SPARSE SENSING IN BIG DATA

by

JUNJIE CHEN

Presented to the Faculty of the Graduate School of

The University of Texas at Arlington in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

THE UNIVERSITY OF TEXAS AT ARLINGTON

May 2014

To my parents, Jiuhao and Zhuying,

My brother Lin, and husband Xin

For their love, support and encouragement.

Acknowledgements

First and foremost I want to thank my supervising professor Dr. Qilian Liang. It has been an honor to be his Ph.D. student. I appreciate all his contributions of time, ideas, and funding to make my Ph.D. experience productive and stimulating. This dissertation would not have been possible without the help, support and patience of Dr. Liang, not to mention his advice and unsurpassed knowledge. The joy and enthusiasm he has for his research was contagious and motivational for me. I would like to thank him for his invaluable advice during my doctoral studies, and also for the excellent example he has provided as a successful professor.

I wish to thank my academic advisors Dr. Jean Gao, Dr. Ioannis Schizas, Dr. Junzhou Huang, Dr. Yuze (Alice) Sun, and Dr. Soontorn Oraintara, for their interest in my research, for their helpful discussions and invaluable comments and for taking time to serve in my dissertation committee.

I am grateful to Zhuo Li, Ishrat Maherin, Xin Wang, Na Li, Qiong Wu, Ji Wu, Lei Xu, Jing Liang, Zongjie Cao and Sheng Su for their helpful discussions and invaluable comments. I would like to acknowledge Dr. Qilian Liang, Dr. Ioannis Schizas, Dr. Jing Liang and Dr. Peter Wang for the courses on related topics that helped me improve my knowledge in the area.

I would also like to thank the following staff at University of Texas at Arlington, for their various forms of support during my graduate study–Ann Lewiston, Gail Paniuski, Janice Moore, Pauline Mason and so on.

Most importantly, none of this would have been possible without the love and patience of my family. My family to whom this dissertation is dedicated to, has been a constant source of love, concern, support and strength all these years. I would like to express my heart-felt gratitude to my family members - my parents, my brother Lin, my husband Xin, my uncle Changde Yan and aunt Shaying Yang, etc.

This work was supported in part by U.S. National Science Foundation under Grants CNS-1247848, CNS-1116749, CNS-0964713, and U.S. Office of Naval Research under Grants N00014-13-1-0043, N00014-11-1-0865, N00014-11-1-0071.

March 28, 2014

Abstract

SPARSE SENSING IN BIG DATA

Junjie Chen, Ph.D.

The University of Texas at Arlington, 2014

Supervising Professor: Qilian Liang

Big data requires exceptional technologies to efficiently process large quantities of data within tolerable elapsed times. While the potential benefits of Big Data are significant, there are still a lot of technical challenges that must be solved to fully realize this potential for Big Data.

For processing, transporting and storing large data sets of enormous sizes, data need to be greatly compressed. In this thesis, several sparse sensing algorithms - compressive sensing (CS), co-prime sampling and nested sampling are studied for Big Data, in theory and applications.

Error performance bounds of noisy compressive sensing are derived based on information theory and estimation theory. Information rate distortion function is a measure as the number of bits per symbol to be stored or transmitted under the constraint of a distortion. Rate distortion performance for scalar quantization of measurement observation is derived. Based on this, reconstruction rate distortion is also studied for CS. In addition, we study the real-world applications of CS in Big Data, to Synthetic Aperture Radar (SAR), radar sensor networks (RSNs), and underwater acoustic sensor networks (UWASNs).

Besides, properties of two new sparse sampling schemes, i.e., coprime sampling and nested sampling are investigated, such as rate distortion function, since sparse sampling can cause possible distortion because less number of samples are used. Theoretical analysis of how these two sparse sampling methods affect the power spectral density is given as well. A secure transmission scheme for Big Data based on coprime sampling and nested sampling is provided as well.

In addition, a hybrid sparse sampling approach is proposed, which combines nested sampling and compressive sensing to reduce the number of symbols, and rate distortion function is used as a criteria to determine how many bits should be used to represent the symbols during this process. We show that with this hybrid approach, less number of bits is required to represent the sensed information.

Finally, since LTE has been a Big Data consumer with ample data, how to allocate resources in the era of Big Data in telecommunications becomes a new issue. A bandwidth allocation method based on smartphone users personality traits and channel condition is studied in a unified mathematical framework in this dissertation.

In conclusion, facing "huge storage and bandwidth costs" challenges for Big Data, several approaches of sparse sensing in Big Data are studied, in theory and applications. Summary of contributions in this dissertation and future works are provided at the end.

Table of Contents

List of Illustrations

List of Tables

Chapter 1

Introduction

We are awash in a flood of data today. Data are flooding in at rates never seen before-doubling every 18 months [1], as a result of greater access to customer data from public, proprietary, and purchased sources, as well as new information gathered from Web communities and newly deployed smart assets. Data is being collected and transmitted at unprecedented scale [2] [4] in a wide range of application areas nowadays. Oil companies, telecommunication companies, and other data-centric industries have had huge data for long time. The phrase "Big Data" refers to large, diverse, complex, distributed data sets generated from instruments, sensors, Internet transactions, email, video, click streams, and all other digital sources available today and in the future, as defined by U.S. National Science Foundation in its recent solicitation.

Big Data is an emerging phenomenon characterized by the three Vs [3]: volume, velocity, and variety. The volume of data has increased from terabytes to petabytes and is encroaching on exabytes. Some pundits are suggesting that zettabytes (10^{21}) are reachable within the next several years. Velocity is concerned with not only how fast we accumulate data, but also how fast some of the data that we already have is changing. Some systems accumulate data at the rate of multiple petabytes per year; some systems have stored data that changes at the rate of terabytes per year. Changing data usually lags accumulating data by several orders of magnitude. Data accumulating at a multiple petabyte rate requires terabits to petabits of transport

capacity. Finally, the variety and modality of data is continually evolving; it may be both structured and unstructured.

Nowadays, decisions can be made based on the data itself, which is different from previously, when decisions were made based on guesswork, or on all kinds of designed models of reality. Such Big Data analysis now drives nearly every aspect of our modern society. How we handle the emergence of an era of Big Data is critical. Research on Big Data could be potentially improved by some inferential techniques. While the potential benefits of Big Data are significant, and some initial successes have already been achieved, there are still a lot of technical challenges that must be solved to fully realize this potential for Big Data. For example, a large-scale smartphone data collection campaign named as LDCC, was described in [7], and the collected data was first stored in the device and then uploaded to server. Approximately 270 TB of raw data would be generated in [8], which was a big challenge for data storage or analysis. Facing "huge storage and bandwidth costs" problem, Bennett et al discussed research challenges for big data and information systems in [9]. And recently, many efforts have been made to develop appropriate compression techniques for Big Data to enable storage and transmission requirements. As pointed by Leavitt [10], one big problem for Big Data is how organizations store and keep up with this tsunami of information. And scaling storage efficiently with no effect on performance is very challenging. Since data collection has been the principle bottleneck, challenges related to data storage, processing, analysis for Big Data was also emphasized in [11]. Chen et al. foresaw the emergence of Business Intelligence and Analytics, which will require techniques for collecting, analyzing, processing, and visualizing large scale mobile and sensor data in [12]. Several solutions were proposed to store and retrieve large amount of data in Big Data [13]. For example, Google File System (GFS) attempt to provide the robustness, scalability, and reliability [13].

Cohen et al. provided a parallel database design that supports SQL and MapReduce scripting on top of a DBMS to integrate multiple data sources [14]. Sun et al. [15] proposed cost-effective approaches that enable multi-tenancy at several levels to make Big Data analytic solutions more affordable.

Traditional compression methods are based on Nyquist sampling rate, which will have poor efficiency in terms of both sampling rate and computational complexity. Unlike traditional compression techniques, some sparse sampling algorithms have been proposed to overcome Nyquist sampling requirement, like compressive sensing, nested sampling and coprime sampling [135][136]. In this thesis, several approaches of sparse sensing in Big Data - compressive sensing, co-prime sampling and nested sampling are studied in theory and applications.

This thesis is organized as follows. Chapter 2 discusses compressive sensing theoretically for the probability of error and rate distortion performance. Besides, applications of CS to Synthetic Aperture Radar (SAR), radar sensor networks (RSNs), and Underwater Acoustic Sensor Networks (UWASNs) are also provided. Chapter 3 shows two new sparse sampling algorithms - co-prime sampling and nested sampling. Information-theoretic rate distortion performance is analyzed for these two sparse sampling methods. The analytical derivation of their spectrum efficiency are also presented in this chapter. A secure transmission for Big Data based on coprime sampling and nested sampling is also given in this chapter. Based on their unique advantages, a hybrid approach of nested sampling and compressive sensing is proposed in Chapter 4, which is efficient to represent huge amount of data, especially in Big Data, while keeps the signal's statistical information. And theoretical rate distortion performance of the proposed hybrid approach is analyzed as well. Since LTE has been a Big Data consumer with ample data, how to allocate resource in the new era of Big Data in telecommunications becomes a new issue. A bandwidth

allocation method based on smartphone users personality traits and channel condition is studied in a unified mathematical framework in Chapter 5. Conclusions and future works are provided in Chapter 6.

Chapter 2

Compressive Sensing

2.1 Introduction

Compressive sensing (CS) is a signal processing technique for efficiently acquiring and reconstructing a signal, by finding solutions to underdetermined linear systems. This takes advantage of the signal's sparseness or compressibility in some domain, allowing the entire signal to be determined from relatively few measurements.

This chapter is organized as follows. Section 2.2 gives a brief introduction of CS. Section 2.3 gives the detailed theoretical analysis of CS, in which Section 2.3.1 shows the derivation of the theoretical bounds of the probability of error. The analytical derivation of the rate distortion performance are presented in Section 2.3.2. Section 2.4 gives the real-world applications of CS in Big Data, to Synthetic Aperture Radar (SAR), radar sensor networks (RSNs), and underwater acoustic sensor networks (UWASNs). Section 2.5 summarizes the results.

2.2 Compressive Sensing: An Overview

For an N-dimensional signal $\underline{x}$, which is assumed to be K-sparse with respect to some basis matrix Ψ [16] [17] [19], it can be represented as $\underline{x} = \Psi\underline{\theta}$, or $\underline{\theta} = \Psi^T\underline{x}$, where Ψ is some orthogonal N-by-N sensing matrix and $\underline{\theta} \in \mathbb{R}^N$ has only K nonzero entries.

The signal will be sampled using the measurement vectors $\phi_i, i = 1, \cdots, M$:

$$y_i = \langle \underline{x}, \phi_i \rangle \tag{2.1}$$

Formula (2.1) could be denoted using $\underline{y} = \Phi\underline{x} = \Phi\Psi\underline{\theta}$, where $\underline{y}$ is the vector of observations and Φ (M-by-N, $M < N$) is the measurement operator which models the measurement system.

Compressive sensing is based on recovering $\underline{\theta}$ via convex optimization [19] [20]. When $\underline{y} = \Phi\Psi\underline{\theta}$ and $\underline{\theta}$ is sparse with respect to Ψ, we are seeking $\underline{\theta}$ consistent with $\underline{y}$ and such that $\Psi^{-1}\underline{x}$ has few nonzero entries.

In order to recover a K-sparse vector θ [16] [17][19], the number of measurements M must be at least greater than K but can be significantly smaller than the signal dimension, i.e., $K < M \ll N$ ($M \geq K\log(N/K)$).

The reconstruction of the sparse vector $\underline{\theta}$ can be achieved by searching for the sparsest vector $\hat{\underline{\theta}}$ consistent with the measurements, given the measurements $\underline{y}$. This is usually referred to as the l_0 optimization problem,

$$\hat{\underline{\theta}} = argmin\|\underline{\theta}\|_{l_0} \; subject \; to \; \underline{y} = \Phi\Psi\underline{\theta} \tag{2.2}$$

where the l_0 pseudo-norm corresponds to the number of non zero elements. As it is well known, this is a combinatorial problem which cannot be solved directly in practice.

And researchers proved that if the RIP (restricted isometry property) holds, then the following linear program gives an accurate reconstruction [16] [17] [19]:

$$\hat{\underline{\theta}} = argmin\|\underline{\theta}\|_{l_1} \; subject \; to \; \underline{y} = \Phi\Psi\underline{\theta} \tag{2.3}$$

With noise introduced,

$$\underline{y} = \Phi\Psi\underline{\theta} + \underline{w} = \Phi\underline{x} + \underline{w} \quad s.t. \quad \|w\|_{l_2} \leq \epsilon \tag{2.4}$$

where $\underline{w}$ is a random noise whose size can be bounded $\|\underline{w}\|_{l2} \leq \epsilon$. Therefore, the recovery process here is solving the problem:

$$\hat{\underline{\theta}} = \arg\min \|\underline{\theta}\|_{l1} \quad s.t. \quad \|\underline{y} - \Phi\Psi\hat{\underline{\theta}}\|_{l2} \leq \epsilon \tag{2.5}$$

2.3 Theoretical Analysis

2.3.1 Theoretical Bounds for the Probability of Error

The results in this chapter are derived under the following assumptions:

(1) The source information $\underline{x} \in \mathbb{R}^N$ is $K-$sparse in the sparsity basis Ψ (there are only K non-zero entries in $\underline{\theta} = \Psi^T \underline{x}$), and each element of $\underline{x}$ is an i.i.d. Gaussian random variable with zero mean and variance σ_x^2.

(2) This chapter analyzes both the general case without any assumption of the measurement matrix and the special case with Bernoulli measurement matrix of compressive sensing. For the special case, the measurement matrix Φ is Bernoulli matrix, that is, the entries of $\phi_{m,n}$ are i.i.d. with $Pr(\phi_{m,n} = \pm 1) = 0.5$. In addition, the measurement matrix Φ is known for both the transmitter and the receiver in advance.

$$\phi_{m,n} = \begin{cases} +1, prob = \frac{1}{2} \\ -1, prob = \frac{1}{2} \end{cases} \tag{2.6}$$

(3) Random Noise: each elements of the noise vector $\underline{w} \in \mathbb{R}^M$ is also an i.i.d. Gaussian random variable with zero mean and variance σ_w^2.

(4) The observations are $\underline{y} = \Phi \underline{x} + \underline{w}$, for the special case with Bernoulli measurement matrix,

$$y_m = \sum_{n=1}^{N} \phi_{m,n} x_n + w_m = \sum_{i=1}^{N_1} (+1) x_i + \sum_{j=1}^{N_2} (-1) x_j + w_m \tag{2.7}$$

$$N_1 + N_2 = N \tag{2.8}$$

Because $Pr(\phi_{m,n} = \pm 1) = 0.5$, if the length N of x is large enough, from the Strong Law of Large Number Theorem, with high probability, $N_1 = N_2 = \frac{N}{2}$,

$$y_m = \frac{N}{2} \sum_{i=1}^{N_1} x_i - \frac{N}{2} \sum_{j=1}^{N_2} x_j + w_m \tag{2.9}$$

As each of x_n is an i.i.d. Gaussian random variable with zero mean and variance σ_x^2, and each noise w_m also follows Gaussian with zero mean and variance σ_w^2, y_m is also a Gaussian random variable.

Since $\underline{y} = \Phi\underline{x} + \underline{w}$, we could notice that the mean of $\underline{y}$ is $\underline{0}$, with covariance matrix C_{yy}.

(5) Quantization: Let a finite discrete set $\mathcal{C} \subset \mathbb{R}^M$ be a codebook. A mapping from $\mathbb{R}^M$ to the codebook $\mathcal{C}$ is a quantizer as $\mathbb{R}^M \to \mathcal{C}$ with $y \to \mathcal{L} \in \mathcal{C}$, where $\mathcal{L}$ is a quantization level. Quantization of a signal results in some distortion of the signal. The performance of a quantizer is often described by its rate distortion function[114][112].

(6) Rate distortion theory: Information rate distortion function is a measure of distortion between the original source and its representation. From [36], we know that information rate distortion function is defined as

$$R(D) = \min_{Ed(X^L, \hat{X}^L) \leq D} I(X^L; \hat{X}^L) \tag{2.10}$$

where $I(X^L; \hat{X}^L)$ is the mutual information between X^L and $\hat{X}^L$.

2.3.1.1 Lower Bound for the Probability of Error

Theorem 1. *In noisy CS, consider* $\underline{x} \sim \mathcal{N}(0, \sigma_x^2 I_N)$, $\underline{w} \sim \mathcal{N}(0, \sigma_w^2 I_M)$, *and the entries of* $\phi_{m,n}$ *are i. i. d. with* $Pr(\phi_{m,n} = \pm 1) = 0.5$, *then the lower bound of the probability of error will be as follows,*

$$P_e \geq 1 - \frac{\frac{1}{2}\log(\frac{N\sigma_x^2 + \sigma_w^2}{\sigma_w^2})^M + 1}{\frac{1}{2}\log[(2\pi e\sigma_x^2)^N]} \tag{2.11}$$

Proof: As each x_n is a Gaussian i.i.d random variable, and $C_x = \sigma_x^2 I_N$, the entropy of $\underline{x}$ can be expressed as [36],

$$h(\underline{x}) = \frac{1}{2}\log(2\pi e\sigma_x^2)^N \tag{2.12}$$

As shown above, $\underline{y} = \Phi\underline{x} + \underline{w}$, we know that $\underline{x} \sim \mathcal{N}(0, \sigma_x^2 I_N)$, and $\underline{w} \sim \mathcal{N}(0, \sigma_w^2 I_M)$, then the covariance matrix of $\underline{y}$ should be calculated as,

$$\begin{aligned} C_{yy} &= E[(\underline{y} - \Phi\underline{\mu_x})(\underline{y} - \Phi\underline{\mu_x})^T] \\ &= E[(\Phi(\underline{x} - \underline{\mu_x}) + \underline{w})(\Phi(\underline{x} - \underline{\mu_x}) + \underline{w})^T] \\ &= \Phi E[\underline{x}\underline{x}^T]\Phi^T + C_w \\ &= \Phi C_{xx}\Phi^T + C_w \\ &\overset{(a)}{=} (N\sigma_x^2)I_M + \sigma_w^2 I_M \\ &= (N\sigma_x^2 + \sigma_w^2)I_M \end{aligned} \tag{2.13}$$

where equation (a) follows from the fact that $Pr(\phi_{m,n} = \pm 1) = 0.5$, if the length N of $\underline{x}$ is large enough, when $m \neq m'$, $E[(\sum_{n=1}^{N} \phi_{m,n}\phi_{m',n}] = 0$, therefore, $\Phi\Phi^T = N \cdot I_M$.

Since $\underline{y} = \Phi\underline{x} + \underline{w}$, we could notice that the mean of $\underline{y}$ is $\underline{0}$, with covariance matrix C_{yy}, i.e., $\underline{y} \sim \mathcal{N}(0, (N\sigma_x^2 + \sigma_w^2)I_M)$.

The joint likelihood function is,

$$p_{y|x}(\underline{y}; \underline{x}) = \frac{1}{2\pi(N\sigma_x^2 + \sigma_w^2)^{\frac{M}{2}}} exp(-\frac{1}{2}\underline{y}^T \cdot C_{yy}^{-1} \cdot \underline{y}) \tag{2.14}$$

$$p_{y|x}(\underline{y}; \underline{x}) = \frac{1}{2\pi(N\sigma_x^2 + \sigma_w^2)^{\frac{M}{2}}} exp(-\frac{1}{2(N\sigma_x^2 + \sigma_w^2)} \parallel \underline{y} - \Phi\underline{x} \parallel^2) \tag{2.15}$$

Fisher's Information could be calculated as [113]

$$[I(x)]_{i,j} = -E[\frac{\partial^2 \ln p(\underline{y}; \underline{x})}{\partial x_i \partial x_j}] = \frac{1}{N\sigma_x^2 + \sigma_w^2}\Phi^T\Phi \tag{2.16}$$

Therefore, the Cramer-Rao Lower Bound [113] for the estimator $\hat{x}$ is

$$E\{\parallel \underline{\hat{x}} - \underline{x} \parallel^2\} \geq Tr\{I^{-1}\} = (N\sigma_x^2 + \sigma_w^2)Tr\{(\Phi^T\Phi)^{-1}\} \tag{2.17}$$

the entropy of $\underline{y}$ can be expressed as,

$$h(\underline{y}) = \frac{1}{2}\log(2\pi e(N\sigma_x^2 + \sigma_w^2))^M \tag{2.18}$$

Since $\underline{x}$ and $\underline{y}$ are joint Gaussian random variables, with covariance matrix C_{xy}

$$C_{xy} = \begin{pmatrix} C_{xx} & Cov(\underline{x}, \underline{y}) \\ Cov(\underline{y}, \underline{x}) & C_{yy} \end{pmatrix}$$

$Cov(\underline{x}, \underline{y})$ is the covariance matrix between $\underline{x}$ and $\underline{y}$, where

$$Cov(x_n, y_m) = \phi_{m,n}\sigma_x^2 \tag{2.19}$$

The determinant of a block matrix: Suppose A, B, C, and D are matrices of dimension $(n \times n)$, $(n \times m)$, $(m \times n)$ and $(m \times m)$respectively, when A or D is invertible, the determinant of the block matrix could be calculated using

$$det\begin{pmatrix} A & B \\ C & D \end{pmatrix} = det(A)det(D - CA^{-1}B)$$

or,

$$det\begin{pmatrix} A & B \\ C & D \end{pmatrix} = det(D)det(A - BD^{-1}C)$$

As both C_{xx} and C_{yy} are invertible matrices, the calculation of the determinant of C_{xy} will be

$$\begin{aligned} det(C_{xy}) &= det(C_{xx}) \cdot det(C_{yy} - Cov(\underline{y}, \underline{x})C_{xx}^{-1}Cov(\underline{x}, \underline{y})) \\ &= (\sigma_x^2)^N \cdot det(C_{yy} - Cov(\underline{y}, \underline{x})(\sigma_x^2 I_N)^{-1}Cov(\underline{x}, \underline{y})) \\ &= (\sigma_x^2)^N \cdot det(C_{yy} - \frac{1}{\sigma_x^2}Cov(\underline{y}, \underline{x})Cov(\underline{x}, \underline{y})) \\ &\overset{(b)}{=} (\sigma_x^2)^N \cdot det(C_{yy} - \frac{1}{\sigma_x^2}(N\sigma_x^4)I_M) \\ &= (\sigma_x^2)^N \cdot det((N\sigma_x^2 + \sigma_w^2)I_M - (N\sigma_x^2)I_M) \\ &= (\sigma_x^2)^N(\sigma_w^2)^M \end{aligned} \tag{2.20}$$

where equation (b) follows from the fact that $Cov(\underline{y}, \underline{x})_{M\times N} = Cov(\underline{x}, \underline{y})^T_{N\times M}$, as $Pr(\phi_{m,n} = \pm 1) = 0.5$, if the length N of x is large enough,

$$Cov(y_m, x_n) \cdot Cov(x_n, y'_m)^T = \begin{cases} \sigma_x^2, \text{if } m = m' \\ 0, \text{if } m \neq m' \end{cases} \tag{2.21}$$

Therefore, we could obtain

$$Cov(\underline{y}, \underline{x})_{M\times N} \cdot Cov(\underline{x}, \underline{y})_{N\times M} = (N\sigma_x^2) I_M \tag{2.22}$$

From our derivation above, we could achieve the entropy of the joint Gaussian variables of $\underline{x}$ and $\underline{y}$,

$$\begin{aligned} h(\underline{x}, \underline{y}) &= \frac{1}{2} \log[(2\pi e)^{M+N} det(C_{xy})] \\ &= \frac{1}{2} \log[(2\pi e)^{M+N} (\sigma_x^2)^N ((\sigma_w^2)^M] \end{aligned} \tag{2.23}$$

As the process of compressive sensing forms a Markov chain with $\underline{x} \longrightarrow \underline{y} \longrightarrow \hat{\underline{x}}$, where $\underline{x}$ is the information source, $\underline{y} = \Phi\underline{x} + \underline{w}$ is the compressed observation using compressive sensing, and $\hat{\underline{x}}$ is the estimated information.

By the data-processing inequality, since the process of compressive sensing forms a Markov chain with $\underline{x} \longrightarrow \underline{y} \longrightarrow \hat{\underline{x}}$, $I(\underline{x}; \hat{\underline{x}}) \leq I(\underline{x}; \underline{y})$, therefore, $h(\underline{x}|\hat{\underline{x}}) \geq h(\underline{x}|\underline{y})$. Therefore, we could get,

$$h(\underline{x}|\underline{y}) \leq 1 + P_e \cdot h(\underline{x}) \tag{2.24}$$

From Fano's inequality, we could get the lower bound of the probability of error, as shown

$$\begin{aligned} P_e &\geq \frac{h(\underline{x} \mid \underline{y}) - 1}{h(\underline{x})} \\ &\overset{(c)}{=} \frac{h(\underline{x}, \underline{y}) - h(\underline{y}) - 1}{h(\underline{x})} \\ &= \frac{\frac{1}{2}\log[(2\pi e)^{M+N}(\sigma_x^2)^N((\sigma_w^2)^M]}{\frac{1}{2}\log[(2\pi e\sigma_x^2)^N]} - \frac{\frac{1}{2}\log[(2\pi e)^M(N\sigma_x^2+\sigma_w^2)^M] + 1}{\frac{1}{2}\log[(2\pi e\sigma_x^2)^N]} \\ &= 1 - \frac{\frac{1}{2}\log(\frac{N\sigma_x^2+\sigma_w^2}{\sigma_w^2})^M + 1}{\frac{1}{2}\log[(2\pi e\sigma_x^2)^N]} \end{aligned} \tag{2.25}$$

where equation (c) comes from the fact of entropy property $h(\underline{x}|\underline{y}) = h(\underline{x}, \underline{y}) - h(\underline{y})$.

This lower bound of the probability of error could be validated from a different proof as follows:

From information theory, we know the relationship between the mutual information and entropy as

$$I(\underline{x}; \underline{y}) = h(\underline{y}) - h(\underline{y}|\underline{x}) \tag{2.26}$$

As the measurement matrix Φ is known in advance, and conditioning reduces mutual information [36]

$$\begin{aligned} I(\underline{x}; \underline{y}|\Phi) &\leq I(\underline{x}; \underline{y}) \\ &= h(\underline{y}) - h(\underline{y}|\underline{x}) \\ &\overset{(d)}{=} h(\underline{y}) - h(\underline{w}) \\ &= \frac{1}{2}\log(2\pi e(N\sigma_x^2 + \sigma_w^2))^M - \frac{1}{2}\log(2\pi e\sigma_w^2)^M \\ &= \frac{1}{2}\log(\frac{N\sigma_x^2 + \sigma_w^2}{\sigma_w^2})^M \end{aligned} \tag{2.27}$$

where equation (d) follows from the fact that when the source information $\underline{x}$ and measurement matrix Φ are given, $h(\underline{y}|\underline{x}) = h(\underline{w})$.

Fano's inequality could be expressed using mutual information as,

$$\begin{aligned} P_e(\underline{x} \neq \hat{\underline{x}}|\Phi) &\geq \frac{h(\underline{x}|\underline{y}) - 1}{h(\underline{x})} \\ &\overset{(e)}{=} 1 - \frac{I(\underline{x};\underline{y}|\Phi) + 1}{h(\underline{x})} \\ &\geq 1 - \frac{\frac{1}{2}\log(\frac{N\sigma_x^2+\sigma_w^2}{\sigma_w^2})^M + 1}{\frac{1}{2}\log[(2\pi e\sigma_x^2)^N]} \end{aligned} \tag{2.28}$$

where equation (e) follows from the fact that $I(\underline{x};\underline{y}) = h(\underline{x}) - h(\underline{x}|\underline{y})$

From analysis, we notice that if $\sigma_x^2 \geq 1$, and the length N of information is sufficient large, $\frac{1}{\frac{1}{2}\log[(2\pi e\sigma_x^2)^N]} \approx 0$, therefore,
(1). If $(2\pi e\sigma_x^2)^N \geq (\frac{N\sigma_x^2+\sigma_w^2}{\sigma_w^2})^M$, then $P_{e_{lb}} \geq 0$, this means that perfect reconstruction of the information vector is impossible, as there will always be certain error.
(2). If $(2\pi e\sigma_x^2)^N \leq (\frac{N\sigma_x^2+\sigma_w^2}{\sigma_w^2})^M$, then $P_{e_{lb}} \leq 0$, this means that perfect reconstruction of the information vector is possible.

We could observe that Theorem 1 provides a special case of the lower bound of the probability of error for compressive sensing when the measurement matrix Φ follows Bernoulli distribution with $Pr(\phi_{m,n} = \pm 1) = 0.5$. We will give a more general lower bound of the probability of error without such special assumption of the matrix matrix Φ, as shown in Theorem 2:

Theorem 2. *In noisy CS, consider* $\underline{x} \sim \mathcal{N}(0, \sigma_x^2 I_N)$, $\underline{w} \sim \mathcal{N}(0, \sigma_w^2 I_M)$, *and the measurement matrix is* Φ, *then the lower bound of the probability of error will be as follows,*

$$P_e \geq \frac{N\log[2\pi e\lambda_{xy_{min}}] + M\log(\frac{\lambda_{xy_{min}}}{\lambda_{y_{max}}})}{N\log[2\pi e\sigma_x^2]} \tag{2.29}$$

Proof: Since the information vector $\underline{x}$ follows Gaussian distribution, and the noise vector $\underline{w}$ also follows Gaussian distribution, no matter what kind of measurement matrix Φ, the observation vector $\underline{y} = \Phi\underline{x} + \underline{w}$ should also follows Gaussian distribution.

$$h(\underline{y}) = \frac{1}{2}\log[(2\pi e)^M \cdot det(C_{yy})] \tag{2.30}$$

$$h(\underline{x},\underline{y}) = \frac{1}{2}\log[(2\pi e)^{N+M} \cdot det(C_{xy})] \tag{2.31}$$

As we know that the determinant of a matrix is the product of its all eigenvalues, therefore, we could express the determinants of the covariance matrices as,

$$det(C_{yy}) = \prod_{i=1}^{M} \lambda_y^i, \qquad det(C_{xy}) = \prod_{j=1}^{N+M} \lambda_{xy}^j \tag{2.32}$$

Therefore, from Fano's inequality, we could achieve the lower bound of the probability of error in a more general case,

$$\begin{aligned}
P_e &\geq \frac{h(\underline{x},\underline{y}) - h(\underline{y}) - 1}{h(\underline{x})} \\
&= \frac{\frac{1}{2}\log[(2\pi e)^{M+N}\prod_{j=1}^{N+M}\lambda_{xy}^j]}{\frac{1}{2}\log[(2\pi e\sigma_x^2)^N]} - \frac{\frac{1}{2}\log[(2\pi e)^M\prod_{i=1}^{M}\lambda_y^i] + 1}{\frac{1}{2}\log[(2\pi e\sigma_x^2)^N]} \\
&\overset{(f)}{\geq} \frac{\frac{1}{2}\log[(2\pi e)^{M+N}(\lambda_{xy_{min}})^{M+N}]}{\frac{1}{2}\log[(2\pi e\sigma_x^2)^N]} - \frac{\frac{1}{2}\log[(2\pi e)^M(\lambda_{y_{max}})^M] + 1}{\frac{1}{2}\log[(2\pi e\sigma_x^2)^N]} \\
&= (1+\frac{M}{N})\frac{\log(\frac{\lambda_{xy_{min}}}{\lambda_{y_{max}}})}{\log(2\pi e\sigma_x^2)} + \frac{\log(2\pi e\lambda_{y_{max}})}{\log(2\pi e\sigma_x^2)} - \frac{1}{\frac{N}{2}\log(2\pi e\sigma_x^2)} \\
&\overset{(g)}{\approx} \frac{N\log[2\pi e\lambda_{xy_{min}}] + M\log(\frac{\lambda_{xy_{min}}}{\lambda_{y_{max}}})}{N\log[2\pi e\sigma_x^2]}
\end{aligned} \tag{2.33}$$

where inequality (f) follows from the fact that $\prod_{j=1}^{N+M}\lambda_{xy}^j \geq (\lambda_{xy_{min}})^{M+N}$, and $\prod_{i=1}^{M}\lambda_y^i \leq (\lambda_{y_{max}})^M$. Approximation (g) follows from the fact that when the length N of $\underline{x}$ is sufficient large, and $\sigma_x^2 \geq 1$, then $\frac{1}{\frac{N}{2}\log(2\pi e\sigma_x^2)} = 0$.

From analysis, we conclude that:

(1). If $(2\pi e\lambda_{xy_{min}})^N \geq (\frac{\lambda_{y_{max}}}{\lambda_{xy_{min}}})^M$, then $P_{e_{lb}} \geq 0$, this means that perfect reconstruction of the information vector is impossible, as there will always be certain error.

(2). If $(2\pi e\lambda_{xy_{min}})^N \leq (\frac{\lambda_{y_{max}}}{\lambda_{xy_{min}}})^M$, then $P_{e_{lb}} \leq 0$, this means that perfect reconstruction of the information vector is possible.

2.3.1.2 Upper Bound for the Probability of Error

Theorem 3. *In noisy CS, consider* $\underline{x} \sim \mathcal{N}(0, \sigma_x^2 I_N)$, $\underline{w} \sim \mathcal{N}(0, \sigma_w^2 I_M)$, *and the measurement matrix is* Φ, *then the upper bound of the probability of error will be as follows,*

$$P_e \leq \frac{1}{(2\pi)^{N/2} \cdot det((\frac{1}{\sigma_x^2} I_N + \frac{1}{\sigma_w^2}\Phi^T\Phi)^{-1})^{1/2}} \tag{2.34}$$

Proof: $\underline{y} = \Phi\underline{x} + \underline{w}$. As each of x_n is an i. i. d. Gaussian random variable with zero mean and variance σ_x^2, and each noise w_m also follows Gaussian with zero mean and variance σ_w^2, we know that $\underline{x} \sim \mathcal{N}(0, \sigma_x^2 I_N)$, and $\underline{w} \sim \mathcal{N}(0, \sigma_w^2 I_M)$.

Using Bayesian estimator[113],

$$B_{MSE}(\underline{\hat{x}}) = E[(\underline{x} - \underline{\hat{x}})^2] = \iint (\underline{x} - \underline{\hat{x}})^2 p(\underline{y}; \underline{x}) \, \mathrm{d}\underline{y} \, \mathrm{d}\underline{x} \tag{2.35}$$

The estimator tries to find $\underline{\hat{x}}$ that minimizes $B_{MSE}(\underline{\hat{x}})$, that is,

$$\underline{\hat{x}} = arg \min_x B_{MSE}(\underline{\hat{x}}) \tag{2.36}$$

The following estimation could be derived

$$\underline{\hat{x}} = E[\underline{x} \mid \underline{y}] \tag{2.37}$$

For the assumed compressive sensing scheme, we could get the estimated $\underline{\hat{x}}$ as

$$\begin{aligned} \underline{\hat{x}} &= E[\underline{x}|\underline{y}] \\ &= E[\underline{x}] + C_{xy}C_{yy}^{-1}(\underline{y} - E[\underline{x}]) \\ &= (\frac{1}{\sigma_x^2} I_N + \frac{1}{\sigma_w^2}\Phi^T\Phi)^{-1}\Phi^T\underline{y}/\sigma_w^2 \end{aligned} \tag{2.38}$$

Besides, we could also get the following covariance matrix[113]

$$\begin{aligned} C_{\underline{x}|\underline{y}} &= E[(\underline{x} - E[\underline{x}|\underline{y}])(\underline{x} - E[\underline{x}|\underline{y}])^T|\underline{y}] \\ &= (\frac{1}{\sigma_x^2} \cdot I_N + \frac{1}{\sigma_w^2}\Phi^T\Phi)^{-1} \end{aligned} \tag{2.39}$$

The error between $\underline{x}$ and $\hat{\underline{x}}$ is

$$\underline{\varepsilon} = \underline{x} - \hat{\underline{x}} = \underline{x} - E[\underline{x}|\underline{y}] \tag{2.40}$$

Since estimator in (2.38) is an affine transformation of Gaussian random vector $\underline{y}$, hence $\hat{\underline{x}}$ is multivariate Gaussian. Estimation error $\underline{\varepsilon} = \underline{x} - \hat{\underline{x}}$ is an affine transformation of jointly Gaussian vectors $\underline{x}$ and $\underline{y}$, hence $\underline{\varepsilon}$ is also Gaussian.

The mean of $\underline{\varepsilon}$ is calculated as

$$\begin{aligned} E_{x,y}[\underline{\varepsilon}] &= E_{x,y}[\underline{x} - E[\underline{x}|\underline{y}] \\ &= \iint (\underline{x} - E[\underline{x}|\underline{y}])p(\underline{y};\underline{x})\,\mathrm{d}\underline{y}\,\mathrm{d}\underline{x} \\ &= \int [\int (\underline{x} - E[\underline{x}|\underline{y}])p(\underline{y}|\underline{x})\,\mathrm{d}\underline{x}]p(\underline{y})\,\mathrm{d}\underline{y} = \underline{0} \end{aligned} \tag{2.41}$$

The covariance of $\underline{\varepsilon}$ is

$$E[\underline{\varepsilon}\underline{\varepsilon}^T] = E_{\underline{x},\underline{y}}[(\underline{x} - E[\underline{x}|\underline{y}])(\underline{x} - E[\underline{x}|\underline{y}])^T] = C_{\underline{x}|\underline{y}} \tag{2.42}$$

As $\underline{y} = \Phi\underline{x} + \underline{w}$, and $\underline{x}$ and $\underline{y}$ are jointly Gaussian random variables, the error vector $\underline{\varepsilon}$ follows Gaussian distribution with mean $\underline{0}$ and covariance matrix $C_{\underline{x}|\underline{y}}$.

We could get the error vector's probability distribution function (pdf)[113] as follows,

$$p(\underline{\varepsilon}) = \frac{1}{(2\pi)^{N/2} \cdot det(C_{\underline{x}|\underline{y}})^{1/2}} exp(-\frac{1}{2}\underline{\varepsilon}^T \cdot C_{\underline{x}|\underline{y}} \cdot \underline{\varepsilon}) \tag{2.43}$$

For Gaussian distribution, we know that the probability achieve the maximum value at its mean value, therefore,

$$\begin{aligned} p(\underline{\varepsilon}) \leq p(\underline{0}) &= \frac{1}{(2\pi)^{N/2} \cdot det(C_{\underline{x}|\underline{y}})^{1/2}} \\ &= \frac{1}{(2\pi)^{N/2} \cdot det((\frac{1}{\sigma_x^2} I_N + \frac{1}{\sigma_w^2}\Phi^T\Phi)^{-1})^{1/2}} \end{aligned} \tag{2.44}$$

Theorem 4. *In noisy CS, if* $\underline{x} \sim \mathcal{N}(0, \sigma_x^2 I_N)$, $\underline{w} \sim \mathcal{N}(0, \sigma_w^2 I_M)$, *and the entries of the measurement matrix* $\phi_{m,n}$ *follows Bernoulli distribution with* $Pr(\phi_{m,n} = \pm 1) = 0.5$, *then the upper bound of the probability of error will be as follows,*

$$P_e \leq \frac{1}{(\frac{2\pi\sigma_x^2\sigma_w^2}{\sigma_w^2 + M\sigma_x^2})^{N/2}} \tag{2.45}$$

When the measurement matrix Φ is Bernoulli matrix, the entries of $\phi_{m,n}$ are i.i.d. with $Pr(\phi_{m,n} = \pm 1) = 0.5$, and if the length N of x is sufficiently large, $\Phi^T\Phi = M \cdot I_N$, from which we could achieve the result in Theorem 4.

Theorem 5. *In noisy CS, if* $\underline{x} \sim \mathcal{N}(0, \sigma_x^2 I_N)$, $\underline{w} \sim \mathcal{N}(0, \sigma_w^2 I_M)$, *and the entries of the measurement matrix* $\phi_{m,n}$ *follows Bernoulli distribution with* $Pr(\phi_{m,n} = \pm 1) = 0.5$, *then the upper bound of the probability of error will be as follows,*

$$P_e \leq 1 - (\pi e \sigma_x^2)^{-\frac{N}{2}} 2^{-\frac{N}{2}(\frac{N\sigma_x^2+\sigma_w^2}{M\sigma_x^2} - \ln(\frac{N\sigma_x^2+\sigma_w^2}{M\sigma_x^2}))} \tag{2.46}$$

Proof: From a linear minimum mean square error (MMSE) estimator, because $\underline{y} = \Phi\underline{x} + \underline{w}$, $\underline{y} \in \mathbb{R}^M$, $\Phi \in \mathbb{R}^{M \times N}$, and $\underline{w} \in \mathbb{R}^M$, $\underline{x} \sim \mathcal{N}(0, \sigma_x^2 I_N)$, and $\underline{w} \sim \mathcal{N}(0, \sigma_w^2 I_M)$ are independent.

Since estimator in (2.38) is an affine transformation of Gaussian random vector $\underline{y}$, hence $\underline{\hat{x}}$ is multivariate Gaussian. Thus the estimated expression for linear MMSE estimator also follows Gaussian distribution, with mean and auto-covariance given by,

$$\underline{\hat{x}} = E[\underline{x}|\underline{y}] = (\frac{1}{\sigma_x^2} I_N + \frac{1}{\sigma_w^2}\Phi^T\Phi)^{-1}\Phi^T\underline{y}/\sigma_w^2 \tag{2.47}$$

$$E[\underline{\hat{x}}] = E[\underline{x}] = \underline{0} \tag{2.48}$$

$$\begin{aligned} C_{\underline{\hat{x}}} &= C_{xy}C_{yy}^{-1}C_{yx} \\ &= C_{xx}\Phi^T(\Phi C_{xx}\Phi^T + C_w)^{-1}\Phi C_{xx} \\ &= \sigma_x^4\Phi^T(\sigma_x^2\Phi\Phi^T + \sigma_w^2 I_M)^{-1}\Phi \end{aligned} \tag{2.49}$$

Therefore, $\underline{\hat{x}} \sim \mathcal{N}(\underline{0}, C_{\underline{\hat{x}}})$, and its probability density function could be expressed as,

$$p(\underline{\hat{x}}) = \frac{1}{(2\pi)^{N/2} det^{1/2}(C_{\underline{\hat{x}}})} exp(-\frac{1}{2}\underline{\hat{x}}^T \cdot C_{\underline{\hat{x}}}^{-1} \cdot \underline{\hat{x}}) \tag{2.50}$$

As we know, the source signal $\underline{x} \sim \mathcal{N}(0, \sigma_x^2 I_N)$, its probability density function is,

$$p(\underline{x}) = \frac{1}{(2\pi\sigma_x^2)^{N/2}} exp(-\frac{1}{2}\underline{x}^T \cdot (\sigma_x^2 I_N)^{-1} \cdot \underline{x}) \tag{2.51}$$

From information theory [36], we know that if $\underline{x}$ and $\underline{\hat{x}}$ are independent with $\underline{x} \sim p(\underline{x})$, and $\underline{\hat{x}} \sim p(\underline{\hat{x}})$, then

$$Pr(\underline{x} = \underline{\hat{x}}) \geq 2^{-H(p(\underline{x}))-D(p(\underline{x})\|p(\underline{\hat{x}}))} \tag{2.52}$$

where $D(p(\underline{x})\|p(\underline{\hat{x}}))$ is the Kullback-Leibler distance between two probability mass functions $p(\underline{x})$ and $p(\underline{\hat{x}})$ which is defined as

$$D(p(\underline{x})\|p(\underline{\hat{x}})) = \sum p(\underline{x}) \log \frac{p(\underline{x})}{p(\underline{\hat{x}})} \tag{2.53}$$

Therefore, the probability of error between $\underline{x}$ and $\underline{\hat{x}}$ is upper bounded by

$$\begin{aligned} P_e &= Pr(\underline{x} \neq \underline{\hat{x}}) = 1 - Pr(\underline{x} = \underline{\hat{x}}) \\ &\leq 1 - 2^{-H(p(\underline{x}))-D(p(\underline{x})\|p(\underline{\hat{x}}))} \\ &= 1 - 2^{-\frac{1}{2}\log(2\pi e\sigma_x^2)^N} \cdot 2^{-D(p(\underline{x})\|p(\underline{\hat{x}}))} \\ &= 1 - (2\pi e\sigma_x^2)^{-N/2} 2^{-D(p(\underline{x})\|p(\underline{\hat{x}}))} \end{aligned} \tag{2.54}$$

When the measurement matrix Φ is Bernoulli matrix, the entries of $\phi_{m,n}$ are i.i.d. with $Pr(\phi_{m,n} = \pm 1) = 0.5$, and if the length N of x is sufficiently large, $\Phi^T\Phi = M \cdot I_N$, and $\Phi\Phi^T = N \cdot I_M$, therefore,

$$\begin{aligned} C_{\underline{\hat{x}}} &= \sigma_x^4 \Phi^T (\sigma_x^2 \Phi\Phi^T + \sigma_w^2 I_M)^{-1} \Phi \\ &= \frac{M\sigma_x^4}{N\sigma_x^2 + \sigma_w^2} I_N \end{aligned} \tag{2.55}$$

and

$$det(C_{\underline{\hat{x}}}) = (\frac{M\sigma_x^4}{N\sigma_x^2 + \sigma_w^2})^N \tag{2.56}$$

For compressive sensing, as $M < N$, $\sigma_x^2 > 0$ and $\sigma_w^2 > 0$, we could conclude that

$$\frac{M\sigma_x^4}{N\sigma_x^2 + \sigma_w^2} < \frac{M\sigma_x^4}{N\sigma_x^2} = \frac{M}{N}\sigma_x^2 < \sigma_x^2 \tag{2.57}$$

As we know, The Kullback-Leibler divergence between two multivariate normal distributions of the dimension N with the means μ_0, μ_1 and their corresponding nonsingular covariance matrices Σ_0, Σ_1 is:

$$D_{KL}(\mathcal{N}_0 \| \mathcal{N}_1) = \frac{1}{2}(tr(\Sigma_1^{-1}\Sigma_0) + (\mu_1 - \mu_0)^T\Sigma_1^{-1} \cdot (\mu_1 - \mu_0) - N - ln(\frac{det\Sigma_0}{det\Sigma_1})) \tag{2.58}$$

Therefore, the Kullback-Leibler divergence between the source signal $\underline{x}$ and the estimated signal $\underline{\hat{x}}$ could be achieved as,

$$\begin{aligned} D(p(\underline{x}) \| p(\underline{\hat{x}})) &= \frac{1}{2}(N\frac{N\sigma_x^2 + \sigma_w^2}{M\sigma_x^2} - N - \ln(\frac{\sigma_x^{2N}}{(\frac{M\sigma_x^4}{N\sigma_x^2 + \sigma_w^2})^N})) \\ &= \frac{1}{2}(N\frac{N\sigma_x^2 + \sigma_w^2}{M\sigma_x^2} - N - N\ln(\frac{N\sigma_x^2 + \sigma_w^2}{M\sigma_x^2})) \\ &= \frac{N}{2}(\frac{N\sigma_x^2 + \sigma_w^2}{M\sigma_x^2} - 1 - \ln(\frac{N\sigma_x^2 + \sigma_w^2}{M\sigma_x^2})) \end{aligned} \tag{2.59}$$

Since

$$\frac{N\sigma_x^2 + \sigma_w^2}{M\sigma_x^2} = \frac{N}{M} + \frac{\sigma_w^2}{M\sigma_x^2} > \frac{N}{M} > 1 \tag{2.60}$$

In compressive sensing, when the number of M and N are sufficiently large, $\frac{\sigma_w^2}{M\sigma_x^2} \to 0$.

We notice that when $a > 1$, then $a - 1 > \ln a$, therefore, from the analysis above, $D(p(\underline{x}) \| p(\underline{\hat{x}})) > 0$ is concluded, which satisfy the requirement of the Kullback-Leibler distance.

Hence, the upper bound of the probability of error will become

$$\begin{aligned} P_e &\leq 1-(2\pi e\sigma_x^2)^{-N/2}2^{-D(p(\underline{x})\|p(\underline{\hat{x}}))} \\ &= 1-(2\pi e\sigma_x^2)^{-N/2}2^{-\frac{N}{2}(\frac{N\sigma_x^2+\sigma_w^2}{M\sigma_x^2}-1-\ln(\frac{N\sigma_x^2+\sigma_w^2}{M\sigma_x^2}))} \\ &= 1-(\pi e\sigma_x^2)^{-N/2}2^{-\frac{N}{2}(\frac{N\sigma_x^2+\sigma_w^2}{M\sigma_x^2}-\ln(\frac{N\sigma_x^2+\sigma_w^2}{M\sigma_x^2}))} \end{aligned} \tag{2.61}$$

2.3.1.3 Numerical Results

In the simulation, we choose the signal variance normalized to $\sigma_x^2 = 1$, the sampling rate is M/N.

Fig. 2.1 shows the lower bound of the probability of error for different N with $SNR = 20dB$, from which we could observe that with the increase of source dimension N, the lower bound of the error probability becomes lower, and when the lower bound of the probability achieves 0, the sampling rate almost converges at around 0.2. When the sampling rate is less than 0.2, it is never possible for compressive sensing to perfectly recover all information, as the lower bound of the probability of error is greater than 0. When the sampling rate is larger than 0.2, the lower bound of the probability of error is non-positive, which implies that it is possible for compressive sensing to perfectly reconstruct all information.

This result provides us an reference of the probability of error of compressive sensing, when designing the measurement matrix, to make sure the perfect reconstruction of information, the sampling rate should choose no less than 0.2 when $SNR = 20dB$.

Fig. 2.2 shows the lower bound of the probability of error for different N with different Signal-to-Noise Ratio (SNR), from which we could observe that with the increase of SNR, the lower bound of the error probability becomes lower. For example, if $SNR = 0dB$, when the sampling rate $M/N \leq 0.3$, it is impossible

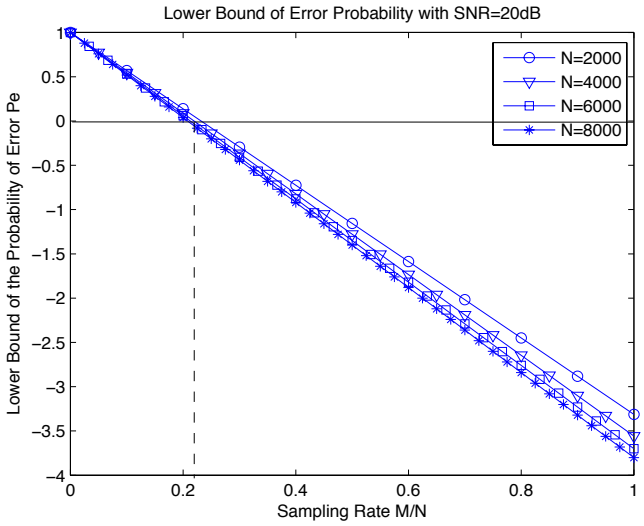

Figure 2.1. Lower Bound of P_e with $SNR = 20dB$.

to perfect reconstruct the original information, however, if the SNR is increased to $SNR = 40dB$, at the sampling rate of $M/N = 0.3$, perfect reconstruction is possible, as the lower bound of the probability of error is non-positive at this point, and perfect recovery could be achieved when $M/N \geq 0.18$. This implies that when the SNR is higher, the compression of the original information source could be very high, i.e., much less number of M is required, with the error probability performance guaranteed, while when the SNR is low, compressive sensing could not perform well, M should not be chosen very small.

Fig. 2.3 gives the relationship of SNR and the lower bound of the probability of error for compressive sensing with different sampling rate. It is obvious that when sampling rate $M/N = 0.1$, in the range of $-10dB \leq SNR \leq 40dB$, the lower bound of the probability of error is always a high positive value, which means that this compression using sampling rate $M/N = 0.1$ is too aggressive, and perfect reconstruction is impossible in this case. With the increase of the sampling rate

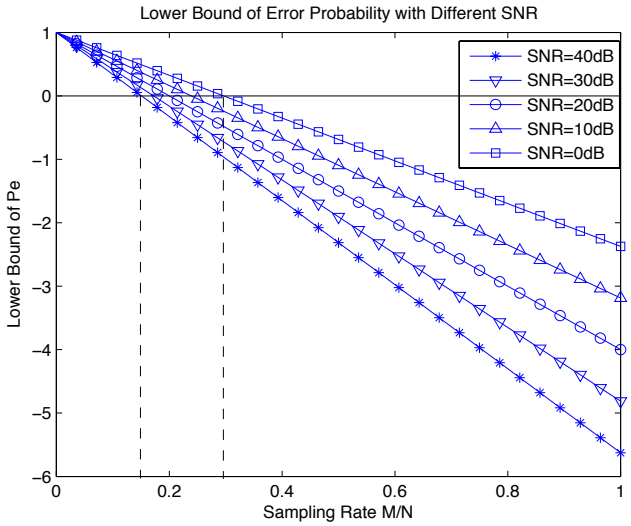

Figure 2.2. Lower Bound of P_e with Different SNR.

to $M/N = 0.2$, when $SNR \geq 20dB$, the lower bound of the probability of error is a non-positive value, and perfect recovery is possible. When $M/N = 0.3$, then if $SNR \geq 0dB$, perfect recovery is possible. We can notice that when the sampling rate is high, with $M/N = 0.4$, or $M/N = 0.5$, in the range of $-10dB \leq SNR \leq 40dB$, the lower bound of the probability of error is always a negative value, and it is always possible to perfect reconstruct the original information.

These results provide us some kind of reference of the probability of error of compressive sensing, from our analysis, we conclude that when using compressive sensing, to make sure the perfect reconstruction of information, the dimension M of the measurement matrix should be determined based on the SNR, this is a tradeoff between the sampling rate and the error probability performance.

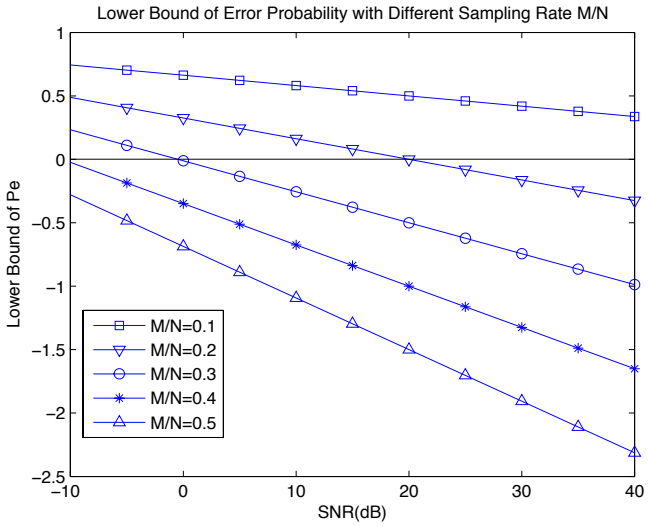

Figure 2.3. SNR v.s. P_e with Different Sampling Rate.

2.3.2 Rate Distortion Performance

The Gaussian assumption is a classical modeling assumption heavily used in areas such as Signal Processing and Communication System [41]. Assume the components of $\underline{x}$ to be identical and independent (i.i.d.) Gaussian $N(0, \sigma_x^2 I_N)$. Treat the error to be the unknown perturbation $\underline{n}$ bounded by ϵ, i.e., $\|\underline{n}\|_{l_2} \leq \epsilon$. In this chapter, we always use underline symbols to represent the random variables (r.v.s), for example, the original signal $\underline{x}$ and the compressed signal $\underline{y}$.

2.3.2.1 Rate Distortion Performance for Scalar Quantization of Measurement Observation

As we know, the random source $\underline{y} \in \mathbb{R}^M$, and the distortion associated with the quantizer [45] is $D(\mathcal{C}) = E[\| \underline{y} - q(\underline{y}) \|_2^2]$. Here, squared-error distortion is used as the distortion measure.

Let the rate of the codebook $\mathcal{C}$ be $R = \frac{1}{M}\log_2|\mathcal{C}|$. For a given code rate R, the distortion rate function is given by [45]

$$D_y(R) = \inf_{\mathcal{C}:\frac{1}{M}\log_2|\mathcal{C}|\leq R} D(\mathcal{C}) = \inf_{\mathcal{C}:\frac{1}{M}\log_2|\mathcal{C}|\leq R} E[\| \underline{y} - q(\underline{y}) \|_2^2] \tag{2.62}$$

For the quantization of compressive sensing, the quantization functions for all the coordinate of $\underline{y}$ are assumed to be the same. The corresponding distortion rate function is,

$$D_y(R) = \inf_{\mathcal{C}:\frac{1}{M}\log_2|\mathcal{C}|\leq R} E[\sum_{i=1}^{M} |y_i - q(y_i)|^2] \tag{2.63}$$

The high-rate operational distortion rate function for scalar quantization is given in [29], as $D(R) = \frac{1}{12}2^{2h(X)-2R}$, where $h(X)$ is the differential entropy of the source, for a given distortion constraint $\sum_i D_{q_i} \leq D$. For Gaussian source with zero mean and variance σ^2, $h(X) = \frac{1}{2}log(2\pi e\sigma^2)$, then $R(D) = \frac{1}{2}\log(\frac{\pi e\sigma^2}{6D})$, i.e., $D(R) = \frac{\pi e}{6}\sigma^2 2^{-2R}$. Under the fact that for high-rate uniform scalar quantization, the step size [30] is $\Delta_i = \sqrt{\frac{6\log_2 e}{\lambda}}, if \frac{1}{2\lambda} \leq \sigma_i^2$.

As shown before, $\underline{y}$ follows Gaussian distribution with $\underline{y} \sim \mathcal{N}(0, (N\sigma_x^2+\sigma_w^2)I_M)$. For scalar quantization of $\underline{y}$, its corresponding distortion rate function is given as,

$$D_y(R) = \frac{\pi e}{6}(N\sigma_x^2 + \sigma_w^2)2^{-2R} \tag{2.64}$$

uniform scalar quantization step size $\Delta_i = \sqrt{\frac{6\log_2 e}{\lambda}}$, if $\frac{1}{2\lambda} \leq N\sigma_x^2 + \sigma_w^2$.

The distortion rate function of uniform scalar quantization of a Gaussian random variable was shown in [31][28][111] as $\lim_{R\to\infty} \frac{2^{2R}}{R} D^*_{u,s,g}(R) = \frac{4}{3}\sigma^2 \log 2$.

Therefore, the distortion rate function for uniform scalar quantization[28] of the compressed signal $\underline{y} \sim \mathcal{N}(0, (N\sigma_x^2 + \sigma_w^2)I_M)$ could be expressed as,

$$\lim_{R\to\infty} \frac{2^{2R}}{R} D^*_{y,u}(R) = \frac{4}{3}(N\sigma_x^2 + \sigma_w^2)\log 2 \tag{2.65}$$

The received signal after quantization has noise introduced, indicated as $\underline{v}$, which also follows Gaussian distribution $\underline{v} \sim \mathcal{N}(0, \sigma_v^2 I_M)$. Therefore, the received signal at the decoder is

$$r_i = q(y_i) + v_i, i = 1, 2, \cdots, M \tag{2.66}$$

Theorem 6. *The measurement rate distortion function for the scalar quantization of the measurement vector of compressive sensing with random noise is*

$$R_y(D) = \sum_{i=1}^{M} (p(r_i) \log \frac{1}{p(r_i)} - \frac{1}{2} \log 2\pi e \frac{D}{M}) \tag{2.67}$$

with $p(r_i)$ shown in expression (2.68) and (2.71) for uniform quantization and non-uniform quantization of the measurement observation respectively.

(1). For Uniform Quantization

Assume the measurement vector $\underline{y}$ is uniformly quantized into L levels, where Δ is the step size. For a uniform quantizer [45][114][112], for an input signal amplitude in the range $(l-1)\Delta \leq y_i < l\Delta$, the output levels are specified as $q(y_i)_l = \frac{1}{2}(2l - 1)\Delta, l = 1, \cdots, L$.

The probability distribution function of the received signal at the decoder could be calculated as,

$$\begin{aligned} p(r_i) &= \sum_{l=1}^{L} P\{r_i | q(y_i) = \frac{(2l-1)\Delta}{2}\} P\{(q(y_i) = \frac{(2l-1)\Delta}{2}\} \\ &= \sum_{l=1}^{L} \mathcal{N}(\frac{(2l-1)\Delta}{2}, \sigma_v^2) Pr\{((l-1)\Delta \leq y_i < l\Delta)\} \\ &= \sum_{l=1}^{L} \mathcal{N}(\frac{(2l-1)\Delta}{2}, \sigma_v^2) [\mathbf{Q}(\frac{(l-1)\Delta}{\sqrt{N\sigma_x^2 + \sigma_w^2}}) + \mathbf{Q}(\frac{-l\Delta}{\sqrt{N\sigma_x^2 + \sigma_w^2}})] \end{aligned} \tag{2.68}$$

where $\mathbf{Q}(\cdot)$ is the Gaussian Q-function.

(2). For Non-uniform Quantization

Assume the measurement vector $\underline{y}$ is quantized into L non-uniform levels. For non-uniform quantization [45][114], we still represent the output levels as $q(y_i)_l$, when the input signal amplitude is in the range $y_{l-1} \leq y_i < y_l$. For an L-level quantizer, the end points are $y_0 = -\infty$ and $y_L = \infty$. The resulting distortion is

$$D = \sum_{l=1}^{L} \int_{y_{k-1}}^{y_k} f(\tilde{y}_k - y) p(y) dy \tag{2.69}$$

In the case of mean square value of the distortion,

$$y_l = \frac{1}{2}(\tilde{y}_l + \tilde{y}_{l+1}) \tag{2.70}$$

The probability distribution function of the received signal at the decoder could be calculated as,

$$\begin{aligned} p(r_i) &= \sum_{l=1}^{L} P\{r_i | q(y_i) = q(y_i)_l\} P\{(q(y_i) = q(y_i)_l\} \\ &= \sum_{l=1}^{L} \mathcal{N}(q(y_i)_l, \sigma_v^2) Pr\{(y_{l-1} \leq y_i < y_l)\} \\ &= \sum_{l=1}^{L} \mathcal{N}(q(y_i)_l, \sigma_v^2) [\mathbf{Q}(\frac{y_{l-1}}{\sqrt{N\sigma_x^2 + \sigma_w^2}}) + \mathbf{Q}(\frac{-y_l}{\sqrt{N\sigma_x^2 + \sigma_w^2}})] \end{aligned} \tag{2.71}$$

Proof. The measurement rate distortion function with squared-error distortion measure in this quantization process could be expressed as,

$$\begin{aligned} R_y(D) &= \min_{p(r|y): \sum_{(y,r)} p(y,r) d(y,r) \leq D} I(\underline{r}; \underline{y}) \\ &= \min_{p(r|y): \sum_{(y,r)} p(y,r) d(y,r) \leq D} h(\underline{r}) - h(\underline{r}|\underline{y}) \\ &\geq \min_{p(r|y): \sum_{(y,r)} p(y,r) d(y,r) \leq D} h(\underline{r}) - h(\underline{r} - \underline{y}) \\ &\geq \min_{p(r|y): \sum_{(y,r)} p(y,r) d(y,r) \leq D} \sum_{i=1}^{M} (h(r_i) - h(r_i - y_i)) \\ &\geq \min_{p(r|y): Ed(y,r) \leq D} \sum_{i=1}^{M} (h(r_i) - h(\mathcal{N}(0, E(r_i - y_i)^2))) \\ &\geq \min_{p(r|y): Ed(y,r) \leq D} \sum_{i=1}^{M} (h(r_i) - \frac{1}{2} \log 2\pi e D_i) \end{aligned} \tag{2.72}$$

where $E(r_i - y_i)^2 \leq D_i$, with $\sum_{i=1}^{M} D_i = D$, and $h(r_i) = -p(r_i)\log p(r_i)$.

The problem of finding the rate distortion function could be solved by solving the following optimization problem,

$$R_y(D) = \min_{\sum D_i = D} \sum_{i=1}^{M} \max\{(p(r_i)\log\frac{1}{p(r_i)} - \frac{1}{2}\log 2\pi e D_i), 0\} \tag{2.73}$$

Construct the function below using Lagrange multipliers,

$$J(D) = \sum_{i=1}^{M}(p(r_i)\log\frac{1}{p(r_i)} - \frac{1}{2}\log 2\pi e D_i) + \lambda\sum_{i=1}^{M} D_i \tag{2.74}$$

Differentiating with respect to D_i, and setting equal to 0, we will get

$$\frac{\partial J(D)}{\partial D_i} = -\frac{1}{2}\frac{1}{D_i} + \lambda = 0 \tag{2.75}$$

or

$$D_i = \lambda^* \tag{2.76}$$

Therefore, equal distortion for each random variable is achieved.

As $\sum_{i=1}^{M} D_i = D$, we could obtain that

$$D_i = \frac{D}{M} \tag{2.77}$$

Hence, the rate distortion function for the uniform quantization process of the measurement vector $\underline{y}$ is

$$R_y(D) = \sum_{i=1}^{M}(p(r_i)\log\frac{1}{p(r_i)} - \frac{1}{2}\log 2\pi e\frac{D}{M}) \tag{2.78}$$

with $p(r_i)$ shown in expression (2.68) and (2.71) for uniform quantization and non-uniform quantization of the measurement observation respectively. □

2.3.2.2 Reconstruction Rate Distortion - Squared Error Distortion

From CS literature, we know that the reconstruction distortion is dependent on the distortion in the measurement [28].

The received signal before reconstruction is described in equation (2.66)

$$\underline{r} = q(\underline{y}) + \underline{v} = \Phi\underline{x} + \underline{e} + \underline{v} = \Phi\underline{x} + \underline{n} \tag{2.79}$$

where $\underline{e} \in \mathbb{R}^M$ is the quantization error, and $\underline{v} \in \mathbb{R}^M$ is the noise introduced to the received signal, $\underline{n} = \underline{e} + \underline{v}$.

Theorem 7. *The reconstruction rate distortion function of noisy compressive sensing could be expressed as,*

$$R_x(D) = \sum_{i=1}^{N}(\frac{D_i N\Phi_{s,i}^2}{\epsilon^2 M}\log(\frac{2\pi e\sigma^2 D_i N\Phi_{s,i}^2}{\epsilon^2 M}) + \frac{\epsilon^2 M - 2D_i N\Phi_{s,i}^2}{2\epsilon^2 M}\log 2\pi e\sigma_x^2) \tag{2.80}$$

where

$$D_i = \begin{cases} \frac{\epsilon^2 M}{N\Phi_{s,i}^2} 2^{-[\frac{1}{\ln 2} + \lambda\frac{\epsilon^2 M}{N\Phi_{s,i}^2}]}, & \text{if } \lambda > -\frac{N\Phi_{s,i}^2}{\epsilon^2 M \ln 2} \\ \frac{\epsilon^2 M}{2N\Phi_{s,i}^2}, & \text{if } \lambda \leq -\frac{N\Phi_{s,i}^2}{\epsilon^2 M \ln 2} \end{cases} \tag{2.81}$$

where λ is chosen so that $\sum_{i=1}^{N} D_i = D$.

Proof. From noisy compressive sensing, we know that[114]

$$\|\underline{n}\|_{l_2} = \|\underline{r} - \Phi\underline{\hat{x}}\|_{l_2} \leq \epsilon \tag{2.82}$$

and $\underline{r}$ is a M-dimensional signal, the measurement matrix Φ's dimension is $M \times N$, and the dimension of $(\underline{x} - \underline{\hat{x}})$ is $N \times 1$. Using $(\cdot)_k$ to represent the $k-$th element of the $M \times 1$ vector $(\Phi(\underline{x} - \underline{\hat{x}}))$, we could get the following

$$\|\underline{r} - \Phi\underline{\hat{x}}\|_{l_2} = \sqrt{\sum_{k=1}^{M} |(\Phi(\underline{x} - \underline{\hat{x}}))_k|^2} \leq \epsilon \tag{2.83}$$

which means that

$$\sum_{k=1}^{M} |(\Phi(\underline{x} - \underline{\hat{x}}))_k|^2 \leq \epsilon^2 \tag{2.84}$$

Let $\Phi_k \in \mathbb{R}^{1\times N}$ $(k = 1, \ldots, M)$ denote the $k-$th row of the measurement matrix Φ, i.e.,

$$\Phi_{M\times N} = \begin{pmatrix} \Phi_1 \\ \Phi_2 \\ \vdots \\ \Phi_M \end{pmatrix}$$

The above inequality could be extended as[112],

$$(\Phi_1(\underline{x} - \underline{\hat{x}}))^2 + (\Phi_2(\underline{x} - \underline{\hat{x}}))^2 + \cdots + (\Phi_M(\underline{x} - \underline{\hat{x}}))^2 \leq \epsilon^2 \tag{2.85}$$

We could further derive the above inequality as

$$\begin{aligned} &\frac{(\Phi_1(\underline{x} - \underline{\hat{x}}) + \Phi_2(\underline{x} - \underline{\hat{x}}) + \cdots + \Phi_M(\underline{x} - \underline{\hat{x}}))^2}{M} \\ &\overset{(i)}{\leq} \quad (\Phi_1(\underline{x} - \underline{\hat{x}}))^2 + (\Phi_2(\underline{x} - \underline{\hat{x}}))^2 + \cdots + (\Phi_M(\underline{x} - \underline{\hat{x}}))^2 \\ &\leq \quad \epsilon^2 \end{aligned} \tag{2.86}$$

Here, (i) comes from the inequality that $a_1^2 + a_2^2 + \cdots + a_n^2 \geq \frac{(a_1+a_2+\cdots+a_n)^2}{n}$.

Extracting all the $(\underline{x} - \underline{\hat{x}})$ on the left side of the inequality,

$$\frac{((\Phi_1 + \Phi_2 + \cdots + \Phi_M)(\underline{x} - \underline{\hat{x}}))^2}{M} \leq \epsilon^2 \tag{2.87}$$

Using Φ_s to represent the summation of all the rows of Φ, i.e., $\Phi_{s,1\times N} = \Phi_1 + \cdots + \Phi_M$,

$$\frac{(\Phi_s(\underline{x} - \underline{\hat{x}}))^2}{M} \leq \epsilon^2 \tag{2.88}$$

Therefore,

$$\sum_{i=1}^{N} (\Phi_{s,i}(x_i - \hat{x}_i))^2 \leq M\epsilon^2 \tag{2.89}$$

$\Phi_{s,i}$ is the $i-$th element of $\Phi_{s,1\times N}$. Similarly, x_i and $\hat{x}_i$ are the $i-$th element of $\underline{x}_{N\times 1}$ and $\underline{\hat{x}}_{N\times 1}$. Which is equivalent to

$$\sum_{i=1}^{N} \Phi_{s,i}^2 (x_i - \hat{x}_i)^2 \leq M\epsilon^2 \tag{2.90}$$

As compressive sensing is used to compress huge amounts of data, which means that the length N of original signal $\underline{x}_{N\times 1}$ is large enough. Since the measurement matrix Φ is a randomly generated Bernoulli matrix, with the entries of $\phi_{m,n}$ i.i.d, $Pr(\phi_{m,n} = \pm 1) = 0.5$, when the dimension of the measurement matrix is sufficient large, each element $\Phi_{s,i}, i = 1, 2, \cdots, N$ of the summation of all the rows of Φ will tend to a constant $\Phi_{s,c}$.

$$\Phi_{s,c}^2 \sum_{i=1}^{N} (x_i - \hat{x}_i)^2 \leq M\epsilon^2 \tag{2.91}$$

Hence,

$$E(\underline{\hat{x}} - \underline{x})^2 \leq \frac{\epsilon^2 M}{N\Phi_{s,c}^2} \tag{2.92}$$

This provides us a numerical upper bound of the mean squared error (MSE) of compressive sensing, which depends on our choice of Φ, the dimension N of the original signal $\underline{x}$, the dimension M of the compressed signal $\underline{y}$, and the noise bound $\|n\|_{l2} \leq \epsilon$.

For scalar quantization of $\underline{y}$, the corresponding reconstruction distortion rate function is upper bounded by,

$$D_x(R) \leq \frac{\pi e 2^{-2R}}{6} \frac{\epsilon^2 M}{N\Phi_{s,c}^2} \tag{2.93}$$

Since $\underline{y} = \Phi\underline{x} + \underline{w}$, we could notice that the mean of $\underline{y}$ is $\underline{0}$, with covariance matrix C_{yy}, i.e., $\underline{y} \sim \mathcal{N}(0, (N\sigma_x^2 + \sigma_w^2)I_M)$.

The joint likelihood function is,

$$p_{y|x}(\underline{y}; \underline{x}) = \frac{1}{2\pi(N\sigma_x^2 + \sigma_w^2)^{\frac{M}{2}}} exp(-\frac{1}{2}\underline{y}^T \cdot C_{yy}^{-1} \cdot \underline{y}) \tag{2.94}$$

which is

$$p_{y|x}(\underline{y}; \underline{x}) = \frac{1}{2\pi(N\sigma_x^2 + \sigma_w^2)^{\frac{M}{2}}} \cdot exp(-\frac{1}{2(N\sigma_x^2 + \sigma_w^2)} \cdot \| \underline{y} - \Phi\underline{x} \|^2) \tag{2.95}$$

Lemma 1. For any unbiased estimator $\hat{\underline{x}}$ of compressed signal $\underline{x}$ in noisy compressive sensing with random noise $\underline{x}$, i.e., $\underline{y} = \Phi\underline{x} + \underline{w}$, the Cramer-Rao Lower Bound for the estimator $\hat{\underline{x}}$ is

$$E\{\| \hat{\underline{x}} - \underline{x} \|^2\} \geq Tr\{I^{-1}\} = (N\sigma_x^2 + \sigma_w^2)Tr\{(\Phi^T\Phi)^{-1}\} \tag{2.96}$$

Proof. Fisher's Information could be calculated as [113]

$$[I(x)]_{i,j} = -E[\frac{\partial^2 \ln p(\underline{y};\underline{x})}{\partial x_i \partial x_j}] = \frac{1}{N\sigma_x^2 + \sigma_w^2}\Phi^T\Phi \tag{2.97}$$

Therefore, the Cramer-Rao Lower Bound [113] for the estimator $\hat{\underline{x}}$ is

$$E\{\| \hat{\underline{x}} - \underline{x} \|^2\} \geq Tr\{I^{-1}\} = (N\sigma_x^2 + \sigma_w^2)Tr\{(\Phi^T\Phi)^{-1}\} \tag{2.98}$$

□

From expression (2.98) we know the Cramer-Rao Lower Bound for the estimator $\hat{x}$ is

$$E\{\| \hat{\underline{x}} - \underline{x} \|^2\} \geq var(\hat{\underline{x}}) \geq (N\sigma_x^2 + \sigma_w^2)Tr\{(\Phi^T\Phi)^{-1}\} \tag{2.99}$$

For scalar quantization of $\underline{y}$, the corresponding reconstruction distortion rate function is lower bounded by,

$$D_x(R) \geq \frac{\pi e 2^{-2R}}{6}(N\sigma_x^2 + \sigma_w^2)Tr\{(\Phi^T\Phi)^{-1}\} \tag{2.100}$$

Similarly, for uniform scalar quantization of $\underline{y}$, the corresponding reconstruction distortion rate function is bounded by,

$$\lim_{R\to\infty} \frac{2^{2R}}{R} D_{x,u}(R) \geq \frac{4}{3}(N\sigma_x^2 + \sigma_w^2)Tr\{(\Phi^T\Phi)^{-1}\}\log 2 \tag{2.101}$$

and

$$\lim_{R\to\infty} \frac{2^{2R}}{R} D_{x,u}(R) \leq \frac{4}{3}\frac{\epsilon^2 M}{N\Phi_{s,c}^2}\log 2 \tag{2.102}$$

Using the definition of rate distortion function, we can get the rate distortion function of compressive sensing,

$$\begin{aligned} R_x(D) &= \min_{p(\hat{x}|x):\sum_{(x,\hat{x})} p(x,\hat{x})d(x,\hat{x})\leq D} I(\underline{x};\underline{\hat{x}}) \\ &= h(\underline{x}) - \max_{E\{d(x,\hat{x})\}\leq D} h(\underline{x}-\underline{\hat{x}}|\underline{\hat{x}}) \\ &\geq h(\underline{x}) - \max_{E\{d(x,\hat{x})\}\leq D} h(\underline{x}-\underline{\hat{x}}) \\ &= \sum_{i=1}^{N}(h(x_i) - \max_{E\{d(x_i,\hat{x}_i)\}\leq D_i} h(x_i-\hat{x}_i)) \end{aligned} \tag{2.103}$$

For squared-error distortion, $Ed(\underline{x}^N, \underline{\hat{x}}^N) \leq D$, where $d(\underline{x}^N, \underline{\hat{x}}^N) = \sum_{i=1}^{N}(x_i - \hat{x}_i)^2$ and $\sum_{i=1}^{N} D_i = D$.

$$Ed(x_i,\hat{x}_i) = p(x_i,\hat{x}_i)(x_i-\hat{x}_i)^2 \leq D_i \tag{2.104}$$

So that

$$\max p(x_i,\hat{x}_i)(x_i-\hat{x}_i)^2 = p(x_i,\hat{x}_i)\frac{\epsilon^2 M}{N\Phi_{s,i}^2} \leq D_i \tag{2.105}$$

Therefore, we could obtain that,

$$p(x_i,\hat{x}_i) \leq \frac{N\Phi_{s,c}^2 D_i}{\epsilon^2 M} \tag{2.106}$$

There are two cases[114][112]: $\frac{N\Phi_{s,c}^2 D_i}{\epsilon^2 M} < \frac{1}{2}$, or $\frac{N\Phi_{s,c}^2 D_i}{\epsilon^2 M} \geq \frac{1}{2}$. The entropy function of a source is a concave function, which means that it will reach its global maximum value when the probability of the source equals to 1/2. If the probability is less than 1/2, the entropy function is a monotone increasing function. And if the probability is greater than 1/2, the entropy function function is a monotone decreasing function. Since compressive sensing is used to compress huge number of data in practice, the joint probability of each elements is usually very small, i.e., $p(x_i,\hat{x}_i) << 1/2$. To find the maximum value of the entropy $h(\underline{x}-\underline{\hat{x}})$ with distortion constraint D, we conclude that if $\frac{N\Phi_{s,c}^2 D_i}{\epsilon^2 M} \geq \frac{1}{2}$, $\max h(x_i-\hat{x_i}) = h(\frac{1}{2}) = \frac{1}{2}$; and if $\frac{N\Phi_{s,c}^2 D_i}{\epsilon^2 M} < \frac{1}{2}$, then

$$\max h(x_i-\hat{x_i}) = h(\frac{N\Phi_{s,c}^2 D_i}{\epsilon^2 M}) = -\frac{N\Phi_{s,c}^2 D_i}{\epsilon^2 M}\log(\frac{N\Phi_{s,c}^2 D_i}{\epsilon^2 M}) \tag{2.107}$$

Therefore, the corresponding reconstruction rate distortion function of noisy compressive sensing is,

$$\begin{aligned} R_x(D) &\geq \sum_{i=1}^{N}(h(x_i) - \max_{E\{d(x_i,\hat{x}_i)\}\leq D_i} h(x_i - \hat{x}_i)) \\ &= \sum_{i=1}^{N}(\frac{\log 2\pi e\sigma_x^2}{2} + \frac{N\Phi_{s,c}^2 D_i}{\epsilon^2 M}\log(\frac{N\Phi_{s,c}^2 D_i}{\epsilon^2 M})) \\ &\quad s.t. \quad \sum_{i=1}^{N} D_i = D \end{aligned} \tag{2.108}$$

Using Lagrange multipliers to construct the function,

$$J(D) = \sum_{i=1}^{N}(\frac{\log 2\pi e\sigma_x^2}{2} + \frac{N\Phi_{s,c}^2 D_i}{\epsilon^2 M}\log(\frac{N\Phi_{s,c}^2 D_i}{\epsilon^2 M})) + \lambda\sum_{i=1}^{N} D_i \tag{2.109}$$

Differentiating with respect to D_i, and setting equal to 0,

$$\frac{\partial J(D)}{\partial D_i} = \frac{N\Phi_{s,c}^2}{\epsilon^2 M}\log(\frac{N\Phi_{s,c}^2 D_i}{\epsilon^2 M}) + \frac{N\Phi_{s,c}^2}{\epsilon^2 M\ln 2} + \lambda = 0 \tag{2.110}$$

Solving the equation above will get D_i as,

$$D_i = \frac{\epsilon^2 M}{N\Phi_{s,c}^2} 2^{-[\frac{1}{\ln 2} + \lambda\frac{\epsilon^2 M}{N\Phi_{s,c}^2}]}, s.t. \sum_{i=1}^{N} D_i = D \tag{2.111}$$

λ could be solved by this condition,

$$\frac{\epsilon^2 M}{N\Phi_{s,1}^2} 2^{-[\frac{1}{\ln 2} + \lambda\frac{\epsilon^2 M}{N\Phi_{s,1}^2}]} + \cdots + \frac{\epsilon^2 M}{N\Phi_{s,N}^2} 2^{-[\frac{1}{\ln 2} + \lambda\frac{\epsilon^2 M}{N\Phi_{s,N}^2}]} = D \tag{2.112}$$

If $\max h(x_i - \hat{x_i}) = h(\frac{1}{2}) = \frac{1}{2}$, set $\frac{N\Phi_{s,i}^2 D_i}{\epsilon^2 M} = \frac{1}{2}$, we will get the choice of $\lambda = -\frac{N\Phi_{s,i}^2}{\epsilon^2 M\ln 2}$.

From all the above derivation and with λ obtained, we can get the rate distortion function of compressive sensing as shown in Theorem 7. □

As $\underline{x}$ follow i.i.d. Gaussian distribution $N(0, \sigma_x^2 I_N)$, assume $\sigma_x^2 > \frac{1}{2\pi e}$, i.e., $\log(2\pi e\sigma_x^2) > 0$.

From the above result, we could conclude that if D_i falls in the following range, the minimum rate in equation (2.80) is a positive value, which means that we could not use any bit rate less than this minimum rate.

$$\frac{\epsilon^2 M}{2\pi e\sigma^2 N\Phi_{s,i}^2} < D_i < \frac{\epsilon^2 M}{2N\Phi_{s,i}^2} \tag{2.113}$$

This conclusion is very important, for the reason that it provides a theoretical minimal useful bit rate, i.e., a meaningful lowest data rate for noisy compressive sensing with particular distortion constraint. And this has great significance in the design and transmission in real application of compressive sensing.

Theorem 8. *The reconstruction rate distortion function[114][110][109] with uniform scalar quantization of the measurement observation for noisy compressive sensing is*

$$R_{x,u}(D) = \sum_{i=1}^{N}(\frac{4D_i N\Phi_{s,i}^2}{M^2\Delta^2}\log(\frac{8\pi e\sigma^2 D_i N\Phi_{s,i}^2}{M^2\Delta^2}) + \frac{M^2\Delta^2 - 8D_i N\Phi_{s,i}^2}{2M^2\Delta^2}\log 2\pi e\sigma_x^2) \tag{2.114}$$

where

$$D_i = \begin{cases} \frac{M^2\Delta^2}{4N\Phi_{s,i}^2} 2^{-[\frac{1}{\ln 2}+\lambda\frac{M^2\Delta^2}{4N\Phi_{s,i}^2}]}, & \text{if } \lambda > -\frac{4N\Phi_{s,i}^2}{M^2\Delta^2\ln 2} \\ \frac{M^2\Delta^2}{8N\Phi_{s,i}^2}, & \text{if } \lambda \le -\frac{4N\Phi_{s,i}^2}{M^2\Delta^2\ln 2} \end{cases} \tag{2.115}$$

where Δ is the step size, with $|y_i - q(y_i)| \in [\frac{-\Delta}{2}, \frac{\Delta}{2})$, λ is chosen so that $\sum_{i=1}^{N} D_i = D$.

Proof. If we assume the received signal after quantization has no other noise introduced, i.e., $\underline{v} = 0$, therefore, the signal at the decoder will be, $\underline{r} = q(\underline{y}) = \Phi\underline{x} + \underline{e}$, where $\underline{e}$ is the quantization error.

$$\|\underline{n}\|_{l_2} = \|\underline{r} - \Phi\hat{\underline{x}}\|_{l_2} = \|q(\underline{y}) - \Phi\hat{\underline{x}}\|_{l_2} \le \epsilon \tag{2.116}$$

For uniform scalar quantization, it is well known that,

$$|y_i - q(y_i)| \in [\frac{-\Delta}{2}, \frac{\Delta}{2}) \tag{2.117}$$

Therefore,

$$\|\underline{y} - q(\underline{y})\|_{l_2} = \sqrt{\sum_{M}|y_i - q(y_i)|^2} \le \sqrt{M(\frac{\Delta}{2})^2} = \frac{\Delta}{2}\sqrt{M} \tag{2.118}$$

In this case, the quantization error $\underline{e}$ is treated as noise $\underline{n}$, and the decoding is to reconstruct the source signal from $\underline{r} = q(\underline{y}) = \Phi\underline{x} + \underline{e}$ with constraint that $\|\underline{n}\|_{l_2} \leq \epsilon$, hence, it could be concluded that $\epsilon = \frac{\Delta}{2}\sqrt{M}$.

Therefore, the reconstruction rate distortion function of uniform scalar quantization will be expressed as Theorem 8. □

Similarly, if D_i falls in the following range, the minimum rate in equation (2.114) is always a positive value.

$$\frac{M^2\Delta^2}{8\pi e\sigma^2 N\Phi_{s,i}^2} < D_i < \frac{M^2\Delta^2}{8N\Phi_{s,i}^2} \tag{2.119}$$

(1). For Uniform Quantization

Table 2.1 shows Max(1960)'s results [45] for optimum step sizes of uniform quantization for a Gaussian random variable.

Table 2.1. Optimum Step Sizes for Uniform Quantization of a Gaussian Random Variable

output levels	$\triangle_{opt}$	D_{min}	$10logD_{min}(dB)$
2	1.596	0.3634	-4.4
4	0.9957	0.1188	-9.25
8	0.5860	0.03744	-14.27
16	0.3352	0.01154	-19.38
32	0.1881	0.00349	-24.57

The optimum step sizes' $\triangle_{opt}$ values for different numbers of output levels (L) could be read from Table 2.1. For example, for the uniform quantization of the compressed signal $\underline{y}$, Gaussian r.v.s with zero mean and unit variable, $\sigma_y^2 = N\sigma_x^2 + \sigma_w^2 = 1$, with the quantization level $L = 8$, ϵ equals to $0.293\sqrt{M}$, the corresponding reconstruction rate distortion function for this eight-level uniform quantized compressive sensing could be achieved.

(2). For Non-uniform Quantization

Table 2.2 shows Max(1960)'s results [45] for the optimum eight-level quantizer of a Gaussian distributed signal amplitude with zero mean and unit variance. It is obvious that, the step size of non-uniform quantizer is much less than that of uniform quantizer for smaller signal; and for larger signal, the case is opposite. Here, only small signal, i.e., $k = 1, 2, \cdots, L-1$, is considered, except the cases when $k = 0$ and $k = L$.

Table 2.2. Optimum Eight-level Quantizer for a Gaussian Random Variable

level k	x_k	$\tilde{x}_k$
1	-1.748	-2.152
2	-1.050	-1.344
3	-0.5006	-0.7560
4	0.0	-0.2451
5	0.5006	0.2451
6	1.050	0.7560
7	1.748	1.344
8	∞	2.152

For eight-level non-uniform quantization of zero mean unit variance Gaussian r.v.s, when $k = 1, 2, \cdots, 7$,

$$|Y - Y_q| \leq \max\{\frac{-1.344 - (-1.748)}{2}, \frac{(-0.7560) - (-1.050)}{2}, \cdots\} = 0.202 \quad (2.120)$$

$$\|Y - Y_q\|_{l_2} = \sqrt{\sum_M |Y - Y_q|^2} \leq 0.202\sqrt{M} \quad (2.121)$$

In this case, $\epsilon = 0.202\sqrt{M}$.

It is clear that for eight-level uniform quantization, $\epsilon = 0.293\sqrt{M}$, and for eight-level non-uniform quantization, $\epsilon = 0.202\sqrt{M}$. As $0.202 < 0.293$, and *log* func-

tion is a monotone increasing function, the lower bound of $R(D)$ for non-uniform quantization is smaller than that of uniform quantization, which indicates that using non-uniform quantization, a much lower bit rate for compression or transmission for compressive sensing could be used than uniform quantization under our assumption that the original source $\underline{x}$ follows Gaussian distribution. This property also proves that non-uniform quantization is much more efficient in the application of compressive sensing[114].

2.3.2.3 Reconstruction Rate Distortion - Hamming Distortion

Hamming distortion is worth analyzing in compressive sensing, as for the sparse source (for example, in cognitive ratio, for unlicensed users, the spectrum is sparse), the support detection is also very important.

Using the definition of rate distortion function, we can get the rate distortion function as,

$$\begin{aligned} R(D) &= \min_{p(\hat{x}|x):\sum_{(x,\hat{x})} p(x,\hat{x})d(x,\hat{x})\leq D} I(\underline{x};\underline{\hat{x}}) \\ &= \min_{p(\hat{x}|x):\sum_{(x,\hat{x})} p(x,\hat{x})d(x,\hat{x})\leq D} h(\underline{x}) - h(\underline{x}|\underline{\hat{x}}) \end{aligned} \tag{2.122}$$

From Fano's Inequality, we know that for the compressive sensing estimator $\underline{\hat{x}}$ such that $\underline{x} \to \underline{y} \to \underline{\hat{x}}$, with $Pe = Pr(\underline{x} \neq \underline{\hat{x}})$, we have

$$1 + Peh(\underline{x}) \geq h(\underline{x}|\underline{\hat{x}}) \tag{2.123}$$

For Hamming distortion, we know that $D = Ed(\underline{x}, \underline{\hat{x}}) = Pr(\underline{x} \neq \underline{\hat{x}}) = Pe$, which could result this

$$1 + Dh(\underline{x}) \geq h(\underline{x}|\underline{\hat{x}}) \tag{2.124}$$

Therefore, we have

$$\begin{aligned} I(\underline{x};\hat{\underline{x}}) &= H(\underline{x}) - H(\underline{x}|\hat{\underline{x}}) \\ &\geq h(\underline{x}) - (1 + Dh(\underline{x})) \\ &= \frac{1-D}{2}\log(2\pi e\sigma_x^2)^N - 1 \end{aligned} \tag{2.125}$$

We could conclude that using Hamming distortion, the rate distortion is

$$R(D) \geq \frac{1-D}{2}\log(2\pi e\sigma_x^2)^N - 1 \tag{2.126}$$

2.3.2.4 Reconstruction Rate Distortion - Upper Bound

As we know, rate is always less than the capacity for a valid communication system, i.e.,

$$R(D) < C = \frac{1}{2}\log(1 + \frac{S}{N}) \tag{2.127}$$

Assume the components of X be i.i.d. Gaussian $N(0, \sigma^2)$. Because X is K sparse, we can get that the efficient power for X is $S = K\sigma^2$. And the noise power equals to the total distortion D. From this analysis, we can get the upper bound of rate distortion function $R(D)$,

$$R(D) < \frac{1}{2}\log(1 + \frac{K\sigma_x^2}{D}) \tag{2.128}$$

2.4 Applications of CS in Big Data

Some applications of compressive sensing in Big Data will be shown in this Section.

First, the compression of real Synthetic Aperture Radar (SAR) Raw Data will be discussed. Due to the low computational resources of the acquisition platforms and the steadily increasing resolution of SAR systems, huge amounts of data are collected and stored, which cannot generally be processed on board and must be

transmitted to the ground to be processed and archived. we test the sparsity of the real SAR raw data (obtained by University of Kansas in Greenland, 2010), compress it using compressive sensing, and then recover the original signal using several CS recovery algorithms, and compare these methods' performance. This is of great significance to help us perform further research in the applications of CS to real SAR raw data.

Radar sensor networks (RSNs) have been recently considered to overcome the performance degradation of a single radar. Due to the expansion of data introduced to RSNs, the compression of received data is a design challenge of future RSNs. A new efficient and effective sampling method for real Radar Sensor Networks based on compressive sensing is also introduced in this Section. Numerical results will show that our proposed algorithm's performance is more efficient with correlations among the sensor data considered, without introduce any computation complexity. With more sensor nodes, our algorithm is more efficient, which significantly reduces the data gathered.

Due to limited bandwidth, poor-quality communication channels, and limited storage capacity for underwater acoustic sensor nodes in underwater acoustic sensor networks (UWASNs), compressive sensing was applied to UWASNs to compress the UWASNs data to enable storage and transmission requirements. The rate distortion performance in UWASNs was studied, with or without correlation among sensor readings. The compression of real-world underwater acoustic sensor network data was applied to verify the theoretical results derived.

2.4.1 Synthetic Aperture Radar (SAR)

This work addresses the use of compressive sensing to compress real Synthetic Aperture Radar (SAR) raw data. Due to the low computational resources of the

acquisition platforms and the steadily increasing resolution of SAR systems, huge amounts of data are collected and stored, which cannot generally be processed on board and must be transmitted to the ground to be processed and archived. Although compressive sensing (CS) has been proposed and studied by a lot of researchers, almost none of them touches the real application of it. While, in this section, we test the sparsity of the real SAR raw data (obtained by University of Kansas in Greenland, 2010), compress it using compressive sensing, and then recover the original signal using several CS recovery algorithms (Basis Pursuit, Matching Pursuit and Orthogonal Matching Pursuit), and compare these methods' performance. Simulation results are presented to prove the successful application of CS to real SAR raw data. When proper sparsity matrix is chosen, the real SAR data could be transformed to sparse signal. Using our designed algorithm, the positions and the exact values of the SAR raw data can be almost perfectly recovered with a very low MSE at a sampling ratio of 1/8. This is of great significance to help us perform further research in the applications of CS to real SAR raw data.

Synthetic Aperture Radar (SAR) [47] is a well established imaging technology that has become critical for remote sensing and observation applications. SAR exploits the motion of a moving platform, such as a plane or a satellite, to create a synthetic aperture much larger than the real aperture of the antenna carried by the moving platform. This allows SAR systems to image very large land swaths with extremely high precision.

The success of SAR has resulted in a demand for increasing resolution and fidelity from the imaged area [48]. One of the main challenges in satisfying this demand is dealing with the huge amount of data acquired. It is necessary to develop effective compression algorithms that can efficiently and effectively compress this huge amount of SAR data. Besides, an important challenge with typical satellite-borne

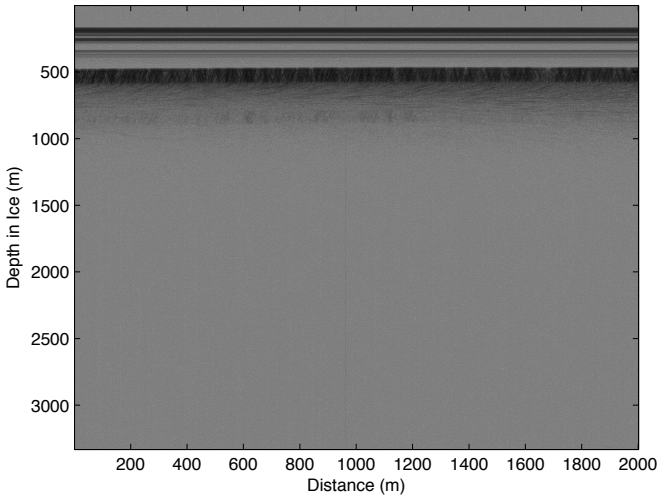

Figure 2.4. SAR image of University of Kansas - Greenland - 2010.

SAR data compression is that the acquired data is not processed on the satellite, but transmitted to a ground station for processing. The reason is that the data processing step, is hard to be performed on-board. Therefore, appropriate techniques, which compress the SAR raw data to enable storage and transmission requirements, have to be developed. A significant consideration for such algorithms is their computational complexity, as they are required to operate using the limited hardware on-board a satellite.

The state-of-the-art in SAR raw data compression is driven by the requirement of high compression ratio at low complexity. And many efforts have been made to develop suitable compression techniques [52] of the bit stream necessary for SAR raw data coding [49], such as adaptive scalar [50] and vector quantization [51] techniques. Transform coding algorithms including discrete cosine transform (DCT) and wavelet transform (WT) based methods [54] [55] also have been used to compress the SAR data by many researchers in recent years.

However, these traditional compression methods [52] are based on Nyquist rate − sampling at a high rate and then eliminating redundancy by compressing a large amount of data, which will have poor efficiency in terms of both sampling rate and computational complexity. Unlike traditional compression methods, compressive sensing provides us a new point of view, with a high compression ratio, which could compress the real SAR data significantly with small computational complexity. The key point of compressive sensing is to exploit redundancy in the data modeled as sparsity in an appropriate basis.

Although compressive sensing (CS) has been proposed and studied by a lot of researchers, almost none of them touches the real application of it. While, in this section, we test the sparsity of the real SAR raw data, compress it using compressive sensing, and then recover the original signal using several CS recover algorithms (Basis Pursuit, Matching Pursuit and Orthogonal Matching Pursuit), and compare these methods' performance.

The real SAR raw data we used in this section was obtained by University of Kansas from Greenland island in 2010. Here we express our gratitude to University of Kansas for sharing these SAR raw data for our research. They developed and deployed a wideband 8-channel synthetic aperture radar (SAR) at Summit Camp, Greenland (72.5783° N and 38.4596° W)[53]. We used the channel 2's SAR raw data obtained in 2010, $Data\_20100510\_07.mat$. Figure 1 shows the SAR raw data image of channel 2. The vertical axis represents the depth in ice, while the horizontal axis represents the distance the radar travels.

Our simulation results are presented to prove the successful application of CS to SAR raw data. Using our designed algorithm, we could almost perfectly recover the SAR raw data at a compression ratio of 1/8. This is of great significance to help us do further research in CS and the real application of CS. Besides that, this

research provides some real instruction of apply CS to compress huge amount of SAR raw data.

For the sparse properties of DWT, we use wavelet transform to make the real SAR raw data sparse. The most commonly used set of discrete wavelet transforms was formulated by the Belgian mathematician Ingrid Daubechies in 1988 [56]. The Daubechies wavelets are a family of orthogonal wavelets defining a discrete wavelet transform and characterized by a maximal number of vanishing moments for some given support. With each wavelet type of this class, there is a scaling function which generates an orthogonal multiresolution analysis. This formulation is based on the use of recurrence relations to generate progressively finer discrete samplings of an implicit mother wavelet function; each resolution is twice that of the previous scale. In our research, we use 6-level "Daubechies" wavelet decomposition of the real SAR raw data.

As random matrices are largely incoherent [18] with any fixed basis Ψ. Select an orthobasis Φ uniformly at random, which can be done by orthonormalizing n vectors sampled independently and uniformly on the unit sphere. Then with high probability, the coherence between Ψ and Φ is about $\sqrt{2}\log n$. By extension, random waveforms with independent identically distributed (i.i.d.) entries, e.g., Gaussian or ± 1 binary entries, will also exhibit a very low coherence with any fixed representation Ψ. Therefore, in this section, we choose Gaussian Random Matrix as the measurement matrix Φ.

In this section, we use three different reconstruction algorithms to recover the real SAR raw data: basis pursuit (BP), matching pursuit (MP), and orthogonal matching pursuit (OMP).

Basis pursuit [57] is the mathematical optimization problem of the form:

$$min\|x\|_1 \; subject \; to \; y = Ax \tag{2.129}$$

where x is a $N \times 1$ solution vector, y is a $M \times 1$ vector of observations, A is a $M \times N$ transform matrix, and $M < N$.

Basis Pursuit is usually applied in cases where there is an underdetermined system of linear equations $y = Ax$ that must be satisfied exactly, and the sparsest solution in the l_1 sense is desired. It can be thought of as a least square problem with an l_1 regularizer.

Matching pursuit (MP) was proposed in [58] and [59]. Matching pursuit is a type of numerical technique which involves finding the "best matching" projections of multidimensional data onto an over-complete dictionary D.

The first extension of Matching Pursuit (MP) is its orthogonal version : Orthogonal Matching Pursuit (OMP) [60]. The main difference with MP is that coefficients are the orthogonal projection of the signal on the dictionary D.

2.4.1.1 Algorithm

The procedure of our algorithm can be described as follows,

Step 1: Obtain the original SAR raw datas's amplitude x;

Step 2: Perform a DWT to the signal x, to get the sparse signal x_s.

Step 3: Using a random measurement matrix (Random Gaussian Matrix, random binary matrix, etc.) to compress the sparse signal x_s and get the compressed data y.

Step 4: Recover sparse $\hat{x}_s$ from y using several recovery algorithms (BP, MP or OMP).

Step 5: Get the recovered SAR signal $\hat{x}$ using the inverse wavelet transform.

Step 6: Compare the MSE (Mean Square Error) of the original SAR raw data x and recovered $\hat{x}$; original sparse signal x_s and the recovered $\hat{x}_s$.

2.4.1.2 Simulation Results

Because the phase of the SAR raw data is either 2.5656 or -0.5760 in radians, in the data process, we can use a particular method to store these phases. In this part, we get the amplitude of the SAR raw data to do the compression and reconstruction. In Figures 2.5, and 2.6, we give an example of the sparsity properties of the SAR raw data. Here, the total number of the original data $N = 3300$, $M = N/8$ and $K = 79$ (N is the original SAR raw data's dimension, M is the compressed data's dimension, while K is the number of non-zero elements, which satisfies $M \geq K \log(N/K))$, with the compression ratio equaling to 1/8. Figure 2.6 shows the recovered sparse signal using the Basis Pursuit(BP).

From our results, we observe that after the wavelet transform, the original SAR raw data is transformed to K non-zero elements, $K = 79$. Here we use Random Gaussian Matrix as the random measurement matrix to do the compression. In the reconstruction process, comparing Figures 2.5 and 2.6, we find both the positions and the exact values of the K non-zero elements can be almost perfectly recovered, with very low MSE as shown in the later part of this section.

Using the algorithm discussed above, in Figures 2.7, 2.8, 2.9, and 2.10 we show SAR data's compression and reconstruction performance. Figure 2.7 gives the original SAR raw data, while the other three figures show the reconstructed SAR raw data using BP, MP and OMP. Here, the three figures all use compression ratio of 1/8. We can analyze from these figures that using our algorithm, almost all these three reconstruction methods work well for this real SAR raw data, and BP and

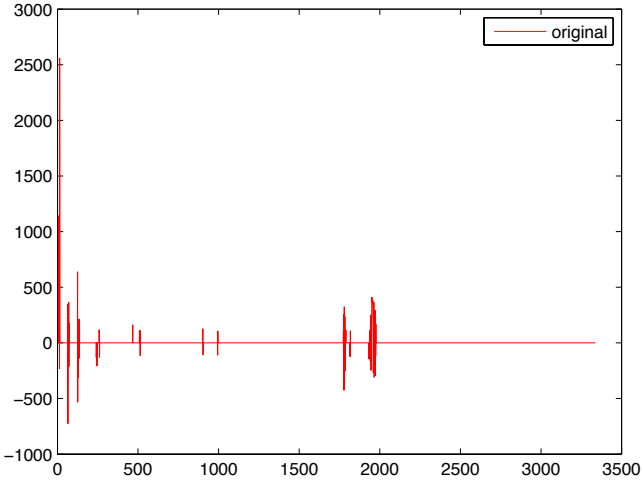

Figure 2.5. SAR signal in sparse basis-Original.

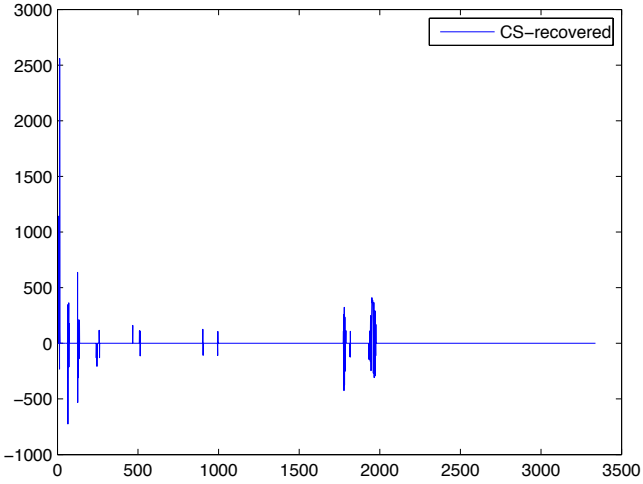

Figure 2.6. SAR signal in sparse basis-Recovered.

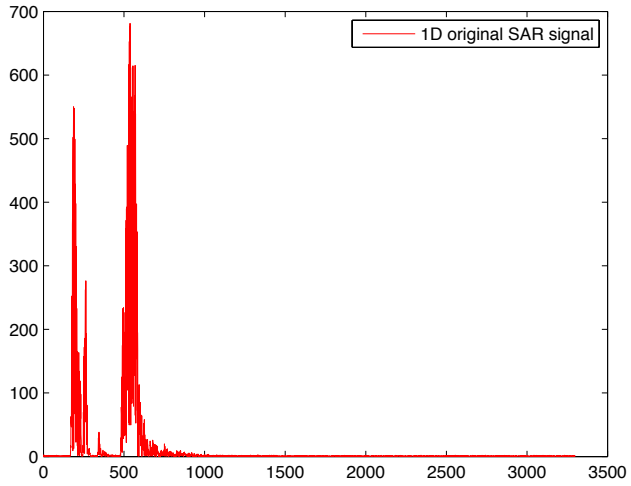

Figure 2.7. SAR signal - Original.

OMP give a little better performance than MP, because we could notice there are some noise features in the recovered signal using MP.

In this part, we simulate SAR raw data using three different recovery algorithms − basis pursuit (BP), matching pursuit (MP), and orthogonal matching pursuit (OMP). In this simulation process, we did not use a large number of samples, and thus time saving looks not that big; while in real applications, facing to a huge amount of SAR data, time cost is really an essential element for us to consider. Besides the time cost, reconstruction performance is also very important. From the elapsed CPU time shown in Table 2.3, we can observe that BP costs the longest time to implement. MP and OMP use much less time. Here, the elapsed CPU time [1] was obtained using tic and toc in MATLAB2010. While from the performance point of view, we notice that MP gives the worst performance, while BP and OMP are much better. We could also observe from these figures that, BP would achieve

[1]Windows Vista, 32-bit Operating System with Intel Core 2 Duo CPU E8600 @3.33GHz and 4.00 GB RAM

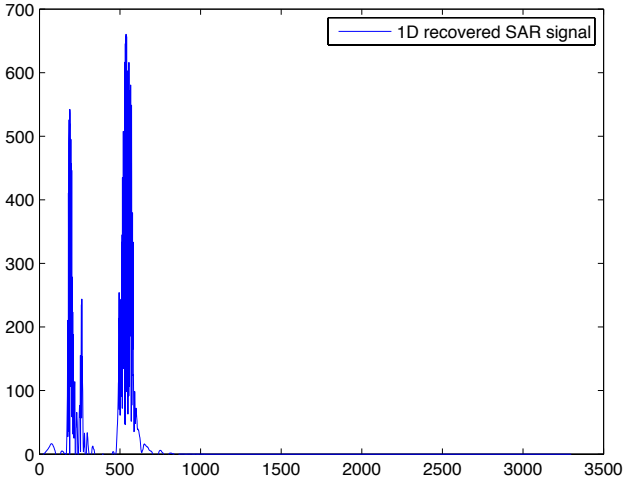

Figure 2.8. SAR signal recovered-BP.

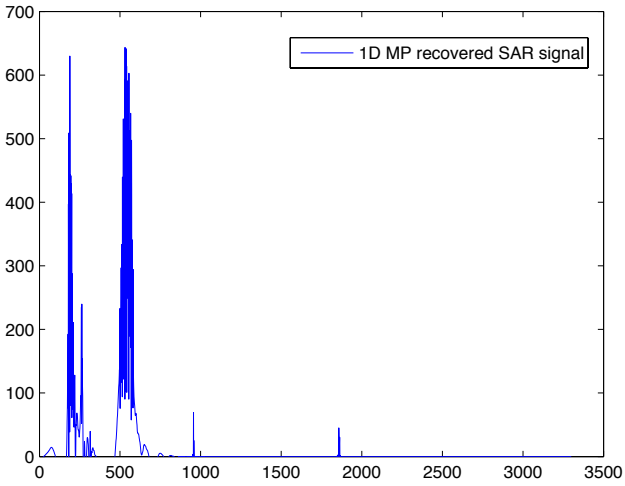

Figure 2.9. SAR signal recovered-MP.

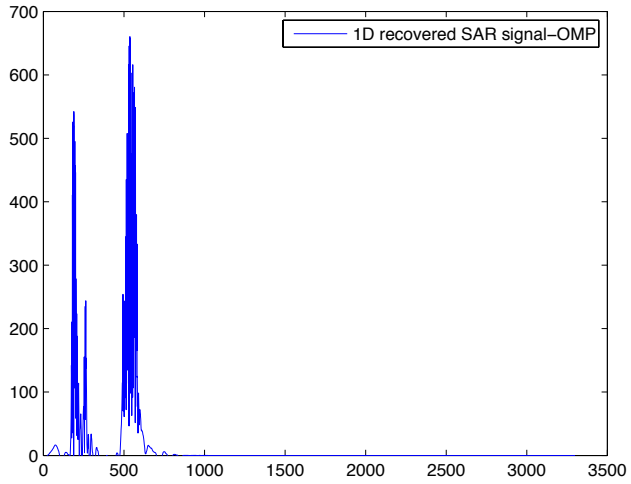

Figure 2.10. SAR signal recovered-OMP.

good reconstruction performance when the compression ratio is greater than 0.15 (with -60dB MSE), MP is nearly 0.18 (with -60dB MSE), while OMP is 0.12 (with MSE as low as -270dB). Considering both aspects of time consumption and MSE performance, we could conclude OMP is the best algorithm for SAR data recovery in these three methods.

This comparison of different reconstruction algorithms is significantly meaningful, as it can give us further instructions in the compression and reconstruction of a huge amount of real SAR raw data.

Table 2.3. Elapsed CPU time of three different recovery algorithms for SAR raw data: BP, MP, and OMP

Methods	BP (seconds)	MP (seconds)	OMP (seconds)
Elapsed CPU time	133.327777	61.399212	27.507376

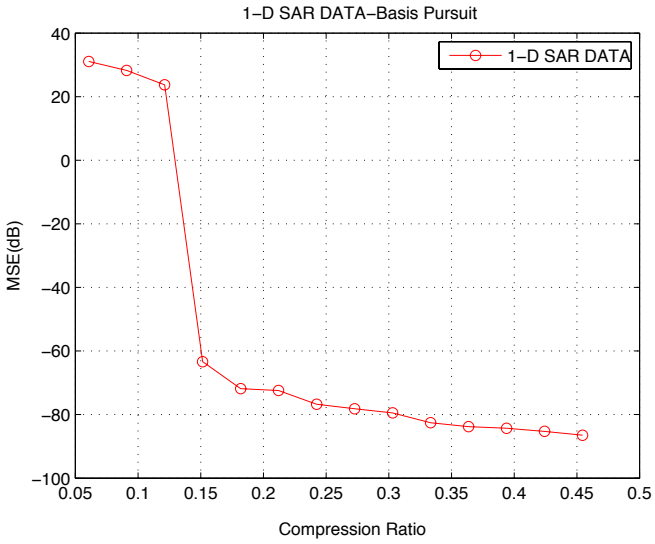

Figure 2.11. SAR data reconstruction performance-BP .

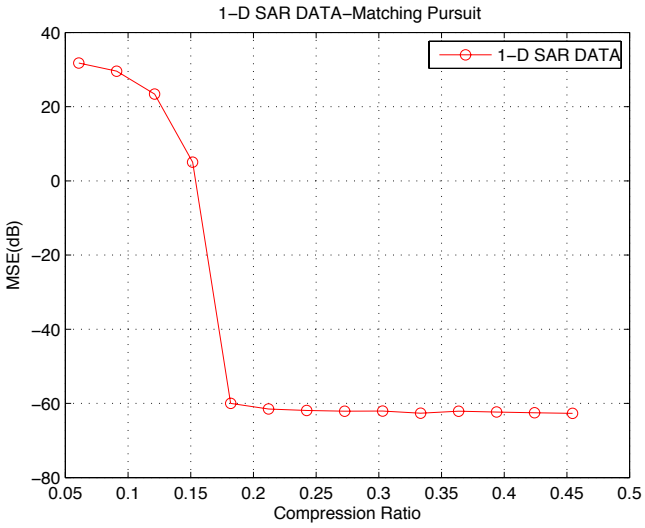

Figure 2.12. SAR data reconstruction performance-MP .

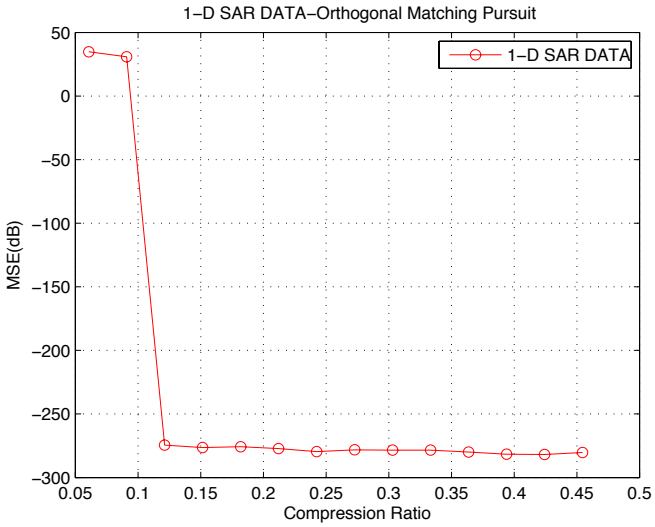

Figure 2.13. SAR data reconstruction performance-OMP .

2.4.2 Radar Sensor Networks (RSNs)

Radar sensor networks (RSNs) have attracted a lot of interest in many civilian, military, biomedical applications, and academia [80] [67] [82] [83]. RSNs consists of multiple collaboratively operating radar sensors which have capabilities of sensing, signal processing and wireless communication [62]. Radar sensors are deployed ubiquitously and networked together in an ad-hoc form to perform a lot of tasks such as surveillance, battlefield, disaster relief, search and rescue, etc. Realistic RSNs are described in details in the literature [61]. Each radar sensor in a RSNs is monostatic and contains only one transmitting and receiving element. Compared to a single radar sensor, RSNs has advantages in improving the system sensitivity, reducing obscuration effects and vulnerability, and increasing the detection performance.

However, some challenging problems such as energy efficiency, interference suppression, networking between sensors, and reducing communication cost need to be considered. And a lot of work have been done to try to solve these problems [72]

[76] [66] [67]. The importance of global communication cost reduction is obvious because radar sensor networks are typically composed of hundreds to thousands of radar sensors, generating tremendous amount of radar sensor data. Due to the resource limitations of sensor nodes [63](processing, memory, bandwidth and energy, etc), the gathered data of sensor nodes has to be compressed quickly and precisely for transmission and recovery. It is necessary to develop effective compression algorithms that can efficiently and effectively compress these huge amounts of highly redundant data [79].

The states of art radar systems [68] apply a large bandwidth and an increasing number of channel produce huge amount of data, based on Nyquist rate - sampling at a high rate and then eliminating redundancy by compressing a large amount of data, which will have poor efficiency in terms of both sampling rate and computational complexity. Achieving adequate wideband radar signal (which can be compressed into a short duration pulse) requires both a high sampling frequency and a large dynamic range. Currently available A/D conversion technology is a limiting factor in the design of ultra wideband radar systems, because in many cases the required performance is either beyond what is technologically possible or too expensive. Moreover, radar sensor networks (RSNs) have been recently considered to overcome the performance degradation of a single radar. Due to the expansion of data introduced to RSNs, the compression of received data is a design challenge of future RSNs. The emerging theory of Compressive Sensing (CS) [77][26] has provided a new framework for signal acquisition and it has gained increased interests over the past few years. CS is an alternative to the Nyquist rate for the acquisition of 'sparse' signals. Consider a signal that is sparse in some basis (often using a wavelet-based transform coding scheme).The basic idea of compressive sensing is projecting the high dimensional signal onto a measurement matrix, which is incoherent with the sparsifying basis,

resulting to a low dimensional sensed sequence. Then with a relatively small number of appropriately designed projection measurements, the underlying signal may be recovered exactly [18]. In contrast to the common framework of first collecting as much data as possible and then discarding the redundant data by digital compression techniques, CS seeks to minimize the collection of redundant data in the acquisition step.

As the nodes in sensor networks have limited computing ability and storage capacity, some researchers focus on finding new data compression algorithm. For example, [63] proposed a sampling frequency control algorithm which adjusted the sampling frequency on sensed data dynamically, and a data compression algorithm was adopted to reduce the amount of transmitted data. [64], proposed to compress the historical information from the sensor nodes using the Adaptive Learning Vector Quantization (ALVQ) algorithm, which captured the prominent features of the data and further compress all the other data with these features. In [65], the authors proposed a Compressed Kalman Filtering (CKF) algorithm to reduce the transmitted data for wireless sensor networks. CS has the advantage of compress the redundant data into a much smaller amount, therefore, a lot of researchers have already begun to apply CS to compress the tremendous amount of radar sensor data in RSNs. However, most existing work did not fully consider the correlations among sensors. Taking full advantage of the correlations among the sensor data is a demanding approach to reduce the cost of communication. In this section, we mainly consider how to compress the huge amount of data gathered by multiple sensors in RSNs. Although there are some existing work trying to compress the tremendous amount of data in RSNs, as we will show, their results are not that efficient. In [74], the authors compress large-scale data in wireless sensor networks (WSNs), while, they directly let all sensors' reading form a large vector to compress, without considering the

correlation between sensor signals. Giorgio [73] recently proposed a joint Principal Component Analysis (PCA) and Compressive Sensing (CS) algorithm to sense and compress the large data sets in real WSN scenarios. However, we will show in this section that this algorithm is not efficient although it considers the correlation of sensor signals. In [75], the proposed joint recovery for distributed compressive sensing is a good and mature method for sensor networks. However, this recovery method uses the optimization of both common component and innovation component, which introduces more computation complexity in the recovery process. [79] assume that only a subset of all the sensors is active at each sampling instant, it is possible to estimate the sample values at all the sensors. While this result is based on certain assumptions, which is not always realistic in real case.

In this section, we propose a new efficient and effective sampling method for real Radar Sensor Networks based on compressive sensing. Our algorithm neither requires any new optimization method, nor needs complex pre-processing before compression. We will compare our algorithm's performance and complexity with some existing work, such as joint PCA & CS, DCS, and traditional CS. Numerical results will show that our proposed algorithm's performance is more efficient with correlations among the sensor data considered, without introduce any computation complexity. With more sensor nodes, our algorithm is more efficient, which significantly reduces the data gathered. Besides that, our algorithm requires no new recovery algorithm, and regular recovery of CS is enough. This is a big advantage of our algorithm, especially for the hardware design.

2.4.2.1 Radar Sensor Networks Description

As lower frequencies have better penetrating properties. UWB radar uses a large spectrum in combination with lower frequencies which makes it suitable for

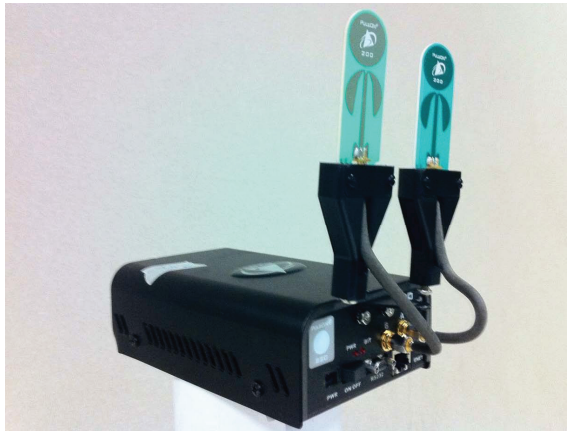

Figure 2.14. Radar - P220 in Monostatic mode.

applications such as ground penetrating radar, foliage penetrating radar [83] and short-range radar to detect hidden objects behind walls. This penetration property is also of great importance for indoor location systems.

The RSNs signals we use in this section are obtained from real-world experiment. Although a lot of researchers did related work on the signal sampling in sensor networks, almost all of them used simulated signals. How the real signals perform? This is quite significant for the research. For each radar sensor, we use P220 UWB radar, as shown in Figure 2.14, in monostatic mode where waveform pulses are transmitted from a single Omni-directional antenna and the scattered waveforms are received by a collocated Omnidirectional antenna. The two antenna ports on the P220 are used for the transmit and receive antennas.

The scan parameters tab contains the settings for the waveform scan. The default scan parameters for P220 are detailed in Table 2.4, where EIRP represents

Effective Isotropic Radiated Power. For each radar sensor, each scan is shown as Figure 2.15.

Table 2.4. PulsOn220 Specifications

Parameter	UWB (P220)
Central Freq.	4.3 GHz
Bandwidth%	53.48
10dB Bandwidth	2.3 GHz
Pulse type	1st order Gaussian Monocycle
Pluse Duration	430 pS
Pluse Length	0.13 m
Resolution	6.5 cm
Power Consumption	5.7 Watts
Pulse Repetition Freq.	9.6 MHz
EIRP	-12.8 dBm

For sensor networks, a number of sensors measure signals which are correlated both spatially and temporally. Traditionally, researchers usually independently project each sensor's signal to some basis and then transmits just a few of the resulting coefficients to a collection point, like in paper[74], the authors use all the J sensors' signals as a large vector $x = [x_1\ x_2\ \ldots\ x_J]^T$ and compress it using conventional compressive sensing method. In that case, the authors did not consider the correlation property between the signals of sensor networks. Some research begin to dig this property in sensor networks, but as we will discuss in this section, their algorithm is not that efficient, that is, their algorithms are correspondingly complex with correlation property introduced, which is not acceptable in real-world application. However, in this section, we propose a new efficient algorithm which considers the correlation between sensors without introducing any complexity, which even reduces the whole complexity of algorithm.

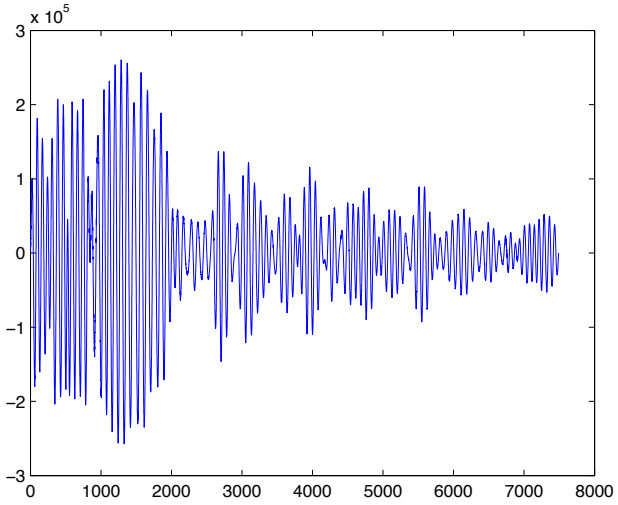

Figure 2.15. One Radar Sensor Signal.

[73] gives a detailed analysis of both inter-node and intra-node correlation for Temperature, Humidity, Solar, Luminosity, Wind, and Voltage sensors. Here, we borrow their definition to see the correlation property of Radar Sensors. The inter-node (spatial) and inter-node (temporal) correlation for Radar Sensor Network signals are shown in Figures 2.16 and 2.17. In Figure 2.16, the x-axis is the number of sensors separated, and the y-axis is the spatial correlation. While in Figure 2.17, the x-axis m is the time shifted for one sensor signal. It obvious that all radar sensor signals are highly correlated both spatially and temporally.

2.4.2.2 Proposed Compression Algorithm

Similarly as discussed in [75], where the authors introduced Distributed Compressive Sensing (DCS) for multi-signal ensembles which exploit both intra- and inter-signal correlation structures in sensor networks, we also use the Joint Sparsity

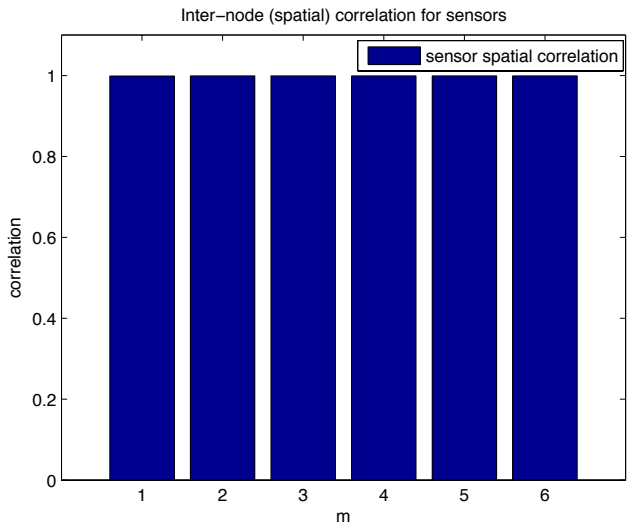

Figure 2.16. Inter-node correlation for radar sensor signals.

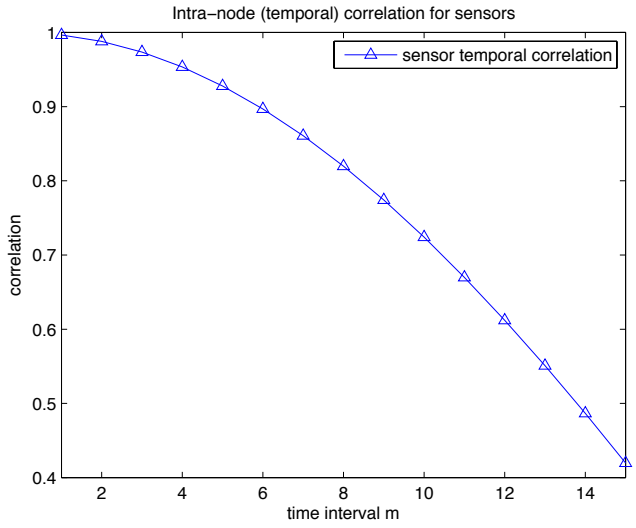

Figure 2.17. Intra-node correlation for radar sensor signals.

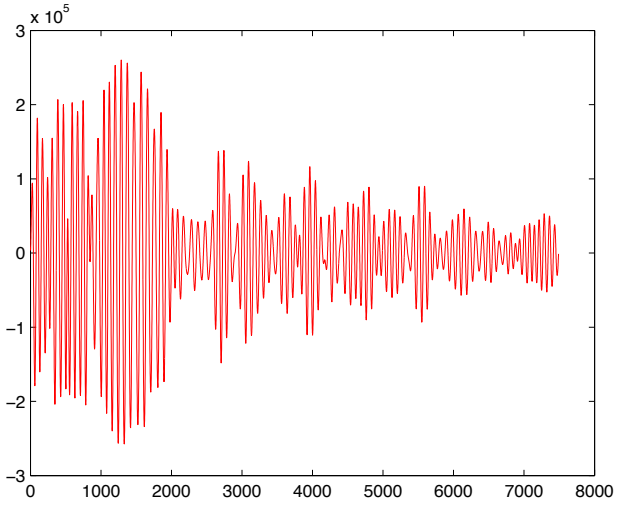

Figure 2.18. Common Component of sensor signals.

Model (JSM) in this section, where all signals share a common component while each individual signal contains a sparse innovation component, as shown below,

$$x_j = x_c + x_i \tag{2.130}$$

in which,

$$x_c = mean(x_1, x_2, \cdots, x_J) \tag{2.131}$$

where $j \in \{1, 2, \ldots, J\}$, J is the total number of sensors, x_c is the common component, and x_i is the innovation component for x_j. Here, we totally have 100 sensors, and the common component is shown as in Figure 2.18.

For each individual sensor's signal, subtract the common part, we could get the corresponding innovation component as illustrated in Figures 2.19 and 2.20. The interesting thing we could observe is that almost all elements of these innovation are around zero, therefore, these innovation components could be treated as sparse with a proper threshold set. Compressive sensing algorithm could be directly used to compress these innovation components, as they are sparse as we see.

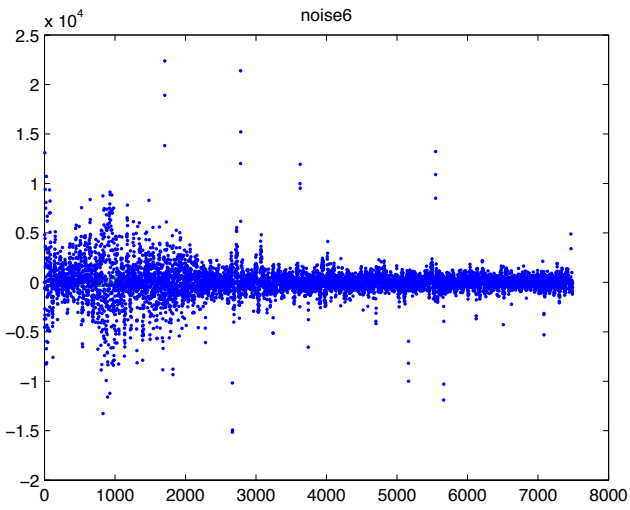

Figure 2.19. Innovation Component-1.

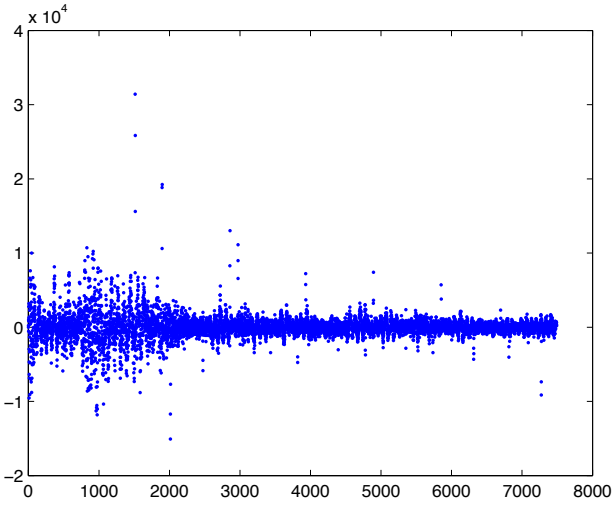

Figure 2.20. Innovation Component-2.

However, we could also observe the dimension of these innovation components are still too large, which could not reduce the compression complexity significantly if we directly use compressive sensing to compress them. A new and efficient algorithm is required to propose to compress these huge amount of data in sensor networks. We will discuss our proposed algorithm in next Section and then compare the performance of our algorithm with some existing methods, such as joint PCA and CS algorithm.

Based on what we observed that the innovation parts are sparse, and our discussion above, we propose a new compression algorithm for Radar Sensor Networks. The proposed compressive data gathering is able to reduce global scale communication cost without introducing intensive computation or complicated transmission control. The algorithm is shown as,

1). Find the common component x_c for J sensors' signals;

$$x_j = x_c + x_i, j \in 1, 2, \cdots, J \tag{2.132}$$

in which,

$$x_c = mean(x_1, x_2, \cdots, x_J) \tag{2.133}$$

2). For each sensor's data, subtract the common component, get each individual sensor's innovation component, i.e., x_i;

3). Observe that each sensor's innovation component x_i is sparse. One method is directly compress each innovation component using CS, however, as we will give in numerical results section, a large dimensional vector is still required to represent each innovation component. Therefore, we try to compress the whole RSNs signals further;

4). Record all the non-zero values v_i and their positions for each sensor's innovation part, get a new vector v_i. This is for the reason that each innovation component is

sparse, only a few non-zero values exists, recording the few positions for non-zero values does not introduce more computing complexity;

5). Find a basis to map the new vector v_i to a sparse vector $v_{s,i}$, here we use Discrete Wavelet Transform (DWT) or Discrete Cosine Transform (DCT);

6). Compress $v_{s,i}$ using compressive sensing to get a much smaller amount of data for communication or storage;

7). To recover the RSNs signal, based on all stored innovation components' position and the compressed values, recover and get each innovation component $\hat{x}_i$;

8). To get each radar sensor's signal, add the corresponding innovation component $\hat{x}_i$ and the common component $\hat{x}_c$ get $\hat{x}_j = \hat{x}_c + \hat{x}_i$;

We will show in the numerical results section that our algorithm is efficient and effective. Although it requires a vector to store the positions of each innovation component's non-zero values, it just is stored there, not involved in any calculation, therefore, it does not introduce any computation complexity. What's more, we observe that the signal dimension for each sensor node is 7489, but the non-zero ones for each innovation component is less than 1000, the vector for position storage is acceptable. Our algorithm fully utilize this property of each innovation component to further compress it.

2.4.2.3 Simulation Results

This section presents some numerical results for our proposed efficient compression algorithm. Due to the correlation property of sensor signals, we could use joint sparsity model to get the common component x_c and each sensor's innovation component $x_{i,j}, j \in \{1, 2, \ldots, J\}$ for the RSNs. The common component has already shown in early part of this section. The compression of the common component x_c

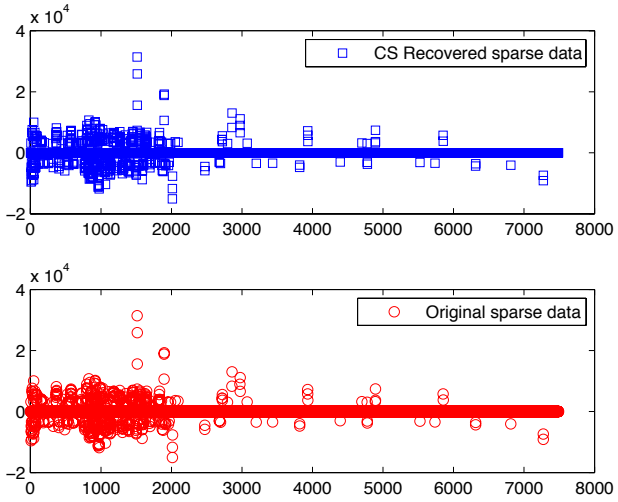

Figure 2.21. Innovation component-Original & Recovered.

is as Figure 2.18, here, we use wavelet transform to convert the common component x_c to be sparse, and use random Gaussian matrix as measurement matrix.

For each sensor's innovation component, as we discussed in the sparsity analysis section, it is sparse. Absolutely, we could directly use CS to compress the sensor signal, as shown in Figure 2.21. However, although these innovation components are sparse, their dimension are still the same as each sensor's signal. The compression of these innovation could not reduce significantly, as shown in Figure 2.22, where for a good recovery, we still get $M = 2000$ compressed signals after CS.

How to reduce further? As these innovation components are sparse, we could consider record all non-zero values and their positions. In this case, as we will discuss, these non-zero values, composing a new vector, v_i, could be found sparse in some basis, and be compressed further using CS. With their positions stored and each value recovered, each innovation component will be easily reconstructed back. This process significantly reduce the computation complexity introduced by CS, as

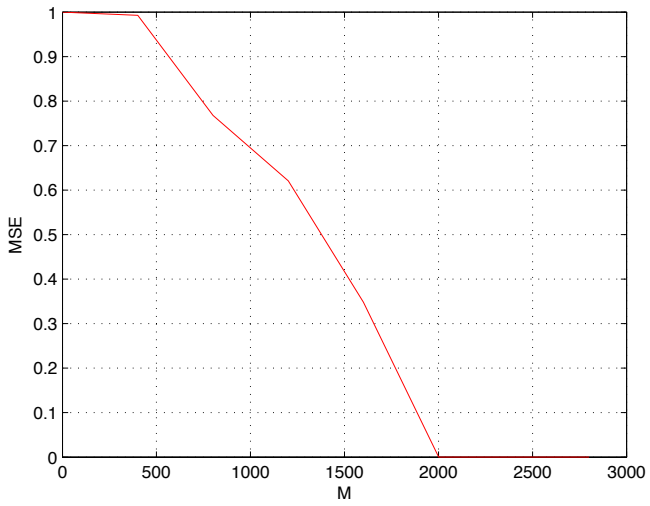

Figure 2.22. Directly compression of innovation component.

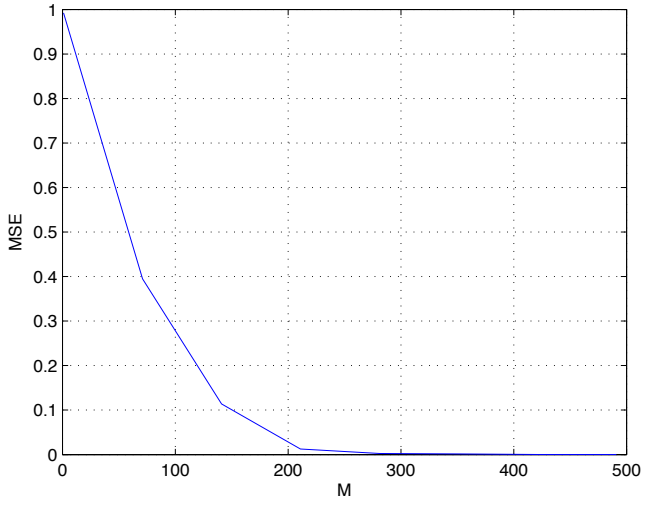

Figure 2.23. Compression of non-zero Innovation Component.

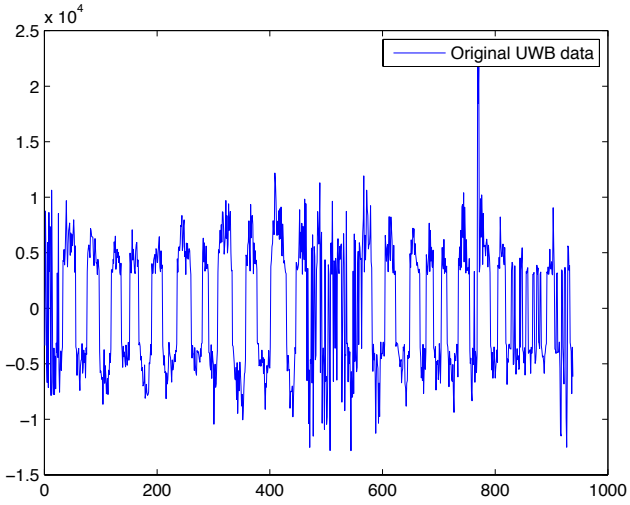

Figure 2.24. Original Non-zero Innovation Component v_i.

the compression of each innovation component only requires quite small number of measurement "M". For example, Figure 2.24 shows all the 941 non-zero values of the innovation component in Figure 2.20, which form a new vector, v_i. We use an independent matrix (with dimension 941 $\times$ 1) to record all the positions for the non-zero values. We notice the new vector v_i is sparse in some basis, such as Discrete Cosine Transform (DCT) or Wavelet Transform, therefore, it could be further compressed and well recovered using compressive sensing. As shown in Figure 2.23, for the same innovation component above, this algorithm only requires $M = 300$ for a good recovery. Although it still need a vector to store all the 941 non-zero values of the innovation component, this vector doesn't involve in any calculation, without introducing any computation cost. The recovered non-zero vector $\hat{v}_i$ is shown in Figure 2.25.

Although there is a threshold set to treat most elements in innovation components as zero, as the threshold is much small compared with the common component

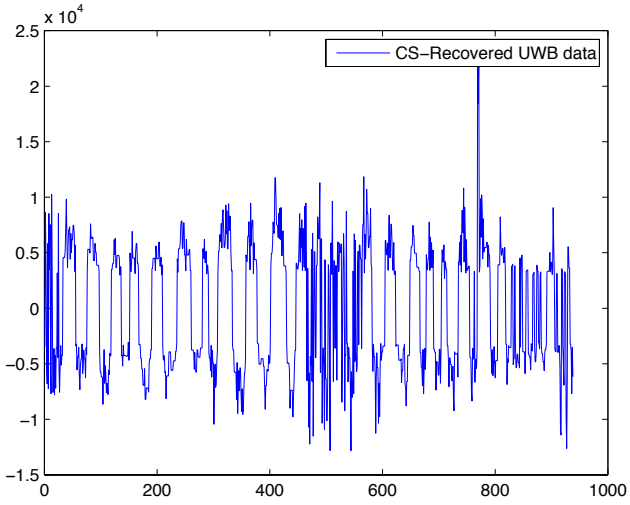

Figure 2.25. Recovered Non-zero Innovation Component $\hat{v}_i$.

of the sensors' signals, the final error for each sensor is still at an acceptable level. The Figure 2.26 below give one sensor's original signal and its corresponding recovered signal using our algorithm, i.e., common component and the innovation component.

Figure 2.27 shows the performance with different number of sensors. The x-axis is the required M-dimension per sensor after compression using our proposed algorithm, and y-axis is the Mean Squared Error (MSE). It is obvious that with more sensors involved, this algorithm contributes less communication cost for each sensor's signal without introducing any computation complexity, which is much more efficient compared with traditional CS or some existing work. We give the detailed comparison of this proposed algorithm with some existing work in next section.

2.4.2.4 Comparison with Existing Work

In [74], the authors let J sensor readings form a new vector $x = [x_1\ x_2\ \ldots\ x_J]^T$, x is a K-sparse signal in a particular domain Ψ, and then compress it using conven-

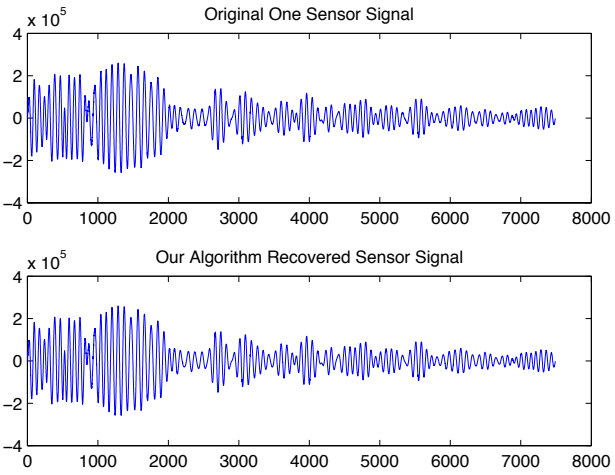

Figure 2.26. Original & Recovered Signal using our algorithm.

tional compressive sensing method. This algorithm did not fully use the correlation property of sensor signals. No matter how many sensors in the sensor networks, the dimension 'M' required for compression of each sensor's signal is almost the same, as this algorithm treats all sensors' signals independent. The performance of this algorithm is indicated as 'CS' in Figure 2.29. We notice no matter how many sensors involved, the dimension 'M' required is almost 2000 per sensor using this method.

Joint CS & PCA [73] is an algorithm that maps each sensor signal x_j into a sparse vector s_j using Principal Component Analysis (PCA), which compresses each sensor's signal into a smaller number of vector. This algorithm could reduce the communication cost, while on the other hand, introduces intensive computation for the pre-processing. We show the comparison of Joint CS & PCA algorithm and traditional CS in Figure 2.29. Here, the joint CS & PCA algorithm first use the following PCA principal to convert the $j-$th sensor's signal $x_j, j \in \{1, 2, \ldots, J\}$ into a sparse vector s_j,

$$x_j - \bar{x}_j = U_j s_j \tag{2.134}$$

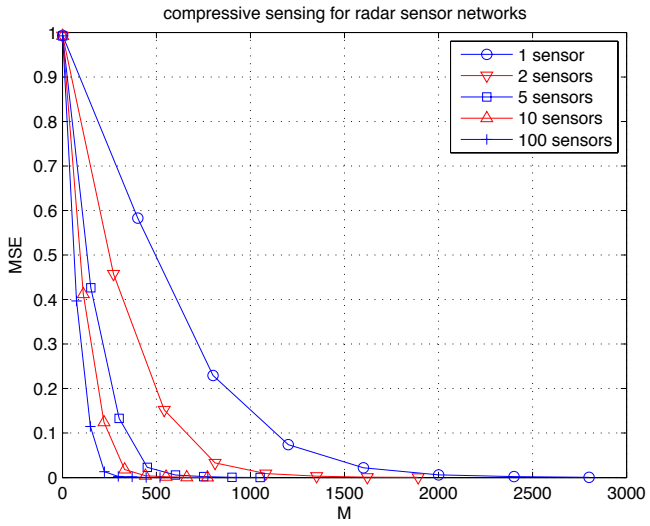

Figure 2.27. Efficient Sampling using CS for RSNs.

where $\bar{x}_j$ is the sample mean vector and U_j is the orthonormal matrix whose columns are the unitary eigenvectors of the sample covariance matrix of x_j according to the decreasing order of the corresponding eigenvalues.

Here the joint CS & PCA algorithm directly use the obtained sparse vector s_j to do the compressive sensing. We could observe that their performance are similar from Figure 2.29 for one sensor signal. With the number of sensors increases, the joint CS & PCA algorithm uses almost the same number of data to represent each sensor's signal, without significantly reduce the communication cost. Furthermore, from Table 2.5, we could see that the pre-processing of joint CS & PCA algorithm is time-consuming, that is, it requires much more time to get the sparse vector s_j using PCA before CS, as there are a lot of matrix computation in PCA. Although its performance is acceptable, but the time inefficient is not tolerant in real application. Comparison of the method in [74] and joint CS & PCA algorithm is shown in Figure 2.29. The x-axis M is the number of measurement per sensor required. As these two

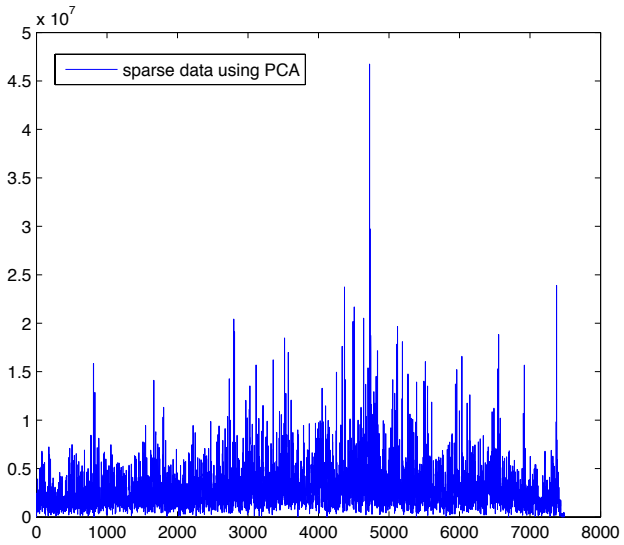

Figure 2.28. Sparse vector obtained from joint PCA & CS.

algorithms process each sensor signal independently, with more sensors involved, the number of measurement per sensor almost remains the same. Here, the elapsed CPU time (Windows 7, 64-bit Operating System with Intel(R) Core(TM) i7-2600s CPU @ 2.80GHz and 8.00 GB RAM) is obtained using tic and toc in MATLAB2010. x_j is the $j-$th radar sensor signal, as described before, $x_j = x_c + x_i$, x_c is the common component for all sensors, and x_i is each radar sensor's innovation component. v_i is a new vector formed by all non-zero values for each sensor's innovation component, x_i. s_j is the sparse vector related to each radar sensor signal x_j, obtained from equation (2.134). DWT represents discrete wavelet transform, in this section, we use 6-level "Daubechies" wavelet decomposition of the real radar signal. DCT is discrete cosine transform. It means that the signal is sparse, if we mention "directly CS", as it's not necessary to find a sparse basis, directly mapping the sparse signal to the measurement matrix is enough. We always use random Gaussian matrix as measurement matrix in this section. For our comparison, in this table, except v_i,

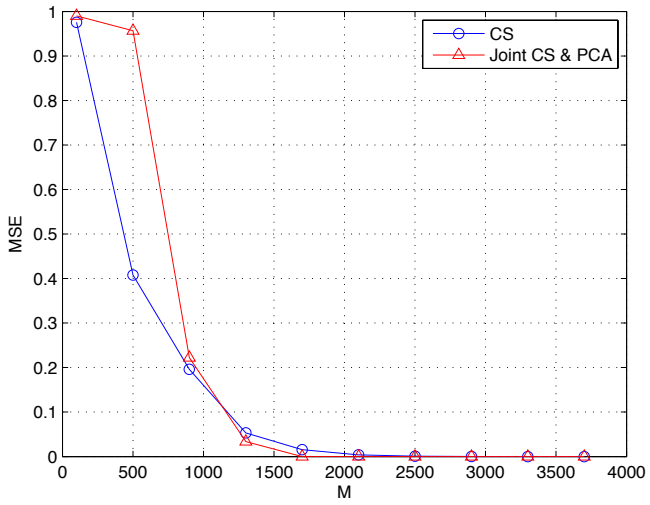

Figure 2.29. Comparison of traditional CS and joint PCA & CS.

all the others use $M = 2000$ measurements for good recovery as shown in Figures 2.22, 2.27, and 2.29, with the signal dimension $N = 7489$. For v_i, its dimension is 941, and we use 300 measurements as shown in Figure 2.23 to compress and recover it. The last row in Table 2.5 shows the time consumption of representing each x_j to a sparse vector s_j using equation (2.134). With matrix U_j introduced, joint CS & PCA algorithm requires too much time for pre-processing before compression. For our algorithm, after we get v_i, the compression is easier and quickly. It is obvious that our algorithm consumes the least time compared with the compression in [74] or joint CS&PCA algorithm. Comparing Figures 2.29 and 2.27, we could conclude that our proposed algorithm is more efficient for RSNs, especially with the number of radar sensors increases. Combined with our performance result, we could conclude that our proposed algorithm is not only time efficient but also performance guaranteed.

Joint recovery [75] for distributed compressive sensing is a good and mature method for sensor networks. However, recover each signal based on both common

Table 2.5. Elapsed CPU time of different compression algorithms for different data in RSNs.

Signal	Compression	CPU time (seconds)
x_j	DWT+CS	92.021605
x_c	DWT+CS	91.913748
x_i	directly CS	62.927351
x_i	DWT+CS	37.560715
v_i	DCT+CS	13.188225
s_j	Directly CS	77.983813
s_j	N/A	1150.283037

component and the corresponding innovation component using the optimization, as shown in equations (2.135) and (2.136), which introduces more computation complexity in the recovery process. They recover each signal based on both

$$\hat{x} = \ arg\ min\|x\|_{l1},\ s.t.\ y = \Phi\Psi x \tag{2.135}$$

$$\|x\|_{l1} = \|x_c\|_{l1} + \|x_i\|_{l1} \tag{2.136}$$

While our algorithm reconstructs common component and innovation component separately. That means that using our algorithm, no new recovery algorithm has to be found, regular recovery of CS is enough. This is a big advantage of our algorithm, especially for the hardware design.

2.4.3 Underwater Acoustic Sensor Networks (UWASNs)

2.4.3.1 Introduction

Underwater Acoustic Sensor Network (UWASN) [84] [85] [86] [87] [88] [89] [90] [91] is a distributed underwater system with networked wireless sensors with sensing, processing, and communication capabilities that are deployed underwater to support a broad range of applications. There has been a growing interest in monitoring

underwater mediums for scientific, environmental, commercial, safety, homeland security and military applications nowadays. UWASN consists of a variable number of sensors or vehicles, which self-organize in an autonomous network which can adapt to the characteristics of the ocean environment, that are deployed underwater to perform collaborative monitoring tasks. Because of the characteristic (large delay, long distance of communication) of network, UWASN mainly differs in the communication media employed for information transmission. And acoustic signal is used in underwater due to less attenuation and further distances in water compared to radio signals. However it has several challenges [84] [88] [89] [92] [93] [94] [95][96] [97] [98] [99] [100] [101] [102] [103] [104], such as limited bandwidth and high propagation delays. Underwater acoustic communications [101][102] are constrained by limited and distance-dependent bandwidth, time-varying multi-path propagation and low speed of sound. Todays off-the-shelf acoustic modems typically have the bandwidth between $5-20Kb/s$. Bandwidth limitations make reducing data counts in UWASNs even more important than for radio networks [89]. This stimulates us to find a solution to reduce information stored or transmitted by UWASNs.

Besides, underwater acoustic sensor nodes have limited storage capacity. From [87][88], we know that if data collecting rate exceeds the storage capacity of the sensor node during the expected battery lifetime, then the data must be transmitted to another location or the node must be retrieved prior to full battery expenditure, which indicates that the amount of data that can be collected over a period of time is dependent on the maximum data storage capacity of a sensor node and the battery lifetime. In [85], the authors assumed that sensor nodes could buffer data with temporary data storage, which did not solve the data storage problem from the root. With applications like seismic imaging, UWASN nodes will collect and send large amount of data that can easily overwhelm the network capacity. Due

to limited bandwidth, poor-quality communication channels, and limited storage capacity for underwater acoustic sensor nodes in UWASNs, appropriate techniques to compress the UWASNs data to enable storage and transmission requirements should be developed. The emerging method compressive sensing (CS) [16] [17] [19] has high compression ratio, which could compress the real UWASNs data significantly with less computational complexity. And CS theory can be very promising, since sparsity is an attribute present in natural as well as man-made systems [19]. Therefore, in this section, noisy compressive sensing with application to UWASNs is studied theoretically and in real-world data.

Some pioneering works have discussed the application of CS in wireless sensor networks (WSNs). Luo et al, presented the first complete design to apply CS theory to sensor data gathering for large-scale WSNs in [105]. Some applications of compressive sensing over networks was demonstrated in [106]. An efficient and effective signal compression algorithm based on CS principles for applications in real-world radar sensor networks was given in [136]. Distributed CS which exploits both intra- and inter- signal correlation structures for jointly sparse signals was introduced in [107]. And a novel CS based approach for sparse target counting and positioning in WSNs was proposed in [108]. Although compressive sensing (CS) has been proposed and studied by many researchers, almost no one has applied it to real-world UWASNs. Besides that, most previous works focused on the gathering, compression and reconstruction of sensor data, while our work focuses more on the theoretical rate distortion analysis of the application of CS in UWASNs, i.e., how much distortion between the original sensor data and its representation using CS.

Since Gaussian assumption is a classical modeling assumption heavily used in areas such as signal processing and communication system [41], in this section, we analyze the performance of noisy compressive sensing with application to UWASNs.

Due to limited bandwidth, poor-quality communication channels, and limited storage capacity for underwater acoustic sensor nodes in UWASNs, appropriate techniques to compress the UWASNs data to enable storage and transmission requirements should be developed, which motivates us to apply CS to UWASNs, and study the theoretical and real-world performance of noisy CS with application to UWASNs. We have to mention that although the theoretical performance results are derived with the assumption that the source and noise follow Gaussian distribution, these results still could be used as theoretical reference of real application of noisy compressive sensing to UWASNs. The reason is that the Gaussian distribution maximizes the entropy over all distributions with the same variance[36]. The entropy is a measure of the average uncertainty in the random variable, which indicates that our results provide the worst-case information-theoretic performance. The performance of noisy compressive sensing is studied theoretically in the presence of random noise. Since the reconstruction distortion is dependent on the distortion in the measurement, we first study the measurement rate distortion performance of compressive sensing with random noise. We discuss the measurement rate distortion performance with uniform quantization and non-uniform quantization of the measurement observation respectively. Then the reconstruction rate distortion function of noisy compressive sensing is detailed. The corresponding reconstruction rate distortion performance with uniform and non-uniform quantization are given as well. The rate distortion performance in underwater acoustic sensor networks is studied, with or without correlation among sensor readings. These results provide some theoretical reference of the real application of compressive sensing in the presence of noise. The compression of real-world underwater acoustic sensor network data is applied to verify the theoretical results derived in this section.

2.4.3.2 Rate-Distortion Performance of Noisy Compressive Sensing in Underwater Acoustic Sensor Networks

For sensor networks with S sensors, $X_1, X_2, \cdots, X_S$, the rate distortion function can be represented as[115][116],

$$\begin{aligned} & R_{X_1,X_2,\cdots,X_S}(D_1, D_2, \cdots, D_S) \\ & = \min_{E(D(X_1,\hat{X_1}))\leq D_1,\cdots,E(D(X_S,\hat{X_S}))\leq D_S} \\ & I(X_1, X_2, \cdots, X_S; \hat{X}_1, \hat{X}_2, \cdots, \hat{X}_S) \end{aligned} \tag{2.137}$$

The mutual information between the sensed information before and after CS[115][116] is

$$\begin{aligned} & I(X_1, X_2, \cdots, X_S; \hat{X}_1, \hat{X}_2, \cdots, \hat{X}_S) \\ & = H(X_1, X_2, \cdots, X_S) - H(X_1, X_2, \cdots, X_S|\hat{X}_2, \cdots, \hat{X}_S) \\ & = I(X_1; \hat{X}_1) + I(X_2; \hat{X}_2) + \cdots + I(X_S; \hat{X}_S) \\ & - I(X_1; X_2) - I(X_1; X_3) - \cdots - I(X_1; X_S) \\ & - I(X_2; X_3) - I(X_2; X_4) - \cdots - I(X_2; X_S) - \cdots - I(X_{S-1}; X_S) \\ & = \sum_{i=1}^{S} I(X_i; \hat{X}_i) - \sum_{i=1}^{S} \sum_{j=2, j<i}^{S} I(X_i; X_j) \end{aligned} \tag{2.138}$$

Underwater sensor networks have attracted many research efforts from academy and industry. The differences between underwater sensor networks and terrestrial networks were addressed in [88][96], from which we know that in terrestrial sensor networks, nodes are densely deployed, and thus there is spatial correlation between sensor nodes because nodes are spatially close to each other. Usually, underwater sensor nodes are deployed sparsely[88], because they may be more expensive and complex. Therefore, correlation among sensor readings from different sensors may

not be significant in UWASNs. In this case, without correlation among sensors, all sensor readings $X_1, X_2, \cdots, X_S$ are independent, therefore

$$I(X_i; X_j) = 0, \quad i, j = 1, \cdots, S, \quad i \neq j \tag{2.139}$$

So that the rate distortion function can be represented as follows, if all sensors are independent[115][116],

$$R_{X_1,X_2,\cdots,X_S}(D_1, D_2, \cdots, D_S) = \sum_{i=1}^{S} R_{X_i}(D_i) \tag{2.140}$$

where the rate distortion function for each sensor signal could be calculated based on what we described in Section 4.

However, in our study, the UWASNs signals are obtained as shown in [117], where 32 sensors were evenly distributed within the exploration cable, and each node is responsible for a signal acquisition of 16 hydrophones, with hydrophone space 2 m. Sensors are spatially close to each other, from which we could conclude that UWASN signals are correlated in our case. Real experimental acquisition of the UWASN waveform signals are shown in Fig. 2.30. A point-like sound source is placed in the not far distance from the array as the sound excitation source in the experiment.

It would need less number of bits to represent the sensed information if sensors are correlated than that if they are independent[115][116], in which case

$$I(X_i; X_j) > 0, \quad i, j = 1, \cdots, S, \quad i \neq j \tag{2.141}$$

Therefore, the corresponding rate distortion function is represented as

$$\begin{aligned} &R_{X_1,X_2,\cdots,X_S}(D_1, D_2, \cdots, D_S) \\ &\quad \geq \sum_{i=1}^{S} R_{X_i}(D_i) - \sum_{i=1}^{S} \sum_{j=2, j<i}^{S} I(X_i; X_j) \end{aligned} \tag{2.142}$$

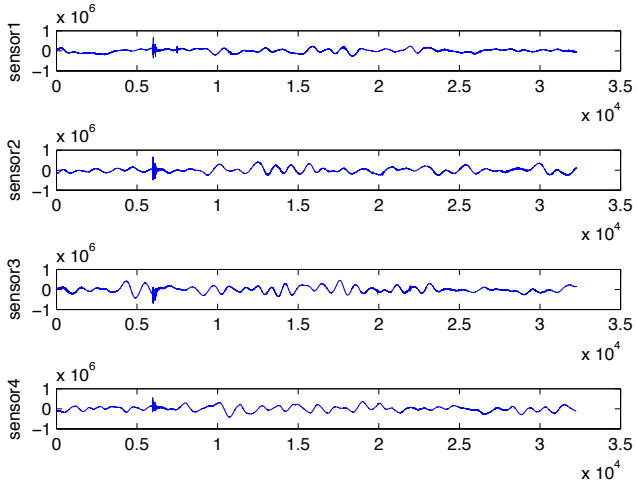

Figure 2.30. Underwater Acoustic Sensors data.

2.4.3.3 Simulation and Experimental Results

The measurement rate distortion performance is shown in Figures 2.31 and 2.32 with different length N of signal and different dimension of measurement matrix M. For compressive sensing, if we increase the dimension N of the original signal, with M fixed, the dimension of the measurement matrix $\Phi_{M\times N}$ also increases. This makes the whole compressive sensing more complex and time-consuming, i.e., more bit rate $R(D)$ is needed to process the signal. Therefore, we will only use the bit per dimension rate $R(D)/N$ in the latter part. On the other hand, if we fix the length of N, but increase the sampling rate M/N, as shown in Fig. 2.32, the bit rate required to transmit and process in compressive sensing will also increases.

The reconstruction rate distortion performance of noisy compressive sensing is shown in Fig. 2.34 with different length of original signal $\underline{x}$, the sampling rate $M/N = 1/4$, and $K = 100$, and Fig. 2.33 with different dimension of measurement matrix M . The original signal $\underline{x}$ follows Gaussian distribution with $\sigma_x^2 = 1$. As random

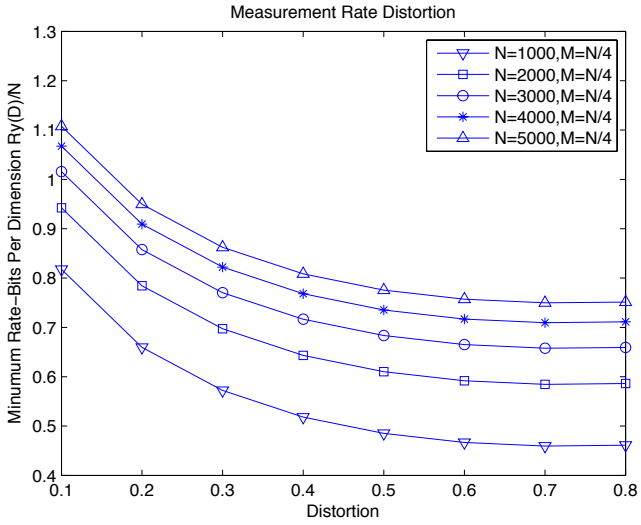

Figure 2.31. Measurement Rate Distortion Performance.

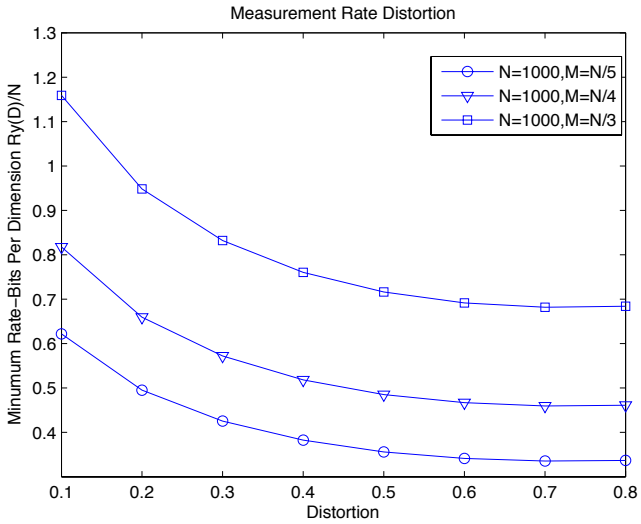

Figure 2.32. Measurement Rate Distortion Performance.

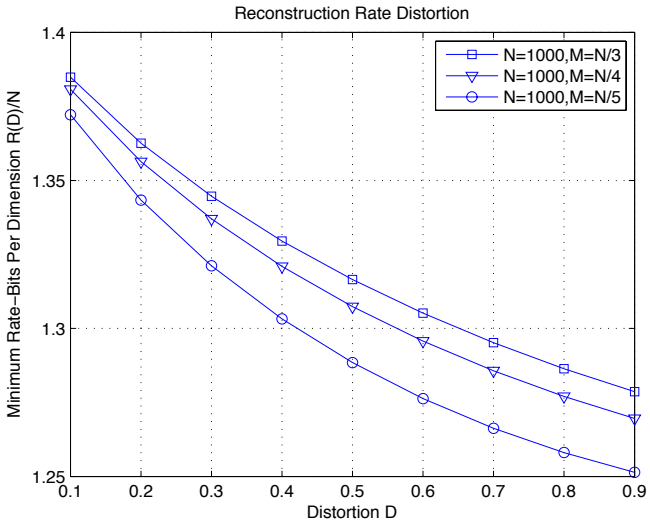

Figure 2.33. Reconstruction Rate Distortion Performance.

waveforms with independent identically distributed (i.i.d.) entries, e.g., Gaussian or ± 1 binary entries, exhibit a very low coherence with any fixed representation Ψ. Therefore, we choose Gaussian Random Matrix as the measurement matrix Φ. From [34], we know that $|\epsilon|^2 \leq 8K/N^2$. We assume the noise bound $\epsilon = 0.01$.

Similarly as in measurement rate distortion performance, if we increase the dimension N of the original signal, with M fixed, the dimension of the measurement matrix $\Phi_{M\times N}$ also increases. This makes the whole compressive sensing more complex and time-consuming, i.e., more bit rate $R(D)$ is needed to process the signal. The rate $R(D)$ in Figures 2.33 and 2.34 use bits per dimension, i.e., $R(D)$ we obtained from the expression divide by N, $R(D)/N$. In this simulation, we set $M = N/4$ and have $N = 1000$, $N = 2000$ until $N = 5000$ several different values. It is obvious that with the increase of distortion, the bit rate decreases and with the increase of N, the rate increases. We can also notice that, although N increases from 1000 to 5000, the bit per dimension $R(D)/N$ does not increase severely. From this

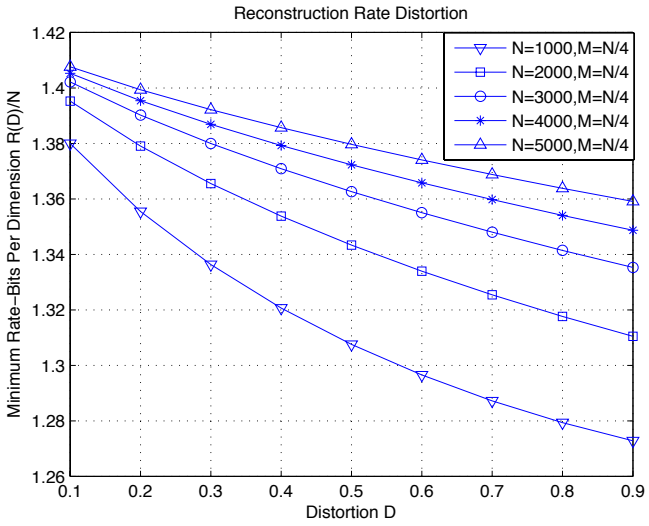

Figure 2.34. Reconstruction Rate Distortion Performance.

observation, we get the conclusion of the advantage of compressive sensing that it uses less number of bits to represent the information.

The reconstruction rate distortion performance of compressive sensing with quantization is shown in Fig. 2.35. For example, with quantization error $\epsilon = 0.293\sqrt{M}$ for eight-level uniform quantization, and $\epsilon = 0.202\sqrt{M}$ for eight-level non-uniform quantization. It is obvious that for non-uniform quantization, the lower bound of rate $R(D)$ is much lower than that of uniform quantization, which indicates that non-uniform quantization is much more efficient.

In our analysis, we assume the source signal follows Gaussian distribution, and provide the corresponding results. However, in real world, not all the source information follows Gaussian distribution. For example, for underwater acoustic sensor networks[117], the data gathered looks like in Fig. 2.30, with the probability density function (PDF) for each sensor signal shown in Fig. 2.36.

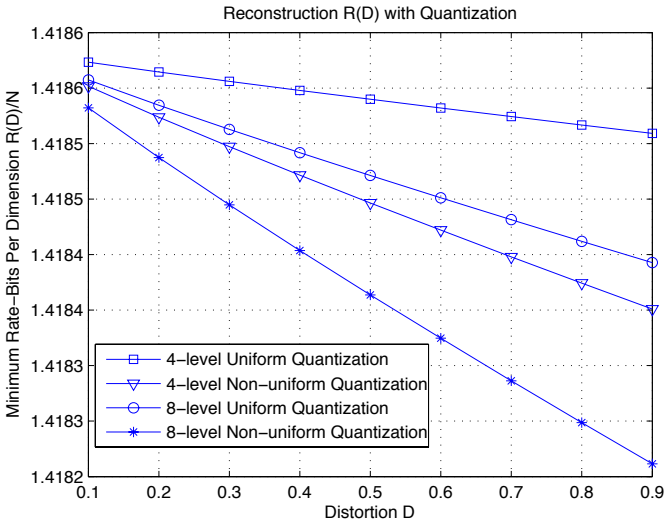

Figure 2.35. Reconstruction Rate Distortion with Quantization.

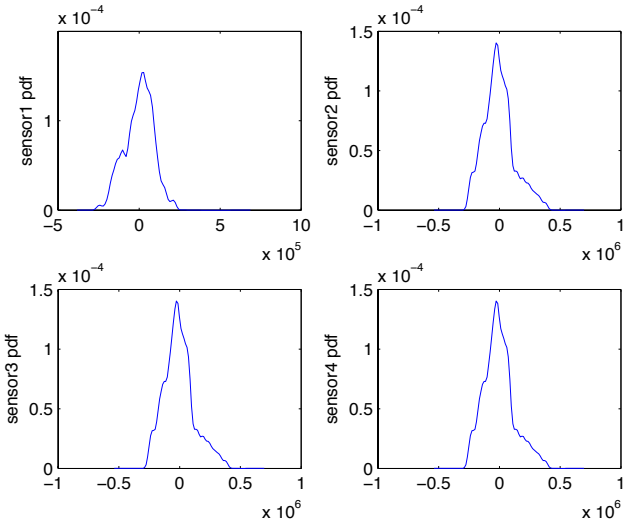

Figure 2.36. Underwater Acoustic Sensors data-PDF.

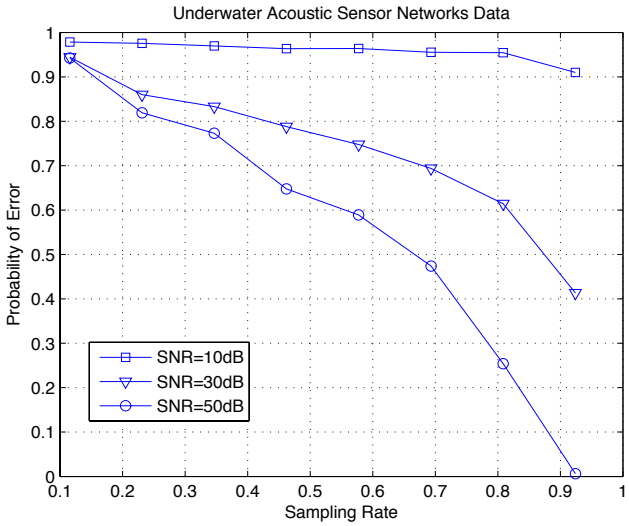

Figure 2.37. Underwater Acoustic data- Probability of Error.

From information theory, we know that the Gaussian distribution maximizes the entropy over all distributions with the same variance[36]. As we know, The entropy is a measure of the average uncertainty in the random variable. This indicates that our results provide the bound of the worst case. For real application of compressive sensing, even though the source does not follow Gaussian distribution, our results could be used as an reference to judge the performance of compressive sensing.

Figure 2.37 shows the compression performance of real-world underwater acoustic data with different SNR using compressive sensing. We observe that with the increase of sampling rate M/N, the probability of error decreases. This property indicates that in order to get better reconstruction performance with lower probability of error, the sampling rate should be chosen as high as possible. However, higher sampling rate means higher computation complexity. To balance this, in the

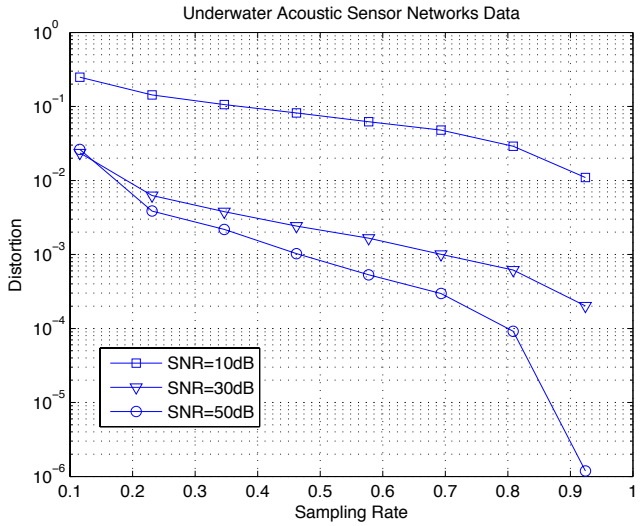

Figure 2.38. Underwater Acoustic data-Sampling Rate v.s. Distortion.

real application of compressive sensing, Fig. 2.37 could be referred for the chosen of certain sampling rate at an acceptable probability of error.

Fig. 2.38 is the relationship between sampling rate M/N and distortion D. From which, we could see that with the increase of sampling rate M/N, the distortion D decreases, and with the increase of signal-to-noise ratio (SNR), the distortion also decreases significantly.

In Fig. 2.39, the rate distortion performance of real-world underwater acoustic data is given with different quantization levels. We have to mention here that the performance in Fig. 2.39 is under the real-world UWASNs data, with quantization of the compressed observation. The curves look not smooth, which is because the quantization process of real-world UWASN data makes some useful information lost. As discussed in the theoretical part, the quantization process is done to the compressed observations y. First, Lloyd algorithm is used to optimize the scalar quantization parameters, and then non-uniform quantization is performed. From Fig. 2.39, it

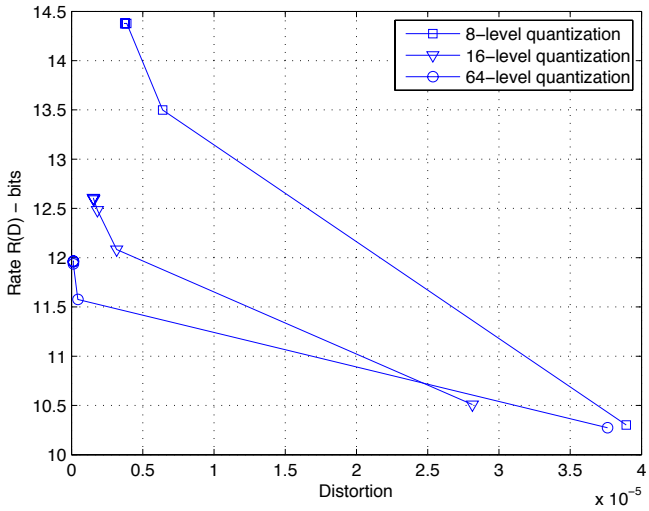

Figure 2.39. UWASNs Rate Distortion with Quantization.

is obvious that with the increase of quantization level from 8 to 64, the minimum distortion could achieve very small, and the minimum bit rate $R(D)$ also becomes a little smaller, which verifies the theoretical results proved in Section IV.

To see the real rate distortion performance in real-world underwater acoustic sensor networks, we select 6 sensors' signal in the same time period, with three of them plotted in Fig. 2.30. As all these sensors monitor the same shallow water environment at different locations, the signals they collected have high correlations, which stimulates us to take advantage of this property in the rate distortion analysis. There is a common part X_{co}, which is the average of all sensor signals, and there is also a innovation part X_{in} for each sensor, which is the signal without the common part. In the compression process, we only need to compress the common part once for all sensors. Since each innovation part is sparse, which is easier to be compressed. The corresponding rate distortion performance could be obtained in this process of compressive sensing.

As we described above, all sensors monitor the same shallow water environment at different locations, the signals they collected have high correlations, which stimulates us to take advantage of this property in the rate distortion analysis. We noticed that there are inter-node (spatial) and inter-node (temporal) correlation for each reading of UWASNs. Therefore, in UWASNs, more sensor readings involved does not mean more data rate $R(D)$ required. Fig. 2.40 gives the change of data rate (bits) with the increase of sampling rate in UWASNs. We can conclude that with the increase of sampling rate M/N (with N fixed, N is the dimension of underwater acoustic signal), data rate $R(D)$ also increases. Take the correlation of UWASNs into account, with more number of sensors, data rate $R(D)$ does not increase significantly, as shown in Fig. 2.40, the data rate per sensor becomes lower than one sensor. There are only 6 sensors calculated in Fig. 2.40. If the number of sensors of UWASNs increases to thousands or above, the advantage will be much more clear that the required data rate does not increase as the whole sensor data increases, which is detailed in [136][135]. This also indicates that the computation complexity does not get higher with more sensor nodes involved, which is also because of the correlation between sensors in UWASNs. Here the real data rate for each sensor $R_{X_i}(D_i)$ is calculated based on the original UWASNs signal's probability mass function (PMF), and the difference between the original and recovered signal's PMF, i.e., $R_{X_i}(D_i) = \min \sum (H(X_i) - H(X_i - \hat{X}_i))$. The mutual information between different sensors is $I(X_i; X_j) = H(X_i) - H(X_i|X_j) = H(X_i) - H(X_i - X_j|X_j)$.

Fig. 2.41 gives the relationship between rate $R(D)$ and distortion. It is clear that with the increase of distortion, the minimum information rate processed decreases. With the number of sensors increasing, for the same distortion, less data rate will be required, which significantly improves the efficiency of processing.

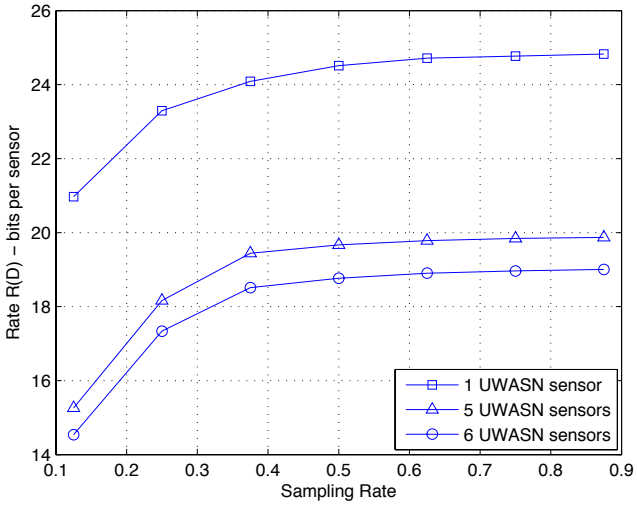

Figure 2.40. UWASNs - Sampling rate v.s. Rate.

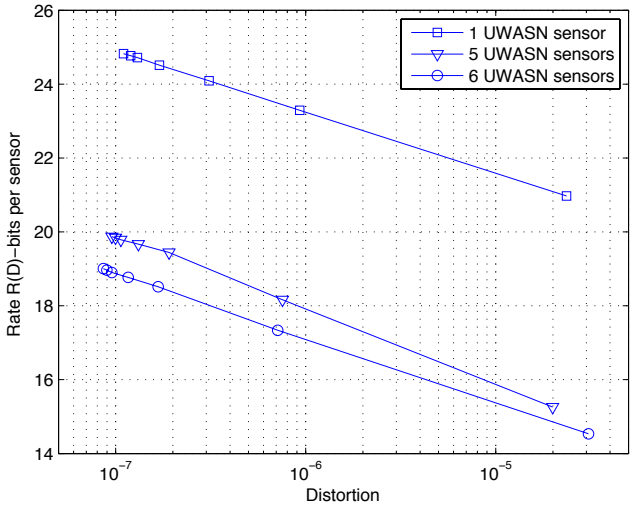

Figure 2.41. UWASNs - Rate v.s. Distortion.

These results based on real-world UWASNs data prove the validity of our theoretical results derived. Although the performance results are derived with the assumption that the source and noise follow Gaussian distribution, these results still could be used as theoretical reference of real application of compressive sensing, as we explained in earlier part that the Gaussian distribution, as the worst case, maximizes the entropy over all distributions with the same variance. The results provided in this section could be used as theoretical references when apply compressive sensing to real data compression and reconstruction.

2.5 Conclusions

In this chapter, error performance bounds of noisy compressive sensing are derived based on information theory and estimation theory first. Information rate distortion function is a measure as the number of bits per symbol to be stored or transmitted under the constraint of a distortion. Rate distortion performance for scalar quantization of measurement observation is derived. Based on this, reconstruction rate distortion is also studied for CS.

In addition, we study the real-world applications of CS in Big Data, to Synthetic Aperture Radar (SAR), radar sensor networks (RSNs), and underwater acoustic sensor networks (UWASNs). All these prove that compressive sensing is an efficient algorithm for Big Data both in theory and applications.

Chapter 3

Coprime Sampling and Nested Sampling

3.1 Introduction

Traditional sampling methods are based on Nyquist rate sampling, which will have poor efficiency in terms of both sampling rate and computational complexity. Nowadays, more and more techniques are proposed to overcome the Nyquist sampling. Compressive sensing [17] provides us a new point of view, which could only use much less samples to perfectly recover the original signal at a high compression ratio. The authors give a new idea of co-prime sampling and nested sampling in [124], which use sparse sampling to estimate the autocorrelation for all lags.

Nested sampling is an non-uniform sampling, using two different samplers in each period. Although the signal is sampled sparsely and non-uniformly at $1 \leq l \leq N_1T$ and $(N_1+1)mT$,$1 \leq m \leq N_2$ for one period, the autocorrelation $R_c(\tau)$ of the signal $x_c(t)$ could be estimated at all lags. Hence, nested sampling can be used to estimate power spectrum even though the samples in the time domain can be arbitrarily sparse [124]. While coprime sampling uses two uniform samplers, with sample spacings PT and QT respectively, where P and Q are coprime integers. Similar as nested sampling case, the authors in [124] proved that the estimates of all lags of autocorrelation $R_c(kT)$ could be obtained from these two sets of samples of the signal $x_c(t)$.

In this chapter, properties of these two sparse sampling schemes are investigated. This chapter is organized as follows. The principle of nested sampling and coprime sampling are given in Section 3.2. Rate distortion function of both nested

sampling and coprime sampling is given in Section 3.3, since sparse sampling can cause possible distortion because less number of samples are used. Theoretical analysis of how these two sparse sampling methods affect the power spectral density is given in section 3.4. A secure transmission scheme for Big Data based on coprime sampling and nested sampling is provided in Section 3.5, as Big Data presents critical requirements for security in data collection and transmission of selected data through a communication network. Section 3.6 summarizes the results.

3.2 Coprime Sampling and Nested Sampling

3.2.1 Nested Sampling

The nested array was introduced in [123] as an effective approach to array processing with enhanced degrees of freedom [126]. The time domain autocorrelation could also be obtained from sparse sampling with nested sampling structure [125]. And the samples of the autocorrelation can be computed at any specified rate, although the samples from this nested sampling are sparsely and nonuniformly located.

In the simplest form, the nested array [125] has two levels of sampling density, with the level 1 samples at the N_1 locations and the level 2 samples at the N_2 locations.

$$1 \leq l \leq N_1, \text{for level 1}$$

$$(N_1 + 1)m, 1 \leq m \leq N_2, \text{for level 2}$$

Fig. 3.1 shows an example of periodic sparse sampling using nested sampling structure with $N_1 = 3$ and $N_2 = 5$. The cross-differences are given by

$$k = (N_1 + 1)m - l, 1 \leq m \leq N_2, 1 \leq l \leq N_1 \tag{3.1}$$

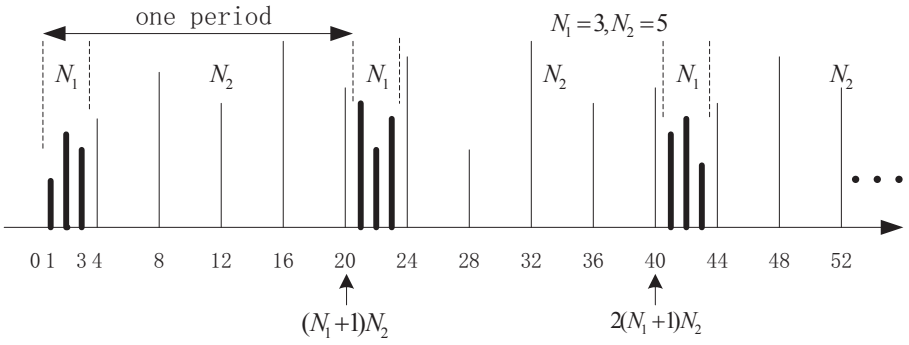

Figure 3.1. Nested sampling with $N_1 = 3, N_2 = 5$..

The cross-differences [125] are in the following range with the maximum value $(N_1 + 1)N_2 - 1$, except the integers and the corresponding negated versions shown in (3.3).

$$-[(N_1 + 1)N_2 - 1] \leq k \leq [(N_1 + 1)N_2 - 1] \tag{3.2}$$

$$(N_1 + 1), 2(N_1 + 1), \cdots, (N2 - 1)(N_1 + 1) \tag{3.3}$$

For example, consider the example in fig. 3.1, where $1 \leq m \leq 5$ and $1 \leq l \leq 3$, the cross differences $k = (N_1 + 1)m - l$ will achieve these values

$$1, 2, 3, (), 5, 6, 7, (), 9, 10, 11, (), 13, 14, 15, (), 17, 18, 19$$

with $4, 8, 12, 16$ missing.

Besides these integers, the difference 0 is also missing, for the reason that m and l are nonzero. While, we notice that the self differences among the second array could cover all of the missing differences, as shown

$$(N_1 + 1)(m_1 - m_2), 1 \leq m_1, m_2 \leq N_2 \tag{3.4}$$

The difference-co-array could be obtained from the cross-differences and the self-differences, which is a filled difference co-array as shown in (3.2). This means

that using nested array structure, with sparse samples, we could obtain the degrees of freedom as

$$2[(N_1+1)N_2-1]+1 = 2(N_1+1)N_2-1 \tag{3.5}$$

Using the above principle, we could get a sparse sampling using nested sampling structure as shown in fig. 3.1. We have two levels of nesting, with N_1 level-1 samples and N_2 level-2 samples in each period, with period $(N_1+1)N_2$. This shows that nested sampling is non-uniform and the samples obtained are very sparse.

Therefore, in $(N_1+1)N_2T$ seconds, there are totally N_1+N_2 samples. The average sampling rate is

$$f_{s,nested} = \frac{N_1+N_2}{(N_1+1)N_2T} \approx \frac{1}{N_1T} + \frac{1}{N_2T} < \frac{1}{T} \tag{3.6}$$

Here, $T = 1/f_n$, $f_n \geq 2f_{max}$ is the Nyquist sampling frequency, which is greater than twice of the maximum frequency. As the Nyquist sampling rate is $1/T$, the average sampling rate of nested sampling is smaller than the conventional Nyquist sampling rate.

If we set N_1 and N_2 larger, the average sampling rate f_s would be arbitrarily smaller. In the theoretical and numerical results sections, we will show that with N_1 and N_2 becoming larger, the bandwidth of the power spectrum density goes narrower, i.e., the spectrum gets more efficiently used.

3.2.2 Co-Prime Sampling

Different with nested sampling, co-prime sampling involves two sets of uniformly spaced samplers as shown in Figure 3.2.

The coprime sampling uniformly sample the Wide-Sense Stationary (WSS) process $x_c(t)$ using two sub-Nyquist samplers, with sample spacing PT and QT respectively, where P and Q are coprime integers with $P < Q$. $1/T$ Hz is the

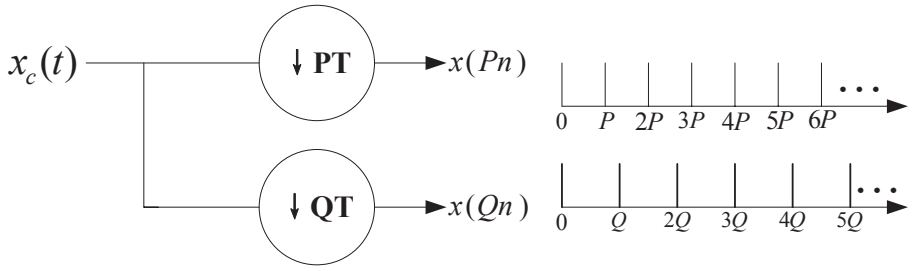

Figure 3.2. Co-Prime sampling in the time domain..

Nyquist rate for a bandlimited process, i.e., $1/T = 2f_{max}$, f_{max} being the highest frequency.

$$x(n) = x_c(nT) \tag{3.7}$$

Consider the product

$$x(Pn_1)x^*(Qn_2) \tag{3.8}$$

where $x(Pn_1)$ and $x(Qn_2)$ comes from the first and the second sampler. Set the difference as

$$k = Pn_1 - Qn_2 \tag{3.9}$$

The authors in [125] have shown that k can achieve any integer value in the range $0 \leq k \leq PQ - 1$, if n_1 and n_2 in the ranges $0 \leq n_1 \leq 2Q - 1$ and $0 \leq n_2 \leq P - 1$.

For coprime sampling, the two samplers collect $P+Q$ samples in PQT seconds, the average sampling rate is

$$f_{s,coprime} = \frac{P+Q}{PQT} = \frac{1}{PT} + \frac{1}{QT} < \frac{1}{T} \tag{3.10}$$

Same as in nested sampling, $T = 1/f_n$, $f_n \geq 2f_{max}$ is the Nyquist sampling frequency. We could notice the average sampling rate of coprime sampling is much smaller than the conventional Nyquist sampling rate of $1/T$.

Similar as stated in nested sampling, if we set P and Q larger, the average sampling rate would be arbitrarily smaller. We will show that with P and Q becoming larger, the bandwidth of the power spectrum density goes narrower.

3.2.3 PSD Estimation

In this part, we will detail the estimation of PSD using nested and co-prime sampling structure. In signal and systems analysis, the autocorrelation plays a very important role. The autocorrelation function of a random signal describes the general dependence of the values of the samples at one time on the values of the samples at another time.

The autocorrelation [128] of a real and stationary signal $x_c(t)$ is defined by this averaging

$$R_c(\tau) = E[x_c(t)x_c^*(t-\tau)] \tag{3.11}$$

$R_c(\tau)$ is always real-valued and an even function with a maximum value at $\tau = 0$.

For sampled signal, define $x(n) = x_c(nT)$, for some fixed spacing T. For the autocorrelation samples, $R(k) = R_c(kT)$, where $R_c(\cdot)$ as shown in (3.11). Therefore,

$$R(k) = E[x_c(nT)x_c^*((n-k)T)] = E[x(n)x^*(n-k)] \tag{3.12}$$

$R(k)$ can be computed from samples of $x_c(t)$ taken at an arbitrarily lower rate using nested or coprime sparse sampling.

And here we only list some important autocorrelation properties which will be used in this chapter:

(1) Maximum Value The magnitude of the autocorrelation function of a wide sense stationary random process at lag m is not greater than its value at lag $m = 0$, i.e.,

$$R(0) \geq \mid R(k) \mid, k \neq 0 \tag{3.13}$$

(2) The autocorrelation function of a periodic signal is also periodic.

(3) The autocorrelation function of WSS process is a conjugate symmetric function of k:

$$R(k) = R^*(-k) \tag{3.14}$$

The power spectral density (PSD) describes how the power of a signal or time series is distributed with frequency. The PSD is the Fourier transform of the autocorrelation function of the signal if the signal is treated as a wide-sense stationary random process [129]. Therefore, the Fourier transform of $R_c(\tau)$ is the PSD $S(f)$,

$$S(f) = \int_{-\infty}^{\infty} R_c(\tau) e^{-2\pi i f \tau} d\tau \tag{3.15}$$

$S(f)$ is a real-valued, nonnegative function. Definition (3.15) shows that $S(-f) = S(f)$, i.e., the PSD is an even function of frequency f.

Taking Discrete Fourier Transform (DFT) of these lags of autocorrelation values, we could obtain the power spectral density as

$$S(n) = \sum_{k=0}^{N-1} \hat{R}(k) e^{-i\frac{2\pi}{N}kn}, k = 0, 1, \ldots, N-1 \tag{3.16}$$

Next, we will separately describe how nested sampling and co-prime sampling estimate the autocorrelation function.

3.2.3.1 For Nested Sampling

For the samples obtained from nested sampling, consider the product $x(n_1)x^*(n_2)$, with n_1 and n_2 belong to the first period in fig. 3.1. We will get the samples at the following locations

$$1, 2, \ldots, N_1, (N_1+1), 2(N_1+1), \cdots, N_2(N_1+1) \tag{3.17}$$

The set of differences $n_1 - n_2$ are exactly the difference-co-array described in (3.2), that is, $n_1 - n_2$ will achieve all integer values in (3.2).

we can see that although the signal is sampled sparsely and nonuniformly at $1 \leq l \leq N_1$ and $(N_1+1)m$,$1 \leq m \leq N_2$ for one period, the autocorrelation $R_c(\tau)$ of the signal $x_c(t)$ could be estimated at all lags $\tau = k$.

An estimate of the autocorrelation samples for all k could be obtained [125] by averaging the products $x(n_1)x^*(n_2)$ over L periods,

$$\hat{R}(k) = \frac{1}{L}\sum_{l=0}^{L-1} x(n)x^*(n-k) \tag{3.18}$$

3.2.3.2 For Coprime Sampling

As P and Q are co-prime, there exist integers $0 \leq n_1 \leq 2Q-1$ and $0 \leq n_2 \leq P-1$, such that the difference in equation (3.9) can achieve any integer value $k = Pn_1 - Qn_2$ in the range of $0 \leq k \leq PQ-1$. Since $k = P(n_1+Ql) - Q(n_2+Pl)$ for any l, we can average l to obtain an estimate of the autocorrelation $R(k)$, that is,

$$\hat{R}(k) = \frac{1}{L}\sum_{l=0}^{L-1} x(P(n_1+Ql))x^*(Q(n_2+Pl)) \tag{3.19}$$

3.3 Rate Distortion Performance

Information rate distortion function is a measure of distortion between the original source and its representation. Sparse sampling can cause possible distortion because less number of samples are used. Our purpose is to construct a distortion function which can measure the distortion because of these two sparse sampling algorithms, either nested sampling (NS) or coprime sampling (CS). A wide variety of distortion functions, such as Euclidean distance, Hamming distance, Mahalanobis distance, Itakura-Saito distance have been used. Squared-error distortion is used here. The original samples are denoted as $x_i, i = 1, \cdots, L$, where L is the total number of samples. Assume that all original information from L samples is $X^L = [x_1, x_2, \cdots, x_L]$, the selected information after sparse sampling can be represented as [116]

$$\hat{X}^{L'} = S(X^L) \tag{3.20}$$

where $S(\cdot)$ denotes sparse sampling, either nested sampling or coprime sampling. $\hat{X}^{L'} = [\hat{x}_1, \hat{x}_2, \cdots, \hat{x}_{L'}]$ and $L' < L$. The distortion associated with the sparse sampling between all original samples and the selected samples is

$$D = Ed(X^L, \hat{X}^{L'}) \tag{3.21}$$

where $d(\cdot)$ is the distortion function.

The expectation in (4.7) is with respect to the probability distribution on X^L. The rate distortion function $R(D)$ is the minimum of data rates R such that (R, D) is in the rate distortion region for a given distortion. From [36][114], we know that information rate distortion function is defined as

$$R(D) = \min_{Ed(X^L, \hat{X}^{L'}) \leq D} I(X^L; \hat{X}^{L'}) \tag{3.22}$$

where $I(X^L; \hat{X}^{L'})$ is the mutual information between X^L and $\hat{X}^{L'}$.

$$\begin{aligned} I(X^L; \hat{X}^{L'}) &= H(X^L) - H(X^L|\hat{X}^{L'}) \\ &= H(X^L) - H(X^L - \hat{X}^{L'}|\hat{X}^{L'}) \\ &\overset{(h)}{\geq} H(X^L) - H(X^L - \hat{X}^{L'}) \end{aligned} \tag{3.23}$$

where inequality (h) follows from the fact that condition reduces the entropy.

From formula (3.22), we know that

$$Ed(X^L, \hat{X}^{L'}) \leq D \tag{3.24}$$

For squared error distortion,

$$\begin{aligned} Ed(X^L, \hat{X}^{L'}) &= E(X^L - \hat{X}^{L'})^2 \\ &= \sum E(x_i - \hat{x}_j)^2 \\ &\overset{(i)}{=} \sum_k D_k \leq D \end{aligned} \tag{3.25}$$

where $i = 1, \cdots, L$ and $j = 1, \cdots, L'$, and (i) follows from the definition that $E(x_i - \hat{x}_j)^2 = D_k$.

Since Gaussian assumption is a classical modeling assumption heavily used in areas such as signal processing and communication system [41], from [36], the rate distortion function for a single Gaussian source $N(0, \sigma^2)$ with squared error distortion is

$$R(D) = \begin{cases} \frac{1}{2} \log \frac{\sigma^2}{D} & 0 \leq D \leq \sigma^2, \\ 0 & D > \sigma^2 \end{cases} \tag{3.26}$$

For L independent zero-mean Gaussian sources $x_1, \cdots, x_L$ with variance $\sigma_1^2, \sigma_2^2, \cdots, \sigma_L^2$, the rate distortion performance with squared-error distortion is given by [36] [114] [112] [109]

$$R(D) = \sum_{i=1}^{L} \frac{1}{2} \log \frac{\sigma_i^2}{D_i} \tag{3.27}$$

where

$$D_i = \begin{cases} \lambda & \text{if } \lambda < \sigma_i^2, \\ \sigma_i^2 & \text{if } \lambda \geq \sigma_i^2 \end{cases} \tag{3.28}$$

where λ is chosen so that $\sum_i^L D_i = D$, and $D_i = E(x_i - \hat{x}_i)^2$. This gives rise to a kind of reverse waterfilling. We choose a constant λ and only describe those random variables with variance greater than λ, and no bits are used to describe random variables with variance less than λ.

3.3.1 For Nested Sampling

Theorem 9. *(Rate distortion for nested sampling of Gaussian source) Let $x_i \sim N(0, \sigma_i^2)$, $i = 1, 2, \cdots, L$, be independent Gaussian random variables, and under squared error distortion. The rate distortion between the original Gaussian source and after nested sampling of these Gaussian random variables is given by*

$$R_{NS}(D) = \sum_{i=1}^{L} \frac{1}{2} \log 2\pi e \sigma_i^2 - \sum_{k=1}^{K_{NS}} \frac{1}{2} \log 2\pi e D_k \tag{3.29}$$

where K_{NS} is given in (3.33) and

$$D_k = \begin{cases} \lambda & \text{if } \lambda < \sigma_k^2, \\ \frac{K_{NS}}{L} \sigma_k^2 & \text{if } \lambda \geq \sigma_k^2. \end{cases} \tag{3.30}$$

where λ is chosen so that $\sum_{k=1}^{K_{NS}} D_k = D$.

Proof. For nested sampling (NS), all L original information is $X^L = [x_1, x_2, \cdots, x_L]$.

And less number of samples $L^{'}$ will be selected based on nested sampling as described,

$$\begin{aligned} \hat{X}_{NS}^{L^{'}} &= [\hat{x}_1, \hat{x}_2, \cdots, \hat{x}_{L^{'}}] \\ &= [x_1, \cdots, x_{N_1}, x_{(N_1+1)}, \cdots, x_{N_2(N_1+1)}, \cdots] \end{aligned} \tag{3.31}$$

Therefore, (3.25) becomes

$$Ed(X^L, \hat{X}_{NS}^{L'}) = \sum_{k=1}^{K_{NS}} D_k \tag{3.32}$$

where the length of K_{NS} could be determined based on the following formula, here we assume $Y = L \pmod{(N_1+1)N_2}$

$$K_{NS} = \begin{cases} \lfloor \frac{L}{(N_1+1)N_2} \rfloor N_1(N_2-1), & \text{if } Y \leq (N_1+1) \\ \lfloor \frac{L}{(N_1+1)N_2} \rfloor N_1(N_2-1) + Z, & \text{if } Y > (N_1+1) \end{cases} \tag{3.33}$$

where

$$Z = \lfloor (\frac{Y}{N_1+1} - 1) \rfloor * N_1 + U \tag{3.34}$$

in which $U = (Y - (N_1+1)) \pmod{(N_1+1)}$.

If all samples are assumed to be independent Gaussian $N(0, \sigma_i^2)$, hence, the corresponding rate distortion function for nested sampling will be

$$\begin{aligned} R_{NS}(D) &= \min_{Ed(X^L, \hat{X}_{NS}^{L'}) \leq D} I(X^L; \hat{X}_{NS}^{L'}) \\ &\geq \min_{Ed(X^L, \hat{X}_{NS}^{L'}) \leq D} H(X^L) - H(X^L - \hat{X}_{NS}^{L'}) \\ &\overset{(j)}{\geq} \min_{\sum_{k=1}^{K_{NS}} D_k = D} H(X^L) - H(N(0, E(X^L - \hat{X}_{NS}^{L'})^2) \\ &= \min_{\sum_{k=1}^{K_{NS}} D_k = D} \sum_{i=1}^{L} \frac{1}{2} \log 2\pi e \sigma_i^2 - \sum_{k=1}^{K_{NS}} \frac{1}{2} \log 2\pi e D_k \end{aligned} \tag{3.35}$$

where inequality (j) follows from the fact that the normal distribution maximizes the entropy for a given second moment, and $\sum_{k=1}^{K_{NS}} D_k = D$.

To find the minimum value, we could using Lagrange multipliers

$$J(D) = \sum_{i=1}^{L} \frac{1}{2} \log 2\pi e \sigma_i^2 - \sum_{k=1}^{K_{NS}} \frac{1}{2} \log 2\pi e D_k + \sum_{k=1}^{K_{NS}} D_k \tag{3.36}$$

and differentiating with respect to D_k and setting equal to 0, we have

$$\frac{\partial J}{\partial D_k} = -\frac{1}{D_k 2 \ln 2} + \lambda = 0 \tag{3.37}$$

or

$$D_k = \frac{1}{\lambda 2 \ln 2} = \lambda^{'} \tag{3.38}$$

which results in an equal distortion for each random variable, if the constant $\lambda^{'}$ is less than σ_i^2 for all i. As the increase of the total allowable distortion D, the constant $\lambda^{'}$ increases until it exceeds σ_i^2 for some i. Kuhn-Tucker conditions could be used to find the minimum in (3.35) if we increase the total distortion D. In this case, the Kuhn-Tucker conditions yield

$$\frac{\partial J}{\partial D_k} = -\frac{1}{D_k 2 \ln 2} + \lambda \tag{3.39}$$

Therefore,

$$D_k = \begin{cases} \lambda & \text{if } \lambda < \sigma_k^2, \\ \frac{K_{NS}}{L}\sigma_k^2 & \text{if } \lambda \geq \sigma_k^2. \end{cases} \tag{3.40}$$

where λ is chosen so that $\sum_{k=1}^{K_{NS}} D_k = D$. □

3.3.2 For Coprime Sampling

Theorem 10. *(Rate distortion for coprime sampling of Gaussian source) Let $x_i \sim N(0, \sigma_i^2)$, $i = 1, 2, \cdots, L$, be independent Gaussian random variables, and under squared error distortion. The rate distortion between the original Gaussian source and after coprime sampling of these Gaussian random variables is given by*

$$R_{CS}(D) = \sum_{i=1}^{L} \frac{1}{2} \log 2\pi e \sigma_i^2 - \sum_{k=1}^{K_{CS}} \frac{1}{2} \log 2\pi e D_k \tag{3.41}$$

where K_{CS} is given in (3.45) and

$$D_k = \begin{cases} \lambda & \text{if } \lambda < \sigma_k^2, \\ \frac{K_{CS}}{L}\sigma_k^2 & \text{if } \lambda \geq \sigma_k^2. \end{cases} \tag{3.42}$$

where λ is chosen so that $\sum_{k=1}^{K_{CS}} D_k = D$.

Proof. For coprime sampling (CS), we still assume the original information with length L, i.e., $X^L = [x_1, x_2, \cdots, x_L]$.

And based on coprime sampling, less number of samples $L^{''}$ will be selected,

$$\begin{aligned} \hat{X}_{CS}^{L^{''}} &= [\hat{x}_1, \hat{x}_2, \cdots, \hat{x}_{L^{''}}] \\ &= [x_P, x_Q, x_{2P}, x_{2Q}, \cdots] \end{aligned} \tag{3.43}$$

Similarly, (3.25) becomes

$$Ed(X^L, \hat{X}_{CS}^{L^{''}}) = \sum_{k=1}^{K_{CS}} D_k \tag{3.44}$$

where the length of K_{CS} could be determined based on the following formula

$$K_{CS} = L - \lfloor \frac{L}{P} \rfloor - \lfloor \frac{L}{Q} \rfloor + \lfloor \frac{L}{PQ} \rfloor \tag{3.45}$$

Therefore, the corresponding rate distortion function for coprime sampling of independent Gaussian source $N(0, \sigma_i^2)$ is

$$\begin{aligned} R_{CS}(D) &= \min_{Ed(X^L, \hat{X}_{CS}^{L^{''}}) \leq D} I(X^L; \hat{X}_{CS}^{L^{''}}) \\ &\geq \min_{Ed(X^L, \hat{X}_{CS}^{L^{''}}) \leq D} H(X^L) - H(N(0, E(X^L - \hat{X}_{CS}^{L^{''}})^2) \\ &= \min_{\sum D_k = D} \sum_{i=1}^{L} \frac{1}{2} \log 2\pi e \sigma_i^2 - \sum_{k=1}^{K_{CS}} \frac{1}{2} \log 2\pi e D_k \end{aligned} \tag{3.46}$$

The minimum value could be obtained using the similar procedure as described in nested sampling. □

3.3.3 Theoretical Analysis

Without sparse sampling, the rate distortion function would be

$$\begin{aligned} R_{WSS}(D) &= \min_{Ed(X^L, X^L) \leq D} I(X^L; X^L) \\ &\geq H(X^L) - H(N(0, E(X^L - X^L)^2) \\ &= \sum_{i=1}^{L} \frac{1}{2} \log 2\pi e \sigma_i^2 - 0 \\ &= \sum_{i=1}^{L} \frac{1}{2} \log 2\pi e \sigma_i^2 \end{aligned} \tag{3.47}$$

which is much greater than that with sparse sampling.

From the above derivation of rate distortion function of nested sampling and coprime sampling, we could notice that if the sampling spacings are assumed to be the same, i.e., $N_1 = P$ and $N_2 = Q$ for these two sparse sampling methods, then the minimum value of $K_{NS_{min}}$ could be achieved when $Y = L \pmod{(N_1 + 1)N_2} = 0$, therefore

$$K_{NS_{min}} = \frac{N_1(N_2 - 1)L}{(N_1 + 1)N_2} = \frac{P(Q-1)L}{(P+1)Q} = \frac{(P^2Q - P^2)L}{PQ(P+1)} \tag{3.48}$$

While for coprime sampling, the minimum value of $K_{CS_{min}}$ could be achieved when $L \pmod{P} = 0$, $L \pmod{Q} = 0$, and $L \pmod{PQ} = 0$, therefore

$$K_{CS_{min}} = \frac{(P-1)(Q-1)L}{PQ} = \frac{(P^2Q - P^2 - Q + 1)L}{PQ(P+1)} \tag{3.49}$$

As we know that for these two sparse sampling algorithms, the sampling interval is for sure greater than Nyquist sampling spacing, which indicates that $Q > 1$, therefore,

$$K_{NS_{min}} > K_{CS_{min}} \tag{3.50}$$

which indicates that in most cases, $K_{NS} > K_{CS}$. Table 3.1 shows some example of K_{NS} and K_{CS} with respect to sampling intervals when $N_1 = P$, $N_2 = Q$, and

$L = 1000$. It is clear that with the increase of sampling spacings, samples are selected more sparsely by both nested sampling and coprime sampling, which results in a increase of K_{NS} and K_{CS}. In addition, we could notice that $K_{NS} > K_{CS}$ as proved.

Table 3.1. K_{NS} and K_{CS}, when $N_1 = P$, $N_2 = Q$, and $L = 1000$.

$N_1 = P$	$N_2 = Q$	K_{NS}	K_{CS}
3	4	561	500
3	5	600	533
3	7	642	572
3	11	681	607
3	13	690	616
3	17	705	628
3	23	717	638
5	23	794	765
7	23	833	821
11	23	873	870

With our assumption that all samples are independent Gaussian $N(0, \sigma_i^2)$, we could conclude that

$$R_{NS}(D) < R_{CS}(D) < R_{WS}(D) \tag{3.51}$$

which indicates that both nested sampling and coprime sampling use less number of bits to describe the information compared that without sparse sampling (WS).

As we know from the introduction part, in $(N_1 + 1)N_2T$ seconds, there are totally $N_1 + N_2$ samples for nested sampling, while coprime sampling totally collect $P + Q$ samples in PQT seconds. If the sampling intervals are the same, i.e., $N_1 = P$ and $N_2 = Q$, it is obvious that nested sampling is a little sparser than coprime sampling method. $R_{NS}(D) < R_{CS}(D)$ is because nested sampling collects a little less number of samples than coprime sampling with the same length L of data. The

rate $R(D)$ at a given distortion for both sparse sampling algorithms are less than that without sparse sampling. The reason is that with sparse sampling, less number of bits is used to describe the original information.

3.3.4 Numerical Results

The total length of the information is set to be $L = 1000$. Each sample is assumed to follow a Gaussian distribution $N(0, 1)$ with zero mean and unit variance. We also assume $D_k = \lambda < \sigma^2 = 1$, which is equal distortion for each random variable.

Fig. 3.3 shows the rate distortion performance of nested sampling with different sampling spacings. It is clear that with the increasing of distortion, the rate decreases. When the sampling intervals N_1 and N_2 becomes larger, i.e., less samples are acquired, the rate becomes smaller. For example, when $D = 0.3$, $N_1 = 3, N_2 = 5$, the data rate $R(D) \approx 1350$, while with the increase of sampling pairs to $N_1 = 3, N_2 = 11$, then $R(D) \approx 1220$, which is much smaller. This is because with more sparse sampling, less number of bits is required to represent the information.

The rate distortion performance of coprime sampling with different sampling spacings is shown in Fig. 3.4. Similarly as nested sampling, with the increasing of distortion, the rate $R(D)$ decreases. When the sampling intervals P and Q becomes larger, the rate becomes smaller.

Fig. 3.5 compares the rate distortion performance between nested sampling and coprime sampling, where D is the distortion between the original source and its sparse-sampled representation, and $R(D)$ is the corresponding rate at a particular distortion D. With the same sampling spacings chosen, $N_1 = P$, and $N_2 = Q$, at the same distortion, the rate of nested sampling is less than that of coprime sampling. For example, when $N_1 = P = 3$, and $N_2 = Q = 17$, when $D = 0.3$, the rate for nested sampling is $R_{NS}(D) \approx 1200$, while the rate for coprime sampling is $R_{CS}(D) \approx 1300$.

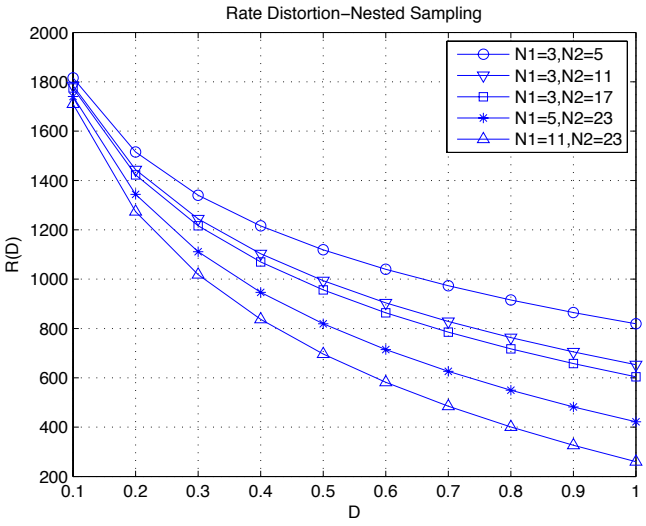

Figure 3.3. Rate Distortion Performance-Nested Sampling.

This verifies the result that $R_{NS}(D) < R_{CS}(D)$, because nested sampling collects a little less number of samples than coprime sampling with the same length L of data, which is a little sparser than coprime sampling.

3.4 Spectrum Efficiency

3.4.1 Theoretical Analysis

As nested sampling and coprime sampling are similar, in this part, nested sampling will be used to state the theoretical analysis.

From the property of DFT, we know that, if $\hat{R}(k)$ are real, then $S(N-n)$ and $S(n)$ are related by

$$S(N-n) = \bar{S}(n) \tag{3.52}$$

for $n = 0, 1, \ldots, N-1$, where $\bar{S}(n)$ denotes the complex conjugate. This also means that the component $S(0)$ is always real for real data.

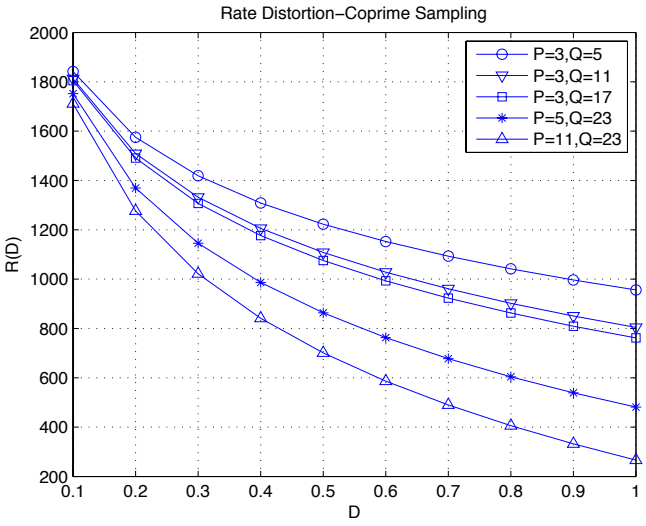

Figure 3.4. Rate Distortion Performance-Coprime Sampling.

In our simulation, we always consider the amplitude of the PSD, which indicates that

$$|S(N-n)| = |\bar{S}(n)| \tag{3.53}$$

This gives the reason of why the PSD figure is always symmetric.

From the simulation, we observe the absolute values of the autocorrelation $\hat{R}(k)$ are the same for the QPSK signal, which obtain positive or negative of a fixed value as shown in Figure 3.6 for different N_1 and N_2 of nested sampling, and Figure 3.7 for different P and Q of coprime sampling. This could make the calculation of the PSD easier. In our analysis, for simplicity, we assume all the $\hat{R}(k)$ have the same absolute value R, i.e., $R = |\hat{R}(k)| = \hat{R}(0) = -\hat{R}(1) = \hat{R}(2) = \ldots$. Therefore, we set $\hat{R}(k) = (-1)^k R$. The estimated autocorrelation satisfies those properties we stated before, i.e., $\hat{R}(0) \geq |\hat{R}(k)|, k \neq 0$, and as the QPSK signal we used is periodic, the estimated autocorrelation function $\hat{R}(k)$ is also periodic.

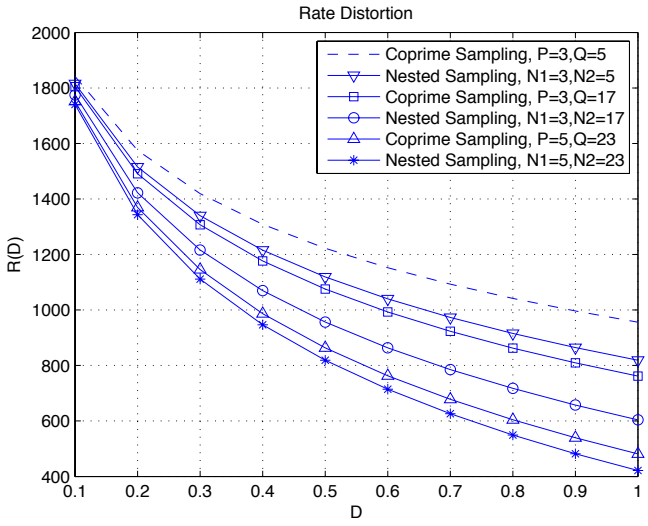

Figure 3.5. Comparison of Nested Sampling and Coprime Sampling.

Let's see how the PSD changes with the increase of N_1 and N_2. As stated in the principle of nested sampling, k falls in the range of (3.2). Here we only use those positive, i.e., $k = 0, 1, \ldots, (N_1+1)N_2 - 1$, that is, $N = (N_1+1)N_2$. We could get the PSD by taking the Fourier transform of the estimated autocorrelation,

$$\begin{aligned} S(n) &= \sum_{k=0}^{N-1} \hat{R}(k) e^{-i\frac{2\pi}{N}kn} \\ &= \sum_{k=0}^{N-1} (-1)^k R e^{-i\frac{2\pi}{N}kn} \\ &= R \sum_{k=0}^{N-1} (-e^{-i\frac{2\pi n}{N}})^k \end{aligned} \tag{3.54}$$

As stated in the principle of nested sampling, k falls in the range of (3.2), $N = (N_1+1)N_2$. For coprime sampling, we show that $k = P(n_1 + Ql) - Q(n_2 + Pl)$ can achieve any integer value in the range of $0 \le k \le PQ - 1$, i.e., in this case $N = PQ$. N could either be even or odd. For example, for nested sampling, if $N_1 = 2$ and $N_2 = 5$, then $N = (N_1+1)N_2 = 15$ is odd, while if $N_1 = 3$ and $N_2 = 5$, then $N = (N_1+1)N_2 = 20$ is even. For coprime sampling, if $P = 2$, $Q = 5$, then

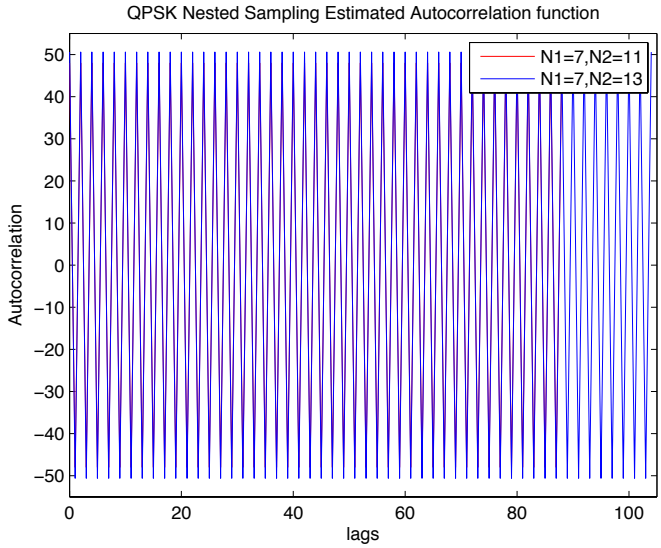

Figure 3.6. Nested Sampling Estimated autocorrelation.

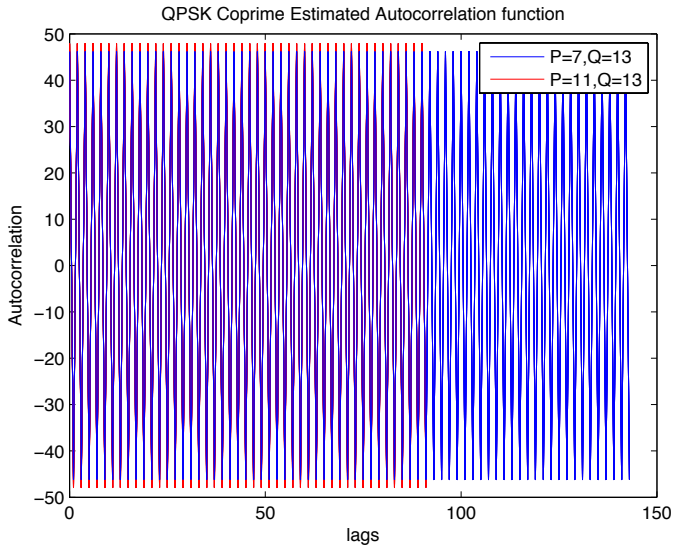

Figure 3.7. Coprime Sampling Estimated autocorrelation.

$N = PQ = 10$ is even, while if $P = 3$, $Q = 5$, then $N = PQ = 13$ is odd. First, we assume $N = (N_1 + 1)N_2$ is odd,

$$\begin{aligned} S(n) &= R\sum_{k=0}^{N-1}(-e^{-i\frac{2\pi n}{N}})^k \\ &= R\frac{1+e^{-i\frac{2\pi n}{N}N}}{1+e^{-i\frac{2\pi n}{N}}} \\ &= R\frac{e^{-i\pi n}(e^{i\pi n}+e^{-i\pi n})}{e^{-i\frac{\pi n}{N}}(e^{i\frac{\pi n}{N}}+e^{-i\frac{\pi n}{N}})} \\ &= R\cdot e^{-i\pi n\frac{N-1}{N}}\frac{cos(\pi n)}{cos(\frac{\pi n}{N})} \end{aligned} \tag{3.55}$$

And if $N = (N_1 + 1)N_2$ is even,

$$\begin{aligned} S(n) &= R\sum_{k=0}^{N-1}(-e^{-i\frac{2\pi n}{N}})^k \\ &= R\frac{1-e^{-i\frac{2\pi n}{N}N}}{1+e^{-i\frac{2\pi n}{N}}} \\ &= R\frac{e^{-i\pi n}(e^{i\pi n}-e^{-i\pi n})}{e^{-i\frac{\pi n}{N}}(e^{i\frac{\pi n}{N}}+e^{-i\frac{\pi n}{N}})} \\ &= R\cdot e^{-i\pi n\frac{N-1}{N}}\frac{isin(\pi n)}{cos(\frac{\pi n}{N})} \end{aligned} \tag{3.56}$$

The corresponding amplitude of the PSD is

$$|S(n)| = R|\frac{cos(\pi n)}{cos(\frac{\pi n}{N})}|(Nodd), or, R|\frac{sin(\pi n)}{cos(\frac{\pi n}{N})}|(Neven) \tag{3.57}$$

We could draw these two expressions in (3.57) as shown in Figures 3.8, 3.9, 3.10, and 3.11. It's obvious that no matter N is odd or even, with the increase of N, the mainlobe becomes narrower and the number of sidelobes increases. In next paragraph, we will prove the central of the mainlobe represents the central frequency.

In the simulation, if we take N-point Fast Fourier Transform (FFT), we will get N PSD values. Let f_n represent the Nyquist sampling frequency, $f_n = 2f_c$ (f_c is the carrier frequency), using $f = f_n \cdot (0 : N-1)/N$, we could map these PSD values

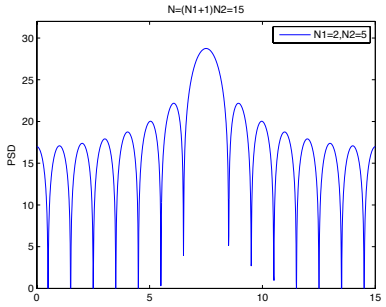

Figure 3.8. PSD, N=15.

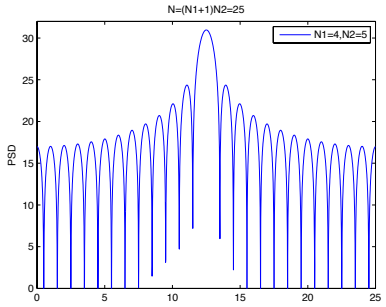

Figure 3.9. PSD, N=25.

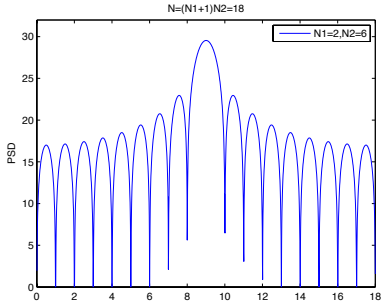

Figure 3.10. PSD, N=18.

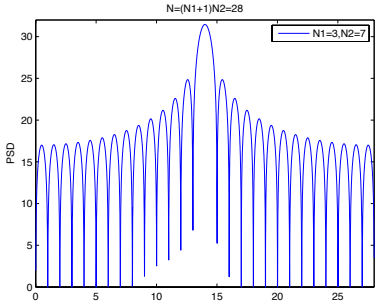

Figure 3.11. PSD, N=28.

to the frequency. It is obvious that when $n = N/2$, the PSD gets its central value of $S(\frac{N}{2})$ at $f = \frac{1}{2}f_n = f_c$. This could be represent as

$$\begin{aligned} S(\frac{N}{2}) &= \sum_{k=0}^{N-1} \hat{R}(k) e^{-i\frac{2\pi}{N}k\frac{N}{2}} \\ &= R\sum_{k=0}^{N-1}(-e^{-i\pi})^k \\ &= R\sum_{k=0}^{N-1} 1 = NR \end{aligned} \tag{3.58}$$

From this derivation, we also notice that with the increase of N, besides the mainlobe becomes narrower, the central value of the PSD gets higher, which results in a higher spectrum efficiency.

In the numerical results part, we will show that with the same sampling spacings chosen for both nested and coprime sampling, i.e., $N_1 = P$, $N_2 = Q$, we could achieve $N = (N_1 + 1)N_2$ for nested sampling will be larger than that of $N = PQ$ for coprime sampling, which will result in a better spectrum efficiency for nested sampling.

3.4.2 Simulation Results

This section presents some numerical results for the autocorrelation and power spectrum density estimation using nested sampling structure. We use QPSK modulated signal with carrier frequency $f_c = 400Hz$, which could be expressed as [118]

$$s_{QPSK}(t) = \sqrt{\frac{2E_s}{T_s}} cos[2\pi f_c t + (i-1)\frac{\pi}{2}] \tag{3.59}$$

where T_s is the symbol duration. In our simulation, we set $E_s = 1$ and $T_s = 1/50$.

The power spectrum density [118] of a QPSK signal using rectangular pulses can be expressed as

$$P_{QPSK}(f) = \frac{E_s}{2}[(\frac{sin\pi(f-f_c)T_s}{\pi(f-f_c)T_s})^2 + (\frac{sin\pi(-f-f_c)T_s}{\pi(-f-f_c)T_s})^2] \tag{3.60}$$

Fig. 3.12 shows the PSD of a QPSK signal for rectangular and raised cosine filtered pulses. The x-axis refers to the frequency in Hz, and the y-axis are the normalized power spectral density in dB. It can be observed the PSD centers at $f_c = 400Hz$ with symmetric sidelobes on both sides.

If we zoom in fig. 3.12, as shown in fig. 3.13, we could notice bandwidth for the original QPSK signal is about $416 - 384 \approx 32Hz$.

The estimated autocorrelation using nested sampling and co-prime sampling structures are plotted in fig. 3.14 and 3.15. In the simulation, for nested sampling, we use $N_1 = 7$, $N_2 = 11$, and $L = 10$. Therefore, $\hat{R}(k)$ can be estimated for

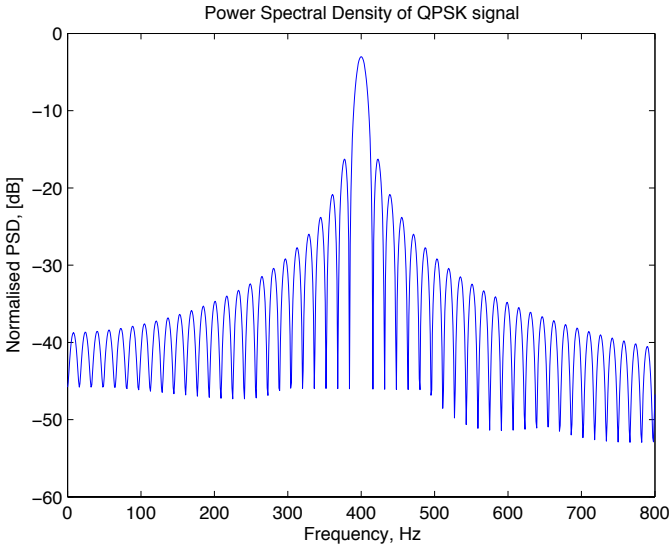

Figure 3.12. PSD of the QPSK signal.

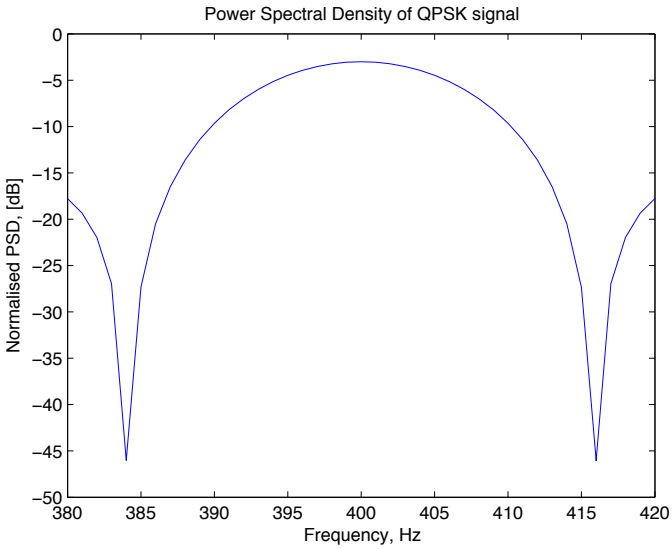

Figure 3.13. Zoom in the main lobe of PSD for QPSK signal.

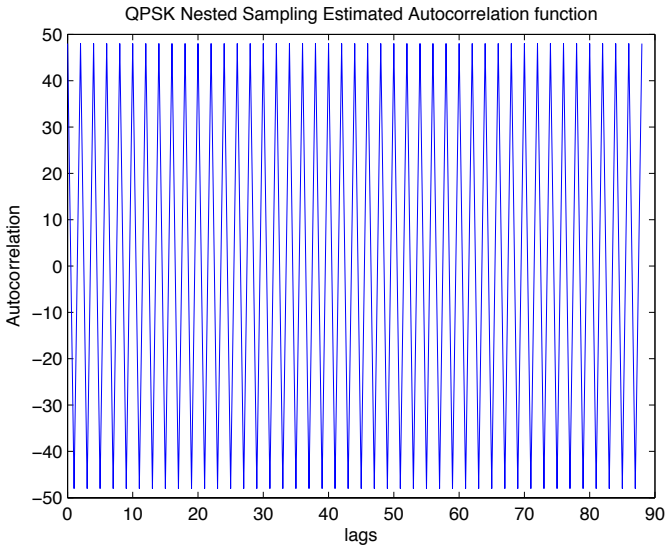

Figure 3.14. Nested Sampling Estimated Autocorrelation of QPSK Signal ($N_1 = 7$, $N_2 = 11$).

$| k | \leq (N_1+1)N_2 - 1$. For each period, we totally get $(N_1+1)N_2 = (7+1) \times 11 = 88$ samples. While for coprime sampling, we set $P = 7$ and $Q = 11$, for each period, we get $PQ = 7 \times 11 = 77$ lags of $\hat{R}(k)$.

Using the relationship of autocorrelation and the PSD described in section 3, we could obtain the estimated PSD using nested sampling structure for this QPSK signal as shown in fig. 3.16. In the simulation, we use 1024 point fast Fourier transform and normalize the PSD. We can see that the estimated PSD is also centered at $f_c = 400Hz$ with symmetric sidelobes on both sides. As stated in section 3, we can see the PSD is an even function.

Similarly, if we zoom in this PSD around the central frequency f_c, in fig. 3.17, we could find the main lobe, i.e., the bandwidth occupied is approximately $409 - 391 \approx 18Hz$, which is much narrower than that $32Hz$ of the PSD of the

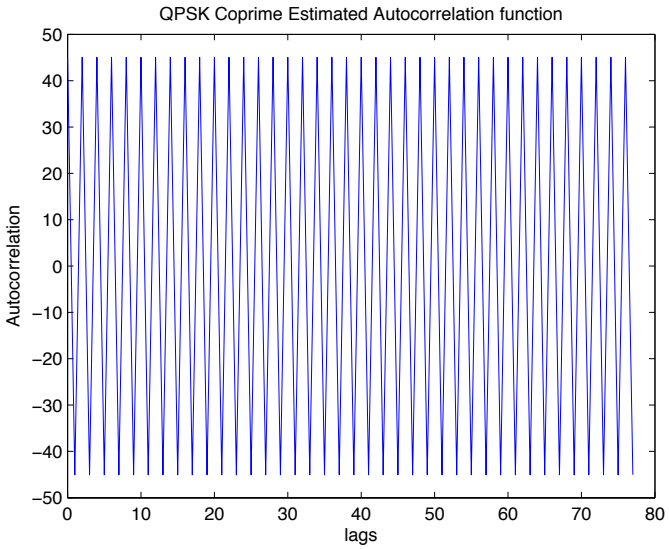

Figure 3.15. Coprime Sampling Estimated Autocorrelation of QPSK Signal ($P = 7$, $Q = 11$).

original QPSK signal. Hence, the spectrum efficiency is improved in the estimation using nested sampling structure.

We could also get the estimated PSD using co-prime sampling structure for this QPSK signal as shown in fig. 3.18. In the simulation, we use 1024 point fast Fourier transform and normalize the PSD. We can see that the estimated PSD is also an even function centered at $f_c = 400Hz$ with symmetric sidelobes on both sides.

If we zoom in this PSD around the central frequency f_c, in fig. 3.19, we could find the main lobe, i.e., the bandwidth occupied is approximately $411 - 389 \approx 22Hz$, which is near to that estimated using nested sampling and is much narrower than that $32Hz$ of the PSD of the original QPSK signal. Hence, the spectrum efficiency is improved in the estimation using co-prime sampling structure as well.

Another interesting observation is that the bandwidth of the PSD estimated using coprime sampling is a little larger than that estimated by nested sampling, as

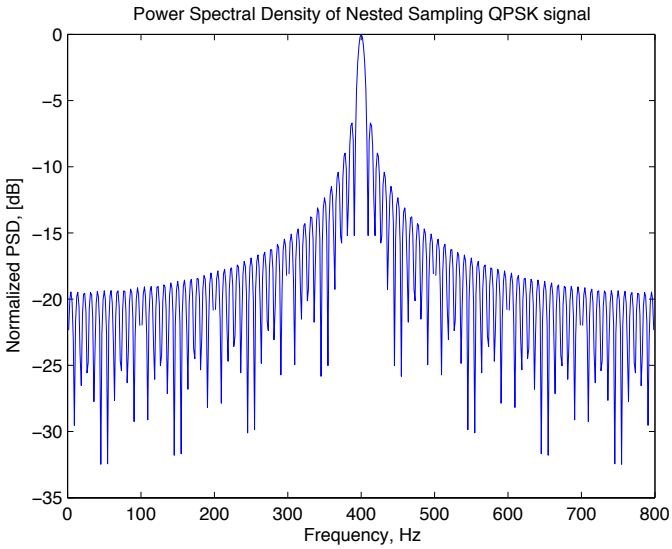

Figure 3.16. PSD of Nested Sampling QPSK signal (N1=7,N2=11).

shown in our example, the bandwidth for coprime estimated PSD is $411 - 389 \approx 22Hz$, while it is $409 - 391 \approx 18Hz$ for nested sampling. This is because for the same number of P and Q (or N_1 and N_2), the nested sampling could achieve $N = (N_1 + 1)N_2$, while coprime sampling could only get $N = PQ$. If $N_1 = P$ and $N_2 = Q$, it is obvious that the nested sampling estimate a larger number of N than coprime sampling. Refer to the theoretical analysis, we could conclude that larger N results in narrower bandwidth, which indicates that if $N_1 = P$ and $N_2 = Q$ for nested and coprime sampling, nested sampling would have a more efficient spectrum performance.

By changing different N_1 and N_2 pairs, as shown in fig. 3.20, it is obvious that for N_1 fixed to $N_1 = 3$, with the increase of the value of N_2 from $5, 7$ to 13, the main lobe of the estimated PSD using nested sampling structure becomes narrower significantly, i.e., the bandwidth occupied gets smaller. Here, in the simulation, we normalize the PSD values.

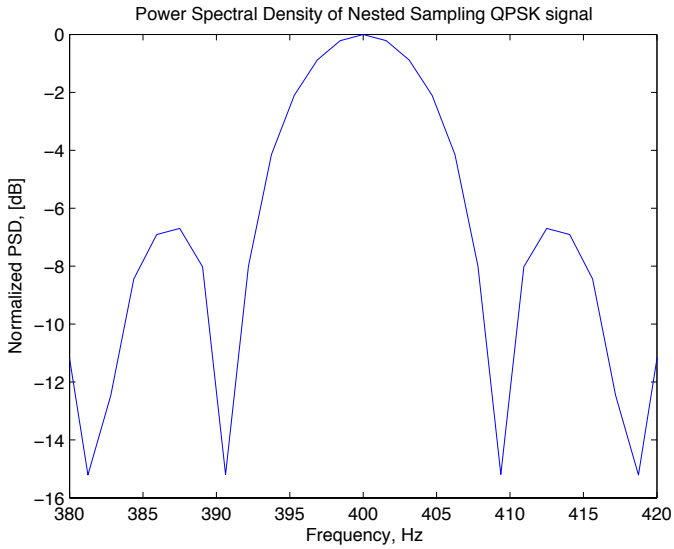

Figure 3.17. Zoom-in the mainlobe of PSD-Nested Sampling QPSK signal(N1=7,N2=11).

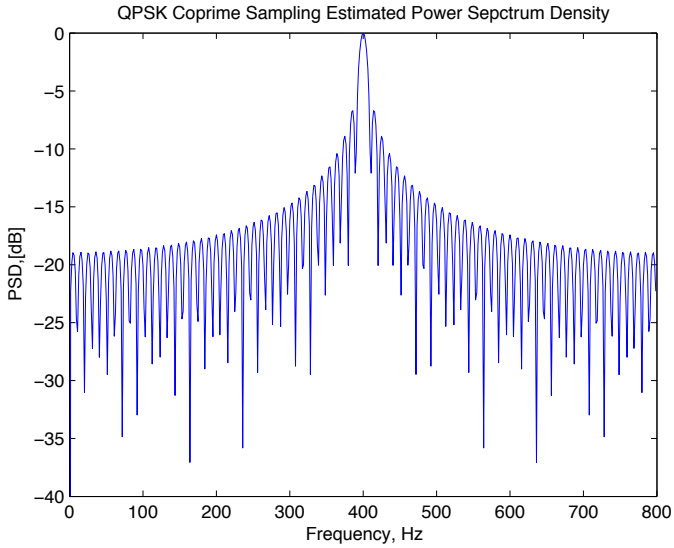

Figure 3.18. PSD of Co-Prime Sampling QPSK signal (P=7,Q=11).

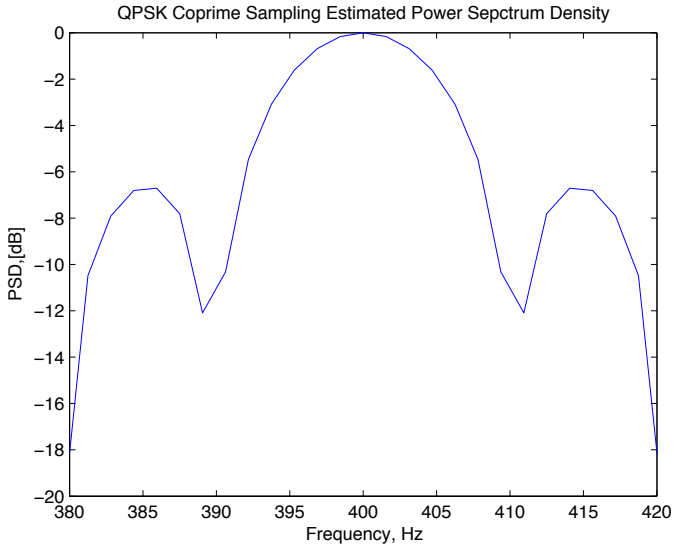

Figure 3.19. Zoom-in the mainlobe of PSD-Coprime Sampling QPSK signal(P=7,Q=11).

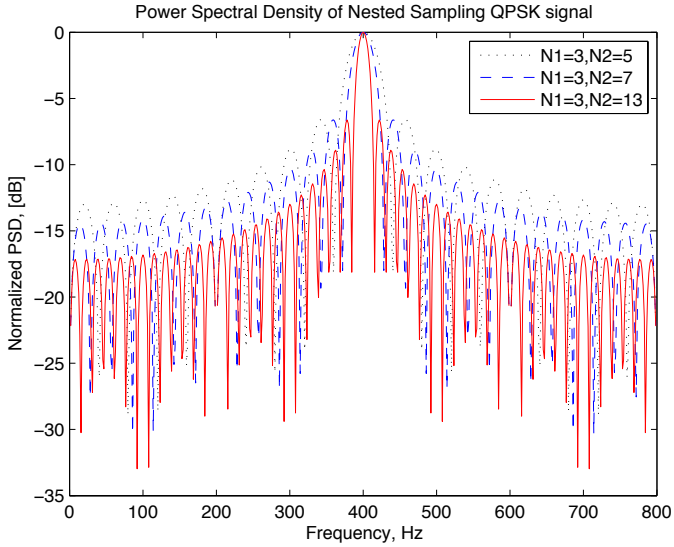

Figure 3.20. PSD of Nested Sampling QPSK signal with different N2.

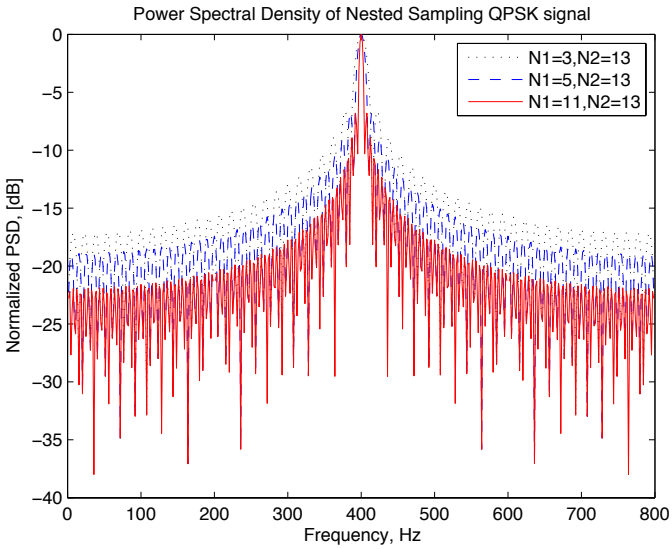

Figure 3.21. PSD of Nested Sampling QPSK signal with different N1.

Similarly, Fig. 3.21 shows that with the increase of N_1 from $N_1 = 3, 5$ to $N_1 = 11$, while N_2 fixed to $N_2 = 13$, the main lobe also gets narrower, which also results in the increase of spectrum efficiency. From the results got from Figures 3.20 and 3.21, we conclude that in the nested sampling process, besides its advantage of less samplers, with N_1 and N_2 chosen larger, the bandwidth of the PSD occupied will becomes narrower, which increases the spectrum efficiency.

Similar as nested sampling, as P and Q increase for co-prime sampling, the mainlobe of the estimated PSD narrows down as well, which also indicates smaller bandwidth and higher spectrum efficiency as shown in Fig. 3.22, where we increase the second sampler's sampling interval of Q from 5, 7, to 13, and in Fig. 3.23, where we increase the first sampler's sampling interval of P from 3, 5, to 11.

From Figures 3.20 to 3.23, we could observe nested sampling and co-prime sampling could obtain similar estimated PSD and both are spectrum efficient as the

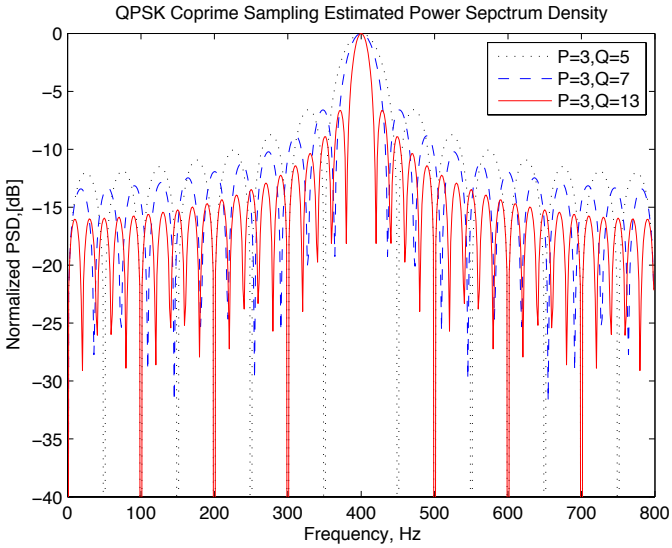

Figure 3.22. PSD of Co-Prime Sampling QPSK signal with different Q.

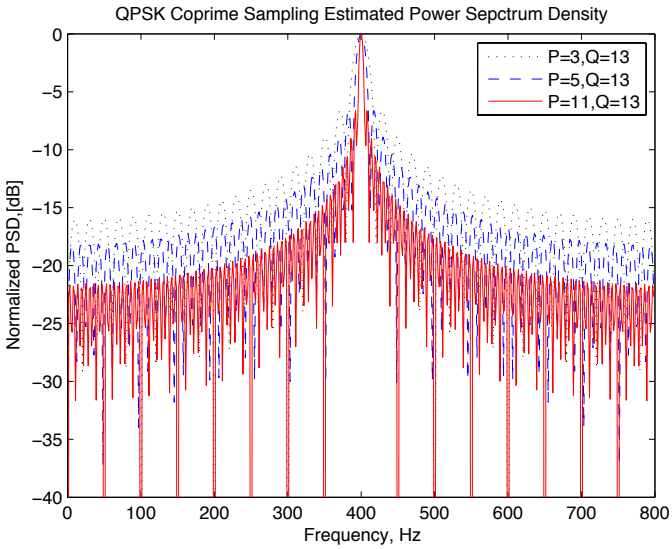

Figure 3.23. PSD of Co-Prime Sampling QPSK signal with different P.

sampling intervals increase, i.e., N_1 and N_2 for nested sampling, and P and Q for co-prime sampling.

3.5 Secure Transmission for Big Data

The security is a significant issue in modern data processing. Critics worry about Big Data may be misused and abused, especially for large corporations. Data experts and critics worry that potential abuses of inferential data may imperil personal privacy, and consumer freedoms[4]. Big Data presents critical requirements for security in data collection and transmission of selected data through a communication network. This section addresses a new secure transmission of Big Data based on nested sampling and coprime sampling.

As both nested sampling and coprime sampling could keep the statistical property of the original signal, these two sampling algorithms could be applied to Big Data to highly reduce the transmission cost of Big Data. In this section, we will prove that these two sampling methods could provide a secure transmission for Big Data based on their properties.

In this section, we give the principle of nested sampling and coprime sampling first. Both the theoretical analysis and the numerical results show that with the sampling spacings larger for both nested sampling and coprime sampling, the mainlobe of PSD (Power Spectral Density) obtained from these two sampling will be much narrower than the original BFSK signal. That is, besides the much less time consumption, the occupied bandwidth B is smaller. With the sampling spacings large enough, the mainlobe of the PSD will becomes as narrow as possible, performs like frequency hopping (FH). Furthermore, we will prove that for such kind of FH/BFSK signal generated by higher sampling spacing pairs, the error probability will be much smaller compared with the original signal with independent multitone interference.

It offers an improvement in performance when the communication system is attacked by jamming. This provide a secure transmission method for Big Data based on nested sampling and coprime sampling.

We also prove that [138] with the increase of N, ($N = (N_1 + 1)N_2$ for nested sampling, and $N = PQ$ for coprime sampling), besides the mainlobe becomes narrower, the central value of the PSD gets higher, which results in a higher spectrum efficiency. It could also be observed that with the same sampling spacings chosen for both nested and coprime sampling, i.e., $N_1 = P$, $N_2 = Q$, $N = (N_1 + 1)N_2$ for nested sampling will be larger than that of $N = PQ$ for coprime sampling, which will result in a higher PSD for nested sampling.

3.5.1 FH/BFSK with Independent Multitone Interference

The BFSK signal is

$$v(t) = \sqrt{2P}cos(\omega_o + d(t)\Omega)t \tag{3.61}$$

$$v_H(t) = \sqrt{2P}cos(\omega_o + \Omega)t \ (binary1) \tag{3.62}$$

$$v_L(t) = \sqrt{2P}cos(\omega_o - \Omega)t \ (binary0) \tag{3.63}$$

The higher frequency $\omega_H = \omega_o + \Omega$, and the lower frequency $\omega_L = \omega_o - \Omega$. Therefore, the BFSK signal is

$$v(t) = \sqrt{2P}H(t)cos(\omega_H t) + \sqrt{2P}L(t)cos(\omega_L t) \tag{3.64}$$

The carrier signals are

$$x(h) = \sqrt{2P}cos(\omega_H t) \tag{3.65}$$

$$x(l) = \sqrt{2P}cos(\omega_L t) \tag{3.66}$$

PSD of the carrier signals are

$$G_H(f) = \frac{P}{2}[\delta(f - f_H) + \delta(f + f_H)] \tag{3.67}$$

$$G_L(f) = \frac{P}{2}[\delta(f - f_L) + \delta(f + f_L)] \tag{3.68}$$

where $\delta()$ is Dirac function. $H(t)$ and $L(t)$ are Unipolar signals.

PSD of Unipolar signal is

$$G_s(f) = \frac{V^2 T_b}{4} sinc^2(fT_b) + \frac{V^2}{4}\delta(f) \tag{3.69}$$

Therefore, PSD of $H(t)$ and $L(t)$ are,

$$G_H(f) = G_l(f) = \frac{T_b}{4} sinc^2(fT_b) + \frac{1}{4}\delta(f) \tag{3.70}$$

Hence, PSD of the BFSK signal will be

$$\begin{aligned} G_v(f) &= G_H(f) * G_{x(h)}(f) + G_L(f) * G_{x(L)}(f) \\ &= \frac{P}{8}[sinc^2[(f - f_H)T_b] + sinc^2[(f + f_H)T_b]] \\ &+ \frac{P}{8}[sinc^2[(f - f_L)T_b] + sinc^2[(f + f_L)T_b]] \\ &+ \frac{P}{8}[\delta(f - f_H) + \delta(f + f_H)] + \frac{P}{8}[\delta(f - f_L) + \delta(f + f_L)] \end{aligned} \tag{3.71}$$

The signal we use here are BFSK signal with $\omega_o = 200Hz$, and frequency offset $\Omega = 10Hz$. $\sqrt{2P} = 1$, so that $P = 1/2$. The PSD of this BFSK signal is obtained as shown in Fig. 3.24.

We assume there are N_S nonoverlapping FH bands in the FH/BFSK communication system [131] [133]. Each FH band occupies B_h bandwidth, that is, B_h is the required bandwidth to transmit an BFSK signal in the absence of FH. Each FH band contains 2 orthogonal signaling tones. We also assume that each 2 signal

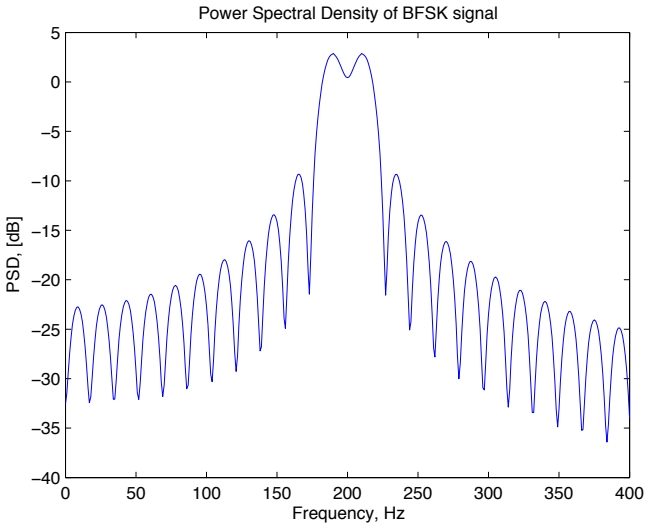

Figure 3.24. Original PSD for BFSK signal..

tones in each FH band are orthogonal to the signaling tones in all other FH bands. Therefore, thare are totally $2 \times N_S$ possible transmitted signaling tones.

The Power Spectral Density (PSD) of Additive White Gaussian Noise (AWGN) that corrupt the channel is defined as $N_0/2$. We assume that the independent multitone interference tones have a total power of P_{JT}, which is transmitted in the total of q equal power interfering tones and is spread uniformly over the spread spectrum bandwidth of the FH/BFSK communication system. Each interfering tone has a power of $P_J = P_{JT}/q$. We also assume that the multiple interfering tones are transmitted at the frequencies exactly corresponding to the possible $2 \times N$ signaling tones, and none of them are transmitted at the same frequency.

As we know from the theoretical analysis section, with the increase of N, besides the mainlobe becomes narrower, the central value of the PSD gets higher for both nested sampling and coprime sampling. It could also be observed that with the same sampling spacings chosen for both nested and coprime sampling, i.e., $N_1 = P$,

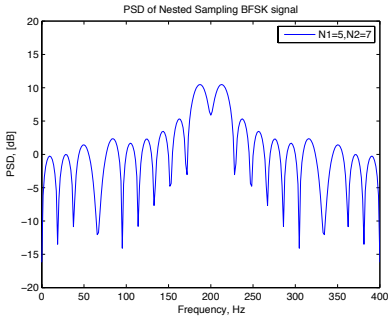

Figure 3.25. $N1 = 5, N2 = 7$.

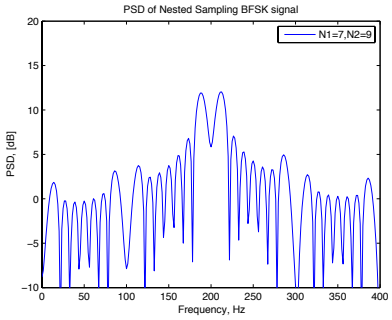

Figure 3.26. $N1 = 7, N2 = 9$.

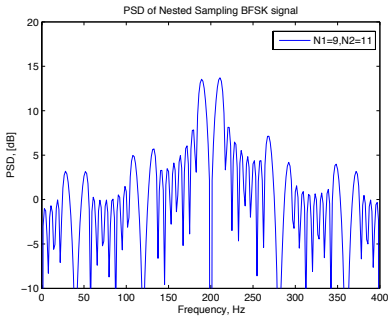

Figure 3.27. $N1 = 9, N2 = 11$.

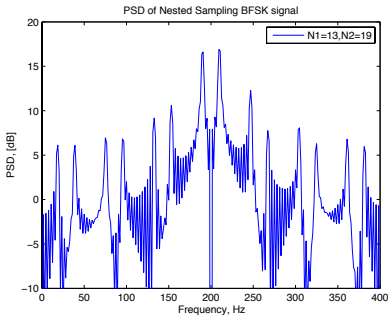

Figure 3.28. $N1 = 13, N2 = 19$.

$N_2 = Q$, $N = (N_1 + 1)N_2$ for nested sampling will be larger than that of $N = PQ$ for coprime sampling, which will result in a higher PSD for nested sampling. If we choose the sampling spacings higher, N will becomes higher, and the mainlobe will becomes much narrower, and the bandwidth will becomes much smaller. When N is high enough, the bandwidth will be as narrow as possible, which performs like frequency hopping. The following numerical results will show this performance using BFSK using both nested sampling as shown in Figures 3.25, 3.26, 3.27, and 3.28 and coprime sampling as shown in Figures 3.29, 3.30, 3.31, and 3.32. Besides the narrower bandwidth, as proved in the theoretical analysis, the PSD is higher with the increase of N, which is obvious from our numerical results.

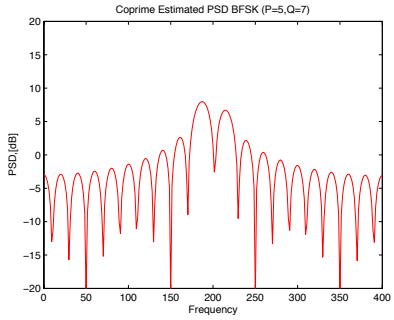

Figure 3.29. $M = 5, N = 7$.

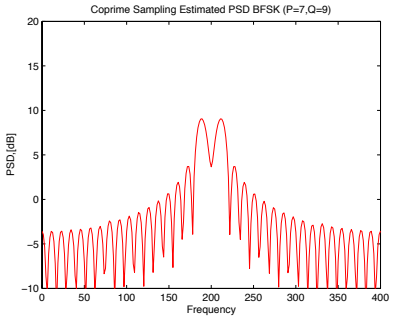

Figure 3.30. $M = 7, N = 9$.

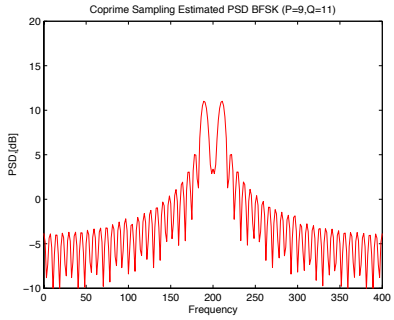

Figure 3.31. $M = 9, N = 11$.

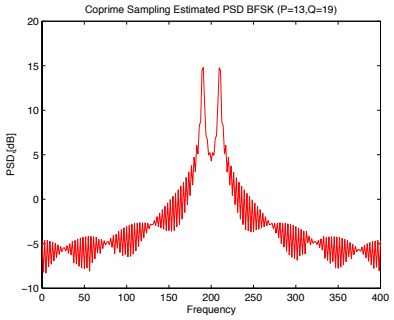

Figure 3.32. $M = 13, N = 19$.

3.5.2 Error Probability Analysis

The error probability for Rician fading multichannel reception of BFSK signal was discussed by William in [130]. For noncoherent receiver, we obtain the error probability of BFSK as the formula (54) in [130].

$$P_e(M) = [\frac{1}{2+\beta}]^M exp[\frac{-L\beta}{2+\beta}] \cdot \sum_{m=0}^{M-1} \sum_{n=0}^{m} \binom{m+M-1}{m-n} (\frac{1+\beta}{2+\beta})^m \frac{x^m}{m!} \tag{3.72}$$

where

$$x = \frac{L\beta}{(1+\beta)(2+\beta)} \tag{3.73}$$

where M is the number of multichannels. α_m $(m = 1, 2, \ldots, M)$ is the instantaneous voltage gain of the $m-$ th channel, $P = \sum_{m=1}^{M} \alpha_m^2$, $2\sigma^2$ is the mean squared value of the random or scatter component in each channel. $R_i = E_i/N_0$ $(i = 1, 2, \ldots, M)$

where, E_i is the energy of each signal transmitted by the $i-th$ transmitter. N_0 is the noise energy. For simplified, we assume $E_i = E$, therefore, $R = E/N_0$. $\beta = 2\sigma^2 R$, $\rho_m = \alpha_m^2 R$, and $L\beta = \sum_{m=1}^{M} \rho_m$.

As we discussed in earlier parts of this section, we prove that with the increase of the sampling space pairs for both nested sampling and coprime sampling, the PSD becomes higher, that is, the signal-to-noise ratio (SNR) gets higher for the same bandwidth BFSK signal. We could also observe from Numerical Results Section that with the sampling pairs increasing, the PSD becomes higher for both nested sampling and coprime sampling. The R increases compared to the original BFSK signal after nested sampling and coprime sampling, calculated using formula (3.72), the BER decreases as shown in Figures 3.33 and 3.36. With the increase of number of multichannels M, the error rates for the noncoherent BFSK receiver based on nested sampling and coprime sampling are shown. It's obvious that with nested sampling and coprime sampling, the signal could achieve higher PSD, which indicates higher SNR for the same bandwidth BFSK signal. Therefore, BFSK signals sampled by nested sampling or coprime sampling achieves lower error probability compared with the original BFSK signal. We could also observe that with the sampling spacing pairs increasing, i.e., N_1 and N_2 for nested sampling, and P and Q for coprime sampling, the error probability becomes lower and lower.

The performance for FH/BFSK with independent multitone interference was discussed in [131]. From formula (34) in [131], we could obtain the probability of error for FH/BFSK with independent multitone interference as

$$\begin{aligned} P_b = & \frac{q}{2N_S}(1 - \frac{q-1}{2N_S-1})P_b(hop\ jammed \mid 1\ jamming\ tone) \\ & + \frac{q}{2N_S}(\frac{q-1}{2N_S-1})P_b(hop\ jammed \mid 2\ jamming\ tones) \\ & + (1 - \frac{q}{2N_S})(1 - \frac{q}{2N_S-1})P_b(hop\ not\ jammed) \end{aligned} \tag{3.74}$$

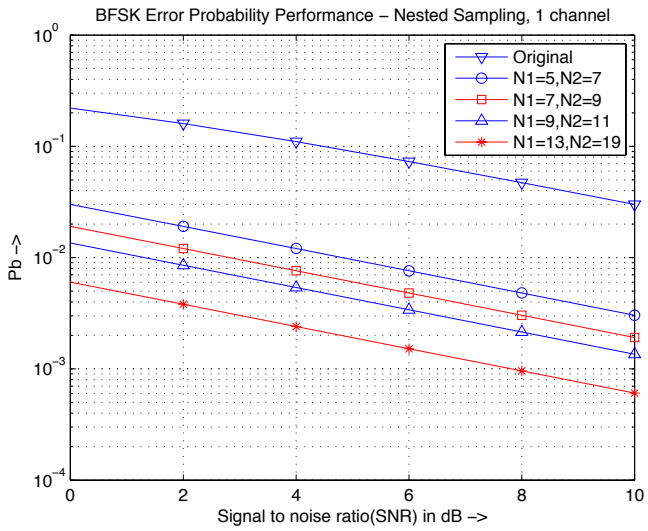

Figure 3.33. Nested Sampling 1 Channel BER.

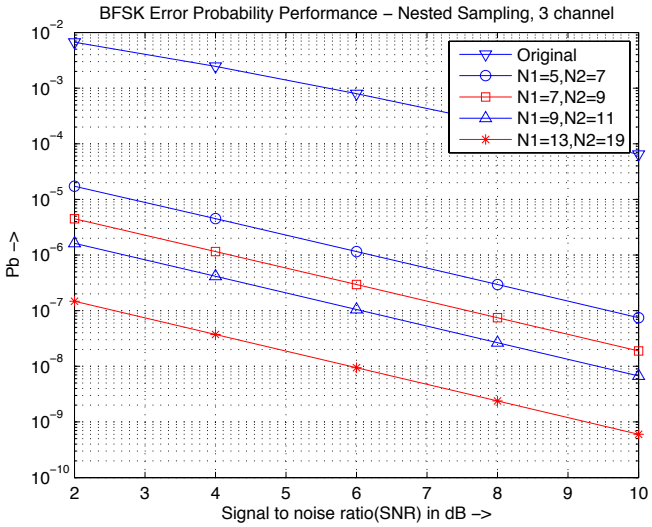

Figure 3.34. Nested Sampling 3 Channels BER.

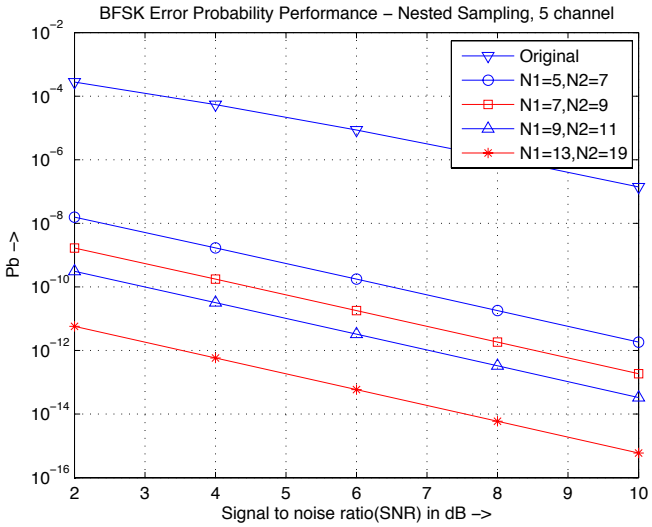

Figure 3.35. Nested Sampling 5 Channels BER.

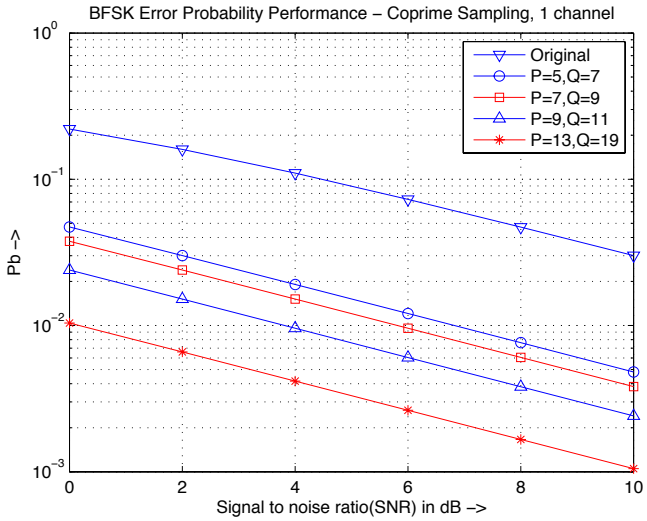

Figure 3.36. Coprime Sampling 1 Channel BER.

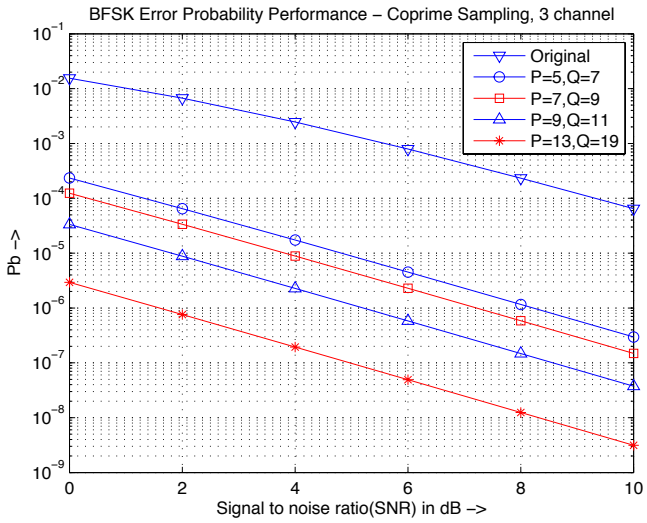

Figure 3.37. Coprime Sampling 3 Channels BER.

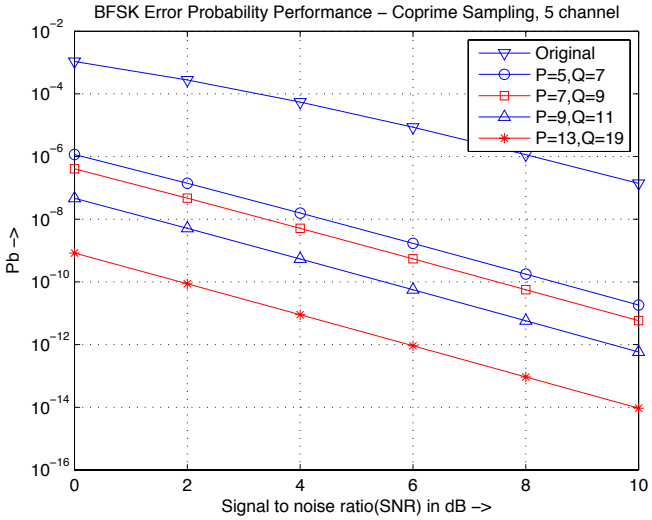

Figure 3.38. Coprime Sampling 5 Channels BER.

where q is the total number of interference tones, here, we set $q = 100$. $N_S = 1000$ is the total number of nonoverlapping FH bands.

From formula (8) in [132], we could obtain

$$\begin{aligned} &P_b(hop\ jammed \mid n\ jamming\ tones) \\ &\quad = (\frac{2-n}{2})P_b(hop\ jammed \mid signal\ is\ not\ jammed) \\ &\quad +(\frac{n}{2})P_b(hop\ jammed \mid signal\ is\ jammed) \end{aligned} \tag{3.75}$$

We assume the Rician Fading Factor K equals to $10dB$ for information signal i.e., $K_S = 10dB$. Following the same procedure described in [133] and [134], we could obtain the error probability v.s. SJR (Signal-to-Jamming-Ratio), $SJR = P_S/P_{JT}$, where P_S is the average information signal power, and P_{JT} is the total average interference power.

As we discussed, both nested sampling and coprime sampling could enhance PSD of the signal, furthermore, increase SNR on the same signal bandwidth and the same noise power. With the results we get, we could obtain the corresponding error probability of FH/BFSK signal with nocoherent detection in Rician fading channel with multitone jamming with total $q = 100$ jammings in $N_S = 1000$ frequency hops, for $K_S = 10dB$.

The results are shown in Figures 3.39 and 3.40 for nested sampling and coprime sampling. We can observe that with the sampling spacing pairs increasing, i.e., N_1 and N_2 for nested sampling, and P and Q for coprime sampling, the error probability gets lower. It is also obvious that nested sampling could achieve lower error probability compared with coprime sampling with the same sampling spacing pairs chosen, i.e., $P = N_1$ and $Q = N_2$, which is also proved in [138], as $N = (N_1+1)N_2$ for nested sampling will be larger than that of $N = PQ$ for coprime sampling, with the

same sampling spacings chosen for both nested and coprime sampling, i.e., $N_1 = P$, $N_2 = Q$, which will result in a higher PSD for nested sampling.

From Fig. 3.39, for the same error probability, for example, $Pb = 4 \times 10^{-3}$, for nested sampling, when $N_1 = 5, N_2 = 7$, it has almost $7dB$ gain compared with the original FH/BFSK signal. With the sampling spacing pairs N_1 and N_2 increase, when $N_1 = 7, N_2 = 9$, it has about $2.5dB$ gain compared with the case when $N_1 = 5, N_2 = 7$. Similarly, we can see that, the case when $N_1 = 9, N_2 = 11$ has about $2dB$ gain compared with that when $N_1 = 7, N_2 = 9$, and when $N_1 = 13, N_2 = 19$, it has about $3dB$ gain compared with that when $N_1 = 9, N_2 = 11$.

Same as stated above, we could see the advantage of coprime sampling with independent multitone interference. However, as we explained, coprime sampling the SNR gain is not that high as nested sampling, for example, when $P = 5, Q = 7$, it only has $5dB$ gain compared with the original FH/BFSK signal, as shown in Fig. 3.40.

3.6 Conclusions

In this chapter, properties of two new sparse sampling schemes, i.e., coprime sampling and nested sampling are investigated, such as rate distortion function, since sparse sampling can cause possible distortion because less number of samples are used. It is showed that with these two sparse sampling algorithms, the data rate is proved to be much less than that without sparse sampling at a given distortion. This is because with sparser sampling, less number of bits is required to represent the information. We also show that with the same sampling pairs, the rate of nested sampling is less than that of coprime sampling at the same distortion.

The procedures of using nested sampling and coprime sampling these two sparse sampling structures to estimate the QPSK signal's autocorrelation and power

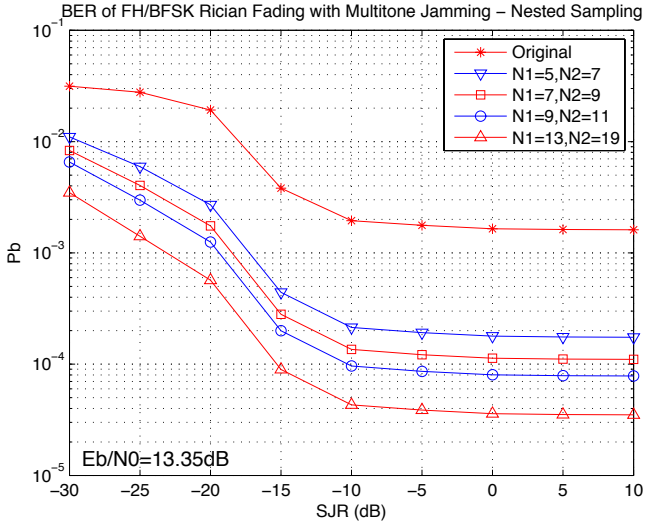

Figure 3.39. BER of FH/BFSK Rician Fading with Multitone Jamming - Nested Sampling, with Rician fading factor $K = 10dB$, total number of jammings q=100, frequency hops $N_S = 1000$, and $E_b/N_0 = 13.35dB$..

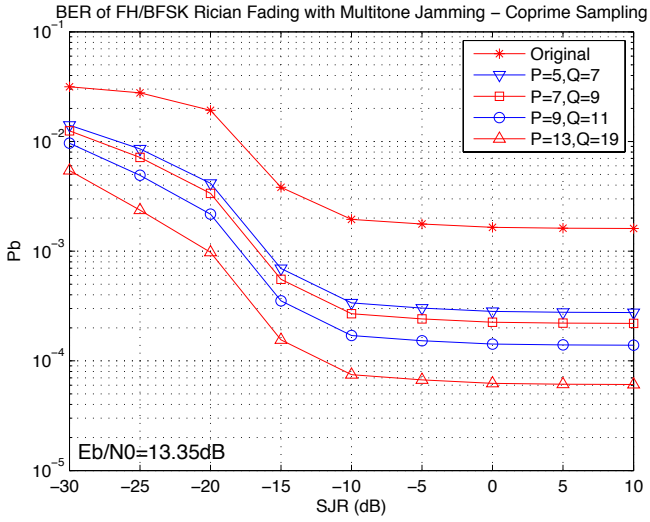

Figure 3.40. BER of FH/BFSK Rician Fading with Multitone Jamming - Coprime Sampling, with Rician fading factor $K = 10dB$, total number of jammings q=100, frequency hops $N_S = 1000$, and $E_b/N_0 = 13.35dB$..

spectral density (PSD) is given. The theoretical analysis of how these two sparse sampling methods effect the power spectral density is provided as well. Our simulation results show that with if we choose the sampling spacings larger, the main lobe of PSD obtained from these two sampling will be much narrower than the original QPSK signal. Besides the smaller average rate, the increased spectrum efficiency is a new advantage of these two sparse sampling algorithms.

A secure transmission scheme for Big Data based on coprime sampling and nested sampling is also provided. With nested sampling and coprime sampling, Big Data could also achieve higher PSD for BFSK signal. When the sampling spacing pairs bigger enough, the spectrum of BFSK signal performs like frequency hopping. This property has great advantage in the security of Big Data collection and transmission using FH/BFSK, as it could achieve low error probability in Rician fading channels. This proves that both nested sampling and coprime sampling could be used in Big Data transmission to resist interference, while guaranteeing the transmission performance.

Chapter 4

A Hybrid Approach of Sparse Sampling

4.1 Introduction

Nested sampling and compressive sensing sense information from two different aspects. Nested sampling [123] is an non-uniform sampling, using two different samplers in each period. Although the signal is sampled sparsely and nonuniformly, the autocorrelation of signal could be estimated at all lags. Therefore, although the samples can be arbitrarily sparse, it keeps the signal's statistical information [124]. While in compressive sensing, the signal, which are sparse in some basis, is sampled using much less number of measurements. The signal could be reconstructed based on the compressed measurements using optimization.

Based on their unique advantages, in this chapter, a hybrid approach of nested sampling and compressive sensing is proposed, which is efficient to represent huge amount of data, especially in Big Data, while keeps the signal's statistical information. For some application, which depends on the difference co-array, or autocorrelation, like Direction-of-arrival (DOA) estimation and beamforming, for the large amount of data acquired, it is not necessary to keep all data. As long as some useful information is kept, such as the signal's statistical information, the requirement is satisfied. In this case, nested sampling could first be applied, with less number of samples selected, to reduce the amount of data significantly while keeping the required information. After that, compressive sensing could be used to further reduce the amount of data in storage or transmission. This hybrid approach has great advantage in the application of Big Data nowadays.

Information rate distortion function is a measure as the number of bits per data sample to be stored or transmitted under the constraint of a distortion. In this chapter, a hybrid approach for Big Data analysis and compression based on nested sampling and compressive sensing is proposed. And theoretical rate distortion performance of the proposed hybrid approach is analyzed. We will show that with this hybrid approach, less number of bits is required to represent the sensed information. And larger sampling intervals for nested sampling will result in less samples kept after nested sampling, and at the same rate, there is more distortion. This indicates us that in real application of this hybrid approach, there is a tradeoff of the sensed information and the performance.

The rest of this chapter is organized as follows. The efficient hybrid sparse sampling approach for Big Data is described in Section 4.2. Theoretical derivation of rate distortion performance of the hybrid approach based on nested sampling and compressive sensing is detailed as well. In Section 4.3, experimental and simulation results are provided to show the theoretical rate distortion performance of our proposed hybrid approach. Conclusions are given in Section 4.4.

4.2 Hybrid Sparse Sampling

For some application, for example, some applications depending on the difference co-array, or autocorrelation, like Direction-of-arrival (DOA) estimation and beamforming, for the large amount of data acquired, it is not necessary to keep all data. As long as some useful information is kept, such as the signal's statistical information, the requirement is satisfied. In this case, nested sampling could first be applied, with less number of samples selected, to reduce the amount of data significantly while keeping the required information. After that, compressive sensing

could be used to further reduce the amount of data in storage or transmission. This hybrid approach is shown in Fig. 4.1.

The original samples are denoted as $x_i, i = 1, \cdots, L$, where L is the total number of samples. Assume that all original information from L samples is $X^L = [x_1, x_2, \cdots, x_L]$, the selected information after nested sampling (NS) can be represented as

$$X^{L^{'}} = NS(X^L) \tag{4.1}$$

where $NS(\cdot)$ denotes nested sampling.

And less number of samples $L^{'} < L$ will be selected based on nested sampling,

$$X^{L^{'}} = [x_1, x_2, \cdots, x_{L'}] = [x_1, \cdots, x_{N_1}, x_{(N_1+1)}, \cdots, x_{N_2(N_1+1)}, \cdots] \tag{4.2}$$

The length $L^{'}$ could be calculated as,

$$L^{'} = \begin{cases} \lfloor \frac{(N_1+N_2)L}{(N_1+1)N_2} \rfloor + Y, & \text{if } Y \leq (N_1+1) \\ \lfloor \frac{(N_1+N_2)L}{(N_1+1)N_2} \rfloor + \lfloor \frac{Y}{N_1+1} \rfloor + N_1, & \text{if } Y > (N_1+1) \end{cases} \tag{4.3}$$

where $Y = L \pmod{(N_1+1)N_2}$, with $\pmod{}$ the modulo operation to find the remainder of division, and $\lfloor \cdot \rfloor$ is the floor function, i.e., the greatest integer function or integer value.

For example, if the original length of source signal is $L = 10008$, with $N_1 = 3$, and $N_2 = 5$, after nested sampling, only $L^{'}$ samples will be selected. Here, $Y = 8$, which is $(N_1 + 1) = 4$. So that $\lfloor \frac{Y}{N_1+1} \rfloor = 2$, and $L^{'} = 4000 + 2 + 3 = 4005$. It is obvious that with the increase of sampling intervals N_1 and N_2, samples will be selected more sparsely, and less number of samples will be selected.

For the hybrid approach, the nested sampling selected samples will be further compressed by compressive sensing as shown

$$Y = \Phi X^{L^{'}} \tag{4.4}$$

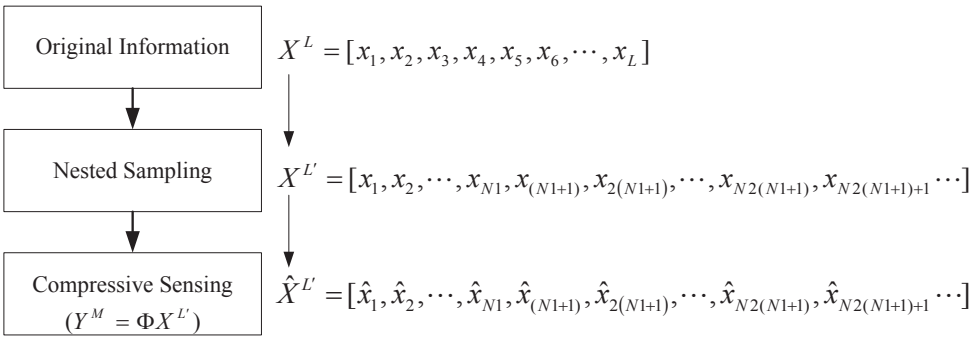

Figure 4.1. Hybrid Approach of Nested Sampling and Compressive Sensing.

where Φ is a $M \times K_{NS}$ measurement matrix, with $M < K_{NS}$, so that the dimension of the observation Y is much less than that of $X^{L'}$.

Based on lots of developed reconstruction algorithms in compressive sensing, although much less observation is obtained, the signal $\hat{X}^{L'}$ could be reconstructed, where

$$\hat{X}^{L'} = [\hat{x}_1, \cdots, \hat{x}_{N_1}, \hat{x}_{(N_1+1)}, \cdots, \hat{x}_{N_2(N_1+1)}, \cdots] \tag{4.5}$$

Information rate distortion function is a measure as the number of bits per symbol to be stored or transmitted under the constraint of a distortion. Our purpose is to construct a distortion function which can determine the bits per symbol because of these two sparse sampling algorithms, nested sampling (NS) and compressive sensing (CS). Sparse sampling can cause possible distortion because less number of samples are used. The original samples are denoted as $x_i, i = 1, \cdots, L$, where L is the total number of samples. Assume that all original information from L samples is $X^L = [x_1, x_2, \cdots, x_L]$, the selected information after hybrid sparse sampling can be represented as [116]

$$\hat{X}^{L'} = S(X^L) \tag{4.6}$$

where $S(\cdot)$ denotes our proposed hybrid approach of sparse sampling as shown in Fig. 4.1 (nested sampling and compressive sampling). $\hat{X}^{L^{'}} = [\hat{x}_1, \hat{x}_2, \cdots, \hat{x}_{L'}]$ and $L^{'} < L$. The distortion associated with the sparse sampling between all original samples and the selected samples is

$$D = Ed(X^L, \hat{X}^{L^{'}}) \tag{4.7}$$

where $d(\cdot)$ is the distortion function.

The expectation in (4.7) is with respect to the probability distribution on X^L. The rate distortion function $R(D)$ is the minimum of data rates R such that (R, D) is in the rate distortion region for a given distortion. From [36], we know that information rate distortion function is defined as

$$R(D) = \min_{Ed(X^L, \hat{X}^{L^{'}}) \leq D} I(X^L; \hat{X}^{L^{'}}) \tag{4.8}$$

where $I(X^L; \hat{X}^{L^{'}})$ is the mutual information between X^L and $\hat{X}^{L^{'}}$.

Since Gaussian assumption is a classical modeling assumption heavily used in areas such as signal processing and communication system [41], from [36], the rate distortion function for a single Gaussian source $N(0, \sigma^2)$ with squared error distortion is

$$R(D) = \begin{cases} \frac{1}{2} \log \frac{\sigma^2}{D} & 0 \leq D \leq \sigma^2, \\ 0 & D > \sigma^2 \end{cases} \tag{4.9}$$

For L independent zero-mean Gaussian sources $x_1, \cdots, x_L$ with variance $\sigma_1^2, \sigma_2^2, \cdots, \sigma_L^2$, the rate distortion performance with squared-error distortion is given by [36] [114][112][137]

$$R(D) = \sum_{i=1}^{L} \frac{1}{2} \log \frac{\sigma_i^2}{D_i} \tag{4.10}$$

where

$$D_i = \begin{cases} \lambda & \text{if } \lambda < \sigma_i^2, \\ \sigma_i^2 & \text{if } \lambda \geq \sigma_i^2 \end{cases} \tag{4.11}$$

where λ is chosen so that $\sum_i^L D_i = D$, and $D_i = E(x_i - \hat{x}_i)^2$. This gives rise to a kind of reverse waterfilling. We choose a constant λ and only describe those random variables with variance greater than λ, and no bits are used to describe random variables with variance less than λ.

The rate distortion performance associated with the nested sampling between all original information and sensed information is detailed in [137].

And the information rate distortion function after the hybrid sparse sensing process shown in Fig. 4.1 could be calculated as [36]

$$R(D) = \min_{Ed(X^L, \hat{X}^{L'}) \leq D} I(X^L; \hat{X}^{L'}) \tag{4.12}$$

where $I(X^L; \hat{X}^{L'})$ is the mutual information between X^L and $\hat{X}^{L'}$.

$$\begin{aligned}
I(X^L; \hat{X}^{L'}) &= H(X^L) - H(X^L|\hat{X}^{L'}) \\
&= H(X^L) + H(\hat{X}^{L'}) - H(X^L, \hat{X}^{L'}) \\
&= H(X^L) + H(\hat{X}^{L'}) - H(X^L|X^{L'}) + H(X^L|X^{L'}) \\
&- H(\hat{X}^{L'}|X^{L'}) + H(\hat{X}^{L'}|X^{L'}) - H(X^L, \hat{X}^{L'}) \\
&\overset{(k)}{=} I(X^L; X^{L'}) + I(X^{L'}; \hat{X}^{L'}) + H(X^L|X^{L'}) + H(\hat{X}^{L'}|X^{L'}) - H(X^L, \hat{X}^{L'}) \\
&= I(X^L; X^{L'}) + I(X^{L'}; \hat{X}^{L'}) + H(X^L|X^{L'}) \\
&+ H(\hat{X}^{L'}|X^{L'}) - H(X^L, \hat{X}^{L'}|X^{L'}) \\
&+ H(X^L, \hat{X}^{L'}|X^{L'}) - H(X^L, \hat{X}^{L'}) \\
&\overset{(l)}{=} I(X^L; X^{L'}) + I(X^{L'}; \hat{X}^{L'}) - (I(X^L, \hat{X}^{L'}; X^{L'}) - I(X^L; \hat{X}^{L'}|X^{L'}))
\end{aligned} \tag{4.13}$$

where equality (k) follows from the fact $I(X^L; X^{L'}) = H(X^L) - H(X^L|X^{L'})$, and $I(X^{L'}; \hat{X}^{L'}) = H(\hat{X}^{L'}) - H(\hat{X}^{L'}|X^{L'})$.

Equality (l) follows from the fact that

$$I(X^L;\hat{X}^{L'}|X^{L'}) = H(X^L|X^{L'}) + H(\hat{X}^{L'}|X^{L'}) - H(X^L,\hat{X}^{L'}|X^{L'}) \tag{4.14}$$

and

$$I(X^L,\hat{X}^{L'};X^{L'}) = H(X^L,\hat{X}^{L'}) - H(X^L,\hat{X}^{L'}|X^{L'}) \tag{4.15}$$

Therefore, the rate distortion function could be represented as

$$\begin{aligned}
R(D) &= \min_{Ed(X^L,\hat{X}^{L'})\leq D} I(X^L;\hat{X}^{L'}) \\
&= \min_{Ed(X^L,\hat{X}^{L'})\leq D} I(X^L;X^{L'}) + I(X^{L'};\hat{X}^{L'}) - (I(X^L,\hat{X}^{L'};X^{L'}) - I(X^L;\hat{X}^{L'}|X^{L'})) \\
&= \min_{D_{NS}} I(X^L;X^{L'}) + \min_{D_{CS}} I(X^{L'};\hat{X}^{L'}) - (I(X^L,\hat{X}^{L'};X^{L'}) - I(X^L;\hat{X}^{L'}|X^{L'})) \\
&= R(D_{NS}) + R(D_{CS}) - (I(X^L,\hat{X}^{L'};X^{L'}) - I(X^L;\hat{X}^{L'}|X^{L'}))
\end{aligned} \tag{4.16}$$

where $R(D_{NS})$ is the rate distortion function for nested sampling, which has been analyzed by authors in [137], and $R(D_{CS})$ is the reconstruction rate distortion function for compressive sensing, analyzed by a lot of authors, such as in [114][112][109][145][146][147][24].

For the last part,

$$\begin{aligned}
&I(X^L,\hat{X}^{L'};X^{L'}) - I(X^L;\hat{X}^{L'}|X^{L'}) \\
&\quad = H(X^L,\hat{X}^{L'}) - H(X^L,\hat{X}^{L'}|X^{L'}) - (H(\hat{X}^{L'}|X^{L'}) - H(\hat{X}^{L'}|X^L,X^{L'})) \\
&\quad \overset{(m)}{=} H(\hat{X}^{L'}) + H(X^L|\hat{X}^{L'}) - (H(X^L|X^{L'}) + H(\hat{X}^{L'}|X^L,X^{L'})) \\
&\quad - (H(\hat{X}^{L'}|X^{L'}) - H(\hat{X}^{L'}|X^L,X^{L'})) \\
&\quad = H(\hat{X}^{L'}) - H(\hat{X}^{L'}|X^{L'}) + H(X^L|\hat{X}^{L'}) - H(X^L|X^{L'}) \\
&\quad \overset{(n)}{\geq} 0
\end{aligned} \tag{4.17}$$

where equality (m) follows from the fact $H(X^L,\hat{X}^{L'}) = H(\hat{X}^{L'}) + H(X^L|\hat{X}^{L'})$, and $H(X^L,\hat{X}^{L'}|X^{L'}) = H(X^L|X^{L'}) + H(\hat{X}^{L'}|X^L,X^{L'})$.

Inequality (n) follows from the fact that condition reduces entropy, i.e., $H(\hat{X}^{L'}) \geq H(\hat{X}^{L'}|X^{L'})$, and since the whole process forms a Markov chain with $X^L \longrightarrow$

$X^{L^{'}} \longrightarrow \hat{X}^{L^{'}}$, by the data-processing inequality, $I(X^L; X^{L^{'}}) \geq I(X^L; \hat{X}^{L^{'}})$. As we know that $I(X^L; X^{L^{'}}) = H(X^L) - H(X^L|X^{L^{'}})$, and $I(X^L; \hat{X}^{L^{'}}) = H(X^L) - H(X^L|\hat{X}^{L^{'}})$, so that $H(X^L|\hat{X}^{L^{'}}) \geq H(X^L|X^{L^{'}})$.

From the above analysis, we could conclude that for the hybrid approach bases on nested sampling and compressive sensing, the rate distortion

$$R(D) \leq R(D_{NS}) + R(D_{CS}) \tag{4.18}$$

which indicates that it would need less bits to represent the sensed information with the hybrid sparse sensing.

Therefore, we can draw the following conclusion as a theorem:

Theorem 11. *(Rate distortion for the hybrid approach based on Nested Sampling (NS) and Compressive Sensing (CS)) Let $x_i \sim N(0, \sigma_i^2)$, $i = 1, 2, \cdots, L$, be independent Gaussian random variables, and under squared error distortion, it would need less number of bits to represent the sensed information for the hybrid approach based on nested sampling (NS) and compressive sensing (CS). The rate distortion function for the hybrid approach is given by*

$$R(D) \geq R(D_{NS}) + R(D_{CS}) - M(X^L, X^{L^{'}}, \hat{X}^{L^{'}}) \tag{4.19}$$

with

$$M(X^L, X^{L^{'}}, \hat{X}^{L^{'}}) = I(X^L, \hat{X}^{L^{'}}; X^{L^{'}}) - I(X^L; \hat{X}^{L^{'}}|X^{L^{'}}) \geq 0 \tag{4.20}$$

where $R(D_{NS})$ is the rate distortion function for nested sampling, and $R(D_{CS})$ is the reconstruction rate distortion function for compressive sensing.

4.3 Simulation and Experimental Results

According to a new study by GSMA (GSM Association) Intelligence [142], the number of 4G-LTE (Long Term Evolution) connections worldwide is forecast to

pass one billion by 2017. Mobile data software company Mobidia found that LTE users used far more data than those using 3G. For example, in South Korea, LTE smartphone users consumed on average almost 2.2GB of data per month [142]. With more and more LTE users, covering more than 90% of the population in the United States, LTE has been a Big Data consumer with ample data. With 5G wireless networks [143] supporting a 10 Gb/s individual user experience capable of extremely low latency and response times, which will be deployed between 2020 and 2030, the data consumption will be extremely high at that time. How to deal with these Big Data in telecommunications? A number of use cases were provided by the whitepaper [144].

As we know, QPSK (Quadrature Phase Shift Keying) is the most popular modulation in 4G, firstly, QPSK signal in Big Data will be studied. QPSK modulated signal could be expressed as [118]

$$s_{QPSK}(t) = \sqrt{\frac{2E_s}{T_s}} cos[2\pi f_c t + (i-1)\frac{\pi}{2}] \tag{4.21}$$

where T_s is the symbol duration, E_s is the energy-per-symbol, and f_c is the carrier frequency.

Fig.4.2 gives the rate distortion performance curve after the hybrid approach of nested sampling and compressive sensing for QPSK signal with $E_s = 1$, $f_c = 400Hz$, and $T_s = 1/f_c$ as the source. The total length of the original QPSK signal is $L = 1800$. The $x-$ axis is the distortion represented with MSE (Mean Squared Error). And $y-$ axis is the corresponding rate, the bits required to represent each QPSK symbol for each sample in this whole hybrid sampling process. Two different sampling intervals for nested sampling are shown, with $N_1 = 2, N_2 = 3$, and $N_1 = 3, N_2 = 5$. With $N_1 = 2, N_2 = 3$, in each period of $(N_1 + 1)N_2 = 9$, only $N_1 + N_2 = 5$ samples are kept, therefore, for the original data length $L = 1800$, after nested sampling, the

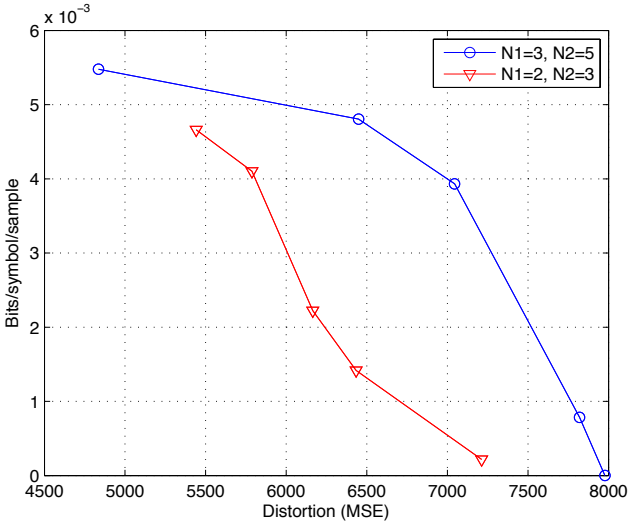

Figure 4.2. Hybrid Sparse Sensing-QPSK signal.

data length $L_1^{'} = 1000$. And a more sparse sampling intervals with $N_1 = 3, N_2 = 5$, $N_1 + N_2 = 8$ samples are selected for each period of $(N_1 + 1)N_2 = 20$, therefore, nested sampling keeps $L_2^{'} = 720$ samples for the whole original signal. It is obvious that with more sparse sampling, i.e., $N_1 = 3, N_2 = 5$, less samples will be kept after nested sampling. Then, with compressive sensing, at the same data rate, there will be more distortion. This is because larger sampling intervals for nested sampling will lose more information, which cause higher distortion.

Now, let's look at how the real rate distortion of hybrid sparse sensing approach performs for real-world data. The experimental settings for real-world sense-through-foliage are described in [116][141]. The foliage measured include late summer, fall and winter foliage. Late summer foliage involved foliage with decreased water content, because of the limited rainfall. While late fall and winter measurements involved largely defoliated but dense forest, which provided a rich scattering environment. Its also a time-varying environment, because of wind or different temperatures in

dense forest. The target is a trihedral reflector located 300 feet away. The foliage experiment was constructed on a seven-ton man lift, with a total lifting capacity of 450 kg. The principle pieces of equipment secured on the lift are: Barth pulse generator (Barth Electronics, Inc. model 732 GL), Tektronix model 7704B oscilloscope, HP signal generator, two antennas, dual-antenna mounting stand, rack system, IBM laptop, custom RF switch, power supply and weather shield. A Barth pulse source generator was used to provide pulses of less than 50 picoseconds (ps) rise time, with amplitude from 150 V to greater than 2 KV into any load impedance through a 50 ohm coaxial line. The generator is capable of producing pulses with a minimum width of 750 ps and a maximum of 1 microsecond. This output pulse width is determined by charge line length for rectangular pulses, or by capacitors for 1/e decay pulses. Each sample is spaced at 50 picosecond interval, and 16,000 samples were collected for each collection for a total time duration of 0.8 microseconds at a rate of approximately 20 Hz. The pulse source was operated at low amplitude and 35 pulses reflected signal were averaged for each collection [116][141]. For illustration purposes, in Fig. 4.3, we plot part of the UWB radar received echo based on one experiment with length $L = 1800$.

Following the procedures described in Fig. 4.1, nested sampling is performed to the original UWB data X^L. Fig. 4.4 provides the data $X^{L'}$ after nested sampling with $N_1 = 2, N_2 = 3$. For the original data length $L = 1800$, after nested sampling with $N_1 = 2, N_2 = 3$, the data length $L_1' = 1000$. And a more sparse sampling intervals with $N_1 = 3, N_2 = 5$ is shown in Fig. 4.5. For this case, nested sampling keeps $L_2' = 720$ samples for the whole original signal.

From Fig. 4.1, after nested sampling, the selected signals $X^{L'}$ is compressed using compressive sensing, and reconstructed as $\hat{X}^{L'}$ in our hybrid approach of sparse sensing system. As we described, a basis matrix Ψ could be found to represent

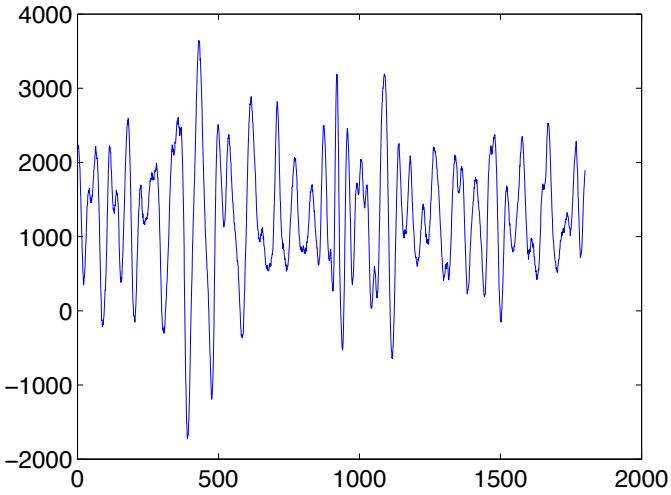

Figure 4.3. Original Sense-through-foliage signal.

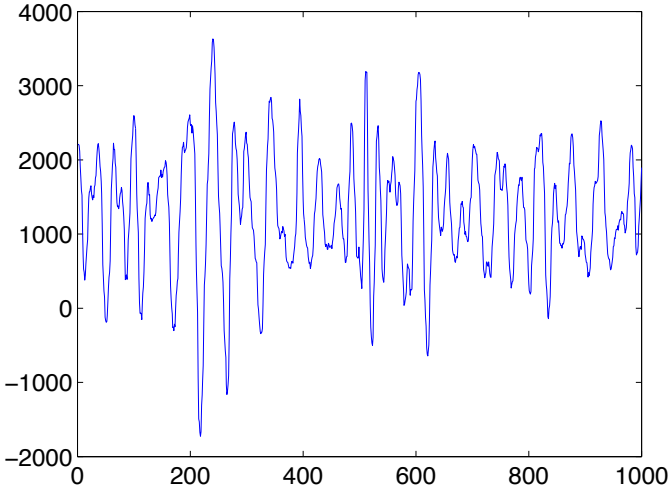

Figure 4.4. Data acquired after nested sampling with $N_1 = 2, N_2 = 3$.

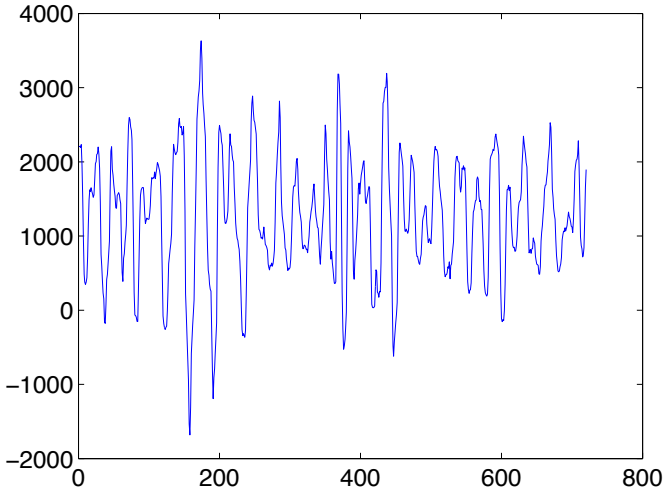

Figure 4.5. Data acquired after nested sampling with $N_1 = 3, N_2 = 5$.

$\theta = (\Psi)^{-1}x$, where θ has only a few nonzero entries as shown in Fig. 4.6. It is clear that compressive sensing could be used in this case. And random gaussian matrix is used as measurement matrix.

The reconstructed signal based on basis pursuit (BP) of the CS compressed signal is shown in Fig. 4.7.

Rate distortion performance for the hybrid sparse sensing approach of real-world sense-through-foliage data is given in Fig. 4.8. Here $x-$ axis is the distortion represented with MSE (Mean Squared Error), and $y-$ axis gives the corresponding rate, bits per sample, to represent how many bits is required for each sample in this whole hybrid sampling process. Similarly to ideal QPSK signal performance in Fig. 4.2, two different sampling intervals for nested sampling are shown as well, with $N_1 = 2, N_2 = 3$, and $N_1 = 3, N_2 = 5$. With $N_1 = 3, N_2 = 5$, less samples will be kept after nested sampling than that of $N_1 = 2, N_2 = 3$, and at the same rate, there is more distortion. Larger sampling intervals for nested sampling loses more

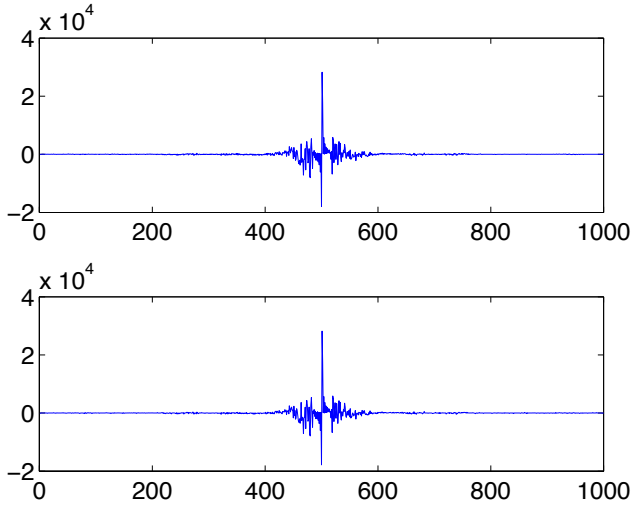

Figure 4.6. Sparse signal for CS.

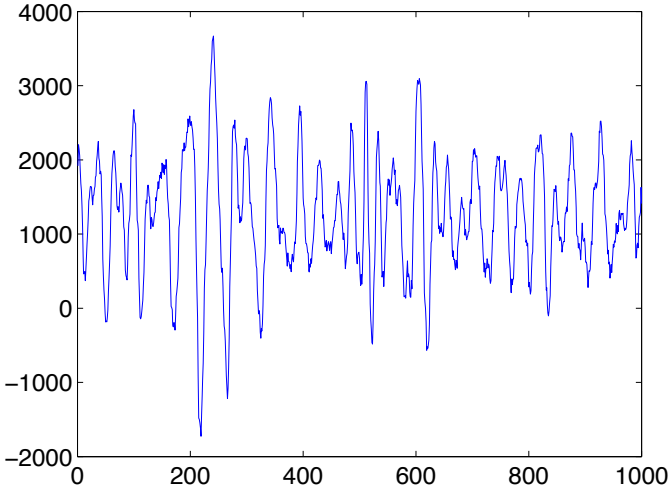

Figure 4.7. CS reconstructed signal.

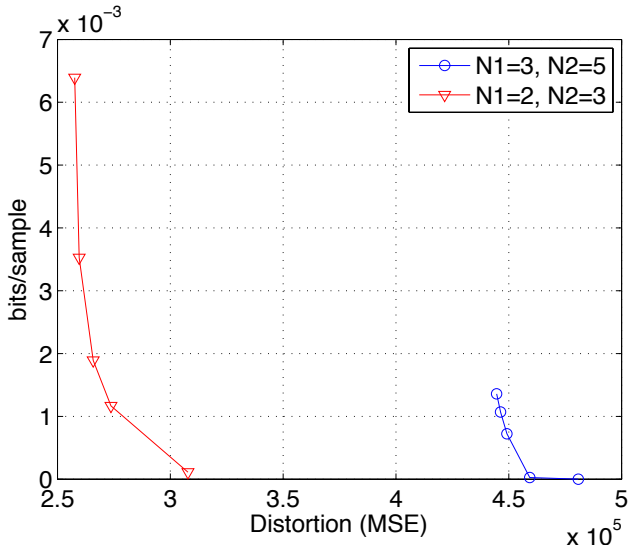

Figure 4.8. Hybrid Sparse Sensing-Sense-through-foliage signal.

information, which cause higher distortion. From Fig. 4.8, we could also notice that for $N_1 = 3, N_2 = 5$, the real-world data sensing could not achieve as low distortion as $N_1 = 2, N_2 = 3$, and the same for rate. This is because a lot of useful information are lost because of ambiguous nested sampling. This indicates us that in real application of nested sampling, there is a tradeoff of sampling intervals (the number of sensed data, which effects data rates) and the performance (which effects distortion).

4.4 Conclusions

In an era of Big Data, resources are intertwined in complex ways with data resources. It is critical to determine how many bits/symbols should be kept in Big Data collection and measurement. For some application, which depends on the difference co-array, or autocorrelation, like Direction-of-arrival (DOA) estimation and beamforming, it is not necessary to keep all data. In this chapter, a hybrid approach was proposed, which combines nested sampling and compressive sensing

to reduce the number of symbols, and rate distortion function is used as a criteria to determine how many bits should be used to represent the symbols during this process. We showed that with this hybrid approach, less number of bits is required to represent the sensed information. This hybrid approach has great advantage in the application of Big Data nowadays. And larger sampling intervals for nested sampling will result in less samples kept after nested sampling, and at the same rate, there will be more distortion. This indicates us that in real application of this hybrid approach, there is a tradeoff of the number of sensed information and the performance.

In future research, we may not only analyze how the rate distortion performance will change with different nested sampling spacing pairs N_1, N_2 (one special case is that $N_1 + N_2$ keeps the same, with different N_1, N_2 combination), but also analyze the real application performance of the proposed hybrid approach, for example, in some real estimation of DOA, to see how the estimation error rate of DOA changes with different nested sampling intervals.

Chapter 5

Bandwidth Allocation

5.1 Introduction

Smartphones are everywhere these days, and they have become an indispensable part of our daily lives. As of July 18, 2013, 90% of global handset sales are attributed to the purchase of Android and iPhone smartphones[148]. The smartphone has many more advanced computing capability and connectivity than the traditional cellphone. Modern smartphones include portable media players, low-end compact digital cameras, pocket video cameras, GPS navigation units, high-resolution touchscreens and web browsers, which allow its owners to e-mail, surf the web, play music and games, and perform a variety of other functions. Wi-Fi and mobile broadband provide high-speed data access for smartphone.

According to a new study by GSMA (GSM Association) Intelligence [142], the number of 4G-LTE (Long Term Evolution) connections worldwide is forecast to pass one billion by 2017. Mobile data software company Mobidia found that LTE users used far more data than those using 3G. For example, in South Korea, LTE smartphone users consumed on average almost 2.2GB of data per month [142]. With more and more LTE users, covering more than 90% of the population in the United States, LTE has been a Big Data consumer with ample data. With 5G wireless networks [143] supporting a 10 Gb/s individual user experience capable of extremely low latency and response times, which will be deployed between 2020 and 2030, the data consumption will be extremely high at that time. How to deal with these Big Data in telecommunications? A bandwidth allocation method based on smartphone

users' personality traits and channel condition is studied in a unified mathematical framework in this chapter.

Smartphone data might be a function of personality, as the smartphone supports interpersonal interaction. Studying the psychological and social implications of smartphone usage has gained an increased importance. Some researchers have investigated the demographic characteristics of smartphone users. Different from traditional self-reported personality testing, which may be not accurate in the measurement of traits if people represent themselves falsely in terms of their personality, since smartphones are programmable, the development of data collection tools to record various behavioral aspects of the users is practical nowadays, which is more accurate and objective.

A new industry–the mobile phone app has been created since the tremendous increase in smartphone ownership, and some demographic data have been collected regarding these types of applications. The relationship between mobile phone usage and user personality has attracted a lot of researchers. Personality traits can predict the patterns of mobile phone use was concluded in [149] by self reported data. For smartphones, this relationship could be investigated based on more accurate data collected by programmable tools. The relationship between five main factors of personality and addiction to SMS in high school students was studied in [150], and neuroticism and addiction to text messaging were found positively related. Concerning changes in consumer privacy in relation to smartphones, the authors in [151] discussed research that found correlations between styles of phone usage and personality traits of users. Papers [152] and [153] studied the impact of the "Big Five" personality traits on smartphone ownership and use. Extraverted individuals were found more likely to own a smartphone in [152]. Extraverted individuals were noticed to report greater importance on gaming applications in [153]. In [154], the

authors proposed to leverage mobile phone usage-derived variables to effectively and automatically infer the users' personality traits as defined by the big five model.

These researches provide a new opportunity of understanding the impact of context on user behavior and studying individual differences such as user personality. These studies also facilitate further research on the usage of personality traits for personalizing services on smartphone users. The Bayesian Network (BN) could be drawn based on the relationship concluded in [151][152][153]. Personality traits of users could be estimated, using inference from Bayesian Network (BN). Personalizing bandwidth allocation could be done based on smartphone users' personality traits, resulting in a smart and efficient usage of the limited bandwidth.

A bandwidth allocation method based on smartphone users' personality traits and channel condition is studied in a unified mathematical framework in this chapter. Based on the relationship between user behavioral characteristics extracted from rich smartphone data and self-reported Big-Five personality traits, the Bayesian Network could be drawn. Further, the service provider could estimate each users probability of having each personality trait using diagnostic inference, and then based on predictive inference to calculate each user's usage of bandwidth. Personalizing bandwidth allocation could be done based on smartphone users' personality traits, resulting in a smart and efficient usage of the limited bandwidth. For our proposed smart bandwidth allocation scheme, both the outage capacity and the outage probability are studied in fading channel. The service provider could adjust the bandwidth allocated further on account of the real channel condition, which makes our proposed algorithm more robust.

The remainder of this chapter is organized as follows. In Section 5.2, we give a brief introduction of Bayesian Network. In Section 5.3, personality traits are analyzed based on smartphone usage and personality psychology. Bayesian network

modeling based on smartphone users' personality traits is given in Section 5.4. The bandwidth allocation scheme based on Bayesian network inference of smartphone usage is detailed in Section 5.5, with diagnostic inference, predictive inference, and approximate inference. The outage throughput capacity in fading channel is analyzed as well in Section 5.6, which helps the service provider to adjust the allocated bandwidth based on real channel condition. Numerical results are shown in Section 5.7, and conclusions are given in Section 5.8.

5.2 Bayesian Network

Bayesian Network (BN) [155] [156] [157] [158], is a directed acyclic graph (DAG) which represents a set of random variables and their conditional dependencies. BNs are both mathematically rigorous and intuitively understandable. An effective representation and computation of the joint probability distribution (JPD) over a set of random variables could be enabled by BNs. Using local conditional probability tables (CPTs), the joint distribution of a collection of variables can be determined uniquely.

Each node in directed acyclic graphs represents a random variable in the Bayesian sense, which may be discrete or continuous. Pairs of nodes are connected by a set of directed links or arrows. These edges represent direct dependence among the variables. And nodes not connected represent variables which are conditionally independent of each other. Each node has a probability function that takes as input a particular set of values for the node's parent variables and gives the probability of the variable represented by the node.

In a BN, an edge from node N_i to node N_j represents a statistical dependence between the corresponding random variables X_i and X_j. Thus, the arrow indicates that a value taken by random variable X_j depends on the value taken by variable

X_i, or variable X_i "direct influences" X_j, which means that causes should be parents of effects. Therefore, node N_i is referred to as a parent of N_j and, similarly, N_j is referred to as the child of N_i. An extension of these genealogical terms is often used to define the sets of "descendants" nodes − the set of nodes from which the node can be reached on a direct path, or "ancestor" nodes − the set of nodes that can be reached on a direct path from the node[158].

A Bayesian network B is an annotated directed acyclic graph that represents a JPD over a set of random variables V. The BN is defined by a pair G and Θ, i.e., $B = \langle G, \Theta \rangle$, where G is the DAG whose nodes $X_1, X_2, \cdots, X_n$ represents random variables, and whose edges represent the direct dependencies between these variables. The graph G encodes independence assumptions, by which each random variable X_i is independent of its nondescendents given its parents in G. The second component Θ represents the set of parameters of the network. This set contains the parameter $\theta_{x_i|\pi_i} = P_B(x_i|\pi_i)$ for each realization x_i of X_i conditioned on π_i, which is the set of parents of X_i in G. Therefore, B defines a unique JPD over V, as

$$P_B(X_1, X_2, \cdots, X_n) = \prod_{i=1}^{n} P_B(X_i|\pi_i) = \prod_{i=1}^{n} \theta_{X_i|\pi_i} \tag{5.1}$$

Given a BN specified the JPD, all possible inference queries could be evaluated by marginalization, that is, summing out over "irrelevant" variables. Two types of inference support are often considered: predictive support for node X_i, based on evidence nodes connected to X_i through its parent nodes, and diagnostic support for node X_i, based on evidence nodes connected to X_i through its children nodes. In this chapter, both kinds of inferences will be estimated based on the bayesian network of smartphone usage and personality traits.

5.3 Personality Traits in Smartphone Usage

The relationship between user behavioral characteristics extracted from rich smartphone data and Big-Five personality traits (Extraversion, Agreeableness, Conscientiousness, Neuroticism and Openness to Experience) was investigated in [151][152][153].

From personality psychology[170], we know that there are two possible explanations for a person's behavior:

(1). Behavior is a function of personality traits, $B = f(P)$.

(2). Behavior is a function of situational forces, $B = f(S)$.

where B is behavior, P represents personality traits, S stands for situational forces, and $f(\cdot)$ represents a function.

From these two points of view, it is obvious that both personality and situations interact to produce behavior, or [158]

$$B = f(P \times S) \tag{5.2}$$

This formula suggests that behavior is a function of the interaction between personality traits and situation forces. For the smartphone usage behavior, for sure, it depends on the user's personality trait and situation force.

Trait theories of personality [151] [152] [153] [158] offer a collection of viewpoints about human nature. Trait theories have three important assumptions about personality traits, which form the basic foundation for trait psychology. These three important assumptions are:

(1). meaningful individual differences: Every personality, no matter how complex or unusual, is the product of a particular combination of a few basic and primary traits.

(2). stability or consistency over time: There is a degree of consistency or stability in personality over time. If someone is highly extraverted during one period of observation, he or she will be highly extraverted in the future. The point that many

broad-based personality traits show considerable stability over time has been supported by a large number of research studies.

(3). consistency across situations: Traits will also exhibit some consistency across situations. Trait psychologists believe that people's personalities show consistency from situation to situation.

In the past two decades, the Big-Five personality traits has received the most attention and support from personality researchers. This five-factor model is a hierarchical model of personality traits that represent personality at the broadest level of abstraction [159]. It consists of five board traits, namely extraversion, agreeableness, conscientiousness, neuroticism, and openness to experience [160]. These factors summarize several more specific traits and are believed to capture most of the individual differences in human personality [159].

Extraverts are sociable, gregarious and ambitious, who are optimistic and seek out new opportunities and excitement[161]. People score high in extraversion are social, active, outgoing, and place a high value on close and warm interpersonal relationships[162]. Agreeable people are sympathetic, good natured, cooperative and forgiving[161]. The authors [167] [166] discovered that more agreeable individuals spend more time on calls and that disagreeable individuals with lower self-esteem spent more time using instant messaging and games. Conscientious people are strong-willed, deliberate, and reliable, who actively plan, organize and carry out tasks[161]. The hallmark of their personality is self-control, reflected in a need for achievement, order, and persistence[163]. Conscientious people look for ways in which the use of technology would allow them to be more efficient and perform at a higher level at work[165]. Neurotic people are anxious, self-conscious and paranoid[168]. People score high in neuroticism tend to be fearful, sad, embarrassed, distrustful, and have difficulty managing stress[161]. Neurotic individuals are found to spend more time

text messaging and reported stronger mobile phone additive tendencies[167]. People high in openness have flexibility of thought and tolerance of new ideas, and they actively seek out new and varied experiences and value change[164]. They are associated with training proficiency and engaging in learning experiences[165], which make them more likely to hold positive attitudes and cognitions toward accepting job-related technology[168].

Based on the relationship concluded in [151][152][153], "Youtube" was found to be more likely to be used by extraverts and non-conscientious users. The "Video/Audio/Music" apps was found to be more likely to be used by users who score higher on openness and low on conscientiousness. The "SMS" apps was found to be more likely to be used by disagreeable users who are conscientious and less open. Further, users scoring high on extraversion and emotional stability and low on agreeableness and openness were more likely to receive SMS. The "Mail" app was found to be more likely to be used by users scoring low on agreeableness and high on conscientiousness. The "Internet" app was found to be more likely to be used by users scoring low on agreeableness and high on conscientiousness. Extroverts had heavy usage of calendars and office apps and their smartphone habits were with less games and Internet functions. Extroverts spent more time on calls, which were greater in number than those received by introverts. They also had more unique contacts in those calls. Less agreeable people were heavier users of office, Internet, Video/Audio/Music, mail, calendar and SMS apps. Agreeable people were seen to have more number of voice calls. Less conscientious people were less likely to use multimedia and YouTube applications. The more conscientious users sent shorter SMS messages, and longer SMS messages were generally made by more emotionally stable users. Emotional stable people were found to use less of office and calendar apps. The length of SMS were higher for emotional stable users. It was seen that the duration of incoming calls

were much small for emotional stable users. Curious people had less usage of office, calendar and SMS apps.

Based on the analysis above, it is thus clear that the personality traits based on smartphone usage is meaningful and consistent over time and situation, which could help us to predict the future smartphone usage.

Psychologists also notice that individual differences in personality emerge very early in life. These individual differences in personality tend to be moderately stable over time, so that people who are high on a particular trait tend to remain high on that dimension. Trait psychologists conclude that people will be relatively consistent over time in their behavior because of the various traits they possess. And personality trait scores refer primarily to average tendencies in behavior. A score on a trait measure refers to how a person is likely to behave, on average, over a number of occasions and situations. For example, based on our analysis, from a person's high score on a measure of trait openness and extraversion, the service provider could confidently predict that a person is more likely to use more bandwidth in the next few billing cycle than a person with a lower score on openness and extraversion.

Trait psychologists are also interested in the accuracy of measurement. Trait psychologists make efforts to improve the measurement of traits, particularly through self-report questionnaire measures. Psychologists who devise questionnaires work hard at making them less susceptible to lying, faking, and careless responding. While with the advent of smartphones, a new view to investigate smartphone users' personality traits could be done based on more accurate data collected by programmable tools.

These measurements and prediction have stimulated the researchers to apply them to a lot of situations in which personality might make a difference. In this chapter, as we will discuss, the service provider could use personality traits measure-

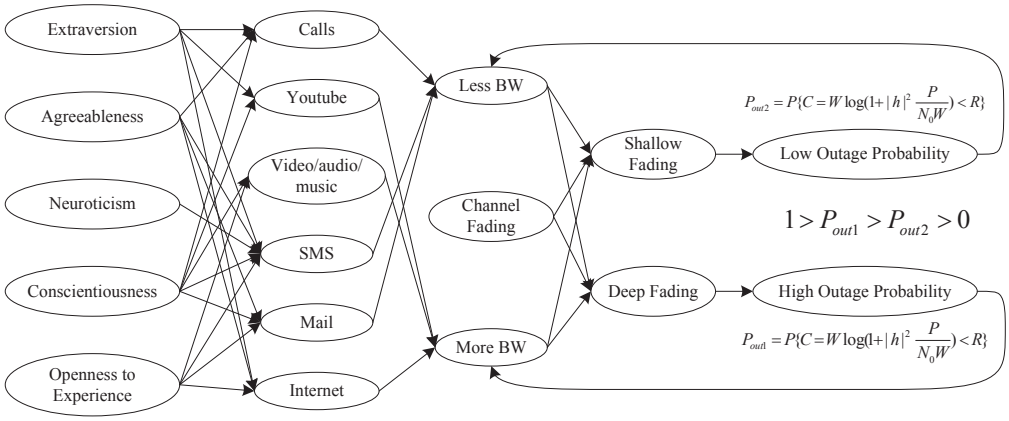

Figure 5.1. Bandwidth Allocation Scheme based on Personality Traits and Channel Condition.

ments and prediction to personalize bandwidth allocation to help smartphone users for a better data service.

Therefore, in this chapter, based on the measures of personality, the usage of smartphone in one billing cycle could be predict in their future usage. We adopt the concluded Big-Five personality traits as inputs to Bayesian network to study the bandwidth allocation problem in a unified mathematical framework.

5.4 Bayesian Network Modeling in Smartphone Usage

The Bayesian Network (BN) could be drawn based on the relationship concluded in [151][152][153] as shown in Fig. 5.1.

Based on the results above [151][152][153], the CPT of node "Calls" is listed in Tables 5.5, and 5.6. For short, we only use the first letter to represent each node, for example, "C" is the abbreviation of "Calls". Similarly, the tables of CPT of "Youtube", "Video/Audio/Music", "SMS", "Mail" and "Internet" could be calculated based on the results in [151] [152] [153]. As the principals are the same, we

only show the CPT of node 'Calls", and we will not give the details of the CPT for all the other application usage for the lack of space in this chapter.

In this chapter, we only use the first three letters to represent each personality trait, for example, "Ext" is the abbreviation of "Extraversion". $Ext = y$ is used to represent "Outgoing", and $Ext = n$ is used to represent "Solitary", as shown in Fig. 5.2. And the other four personality traits are represented in the same way. Similarly, only the first letter is used to represent each application, that is, "C", "Y", "V", "S", "M" and "I" for "Calls", "Youtube", "Video/Audio/Music", "SMS", "Mail" and "Internet" respectively. "LBW" stands for "Less BW". The CPT of "Less BW" is listed in Tables 5.7, 5.8, 5.9, and 5.10. The CPT of "More BW" could be listed in a similar method. Based on the smartphone usage, users' personality could be estimated based on diagnostic inference of Bayesian Network in Fig. 5.1. With users' personality traits estimated, using predictive inference from this BN, the service provider could predict the future smartphone usage of each user in the next billing cycle, so that flexible bandwidth allocation could be predicted. This study facilitates further research on the automated classification and usage of personality traits for personalizing services on smartphones. Personalizing bandwidth allocation could be done based on smartphone users' personality traits, resulting in a smart and efficient usage of the limited bandwidth.

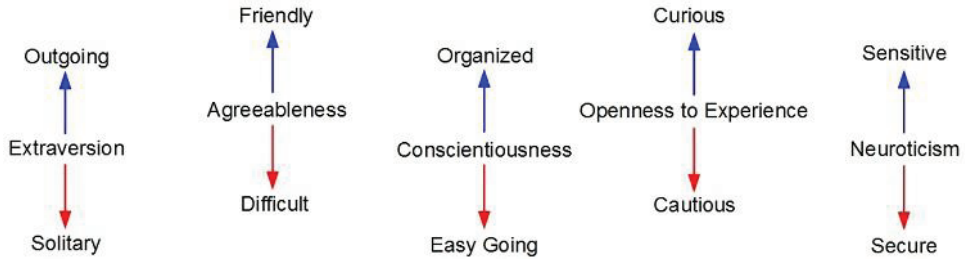

Figure 5.2. Big-Five Personality Traits.

5.5 Bandwidth Allocation: Bayesian Network Inference

In this chapter, both diagnostic inference and predictive inference are calculated based on personality traits and their usage of smartphone apps in Fig. 5.1. Diagnostic inference[156] provides the probability of each personality trait with the data of each user's bandwidth usage, for example, what is the probability of one user's personality traits if this smartphone user occupies above 10GB data per bill cycle? Differently, Predictive inference calculate the probability of the usage of bandwidth based on each node's parent nodes, i.e., for example, if a user score high on extraversion, what is the probability of this user occupies more bandwidth? Based on one user's data usage in one billing cycle, the service provider could calculate this user's probability of having each personality trait using diagnostic inference, and then based on predictive inference to calculate this user's usage of bandwidth in the next billing cycle. We can see that both diagnostic inference and predictive inference could help the service provider to better allocate the smartphone bandwidth based on each user's personality.

Since the structure of a BN implies that the value of a particular node is conditional only on the values of its parent nodes, as shown in equation (5.1), we can simplify its joint probability expressions in our case that the joint probability $P(Ext, Agr, Neu, Con, Ope, C, Y, V, S, M, I, LessBW)$

$$
\begin{aligned}
&P(Ext, Agr, Neu, Con, Ope, C, Y, V, S, M, I, LessBW) \\
&\quad = P(Ext)P(Agr)P(Neu)P(Con)P(Ope) \cdot P(C|Ext, Agr, Neu, Con, Ope) \\
&\quad \cdot P(Y|Ext, Agr, Neu, Con, Ope) \cdot P(V|Ext, Agr, Neu, Con, Ope) \\
&\quad \cdot P(S|Ext, Agr, Neu, Con, Ope) \cdot P(M|Ext, Agr, Neu, Con, Ope) \\
&\quad \cdot P(I|Ext, Agr, Neu, Con, Ope) \cdot P(LessBW|C, Y, V, S, M, I) \qquad (5.3)
\end{aligned}
$$

5.5.1 Diagnostic Inference

First, let's look at diagnostic inference. The probability could be calculated based on BN. Here, we only show one example of given "Less BW", the probability of Extraversion. For other probability with other personality traits, the principal is the same.

$$P(Ext = y|LessBW = y) = \frac{P(Ext = y, LessBW = y)}{P(LessBW = y)} \tag{5.4}$$

where

$$\begin{aligned}
&P(Ext = y, LessBW = y) \\
&\quad = \sum P(Ext = y)P(Agr)P(Neu)P(Con)P(Ope) \\
&\quad \cdot P(Cal|Ext = y, Agr, Neu, Con, Ope) \cdot P(You|Ext = y, Agr, Neu, Con, Ope) \\
&\quad \cdot P(Vid|Ext = y, Agr, Neu, Con, Ope) \cdot P(SMS|Ext = y, Agr, Neu, Con, Ope) \\
&\quad \cdot P(Mai|Ext = y, Agr, Neu, Con, Ope) \cdot P(Int|Ext = y, Agr, Neu, Con, Ope) \\
&\quad \cdot P(LessBW = y|Cal, You, Vid, SMS, Mai, Int)
\end{aligned} \tag{5.5}$$

and

$$\begin{aligned}
&P(LessBW = y) \\
&\quad = \sum P(Ext)P(Agr)P(Neu)P(Con)P(Ope) \\
&\quad \cdot P(Cal|Ext, Agr, Neu, Con, Ope) \cdot P(You|Ext, Agr, Neu, Con, Ope) \\
&\quad \cdot P(Vid|Ext, Agr, Neu, Con, Ope) \cdot P(SMS|Ext, Agr, Neu, Con, Ope) \\
&\quad \cdot P(Mai|Ext, Agr, Neu, Con, Ope) \cdot P(Int|Ext, Agr, Neu, Con, Ope) \\
&\quad \cdot P(LessBW = y|Cal, You, Vid, SMS, Mai, Int)
\end{aligned} \tag{5.6}$$

5.5.2 Predictive Inference

Similarly, the predictive inference could be calculated for each personality trait. We also only show one example of knowing a user's personality of extraversion, the probability of less BW assignment as

$$P(LessBW = y|Ext = y) = \frac{P(Ext = y, LessBW = y)}{P(Ext = y)} \tag{5.7}$$

where $P(Ext = y, LessBW = y)$ could be calculated the same as in formula (5.5).

We assume the original probability for all personality traits is unknown, that is $P(Ext = y) = 0.5$, $P(Agr = y) = 0.5$, $P(Neu = y) = 0.5$, $P(Con = y) = 0.5$, and $P(Ope = y) = 0.5$. Therefore, all the probability in Tables 5.1 and 5.2 could be calculated.

5.5.3 Approximate Inference

We are also interested in sampling applied to the computation of posterior probabilities. Here we consider direct sampling method as described in[158], referred as PRIOR-SAMPLE. Its operation on the network in Fig. 5.1 is illustrated with an ordering $[Ext, Agr, Neu, Con, Ope, C, Y, V, S, M, I, LessBW]$. In this case, PRIOR-SAMPLE returns the event $[y, y, y, y, y, y, y, n, y, y, n, n]$, with values shown in formula (5.8), where sample from $P(Ext) =< 0.5, 0.5 >$, value is y, sample from $P(C|Ext = y, Agr = y, Neu = y, Con = y, Ope = y) =< 0.69, 0.31 >$, value is y, and so on.

$$[0.5, 0.5, 0.5, 0.5, 0.5, 0.69, 0.55, 0.65, 0.72, 0.52, 0.99, 0.8] \tag{5.8}$$

From [158], $S_{PS}(X_1, \cdots, X_n)$ is the probability that a specific event is generated by the PRIOR-SAMPLE algorithm, and each sampling step depends only on the parent values,

$$S_{PS}(X_1, \cdots, X_n) = \prod_{i=1}^{n} P(X_i|\pi(X_i)) \tag{5.9}$$

Suppose N total samples, and the number of times the specific event $X_1, \cdots, X_n$ occurs in the set of samples are $N_{PS}(X_1, \cdots, X_n)$[158].

$$\lim_{N\to\infty} \frac{N_{PS}(X_1, \cdots, X_n)}{N} = S_{PS}(X_1, \cdots, X_n) = P_B(X_1, X_2, \cdots, X_n) \tag{5.10}$$

Therefore, the sampling probability for this event is

$$S_{PS}(y, y, y, y, y, y, y, n, y, y, n, y, n) = 0.5 \times \cdots \times 0.8 = 0.00228578922 \tag{5.11}$$

that is, in the limit of large N, we only expect 0.228578922% of the samples to be of this event.

5.6 Bandwidth Allocation in Fading Channel

As we know that the capacity of wireless channels is also based on channel conditions[169]. We can build a smart bandwidth allocation scheme that adaptively allocates channel bandwidth to each smartphone user depending on what the current channel condition is. Taking advantage of the better channel conditions, less bandwidth could be assigned with performance guarantee.

For a bandwidth-limited system, the capacity is an increasing, concave function of the bandwidth W[169]. And for slow fading channel, the capacity is

$$C = Wlog(1 + |h|^2 \frac{P}{N_0 W}) \tag{5.12}$$

where h is the fading channel gain, and $|h|$ is the amplitude of h. W is the channel bandwidth. P is a given received power, and additive white Gaussian noise power spectral density is $N_0/2$.

Suppose the transmitter encodes data at a rate R bits/s. If the channel gain h is small that $C < R$, then the system is in outage, and the decoding error probability can not be made arbitrarily small. The outage probability is

$$P_{out}(R) = Pr\{Wlog(1 + |h|^2 \frac{P}{N_0 W}) < R\} \tag{5.13}$$

For fading channels, we consider ϵ-outage capacity C_ϵ, which is the largest transmission rate R such that $P_{out}(R)$ is less than ϵ. And solving $P_{out}(R) = \epsilon$, we can get

$$C_\epsilon = Wlog(1 + F^{-1}(1-\epsilon)\frac{P}{N_0 W}) \quad bit/s \tag{5.14}$$

where $F(x) = Pr\{|h|^2 > x\}$ is the complementary cumulative distribution function (CDF) of $|h|^2$.

When h is complex Gaussian $\mathcal{CN}(0,1)$, the outage probability $P_{out}(R)$ for Rayleigh fading can be derived as

$$\begin{aligned} P_{out}(R) &= Pr\{Wlog(1+|h|^2\frac{P}{N_0 W}) < R\} \\ &= Pr\{|h| < \sqrt{\frac{(2^{\frac{R}{W}}-1)N_0 W}{P}}\} \\ &\overset{(a)}{\approx} \frac{(2^{\frac{R}{W}}-1)N_0 W}{P} \end{aligned} \tag{5.15}$$

where (a) follows from the fact that $|h|$ follows Rayleigh distribution, with the probability density function (PDF) of $|h|$,

$$f(|h|) = \frac{2|h|}{\sigma^2} exp\{-\frac{|h|^2}{\sigma^2}\}, |h| > 0 \tag{5.16}$$

so that

$$\begin{aligned} &Pr\{|h| < \sqrt{\frac{(2^{\frac{R}{W}}-1)N_0 W}{P}}\} \\ &= \int_0^{\sqrt{\frac{(2^{\frac{R}{W}}-1)N_0 W}{P}}} f(|h|)d|h| \\ &= 1 - exp(-\frac{(2^{\frac{R}{W}}-1)N_0 W}{P}) \\ &\approx \frac{(2^{\frac{R}{W}}-1)N_0 W}{P} \end{aligned} \tag{5.17}$$

where the last step follows from the fact of the Taylor series for the exponential function e^{-x} that $e^{-x} \approx 1 - x$.

Observe that both the ϵ-outage capacity C_ϵ and the outage probability is related to the bandwidth W. With the increase of bandwidth W, the outage probability $P_{out}(R)$ decreases and the ϵ-outage capacity C_ϵ increases. Therefore, for the smart bandwidth allocation, when less bandwidth is assigned, in the fading channel, the outage probability increases, which means that the bandwidth could not be arbitrarily small with service performance guarantee. While with more bandwidth allocated, the ϵ-outage capacity C_ϵ increases, and $P_{out}(R)$ decreases, so that the service performance could be guaranteed for higher data transmission requirement.

Above is the relationship of the outage probability with bandwidth W in our smart bandwidth allocation scheme. Next, let's think about the effect of channel fading.

(1). For deep fading, $|h| \to 0$, with high probability that the channel capacity will be arbitrarily small that $C < R$, and the system will be in outage, therefore, more bandwidth W could be allocated to reduce the outage probability, as shown in Fig. 5.1.

$$P_{out1}(R) = Pr\{Wlog(1 + |h|^2 \frac{P}{N_0 W}) < R\} \tag{5.18}$$

(2). For shallow fading, the channel quality is good, so with high probability that $C > R$, reliable communication with small error probability could be achieved, and with very low probability that the system is in outage. Therefore, taking advantage of the better channel conditions, less bandwidth W could be assigned, as shown in Fig. 5.1.

$$P_{out2}(R) = Pr\{Wlog(1 + |h|^2 \frac{P}{N_0 W}) < R\} \tag{5.19}$$

with

$$0 < P_{out2}(R) < P_{out1}(R) < 1 \tag{5.20}$$

5.7 Numerical Results

A bandwidth allocation scheme is proposed in this chapter. Based on one smartphone user's data usage in one billing cycle, the service provider could calculate this user's probability of having each personality trait using diagnostic inference, and then based on predictive inference to calculate this user's usage of bandwidth in the next billing cycle. We can see that both diagnostic inference and predictive inference could help the service provider to better allocate the smartphone bandwidth based on each user's personality.

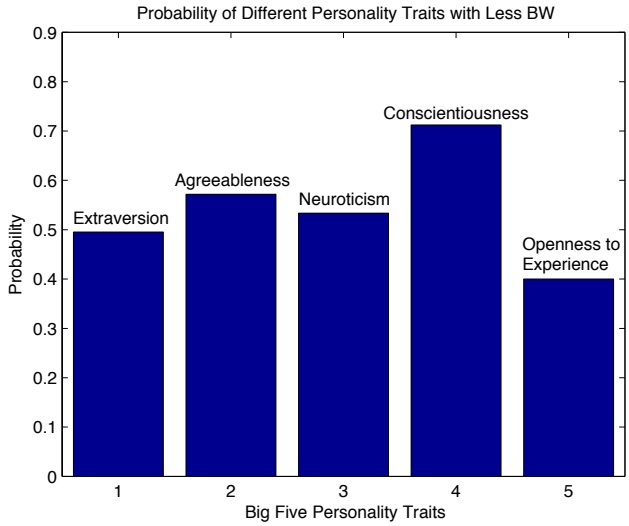

Figure 5.3. Diagnostic Inference for Personality Probability-LessBW.

Table 5.1 and Fig. 5.3 show the diagnostic inference for the probability of personality traits with less bandwidth allocated. We can observe that $P(Con = y|LessBW = y) = 0.7123$ with the highest probability, which means that when less bandwidth is occupied, with the probability of 0.7123 that the user scoring high on

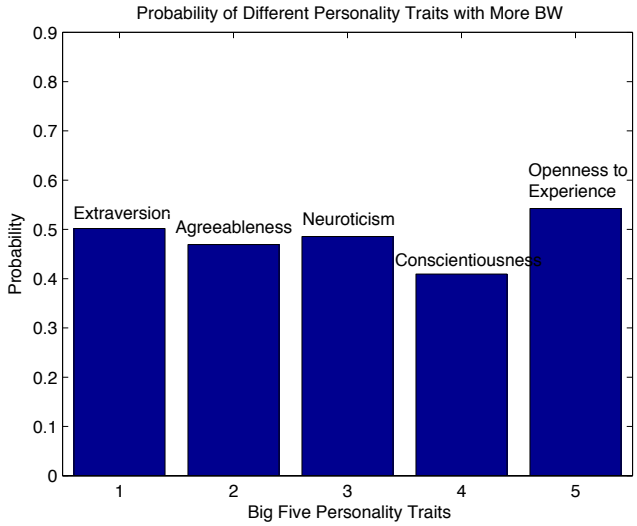

Figure 5.4. Diagnostic Inference for Personality Probability-MoreBW.

conscientiousness. We can notice that with less bandwidth allocated, users are most likely to score high on conscientiousness, least likely to be open-minded.

Table 5.1. Probability of Personality Traits with Less BW

Probability of Personality Traits with Less BW
$P(Ext = y \mid LessBW = y) = 0.4954$
$P(Agr = y \mid LessBW = y) = 0.5717$
$P(Neu = y \mid LessBW = y) = 0.5335$
$P(Con = y \mid LessBW = y) = \mathbf{0.7123}$
$P(Ope = y \mid LessBW = y) = 0.4000$

In the simulation, we assume the original probability for all personality traits is unknown, that is $P(Ext = y) = 0.5$, $P(Agr = y) = 0.5$, $P(Neu = y) = 0.5$, $P(Con = y) = 0.5$, and $P(Ope = y) = 0.5$.

Table 5.2 and Fig. 5.4 give the diagnostic inference for the probability of personality traits with more bandwidth. It can be observed that $P(Ope = y \mid MoreBW =$

Table 5.2. Probability of Personality Traits with More BW

Probability of Personality Traits with More BW
$P(Ext = y \mid MoreBW = y) = 0.5019$
$P(Agr = y \mid MoreBW = y) = 0.4694$
$P(Neu = y \mid MoreBW = y) = 0.4857$
$P(Con = y \mid MoreBW = y) = 0.4095$
$P(Ope = y \mid MoreBW = y) = \mathbf{0.5426}$

$y) = 0.5426$, which means that when more bandwidth is allocated, with the probability of 0.5426 that the user scoring high on openness. We can notice that with more bandwidth allocated, users are more likely to score high on openness.

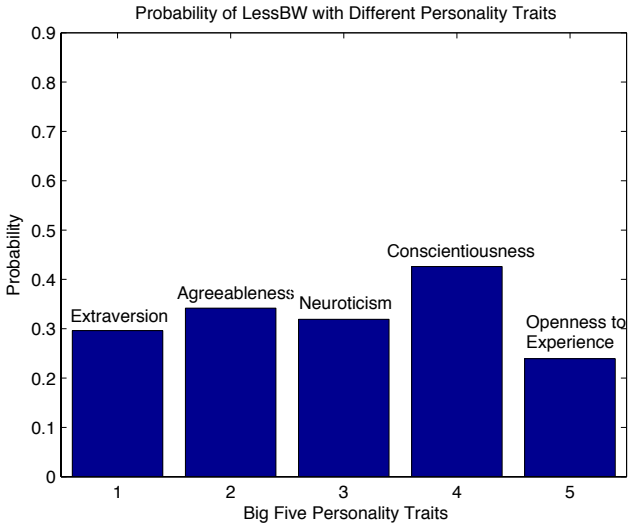

Figure 5.5. Predictive Inference for Personality Probability-LessBW.

Tables 5.3 and 5.4 and Figures 5.5 and 5.6 provide the predictive inference probability. Given a user's personality trait, the probability of this user occupying less bandwidth is shown. For an extravert person, the probability of less bandwidth occupation is 0.2962, while the probability of more bandwidth occupation is 0.7038.

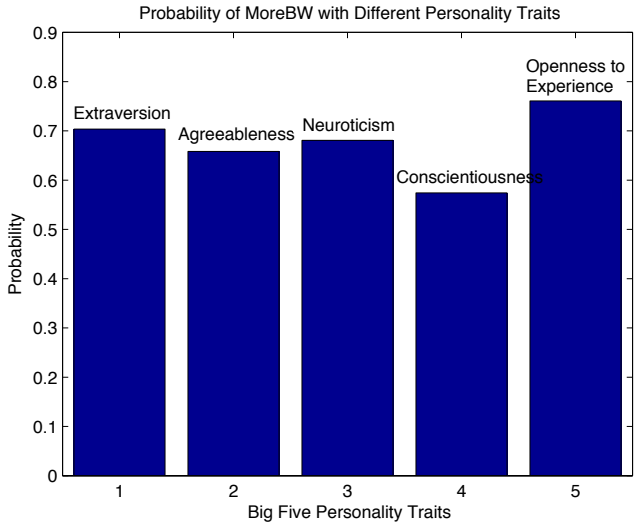

Figure 5.6. Predictive Inference for Personality Probability-MoreBW.

For an open-minded user, the case of less bandwidth is 0.2392, and more bandwidth is 0.7608. Open-minded user has the highest probability to occupy more bandwidth.

Table 5.3. Probability of LessBW with diff. Personality Traits

Probability of LessBW with diff. Personality Traits
$P(LessBW = y \mid Ext = y) = 0.2962$
$P(LessBW = y \mid Agr = y) = 0.3418$
$P(LessBW = y \mid Neu = y) = 0.3190$
$P(LessBW = y \mid Con = y) = 0.4259$
$P(LessBW = y \mid Ope = y) = 0.2392$

We can also observe that with the spread of smartphone usage, no matter what kind of personality traits, all kinds of personality have high probability of occupying larger bandwidth. That's why bandwidth is so precious nowadays, and smart bandwidth allocation is a good solution for smartphone users.

Table 5.4. Probability of MoreBW with diff. Personality Traits

Probability of MoreBW with diff. Personality Traits
$P(MoreBW = y \mid Ext = y) = \mathbf{0.7038}$
$P(MoreBW = y \mid Agr = y) = 0.6582$
$P(MoreBW = y \mid Neu = y) = 0.6810$
$P(MoreBW = y \mid Con = y) = 0.5741$
$P(MoreBW = y \mid Ope = y) = \mathbf{0.7608}$

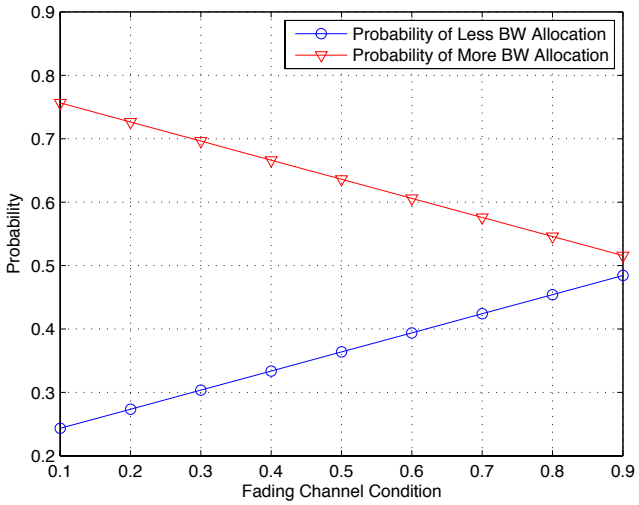

Figure 5.7. Probability of BW Allocation v.s. Fading Channel Condition.

Fig. 5.7 shows the probability of bandwidth allocation with fading channel conditions, with the assumption that the original probability for all personality traits is unknown, that is $P(Ext = y) = 0.5$, $P(Agr = y) = 0.5$, $P(Neu = y) = 0.5$, $P(Con = y) = 0.5$, and $P(Ope = y) = 0.5$. The x-axis is the probability of good channel condition (shallow fading), and the higher the better. $x = 0.1$ represents very deep channel fading, while $x = 0.9$ is shallow fading. It is obvious that with worse channel quality (deep fading), the probability is higher to assign more bandwidth. While it is highly likely to allocate less channel bandwidth when the channel condition

is good. This helps the service provider to better allocate the limited bandwidth for smartphone users as well as smartphone users' personality traits.

Manufacturers and service providers in the cell phone industry could provide more personal service to each user based on the results calculated. Personality traits of users could be estimated, using inference from this BN. For example, for two smartphone users A and B, in last bill cycle, user A used 2GB data, and the other one used 20MB data, with high probability that user A is open-minded, and user B is conscientious. Smartphone service provider could store each user's personality trait and provide the corresponding service to each user. Personalizing bandwidth allocation could be done based on smartphone users' personality traits, resulting in a smart and efficient usage of the limited bandwidth.

5.8 Conclusions

A bandwidth allocation method based on smartphone users' personality traits and channel condition is studied in a unified mathematical framework. Based on the relationship between user behavioral characteristics extracted from rich smartphone data and self-reported Big-Five personality traits, further research and usage of personality traits for personalizing services on smartphones could be practical using Bayesian Network. Based on one user's data usage, the service provider could estimate this user's probability of having each personality trait using diagnostic inference, and then based on predictive inference to calculate this user's usage of bandwidth in the future. We can see that both diagnostic inference and predictive inference could help the service provider to better allocate smartphone bandwidth based on each user's personality. Personalizing bandwidth allocation could be done based on smartphone users' personality traits, resulting in a smart and efficient usage of the limited bandwidth.

For our proposed smart bandwidth allocation scheme, both the ϵ-outage capacity C_ϵ and the outage probability are studied in fading channel. For a smartphone user with high probability of more bandwidth requirement, the outage capacity C_ϵ increases, and $P_{out}(R)$ decreases, so that the service performance could be guaranteed for higher data transmission requirement. While in deep fading scenario, the outage probability will become too high to support the service. One possible solution is to assign more bandwidth to guarantee the service performance.

Table 5.5. CPT of Calls

Ext	Agr	Neu	Con	Ope	C	$P(C\|Ext,\cdots,Ope)$
y	y	y	y	y	y	0.69
y	y	y	y	y	n	0.31
y	y	y	y	n	y	0.6
y	y	y	y	n	n	0.4
y	y	y	n	y	y	0.99
y	y	y	n	y	n	0.01
y	y	n	y	y	y	0.7
y	y	n	y	y	n	0.3
y	n	y	y	y	y	0.54
y	n	y	y	y	n	0.46
n	y	y	y	y	y	0.24
n	y	y	y	y	n	0.76
y	y	y	n	n	y	0.9
y	y	y	n	n	n	0.1
y	y	n	y	n	y	0.61
y	y	n	y	n	n	0.39
y	y	n	n	y	y	0.999
y	y	n	n	y	n	0.001
y	y	n	n	n	y	0.91
y	y	n	n	n	n	0.09
y	n	y	y	n	y	0.45
y	n	y	y	n	n	0.55
y	n	y	n	y	y	0.84
y	n	y	n	y	n	0.16
y	n	y	n	n	y	0.75
y	n	y	n	n	y	0.25
y	n	n	y	y	y	0.55
y	n	n	y	y	n	0.45
y	n	n	y	n	y	0.46
y	n	n	y	n	n	0.54
y	n	n	n	y	y	0.85
y	n	n	n	y	n	0.15
y	n	n	n	n	y	0.76
y	n	n	n	n	n	0.24
n	y	y	y	n	y	0.15
n	y	y	y	n	n	0.85
n	y	y	n	y	y	0.54
n	y	y	n	y	n	0.46
n	y	y	n	n	y	0.45
n	y	y	n	n	n	0.55

Table 5.6. CPT of Calls

Ext	Agr	Neu	Con	Ope	C	$P(C\|Ext, \cdots, Ope)$
n	y	n	y	y	y	0.25
n	y	n	y	y	n	0.75
n	y	n	y	n	y	0.16
n	y	n	y	n	n	0.84
n	y	n	n	y	y	0.55
n	y	n	n	y	n	0.45
n	y	n	n	n	y	0.46
n	y	n	n	n	n	0.54
n	n	y	y	y	y	0.09
n	n	y	y	y	n	0.91
n	n	y	y	n	y	0.001
n	n	y	y	n	n	0.999
n	n	y	n	y	y	0.39
n	n	y	n	y	n	0.61
n	n	y	n	n	y	0.3
n	n	y	n	n	n	0.7
n	n	n	y	y	y	0.1
n	n	n	y	y	n	0.9
n	n	n	y	n	y	0.01
n	n	n	y	n	n	0.99
n	n	n	n	y	y	0.4
n	n	n	n	y	n	0.6
n	n	n	n	n	y	0.31
n	n	n	n	n	n	0.69

Table 5.7. CPT of Less BW

C	Y	V	S	M	I	LBW	$P(LBW\|C,\cdots,I)$
y	y	y	y	y	y	y	0.001
y	y	y	y	y	y	n	0.999
y	y	y	y	y	n	y	0.1
y	y	y	y	y	n	n	0.9
y	y	y	y	n	y	y	0.05
y	y	y	y	n	y	n	0.95
y	y	y	y	n	n	y	0.1
y	y	y	y	n	n	n	0.9
y	y	y	n	y	y	y	0.05
y	y	y	n	y	y	n	0.95
y	y	y	n	y	n	y	0.1
y	y	y	n	y	n	n	0.9
y	y	y	n	n	y	y	0.05
y	y	y	n	n	y	n	0.95
y	y	y	n	n	n	y	0.1
y	y	y	n	n	n	n	0.9
y	y	n	y	y	y	y	0.15
y	y	n	y	y	y	n	0.85
y	y	n	y	y	n	y	0.2
y	y	n	y	y	n	n	0.8
y	y	n	y	n	y	y	0.15
y	y	n	y	n	y	n	0.85
y	y	n	y	n	n	y	0.2
y	y	n	y	n	n	n	0.8
y	y	n	n	y	y	y	0.15
y	y	n	n	y	y	n	0.85
y	y	n	n	y	n	y	0.2
y	y	n	n	y	n	n	0.8
y	y	n	n	n	y	y	0.15
y	y	n	n	n	y	n	0.85
y	y	n	n	n	n	y	0.2
y	y	n	n	n	n	n	0.8

Table 5.8. CPT of Less BW

C	Y	V	S	M	I	LBW	$P(LBW\|C,\cdots,I)$
y	n	y	y	y	y	y	0.15
y	n	y	y	y	y	n	0.85
y	n	y	y	y	n	y	0.2
y	n	y	y	y	n	n	0.8
y	n	y	y	n	y	y	0.15
y	n	y	y	n	y	n	0.85
y	n	y	y	n	n	y	0.2
y	n	y	y	n	n	n	0.8
y	n	y	n	y	y	y	0.15
y	n	y	n	y	y	n	0.85
y	n	y	n	y	n	y	0.2
y	n	y	n	y	n	n	0.8
y	n	y	n	n	y	y	0.15
y	n	y	n	n	y	n	0.85
y	n	y	n	n	n	y	0.2
y	n	y	n	n	n	n	0.8
y	n	n	y	y	y	y	0.45
y	n	n	y	y	y	n	0.55
y	n	n	y	y	n	y	0.999
y	n	n	y	y	n	n	0.001
y	n	n	y	n	y	y	0.45
y	n	n	y	n	y	n	0.55
y	n	n	y	n	n	y	0.99
y	n	n	y	n	n	n	0.01
y	n	n	n	y	y	y	0.45
y	n	n	n	y	y	n	0.55
y	n	n	n	y	n	y	0.95
y	n	n	n	y	n	n	0.05
y	n	n	n	n	y	y	0.45
y	n	n	n	n	y	n	0.55
y	n	n	n	n	n	y	0.99
y	n	n	n	n	n	n	0.01

Table 5.9. CPT of Less BW (Cont.)

C	Y	V	S	M	I	LBW	$P(LBW\|C,\cdots,I)$
n	y	y	y	y	y	y	0.05
n	y	y	y	y	y	n	0.95
n	y	y	y	y	n	y	0.1
n	y	y	y	y	n	n	0.9
n	y	y	y	n	y	y	0.05
n	y	y	y	n	y	n	0.95
n	y	y	y	n	n	y	0.1
n	y	y	y	n	n	n	0.9
n	y	y	n	y	y	y	0.05
n	y	y	n	y	y	n	0.95
n	y	y	n	y	n	y	0.1
n	y	y	n	y	n	n	0.9
n	y	y	n	n	y	y	0.001
n	y	y	n	n	y	n	0.999
n	y	y	n	n	n	y	0.1
n	y	y	n	n	n	n	0.9
n	y	n	y	y	y	y	0.15
n	y	n	y	y	y	n	0.85
n	y	n	y	y	n	y	0.2
n	y	n	y	y	n	n	0.8
n	y	n	y	n	y	y	0.15
n	y	n	y	n	y	n	0.85
n	y	n	y	n	n	y	0.25
n	y	n	y	n	n	n	0.75
n	y	n	n	y	y	y	0.15
n	y	n	n	y	y	n	0.85
n	y	n	n	y	n	y	0.2
n	y	n	n	y	n	n	0.8
n	y	n	n	n	y	y	0.15
n	y	n	n	n	y	n	0.85
n	y	n	n	n	n	y	0.2
n	y	n	n	n	n	n	0.8
n	n	y	y	y	y	y	0.15
n	n	y	y	y	y	n	0.85
n	n	y	y	y	n	y	0.2
n	n	y	y	y	n	n	0.8

Table 5.10. CPT of Less BW (Cont.)

C	Y	V	S	M	I	LBW	$P(LBW\|C,\cdots,I)$
n	n	y	y	n	y	y	0.15
n	n	y	y	n	y	n	0.85
n	n	y	y	n	n	y	0.2
n	n	y	y	n	n	n	0.8
n	n	y	n	y	y	y	0.15
n	n	y	n	y	y	n	0.85
n	n	y	n	y	n	y	0.2
n	n	y	n	y	n	n	0.8
n	n	y	n	n	y	y	0.15
n	n	y	n	n	y	n	0.85
n	n	y	n	n	n	y	0.2
n	n	y	n	n	n	n	0.8
n	n	n	y	y	y	y	0.45
n	n	n	y	y	y	n	0.55
n	n	n	y	y	n	y	0.99
n	n	n	y	y	n	n	0.01
n	n	n	y	n	y	y	0.45
n	n	n	y	n	y	n	0.55
n	n	n	y	n	n	y	0.99
n	n	n	y	n	n	n	0.01
n	n	n	n	y	y	y	0.45
n	n	n	n	y	y	n	0.55
n	n	n	n	y	n	y	0.9
n	n	n	n	y	n	n	0.1
n	n	n	n	n	y	y	0.45
n	n	n	n	n	y	n	0.55
n	n	n	n	n	n	y	0.999
n	n	n	n	n	n	n	0.001

Chapter 6

Conclusions and Future Works

This chapter concludes this dissertation. It begins with a summary of the dissertation results and contributions, follows with a discussion of future research directions of sparse sensing in Big Data.

6.1 Summary

In this dissertation, several approaches of sparse sensing in Big Data - compressive sensing, co-prime sampling and nested sampling are studied, in theory and applications. The contributions of this dissertation are:

- Error performance bounds of noisy compressive sensing (CS) were derived based on information theory and estimation theory. Numerical results showed that the probability of error is related with the sampling ratio and the signal-to-noise ratio (SNR).
- Information rate distortion function is a measure as the number of bits per symbol to be stored or transmitted under the constraint of a distortion. Rate distortion performance for scalar quantization of measurement observation was derived. Based on this, reconstruction rate distortion was studied for compressive sensing as well.
- Real-world applications of CS in Big Data were studied, to Synthetic Aperture Radar (SAR), radar sensor networks (RSNs), and underwater acoustic sensor networks (UWASNs). By comparing different reconstruction methods, we concluded that OMP algorithm could be chosen for the reconstruction of a huge

amount of real-world Big Data. An efficient and effective sampling algorithm was proposed to compress the large-scale data gathered in real-world Radar Sensor Networks based on compressive sensing. Numerical results showed that our algorithm is more efficient with correlations among the sensor data considered, without introducing more computation complexity. The rate distortion performance in UWASNs was studied, with or without correlation among sensor readings.

- Properties of two new sparse sampling algorithms - coprime sampling and nested sampling were investigated. Rate distortion function was constructed and analyzed which can determine the bits per symbol, since sparse sampling can cause possible distortion because less number of samples are used.
- Theoretical analysis of how coprime sampling and nested sampling effect the power spectral density was given. Simulation results showed that with if we choose the sampling spacings larger, the main lobe of PSD obtained from these two sampling will be much narrower than the original QPSK signal with increased spectrum efficiency.
- A secure transmission scheme for Big Data based on coprime sampling and nested sampling was provided. When the sampling spacing pairs bigger enough, the spectrum of BFSK signal performs like frequency hopping. This property has great advantage in the security of Big Data collection and transmission using FH/BFSK, as it could achieve low error probability in Rician fading channels.
- A hybrid approach of sparse sensing was proposed, which combines nested sampling and compressive sensing to reduce the number of symbols, and rate distortion function was used as a criteria to determine how many bits should be used to represent the symbols during this process. We showed that with

this hybrid approach, less number of bits was required to represent the sensed information. This hybrid approach has great advantage in the application of Big Data nowadays, especially for some application which depends on the difference co-array, or autocorrelation, like Direction-of-arrival (DOA) estimation and beamforming.

- A bandwidth allocation method based on smartphone users personality traits and channel condition was studied in a unified mathematical framework. Since LTE has been a Big Data consumer with ample data, how to deal with these Big Data in telecommunications becomes a new issue. Personality traits of smartphone users could be estimated. Manufacturers and service providers in the cell phone industry could provide more personal service to each user. Therefore, personalizing bandwidth allocation could be done based on smartphone users personality traits and channel condition, resulting in a smart and efficient usage of the limited bandwidth.

6.2 Future Directions

In future research, we would like to continue in the following aspects.

6.2.1 For nested sampling and coprime sampling

We would like study the effect of sampling spacings for nested sampling and coprime sampling theoretically. For example, with $N_1 \times N_2 = 30$, there are several combinations of N_1, N_2, like $N_1 = 5, N_2 = 6$, $N_1 = 3, N_2 = 10$, and $N_1 = 2, N_2 = 15$. This indicates that for the same interval, a lot of sampling schemes exist, for nested sampling and coprime sampling. For the case of $N_1 = 2, N_2 = 15$, the chosen samples are much sparse than that of $N_1 = 5, N_2 = 6$. How these different combinations of sampling spacings will affect the performance of nested sampling and coprime

sampling? One we would like to analyze is the rate distortion performance, and the other one is spectrum efficiency performance.

(1). Rate Distortion Performance

we may not only analyze how the rate distortion performance will change with different nested sampling spacing pairs N_1, N_2 (one special case is that $N_1 + N_2$ keeps the same, with different N_1, N_2 combination), but also analyze the real application performance of the hybrid approach, for example, in some real estimation of DOA, to see how the estimation error rate of DOA changes with different nested sampling intervals.

(2). Spectrum Efficiency We noticed that with the proper choice of sampling intervals, i.e., making them large enough, the main lobe of PSD obtained from both nested sampling and coprime sampling is much narrower than the original QPSK signal. If the sampling intervals are chosen larger, the bandwidth occupied will be narrower, which improves the spectrum efficiency.

We would like to continue the research to analyze how sampling intervals will affect spectrum efficiency with different combination. For example, Figures 6.1 and 6.2 give the corresponding PSD with nested sampling $N_1 = 5, N_2 = 6$, and $N_1 = 3, N_2 = 10$. And Figures 6.3 and 6.4 show the corresponding PSD with coprime sampling $M = 5, N = 6$, and $M = 3, N = 10$. It is obvious that with $N_1 = 3, N_2 = 10$, and $M = 3, N = 10$, the chosen samples are much sparse than that of $N_1 = 5, N_2 = 6$ and $M = 3, N = 10$ for both nested sampling and coprime sampling.

6.2.2 Dynamic Bandwidth Allocation

For bandwidth allocation, in our current research, based on one user's data usage, the service provider estimate each user's probability of having each personality trait using diagnostic inference, and then based on predictive inference to calculate

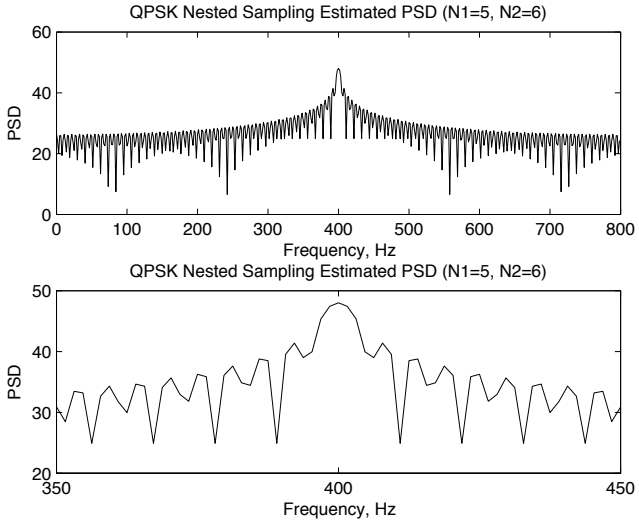

Figure 6.1. Nested Sampling with $M = 5, N = 6$.

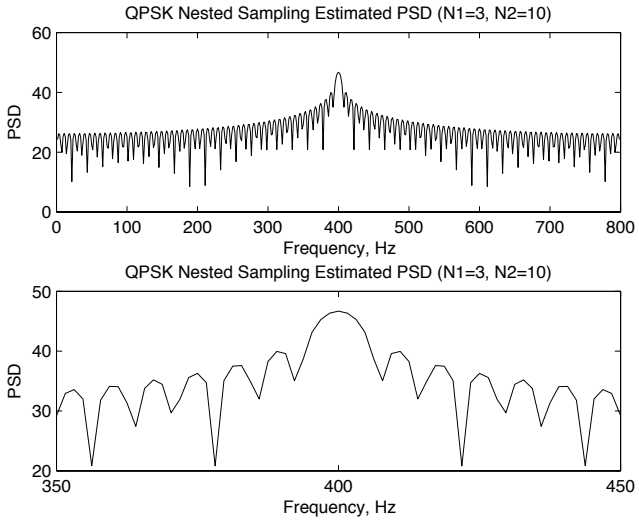

Figure 6.2. Nested Sampling with $N1 = 3, N2 = 10$.

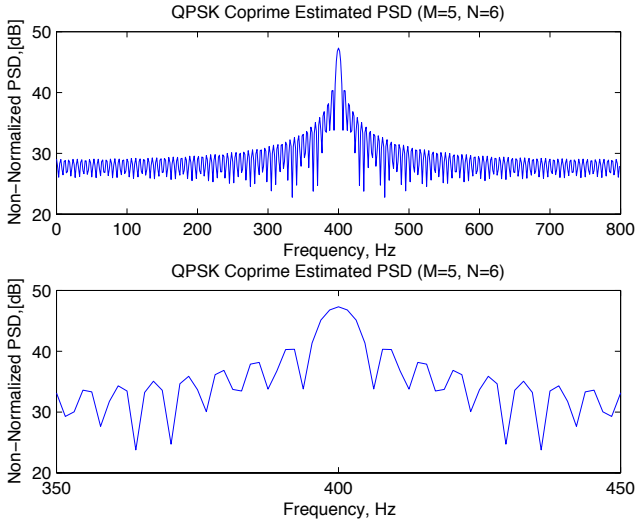

Figure 6.3. Coprime Sampling with $M = 5, N = 6$.

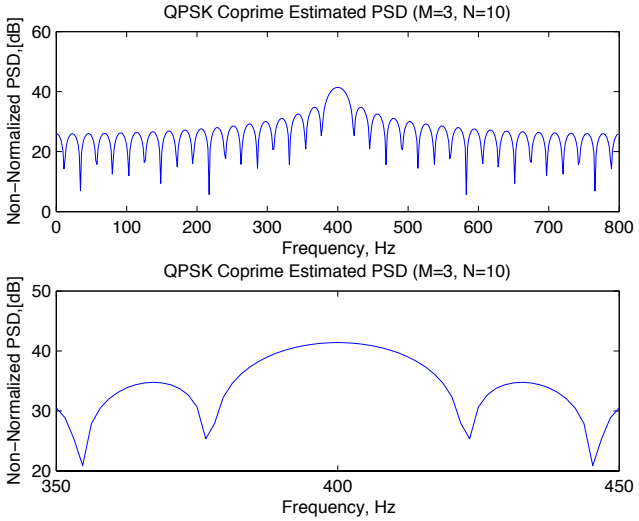

Figure 6.4. Coprime Sampling with $M = 3, N = 10$.

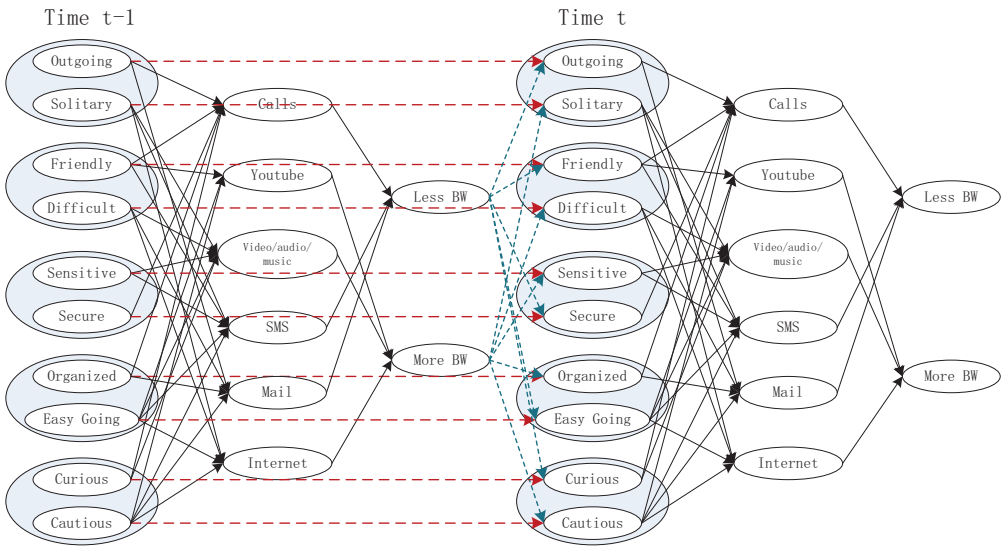

Figure 6.5. Dynamic Bayesian Network Inference of Smartphone Usage.

this user's usage of bandwidth in next billing cycle. Both diagnostic inference and predictive inference are calculated based on Bayesian network in one billing cycle to allocate smartphone bandwidth. In the future work, to get more accurate prediction, dynamic Bayesian network (DBN) could be considered to dynamic adjust each user's bandwidth real-timely, as shown in Fig. 6.5. For example, one person has recognized that he/she has already used 2G bits data in the first week of billing cycle (time $t-1$), he/she may act more organize, and more sensitive after that (time t), to try to control his data usage.

Appendix A

Publications

1. J. Chen, and Q. Liang, "Noisy Compressive Sensing with Application to Underwater Acoustic Sensor Networks," *submitted to IEEE Transactions on Parallel and Distributed Systems*.
2. J. Chen, and Q. Liang, "Theoretical Performance Bounds for Compressive Sensing with Random Noise," *submitted to IEEE Transactions on Signal Processing*.
3. J. Chen, and Q. Liang, "Bandwidth Allocation Based on Personality Traits on Smartphone Usage and Channel Condition," *submitted to IEEE Transactions on Neural Networks and Learning Systems*.
4. J. Chen, and Q. Liang, "A Hybrid Approach of Sparse Sampling for Big Data Analysis and Compression," *submitted to IEEEJournal of Selected Topics in Signal Processing Special Issue on Signal Processing for Big Data*.
5. J. Chen, and Q. Liang, "Rate Distortion Performance Analysis of Nested Sampling and Coprime Sampling," *accepted for publishing by EURASIP Journal on Advances in Signal Processing on 15 January 2014*.
6. J. Chen, Q. Liang and J. Wang, "Secure transmission for big data based on nested sampling and coprime sampling with spectrum efficiency," *Security and Communication Networks*, Wiley, Security Comm. Networks (2013), DOI: 10.1002/sec.785.
7. J. Chen, and Q. Liang, "Efficient Sampling for Radar Sensor Networks," *accepted for publishing by International Journal of Sensor Networks 2013*.
8. J. Chen, Q. Liang, B. Zhang, X. Wu, "Spectrum Efficiency of Nested Sparse Sampling and Co-Prime Sampling," *EURASIP Journal on Wireless Communications and Networking 2013*, no. 1 (2013): 47.

9. J. Chen, and Q. Liang, "Information-Theoretic Performance Analysis for Compressive Sensing with Random Noise," *submitted to IEEE Global Telecommunications Conference 2014 (GLOBECOM 2014).*
10. J. Chen, and Q. Liang, "Bandwidth Allocation Based on Personality Traits on Smartphone Usage and Channel Condition," *submitted to the 3rd International Conference on Communications, Signal Processing, and Systems (CSPS'14).*
11. J. Chen, Q. Liang, B. Zhang and X. Wu, "A New Secure Transmission for Big Data Based on Nested Sampling and Coprime Sampling," *accepted by the 2rd International Conference on Communications, Signal Processing, and Systems (CSPS'13).*
12. J. Chen, and Q. Liang, "Theoretical Performance Limits for Compressive Sensing with Random Noise," *accepted by Global Telecommunications Conference (GLOBECOM 2013)*, Dec. 2013, Atlanta, GA, USA.
13. J. Chen, Q. Liang, B. Zhang and X. Wu, "Information Theoretic Performance Bounds for Noisy Compressive Sensing," *IEEE ICC 2013 Workshop on Radar and Sonar Networks (RSN)*, June 2013, Budapest, Hungary.
14. J. Chen, Q. Liang, J. Paden, and P. Gogineni, "Compressive Sensing Analysis of Synthetic Aperture Radar Raw Data," *2012 IEEE International Conference on Communications (ICC)*, 2012, pp. 6362-6366.
15. J. Chen, Q. Liang, J. Wang, and H.-A. Choi, "Spectrum Efficiency of Nested Sparse Sampling," *Wireless Algorithms, Systems, and Applications, Springer Berlin Heidelberg*, 2012, pp. 574-583.
16. J. Chen, and Q. Liang, "Rate Distortion Performance Analysis of Compressive Sensing," *Global Telecommunications Conference (GLOBECOM 2011)*, Dec. 2011, Houston, TX, USA, pp. 1-5.

17. J. Chen, J. Liang, Q. Liang, W. Wang, B. Zhang, "Passive RFID Customer Service Enhancement Using Fuzzy Logic," *IEEE 11th International Symposium on Communications and Information Technologies 2011 (ISCIT'11)*, pp. 194-198.

18. T. Liu, and J. Chen, "Joint downlink transmit and receive beamforming under per-antenna power constraints," *IEEE 7th International Conference on Information, Communications and Signal Processing, 2009 (ICICS 2009)*, pp. 1-5.

19. X. Li, T. Wang, J. Chen, Z. Qian, J. K. Pollard, S. Liu, and J. Yu, "Customer service enhancement using passive RFID," *IEEE International Conference on Communications Technology and Applications, 2009 (ICCTA'09)*, pp. 5-9.

References

[1] J. Bughin, M. Chui, J. Manyika, "Clouds, big data, and smart assets: Ten tech-enabled business trends to watch," *McKinsey Quarterly, 2010*.

[2] Computing Research Association, "Challenges and Opportunities with Big Data," *Big Data White Paper*.

[3] S. Kaisler, F. Armour, and A. Espinosa, "Introduction to Big Data: Scalable Representation and Analytics for Data Science Minitrack," In System Sciences (HICSS), 2013 46th Hawaii International Conference on, pp. 984-984. IEEE, 2013.

[4] D. Bollier, "The Promise and Peril of Big Data, " *The Aspen Institute, 2010*.

[5] S. Shekhar, M. R. Evans, V. Gunturi, K. Yang, "Spatial Big-Data Challenges Intersecting Mobility and Cloud Computing,"*MobiDE '12, Proceedings of the Eleventh ACM International Workshop on Data Engineering for Wireless and Mobile Access*, pp. 1-6.

[6] H. Herodotou, H. Lim, G. Luo, N. Borisov, L. Dong, F. B. Cetin, S. Babu, "Starfish: A Self-tuning System for Big Data Analytics," *Proc. of the 5th Conference on Innovative Data Systems Research (CIDR '11)*, Jan. 2011, pp. 261-272.

[7] J. K. Laurila, D. Gatica-Perez, I. Aad, J. Blom, O. Bornet, T.-M.-T. Do, O. Dousse, J. Eberle, and M. Miettinen, "The mobile data challenge: Big data for mobile computing research," *the Workshop on the Nokia Mobile Data Challenge, in Conjunction with the 10th International Conference on Pervasive Computing, 2012*, pp. 1-8.

[8] J. Sanyal, and J. New, "Simulation and big data challenges in tuning building energy models," *IEEE, 2013 Workshop on Modeling and Simulation of Cyber-Physical Energy Systems (MSCPES)*, pp. 1-6.

[9] P. Bennett, L. Giles, A. Halevy, J. Han, M. Hearst, and J. Leskovec, "Channeling the deluge: research challenges for big data and information systems," *the 22nd ACM international conference on Conference on information & knowledge management*, ACM, 2013. pp. 2537-2538.

[10] N. Leavitt,"Storage Challenge: Where Will All That Big Data Go?," *Computer 46*, no. 9 (2013): 22-25.

[11] R. M. Ward, R. Schmieder, G. Highnam, and D. Mittelman, "Big data challenges and opportunities in high-throughput sequencing," *Systems Biomedicine 1*, no. 1 (2013): 0-1.

[12] H. Chen, R. H. L. Chiang, V. C. Storey, "Business intelligence and analytics: From big data to big impact," *MIS Quarterly* 36 (4) (2012), pp. 1165-1188.

[13] M. D. Assuncao, R. N. Calheiros, S. Bianchi, M. A. Netto, and R. Buyya, "Big Data Computing and Clouds: Challenges, Solutions, and Future Directions," *arXiv preprint arXiv:1312.4722 (2013)*.

[14] J. Cohen, B. Dolan, M. Dunlap, J. M. Hellerstein, C. Welton, "MAD skills: new analysis practices for big data," *Proc. VLDB Endow. 2 (2) (2009)*, pp. 1481-1492.

[15] X. Sun, B. Gao, L. Fan, W. An, "A cost-effective approach to delivering analytics as a service," *19th IEEE International Conference on Web Services (ICWS 2012)*, Honolulu, USA, 2012, pp. 512-519.

[16] R. G. Baraniuk, "Compressive Sensing," *IEEE Signal Processing Magazine*, July 2007, PP. 118-124.

[17] E. J. Candes, "Compressive sampling," *Proceedings of the International Congress of Mathematicians*, Madrid, Spain, 2006.

[18] E. J. Candes, M. B. Wakin, "An Introduction To Compressive Sampling", *IEEE Signal Processing Magazing*, March 2008, pp. 21-30.

[19] R. G. Baraniuk, E. J. Candes, R. Nowak, and M. Vetterli, "Compressive Sampling," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 12-13, Mar. 2008.

[20] R. G. Baraniuk, V. Cevher, M. F. Duarte, and C. Hegde, "Model-based compressive sensing," *IEEE Transactions on Information Theory*, vol. 56, no. 4 (2010), pp. 1982-2001.

[21] G. Reeves, and M. Gastpar, "Approximate sparsity pattern recovery: Information-theoretic lower bounds," arXiv preprint arXiv:1002.4458 (2010).

[22] B. Babadi, N. Kalouptsidis, and V. Tarokh, "Asymptotic achievability of the Cramer-Rao bound for noisy compressive sampling," *Signal Processing, IEEE Transactions on 57*, no. 3 (2009): 1233-1236.

[23] R. Niazadeh, M. Babaie-Zadeh, and C. Jutten, "On the Achievability of Cramer-Rao Bound in Noisy Compressed Sensing," *Signal Processing, IEEE Transactions on 60*, no. 1 (2012): 518-526.

[24] M. Akcakaya, and V. Tarokh, "Shannon-theoretic limits on noisy compressive sampling," *Information Theory, IEEE Transactions on 56*, no. 1 (2010): 492-504.

[25] Z. Ben-Haim, and Y. C. Eldar, "Performance bounds for sparse estimation with random noise," *Statistical Signal Processing, 2009, IEEE/SP 15th Workshop on*, pp. 225-228.

[26] V. K Goyal, A. K. Fletcher, and S. Rangan, "Compressive Sampling and Lossy Compression," *IEEE Signal Processing Magazine*, March 2008, PP. 48-56.

[27] D. Kirachaiwanich, and Q. Liang, "Compressive sensing: to compress or not to compress," *Signals, Systems and Computers (ASILOMAR), 2011 Conference Record of the Forty Fifth Asilomar Conference on*, pp. 809-813.

[28] W. Dai, H. V. Pham, and O. Milenkovic, "Distortion-Rate Functions for Quantized Compressive Sensing," *2009 IEEE Information Theory Workshop on Networking and Information Theory (ITW 2009)*, Volos, Greece, June 10 - 12, 2009, PP. 171-175.

[29] H. Gish, and J. Pierce, "Asymptotically efficient quantizing," *IEEE Transactions on Information Theory* 14, no. 5 (1968), 676-683.

[30] V. Sheinin, and A. Jagmohan, "Low Rate Uniform Scalar Quantization of Memoryless Gaussian Sources," *2006 IEEE International Conference on Image Processing*, pp. 793-796. IEEE, 2006.

[31] P. Zador, "Development and evaluation of procedures for quantizing multivariate distributions," PhD thesis, Stanford University, Stanford, CA, 1964.

[32] A. Schulz, L. Velho, E. A. B. da Silva, "On the Empirical Rate-Distortion Performance of Compressive Sensing," *ICIP 2009*, PP. 3049-3052.

[33] A. K. Fletcher, S. Rangan, and V. K Goyal, "On the Rate-Distortion Performance of Compressed Sensing," *Acoustics, Speech and Signal Processing, 2007*, pp. III 885-III 888.

[34] A. K. Fletcher, S. Rangan, and V. K Goyal, "Rate-Distortion Bounds For Sparse Approximation," *2007 IEEE/SP 14th Workshop on Statistical Signal Processing*, pp. 254-258.

[35] B. Mulgrew and M. E. Davies, "Approximate lower bounds for rate-distortion in compressive sensing systems," *ICASSP 2008*, pp. 3849-3852.

[36] T. M. Cover, J. A. Thomas, "Elements of Information Theory," Second Edition, 2006, John Wiley & Sons, Inc., Hoboken, New Jersey.

[37] S. Kay, "Fundamentals of Statistical Signal Processing: Estimation Theory," Prentice-Hall Signal Processing Series, 1993.

[38] J. Haupt and P. Nowak, "signal reconstruction from noisy random projections," *IEEE Transactions Information Theory*, Sept. 2006, pp. 4036-4048.

[39] A. K. Fletcher, S. Rangan, V. K. Goyal, and K. Ramchandran, "Denoising by sparse approximation: Error bounds based on rate-distortion theory," *EURASIP Journal on Applied Signal Processing 2006*, vol. 10, pp. 1-19.

[40] S. Aeron, V. Saligrama, and M. Zhao, "Information theoretic bounds for Compressed Sensing," *IEEE Trans. on Information Theory*, Oct. 2010, vol. 56, pp. 5111-5130.

[41] M. Capdevila, O.W.M. Florez, "A Communication Perspective on Automatic Text Categorization", *IEEE Transactions on Knowledge and Data Engineering*, July 2009, vol 21, pp. 1027-1041.

[42] J. Romberg, M. Wakin, "Compressed Sensing: A Tutorial", *IEEE Statistical Signal Processing Workshop*, Madison, Wisconsin, August 26, 2007.

[43] B. Babadi, N. Kalouptsidis, and V. Tarokh, "Asymptotic Achievability of the CramrCRao Bound forNoisy Compressive Sampling", *IEEE Transactions on Signal Processing*, vol. 57, no. 3, March 2009. pp.1233 - 1236.

[44] R. Niazadeh, M. Babaie-Zadeh, C. Jutten, "On the Achievability of Cramer-Rao Bound In Noisy Compressed Sensing", eprint arXiv:1006.2513,06/2010.

[45] J. G. Proakis, *Digital Communications*, Fourth Edition, McGraw-Hill.

[46] Z. Charbiwala, S. Chakraborty, S. Zahedi, Y. Kim, M. B. Srivastava, T. He, C. Bisdikian, "Compressive Oversampling for Robust Data Transmission in Sensor Networks," *IEEE INFOCOM, March 2010*, pp. 1-9.

[47] J. C. Curlander, R. N. McDonough, "Synthetic Aperture Radar: Systems and Signal Processing", 1991 by John Wiley & Sons Inc.

[48] S. Ranea, P. Boufounosa, A. Vetroa and Y. Okadab, "Low Complexity Efficient Raw SAR Data Compression", *Mitsubishi Electric Research Laboratories*, Inc., 2011.

[49] R. Kwok and W. T. K. Jhonson, "Block adaptive quantization of Magellan SAR data", *IEEE Transactions on Geoscience and Remote Sensing*, Jul 1989, vol 27, PP. 375-383.

[50] U. Benz, K. Strodl, and A. Moreira, "A comparison of several algorithms for SAR raw data compression", *IEEE Transactions on Geoscience and Remote Sensing*, Sep 1995, vol 33, PP. 1266-1276.

[51] J. M. Moureaux, P. Gauthier, M. Barlaud, and P. Bellemain, "Raw SAR data compression using vector quantization", *International Journal of Remote Sensing*, 1995, vol 16, PP. 3179-3187.

[52] M. Wang, "Raw SAR data compression by structurally random matrix based compressive sampling", *APSAR 2009*, PP. 1119-1122.

[53] J. Paden, C. Allen, P. Gogineni, "3D IMAGING OF ICE SHEETS", *2010 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, PP. 2611 - 2613.

[54] E. Magli, G. Olmo, B. Penna, "Wavelet-based compression of SAR raw data", *Geoscience and Remote Sensing Symposium, IGARSS '02*, June 2002, vol.2, PP. 1129-1131.

[55] V. Pascazio, G. Schirinzi, "SAR Phase History Data Compression by Using Wavelet Packets", *IGARSS 2000*, vol.6, PP. 2639-2641.

[56] N. Terzija, M. Repges, K. Luck, and W. Geisselhardt, "Digital Image Watermarking using Discrete Wavelet Transform: Performance Comparison of Error Correction Codes", *Visualization, Imaging, and Image Processing*, 2002.

[57] S. Chen, D. Donoho, "Basis Pursuit", *Conference Record of the Twenty-Eighth Asilomar Conference on Signals, Systems and Computers*, 1994, vol.1, PP. 41-44.

[58] S.e Mallat, Z. Zhang,"Matching pursuit with time-frequency dictionaries," *IEEE Transactions on Signal Processing*, 41(12):3397-3415, Dec 1993.

[59] S. Qian, D. Chen, and K.Chen. "Signal approximation via data-adaptive normalized gaussian function and its applications for speech processing", *In ICASSP-1992*, March 23-26 1992, pages 141-144.

[60] Y. Pati, R. Rezaiifar, P. Krishnaprasad, "Orthogonal Matching Pursuit : recursive function approximation with application to wavelet decomposition", *Conference Record of The Twenty-Seventh Asilomar Conference on Signals, Systems and Computers*, 1993, vol.1, PP. 40-44.

[61] P.K. Dutta, A.K. Arora, and S.B. Bibyk, (2006) "Towards Radar-Enabled Sensor Networks," *Fifth Intl Conf. Information Processing in Sensor Networks*, pp. 467-474.

[62] H. D. Ly, Q. Liang, (2009) "Diversity in Radar Sensor Networks: Theoretical Analysis and Application to Target Detection," *Int. J. Wireless Inf Networks*, Vol. 16, pp. 209-216.

[63] J. Li, and J. Li, "Data sampling control, compression and query in sensor networks," *International Journal of Sensor Networks* 2, no. 1 (2007): 53-61.

[64] S. Lin, V. Kalogeraki, D. Gunopulos, and S. Lonardi, "Efficient information compression in sensor networks," *International Journal of Sensor Networks* 1, no. 3 (2006): 229-240.

[65] J. Lin, L. Xie, and W. Xiao, (2009) "Target tracking in wireless sensor networks using compressed Kalman filter," *Int. J. of Sensor Networks*, Vol. 6, pp. 251-262.

[66] X. Xia, Q. Liang, (2010) "Latency-aware and energy efficiency tradeoffs for wireless sensor networks," *Int. J. of Sensor Networks*, Vol. 8, pp. 1-7.

[67] Q. Liang, L. Wang, Q. Ren, (2007) "Fault-tolerant and energy efficient cross-layer design for wireless sensor networks," *Int. J. of Sensor Networks*, Vol. 2, pp. 248-257.

[68] J. H. Ender, (2010) "On compressive sensing applied to radar," *Signal Processing*, Vol 90(5), pp. 1402-1414.

[69] R. Baraniuk, P. Steeghs, "Compressive radar imaging," *IEEE Radar Conference*, April 2007, Boston, MA, pp. 128-133.

[70] J. Romberg, M. Wakin, (2007) "Compressed Sensing: A Tutorial", *IEEE Statistical Signal Processing Workshop*, Madison, Wisconsin.

[71] M. A. Herman, T. Strohmer, "High-resolution radar via compressed sensing," *IEEE Transactions on Signal Processing*, June 2009, Vol. 57(6), pp. 2275-2284.

[72] L. Song, K. Vana, G. Dimitrios, L. Stefano, (2006) "Efficient Information Compression in Sensor Networks", *Int. J. of Sensor Networks*, 2006, Vol. 1, No.3/4 pp. 229 - 240.

[73] Q. Giorgio, M. Riccardo, P. Gianluigi, R. Michele, and Z. Michele, (2012) "Sensing, Compression, and Recovery for WSNs: Sparse Signal Modeling and Monitoring Framework", *IEEE Trans. on Wireless Communications*, Vol. 11, No. 10, pp. 3447-3461.

[74] L. Chong, F. Wu, J. Sun, and C. W. Chen, (2009) "Compressive Data Gathering for Large-Scale Wireless Sensor Networks", *the 15th annual international conference on Mobile computing and networking*, ACM, 2009, pp. 145-156.

[75] M. F. Duarte, S. Sarvotham, D. Baron, M. B. Wakin, and R. G. Baraniuk, (2005) "Distributed compressed sensing for jointly sparse signals," *Thirty-ninth Asilomar conference on signals, systems and computers*, vol. 24, pp. 1537-1541.

[76] S. Kafetzoglou, and S. Papavassiliou, (2011) "Energy-efficient framework for data gathering in wireless sensor networks via the combination of sleeping MAC and data aggregation strategies", *Int. J. of Sensor Networks*, Vol.10, No.1/2, pp.3 - 13.

[77] Z. Charbiwala, S. Chakraborty, S. Zahedi, Y. Kim, M.B. Srivastava, T. He, C. Bisdikian, (2010) "Compressive Oversampling for Robust Data Transmission in Sensor Networks," *IEEE INFOCOM*, pp. 1-9.

[78] S. Aeron, V. Saligrama, and M. Zhao, (2010) "Information theoretic bounds for Compressed Sensing," *IEEE Trans. on Information Theory*, vol. 56, pp. 5111-5130.

[79] A. Majumdar, and R. K. Ward, (2011) "Increasing energy efficiency in sensor networks: blue noise sampling and non-convex matrix completion," *Int. J. of Sensor Networks*, Vol. 9, Number 3-4/2011, pp. 158-169.

[80] Q. Liang, (2006) "Waveform Design and Diversity in Radar Sensor Networks: Theoretical Analysis and Application to Automatic Target Recognition," *Third Ann. IEEE Comm. Soc. on Sensor and Ad Hoc Comm. and Networks, (SECON 06)* vol. 2, pp. 684-689.

[81] H. Shu, and Q. Liang, (2007) "Data Fusion in a Multi-Target Radar Sensor Network," *IEEE Radio and Wireless Symp.*, pp. 129-132.

[82] J. Liang, and Q. Liang, (2006) "orthogonal Waveform Design and Performance Analysis in Radar Sensor Networks," *IEEE Military Comm. Conf.*, pp. 1-6.

[83] J. Liang, and Q. Liang, (2010) "Sense-through-Foliage Target Detection Using UWB Radar Sensor Networks," *Pattern Recognition Letter*, vol. 31, no. 11, pp. 1412-1421.

[84] J. Partan, J.Kurose, and B. N. Levine, "A survey of practical issues in underwater networks," *ACM SIGMOBILE Mobile Computing and Communications Review*, vol. 11, no. 4 (2007): 23-33.

[85] J. Heidemann, W. Ye, J. Wills, A. Syed, Y. Li, "Research Challenges and Applications for Underwater Sensor Networking" *IEEE Wireless Communications and Networking Conference (WCNC) 2006*, Las Vegas, USA, April 2006, vol. 1, pp. 228-235.

[86] M. Erol-Kantarci, H. T. Mouftah, S. Oktug, "A survey of architectures and localization techniques for underwater acoustic sensor networks," *Communications Surveys and Tutorials*, IEEE (2011), 13(3), 487-502.

[87] I. Vasilescu, et al, "Data collection, storage, and retrieval with an underwater sensor network," *ACM, 2005, the 3rd international conference on Embedded networked sensor systems*, pp. 154-165.

[88] I. F. Akyildiz, D. Pompili, and T. Melodia, "Underwater acoustic sensor networks: research challenges," *Ad hoc networks* 3, no. 3 (2005): 257-279.

[89] J. Heidemann, Mi. Stojanovic, and Mi. Zorzi, "Underwater sensor networks: applications, advances and challenges," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 370, no. 1958, doi: 10.1098/rsta.2011.0214, pp. 158-175.

[90] I. F. Akyildiz, D. Pompili, and T. Melodia, "Challenges for efficient communication in underwater acoustic sensor networks," *ACM Sigbed Review 1*, no. 2 (2004): 3-8.

[91] I. F. Akyildiz, D. Pompili, and T. Melodia, "State-of-the-art in protocol research for underwater acoustic sensor networks," *1st ACM international workshop on Underwater networks 2006*, pp. 7-16.

[92] E. M. Sozer, M. Stojanovic, and J. G. Proakis, "Underwater acoustic networks," *IEEE Journal of Oceanic Engineering*, vol. 25, no. 1 (2000): 72-83.

[93] X. Lurton, *An introduction to underwater acoustics: principles and applications*, Springer, 2002.

[94] P. C. Etter, *Underwater acoustic modeling and simulation*, CRC Press, 2013.

[95] T. Melodia, H. Kulhandjian, L.-C. Kuo, and E. Demirors, "Advances in Underwater Acoustic Networking," *Mobile Ad Hoc Networking: Cutting Edge Directions*, Second Edition (2013): 804-852.

[96] M. R.B, S. S. Manvi, "Issues in Underwater Acoustic Sensor Networks," *International Journal of Computer and Electrical Engineering*, Vol.3, No.1, February, 2011, pp. 1793-8163.

[97] D. B. Kilfoyle, and A. B. Baggeroer, "The state of the art in underwater acoustic telemetry," *IEEE Journal of Oceanic Engineering*, vol. 25, no. 1 (2000): 4-27.

[98] D. Pompili, T. Melodia, and Ian F. Akyildiz, "Distributed routing algorithms for underwater acoustic sensor networks," *IEEE Transactions on Wireless Communications*, vol. 9, no. 9 (2010): 2934-2944.

[99] D. Pompili, and Ian F. Akyildiz, "A multimedia cross-layer protocol for underwater acoustic sensor networks," *IEEE Transactions on Wireless Communications*, vol. 9, no. 9 (2010): 2924-2933.

[100] D. Pompili, T. Melodia, and Ian F. Akyildiz, "Three-dimensional and two-dimensional deployment analysis for underwater acoustic sensor networks," *Ad Hoc Networks*, vol. 7, no. 4 (2009): 778-790.

[101] M. Stojanovic, and J. Preisig, "Underwater acoustic communication channels: Propagation models and statistical characterization," *IEEE Communications Magazine*, vol. 47, no. 1 (2009): 84-89.

[102] M. Stojanovic, "On the relationship between capacity and distance in an underwater acoustic communication channel," *ACM SIGMOBILE Mobile Computing and Communications Review* vol. 11, no. 4 (2007): 34-43.

[103] M. Erol-Kantarci, H. T. Mouftah, and S. Oktug. "A survey of architectures and localization techniques for underwater acoustic sensor networks," *IEEE Communications Surveys & Tutorials*, vol. 13, no. 3 (2011): 487-502.

[104] J. Xu, Keqiu Li, and Geyong Min, "Reliable and energy-efficient multipath communications in underwater sensor networks," *IEEE Transactions on Parallel and Distributed Systems*, vol. 23, no. 7 (2012): 1326-1335.

[105] C. Luo, et al. "Compressive data gathering for large-scale wireless sensor networks," *Proceedings of the 15th annual international conference on Mobile computing and networking*, ACM, 2009, pp. 145-156.

[106] S. Feizi, M. Mdard, and M. Effros, "Compressive sensing over networks.," *2010 IEEE 48th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pp. 1129-1136.

[107] M. F. Duarte, S. Sarvotham, D. Baron, M. B. Wakin, and R. G. Baraniuk, "Distributed compressed sensing of jointly sparse signals," *2005 Asilomar Conf. Signals, Sys., Comput*, pp. 1537-1541.

[108] B. Zhang, et al, "Sparse target counting and localization in sensor networks based on compressive sensing," *2011 IEEE INFOCOM*, pp. 2255-2263.

[109] J. Chen, and Q. Liang, "Theoretical Performance Limits for Compressive Sensing with Random Noise," *Accepted by Global Telecommunications Conference (GLOBECOM 2013)*, Dec. 2013, Atlanta, GA, USA.

[110] J. Chen, and Q. Liang, "Rate Distortion Performance Analysis of Nested Sampling and Coprime Sampling," *Accepted to EURASIP Journal on Advances in Signal Processing on 15 January 2014.*

[111] D. Hui and D. Neuhoff, "Asymptotic analysis of optimal fixed-rate uniform scalar quantization," *IEEE Transactions on Information Theory*, vol. 47, no. 3 (2001), pp. 957-977.

[112] J. Chen, Q. Liang, B. Zhang and X. Wu, "Information Theoretic Performance Bounds for Noisy Compressive Sensing," *IEEE ICC 2013 Workshop on Radar and Sonar Networks (RSN)*, June 2013, Budapest, Hungary.

[113] S. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*, Prentice-Hall Signal Processing Series, 1993.

[114] J. Chen, and Q. Liang, "Rate Distortion Performance Analysis of Compressive Sensing," *Global Telecommunications Conference (GLOBECOM 2011)*, Dec. 2011, Houston, TX, USA, pp. 1-5.

[115] Q. Liang, et al, "Opportunistic Sensing in Wireless Sensor Networks: Theory and Application," *2011 IEEE Global Telecommunications Conference* (GLOBECOM 2011), pp. 1-5.

[116] Q. Liang, et al, "Opportunistic Sensing in Wireless Sensor Networks: Theory and Application," *IEEE Transactions on Computers 2013*, Digital Object Indentifier 10.1109/TC.2013.85.

[117] J. Chen, F. Duan, J. Jiang, B. Zhang, Y. Tong, and J. He, "Self-similarity property of acoustic data acquired in shallow water environment," *EURASIP Journal on Wireless Communications and Networking 2013*, no. 1 (2013): 1-6. 2013:91 doi:10.1186/1687-1499-2013-91.

[118] T. S. Rappaport, "Wireless Communications: Principles and Practice," Second Edition, Prentice Hall PTR, pp. 278-302.

[119] H. Mehdi, K. Feher, "FQPSK, Power and Spectral Efficient Family of Modulations for Wireless Communication Systems," *IEEE Vehicular Technology Conference*, June 1994, pp. 1562-1566.

[120] D. Li, "A High Spectrum Efficient Multiple Access Code," *APCC/OECC*, Oct. 1999, pp. 18-22.

[121] M-S. Alouini, A. J. Goldsmith, "Area Spectral Efficiency of Cellular Mobile Radio Systems," *IEEE Transactions on Vehicular Technology*, Vol. 48, No. 4, Jul. 1999, pp. 1047-1066.

[122] J. W. Burns, "Measuring Spectrum Efficiency-The Art of Spectrum Utilisation Metrics," *IEE Two Day Conference. Getting the Most Out of the Radio Spectrum*, London, UK, 24-25 Oct. 2002, pp. 1-3.

[123] P. Pal, P. P. Vaidyanathan, "Nested Arrays: A Novel Approach to Array Processing With Enhanced Degrees of Freedom," *IEEE Transactions on Signal Processing*, Vol. 58, No. 8, Aug. 2010, pp.4167-4181.

[124] P. Pal, P. P. Vaidyanathan, "Coprime Sampling and the MUSIC Algorithm," *Digital Signal Processing Workshop and IEEE Signal Processing Education Workshop*, Jan 2011, pp. 289-294.

[125] P. P. Vaidyanathan, P. Pal, "Sparse Sensing with Co-Prime Samplers and Arrays," *IEEE Transactions on Signal Processing*, wol. 59, No. 2, Feb. 2011, pp. 573-586.

[126] P. Pal, P. P. Vaidyanathan, "A Novel Array Structure for Directions-of-Arrival Estimation with Increased Degrees of Freedom," *Acoustics Speech and Signal Processing*, Mar. 2010, pp. 2606 - 2609.

[127] P. Pal, Piya, and P. P. Vaidyanathan, "Two dimensional nested arrays on lattices," *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 2548-2551.

[128] D. W. Ricker, "Echo Signal Processing," Springer, ISBN 1-4020-7395-X, pp. 23-26.

[129] P. Stoica, R. L. Moses, "Introduction to Spectral Analysis," Upper Saddle River, New Jersey, 1997, Prentice Hall, 1 edition, pp. 1-13.

[130] W. C. Lindsey, "Error probabilities for Rician fading multichannel reception of binary and n-ary signals," *IEEE Transactions on Information Theory*, Vol. 10, Issue 4, Oct. 1964, pp. 339 - 350.

[131] R. C. Robertson, J. F. Sheltry, "Multiple tone interference of frequency-hopped noncoherent MFSK signals transmitted over Ricean fading channels," *IEEE Transactions on Communications*, Vol. 44, Issue 7, Jul. 1996, pp. 867-875.

[132] D. Kirachaiwanich, Q. Liang, "The Combined-Jamming Model for IEEE 802.11 FH/MFSK Networks," *European Transactions on Telecommunications*, Vol. 22, Issue 1, Jan. 2011, pp. 14-24.

[133] T. T. Tjhung, C. C. Chai, "Multitone Jamming of FH/BFSK in Rician Channels," *IEEE Transactions on Communications*, Vol. 47, No. 7, Jul. 1999, pp. 974-978.

[134] Z. Yu, T. T. Tjhung, C. C. Chai, "Independent Multitone Jamming of FH/MFSK in Rician Channels," *IEEE Transactions on Communications*, Vol. 49, No. 11, Nov. 2001, pp. 2006-2015.

[135] J. Chen, Q. Liang, J. Paden, and P. Gogineni, "Compressive sensing analysis of Synthetic Aperture Radar raw data," *2012 IEEE International Conference on Communications (ICC)*, 2012, pp. 6362-6366.

[136] J. Chen, and Q. Liang, "Efficient Sampling for Radar Sensor Networks," *accepted for publishing by International Journal of Sensor Networks 2013.*

[137] J. Chen, and Q. Liang, "Rate Distortion Performance Analysis of Nested Sampling and Coprime Sampling," *Accepted to EURASIP Journal on Advances in Signal Processing on 15 January 2014.*

[138] J. Chen, Q. Liang, B. Zhang, X. Wu, "Spectrum Efficiency of Nested Sparse Sampling and Co-Prime Sampling," *EURASIP Journal on Wireless Communications and Networking 2013*, no. 1 (2013): 47.

[139] J. Chen, Q. Liang, J., and H.-A. Choi, "Spectrum Efficiency of Nested Sparse Sampling," *Wireless Algorithms, Systems, and Applications, Springer Berlin Heidelberg*, 2012, pp. 574-583.

[140] J. Chen, Q. Liang and J. Wang, "Secure transmission for big data based on nested sampling and coprime sampling with spectrum efficiency," *Security and Communication Networks*, Wiley, Security Comm. Networks (2013), DOI: 10.1002/sec.785.

[141] C. Dill, "Foliage Penetration (Phase II) Field Test: Narrowband versus Wideband Foliage Penetration," Final Report of Contract Number F41624-03-D-7001/04, July 2005 to Feb 2006.

[142] Ma. Bryant, "LTE users consume more data than those on 3G, but its big data plans that drive use more than speed," *The Next Web*, 3 Jan'13.

[143] *5G: A Technology Vision*, by Huawei.

[144] *Big Data LTE. What is Big Data in Telecommunications?* Whitepaper, EMPIRIX.

[145] Martin J. Wainwright, "Information-theoretic limits on sparsity recovery in the high-dimensional and noisy setting," *IEEE Transactions on Information Theory*, vol. 55, no. 12 (2009), pp. 5728-5741.

[146] W. Wang, Martin J. Wainwright, and K. Ramchandran, "Information-theoretic limits on sparse signal recovery: Dense versus sparse measurement matrices," *IEEE Transactions on Information Theory*, vol. 56, no. 6 (2010), pp. 2967-2979.

[147] W. Dai, and O. Milenkovic, "Information theoretical and algorithmic approaches to quantized compressive sensing," *IEEE Transactions on Communications*, vol. 59, no. 7 (2011), pp. 1857-1866.

[148] C. Arthur (18 July 2013). "Nokia revenues slide 24% but Lumia sales rise offers hope," *The Guardian*, Retrieved 19 July 2013.

[149] S. Butt, and J. G. Phillips, "Personality and self reported mobile phone use," *Computers in Human Behavior 24*, no. 2 (2008): 346-360.

[150] A. Beydokhti, R. Hassanzadeh, and B. Mirzaian, "The Relationship between Five Main Factors of Personality and Addiction to SMS in High School Students," *Current Research Journal of Biological Sciences* 2012, 4(6): 685-689.

[151] G. Chittaranjan, J. Blom, and D. Gatica-Perez, "Mining large-scale smartphone data for personality studies," *Personal and Ubiquitous Computing 17*, no. 3 (2013): 433-450.

[152] W. Lane, and C. Manner, "The impact of personality traits on smartphone ownership and use," *International Journal of Business and Social Science* 2, no. 17 (2011): 22-28.

[153] W. Lane, and C. Manner, "The Influence of Personality Traits on Mobile Phone Application Preferences," *Journal of Economics and Behavioral Studies*, Vol. 4, No. 5, May 2012, pp. 252-260.

[154] R. de Oliveira, A. Karatzoglou, P. Concejero Cerezo, A. Armenta Lopez de Vicuna, and N. Oliver, "Towards a psychographic user model from mobile phone usage," *CHI'11 Extended Abstracts on Human Factors in Computing Systems*, ACM, 2011, pp. 2191-2196.

[155] R. E. Neapolitan, *Learning bayesian networks*, Upper Saddle River: Pearson Prentice Hall, 2004.

[156] J. Pearl, (2000), *Causality: Models, Reasoning, and Inference*, Cambridge University Press, ISBN 0-521-77362-8.

[157] O. Pourret, P. Naim, and B. Marcot, *Bayesian networks: a practical guide to applications*, Wiley, 2008, ISBN: 978-0-470-06030-8.

[158] S. J. Russell, P. Norvig, *Artificial Intelligence: A Modern Approach*, 3rd Edition, Pearson Education Inc., Prentice Hall, Upper Saddle River, New Jersey, 2010.

[159] S. D. Gosling, P. J. Rentfrow, and W. B. Swann Jr., "A very brief measure of the Big-Five personality domains," *Journal of Research in personality* 37, no. 6 (2003): 504-528.

[160] R. R. McCrae, and O. P. John, "An introduction to the five-factor model and its applications," *Journal of personality* 60, no. 2 (1992): 175-215.

[161] J. C. McElroy, A. R. Hendrickson, A. M. Townsend, and S. M. DeMarie, "Dispositional factors in internet use: personality versus cognitive style," *MIS Quarterly* (2007): 809-820.

[162] D. Watson, L. A. Clark, (1997) "Extraversion and its positive emotional core," *Handbook of Personality Psychology*, Academic Press, San Diego, 767C793.

[163] P. T. Costa, R. R. McCrae, D. A. Dye, (1991) "Facet scales for agreeableness and conscientiousness: A revision of the NEO personality inventory," *Personality and Individual Differences*, 12, no. 9 (1991): 887-898.

[164] P. T. Costa, R. R. McCrae, *Revised neo personality inventory (neo pi-r) and neo five-factor inventory (neo-ffi)*, Odessa, FL: Psychological Assessment Resources, 1992.

[165] M. R. Barrick, M. K. Mount, 1991, "The big five personality dimensions and job performance: A meta-analysis," *Personnel Psychology*, 441, 1-26.

[166] J. G. Phillips, S. Butt, and A. Blaszczynski, "Personality and self-reported use of mobile phones for games," *CyberPsychology & Behavior* 9, no. 6 (2006): 753-758.

[167] A. Ehrenberg, S. Juckes, K. M. White, and S. P. Walsh, "Personality and self-esteem as predictors of young people's technology use," *CyberPsychology & Behavior* 11, no. 6 (2008): 739-741.

[168] S. Devaraj, R. F. Easley, and J. M. Crant, "Research note - how does personality matter? Relating the five-factor model to technology acceptance and use," *Information Systems Research* 19, no. 1 (2008): 93-105.

[169] D. Tse, P. Viswanath, *Fundamentals of Wireless Communication*, ISBN 0521845270, Cambridge University Press, 2005.

[170] R. J. Larsen, D. M. Buss, *Personality Psychology: Domains of Knowledge About Human Nature*, Fourth Edition, McGraw Hill, 2010, ISBN: 978-0-07-337068-2.

Biographical Statement

Junjie Chen received her B.S. degree in Telecommunications Engineering from Xidian University, China, in 2007, and M.S. degree in Signal and Information Processing from Beijing University of Posts and Telecommunications, China, in 2010. From 2010, she works toward the Ph.D. degree in Electrical Engineering at the University of Texas at Arlington. Her research interests include sparse sensing in Big Data, wireless communications and networking, wireless sensor networks, signal processing, and information theory.