Data gathering and anomaly detection in wireless sensors networks

Mohamed Ali Moussa

Université Paris-Est

Lab. Informatique Gaspard Monge

UMR CNRS 8049

THESIS

presented by

Mohamed Ali Moussa

10 November 2017, Paris

Data gathering and anomaly detection in Wireless sensor networks

---

Reporter	Mohamed Mosbah	Prof. Institut Polytechnique de Bordeaux
Reporter	Nadjib Achir	Assoc.Prof. Institut Galilée Université Paris 13
Examiner	Rami LANGAR	Prof. Université Paris-Est Marne-la-Vallée
Examiner	Sandrine Vaton	Prof. Télécom Bretagne, Brest
PhD supervisor	Yacine Ghamri-Doudane	Prof. Université de La Rochelle

Resumé

L'utilisation des réseaux de capteurs sans fil (WSN)s ne cesse d'augmenter au point de couvrir divers domaines et applications. Cette tendance est supportée par les avancés techniques achevées dans la conception des capteurs, qui ont permis de réduire le coût ainsi que la taille de ces composants. Toutefois, il reste plusieurs défis qui font face au déploiement et au bon fonctionnement de ce type de réseaux et qui sont liés principalement aux limitations des ressources des capteurs ainsi qu'à l'imperfection des données collectées.

Dans cette thèse, nous nous intéressons au problème de collecte de données et de détection d'anomalies dans les réseaux de capteurs. Nous visons à assurer ces deux fonctionnalités tout en économisant l'utilisation des ressources des capteurs et en prolongeant la durée de vie de ces réseaux.

Tout au long de ce travail, nous présentons plusieurs solutions qui permettent une collecte efficace de données de capteurs ainsi qu'une capacité de détection des éventuelles anomalies. Notre objectif est de discuter celles-ci et de montrer leurs spcificités et leurs apports.

Dans notre première contribution, nous décrivons une solution basée sur la technique *Compressive Sensing* (CS) qui permet d'équilibrer le trafic transmis par les nœuds dans le réseau. Notre approche diffère des solutions existantes par la prise en compte de la corrélation temporelle ainsi que de la corrélation spatiale dans le processus de décompression des données. De plus, nous proposons d'intgrer dans ce processus une nouvelle formulation pour détecter les anomalies. Les simulations réalisées sur des données réelles démontrent l'efficacité de notre approche en terme de capacités combinées de reconstruction de données et de détection d'anomalies par rapport à ce qui est atteint par d'autres approches existantes.

Pour mieux optimiser l'utilisation des ressources des WSNs, nous proposons dans une deuxième contribution une solution de collecte de données et de détection d'anomalies basée sur la technique *Matrix Completion* (MC) qui consiste à transmettre un sous-ensemble aléatoire de données de capteurs. Nous développons un algorithme qui estime les mesures manquantes en se basant sur plusieurs propriétés des données. L'algorithme développé permet également de dissimuler les anomalies de la structure normale des données. Cette solution est par la suite substantiellement améliorée

dans notre troisième contribution. Dans cette dernière, nous proposons une formulation différente du problème de collecte de données et de détection d'anomalies. Plus précisément, Nous reformulons les connaissances à priori sur les données cibles par des contraintes convexes. Ainsi, les paramètres impliqués dans l'algorithme développé sont liés a certaines propriétés physiques du phénomène observé et deviennent donc mieux maitisable et ainsi plus faciles à ajuster. Nos deux approches montrent de bonnes performances par rapport à des approches concurrentes lorsqu'appliquées sur des données réelles.

Enfin, nous proposons dans la dernière contribution une nouvelle technique de collecte de données qui consiste à n'envoyer que les positions les plus importantes dans la représentation parcimonieuse des données. Nous considérons dans cette approche le bruit qui peut s'additionner aux données reçues par le nœud collecteur. Cette solution permet aussi de détecter les pics qui peuvent répresenter des changements brusques dans les mesures prélevées. Nous validons l'efficacité de notre solution par une analyse théorique corroborée par des simulations sur des données réelles.

Abstract

The use of Wireless Sensor Networks (WSN)s is steadily increasing to cover various applications and domains. This trend is supported by the technical advancements in sensor manufacturing process which allow a considerable reduction in the cost and size of these components. However, there are several challenges facing the deployment and the good functioning of this type of networks. Indeed, WSNs' applications have to deal with the limited energy, memory and processing capacities of sensor nodes as well as the imperfection of the probed data.

This dissertation addresses the problem of collecting data and detecting anomalies in WSNs. The aforementioned functionality needs to be achieved while ensuring a reliable data quality at the collector node, a good anomaly detection accuracy, a low false alarm rate as well as an efficient energy consumption solution.

Throughout this work, we provide different solutions that allow to meet these requirements. Foremost, we propose a Compressive Sensing (CS)-based solution that allows to equilibrate the traffic carried by nodes regardless their distance from the sink. This solution promotes a larger lifespan of the WSN since it balances the energy consumption between sensor nodes. Our approach differs from existing CS-based solutions by taking into account the sparsity of sensory representation in the temporal domain in addition to the spatial dimension. Moreover, we propose a new formulation to detect aberrant readings. The simulations carried on real datasets prove the efficiency of our approach in terms of data recovering and anomaly detection compared to existing solutions.

Aiming to further optimize the use of WSN resources, we propose in our second contribution a Matrix Completion (MC)-based data gathering and anomaly detection solution where an arbitrary subset of nodes contributes at the data gathering process at each operating period. To fill the missing values, we mainly relay on the low rank structure of sensory data as well as the sparsity of readings in some transform domain. The developed algorithm also allows to dissemble anomalies from the normal data structure. This solution is enhanced in our third contribution where we propose a constrained formulation of the data gathering and anomalies detection problem. We reformulate the *a prior* knowledge about the target data as hard convex constraints. Thus, the involved parameters into the developed algorithm become easy to adjust

since they are related to some physical properties of the treated data. Both MC-based approaches are tested on real datasets and demonstrate good capabilities in terms of data reconstruction quality and anomaly detection performance.

Finally, we propose in the last contribution a position based compressive data gathering scheme where nodes cooperate to compute and transmit only the relevant positions of their sensory sparse representation. This technique provide an efficient tool to deal with the noisy nature of WSN environment as well as to detect spikes in the sensory data. Note that we denote by a spike a brisk change in the probed values. Furthermore, we validate the efficiency of our solution by a theoretical analysis and corroborate it by a simulation evaluation.

Contents

Introduction

Context and Motivations

Over the last few years, the number of smart devices and sensors has seen an exponential growth due to the considerable advancement in the manufacturing process as well as the continuously extending number of applications relying on these devices. The number of sensors is expected to exceed 1 trillion by 2025 [Masden, 2014]. These devices are capable of measuring a variety of physical, chemical and biomass values, such as temperature, humidity stress, light, sound, magnetism and pH values.

Wireless Sensor Networks (WSNs) are in the center of this revolution benefiting from the advancement in the technologies behind the miniaturization of sensors and their capabilities to probe different magnitudes. The utilization of WSNs is widely spread among many applications where the main goal is to collect data from a large number of battery-powered, low computation and low memory sensor nodes that are randomly deployed in an area of interest. The aforementioned goal needs to be achieved while ensuring a high quality of the collected data as well as an efficient use of WSN resources.

On the other hand, because of the physical limitations of sensor devices and the unpredictable changes in WSN environments, the raw generated data is susceptible to the presence of various anomalies. Furthermore, sending all the probed data to the sink which is the collector node may be a costly operation in terms of WSN resource consumption. These inherent limitations lead to two main challenges: The first challenge comes from the limited energy supply of sensors devices. We need to find a suitable solution that enhances the lifespan of WSN and allows to efficiently gather the sensor readings and avoid sending the whole raw data. In fact, sensor nodes are usually deployed in harsh environments which render the recharging or the replacement of the exhausted batteries a difficult task. Therefore, optimizing sensors' resources while collecting data is of tremendous importance in WSNs. The second challenge comes from the imperfection of sensory data. Indeed, the raw generated data is vulnerable to faults, malicious attacks and missing values due to many reasons such as collisions, interference or unexpected node failures or damages. Thus, it is

crucial to develop efficient techniques to detect and correct the eventual anomalies.

Currently, most research in the WSN area has separately focused on these problems. Most data gathering techniques concentrate on optimizing WSN resources to deliver the probed readings to the sink and neglect the eventual presence of outliers. In the counterpart, anomaly detection techniques promote analyzing data and carry less attention to the data gathering process. The goal of this research is to develop a practical, autonomous and robust multitask solution that allows both, a reliable sensory data collection and an efficient detection of the potential anomalies.

Problem statement and challenges

WSN are often composed of a huge number of tiny devices with limited energy and computation capabilities. The evolution in the sensor manufacturing process has decreased their cost and has leaded to the expansion of WSN applications. However, mastering sensor energy consumption remains a relevant challenge. In fact, because of the harsh nature of some observed environments, recharging or replacing sensor's batteries is either a costly or an impossible task. Hence, optimizing the energy consumption at each senor node and thus, enhancing the network lifespan is a key challenge in WSNs.

Another big challenge is how to efficiently deliver the sensed measurements to the sink with a maximum fidelity to the probed data. Indeed, transmitting all sensors' readings to the collector node is either unpractical or impossible in most WSN systems because of the storage and energy capacity limitations of nodes. In fact, forwarding the probed data to the sink using a muli-hop path requires sensors that are near to the collector node to carry the measurements of their descendant nodes in addition to their own readings. Unfortunately, this trivial solution necessitates a huge data storage and energy capacities available at the sensors that are close to the sink to store and process the received data of their descendant nodes. Thus, the importance of providing a suitable data gathering solutions that overcome WSN limitations.

Finally, providing an efficient solution to prevent anomalies is of tremendous importance in WSN applications. In fact, the raw generated data are often susceptible to the presence of outlying values due to many reasons such as an abrupt changes in the vicinity of the sensor, perturbing the correct measurement, or a technical error in the sensor device itself. Detecting the potential anomalies from a large data set and alerting the data interpreter agent in the convenient time is of crucial importance in many applications such as fire alerting systems for instance.

CONTRIBUTIONS

In this work, we are interested in gathering the sensory data from WSN nodes while detecting the eventual anomalies. In order to provide the best trade-off between the end-user requirements and the network constraints, we provide answers to the following

questions throughout our contributions: how to efficient gather the sensory data? How to reliably identify the data anomalies? And how to achieve this task with an optimal use of WSN resources?

In this thesis, we investigate different approaches to answer these questions, leading to the four contributions described hereafter.

In our first contribution, we propose an on-line Compressive Sensing (CS)-based solution that simultaneously allows to collect data and estimate anomalies. At the difference of existing CS based approaches, our proposed scheme integrates both the spatial and temporal dimensions into the data recovery(*i.e.* decompression) and anomaly detection process. Moreover, we propose an innovative manner to take into account the outlying values without increasing the compression ratio. The association of data gathering and anomaly detection tasks into the CS scheme is made possible by using the primal-dual algorithm classes. In fact, we design a primal-dual based algorithm that allows to incorporate a multitude of data features in order to achieve the previously described goal. Our proposed algorithm is evaluated using real datasets and it demonstrates good data gathering and anomaly detection performance.

In the second contribution, we propose a Matrix Completion (MC) based solution that allows to achieve the same goal using a different data gathering scheme. In fact, the solution proposed in our first contribution requires a static routing scheme and do not offer a suitable tool to deal with the out of order nodes. We proposed in this contribution a dynamic routing scheme where at each operating time slot, only a subset of nodes participates at the data collecting process. Hence, after a predefined observation period, the collector node will dispose of an incomplete data matrix. Using some data features such as the low rank pattern and the sparsity of reading under some transform domain, we develop an algorithm that allows to fill missing measurements and detect the aberrant values. We evaluate the efficiency of our method by extensive simulations on two real datasets. Throughout these simulations, we demonstrate that the proposed algorithm achieves good data recovery and anomaly detection performance and outperforms the state-of-the-art techniques.

Despite that the two previous approaches provide an efficient solution to gather data and detect anomalies in WSNs, they rely on some regularization parameters. In fact, because we consider many data features to refine the data estimation result, The choice of each weighting parameter could affect the resulting solution, and there is no clear strategy to properly set these parameters in the data gathering and anomaly detection process. Hence, we propose in our third contribution an alternative approach that allows to directly incorporate the physical parameters related to the gathered data into the developed algorithm. More precisely, we propose a MC-based solution where the *a prior* knowledge about the target data is modeled via hard convex constraints. The involved parameters are hence related to some physical properties of the data itself and are thus easy to interpret. We develop a primal-dual algorithm to resolve the resulting constrained minimization problem. The simulation results carried on two real datasates demonstrate the robustness of our approach.

In our last contribution, we propose a new data gathering scheme that takes advantage of the structure of the sensory sparse presentation. In fact, In addition to the low rank and the sparsity features, we noticed that sensor spatial representations always share the same support. Starting from this observation, we develop a new data gathering scheme where nodes cooperate to compute and forward the most relevant positions of their sparse representation. The proposed scheme is extended to integrate also the temporal dimension in the data gathering process by applying the compression on both the temporal and spatial dimensions. The extended solution allows to reduce the frequency of data transmissions as well as a better data reconstruction quality. Furthermore, The noisy nature of WSNs is taken into account by considering the perturbations that may affect the sink received data. Finally, we develop a spike detection solution for both versions of our data gathering scheme by analyzing some energetic aspect of the received signal without the need of decompressing the received data. The proposed approach is shown to achieve good performance by running extensive simulations on the same two real datasets used in the previous evaluations.

ORGANIZATION

This work is composed of 5 chapters. Following this introduction, the next chapter is devoted to introduce the reader to the topic of data gathering and anomaly detection in WSNs. We present WSN characteristics and limitations. Then, we present some common solutions to reduce the amount of data carried along WSNs as well as the most known techniques to detect outlying values. Then, the subsequent chapters 2-5 address the problem of jointly gathering data and detecting anomalies in WSNs. These chapters are organized according to the logical contribution progress brought up by our solutions. More precisely, we present in chapter 2 our CS based solution that exploits both spatial and temporal dimensions in order to recover data and detect anomalies. We develop an optimization framework to deal with all considered data feature and we reinforce the efficiency of our solution by simulating different operating scenarios. In chapter 3, we introduce our first MC-based solution to achieve the same goal. We describe the resulting optimization problem and we evaluate the proposed algorithm to gather sensory data and detect the eventual anomalies. In chapter 4, we propose a constrained MC-based solution that extends the previous one by efficiently exploiting the side information about the data structure and we prove the robustness of our approach by running extensive simulation on real datasets. Thereafter, we present our position-based compressive data gathering solution where we describe the data gathering scheme as well as the developed solution to detect anomalies, whose are schematized by spikes, or abrupt changes in the sensory data. Finally, as a conclusion, we evoke our contributions and present some eventual perspectives that can be drawn from this work.

- Chapter 1 -

Background

1.1 Introduction

Since the emergence of WSN applications, the reliability in delivering a precise description about the observed environment becomes of a paramount importance. However, sensor readings are often exposed to a variety of disruptions that may affect the quality of the data. On the other hand, setting an efficient data gathering scheme poses many challenges related to the physical constraints of WSNs.

In this chapter, we present an overview of WSNs as well as the existing approaches to answer the above mentioned challenges. We provide a detailed description about the needed background to understand in depth the contributions presented throughout this thesis. We start by a brief description of WSNs, their applications and the shortcomings encountered in WSN-based systems. Subsequently, we start by presenting the existing data gathering solutions from the literature. Then, we describe the proposed approaches to detect and correct anomalies in WSNs. Thereafter, we present the use of inverse problems in WSNs and the class of algorithms used to tackle the problems approached in the context of this thesis.

1.2 Overview of Wireless Sensor Networks

A Wireless Sensor Network (WSN) is a special type of ad-hoc networks composed of a cooperating set of sensor nodes dispersed in a geographical area in order to monitor physical or environmental phenomenon in an autonomous manner. WSNs are essentially built around two main entities, namely the sensor nodes and the sink. The size of a WSN can vary from few hundred to thousand of connected nodes that communicate over wireless channels. Hence, each node is equipped with a radio transceiver, a processing unit, and at least one sensing unit that allows to probe different magnitudes such as temperature, humidity, pressure, light, etc. Sensor nodes are often powered by limited batteries or energy harvesting sources, and come in different shapes and sizes that may attain the scale of few millimeters. Each single-sensor readings are routed through other nodes based on the network connectivity architecture to reach the sink. Then, the data is treated and forwarded via Internet, or other communication means such as satellite communications, to the final user as depicted in Figure 1.1. Note that the sink is usually equipped with enough energy and computational resources to perform relatively complicated tasks.

Like most wireless communication systems, WSNs were initially developed for military purposes, exactly during the cold war. The first form of WSN was the United States project named the SOund SUrveillance System (SOSUS) deployed to track the Soviet submarines [Nack, 2010]. Then, WSNs found their way to many others military applications upheld by several well sponsored programs. The promising results of the driven projects as well as the evolution in computing and communication systems in the late 1990s raised the interest in WSNs and paved the way for many civil and

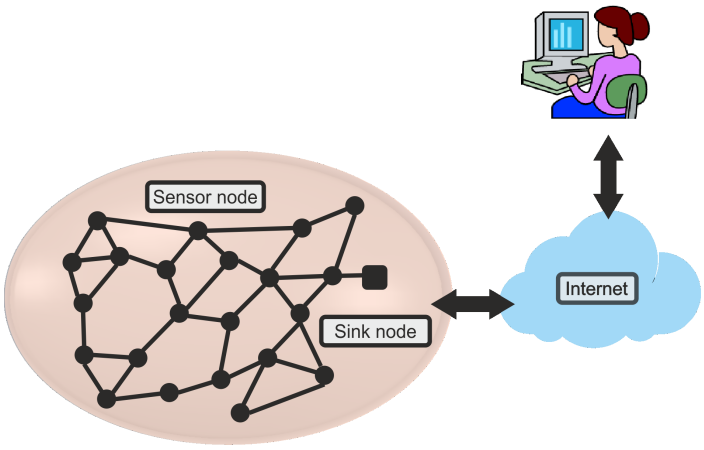


Figure 1.1: *Wireless sensor network architecture*

modern-day applications. The potential of WSNs and their multitude of commercial applications pushed the stakeholders and the manufacturing companies to set several standards such as ZigBee or WirelessHART which are both based on the IEEE 802.15.4 radio standard [Bell, 2004].

1.2.1 Applications of WSNs

Due to the recent advancements in WSNs, their use expands to a wide range of applications, in almost every type of environments, spanning from probing the underwater temperature to monitoring the air pollution. Many of these applications share the same goal of collecting data and detecting anomalies. In the following, we highlight some key applications of this work.

WSNs are used to study and monitor the seabirds on Skomer Island, a UK national nature reserve. The authors of [Mainwaring et al., 2002] design a system of 32 sensors deployed in this small island to investigate the behavior and the spatial ecology of some bird species. They describe the hardware and software design requirements as well as the required architecture to stream the collected data to the web. This system was further investigated in [Polastre et al., 2004]. The performance of the deployed system is evaluated by analysing over one million data readings related to some environmental and health parameters gathered from the deployed nodes. Based on some data mining tools, the authors demonstrate that they could predict some network malfunctionings. They also ameliorate the network design to enhance the performance of the deployed system toward network failures. Authors of [Naumowicz et al., 2008] and [Guilford et al., 2009] update the network hardware and software configuration to gather more parameters such as the movement at the entrance to the burrow or the presence of seabirds in the burrow tagged with RFID sensors.

The use of WSNs is not restricted to monitor isolated island, they are also used in

monitoring many of nowadays complex system such as electrical distribution system. Monitoring this later system consists on observing a multitude of physical parameters such as electricity voltage, power and frequency distortion. In [Lim et al., 2010], the authors describe a WSN system where each pole transformer is equipped with a sensor node that transmits its probed values to a collector node every second to detect and analyse possible system faults. The proposed system is composed of 1800 nodes divided into 6 clusters and communicating their data to a collector center. The challenges addressed in this system are: (i) Gathering data at high data sampling rates, (ii) maintaining a good functioning of the electrical distribution system.

WSNs are also used to monitor water pipelines installation. In [Whittle et al., 2013], the authors describe a WSN system installed in Singapore to locate and detect water leakage. This consists of 50 sensors nodes deployed in an area of over 80 km square to sense many pipeline physical parameters such as the pressure and the water flow. Then, the measured data is forwarded to a collector node where it is processed and analysed to identity the potential water leaks and to locate the hot spots that are more fragile in order to take preventive measures.

The author of [Li and Liu, 2007] describe a WSN system deployed in coal mine in order to monitor the inner surface of tunnels based on the radio signals probed by sensor nodes. The aim of the proposed system is to prevent underground collapses in coal mining by monitoring the brisk changes in the tunnel structures. The provided solution consists on predicting the potential changes based on the gathered data without including any prior knowledge about the structural changes. Another interesting example is given in [Werner-Allen et al., 2006] where a WSN is used to monitor the volcanic activities at Volcan Tingurahua in central Ecuador. The considered network is composed of a multiple of acoustic sensors that continuously transmit their data to a distant base station. Authors of [Werner-Allen et al., 2006] set up some anomaly detection tools to deal with the imperfection of collected data in the same context. For instance, they use the linear regression techniques to fill the missing readings.

To sum up, all these systems, which we take as examples, as well as many other WSN systems share the same goal of collecting sensor measurements in a central node over a long period of time or detecting the possible anomalies. However, the proposed solutions almost always focus on the quality of the gathered data or the anomaly detection capabilities rather than on optimizing WSN resources while ensuring these functionalities.

1.2.2 Shortcomings in WSNs

The size of the sensors can vary from a few centimeters going down to a few millimeters. Naturally, this limits the capacities of sensors in terms of computing, storage, and communications, despite the remarkable advancement in the miniaturization technology. Due to these constraints, we can identify the following shortcomings in WSNs:

- Resource limitations: This constraint comes from the individual limitations of

sensors. The most restricting limitation is the sensors' energy capacities. In fact, the lifespan of the entire network depends on the energy level of the deployed nodes. Indeed, since sensors participate in the data routing process and the network self-organization, the unavailability of some nodes due to power shortage perturbs the good functioning of the whole network. Hence, sensor nodes have to cautiously manage their batteries by limiting the computational and storage tasks and by reducing their transmission rates, while maintaining a good service quality.

- Communication failures: This constraint comes form the vulnerability of radio communications. In fact, the communication channels in WSNs are exposed to many source of perturbations such as moving objects in the observed area or external noise caused by external electrical fields. Collision and interference between nodes may also result in missing readings and can seriously reduce the quality of the collected data.

- Reliability issues: Some sensors can generate faulty readings or transmission errors because of insufficient batteries' level, interference and other physical problems. Moreover, due to the wireless communication channel nature, WSNs are vulnerable to external attacks and easily infiltrated by external listeners/ actors. Typical encryption techniques are greedy in energy resources and hence impractical for WSN systems.

To overcome these shortcomings, many optimized data gathering and anomaly detection techniques were proposed. In the following, we describe the most common techniques.

1.3 Data reduction techniques

In this section, we detail the most relevant data reduction techniques which consist on reducing the amount of data produced, processed and transmitted by sensors while maintaining a good quality level of the gathered data.

1.3.1 Data-compression-based techniques

The correlation between sensory data can be exploited to communicate a compressed form of the probed values. This class of data reduction techniques relies on information theory tools and require an *a priori* knowledge about the probability distribution of probed data. The authors in [Cristescu et al., 2006] propose an entropy coding scheme that exploits the entire available data at sensor nodes to reduce the communication overhead. The proposed solution requires a static routing strategy. In fact, Due to the correlation between readings, entropy coding schemes postulate that we can encode the readings using only few bits by taking into account other nodes' received messages.

However, entropy-based solutions require large sequence of readings to encode the transmitted messages and demand a considerable computation resources to perform the coding and decoding operations.
In [Acimovic et al., 2005; Ciancio et al., 2006], an interactive compressive scheme is proposed where nodes are allowed to exchange messages in order to reduce the resulting size of the transmitted information. The proposed solution relies on Wavelet Transform (WT) decomposition. Although the proposed solution achieves a good in-network compression ratio, it requires advanced computational capacities and a large amount of data exchanges.
Another appealing solution is proposed by distributed source coding techniques [Chou et al., 2003; Cristescu et al., 2004; Hua and Chen, 2008]. The idea here consists on the use of an advanced form of entropy coding schemes where nodes are allowed to take advantages of the global data correlation structure of the whole network data. The proposed approaches rely on the Slepian-Wolf coding theory [Slepian and Wolf, 1973] which postulates that we can separately encode the sensed data at each node and reduce the compression ratio while taking advantage of the correlation between sensors' measurement probability distributions. However, the sink can not decode each node message separately and it has to wait until all nodes' messages arrive to jointly perform the decoding process and extract the useful information. Slepain-Wolf coding schemes have the major advantage of decoupling the coding and routing processes. Unfortunately, they require an *a prori* knowledge about the probability distribution of all sensors' data which remains a tough task in WSNs, especially when the number of nodes goes high.

1.3.2 Prediction-based techniques

Since the sink disposes of the historical collected data, it is possible to predict the future observations without the need of regular data transmissions from network nodes. Prediction-based data gathering solutions are built around the possibility of constructing a data model that can meet the system requirements and estimates the probed data without extensively soliciting the sensing device resources.

An example of data prediction technique is introduced in [Santini, 2006; Santini and Romer, 2006] where the authors present an online solution named AMS to collect sensory data. The prediction model is constructed locally at each node given the assumption that the withdrawn values are accepted within an error margin. Sensor nodes choose the most statistically adapted model among a set of candidate models and transmit it to the sink to predict the future probed values. The choice of the adequate model is performed in order to achieve the best trade-off between the selection complexity, the communication requirements and the quality of the collected data. However, AMS presents some disadvantages since it requires to locally construct a multitude of models that define different data behaviors which is a hardly achievable task in WSN monitored environments. Moreover, locally commuting and storing all possible models can overstep the capacity of WSN nodes.

Another similar solution is proposed by [Chu et al., 2006] where sensor nodes construct a dynamic spatial data prediction model based on Markov chains. This solution has the main advantage of computing the dynamic model in a distributed manner without imposing regular sensor to sensor communications to update the prediction template. The proposed solution operates as follows: the dynamic probabilistic model is stored at both the sink and the sensor nodes and it is kept updated by exchanging some synchronization messages. The predicted values generated by the sink are used as an approximation of the real sensed data. Whereas, nodes run the prediction model and compare the predicted values to the probed ones. If the resulting prediction error exceeds the accepted margin to ensure a good functioning of the system, then the probed data is transmitted to the sink and the dynamic prediction model is updated. However, we can identify two main drawbacks of this solution. The first one is related to the high complexity required to compute and update the dynamic model which surpasses the computation capabilities of sensor nodes. The second shortcoming comes from incorporating Markov chains in the model construction process. Indeed, Markov chains are more efficient in predicting linear behavior and do not fit for the nonlinear and dynamic nature of most sensed phenomena and observed environments.
In [Goel and Imielinski, 2001], Goel et al propose a prediction-based data gathering solution inspired from image processing techniques, namely the MPEG coding technique which allows to enhance the quality of image and decrease its storage size based on some efficient coding and compression tools. To reconstruct the appropriate data model, nodes start by diffusing their readings to the sink. The later computes the prediction model based on the MPEG technique and return it to sensors. By disposing the suitable model, nodes only forward the readings that lie outside the acceptable error bounds to save their energy resources.
A similar approach is proposed by Jain et al in [Jain et al., 2004], where the MPEG model is replaced by a Dual Kalman Filter. In this case, nodes initiate the data prediction process by forwarding their readings to the sink, which make it able to train the filter and estimate the accurate parameters. After that, the calculated parameters are transmitted to local nodes. If the filter output diverges regarding the real sensed data, then sensors transmit their readings and alert the sink to update the filter parameters. Note that filter data prediction methods are not restricted to the use of Kalman filter, other filer models are used such as adaptive filters [Ali et al., 2008].

Despite the fact that prediction-based techniques ensure a god quality of the collected data, they still unpractical for most WSN applications since the construction of generic data model is not an easy achievable task. Moreover, they require a considerable computational capacities at sensor nodes.

1.3.3 Compressive sensing-based techniques

Compressive Sensing (CS) is a signal processing technique that allows to reconstruct a signal from far fewer samples than the required number by the Shannon theorem. This reconstruction is made possible by solving a linear convex problem under the

condition that the coded signal is sparse in some transform domain, *i.e.*, the resulting transformation of the signal contains only few nonzero entries. The high compression rates promoted by the CS theory drive it to data gathering applications. A multitude of attempts have been proposed to adapt the CS technique in the context of WSNs. The most typical solution consists on balancing the traffic carried by nodes by transmitting a spatially projected data. Hence, whatever the distance from the sink is, all nodes carry the same size of information. After receiving the coded information, the sink proceeds by recovering the probed data by solving an inverse optimization problem. In [Chong et al., 2009], authors introduce one of the first CS-based data gathering scheme. In [Luo et al., 2010], a hybrid CS scheme is proposed where only nodes from a certain distance of the sink apply the CS technique to compress their traffic. This technique is proved to achieve a significant gain in terms of throughput. The authors of [Xiang et al., 2013] propose a proper sparse basis design based on wavelets diffusion for arbitrary sensor networks. The proposed solution is shown to achieve good performance. The authors in [Zheng et al., 2013] extend the CS data gathering solution to the case of WSNs with multiple sink collectors and develop a specific routing scheme that allows to accommodate such an extended CS solution.
CS theory is also employed in data gathering in vehicular networks as shown in [Wang et al., 2013] where the historical readings are used to define the sparsity level of sensory data and hence the dimension of each vector navigating through the networks. The authors in [Wang et al., 2012] track the problem of data gathering using CS technique from another perspective where they tried to vary the size of the transmitted vector according to the instantaneous sparsity level of probed data. In fact, the sparsity of sensory readings varies in time and space. However, this proposed technique increase the computational charge on sensor nodes.

Although CS-based data gathering techniques provide an energy efficient solution to collect the probed data, there are some issues when applying CS in practice. In fact, these methods do not suit for low sampling rates since they demand that the minimum number of transmitted data must exceed a certain threshold, depending on the sparsity of the transform of measurements. Furthermore, CS-based technique are quite sensitive toward abnormal readings [Chong et al., 2009].

1.3.4 Incomplete-data-based techniques

Gathering data in WSNs can be achieved by using missing value estimation techniques. These techniques operate as follows. An arbitrary subset of nodes transmit their readings to the collector node at each gathering period. Then, the sink estimates the missing values using some an *a priori* knowledge about the target data. The problem of missing readings is addressed in many works. In [Zhao et al., 2003], the authors propose a communication level solution to deal with missing values. The proposed solution requires to retransmit missing values which may cost additional transmission power and causes some delays in the process of data gathering. Another solution is proposed to estimate missing values in [Fletcher et al., 2004]. The proposed solution

is based on modeling the targeted problem as a linear system and on using Kalman filter for estimating missing data. In this scheme, the filter is trained to determine the adequate regression model in order to fill missing readings. However, this solution requires a training data set and considerable communication and storage capacities. A simplified solution is proposed in [Werner-Allen et al., 2005] where the authors propose an autoregressive model to predict the missing values. The model parameters are set according to the historical collected data.
A more sophisticated approach is proposed by Matrix Completion (MC) based solutions. These solutions take advantage of the data low rank feature which consists on assuming that the data matrix is supported by few dimensions. In [Cheng et al., 2013], low rank and short term stability features are used to estimate the data matrix. The introduced algorithm takes into account only non-negative measurements, and the low rank feature is approximated using the Frobenius norm which is convex and smooth but not a tight approximation of the low rank structure. In [Yi et al., 2015], a variant of proximal gradient methods is proposed to enable the sink estimating the missing readings. The proposed algorithm takes advantage of the low rank structure and the compactness of the sparse representation of measurements under the Discrete Cosine Transform (DCT) domain. Again, the main limitations of the missing-data-based techniques reside in the fact that they do not offer an appropriate tools to deal with the presence of anomalies.

1.4 Anomaly detection in WSNs

The performance of WSNs relies on the quality of the collected data. Most applications set up a certain data accuracy threshold in order to take the most accurate decisions and actions regarding the collected data. An inappropriate information may lead to clumsy reactions and inadequate decision which may load and solicit node capacities. As an example, a fake fire alert would set on home sprinklers or an unreliable health sensor information may lead the patient to take irrelevant medicine and worsen the situation. Hence, providing a suitable tool to deal with the imperfection of data is of tremendous importance to ensure that application requirements are properly meet.

Sensory data in WSNs are often exposed to a variety of perturbations that induce to faulty and inaccurate measurements. The probed values that diverge from the normal behavior are often called *outliers*. Authors of [Hawkins, 1980] define outliers as *an observation, which deviates so much from other observations as to arouse suspicions*. Whereas Barnet et al. [Barnett and Lewis, 1974], define outlier as *observation (or subset of observations) which appears to be inconsistent with the remainder of that set of data*.

Since WSNs are widely used in many applications, a variety of outlier detection techniques had been proposed depending on the requirements of the deployed system. We can find anomaly detection solutions in credit card systems for fraud detection, military system for enemy recognition targets and insurance and health systems, etc.

In all these applications, the boundary between outling readings and normal behavior differs according to the specificities of ensured functions. Therefore, it is important to integrate the application requirements and the environment features into the outlier detection solution. So, the deployed system could dissemble between normal readings and anomalies while meeting the required performance.

Anomalies can occur due to a variety of reasons. Since sensors dispose of limited energy resources, abnormal readings frequently appear when the batteries reach low energy levels. Moreover, the transmission noise as well as the environmental obstacles can deteriorate the quality of the data [Chandola et al., 2007]. Abnormal readings can also occur due to the presence of an event.

The problem of identifying and correcting anomalies had been tackled using many approaches. All existing solutions share the same goal of meeting system requirements while maintaining a low WSN energy consumption level. The identification of outlying values can be made either locally at each sensor node or centrally at an analyzing and monitoring node. The authors in [Bettencourt et al., 2007; Janakiram et al., 2006; Wu et al., 2007] develop some local anomaly detection techniques where each sensor could detect the presence of anomalies by analyzing its past observations. In some cases, observing only one sensor reading is not enough to dissemble an outlying situation from a normal regime. Therefore, sensors recourse to their neighboring nodes to enhance their knowledge about the probed data. The detection operation can be shifted to the sink as described in the works of [Subramaniam et al., 2006; Sheng et al., 2007]. Hence the collector node disposes of enough information to separate normal data from the potential perturbations. However, such approach may be costly and exhausts the WSN resources. The authors in [Subramaniam et al., 2006] propose an alternative approach where the network is organized into clusters and the detection process is performed in each cluster head. The proposed solution lessen the overhead over WSN resources, but it remains costly in terms of resource utilization, especially when the WSN is large.

Regardless the network level where the anomaly detection process is performed, most of the proposed approaches are inspired from existing solutions in the data mining field and they are based on statistics, clustering and machine learning tools. In the following, we present the most common approaches.

1.4.1 Statistical-based Approaches

This type of approaches is built around statistical models where probability distributions are inferred to describe both normal and abnormal regimes. The construction of the probabilistic model can be achieved by using parametric based approaches such as Gaussian and non-Gaussian models [Bettencourt et al., 2007; Jun et al., 2005; Wu et al., 2007] where the data is supposed to fit a known distribution, or non parametric based approaches such as Kernel-based and histogram-based approaches [Palpanas et al., 2003; Sheng et al., 2007]. Non parametric approaches are more convenient to WSN applications since the underlying model is tailored according to the probed data

and the outlying values are detected if they exceed a certain threshold distance from the constructed model. The authors of [Paschalidis and Chen, 2010] propose a statistical anomaly detection approach. They model the data structure using Markov chains. Anomaly detection rules are developed for each Markov model based on the probability law inferred form the past traces. In [Magán-Carrión et al., 2015], a multivariate statistical approach for anomaly detection is proposed. The proposed solution is based on multivariate statistical process control and partial least square tools.
However, statistical-based approaches require large sequences of training readings to infer the probability distributions of normal and abnormal regimes and do not offer an appropriate tool to deal with eccentric data.

1.4.2 Nearest-neighbor-based Approaches

This class of approaches uses some distance metrics such as the Euclidean distance or the Mahalanobis distance to measure the gap between the probed value and the remaining set of data [Knox and Ng, 1998; Ramaswamy et al., 2000]. These techniques operate as follws: For each data instance, the k-nearest-neighbors have to be found. Then, an anomaly score is computed using these neighbors. Note that there are two strategies to compute the anomaly score : Either the distance to the k^{th}-nearest-neighbor is used (a single one) [Ramaswamy et al., 2000] or the average distance to all the k-nearest-neighbors is used [PKDD. et al., 1998]. If the distance that separates the probed measurement form its K-neighboring values exceeds some predefined threshold, then it is declared as outlier [Chandola et al., 2007]. The k-nearest-neighbor approaches are efficient in detecting global anomalies but it fails to detect outliers with small magnitude. Moreover, in practice, its not straightforward to set the appropriate threshold as well as the adequate number of neighbors. Note that k-nearest-neighbor anomaly detection approaches are different form k-nearest neighbor classification techniques which consist on assigning each data instance to a class membership by observing the k-nearest neighboring values in the training dataset.

1.4.3 Clustering and classification-based Approaches

To detect the potential anomalies in WSNs, we can apply clustering-based approaches by grouping the probed data into clusters according to their similarity. The points that do not fit in any conventional cluster are labeled as anomalies [Jain and Dubes, 1988]. In [Xie et al., 2011], a distributed anomaly detection approach is described. The proposed solution is based on a distributed cluster-based anomaly detection algorithm. The communication overhead is minimized by clustering the sensor readings and merging clusters before sending a description of the clusters to other nodes. Another appealing classification-based solution is proposed by Support Vector Machines (SVM)-based technique which is a semi-supervised anomaly detection approach [Rajasegarar et al., 2006] where the decider node learns the normal behavior regime form a training data set. Then, any instance that appears to diverge form the anomaly-free

model is declared as outlier. Authors of [Garcia-Font et al., 2016] propose a comparative study of anomaly detection techniques in the context of smart city WSNs. They demonstrate that one-class Support Vector Machines is the most appropriate technique for this kind of WSN applications. In spite the fact that SVM-based solutions offer good performance in detecting outliers, these techniques require high computational capacities, which play against their use in WSNs.

1.4.4 Spectral decomposition-based approaches

This class of techniques consists on finding the attributes or spaces that best represent and capture the variation of WSN data. Hence, outlying values can be easily dissembled since they do not obey to the normal structure. Principal Component Analysis (PCA) is a common technique used in data reduction and anomaly detection context [Chatzigiannakis et al., 2006]. The aforementioned technique proceeds by computing the most relevant direction in the data space. The principal components are the eigenvectors of the covariance matrix and thus their computation may be expensive [Sheng et al., 2007], especially when the dimension of the considered data matrix is relatively high. An instance is declared as anomaly if it deviates from the normal subspaces. Moreover, these technique require the availability of the whole sensory data to compute the covaraince matrix and thus the normal subspaces.

1.5 Inverse problems in WSNs

Most described data reduction techniques as well as anomaly detection techniques can be seen as the solution of some inverse problems where we aim to estimate an unknown original signal from a transformed (or incomplete) version of it. Mathematically, let the original signal denoted by the vector $\mathbf{x} \in \mathbb{R}^N$, and the received observation denoted by the vector $\mathbf{y} \in \mathbb{R}^M$. The two signals are related by the following expression:

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{w} \tag{1.1}$$

Where $\mathbf{H} \in \mathbb{R}^{N \times M}$ is a bounded linear operator that represents the system transformation. As an example, $\mathbf{H}$ can model the sensing matrix in CS-based data reduction techniques. The vector $\mathbf{w} \in \mathbb{R}^M$ is the additive noise perturbation that represents an acquisition or model errors.

Our objective is to produce an estimate $\hat{\mathbf{x}} \in \mathbb{R}^N$ of the original signal $\mathbf{x}$ based on the received signal $\mathbf{y}$ and the matrix $\mathbf{H}$.

One common approach to find the optimal solution of inverse problems is to define $\hat{\mathbf{x}}$ as the minimizer of a sum of two functions [Demoment, 1989]:

$$\hat{\mathbf{x}} \in \underset{\mathbf{x} \in \mathbb{R}^N}{\text{Argmin}}\ h(\mathbf{x}) + g(\mathbf{x}), \tag{1.2}$$

where $h : \mathbb{R}^N \to]-\infty, +\infty]$ is the data fidelity term that reflects a directly related information to the treated problem such as the appropriate statistic estimator interpreted from the probability distribution of $\mathbf{w}$. The function $g : \mathbb{R}^N \to]-\infty, +\infty]$ represents the data regularization term (penalization term) that allows to integrate the *a priori* knowledge about the target signal $\mathbf{x}$ such as the sparsity of $\mathbf{x}$ under some transform domain. For a better understanding of the functions involved in (1.2), we describe in the following a Bayesian estimator that relates the signal probabilistic models to these parameters.

1.5.1 Maximum A Posteriori estimator

Assume that $\mathbf{x}$ is a realization of random vector $\mathbf{X}$ and $\mathbf{y}$ is a realization of random vector $\mathbf{Y}$. The random variables $\mathbf{X}$ and $\mathbf{Y}$ are supposed independent. Let $p_{\mathbf{X}|\mathbf{Y}=\mathbf{y}}$ denotes the *a posteriori* probability density function of $\mathbf{X}$ given $\mathbf{Y} = \mathbf{y}$. Supposing that we have an *a priori* knowledge about the desired solution, the inverse problem can be solved by using the Maximum *A Posteriori* (MAP) estimator:

$$\hat{\mathbf{x}} \in \underset{\mathbf{x}\in\mathbb{R}^N}{\operatorname{argmax}}\ p_{\mathbf{X}|\mathbf{Y}=\mathbf{y}}(\mathbf{x}). \tag{1.3}$$

By using the Bayes rule, we can rewrite (1.3) as follow:

$$\hat{\mathbf{x}} \in \underset{\mathbf{x}\in\mathbb{R}^N}{\operatorname{argmax}}\ p_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}(\mathbf{y})\frac{p_{\mathbf{X}}(\mathbf{x})}{p_{\mathbf{Y}}(\mathbf{y})}, \tag{1.4}$$

$$\Leftrightarrow \hat{\mathbf{x}} \in \underset{\mathbf{x}\in\mathbb{R}^N}{\operatorname{argmax}}\ p_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}(\mathbf{y})p_{\mathbf{X}}(\mathbf{x}). \tag{1.5}$$

Due to the monotonicity of the logarithm function, the previous equation can be expressed as follows:

$$\hat{\mathbf{x}} \in \underset{\mathbf{x}\in\mathbb{R}^N}{\operatorname{argmin}}\ -\log(p_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}(\mathbf{y})) - \log(p_{\mathbf{X}}(\mathbf{x})). \tag{1.6}$$

From the model in (1.1), we can write $\mathbf{w} = \mathbf{y} - \mathbf{H}\mathbf{x}$. Hence, for all $\mathbf{x} \in \mathbf{R}^N$, we have :

$$p_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}(\mathbf{y}) = p_{\mathbf{W}|\mathbf{X}=\mathbf{x}}(\mathbf{w}), \tag{1.7}$$

$$= p_{\mathbf{W}|\mathbf{X}=\mathbf{x}}(\mathbf{y} - \mathbf{H}\mathbf{x}), \tag{1.8}$$

$$= p_{\mathbf{W}}(\mathbf{y} - \mathbf{H}\mathbf{x}). \tag{1.9}$$

Where the last equality follows from the independency between the random noise $\mathbf{W}$ and $\mathbf{X}$. By injecting the expression of $p_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}(\mathbf{y})$ in (1.6), we get:

$$\hat{\mathbf{x}} \in \underset{\mathbf{x}\in\mathbb{R}^N}{\operatorname{argmin}}\ -\log(p_{\mathbf{W}}((\mathbf{y} - \mathbf{H}\mathbf{x})) - \log(p_{\mathbf{X}}(\mathbf{x})), \tag{1.10}$$

$$\Leftrightarrow \hat{\mathbf{x}} \in \underset{\mathbf{x}\in\mathbb{R}^N}{\operatorname{argmin}}\ h(\mathbf{x}) + g(\mathbf{x}). \tag{1.11}$$

Where $g(\mathbf{x}) = -\log(p_{\mathbf{X}}(\mathbf{x}))$ is the regularization term and $h(\mathbf{x}) = -\log(p_{\mathbf{w}}((\mathbf{y} - \mathbf{H}\mathbf{x}))$ is the data fidelity term. For example, when $\mathbf{W}$ is a white Gaussian noise, *i.e.* for all $\mathbf{w} \in \mathbb{R}^M$, the probability density function of $\mathbf{W}$ has the following expression:

$$p_{\mathbf{W}}(\mathbf{w}) = \left(2\sigma^2\pi\right)^{-M/2} \exp\left(\frac{1}{2\sigma^2}\|\mathbf{w}\|_2^2\right), \tag{1.12}$$

where $\|w\|_2 = \left(\sum_i^M w_i^2\right)^{\frac{1}{2}}$ is the ℓ_2-norm. In this case, the data fidelity term could be written as follows:

$$h(\mathbf{x}) = \frac{1}{2\sigma^2}\|\mathbf{y} - \mathbf{H}\mathbf{x}\|_2^2 + \frac{M}{2}\log(2\sigma^2\pi). \tag{1.13}$$

Without loss of generality, we can choose :

$$h(\mathbf{x}) = \frac{1}{2\sigma^2}\|\mathbf{y} - \mathbf{H}\mathbf{x}\|_2^2. \tag{1.14}$$

The data regularization term depends on the probability distribution of $\mathbf{X}$. The design of $g(\mathbf{x})$ is of paramount importance for a better resolution of the variational problem in (1.2). In the following, we detail the most common penalization terms encountered while dealing with WSN data.

1.5.2 Regularization functions

The regularization term allows to express some *a priori* knowledge about the target data that may cover many aspects. Hence, the function g can be written as the sum of several regularization functions:

$$g(\mathbf{x}) = \sum_{i=1}^{n} \lambda_i g_i(\mathbf{x}), \tag{1.15}$$

where, for each $i \in \{1 \dots n\}$, λ_i is a positive tuning parameter that controls the weight of each considered regularization function. One may expect better estimation results when increasing the number of considered regularization functions, since each of them constrains the estimated solution to obey some property. However, the quality of the estimator depends on the choice of the regularization parameters. The later should be adjusted to achieve the best trade-off between the data fidelity term and each considered regularization factor. In the sequel, we describe some common regularizations used in WSN optimization problems.

1.5.2.1 sparsity

One common property used in WSN optimization problems is the sparsity feature which asserts that the signal of interest only contains few nonzero components. Some signals have a sparse representation in the temporal domain such as the seismic signals (Figure 1.2). More often, The signals encountered in WSNs applications are not sparse

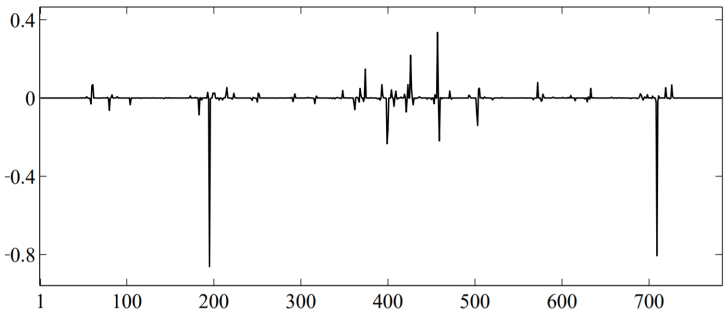


Figure 1.2: *Sparse seismic signal [Repetti et al., 2015]*

in the observation domain, but they have a sparse representation in other transform domain as we will show in the next chapters. These signals need to be transformed in the sparsifying domain using some linear operators. In this case, the regularization function can be expressed as follows:

$$g(\mathbf{x}) = \sum_{j=1}^{J} g_j(\mathbf{F}_j\mathbf{x}), \tag{1.16}$$

where for each $j \in \{1, \ldots, J\}$, $\mathbf{F}_j \in \mathbf{R}^{M_j \times N}$ is a linear operator that sparsifies the signal $\mathbf{x}$, *i.e.* for each $j \in \{1, \ldots, J\}$, the vector $\mathbf{F}_j\mathbf{x}$ contains a dominant number of zeros. The function $g_j : \mathbb{R}^{M_j \times N} \to]-\infty, +\infty]$ is the regularization function that penalizes the sparsity of $\mathbf{F}_j\mathbf{x}$.

The regularization function g_j should be chosen to reinforce the sparsity structure of the data. The most intuitive choice is to minimize the number of nonzero entries of $\mathbf{x}$, which we refer to as the ℓ_0-norm:

$$\|x\|_0 = \text{card}\{i : x_i \neq 0\}. \tag{1.17}$$

Although we use the norm terminology, $\|.\|_0$ is not actually a norm, since it is not homogeneous.

However, solving an optimization problem using the ℓ_0-norm as a regularization function is too complex since it renders the problem combinatorial and NP-hard to solve to a global optimum. To promote the sparsity regularization feature, we can use instead the ℓ_α regularization function defined as follows:

$$\ell\alpha(\mathbf{x}) = \left(\sum_{j=1}^{N} |x_i|^\alpha \right)^{\frac{1}{\alpha}}, \tag{1.18}$$

where $0 \leqslant \alpha < 1$. However, these functions are neither convex nor differentiable resulting in hard solvable optimization problem. A common approach to avoid this

shortcoming is giving by using the ℓ_1-norm as a convex surrogate of the sparsity pattern [Donoho, 2006b; Figueiredo et al., 2007]. The ℓ_1-norm is defined as follows:

$$\|x\|_1 = \sum_{j=1}^{N} |x_i|. \tag{1.19}$$

Despite the ℓ_1-norm is a non differentiable function, it leads to usually tractable convex optimization problem. We use this regularization function to penalize the sparsity of anomalies in our three first contributions.

1.5.2.2 Low rank

The rank is defined as the minimal number of elementary matrices of the form $\mathbf{a}\mathbf{b}^T$, where $\mathbf{a} \in \mathbb{R}^N$ and $\mathbf{b} \in \mathbb{R}^M$, that can be used to additively decompose a linear operator from $\mathbb{R}^N$ to $\mathbb{R}^M$ expressed through the matrix $\mathbf{X} \in \mathbb{R}^{N \times M}$. Formally, we can define the rank of $\mathbf{X}$ as follows:

$$\text{rank}(\mathbf{X}) = \{\|\sigma\|_0 \ : \ \mathbf{X} = \sum_i \sigma_i \mathbf{a}_i \mathbf{b}_i^T, \ \|\mathbf{a}_i\| = \|\mathbf{b}_i\| = 1\}, \tag{1.20}$$

where, $\mathbf{a}_i \in \mathbb{R}^N$, $\mathbf{b}_i \in \mathbb{R}^M$.
The infimum in equation (1.20) is achieved by using the Singular Value Decomposition (SVD). In this case, the vectors $\mathbf{a}_i$ (respectively $\mathbf{b}_i$) are mutually orthogonal.

Due to the correlation between sensory readings, the probed data often exhibits a low rank structure. To integrate this pattern into the optimization problem, one can consider the rank function as a regularization. However, the rank is a non convex matrix functional which complicates the solvability of the resulting optimization problem. As an alternative, we can use the nuclear norm, which is the sum of the singular values, since it is the tightest convex relaxation to the rank function [Fazel, 2002].

1.5.3 Optimization algorithms

Regarding the data features encountered in WSN inverse problems, we are interested in this part in describing some algorithms used to solve the following variational problem:

$$\underset{\mathbf{x} \in \mathbb{R}^N}{\text{Argmin}} \ h(\mathbf{x}) + f(\mathbf{x}) + g(\mathbf{L}\mathbf{x}), \tag{1.21}$$

where h is a convex differentiable data fidelity term. The regularization term is composed of the sum of a non differentiable function f and a penalization of a linear operator $\mathbf{L}$. This sum may express for example a regularization function f that penalizes the data rank structure and another function g that expresses the sparsity of readings under the linear operator $\mathbf{L}$. The association of convex differentiable and convex non differentiable functions renders the optimization problem in (1.21) relevant to

our context of study since we use a similar formulation of the problem of data gathering and anomaly detection in the next chapters. This form of regularization functions makes the problem in (1.21) non trivial. A class of algorithms has bean proposed to deal with this form of regularization. The main tool of these methods is the proximity operator defined as follows:

Definition 1.5.1 Let f be a convex, proper, and lower semicontinuous function from $\mathbb{R}^N$ to $\mathbb{R}$ and let $\lambda > 0$. The proximity operator of λf is defined for every $\mathbf{X} \in \mathbb{R}^N$ as the unique solution of the following minimization problem:

$$\min_{\mathbf{U}\in\mathbb{R}^N} \frac{1}{2}\|\mathbf{U}-\mathbf{X}\|_F^2 + \lambda f(\mathbf{U}), \tag{1.22}$$

and it will be denoted by $\text{prox}_{\lambda f}(\mathbf{X})$.

Proximity operators have many attractive properties that make them interesting tools to develop iterative minimization algorithms. For instance, the proximal operator of the indicator function in some closed convex set reduces to the projection onto this set. Thus, proximal based algorithms seem to be particularly well suitable to address our minimization problem, where we have a non smooth penalization.

In the following, we present some common algorithms used to solve the minimization problem in (1.21).

1.5.3.1 ADMM algorithm

When $L^\star L$ is an isomorphism, where $(.)^\star$ denotes the conjugate function, we can solve the minimization problem depicted in (1.21) using the Alternating Direction Method of Multipliers (ADMM) algorithm [Boyd et al., 2011; Fortin and Glowinski, 2000; Gabay and Mercier, 1976].

Algorithm 1 ADMM algorithm

Initialization

$\gamma > 0$, $\mathbf{y}_0, \mathbf{z}_0 \in \mathbb{R}^M$

Iterations

For $k = 0, ..$

$\mathbf{x}_k = \text{Argmin } \frac{1}{2}\|\mathbf{L}\mathbf{x} - \mathbf{y}_k + \mathbf{z}_k\|^2 + \frac{1}{\gamma}(f(\mathbf{x}) + h(\mathbf{x}))$,

$\mathbf{s}_k = \mathbf{L}\mathbf{x}_k$

$\mathbf{y}_{k+1} = \text{prox}_{g/\gamma}(\mathbf{z}_k + \mathbf{s}_k)$,

$\mathbf{z}_{k+1} = \mathbf{z}_k + \mathbf{s}_k - \mathbf{y}_{k+1}$.

The ADMM algorithm is widely used in signal processing [Afonso et al., 2011; Goldstein and Osher, 2009; Tran-Dinh et al., 2013]. However, there are some practical issues that arise when implementing this algorithm. In fact, the computation of $\mathbf{x}_k$ is non trivial for high vector dimensions or irregular structure of $\mathbf{L}$. Moreover, the ADMM algorithm do not allow to directly include the gradient expression of h.

1.5.3.2 Condat-Vũ primal-dual algorithm

To overcome the shortcoming identified previously for ADMM algorithm, one alternative is to use the primal-dual algorithm proposed by [Condat, 2013; Vũ, 2013]. This algorithm operates as follows:

Algorithm 2 Primal-dual algorithm

Initialization

$\gamma > 0$,$\tau > 0$, $\lambda_k > 0$, $\mathbf{x}_0 \in \mathbb{R}^N$, $\mathbf{v}_0 \in \mathbb{R}^M$

Iterations

For $k = 0, ..$

$$\begin{cases} \mathbf{y}_k = \text{prox}_{\tau f}(\mathbf{x}_k - \tau(\nabla h(\mathbf{x}_k) + \mathbf{b}_k + \mathbf{L}^\star \mathbf{v}_k)) + \mathbf{a}_k, \\ \mathbf{u}_k = \text{prox}_{\gamma g^\star}(\mathbf{v}_k + \sigma \mathbf{L}(2\mathbf{y}_k - \mathbf{x}_k)) + \mathbf{c}_k, \\ \mathbf{x}_{k+1} = \lambda_k \mathbf{y}_k + (1 - \lambda_k)\mathbf{x}_k, \\ \mathbf{v}_{k+1} = \lambda_k \mathbf{u}_k + (1 - \lambda_k)\mathbf{v}_k. \end{cases}$$

Note that, there are no prior conditions about the linear operator $\mathbf{L}$ to ensure the convergence of Algorithm 2 [Condat, 2013]. For each iteration, the terms $\mathbf{a}_k$, $\mathbf{b}_k$ and $\mathbf{c}_k$ represent the eventual errors that may occur during the computation of the proximity operator of τf, the gradient of h and the proximity operator of $\gamma g^\star$.

1.5.3.3 Monotone + Lipschitz Forward Backward Forward algorithm

The author of [Combettes and Pesquet, 2012] presented the Monotone + Lipschitz Forward Backward Forward (M+L FBF) algorithm to solve the minimization problem in 1.21. This algorithm splits the minimization problem and individually proceeds each term involved in 1.21 by alternating the computation of the involved proximity and gradient operators. One main advantage of M+L FBF algorithm is its ability for parallel computing since most of its steps can be executed simultaneously.

Algorithm 3 M+L FBF algorithm

Initialization

$\gamma_k > 0$, $\mathbf{x}_0 \in \mathbb{R}^N$, $\mathbf{v}_0 \in \mathbb{R}^M$

Iterations

For $k = 0, ..$

$$\begin{cases} \mathbf{y}_{1,k} = \mathbf{x}_k - \gamma_k(\nabla h(\mathbf{x}_k) + \mathbf{L}^\star \mathbf{v}_k) \\ \mathbf{p}_{1,k} = \text{prox}_{\gamma_k f}(\mathbf{y}_{1,k}), \\ \mathbf{y}_{2,k} = \mathbf{v}_k + \gamma_k \mathbf{L}\mathbf{x}_k \\ \mathbf{p}_{2,k} = \text{prox}_{\gamma_k g^\star}(\mathbf{y}_{2,k}), \\ \mathbf{q}_{2,k} = \mathbf{p}_{2,k} + \gamma_k \mathbf{L}\mathbf{p}_{1,k}, \\ \mathbf{v}_{k+1} = \mathbf{v}_k + \mathbf{y}_{2,k} + \mathbf{q}_{2,k}, \\ \mathbf{q}_{1,k} = \mathbf{p}_{1,k} - \gamma_k(\nabla h(\mathbf{p}_{1,k}) + \mathbf{L}^\star \mathbf{p}_{2,k}) \\ \mathbf{x}_{k+1} = \mathbf{x}_k - \mathbf{y}_{1,k} + \mathbf{q}_{1,k}. \end{cases}$$

Under a suitable choice of the parameter γ_k, the sequence $\mathbf{x}_k$ converges to the minimum of the optimization problem [Combettes and Pesquet, 2012]. Note that we use an extended form of this algorithm in the next chapters.

1.6 CONCLUSION

In this chapter, we presented an overview of WSN as well as the most common challenges encountered in WSN applications. Then, we described the typical techniques used to gather the sensory data and to detect the presence of anomalies. Thereafter, we initiated the reader to the application of inverse problems in the context of our study. Looking to the related works, we noticed that most of existing solutions separate the problem of data gathering and anomaly detection. In this thesis, we propose to jointly address the problem of data gathering and anomaly detection and correction, *i.e.* how to optimize collecting data in terms of resource consumption while detecting and correcting random anomalies?

To answer these questions, we propose in our first contribution a new CS-based solution that allows to collect data and detect anomalies using an extended version of the M+L FBF algorithm. Then, we propose two incomplete-data based solution for data gathering and anomaly detection using convex optimization techniques. Finally we propose a new approach for collecting data and detecting anomalies that combines some aspects of CS and data compression-based techniques.

- Chapter 2 -

Spatio-temporal compressive sensing technique for data gathering and anomaly detection

2.1 Introduction

In this chapter, we focus on collecting data and detecting anomalies in Wireless Sensor Networks (WSNs) while optimizing the use of sensor computational and energetic resources. Recently, Compressive Sensing (CS)-based solutions had been the subject of extensive studies for the design of efficient data gathering solutions in WSNs. However, existing CS-based approaches are very sensitive to outlying values and do not offer a proper tool to deal with the presence of anomalies. Moreover, CS data gathering

schemes are based only on the spatial correlation pattern between sensory data and ignore an important feature which is the temporal correlation between sensor readings.

This chapter introduces a novel CS-based data gathering solution that allows to integrate the spatial and temporal correlation features into the data gathering scheme allowing to render it even more efficient. Furthermore, the proposed approach is built in such a way that it also allows to detect and correct eventual anomalies. We propose a general formulation of data gathering and anomaly detection problem as a tractable convex optimization problem on the Hilbert space of data measurements and anomalies. Besides, we design a new class of primal-dual algorithms to solve the resulting optimization problem. We evaluate the efficiency of our method by running extensive simulations on two real datasets. We demonstrate that the proposed algorithm achieves good data recovery and anomaly detection performance and outperforms the main state-of-the-art technique addressing the same problem.

This chapter is organized as follows: We introduce the motivations and the context of this study in next section. We formally introduce the studied problem in Section 2.3. In Section 2.4, we detail the proposed variational approach. We describe the proposed algorithm to solve the resulting problem in section 2.5. Section 2.6 is devoted to the performance evaluation. Finally, concluding remarks are given in Section 2.7.

2.2 Context and motivation

The utilization of Wireless Sensor Networks (WSNs) is widely spread throughout many fields and applications that require reliable and efficient data processing techniques to perform some predefined tasks such as gathering the observed data while being able to detect the occurrence of some events or anomalies. To achieve this aforementioned goal, nodes need to transmit their probed measurements to the sink according to a predefined cadence using a multi-hop path. However, this simple approach leads to an unfair energy consumption distribution between nodes, since sensors close to the sink have to carry the measurements of their descendant nodes in addition to their own readings. Thus, intuitively forwarding the sampled signal may be too costly and may reduce the lifespan of WSNs.

To enhance the lifespan of WSNs and reduce the energy consumption of wireless sensors, many data gathering methods are proposed. One efficient solution is given by CS-based techniques which rely on finding a proper base or domain such as Discrete Cosine Transform (DCT) or Discrete Wavelets Transform (DWT) where the signal could be well approximated by a sparse representation [Candès and Wakin, 2008]. Hence, sensor nodes would only transmit the largest coefficients of their sparse representation instead of the whole vector of readings. The temporal correlation between readings probed at different time slots implies the sparse representation of each sensor readings, whilst the spatial correlation induces the sparsity of all sensors' readings at a predefined time slot. However, in the presence of anomalies, sensor measurements are no longer compressible, *i.e.*, sparse in some transform domain [Chong et al., 2009]. More-

over, most CS-based techniques [Wang et al., 2013, 2012; Zheng et al., 2013; Chong et al., 2009] are restrained to one of the correlation patterns, either space or time, depending on which dimension the CS is performed, and they do not offer a suitable tools to integrate both correlation dimensions which may reduce the performance of these techniques, especially, when the compression rate is high or the number of sensor is relatively low.

Another appealing solution is proposed by Matrix Completion (MC)-based techniques. These methods are based on the low rank feature of the data. In fact, in addition to the sparsity, sensor readings exhibit low rank structure due to their spatio-temporal correlation. Using MC-based techniques, sensors only forward an arbitrarily set of their measurements according to a predefined sampling probability. Hence, based on the few received readings, the sink tries to estimate the missing values with an acceptable distortion. However, This class of techniques do not offer an appropriate tool to deal with aberrant values.

To overcome the constraints imposed by the presence of outliers, many data anomaly detection techniques are proposed. Depending on the *a priori* knowledge about anomaly behaviors, we can assort three different detection approaches [Chandola et al., 2009]. All considered solutions assume that the data is stored centrally and there are enough measurements to train the central node about different data behavior regimes. The first approach relies on unsupervised learning where the sink is supposed to be able to disassemble outliers from the normal data structure. The second approach leans on supervised clustering where the sink learns about the different data behavior models from labeled data training sets. The last approach is based on semi-supervised clustering which combines the two previous learning techniques.

Most of anomaly detection solutions in WSNs are inspired from data mining and machine learning fields and they are not well suitable for resource limited WSNs, since they assume that all nodes forward their readings to the sink, which may cause a high energy consumption and limited lifespan of the whole network. Moreover, most of existing approaches separately treat the problem of data gathering and anomaly detection which may results in non-optimal solutions in terms of data accuracy as well as error detection and correction resolution.

In an aim of overcoming these shortcomings, we propose in this chapter to jointly process the problem of data gathering and anomaly detection in WSNs by proposing a CS-based technique that allows to perform efficient data gathering and anomaly detection in real-time. The originality of the idea comes from including the past observations in the process of data recovering and anomaly detection on the current measurements, and thus enhancing the WSN performance. Hence, even if the CS is performed over one dimension, the other dimension is included in the proposed data recovery and anomaly detection solution by exploiting both spatial and temporal correlation features.

In this chapter, we formulate the problem of data gathering in the presence of outliers as a tractable convex optimization problem on the Hilbert space of data mea-

surements and anomalies. We introduce a new primal-dual algorithm to solve the resulting optimization problem. The proposed algorithm takes into account different data features, namely, the sparsity of the projections of readings in some basis and the low rank structure of the data matrix formed by the previous reading estimations as well as the current observation, which allows to capture the spatio-temporal correlation among readings. The presence of outliers is represented by a sparsity constraint in the optimization problem. To asses the performance of our proposed algorithm and prove its efficiency, we conduct simulations on real data where we consider different anomaly densities and different compression rates.

2.3 Preliminaries and problem formulation

2.3.1 Foundations of Compressive Sensing

CS theory suggests that we can recover a sparse signal from far fewer samples than the number required by the Shannon-Nyquist sampling theorem [Candes and Tao, 2006] [Donoho, 2006a]. This reconstruction is made possible by solving a programming optimization problem from non-adaptive linear projections.

An $N-$dimensional signal $\mathbf{d}$ is said to have a sparse representation if there exist some basis where the projection of $\mathbf{d}$ induces a sparse sequence. The representation of $\mathbf{d}$ in the considered domain is given by the following expression:

$$\mathbf{d} = \mathbf{\Psi}^{-1}\mathbf{x} = \sum_{i=1}^{N} \varphi_i x_i, \tag{2.1}$$

where $\mathbf{\Psi}^{-1} = \{\varphi_1, \varphi_2, \ldots, \varphi_n\}$ represents the basis that sparsifies the signal $\mathbf{d}$ and φ_i is the i^{th} component of the basis $\mathbf{\Psi}^{-1}$. In the CS literature, usually the sparsifing domain is set to the DCT or WT operators. The projection coefficients of the signal $\mathbf{d}$ are given by the sparse signal $\mathbf{x}$, *i.e.*, most of the resulting coefficients are zero. Moreover, the signal $\mathbf{d}$ is said to be $K-$sparse if the number of non-zero entries of $\mathbf{x}$ is equal to K.

Compressive sensing theory suggests that we can reconstruct a $K-$sparse signal from a small set of measurements obtained using a specific acquisition system that can be expressed as follows:

$$\mathbf{y} = \mathbf{\Phi}\mathbf{d}, \tag{2.2}$$

where $\mathbf{y}$ is the resulting sequence of the projection of $\mathbf{d}$ using the $M \times N (M \ll N)$ sampling matrix $\mathbf{\Phi}$. This accusation system leads to an undetermined system of linear equations where we have only M linear equations to determine N unknown variables. There are two conditions required to make this recovery possible. The first one states that the sparsity of the signal must satisfy the following inequality:

$$M \geq cK \log \frac{N}{K}, \tag{2.3}$$

where c is a positive constant. In practice $M = 3K \sim 4K$ is sufficient to satisfy this condition. The second condition requires matrices $\mathbf{\Phi}$ and $\mathbf{\Psi}$ to be incoherent. This condition is always ensured by using random measurement matrices since a random matrices are proved to be incoherent with any fixed basis [Candès and Tao, 2005]. Under this two conditions, it has been shown that the signal $\mathbf{d}$ can be exactly recovered by solving the following minimum ℓ_1-norm optimization problem [Candes et al., 2006]:

$$\min_{\mathbf{x}\in\mathbb{R}^N} \|\mathbf{x}\|_1 \quad s.t\ \mathbf{y} = \mathbf{\Phi}\mathbf{d},\ \mathbf{d} = \mathbf{\Psi}^{-1}\mathbf{x}. \tag{2.4}$$

Hence the estimation of the original signal is $\tilde{\mathbf{d}} = \mathbf{\Psi}^{-1}\tilde{\mathbf{x}}$, where $\tilde{\mathbf{x}}$ is the solution of the convex optimization problem (2.4). Many optimization algorithms can be used to solve the above convex problem such as the least absolute shrinkage and selection operator algorithm or the matching pursuit techniques [Potra and Wright, 1997; Donoho et al., 2012; Candès and Romberg, 2005].

2.3.2 Problem Formulation

In order to avoid the bottleneck effect caused by baseline data gathering methods, we can balance the traffic carried by sensors and reduce the communication cost by applying CS theory in collecting data in WSN. Since, measurements probed by sensors at any time slot are correlated, and hence, exhibit the sparsity pattern in some transform domain, one can transmits a linear combination of all sensors' readings instead of the whole sequence of measurements. This projection is performed in a distributed way in sensor nodes and the complexity caused by solving the optimization problem to recover the coded data is shifted to the sink, which is supposed to dispose enough computation, storage and energy resources to perform this kind of operations. Note that the random projection of reading do not require an *a priori* knowledge about the basis that sparsifies the readings since this basis is only required in solving the optimization problem at the sink. We refer to this type of compression by Spatial Compressive Sensing (SCS) data gathering scheme, since the sparse pattern involved in the data recovery process is caused by the spatial correlation pattern between sensors.

In a typical SCS data gathering scheme, the data measurements probed at a time slot t from N sensor nodes can be represented by the vector $\mathbf{d}_t = [d_{1,t}, d_{2,t}, \ldots, d_{N,t}]$, where $d_{i,t}$ is the value sensed by node i at the time slot t. To perform the distributed random projection, we store at each node i the i^{th} component ϕ_i of the random matrix $\mathbf{\Phi}$. After the measurement acquisition step, leaf nodes initiate the transmission process. Each node multiplies the probed reading by the stored vector of the random matrix. Then, the resulting vector is concatenated to the received vectors generated by the children nodes and transmitted to the higher node. Note that the transmission of readings to the sink is performed using an Ad-hoc routing protocol, based on which the set of routing paths presents a tree structure. The M-dimensional auxiliary vector

$\mathbf{y}_{i,t}$ transmitted by node i at the time slot t can be expressed as follows:

$$\mathbf{y}_{i,t} = \phi_i d_{i,t} + \sum_{j \in S_i} \phi_j d_{j,t}, \tag{2.5}$$

where S_i denotes the set of indices of children nodes attached to a sensor i. Thereby, the SCS data gathering scheme allows sensors to carry the same number of messages (M samples) regardless their distance from the sink. After transmitting all compressed readings, the collector node receives the following M-dimensional vector:

$$\mathbf{y}_t = \sum_{i=1}^{N} \phi_i d_{i,t}. \tag{2.6}$$

The data gathering process is depicted in Figure 2.1. In traditional SCS data gathering techniques, the sink applies an ℓ_1-optimization algorithm to recover the coded signal $\mathbf{d}_t$ from the received vector $\mathbf{y}_t$.

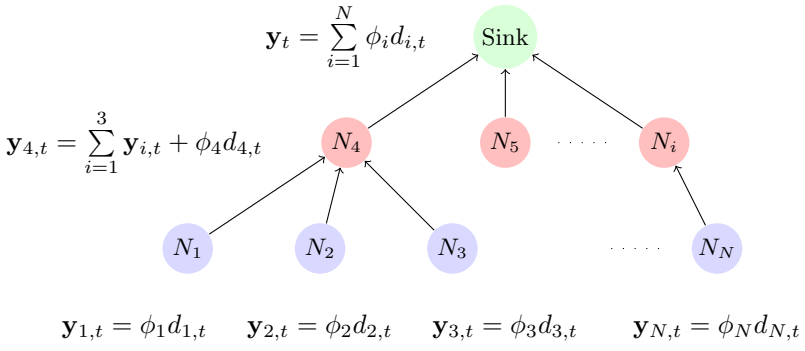


Figure 2.1: *Data gathering process using SCS technique.*

In the following, we propose a new frame of data recovering and anomaly detection algorithms, where the past observations are included into the proposed algorithm to enhance WSN performance.

2.4 Data gathering and anomaly detection thought spatio-temporal compressive sensing

In this section, we propose our CS-based solution that allows to efficiently collect data while detecting anomalies. Our scheme differs from traditional CS solutions by integrating the side information available at the sink in the data recovering and anomaly detection process. The considered side information consists on the l previous data estimations that precede the actual time slot t. These estimations can be arranged in a time window $\mathbf{W}_l$ such that $\mathbf{W}_l = [\tilde{\mathbf{d}}_{t-l}, \tilde{\mathbf{d}}_{t-l+1}, \ldots, \tilde{\mathbf{d}}_{t-1}]$, where $\tilde{\mathbf{d}}_i$ denotes the

estimation of sensors' readings at the time slot i. In the sequel, we detail how to encode the presence of this side information in the data gathering and anomaly detection algorithm.

2.4.1 Spatio-Temporal Compressive Sensing Solution

AS mentioned in chapter 1, in order to recover the coded sensors' readings, standard SCS techniques proceed by solving the optimization problem in (2.12) which could be written as follows:

$$\underset{\mathbf{d}_t}{\text{minimize}}\ h(\mathbf{d}_t) + \xi(\mathbf{d}_t), \tag{2.7}$$

where $h(\mathbf{d}_t) = \frac{1}{2}\|\mathbf{y}_t - \mathbf{\Phi}\mathbf{d}_t\|^2$ is the data fidelity term, whereas $\xi(\mathbf{d}_t) = \lambda\mathbf{\Psi}^{-1}\mathbf{d}_t$ is the term that penalizes the sparsity of readings in the considered domain and λ is a positive regularization parameter.
We propose to include the previous estimations to ameliorate the recovering quality by suitably choosing the regularization term. Indeed, this term models an a prior knowledge about the target data, essentially the sparsity of the data in some domain. We redefine this regularization term so it could model the spatial sparse representation of $\mathbf{d}_t$ and the correlation between the columns of the updated time window $\mathbf{W}'_l = [\mathbf{W}_l, \mathbf{d}_t]$. The sparse representation of the columns of the matrix $\mathbf{W}'_l$ is given by $\mathbf{\Psi}^{-1}\mathbf{W}'_l$. The correlation between points of the same rows is taken into account by jointly processing the components belonging to the same rows which enables to include the temporal dimension in the sparsity penalization term in addition to the spatial dimension. Then, the regularization term in equation (2.7) could be written as follows:

$$\xi(\mathbf{d}_t) = \lambda_1 f(\mathbf{\Psi}^{-1}\mathbf{W}'_l), \tag{2.8}$$

where λ_1 is a positive regularization parameter and f is defined for all $\mathbf{U} \in \mathbb{R}^{N\times(l+1)}$ as follows:

$$f(\mathbf{U}) = \sum_{j=1}^{l+1} g(\|\mathbf{U}_{:,j}\|), \tag{2.9}$$

where $\|.\|$ denotes the ℓ_2 norm, $g : \mathbb{R} \longrightarrow \mathbb{R}^+$ and for all $j \in \{1, \ldots, l+1\}$, $\mathbf{U}_{:,j}$ denotes the j^{th} column of the matrix $\mathbf{U}$. In this work, we use the generalized Gaussian regularization defined for all $t \geqslant 0$ as $g(t) = t^\beta$ where $\beta > 0$ is the shape parameter. It reduces to the neg-log-likelihood of the normal distribution when $\beta = 2$. It promotes tails that are heavier than normal when $\beta < 2$ reflecting the sparsity of the data [Marnissi et al., 2013; Kwitt et al., 2009]. In particular, we obtain the ℓ_2 norm when $\beta = 1$.

The recovery accuracy could be enhanced by investigating the structure of the updated time windows, $\mathbf{W}'_l$. In fact, because of the spatio-temporal correlation between sensors' readings, the probed measurements are supported by few dimensions of the data space. Thus the data matrix $\mathbf{W}'_l$ presents a low rank structure which could be

analyzed using the Singular Value Decomposition (SVD). The data matrix $\mathbf{W}'_l$ can thus be decomposed as follows:

$$\mathbf{W}'_l = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^\top, \tag{2.10}$$

where $\mathbf{U}$ and $\mathbf{V}$ are respectively an $N \times N$ and $(l+1) \times (l+1)$ orthonormal matrices, and $\mathbf{\Sigma}$ is an $N \times (l+1)$ diagonal matrix formed by the singular values $\sigma_1, \sigma_2, \ldots, \sigma_r$ arranged in decreasing order. For low ranked matrices, the data points lie in a space with dimension equals to the number of the highest singular values which is always smaller than the ambient dimension of $\mathbf{W}'_l$. We model the low rank nature of the data using the nuclear norm since it is the closest convex approximation of the low rank structure. Hence, the regularization term can be updated as follows:

$$\xi(\mathbf{d}_t) = \lambda_1 f(\mathbf{\Psi}^{-1}\mathbf{W}'_l) + \lambda_2 \|\mathbf{W}'_l\|_*, \tag{2.11}$$

where λ_2 is a positive tuning parameter and $\|\mathbf{W}'_l\|_*$ denotes the nuclear norm of $\mathbf{W}'_l$ and is defined as the sum of the singular values.

2.4.2 Recovering data in the presence of anomalies

Anomaly readings often occur in WSNs due to various phenomena such as the capacity limitation of sensor devices and the perturbations of the observed environment. While SCS techniques offer a suitable solution to collect data in WSNs, they struggle to manage the presence of anomalies since it affect the sparsity pattern. To illustrate the effect of outliers on the sparsity feature, we plot the DCT of a normal signal spotted with 3 anomalies and the effect of outlying values on the DCT representation in Figure 2.2. Even if the number of anomalies is negligible, the resulting consequences on the sparsity feature are dramatic. Figure 2.2 shows that the spotted signal is no longer compressible.

To overcome this limitation, traditional CS techniques relay on overomplete basis representation [Candes et al., 2006; Candès and Wakin, 2008], which consists on recovering a $2N-$sparse vector that contains the N coefficients of the sparse representation and N others samples that represent the eventual anomalies. Despite that these techniques provide a theoretical solution to detect anomalies, they remain impractical since they double the amount of data to be compressed and hence reduce the data recovery performance. In this work, we propose an alternative approach that jointly determines the normal and outlying readings without sacrificing the compression ratio. We propose to modify the optimization problem in (2.7) in order to jointly seek for the normal vector $\mathbf{d}_t$ and the anomaly vector denoted by $\mathbf{e}_t \in \mathbb{R}^N$. We can rewrite equation (2.7) as follows:

$$\underset{\mathbf{d}_t, \mathbf{e}_t}{\text{minimize}}\ h(\mathbf{d}_t, \mathbf{e}_t) + \xi(\mathbf{d}_t) + \xi_a(\mathbf{e}_t), \tag{2.12}$$

where $h(\mathbf{d}_t, \mathbf{e}_t) = \frac{1}{2}\|\mathbf{y}_t - \mathbf{\Phi}(\mathbf{d}_t + \mathbf{e}_t)\|_F^2$ is the new data fidelity term, whereas $\xi_a(\mathbf{e}_t)$ is the term that penalizes the presence of anomalies. Since outliers are sporadic, we

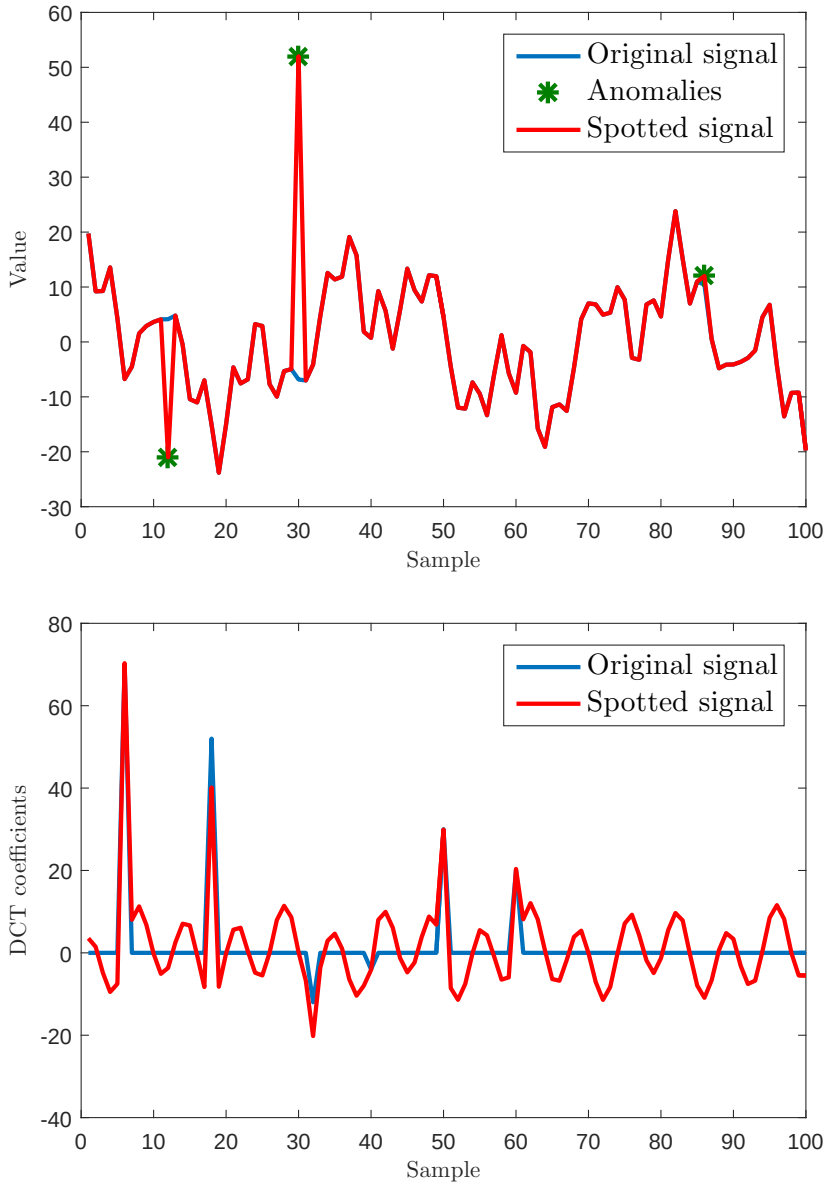


Figure 2.2: *The effect of anomalies on the DCT representation.*

propose to penalize their presence using the ℓ_1-norm. Hence, the anomaly penalization term can be fixed as follows:

$$\xi_a(\mathbf{e}_t) = \lambda_3 \|\mathbf{e}_t\|_1, \tag{2.13}$$

where λ_3 is a positive regularization parameter.
By replacing the regularization terms by their expressions, we get the final expression

of the optimization problem :

$$\underset{\mathbf{d}_t,\mathbf{e}_t}{\text{minimize}}\ h(\mathbf{d}_t,\mathbf{e}_t)+\lambda_1 f(\mathbf{\Psi}^{-1}\mathbf{W}_l')+\lambda_2\|\mathbf{W}_l'\|_*+\lambda_3\|\mathbf{e}_t\|_1, \tag{2.14}$$

In the next section, we detail the proposed algorithm to solve the optimization problem depicted by (2.14).

2.5 Primal-dual splitting algorithm for spatio-temporal Compressive Sensing

Our objective is to provide a solution to the optimization problem in equation (2.14). An effective approach for the minimization of differentiable criteria is gradient-based algorithms. When the diffrentiability property is not satisfied such as in the treated case due to the considered regularization terms, one can still use these algorithms by employing smooth approximations of the non-differentiable functions. Another way to solve these problems is to resort to proximal based methods [Combettes and Pesquet, 2011, 2008].

To provide a numerical solution to the problem depicted by equation (2.14), we propose to apply the primal-dual algorithm [Combettes and Pesquet, 2011; Komodakis and Pesquet, 2015]. This algorithm is specifically designed to minimize a criterion of the form (2.14) which can be written as the sum of a differentiable function h with a Lipschitz gradient, and non-differentiable functions. Note that h is a convex smooth function and it is continuously differentiable with 1-Lipschitz gradient:

$$\nabla h\left(\begin{bmatrix}\mathbf{d}_t\\ \mathbf{e}_t\end{bmatrix}\right)=\begin{bmatrix}\mathbf{\Phi}^\top(\mathbf{\Phi}\mathbf{d}_t-\mathbf{y}_t+\mathbf{\Phi}\mathbf{e}_t)\\ \mathbf{\Phi}^\top(\mathbf{\Phi}\mathbf{e}_t-\mathbf{y}_t+\mathbf{\Phi}\mathbf{d}_t)\end{bmatrix} \tag{2.15}$$

The data fidelity term h is introduced into the algorithm using its gradient expression while the non-smooth components of the regularization function are involved via their proximity operators.
The proximal operator of the nuclear norm is simply given by the shrinkage operator [Cai et al., 2010] expressed as follows:

$$\text{prox}_{\lambda\|.\|_*}(\mathbf{W}_l')=\mathbf{U}\text{diag}\left(\{\max(0,\sigma_i-\lambda)\}_{1\leqslant i\leqslant r}\right)\mathbf{V}^\top, \tag{2.16}$$

where $\mathbf{U}$, $\mathbf{V}$ and $(\sigma_i)_{1\leqslant i\leqslant r}$ are defined as in (2.10).
The proximal operator associated to the anomaly structure is given by the following expression:

$$\text{prox}_{\lambda\|.\|_1}(\mathbf{e}_t)_i=\max\left(0,1-\frac{\lambda}{|e_{i,t}|}\right)e_{i,t}, \tag{2.17}$$

where $e_{i,t}$ denotes the i^{th} position of the vector $\mathbf{e}_t$.
The proximal operator of f_1 depends on the choice of β and can be easily computed using [Combettes and Pesquet, 2010].

Algorithm 4 Spatio-Temporal Compressive Sensing (STCS) algorithm for data gathering and anomaly detection and correction

Initialization

$\gamma > 0$, $\mathbf{d}_0, \mathbf{e}_0, \mathbf{v}_{1,0}, \mathbf{v}_{2,0}, \mathbf{v}_{3,0} \in \mathbb{R}^N$

Iterations

For $k = 0, ..$

Gradient computation

$$\begin{bmatrix} \mathbf{w}_{0,k} \\ \mathbf{w}_{1,k} \end{bmatrix} = \nabla h \left(\begin{bmatrix} \mathbf{d}_k \\ \mathbf{e}_k \end{bmatrix} \right)$$

$$\mathbf{y}_{0,k} = \mathbf{w}_{0,t} + \mathbf{\Psi}^\top \mathbf{v}_{2,K} + \mathbf{v}_{1,k}$$
$$\mathbf{y}_{1,k} = \mathbf{w}_{1,t} + \mathbf{v}_{3,k}$$
$$\mathbf{p}_{0,k} = \mathbf{d}_k - \gamma\ \mathbf{y}_{0,k}$$
$$\mathbf{p}_{1,k} = \mathbf{e}_t - \gamma\ \mathbf{y}_{1,k}$$

Proxy computation

$$\mathbf{y}_{2,1,k} = \mathbf{v}_{1,k} + \gamma\ \mathbf{d}_k$$
$$\mathbf{y}_{2,2,k} = \mathbf{v}_{2,k} + \gamma\ \mathbf{\Psi}\mathbf{d}_k$$
$$\mathbf{y}_{2,3,k} = \mathbf{v}_{3,k} + \gamma\ \mathbf{e}_k$$
$$\mathbf{p}_{2,1,k} = \mathbf{y}_{2,1,t} - \gamma\ \mathrm{prox}_{\gamma^{-1}\lambda_2\|.\|_*}(\gamma^{-1}\left[\mathbf{W}_l, \mathbf{y}_{2,1,k}\right])$$
$$\mathbf{p}_{2,2,k} = \mathbf{y}_{2,2,t} - \gamma\ \mathrm{prox}_{\gamma^{-1}\lambda_1 f}(\gamma^{-1}\left[\mathbf{\Psi}\mathbf{W}_l, \mathbf{y}_{2,2,k}\right])$$
$$\mathbf{p}_{2,3,k} = \mathbf{y}_{2,3,k} - \gamma\ \mathrm{prox}_{\gamma^{-1}\lambda_3\|.\|_1}(\gamma^{-1}\mathbf{y}_{2,3,k})$$

Averaging

$$\mathbf{q}_{2,1,k} = \mathbf{p}_{2,1,k} + \gamma\ \mathbf{p}_{0,k}$$
$$\mathbf{q}_{2,2,k} = \mathbf{p}_{2,2,k} + \gamma\mathbf{\Psi}\mathbf{p}_{0,k}$$
$$\mathbf{q}_{2,3,k} = \mathbf{p}_{2,3,k} + \gamma\ \mathbf{p}_{1,k}$$
$$\mathbf{v}_{1,k+1} = \mathbf{v}_{1,k} - \mathbf{y}_{2,1,k} + \mathbf{q}_{2,1,k}$$
$$\mathbf{v}_{2,k+1} = \mathbf{v}_{2,k} - \mathbf{y}_{2,2,t} + \mathbf{q}_{2,2,k}$$
$$\mathbf{v}_{3,k+1} = \mathbf{v}_{3,k} - \mathbf{y}_{2,3,k} + \mathbf{q}_{2,3,k}$$

Update

$$\begin{bmatrix} \mathbf{w}_{2,k} \\ \mathbf{w}_{3,k} \end{bmatrix} = \nabla h \left(\begin{bmatrix} \mathbf{p}_{0,k} \\ \mathbf{p}_{1,k} \end{bmatrix} \right)$$

$$\mathbf{q}_{1,k} = \mathbf{w}_{2,k} + \mathbf{\Psi}^\top \mathbf{p}_{2,k,t} + \mathbf{p}_{2,1,k}$$
$$\mathbf{q}_{2,k} = \mathbf{w}_{3,t} + \mathbf{p}_{2,3,k}$$
$$\mathbf{d}_{k+1} = \mathbf{d}_k - \mathbf{y}_{0,k} + \mathbf{p}_{0,k} - \gamma\ \mathbf{q}_{1,k}$$
$$\mathbf{e}_{k+1} = \mathbf{e}_k - \mathbf{y}_{1,k} + \mathbf{p}_{1,k} - \gamma\ \mathbf{q}_{2,k}.$$

Return

$d_t = d_k$

$e_t = e_k$

The main steps of the proposed optimization method are summarized in Algorithm 1. It alternatively performs the computation of the gradient and the proximal operators detailed before. Under some conditions about the choice of γ, the convergence of the primal-dual algorithm to a global minimizer of the proposed criterion is guaranteed for $\beta \geqslant 1$ from the results in [Combettes and Pesquet, 2011].

One of the advantages of this algorithm is that it allows us to solve (2.7) for arbitrary choices of the basis $\mathbf{\Psi}$.

2.6 Performance evaluation

In this section, we evaluate the performance of our proposed algorithm by running extensive simulations on two real datasets and under different anomaly configurations. The previously described algorithm is compared to the SCS approach introduced in [Chong et al., 2009] which allows to reconstruct measurements and detect anomalies based only on the spatial sparsity pattern. In the sequel, we start by analyzing the required data structure in the treated datasets. Then, we discuss our simulation configurations and results in terms of data recovering and anomaly detection performance.

2.6.1 Datasets structure

In order to evaluate the performance of our Spatio-Temporal Compressive Sensing (STCS) algorithm, we run multiple simulations on two independent datasets provided by Intel Berkeley Research Lab. Both datasets are composed of measurements gathered from 52 sensors and probed from February 28$^{\text{th}}$ to April 28$^{\text{th}}$, 2004. The nodes are deployed in a residential building located at Charlottetown in Canada to measure the indoor temperature and humidity traces. The readings are reported to the sink each 30 seconds. By analyzing these datasets, we found out that there is a huge portion of missing values, so we rearranged them by taking the most likely values probed each 10 minutes.

Figure 2.3 shows the humidity and temperature measurements observed by the first sensor. The sensed values are almost periodic with a period nearly equals to 100 time slot. Hence, we fix the time window length l to 100, so it captures the eventual variations of the data behavior.

Our proposed approach assumes a certain structure to be present in the treated data, namely, the low rank structure and the sparsity under some transform operators. To verify if these conditions are meet, which is required by our STCS algorithm, we analyze the two considered datasets. We use the following metric to examine the low rank structure:

$$g(d) = \frac{\sum_{i=1}^{d} \sigma_i}{\|\mathbf{W}'_L\|_*} = \frac{\sum_{i=1}^{d} \sigma_i}{\sum_{i=1}^{r} \sigma_i}, \tag{2.18}$$

where d represents the number of considered singular values. This metric represents the distribution of energy captured by the first d dimensions, *i.e.*, the d first singular

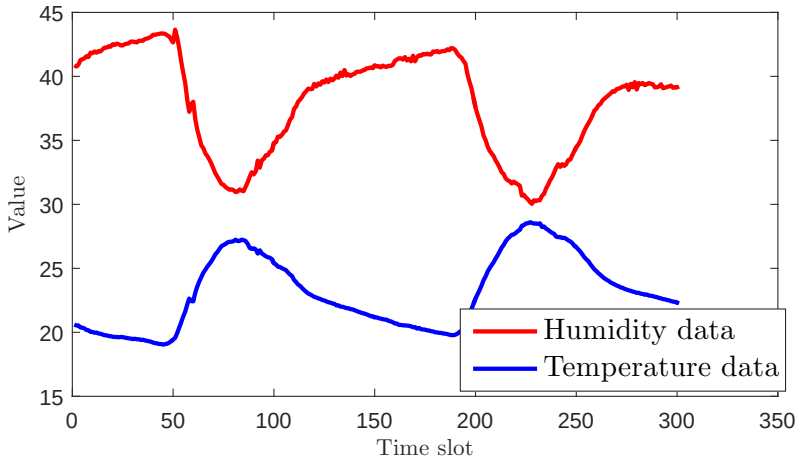


Figure 2.3: *First sensor humidity and temperature readings over 300 time slots.*

values and their associated subspaces.
Concerning the local structure, we propose to investigate the sparsity of the readings under a well-known sensory data sparsifying domain induced by the DCT operator.
Figure 2.4 shows that the first 10% of singular values capture about 92% to 98% of the total energy, which suggests that the considered time window exhibits a good low rank structure. To highlight the spatial correlation among sensors' readings, we plot the DCT coefficients of all sensors' readings probed at the first time slot for both data sets. Figure 2.4 confirms that both considered datasets satisfy the sparsity condition under the DCT domain.

2.6.2 Simulation results

To evaluate the performance of our STCS algorithm, we spot sensors' readings with random anomalies drawn according to the normal distribution since outliers in real environment occur due to a variety of random phenomena. Hence, from the central limit theorem, they behave according to the normal distribution. We fix the variance of the anomaly distribution to 4 and the expectation to 0 for all conducted simulations. Regarding the data feature depicted previously, we set the sparsity operator $\mathbf{\Psi}$ to the DCT operator and we fix $\beta = 1$ to capture the spatio-temporal correlation between readings. The regularization parameters λ_1, λ_2, λ_3 introduced in our STCS algorithm control the weight of each term in the optimization criteria. Thus, the induced performance depends heavily on these regularization parameters. We should find the parameters that give the best trade-off between the data fidelity term, the low-rank feature, the sparsity constraint and the sparsity of anomalies. In our simulations, we set them empirically by choosing the ones giving the best results. Finally, we conduct simulations over 2 successive time windows given that the sink disposes

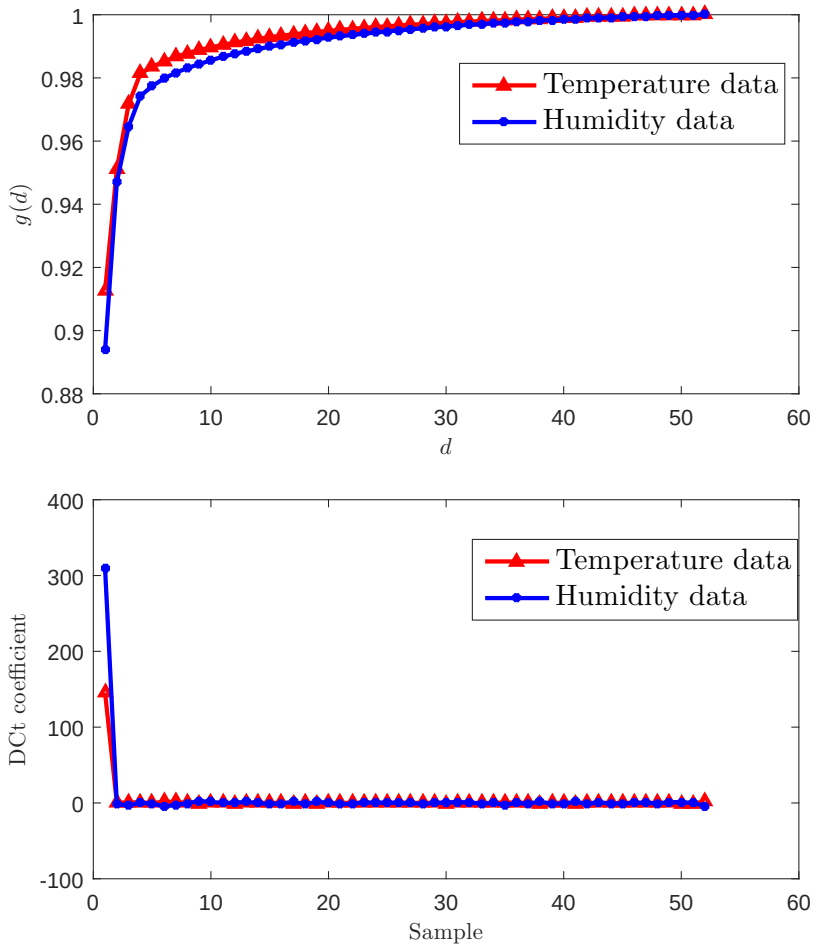


Figure 2.4: *From top to bottom: low rank feature - sparsity of DCT of all sensors' readings at the first time slot.*

at the beginning the necessary readings to initiate the algorithm. Each simulation is conducted for 10 independent random trials and the induced results are averaged over 10 trials.
In the following, we demonstrate the anomaly detection capabilities of our method. Then, we focus on data recovering performance.

2.6.2.1 Anomaly detection performance

To test the capacity of our method in terms of separating outliers from original measurements, we simulate different possible anomaly percentages that vary from 3% to 18% of the probed data. We recall that the considered anomalies are drawn according to a centered normal distribution with variance equals to 4.

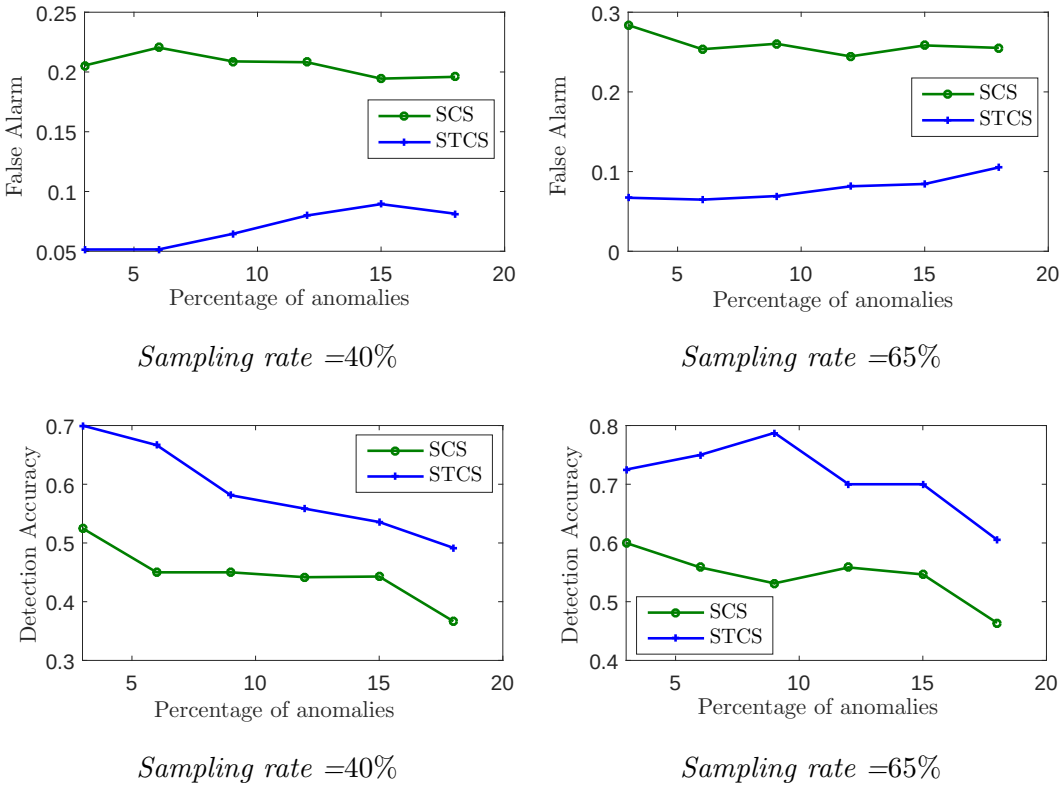


Figure 2.5: *Anomaly detection performance on the Intel the humidity data.*

We propose to evaluate our anomaly detection method in terms of the following metrics:

- Detection accuracy: the ratio of the number of detected anomalies to the total number of anomalies.
- False alarm: the ratio of the number of normal readings that are claimed as anomalies to the total number of normal readings.

We propose to test the performance of our STCS method under two different data compression rates, namely 40% and 65% of the amount of total readings.

As illustrated in Figures 2.5 and 2.6, our proposed approach can effectively identify and estimate the outlying readings with a high detection accuracy and a low false alarm in most treated cases and for both datasets. Our STCS algorithm can reach a high detection accuracy up to 90% while maintaining a negligible false alarm rate less than 8% in average. For example, at the sampling rate of 65%, and in the presence of 9% of anomalies among Intel temperature data, we can reach a detection accuracy higher than 0.9, whilst the normal readings that are claimed as outliers are less than 3% of the

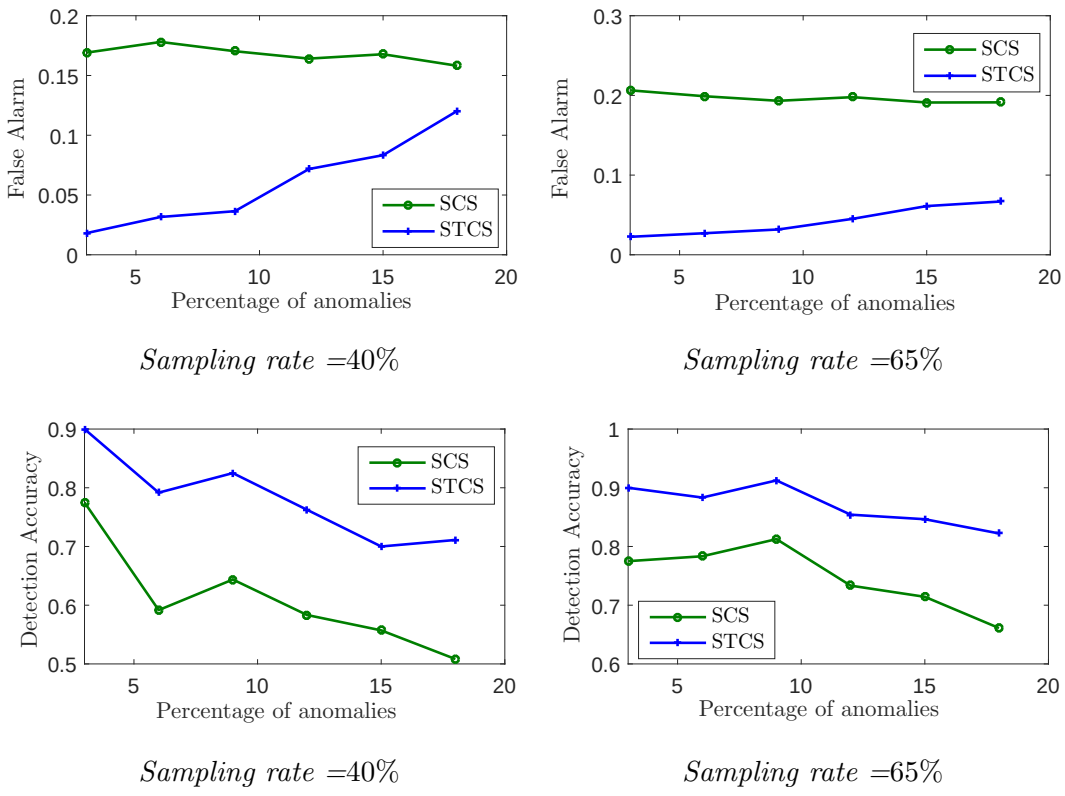


Figure 2.6: *Anomaly detection performance on the Intel temperature data.*

total clean readings which illustrates the robustness of our approach. Alternatively, the SCS method reaches considerable false alarm rate up to 25% and low detection accuracy rate up to 50%. This huge improvement comes from the inclusion of the temporal component in the data process recovery as well as in computing the vector of anomalies. One should also note that this is achieved without doubling the sampling rate such as in the case of the SCS technique.

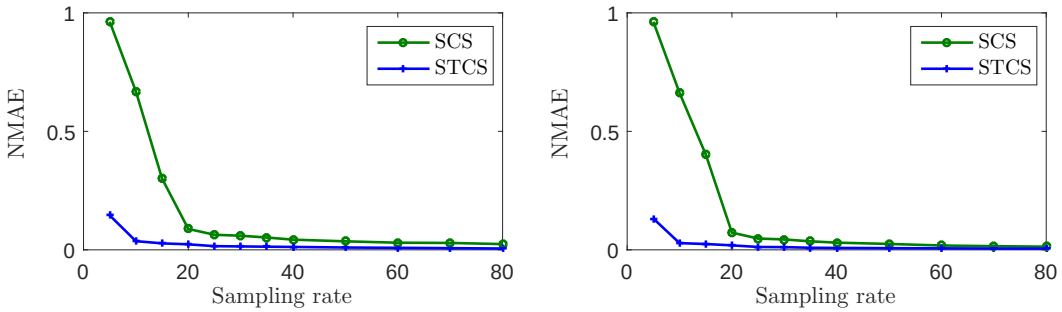
2.6.2.2 Data recovery performance

In order to test the efficiency of our method in collecting data, we use the Normalized Mean Absolute Error (NMAE) metric to measure the resulting recovery performance. this metric is defined as follows:

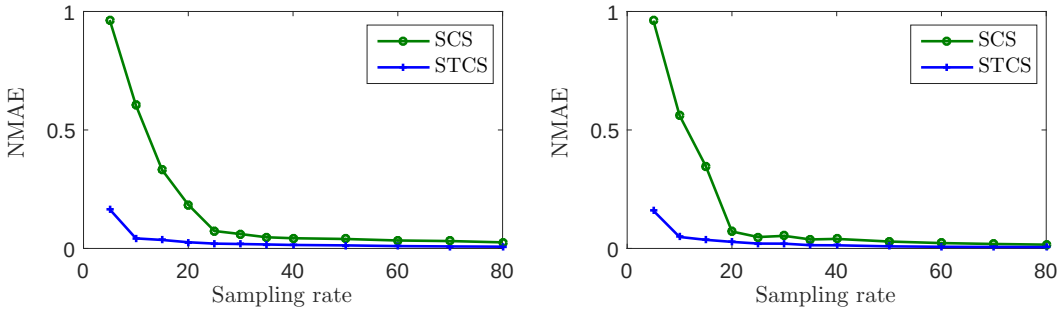
$$\text{NMAE} = \frac{\sum_{i=1}^{N} |d_{i,t} - \hat{d}_{i,t}|}{\sum_{i=1}^{N} |d_{i,t}|}. \tag{2.19}$$

We run our STCS algorithm under different choices of sampling rates ranging from 5% to 80% to illustrate the capabilities of our method in terms of data recovering accuracy. We studied the performance of our method under three different choices of anomaly percentage, namely, 5%, 10% and 15% of measurements. Simulation results are depicted in Figure 2.7. For the temperature dataset, the STCS algorithm succeeds to estimate the transmitted values and correct aberrant readings with a high precision. For all sampling rates higher than 15%, our approach can perfectly estimate the coded measurements with a precision nearly equals to 99% for all considered anomaly percentages. The frequency of outliers has a small effect on the data recovery performance which mostly depends on the sampling rate. For example, from only 5% of raw data, we can estimate missing values with a recovery error less than 10%. Whilst, SCS fails to recover the transmitted measurements and achieve poor performance, especially when the sampling rate is less than 15%. In this case, the recovery precision of the SCS techniques varies between 5% and 75% in the best cases. Again, including the past observations into the data recovery process demonstrates an important improvement in the system performance.

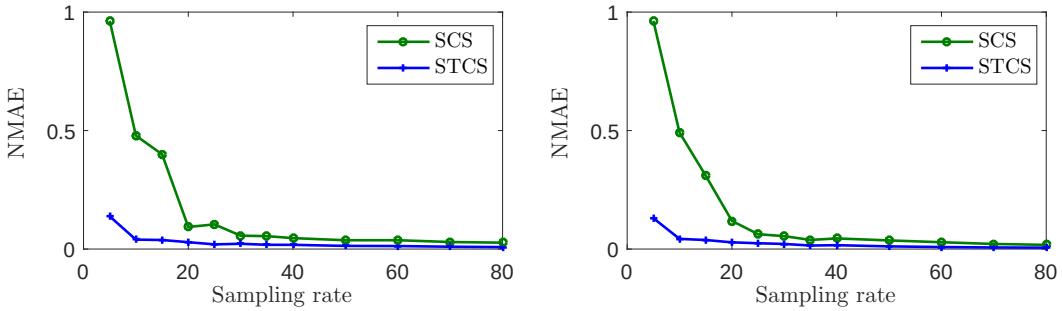
For the humidity dataset and under the same anomaly probability distribution, STCS behaves in a similar way as in the temperature measurements recovering case. Here again, one can easily note that the frequency of outliers has a small effect on the recovery performance in different simulated sampling rates. For instance, from only one quarter of data, we can estimate the original measurements with a fidelity nearly equals to 99% for all tested anomaly percentages.



Anomaly percentage =5%



Anomaly percentage =10%



Anomaly percentage =15%

Figure 2.7: *Recovery accuracy: Left: on the Intel temperature data. Right: on the Intel humidity data.*

2.7 Conclusion

In this chapter, we proposed a novel data gathering and anomaly detection and correction method. The novelty of this work relies on jointly treating the problem of data gathering as well as anomaly detection and correction while including the past observations into this process. We formulated this problem as a convex optimization problem where the proposed regularization term covers a wide class of the data proprieties. Then, we presented a new primal-dual algorithm to derive a solution to the proposed penalized criterion. The experiments carried out on two datasets show the important improvements brought by our solution while compared to SCS, the reference solution in CS-based schemes. These improvements are achieved under different choices of anomaly densities and sampling rates.

One main advantage of our proposed solution is its ability to balance the traffic carried by nodes through the network based on a static routing strategy. However we can identify some drawbacks in this technique. In fact, because of the sensory data projection and transmission is performed in a collaborative way, a node failure may dramatically affect the quality of the recovered information. To overcome this limitation, we propose in the next chapter an incomplete-data-based approach that allows to efficiently collect the sensory readings and detect the potential anomalies.

- Chapter 3 -

Regularized matrix completion solution for efficient data gathering and anomaly detection

3.1 Introduction

We have addressed in the previous chapter the problem of data gathering and anomalies detection using a CS-based technique. We continue in this chapter focusing on the same problem while using an incomplete-data-based technique. In fact, this class of technique allow to better handle the problem of nodes' failure and incomplete datasets.

In this chapter, we introduce a novel scheme that allows to solve the joint problem of data gathering and anomaly detection and correction while optimizing the use of WSN resources. The proposed solution is based on Matrix Completion (MC) approaches and it allows to fill the missing elements in the data matrix as well as to detect potential anomalies. The aforementioned goal is achieved by exploiting the low

rank structure of accurate measurements and the sparsity of the occurrence of outlying values. The robustness of our approach is enhanced by taking full advantage of the local data structure such as the sparsity of readings under some transform domain. The problem of data gathering and anomaly detection is formulated as a tractable convex optimization problem on the Hilbert space of data measurements and anomalies. To solve the resulting optimization problem, we develop a primal-dual based algorithm that allows to incorporate different data features. The efficiency of our method is evaluated using two real datasets. We demonstrate that the proposed algorithm achieves good data recovery and anomaly detection performance while it outperforms state-of-the-art techniques.
This chapter is organized as follows: Section 3.2 introduces our motivation and the context of this work. A formal presentation of the studied problem is introduced in Section 3.3. In Section 3.4, we detail the proposed variational approach and the underlying algorithm. Section 3.6 is devoted to the simulation study. Finally, concluding remarks are given in Section 3.7.

3.2 Context of the study

As we have mentioned in the first chapter, the applications of WSNs are multiple and have multitude of purposes. One main goal of WSN is collecting sensory data and detecting the potential anomalies. However, depending on the application specificities, there is always an order of priorities between these two functionalities; Either collecting data while filtering anomalies to get a high quality of the collected sensory readings, or observing the collected data and identifying anomalies to get the user alerted in a case of emergencies or event occurrence. In both cases, the system requirements should be achieved with the minimum use of WSN resources.
As an example of applications where the data quality requirements overpass the security exigences, the system described in [Naumowicz et al., 2008; Guilford et al., 2009] where a WSN is deployed in Skomer island, a UK national nature reserve, to monitor and observe seabirds. The deployed sensors measure a set of variables such as the ambient temperature, the presence of birds in the borrows or the movement in the entrance of the borrows. The priority in this application as well as most scientific monitoring systems is to collect the most realistic data while optimizing the use of WSN resources and minimizing human interventions.
On the other hand, there are many WSN driven applications where security is in the heart of system requirements and its design. As an example, we can cite the application described in [Whittle et al., 2013] where a WSN is deployed to monitor a water pipeline installation in Singapore to locate and detect water leakage. In this descried system, the priority is given to reliably identify outlying values at each sensor node rather than collecting normal readings.

The technique developed in the previous chapter can achieve both detailed goals since it allows estimating both readings and outlying values. One main advantage of

the STCS algorithm is its ability to balance the traffic carried by sensors. The values probed by each node is coded and transmitted to the sink which would dispose of a compressed information relative to all sensory data. Hence, STCS seems well adapted to detect individual sensor anomalies and fit the applications that privilege the security requirements. For the applications that focus on the data quality rather than on anomaly detection, one can still utilize this technique. However, for a better use of WSN resources, we can use other energy efficient techniques such as MC-based techniques [Cheng et al., 2013]. In fact, this class of techniques conserve sensor energy by promoting node sleeping mode. Moreover, due to the static routing strategy deployed in CS-based techniques, a node failure may induce dramatic effects on the data quality since it carries one piece of the coded data that is necessary to the decoding step. For example, if the sink have two descendant nodes and one of them runs out of energy causing a faulty data transmissions, then half of the useful information will be lost which render the process of data extraction impossible. Finally, STCS technique requires to store each basis vector at a specific node and the order of nodes is fixed in advance. Yet, this operation can be costly or impossible in some observed environments. So, even thought, STCS achieves good performances, especially for the anomaly detection part, it is sensitive to many factors.

In this chapter, we describe an efficient solution that allows to collect the sensory data while correcting the potential anomalies to ensure a high quality of the gathered data. The proposed solution is based on MC techniques where only an arbitrary subset of nodes forward their readings to the sink. To reconstruct the missing values, the collector node takes advantage of some *a priori* knowledge about the target data such as the low rank structure of the normal data or the sparsity of readings under some transform domain. The proposed solution is extended to deal with the presence of anomalies under the condition that the number of outliers is small compared to the number of available measurements.

In the sequel, we formulate the problem of anomaly detection and data gathering as a tractable convex optimization problem on the Hilbert space of data measurements and anomaly matrices. We introduce a primal-dual algorithm to solve the resulting problem. The proposed algorithm takes into account different data features, namely, the sparsity of the projections of readings in some basis, the low rank structure of the regular data matrix and the sparsity of outliers among the available set of readings. We conduct simulations on real data where we considered different anomaly distributions and percentages under different sampling rates to prove the efficiency of our method in terms of anomaly detection and correction and data recovering.

3.3 Problem formulation

In this work, we suppose that we have a WSN composed of N sensor nodes deployed within an environment to monitor. Each node transmits the probed measurements to the sink through a multi-hop path. The process of data acquisition is performed

periodically at each time slot t. During M time slots, the sink would collect $N \times M$ data readings organized into an $N \times M$ matrix $\mathbf{X}$. For all $i \in \{1, \ldots, N\}$ and $j \in \{1, \ldots, M\}$, we denote by X_{ij} the correct value of the data measured by the node i at the time slot j. We assume that a random portion of measurements is corrupted. The considered outliers could capture significant changes in the system state or error measurements. We also suppose that the number of outliers is small compared to the number of available measurements and we denote by $\mathbf{E}$ the $N \times M$ matrix of anomalies.
In order to enhance the lifespan of sensors and reduce their energy consumption, only a random subset of nodes participates at each time slot to the process of data gathering. The sink associates zero to missing values corresponding to the inactive nodes. Let $\mathbf{Y}$ denotes the $N \times M$ matrix of data received by the sink during M time slots which can be expressed as follows:

$$\mathbf{Y} = \mathbf{M} \,.*\, (\mathbf{X} + \mathbf{E}), \tag{3.1}$$

where $.*$ denotes the Hadamard product operator and $\mathbf{M}$ denotes the sampling matrix that indicates the state of sensor nodes at each time slot. It can be defined as follows:

$$M_{ij} = \begin{cases} 1 & \text{if } X_{ij} \text{ is transmitted} \\ 0 & \text{otherwise.} \end{cases} \tag{3.2}$$

Observing $\mathbf{Y}$, the sink aims to reliably estimate the anomaly matrix $\mathbf{E}$ and reconstruct the data matrix $\mathbf{X}$ by exploiting the *a priori* knowledge about the data and the anomaly structures.
In the sequel, we will explore these structures to formulate our optimization problem.

3.4 Regularized Matrix completion in the presence of anomalies

In this section, we formally introduce our regularized MC-based solution in the presence of anomalies. Then we explore some key data features, and based on these features, we propose a mathematical formulation that allows recovering missing sensor readings with maximum fidelity to the real ones while determining the positions and the values of outliers.
In order to fill missing values from an incomplete partially corrupted set of measurements, we propose to solve the following optimization problem:

$$\underset{\mathbf{X},\mathbf{E}}{\text{minimize}}\ h(\mathbf{X}, \mathbf{E}) + \phi(\mathbf{X}, \mathbf{E}), \tag{3.3}$$

where $h(\mathbf{X}, \mathbf{X}) = \frac{1}{2}\|\mathbf{M}. * (\mathbf{X} + \mathbf{E}) - \mathbf{Y}\|_F^2$ is the data fidelity term and $\phi(\mathbf{X}, \mathbf{X})$ is the regularization term. $\|.\|_F$ denotes the Frobenius norm. Such a reconstruction is possible by suitably choosing the regularization term. Indeed, this term models the *a prior* knowledge about the target data, essentially the sparsity of outliers compared to the number of transmitted data and the low rank structure of $\mathbf{X}$. The recovery

accuracy could be enhanced by investigating the local structure of the data such as the sparsity of the projection of one sensor data in some transform domain, and the sparsity of all measurements taken at the same time under the same transform. We propose to decompose the regularization term as follows:

$$\phi(\mathbf{X}, \mathbf{E}) = \phi_r(\mathbf{X}) + \phi_a(\mathbf{E}) + \phi_l(\mathbf{X}), \tag{3.4}$$

where ϕ_r, ϕ_a and ϕ_l are the low rank, the anomaly and the local structure regularization terms, respectively.
In the following, we will explore the data and anomaly properties in order to integrate them into the optimization problem depicted by equation (3.3).

3.4.1 Low rank structure

To analyze the low rank nature of $\mathbf{X}$, we use the Singular Value Decomposition (SVD). The data matrix X can be decomposed as follows:

$$\mathbf{X} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^\top, \tag{3.5}$$

where $\mathbf{U}$ and $\mathbf{V}$ are respectively an $N \times N$ and $M \times M$ orthonormal matrices, and $\mathbf{\Sigma}$ is an $N \times M$ diagonal matrix formed by the singular values $\sigma_1, \sigma_2, \ldots, \sigma_r$ arranged in decreasing order. For low ranked matrices, the data points lie in a space with dimension equals to the number of the highest singular values which is always smaller than the ambiant dimension of $\mathbf{X}$. Based on this property, MC-based techniques postulate that we can recover $\mathbf{X}$ with high fidelity using few samples [Candès and Recht, 2009]. This reconstruction was made possible by minimizing the data fidelity term subject to a convex surrogate of the rank. We model the data low rank nature using the nuclear norm since it is the closest convex approximation of the low rank structure. Hence, the low rank regularization term can be written as follows:

$$\phi_r(X) = \lambda_1 \|\mathbf{X}\|_*, \tag{3.6}$$

where $\|\mathbf{X}\|_*$ denotes the nuclear norm of $\mathbf{X}$ defined as the sum of the singular values and λ_1 is a positive tuning parameter.

3.4.2 Anomaly structure

A simple low rank assumption about the structure of the underlining data is not sufficient to deal with the imperfections corrupting real datasets. Under the realistic assumption that the presence of outliers is not frequent compared to the intact measurements, we can model the irrelevant values using a sparse matrix $\mathbf{E}$. Hence, we incorporate this structure using the ℓ_1-norm since it well promotes the sparsity of $\mathbf{E}$. A relevant choice of the anomaly structure regularization function is:

$$\phi_a(\mathbf{E}) = \lambda_2 \|\mathbf{E}\|_1, \tag{3.7}$$

where $\|\mathbf{E}\|_1$ denotes the ℓ_1-norm of $\mathbf{E}$ and λ_2 is a positive tuning parameter. The ℓ_1-norm based penalty allows to capture points that lie outside the nominal low rank component of $\mathbf{X}$.

3.4.3 Local structure

In spite of the fact that we can recover the data matrix and filter the anomalies based on both previous detailed regularization terms. The accuracy of the reconstruction can be enhanced by exploiting the local data structure, namely the sparsity of readings under some transform domain and/or the inter-columns and inter-rows correlations. Let $\mathbf{R} \in \mathbb{R}^{Q_1 \times M}$ and $\mathbf{C} \in \mathbb{R}^{Q_2 \times N}$ be the associated row and columns operators, respectively. Then, the transform of the columns of the matrix $\mathbf{X}$ is given by $\mathbf{CX}$ while the transform of the rows is given by $\mathbf{XR}^\top$. The correlation between points of the same columns or rows is taken into account by jointly processing the components belonging to the same columns or rows. Then, the local regularization function takes the following form:

$$\phi_l(\mathbf{X}) = \lambda_3 f_1(\mathbf{XR}^\top) + \lambda_4 f_2(\mathbf{CX}), \tag{3.8}$$

where λ_3 and λ_4 are positive regularization parameters and f_1 and f_2 are defined for all $\mathbf{U} \in \mathbb{R}^{N \times M}$ as follows:

$$f_1(\mathbf{U}) = \sum_{i=1}^{N} g(\|\mathbf{U}_{i,:}\|), \tag{3.9}$$

$$f_2(\mathbf{U}) = \sum_{j=1}^{M} g(\|\mathbf{U}_{:,j}\|), \tag{3.10}$$

where $\|.\|$ denotes the ℓ_2 norm, $g : \mathbb{R} \longrightarrow \mathbb{R}^+$ and for all $i \in \{1, \ldots, N\}$ and $j \in \{1, \ldots, M\}$, $\mathbf{U}_{:,j}$ and $\mathbf{U}_{i,:}$ denote respectively the j^{th} column and the i^{th} row of the matrix $\mathbf{U}$. In this work, we use the Generalized Gaussian regularization defined for all $t \geqslant 0$ as $g(t) = t^\beta$ where $\beta > 0$ is the shape parameter. It reduces to the neg-log-likelihood of the normal distribution when $\beta = 2$. It promotes tails that are heavier than normal when $\beta < 2$ reflecting the sparsity of data [Marnissi et al., 2013; Kwitt et al., 2009]. In particular, we obtain the ℓ_2 norm when $\beta = 1$. Besides, the choice of the operators $\mathbf{C}$ and $\mathbf{R}$ defines the adapted penalization strategy. For example, they can represent the DCT transform operator or a frame analysis operator. We can also choose $\mathbf{R}$ to model the temporal smoothness of readings, i.e., the small change between two consecutive time slots, and that by fixing $\mathbf{R}$ as the matrix that computes the discrete difference between horizontal neighboring values. Similarly, $\mathbf{C}$ can be chosen to model the small difference between measurements probed by adjacent sensors by fixing $\mathbf{C}$ as the matrix that expresses which sensors are close to each others. Thus, local regularization functions used in [Zhang et al., 2009; Cheng et al., 2013] can fit in this optimization criterion by carefully choosing the temporal and spatial operators and fixing $\beta = 2$.

3.5 Primal-dual splitting algorithm for Data Gathering and Anomaly Detection (DGAD)

In order to solve the optimization problem depicted in equation (3.4), we propose to apply a primal-dual algorithm. We opt for this class of algorithms since it is specifically designed to minimize a criterion of the form (3.4) which can be written as the sum of a differentiable function h with a Lipschitz gradient, and non-differentiable regularization functions. The data fidelity term h is a convex smooth function and it is continuously differentiable with 1-Lipschitz gradient:

$$\nabla h\left(\begin{bmatrix}\mathbf{X}\\ \mathbf{E}\end{bmatrix}\right)=\begin{bmatrix}\mathbf{M}.*(\mathbf{X}+\mathbf{E})-\mathbf{Y}\\ \mathbf{M}.*(\mathbf{X}+\mathbf{E})-\mathbf{Y}\end{bmatrix} \tag{3.11}$$

The data fidelity term h is introduced into the algorithm using its gradient expression while the non-smooth components of the regularization function are involved via their proximity operators.
The proximal operator of the nuclear norm is simply given by the shrinkage operator [Cai et al., 2010] expressed as follows:

$$\text{prox}_{\lambda\phi_r}(\mathbf{X})=\mathbf{U}\text{diag}\left(\{\max(0,\sigma_i-\lambda)\}_{1\leqslant i\leqslant r}\right)\mathbf{V}^\top, \tag{3.12}$$

where $\mathbf{U}$, $\mathbf{V}$ and $(\sigma_i)_{1\leqslant i\leqslant r}$ are defined as in (3.5).
The proximal operator associated to the anomaly structure is given by the following expression:

$$\text{prox}_{\lambda\phi_a}(E_{i,j})=\begin{cases}E_{i,j}-\lambda & \text{if} \quad E_{i,j}\geqslant\lambda\\ E_{i,j}+\lambda & \text{if} \quad E_{i,j}\leqslant-\lambda\\ 0 & \text{otherwise,}\end{cases} \tag{3.13}$$

where $E_{i,j}$ denotes the anomaly that corresponds to the position (i,j) of the anomaly matrix.
The proximal operators of f_1 and f_2 depend on the choice of β and can be easily computed using [Combettes and Pesquet, 2010]. The proposed algorithm alternatively repeats the computation of the gradient of the data fidelity term and the proximal operators associated to each regularization term. The detailed steps are explained in Algorithm 5. One main advantage of the proposed algorithm is that it allows to solve (3.4) for arbitrary choices of linear operators $\mathbf{R}$ and $\mathbf{C}$ [Combettes and Pesquet, 2011; Komodakis and Pesquet, 2015].
Under some conditions about the choice of γ, the convergence of the sequence $((\mathbf{X}_t)_{t\in\mathbb{N}},(\mathbf{E}_t)_{t\in\mathbb{N}})$ generated by the primal-dual algorithm to a global minimizer of the proposed criterion is guaranteed for $\beta\geqslant 1$ from the results in [Combettes and Pesquet, 2011].

Algorithm 5 Primal-dual algorithm for data gathering and anomaly detection and correction

Initialization

$\gamma > 0$, $\mathbf{X}_0, \mathbf{E}_0, \mathbf{v}_{1,0}, \mathbf{v}_{2,0}, \mathbf{v}_{3,0}, \mathbf{v}_{4,0} \in \mathbb{R}^{N \times M}$

Iterations

For $t = 0, ..$

Gradient computation

$$\begin{bmatrix} \mathbf{w}_{0,t} \\ \mathbf{w}_{1,t} \end{bmatrix} = \nabla h \left(\begin{bmatrix} \mathbf{X}_t \\ \mathbf{E}_t \end{bmatrix} \right)$$

$$\mathbf{y}_{0,t} = \mathbf{w}_{0,t} + \mathbf{v}_{2,t}\mathbf{R} + \mathbf{C}^\top \mathbf{v}_{3,t} + \mathbf{v}_{1,t}$$
$$\mathbf{y}_{1,t} = \mathbf{w}_{1,t} + \mathbf{v}_{4,t}$$
$$\mathbf{p}_{0,t} = \mathbf{X}_t - \gamma\ \mathbf{y}_{0,t}$$
$$\mathbf{p}_{1,t} = \mathbf{E}_t - \gamma\ \mathbf{y}_{1,t}$$

Proxy computation

$$\mathbf{y}_{2,1,t} = \mathbf{v}_{1,t} + \gamma\ \mathbf{X}_t$$
$$\mathbf{y}_{2,2,t} = \mathbf{v}_{2,t} + \gamma\ \mathbf{X}_t\mathbf{R}^\top$$
$$\mathbf{y}_{2,3,t} = \mathbf{v}_{3,t} + \gamma\ \mathbf{C}\mathbf{X}_t$$
$$\mathbf{y}_{2,4,t} = \mathbf{M}\ .*\ (\mathbf{v}_{4,t} + \gamma\ \mathbf{E}_t)$$
$$\mathbf{p}_{2,1,t} = \mathbf{y}_{2,1,t} - \gamma\ \mathrm{prox}_{\gamma^{-1}\lambda_1\phi_r}(\gamma^{-1}\mathbf{y}_{2,1,t})$$
$$\mathbf{p}_{2,2,t} = \mathbf{y}_{2,2,t} - \gamma\ \mathrm{prox}_{\gamma^{-1}\lambda_2 f_1}(\gamma^{-1}\mathbf{y}_{2,2,t})$$
$$\mathbf{p}_{2,3,t} = \mathbf{y}_{2,3,t} - \gamma\ \mathrm{prox}_{\gamma^{-1}\lambda_3 f_3}(\gamma^{-1}\mathbf{y}_{2,3,t})$$
$$\mathbf{p}_{2,4,t} = \mathbf{y}_{2,4,t} - \gamma\ \mathrm{prox}_{\gamma^{-1}\lambda_4\phi_a}(\gamma^{-1}\mathbf{y}_{2,4,t})$$

Averaging

$$\mathbf{q}_{2,1,t} = \mathbf{p}_{2,1,t} + \gamma\ \mathbf{p}_{0,t}$$
$$\mathbf{q}_{2,2,t} = \mathbf{p}_{2,2,t} + \gamma\ \mathbf{p}_{0,t}\mathbf{R}^\top$$
$$\mathbf{q}_{2,3,t} = \mathbf{p}_{2,3,t} + \gamma\mathbf{C}\mathbf{p}_{0,t}$$
$$\mathbf{q}_{2,4,t} = \mathbf{p}_{2,4,t} + \gamma\ \mathbf{p}_{1,t}$$
$$\mathbf{v}_{1,t+1} = \mathbf{v}_{1,t} - \mathbf{y}_{2,1,t} + \mathbf{q}_{2,1,t}$$
$$\mathbf{v}_{2,t+1} = \mathbf{v}_{2,t} - \mathbf{y}_{2,2,t} + \mathbf{q}_{2,2,t}$$
$$\mathbf{v}_{3,t+1} = \mathbf{v}_{3,t} - \mathbf{y}_{2,3,t} + \mathbf{q}_{2,3,t}$$
$$\mathbf{v}_{4,t+1} = \mathbf{v}_{4,t} - \mathbf{y}_{2,4,t} + \mathbf{q}_{2,4,t}$$

Update

$$\begin{bmatrix} \mathbf{w}_{2,t} \\ \mathbf{w}_{3,t} \end{bmatrix} = \nabla h \left(\begin{bmatrix} \mathbf{p}_{0,t} \\ \mathbf{p}_{1,t} \end{bmatrix} \right)$$

$$\mathbf{q}_{1,t} = \mathbf{w}_{2,t} + \mathbf{p}_{2,2,t}\mathbf{R} + \mathbf{C}^\top \mathbf{p}_{2,3,t} + \mathbf{p}_{2,1,t}$$
$$\mathbf{q}_{2,t} = \mathbf{w}_{3,t} + \mathbf{p}_{2,4,t}$$
$$\mathbf{X}_{t+1} = \mathbf{X}_t - \mathbf{y}_{0,t} + \mathbf{p}_{0,t} - \gamma\ \mathbf{q}_{1,t}$$
$$\mathbf{E}_{t+1} = \mathbf{E}_t - \mathbf{y}_{1,t} + \mathbf{p}_{1,t} - \gamma\ \mathbf{q}_{2,t}.$$

3.6 Experimental results and analysis

To evaluate the performance of our proposed solution, we run extensive simulations on two real datasets and under different anomaly configurations. Again, we consider the benchmark datasets of humidity and temperature readings provided by Intel Berkeley Research Lab. We compare our proposed algorithm with two state-of-the-art anomaly detection and estimation techniques that we propose to combine with a MC-based data gathering technique in order to achieve a multitask solution as brought by our method. In the sequel, we present our simulation methodology and we specify the methods that we compare with our approach. Finally, we deliberate the simulation results that deal with different performance evaluation prospects.

3.6.1 Simulation methodology

In this experiments, we construct the data matrix by observing 300 successive time slots. Thus, $\mathbf{X}$ and $\mathbf{E}$ are 52×300 matrices.
The performance of our method relies on three main factors: the data sampling rate, the percentage of anomalies among transmitted readings and the outlying probability distribution. In this experiment, we randomly drop entries from the data matrix $\mathbf{X}$ according to a predefined sampling rate. Then, we randomly generate the set of anomalies according to some predefined probability mass function $p_E(.)$. Finally, we randomly choose the positions of anomalies among active nodes according to a predefined outlier rate. The received data matrix $\mathbf{Y}$ is obtained by combining the incomplete data matrix $\mathbf{X}$ and the anomaly matrix $\mathbf{E}$. Regarding the data features described previously, we propose to take full advantage of the low rank structure and the sparsity of the projection of rows and columns under the DCT domain. Hence, we set $\mathbf{R}$ and $\mathbf{C}$ to the DCT operator and we fix $\beta = 1$ to capture the sparsity of the data transform. Since our proposed approach performs anomaly detection in an unsupervised fashion and for a seek of fairness, we propose to compare our DGAD algorithm with two state-of-the-art unsupervised anomaly detection techniques combined to an efficient MC-based data gathering technique. The raw data is filtered using the considered outlier detection techniques. Then, the incomplete data are estimated using the MC-based technique.
The first considered anomaly detection technique is a statistical based method that consists on predicting each sensor reading based on its predecessor transmitted ones using a linear regression model [Witten et al., 2011]. Based on the statistical behaviour of sensory readings, the sink could predict the value of the transmitted measurement and compare it with the received one at each time slot. If the gap between the two values exceed a certain threshold, then the sink declares the received value as outlier and replaces it with the predicted one for further processing.
The second considered anomaly detection technique is a distance-statistical based method that uses the Hampel identifier to detect and remove anomalies [Liu et al., 2004; Suomela, 2014]. For each received measurement, this technique computes the

median of a window composed of the actual reading and a predefined number of its temporal neighboring readings. Then it estimates the standard deviation of the sample about its window median using the median absolute deviation. If the reading differs from the median by more than three standard deviations, then it is declared as anomaly and replaced by the median.

After filtering anomalies, a Primal-Dual Matrix Completion (PDMC) algorithm for data gathering is applied to fill missing values [Moussa et al., 2016]. This algorithm takes advantage of the same data features as in our data DGAD solution, namely the low rank structure and the sparsity of measurements under the DCT domain. However, PDMC solution do not offer a proper tool to deal with the presence of anomalies. Note that we consider this PDMC algorithm to highlight the improvement brought up by jointly treating the problem of data gathering and anomaly detection rather than treating each individual problem separately. In the sequel, we denote the linear regression anomaly detection method combined to the PDMC algorithm by LR-PDMC, and we denote the Hampel Filter combined to the PDMC algorithm by HF-PDMC.

The regularization parameters λ_1, λ_2, λ_3 and λ_4 introduced in (3.4) control the contribution of each penalized term. Hence, an arbitrary choice of these parameters may leads the algorithm to converge toward a non-optimal solution. We should find the parameters that give the best trade-off between the data fidelity term, the low-rank feature, the sparsity constraints and the sparsity of anomalies. In our simulations, we set them empirically and we choose the ones giving the best results. Concerning the Linear Regression (LR) method, we empirically set the gap that triggers the presence of anomaly to 1. The number of neighbors that define the window size in the Hampel Filter method (HF) is empirically set to 4.

In the following, we demonstrate the anomaly detection capabilities of our method. Then, we focus on the data recovering performance.

3.6.2 Anomaly detection performance

To evaluate the capacity of our method in terms of separating outliers from original measurements, we simulate different possible anomaly sets drawn according to different normal probability distributions. Note that we consider the normal probability distribution since outliers in real conditions occur due to a variety of random phenomena. Hence, from the central limit theorem, they behave according to the normal distribution.

To illustrate the process of anomaly detection and correction proceeded before filling the missing values, we plot in Figure 3.1 the raw transmitted data gathered from one sensor at a sampling rate equals to 50% and spotted with 15% of anomalies drown according to a centred normal distribution with a variance equals to 3. We present the different anomaly detection technique responses toward the treated data. The DGAD technique succeeds to follow the data deviation while capturing almost all the anomalies. While the LR technique confuses between the brisk changes on data

values and the occurrence of anomalies. (The HF method achieves also good anomaly detection performance).

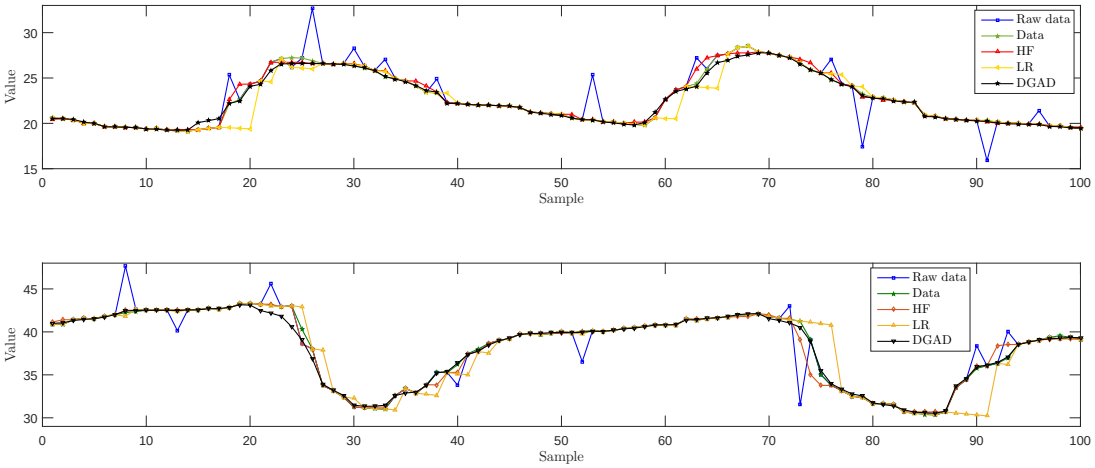


Figure 3.1: *Anomaly detection and correction example. Top: on Intel temperature data. Bottom: on Intel humidity data.*

Since outlier magnitude depends on the considered anomaly distribution, we propose to evaluate the robustness of our method under different normal distribution parameters by varying the mean and the variance of the generator probability mass function. Figure 3.2 and Figure 3.3 illustrate the resolution capabilities of the considered methods in terms of separating normal values from added anomalies. We plot the responses of different techniques to 40% of data corrupted with 15% of normally distributed anomalies. The histogram of the difference between the estimated and the real outlier values and the plots of detected anomalies under different probability distributions sets demonstrate the capacity of our method in achieving good anomaly detection performance compared to the two state-of-the-art techniques for both treated datasets. Beyond the visual-based results, we propose to evaluate our anomaly detection technique using the detection accuracy and the false alarm metrics. We evaluate the performance of our DGAD method under two different available data sizes, namely 10% and 40% of total readings. The anomalies are drawn according to a centered normal distribution with variance equals to 2.

As illustrated in Figures 3.4 and 3.5, our proposed approach can effectively identify and estimate the outlying readings with a high detection accuracy and a low false alarm in most treated cases and for both datasets. Our DGAD algorithm can reach a high detection accuracy even when the amount of available data is considerably low while maintaining a small false alarm rate. contrarily, the false alarm rate reached by the HF method is considerable despite that it achieves a good detection accuracy rate. The LR method fails to distinguish between normal and abnormal readings and achieves

both high false alarm rate and low detection accuracy rate. The poor performance of the LR method is due to the lack of sufficient data, since this class of techniques are sensitive toward missing data.

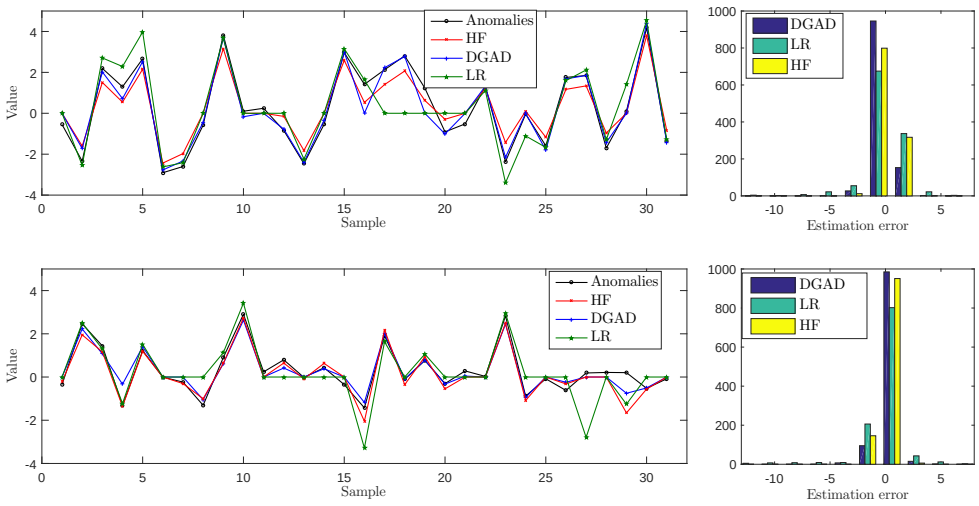


Figure 3.2: *Anomaly detection and estimation on Intel humidity data using a centered anomaly distribution. Top: variance equals to* 2*. Bottom: variance equals to* 1*.*

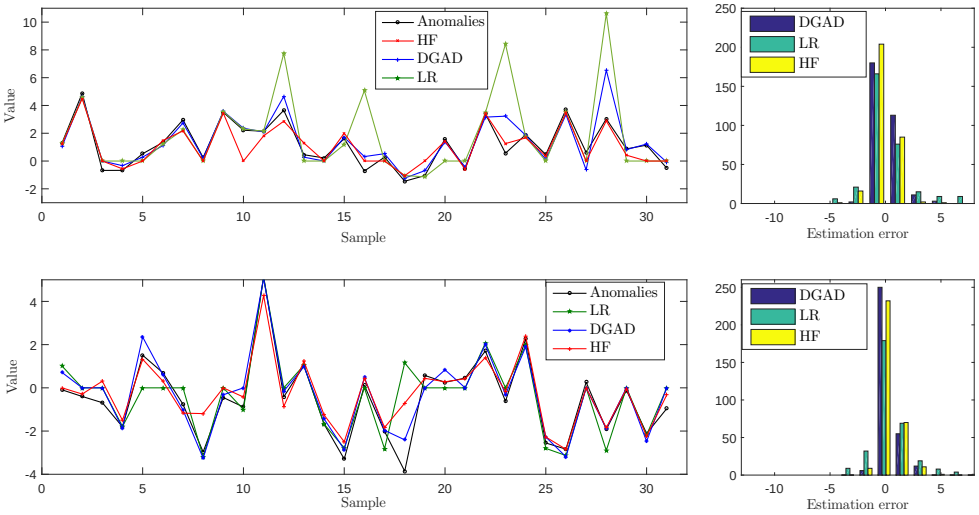


Figure 3.3: *Anomaly detection and estimation on Intel temperature data using a normal anomaly probability distribution. Top: mean equals to* 1 *and variance equals to* 2*. Bottom: mean equals to* 1 *and variance equals to* 2*.*

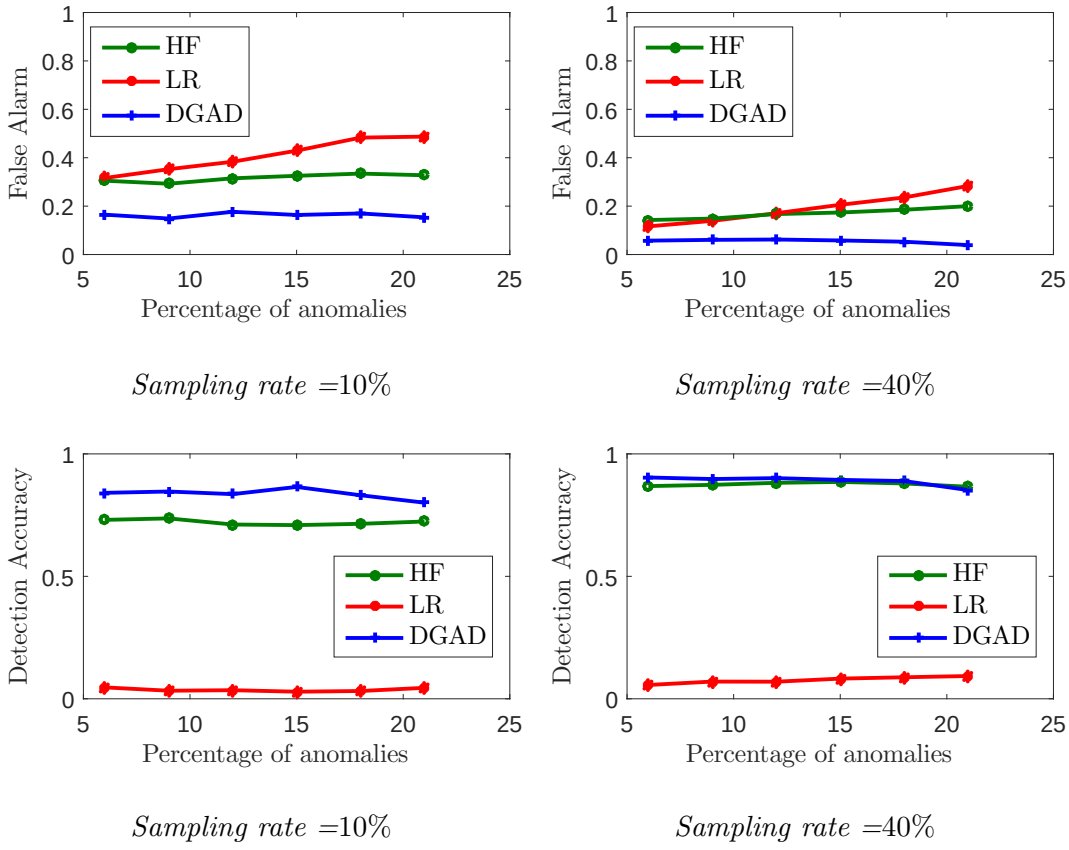


Figure 3.4: *Anomaly detection performance on Intel temperature data.*

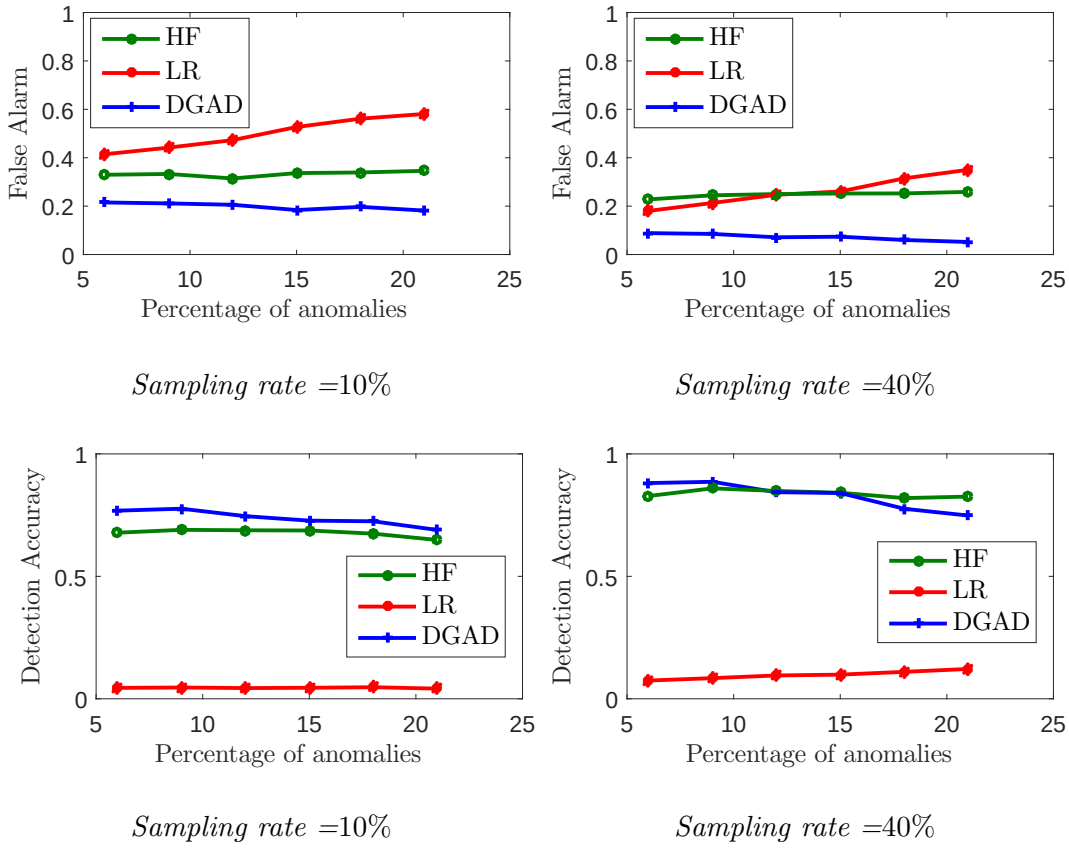


Figure 3.5: *Anomaly detection performance on Intel humidity data.*

3.6.3 Missing data recovery performance

To evaluate the data gathering efficiency of our method, we use the Normalized NMAE metric to measure the resulting recovery performance. In the context of MC-based approaches, this metric express the sum of the absolute difference between the missing values and the estimated ones decided by the sum of the absolute value of missing readings:

$$\text{NMAE} = \frac{\sum_{(i,j)|M_{i,j}=0} |X_{i,j} - \hat{X}_{i,j}|}{\sum_{(i,j)|M_{i,j}=0} |X_{i,j}|}. \tag{3.14}$$

The recovery accuracy of our DGAD algorithm is tested under different choices of missing measurement rates ranging from 5% to 50%. We analyze the performance of our method under three different choices of anomaly percentages, namely, 6%, 12% and 18% of the available data at the sink. Furthermore, we simulate centred normal anomaly distribution with a variance equals to 2 for both datasets. Each simulation is conducted for 10 independent random trials and the recovery errors are averaged over the 10 trials.

Simulation results are depicted in Figure 3.6. For the Temperature dataset, the DGAD algorithm succeeds to reconstruct missing values and correct aberrant readings with a precision higher than 94% for all considered sampling rates and anomaly percentages. The frequency of outliers has a small effect on the data recovery performance that mostly depends on the amount of the available raw data. From only 5% of the raw data, we can estimate missing values with a recovery error less than 5%. Whilst, HF-PDMC and LR-PDMC achieve the twice recovery error for the same amount of available data. When the sampling rate goes higher, the DGAD algorithm reaches better performance for all tested anomaly percentages. W can reconstruct the humidity data matrix with a precision nearly equals to 99% in the presence of 18% of corrupted readings and with a sampling rate higher or equal to 20%.

For the humidity dataset and under the same anomaly probability distribution, DGAD behaves in a similar way as in the temperature measurement recovering case. The DGAD technique achieves good performance on Intel humidity data and outperforms HF-PDMC and LR-PDMC techniques. The frequency of outliers impacts the recovery performance in low sampling rates. In fact, the less the data is corrupted, the better the recovery accuracy is. When the sampling rate increases, DGAD becomes immune to the frequency of outliers. For example, from only one quarter of data, we can estimate the original data matrix with a fidelity nearly equals to 98% for all tested anomaly percentages.

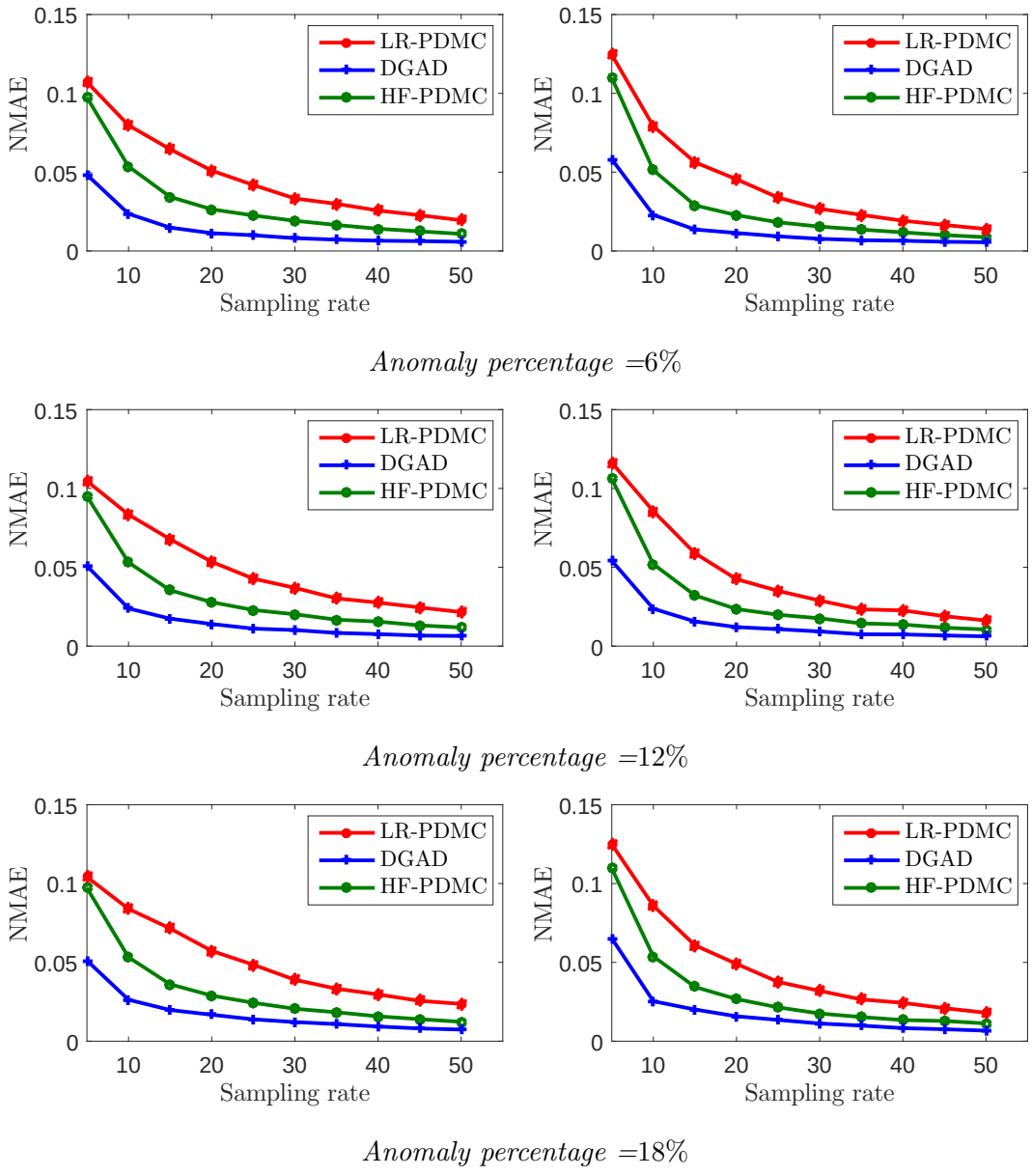


Figure 3.6: *Recovery accuracy. Top: on Intel temperature data. Bottom: on Intel humidity data.*

3.7 Conclusion

In this chapter, we proposed a novel MC-based data gathering and anomaly detection and correction method. The proposed approach exploits the low rank structure of regular sensor readings and the sparsity of the presence of outliers. The novelty of this work relies on jointly treating the problem of data gathering and anomaly detection and correction. We formulated this problem as a convex optimization problem where the proposed regularization term covers a wide class of data proprieties. Then, we presented a new primal-dual algorithm to derive the solution of the proposed penalized criterion. The Experiments carried out on two datasets show the good performance of our method under different choices of anomaly distributions and sampling rates. However, one should note that the performance of our solution depends on the regularization parameters and do not incorporate directly WSN features in the algorithm. In the next chapter we provide a solution where the involved parameters have a physical signification, so they can easily set into the proposed solution.

- Chapter 4 -

Constrained matrix completion approach for data gathering and anomaly detection

4.1 Introduction

In the previous chapter, we have proposed an incomplete-data-based approach to collect the sensory data and detect the potential anomalies. A major and strong point of the previously proposed solution is its ability to integrate different data prospects. However, this strong point involves some weak articulations since the underlying data features are included using some regularization functions and the performance of the proposed algorithm depends heavily on their associated tuning parameters.

In this chapter, we propose a more intuitive formulation of the data gathering and anomaly detection problem using MC-based approaches. We aim to optimize the usage of WSN resources during the data gathering and anomaly detection process while providing a practical algorithm that takes into account various data features. To achieve this goal, we model the *a priori* knowledge about the sensory data via hard convex constraints where the involved regularizations are related to some physical properties of the data itself and are then easy to interpret and adjust. Indeed, the problem of data gathering and anomaly detection is reduced to a tractable constrained minimization one. Thereafter, we extend a new class of primal-dual algorithms to solve the resulting optimization problem. The experiments carried out on the two real datasets previously used allowed to show that this last proposed algorithm outperforms the state-of-the-art methods and achieves an even better recovery accuracy for different data sampling levels, compared to the previously studied regularization approach.

The rest of this chapter is organized as follows: In the next section, we present some motivations that initiate this works. In Section 4.3, we present the proposed variational approach and the underlying algorithm. We evaluate the performance of our solution in Section 4.4. Finally, concluding remarks are given in Section 4.5.

4.2 From regularized approaches to a constrained formulation

Most energy saving techniques in WSNs are built upon optimization problems. These methods involve the minimization of a penalized criterion composed of two terms. First, a well defined data fidelity term that links the solution to the observation so that it will be consistent with the observed data. Second, a regularization term that incorporates an *a priori* knowledge about the target measurements in order to ensure the stability of the obtained solution. This term may be composed of several functions, each one expresses one data feature (e.g. low rank, sparsity, temporal smoothness), and is multiplied with a positive constant namely the regularization parameter. The latter control the relative weights of the data fidelity and the regularization functions. Hence, the underlying optimization problem can be expressed as follows:

$$\underset{\mathbf{X}}{\text{minimize}}\ h(\mathbf{X},\mathbf{Y}) + \sum_{i=1}^{l} \lambda_i \phi_i(\mathbf{X}), \tag{4.1}$$

where $\mathbf{X}$ is the system unknown variable and $\mathbf{Y}$ is the system observation *i.e.*, the received data. The data fidelity term and the regularization functions are denoted respectively by $h(.)$ and $\phi_i(.)$. Each regularization function is controlled by the tuning parameter λ_i. As the number of regularized functions increases, one can expect a better estimation of the data matrix, but since the number of the regularization parameters increases, their choice becomes a more challenging task and will highly affect the quality of the solution. Hence, an important issue in these methods is the setting of

these parameters, which may depend on the properties of the WSN and the measured phenomena.

To overcome the shortcoming of these methods, one may prefer to adopt a constrained formulation rather than a regularized one. In fact, thanks to the Lagrangian duality, any penalized estimation problem can be reformulated as a constrained problem. Hence, the optimization problem depicted in equation (4.1) can be expressed as follows:

$$\underset{\mathbf{X}\in\cap_{i=1}^{l} C_i}{\text{minimize}}\ h(\mathbf{X},\mathbf{Y}), \tag{4.2}$$

where C_i is the equivalent constraint to the regularization function $\phi_i(.)$. The constraint bounds may be related to some knowledge on the observation process or the statistical and physical properties of the target data. Moreover, it has been very often recognized in many fields that when incorporating prior informations directly on the solution, the choice of the involved parameters becomes easier [Youla and Webb, 1982; Teuber et al., 2013].

In this chapter, we propose to address the joint problem of data gathering and anomaly detection as a constrained optimization inverse problem using MC-based approaches. Observing an incomplete and potentially corrupted data matrix, the sink tempts to reliably recover the missing values and detect the potential anomalies by resorting to some constraints that reflect the prior knowledge about the target data. In fact, data features such as the sparsity in some transform domain or the low rank structure can be expressed through hard constraints modelled by nonempty closed convex sets. One of the main advantages of such a formulation is that it facilitates the choice of the related parameters compared to the regularized approach which was investigated in the previous chapter.

4.3 Proposed variational approach

We consider the problem of data gathering while detecting the potential anomalies in WSNs. The proposed solution is based on MC-based techniques. Hence, only a subset of readings is transmitted to the sink at each time slot and based on these few measurements, the collector node tries to estimate missing values while detecting the potential anomalies. More rigorously, the equation relating the data matrix $\mathbf{X}$ and the anomaly matrix $\mathbf{E}$ to the received data matrix $\mathbf{Y}$ can be expressed as follows:

$$\mathbf{Y} = \mathbf{M}.*(\mathbf{X}+\mathbf{E}), \tag{4.3}$$

where $.*$ denotes the Hadamard product operator and $\mathbf{M}$ denotes the sampling matrix that indicates the state of sensor nodes at each time slot:

$$M_{ij} = \begin{cases} 1 & \text{if } X_{ij} \text{ is transmitted} \\ 0 & \text{otherwise.} \end{cases} \tag{4.4}$$

and X_{ij} denotes the value probed by sensor i at the time slot j for all $i \in \{1 \dots N\}$ and $j \in \{1 \dots M\}$.
Observing $\mathbf{Y}$, the sink aims to reliably estimate the data matrix $\mathbf{X}$ and the anomaly matrix $\mathbf{E}$ by minimizing a data fidelity terms subject to some closed convex constraints related to the probed data features. In the following, we explore the set of the considered constraints in order to achieve this goal.

4.3.1 Considered constraints

4.3.1.1 Sensors value range constraint

The range of possible values taken by sensors are always known and indicated in their technical data sheet. This value range depends on the probed physical magnitude and the observed environment. We encode this property by the following constraint:

$$C_1 = \{\mathbf{X} \in \mathbb{R}^{N \times M} \ | X_{i,j} \in [X_{min}, X_{max}]\} \tag{4.5}$$

where $X_{min}$and X_{max} define the bounds of possible measurements.

4.3.1.2 Low rank constraint

To analyze the low rank feature, we use the Singular Value Decomposition (SVD). The data matrix $\mathbf{X}$ can be decomposed as follows:

$$\mathbf{X} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^\top, \tag{4.6}$$

where $\mathbf{U}$ and $\mathbf{V}$ are respectively a $N \times N$ and $M \times M$ orthonormal matrices, and $\mathbf{\Sigma}$ is a $N \times M$ diagonal matrix formed by the singular values $\sigma_1, \sigma_2, \dots, \sigma_r$ arranged in a decreasing order. The rank is defined by the number of non zero singular values. For low ranked matrices, only a few number of singular values captures most of the energy of $\mathbf{X}$. Most data matrices in WSNs exhibit the low rank feature that can be translated by the following constraint:

$$C_2 = \{\mathbf{X} \in \mathbb{R}^{N \times M} | \text{rank}(\mathbf{X}) \leqslant \mu\}, \tag{4.7}$$

where μ is a positive constant.

4.3.1.3 Data sparsity constraint

Sensor readings are very often sparse in some transform domain. Only few coefficients of their sparse representation concentrate the most energy of the data matrix $\mathbf{X}$. Let $\mathbf{F}_1 \in \mathbb{R}^{L \times N}$ and $\mathbf{F}_2 \in \mathbb{R}^{P \times M}$ be the linear operators associated with the considered transform analysis for the rows and columns respectively. Let $\mathbf{D} \in \mathbb{R}^{L \times P}$ denotes the sparse representation of the data matrix $\mathbf{X}$, i.e. $\mathbf{D} = \mathbf{F}_1\mathbf{X}\mathbf{F}_2^\top$. We suppose that there exist two block-selection operators $\mathbf{R}_1 \in \mathbb{R}^{P \times P}$ and $\mathbf{R}_2 \in \mathbb{R}^{P \times P}$ (resp. $\mathbf{C}_1 \in \mathbb{R}^{L \times L}$ and

$\mathbf{C}_2 \in \mathbb{R}^{L\times L}$) for the rows (resp. columns) that allow to split the sparse representation D into two matrices $\mathbf{D}_1 \in \mathbb{R}^{L\times P}$ and $\mathbf{D}_2 \in \mathbb{R}^{L\times P}$ such that $\mathbf{D} = \mathbf{D}_1 + \mathbf{D}_2$, where $\mathbf{D}_1 = \mathbf{C}_1\mathbf{D}\mathbf{R}_1^\top$ and $\mathbf{D}_2 = \mathbf{C}_2\mathbf{D}\mathbf{R}_2^\top$. $\mathbf{D}_1$ captures the highest transform coefficients that occupy the most energy of $\mathbf{X}$ while $\mathbf{D}_2$ captures the vanishing ones (i.e $\mathbf{D}_{1_{i,j}} = \mathbf{D}_{i,j}$ and $\mathbf{D}_{2_{i,j}} = 0$ for all $(i,j) \in \{1,\dots,L\}\times\{1,\dots,P\}$ corresponding to the positions of the highest transform coefficients, otherwise $\mathbf{D}_{1_{i,j}} = 0$ and $\mathbf{D}_{2_{i,j}} = \mathbf{D}_{i,j}$).

The sparsity feature is modeled using the following two constraints:

$$C_3 = \{\mathbf{X} \in \mathbb{R}^{N\times M}, \mathbf{D}_1 = \mathbf{C}_1\mathbf{F}_1\mathbf{X}\mathbf{F}_2^\top\mathbf{R}_1^\top \mid \|\mathbf{D}_1\|_{\mathrm{F}} \leqslant \delta\} \tag{4.8}$$

where $\|.\|_{\mathrm{F}}$ is the Frobenius norm with $\|\mathbf{X}\|_{\mathrm{F}} = \sqrt{\sum_{i,j} X_{ij}^2}$ and δ is fixed depending on the target data and the used operator transform, and

$$C_4 = \{\mathbf{X} \in \mathbb{R}^{N\times M}, \mathbf{D}_2 = \mathbf{C}_2\mathbf{F}_1\mathbf{X}\mathbf{F}_2^\top\mathbf{R}_2^\top \mid |D_{2_{i,j}}| \leqslant \epsilon\} \tag{4.9}$$

where ϵ is a small positive value that penalizes the vanishing values of the data matrix sparse representation.

Different choices of the operator $\mathbf{F}_1$ and $\mathbf{F}_2$ lead to different penalization strategies. For example, these operators can model either a DCT or a DWT. Moreover, we can also choose $\mathbf{F}_2$ as the matrix computing the discrete difference between horizontal neighboring values to model the short-term stability feature. In this case, the constraint C_4 allows to prevent strong variations between two successive measurements provided by each sensor node, i.e:

$$C_4 = \{\mathbf{X} \in \mathbb{R}^{N\times M}, |X_{i,j+1} - X_{i,j}| \leqslant \epsilon\} \tag{4.10}$$

4.3.1.4 Sparsity of outliers

In the absence of outliers, the previous detailed constraints are sufficient to estimate the data matrix by minimizing some constrained optimization problem over the variable $\mathbf{X}$. However, real datasets are far from being pure and anomaly free sets. Hence, an effective approach to deal with the presence of anomalies is by proposing a robust estimator that jointly seek for the data and anomaly matrices.

Unlike the data features that are expressed through hard constraints, anomalies randomly occur in WSNs and can not be modeled using a fixed constraint. Thus, we propose to integrate the presence of outliers using some regularization functions. Based on the realistic assumption that outliers are sporadic and not frequent compared to the accurate measurements, we propose to penalize this structure using the ℓ_1-norm since it promotes the sparsity of the anomaly occurrence. Indeed, the anomaly regularization function can be expressed as follows:

$$\phi(\mathbf{E}) = \lambda\|\mathbf{E}\|_1, \tag{4.11}$$

where $\|.\|_1$ denotes the ℓ_1-norm and λ is a positive tuning parameter. The ℓ_1-norm based penalty allows to capture points that lie outside of the intersection of all considered normal data constraints.

4.3.2 Constrained Robust Matrix Completion (CRMC) algorithm

To best estimate the missing values while detecting the potential anomalies, based on the received data and the considered constraints, the sink has to solve the following constrained minimization problem:

$$\underset{\mathbf{X}\in\cap_{m=1}^{4}C_m,\ \mathbf{E}\in\mathbb{R}^{N\times M}}{\text{minimize}}\ h(\mathbf{X},\mathbf{E})+\lambda\phi(\mathbf{E}) \tag{4.12}$$

where $h(\mathbf{X})=\frac{1}{2}\|\mathbf{M}.*(\mathbf{X}+\mathbf{E})-\mathbf{Y}\|_F^2$ is the data fidelity term. It can be noted that h is a convex smooth function and is continuously differentiable with 1-Lipschitz gradient:

$$\nabla h\left(\begin{bmatrix}\mathbf{X}\\ \mathbf{E}\end{bmatrix}\right)=\begin{bmatrix}\mathbf{M}.*(\mathbf{X}+\mathbf{E})-\mathbf{Y}\\ \mathbf{M}.*(\mathbf{X}+\mathbf{E})-\mathbf{Y}\end{bmatrix} \tag{4.13}$$

Our objective is to provide a numerical solution to the problem described by (4.12). This amounts to minimizing the function h with respect to $\mathbf{X}$ and $\mathbf{E}$ by taking into account that our solution is constrained to belong to the constraint sets defined previously and the outiling values are not frequent campared to the remaining set of data. Some of the considered constraints are expressed via linear operators namely the DCT or DWT. Thus, we propose to adopt primal dual approaches since they avoid some large-size matrix inversions which may be numerically intractable or computationally costly [Combettes and Pesquet, 2010]. Since h is a Lispchitz-differentiable function, we propose to use specifically the Monotone+Lipschitz Forward-Backward-Forward (M+LFBF) algorithm [Combettes and Pesquet, 2011]. The latter alternates the computations of the gradient of h, the projections onto the convex constraints sets and the proximity operator of the anomaly penalization term. The proximity operator of the ℓ_1-norm is obtained as in the previous chapters. The projections onto the first and last constraint are straightforward. The projection onto C_3 is given by the following expression:

$$\text{proj}_{C_3}(\mathbf{D}_1)=\begin{cases}\mathbf{D}_1 & \text{if } \|\mathbf{D}_1\|_{\text{F}}\leqslant\delta\\ \dfrac{\delta}{\|\mathbf{D}_1\|_{\text{F}}}\mathbf{D}_1 & \text{otherwise.}\end{cases} \tag{4.14}$$

We propose to relax the rank with the nuclear norm in order to obtain a convex optimization problem as it is the closest convex approximation of the low rank nature of data. Hence, this constraint can be equivalently expressed as follows:

$$C_2=\{\mathbf{X}\in\mathbb{R}^{N\times M}\mid\|\mathbf{X}\|_*\leqslant\lambda\}. \tag{4.15}$$

where λ is a positive parameter related to the target data.

Algorithm 6 M+LFBF algorithm for Constrained Robust Matrix Completion (CRMC)

Initialization

$\gamma > 0$, $\mathbf{X}_0, \mathbf{E}_0, \mathbf{v}_{1,0}, \mathbf{v}_{2,0}, \mathbf{v}_{3,0}, \mathbf{v}_{4,0} \in \mathbb{R}^{N \times M}$

Iterations

For $t = 0, ..$

Gradient computation

$$\begin{bmatrix} \mathbf{w}_{0,t} \\ \mathbf{w}_{1,t} \end{bmatrix} = \nabla h \left(\begin{bmatrix} \mathbf{X}_t \\ \mathbf{E}_t \end{bmatrix} \right)$$

$$\mathbf{y}_{0,t} = \mathbf{w}_{0,t} + \mathbf{F}_1^\top (\mathbf{C}_1^\top \mathbf{v}_{2,t} \mathbf{R}_1 + \mathbf{C}_2^\top \mathbf{v}_{3,t} \mathbf{R}_2) \mathbf{F}_2 + \mathbf{v}_{1,t}$$

$$\mathbf{y}_{1,t} = \mathbf{w}_{1,t} + \mathbf{v}_{4,t}$$

$$\mathbf{p}_{0,t} = \mathbf{X}_t - \gamma\, \mathbf{y}_{0,t}$$

$$\mathbf{p}_{1,t} = \mathbf{E}_t - \gamma\, \mathbf{y}_{1,t}$$

Projection and proxy computation

$$\mathbf{p}_{0,t} = \mathrm{proj}_{C_1}(\mathbf{p}_{0,t})$$

$$\mathbf{y}_{2,1,t} = \mathbf{v}_{1,t} + \gamma\, \mathbf{X}_t$$

$$\mathbf{y}_{2,2,t} = \mathbf{v}_{2,t} + \gamma\, \mathbf{C}_1 \mathbf{F}_1 \mathbf{X}_t \mathbf{F}_1^\top \mathbf{R}_1^\top$$

$$\mathbf{y}_{2,3,t} = \mathbf{v}_{3,t} + \gamma\, \mathbf{C}_2 \mathbf{F}_1 \mathbf{X}_t \mathbf{F}_1^\top \mathbf{R}_2^\top$$

$$\mathbf{y}_{2,4,t} = \mathbf{M} \,.*\, (\mathbf{v}_{4,t} + \gamma\, \mathbf{E}_t)$$

$$\mathbf{p}_{2,1,t} = \mathbf{y}_{2,1,t} - \gamma\, \mathrm{proj}_{\gamma^{-1} C_2}(\gamma^{-1} \mathbf{y}_{2,1,t})$$

$$\mathbf{p}_{2,2,t} = \mathbf{y}_{2,2,t} - \gamma\, \mathrm{proj}_{\gamma^{-1} C_3}(\gamma^{-1} \mathbf{y}_{2,2,t})$$

$$\mathbf{p}_{2,3,t} = \mathbf{y}_{2,3,t} - \gamma\, \mathrm{proj}_{\gamma^{-1} C_4}(\gamma^{-1} \mathbf{y}_{2,3,t})$$

$$\mathbf{p}_{2,4,t} = \mathbf{y}_{2,4,t} - \gamma\, \mathrm{prox}_{\gamma^{-1} \lambda_4 \phi_a}(\gamma^{-1} \mathbf{y}_{2,4,t})$$

Averaging

$$\mathbf{q}_{2,1,t} = \mathbf{p}_{2,1,t} + \gamma\, \mathbf{p}_{0,t}$$

$$\mathbf{q}_{2,2,t} = \mathbf{p}_{2,2,t} + \gamma\, \mathbf{C}_1 \mathbf{F}_1 \mathbf{p}_{0,t} \mathbf{F}_1^\top \mathbf{R}_1^\top$$

$$\mathbf{q}_{2,3,t} = \mathbf{p}_{2,m,t} + \gamma \mathbf{C}_2 \mathbf{F}_1 \mathbf{p}_{0,t} \mathbf{F}_1^\top \quad \mathbf{R}_2^\top$$

$$\mathbf{q}_{2,4,t} = \mathbf{p}_{2,4,t} + \gamma\, \mathbf{p}_{1,t}$$

$$\mathbf{v}_{1,t+1} = \mathbf{v}_{1,t} - \mathbf{y}_{2,1,t} + \mathbf{q}_{2,1,t}$$

$$\mathbf{v}_{2,t+1} = \mathbf{v}_{2,t} - \mathbf{y}_{2,2,t} + \mathbf{q}_{2,2,t}$$

$$\mathbf{v}_{3,t+1} = \mathbf{v}_{3,t} - \mathbf{y}_{2,3,t} + \mathbf{q}_{2,3,t}$$

$$\mathbf{v}_{4,t+1} = \mathbf{v}_{4,t} - \mathbf{y}_{2,4,t} + \mathbf{q}_{2,4,t}$$

Update

$$\begin{bmatrix} \mathbf{w}_{2,t} \\ \mathbf{w}_{3,t} \end{bmatrix} = \nabla h \left(\begin{bmatrix} \mathbf{p}_{0,t} \\ \mathbf{p}_{1,t} \end{bmatrix} \right)$$

$$\mathbf{q}_{1,t} = \mathbf{w}_{2,t} + \mathbf{F}_1^\top (\mathbf{C}_1^\top \mathbf{p}_{2,2,t} \mathbf{R}_1 + \mathbf{C}_2^\top \mathbf{p}_{2,3,t} \mathbf{R}_2) \mathbf{F}_2 + \mathbf{p}_{2,1,t}$$

$$\mathbf{q}_{2,t} = \mathbf{w}_{3,t} + \mathbf{p}_{2,4,t}$$

$$\mathbf{X}_{t+1} = \mathbf{X}_t - \mathbf{y}_{0,t} + \mathbf{p}_{0,t} - \gamma\, \mathbf{q}_{1,t}$$

$$\mathbf{E}_{t+1} = \mathbf{E}_t - \mathbf{y}_{1,t} + \mathbf{p}_{1,t} - \gamma\, \mathbf{q}_{2,t}.$$

Similarly, we can show that the projection onto the nuclear-ball has the following expression [Chierchia et al., 2014]:

$$\text{proj}_{C_2}(\mathbf{X}) = \mathbf{U}\ \text{diag}\left(\{(\sigma_i - \nu_0)_+\}_{1\leqslant i\leqslant r}\right)\mathbf{V}^\top, \tag{4.16}$$

where $\mathbf{U}$, $\mathbf{V}$ and $(\sigma_i)_{1\leqslant i\leqslant r}$ are defined as in (4.6), for all $u \in \mathbb{R}$, $(u)_+ = \max\{u, 0\}$, $\nu_0 = \left(\frac{\sum_{i=1}^{d_0}\sigma_i - \lambda}{d_0}\right)_+$ and d_0 is the largest integer in $\{1, \ldots, r\}$ that satisfies $\sum_{i=1}^{r}(\sigma_i - \sigma_{d_0})_+ \leqslant \lambda$.

The M+LFBF algorithm chosen to solve the minimization problem is detailed in Algorithm 6. It alternatively performs the computation of the gradient of h and the projection onto $(C_m)_{1\leqslant m\leqslant 4}$ as well as the proximity operator of the anomaly penalization. Under some conditions about the choice of the step size γ, the convergence of the sequence $(\mathbf{X}_t, \mathbf{E})_{t\in\mathbb{N}}$ generated by the primal-dual algorithm to a global minimizer of the proposed criterion is guaranteed from the results in [Combettes and Pesquet, 2011].

4.4 Simulations and results

To demonstrate the efficiency and the robustness of our method, we propose to evaluate it on the same real datasets as in the previous chapters. Our data matrix is composed of 300 columns delivered by 52 sensors. In the following, we first investigate the considered datasets and discuss their features. Then, we describe the performance evaluation of our proposed algorithm.

4.4.1 Dataset features

Our Constrained Robust Matrix Completion (CRMC) solution is essentially based on some particular features represented by the considered constraints. The main feature that allows to fill missing values is the low rank data propriety. Since we consider the same datasets as in the previous chapter, this feature is fully satisfied. We recall that about 96% of the energy is captured by the first 5 singular values for both datasets.

The considered datasets also exhibit the sparsity feature under the DCT domain. The grayscale image depicted in Figure 4.1 represents the DCT transform of the temperature and humidity datasets where the dark colors capture the high coefficients while the light colors are proportional to the small ones. We can easily observe that most of the energy is concentrated in the first few rows. Hence, the sparsity constraint discussed in the previous section is well satisfied under the DCT domain.

Moreover, the two considered datasets exhibit a short-term stability feature. In fact, the difference between two measurements probed by the same sensor at consecutive time slots is always small. Figure 4.2 shows that the registered variation between two consecutive measurements for both datasets is concentrated around zero.

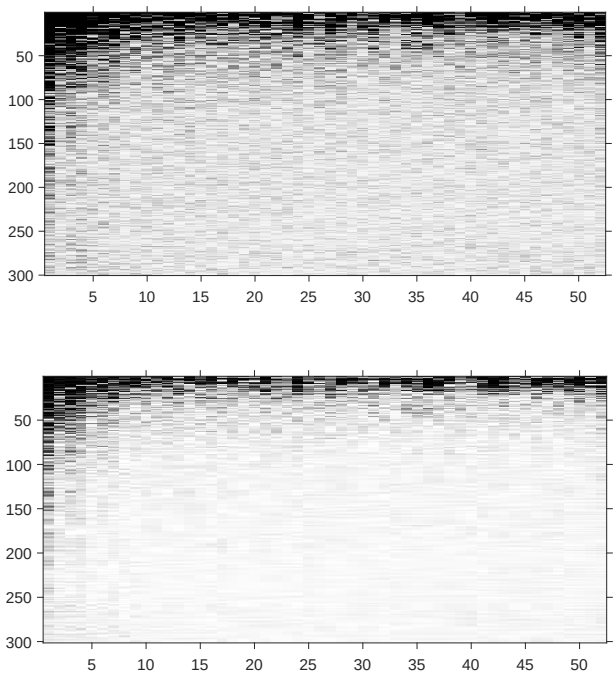


Figure 4.1: *The DCT of the data matrix. Top: Intel humidity dataset. Bottom: Intel temperature dataset.*

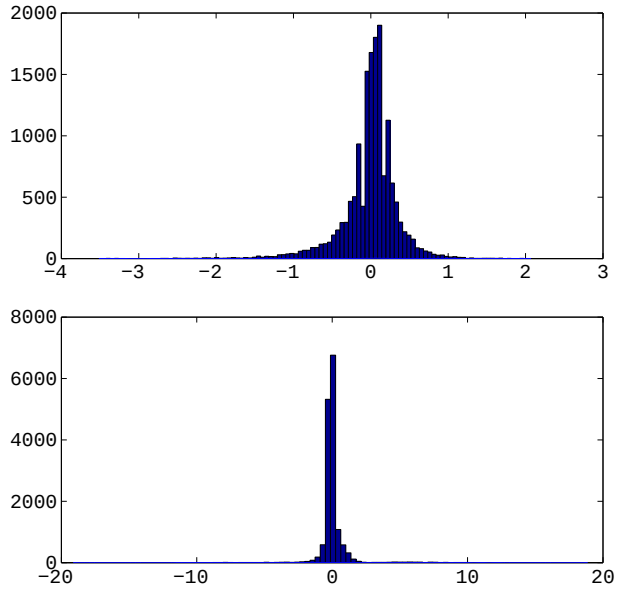


Figure 4.2: *Histogram of the difference between two consecutive. measurements: Top: Intel humidity dataset. Bottom: Intel temperature dataset.*

In the light of the explored features, we propose to evaluate two variants of our CRMC algorithm. The first version of the proposed algorithm integrates the sparsity of data using the DCT transform while the second version takes advantage of the Short Term (ST) stability feature. In the following, CRMC-DCT and CRMC-ST denote the first and the second version of our algorithm respectively.

4.4.2 Evaluation of the Constrained Robust Matrix Completion algorithm

4.4.2.1 Simulation settings

To evaluate the performance of our approach, we randomly drop entries of the data matrix according to a predefined sampling rate which corresponds to the ratio of the active nodes. Then, we spot the remaining dataset by Gaussian distributed anomalies according to a predefined anomaly percentage. We compare our CRMC algorithm with two state-of-the-art techniques. The first one is the solution proposed in the previous chapter named DGAD which is a regularized approach that allows to recover the missing values based on the low rank feature and the sparsity of readings in the DCT domain. We consider this regularized approach to highlight the improvement brought up by considering a constrained formulation, even if the data sparsity is included into the underlying algorithms using different approaches (either considering each row and column of the data matrix or splitting it into two sub-matrices). The second technique is the Spatio-Temporal Compressive-data Gathering (STDG) which is a MC-based approach that relies on the low rank and short-term stability features [Cheng et al., 2013]. The aforementioned technique is coupled with the Hampel filter to remove the potential anomalies before filling the missing values. We consider this later technique to highlight the importance of the constrained formulation on modelling the short term stability feature.

For an accurate performance evaluation, we run different simulations under different sampling rates and anomaly percentages. Each simulation is conducted for 10 independent random trials and the performances are averaged over the number of trials.

Considering the data features described previously, we assume that the maximal rank of the target solution can be chosen to be equal to 5 for both considered datasets. Since we are interested in indoor humidity and temperature traces, the possible values collected by sensors are supposed to range form 0 to 60. We propose also to take full advantage of the sparsity feature in our recovery process. Using CRMC-DCT, the DCT representation of the target signal is split into two parts. The Matrix D_1 is constructed by considering the first 30 rows of the sparse representation, while matrix D_2 contains the remaining coefficients. After observing different data matrices, the parameters δ and ϵ are set around 5000 and 0.01, respectively, for both datasets. When using the short-term stability feature, the matrix D_1 is chosen to store the first column of the considered data matrix while the matrix D_2 represents the first order difference

between the remaining columns. The constraint bounds δ and ϵ are set using the same approach as in CRMC-DCT. In this case, δ represents the norm of the vector containing the probed measurements at the first time slot. It is fixed to 400 and 200 for the humidity and the temperature datasets, respectively. The parameter ϵ reflects the low variation of data over 20 minutes and it is set to 0.2 for both datasets.

4.4.3 Anomaly detection performance

Our CRMC solution is designed to offer a suitable tool to detect faults and anomalies. In order to evaluate the capacity of our algorithm in terms of separating abnormal readings from the data matrix structure, we simulate centred Gaussian outliers with a variance equals to 4. Moreover, our proposed algorithm is tried under different anomaly percentages ranging form 6% to 21% of the amount of the available data. We consider also two different sampling rates, namely 10% and 40% of the total readings.
Again, we propose to evaluate our anomaly detection technique using the detection accuracy and false alarm metrics.

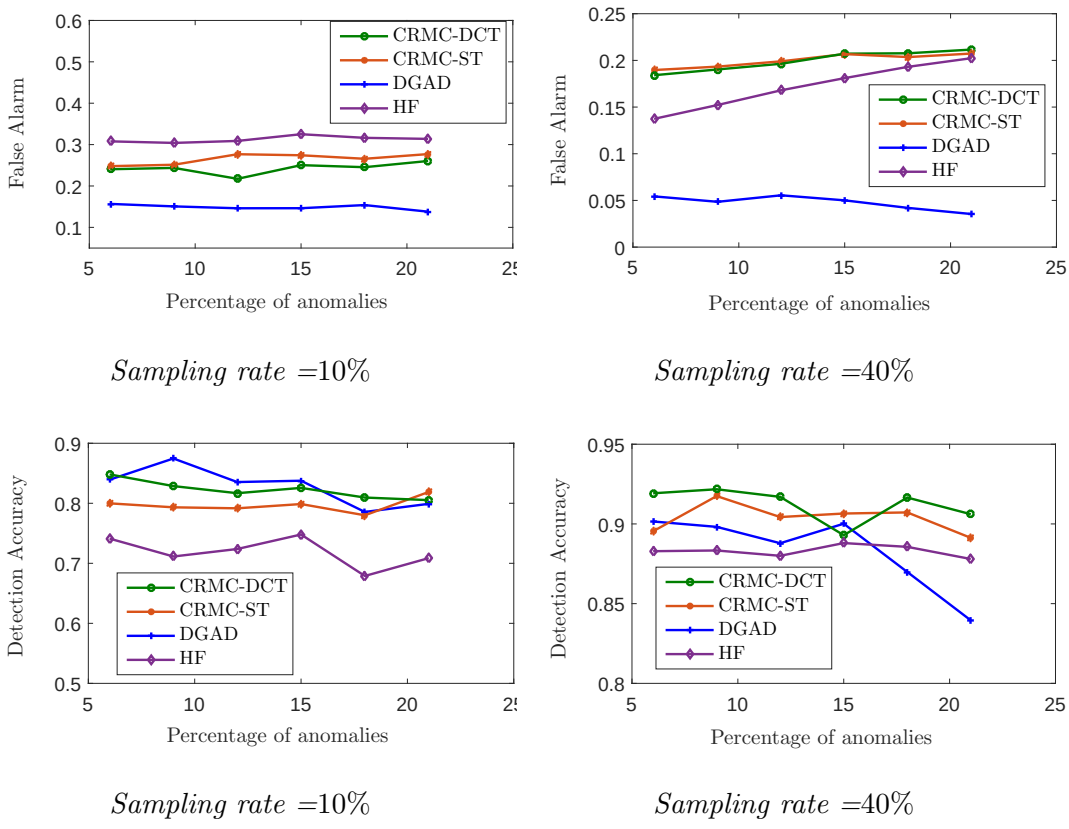


Figure 4.3: *Anomaly detection performance on Intel temperature data.*

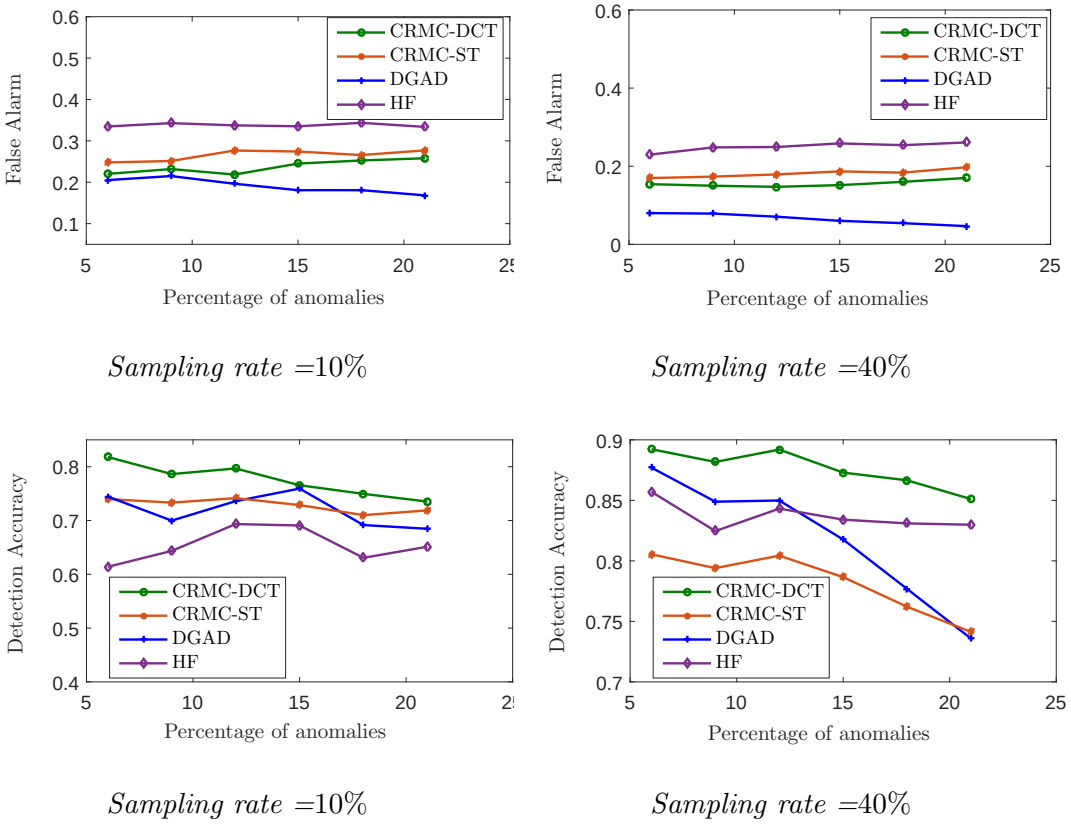


Figure 4.4: *Anomaly detection performance on Intel humidity data.*

The anomaly detection performance of our CRMC technique is depicted in Figures 4.3 and 4.4. For the humidity dataset, the both versions of our CRMC algorithm can achieve a detection accuracy level higher than the other methods for both considered sampling rates. As an example, at the sampling rate of 10% and the outlying percentage of 10%, we can identify nearly 80% (resp. 73%) of anomalies using CRMC-DCT algorithm (resp. CRMC-ST algorithm), while we can achieve only an anomaly detection level equals to 70% using the regularized approach (DGAD) and under the same conditions. By increasing the number of outilers among received readings, the detection accuracy of our method decreases, but it still overpass the other considered techniques.
However, CRMC solution achieve an average false alarm performance compared to the regularized solution. In fact, due to the considered formulation, all data points that lie outside the considered constraints are declared as outlying value, which explain the good detection accuracy level and the average false alarm performance.

For the temperature dataset, our constrained solution behaves in the same way as in the humidity data case and achieves acceptable performance for the different tested

configurations. Here again, Both versions of the CRMC algorithm outperform other techniques in detecting outlying values, especially at high anomaly rates.

4.4.4 recovery accuracy performance

To evaluate the recovery accuracy of our CRMC method, we use the Normalized Mean Absolute Error (NMAE) metric. We recall that the NMAE is defined as follows:

$$\text{NMAE} = \frac{\sum_{(i,j)|M_{i,j}=0} |X_{i,j} - \hat{X}_{i,j}|}{\sum_{(i,j)|M_{i,j}=0} |X_{i,j}|}. \tag{4.17}$$

Figure 4.5 presents the recovery accuracy of all tested algorithms. For the humidity dataset, both simulated variants of our algorithm achieve good performance and especially outperform the STDG technique. The improvement of our approach is notable especially for very low sampling rates. Our solution guaranties a recovery accuracy always higher than 94% for all simulated compression ratio. Both variants of our CRMC solution can achieve a recovery error less than 5% based only on 5% of the entries of the data matrix for all simulated anomaly percentages. On the other hand, using the spatio-temporal correlation between sensory readings, the recovery error achieved by STCDG technique is higher than 23% for the same sampling rate (5%). When the sampling rate goes higher, the recovery accuracy of all methods increases until it almost achieves a comparable performance.

The recovery accuracy performance of our method on the temperature dataset is similar to the one obtained on the humidity dataset for all tested algorithms. In fact, this similarity is due to the physical correlation between the temperature and the humidity magnitudes in an indoor environment.

4.5 Conclusion

In this chapter, we proposed a novel efficient data gathering and anomaly detection algorithm based on robust MC. We modeled the problem of data gathering and anomaly detection as a constrained optimization problem. Our proposed algorithm allows to directly incorporate the physical data features such as the sparsity and the low rank structure into the recovery algorithm through hard convex constraints, which offers a considerable flexibility toward the target data and avoids the problem of adjusting the regularization parameters. The simulation results demonstrate the efficiency of the proposed approach in terms of data recovery quality and anomaly detection capabilities compared to the state-of-the-art algorithms.

Our proposed solution as well as all the previous contributions integrate the spatio-temporal correlation pattern by considering matrix data structures. However, these approaches are demanding in terms or computational resources. In the next chapter, we introduce a novel low complexity solution for data gathering and anomaly detection in WSNs.

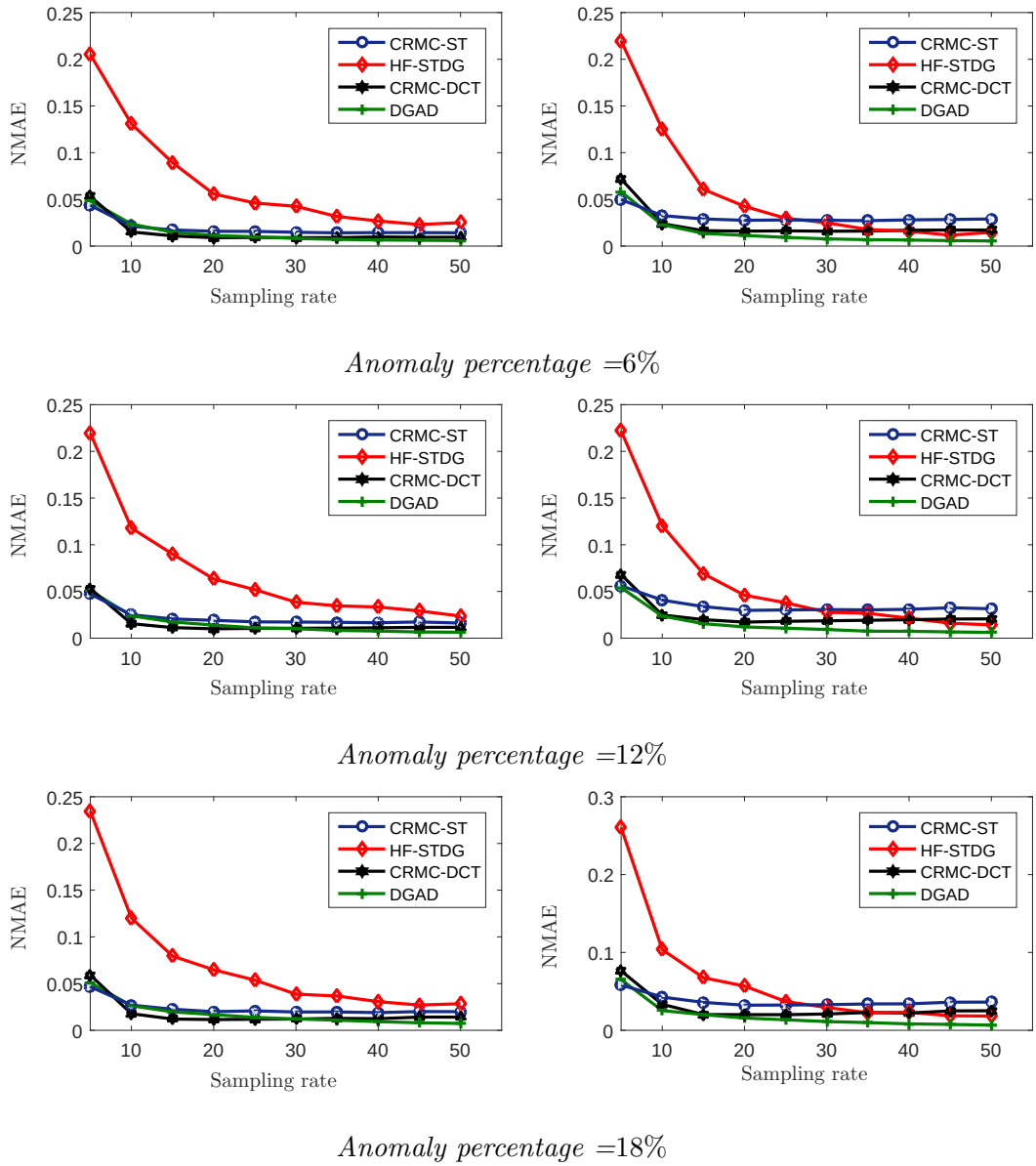


Figure 4.5: *Recovery accuracy: Right: on Intel temperature data. Left: on Intel humidity data.*

- Chapter 5 -

Position-based compressive data gathering

5.1 Introduction

Our previous solutions are based on two main features: the sparsity of the sensory readings in some transform domain and the low rank structure of the resulting data matrix. In addition to these two patterns, we noticed that the sensory sparse representation keep always the same shape where the most relevant coefficients are concentrated in the first positions. Based on this observation, we propose in this chapter a new data gathering scheme that differs from existing solutions. Our idea consists on collaboratively compute and transmit only the largest coefficients of the spatial sparse representation and thus, reduce the number of required samples to reconstruct the original information. Furthermore, we consider a realistic setting where the noisy nature of WSN environments is taken into account. The proposed solution is extended to integrate the temporal sparsity feature as well as the spatial sparsity pattern into the data gathering and recovering process. Thereafter, the recovery performance of

the two versions of our solution is theoretically analyzed. Finally, we describe a peak detection solution that allows to identify spiky data by analyzing the power variation on the compressed information without the need of decoding the forwarded readings at the sink.

This chapter is organized as follows: In the next section, we present the rationale behind our proposed solution. Then, we present in Section 5.3 the links between our approach and CS-based solution. Thereafter, we detail the two versions of our solution in Section 5.4. In Section 5.5, we describe the proposed solution to detect spiky data. Subsequently, we evaluate the performance of our approach in Section 5.6. Finally, Concluding remarks are given in Section 5.7.

5.2 Context and motivation

The *a priori* knowledge about WSN data features has driven to various data gathering and anomaly detection approaches: The sparsity of the sensory data leads to the use of CS-based techniques. Whereas, the low rank pattern due to the correlation between sensor measurements endorses MC-based methods. Both approaches were deeply investigated in the previous chapters where an association between these features (sparsity and low rank features) is used to enhance the recovery and detection performance of the developed algorithms.

Although, the solutions brought by our previous contributions achieve good performance, there is no modifications introduced on the existing routing schemes. In fact, the enhancement in the recovery and detection performance is reached by ameliorating some algorithmic aspects that are proceeded at the sink. However, at a lower level, *i.e.* the sensor nodes, we use the conventional routing schemes depending on whether the proposed solution is a CS or a MC-based approach.

Despite that these routing schemes are well studied in the literature, there is no evidence that they achieve the best performance in terms of optimizing the use of WSN resources. Contrariwise, we can identify some shortcomings: By using CS-based techniques, nodes need to operate at each time slot to transmit their compressed version of readings. Indeed, these techniques incorporate only the spatial sparsity feature in routing sensory readings toward the sink and there are no benefits drawn from the temporal correlation in the routing process, even if the sink use booth sparsity dimensions to recover the compressed data. Therefore, a short sensing period would result in frequent unnecessary sensor transmissions.
On the other hand, using MC-based techniques require the selection of an arbitrarily set of nodes at each time slot which leads to a dynamic change in the network topology. Hence, at each time slot, nodes have to cooperate in order to establish a new routing tree to forward their readings to the sink, which is a non-trivial operation in wireless environments. Moreover, the readings are simply forwarded to the sink using multi-hop paths, which is a simple approach that do not reflect any prior knowledge about the nature of the targeted data.

On top of that, the use of the spatio-temporal correlation pattern in WSN optimization problem is not an easy task. In effect, this feature is usually included by considering matrix data structures and by using a difficult penalization functions such as the SVD decomposition which requires $O(\min\{mn^2, m^2n\})$ operations for a matrix of size $m \times n$ [Holmes et al., 2007]. By using iterative algorithms and for high dimension matrices, recovering sensory data may be costly in terms of execution time and resource utilization.

For all these reasons, we propose in this chapter to develop a low complexity data gathering scheme that is able to achieve a high data recovery performance. The main idea of the proposed solution consists on integrating the *a priori* knowledge about the support set of the sensory sparse representation on the routing and recovering process. In fact, we noticed that relevant positions in the sensory sparse representation are always concentrated in a known support set. For example, the first entries of the DCT of the sensory readings capture most of the signal energy.

Based on this observation, we design a new data gathering scheme called Position-Based Compressive Data Gathering (PBCDG) where nodes cooperatively transmit only the largest coefficients of their sparse representations. The proposed scheme differs from existing CS-based approaches in many aspect: first, we are not constrained to respect CS required conditions to ensure an acceptable signal recovery quality. In fact, CS theory postulates that the minimum number of samples required to estimate the original signal with high probability must exceed a certain threshold [Candes and Tao, 2006]. In general, this threshold is set three times the sparsity level of the target signal. Whereas, our proposed method suggests that we can recover the original signal by using the same number of samples as the sparsity level of the target signal. In other words, CS-based approaches need at least three times more data samples than our PBCDG approach, to achieve the same recovery performance. Second, CS-based techniques require the use of ℓ_1-minimization techniques to recover the coded signal. Whilst, we use a simple projection of the received vector to recover the sensory data. Indeed, knowing the position of the null entries of the sparse representation renders the problem of recovering the sensory readings tractable and easy to solve.

Thereafter, our data gathering solution is extended to incorporate both spatial and temporal correlation dimensions in the data routing process. Moreover, we consider a more realistic setting where the sink is subject to a noisy perturbations. Furthermore, we develop an efficient solution to detect the presence of anomalies without increasing the complexity of the proposed data gathering scheme. The later goal is achieved by analyzing some energetic aspects of the received signal. Indeed, we show that the appearance of spikes in the probed values induces a noticeable variation on the power of the compressed data. Consequently, we can detect the presence of anomalies without decoding the received vector by investigating these properties.

5.3 From CS Data Gathering to Position Based Compressive Data Gathering

Assume that we have a WSN composed of N sensor nodes organized into a particular routing tree that allows them to communicate their readings to the sink. Let $x_{i,t}$ denotes the value probed by sensor i at a given time slot t where $i \in \{1 \dots N\}$. The values probed by all sensors at a given time slot t are represented by the vector $\mathbf{x}_t = [x_{1,t}, \dots, x_{N,t}]$. CS-based techniques are known for their ability to efficiently gather the sensory data based on the spatial sparsity feature. Let $\mathbf{C} \in \mathbb{R}^{N \times N}$ denotes the sparsifying operator, hence we can write:

$$\mathbf{s}_t = \mathbf{C}\mathbf{x}_t, \tag{5.1}$$

where $\mathbf{s}_t$ is the spatial sparse representation of $\mathbf{x}_t$. CS-based techniques postulate that we can recover the vector $\mathbf{x}_t = \mathbf{C}^\top \mathbf{s}_t$ from an observed signal $\mathbf{y}_t \in \mathbb{R}^M$ such that:

$$\mathbf{y}_t = \mathbf{A}\mathbf{s}_t + \mathbf{n}, \tag{5.2}$$

where $\mathbf{n} \sim \mathcal{N}(0, \sigma_n^2 \mathbf{I}_M)$ is a white Gaussian noise and $\mathbf{A} \in \mathbb{R}^{M \times N} (M \ll N)$ is the sensing matrix. To estimate the sonsory data, we need to compute both the locations and the values of the nonzero entries of $\mathbf{s}_t$. Though, if a Genie provides us with $\mathcal{I} = \{i \in \{1 \dots N\} \, | \mathbf{s}_{i,t} \neq 0\}$ the support set of the vector $\mathbf{s}_t$, then the problem is reduced to estimating only the nonzero entries of the target signal. We denote the estimator to this reduced problem by the Genie-Aided Estimator (GAE). This prior knowledge about the support set $\mathcal{I}$ allows the GAE to achieve a Mean Squared Error (MSE) better than any unbiased estimator [Candes and Tao, 2007], where the MSE is defined as follows:

$$MSE(\hat{\mathbf{s}}_t) = E\left[\|\mathbf{s}_t - \hat{\mathbf{s}}_t\|^2\right], \tag{5.3}$$

where $\hat{\mathbf{s}}_t$ is the estimated version of $\mathbf{s}_t$. Many efforts have been made to reliably estimate $\mathbf{x}_t$ from the observed signal $\mathbf{y}_t$ and in the absence of the information relative to the support set of the sparse signal. In [Candes and Tao, 2006], the authors propose an estimator that can achieve the estimation error of the GAE up to a factor of $\log(M)$. The sensing matrix is constructed based on Rademacher projection and a bound-optimization recovery procedure is proposed. Each iteration of the proposed algorithm requires $O(MN)$ operations and the iterations are repeated until convergence is achieved. Another appealing solution is proposed by [Candes and Tao, 2007] where a linear programming based estimator is described. The proposed solution relays on iterative primal-dual interior point methods. Each iteration requires to solve an $M \times M$ system of linear equations and the iterations are repeated until the convergence to an optimal solution. The proposed estimator is able to achieve the estimation error of the GAE up to a factor of $\log(M)$.

However, neither of the previous descried estimators can achieve the Cramér-Rao Bound (CRB) on the estimation error of the GAE which is the theoretical MSE lower

bound that can be achieved by any unbiased estimator. The problem of finding efficient (*i.e.* able to achieve the CRB) and low-complexity estimators that achieve the CRB for noisy compressive sampling remains open [Babadi et al., 2009]. Although, it was proven that the CRB for a noisy CS setting is the same as if a Genie helped the decoder with the support set of the sparse representation of the compressed signal [Babadi et al., 2009], and it is given by:

$$\text{CRB} = \sigma_n^2 \text{Trace}\left\{\left(\mathbf{A}_{\mathcal{I}}^{\top}\mathbf{A}_{\mathcal{I}}\right)^{\dagger}\right\}, \tag{5.4}$$

where $\mathbf{A}_{\mathcal{I}}$ is the submatrix of $\mathbf{A}$ that contains the columns corresponding to the indices in $\mathcal{I}$ (the support set of the sparse representation), $(.)^{\top}$ denotes the transpose operator and $\mathbf{X}^{\dagger} = \mathbf{X}^{\top}(\mathbf{X}\mathbf{X}^{\top})^{-1}$ is a generalized inverse matrix of $\mathbf{X}$ (pseudo-inverse).

Otherwise, the additional information about the support set $\mathcal{I}$ considerably minifies the complexity of the estimation problem and reduces the required number of samples for the recovery process. To recover the compressed signal in this case, we can use the efficient Structural Least Square Estimator (SLSE) described in [Carbonelli et al., 2007] which allows to find the solution of the following problem:

$$\hat{\mathbf{s}}_{t_{\mathcal{I}}} = \underset{\mathbf{s}_{t_{\mathcal{I}}}}{\text{argmin}} \|\mathbf{y}_t - \mathbf{A}_{\mathcal{I}}\mathbf{s}_{t_{\mathcal{I}}}\|^2, \tag{5.5}$$

where $\mathbf{s}_{t_{\mathcal{I}}}$ is a subvector of $\mathbf{s}_t$ that contains only the nonzero entries of $\mathbf{s}_t$.

Inspired from all these works and in attempt to more compress the probed measurements and enhance the WSN performance in noisy environments, we propose a new data gathering scheme that involves the *a priori* knowledge about the support set of the sparsity structure in the process of data routing and recovering. In the sequel, we detail our Position-Based Compressive Data Gathering (PBCDG) solution.

5.4 Position-based compressive data gathering

The goal of our PBCDG solution is to enhance the lifespan of a WSN by balancing the traffic carried by nodes regardless their distance from the sink. To do that, sensors have to collaborate in order to generate a compressed information relative to their probed measurements and thus, reduce the global data traffic through the WSN. The *a priori* knowledge about the support set of the sensory sparse representation not only allows to reach higher compression rates, but also promotes high estimation quality in noisy WSN environments. In fact, the precious information about the positions of the highest sparse coefficients enables us to achieve the CRB for noisy CS settings by using the low complexity SLSE.

To achieve the aforementioned goal, nodes need to take advantage of their spatial correlation by combining the nonzero entries of their spatial sparse representation $\mathbf{s}_t$ while transmitting the coded information toward the sink. Note that real sensory

readings are not perfectly sparse, but they can be approximated using only K entries of $\mathbf{s}_t$. We denote the approximated version of $\mathbf{x}_t$ by $\tilde{\mathbf{x}}_t$.

The process of PBCDG is depicted in Figure 5.1 where the set of routing paths that relays sensor nodes to the sink presents a tree structure. We suppose that nodes are able to memorize their local routing structure. More precisely, each node knows its parent and children nodes. In order to compute the K nonzero entries of the spatial representation of the sensory readings in a distributed way, we store at each node i the $K \times 1$ vector $\mathbf{AC}_i$ where $\mathbf{C}_i$ denotes the i^{th} column of the sparsifying operator $\mathbf{C}$ and $\mathbf{A} \in \mathbb{R}^{K \times N}$ is the sensing matrix that allows to select the nonzero entries of the sparse representation. More preciously, the column i of the matrix $\mathbf{A}$ is set to zero if $i \notin \mathcal{I}$. After the acquisition step, leaf nodes initiate the process of data gathering. Then, each node multiplies the locally probed value by its stored vector and waits till the reception of all its children nodes' vectors. After that, the locally computed vector is concatenated to the received vectors and forwarded to the higher node in the tree structure. The K-dimensional vector produced by the node i at the time slot t, denoted by $\tilde{\mathbf{y}}_{i,t}$, can be expressed as follows:

$$\tilde{\mathbf{y}}_{i,t} = \mathbf{AC}_i x_{i,t} + \sum_{j \in J} \mathbf{AC}_j x_{j,t}, \tag{5.6}$$

where $\mathcal{J}$ denotes the set of indices of the children nodes attached to the sensor i.

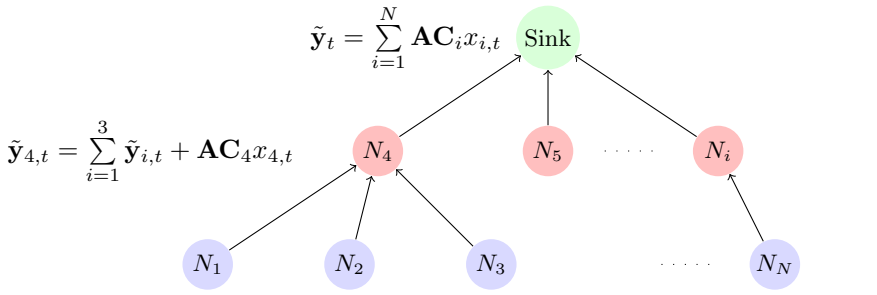


Figure 5.1: *Data gathering process using PBCDG technique.*

The previously described procedure is repeated at each node till the coded data arrive to the sink. After collaboratively transmitting the linear combination of the sensory data projection and in the absence of noise, the sink will receive the k-dimensional vector given by the following expression:

$$\tilde{\mathbf{y}}_t = \sum_{i=1}^{N} \mathbf{AC}_i x_{i,t} = \mathbf{ACx}_t. \tag{5.7}$$

However, the assumption that the received signal is noise-free is not realistic. By taking into account the noisy nature of WSN environments, we can alternatively express the received signal as follows:

$$\mathbf{y}_t = \mathbf{ACx}_t + \mathbf{n} = \mathbf{As}_t + \mathbf{n}, \tag{5.8}$$

where $\mathbf{n} \sim \mathcal{N}(0, \sigma_n^2)$ is a white Gaussian noise. To recover the vector $\mathbf{s}_t$, we propose to use the SLSE. Thus, the estimated projection vector is given by the following expression [Carbonelli et al., 2007]:

$$\hat{\mathbf{s}}_t = \mathbf{A}^\top(\mathbf{AA}^\top)^{-1}\mathbf{y}_t. \tag{5.9}$$

After estimating the transmitted projection vector, the sink can estimate $\tilde{\mathbf{x}}_t$ by projecting $\hat{\mathbf{s}}_t$ into the original basis. Hence, the estimated approximation of the sensory readings $\hat{\tilde{\mathbf{x}}}_t$ can be expressed as follows:

$$\hat{\tilde{\mathbf{x}}}_t = \mathbf{C}^\top\mathbf{A}^\top(\mathbf{AA}^\top)^{-1}\mathbf{y}_t. \tag{5.10}$$

In the presence of a white Gaussian noise, the resulting MSE of the estimated vector $\hat{\tilde{\mathbf{x}}}_t$ can be expressed as follows:

$$\begin{aligned}
MSE(\hat{\tilde{\mathbf{x}}}_t) &= E\left[\|\tilde{\mathbf{x}}_t - \hat{\tilde{\mathbf{x}}}_t\|^2\right] \\
&= E\left[\|\mathbf{C}^\top(\mathbf{s}_t - \hat{\mathbf{s}}_t)\|^2\right] \\
&= E\left[\text{Trace}\left\{\mathbf{C}^\top(\mathbf{s}_t - \hat{\mathbf{s}}_t)(\mathbf{C}^\top(\mathbf{s}_t - \hat{\mathbf{s}}_t))^\top\right\}\right] \\
&= E\left[\text{Trace}\left\{\mathbf{C}^\top(\mathbf{s}_t - \hat{\mathbf{s}}_t)(\mathbf{s}_t - \hat{\mathbf{s}}_t)^\top\mathbf{C}\right\}\right] \\
&= E\left[\text{Trace}\left\{(\mathbf{s}_t - \hat{\mathbf{s}}_t)(\mathbf{s}_t - \hat{\mathbf{s}}_t)^\top\right\}\right] \\
&= E\left[\|\mathbf{s}_t - \hat{\mathbf{s}}_t\|^2\right] \\
&= MSE(\hat{\mathbf{s}}_t).
\end{aligned} \tag{5.11}$$

On the other hand, the SLSE used to recover $\mathbf{s}_t$ achieves the CRB. Thus, using (5.4), the $MSE(\hat{\mathbf{s}}_t)$ can be written as follows:

$$MSE(\hat{\mathbf{s}}_t) = \sigma_n^2 \text{Trace}\left\{\left(\mathbf{A}_\mathcal{I}^\top\mathbf{A}_\mathcal{I}\right)^\dagger\right\} \tag{5.12}$$

For example, by choosing $\mathbf{A}_\mathcal{I}$ as the identity matrix or as an independent and identically distributed Gaussian matrix such that $\mathbf{A}_\mathcal{I}^\top\mathbf{A}_\mathcal{I} = \mathbf{I}_K$, then the resulting MSE is equal to $\sigma_n^2 K$.

Considering the sensor readings as sparse is not really true. By doing so, we introduce a bias in the calculation of the MSE due to a mismatch of the system model.

Hence the MSE between the real probed signal and the recovered one can be expressed as follows:

$$MSE(\hat{\mathbf{x}}_t) = MSE(\hat{\tilde{\mathbf{x}}}_t) + E\left[\|\tilde{\mathbf{x}}_t - \mathbf{x}_t\|^2\right]. \tag{5.13}$$

To achieve the best recovery performance using PBCDG, one should find the best trade-off between $MSE(\hat{\tilde{\mathbf{x}}}_t)$ and $E\left[\|\tilde{\mathbf{x}}_t - \mathbf{x}_t\|^2\right]$ since these two terms are inversely proportional. In fact, by estimating only K position of the sparse representation, we reduce the effect of noise on the estimated signal. In return, small values of K may induce a large approximation of the original signal $\mathbf{x}_t$.

PBCDG is built around the spatial correlation among sensor readings. In the sequel, we present an extension of this solution to incorporate both the spatial and temporal dimensions.

5.4.1 Spatio-temporal Position-based Compressive Data Gathering

Due to the temporal correlation between each sensor readings, node measurements exhibit a temporal sparse representation in addition to the spatial one. To further reduce the number of samples required to reconstruct sensory readings, we propose to take advantage of the temporal sparsity feature in the data gathering process. In the sequel, we present our Spatio-Temporal Position-Based Compressive Data Gathering (ST-PBCDG) scheme.

Assume that the sink aims to collect the sensory data during a period of M time slots. Let $\mathbf{X} = [\mathbf{x}_1 \dots \mathbf{x}_M]$ denotes the measurement data matrix. Again, Let $\mathbf{C} \in \mathbb{R}^{N\times N}$ and $\mathbf{R} \in \mathbb{R}^{M\times M}$ denote, respectively, the column and the row sparsifiying operators. Assume that the data matrix admits the following spatio-temporal sparse representation:

$$\mathbf{S} = \mathbf{C}\mathbf{X}\mathbf{R}^\top, \tag{5.14}$$

where $\mathbf{S}$ is $N \times M$ sparse matrix. Let $\mathcal{I}$ and $\mathcal{J}$ denote, respectively, the index set of the nonzero columns and rows of $\mathbf{S}$. The data matrix $\mathbf{X}$ can be approximated using a submatrix $\tilde{\mathbf{S}} \in \mathbb{R}^{K_1\times K_2}$ of $\mathbf{S}$ obtained by removing the null columns and rows of $\mathbf{S}$. Hence, $\tilde{\mathbf{S}}$ can be expressed as follows:

$$\tilde{\mathbf{S}} = (\mathbf{C}^\top)^\top_{\mathcal{I}}\ \mathbf{X}\ (\mathbf{R}^\top)_{\mathcal{J}}, \tag{5.15}$$

where $()_{\mathcal{I}}$ (respectively $()_{\mathcal{J}}$) denote the operator that allows to create a submatrix that only contains the columns that correspond to the indices in $\mathcal{I}$ (respectively $\mathcal{J}$). The approximated version of $\mathbf{X}$ denoted by $\tilde{\mathbf{X}}$ can be expressed as follows:

$$\tilde{\mathbf{X}} = (\mathbf{C}^\top)_{\mathcal{I}}\ \tilde{\mathbf{S}}\ (\mathbf{R}^\top)^\top_{\mathcal{J}}. \tag{5.16}$$

To enable the sink generating the approximated data matrix $\tilde{\mathbf{X}}$, nodes need to collaborate in order to transmit a linear combination of the submatrix $\tilde{\mathbf{S}}$. We denote by $\mathbf{Y} \in \mathbb{R}^{K_1\times K_2}$ the matrix received by the sink:

$$\mathbf{Y} = \mathbf{A}\ (\mathbf{C}^\top)^\top_{\mathcal{I}}\ \mathbf{X}\ (\mathbf{R}^\top)_{\mathcal{J}} + \mathbf{N} = \mathbf{A}\ \tilde{\mathbf{S}} + \mathbf{N} \tag{5.17}$$

where $\mathbf{N} \sim \mathcal{N}(0, \sigma_n^2)$ is a white Gaussian noise and $\mathbf{A} \in \mathbb{R}^{K_1 \times K_1}$ is the sensing matrix.

The proposed data gathering scheme that allows to achieve the aforementioned goal operates as follows: At first, we store at each node i the temporal sparsifying operator $(\mathbf{R}^\top)^\top_{\mathcal{J}}$ that allows to only take the K_2 nonzero entries of the temporal sparse representation, in addition to the i^{th} column vector of $\mathbf{A}$ $(\mathbf{C}^\top)^\top_{\mathcal{I}}$ required to perform the spatial projection and the linear combination. Then, each node observes its environment during M time slots and generates the k_2 sparse temporal representation of its M-dimensional readings. At the end of the time window M, nodes apply K_2 times the standard PBCDG scheme to transmit the K_2 points available at each sensor.

To recover the probed data matrix, the sink has to estimate the submatrix $\tilde{\mathbf{S}}$ using the noisy received signal which can be written using the following vectorial formulation:

$$\mathbf{y}' = \mathbf{A}' \, \tilde{\mathbf{s}}' + \mathbf{n}', \tag{5.18}$$

where $\mathbf{y}' \in \mathbb{R}^{K_1 K_2}$, $\tilde{\mathbf{s}}' \in \mathbb{R}^{K_1 K_2}$ and $\mathbf{n}' \in \mathbb{R}^{K_1 K_2}$ are the vectors composed from the columns of $\mathbf{Y}$, $\tilde{\mathbf{S}}$ and $\mathbf{N}$, respectively. The matrix $\mathbf{A}' \in \mathbb{R}^{K_1 K_2 \times K_1 K_2}$ is a bloc diagonal matrix that contains $\mathbf{A}$ at each bloc. Thus, the two dimensional estimation problem is reduced to a tractable one dimensional estimation problem. Using the efficient SLSE, the estimated spatio-temporal sparse representation vector is equal to:

$$\hat{\tilde{\mathbf{s}}} = \mathbf{A'}^\top (\mathbf{A}' \mathbf{A'}^\top)^{-1} \mathbf{y}'. \tag{5.19}$$

After estimating $\hat{\tilde{\mathbf{s}}}$, the sink reshapes it into $K_1 \times K_2$ matrix $\hat{\tilde{\mathbf{S}}}$. Thereafter, the approximated data matrix can be computed using (5.16):

$$\hat{\tilde{\mathbf{X}}} = (\mathbf{C}^\top)_{\mathcal{I}} \, \hat{\tilde{\mathbf{S}}} \, (\mathbf{R}^\top)^\top_{\mathcal{J}}. \tag{5.20}$$

The resulting MSE of the estimation problem can be expressed as follows:

$$
\begin{aligned}
MSE(\hat{\tilde{\mathbf{X}}}_t) &= E\left[\|\tilde{\mathbf{X}} - \hat{\tilde{\mathbf{X}}}\|_F^2\right] \\
&= E\left[\|(\mathbf{C}^\top)_{\mathcal{I}}\ \tilde{\mathbf{S}}\ (\mathbf{R}^\top)_{\mathcal{J}}^\top - (\mathbf{C}^\top)_{\mathcal{I}}\ \hat{\tilde{\mathbf{S}}}\ (\mathbf{R}^\top)_{\mathcal{J}}^\top\|_F^2\right] \\
&= E\left[\|(\mathbf{C}^\top)_{\mathcal{I}}\ (\tilde{\mathbf{S}} - \hat{\tilde{\mathbf{S}}})\ (\mathbf{R}^\top)_{\mathcal{J}}^\top\|_F^2\right] \\
&= E\left[\text{Trace}\left\{(\mathbf{C}^\top)_{\mathcal{I}}(\tilde{\mathbf{S}} - \hat{\tilde{\mathbf{S}}})(\mathbf{R}^\top)_{\mathcal{J}}^\top(\mathbf{R}^\top)_{\mathcal{J}}(\tilde{\mathbf{S}} - \hat{\tilde{\mathbf{S}}})^\top\ (\mathbf{C}^\top)_{\mathcal{I}}^\top\right\}\right] \\
&= E\left[\text{Trace}\left\{(\mathbf{C}^\top)_{\mathcal{I}}(\tilde{\mathbf{S}} - \hat{\tilde{\mathbf{S}}})(\tilde{\mathbf{S}} - \hat{\tilde{\mathbf{S}}})^\top\ (\mathbf{C}^\top)_{\mathcal{I}}^\top\right\}\right] \\
&= E\left[\text{Trace}\left\{(\tilde{\mathbf{S}} - \hat{\tilde{\mathbf{S}}})(\tilde{\mathbf{S}} - \hat{\tilde{\mathbf{S}}})^\top\right\}\right] \\
&= E\left[\|\tilde{\mathbf{S}} - \hat{\tilde{\mathbf{S}}}\|_F^2\right] \\
&= E\left[\|\tilde{\mathbf{s}} - \hat{\tilde{\mathbf{s}}}\|^2\right].
\end{aligned}
\tag{5.21}
$$

Since the SLSE used to compute $\hat{\tilde{\mathbf{s}}}$ is an efficient estimator and by taking into account the residual error resulting from approximating the data matrix $\mathbf{X}$ using only K_1K_2 coefficients. The MSE of estimating the original probed data can be expressed as follows:

$$
MSE(\hat{\mathbf{X}}) = \sigma_n^2 \text{Trace}\left\{\left(\mathbf{A'}^\top \mathbf{A'}\right)^\dagger\right\} + E\left[\|\tilde{\mathbf{X}} - \mathbf{X}\|^2\right], \tag{5.22}
$$

where the first term is due to the noisy perturbation affecting the data and it is computed using equation (5.4), while the second term represents the error of representing the original data matrix using only K_1K_2 coefficients. If the sensing matrix $\mathbf{A}$ satisfies $\mathbf{A}^\top\mathbf{A} = \mathbf{I}_{k_1}$, then the resulting MSE can be expressed as follows:

$$
MSE(\hat{\mathbf{X}}) = \sigma_n^2 k_1 k_2 + E\left[\|\tilde{\mathbf{X}} - \mathbf{X}\|^2\right]. \tag{5.23}
$$

Hence, The resulting MSE is a sum of two inversely proportional terms where the first one depends on the quality of the transmissions whereas the second one is related to the sparse model error.

PBCDG and ST-PBCDG allow to efficiently gather the sensory measurements based on the *a priori* knowledge about the support set of the sparse data representation. However, the sparsity feature is sensitive to the presence of anomalies. In the sequel we extend our proposed solutions to deal with the presence of anomalies while maintaining a low complexity recovering algorithm.

5.5 SPIKE DETECTION USING PBCDG AND ST-PBCDG

The main advantage of PBCDG approach compared to CS-based data gathering techniques comes from its low complexity recovering algorithm and its ability to efficiently deal with noisy environments. Yet, our proposed model do not provide an efficient tool to detect the potential presence of anomalies since the sparsity structure of sensory data is vulnerable to aberrant values.
In this section, we provide an efficient approach to detect spikes which describe abrupt raises or falls in the probed measurements. At the difference of typical CS-based anomaly detection approaches, our PBCDG solution allows to identify anomalies without decoding the received signal, and this by analyzing the power variation on the coded sensory data over time. In the sequel, we prove that the presence of spikes in sensory data results in a jump in the received signal power.

Again, let $\mathbf{x}_t$ denotes the vector of N sensory readings without spikes and let $\check{\mathbf{x}}_t$ the clone of $\mathbf{x}_t$ except at the position i where a spike of value Δ occurs, *i.e.* $\check{\mathbf{x}}_t = [x_{1,t}, \ldots, x_{i-1,t}, x_{i,t} + \Delta, x_{i+1,t}, \ldots, x_{N,t}]$. Let, $\mathbf{y}_t = \mathbf{AC}\mathbf{x}_t + \mathbf{n}$ and $\check{\mathbf{y}}_t = \mathbf{AC}\check{\mathbf{x}}_t + \mathbf{n}$ denote the compressed received signals that correspond to $\mathbf{x}$ and $\check{\mathbf{x}}_t$, respectively. The power of the free-spike compressed signal can be expressed as follows:

$$P_{\mathbf{y}_t} = E\{\|\mathbf{AC}\mathbf{x}_t + \mathbf{n}\|^2\} = \|\mathbf{AC}\mathbf{x}_t\|^2 + K\sigma_n^2. \tag{5.24}$$

On the other hand, the power of the received signal in the presence of a data spike can be expressed as follows

$$\begin{aligned} P_{\check{\mathbf{y}}_t} &= E\{\|\mathbf{AC}\check{\mathbf{x}}_t + \mathbf{n}\|^2\} \\ &= E\{\|\mathbf{AC}\mathbf{x}_t + \Delta(\mathbf{AC})_i + \mathbf{n}\|^2\} \\ &= \|\mathbf{AC}\mathbf{x}_t + \Delta(\mathbf{AC})_i\|^2 + K\sigma_n^2, \end{aligned} \tag{5.25}$$

where $(\mathbf{AC})_i$ denotes the i^{th} column of the matrix $\mathbf{AC}$. Thus, the difference of power between the spiky and the original information can be bounded as follows:

$$\begin{aligned} |P_{\check{\mathbf{y}}_t} - P_{\mathbf{y}_t}| &= |\|\mathbf{AC}\mathbf{x}_t + \Delta(\mathbf{AC})_i\|^2 - \|\mathbf{AC}\mathbf{x}_t\|^2| \\ &\leqslant \Delta^2 \|(\mathbf{AC})_i\|^2 \end{aligned} \tag{5.26}$$

Thus, the power variation due to the presence of spikes is independent of the noisy nature of a WSN environment and it is always bounded by the square magnitude of the spike times a certain constant that depends on the system model. Starting from this observation, we conceive our spike detection solution. In the absence of anomalies, we assume that the power variation between two successive coded vectors can be bounded using a certain threshold C that can be learned using a training data set. In order to detect the presence of spikes at a specific time slot t, it is sufficient to compare the

power variation $|P_{\mathbf{y}_t} - P_{\mathbf{y}_{t-1}}|$ to the threshold C. Note that we assume that the sink disposes of a free-spikes data at the beginning of the anomaly detection procedure to compute the adequate value of C. If the considered threshold is respected, then we can recover the coded data using our solution. Otherwise, the sink declares the presence of anomalous data.

To compute the value and the position of anomalies, we can rely on the ℓ_1-minimization algorithm for overcomplete system [Candes and Tao, 2006]. In fact, the spotted signal can be decomposed into the sum of the normal signal and another vector that represents the anomalies as follows:

$$\check{\mathbf{x}}_t = \mathbf{x}_t + \mathbf{e}_t, \tag{5.27}$$

where $\mathbf{e}_t$ is the vector representing the occurring spikes. We suppose that abnormal readings appear in a small set of positions and hence the vector $\mathbf{e}_t$ is supposed to be sparse. The vector of spotted readings $\check{\mathbf{x}}_t$ can be expressed as follows:

$$\check{\mathbf{x}}_t = \mathbf{C}^\top \mathbf{s}_t + \mathbf{I}_N \mathbf{e}_t, \tag{5.28}$$

where $\mathbf{I}_N$ is the identity matrix. Thus, the spotted signal can be decomposed into two signals which are sparse in different domains. By constructing an overcomplete basis $\mathbf{C}' = \left[\mathbf{C}^\top \mathbf{I}_N\right]$, then the spiky signal have a sparse representation in the new domain:

$$\check{\mathbf{x}}_t = \mathbf{C}'\mathbf{d}_t, \quad \mathbf{d}_t = \left[\mathbf{s}_t^\top \mathbf{e}_t^\top\right]^\top, \tag{5.29}$$

Using the previous notation, the noisy compressed information received by the sink can be expressed as follows:

$$\check{\mathbf{y}}_t = \mathbf{A}\mathbf{C}'\mathbf{d}_t + \mathbf{n}_t, \tag{5.30}$$

In order to recover the vectors of normal readings and the potential anomalies, the sink have to estimate $\mathbf{d}_t$ by solving the following optimization problem:

$$\underset{\mathbf{d}_t}{\text{minimize}} \|\check{\mathbf{y}}_t - \mathbf{A}\mathbf{C}'\mathbf{d}_t\|^2 + \lambda \|\mathbf{e}_t\|_1, \tag{5.31}$$

The optimum solution to the previous optimization problem $\hat{\mathbf{d}}_t$ can be found by using standard ℓ_1-optimization algorithm [Donoho et al., 2012; Candes and Tao, 2006; Donoho et al., 2006]. Hence, the original readings are obtained by projecting the first N entries of $\hat{\mathbf{d}}_t$ into the spatial sparsifying domain, while the last N elements of $\hat{\mathbf{d}}_t$ represent the vector of anomalies.

Our anomaly detection solution can be extended to detect spiky values while using ST-PBCDG. The first step of the extended solution consists on removing the compression on the temporal component by projecting the received signal on the temporal sparsifying basis. Let $\mathbf{Y}' \in \mathbb{R}^{K_1 \times M}$ denotes the resulting matrix:

$$\mathbf{Y}' = \mathbf{A}\ (\mathbf{C}^\top)^\top_{\mathcal{I}}\ \mathbf{X} + \mathbf{N}\ (\mathbf{R}^\top)^\top_{\mathcal{J}}. \tag{5.32}$$

After that, we apply the power variation test on each column of $\mathbf{Y'}$ separately. To theoretically validate our approach, we suppose that a spike Δ occurs on the i^{th} element of the j^{th} column of $\mathbf{X}$. Let $\mathbf{y}'_j$ and $\check{\mathbf{y}}'_j$ denote the j^{th} column of $\mathbf{Y'}$ and $\check{\mathbf{Y}}'$, respectively, where the matrix $\check{\mathbf{Y}}'$ is the analogous of $\mathbf{Y'}$ in the presence of spiky data. The power of the free-spike signal can be expressed as follows:

$$\begin{aligned} P_{\mathbf{y}_j} &= E\{\|\mathbf{A}\ (\mathbf{C}^\top)^\top_{\mathcal{I}}\ \mathbf{x}_j + (\mathbf{N}\ (\mathbf{R}^\top)^\top_{\mathcal{J}})_j\|^2\} \\ &= \|\mathbf{A}\ (\mathbf{C}^\top)^\top_{\mathcal{I}}\ \mathbf{x}_j\|^2 + E\{\|(\mathbf{N}\ (\mathbf{R}^\top)^\top_{\mathcal{J}})_j\|^2\}, \end{aligned} \tag{5.33}$$

where the last equality follows from the decorrelation between the resulting noise and the reference signal. In the presence of spike, the power of $\check{\mathbf{y}}'_j$ is equal to:

$$\begin{aligned} P_{\check{\mathbf{y}}_j} &= E\{\|\mathbf{A}\ (\mathbf{C}^\top)^\top_{\mathcal{I}}\ \check{\mathbf{x}}_j + (\mathbf{N}\ (\mathbf{R}^\top)^\top_{\mathcal{J}})_j\|^2\} \\ &= E\{\|\mathbf{A}\ (\mathbf{C}^\top)^\top_{\mathcal{I}}\ \mathbf{x}_j + \Delta(\mathbf{A}\ (\mathbf{C}^\top)^\top_{\mathcal{I}})_i + (\mathbf{N}\ (\mathbf{R}^\top)^\top_{\mathcal{J}})_j\|^2\} \\ &= \|\mathbf{A}\ (\mathbf{C}^\top)^\top_{\mathcal{I}}\ \mathbf{x}_j + \Delta(\mathbf{A}\ (\mathbf{C}^\top)^\top_{\mathcal{I}})_i\|^2 + E\{\|(\mathbf{N}\ (\mathbf{R}^\top)^\top_{\mathcal{J}})_j\|^2\}. \end{aligned} \tag{5.34}$$

Hence, we can bound the power variation between the spiky and non-spiky data as follows:

$$\begin{aligned} |P_{\check{\mathbf{y}}_j} - P_{\mathbf{y}_j}| &= \|\mathbf{A}\ (\mathbf{C}^\top)^\top_{\mathcal{I}}\ \mathbf{x}_j + \Delta(\mathbf{A}\ (\mathbf{C}^\top)^\top_{\mathcal{I}})_i\|^2 - \|\mathbf{A}\ (\mathbf{C}^\top)^\top_{\mathcal{I}}\ \mathbf{x}_j\|^2 \\ &\leqslant \Delta^2\|(\mathbf{A}\ (\mathbf{C}^\top)^\top_{\mathcal{I}})_i\|^2. \end{aligned} \tag{5.35}$$

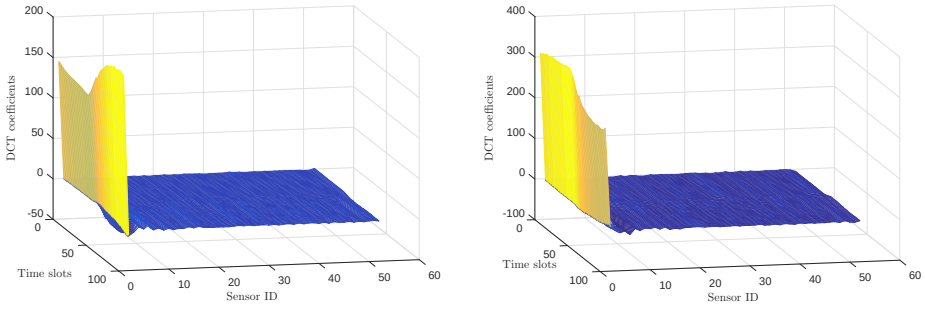
Consequently, The power variation using ST-PBCDG is bounded by some constant times Δ^2 which allows us to adapt the same methodology to detect spikes as in the one dimensional PBCDG case.

5.6 Performance evaluation

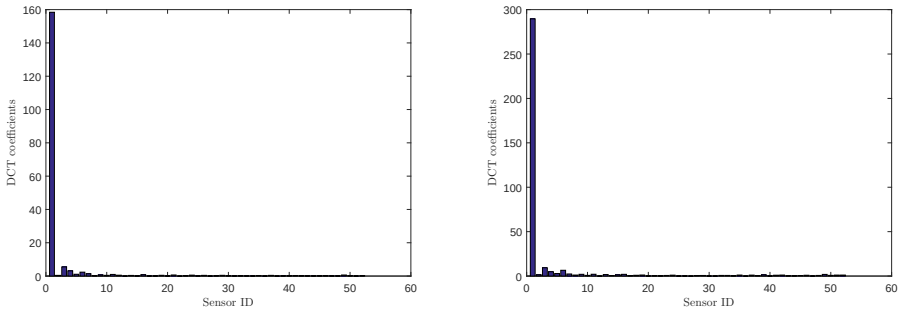
In this section we propose to evaluate the performance of the two versions of our PBCDG solution. At first, we describe our simulation setting. Then we evaluate the data recovery performance of our solutions. Finally, we present the performance of PBCDG solution while dealing with spiky data.

5.6.1 Simulation setting

To evaluate the performance of our PBCDG approach, we consider the same humidity and temperature datasets provided by Intel Berkeley research lab as in the previous chapters. As a reminder, these datasets represent the values probed by 52 sensors during 300 successive time slots. Our PBCDG solution requires the data to have a sparse representation in some transform domain. We propose to study the sparsity pattern



The spatial DCT representation.

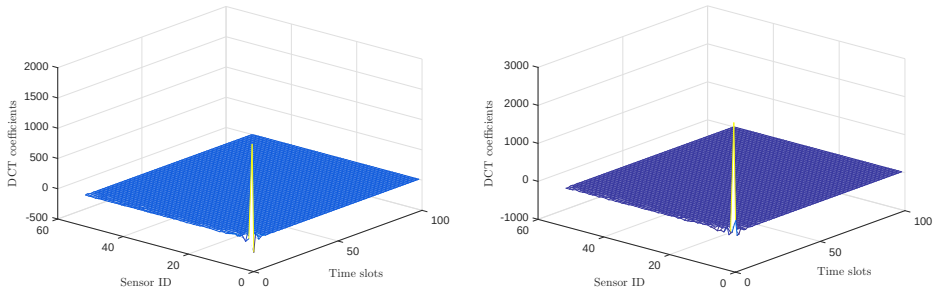


The dominant positions in the spatial DCT representation.

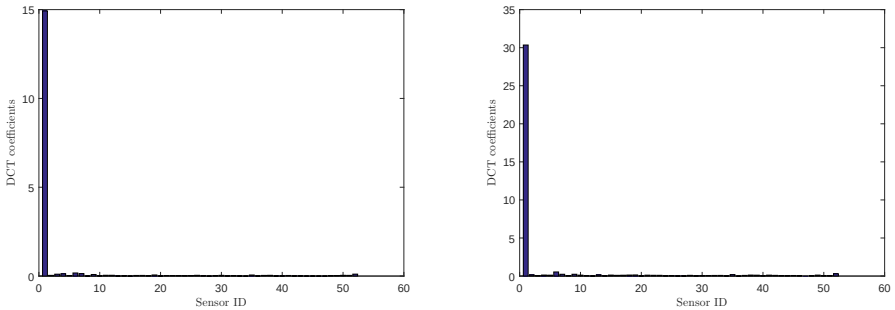
Figure 5.2: *The spatial sparsity feature in: Left: Intel temperature data. Right: Intel humidity data.*

under the DCT domain. Thus, the sparsifying operators $\mathbf{C}$ and $\mathbf{R}$ are represented by the DCT operator. To highlight the spatial sparsity structure of the considered sensory readings, we plot in Figure 5.2 the DCT representation of the spatial readings during 100 times slots. We can easily distinguish that the first frequencies monopolize most of the energy of the transformed vector. However, the order of the most dominant positions differs slightly from the readings' order in the spatial sensory vector. Thence, we arrange the support set of the sparse representation according to the magnitude of each position. This allows us to ensure that for each sampling rate, the most important positions are transmitted at first.

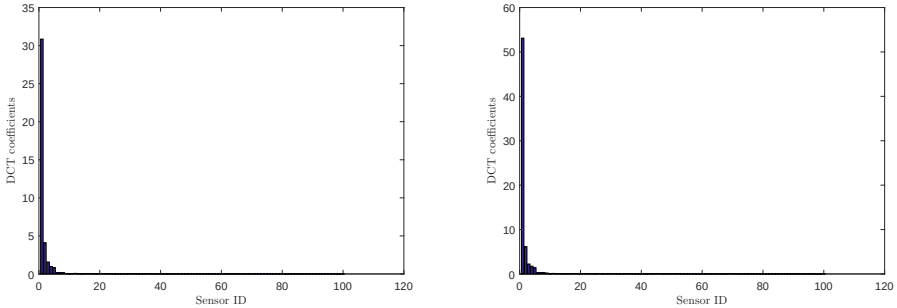
Our ST-PBCDG requires the treated data to have a spatio-temporal sparse representation. This property is well satisfied by our treated data as depicted in Figure 5.3 where we plot the spatio-temporal sparse representation of a data matrix gathered during 100 time slots. The support set of the spatial component driven form the spatio-temporal sparse representation is similar to the previous case. Therefore, the dominant position are arranged according to their order of magnitude in a decreas-



The spatio-temporal DCT representation.



The dominant positions in the spatial component of the spatio-temporal DCT representation.



The dominant positions in the temporal component of the spatio-temporal DCT representation.

Figure 5.3: *The spatio-temporal sparsity feature in: Left: Intel temperature data. Right: Intel humidity data.*

ing order to form the spatial support set. Whereas, The dominant positions of the temporal component of the spatio-temporal representation are almost arranged in a decreasing order in their raw form. Thus, according to the temporal sampling rates,

nodes have always to transmit the first positions. Finally, we set the sensing matrix as the identity matrix. Note that, we can use instead Gaussian matrices when the dimension of the data matrix is large enough.

5.6.2 Recovery performance

To evaluate the recovery performance of the two version of our PBCDG solution, We consider a set of sampling rates ranging from 5% to 90% of the amount of sensory readings. At first, we simulate a noisy free environment. Thus, our setting is reduced to the system setting described in the first chapter which allows us to compare our data recovering approach to the STCS solution described previously. Furthermore, we propose to compare our solution with the reference solution in CS-data gathering schemes, namely, the Spatial Compressive Sensing solution [Candes et al., 2006; Chong et al., 2009]. Again, we use the NMAE metric to measure the recovery accuracy in the absence of noise.
Note that the sampling rate used in ST-PBCDG hides two individual sampling rates that represent the amount of compression on the temporal dimension as well as on the spatial dimension. In our simulations, we equally set these compression rates to get the global sampling rate. Indeed, optimizing the amount of compression on each dimension to get the most representative information is beyond the scope of this work.

The recovery performance for the noisy free environment is depicted in Figure 5.4. For the temperature dataset, both versions of our method reach a stable NMAE of the order of 3%. The improvement of our solutions is noticeable in low sampling rates. Even by including the temporal component in the recovery process using STCS, the *a prior* knowledge about the support set of the sparse representation allows to better reconstruct the sensory reading. For example, using only 5% of the sparse representation entries, we can estimate the data matrix with a precision nearly equals to 97%, while the STCS method allows a reconstruction precision of the order of 89% for the same number of samples. As the sampling rate increases, all considered methods reach a good recovery performance.

The PBCDG reaches similar performance on the humidity dataset. However, in high sampling rates, the performance of the spatio-temporal version of our solution achieves higher NMAE than other considered algorithm because of the lack of some details while applying the compression on the two dimensions and because the assumption that the temporal sparse representation is arranged in a decreasing order which is not perfectly true.

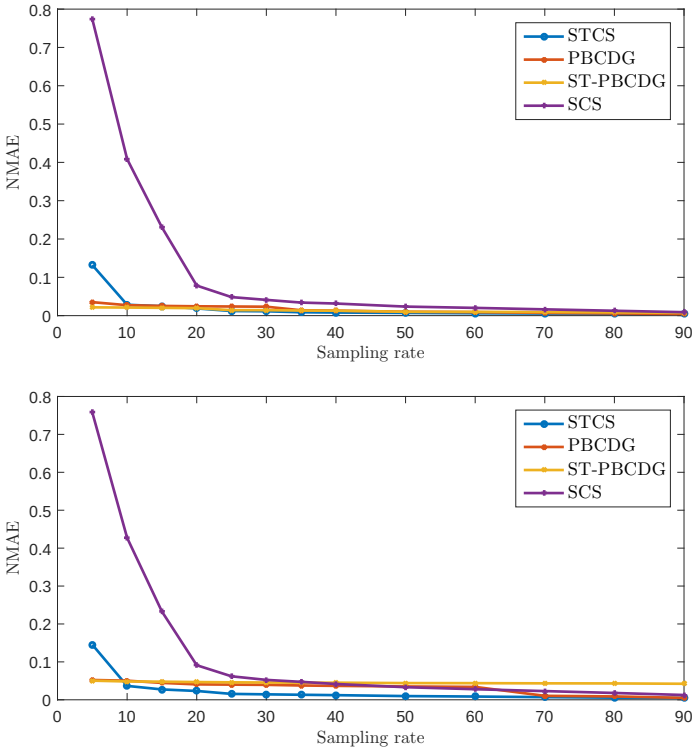


Figure 5.4: *Recovery performance in a noise free environment. Top: on Intel temperature dataset. Bottom: on Intel humidity dataset.*

The main advantage of our solution is its higher ability to deal with noisy environments. To evaluate the improvement brought up by our approach, we simulate the data gathering process under different noise level. We use the relative Signal to Noise Ratio (SNR) metric defined as follows:

$$\text{SNR} = \frac{\sigma_s^2}{\sigma_n^2}, \tag{5.36}$$

where $\sigma_s^2 = E\{\|\mathbf{x}_t - E(\mathbf{x}_t)\|^2\}$. We propose to compare the two versions of our solution with the SCS algorithm. We do not consider STCS solution since it does not cover the noisy setting. Due to the considered metric used in the theoretical evaluation of our solution, we consider the per sample MSE to analyse the estimation performance of our approach.

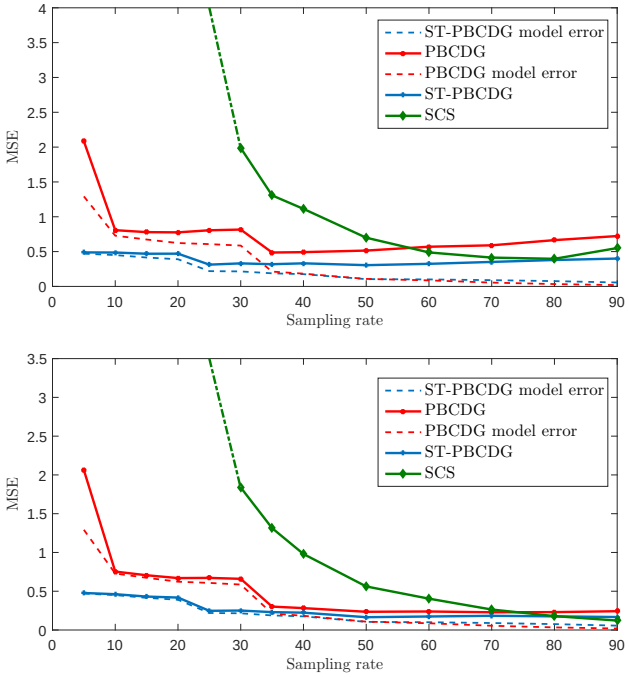


Figure 5.5: *recovery performance on Intel temperature data for a noisy setting with. Top; SNR=5 db. Bottom:20 db.*

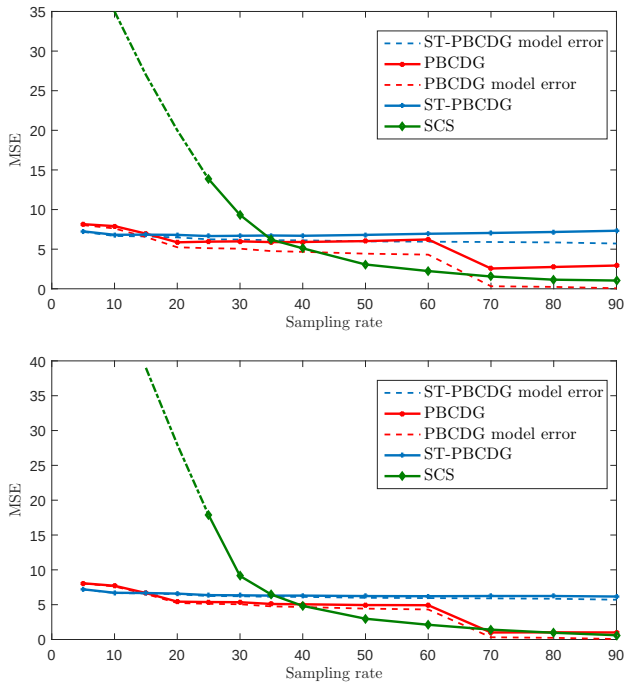


Figure 5.6: *recovery performance on Intel humidity data for a noisy setting with. Top; SNR=5 db. Bottom:20 db.*

Figure 5.5 illustrates the recovery performance on the Intel temperature data under different SNR levels. The dotted plots represent the theoretical MSE limits due to the model mismatch. In fact, these plots correspond to the resulting MSE when considering only few samples from the sparse representation in the absence of noise. By taking into account the noisy characteristic of WSN environments, an additional estimation error is added to the system model error. We can easily spot that the additional error is a linear function of the noise variance which is in concordance with the theoretical results established previously. The two versions of our data gathering solution outperform the SCS technique, especially when the sampling rate is less than 50%.

The recovery performance on the humidity data set is depicted in Figure 5.6. The two variants of our proposed solution have a similar behaviour as in the temperature dataset case and again outperform the SCS technique, especially when the sampling rate is very low. When the sampling rate is less than 15%, the ST-PBCDG solution outperforms the other methods, then it keeps a constant MSE. Whereas the performance of the BPCDG is an increasing function of the sampling rate and achieves good recovery quality for both low and high available coefficients of the sparse representation.

5.6.3 spike detection performance

Our method of detecting spikes is based on the observation that the power of the received signal slightly varies between two successive time slots in the absence of abnormal readings. This assumption is well satisfied by our treated datasets as depicted in Figure 5.7 where we simulated different spikes values in different positions. We can easily spot that the presence of spikes induces a large variation on the resulting per sample power variation. To evaluate the spike detection performance of our proposed

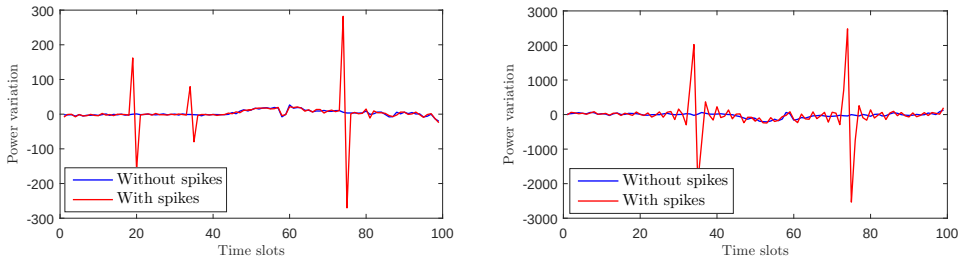


Figure 5.7: *The effect of spikes on the power variation between two successive time slots. Left: on the temperature dataset using PBCDG and an SNR=20 db. Right: on the humidity dataset using ST-PBCDG and an SNR=20 db.*

approach, we use a training dataset to learn the maximum possible power variation in the absence of spikes. Then, a random set of spikes is generated and added to the tested dataset. In our simulations, we define a spike as a value that exceeds three times the mean of the sensory data. Again, we use the detection accuracy and the

false alarm metrics to evaluate the performance of our approach.

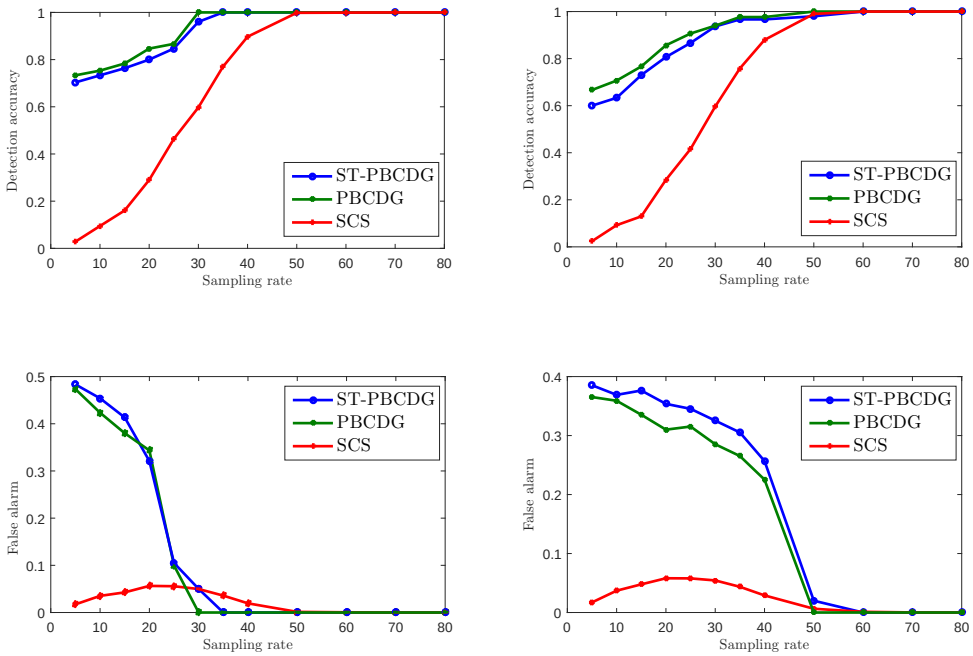


Figure 5.8: *Spike detection performance. Left: on the temperature dataset using PBCDG and an SNR=10 db. Right: on the humidity dataset using ST-PBCDG and an SNR=20 db.*

The spike detection performance of the two variants of our approach under different noise levels is depicted in Figure 5.8. Both versions of our solution achieve high detection accuracy and low false alarm rates when the sampling rate exceeds 40%. For low sampling rates, we can identify the spiky data better than the SCS technique. When the available number of samples is relatively low, the SCS algorithm generates a false estimation of the sensory data and nearly zero vector that corresponds to the estimated anomalies. Thus, it achieves a low false alarm rate and a vanishing detection accuracy. On the other hand, and under the same conditions, the two versions of our solution succeed to detect more than the half of the spiky sequences, but with high false alarm rates.

5.7 CONCLUSION

In this chapter, we presented a new data gathering scheme for noisy WSN environments that allows an efficient data reconstruction based on the *a priori* knowledge about the support set of the sensory sparse representation. The proposed solution is extended to take advantage on the spatio-temporal correlation between nodes. Furthermore, we

proposed a spike detection scheme that allows to detect the abrupt changes on the sensory readings by observing the power variation of the compressed sequences. We evaluated the theoretical recovery performance of the two versions of our solution.
In addition to its ability to deal with noisy data, our proposed approach allows to recover the sensory data using low complexity operations, which make it suitable for large WSNs where the amount of generated data could not be hold using our MC or SC-based approaches.

Conclusion

In this thesis, we treated the problem of data gathering and anomaly detection in WSNs. In the literature, these two issues were often treated separately in spite of their appearing complementarity. We aimed in this work to jointly address the problem of collecting data and detecting anomalies while reducing the use of WSNs' resources and enhancing their lifespan.

Throughout our contributions, we proposed different approaches that allow to perform both tasks. We started by proposing a CS-based technique. In fact, this class of techniques foster the network lifespan thanks to their ability to balance the network traffic among nodes. In order to enhance the data recovery performance compared to existing CS-based solutions, we proposed to include the spatio-temporal correlation between sensors' readings in the data recovering algorithm. Hence, each received observation is decoded by taking into account the previous estimations. Moreover, we proposed a novel formulation to decorticate anomalies from the normal data structure using the class of primal-dual algorithms. This approach was validated by running extensive simulations on real datasets. To sum up, our main conclusion here was that this first contribution provides an efficient tool to collect sensory readings from all sensors as well as to detect the occurrence of outlying values everywhere in the observed environment. Thus, it is well adapted for the applications that privilege the fault detection over the data gathering functionality. Otherwise, when optimizing the WSN resources while collecting the sensory data is the first priority of the deployed system, we can use instead MC-based methods since they achieve better lifespan for the same sampling rate [Cheng et al., 2013].

In our second and third contributions, we proposed two MC-based solutions where only a subset of nodes participates at the data gathering process at each time slot. In the second contribution, we used the low rank structure of the sensory data as well as the sparsity pattern to enable the sink estimating missing values. We extended this approach to detect the probable anomalies. The proposed algorithm demonstrates a good data recovery and anomaly detection performance. In the aim of more refining the estimation quality and facilitating the interpretation of the involved parameters in the deployed algorithm, we proposed in the third contribution a constrained formulation of the data gathering and anomaly detection problem. Hence, the *a priori*

knowledge about the target data were modeled through convex constraints which are related to some physical proprieties of the sensory data. We applied this new formulation to model some data proprieties such as the sparsity and the low rank patterns. The simulations carried on real data sets demonstrate the efficiency of our proposed approach in terms of data recovering quality and anomaly detection capacities.
Even the previous techniques allow to ensure a reliable data gathering and anomaly detection solutions, they require to manipulate a complex data structures and do not offer a proper tool to deal with the noisy nature of WSNs' environments.

In our last contribution, we proposed a different approach that allows to deal with the noisy nature of WSNs' environments. This solution do not require a huge computation capacities in the collector node and achieves the optimum recovery performance given the assumption that sensory data are perfectly sparse. This was made possible by sending only the relevant positions in the sensory sparse representation and by applying an efficient estimator at the sink. Our solution was extended to integrate both the temporal and spatial sparsity pattern. Furthermore, we described a distributed strategy to compute and transmit the most important positions in the sensory sparse representation which allows to equilibrate the traffic carried by the intermediate nodes and hence to improve the lifespan of the WSN. Finally, we presented a low complexity solution to detect the spiky data and the potential changes in the sensory data. The efficiency of our solution was proved by theoretical analysis as well as extensive simulations on real datasets.

To summarize, we proposed throughout our contributions different answers to the questions posed at the beginning of this thesis, namely: How to efficient gather the sensory data? How to reliably identify the data anomalies? And how to achieve this task with an optimal use of WSN resources? Depending on the system requirements, one can opt for our CS-based solution if the anomaly detection is privileged over the data gathering Functionality. Otherwise, we can use our MC-based solutions for a better use of WSN resources. When the size of WSNs goes high, resulting on large datasets, we can rely on our low complexity PBCDG solution which offers the best trade-off between complexity and performance.

PERSPECTIVES

The solutions proposed throughout this work allow the emergence of some insights that may further enhance the performance of WSNs. In all our contributions, we assumed that the sensory readings are univariate. On the other hand, while observing the simulation results on the humidity and temperature datasets, we can easily spot some correlations. Starting from this remark, we can guess that dealing with multivariate data can further enhance the recovering performance. Indeed, we can jointly treat the sensors that probe some correlated phenomena and reformulate the problem of data gathering and anomaly detection in a way that it incorporates the multivariate data structure. This technique will probably increase the ability of our system to manage

outlying readings since we will dispose of a multitude of correlated values probed at the same time and in the same positions. This extra side information will enable us to better dissemble anomalies form the multivariate data structure unless the outlying values are affecting all the probed magnitudes at the same time.

Another promising prospect is given by investigating the functionality of WSNs. In fact, all along this thesis, the sensor nodes were deprived of any decisive power and only the sink have the ability to detect anomalies and reconstruct data. We can potentially improve the WSN performance by attributing some freedom degree to local sensors, so they can decide whether their readings are erroneous or not. For example, we can allow sensors to communicate with their neighboring nodes to assign a reliability coefficient to their probed data, which would be used by the sink to further optimize the data gathering and anomaly detection performance. Having in mind the energy efficiency and the low computational as well as low memory sensor nodes while performing this, render such a target a very challenging research issue.

Finally, minutely including the wireless transmission conditions in the data collecting process is a demanding perspective of this thesis. In fact, we only considered the white noise that could affect the transmission in WSNs. We can propose a general model that matches the wireless channel proprieties, the environment topology and the sensor deployment strategy. The generation of such a model is feasible since we consider static sensors. We can utilize this model in many ways to enable a better estimation of the sensory signals. For example, this model can be used to select the subset of nodes that have a better transmission quality to forward their readings toward the sink in MC-based solutions, rather than relying on random sampling which may leads to the selection of a subset of nodes that have a low channel quality, and hence a low quality of the transmitted data.

List of Figures

Bibliography

Jugoslava Acimovic, Baltasar Beferull-Lozano, and Razvan Cristescu. Adaptive distributed algorithms for power-efficient data gathering in sensor networks. In *Wireless Networks, Communications and Mobile Computing, 2005 International Conference on*, volume 2, pages 946–951. IEEE, 2005. 10

Manya V Afonso, José M Bioucas-Dias, and Mário AT Figueiredo. An augmented lagrangian approach to the constrained optimization formulation of imaging inverse problems. *IEEE Transactions on Image Processing*, 20(3):681–695, 2011. 21

Bakhtiar Qutub Ali, Niki Pissinou, and Kia Makki. Approximate replication of data using adaptive filters in wireless sensor networks. In *Wireless Pervasive Computing, 2008. ISWPC 2008. 3rd International Symposium on*, pages 365–369. IEEE, 2008. 11

Behtash Babadi, Nicholas Kalouptsidis, and Vahid Tarokh. Asymptotic achievability of the cramér–rao bound for noisy compressive sampling. *IEEE Transactions on Signal Processing*, 57(3):1233–1236, 2009. 81

Vic Barnett and Toby Lewis. *Outliers in statistical data*. Wiley, 1974. 13

G. Bell. A time and a place for standards. *Queue*, 2:66 –74, 2004. 7

Luís MA Bettencourt, Aric A Hagberg, and Levi B Larkey. Separating the wheat from the chaff: Practical anomaly detection schemes in ecological applications of distributed sensor networks. In *International Conference on Distributed Computing in Sensor Systems*, pages 223–239. Springer, 2007. 14

Stephen Boyd, Neal Parikh, Eric Chu, Borja Peleato, and Jonathan Eckstein. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends® in Machine Learning*, 3(1):1–122, 2011. 21

J.-F Cai, E. J. Candès, and Z. Shen. A Singular value Thresholding Algorithm for Matrix Completion. *SIAM Journal on Optimization*, 20(4):1956–1982, March 2010. ISSN 1052-6234. 34, 51

E. J. Candès and Benjamin Recht. Exact Matrix Completion via Convex Optimization. *Foundations of Computational Mathematics*, 9(6):717–772, December 2009. 49

E. J. Candès and J. Romberg. Practical Signal Recovery from Random Projections. *Wavelet Applications in Signal and Image Processing XI, Proc. SPIE Conf.*, 5914, 2005. 29

E. J. Candes and T. Tao. Near-optimal signal recovery from random projections: Universal encoding strategies? *IEEE Trans. Inf. Theor.*, 52(12):5406–5425, December 2006. ISSN 0018-9448. 28, 79, 80, 88

Emmanuel Candes and Terence Tao. The dantzig selector: Statistical estimation when p is much larger than n. *The Annals of Statistics*, pages 2313–2351, 2007. 80

Emmanuel J. Candès and Terence Tao. Decoding by linear programming. *IEEE Trans. Information Theory*, 51(12):4203–4215, 2005. 29

Emmanuel J. Candes, Justin K. Romberg, and Terence Tao. Stable signal recovery from incomplete and inaccurate measurements. *Communications on Pure and Applied Mathematics*, 59(8):1207–1223, 2006. 29, 32, 92

J. Candès and M. B. Wakin. An introduction to Compressive Sampling – A sensing/sampling paradigm that goes against ... 25:21–30, Mar. 2008. 26, 32

Cecilia Carbonelli, Satish Vedantam, and Urbashi Mitra. Sparse channel estimation with zero tap detection. *IEEE Transactions on Wireless Communications*, 6(5), 2007. 81, 83

Varun Chandola, Arindam Banerjee, and Vipin Kumar. Outlier detection: A survey. *ACM Computing Surveys*, 2007. 14, 15

Varun Chandola, Arindam Banerjee, and Vipin Kumar. Anomaly detection: A survey. *ACM Comput. Surv.*, 41(3):15:1–15:58, July 2009. ISSN 0360-0300. doi: 10.1145/1541880.1541882. 27

Vasilis Chatzigiannakis, Symeon Papavassiliou, Mary Grammatikou, and B Maglaris. Hierarchical anomaly detection in distributed large-scale sensor networks. In *Computers and Communications, 2006. ISCC'06. Proceedings. 11th IEEE Symposium on*, pages 761–767. IEEE, 2006. 16

J. Cheng, Q. Ye, H. Jiang, D. Wang, and C. Wang. STCDG: An efficient data gathering algorithm based on matrix completion for wireless sensor networks. *IEEE Transactions on Wireless Communications*, 12(2):850–861, 2013. 13, 47, 50, 72, 99

G. Chierchia, N. Pustelnik, B. Pesquet-Popescu, and J-C. Pesquet. A nonlocal structure tensor-based approach for multicomponent image recovery problems. *Image Processing, IEEE Transactions on*, 23(12):5531–5544, 2014. 70

L. Chong, W. Feng, S. Jun, and W. C. Chang. Compressive Data Gathering for Large-Scale Wireless Sensor Networks. In *ACM Mobicom*, pages 145–156. Association for Computing Machinery, Sep. 2009. 12, 26, 27, 36, 92

Jim Chou, Dragan Petrovic, and Kannan Ramachandran. A distributed and adaptive signal processing approach to reducing energy consumption in sensor networks. In *INFOCOM 2003. Twenty-Second Annual Joint Conference of the IEEE Computer and Communications. IEEE Societies*, volume 2, pages 1054–1062. IEEE, 2003. 10

David Chu, Amol Deshpande, Joseph M Hellerstein, and Wei Hong. Approximate data collection in sensor networks using probabilistic models. In *Data Engineering, 2006. ICDE'06. Proceedings of the 22nd International Conference on*, pages 48–48. IEEE, 2006. 11

Alexandre Ciancio, Sundeep Pattem, Antonio Ortega, and Bhaskar Krishnamachari. Energy-efficient data representation and routing for wireless sensor networks based on a distributed wavelet compression algorithm. In *Proceedings of the 5th international conference on Information processing in sensor networks*, pages 309–316. ACM, 2006. 10

P. L. Combettes and J.-C. Pesquet. A proximal decomposition method for solving convex variational inverse problems. *Inverse Problems*, 24(6), Dec. 2008. 34

P. L. Combettes and J.-C. Pesquet. Proximal Splitting Methods in Signal Processing. In Burachik R.S. Combettes P.L. Elser V. Luke D.R.; Wolkowicz H. (Eds.) Bauschke, H.H., editor, *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*, pages 185–212. Springer, New York, 2010. 34, 51, 68

P. L. Combettes and J.-C. Pesquet. Primal-dual splitting algorithm for solving inclusions with mixtures of composite, lipschitzian, and parallel-sum type monotone operators. *Set-Valued and Variational Analysis*, 20(2):307–330, june 2011. 34, 36, 51, 68, 70

Patrick L Combettes and Jean-Christophe Pesquet. Primal-dual splitting algorithm for solving inclusions with mixtures of composite, lipschitzian, and parallel-sum type monotone operators. *Set-Valued and variational analysis*, 20(2):307–330, 2012. 22, 23

Laurent Condat. A primal–dual splitting method for convex optimization involving lipschitzian, proximable and linear composite terms. *Journal of Optimization Theory and Applications*, 158(2):460–479, 2013. 22

Razvan Cristescu, Baltasar Beferull-Lozano, and Martin Vetterli. On network correlated data gathering. In *INFOCOM 2004. Twenty-third AnnualJoint Conference of the IEEE Computer and Communications Societies*, volume 4, pages 2571–2582. IEEE, 2004. 10

Razvan Cristescu, Baltasar Beferull-Lozano, Martin Vetterli, and Roger Wattenhofer. Network correlated data gathering with explicit communication: Np-completeness and algorithms. *IEEE/ACM Transactions On Networking*, 14(1):41–54, 2006. 9

Guy Demoment. Image reconstruction and restoration: Overview of common estimation structures and problems. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 37(12):2024–2036, 1989. 16

D. L. Donoho. Compressed sensing. *IEEE Trans. Inf. Theor.*, 52(4):1289–1306, April 2006a. ISSN 0018-9448. 28

D. L. Donoho, Y. Tsaig, I. Drori, and J. L. Starck. Sparse solution of underdetermined systems of linear equations by stagewise orthogonal matching pursuit. *IEEE Trans. Inf. Theor.*, 58(2):1094–1121, February 2012. ISSN 0018-9448. 29, 88

David L Donoho. Compressed sensing. *IEEE Transactions on information theory*, 52 (4):1289–1306, 2006b. 20

David L Donoho, Michael Elad, and Vladimir N Temlyakov. Stable recovery of sparse overcomplete representations in the presence of noise. *IEEE Transactions on information theory*, 52(1):6–18, 2006. 88

Maryam Fazel. *Matrix rank minimization with applications*. PhD thesis, PhD thesis, Stanford University, 2002. 20

Mário AT Figueiredo, Robert D Nowak, and Stephen J Wright. Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems. *IEEE Journal of selected topics in signal processing*, 1(4):586–597, 2007. 20

Alyson K Fletcher, Sundeep Rangan, and Vivek K Goyal. Estimation from lossy sensor data: Jump linear modeling and kalman filtering. In *Information Processing in Sensor Networks, 2004. IPSN 2004. Third International Symposium on*, pages 251–258. IEEE, 2004. 12

Michel Fortin and Roland Glowinski. *Augmented Lagrangian methods: applications to the numerical solution of boundary-value problems*, volume 15. Elsevier, 2000. 21

Daniel Gabay and Bertrand Mercier. A dual algorithm for the solution of nonlinear variational problems via finite element approximation. *Computers & Mathematics with Applications*, 2(1):17–40, 1976. 21

Victor Garcia-Font, Carles Garrigues, and Helena Rifà-Pous. A comparative study of anomaly detection techniques for smart city wireless sensor networks. *Sensors*, 16 (6):868, 2016. ISSN 1424-8220. doi: 10.3390/s16060868. 16

Samir Goel and Tomasz Imielinski. Prediction-based monitoring in sensor networks: taking lessons from mpeg. *ACM SIGCOMM Computer Communication Review*, 31 (5):82–98, 2001. 11

Tom Goldstein and Stanley Osher. The split bregman method for l1-regularized problems. *SIAM journal on imaging sciences*, 2(2):323–343, 2009. 21

T Guilford, J Meade, J Willis, Richard A Phillips, D Boyle, S Roberts, M Collett, R Freeman, and CM Perrins. Migration and stopover in a small pelagic seabird, the manx shearwater puffinus puffinus: insights from machine learning. *Proceedings of the Royal Society of London B: Biological Sciences*, pages rspb–2008, 2009. 7, 46

Douglas M Hawkins. *Identification of outliers*, volume 11. Springer, 1980. 13

Michael Holmes, Alexander Gray, and Charles Isbell. Fast svd for large-scale matrices. In *Workshop on Efficient Machine Learning at NIPS*, volume 58, pages 249–252, 2007. 79

Guogang Hua and Chang Wen Chen. Correlated data gathering in wireless sensor networks based on distributed source coding. *International Journal of Sensor Networks*, 4(1-2):13–22, 2008. 10

AK Jain and RC Dubes. Algorithms for clustering data, prentice-hall, inc. upper saddle river. *NJ, USA*, 1988. 15

Ankur Jain, Edward Y Chang, and Yuan-Fang Wang. Adaptive stream resource management using kalman filters. In *Proceedings of the 2004 ACM SIGMOD international conference on Management of data*, pages 11–22. ACM, 2004. 11

Dharanipragada Janakiram, VA Reddy, and AVU Phani Kumar. Outlier detection in wireless sensor networks using bayesian belief networks. In *Communication System Software and Middleware, 2006. Comsware 2006. First International Conference on*, pages 1–6. IEEE, 2006. 14

Minwook C Jun, H Jeong, and C-C Jay Kuo. Distributed spatio-temporal outlier detection in sensor networks. In *Proceedings of SPIE*, volume 5819, pages 273–284, 2005. 14

Edwin M Knox and Raymond T Ng. Algorithms for mining distancebased outliers in large datasets. In *Proceedings of the International Conference on Very Large Data Bases*, pages 392–403. Citeseer, 1998. 15

Ni. Komodakis and J-C. Pesquet. Playing with Duality: An Overview of Recent Primal-Dual Approaches for Solving Large-Scale Optimization Problems. *IEEE Signal Processing Magazine*, 32(6):31–54, November 2015. 34, 51

R. Kwitt, P. Meerwald, and A. Uhl. Color-image watermarking using multivariate power-exponential distribution. In *IEEE International Conference on Image Processing*, pages 4245 –4248, Cairo, Egypt, Nov 2009. 31, 50

Mo Li and Yunhao Liu. Underground structure monitoring with wireless sensor networks. In *Proceedings of the 6th international conference on Information processing in sensor networks*, pages 69–78. ACM, 2007. 8

Yujin Lim, Hak-Man Kim, and Sanggil Kang. A design of wireless sensor networks for a power quality monitoring system. *Sensors*, 10(11):9712–9725, 2010. 8

Hancong Liu, Sirish Shah, and Wei Jiang. On-line outlier detection and data cleaning. *Computers & chemical engineering*, 28(9):1635–1647, 2004. 53

J. Luo, L. Xiang, and C. Rosenberg. Does Compressed Sensing Improve the Throughput of Wireless Sensor Networks? In *IEEE, International Conference on Communications*, pages 1–6, May 2010. 12

Roberto Magán-Carrión, JosÉ Camacho, and Pedro García-Teodoro. Multivariate statistical approach for anomaly detection and lost data recovery in wireless sensor networks. *International Journal of Distributed Sensor Networks*, 11(6), 2015. doi: 10.1155/2015/672124. 15

Alan Mainwaring, David Culler, Joseph Polastre, Robert Szewczyk, and John Anderson. Wireless sensor networks for habitat monitoring. In *Proceedings of the 1st ACM international workshop on Wireless sensor networks and applications*, pages 88–97. Acm, 2002. 7

Y. Marnissi, A. Benazza-Benyahia, E. Chouzenoux, and J-C. Pesquet. Generalized multivariate exponential power prior for wavelet-based multichannel image restoration. In *IEEE International Conference on Image Processing*, pages 2402–2406, Melbourne, Australia, Sep 2013. 31, 50

Chris Masden. The paradigm shift of the internet of things and its impact on and in the mems and sensors industry. In *MEMS Executive Congress. MEMS Industry Group*, 2014. 1

M. A. Moussa, Y. Marnissi, and Y. Ghamri-Doudane. A Primal-Dual algorithm for data gathering based on matrix completion for wireless sensor networks. In *IEEE, International Conference on Communications*, Kuala Lumpur, Malaysia, May 2016. 54

Fabian Nack. An overview on wireless sensor networks. *Institute of Computer Science (ICS), Freie Universität Berlin*, 2010. 6

Tomasz Naumowicz, Robin Freeman, Andreas Heil, Martin Calsyn, Eric Hellmich, Alexander Brändle, Tim Guilford, and Jochen Schiller. Autonomous monitoring of vulnerable habitats using a wireless sensor network. In *Proceedings of the workshop on Real-world wireless sensor networks*, pages 51–55. ACM, 2008. 7, 46

Themistoklis Palpanas, Dimitris Papadopoulos, Vana Kalogeraki, and Dimitrios Gunopulos. Distributed deviation detection in sensor networks. *ACM SIGMOD Record*, 32(4):77–82, 2003. 14

Ioannis Ch. Paschalidis and Yin Chen. Statistical anomaly detection with sensor networks. *ACM Trans. Sen. Netw.*, 7(2):17:1–17:23, September 2010. ISSN 1550-4859. doi: 10.1145/1824766.1824773. 15

European Symposium PKDD., Jan M Zytkow, and Mohamed Quafafou. *Principles of data mining and knowledge discovery*. Springer, 1998. 15

Joseph Polastre, Robert Szewczyk, Alan Mainwaring, David Culler, and John Anderson. Analysis of wireless sensor networks for habitat monitoring. In *Wireless sensor networks*, pages 399–423. Springer, 2004. 7

Florian A. Potra and Stephen J. Wright. Primal-dual interior-point methods. SIAM, 1997. 29

Sutharshan Rajasegarar, Christopher Leckie, Marimuthu Palaniswami, and James C Bezdek. Distributed anomaly detection in wireless sensor networks. In *Communication systems, 2006. ICCS 2006. 10th IEEE Singapore International Conference on*, pages 1–5. IEEE, 2006. 15

Sridhar Ramaswamy, Rajeev Rastogi, and Kyuseok Shim. Efficient algorithms for mining outliers from large data sets. In *ACM Sigmod Record*, volume 29, pages 427–438. ACM, 2000. 15

Audrey Repetti, Mai Quyen Pham, Laurent Duval, Emilie Chouzenoux, and Jean-Christophe Pesquet. Euclid in a taxicab: Sparse blind deconvolution with smoothed $\ell_1\backslash\ell_2$ regularization. *IEEE Signal Processing Letters*, 22(5):539–543, 2015. 19, 103

Silvia Santini. Towards adaptive wireless sensor networks. In *3rd European Workshop on Wireless Sensor Networks (EWSN 2006)*, pages 13–15, 2006. 10

Silvia Santini and Kay Romer. An adaptive strategy for quality-based data reduction in wireless sensor networks. In *Proceedings of the 3rd international conference on networked sensing systems (INSS 2006)*, pages 29–36, 2006. 10

Bo Sheng, Qun Li, Weizhen Mao, and Wen Jin. Outlier detection in sensor networks. In *Proceedings of the 8th ACM international symposium on Mobile ad hoc networking and computing*, pages 219–228. ACM, 2007. 14, 16

David Slepian and Jack Wolf. Noiseless coding of correlated information sources. *IEEE Transactions on information Theory*, 19(4):471–480, 1973. 10

Sharmila Subramaniam, Themis Palpanas, Dimitris Papadopoulos, Vana Kalogeraki, and Dimitrios Gunopulos. Online outlier detection in sensor data using non-parametric models. In *Proceedings of the 32nd international conference on Very large data bases*, pages 187–198. VLDB Endowment, 2006. 14

Jukka Suomela. Median filtering is equivalent to sorting. *arXiv preprint arXiv:1406.1717*, 2014. 53

Tanja Teuber, Gabriele Steidl, and Raymond Honfu Chan. Minimization and parameter estimation for seminorm regularization models with i-divergence constraints. *Inverse Problems*, 29(3):035007, 2013. 65

Quoc Tran-Dinh, Anastasios Kyrillidis, and Volkan Cevher. Composite self-concordant minimization. *arXiv preprint arXiv*, 1308, 2013. 21

Bng Công Vũ. A splitting algorithm for dual monotone inclusions involving cocoercive operators. *Advances in Computational Mathematics*, 38(3):667–681, 2013. 22

Hongjian Wang, Yanmin Zhu, and Qian Zhang. Compressive sensing based monitoring with vehicular networks. In *INFOCOM, 2013 Proceedings IEEE*, pages 2823–2831. IEEE, 2013. 12, 27

Jin Wang, Shaojie Tang, Baocai Yin, and Xiang-Yang Li. Data gathering in wireless sensor networks through intelligent compressive sensing. In *INFOCOM, 2012 Proceedings IEEE*, pages 603–611. IEEE, 2012. 12, 27

Geoff Werner-Allen, Konrad Lorincz, Jeff Johnson, Jonathan Lees, and Matt Welsh. Fidelity and yield in a volcano monitoring sensor network. In *Proceedings of the 7th symposium on Operating systems design and implementation*, pages 381–396. USENIX Association, 2006. 8

Geoffrey Werner-Allen, Jeff Johnson, Mario Ruiz, Jonathan Lees, and Matt Welsh. Monitoring volcanic eruptions with a wireless sensor network. In *Wireless Sensor Networks, 2005. Proceeedings of the Second European Workshop on*, pages 108–120. IEEE, 2005. 13

AJ Whittle, M Allen, A Preis, and M Iqbal. Sensor networks for monitoring and control of water distribution systems. 2013. 8, 46

Ian H. Witten, Eibe Frank, and Mark A. Hall. *Data Mining: Practical Machine Learning Tools and Techniques*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 3rd edition, 2011. ISBN 0123748569, 9780123748560. 53

Weili Wu, Xiuzhen Cheng, Min Ding, Kai Xing, Fang Liu, and Ping Deng. Localized outlying and boundary data detection in sensor networks. *IEEE transactions on knowledge and data engineering*, 19(8):1145–1157, 2007. 14

L. Xiang, J. Luo, and C. Rosenberg. Compressed data aggregation: Energy efficient and high fidelity data collection. *IEEE Transactions on Networking*, 21(6):1722–1735, 2013. 12

Miao Xie, Song Han, Biming Tian, and Sazia Parvin. Anomaly detection in wireless sensor networks: A survey. *J. Netw. Comput. Appl.*, 34(4):1302–1325, July 2011. ISSN 1084-8045. doi: 10.1016/j.jnca.2011.03.004. 15

K. Yi, J. Wan, T. Bao, and L. Yao. A DCT regularized matrix completion algorithm for energy efficient data gathering in wireless sensor networks. *International Journal of Distributed Sensor Networks*, 19(11):54–57, 2015. 13

Dan C Youla and Heywood Webb. Image restoration by the method of convex projections: Part 1: Theory. *Medical Imaging, IEEE Transactions on*, 1(2):81–94, 1982. 65

Y. Zhang, M. Roughan, W. Willinger, and L. Qiu. Spatio-temporal compressive sensing and internet traffic matrices. In *The ACM SIGCOMM conference on Data communication*, pages 267–278, 2009. 50

Jerry Zhao, Ramesh Govindan, and Deborah Estrin. Computing aggregates for monitoring wireless sensor networks. In *Sensor Network Protocols and Applications, 2003. Proceedings of the First IEEE. 2003 IEEE International Workshop on*, pages 139–148. IEEE, 2003. 12

Haifeng Zheng, Shilin Xiao, Xinbing Wang, Xiaohua Tian, and Mohsen Guizani. Capacity and delay analysis for data gathering with compressive sensing in wireless sensor networks. *IEEE Transactions on Wireless Communications*, 12(2):917–927, 2013. 12, 27