

**EXPLOITING PRIOR KNOWLEDGE IN COMPRESSED SENSING  
WIRELESS ECG SYSTEMS**

by

Luisa F. Polania

A dissertation submitted to the Faculty of the University of Delaware in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Electrical and Computer Engineering

Summer 2015

© 2015 Luisa F. Polania  
All Rights Reserved

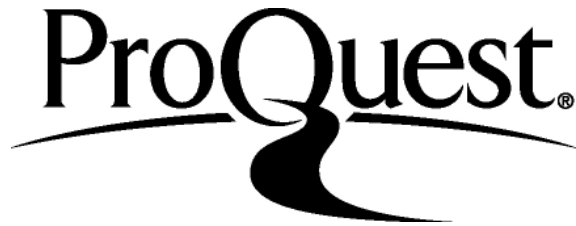
ProQuest Number: 3730214

All rights reserved

INFORMATION TO ALL USERS

The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if material had to be removed, a note will indicate the deletion.



ProQuest 3730214

Published by ProQuest LLC (2015). Copyright of the Dissertation is held by the Author.

All rights reserved.

This work is protected against unauthorized copying under Title 17, United States Code  
Microform Edition © ProQuest LLC.

ProQuest LLC.  
789 East Eisenhower Parkway  
P.O. Box 1346  
Ann Arbor, MI 48106 - 1346

**EXPLOITING PRIOR KNOWLEDGE IN COMPRESSED SENSING  
WIRELESS ECG SYSTEMS**

by

Luisa F. Polania

Approved: \_\_\_\_\_  
Kenneth E. Barner, Ph.D.  
Chair of the Department of Electrical and Computer Engineering

Approved: \_\_\_\_\_  
Babatunde Ogunnaike, Ph.D.  
Dean of the College of Engineering

Approved: \_\_\_\_\_  
James G. Richards, Ph.D.  
Vice Provost for Graduate and Professional Education

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed: \_\_\_\_\_  
Kenneth E. Barner, Ph.D.  
Professor in charge of dissertation

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed: \_\_\_\_\_  
Charles G. Boncelet, Ph.D.  
Member of dissertation committee

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed: \_\_\_\_\_  
Ryan Zurakowski, Ph.D.  
Member of dissertation committee

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed: \_\_\_\_\_  
Manuel Blanco-Velasco, Ph.D.  
Member of dissertation committee

## ACKNOWLEDGEMENTS

My deep gratitude goes first to Prof. Barner for providing the right combination of supervision, encouragement, and support during my Ph.D. studies. I thank him for his efforts on helping me become a better researcher and writer. The freedom that Prof. Barner gave me to approach research problems and to study topics of my interest was extremely important to keep me motivated during the past five years. I thank my thesis committee for providing valuable suggestions and contributions to improve the quality of my work. I would also like to thank my collaborators, whose significant input is reflected in this document: Rafael Carrillo, Manuel Blanco-Velasco, and Sherin Mathews. Rafael introduced me to the area of compressed sensing. I thank Manuel for sharing his expertise on ECG compression and for his continuous encouragement. I also thank Sherin for the hours she enthusiastically devoted to study our research problems.

My appreciation also extends to my labmates. Thanks to Xin Guo, Sherin, Yin Zhou, Jinglun Gao, and Rui Hu. You all contributed to the positive atmosphere that made our lab a nice place to work. I must thank my wonderful friends for their continuous support over the years. You filled my life at Delaware with joyful hours. Special thanks to: Rafael Plaza, Qi Wang, Irene Manotas, Elpiniki, Elkin, Anagha, Iñaki Iglesias, Sergio Pino, Irene de Teresa, Jesus, and Liuda. Thanks to my childhood friends, Ximena, Angela and Camila, for their continuous support and encouragement regardless of the distance.

Finally, I am deeply grateful to my loving family. My parents are my role models of hard work and dedication. Thanks to my brother for his enthusiasm, good humor, and helping me keep things in perspective.

## TABLE OF CONTENTS

<b>LIST OF TABLES</b> . . . . .	<b>ix</b>
<b>LIST OF FIGURES</b> . . . . .	<b>x</b>
<b>ABSTRACT</b> . . . . .	<b>xiii</b>
 <b>Chapter</b>	
<b>1 INTRODUCTION</b> . . . . .	<b>1</b>
1.1 Motivation and Background . . . . .	1
1.2 Contributions and Dissertation Outline . . . . .	4
1.3 Notation . . . . .	8
1.4 Brief Review of Compressed Sensing . . . . .	9
1.4.1 Introduction . . . . .	9
1.4.2 The Compressed Sensing Problem . . . . .	10
1.4.2.1 The null-space property . . . . .	13
1.4.2.2 The restricted isometry property . . . . .	14
1.4.2.3 Coherence . . . . .	15
1.4.3 Reconstruction Methods . . . . .	17
1.4.3.1 Convex optimization methods . . . . .	17
1.4.3.2 Greedy algorithms . . . . .	18
1.4.4 The Sparsity Basis . . . . .	21
<b>2 EXPLOITING PRIOR KNOWLEDGE IN COMPRESSED SENSING WIRELESS ECG SYSTEMS</b> . . . . .	<b>24</b>
2.1 Introduction . . . . .	24
2.2 Background . . . . .	25
2.2.1 Brief Review of Model-based Compressed Sensing . . . . .	25

2.2.2	Tree-structured Sparsity Models . . . . .	27
2.3	Motivation . . . . .	28
2.3.1	Connected Subtree Structure of ECG Wavelet Coefficients . . . . .	29
2.3.2	Common Support Between Consecutive ECG Segments . . . . .	32
2.4	Methods . . . . .	33
2.4.1	Sampling and Encoding . . . . .	33
2.4.2	Reconstruction Algorithm using Prior Information . . . . .	34
2.5	Experimental Results . . . . .	37
2.5.1	Performance Evaluation . . . . .	38
2.5.2	Practical Considerations . . . . .	38
2.5.3	Evaluation of ECG Reconstruction using the Proposed Method . . . . .	43
2.5.4	Evaluation of Compression Performance . . . . .	44
2.5.5	Visual Evaluation . . . . .	48
2.6	Conclusions . . . . .	51
<b>3</b>	<b>MULTI-SCALE DICTIONARY LEARNING FOR COMPRESSIVE SENSING ECG . . . . .</b>	<b>52</b>
3.1	Introduction . . . . .	52
3.2	Background and Related Work . . . . .	53
3.2.1	Traditional Dictionary Learning . . . . .	53
3.2.2	Multi-scale Dictionary Learning using Wavelets . . . . .	54
3.3	Methods . . . . .	55
3.3.1	Dictionary Learning in the Wavelet Domain . . . . .	56
3.3.2	Compressed Sensing Reconstruction of ECG Signals . . . . .	56
3.4	Experimental Results . . . . .	58
3.5	Conclusions . . . . .	63
<b>4</b>	<b>A WEIGHTED <math>\ell_1</math> MINIMIZATION ALGORITHM FOR COMPRESSED SENSING ECG . . . . .</b>	<b>64</b>
4.1	Introduction . . . . .	64

4.2	Motivation and Background . . . . .	65
4.2.1	Problem Formulation . . . . .	65
4.2.2	Wavelet representation . . . . .	65
4.2.3	Motivation . . . . .	66
4.3	Methods . . . . .	68
4.4	Experimental Results . . . . .	70
4.5	Conclusion . . . . .	74
<b>5</b>	<b>COMPRESSIVE SENSING FOR ECG SIGNALS IN THE PRESENCE OF ELECTROMYOGRAPHIC NOISE . . . . .</b>	<b>75</b>
5.1	Introduction . . . . .	75
5.2	Background . . . . .	77
5.2.1	Noise in Ambulatory ECG . . . . .	77
5.2.2	Review of $S\alpha S$ Distributions . . . . .	78
5.3	Robust Reconstruction Algorithm . . . . .	79
5.3.1	Problem Formulation . . . . .	79
5.3.2	An Iterative Reweighted Least Squares Approach . . . . .	80
5.3.3	Algorithm . . . . .	81
5.4	Experimental Results . . . . .	82
5.5	Conclusions . . . . .	87
<b>6</b>	<b>EXPLOITING RESTRICTED BOLTZMANN MACHINES AND DEEP LEARNING ARCHITECTURES IN COMPRESSED SENSING WIRELESS ECG SYSTEMS . . . . .</b>	<b>88</b>
6.1	Introduction . . . . .	88
6.2	Background on Deep Belief Networks and Restricted Boltzmann Machines . . . . .	90
6.3	Proposed Compressed Sensing Scheme . . . . .	92
6.3.1	Compressed Sensing Stage . . . . .	93
6.3.1.1	Problem formulation . . . . .	93
6.3.1.2	Prior distribution . . . . .	95



6.3.1.3	MAP estimator via a greedy approach . . . . .	97
6.3.2	Training Stage using Overcomplete Learned Dictionaries . . .	98
6.3.2.1	Traditional dictionary learning . . . . .	98
6.3.2.2	Estimation of model parameters . . . . .	99
6.3.3	Training Stage using Orthonormal Bases . . . . .	100
6.4	Experimental Results . . . . .	101
6.4.1	Performance Evaluation . . . . .	101
6.4.2	Qualitative Observations . . . . .	105
6.4.3	Robustness to Noise . . . . .	109
6.4.4	Parameter Evaluation . . . . .	109
6.5	Conclusions . . . . .	111
<b>7</b>	<b>SUMMARY AND FUTURE WORK . . . . .</b>	<b>113</b>
7.1	Summary . . . . .	113
7.2	Future Work . . . . .	114
	<b>BIBLIOGRAPHY . . . . .</b>	<b>117</b>
	<b>Appendix</b>	
	<b>COPYRIGHT FORMS . . . . .</b>	<b>125</b>

## LIST OF TABLES

2.1	PRD obtained by MMB-IHT and MMB-CoSaMP for different sensing matrices . . . . .	42
2.2	Performance of MMB-IHT and MMB-CoSaMP for CR=6.4 . . . . .	47

## LIST OF FIGURES

1.1	Best approximation of a signal in $\mathbb{R}^2$ by a one-dimensional subspace using the $\ell_1$ norm (left) and the $\ell_2$ norm (right). . . . .	18
1.2	Sparsity of ECG signals in the wavelet domain. (a) Original ECG sequence. (b) Wavelet representation. (c) Sorted wavelet coefficients in order of decreasing absolute value (logarithmic scale). (d) Recovered signal from the 10% largest (in magnitude) wavelet coefficients. . . . .	23
2.1	Daubechies-4 coefficients subvectors $d_j$ , $j = 1 \dots 5$ , and $a_5$ for ECG time series. . . . .	30
2.2	Fraction of common support between consecutive ECG segments. The results are averaged over the selected data set from the MIT-BIH Arrhythmia Database. . . . .	31
2.3	Block Diagram of the proposed ECG compression method. . . . .	33
2.4	Wavelet decomposition of level $L = 5$ for ECG time series using Daubechies-4. . . . .	36
2.5	Sequence of residual energy averaged over the selected set of records from the MIT-BIH Arrhythmia Database. . . . .	40
2.6	Reconstruction SNR as a function of the wavelet decomposition level. Number of measurements $m = 3K$ . . . . .	42
2.7	Comparison of MMB-CoSaMP and MMB-IHT with CoSaMP, IHT, and the oracle estimate. Reconstruction SNR averaged over all the records of the selected data set for different number of measurements. . . . .	45
2.8	Compression performance evaluation of the compression scheme using the proposed algorithms (MMB-CoSaMP, MMB-IHT), the BO-BSBL, the MS-BPDN, and the BPDN algorithms. The results of SPIHT are included as a baseline for comparisons. . . . .	46

2.9	Visual evaluation of the reconstruction of record 119 using MMB-CoSaMP and MMB-IHT. $CR = 6.4$ . PRD=2.61 (MMB-CoSaMP), PRD=2.29 (MMB-IHT). . . . .	49
2.10	Visual evaluation of the reconstruction of record 118 using MMB-CoSaMP and MMB-IHT. $CR = 6.4$ . PRD=2.86 (MMB-CoSaMP), PRD=2.54 (MMB-IHT). . . . .	50
3.1	Block Diagram of the proposed method . . . . .	55
3.2	Comparison of CS-based methods using different dictionaries. ECG compression is evaluated using PRD as a function of the compression ratio. . . . .	59
3.3	Reconstruction results of wavelet subbands $V_2, V_3$ , and $V_4$ using CS-based methods with different dictionaries. (a) Approximation subband, $V_4$ . (b) Detail subband $V_3$ . (c) Detail subband $V_2$ . . . . .	61
3.4	Reconstruction of record 100 using using 65 measurements. PRD=4.35. (a) Original signal. (b) Reconstructed signal. (c) Error signal. . . . .	62
4.1	Reconstruction of ECG signals using BPDN (a) Original signal. (b) Wavelet representation. (c) Relative error signal. . . . .	67
4.2	Scatter plot of $j$ versus $\log_2 \sigma_j^2$ and fitted regression line. . . . .	71
4.3	Comparison of the proposed method with BPDN. . . . .	73
4.4	Comparison of the proposed method with BPDN. PRD as a function of the compression ratio. . . . .	73
5.1	Muscle artifact record from the MIT-BIH Noise Stress Test Database. . . . .	78
5.2	Comparison of the proposed method with IHT-PKS. . . . .	85
5.3	Visual evaluation of the reconstruction of record 117 using the proposed recovery algorithm. . . . .	86
6.1	(a) Schematic of a restricted Boltzmann machine. (b) Schematic of a deep belief network of one visible and three hidden layers. . . . .	92
6.2	Block diagram of the proposed CS scheme. . . . .	93

6.3	Comparison of the reconstruction of ECG signals using the OMP, the RBM-OMP-like algorithm, and the DBN-OMP-like algorithm for both wavelets and overcomplete learned dictionaries. . . . .	102
6.4	Visual evaluation of the RBM-OMP-like algorithm. Record 119, $M = 0.35N$ . First row: (a) Original signal. Second and third rows: (b-e) Reconstructed signals using wavelets (b-c) OMP reconstruction and error, R-SNR=18.32, (d-e) RBM-OMP-like reconstruction and error, R-SNR=25. Fourth and fifth rows: (f-i) Reconstructed signals using an overcomplete dictionary (f-g) OMP reconstruction and error, R-SNR=24.59, (h-i) RBM-OMP-like reconstruction and error, R-SNR=31.67. . . . .	103
6.5	Comparison of the reconstruction of ECG signals using the RBM-OMP-like algorithm, the DBN-OMP-like algorithm, MB-CoSaMP, MB-IHT, and the BO-BSBL algorithm. PRD as a function of the compression ratio. . . . .	105
6.6	(a) Visible layer bias terms of the RBM model using learned overcomplete dictionaries. (b) Dictionary atom sharing the same index as the largest bias term. . . . .	106
6.7	(a) Visible layer bias terms of the RBM model using wavelets. (b) Wavelet representation of an ECG segment of 128 samples. . . . .	107
6.8	(a) Weights associated with the hidden unit with the largest bias term. (b-c) Dictionary atoms sharing the same indexes as the visible units with the 2 most positive weights. (d) Dictionary atom sharing the same index as the visible unit with the most negative weight. . . . .	108
6.9	Performance of the RBM-OMP-like and OMP algorithms in the presence of measurement noise . . . . .	110
6.10	Reconstruction performance of the RBM-OMP-like algorithm using an overcomplete dictionary as a function of the dictionary size. . . . .	111
6.11	Reconstruction performance of the RBM-OMP-like algorithm using an overcomplete dictionary as a function of the sparsity threshold. . . . .	112

## ABSTRACT

Wireless body area networks promise to revolutionize health monitoring by allowing the transition from centralized health care services to ubiquitous and pervasive health monitoring in every-day life. One of the major challenges in the design of such systems is the energy consumption as wireless body area networks are battery-powered. Recent results in telecardiology show that compressed sensing (CS) is a promising tool to lower energy consumption in wireless body area networks for electrocardiogram (ECG) monitoring. However, the performance of current CS-based algorithms, in terms of compression rate and reconstruction quality of the ECG, still falls behind the performance attained by state-of-the-art wavelet-based algorithms. This is mainly because current CS-based algorithms exploit only the sparsity of the signal, ignoring important signal structure information that can be known *a priori* and lead to enhanced reconstruction results.

This dissertation presents methods to exploit prior knowledge of the ECG in order to improve the reconstruction quality and to increase the compression rates offered by current CS-based algorithms. First, we describe an algorithm that exploits prior information about the wavelet dependencies across scales and the high fraction of common support of the wavelet coefficients of consecutive ECG segments.

One of the main challenges in the reconstruction of ECG signals via CS-based algorithms is the recovery of the small-magnitude wavelet coefficients. This dissertation also presents a weighted  $\ell_1$  minimization algorithm, based on a maximum *a posteriori* (MAP) approach, that exploits the exponentially decaying magnitude of the detail coefficients across scales and the accumulation of signal energy in the approximation subband.

In real scenarios, ECG recordings are often corrupted by artifacts. This dissertation also presents a robust reconstruction method for ECG signals in the presence of electromyographic noise. To achieve this objective, robust statistics are used to develop appropriate methods addressing the problem of electromyographic noise, which can be modeled as impulsive noise.

Most prior work in CS ECG has employed analytical sparsifying transforms such as wavelets. Another contribution of this dissertation is to adaptively learn a sparsifying transform (overcomplete dictionary) that exploits the multi-scale sparse representation of ECG signals. By calculating subdictionaries at different data scales, we are able to exploit the correlation within each wavelet subband and, subsequently, represent the data in a more efficient manner.

Generic sparsity models that are not tied to a specified structure are also explored in this dissertation. More precisely, restricted Boltzmann machines and deep belief networks are employed to model the sparsity pattern of ECG signals with the goal of exploiting higher-order statistical dependencies between sparse coefficients.

The effectiveness of the proposed algorithms is demonstrated on real ECG signals from the MIT-BIH Arrhythmia Database. Results show that the proposed algorithms require fewer measurements and offer superior reconstruction accuracy than existing CS-based methods for ECG compression.

# Chapter 1

## INTRODUCTION

### 1.1 Motivation and Background

A wireless body area network (WBAN) is a radio frequency-based wireless networking technology that connects small nodes with sensor capabilities in, on, or around a human body. These biosensors are capable of measuring physiological signals, such as heart rate, blood pressure, electrocardiogram, blood oxygen level, respiration rate, etc. The obtained measurements are transmitted to a central processing system, which transfers the information to a medical network where health-care professionals can assess the users health condition [16]. The main motivation behind the advancement of WBANs is the need to provide pervasive, long-term and continuous healthcare monitoring without restricting the user from their daily life activities.

The energy consumption is a primary design constraint in WBANs given that they are battery powered [16]. The WBAN energy consumption can be divided into three main processes: sensing, wireless communication and data processing. However, the process that consumes most of the energy is the wireless transmission of data [16], which indicates that some data reduction operation should be performed at the sensor node to reduce the energy cost of the network. In addition, data reduction can supplement the bandwidth constraints of the WBAN when many sensor nodes are required to measure different physiological signals. Mamaghanian *et al.* [56] recently proposed to use CS to lower energy consumption and complexity in WBAN-enabled ECG monitors. Compressed sensing is an emerging field that exploits the structure of signals to acquire data at a rate proportional to the information content rather than the frequency content, therefore allowing sub-Nyquist sampling [11, 13, 27]. When this



sampling strategy is introduced in WBANs, it gives rise to sensor nodes that efficiently acquire a small group of random linear measurements that are wirelessly transmitted to remote terminals. Indeed, sensor nodes can achieve high compression of ECG data with low computational cost when using a sparse binary sensing matrix [56]. Mamaghanian *et al.* [56] performed system-level comparisons between a CS-based algorithm and a state-of-the-art wavelet transform-based algorithm for ECG compression. Both algorithms were implemented on the Shimmer embedded platform and the comparison was based on signal reconstruction metrics, embedded memory usage, energy consumption, and CPU execution time.

In another publication [21], Chen *et al.* presented a hardware implementation of a CS encoder, fabricated in a 90 nm CMOS. Their architecture was presented in the context of medical sensors and is the first fully integrated circuit realization of a CS encoder. They employed a pseudo-random Bernoulli sampling matrix in order to simplify matrix multiplication operations and to lower complexity and data storage. A compression factor of 10X was attained without the need for any general purpose memory or processing at the sensor node. An important contribution of their work is a comparison between implementations of a CS encoder in the analog and digital domains. Analog domain implementations are usually preferred in CS scenarios since they lead to sampling frequency reduction of ADC converters. However, ADC performance is not a limitation in WBAN, and therefore the criteria for best performance is determined by the power consumption of the implementation. Chen *et al.* presented a power/performance analysis of analog and digital implementations of the proposed CS encoder and found that the digital implementation is more energy-efficient and suitable for WBANs. Similarly, Dixon *et al.* [26] presented a digital-domain CS encoder architecture with applications to ECG and electroencephalogram signals. Bernoulli matrices with different structures, such as Toeplitz, circulant and triangular, were considered in this work. The study indicates that the use of Bernoulli circulant matrices leads to compression factors of 16X and offers superior reconstruction performance for ECG signals and reduction in hardware complexity and energy consumption. The

reconstruction algorithms employed in their work were orthogonal matching pursuit, compressive sampling matching pursuit, and normalized iterative hard thresholding.

In a recent work [93], Zhang *et al.* proposed the application of block sparse Bayesian learning (BSBL) for the reconstruction of fetal ECG signals from compressed random measurements. The BSBL algorithm is a CS algorithm that exploits the spatial and structural information of block-sparse signals and achieves accurate reconstruction regardless of the non-sparsity of fetal ECG signals. Their study focuses on the reconstruction algorithm instead of the CS encoder hardware implementation. One of the advantages of the algorithm is that it exploits correlation within each block. The BSBL algorithm is based on a maximum *a posteriori* estimation that adopts a multi-variate Gaussian prior for the block sparse signal. The experimental results of their work suggest that incorporating signal structure information can significantly boost the performance of CS reconstruction algorithms. The successful application of CS to fetal ECG motivates the development of new algorithms that exploit structural information of adult ECG signals. The energy efficiency property associated to CS architectures is useful for wireless body area networks-enabled ECG systems.

Most state-of-the-art algorithms for ECG compression are based on wavelet transforms because of their straightforward implementation and desirable time and frequency localization properties [41, 44, 54]. The main works in this area are characterized by hierarchical tree structures, such as embedded zero-tree wavelet (EZW) [41] and set partitioning in hierarchical tree (SPIHT) [54], which leverage the correlations of the wavelet coefficients across scales within a hierarchically decomposed wavelet tree.

The EZW algorithm prioritizes the transmission of the location and sign of the most significant wavelet coefficients. For the transmission of the insignificant coefficients, the EZW algorithm exploits the self-similarity of the wavelet transform across scales, meaning that if a wavelet coefficient at a coarse scale has low magnitude, then it is highly likely that the wavelet coefficients in the same spatial location at finer scales also have low magnitude. The embedded coding scheme places the most important bits at the beginning of the bit-stream so that the encoding or decoding process can

terminate when a desired bit-count or distortion metric is met. A wavelet coefficient is considered significant if its magnitude exceeds the threshold of the current iteration. SPIHT is a modified and enhanced version of EZW. SPIHT differs from EZW in that significant coefficients are uncovered by examining entire sets and subsets until the algorithm narrows in on the exact coefficient value.

The traditional ECG compression methods based on wavelets, such as SPIHT and EZW, still outperform the CS-based methods in terms of compression rate and reconstruction quality. The reason behind this result is that current CS-based methods for ECG only consider the simplistic sparsity model, without exploiting all the structural information present in the ECG.

Most of the previous work related to the application of CS to ECG refers to the design of the sampling matrix and the hardware implementation of the CS encoder. There is limited work regarding the development of novel CS reconstruction algorithms especially adapted to ECG. Therefore, the motivation of this dissertation is the need for development of new CS reconstruction algorithms that can be applied to ECG signals and exploit their structural information.

To the best of our knowledge, there is no work that studies the ECG compressive sensing problem in the presence of real noise. However, there are numerous non-cardiac contaminants, such as electromyographic noise, that unavoidably corrupt the ECG during data acquisition. This fact leads to the second motivation of this dissertation: the need for developing CS techniques that address ECG reconstruction in the presence of electromyographic noise.

## **1.2 Contributions and Dissertation Outline**

This dissertation focuses on CS reconstruction algorithms for ECG signals by exploiting prior knowledge. Fundamental questions addressed include how to leverage the rich structure and quasi-periodicity of the ECG, as well as how to exploit knowledge of the noise sources corrupting the ECG acquisition in order to improve the performance of CS reconstruction algorithms as applied to ECG signals. The work presented in this

thesis is one of the first attempts on the application of CS to ECG signals. This work has led to nine conference papers [17, 19, 66, 67, 70–74], one journal publication [75], and two submitted journal articles [68, 69]. Our key contributions include:

- Exploration and analysis of the properties of the ECG wavelet representation that can be employed as prior information in CS reconstruction algorithms.
- Development of CS reconstruction algorithms that exploit the structure of the ECG wavelet representation. Specifically, the algorithms exploit the following properties:
  - Wavelet domain dependencies across scales.
  - Accumulation of energy in the approximation wavelet subband.
  - Magnitude decay of the wavelet coefficients across scales.
  - High fraction of common support between the wavelet representation of consecutive ECG segments.
- Development of a CS reconstruction algorithm that exploits higher-order statistical dependencies between coefficients of the ECG sparse representation using wavelets and overcomplete learned dictionaries.
- Design of a block CS scheme that employs a different overcomplete learned dictionary for each scale of signal representation.
- Development of robust CS methods for reconstruction of ECG signals in the presence of electromyographic noise.
- Evaluation performance of the proposed methods and comparison with state-of-the-art algorithms for ECG compression through extensive simulations.

These contributions are organized into five chapters:

In Chapter 2, we design, discuss and evaluate ECG reconstruction algorithms that exploit wavelet domain dependencies across subbands and common support between consecutive ECG segments. First, we study the ECG wavelet representation and note important properties such as the high fraction of common support shared by consecutive ECG segments of fixed length and the tree structure formed by the largest wavelet coefficients. We propose reconstruction algorithms that are based on the framework of model-based CS, a new CS framework that enables reduction in the number of measurements by exploiting prior knowledge of the signal support. The performance of the proposed algorithms is evaluated and compared with state-of-the-art algorithms for ECG compression.

Chapter 3 introduces the use of multi-scale dictionaries to the reconstruction of ECG signals. By employing subdictionaries at different data scale, we are able to exploit the correlation within each wavelet subband. This chapter also presents a block compressed sensing scheme that allows for independent sampling of each wavelet subband. This sampling scheme is very efficient as it offers the flexibility of using different number of measurements for different data scales according to the corresponding sparsity level. Simulations with the MIT-BIH Arrhythmia Database show the results of sensing and reconstructing the ECG sparse representation at each scale.

Chapter 4 addresses the CS reconstruction of ECG signals by exploiting the magnitude decay of the ECG wavelet coefficients across scales. We first note that the traditional  $\ell_1$  minimization fails to reconstruct the ECG wavelet representation as it ignores the magnitude disparity of the wavelet coefficients across subbands. A weighted  $\ell_1$  minimization algorithm, based on a maximum *a posteriori* approach, is presented in this chapter. The resulting weights depend on the exponential magnitude decay rate of the ECG wavelet coefficients. A training stage is also proposed for estimation of the decay rate via maximum likelihood. Simulation results show that the proposed weighted version of  $\ell_1$  minimization significantly outperforms the traditional version.

In Chapter 5, a robust CS method for the reconstruction of ECG signals in the presence of electromyographic noise is presented. It is known that the Gaussian distribution fails to model electromyographic noise due to its impulsive behavior. A more realistic approach is to model muscle noise with the symmetric  $\alpha$ -stable distribution, which is the model adopted in this chapter. For the reconstruction of the ECG sparse representation, an optimization problem that minimizes the  $\ell_p$  norm of the residual error subject to a constraint on the  $\ell_0$  norm of the sparse representation is proposed. Most CS algorithms use the  $\ell_2$  norm as they assume a Gaussian model for the noise; however, we employ the  $\ell_p$  norm because it is related to the dispersion, which measures the spread of the density around its location parameter for the case of symmetric  $\alpha$ -stable distributions. This chapter also introduces an iterative hard thresholding-based algorithm to solve the  $\ell_p$  minimization problem.

Chapter 6 presents a statistical approach that uses restricted Boltzmann machines (RBMs) and deep belief networks (DBNs) to model the prior distribution of the sparsity pattern of ECG signals. Such a prior is employed by a maximum *a posteriori* estimator for the reconstruction. It is shown through simulations that the proposed scheme leads to significantly superior reconstruction results when compared with CS methods that do not exploit any statistical dependencies between sparse coefficients. The proposed approach also outperforms CS methods that only exploit pair-wise correlations between sparse coefficients.

### 1.3 Notation

Let  $\Phi \in \mathbb{R}^{N \times M}$ ,  $x \in \mathbb{R}^N$ ,  $r \in \mathbb{R}$ , and  $\Gamma \subseteq \{1, 2, \dots, N\}$ , then

$\Phi^\dagger$	Moore-Penrose pseudo-inverse of matrix $\Phi$
$\Phi_\Gamma$	Submatrix obtained by extracting the columns of matrix $\Phi$ corresponding to the indexes in $\Gamma$
$\Phi^T$	Transpose of matrix $\Phi$
$\text{supp}(x)$	Support of vector $x$
$\Gamma^c$	Complement of the set $\Gamma$
$ \Gamma $	Cardinality of the set $\Gamma$
$ r $	Absolute value of real number $r$

## 1.4 Brief Review of Compressed Sensing

### 1.4.1 Introduction

For nearly six decades, the sampling rate of signal acquisition systems has been determined by the Shannon-Nyquist theorem, which states that the sampling rate must be at least twice the signal's bandwidth. This is commonly known as the Nyquist rate. However, we are living in an era of data deluge that calls for developing new acquisition systems that go beyond the traditional limits of sampling theory. There are modern applications where the resulting Nyquist rate is either too high to be physically realizable or too costly. In addition, the traditional sampling scheme is highly inefficient as it is usually followed by a compression stage where a large number of samples are discarded in order to reduce transmission rates and storage.

Compressed sensing has recently emerged as a response to the need for new sensing techniques. Compressed sensing is a new framework for signal acquisition and sensor design, which enables a significant reduction in the sampling and computation costs for sensing sparse or compressible signals. Under the CS framework, signals are directly sampled in a compressed form, which makes the sensing process much more efficient. The theory of CS was initially developed by Candès, Romberg, Tao [11, 13] and Donoho [27], who showed that a finite-dimensional sparse or compressible signal can be recovered from a small set of linear, nonadaptive measurements. Unlike traditional sampling that acquires samples uniformly in time, CS systems typically acquire measurements in the form of inner products between the signal and a set of vectors incoherent with the sparsity basis. Randomness plays a key role in the CS theory as random matrices satisfy the incoherence property with any orthonormal sparsity basis with high probability.

The CS reconstruction process also differs from traditional approaches. In the Nyquist-Shannon framework, signal reconstruction is achieved through sinc interpolation; while in CS, it involves nonlinear methods. Current reconstruction algorithms can be classified into three main categories: convex methods [14, 22, 84], greedy algorithms [8, 59, 85], and Bayesian methods [5, 81, 94]. Examples of convex methods include



basis pursuit denoising [22], which can recover the signal exactly under mild conditions and is robust to measurement noise. Greedy algorithms offer the advantage of being computationally efficient, but usually at the expense of reconstruction accuracy. Bayesian approaches exploit the sparsity condition by assuming a sparsity-inducing prior for the signal of interest that is then estimated via Bayesian inference. During the past 10 years, CS has been extensively applied by mathematicians, computer scientists and engineers in different different fields, such as medical imaging, computational biology, geoscience, radar, astronomy, and communications [48] .

This section summarizes the theory of CS. The organization of the section is as follows: In section 1.4.2, the problem of sensing and reconstruction of sparse signals is presented. The conditions that the sampling matrix need to satisfy to ensure signal recoverability are presented in this section as well. Section 1.4.3 introduces some of the main CS reconstruction algorithms in the literature. We focus on the algorithms that have been applied to the reconstruction of ECG signals. Finally, section 1.4.4 introduces the concept of a sparsity basis, which enables the extension of the CS theory to a larger family of signals.

### 1.4.2 The Compressed Sensing Problem

Under the traditional sampling scheme, there are applications (*e.g.*, digital images and video cameras) that require large efforts in order to sample the signal at the Nyquist rate, and afterwards most of the information is discarded at the compression stage because the number of samples exceeds the storage and transmission rates capabilities of current systems. There are other applications (*e.g.*, medical scanners and radars) where sampling at the Nyquist rate is too expensive or difficult to be physically realizable. Therefore, a question arises at this point: is it possible to accurately reconstruct a signal from a number of samples much lower than what is dictated by the Nyquist-Shannon theorem? The theory of CS offers a positive answer to this question in the case of sparse and compressible signals.

A signal  $x$  is  $K$ -sparse when it has at most  $K$  nonzeros, *i.e.*,  $\|x\|_0 = \text{supp}(x) = K$ . The set of all  $K$ -sparse signals is defined as

$$\Sigma_K = \{x : \|x\|_0 = K\}. \quad (1.1)$$

In real scenarios, it is rare to find exactly sparse signals. Instead, it is common to encounter signals that are compressible. A compressible signal is one which can be well approximated by a sparse signal (*i.e.* most of its entries are approximately zero). The compressibility of a signal  $x$  is measured in terms of the error produced by the best  $K$ -sparse approximation  $\hat{x} \in \Sigma_K$  in terms of the  $\ell_p$  norm

$$\sigma_K(x)_p = \min_{\hat{x} \in \Sigma_K} \|x - \hat{x}\|_p. \quad (1.2)$$

The compression of  $x$  reduces to storing the  $K$  largest entries of  $x$ .

Let  $x \in \mathbb{R}^N$  be a signal that is either  $K$ -sparse or compressible. Consider a sensing system that acquires  $M < N$  linear and nonadaptive measurements. This process can be mathematically represented as

$$y = \Phi x, \quad (1.3)$$

where  $\Phi \in \mathbb{R}^{M \times N}$  represents a dimensionality reduction and  $y \in \mathbb{R}^M$  represents the measurement vector. It seems impossible to recover  $x$  from  $y$  since the linear system (1.3) is underdetermined, and therefore, it has infinitely many solutions. However, the assumption that the signal is sparse or compressible changes the character of the problem. More precisely, the signal  $x$  can be reconstructed from  $y$  if the matrix  $\Phi$  satisfies some special conditions, *e.g.*, the null space property [27], the restricted isometry property [15], or the coherence [9].

The reconstruction of  $x$  is related to the  $\ell_p$  norm of a vector, which is defined for  $p \in [1, \infty]$  as

$$\|x\|_p = \begin{cases} (\sum_{i=1}^N |x_i|^p)^{\frac{1}{p}} & p \in [1, \infty) \\ \max_{i=1,2,\dots,N} |x_i| & p = \infty. \end{cases}$$

The signal  $x$  can be reconstructed from  $y$  by solving an  $\ell_1$  minimization problem

$$\arg \min_x \|x\|_1 \text{ s.t. } y = \Phi x. \quad (1.4)$$

Problem (1.4) is referred to as basis pursuit.

Real data acquisition systems are always affected by the presence of noise. Therefore, a more realistic setting is to assume noisy measurements of the form

$$y = \Phi x + r, \quad (1.5)$$

where  $r$  is white Gaussian noise. There are variations of the basis pursuit algorithm (1.4) that reliably approximate the signal of interest from noisy measurements. For example, the basis pursuit with  $\ell_2$  constraint relaxes the condition that the recovered signal explains exactly the measurements [10]. The reconstruction is attained by solving the optimization problem

$$\arg \min_x \|x\|_1 \text{ s.t. } \|y - \Phi x\|_2 \leq \epsilon \quad (1.6)$$

for some  $\epsilon > 0$ . Reconstruction guarantees for the basis pursuit with  $\ell_2$  constraint are provided in [10]. Basis pursuit denoising (BPDN), which refers to the solution of

$$\arg \min_x \frac{1}{2} \|y - \Phi x\|_2^2 + \lambda \|x\|_1, \quad (1.7)$$

yields the same result as the constrained problem (1.6) for some choice of the parameter  $\lambda$ , which is unknown *a priori*. The selection of the parameter  $\lambda$  balances the trade-off between sparsity and reconstruction fidelity. Note that as  $\lambda \rightarrow 0$ , the solution of (1.7) behaves like basis pursuit.

### 1.4.2.1 The null-space property

The null space of  $\Phi$  is defined as

$$\mathcal{N}(\Phi) = \{z : \Phi z = 0\}. \quad (1.8)$$

To recover all sparse signals  $x$  from the measurements  $\Phi x$ , then any pair of distinct vectors  $x, x' \in \Sigma_K$  need to satisfy  $\Phi x \neq \Phi x'$ . If  $\Phi x = \Phi x'$  then  $\Phi(x - x') = 0$  with  $x - x' \in \Sigma_{2K}$ , and therefore,  $\Phi$  uniquely represents all  $x \in \Sigma_K$  if and only if  $\mathcal{N}(\Phi)$  contains no vectors in  $\Sigma_{2K}$ . Matrices that satisfy this condition can be characterized by the spark concept. The spark of a matrix is the smallest number of linearly dependent columns. Based on this concept, the following theorem provides guarantees for sparse signal recoverability.

**Theorem 1.** (Corollary 1 of [29]) *For any vector  $y \in \mathbb{R}^M$ , there exists at most one signal  $x \in \Sigma_K$  such that  $y = \Phi x$  if and only if  $\text{spark}(\Phi) > 2K$ .*

Since  $\text{spark}(\Phi) \in [2, M + 1]$ , theorem 1 yields the requirement  $M \geq 2K$ . Theorem 1 provides guarantees in the case of exactly sparse signals, but additional conditions need to be imposed in the case of compressible signals. The null space property (NSP) is applicable to compressible signals and it formally presents the notion that vectors in the null space of  $\Phi$  should not be too concentrated on a small subset of indices.

**Definition 1.** *A matrix  $\Phi$  is said to satisfy the NSP of order  $K$  with constant  $C > 0$  if*

$$\|\eta_T\|_2 \leq C \frac{\|\eta_{T^c}\|_1}{\sqrt{K}} \quad (1.9)$$

for every  $\eta \in \mathcal{N}(\Phi)$  and for all  $T$  such that  $|T| \leq K$ .

Suppose that  $X$  is an  $\ell_p^N$  space. To measure the reconstruction quality of compressible signals  $x \in \mathbb{R}^N$  by using a specific recovery algorithm  $\Delta : \mathbb{R}^M \rightarrow \mathbb{R}^N$ , it is necessary to define reconstruction guarantees of the form

$$\|\Delta(\Phi x) - x\|_X \leq C_0 \sigma_K(x)_X, \quad (1.10)$$

where  $\|\cdot\|_X$  denotes the norm in the  $X$ -space,  $\sigma_K(x)_X$  is as defined in (1.2), and  $C_0$  is a constant independent of  $N$  and  $K$ . Theorem 2 states that the NSP is a necessary and sufficient condition for having reconstruction guarantees of the form (1.10).

**Theorem 2.** (Corollary 3.3 of [23]) *Suppose that  $K > 0$  and integer,  $\Phi$  a sampling matrix, and  $\Delta : \mathbb{R}^M \rightarrow \mathbb{R}^N$  denotes an arbitrary reconstruction algorithm. If  $\Phi$  satisfies the NSP (1.9) in  $X$  of order  $2K$  with constant  $C_0/2$ , then the pair  $(\Phi, \Delta)$  satisfies (1.10) with constant  $C_0$ . Conversely, if  $\Delta$  satisfies (1.10), then  $\Phi$  has the NSP (1.9) in  $X$  of order  $2K$  with constant  $C_0$ .*

#### 1.4.2.2 The restricted isometry property

The fact that the null space property does not account for noise makes it impractical for real scenarios. Instead, the restricted isometry property [15] proposed by Donoho and Tao implies stability under noise.

**Definition 2.** *A matrix  $\Phi$  satisfies the restricted isometry property (RIP) of order  $K$  if there exists a  $\delta_K \in (0, 1)$  such that*

$$(1 - \delta_K)\|x\|_2^2 \leq \|\Phi x\|_2^2 \leq (1 + \delta_K)\|x\|_2^2, \quad (1.11)$$

*holds for all  $x \in \Sigma_K$ .*

A matrix  $\Phi$  having a small restricted isometry constant means that every subset of  $K$  or fewer columns is approximately an orthonormal system. If a matrix  $\Phi$  satisfies the restricted isometry property of order  $2K$ , then  $\Phi$  approximately preserves the distance between any two  $K$ -sparse vectors. It is worth noting that if  $\Phi$  satisfies the RIP of order  $K$  with constant  $\delta_K$ , then  $\Phi$  also satisfies the RIP of order  $K'$  with constant  $\delta_{K'} \leq \delta_K$  for any  $K' < K$ . It was shown in [12] that if  $\Phi$  has restricted isometry constants such that  $\delta_K + \delta_{2K} + \delta_{3K} < 1$ , then the  $\ell_1$  minimization problem (1.4) reconstructs any sparse signal with support size of at least  $K$ . Similar reconstruction guarantees, in terms of the RIP, are obtained for the reconstruction of signals contaminated by noise via the basis pursuit with  $\ell_2$  constraint [10].

It was shown by Candes and Tao [12], as well as Donoho [27], that certain families of random matrices, such as Gaussian or Bernoulli matrices, satisfy the RIP of order  $K$  with  $\delta_K < \delta < 1$  for some value of  $\delta$  independent of  $N$  provided that  $M$  is of the order  $\mathcal{O}(K \log(N/K))$ .

**Theorem 3.** (Theorem 2.12 of [77]) *Let  $\Phi \in \mathbb{R}^{M \times N}$  be a Gaussian matrix having i.i.d. entries of mean 0 and variance  $1/M$  or a Bernoulli matrix having i.i.d. entries equal to  $\pm 1/\sqrt{M}$  with probability  $1/2$ . Let  $\epsilon, \delta \in (0, 1)$  and assume*

$$M \geq C\delta^{-2}(K \log(N/K) + \log(\epsilon^{-1})) \quad (1.12)$$

for  $C > 0$ . Then the RIP constant of  $\Phi$  satisfies  $\delta_K \leq \delta$  with probability at least  $1 - \epsilon$ .

The matrix multiplication operations with Gaussian matrices are computationally expensive for signals of high dimension (e.g. images or videos),  $\mathcal{O}(MN)$ . An alternative that enables efficient sampling is the use of random Fourier ensembles, as the matrix multiplication can be replaced by the Fast Fourier Transform (FFT) Cooley-Tukey algorithm, and therefore, the complexity can be reduced to  $\mathcal{O}(N \log N)$ . In this case, the sampling matrix is given by  $\Phi = SF$ , where  $F \in \mathbb{R}^{N \times N}$  is the discrete Fourier transform on  $\mathbb{R}^N$  and the matrix  $S \in \mathbb{R}^{M \times N}$  picks  $M$  elements of any  $N$ -dimensional vector at random. Deterministic constructions of sampling matrices satisfying the RIP are also possible [25] but they suffer from high lower bounds on the number of measurements with respect to the RIP of order  $K$ .

### 1.4.2.3 Coherence

It is typically difficult to verify if a given matrix satisfies the RIP or the null space property. In fact, it has combinatorial computational complexity. The concept of coherence of a sampling matrix provides a computable guarantee for signal recoverability.

**Definition 3.** The coherence  $\mu(\Phi)$  of a matrix  $\Phi \in \mathbb{R}^{M \times N}$  is the largest absolute normalized inner product between different pair of columns of  $\Phi$

$$\mu(\Phi) = \max_{1 \leq i, j \leq M, i \neq j} \frac{|\phi_i^T \phi_j|}{\|\phi_i\|_2 \|\phi_j\|_2}, \quad (1.13)$$

where  $\phi_i$  is the  $i$ th column of  $\Phi$ .

The coherence is always in the range  $\mu(\Phi) \in \left[ \sqrt{\frac{N-M}{M(N-1)}}, 1 \right]$ . The lower bound is usually known as the Welch bound and is approximately  $\mu(\Phi) \geq 1/\sqrt{M}$  for  $N \gg M$ .

Theorem 4 poses a condition on  $\Phi$ , based on the coherence, that guarantees uniqueness.

**Theorem 4.** (Theorem 12 of [29]) *If*

$$K < \frac{1}{2} \left( 1 + \frac{1}{\mu(\Phi)} \right), \quad (1.14)$$

*then there exists at most one signal  $x \in \Sigma_K$  for each measurement vector  $y \in \mathbb{R}^M$  such that  $y = \Phi x$ .*

Reconstruction guarantees for the BPDN algorithm (1.7) in terms of the coherence are presented in Theorem 5.

**Theorem 5.** (Corolary 1 of [6]) *Assume that  $\Phi$  has coherence  $\mu$  and that  $x \in \Sigma_K$  with  $K \leq 1/(3\mu)$ . Also, assume that measurements  $y$  are of the form  $y = \Phi x + r$ , where the entries of  $r$  are i.i.d.  $N(0, \sigma^2)$ . Let  $\lambda = \sqrt{8\sigma^2(1 + \alpha)\log(N - K)}$  for some small  $\alpha > 0$ . Then with probability greater than*

$$\left( 1 - \frac{1}{(N - K)^\alpha} \right) (1 - \exp(-K/7))$$

*the solution  $\hat{x}$  to 1.7 is unique,  $\text{supp}(\hat{x}) \subset \text{supp}(x)$ , and*

$$\|\hat{x} - x\|_2^2 \leq \left( \sqrt{3} + 3\sqrt{2(1 + \alpha)\log(N - K)} \right)^2 K \sigma^2$$

### 1.4.3 Reconstruction Methods

In this section, we describe some of the main reconstruction algorithms to solve the CS problem presented in Chapter 1.4.2. We focus on two families of recovery algorithms: convex optimization methods and greedy algorithms. Most of the previously proposed CS recovery algorithms for ECG fall into one of these categories.

#### 1.4.3.1 Convex optimization methods

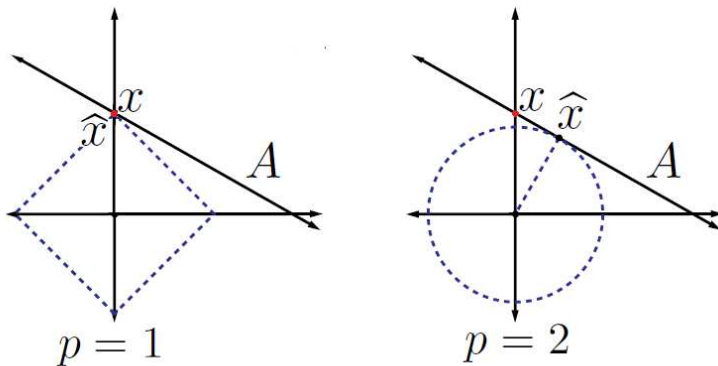
Since CS focuses on the reconstruction of sparse signals, it is natural to try to recover sparse signals by solving an  $\ell_0$  minimization problem

$$\arg \min_x \|x\|_0 \text{ s.t. } y = \Phi x. \quad (1.15)$$

Problem (1.15) is an NP-complete combinatorial optimization problem. It is difficult to solve due to the non-convexity of the  $\ell_0$  norm. Instead, Chen, Donoho and Saunders [22] proposed to solve a different optimization problem, coined Basis Pursuit (1.4), which was introduced in Chapter 1.4.2. Basis pursuit attains the same solution as  $\ell_0$  minimization under certain conditions on the matrix  $\Phi$ , *e.g.*, the null space property, the RIP, or the coherence, which were introduced in Chapter 1.4.2. Problem (1.4) is tractable since it can be posed as a linear program. It is worth noting that (1.4) corresponds to a convex optimization problem and can be seen as a convex relaxation of (1.15). The computational complexity of problem (1.4) is  $O(N^3)$ .

An intuitive understanding that the  $\ell_1$  minimization can provide a sparse solution is provided by Fig. 1.1 For this example, let  $x$  be 1-sparse in the canonical basis of  $\mathbb{R}^2$ . The solid lines  $V_y = \{v \in \mathbb{R}^2 : \Phi v = y\}$ , intersecting the axes, represent the linear constraint and the dotted lines represent the  $\ell_p$  ball with  $p = 1$  and  $p = 2$ . The solutions that minimize the  $\ell_p$  norm,  $p = 1, 2$ , subject to the linear constraint are represented by  $\hat{x}$ . The solution of the  $\ell_1$  minimization lies on one of the axes, while the  $\ell_2$  minimization solution contains non-zero coordinates and does not match the original  $x$ .





**Figure 1.1:** Best approximation of a signal in  $\mathbb{R}^2$  by a one-dimensional subspace using the  $\ell_1$  norm (left) and the  $\ell_2$  norm (right).

### 1.4.3.2 Greedy algorithms

In addition to convex optimization methods, a family of iterative greedy algorithms [8, 59, 85] has received significant attention due to their algorithmic simplicity and low complexity. Greedy algorithms iteratively build up an approximate solution to the  $\ell_0$  minimization problem (1.15) by updating the support set at each iteration. It has been shown that greedy algorithms may outperform  $\ell_1$  minimization in some cases. Three of the main iterative greedy algorithms are Orthogonal Matching Pursuit (OMP) [85], Compressive Sampling Matching Pursuit (CoSaMP) [59], and Iterative Hard Thresholding (IHT) [8].

- **Orthogonal Matching Pursuit:** OMP iteratively tries to estimate the support  $\Gamma$  of the sparse signal  $x$  by starting with  $\Gamma = \emptyset$  and then adding a new element to  $\Gamma$  at each iteration. The criteria for the selection of the new element depends on the residual vector. The residual vector represents the component of the measurement vector  $y$  that cannot be explained by the columns of  $\Phi_\Gamma$ . Let  $\Gamma^j$  and  $\hat{x}^j$  be the current estimate of the support and the reconstructed signal, respectively, at the  $j$ th iteration. Then the residual vector  $r^j$  can be defined as  $r^j = y - \Phi \hat{x}^j$ , where  $\text{supp}(\hat{x}^j) \subseteq \Gamma^j$ . The new element added to  $\Gamma^j$  corresponds to the index of the column of  $\Phi$  that produces the largest inner product with  $r^j$ .

---

**Algorithm 1** Orthogonal Matching Pursuit (OMP)

---

**Require:** Matrix  $\Phi$  and measurements  $y$ .

- 1: Initialize  $j = 0$ ,  $r^0 = y$ ,  $\Gamma^0 = \emptyset$ .
  - 2: **while** halting criterion false **do**
  - 3:    $j \leftarrow j + 1$
  - 4:    $\gamma^j \leftarrow \arg \max_i |\langle \Phi_i, r^j \rangle|$
  - 5:    $\Gamma^j \leftarrow \Gamma^{j-1} \cup \{\gamma^j\}$
  - 6:    $\hat{x}^j|_{\Gamma^j} \leftarrow \Phi_{\Gamma^j}^\dagger y$ ,  $\hat{x}^j|_{\Gamma^{j^c}} \leftarrow 0$
  - 7: **end while**
  - 8: **return**  $\hat{x} \leftarrow \hat{x}^j$
- 

The signal  $\hat{x}^j|_{\Gamma^j}$  is updated by solving a least squares problem  $\hat{x}^j|_{\Gamma^j} = \Phi_{\Gamma^j}^\dagger y$  and the remaining components are set to zero,  $\hat{x}^j|_{(\Gamma^j)^c} = 0$ . Algorithm 1 summarizes the OMP algorithm.

Trop and Hilbert [85] showed that the OMP algorithm is capable of recovering sparse signals with high probability.

**Theorem 6.** (Theorem 2.3 of [59]) *Let  $\Phi \in \mathbb{R}^{M \times N}$  be a sub-Gaussian matrix and  $x \in \mathbb{R}^N$  be a  $K$ -sparse signal. Then the OMP algorithm recovers  $x$  from the measurements  $y = \Phi x$  with high probability, provided that  $M = O(K \log N)$ .*

Even though OMP does not have strong recovery guarantees as compared to  $\ell_1$  minimization, it offers great advantages in terms of speed.

- **Compressive Sampling Matching Pursuit:** CoSaMP is an iterative greedy algorithm that offers the same theoretic performance guarantees as even the best convex optimization approaches [59]. At each iteration, several components of the vector  $x$  are selected based on the largest correlation values between the columns of  $\Phi$  and the residual vector. If they are found sufficiently reliable, their indices are added to the current support estimate of  $x$ . CoSaMP solves the least squares on the current support to obtain a signal approximation. This is repeated until all the recoverable portion of the signal is found. Algorithm 2 describes the CoSaMP method.

Theorem 7 presents the optimal recovery guarantees and running time of CoSaMP.

---

**Algorithm 2** Compressive Sampling Matching Pursuit (CoSaMP)

---

**Require:** Matrix  $y = \Phi x$  and measurements  $y$ .

- 1: Initialize  $j = 0$ ,  $r^0 = y$ ,  $x^0 = 0$ .
  - 2: **while** halting criterion false **do**
  - 3:    $j \leftarrow j + 1$
  - 4:    $e \leftarrow \Phi^T r^{j-1}$
  - 5:    $\Omega \leftarrow \text{supp}(e_{2K})$
  - 6:    $T \leftarrow \Omega \cup \text{supp}(\hat{x}^{j-1})$
  - 7:    $b|_T \leftarrow \Phi_T^\dagger y$ ,  $b|_{T^c} \leftarrow 0$
  - 8:    $\hat{x}^j \leftarrow b_K$
  - 9:    $r^j \leftarrow y - \Phi \hat{x}^j$
  - 10: **end while**
  - 11: **return**  $\hat{x} \leftarrow \hat{x}^j$
- 

**Theorem 7.** (Theorem 4.1 of [59]) *Let  $\Phi \in \mathbb{R}^{M \times N}$  be a matrix satisfying the RIP of order  $2K$  with  $\delta_{2K} < c$ ,  $y = \Phi x + r$  be a noisy observation of an arbitrary signal  $x$ , and  $x_{K/2}$  be the best  $K/2$ -sparse approximation to  $x$ . Given a precision parameter  $\eta$ , the CoSaMP algorithm produces a  $K$ -sparse approximation  $x'$  such that*

$$\|x - x'\|_2 \leq C \max \left\{ \eta, \frac{1}{\sqrt{K}} \|x - x_{K/2}\|_1 + \|r\|_2 \right\}. \quad (1.16)$$

*The running time of CoSaMP is  $O(L \log(\|x\|_2/\eta))$ , where  $L$  bounds the cost of the matrix-vector multiply operation with  $\Phi$ .*

Needell *et al.* [59] showed that only a fixed number of iterations is necessary to reduce the reconstruction error to an optimal amount.

- Iterative Hard Thresholding: IHT is a powerful method for sparse recovery that converges to a local minimum of the problem statement

$$\min_x \|y - \Phi x\|_2^2 \text{ subject to } \|x\|_0 \leq K, \quad (1.17)$$

by using the recursion

$$x^{j+1} = H_K(x^j + \Phi^T(y - \Phi x^j)) \quad (1.18)$$

where  $H_K$  is a non-linear operator that sets to zero all elements other than the  $K$  largest elements (in magnitude). IHT is a very simple algorithm that recovers sparse and compressible vectors with a minimal number of observations and with near optimal accuracy, whenever the matrix  $\Phi$  has a small RIP [8].

Theorem 8 shows that IHT reduces the estimation error at each iteration and is guaranteed to come within a constant factor of the best  $K$ -sparse approximation of the original signal.

**Theorem 8.** (Theorem 4 of [8]) *Let  $\Phi \in \mathbb{R}^{M \times N}$  be a matrix satisfying the RIP of order  $3K$  with  $\delta_{3K} < 1/\sqrt{32}$ ,  $y = \Phi x + r$  be a noisy observation of an arbitrary signal  $x$ , and  $x_K$  be the best  $K$ -sparse approximation to  $x$ . At each iteration  $j$ , the IHT algorithm produces a  $K$ -sparse approximation  $\hat{x}^j$  such that*

$$\|x - \hat{x}^j\|_2 \leq 2^{-j} \|x_K\|_2 + 6\epsilon \quad (1.19)$$

where

$$\epsilon = \|x - x_K\|_2 + \frac{1}{\sqrt{K}} \|x - x_K\|_1 + \|r\|_2. \quad (1.20)$$

#### 1.4.4 The Sparsity Basis

Signals of interest are not necessarily sparse in the canonical basis; however, they may have a concise representation when expressed in a convenient basis. Let  $x \in \mathbb{R}^N$  be expanded in an orthonormal basis  $\Psi = [\Psi_1 \Psi_2 \dots \Psi_N]$  as follows

$$x = \sum_{i=1}^N s_i \psi_i, \quad (1.21)$$

where  $s$  is the  $N \times 1$  column vector of transform coefficients,  $s_i = \langle x, \psi_i \rangle = \psi_i^T x$ .

As described in Chapter 1.4.2, the compressed measurements are of the form  $y = \Phi x$ . By replacing  $x$  from (1.21),  $y$  can be written as

$$y = \Phi \Psi s = \Theta s, \quad (1.22)$$

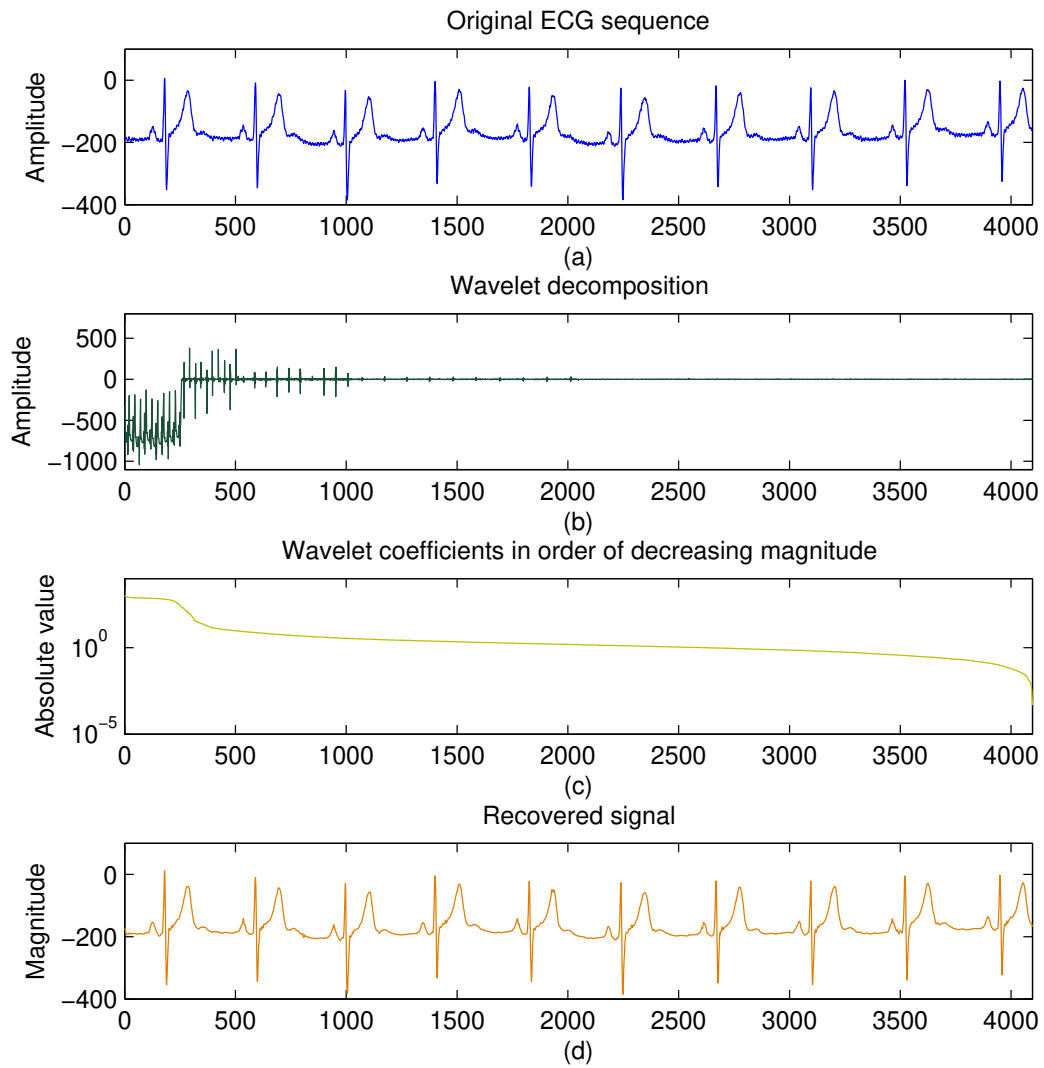
where  $\Theta = \Phi\Psi$ . Therefore, in this case, it is required that the matrix  $\Theta$  satisfies the RIP condition. If  $\Phi$  is a sub-Gaussian matrix and  $\Psi$  is an orthogonal matrix, then the matrix  $\Theta = \Phi\Psi$  will also have a sub-Gaussian distribution. If, in addition, the number of measurements  $M$  is sufficiently high, then the matrix  $\Theta$  also satisfies the RIP with high probability [2].

The signal  $x$  is  $K$ -sparse if only  $K$  of the  $s_i$  coefficients in (1.22) are nonzero, where  $K \ll N$ . This definition is consistent with (1.1) when  $\Psi = I$ . The signal  $x$  is compressible if their coefficients  $s_i$ , sorted in order of decreasing magnitude, exhibit a power-law decay. This is also consistent with the definition of compressible signals presented in Chapter 1.4.2 when  $\Psi = I$ .

Consider the case of ECG signals. It is known that ECG signals are compressible in the wavelet domain. Indeed, state-of-the-art algorithms for ECG compression are wavelet-based methods [41, 54]. Figure 1.2 (a) illustrates an ECG sequence from record 117 of the MIT-BIH Arrhythmia Database. The corresponding wavelet representation using Daubechies-4 and a decomposition level  $L = 4$  is plotted in Fig. 1.2 (b). Most of the wavelet coefficients have very low magnitude, being nearly zero for the highest-frequency wavelet subband. By organizing the wavelet coefficients in order of decreasing absolute value, it is noted in Fig. 1.2 (c) that they decay rapidly. Figure 1.2 (d) illustrates the resulting signal after keeping only the 10% largest (in magnitude) wavelet coefficients and setting the others to zero.

The fact that compressible signals can be well approximated by  $K$ -sparse representations is the foundation of transform coding [55]. A standard transform coding algorithm encodes only the magnitude and location of the  $K$  most significant transform coefficients. Modern lossy coders such as MP3, JPEG, and JPEG 2000 exploit the principles of transform coding [47].

Sparsity plays a prominent role in other applications as well. For signal denoising [28], the small-magnitude transform coefficients are set to zero as they are considered as pure noise. In statistics and learning theory, sparsity is employed to avoid overfitting [87].



**Figure 1.2:** Sparsity of ECG signals in the wavelet domain. (a) Original ECG sequence. (b) Wavelet representation. (c) Sorted wavelet coefficients in order of decreasing absolute value (logarithmic scale). (d) Recovered signal from the 10% largest (in magnitude) wavelet coefficients.

## Chapter 2

# EXPLOITING PRIOR KNOWLEDGE IN COMPRESSED SENSING WIRELESS ECG SYSTEMS

### 2.1 Introduction

Even though previous works validate the potential of compressed sensing (CS) for energy-efficient electrocardiogram (ECG) compression [26, 56, 93], the performance of CS in terms of compression rate and ECG reconstruction quality is still unsatisfactory when compared to the results attained by state-of-the-art algorithms based on exploiting magnitude correlation across wavelet subbands [41, 54]. Most previous CS-based ECG compression works exploit only the sparsity of the signal, ignoring important signal structure information that can be known *a priori* and lead to enhanced reconstruction results.

This chapter describes the application of a CS algorithm that enables exploitation of ECG signal structure in the reconstruction process. Two key signal structure properties are incorporated into the proposed algorithm. The first property captures the wavelet domain dependencies across subbands and the second exploits the significant fraction of common wavelet coefficient support for consecutive ECG segments. The proposed algorithm falls within the framework of model-based CS [3]—a new CS framework based on unions of subspaces—that can enhance signal reconstruction while reducing the number of measurements. The motivation for using a model-based approach is that it enables the incorporation of structural dependencies between the locations of the signal coefficients caused by R wave events. However, it is worth mentioning that the proposed algorithm differs from traditional model-based recovery algorithms [3, 57] in two respects. First, it uses prior support estimate knowledge to

leverage the small variation in the support set of adjacent data sequences. Second, it excludes the selection of coefficients from the lowest-energy wavelet subband.

The performance of the proposed method is evaluated using the MIT-BIH Arrhythmia Database [39]. Given that most CS ECG application works use the basis pursuit denoising (BPDN) algorithm [26, 56], system-level comparisons are provided based on implementations of the BPDN and proposed reconstruction algorithms. Simulations are also performed with the bound-optimization-based BSBL algorithm, previously employed by Zhilin *et al.* [93] for non-invasive fetal ECG, in order to compare with other structured sparsity-based CS reconstruction algorithms. Results indicate that the proposed algorithms outperform both BPDN and bound-optimization-based BSBL in terms of compression rate and reconstruction quality.

The organization of this chapter is as follows. Section 2.2 presents a brief review of model-based CS [3]. Section 2.3 describes the connected subtree structure encountered in the wavelet representation of the ECG and studies the support variation across consecutive ECG segments. In Section 2.4, the proposed method is presented. Numerical results for the proposed method and comparisons with a benchmark state-of-the-art algorithm for ECG compression, SPIHT, are presented in 2.5. Finally, we conclude in Section 2.6 with closing remarks.

## 2.2 Background

### 2.2.1 Brief Review of Model-based Compressed Sensing

Model-based compressive sensing is a new paradigm that aims to capture the inter-dependency structure in the support of the large signal coefficients using a union-of-subspaces model [3]. This model decreases the degrees of freedom of the signal by allowing only some configurations of support for the largest coefficients.

Let  $x$  be a  $K$ -sparse signal. Then,  $x$  lies in  $\Sigma_K \subset \mathbb{R}^N$ , which is a union of  $\binom{N}{K}$  subspaces of dimension  $K$ . A union-of-subspaces model allows only some  $K$ -dimensional subspaces in  $\Sigma_K$  and leads to representations that incorporate signal structure. Let  $x|_{\Omega}$  denote the coordinates of  $x$  corresponding to the set of indices  $\Omega \subseteq 1, \dots, N$ , and



let  $\Omega^c$  represent the complement of  $\Omega$ . Then, the signal model  $\mathcal{M}_K$  is defined as the union of  $m_K$  canonical  $K$ -dimensional subspaces

$$\mathcal{M}_K = \bigcup_{m=1}^{m_K} \mathcal{X}_m, \mathcal{X}_m = \{x : x|_{\Omega_m} \in \mathbb{R}^K, x|_{\Omega_m^c} = 0\} \quad (2.1)$$

where  $\{\Omega_1, \dots, \Omega_{m_K}\}$  is the set of all allowed supports with  $|\Omega_m| = K$  for each  $m = 1 \dots m_K$ . It should be noted that  $\mathcal{M}_K \subseteq \Sigma_k$  and that  $\mathcal{M}_K$  contains  $m_K \leq \binom{N}{K}$  subspaces. Signals from  $\mathcal{M}_K$  are called  $K$ -model sparse.

A similar treatment is applied to compressible signals. A compressible signal  $x \in \mathbb{R}^N$  that is nearly  $K$ -model sparse can be approximated to the best model-based approximation in  $\mathcal{M}_K$ . The  $\ell_2$  error produced by the approximation is given by

$$\sigma_{\mathcal{M}_K} = \inf_{\bar{x} \in \mathcal{M}_K} \|x - \bar{x}\|_2. \quad (2.2)$$

The algorithm that provides the best  $K$ -term approximation of the signal  $x$  under the model  $\mathcal{M}_K$  is denoted as  $\mathbb{M}(x, K)$ . Thus, the error  $\sigma_{\mathcal{M}_K}$  can also be written as  $\sigma_{\mathcal{M}_K} = \|x - \mathbb{M}(x, K)\|_2$ . A sparsity model  $\mathcal{M} = \{\mathcal{M}_1, \mathcal{M}_2, \dots\}$  produces nested approximations if the support of  $\mathbb{M}(x, K^*)$  contains the support of  $\mathbb{M}(x, K)$  for all  $K < K^*$ . If the signal model produces nested approximations, then the support of the difference vector  $\mathbb{M}(x, jK) - \mathbb{M}(x, (j-1)K)$  lies in a small union of subspaces. These difference vectors form sets of residual subspaces. The  $j^{\text{th}}$  set of residual subspaces of size  $K$  is defined as

$$\mathcal{R}_{j,K}(\mathcal{M}) = \{u \in \mathbb{R}^N \text{ s.t. } u = \mathbb{M}(x, jK) - \mathbb{M}(x, (j-1)K) \text{ for some } x \in \mathbb{R}^N\}, \quad (2.3)$$

for  $j = 1, \dots, \lceil N/K \rceil$ . A structured compressible signal  $x$  can be robustly recovered from the compressive measurements  $y = \Phi x$  if the matrix  $\Phi$  satisfies the restricted amplification property (RAmP) [3]. A matrix  $\Phi$  has the  $(\epsilon_K, r)$ -restricted amplification

property for the residual subspaces  $R_{j,K}$  of model  $\mathcal{M}$  if

$$\|\Phi u\|_2^2 \leq (1 + \epsilon_K) j^{2r} \|u\|_2^2 \quad (2.4)$$

for any  $u \in R_{j,K}$  and for each  $1 \leq j \leq \lceil N/K \rceil$ .

Baraniuk *et al.* [3] incorporated the union-of-subspaces models into two well-known CS recovery algorithms, CoSaMP and IHT, through a single modification in the algorithms. The modification, in practice, replaces the best  $K$ -term approximation with a best  $K$ -term model-based approximation.

### 2.2.2 Tree-structured Sparsity Models

One example of a structured sparsity model is that encountered in signals whose most significant wavelet coefficients are organized into a tree structure, and where the largest coefficients cluster along the branches of the tree [3].

Consider an  $N$ -dimensional signal  $x$ . Given a wavelet function  $\psi$  and a scaling function  $\varphi$ , the wavelet representation of  $x$  is defined in terms of shifted versions of  $\varphi$  and shifted and dilated versions of  $\psi$

$$x = \sum_{i=0}^{N_L-1} a_{L,i} \varphi_{L,i} + \sum_{j=1}^L \sum_{i=0}^{N_j-1} d_{j,i} \psi_{j,i}, \quad (2.5)$$

where  $j$  denotes the scale of analysis and  $L$  denotes the coarsest scale.  $N_j = N/2^j$  indicates the number of coefficients at scale  $j \in \{1, \dots, L\}$  and  $i$  represents the position,  $0 \leq i \leq N_j - 1$ . The wavelet transform consists of the scaling coefficients  $a_{L,i}$  and wavelet coefficients  $d_{j,i}$ . Using the previous notation, we write  $x = \Psi s$ , where  $\Psi$  is the orthogonal matrix containing the wavelet and scaling functions as columns and  $s = [d_{1,0} \dots d_{1,N_1-1} \dots d_{L,0} \dots d_{L,N_L-1} a_{L,1} \dots a_{L,N_L-1}]^T$  is the vector of scaling and wavelet coefficients. The vector  $s$  can be decomposed into  $L + 1$  subvectors. The first  $L$  subvectors are denoted by  $d_j$ ,  $j = 1, \dots, L$ , and the  $j$ th subvector contains all of the wavelet coefficients for scale  $j$ . The last subvector corresponds to the scaling coefficients

and is denoted as  $a_L$ . Thus,  $s$  can also be written as  $s = [d_1 d_2 \dots d_L a_L]^T$ .

The wavelet atoms form a binary tree structure where the wavelet coefficient  $d_{j,i}$  is the parent of its two children  $d_{j+1,2i}$  and  $d_{j+1,2i+1}$ . This nesting property causes rapid transitions and other singularities to manifest as chains of large coefficients along the branches of the wavelet tree [24]. This gives rise to the concept of a connected subtree, which refers to a connected set of nodes  $\Omega$  meeting the condition that whenever a coefficient  $d_{j,i} \in \Omega$ , then its parent also belongs to  $\Omega$ . In [3], Baraniuk *et al.* assumed  $N_L = 1$  for simplicity and defined the structured sparsity model  $\mathcal{T}_K$  as the union of all  $K$ -dimensional subspaces corresponding to supports  $\Omega$  that form connected subtrees,

$$\mathcal{T}_K = \left\{ x = a_{L,0} \varphi_{L,0} + \sum_{j=1}^L \sum_{i=0}^{N_j-1} d_{j,i} \psi_{j,i} : d|_{\Omega^c} = 0, |\Omega| = K, \right. \\ \left. \Omega \text{ forms a connected subtree} \right\}, \quad (2.6)$$

where  $\Omega^c$  denotes the complement of the set  $\Omega$ . To find the best  $K$ -term tree-based approximation, Baraniuk *et al.* used the condensing sort and select algorithm (CSSA) [4], which solves for

$$x^* = \arg \min_{\bar{x} \in \mathcal{T}_k} \|x - \bar{x}\|_2 \quad (2.7)$$

by using a two-stage process. The first stage merges the non-monotonic segments of the tree branches with an iterative sort-and-average routine. The second stage simply sorts the wavelet coefficients once they are organized in a monotonic non-increasing sequence along the branches out from the root.

### 2.3 Motivation

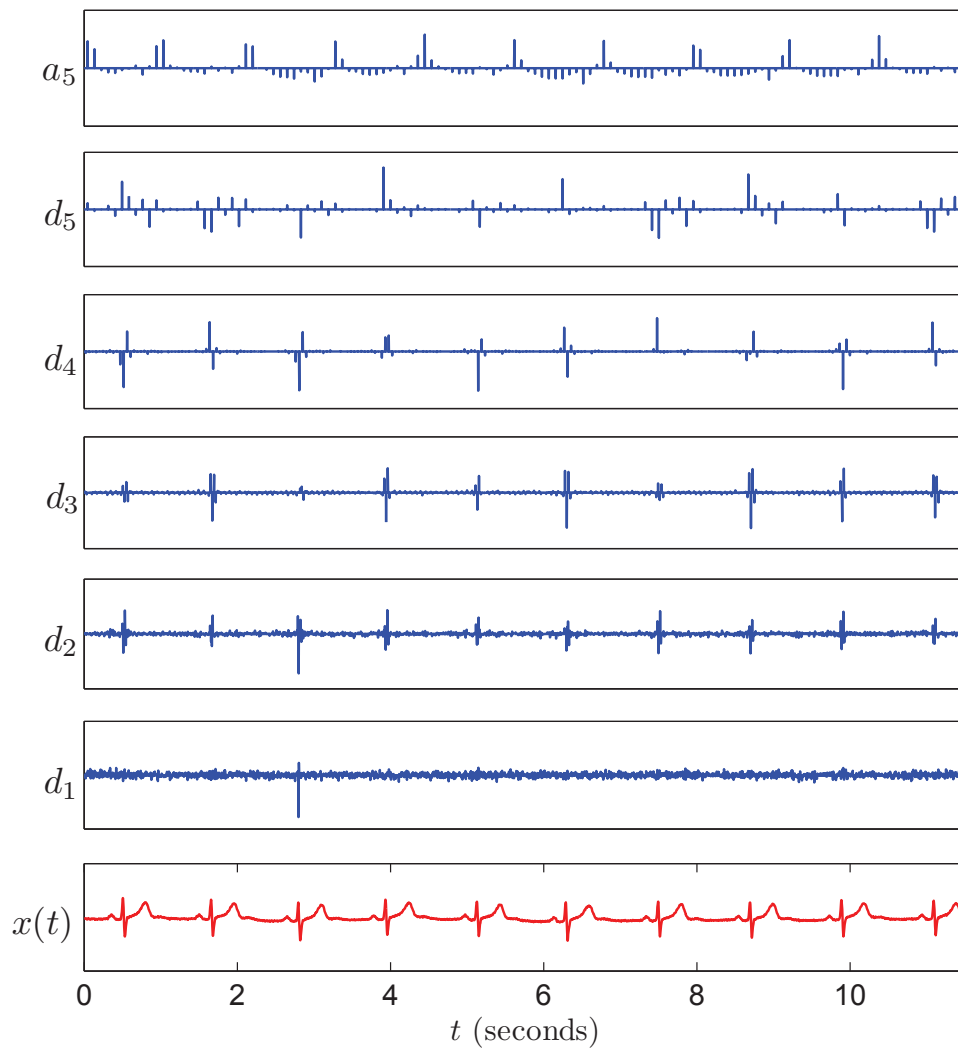
In this section, we analyze the structure of the wavelet representation of ECG signals to motivate the incorporation of prior information in CS-based recovery algorithms for ECG reconstruction. We concentrate on exploiting two key properties of the ECG wavelet coefficients. The first property is the connected subtree structure formed

by the largest (in magnitude) wavelet coefficients, and the second property is the high fraction of common support between adjacent ECG segments.

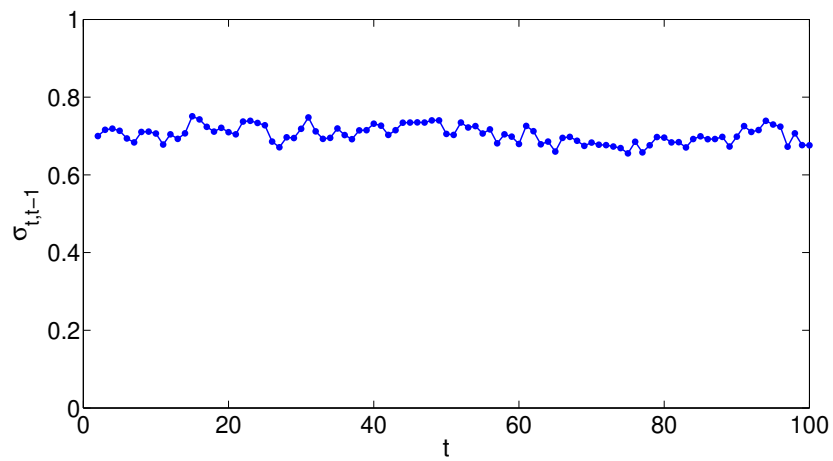
### 2.3.1 Connected Subtree Structure of ECG Wavelet Coefficients

Sharp transition regions in the time domain generate large magnitude wavelet coefficients that persist along the branches of the wavelet tree, forming connected rooted subtrees [24]. This behavior is also present in the wavelet representation of ECG signals and is connected with R wave events. This idea is illustrated in Fig. 2.1, which shows that coefficients subvectors  $d_j$ ,  $j = 1, \dots, 5$ , and  $a_5$ , corresponding to the Daubechies-4 wavelet transform of an ECG time series using a decomposition level  $L = 5$ . The subvectors are plotted as rows stacked on top of each other. The ECG time series, located at the bottom of Fig. 2.1, corresponds to an extract of 11.5 seconds from record 117 of the MIT-BIH Arrhythmia Database. Each coefficient vector is time-shifted so that the tree structure can be clearly identified. For visualization purposes, the magnitude of the coefficients is normalized so that the Euclidean norm of each wavelet subband is unity.

Examining the wavelet representation in Fig. 2.1, it is noticed that the large coefficients are aligned and propagate across scales, forming a connected tree structure. This persistence property is mainly noticed in subbands  $d_4$ ,  $d_3$ , and  $d_2$ . These results suggest that the tree-structured sparsity model described in Section 2.2.2 is an appropriate model to represent the support of the most significant ECG wavelet coefficients. The tree structure is intrinsically related to the shape of the ECG cycles. When compared with the ECG time series, it is noted that the large coefficients are connected with the QRS complexes, which can be regarded as sharp transition regions of the signal. We propose to exploit the tree-structured sparse representation as additional prior information for the CS reconstruction of ECG signals. As stated by the model-based CS framework [3], knowledge of the signal support potentially leads to high quality reconstruction using fewer measurements, and thus higher compression performance.



**Figure 2.1:** Daubechies-4 coefficients subvectors  $d_j$ ,  $j = 1 \dots 5$ , and  $a_5$  for ECG time series.



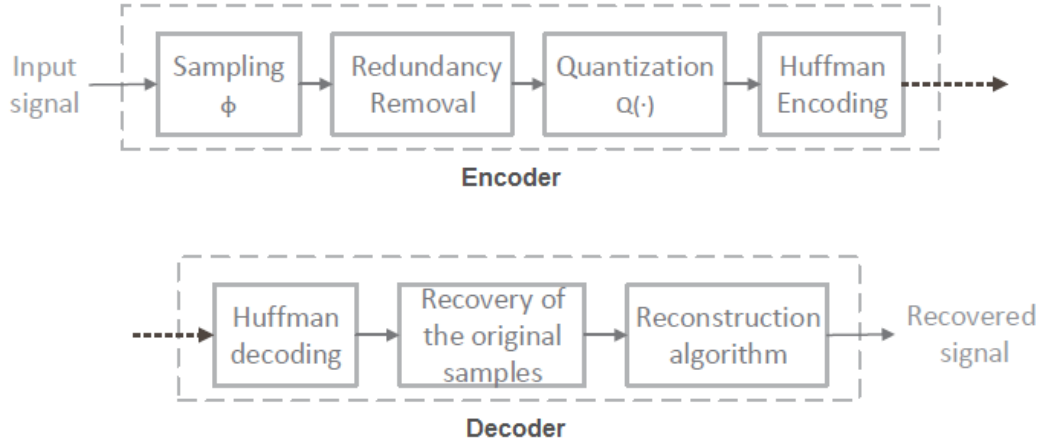
**Figure 2.2:** Fraction of common support between consecutive ECG segments. The results are averaged over the selected data set from the MIT-BIH Arrhythmia Database.

### 2.3.2 Common Support Between Consecutive ECG Segments

In this section, we study how the support of two consecutive ECG segments varies. Consider the 10-min long single leads extracted from records 100, 101, 102, 107, 109, 111, 115, 117, 118 and 119 in the MIT-BIH Arrhythmia Database, which are sampled at 360 Hz, and form 100 consecutive sequences of 2048 samples for each record. Let  $\Omega_t$  denote the support of the  $K = 225$  largest (in magnitude) wavelet coefficients of sequence  $s_t$ , where  $t = 1, \dots, 100$  denotes the order of the sequences. Consider the fraction of common support between two consecutive sequences, denoted as  $\sigma_{t,t-1}$ , which is defined as

$$\sigma_{t,t-1} = \frac{|\Omega_t \cap \Omega_{t-1}|}{|\Omega_t|}, \quad t = 2, \dots, 100. \quad (2.8)$$

The averaged results over the entire ECG data set are illustrated in Fig. 2.2, where the data points are connected for visualization purposes only. The fraction of common support is high, always greater than 0.65, which indicates that we can use the support of the previous data sequence to improve the estimate of the support of the current data sequence. In designing the CS-based reconstruction algorithm, we propose to include support information of previous reconstructed data sequences to improve the performance. A related work by Wang *et al.* [91] shows how to iteratively enhance recovery of a signal by solving basis pursuit in the first iteration to obtain a support estimate, solving the problem of basis pursuit with partially known support [88] with this support estimate, and repeating the steps. Their work attains exact reconstruction guarantees for a single iteration even if the initial support estimate includes a small fraction of indices outside the true support.



**Figure 2.3:** Block Diagram of the proposed ECG compression method.

## 2.4 Methods

The proposed ECG compression method based on compressive sensing is described in this section. The method is summarized in Fig. 2.3. The process initiates with the sampling of the ECG signals through linear random measurements, followed by a redundancy removal module and a quantization stage. The quantized samples are entropy coded and transmitted to a remote terminal where the reconstruction is performed. The presented contribution focuses on the reconstruction algorithm that relies on prior knowledge of ECG wavelet coefficient structure. A more detailed explanation of each stage is given below.

### 2.4.1 Sampling and Encoding

The first step addresses the sampling of consecutive ECG segments of length  $N$ . Let  $x$  denote an ECG segment. The information we acquire about  $x$  can be described by  $y = \Phi x$ , where  $\Phi$  is a  $M \times N$  matrix. In order to recover the best  $K$ -term model-based approximation of the original signal, the matrix  $\Phi$  needs to satisfy the restricted amplification property (RAM-P). It is known that sub-Gaussian matrices meet this condition with high probability [3]. Here we build the entries of the matrix  $\Phi$  by independently



sampling from a symmetric Bernoulli distribution ( $P(\Phi_{i,j} = \pm 1/\sqrt{M}) = 1/2$ ) in order to facilitate an efficient hardware implementation. The use of Bernoulli matrices, as compared to other sub-Gaussian matrices, results in simpler circuit complexity, data storage, and computation requirements [21].

To realize the analog hardware implementation of CS matrices, Mamaghanian *et al.* [58] recently proposed the spread spectrum random modulator pre-integrator. This architecture starts with a pre-modulation block, followed by a random demodulation pre-integrator architecture, which is composed of parallel channels of random demodulators. Each random demodulator is, in turn, composed of three stages. The first stage refers to the multiplication of the input signal with a pseudo-random sequence that takes values  $\pm 1$  with equal probability. The second stage incorporates a low-pass filter to avoid aliasing and the final stage corresponds to a standard ADC.

The use of the described sampling procedure results in similar adjacent measurement vectors. This is caused by the quasi-periodic nature of the ECG signal and the use of a fixed sampling matrix. To further improve the compression and reduce the amount of redundant information, we implemented the same redundancy removal and entropy coding stages proposed by Mamaghanian *et al.* [56]. The redundancy removal stage computes the difference between two consecutive measurement vectors and only transmits this difference to the quantization module. In [56], it was shown that the variance of the difference signals between consecutive measurement vectors is lower than the variance of the original measurement vectors, which leads to a reduction in the number of bits for signal representation, from 12 to 9 bits. An 8-bit optimal scalar quantizer designed with the Lloyd-Max algorithm is utilized [38] and entropy coding stage uses Huffman coding to further increase the compression ratio.

#### 2.4.2 Reconstruction Algorithm using Prior Information

The original samples must be recovered from the transmitted difference between consecutive vectors. Therefore, at the remote terminal, Huffman decoding followed by recovery of the original samples is employed. For reconstruction of the ECG signal,

two iterative algorithms that are easily modified to incorporate prior ECG wavelet representation structure information are proposed. The algorithms are CoSaMP and IHT, which were previously modified by Baraniuk *et al.* [3] to incorporate structured sparsity models and which are described in more detail in Section 1.4.3.2. Their modification results in replacing the nonlinear sparse approximation step with a structured sparse approximation. These algorithms are known in the literature as model-based CoSaMP and model-based IHT; both have provable robust guarantees. One of the main properties of model-based CoSaMP is that robust signal recovery requires only a number of measurements that is proportional to the sparsity level of the signal.

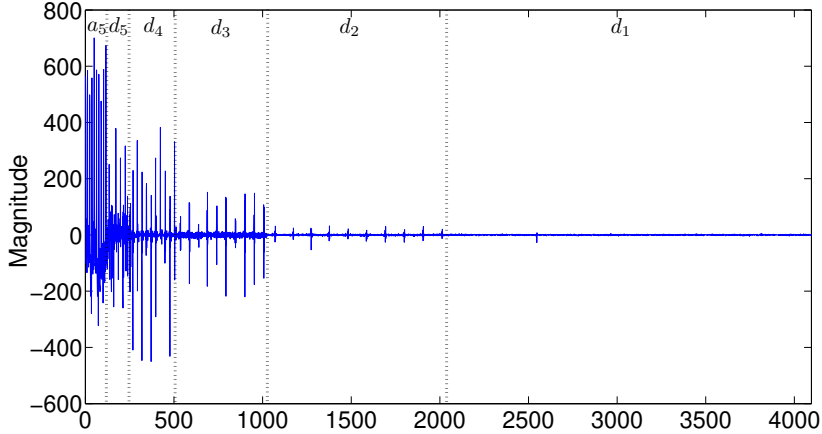
The sparsity model employed in this work corresponds to a modified version of the tree-structured sparsity model described in Section 2.2.2. The modification is based on the fact that ECG wavelet coefficients at scale  $j = 1$  correspond to coefficients with nearly zero magnitude, as illustrated in Fig. 2.4, and should therefore not be included in the best  $K$ -term model-based signal approximation. Thus, we select  $N_L$  scaling coefficients and define  $\mathcal{T}_K^*$  as

$$\mathcal{T}_K^* = \left\{ x = \sum_{i=0}^{N_L-1} a_{L,i} \varphi_{L,i} + \sum_{j=1}^L \sum_{i=0}^{N_j-1} d_{j,i} \psi_{j,i} : \text{supp}(d_{1,i}) \in \Omega^c \text{ for all } i, d|_{\Omega^c} = 0, \right. \\ \left. |\Omega| = K, \Omega \text{ forms a connected subtree} \right\}. \quad (2.9)$$

Denote the algorithm that finds the best  $K$ -term tree-based approximation by  $\mathbb{T}(x, K)$ . That is,

$$\mathbb{T}(x, K) = \arg \min_{\bar{x} \in \mathcal{T}_K^*} \|x - \bar{x}\|_2. \quad (2.10)$$

The model-based CoSaMP and the model-based IHT, using the tree-structured sparsity model  $\mathcal{T}_K^*$ , are summarized in Algorithms 3 and 4, respectively, and are referred to as modified model-based CoSaMP (MMB-CoSaMP) and modified model-based IHT (MMB-IHT). The halting criterion for both algorithms can be a fixed number of iterations or a bound on the residual norm,  $\|r_j\|_2 \leq \epsilon$ , for some predetermined  $\epsilon > 0$ . Note



**Figure 2.4:** Wavelet decomposition of level  $L = 5$  for ECG time series using Daubechies-4.

that in the algorithm tables, the Moore-Penrose pseudo-inverse of  $\Theta$  is denoted by  $\Theta^\dagger$ ,  $\Theta_T$  denotes the submatrix obtained by extracting the columns of  $\Theta$  corresponding to the indexes in  $T$ ,  $b|_T$  represents the entries of  $b$  corresponding to the set of indices  $T$ , and the support of  $s$  is denoted by  $\text{supp}(s)$ . Also, we use the condensing sort and select algorithm [4] to solve for the optimization problem in (2.10).

As in model-based CoSaMP, MMB-CoSaMP starts the iteration by calculating the correlation values between the columns of  $\Theta$  and the residual of the previous iteration  $r_{j-1}$  (step 4). The correlation values are used to find support of the best  $2K$ -term tree-based approximation (step 5), which is subsequently merged with the support of the previous iteration (step 6). The next step refers to solving a least-squares problem to approximate the target signal on the updated support (step 7). To enforce a sparse solution, the algorithm finds the best  $K$ -term tree-based approximation of the least squares solution (step 8) and finalizes with the update of the residual (step 9).

Similarly, MMM-IHT also resembles model-based IHT. The algorithm iteratively solves (1.17) by moving in the opposite direction of the gradient of  $\|y - \Theta s\|_2^2$  at each iteration (step 4). MMM-IHT enforces a sparse solution by selecting the best  $K$ -term tree-based approximation of  $b$  (step 5) and finalizes with an update of the residual

(step 6). It is worth emphasizing that the proposed algorithms exploit the wavelet tree structure in the steps that refer to the best  $K$ -term tree-based approximation defined in (2.10). Both MMB-IHT and MMB-CoSaMP differ from the traditional model-based approaches in the initialization step and in the  $K$ -term tree-based approximation of the signal.

Unlike the traditional model-based approaches, the proposed algorithms always include the scaling coefficients in the  $K$ -term tree-based approximation of the signal, which is of great importance given that the scaling coefficients accumulate most of the ECG signal energy. The initialization step is a modification with respect to traditional approaches because it incorporates support information from the previously reconstructed signal. In Section 2.3.2, it is noted that a high fraction of support is shared between two consecutive ECG segments. The two consecutive ECG segments in the wavelet domain are denoted by  $s^t$  and  $s^{t-1}$ . With the aim of exploiting this information, the first signal estimate is determined by solving a least squares problem using the support of the previously reconstructed ECG segment,  $\Lambda = \text{supp}(s^{t-1})$ . The contribution of this estimate is subtracted from the measurement vector and iterated on the residual. These steps are implemented in the initialization of Algorithms 3 and 4. In both algorithms, MMB-IHT and MMB-CoSaMP, the support set is refined through iterations by addition of promising new atoms and deletion of unnecessary atoms. In this way, we expect the algorithms to preserve the common support ( $\text{supp}(s^t) \cap \text{supp}(s^{t-1})$ ) and to replace the unnecessary atoms ( $\text{supp}(s^{t-1}) \setminus \text{supp}(s^t)$ ) with the innovation support of the current ECG sequence, *i.e.* ( $\text{supp}(s^t) \setminus \text{supp}(s^{t-1})$ ).

## 2.5 Experimental Results

To validate the proposed methods, compressed measurements are generated using a set of records from the MIT-BIH Arrhythmia Database [39] as the original signals. Every file in the database contains two lead recordings sampled at 360 Hz with 11 bits per sample of resolution. However, since body area networks adopt lower sampling frequencies than 360 Hz, each ECG recording is resampled at 250 Hz. The

sampling frequency of 250 Hz is commonly used for ECG monitoring in body area networks [62]. Experiments are carried out and averaged over 10-min long single leads extracted from records 100, 101, 102, 103, 107, 109, 111, 115, 117, 118 and 119. This data set was proposed in [54]; it consists of a variety of signals with different rhythms, wave morphologies and abnormal heartbeats. Results are presented for averages of 100 repetitions of each experiment, with a different realization of the random measurement matrix at each time.

### 2.5.1 Performance Evaluation

The compression ratio (CR), the percentage root-mean-square difference (PRD), the normalized version of PRD (PRDN), the quality score (QS), and the reconstruction SNR (R-SNR) are used as performance measures. The CR is defined as the number of bits required for the original signal over the number of bits required for the compressed signal. Here the original signals refer to the resampled ECG records at 250 Hz. The reconstruction SNR is defined as

$$\text{R-SNR} = 10 \log_{10} \frac{\|x\|_2^2}{\|x - \hat{x}\|_2^2}, \quad (2.11)$$

where  $x$  and  $\hat{x}$  denote the  $N$ -dimensional original and reconstructed signals, respectively. The PRD is defined as  $\text{PRD} = (\|x - \hat{x}\|_2 / \|x\|_2) \times 100$ . Let  $e$  denote an  $N$ -dimensional vector of ones and  $\bar{x}$  be the mean value of  $x$ . The PRDN is defined as  $\text{PRDN} = (\|x - \hat{x}\|_2 / \|x - \bar{x}e\|_2) \times 100$ , and the QS as  $\text{QS} = \text{CR} / \text{PRD}$  [36].

### 2.5.2 Practical Considerations

The length of the ECG segments is set to  $N = 256$ , so that the acquisition time can be sufficiently short (approximately 1 sec at the rate of 250 Hz) for real-time monitoring. The orthogonal Daubechies-4 wavelets is chosen as the sparsifying transform. For the reconstruction algorithms, the halting criterion is either a maximum number of iterations (we selected 70 for our simulations) or a bound on the residual norm,  $\|r_t\|_2 \leq \epsilon$ . We selected  $\epsilon = 1 \times 10^{-3} \|y\|_2$ .

---

**Algorithm 3** MMB-CoSaMP

---

**Require:** Matrices  $\Theta = \Phi\Psi$  and  $\Psi$ , measurements  $y$ , sparsity level  $K$ , support of previously reconstructed ECG segment  $\Lambda$ , structured sparse approximation algorithm  $\mathbb{T}$ .

- 1: Initialize  $\hat{s}_0|_\Lambda = \Theta_\Lambda^\dagger y$ ,  $\hat{s}_0|_{\Lambda^c} = 0$ ,  $r_0 = y - \Theta\hat{s}_0$ ,  $j = 0$ .
  - 2: **while** halting criterion false **do**
  - 3:    $j \leftarrow j + 1$
  - 4:    $e \leftarrow \Theta^T r_{j-1}$
  - 5:    $\Omega \leftarrow \text{supp}(\Psi^T \mathbb{T}(\Psi e, 2K))$
  - 6:    $T \leftarrow \Omega \cup \text{supp}(\hat{s}_{j-1})$
  - 7:    $b|_T \leftarrow \Theta_T^\dagger y$ ,  $b|_{T^c} \leftarrow 0$
  - 8:    $\hat{s}_j \leftarrow \Psi^T \mathbb{T}(\Psi b, K)$
  - 9:    $r_j \leftarrow y - \Theta\hat{s}_j$
  - 10: **end while**
  - 11: **return**  $\hat{x} \leftarrow \Psi\hat{s}_j$
- 

For the sparsity level, a sequence of residual energy is defined. The elements in the vector of wavelet coefficients  $s$  are ordered according to their square magnitudes, such that

$$|s_{(1)}|^2 \geq |s_{(2)}|^2 \geq \dots \geq |s_{(N-1)}|^2 \geq |s_{(N)}|^2. \quad (2.12)$$

The sequence of residual energy is defined as

$$C_K = \frac{\sum_{j=1}^N |s_{(j)}|^2 - \sum_{j=1}^K |s_{(j)}|^2}{\sum_{j=1}^N |s_{(j)}|^2}, \quad K = 1, \dots, N. \quad (2.13)$$

From each record of the data set, 300 ECG segments of length  $N = 256$  are selected. The sequence of residual energy is averaged over all the selected ECG segments and over all the different records. The results are plotted in Fig. 2.5 in logarithmic scale. The sparsity level is selected as the value of  $K$  that satisfies averaged  $C_K = 0.001$ , which corresponds to  $K = 34$  and is indicated in Fig. 2.5 with dotted lines. This result indicates that the most significant 34 wavelet coefficients approximately accumulate 99.999% of the total signal energy.

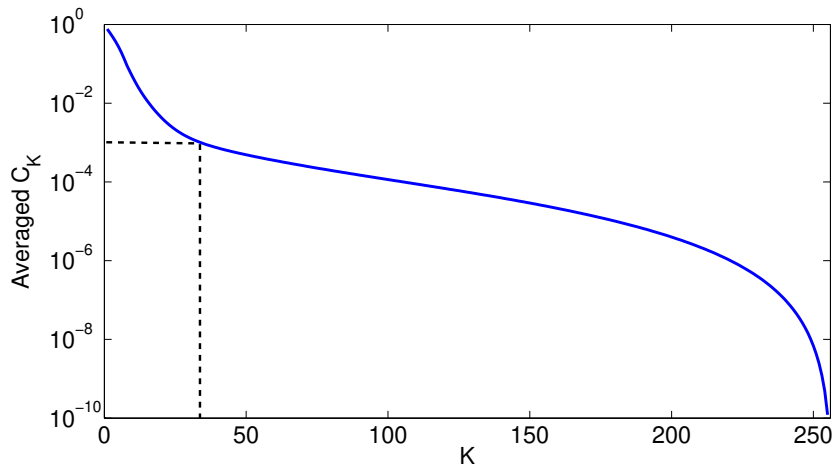
---

**Algorithm 4** MMB-IHT

---

**Require:** Matrices  $\Psi$ ,  $\Phi$ , and  $\Theta = \Phi\Psi$ , measurements  $y$ , sparsity level  $K$ , support of previously reconstructed ECG segment  $\Lambda$ , structured sparse approximation algorithm  $\mathbb{T}$ .

- 1: Initialize  $\hat{s}_0|_{\Lambda} = \Theta_{\Lambda}^{\dagger}y$ ,  $\hat{s}_0|_{\Lambda^c} = 0$ ,  $\hat{x}_0 = \Psi\hat{s}_0$ ,  $r_0 = y - \Phi\hat{x}_0$ ,  $j = 0$
  - 2: **while** halting criterion false **do**
  - 3:    $j \leftarrow j + 1$
  - 4:    $b \leftarrow \hat{x}_{j-1} + \Phi^T r_{j-1}$
  - 5:    $\hat{x}_j \leftarrow \mathbb{T}(b, K)$
  - 6:    $r_j \leftarrow y - \Phi\hat{x}_j$
  - 7: **end while**
  - 8: **return**  $\hat{x} \leftarrow \hat{x}_j$
- 



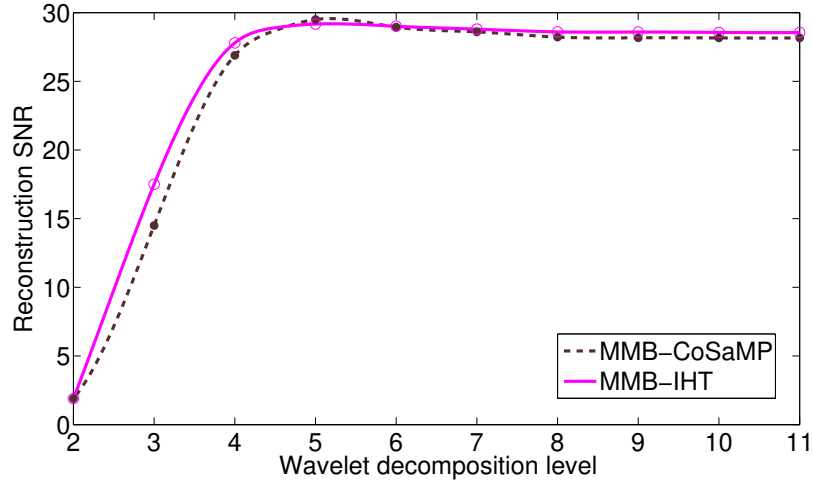
**Figure 2.5:** Sequence of residual energy averaged over the selected set of records from the MIT-BIH Arrhythmia Database.

Given that the ECG recordings are sampled at 250Hz, the sampling interval becomes  $\Delta t = 1/250$  seconds. A decomposition level  $L = 5$  is utilized so that the scaling coefficients vector  $a_5$ , associated with a physical scale of  $2^5\Delta t = 2^5/250 = 0.13$  seconds, approximately isolate the T waves, and thus the detail coefficients capture the QRS complex. In this way, the largest (in magnitude) detail coefficients are expected to exhibit a connected subtree structure caused by the QRS complexes, as shown in Fig. 2.1. The ECG reconstruction quality, using both MMB-CoSaMP and MMB-IHT algorithms, is evaluated as a function of the wavelet decomposition level. The results

are averaged over the selected set of records for a number of measurements  $m = 3K$ . The results in Fig. 2.6 indicate that the selection of  $L = 5$  is indeed a good choice for the decomposition level. The data points are connected for visualization purposes only.

Sparse binary matrices have recently been proposed for CS ECG since they lead to fast computations and low-memory requirements [56]. An experiment is designed to test the performance of the proposed algorithms when using sparse binary matrices. Similarly to the work in [56], two types of sparse binary matrices, matrix I and matrix II, are employed. Matrix I has only  $q \ll N$  nonzero elements in each column. Each nonzero element takes the value  $1/\sqrt{q}$  and its location is chosen randomly. Matrix II has only  $q \ll N$  nonzero elements in each column. Each nonzero element takes the value  $\pm 1/\sqrt{q}$  with equal probability and its location is chosen randomly. In [56], the value of  $q = \lfloor 0.025 \times N \rfloor$  provides a good tradeoff between execution time and reconstruction quality. The same relation between the value of  $q$  and  $N$  is selected for the proposed experiment,  $q = \lfloor 0.025 \times N \rfloor = 6$ . The reconstruction performance of the proposed algorithms using the Bernoulli matrix and the sparse binary matrices is shown in Table I. The results correspond to averaged PRD values over the entire set of selected records. Even though all the matrices have a similar performance for  $CR \leq 6$ , the performance of the sparse binary matrices deteriorates for  $CR > 6$ . Hence, the dense Bernoulli matrix is selected as the sampling matrix for the proposed methods. Nevertheless, the minimal detriment to performance suggests that sparse binary sensing matrices are a plausible alternative for CS ECG.





**Figure 2.6:** Reconstruction SNR as a function of the wavelet decomposition level. Number of measurements  $m = 3K$ .

**Table 2.1:** PRD obtained by MMB-IHT and MMB-CoSaMP for different sensing matrices

Reconstruction algorithm	Sampling matrix	CR									
		3.5	4	4.5	5	5.5	6	6.5	7	7.5	8
MMB-IHT	Bernoulli	3.31	3.32	3.32	3.33	3.35	3.55	3.81	4.05	4.43	4.91
	Matrix I	3.42	3.42	3.47	3.7	3.85	4.23	4.8	5.71	6.7	7.79
	Matrix II	3.42	3.43	3.45	3.55	3.67	3.99	4.32	4.68	5.2	5.82
MMB-CoSaMP	Bernoulli	2.98	2.99	2.99	3.11	3.24	3.49	4.02	4.8	9.74	25.58
	Matrix I	2.99	2.99	3.04	3.12	3.62	3.92	5.12	6.9	14.01	28.15
	Matrix II	3.01	3.01	3.03	3.07	3.4	3.7	4.8	6.2	12.9	27.74

### 2.5.3 Evaluation of ECG Reconstruction using the Proposed Method

This section presents an experiment to evaluate how the reconstruction of ECG signals changes as the amount of prior knowledge of the support set varies. The best performance is achieved with the oracle estimate. Assume an oracle reveals the support set  $\Omega$  of the  $K$  most significant wavelet coefficients of the signal of interest. The oracle estimate corresponds to the least squares projection onto the subspace spanned by the columns of  $\Theta$  with indices in  $\Omega$ . This experiment also considers the case where no prior information is known about the support set  $\Omega$ , and the signal is reconstructed using the traditional CoSaMP [59] and the traditional IHT [8]. The performance of all these methods is compared with the performance attained by the proposed algorithms, MMB-CoSaMP and MMB-IHT.

The results are averaged over the set of selected records and illustrated in Fig. 2.7. The error bars indicate plus and minus one standard deviation across ECG records. Given that the objective of this first experiment is only to evaluate the reconstruction of the proposed scheme, we restrict our method to the sampling and reconstruction of the signals and exclude the redundancy removal, quantization, and entropy coding stages. For this experiment, the reconstruction SNR is used to evaluate the quality of the recovered signals as a function of the oversampling ratio  $M/K$ .

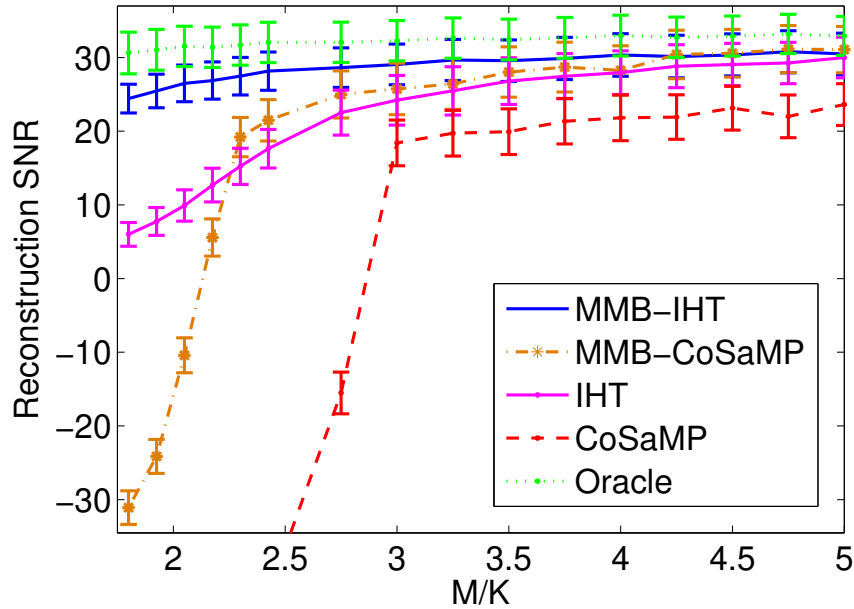
As shown in Fig. 2.7, the proposed algorithms outperform the traditional CoSaMP and IHT, indicating that exploiting the connected subtree structure of the most significant wavelet coefficients, as well as the common support between consecutive ECG segments, results in a large reduction of the required number of measurements to achieve successful recovery of ECG signals. It is also noted from Fig. 2.7 that the algorithms based on IHT require fewer measurements than the algorithms based on CoSaMP to attain good reconstruction. The results of the proposed methods are the closest to the best achievable performance obtained by the oracle estimate.

#### 2.5.4 Evaluation of Compression Performance

For the second experiment, the reconstruction algorithm in the compression scheme, Fig. 2.8, is varied. The proposed MMB-CoSaMP and MMB-IHT are first used as reconstruction algorithms, and their results are compared with BPDN, the bound-optimization-based BSBL algorithm, and an overcomplete dictionary-based reconstruction algorithm. The results are also compared with SPIHT, a state-of-the-art algorithm for ECG compression. The same entropy coding stage of the proposed method is added to SPIHT to ensure fairness in the comparison. The results of the PRD as a function of the compression ratio are illustrated in Fig. 2.8. The results are averaged over the entire set of selected records.

Basis pursuit denoising is the reconstruction algorithm selected by Mamaghani *et al.* [56] and Dixon *et al.* [26] for the recovery of ECG signals. However, the results in Fig. 2.8 indicate that the proposed compression scheme works significantly better when MMB-IHT and MMB-CoSaMP are used as reconstruction algorithms than when BPDN is employed. These results are expected, as the proposed reconstruction algorithms exploit prior knowledge of the signal structure, unlike BPDN, which only leverages the sparsity of the signals.

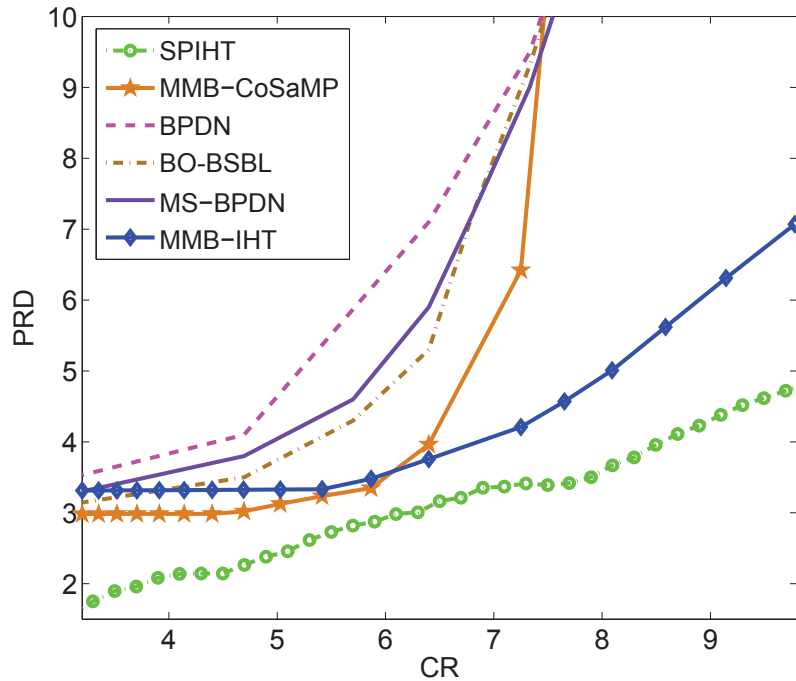
Unlike BPDN, the bound-optimization-based BSBL algorithm, denoted as BO-BSBL, provides flexibility to exploit the block structure and intra-block correlation of the signal sparsity pattern. Even though it was previously employed to reconstruct non-invasive fetal ECG [93], it is also successfully applied in the recovery of adult ECG in the wavelet domain. The reason is the clustering property of the ECG wavelet coefficients that suggests the use of a block-sparsity model. This property refers to the tendency of the large coefficients to cluster together into blocks, as it is suggested by Fig. 2.1. For the implementation of the BO-BSBL algorithm, the block partition was set to  $h = 15$  and the maximum number of iterations to 30. Even though the BO-BSBL algorithm offers performance superior to BPDN, it is outperformed by MMB-IHT and MMB-CoSaMP. This result suggests that the connected subtree structured sparsity model may be more appropriate to represent the largest (in magnitude) ECG



**Figure 2.7:** Comparison of MMB-CoSaMP and MMB-IHT with CoSaMP, IHT, and the oracle estimate. Reconstruction SNR averaged over all the records of the selected data set for different number of measurements.

wavelet coefficients than is the block sparsity model. In addition, the incorporation of prior support knowledge also contributes to the superior performance attained by the proposed methods.

It is of interest to compare the performance of the proposed algorithms with overcomplete dictionary-based reconstruction methods. In a previous work [66], we propose the use of a multi-scale dictionary  $D \in \mathbb{R}^{N \times J}$ ,  $J > N$ , for CS ECG, with the aim of combining the advantages of multi-scale representations using wavelets with the benefits of dictionary learning. The dictionary  $D$  is divided into subdictionaries according to the corresponding wavelet subband and each subdictionary is learned separately. The idea behind this approach is to exploit correlations within each wavelet subband. The multi-scale dictionary-based algorithm, denoted as MS-BPDN, aims to solve problem (1.1) with  $\Theta = \Phi\Psi D$ , instead of  $\Theta = \Phi\Psi$ , by using basis pursuit



**Figure 2.8:** Compression performance evaluation of the compression scheme using the proposed algorithms (MMB-CoSaMP, MMB-IHT), the BO-BSBL, the MS-BPDN, and the BPDN algorithms. The results of SPIHT are included as a baseline for comparisons.

denoising. According to the results in Fig. 2.8, the proposed algorithms outperform MS-BPDN because they exploit additional signal structure information. However, it is also noticed that MS-BPDN outperforms the traditional wavelet based-BPDN algorithm, which suggests the promising application of adaptive overcomplete dictionaries to CS ECG.

The results of SPIHT are illustrated in Fig. 2.8, with the aim of evaluating how the proposed compression scheme compares with current ECG compression methods. SPIHT is a benchmark state-of-the-art algorithm for ECG compression. SPIHT utilizes an optimized embedded coding of the wavelet coefficients that increases the compression rates and makes it outperform the results of the proposed methods, at the

**Table 2.2:** Performance of MMB-IHT and MMB-CoSaMP for CR=6.4

Record	MMB-IHT			MMB-CoSaMP		
	PRD	PRDN	QS	PRD	PRDN	QS
100	3.65	7.73	1.75	3.86	8.18	1.66
101	5.86	7.79	1.09	5.83	7.76	1.09
102	4.84	8.34	1.32	5.21	8.98	1.22
103	3.94	4.8	1.62	4.17	5.08	1.53
107	3.56	3.68	1.79	3.99	4.13	1.6
109	3.91	4.51	1.63	4.13	4.75	1.54
111	5.91	6.99	1.08	6.21	7.35	1.02
115	2.74	6.62	2.33	2.57	6.19	2.49
117	2.11	7.84	3.03	2.11	7.86	3.02
118	2.54	4.61	2.51	2.86	5.2	2.23
119	2.29	4.5	2.78	2.61	5.12	2.45

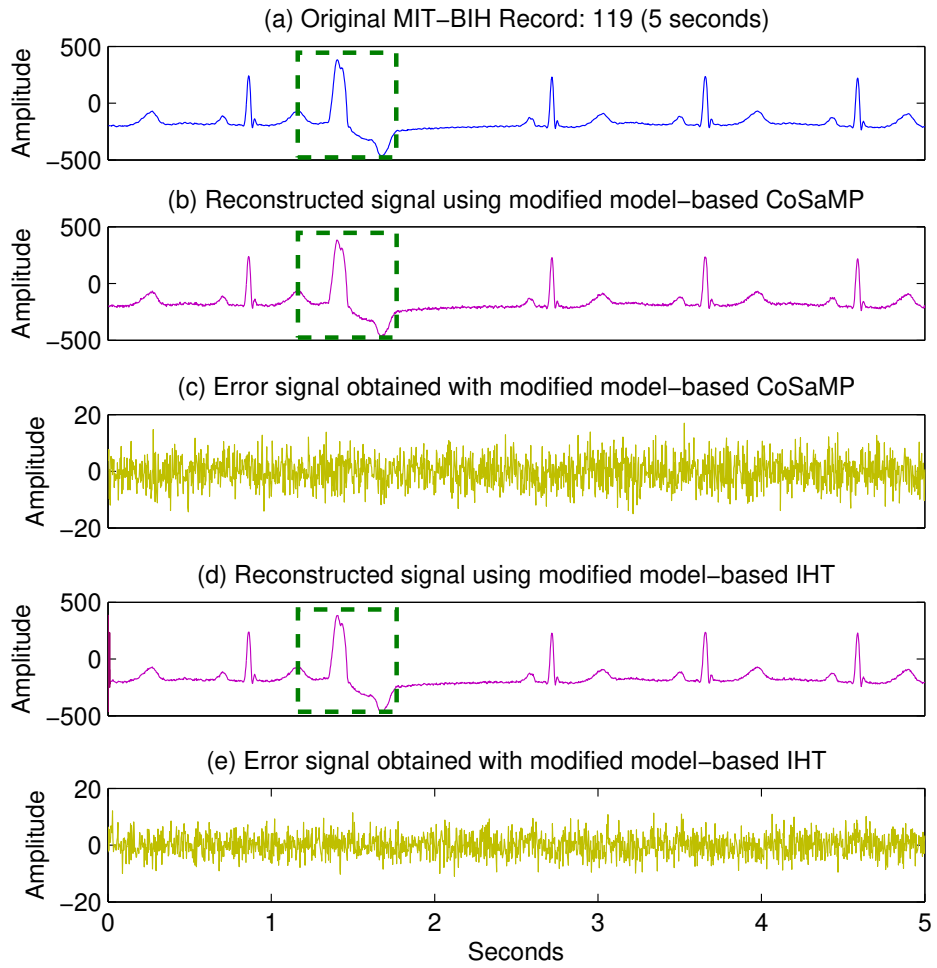
expense of requiring more computations. The time complexity of the SPIHT encoder is  $\mathcal{O}(N\log N)$  [63], while the matrix-vector multiplication of the CS encoder,  $\Phi x$ , requires only  $\mathcal{O}(N)$  operations when  $\Phi$  is a random symmetric Bernoulli matrix [53]. Therefore, the use of a CS encoder offers a substantial complexity reduction.

The difference in compression performance between wavelet transform-based methods and CS-based methods was previously noted by Mamaghanian *et al.* [56], who emphasized the substantial lower power consumption offered by CS-based methods. Even though the proposed algorithm does not outperform SPIHT in terms of compression ratio, it offers better compression and reconstruction performance than previously proposed CS-based methods for ECG compression, such as BPDN and BOBSBL.

There is significant variability among the set of ECG records, and therefore, it is instructive to calculate the reconstruction performance of the proposed algorithms for each ECG recording. The results are illustrated in Table II. For almost all the records, MMB-IHT provides better performance than MMB-CoSaMP.

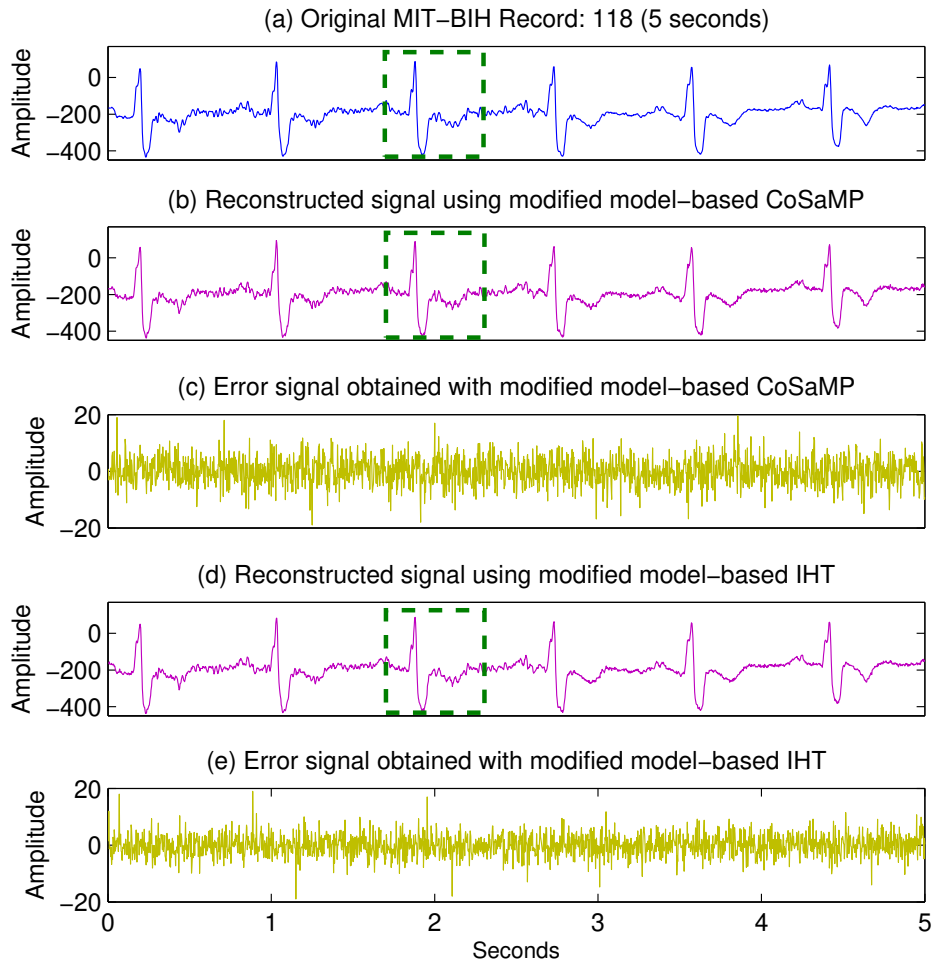
### 2.5.5 Visual Evaluation

Finally, visual study of the reconstructed and error signals using the proposed compression scheme is also presented. Two records with different clinical characteristics are selected for the study: 118 and 119. Record 119 contains ventricular ectopic heartbeats while the record 118 contains right bundle branch block heartbeats. The results for a  $CR = 6.4$  are shown in Fig. 2.9 for record 119 and in Fig. 2.10 for record 118. For the two records, the recovered signals are a good estimate of the original signals and they preserve detailed information for clinical diagnosis. It should be noted that the reconstructed signals using MMB-IHT exhibit less artifacts than the reconstructed signals using MMB-CoSaMP.



**Figure 2.9:** Visual evaluation of the reconstruction of record 119 using MMB-CoSaMP and MMB-IHT.  $CR = 6.4$ .  $PRD=2.61$  (MMB-CoSaMP),  $PRD=2.29$  (MMB-IHT).





**Figure 2.10:** Visual evaluation of the reconstruction of record 118 using MMB-CoSaMP and MMB-IHT.  $CR = 6.4$ .  $PRD=2.86$  (MMB-CoSaMP),  $PRD=2.54$  (MMB-IHT).

## 2.6 Conclusions

The universal applicability of compressed sensing to sparse signals lies in the fact that no specific prior information about the signals is assumed, apart from the sparsity condition. However, for a particular application, some prior information about the signals is typically available. In this chapter, we showed that for the specific application of compressed sensing to ECG compression, the appropriate incorporation of prior information into the reconstruction procedures leads to more accurate reconstruction and higher compression rates. More precisely, we exploit prior information on the connected subtree structure formed by largest (in magnitude) wavelet coefficients and the common support of the wavelet representation of consecutive ECG segments are exploited. The model-based CoSaMP and model-based IHT algorithms are modified to incorporate support knowledge from the previously reconstructed data sequence. The tree-structured sparsity model is also modified to exclude the selection of atoms from the lowest-energy wavelet subband. We justified the application of a model-based compressed sensing approach with the fact that R wave events cause a connected subtree structure of large magnitude wavelet coefficients. In addition, we also selected an appropriate wavelet decomposition level to enable the formation of such structure.

The proposed scheme was evaluated for the compression of a set of eleven ECG records from the MIT-BIH Arrhythmia Database encompassing a variety of signals with different rhythms, wave morphologies and abnormal heartbeats. The experimental results were evaluated in terms of PRD and compression ratio. The proposed method outperformed the results of previously proposed CS-based methods for ECG compression while still maintaining the low-complexity and energy-efficient implementation inherent to CS-based approaches.

## Chapter 3

# MULTI-SCALE DICTIONARY LEARNING FOR COMPRESSIVE SENSING ECG

### 3.1 Introduction

Compressed sensing as applied to ECG utilizes the sparsity of ECG signals to enable accurate reconstruction from undersampled data. Most prior work in CS ECG has employed analytical sparsifying transforms such as wavelets. Even though analytical and pre-defined transformations, such as wavelets, offer fast implementations and a highly sparse representation of natural signals, recent work on adaptively learning the basis functions has demonstrated improved performance over analytical functions [1, 33]. Motivated by this, sparse dictionary learning has recently been applied to ECG reconstruction. For instance, in [90], the authors used K-SVD [1] to learn an overcomplete dictionary from the ECG data itself. Based on this dictionary, the electrocardiogram was reconstructed using the orthogonal matching pursuit algorithm [85]. Similarly, the authors in [35] also focused on reconstructing ECG signals using dictionary learning algorithms. However, in this case, they used a family of iterative least squares-based dictionary learning algorithms using block oriented dictionaries, unrestricted overlapping dictionaries, and overlapping dictionaries with predefined structure.

Recently, the idea of multi-scale dictionary learning was introduced in [60], inspired by the fact that natural signals frequently come at different scales. This approach combines the advantages of multi-scale representations using wavelets with the benefits of dictionary learning. In this chapter, we use multi-scale dictionary learning to create a sparsifying transform for the ECG that allow us to exploit the correlation within each wavelet subband and, subsequently, represent the data in a more efficient manner. The

learned dictionary is later used in a block-based compressive sensing framework where each wavelet subband is sampled and reconstructed independently.

## 3.2 Background and Related Work

This section gives an overview of the traditional dictionary learning problem and of multi-scale dictionary learning using wavelets [60].

### 3.2.1 Traditional Dictionary Learning

We first describe the mathematical model for the dictionary learning problem. Let  $Y = [y_1 \dots y_L] \in \mathbb{R}^{N \times L}$  denotes the set of  $N$ -dimensional training samples. The classic methodology for constructing the overcomplete dictionary  $D = [d_1 \dots d_J] \in \mathbb{R}^{N \times J}$  ( $J > N$ ) tries to solve the following optimization problem

$$\arg \min_{X, D} \|Y - DX\|_F^2 \text{ s. t. } \|x_i\|_0 < T, \quad \forall i \quad (3.1)$$

with  $X = [x_1 \dots x_L] \in \mathbb{R}^{J \times L}$  denoting the sparse codes of input signals  $Y$ , and  $T$  the pre-specified sparsity threshold.

The Method of optimal directions [34] is essentially a dictionary learning algorithm inspired by the generalized Lloyd algorithm (GLA) [38] and used to solve the problem in (3.1). GLA is an iterative algorithm designed for optimizing vector quantization codebooks. At each iteration, MOD alternates between sparse coding of the examples based on the current dictionary and an update process for the dictionary atoms so as to better fit the data.

In the dictionary update stage, MOD assumes the coefficient vector  $X$  as known, and seeks an update of the dictionary that minimizes the mean square error given by

$$\|E\|_F^2 = \|[e_1, e_2, \dots, e_N]\|_F^2 = \|Y - DX\|_F^2. \quad (3.2)$$

Taking the derivative of the above formula with respect to  $D$  ( $X$  and  $Y$  are fixed),  $(Y - DX)X^T = 0$ , results in the following dictionary update expression

$$D^{(t+1)} = YX^{(t)T}(X^{(t)}X^{(t)T})^{-1}. \quad (3.3)$$

MOD derives the best possible dictionary in the mean square error sense for a fixed coefficient matrix  $X$ .

In the sparse coding stage,  $D$  is fixed, and the problem reduces to searching for sparse representations with coefficients summarized in the matrix  $X$ . The problem can be decoupled to  $L$  distinct problems of the form

$$\arg \min_{x_i} \|y_i - Dx_i\|_2^2 \quad \text{s. t.} \quad \|x_i\|_0 < T, \quad \text{for } i = 1, 2, \dots, L. \quad (3.4)$$

In this chapter, problem (3.4) is addressed by using the orthogonal matching pursuit algorithm [85] described in Section 1.4.3.2.

### 3.2.2 Multi-scale Dictionary Learning using Wavelets

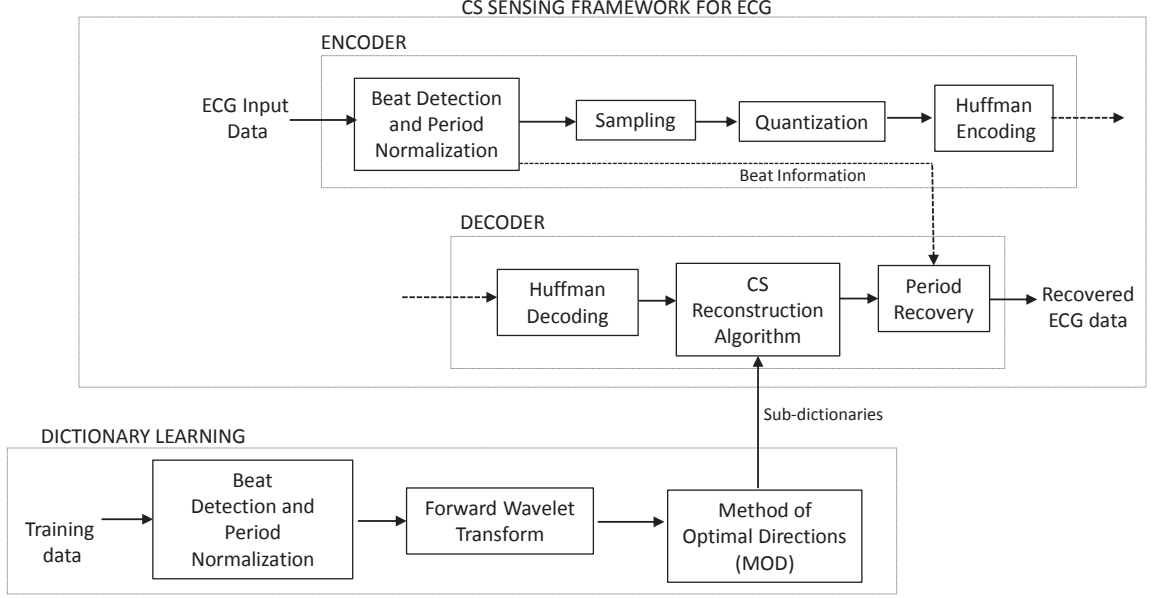
Each atom in  $D$  can be expressed as a linear combination of pre-atoms extracted from the wavelet synthesis matrix  $W_s$ , *i.e.*  $D = W_s A$ . Based on this representation, the learning problem can be expressed by the following modification to (3.1)

$$\arg \min_{X,A} \|Y - W_s AX\|_F^2 \quad \text{s. t.} \quad \|x_i\|_0 \leq T, \quad \forall i \quad (3.5)$$

This model suggests that the data can be expressed by a sparse combination of atoms, which are themselves combinations of atoms from the wavelet dictionary. Since  $W_s$  is square and unitary, the optimization problem in (3.5) can equivalently be written as

$$\arg \min_{X,A} \|W_a Y - AX\|_F^2 \quad \text{s. t.} \quad \|x_i\|_0 \leq T, \quad \forall i \quad (3.6)$$

where  $W_a$  denotes the wavelet analysis operator (forward wavelet transform). The signal representation in the wavelet domain can be seen as a collection of coefficient



**Figure 3.1:** Block Diagram of the proposed method

subbands, and subsequently, we can train different subdictionaries  $A_b$  for each of these subbands

$$\forall b \quad \arg \min_{A_b, X_b} \|(W_a Y)_b - A_b X_b\|_F^2 \quad \text{s. t.} \quad \|x_{i,b}\|_0 \leq T_b, \quad \forall i \quad (3.7)$$

with  $b$  denoting the different wavelet coefficient subbands. Problem (3.7) can be solved using the MOD algorithm.

### 3.3 Methods

This chapter proposes the reconstruction of ECG signals using a CS approach where the dictionary is adaptively learned from ECG training data. The proposed scheme is divided into two main stages: the dictionary learning stage and the CS stage. Both stages require preprocessing of ECG signals, which refers to period normalization of ECG cycles. The first step is the detection of the R peaks. Since each ECG cycle may have a different duration, all cycles are normalized to the same length  $N$  by using cubic-spline interpolation [70]. The original periods are sent to the decoder as side

information since they are needed for the recovery stage. The normalized cycles are organized column-wise in a matrix. The dictionary learning and the CS stage are described thoroughly in this section. The block diagram of the proposed compression scheme is presented in Fig. 3.1.

### 3.3.1 Dictionary Learning in the Wavelet Domain

After preprocessing the training data, we get a matrix  $Y \in \mathbb{R}^{N \times L}$  whose columns correspond to normalized heartbeats. The length and the number of normalized ECG cycles are denoted by  $N$  and  $L$ , respectively. For the dictionary learning, we first build a matrix  $S \in \mathbb{R}^{N \times L}$  by applying the one-dimensional wavelet transform to each of the columns in  $Y$ . That is,  $S = W_{ac}Y$ , where the operator  $W_{ac}$  is the one-dimensional forward wavelet transform applied in a column-wise manner.

Assuming the wavelet decomposition level is  $B$ , the matrix  $S$  can be decomposed into  $B + 1$  submatrices, each representing a different wavelet subband. That is  $S = [S_1 \dots S_{B+1}]^T$ , where the first  $B$  submatrices,  $S_b \in \mathbb{R}^{N/2^b \times L}$ , correspond to the discrete wavelet coefficients at scales  $b = 1, \dots, B$ , and the submatrix  $S_{B+1} \in \mathbb{R}^{N/2^B \times L}$  corresponds to the approximation subband. For each wavelet subband  $S_b$ , a sub-dictionary  $A_b$  is trained with a number of atoms  $K_b$ . The training is performed by solving the following optimization problem with the MOD algorithm

$$\forall b \quad \arg \min_{A_b, X_b} \|S_b - A_b X_b\|_F^2 \quad \text{s. t.} \quad \|x_{i,b}\|_0 < T_b, \quad \forall i. \quad (3.8)$$

where  $X_b = [x_{1,b} \dots x_{L,b}] \in \mathbb{R}^{K_b \times L}$  denotes the sparse code of the input signal  $S_b$ , and  $T_b$  the pre-specified sparsity threshold.

### 3.3.2 Compressed Sensing Reconstruction of ECG Signals

As a result of preprocessing the ECG data, we obtain a matrix  $F \in \mathbb{R}^{N \times P}$  whose columns correspond to normalized ECG cycles.  $N$  and  $P$  are the length and the number of normalized ECG cycles, respectively. Let the operator  $W_{sc}$  be the one-dimensional inverse wavelet transform applied in a column-wise manner, the matrix

$F$  can be expressed as  $F = W_{sc}V$ , where  $V \in \mathbb{R}^{N \times P}$  is the corresponding wavelet representation.

As with matrix  $S$  in the dictionary learning stage, we can decompose  $V$  into different wavelet subbands. That is  $V = [V_1 \dots V_{B+1}]^T$ , where the first  $B$  submatrices,  $V_b \in \mathbb{R}^{N/2^b \times P}$ ,  $b = 1, \dots, B$ , correspond to the wavelet coefficients at scales  $b = 1, \dots, B$ , and the submatrix  $V_{B+1} \in \mathbb{R}^{N/2^B \times P}$  corresponds to the approximation subband.

The information we gather about  $V_b = [v_{1,b} \dots v_{P,b}]$ ,  $b = 1, \dots, B$ , can be described by  $G_b = [g_{1,b} \dots g_{P,b}] = \Phi_b V_b$ , where  $\Phi_b$  is a  $m_b \times N/2^b$  matrix, that needs to satisfy the restricted isometry property [15] in order to recover  $V_b$ . It is known that random matrices satisfy this condition with overwhelming probability. Here we assume that the entries of the matrix  $\Phi_b$ ,  $b = 1, \dots, B$ , are independently sampled from the normal distribution with mean zero and variance  $1/m_b$ . Similarly, the matrix  $V_B$  can be sampled with a random matrix  $\Phi_B \in m_B \times N/2^B$ , such that  $G_B = \Phi_B V_B$ . The effective matrix  $\Phi$ , such that  $G = [G_1 \dots G_{B+1}]^T = \Phi V$ , is a block diagonal matrix whose diagonal elements correspond to  $\Phi_b$ ,  $b = 1, \dots, B + 1$ ,

$$\Phi = \begin{bmatrix} \Phi_1 & & & & \\ & \Phi_2 & & & \\ & & \ddots & & \\ & & & \ddots & \\ & & & & \Phi_{B+1} \end{bmatrix}$$

We propose to sample the matrix of normalized heartbeats,  $F$ , with the matrix  $\Phi W_{ac}$ . That is,

$$\begin{aligned} \Phi W_{ac} F &= \Phi W_{ac} W_{sc} V \\ &= \Phi V = G \end{aligned} \tag{3.9}$$

To further improve the compression and reduce the amount of redundant information, we use Huffman coding. An 9-bit optimal scalar quantizer designed with the Lloyd-Max algorithm is utilized [38]. The matrices  $V_b$ ,  $b = 1, \dots, B + 1$ , can be sparsely represented using the dictionaries built in the dictionary learning stage, *i.e.*



$V_b = A_b X_b$ , where  $X_b = [x_{1,b} \dots x_{L,b}]$  denotes the corresponding sparse code. For the reconstruction of the sparse codes, the basis pursuit algorithm with  $\ell_2$  constraint is employed as follows

$$\min \|x_{i,b}\|_1 \text{ s. t. } \|g_{i,b} - \Phi_b A_b x_{i,b}\|_2 \leq \epsilon, \quad \forall i, 1 \leq b \leq B + 1 \quad (3.10)$$

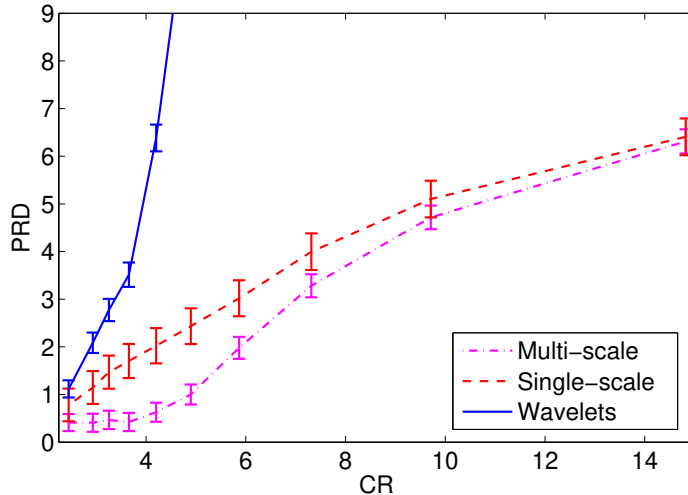
where  $\epsilon$  bounds the amount of noise unavoidably corrupting the data.

Each matrix  $V_b$  is reconstructed by calculating  $V_b = A_b X_b$ . The matrix of normalized ECG cycles  $F$  is reconstructed by applying the one-dimensional inverse wavelet transform to the columns of  $V$ , *i.e.*  $F = W_{sc} V$ . The original cycles are recovered by using cubic spline interpolation [70].

### 3.4 Experimental Results

This section describes three experiments employed to evaluate the performance of the proposed scheme. The ECG data in our experiments are from the MIT-BIH Arrhythmia Database, sampled at 360 Hz with a resolution of 11 bits/sample. We use the Daubechies db4 wavelets at decomposition level  $B = 3$ . The detail subbands are denoted as  $V_1, V_2$ , and  $V_3$ , while the approximation subband is denoted as  $V_4$ . For our simulations, results are averaged over 50 trials.

The PRD and CR are used as performance measures in the simulations. After the preprocessing stage, all ECG cycles are normalized to the length  $N = 256$ . For the dictionary learning stage, the training data are constructed as a set of  $L = 30N = 7680$  ECG cycles from records 100, 102, 230, and 109. We execute the MOD algorithm for a total of 50 iterations, training subdictionaries containing  $3N$  atoms for both the approximation subband,  $V_4$ , and the detail subband at scale  $b = 3$ ,  $V_3$ . Since the subdictionary for the detail subband at scale  $b = 2$  needs to express a higher level of variability than the ones for scales 3 and 4, we train a subdictionary with  $4N$  atoms for  $V_2$ . The coefficients of the subband  $V_1$  are close to zero, and therefore there is no significant need to sample them and recover their magnitude. Thus, we opt for approximating the coefficients of  $V_1$  to zero.



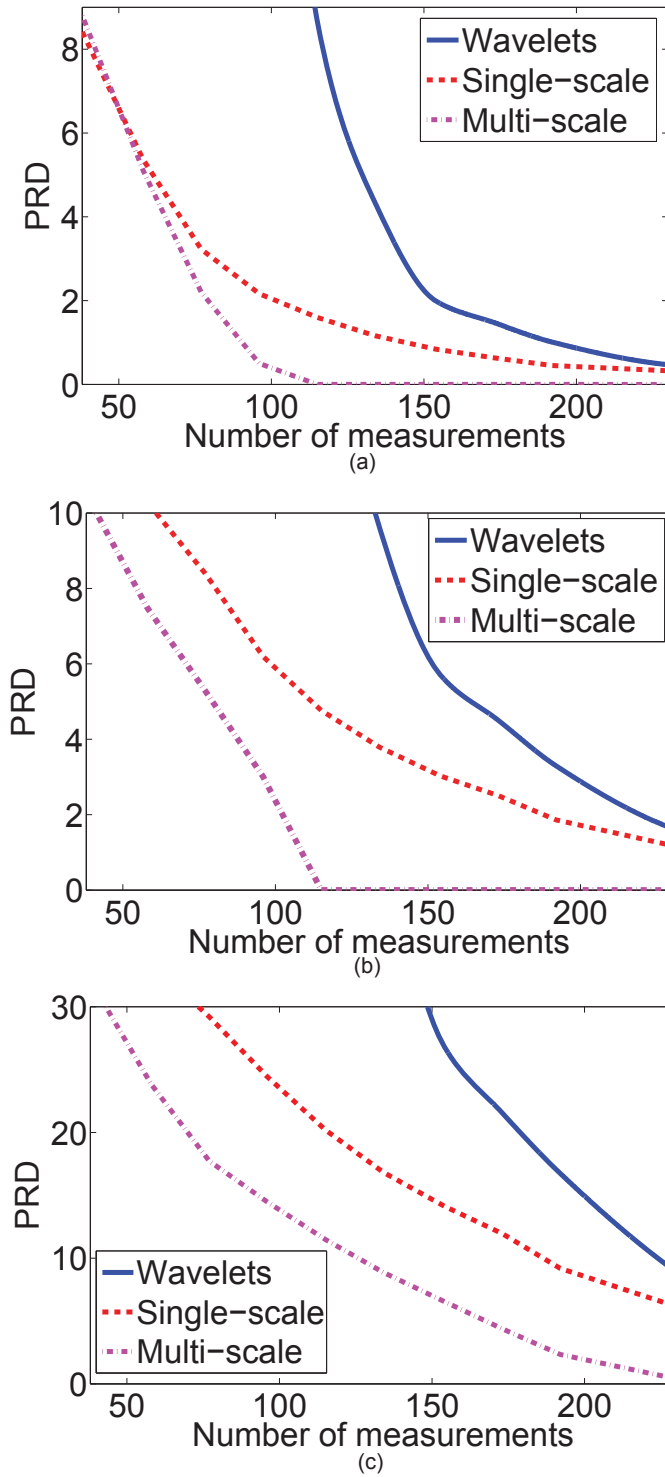
**Figure 3.2:** Comparison of CS-based methods using different dictionaries. ECG compression is evaluated using PRD as a function of the compression ratio.

The proposed method is evaluated with 4000 ECG cycles that are not included in the training set. Even though the length of the wavelet subbands decreases across scales, we sample all the subbands  $V_b$ ,  $b = 2, \dots, 4$ , with the same number of measurements because the magnitude of the wavelet coefficients tends to increase across scales. Each subband is sampled independently with matrices having i.i.d. entries drawn from a standard normal distribution with normalized columns.

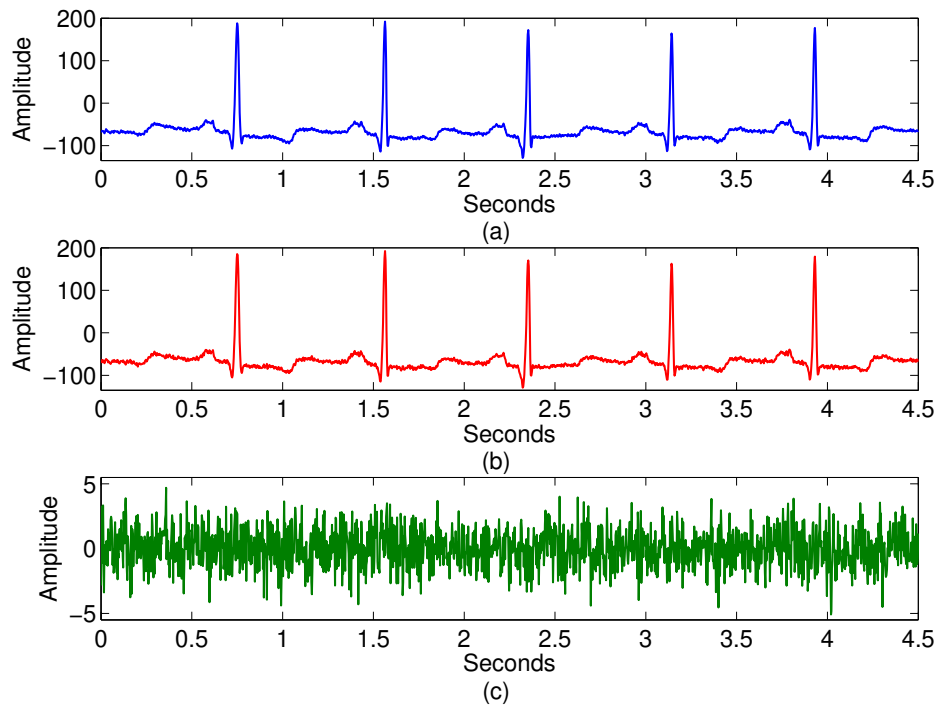
We compare our approach with the single-scale version of the proposed method, which is described in [60]. We also compare the results with the CS reconstruction using the standard wavelet basis as the sparsifying transform. In Fig. 3.2, the results are shown in terms of the PRD as a function of the compression ratio. The error bars indicate plus and minus one standard deviation across ECG records. Since the objective is to attain accurate reconstruction of signals, the range of interest is restricted to PRD values between 0 and 9. In this range, dictionary learning-based methods offer better performance than the CS method that uses standard wavelet dictionaries. This figure also shows that introducing multi-scale dictionaries in CS algorithms improves the quality of the reconstruction of ECG signals compared to single-scale approaches.

Evaluation of the reconstruction of each wavelet subband with the proposed algorithm and comparison with the results of the single-scale and wavelet dictionaries are also presented. For this experiment, we restrict our method to sampling and reconstruction of the signals and exclude the quantization and entropy coding stages. Figure 3.3 illustrates the results. The proposed approach provides the best performance for all the different wavelet subbands. Of special interest is the performance attained for  $V_3$  and  $V_4$ , the wavelet subbands that accumulate most of the signal energy. The proposed method achieves almost perfect reconstruction of these subbands when using a number of measurements greater than 115.

In this work, visual study of error signals is also considered. Error signals are calculated as the difference between original and reconstructed signals. In the last experiment, the waveforms of record 100 produced by the proposed reconstruction scheme are visually evaluated in Fig. 3.4. The reconstructed signal remains close to the original signal and the error is equally distributed along the temporal axis. This implies that the proposed method performs well locally and does not incorporate outliers in the reconstruction.



**Figure 3.3:** Reconstruction results of wavelet subbands  $V_2$ ,  $V_3$ , and  $V_4$  using CS-based methods with different dictionaries. (a) Approximation subband,  $V_4$ . (b) Detail subband  $V_3$ . (c) Detail subband  $V_2$ .



**Figure 3.4:** Reconstruction of record 100 using using 65 measurements.  $PRD=4.35$ .  
(a) Original signal. (b) Reconstructed signal. (c) Error signal.

### 3.5 Conclusions

In this chapter, a novel CS-based method for the acquisition and reconstruction of the ECG is presented. The proposed scheme exploits the fact that different wavelet subbands contain data at different scales to efficiently learn dictionaries for sparse and redundant representation of ECG data. This chapter also proposes an efficient acquisition method that allows for independent sampling of each wavelet subband. Consequently, the number of measurements for each wavelet subband can be selected according to the location of the largest wavelet coefficients. Simulation results demonstrate the superior performance of the proposed approach as compared to other CS-based methods, and also show dramatic reconstruction improvement of the wavelet subbands that accumulate most of the signal energy.

## Chapter 4

### A WEIGHTED $\ell_1$ MINIMIZATION ALGORITHM FOR COMPRESSED SENSING ECG

#### 4.1 Introduction

Compressive sensing, as detailed in previous chapters, has recently been applied to ECG acquisition and reconstruction with the aim of lowering energy consumption and sampling rates in WBANs for ambulatory ECG monitoring. However, most current methods only adopt a sparse prior on the ECG wavelet representation, without exploiting the rich structure of the ECG wavelet coefficients. In this chapter, we propose a reconstruction algorithm that goes beyond simply assuming that ECG signals are compressible in a wavelet basis by exploiting the *a priori* statistical information that ECG signals exhibit in the wavelet domain.

Recent works have exploited the incorporation of prior information about the signal of interest into CS-based reconstruction algorithms. In [3], for instance, a connected tree structure was assumed in the wavelet domain, which restricted the class of signals that could be reconstructed to be piecewise smooth. A different approach was proposed by He *et al.* [40], in a work that imposed the structure statistically using a Bayesian prior, instead of making an explicit imposition of the coefficients structure. In [17], we showed that iterative CS algorithms can be easily modified to incorporate known support in the recovery process.

A latent problem when trying to reconstruct ECG signals using CS-based methods is the inability to accurately recover the low-magnitude coefficients of the wavelet representation [93]. To alleviate this problem, we propose to incorporate prior information about the magnitude decay of the wavelet coefficients across subbands in the reconstruction algorithm. More precisely, we derive a weighted  $\ell_1$  minimization

algorithm, based on a MAP approach, with a weighting scheme based on the standard deviation of the wavelet coefficients at different scales. In addition, the weighting scheme also takes into consideration the fact that the approximation subband coefficients accumulate most of the signal energy. Experimental results on ECG records from the MIT-BIH Arrhythmia Database validate the superior performance of the proposed algorithm, in terms of reconstruction quality and number of measurements, compared to current CS-based methods with application to wireless ECG systems.

## 4.2 Motivation and Background

### 4.2.1 Problem Formulation

Let  $x \in \mathbb{R}^N$  and  $\Psi \in \mathbb{R}^{N \times N}$  be an ECG signal and a wavelet basis, respectively. Thus, signal  $x$  can be well approximated by a linear combination of a small set of vectors from  $\Psi$ , *i.e.*  $x \approx \sum_{i=1}^K s_i \psi_i$ , where  $K \ll N$ . Let  $\Phi$  be an  $M \times N$  sensing matrix,  $M < N$ . The problem addressed in this chapter is the recovery of  $x$  from undersampled linear measurements of the form  $y = \Phi x + n = \Phi \Psi s + n$ , where  $n$  accounts for the additive Gaussian sampling noise. If we define  $\Theta = \Phi \Psi$ , then the measurement vector becomes  $y = \Theta s + n$ . As described in previous chapters, the reconstruction of  $s$  is typically attained via the BPDN algorithm:

$$\min_s \frac{1}{2} \|y - \Theta s\|_2^2 + \lambda \|s\|_1, \quad (4.1)$$

with  $\lambda$  a parameter that controls the trade-off between sparsity and reconstruction fidelity. Problem (4.1) can be viewed as a MAP estimate for  $s$  under the assumption that each component of  $s$  is drawn i.i.d. from a Laplace prior [52].

### 4.2.2 Wavelet representation

Given a scaling function  $\varphi$  and a wavelet function  $\psi$ , the wavelet representation of  $x$  can be expressed in terms of shifted versions of  $\varphi$  and shifted and dilated versions



of  $\psi$

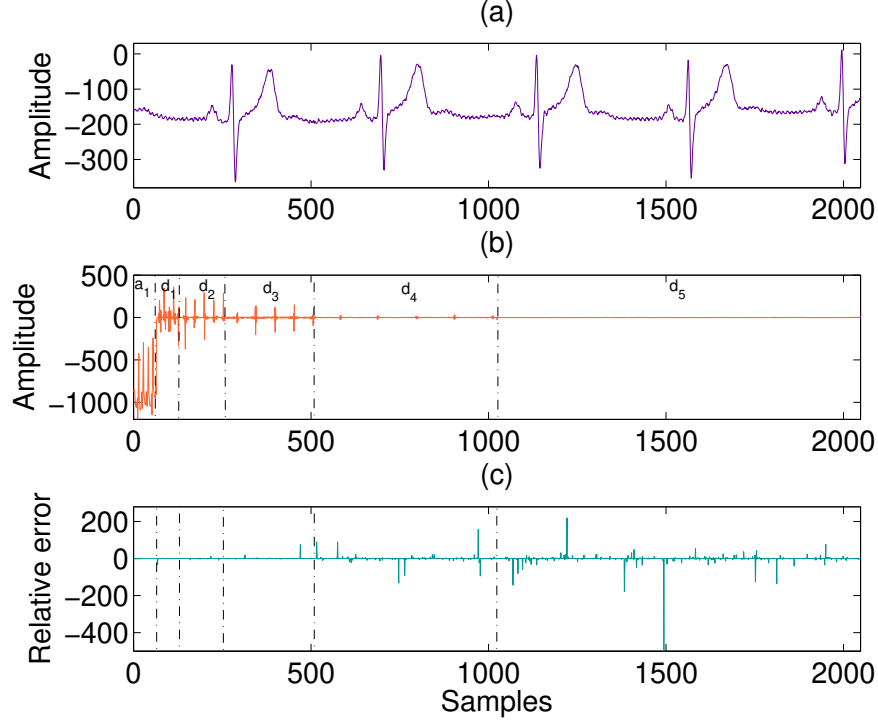
$$x = \sum_{i=0}^{N_1-1} a_{1,i} \varphi_{1,i} + \sum_{j=1}^L \sum_{i=0}^{N_j-1} d_{j,i} \psi_{j,i}, \quad (4.2)$$

where  $j$  denotes the scale of analysis and  $L$  indicates the finest scale.  $N_j = N/2^{L-j+1}$  corresponds to the number of coefficients at scale  $j \in \{1, \dots, L\}$  and  $i$  represents the position,  $0 \leq i \leq N_j - 1$ . The wavelet transform consists of the scaling coefficients  $a_{1,i}$  and wavelet coefficients  $d_{j,i}$ . Using vector notation, we write  $x = \Psi s$ , where  $s = [a_{1,1} \dots a_{1,N_1-1} d_{1,0} \dots d_{1,N_1-1} \dots d_{L,0} \dots d_{L,N_L-1}]^T$  is the vector of scaling and wavelet coefficients and  $\Psi$  is the orthogonal matrix containing the wavelet and scaling functions as columns. The vector  $s$  can be decomposed into  $L+1$  subvectors. The first subvector corresponds to the scaling coefficients and is denoted as  $a_1$ . The next  $L$  subvectors are denoted by  $d_j$ ,  $j = 1, \dots, L$ , and the  $j$ th subvector contains all of the wavelet coefficients for scale  $j$ . Thus,  $s$  can also be written as  $s = [a_1 d_1 d_2 \dots d_L]^T$ .

### 4.2.3 Motivation

The optimization problem in (4.1) only considers the sparsity of the ECG wavelet representation and, therefore, it does not exploit all of the rich structure present in the ECG wavelet representation. Recent CS efforts show that the incorporation of prior knowledge into standard sparse recovery algorithms can boost their performance [17, 89]. Along these lines, we aim to find properties of the ECG wavelet representation that can be incorporated into CS-based algorithms to improve the reconstruction and reduce the number of necessary measurements.

An experiment is performed to evaluate the recovery of ECG signals when only the sparsity property is exploited in the reconstruction. The selected signal, denoted as  $x$  and illustrated in Fig. 4.1(a), corresponds to a sequence of the record 117 from the MIT-BIH Arrhythmia Database, formed by  $N = 2048$  samples. Figure 4.1(b) presents its corresponding wavelet representation, using Daubechies-4 and a decomposition level  $L = 5$ . The dotted lines indicate the separation between consecutive wavelet subbands. Note that the approximation subband accumulates most of the signal energy. The



**Figure 4.1:** Reconstruction of ECG signals using BPDN (a) Original signal. (b) Wavelet representation. (c) Relative error signal.

sparsity level is selected as the number of coefficients that accumulates 99.99% of the signal energy, which corresponds to  $K = 210$  for the selected sequence. The sequence is sensed with a random CS matrix satisfying the RIP, and with the number of measurements set as  $M = 630$ . The reconstruction is performed with the traditional BPDN algorithm presented in (4.1) and the performance, in terms of the relative reconstruction error, is illustrated in Fig. 4.1(c). The relative reconstruction error is defined as  $(x - \hat{x})/x$ , where  $\hat{x}$  is the recovered signal. From Fig. 4.1(c), it is noted that the error in the reconstruction of the detail subbands  $d_4$  and  $d_5$  has the highest magnitude, which is detrimental for medical diagnosis purposes.

The unsatisfactory performance of BPDN in reconstructing ECG signals motivates the development of new CS-based algorithms for ECG reconstruction. This chapter aims at advancing CS ECG by exploiting two properties of the ECG wavelet

representation. More precisely, we exploit the fact that the approximation subband coefficients accumulate most of the signal energy and that the magnitude of the detail wavelet coefficients decreases across scales.

### 4.3 Methods

The same compression scheme presented in Chapter 2.4 and illustrated in Fig. 2.3 is employed here. The entries of the sampling matrix  $\Phi$  are independently sampled from a symmetric Bernoulli distribution ( $P(\Phi_{i,j} = \pm 1/\sqrt{M}) = 1/2$ ) to facilitate an efficient hardware implementation [21]. An 8-bit optimal scalar quantizer designed with the Lloyd-Max algorithm is utilized. The contribution of this chapter is the reconstruction algorithm, which is different from that presented in Chapter 2.4.

An important property of CS is its ability to reconstruct signals in total blindness of any structure beyond standard sparsity on a given basis. However, this property does not discard the possibility of adapting or optimizing the CS recovery process to the specific signal at hand in order to reduce the number of measurements needed to faithfully represent it. This section aims to reconstruct the vector  $s$  from the measurements  $y = \Theta s + n$ , where  $n$  is the noise corrupting the data. A maximum *a posteriori* (MAP) approach is proposed for the reconstruction of  $s$ . To favor a sparse estimate, a Laplacian distribution with standard deviation  $\sigma_i$  is adopted for each entry  $s_i$  of  $s$ ; that is

$$p(s_i) = \frac{1}{\sqrt{2}\sigma_i} \exp\left(-\frac{\sqrt{2}\|s_i\|_1}{\sigma_i}\right). \quad (4.3)$$

The entries of  $s$  are assumed to be independent. The noise  $n$  is modeled as independent and Gaussian with zero mean and variance equal to  $\sigma_n^2$ . To infer  $s$  from  $y$ , we maximize the conditional probability distribution  $p(s|y, \Theta)$ , which can be expressed by means of Bayes's rule as

$$p(s|y, \Theta) \propto p(y|\Theta, s)p(s). \quad (4.4)$$

Because the noise is assumed to be Gaussian, the likelihood function is given by

$p(y|\Theta, s) \propto \exp(-\|y - \Theta s\|_2^2 / (2\sigma_n^2))$ . Therefore, maximizing the posterior distribution  $p(s|y, \Theta)$  leads to

$$s_{\text{MAP}} = \arg \max_s p(s|y, \Theta) \quad (4.5)$$

$$= \arg \max_s \left( \log p(y|\Theta, s) + \sum_i \log p(s_i) \right) \quad (4.6)$$

$$= \arg \min_s \left( \frac{\|y - \Theta s\|_2^2}{2\sigma_n^2} + \sum_i \frac{\sqrt{2}\|s_i\|_1}{\sigma_i} \right). \quad (4.7)$$

The problem in (4.7) can also be expressed as

$$s_{\text{MAP}} = \arg \min_s \left( \frac{1}{2}\|y - \Theta s\|_2^2 + \lambda\|W s\|_1 \right), \quad (4.8)$$

where  $W$  is a diagonal matrix, whose  $i$ th diagonal element is of the form  $1/\sigma_i$ , and  $\lambda$  is a tuning parameter. This problem is equivalent to BPDN when all  $\sigma_i$  are equal.

The solution of problem (4.8) requires the standard deviations for each coefficient of the wavelet representation, which are unknown in our case. To cope with this issue, we propose to model the variance variation across scales with exponential decay functions. As presented in Section 4.2.3, the detail wavelet coefficients of ECG signals tend to decrease across scales, and this behavior can be enforced by modeling the variances so that they decay exponentially as the scale becomes finer. This corresponds to the model proposed by Romberg *et al.* [79]:

$$\sigma_j^2 = C2^{-j\alpha} \quad j = 1, \dots, L, \quad (4.9)$$

where  $C$  and  $\alpha$  are the model parameters and  $j$  is the scale of analysis. In this model, the  $\sigma_i$ ,  $i = 1, \dots, N$  values are made equal for all coefficients within a scale, and therefore,  $\sigma_j^2$  refers to the variance of the coefficients at scale  $j$ .

Define  $W^* = \sqrt{C}W$  and  $\lambda^* = \lambda/\sqrt{C}$ . Then, problem (4.8) can be reformulated

as the following  $\ell_1$  weighted minimization

$$s_{\text{MAP}} = \arg \min_s \left( \frac{1}{2} \|y - \Theta s\|_2^2 + \lambda^* \|W^* s\|_1 \right), \quad (4.10)$$

where  $\lambda^*$  is regarded as a tuning parameter. Using eq. (4.9) and the fact that each diagonal entry of  $W$  satisfies  $W_{i,i} = 1/\sigma_i$ , we infer that the diagonal elements of  $W^*$  corresponding to scale  $j$  are of the form  $2^{j\alpha/2}$ . Traditional  $\ell_1$  minimization is therefore a special case where these standard deviations are assumed to be equal.

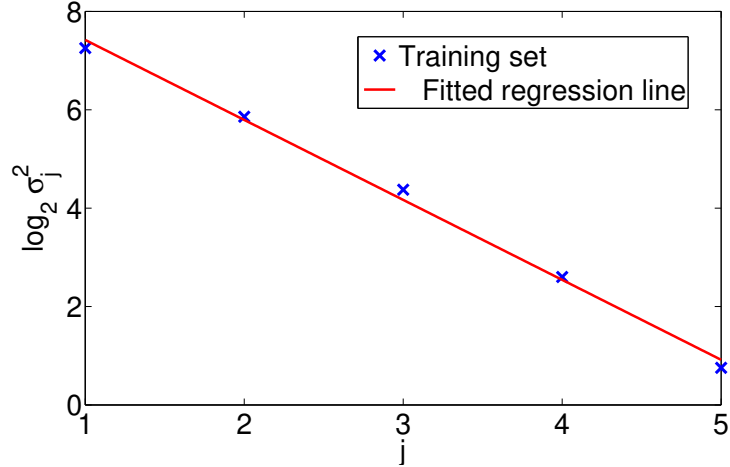
As presented in Chapter 4.2.3, the approximation subband coefficients accumulate most of the signal energy and, therefore, should be included in the sparse representation of the ECG. To exploit this property, we employ an approach similar to that of Vaswani *et al.* [89] to reconstruct a sparse signal when part of the support is known *a priori*. This approach consists of finding the signal that satisfies the data fidelity constraint and is the sparsest outside of the known support. In our approach, this idea is implemented by setting to zero the diagonal entries of  $W^*$  corresponding to the approximation subband, *i.e.*  $W_{i,i}^* = 0$  for  $i = 1, \dots, N_1$ .

As the diagonal entries of  $W^*$  only depend on the value of  $\alpha$ , problem (4.10) can be solved after  $\alpha$  is calculated. This leads to the need for a training stage to estimate the value of  $\alpha$ . The first step in this training estimates the standard deviations  $\sigma_j$ ,  $j = 1, \dots, L$  using maximum likelihood estimation. Once the variances are estimated, simple linear regression is employed to solve for  $\alpha$  in the following equation, derived from (4.9),

$$\log_2 \sigma_j^2 = \log_2 C - j\alpha, \quad j = 1, \dots, L. \quad (4.11)$$

#### 4.4 Experimental Results

To validate the proposed method, the MIT-BIH Arrhythmia Database [39] is employed for both training and testing. The data set proposed by Lu *et al.* [54] is used in our experiments. It consists of records 100, 101, 102, 103, 107, 109, 111, 115, 117, 118 and 119, which encompasses a variety of signals with different rhythms, QRS



**Figure 4.2:** Scatter plot of  $j$  versus  $\log_2 \sigma_j^2$  and fitted regression line.

complex morphologies and ectopic beats. The length of the sliding window is set to  $N = 2048$ , a commonly used segment length value for ECG processing. The orthogonal Daubechies-4 wavelets is employed as the sparsifying transform and the decomposition level is set to  $L = 5$ .

The reconstruction SNR and the PRD are employed as performance measures for experiments. The first experiment aims at finding the parameter  $\alpha$  in (4.11) through simple linear regression. The training data consists of the wavelet representation of 330 ECG sequences of length  $N = 2048$  from the selected set of records; 30 sequences per record. Let  $c_j$  denote the vector formed by the concatenation of the wavelet subbands from the training data at scale  $j$ . The standard deviations  $\sigma_j$ ,  $j = 1, \dots, L$ , are first calculated using the maximum likelihood estimate of the standard deviation of a Laplacian distribution,

$$(\sigma_j)_{\text{ML}} = \frac{\sqrt{2} \|c_j\|_1}{P_j} \quad j = 1, \dots, L, \quad (4.12)$$

where  $P_j$  refers to the dimension of  $c_j$ . Figure 4.2 shows the scatter plot with scale  $j$  displayed on the horizontal axis and  $\log_2 \sigma_j^2$  on the vertical axis. The continuous line

in Fig. 4.2, with slope  $-\alpha = -1.62$ , corresponds to the fitted regression line.

The second experiment is performed to compare the proposed reconstruction algorithm with the BPDN algorithm, which is the reconstruction algorithm selected by Mamaghanian *et al.* [56], Dixon *et al.* [26], and Chen *et al.* [21] for the recovery of ECG signals. The experiment is carried out and averaged over 10-min long single leads extracted from the selected set of records. Results are shown in Fig. 4.3. The reconstruction SNR is used to evaluate the quality of the recovered signals as a function of the number of measurements  $M$ . The measurements are corrupted by additive white Gaussian noise with  $\sigma_r = 0.05$ . Chen *et al.* [22] proposed to set the tuning parameter to the value  $\lambda = \sigma_r \sqrt{2 \log(N)}$ . The same criteria is adopted for the selection of the tuning parameters in our experiments. The results of the entire compression scheme, including redundancy removal of measurements, quantization, and Huffman coding, are presented in Fig. 4.4. The error bars indicate plus and minus one standard deviation across ECG records.

As shown in Figs. 4.3 and 4.4, the proposed algorithm outperforms the traditional BPDN algorithm as it requires fewer measurements while achieving superior reconstruction quality. These results are expected as the proposed method exploits prior knowledge of the signal structure, unlike BPDN, which only leverages the signal sparsity.

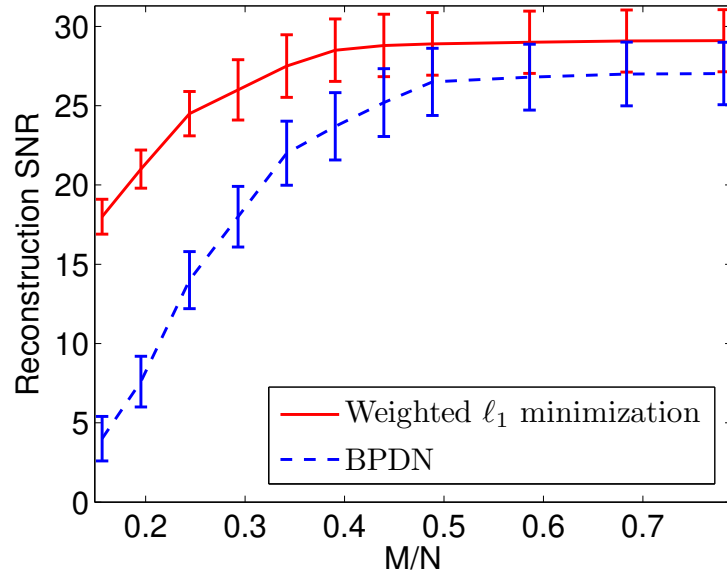


Figure 4.3: Comparison of the proposed method with BPDN.

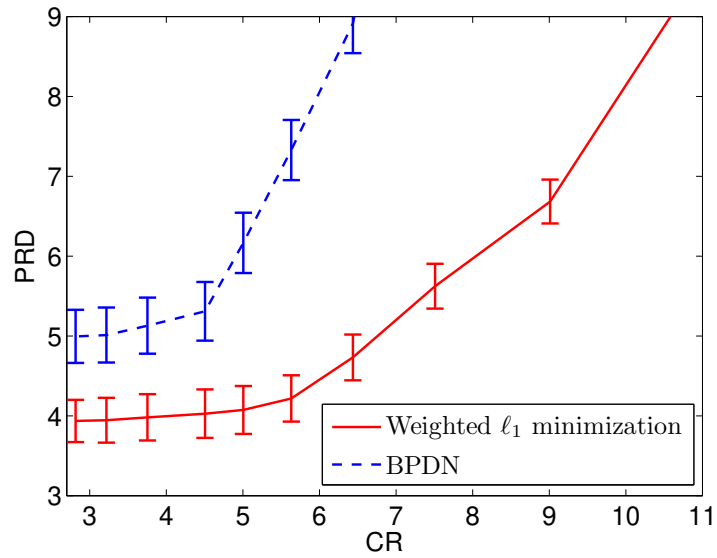


Figure 4.4: Comparison of the proposed method with BPDN. PRD as a function of the compression ratio.



## 4.5 Conclusion

This chapter proposes an ECG signal reconstruction scheme based on a weighted  $\ell_1$  minimization method. The proposed weighting scheme allows the efficient use of information on two important properties of the ECG wavelet representation: energy concentration in the approximation subband and exponential magnitude decay of the detail coefficients across scales. The proposed algorithm was evaluated for the reconstruction of a set of eleven ECG records from the MIT-BIH Arrhythmia Database encompassing a variety of signals with different rhythms, QRS complex morphologies and ectopic beats. Results show significant performance gains over the traditional basis pursuit denoising algorithm.

## Chapter 5

### COMPRESSIVE SENSING FOR ECG SIGNALS IN THE PRESENCE OF ELECTROMYOGRAPHIC NOISE

#### 5.1 Introduction

Existing works on the CS to the ECG ignore the presence of real noise that severely affects the reconstruction. Of particular interest is the electromyographic (EMG) noise, which is difficult to remove due to the significant overlapping of its spectral content with that of the ECG. Electromyographic noise is caused by the contraction of other muscles besides the heart. When other muscles in the vicinity of the electrodes contract, they generate depolarization and repolarization waves that can also be captured by the ECG sensors. Electromyographic noise is unavoidable in ambulatory ECG monitoring, which requires recording of the heart electrical activity without interrupting the daily life activities of the user. In this scenario, each movement may result in large disturbance and noise in the recorded signal. This chapter presents a CS-based method to reconstruct ECG signals in the presence of EMG noise.

A variety of algorithms have been proposed for EMG noise removal [50, 64, 83]. Since ECG signals are quasi-periodic, ensemble averaging has been proposed for EMG suppression [50]. However, this method is restricted to one particular QRS morphology at a time and requires a large number of heartbeats. A different approach for EMG noise removal employs time-varying lowpass filtering using a filter with variable frequency response [83]. A third approach for suppression of EMG interference on the ECG, based on the application of the discrete cosine transform in conjunction with singular value decomposition, is proposed in [64].

White Gaussian noise is usually employed to model EMG signals. Muscle noise however, frequently shows an impulsive behavior. Thus, the Gaussian model fails to

accurately represent the EMG noise distribution. This impulsive behavior of muscle artifacts distorts low-amplitude ECG wave components, such as the P wave. Impulsiveness is more evident in recordings acquired during exercise and sports practice [64]. As a better fitting model for EMG noise, the  $\alpha$ -stable distribution was introduced in [61]. In this chapter, the symmetric  $\alpha$ -stable ( $S\alpha S$ ) distribution is used to model the EMG interference and the proposed reconstruction algorithm is based on fractional lower-order moments, which efficiently deal with the large variance of the noise. Several studies have shown that the family of  $\alpha$ -stable distributions, and particularly the class of  $S\alpha S$  distributions, is a powerful statistical tool for modeling highly impulsive signals [46, 49].

To the best of our knowledge, the family of  $S\alpha S$  distributions has never been employed to model measurement noise in a CS framework due to the lack of closed-form expressions for the density functions and the lack of well-defined second-order statistics. Most algorithms for reconstruction of sparse signals from noisy compressive measurements provide bounds for the  $\ell_2$  reconstruction error based on the assumption that the noise is Gaussian, bounded, or has finite variance [46, 49]. However, such assumption does not apply for heavy-tailed processes exhibiting very large or infinite variance.

Recent works address the recovery of sparse signals from measurements corrupted by impulsive noise. In [18], Carrillo *et al.* proposed a non-convex reconstruction algorithm based on robust statistics theory that seeks a solution that minimizes the  $\ell_1$  norm subject to a nonlinear constraint based on the Lorentzian norm. The drawback of this method is that it is computationally slow and complex. Robust versions of CoSaMP, IHT, and OMP, using robust weighting of the residuals and robust regression, in place of least squares regression, was shown to succeed in the reconstruction of sparse signals in the presence of impulsive noise. A more recent work [45] proposes a Bayesian approach for the reconstruction of sparse signals in the presence of impulsive noise. This approach is based on Bayesian theory, robust statistics, and fast relevance vector machine methods.

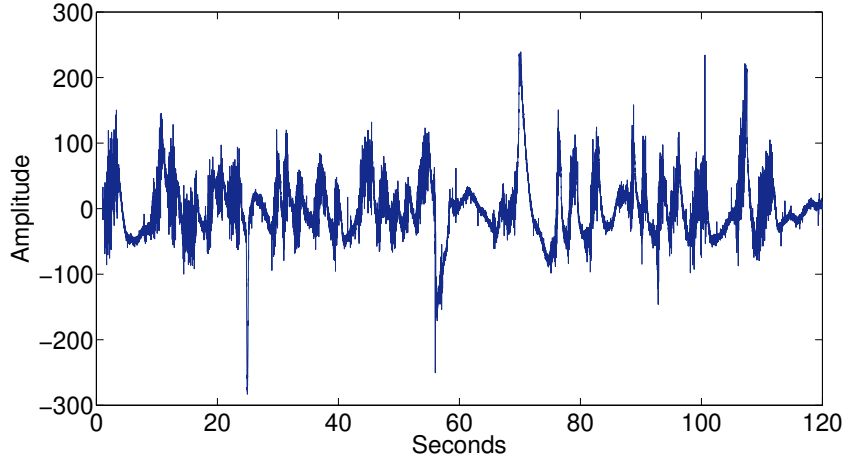
Our proposed reconstruction algorithm for ECG signals in the presence of muscle artifacts seeks a solution that minimizes the  $\ell_p$  norm of the noise subject to a constraint on the  $\ell_0$  norm of the sparse representation of the ECG, thereby defining a feasible set that diminishes the effect of gross errors. The proposed method is validated using real EMG data from the MIT-BIH Noise Stress Test Database and ECG data from the MIT-BIH Arrhythmia Database [39].

## 5.2 Background

### 5.2.1 Noise in Ambulatory ECG

Ambulatory ECG recordings are usually contaminated by both biologic and environmental sources. Examples of environmental noise include the 60 (or 50) Hz interference and its harmonics generated by power lines, instrumentation noise, and radio-frequency and electrosurgical noise. Examples of biologic interference include the baseline wander mainly caused by respiration, motion artifact, and electromyogram caused by the muscle activity.

Electromyographic noise is usually the most difficult form of noise to remove from ambulatory ECG signals. One of the reasons is that the spectrum of the EMG overlaps that of the ECG, particularly in the 0.01 Hz-100 Hz frequency range. Another difficulty is that the EMG morphology is similar to that of the P, QRS, and T waves. Figure 5.1 shows a muscle artifact record from the MIT-BIH Noise Stress Test Database of 120-second duration.



**Figure 5.1:** Muscle artifact record from the MIT-BIH Noise Stress Test Database.

### 5.2.2 Review of $S\alpha S$ Distributions

A class of  $S\alpha S$  distributions can be characterized by their distribution having a characteristic function given by the exponential

$$\exp(i\delta t - \gamma|t|^\alpha), \quad (5.1)$$

where  $\alpha$  is the characteristic exponent restricted to the range  $0 < \alpha \leq 2$ ,  $\delta$  is the real-valued location parameter,  $\gamma$  is the dispersion of the distribution that determines the spread of the density around its location parameter. The characteristic exponent  $\alpha$  controls the heaviness of the distribution tails. Thus,  $S\alpha S$  random variables with small  $\alpha$  values are highly impulsive. Note that the case of  $\alpha = 2$  corresponds to Gaussian random variables.

An important characteristic of  $S\alpha S$  distributions with  $\alpha < 2$  is the infinite second-order moments. Consequently, all moments of order  $p < \alpha$  are defined and are called fractional lower-order moments (FLOMs). In particular, the FLOMs of

$X \sim f_\alpha(\gamma_X, \delta = 0)$  are given by

$$\mathbb{E}\{|X|^p\} = (C(p, \alpha)\gamma_X)^p, \quad 0 < p < \alpha, \quad (5.2)$$

where

$$(C(p, \alpha))^p = \frac{\Gamma(1 - \frac{p}{\alpha})}{\cos(\frac{\pi}{2}p)\Gamma(1 - p)} \quad (5.3)$$

From (5.3), we get an expression for the dispersion of  $X$  of the form

$$\gamma_X = (\mathbb{E}\{|X|^p\})^{1/p}(C(p, \alpha))^{-1} \quad (5.4)$$

### 5.3 Robust Reconstruction Algorithm

#### 5.3.1 Problem Formulation

The additive noise model of the acquired ECG signal is

$$x = x_0 + n, \quad (5.5)$$

where  $n$  is the EMG noise and  $x_0$  is the original signal, represented by a linear superposition of  $K$  elements of an orthonormal wavelet basis,  $\Psi$ . That is  $x_0 = \sum_{i=1}^K s_i \psi_i = \Psi s$ . If we consider linear random measurements as in the traditional CS literature, then

$$y = \Phi x_0 + r, \quad (5.6)$$

with  $r = \Phi n$ . From the properties of  $S\alpha S$  distributions, if  $X \sim f_\alpha(\gamma_X, 0)$  and  $Y \sim f_\alpha(\gamma_Y, 0)$  are independent  $S\alpha S$  random variables, then

$$cX \sim f_\alpha(|c|\gamma_X, 0) \quad (5.7)$$

for  $c \neq 0$  and  $X + Y \sim f_\alpha((\gamma_X^\alpha + \gamma_Y^\alpha)^{1/\alpha}, 0)$ . Thus, it follows that  $r$  is also  $S\alpha S$  distributed with the same  $\alpha$  as  $n$ .

The dispersion parameter of  $S\alpha S$  distributions, defined in (5.4), is closely related to the variance of Gaussian distributions in the sense that it is also a measure of the spread of the variable. From (5.4), we know that the dispersion depends linearly on the fractional lower-order moments. Therefore, minimizing the  $p$ th FLOM of a  $S\alpha S$ -distributed random variable  $r$  results in the minimization of the dispersion of its distribution. The proposed algorithm for the reconstruction of ECG signals contaminated by EMG noise corresponds to an optimization problem that minimizes the  $\ell_p$  norm of the residual error subject to a sparsity constraint as follows

$$\min_{s \in \mathbb{R}^N} \|y - \Phi\Psi s\|_p^p \quad \text{s. t.} \quad \|s\|_0 \leq \epsilon, \quad (5.8)$$

where  $\epsilon$  is a sparsity threshold. For choosing the value of  $p$ , we use the method proposed in [86], which is based on minimizing the standard deviation of a FLOM-based covariation estimator.

### 5.3.2 An Iterative Reweighted Least Squares Approach

Iterative reweighted least squares has been successfully applied to the recovery of sparse signals [20]. In this chapter, we propose to use an iterative reweighted least squares approach to replace the  $\ell_p$  norm of equation (5.8) by a weighted  $\ell_2$  norm. First, a diagonal matrix  $W \in \mathbb{R}^{N \times N}$  is defined. Each diagonal element of  $W$  is of the form  $W_i = |r_i|^{(p-2)/2}$ , where  $r_i$  corresponds to the  $i$ th element of the residual vector  $r = y - \Phi\Psi s$ . This definition of the matrix  $W$  leads to the replacement of the  $\ell_2$  norm of the weighted residual to the  $\ell_p$  norm of the residual,  $\|Wr\|_2^2 = \|r\|_p^p$ . Therefore, problem (5.8) can be reformulated as

$$\min_{s \in \mathbb{R}^n} \|W(y - \Phi\Psi s)\|_2^2 \quad \text{s. t.} \quad \|s\|_0 \leq \epsilon, \quad (5.9)$$

Given that  $p-2$  is negative, the weights  $W_i$  are undefined whenever  $r_i = 0$ . The traditional approach for dealing with this kind of problem is to add a small  $\delta > 0$  to the residual. The same approach is adopted in this chapter. Also, note that  $0 \leq W_i < 1$

since the range of  $p$  is  $0 < p < 2$ . As expected, the weights tend to zero for large residual error components.

### 5.3.3 Algorithm

An IHT algorithm is employed to solve the problem in (5.9). Let  $s^{(0)} = 0$  and use the recursion

$$s^{(t+1)} = H_{K-L}(s^{(t)} + \mu g^{(t)}) \quad (5.10)$$

where  $H_{K-L}(\cdot)$  is a non-linear operator that sets to zero all the elements other than the  $L$  coefficients in the low-pass approximation wavelet subband together with the largest  $K - L$  remaining elements,  $\mu$  is a step size,  $t$  is the iteration index, and

$$g = -\nabla_s \|W(y - \Phi\Psi s)\|_2^2 \quad (5.11)$$

$$= (\Phi\Psi)^T W^2(y - \Phi\Psi s) \quad (5.12)$$

is the negative gradient of the objective function. The thresholding operator  $H_{K-L}(\cdot)$  ensures the selection of the subband that accumulates the majority of the signal energy at each iteration. The computational complexity of (5.10) is identical to the complexity of the traditional IHT (1.18), except for the additional cost of the multiplication by the matrix of weights.

Let  $\Gamma^{(t)}$  denote the support of  $s^{(t)}$ . It is assumed that  $\Gamma^{(t)}$  corresponds to the right support. Therefore, the objective becomes the minimization of  $\|W(y - (\Phi\Psi)_{\Gamma^{(t)}} s_{\Gamma^{(t)}})\|_2^2$ , where  $s_{\Gamma^{(t)}}$  denotes the entries of vector  $s$  corresponding to the set of indices  $\Gamma^{(t)}$ . A gradient descent approach using the iteration

$$s_{\Gamma^{(t)}}^{(t+1)} = s_{\Gamma^{(t)}}^{(t)} + \mu (\Phi\Psi)_{\Gamma^{(t)}}^T W^{(t)2}(y - (\Phi\Psi)_{\Gamma^{(t)}} s_{\Gamma^{(t)}}^{(t)}) \quad (5.13)$$

is employed to solve this minimization problem. If the support is fixed, the optimal



step size is easily calculated as the value of  $\mu$  that minimizes the objective function

$$\|W^{(t)}(y - (\Phi\Psi)_{\Gamma^{(t)}})s_{\Gamma^{(t)}}^{(t+1)}\|_2^2 \quad (5.14)$$

at each iteration. That is,

$$\mu^{(t)} = \frac{\|g_{\Gamma^{(t)}}^{(t)}\|_2^2}{\|W^{(t)}(\Phi\Psi)_{\Gamma^{(t)}}g_{\Gamma^{(t)}}^{(t)}\|_2^2}. \quad (5.15)$$

We propose to calculate the step size using (5.15) at each iteration and update  $s^{(t+1)}$  as follows

$$s^{(t+1)} = H_{K-L}(s^{(t)} + \mu^{(t)}g^{(t)}). \quad (5.16)$$

If the support of  $s^{(t+1)}$  differs from the support of  $s^{(t)}$ , then the optimality of  $\mu^{(t)}$  is no longer guaranteed. If

$$\|y - (\Phi\Psi)s^{(t+1)}\|_p^p > \|y - (\Phi\Psi)s^{(t)}\|_p^p, \quad (5.17)$$

we use a backtracking algorithm that sets  $\mu^{(t)} \leftarrow \mu^{(t)}/2$  until a decrease in the objective function of (5.8) is observed.

## 5.4 Experimental Results

Experiments are carried out over a 10-min long single-ECG lead from the MIT-BIH Arrhythmia Database. The EMG noise is taken from the MIT-BIH Noise Stress Test Database, which contains three half-hour recordings of noise typical in ambulatory ECG recordings. Simulated noisy records with different SNR values are generated by scaling the noise records from the MIT-BIH Noise Stress Test Database and adding them to the ECG signals from the MIT-BIH Arrhythmia Database.

The normalized mean square error (NMSE) and the reconstruction SNR (R-SNR) are used as performance measures. The NMSE is defined as

$$\text{NMSE} = \frac{\|x - \hat{x}\|_2^2}{\|x\|_2^2}, \quad (5.18)$$

where  $x$  and  $\hat{x}$  denote the  $N$ -dimensional original and reconstructed signals, respectively. The reconstruction SNR is defined as

$$\text{R-SNR} = 10 \log_{10} \frac{\|x\|_2^2}{\|x - \hat{x}\|_2^2}. \quad (5.19)$$

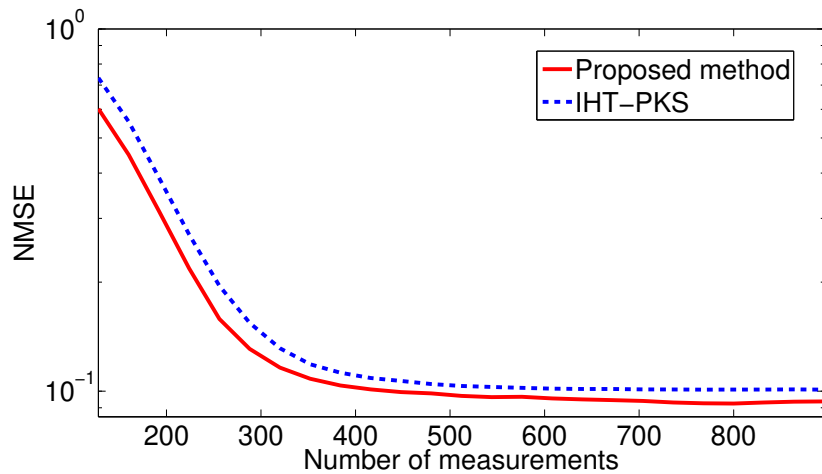
All the results are averaged over the set of records 117, 100, 115 and 119. The signal is sampled and recovered in a block-by-block manner. The length of each block is set to  $N = 1024$ . Bernoulli sensing matrices are employed and Daubechies db4 wavelets are used as the sparsifying transform. The results are averaged over one hundred different realizations of the sensing matrix.

In the first experiment, the performance of the proposed algorithm is compared with iterative hard thresholding with partially known support (IHT-PKS) [19]. Unlike the least squares IHT (1.18), IHT-PKS uses a thresholding operator  $H_{K-L}(\cdot)$  that ensures the selection of the approximation wavelet subband at each iteration. That is, the thresholding operator of IHT-PKS is the same as the thresholding operator of the proposed algorithm. The SNR of the noisy input signal is set to 10 dB while the number of samples is varied. The result of this experiment is shown in Fig. 5.2(a). The proposed method is preferred since it outperforms IHT-PKS in the reconstruction of ECG signals.

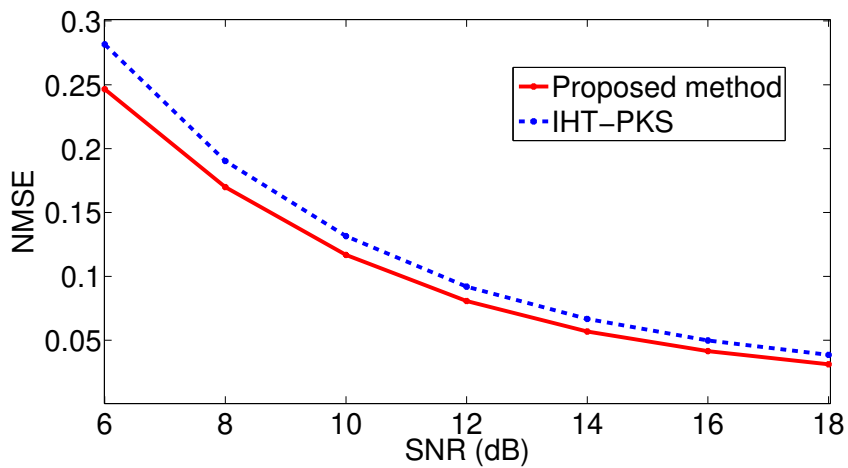
The proposed algorithm is also compared with IHT-PKS when the SNR of the input signal is varied from 6 dB to 18 dB and the number of measurements is fixed to  $M = 320$ . In Fig. 5.2(b), the proposed algorithm exhibits superior performance even when the acquired signals are highly contaminated with EMG noise.

Figure 5.3 illustrates the reconstruction of record 117 when contaminated with

EMG noise at an SNR of 16 dB. The number of measurements for this experiment is set to  $M = 640$ . Figure 5.3 (a) and (b) show the original record and the corrupted record, respectively. Figure 5.3 (c) shows the recovered signal, which attains a R-SNR of 23 dB. The proposed recovery algorithm removes the impulsiveness of the contaminated signal and attains accurate reconstruction.

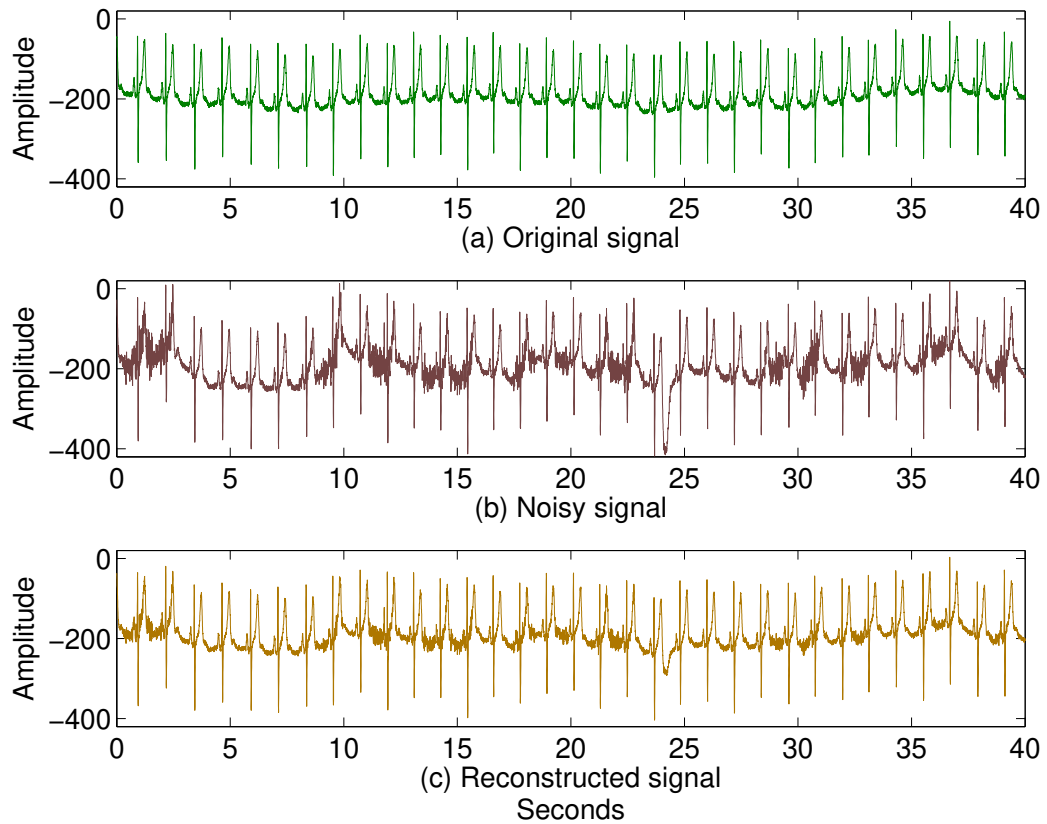


(a) Averaged NMSE over the records 117, 115, 100 and 119 as a function of the number of measurements.



(b) Averaged NMSE over the records 117, 115, 100 and 119 as a function of the SNR of the noisy input signal.

**Figure 5.2:** Comparison of the proposed method with IHT-PKS.



**Figure 5.3:** Visual evaluation of the reconstruction of record 117 using the proposed recovery algorithm.

## 5.5 Conclusions

This chapter develops an algorithm to reconstruct ECG signals contaminated by EMG noise from compressive measurements. This type of noise is a serious limitation in ambulatory ECG recording systems since each movement may result in large disturbance. Modeling the noise as symmetric  $\alpha$ -stable and using an  $\ell_p$ -based iterative hard thresholding algorithm for ECG reconstruction lead to superior recovery performance compared to the least squares iterative hard thresholding. Simulation results with ECG records from the MIT-BIH Arrhythmia Database and muscle artifacts from the MIT-BIH Noise Stress Test Database are presented.

## Chapter 6

# EXPLOITING RESTRICTED BOLTZMANN MACHINES AND DEEP LEARNING ARCHITECTURES IN COMPRESSED SENSING WIRELESS ECG SYSTEMS

### 6.1 Introduction

In this chapter, we aim at reducing the number of necessary measurements to achieve faithful ECG reconstruction by exploiting the representational power of restricted Boltzmann machines (RBMs) and deep belief networks (DBNs) to model the probability distribution of the sparsity pattern of ECG signals. The determined probability distribution is then employed in a MAP approach for the reconstruction. Exact solution of the MAP estimator can become computational unfeasible since the complexity increases exponentially with the signal length. To overcome this limitation, we propose an orthogonal matching pursuit-based algorithm to maximize the posterior distribution of the sparsity pattern. The motivation for using this approach is to capture the higher-order statistical dependencies between the coefficients of the ECG sparse representation, which in turn, leads to superior reconstruction accuracy and reduction in the number of measurements, as it is shown via experiments.

There are two main reasons for using RBMs and DBNs. First, they possess tremendous representational power and, second, inference and parameter learning are easily achieved [7, 51, 82]. Indeed, Le Roux *et al.* [51] showed that a RBM can model any discrete distribution and that adding hidden units yields strictly enhanced modeling performance unless the RBM already perfectly models the data. Similarly, Sutskever *et al.* [82] showed that deep belief networks can approximate any distribution over binary vectors to an arbitrary level of accuracy, even when the width of each layer is limited to the dimensionality of the data. Deep belief networks are the main architectures in deep

learning, the new ground breaking paradigm in artificial intelligence [7]. Therefore, this chapter links deep learning with CS by exploring the capabilities of deep learning architectures in modeling the statistical dependencies in the sparsity pattern of ECG signals.

In this chapter, overcomplete dictionaries and wavelets are selected to sparsely represent ECG signals. Even though the bulk of CS theory has been developed for signals that have a sparse representation in an orthonormal basis, recent efforts have been made to extend CS theory to signals that are sparse with respect to an overcomplete dictionary [31, 32, 78]. The coherence between the columns of an overcomplete dictionary poses some limitations in extending the CS theory to sparse overcomplete representations [27, 78]. However, Raught *et al.* [78] showed that CS is viable in the context of signals that are sparse in an overcomplete dictionary. They studied the conditions on the overcomplete dictionary that, in combination with a random sampling matrix, results in small restricted isometry constants.

Training of the dictionaries results in a set of sparse codes associated with the training data, which are often discarded after training since the main interest lies in the dictionary. Instead, we propose to utilize the sparse codes support to train either a RBM or a DBN, which is later used by a CS reconstruction algorithm that fully exploits the model. The reason for incorporating deep learning architectures into the reconstruction is to exploit higher-order dependencies between sparse coefficients. The proposed scheme falls within the structured CS framework [30], which aims at exploiting signal structure by considering more elaborate priors that go beyond the simplistic sparsity prior. Even though ECG signals have a rich structure, most of the previous CS ECG works only exploit signal sparsity.

The performance of the proposed scheme is evaluated using the MIT-BIH Arrhythmia Database [39]. Experimental results include comparisons with the BPDN algorithm, the most widely used reconstruction algorithm in the CS ECG literature [26, 56]. Experiments are also performed with the BO-BSBL algorithm [93], and with the model-based CoSaMP and model-based Iterative Hard Thresholding algorithms [75] in order



to compare with other CS reconstruction algorithms that also exploit signal structure. Simulation results indicate that the proposed algorithm offers superior reconstruction accuracy for the low-measurement regime.

## 6.2 Background on Deep Belief Networks and Restricted Boltzmann Machines

Deep learning is the new ground breaking paradigm in artificial intelligence [7]. It aims at learning hierarchical feature representations with higher level features formed by the composition of lower level features. Deep learning is inspired by biological structures in human brain mechanisms for processing of natural signals. The most extensively investigated and widely deployed deep learning architecture, the DBN [42], is presented in this section.

Restricted Boltzmann machines are the building blocks of DBNs. They are probabilistic generative models that learn a joint probability distribution of training data. A RBM is composed of a single visible layer and a single hidden layer. The visible units  $v = [v_1 v_2 \dots v_J]^T$  correspond to the input variables of the data that needs to be modeled. The hidden units  $h = [h_1 h_2 \dots h_P]^T$  are trained to capture higher-order data correlations that are observed at the visible units. Symmetric connections between the layers are represented by a weight matrix  $W$ . The structure of a RBM forms a bipartite graph, as shown in Fig. 6.1(a).

In a RBM, units within the same layer are not connected. Therefore, the posterior distribution over hidden vectors factorizes into a product of independent distributions for each hidden unit. The RBM parameters can be optimized by performing stochastic gradient ascent on the log-likelihood of training data. Given that computing the exact gradient of the log-likelihood is intractable, the contrastive divergence approximation [43] is typically employed.

In a RBM, the joint distribution  $p(v, h)$  over the visible units  $v$  and the hidden units  $h$  is defined as

$$p(v, h) = \frac{\exp(-E(v, h))}{Z}, \quad (6.1)$$

where  $E(v, h)$  is the energy function and  $Z = \sum_v \sum_h \exp(-E(v, h))$  is the normalization term. For a Bernoulli (visible)–Bernoulli (hidden) RBM, the energy function takes the form

$$E(v, h) = -v^T W h - b_v^T v - b_h^T h, \quad (6.2)$$

where  $W$  denotes the weights between visible and hidden units, and  $b_v$  and  $b_h$  are the bias terms. The conditional distributions over hidden and visible units take the form

$$p(h_j = 1|v) = \sigma((b_h)_j + W_{.j}^T v), \quad (6.3)$$

$$p(v_i = 1|h) = \sigma((b_v)_i + W_{i.} h), \quad (6.4)$$

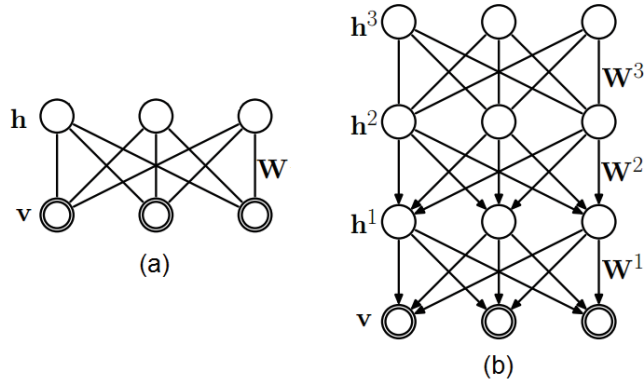
where  $\sigma(x) = (1 + e^{-x})^{-1}$ , and  $W_{.j}$  and  $W_{i.}$  correspond to the  $j$ th column and  $i$ th row of matrix  $W$ , respectively.

A DBN architecture is composed of a stack of RBMs. The lowest-level RBM learns a shallow model of the data. The RBM at the next level captures high-order correlations between the hidden units of the first, and so on. A DBN with  $L$  layers models the joint distribution between the visible layer  $v$  and the hidden layers  $h^l$ ,  $l = 1, \dots, L$  as follows

$$p(v, h^1, \dots, h^L) = p(v|h^1) \left( \prod_{l=1}^{L-2} p(h^l|h^{l+1}) \right) p(h^{L-1}, h^L). \quad (6.5)$$

The log-probability of the training data can be improved by adding layers to the network, which, in turn, increases the true representational power of the network [42].

Let  $v = h^0$ . The bias vector of layer  $l$  and the weight matrix that represents the connections between layer  $l - 1$  and layer  $l$  are denoted by  $b_{h^l}$  and  $W^l$ , respectively. A schematic representation of a DBN with one visible and three hidden layers is shown in Fig. 6.1(b). The top two layers form a restricted Boltzmann machine, which is an undirected graphical model, and the lower layers form a directed generative model.

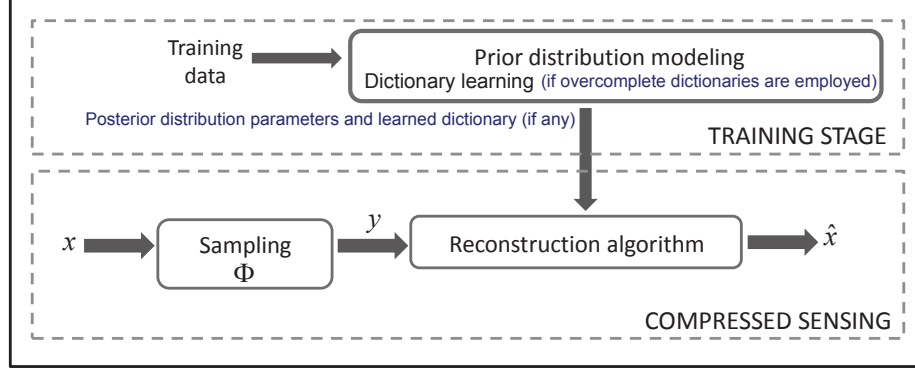


**Figure 6.1:** (a) Schematic of a restricted Boltzmann machine. (b) Schematic of a deep belief network of one visible and three hidden layers.

The main breakthrough introduced by Hinton *et al.* [42] was a greedy, layer-wise unsupervised learning algorithm that allows efficient training of DBNs. This algorithm trains each RBM separately, making the time complexity of the DBN learning linear to the size and depth of the networks.

### 6.3 Proposed Compressed Sensing Scheme

The proposed scheme requires training data of the same class as the signal to be reconstructed. A training stage is employed to learn a prior model for the sparsity pattern of the signal class. The proposed CS reconstruction algorithm exploits the determined prior in a MAP approach. The CS and training stages are described thoroughly in this section. The training stage varies depending on the employed sparsifying transform, either orthonormal bases or overcomplete learned dictionaries. The block diagram of the proposed CS scheme is presented in Fig. 6.2.



**Figure 6.2:** Block diagram of the proposed CS scheme.

### 6.3.1 Compressed Sensing Stage

#### 6.3.1.1 Problem formulation

Let  $D \in \mathbb{R}^{N \times Q}$  denote the sparsifying transform employed to represent a signal  $x \in \mathbb{R}^N$ , *i.e.*  $x = Ds + r$ , where  $s$  and  $r$  are the sparse representation and the representation error, respectively. A Gaussian distribution with zero mean and covariance  $\Sigma_r$  is assumed for  $r$ . In this thesis, we consider the traditional synthesis-based CS approach that aims at reconstructing the sparse representation  $s$  of a signal  $x$  from undersampled and noisy measurements of the form  $y = \Phi x + n$ , where  $\Phi \in \mathbb{R}^{M \times N}$  is the sampling matrix and  $n$  accounts for the additive Gaussian sampling noise of zero mean and variance  $\sigma_n^2$ . Vector  $y$  can also be written as

$$y = \Phi Ds + \Phi r + n. \quad (6.6)$$

Let  $\eta = \Phi r + n$  and  $\Xi = \Phi D$ , then vector  $y$  takes the form

$$y = \Xi s + \eta. \quad (6.7)$$

As both  $r$  and  $n$  are Gaussian distributed, vector  $\eta$  is also Gaussian distributed with zero mean and covariance  $\Sigma_\eta = \Phi\Sigma_r\Phi^T + \sigma_n^2 I$ . We adopt the commonly used assumption that the sampling noise variance  $\sigma_n^2$  is known [65, 76, 92].

The approach proposed in [65] is adopted in this thesis, namely first calculating the MAP estimator of the sparsity pattern and then calculating the MAP estimator of the sparse vector. The support of  $s$ , of cardinality  $K$ , is denoted as  $\theta$ . Let  $s_\theta$  denote the nonzero coefficients of  $s$ . A Gaussian distribution with zero mean and variance  $\sigma_{s_i}^2$  is assumed for each nonzero coefficient  $s_i$ ,  $i \in \theta$ . The same probability distribution is employed in [37, 65] for nonzero sparse coefficients. Then, the conditional distribution of  $s_\theta$  given  $\theta$  is given by

$$s_\theta|\theta \sim \mathcal{N}(\mathbf{0}, \Sigma_\theta), \quad (6.8)$$

where  $\Sigma_\theta \in \mathbb{R}^{K \times K}$  is a diagonal matrix, whose diagonal is formed by the variances of the nonzero coefficients  $\sigma_{s_i}^2$ ,  $i \in \theta$ .

The Gaussian distribution of  $\eta$  leads to the following distribution for the likelihood  $p(y|s_\theta, \theta)$ :

$$y|s_\theta, \theta \sim \mathcal{N}(\Xi_\theta s_\theta, \Sigma_\eta). \quad (6.9)$$

By integrating the product of  $p(s_\theta|\theta)$  and  $p(y|s_\theta, \theta)$  over all possible  $s_\theta$ , an expression for the probability distribution  $p(y|\theta)$  is obtained,

$$p(y|\theta) = C \times \det(\Xi_\theta^T \Sigma_\eta^{-1} \Xi_\theta \Sigma_\theta + I)^{-1/2} \exp\left\{\frac{1}{2} y^T \Sigma_\eta^{-1} \Xi_\theta P^{-1} \Xi_\theta^T \Sigma_\eta^{-1} y\right\}, \quad (6.10)$$

where  $C = \det(2\pi\Sigma_\eta)^{-1/2} \exp\{-\frac{1}{2} y^T \Sigma_\eta^{-1} y\}$  and  $P = \Xi_\theta^T \Sigma_\eta^{-1} \Xi_\theta + \Sigma_\theta^{-1}$ .

The MAP estimator of  $\theta$ , denoted by  $\hat{\theta}$ , can be calculated as

$$\hat{\theta} = \arg \max_{\theta} p(\theta|y) = \arg \max_{\theta} p(y|\theta)p(\theta). \quad (6.11)$$

The posterior distribution  $p(s_\theta|y, \hat{\theta})$  has a Gaussian distribution with mean  $\mu_s$  and

covariance  $\Sigma_s$ , such that

$$\mu_s = \Sigma_{\hat{\theta}} \Xi_{\hat{\theta}}^T (\Xi_{\hat{\theta}} \Sigma_{\hat{\theta}} \Xi_{\hat{\theta}}^T + \Sigma_{\eta})^{-1} y \quad (6.12)$$

$$\Sigma_s = \Sigma_{\hat{\theta}} - \Sigma_{\hat{\theta}} \Xi_{\hat{\theta}}^T (\Xi_{\hat{\theta}} \Sigma_{\hat{\theta}} \Xi_{\hat{\theta}}^T + \Sigma_{\eta})^{-1} \Xi_{\hat{\theta}} \Sigma_{\hat{\theta}}. \quad (6.13)$$

Therefore, the MAP estimate of  $s$ , denoted as  $\hat{s}_{\hat{\theta}}$ , is directly obtained from the mean of the posterior, *i.e.*,

$$\begin{aligned} \hat{s}_{\hat{\theta}} &= \arg \max_{s_{\hat{\theta}}} p(s_{\hat{\theta}} | y, \hat{\theta}), \\ &= \Sigma_{\hat{\theta}} \Xi_{\hat{\theta}}^T (\Xi_{\hat{\theta}} \Sigma_{\hat{\theta}} \Xi_{\hat{\theta}}^T + \Sigma_{\eta})^{-1} y. \end{aligned} \quad (6.14)$$

An expression for  $p(\theta)$  needs to be calculated to solve for (6.11). Define the sparsity pattern  $S^{\{\theta\}}$  associated with the support  $\theta$  as  $S_i^{\{\theta\}} = \mathbf{1}_{\theta}(i)$  for  $i = 1, \dots, N$ , where  $\mathbf{1}_{\theta}(\cdot)$  denotes the indicator function of  $\theta$ , and note that  $p(\theta) = p(S^{\{\theta\}})$ . We propose to use the probability distribution over visible units  $p(v)$  of restricted Boltzmann machines (RBMs) and deep belief networks (DBNs) to model the prior distribution  $p(S^{\{\theta\}})$ , or equivalently,  $p(\theta)$ .

### 6.3.1.2 Prior distribution

In a RBM, the probability distribution over visible units is obtained by marginalizing (6.1) over the hidden units

$$p(v) = \int p(v, h) dh = \frac{1}{Z} \exp(-E(v)), \quad (6.15)$$

where

$$E(v) = - \sum_j \log \left( 1 + e^{W_{\cdot j}^T v + b_{h_j}} \right) - b_v^T v. \quad (6.16)$$

A DBN can be seen as a probabilistic generative model. To calculate the probability distribution of the visible units of a DBN, we start with a random configuration

at the top hidden layer  $h^L$  and let the top-level RBM converge to a stationary distribution using alternating Gibbs sampling. Alternating Gibbs sampling iterates between updating the hidden units in parallel using (6.3) and updating the visible units in parallel using (6.4). Next, it performs a top-down pass in which the state of each variable in a layer is chosen from a Bernoulli distribution with the probability that a variable has a value of one depending on the states of the layer above. That is,

$$p(h_i^l = 1|h^{l+1}) = \sigma((b_{h^l})_i + W_{i \cdot}^{l+1}h^{l+1}), \quad (6.17)$$

where as before,  $v = h^0$ .

Repeated top-down passes generate a full set of data vectors at each layer of the DBN. Let  $H$  be the total number of top-down passes. The sequence of data vectors for the hidden layers is denoted as  $h^{l(1)}, \dots, h^{l(H)}$ ,  $l = 1, \dots, L$  and  $v^{(1)}, \dots, v^{(H)}$  for the visible layer. Such sequence of data vectors assigned to the visible layer can be employed to give a rough approximation of the marginal distribution  $p(v)$ . However, the conditional density function  $p(v|h^1)$  contains more information about the shape of the distribution  $p(v)$  than the sequence of the individual realizations since  $p(v) = E_{h^1}[p(v|h^1)]$ , where  $E[\cdot]$  denotes the expected value of a random variable. Therefore, the marginal density is approximated by  $\hat{p}(v) = \frac{1}{H} \sum_{k=1}^H p(v|h^{1(k)})$ . Since the visible units are conditionally independent given the hidden states  $h^1$ , the approximation of  $p(v)$  takes the form

$$\hat{p}(v) = \frac{1}{H} \sum_{k=1}^H \prod_{i=1}^N p(v_i|h^{1(k)}). \quad (6.18)$$

Using the probability distribution (6.15) to model  $p(\theta)$  in (6.11), leads to the following MAP estimator of  $\theta$ :

$$\hat{\theta} = \arg \max_{\theta} \left( \frac{1}{2} y^T \Sigma_{\eta}^{-1} \Xi_{\theta} P^{-1} \Xi_{\theta}^T \Sigma_{\eta}^{-1} y - \frac{1}{2} \log(\det(P \Sigma_{\theta})) + \sum_j \log \left( 1 + e^{W_{\cdot j}^T v + b_{h_j}} \right) + b_v^T S\{\theta\} \right). \quad (6.19)$$

Similarly, when using (6.18) to model  $p(\theta)$  in (6.11), the MAP estimator of  $\theta$  becomes

$$\hat{\theta} = \arg \max_{\theta} \left( \frac{1}{2} y^T \Sigma_{\eta}^{-1} \Xi_{\theta} P^{-1} \Xi_{\theta}^T \Sigma_{\eta}^{-1} y - \frac{1}{2} \log(\det(P \Sigma_{\theta})) \right. \\ \left. + \log \left( \sum_{k=1}^H \prod_{i=1}^N p \left( S_i^{\{\theta\}} | h^{1(k)} \right) \right) \right) \quad (6.20)$$

### 6.3.1.3 MAP estimator via a greedy approach

Peleg *et al.* [65] proposed a greedy pursuit algorithm based on OMP to approximate the MAP estimator of a sparse representation. The same algorithm is employed for reconstructing the sparse representation  $s$ , although using a different posterior distribution than that in [65]. The support is initialized to the empty set. At each iteration, the algorithm searches for the element  $\bar{i}$  that can be added to the support in order to maximize  $p(\theta|y)$ . The algorithm stops when the number of iterations exceeds the pre-defined sparsity level. Once the support is recovered, the sparse representation  $s$  is calculated using (6.14). A summary of the algorithm is presented in Algorithm 5. The functions  $g_{\text{RBN}}(\cdot)$  and  $g_{\text{DBN}}(\cdot)$  refer to the objective functions in (6.19) and (6.20), respectively. If a RBM is employed to model the prior distribution  $p(\theta)$ , the function  $g_{\text{RBN}}(\cdot)$  is used and the algorithm is referred to as the RBM-OMP-like algorithm. If a DBN is employed instead, the function  $g_{\text{DBN}}(\cdot)$  is used and the algorithm is referred to as the DBN-OMP-like algorithm.



---

**Algorithm 5** RBM-OMP-like/DBN-OMP-like algorithm

---

**Require:** Matrix  $\Xi = \Phi D$ , measurements  $y$ , model parameters defining the probability distribution  $p(\theta|y)$ .

- 1: Initialize  $t = 0$ ,  $\theta^{(0)} = \emptyset$ .
  - 2: **while** halting criterion false **do**
  - 3:    $t \leftarrow t + 1$
  - 4:   **for**  $i \notin \theta^{(t-1)}$  **do**
  - 5:      $\bar{\theta}^{(t)} \leftarrow \theta^{(t-1)} \cup i$
  - 6:      $f(i) \leftarrow g_{\text{RBN}}(\bar{\theta}^{(t)})$  or  $f(i) \leftarrow g_{\text{DBN}}(\bar{\theta}^{(t)})$
  - 7:   **end for**
  - 8:    $\bar{i} \leftarrow \arg \max_i f(i)$
  - 9:    $\theta^{(t)} \leftarrow \theta^{(t-1)} \cup \bar{i}$
  - 10: **end while**
  - 11:  $\hat{\theta} \leftarrow \theta^{(t)}$
  - 12: **return**  $\hat{s}_{\hat{\theta}} \leftarrow \Sigma_{\hat{\theta}} \Xi_{\hat{\theta}}^T (\Xi_{\hat{\theta}} \Sigma_{\hat{\theta}} \Xi_{\hat{\theta}}^T + \Sigma_{\eta})^{-1} y$ .
- 

### 6.3.2 Training Stage using Overcomplete Learned Dictionaries

In this section, the sparsifying transform  $D$  from (6.6) is assumed to be an overcomplete dictionary. We learn  $D$  from a set of ECG training data. The resulting sparse codes and the representation error are employed to estimate the model parameters defining  $p(\theta|y)$ .

#### 6.3.2.1 Traditional dictionary learning

First, the dictionary learning problem is described. Let  $G = [g^1 \dots g^B] \in \mathbb{R}^{N \times B}$  denotes the set of  $N$ -dimensional training samples, which is referred to as training data set I. One methodology for building the overcomplete dictionary  $D = [d^1 \dots d^J] \in \mathbb{R}^{N \times J}$  ( $J > N$ ) is to solve the following optimization problem

$$\{\hat{D}, \hat{A}\} = \arg \min_{D, A} \|G - DA\|_F^2 \quad \text{s. t.} \quad \|a^j\|_0 < K, \quad \forall j \quad (6.21)$$

with  $A = [a^1 \dots a^B] \in \mathbb{R}^{J \times B}$  denoting the sparse codes of  $G$ , and  $K$  the pre-specified sparsity level. The representation error is defined as  $E = G - \hat{D}\hat{A}$ . Several algorithms have been proposed to solve (6.21). Here, the K-SVD algorithm proposed by Aharon

*et al.* [1] is employed.

### 6.3.2.2 Estimation of model parameters

This section addresses the parameter estimation of  $p(\theta|y)$ , which comprises the parameter estimation of  $p(\theta)$ , the estimation of the variances  $\sigma_{s_i}^2, \forall i$ , and the estimation of the covariance  $\Sigma_\eta$ . Let  $u^j$  denote the sparsity pattern of the sparse code  $\hat{a}^j, j = 1, \dots, B$ . The  $i$ th element of  $u^j$  is defined as  $u_i^j = \mathbf{1}_{\text{supp}(\hat{a}^j)}(i)$ , where  $\text{supp}(\hat{a}^j)$  denotes the support of  $\hat{a}^j$ . The set of vectors  $U = [u^1 \dots u^B]$  can be employed to model a prior distribution for the sparsity pattern of signals belonging to the same class as the training data. Since the signal to be reconstructed,  $x$ , also belongs to the class of ECG signals, the set of column vectors of  $U$ , which are referred to as training data set II, are used to learn the parameters of  $p(\theta)$ . As mentioned in Section 6.3.1.1, RBMs and DBNs are used for modeling such a prior distribution.

In the case of the RBM model, the probability distribution parameters that need to be estimated are the weight matrix  $W$  and the bias terms  $b_v$  and  $b_h$ . They can be optimized by performing stochastic gradient ascent on the log-likelihood of the training data set II. However, computing the exact gradient of the log-likelihood is intractable. Here, we use contrastive divergence, which approximates the gradient of the log-likelihood by using a Markov chain. For more details on contrastive divergence, the reader is referred to [43].

When the prior distribution of the sparsity pattern is modeled with a DBN, the parameters that need to be estimated are the weights and bias terms of each layer. Hinton recently presented a powerful greedy layer-wise method to learn these parameters [42]. The weight matrix  $W^1$ , the bias terms of the visible layer  $b_v$ , and the bias terms of the lowest hidden layer  $b_{h^1}$  are learned by training a RBM with the training data set II as input. Then, the inferred hidden values of  $h^1$  can be used as input for training another RBM that learns the parameters of the layer above. This bottom-up process is repeated at the next layers until all the parameters of the network are learned.

Additionally, the set of sparse codes  $\hat{A}$  can be employed to estimate the variance  $\sigma_{s_i}^2$  of each  $i$ th element of the sparse representation  $s$  of signal  $x$  (see [65]):

$$\hat{\sigma}_{s_i}^2 = \frac{\sum_{j=1}^B (\hat{a}_i^j)^2}{\sum_{j=1}^B \mathbf{1}_{\text{supp}(\hat{a}^j)}(i)}. \quad (6.22)$$

For the estimation of the covariance matrix  $\Sigma_r$ , independence is assumed between the representation error coefficients  $r_i$  and  $r_j$  for  $i \neq j$ . Therefore, the covariance matrix  $\Sigma_r$  is a diagonal matrix, whose diagonal is formed by the variances of the representation error coefficients  $\sigma_{r_i}^2$ , for  $i = 1, \dots, N$ . For the estimate of  $\Sigma_r$ , denoted as  $\hat{\Sigma}_r$ , the representation error of the learned dictionary  $E = [e^1 \dots e^B]$  is employed. More precisely, each  $i$ th diagonal element of  $\Sigma_r$  is estimated as

$$\hat{\sigma}_{r_i}^2 = \frac{1}{B} \sum_{j=1}^B (e_i^j)^2. \quad (6.23)$$

The estimate of  $\Sigma_\eta$  is directly calculated as  $\hat{\Sigma}_\eta = \Phi \hat{\Sigma}_r \Phi^T + \sigma_n^2 I$ .

### 6.3.3 Training Stage using Orthonormal Bases

As in Section 6.3.2.1, let  $G = [g^1 \dots g^B] \in \mathbb{R}^{N \times B}$  denote the set of  $N$ -dimensional training ECG samples. In this section,  $D$  does not denote an overcomplete dictionary, but instead, it denotes an orthonormal basis. Each vector  $g^j$  can be expressed as  $g^j = D\bar{a}^j$ , where  $\bar{a}^j$  is the representation of the signal  $g^j$  in the  $D$  domain. Let  $a^j$  denote the best  $K$ -term approximation of  $\bar{a}^j$ , which is obtained by keeping only the  $K$  largest (in magnitude) coefficients in  $\bar{a}^j$  and setting the others to zero. Therefore, the signal  $g^j$  can be modeled as  $g^j = Da^j + e^j$ , where  $e^j$  is the representation error. Let  $u^j$  denote the sparsity pattern of the sparse code  $a^j$ ,  $u_i^j = \mathbf{1}_{\text{supp}(a^j)}(i)$ . As in the case of overcomplete dictionaries, the sparse codes  $A = [a^1 \dots a^B]$ ,  $U = [u^1 \dots u^B]$ , and the representation error  $E = [e^1 \dots e^B]$  are used to learn the model parameters defining  $p(\theta|y)$ . That is, the set of vectors  $U$  is used to train either the RBM or the DBN that

models  $p(\theta)$ , the sparse codes  $A$  are used to estimate the variances  $\sigma_{s_i}^2, \forall i$ , using (6.22), and  $E$  is used to estimate  $\sigma_{r_i}^2, \forall i$ , using (6.23).

Unlike the case of overcomplete dictionaries, the training stage does not require optimization of the sensing matrix. For the case of orthonormal bases, the entries of the sampling matrix  $\Phi \in \mathbb{R}^{M \times N}$  are independently sampled from a normal distribution with mean zero and variance  $1/M$ .

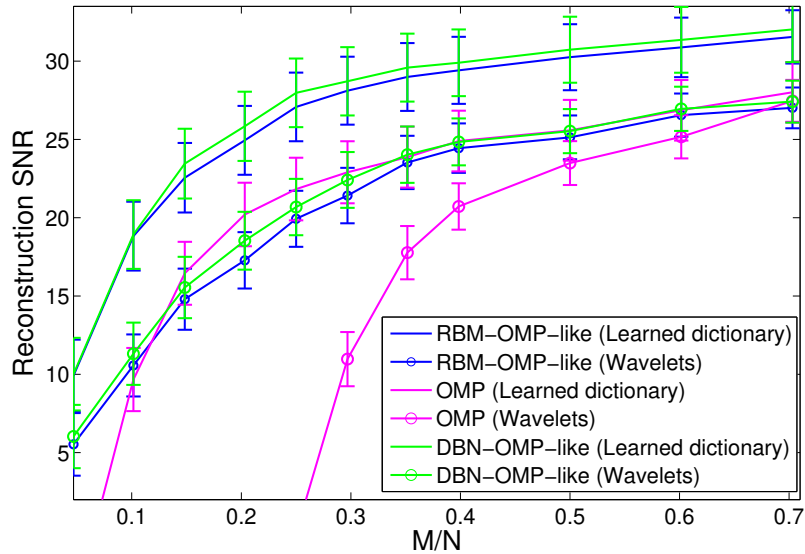
## 6.4 Experimental Results

To validate the potential of the proposed algorithms to improve the performance of CS ECG systems, experiments are performed on records from the MIT-BIH Arrhythmia Database [39]. Single leads from records 100, 101, 102, 107, 109, 111, 115, 117, 118 and 119 are employed for the experiments. This data set consists of a variety of signals with different morphologies, rhythms and abnormal heartbeats.

The entries of the sampling matrix  $\Phi$  are independently sampled from a symmetric Bernoulli distribution ( $P(\Phi_{i,j} = \pm 1/\sqrt{M}) = 1/2$ ) in order to build an efficient hardware implementation. The use of Bernoulli matrices, as compared to other sub-Gaussian matrices, results in simpler circuit complexity, data storage, and computation requirements [21]. Presented results correspond to averages of 50 repetitions of each experiment, with a different realization of the random measurement matrix at each time. Both the wavelet transform and overcomplete learned dictionaries are employed as sparsifying transforms in the experiments. Specifically, the Daubechies-4 wavelet transform, using a decomposition level  $L = 4$ , is employed. The reconstruction SNR, the PRD, and the CR are used as performance measures for our experiments.

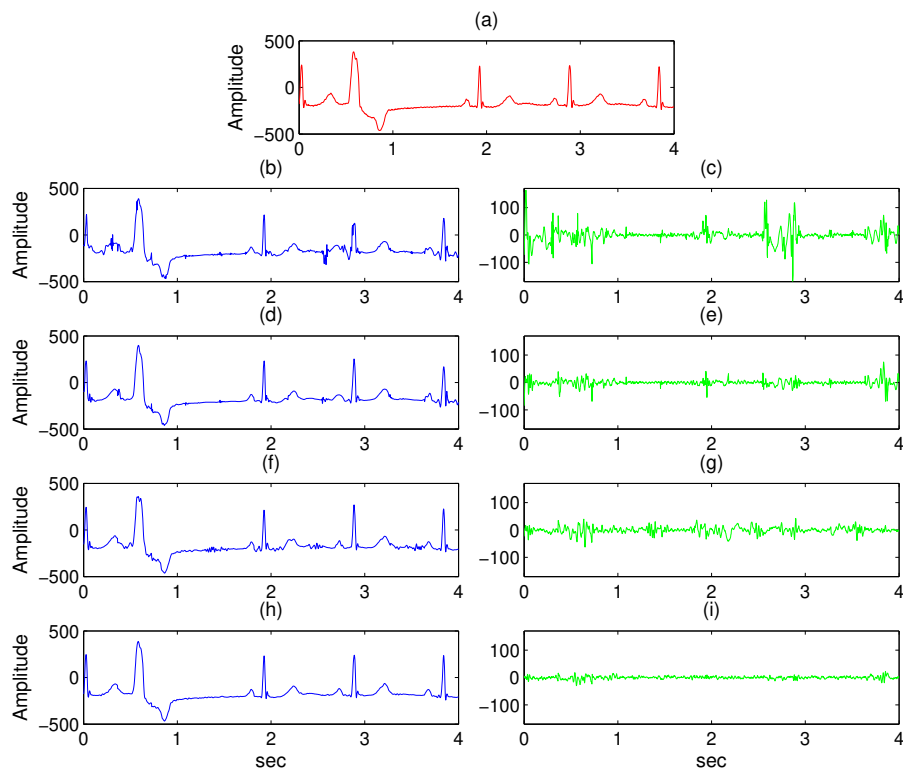
### 6.4.1 Performance Evaluation

The first experiment compares the ECG reconstruction performance of the RBM/DBN-OMP-like algorithms with the traditional OMP algorithm. The training and testing data sets for this experiment consist of 92800 and 5000 segments of size  $N = 128$ , extracted from the selected ECG records, respectively. The same number of



**Figure 6.3:** Comparison of the reconstruction of ECG signals using the OMP, the RBM-OMP-like algorithm, and the DBN-OMP-like algorithm for both wavelets and overcomplete learned dictionaries.

segments is extracted from each record. A RBM with the same number of hidden units as of visible units and a 2-layer DBN are employed to model the probability distribution of the sparsity pattern. The number of hidden units per layer of the DBN is set half the number of visible units. Compressed measurements are artificially contaminated with Gaussian noise of variance  $\sigma_n^2 = 0.25$ . The number of dictionary atoms is set to  $3N$ . The sparsity level is set to  $K = 0.08N$  and  $K = 0.1N$  for overcomplete dictionaries and wavelets, respectively. The results of the comparison in Fig. 6.3 indicate that the proposed algorithms have superior reconstruction performance than OMP and require significantly less number of measurements to achieve accurate reconstruction. The performance gain is larger in the case of overcomplete dictionaries than in the case of wavelets. Even though the DBN uses only half the number of hidden units per layer of the RBM, and therefore requires less number of parameters to be learned, the performance of the DBN-OMP-like algorithm is slightly superior to that of the RBM-like



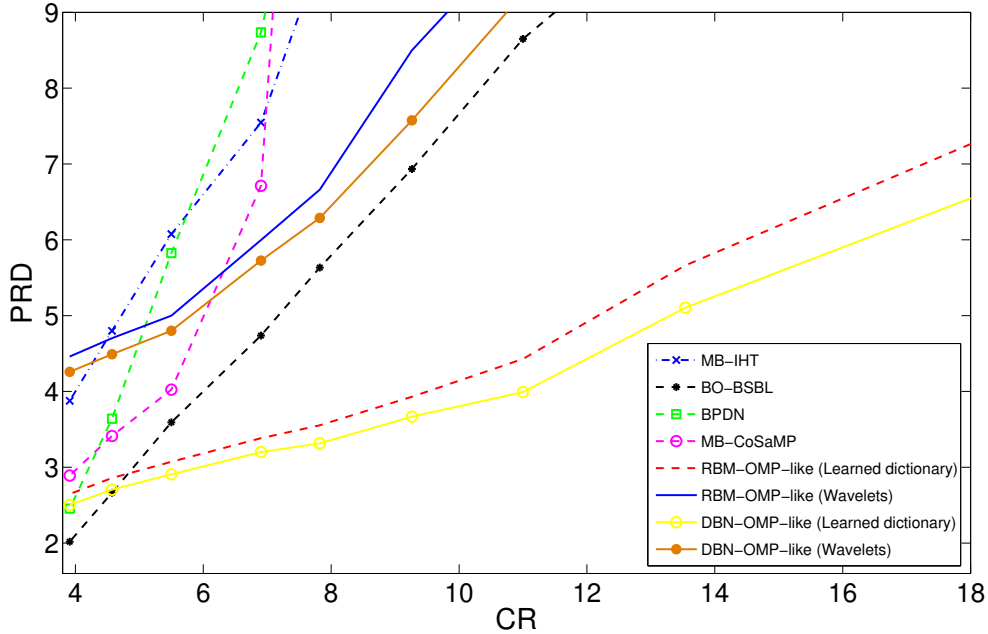
**Figure 6.4:** Visual evaluation of the RBM-OMP-like algorithm. Record 119,  $M = 0.35N$ . First row: (a) Original signal. Second and third rows: (b-e) Reconstructed signals using wavelets (b-c) OMP reconstruction and error, R-SNR=18.32, (d-e) RBM-OMP-like reconstruction and error, R-SNR=25. Fourth and fifth rows: (f-i) Reconstructed signals using an overcomplete dictionary (f-g) OMP reconstruction and error, R-SNR=24.59, (h-i) RBM-OMP-like reconstruction and error, R-SNR=31.67.

algorithm due to the multi-layer structure of the DBN. As the reconstruction performance is related to the representational power of the model, the results in Fig. 6.3 are consistent with the recent work by Le Roux et al. [51], which shows that, even though a DBN and a RBM can have the same representational power, a DBN offers a more compact representation in terms of number of parameters. The error bars in Fig. 6.3 indicate plus and minus one standard deviation across ECG records.

Fig. 6.4 visually illustrates the reconstruction of an ECG signal using the RBM-OMP-like algorithm. Fig. 6.4(a) corresponds to a 4-second duration segment from

record 119, which contains ventricular ectopic beats. The signal is divided into segments of 128 samples. Each segment is sampled and reconstructed separately. Then, the segments are concatenated to reconstruct the original signal. For this experiment, the number of measurements is set to  $M = 0.35N$ . Figs. 6.4(b-e) and Figs. 6.4(f-i) are obtained when wavelets and learned overcomplete dictionaries are employed as the sparsifying transform, respectively. Figs. 6.4(b-c) and Figs. 6.4(f-g) show the obtained reconstructed (left) and error (right) signals when using the OMP-like algorithm. Similarly, Figs. 6.4(d-e) and Figs. 6.4(h-i) show the obtained reconstructed (left) and error (right) signals when using the RBM-OMP-like algorithm. The recovered signals using the RBM-OMP-like algorithm are better estimates of the original signals than those obtained with the traditional OMP algorithm. It should be noted that using overcomplete dictionaries leads to preservation of detailed information for clinical diagnosis and less number of artifacts in the reconstruction.

The second experiment aims at comparing the performance of the RBM/DBN-OMP-like algorithms with previously applied CS algorithms to the problem of ECG reconstruction, such as BPDN [26,56], model-based CoSaMP, denoted as MB-CoSaMP, model-based Iterative Hard Thresholding, denoted as MB-IHT [75], and the BO-BSBL algorithm [93]. For this experiment, a quantization stage and Huffman coding are added to the scheme presented in Section 6.3. An 8-bit optimal scalar quantizer designed with the Lloyd-Max algorithm is utilized. Results of the obtained PRD values as a function of the compression ratio are shown in Fig. 6.5. Wavelets are employed as the sparsifying transform in all the reconstruction algorithms. The RBM/DBN-OMP-like algorithms are the only algorithms that use both wavelets and overcomplete dictionaries in this experiment. The second experiment uses the same training and testing datasets of the first experiment. For  $CR > 6$ , the proposed algorithms using wavelets outperform MB-CoSaMP, MB-IHT, and the BPDN algorithms. The RBM/DBN-OMP-like algorithms using an overcomplete dictionary exhibit the best performance for  $CR > 5$ . It is only outperformed by the BO-BSBL algorithm for low CR values ( $CR < 5$ ), which is not the case of interest for WBAN-enabled ECG monitors.



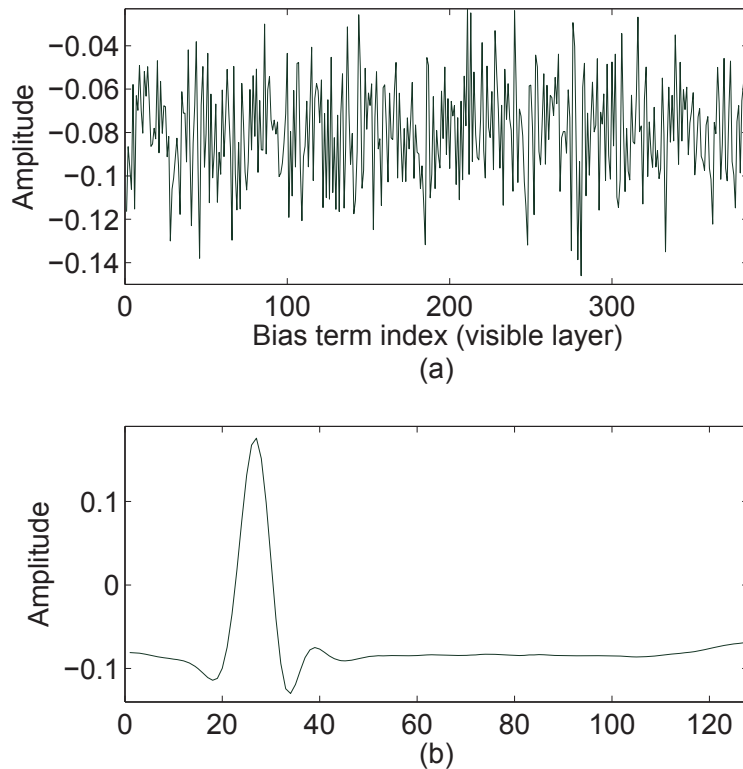
**Figure 6.5:** Comparison of the reconstruction of ECG signals using the RBM-OMP-like algorithm, the DBN-OMP-like algorithm, MB-CoSaMP, MB-IHT, and the BO-BSBL algorithm. PRD as a function of the compression ratio.

### 6.4.2 Qualitative Observations

This section presents a qualitative assessment of the trained RBM models. Fig. 6.6(a) illustrates the visible layer bias terms of the RBM using learned over-complete dictionaries. Note that the bias terms are negative since the elements of the sparse representation are zero most of the time. Fig. 6.6(b) illustrates the dictionary atom sharing the same index as the largest bias term. Note the resemblance of the dictionary atom with the QRS complex of an ECG cycle.

Fig. 6.7(a) illustrates the visible layer bias terms of the RBM using wavelets. Most of the bias terms are negative, with the exception of 6 elements. Fig. 6.7(b) illustrates the magnitude of the wavelet representation of a segment from record 117, where the dashed lines separate each wavelet subband. Sparse representation coefficients sharing the same indexes as the largest bias terms of the visible layer are likely

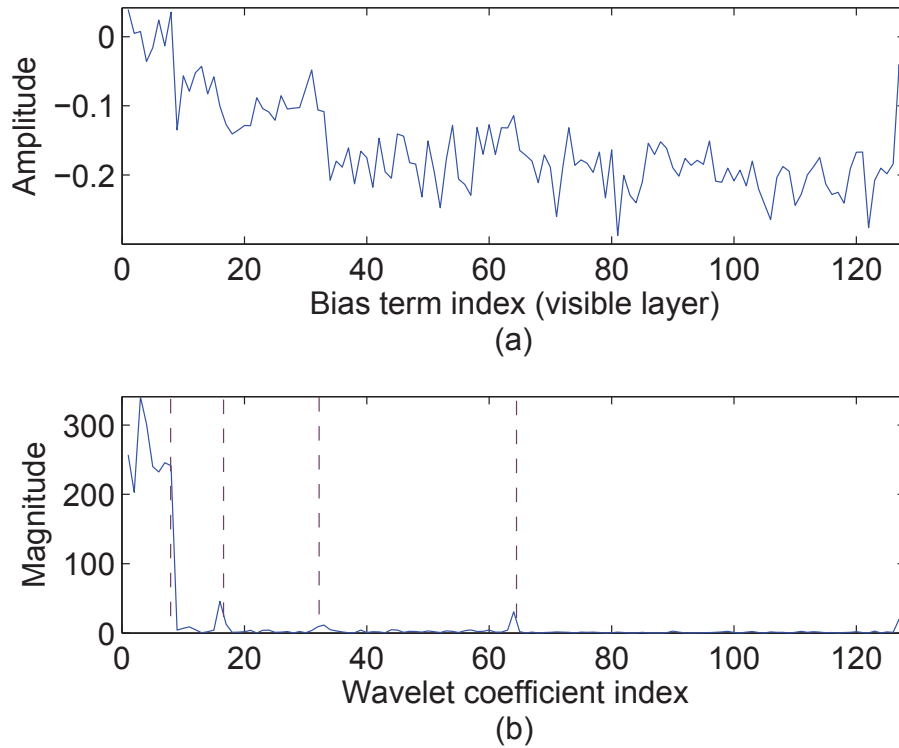




**Figure 6.6:** (a) Visible layer bias terms of the RBM model using learned overcomplete dictionaries. (b) Dictionary atom sharing the same index as the largest bias term.

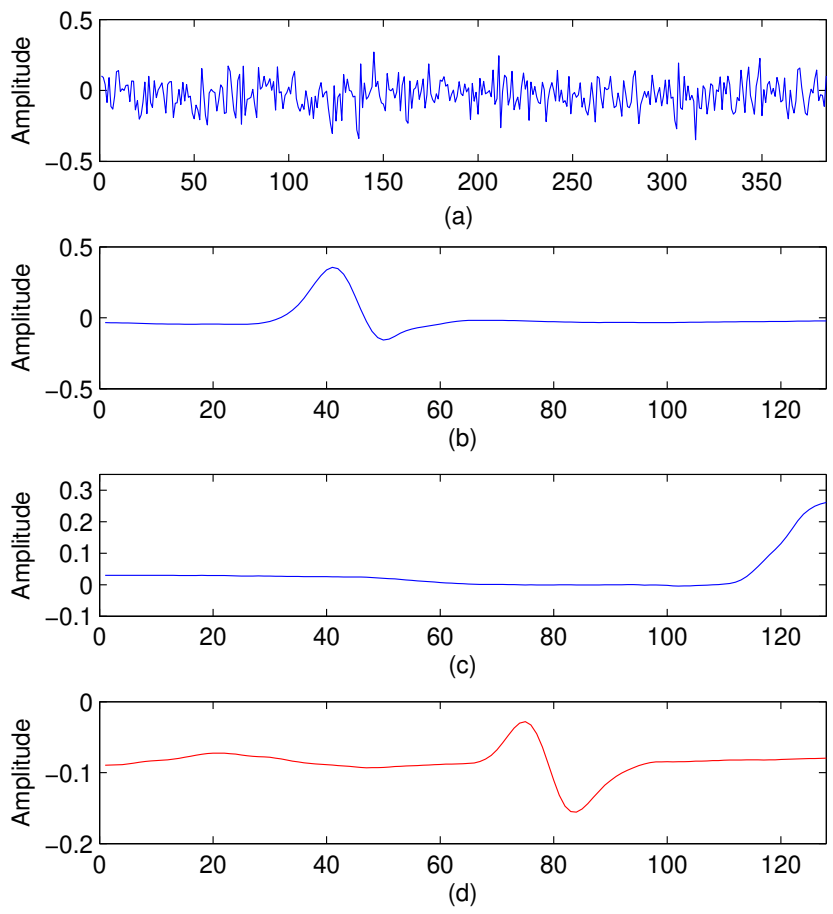
to belong to the best  $K$ -term sparse approximation. This is consistent with the result in Fig. 6.7(a) showing that the largest bias term indexes correspond to the indexes of the scaling wavelet coefficients (8 first indexes of the wavelet representation), which accumulate most of the ECG signal energy [70], and therefore, are likely to belong to the best  $K$ -term sparse approximation. Note that the magnitude of the wavelet coefficients tends to decay across scales. A similar behavior is observed for the bias terms of the visible layer (Fig. 6.7(a)).

The weight matrix of the RBM model reveals coefficient dependencies. Let  $h_*$  denote the hidden unit with the largest bias term of the RBM model using the learned overcomplete dictionary. Fig. 6.8(a) shows the weights associated with  $h_*$ . This hidden unit is connected to a group of visible units via large positive weights, which tend to



**Figure 6.7:** (a) Visible layer bias terms of the RBM model using wavelets. (b) Wavelet representation of an ECG segment of 128 samples.

be active simultaneously with  $h_*$ , in order to lower the system energy. Figs. 6.8(b) and (c) show the dictionary atoms sharing the same indexes as the 2 visible units with the most positive connections to  $h_*$ . Contrarily, Fig. 6.8(d) corresponds to the dictionary atom that shares the same index as the visible unit connected to  $h_*$  via the most negative weight. Note that both the patterns in Figs. 6.8(b) and (d) resemble the QRS complex, and therefore, are unlikely to happen simultaneously in an ECG segment of 128 samples (sampling frequency of 360 Hz).



**Figure 6.8:** (a) Weights associated with the hidden unit with the largest bias term. (b-c) Dictionary atoms sharing the same indexes as the visible units with the 2 most positive weights. (d) Dictionary atom sharing the same index as the visible unit with the most negative weight.

### 6.4.3 Robustness to Noise

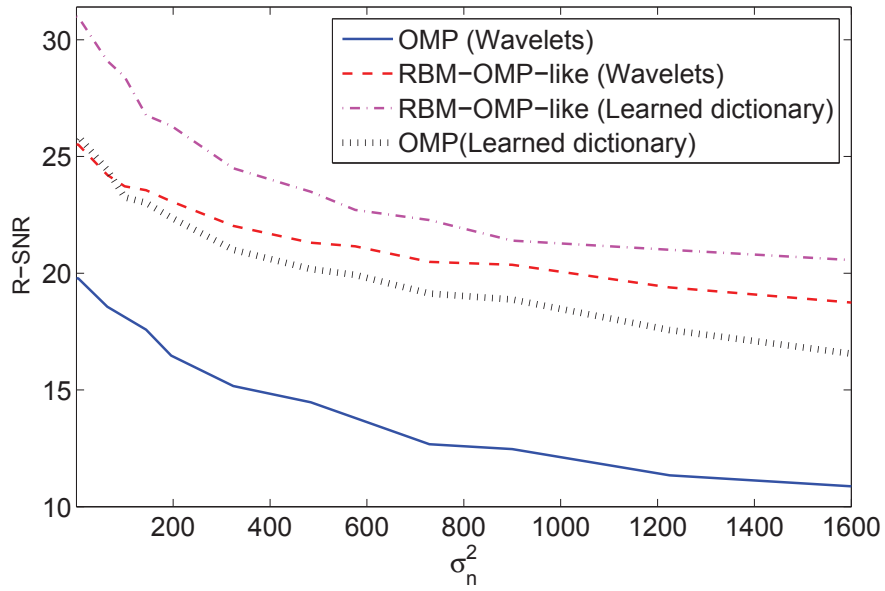
The third experiment evaluates the robustness of the RBM-OMP-like algorithm to measurement noise. Results are shown in Fig. 6.9. All the other parameters are set the same as in the first experiment. The Gaussian sensing matrix is kept fixed while the sampling noise variance is varied in the range  $[0.25, 1600]$ . When using wavelets as the sparsifying transform, the R-SNR as a function of the noise variance decays at a lower rate for the RBM-OMP-like algorithm than for OMP. In the case of overcomplete dictionaries, the decay rate of the RBM-OMP-like algorithm and OMP is comparable. The successful performance of the RBM-OMP-like algorithm in the presence of sampling noise is expected as the noise variance is explicitly taking into account in the algorithm.

### 6.4.4 Parameter Evaluation

We evaluate the sensitivity of the RBM-OMP-like algorithm using overcomplete dictionaries to major parameters, such as the dictionary size and the sparsity threshold. In each experiment, only one parameter is changed and the others are kept fixed. The default values are set the same as in the first experiment. Fig. 6.10 illustrates the performance of the RBM-OMP-like algorithm for dictionaries of size  $2N$ ,  $3N$  and  $4N$ . The increase in the number of atoms improves the capability of capturing details, and thus leads to better representations. Even though the number of atoms is bounded by the amount of training segments, it is seen that the R-SNR change from  $3N$  to  $4N$  is rather small.

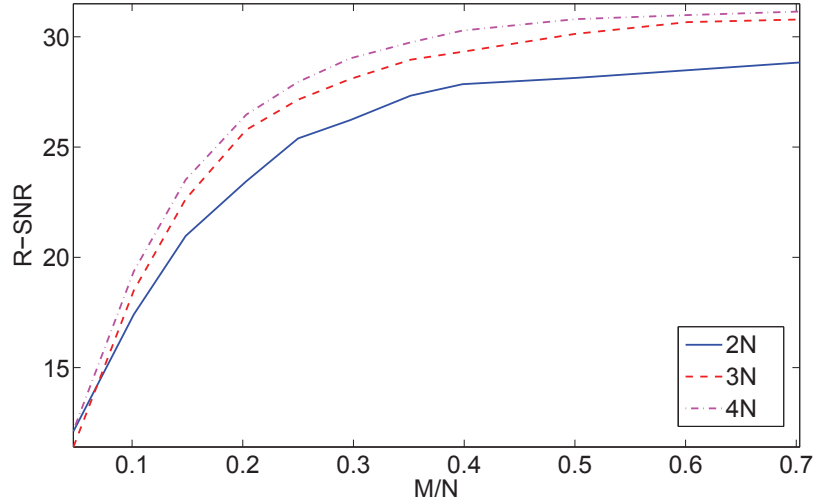
Fig. 6.11 illustrates how the sparsity threshold affects the performance of the RBM-OMP-like algorithm when overcomplete dictionaries are employed as the sparsifying transform. For  $M > 0.3N$ , the R-SNR decreases rapidly as the sparsity threshold decreases from  $0.2N$  to  $0.05N$ . The reason is that useful information is lost when the sparsity level decreases.

These results demonstrate that the RBM-OMP-like algorithm is not extremely sensitive to variations of the selected parameters. Even though we are aware that



**Figure 6.9:** Performance of the RBM-OMP-like and OMP algorithms in the presence of measurement noise

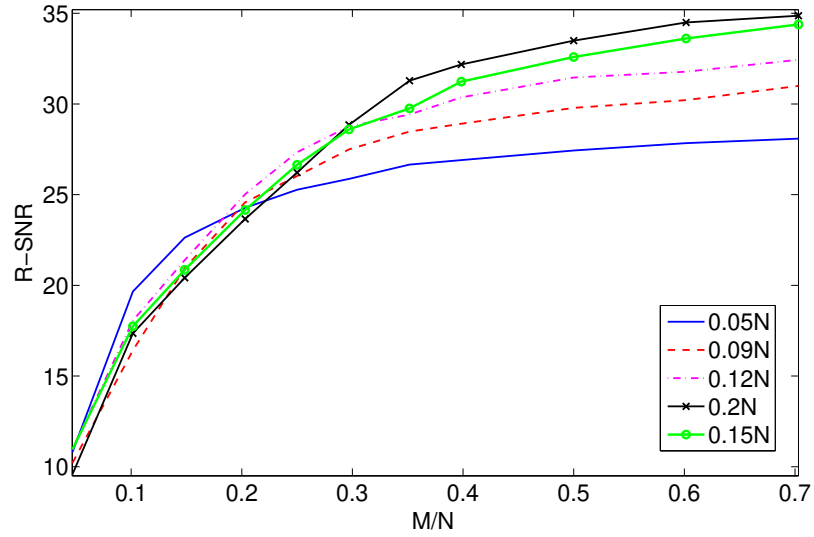
optimizing the parameters may lead to reconstruction improvement, how to perform the optimization is still an open problem that is beyond the scope of this thesis.



**Figure 6.10:** Reconstruction performance of the RBM-OMP-like algorithm using an overcomplete dictionary as a function of the dictionary size.

## 6.5 Conclusions

The potential of CS to lower energy consumption in WBAN-enabled ECG monitors lies in its ability to reduce the number of samples at the sensor node. With the goal of further reducing the number of samples, we propose to exploit the representational power of restricted Boltzmann machines and deep learning architectures to model the sparsity pattern of ECG signals. This scheme fully exploits signal structure in a statistical fashion, and subsequently, leads to stable reconstruction with low number of measurements. Even though other CS algorithms that exploit signal structure have been applied to ECG reconstruction, they all assume a specific type of structure, *e.g.*, tree-structured sparsity model and block-sparsity model, which is also tied to a specific sparsifying transform. Instead, the RBM/DBN-OMP like algorithms use a general model that accounts for a large set of structures and sparsifying transforms. It was shown that significant performance gains in the low-measurement regime can be obtained by using the proposed algorithms relative to state-of-the-art algorithms for CS ECG. Simulations using real ECG data from the MIT-BIH Arrhythmia database revealed that the performance of the algorithm is better when using overcomplete learned



**Figure 6.11:** Reconstruction performance of the RBM-OMP-like algorithm using an overcomplete dictionary as a function of the sparsity threshold.

dictionaries than when using wavelets.

The proposed algorithms operate over signals belonging to a certain signal class. In this thesis, the signal class of ECG signals was selected, but the scheme can also be applied to other signal classes; *e.g.* radar, natural images, speech signals, *etc.*

## Chapter 7

### SUMMARY AND FUTURE WORK

#### 7.1 Summary

The potential of CS to lower energy consumption in WBAN-enabled ECG lies in its ability to reduce the number of necessary measurements to attain faithful reconstruction. In this thesis, we proposed and studied a group of sparsity and compressibility models to further reduce the number of necessary measurements to reconstruct ECG signals using the CS framework. In addition, we also proposed an algorithm to introduce knowledge of the ECG real noise sources in the reconstruction.

In Chapter 2, the tree-structured sparsity model was employed. The largest (in magnitude) ECG wavelet coefficients form a connected tree structure caused by the piecewise smooth nature of ECG signals. Such structure was introduced in Chapter 2 to constrain the selection of the ECG signal support. Two iterative algorithms, based on CoSaMP and IHT, were developed to fully exploit the structured sparsity model. The algorithms also account for the quasi-periodicity property of ECG signals by exploiting the common support between consecutive ECG segments.

In Chapter 3, a novel approach to reconstruct ECG signals was presented. The main idea was to combine multi-scale wavelet representations and overcomplete learned dictionaries to exploit the magnitude correlation within each wavelet subband. In Chapter 3, the  $\ell_1$  minimization algorithm with quadratic constraints was employed for the reconstruction.

The ECG reconstruction algorithm presented in Chapter 4 stemmed from the need of improving the recovery of the detail wavelet coefficients, which provide critical information for medical diagnosis. Specifically, a weighted  $\ell_1$  minimization algorithm



that accounts for the magnitude decay of the ECG wavelet coefficients across scales was presented in Chapter 4.

Traditional CS reconstruction algorithms assume a Gaussian noise model. However, real EMG noise is better modeled by the symmetric  $\alpha$ -stable distribution than by the Gaussian distribution. In Chapter 5, an IHT-based algorithm, that accounts for the impulsiveness of the EMG noise by using the  $\ell_p$  norm, was developed.

In Chapter 6, a sparsity model that does not focus on a specific type of structure, but allows for a wide set of structures, was presented. The amplitude of the nonzero sparse representation coefficients was modeled with a joint Gaussian distribution and the support was modeled with a RBM and a DBN to exploit higher-order coefficient dependencies. The ECG reconstruction problem was formulated as a particular maximum *a posteriori* problem on the support of the sparse vector. The problem was solved via a greedy algorithm that shares some structural similarities with OMP. It was shown via simulations that the huge representational power of RBMs and DBNs leads to a significant reduction in the number of necessary measurements to achieve stable reconstruction.

## 7.2 Future Work

Some of the algorithms developed in this thesis for ECG can be extended to other type of signals. For example, the wavelet coefficient exponential decay property, exploited by the weighted  $\ell_1$  minimization algorithm presented in Chapter 4, is inherent to piecewise smooth signals. Therefore, the developed algorithm could be successfully applied to other type of piecewise smooth signals, such as natural images. Even though the RBM/DBN-OMP-like algorithm, presented in Chapter 6, was proposed for ECG reconstruction, it can be applied to a far richer class of signals since it is based on a generic sparsity model that is neither tied to a specific structure, nor to a specific sparsifying transform.

Regarding the area of CS ECG, there are many avenues for future work. For

example, real ECG noise sources should be taken into consideration in the reconstruction algorithms. In this work, we only considered EMG noise. As future work, it remains to consider other noise sources, such as baseline wander and electrode motion artifacts. The CS ECG framework presented in this thesis assumes that the sampling matrix is fixed and in particular, it is non-adaptive to the ECG signal structure. The benefit of such approach is that it leads to simple implementations, which is favorable in terms of computational cost and hardware design [21, 56]. However, adaptive measurement schemes may lead to significant reconstruction performance gains. Therefore, future work on CS ECG also relates to exploring novel sensing matrices that exploit the structural models presented in this dissertation.

Three different redundancy sources can be distinguished in the sparse representation of ECG signals: correlation between the sparse coefficients of a single beat, beat-to-beat quasi-periodic behavior, and inter-lead correlation. The first two types of correlation have been exploited in this thesis. As future work, it remains to exploit inter-lead correlation as we have only focused on single-lead ECG. For multi-lead ECG, the waveform morphology and the signal level of each channel is different. However, there is correlation among various channels because they record the same heart activity simultaneously.

In this dissertation, most of the proposed CS reconstruction algorithms that leverage structure in the sparsity pattern fall into the category of iterative greedy algorithms. There is an unexplored line of research regarding structured sparse regularization problems. Penalized loss function minimizations, such as the  $\ell_1$ -penalized squared error minimization, have shown to be effective for extracting useful information from high-dimensional data [22, 80]. However, the  $\ell_1$  penalty only deals with the signal sparsity. It would be desirable to design intelligent penalties in order to add structure information to the fitting of a regression model so as to allow groupings and hierarchical relationships between the sparsity basis coefficients. This idea would be beneficial for ill-posed estimation problems where sparsity alone may not be sufficient to obtain stable estimates.

In Chapter 7, it was shown via simulations that overcomplete learned dictionaries lead to better performance of CS ECG reconstruction algorithms compared to orthonormal bases. It was also shown that exploiting structured sparsity models leads to significant performance gains in the reconstruction. A future avenue of research refers to designing dictionaries that generate structured sparse representations. In that case, the structure would be known *a priori* and could be easily incorporated into the reconstruction algorithm for optimal performance.

## BIBLIOGRAPHY

- [1] M. Aharon, M. Elad, and A. Bruckstein. K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Transactions on Signal Processing*, 54(11):4311–4322, Nov. 2006.
- [2] R. Baraniuk, M. Davenport, R. DeVore, and M. Wakin. A simple proof of the restricted isometry property for random matrices. *Constructive Approximation*, 28(3):253–263, 2008.
- [3] R.G. Baraniuk, V. Cevher, M.F. Duarte, and C. Hegde. Model-based compressive sensing. *IEEE Transactions on Information Theory*, 56(4):1982–2001, Apr. 2010.
- [4] R.G. Baraniuk, R.A. DeVore, G. Kyriazis, and X.M. Yu. Near best tree approximation. *Advances in Computational Mathematics*, 16(4):357–373, May 2002.
- [5] D. Baron, S. Sarvotham, and R.G. Baraniuk. Bayesian compressive sensing via belief propagation. *IEEE Transactions on Signal Processing*, 58(1):269–280, Jan 2010.
- [6] Z. Ben-Haim, Y.C. Eldar, and M. Elad. Coherence-based performance guarantees for estimating a sparse vector under random noise. *IEEE Transactions on Signal Processing*, 58(10):5030–5043, Oct 2010.
- [7] Y. Bengio. Learning deep architectures for ai. *Foundations and Trends in Machine Learning*, 2(1):1–127, January 2009.
- [8] T. Blumensath and M Davies. Iterative hard thresholding for compressed sensing. *Applied and Computational Harmonic Analysis*, 27(3), Nov. 2009.
- [9] A.M. Bruckstein, D.L. Donoho, and M. Elad. From sparse solutions of systems of equations to sparse modeling of signals and images. *SIAM Review*, 51(1):34–81, Feb. 2009.
- [10] E.J. Candès. The restricted isometry property and its implications for compressed sensing. *Comptes Rendus Mathématique*, 346(9):589 – 592, 2008.
- [11] E.J. Candès, J. Romberg, and T. Tao. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *IEEE Transactions on Information Theory*, 52(2):489–509, Feb 2006.

- [12] E.J. Candès and T. Tao. Decoding by linear programming. *IEEE Transactions on Information Theory*, 51(12):4203–4215, Dec 2005.
- [13] E.J. Candès and T. Tao. Near-optimal signal recovery from random projections: Universal encoding strategies? *IEEE Transactions on Information Theory*, 52(12):5406–5425, Dec 2006.
- [14] E.J. Candès and T. Tao. The dantzig selector: Statistical estimation when  $p$  is much larger than  $n$ . *The Annals of Statistics*, pages 2313–2351, 2007.
- [15] E.J. Candès and M.B. Wakin. An introduction to compressive sampling. *IEEE Signal Processing Magazine*, 25(2):21–30, March 2008.
- [16] H. Cao, V. Leung, C. Chow, and H. Chan. Enabling technologies for wireless body area networks: A survey and outlook. *IEEE Communications Magazine*, 47(12):84–93, Dec. 2009.
- [17] R. E. Carrillo, L. F. Polania, and K. E. Barner. Iterative algorithms for compressed sensing with partially known support. In *Proc., IEEE Int. Conf. on Acoustics, Speech, and Signal Processing*, pages 3654 –3657, Dallas, TX, March 2010.
- [18] R.E. Carrillo, K.E. Barner, and T.C. Aysal. Robust sampling and reconstruction methods for sparse signals in the presence of impulsive noise. *IEEE Journal of Selected Topics in Signal Processing*, 4(2):392–408, April 2010.
- [19] R.E. Carrillo, L.F. Polania, and K.E. Barner. Iterative hard thresholding for compressed sensing with partially known support. In *Proc., IEEE Int. Conf. on Acoustics, Speech, and Signal Processing*, pages 4028 –4031, may 2011.
- [20] R. Chartrand and Y. Wotao. Iteratively reweighted algorithms for compressive sensing. In *Proc., IEEE Int. Conf. on Acoustics, Speech, and Signal Processing*, pages 3869 –3872, april 2008.
- [21] F. Chen, A.P. Chandrakasan, and V.M. Stojanovic. Design and analysis of a hardware-efficient compressed sensing architecture for data compression in wireless sensors. *IEEE Journal of Solid-State Circuits*, 47(3):744–756, Mar. 2012.
- [22] S.S. Chen, D.L. Donoho, and M.A. Saunders. Atomic decomposition by basis pursuit. *SIAM Journal on Scientific Computing*, 20(1):33–61, Dec. 1998.
- [23] A. Cohen, W. Dahmen, and R. DeVore. Compressed sensing and best  $k$ -term approximation. *Journal of the American Mathematical Society*, 22(1):211–231, 2009.
- [24] M.S. Crouse, R.D. Nowak, and R.G. Baraniuk. Wavelet-based statistical signal processing using hidden markov models. *IEEE Transactions on Signal Processing*, 46(4):886–902, Apr. 1998.

- [25] R.A. DeVore. Deterministic constructions of compressed sensing matrices. *Journal of Complexity*, 23(4-6):918–925, August 2007.
- [26] A.M.R. Dixon, E.G. Allstot, D. Gangopadhyay, and D.J. Allstot. Compressed sensing system considerations for ECG and EMG wireless biosensors. *IEEE Transactions on Biomedical Circuits and Systems*, 6(2):156–166, Apr. 2012.
- [27] D. Donoho. Compressed sensing. *IEEE Transactions on Information Theory*, 52(4):1289–1306, Sept. 2006.
- [28] D.L. Donoho. De-noising by soft-thresholding. *IEEE Transactions on Information Theory*, 41(3):613–627, 1995.
- [29] D.L. Donoho and M. Elad. Optimally sparse representation in general (non-orthogonal) dictionaries via L1 minimization. In *Proceedings of the National Academy of Sciences*, volume 100, pages 2197–2202, 2003.
- [30] M.F. Duarte and Y.C. Eldar. Structured compressed sensing: From theory to applications. *IEEE Transactions on Signal Processing*, 59(9):4053–4085, 2011.
- [31] J.M. Duarte-Carvajalino and G. Sapiro. Learning to sense sparse signals: Simultaneous sensing matrix and sparsifying dictionary optimization. *IEEE Transactions on Image Processing*, 18(7):1395–1408, July 2009.
- [32] M. Elad. Optimized projections for compressed sensing. *IEEE Transactions on Signal Processing*, 55(12):5695–5702, Dec 2007.
- [33] M. Elad and M. Aharon. Image denoising via sparse and redundant representations over learned dictionaries. *IEEE Transactions on Image Processing*, 15(12):3736–3745, Dec. 2006.
- [34] K. Engan, S. O. Aase, and J. H. Husøy. Multi-frame compression: Theory and design. *Signal Processing*, 80(10):2121–2140, Oct. 2000.
- [35] K. Engan, K. Skretting, and J.H. Husøy. Family of iterative LS-based dictionary learning algorithms, ils-dla, for sparse signal representation. *Digital Signal Processing*, 17(1):32–49, Jan. 2007.
- [36] C.M. Fira and L. Goras. An ECG signals compression method and its validation using NNs. *IEEE Transactions on Biomedical Engineering*, 55(4):1319–1326, April 2008.
- [37] P. Garrigues and B.A. Olshausen. Learning horizontal connections in a sparse coding model of natural images. In *Advances in Neural Information Processing Systems*, pages 505–512, 2008.
- [38] A. Gersho and R. M. Gray. *Vector quantization and signal compression*, volume 159. Springer, 1992.

- [39] A.L. Goldberger et al. Physiobank, Physiotoolkit, and Physionet: Components of a new research resource for complex physiologic signals. *Circulation*, 101(23):e215–e220, 2000.
- [40] L. He and L. Carin. Exploiting structure in wavelet-based Bayesian compressive sensing. *IEEE Transactions on Signal Processing*, 57(9):3488–3497, 2009.
- [41] M.L. Hilton. Wavelet and wavelet packet compression of electrocardiograms. *IEEE Transactions on Biomedical Engineering*, 44(5):394–402, May 1997.
- [42] G. Hinton, S. Osindero, and Y. Teh. A fast learning algorithm for deep belief nets. *Neural computation*, 18(7):1527–1554, 2006.
- [43] G.E. Hinton. Training products of experts by minimizing contrastive divergence. *Neural Comput.*, 14(8):1771–1800, August 2002.
- [44] R.S.H. Istepanian, L.J. Hadjileontiadis, and S.M. Panas. ECG data compression using wavelets and higher order statistics methods. *IEEE Transactions on Information Technology in Biomedicine*, 5(2):108–115, 2001.
- [45] Y. Ji, Z. Yang, and W. Li. Bayesian sparse reconstruction method of compressed sensing in the presence of impulsive noise. *Circuits, Systems, and Signal Processing*, 32(6):2971–2998, 2013.
- [46] R. Kapoor, A. Banerjee, G.A. Tsihrintzis, and N. Nandhakumar. UWB radar detection of targets in foliage using alpha-stable clutter models. *IEEE Transactions on Aerospace and Electronic Systems*, 35(3):819–834, jul 1999.
- [47] J. Kovacevic and M. Vetterli. Transform coding: past, present, and future. *IEEE Signal Processing Magazine*, 18(5):6–6, Sep 2001.
- [48] G. Kutyniok. Theory and applications of compressed sensing. *GAMM-Mitteilungen*, 36(1):79–101, 2013.
- [49] Shu L. and A.C.S. Chung. Feature based nonrigid brain MR image registration with symmetric alpha stable filters. *IEEE Transactions on Medical Imaging*, 29(1):106–119, jan. 2010.
- [50] P. Lander and E.J. Berbari. Time-frequency plane Wiener filtering of the high-resolution ECG: background and time-frequency representations. *IEEE Transactions on Biomedical Engineering*, 44(4):247–255, April 1997.
- [51] N. Le Roux and Y. Bengio. Representational power of restricted Boltzmann machines and deep belief networks. *Neural Computation*, 20(6):1631–1649, 2008.
- [52] M.S. Lewicki and B.A. Olshausen. Probabilistic framework for the adaptation and comparison of image codes. *Journal of the Optical Society of America A*, 16(7):1587–1601, Jul 1999.

- [53] E. Liberty and S.W. Zucker. The Mailman algorithm: A note on matrix-vector multiplication. *Information Processing Letters*, 109(3):179 – 182, 2009.
- [54] Z. Lu, D. Y. Kim, and W. A. Pearlman. Wavelet compression of ECG signals by the set partitioning in hierarchical trees algorithm. *IEEE Transactions on Biomedical Engineering*, 47(7):849–856, Jul. 2000.
- [55] Stéphane Mallat. *A Wavelet Tour of Signal Processing, Third Edition: The Sparse Way*. Academic Press, 3rd edition, 2008.
- [56] H. Mamaghanian, N. Khaled, D. Atienza, and P. Vanderghenst. Compressed sensing for real-time energy-efficient ECG compression on wireless body sensor nodes. *IEEE Transactions on Biomedical Engineering*, 58(9):2456–2466, Sept. 2011.
- [57] H. Mamaghanian, N. Khaled, D. Atienza, and P. Vanderghenst. Structured sparsity models for compressively sensed electrocardiogram signals: A comparative study. In *2011 IEEE Biomedical Circuits and Systems Conference*, pages 125–128, Nov. 2011.
- [58] H. Mamaghanian, N. Khaled, D. Atienza, and P. Vanderghenst. Design and exploration of low-power analog to information conversion based on compressed sensing. *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, 2(3):493–501, Sept. 2012.
- [59] D. Needell, J. Tropp, and R. Vershynin. Greedy signal recovery review. In *Asilomar Conference on Signals, Systems and Computers*, pages 1048–1050, Oct 2008.
- [60] B. Ophir, M. Lustig, and M. Elad. Multi-scale dictionary learning using wavelets. *IEEE Journal of Selected Topics in Signal Processing*, 5(5):1014–1024, Sept. 2011.
- [61] T.P. Pander. A suppression of an impulsive noise in ecg signal processing. In *IEMBS*, volume 1, pages 596 –599, sept. 2004.
- [62] R. Paradiso, G. Loriga, and N. Taccini. A wearable health care system based on knitted integrated sensors. *IEEE Transactions on Information Technology in Biomedicine*, 9(3):337–344, Sept. 2005.
- [63] Y Park. Time complexity analysis of SPIHT (set partitioning in hierarchy trees) image coding algorithm. *Journal of the Institute of Signal Processing and Systems*, 4(1):36–40, 2003.
- [64] J.S. Paul, M.R. Reddy, and V.J. Kumar. A transform domain SVD filter for suppression of muscle noise artefacts in exercise ECG’s. *IEEE Transactions on Biomedical Engineering*, 47(5):654–663, May 2000.



- [65] T. Peleg, Y.C. Eldar, and M. Elad. Exploiting statistical dependencies in sparse representations for signal recovery. *IEEE Transactions on Signal Processing*, 60(5):2286–2303, 2012.
- [66] L.F. Polania and K.E. Barner. Multi-scale dictionary learning for compressive sensing ECG. In *2013 IEEE Digital Signal Processing and Signal Processing Education Meeting*, pages 36–41, Aug. 2013.
- [67] L.F. Polania and K.E. Barner. A weighted  $\ell_1$  minimization algorithm for compressed sensing ECG. In *IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 4413–4417, May 2014.
- [68] L.F. Polanía and K.E. Barner. Compressed sensing meets deep learning. *IEEE Transactions on Signal Processing*, 2015. Submitted for publication.
- [69] L.F. Polanía and K.E. Barner. Exploiting restricted Boltzmann machines in compressed sensing wireless ECG systems. *IEEE Transactions on Biomedical Engineering*, 2015. Submitted for publication.
- [70] L.F. Polania, R.E. Carrillo, M. Blanco-Velasco, and K.E. Barner. Compressed sensing based method for ECG compression. In *Proc., IEEE Int. Conf. on Acoustics, Speech, and Signal Processing*, pages 761 –764, May 2011.
- [71] L.F. Polania, R.E. Carrillo, M. Blanco-Velasco, and K.E. Barner. Matrix completion based ECG compression. In *Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 1757–1760. IEEE, 2011.
- [72] L.F. Polania, R.E. Carrillo, M. Blanco-Velasco, and K.E. Barner. Compressive sensing exploiting wavelet-domain dependencies for ECG compression. In *SPIE Defense, Security, and Sensing*, pages 83650E–83650E, Apr. 2012.
- [73] L.F. Polania, R.E. Carrillo, M. Blanco-Velasco, and K.E. Barner. Compressive sensing for ECG signals in the presence of electromyographic noise. In *38th Annual Northeast Bioengineering Conference*, pages 295–296, Mar. 2012.
- [74] L.F. Polania, R.E. Carrillo, M. Blanco-Velasco, and K.E. Barner. On exploiting interbeat correlation in compressive sensing-based ECG compression. In *SPIE Defense, Security, and Sensing*, pages 83650D–83650D, Apr. 2012.
- [75] L.F. Polania, R.E. Carrillo, M. Blanco-Velasco, and K.E. Barner. Exploiting prior knowledge in compressed sensing wireless ECG systems. *IEEE Journal of Biomedical and Health Informatics*, 19(2):508–519, March 2015.
- [76] R. Prasad and C.R. Murthy. Cramér-Rao-type bounds for sparse Bayesian learning. *IEEE Transactions on Signal Processing*, 61(3):622–632, Feb 2013.

- [77] H. Rauhut. Compressive sensing and structured random matrices. *Theoretical foundations and numerical methods for sparse recovery*, 9:1–92, 2010.
- [78] H. Rauhut, K. Schnass, and P. Vandergheynst. Compressed sensing and redundant dictionaries. *IEEE Transactions on Information Theory*, 54(5):2210–2219, 2008.
- [79] J.K. Romberg, Hyeokho Choi, and R.G. Baraniuk. Bayesian tree-structured image modeling using wavelet-domain hidden Markov models. *IEEE Transactions on Image Processing*, 10(7):1056–1068, 2001.
- [80] C. Shaobing and D. Donoho. Basis pursuit. In *Conference Record of the Twenty-Eighth Asilomar Conference on Signals, Systems and Computers, 1994.*, volume 1, pages 41–44 vol.1, Oct-2 Nov.
- [81] J. Shihao, Ya X., and L. Carin. Bayesian compressive sensing. *IEEE Transactions on Signal Processing*, 56(6):2346–2356, June 2008.
- [82] I. Sutskever and G.E. Hinton. Deep, narrow sigmoid belief networks are universal approximators. *Neural Computation*, 20(11):2629–2636, 2008.
- [83] J. Talmon, J. Kors, and J. Van Bommel. Adaptive gaussian filtering in routine ECG/VCG analysis. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 34(3):527–534, Jun 1986.
- [84] R. Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society, Series B*, 58:267–288, 1994.
- [85] J.A. Tropp and A.C. Gilbert. Signal recovery from random measurements via orthogonal matching pursuit. *IEEE Transactions on Information Theory*, 53(12):4655–4666, Dec. 2007.
- [86] G. Tzagkarakis. *Bayesian Compressed Sensing using Alpha-Stable Distributions*. PhD thesis, University of Crete, Greece, 2009.
- [87] V.N. Vapnik. *The nature of statistical learning theory*. Springer-Verlag New York, Inc., New York, NY, USA, 1995.
- [88] N. Vaswani and Wei L. Modified-CS: Modifying compressive sensing for problems with partially known support. *IEEE Transactions on Signal Processing*, 58(9):4595–4607, Sept. 2010.
- [89] N. Vaswani and Wei L. Modified-CS: Modifying compressive sensing for problems with partially known support. *IEEE International Symposium on Information Theory*, 58(9):4595–4607, 2010.
- [90] C. Wang, J. Liu, and J. Sun. Compression algorithm for electrocardiograms based on sparse decomposition. *Frontiers of Electrical and Electronic Engineering in China*, 4:10–14, 2009.

- [91] L. Wang and W. Yin. Sparse signal reconstruction via iterative support detection. *SIAM Journal on Imaging Sciences*, 3(3):462–491, Jul. 2010.
- [92] Z. Yang, L. Xie, and C. Zhang. Variational bayesian algorithm for quantized compressed sensing. *IEEE Transactions on Signal Processing*, 61(11):2815–2824, June 2013.
- [93] Z. Zhang, T. Jung, S. Makeig, and B.D. Rao. Compressed sensing for energy-efficient wireless telemonitoring of noninvasive fetal ECG via block sparse bayesian learning. *IEEE Transactions on Biomedical Engineering*, 60(2):300–309, Feb. 2013.
- [94] Z. Zhang and B.D. Rao. Sparse signal recovery with temporally correlated source vectors using sparse Bayesian learning. *IEEE Journal of Selected Topics in Signal Processing*, 5(5):912–926, Sept 2011.

**Appendix**  
**COPYRIGHT FORMS**

The IEEE copyright form for the paper entitled *Exploiting Prior Knowledge in Compressed Sensing Wireless ECG Systems* and published in the IEEE Journal of Biomedical and Health Informatics is presented in the next pages.

## IEEE COPYRIGHT AND CONSENT FORM

To ensure uniformity of treatment among all contributors, other forms may not be substituted for this form, nor may any wording of the form be changed. This form is intended for original material submitted to the IEEE and must accompany any such material in order to be published by the IEEE. Please read the form carefully and keep a copy for your files.

TITLE OF PAPER/ARTICLE/REPORT, INCLUDING ALL CONTENT IN ANY FORM, FORMAT, OR MEDIA (hereinafter, "The Work"): **Exploiting Prior Knowledge in Compressed Sensing Wireless ECG Systems**

COMPLETE LIST OF AUTHORS: **Polania, Luisa; Barner, Kenneth; Carrillo, Rafael; Blanco-Velasco, Manuel**

IEEE PUBLICATION TITLE (Journal, Magazine, Conference, Book): **IEEE Journal of Biomedical and Health Informatics**

### COPYRIGHT TRANSFER

1. The undersigned hereby assigns to The Institute of Electrical and Electronics Engineers, Incorporated (the "IEEE") all rights under copyright that may exist in and to: (a) the above Work, including any revised or expanded derivative works submitted to the IEEE by the undersigned based on the Work; and (b) any associated written or multimedia components or other enhancements accompanying the Work.

### CONSENT AND RELEASE

2. In the event the undersigned makes a presentation based upon the Work at a conference hosted or sponsored in whole or in part by the IEEE, the undersigned, in consideration for his/her participation in the conference, hereby grants the IEEE the unlimited, worldwide, irrevocable permission to use, distribute, publish, license, exhibit, record, digitize, broadcast, reproduce and archive, in any format or medium, whether now known or hereafter developed: (a) his/her presentation and comments at the conference; (b) any written materials or multimedia files used in connection with his/her presentation; and (c) any recorded interviews of him/her (collectively, the "Presentation"). The permission granted includes the transcription and reproduction of the Presentation for inclusion in products sold or distributed by IEEE and live or recorded broadcast of the Presentation during or after the conference.

3. In connection with the permission granted in Section 2, the undersigned hereby grants IEEE the unlimited, worldwide, irrevocable right to use his/her name, picture, likeness, voice and biographical information as part of the advertisement, distribution and sale of products incorporating the Work or Presentation, and releases IEEE from any claim based on right of privacy or publicity.

4. The undersigned hereby warrants that the Work and Presentation (collectively, the "Materials") are original and that he/she is the author of the Materials. To the extent the Materials incorporate text passages, figures, data or other material from the works of others, the undersigned has obtained any necessary permissions. Where necessary, the undersigned has obtained all third party permissions and consents to grant the license above and has provided copies of such permissions and consents to IEEE.

Please check this box if you do not wish to have video/audio recordings made of your conference presentation.

See below for Retained Rights/Terms and Conditions, and Author Responsibilities.

### AUTHOR RESPONSIBILITIES

The IEEE distributes its technical publications throughout the world and wants to ensure that the material submitted to its publications is properly available to the readership of those publications. Authors must ensure that their Work meets the requirements as stated in section 8.2.1 of the IEEE PSPB Operations Manual, including provisions covering originality, authorship, author responsibilities and author misconduct. More information on

IEEE's publishing policies may be found at [http://www.ieee.org/publications\\_standards/publications/rights/authorrightsresponsibilities.html](http://www.ieee.org/publications_standards/publications/rights/authorrightsresponsibilities.html). Authors are advised especially of IEEE PSPB Operations Manual section 8.2.1.B12: "It is the responsibility of the authors, not the IEEE, to determine whether disclosure of their material requires the prior consent of other parties and, if so, to obtain it." Authors are also advised of IEEE PSPB Operations Manual section 8.1.1B: "Statements and opinions given in work published by the IEEE are the expression of the authors."

## **RETAINED RIGHTS/TERMS AND CONDITIONS**

### **General**

1. Authors/employers retain all proprietary rights in any process, procedure, or article of manufacture described in the Work.
2. Authors/employers may reproduce or authorize others to reproduce the Work, material extracted verbatim from the Work, or derivative works for the author's personal use or for company use, provided that the source and the IEEE copyright notice are indicated, the copies are not used in any way that implies IEEE endorsement of a product or service of any employer, and the copies themselves are not offered for sale.
3. In the case of a Work performed under a U.S. Government contract or grant, the IEEE recognizes that the U.S. Government has royalty-free permission to reproduce all or portions of the Work, and to authorize others to do so, for official U.S. Government purposes only, if the contract/grant so requires.
4. Although authors are permitted to re-use all or portions of the Work in other works, this does not include granting third-party requests for reprinting, republishing, or other types of re-use. The IEEE Intellectual Property Rights office must handle all such third-party requests.
5. Authors whose work was performed under a grant from a government funding agency are free to fulfill any deposit mandates from that funding agency.

### **Author Online Use**

6. Personal Servers. Authors and/or their employers shall have the right to post the accepted version of IEEE-copyrighted articles on their own personal servers or the servers of their institutions or employers without permission from IEEE, provided that the posted version includes a prominently displayed IEEE copyright notice and, when published, a full citation to the original IEEE publication, including a link to the article abstract in IEEE Xplore. Authors shall not post the final, published versions of their papers.
7. Classroom or Internal Training Use. An author is expressly permitted to post any portion of the accepted version of his/her own IEEE-copyrighted articles on the authors personal web site or the servers of the authors institution or company in connection with the authors teaching, training, or work responsibilities, provided that the appropriate copyright, credit, and reuse notices appear prominently with the posted material. Examples of permitted uses are lecture materials, course packs, e-reserves, conference presentations, or in-house training courses.
8. Electronic Preprints. Before submitting an article to an IEEE publication, authors frequently post their manuscripts to their own web site, their employers site, or to another server that invites constructive comment from colleagues. Upon submission of an article to IEEE, an author is required to transfer copyright in the article to IEEE, and the author must update any previously posted version of the article with a prominently displayed IEEE copyright notice. Upon publication of an article by the IEEE, the author must replace any previously posted electronic versions of the article with either (1) the full citation to the IEEE work with a Digital Object Identifier (DOI) or link to the article abstract in IEEE Xplore, or (2) the accepted version only (not the IEEE-published version), including the IEEE copyright notice and full citation, with a link to the final, published article in IEEE Xplore.

## **INFORMATION FOR AUTHORS**

### **IEEE Copyright Ownership**

It is the formal policy of the IEEE to own the copyrights to all copyrightable material in its technical publications and to the individual contributions contained therein, in order to protect the interests of the IEEE, its authors and their employers, and, at the same time, to facilitate the appropriate re-

use of this material by others. The IEEE distributes its technical publications throughout the world and does so by various means such as hard copy, microfiche, microfilm, and electronic media. It also abstracts and may translate its publications, and articles contained therein, for inclusion in various compendiums, collective works, databases and similar publications.

### **Author/Employer Rights**

If you are employed and prepared the Work on a subject within the scope of your employment, the copyright in the Work belongs to your employer as a work-for-hire. In that case, the IEEE assumes that when you sign this Form, you are authorized to do so by your employer and that your employer has consented to the transfer of copyright, to the representation and warranty of publication rights, and to all other terms and conditions of this Form. If such authorization and consent has not been given to you, an authorized representative of your employer should sign this Form as the Author.

### **GENERAL TERMS**

1. The undersigned represents that he/she has the power and authority to make and execute this form.
2. The undersigned agrees to identify and hold harmless the IEEE from any damage or expense that may arise in the event of a breach of any of the warranties set forth above.
3. In the event the above work is not accepted and published by the IEEE or is withdrawn by the author(s) before acceptance by the IEEE, the foregoing grant of rights shall become null and void and all materials embodying the Work submitted to the IEEE will be destroyed.
4. For jointly authored Works, all joint authors should sign, or one of the authors should sign as authorized agent for the others.

**Luisa F. Polania**

**Author/Authorized Agent For Joint Authors**

**14-05-2014**

**Date(dd-mm-yy)**

THIS FORM MUST ACCOMPANY THE SUBMISSION OF THE AUTHOR'S MANUSCRIPT.

Questions about the submission of the form or manuscript must be sent to the publication's editor. Please direct all questions about IEEE copyright policy to:

IEEE Intellectual Property Rights Office, [copyrights@ieee.org](mailto:copyrights@ieee.org), +1-732-562-3966 (telephone)