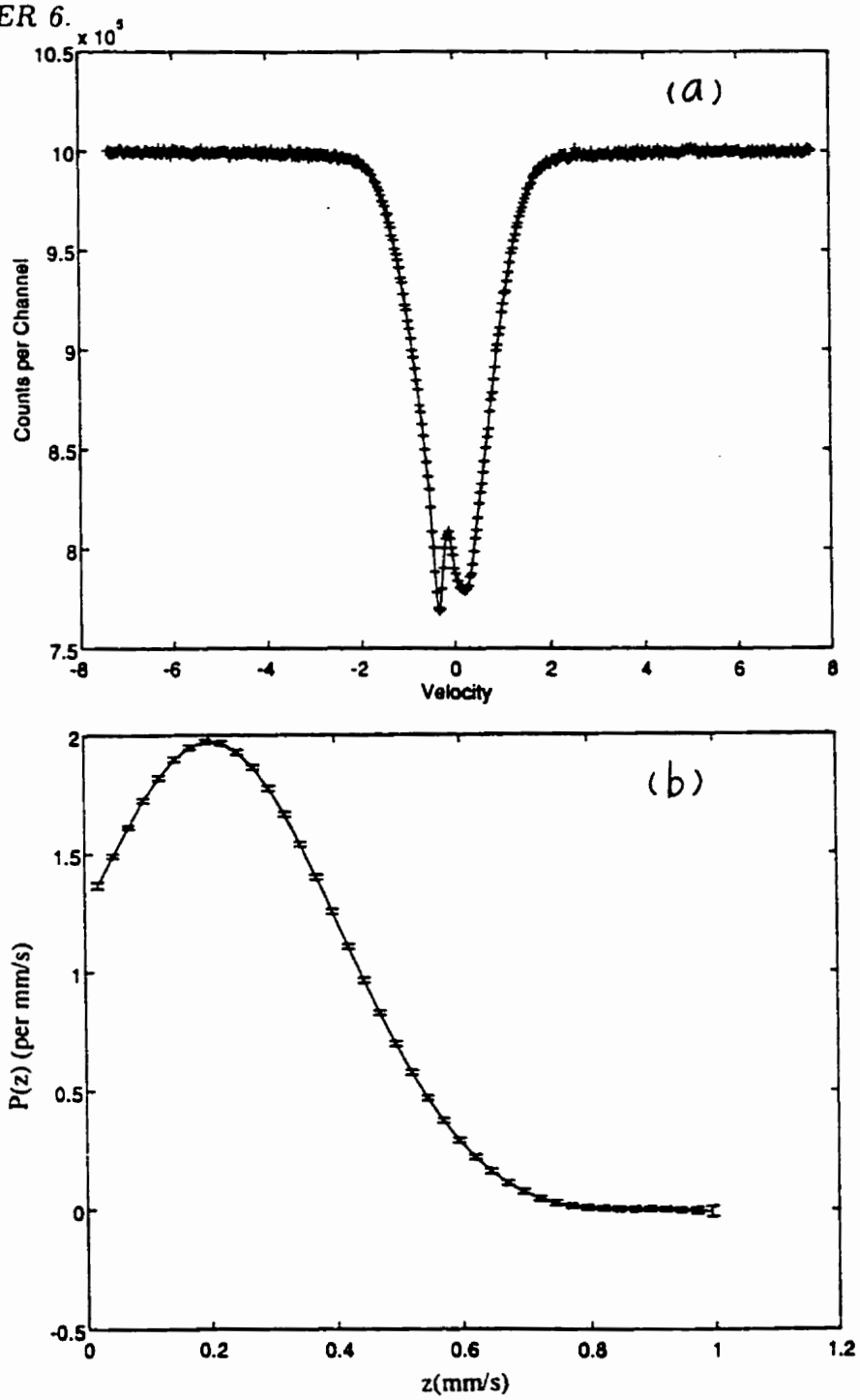

Figure 6.17: The extracted hyperfine field distribution from the simulated spectrum using Blaes' line shape model. (a) The simulated spectrum (solid line) and the recalculated spectrum (dots); (b) The exact distribution centered at $\mid g_{3/2}\mu_{NH} \mid = 0.2$ mm/s (solid line) and the extracted distribution with error bars.

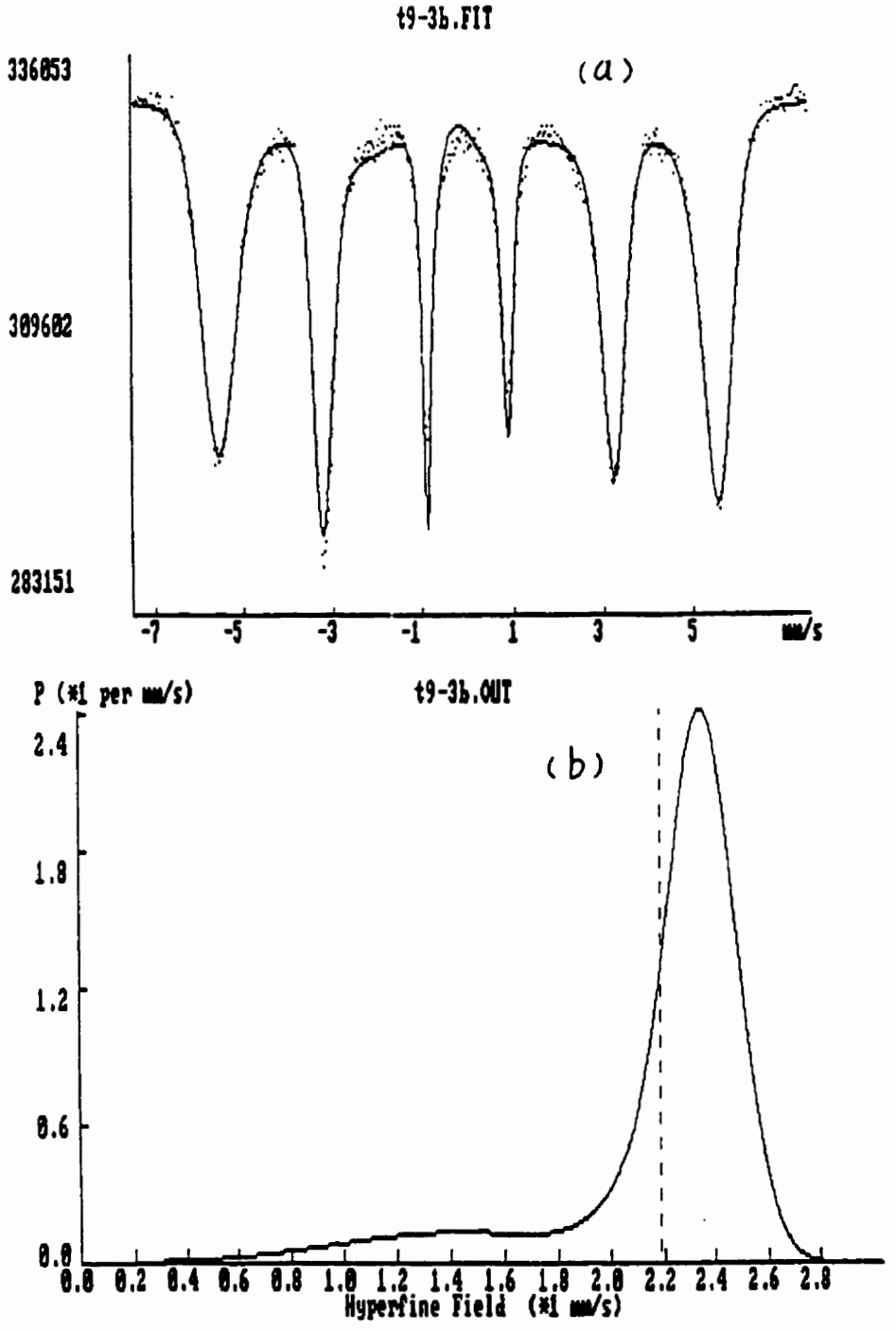

Figure 6.18: (a) Fitted thin limit spectrum at $T = 4.2K$ and (b) corresponding HFD for Fe_65Ni_35 Invar using the VB-F method.

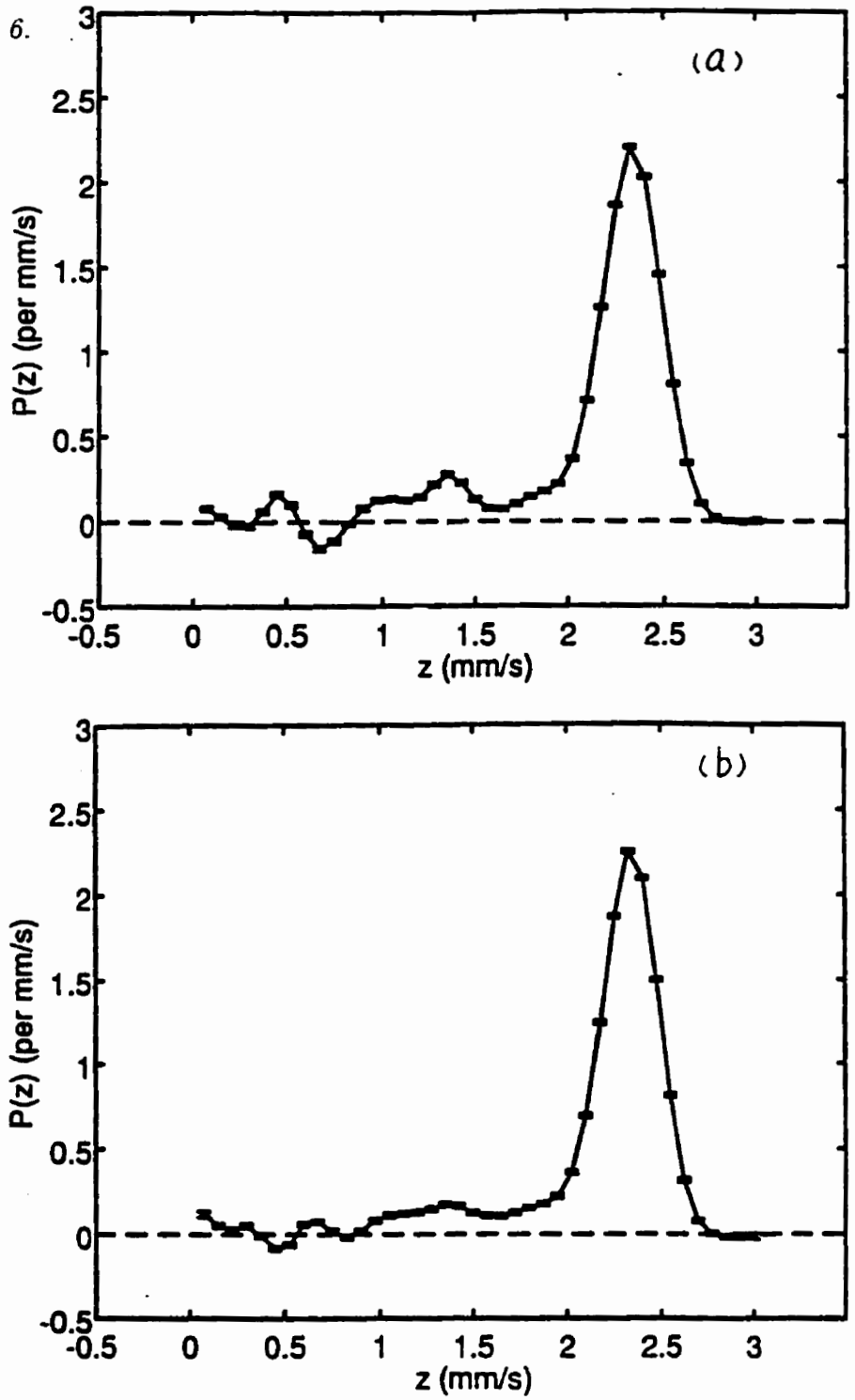

Figure 6.19: Extracted HFDs using BIT-DE method and perturbation sextet model. (a) $\delta_0 = 0.04716$ mm/s; $\delta_1 = -0.00969$; $\epsilon_0 = 0.40516$ mm/s; $\epsilon_1 = -0.17221$. (b) $\delta_0 = 0.026$ mm/s; $\delta_1 = 0$; $\epsilon_0 = 0.029$ mm/s; $\epsilon_1 = 0$.

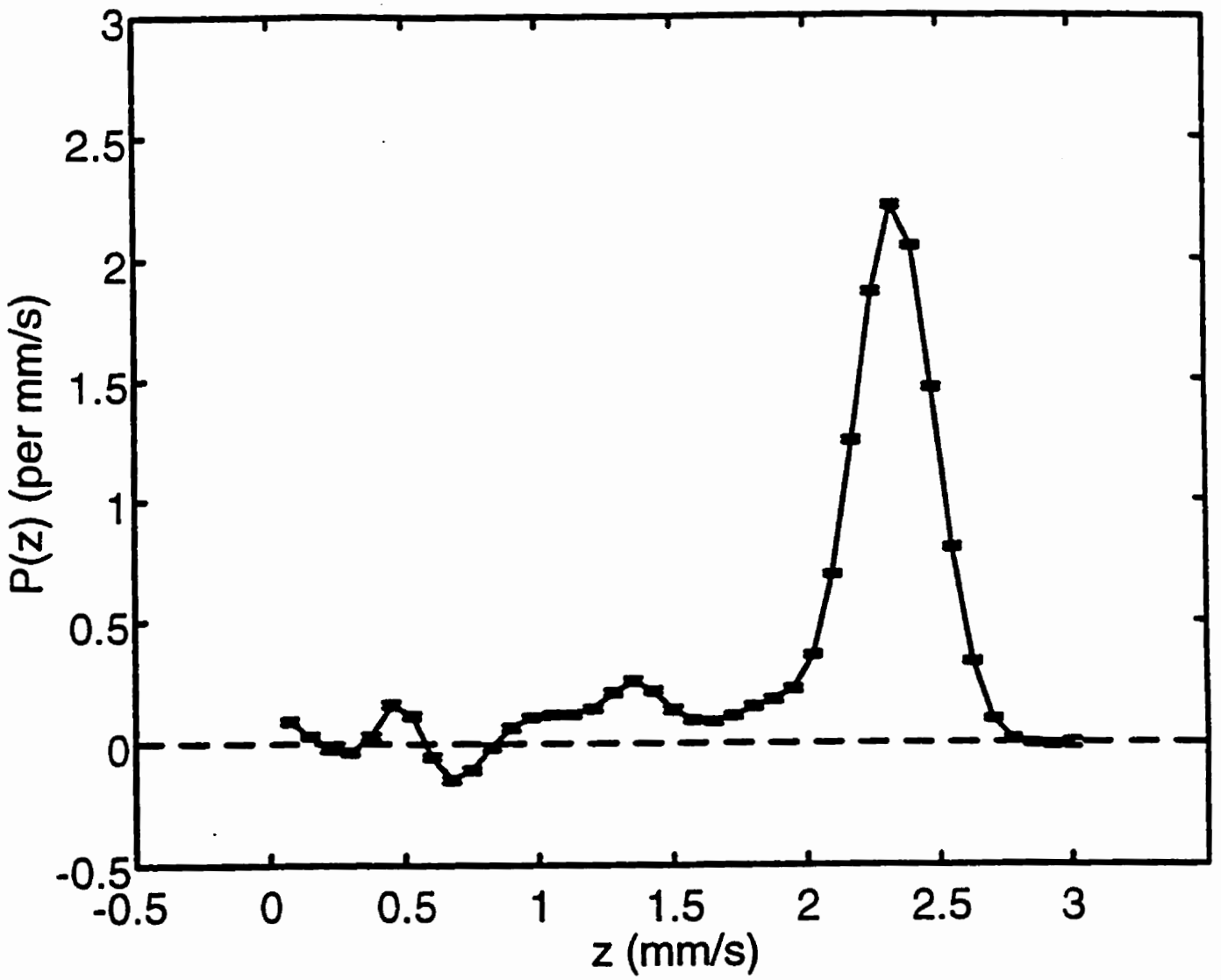

Figure 6.20: Extracted HFDs using BIT-DE method and Blaes line shape model described in C with $\beta = 45$ degrees and single value for K.

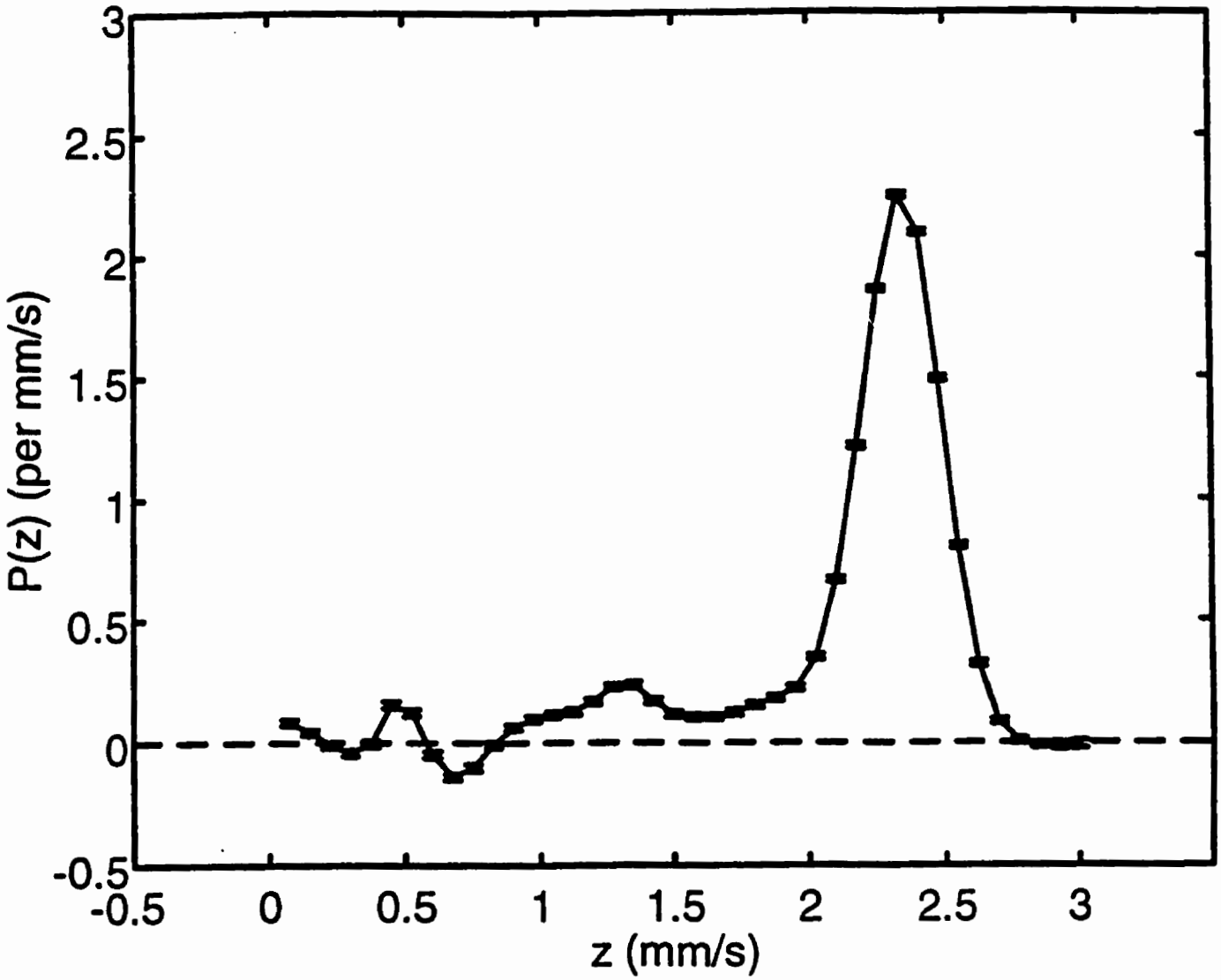

Figure 6.21: Extracted HFD using BIT-DE and Blaes line shape model described in D with single value for K.

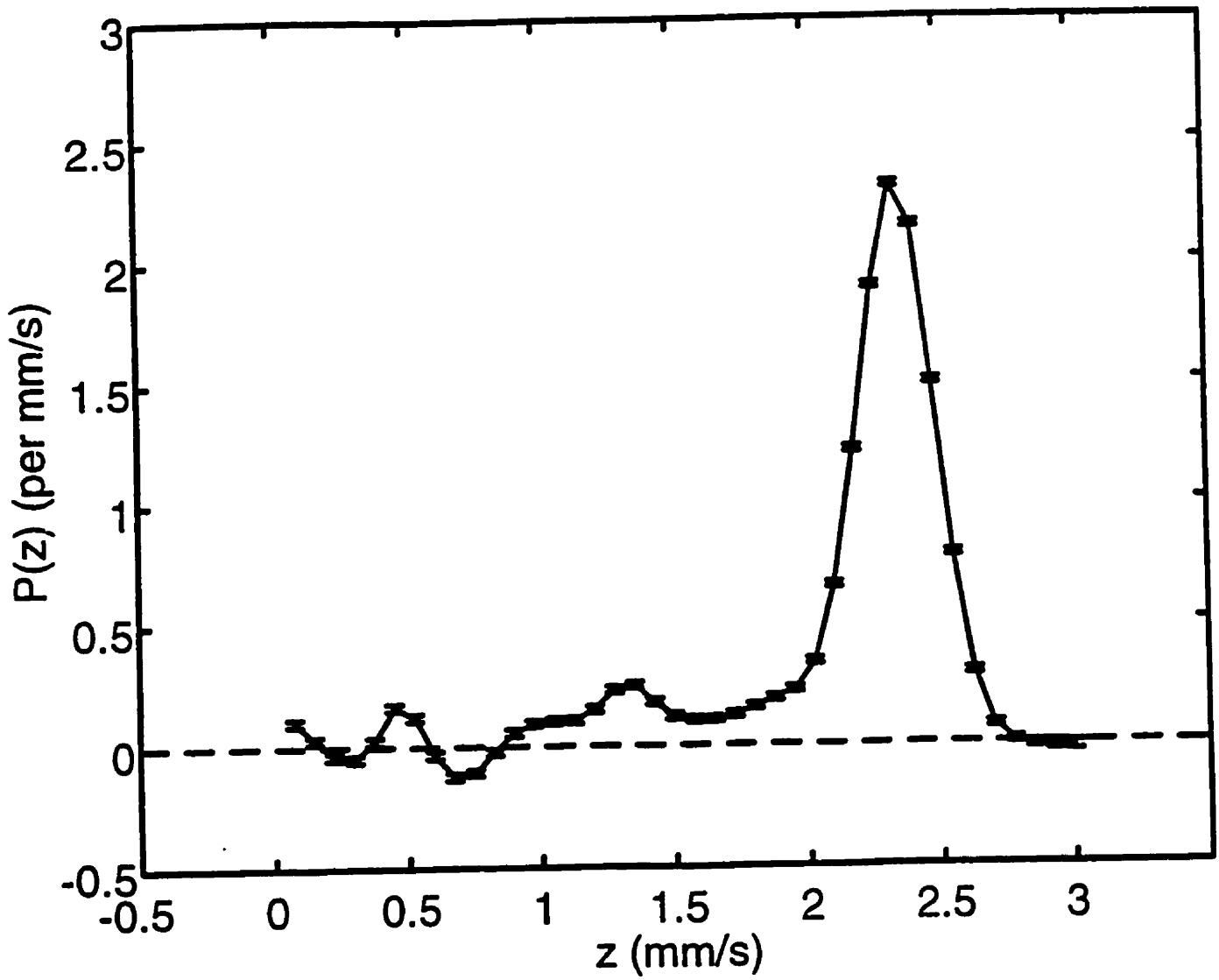

Figure 6.22: Extracted HFD using BIT-DE and Blaes line shape model described in E with independent Gaussian distribution for K.

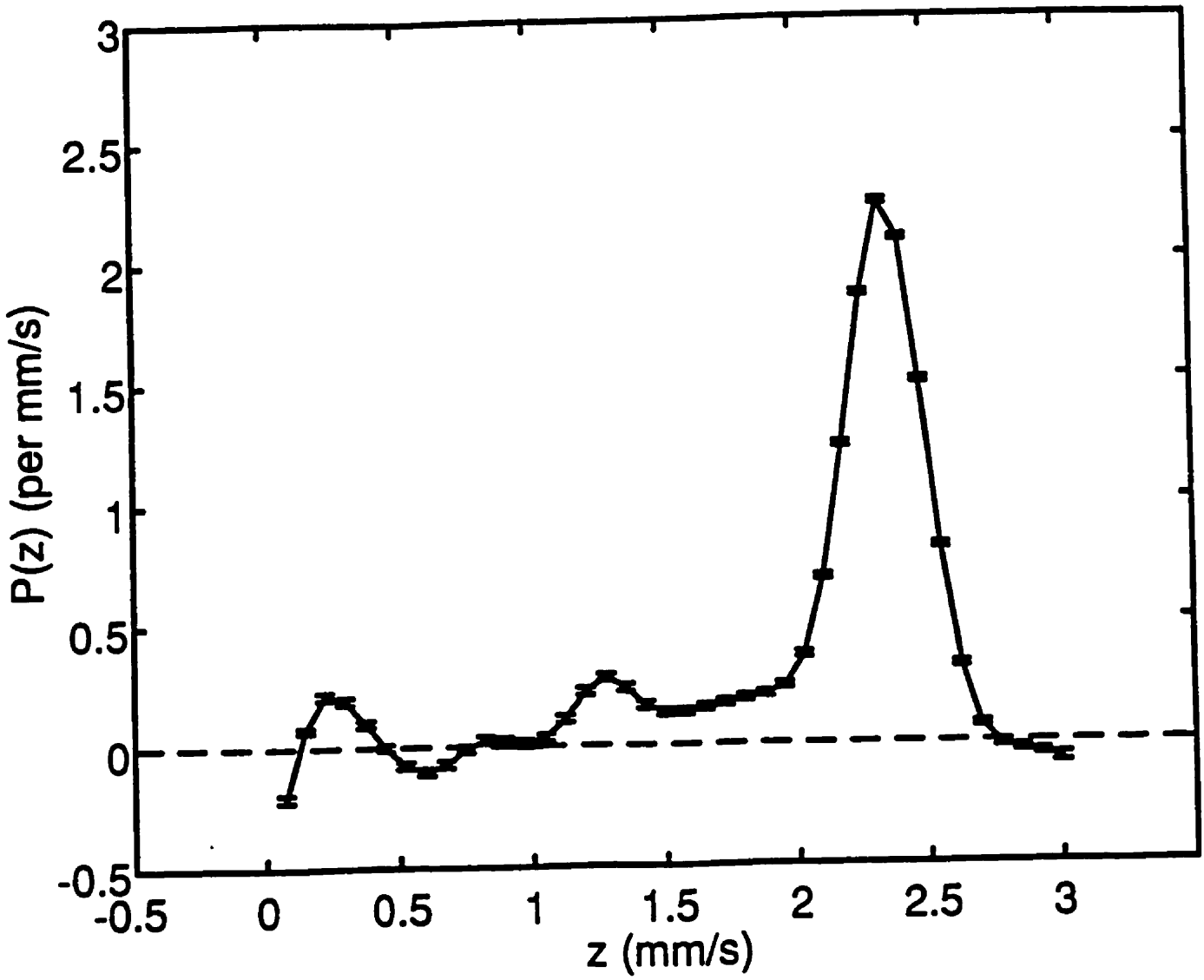

Figure 6.23: Extracted HFD using BIT-DE and Blaes line shape model described in E with linearly coupled distibution to H for K.

Chapter 7. Conclusion and Discussion

Two very common and well known problems in applied physics have been studied in this thesis by using the Bayesian inference theory and Gibbs sampling technology. The first one is from the arena of discrete time signal analysis and the second is from Mössbuaer spectroscopy which is very useful not only in physics but also in material science and mineralogy.

From the applications to both spectral analysis and parameter estimation in discrete time signals and hyperfine parameter and field distribution extraction in Mössbauer spectroscopy it can be seen that the whole process fits into the plausible reasoning and follows the iteration steps depicted in Figure 1 and Figure 2. The Bayesian inference theory provides a unique mathematical language to describe the essential features of the plausible reasoning.

The results from the first application showed that the Bayesian approach to spectral analysis is very promising and robust to the noise and non-honmornic features in the signals. Also from the results, the shortcomings in the FFT technique which is widely used in the signal processing field are overcome. Based on this very promising study, potentially the method can be applied to many other areas such as the sunspot data analysis , nuclear magnetic resonance data analysis, etc.. The speed of the Bayesian approach in spectral analysis, however, can not be compared with FFT due to the many steps of sampling and minimization. Higher speed algorithms need to be developed in the future. But with more and more powerful computers, this shortcoming can be overcome over time. All the applications presented in this

study used the non-informative prior which is the most simple type of prior. For the applications where a more informative prior is available, studies need to be focused on how to tune the prior information into the posterior in the way that the rest of the calculations are numerically durable.

This is the first time that BIF-DE method has been successfully applied to Mössbauer spectroscopy. The BIT gives a general framework which can accommodate more complicated situations involving complex line shape functions, multiple distributions, more highly tuned prior information functions which the other methods cannot easily deal with. It is the first time that the error propagation to the distribution is performed in the extraction. This unique feature can help physicists to distinguish the true features of the distributions from the experimental/analysis artifacts.

The method described in this work can be applied to many problems involving detailed analysis of time-series data with noise and inverse problems involving differential and integral equations. They include varies fields from medical signal/image to dynamic systems, speech processing to astronomy.

The beauty of the Bayesian inference theory is that it allows us to tune more and more precise prior information to the investigation with each step we move on to. Indeed the Bayesian inference theory provides a system of statistical inference intrinsically suited to iterative model building, which is in turn an essential part of scientific investigation.

References

[1] Box, George E.P., and Tiao, George C., (1973), Bayesian Inference in Statistical Analysis, pp.25-60.

[2] Rao, C.R., (1965), Linear Statistical Inference ans its Applications

[3] Skilling, J., Classic Maximum Entropy, In Maximum Entropy and Bayesian Methods. Combridge, England, 1988 (e.d. by J. Skilling), pp45-52, Kluwer, Dordrecht.

[4] Bretthorst, G, Larry, (1987), Bayesian Spectrum Analysis and Parameter Estimation, Lecture Notes in Statistics **48**, Berger, J., ed., Spinger-Verlag.

[5] Jaynes, E.T., Probability Theory: the Logic of Science.

[6] Naylor, J.C., anf Smith, A.F.M. (1982), Applied Statistics, **31**, pp.214-225.

[7] Smith, A.F.M., Skene, A.M., Shaw, J.E.H., Naylor, J.C., and Dransfield, M. (1985), Communications in Statistics, Part A-Theory and Methods, **14**, pp.1079-1102.

[8] Smith, A.F.M., Skene, A.M., Shaw, J.E.H., and Naylor, J.C. (1987), The Statistician, **36**, pp.75-82.

[9] Tierney, L., and Kadane, J.B. (1986), Journal of the American Statistical Association, **81**, pp.82-86.

[10] Shaw, J.E.H. (1988), The Annals of Statistics, **16**, pp.895-914.

[11] Gewek, J. (1988), Journal of Econometrics, **38**, pp.73-90.

[12] Tanner, M., and Wang, W. (1987), Journal of the American Statistical Association, **82**, pp.528-550.

[13] Tanner, M.A. (1993), Tools for Statistical Inference, Springer-Verlag.

[14] Geman, S. and Geman, D. (1984) IEEE Trans. Pattn Anal. Mach. Intell., **6**, 721-741.

[15] Rubin, D.B. (1987), Journal of the American Statistical Association, **82**, pp.543-546.

[16] Rubin, D.B. (1988), "Using the SIR Algorithm to Simulate Posterior Distributions", in Bayesian Statistics 3, eds. Bernado, J. M., DeGroot, M.H., Lindely, D.V., and Smith, A.F.M., Oxford, U.K.: Oxford University Press, PP.395-402.

[17] Gelfand, Alan E., and Smith, Adrian F.M., (1990), J. Am. Statis. Assoc., **85**, pp.398-409.

[18] Gelfand, Alan E., et al., (1990), J. Am. Statis. Assoc., **85**, pp.972-985.

[19] Robinson, E.A., (1982), "A Historical Perspective of Spectrum Estimation," Proceedings of the IEEE, **70**, pp.855-906.

[20] Marple, S.L., (1987), Digital Spectral Analysis with Applications, Prentice-Hall, New Jersey.

[21] Laplace, P.S., (1812), Théorie Analytique des Probabilitiés, Paris, (2nd edition, 1814; 3rd edition, 1820).

[22] Legendre, A.A., (1860), "Nouvelles Méthodes pour la Détermination des Orbits des Cométes,", Paris.

[23] Gauss, K.F., (1963 reprint) Theory of the Motion of the Heavenly Bodies Moving About the Sun in Conic Sections, Dover Publications, Inc., New York.

[24] Cooley, J.W., and Tukey, J.W.,(1965), Mathe Comput., **19**, pp.297-301.

[25] Schuster, A., (1905), "The Periodogram and its Optical Analogy," Proceedings of the Royal Society of London, **77**, pp.136.

[26] Bretthorst, G, Larry, (1990), J. Magn. Reson., **88**, pp.571-595.

[27] Jaynes, E.T., (1987), "Bayesian Spectrum and Chirp Analysis", in Maximum Entropy and Bayesian Analysis and Estimation Problems, pp.1-37, Smith, C. Ray and Erickson, G.J., ed, D. Reidel, Dordrecht-Holland.

[28] Bretthorst, G, Larry, and Smith, C.R., (1989), Prceedings of the SPIE - The International Society for Optical Engineering, **1050**, pp.93-104.

[29] Bretthorst, G, Larry, (1990), J. Magn. Reson., **88**, pp.533-551.

[30] Bretthorst, G, Larry, (1990), J. Magn. Reson., **88**, pp.552-570.

[31] Bretthorst, G, Larry, (1991), J. Magn. Reson., **93**, pp.369-394.

[32] Bretthorst, G, Larry, (1992), J. Magn. Reson., **98**, pp.501-523.

[33] Bretthorst, G, Larry, et al, (1988), J. Magn. Reson., **79**, pp.369-376

[34] Bretthorst, G, Larry, et al, (1989), Magn. Reson. in Medicine, **9**, pp.282-287.

[35] Kotyk, John J., et al, (1992), J. Magn. Reson., **98**, pp.483-500.

[36] Neil, Jeffrey, J., and Bretthorst, G, Larry, (1993), Magn. Reson. in Medicine, **29**, pp.642-647.

[37] Cox, R. T, (1961), The algebra of probable inference, Johns Hopkins Press, Baltimore.

[38] Djuric P M and Kay S M 1990 it IEEE Trans. Acoust., Speech, Signal Processing **38** 2118-2126.

[39] Dou L, Hodgson RJW and Rancourt D 1995 Bayesian Inference Theory Applied to Hyperfine Parameter Distribution Extraction in Mossbauer Spectroscopy *Nucl. Instr. & Methods* **B100** 511-518 .

[40] Harris, Robin K., (1983), Nuclear Magnetic Resonance Spectroscopy, pp.66-94.

[41] Smith, A.F.M. and Gelfand, A.E. (1992) Bayesian statistics without tears: a sampling -resampling perspective. *Am. Statistn.*, **46**, 84-88.

[42] Tierney, L. (1991) Markov chains for exploring posterior distributions. Submitted to *Ann. Statist.*

[43] Greenwood, N.N. and Gibb, T.C., *Mössbauer spectroscopy*, (Chapman and Hall Ltd. 1971).

[44] H. Hochstadt, *Integral Equations*, (Wily, New York, 1973).

[45] F. Varret, A. Gerard and P. Imbert, Phys. Stat. Solidi b 43 (1971) 723.

[46] B. Window, J. Phys. E: Sci. Instrum. 4 (1971) 401.

[47] J. Hesse and A. Rübartsch, J. Phys. E: Sci. Instrum. 7 (1974) 526.

[48] G. Le Caër and J.M. Dubois, J. Phys. E: Sci. Instrum. 12 (1979) 1083.

[49] C. Wivel and S. Mørup, J. Phys. E: Sci. Instrum. 14 (1981) 605.

[50] P. Levitz, D. Bonnin, G. Callas and A.P. Legrand, J. Phys. E: Sci. Instrum. 13 (1980) 427.

[51] I. Vincze, Nucl. Instr. Meth. 199 (1982) 247.

[52] G. Le Caër, J.M. Dubois, H. Fischer, U. Gonser and H.G. Wagner, Nucl. Instr. Meth. Phys. Res. B5 (1984) 25.

[53] R.A. Brand and G. Le Caër, Nucl. Instr. Meth. Phys. Res. B34 (1988) 272.

[54] P.M.A. de Bakker, E. De Grave, R.M. Persoons, L.H. Boweu and R.E. Vandenberghe, Meas. Sci. Technol. 1 (1990) 954.

[55] D.G. Rancourt and J.Y. Ping, Nucl. Instr. Meth. Phys. Res. B58 (1991) 85.

[56] J.Y. Ping and D.G. Rancourt, Hyp. Int. 92 (1994) 1209.

[57] J.Y. Ping and D.G. Rancourt, Hyp. Int. 92 (1994) 1203.

[58] S.J. Campbell, in: Trends in Mössbauer Spectroscopy, Proc. 2nd Seeheim Workshop on Mössbauer Spectroscopy May 1983, eds. P. Gutlitch and G.M. Kalvius, University of Mainz (1983) p.117.

[59] F. Varret, J.M. Greneche and J. Teillet, in: Trends in Mössbauer Spectroscopy, Proc. 2nd Seeheim Workshop on Mössbauer Spectroscopy May 1983, eds. P. Gutlitch and G.M. Kalvius, University of Mainz (1983) p. 105

[60] G. Longworth, in: Mössbauer Spectroscopy Applied to Inorganic Chemistry, ed. G.J. Long (Plenum, New York, 1987) vol. 2, p. 289.

[61] S.J. Campbell and F. Aubertin, in: Mössbauer Spectroscopy Applied to Inorganic Chemistry, ed. G.J. Long (Plenum, New York, 1989) vol. 3, p. 183.

[62] G. Le Caër and R.A. Brand, Hyp. Int. 71 (1992) 1507.

[63] R.E. Vandenberghe, E. De Grave and P.M.A. de Bakker, Hyp. Int. 83 (1994) 29.

[64] D.G. Rancourt, in: Mössbauer Spectroscopy Applied to Magnetism and Materials Science, eds. G.J. Long and F. Grandjean (Plenum, New York, 1996) vol. 2, In Press.

[65] G.F. Knoll, *Radiation Detection and Measurement*, (John Wiley & Sons, New York, 1989).

[66] G. Talenti, *Inverse Problems, Lecture Notes in Mathematics 1225*, (Springer Verlag, Berlin, 1986).

[67] I.P. Nedelkov, Computer Physics Communications, 4 (1972) 157.

[68] J.N. Kapur and H.K. Kesavan, *Entropy Optimization Principles with Applications*, (Academic Press, New York, 1992).

[69] A.N. Tikhonov and V.Y. Arsenin, *Solutions of Ill-Posed Problems*, (Winston & Sons, Washington, D.C. 1977).

[70] P.C. Hansen and D.P. O'Leary, SIAM J. Sci. Comput. 14 (1993) 1487.

[71] P.C. Hansen, BIT 29 (1989) 491.

[72] Puerta, J. and Martin, P., (1981), Applied Optics, **20**, pp. 3923-3928.

[73] Rancourt, D.G., Ping, J.Y., and Berman, R.G. (1994), Phys. Chem. Minerals, **21**, pp.258-267.

[74] Ping,J.Y., Rancourt, D.G., and Stadnik, Z.M., (1991), Hyp. Interact. **69**, pp.493-496.

[75] Sabina, A.P. (1986) Geological Survey og Canada, (Miscellaneous report 39), Ottawa.

[76] Rancourt, D.G., Tume, P., and Lalonde, A.E., (1993), Phys. Chem. Minerals, **20**, pp.276-284.

[77] D.G. Rancourt, M.Z. Dang, and A.E. Lalonde, Am. Mineral. 77 (1992) 34.

[78] M. Shiga, in *Materials Science and Technology*, edited by R.W. Cahn, P. Haasen, and E.J. Kramer (VCH, Weinheim, 1994), Vol. 3B, Part II, p. 159; and references therein.

[79] D.G. Rancourt, S. Chehab, and G. Lamarche, J. Magn. Magn Mat. **78**, 129 (1989).

[80] K.B. Reuter, D.B. Williams, and J.I. Goldstein, Metall. Trans. **20A**, 719 (1988); and references therein.

[81] D.G. Rancourt and R.B. Scorzelli, J. Magn. Magn. Mat. (in press).

[82] D.G. Rancourt, H.H.A. Smit, and R.C. Thiel, J. Magn. Magn. Mat. **66**, 121 (1987); and references therein.

[83] R.J. Weiss, Proc. Phys. Soc. **82**, 281 (1963).

[84] M. Matsui, K. Adachi, and S. Chikazumi, J. Appl. Phys. **51**, 6319 (1980).

[85] S. Chikazumi, J. Magn. Magn. Mat. **10**, 113 (1979).

[86] E.P. Wohlfarth, J. Magn. Magn. Mat. **10**, 120 (1979).

[87] A.Z. Menshikov, Physica B **161**, 1 (1989).

[88] W.J. Carr Jr., Phys. Rev. **85**, 590 (1952).

[89] D. Bagayoko and J. Callaway, Phys Rev. B, **28**, 5419 (1983).

[90] M. Shimizu and A. Kunibara, Phys. Lett. **100A**, 218 (1984).

[91] F.J. Pinski, J. Staunton, B.L. Gyorffy, D.D. Johnson and G.M. Stocks, Phys. Rev. Lett. **56**, 2096 (1986).

[92] M. Uhl, L.M. Sandratskii and J. Kubler, J. Magn. Magn. Mat. **103**, 314 (1992).

[93] O.K. Andersen, J. Madsen, U.K. Poulsen, O. Jepsen and J. Kollar, Physica, **86-88B**, 249 (1977).

[94] J. Kubler, Phys. Lett. **81A**, 81 (1981).

[95] A.R. Williams, V.L. Moruzzi, C.D. Gelatt Jr., J. Kubler and K. Schwartz, J. Appl. Phys. **53**, 2019 (1982).

[96] H. Hasegawa and D.G. Pettifor, Phys. Rev. Lett. **50**, 130 (1983).

[97] C.S. Wang, B.M. Klein and H. Krakauer, Phys. Rev. Lett. **54**, 1852 (1985).

[98] D. Guenzburger, and D.E. Ellis Phys. Rev. B. **51**, 12519 (1995).

[99] E.G. Moroni, and T. Jarlborg, Phys. Rev. B, **41**, 9600 (1990).

[100] P. Entel, E. Hoffmann, P. Mohn, K. Schwarz, and V.L. Moruzzi, Phys. Rev. B, **47**, 8760 (1993).

[101] E. Hoffmann, H. Herper, P. Entel, S.G. Mishra, P. Mohn, and K. Schwarz, Phys. Rev. B, **47**, 5589 (1993).

[102] S. Lipinski, J. Magn. Magn. Mat. **140-144**, 233 (1995).

[103] I.A. Abrikosov, O. Erikson, P. Söderlind, H.L. Skriver, and B. Johansson, Phys. Rev. B, **51**, 1058 (1995).

[104] M. Dubé, P.R.L. Heron, and D.G. Rancourt, J. Magn. Magn. Mat. **147**, 122 (1995).

[105] M.-Z. Dang, M. Dubé, and D.G. Rancourt, J. Magn. Magn. Mat. **147**, 133 (1995).

[106] D.G. Rancourt, M. Dubé, and P.R.L. Heron, J. Magn. Magn. Mat. **125**, 39 (1993).

[107] D.G. Rancourt, P. Hargraves, G. Lamarche, and R.A. Dunlap, J. Magn. Magn. Mat. **87**, 71 (1990).

[108] J.Y. Ping, D.G. Rancourt, and R.A. Dunlap, J. Magn. Magn. Mat. **103**, 285 (1992); and references therein.

[109] D.G. Rancourt, Nucl. Instr. Meth. Phys. Res. B. **44**, 199 (1989).

[110] D.G. Rancourt, in *Mössbauer Spectroscopy Applied to Magnetism and Materials Science*, edited by G.J. Long and F. Grandjean (Plenum, New York, 1996), in press.

[111] M.M. Abd-Elmeguid, U. Hobuss, H. Micklitz, B. Huck, and J. Hesse, Phys. Rev. B, **35**, 4796 (1987).

[112] F. Van der Woude, and A.J. Dekker, Phys. Stat. Sol. **9**, 775 (1965).

[113] D.G. Rancourt, S.R. Julian, and J.M. Daniels, J. Magn. Magn. Mat. **51**, 83 (1985).

[114] D.G. Rancourt, Hyp. Int. **40**, 183 (1988).

[115] D.G. Rancourt, and J.Y. Ping, Nucl. Instr. Meth. Phys. Res. B, **58**, 85 (1991).

[116] L. Dou, R.J.W. Hodgson, and D.G. Rancourt, Nucl. Instr. Meth. Phys. Res. B, **100**, 511 (1995).

[117] G.F. Knoll, in *Radiation Detection and Measurement*, (John Wiley & Sons, New York, 1989).

[118] G. Longworth, in *Mössbauer Spectroscopy Applied to Inorganic Chemistry*, edited by G.J. Long, (Plenum, New York, 1987), vol. 2, p. 289.

[119] N. Blaes, H. Fischer, and U. Gonser, Nucl. Instr. Meth. Phys. Res. B, **9**, 201 (1985).

[120] K. Binder, in *Statics and Dynamics of Alloy Phase Transformations*, edited by P.E.A. Turchi and A. Gonis (Plenum, New York, 1994) p.467.

[121] K. Binder and D.W. Heermann, in *Monte Carlo Simulation in Statistical Physics* (Springer Verlag, Berlin, 1988).

[122] C.E. Johnson et al., Proc. Phys. Soc. **81**, 1079 (1963).

[123] M. Shiga and Y. Nakamura, J. Magn. Magn. Mat. **40**, 319 (1984).

[124] J.B. Müuller and J. Hesse, Z. Phys. B **54**, 43 (1983).

[125] P. Panissod, J. Durand and J.I. Budnick, Nucl. Instr. Meth. **199**, 99 (1982)

[126] B. Fultz, in *Mössbauer Spectroscopy Applied to Materials and Magnetism.* edited by G.J. Long and F. Grandjean, (Plenum, 1993).

[127] M.Z. Dang and D.G. Rancourt, Proceedings of the Int. Conf. Appl. Mössbauer Effect 1996.

[128] M.-Z. Dang, thesis, University of Ottawa, 1995.

[129] L. Dou, M.-Z. Dang, R.J.W. Hodgson and D.G. Rancourt, "Ground State Hyperfine Field Distribution and Spin Structure of $Fe_{65}Ni_{35}$ Invar" in the process of writing.