Fusion of Sparse Reconstruction Algorithms in Compressed Sensing

A thesis submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in the
Faculty of Engineering

by

Sooraj K. Ambat

Department of Electrical Communication Engineering
Indian Institute of Science
Bangalore - 560 012

April 2015
THESIS ADVISOR

Prof. K.V.S. Hari

THESIS EXAMINERS

Prof. S. D. Joshi
Department of Electrical Engineering
Indian Institute of Technology Delhi
New Delhi-110 016, India.

Prof. Anamitra Makur
School of Electrical & Electronic Engineering
Nanyang Technological University
Singapore 639798
Acknowledgements

This thesis is a major milestone on my journey towards my Ph.D. which has been kept on track and been seen through to completion with the support and encouragement of numerous people including my teachers, well wishers, friends, and colleagues. At the end of my thesis, I would like to thank all those people who made this thesis possible and an unforgettable experience for me. It gives me great pleasure to express my gratitude to all those who have supported me and had their contributions in making this thesis possible.

With great esteem, gratitude, and love I express my heartfelt obligations to my research adviser Prof. K.V.S. Hari. I have been very fortunate to have an adviser who gave me the freedom to choose and explore the area my own. He was perpetually there to help me decide my next step, as a true ‘bouncing board’, whenever I was at a crossroads. I warmly thank Prof. Hari, for his valuable advice, constructive criticism, and the extensive discussions around my work.

I have been fortunate to have enjoyed the support and collaboration of Dr. Saikat Chatterjee, who showed me the art of replying to reviewers. I thank him for his support, intuition, and valuable suggestions toward improving my work.

I am indebted to the members of Statistical Signal Processing (SSP) Lab during 2010 – 2014 who made my stay at IISc an unforgettable experience. I specifically thank Amit Kumar Dutta, Dinesh Dileep Gaurav, Renu Jose, Rakshith Mysore Rajasekhar, Mukund Sriram, Prateek G. V., K. V. S. Sastry, Imtiaz Pasha, Rakshith Jagannath, Dr. Satya Sudhakar Yedlapalli, Shree Ranga Raju, Sachin Ambedkar, Deepa K. G., and Karthik Upadhya for their support in
various capacities during my life at SSP Lab. I would also like to extend my gratitude to the all the faculty and staff in the Department of ECE, IISc for always being helpful and friendly.

I take this opportunity to sincerely acknowledge Naval Physical and Oceanographic Laboratory (NPOL) and Defence R&D Organization (DRDO), India for sanctioning me a study leave for three years which buttressed me to perform my Ph.D. work comfortably. I gratefully acknowledge S. Anantha Narayanan, Director (Rtd.), NPOL, and V. P. Felix, Group Head (Rtd.), Signal Processing Algorithms Group, NPOL, for sparing me for three years to join IISc as a regular Ph.D. scholar. I also express my deep gratitude to Dr. A. Unnikrishnan, Associate Director (Rtd.), NPOL, for his constant encouragement and support for my decision to go for Ph.D. My warm thanks are also due to Dr. N. R. Manoj, NPOL, for giving me the inspiration, at the right time, to cross the red tape hurdles to get my study leave. I also like to thank Chitharanjan, Admin. Officer (Rtd.), NPOL, Anoobkumar A. A., Senior Admin. Asst., NPOL, and Ajithkumar P., Senior Admin. Asst., NPOL for their support which enabled me to understand the service rules in a better way. This thesis would not have been possible without the encouragement and support of the friends and colleagues at NPOL that surround me. I heartily thank them all, including - although I am sure the list is incomplete - Nissar K. E., Pradeepa R., Sarath Gopi, MuraliKrishna, Sreedavy Prakash, Aparna V., Hareesh, Ajan, Mercy Paul, Lasitha Ranjith, Maria Sparanica, Sreekanth Raja, Remadevi, Jojish Joseph, Sijomon, Nirmal Mohan, Naishab, Bibin, and Gopan. I also like to thank K. C. Unnikrishnan and Baiju M. Nair for the endless discussions in late hours, most of them occurred after playing badminton, which boosted my desire to do a full-time Ph. D.

I thank Prof. Chandra R. Murthy, IISc for encouraging me to take study leave and join IISc as a regular student. I would like
to acknowledge Dr. Rafael E. Carrillo, Ms. Luisa F. Polania, and Prof. Kenneth E. Barner, the authors of [139],[140], for sharing their Matlab code which was useful for validating our proposed methods with real signals. I also thank Prof. Bhaskar D. Rao for the fruitful discussions during his visit to Statistical Signal Processing Lab and pointing out the similarities between our work [148] and the committee machine approach used in neural networks. I sincerely thank Igor Carron for maintaining such a wonderful blog [78] which helped me a great extent to update the happenings in Compressed Sensing (CS) and related areas. Often, my Ph.D. days started by reading his excellent blog entries on the cutting edge research in CS.

I thank my comprehensive examination committee members Prof. A. G. Ramakrishnan, Prof. Soumyendu Raha, and Prof. Chandrachud R. Murthy, for spending their time and giving valuable advice during the initial phase of my research.

I am also thankful to my cousin Vijay Sankar for his support during my stay at IISc. I gratefully acknowledge Jis Joseph, IISc for the wake up calls. I take this opportunity to acknowledge all the wake up calls in my previous endeavours!

Finally, I would like to thank all my family members: my mother K. Ammini Amma and father A. P. Krishnan Nair, for their unconditional love and endless encouragement & support. Many thanks to my in-laws (Santha and C. S. Gopalakrishna Pillai), brother (Subhash K. Ambat), brother-in-law (Manoj G. Pillai), sisters-in-law (Amrutha Subhash and Manjusha Manoj), aunts, uncles and cousins (oh! we are a big family to list out all the names) for their consistent love and support. I would also wish to thank my late grandparents Gouri Amma, Parameswaran Nair, Kamalakshi Amma, and Parameswaran Nair, for the endless love and affection
they have always offered me. I am extremely grateful to my wife, Manju, and our boys Gourisankar and Harisankar for their invaluable love, support, and encouragement. They all have always believed in me, more than I myself do, and have been fully supportive of my decision to go for a full-time Ph.D. I dedicate this thesis to my family, specifically to my late grandparents, for their endless love, support, and self-sacrifices.
Abstract

Compressed Sensing (CS) is a new paradigm in signal processing which exploits the sparse or compressible nature of the signal to significantly reduce the number of measurements, without compromising on the signal reconstruction quality. Recently, many algorithms have been reported in the literature for efficient sparse signal reconstruction. Nevertheless, it is well known that the performance of any sparse reconstruction algorithm depends on many parameters like number of measurements, dimension of the sparse signal, the level of sparsity, the measurement noise power, and the underlying statistical distribution of the non-zero elements of the signal. It has been observed that a satisfactory performance of the sparse reconstruction algorithm mandates certain requirement on these parameters, which is different for different algorithms. Many applications are unlikely to fulfil this requirement. For example, imaging speed is crucial in many Magnetic Resonance Imaging (MRI) applications. This restricts the number of measurements, which in turn affects the medical diagnosis using MRI. Hence, any strategy to improve the signal reconstruction in such adverse scenario is of substantial interest in CS.

Interestingly, it can be observed that the performance degradation of the sparse recovery algorithms in the aforementioned cases does not always imply a complete failure. That is, even in such adverse situations, a sparse reconstruction algorithm may provide partially correct information about the signal. In this thesis, we
study this scenario and propose a novel fusion framework and an iterative framework which exploit the partial information available in the sparse signal estimate(s) to improve sparse signal reconstruction.

The proposed fusion framework employs multiple sparse reconstruction algorithms, independently, for signal reconstruction. We first propose a fusion algorithm viz. Fusion of Algorithms for Compressed Sensing (FACS) which fuses the estimates of multiple participating algorithms in order to improve the sparse signal reconstruction. To alleviate the inherent drawbacks of FACS and further improve the sparse signal reconstruction, we propose another fusion algorithm called Committee Machine Approach for Compressed Sensing (CoMACS) and variants of CoMACS. For low latency applications, we propose a latency friendly fusion algorithm called progressive Fusion of Algorithms for Compressed Sensing (pFACS). We also extend the fusion framework to the Multiple Measurement Vector (MMV) problem and propose the extension of FACS called Multiple Measurement Vector Fusion of Algorithms for Compressed Sensing (MMV-FACS). We theoretically analyse the proposed fusion algorithms and derive guarantees for performance improvement. We also show that the proposed fusion algorithms are robust against both signal and measurement perturbations. Further, we demonstrate the efficacy of the proposed algorithms via numerical experiments: (i) using sparse signals with different statistical distributions in noise-free and noisy scenarios, and (ii) using real-world ECG signals. The extensive numerical experiments show that, for a judicious choice of the participating algorithms, the proposed fusion algorithms result in a sparse signal estimate which is often better than the sparse signal estimate of the best participating algorithm.

The proposed fusion framework requires to employ multiple
sparse reconstruction algorithms for sparse signal reconstruction. We also propose an iterative framework and algorithm called Iterative Framework for Sparse Reconstruction Algorithms (IFSRA) to improve the performance of a given arbitrary sparse reconstruction algorithm. We theoretically analyse IFSRA and derive convergence guarantees under signal and measurement perturbations. Numerical experiments on synthetic and real-world data confirm the efficacy of IFSRA. The proposed fusion algorithms and IFSRA are general in nature and does not require any modification in the participating algorithm(s).
Glossary

Bold upper case and bold lower case Roman letters denote matrices and vectors, respectively. Calligraphic letters and upper case Greek alphabets are used to denote sets.

\[ \| \cdot \|_p \] The $p^{th}$ norm.
\[ \| X \|_{(p,q)} \] The $(p, q)$ mixed norm of the matrix $X$.
\[ \| A \|_F \] The Frobenius norm of matrix $A$.
$A$ The measurement matrix.
$A^H$ Hermitian of matrix $A$.
$A^{-1}$ Inverse of matrix $A$.
$alg^{(i)}$ The $i^{th}$ participating algorithm.
$A^\dagger$ Moore-Penrose pseudo-inverse of matrix $A$.
$A_T$ The column sub-matrix of $A$ with column indices listed in $T$.
$A^T$ Transpose of matrix $A$.
$b$ The measurement vector.
$|c|$ The magnitude of $c$.
$K$ The sparsity level (The number of non-zero elements).
$M$ The number of measurements.
$N$ The dimension of the sparse vector.
$\mathcal{N}(A)$ The Nullspace of the matrix $A$.
$R$ The cardinality of the joint support-set $\Gamma$.
$\mathcal{T}_1 \setminus \mathcal{T}_2$ The set difference between the sets $\mathcal{T}_1$ and $\mathcal{T}_2$ defined as $\mathcal{T}_1 \cap \mathcal{T}_2^c$.
$\text{supp}(x)$ The set of indices of non-zero elements of $x$.
$\text{supp}(X)$ The set of indices of non-zero rows of $X$. 
<table>
<thead>
<tr>
<th>Symbol</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>$</td>
<td>\mathcal{T}</td>
</tr>
<tr>
<td>$\mathcal{T}^c$</td>
<td>The complement of the set $\mathcal{T}$ with respect to the set {1, 2, \ldots, N}.</td>
</tr>
<tr>
<td>$\hat{T}_i$</td>
<td>The support-set estimated by the $i^{th}$ participating algorithm.</td>
</tr>
<tr>
<td>$w$</td>
<td>The additive measurement noise.</td>
</tr>
<tr>
<td>$x$</td>
<td>The unknown signal.</td>
</tr>
<tr>
<td>$\hat{X}_i$</td>
<td>The matrix reconstructed by the $i^{th}$ participating algorithm.</td>
</tr>
<tr>
<td>$\hat{x}_i$</td>
<td>The signal estimated by the $i^{th}$ participating algorithm.</td>
</tr>
<tr>
<td>$x^K$</td>
<td>The best $K$-term approximation of $x$, obtained from $x$ by keeping its entries with $K$ largest magnitudes and by setting all other magnitudes to zero.</td>
</tr>
<tr>
<td>$(x^K)_\mathcal{T}$</td>
<td>The sub-vector formed from $x^K$ whose indices are listed in the set $\mathcal{T}$.</td>
</tr>
<tr>
<td>$x^{(l)}$</td>
<td>The $l^{th}$ column vector of $X$.</td>
</tr>
<tr>
<td>$x^\Lambda$</td>
<td>The vector obtained from $x$ by keeping only the elements in indices listed in the set $\Lambda$, and setting rest of the elements zeros.</td>
</tr>
<tr>
<td>$x_\mathcal{T}$</td>
<td>The sub-vector formed from $x$ whose indices are listed in $\mathcal{T}$.</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>The fraction of measurements ($\frac{M}{N}$).</td>
</tr>
<tr>
<td>$\mathbb{E}{\cdot}$</td>
<td>The mathematical expectation operator.</td>
</tr>
<tr>
<td>$\Gamma$</td>
<td>The union of the support-sets estimated by the participating algorithms.</td>
</tr>
<tr>
<td>$\Sigma_K$</td>
<td>The set of all $K$-sparse signals.</td>
</tr>
</tbody>
</table>
Acronyms

ASRER  Average Signal-to-Reconstruction-Error Ratio.
BP     Basis Pursuit.
BPDN   Basis Pursuit De-Noising.
CoMACS Committee Machine Approach for Compressed Sensing.
CoSaMP Compressive Sampling Matching Pursuit.
CP     Chaining Pursuits.
CRM    Convex Relaxation Methods.
CS     Compressed Sensing.
DS     Dantzig Selector.
FACS   Fusion of Algorithms for Compressed Sensing.
FOCUSS FOcal Underdetermined System Solver.
FSA    Fourier Sampling Algorithm.
GSS    Gaussian Sparse Signals.
HHS    Heavy Hitters on Steroids.
ICoMACS Iterative CoMACS.
IHT    Iterative Hard Thresholding.
i.i.d. independently and identically distributed.
IRL1   Iterative Re-weighted L1.
IRLS   Iterative Re-weighted Least-Squares.
ISD    Iterative Support Detection.
LARS   Least Angle Regression.
LASSO  Least Absolute Shrinkage and Selection Operator.
LS     Least-Squares.
<table>
<thead>
<tr>
<th>Acronyms</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>MMV</td>
<td>Multiple Measurement Vector.</td>
</tr>
<tr>
<td>MMV-FACS</td>
<td>Multiple Measurement Vector Fusion of Algorithms for Compressed Sensing.</td>
</tr>
<tr>
<td>MOMP</td>
<td>Multiple Measurement Vector Orthogonal Matching Pursuit.</td>
</tr>
<tr>
<td>MP</td>
<td>Matching Pursuit.</td>
</tr>
<tr>
<td>MRI</td>
<td>Magnetic Resonance Imaging.</td>
</tr>
<tr>
<td>MSE</td>
<td>Mean-Square Error.</td>
</tr>
<tr>
<td>MSP</td>
<td>Multiple Measurement Vector Subspace Pursuit.</td>
</tr>
<tr>
<td>NSP</td>
<td>Null Space Property.</td>
</tr>
<tr>
<td>OMP</td>
<td>Orthogonal Matching Pursuit.</td>
</tr>
<tr>
<td>pFACS</td>
<td>progressive Fusion of Algorithms for Compressed Sensing.</td>
</tr>
<tr>
<td>RADAR</td>
<td>Radio Detection and Ranging.</td>
</tr>
<tr>
<td>RIC</td>
<td>Restricted Isometry Constant.</td>
</tr>
<tr>
<td>RIP</td>
<td>Restricted Isometry Property.</td>
</tr>
<tr>
<td>RSS</td>
<td>Rademacher Sparse Signals.</td>
</tr>
<tr>
<td>SAR</td>
<td>Synthetic Aperture RADAR.</td>
</tr>
<tr>
<td>SBL</td>
<td>Sparse Bayesian Learning.</td>
</tr>
<tr>
<td>SMV</td>
<td>Single Measurement Vector.</td>
</tr>
<tr>
<td>SONAR</td>
<td>Sound Navigation and Ranging.</td>
</tr>
<tr>
<td>SP</td>
<td>Subspace Pursuit.</td>
</tr>
<tr>
<td>SRA</td>
<td>Sparse Reconstruction Algorithm.</td>
</tr>
<tr>
<td>SRER</td>
<td>Signal-to-Reconstruction-Error Ratio.</td>
</tr>
<tr>
<td>SSR</td>
<td>Sparse Signal Reconstruction.</td>
</tr>
<tr>
<td>StCoMACS</td>
<td>Stage-wise CoMACS.</td>
</tr>
<tr>
<td>StOMP</td>
<td>Stage-wise Orthogonal Matching Pursuit.</td>
</tr>
</tbody>
</table>
# Contents

Acknowledgements i  
Abstract v  
Glossary ix  
Acronyms xi  
List of Publications xvii  
List of Algorithms xx  
List of Matlab Codes for Reproducible Research xxiii  
List of Figures xxv  
List of Tables xxix  

1 Introduction 1  
1.1 Compressed Sensing and Sparse Signal Processing 3  
1.1.1 How Good is the Sparse Assumption? . . . . 3  
1.2 Sparse Reconstruction Algorithms . . . . . . . 5  
1.2.1 $\ell_p$ Norm: Building the Intuition for Sparse Signal Reconstruction . . . . . . . 7  
1.2.2 Convex Relaxation Methods . . . . . . . . 10  
1.2.3 Greedy Pursuits . . . . . . . . . . . . . 11  
1.2.4 Combinatorial Algorithms . . . . . . . . 12  
1.2.5 Non Convex Minimization Algorithms . . . . 12  
1.3 Applications of CS . . . . . . . . . . . . . 12  
1.3.1 Compressive Imaging . . . . . . . . . . . 13  
1.3.2 Compressive RADAR/SONAR . . . . . . . 14  

xiii
4.1 CoMACS: Algorithm .......................... 74
  4.1.1 Theoretical Analysis for CoMACS .......... 76
    4.1.1.1 Sparse Signals ..................... 76
    4.1.1.2 Extension to Arbitrary Signals ...... 79
  4.1.2 CoMACS for Multiple Participating Algorithms 81
4.2 Iterative CoMACS ............................ 83
4.3 Numerical Experiments and Results .............. 87
  4.3.1 Synthetic Sparse Signals ................. 88
    4.3.1.1 Large Dimensional Problems ........ 93
    4.3.1.2 Reproducible Research ............. 94
  4.3.2 Real Compressible Signals ............... 94
4.4 Summary .................................... 96
  4.4.1 Relevant Publications ................... 96
4.A Proof of Proposition 4.1 ..................... 97
4.B Proof of Theorem 4.2 (Analysis of Signal and Measurement Perturbations) ................ 98

5 Progressive Fusion for Low Latency Applications 101
  5.1 Progressive Fusion of Algorithms for Compressed Sensing (pFACS) .................... 102
    5.1.1 Proposed Progressive FACS (pFACS) .... 103
    5.1.2 Theoretical Analysis of pFACS .......... 104
    5.1.3 On Latency of pFACS .................. 105
    5.1.4 pFACS vis-a-vis FACS .................. 106
  5.2 Numerical Experiments and Results ............ 107
    5.2.1 Experiment 1 (Synthetic Signals) ....... 107
      5.2.1.1 Reproducible Research ............ 110
    5.2.2 Experiment 2 (Real-World Signals) .... 110
  5.3 Summary ................................ 111
    5.3.1 Relevant Publication .................. 112

6 Fusion of Algorithms for Multiple Measurement Vectors113
  6.1 Problem Formulation ........................ 115
  6.2 Fusion of Algorithms for Multiple Measurement Vector Problem ....................... 116
  6.3 Theoretical Studies of MMV-FACS .............. 118
    6.3.1 Performance Analysis .................. 119
    6.3.2 Exactly K-sparse Matrix ............... 126
List of Publications

Journal Papers ............................


**Conference Papers**

   *Cited by: 9*.

   *Cited by: 10*.

   *Cited by: 4*.


* Number of citations listed in http://www.google.co.in as on 9th Mar 2015.
<table>
<thead>
<tr>
<th>Section</th>
<th>Algorithm Name</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Matching Pursuit (MP)</td>
<td>35</td>
</tr>
<tr>
<td>2.2</td>
<td>Orthogonal Matching Pursuit (OMP)</td>
<td>37</td>
</tr>
<tr>
<td>2.3</td>
<td>Subspace Pursuit (SP)</td>
<td>38</td>
</tr>
<tr>
<td>2.4</td>
<td>Compressive Sampling Matching Pursuit (CoSaMP)</td>
<td>41</td>
</tr>
<tr>
<td>3.1</td>
<td>Fusion of Algorithms for Compressed Sensing (FACS)</td>
<td>50</td>
</tr>
<tr>
<td>4.1</td>
<td>Committee Machine Approach for Compressed Sensing (CoMACS)</td>
<td>76</td>
</tr>
<tr>
<td>4.2</td>
<td>Stage-wise CoMACS (StCoMACS)</td>
<td>81</td>
</tr>
<tr>
<td>4.3</td>
<td>Iterative CoMACS (ICoMACS)</td>
<td>85</td>
</tr>
<tr>
<td>5.1</td>
<td>Progressive FACS (pFACS)</td>
<td>104</td>
</tr>
<tr>
<td>6.1</td>
<td>MMV Fusion of Algorithms for Compressed Sensing (FACS)</td>
<td>117</td>
</tr>
<tr>
<td>7.1</td>
<td>Iterative Framework for Sparse Reconstruction Algorithms (IFSRA)</td>
<td>152</td>
</tr>
</tbody>
</table>
List of Matlab Codes for Reproducible Research

<table>
<thead>
<tr>
<th>URL</th>
<th>Reference(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td><a href="http://www.ece.iisc.ernet.in/~ssplab/Public/FACS.tar.gz">http://www.ece.iisc.ernet.in/~ssplab/Public/FACS.tar.gz</a></td>
<td>[91], [154], Chapter 3</td>
</tr>
<tr>
<td><a href="http://www.ece.iisc.ernet.in/~ssplab/Public/CoMACS.tar.gz">http://www.ece.iisc.ernet.in/~ssplab/Public/CoMACS.tar.gz</a></td>
<td>[92], Chapter 4</td>
</tr>
<tr>
<td><a href="http://www.ece.iisc.ernet.in/~ssplab/Public/pFACS.tar.gz">http://www.ece.iisc.ernet.in/~ssplab/Public/pFACS.tar.gz</a></td>
<td>[93], Chapter 5</td>
</tr>
<tr>
<td><a href="http://www.ece.iisc.ernet.in/~ssplab/Public/MMVFACS.tar.gz">http://www.ece.iisc.ernet.in/~ssplab/Public/MMVFACS.tar.gz</a></td>
<td>Chapter 6</td>
</tr>
<tr>
<td><a href="http://www.ece.iisc.ernet.in/~ssplab/Public/IFSRA.tar.gz">http://www.ece.iisc.ernet.in/~ssplab/Public/IFSRA.tar.gz</a></td>
<td>[94], Chapter 7</td>
</tr>
<tr>
<td><a href="http://www.ece.iisc.ernet.in/~ssplab/Public/FuGP.tar.gz">http://www.ece.iisc.ernet.in/~ssplab/Public/FuGP.tar.gz</a></td>
<td>[148]</td>
</tr>
<tr>
<td><a href="http://www.ece.iisc.ernet.in/~ssplab/Public/modifiedCoSaMP.tar.gz">http://www.ece.iisc.ernet.in/~ssplab/Public/modifiedCoSaMP.tar.gz</a></td>
<td>[133]</td>
</tr>
<tr>
<td><a href="http://www.ece.iisc.ernet.in/~ssplab/Public/SPEmbeddedOMP.tar.gz">http://www.ece.iisc.ernet.in/~ssplab/Public/SPEmbeddedOMP.tar.gz</a></td>
<td>[132]</td>
</tr>
</tbody>
</table>
## List of Figures

1.1 Comparison of the distribution of pixel values and wavelet coefficients........................................ 4
1.2 Different stages of undersampling and reconstruction of the sparse signal........................................ 6
1.3 Unit spheres (\( \{ x : \| x \|_p = 1 \} \)) in \( \mathbb{R}^2 \) induced by different \( \ell_p \) norms.................................................. 8
1.4 Best \( \ell_p \) norm approximations of \( x \in \mathbb{R}^2 \) by a one dimensional subspace \( D \)........................................ 9
1.5 A broad classification of CS Sparse Reconstruction Algorithms...................................................... 10
1.6 The block diagram of Single-pixel Camera......................................................................................... 13
1.7 SAR image recovery using Notch filtering and sparsity based method.............................................. 15
1.8 A context diagram of the Thesis contribution..................................................................................... 18
2.1 A pictorial demonstration of an underdetermined measurement system, \( \Phi \), acting on a signal \( x_0 \) which has a sparse representation \( x \) in an orthonormal basis \( \Psi \)..................................................... 25
2.2 A pictorial demonstration of an underdetermined measurement system, \( A \), acting on a \( K \)-sparse signal \( x \).......................................................... 25
2.3 A schematic block diagram representing \( k^{th} \) iteration of Orthogonal Matching Pursuit (OMP) algorithm.......................................................... 37
2.4 A schematic block diagram representing the \( k^{th} \) iteration of Subspace Pursuit (SP) algorithm.......................................................... 40
3.1 Schematic block diagram representing the Fusion Framework for Compressed Sensing.............................. 49
3.2 Fusion of two participating algorithms: Gaussian sparse signals.......................................................... 63
3.3 Fusion of two participating algorithms: Rademacher sparse signals.......................................................... 63
<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.4</td>
<td>Fusion of three participating algorithms: Gaussian sparse signal</td>
<td>64</td>
</tr>
<tr>
<td>3.5</td>
<td>Performance of FACS on Real Compressible signals</td>
<td>66</td>
</tr>
<tr>
<td>4.1</td>
<td>Performance comparison of FACS and CoMACS</td>
<td>88</td>
</tr>
<tr>
<td>4.2</td>
<td>Performance of the CoMACS and variants for Gaussian Sparse Signals (GSS) and Rademacher Sparse Signals (RSS)</td>
<td>90</td>
</tr>
<tr>
<td>4.3</td>
<td>Performance of CoMACS, $\ell_1$ based CoMACS (CoMACS$_{L1}$) for Gaussian Sparse Signals</td>
<td>92</td>
</tr>
<tr>
<td>4.4</td>
<td>Performance of CoMACS, $L_1$ based CoMACS (CoMACS$_{L1}$) for large dimensional problems</td>
<td>94</td>
</tr>
<tr>
<td>4.5</td>
<td>Performance of CoMACS based methods for ECG signals</td>
<td>95</td>
</tr>
<tr>
<td>5.1</td>
<td>Progressive performance of pFACS for Rademacher sparse signals</td>
<td>108</td>
</tr>
<tr>
<td>5.2</td>
<td>Performance of pFACS for real-world ECG signals</td>
<td>111</td>
</tr>
<tr>
<td>6.1</td>
<td>Performance of MMV-FACS for Gaussian sparse signal matrices, varying the number of measurements ($M$)</td>
<td>136</td>
</tr>
<tr>
<td>6.2</td>
<td>Performance of MMV-FACS for Rademacher sparse signal matrices, varying the number of measurements ($M$)</td>
<td>137</td>
</tr>
<tr>
<td>6.3</td>
<td>Performance of MMV-FACS for Gaussian sparse signal matrices, varying the number of measurement vectors ($L$)</td>
<td>138</td>
</tr>
<tr>
<td>6.4</td>
<td>Performance of MMV-FACS for Gaussian sparse signal matrices, varying SMNR</td>
<td>139</td>
</tr>
<tr>
<td>6.5</td>
<td>Performance of MMV-FACS for real compressible signals</td>
<td>140</td>
</tr>
<tr>
<td>7.1</td>
<td>Progression of IFSRA(MP) over iterations for Gaussian sparse signal</td>
<td>153</td>
</tr>
<tr>
<td>7.2</td>
<td>Performance of IFSRA(MP) under measurement perturbations</td>
<td>167</td>
</tr>
<tr>
<td>7.3</td>
<td>Performance of IFSRA(CoSaMP) under measurement perturbations</td>
<td>167</td>
</tr>
<tr>
<td>7.4</td>
<td>Performance of IFSRA(BPDN) under measurement perturbations</td>
<td>168</td>
</tr>
</tbody>
</table>
7.5 Performance comparison of IFSRA(MP), IFSRA(CoSaMP), and IFSRA(BPDN) under measurement perturbations 169
7.6 Performance of IFSRA for ECG signals from MIT-BIH Arrhythmia Database . . . . . . . . . . . . . . . . . 170
List of Tables

3.1 Average number of correctly estimated atoms by OMP and SP, for GSS, in clean measurement case .......................... 46
3.2 Performance of FACS on a highly coherent dictionary matrix: .......................................................... 67
4.1 Ratio of average number of true atoms in common support-set and average cardinality of common support-set, for GSS, in clean measurement case ........................................... 75
5.1 Comparison of average computation time (in seconds) by different fusion algorithms for Rademacher sparse signals (RSS) ................................................................. 110
To my grandparents
Introduction

“The last thing that we find in making a book is to know what we must put first.”
Blaise Pascal [1623-1662]

Signal processing community has been interested in sparse signal reconstruction for a long time. The vast activity in the years 1980 – 2000 on transforms and transform coding, particularly in wavelets and frame theory, significantly contributed to this field. The highly influential works in the mid-nineties [1–3] brought out the importance of treating sparse signal reconstruction as an individual research area. Inspired by this, a lot of extensive works on sparse signal modelling took place in the past decade. Donoho and Huo [4] established a theoretical connection, for the first time, between the sparsity seeking transforms and the $\ell_1$-norm measure. The seminal works by Donoho [5,6] and Candès et al. [7–9] ignited a burst of interest in sparse signal modelling by mathematicians (theorists and applied), statisticians, physicists, geophysicists, neuropsychologists, engineers from various fields, computer science theoreticians, and others. A large number of conferences, workshops, and special sessions have been organized in these topics in
recent years. Prestigious journals in the field have allocated special issues for sparse signal modelling and related topics. For example, IEEE-SPL EDICS recently added many entries related to sparse signal modelling. All of these are testimonials for the great attention received by this field.

In this thesis, we deal with a special class of sparse signal modelling problems called Compressed Sensing (CS), which has received wide attention after the pioneering works of Donoho [5, 6] and Candès et al. [7–9]. CS combines sampling and compression of the signal of interest through a non-adaptive, under-sampled linear measurement setup. Though CS was introduced only in the last decade, it has been manifested as a revolution and still one of the most intensively researched topic in many areas like signal processing, sensor systems, and machine learning. Next, we briefly discuss an application where CS has already made life-altering impacts.

One of the key advantages of CS is that it offloads the complexity from data acquisition to data reconstruction. In many applications, the data acquisition is critical as the acquisition time and other resources, including hardware, are very limited. For example, CS has already made life-altering impacts in Magnetic Resonance Imaging (MRI), which is an essential tool of modern medical imaging. MRI plays a key role in investigating the anatomy and function of the body in both health and disease. High resolution MRI requires the patient to lie very still, even the heartbeat need to be stopped, during the measurement process and the imaging speed is often critical for the life of the patient in many MRI applications. Also, in many situations a slow MRI scan may not be feasible due to other reasons [10]. Hence, any significant speed improvement in MRI data acquisition will be life altering. However, the physical (gradient amplitude and slew-rate) and physiological (nerve stimulation) constraints fundamentally limit the speed of MRI data
acquisition [11]. It has been shown that, by exploiting the transform sparsity inherent in MR images, the data acquisition can be sped up by a factor of 7 [12] using the principles of CS.

1.1 Compressed Sensing and Sparse Signal Processing

In CS setup, we assume that the signal-of-interest is sparse/compressible in some orthonormal basis and the task is to recover the signal from under-sampled measurements by exploiting the sparse/compressible nature of the signal. That is, in CS the signal $x_0 \in \mathbb{R}^{N \times 1}$ is modelled as a superposition of a few columns of a given matrix $\Psi \in \mathbb{R}^{N \times N}$. In other words, we have, $x_0 = \Psi x$ where $x$ is sparse. Consider a linear measurement setup $b = Ax$ where $A \in \mathbb{R}^{M \times N}$ represents the measurement system. For example, in MRI applications, $A$ may be formed from a few rows of the DFT matrix and $\Psi$ may be chosen as a wavelet basis matrix. In this thesis, we assume that $\Psi$ is an identity matrix so that $x$ itself is sparse. In CS, the strive is to reduce the number of measurements, $M$, without compromising on the reconstruction quality. In this thesis, we focus on the methods to improve the estimate of the sparse signal given $A$ and $b$.

1.1.1 How Good is the Sparse Assumption?

In practice, we rarely meet exactly sparse signals. However, many signals we deal in applications are found to be sparse or compressible in some transform domain. For example, many images, especially natural images, are highly compressible in the wavelet domain. To elaborate more on this claim, consider the image shown
in Figure 1.1(a). This is an aerial view of the main building of Indian Institute of Science (IISc).

(Figure 1.1(a) courtesy: http://ipc.iisc.ernet.in/~raman/).

The single-level two-dimensional wavelet decomposition with respect to the Haar wavelet [13] of Figure 1.1(a) is shown in Figure 1.1(b). The wavelet decomposition was done independently on the three colour panes (red, green, and blue) of the original image. The upper left sub-image in Figure 1.1(b) shows the image reconstructed from the approximation coefficient matrix and the
other three sub-images are reconstructed from the details coefficients matrices of the single-level two-dimensional wavelet decomposition.

The normalized histogram of the pixel values of the image shown in Figure 1.1(a) is given in Figure 1.1(c). Figure 1.1(d) shows the normalized histogram of the coefficients of the single-level two-dimensional wavelet decomposition of the original image. It may be observed from Figure 1.1(c) and Figure 1.1(d) that the wavelet representation of this image is approximately sparse. That is, most of the wavelet coefficients of the image are close to zero. Sparsity is one of the key assumptions in CS and this is an example to show that sparse signals are ubiquitous in practice.

### 1.2 Sparse Reconstruction Algorithms

In a CS setup, the number of measurements available may be significantly smaller than the dimension of the sparse signal. The reconstruction requires to solve a highly underdetermined system of equations which will result in an infinite number of solutions, in general [14]. However, using CS theory, it has been shown that robust signal reconstruction is possible by exploiting sparse nature of the signal [6–8, 15].

According to the celebrated Whittaker-Nyquist-Kotelnikov-Shannon theorem [16–19], a signal can be reconstructed if the sampling rate is at least twice its bandwidth. If this criterion is not satisfied, aliasing occurs, and in general it is not possible to discern an unambiguous signal. To develop an intuition for reconstructing the signals from the undersampled data, consider the scenario presented in Figure 1.2 [20].
In this figure, we consider a signal which is sparse in the frequency domain. Figure 1.2(a) shows the frequency spectrum of a sparse signal with 3 non-zero elements. We use two sampling criteria: uniform undersampling (lower red bubbles) and random undersampling (upper red bubbles) to sample the signal as shown in Figure 1.2(b). The results of random under sampling and uniform undersampling are depicted in Figure 1.2(c) and Figure 1.2(d) respectively. It may be observed from Figure 1.2(d) that the uniform undersampling resulted in severe aliasing preventing signal reconstruction. However, though aliasing also occurs in the case of random undersampling as shown in Figure 1.2(c), the two strong signal components appear well above the interference level caused by aliasing. These two strong signal components are detected and identified using thresholding as shown in Figure 1.2(e) and Figure 1.2(f). The interference due to the two signal components are calculated and shown in Figure 1.2(h). This estimated interference
is subtracted from the aliased signal to get the signal shown in Figure 1.2(g). Another thresholding on the signal on Figure 1.2(g) recovers the remaining (weak) component of the sparse signal.

Figure 1.2 shows that a random sampling strategy preserves the information of a sparse signal even in an undersampled data and using efficient reconstruction algorithms perfect (near-perfect) signal reconstruction is possible.

1.2.1 $\ell_p$ Norm: Building the Intuition for Sparse Signal Reconstruction

In this section, we briefly discuss the role of $\ell_p$ norm ($p \in [0, \infty]$) in CS which will help to build an intuition behind the principles of Sparse Reconstruction Algorithms (SRAs).

Let $\hat{x}$ denote an estimate of the signal $x$. Generally we use $\ell_p$ norm to measure the approximation error. For a signal $x = [x_1, x_2, \ldots, x_N]^T$, $\ell_p$ norm is defined as

$$
\|x\|_p = \begin{cases} 
|\text{supp}(x)|, & p = 0; \\
\left( \sum_{i=1}^{N} |x_i|^p \right)^{\frac{1}{p}}, & p \in (0, \infty); \\
\max_{i=1:N} |x_i|, & p = \infty.
\end{cases}
$$

(1.1)

Note that $\ell_p$-norm satisfies all the properties of norm iff $p \geq 1$. For $p \in (0, 1)$, $\ell_p$ norm is only a quasi-norm. The $\ell_0$ norm is not even a quasi-norm. However, we can show that $\lim_{p \rightarrow 0} \|x\|_p^p = |\text{supp}(x)| = \|x\|_0$, which justifies the choice of the notation used. Different $\ell_p$ norms have different properties as illustrated in Figure 1.3, and as
we describe next, the choice of $\ell_p$-norm plays a major role in sparse reconstruction problems.

![Unit spheres in $\mathbb{R}^2$ induced by different $\ell_p$ norms](image)

**Figure 1.3:** Unit spheres ($\{x : \|x\|_p = 1\}$) in $\mathbb{R}^2$ induced by different $\ell_p$ norms ($p = 1, 2, \infty, 1/2$) (figure courtesy: [21]).

To illustrate the role of different $\ell_p$ norms in sparse signal reconstruction, consider a sparse signal $x \in \mathbb{R}^2$ which has only one non-zero element. Consider the problem of finding an approximation for $x$, using a point in a one-dimensional affine space $D$. In $\ell_p$ norm sense, this can be achieved by finding $\hat{x} \in D$ which minimizes $\|x - \hat{x}\|_p$. Finding the closest approximation of $x$ in $D$ using $\ell_p$ norm may be viewed as growing an $\ell_p$ sphere (more precisely, a circle in the $\mathbb{R}^2$ case) centered on $x$ until it touches $D$. The intersecting point will be the closest point to $x$ in the chosen $\ell_p$ norm sense. This scenario for different $\ell_p$ norms are illustrated in Figure 1.4. We may observe that $\ell_p$-norm intersects with $D$ at different values for $p = 1, 2, \infty$. It may be observed from Figure 1.4 that the evenness of the distribution of error among the two coefficients are directly proportional to $p$. That is, a smaller $p$ promotes sparsity. For example $\ell_1$-norm intersects with $D$ on the axis where
x lies. Hence \( \hat{x} \) will be sparse as \( x \) is sparse. However, as it may be observed from Figure 1.4, \( \ell_2 \)-norm and \( \ell_\infty \) norm will not yield a sparse solution. A generalization of this behaviour may also be observed in higher dimensional problems and it plays an important role in sparse signal reconstruction problems.

![Diagram of \( \ell_p \)-norm approximations](image)

**Figure 1.4:** An illustration of the best \( \ell_p \)-norm approximations, for \( p = 0, 1, 2, \infty, 1/2 \), of \( x \in \mathbb{R}^2 \) by a one dimensional subspace \( D \) (figure courtesy: [21]).

For CS signal reconstruction, the optimal method is solving an \( \ell_0 \)-minimization problem, which is Non-deterministic Polynomial-time hard (NP-hard). For practical implementations, many suboptimal algorithms are introduced in recent years which may be broadly categorized into four classes as shown in Figure 1.5.
FIGURE 1.5: A broad classification of CS Sparse Reconstruction Algorithms.

Next, we briefly discuss each category. A more elaborated discussion on a few SRAs, relevant for the thesis, is given in Chapter 2.

1.2.2 Convex Relaxation Methods

Convex relaxation methods are so popular in CS literature that many consider it as a synonym for SRAs. This is mainly due to the fact that Convex Relaxation Methods (CRM) were the first to
provide elegant theoretical guarantees for sparse signal reconstruction in the CS framework. Another factor is the off-the-shelf availability of excellent toolboxes to solve convex problems efficiently and accurately. In convex relaxation methods, the $\ell_0$ minimization problem is relaxed using an $\ell_1$ minimization problem.

Examples of popular CRM include Basis Pursuit (BP) [2] and Basis Pursuit De-Noising (BPDN) [36], modified BPDN [40], Least Absolute Shrinkage and Selection Operator (LASSO) [37,41], Least Angle Regression (LARS) [39], and Dantzig Selector (DS) [38]. The number of measurements required by CRM for exact signal reconstruction is small. However, CRM are computationally very expensive which make them less attractive for practical applications.

1.2.3 Greedy Pursuits

Greedy pursuits find the estimate of the sparse signal step by step, in an iterative fashion. They possess simple geometric interpretations and like CRM, many of them show elegant theoretical guarantees. The advantages in terms of computational complexity and memory requirements make them more attractive for applications. The popular greedy pursuits include Matching Pursuit (MP) [1], Orthogonal Matching Pursuit (OMP) [26, 27], Stage-wise Orthogonal Matching Pursuit (StOMP) [31], Subspace Pursuit (SP) [28], Compressive Sampling Matching Pursuit (CoSaMP) [29], and Iterative Hard Thresholding (IHT) [30].
1.2.4 Combinatorial Algorithms

Greedy pursuits provide computational advantage over CRM, both empirically and theoretically. Combinatorial methods are significantly fast and efficient than both CRM and greedy pursuits. However, these methods require specific pattern in the measurements which may not be feasible to realize in many applications. They reconstruct sparse signals following the principles of group testing [42]. The popular algorithms in this area include Heavy Hit- ters on Steroids (HHS) [24], Chaining Pursuits (CP) [23], Fourier Sampling Algorithm (FSA) [22], and Sudocodes [25].

1.2.5 Non Convex Minimization Algorithms

It has been shown that, instead of relaxing $\ell_0$ minimization problem to an $\ell_1$ minimization problem, we may also solve non convex relaxation problems for efficient sparse signal reconstruction. A popular method is to replace $\ell_0$ minimization with $\ell_q$ minimization problem where $0 < q < 1$. Another strategy is to use a Bayesian framework which exploits the sparse nature of the signal. Examples of the popular work in this family are Iterative Re-weighted L1 (IRL1) [33], Iterative Support Detection (ISD) [35], and Sparse Bayesian Learning (SBL) [34, 43–46].

1.3 Applications of CS

In many applications it is highly desirable to reduce the number of measurements without reducing the reconstruction quality since it gives several advantages like reduction in the number of sensors, simpler hardware design, faster acquisition time, and less power
consumption. Due to these potential advantages, though CS is a relatively new area, it has been successfully applied in many fields. In this section, we briefly discuss some of the applications.

### 1.3.1 Compressive Imaging

CS has far reaching implications in imaging as it reduces the number of measurements and hence cut down power consumption, storage space, hardware complexity, and acquisition time. The single pixel camera [47] developed by Rice university is one of the first applications built using CS principles. The block diagram of single pixel camera is given in Figure 1.6. Recently Huang et al. [48] proposed a lensless compressive imaging architecture for capturing images of visible and other spectra such as infrared, or millimeter waves.

![Figure 1.6: The block diagram of Single-pixel Camera.](source: http://www dsp ece rice edu/cscamera/)

CS has also found life altering applications in the field of medical imaging. It has been used to reduce the MRI scanning time [11, 20]. CS has also been successfully applied to seismic imaging [49].
Chapter 1

1.3.2 Compressive RADAR/SONAR

Radio Detection and Ranging (RADAR) and Sound Navigation and Ranging (SONAR) are widely used in many civilian, military, and bio-medical applications. However, the resolution in these applications are limited by the classic time-frequency uncertainty principles. CS has been shown to produce promising results in images of RADAR/SONAR using relatively a smaller number of measurements than the conventional methods. Compressive RADAR eliminated the need of pulse compression matched filter at receiver and reduces A/D conversion bandwidth which simplifies the hardware design [50]. Sparse sampling has been applied to both time and frequency domain to enhance pulse compression technique in order to efficiently compress, restore and recover the RADAR data [51–53]. The optimization of waveforms for CS application in RADAR is discussed by He et al. [54] and Kyriakides et al. [55]. CS has also found applications in Passive coherent location (PCL) [56, 57], Synthetic Aperture RADAR (SAR) [58–60], through-the-wall RADAR [61], and SONAR and ground penetrating RADAR [62–64]. The advantage of using CS in SAR image recovery is shown in Figure 1.7.

CS has also been widely used in many other applications like Compressive micorarrays [65], group testing [66], A/D converters [67,68], communication and networks. Examples include sparse channel estimation [69,70], spectrum sensing [71,72], ultra wideband systems [73], wireless sensor networks [74], network management [75], network data mining [76] and network security [77]. A more elaborated list of applications and references can be found at [78] and [79].
1.4 Challenges/ Problems Identified

Though many SRAs possess elegant theoretical guarantees for sparse signal recovery, it is well known that the performance of any sparse recovery algorithm depends on several factors like signal dimension, sparsity level of signal, and measurement noise power [8, 80–82]. Empirically, it has been also observed that the recovery performance varies significantly and depends on the underlying statistical distribution of the non-zero elements of the sparse signal [81, 82]. If this distribution is known \textit{a priori}, we can employ the best recovery algorithm suitable for that type of signal and get the best sparse signal estimate. In many practical applications, we may not have this prior knowledge and hence, we cannot use the appropriate method to achieve the best performance.

It can be also seen that every sparse recovery algorithm requires a minimum number of measurements (algorithm dependent) for sparse signal recovery and performs poorly in a very low dimension measurement scenario [81–84]. The reduction in number of measurements leads to reduction in the number of sensors and/or
reduction in the measurements time. Hence in many applications, it is highly desirable to have a reduced number of measurements. For example, in IR camera [85], the sensors are very costly. In medical applications like MRI, where we have to even stop the heartbeat of the patient during the measurements process to get a high resolution MRI, the reduction in measurement time is often critical for the life of the patient [11].

Though it is evident that the performance of the sparse recovery algorithms degrades in cases where only a limited number of measurements are available or the statistical distribution of the sparse signal is unknown, it is interesting to observe (empirically) that this degradation does not always imply a complete failure [81,82]. The estimate obtained by the algorithm will often contain partially correct information about the sparse signal. By exploiting the partial information about the target signal, it may be possible to get a better sparse signal estimate. In this thesis, we explore this possibility and propose novel methods to improve the performance of arbitrary sparse signal reconstruction algorithms.

1.5 Contributions

In this thesis, we propose novel frameworks and algorithms to improve the performance of any arbitrary SRA.

- We propose a fusion framework which fuses the estimates of multiple participating algorithms to result in a better sparse signal estimate.

- We also propose an iterative framework which improves the performance of any arbitrary sparse signal reconstruction algorithm without modifying the underlying algorithm.
Fusion Framework

To improve the sparse signal estimate, we propose a fusion framework where we employ multiple SRAs which are run in parallel, independently. The estimates obtained by these participating algorithms are fused efficiently to get a sparse signal estimate which is often better than the best estimate provided by the participating algorithms independently. We propose different schemes for fusion. The proposed schemes use the participating algorithm as a black box, and does not require any change in the underlying participating algorithm. We mathematically analyse the proposed schemes and verify the robustness against signal and measurement perturbations. We demonstrate the efficiency and effectiveness of the proposed methods in applications through extensive numerical experiments on both synthetic and real-world data.

Iterative Framework:

In the iterative framework, we exploit the partial information about the sparse signal available in the estimate obtained by a given arbitrary SRA. We use this information in the subsequent iterations to improve the sparse signal reconstruction iteratively. The proposed iterative algorithm is also general in nature, which can incorporate any SRA as a participating algorithm. We derive convergence guarantees for the proposed algorithm, and demonstrate its advantage in applications using simulations.
A context diagram of the Thesis contribution is shown in Figure 1.8.

Figure 1.8: A context diagram of the Thesis contribution.

1.6 Organization of the Thesis

In this section, we give an overview of the organization of the thesis and briefly discuss the contributions in each chapter.
Chapter 2

In Chapter 2, we briefly introduce CS and discuss a few desirable properties of the measurement system which are sufficient to guarantee signal reconstruction. We also illustrate a few popular SRAs widely used in subsequent chapters of this thesis. We also provide some existing theoretical results which will be useful while theoretically analysing the proposed methods.

Chapter 3

In Chapter 3, we explain the motivation behind this research and develop a framework to fuse the estimates of multiple participating algorithms. We also propose a fusion algorithm called Fusion of Algorithms for Compressed Sensing (FACS) and derive theoretical guarantees for performance improvement. Then we perform extensive numerical experiments using some of the well known SRAs in CS which confirms the effectiveness of the proposed scheme. The work described in this chapter has been published in *IEEE Transactions of Signal Processing* [91].

Chapter 4

We propose another fusion algorithm called Committee Machine Approach for Compressed Sensing (CoMACS) in Chapter 4. Variants of CoMACS are also proposed here to further enhance the sparse reconstruction quality. We also study the theoretical aspects of the proposed methods. We also show that the proposed algorithms produce better sparse signal estimates compared to the participating algorithms. The work summarized in this chapter has been published in *IEEE Transactions on Signal Processing* [92].
Chapter 5

FACS and CoMACS are not latency friendly algorithms. For low latency applications, we propose a latency friendly fusion algorithm called progressive Fusion of Algorithms for Compressed Sensing (pFACS) in Chapter 5. We also discuss the theoretical guarantees of pFACS and show the efficacy of pFACS using extensive numerical experiments. This work has been published in Signal Processing [93].

Chapter 6

In Chapter 6, we extend the fusion framework and FACS to the Multiple Measurement Vector (MMV) problem. We theoretically analyse the proposed algorithm and derive sufficient conditions for improving the performance. Further, we corroborate the claim with the average case analysis. Using extensive simulations, we show that the proposed method provides significant performance improvement.

Chapter 7

In Chapter 7, we develop a novel iterative framework called Iterative Framework for Sparse Reconstruction Algorithms (IFSRA) which can be used to improve the performance of any arbitrary SRA. We theoretically analyse IFSRA and derive convergence guarantees. Using numerical experiments, we show that IFSRA improves performance of SRAs. This work that has been published in Signal Processing [94].
Chapter 8

In Chapter 8, we give the conclusions we have drawn from our research and suggest a few ideas for related future work.

1.7 Summary

In this chapter, we briefly discussed the importance of sparse signal modelling, sparse signal processing and CS. The we discussed a few applications to motivate about the significance of compressed sensing and sparse signal processing. We also used a few examples to give the intuition behind the working principles of CS and sparse signal reconstruction. A few challenges identified in this field and the solutions offered in this thesis are also briefly discussed in the later part of this chapter.
CS and Sparse Signal Reconstruction: Background

“I may not agree with what you say, but I’ll defend to the death your right to say it.”
Voltaire [1694-1778]

First, we briefly review Compressed Sensing (CS). Then we discuss some theoretical results and a few popular Sparse Reconstruction Algorithms (SRAs) in CS which are widely used in the subsequent chapters.

2.1 Signal Models

With the seminal works of Donoho [5] and Candès et al. [8, 15], CS has emerged as a new framework for signal acquisition which allows large reduction in the cost of acquiring signals that have a sparse or compressible representation in some transform domain. Consider a standard measurement system that produces $M$ linear
measurements of the signal $x_0 \in \mathbb{R}^{N \times 1}$ which can be mathematically expressed as

$$b = \Phi x_0 + w, \quad (2.1)$$

where $\Phi \in \mathbb{R}^{M \times N}$ represents the measurement system, $b \in \mathbb{R}^{M \times 1}$ represents the measurement vector, and $w \in \mathbb{R}^{M \times 1}$ represents the additive measurement noise present in the system. Let the signal $x_0$ have a $K$-sparse representation in some transform domain with orthonormal basis $\Psi \in \mathbb{R}^{N \times N}$. Let $x \in \mathbb{R}^{N \times 1}$ denote the $K$-sparse representation of $x_0$ such that

$$x_0 = \Psi x. \quad (2.2)$$

Note that, $x$ is a $K$-sparse signal. Mathematically, a signal is said to be $K$-sparse, if it has at most $K$ non-zero elements ($\|x\|_0 \leq K$). Let $\mathcal{T}$ denote the support-set of $x$ with $|\mathcal{T}| \leq K$.

Combining (2.1) and (2.2), we get

$$b = \Phi \Psi x + w. \quad (2.3)$$

A pictorial representation of (2.3) is given in Figure 2.1.

We can re-write (2.3) as

$$b = Ax + w,$$

where $A = \Phi \Psi \in \mathbb{R}^{M \times N}$. Unless otherwise stated, throughout this thesis, we assume $\Psi = I$ so that $A$ represents the measurement system. That is, in this thesis, we consider the standard CS measurement setup which acquires a $K$-sparse signal $x \in \mathbb{R}^{N \times 1}$ using

---

1Support-set of a vector is defined as the set of indices of non-zero elements of the vector.
$\begin{align*}
\mathbf{b}_{M \times 1} &= \Phi_{M \times N} \mathbf{x}_{N \times 1} + \mathbf{w}_{M \times 1}.
\end{align*}$

Figure 2.1: A pictorial demonstration of an underdetermined measurement system, $\Phi$, acting on a signal $\mathbf{x}_0$ which is has a sparse representation $\mathbf{x}$ in an orthonormal basis $\Psi$ (original source: [95]).

$M(\ll N)$ linear measurements via the following relation

$$
\mathbf{b} = \mathbf{A}\mathbf{x} + \mathbf{w},
$$

(2.4)

where $\mathbf{b} \in \mathbb{R}^{M \times 1}$ denotes the measurement vector, $\mathbf{A} \in \mathbb{R}^{M \times N}$ denotes the measurement matrix, and $\mathbf{w} \in \mathbb{R}^{M \times 1}$ denote the additive measurement noise. The measurement matrix $\mathbf{A}$ is also called projection matrix in literature. A pictorial representation of (2.4) is shown in Figure 2.2.

Figure 2.2: A pictorial demonstration of an underdetermined measurement system, $\mathbf{A}$, acting on a $K$-sparse signal $\mathbf{x}$ (original source: [95]).
Chapter 2  
CS and Sparse Signal Reconstruction: Background

Note that (2.4) is an underdetermined system and solving \( x \) from (2.4) is an ill-posed problem, in general [14]. In the noiseless case \((w = 0)\), (2.4) may be also viewed as dimensionality reduction of a high dimensional sparse vector.

There are two main theoretical questions in CS:

i) How should we design the measurement matrix \( A \) to ensure that it preserves the information in the signal \( x \)?

ii) How can we reconstruct the original signal \( x \) from the measurements \( b \)?

2.2 Measurement System

The measurement matrix \( A \) in (2.4) may be viewed as a dimensionality reduction operator which maps signals in \( \mathbb{R}^{N \times 1} \) to \( \mathbb{R}^{M \times 1} \) \((M \ll N)\). CS strives to reduce the measurements without compromising on the reconstruction quality. Hence, efficient design of the measurement system \( A \) involves two main tasks:

i) To reduce the number of measurements as far as possible,

ii) To preserve the information of the signal in the measurements.

A few desirable properties for efficient designs of \( A \) are discussed below.

2.2.1 Null Space Property (NSP)

Let \( \Sigma_K = \{ x \in \mathbb{R}^{N \times 1} : \| x \|_0 \leq K \} \) denote the set of all \( K \)-sparse signals in the \( N \)-dimensional space. Let \( \mathcal{N}(A) = \{ z : A z = 0 \} \)
denote the null space of $A$. To recover all $K$-sparse signals from the measurements $b$, it is necessary and sufficient that, for any pair of distinct $K$-sparse vectors $x_1$ and $x_2$, we must have, $Ax_1 \neq Ax_2$. Formally, $A$ uniquely represents all $x \in \Sigma_K$ iff $\mathcal{N}(A)$ does not contain any $2K$ sparse vectors. This property is widely characterized by spark [96] which is defined as follows.

**Definition 2.1 (Spark [96]).** Spark of matrix $A$ is defined as the smallest number of linearly dependent columns of $A$.

**Theorem 2.1 (Corollary 1 of [96]).** For any vector $b \in \mathbb{R}^{M \times 1}$, there exists at most one signal $x \in \Sigma_K$ such that $b = Ax$ if and only if $\text{spark}(A) > 2K$.

For $A \in \mathbb{R}^{M \times N}$ ($2 \leq M < N$), we have, $2 \leq \text{spark}(A) \leq M + 1$. Hence, as a consequence of Theorem 2.1, we have $M \geq 2K$ is a necessary condition for unique sparse signal recovery. Though for exact $K$-sparse signals spark provides a necessary and sufficient condition for sparse signal recovery, for approximately sparse signals (compressible signals) more restrictive conditions on null space, called Null Space Property (NSP) [97], is required.

**Definition 2.2.** A matrix $A$ satisfies the NSP of order $K$ if there exist a $C > 0$ such that

$$\|z_{T_1}\|_2 \leq C \frac{\|z_{T_1}\|_1}{\sqrt{K}}$$

holds for all $z \in \mathcal{N}(A)$ and for all $T_1 \subset \{1, 2, 3, \ldots, N\}$ with $|T_1| \leq K$.

While NSP provides a necessary and sufficient condition for establishing convergent guarantees (typically upper bounds on reconstruction errors) for arbitrary signals, these guarantees do not
cater for errors due to measurement noise or quantization. Candès et al. [7–9] introduced an isometric condition on the measurement matrix called Restricted Isometry Property (RIP). In this thesis, we extensively use RIP while theoretically analysing our proposed methods.

### 2.2.2 Restricted Isometry Property (RIP)

**Definition 2.3.** A matrix $A$ satisfies RIP [7–9] if there exist a constant $\delta \in [0, 1)$ such that

\[
(1 - \delta) \|x\|^2_2 \leq \|Ax\|^2_2 \leq (1 + \delta) \|x\|^2_2 \tag{2.5}
\]

holds for any $K$-sparse vector $x$. The Restricted Isometry Constant (RIC) $\delta_K \in [0, 1)$ is defined as the smallest constant for which RIP property holds for all $K$-sparse vectors.

If $A$ satisfies RIP of order $2K$, then $A$ approximately preserves the distance between any pair of $K$-sparse vectors. RIP may be viewed as a less general form of stable embedding property [98] of sparse vectors.

Next we discuss a few results due to RIP, which will be widely used in the theoretical discussions of subsequent chapters of this thesis.

**Proposition 2.1.** (Proposition 3.1 in [29]) Let $A$ have RIC $\delta_r$ and let $\mathcal{T}$ denote a set of $r$ indices or fewer. Then, for an arbitrary $z$, we have

\[
\left\| \left( A_H^T A_T \right)^{-1} z \right\|_2 \leq \frac{1}{1 - \delta_r} \|z\|_2, \tag{2.6}
\]

and

\[
\left\| A_T^\dagger z \right\|_2 \leq \frac{1}{\sqrt{1 - \delta_r}} \|z\|_2. \tag{2.7}
\]
Proposition 2.2. (Proposition 3.2 in [29]) Let $A$ have RIC $\delta_r$. Let $\mathcal{T}_1$ and $\mathcal{T}_2$ be two disjoint sets of indices of columns of $A$ such that $|\mathcal{T}_1 \cup \mathcal{T}_2| \leq r$. Then
\[
\|A_{\mathcal{T}_1}^H A_{\mathcal{T}_2}\|_2 \leq \delta_r. \tag{2.8}
\]

Corollary 2.1. (Corollary 3.3 in [29]) Let $A$ have RIC $\delta_r$ and let $S$ denote a set of column indices from $A$. Let $x$ be a sparse vector with support-set $\mathcal{T}$ such that $r \geq |\mathcal{T} \cup S|$. Then we have
\[
\|A_{S}^H A_{S^c} x_{S^c}\|_2 \leq \delta_r \|x_{S^c}\|_2. \tag{2.9}
\]

Proposition 2.3. (Lemma 2 in [28]) Consider $A \in \mathbb{R}^{M \times N}$, and let $\mathcal{T}_1, \mathcal{T}_2$ be two subsets of $\{1, 2, \ldots, N\}$ such that $\mathcal{T}_1 \cap \mathcal{T}_2 = \emptyset$. Assume that $\delta_{|\mathcal{T}_1|+|\mathcal{T}_2|} < 1$, and let $y \in \text{span}(A_{\mathcal{T}_1})$ and $r = y - A_{\mathcal{T}_2} A_{\mathcal{T}_2}^\dagger y$, then we have
\[
\left(1 - \frac{\delta_{|\mathcal{T}_1|+|\mathcal{T}_2|}}{1 - \delta_{\text{max}(|\mathcal{T}_1|, |\mathcal{T}_2|)}}\right) \|y\|_2 \leq \|r\|_2 \leq \|y\|_2. \tag{2.10}
\]

Lemma 2.1. (Lemma 3 in [28]) Consider the measurement system $b = Ax + w$, where $x \in \mathbb{R}^N$ is a $K$-sparse signal vector, $w \in \mathbb{R}^M$ denote the additive measurement noise, and $A \in \mathbb{R}^{M \times N}$ represent the sampling matrix with RIC $\delta_K$. For an arbitrary set $\mathcal{T}_1 \subset \{1, 2, \ldots, N\}$ with $|\mathcal{T}_1| \leq K$, define $x_{\mathcal{T}_1} = A_{\mathcal{T}_1}^\dagger b$ and $x_{\mathcal{T}_1^c} = 0$. Then, the following inequality holds.
\[
\|x - \hat{x}\|_2 \leq \frac{1}{1 - \delta_{2K}} \|x_{\mathcal{T}_1^c}\|_2 + \frac{1 + \delta_{2K}}{1 - \delta_{2K}} \|w\|_2. 
\]
Note that though the result in Lemma 3 of [28] contains only $\delta_{3K}$, the lemma is even valid for $\delta_{2K}$ (see proof of Lemma 3 in [28] for more details).
Lemma 2.2. (Lemma 15 in [24]) Let the measurement matrix $A \in \mathbb{R}^{M \times N}$ have RIC $\delta_K$. Then for an arbitrary $x \in \mathbb{R}^{N \times 1}$, we have

$$\|Ax\|_2 \leq \sqrt{1 + \delta_K} \left(\|x\|_2 + \frac{1}{\sqrt{K}} \|x\|_1\right).$$

2.2.2.1 Measurement bounds using RIP

The following theorem gives a lower bound on the number of measurements ($M$) to achieve RIP.

**Theorem 2.2** (Theorem 3.5 of [99]). Let $A \in \mathbb{R}^{M \times N}$ satisfies RIP of order $2K$ with RIC $\delta_{2K} \in (0, \frac{1}{2}]$. Then

$$M \geq CK \log\left(\frac{N}{K}\right),$$

(2.11)

where $C = \frac{1}{2} \log(\sqrt{24} + 1) \approx 0.28$.

2.2.3 Coherence

Though spark, NSP, and RIP provide theoretical guarantees for the recovery of sparse signals, verifying whether a given matrix satisfies these properties is a highly combinatorial problem which requires extensive search over all $\binom{N}{K}$ sub-matrices. Unlike these properties, coherence [27, 96] of a matrix is easily computable which can also provide convergence guarantees for the recovery of sparse signals.

**Definition 2.4** (Coherence). The coherence of a matrix $A \in \mathbb{R}^{M \times N}$, denoted by $\mu(A)$, is defined as the largest absolute inner product between any pair of columns $a_i$ and $a_j$ ($i \neq j$) of $A$:

$$\mu(A) = \max_{1 \leq i < j \leq N} \frac{|\langle a_i, a_j \rangle|}{\|a_i\|_2 \|a_j\|_2}$$
Chapter 2 CS and Sparse Signal Reconstruction: Background

We have, for any matrix $A \in \mathbb{R}^{M \times N}$, $\sqrt{\frac{N-M}{M(N-1)}} \leq \mu(A) \leq 1$. The lower bound is known as the Welch bound [100–102] which can be approximated by $\frac{1}{\sqrt{M}}$ for $M \ll N$. Under certain conditions coherence can be related to the spark, NSP, and RIP. For example, $\text{spark}(A) \geq 1 + \frac{1}{\mu(A)}$.

2.2.4 Measurement Matrix Constructions

Though many schemes have been proposed to construct measurement matrix, the most optimal construction in terms of the number of measurements is provided by the random matrix constructions. It has been also shown that the random matrices will satisfy the RIP with high probability if the entries are chosen from a sub-Gaussian distribution [5, 15]. The linkage between the RIP property of random matrices and Johnson-Lindenstrauss (JL) lemma [103] was studied by Baraniuk et al. [104] and derived a simpler proof of the RIP for random matrices. The measurements obtained using random matrices are democratic, which in general contains equal information about the measured signal. Hence random measurement matrices are robust to the measurement noise up to a reasonable level. Examples of sub-Gaussian distributions include Gaussian and Bernoulli distributions. In this thesis, we use Gaussian random matrices as the measurement matrices in the numerical experiments. In practice, we may be interested in acquiring a signal which is sparse w.r.t. some orthonormal basis $\Phi$. If $A$ is a Gaussian random matrix, similar to $A$, $A\Phi$ will also satisfy RIP with high probability for sufficiently large $M$ [104].
2.3 Reconstruction Algorithms

Let us assume that our measurements are noise-free such that (2.4) reduces to

\[ b = Ax. \]  \hspace{1cm} (2.12)

Reconstructing \( x \) from a small number of measurements from (2.12) is an ill-posed problem and unique signal reconstruction is not possible, in general. However, for a \( K \)-sparse signal, unique signal reconstruction is possible by solving an optimization problem of the form

\[
\min \|x\|_0 \quad \text{subject to} \quad Ax = b, \quad (2.13)
\]

provided \( \text{spark}(A) > 2K \).

We rarely meet exactly sparse signals in practice. However, many signals found in real life are compressible or compressible in some transform domain. Also, in many applications, we need to cater for the noise present in the measurement system. In such situations, we rather solve a robust version of (2.13), given as

\[
\min \|x\|_0 \quad \text{subject to} \quad \|Ax - b\|_2 \leq \epsilon, \quad (2.14)
\]

where \( \epsilon \) represents the tolerance factor for signal and measurement contamination. It is interesting to note that (2.14) may be expressed in two alternate, but equivalent ways. An unconstrained alternative for (2.14) is given by

\[
\min \|x\|_0 + \lambda \|Ax - b\|_2, \quad (2.15)
\]
where $\lambda$ represents a regularization parameter which controls the trade-off between the sparsity of the solution and its fidelity with the measurements.

Another equivalent alternative can be written as

$$\min_x \|Ax - b\|_2 \quad \text{subject to } \|x\|_0 \leq K, \quad (2.16)$$

where we specify the desired level of sparsity, $K$.

Solving these optimization problems require an extensive search through all possible sets of $K$ columns of $A$. Unfortunately, this is a Non-deterministic Polynomial-time hard (NP-hard) problem [105] which leaves $\binom{N}{K}$ possibilities. Hence this strategy is not feasible even for the modest values of $N$ and $K$.

Many sub-optimal algorithms and heuristics have been proposed over the past few years to solve this and closely related problems in signal processing and other areas. Next we provide a brief overview of the key reconstruction algorithms we have used in this thesis. Interested reader may refer Tropp and Wright [106], and Eldar and Kutyniok [21], and the references therein for a more thorough survey of CS reconstruction methods.

### 2.3.1 Convex Relaxation Methods

It may be observed that the difficulty in solving (2.13) and equivalent alternatives arises from the fact that the $\ell_0$-norm is a highly non-convex function. One of the popular approaches to make these problems tractable is to replace the $\ell_0$-norm with a convex relaxation function. By replacing $\ell_0$-norm with the closest convex norm function, $\ell_1$-norm, we obtain (2.13) as

$$\min \|x\|_1 \quad \text{subject to } Ax = b, \quad (2.17)$$
Chapter 2 CS and Sparse Signal Reconstruction: Background

and (2.14) as

$$\min_x ||x||_1 \text{ subject to } ||Ax - b||_2 \leq \epsilon.$$  \hfill (2.18)

In literature (2.17) is referred as Basis Pursuit (BP) and (2.18) is referred as Basis Pursuit De-Noising (BPDN) [2, 36]. The $\ell_1$ variant of (2.16), known as Least Absolute Shrinkage and Selection Operator (LASSO) [37, 41, 107], received wide attention in Statistics literature for variable selection in regression.

Another popular convex relaxation method known as Dantzig Selector (DS) [38] solves

$$\min_x ||x||_1 \text{ subject to } ||Ax - b||_\infty \leq \epsilon.$$ \hfill (2.19)

Many efficient algorithms have been proposed in recent literature to solve these problems efficiently [108–115]. Many of these are iterative algorithms which are proven to solve the respective convex objective function.

### 2.3.2 Greedy Methods

Convex Relaxation Methods (CRM) are often used as synonymous with Sparse Signal Reconstruction (SSR) methods in CS. However, CRM are often very expensive in terms of computation and memory requirements and hence not suitable for large dimensional/real-time applications. Many of these methods are neither flexible nor easy to implement on hardware. An alternate strategy for approximating (2.14) is using greedy family of algorithms. Greedy algorithms have simple geometric interpretation, and like CRM, many of them possess elegant theoretical guarantees.
Chapter 2 CS and Sparse Signal Reconstruction: Background

Let \( A = [a_1, a_2, \ldots, a_N] \), where \( a_i \) denotes the \( i^{\text{th}} \) column (often called ‘atom’ in literature) of \( A \). Using these notations, we can re-write (2.4) as

\[
b = Ax + w = \sum_{i \in \mathcal{T}} a_i x_i + w = A_T x_T + w. \tag{2.20}
\]

Popular greedy algorithms try to exploit the structure in (2.20) to estimate the sparse signal. Next, we discuss a few greedy algorithms which will be extensively used in later chapters of this Thesis.

2.3.3 Matching Pursuit (MP) [1, 116]

Matching Pursuit (MP) [1,116] is one of the early proposed greedy algorithms. MP algorithm is described in Algorithm 2.1. Qian and Chen [117] proposed a similar algorithm and applied it to the Gabor dictionary.

Algorithm 2.1: Matching Pursuit (MP) [1,116]

**Inputs:** \( A_{M \times N}, b_{M \times 1} \), and \( K \).

**Initialization:** \( k = 0 \), \( r_0 = b \), and \( \hat{x}_0 = 0 \in \mathbb{R}^{N \times 1} \)

1: \textbf{repeat}
2: \hspace{1em} \( k = k + 1 \);
3: \hspace{1em} \( i_k = \arg \max_{i=1:N} |a_i^T r_{k-1}| \);
4: \hspace{1em} \( \hat{x}_{i_k} = \hat{x}_{i_k} - a_{i_k}^T r_{k-1} \);
5: \hspace{1em} \( r_k = b - A \hat{x}_k \);
6: \textbf{until} \( (k \geq K) \);
7: \( \hat{x} = \hat{x}_K \);

**Output:** \( \hat{x} \).

MP starts with initial solution \( \hat{x} = 0 \) and the residual \( r_0 = b \). In each iteration the signal is approximated as \( \hat{x}_{i_k} = \hat{x}_{i_k} - a_{i_k}^T r_{k-1} \) where \( a_{i_k} \) denotes the atom of \( A \) which gives maximum correlation
Chapter 2 CS and Sparse Signal Reconstruction: Background

With the residual \( r_{k-1} \). Then MP regularizes the measurement vector \( b \) using \( \hat{x}_k \), the current estimate of \( x \), to get the new residual \( r_k \).

2.3.4 Orthogonal Matching Pursuit (OMP) [26, 27]

The main disadvantage of MP is that the convergence of MP heavily relies on the orthogonality of the residual to the dictionary [1, 26]. As a result, although asymptotic convergence is guaranteed, MP could be sub-optimal even after any finite number of iterations [26].

Though many extensions have been proposed to MP, the most popular extension is Orthogonal Matching Pursuit (OMP) [26, 27]. Throughout the iterations, OMP maintains the full backward orthogonality of the residual which help OMP to provide a better convergence guarantee [26, 27]. To estimate a \( K \)-sparse signal, like MP, OMP works in a serial fashion and identifies one atom per iteration to estimate \( K \) support atoms in \( K \) iterations. In each iteration, OMP selects the atom which gives the highest correlation with the regularized measurement vector. The regularized measurement vector is obtained by subtracting the contribution of a partial estimate of the signal, obtained in the previous iteration, from the original measurement vector. In each iteration, the partial estimate of the signal \( x \) is obtained by projecting \( b \) orthogonally onto the columns of \( A \) associated with the current estimated support-set. A schematic block diagram illustrating the main steps involved in the \( k^{th} \) iteration of OMP is shown in Figure 2.3. The main difference between MP and OMP is that, MP can repeatedly select the same atom whereas the Least-Squares (LS) step in OMP (Step 5 in Algorithm 2.2) ensures that an already selected atom
Chapter 2  CS and Sparse Signal Reconstruction: Background

![Diagram](image)

**Figure 2.3:** A schematic block diagram representing \( k \)th iteration of OMP algorithm.

will not be selected again in later iterations. The OMP algorithm is summarized in Algorithm 2.2.

**Algorithm 2.2:** Orthogonal Matching Pursuit (OMP) [26, 27]

*Inputs:* \( A_{M \times N}, b_{M \times 1}, \) and \( K. \)

*Initialization:* \( k = 0, r_0 = b, \) and \( \hat{T}_0 = \emptyset; \)

1. repeat
2. \( k = k + 1; \)
3. \( i_k = \arg \max_{i=1:N, i \notin \hat{T}_{k-1}} |a_i^T r_{k-1}|; \)
4. \( \hat{T}_k = i_k \cup \hat{T}_{k-1}; \)
5. \( r_k = b - A_{\hat{T}_k} A_{\hat{T}_k}^\dagger b; \)
6. until \( (k \geq K); \)
7. \( \hat{T} = \hat{T}_k; \)
8. \( \hat{x}_{\hat{T}} = A_{\hat{T}}^\dagger b, \hat{x}_{\hat{T}_c} = 0; \)

*Outputs:* \( \hat{x} \) and \( \hat{T}. \)

It has been shown that, in a noiseless measurement case \( (w = 0) \) OMP recovers a \( K \)-sparse signal in exactly \( K \) iterations. This result has been shown for both the measurement matrices satisfying RIP [118] and the measurement matrices satisfying bounded coherence [90]. Both the results hold only when \( M \geq K^2 \log(N). \) Many improvements have been suggested recently to improve the
basic results using RIP [119, 120]. The convergence of OMP using RIP has also been extended for non-sparse signals and measurement perturbations cases [121, 122].

### 2.3.5 Subspace Pursuit (SP) [28]

An obvious drawback of MP and OMP is that both algorithms lack a backtracking mechanism to identify and remove the wrong atoms included in their estimated support-set. In each iteration, a forward-backward algorithm can provide the inclusion of potential atoms and removal of outdated atoms in the estimated support-set. Subspace Pursuit (SP), proposed by Dai and Milenkovic [28], is a popular forward-backward greedy algorithm which works in a parallel way. For a $K$-sparse signal, in each iteration, SP identifies a $K$-dimensional subspace and tries to improve the estimated subspace in subsequent iterations. SP algorithm is given in Algorithm 2.3.

**Algorithm 2.3: Subspace Pursuit (SP) [28]**

**Inputs:** $A_{M \times N}$, $b_{M \times 1}$, and $K$.

**Initialization:** $k = 0$, $r_0 = b$, $\hat{T}_0 = \emptyset$;

1: repeat  
2: $k = k + 1$;  
3: $u = A^T r_{k-1}$;  
4: $J = \text{supp}(u^K)$;  
5: $\bar{T} = J \cup \hat{T}_{k-1}$;  
6: $v_{\bar{T}} = A_{\bar{T}}^\dagger b$, $v_{\bar{T}^c} = 0$;  
7: $\hat{T}_k = \text{supp}(v^K)$;  
8: $r_k = b - A_{\hat{T}_k} A_{\hat{T}_k}^\dagger b$;  
9: until $(\|r_k\|_2 \geq \|r_{k-1}\|_2)$;  
10: $\hat{T} = \hat{T}_{k-1}$;  
11: $\hat{x}_{\hat{T}} = A_{\hat{T}}^\dagger b$, $\hat{x}_{\hat{T}^c} = 0$;  

**Outputs:** $\hat{x}$ and $\hat{T}$.  

$\triangleright K \leq |\hat{T}| \leq 2K$  

$\triangleright v \in \mathbb{R}^{N \times 1}$  

$\triangleright \hat{x} \in \mathbb{R}^{N \times 1}$
Like MP and OMP, SP also relies on the matched filter to identify the potential atoms. SP iterates over the following three steps:

i) Expansion: In $k^{th}$ iteration, SP identifies the $K$ atoms of $A$ which gives maximum correlation with the regularized measurement vector (also known as residue), $r_{k-1}$, obtained during the $(k - 1)^{th}$ iteration. These atoms are added to the support-set estimated in the $(k - 1)^{th}$ iteration to get a larger set $\tilde{T}$ which has a maximum cardinality $2K$.

ii) Contraction: To retain only $K$ atoms in the end of the $k^{th}$ iteration, SP projects the observation $b$ onto the set $\tilde{T}$ to give a vector $v$. Now, the indices of the $K$ largest magnitudes of $v$ yield the estimate of the support-set, $\hat{T}_k$, in the current iteration.

iii) Regularization: Projecting $b$ onto the submatrix formed by the columns of $A$ listed in the estimated support-set $\hat{T}_k$ gives an estimate of the sparse signal. This sparse signal estimate is then used to regularize the measurement matrix to get the residual.

A block diagram representing the main steps in the $k^{th}$ iteration of the SP algorithm is shown in Figure 2.4. For $K \leq O(\sqrt{N})$, SP algorithm has a computational complexity $O(MN \log(K))$ for both Gaussian Sparse Signals (GSS) and Rademacher Sparse Signals (RSS) [28], which is significantly less as compared to $O(M^2N^{3/2})$, the computational complexity of $\ell_1$ algorithms based on the interior point methods [123] in the same asymptotic region.

Using RIP, SP is shown to have elegant theoretical guarantees for convergence under both signal and measurement perturbations [28, 124]. It has been shown that, for a clean measurement case ($b =$...
\[ J = \text{set of } K \text{ atoms with largest projection on residue } r_{k-1} \]

\[ \tilde{T} = J \cup \hat{T}_{k-1} \]

\[ v_{\tilde{T}} = A^\dagger b \]

\[ v_{\tilde{c}} = 0 \]

\[ \hat{T}_k = \text{supp}(v) \]

\[ \hat{x}_{\hat{T}_k} = A_{\hat{T}_k}^\dagger b; \hat{x}_{\hat{T}_c} = 0. \]

**Figure 2.4:** A schematic block diagram representing the \( k \)th iteration of SP algorithm.

Ax) where the measurement matrix \( A \) has RIC \( \delta_{3K} < 0.205 \), SP recovers any \( K \)-sparse vector in a finite number of iterations [28]. For a noisy measurement scenario, the performance bounds on reconstruction distortion have been derived, which hold true for \( \delta_{3K} < 0.083 \) [28].

### 2.3.6 Compressive Sampling Matching Pursuit (CoSaMP) [29]

Compressive Sampling Matching Pursuit (CoSaMP), proposed by Needell and Tropp [29], is one of the first greedy algorithms to show similar performance guarantees as shown by CRM. CoSaMP is very similar to SP as shown in Algorithm 2.4. The main difference is that CoSaMP selects \( 2K \) atoms from the matched filter whereas SP selects only \( K \) atoms. Another difference is that SP requires an additional LS step to compute the sparse signal estimate and the residual. CoSaMP produces a \( K \)-sparse approximation of the signal whose \( \ell_2 \) approximation error is comparable with the scaled \( \ell_1 \) approximation error of the best \( \frac{K}{2} \)-sparse approximation.
to the signal. The performance bounds on the reconstruction error provided by CoSaMP have been derived for $\delta_{4K} < 0.1$.

Algorithm 2.4: Compressive Sampling Matching Pursuit (CoSaMP) [29]

**Inputs:** $A_{M \times N}$, $b_{M \times 1}$, and $K$.

**Initialization:** $k = 0$, $r_0 = b$, $x_0 = 0$, $T_0 = \emptyset$;

1: repeat
2: $k = k + 1$;
3: $u = A^T r_{k-1}$;
4: $J = \text{supp}(u^{2K})$;
5: $T = J \cup T_{k-1}$;
6: $v_T = A_T^+ b$, $v_T = 0$;
7: $T_k = \text{supp}(v^K)$;
8: $\hat{x}_k = v_{T_k}$;
9: $r_k = b - A \hat{x}_k$;
10: until ($\|r_k\|_2 \geq \|r_{k-1}\|_2$);
11: $T = T_{k-1}$;
12: $x = \hat{x}_{k-1}$;

**Outputs:** $x$ and $T$.

Though CoSaMP is very similar to SP, Needell and Tropp [29] discuss many interesting properties of CoSaMP. This makes CoSaMP more appealing. For example, three gradient based methods are proposed in [29] to replace the computationally demanding LS step in CoSaMP. This modification has been also shown to retain a similar theoretical guarantee as provided by CoSaMP. Each iteration of CoSaMP requires only $O(L)$ time, where $L \geq N$ denotes the maximum cost of a multiplication with $A$ or $A^T$. A typical value of $L$ is given as $O(N \log N)$, which is satisfied by a partial Fourier matrix. The storage requirement of CoSaMP is $O(N)$.

We have briefly described CS and sparse signal reconstruction algorithms in this chapter. In the following chapters, we extensively use the theoretical results and algorithms discussed in this chapter.
“Make everything as simple as possible, but not simpler.”
Albert Einstein [1879-1955]

Though many Sparse Reconstruction Algorithms (SRAs) possess elegant theoretical guarantees for sparse signal recovery, it is important to note that the mathematical analysis is usually borne out in the asymptotic regimes which seldom occur in applications. In practice, different SRAs provide different performance, which is highly dependent on the measurement scenario. For example, it is well known that the performance of any sparse recovery algorithm depends on several factors like signal dimension, sparsity level of signal, and measurement noise power [8, 80–82]. It has been also reported that the reconstruction performance varies significantly depending on the underlying statistical distribution of the non-zero elements of the sparse signal [81, 82]. Hence, if this distribution is known a priori, we can get the best sparse signal estimate using the best recovery algorithm suitable for that type of signal. In practice,
we may not have this prior knowledge and thus, we cannot use the appropriate method.

The number of measurements also plays a substantial role in sparse signal reconstruction as it can influence the information content of the signal captured by the measurement process. It has been observed that every sparse recovery algorithm requires a minimum number of measurements (algorithm dependent) for sparse signal recovery and performs poorly in a very low dimension measurement scenario [81–84]. On the other hand, the reduction in number of measurements is highly preferred in many applications [11, 85] as it helps to reduce many key parameters like the number of sensors, measurement time, noise power, and hardware complexity.

Though it is evident that the performance of the sparse recovery algorithms degrades in the aforementioned cases, it is worthwhile to note (empirically) that this degradation does not always imply a complete failure [81]. That is, a SRA may yield an estimate with partially correct information about the sparse signal even in such disastrous situations. Naturally, we may expect that different SRAs, operating with different principles, will yield different information (correct) about the sparse signal. It may be possible to fuse these estimates to get a sparse signal estimate which is better than the best estimates of any of the SRAs used.

In data fusion [125], we use data from multiple sensors (often operate according to different principles) and fuse these data in order to extract information which will be more efficient than if they were achieved by means of any of the sensor alone. We explore this possibility here and show that fusion can indeed give a better sparse signal estimate as compared to the estimates of the
participating algorithms. This idea of fusion or mixing of multiple estimators to get a better estimate has been proposed recently in different contexts. For signal denoising, a random version of Orthogonal Matching Pursuit (OMP) was proposed by Elad and Yavneh [126] to get several signal representations and fusion was performed by plain averaging. Fusion of multiple estimators which use different dictionaries was discussed by Fadili et al. [127] and Starck et al. [128]. Recent Machine-learning and Statistics literature [129–131] also suggested fusion of a group of competing estimators using exponential weights, which often lead to an estimate better than the best in the group.

We propose a general fusion framework which employs multiple sparse recovery algorithms and fuse the resultant estimates to obtain a better sparse signal estimate. We also propose different algorithms to fuse the estimates of the participating algorithms which are described in Chapter 3, Chapter 4, and Chapter 5 of this thesis.

Next, we describe an exploratory experiment which shows the motivation of the proposed fusion framework and its significance in sparse recovery.

### 3.1 Exploratory Experiment

Let us consider the standard Compressed Sensing (CS) measurement setup given in (2.4) for Gaussian Sparse Signals (GSS) with signal dimension $N = 500$ and sparsity level $K = 20$. For GSS, the non-zero values are generated from independently and identically distributed (i.i.d.) $\mathcal{N}(0, 1)$. We assume a clean measurement setup (i.e., $w = 0$). Let us use two sparse recovery algorithms viz.
OMP [27] and Subspace Pursuit (SP) [28] for sparse signal recovery. Let $\mathcal{T}$ denote the actual support-set of $x$. Also, let $\hat{\mathcal{T}}^{(\text{OMP})}$ and $\hat{\mathcal{T}}^{(\text{SP})}$ denote the support-sets estimated by OMP and SP, respectively. Let $\hat{\mathcal{T}}_{\text{true}}^{(\text{OMP})} = \mathcal{T} \cap \hat{\mathcal{T}}^{(\text{OMP})}$ and $\hat{\mathcal{T}}_{\text{true}}^{(\text{SP})} = \mathcal{T} \cap \hat{\mathcal{T}}^{(\text{SP})}$ represent the sets of correct (true) atoms estimated by OMP and SP respectively. We have, $|\mathcal{T}| = |\hat{\mathcal{T}}^{(\text{OMP})}| = |\hat{\mathcal{T}}^{(\text{SP})}| = K$, $0 \leq |\hat{\mathcal{T}}_{\text{true}}^{(\text{OMP})}| \leq K$, and $0 \leq |\hat{\mathcal{T}}_{\text{true}}^{(\text{SP})}| \leq K$. We define the fraction of measurements denoted by $\alpha$ as

$$\alpha \triangleq \frac{M}{N}, \quad (3.1)$$

and simulated the experiment for small values of $\alpha$. In CS, the goal is to reduce the number of measurements and hence small values of $\alpha$ carry a special interest in CS [11, 85]. The results, computed by averaging over 10,000 trials, are summarized in Table 3.1. The details of the simulations are explained in Section 3.5.1.

<table>
<thead>
<tr>
<th>$\alpha = \frac{M}{N}$</th>
<th>0.10</th>
<th>0.11</th>
<th>0.12</th>
<th>0.13</th>
<th>0.14</th>
</tr>
</thead>
<tbody>
<tr>
<td>Avg$</td>
<td>\hat{\mathcal{T}}_{\text{true}}^{(\text{OMP})}</td>
<td>$</td>
<td>5.6</td>
<td>6.7</td>
<td>8.1</td>
</tr>
<tr>
<td>Avg$</td>
<td>\hat{\mathcal{T}}_{\text{true}}^{(\text{SP})}</td>
<td>$</td>
<td>5.8</td>
<td>7.9</td>
<td>10.5</td>
</tr>
<tr>
<td>Avg$</td>
<td>\hat{\mathcal{T}}<em>{\text{true}}^{(\text{OMP})} \cup \hat{\mathcal{T}}</em>{\text{true}}^{(\text{SP})}</td>
<td>$</td>
<td>7.9</td>
<td>9.9</td>
<td>12.4</td>
</tr>
</tbody>
</table>

**Table 3.1**: Average number of correctly estimated atoms by OMP and SP, for GSS, in clean measurement case, averaged over 10,000 trials ($N = 500$, $K = 20$).

For $\alpha = 0.13$ ($M = 65$) (refer Table 3.1), the average number of correctly estimated atoms by OMP is 10.1, and by SP is 13.2. Interestingly, the average number of true atoms in the set formed by the union of the support-sets estimated by OMP and SP is 15, closer to the true value of 20. A similar observation is also valid for other values of $\alpha$. Note that the union-set always contains at least as many true atoms as in the support-set estimated by the best performing algorithm (SP in this case).
Chapter 3  Fusion of Algorithms for Compressed Sensing

From these observations it is evident that the union-set is richer in true information about the support-set, as compared to both OMP and SP. If we can pick all true atoms from the union-set, then we can have a better reconstruction than both OMP and SP. In our proposed fusion framework, we explore this possibility and focus only on the union-set for estimating the support atoms.

The result in Table 3.1 confirms our intuition that the union-set is rich in necessary information. This leads to the possibility of estimating more true atoms from the union-set than the true atoms identified by OMP and SP algorithms individually. In the experiment, the union-set can have the highest cardinality as $2K = 40$. The worst case exhaustive search for the $K = 20$ true atoms from the union-set requires $\binom{2K}{K} = \binom{40}{20}$ searches. Though the dimension of search space has reduced significantly from $\binom{500}{20}$ to $\binom{40}{20}$, an exhaustive search is still not feasible. Hence, we need an efficient scheme to select $K$ atoms from $2K$ atoms. Note that OMP and SP are merely two examples and can be replaced by any pair of CS reconstruction algorithms. Also we can fuse information from more than two CS SRAs. In that case, the union-set can have a higher cardinality than $2K$ and the problem of finding $K$ indices from the union-set requires more searches. Therefore, in fusion, we must need to use a computationally simple, but efficient search strategy.

3.2 Proposed Fusion Framework

We develop the fusion framework in order to fuse the information from two or more arbitrary CS reconstruction algorithms. Let us assume that we use $P \geq 2$ different algorithms in parallel (independently) for signal reconstruction. We define an algorithmic function which symbolically represents the $i^{th}$ participating sparse
recovery algorithm for estimating the support-set of a $K$-sparse signal $x$ from (2.4), as

$$[\hat{x}_i, \hat{T}_i] = alg^{(i)}(A, b, K),$$

(3.2)

where $\hat{x}_i$ and $\hat{T}_i$ respectively denote the signal and the support-set, estimated by the $i^{th}$ participating algorithm $(i = 1, 2, \ldots, P)$ where $|\hat{T}_i| = |\text{supp}(\hat{x}_i)| = K$. Note that many sparse recovery algorithms (for example, $\ell_1$-minimization methods and Bayesian methods) may not directly find the support-set or the estimated signal may not be $K$-sparse. In such cases, we choose $\hat{x}_i$ as the best $K$-term approximation of the estimated signal and $\hat{T}_i$ as the set of indices corresponding to the $K$ largest magnitudes of the estimated sparse signal. We define the joint support-set as the union of the estimated support-sets, denoted by $\Gamma$, as

$$\Gamma = \bigcup_{i=1}^{P} \hat{T}_i.$$

We hope that $R \triangleq |\Gamma|$ is significantly lower than the signal dimension $N$, and most of the true atoms in $\mathcal{T}$ are included in $\Gamma$. Now, our goal is to identify all true atoms which are included in $\Gamma$. For this, we shall solve the following problem in the fusion framework

$$b = A_{\Gamma} x_{\Gamma} + \tilde{w},$$

(3.3)

where $A_{\Gamma} \in \mathbb{R}^{M \times R}$, $x_{\Gamma} \in \mathbb{R}^{R \times 1}$, and $\tilde{w} = w + A_{\Gamma^c} x_{\Gamma^c}$. Note that the problem dimension of (3.3) is significantly lower as compared to the problem dimension of (2.4). However, exhaustive search to solve (3.3) is still not feasible for applications, and hence we need to develop efficient schemes to identify the true atoms included in $\Gamma$. 

48
Chapter 3 Fusion of Algorithms for Compressed Sensing

A schematic diagram of the fusion framework is shown in Figure 3.1.

We propose different schemes for fusion and evaluate their performances below as well as in later chapters.

3.3 FACS Scheme

Next, we develop a novel fusion algorithm, which we refer to as Fusion of Algorithms for Compressed Sensing (FACS), to solve (3.3). In principle, we can use any sparse signal reconstruction algorithm to solve (3.3). For engineering simplicity as well as analytical tractability, we use a Least-Squares (LS) approach for the solution of (3.3). The LS approach is to use pseudo-inverse and the required assumption is that $R \leq M$. Using the LS approach, the FACS scheme is shown in Algorithm 3.1.

Let us discuss a few advantages and disadvantages of the Fusion of Algorithms for Compressed Sensing (FACS). The FACS is
Chapter 3  Fusion of Algorithms for Compressed Sensing

Algorithm 3.1: Fusion of Algorithms for CS (FACS)

Inputs: $A \in \mathbb{R}^{M \times N}$, $b \in \mathbb{R}^{M \times 1}$, $K$, and $\left\{ \hat{T}_i \right\}_{i=1:P}$.

Assumption: $| \bigcup_{i=1}^{P} \hat{T}_i | \leq M$.

Initialization: $v = 0 \in \mathbb{R}^{N \times 1}$.

Fusion:
1. $\Gamma = \bigcup_{i=1}^{P} \hat{T}_i$;
2. $v_\Gamma = A_\Gamma^\dagger b$, $v_\Gamma^c = 0$;
3. $\hat{T} = \text{supp}(v^K)$; $\triangleright v^K$ is the best $K$-term approximation of $v$

Outputs: $\hat{T}$ and $\hat{x}$ (where $\hat{x}_\hat{T} = A_\hat{T}^\dagger b$ and $\hat{x}_{\hat{T}^c} = 0$).

powerful due to its scalability. Any CS reconstruction algorithm can be easily incorporated into an existing FACS scheme due to the strategy of parallel (or independent) execution of participating algorithms. A weak aspect of FACS is that it is blind to the information not captured in the joint support-set. Also FACS requires high computational complexity due to the execution of several participating algorithms. We explain theoretical studies of FACS in the next section.

3.4 Theoretical Studies of FACS

Using Restricted Isometry Property (RIP), we analyse the FACS scheme shown in Algorithm 3.1 and derive a sufficient condition to get an improved performance over any participating algorithm. The performance analysis is characterized by a measure called Signal-to-Reconstruction-Error Ratio (SRER) [132, 133] defined as

$$SRER \triangleq \frac{\|x\|_2^2}{\|x - \hat{x}\|_2^2}, \quad (3.4)$$
where \(x\) and \(\hat{x}\) denote the actual and reconstructed signal vector respectively.

We start analysis with the poor condition that \(\mathcal{T} \not\subseteq \Gamma\), that means \(\|x_{\mathcal{T}^c}\|_2 \neq 0\). In this poor condition, Theorem 3.1 provides a sufficient condition for improved reconstruction performance of FACS over each participating algorithm.

**Theorem 3.1.** For the CS setup (2.4), let the sparse signal \(x\) have the support-set \(\mathcal{T}\) such that \(|\mathcal{T}| = K\). In FACS scheme (Algorithm 3.1), we use \(P \geq 2\) independent algorithms in parallel. Let the \(i^{th}\) participating algorithm provides the reconstructed signal \(\hat{x}_i\) and the associated support-set \(\hat{T}_i\) where \(|\hat{T}_i| = K\). In FACS scheme, we use the joint support-set \(\Gamma = \bigcup_{i=1}^{P} \hat{T}_i\) where \(|\Gamma| = R \leq M\). Using \(\Gamma\) and LS estimation, the FACS scheme provides the reconstructed signal \(\hat{x}\) and the associated support-set \(\hat{T}\) where \(|\hat{T}| = K\). Assume that \(\|x_{\mathcal{T}^c}\|_2 \neq 0\), \(\|x_{\mathcal{T}^c}\|_2 \neq 0\), and the CS measurement matrix \(A\) holds RIP with the Restricted Isometry Constant (RIC) \(\delta_{R+K}\). By defining \(\eta_i = \frac{\|x_{\mathcal{T}^c}\|_2}{\|x_{\mathcal{T}^c}\|_2}\) and \(\zeta = \frac{\|w\|_2}{\|x_{\mathcal{T}^c}\|_2}\), we have the following results.

i) \(\forall i \in \{1, 2, \ldots, P\}, 0 < \eta_i \leq 1\).

ii) FACS provides at least SRER gain of \(\left(\frac{(1-\delta_{R+K})^2}{(1+\delta_{R+K}+3\zeta)\eta_i}\right)^2\) over the \(i^{th}\) participating algorithm if \(\eta_i < \frac{(1-\delta_{R+K})^2}{1+\delta_{R+K}+3\zeta}\).

iii) Let \(r = b - A_{\hat{T}}A_{\hat{T}}^\dagger b\) and \(r_i = b - A_{\hat{T}_i}A_{\hat{T}_i}^\dagger b\) denote the residues of FACS and the \(i^{th}\) algorithm, respectively. Then \(\|r\|_2 < \|r_i\|_2\) if \(\eta_i < \frac{(1 - \delta_{R+K})(1 - 2\delta_{R+K})}{(1 + \delta_{R+K})^2 + 4\zeta}\).

**Proof:**

1) We have, \(\|x_{\mathcal{T}^c}\|_2 > 0\) and \(\|x_{\mathcal{T}^c}\|_2 > 0\) (using the property of norm).
Hence \( \eta_i = \frac{\|x_{\Gamma^c}\|_2}{\|x_{\hat{T}^c_i}\|_2} > 0 \).

The claim \( \eta_i \leq 1 \) follows from the relation \( \|x_{\Gamma^c}\|_2 \leq \|x_{\hat{T}^c_i}\|_2 \) \( \because \Gamma^c \subset \hat{T}^c_i, i = 1, 2, \ldots, P \).

2) To show the improvement in SRER, we first consider

\[
\|x - \hat{x}\|_2 \leq \|x_{\hat{T}} - \hat{x}_{\hat{T}}\|_2 + \|x_{\hat{T}^c}\|_2 \quad (\because \hat{x}_{\hat{T}^c} = 0) \tag{3.5}
\]

Then we have,

\[
\|x_{\hat{T}} - \hat{x}_{\hat{T}}\|_2 = \|x_{\hat{T}} - A_{\hat{T}}^T b\|_2
\]

\[
= \|x_{\hat{T}} - A_{\hat{T}}^T (Ax + w)\|_2 \quad (\because b = Ax + w)
\]

\[
= \|x_{\hat{T}} - A_{\hat{T}}^T (A_{\hat{T}}x_{\hat{T}} + A_{\hat{T}^c}x_{\hat{T}^c} + w)\|_2
\]

\[
= \|A_{\hat{T}}^T A_{\hat{T}^c} x_{\hat{T}^c} + A_{\hat{T}}^T w\|_2 \quad (\because A_{\hat{T}}^T A_{\hat{T}} = I)
\]

\[
\leq \left\| (A_{\hat{T}}^H A_{\hat{T}})^{-1} A_{\hat{T}}^H A_{\hat{T}^c} x_{\hat{T}^c} \right\|_2 + \left\| A_{\hat{T}}^T w \right\|_2
\]

\[
\leq \frac{1}{1 - \delta_{R+K}} \|A_{\hat{T}}^H A_{\hat{T}^c} x_{\hat{T}^c}\|_2 + \frac{1}{\sqrt{1 - \delta_{R+K}}} \|w\|_2 \tag{3.6}
\]

\( (\because R + K \geq 2K \geq |\hat{T} \cup \hat{T}| \text{ and using (2.9))} \)

(a) follows from (2.6) & (2.7) and using \( |\hat{T}| = K \leq R + K \), \( A \) has RIC \( \delta_{R+K} \).

Substituting (3.6) in (3.5), we get

\[
\|x - \hat{x}\|_2 \leq \left( 1 + \frac{\delta_{R+K}}{1 - \delta_{R+K}} \right) \|x_{\hat{T}^c}\|_2 + \frac{1}{\sqrt{1 - \delta_{R+K}}} \|w\|_2
\]

\[
= \frac{1}{1 - \delta_{R+K}} \|x_{\hat{T}^c}\|_2 + \frac{1}{\sqrt{1 - \delta_{R+K}}} \|w\|_2 \tag{3.7}
\]

Next, we will find an upper bound for \( \|x_{\hat{T}^c}\|_2 \). Let us define
\( \hat{T}_\Delta \triangleq \Gamma \setminus \hat{T} \). That is, \( \hat{T}_\Delta \) is the set formed by the atoms in \( \Gamma \) which are discarded by Algorithm 3.1. Now, since \( \hat{T} \subset \Gamma \), we have \( \hat{T}^c = \Gamma^c \cup \hat{T}_\Delta \) and hence

\[
\| x_{\hat{T}} \|_2 \leq \| x_{\Gamma^c} \|_2 + \| x_{\hat{T}_\Delta} \|_2. \tag{3.8}
\]

Let us consider

\[
\left\| (v_{\Gamma})_{\hat{T}_\Delta} \right\|_2 = \left\| x_{\hat{T}_\Delta} + (v_{\Gamma} - x_{\Gamma})_{\hat{T}_\Delta} \right\|_2 \\
\geq \left\| x_{\hat{T}_\Delta} \right\|_2 - \left\| (v_{\Gamma} - x_{\Gamma})_{\hat{T}_\Delta} \right\|_2
\]

(using reverse triangle inequality)

\[
\Rightarrow \left\| x_{\hat{T}_\Delta} \right\|_2 \leq \left\| (v_{\Gamma})_{\hat{T}_\Delta} \right\|_2 + \left\| (v_{\Gamma} - x_{\Gamma})_{\hat{T}_\Delta} \right\|_2 \\
\leq \left\| (v_{\Gamma})_{\hat{T}_\Delta} \right\|_2 + \left\| v_{\Gamma} - x_{\Gamma} \right\|_2. \tag{3.9}
\]

Note that \( (v_{\Gamma})_{\hat{T}} \) contains the \( K \)-elements of \( v_{\Gamma} \) with highest magnitudes. Hence, using \( |\hat{T}| = |T| = K \), we can write \( \|(v_{\Gamma})_{\hat{T}}\|_2^2 \geq \|(v_{\Gamma})_{\hat{T}}\|_2^2 \) and hence \( \|(v_{\Gamma})_{\hat{T}}\|_2^2 - \|(v_{\Gamma})_{\hat{T}}\|_2^2 \leq 0 \). Now we have

\[
\left\| (v_{\Gamma})_{\hat{T}_\Delta} \right\|_2^2 = \left\| (v_{\Gamma})_{\hat{T}_\Delta} \right\|_2^2 + \left\| (v_{\Gamma})_{\hat{T}} \right\|_2^2 - \left\| (v_{\Gamma})_{\hat{T}} \right\|_2^2 \\
= \left\| (v_{\Gamma}) \right\|_2^2 - \left\| (v_{\Gamma})_{\hat{T}} \right\|_2^2 \\
= \left\| (v_{\Gamma})_{\Gamma \setminus T} \right\|_2^2 + \left\| (v_{\Gamma})_{T} \right\|_2^2 - \left\| (v_{\Gamma})_{\hat{T}} \right\|_2^2 \\
\leq \left\| (v_{\Gamma})_{\Gamma \setminus T} \right\|_2^2,
\]

and hence we have

\[
\left\| (v_{\Gamma})_{\hat{T}_\Delta} \right\|_2 \leq \left\| (v_{\Gamma})_{\Gamma \setminus T} \right\|_2 \\
= \left\| (v_{\Gamma})_{\Gamma \setminus T} - x_{\Gamma \setminus T} \right\|_2 \quad (\because x_{\Gamma \setminus T} = 0) \\
= \left\| (v_{\Gamma} - x_{\Gamma})_{\Gamma \setminus T} \right\|_2 \\
\leq \left\| (v_{\Gamma} - x_{\Gamma}) \right\|_2. \tag{3.10}
\]
Chapter 3  fusion of Algorithms for Compressed Sensing

Substituting (3.10) in (3.9), we get

\[ \|x_{\hat{\gamma}}\|_2 \leq 2 \| (v_{\Gamma} - x_{\Gamma}) \|_2. \]  (3.11)

Then we have

\[ \| (v_{\Gamma} - x_{\Gamma}) \|_2 = \| \hat{A}^\dagger b - x_{\Gamma} \|_2 \]
\[ = \| \hat{A}^\dagger (A_{\Gamma} x_{\Gamma} + A_{\Gamma} x_{\Gamma}^e + w) - x_{\Gamma} \|_2 \quad (\because b = Ax + w) \]
\[ \leq \| \hat{A}^\dagger A_{\Gamma} x_{\Gamma}^e \|_2 + \| \hat{A}^\dagger w \|_2 \quad (\because \hat{A}^\dagger A_{\Gamma} = I) \]
\[ = \| (A_{\Gamma}^H A_{\Gamma})^{-1} A_{\Gamma}^H A_{\Gamma} x_{\Gamma}^e \|_2 + \| \hat{A}^\dagger w \|_2 \]

(using definition of \( \hat{A}^\dagger \))
\[ \leq \frac{1}{1 - \delta_R} \| A_{\Gamma}^H A_{\Gamma} x_{\Gamma}^e \|_2 + \frac{1}{\sqrt{1 - \delta_R}} \| w \|_2 \]

(using (2.6)) & (2.7)
\[ \leq \frac{\delta_{R+K}}{1 - \delta_R} \| x_{\Gamma}^e \|_2 + \frac{1}{\sqrt{1 - \delta_R}} \| w \|_2 \quad (\text{using (2.9)}) \]
\[ \leq \frac{\delta_{R+K}}{1 - \delta_{R+K}} \| x_{\Gamma}^e \|_2 + \frac{1}{1 - \delta_{R+K}} \| w \|_2 \]

(\because \delta_R \leq \delta_{R+K}, 0 \leq 1 - \delta_R \leq 1)

(3.12)

Now, using (3.11), and (3.12) in (3.8), we get

\[ \| x_{\hat{\gamma}} \|_2 \leq \left( 1 + \frac{2\delta_{R+K}}{1 - \delta_{R+K}} \right) \| x_{\Gamma}^e \|_2 + \frac{2}{\sqrt{1 - \delta_{R+K}}} \| w \|_2 \]
\[ = \frac{1 + \delta_{R+K}}{1 - \delta_{R+K}} \| x_{\Gamma}^e \|_2 + \frac{2 \| w \|_2}{1 - \delta_{R+K}}. \]  (3.13)

Using (3.13) in (3.7), we get

\[ \| x - \hat{x} \|_2 \leq \frac{1 + \delta_{R+K}}{(1 - \delta_{R+K})^2} \| x_{\Gamma}^e \|_2 + \left[ \frac{1}{\sqrt{1 - \delta_{R+K}}} + \frac{2}{(1 - \delta_{R+K})^2} \right] \| w \|_2 \]
\[ \leq \frac{1 + \delta_{R+K}}{(1 - \delta_{R+K})^2} \| x_{\Gamma}^e \|_2 + \frac{3 \| w \|_2}{(1 - \delta_{R+K})^2} \]  (3.14)
\[ (0 \leq 1 - \delta_{R+K} \leq 1) \]
\[ = \frac{1 + \delta_{R+K} + 3\zeta}{(1 - \delta_{R+K})^2} \eta_i \left\| x_{\hat{c}}^i \right\|_2 \]
\[ \text{using } \zeta = \frac{\|w\|_2}{\|x_{\hat{c}}^i\|_2} \text{ and } \eta_i = \frac{\|x_{\Gamma c}\|_2}{\left\| x_{\hat{c}}^i \right\|_2} \]
\[ = \frac{1 + \delta_{R+K} + 3\zeta}{(1 - \delta_{R+K})^2} \eta_i \left\| x - \hat{x}_i \right\|_2 \quad (\because (\hat{x}_i)_{\hat{c}} = 0) \]
\[ \leq \frac{1 + \delta_{R+K} + 3\zeta}{(1 - \delta_{R+K})^2} \eta_i \left\| x - \hat{x}_i \right\|_2. \quad (3.15) \]

Finally, we derive SRER for FACS as,
\[ \text{SRER}_{\text{FACS}} = \frac{\|x\|_2^2}{\|x - \hat{x}\|_2^2} \]
\[ \geq \frac{\|x\|_2}{\|x - \hat{x}_i\|_2} \times \left( \frac{(1 - \delta_{R+K})^2}{(1 + \delta_{R+K} + 3\zeta)\eta_i} \right)^2 \]
\[ = \text{SRER}_{i^{th} \text{ algorithm}} \times \left( \frac{(1 - \delta_{R+K})^2}{(1 + \delta_{R+K} + 3\zeta)\eta_i} \right)^2. \]

Hence FACS provides at least SRER gain of \( \left( \frac{(1 - \delta_{R+K})^2}{(1 + \delta_{R+K} + 3\zeta)\eta_i} \right)^2 \) over \( i^{th} \) algorithm if \( \eta_i < \frac{(1 - \delta_{R+K})^2}{1 + \delta_{R+K} + 3\zeta} \). Note that \( \frac{(1 - \delta_{R+K})^2}{1 + \delta_{R+K} + 3\zeta} < 1. \)

3) We have,
\[ \|r\|_2 = \left\| b - A_{\hat{T}}A_{\hat{T}}^\dagger b \right\|_2 \]
\[ = \left\| (Ax + w) - A_{\hat{T}}A_{\hat{T}}^\dagger (Ax + w) \right\|_2 \]
\[ = \left\| (A_{\hat{T}c}x_{\hat{T}c} - A_{\hat{T}}A_{\hat{T}}^\dagger x_{\hat{T}c}) + (w - A_{\hat{T}}A_{\hat{T}}^\dagger w) \right\|_2 \]
\[ \leq \left\| A_{\hat{T}c}x_{\hat{T}c} - A_{\hat{T}}A_{\hat{T}}^\dagger x_{\hat{T}c} \right\|_2 + \| w - A_{\hat{T}}A_{\hat{T}}^\dagger w \|_2 \]
\[ \leq \left\| A_{\hat{T}c}x_{\hat{T}c} \right\|_2 + \| w \|_2 \quad \text{(using (2.10))} \]
\[ = \left\| A_{(\hat{T} \setminus \hat{T})}x_{(\hat{T} \setminus \hat{T})} \right\|_2 + \| w \|_2 \quad (\because \hat{T} = \text{supp}(x)) \]
From (3.16) and (3.17) we can see that

\[
\| r_i \|_2 = \left\| b - A_{\tilde{T}_i} A_{\tilde{T}_i}^\dagger b \right\|_2
\]

= \left\| (A x + w) - A_{\tilde{T}_i} A_{\tilde{T}_i}^\dagger (A x + w) \right\|_2
\]

\[
\geq \left\| A_{\tilde{T}_i} x_{\tilde{T}_i} - A_{\tilde{T}_i} A_{\tilde{T}_i}^\dagger A_{\tilde{T}_i} x_{\tilde{T}_i} \right\|_2 - \left\| w - A_{\tilde{T}_i} A_{\tilde{T}_i}^\dagger w \right\|_2
\]

(using triangular inequality)

\[
= \left\| A_{(T \setminus \tilde{T}_i)} x_{(T \setminus \tilde{T}_i)} - A_{\tilde{T}_i} A_{\tilde{T}_i}^\dagger A_{(T \setminus \tilde{T}_i)} x_{(T \setminus \tilde{T}_i)} \right\|_2
\]

- \left\| w - A_{\tilde{T}_i} A_{\tilde{T}_i}^\dagger w \right\|_2
\]

\[
\geq \left( 1 - \frac{\delta_{2K}}{1 - \delta_K} \right) \left\| A_{(T \setminus \tilde{T}_i)} x_{(T \setminus \tilde{T}_i)} \right\|_2 - \left\| w \right\|_2
\]

(using (2.10), \(|T \setminus \tilde{T}_i| \leq K\))

\[
\geq \left( 1 - \frac{\delta_{2K}}{1 - \delta_K} \right) \sqrt{1 - \delta_K} \left\| x_{(T \setminus \tilde{T}_i)} \right\|_2 - \left\| w \right\|_2
\]

(using \(|T \setminus \tilde{T}_i| \leq K\), and (2.5))

\[
\geq \left( 1 - \frac{\delta_{2K}}{1 - \delta_K} \right)^2 \left\| x_{(T \setminus \tilde{T}_i)} \right\|_2 - \left\| w \right\|_2 \hspace{1cm} (\because \sqrt{1 - \delta_K} < 1)
\]

\[
= (1 - \delta_K - \delta_{2K}) \left\| x_{\tilde{T}_i} \right\|_2 - \left\| w \right\|_2
\]

\[
\geq (1 - 2\delta_{R+K}) \frac{1}{\eta_i} \left\| x_{\Gamma^c} \right\|_2 - \zeta \left\| x_{\Gamma^c} \right\|_2 \hspace{1cm} (\because 1 - 2\delta_{R+K} < 1 - \delta_K - \delta_{2K})
\]

\[
= \left( 1 - 2\delta_{R+K} \right) \frac{1}{\eta_i} - \zeta \right\| x_{\Gamma^c} \right\|_2.
\]

From (3.16) and (3.17) we can see that \( \| r_i \|_2 < \| r \|_2 \), if
\[
\left[ \frac{(1+\delta R+K)^2}{1-\delta R+K} + \zeta^{3+\delta R+K} \right] < \left( (1-2\delta R+K) \eta_i - \zeta \right)
\]
or \( \eta_i < \frac{(1-\delta R+K)(1-2\delta R+K)}{(1+\delta R+K)^2 + 4\zeta} \).

Note that Theorem 3.1 on page 51 considers the poor case when \( \|x^\hat{c}_T\|_2 \neq 0 \) and \( \|x^\Gamma_c\|_2 \neq 0 \). When \( \|x^\hat{c}_T\|_2 = 0 \), the support-set is already correctly estimated by \( i \)th algorithm and further performance improvement is not possible in the FACS. Note that \( \|x^\hat{c}_T\|_2 = 0 \) implies \( \|x^\Gamma_c\|_2 = 0 \). Hence we consider only the general case \( \|x^\Gamma_c\|_2 = 0 \) in Proposition 3.1. Note that \( \|x^\hat{c}_T\|_2 \neq 0 \) and \( \|x^\Gamma_c\|_2 = 0 \) together imply \( T \subset \Gamma \) and \( T \neq \Gamma \).

**Proposition 3.1.** Assume that all the conditions in Theorem 3.1 on page 51 hold except \( \|x^\Gamma_c\|_2 \neq 0 \). The assumption is that \( \|x^\Gamma_c\|_2 = 0 \). Then, in the clean measurement case (\( w = 0 \)), FACS estimates the support-set correctly, in turn providing exact reconstruction.

**Proof:** From Algorithm 3.1 on page 50 (step 2), we get

\[
v_\Gamma = A_\Gamma^\dagger b = A_\Gamma^\dagger (A_\Gamma x_\Gamma + w) \\
= x_\Gamma + A_\Gamma^\dagger w = x_\Gamma + e_\Gamma.
\] (3.18)

where \( e \in \mathbb{R}^{N \times 1} \) with \( e_\Gamma = A_\Gamma^\dagger w \) and \( e_{\Gamma^c} = 0 \). \( e \) denote the “error” due to the projection of measurement noise \( w \) onto the space spanned by the columns of \( A \) whose indices are listed in \( \Gamma \). Note that if \( w = 0 \), then \( e = 0 \) and \( v = x \) (\( \therefore T \subset \Gamma \)). Therefore FACS estimates the support-set correctly from \( v \).

It may be noted that FACS is not guaranteed to give performance improvement in all the cases. In fact, in principle, a performance degradation is also possible with FACS. It should be noted that a similar performance degradation is also possible in the celebrated data fusion framework. However, our extensive numerical experiments (refer Section 3.5) show that FACS outperforms
even the best participating algorithms in most of the cases. From (3.18), we can see that FACS will pick up all the correct atoms iff
\[
\min_{i \in T} |x_i + e_i| > \max_{i \in \Gamma \setminus T} |e_i|
\]
where \(x_i\) and \(e_i\) denote the \(i^{th}\) element of \(x\) and \(e\) respectively.

### 3.4.1 Extension to Arbitrary Signals

We analyse the performance of FACS for arbitrary signals in this section. The signals are either not at all sparse signals or sparse signals with the sparsity level more than \(K\). These kind of signal models are motivated from practical scenarios: for example, many signals are compressible in nature [80].

**Theorem 3.2.** (Performance for arbitrary signals): For the standard CS measurement setup (2.4), let the signal \(x \in \mathbb{R}^{N \times 1}\) be an arbitrary signal. In FACS scheme (Algorithm 3.1), we use \(P \geq 2\) independent algorithms in parallel. Let the \(i^{th}\) participating algorithm provide the reconstructed signal \(\hat{x}_i\) and the associated support-set \(\hat{T}_i\) where \(|\hat{T}_i| = K\). In FACS scheme, we use the joint support-set \(\Gamma = \bigcup_{i=1}^{P} \hat{T}_i\) where \(|\Gamma| = R \leq M\). Using \(\Gamma\) and LS estimation, FACS scheme provides the reconstructed \(K\)-sparse signal \(\hat{x}\) and the associated support-set \(\hat{T}\) where \(|\hat{T}| = K\). The assumption is that the CS measurement matrix \(A\) holds RIP with the RIC \(\delta_{R+K}\).

i) Upper bound of reconstruction error: We have,
\[
\|x - \hat{x}\|_2 \leq c_1 \|x - x^K\|_2 + c_2 \|x - x^K\|_1 + c_3 \|x_{\Gamma^c}\|_2 + \nu \|w\|_2,
\]
where \(\nu = \frac{3 - \delta_{R+K}^2}{(1 - \delta_{R+K})^2}\), \(c_1 = \left(1 + \frac{\sqrt{1 + \frac{\nu}{R + K}}}{\nu}\right)\),
\(c_2 = \frac{\nu \sqrt{1 + \delta_{R+K}}}{\sqrt{R + K}}\), and \(c_3 = \frac{1 + \delta_{R+K}}{(1 - \delta_{R+K})^2}\).
ii) SRER gain: Assuming \( \|x_{\hat{T}_i}\|_2 \neq 0 \), and \( \|x_{\Gamma_c}\|_2 \neq 0 \), define \( \eta_i = \frac{\|x_{\Gamma_c}\|_2}{\|x_{\hat{T}_i}\|_2} \), \( \zeta = \frac{\|w\|_2}{\|x_{\Gamma_c}\|_2} \) and \( \xi = \frac{1 + 3\sqrt{1 + \delta_{R+K}}}{3\|x_{\Gamma_c}\|} \|x - x^K\|_2 \) + \( \sqrt{\frac{1 + \delta_{R+K}}{R + K}} \frac{\|x - x^K\|_1}{\|x_{\Gamma_c}\|} \). FACS provides at least an SRER gain of \( \left( \frac{(1 - \delta_{R+K})^2}{(1 + \delta_{R+K} + 3\xi + 3\zeta)\eta_i} \right) \left( \frac{(1 - \delta_{R+K})^2}{1 + \delta_{R+K} + 3\xi + 3\zeta} \right) \) over \( i^{th} \) participating algorithm if \( \eta_i < \frac{(1 - \delta_{R+K})^2}{1 + \delta_{R+K} + 3\xi + 3\zeta} \).

Proof: The proof is given in Appendix 3.A. ■

For \( \|x_{\Gamma_c}\|_2 = 0 \) (i.e., if \( x \) is a sparse signal and the associated support-set \( \subset \Gamma \)), we can comment on the tightness of the upper bound of the reconstruction error shown in Theorem 3.2 on the facing page. Note that, for \( \|x_{\Gamma_c}\|_2 = 0 \), if the signal \( x \) is \( K \)-sparse (i.e., \( \|x - x^K\|_2 = \|x - x^K\|_1 = 0 \)) and the measurement noise \( w = 0 \), then we get the exact reconstruction.

### 3.5 Numerical Experiments and Results

We conducted simulations to evaluate the proposed FACS using OMP, SP, and Basis Pursuit (BP)/ Basis Pursuit De-Noising (BPDN) as the participant algorithms. These three algorithms work with different principles. We compared FACS vis-a-vis the participating algorithms. In the following sections, simulation setups and results are discussed.
3.5.1 Synthetic Sparse Signals

To evaluate the performance of FACS in numerical experiments, we use a measure called Average Signal-to-Reconstruction-Error Ratio (ASRER) which is defined as

\[
\text{ASRER} = \left( \frac{\sum_{j=1}^{n_{\text{trials}}} \|x_j\|_2^2}{\sum_{j=1}^{n_{\text{trials}}} \|x_j - \hat{x}_j\|_2^2} \right),
\]

where \(x_j\) and \(\hat{x}_j\) respectively denote the actual and reconstructed sparse signal in the \(j^{\text{th}}\) trial, and \(n_{\text{trials}}\) denotes the total number of trials. Note that ASRER is not defined as the average of the SRER defined in (3.4). This is intentionally chosen as ASRER defined in this fashion will be very large, biasing the average of the SRER, whenever the estimation error is close to zero.

In the presence of a measurement noise in (2.4), it is impossible to achieve perfect CS reconstruction. On the other hand, for the clean measurement case, perfect CS reconstruction is possible if the fraction of measurements, \(\alpha\) (defined in (3.1)), exceeds a certain threshold. In the spirit of using CS for practical applications with a low number of measurements at clean and noisy conditions, we are mainly interested in a lower range of \(\alpha\) where performances of the contesting methods can be fairly compared. Using \(\alpha\), the main steps of the simulation are as follows:

i) Fix \(K\), \(N\) and choose an \(\alpha\) so that the number of measurements \(M\) is an integer.

ii) Generate elements of \(A_{M \times N}\) independently from \(\mathcal{N}(0, \frac{1}{M})\) and normalize each column norm to unity.
iii) Choose $K$ locations uniformly over the set $\{1, 2, \ldots, N\}$ and fill these non-zero values of $x$ based on the choice of signal characteristics:

(a) *Gaussian Sparse Signals (GSS)*: non-zero values are independently chosen from $\mathcal{N}(0, 1)$.

(b) *Rademacher Sparse Signals (RSS)*: non-zero values are set to $+1$ or $-1$ with probability $\frac{1}{2}$. They are also known as *constant amplitude random sign* signals.

Set remaining $N - K$ locations of $x$ as zeros.

iv) For noisy case, the additive noise $w$ is a Gaussian random vector whose elements are independently chosen from $\mathcal{N}(0, \sigma_w^2)$, and for a clean measurement case, $w$ is set to zero.

v) Calculate the measurement vector $b = Ax + w$.

vi) Apply the reconstruction methods independently.

vii) Repeat steps iii-v $T$ times. $T$ indicates the number of times $x$ is independently generated, for a realization of $A$.

viii) Repeat steps ii-vi $S$ times. $S$ indicates the number of times $A$ is independently generated.

ix) Calculate ASRER using (3.19).

x) Repeat steps ii-viii for a new $\alpha$.

Considering the measurement noise $w \sim \mathcal{N}(0, \sigma_w^2 I_M)$, for noisy measurement simulations, we define the Signal-to-Measurement-Noise Ratio (SMNR) as

$$\text{SMNR} = \frac{\mathbb{E}\{\|x\|_2^2\}}{\mathbb{E}\{\|w\|_2^2\}},$$

(3.20)
where \( \mathbb{E}\{\|w\|_2^2\} = \sigma_w^2 M \).

We conducted the experiments with \( N = 500, K = 20, S = 100, \) and \( T = 100. \) That means, we used sparse signals with dimension 500 and sparsity level 20. This 4\% level of sparsity is intentionally chosen as it closely resembles many real application scenarios. For example, it is empirically observed that most of the energy of any natural image in the wavelet domain is concentrated within 2\% – 4\% of the coefficients [80]. We used 100 realizations of \( A \) (i.e., \( S = 100 \)) and for each realization of \( A \), we randomly generated 100 sparse signal vectors (i.e., \( T = 100 \)).

**Experiment 1**

In the first experiment, we show the robustness of the FACS scheme for signals with arbitrary statistics. We use OMP and SP as the participant algorithms in FACS and the corresponding FACS scheme is denoted as FACS(OMP,SP). For GSS, in both clean and noisy measurement (SMNR = 20 dB) cases, the SRER results are shown in Figure 3.2(a) and Figure 3.2(b), respectively. It can be observed that for all values of \( \alpha \), FACS(OMP,SP) performed better than both OMP and SP. In the clean measurement case, at \( \alpha = 0.18 \), FACS(OMP,SP) gave 10 dB improvement over SP and 6.5 dB improvement over OMP. In the noisy measurement case, at \( \alpha = 0.18 \), FACS(OMP,SP) showed 5 dB improvement over SP and 2.5 dB improvement over OMP. Next, for Rademacher sparse signals in both clean and noisy measurement (SMNR = 20 dB) cases, the SRER results are shown in Figure 3.3(a) and Figure 3.3(b). We observe that FACS(OMP,SP) performed better than the participating algorithms.
Experiment 2

Now, we verify the scalability of FACS using another CS reconstruction algorithm in the existing FACS(OMP,SP). The new algorithm is either BP or BPDN, according to either clean or noisy measurement case. Henceforth, for BP/ BPDN, we use the notation BP. The software code of BP was taken from the $\ell_1$-magic toolbox [110].
Figure 3.4: Fusion of three participating algorithms: Performance of FACS(OMP, SP, BP) for the Gaussian sparse signal ($N = 500, K = 20$).

BP does not provide an estimate of the support-set, but provides an estimate of sparse signal directly. Also, BP is not guaranteed to provide a $K$-sparse signal estimate. Hence, to use BP as the participant algorithm in FACS, we use the approach by Chatterjee et al. [83] where the support-set was formed by the indices corresponding to the $K$ highest amplitude entries in the estimated sparse signal of BP. The new $K$-sparse signal estimate of BP is found by orthogonally projecting $b$ onto the range space of the columns of $A$ indexed by the estimated support-set. When the sparsity level $K$ is known a priori, such an approach has been shown to improve the performance of BP [83, 134, 135]. The inclusion of BP leads to the new FACS which is denoted as FACS(OMP, SP, BP). For GSS, in both clean and noisy measurement ($\text{SNMR} = 20\text{dB}$) cases, the SRER results are shown in Figure 3.4(a) and Figure 3.4(b). It may be also noted from Figure 3.4 that, for smaller values of $\alpha$ ($0.1 \leq \alpha \leq 0.12$), FACS(OMP, SP, BP) gave a less ASRER than the participating algorithm BP. This an example which shows that fusion always need not result in a better ASRER. However, for all other values of $\alpha$, FACS(OMP, SP, BP) resulted in a better ASRER compared to BP.
More importantly, it is also observed that the FACS(OMP,SP,BP) performs better than the FACS(OMP,SP). Similar performance improvement was also noticed for the RSS, but we do not report the results for brevity.

3.5.1.1 Reproducible Research

In the spirit of reproducible research [136, 137], we provide necessary Matlab codes publicly downloadable at http://www.ece.iisc.ernet.in/~ssplab/Public/FACS.tar.gz. The code reproduces the simulation results shown in Figure 3.2, Figure 3.3, and Figure 3.4.

3.5.2 Real Compressible Signals

Most of the signals we often meet in applications are not exactly sparse. However, many of them including natural signals are found to be compressible which can be well approximated by their sparse versions. In this section, we evaluate the efficacy of FACS for real-world compressible signals.

Experiment 3

We conducted experiments on real-world ECG signals selected from MIT-BIH Arrhythmia Database [138]. ECG signals are compressible and have a good structure for sparse decompositions. We used a similar simulation setup used by Carrillo et al. [139, 140]. As earlier, here also we used OMP, SP, and BP as the participating algorithms. Similar to the synthetic sparse signal simulation setup, we used Gaussian measurement matrices with appropriate sizes to
vary the number of measurements, $M$, from 256 to 480 with an increment of 32. We assumed a sparsity level 128 and the reconstruction results are shown in Figure 3.5.

![Graph showing performance of different algorithms](image)

**Figure 3.5:** Real Compressible signals: Performance of FACS (Signal-to-Reconstruction-Error Ratio (SRER) vs. Number of Measurements) for ECG signals selected from MIT-BIH Arrhythmia Database [138, 141].

As in the case of synthetic signals, for real-world ECG signals also FACS(OMP,SP) resulted in a better SRER as compared to both the participating algorithms OMP and SP. For example, at $M = 288$, FACS(OMP,SP) gave 5.8 dB and 6.8 dB SRER improvement over OMP and SP respectively. By using BP as the third participating algorithm, FACS(OMP,SP,BP) further improved the SRER by 1.8 dB than FACS(OMP,SP) for $M = 288$. Also Note that, except for $M = 256$, FACS(OMP,SP,BP) resulted in a better ASRER than BP. A similar trend can be observed for other values of $M$ in Figure 3.5, showing the advantage of using FACS in real-life applications.

### 3.5.3 Highly Coherent Dictionary

In many applications like RADAR and SONAR [52,142], and Direction-of-Arrival (DOA) estimation [143], the measurement matrix (also
known as dictionary matrix) is often highly coherent. Sparse signal recovery has found wide applications in such cases. Next, we evaluate the performance of the proposed FACS in such a situation where the matrix is highly coherent.

**Experiment 4**

A comparison result of twelve typical sparse signal recovery algorithms with a highly coherent dictionary matrix, which is a simplified real-world lead-field matrix in EEG source localization, was reported by Zhang [144]. The maximum coherence of the columns of the dictionary matrix was $0.9983$. T-MSBL [46] and FBMP [145] were reported as the best and second best algorithms in terms of failure rate [146] and Mean-Square Error (MSE) [144]. We repeated the same experiment and used FACS with T-MSBL and FBMP as the participating algorithms. The results averaged over 1,000 trials are shown in Table 3.2. It can be seen that, in terms of both failure rate and MSE, FACS resulted in a better performance than the best reported algorithm, T-MSBL.

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Average failure rate</th>
<th>Average MSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>FBMP [145]</td>
<td>0.2200</td>
<td>0.2528</td>
</tr>
<tr>
<td>T-MSBL [46]</td>
<td>0.0980</td>
<td>0.0991</td>
</tr>
<tr>
<td>FACS(T-MSBL,FBMP)</td>
<td><strong>0.0630</strong></td>
<td><strong>0.0920</strong></td>
</tr>
</tbody>
</table>

**Table 3.2**: Performance of FACS, averaged over 1,000 trials, on a highly coherent dictionary matrix: a simplified real-world matrix in EEG source localization [144].

**3.6 Summary**

It is shown that a judicious fusion of outputs obtained from several algorithms leads to a better compressed sensing reconstruction
performance. The fusion in algorithmic level results in seamless scalability and robustness. While it is possible to engineer several sophisticated fusion strategies, we use a simple LS based approach. The approach is not only theoretically tractable, but provides significant performance improvement in practice. Naturally we can expect that the use of a sophisticated fusion strategy will provide further performance improvement.

### 3.6.1 Relevant Publications


### 3.A Proof of Theorem 3.2 on page 58 (Extension to Arbitrary Signals)

i) We have

\[
\|x - \hat{x}\|_2 = \| (x - x^K) + (x^K - \hat{x}) \|_2 \\
\leq \| x - x^K \|_2 + \| x^K - \hat{x} \|_2 \tag{3.21}
\]
Now, consider

\[ b = Ax + w = A x^K + A(x - x^K) + w \]  \hspace{1cm} (3.22) \]

Note that (3.22) can be viewed as a standard CS measurement system with \( x^K \) as the \( K \)-sparse signal and \( A(x - x^K) + w \) as the measurement perturbations. Hence, using Lemma 2.1 on page 29, we get

\[
\| x^K - \hat{x} \|_2 \leq \frac{1}{1 - \delta_{2K}} \| (x^K)_{\hat{T}} \|_2 + \frac{1 + \delta_{2K}}{1 - \delta_{2K}} \| A(x - x^K) + w \|_2 \\
\leq \frac{1}{1 - \delta_{2K}} \| (x^K)_{\hat{T}} \|_2 + \frac{1 + \delta_{2K}}{1 - \delta_{2K}} \| A(x - x^K) \|_2 \\
+ \frac{1 + \delta_{2K}}{1 - \delta_{2K}} \| w \|_2 \\
\leq \frac{1}{1 - \delta_{R+K}} \| (x^K)_{\hat{T}} \|_2 + \frac{1 + \delta_{R+K}}{1 - \delta_{R+K}} \| A(x - x^K) \|_2 \\
+ \frac{1 + \delta_{R+K}}{1 - \delta_{R+K}} \| w \|_2. 
\]  \hspace{1cm} (3.23) \]

We have, using (3.13)

\[
\| (x^K)_{\hat{T}} \|_2 \leq \frac{1 + \delta_{R+K}}{1 - \delta_{R+K}} \| (x^K)_{\hat{T}_c} \|_2 + \frac{2}{1 - \delta_{R+K}} \| A(x - x^K) + w \|_2 \\
\leq \frac{1 + \delta_{R+K}}{1 - \delta_{R+K}} \| x_{\hat{T}_c} \|_2 + \frac{2}{1 - \delta_{R+K}} \left( \| A(x - x^K) \|_2 + \| w \|_2 \right). 
\]  \hspace{1cm} (3.24) \]

\[ (: : \| (x^K)_{\hat{T}_c} \|_2 \leq \| x_{\hat{T}_c} \|_2) \hspace{1cm} (3.25) \]

Using Lemma 2.2 on page 29, we obtain,

\[
\| A(x - x^K) \|_2 \leq \sqrt{1 + \delta_{R+K}} \left( \| x - x^K \|_2 + \frac{1}{\sqrt{R+K}} \| x - x^K \|_1 \right) 
\]  \hspace{1cm} (3.26) \]
Substituting (3.24) and (3.26) in (3.23), we get

\[
\|x^K - \hat{x}\|_2 \leq \frac{(3 - \delta^2_{R+K})\sqrt{1 + \delta_{R+K}}}{(1 - \delta_{R+K})^2} \|x - x^K\|_2 \\
+ \frac{(3 - \delta^2_{R+K})\sqrt{1 + \delta_{R+K}}}{(1 - \delta_{R+K})^2\sqrt{R + K}} \|x - x^K\|_1 \\
+ \frac{1 + \delta_{R+K}}{(1 - \delta_{R+K})^2} \|x_{\Gamma^c}\|_2 + \frac{3 - \delta^2_{R+K}}{(1 - \delta_{R+K})^2} \|w\|_2 \\
= \nu \sqrt{1 + \delta_{R+K}} \|x - x^K\|_2 + \frac{\nu \sqrt{1 + \delta_{R+K}}}{\sqrt{R + K}} \|x - x^K\|_1 \\
+ \frac{1 + \delta_{R+K}}{(1 - \delta_{R+K})^2} \|x_{\Gamma^c}\|_2 + \nu \|w\|_2
\]  

(3.27)

where \(\nu = \frac{3 - \delta^2_{R+K}}{(1 - \delta_{R+K})^2}\).

Substituting (3.27) in (3.21), we get

\[
\|x - \hat{x}\|_2 \leq \left(1 + \nu \sqrt{1 + \delta_{R+K}}\right) \|x - x^K\|_2 + \frac{\nu \sqrt{1 + \delta_{R+K}}}{\sqrt{R + K}} \|x - x^K\|_1 \\
+ \frac{1 + \delta_{R+K}}{(1 - \delta_{R+K})^2} \|x_{\Gamma^c}\|_2 + \nu \|w\|_2 \\
= c_1 \|x - x^K\|_2 + c_2 \|x - x^K\|_1 + c_3 \|x_{\Gamma^c}\|_2 + \nu \|w\|_2.
\]

(3.28)

(\text{using definition of } c_1, c_2 \text{ and } c_3)

\[\text{ii) Using the definition of } \xi, \zeta, \text{ and } \nu \text{ in (3.28), we obtain}\]

\[
\|x - \hat{x}\|_2 \leq \frac{1 + \delta_{R+K} + 3\xi + 3\zeta}{(1 - \delta_{R+K})^2} \|x_{\Gamma^c}\|_2 \\
= \frac{1 + \delta_{R+K} + 3\xi + 3\zeta}{(1 - \delta_{R+K})^2} \eta_i \|x_{\Gamma^c}\|_2 \quad \text{(using definition of } \eta_i) \\
= \frac{1 + \delta_{R+K} + 3\xi + 3\zeta}{(1 - \delta_{R+K})^2} \eta_i \|(x - \hat{x}_i)_{\hat{\Gamma}^c}\|_2 \quad \text{(} \because (\hat{x}_i)_{\hat{\Gamma}^c} = 0 \text{)}
\]
Now, we have SRER for FACS in case of arbitrary signals

\[
\text{SRER}_{\text{FACS}} = \frac{\|x\|^2}{\|x - \hat{x}\|^2} \geq \frac{\|x\|^2}{\|x - \hat{x}_i\|^2} \times \left( \frac{(1 - \delta_{R+K})^2}{(1 + \delta_{R+K} + 3\xi + 3\zeta)\eta_i} \right)^2
\]

\[
= \text{SRER}_{i^{th \text{ algorithm}}} \times \left( \frac{(1 - \delta_{R+K})^2}{(1 + \delta_{R+K} + 3\xi + 3\zeta)\eta_i} \right)^2.
\]

Hence FACS provides at least an SRER gain of

\[
\left( \frac{(1 - \delta_{R+K})^2}{1 + \delta_{R+K} + 3\xi + 3\zeta} \right)^2
\]

over \(i^{th}\) participating algorithm if

\[
\eta_i < \frac{(1 - \delta_{R+K})^2}{1 + \delta_{R+K} + 3\xi + 3\zeta}.
\]

Note that \(1 + \delta_{R+K} + 3\xi + 3\zeta < 1\). ■
“Alone we can do so little; together we can do so much.”
Helen Keller [1980-1968]

Though Fusion of Algorithms for Compressed Sensing (FACS) improves the sparse signal recovery, as compared to the participating algorithms, it is completely blind about the true atoms which are not included in the joint support-set, $\Gamma$. To address this problem, we propose novel fusion schemes in this chapter. To fuse estimates of two participating algorithms we propose an algorithm which we referred to as Committee Machine Approach for Compressed Sensing (CoMACS). We also propose two variations of CoMACS to further improve sparse signal recovery.
4.1 CoMACS: Algorithm

To develop CoMACS, let us assume that we use two participating algorithms independently for signal reconstruction from the Compressed Sensing (CS) measurement setup (2.4). We follow the notations used in Chapter 3. Let \( \text{alg}^{(i)} \) provide the reconstructed signal \( \hat{x}_i \) and the associated support-set be \( \hat{T}_i \), where \( |\hat{T}_i| = K \) \((i = 1, 2)\). We have, \( \Gamma = \hat{T}_1 \cup \hat{T}_2 \) and \( R = |\Gamma| \). Let us denote the intersection of the estimated support-sets, called common support-set, as \( \Lambda = \hat{T}_1 \cap \hat{T}_2 \) and let \( S \triangleq |\Lambda| \). We have, \( 0 \leq S \leq K \leq R \leq 2K \).

As discussed in Section 3.1, we decided to estimate the support atoms only from the joint support-set \( \Gamma \), which reduces (2.4) to a comparatively lower dimensional problem (3.3).

Note that the two participating algorithms used here play the role of ‘experts’ in a committee machine approach [147]. In a committee, a natural strategy is to accept the part where both the ‘experts’ agree. We follow this simple rule in our work and include the intersection set, \( \Lambda \), in our estimated support-set. Notice that the intersection set, \( \Lambda \), has at least the ‘higher accuracy’ as of the intersection of subsets of both the sets with same cardinality. The ratio of the number of true atoms included in \( \Lambda \) and the number of atoms in \( \Lambda \) is a measure of the efficacy of the strategy. The results given in Table 4.1 are taken from the exploratory experiment discussed in Section 3.1, which also justifies this strategy. For example, for \( \alpha = 0.13 \), on an average, 92% elements of \( \hat{T}(OMP) \cap \hat{T}(SP) \) are true atoms. Hence, we have \( \Lambda \subset \hat{T} \), by choice where \( \hat{T} \) denote the support-set estimated by our proposed method.

The decision to include \( \Lambda \) in \( \hat{T} \) further reduces the dimension of the problem from \( \binom{R}{K} \) to \( \binom{R-S}{K-S} \). That is, we need to estimate only \( K - S \) atoms from (3.3) to complete the estimated support-set.
using the same number of measurements $M$ which gives a better sparsity-measurement trade-off as compared to the problems given in (2.4) and (3.3). To devise a simple strategy, let us assume that $R \leq M$, which is reasonable in many practical situations. Using this assumption, we have an overdetermined system in (3.3). We propose a least-squares based method, which is simple yet powerful, to estimate $K - S$ atoms from (3.3). The algorithm is summarized in Algorithm 4.1. We refer to this algorithm as Committee Machine Approach for Compressed Sensing (CoMACS). The term Committee Machine is borrowed from Neural Network literature [147]. In a Committee Machine, a complex estimation task is solved by a number of experts. The combination of these experts constitutes a Committee Machine, which fuses the estimates obtained by the experts to get an estimate which is likely to be superior to that attained by any one of the experts, acting alone. It may be noted that the nomenclature ‘Committee Machines’ we used here is by and large superficial and our proposed method has no direct connection with it. A variation of CoMACS was first proposed in [148].

When $\Lambda = \emptyset$, CoMACS estimates all the $K$ atoms using Least-Squares (LS) (step 4 in Algorithm 4.1). If $S = K$, i.e., when both the algorithms agree on all $K$ atoms, CoMACS finds $\Lambda$ as the estimated support-set.
## Algorithm 4.1: Committee Machine Approach for Compressed Sensing (CoMACS)

**Inputs:** $A_{M \times N}$, $b_{M \times 1}$, $K$, $\tilde{T}_1$, and $\tilde{T}_2$.

**Ensure:** $|\tilde{T}_1 \cup \tilde{T}_2| \leq M$

**Initialization:** $v = 0 \in \mathbb{R}^N$, $\hat{x} = 0 \in \mathbb{R}^N$;

1. $\Lambda = \tilde{T}_1 \cap \tilde{T}_2$; \hspace{1cm} ▶ $0 \leq |\Lambda| \leq K$
2. $\Gamma = \tilde{T}_1 \cup \tilde{T}_2$; \hspace{1cm} ▶ $K \leq |\Gamma| \leq 2K$
3. $v_\Gamma = A_\Gamma^\dagger b$, $v_{\Gamma^c} = 0$; \hspace{1cm} ▶ $v \in \mathbb{R}^{N \times 1}$
4. $\tilde{T} = \text{indices corresponding to the } (K - |\Lambda|) \text{ largest magnitude entries in } v \text{ which are not in } \Lambda$;
5. $\hat{T} = \tilde{T} \cup \Lambda$;
6. $\hat{x}_\hat{T} = v_\hat{T}$, $\hat{x}_{\hat{T}^c} = 0$; \hspace{1cm} ▶ $\hat{x} \in \mathbb{R}^{N \times 1}$

**Outputs:** $\hat{x}$ and $\hat{T}$.

Note that, in principle, we can apply any CS reconstruction algorithm (for example, $\ell_1$-minimization methods) to identify $K$ atoms from $\Gamma$. We explore this option empirically in Section 4.3.1 on page 91 and compare the performance with the proposed methods.

### 4.1.1 Theoretical Analysis for CoMACS

In this section, we theoretically analyse CoMACS (Algorithm 4.1) using Restricted Isometry Property (RIP). We will consider the analysis for two cases: (a) $x$ is exactly sparse, and (b) $x$ is not exactly sparse. The second case shows the robust nature of CoMACS. The performance analysis is characterized by Signal-to-Reconstruction-Error Ratio (SRER).

#### 4.1.1.1 Sparse Signals

Here, we consider signals which are exactly sparse and derive sufficient conditions for performance improvement of CoMACS, in
terms of SRER. The results are summarized in Theorem 4.1. For this, we also use the results from Proposition 4.1 on page 79.

**Theorem 4.1.** For the CoMACS framework discussed in Section 4.1, assume that \( \|x_{\tilde{T}_i}\|_2 \neq 0 \), \( \|x_{\Gamma}\|_2 \neq 0 \), and the CS measurement matrix \( A \) holds RIP with the Restricted Isometry Constant (RIC) \( \delta_{R+K} \).

By defining \( \eta_i = \frac{\|x_{\Gamma}\|_2}{\|x_{\tilde{T}_i}\|_2} \), \( \zeta = \frac{\|w\|_2}{\|x_{\Gamma}\|_2} \), and \( \upsilon = \frac{\|x_{\Lambda}\|_2}{\|x_{\Gamma}\|_2} \), we have the following results.

i) \( 0 < \eta_i \leq 1 \), \( \forall i = 1, 2 \).

ii) CoMACS provides a minimum SRER gain of

\[
\left( \frac{1 - \delta_{R+K}}{(3 + 3\zeta + 2\upsilon(1 - \delta_{R+K}))\eta_i} \right)^2 \text{ over } \text{alg}^{(i)}
\]

if \( \eta_i < \frac{1 - \delta_{R+K}}{3 + 3\zeta + 2\upsilon(1 - \delta_{R+K})} \).

**Proof.**

i) Using properties of norm, we have \( \|x_{\Gamma}\|_2 > 0 \) and \( \|x_{\tilde{T}_i}\|_2 > 0 \). Hence we have, \( \eta_i = \frac{\|x_{\Gamma}\|_2}{\|x_{\tilde{T}_i}\|_2} > 0 \).

The claim \( \eta_i \leq 1 \) follows from the following relation \( \|x_{\Gamma}\|_2 \leq \|x_{\tilde{T}_i}\|_2 \) (\( \therefore \Gamma \subset \tilde{T}_i, i = 1, 2 \)).

ii) We have,

\[
\|x - \hat{x}\|_2 \leq \|x - v\|_2 + \|v - \hat{x}\|_2. \tag{4.1}
\]

From Algorithm 4.1, we have,

\[
\hat{x} = v^{\tilde{T}} = v^{\Lambda} + (v - v^{\Lambda})^{K - |\Lambda|}, \tag{4.2}
\]

where \( v \) and \( \tilde{T} \) are defined in Step 3 and Step 5 of Algorithm 4.1 respectively. Using this, we get,

\[
\|v - \hat{x}\|_2 = \left\| v - v^{\Lambda} - (v - v^{\Lambda})^{K - |\Lambda|} \right\|_2
\]
\[
\begin{align*}
\leq & \|v - v^\Lambda\|_2 \\
\leq & \|v - x\|_2 + \|x^\Lambda - v\|_2 \\
\leq & 2\|v - x\|_2 \\
= & 2\|v - x + x^\Lambda\|_2 \\
\leq & 2\|v - x\|_2 + 2\|x^\Lambda\|_2 \\
= & 2\|x - v\|_2 + 2\|x^\Lambda\|_2. 
\end{align*}
\] 

(4.3)

The first inequality follows from the fact that, for any vector \(s \in \mathbb{R}^{N \times 1}\) and any positive integer \(N_1 \leq N\), \(\|s - s^{N_1}\|_2 \leq \|s\|_2\).

Substituting (4.3) in (4.1), we get

\[
\|x - \hat{x}\|_2 \leq 3\|x - v\|_2 + 2\|x^\Lambda\|_2
\]

\[
\leq \frac{3(\|x_{i^c}\|_2 + \|w\|_2)}{1 - \delta_{R+K}} + 2\|x_{i^c}\|_2 \tag{4.4}
\]

\[
\leq \frac{3 + 3\zeta + 2\upsilon(1 - \delta_{R+K})}{1 - \delta_{R+K}}\eta_i\|x_{\hat{t}_i^c}\|_2
\]

\[
\leq \frac{3 + 3\zeta + 2\upsilon(1 - \delta_{R+K})}{1 - \delta_{R+K}}\eta_i\|(x - \hat{x}_i)_{\hat{t}_i^c}\|_2
\]

\[
\leq \frac{3 + 3\zeta + 2\upsilon(1 - \delta_{R+K})}{1 - \delta_{R+K}}\eta_i\|x - \hat{x}_i\|_2. \tag{4.5}
\]

(a) follows by using Proposition 4.1 and definition of \(\upsilon\), (b) follows by using the definition of \(\zeta\) and \(\eta_i\), and finally (c) follows from the fact \((\hat{x}_i)_{\hat{t}_i^c} = 0\).

Now using (4.5), we have SRER for CoMACS given by

\[
\text{SRER}_{\text{CoMACS}} = \frac{\|x\|_2^2}{\|x - \hat{x}\|_2^2}
\]

\[
\geq \frac{\|x\|_2^2}{\|x - \hat{x}_i\|_2^2} \times \left( \frac{1 - \delta_{R+K}}{3 + 3\zeta + 2\upsilon(1 - \delta_{R+K})\eta_i} \right)^2
\]

\[
= \text{SRER}_{\text{alg}}(i) \times \left( \frac{1 - \delta_{R+K}}{3 + 3\zeta + 2\upsilon(1 - \delta_{R+K})\eta_i} \right)^2.
\]
Chapter 4  A Committee Machine Approach

**Proposition 4.1.** Assume that the conditions in Theorem 4.1 holds. Let \( \Gamma \subset \{1, 2, \ldots, N\} \) with \( R = |\Gamma| \leq M \) and \( v_\Gamma = A_\Gamma^\dagger b \) with \( v_{\Gamma^c} = 0 \). Then we have,

\[
\|x - v\|_2 \leq \frac{1}{1 - \delta_{R+K}}\|x_{\Gamma^c}\|_2 + \frac{1}{1 - \delta_{R+K}}\|w\|_2.
\]


Theorem 4.1 provides sufficient conditions for CoMACS to improve the sparse recovery performance. It may be noted that the theoretical conditions are, in general, ‘pessimistic’ worst case conditions. As we will see in Section 4.3, CoMACS performs better than Orthogonal Matching Pursuit (OMP) and Subspace Pursuit (SP) algorithms especially in lower measurement cases.

**4.1.1.2 Extension to Arbitrary Signals**

In Theorem 4.1, we assumed that the signal under consideration is \( K \)-sparse. This condition is not always met in practice and when \( x \) is not strictly sparse, the exact signal reconstruction is not possible. We consider the case of arbitrary signals and analyse the properties of the estimates obtained using CoMACS. We present an upper bound on reconstruction error and derive sufficient conditions for SRER gain in Theorem 4.2.

**Theorem 4.2.** (Performance for arbitrary signals) Consider the measurement setup (2.4) for an arbitrary signal \( x \in \mathbb{R}^{N \times 1} \) with the CoMACS framework described in Section 4.1. Assuming that the measurement matrix \( A \) satisfies RIP with RIC \( \delta_{R+K} \), we have the following results:
Chapter 4  A Committee Machine Approach

i) Upper bound on reconstruction error: \( \| x - \hat{x} \|_2 \leq c_1 \| x - x^K \|_2 + c_2 \| x - x^K \|_1 + c_3 \| x_\Gamma^c \|_2 + c_4 \| w \|_2 \), where \( c_1 = \frac{3 \sqrt{1 + \delta_{R+K}}}{1 - \delta_{R+K}} \), \( c_2 = \frac{3 \sqrt{1 + \delta_{R+K}}}{(1 - \delta_{R+K}) \sqrt{R + K}} \), \( c_3 = \frac{3 + 2 \nu(1 - \delta_{R+K})}{1 - \delta_{R+K}} \), and \( c_4 = \frac{3}{1 - \delta_{R+K}} \).

ii) SRER gain: Assuming \( \| x_{\Gamma^c}^T \|_2 \neq 0 \), and \( \| x_\Gamma^c \|_2 \neq 0 \), define

\[
\eta_i = \frac{\| x_\Gamma^c \|_2}{\| x_{\Gamma^c}^T \|_2}, \quad \zeta = \frac{\| w \|_2}{\| x_\Gamma^c \|_2}, \quad \nu = \frac{\| x_{\Lambda^c} \|_2}{\| x_\Gamma^c \|_2}
\]

and

\[
\xi = \frac{2 \sqrt{1 + \delta_{R+K}}}{\| x_\Gamma^c \|_2} \left( \| x - x^K \|_2 + \frac{1}{\sqrt{R + K}} \| x - x^K \|_1 \right),
\]

CoMACS will result in a minimum SRER gain of

\[
\frac{(1 - \delta_{R+K})^2}{(3 \xi + 3 + 2 \nu(1 - \delta_{R+K}) + 3 \zeta)^2 \eta_i^2}
\]

as compared to \( \text{alg}^{(i)} \) if \( \eta_i < \frac{1}{1 - \delta_{R+K}} \).

Proof. The proof is presented in Appendix 4.B on page 98. ■

From Theorem 4.2 on the previous page we can see that, for a \( K \)-sparse signal, if \( \| x_{\Gamma^c} \|_2 = 0 \) we get a perfect signal reconstruction using CoMACS in a clean measurement case. This shows the tightness of CoMACS for a \( K \)-sparse signal.

It may be observed that the computational and memory requirements of CoMACS are mainly due to the participating algorithms. The computational complexity of CoMACS is slightly more than the added computational complexity of the individual participating algorithms. The main additional computation required by CoMACS is in finding the LS solution (Step 3 in Algorithm 4.1 on page 76).
4.1.2 CoMACS for Multiple Participating Algorithms

Now, we consider extension of CoMACS for more than two participating algorithms. Let there be \( P \) participating algorithms employed to estimate the sparse signal \( x \) from (2.4). Let \( \hat{x}_i \) and \( \hat{T}_i \) respectively denote the sparse signal and support-set estimated by the \( i \)th participating algorithm, \( \text{alg}^{(i)} \) \((i = 1, 2, \ldots, P)\). We propose a stage-wise fusion strategy for fusing the participating algorithms using CoMACS which we referred to as Stage-wise CoMACS (StCoMACS).

In the first stage of StCoMACS, we fuse the estimates of \( \text{alg}^{(1)} \) and \( \text{alg}^{(2)} \) using CoMACS. The resultant \emph{fused} estimate is then fused with the estimate of \( \text{alg}^{(3)} \) in the second stage. The procedure continues till the fusion of all \( P \) participating algorithms completes in \( P - 1 \) stages. Let \( \text{StCoMACS}(j) \) denote the \( j \)th stage of StCoMACS, and let \( \tilde{x}_j \) and \( \tilde{T}_j \) denote the sparse signal and support-set estimated by \( \text{StCoMACS}(j) \) \((j = 1, 2, \ldots, P - 1)\). In \( \text{StCoMACS}(j) \) the estimates of \( \text{StCoMACS}(j - 1) \) and \( \text{alg}^{(j+1)} \) are fused using CoMACS. The StCoMACS algorithm for \( P \) participating algorithms is given in Algorithm 4.2. The algorithmic function \( \text{CoMACS}(A, b, K, \hat{T}_1, \hat{T}_2) \) used in Algorithm 4.2 calls Algorithm 4.1 on page 76 with respective inputs.

**Algorithm 4.2 :** Stage-wise CoMACS (StCoMACS)

**Inputs:** \( A_{M \times N}, b_{M \times 1}, K, \) and \( \{\hat{T}_i\}_{i=1:P}. \)

1. \( \tilde{T}_0 = \hat{T}_1; \)
2. for \( j = 1: P - 1 \) do
3. \[ [\tilde{x}_j, \tilde{T}_j] = \text{CoMACS}(A, b, K, \hat{T}_{j-1}, \hat{T}_{j+1}); \]
4. end for

**Outputs:** \( \hat{x} = \tilde{x}_{P-1} \) and \( \hat{T} = \tilde{T}_{P-1}. \)

Next, we extend Theorem 4.2 for StCoMACS.
**Proposition 4.2.** Assume that $P \geq 2$ participating algorithms are employed independently to reconstruct an arbitrary signal $x$ from (2.4). Let $\hat{x}_i$ and $\hat{T}_i$ respectively denote the sparse signal and support-set estimated by the $i^{th}$ participating algorithm, $\text{alg}^{(i)}$ ($i = 1, 2, \ldots, P$). Let $\text{StCoMACS}(j)$ denote StCoMACS algorithm with the first $j + 1$ algorithms as participating algorithms. Let $\tilde{x}_0 = \hat{x}_1$ and $\tilde{T}_0 = \hat{T}_0$, and let $\tilde{x}_j$ and $\tilde{T}_j$ respectively denote the estimate of sparse signal and support-set obtained by $\text{StCoMACS}(j)$, $(j = 1, 2, \ldots, P - 1)$. Let the measurement matrix $A$ have RIC $\delta_{R+K}$. Let $\Gamma_j = \tilde{T}_{j-1} \cup \tilde{T}_{j+1}$, $\Lambda_j = \tilde{T}_{j-1} \cap \tilde{T}_{j+1}$, $\eta_j = \frac{\|x_{\hat{T}_j^c}\|_2}{\|x_{\hat{T}_j^c}\|_2}$, $\zeta_j = \frac{\|w\|}{\|x_{\hat{T}_j^c}\|_2}$, $\xi_j = \frac{2\sqrt{1 + \delta_{R+K}}}{\|x_{\hat{T}_j^c}\|_2} \left(\|x - x^K\|_2 + \frac{1}{\sqrt{R + K}}\|x - x^K\|_1\right)$, and $\upsilon_j = \frac{\|x_{\hat{\Lambda}_j^c}\|_2}{\|x_{\hat{\Lambda}_j^c}\|_2}$. Then, $\text{StCoMACS}(j)$ provides at least SRER gain of
\[
\frac{(3\xi_j + 3 + 2\upsilon_j(1 - \delta_{R+K}) + 3\zeta_j)^2\eta_j^2}{(1 - \delta_{R+K})^2}
\]
as compared to $\text{alg}^{(j+1)}$ if $\eta_j < \frac{1 - \delta_{R+K}}{3\xi_j + 3 + 2\upsilon_j(1 - \delta_{R+K}) + 3\zeta_j}$.

**Proof:** To prove this, we use the fact that in $j^{th}$ stage, StCoMACS uses CoMACS to fuse the estimated support-sets $\tilde{T}_{j-1}$ and $\tilde{T}_{j+1}$. Since $|\tilde{T}_{j-1}| = |\tilde{T}_{j+1}| = K$, we have $|\Gamma_j| = |\tilde{T}_{j-1} \cup \tilde{T}_{j+1}| \leq 2K$. Now, the result follows from Theorem 4.2 by setting $P = 2$, $\hat{x}_1 = \hat{x}_{j-1}$, $\hat{x}_2 = \hat{x}_{j+1}$, $\hat{T}_1 = \tilde{T}_{j-1}$, and $\hat{T}_2 = \tilde{T}_{j+1}$. 

For $P$ participating algorithms, StCoMACS can be employed in $\binom{P}{2}$ different ways. Empirically we found that the order in which we fuse the participating algorithms is not important (the average sparse recovery performance in all cases gave similar results).
Limitations of CoMACS

Though CoMACS can effectively fuse the estimates of multiple participating algorithms, it has mainly two limitations.

- It can be observed that the performance of CoMACS crucially depends on the ‘quality’ of the joint support-set $\Gamma$. CoMACS is totally blind about the true atoms which are not included in $\Gamma$ and hence it cannot identify those atoms.

- The common support-set $\Lambda$ may contain some wrong atoms. Inclusion of $\Lambda$ in final estimated support-set leads to inclusion of those wrong atoms. Hence CoMACS is also blind to wrong atoms collected in $\Lambda$.

To alleviate these limitations and improve the performance we develop a new scheme in the next section.

4.2 Iterative CoMACS

Based on the approach of partial support recovery [33, 35, 40, 139, 140, 149], the new scheme uses CoMACS iteratively. We refer the scheme as Iterative CoMACS (ICoMACS).

- *Include potential atoms outside $\Gamma$: It has been shown that the sparsity-measurement trade-off of existing sparse recovery algorithms can be improved by incorporating partial knowledge about the support set [40, 139, 140, 149]. Even in the absence of such partial knowledge, extending the partial support recovery principles, iterative schemes are shown to improve the sparsity-measurement trade-off [33, 35]. In such
schemes, the information extracted from the sparse signal estimate of the previous iteration is used in the current iteration to improve the sparse signal estimate.

We use a similar strategy in ICoMACS to identify the true atoms not included in $\Gamma$. We start with CoMACS in the first iteration. In the subsequent iteration, the common support-set estimated in the previous iteration is used as a partially known support-set. In ICoMACS also, we continue to rely on the atoms in the common support-set (i.e, we agree with the common decision of the ‘committee’) and include all the atoms in common support-set in the estimated support-set. Now, we need to identify only a reduced dimensional subspace. For this, the participating algorithms are run again using this partially known support-set and the results are fused using CoMACS.

- **Discard outdated atoms in $\Lambda$:** The decision to include $\Lambda$ in the estimated support-set of CoMACS evolved from a natural engineering intuition. This is an ad-hoc scheme. In the worst case, $\Lambda$ may not contain any true atoms. Considering this worst case scenario, we need to incorporate a sanity check for the atoms in $\Lambda$ to discard the outdated atoms. We use a LS based scheme to remove the outdated atoms from $\Lambda$.

These two procedures continue as long as the $\ell_2$-norm of the residue decreases. A few of the other popular halting criteria includes: (i) stop algorithm after a fixed number of iterations, and (ii) stop when $\ell_2$-norm of the residue is less than a pre-fixed threshold. We summarize ICoMACS in Algorithm 4.3.

In the $(k+1)^{th}$ iteration of ICoMACS we use the common support-set estimated in $k^{th}$ iteration, $\Lambda_k$, as a partially known support-set. Let $S_k \triangleq |\Lambda_k|$, $(0 \leq S_k \leq K)$. In a clean measurement case ($w = 0$),
Algorithm 4.3: Iterative CoMACS (ICoMACS)

Inputs: $A_{M \times N}$, $b_{M \times 1}$, and $K$.

Initialization: $k = S_0 = 0$, $A_0 = A$, $r_0 = e_0 = b$, $\Lambda_0 = \emptyset$;

1: repeat
   2: $k = k + 1$;
   3: $\hat{T}^{(1)} = \text{alg}^{(1)}(A_{k-1}, e_{k-1}, K)$; $\hat{T}^{(1)} = K$
   4: $\Upsilon_1 = \{\text{indices of atoms of } A \text{ listed in } \hat{T}^{(1)}\} \cup \Lambda_{k-1}$;
   5: $\check{u}_{\Upsilon_1} = A_{\Upsilon_1}^\dagger b$, $\check{u}_{\Upsilon_1} = 0$;
   6: $\hat{T}^{(2)} = \text{alg}^{(2)}(A_{k-1}, e_{k-1}, K)$; $\hat{T}^{(2)} = K$
   7: $\Upsilon_2 = \{\text{indices of atoms of } A \text{ listed in } \hat{T}^{(2)}\} \cup \Lambda_{k-1}$;
   8: $\check{v}_{\Upsilon_2} = A_{\Upsilon_2}^\dagger b$, $\check{v}_{\Upsilon_2} = 0$;
   9: $\hat{T}^{(1)}_k = \text{alg}^{(1)}(A_{k-1}, e_{k-1}, K)$; $\hat{T}^{(1)}_k = K$
   10: $\hat{T}^{(2)}_k = \text{alg}^{(2)}(A_{k-1}, e_{k-1}, K)$; $\hat{T}^{(2)}_k = K$
   11: $[\hat{x}_k, \hat{T}_k] = \text{CoMACS}(A, b, K, \hat{T}^{(1)}_k, \hat{T}^{(2)}_k)$;
   12: $\Lambda_k = \hat{T}^{(1)}_k \cap \hat{T}^{(2)}_k$; $\Lambda_k \leq K$
   13: $U_k = (I - A_{\Lambda_k} A_{\Lambda_k}^\dagger)$;
   14: $A_k = U_k A_{\Lambda_k}$;
   15: $e_k = U_k b$;
   16: $S_k = |\Lambda_k|$;
   17: $r_k = b - A_{\Lambda_k} x_k$;
   18: until $(\|r_k\|_2 \geq \|r_{k-1}\|_2)$;

Outputs: $\hat{T} = \hat{T}_{k-1}$ and $\hat{x} = \hat{x}_{k-1}$.

we have,

$$b = A x = A_{\Lambda_k} x_{\Lambda_k} + A_{\Lambda_k} x_{\Lambda_k}.$$  \hfill (4.6)

Given $\Lambda_k$ (the partially known support-set), the aim is to identify the remaining $K - S_k$ non-zero locations of $x_{\Lambda_k}$ from (4.6). This can be re-stated as a problem of estimating the support-set of a $(K - S_k)$ sparse vector $x_{\Lambda_k}$ satisfying

$$(U_k A_{\Lambda_k}^\dagger) x_{\Lambda_k} = U_k b,$$  \hfill (4.7)

where $U_k$ is the matrix of the orthogonal projection from $\mathbb{R}^M$ onto $\mathcal{R}(A_{\Lambda_k})$ defined as $U_k \triangleq I - A_{\Lambda_k} A_{\Lambda_k}^\dagger$ [150]. In a CS setup, it
reasonable to assume that $A_{\Lambda_k}$ is full column rank. Then we have

$$U_k = I - A_{\Lambda_k} A_{\Lambda_k}^\dagger = I - A_{\Lambda_k} (A_{\Lambda_k}^T A_{\Lambda_k})^{-1} A_{\Lambda_k}^T. \quad (4.8)$$

(4.7) may be easily verified as follows. We have,

$$U_k A_{\Lambda_k} x_{\Lambda_k} = (I - A_{\Lambda_k} A_{\Lambda_k}^\dagger) A_{\Lambda_k} x_{\Lambda_k}$$

(post-multiplying (4.8) with $A_{\Lambda_k} x_{\Lambda_k}$)

$$= b - A_{\Lambda_k} A_{\Lambda_k}^\dagger A_{\Lambda_k} x_{\Lambda_k} - A_{\Lambda_k} x_{\Lambda_k}$$

$$= b - A_{\Lambda_k} A_{\Lambda_k}^\dagger b = U_k b = e_k,$$

where $e_k$ is defined as in Algorithm 4.3.

It has been shown recently that many sparse recovery algorithms can identify more true atoms in the presence of a partially known support-set [33, 35, 40, 139, 140, 149]. Hence using the partially known support-set $\Lambda_k$, the participating algorithms can identify more number of true atoms in $(k+1)^{th}$ iteration. Then, the joint support-set, $\Gamma_{k+1} = \hat{T}^{(1)}_{k+1} \cup \hat{T}^{(2)}_{k+1}$, will contain more number of true atoms than the earlier iterations. This procedure will eventually lead to a better sparse signal estimate by CoMACS in the $(k+1)^{th}$ iteration of ICoMACS. Note that, in the worst case, $\Lambda_k$ may not contain any true atoms and in such cases all $K$ true atoms need to be identified by the participating algorithm, in the $(k+1)^{th}$ iteration. Hence, in the $(k+1)^{th}$ iteration, we use the participating algorithms to identify $K$ atoms. If $\Lambda_k$ contains at least one true atom, the $K$ atoms estimated by the participating algorithm will contain at least one wrong atom. We use LS method to identify the potential atoms from the support-set newly estimated by the participating algorithm and discard the false atoms from $\Lambda_k$. Hence, in the $(k+1)^{th}$ iteration, ICoMACS may include more potential atoms.
in the union-set $\hat{T}^{(1)}_{k+1} \cup \hat{T}^{(2)}_{k+1}$ as compared to $\hat{T}^{(1)}_{k} \cup \hat{T}^{(2)}_{k}$ and discard outdated atoms from $\Lambda_k$. Proceeding in this iterative manner ICoMACS results in a better reconstruction performance than the non-iterative CoMACS. Also we mention that a straightforward iterative extension of StCoMACS is possible to develop in a similar manner.

### 4.3 Numerical Experiments and Results

The sparse reconstruction performance is evaluated using Average Signal-to-Reconstruction-Error Ratio (ASRER), defined in (3.19). To compare the computational requirement by each method, we calculated the average computation time where computation time was calculated using the function ‘cputime’ available in Matlab. To avoid Matlab favouring methods using multiple computational threads, we used the option ‘singleCompThread’ in Matlab. This will limit Matlab to a single computational thread. The specifications of the Desktop machine used to run the simulations are as follows. Matlab version: R2010b (64-bit), Operating System: Ubuntu 13.04 (64-bit), Processor: Intel(R) Core(TM) i7-2600K CPU @ 3.40GHz, and RAM: 16 GB.

CoMACS and ICoMACS with OMP and SP as the participating algorithms are denoted respectively by CoMACS(OMP, SP) and ICoMACS(OMP, SP). We used Basis Pursuit (BP)/Basis Pursuit De-Noising (BPDN) [36] as the third participating algorithm for StCoMACS denoted by StCoMACS(OMP, SP, BP). For BP, we used the function ‘SolveBP’ available in SparseLab [151]. In StCoMACS(OMP, SP, BP), the estimates of OMP and SP are fused first. The resultant estimate is then fused with the estimate of BP. For brevity, we often
drop the name of the participating algorithms and simply use Co-
MACS, ICoMACS, and StCoMACS. We show simulation results for
synthetic as well as real signals.

4.3.1 Synthetic Sparse Signals

For evaluating the performance of the proposed methods, we use
the simulation setup and performance measures described in Sec-
tion 3.5.1 on page 60.

We performed Monte-Carlo simulations with following param-
eters: $N = 500$, $K = 20$, $S = 100$, and $T = 100$. That is, we
generated measurement matrix $A$ 100 times and for each realiza-
tion of $A$, we generated sparse signals with ambient dimension 500
and sparsity level $K = 20$, 100 times.

![Figure 4.1: Performance comparison of FACS and CoMACS, in terms of
Average Signal-to-Reconstruction-Error-Ratio (ASRER) , averaged over
10,000 trials, for Gaussian Sparse Signals (GSS) and Rademacher Sparse
Signals (RSS) in noisy measurement case ($N = 500$, $K = 20$, SMNR = 20
dB).](image)

(a) Average SRER: GSS, SMNR = 20 dB

(b) Average SRER: RSS, SMNR = 20 dB
**Experiment 1**

First, we compare the performance of FACS and CoMACS for GSS and RSS in the presence of measurement noise. The results are shown in Figure 4.1. It may be observed that FACS and CoMACS resulted in a comparable performance in terms of ASRER. We observed a similar behaviour in the rest of the experiments in this chapter and hence for brevity, we have not shown the results for FACS further.

**Experiment 2**

In this experiment we evaluate the performance of the proposed methods CoMACS, ICoMACS, and StCoMACS for GSS and RSS.

*Gaussian Sparse Signals (GSS):* The performance of the proposed methods in terms of ASRER and computation time for GSS, in the presence of measurement noise (SMNR = 20 dB), are shown in Figure 4.2(a) and Figure 4.2(b), respectively. It can be observed that CoMACS gave a significant ASRER improvement as compared to both OMP and SP and ICoMACS further improved the ASRER. For example, at $\alpha = 0.16$, CoMACS resulted in 2.5 dB and 4.7 dB ASRER improvement respectively over OMP and SP which is further improved by 1.7 dB using ICoMACS. Fusing the third participating algorithm, BP, StCoMACS also resulted in further improvement in ASRER as compared to CoMACS, showing the scalability using StCoMACS. At $\alpha = 0.16$, StCoMACS showed an ASRER improvement by 1.7 dB as compared to CoMACS. The improvement in ASRER is gained at the price of higher computational complexity which is shown in Figure 4.2(b). For $\alpha = 0.16$, CoMACS took 3 additional milliseconds as compared to both OMP and SP, and ICoMACS further took 12.8 milliseconds. StCoMACS used 0.24 seconds
FIGURE 4.2: Performance of the proposed CoMACS and variants in terms of Average Signal-to-Reconstruction-Error Ratio (ASRER) and Average Computation Time, averaged over 10,000 trials, for Gaussian Sparse Signals (GSS) and Rademacher Sparse Signals (RSS) in noisy measurement case ($N = 500$, $K = 20$, SMNR = 20 dB).

more as compared to CoMACS. It may be also noted that, for lower values of $\alpha$ (0.10, 0.11, 0.12), StCoMACS(OMP,SP,BP) showed a less ASRER than the participating algorithm BP. It may be also noted that, for $\alpha = 0.1$, ICoMACS(OMP,SP) showed a lesser ASRER than CoMACS(OMP,SP). These exceptional cases show that the fusion strategies may not always give improvement in ASRER. However, from our extensive numerical experiments, we have observed that
in majority of the cases the fusion strategies improve ASRER significantly.

*Rademacher Sparse Signal (RSS)*: The simulation results for RSS with SMNR = 20 dB are given in Figure 4.2(c) and Figure 4.2(d). For RSS also, the proposed method showed ASRER improvement over OMP and SP (refer Figure 4.2(c)). For example, at $\alpha = 0.23$, CoMACS showed ASRER improvement of 10.5 dB and 1.6 dB as compared to OMP and SP, respectively. For this ASRER improvement, CoMACS took only 3 milliseconds additionally, as compared to both OMP and SP. ICoMACS showed 3.1 dB further improvement in ASRER over CoMACS by using additional 10 milliseconds. StCoMACS used 28 milliseconds additionally and showed 7.3 dB improvement over CoMACS.

It may be observed from Figure 4.2(b) and Figure 4.2(d) that the additional computational overhead for StCoMACS, as compared to CoMACS, is mainly due to the computationally intensive participating algorithm BP.

It may be also observed from Figure 4.2 that OMP gave a better ASRER than SP for GSS and *vice-versa* for RSS. Hence if the *a priori* knowledge of the underlying statistical distribution is not available in advance, we cannot get the best sparse recovery performance. Note that CoMACS(OMP, SP) resulted in a better ASRER than both OMP and SP for both GSS and RSS, which clearly shows the advantage of using CoMACS in situations where the underlying statistical distribution is not known *a priori*.

**Experiment 3**

As we mentioned in Section 4.1 on page 74, in principle, we can use any sparse recovery algorithm for estimating $K$ non-zero elements
from $\Gamma$. In this experiment, we explore this option empirically and compare the performance with our proposed least-squared based CoMACS. We use BP as an alternate method for estimating correct atoms from $\Gamma$, which we denote by CoMACS_L1(OMP, SP). That is in CoMACS, we solve the following problem using BP.

$$v_\Gamma = \left\{ \min_{y \in \mathbb{R}^{N \times 1}} \|y\|_1 \text{ s.t. } \|A_\Gamma y - b\|_2 \leq \epsilon \right\},$$

and set $v_{\Gamma^c} = 0$ where $v \in \mathbb{R}^{N \times 1}$. Using this settings, we choose estimated support-set $\hat{x} = v^K$ and $\hat{T} = \text{supp}(v^K)$.

We used the function ‘SolveBP’ available in SparseLab [151] for CoMACS_L1 implementation. We compare the performance of these algorithms with our proposed LS based CoMACS which we denote by CoMACS. The results in a clean measurement case for GSS is given in Figure 4.3(a) and Figure 4.3(b).

**Figure 4.3**: Performance of CoMACS, $\ell_1$ based CoMACS (CoMACS_L1) in terms of Average Signal-to-Reconstruction-Error-Ratio (ASRER) and Average Computation Time, averaged over 10,000 trials, for Gaussian Sparse Signals (GSS) in clean measurement case ($N = 500$, $K = 20$).
From Figure 4.3(a) it can be seen that CoMACS_L1 and ICoMACS_L1 gave a similar ASRER as CoMACS and ICoMACS respectively. But Figure 4.3(b) reveals that CoMACS_L1 and ICoMACS_L1 took more than twice the computation time, on an average, as compared to CoMACS and ICoMACS respectively. A similar result was also observed for RSS and also for noisy measurement cases which are not shown here for brevity. For reproducible codes to repeat these experiments, please refer Section 4.3.1.2. This experiment shows that the proposed strategies (CoMACS and ICoMACS) provide a better trade-off between computation and performance, when compared to the alternate strategy of $\ell_1$ approach.

### 4.3.1.1 Large Dimensional Problems

To compare the performance of the proposed methods with the alternate $\ell_1$ strategies for large dimensional problems, we repeated the above experiment for GSS in clean measurement case with $N = 50,000$ and $K = 2,000$. As the experiment was very time consuming, we conducted only 500 trials. In each trial, both $A$ and $x$ were newly generated. Here also, CoMACS(OMP, SP) and CoMACS_L1(OMP, SP) gave a similar ASRER, which is better than the ASRER of both OMP and SP. For brevity, we have not presented the ASRER results. The average computation time taken by each method is shown in Figure 4.4.

Here, CoMACS has only a marginal advantage over CoMACS_L1 in terms of average computation time. But ICoMACS again showed computational advantage over ICoMACS_L1. For example, at $\alpha = 0.16$, on an average ICoMACS(OMP, SP) took only 5.1 hours whereas ICoMACS_L1(OMP, SP) took 10 hours for sparse signal reconstruction. Note that, in addition to the computational advantage, proposed CoMACS provides theoretical guarantees.
FIGURE 4.4: Performance of CoMACS, L1 based CoMACS (CoMACS_L1) in terms of Average Computation Time, averaged over 500 trials, for Gaussian Sparse Signals (GSS) of large dimension in clean measurement case \((N = 50,000, K = 2,000)\).

4.3.1.2 Reproducible Research

We have also done simulations for other values of SMNR s which showed a similar performance advantage for the proposed methods (CoMACS, ICoMACS, and StCoMACS) in terms of ASRER. In the spirit of reproducible research [136, 137], we provide necessary Matlab codes publicly available. It is downloadable freely at http://www.ece.iisc.ernet.in/~ssplab/Public/CoMACS.tar.gz. The code may be used to reproduce the results shown in Figure 4.2, Figure 4.3, and Figure 4.4.

4.3.2 Real Compressible Signals

To evaluate the performance of the proposed schemes on compressible signals (which are not sparse) and real-world applications, we also conducted experiments on real-world signals. We used ECG signals due to their good structure for sparse decomposition. We use a similar setup as explained in [139,140] for this purpose. The
experiments were carried out over 10-minute long leads extracted from records 100, 101, 102, 103, 107, 109, 111, 115, 117, 118, and 119 from the MIT-BIH Arrhythmia Database [138, 141]. We used cosine modulated filter banks to determine a sparse representation of the signal [152]. 1024 samples of ECG data were processed to determine the sparse signal approximation, setting the number of channels to 16. Here also, we used OMP, SP, and BP as the participating algorithms. We assumed a sparsity level 128, and the reconstruction results and computation time averaged over 20 trials are respectively shown in Figure 4.5(a) and Figure 4.5(b). As in Section 4.3.1, we used Gaussian measurement matrices with appropriate dimensions to vary the fraction of measurements, \( \alpha \), from 0.25 to 0.49 with an increment of 0.03.

**Figure 4.5:** Performance of the proposed methods in terms of Average Signal-to-Reconstruction-Error-Ratio (ASRER) and Average Computation Time for ECG Signals, averaged over 20 trials, selected from MIT-BIH Arrhythmia Database [138, 141] \((N = 1024, K = 128)\)

From Figure 4.5(a), it can be seen that as in the case of synthetic sparse signals, here also the proposed schemes resulted in a better reconstructed signal as compared to the participating algorithms. For example, for \( \alpha = 0.31 (M = 317) \), CoMACS(OMP,SP) gave 4.4 dB and 7.6 dB improvement over OMP and SP respectively.
ICoMACS(OMP,SP) and StCoMACS(OMP,SP,BP) further improved ASRER respectively by 3.4 dB and 2.1 dB over CoMACS(OMP,SP). The improved ASRER of the proposed methods comes with the cost of additional computational complexity as depicted in Figure 4.5(b). An exception can be found at $\alpha = 0.25$, where StCoMACS(OMP,SP,BP) resulted in a lesser ASRER than the participating algorithm BP.

### 4.4 Summary

We introduced a framework to fuse the estimates of multiple sparse recovery algorithms and proposed different methods for fusion. The proposed methods are general in nature and can accommodate any sparse signal reconstruction algorithm as a participating algorithm. We derived performance guarantees for the proposed methods using RIP. Using OMP, SP, and BP as the participating algorithms, we conducted numerical experiments with synthetic data (both continuous and discrete amplitudes) and real ECG signals. The simulation results showed the robustness of the proposed schemes for different types of signals under measurement perturbations and its efficacy in real-world applications.

### 4.4.1 Relevant Publications

4.A Proof of Proposition 4.1

We have,

$$\|x - v\|_2 \leq \|x_{\Gamma} - v_{\Gamma}\|_2 + \|x_{\Gamma^c}\|_2 \quad (\because v_{\Gamma^c} = 0). \quad (4.9)$$

Consider

$$\|x_{\Gamma} - v_{\Gamma}\|_2 = \|x_{\Gamma} - A_{\Gamma}^\dagger (Ax + w)\|_2 \quad \text{(using definition of b)}$$

$$= \|x_{\Gamma} - A_{\Gamma}^\dagger (A_{\Gamma}x_{\Gamma} + A_{\Gamma^c}x_{\Gamma^c} + w)\|_2$$

$$= \|A_{\Gamma}^\dagger (A_{\Gamma^c}x_{\Gamma^c} + w)\|_2 \quad (\because A_{\Gamma}^\dagger A_{\Gamma} = I)$$

$$\leq \|(A_{\Gamma}^H A_{\Gamma})^{-1} A_{\Gamma}^H A_{\Gamma} x_{\Gamma^c}\|_2 + \|A_{\Gamma}^\dagger w\|_2$$

(using definition of $A_{\Gamma}^\dagger$)

$$\leq \frac{1}{1 - \delta_R} \|A_{\Gamma}^H A_{\Gamma} x_{\Gamma^c}\|_2 + \frac{1}{\sqrt{1 - \delta_R}} \|w\|_2$$

(using (2.6), (2.7))

$$\leq \frac{\delta_{R+K}}{1 - \delta_{R+K}} \|x_{\Gamma^c}\|_2 + \frac{1}{1 - \delta_{R+K}} \|w\|_2 \quad (4.10)$$

(using (2.9), $\delta_R \leq \delta_{R+K}$)

Substituting (4.10) in (4.9), we get

$$\|x - v\|_2 \leq \frac{1}{1 - \delta_{R+K}} \|x_{\Gamma^c}\|_2 + \frac{1}{1 - \delta_{R+K}} \|w\|_2. \quad \blacksquare$$
4.B Proof of Theorem 4.2 (Analysis of Signal and Measurement Perturbations)

i) Consider

\[ b = Ax + w = Ax^K + A(x - x^K) + w = Ax^K + \tilde{w} \] (4.11)

where \( \tilde{w} = A(x - x^K) + w \). Observe (4.11) as a standard CS measurement setup for a \( K \)-sparse signal given in (2.4) with \( \tilde{w} = w \). Hence using (4.4), we get

\[
\|x - \hat{x}\|_2 \leq \frac{3}{1 - \delta_{R+K}} \left( \|x_{\Gamma^c}\|_2 + \|\tilde{w}\|_2 \right) + 2\nu \|x_{\Gamma^c}\|_2 \\
= \frac{3 + 2\nu(1 - \delta_{R+K})}{1 - \delta_{R+K}} \|x_{\Gamma^c}\|_2 + \frac{3\|\tilde{w}\|_2}{1 - \delta_{R+K}}
\] (4.12)

We have,

\[
\|\tilde{w}\|_2 = \|A(x - x^K) + w\|_2 \\
\leq \|A(x - x^K)\|_2 + \|w\|_2 \\
\leq \sqrt{1 + \delta_{R+K}} \left( \|x - x^K\|_2 + \frac{\|x - x^K\|_1}{\sqrt{R + K}} \right) + \|w\|_2
\] (4.13)

(using Lemma 2.2)

Substituting (4.13) in (4.12), we get

\[
\|x - \hat{x}\|_2 \leq \frac{3\sqrt{1 + \delta_{R+K}}}{1 - \delta_{R+K}} \left( \|x - x^K\|_2 + \frac{\|x - x^K\|_1}{\sqrt{R + K}} \right)
\]
ii) Using definition of $\xi$ and $\zeta$ in (4.14), we have,

$$
\|\mathbf{x} - \hat{\mathbf{x}}\|_2 \leq \frac{3\xi + 3 + 2\nu(1 - \delta_{R+K}) + 3\zeta}{1 - \delta_{R+K}} \|\mathbf{x}_{\Gamma^c}\|_2
$$

$$
= \frac{3\xi + 3 + 2\nu(1 - \delta_{R+K}) + 3\zeta}{1 - \delta_{R+K}} \eta_i \|\mathbf{x}_{T_i^c}\|_2
$$

$$
= \frac{3\xi + 3 + 2\nu(1 - \delta_{R+K}) + 3\zeta}{1 - \delta_{R+K}} \eta_i \|\mathbf{x} - \hat{\mathbf{x}}_i\|_{T_i^c} \|_2
$$

$$
\leq \frac{3\xi + 3 + 2\nu(1 - \delta_{R+K}) + 3\zeta}{1 - \delta_{R+K}} \eta_i \|\mathbf{x} - \hat{\mathbf{x}}_i\|_2
$$

Hence, we get, SRER for CoMACS in case of arbitrary signals

$$
\text{SRER}_{\text{CoMACS}} = \frac{\|\mathbf{x}\|_2^2}{\|\mathbf{x} - \hat{\mathbf{x}}\|_2^2} \geq \frac{(1 - \delta_{R+K})^2}{\|\mathbf{x} - \hat{\mathbf{x}}_i\|_2^2 (3\xi + 3 + 2\nu(1 - \delta_{R+K}) + 3\zeta)^2 \eta_i^2}.
$$
CHAPTER 5

Progressive Fusion for Low Latency Applications

“It does not matter how slowly you go as long as you do not stop.”
Confucius [551-479 BC]

In the previous chapters we developed a fusion framework where a set of algorithms participates. The participating algorithms provide estimates of the sparse signal independently and then the estimates are fused to achieve a final better estimate. Fusion of Algorithms for Compressed Sensing (FACS) discussed in Chapter 3 work in a batch mode which require availability of all the estimates of participating algorithms before performing the fusion. It is known that sparse signal reconstruction algorithms have varying complexity leading to varying latency for providing estimates. The total latency requirement of existing fusion strategies is decided by the participating algorithm that has the highest computational complexity. For example, a convex relaxation based algorithm requires more complexity (in turn a high latency) than a greedy algorithm, in general. In many applications with a low latency requirement, it
is desirable to achieve a progressive improvement of reconstruction quality (or estimation quality). For example, image coding schemes with progressive improvement of quality are used in internet based image browsing applications [153]; the image coding schemes satisfy a low latency requirement for a good viewing experience.

Considering a low latency requirement, we develop a progressive fusion strategy for sparse signal reconstruction in a standard Compressed Sensing (CS) setup. In this progressive fusion strategy, the fusion is performed according to a rule: estimates of low latency algorithms are fused first and then the estimates of high latency algorithms are progressively fused. Naturally the reconstruction satisfies a low latency requirement with the aspect of quality improvement in progression. Our proposed scheme is referred to as progressive Fusion of Algorithms for Compressed Sensing (pFACS). For the pFACS, we theoretically characterize the progressive fusion strategy for reconstruction quality improvement by similar tools developed in Chapter 3 and show its advantages by simulations.

5.1 Progressive Fusion of Algorithms for Compressed Sensing (pFACS)

In this section, we develop low latency pFACS based on our earlier method FACS [91,154] developed in Chapter 3. FACS has a high latency requirement. FACS uses a set of Sparse Reconstruction Algorithms (SRAs) independently and fuses their estimates to improve the sparse signal reconstruction performance. The fusion strategy is based on a Least-Squares (LS) approach. Let us assume that $P$ SRAs are independently used to recover the $K$-sparse signal $x$ in the CS setup (2.4). For the $i^{th}$ participating algorithm, let $\hat{x}_i$ and
\( \hat{T}_i \) denote the estimated sparse signal and the estimated support-set, respectively. We denote the \( i^{th} \) participating algorithm by \( \text{alg}^{(i)} \). We also assume that \( \| \hat{x}_i \|_0 = | \hat{T}_i | = K \) (\( i = 1, 2, \ldots, P \)). FACS estimates the support-set from \( \Gamma \triangleq \bigcup_{i=1}^{P} \hat{T}_i \), where \( \Gamma \) is the union of the support-sets \( \{ \hat{T}_i \} \) estimated by the \( P \) participating algorithms (see Chapter 3 for more details on this choice). We assume that \( R \triangleq |\Gamma| \leq M \).

### 5.1.1 Proposed Progressive FACS (pFACS)

Without loss of generality, let us assume that the participating algorithms are ordered in ascending order of their computational requirement (latency requirement). I.e., among the \( P \) participating algorithms, \( \text{alg}^{(1)} \) is the least computationally demanding, \( \text{alg}^{(2)} \) is the second least computationally demanding, and so on and \( \text{alg}^{(P)} \) is the most computationally demanding. Similar to FACS, here also we employ \( P \) participating algorithms independently, in parallel. But in pFACS, we will not wait for all participating algorithms to terminate. As soon as the estimate of \( \text{alg}^{(2)} \) is available, we fuse the estimates of \( \text{alg}^{(1)} \) and \( \text{alg}^{(2)} \) using FACS. The resultant fused estimate is then fused with estimate of \( \text{alg}^{(3)} \) (depending on availability), and so on till the fusion of estimate of \( \text{alg}^{(P)} \). Therefore, in pFACS, fusion is done progressively in \( P - 1 \) stages. We denote the \( i^{th} \) stage of pFACS by \( \text{pFACS}(i) \). For \( \text{pFACS}(i) \) (\( i = 1, 2, \ldots, P - 1 \)), let \( \hat{x}_i \) and \( \hat{T}_i \) denote the estimated sparse signal and support-set respectively. In \( \text{pFACS}(i + 1) \), the estimates of \( \text{pFACS}(i) \) and \( \text{alg}^{(i+1)} \) are fused using FACS. Note that at each stage pFACS always fuses only two estimates using FACS. pFACS algorithm is given in Algorithm 5.1.


**Chapter 5 Progressive Fusion for Low Latency Applications**

**Algorithm 5.1 : progressive Fusion of Algorithms for Compressed Sensing (pFACS)**

**Inputs:** $\mathbf{A} \in \mathbb{R}^{M \times N}$, $\mathbf{b} \in \mathbb{R}^{M \times 1}$, $K$, and $\{\mathbf{T}_i\}_{i=1}^P$.

1: $\hat{T}_0 = \mathbf{T}_1$;
2: for $j = 1 : P - 1$ do
3: $(\hat{x}_j, \hat{\mathbf{T}}_j) = \text{FACS}(\mathbf{A}, \mathbf{b}, K, \hat{T}_{j-1}, \hat{T}_{j+1})$; ▶ realizes pFACS($j$)
4: end for

**Outputs:** $\hat{x} = \hat{x}_{P-1}$ and $\hat{T} = \hat{T}_{P-1}$.

### 5.1.2 Theoretical Analysis of pFACS

Next, we theoretically analyse pFACS (Algorithm 5.1) using Restricted Isometry Property (RIP) of the measurement matrix $\mathbf{A}$. The performance analysis is characterized by Signal-to-Reconstruction-Error Ratio (SRER). Proposition 5.1 provides a sufficient condition for pFACS to provide SRER improvement over alg$(i)$ in the $(i-1)^{th}$ stage of fusion ($i = 2, \ldots, P$).

**Proposition 5.1.** Assume that we have employed $P \geq 2$ participating algorithms independently to reconstruct the $K$-sparse signal $\mathbf{x}$ from (2.4). Let $\hat{x}_i$ and $\hat{\mathbf{T}}_i$ respectively denote the sparse signal and support-set estimated by alg$^{(i)}$ ($i = 1, 2, \ldots, P$). Let pFACS$(j)$ denote the pFACS algorithm with the least computationally demanding $j + 1$ algorithms as participating algorithms. Let $\hat{x}_0 = \hat{x}_1$ and $\hat{T}_0 = \hat{T}_0$, and let $\hat{x}_j$ and $\hat{T}_j$ respectively denote the estimate of sparse signal and support-set obtained by pFACS$(j)$, $(j = 1, 2, \ldots, P - 1)$. Let the CS measurement matrix $\mathbf{A}$ hold RIP with the Restricted Isometry Constant (RIC) $\delta_{3K}$. Let $\Gamma_j = \hat{T}_{j-1} \cup \hat{T}_{j+1}$, $(j = 1, 2, \ldots, P - 1)$. Assuming $\|\mathbf{x}_{\hat{T}_{j+1}^c}\|_2 \neq 0$, $\|\mathbf{x}_{\hat{T}_j^c}\|_2 \neq 0$, define $\eta_j = \frac{\|\mathbf{x}_{\hat{T}_j^c}\|_2}{\|\mathbf{x}_{\hat{T}_{j+1}^c}\|_2}$ and $\zeta_j = \frac{\|\mathbf{w}\|_2}{\|\mathbf{x}_{\hat{T}_j^c}\|_2}$. Then, pFACS$(j)$ provides at least SRER gain of $\left(\frac{(1-\delta_{3K})^2}{(1+\delta_{3K}+3\zeta_j)^2} \eta_j\right)^2$ over the alg$^{(j+1)}$ if $\eta_j < \frac{(1-\delta_{3K})^2}{1+\delta_{3K}+3\zeta_j}$, $(j = 1, 2, \ldots, P - 1)$. 

104
Chapter 5 Progressive Fusion for Low Latency Applications

Proof. To prove this, we use the fact that in $j^{th}$ stage, pFACS uses FACS to fuse the estimated support-sets $\tilde{T}_{j-1}$ and $\tilde{T}_{j+1}$. Since $|\tilde{T}_{j-1}| = |\tilde{T}_{j+1}| = K$, we have $|T_j| = |\tilde{T}_{j-1} \cup \tilde{T}_{j+1}| \leq 2K$. Now, the result follows from Theorem 3.1 by setting $P = 2$, $\hat{x}_1 = \hat{x}_{j-1}$, $\hat{x}_2 = \hat{x}_{j+1}$, $\hat{T}_1 = \tilde{T}_{j-1}$, and $\hat{T}_2 = \tilde{T}_{j+1}$. 

5.1.3 On Latency of pFACS

To discuss the advantage of pFACS over FACS in terms of latency, we consider two popular family of algorithms, greedy pursuit algorithms and convex relaxation algorithms, widely used for sparse signal reconstruction in CS. In general, the average reconstruction performances are in decreasing trends for convex and greedy algorithms. However, the computational cost is also in a decreasing trend for the mentioned order of algorithms. In this work, we consider four popular sparse signal reconstruction algorithms viz. Matching Pursuit (MP), Orthogonal Matching Pursuit (OMP), Subspace Pursuit (SP), and Basis Pursuit De-Noising (BPDN) as participating algorithms to discuss the advantage of pFACS over FACS in terms of low latency. The algorithms are listed in the ascending order of their computational complexity. We use pFACS(MP,OMP), pFACS(MP,OMP,SP), and pFACS(MP,OMP,SP,BPDN) respectively to denote pFACS with the participating algorithms listed in the brackets, fused progressively in the order of their appearance in the list. The computational complexities, in general, for MP, SP, OMP and BPDN are $O(KMN)$, $O(K(MN + K^2 + KM))$, $O(K(MN + K^2M))$, and $O(N^3)$ respectively [83]. The computational complexity of FACS with these algorithms as the participating algorithms will be at least $O(N^3)$. Note that the computational complexities of pFACS(MP,OMP) and pFACS(MP,OMP,SP) are a little more than $O(K(MN + K^2 + KM))$ and $O(K(MN + K^2M))$ respectively which
are significantly smaller as compared to the computational complexity of FACS, for large values of \( N \).

### 5.1.4 pFACS vis-a-vis FACS

In this section, we provide a list of remarks on pros-and-cons of pFACS vis-a-vis FACS as follows:

- In pFACS, fused estimates are available in a progressive manner. pFACS provides successive refinements of the estimates and give quick interim results during the fusion process.

- In many applications, it is possible to measure the reconstruction quality of the estimated signal (for example, using Peak Signal-to-Noise Ratio (PSNR) in image processing applications). In such applications, we can stop pFACS at any interim stage, as soon as the required reconstruction quality is met.

- pFACS gives flexible control over whole fusion process and provides a flexible mechanism to trade reconstruction quality and latency.

- Since the interim results are available and the quality can be assessed in many applications, we can, on the fly, change participating algorithms in pFACS. FACS operates in a batch mode where we have to fix the participating algorithms in advance.

- pFACS can handle any number of participating algorithms whereas FACS requires \(|\Gamma| \leq M |\) which in turn limits the number of participating algorithms.
• As compared to FACS, pFACS demands the measurement matrix $A$ to have a smaller RIC $\delta_{3K}$ which is independent of the number of participating algorithms.

• On the down side, pFACS requires to solve $P - 1$ LS problems whereas FACS requires solution for only one LS problem.

Using the simulations explained in Section 5.2 we show that pFACS and FACS give a similar sparse reconstruction performance in terms of SRER in an average sense.

## 5.2 Numerical Experiments and Results

We evaluated the performance of pFACS using synthetic signals and real-world signals. We used MP, OMP, SP, and BPDN as the participating algorithms. MP, OMP, and SP codes were realized in Matlab and for BPDN, we used $\ell_1$-magic toolbox [110]. Note that BPDN will not directly estimate the support-set. We choose the indices corresponding to the $K$-largest magnitudes of the signal estimate as the estimated support-set of BPDN. The sparse reconstruction performance is evaluated using Average Signal-to-Reconstruction-Error Ratio (ASRER), defined in (3.19).

### 5.2.1 Experiment 1 (Synthetic Signals)

To evaluate the performance of pFACS we conducted simulations using RSS with signal dimension $N = 500$ and sparsity level $K = 20$. We followed the simulation setup described in Section 3.5.1 on page 60. We consider measurement noise $w \sim \mathcal{N}(0, \sigma_w^2 I_M)$. The simulations were carried out for small values of $\alpha$, defined in
Figure 5.1: Progressive performance of pFACS in terms of ASRER for Rademacher Sparse Signals (RSS) ($N = 500$, $K = 20$, and Signal-to-Measurement-Noise Ratio (SMNR) = 20 dB).

(7.28). To benchmark the performance, we also use an oracle estimator in the simulations. The oracle estimator assumes knowledge about the true support-set and with help of least-squares, estimates the non-zero magnitudes of the sparse signal.

Figure 5.1 shows the simulation results for RSS with SMNR = 20 dB averaged over 10,000 trails. Figure 5.1(a)-(d) show the progressive improvement in ASRER resulted by employing pFACS. For example, at $\alpha = 0.24$, pFACS(MP,OMP), pFACS(MP,OMP,SP),
and pFACS(MP,OMP,SP,BPDN) showed 4.6 dB, 16.4 dB, and 22.8 dB SRER improvement respectively over MP. It may be noted from Figure 5.1(c) that, pFACS(MP,OMP,SP,BPDN) showed a lesser ASRER as compared to the participating algorithm BPDN for smaller values of $\alpha$ (0.17, 0.18). Except for these values, pFACS provided ASRER improvement over all the participating algorithms. It may be also observed that, FACS provided a similar ASRER as that given by pFACS(MP,OMP,SP,BPDN). Note that, pFACS always need to deal only with at most $2K$ atoms whereas FACS need to deal with at most $PK$ atoms to estimate the $K$ true atoms. This gives pFACS as slight advantage over FACS in numerical computations. This advantage is visible in Figure 5.1(d).

To evaluate the computational advantage of pFACS over FACS, we also show the average computation time taken by different algorithms, in Table 5.1. The computation time was measured using the function ‘cputime’ available in Matlab. The specifications of the Desktop machine used to run the simulations are given below. Matlab version: R2010b (64-bit), Operating System: Ubuntu 13.04 (64-bit), Processor: Intel(R) Core(TM) i7-2600K CPU @ 3.40GHz, and RAM: 16 GB. To avoid Matlab favouring methods using multiple computational threads, we used the option ‘single-CompThread’ in Matlab. This option forces Matlab to use only single thread for computation.

It can be observed that pFACS(MP,OMP) and pFACS(MP,OMP,SP) provided quick interim estimates as compared to the fusion algorithm FACS(MP,OMP,SP,BPDN), which is a highly desirable behaviour for low latency applications. For example, at $\alpha = 0.30$ (refer Figure 5.1 and Table 5.1), pFACS scheme was able to produce the first output (pFACS(MP,OMP)) in 0.036 seconds giving 10.5 dB SRER gain over MP, the next output (pFACS(MP,OMP,SP)) was provided in another 0.035 seconds which showed a SRER gain
Table 5.1: Comparison of average computation time (in seconds) by different algorithms for Rademacher Sparse Signals (RSS) ($N = 500$, $K = 20$) with SMNR = 20 dB, averaged over 10,000 trials.

<table>
<thead>
<tr>
<th>Fraction of Measurements ($\alpha$)</th>
<th>0.18</th>
<th>0.22</th>
<th>0.26</th>
<th>0.30</th>
</tr>
</thead>
<tbody>
<tr>
<td>MP</td>
<td>0.005</td>
<td>0.005</td>
<td>0.005</td>
<td>0.005</td>
</tr>
<tr>
<td>pFACSM(P,OMP)</td>
<td>0.032</td>
<td>0.033</td>
<td>0.035</td>
<td>0.036</td>
</tr>
<tr>
<td>pFACSM(P,OMP,SP)</td>
<td>0.057</td>
<td>0.060</td>
<td>0.061</td>
<td>0.061</td>
</tr>
<tr>
<td>pFACSM(P,OMP,SP,BPDN)</td>
<td>4.311</td>
<td>4.835</td>
<td>5.595</td>
<td>6.045</td>
</tr>
<tr>
<td>FACSM(P,OMP,SP,BPDN)</td>
<td>4.299</td>
<td>4.822</td>
<td>5.581</td>
<td>6.031</td>
</tr>
</tbody>
</table>

of 10.6 dB as compared to the first output (pFACSM(P,OMP)). Note that pFACSM(P,OMP,SP) gave a similar ASRER as that provided by FACSM(P,OMP,SP,BPDN) saving 5.91 seconds.

5.2.1.1 Reproducible Research

In the spirit of reproducible research, we provide the Matlab codes at http://www.ece.iisc.ernet.in/~ssplab/Public/pFACSM.tar.gz, which reproduces the results shown in Figure 5.1.

5.2.2 Experiment 2 (Real-World Compressible Signals)

To evaluate the performance of pFACSM on real-world applications, we also conducted experiments on ECG signals from MIT-BIH Arrhythmia Database [138]. ECG signals are compressible and have good structure for sparse decomposition.

We used the same simulation setup as used in [139] and [140]. We assumed a sparsity level, $K = 128$, and the reconstruction results are shown in Figure 5.2. Similar to the synthetic sparse signal
Figure 5.2: Performance of pFACS in terms ASRER for real-world ECG signals ($N = 1024$ and $K = 128$) from MIT-BIH Arrhythmia database [138]

From Figure 5.2, it can be observed that pFACS progressively improved SRER for ECG signals also. For example, for $M = 320$, pFACS(MP,OMP) and pFACS(MP,OMP,SP) respectively gave 12.7 dB and 16.1 dB ASRER improvement over MP. pFACS(MP,OMP,SP) and pFACS(MP,OMP,SP,BPDN) resulted in a similar ASRER as given by FACS(MP,OMP,SP,BPDN).

### 5.3 Summary

In this chapter, we proposed a progressive scheme for fusion of sparse signal reconstruction algorithms viz. pFACS, which is suitable for low latency applications while enjoying the advantages of an earlier proposed fusion algorithm called FACS. For large dimensional problems, pFACS produces quick interim estimates by fusing the least computationally complex participating algorithms.
We theoretically analysed the performance of pFACS and showed that pFACS requires a smaller RIC for the measurement matrix, as compared to FACS. Using numerical experiments we showed that pFACS improves the sparse signal reconstruction progressively.

5.3.1 Relevant Publication

Fusion of Algorithms for Multiple Measurement Vectors

“There is no harm in repeating a good thing.”
Plato [428-348 BC]

The problem discussed so far in the previous chapters, described in (2.4), involves only a single measurement vector. This problem is known as the Single Measurement Vector (SMV) problem. A natural extension of the SMV problem is the Multiple Measurement Vector (MMV) problem where a set of $L$ measurements are given:

$$b^{(l)} = Ax^{(l)}, \quad l = 1, 2, 3, \ldots, L$$

(6.1)

The vectors $\{x^{(l)}\}_{l=1}^L$ are assumed to have a common sparse support-set. The problem is to estimate $x^{(l)}$ ($l = 1, 2, \ldots, L$) from (6.1). When $L = 1$, this problem reduces to (2.4), the SMV problem. Instead of recovering the $L$ signals individually, the attempt in the MMV problem is to simultaneously recover all the $L$ signals. MMV
Chapter 6 Fusion of Algorithms for Multiple Measurement Vectors

problem arises in many applications such as the neuromagnetic inverse problem in Magnetoencephalography (a modality for imaging the brain) [32, 155], array processing [156], non-parametric spectrum analysis of time series [157], and equalization of sparse communication channels [158].

Recently many algorithms have been proposed to recover signal vectors with a common sparse support. Some among them are algorithms based on diversity minimization methods like $\ell_{2,1}$ minimization [159], and M-FOCUSS [160], greedy methods like M-OMP and M-ORMP [160], and Bayesian methods like MSBL [146] and T-MSBL [46]. The ReMBo algorithm [161] linearly combines the multiple measurement vectors into a single measurement vector and then solves the resultant single measurement vector problem.

In this chapter, we extend the fusion framework developed in earlier chapters for MMV problem. Like the SMV problem, several MMV reconstruction algorithms participate and combine their estimates to determine the final signal estimate. We refer to this scheme as Multiple Measurement Vector Fusion of Algorithms for Compressed Sensing (MMV-FACS). We theoretically analyse this fusion based scheme and derive sufficient condition for achieving a better reconstruction performance than any individual participating algorithm. We derive an upper bound on the reconstruction error by MMV-FACS. Also we analyse the average-case performance of MMV-FACS. By numerical experiments we show that fusion of viable algorithms leads to improved reconstruction performance for the MMV problem.

114
Chapter 6 Fusion of Algorithms for Multiple Measurement Vectors

6.1 Problem Formulation

The MMV problem involves solving the following $L$ under-determined systems of linear equations

$$b^{(l)} = Ax^{(l)} + w^{(l)}, \quad l = 1, 2, 3, \ldots, L$$  \hspace{1cm} (6.2)

where $A \in \mathbb{R}^{M \times N} (M \ll N)$ represents the measurement matrix, $b^{(l)} \in \mathbb{R}^{M \times 1}$ represents the $l$th measurement vector, and $x^{(l)} \in \mathbb{R}^{N \times 1}$ denote the corresponding $K$-sparse source vector. That is, $|\text{supp}(x^{(l)})| \leq K$ and $x^{(l)}$ share a common support-set for $l = 1, 2, \ldots, L$. $w^{(l)} \in \mathbb{R}^{M \times 1}$ represents the additive measurement noise. We can rewrite (6.2) as

$$B = AX + W$$  \hspace{1cm} (6.3)

where $X = [x^{(1)}, x^{(2)}, \ldots, x^{(L)}]$, $W = [w^{(1)}, w^{(2)}, \ldots, w^{(L)}]$, and $B = [b^{(1)}, b^{(2)}, \ldots, b^{(L)}]$. For a matrix $X$, we define

$$\text{supp}(X) = \bigcup_{i=1}^{L} \text{supp}(x^{(i)}).$$

$X$ is a $K$ jointly sparse matrix. That is, $|\text{supp}(X)| \leq K$. There are at most $K$ rows in $X$ that contain non-zero elements. Like in the SMV case, here also we assume that $K < M$ and $K$ is known a priori.
6.2 Fusion of Algorithms for Multiple Measurement Vector Problem

Let $P \geq 2$ denote the number of different participating algorithms employed to estimate the sparse signal. Let $\hat{T}_i$ denote the support-set estimated by the $i^{th}$ participating algorithm and let $\mathcal{T}$ denote the true-support-set. Denote the union of the estimated support-sets as $\Gamma$, i.e., $\Gamma \triangleq \bigcup_{i=1}^{P} \hat{T}_i$, assume that $R \triangleq |\Gamma| \leq M$. Since we are estimating the support atoms only from $\Gamma$, we need to only solve the following problem which is lower dimensional as compared to the original problem (6.3):

$$B = A_{\Gamma}X_{\Gamma,:} + \tilde{W},$$

(6.4)

where $A_{\Gamma}$ denotes the sub-matrix formed by the columns of $A$ whose indices are listed in $\Gamma$, $X_{\Gamma,:}$ denotes the submatrix formed by the rows of $X$ whose indices are listed in $\Gamma$, and $\tilde{W} = W + A_{\Gamma^c}X_{\Gamma^c,:}$. The matrix equation (6.4) represents a system of $L$ linear equations which are over-determined in nature. We use the method of Least-Squares (LS) to find an approximate solution to the overdetermined system of equations in (6.4). Let $V_{\Gamma,:}$ denote the LS solution of (6.4). We choose the support-set estimate of MMV-FACS as the support of $V^K$, i.e., indices of those rows having the largest $\ell_2$-norm. Once the non-zero rows are identified, solving the resultant overdetermined solution using LS we can estimate the non-zero entries of $\hat{X}$. MMV-FACS is summarized in Algorithm 6.1.

Remark:

An alternate approach for solving an MMV problem is to stack all the columns of $B$ to get a single measurement vector. Then (6.3)
Chapter 6  Fusion of Algorithms for Multiple Measurement Vectors

Algorithm 6.1: MMV-FACS

**Inputs:** \( A \in \mathbb{R}^{M \times N}, B \in \mathbb{R}^{M \times L}, K \), and \( \left\{ \hat{T}_i \right\}_{i=1:P} \).

**Assumption:** \( |\bigcup_{i=1}^{P} \hat{T}_i| \leq M \).

**Initialization:** \( V = 0 \in \mathbb{R}^{N \times 1} \).

**Fusion:**
1. \( \Gamma = \bigcup_{i=1}^{P} \hat{T}_i; \)
2. \( V_{\Gamma,:} = A_{\Gamma,:}^\dagger B, V_{\Gamma^c,:} = 0; \)
3. \( \hat{T} = \text{supp}(V^K); \)

**Outputs:** \( \hat{T} \) and \( \hat{X} \) (where \( \hat{X}_{\hat{T},:} = A_{\hat{T},:}^\dagger B \) and \( \hat{X}_{\hat{T}^c,:} = 0 \))

In a noiseless case becomes

\[
\begin{bmatrix}
  b_1 \\
  b_2 \\
  \vdots \\
  b_L
\end{bmatrix}_{ML \times 1} =
\begin{bmatrix}
  A & 0 \\
  A & \cdots & \ddots & 0 \\
  0 & \cdots & \ddots & A \\
\end{bmatrix}_{ML \times NL}
\begin{bmatrix}
  x_1 \\
  x_2 \\
  \vdots \\
  x_L
\end{bmatrix}_{NL \times 1},
\]

where \( b_i \) and \( x_i (i = 1, 2, \ldots, L) \) denote the \( i \)th column of \( B \) and \( X \) respectively. Now, we have the following SMV problem.

\[
\begin{bmatrix}
  b_1 \\
  b_2 \\
  \vdots \\
  b_{NL}
\end{bmatrix}_{ML \times 1} =
\begin{bmatrix}
  A & 0 \\
  A & \cdots & \ddots & 0 \\
  0 & \cdots & \ddots & A \\
\end{bmatrix}_{ML \times NL}
\begin{bmatrix}
  x_1 \\
  x_2 \\
  \vdots \\
  x_{NL}
\end{bmatrix}_{NL \times 1},
\] (6.5)

In principle, we can solve (6.5) using Fusion of Algorithms for Compressed Sensing (FACS) with sparsity level \( LK \). Note that, after stacking \( X \) column-wise, we lost the joint sparsity constraint
imposed on $X$ in the MMV problem in (6.3). The $LK$ non-zero elements estimated from (6.5) using FACS can be from more than $K$ different rows of $X$. In the worst case, the estimate of FACS may include non-zero elements from $\min(LK, M)$ different rows of $X$. Then we will end up with an estimate of $X$ with $LK$ non-zero rows, which is highly undesirable. Hence stacking the columns of the observation matrix $B$ and solving it using FACS is not advisable. Note that Step 3 in Algorithm 6.1 ensures that MMV-FACS estimates only $K$ non-zero rows of $X$.

### 6.3 Theoretical Studies of MMV-FACS

In this section, we will theoretically analyse the performance of MMV-FACS. We consider the general case for an arbitrary signal matrix. We also study the average case performance of MMV-FACS subsequently.

The performance analysis is characterized by Signal-to-Reconstruction-Error Ratio (SRER) extended for MMV which is defined as

$$\text{SRER} \triangleq \frac{\|X\|_F^2}{\|X - \hat{X}\|_F^2},$$

where $X$ and $\hat{X}$ denote the actual and reconstructed signal matrix respectively.

**Lemma 6.1.** Suppose that $A$ satisfies the relation, for some constant $\delta_{R+K} \in (0, 1)$,

$$\|AX\|_F \leq \sqrt{1 + \delta_{R+K}}\|X\|_F,$$

where $\|X\|_0 \leq R + K$ and $\delta_{R+K} \in (0, 1)$. Here $\|X\|_0$ denotes the number of non-zero rows of the matrix $X$. Then, for every matrix
Chapter 6 Fusion of Algorithms for Multiple Measurement Vectors

\[ \|AX\|_F \leq \sqrt{1 + \delta_{R+K}} \left[ \|X\|_F + \frac{1}{\sqrt{R + K}} \|X\|_{2,1} \right] \]

**Proof:** Proof is given in Appendix 6.A on page 140.

**Lemma 6.2.** Consider \( A \in \mathbb{R}^{M \times N} \) and let \( T_1 \) & \( T_2 \) be two subsets of \( \{1, 2, \ldots, N\} \) such that \( T_1 \cap T_2 = \emptyset \). Assume that \( \delta_{|T_1|+|T_2|} \leq 1 \) and let \( Y \) be any matrix, such that \( \text{span}(Y) \in \text{span}(A_{T_1}) \) and \( R = Y - A_{T_2}A_{T_2}^\dagger Y \). Then we have

\[
\left(1 - \frac{\delta_{|T_1|+|T_2|}}{1 - \delta_{|T_1|+|T_2|}}\right) \|Y\|_2 \leq \|R\|_2 \leq \|Y\|_2.
\]

**Proof:** Proof is given in Appendix 6.B on page 142.

### 6.3.1 Performance Analysis for Arbitrary Signals under Measurement Perturbations

We analyse the performance of MMV-FACS for arbitrary signals and give an upper bound on the reconstruction error in Theorem 6.1. We also derive a sufficient condition to get an improved performance of MMV-FACS scheme over any given participating algorithm.

**Theorem 6.1.** Let \( X \) be an arbitrary signal with \( T = \text{supp}(X^K) \). Consider the MMV-FACS setup discussed in Section 6.2 on page 116, and assume that the measurement matrix \( A \) satisfies Restricted Isometry Property (RIP) with Restricted Isometry Constant (RIC) \( \delta_{R+K} \). We have the following results:
Chapter 6 Fusion of Algorithms for Multiple Measurement Vectors

i) Upper bound on reconstruction error: We have,

\[ \| \mathbf{X} - \hat{\mathbf{X}} \|_F \leq C_1 \| \mathbf{X} - \mathbf{X}^K \|_F + C_2 \| \mathbf{X} - \mathbf{X}^K \|_{2,1} + C_3 \| \mathbf{X} \Gamma_c, : \|_F + \nu \| \mathbf{W} \|_F \]

where \( C_1 = \left( 1 + \nu \sqrt{1 + \delta R + K} \right) \), \( C_2 = \frac{\nu \sqrt{1 + \delta R + K}}{\sqrt{R + K}} \), \( C_3 = \frac{1 + \delta R + K}{(1 - \delta R + K)^2} \), and \( \nu = \frac{3 - \delta R + K}{(1 - \delta R + K)^2} \).

ii) SRER gain:

For \( \| \mathbf{X} \Gamma_{c, i} : \|_F \neq 0 \) and \( \| \mathbf{X}_{\Gamma c, :} \|_F \neq 0 \), MMV-FACS provides at least SRER gain of

\[ \left( \frac{(1 - \delta R + K)^2}{(1 + \delta R + K + 3\zeta + 3\xi)\eta_i} \right)^2 \]

over the \( i^{th} \) participating algorithm if

\[ \eta_i < \frac{(1 - \delta R + K)^2}{(1 + \delta R + K + 3\zeta + 3\xi)} \],

where \( \eta_i = \frac{\| \mathbf{X}_{\Gamma_{c, i}} : \|_F}{\| \mathbf{X}_{\hat{\Gamma}_{c, i}} : \|_F} \), \( \zeta = \frac{\| \mathbf{W} \|_F}{\| \mathbf{X}_{\Gamma_{c, i}} : \|_F} \),

and

\[ \xi = \left( 3\sqrt{1 + \delta R + K} + 1 \right) \frac{\| \mathbf{X} - \mathbf{X}^K \|_F}{3 \| \mathbf{X}_{\Gamma c, :} \|_F} + \sqrt{\frac{1 + \delta R + K}{R + K}} \frac{\| \mathbf{X} - \mathbf{X}^K \|_{2,1}}{\| \mathbf{X}_{\Gamma c, :} \|_F} \].

Proof:

i) We have,

\[ \| \mathbf{X} - \hat{\mathbf{X}} \|_F \leq \| \mathbf{X} - \mathbf{X}^K \|_F + \| \mathbf{X}^K - \hat{\mathbf{X}} \|_F \]  \hspace{1cm} (6.7)

Consider,

\[ \| \mathbf{X}^K - \hat{\mathbf{X}} \|_F \leq \| (\mathbf{X}^K)_{\hat{T}_{c, i}} : - \hat{\mathbf{X}}_{\hat{T}_{c, i}} : \|_F + \| (\mathbf{X}^K)_{\hat{T}_{c, i}} : - \hat{\mathbf{X}}_{\hat{T}_{c, i}} : \|_F \\
\leq \| (\mathbf{X}^K)_{\hat{T}_{c, i}} : - \hat{\mathbf{X}}_{\hat{T}_{c, i}} : \|_F + \| (\mathbf{X}^K)_{\hat{T}_{c, i}} : \|_F \]  \hspace{1cm} (\because \hat{\mathbf{X}}_{\hat{T}_{c, i}} = 0)

(6.8)
Chapter 6 Fusion of Algorithms for Multiple Measurement Vectors

Using the relations \( \hat{X}_{\hat{T}} = A_{\hat{T}}^\dagger B \) (from Algorithm 6.1) and \( A_{\hat{T}}^\dagger A_{\hat{T}} = I \), we get

\[
\| (X^K)_{\hat{T}} : - \hat{X}_{\hat{T}} \|_F \\
= \| (X^K)_{\hat{T}} : - A_{\hat{T}}^\dagger B \|_F \\
= \| (X^K)_{\hat{T}} : - A_{\hat{T}}^\dagger (AX + W) \|_F \quad (\because B = AX + W) \\
= \| (X^K)_{\hat{T}} : - A_{\hat{T}}^\dagger (AX^K + A(X - X^K) + W) \|_F \\
= \| A_{\hat{T}}^\dagger A_{\hat{T}}c (X^K)_{\hat{T}c} : + A_{\hat{T}}^\dagger A(X - X^K) + A_{\hat{T}}^\dagger W \|_F \\
\leq \| (A_{\hat{T}}^H A_{\hat{T}})^{-1} A_{\hat{T}}^H A_{\hat{T}c} (X^K)_{\hat{T}c} : \|_F + \| A_{\hat{T}}^\dagger A(X - X^K) \|_F + \| A_{\hat{T}}^\dagger W \|_F \\
(6.9)
\]

Let \( x^{(i)} \) denote the \( i \)th column of matrix \( X \) and \( w^{(i)} \) denote the \( i \)th column of matrix \( W \), \( i = 1, 2, \ldots L \). Now from Proposition 2.1 on page 28 and Corollary 2.1 on page 29 we obtain the following relations.

\[
\| A_{\hat{T}}^\dagger w^{(i)} \|_2 \leq \frac{\| w^{(i)} \|_2}{\sqrt{1 - \delta_{R+K}}} \quad (6.10)
\]

\[
\| A_{\hat{T}}^\dagger A \left( x^{(i)} - (x^{(i)})^K \right) \|_2 \leq \frac{\| A \left( x^{(i)} - (x^{(i)})^K \right) \|_2}{\sqrt{1 - \delta_{R+K}}} \quad (6.11)
\]

\[
\| (A_{\hat{T}}^H A_{\hat{T}})^{-1} A_{\hat{T}}^H A_{\hat{T}c} \left( (x^{(i)})^K \right)_{\hat{T}c} \|_2 \leq \frac{\delta_{R+K}}{1 - \delta_{R+K}} \| (x^{(i)})^K \|_{\hat{T}c} \quad (6.12)
\]
Consider (6.10), we get

$$\left\| A^\dagger T w^{(i)} \right\|_2^2 \leq \frac{\left\| w^{(i)} \right\|_2^2}{1 - \delta_{R+K}} \quad \forall \ i = 1, 2, \ldots L$$

Summing the above equation over \( i = 1, 2, \ldots L \), we obtain

$$\sum_{i=1}^{L} \left\| A^\dagger T w^{(i)} \right\|_2^2 \leq \frac{1}{1 - \delta_{R+K}} \sum_{i=1}^{L} \left\| w^{(i)} \right\|_2^2$$

$$\left\| A^\dagger T W \right\|_F \leq \frac{1}{1 - \delta_{R+K}} \left\| W \right\|_F.$$ \( (6.13) \)

Similarly, summing the relations in (6.11) and (6.12), we obtain

$$\left\| A^\dagger T A (X - X^K) \right\|_F \leq \frac{\left\| A (X - X^K) \right\|_F}{\sqrt{1 - \delta_{R+K}}}$$ \( (6.14) \)

$$\left\| (A^H_T A_T)^{-1} A^H_T A_T c (X^K)_{T^c} : \right\|_F \leq \frac{\delta_{R+K}}{1 - \delta_{R+K}} \left\| (X^K)_{T^c} : \right\|_F.$$ \( (6.15) \)

Substituting (6.13), (6.14) and (6.15) in (6.9), we get

$$\left\| (X^K)_{T^c} : - \hat{X}_{T^c} : \right\|_F \leq \frac{\delta_{R+K}}{1 - \delta_{R+K}} \left\| (X^K)_{T^c} : \right\|_F + \frac{\left\| A (X - X^K) \right\|_F}{\sqrt{1 - \delta_{R+K}}} + \frac{\left\| W \right\|_F}{\sqrt{1 - \delta_{R+K}}}$$

$$\leq \frac{\delta_{R+K}}{1 - \delta_{R+K}} \left\| (X^K)_{T^c} : \right\|_F + \frac{1}{1 - \delta_{R+K}} \left( \left\| A (X - X^K) \right\|_F + \left\| W \right\|_F \right)$$ \( (6.16) \)

Substituting (6.16) in (6.8), we get

$$\left\| X^K - \hat{X} \right\|_F \leq \frac{1}{1 - \delta_{R+K}} \left\| (X^K)_{T^c} : \right\|_F + \frac{1}{1 - \delta_{R+K}} \left( \left\| A (X - X^K) \right\|_F + \left\| W \right\|_F \right)$$ \( (6.17) \)
Next, we will find an upper bound for \( \| (X^K)_{\hat{T}_c} \|_F \).

Define \( \hat{T}_\Delta \triangleq \Gamma \setminus \hat{T} \). That is, \( \hat{T}_\Delta \) is the set formed by the atoms in \( \Gamma \) which are discarded by Algorithm 6.1. Since \( \hat{T} \subset \Gamma \), we have \( \hat{T}^c = \Gamma^c \cup \hat{T}_\Delta \) and hence we obtain

\[
\| (X^K)_{\hat{T}_c} \|_F \leq \| (X^K)_{\hat{T}^c} \|_F + \| (X^K)_{\hat{T}_\Delta} \|_F \tag{6.18}
\]

We also have,

\[
\| (X^K)_{\hat{T}_\Delta} \|_F \leq \| (V_{\Gamma^c})_{\hat{T}_\Delta} \|_F + \| (V_{\Gamma^c} - (X^K)_{\hat{T}_\Delta}) \|_F \\
\leq \| (V_{\Gamma^c})_{\hat{T}_\Delta} \|_F + \| V_{\Gamma^c} - (X^K)_{\hat{T}_\Delta} \|_F \tag{6.19}
\]

Note that \( (V_{\Gamma^c})_{\hat{T}_\Delta} \) contains the \( K \)-rows of \( V_{\Gamma^c} \) with highest row \( \ell_2 \)-norm. Therefore, using \( |\hat{T}| = |\mathcal{T}| = K \), we get

\[
\| (V_{\Gamma^c})_{\hat{T}_\Delta} \|_F \leq \| (V_{\Gamma^c})_{\Gamma \setminus \hat{T}_\Delta} \|_F \\
= \| V_{\Gamma \setminus \hat{T}_\Delta} - (X^K)_{\Gamma \setminus \hat{T}_\Delta} \|_F \quad (\because (X^K)_{\Gamma \setminus \hat{T}_\Delta} = 0) \\
\leq \| (V_{\Gamma^c} - (X^K)_{\hat{T}_\Delta}) \|_F \tag{6.20}
\]

Substituting (6.20) in (6.19), we get

\[
\| (X^K)_{\hat{T}_\Delta} \|_F \leq 2 \| (V_{\Gamma^c} - (X^K)_{\hat{T}_\Delta}) \|_F \tag{6.21}
\]

Now, consider

\[
\| V_{\Gamma^c} - (X^K)_{\Gamma^c} \|_F \\
= \| A_{\Gamma^c}B - (X^K)_{\Gamma^c} \|_F \\
= \| A_{\Gamma^c}(AX + W) - (X^K)_{\Gamma^c} \|_F \\
= \| A_{\Gamma^c}((AX^K)_{\Gamma^c} + A(X - X^K) + W) - (X^K)_{\Gamma^c} \|_F \\
= \| A_{\Gamma^c}(A_{\Gamma^c}X_{\Gamma^c} + A_{\Gamma^c}(X_{\Gamma^c})_{\Gamma^c} + A(X - X^K) + W) - (X^K)_{\Gamma^c} \|_F
\]
\[
\begin{align*}
\mathbf{V}_{\Gamma, :} - (\mathbf{X}^K)_{\Gamma, :} & \leq \frac{\delta_{R+K}}{1 - \delta_{R+K}} \| (\mathbf{X}^K)_{\Gamma, :} \|_F + \frac{\| (\mathbf{X}^K - \mathbf{X}^K) \|_F}{\sqrt{1 - \delta_{R+K}}} + \frac{\| \mathbf{W} \|_F}{\sqrt{1 - \delta_{R+K}}} \\
\end{align*}
\]

\[
(\therefore 0 < 1 - \delta_{R+K} < 1)
\]

Using (6.21) and (6.23) in (6.18), we get

\[
\| (\mathbf{X}^K)_{\hat{\Gamma}, :} \|_F \leq \frac{1 + \delta_{R+K}}{1 - \delta_{R+K}} \| (\mathbf{X}^K)_{\Gamma, :} \|_F + \frac{2}{1 - \delta_{R+K}} (\| (\mathbf{X}^K - \mathbf{X}^K) \|_F + \| \mathbf{W} \|_F) 
\]

\[
(6.24)
\]

Let \( \mathbf{x}_{1}^{(i)} \) denote the \( i \)th column of matrix \( \mathbf{X}^K \). The, we have,

\[
\| \mathbf{A} \mathbf{x}_{1}^{(i)} \|_2^2 \leq (1 + \delta_{R+K}) \| \mathbf{x}_{1}^{(i)} \|_2^2 \quad (\therefore \mathbf{A} \text{ satisfies RIP})
\]

\[
\sum_{i=1}^{L} \| \mathbf{A} \mathbf{x}_{1}^{(i)} \|_2^2 \leq \sum_{i=1}^{L} (1 + \delta_{R+K}) \| \mathbf{x}_{1}^{(i)} \|_2^2
\]

\[
\| \mathbf{A} \mathbf{X}^K \|_F^2 \leq (1 + \delta_{R+K}) \| \mathbf{X}^K \|_F^2
\]

\[
(6.25)
\]

Using Lemma 6.1 and (6.25), we get

\[
\| \mathbf{A} (\mathbf{X} - \mathbf{X}^K) \|_F \leq \sqrt{1 + \delta_{R+K}} \left[ \| \mathbf{X} - \mathbf{X}^K \|_F + \frac{1}{\sqrt{R + K}} \| (\mathbf{X} - \mathbf{X}^K) \|_{2,1} \right]
\]

\[
(6.26)
\]
Substituting (6.24) in (6.17), we get

\[
\|X^K - \hat{X}\|_F \leq \frac{1 + \delta_{R+K}}{(1 - \delta_{R+K})^2} \|(X^K)_{\Gamma^c,:}\|_F \\
+ \frac{3 - \delta_{R+K}}{(1 - \delta_{R+K})^2} \left[ \|A(X - X^K)\|_F + \|W\|_F \right] \\
\leq \nu \sqrt{1 + \delta_{R+K}} \|X - X^K\|_F + \nu \frac{\sqrt{1 + \delta_{R+K}} \|X - X^K\|_{2,1}}{\sqrt{R + K}} \\
+ \frac{1 + \delta_{R+K}}{(1 - \delta_{R+K})^2} \|X_{\Gamma^c,:}\|_F + \nu \|W\|_F, \quad \text{(using (6.26))}
\]

(6.27)

where \(\nu = \frac{3 - \delta_{R+K}}{(1 - \delta_{R+K})^2}\).

Substituting (6.27) in (6.7) and using the definitions of \(C_1, C_2,\) and \(C_3\), we get

\[
\|X - \hat{X}\|_F \leq C_1 \|X - X^K\|_F + C_2 \|X - X^K\|_{2,1} + C_3 \|X_{\Gamma^c,:}\|_F + \nu \|W\|_F.
\]

(6.28)

ii) Using (6.28) and the definitions of \(\xi\) and \(\eta_i\), we get

\[
\|X - \hat{X}\|_F \leq \frac{1 + \delta_{R+K} + 3\zeta + 3\xi}{(1 - \delta_{R+K})^2} \|X_{\Gamma^c,:}\|_F \\
= \frac{1 + \delta_{R+K} + 3\zeta + 3\xi}{(1 - \delta_{R+K})^2} \eta_i \|X - \hat{X}_i\|_F^c, (\therefore (\hat{X}_i)_{\hat{\Gamma}^c,:} = 0) \\
\leq \frac{1 + \delta_{R+K} + 3\zeta + 3\xi}{(1 - \delta_{R+K})^2} \eta_i \|X - \hat{X}_i\|_F.
\]

Hence, we obtain the relation for SRER for MMV-FACS, in case of arbitrary signals, as

\[
\text{SRER}_{\text{MMV-FACS}} = \frac{\|X\|^2_F}{\|X - \hat{X}\|^2_F}
\]

125
Chapter 6 Fusion of Algorithms for Multiple Measurement Vectors

\[
\geq \frac{\|X\|_F^2}{\|X - \hat{X}_i\|_F^2} \times \left(\frac{(1 - \delta_{R+K})^2}{(1 + \delta_{R+K} + 3\zeta + 3\xi)\eta_i}\right)^2
\]

Hence MMV-FACS provides at least SRER gain of \(\left(\frac{(1 - \delta_{R+K})^2}{(1 + \delta_{R+K} + 3\zeta + 3\xi)\eta_i}\right)^2\)
over \(i^{th}\) algorithm if \(\eta_i < \frac{(1 - \delta_{R+K})^2}{(1 + \delta_{R+K} + 3\zeta + 3\xi)}\).

Note that \(\frac{(1 - \delta_{R+K})^2}{(1 + \delta_{R+K} + 3\zeta + 3\xi)} < 1\).

6.3.2 Exactly \(K\)-sparse Matrix

Theorem 6.1 considered the case when \(X\) is an arbitrary matrix. If \(X\) is a \(K\)-sparse matrix then we have \(X = X^K\) and \(\zeta = 0\). Thus, it follows from Theorem 6.1 that, MMV-FACS provides at least SRER gain of \(\left(\frac{(1 - \delta_{R+K})^2}{(1 + \delta_{R+K} + 3\zeta + 3\xi)\eta_i}\right)^2\) over \(i^{th}\) participating algorithm if \(\eta_i < \frac{(1 - \delta_{R+K})^2}{(1 + \delta_{R+K} + 3\zeta + 3\xi)}\). Thus, the improvement in the SRER gain provided by MMV-FACS over the \(i^{th}\) Algorithm for a \(K\)-sparse matrix is greater than that of an arbitrary matrix by a factor of \(\left(1 + \frac{3\xi}{(1 + \delta_{R+K} + 3\zeta)}\right)^2\).

The second part of Theorem 6.1 considers the case when \(\|X_{\hat{T}_i,:}\|_F \neq 0\) and \(\|X_{\Gamma,:}\|_F \neq 0\). If \(\|X_{T_i,:}\|_F \neq 0\), then \(\hat{T}_i \notin \mathcal{T}\). Also, \(\|X_{\Gamma,:}\|_F = 0\) implies \(\mathcal{T} \subseteq \Gamma\). Suppose \(\|X_{T_i,:}\|_F = 0\), then the support-set is correctly estimated by \(i^{th}\) algorithm and further performance improvement is not possible by MMV-FACS. Hence we consider the case where \(\|X_{\Gamma,:}\|_F = 0\), and derive the condition for exact reconstruction by MMV-FACS in the following proposition.
Chapter 6 Fusion of Algorithms for Multiple Measurement Vectors

**Proposition 6.1.** Assume that $\|X_{\Gamma^c,:}\|_F = 0$ and all other conditions in Theorem 6.1 hold good. Then, in clean measurement case ($W = 0$), MMV-FACS estimates the support-set correctly and provides exact reconstruction.

**Proof:** We have

\[ X_{\Gamma^c,:} = 0 \Rightarrow \mathcal{T} \subset \Gamma \quad (6.29) \]
\[ B = A_TX_{\hat{\mathcal{T}},:} + W \quad (6.30) \]

From Algorithm 6.1, we have $V \in \mathbb{R}^{N \times L}$ where $V_{\Gamma^c,:} = 0$, and

\[ V_{\Gamma,:) = A_T^*B \]
\[ = A_T^*(A_TX_{\hat{T},:} + W) \quad \text{(using (6.30))} \]
\[ = A_T^*(A_{\Gamma}X_{\Gamma,:} + W) \quad \text{(using (6.29))} \]
\[ = X_{\Gamma,:} + A_T^*W. \]

If $W = 0$, then $V_{\Gamma,:) = X_{\Gamma,:}$ and $V = X$ ($\therefore \mathcal{T} \subset \Gamma$). Thus MMV-FACS estimates the support-set correctly from $V$. $\blacksquare$

In practice, the original signal is not known and hence it is not possible to evaluate the performance w.r.t. the true signal. Hence in applications, the decrease in energy of the residual is often treated as a measure of performance improvement. Proposition 6.2 gives a sufficient condition for decrease in the energy of the residual matrix obtained by MMV-FACS over the $i^{\text{th}}$ participating algorithm.

**Proposition 6.2.** For a $K$-sparse matrix $X$, let $R = B - A_T^*A_T^iB$ and $R_i = B - A_{\hat{T}_i}^*A_{\hat{T}_i}^iB$ represent the residue matrix of MMV-FACS and $i^{\text{th}}$ Algorithm respectively. Assume that

\[ \frac{\sqrt{1+\delta_{R+K}}}{1-\delta_{R+K}}(1 + \delta_{R+K} + \frac{3\zeta}{\eta_i \sqrt{1 - \delta_{R+K}}} - \zeta) \leq \frac{1 - 2\delta_{R+K}}{\eta_i \sqrt{1 - \delta_{R+K}}} \]

is satisfied then we have, $\|R\|_F \leq \|R_i\|_F$. 127
Using (6.24) we have,

\[ \|R\|_F = \left\| \mathbf{B} - \mathbf{A}_\mathcal{T} \mathbf{A}_\mathcal{T}^\dagger \mathbf{B} \right\|_F \]
\[ = \left\| \mathbf{A} \mathbf{X} + \mathbf{W} - \mathbf{A}_\mathcal{T} \mathbf{A}_\mathcal{T}^\dagger (\mathbf{A} \mathbf{X} + \mathbf{W}) \right\|_F \quad (\because \mathbf{B} = \mathbf{A} \mathbf{X} + \mathbf{W}) \]
\[ = \left\| \mathbf{A}_\mathcal{T} \mathbf{X}_\mathcal{T}: + \mathbf{A}_\mathcal{T} \mathbf{X}_\mathcal{T}^c: + \mathbf{W} - \mathbf{A}_\mathcal{T} \mathbf{A}_\mathcal{T}^\dagger (\mathbf{A}_\mathcal{T} \mathbf{X}_\mathcal{T}: + \mathbf{A}_\mathcal{T} \mathbf{X}_\mathcal{T}^c: + \mathbf{W}) \right\|_F \]
\[ = \left\| \mathbf{A}_\mathcal{T} \mathbf{X}_\mathcal{T}^c: - \mathbf{A}_\mathcal{T} \mathbf{A}_\mathcal{T}^\dagger \mathbf{A}_\mathcal{T} \mathbf{X}_\mathcal{T}^c: + \mathbf{W} - \mathbf{A}_\mathcal{T} \mathbf{A}_\mathcal{T}^\dagger \mathbf{W} \right\|_F \quad (\because \mathbf{A}_\mathcal{T} \mathbf{A}_\mathcal{T}^\dagger = \mathbf{I}) \]
\[ \leq \left\| \mathbf{A}_\mathcal{T} \mathbf{X}_\mathcal{T}^c: - \mathbf{A}_\mathcal{T} \mathbf{A}_\mathcal{T}^\dagger \mathbf{A}_\mathcal{T} \mathbf{X}_\mathcal{T}^c: \right\|_F + \left\| \mathbf{W} - \mathbf{A}_\mathcal{T} \mathbf{A}_\mathcal{T}^\dagger \mathbf{W} \right\|_F \]
\[ \leq \left\| \mathbf{A}_\mathcal{T} \mathbf{X}_\mathcal{T}^c: \right\|_F + \left\| \mathbf{W} \right\|_F \quad (\because \mathcal{T} = \supp(\mathbf{X})) \]
\[ \leq \sqrt{1 + \delta_{R+K}} \left\| \mathbf{X}_\mathcal{T}^c: \right\|_F + \left\| \mathbf{W} \right\|_F \quad (\because |\mathcal{T} \setminus \hat{\mathcal{T}}| \leq K \quad \& \quad \delta_K \leq \delta_{R+K}) \].

Using (6.24) we have,

\[ \|\mathbf{R}\|_F \leq \sqrt{1 + \delta_{R+K}} \left( \frac{1 + \delta_{R+K}}{1 - \delta_{R+K}} \|\mathbf{X}_{\mathcal{T}^c}:\|_F + \frac{2}{1 - \delta_{R+K}} \|\mathbf{W}\|_F \right) + \|\mathbf{W}\|_F \]
\[ \leq \frac{\sqrt{1 + \delta_{R+K}}}{1 - \delta_{R+K}} \left( 1 + \delta_{R+K} + 2\zeta + \frac{1 - \delta_{R+K}}{\sqrt{1 + \delta_{R+K}} \zeta} \right) \|\mathbf{X}_{\mathcal{T}^c}:\|_F \]
\[ \leq \frac{\sqrt{1 + \delta_{R+K}}}{1 - \delta_{R+K}} (1 + \delta_{R+K} + 3\zeta) \|\mathbf{X}_{\mathcal{T}^c}:\|_F. \quad (6.31) \]

Now, consider

\[ \|\mathbf{R}_i\|_F = \left\| \mathbf{B} - \mathbf{A}_{\mathcal{T}_i} \mathbf{A}_{\mathcal{T}_i}^\dagger \mathbf{B} \right\|_F \]
\[ = \left\| \mathbf{A} \mathbf{X} + \mathbf{W} - \mathbf{A}_{\mathcal{T}_i} \mathbf{A}_{\mathcal{T}_i}^\dagger (\mathbf{A} \mathbf{X} + \mathbf{W}) \right\|_F \quad (\because \mathbf{B} = \mathbf{A} \mathbf{X} + \mathbf{W}) \]
\[ = \left\| \mathbf{A}_{\mathcal{T}_i} \mathbf{X}_{\mathcal{T}_i}: + \mathbf{A}_{\mathcal{T}_i} \mathbf{X}_{\mathcal{T}_i}^c: + \mathbf{W} - \mathbf{A}_{\mathcal{T}_i} \mathbf{A}_{\mathcal{T}_i}^\dagger (\mathbf{A}_{\mathcal{T}_i} \mathbf{X}_{\mathcal{T}_i}: + \mathbf{A}_{\mathcal{T}_i} \mathbf{X}_{\mathcal{T}_i}^c: + \mathbf{W}) \right\|_F \]
\[ = \left\| \mathbf{A}_{\mathcal{T}_i} \mathbf{X}_{\mathcal{T}_i}^c: + \mathbf{W} - \mathbf{A}_{\mathcal{T}_i} \mathbf{A}_{\mathcal{T}_i}^\dagger \mathbf{A}_{\mathcal{T}_i} \mathbf{X}_{\mathcal{T}_i}^c: - \mathbf{A}_{\mathcal{T}_i} \mathbf{A}_{\mathcal{T}_i}^\dagger \mathbf{W} \right\|_F \]
\[ \geq \left\| \mathbf{A}_{\mathcal{T}_i} \mathbf{X}_{\mathcal{T}_i}^c: - \mathbf{A}_{\mathcal{T}_i} \mathbf{A}_{\mathcal{T}_i}^\dagger \mathbf{A}_{\mathcal{T}_i} \mathbf{X}_{\mathcal{T}_i}^c: \right\|_F - \left\| \mathbf{W} - \mathbf{A}_{\mathcal{T}_i} \mathbf{A}_{\mathcal{T}_i}^\dagger \mathbf{W} \right\|_F \]

(Using reverse triangle inequality)
= \left\| \mathbf{A}_{T \setminus \hat{T}_i} \mathbf{X}_{T \setminus \hat{T}_i} - \mathbf{A}_{\hat{T}_i} \mathbf{A}_{T \setminus \hat{T}_i} \mathbf{X}_{T \setminus \hat{T}_i} \right\|_F - \left\| \mathbf{W} - \mathbf{A}_{\hat{T}_i} \mathbf{A}_{T \setminus \hat{T}_i} \mathbf{W} \right\|_F \\
\geq \left( 1 - \frac{\delta_{R+K}}{1 - \delta_{R+K}} \right) \left\| \mathbf{A}_{T \setminus \hat{T}_i} \mathbf{X}_{T \setminus \hat{T}_i} \right\|_F - \left\| \mathbf{W} \right\|_F \\
(\text{Using Lemma 6.2} \& \delta_{2K} \leq \delta_{R+K}) \\
\geq \frac{1 - 2\delta_{R+K}}{1 - \delta_{R+K}} \sqrt{1 - \delta_{R+K}} \left\| \mathbf{X}_{\hat{T}_i,c} \right\|_F - \left\| \mathbf{W} \right\|_F \\
(: \cdot \left| T \setminus \hat{T} \right| \leq K \ \& \ \delta_K \leq \delta_{R+K}) \\
= \left( \frac{1 - 2\delta_{R+K}}{\eta_i \sqrt{1 - \delta_{R+K}} - \zeta} \right) \left\| \mathbf{X}_{\hat{T}_i,c} \right\|_F. \tag{6.32}

From (6.31) and (6.32) we get a sufficient condition for \( \| \mathbf{R} \|_F \leq \left\| \mathbf{R}_i \right\|_F \) as

\[
\frac{\sqrt{1 + \delta_{R+K}}}{1 - \delta_{R+K}} (1 + \delta_{R+K} + 3\zeta) \leq \left( \frac{1 - 2\delta_{R+K}}{\eta_i \sqrt{1 - \delta_{R+K}} - \zeta} \right) . \tag{6.33}
\]

Thus, if (6.33) is satisfied, MMV-FACS produces a smaller residual matrix (in the Frobenius norm sense) than that of the \( i \)th participating algorithm.

6.3.3 Average Case Analysis

Intuitively, we expect multiple measurement vector problem to perform better than the single measurement vector case. However, if each measurement vector is the same, i.e., in the worst case, we have \( \mathbf{x}^{(i)} = \mathbf{c}, \ \forall \ i = 1, \ldots, L \), then we do not have any additional information on \( \mathbf{X} \) than that provided by a single measurement vector \( \mathbf{x}^{(1)} \). So far we have carried out only the worst case analysis, i.e., conditions under which the algorithm is able to recover any joint sparse matrix \( \mathbf{X} \). This approach does not provide insight into the superiority of sparse signal reconstruction with multiple measurement vectors compared to the single measurement vector case.
Chapter 6 Fusion of Algorithms for Multiple Measurement Vectors

To notice a performance gain with multiple measurement vectors, next we proceed with an average case analysis. Here we impose a probability model on the $K$ sparse $X$ as suggested by Remi et al. [162]. In particular, on the support-set $\mathcal{T}$, we impose that $X_{\mathcal{T}} = \Sigma \Phi$, where $\Sigma$ is a $K \times K$ diagonal matrix with positive diagonal entries and $\Phi$ is a $K \times L$ random matrix with independently and identically distributed (i.i.d.) Gaussian entries. Our goal is to show that, under this signal model, the typical behavior of MMV–FACS is better than in the worst case.

**Theorem 6.2.** Consider the MMV-FACS setup discussed in Section 6.2. Assume a Gaussian signal model, i.e., $X_{\mathcal{T}} = \Sigma \Phi$, where $\Sigma$ is a $K \times K$ diagonal matrix with positive diagonal entries and $\Phi$ is a $K \times L$ random matrix with i.i.d. Gaussian entries. Let $e_i$ denote a $|\Gamma| \times 1$ vector with a ‘1’ in the $i$th coordinate and ‘0’ elsewhere. Let

$$\eta = \min_{i \in (\mathcal{T} \cap \Gamma)} \left\| e_i^T A^\dagger_\Gamma W \right\|_2 + \max_{j \in (\Gamma \setminus \mathcal{T})} \left\| e_j^T A^\dagger_\Gamma W \right\|_2$$(1)

and

$$\gamma = \frac{\min_{i \in (\mathcal{T} \cap \Gamma)} \left\| e_i^T A^\dagger_\Gamma A_\Gamma \Sigma \right\|_2 - \max_{j \in (\Gamma \setminus \mathcal{T})} \left\| e_j^T A^\dagger_\Gamma A_\Gamma \Sigma \right\|_2 - \eta}{C_2(L)}.$$ (2)

where $C_2(L) = \mathbb{E} \left\| Z \right\|_2$ with $Z = (Z_1, \ldots, Z_L)$ being a vector of independent standard normal variables. Assume that

$$\min_{i \in (\mathcal{T} \cap \Gamma)} \left\| e_i^T A^\dagger_\Gamma A_\Gamma \Sigma \right\|_2 - \max_{j \in (\Gamma \setminus \mathcal{T})} \left\| e_j^T A^\dagger_\Gamma A_\Gamma \Sigma \right\|_2 \geq \frac{\eta}{C_2(L)}.$$ (3)

Let $\Theta$ denote the event that MMV–FACS picks all correct indices from the union-set $\Gamma$. Then, we have

$$P(\Theta) \geq 1 - K \exp(-2A_2(L)\gamma^2),$$ (4)

where $A_2(L) = \left( \frac{\hat{\Gamma}(L+1)}{\hat{\Gamma}(L/2)} \right)^2 \approx \frac{L}{2}$. $\hat{\Gamma}(\cdot)$ denotes the Gamma function.
Influenced by the concentration of measure results in [162], we set
\[ \min_{i \in (\mathcal{T} \cap \Gamma)} \left\| e_i^T A_i \right\|^2_2 > \max_{j \in (\mathcal{G} \setminus \mathcal{T})} \left\| e_j^T A_j \right\|^2_2 \]

Using (5.5) in [162], we get,
\[ P(\Theta) = P \left( \min_{i \in (\mathcal{T} \cap \Gamma)} \left\| e_i^T A_i \right\|^2_2 > \max_{j \in (\mathcal{G} \setminus \mathcal{T})} \left\| e_j^T A_j \right\|^2_2 \right) \]
\[ > P \left( \min_{i \in (\mathcal{T} \cap \Gamma)} \left\| e_i^T A_i \right\|^2_2 + \max_{j \in (\mathcal{G} \setminus \mathcal{T})} \left\| e_j^T A_j \right\|^2_2 \right) \]
\[ \geq \max_{j \in (\mathcal{G} \setminus \mathcal{T})} \left( \left\| e_j^T A_j \right\|^2_2 - \left\| e_j^T A_j \right\|^2_2 \right) \]

(Using reverse triangle inequality and triangle inequality respectively)
\[ = P \left( \min_{i \in (\mathcal{T} \cap \Gamma)} \left\| e_i^T A_i \right\|^2_2 \right) \]
\[ > \min_{i \in (\mathcal{T} \cap \Gamma)} \left\| e_i^T A_i \right\|^2_2 + \max_{j \in (\mathcal{G} \setminus \mathcal{T})} \left\| e_j^T A_j \right\|^2_2 \]
\[ = P \left( \min_{i \in (\mathcal{T} \cap \Gamma)} \left\| e_i^T A_i \right\|^2_2 \right) \]
\[ = 1 - P \left( \min_{i \in (\mathcal{T} \cap \Gamma)} \left\| e_i^T A_i \right\|^2_2 \leq C \right) \]
\[ \geq 1 - P \left( \min_{i \in (\mathcal{T} \cap \Gamma)} \left\| e_i^T A_i \right\|^2_2 \geq C - \eta \right) \]  \hspace{1cm} (6.34)

Now, let us derive an upper bound for \( P \left( \min_{i \in (\mathcal{T} \cap \Gamma)} \left\| e_i^T A_i \right\|^2_2 \leq C \right) \).

Influenced by the concentration of measure results in [162], we set
\[ C = (1 - \epsilon_1)C_2(L) \min_{i \in (\mathcal{T} \cap \Gamma)} \left\| e_i^T A_i \right\|^2_2 \]  \hspace{1cm} (6.35)

where \( 0 < \epsilon_1 < 1 \).

Using (5.5) in [162], we get,
\[ P \left( \min_{i \in (\mathcal{T} \cap \Gamma)} \left\| e_i^T A_i \right\|^2_2 \leq C \right) \leq |\mathcal{T}| \exp(-A_2(L)\epsilon_1^2). \]  \hspace{1cm} (6.36)
To bound the second probability, consider

\[
P \left( \max_{j \in (\Gamma \setminus T)} \left\| e_j^T A^\dagger_1 A_\tau \Sigma \Phi \right\|_2 \geq C - \eta \right) \cap 
\left( \max_{j \in (\Gamma \setminus T)} \left\| e_j^T A^\dagger_1 A_\tau \Sigma \right\|_2 \geq (C - \eta) \right) = \left( \max_{j \in (\Gamma \setminus T)} \left\| e_j^T A^\dagger_1 A_\tau \Sigma \Sigma_\Phi \right\|_2 \geq (C - \eta) \right)
\]

Let

\[
1 + \epsilon_2 = \frac{C - \eta}{C_2(L) \max_{j \in (\Gamma \setminus T)} \left\| e_j^T A^\dagger_1 A_\tau \Sigma \right\|_2}.
\]

Using equation (5.3) in [162]

\[
P \left( \max_{j \in (\Gamma \setminus T)} \left\| e_j^T A^\dagger_1 A_\tau \Sigma \Phi \right\|_2 \geq (1 + \epsilon_2) \right) \cap 
\left( \max_{j \in (\Gamma \setminus T)} \left\| e_j^T A^\dagger_1 A_\tau \Sigma \right\|_2 \geq (C - \eta) \right) \leq \left| T \right| \exp \left( -A_2(L) \epsilon_2^3 \right).
\]

For the above inequality to hold, it is required that \( \epsilon_2 > 0 \). By setting \( \epsilon_2 = \epsilon_1 \), and using (6.35) and (6.37), we get

\[
\epsilon_1 = \frac{(1 - \epsilon_1) C_2(L) \min_{i \in (T \cap \Gamma)} \left\| e_i^T A^\dagger_1 A_\tau \Sigma \right\|_2 - \eta}{C_2(L) \max_{j \in (\Gamma \setminus T)} \left\| e_j^T A^\dagger_1 A_\tau \Sigma \right\|_2} - 1.
\]

Now, solving for \( \epsilon_1 \), we get

\[
\epsilon_1 = \frac{\min_{i \in (T \cap \Gamma)} \left\| e_i^T A^\dagger_1 A_\tau \Sigma \right\|_2 - \max_{j \in (\Gamma \setminus T)} \left\| e_j^T A^\dagger_1 A_\tau \Sigma \right\|_2 - \frac{\eta}{C_2(L)}}{\min_{i \in (T \cap \Gamma)} \left\| e_i^T A^\dagger_1 A_\tau \Sigma \right\|_2 + \max_{j \in (\Gamma \setminus T)} \left\| e_j^T A^\dagger_1 A_\tau \Sigma \right\|_2}.
\]

Clearly \( \epsilon_1 < 1 \) and by the assumption in the theorem \( \epsilon_1 > 0 \). Hence we have \( 0 < \epsilon_1 < 1 \). Also, note that \( \gamma = \epsilon_1 \). Substituting (6.36) and
Chapter 6 Fusion of Algorithms for Multiple Measurement Vectors

(6.38) in (6.34), we get

\[
P(\Theta) \geq 1 - K \exp(-2A_2(L)\gamma^2).
\]

Since \( A_2(L) \approx \frac{L}{2} \), the probability that MMV-FACS selects all correct indices from the union set increases as \( L \) increases. Thus, more than one measurement vector improves the performance.

6.4 Numerical Experiments and Results

We conducted numerical experiments using synthetic data and real signals to evaluate the performance of MMV-FACS. The performance is evaluated using Average Signal-to-Reconstruction-Error Ratio (ASRER) which is defined as

\[
\text{ASRER} = \frac{\sum_{j=1}^{n_{\text{trials}}} \|X_j\|_F^2}{\sum_{j=1}^{n_{\text{trials}}} \|X_j - \hat{X}_j\|_F^2}, \quad (6.39)
\]

where \( X_j \) and \( \hat{X}_j \) denote the actual and reconstructed jointly sparse signal matrix in the \( j \)th trial respectively, and \( n_{\text{trials}} \) denotes the total number of trials.

6.4.1 Synthetic Sparse Signals

For noisy measurement simulations, we define the Signal-to-Measurement-Noise Ratio (SMNR) as

\[
\text{SMNR} \triangleq \frac{\mathbb{E}\{\|x^{(i)}\|_2^2\}}{\mathbb{E}\{\|w^{(i)}\|_2^2\}},
\]

133
Chapter 6 Fusion of Algorithms for Multiple Measurement Vectors

where \( \mathbb{E}\{\cdot\} \) denotes the mathematical expectation operator. The simulation set-up is described below.

6.4.1.1 Experimental Setup

Following steps are involved in the simulation:

i) Generate elements of \( A_{M \times N} \) independently from \( \mathcal{N}(0, \frac{1}{M}) \) and normalize each column norm to unity.

ii) Choose \( K \) nonzero locations uniformly at random from the set \( \{1, 2, \ldots, N\} \) and fill those \( K \) rows of \( X \) based on the choice of signal characteristics:

(a) Gaussian sparse signal matrix: Non-zero values independently from \( \mathcal{N}(0, 1) \).

(b) Rademacher sparse signal matrix: Non-zero values are set to +1 or -1 with probability \( \frac{1}{2} \).

Remaining \( N - K \) rows of \( X \) are set to zero.

iii) The MMV measurement matrix \( B \) is computed as \( B = AX + W \), where the columns of \( W \), \( w^{(i)} \)'s are independent and their elements are i.i.d. as Gaussian with variance determined from the specified SMNR.

iv) Apply the MMV sparse recovery method.

v) Repeat steps i-iv, \( S \) times.

vi) Find ASRER using (6.39).

We used M-OMP, M-SP, M-BPDN [113], and M-FOCUSS [160] as the participating algorithms in MMV-FACS. The software code for
M-BPDN was taken from SPGL1 software package [163]. Since M-FOCUSS and M-BPDN algorithms may not yield an exact $K$-sparse solution, we estimate the support-set as the indices of the $K$ rows with largest $\ell_2$ norm. We fixed the sparse signal dimension $N = 500$ and sparsity level $K = 20$ in the simulation the result were calculated by averaging over 1,000 trials ($S = 1,000$). We use an oracle estimator for performance benchmarking. The oracle estimator is aware of the true support-set and finds the non-zero entries of the sparse matrix by solving LS.

The empirical performance of MMV reconstruction algorithms for different values of $M$ is shown in Figure 6.1. The simulation parameters are $L = 20$, $\text{SMNR} = 20$ dB and $X$ is chosen as Gaussian sparse signal matrix. For $M = 35$, MMV-FACS (M-BPDN,M-FOCUSS) gave 10.67 dB and 4.27 dB improvement over M-BPDN and M-FOCUSS respectively.

### 6.4.1.2 Results and Discussions

Figure 6.2 depicts the performance of Rademacher sparse signal matrix for different values of $M$ where we set $L = 20$ and $\text{SMNR} = 20$ dB. We again observe similar performance improvement as in the case of Gaussian sparse signal matrix. For example, for $M = 35$, MMV-FACS(M-OMP,M-BPDN) showed 7.56 dB and 4.32 dB over M-OMP and M-BPDN respectively.

A comparison of MMV reconstruction techniques is shown in Figure 6.3 for Gaussian sparse signal matrix for different values of $L$ where we set $M = 50$ and $\text{SMNR} = 20$ dB. It may be observed that MMV-FACS gave a significant performance improvement over
the participating algorithms. Specifically, MMV-FACS(M-OMP,M-SP) improved the performance by 5.77 dB and 4.94 dB over M-OMP and M-SP respectively.

To show the dependency of recovery performance on SMNR, we conducted simulations for different values of SMNR. Figure 6.4 illustrates the performance for Gaussian sparse signal matrix where \( L = 10 \) and \( M = 45 \). An additional ASRER improvement of 2.51 dB
Chapter 6 Fusion of Algorithms for Multiple Measurement Vectors

![Graphs](image)

**Figure 6.2:** Performance of MMV-FACS, averaged over 1000 trials, for Rademacher sparse signal matrices with SMNR = 20 dB. Sparse signal dimension $N = 500$, sparsity level $K = 20$ and number of measurement vectors $L = 20$.

and 2.08 dB were achieved as compared to M-OMP and M-FOCUSS respectively for SMNR= 10 dB. This shows the robustness of MMV–FACS to noisy measurements.

From the above simulation results it can be seen that MMV–FACS improved the sparse signal recovery compared to participating algorithms.
6.4.1.3 Reproducible Research

We provide necessary Matlab codes to reproduce all the figures at http://www.ece.iisc.ernet.in/~ssplab/Public/MMVFACS.tar.gz.

6.4.2 Real Compressible Signals

To evaluate the performance of MMV-FACS on compressible signals and real world data, we used the data set ‘05091.dat’ from MIT-BIH Atrial Fibrillation Database [164]. The recording is of 10 hours in duration, and contains two ECG signals each sampled at 250 samples per second with 12-bit resolution over a range of ±10 millivolts. We selected the first 250 time points of the recording as the data set used in our experiment. We used a randomly generated Gaussian sensing matrix of size $M \times 250$, with different values of $M$ in the experiment. We assumed sparsity level $K = 50$ and
used M-OMP and M-SP as the participating algorithms. The reconstruction results are shown in Figure 6.5.

Similar to synthetic signals, MMV-FACS shows a better ASRER compared to the participating algorithms M-OMP and M-SP. This demonstrates the advantage of MMV-FACS in real-life applications, requiring fewer measurement samples to yield an approximate reconstruction.
6.5 Summary

In this chapter, we extended FACS to the MMV case and showed that MMV-FACS improves sparse signal matrix reconstruction. Using RIP, we theoretically analysed the proposed scheme and derived sufficient conditions for the performance improvement over the participating algorithm. Using Monte-Carlo simulations, we showed the performance improvement of the proposed scheme over the participating methods. Though this chapter discusses only the extension of FACS for MMV problem, a similar approach can be used to extend the other fusion algorithms developed in this thesis.

6.A Proof of Lemma 6.1 on page 118

The proof is inspired by Proposition 3.5 by Needell and Tropp [29]. Define set S as the convex combination of all matrices which are...
$R + K$ sparse and have unit Frobenius norm.

$$S = \text{conv} \left\{ \mathbf{X} : \|\mathbf{X}\|_0 \leq R + K, \|\mathbf{X}\|_F = 1 \right\}$$

Using the relation $\|A\mathbf{X}\|_F \leq \sqrt{1 + \delta_{R+K}} \|\mathbf{X}\|_F$, we get,

$$\arg\max_{\mathbf{X} \in S} \|A\mathbf{X}\|_F \leq \sqrt{1 + \delta_{R+K}}.$$ 

Define

$$Q = \left\{ \mathbf{X} : \|\mathbf{X}\|_F + \frac{1}{\sqrt{R + K}} \|\mathbf{X}\|_2,1 \leq 1 \right\}.$$ 

The lemma essentially claims that

$$\arg\max_{\mathbf{X} \in Q} \|A\mathbf{X}\|_F \leq \arg\max_{\mathbf{X} \in S} \|A\mathbf{X}\|_F.$$ 

To prove this, it is sufficient to ensure that $Q \subset S$.

Consider a matrix $\mathbf{X} \in Q$. Partition the support of $\mathbf{X}$ into sets of size $R + K$. Let set $\mathcal{I}_0$ contain the indices of the $R + K$ rows of $\mathbf{X}$ which have largest row $\ell_2$-norm, breaking ties lexicographically. Let set $\mathcal{I}_1$ contain the indices of the next largest (row $\ell_2$-norm) $R + K$ rows and so on. The final block $\mathcal{I}_J$ may have lesser than $R + K$ components. This partition gives rise to the following decomposition:

$$\mathbf{X} = \mathbf{X}|_{\mathcal{I}_0} + \sum_{j=1}^{J} \mathbf{X}|_{\mathcal{I}_j} = \lambda_0 \mathbf{Y}_0 + \sum_{j=1}^{J} \lambda_j \mathbf{Y}_j,$$

where $\lambda_j = \|\mathbf{X}|_{\mathcal{I}_j}\|_F$ and $\mathbf{Y}_j = \lambda_j^{-1} \mathbf{X}|_{\mathcal{I}_j}$. 

141
Chapter 6 Fusion of Algorithms for Multiple Measurement Vectors

By construction each matrix $Y_j$ belongs to $S$ because it is $R + K$ sparse and has unit Frobenius norm. We will show that $\sum \lambda_j \leq 1$. This implies that $X$ can be written as a convex combination of matrices from the set $S$. As a result $X \in S$. Therefore, $Q \subset S$.

Fix some $j$ in the range $\{1, 2, \ldots, J\}$. Then, $I_j$ contains at most $R + K$ elements and $I_j - 1$ contains exactly $R + K$ elements. Therefore,

$$\lambda_j = \|X|_{I_j}\|_F \leq \sqrt{R + K} \|X|_{I_j}\|_{2, \infty} \leq \sqrt{R + K} \cdot \frac{1}{R + K} \|X|_{I_j - 1}\|_{2, 1}.$$

The last inequality holds because the row $\ell_2$-norm of $X$ on the set $I_{j-1}$ dominates its largest row $\ell_2$-norm in $I_j$. Summing these relations, we get

$$\sum_{j=1}^{J} \lambda_j \leq \frac{1}{\sqrt{R + K}} \sum_{j=1}^{J} \|X|_{I_j - 1}\|_{2, 1} \leq \frac{1}{\sqrt{R + K}} \|X\|_{2, 1}.$$

Also, we have $\lambda_0 = \|X|_{I_0}\|_F \leq \|X\|_F$. Since $X \in Q$, we conclude that

$$\sum_{j=0}^{J} \lambda_j \leq \left[\|X\|_F + \frac{1}{\sqrt{R + K}} \|X\|_{2, 1}\right] \leq 1.$$

6.B Proof of Lemma 6.2 on page 119

Let $y^{(i)}$ denote the $i^{th}$ column of matrix $Y$ and $r^{(i)}$ denote the $i^{th}$ column of matrix $R$, $i = 1, 2, \ldots L$. Then we have from Proposition 2.3

$$\left(1 - \frac{\delta_{|T_1| + |T_2|}}{1 - \delta_{|T_1| + |T_2|}}\right) \|y^{(i)}\|_2 \leq \|r^{(i)}\|_2 \leq \|y^{(i)}\|_2.$$
Summing the above relation, we obtain

\[
\left(1 - \frac{\delta |T_1| + |T_2|}{1 - \delta |T_1| + |T_2|}\right) \sum_{i=1}^{L} \|y^{(i)}\|_2^2 \leq \sum_{i=1}^{L} \|r^{(i)}\|_2^2 \leq \sum_{i=1}^{L} \|y^{(i)}\|_2^2.
\]

Equivalently, we have,

\[
\left(1 - \frac{\delta |T_1| + |T_2|}{1 - \delta |T_1| + |T_2|}\right) \|Y\|_F \leq \|R\|_F \leq \|Y\|_F.
\]
Partial information about the non-zero locations and the non-zero values of the sparse signal of interest may be available \textit{a priori} in many applications. For example, in signals such as video, the adjacent temporal frames will be highly coherent and a partial knowledge about the support-set of the current frame can be obtained from the estimate of the previously reconstructed frames. In such situations, it has been shown that a better sparsity-measurement trade-off than conventional Convex Relaxation Methods (CRM) can be achieved by incorporating this knowledge in the CRM framework [40, 149, 165, 166]. This idea has been also extended successfully for other methods to improve the sparsity-measurement trade-off of the existing algorithms [139, 140].

The seminal work by Candés et al. [33] showed that, even in the absence of any \textit{a priori} information, a re-weighted strategy can
improve the reconstruction performance of CRM. This method was referred to as Iterative Re-weighted L1 (IRL1). IRL1 exploits the information from the estimated signal in the current iteration to improve the signal reconstruction quality in the subsequent iteration by selectively penalizing the atoms. Many variations of the iterative re-weighted strategies have been proposed recently [35, 115, 167]. Unfortunately, none of these iterative strategies are easily extendable for an arbitrary Sparse Reconstruction Algorithm (SRA). To the best of our knowledge, there does not exist any general framework for improving the performance of arbitrary SRA, iteratively. In this chapter, we propose a general iterative framework to improve the performance of any arbitrary SRA, which we referred to as Iterative Framework for Sparse Reconstruction Algorithms (IF-SRA). Similar to IRL1, IFSRA exploits the information from the signal estimate in the current iteration to get a better reconstruction quality in the subsequent iteration.

### 7.1 Background

Consider the standard Compressed Sensing (CS) measurement setup described in (2.4). Though (2.4) is an underdetermined system, CS theory showed that stable and robust reconstruction of $x$ is possible if $x$ is sufficiently sparse and $A$ satisfies some incoherence conditions [5, 8]. For example, we can solve the following convex optimization problem to get an estimate of $x$:

$$
\min_x \gamma \|x\|_1 + \frac{1}{2}\|Ax - b\|_2^2,
$$

(7.1)

where $\gamma > 0$ is a pre-fixed regularization parameter. The optimization problem in (7.1) is widely known as Basis Pursuit De-Noising (BPDN) [36] which provides good numerical results and elegant
theoretical guarantees. In BPDN, the $\ell_1$-term promotes sparsity in the solution whereas the $\ell_2$-term ensures consistency in the solution.

In many applications, some partial knowledge about the signal may be available a priori. It has been shown that a weighted version of (7.1) often promotes sparsity better in the solution and improves the reconstruction performance in such cases [40, 107, 149, 165, 166, 168]. The weighted $\ell_1$-norm minimization form of (7.1) can be written as

$$\min_x \sum_{i=1}^{N} u_i |x_i| + \frac{1}{2} \| Ax - b \|_2^2,$$  

(7.2)

where $u_i \geq 0$ denotes the weight at index $i$. The partial knowledge about the signal can be used for setting different weights, which in turn selectively penalizes different coefficients of the signal.

Even in the absence of such prior information, it has been shown that an iterative re-weighting strategy can result in a better sparsity-measurement trade-off than BPDN. IRL1 [33] is one of the early proposed methods in this direction which received wide attention. In the first iteration, IRL1 sets all weights to unity and solves (7.2). In other words, in the first iteration IRL1 solves (7.1) (BPDN). Let $\hat{x}_k$ denote the sparse signal estimated by IRL1 in the $k$th iteration. In the $(k+1)^{th}$ iteration, IRL1 solves (7.2) with $u_i = \frac{1}{\hat{x}_i + \eta}$ where $\eta > 0$ is a pre-fixed parameter. The iteration continues till some halting condition is reached. Though IRL1 shows significant performance improvement over BPDN, in each iteration IRL1 needs to solve a weighted BPDN and hence IRL1 is computationally much more demanding as compared to BPDN. Many variations of IRL1 have been proposed in literature to improve the performance and
reduce the computational cost. For example, Iterative Support Detection (ISD) [35] uses only binary values (0 or 1) as weights. In each iteration, ISD estimates the indices of the dominant part of the signal known as active-set using thresholding or by a more sophisticated first significant jump rule. The atoms in the active-set are given weights equal to zero and weights of the remaining atoms are set to unity to solve a weighted BPDN in the subsequent iteration. ISD showed a better performance than IRL1 in both computation time and reconstruction quality.

This idea of exploiting the partial knowledge about the signal to improve the sparse reconstruction has been also extended to other types of SRAs to improve the sparsity-measurement trade-off [139, 140]. However, to the best of our knowledge, iterative strategies similar to IRL1 are not available for an arbitrary SRA. Next, we develop a general framework which can be used to iteratively improve the sparse reconstruction quality of any SRA.

### 7.2 Iterative Framework for Sparse Signal Reconstruction

In general, solving (2.4) involves three tasks related to the elements of $x$: (i) estimating the sparsity level, (ii) identifying the indices of non-zero elements, and (iii) estimating non-zero values. In this work, we assume that the sparsity level $K$ is known. Needell and Tropp [29] discuss various strategies for choosing $K$ in applications. Once the true atoms are estimated with reasonable accuracy, the non-zero values can be estimated with the help of Least-Squares (LS) method. Hence a better estimate of the support-set will naturally lead to a better sparse signal estimate.
Our aim is to develop a general iterative framework which can be used to improve the performance of any SRAs, even in the absence of any \textit{a priori} information. Like IRL1, we will try to exploit the information available in the estimate of the current iteration to enhance the sparse reconstruction performance in the subsequent iteration(s).

To develop IFSRA, let us define an algorithmic function which is used to denote the general functional form of any SRA to solve (2.4).

\textbf{Definition 7.1.} We define an algorithmic function \textit{SRA} as

\[ [\hat{x}, \hat{T}] = \text{SRA}(A, b, K, \ldots), \]  

(7.3)

where \( \hat{x} \) denotes the estimated sparse signal which is \( K \)-sparse and \( \hat{T} = \text{supp}(\hat{x}) \). The dots in (7.3) indicate the optional parameters used by the SRA, if any.

In (2.4), \( A \) and \( b \) are the essential parameters required for any SRA. Additionally, we keep \( K \) (which is assumed to be known in this work) also as an essential parameter in (7.3). It has been reported that using the knowledge of \( K \) we can de-bias the estimates with the help of LS to result in a better sparse signal estimate [38]. Using the algorithmic function ‘SRA’, now we introduce IFSRA.

To develop the iterative framework, let \( \hat{x}_k \) denote the \( K \)-sparse signal estimate obtained in the \( k \)-th iteration and \( \hat{T}_k = \text{supp}(\hat{x}_k) \) (\( |\hat{T}_k| = K \)). Note that \( x_{\hat{T}_k} \in \mathbb{R}^{K \times 1} \) is possibly dense and \( x_{\hat{T}_c} \in \mathbb{R}^{(N-K) \times 1} \) is sparse. Now, considering a clean measurement case (\( w = 0 \)), we can re-write (2.4) as

\[ b = A_{\hat{T}_c} x_{\hat{T}_c} + A_{\hat{T}_k} x_{\hat{T}_k}. \]  

(7.4)
Let $S_k$ denote the number of true atoms in $\hat{T}_k$. We have, $0 \leq S_k \leq K$. It may be observed from (7.4) that, in the $(k+1)^{th}$ iteration we need to identify only $K - S_k$ true atoms from $N - K$ atoms listed in $\hat{T}_k^c$. Note that this new problem in the $(k+1)^{th}$ iteration is a reduced dimensional problem and has a sparser signal than $x$, with the same number of measurements, $M$. Hence, we are likely to improve the sparse signal reconstruction in the $(k+1)^{th}$ iteration by exploiting the information in $\hat{T}_k$.

Now, we formally define the new sparse reconstruction problem in the $(k+1)^{th}$ iteration, given $\hat{T}_k$. For this, we define

$$U_k \triangleq I - A_{\hat{T}_k} A_{\hat{T}_k}^\dagger.$$  

(7.5)

Note that $U_k$ is the projection matrix from $\mathbb{R}^M$ onto $\mathcal{R}(A_{\hat{T}_k})^\perp$. Let $r_k$ denote the measurement vector regularized by the partially known support-set $\hat{T}_k$, which is defined as

$$r_k = b - A_{\hat{T}_k} A_{\hat{T}_k}^\dagger b = U_k b.$$  

(7.6)

From (7.4), (7.5), and (7.6), we obtain

$$(U_k A_{\hat{T}_k^c}) x_{\hat{T}_k^c} = A_k x_{\hat{T}_k^c} = r_k,$$  

(7.7)

where $A_k = U_k A_{\hat{T}_k^c}$.

We may view (7.7) as a standard CS setting described in (2.4) with $A_k$ as the measurement matrix and $r_k$ as the measurement vector where we need to identify $K - S_k$ atoms from $N - K$ candidate atoms listed in $\hat{T}_k^c$. Note that the sparse reconstruction problem in (7.7) is a reduced dimensional problem as compared to (2.4) with the same number of measurements $M$. In other words, we have a better sparsity-measurement trade-off in (7.7) than in (2.4). $A_k$ may be viewed as a regularization of the measurement matrix.
A using the partially known support-set $\hat{T}_k$. Note that in the worst case we may not have any correct atoms in $\hat{T}_k$. Hence, considering this worst case scenario, in the $(k+1)^{th}$ iteration, we need to target to estimate all $K$ support atoms and do a sanity check to discard the incorrect atoms. We can solve (7.7) by using the SRA yielding

$$[\tilde{x}_{k+1}^*, \tilde{T}_{k+1}^*] = \text{SRA}(A_k, r_k, K),$$

(7.8)

where $\tilde{T}_{k+1}$ is a subset of indices of columns of $A_k$. Since we are solving a reduced dimensional problem in (7.8), we need to map the indices listed in $\tilde{T}_{k+1}$ to the indices of $A$. Let $\tilde{T}_{k+1}$ denote the mapped indices.

Note that we obtained the new estimate using the partially known support-set $\hat{T}_k$ which may include many false atoms. Similarly $\tilde{T}_{k+1}$ may also contain many false atoms. At the end of $(k+1)^{th}$ iteration, we need to retain only $K$ potential atoms from $\hat{T}_k$ and $\tilde{T}_{k+1}$. Recently Ambat et al. [91] proposed a fusion scheme based on LS to efficiently identify the true atoms from the union of the estimated support-sets. We use this fusion strategy here to select $K$ potential atoms from $\hat{T}_k$ and $\tilde{T}_{k+1}$.

Combining these ideas, the proposed IFSRA algorithm is summarized in Algorithm 7.1. During each iteration, IFSRA performs the following tasks to successively improve the solution:

i) *Estimation:* Solve the regularized sparse reconstruction problem using SRA (steps 3-4 in Algorithm 7.1).

ii) *Fusion:* Fuse the estimate of this regularized sparse reconstruction problem and the estimate in the previous iteration, and retain only $K$ potential atoms in the estimated support-set (steps 5-7 in Algorithm 7.1).
Algorithm 7.1: IFSRA

Inputs: $A_{M \times N}$, $b_{M \times 1}$, and $K$.

Initialization: $k = 0$, $A_0 = A$, $r_0 = b$, $\hat{T}_0 = \phi$;

1: repeat
2: $k = k + 1$;

Estimation Steps:
3: $[\tilde{x}_k^*, \tilde{T}_k^*] = \text{SRA}(A_{k-1}, r_{k-1}, K, \ldots)$;
4: $\hat{T}_k = \text{set of indices of atoms of } A \text{ listed in } \tilde{T}_k^*$;

Fusion Steps:
5: $\Gamma_k = \hat{T}_k \cup \hat{T}_{k-1}$; $\blacktriangleright K \leq |\Gamma_k| \leq 2K$
6: $v_{\Gamma_k} = A_{\Gamma_k}^\dagger b$, $v_{\Gamma_k^c} = 0$; $\blacktriangleright v \in \mathbb{R}^{N \times 1}$
7: $\hat{T}_k = \text{supp}(v^K)$; $\blacktriangleright |T| = K$

Regularization Steps:
8: $U_k = I - A_{\hat{T}_k} A_{\hat{T}_k}^\dagger$;

9: $r_k = U_k b$; $\blacktriangleright$ regularize $b$ using $\hat{T}_k$
10: $A_k = U_k A_{\hat{T}_k^c}$; $\blacktriangleright$ regularize $A$ using $\hat{T}_k$

11: until ($\|r_k\|_2 \geq \|r_{k-1}\|_2$);
12: $\hat{T} = \hat{T}_{k-1}$;
13: $\hat{x}_{\hat{T}} = A_{\hat{T}}^\dagger b$, $\hat{x}_{\hat{T}^c} = 0$; $\blacktriangleright \hat{x} \in \mathbb{R}^{N \times 1}$

Outputs: $\hat{x}$ and $\hat{T}$.

iii) Regularization: Regularize the sparse reconstruction problem in (2.4) by regularizing both the measurement matrix $A$ and the measurement vector $b$ in order to remove the effect of the columns of $A$ which are listed in the estimated support-set (steps 9-10 in Algorithm 7.1).

These steps continues as long as the $\ell_2$-norm of the regularized measurement vector decreases. We may also use any other halting criteria while realizing IFSRA.

Next, we give a small example to demonstrate the progression of IFSRA over iterations.
7.2.1 A Demonstration

Matching Pursuit (MP) [1] is one of the simplest SRA. Using MP as the parent SRA in IFSRA, we now show that Algorithm 7.1 works for a toy experiment in a clean measurement case \((w = 0)\). We created a sparse signal \(x\) of size 20 with 5 non-zero elements (i.e., \(N = 20\) and \(K = 5\)). The 5 non-zero values of \(x\) were generated independently from the standard Gaussian distribution. Setting \(M = 12\), we generated an \(M \times N\) Gaussian random matrix \(A\) whose columns were normalized to unity and set \(b = Ax\). We applied IFSRA with MP as the parent SRA to reconstruct the sparse signal \(x\) from \(A\) and \(b\) which we denote by IFSRA(MP) and the results are illustrated in Figure 7.1.

![Figure 7.1: Progression of IFSRA(MP) over iterations for Gaussian sparse signal \((N = 20, K = 5, \text{ and } M = 12)\): (a) true sparse signal, (b) result of first iteration of IFSRA(MP) which is same as MP, (c) result of iteration 2 of IFSRA(MP), and (d) result of iteration 3 of IFSRA(MP).](image)

Note that in the first iteration, IFSRA executes SRA without any modifications to (2.4). Hence the solution obtained in the first iteration of IFSRA(MP) is same as the solution of MP. MP could
identify only 3 true atoms out of 5 and the corresponding sparse signal estimate \( \hat{x} \) resulted in an error \( \|x - \hat{x}\|_2 = 0.8743 \) (refer Figure 7.1(b)). In the second iteration, IFSRA(MP) identified 4 true atoms resulting in a lesser error \( \|x - \hat{x}\|_2 = 0.4131 \). IFSRA(MP) estimates all 5 true atoms correctly in the third iteration leading to a perfect sparse signal reconstruction (Figure 7.1(d)).

The example illustrates a scenario where IFSRA provides a perfect signal reconstruction whereas the parent SRA fails. Next, we study the theoretical properties of IFSRA.

### 7.3 Theoretical Analysis of IFSRA

We present theoretical analysis of the proposed scheme and derive sufficient conditions for enhancing the sparse signal estimate, using the Restricted Isometry Property (RIP) \([7, 8]\).

The following lemma is derived from Lemma 3 in \([28]\).

**Lemma 7.1.** Consider the measurement system in (2.4) for a \( K \)-sparse signal \( x \in \mathbb{R}^{N \times 1} \) with support-set \( T \). Let the measurement matrix \( A \in \mathbb{R}^{M \times N} \) have Restricted Isometry Constant (RIC) \( \delta_{3K} \). For an arbitrary set \( T_1 \subset \{1, 2, \ldots, N\} \) with \( |T_1| \leq K \), define \( \hat{x}_{T_1} = A_{T_1}^\dagger b \) and \( \hat{x}_{T_1^c} = 0 \). Then, we have,

\[
\|x - \hat{x}\|_2 \leq \frac{1}{1 - \delta_{3K}} \|x_{T_1^c}\|_2 + \frac{1}{1 - \delta_{3K}} \|w\|_2.
\]

**Proof.** We have,

\[
\|x - \hat{x}\|_2 \leq \|x_{T_1} - A_{T_1}^\dagger b\|_2 + \|x_{T \setminus T_1}\|_2
\]

\[
= \|x_{T_1} - A_{T_1}^\dagger (A_T x_T + w)\|_2 + \|x_{T \setminus T_1}\|_2
\]
\[
\begin{align*}
\|x_{\mathcal{T}_1} - A_{T_1}^\dagger (A_{T_1} x_{T_1} + A_{T\setminus T_1} x_{T\setminus T_1} + w)\|_2 + \|x_{T\setminus T_1}\|_2 \\
(\because x_{T_1\setminus T} = 0) \\
\leq \|A_{T_1}^\dagger A_{T\setminus T_1} x_{T\setminus T_1}\|_2 + \|A_{T_1}^\dagger w\|_2 + \|x_{T\setminus T_1}\|_2 \\
(\because A_{T_1} A_{T_1}^\dagger = I) \\
\leq \|(A_{T_1}^T A_{T_1})^{-1} A_{T_1}^T A_{T\setminus T_1} x_{T\setminus T_1}\|_2 + \|x_{T\setminus T_1}\|_2 + \frac{\|w\|_2}{\sqrt{1 - \delta_{3K}}} \\
& \quad \text{(using (2.7))} \\
\leq \frac{1}{1 - \delta_{3K}} \|A_{T_1}^T A_{T\setminus T_1} x_{T\setminus T_1}\|_2 + \|x_{T\setminus T_1}\|_2 + \frac{\|w\|_2}{\sqrt{1 - \delta_{3K}}} \\
& \quad \text{(using (2.6))} \\
\leq \left( \frac{\delta_{3K}}{1 - \delta_{3K}} + 1 \right) \|x_{T\setminus T_1}\|_2 + \frac{\|w\|_2}{1 - \delta_{3K}} \\
& \quad \text{(using (2.9))} \\
= \frac{1}{1 - \delta_{3K}} \|x_{T_1}\|_2 + \frac{1}{1 - \delta_{3K}} \|w\|_2.
\end{align*}
\]

\[\blacksquare\]

**Lemma 7.2.** Let \( A \in \mathbb{R}^{M \times N} \) have RIP of order \( K \) with RIC \( \delta_K \). Consider \( T \subset \{1, 2, 3, \ldots, N\} \) such that \( |T| = K \) and define \( U \triangleq (I - A_T A_T^\dagger) \). Then \( UA_{T^c} \) satisfies RIP with RIC \( \delta_K \).

**Proof.** Since \( A \) has RIP with RIC \( \delta_K \), by setting \( x_{T^c} = 0 \), we can observe that \( A_T \) has full column rank.

Using (2.5), we have,

\[ (1 - \delta_K) \|x_{T^c}\|_2^2 \leq (1 - \delta_K) \|x_T\|_2^2 + \|x_{T^c}\|_2^2 \leq \|A_T x_T + A_{T^c} x_{T^c}\|_2^2. \] (7.9)

Now, for any \( K - S \) sparse signal \((0 < S < K) x_{T^c} \in \mathbb{R}^{(N-K) \times 1}\), by setting \( x_T = -A_T^\dagger A_{T^c} x_{T^c} = -(A_T^T A)^{-1} A_T^T A_{T^c} x_{T^c} \), we get \( \|A_T x_T + A_{T^c} x_{T^c}\|_2^2 = \|UA_{T^c} x_{T^c}\|_2^2 \). Substituting this in
(7.9), we get

\[(1 - \delta_K)\|x_{T^c}\|^2 \leq \|UA_{T^c}x_{T^c}\|^2. \quad (7.10)\]

Setting \(x_{T} = 0\), we get

\[\|UA_{T^c}x_{T^c}\|^2 = \|A_{T^c}x_{T^c} - A_{T^c}A_T^\dagger A_{T^c}x_{T^c}\|^2 \leq \|A_{T^c}x_{T^c}\|^2 \quad \text{(using (2.10))}\]

\[\leq (1 + \delta_K)\|x_{T^c}\|^2 \quad \text{(using (2.5))} \quad (7.11)\]

Combining (7.10) and (7.11) we get the result. ■

Lemma 7.2 guarantees that, \(UA_{T^c}\) will satisfy RIP with the same RIC as that of \(A\). We use this property during the theoretical analysis of IFSRA.

### 7.3.1 Performance of IFSRA for Sparse Signals under Measurement Perturbations

The theoretical analysis of IFSRA using RIP, in the case of \(K\) sparse signal in the presence of measurement perturbations, is given in Theorem 7.1.

**Theorem 7.1.** (Measurement Perturbations) Let the measurement matrix \(A \in \mathbb{R}^{M \times N}\) have RIC \(\delta_{3K}\) and assume that we use an arbitrary SRA to reconstruct a \(K\)-sparse signal \(x\) from (2.4). Using the algorithmic function ‘SRA’ (please refer Definition 1), we have, \([\tilde{x}, \tilde{T}] = SRA(A, b, K, \ldots)\). Now, consider IFSRA (Algorithm 7.1) which uses the given SRA to iteratively improve the performance. Assume that the given SRA satisfies the relation

\[\|x_{T^c}\|_2 \leq C\|x\|_2 + D\|w\|_2, \quad (7.12)\]
where \( C = \beta \frac{(1-2\delta_{3K})^2}{(1+\delta_{3K})^2} \) with \( 0 \leq \beta < 1 \), and \( D \geq 0 \). Then, following the notations used in Section 7.2 and Algorithm 7.1, in the \( k \)th iteration of IFSRA, we have,

i) \[ \| r_k \|_2 \leq \beta \| r_{k-1} \|_2 + C_1 \| w \|_2, \]

ii) \[ \| x - \hat{x}_k \|_2 \leq \beta^k \| x \|_2 + C_2 \| w \|_2, \]

where the constants \( C_1 \) and \( C_2 \) are defined as

\[ C_1 = \frac{\beta + D(1 + \delta_{3K})^2 + 3 + \delta_{3K}}{(1 - 2\delta_{3K})^2} \]

and

\[ C_2 = \frac{D(1 + \delta_{3K}) + 3 - \delta_{3K}}{(1 - \beta)(1 - \delta_{3K})^2}. \]

Proof: To prove the theorem, first we show that, in the \( k \)th iteration, the given SRA satisfies the inequality:

\[ \| x_{\hat{T}_k^c} \|_2 \leq C \| x_{T_k^c} \|_2 + D \| w \|_2. \]

Then, we establish a relation between \( \| x_{\hat{T}_k^c} \|_2 \) and \( \| x_{\Gamma^c} \|_2 \).

The given SRA satisfies (7.12) for reconstructing \( x \) from (2.4). Note that, in the \( k \)th iteration of IFSRA, we use the given SRA to solve the system \( A_{k-1} x_{\hat{T}_k^c} = r_{k-1} \) where \( A_{k-1} = U_{k-1} A_{\hat{T}_k^c} \). Since \( A \) satisfies RIP with RIC \( \delta_{3K} \), using Lemma 7.2 on page 155 we get, \( A_{k-1} \) has RIP with RIC \( \delta_{3K} \) (here we also used the fact that a \( K \)-sparse signal also belongs to the family of \( 3K \) sparse signals.).

Hence, in the \( k \)th iteration, using (7.12), the SRA will satisfy the following relation:

\[ \| x_{\hat{T}_k^c} \|_2 \leq C \| x_{T_k^c} \|_2 + D \| w \|_2. \]

Next, we will derive a relationship between \( \| x_{\hat{T}_k^c} \|_2 \) and \( \| x_{\Gamma^c} \|_2 \). Let us define \( \Delta_k \triangleq \Gamma_k \setminus \hat{T}_k \). That is, \( \Delta_k \) is the set formed by the atoms in \( \Gamma_k \) which are discarded by Algorithm 7.1. Since \( \hat{T}_k \subset \Gamma_k \), we have \( \hat{T}_k^c = \Gamma_k^c \cup \Delta_k \) and hence we get

\[ \| x_{\hat{T}_k^c} \|_2 \leq \| x_{\Gamma_k^c} \|_2 + \| x_{\Delta_k} \|_2. \]
Let us consider

$$
\| (\mathbf{v}_k^\Gamma)_{\Delta_k} \|_2^2 = \| \mathbf{x}_{\Delta_k} + (\mathbf{v}_k^\Gamma - \mathbf{x}_k^\Gamma)_{\Delta_k} \|_2^2 \\
\geq \| \mathbf{x}_{\Delta_k} \|_2^2 - \| (\mathbf{v}_k^\Gamma - \mathbf{x}_k^\Gamma)_{\Delta_k} \|_2^2 \\
(\text{using reverse triangle inequality})
$$

$$
\Rightarrow \| \mathbf{x}_{\Delta_k} \|_2 \leq \| (\mathbf{v}_k^\Gamma)_{\Delta_k} \|_2 + \| (\mathbf{v}_k^\Gamma - \mathbf{x}_k^\Gamma)_{\Delta_k} \|_2 \\
\leq \| (\mathbf{v}_k^\Gamma)_{\Delta_k} \|_2 + \| \mathbf{v}_k^\Gamma - \mathbf{x}_k^\Gamma \|_2. \quad (7.15)
$$

Note that $\hat{T}_k$ contains the $K$-elements of $v_k^\Gamma$ with the highest magnitudes. Therefore, using $|\hat{T}_k| = |T| = K$, we have, $\| (v_k^\Gamma)_{\hat{T}_k} \|_2^2 \geq \| (v_k^\Gamma)_{T} \|_2^2$ and hence we obtain

$$
\| (v_k^\Gamma)_{T} \|_2^2 - \| (v_k^\Gamma)_{\hat{T}_k} \|_2^2 \leq 0. \quad (7.16)
$$

Now consider

$$
\| (v_k^\Gamma)_{\Delta_k} \|_2^2 = \| (v_k^\Gamma)_{\Delta_k} \|_2^2 + \| (v_k^\Gamma)_{\hat{T}_k} \|_2^2 - \| (v_k^\Gamma)_{\hat{T}_k} \|_2^2 \\
= \| (v_k^\Gamma)_{\Delta_k} \|_2^2 + \| (v_k^\Gamma)_{\hat{T}_k} \|_2^2 - \| (v_k^\Gamma)_{\hat{T}_k} \|_2^2 \\
= \| (v_k^\Gamma)_{\Delta_k} \|_2^2 + \| (v_k^\Gamma)_{\hat{T}_k} \|_2^2 - \| (v_k^\Gamma)_{\hat{T}_k} \|_2^2 \\
\leq \| (v_k^\Gamma)_{\hat{T}_k} \|_2^2 \quad (\text{using (7.16)})
$$

Therefore we have,

$$
\| (v_k^\Gamma)_{\Delta_k} \|_2 \leq \| (v_k^\Gamma)_{\hat{T}_k} \|_2 \\
= \| (v_k^\Gamma)_{\hat{T}_k} - \mathbf{x}_{\hat{T}_k} \|_2 \quad (\therefore \mathbf{x}_{\hat{T}_k} = \mathbf{0}) \\
= \| (v_k^\Gamma - \mathbf{x}_{\hat{T}_k})_{\hat{T}_k} \|_2 \\
\leq \| (v_k^\Gamma - \mathbf{x}_{\hat{T}_k}) \|_2. \quad (7.17)
$$

Substituting (7.17) in (7.15), we get

$$
\| \mathbf{x}_{\Delta_k} \|_2 \leq 2 \| (v_k^\Gamma - \mathbf{x}_{\hat{T}_k}) \|_2. \quad (7.18)
$$

158
Chapter 7  Iterative Framework

Now consider

\[
\| (v_k - x_k) \|_2 = \left\| A_{i_k}^\dagger b - x_k \right\|_2 \\
= \left\| A_{i_k}^\dagger (A_{i_k} x_k + A_{i_k}^e x_{i_k}^e + w) - x_k \right\|_2
\]

(\because b = Ax + w)

\[
\leq \left\| A_{i_k}^\dagger A_{i_k} x_k^e \right\|_2 + \left\| A_{i_k}^\dagger w \right\|_2
\]

(\because A_{i_k}^\dagger A_{i_k} = I)

(\text{using definition of } A_{i_k}^\dagger)

\leq \frac{1}{1 - \delta_{2K}} \left\| A_{i_k}^\dagger A_{i_k} x_k^e \right\|_2 + \frac{1}{\sqrt{1 - \delta_{2K}}} \| w \|_2

(\text{using (2.6)) & (2.7})

\leq \frac{\delta_{3K}}{1 - \delta_{2K}} \| x_k^e \|_2 + \frac{1}{\sqrt{1 - \delta_{2K}}} \| w \|_2

(\text{using (2.9))}

\leq \frac{\delta_{3K}}{1 - \delta_{2K}} \| x_k^e \|_2 + \frac{1}{1 - \delta_{3K}} \| w \|_2

(\because \delta_{2K} \leq \delta_{3K}, \ 0 \leq 1 - \delta_{2K} \leq 1)

(7.19)

Now, using (7.18), and (7.19) in (7.14), we obtain

\[
\| x_{\tilde{T}_k}^e \|_2 \leq \left(1 + \frac{2\delta_{3K}}{1 - \delta_{3K}}\right) \| x_{i_k}^e \|_2 + \frac{2}{1 - \delta_{3K}} \| w \|_2
\]

\[
= \frac{1 + \delta_{3K}}{1 - \delta_{3K}} \| x_{i_k}^e \|_2 + \frac{2 \| w \|_2}{1 - \delta_{3K}}.
\]

(7.20)

With the results in (7.13) and (7.20) we now prove the theorem.

i) Since \( \tilde{T}_k \subset \Gamma_k \), we have

\[
\| x_{i_k}^e \|_2 \leq \| x_{\tilde{T}_k}^e \|_2 \leq C \| x_{\tilde{T}_k}^e \|_2 + D \| w \|_2
\]

(7.21)
Chapter 7  
Iterative Framework

Using (7.20) and (7.21), we get

\[ \| x_{\hat{r}_k} \|_2 \leq C \frac{1 + \delta_{3K}}{1 - \delta_{3K}} \| x_{\hat{r}_{k-1}} \|_2 + \frac{D(1 + \delta_{3K}) + 2}{1 - \delta_{3K}} \| w \|_2 \] (7.22)

Using (7.22) we can find an upper bound for \( \| r_k \|_2 \) as follows.

We have,

\[ \| r_k \|_2 = \| b - A_{\hat{r}_k} A_{\hat{r}_k}^\dagger b \|_2 \]
\[ \leq \| A_{\hat{r}_k} x_{\hat{r}_k} - A_{\hat{r}_k} A_{\hat{r}_k}^\dagger A_{\hat{r}_k} x_{\hat{r}_k} \|_2 + \| w - A_{\hat{r}_k} A_{\hat{r}_k}^\dagger w \|_2 \] (using (2.4))
\[ \leq \| A_{\hat{r}_k} x_{\hat{r}_k} \|_2 + \| w - A_{\hat{r}_k} A_{\hat{r}_k}^\dagger w \|_2 \] (using (2.10))
\[ = \| A_{\hat{r}_k} x_{\hat{r}_k} \|_2 + \| w - A_{\hat{r}_k} A_{\hat{r}_k}^\dagger w \|_2 \] (∵ \( x_{\hat{r}_k} = 0 \))
\[ \leq \sqrt{1 + \delta_K} \| x_{\hat{r}_k} \|_2 + \| w \|_2 \] (using (2.5) and (2.10))
\[ \leq (1 + \delta_{3K}) \| x_{\hat{r}_k} \|_2 + \| w \|_2 \] (∵ \( \delta_K \leq \delta_{3K} \))
\[ \leq \frac{C(1 + \delta_{3K})^2}{1 - \delta_{3K}} \| x_{\hat{r}_{k-1}} \|_2 \] (using (7.22))
\[ + \frac{D(1 + \delta_{3K})^2 + 3 + \delta_{3K}}{1 - \delta_{3K}} \| w \|_2 \] (7.23)

Next, we will derive a lower bound for \( \| r_{k-1} \|_2 \).

We have,

\[ \| r_{k-1} \|_2 = \| b - A_{\hat{r}_{k-1}} A_{\hat{r}_{k-1}}^\dagger b \|_2 \]
\[ \geq \| A_{\hat{r}_{k-1}} x_{\hat{r}_{k-1}} - A_{\hat{r}_{k-1}} A_{\hat{r}_{k-1}}^\dagger A_{\hat{r}_{k-1}} x_{\hat{r}_{k-1}} \|_2 - \| w - A_{\hat{r}_{k-1}} A_{\hat{r}_{k-1}}^\dagger w \|_2 \]
\[ \geq \frac{1 - \delta_{2K}}{1 - \delta_{3K}} \| A_{\hat{r}_{k-1}} x_{\hat{r}_{k-1}} \|_2 - \| w \|_2 \]
\[ \geq \frac{1 - \delta_{2K}}{\sqrt{1 - \delta_K}} \| x_{\hat{r}_{k-1}} \|_2 - \| w \|_2 \]
\[ \geq \frac{1 - 2\delta_{3K}}{\sqrt{1 - \delta_K}} \| x_{\hat{r}_{k-1}} \|_2 - \| w \|_2 \]
\[ \geq (1 - 2\delta_{3K}) \| x_{\hat{r}_{k-1}} \|_2 - \| w \|_2 \] (7.24)
Using (7.23) and (7.24), we get
\[
\|r_k\|_2 \leq \frac{C(1 + \delta_{3K})^2}{(1 - \delta_{3K})(1 - 2\delta_{3K})} \|r_{k-1}\|_2 \\
+ \frac{(C + D)(1 + \delta_{3K})^2 + 3 + \delta_{3K}}{(1 - \delta_{3K})(1 - 2\delta_{3K})} \|w\|_2.
\]

Substituting the value of \(C\), we get
\[
\|r_k\|_2 \leq \frac{\beta}{1 - 2\delta_{3K}} \|r_{k-1}\|_2 + \frac{\beta \delta_{3K} + D(1 + \delta_{3K})^2 + 3 + \delta_{3K}}{(1 - \delta_{3K})^2} \|w\|_2 \tag{7.25}
\]

ii) Using Lemma 7.1, we have,
\[
\|x - \hat{x}_k\|_2 \leq \frac{1}{1 - \delta_{3K}} \|\hat{x}_k\|_2 + \|w\|_2 \frac{1}{1 - \delta_{3K}} \\
\leq \frac{(1 + \delta_{3K})\|x_{\Gamma_k}\|_2 + (3 - \delta_{3K})\|w\|_2}{(1 - \delta_{3K})^2} \tag{using (7.20))
\leq \frac{C(1 + \delta_{3K})}{(1 - \delta_{3K})^2} \|\hat{x}_{k-1}\|_2 + \frac{D(1 + \delta_{3K}) + 3 - \delta_{3K}}{(1 - \delta_{3K})^2} \|w\|_2 \tag{using (7.21))
\leq \frac{C(1 + \delta_{3K})}{(1 - \delta_{3K})^2} \|x - \hat{x}_{k-1}\|_2 + \frac{D(1 + \delta_{3K}) + 3 - \delta_{3K}}{(1 - \delta_{3K})^2} \|w\|_2 \\
\leq \beta \|x - \hat{x}_{k-1}\|_2 + \frac{D(1 + \delta_{3K}) + 3 - \delta_{3K}}{(1 - \delta_{3K})^2} \|w\|_2 \tag{7.26}
\]

\[
\leq \beta \|x - \hat{x}_{k-1}\|_2 + \frac{D(1 + \delta_{3K}) + 3 - \delta_{3K}}{(1 - \delta_{3K})^2} \|w\|_2 \tag{7.27}
\]

Hence, setting \(\hat{x}_0 = 0\), we obtain
\[
\|x - \hat{x}_k\|_2 \leq \beta^k \|x\|_2 + \frac{1 - \beta^k}{1 - \beta} \frac{D(1 + \delta_{3K}) + 3 - \delta_{3K}}{(1 - \delta_{3K})^2} \|w\|_2
\]

161
\[
\leq \beta^k \|x\|_2 + C_2 \|w\|_2.
\]

Convergence Guarantees of IFSRA: Since \(0 < \beta < 1\), \(\beta^k\) vanishes as \(k\) increases. Hence from Theorem 7.1 on page 156, we can interpret the error guarantees of IFSRA as follows. IFSRA can recover a \(K\)-sparse signal to arbitrarily high precision in a clean measurement case (\(w = 0\)). Any SRA cannot resolve the uncertainty due to the additive measurement noise. The performance of IFSRA also degrades gracefully as the energy in the noise increases.

The other important consequences of Theorem 7.1 are listed below.

- Theorem 7.1 gives sufficient conditions for progression of IFSRA during each iteration in terms of reconstruction error.

- In practice where the signal \(x\) is unknown, \(\ell_2\)-norm of the regularized measurement vector \((r_k)\) is often used as a measure of performance. Theorem 7.1 also discusses this case and gives sufficient conditions for reducing the \(\ell_2\)-norm of the residue over iterations.

- Consider a clean measurement case (\(w = 0\)) where the parent SRA fails to get a perfect reconstruction, but satisfies the relation \(\|x_\tilde{T}\|_2 \leq C\|x\|_2\). Then Theorem 7.1 guarantees that IFSRA will converge to the true sparse signal (\(\because \beta < 1\)). This clearly guarantees the advantage of using IFSRA over the parent SRA in such cases.

### 7.3.2 Performance of IFSRA for Arbitrary Signals under Measurement Perturbations

In practice, most of the signals we often deal in applications are not strictly sparse. Fortunately many natural signals are found to
be compressible in some transformed domain, which can be well approximated by a sparse representation of the same [80]. Hence, the performance of any scheme under signal perturbations carries significant interest in CS. Next, we study the performance of IFSRA for an arbitrary signal $x \in \mathbb{R}^{N \times 1}$.

**Theorem 7.2.** (Signal and Measurement Perturbations) Let $x \in \mathbb{R}^{N \times 1}$ be an arbitrary signal with all other conditions in Theorem 7.1 on page 156 hold good. Then, we have,

\[
\begin{align*}
\text{i)} \quad & \|r_k\|_2 \leq \beta \|r_{k-1}\|_2 + C_1 \sqrt{1 + \delta_{3K}} \|x - x^K\|_2 + C_1 \|x - x^K\|_1 + C_1 \|w\|_2, \text{ and} \\
\text{ii)} \quad & \|x - \hat{x}_k\|_2 \leq \beta_k \|x\|_2 + C_2 \sqrt{1 + \delta_{3K}} \|x - x^K\|_2 + C_2 \|x - x^K\|_1 + C_2 \|w\|_2.
\end{align*}
\]

*Proof.* Proof is given in Appendix 7.A. ■

It may be observed that the results given Theorem 7.2 is tight. For a $K$-sparse signal, the results in Theorem 7.2 reduce to the result given in Theorem 7.1.

### 7.3.3 Remarks on IFSRA

The proposed IFSRA has the following properties which are highly desirable for any iterative framework for SRA.

- IFSRA can accommodate any SRA without any modification of the SRA. At the end of each iteration, IFSRA regularizes the original CS problem using the currently estimated support-set and applies the parent SRA on the new regularized problem.
• Unlike the iterative re-weighted algorithms like IRL1, which has an *ad-hoc* tuning parameter, IFSRA does not have any such tuning parameters.

• In an ideal measurement setup \((w = 0)\), for a \(K\)-sparse signal, IFSRA shows elegant convergent guarantees.

• IFSRA is robust against both signal and measurement perturbations. The performance of IFSRA degrades gracefully in the presence of perturbations.

• IFSRA also provides error guarantees for arbitrary signals under measurement perturbations.

• Along with the elegant theoretical guarantees, as we show in Section 7.4, IFSRA also shows significant sparse signal reconstruction performance improvement for both synthetic and real-world signals.

It may be noted that, as IFSRA runs the parent SRA multiple times to enhance the sparse signal reconstruction, it is computationally more demanding than the parent SRA.

### 7.4 Numerical Experiments and Results

In general, the theoretical results in CS give *pessimistic* worst case bounds. While dealing with real-world applications, we are often interested in the simulation results for benchmarking. To evaluate the performance of the proposed IFSRA we conducted numerical experiments using different SRAs. We choose three popular SRAs *viz*. MP [1], Compressive Sampling Matching Pursuit (CoSaMP) [29], and BPDN [36].
These three algorithms work with different principles. MP is one of the simplest and early proposed SRAs which work iteratively in a greedy fashion. In each iteration, MP estimates and updates one non-zero value in the sparse signal. CoSaMP is also an iterative greedy algorithm which received much attention due to its elegant theoretical guarantees and simple geometrical interpretation. CoSaMP estimates a support-set of cardinality $K$ in each iteration and the estimated support-set is refined in successive iterations. To evaluate the performance of IFSRA with these algorithms we conducted experiments on both synthetic and real signals.

Note that, in general, a SRA (for example, BPDN) may not give an exact $K$-sparse solution. Hence we choose $\hat{T} = \text{supp}(\hat{x}^K)$ and $\hat{x}_{\hat{T}}$ is re-estimated by solving $b = A_{\hat{T}} \hat{x}_{\hat{T}}$ and set $\hat{x}_{\hat{T}^c} = 0$. It has been observed that this de-biasing operation will also improve the sparse signal estimate [38]. For realizing BPDN, we used the function ‘SolveBP’ available in the package ‘SparseLab’ [151].

One of the main objectives in CS is to reduce the number of measurements without degrading the sparse reconstruction performance. Hence, the performance of SRA in the lower measurement cases (smaller values of $M$) carry a significant interest in CS. To measure the level of under-sampling in CS, we define the fraction of measurements, denoted by $\alpha$, as

$$\alpha = \frac{M}{N}. \quad (7.28)$$

We evaluate the performance of IFSRA in numerical experiments using Average Signal-to-Reconstruction-Error Ratio (ASRER) defined in (3.19).
7.4.1 Synthetic Sparse Signals

We used the simulation setup described in Section 5.2 on page 107 and conducted the experiments with \( N = 500, K = 20, S = 100, \) and \( T = 100 \). That is, we used sparse signals with dimension 500 and sparsity level 20. This 4\% level of sparsity is intentionally chosen, as it closely resembles many real application scenarios. For example, it is empirically observed that most of the energy of any natural image in the wavelet domain is concentrated within 2\%–4\% of the coefficients [80]. We used 100 realizations of \( A \) (i.e., \( S = 100 \)) and for each realization of \( A \), we randomly generated 100 sparse signal vectors (i.e., \( T = 100 \)).

The simulation results by using IFSRA with MP as the parent algorithm is shown in Figure 7.2. For both GSS and RSS, IFSRA(MP) showed significant performance improvement over MP. For example, at \( \alpha = 0.22 \), IFSRA(MP) resulted in 3.5 dB ASRER improvement over MP for GSS (refer Figure 7.2(a)). Similarly, for RSS, at \( \alpha = 0.29 \), IFSRA(MP) gave 16 dB ASRER improvement as compared to MP (refer Figure 7.2(b)).

CoSaMP

Figure 7.3 depicts the performance of IFSRA with CoSaMP as the parent algorithm which shows improved ASRER as compared to CoSaMP. For example, at \( \alpha = 0.20 \), IFSRA(CoSaMP) showed 2.8 dB performance improvement over CoSaMP, for GSS.
Figure 7.2: Performance of IFSRA(MP) under measurement perturbations (SMNR = 20 dB) in terms of Average Signal-to-Reconstruction-Noise-Ratio (ASRER) vs. Fraction of Measurements ($\alpha$), averaged over 10,000 trials. Sparse signal dimension $N = 500$ and sparsity level $K = 20$.

Figure 7.3: Performance of IFSRA(CoSaMP) under measurement perturbations (SMNR = 20 dB) in terms of Average Signal-to-Reconstruction-Noise-Ratio (ASRER) vs. Fraction of Measurements ($\alpha$), averaged over 10,000 trials. Sparse signal dimension $N = 500$ and sparsity level $K = 20$.

IFSRA(BPDN) also shows a significant ASRER improvement over the parent algorithm, which is shown in Figure 7.4. For $\alpha = 0.20$,
IFSRA(BPDN) provided 3.5 dB ASRER improvement over BPDN for GSS (refer Figure 7.4(a)). Similarly for RSS, at $\alpha = 0.22$, IFSRA(BPDN) showed 7.3 dB ASRER improvement as compared to BPDN (refer Figure 7.4(b)).

Figure 7.4: Performance of IFSRA(BPDN) under measurement perturbations (SMNR = 20 dB) in terms of Average Signal-to-Reconstruction-Noise-Ratio (ASRER) vs. Fraction of Measurements ($\alpha$), averaged over 10,000 trials. Sparse signal dimension $N = 500$ and sparsity level $K = 20$.

The performance comparison of IFSRA(MP), IFSRA(CoSaMP), and IFSRA(BPDN) are shown in Figure 7.5. For both Gaussian Sparse Signals (GSS) and Rademacher Sparse Signals (RSS), IFSRA(BPDN) resulted in a better performance than both IFSRA(CoSaMP) and IFSRA(BPDN).

### 7.4.1.1 Reproducible Research

In the spirit of reproducible research [136, 137], we provide codes at: [http://www.ece.iisc.ernet.in/~ssplab/Public/IFSRA.tar.gz](http://www.ece.iisc.ernet.in/~ssplab/Public/IFSRA.tar.gz). The codes reproduces the simulation results shown in Figure 7.2, Figure 7.3, and Figure 7.4. The folder also includes codes for clean
FIGURE 7.5: Performance comparison of IFSRA(MP), IFSRA(CoSaMP), and IFSRA(BPDN) under measurement perturbations (SMNR = 20 dB) in terms of Average Signal-to-Reconstruction-Noise-Ratio (ASRER) vs. Fraction of Measurements (α), averaged over 10,000 trials. Sparse signal dimension \(N = 500\) and sparsity level \(K = 20\).

measurement cases and noisy measurement cases with different SMNRS.

### 7.4.2 Real Compressible Signals

Most of the signals we often meet in applications are not exactly sparse. However, many of them including natural signals are found to be compressible which can be well approximated by their sparse versions. In this section, we evaluate the performance of the proposed IFSRA using real-world compressible signals.

ECG signals selected from MIT-BIH Arrhythmia Database [138] were used to conduct the experiments. ECG signals are compressible and have a good structure for sparse decompositions. We used a similar simulation setup used in [139] and [140]. ECG signals of length \(N = 1024\) were processed. Gaussian measurement matrices
with appropriate sizes were used to vary the fraction of measurements, $\alpha$, from 0.25 and 0.49 with an increment of 0.03. We assumed a sparsity level $K = 128$ and the reconstruction results are shown in Figure 7.6.

![Figure 7.6](image)

**Figure 7.6**: Performance of IFSRA compared with the parent algorithms in terms of Average Signal-to-Reconstruction-Noise-Ratio (ASRER) vs. fraction of measurements ($\alpha$) for ECG signals from MIT-BIH Arrhythmia Database [138, 141] (Signal dimension $N = 1024$ and sparsity level $K = 128$).

As in the case of synthetic signals, here also IFSRA continued to give significant ASRER improvement over the respective parent algorithms. For example, at $\alpha = 0.37$, IFSRA(MP) resulted in 14.7 dB ASRER improvement as compared to MP. IFSRA(CoSaMP) improved the ASRER by 10.6 dB as compared to CoSaMP, at $\alpha = 0.40$. At $\alpha = 0.34$, IFSRA(BPDN) showed 2.7 dB ASRER improvement over BPDN. We can observe a similar trend of ASRER improvement for other values of $\alpha$ in Figure 7.6.

### 7.5 Summary

To enhance the performance of any given arbitrary SRA, we proposed a novel iterative framework and devised a new algorithm...
referred to as IFSRA. In each iteration of IFSRA, we solve a regularized version of the original problem using the given SRA. Using RIP we derived sufficient conditions for performance improvement and convergence of IFSRA. The efficacy of IFSRA in applications was shown using extensive numerical simulations on both synthetic and real-world signals.

7.5.1 Relevant Publication


7.A Proof of Theorem 7.2 (Signal and Measurement Perturbations)

For any arbitrary signal $x$, we have

$$b = Ax + w = Ax^K + \tilde{w},$$

where $\tilde{w} = A(x - x^K) + w$. Now (7.29) may be viewed as a standard CS measurement system given in (2.4) with $x^K$ as the $K$-sparse signal and $A(x - x^K) + w$ as the measurement perturbations. Hence using Theorem 7.1 on page 156 we get

$$\|r_k\|_2 \leq \beta \|r_{k-1}\|_2 + C_1 \|\tilde{w}\|_2,$$

$$\leq \beta \|r_{k-1}\|_2 + C_1 \|A(x - x^K)\|_2 + C_1 \|w\|_2,$$

$$\leq \beta \|r_{k-1}\|_2 + C_1 \sqrt{1 + \delta_{3K}} \|x - x^K\|_2 + C_1 \frac{\sqrt{1 + \delta_{3K}}}{\sqrt{K}} \|x - x^K\|_1$$

$+ C_1 \|w\|_2$
\[
\beta \|r_{k-1}\|_2 + C_1 \sqrt{1 + \delta_{3K}} \|x - x^K\|_2 + C_1 \sqrt{1 + \delta_{3K}} \|x - x^K\|_1 + C_1 \|w\|_2
\]

(\text{using Lemma 2.2})

We have,
\[
\|x - \hat{x}_k\|_2 \leq \beta^k \|x\|_2 + C_2 \|\tilde{w}\|_2
\]
\[
\leq \beta^k \|x\|_2 + C_2 \sqrt{1 + \delta_{3K}} \|x - x^K\|_2 + C_2 \sqrt{1 + \delta_{3K}} \|x - x^K\|_1 + C_2 \|w\|_2
\]

(\text{using Lemma 2.2})
Conclusions and Future Work

“Problems worthy of attack prove their worth by hitting back.”

Piet Hein [1905-1996]

It is well known that the reconstruction quality of any Sparse Reconstruction Algorithm (SRA) depends on many parameters like dimension of the signal, level of sparsity of the signal, number of measurements, noise power, and the underlying statistical distribution of the non-zero elements of the signal. Though the performance of the SRA deteriorates in such adverse situations where these parameters do not meet the minimum requirements (which often varies from algorithm to algorithm), the SRAs may still be able to obtain a partial information about the sparse signal. In this thesis, we have proposed a novel fusion framework which employs multiple sparse reconstruction algorithms independently for signal reconstruction. We have also proposed different fusion algorithms for efficiently fusing the estimates of the participating algorithms. The analysis of the proposed fusion algorithms shows that, a judicious choice of the participating algorithms often leads to an improved signal reconstruction. A rule of thumb, that may be used
in practice, is to select the algorithms working with different principles as the participating algorithms. Based on the fusion idea, we have also proposed an iterative framework for improving the performance of any arbitrary SRA.

We developed the fusion algorithm, Fusion of Algorithms for Compressed Sensing (FACS), in Chapter 3 and showed the effectiveness and efficiency of FACS using comprehensive numerical experiments. Though FACS is shown to improve the sparse signal reconstruction, it is completely blind about the true atoms which are not included in the union of the support-sets estimated by the participating algorithms. To alleviate this drawback, another fusion algorithm \textit{viz.} Committee Machine Approach for Compressed Sensing (CoMACS) and variations of CoMACS were proposed in Chapter 4. Though FACS and CoMACS improve the signal reconstruction, the higher computational requirement makes them less attractive for low latency applications. Hence, for low latency applications, we developed a progressive fusion scheme called progressive Fusion of Algorithms for Compressed Sensing (pFACS) in Chapter 5. Unlike the other fusion algorithms, pFACS provides quick interim results and successive refinements during the fusion process, which is highly desirable in low latency applications. In Chapter 6, we extended the fusion framework and FACS for Multiple Measurement Vector (MMV) problem. Motivated by the fusion principles, we proposed an iterative framework \textit{viz.} Iterative Framework for Sparse Reconstruction Algorithms (IFSRA) in Chapter 7, which iteratively improves the sparse reconstruction performance of any arbitrary SRA.

The proposed fusion algorithms were theoretically analysed and
Chapter 8 Conclusions and Future Work

Performance guarantees were derived using the Restricted Isometry Property (RIP). The proposed algorithms were shown to be robust under both signal and measurement perturbations. It is worthwhile to note that the proposed algorithms are kept general in nature, and does not require any non-trivial modification in the underlying participating algorithm. Hence, any off-the-shelf SRA can be used as a participating algorithm in the proposed schemes. Extensive numerical experiments were carried out on both synthetic and real-world signals to show the efficacy of the proposed schemes in applications.

8.1 Scope for Future Work

While it is possible to engineer several sophisticated fusion strategies, we used a simple Least-Squares (LS) based approach in the proposed schemes. A more sophisticated fusion strategy should provide further performance improvement. In the proposed fusion algorithms, we have considered only the support-sets estimated by the participating algorithms. It would be interesting to exploit the non-zero magnitudes of the sparse signal estimated by the participating algorithm in the fusion framework. For example, we may incorporate the magnitudes of the non-zero entries of the estimates of the participating algorithms in a Bayesian framework, which may result in a better sparse signal estimate.

The theoretical analysis presented in this thesis relies on the RIP of the measurement matrix. It may be interesting to obtain tighter performance bounds using other variants of RIP like D-RIP [169] and fusion RIP [170]. Another possibility is to derive the performance guarantees using other properties of the measurement matrix, like Null Space Property (NSP) and incoherence.
An important extension of Compressed Sensing (CS) is the matrix completion problem \cite{171}, where a low-rank matrix is required to be estimated from the incomplete information. It may be possible to extend the fusion-idea to the matrix completion problem to yield a better result. It is also worthwhile to explore the possibility of applying fusion to other closely related problems such as co-sparse analysis model \cite{172} and dictionary learning \cite{173}.
Bibliography


[18] V. Kotelnikov, “On the carrying capacity of the ether and wire in telecommunications,” in *First All-Union Conference on the technological reconstruction of the communications sector and the development of low-current engineering*, Moscow, Russia, 1933. (cited on page 5).


The Twenty-Seventh Asilomar Conference on, 1993, pp. 40–44 vol.1. (cited on pages 10, 11, 36, 37).


