SIGNAL RECONSTRUCTION THROUGH DISTRIBUTED COMPRESSIVE DATA GATHERING IN WIRELESS SENSOR NETWORK

A DISSERTATION

Submitted in partial fulfillment of the
requirements for the award of the degree
of

INTEGRATED DUAL DEGREE

in

ELECTRONICS AND COMMUNICATION ENGINEERING
(With Specialization in Wireless Communication)

By

CHARUL AGRAWAL

M.G. CENTRAL LIBRARY
ACC No. G21478
Date 17/8/12
I.I.T. ROORKEE

DEPARTMENT OF ELECTRONICS AND COMPUTER ENGINEERING
INDIAN INSTITUTE OF TECHNOLOGY ROORKEE
ROORKEE-247 667 (INDIA)

JUNE, 2012

Candidate's Declaration

I hereby declare that the work, which is presented in this dissertation report entitled, **"SIGNAL RECONSTRUCTION THROUGH DISTRIBUTED COMPRESSIVE DATA GATHERING IN WIRELESS SENSOR NETWORK"** towards the partial fulfillment of the requirements for the award of the degree of **MASTER OF TECHNOLOGY** with specialization in **Wireless Communication**, submitted in the Department of Electronics and Computer Engineering, Indian Institute of Technology Roorkee, Roorkee (India) is an authentic record of my own work carried out during the period from June 2011 to June 2012, under the guidance of **Dr. Debashis Ghosh, Associate Professor, Department of Electronics and Computer Engineering, Indian Institute of Technology Roorkee.**

I have not submitted the matter embodied in this dissertation for the award of any other Degree or Diploma.

Date: 14.06.2012

Place: Roorkee

CHARUL AGRAWAL

CERTIFICATE

This is to certify that the above statement made by the candidate is correct to the best of my knowledge and belief.

Date: 14.06.2012

Place: Roorkee

Dr. Debashis Ghosh

Associate Professor

E&C Department

IIT Roorkee

Acknowledgement

During the course of the past five years a lot many people have helped me in a lot many ways. I owe this dissertation to their continued support and blessings. Each of their contribution means a lot to me and I will try to fit within this page to the best of my capability a token of gratitude for their efforts.

First and foremost I thank my guide, Dr. Debashis Ghosh for all his encouragement, support, positive criticism and most of all his patience. I am greatly indebted to him for his constant motivation and invaluable advice. He was always available whenever I approached him for the whole year clearing smallest of my doubts.

I would also like to thank all the faculty of E&CE department who helped me understand the basics of Electronics and Communication Engineering in these years. Special thanks for the staff of Signal Processing Lab, Indian Institute of Technology Roorkee for providing such a well maintained laboratory that always provided a congenial atmosphere.

Next, I thank all my friends and batch-mates who immensely helped me in their own little way.

And above all my heartfelt gratitude is for my parents Shri Sanjay Agrawal and Smt. Sushma Agrawal, and my brother Rahul Agrawal. They have stood by me at each and every moment and made me believe in myself when I had doubts over it. Finally I thank God for his blessings.

Abstract

Wireless sensor network (WSN) is an emerging technology with unprecedented opportunities for wide variety of applications in the present world. The essential task in many applications of sensor networks is to extract relevant information about the sensed data and deliver it with a desired fidelity to a central collection point or sink. WSNs, or more specifically each sensor node, are resource constrained. They have limited power supply, bandwidth for communication, processing speed, and memory space which make the reduction of communication critical to increase the network's performance and lifetime. Data compression is one effective method to utilize limited resources of WSNs. Compressive Sensing (CS) is a novel data compression technique that exploits the inherent correlation in the input data to compress it by means of quasi-random matrices. Distributed Compressed Sensing (DCS) is an extension of CS to multiple-signal case. Since sensors presumably observe related phenomena, the ensemble of signals they acquire may be expected to possess some joint structure, or inter-signal correlation, in addition to the intra-signal correlation in each individual sensor's measurements. DCS enables new distributed coding algorithms that exploit both intra- and inter-signal correlation structures.

Also, nodes close to the sink transmit more data and consume more energy than those at the peripheral of the network. The unbalanced energy consumption has a major impact on network lifetime. Compressive data gathering (CDG) leverages compressive sensing (CS) principle to efficiently reduce communication cost and prolong network lifetime for large scale monitoring sensor networks by balancing the energy consumption and reducing the transmissions. With the recent developments in DCS reducing the communication costs in sensor networks, we propose Distributed Compressive Data Gathering (DCDG) to further reduce the communication costs in data gathering and number of measurements in WSNs.

Contents

List of Figures

List of Abbreviations

CS	*Compressive Sensing*
CDG	*Compressive Data Gathering*
DCDG	*Distributed Compressive Data Gathering*
DCS	*Distributed Compressed Sensing*
FOCUSS	*FOCal Underdetermined System Solver*
JPEG	*Joint Photographic Experts Group*
JSM	*Joint Sparsity Model*
MAP	*Maximum aposteriori*
MMSE	*Minimum Mean Square Error*
MMV	*Multiple Measurement Vectors*
MPEG	*Moving Picture Experts Group*
MSE	*Mean Square Error*
OMP	*Orthogonal Matching Pursuit*
OSGA	*One-Step Greedy Algorithm*
PCA	*Principal Component Analysis*
RIP	*Restricted Isometry Property*
SOMP	*Simultaneous Orthogonal Matching Pursuit*
SMV	*Single Measurement Vector*
WSN	*Wireless Sensor Network*

1 Introduction

Wireless sensor network (WSN) is an emerging technology with wide variety of applications in the present world. It is composed of a large number of sensor nodes spatially distributed over a region of interest that can sense the environment in various modalities, such as acoustic, seismic, thermal and infra-red. Cheap, smart devices with multiple onboard sensors, networked through wireless links and the Internet are deployed in large numbers. It provides unprecedented opportunities for instrumenting and controlling homes, cities, and the environment, disaster relief, border monitoring, and surveillance in battlefield scenarios. A sensor is generally equipped with at least a power supply, sensing unit, processing unit to process the sensed data, and transmitter-receiver unit. All nodes are connected by radio frequency, infrared, or other wireless medium. The data collected by nodes traverse among the nodes in wireless medium. In order to realize WSNs, peer-to-peer network techniques are widely used so that it allows direct communication between any two nodes. If two devices cannot communicate directly, other intermediate nodes relay data packets from the source node to the destination node. This is called multi-hop routing. The sensors coordinate among themselves to form a communication network such as a single multi-hop network or a hierarchical organization with several clusters and cluster heads. Because of their peer-to-peer communication style, no centralized point, which controls a network formation like a base station for a cellular system, is required for the network. Since no fixed infrastructure is necessary for WSNs, a network is constructed inexpensively. Also, nodes may be added to and removed from the network easily. On the other hand, the network topology in a WSN may change drastically since nodes can be added and removed easily. The sensors periodically sense the data, process it and transmit it to (usually) distant destination, termed as the fusion centre (FC) or sink. The sink may be connected to the outside world through Internet or satellite.

The essential task in many applications of sensor networks is to extract relevant information about the sensed data and deliver it with a desired fidelity to the sink. A large number of sensor nodes are often deployed to the locations where it is hard to access. It is not practical to perform maintenance operations, such as changing batteries, on deployed sensor nodes. Because of the above reasons, WSNs, or more specifically the sensor nodes, are resource constrained. They have limited power supply, bandwidth for communication, processing speed, and memory space. In many sensor networks, and in particular battery-powered ones, these factors make the reduction of communication critical to increase the network's performance and lifetime. Data compression is one effective method to utilize limited resources of WSNs. Compressing the sensed data will reduce the power consumption due to processing and transmitting data in each node, and thus extend the life time of sensor network. Also, by reducing data size less bandwidth is required for sending and receiving data. Our objective is to measure large data sets with high accuracy through the collection of a small number of readings. However, most existing data compression algorithms are not feasible for WSNs due to the size of the algorithms and processing speed of the nodes.

It is possible to avoid the transmission of any "redundant" information if the sensors could communicate with one another. A number of distributed coding algorithms have been developed that involve collaboration amongst the sensors. However, this increases communication overhead. Slepian-Wolf coding has the distinct advantage that the sensors need not collaborate while encoding their measurements thereby saving valuable communication overhead. In the Slepian-Wolf framework for lossless distributed coding [1, 2], the availability of correlated side information at the collection point / decoder enables each sensor node to communicate losslessly at its conditional entropy rate rather than at its individual entropy rate. Unfortunately, however, most existing coding algorithms [2] exploit only inter-signal correlations and not intra-signal correlations. To date there has been only limited progress on distributed coding of so-called "sources with memory." The direct implementation for such sources would require huge lookup tables. Furthermore, approaches combining pre- or post-processing of the data to remove intra-signal correlations combined with Slepian-Wolf coding for the inter-signal correlations appear to have limited applicability. This entails the design of distributed algorithms for the joint gathering and compression of data and the exploitation, at the sink, of signal processing techniques for the approximation of the signal in space and time. The area of communication and protocol design for Wireless Sensor Networks (WSNs) has been widely researched in the past few years. An important research topic which needs further investigation is in-network aggregation and data management to increase the efficiency of data gathering solutions (in terms of energy cost) while being able to measure large amount of data with high accuracy. Before going into these details, first let us understand the concept of Compressive Sensing (CS) [3]–[5] which is a novel data compression technique that exploits the inherent correlation in the input data to compress it by means of quasi-random matrices.

1.1 Compressive Sensing

A new framework for single-signal sensing and compression has developed recently under the rubric of Compressive Sensing (CS). CS builds on the ground-breaking work of Candes, Romberg, and Tao [3] and Donoho [4], who showed that if a signal has a sparse representation in one basis then it can be recovered from a small number of projections onto a second basis that is incoherent with the first. A sparse signal is a signal with a very few non-zero coefficients/values in its representation. A large number of real and generated signals are either sparse in their original form or may be represented as a sparse signal in transform domain. Sparse signals are present everywhere. The dogma of signal processing maintains that a signal must be sampled at the Nyquist rate at least twice its bandwidth in order to be represented without error. However, in practice, we often compress the data soon after sensing, trading off signal representation complexity (bits) for some error (consider JPEG image compression in digital cameras, for example). Clearly, this is wasteful of valuable sensing/sampling resources. In compressive sensing, the signal is sampled (and simultaneously compressed) at a greatly reduced rate.

Compressive Sensing problem [3]-[6] may be stated as recovery of vector $x \in R^{N}$ from the measurement vector $y \in R^{M}$ such that $M \ll N$ and x is a K-sparse vector (K out of N coefficients of x are nonzero in some basis ψ, $K < M$) such that $y = \phi x$ where ϕ is a $M \times N$ measurement matrix. The M rows of ϕ may be considered as basis vector. To ensure the recoverability of any such x, the measurement matrix should satisfy the conditions of incoherence and restricted isometry property (RIP) [7]. It is also proved that the measurement matrix whose coefficients are chosen randomly can satisfy these conditions with high probability. Using such a matrix it is possible, with high probability, to recover any signal that is

K-sparse in the basis ψ from its image under ϕ, where ψ and ϕ are incoherent. For signals that are not K-sparse but compressible, meaning that their coefficient magnitudes decay exponentially, there are tractable algorithms that reconstruct signals with error not more than a multiple of the error of the best K-term approximation of the signal.

Many reconstruction algorithms have been developed based on minimizing the error coupled with the sparse constraint. The algorithms mainly fall in two categories: convex relaxation and greedy pursuits. Examples of convex relaxation include interior point methods like l_1 minimization, Primal-Dual interior methods for convex objectives (PDCO) using conjugate gradients and Iteratively Reweighted Least Square (IRLS). Examples of Greedy pursuits include Orthogonal Matching Pursuit (OMP), Stagewise OMP (StOMP), and Iterative hard thresholding method (IHT). Since CS is comparatively a new field, a lot of research till date is directed at theoretical aspects of CS and at improving CS recovery algorithms to operate faster with minimum possible number of measurements. Techniques from other fields are being invoked for finding newer and better methods for CS Recovery. A large number of CS recovery algorithms taking advantage of structure present in sparse signals were developed in the last few years to achieve the twin objectives of speed and minimizing the number of measurements required [8]. Carin et al. [9] extended Bayesian framework to solve for CS problem to obtain maximum aposteriori (MAP) estimate for the sparse signal based on the measurements/observations.

In [10], Principal Component Analysis (PCA) was used to find transformations (ψ) that sparsify the signal, which are required for CS to retrieve, with good approximation, the original signal from a small number of samples.

1.2 Distributed Compressed Sensing

Distributed Compressed Sensing (DCS) [11] is an extension of CS to multiple-signal case. Since sensors presumably observe related phenomena, the ensemble of signals they acquire may be expected to possess some joint structure, or inter-signal correlation, in addition to the intra-signal correlation in each individual sensor's measurements. For example, imagine a microphone network recording a sound field at several points in space. The time-series acquired by a given sensor generally have considerable intra-signal (temporal) correlation and might be sparsely represented in a local Fourier basis. In addition, since all microphones listen to the same sources the ensemble of time-series acquired at all sensors might have considerable inter-signal (spatial) correlation.

Distributed compressed sensing (DCS) enables new distributed coding algorithms that exploit both intra- and inter-signal correlation structures. The DCS theory rests on a concept termed as joint sparsity of a signal ensemble. Each sensor independently encodes its signal by projecting it onto another incoherent basis (such as a random one) and then transmits just a few of the resulting coefficients to a collection point. Under the right conditions, a decoder at the collection point can jointly reconstruct all the signals precisely. This allows WSNs to save on the communication costs involved in exporting the ensemble of signals to the collection point. This entails the design of distributed algorithms for the joint gathering and compression of data and the exploitation of signal processing techniques at the sink for the approximation of the signal in space and time. Baron et al. [12] have studied three joint sparsity models (JSMs) and proposed tractable algorithms, namely One step greedy algorithm (OSGA), DCS-SOMP (Simultaneous Orthogonal Matching Pursuit) and Alternating Common and Innovation Estimation (ACIE), for joint recovery of signal ensembles from incoherent projections, and characterized theoretically and empirically the number of measurements per sensor required for accurate reconstruction.

In [13], Cotter et al. addressed the problem of finding sparse solutions to an underdetermined system of equations when there are multiple measurement vectors having the same, but unknown, sparsity structure. It extends two classes of algorithms, Orthogonal Matching Pursuit (OMP) and FOCal Underdetermined System Solver (FOCUSS), to the multiple measurement vectors (MMV) case –M-OMP and M-FOCUSS, so that they may be used in applications such as neuromagnetic imaging, where multiple measurement vectors are available, and solutions with a common sparsity structure must be computed.

1.3 Compressive Sensing in Communication Networks

In 2004, CS was first proposed for efficient storage and compression of digital images, which show high space correlation. In the following few years, CS has expressed many advantages and its application has prevailed in these fields, with the development of plenteous novel techniques, such as developing simpler, smaller, and cheaper digital cameras, novel analog-to-digital (A/D) converter architectures and so on. Recently, CS has been earning more and more interests in the area of wireless communication networks and a plenty of researches focused on how to utilize CS efficiently in this area have been carried out.

We mention the researches in the four layers according to the OSI (Open Systems Interconnection) network model, respectively [14].

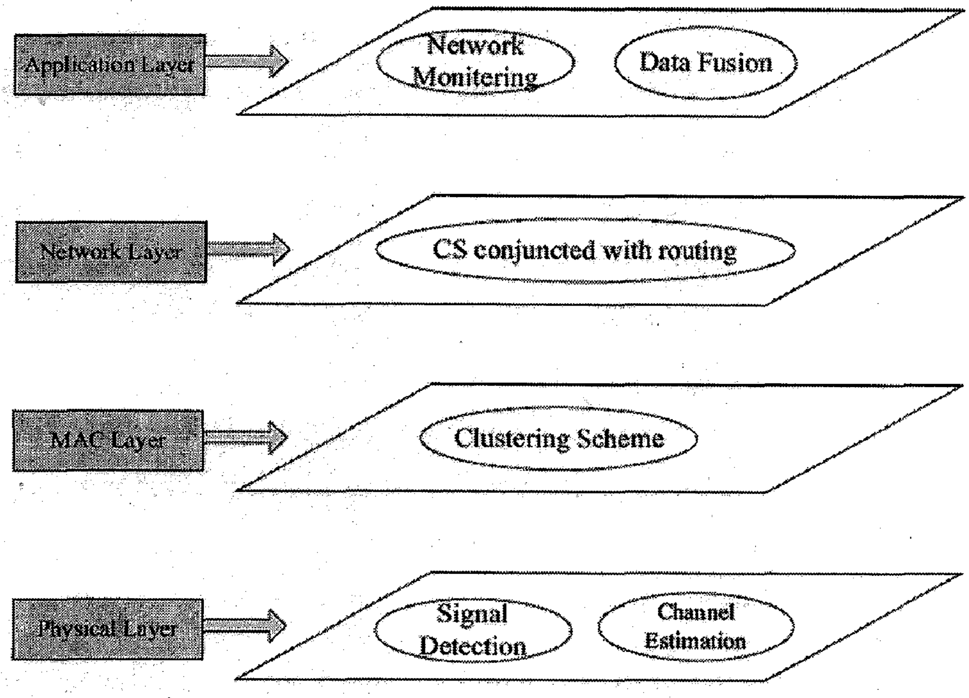

Figure 1.1 Applications of CS in communication networks according to OSI network model [14].

In the physical layer, CS has been used in data detection and channel estimation, especially in ultra-wideband (UWB) communications, underwater acoustic (UWA) communications and Cognitive Radio (CR). In particular, CS can be used in the identification of Frequency Hopping signals, detection in On-Off Random Access Channels, Spectrum Sensing in Cognitive Radio, and Sparse Event Detection in WSNs. However, in case the number of channels is not large enough in signal detection and channel estimation, the requirement of sparsity cannot be guaranteed.

In MAC layer, CS can be used in WSNs to reduce measurement cost by minimizing the number of measurements. The studies are focused on addressing the problem of high costs caused by dense measurement when CS is utilized. Using CS in multi-hop networks is a serious problem as each measurement in CS is a linear combination of many (or all) samples of the signal to be reconstructed, which may result in significant transport costs even though the number of measurements has been minimized. Cluster-based technique is used in WSN to solve this problem. However, design of efficient clusters is still an open problem as it influences the efficiency of reconstruction of data in CS.

In network layer, CS is very promising for jointly acquiring and aggregating data from distributed data sources in a multi-hop WSN. The fact that the number of sensing nodes in WSNs is very huge, while the information generated by the nodes is almost the same indicates that the information of the whole network is compressible. Based on this, CS shows its great potential in joint data compression and transmission without the knowledge of correlation properties of the input signal over the entire network. Also, CS makes the reconstruction of all sensor readings of the network possible, using much fewer transmissions than traditional routing or aggregation schemes, thereby increasing the efficiency of data gathering solutions.

CS is utilized in data fusion and network monitoring in the application layer. It is valuable to apply CS to network monitoring, addressing the problem of efficient end-to-end network monitoring of the path metrics in large-scale wireless communication networks. Also, CS shows great advantages in data fusion. Different from the other decentralized compression strategies, such as Slepian-Wolf coding, which need a prior knowledge of the correlations between data at different nodes, CS needs no prior knowledge, which has been of increasing interests recently [11], [15], and CS offers two highly desirable advantages for networked data analysis: one is decentralized, meaning that distributed data can be encoded without a central controller; the other is universal, in the sense that sampling does not require a priori knowledge or assumptions about the data. Several approaches [15]-[17] in the action of networked data compression are presented in the recent researches. However, a lot of computing is required at the fusion center for the reconstruction of the data.

1.4 Statement of the Problem

Most of the existing studies conducted so far focuses on how to achieve the maximum utilization of limited sensor resources. One field of resource utilization studies for sensor networks is data compression. Researchers seek the optimal way to compress the sensing data. In this dissertation we

1) Implement M-FOCUSS and DCS-SOMP on the real world signals gathered from WSNs to reduce the number of measurements.
2) Recover real signal data through joint Distributed Compressed Sensing (DCS) and Principal Component Analysis (PCA).
3) Propose Distributed Compressive Data Gathering scheme based on Compressive Data Gathering and Distributed Compressed Sensing for efficient data aggregation, hence reducing number of measurements and communication cost in WSN.
4) Implement the above scheme on the real world data gathered from WSNs.

1.5 Organization of the Dissertation

In the next chapter, theory behind CS is discussed. The concept of sparsity on which CS is based upon, problem formulation to find sparse representation and various algorithms to solve the under-determined system of equations to find sparse representation are mentioned. Also, motivation behind CS and the exact CS problem is presented. The detailed theory includes conditions required to be satisfied by measurement matrix to ensure recoverability of any K-sparse signal. Principal Component Analysis (PCA) used to dynamically find transformation matrix for signals with time varying correlation is also discussed.

In Chapter 3, we discuss the concept of Distributed Compressed Sensing which uses the temporal and spatial correlation to reduce the number of measurements for joint recovery of signals in sensor networks and its applications. Different joint sparsity models and their recovery algorithms have been discussed. Also, M-FOCUSS used for recovery from multiple measurement vectors has been used for JSM-2 model. Simulation results of the implementation of M-FOCUSS and JSM-2 recovery algorithms OSGA and DCS-SOMP are shown. Also, M-FOCUSS and DCS-SOMP are compared. Simulation results of recovery of signals using joint DCS and PCA are also shown.

Chapter 4 mentions different data gathering schemes in WSN. Compressive Data Gathering (CDG) proposed by Luo et al. [16] is discussed in detail. It leverages compressive sampling (CS) principle to efficiently reduce communication cost and prolong network lifetime. IR-CDG is another scheme for data gathering with even less number of measurements in WSN. Simulation results of CDG and IR-CDG are shown.

We propose Distributed Compressive Data Gathering (DCDG) to further reduce the communication cost and measurements in WSN. This is detailed in chapter 4. Simulation results for DCDG are shown for real world signals.

Chapter 5 concludes the dissertation thesis mentioning the results and area of further studies.

2 Compressive Sensing

According to Nyquist/Shannon sampling theory, signals, images, videos, and other data can be exactly recovered from a set of uniformly spaced samples taken at the so-called Nyquist rate of twice the highest frequency present in the signal of interest. Capitalizing on this discovery, much of signal processing has moved from the analog to the digital domain. Digitization has enabled the creation of sensing and processing systems that are more robust, flexible, and cheaper and, consequently, more widely used than their analog counterparts. As a result, in recent years, the amount of data generated by the sensing systems has grown drastically. For some of the important and emerging applications, the required Nyquist rate is so high that a large number of samples are generated and stored for efficient representation of the generated data. Along with the large amount of data generated, it may be either costly or physically impossible to achieve such high acquisition rates in some fields like imaging, video, medical imaging, remote surveillance, and spectroscopy.

To deal with the first challenge of large amount of data generated, we depend on compression. Also, data often need to be transmitted through a channel or a network such as in wireless sensor networks. Prior to transmission, it is desirable to compress the data for efficient usage of storage resources and/or bandwidth of the communication channels. Compression is basically finding the most concise representation of a signal within level of acceptable distortion. One of the most popular compressing techniques is Transform coding which relies on finding some basis which makes the signal sparse or compressible. By a sparse representation, we mean that for a signal of length N, we can represent it with $K \ll N$ nonzero coefficients; by a compressible representation, we mean that the signal is well-approximated by a signal with only K nonzero coefficients. This is attained by preserving only the largest coefficients of the transformed signal, which contain most of the information, without much numerical or perceptual loss. This process is the basis behind many compression schemes like JPEG, JPEG2000, MPEG and MP3 standards. Since in transform coding most of the low information carrying small coefficients are discarded, acquisition of so many samples and then calculating their equivalent representation in transform domain seems to be a big loss. This raises the question: "Why go to so much effort to acquire **all** the data when **most** of what we get will be thrown away? Can we not just **directly measure** the part that will not end up being thrown away?" This became the central idea behind CS: rather than sampling the signal at a high rate and then compressing it, we would like to find ways to directly sense the signal in compressed form. The term CS was coined in the separate works of Candes, Romberg and Tao [3] and Donoho [4], who showed that a finite-dimensional signal having a sparse representation can be recovered from a set of linear, nonadaptive measurements. Moreover, the acquisition does not require knowledge of the signal/image to be acquired in advance-other than knowledge that the data will be compressible. The design of these measurement schemes and their extension to practical data models and acquisition systems are central challenges in the field of CS.

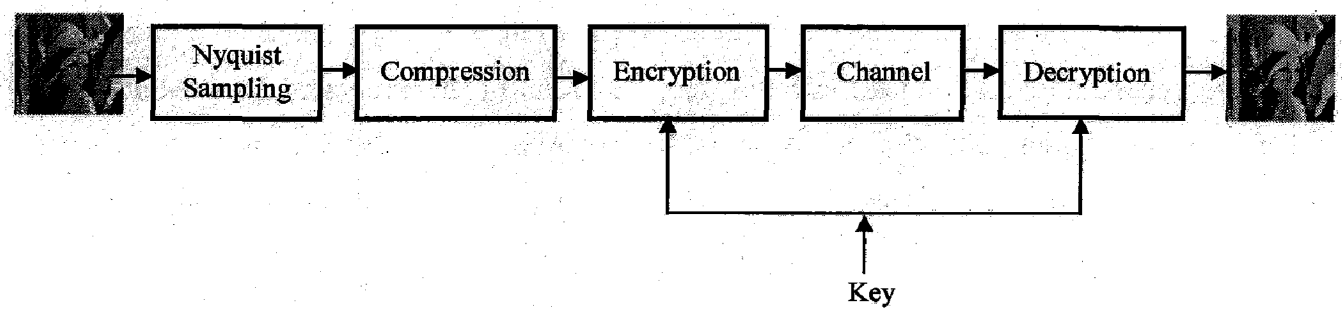

Figure 2.1 Flow diagram of a conventional sampling, compression and encryption scheme.

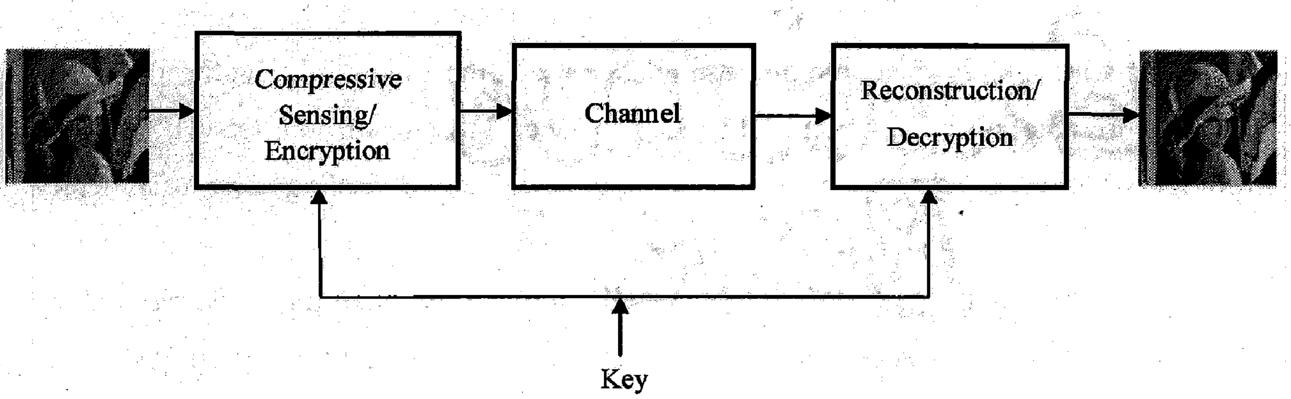

Figure 2.2 Flow diagram of compressive sensing for unified sampling, compression and encryption.

Compressive Sensing differs from classical sampling theory in three major aspects [18]:

1. Sampling theory typically considers infinite length, continuous-time signals. In contrast, CS is a mathematical theory focused on measuring finite-dimensional vectors in R^N.

2. Rather than sampling the signals at specific points in time, CS systems typically acquire measurements in the form of inner products between the signal and more general test functions.

3. The two differ in the manner in which they deal with signal recovery. In the Nyquist- Shannon framework, signal recovery is achieved through sinc interpolation, In CS, signal recovery is achieved using highly nonlinear methods.

2.1 Sparsity

Before moving to the theory of CS, let us first understand the concept of sparsity and how to find sparse representations. Few definitions should be understood first.

Definition 1. l_p norm: In the case of a discrete, finite domain, signals can be viewed as vectors in an N-dimensional Euclidean space, denoted by R^N. *lp* norms for such signals are defined for $p \in [1, \infty]$ as

$$\|x_p\| = \begin{cases} (\sum_{i=1}^{p} |x_i|^p)^{1/p}, & p \in [1, \infty) \\ \max_{i=1,2,\ldots,n} |x_i|, & p = \infty \end{cases}$$

Definition 2. Support: Support of a vector θ is defined as the locations of nonzero entries in the vector θ. Mathematically,

$$\|\theta\|_0 = |supp(\theta)|, \qquad where\ supp(\theta) = \{i : \theta_i \neq 0\}$$

Definition 3. Bases and Frames: A set $\{\psi_i\}_{i=1}^{N}$ is called a basis for R^N if the vectors span R^N and are linearly independent. It implies that if we let ψ denote the N×N matrix with columns given by ψ_i and let θ denote the length-N vector with entries θ_i, then we can represent this relation more compactly as

$$x = \psi\theta \tag{2.1}$$

It is often more useful to generalize the concept of a basis to allow sets of possibly linearly dependent vectors, resulting in what is known as a frame. A frame is a set of vectors $\{\psi_i\}_{i=1}^{N} \in R^d$, d < N corresponding to a matrix $\psi \in R^{d \times N}$, such that for all vectors $x \in R^d$,

$$A\|x\|_2^2 \leq \|\psi^T x\|_2^2 \leq B\|x\|_2^2 \quad with\, 0 < A \leq B < \infty.$$

Frames can provide richer representation of data due to their redundancy. For a given signal x, there exist, infinitely many coefficient vectors θ such that $x = \psi\theta$. In order to obtain a set of feasible coefficients we exploit the dual frame $\tilde{\psi}$. Specifically, any frame satisfying

$$\psi\tilde{\psi}^T = \psi^T\tilde{\psi} = I$$

is called an alternate frame. The particular choice $\tilde{\psi} = (\psi\tilde{\psi}^T)^{-1}\psi$ is referred to as the canonical dual frame. It is also common for a basis or frame to be referred to as a dictionary or overcomplete dictionary respectively, with the dictionary elements being called atoms.

Definition 4. K-sparse signal: A signal is K-sparse when it has at most K nonzero coefficients i.e.$\|x\|_0 \leq$ K. And let

$$\Sigma_K = \{x : \|x\|_0 \leq K\}$$

denote the set of all K-sparse signals. We say a signal x is K-sparse in the basis or frame ψ if there exists a vector $\theta \in R^N$ with only K ≪ N nonzero entries such that $x = \psi\theta$. We call the set of indices corresponding to the nonzero entries the support of θ and denote it by supp(θ).

Definition 5. Compressible Signals: Compressible signals are those which can be well approximated by a sparse signal. This well-approximation can be quantified by calculating the error incurred by approximating a signal x by $\tilde{x} \in \Sigma_K$ as

$$\sigma_K(x)_p = \min_{\tilde{x}\in\Sigma_K} \|x - \tilde{x}\|_p$$

Another way to think about compressible signals is to consider the rate of decay of their coefficients. For many important classes of signals there exist bases such that the coefficients obey a power law decay, in which case the signals are highly compressible. Specifically, if $x = \psi\theta$ and we sort the coefficients θ_i such that $|\theta_1| \geq |\theta_2| \geq \cdots \geq |\theta_N|$, then we say that the coefficients obey a power law decay if there exist constants $C_1, q > 0$ such that

$$|\theta_i| \leq C_1 i^{-q}.$$

2.1.1 Finding Sparse Representations

It is useful to determine whether a signal has a sparse representation in a given basis or frame. If an orthonormal basis ψ is used, then a signal x has a unique representation $\theta = \psi^{-1}x$ and we can learn whether x is K-sparse in θ simply by inspecting this vector. When ψ is a frame, however, there are infinitely many solutions to the underdetermined problem $x = \psi\theta$ and hence infinitely many representations θ for x, making it more difficult to answer this question.

Thus, this underdetermined system needs additional constraints or conditions to find an exact solution that meets a given set of requirements. These conditions might be the minimization or maximization of certain parameters associated with the system. The problem of finding an extreme value of a function subject to some given constraints is quite a popular one and can be put as

$$\min f(x) \; such \; that \; x = \psi\theta \tag{2.2}$$

As is expected the function $f(x)$ can take any form, it might be the distance of the given vector from a point, its length or the number of nonzero elements in it. At this point the following question arises, how can sparse solutions to underdetermined systems of equations be obtained?

l_0-norm Minimization

To find the sparsest representation of the signal, we need to find a vector with minimum number of nonzero elements. So the problem may now be stated as

$$\min \|\theta\|_0 \; such \; that \; x = \psi\theta \tag{2.3}$$

with

$$\|\theta\|_0 = \sum_{i=1}^{N} |\theta_i|^0 \tag{2.4}$$

where,

$$|\theta_i|^0 = \begin{cases} 0, & if\ \theta_i = 0 \\ 1, & if\ \theta_i \neq 0 \end{cases}$$

It can thus be said that the sparse solution of a system of equations is the one that has the minimum l_0 norm. While this algorithm will - by construction - find the sparsest representation of the signal x in the frame ψ, its computational complexity is combinatorial; it must search whether the signal x is in the span of any of the columns of ψ then whether it is in the span of any pair of columns of ψ, then repeat for any set of three columns, etc., until a combination of columns for which x is in their span is found. But this aspect does not deter people from approximating and this task has been achieved with much success during the past many years of research. This has led to considerable effort being put into the development of many sub-optimal schemes.

Sub-optimal Algorithms

Instead of attempting to solve the problem exactly by brute force, approximate solutions have been developed that tend to approach the exact solution. The first relaxation is given in terms of the error in solving the system of equations. Instead of exactly solving the system, certain error is allowed so that a sparse solution may be achieved with some arbitrarily small error, represented in the following equation by ϵ.

$$\min \|\theta\|_0 \text{ such that } \|\psi\theta - x\|_2 < \epsilon \tag{2.5}$$

Various approximation algorithms have been put forward for the recovery of an approximate solution to the above equations. Some of these recovery methods may broadly be classified into different groups listed as follows.

Greedy pursuit: Iteratively refine a sparse solution by successively identifying one or more components that yield the greatest improvement in quality. These start from an all zero solution and add components to x one at a time based on selecting the best out of the available options. Some of such approaches are Matching pursuit [19], Orthogonal Matching Pursuit [20], Stagewise Orthogonal Matching Pursuit (StOMP) [21] etc. These have been shown to converge by Tropp in [22]. Matching Pursuit decomposes any signal into a linear expansion of waveforms that are selected from a redundant dictionary of functions. These waveforms are chosen in order to best match the signal structures. Matching pursuits are general procedures to compute adaptive signal representations.

Convex Optimization: The above is an optimization problem but the l_0 norm is not a convex function so it may be replaced by some equivalent convex function and the problem may be solved by convex optimization. The methods developed from this point of view are Basis Pursuit [23], FOCal Underdetermined System Solver (FOCUSS) [24] and Iteratively Reweighted Least Square (IRLS) [25]. As these methods are based upon standard optimization techniques so they are guaranteed to converge, the only thing that is needed to be proven in such cases is that the convex function chosen to replace the l_0 norm is adequate. Basis pursuit (BP) is a principle for decomposing a signal into an "optimal" superposition of dictionary elements, where optimal means having the smallest l_1 norm of coefficients among all such decompositions. FOCUSS has two integral parts: a low-resolution initial estimate of the real signal and the iteration process that refines the initial estimate to the final localized energy solution. The iterations are based on weighted norm minimization of the dependent variable with the weights being a function of the preceding iterative solutions.

Statistical approaches: As the least squares solution provides the ML (Maximum Likelihood) estimate of the system equations so in a probabilistic approximation of the sparsest solution an appropriate prior distribution of the elements of x is chosen and then a MAP [26] or an MMSE [27] estimate of x is found. The success of these methods depends mainly upon the appropriateness of the chosen apriori distribution.

Algorithmic Performance

To provide a guarantee for the performance of these algorithms, we define a metric of the frame ψ known as *coherence.*

Definition 6. Coherence: The coherence of a matrix ψ, $\mu(\psi)$, is the largest absolute inner product between any two columns ψ_i, ψ_j of ψ :

$$\mu(\psi) = \max_{1 \le i,j \le N} \frac{|\langle \psi_i, \psi_j \rangle|}{\|\psi_i\|_2 \|\psi_j\|_2}$$

It can be shown that the coherence of a matrix lies in the range $\mu(\psi) \in \left[\sqrt{\frac{N-M}{M(N-1)}}, 1\right]$ where the lower bound is known as Welch bound.

The coherence then dictates the maximum sparsity $||\theta||_0$ for which the BP and OMP algorithms obtain the sparse representation of $x = \psi\theta$ [3]. The BP and OMP algorithms can obtain the sparse representation of any K-sparse signal in ψ if $K < \frac{1}{2}(\frac{1}{\mu(\psi)} + 1)$.

However another question arises here, that pertains to the uniqueness of this sparse solution. It is known that the infinite solutions exist for the system of equations so the uniqueness of a solution of this problem becomes dubious. Uniqueness can be guaranteed by defining a relevant metric.

Definition 7. Spark: The spark of a given matrix ψ is the smallest number of columns of ψ that are linearly dependent.

Definition 8. Null Space: Null space of a matrix ψ is defined as $\mathcal{N}(\psi) = \{z : \psi z = 0\}$.

Theorem 1. If a signal x has a sparse representation $x = \psi\theta$ with $||\theta||_0 = K$ and

$$K < spark(\psi)/2$$

then θ is the unique sparsest representation of x in ψ.

Proof. Let there exist another vector θ' such that $\psi\theta' = x$ and $||\theta'||_0 < spark(\psi)/2$ so now $\psi(\theta - \theta') = 0$ and lies in $\mathcal{N}(\psi)$. Using the definition of the spark

$$||\theta||_0 + ||\theta'||_0 \geq ||\theta - \theta'||_0 > spark(\psi)$$

as any vector must have at least spark(ψ) non-zero components to lie in the null space of ψ. Also as the number of nonzero terms in $\theta - \theta'$ can not exceed the sum of the number of non zero terms in θ and θ' separately. And as $||\theta||_0 < spark(\psi)/2$ hence such a θ' does not exist. So a vector θ with the above mentioned properties is indeed unique.

2.2 Compressive Sensing

Consider the general problem of reconstructing a vector $x \in R^N$ from linear measurements $y \in R^M$ of x of the form

$$y_m = \langle x, \varphi_m \rangle, \qquad m = 1, 2, \dots, M, \qquad or \quad y = \phi x$$

where ϕ is a $M \times N$ transform matrix with $M \ll N$. That is, we acquire information about the unknown signal by sensing x against M vectors $\varphi_m \in R^N$. We are interested in the "underdetermined" case $M \ll N$, where we have many fewer measurements than unknown signal values. At first glance, solving the underdetermined system of equations appears hopeless. But if the signal x is *sparse* or *compressible* in basis ψ, meaning that it essentially depends on a number of degrees of freedom which is smaller than N, that is, it can be written either exactly or accurately as a superposition of a small number of vectors in some fixed basis, i.e.

$$x = \psi\theta$$

where, ψ is $N \times N$ matrix and θ is $N \times 1$ column vector and hence,

$$y = \phi\psi\theta$$

Then this radically changes the problem, making the search for solutions feasible. In fact, accurate and sometimes exact recovery is possible by solving a simple convex optimization problem. In other words, instead of sensing an N dimensional signal x with sparsity K, we can measure M random linear functionals of x where $M \ll N$ and find x by solving the under-determined system of equations as above with the extra condition that x is K sparse in basis ψ.

To recover the signal representation θ from its measurements y, we can exploit the fact that y will be sparse in the frame $\phi\psi$. However, a distinguishing feature of CS is that we do not want to find just a sparse representation of y, but rather we aim for the *correct representation* θ that yields our signal $x = \psi\theta$. Therefore, the requirements, guarantees, and algorithms relevant to CS signal recovery are slightly different from, although based on, the sparse representation and approximation algorithms mentioned earlier. For brevity, we define the matrix product $A = \phi\psi$ so that $y = A\theta$.

A major question arises out of this problem statement, i.e., how should the measurement/sensing matrix ϕ be designed so that it preserves the information in the signal x? Following section attempts to answer this question.

2.3 How to construct Measurement Matrix?

The measurement matrix ϕ represents a dimensionality reduction, i.e., it maps R^N, where N is generally large, into R^M, where M is typically much smaller than N. Here, we assume that the measurements are non-adaptive, meaning that the rows of ϕ are fixed in advance and do not depend on the previously acquired measurements.

The CS theory states that it is possible to construct an M × N *measurement* matrix ϕ, where M<<N, yet the measurements $y = \phi x$ preserve the essential information about x, if it satisfies some properties to be discussed below. For example, let ϕ be a cK ×N random matrix with i.i.d. Gaussian entries, where c = c(N,K) is an oversampling factor. Using such a matrix it is possible, with high probability, to recover any signal that is K-sparse in the basis ψ from its image under ϕ. For signals that are not K-sparse but compressible, meaning that their coefficient magnitudes decay exponentially, there are tractable algorithms that reconstruct signals with error not more than a multiple of the error of the best K-term approximation of the signal.

Some conditions/attributes that ϕ should satisfy for preserving the information and helping recovery are given below [18].

2.3.1 Restricted Isometry Property

In [3], Candes and Tao introduced the following isometry condition on matrix A and established its important role in CS.

Definition 9. Restricted Isometry Property (RIP): A matrix ϕ satisfies the restricted isometry property (RIP) of order K if there exists a $\delta_K \in (0,1)$ such that

$$(1-\delta_K)\|x\|_2^2 \leq \|\phi x\|_2^2 \leq (1+\delta_K)\|x\|_2^2 , \tag{2.6}$$

This property essentially requires that every set of columns with cardinality less than K approximately behaves like an orthonormal system. An important result is that if the columns of the measurement matrix are approximately orthogonal, then the exact recovery phenomenon occurs [3]. If a matrix ϕ satisfies the RIP of order 2K, then we can interpret (2.1) as saying that ϕ approximately preserves the distance between any pair of K-sparse vectors. Also, if ϕ satisfies the RIP of order K with constant δ_K, then for any $K' < K$ we automatically have that ϕ satisfies the RIP of order $\delta_{K'}$ with constant $\delta_{K'} \leq \delta_K$.

RIP & the stability:

Definition 10. Let $\phi : R^N \rightarrow R^M$ denote a measurement matrix and $\Delta: R^M \rightarrow R^N$ denote a recovery algorithm. We say that the pair $(\phi; \Delta)$ is C -stable if for any $x \in \Sigma_K$ and any $e \in R^M$, we have that

$$\|\Delta(\phi x + e) - x\|_2 \leq C\|e\|_2$$

.

This definition simply says that if we add a small amount of noise to the measurements, then the impact of this on the recovered signal should not be arbitrarily large.

Practical recovery algorithms typically require that ϕ have a slightly stronger 2K-RIP, 3K-RIP, or higher-order RIP [3]. In fact, the uniqueness requirement is implied when the matrix has the 2K-RIP with $\delta_{2K} > 0$ as this implies that all sets of 2K columns be linearly independent, putting spark(ϕ) > 2K.

2.3.2 Mutual Coherence

In particular cases, the choice of measurements that can be taken from the signal are limited to a transformation, such as the Fourier/Radon transform performed in magnetic resonant imaging. Thus, we can assume that a basis $\phi \in R^{N\times N}$ is provided for measurement purposes, and we can choose a subset of the signal's coefficients in this transform as measurements. That is, let $\bar{\phi}$ be an N x M submatrix of ϕ that preserves the basis vectors with indices Γ and $y = \bar{\phi}^T x$. Under this setup, a different metric arises to evaluate the performance of CS.

Definition 11. Mutual Coherence: The mutual coherence of the N-dimensional orthonormal bases ϕ and ψ is the maximum absolute value for the inner product between elements of the two bases:

$$\mu(\phi, \psi) = \max_{1 \leq i,j \leq N} |\langle \phi_i, \psi_j \rangle|$$

.

2.3.3 Random Matrices

Fortunately, these conditions can be achieved by randomizing the matrix construction. It can be shown that random matrices will satisfy the RIP with high probability if the entries are chosen according to a Gaussian, Bernoulli, or more generally any sub-Gaussian distribution [3].

Using random matrices to construct ϕ has a number of additional benefits. To illustrate these, the focus will be on the RIP.

1. For random constructions the measurements are *democratic*, meaning that it is possible to recover a signal using any sufficiently large subset of the measurements. Thus, by using random ϕ one can be robust to the loss or corruption of a small fraction of the measurements.

2. In practice, we are often more interested in the setting where x is sparse with respect to some basis ψ. In this case what we actually require is that the product $A = \phi\psi$ satisfies the RIP. If we were to use a deterministic construction then we would need to explicitly take ψ into account in our construction of ϕ, but when ϕ is chosen randomly we can avoid this consideration. For example, if ϕ is chosen according to a Gaussian distribution and ψ is an orthonormal basis then one can easily show that A will also have a Gaussian distribution, and so provided that M is sufficiently high A will satisfy the RIP with high probability, just as before.

2.3.4 Measurement bounds

A K-sparse signal can be reconstructed from M measurements if M satisfies the following conditions:

$$\mathrm{M} \geq c.\mu^2(\phi,\psi).\mathrm{K}.\log \mathrm{N}$$

where, c is a positive constant. The smaller the coherence between ϕ & ψ is, the lesser measurements are required to reconstruct the signal. A random basis has been shown to be largely incoherent with any fixed basis, and M = 3K~4K is usually sufficient to reconstruct the signal.

2.4 Principal Component Analysis (PCA)

Transformation domain that make the signal sparse (ψ) needs to be known for the reconstruction of the signals using CS. Using a fixed basis with a time varying correlation structure as in WSNs may not provide the expected results. Transformation domain should adapt to the signals. PCA allows to dynamically learn the optimal transformation to be used by CS recovery, effectively accounting for the time varying correlation affecting real signals as in the case of WSNs [10].

Let $x^{(n)} \in \mathrm{R}^{\mathrm{J}}$ be the vector of measurements from our WSN at a given time instant n, where the network consists of J nodes. We collect measurements according to a fixed sampling rate at discrete times $n = 1, 2, \ldots, \mathrm{N}$. From a geometrical point of view, we consider $x^{(n)}$ as a point in R^{J} and look for the K-dimensional plane (with $\mathrm{K} \ll \mathrm{J}$) which best matches the points in $x^{(n)}$ in terms of minimum Euclidean distance. The sample mean vector $\bar{x}$ and the sample covariance matrix $\hat{\Sigma}$ of $x^{(n)}$ are given as:

$$\bar{x} = \frac{1}{\mathrm{N}} \sum_{n=1}^{\mathrm{N}} x^{(n)}, \qquad \hat{\Sigma} = \frac{1}{\mathrm{N}} \sum_{n=1}^{\mathrm{N}} \left(x^{(n)} - \bar{x}\right)\left(x^{(n)} - \bar{x}\right)^T$$

Given the above equations, we consider the orthonormal matrix $\mathbf{U}$ whose columns are the eigenvectors of the covariance matrix $\hat{\Sigma}$ placed in decreasing order with respect to the corresponding eigenvalues. If we define the vector $s^{(n)}$ as:

$$s^{(n)} = \mathbf{U}^T\left(x^{(n)} - \bar{x}\right)$$

Assuming that the instances $x^{(1)}, x^{(2)}, \ldots, x^{(\mathrm{N})}$ of the process x are correlated, as is often the case in WSN monitoring applications, there exists an $\mathrm{K} \leq \mathrm{J}$ such that all the component $s_i^{(n)}$ with $i = \mathrm{K} + 1, \ldots, \mathrm{J}$ are negligible with respect to the average energy, where the actual value of K depends on the spatial correlation of the signal. We can write:

$$x^{(n)} = \bar{x} + \psi s^{(n)}$$

where, we have defined the sparsifying matrix $\psi = \mathbf{U}$ [28]. The J-dimensional vector $s^{(n)}$ obtained through PCA turns out to be K-sparse, so it can be efficiently recovered with CS.

We extend the use of PCA for multiple measurement vectors and recover signals using joint Distributed Compressed Sensing (DCS) and PCA. But, before that, let us understand the theory behind DCS and various recovery algorithms used for DCS in chapter 3.

2.5 Summary

In this chapter, theory behind CS is discussed. The concept of sparsity on which CS is based upon, problem formulation to find sparse representation and various algorithms to solve the under-determined system of equations to find sparse representation are mentioned. Also, motivation behind CS and the exact CS problem is presented. The detailed theory includes conditions required to be satisfied by measurement matrix to ensure recoverability of any K-sparse signal. Principal Component Analysis (PCA) used to dynamically find transformation matrix for signals with time varying correlation is also discussed.

M.G. CENTRAL LIBRARY
ACC No. G21478
Date 17/8/12
I.I.T. ROORKEE

3 Distributed Compressed Sensing

While the theory and practice of compression have been well developed for individual signals, there has been less progress for multiple signals. One such application involving multiple signals is sensor network in which a potentially large number of distributed sensor nodes are programmed to perform a variety of data acquisition tasks as well as to network themselves to communicate their results to a central collection point. In many sensor networks, and in particular battery-powered ones, communication energy and bandwidth are scarce resources which necessitate the reduction of amount of data to be transmitted. Fortunately, since the sensors presumably observe related phenomena, the ensemble of signals they acquire are expected to possess some joint structure resulting in inter-signal correlation, in addition to the intra-signal correlation in each individual sensor's measurements.

Duarte et al. [11] introduced a new theory for distributed compressed sensing (DCS) that enables new distributed coding algorithms that exploit both intra- and inter-signal correlation structures. In a typical DCS scenario, a number of sensors measure signals (of any dimension) that are each individually sparse in some basis and are also correlated amongst themselves. Each sensor independently encodes its signal by projecting it onto another incoherent basis (such as a random one) and then transmits the resulting coefficients to a single collection point. Under the right conditions, a decoder at the collection point is able to reconstruct each of the signals precisely. The DCS theory rests on a concept termed as the joint sparsity of a signal ensemble.

3.1 Advantages of DCS

In addition to offering substantially reduced measurement rates in multi-signal applications, DCS employ random projections at the sensors. As in single-signal CS, random measurement bases are universal in the sense that they may be paired with any sparsifying basis. This allows exactly the same encoding strategy to be applied in a variety of different sensing environments;

knowledge of the nuances of the environment is needed only at the decoder. Moreover, random measurements are also future-proof: if a better sparsity-inducing basis is found for the signals, then the same random measurements may be used to reconstruct an even more accurate view of the environment. A pseudorandom basis as measurement basis may be generated using a simple algorithm according to a random seed. Such encoding effectively implements a form of encryption: the randomized measurements will themselves resemble noise and are meaningless to an observer who does not know the measurement basis. Random coding is also robust: the randomized measurements coming from each sensor have equal priority, unlike the Fourier or wavelet coefficients in current coders. Thus, they allow a progressively better reconstruction of the data as more measurements are obtained; one or more measurements may also be lost without corrupting the entire reconstruction.

Two additional properties of DCS make it well-matched to distributed applications such as sensor networks and arrays [12]. First, each sensor encodes its measurements independently, which reduces inter-sensor communication overhead to zero. Second, DCS distributes its computational complexity asymmetrically, placing most of it in the joint decoder, which often has more substantial computational resources than any individual sensor node. The encoders are very simple; they merely compute incoherent projections with their signals and make no decisions.

3.2 Joint Sparsity Models

A joint sparsity model (JSM) encodes the correlation between the values and locations of the coefficients for a group of sparse signals. Three different joint sparsity models (JSMs) have been proposed in [12] that apply in different situations. In the first two models, each signal itself is sparse, and so the CS framework may be used to encode and decode each one separately (independently). In the third model, no signal is itself sparse, yet there still exists a joint sparsity among the signals that allows recovery from significantly fewer measurements per sensor.

3.2.1 JSM-1: Sparse common component + innovations

In this model, each signal consists of a sum of two components: a common component that is present in all the signals and an innovation component that is unique to each signal. Both the common and innovation components may be sparsely represented in some basis.

$$x_j = z_C + z_j, \qquad j \in \{1,2,\ldots,J\}$$

with

$$z_C = \psi\theta_C, \qquad \|\theta_C\|_0 = \mathrm{K}$$

$$z_j = \psi\theta_j, \qquad \|\theta_j\|_0 = \mathrm{K_j}$$

Thus, the signal z_C is common to all the x_j and has sparsity K in basis ψ. The signals z_j are the unique portions of the x_j and have sparsity $\mathrm{K_j}$ in the same basis [11].

Such signals may arise in settings where large-scale phenomena affect all sensors and local phenomena affect individual sensors. A practical situation well-modelled by JSM-1 is a group of sensors measuring temperatures at a number of outdoor locations throughout the day. The temperature readings x_j have both temporal (intra-signal) and spatial (inter-signal) correlations. Global factors, such as the sun and prevailing winds, have an effect z_C that is both common to all sensors and structured enough to permit sparse representation. More local factors, such as shade, water, or animals, contribute localized innovations z_j that are also structured (and hence sparse). A similar scenario may be imagined for a network of sensors recording light intensities, air pressure, or other phenomena. All these scenarios correspond to measuring properties of physical processes that change smoothly in time and in space and thus are highly correlated.

For JSM-1, there exists an analytical framework inspired by principles of information theory. This allows characterizing the measurement rates $\mathrm{M_j}$ required to *jointly* recover the signals x_j. The measurement rates relate directly to the signals' *conditional sparsities*. The recovery technique is based on a single execution of a weighted linear program that seeks the sparsest components $[z_C; z_1; \ldots; z_J]$ that account for the observed measurements. Theoretical analysis and numerical experiments confirm that the rates M_j required for joint CS recovery are well below those required for independent CS recovery of each signal x_j [12].

3.2.2 JSM-2: Common sparse supports

In this model, all signals are constructed from the same sparse set of basis vectors, but with different coefficient values:

$$x_j = \psi\theta_j, \qquad j \in \{1, 2, \dots, J\}$$

where, each θ_j is supported on the same $\Omega \subset \{1, 2, \dots, N\}$ with $|\Omega| = K$, $K \ll N$. Hence, all signals are K-sparse and are constructed from the same K elements of ψ, but with arbitrarily different coefficients [11]. This model may be viewed as a special case of JSM-1 (with $K_C = 0$ and $K_j = K$ for all j) but features additional correlation structure that suggests distinct recovery algorithms.

A practical situation well-modelled by JSM-2 is where multiple sensors acquire replicas of the same Fourier-sparse signal but with different phase shifts and attenuations caused by signal propagation. In many cases it is critical to recover each of the sensed signals, such as in many acoustic localization and array processing algorithms. Another useful application for JSM-2 is MIMO communication [12].

For JSM-2, Duarte et al. [11] proposed two techniques based on iterative greedy pursuit for signal ensemble reconstruction from independent, incoherent measurements inspired by conventional greedy algorithms (such as OMP) that can substantially reduce the number of measurements when compared with independent recovery. In the single-signal case, OMP iteratively constructs the sparse support set Ω; decisions are based on inner products between the columns of $\phi\psi$ (ϕ is the measurement matrix) and a residual. In the multi-signal case, there are more clues available for determining the elements of Ω. For a large number of sensors J, joint recovery is possible with the number of measurements per signal close to K (that is, oversampling factor $c \to 1$ as $J \to \infty$); see Figure 3.1 for an example of improving performance as J increases. On the contrary, with independent CS recovery, perfect recovery of all signals requires increasing each M_j in order to maintain the same probability of recovery of the signal ensemble. This is due to the fact that each signal experiences an independent probability $p < 1$ of successful recovery; therefore the overall probability of complete success is p^J. Consequently, each sensor must compensate by making additional measurements.

It is noted that when the supports of the innovations of the signals are small, signals that are well modelled by JSM-1 may also be modelled by JSM-2 by selecting a global support that contains all of the individual supports. Such approximation allows for a simpler recovery algorithm, while incurring a slight increase in the number of measurements required for recovery [29].

One-step Greedy Algorithm (OSGA)

Given all of the measurements $y_j = \phi x_j$, $j = \{1, 2, \dots, J\}$, where $x_j \in \mathrm{R}^N$ compute the test statistics [11]

$$\xi_n = \frac{1}{J}\sum_{j=1}^{J}\langle \boldsymbol{y}_j, \boldsymbol{\Phi}_{j,n}\rangle^2, \quad n \in \{1, 2, 3, \dots, N\}$$

and estimate the elements of the common coefficient support set by

$$\hat{\Omega} = \{n \text{ having K largest } \xi_n\}.$$

DCS-SOMP (Simultaneous Orthogonal Matching Pursuit)

Simultaneous Orthogonal Matching Pursuit (SOMP), proposed by Tropp and Gilbert [30], is a variant of OMP that seeks to identify one element at a time. It has been extended in [11] to DCS-SOMP to adapt it to JSM-2. In each iteration, the column index $n \in \{1, 2, \dots, N\}$ is selected that accounts for the greatest amount of residual energy across all signals.

First, the iteration counter is set as $l = 1$. For each signal index $j \in \{1, 2, \dots, J\}$, the orthogonalized coefficient vectors $\hat{\boldsymbol{\beta}}_j = 0$, $\hat{\boldsymbol{\beta}}_j \in \mathrm{R}^M$ are initialized. The set of selected indices $\hat{\Omega} = \emptyset$ is also initialized. If $\boldsymbol{r}_{j,l}$ denote the residual of the measurement $\boldsymbol{y}_j$ remaining after the first l iterations, initializing $\boldsymbol{r}_{j,0} = \boldsymbol{y}_j$, we have

$$n_l = \arg \max_{n=1,2,\dots,N} \sum_{j=1}^{J} \frac{|\langle \boldsymbol{r}_{j,l-1}, \boldsymbol{\Phi}_{j,n}\rangle|}{\|\boldsymbol{\Phi}_{j,n}\|_2}$$

$$\hat{\Omega} = [\hat{\Omega}\ n_l]$$

Next, the selected basis vector is orthogonalized against the orthogonalized set of previously selected dictionary vectors

$$\boldsymbol{\gamma}_{j,l} = \boldsymbol{\Phi}_{j,n_l} - \sum_{t=0}^{l-1} \frac{\langle \boldsymbol{\Phi}_{j,n_l}, \boldsymbol{\gamma}_{j,t} \rangle}{\left\| \boldsymbol{\gamma}_{j,t} \right\|_2^2} \boldsymbol{\gamma}_{j,t}$$

The estimate of the coefficients are then updated for the selected vector and residual as

$$\hat{\boldsymbol{\beta}}_j(l) = \frac{\langle \boldsymbol{r}_{j,l-1}, \boldsymbol{\gamma}_{j,l} \rangle}{\left\| \boldsymbol{\gamma}_{j,l} \right\|_2^2}$$

$$\boldsymbol{r}_{j,l} = \boldsymbol{r}_{j,l-1} - \frac{\langle \boldsymbol{r}_{j,l-1}, \boldsymbol{\gamma}_{j,l} \rangle}{\left\| \boldsymbol{\gamma}_{j,l} \right\|_2^2} \boldsymbol{\gamma}_{j,l}$$

If $\left\| \boldsymbol{r}_{j,l} \right\|_2 > \epsilon \left\| \boldsymbol{y}_j \right\|_2$ for all j, the iteration index l is incremented and next iteration is performed. The parameter ϵ determines the target error power level allowed for algorithm convergence.

3.2.3 JSM-3: Nonsparse common component + sparse innovations

This model extends JSM-1 so that the common component needs no longer be sparse in any basis, that is,

$$x_j = z_C + z_j, \qquad j \in \{1,2,\dots,J\}$$

with

$$z_C = \psi\theta_C,$$

$$z_j = \psi\theta_j, \qquad \left\| \theta_j \right\|_0 = K_j$$

but z_C is not necessarily sparse in the basis ψ [12].

A practical situation well-modelled by JSM-3 is where several sources are recorded by different sensors together with a background signal that is not sparse in any basis. Consider, for example, an idealized computer vision-based verification system in a device production plant that checks for failures in the devices for quality control purposes. Cameras acquire snapshots of components in the production line from different viewing points. While each image could be extremely complicated, the ensemble of images is highly correlated, since each camera is observing the same device with minor (sparse) variations.

For JSM-3, since the common component is not sparse, no individual signal contains enough structure to permit efficient compression for CS; in general N measurements would be required for each individual N-sample signal. However, it is demonstrated in [12] that the common structure shared by the signals permits a drastic reduction in the required measurement rates. This is the main concept behind the Alternating Common and Innovation Estimation (ACIE) recovery algorithm [12], which alternates between two steps: (1) Estimate the common component z_C by combining all measurements and treating the innovations z_j as noise that may be averaged out; (2) Estimate the innovations z_j from each sensor by subtracting the estimated common component z_C and then applying standard CS recovery techniques. In fact, asymptotically, the required measurement rates relate simply to the sparsity K of the innovation components; as the number of sensors grows, each sensor may again reduce its oversampling factor to c = 1. Thus, for a large number of sensors J, the impact of the common nonsparse component z_C is eliminated.

3.3 M-FOCUSS

In [13], Cotter et al. addressed the problem of finding sparse solutions to an underdetermined system of equations when there are multiple measurement vectors having the same, but unknown, sparsity structure. It extends two classes of algorithms–Matching Pursuit (MP) and FOCal Underdetermined System Solver (FOCUSS) - to the multiple measurement vectors (MMV) case so that they may be used in applications such as neuromagnetic imaging, where multiple measurement vectors are available, and solutions with a common sparsity structure must be computed.

3.3.1 Problem Formulation

MMV problem may be stated as solving the following L underdetermined systems of equations:

$$Ax^{(l)} = y^{(l)}, \qquad l = 1, 2, \ldots, L$$

where, $A = \phi\psi \in R^{M\times N}$, M<N. L is the number of measurement vectors and it is usually assumed that L<M. The quantities $y^{(l)} \in R^M$, $l = 1,2, \ldots L$ are the measurement vectors and $x^{(l)} \in R^N$, $l = 1, 2, \ldots, L$ are the corresponding source vectors. We may also write it as:

$$AX = Y$$

where, $X = [x^{(1)}, x^{(2)}, \ldots, x^{(L)}]$ and $Y = [y^{(1)}, y^{(2)}, \ldots, y^{(L)}]$. It is assumed that $x^{(l)}$, $l = 1, 2, \ldots, L$ are sparse and have the same sparsity profile so that the indices of the nonzero entries are independent of l. The problem is to find the maximally sparse solution from among the infinite solutions.

3.3.2 Algorithm

One of the sub-optimal algorithms for single measurement vector (SMV), FOCUSS [24], has been extended for the case of multiple measurement vectors [13]. It is referred as M-FOCUSS. It is summarized as follows:

$$W_{k+1} = diag\left(c_k[i]^{1-\frac{p}{2}}\right),$$

where,

$$c_k[i] = \|x_k[i]\| = (\sum_{l=1}^{L} (x_k^{(l)}[i])^2)^{1/2}, \qquad p \in [0, 1]$$

where, $x[i] = \left[x^{(1)}[i], x^{(2)}[i], \ldots, x^{(l)}[i]\right]$ is the ith row of X.

$$Q_{k+1} = A_{k+1}^{\dagger} Y, \text{ where } A_{k+1} = AW_{k+1}$$

$$X_{k+1} = W_{k+1} Q_{k+1}$$

The algorithm is terminated once the convergence criterion has been satisfied, which is given as

$$\frac{\|X_{k+1} - X_k\|_F}{\|X_k\|_F} < \delta$$

3.4 Simulation Results

With Synthetic Data

Algorithms for JSM-2 are simulated in MATLAB and their results are shown below. OSGA and DCS-SOMP are used to reconstruct the support set of synthetic signals. Probability of exact reconstruction (Probability of recovering the support set) v/s Number of measurements per sensor for different number of sensors is plotted as shown in Fig 3.1 and Fig 3.2, respectively.

As expected, the average number of measurements per sensor, M, required for perfect reconstruction decreases if signals from more sensors are used for joint decoding.

Simulation results for M-FOCUSS are plotted in Fig. 3.3. It is a plot of Root Mean Square Error (RMSE) defined as

$$\text{RMSE} = \frac{\|X_{rec} - X\|_2}{\|X\|_2}$$

where, X is the data generated and X_{rec} is the data recovered, v/s Number of measurements, M.

Again, as number of sensors L increases, average number of measurements per sensor required for perfect reconstruction decreases.

With Real Data

Fig 3.4 shows the comparison between DCS-SOMP and M-FOCUSS. This simulation is performed using the real signals from EPFL WSN deployment LUCE (Available: http://sensorscope.epfl.ch/). Here, we use Haar wavelet transform for the sparse representation of the signals.

It can be seen that DCS-SOMP requires lesser number of measurements per sensor than M-FOCUSS for reconstruction of signals.

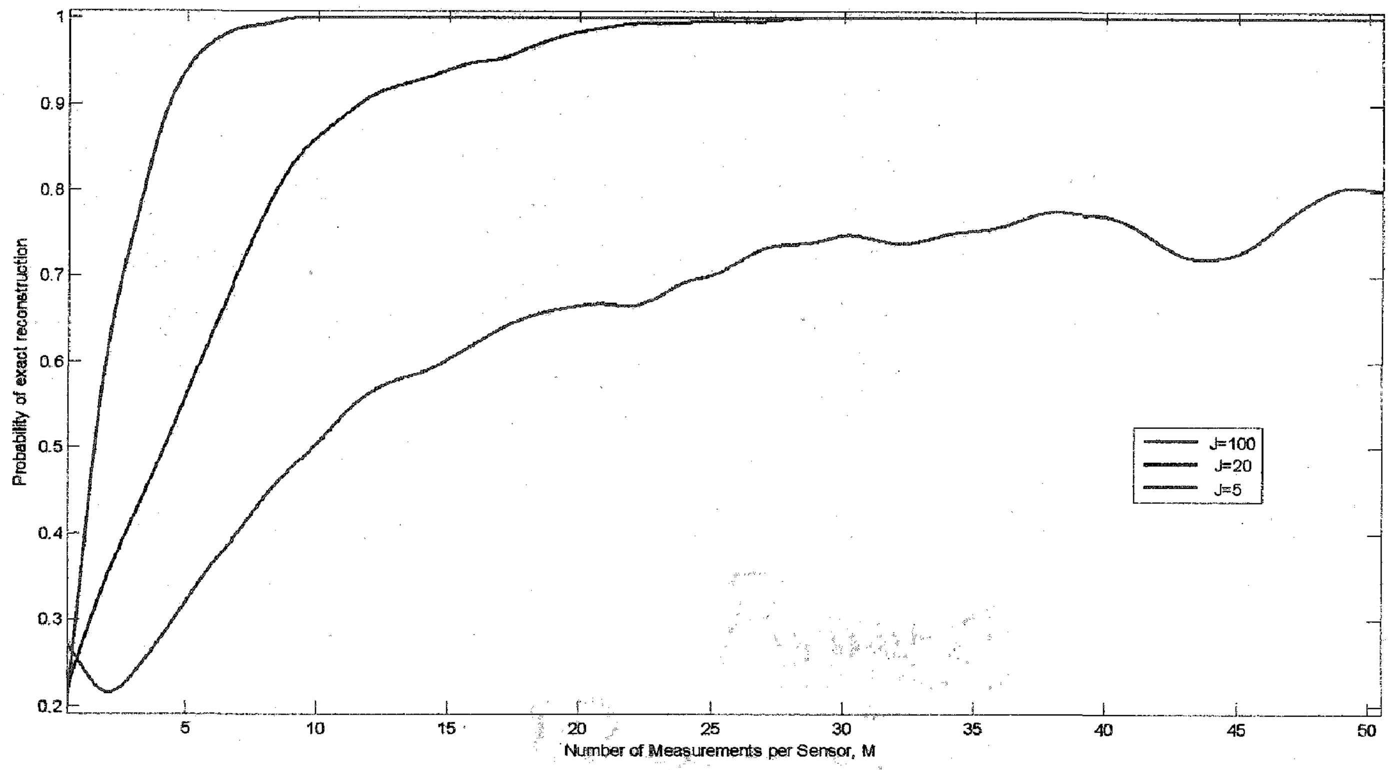

Figure 3.1 Plot of probability of exact reconstruction via OSGA as a function of the number of measurements per sensor M and the number of sensors J. Signal length N = 50, Sparsity K = 5.

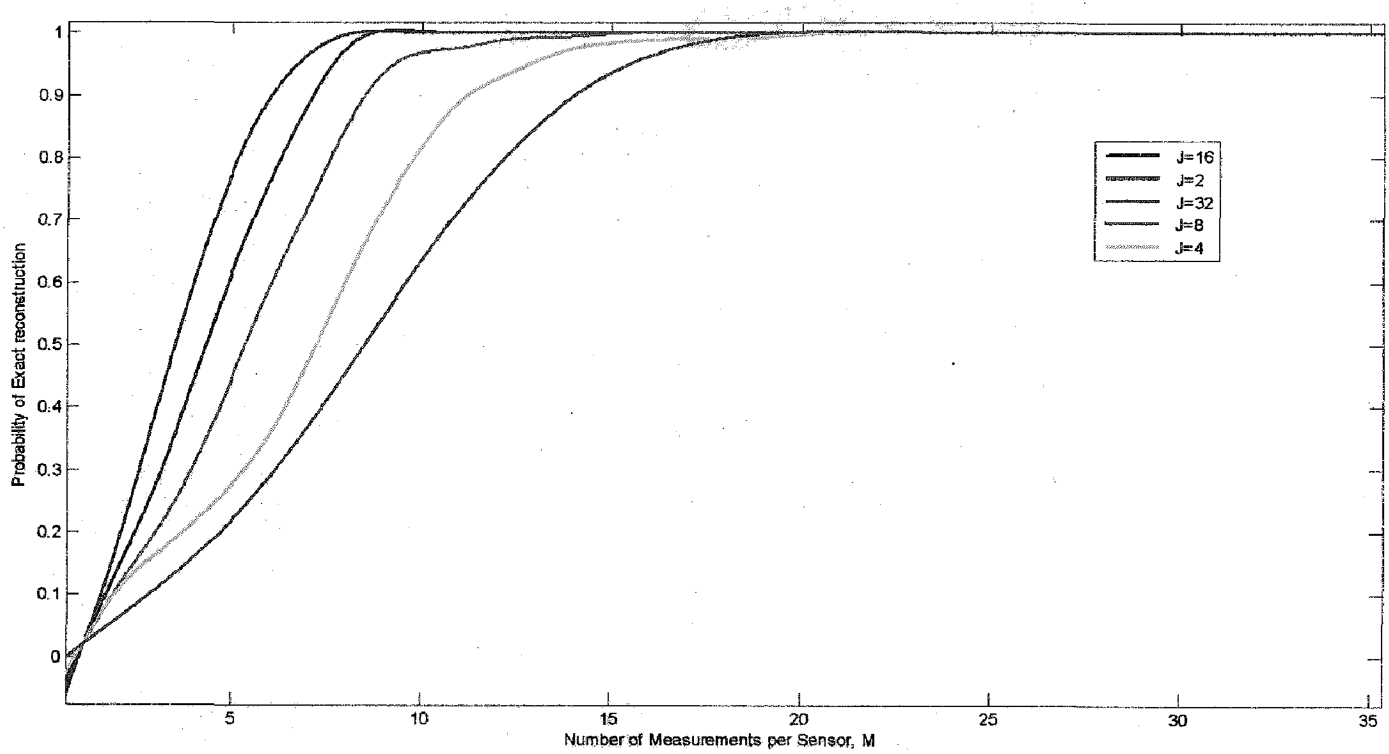

Figure 3.2 Plot of probability of exact reconstruction via DCS-SOMP as a function of the number of measurements per sensor M and the number of sensors J. Signal length N = 50, Sparsity K = 5.

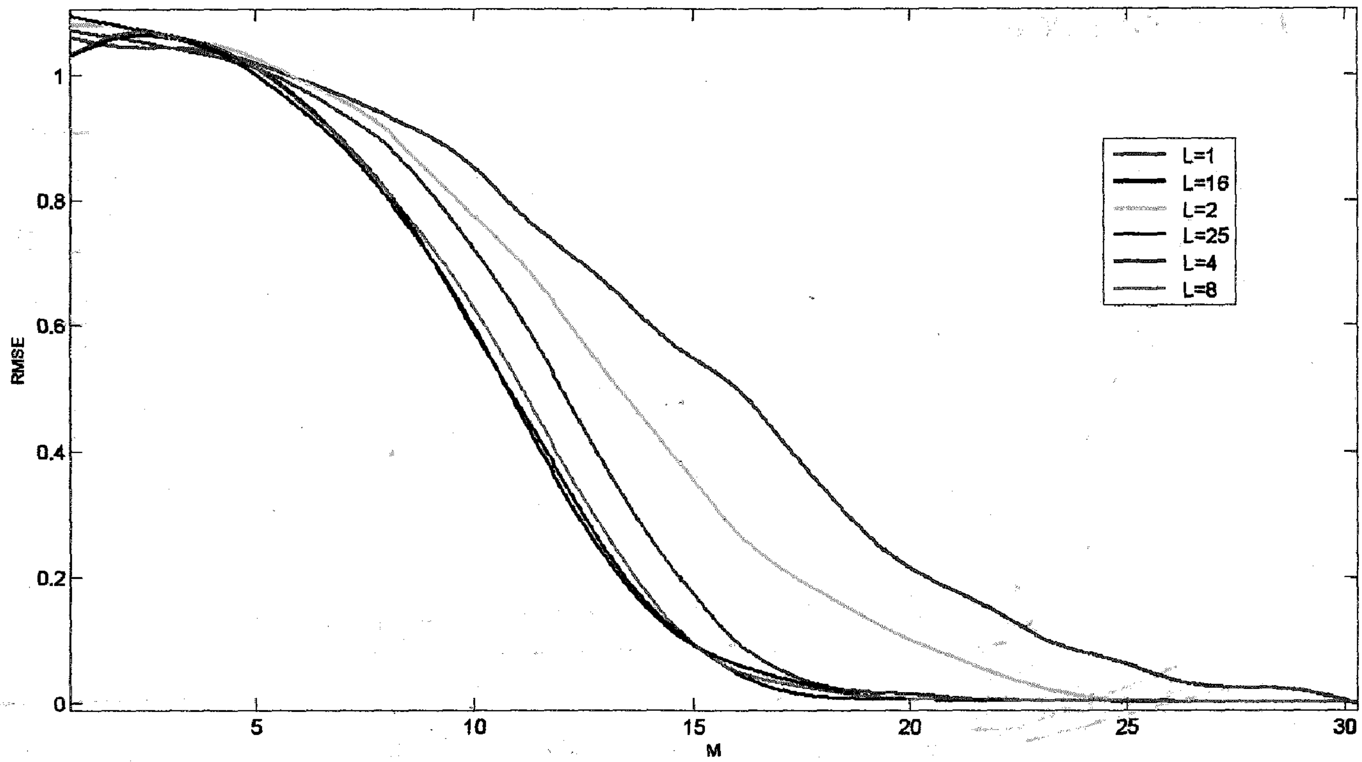

Figure 3.3 Plot of Root Mean square error via M-FOCUSS as a function of number of measurements per sensor M and the number of sensors L. Signal length N=50, Sparsity K=7.

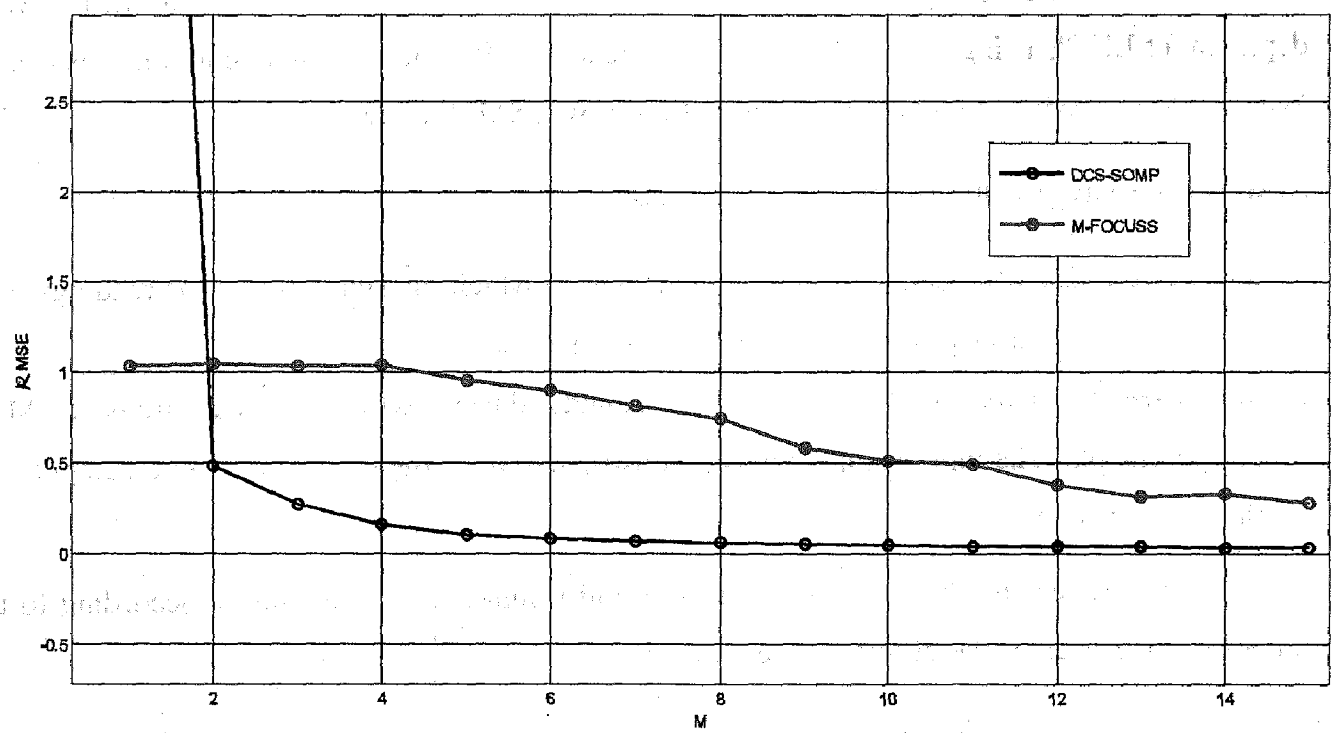

Figure 3.4 Plot of Root Mean square error via DCS-SOMP and M-FOCUSS as a function of number of measurements M. Signal length N=64 and number of sensors J=16.

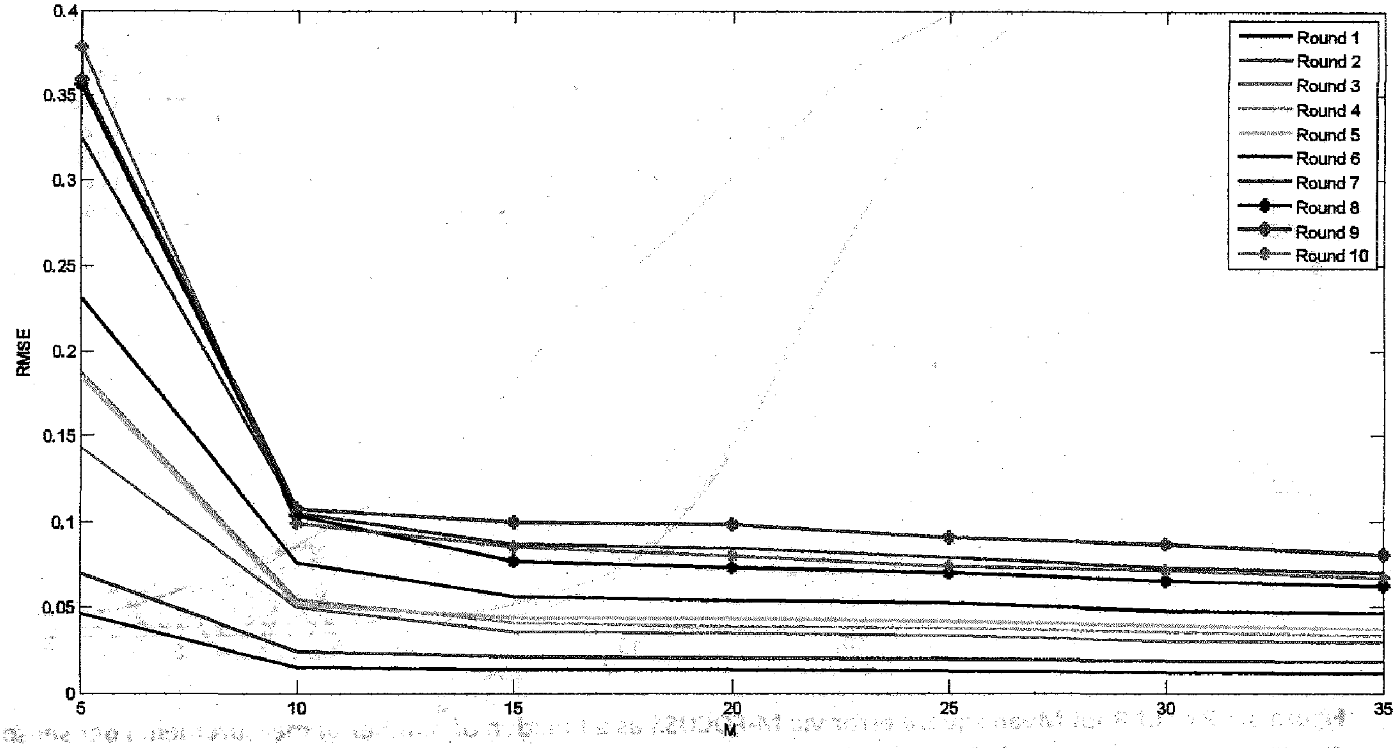

Figure 3.5 Plot of Root Mean square error via SOMP and PCA as a function of number of measurements per sensor M for ambient temperature. Signal length N=64, J=16.

We reconstruct the sensor readings of Ambient Temperature (Fig 3.5) from EPFL WSN deployment LUCE using the sparsifying matrix learnt by Principal Component Analysis (PCA). We use DCS-SOMP for reconstruction as data follows JSM-2 model.

However, we have to alternate between two phases:

1. a *training phase* of N data collection rounds during which the sink collects the readings from all J sensors and uses this information to compute $\bar{x}$ and $\hat{\Sigma}$;
2. a subsequent *monitoring phase* of ςN rounds during which we can transmit M<N measurements. The input signal is thus reconstructed using the statistics $\bar{x}$ and $\hat{\Sigma}$ computed in the previous phase.

The ratio ς between the duration of monitoring and training phases is chosen according to the temporal correlation of the observed phenomena.

We can see from Fig 3.5 that for M=10, N=64, we can reconstruct data with less than 10% RMSE for $\varsigma = 6$ rounds of monitoring phase after 1 training phase.

3.5 Summary

In this chapter, we discuss the motivation behind and the concept of Distributed Compressed Sensing which uses the temporal and spatial correlation to reduce the number of measurements for joint recovery of signals in sensor networks. Different joint sparsity models and their recovery algorithms are also discussed. Also, M-FOCUSS used for recovery from multiple measurement vectors has been used for JSM-2. Simulation results of M-FOCUSS and JSM-2 recovery algorithms OSGA and DCS-SOMP are shown. Further, a comparison in terms of number of measurements per sensor required for reconstruction of signals between M-FOCUSS and DCS-SOMP has been done. Simulation results for signal recovery through joint DCS and PCA are also shown.

4 Distributed Compressive Data Gathering in Wireless Sensor Network

4.1 Data gathering using CS

In general, data transmissions are accomplished through multi-hop routing from individual sensor nodes to the data sink as shown in Fig 4.1. Successful deployment of such large scale sensor network faces two major challenges in effective global communication cost reduction and in energy consumption load balancing. In Fig. 4.1, Node s_1 transmits its reading d_1 to s_2, and s_2 transmits both its reading d_2 and the relayed reading d_1 to s_3. At the end of the route, s_N transmits all N readings to the sink. It may be observed that the closer a sensor is to the sink, the more energy it consumes. Clearly, the sensor nodes closer to the data sink will soon run out of energy and consequently, lifetime of the sensor network will be significantly shortened.

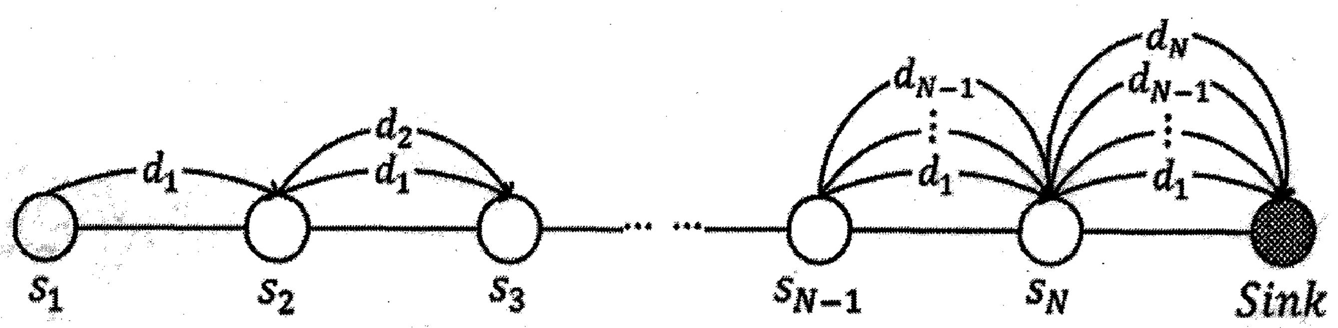

Figure 4.1 Baseline data gathering [31]

4.1.1 Randomised Gossip

In any general multi-hop network, each of the sensors locally generates its information by multiplying its data with the corresponding element of the compressing matrix. The local information is simultaneously aggregated and distributed across the network by randomized gossip algorithm [17]. When randomized gossip terminates, each node in the network knows all the data in the network, and then M of the nodes are randomly chosen to transmit their information to the sink.

4.1.2 Compressive Wireless Sensing

In [15], Bajwa et al. introduced the concept of Compressive Wireless Sensing for single-hop sensor networks in which a fusion centre retrieves signal field information from an ensemble of spatially distributed sensor nodes based on a distributed matched source-channel communication architecture. It requires no prior knowledge about the sensed data. It is shown to be able to reduce the latency of data gathering by delivering linear projections of sensor readings through synchronized amplitude-modulated analogue transmissions. In this case, analog mode of transmission is used, which corresponds to a completely decentralized way of delivering M random projections of the sensed data to the fusion centre by employing M transmissions. However, it requires accurate synchronization of sensor nodes.

4.1.3 Compressive Data Gathering

In baseline data gathering scheme, data transmissions are generally accomplished through multi-hop routing from individual sensor nodes to the data sink as shown in Fig 4.1. Nodes close to the sink will transmit more data and consume more energy than those at the peripheral of the network. The unbalanced energy consumption has a major impact on the network lifetime, defined as the time till the first node fails.

For large scale monitoring sensor networks, Luo et al. have proposed Compressive Data Gathering (CDG) in [16] and IR-CDG in [31] that leverages compressive sampling (CS) principle to efficiently reduce communication cost and prolong network lifetime. In a densely deployed sensor networks, sensors have spatial correlations in their readings. Let J sensor readings form a vector $d = [d_1\ d_2\ ...\ d_J]^T$, which is a K-sparse signal in a particular domain ψ. Then, we have $d = \psi\theta$, where θ is a sparse vector. As the sensor readings are compressible, CS may be used for recovery of d from $M < N$ incoherent measurements $y = \phi d$ using various recovery algorithms, where ϕ is the measurement matrix.

The data gathering process of CDG is depicted in Fig. 4.2 through a simple chain-type topology. Comparing this with the baseline data gathering scheme in Fig. 4.1, CDG delivers weighted sums (or measurements) of sensor readings, instead of individual readings, to the data sink. To transmit the i^{th} measurement to the sink, s_1 multiplies its reading d_1 with a random coefficient ϕ_{i1} and sends the product to s_2. Then s_2 multiplies its reading d_2 with a random coefficient ϕ_{i2} and sends the sum $\phi_{i1}\,d_1 + \phi_{i2}\,d_2$ to s_3. Similarly, each node s_j contributes to the relayed message by adding its own product. Finally, the sink receives $\sum_{j=1}^{N} \phi_{ij}\,d_j$, a weighted sum of all the readings. This process is repeated using M sets of different weights so that the sink receives M measurements. With such design, all nodes transmit M messages and consume same amount of energy. This is able to achieve substantial sensor data compression without introducing excessive computation. With elegant design, this scheme is also able to disperse the communication costs to all sensor nodes along a given sensor data gathering route. This results in a natural load balancing and extends the lifetime of the sensor network.

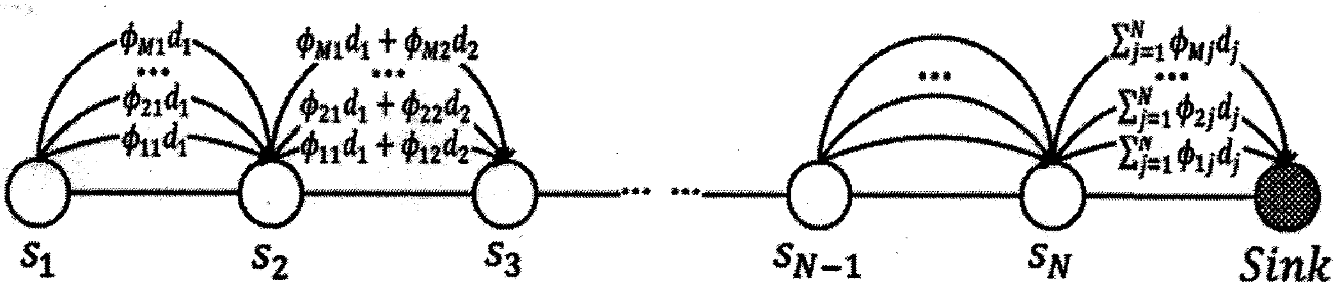

Figure 4.2 Basic CDG [31]

To further reduce the communication cost, IR-CDG was proposed in [31] which uses [**I R**] as measurement matrix ϕ where **R** is the original measurement matrix with its entries being i.i.d. Gaussian random numbers drawn according to $\mathcal{N}(0, \frac{1}{M})$ and **I** is the identity matrix of size

M×M. Matrix [I R] preserves the restricted isometry property and incurs the minimum communication cost in multi-hop networks and thus is a favourable choice for data gathering. By using [I R] as the measurement matrix, the first M sensor nodes simply transmit their original sensor readings to node s_{M+1}. Upon receiving the reading from sensor s_i, s_{M+1} computes the i^{th} product and transmits $d_i + \phi_{iM+1} d_{M+1}$ to the next node. In IR-CDG, the first M nodes do not have any computation load, and the rest of nodes have the same computation and communication load as in the basic CDG scheme.

4.2 Proposed scheme for Distributed Compressive Data Gathering in WSN

By using Distributed Compressive Data Gathering for data aggregation in WSN, we may further reduce the number of measurements required for the reconstruction of the data. Here, we use CDG for multiple linear projections (M_1) of the sensed data. That is, we first take M_1 projections of the data and then, instead of collecting all M_1 readings from all the J sensors, we use the above mentioned CDG scheme for its transmission to the sink. Let

$$d_j = \phi_1 x_j, \qquad j = \{1, 2, \dots, J\}$$

where x_j is N×1 vector containing sensor readings at discrete times n=1, 2, …, N for j^{th} sensor, ϕ_1 is M_1×N measurement matrix and d_j is M_1×1 measurements vector for j^{th} sensor. Let each of these M_1 readings from J sensors form a vector $v_i = [d_{1i}\ d_{2i}\ \dots\ d_{Ji}]^T$, $i = \{1, 2, \dots M_1\}$. We then use CDG to collect M linear projections of these M_1 v_i's as

$$y_i = \phi v_i, \qquad i = \{1, 2, \dots, M_1\}$$

where, $\phi \in R^{M \times J}$ is the measurement basis and $y_i \in R^M$, $i = \{1, 2, \dots, M_1\}$. $Y = [y_1, y_2, \dots, y_{M_1}]$ is the aggregated data from all the J sensors of size M×M_1.

Subsequently, we recover x from y in two steps: first recover v from Y and then x from v using the recovery algorithms for DCS as discussed in Chapter 3.

4.3 Simulation Results

Compressive data gathering was simulated using FOCUSS and OMP. Plots of reconstruction error (Root mean square error) are shown for increasing number of measurements in Fig 4.3 and Fig 4.4 using the real signal data (ambient temperature readings) from EPFL WSN deployment LUCE (Available at: http://sensorscope.epfl.ch/) gathered from 24 sensors.

We see that by using only M=7 linear projections for J=24, we are able to reconstruct the data with less than 10% mean square error in case of FOCUSS as reconstruction algorithm. With OMP as recovery algorithm we can reconstruct the data with lesser error.

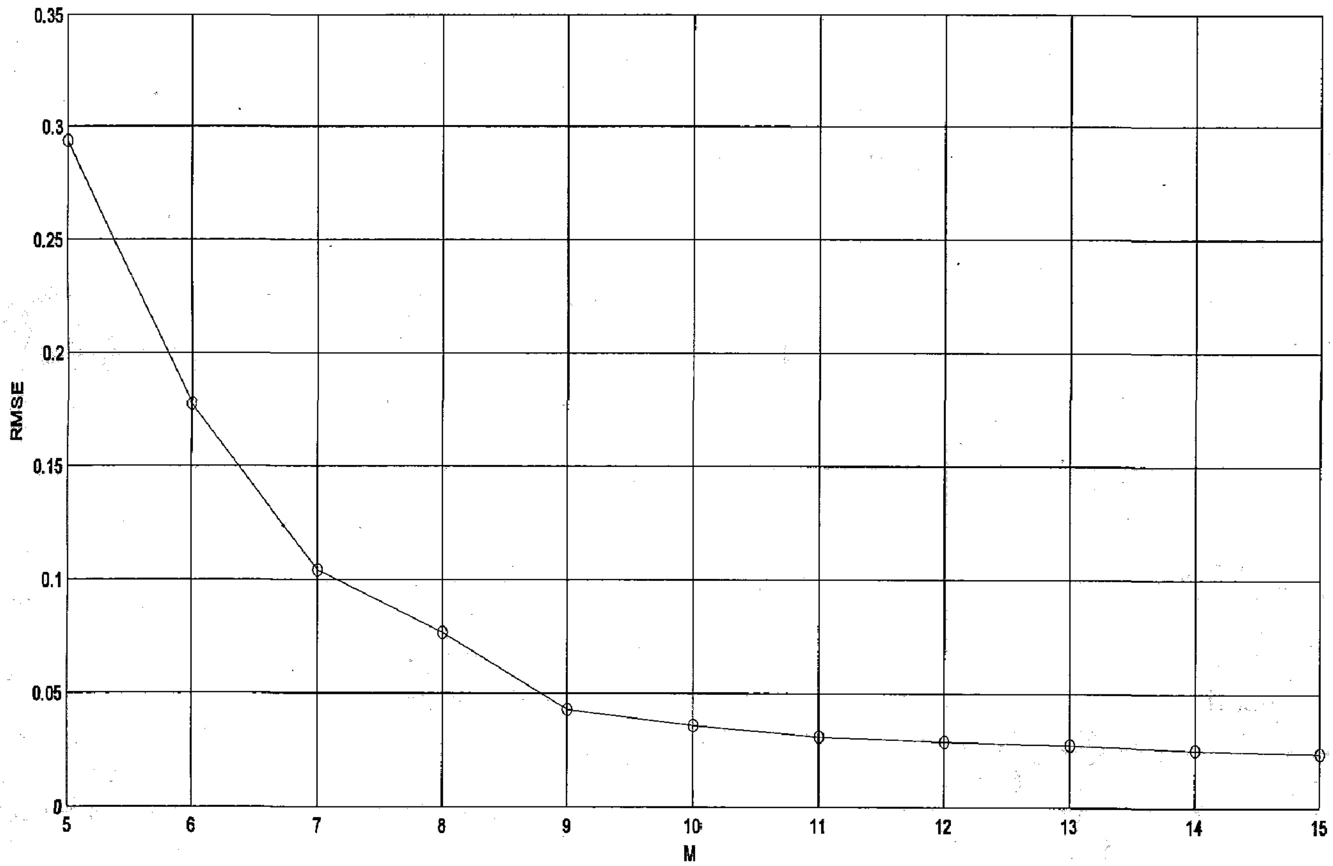

Figure 4.3 Plot of RMSE vs. M for CDG using FOCUSS. No of sensors, J=24.

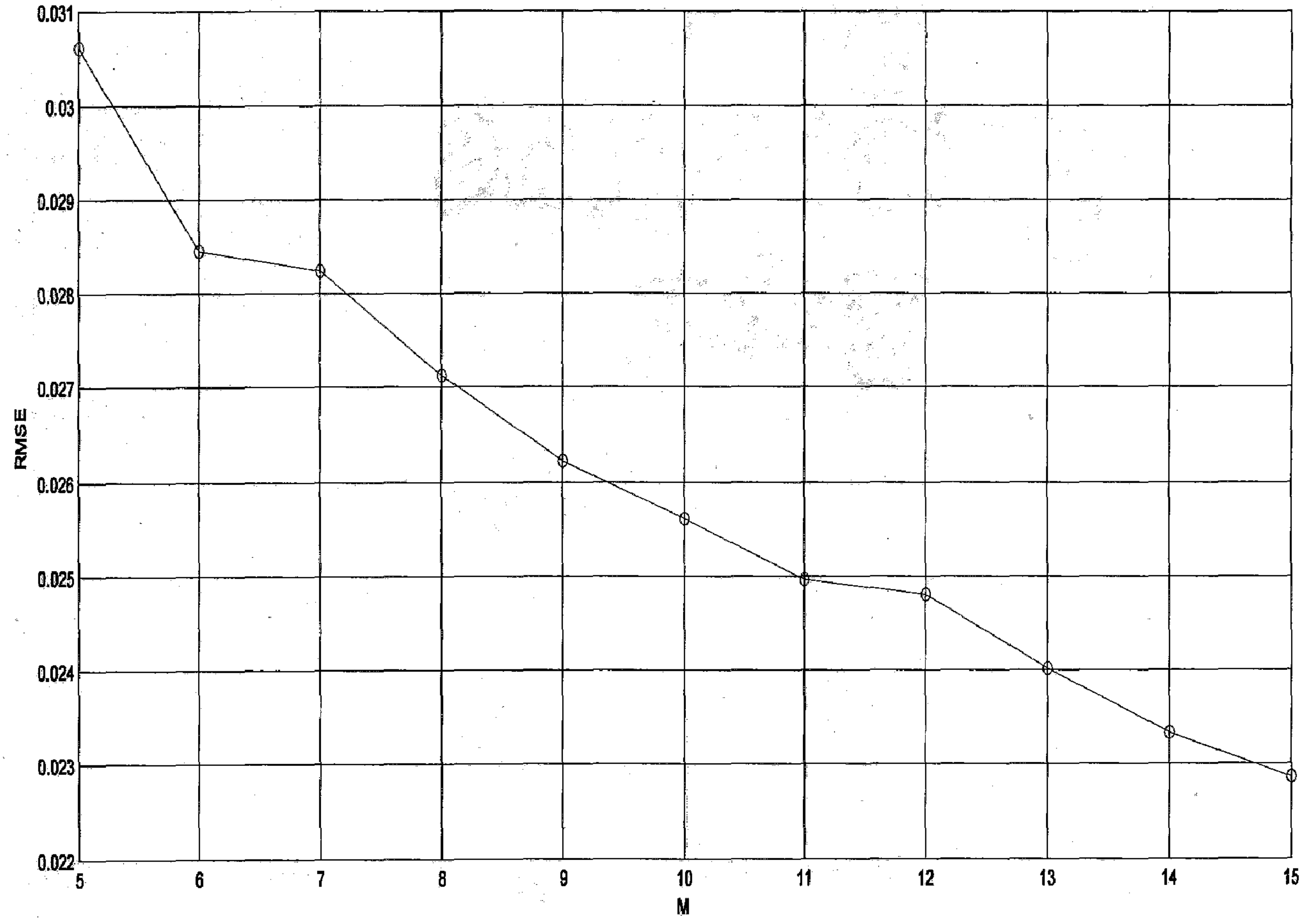

Figure 4.4 Plot of RMSE vs. M for CDG using OMP. No of sensors, J=24.

We, also, simulated IR-CDG using the set of CTD (Conductivity, Temperature, and Depth) data from National Oceanic and Atmospheric Administration's (NOAA) National Data Buoy Center (NDBC). Figure 4.5 shows the temperature data collected in the Pacific Sea at (7.0N, 180W) on March 29, 2008 (Available at: http://tao.noaa.gov/refreshed/ctd delivery.php). The data set contains 1000 readings obtained at different depth of sea, ranging from 4.579°C to 27.87°C. It is clear that the readings are piece-wise smooth, and should be sparse in wavelet domain. We use Haar wavelet for sparsifying transform and FOCUSS for reconstruction.

We observe that we have successfully reconstructed data using IR-CDG and minimum communication cost for multi-hop network (M=150 and M=200 for N=768).

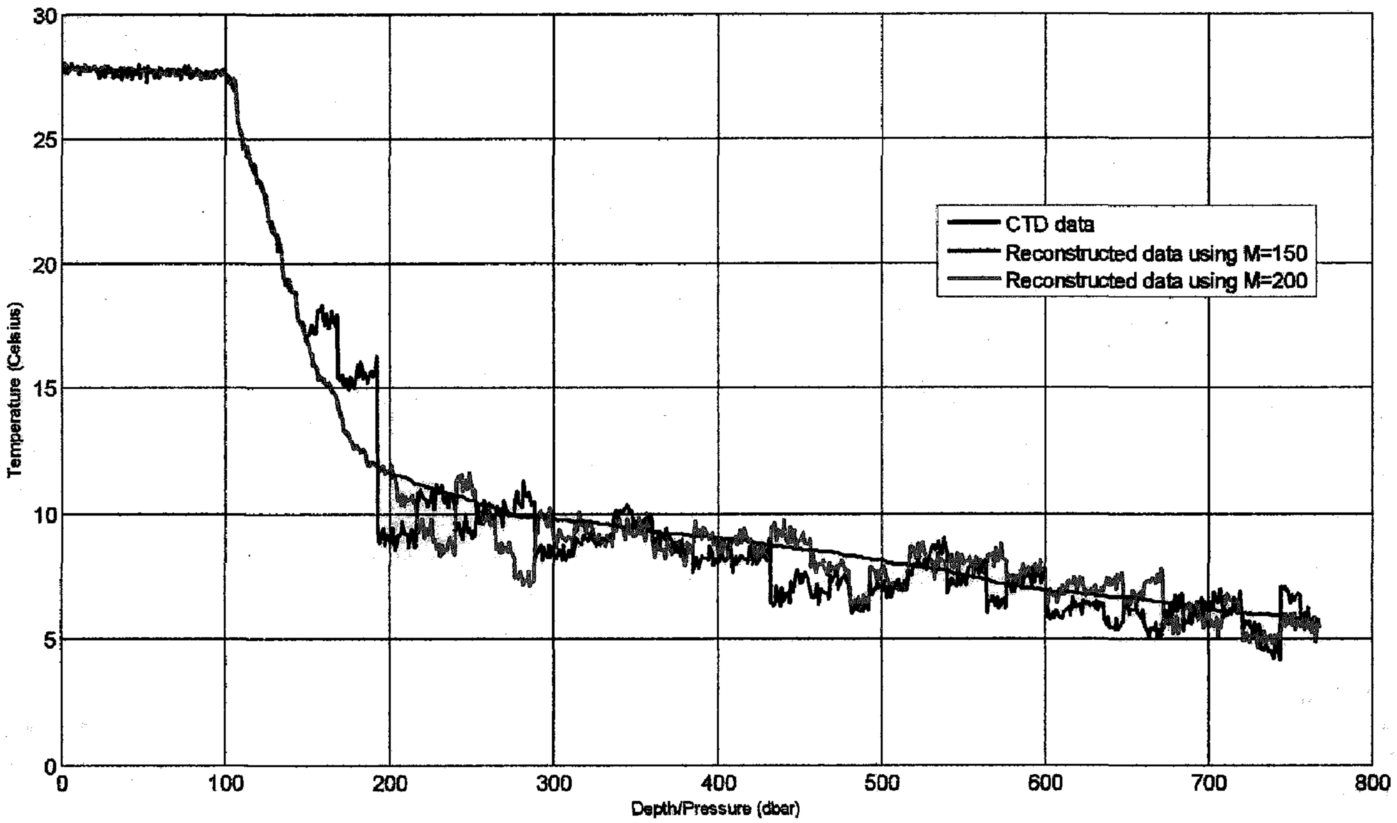

Figure 4.5 Figure showing actual and reconstructed temperature readings using IR-CDG.

We simulated the above proposed scheme (DCDG) using the real signal data from EPFL WSN deployment LUCE of Ambient Temperature (Fig 4.6) and Relative Humidity (Fig 4.7). Measurement matrices are generated with their entries being i.i.d. Gaussian random numbers drawn according to $\mathcal{N}(0,\frac{1}{M})$ and $\mathcal{N}(0,\frac{1}{M_1})$. We plot average root mean square error (RMSE) in the reconstruction of a signal of length N=512 from sensors J=24 for M=3 versus M_1 as shown in Fig. 4.6 and Fig 4.7. We use Haar wavelet transform for the sparse representation of the signals and use DCS-SOMP for reconstruction.

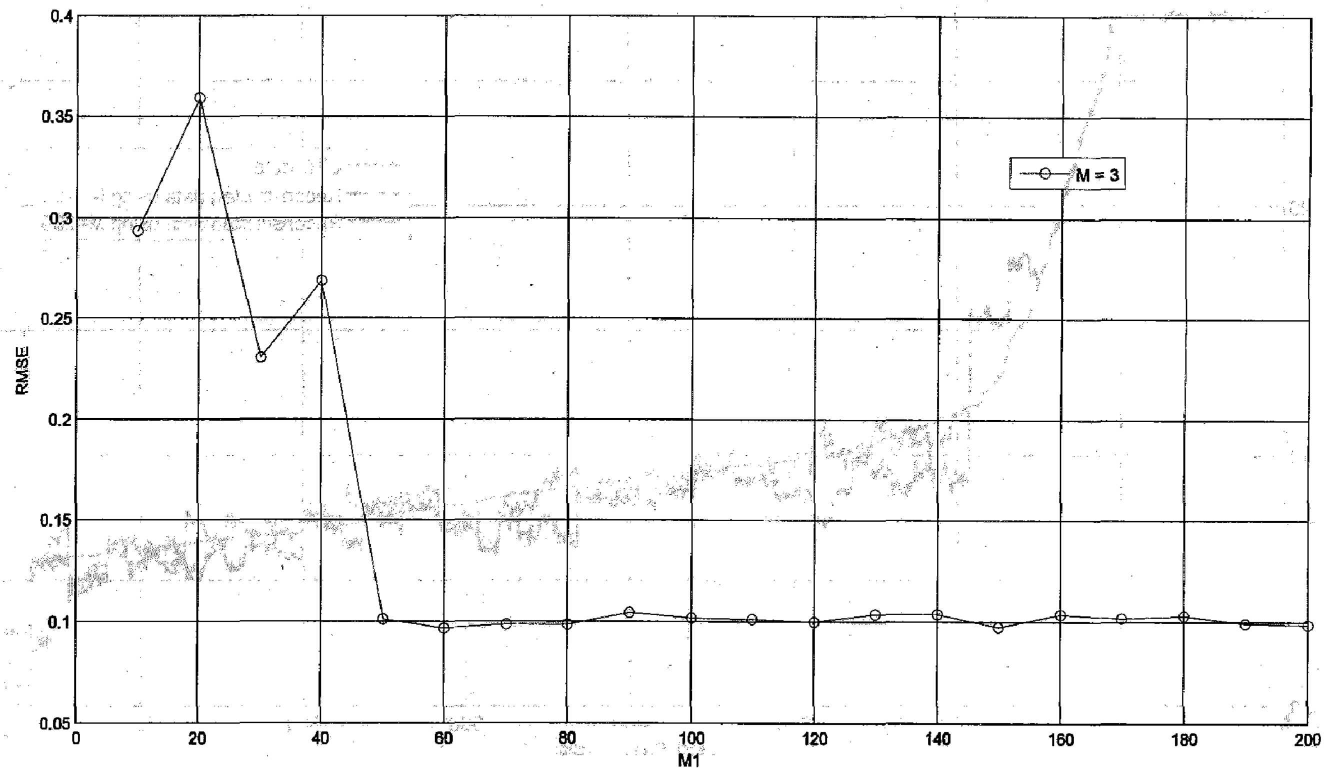

Figure 4.6 Plot of Root Mean square error v/s M for DCDG for ambient temperature. M=3, N=512, J=24.

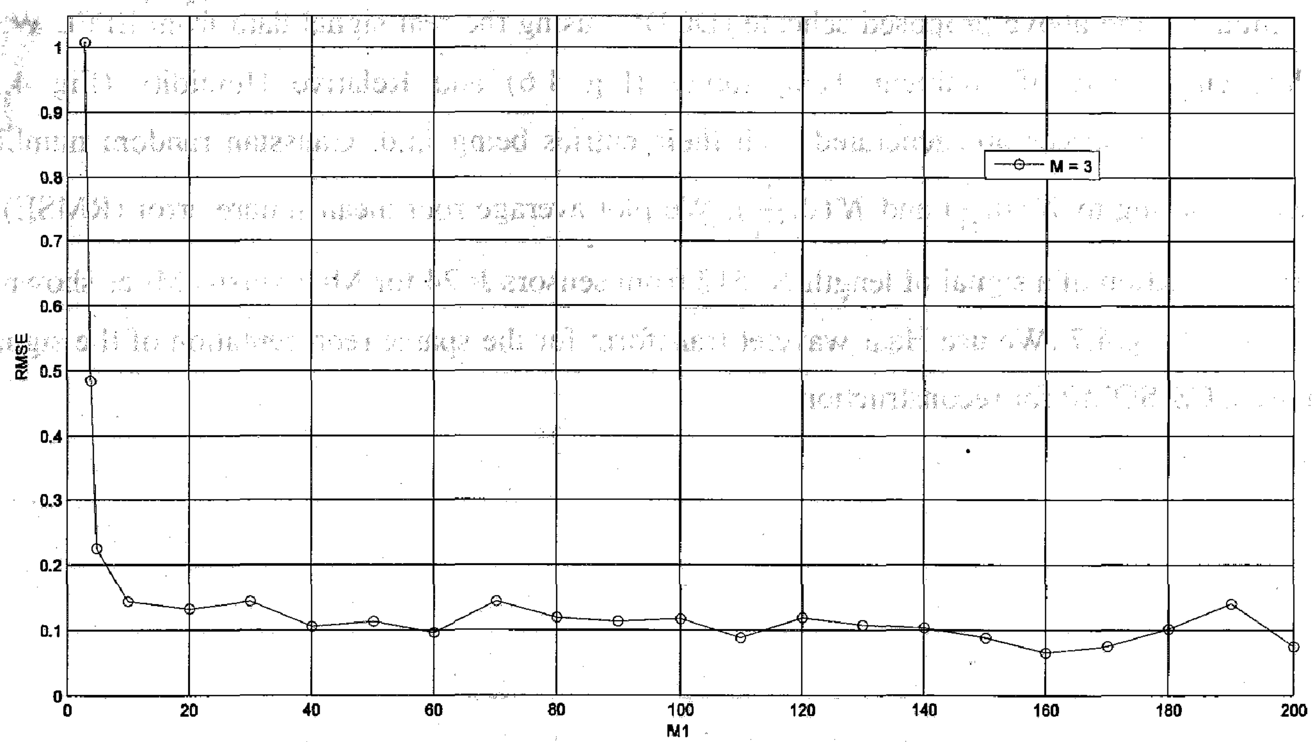

Figure 4.7 Plot of Root Mean square error v/s M for DCDG for relative humidity. M=3, N=512, J=24.

As we see from Fig. 4.6, reconstruction error (RMS error) for $M_1 \geq 50$ is around 10% for temperature readings. Hence, instead of sending 512×24 values, we may transmit only 50×3 values to the sink and reconstruct the data almost perfectly.

For relative humidity (Fig 4.7), with $M_1 \geq 10$ we can reconstruct data with approx. 15% error which means a drastic reduction in number of measurements and transmissions required for successful reconstruction.

4.4 Summary

This chapter mentions different data gathering schemes in WSN. Compressive Data Gathering (CDG) proposed by Luo et al. [16] has been discussed in detail. It leverages compressive sampling (CS) principle to efficiently reduce communication cost and prolong network lifetime. IR-CDG is another scheme for data gathering with even less number of measurements in WSN. Simulation results of CDG and IR-CDG are shown.

Thereafter, we proposed to use Distributed Compressive Data Gathering to further reduce the communication cost and observed that we are able to reconstruct a signal almost accurately with very less number of measurements.

5 Conclusion and Future Work

This dissertation attempts to study the emerging topic of Compressive Sensing and Distributed Compressed sensing and their application to Wireless Sensor Networks. With the increasing use of WSNs, it becomes necessary to efficiently utilize the scarce resources of power and bandwidth. DCS shows promising results for this by reducing the measurements to be acquired and transmitted using the spatial and temporal correlation structure of the data being sensed, i.e. inter- and intra-signal correlation. It shifts the processing from the sensors to the sink which often has more substantial computational resources. As from the results in Chapter 3, we see that a sparse or compressible signal may be reconstructed with only few linear projections of it on an incoherent basis. DCS-SOMP requires fewer measurements than M-FOCUSS for the reconstruction and provides better results. All the sensors may work independently and hence reduce the in-network communication. In-network communication is required only to support multi-hop networking to the data collection point.

This communication for data collection in multi-hop network may further be reduced if Compressive data gathering (CDG) is combined with DCS. In CDG, we use CS while collecting the data at a time instant from sensors and reduce the number of transmissions required, thereby, reducing the communication cost and prolonging the network lifetime.

From the results obtained in Chapter 4, we see that with Distributed Compressive Data Gathering, we may reduce the number of measurements both for the inter- and intra-signals to a large extent and still be able to reconstruct the signals with great accuracy.

We have also extended the use of Principal Component Analysis (PCA) for DCS. Simulation results show that we can reconstruct data successfully using the transformation matrix learnt through PCA. For N=64, we are able to reconstruct data for 6 monitoring rounds after 1 training phase.

In future, we may extend other algorithms like Basis Pursuit, Sparse Bayesian learning, StOMP etc. for DCDG. Many algorithms can be first extended for multiple measurement vectors and then used for DCDG. MMV recovery algorithms in [13] may also be extended to JSM-1 and JSM-3 models. DCDG may be combined with PCA for dynamically learning the transformation matrix for WSN data with time varying correlation. Hence, signals may be reconstructed through joint DCDG and PCA.

References

[1] D. Slepian and J. K. Wolf, "Noiseless coding of correlated information sources," *IEEE Transactions on Information Theory*, vol.19, no.4, pp. 471–480, July 1973.

[2] Z. Xiong, A. Liveris, and S. Cheng, "Distributed source coding for sensor networks," *IEEE Signal Processing Magazine*, vol. 21, pp. 80–94, Sept. 2004.

[3] E. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information," *IEEE Transactions on Information Theory,* vol. 52, no. 2, pp. 489–509, Feb. 2006.

[4] D.L. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory,* vol.52, no.4, pp.1289-1306, April 2006.

[5] E. Candès, "Compressive sampling," *Proceedings of International Congress of Mathematicians,* vol. 3, pp. 1433-1452, 2006.

[6] R. Baraniuk, "Compressive sensing [lecture notes]," *IEEE Signal Processing Magazine*, vol. 24, no. 4, pp. 118 –121, July 2007.

[7] E. Candès, and T. Tao, "Decoding by linear programming," *IEEE Transactions on Information Theory*, vol. 51, no.12, pp. 4203–4215, Dec. 2005.

[8] R. Baraniuk, V. Cevher, M. Duarte, and C. Hegde, "Model-based compressive sensing," *IEEE Transactions on Information Theory*, vol.56, no.4, pp. 1982 –2001, April 2010.

[9] S. Ji, Y. Xue, and L. Carin, "Bayesian compressive sensing," *IEEE Transactions on Signal Processing*, vol. 56, no.6, pp. 2346 –2356, June 2008.

[10] R. Masiero, G. Quer, D. Munaretto, M. Rossi, J. Widmer, and M. Zorzi, "Data acquisition through joint compressive sensing and principal component analysis," *IEEE Global Telecommunications Conference*, pp.1-6, Nov.-Dec. 2009.

[11] M.F. Duarte, S. Sarvotham, D. Baron, M.B. Wakin, and R. Baraniuk, "Distributed compressed sensing of jointly sparse signals," *Proceedings of the 39th ACSSC*, pp. 1537 – 1541, Oct. 2005.

[12] D. Baron, M.B. Wakin, M.F. Duarte, S. Sarvotham, and R. Baraniuk, "Distributed compressed sensing," *Preprint*, 2009.

[13] S.F. Cotter, B.D. Rao, K. Engan, and K. Kreutz-Delgado, "Sparse solutions to linear inverse problems with multiple measurement vectors," *IEEE Transactions on Signal Processing*, vol. 53, no.7, pp. 2477- 2488, July 2005.

[14] X. Wang, Z. Zhao, N. Zhao, and H. Zhang, "On the application of compressed sensing in communication networks," *5th International ICST Conference on Communications and Networking in China (CHINACOM)*, pp.1-7, Aug. 2010.

[15] W. Bajwa, J. Haupt, A. Sayeed, and R. Nowak, "Compressive wireless sensing," *The Fifth International Conference on Information Processing in Sensor Networks (IPSN)*, pp. 134–142, 2006.

[16] C. Luo, F. Wu, J. Sun, and C.W. Chen, "Compressive data gathering for large-scale wireless sensor networks," *Proceedings of ACM Mobicom*, pp. 145–156, Sep. 2009.

[17] S. Boyd, A. Ghosh, B. Prabhakar, and D. Shah, "Randomized gossip algorithms," *IEEE Transactions on Information Theory*, vol. 52, no. 6, pp. 2508–2530, June 2006.

[18] M. A. Davenport, M. F. Duarte, Y. C. Eldar, and G. Kutyniok, "Introduction to Compressed Sensing," *Compressed Sensing: Theory and Applications*, Cambridge University Press, 2012.

[19] S. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Transactions on Signal Processing*, vol.41, no.12, pp. 3397–3415, Dec. 1993.

[20] G. Davis, S. Mallat, and M. Avellaneda, "Adaptive greedy approximations," *Constructive Approximation*, vol. 13, pp. 57–98, 1997.

[21] D.L. Donoho, Y. Tsaig, "Fast solution of l_1-norm minimization problems when the solution may be sparse," *IEEE Transactions on Information Theory*, vol. 54, no.11, pp.4789-4812, Nov. 2008.

[22] J. Tropp, "Greed is good: algorithmic results for sparse approximation," *IEEE Transactions on Information Theory*, vol. 50, no. 10, pp. 2231 –2242, Oct. 2004.

[23] S.S. Chen, D.L. Donoho, Michael, and A. Saunders, "Atomic decomposition by basis pursuit," *SIAM Journal on Scientific Computing*, vol. 20, pp. 33–61, 1998.

[24] I. Gorodnitsky and B. Rao, "Sparse signal reconstruction from limited data using FOCUSS: a re-weighted minimum norm algorithm," *IEEE Transactions on Signal Processing*, vol. 45, no.3, pp. 600 –616, Mar. 1997.

[25] M. Elad, "Why simple shrinkage is still relevant for redundant representations?," *IEEE Transactions on Information Theory,* vol. 52, no. 12, pp. 5559 –5569, Dec. 2006.

[26] D. Wipf and B. Rao, "Sparse bayesian learning for basis selection," *IEEE Transactions on Signal Processing*, vol. 52, no. 8, pp. 2153 – 2164, Aug. 2004.

[27] E.G. Larsson and Y. Selen, "Linear regression with a sparse parameter vector," *IEEE Transactions on Signal Processing*, vol. 55, no. 2, pp. 451–460, Feb. 2007.

[28] G. Quer, D. Zordan, R. Masiero, M. Zorzi, and M. Rossi, "WSN-control: Signal reconstruction through Compressive Sensing in Wireless Sensor Networks," *IEEE 35th Conference on Local Computer Networks (LCN),* pp. 921-928, Oct. 2010.

[29] M.F. Duarte, "Compressive sensing for signal ensembles," Rice University, July 2009.

[30] J. Tropp, A. C. Gilbert, and M. J. Strauss, "Simultaneous sparse approximation via greedy pursuit," *IEEE International Conference on Acoustics, Speech, Signal Processing (ICASSP)*, vol. 5, pp. 721-724, Mar. 2005.

[31] C. Luo, F. Wu, J. Sun, and C.W. Chen, "Efficient measurement generation and pervasive sparsity for Compressive Data Gathering," *IEEE Transactions on Wireless Communications,* vol. 9, no. 12, pp. 3728-3738, Dec. 2010.