Processing of Randomly Obtained Seismic Data

by

Youshun Sun

M.S. Electrical Engineering, Wright State University, 1996

B.S. Biomedical Engineering, Tsinghua University, 1992

Submitted to the Department of Earth, Atmospheric, and Planetary Sciences

in partial fulfillment of the requirements for the degree of

Master of Science

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

May 5, 1998

Author ....

Department of Earth, Atmospheric, and Planetary Sciences

May 5, 1998

Certified by ....

Nafi Toksöz

Professor, Department of Earth, Atmospheric, and Planetary Sciences

Thesis Supervisor

Certified by ....

Thomas H. Jordan

Professor, Department of Earth, Atmospheric, and Planetary Sciences

Geosystems Supervisor

Accepted by ....

Thomas H. Jordan

Professor, Department of Earth, Atmospheric, and Planetary Sciences

Department Head

Processing of Randomly Obtained Seismic Data

by

Youshun Sun

Submitted to the Department of Earth, Atmospheric, and Planetary Sciences (EAPS)
on May 5, 1998 in partial fulfillment of the
requirements for the degree of
Master of Science

Abstract

This thesis involves processing randomly obtained seismic data. The main contributions of this research are: the improvement in reconstructing the evenly spaced seismic signal from randomly sampled data; the significant reduction in the sampling frequency (Nyquist frequency or below). To achieve these objectives, we have made use of the Shannon sampling theorem and compared both linear and higher-order random sampling and reconstruction techniques. Both techniques can successfully recover the original signal from the randomly obtained data at very high sampling frequencies. When the average sampling frequency is only at Nyquist frequency or below, the linear technique performs poorly on the signal recovery. Only the higher-order technique is a good choice for reconstructing the original signal with average sampling rates at or below the Nyquist rate.

The higher-order technique has been used successfully in the time domain. Likewise, this technique can also be applied to the space domain. We incorporate this technique to reduce the seismic data volume, the number of geophones being used in the field, and to improve the degree of freedom in arranging geophones.

Thesis supervisor: Nafi Toksöz
Title: Professor of Geophysics at EAPS

Acknowledgments

I would like to thank, first of all, Professor Tom Jordan whose encouragement throughout the time this work was being produced, was the most inspiring reason for my persistence and hard work. I also wish to express my deepest gratitude to Tom for his hard work on behalf of the Geosytems program, and for all the academic and financial help I received from him when I was really in need of them.

I am very grateful to Professor Nafi Toksöz, my thesis supervisor, for his openness, encouragement, financial support, and good advice when I needed them the most.

Special thanks to Professor Jochem Marotzke, Dr. Scott Sewell, Ms. Stacey Frangos and Dr. Yunpeng Wang for their help with our class and administration.

I would also like to thank my office-mates for their considerance and coorperation.

Finally, I thank my family for their love, encouragement, and patience. This thesis, like all of my other accomplishments, would not have been possible without their support.

Contents

Chapter 1

Introduction

Seismic data obtained in the field are oversampled by geophones and therefore needed to be downsampled at nonuniform (random) frequencies to reduce the data capacity. Also, while the geophones are geographically needed to be arranged randomly, it's needed to develop the techniques to process those randomly obtained seismic data. The goal of this thesis is to develop techniques which reconstruct the original signal from the randomly obtained time-domain or space-domain data. This chapter motivates this work and describes the organization of the thesis. First the problem of processing randomly obtained data is described. Then, the general approach taken to solve this problem is introduced. Finally, an outline of the thesis is provided.

1.1 Problem Description

In many cases, seismic data is oversampled and requires a big capacity of disk storage. In order to reduce the data volume without distorting the authentic information, we can resample the data with random frequencies. Those random frequencies can be uniformly distributed, Gaussian distributed, etc. The average of the random frequencies can also be the Nyquist frequency or below. Theoretically, the sampled data cannot be completely recovered if the lowest sampling frequency is below the Nyquist rate [7]. However, if the average sampling frequency is chosen to be the Nyquist rate, the original signal can be reconstructed very well from the randomly sampled data even though some sampling frequencies are below the Nyquist rate. It's promising to develop reconstruction techniques that recover the original signal from the random samples and make the error between the original signal and the reconstructed signal small.

Moreover, the reconstruction techniques can be used to recover the randomly sampled space-domain signal. In the field, geophones are usually arranged as evenly spaced as possible. We can place geophones randomly or as needed if we can apply the reconstruction techniques to process the randomly sampled space-domain data. The number of geophones can also be reduced without losing significant information after these reconstruction techniques are developed.

This thesis attempts to develop techniques which reconstruct the original signal from the randomly obtained samples.

1.2 Proposed Solution

The techniques studied in this thesis are motivated by the fact that the randomly sampled data can be satisfactorily reconstructed if the average sampling frequency is the Nyquist rate or below. We consider the linear reconstruction technique and the higher-order reconstruction technique in this thesis.

Conceptually, signal reconstruction requires that the lowest sampling frequency is at least the Nyquist sampling rate to completely recover the original signal. This is the so called

Nyquist Sampling theorem. According to this theorem, if the sampling frequency is smaller than the Nyquist rate, the Fourier spectrum of the sampled signal is going to overlap (aliasing). It's impossible to recover the original signal from an overlapped Fourier response.

However, in practice, if the Fourier spectrum of the original is small enough over a certain frequency, we can specify that frequency is the highest signal frequency by applying a low pass filter [14]. By doing so, we can still be capable of recovering the original signal because the Fourier spectrum is not corrupted.

This thesis is an experimental study of the processing of the randomly obtained seismic data. We study the efficiency and accuracy of both the linear and the higher-order reconstruction techniques by theoretical analysis and computer simulation. We also study both the time-domain and space-domain signal reconstruction by reconstructing the real seismic data from the field.

1.3 Thesis Outline

The body of the thesis is divided into five chapters:

Chapter 2 covers the background information useful for the discussion of signal reconstruction presented later in the thesis. The Shannon reconstruction formula is the classical formula in signal processing. All the signal reconstruction techniques are developed based on this formula.

Chapter 3 presents a detailed description of both the linear reconstruction technique and the higher-order reconstruction technique. Procedures to perform reconstruction are also described.

Chapter 4 presents the computer simulations of both the linear and the higher-order reconstruction techniques. The average sampling frequencies are chosen to be as low as the Nyquist rate and three times as high as the Nyquist rate. The simulations show that the higher-order reconstruction technique is a better method to recover the original signals from the random samples with the average sampling frequency at the Nyquist rate or below.

Chapter 5 presents an experimental study of the effectiveness of the higher-order reconstruction technique. The original signal was resampled with low average sampling rate to reduce the requirement for data storage.

Chapter 6 concludes the thesis with a summary and directions for future work. The techniques and experiments that are presented in the thesis have left many open issues. It is hoped that this work will stimulate further investigations which may address these issues.

Chapter 2

Background

The reconstruction of original signal from a uniformly sampled data set is quite well known: the Shannon reconstruction formula [4] is given by,

$$s(t) = \sum_{-\infty}^{\infty} s(kT) Sinc(t - kT) \qquad (1)$$

where T is the sampling period. In random sampling, t of equation (1) is not an integer multiple of the average sampling rate. The reconstruction formula in this case is derived by defining a deviation of the sampling point from the sampling time corresponding to the average sampling interval. Thus it is no longer an impulse response of one ideal low-pass filter; the filter response function is a modulated one. This is represented in equation (2).

$$s(t) = \sum_{-\infty}^{\infty} s(t_k Sinc(t - t_k)) \qquad (2)$$

where t_k is the random k-th sampling instant. However, reconstruction from a random sampling set is not possible for any arbitrary deviation from the average sampling set. According to the Nyquist sampling theorem [5], the lowest sampling frequency should at least double the highest frequency of the sampled signal to avoid aliasing. The Nyquist frequency, therefore, means the sampling frequency which is exactly twice as big as the highest signal frequency. In the other words, any sampling frequency below the Nyquist frequency will not be able to recover the original signal completely. In this thesis, we are trying to reconstruct (approximate) the original signal from the data randomly sampled with the Nyquist frequency or below.

Chapter 3

Reconstruction Techniques

The signal can be reconstructed from randomly sampled data by using the Shannon reconstruction formula given by equation (2), and the sampling instants are known. However, a better and robust reconstruction procedure can be obtained by using many reconstruction approaches. Some reconstruction methods were discussed by Wunsch [22] and Goff and Jordan [23] under different data conditions. In this thesis, we only introduce and focus on two techniques: the linear reconstruction method and the higher-order reconstruction method.

3.1 The Linear Reconstruction Method

The linear reconstruction method is based upon the fluency model [1]. The fluency model clarifies the relationship between continuous-time and discrete-time systems by utilizing a group of piecewise polynomial functions called fluency digital/analog functions (common sampling function). By selecting a fluency sampling function of an appropriate class according to the characteristics of the signals in question, we can interpolate and accurately approximate continuous-time signals from sampled data. In order to discuss the linear reconstruction technique, we have to discuss the fluency model and fluency sampling function first.

3.1.1 The Fluency Model and the Fluency Sampling Function

In the fluency model, signals are categorized in terms of signal space. A signal space, ${}^{m}S$, is defined to be a space composed of piecewise polynomial functions of degree (m-1) with a parameter of (m-2) times continuous differentiability. It has been proven that a signal space of (m-1)th order piecewise polynomial becomes equivalent to those of staircase functions when m = 1, and to Fourier exponential functions when m goes to infinity [1]. The fluency model, describes the relationship between these signals belonging to ${}^{m}S$ and discrete-time signals, by introducing a group of functions called fluency sampling functions.

Fluency digital/analog conversion functions (common sampling function) are piecewise polynomial functions with degree of (m-1) and a parameter of (m-2) times continuous differentiability. Figure 1-3 show class m=2, m=3, and m=∞ sampling functions, respectively.

Any signal belonging to ${}^{m}S$ can be expressed in terms of linear combinations of discrete-time signal and class m fluency sampling function. From the view of fluency model, staircase signals of signal space ${}^{1}S$, can be represented by a linear combination of step functions (class m=1 sampling function). Likewise, Fourier exponential function space class can be represented by Sinc.

There are several advantages in using the fluency model to describe signals. First, we are able to deal with a variety of signals of different signal space. In real world, signals are

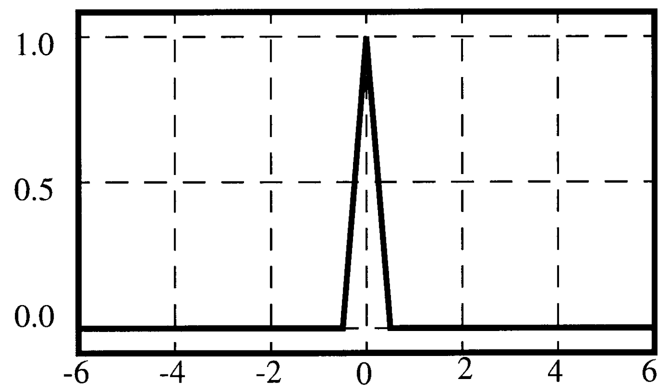

Figure 1: class m=2 fluency sampling function

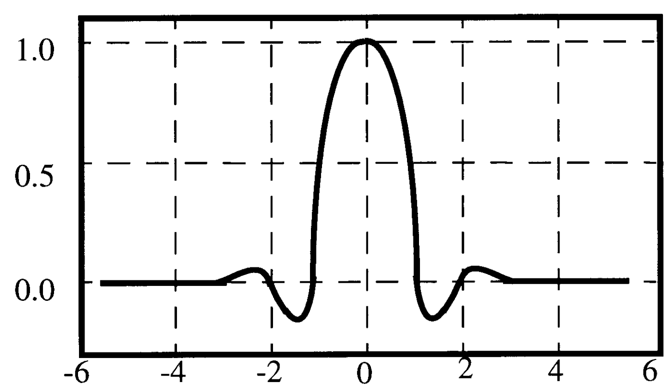

Figure 2: class m=3 fluency sampling function

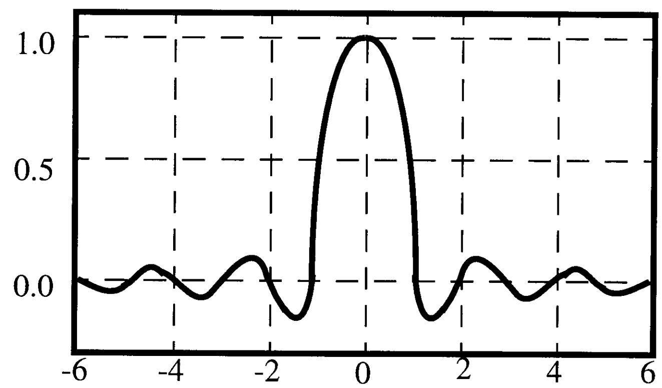

Figure 3: class $m=\infty$ fluency sampling function

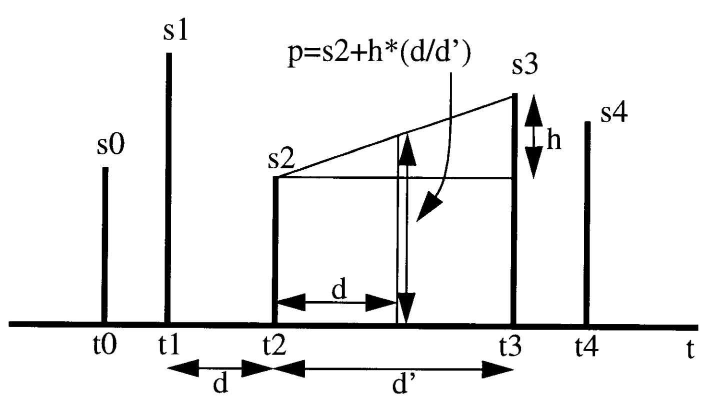

Figure 4: obtaining pseudo samples for interpolating values between t2 and t3

time-varying, that is they continuously change their forms along with time. At some point, a signal may be smooth and belong to ${}^4S$, whereas at another point, it may be rugged and belong to ${}^1S$. With fluency sampling functions, we are able to correlate these time-varying signals by changing their classes, while with the conventional Fourier model, we need to assume that signals are only composed of polynomials that belong to ${}^{\infty}S$.

Second, fluency functions provide us with the means of piecewise polynomial approximation, without the need for heavy calculation. Conventionally, approximation by piecewise polynomial methods are done by solving coefficients for B-spline basis. However, this requires solving linear equations. With fluency sampling functions, sample values only need to be convoluted.

Third, amplitude of fluency sampling functions of lower classes ($3<=m<\infty$) attenuate exponentially. Thus, truncation is possible with small approximation error. This is useful for implementation, for it is obviously not realistic to deal with functions having infinite support. The truncation error will be discussed further in the following subsection.

3.1.2 The Fluency Function

In this section, we will discuss briefly the mathematics of fluency functions [8].

First let the sample points on the time axis be $\{t_k\}_{k=-\infty}^{\infty}$ where $t_k :=$ kh, $k = 0, \pm 1, \pm 2, \ldots$ Here, h is the sampling interval. Then define the sample value at a sample point t_k to be v_k. Thus the relation between continuous-time signal s and discrete-time ν is

$$s(t_k) = v_k,\ k = 0, \pm 1, \pm 2, \ldots \qquad (3)$$

Under these conditions, the sampling basis for ^{m}S is defined as the system of functions $\{{}^{m}_{[s]}\varphi_k\}_{k=-\infty}^{\infty}$ that satisfies

$$s(t) = \sum_{k=-\infty}^{\infty} v_k \cdot {}^{m}_{[s]}\varphi_k(t), \forall s \in {}^{m}S \tag{4}$$

The equation for the sampling basis is derived as

$$ {}^{m}_{[s]}\varphi_k(t) = \sum_{l=-\infty}^{\infty} {}^{m}\beta(l-k) {}^{m}_{[b]}\varphi_l(t) \quad k = 0, \pm 1, \pm 2, \ldots \tag{5}$$

The function, $\{{}^{m}_{[b]}\varphi_l\}_{l=-\infty}^{\infty}$, is called the B-spline basis of degree (m-1) with a parameter that is (m-2) times continuously differentiable. They are shift-invariant, symmetric functions.

The coefficients $\{{}^{m}\beta(p)\}_{p=-\infty}^{\infty}$ are derived as

$$ {}^{m}\beta(p) = h \int_{-1/2h}^{1/2h} {}^{m}_{f}B(f) e^{j2\pi fph} df \tag{6}$$

and

$$ {}^{m}_{f}B(f) = \frac{1}{\sum_{q=-\lceil (m-1)/2 \rceil}^{\lceil (m-1)/2 \rceil} {}^{m}_{[b]}\varphi_0(qh) e^{-j2\pi fqh}} \tag{7}$$

where $\lceil \alpha \rceil$ denotes the greatest integer not exceeding α.

3.1.3 Truncation Error of Fluency Sampling Functions

When v_k, $k = 0, \pm 1, \pm 2, \ldots$ are the sample values for $x \in L_2(R)$, the least squares approximation $s_0 \in {}^m S$ for $x \in L_2(R)$ is represented by

$$s_0 = \sum_{-\infty}^{\infty} v_k \, {}_{[s]}^{m}\varphi_0(t - kh). \tag{8}$$

x belongs to $L_2(R)$ when $\int_{-\infty}^{\infty} |x(t)|^2 dt < +\infty$.

The truncation error E between the proposed approximation $\tilde{s}_0$ and the least squares approximation s_0 for $x \in L_2(R)$ is defined as follows:

$$E^2 = \frac{\|\tilde{s}_0 - s_0\|_{L_2}}{\|x\|_{L_2}} = \frac{\left[\int_{-\infty}^{\infty} |\tilde{s}_0(t) - s_0(t)|^2 dt\right]^{\frac{1}{2}}}{\left[\int_{-\infty}^{\infty} |x(t)|^2 dt\right]^{\frac{1}{2}}} \tag{9}$$

When the sampling interval is normalized to 1, it is known that for m=3 class, truncation at ± 5 will keep the truncation error within only -60dB. For, m=4, truncation at ± 7, will give the same precision [3]. E is expressed in the form of $20 \log_{10} E(dB)$.

3.1.4 Random Sampling and Interpolation

As shown above, fluency sampling functions can be truncated with small amount of approximation error. Thus, if fluency sampling functions are applied somehow to randomly obtained discrete-time signals, original signals may be approximated with accuracy without the need for solving linear equations. Here, we propose a method for randomly sampling and interpolation based on fluency sampling functions. Although this technique is valid for fluency functions of all classes, for simplicity, we explain by using class m=3.

3.1.4.1 Random Sampling and Interpolation

To interpolate randomly sampled signals, sampling functions are convoluted with sample values. However, when it comes to interpolating randomly sampled signals, simply dilating the sampling functions so that they intersect the time axis at the place of sample points will not give good interpolation. This is because information concerning the lengths of intervals between randomly sampled data are not reflected. In other words, upon interpolation, we would like the sample points that are further away in time to give smaller influence compared with the points that are nearer. Hence, we make approximations for each randomly sampled data, the values that they would take if they were rearranged in randomly distributed interval. We call these approximated samples pseudo samples.

3.1.4.2 Pseudo Samples

As mentioned before, fluency functions can be truncated without resulting in large approximation error. If we truncate at ± 5, class m=3 function will intersect the time axis 4 times at either side of the origin, which means only a total of 8 sample points will be needed for interpolation (i.e. convolution). Accordingly, only 8 pseudo samples are needed. These 8 pseudo samples, instead of the original randomly sampled data, are convoluted with the sampling function.

In Figure 4, s0, s1, s2, s3, and s4 represent sampled signals at times t0, t1, t2, t3 and t4, respectively. The figure shows how to obtain one of pseudo samples needed to interpolate values between t2 and t3. The pseudo sample value, p, is obtained by a linear approximation between s2 and s3 at time t2+d. The distance between t1 and t2, called d, is the base interval.

We have shown the method for obtaining one pseudo sample. The rest are calculated at times t2+2*d, t2+3*d, t2+4*d, t1-d, t1-2*d, t1-3*d, and t1-4*d, using different pairs of sampled signal. For example, to obtain pseudo samples at times t2+2*d, and t1-2*d, linear approximations are done between s2 and s4, and between s0 and s1, respectively.

As it could be easily understood, the base interval changes according to the distance between the intervals of sampled signals. The sampling function is dilated according to the length of the base interval so that it will across the time axis at points where pseudo samples exist.

3.1.4.3 Random Sampling

Given this method for random interpolation, points that best approximate the original signal are extracted as feature points. It is known that midpoints of two adjacent inflection points give the best approximations. This is because fluency sampling functions are designed in such way that their points of inflection come half way in between two adjacent sample points. The selection of signal's feature points, is considered random sampling.

From the above description of the reconstruction method, we can see that linear approximations are done to interpolate the randomly sampled data. Therefore this technique is called the linear reconstruction method. The simulation and experiment based on this method will be discussed further in the following sections. In order to compare the reconstruction quality, let's introduce the other reconstruction technique discussed in this thesis: the higher-order reconstruction method.

3.2 The Higher-order Reconstruction Method

Equation (1) and equation (2) show mathematically how to recover original signals from randomly sampled data. In equation (2), since t_k is the random k-th sampling instant, reconstruction from a random sampling set is possible for any arbitrary deviation from the average sampling set. The bounds on the nonuniformity, allowed from an average sampling rate in a particular sampling set, has been established by various methods as discussed in many papers [4]. The bounds utilized here are derived by the restrictions introduced for a one-to-one mapping from the sampling set to the original signal and vice versa. There are two bounds on the permissible nonuniformity based on the necessary and sufficient condition of the mapping in the bandwidth corresponding to the average sampling rate [4]. The two bounds are given by the following equations:

$$t_k - kT \leq T/\pi \qquad (10)$$

$$\sum_{0}^{N-1}(t_k - kT) \leq 3T^2/\pi^2 \qquad (11)$$

where T is the sampling interval. Now, the samples in particular sampling set can be considered as samples taken at the average sampling rate, corresponding to that sampling set. Each sampling section is reconstructed separately, therefore the stability of all the sampling sets together with respect to a single average sampling rate does not arise. However, the continuity of the consecutive sampling sections has to be maintained. To fulfill this requirement the average sampling rates of the consecutive sampling sets must satisfy condition of equation (10).

The criterion chosen to increase or decrease the sampling rate is the first derivative of the signal. The sampling rate is increased as the normalized first derivative increases. To implement this criterion, first derivative is approximated by a first difference or a bilinear approximation. The signals sampled at the Nyquist rate has unity normalized first difference, given by equation (12).

$$ds(k) = \frac{s(k) - s(k-1)}{|s(k)| + |s(k-1)| + 0.001} \tag{12}$$

The denominator used in this equation is provided an offset to avoid numerical instability. The first difference given by equation (12) is sensitive to random noise; a three point averager given by equation (13) is operated on the signal before estimating the first difference.

$$s_{av}(k) = \frac{0.5s(k-1) + s(k) + 0.5s(k+1)}{2} \tag{13}$$

The discarding basis can be derived from the knowledge of the first derivative of the signal. In this discrete sampling process the sampling is carried out at a sampling rate corresponding to the maximum frequency expected in the signal and validation of each sample is done by checking whether the first derivative is above a threshold level. If the first derivative is below the threshold level the samples continue to be discarded unless it meets the bounds of the sampling set. The threshold value is chosen according to the oversampling requirement.

3.2.1 Reconstruction by Low-pass Filtering

The signal can be reconstructed from a randomly sampled data by using the Shannon reconstruction formula given by equation (2), when the sampling instants are known. However, a better and robust reconstruction procedure can be obtained by using a practical Finite Impulse Response (FIR) low-pass filter [7].

3.2.1.1 Design of A Low Pass Filter Using Kaiser Window

The Kaiser window used here is given by

$$w(k) = I_0\left[\beta\left(1-(k/N)^2\right)^{0.5}\right] \Big/ I_0(\beta) \qquad |k| \le N \qquad (14)$$

$$= 0 \qquad |k| > N \qquad (15)$$

Where I_0 is the zeroth-order modified Bessel function of the first kind

$$I_0(x) = 1 + \sum_{m=1}^{M}\left[(x/2)^m / m!\right]^2 \qquad (16)$$

and β is the scaling factor. However, the impulse response of the low pass filter is time-varying, depending on the number of samples discarded. In the case of sampling with digital means with discarding basis, the base sampling rate is the highest corresponding to the highest frequency; the discarding of samples reduces the sampling rate only at the integer multiples of the base rate. Therefore, the requirement of a time varying impulse response is met with the help of an interpolator. The well known formula of an interpolator, having an integer factor L is given by [7],

$$y(m) = \sum_{k=-\infty}^{\infty} h(m-k)s(k/L) \quad \text{if k is real} \qquad (17)$$

$$y(m) = \sum_{r=-\infty}^{\infty} h(m - rL)s(r) \tag{18}$$

3.2.2 Implementation

For practical implementation, the algorithm for selecting a valid sample is given below following Ghosh and Dutta [4]:

I. input the first sample

II. initialize the average sampling rate equal to base rate

III. initialize the bound according to the average sampling rate

IV. take a new sample and estimate the normalized first derivative with the help of equation (12)

V. if the first derivative is below the threshold discard the sample

VI. calculate the new sampling rate

VII. calculate the new bound according to the new average rate

VIII. calculate the sum-of-deviation

IX. check whether sum-of-deviation agrees with the bound

X. if not close the section and compute the average rate

XI. compute average of the average sampling rates corresponding to all the previous average rate

XII. compute the average rate of the new section according to the bound of equation (10)

XIII. repeat from step (III)

Actually, the sum-of-deviation bound can be obtained through a look-up table.

3.2.3 Reconstruction

For reconstructing the original signal the discarded samples are estimated by interpolation, using the current sample which contains the number of discarded samples and previous (N-1) samples, where N is the selected filter length. The filter has a cutoff frequency π/L and gain L. In order to avoid aliasing the original signal has to be sampled at least L times the Nyquist rate. In the sampling process implemented through discarding basis this condition is satisfied automatically. As the number of zeros at different portions of the sampled data is variable, the modified impulse response estimated accordingly is given by,

$$h_L(k) = L.(\pi/L).1/\pi.Sinc(k\pi/L) \quad (19)$$

$$= Sinc(k\pi/L) \quad (20)$$

Unlike the linear method, this reconstruction algorithm requires more than two random samples to interpolate one even sample, and is therefore called higher-order reconstruction method and the actual steps are given below:

I. take the first sample
II. input the next sample
III. obtain the information of the number of discarded prior to the current sample
IV. choose the interpolation factor as equal to the number of discarded samples and estimate the impulse response function
V. compute the interpolated samples according to equation (17)

By far, we have introduced two methods of reconstructing the original signal from the randomly sampled data. In the following sections, we will show the computer simulations and real seismic data processing.

Chapter 4

The Computer Simulation

In order to test both the linear reconstruction and higher-order reconstruction methods, we construct a signal s(t) consisting of 10 frequencies:

$$s(t) = \sum_{k=1}^{10} \sin(2\pi kft) \qquad (21)$$

where the base frequency $f = 1000Hz$. The highest frequency, therefore, is 10,000 Hz, and the Nyquist frequency is 20,000 Hz. We will resample s(t) with randomly distributed intervals. Two average sampling frequencies will be used in this thesis. One is the Nyquist frequency, and the other is a sampling frequency which is three times as big as the Nyquist frequency. The distribution of the random sampling frequencies is assumed to be a uniform distribution in this paper. After resampling the original signal, both linear and higher-order reconstruction techniques will be applied to recover the original signal from the randomly obtained data.

4.1 The Linear Reconstruction Method

As we mentioned before, by selecting a fluency sampling function of an approximate class according to the characteristics of the signals in question, we can interpolate and accurately approximate continuous-time signals from the sampled data. By default, the sampling function used in this paper is class m=3. This reconstruction technique can recover the original signal very well when the averaging sampling frequency is high.

4.1.1 Linear Reconstruction with High Average Sampling Rate

The original signal given in equation (21) is shown is Figure 5(A). The Fourier spectrum of the original signal is given in Figure 5(B). In Figure 5(B), we can see 10 distinct frequencies from 1k to 10k Hz. The Nyquist frequency is 20k Hz. When the average sampling frequency is 60k Hz, the uniformly distributed histogram of the sampling frequencies is shown in Figure 6. We assume the sampling frequencies are uniformly distributed from 40k to 80k.

The randomly sampled data are shown in Figure 7. We can see clearly from Figure 7 that the sampling intervals are not evenly spaced. Random samples exactly match the original signal.

When the linear reconstruction technique is applied, the evenly spaced samples are obtained from the random samples. In Figure 8, the even samples fit the original signal very well. After the even samples are obtained, the uniform interpolation can be readily applied to reconstruct the original signal. Figure 9 shows that the recovered signal fits the original signal with very small error. Apparently the linear reconstruction technique is a very good method for recovering the original signal if the average sampling frequency is

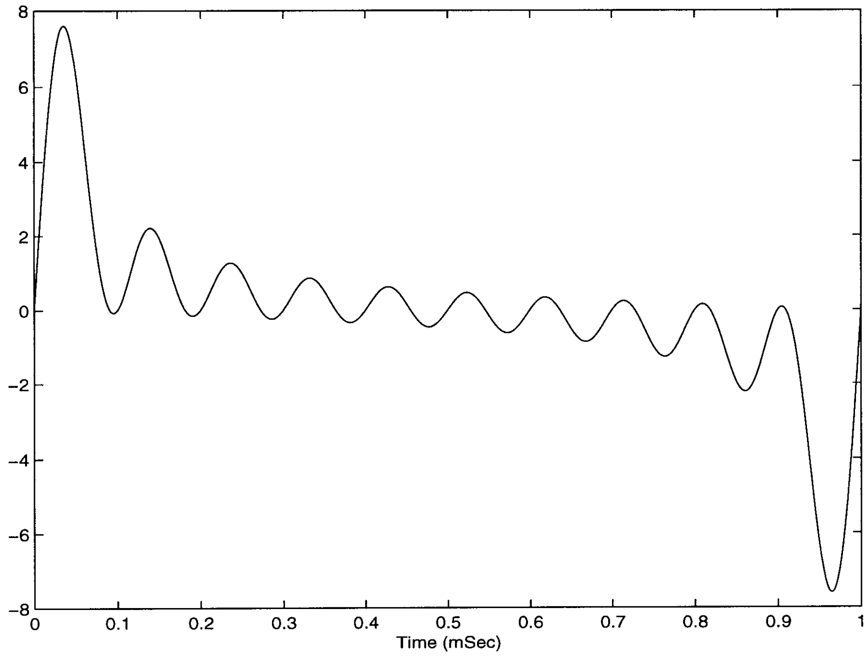

Figure 5(A): the original signal s(t)

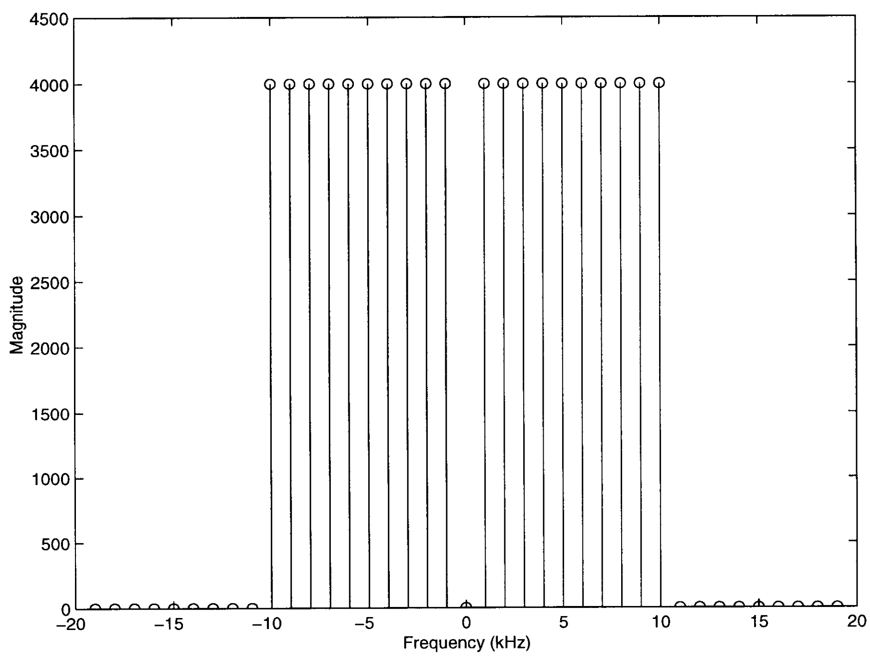

Figure 5(B): the Fourier spectrum of s(t)

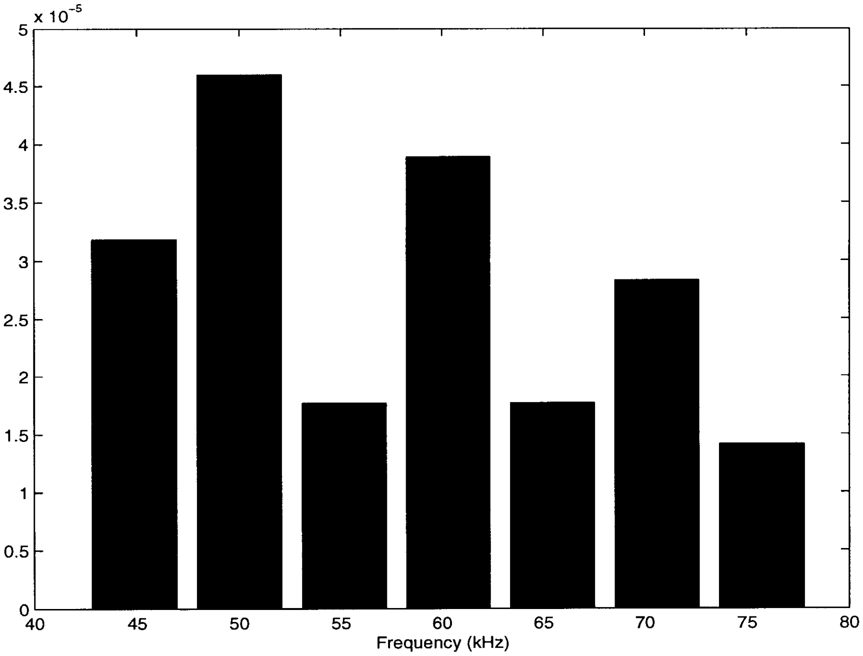

Figure 6: the uniformly distributed histogram of the sampling frequencies

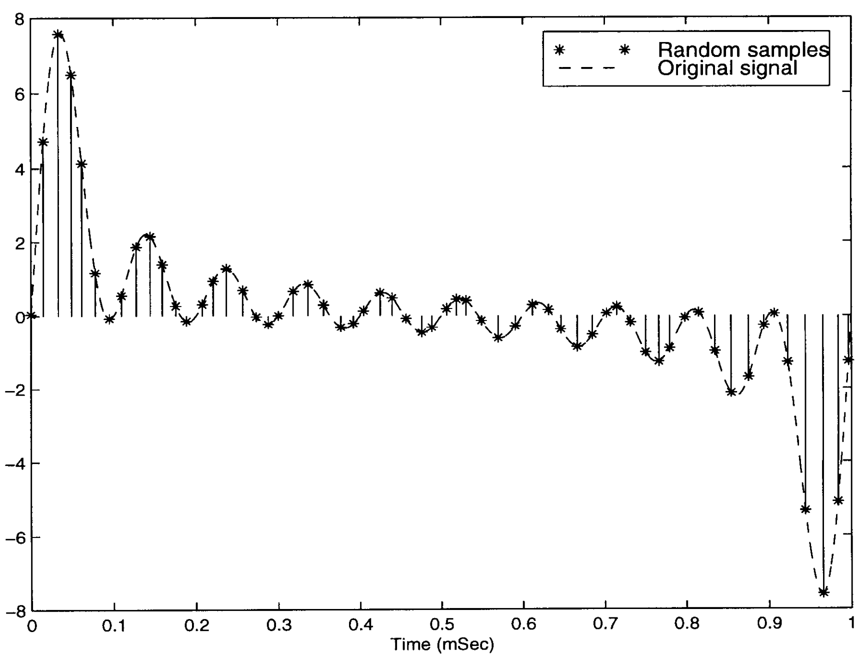

Figure 7: the randomly sampled data

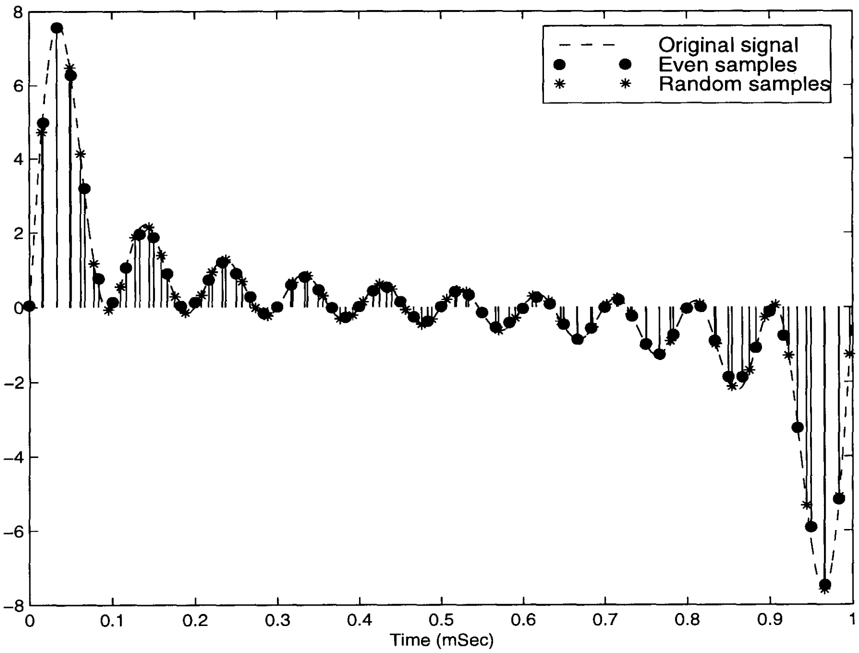

Figure 8: the recovered even samples

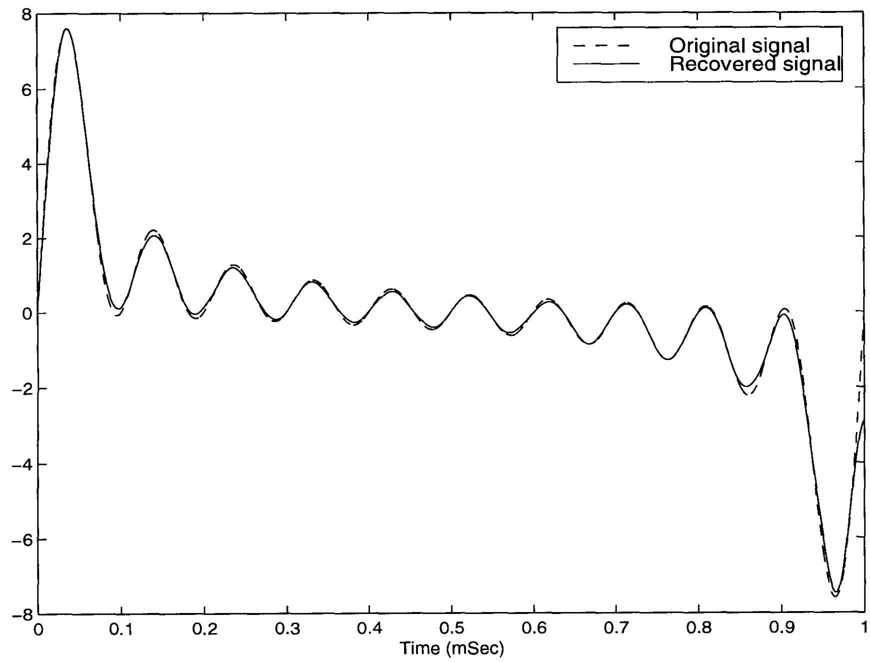

Figure 9: the recovered signal

high enough. The question is: what if the average sampling rate is low? For example, what if the average sampling rate is equal to or below the Nyquist sampling rate?

4.1.2 Linear Reconstruction with Low Average Sampling Rate

Let's reduce the average sampling rate to the Nyquist rate. By doing so, we can reduce the capacity of data storage if the reconstruction is still successful.

Figure 10 shows that the random samples match the original signal well when the average sampling frequency is exactly the Nyquist sampling frequency. However, since the average sampling frequency is too low, the recovered evenly spaced samples are located correctly with their amplitudes mismatched. Figure 11 & 12 show that the linear reconstruction does a poor job when the average sampling frequency is the Nyquist sampling rate.

4.2 The Higher-order Reconstruction Method

As it has been seen, the linear interpolation is not a good choice when the original signal is sampled with a low average sampling rate. This is systematically determined by the linear technique itself. As we know, the linear reconstruction is very easy to implement and requires less calculation compared to the higher-order reconstruction technique. The linear method can only give an approximation which linearly interpolates the original sample between its two adjacent random samples. This reconstruction turns out to be unsuccessful when the sampling intervals are too large, i.e. the sampling frequencies are too low.

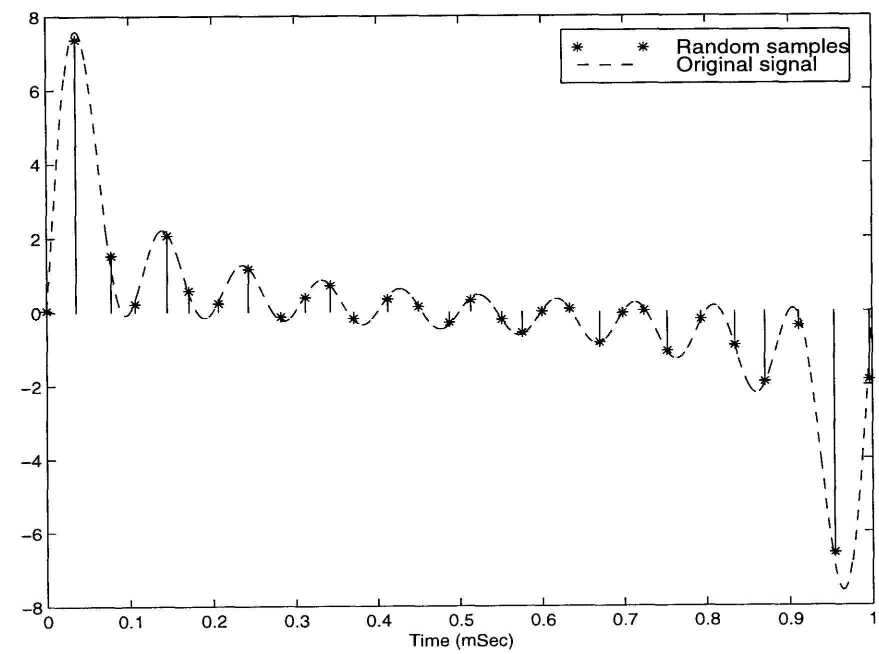

Figure 10: the random samples

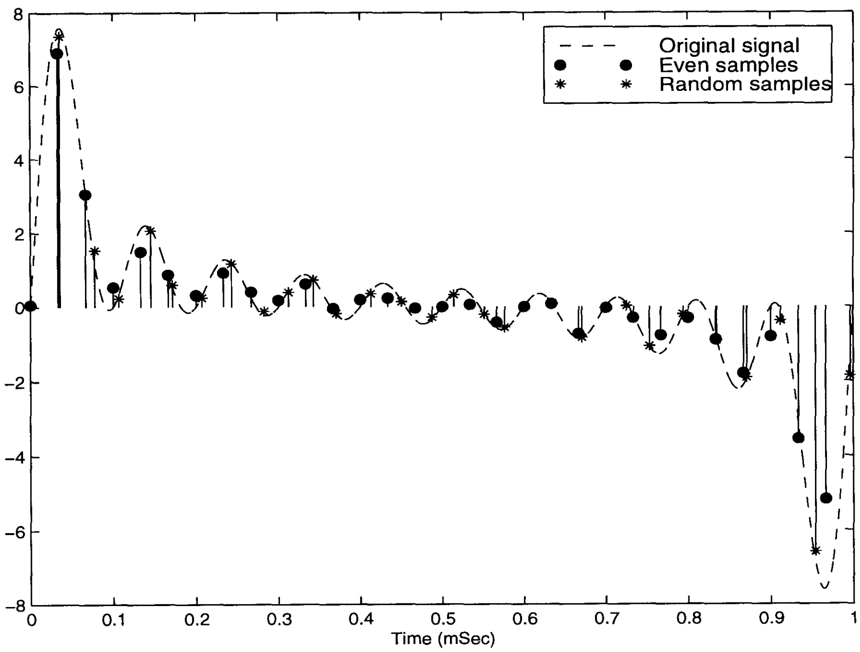

Figure 11: the recovered even samples

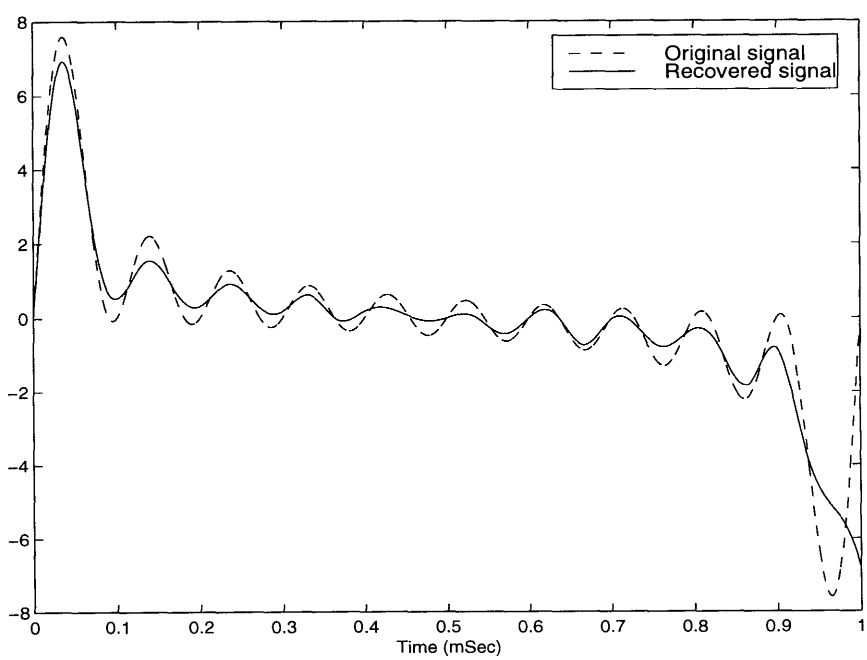

Figure 12: the recovered signal

4.2.1 Higher-order Method with High Average Sampling Rate

The higher-order technique, however, can do a better job than the linear one. Figure 13-15 show that the higher-order reconstruction can reconstruct perfectly the original signal from the randomly sampled data. Therefore, it will be more interesting for us to take a look at the case when the average sampling frequency is only the Nyquist sampling rate.

4.2.2 Higher-order Method with Low Average Sampling Rate

Unlike the linear reconstruction technique, the higher-order reconstruction method can recover the original signal from a low average sampling rate very well. Figure 16-17 show the similar graphs we discussed before. In Figure 18, we still see a very good reconstruction of the original signal from the data sampled with the average Nyquist rate.

Due to the outstanding performance of the higher-order reconstruction technique, we will only utilize the higher-order method in our tests with real seismic data.

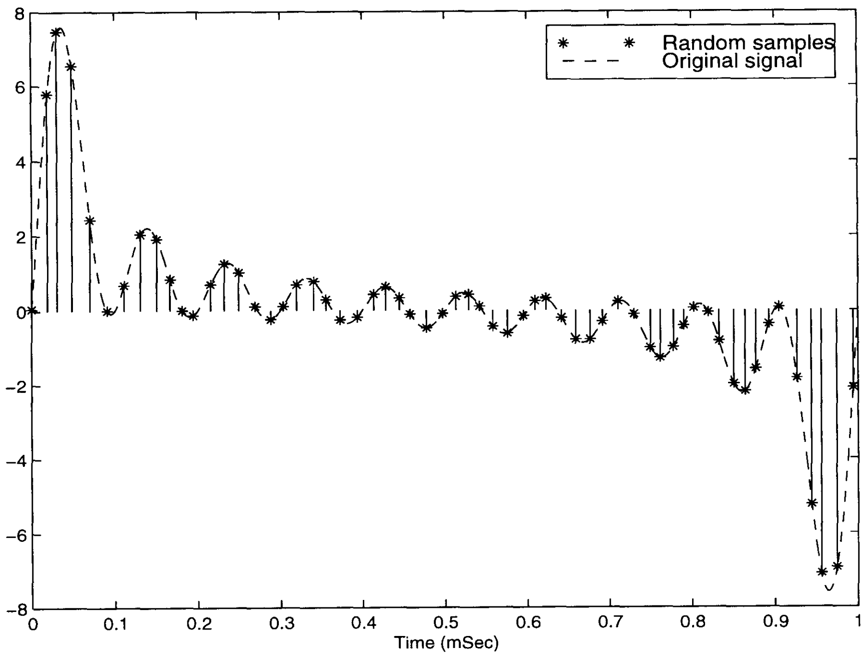

Figure 13: the random samples

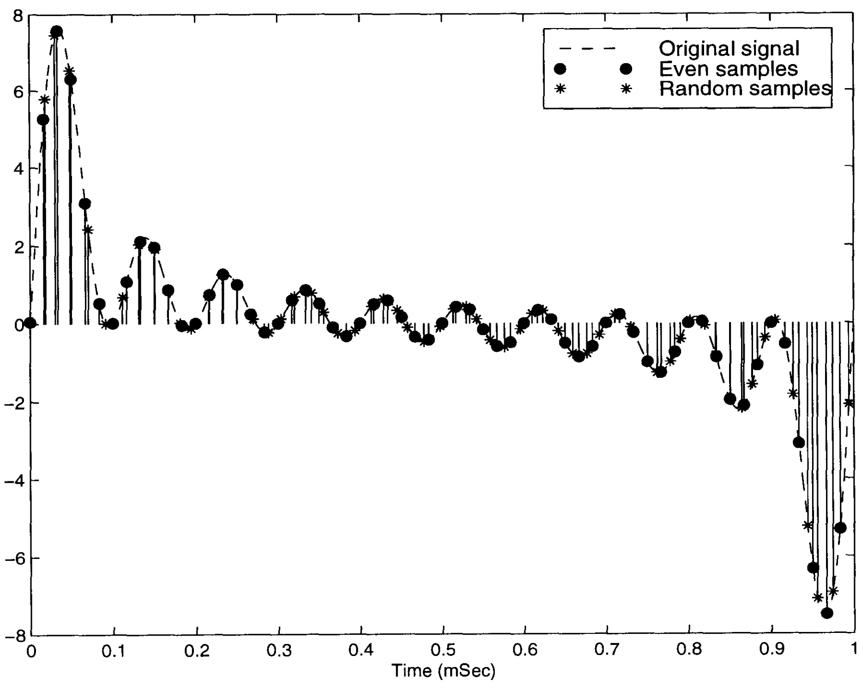

Figure 14: the recovered even samples

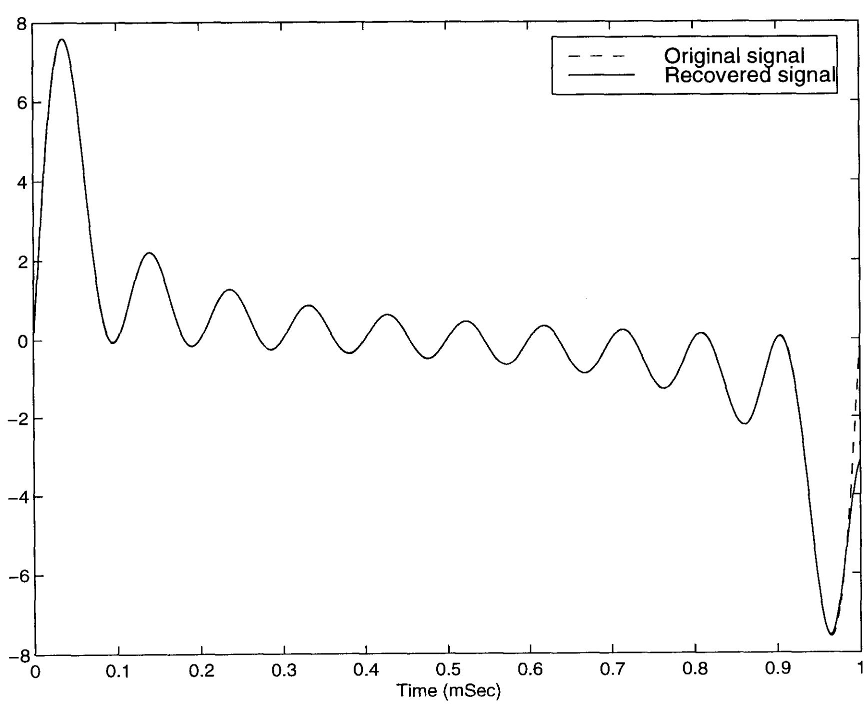

Figure 15: the recovered signal

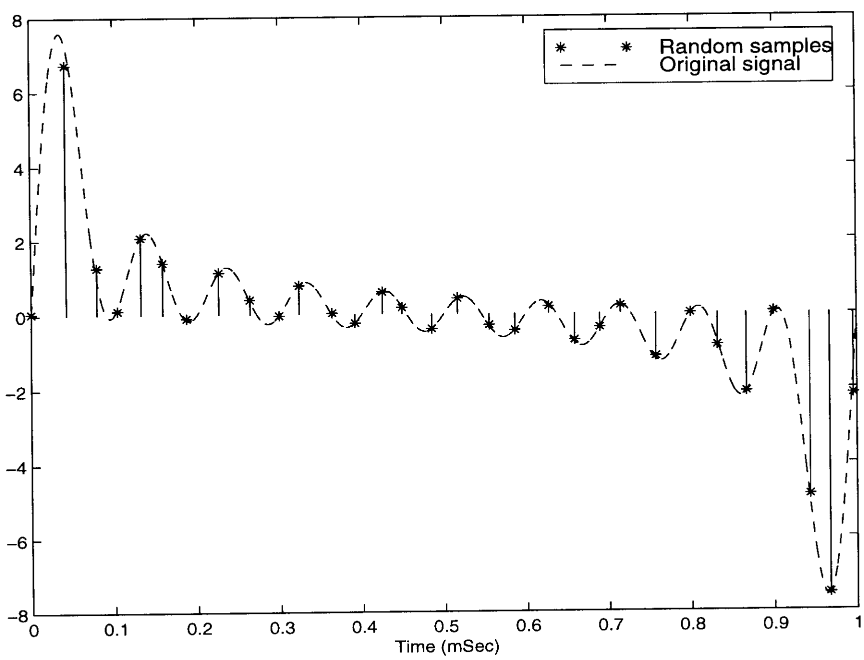

Figure 16: the random samples

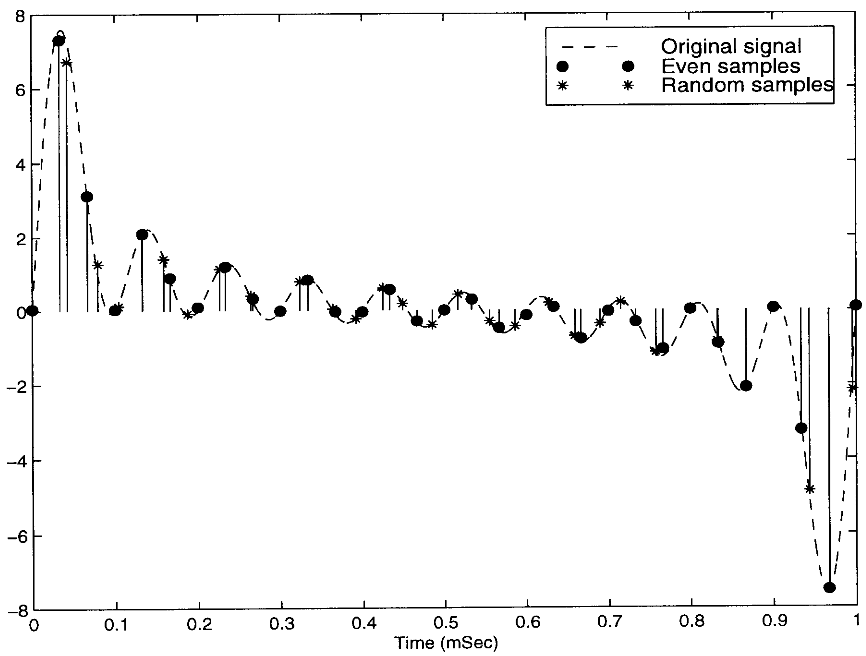

Figure 17: the recovered even samples

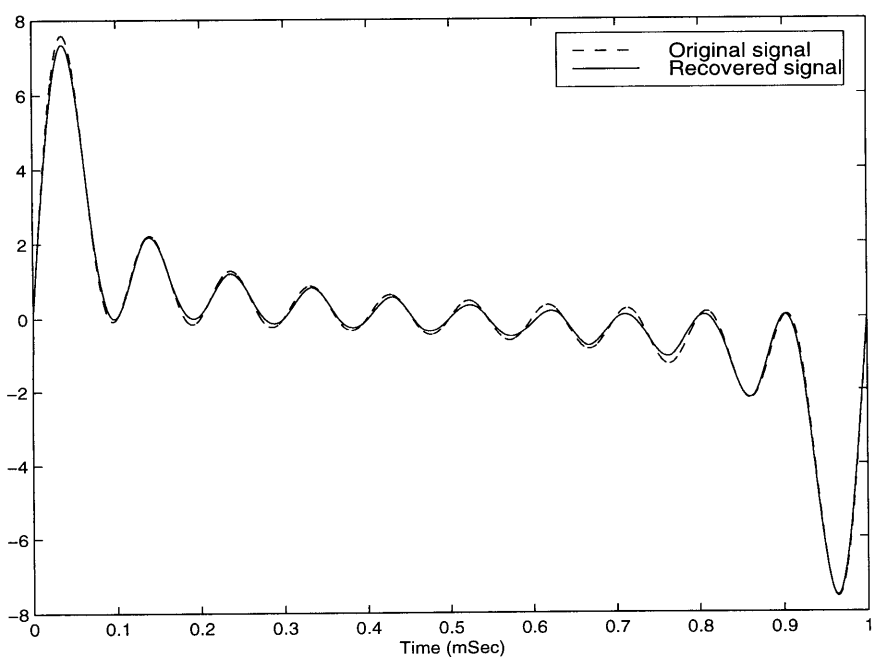

Figure 18: the recovered signal

Chapter 5

The Results of Experiments with Real Data

In chapter 4, we proved that the higher-order reconstruction technique is a better method for recovering the original signal from the randomly sampled data than the linear reconstruction technique. We will now apply this technique to real seismic data.

5.1 The Real Seismic Data

A set of real seismic data composed of 36 traces is shown in Figure 19. Each trace has 1750 samples. The sampling frequency is 500 Hz, which means the sampling interval here is 2 ms. The 30th trace is chosen to be resampled randomly and then to be reconstructed by applying the higher-order reconstruction technique.

The 1750 samples of the 30th trace are displayed in Figure 20. The total time period of those samples is 3.5 seconds. By taking the Fourier transform, in Figure 21, we can see that most energy falls into low frequency band ($f < 50$ Hz). When frequency is higher than 50 Hz, the energy is almost negligible. Therefore, we can choose 50 Hz to be the

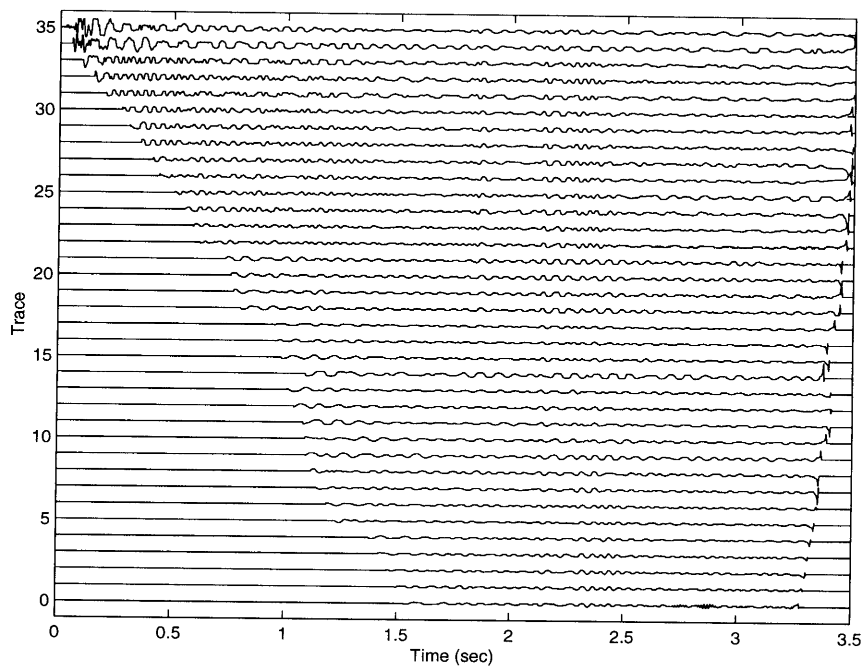

Figure 19: the real seismic data

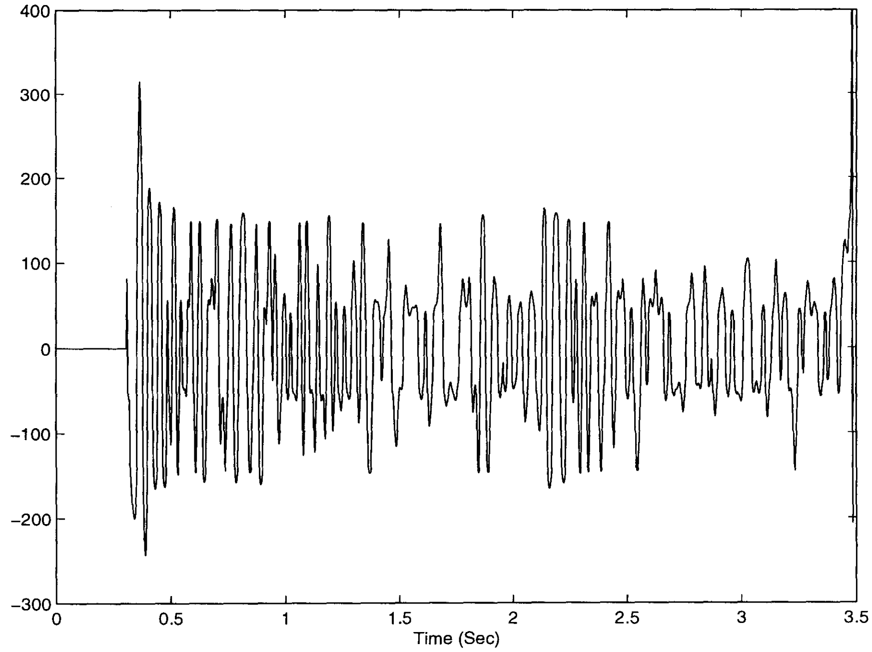

Figure 20: the 30th trace

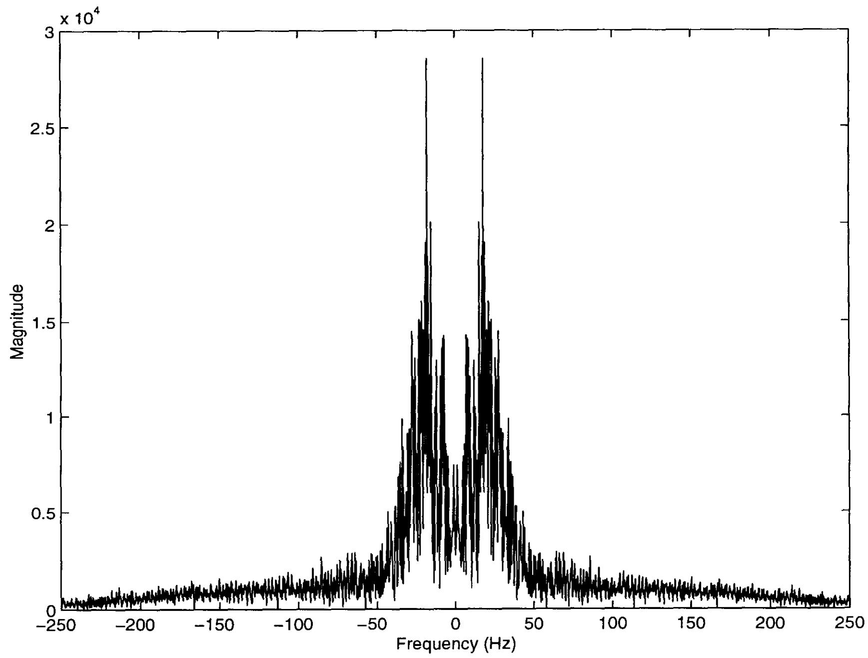

Figure 21: the Fourier spectrum of the 30th trace

highest frequency of the original signal and resample it with average sampling frequency at Nyquist rate (100 Hz).

5.2 Resampling the Original Data

The histogram of the sampling frequencies is shown in Figure 22. The average sampling frequency is 100 Hz. The sampling frequencies vary from 70 Hz to 130 Hz. Figure 23 shows that the resampled signal (random samples) matches the original signal very well. We will reconstruct the original signal from the random samples by applying the higher-order reconstruction method.

5.3 Reconstruction of the Original Signal

After applying the higher-order reconstruction method, the even samples can be obtained from the random samples. Figure 24 shows that the recovered even samples match the original signal very well. In Figure 25, after uniform interpolation, the even samples can have a 2 ms sample interval. The recovered signal fits the original signal well.

From the experiment above, we further believe that the higher-order reconstruction technique is a very good method to realistically recover the original signal from the randomly obtained data. So far, we have applied the reconstruction technique to recover signals in the time domain. We also can reconstruct the original space domain signal from the random samples by applying the higher-order reconstruction method.

5.4 Reconstruction of the Space Domain Signal

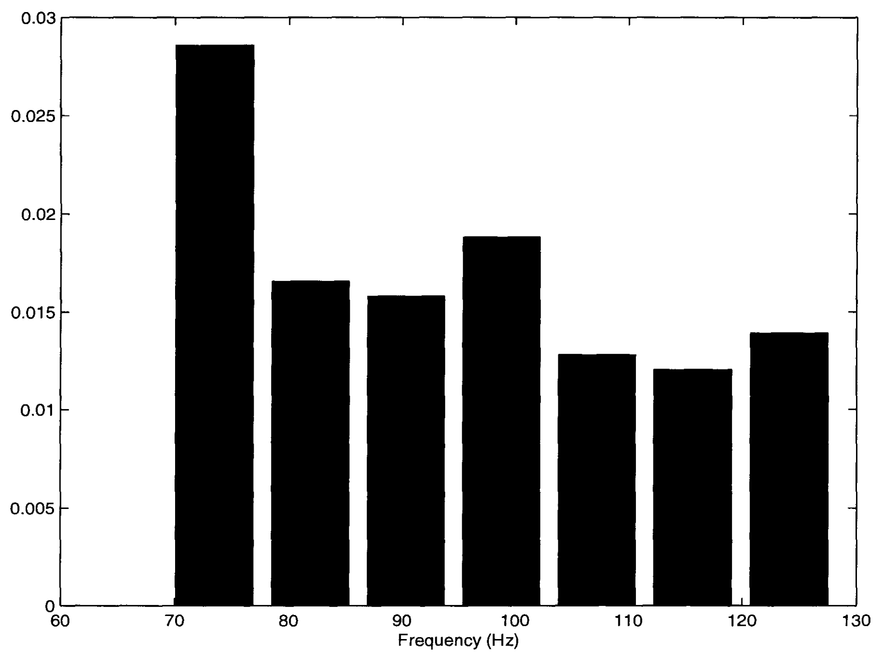

Figure 22: the histogram of the sampling frequencies

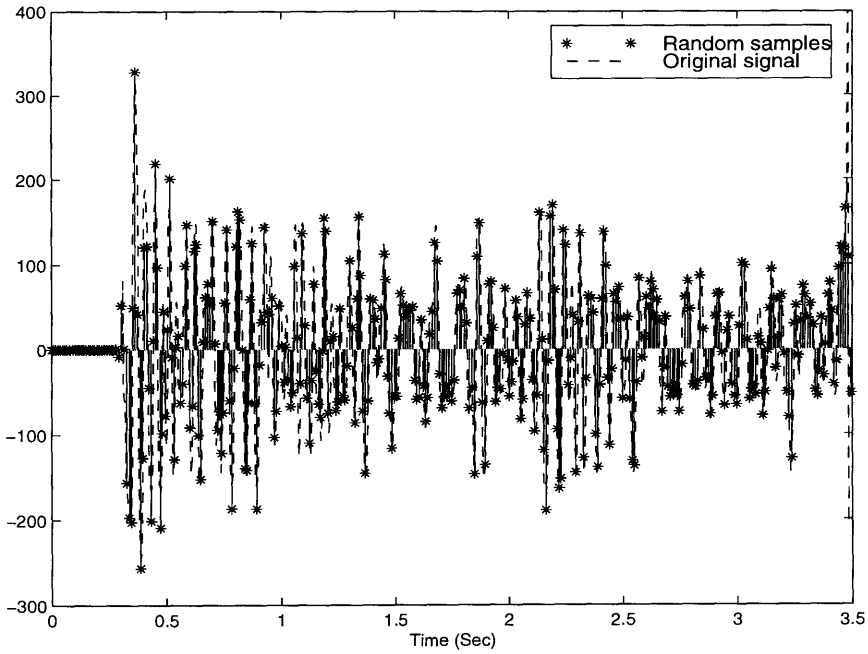

Figure 23(A): the random samples

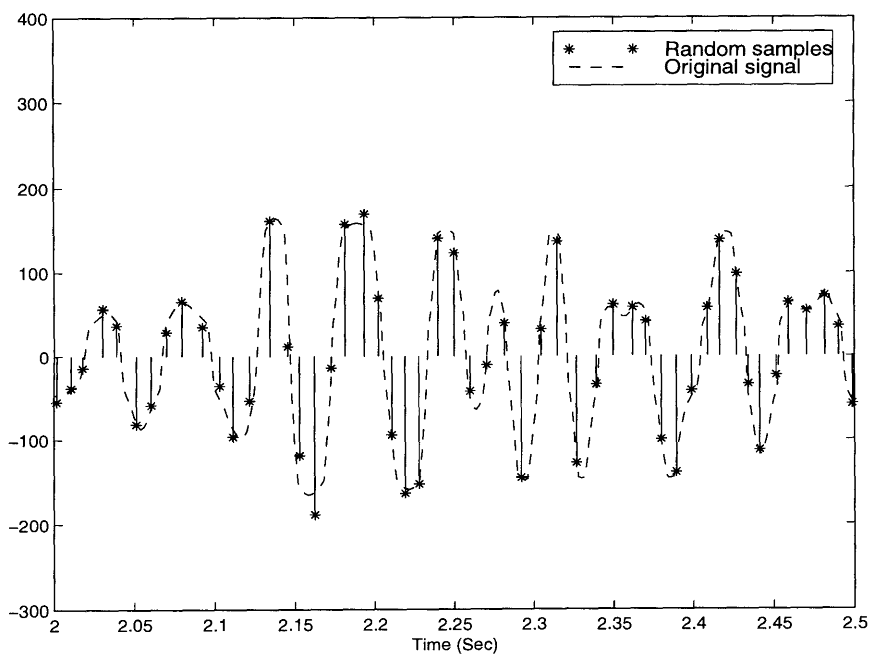

Figure 23(B): the zoomed plot of Figure 23(A)

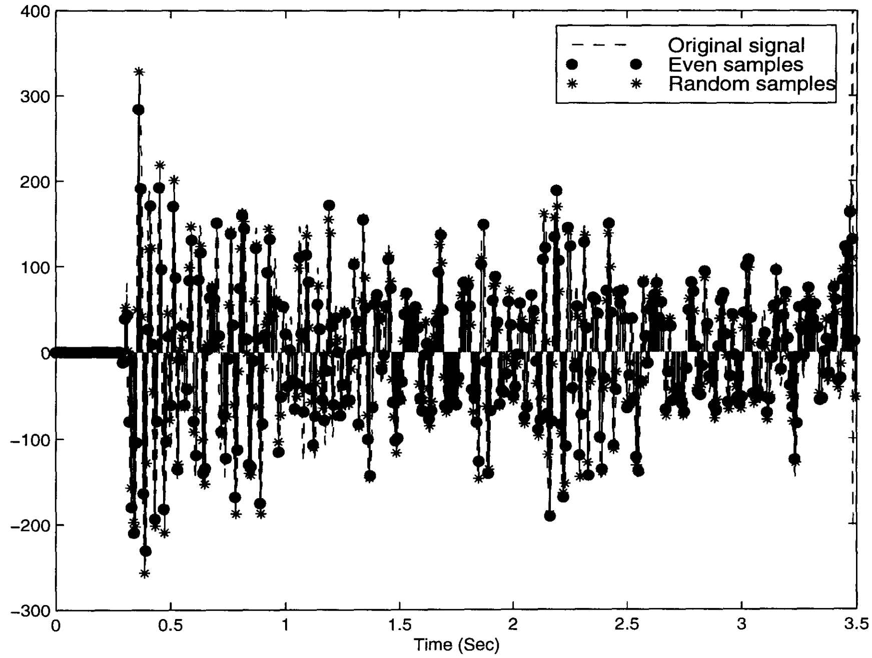

Figure 24(A): the recovered even samples

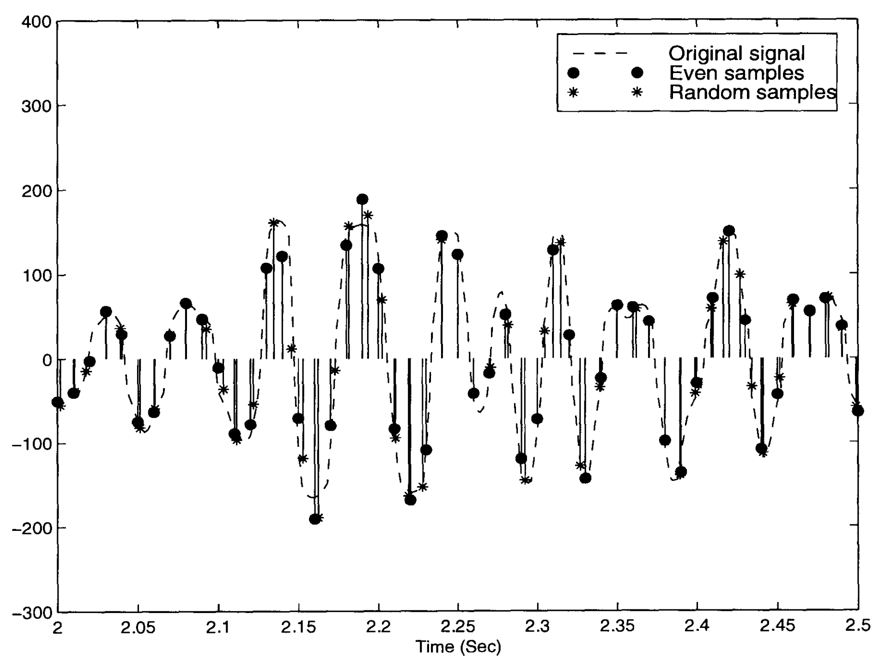

Figure 24(B): the zoomed plot of Figure 24(A)

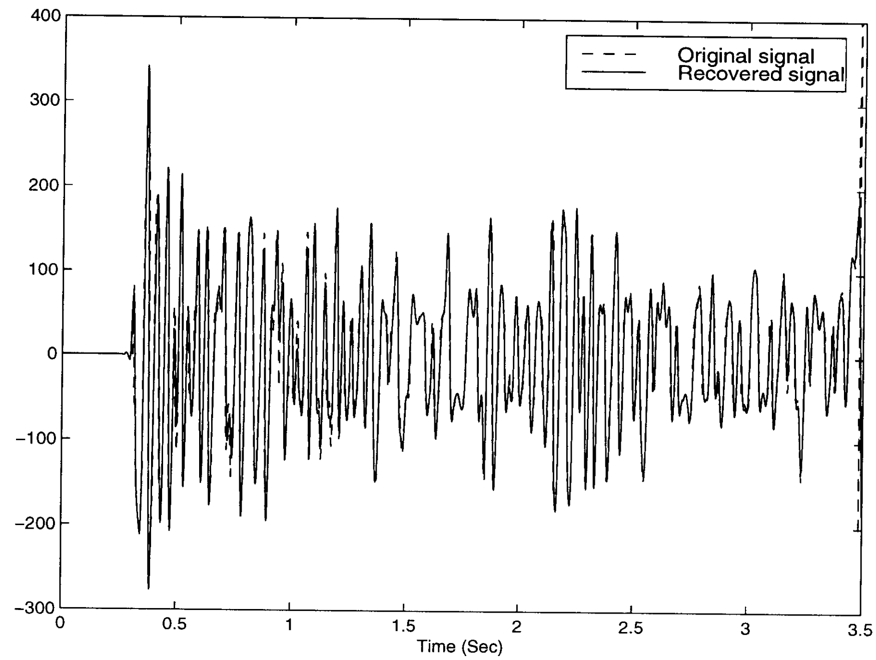

Figure 25(A): the recovered signal

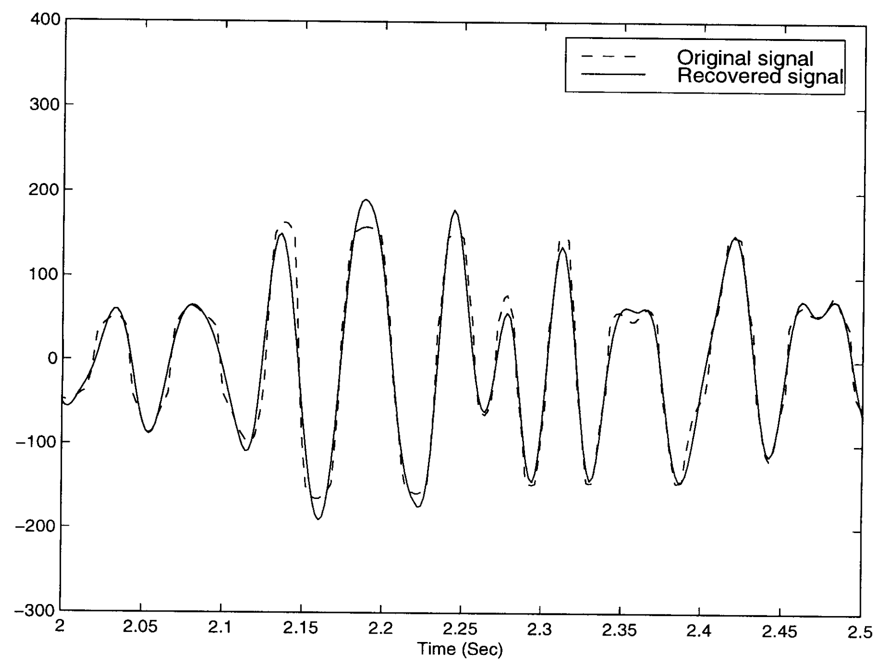

Figure 25(B): the zoomed plot of Figure 25(A)

An approach similar to recovery of time-domain signals can be applied to unevenly sampled space-domain signals. We can apply our time-domain techniques to process the space-domain signals by treating the space variable in a similar way we treated the time variable.

In Figure 19, we have 36 traces of real seismic data, and there are 1750 samples on each trace. For each time instant, there are 36 samples varying in space between 36 different traces. We pick the 1000th sample from each trace (that is t=2.0 sec in Figure 19) and combine them to be a space-domain signal. The space interval of those samples is 122 ft. In this paper, we simply normalize the space interval to be one. The original space signal is shown in Figure 26.

After taking the Fourier transform, we can see the frequency spectrum of the space signal in Figure 27. We resample the space signal by the random spatial frequencies with average sampling frequency at one. The histogram of the random sampling frequencies is shown in Figure 28. After resampling the original spatial signal, we have the randomly obtained data shown in Figure 29. Figure 30 gives the reconstructed evenly spaced samples after the reconstruction technique is applied. The finally reconstructed signal after the uniform interpolation is shown in Figure 31. There is a very good match between the original signal and the recovered signal.

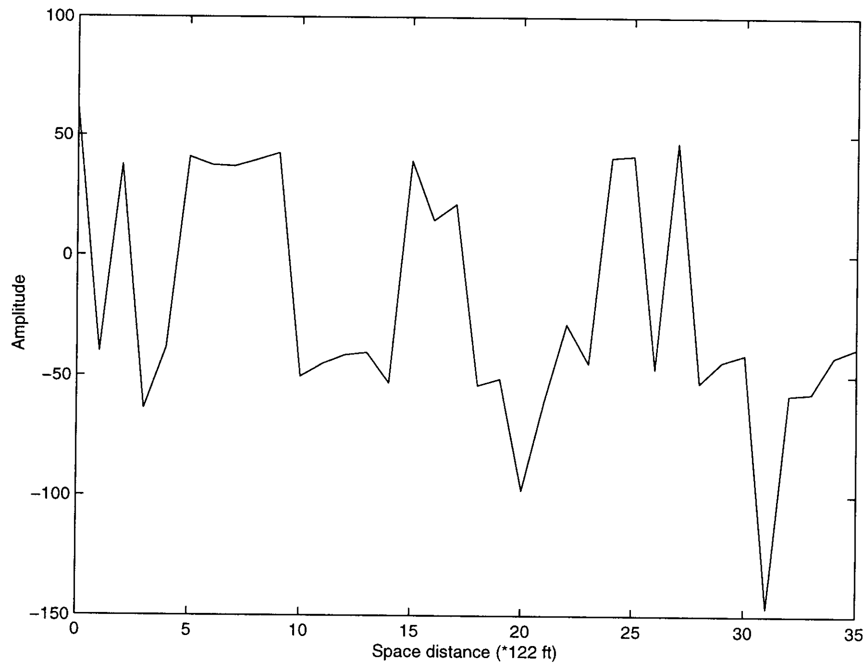

Figure 26: the space signal

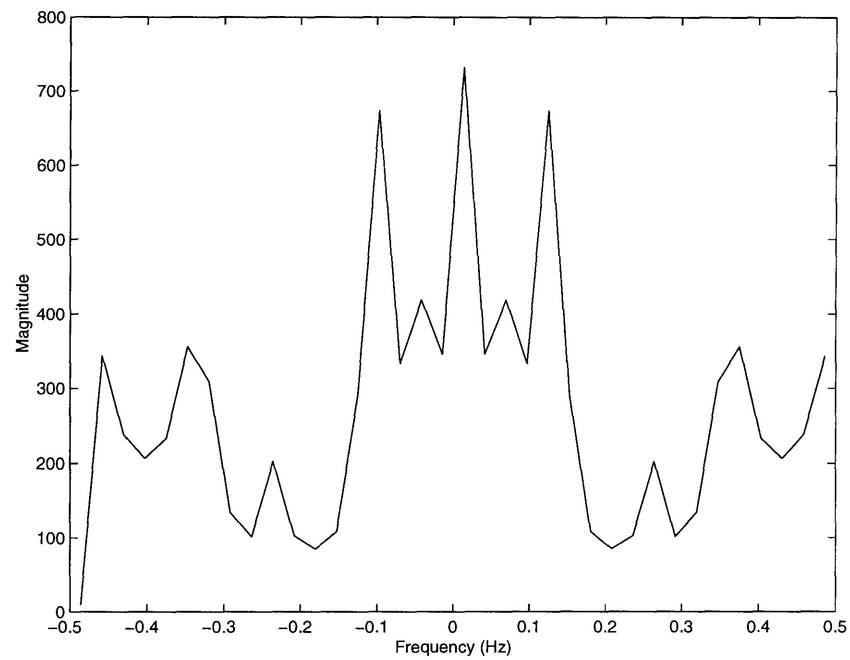

Figure 27: the Fourier spectrum of the space signal

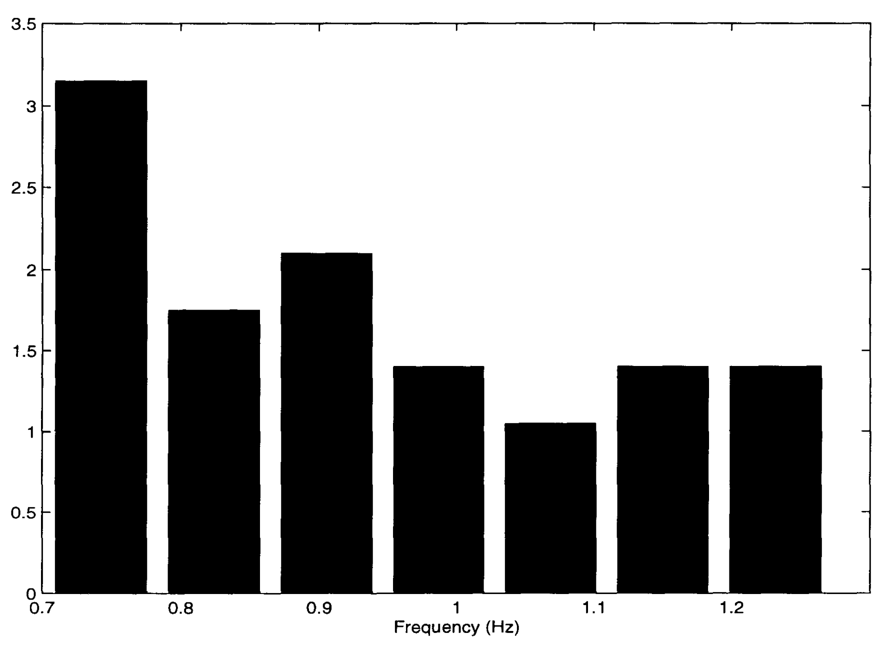

Figure 28: the histogram of the sampling frequencies

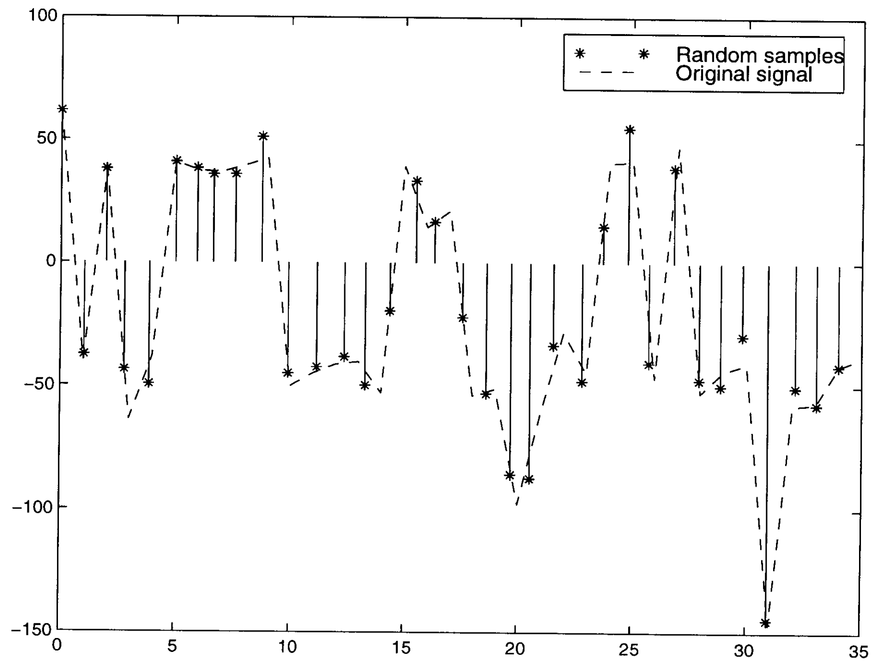

Figure 29: the random samples

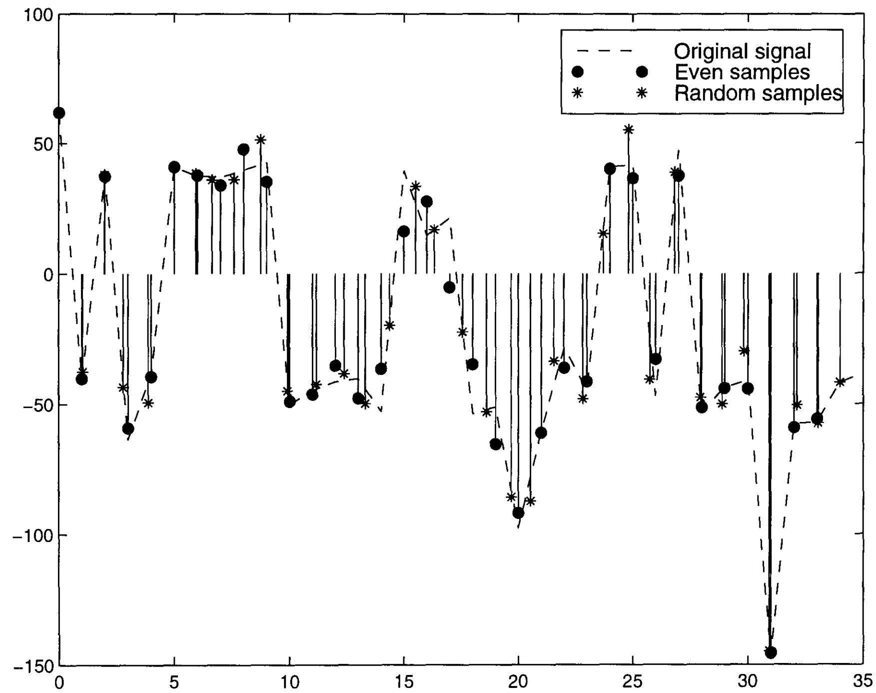

Figure 30: the recovered even samples

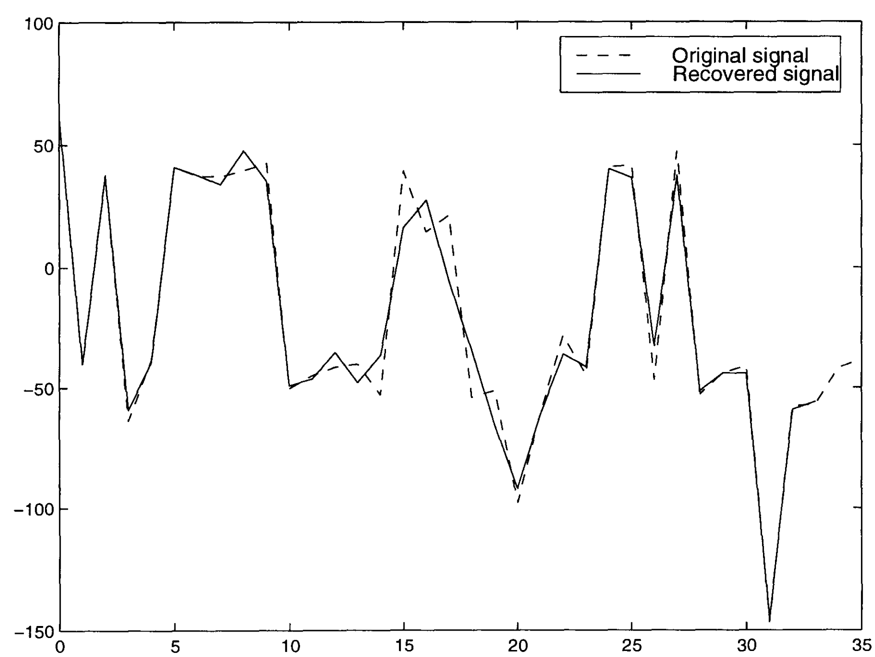

Figure 31: the recovered space signal

Chapter 6

Conclusions

6.1 Summary

In this thesis, we developed and compared two reconstruction techniques. Both the linear and higher-order reconstruction techniques can recover the original signals very well from the randomly obtained samples when the average sampling frequency is high enough. However, when the average sampling frequency is only the Nyquist rate or below, the linear reconstruction method behaves poorly. The higher-order technique, on the contrary, can reconstruct the original signal very well even if the average sampling frequency is only the Nyquist rate or below.

As far as the space-domain signal is concerned, the higher-order reconstruction technique can also be applied to recover the original space signal from the random samples. This is very helpful when we have to place the geophones randomly. The technique and conclusions obtained here open the door for future research in the processing of randomly sampled (in time or in space domain) seismic data.

6.2 Future Work

The experimental study and analysis presented in this thesis have shown the higher-order reconstruction technique to be a promising way to recover the original signal from randomly obtained samples. However, a number of issues were not addressed. Further investigations into these issues may yield interesting results and insights into improved reconstruction techniques. In this section, we give a few ideas of possible directions for future work.

First, this study applied the reconstruction to seismic data randomly obtained with the average sampling frequency at the Nyquist rate or below. However, since the random sampling frequencies are assumed to be uniformly distributed from $0.7*f_{Nyquist}$ to $1.3*f_{Nyquist}$, we are not quite sure how well the higher-order reconstruction technique can recover the original signal if random frequencies are Gaussian distributed. As we know, if the sampling frequencies are normally distributed, there must be some frequencies close to zero. Those low sampling frequencies will sample the original signal at a very long time or space interval. The reconstruction of the original signal from the random samples obtained by the normally distributed sampling frequencies will be even more difficult.

Second, in order to reconstruct the original space signal from the randomly spaced samples, there needs to be enough spatial samples. The seismic data we used in this thesis contain only 36 samples. The reconstruction would be harder if even fewer randomly placed goephones were used. More important however is when geophones are distributed aerially, as is the case in 3-D seismic acquisition. For this we need to develop the method for random sampling along both x and y coordinates.

Third, in this thesis all the computer simulations and real data experiments were done using Matlab. As we know, in the seismic data processing field, Promax is a more commonly used package. Therefore it's a better idea to develop a simulation and experiment method using Promax. Matlab was chosen due to time constraints and my familiarity with it.

Appendix: Explanation Regarding Reconstruction Algorithms

A.1 Reconstruction Implementation

There are two reconstruction techniques introduced in this thesis. In chapter 3, more than 20 equations are used to elaborate how the two reconstruction methods work. Some formulas are classical and commonly used in signal processing. One thing I have to point out is: there are some differences between the actual algorithm I used in this thesis and the theoretical methods discussed in chapter 3. The modifications were made according to my familiarity with Matlab and the Sparc 5 Sun station. There are some technical ways to improve the calculation efficiency. These modifications are not different methods than those in chapter 3. Here I will restate the key points of the procedures described in chapter 3.

As mentioned in chapter 3, the linear method reconstructs the original signal from random samples by linearly interpolating random samples to be even samples. Any even sample is obtained by the linear interpolation of its two adjacent random samples. The related interpretation is shown in Figure 4. The higher-order reconstruction method, however, reconstructs the original signal from random samples by applying Sinc function convolution with several neighboring random points of the interpolated even sample. The mechanism of the higher-order interpolation is shown in Figure A.1. The step (V) of the higher-order reconstruction on Page 25 shows the related operation. The equation corresponding to the higher-order interpolation is given in equation 17-20. Since performing the interpolation by Sinc is not quite efficient, in this thesis, we actually apply the curve fitting (data fitting) technique to interpolate the even samples from random samples. The data fitting technique is shown in Figure A.2. Basically, we can find a polynomial which fits the given random samples very well and interpolate (calculate by the polynomial) the even samples after that. We discovered that the quality of reconstruction is very similar by applying Sinc convolution and data fitting method separately. However, the method is faster using data fitting method rather than Sinc convolution. In next section, we will compare the computation speed of reconstruction. We make use of the data fitting method to evaluate the computation speed of higher-order reconstruction.

A.2 Computation speed of Reconstruction

Second, as mentioned in chapter 3, the higher-order method can do a better job than the linear method, but is less efficient for calculations. We verified this statement by measuring the time elapsed when we apply the linear method and the higher-order method separately to reconstructing the original signal from the same random samples. It takes 13.9 seconds for the higher-order method to process 1750 real seismic data points, but only takes 11.2 seconds for the linear-order method. We therefore believe that the time difference will be even larger when we process an even larger volume of data. In general, the linear method does the job faster however with poorer quality, and the higher-order method is the opposite. Therefore, there is a tradeoff when you have a large capacity of data and want to process them very fast. However, if the average sampling frequency is higher than the Nyquist rate, the linear method can still do a good job in reconstruction. Therefore, it is wise to choose the linear technique to deal with a large number of data as in this case.

A.3 Random versus Even Samples

Finally, we want to compare the accuracy of reconstruction from random samples with that from even samples. In chapter 5, we chose 50 Hz to be the highest frequency of the real seismic data. The average sampling frequency was chosen to be the Nyquist rate (100 Hz), and the real seismic data were sampled at 500 Hz. Therefore, downsampling the original signal by 5 is equivalent to resampling the real data at the Nyquist rate. Figure A.3-A.6 show that the reconstruction from uniform samples is also successful when the higher-order reconstruction method is applied. By equation (9) in chapter 3, we calculated the error E to be -12.45 dB for the random reconstruction and -11.30 dB for the uniform reconstruction, respectively. Therefore, we can see that random reconstruction can have almost the same accuracy as uniform reconstruction when the sampling frequency and the average sampling frequency are the Nyquist rate.

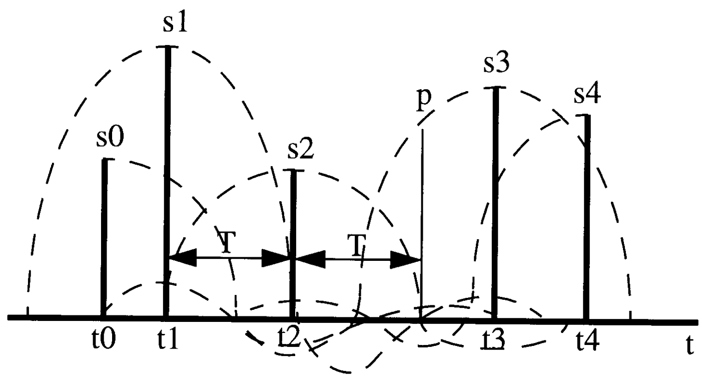

Figure A.1: obtaining pseudo samples for interpolating values between t2 and t3

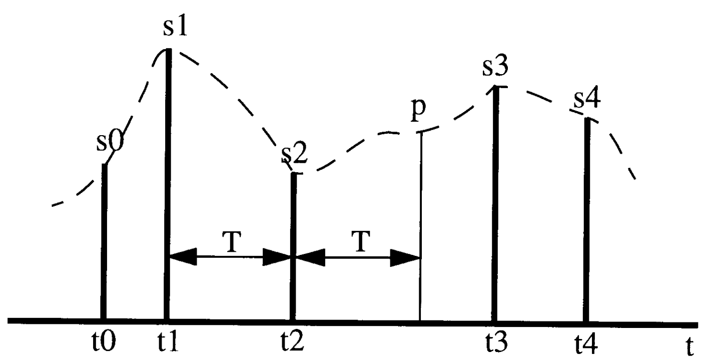

Figure A.2: obtaining pseudo samples for interpolating values between t2 and t3

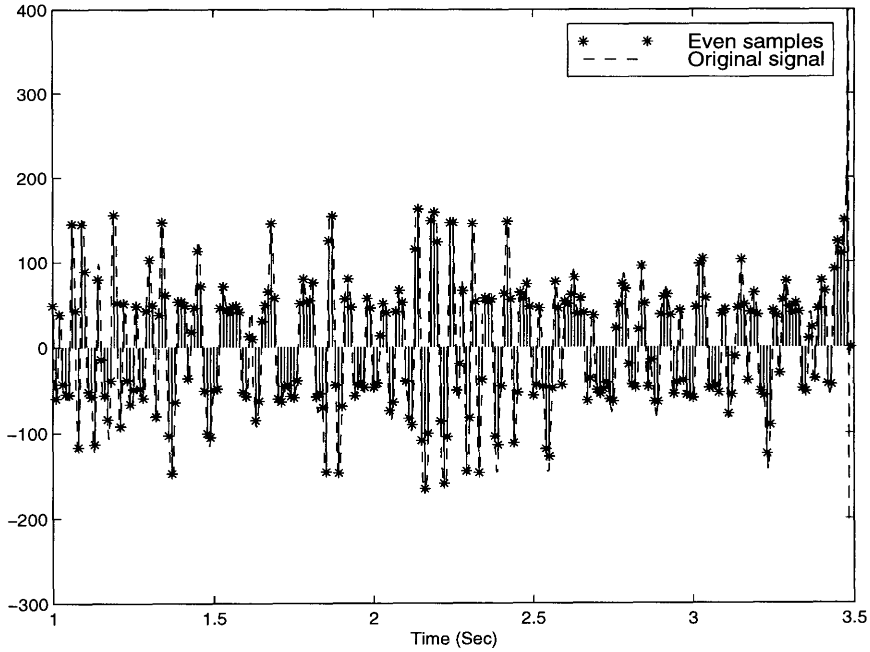

Figure A.3: the even samples by downsampling the original signal by 5

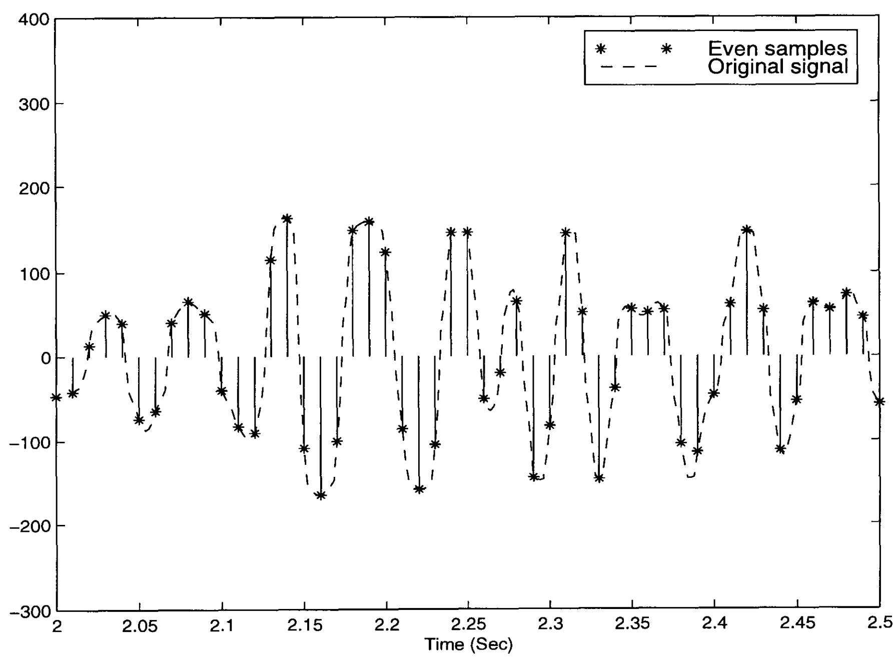

Figure A.4: the Zoomed plot of Figure A.3

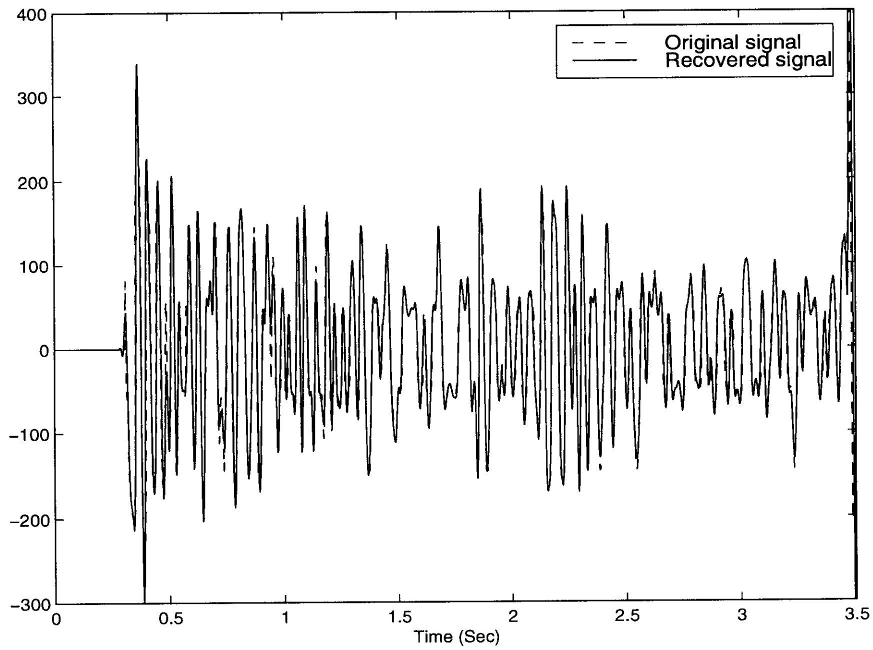

Figure A.5 the recovered signal

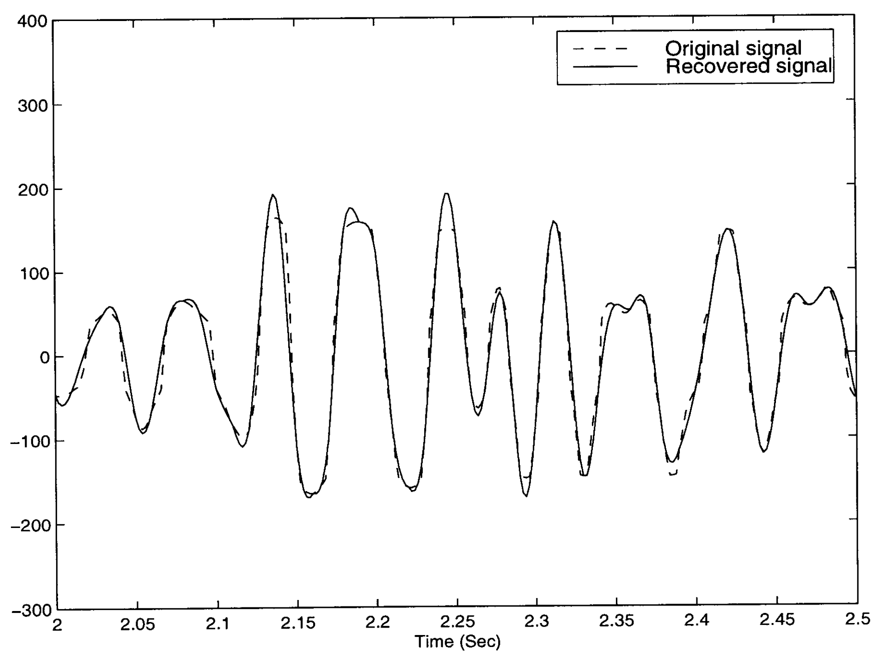

Figure A.6: the zoomed plot of Figure A.5

Bibliography

[1] Obata, M., and Mori, K., 1997, *Nonuniform sampling and interpolation based on partially dilated sampling function*: 1997 IEEE Pacific Rim Conference on Communications, Computer and Signal Processing, (Cat. No. 97CH36060), p. 2 vol.

[2] Toraichi, K. and Kamada, M., 1990, *A note on connection between spline signal spaces and band-limited signal spaces*: trans. IEICE, vol.J73-A, no.9, pp.1501-1508.

[3] Katagishi, K., Toraichi, K., Obata, M., and Wada, K., 1997, *A practical piecewise polynomials approximation based on compactly supported biorthonomal expansions*: to be appeared in the Proceedings of ICSPAT.

[4] Ghosh, K., and Dutta, P., 1994, *Nonuniform sampling and reconstruction of waveforms and its extension to wide band signal*: 1994 IEEE Instrumentation and Measurement Technology Conference (Cat. No 94CH3424-9), p. 3 vol.+suppl. 708-11 vol.2.

[5] Marvasti, F., Analoni, M., and Gamshadzhim M., 1991, *Recovery of signals from nonuniform samples using iterative method*: IEEE Transactions on Signal Processing, vol. -39, No. -4, pp 872-878.

[6] Jerri, J. A., 1977, *The Shannon sampling theorem - its various extensions and application*: A tutorial review, proc. IEEE, pp 1565 - 1596.

[7] Oppenheim A. V., and Schafer, R. W., 1989, *Discrete-Time Signal Processing*, New Jersey: McGraw-Hill.

[8] Kamada, M., Toraichi, K., and Mori, R., 1988, *Sampling bases for the signal space composed of spline functions*: Trans. IEICE, vol.J71-A, no.3, pp.875-881.

[9] Balakrishnan, A. V., 1965, *Essentially band-limited stochastic processes*: IEEE Trans. Inform. Theory, vol. IT-11, pp. 154-156.

[10] Sankur, B. and Gerhardt, L., 1973, *Reconstruction of signals from nonuniform samples*: IEEE Int. Conf. Commum., Conf,., Rec., pp. 15.13-15.18, June 11-13.

[11] Cipra, B. A. 1991, *Breaking the curse of dimensionality*: Science, 165.

[12] French, W. S., 1974, *Two-dimensional and three-dimensional migration*: Geophysics, 39, 265-277.

[13] French, W. S., 1992, *Implications of parallel computation in seismic data processing*: Leading Edges, 11, No. 22-25.

[14] Oppenheim, A., and Wilsky, A., 1983, *Signals and Systems*, Prentice-Hall, Inc.

[15] Press, W., Teukolsky, S., Vettering, W., and Flannery, B., 1992, *Numerical recipes in fortran*: Cambridge Univ. Press.

[16] Sikorski, K., and Schuster, G. T., 1991, *Hammersley quadrature applied to 3-D seismic migration*: Annual tomography and modeling report: Univ. of Utah, 234-259.

[17] Steinberg, B., 1983, *Microwave imaging with large antenna arrays*: J. Wiley and Sons.

[18] Steinberg, B., and Subbaram, H., 1991, *Microwave imaging techniques*: J. Wiley Co.

[19] Traub, J. E., and Wozniakowski, H., 1994, *Breaking intractability*: Scientific American, 270, 102-107.

[20] Wozniakowski, H., 1991, *Average case complexity of multivariate integration*: Bull. AMS, 24, 185-194.

[21] Zwillinger, D., 1992, *Integration*: Jones and Bartlett Publishers.

[22] Wunch, C., 1996, *The Ocean Circulation Inversion Problem: New York:* Cambridge University Press.