Scalable Sensor Network Field Reconstruction with Robust Basis Pursuit

Submitted in partial fulfillment of the requirements for
the degree of
Doctor of Philosophy
in
Electrical and Computer Engineering

Aurora C. Schmidt

M.Eng., Electrical Eng. and Computer Sci., Massachusetts Institute of Technology
S.B., Electrical Eng. and Computer Sci., Massachusetts Institute of Technology

Carnegie Mellon University
Pittsburgh, PA

May 2013

Keywords: sensor network information fusion, decentralized field inversion, ℓ_1 optimization, basis pursuit, sparse estimation, distributed average consensus, compressed sensing, structured model uncertainty

dedicated to my family,
especially to Mom and Dad for their wisdom and guidance
and to Brian my biggest supporter

Abstract

We study a scalable approach to information fusion for large sensor networks. The algorithm, field inversion by consensus and compressed sensing (FICCS), is a distributed method for detection, localization, and estimation of a propagating field generated by an unknown number of point sources. The approach combines results in the areas of distributed average consensus and compressed sensing to form low dimensional linear projections of all sensor readings throughout the network, allowing each node to reconstruct a global estimate of the field. Compressed sensing is applied to continuous source localization by quantizing the potential locations of sources, transforming the model of sensor observations to a finite discretized linear model. We study the effects of structured modeling errors induced by spatial quantization and the robustness of ℓ_1 penalty methods for field inversion. We develop a perturbations method to analyze the effects of spatial quantization error in compressed sensing and provide a model-robust version of noise-aware basis pursuit with an upperbound on the sparse reconstruction error. Numerical simulations illustrate system design considerations by measuring the performance of decentralized field reconstruction, detection performance of point phenomena, comparing trade-offs of quantization parameters, and studying various sparse estimators. The method is extended to time-varying systems using a recursive sparse estimator that incorporates priors into ℓ_1 penalized least squares. This thesis presents the advantages of inter-sensor measurement mixing as a means of efficiently spreading information throughout a network, while identifying sparse estimation as an enabling technology for scalable distributed field reconstruction systems.

Acknowledgments

I would like to acknowledge each of the members of the thesis committee, Professors José M. F Moura, Rohit Negi, Bruno Sinopoli, João M. F. Xavier, and Petros T. Boufounos, for their support, insights, and brilliant guidance. I have learned a great deal from each of you. Thank you to Dr. Boufounos for teaching me DSP in my formative years, and for the continuing amazing advice and feedback. Thanks to Professor Negi for sharing his passion for information theory and research, Professor Sinopoli for sharing his experience with networked control systems, and Professor Xavier for introducing the beautiful world of nonlinear optimization and sharing insights in so many connected fields of signal processing and control.

Above all, I am grateful to my graduate advisor, Professor Moura, for sharing his inspiring philosophy and approach to discovery, and without whose dedication and encouragement this work would not have been possible.

I would also like to thank a few of the numerous friends and colleagues whose assistance has been invaluable. To Prof.'s Jelena Kovačević and Marija Illić, thank you for the wonderful insights and encouragement. I will always be thankful for having a chance to learn from you. Thanks very much to Carolyn Patterson and Claire Bauerle who made complex tasks seem simple while never failing to brighten my spirits. Thanks to my officemates and colleagues who have shared so much wisdom and good times with me; Soummya Kar, Divyanshu Vats, Marek Telgarsky, Kyle Anderson, Subhro Das, Stephen Kruzick, Matthew Baron, and Liangyan Gui; as well as all the members of the Signal Processing/Networks group, Aliaksei Sandryhaila, João Mota, Nicholas O'Donoughue, Usman Khan, Joel Harley, June Zhang, Joya Deri, Sérgio Pequito, Dusan Jakovetic, Dragana Bajovic, and Augusto Santos. These group members, and all in Porter Hall B-Level, made for an interesting, fun, and inspiring work environment. Particular thanks to Akshay Rajhans, Rohan Chabukswar, Nikos Aréchiga, Kyri Baker, Sanja Cvijic, Jhi-Young Joo, Andrew Hsu, Nipun Popli, Jon Donadee, Javad Mohammadi, Matthias Althoff, and Jim Weimer, for always being ready to run a practice talk, have coffee break, or participate in a Markov gift swap. Thanks also to Kevin Gimpel, Sarah Loos, Min Xu, Yuko Ishii, Joseph Bradley, and Ian Wong, for the all the dancing/games/kayaking. Thank you, dear friends. I will continue to draw inspiration and joy from these memories.

This work was partially supported by the Office of Naval Research under the Multidisciplinary University Research Initiative, MURI N000140710747.

Contents

1 Introduction **1**
- 1.1 Contributions . . . 2
- 1.2 Motivation . . . 3
- 1.3 Previous Work . . . 5
- 1.4 Outline . . . 7

2 Observation System Modeling for Field Inversion **11**
- 2.1 Example Application: Acoustic Source Localization . . . 14
- 2.2 Conclusion . . . 15

3 Field Inversion by Consensus and Compressed Sensing **17**
- 3.1 The FICCS Algorithm . . . 17
  - 3.1.1 Network Set-up . . . 18
  - 3.1.2 Sensing Mode using Consensus . . . 19
  - 3.1.3 Field Inversion Mode using Compressed Sensing . . . 20
- 3.2 Design of Measurement Projection Weights . . . 22
- 3.3 Consensus Weight Design and Scaling . . . 23
- 3.4 Example Recovery Using FICCS . . . 26
- 3.5 Simulation Testing of FICCS . . . 28
- 3.6 Summary . . . 31

4 Structured Model Error and Model-Robust Basis Pursuit **33**
- 4.1 Perturbations to the Linearized Model . . . 35
- 4.2 Lower Bound on Error of Basis Pursuit . . . 37
- 4.3 Upper Bound on the Error of Model-Robust Basis Pursuit . . . 40
- 4.4 Analysis of Spatial Quantization Trade-offs in Power Law Propagation . . . 42
  - 4.4.1 Model Error for Inverse Square Distance Propagation: One Dimension . . . 42
  - 4.4.2 Coherence for Inverse Square Distance Propagation: One Dimension . . . 44
  - 4.4.3 Numerical Evaluation of $g_{\max}$ and μ versus 2D Lattice Spacing and Power Law γ 48
- 4.5 Conclusions . . . 49

5 Motivating Applications **53**
- 5.1 Acoustic Sources Localization and Infrared Fields, $\gamma = 2$ . . . 53
- 5.2 Radar Surveillance Network, $\gamma = 4$ . . . 55
- 5.3 Guided Wave Applications, $\gamma = 1$ . . . 57
- 5.4 Conclusions . . . 58

6 Numerical Simulation Testing in Surveillance Applications **59**
6.1 Overview of Experiments and Tradeoffs . . . 60
6.2 Studies of Estimation Tradeoffs . . . 61
6.2.1 Cone Expansion Factor β . . . 62
6.2.2 ℓ_1-Norm Penalization Factor λ . . . 66
6.2.3 Quantization Lattice Spacing δ . . . 68
6.2.4 Positivity Constraints . . . 70
6.2.5 Normalization of Observation Matrices . . . 72
6.3 Studies of System Tradeoffs . . . 76
6.3.1 Signal-to-noise Ratio . . . 76
6.3.2 Random versus Regular Sensor Placement . . . 77
6.3.3 Propagation Power Law . . . 81
6.4 Conclusions . . . 86

7 Time-Varying Field Inversion Using Dynamically Re-Weighted ℓ_1 Penalized Least Squares **89**
7.1 Dynamic Target Support Model . . . 90
7.2 Dynamically Re-weighted ℓ_1 Penalized Least Squares . . . 91
7.3 Example Simulation Experiment . . . 92
7.4 Simulation Experiments . . . 93
7.5 Conclusions . . . 94

8 On the Benefits of Measurement Mixing **99**
8.1 Problem Formulation . . . 100
8.2 Linear Gaussian Model . . . 101
8.3 Bernoulli Locations and Gaussian Intensities . . . 107
8.4 Conclusions . . . 110

9 Conclusions and Directions for Future Work **113**

Appendices **115**

A Scaling of mixing time with N in the 2D 8-Nearest-Neighbor Graph **117**

B Error Stability of Model Robust Basis Pursuit **121**

C Model Error for 1D System **125**
C.1 Evaluating $2\int_0^{\pi/T} e^{-2\omega}\cos(t_0\omega)d\omega$ by parts . . . 127

D Coherence for 1D System **129**
D.1 Energy of $x(nT)$ on the Infinite 1D Lattice . . . 129
D.2 Inner product of adjacent bases, $x(nT)$ and $x((n-1)T)$ . . . 131
D.3 Expression for μ . . . 133

Bibliography **135**

List of Figures

2.1 Inverse square distance propagation field with three continuously located sources (panel a) and the approximation to a discrete lattice (panel b). Dashed squares represent the regions that each source is associated with. The corresponding field values are shown in contour plots in panels c and d, respectively. 13

3.1 Two ways to deploy sensors with respect to the discrete lattice of potential source locations. The left panel shows the optimal choice of regular placement, while the right panel shows random, but even, placements. 18

3.2 Example recovery estimates for a sensor network with $N = 1024$ sensors placed in a 32-by-32 grid, where intersensor spacing is $\delta = 2$. The left panel shows all sensor readings; additive noise has variance $\sigma^2 = 0.01$. Middle and right panels show basis pursuit estimates for 300 randomly selected sensor measurements and 300 random orthonormal projections computed via consensus. The true source locations of the 10 sources are overlayed in white crosses. The ℓ_2 recovery errors were 20.75 and 2.09 for panels (b) and (c), respectively. 27

3.3 Field inversion performance as a function of the number of projections m, for noiseless measurements (left) and $\sigma_2 = 0.2$ (right). Mean squared error of the source vector is plotted for various choices of projection matrices, Φ, and compared to inversion using m point samples or randomly mixed samples. 29

3.4 Comparison of the asymptotic convergence rates of FDLA in the 8 nearest-neighbor grid graph versus a random Erdös-Rényi, and Ramanujan graphs with the same number of edges. Each were computed by numerically, except in the case of the Ramanujan graph with represents a lower bound on the number of mixing in an 8-regular graph, vs. N. 30

3.5 A comparison of center source detection performance in 10 random sensor deployments with the restriction that inter-sensor distance is at least $\frac{1}{2}$, Left panel shows an example deployment. Each series in (b) and (c) corresponds to a different random network deployment. Receiver operator characteristic for the center source is averaged over 100 experiments. 31

4.1 Histogram for 100 trials of computed basis pursuit error lower bounds for 3 randomly located acoustic sources each with magnitude 15. One hundred sensors cover the area in a lattice arrangement with spacing of 2. They perform consensus to find m projections of a randomly drawn orthonormal matrix, Φ. Three values of m are compared. The sensor noise vector $\mathbf{z}$ is drawn uniformly from vectors with norm 0.3526. Values of g_i range from 0.7526 to 0.7532. 39

4.2 Relationships between the four compared sparse recovery methods. All methods shown include the optional positivity constraint. 41

4.3 Illustration of the propagated field at the 0 position (upper) and at the 0 position plus an offset t_0 (lower). 42

4.4 Worst case model error versus discretization spacing T for one-dimensional lattice and inverse square distance propagation. Left panels show squared error for the worst case offset, $t_0 = T/2$. The red asterisk series using the formula of (4.31), assuming no aliasing, while the blue dots series is the result of sampling and computing of the error for 2 million samples. Right panels: showing the ratio of the model error to the total squared energy $x(nT)$. The upper panels zoom in to $0 < T \leq 2$, while the lower left panels show a larger range of T. 45

4.5 Coherence study for one dimensional lattice with inverse square propagation law: Each panel compares the derived quantity versus the measured value computed using 10,000 sensor samples and spacing T. Panel (a) shows the squared energy of $x(nT)$, panel (b) shows the innerproduct of adjacent columns of Ψ, and panel (c) shows the resulting mutual coherence versus T. Panel (d) shows measured coherence for a larger range of T. 47

4.6 Worst case model error and coherence versus discretization spacing T for a 1D lattice and various inverse power propagation laws. The power γ is varied from 0.5 to 4, and 10,000 sensors are modeled. The left plot of panel (a) shows squared error for the worst case offset, $t_0 = T/2$, versus lattice spacing T, while the right shows the ratio of the model error to the total squared energy $x(nT)$. The left plot of panel (b) shows signal energy and innerproduct of adjacent bases, while the right plot shows the resulting mutual coherence vs. T 49

4.7 Illustration of simulation of worst case modeling error in the 2D lattice for $T = 1/2$. Panel (a) shows a source exactly at a gridpoint with contours based in inverse squared distance, Panel (b) shows the source at location $[T/2, T/2]$ away from the grid point. The black box denotes the quantization region corresponding to the origin. Selected contours are labeled with the field level values. 50

4.8 Worst case model error and coherence versus discretization spacing T for a 2D lattice and various inverse power propagation laws. The power γ is varied from 0.5 to 4, using 160,000 sensors. The left plot of panel (a) shows squared error for the worst case offset, $(T/2, T/2)$, versus lattice spacing T, while the right shows the ratio of the model error to the total squared energy $x(nT)$. The left plot of panel (b) shows signal energy and innerproduct of adjacent bases, while the right plot shows the resulting mutual coherence vs. T 51

5.1 Depiction of the distributed radars and monitored discretized space with 3 targets, red circles, in (a), and, in (b), a close-up showing the potential location variation within a voxel. 56

6.1 Study of Feasibility of $\mathbf{v}_{\text{true}}$ vs. β: Panel (a) illustrates the set-up of the experiments, which use a regular grid of 25 sensors spaced by distance 2 and source lattice spacing $\delta = 1.5$. The maximum perturbation constant and its source location are also shown. Panels (b-d) show probability of feasibility over 1000 trials of the original source vector for residual cone constraints expanded using varying β, using $m = 18$ random orthonormal network projections. The series in (b) compare various propagation factors, for Gaussian measurement noise with variance 0.01. Panels (c) and (d) show the feasibility for varying measurement noise levels σ_n^2 in the case of Gaussian noise, and random noise with norm b_{σ^2}, respectively. 63

6.2 Number of Feasible Trials and System Parameters vs. m, for two source lattice sampling rates, $u = 1, 3$ with $\delta = 2, 2/3$, $N = 150$ randomly placed sensors, random additive noise of norm $b_{0.001}$. Panel (a) shows feasibility of BP and (P)MRBP for various β at each lattice sample rate. Panel (b) compares average coherence μ to model perturbation constants g_{max} for the two lattice sample rates. 64

6.3 β Study: Recovery error and detection rates of FICCS for two source lattice sampling rates, $u = 1, 3$ with $\delta = 2, 2/3$, over 100 trials, 150 randomly-placed sensors, random additive noise of norm $b_{0.001}$; $\lambda = 1$ and $m = 30$ projections are used. Top panels show shows the average ℓ_2 error and detection rate vs. β for estimators using the un-normalized system matrix A and fixed false alarm rate 0.01. Bottom panels show average ℓ_2 error and detection rates vs. β for estimators using a normalized system matrix A. 65

6.4 λ Study : Recovery error and detection rates of FICCS for two source lattice sampling rates, $u = 1, 3$ with $\delta = 2, 2/3$, 100 trials. $N = 150$ randomly-placed sensors, random additive noise of norm $b_{0.001}$; $\beta = 0.5$ and $m = 30$ projections are used. Top panels show shows the average ℓ_2 error and detection rate vs. λ for estimators using the un-normalized system matrix A and fixed false alarm rate 0.01. Bottom panels show average ℓ_2 error and detection rates vs. λ for estimators using a normalized system matrix A. 67

6.5 Performance of FICCS, averaged over 100 trials, vs. quantization spacing. In both, $N = 100$ sensors are distributed in a regular grid, spaced by the same δ as the source quantization lattice, using $m = 30$ network projections and additive noise has an exact norm $b_{0.01}$. Columns of A are not normalized, in estimators. Panel (a) shows average ℓ_2 estimation error of the sparse estimators, using $\lambda = 2$, while (b) shows the detection rates for a fixed false alarm rate of 0.001. 68

6.6 μ and $g_{\max}$ versus quantization lattice spacing. 69

6.7 Challenging Case: 100 Trial Average FICCS ℓ_2 error and detection rates vs. number of network projections m, for 100 sensors in a 10-by-10 grid, source lattice spacing $\delta = 4$. The propagation law corresponds to a Radar test, with $\gamma = 4$. Additive noise has an exact norm $b_{0.01}$, and penalization parameter is $\lambda = 2$. 70

6.8 Effects of Positivity Constraints: Propagation factor $\gamma = 1$ (data from study of figures 6.13 and 6.14), source lattice spacing $\delta = 2$, 100 recovery trials, additive noise with norm $b_{0.01}$, and $\lambda = 2$. Left panels comparing avg. ℓ_2 error or detection rate of sparse estimators using the positivity constraint on $\mathbf{v}$ versus none, plot the performance of both with unconstrained in dashed lines. Right panels show direct differences due to requiring $\mathbf{v} \geq 0$. Upper panels (a-d) use a system with 100 sensors in a regular grid, while lower panels (e-h) have 150 randomly placed sensors. 71

6.9 Effects of Positivity Constraints on Estimation Performance in the propagation factor $\gamma = 2$ study. Upper panels show mean ℓ_2 error over 100 trials, while lower panels show detection rates at a fixed false alarm rate. Panels (b) and (c) show the increase of detection rate using positivity constraints over unconstrained optimization. Panels (e) and (f) show the average of reduction of ℓ_2 recovery error of using the positivity constraint. Simulations correspond to those the study of figure 6.16, which uses inverse squared distance propagation loss, 150 randomly placed sensors, source lattice spacing $\delta = 2$, 100 recovery trials, with additive noise with an exact norm b for $\sigma^2 = 0.01$. Sparse estimators use $\lambda = 2$ and $\beta = 0.40$. 73

6.10 Un-normalized vs. Normalized A in for $\gamma = 4$, 150 randomly-placed sensors, (study also in figure 6.18). Lattice spacing $\delta = 2.0$, 100 recovery trials, with random additive noise with norm $b_{0.01}$. Sparse estimators use $\lambda = 2$ and $\beta = 0.10$, and contrained $\mathbf{v} \geq 0$. Upper panels show average ℓ_2 estimation error and lower panels show detection rates, as a function of the number of network projections m. 74

6.11 Un-normalized vs. Normalized A for $\gamma = 1$ Study, (from figure 6.14): Recovery error and detection Performance, as a function of the number of network projections m. Lower panels show detection rates at a fixed false alarm rate, upper panels show average ℓ_2 error. 150 Sensors are randomly placed, 100 simulations trials use lattice spacing $\delta = 2.0$, additive noise with norm $b_{0.01}$, $\lambda = 2$, and $\beta = 0.10$, and no constraints on the signs of $\mathbf{v}$. 75

6.12 SNR Study: Performance of FICCS averaged over 100 trials, with $N = 100$ regularly spaced sensors with $\delta = 2$, using $m = 30$ projections, additive noise has an exact norm b, and propagation constant $\gamma = 4$. The constant b is adjusted to provide the various SNR levels. Columns of A are not normalized in the estimators. Panel (a) shows average ℓ_2 estimation error of the sparse estimators, using $\lambda = 2$, while (b) shows the detection rates for a fixed false alarm rate of 0.001. 77

6.13 Study with $\gamma = 1$, Lattice vs. Random Sensor Placement, Estimators using Positivity constraints. Left panels show detection rates at a fixed false alarm rate, while right panels show the difference of the detection rate using positivity constraints minus the rates when the signs of $\mathbf{v}$ are unconstrained. The upper panels use a system with 100 sensors in a regular grid, while lower panels have 150 randomly placed sensors. Simulations correspond to source lattice spacing $\delta = 2$, 100 recovery trials, with additive noise of norm $b_{0.01}$. Estimators use $\lambda = 2$ and $\beta = 0.10$. 79

6.14 Study with $\gamma = 1$, Lattice vs. Random Sensor Placement, Estimators not using positivity constraints. Left panels show detection rates at a fixed false alarm rate, while right panels show the difference of the detection rate using positivity constraints minus the rates when the signs of $\mathbf{v}$ are unconstrained. The upper panels use a system with 100 sensors in a regular grid, while lower panels have 150 randomly placed sensors. Simulations correspond to source lattice spacing $\delta = 2$, 100 recovery trials, with additive noise of norm $b_{0.01}$. Estimators use $\lambda = 2$ and $\beta = 0.10$. 80

6.15 Coherence μ and perturbation constants $g_{\max}$ for each power law propagation exponent $\gamma = 2, 1, 4$, using lattice spacing $\delta = 2$. Left panels show these for a system of 100 sensors in a regular grid. Right panels show these for 150 randomly placed sensors. . . . 82

6.16 Study with $\gamma = 2$, Lattice vs. Random Sensor Placement, Estimators Using positivity constraints. Left panels show detection rates at a fixed false alarm rate, while right panels show the difference of the detection rate using positivity constraints minus the rates when the signs of $\mathbf{v}$ are unconstrained. The upper panels use a system with 100 sensors in a regular grid, while lower panels have 150 randomly placed sensors. Simulations correspond to source lattice spacing $\delta = 2$, 100 recovery trials, with additive noise of norm $b_{0.01}$. Estimators use $\lambda = 2$ and $\beta = 0.10$, and normalized A. 83

6.17 Study with $\gamma = 2$, Lattice vs. Random Sensor Placement, Estimators not using positivity constraints. Left panels show detection rates at a fixed false alarm rate, while right panels show the difference of the detection rate using positivity constraints minus the rates when the signs of $\mathbf{v}$ are unconstrained. The upper panels use a system with 100 sensors in a regular grid, while lower panels have 150 randomly placed sensors. Simulations correspond to source lattice spacing $\delta = 2$, 100 recovery trials, with additive noise of norm $b_{0.01}$. Estimators use $\lambda = 2$ and $\beta = 0.10$, and normalized A. 84

6.18 Study with $\gamma = 4$, Lattice vs. Random Sensor Placement, Estimators Using positivity constraints. Left panels show detection rates at a fixed false alarm rate, while right panels show the difference of the detection rate using positivity constraints minus the rates when the signs of $\mathbf{v}$ are unconstrained. The upper panels use a system with 100 sensors in a regular grid, while lower panels have 150 randomly placed sensors. Simulations correspond to source lattice spacing $\delta = 2$, 100 recovery trials, with additive noise of norm $b_{0.01}$. Estimators use $\lambda = 2$ and $\beta = 0.10$, and un-normalized A. 85

6.19 Study with $\gamma = 4$, Lattice vs. Random Sensor Placement, Estimators not using positivity constraints. Left panels show detection rates at a fixed false alarm rate, while right panels show the difference of the detection rate using positivity constraints minus the rates when the signs of $\mathbf{v}$ are unconstrained. The upper panels use a system with 100 sensors in a regular grid, while lower panels have 150 randomly placed sensors. Simulations correspond to source lattice spacing $\delta = 2$, 100 recovery trials, with additive noise of norm $b_{0.01}$. Estimators use $\lambda = 2$ and $\beta = 0.10$, and un-normalized A. 86

7.1 Diagram of the slow observation time-scale, in which a single source may disappear or appear, versus the fast network communication time-scale, in which the m distributed averages are computed using distributed consensus 91

7.2 Example Simulation Test: The number of sources, panel a, and ℓ_2 estimation error, panel b, versus time-step of FICCS estimates. Panel (b) compares the mean ℓ_2 estimation error of performing independent L1LS at each new time-step with only $m = 15$ network projections versus the dynamically re-weighted L1LS which uses the estimate of the previous time-step as a prior. 92

7.3 Example Simulation Test: Comparison of the source field estimates of L1Ls, panels (b) and (d), vs. the sequential estimator Wtd. L1LS, panels (c) and (e), at time-steps 5 and 20. Both estimators start with the initial L1LS $\widehat{\mathbf{v}}_0$, shown in panel (a), which uses 30 network projections to compute the initial estimate. Correct source locations as white crosses. 93

7.4 Tests of dynamic reweighted ℓ_1 penalty, for acoustic source simulation, source and sensor lattice spacing of $\delta = 2$, averaged over 100 trials. The initial number of network projections at $t = 0$ is $m_0 = 30$, with each subsequent time-step using only 15 network projections. Panel (a) shows the average number of sources versus time-step, in blue $\times$'s, with black error bars marking the $\pm$ one standard deviation and light blue series showing individual traces of number of sources vs. time. Panels (b) and (c) show average estimation error in the case where source fall exactly on grid points, on the left, and uniformly continuously in the space (i.e. off-grid) on the right. The blue $\times$'s show the performance of ℓ_1 penalty with no information from prior estimates, while green $+$'s denote the performance of the dynamically reweighted ℓ_1 penalty estimator. Panels (d) and (e) show detection rates for average false alarm rate of 0.0001 and 0.01, respectively. 95

7.5 Tests of dynamic reweighted ℓ_1 penalty, for acoustic source simulation, source and sensor lattice spacing of $\delta = 1$, averaged over 100 trials. The initial number of network projections at $t = 0$ is $m_0 = 40$, with each subsequent time-step using 20 network projections. Panel (a) shows the average number of sources versus time-step, in blue $\times$'s, with black error bars marking the $\pm\sigma$ and light blue series showing individual traces of number of sources vs. time. Panels (b) and (c) show average estimation error in the case where source fall exactly on grid points, on the left, and uniformly continuously in the space (i.e. off-grid) on the right. The blue $\times$'s show the performance of ℓ_1 penalty with no information from prior estimates, while green $+$'s denotes the performance of the dynamically reweighted ℓ_1 penalty estimator using the values of the previous time-step to compute weights. Panels (d) and (e) show detection rates for average false alarm rate of 0.001. 96

8.1 The two error systems; in the upper system the observation matrix is restricted to a simple selection matrix H, in the lower system, Φ allows the production of any linear mapping. 101

8.2 Numerical tests for 10 Known Source Locations, 2 noise levels; left plots: $\sigma_n^2 = 0.01$, right plots: $\sigma_n^2 = 1$. $N = 1024$, m is varied, $\mu_s = 20$, $\sigma_s^2 = 1$. Ψ the inverse squared distance matrix. Upper panels compare bounds and estimated mutual information for mixed versus selected measurements. In the selection case, both random and best greedy selection are shown. Lower panels compare squared estimation error for the MMSE estimate. 106

8.3 Numerical tests for 2 noise levels; left plots: $\sigma_n^2 = 0.01$, right plots: $\sigma_n^2 = 1$. $N = 1024$, m is varied, $\mu_s = 20$, $\sigma_s^2 = 1$. Ψ the inverse squared distance matrix. Upper panels compare bounds and estimated mutual information (5000 samples) for various choices of Φ and Ψ. Lower panels compare average squared estimation error, 1000 trials, for MMSE, MAP, and ℓ_1-minimization estimators, for orthonormal Φ versus random selection. 109

List of Tables

3.1 Order of algorithm scalings with N, the number of nodes in the 8 nearest neighbor 2D lattice graph. 26

Chapter 1

Introduction

Our interest is in the problem of coordinating networked sensors that monitor a large area to reconstruct spatially distributed fields. Applications such as environmental monitoring, distributed surveillance, and early warning systems depend on information fusion strategies that scale with increasingly large networks. We focus on distributed inference techniques for their resilience to individual node failures as well as their avoidance of the centralized collection of increasingly large data sets. The approach we develop, called field inversion by consensus and compressed sensing (FICCS), is in two steps; 1) a consensus step, where sensors cooperate with neighbors to iteratively update their states by exchanging low-dimensional projections of the measurements of all sensors; and 2) a compressed sensing step, in which any sensor may invert the field parameters using efficient sparse optimization methods. Often, the models for physical fields are described by a small number of system parameters. We leverage advances in sparse estimation to allow for the accurate reconstruction of large fields using efficiently computed projections of network measurements.

This work considers distributed field inversion; i.e., determining the locations of an unknown number of sources with unknown amplitudes from measurements collected by a sensor network. The classical solution to field inversion collects all sensor measurements to a centralized location, or *fusion center*, and applies least-squares (LS) or maximum likelihood (ML) optimization strategies to estimate the parameters of the field. Traditional least-squares approaches are not robust to noise or undersampling, taking no advantage when the number of sources is small. On the other hand, ML methods can give preference

to sparse solutions, but lead to high-dimensional nonlinear optimizations, which are not scalable as the size of networks grow. We depart from these approaches by proposing a distributed algorithm combining consensus and compressed sensing.

1.1 Contributions

This thesis makes contributions toward the goal of scalable distributed network inference along the following dimensions: 1) modeling and casting distributed field inversion problems into the framework of average consensus and compressed sensing; 2) describing a scalable network inversion algorithm; 3) addressing quantization error that arises as structured uncertainty to the linear model; 4) proposing a model-robust version of basis pursuit; 5) studying tradeoffs in system parameter choices in various settings; 6) extending the distributed field inversion approach to time-varying systems; and 7) demonstrating the improvement of information content of observations when sensors mix their values with others sensors.

We detail a fully decentralized scalable algorithm for field inversion, FICCS, that combines consensus and compressed sensing. The strategy follows from early observations that, by iteratively averaging their own values with those of their neighbors, sensor nodes in a network may increase the information locally available. Equipped with a generative linear model for point phenomena, sensors then use low dimensional projections of global information to invert the field by casting reconstruction as a sparse optimization problem. Using this approach, the total number of messages passed in the network may grow with order $N \log(N)$ in the number of nodes in the network, representing a comparatively slow growth rate for a strategy that is inherently broadcast in nature.

We address an issue fundamental to the application of compressed sensing to many real world problems; the modeling error induced by quantizing continuous parameter spaces. The issue arises when discretizing a model of sparse phenomena that reside in a continuous location space. To cast the problem as a finite dimensional discrete linear one, generally the location space is quantized, and the system is modeled based on phenomena approximated at the quantized locations. This introduces a structured uncertainty into the linear system, for which we devise a method for analyzing the structured errors from a perturbations perspective.

A model-robust version of noise-aware basis pursuit is presented that allows for use of a residual cone constraint in sparse optimizations. This model-robust basis pursuit (MRBP) estimator comes with a stability result derived from easily verifiable coherence conditions of the observation system. In practice, the popular ℓ_1 penalized least squares (ℓ_1 penalty) shows robustness to this type of modeling error, and a hybrid version of penalized least squares, penalized model-robust basis pursuit (PMRBP), trades between the performance of the two approaches while displaying lower parameter sensitivity.

Through the use of comprehensive numerical studies, we demonstrate the distributed field inversion technique and benefits of model robust sparse estimation methods. We characterize the impacts of the number of network projections computed, choice of spatial quantization parameters, optimization trade-off parameters, and measurement noise levels.

We present a recursive version of of the distributed field inversion approach, called dynamically re-weighted ℓ_1 penalized least squares, that adapts estimation to time-varying systems where point phenomena may randomly appear and dissappear.

1.2 Motivation

To place this work in the context of real-world problems, we give a few example applications motivating the development of efficient distributed field reconstruction and estimation techniques, such as wide area surveillance with distributed radar or infrared camera networks and tsunami detection using waterheight samples of distributed buoys. In each case, we imagine large networks where the overhead of centrally collecting all sensor measurements becomes too burdensome. Rather, we look for a decentralized method of passing information between neighboring sensor nodes and inverting the global field.

In distributed radar and infrared sensor networks, there are several potential approaches to cooperative detection. We envision a simple case in which sensors aggregate angular information into an omni-directional total received energy sample. This necessitates cooperation among the sensors since multiple targets in the monitored field will superimpose, making inference not possible from the information of a single sensor. Collectively, the whole network of sensors has enough information to disambiguate the locations and intensities of a small number of targets in the field. Assuming that sensors communicate wirelessly, we seek approaches where the number of messages needed for distributed reconstruction of

the target parameters does not grow too large as additional sensors are added or the monitored area is increased.

Another type of application is in distributed geophysical sensing. A pressing example is that of giving accurate warnings of rare but dangerous tsunami events to coastal communities. The National Oceanographic and Atmospheric Administration (NOAA), through its Deep-ocean Assessment and Reporting of Tsunamis (DART) system, deploys networked surface buoys to monitor ocean height and aid in detecting tsunami waves. The current system uses seismic information with event-triggered ocean height collections to estimate the origin, magnitude, and long-distance propagation details of tsunami waves across the ocean, [1, 2]. However, currently the locations of origin of seismic events must be modeled at a number of pre-selected locations, and the small number of buoys are used only to verify the presence of large waves.

We can imagine a future system with increasing large networks of buoys capable of detecting destructive waves purely based on collective cooperative inference using ocean height readings, thereby reducing sensitivity to selected origin locations. In addition, some events, such as coastal landslides, have been known in rare cases to cause tsunamis without the involvement of underwater seismic activity. A large system would have to deal with increasing amounts of data to be transmitted as well as sensitivity to single point failures in the network or at data fusion centers. To that end, developing a distributed method for sensors to communicate with neighbors and reconstruct source parameters of hazardous waves would allow the size of the system to be scaled up without incurring ballooning costs, improving the lifetime of buoys, and reducing the amount of data transmitted through satellite networks.

Distributed networked sensors are used in a wide variety of applications, from temperature monitoring of crops to structural health monitoring of buildings and bridges. Our focus is on applications that require sensors to coordinate and exchange information in order to achieve the desired detection or estimation objectives.

In much of our analysis, we will study the toy problem of localizing acoustic sources observed by distributed isotropic sensors, with simplified physical models. This example is useful for its low single-sensor observability, necessitating coordination among the sensors to achieve acceptable reconstruction, detection, and localization performance. We propose methods for this class of distributed estimation

problems that scale with increasing numbers of sensors, avoiding problems associated with centralized approaches, such as bottlenecks that increase latency or brittleness of the system.

1.3 Previous Work

Much of early multisensor estimation work falls into the fields of array processing or decentralized detection. While major innovations in array processing have enabled a host of revolutionizing technologies, most require a centralized collection of sensor measurements. In this work, we aim to find alternatives to centralized schemes. The idea of decentralized detection using a network of sensors was studied by Tsitsiklis, [3], and there are many results framed in this context of hypothesis testing for conditionally independent sensor measurements. This challenging field has opened the door to many advances, leveraging a variety of disciplines such as filtering theory, graph theory, distributed optimization, and Markov processes. By studying an example problem of acoustic source localization within a sensor network and borrowing from new techniques in sparse signal reconstruction, we further characterize and expand the set of problems that can be treated by decentralized detection and estimation methods.

The problem of localizing an acoustic source using received signal strength measurements from multiple sensors has been well studied for the case of a single source. There are a number of distributed algorithms that may be performed using local communications in the sensor network to localize a single source, such as in [4, 5, 6]. In [5] Rabbat, Nowak, and Bucklew show how localization can be posed as an averaging problem of the locations of sensors with detections, which are then computed using distributed averaging gossip techniques. However, this method along with requiring a single source, assumes that source presence can be reliably detected with measurements local to each sensor. In more complex problems where coordination is needed to gain observability, such as the tsunami example above, this approach is not general enough. In [6], Blatt and Hero note the nonconvexity of computing the maximum likelihood estimate of a source's location and magnitude, but show how a reformulation of the problem to projection onto convex sets yields efficient global solutions that can be implemented in a distributed way. However, it is not clear how to extend these methods to handle multiple active sources.

We will consider cases where there may be a finite number of sources present and the number is not known *a priori*. We assume a simplified propagation model, namely free-space propagation that follows

the inverse square power law, so that the observation model is well known as a function of each source location and magnitude. First, we consider the case where the sources can only lie on a discrete lattice of positions, later generalizing to continuously located sources. Since the discrete lattice has only a finite set of locations, one approach is to treat localization as an M-ary hypothesis testing problem, using a generalized likelihood ratio test to determine the most likely locations of a set of sources.

Let H_0 represent the case of no sources present, H_1 a source at the first location, H_2 a source at the second location, ..., H_{L+1} two sources at location 1 and 2, and so on up to H_M the case of sources present everywhere in the lattice. There are $M = 2^L$ such hypotheses to consider, where L is the number of lattice points. This represents an exponentially growing problem that is quickly intractable for moderate L. Let the vector, $\mathbf{v} \in \mathbb{R}^L$, correspond to the field magnitudes at all of the locations in the lattice. The observation model for this discrete problem is written as the linear superposition of sources.

$$\mathbf{y} \quad = \quad A\mathbf{v} + \mathbf{n} \tag{1.1}$$

where the columns of A contain the predicted sensor readings for a unit magnitude source at each location, and $\mathbf{n}$ represents additive independent zero-mean Gaussian noise. The maximum likelihood hypothesis test solves

$$\min_{i \in [1,2^n]} \left\{ J_i = \min_{\mathbf{v}_{T_i}} \|\mathbf{y} - A_{T_i}\mathbf{v}_{T_i}\|_2^2 \right\} \tag{1.2}$$

where T_i represents the set of indices where sources are present under H_i, and the matrix A_{T_i} and vector $\mathbf{v}_{T_i}$ represent restrictions of the columns of A and the nonzeros of $\mathbf{v}$ to the set T_i. Note that an assumption here is that the number of sensors is at least as large as L, so that the minimization J_i is well-posed. We can see that considering all M hypotheses is not necessary for this minimization, since the value of J_M must always be less than or equal to all other J_i. In this formulation, the system will choose the most complicated hypothesis, corresponding to having sources everywhere, to explain the data.

In practical systems, we have strong reasons to prefer sparse hypotheses; i.e., explanations involving only a small number of sources. In a Bayesian estimation context, this may correspond to the prior knowledge that observing many sources is unlikely as compared with a few active sources. One way to account for this is to add a penalty to the cost, shown in (1.3), related to the number of sources

hypothesized.

$$\min_{i\in[1,2^n]}\left\{J_i=\min_{\mathbf{v}_{T_i}}\|\mathbf{y}-A_{T_i}\mathbf{v}_{T_i}\|_2^2+\lambda|T_i|\right\} \tag{1.3}$$

Another possibility is to use an upperbound on the expected norm of the noise vector, $\|\mathbf{n}\|_2 \leq \epsilon$, and then solve for the simplest hypothesis that satisfies the error bound, as in (1.4).

$$\min_{i\in[1,2^n]}\left\{|T_i| \text{ subject to } \min_{\mathbf{v}_{T_i}}\|\mathbf{y}-A_{T_i}\mathbf{v}_{T_i}\|_2\leq\epsilon\right\} \tag{1.4}$$

Both (1.3) and (1.4) are combinatorial problems in that, at worst, all 2^L hypotheses must be considered, and, at best, at least $\binom{L}{S}$ must be considered where S is the smallest integer corresponding to the number of sources required to meet the feasibility condition of (1.4). However, the insights of compressed sensing have shown that, for certain A, a sparse $\mathbf{v}$ can be approximately recovered using relaxations of these minimizations, where the cardinality of the support is replaced by the ℓ_1-norm of $\mathbf{v}$. We will leverage these efficient estimation methods for sensor networks observing a sparse set of sources and discuss the details of this in the following sections.

It is worth mentioning here another efficient and well-known method for estimating the locations of sources, which is a generalization of the matched filter or correlation filter, originally developed by D. O. North for radar target detection, [7]. The correlation filtering procedure corresponds to computing the correlations between columns of A and the received measurement vector $\mathbf{y}$, and then choosing the detections to be at locations where $\mathbf{r} = A^T\mathbf{y}$ has magnitude above some threshold. For well-chosen thresholds this would yield sparse estimates of the source locations, and, as detailed by Bajwa et. al. in [8], this procedure, referred therein as one step thresholding, requires conditions for recovery that are stronger than the conditions for compressed sensing recovery algorithms. Various iterative versions of this correlation procedure, called Matching Pursuit or Orthogonal Matching Pursuit, also have similar compressed-sensing type conditions for reliable estimation, [9]. We will present these as alternative estimators in our approach in subsequent sections.

1.4 Outline

In chapter 2, we give background on general field inversion and the set-up for quantizing source locations to approximate the source propagation model by a discrete linear observation model. In chapter 3,

we show that a combination of consensus message passing and sparse estimation methods, which we call field inversion by consensus and compressed sensing (FICCS), can yield efficient distributed field reconstruction and estimation algorithms that scale with network size. Using results from the field of compressed sensing, we describe the design considerations for implementing this algorithm with sensors placed on a 2D lattice referring to [10, 11, 12] for predictions of sparse estimation performance. We illustrate the approach with the example of 2D localization of an unknown number of acoustic sources given a single snapshot of superimposed received signal strength (RSS) measurements, and discuss the scaling of the algorithm in common used network structures.

In chapter 4 we address the assumption that the sources must lie on the discrete grid and propose a robust approach to sparse linear parameter estimation. The robust estimator takes into account the uncertainty of the system model. Since this type of error is induced by spatial quantization, we use this knowledge to bound this type of error and adjust the noise-aware basis pursuit estimator to account for the structured errors. We derive a bound on estimation error using model-robust basis pursuit, as it depends on model error constants and the maximum magnitude of sources. We also compute lower bounds for the error of basis pursuit estimates that do not account for structural errors. We also analyze the relationship of structured modeling error constants and system coherence with the spacing of the spatial quantization lattice.

Chapter 5 describes several motivating applications in distributed sensing and oulines our approach to applying FICCS to these areas. Chapter 6 presents studies of the performance sensitivity of FICCS in numerical simulations using sparse estimation approaches for localization experiments with inverse power propagation laws. Both ℓ_2 errors and detectability of sparse signal support are compared using various sparse estimators as a function of a comprehensive set of system and estimation tradeoff parameters.

Chapter 7 presents an adaptation of FICCS to time-varying problems using dynamically re-weighted ℓ_1 penalized least-squares to incorporate knowledge of previously located sources. The estimator is tested in an acoustic source localization system where sources may randomly appear and disappear anywhere in the monitored area. The recursive estimator is shown to require few global network projections of measurements to accurately track of the field.

Chapter 8 analyzes how mixing pointwise measurements in a network through linear projection in-

creases the information available to local sensors. We derive bounds on the improvement of the mutual information of the mixed measurements in situations where sparse parameter locations are known, as well as unknown and random, showing that matrices with low coherence approach these bounds. In addition, we compare ℓ_1 recovery with minimum mean-squared estimators, showing that in test cases with low noise, there is little gap in recovery error.

We conclude in chapter 9 with a discussion of remaining challenges and future work.

Chapter 2

Observation System Modeling for Field Inversion

This chapter formulates the field inversion problem with continuous parameter locations as a discrete linear model with a sparse state parameter vector. We consider a physical scalar field f, defined in continuous space; e.g., temperature, acoustic signal strength, or water height. We assume propagating fields where at each time-snapshot the field is described by a small number of point sources, and the principle of superposition applies to fields composed of multiple point sources. For a real-valued two-dimensional (2D) field, this means that the field $f(t_1, t_2)$, at location (t_1, t_2), can be represented as a sum of scaled location-dependent basis functions, $\psi_{\mathbf{p}_k}(t_1, t_2)$,

$$f(t_1, t_2) \;=\; \sum_{k=1}^{S} \alpha_k \psi_{\mathbf{p}_k}(t_1, t_2) \tag{2.1}$$

where there are S sources with magnitudes $\alpha_k \in \mathbb{R}$, at 2D locations $\mathbf{p}_k \in \mathbb{R}^2$. The field locations may also be written as a vector, $[t_1, t_2]^T = \mathbf{t}$. The ψ's are identical except shifted by the source location $\mathbf{p}_k$; i.e., $\psi_{\mathbf{p}_k}(\mathbf{t}) = \psi_{\mathbf{0}}(\mathbf{t}) * \delta(\mathbf{t} - \mathbf{p}_k)$, and we refer to them as propagation functions.

We now formalize the network observation model. A distributed sensor network of N sensors samples the field at a discrete set of locations, $\mathbf{t}_j \in \mathbb{R}^2$. Each sensor's local observations, $\{x_j\}_{j=1}^N$, are corrupted by additive independent measurement noise, $\{n_j\}_{j=1}^N$. We have

$$x_j \;=\; f(\mathbf{t}_j) + n_j \;=\; \sum_{k=1}^{S} \alpha_k \psi_{\mathbf{p}_k}(\mathbf{t}_j) + n_j \tag{2.2}$$

With continuously located sources, the observation model is characterized by location-dependent basis functions that can be arbitrarily located in the continuum. As such, inverting the field to find the number of sources and their locations is a difficult problem. In cases in which the basis functions may be described through convolution of a propagation function with delta functions located at a finite set of locations $\{\mathbf{p}_k\}_{k=1}^{S}$, this corresponds to a deconvolution problem. When there is just one source, its location may be estimated through a variety of maximum likelihood or cost-minimization schemes.

In this work, we assume multiple sources may be present in the field. We also assume that all sources lie within the finite area monitored by the sensor network. Since our motivation is to find distributed estimation approaches that scale with increasing numbers of sensors, corresponding to an expanding area monitored, we think of the number of sensors as large. In particular, the number of sensors is assumed to be larger than the number of sources present in the field.

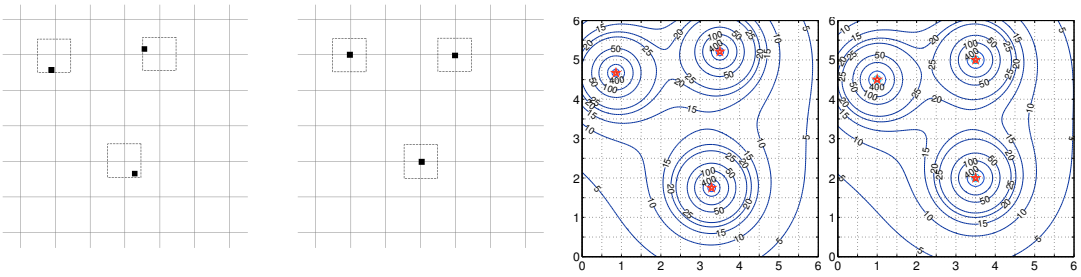
To formulate the observation system as a linear function of a discrete state vector, we approximate (2.2) by discretely sampling or *spatially quantizing* the monitored area using a 2D lattice with spacing parameter, δ, and $\sqrt{L} \times \sqrt{L}$ grid points. A key assumption is that, due to the similarity of responses to sources from nearby locations, the prediction errors of modeling the field as the superposition of sources ar a discrete set of locations can be bounded to moderate values. We index the discrete source vector, $\mathbf{v} = \left[v_i \right]_{i=1}^{L}$, by grid-point location, $i = 1, 2, \ldots, L$, which we order by stacking columns of the lattice. Each index i corresponds to a discrete location on the lattice. The source locations, previously indexed by source number k, are now indexed by the lattice index i. We write

$$x_j \approx \sum_{i=1}^{L} v_i \psi_{\mathbf{p}_i}(\mathbf{t}_j) + n_j \tag{2.3}$$

$$v_i = \begin{cases} \alpha_k & \text{for } \mathbf{p}_k \approx \mathbf{p}_i \\ 0 & \text{otherwise} \end{cases} \tag{2.4}$$

The source vector $\mathbf{v}$ is now of finite dimension with entries that are nonzero only when there is a source that is close to the corresponding location in the sampled lattice.

This process of quantizing the spatial dimensions of the field may be equivalently thought of as sampling in the spatial domain and "accumulating" the continuous point-source field in the 2D square



(a) actual locations (b) quantized locations (c) actual field values (d) approx. field values

Figure 2.1: Inverse square distance propagation field with three continuously located sources (panel a) and the approximation to a discrete lattice (panel b). Dashed squares represent the regions that each source is associated with. The corresponding field values are shown in contour plots in panels c and d, respectively.

region given by,

$$R_i = \left\{ \; (t_1, t_2) : \left(p_i(1) - \frac{\delta}{2}\right) < t_1 \leq \left(p_i(1) + \frac{\delta}{2}\right), \left(p_i(2) - \frac{\delta}{2}\right) < t_2 \leq \left(p_i(2) + \frac{\delta}{2}\right) \; \right\} \quad (2.5)$$

where δ is the distance between each lattice point, and $p_i(1)$ and $p_i(2)$ are the first and second coordinates of the location of lattice point i. If more that one source falls in the same region, their contributions will be summed, and therefore indistinguishable in the discretized source vector.

Figure 2.1 shows this process in an example field with 3 sources. Each source is associated with the region of just one of the lattice points. The field contours are compared for actual source locations versus sources locations at the nearest lattice points. Based on the particular field basis functions and the spatial sampling rates of the lattice, this approximation can be more or less accurate.

Written in matrix form, the approximation of (2.3) becomes the following system.

$$\mathbf{x} \quad = \quad \Psi\mathbf{v} + \mathbf{n}, \qquad \text{where } \Psi \in \mathbb{R}^{N \times L} : \Psi = \left[\psi_{\mathbf{p}_i}(\mathbf{t}_j)\right]_{j,i} \quad (2.6)$$

By approximating the system with discretely sampled source locations, the field inversion problem considered becomes the problem of estimating an unknown discrete vector $\mathbf{v}$ that can be described as a sparse linear combination of parameters, using observations $\mathbf{x}$ collected by a network of sensors with a known linear model, $\Psi : \mathbb{R}^L \to \mathbb{R}^N$. By construction, the source vector is known to be sparse; i.e., $\|\mathbf{v}\|_0 \leq S$, where $\|\cdot\|_0$ is called the ℓ_0 quasi-norm and counts the number of nonzero entries.

The observation model of (2.6) will be further modified through a message-passing method of mea-

surement combination throughout the network, which will be described in chapter 3 section 3.1, creating a low dimensional projection of the field.

$$\mathbf{y} = \Phi\left(\Psi\mathbf{v} + \mathbf{n}\right) = \Phi\Psi\mathbf{v} + \mathbf{z}, \quad \text{for } \Phi \in \mathbb{R}^{m\times N} \quad \text{or} \quad \mathbb{C}^{m\times N} \tag{2.7}$$

The important thing to note is that this system now corresponds to a typical noisy compressed sensing set-up. Hence, when Φ and Ψ satisfy certain conditions, the parameter vector $\mathbf{v}$ may be approximately recovered through noise-aware basis pursuit, even for $m \ll N$, [13, 14, 15, 12]. Next, we show how the discrete linear approximation is applied to the observation model of acoustic sources.

2.1 Example Application: Acoustic Source Localization

We use acoustic source localization as a concrete application to test the methods presented. The goal is to localize point sources to a 2D lattice of potential locations using a single time snapshot of received signal strength (RSS) measurements. Optimally placed sensor positions will occur on the same lattice. Each sensor samples the acoustic field, which is the superposition of the signals radiated by all sources. A simple approximate model for acoustic propagation of received intensity is an inverse power law. We refer the reader to [16] for more details on the free-space far field approximation leading to the following model for acoustic intensity.

$$\psi_{\mathbf{p}_i}(\mathbf{l}_j) = \frac{1}{(\|\mathbf{l}_j - \mathbf{p}_i\|^{\gamma} + 1)} \tag{2.8}$$

where $\mathbf{l}_j$ represents the 2D position of sensor j, $\mathbf{p}_i$ the 2D position of source i, and $\gamma \geq 1$ is a known environmental parameter. The addition of 1 to the denominator serves to avoid numerical issues at small distances. Under this model and with $\gamma = 2$, the form of the Ψ matrix of (2.6) and (2.7) is specified by inverse square distances between each source and sensor location.

$$\Psi = \left[\frac{1}{d_{j,i}^2 + 1}\right]_{1\leq i\leq N, 1\leq j\leq L} \tag{2.9}$$

We expect a small number of active point sources compared to the number of grid locations, so the vector of source magnitudes $\mathbf{v}$ is sparse by construction. A key parameter in this model is the lattice spacing. This affects the resolution of the source location estimates, as well as the prediction error due to

quantization of source location. However, the conditioning of the original linear observation system (2.6) worsens with finer lattice spacing, pointing to a tradeoff between a grid that is too fine or too coarse. Sparse recovery algorithms can overcome certain conditioning issues, although their own version of a conditioning metric, called coherence, will come into play. These issues will be discussed in section 3.2.

2.2 Conclusion

This chapter showed the observation model that is generated by quantizing the space of potential source locations to a 2D lattice. Because of this construction, the source field vector in this system has only a small number of nonzeros. This model considers field inversion using a single time snap-shot of information.

In chapter 3, we propose the FICCS algorithm, which distributes a low dimension projection of the field using a robust type of message passing, and describe how the full field is estimated at each node using techniques from compressed sensing.

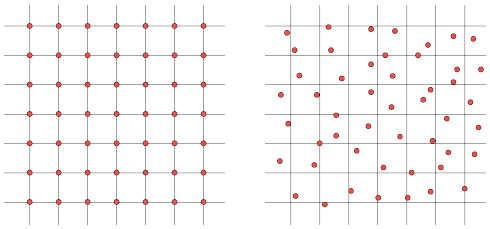
Chapter 3

Field Inversion by Consensus and Compressed Sensing

In this chapter, we present, FICCS, a distributed algorithm for field inversion using consensus and compressed sensing. The idea of using local communications to distribute projections of sensor network measurements for use in sparse recovery algorithms was proposed by Rabbat, Haupt, Singh and Nowak in a method for compressing and distributing network data [17]. Gossip and distributed average consensus have many applications in network signal processing; for an extensive review of these see [18]. In studying the performance of sparse field inversion using randomly weighted iterative local mixings, we found that this method for distributing low dimensional projections of measurements throughout the network enabled the design of weights with good sparse inversion properties for the discrete acoustic source localization problem, [19, 20].

3.1 The FICCS Algorithm

The steps of the field inversion by consensus and compressed sensing (FICCS) algorithm involve setting up a sensor network to use iterative message passing for computing projections with pre-defined weights. The message passing portion will be called operation in *sensing mode* and uses distributed consensus to pass compressed information throughout the network, after which, each node uses a compressed sensing recovery algorithm to estimate the field, called *field inversion*.



(a) lattice placement (b) random placement

Figure 3.1: Two ways to deploy sensors with respect to the discrete lattice of potential source locations. The left panel shows the optimal choice of regular placement, while the right panel shows random, but even, placements.

3.1.1 Network Set-up

N sensors are deployed throughout the area to be monitored. Ideally their locations correspond to the $\sqrt{N} \times \sqrt{N}$ lattice of discretized source positions to be detected, but they may be deployed in other evenly-distributed ways. Figure 3.1 illustrates this. Sensors must then be localized with respect to each other. There are efficient decentralized methods for this; for an example see [21, 22].

Having the sensor locations allows us to find the linear observation matrix, Ψ, as well as the undirected communication graph topology, as determined by a uniform sensor communication radius. During network initialization, the nodes compute their local distributed averaging weights, potentially using aspects of the graph structure. Every node will associate a weight for itself and each of its neighbors. Collected together in a matrix, the set of all consensus weights has the structure

$$W \in \mathbb{R}^{N \times N} \ : w_{i,k} = 0 \ \text{ if } \ (i,k \notin \mathcal{E} \text{ and } i \neq k) \tag{3.1}$$

where $\mathcal{E}$ is the set of available direct communication channels among sensors, represented as edges in an undirected graph. The entries of W are chosen such that iterative weighted combinations of each sensor's value with its neighbors will converge to the average of all N sensor values. The method for this is discussed in 3.3. Each node is also assigned a set of m projection weights, which are chosen so that the collection has low "coherence" with the sparsifying basis for the field, Ψ. Considerations for choosing projections are discussed in 3.2. The collection of all projection weights forms the matrix $\Phi \in \mathbb{R}^{m \times N}$; each column of which is assigned to a different node.

3.1.2 Sensing Mode using Consensus

The sensors collect noisy measurements of a static or concurrently observed field, $\mathbf{x}$, defined in (2.6). The sensor i initializes its local m dimensional state equal to $(Nx_i) \otimes \phi_i^T$, where ϕ_i is the i^{th} column of the projection matrix Φ and $\otimes$ is the Kronecker product, which multiplies each entry in ϕ_i^T by the scalar Nx_i. It will be convenient to represent each projection value in $\mathbf{y} = \Phi\mathbf{x}$, using rows of Φ, which we refer to as $\mathbf{p}_j$ for $j \in \{1, \ldots, m\}$.

$$\mathbf{y} = \Phi\mathbf{x} = \begin{bmatrix} \phi_1 & \cdots & \phi_N \end{bmatrix} \mathbf{x} = \begin{bmatrix} - & \mathbf{p}_1^T & - \\ & \vdots & \\ - & \mathbf{p}_m^T & - \end{bmatrix} \mathbf{x} \tag{3.2}$$

All states in the network can be represented by the $N \times m$ dimensional matrix S_0.

$$S_0 = \begin{bmatrix} (Nx_1) \otimes \phi_1^T \\ \vdots \\ (Nx_N) \otimes \phi_N^T \end{bmatrix} \tag{3.3}$$

$$= \begin{bmatrix} Np_1(1)x_1 & Np_2(1)x_1 & \cdots & Np_m(1)x_1 \\ & \vdots & & \\ Np_1(N)x_N & Np_2(N)x_N & \cdots & Np_m(N)x_N \end{bmatrix} \tag{3.4}$$

We assume a deterministic consensus protocol using constant symmetric weights, W. Randomized gossip methods can also be applied as in [23, 17]. At each consensus iteration every sensor updates its state by a weighted linear combination of its own and the states of its neighbors, using weights in W. With this matrix representation of the consensus state, the projection values at each node at the k^{th} mixing is given in terms of the previous state.

$$S_k = WS_{k-1} = W^k S_0 \tag{3.5}$$

$$\lim_{k\to\infty} W^k = \frac{\mathbf{1}\mathbf{1}^T}{N} \tag{3.6}$$

The states of all the nodes converge to

$$\frac{\mathbf{1}\mathbf{1}^T}{N} S_0 = \frac{\mathbf{1}\mathbf{1}^T}{N} \begin{bmatrix} N\text{diag}(\mathbf{p}_1)\mathbf{x} & \cdots & N\text{diag}(\mathbf{p}_m)\mathbf{x} \end{bmatrix} \tag{3.7}$$

$$= \begin{bmatrix} \mathbf{1}\mathbf{p}_1^T\mathbf{x} & \cdots & \mathbf{1}\mathbf{p}_m^T\mathbf{x} \end{bmatrix} = \mathbf{1} \otimes \mathbf{y}^T \tag{3.8}$$

where the diag$(\cdot)$ operation creates a square zero matrix and places the vector argument on the diagonal. After a large enough number of mixings, all nodes will approximately have the same projection $\mathbf{y} = \Phi\mathbf{x}$. In practice, sensors will stop at a number k agreed upon during set-up. We denote the exact projection matrix of node i as $\Phi_{\text{eff}(i)}$, which is known to each node.

3.1.3 Field Inversion Mode using Compressed Sensing

We now discuss the ways in which local sensors may reconstruct estimates of the global field using their m low dimensional projections. The projection at node i represents

$$\begin{aligned} \mathbf{y}(i) &= \Phi_{\text{eff}(i)} \left(\Psi\mathbf{v} + \mathbf{n}\right) && (3.9) \\ &= A_{\text{eff}(i)}\mathbf{v} + \mathbf{z} && (3.10) \end{aligned}$$

For $\mathbf{n}$ being additive independent zero-mean Gaussian noise, $\mathbf{z} \sim \mathcal{N}(0, \sigma_n^2 \Phi_{\text{eff}(i)} \Phi_{\text{eff}(i)}^T)$. A noise-aware basis pursuit approach places a bound on the norm of the additive noise vector. In that case, we estimate a sparse $\widehat{\mathbf{v}}$ that satisfies the linear model to within an error ϵ, based on, for example, a 99% bound on $\|\mathbf{z}\|_2$.[1] Note that non-Gaussian noise can also be handled by this as long as a reasonable bound on the norm of the noise vector can be used.

There are many alternatives for estimating the arguments of sparse linear systems; see [24, 25, 8] for some comparisons. These methods are based on preferring sparse estimates $\widehat{\mathbf{v}}$, and have varying computational complexities. We list 5 alternative approaches for comparison.

1. ℓ_0 minimization ($P_{0,\epsilon}$):

$$\begin{aligned} &\text{minimize} && \|\mathbf{v}\|_0 \\ &\text{subject to} && \|\mathbf{y} - A\mathbf{v}\|_2 \leq \epsilon \end{aligned} \tag{3.11}$$

   for a chosen ϵ, a bound on the norm of the additive noise $\|\mathbf{n}\|_2 \leq \epsilon$, [11].

2. ℓ_1 minimization ($P_{1,\epsilon}$): (Also called Noise-Aware Basis Pursuit)

$$\begin{aligned} &\text{minimize} && \|\mathbf{v}\|_1 \\ &\text{subject to} && \|\mathbf{y} - A\mathbf{v}\|_2 \leq \epsilon \end{aligned} \tag{3.12}$$

   for a chosen ϵ, a bound on the norm of the additive noise $\|\mathbf{n}\|_2 \leq \epsilon$, [11].

[1] For a Gaussian noise vector, this high probability bound may be computed via the inverse chi-squared function.

3. ℓ_1 penalty:

$$\text{minimize} \quad \gamma\|\mathbf{v}\|_1 + \|\mathbf{y} - A\mathbf{v}\|_2 \tag{3.13}$$

for a chosen γ value, [24].

4. Orthogonal Matching Pursuit (OMP): (Also called Orthogonal Greedy Algorithm)

Initialize $\mathbf{r}_0 = \mathbf{y}$, $\mathbf{v}_0 = \mathbf{0}$, $T_0 = \emptyset$.

Iteration k: compute $\mathbf{q} = A^T\mathbf{y}$, let $i = \text{argmax}_{i \notin T_k} |q_i|$

$T_{k+1} = \{T_k \cup i\}$, $\mathbf{v}_{k+1} = \left(A_{T_k}^T A_{T_k}\right)^{-1} A_{T_k}^T \mathbf{y}$, $\mathbf{r}_{k+1} = \mathbf{y} - A\mathbf{v}_{k+1}$.

Stop when: $\max_{i \notin T_k} |q_i| < \gamma$, [26, 11].

5. Maximum Correlation Thresholding (MC): (Also called One Step Thresholding)

$$\begin{aligned} \text{Compute} \qquad & \mathbf{q} = A^T\mathbf{y} \\ \text{Find} \qquad & T = \{i : i \in [1, N], \ \ |q_i| > \gamma\} \\ & \widehat{\mathbf{v}} = (A_T^T A_T)^{-1} A_T^T \mathbf{y} \end{aligned} \tag{3.14}$$

for a chosen γ value, [25, 8].

The ℓ_0 minimizer, $(P_{0,\epsilon})$, gives the ideal sparse solution, but is combinatorial to compute, and so is infeasible even for modest dimension. The ℓ_1 minimizer, $(P_{1,\epsilon})$, which is a convex relaxation of the former, corresponds to a second order cone program (SOCP), [27], for which there exist efficient methods for solving. For optimization techniques specifically tailored to basis pursuit, see [28]. The ℓ_1 penalty is also a very efficient program, and has good properties for noise that is Gaussian. Depending on the computational resources available at sensor nodes, when the dimension of the source vector is very large, both convex programs, corresponding to methods 2 and 3, may become intractable to compute. In those cases, we may use method 4, which is a sequential greedy pursuit method. For extremely large problems, solving the least squares at each iteration of OMP, may be too taxing, in which case the MC estimator, corresponding to a type of matched filter detection, takes only linear time to estimate the sparse support $\mathbf{v}$, [8].

Our efforts thus far have focused on specifically on the second method, basis pursuit and its variants, due its excellent performance on moderate sized problems and the extensive literature on this type of sparse signal recovery. Some of the strongest compressed sensing results for robust signal recovery

correspond to this approach (as well as the ℓ_1 penalty method). Also, when a bound on noise energy is known, this method requires little parameter tuning.

3.2 Design of Measurement Projection Weights

In this section, we discuss the design of the measurement projections used in the matrix Φ, and how they relate to theoretical predictions of field inversion performance. To address the questions of field inversion performance after network compression of the sensor measurements, we turn to some predictions from the field of compressed sensing. Compressed sensing theory suggests a variety of methods for observing a sparse signal that are likely to lead to stable estimates; i.e., ℓ_2 errors linearly related to the additive noise norm bound ϵ. For the noise-aware basis pursuit method of [11], given in (3.12), the ℓ_2 error of recovery estimates is upperbounded by a function of the coherence. This parameter is used to predict the performance of compressed sensing and is defined in [11, 24] for $A = \Phi\Psi$, as

$$\mu(A) \quad = \quad \max_{i \neq j} \frac{\left|\mathbf{a}_i^T \mathbf{a}_j\right|}{\|\mathbf{a}_i\|_2 \cdot \|\mathbf{a}_j\|_2}, \qquad \text{for} \quad A = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_N \end{bmatrix} \tag{3.15}$$

When the number of nonzeros, also called the sparsity of $\mathbf{v}$, denoted S, is less than $\frac{1}{4}\left(\frac{1}{\mu}+1\right)$, one can deterministically bound the recovery error of the basis pursuit estimator of (3.12), with

$$\|\widehat{\mathbf{v}}_{BP} - \mathbf{v}_0\|_2 \quad \leq \quad \frac{2\epsilon}{\sqrt{1-\mu(4S-1)}} \tag{3.16}$$

as proven by Donoho et. al. in [11]. This bound is not tight, and stable recoveries are frequently observed in systems not satisfying the requirement $S < \frac{1}{4}\left(\frac{1}{\mu}+1\right)$. The complimentary analysis of Candès, Romberg, and Tao, [10], derives conditions for stable recovery via (3.12) using the restricted isometry property (RIP). The results suggest the use of random matrices and provide less restrictive conditions for approximate recovery. Testing whether a particular matrix satisfies RIP is intractable, although [29] provides methods of bounding restricted isometry constants of a matrix, as well as sufficient conditions using the null space property (NSP).

Choosing the projections in Φ will determine the coherence of the system. There are a variety of choices investigated in the compressed sensing literature. Compression matrices with entries that are randomly drawn from i.i.d. $\mathcal{N}\left(0, \frac{1}{m}\right)$ or the matrices generated from orthonormalizing these, have low

coherence, defined in (3.15), with high probability, [14]. In the noisy setting, Gabor frames achieve a theoretical minimum on the coherence, defined in (3.15), and can be constructed with continuous subsets of m rows of the discrete Fourier matrix.[2] For real-valued projections, Grassmann frames achieve lower bounds in coherence, but are unfortunately not constructable for all pairs (N, m), [31, 32]. Although additional strategies abound, for simplicity we consider random orthonormal projections or a subset of the Fourier matrix. To use the complex Gabor frame, we compute Ψ^{-1} and then choose

$$\Phi = F_{1:m}\Psi^{-1} \tag{3.17}$$

where $F_{1:m}$ is the first m rows of the Fourier matrix. The overall sensing matrix

$$A = \Phi\Psi \tag{3.18}$$

is then equal to the Gabor frame and will have optimally low coherence. Alternatively, we may use real random projections, such as a matrix i.i.d. Gaussian entries or the first m rows of a randomly generated orthonormal matrix. In all of these cases, the number of projections m needed for reliable sparse recovery is predicted to scale with N at a rate of $O(S\log(N))$, where S is an upperbound on the number of nonzero components to be estimated, [14].

3.3 Consensus Weight Design and Scaling

A key goal of this study investigating the application of sparse estimation and distributed averaging to network field inversion is to find techniques that scale with increasing numbers of sensors N. In this section, we discuss the design of the distributed averaging weights of W and discuss the asymptotic growth of the number of messages and latency of FICCS, specifically in an 8-nearest neighbor 2D lattice graph communication topology. We will assume, for simplicity, that the 2D lattice of source positions also corresponds to the set of sensor positions, meaning the dimensions N and L from the observation equation (2.6) are equal, and the initial observation system is square. This also means that increasing N corresponds to increasing the number of source locations. Under these assumptions, the area monitored grows larger by increasing the size of the sensor/location lattice.

[2] In coding settings, these are known as maximally robust tight frames, [30].

In section 3.2 we estimated how the number of measurement projections m scales with increasing N. To study the convergence rate of distributed consensus, we use the fast distributed linear averaging (FDLA) algorithm, assuming a grid-based network in which nodes on the 2D lattice communicate with 8 neighboring nodes. It is known that consensus in this type of network is slow to converge, [33]. However, this corresponds to a straightforward sensor deployment and minimizes the distance that any single sensor must transmit messages.

For simplicity, we use the results of [34] setting all edge-weights to the uniform value α_{opt} and the self weight for node i to $(1 - \alpha_{\text{opt}} d_i)$, where d_i is the degree of the i^{th} node. The optimal α is equal to $2/\left(\lambda_2(L) + \lambda_N(L)\right)$, where λ_2 and λ_N are the second smallest and largest eigenvalues of the graph Laplacian matrix,

$$L = D - A \tag{3.19}$$

where the D matrix is a diagonal matrix of node degrees, and A is the adjacency matrix with ones only at (i, j) coordinates where $(i, j) \in \mathcal{E}$, and zeros everywhere else. Using the uniform weights strategy,

$$W = I - \alpha_{\text{opt}} L \tag{3.20}$$

This choice of weights yields the asymptotic step-wise convergence factor,

$$\tau_{\text{step}} \;=\; \rho\left(W - \mathbf{1}\mathbf{1}^T/N\right) \;=\; \frac{\lambda_N - \lambda_2}{\lambda_N + \lambda_2} \tag{3.21}$$

For a desired multiplicative error tolerance of $0 < c < 1$, representing $\|\mathbf{y}_k - \mathbf{y}_{\text{avg}}\|_2 / \|\mathbf{y}_0\| \leq c$, the number of mixings k should be

$$k \geq \frac{\log(c)}{\log(\tau_{\text{step}})} \tag{3.22}$$

Characterizing how k grows with increasing N involves describing how τ and therefore the eigenvalues of the Laplacian depend on N.

In practice the graph of sensors on a 2D lattice suffers from boundary effects so that only the nodes in the interior have the regular number of 8 adjacent neighbors. For analysis purposes only, we add edges to the boundary nodes connecting them to the opposite sides of the lattice. This results in a toroid structure, with the advantage of now being regular with a circulant adjacency matrix. These additional edges can

only improve the rates of mixing convergence and so are representative of a slightly better case version of the original grid structure. In a regular graph, the Laplacian simplifies to $dI - A$ where is d is the degree of every node. Using a convention that eigenvalues be ordered, $0 = \lambda_1 \leq \lambda_2 \leq \cdots \lambda_N$,

$$\lambda_i(L) = d - \lambda_{(N-i+1)}(A) \quad \text{for all } i \in [1, N] \tag{3.23}$$

To find the needed eigenvalues of the Laplacian we may solve for the smallest and second largest eigenvalues of A, representing their dependence on the lattice dimension g explicitly.

Theorem 3.3.1 Let g be an even positive integer. Assume $N = g^2$ sensors lie on a 2D torus with dimensions $g \times g$ and each node can communicate with its 8 nearest neighbors. Using the optimal uniform-weight distributed averaging algorithm of [34], the number of consensus iterations k to ensure convergence within the multiplier $c < 1$ of the average grows linearly with N.

The proof of this is detailed in Appendix A and relies on the relationship between the eigenvalues of L and A and on the property that A is circulant, and therefore diagonalizable by the Fourier matrix. This yields a closed form for the eigenvalues of the adjacency matrix.[3] We show that the minimum magnitude eigenvalue of A is constant with g, corresponding to a constant $\lambda_N(L)$. In contrast, the eigenvalue of A with second largest magnitude converges asymptotically with N to the degree, 8. This results in a decreasing $\lambda_N(L)$ and increasing in the stepwise convergence factor. From these, the behavior of

$$k = \frac{\log(c)}{\log\left(\tau_{\text{step}}\right)} = \frac{\log(c)}{\log\left(\frac{\lambda_N - \lambda_2}{\lambda_N + \lambda_2}\right)} \tag{3.24}$$

is characterized for large g and shown to grow at the rate of $g^2 = N$.

Since each deterministic consensus iteration involves approximately $8N$ messages to be exchanged among nodes in the network, the total number of messages required for convergence, applying Theorem 3.3.1, has the scaling $O(N^2)$. Certain randomized gossip algorithms, such as the geographic gossip of [36], can perform better than this deterministic method. A modification of geographic gossip, called randomized path averaging, of [37], has recently shown an "order-optimal" convergence rate. This means the number of messages needed for each projection can be made to scale linearly with N in both grid-based and random geometric graphs.

[3] A similar relationship is given in [35] for a 1D torus with 2-nearest neighbors.

FICCS with	messages	latency
uniform weights, [34]	$SN^2\log(N)$	$SN\log(N)$
path averaging, [37]	$SN\log(N)$	$SN\log(N)$

Table 3.1: Order of algorithm scalings with N, the number of nodes in the 8 nearest neighbor 2D lattice graph.

Since the FICCS algorithm requires $m \sim O(S\log N)$ weighted averages to be computed, which can be done in parallel, we may now conclude that the growth in messages with N in this nearest neighbor graph is $O(SN^2\log N)$ for the uniform-weighted W described, and $O(SN\log N)$ if using the order-optimal path averaging of [37]. To define the latency of FICCS, in an idealized sense, we ignore computation time for estimation and assume that a consensus iteration for one projection can be executed in $O(0)$ time. Then in both approaches, the latency grows with

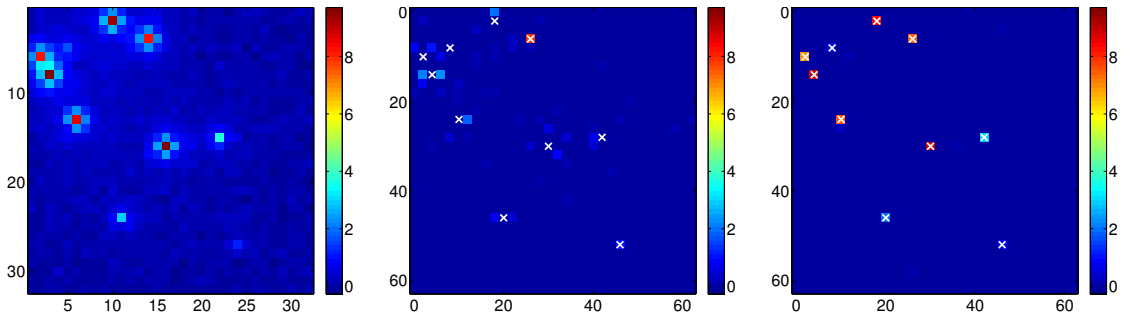
$$m \cdot k \sim O(SN\log(N)) \tag{3.25}$$

Table 3.1 summarizes these scaling relationships.

For the FICCS method with the order-optimal consensus algorithm, we see that the number of messages, our "stand-in" for communication energy burden, grows with a rate at least as large as $N\log(N)$. If we take lattice growth to correspond to incorporating a larger area to be monitored, it is reasonable that S, the maximum number of nonzero components, will also scale linearly with N. This scaling law leaves much to be desired since greater than N^2 growth in messages diverges quickly with network size. This can motivate the study of other topologies, which can drastically improve the scaling of consensus approaches in applications where longer distance links are allowed, as in the work of [33, 38] containing some results on topology optimization. These scaling relationships show the promise of the FICCS algorithm for reducing the scaling factors for communication overhead in growing networks.

3.4 Example Recovery Using FICCS

We now present an example of FICCS used on an acoustic source localization experiment. In this test, sources are assumed to only occur at the sampled grid locations.



(a) all sensor measurements (b) 300 selected sensor readings (c) 300 orthonormal projections

Figure 3.2: Example recovery estimates for a sensor network with $N = 1024$ sensors placed in a 32-by-32 grid, where intersensor spacing is $\delta = 2$. The left panel shows all sensor readings; additive noise has variance $\sigma^2 = 0.01$. Middle and right panels show basis pursuit estimates for 300 randomly selected sensor measurements and 300 random orthonormal projections computed via consensus. The true source locations of the 10 sources are overlayed in white crosses. The ℓ_2 recovery errors were 20.75 and 2.09 for panels (b) and (c), respectively.

Figure 3.2 shows the example sparse recovery using the FICCS algorithm. The left panel displays the set of all sensor observations in a 32-by-32 grid network. These are samples of a field induced by 10 acoustic sources with random positive magnitudes between 1 and 10. Sources are located exactly at grid-points in the potential source location lattice. There is additive independent Gaussian measurement noise with variance 0.01, resulting in signal-to-noise ratio of approximately 16 dB. The spacing of the regular sensor grid is 2, and is coincident with the lattice of potential source locations. Each sensor's communication radius is 3, meaning an interior sensor is connected to its 8 nearest neighbors.

With the full set of observations, the locations and magnitudes of the sources can clearly be estimated by collecting all 1,024 sensor readings at a fusion center and using classical least squares. However, the FICCS approach avoids the communication bottleneck and single-point-of-failure sensitivity of centralized collection. In this example, $m = 300$ random orthogonal measurement projections are computed using the distributed consensus of section 3.1.2. Panel (c) shows the recovery estimate of the noise-aware Basis Pursuit (BP) with $\epsilon = 1.53$, corresponding to the 95% probability bound on the additive noise $\mathbf{z}$. The white crosses show the true source locations. We see that all but one source is detected, and ℓ_2 estimation error is very low, at only 2.09.

In panel (b) we compare with the BP recovery estimate using 300 randomly-selected sensor observations. Because the propagation basis Ψ is not incoherent with the "spike" basis, the sparse estimator

performs very poorly, detecting only one of the sources present, resulting in roughly 10 times the estimation error of the FICCS estimate. This demonstrates the importance of the consensus message-passing phase of the approach. Using distributed averaging computes incoherent projections of the field and allows every node to estimate the global field. In the next section, we give average reconstruction performance as a function of the number of network measurement projections, m.

3.5 Simulation Testing of FICCS

To test the performance of the approach for distributed field inversion, we simulate a network of sensors observing an acoustic field with a small number of point-sources on a 2D lattice. We plot average ℓ_2 error of the basis pursuit estimates over a large set of trials and as a function of the number of measurement projections m, replicating the threshold behavior typical to compressed sensing systems. We also plot the computed asymptotic convergence parameters of distributed consensus in our grid graph topology for various network sizes. This shows numerically the linear trend that was predicted in section 3.3 of the growth in the number of mixings needed. Finally, we compare the performance of the FICCS estimator to the least-squares estimator when sensors are deployed randomly in the monitored area. This demonstrates how an estimator that promotes sparsity may be more robust to ill-conditioned random sensor deployments.

As an example of a FICCS reconstruction, we simulate a square 2D lattice of 1024 sensors. The lattice spacing used is $\delta = 2$, and sources unknown positive magnitudes may be located only at lattice locations. Each sensor exchanges messages with its neighbors within a communication radius that yields a version of a 2D lattice graph with connections to the 8 nearest neighbors. This geographically driven communication topology is very sparse. With only 7,812 edges, there are many fewer edges than in the one million edges of the complete graph.

With this communication topology and using the optimal uniform weights method of [34], the stepwise convergence factor for uniform average consensus is $\tau_{\text{step}} = 0.9953$. To converge to within 0.001 of the average, we set the number of consensus mixing iterations to 1,461.

We compare the performance of two types of projection matrices. The first type is a subset of rows from a randomly generated real orthonormal matrix, called Φ_{orth}. The second matrix Φ_{eff} is complex

and designed to yield a subset of the DFT coefficients when composed with the propagation matrix. We examine the reconstruction performance of these ideal Φ's as compared with the actual Φ_{eff} that results from the finite number of consensus mixings. Basis pursuit estimates were computed using the tool `cvx`, [39]. We have also compared a tool using spectral projected-gradient method detailed in [40], which is optimized to solve basis pursuit. This open-source code package, `SPGL` [41], is much faster, although the results here do not use it.

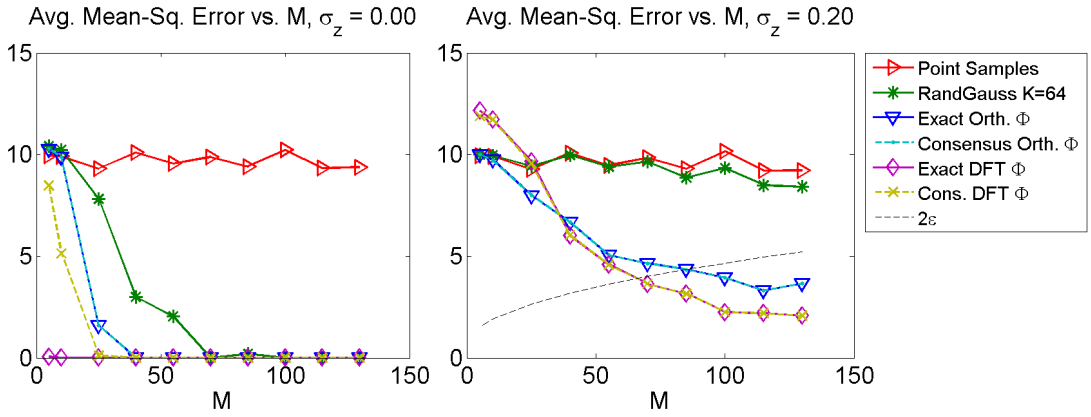


Figure 3.3: Field inversion performance as a function of the number of projections m, for noiseless measurements (left) and $\sigma_2 = 0.2$ (right). Mean squared error of the source vector is plotted for various choices of projection matrices, Φ, and compared to inversion using m point samples or randomly mixed samples.

Figure 3.3 shows average ℓ_2 inversion performance, as it depends on the number of projections, m, and method for choosing Φ. The two panels represent a study with no measurement noise and a study with $\sigma_z = 0.2$. The number of randomly chosen sources is 3, and mean squared error is averaged over 20 independent trials. As the number of projections increases, there is more information to constrain the solutions of the sparse minimization, and so the ℓ_2 error of the estimates fall. We compare estimation performance of using exact projections to the effective ones arrived at through distributed consensus, and show very little difference in error. The top red line, shows a baseline for comparison. We call this point-wise samples, and it represents a projection that simply selects m of the sensors observations for distribution. The point-wise samples show the effects of having a very poorly mixed projection of the sparse sources motivating the use of nonsparse observation bases. The second series, marked by green asterisks, is shown for comparison to a completely random mixing method in which nodes draw new i.i.d.

Gaussian weights for each projection at every consensus iteration. Although this random mixing method performed well in no noise, as seen again on the left, it is not robust to additive noise. The remaining series are tests of the FICCS approach, where the projections are carefully designed to yield either a Φ_{orth} of Φ_{dft}. The results of using the distributed parallel mixings match very well to using the exact projection values. The projections designed to yield the DFT perform very well in the noisy measurements cases, but requires the storage and transmission of complex states.

Figure 3.4 shows how the convergence factor grows with number of nodes N. In this 8-nearest neighbor 2D lattice graph structure, the number of required mixings grows linearly with N, in both the optimal uniform and optimal symmetric strategies of [34]. These are compared with the random Erdös-Rényi graph with the same number of edges, and the predicted performance of an 8-regular Ramanujan graph, described in [33]. This verifies our estimates of the growth rate of required iterations, for our geographically driven communication topology, and highlights the stark difference with communication structures that are optimized for consensus.

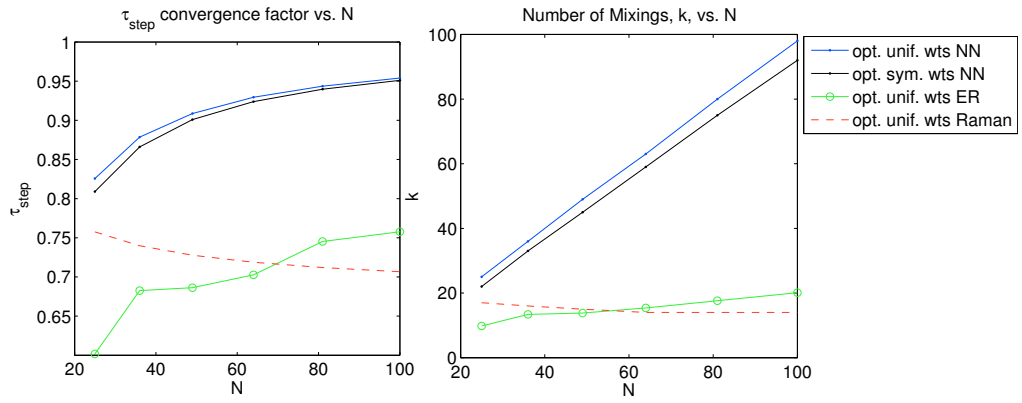
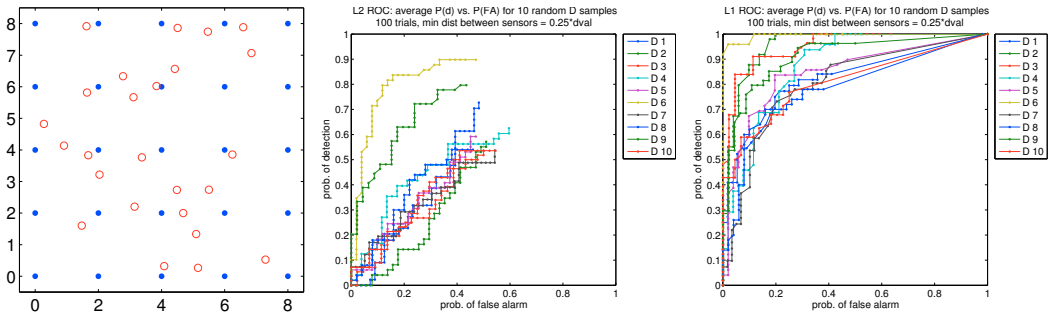


Figure 3.4: Comparison of the asymptotic convergence rates of FDLA in the 8 nearest-neighbor grid graph versus a random Erdös-Rényi, and Ramanujan graphs with the same number of edges. Each were computed by numerically, except in the case of the Ramanujan graph with represents a lower bound on the number of mixing in an 8-regular graph, vs. N.

Lastly, we show the performance of FICCS on a small network, $N = 25$, where sensors are placed randomly, but such that they satisfy a minimum inter-sensor distance of $\frac{\delta}{4}$. The potential source locations again lie on the 2D lattice with $L = N = 25$, and lattice spacing parameter is again $\delta = 2$. The leftmost panel of fig. 3.5 shows an example deployment. Unlike the previous tests, no compression on measurements is done; i.e., $\Phi = I$, and we focus on the middle sensor's ability to detect a source at



(a) example sensor deployment (b) Least Squares ROC (c) Basis Pursuit ROC

Figure 3.5: A comparison of center source detection performance in 10 random sensor deployments with the restriction that inter-sensor distance is at least $\frac{1}{2}$, Left panel shows an example deployment. Each series in (b) and (c) corresponds to a different random network deployment. Receiver operator characteristic for the center source is averaged over 100 experiments.

the center of the grid using measurements from all of the sensors. We simulate the presence sources at each lattice point as random Bernoulli, with the uniform probability $p = \frac{1}{5}$, except at the center location where the source arrives with probability $p = \frac{1}{2}$. Source magnitudes are now equal to 15, when present, and the additive sensor noise has standard deviation of 1.

For the least squares approach, the estimate $\widehat{\mathbf{v}}_{LS} = \Psi^{-1}\mathbf{y}$ is used and the center estimate is thresholded for detection. Varying the threshold generates the receiver operator characteristic (ROC) for detection versus false alarm rates. This is compared with the detection performance of basis pursuit, in which the center value is thresholded. The detection and false alarm rates are averaged over 100 trials, showing the basis pursuit method to be much more robust to the particular sensor deployment. This suggests an added benefit of the sparse estimation approach, in that it can better handle poorly conditioned Ψ.

These examples showed the operation of FICCS in test cases where sources were constrained to lie on the grid-points of the spatial quantization lattice. In practical applications we expect sources or targets to arrive uniformly in the 2D monitored area. The next chapter presents the model for bounding this structured error and methods to adapt basis pursuit estimators to spatial quantization error.

3.6 Summary

This chapter presented the distributed field inversion algorithm, FICCS, detailing the deployment and operation of sensing and field inversion modes using consensus and compressed sensing. We discussed

the design of the weights matrices for distributed averaging and how the number of messages required for consensus scales with the number of nodes in regular lattice graphs. We examined considerations for choosing informative measurement projections, Φ, from the perspective of low mutual coherence. An example field inversion using FICCS was presented, as well as a study of average recovery error in noise-free and noise cases when sources are constrained to be located only at grid points. The next section will address the issue of continuously located sources and the impact of structural model error on the noise-aware basis pursuit estimator.

Chapter 4

Structured Model Error and Model-Robust Basis Pursuit

Our assumption that potential source locations must lie on a discrete lattice can affect whether sensor measurements are well represented by a sparse discrete vector. In this chapter, we describe the perturbations to the system model due to sources actually being distributed in a continuum and propose a method to help deal with these structural perturbations. In general, the impacts of modeling error due to discretization will be application dependent. First, we show an additive matrix perturbation representation for the model error. Then we use this set-up to bound the model fit error, leading to a straightforward modification of the noise-aware basis pursuit algorithm that we call *model-robust basis pursuit*.

Comparison of Existing Approaches

We call attention to a number of approaches to sparse estimation when the field is composed of continuously translated, known propagation functions. There are approaches that seek to recover the continuous location parameters exactly. One example is the finite rate of innovations techniques of Vetterli, Marziliano, and Blu, [42]. The recent work of Tang, Bhaskar, Shah, and Recht addresses continuous localization through minimization of the atomic norm, [43].

We use the approach of applying discretization of spatial location parameters, and controlling for modeling error. Existing work on this subject of handling location discretization error include the *Spar-*

sity Cognizant Total Least Squares approach of Zhu et. al., [44], and the *General Deviants* model of Herman and Strohmer, [45]. In the total least squares work, an additive matrix perturbations model is used and the nonconvex joint optimization of perturbation variables and the sparse parameter vector is treated. In our approach, we do not significantly change the form of the inversion algorithm. Instead of optimizing over the potential model perturbations, we expand the constraints on the error residual. In addition, there are notable results on the potential ℓ_1 recovery stability of off-grid sampling for angle estimation, by Boufounos, Cevher, Gilbert, Li, and Strauss, in [46].

Another robust approach to off-grid error is given by Ekanadham, Tranchina, and Simoncelli in [47], in which an interpolated approximate translational error in one dimension is incorporated with learned offset values in the basis pursuit optimization. A simple comparison is that our residual constraint is expanded by an approximation that is constant with grid-offset value, whereas their approach estimates the offset values using a first-order Taylor approximation of quantization modeling error. Their analysis, found the same tradeoffs that we observe between lower modeling error through fine discretization versus increasing coherence between basis functions, which impacts the chances of stable sparse recovery.

The general deviants method of Herman and Strohmer is similar to our work. This method finds new restricted isometry constants that can still guarantee stable parameter recovery under the approximate observation model. In contrast, our analysis, detailed in [48, 49], follows the approach of Donoho, Elad, and Temlyakov, [11], deriving bounds based on mutual coherence for statements regarding error stability.

Overview

We motivate the approach of model-robust basis pursuit in section 4.1 by deriving a lower-bound on the inversion error due to model mismatch in the usual basis pursuit, under worst case, constant-norm noise vectors. In section 4.3, we upperbound the estimation error of model-robust estimates, showing that local stability results are possible for some compressed sensing problems. We show lower bounds on reconstruction error in the presence of modeling error in section 4.2.

Section 4.4 presents an analysis of the model error and mutual coherence in power law propagation systems as it depends on the quantization lattice spacing parameter. Subsections 4.4.1 and 4.4.2 derive these constants for one dimensional lattice systems and inverse square distance loss functions. Subsection

4.4.3, uses numerical experiments to show the effect of the power law exponent γ and spacing parameter in both one dimensional and two dimensional systems.

4.1 Perturbations to the Linearized Model

We now relax the assumption that points sources must lie exactly on our 2D lattice. Rather, they can occur continuously in the field. Our object remains to localize sources to their nearest lattice point, as well as estimate their magnitudes. We may represent the difference between what sensors measure and what is modeled by the discrete set of lattice locations by an additive perturbation to the observation matrix. The observations at all sensors are now described by

$$\mathbf{x} = \Psi\mathbf{v} + \mathbf{n} = (\Psi_0 + E)\,\mathbf{v} + \mathbf{n} \tag{4.1}$$

Here, the matrix $\Psi \in \mathbb{R}^{L\times N}$ represents the actual observation system with each column corresponding to the sampled propagated field due to the actual continuous source location. The matrix Ψ_0 denotes the model that we assumed by quantizing source locations to a lattice, and E is a matrix of perturbations to this assumed model.

Taking this discretized modeling error into account, the Noise-Aware Basis Pursuit method for reconstructing sparse solutions gives an altered fit constraint.

$$(P_{1,\epsilon}) \qquad \begin{array}{ll} \text{minimize} & \|\mathbf{v}\|_1 \\ \text{subject to} & \|\mathbf{y} - \Phi\Psi_0\mathbf{v} - \Phi E\mathbf{v}\|_2 \leq \epsilon \end{array} \tag{4.2}$$

We do not know *a priori* the particular E that will be observed, and jointly minimizing over both E and $\mathbf{v}$ is a nonconvex problem, as formulated and treated in [44]. Instead, we will try to bound the norm of the additional error vector, noting that the vector $\mathbf{v}$ is sparse. We assume $\mathbf{v}$ has nonzeros on the set of indices T, where $|T| \leq S$.

We recognize structure to the perturbation matrix, namely, the columns of $E = \begin{bmatrix} \mathbf{e}_1 & \cdots & \mathbf{e}_N \end{bmatrix}$, are given by the difference between sensor measurements of a source at the lattice point as compared with the source shifted in each of the spatial dimensions by as much as $\pm\delta/2$, where δ is the lattice step-size. The additional error term $E\mathbf{v}$ can be decomposed as a sum over indices T describing the support of $\mathbf{v}$;

as in $E\mathbf{v} = \sum_{i \in T} \mathbf{e}_i v_i$. The difference vectors $\mathbf{e}_i \in \mathbb{R}^N$ are drawn from the set of allowed column perturbations, which we call Δ_i. Each set is defined as

$$\Delta_i = \{\, \mathbf{e} : \; e_j = \psi_{\mathbf{p}_i + \mathbf{t}}(\mathbf{l}_j) - \psi_{\mathbf{p}_i}(\mathbf{l}_j), \;\; \text{for } 1 \leq j \leq N, \; t_1, t_2 \in (-\delta/2, \delta/2] \,\} \tag{4.3}$$

where $\mathbf{l}_j$ represents the 2D position of sensor j, $(\mathbf{p}_i + \mathbf{t})$ the actual 2D position of the source within range of lattice point i, and δ the distance between adjacent points on the lattice, a sampling resolution parameter.

In cases when the observations are lower dimensional projections of the field measurement vector, as described by (2.7), the perturbation system is correspondingly modified to be a linear transformation of the original set.

$$\mathbf{y} \;\; = \;\; \Phi\Psi\mathbf{v} + \Phi\mathbf{n} = \Phi\left(\Psi_0 + E\right)\mathbf{v} + \Phi\mathbf{n} \tag{4.4}$$

$$\Delta_i \;\; = \;\; \{\, \Phi\mathbf{e} : \; e_j = \psi_{\mathbf{p}_i + \mathbf{t}}(\mathbf{l}_j) - \psi_{\mathbf{p}_i}(\mathbf{l}_j), \;\; \text{for } 1 \leq j \leq N, \; t_1, t_2 \in (-\delta/2, \delta/2] \,\} \tag{4.5}$$

In this case, model perturbations depend both on the propagation model as well as the projection matrix Φ.

By bounding the norm of vectors from each Δ_i, we may bound the overall fit error. Let g_i be an upper bound on the norm of any vector from Δ_i.

$$g_i \;\; = \;\; \max_{\mathbf{e} \in \Delta_i} \|\Phi\mathbf{e}\|_2 \tag{4.6}$$

Performing a search over each set for these extrema may actually be simplified when the type of environmental model is unimodal and isotropic, so that the maximum is achieved at one of the points $\mathbf{t}^T \in \{[\delta/2, \delta/2], [\delta/2, -\delta/2], [-\delta/2, \delta/2], [-\delta/2, -\delta/2]\}$. This computation may be done during network set-up, while computing the observation matrix. From these column-norm upperbounds, we can now bound the overall fit error, which we define as

$$\mathbf{f} \;\; = \;\; \Phi E\mathbf{v} + \mathbf{z} \;\; = \;\; \sum_{i \in T} v_i \Phi\mathbf{e}_i + \mathbf{z} \tag{4.7}$$

We recall that $\mathbf{z}$ is the additive measurement noise vector that we bounded in norm by ϵ. Using the triangle inequality property of norms, the upperbound on the overall fit error is

$$\|\mathbf{f}\|_2 \;\; \leq \;\; \sum_{i \in T} |v_i| g_i + \epsilon \tag{4.8}$$

For source magnitudes that cannot be negative, and defining $\mathbf{g} = [g_i]_{i=1}^N$, this leads to the simple expression

$$\|\mathbf{f}\|_2 \quad \leq \quad \mathbf{g}^T\mathbf{v} + \epsilon \tag{4.9}$$

We now use this to expand the fit constraint of the noise-aware basis pursuit program. We propose the following model-robust formulation of a noisy compressed sensing estimator.

$$\begin{array}{lll} & \text{minimize} & \mathbf{1}^T\mathbf{v} \\ (P^+_{1,\epsilon,\mathbf{g}}) & \text{subject to} & \|\mathbf{y} - A_0\mathbf{v}\|_2 \leq \epsilon + \mathbf{g}^T\mathbf{v} \\ & & \mathbf{v} \geq \mathbf{0} \end{array} \tag{4.10}$$

Even in cases where we do not know the sign of $\mathbf{v}$, we can formulate the optimization by splitting $\mathbf{v}$ into positive and negative parts; i.e., $\mathbf{v} = \mathbf{v}_+ - \mathbf{v}_-$, the cone constraint of the original basis pursuit optimization may be modified to include this model error.

$$\begin{array}{lll} & \text{minimize} & \mathbf{1}^T\mathbf{v}_+ + \mathbf{1}^T\mathbf{v}_- \\ (P_{1,\epsilon,\mathbf{g}}) & \text{subject to} & \|\mathbf{y} - \Phi\Psi_0(\mathbf{v}_+ - \mathbf{v}_-)\|_2 \leq \epsilon + \mathbf{g}^T(\mathbf{v}_+ + \mathbf{v}_-) \\ & & \mathbf{v}_+ \geq \mathbf{0}, \quad \mathbf{v}_- \geq \mathbf{0} \end{array} \tag{4.11}$$

These model-robust formulations of basis pursuit remain second order cone programs, and so can be efficiently solved for their global minimum using off-the-shelf solvers such as `cvx`, [39]. In each program, the cone constraint has been expanded by a linear term involving the magnitudes of source intensities. In section 4.3, we will show what implications this may have on the stability of recovery. In section 4.2, we first demonstrate conditions that would necessitate a modification of the original noise-aware basis pursuit program, by examining lower bounds on the error of basis pursuit estimates as it depends on the extent to which (2.7) is violated. Section 4.4 considers the model error contants and system coherence in power law propagation models. Section 4.5 concludes the chapter.

4.2 Lower Bound on Error of Basis Pursuit

In order to study conditions necessitating the expansion of the model fit constraint in basis pursuit, we first assume that the additive noise vector has a known norm. Rather than this $\mathbf{z}$ being Gaussian, we call it

a worst-case perturbation vector with norm equal to ϵ. We term this worst-case because the ϵ previously referred to a 99% probability bound on the norm of sensor noise. The exact values of this norm-ϵ noise vector are still unknown.

First, we check for when the modeling error could make the optimization of (3.12) infeasible. We recap the noise-aware basis pursuit program from chapter 3, for $A = \Phi\Psi_0$, below.

$$\begin{aligned} &\text{minimize} && \|\mathbf{v}\|_1 \\ &\text{subject to} && \|\mathbf{y} - \Phi\Psi_0\mathbf{v}\|_2 \leq \epsilon \end{aligned} \tag{4.12}$$

The optimization is infeasible when the set $F_1 = \{\mathbf{v} : \|\mathbf{y} - \Phi\Psi_0\mathbf{v}\|_2 \leq \epsilon\}$ is empty. However, the dimensions of the sensing matrix $A \in \mathbb{R}^{m\times N}$ typically satisfy $m \leq N$. In these situations, there is a $\mathbf{v}$ such that $\mathbf{y} = A\mathbf{v}$, and so F_1 is not empty. Infeasibility of basis pursuit becomes a consideration when there are more measurements than the dimension of $\mathbf{v}$, creating an overdetermined system. In that situation, expanding the constraint to include modeling error may be required for feasibility, especially when assuming worst-case noise, or if ϵ is small compared to the magnitudes of sources and the norm of structural errors.

In most situations, the basis pursuit program will remain feasible. However, the feasible set is no longer guaranteed to contain the true parameter vector. Denoting the true source vector as $\mathbf{v}_0$, we may find the distance between $\mathbf{v}_0$ and the feasible set of (4.12), denoted $d_{F_1}(\mathbf{v}_0)$. This provides a lower bound on the estimation error of basis pursuit, since all potential estimates must be in the feasible set. The feasible set can be rewritten as

$$F_1 \quad = \quad \{\mathbf{v} : \|\Phi\left(\Psi_0 + E\right)\mathbf{v}_0 + \mathbf{z} - \Phi\Psi_0\mathbf{v}\|_2 \leq \epsilon\} \tag{4.13}$$

Define the vector, $\mathbf{b} = \mathbf{z} + \Phi E\mathbf{v}_0$, to be the noise vector and model error combined. Let the difference between $\mathbf{v}$ and $\mathbf{v}_0$ be denoted by $\mathbf{w}$.

$$F_1 \quad = \quad \{\mathbf{v} : \|\Phi\Psi_0\mathbf{w} + \mathbf{b}\|_2 \leq \epsilon\} \tag{4.14}$$

Suppose that $\|\mathbf{b}\|_2$ is known to be greater than ϵ. This is especially likely when $\|\mathbf{z}\|_2$ alone equals ϵ. Then, the squared distance of $\mathbf{v}_0$ to the feasible set is the solution to the minimization

$$\begin{aligned} &\text{minimize} && \mathbf{w}^T\mathbf{w} \\ &\text{subject to} && \mathbf{w}^T(\Phi\Psi_0)^T(\Phi\Psi_0)\mathbf{w} - 2\mathbf{w}^T(\Phi\Psi_0)^T\mathbf{b} + \mathbf{b}^T\mathbf{b} - \epsilon^2 \leq 0 \end{aligned} \tag{4.15}$$

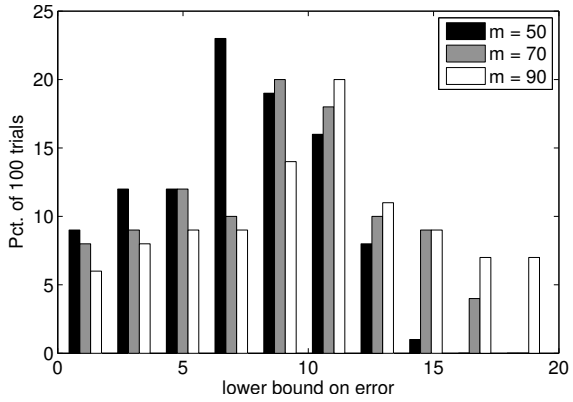


Figure 4.1: Histogram for 100 trials of computed basis pursuit error lower bounds for 3 randomly located acoustic sources each with magnitude 15. One hundred sensors cover the area in a lattice arrangement with spacing of 2. They perform consensus to find m projections of a randomly drawn orthonormal matrix, Φ. Three values of m are compared. The sensor noise vector $\mathbf{z}$ is drawn uniformly from vectors with norm 0.3526. Values of g_i range from 0.7526 to 0.7532.

which is a quadratically constrained quadratic program (QCQP), [27]. Strong duality holds in this problem as long as there is a Slater point; i.e., a strictly feasible $\mathbf{w}$, which is guaranteed when $\Phi\Psi_0$ is not overdetermined. As a result, the problem can be solved through its dual, which is a convex maximization of $g(\lambda)$ over the dual variable λ, in this case a scalar argument.

$$\begin{aligned} g(\lambda) &= \min_{\mathbf{w}} \left\{ \mathcal{L}(\mathbf{w}, \lambda) = \mathbf{w}^T \left(I + \lambda A_0^T A_0 \right) \mathbf{w} - 2\lambda \mathbf{b}^T A_0 \mathbf{w} + \lambda \left(\mathbf{b}^T \mathbf{b} - \epsilon^2 \right) \right\} && (4.16) \\ &= -\lambda^2 \mathbf{b}^T A_0 \left(I + \lambda A_0^T A_0 \right)^{-1} A_0^T \mathbf{b} + \lambda \left(\mathbf{b}^T \mathbf{b} - \epsilon^2 \right) && (4.17) \\ d_{F_1}(\mathbf{v}_0)^2 &= \max_{\lambda} \left\{ g(\lambda) : \lambda \geq 0 \right\} && (4.18) \end{aligned}$$

The solution for the distance $d_{F_1}(\mathbf{v}_0)$, will depend on the particular vector $\mathbf{b}$ observed.

To gauge the effect of model mismatch on the lower bound of the basis pursuit error, we computed the minima of (4.15) via its dual for 100 acoustic source localization tests with $N = 100$. Figure 4.1 shows the histogram of distances for various dimensions m, which corresponds to the number of randomly generated orthogonal projections computed through Φ. The sensors are arranged on a regular 10-by-10 lattice, with lattice spacing $\delta = 2$, and the three sources are sampled from anywhere in the monitored area, each with magnitude 15. The additive measurement noise has norm $\epsilon = 0.35$.

As expected, the model fit errors increased with the dimension of the projection m, resulting in higher lower bounds on the error for a larger number of projections. There is a wide range of lower bounds computed because of variability in the $\mathbf{b}$ observed, since sources may occur closer or further

from the lattice points, depending on where they are located. We have observed larger computed lower bounds when the magnitudes of sources are larger as compared with the norm bound on the additive noise vector. The benefit of accounting for model fit error in this way is dependent on it being an effect that is not overshadowed by the additive sensor noise.

4.3 Upper Bound on the Error of Model-Robust Basis Pursuit

In this section, we discuss predictions on the stability of model-robust basis pursuit (MRBP), similar to those of the original noise-aware basis pursuit summarized in chapter 3 in (3.16). By stability, we refer to finding an upperbound on the ℓ_2 estimation error. Because of the expansion of the fit constraint in MRBP by a term that involves source magnitudes, we will need to impose a restriction on the magnitudes of nonzero terms to get a similar stability result. Intuitively, if we allow elements of the true source vector $\mathbf{v}_0$ to have arbitrarily large magnitude, then the constraint set, denoted as F_2, for $(P^{+}_{1,\epsilon,\mathbf{g}})$ as defined in (4.10) would also grow arbitrarily large.

$$F_2 \;=\; \left\{\mathbf{v} : \|\Phi\Psi_0\left(\mathbf{v}-\mathbf{v}_0\right) - \Phi E\mathbf{v}_0 - \mathbf{z}\|_2 \leq \epsilon + \mathbf{g}^T\mathbf{v}\right\} \tag{4.19}$$

We impose a bound on the elements of $\mathbf{v}_0$.

$$|v_i| \leq v_{\max},\ \ \forall\ i \tag{4.20}$$

In this case, a clear extension to the proof of the stability of basis pursuit in [11] yields an upperbound on the error of MRBP estimates.

Theorem 4.3.1 For a source vector, $\mathbf{v}_0$, satisfying $\|\mathbf{v}\|_0 \leq S$ and $\max_i |v_i| \leq v_{\max}$; and system $\mathbf{y} = \Phi\left(\Psi_0 + E\right)\mathbf{v}_0 + \mathbf{z}$, with coherence $\mu(\Phi\Psi_0)$ given by (3.15), E satisfying $\max_i \|\Phi\mathbf{e}_i\|_2 \leq g_{\max}$, and $\|\mathbf{z}\|_2 \leq \epsilon$, then, for $S < \frac{1}{4}\left(\frac{1}{\mu}+1\right)$, the estimation errors of (4.11) and (4.10), are upperbounded by

$$\|\widehat{\mathbf{v}} - \mathbf{v}_0\|_2 \leq \frac{2\left(\epsilon + S\cdot g_{\max} v_{\max}\right)}{\sqrt{1-\mu(4S-1)}} \tag{4.21}$$

The proof follows from a modification of the stability analysis of BP given by Donoho, Elad, Temlyakov in [11], and is detailed in appendix B. We note that, similar to [11], we find the recovery error bound

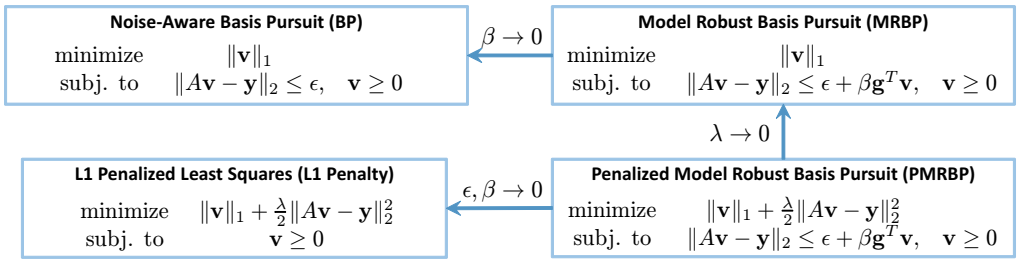


Figure 4.2: Relationships between the four compared sparse recovery methods. All methods shown include the optional positivity constraint.

derived by this method to be very loose and that, in practice, we rarely meet the strict coherence requirement of $S < (1/\mu + 1)/4$, while still observing reasonable reconstruction errors. In comparison, Herman and Strohmer in [45] also model the error due to perturbations of the linear observation model and conditions for stability of noise-aware basis pursuit. However, their analysis is based on modification of RIP constants. While the error bound may be loose, the advantage is that the mutual coherence for the system of interest in our application is easily computable, whereas RIP constants are much more difficult to verify.

The enlarged cone constraint of (4.10) can reduce the estimation performance of MRBP, so we add an ℓ_2 fit penalty to the objective function, calling this method penalized model-robust basis pursuit (PMRBP). In addition, the parameter, $\beta \in [0, 1]$, is added to the cone constraint expansion, which allows adjustment of these worst-case perturbation bounds. Figure 4.2 shows the comparison of sparse estimators, with parameter $\beta \in [0, 1]$ relating the four methods. We see from the diagram the close relationship between the various estimators. The MRBP method allows for expansion of the residual cone constraint to account for the multiplicative-type structured model errors and is tuned by a parameter $\beta \in (0, 1]$ adjusting how much to expand the constraint. The PMRBP method combines the new cone constraint with the traditional penalized objective of ℓ_1 penalty and depends on the trade-off parameter λ. The next section explores the tradeoffs of these various sparse estimators and their performance in the presence of spatial quantization errors.

4.4 Analysis of Spatial Quantization Trade-offs in Power Law Propagation

We examine the impact of lattice spacing δ on the modeling error for observation systems having inverse power law propagation. To compute the worst-case column error of the observation system, we start by analyzing the system for the full collection of sensor observations $\mathbf{x}$, in (2.6). Since the $\Phi \in \mathbb{R}^{m \times N}$ of (2.7) are generally chosen to be selected rows of a random orthonormal matrix, the analysis of model error in the global system indicates the relation of model error to lattice spacing without having to choose the number of network projections or projection matrix.

We first derive, in subsection 4.4.1, the relationship of the worst case constants $g_{\max}$. In addition, we study system coherence μ, as it depends on δ, for the one dimensional case in subsection 4.4.2. Then we quantify the relationship numerically in the two dimensional case, in 4.4.3.

4.4.1 Model Error for Inverse Square Distance Propagation: One Dimension

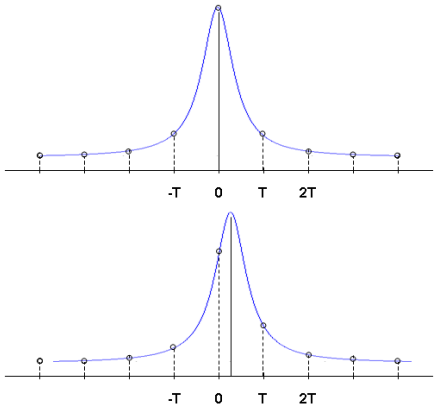


Figure 4.3: Illustration of the propagated field at the 0 position (upper) and at the 0 position plus an offset t_0 (lower).

Here, we derive the equation for model fit error, in a one dimensional sensing array. The propagation model is the inverse square distance from a source, assumed to be at the origin. We assume regularly spaced sensors along the line, spaced by distance T, which also corresponds to the δ of our source lattice. We also assume the line of sensors is infinite. In practice, since the power law energy dies quickly enough with distance, we observe the same result for finite, but large, numbers of sensors. Figure 4.3 illustrates

the set-up. The sensors measure the received signal strength (RSS), based on free space propagation given by

$$x_n = \frac{1}{d_{0,n}^2 + 1} = \frac{1}{(Tn)^2 + 1}, \quad n \in \mathbb{Z} \tag{4.22}$$

where $d_{0,n}$ is the distance from the origin to the n^{th} sensor, which is equal to Tn.

In the discretized model, we assume a source is located exactly at one of the discrete points, here the origin. Now we compare to the same system but now the source location is offset by a t_0, where $-T/2 < t_0 \leq T/2$. We call the actual signal $y(t)$.

$$y_n = \frac{1}{(Tn - t_0)^2 + 1}, \quad n \in \mathbb{Z} \tag{4.23}$$

The total model error of one source is equivalent to the bound on the maximum norm of the error for one column of Ψ, of (2.6). This is also the bound on the norm of one column of the perturbation matrix E, in (4.1). Denoting the error vector, or column of E, as e_n, given by

$$e_n = x_n - y_n \tag{4.24}$$

the squared error may be obtained from $\|\mathbf{e}\|_2^2 = \sum_{n=-\infty}^{\infty} |e(nT)|^2$. We note that $e(nT)$ is a sampled version of the continuous error signal.

$$e(t) = x(t) - y(t) \tag{4.25}$$

$$\text{for } x(t) = \frac{1}{t^2 + 1}, \text{and } y(t) = \frac{1}{(t - t_0)^2 + 1} \tag{4.26}$$

The goal is to compute the squared error difference in the observation vector due to an off-grid location. The approach is to use Parseval's relation for Fourier transform of a signal to evaluate the energy of the continuous-time error signal in the Fourier domain, converting this back to the discrete-time domain by sampling at rate T.

The Fourier transform of $x(t)$ has a well-known form, and $y(t)$ is a version of $x(t)$ shifted by t_0, yielding

$$X(j\omega) = \pi e^{-|\omega|} \tag{4.27}$$

$$Y(j\omega) = \pi e^{-|\omega|} e^{-j\omega t_0} \tag{4.28}$$

We may use the the Fourier Domain representation to compute the squared energy of the error

$$E(j\omega) \;=\; X(j\omega) - Y(j\omega) \;=\; e^{-|\omega|}\left(1 - e^{-j\omega t_0}\right) \tag{4.29}$$

since, by Parseval's relation, we know $\int |e(t)|^2 dt = \left(\int |E(j\omega)|^2 d\omega\right)/2\pi$. Then to apply sampling, we use

$$\int_{t=-\infty}^{t=+\infty} |e(t=nT)|^2\, dt \;=\; \frac{1}{T}\left(\frac{1}{2\pi}\int_{-\frac{\omega_s}{2}}^{\frac{\omega_s}{2}} |E(j\omega)|^2 d\omega\right) \tag{4.30}$$

where ω_s is the sampling frequency, which is $2\pi/T$. An important assumption to this computational approach is that aliasing due to the sampling by T is negligible. While the error signal is technically not bandlimited, it is limited such that sample rates for $0 < T \leq 1/2$ have neglible aliasing.

Appendix C, shows that the computation of the total squared error, $\frac{1}{2\pi T}\int_{-\frac{\pi}{T}}^{\frac{\pi}{T}} |E(j\omega)|^2 d\omega$, yields

$$= \frac{\pi}{T}\left(\frac{t_0^2}{4+t_0^2}\right) + \frac{\pi}{T}e^{\frac{-2\pi}{T}}\left(\frac{4}{(4+t_0^2)}\cos(\pi t_0/T) - \frac{2t_0}{(4+t_0^2)}\sin(\pi t_0/T) - 1\right) \tag{4.31}$$

which depends on T and the shift amount t_0, which can be between $-T/2$ and $T/2$. Because the propagation function is strictly increasing and strictly decreasing before and after its peak, respectively, we can argue the shift corresponding to the largest error is at the boundary of the bin, $t_0 = T/2$.

Figure 4.4 shows the computed squared errors (by the formula, and actual sampling). Recall that the formula assumes no aliasing, which becomes a factor for T larger than 0.5, completely invalidating (4.31) for $T > 1$. The experimentally computed error grows with T both in absolute terms and as a fraction of the total energy of the measured samples $x(nT)$. Although the model error levels off with very large sample rate T, the level or error approaches 100 percent of the energy measured from the source, and so, would not allow for detection.

This section derived an expression for the maximum model fit error due to off-grid location in the one dimensional inverse square law propagation system. It shows that the error is well behaved at small discretization lattice spacings. The next section examines the impact of this spacing on the mutual coherence of the global observation system.

4.4.2 Coherence for Inverse Square Distance Propagation: One Dimension

As described in section 3.2, the mutual coherence of the observation matrix is one way to predict performance of sparse estimators. To illustrate how this changes with the choice of source lattice sampling

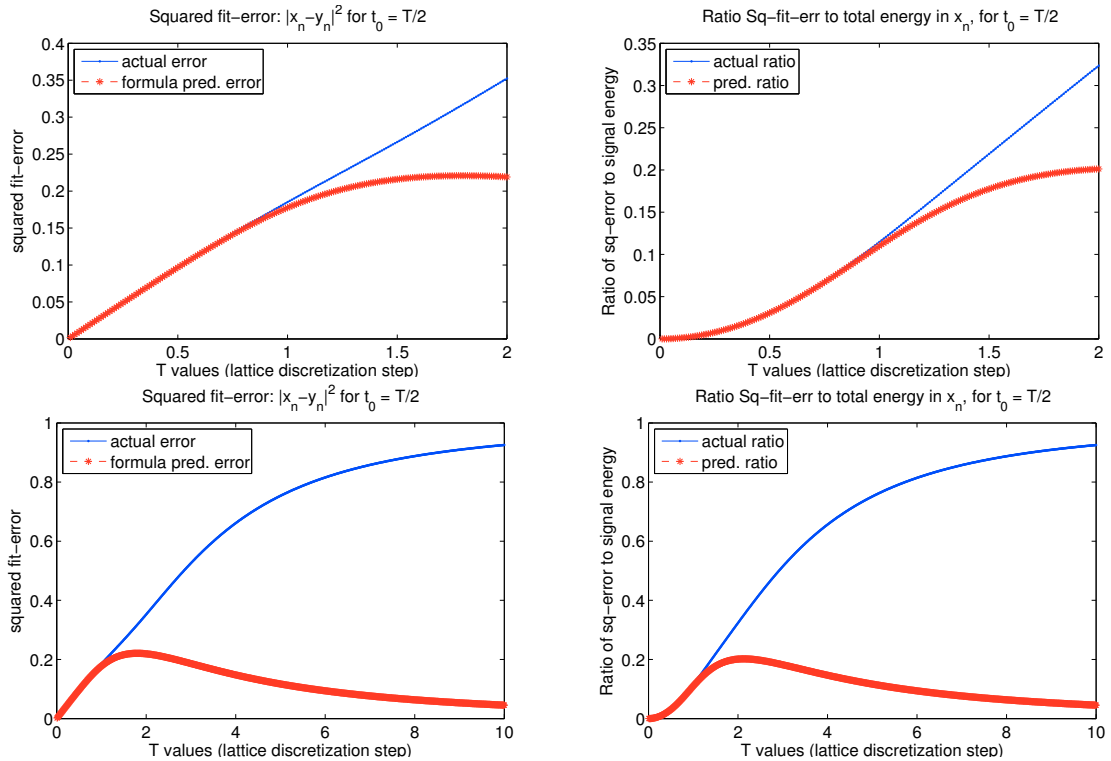


Figure 4.4: Worst case model error versus discretization spacing T for one-dimensional lattice and inverse square distance propagation. Left panels show squared error for the worst case offset, $t_0 = T/2$. The red asterisk series using the formula of (4.31), assuming no aliasing, while the blue dots series is the result of sampling and computing of the error for 2 million samples. Right panels: showing the ratio of the model error to the total squared energy $x(nT)$. The upper panels zoom in to $0 < T \leq 2$, while the lower left panels show a larger range of T.

parameter δ, but herein referred to as T, we derive the relationship of the mutual coherence versus T for the one dimensional lattice with inverse square law propagation.

We examine the mutual coherence $\mu(\Psi)$, for matrix Ψ of (2.6). The matrix Φ of (2.7) is generally chosen as rows of an orthonormal system, so the worst-case coherence is described by the coherence of Ψ. This will show the trend of this parameter, as it depends on the spacing T, without requiring the choice of the number of network projections or a specific projection matrix. We assume an infinite lattice of sensors that sample the source field, arranged in a lattice that coincides with the source lattice, as was done in the previous section.

Again we model the sampled signal due to one source as $x(nT)$,

$$x(nT) \quad = \quad \frac{1}{(Tn)^2 + 1}, \quad \forall \; n \in \mathbb{Z} \tag{4.32}$$

noting that shifts of this signal x make up the columns of the matrix Ψ. We recall that the mutual coherence is the maximum inner product between any two columns of a matrix.

$$\mu(A) \;=\; \max_{i \neq j} \frac{|\langle \mathbf{a}_i, \mathbf{a}_j \rangle|}{\|\mathbf{a}_i\|_2 \cdot \|\mathbf{a}_j\|_2} \tag{4.33}$$

for $\mathbf{a}_i$ and $\mathbf{a}_j$ distinct columns of A.

While in the previous section, we were interested in computing the norm of the difference of $x(nT)$ and $x(nT - t_0)$, here we look for the innerproduct of adjacent columns of A, $x(nT)$ and $x((n-1)T)$, which we denote as $\langle \mathbf{a}_1, \mathbf{a}_2 \rangle$. We again use the method of analysis of the continuous-time Fourier transform of the function, applying Parseval's relation and sampling by T. Appendix D computes the square energy of $x(nT)$ and the innerproduct of x with its shift by finding the $t = 0$ value of the convolution signal $n[0] = (\mathbf{a}_1[n] * \mathbf{a}_2[n])\,[0]$, again using their Fourier transforms.

The result is that the energy of a signal $\sum_{n \in \mathbb{Z}} |x_n|^2 = \|\mathbf{a}_1\|_2^2 = \|\mathbf{a}_2\|_2^2$ is given by

$$\sum_{n \in \mathbb{Z}} |x(nT)|^2 \;=\; \frac{\pi}{2T}\left(1 - e^{-2\pi/T}\right) \tag{4.34}$$

as shown in section D.1

The absolute innerproduct of two adjacent columns of Ψ is given by $|\langle \mathbf{a}_i, \mathbf{a}_j \rangle| = \sum_{n \in \mathbb{Z}} |x(nT)x((n-1)T)|^2$

$$\langle \mathbf{a}_i, \mathbf{a}_j \rangle \;=\; \frac{1}{2\pi}\int N(j\omega)e^{j\omega t}d\omega\Big|_{t=0} \;=\; \frac{1}{2\pi}\int N(j\omega)d\omega \tag{4.35}$$

$$\;=\; \frac{1}{2\pi}\int X(j\omega)^2 e^{-j\omega T}d\omega \tag{4.36}$$

$$\;=\; \frac{\pi}{2T}\left[\frac{1}{(2+jT)}\left(1 + e^{-2\pi/T}\right) + \frac{1}{(2-jT)}\left(1 + e^{-2\pi/T}\right)\right] \tag{4.37}$$

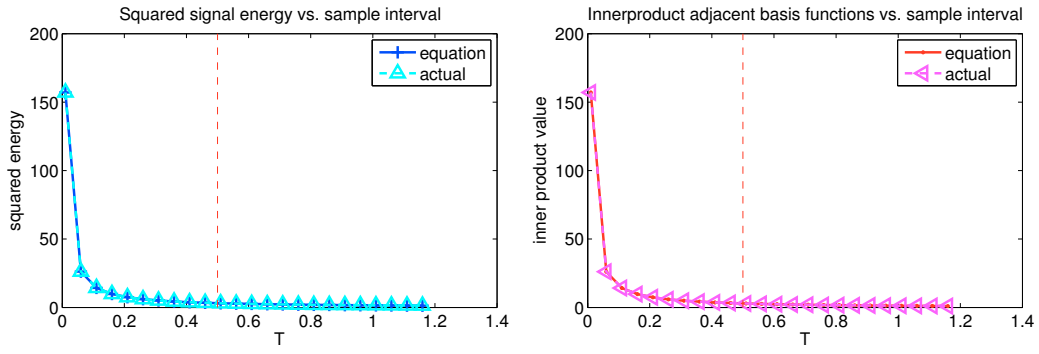
as is shown in section D.2.

Putting the energy and innerproduct expressions together gives an expression for the coherence as a function of T.

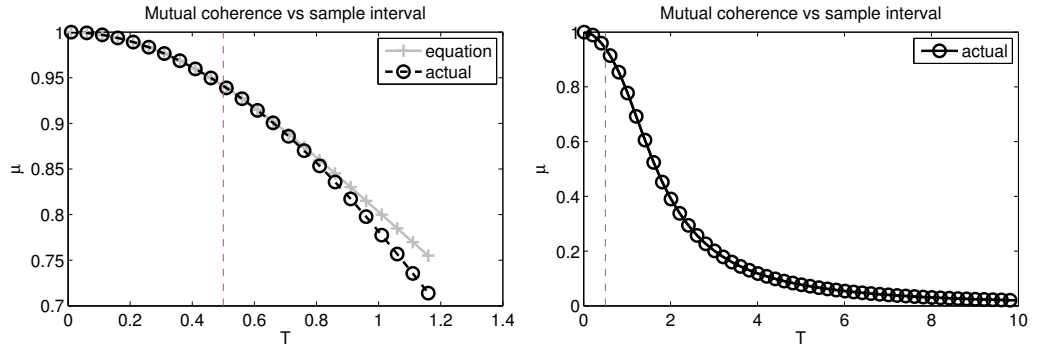
$$\mu \;=\; \frac{|\langle \mathbf{a}_i, \mathbf{a}_j \rangle|}{\|\mathbf{a}_i\|_2 \cdot \|\mathbf{a}_j\|_2} \;=\; \frac{\frac{\pi}{2T}\left(\frac{4}{4+T^2}\right)\left(1 + e^{-2\pi/T}\right)}{\frac{\pi}{2T}\left(1 - e^{-2\pi/T}\right)} \tag{4.38}$$

$$\;=\; \left(\frac{4}{4+T^2}\right)\frac{\left(1 + e^{-2\pi/T}\right)}{\left(1 - e^{-2\pi/T}\right)} \tag{4.39}$$

which we must be careful not to apply when the sample interval T is too large, leading to aliasing.



(a) $\|\mathbf{a}_i\|_2^2$ (b) $\langle \mathbf{a}_i, \mathbf{a}_j \rangle$



(c) coherence for $T \in (0, 1.2)$ (d) coherence for $T \in (0, 10)$

Figure 4.5: Coherence study for one dimensional lattice with inverse square propagation law: Each panel compares the derived quantity versus the measured value computed using 10,000 sensor samples and spacing T. Panel (a) shows the squared energy of $x(nT)$, panel (b) shows the innerproduct of adjacent columns of Ψ, and panel (c) shows the resulting mutual coherence versus T. Panel (d) shows measured coherence for a larger range of T.

Figure 4.5 shows the squared energy of the basis functions and the inner product of adjacent basis functions versus sample rate, T. The bottom panels show the measured mutual coherence metrics for the same range of T in panel (c), on the left, and for a larger range of T in panel (d), on the right. The predicted values of mutual coherence from (4.39) are high since, for $T > 1/2$, the formula is inaccurate as aliasing error becomes significant. The measured coherence in panel (d) falls quickly with large spacing. However, as shown in the previous section the modeling error becomes large, and so lattice spacing parameters above $T = 2$ may not result in good recovery.

These sections, 4.4.1 and 4.4.2, explored the tradeoffs of choosing a lattice spacing parameter in the one dimensional lattice, using the inverse distance-squared propagation law. The relationships of worst case modeling error and mutual coherence versus spacing interval showed the tension between making

a fine lattice to keep error constants low and keeping basis functions that are distinct enough to allow stable recovery by sparse estimation. The next section shows these error and coherence parameters in the two dimensional system of interest, as they depend on lattice spacing as well as the exponent of the propagation power law.

4.4.3 Numerical Evaluation of $g_{\max}$ and μ versus 2D Lattice Spacing and Power Law γ

In the case of a two dimensional lattice, closed-form expressions for model error constants and mutual coherence are not easy to derive. Instead, we simulate the two dimensional lattice as a function of the spacing parameter T, using the inverse power law propagation function of (2.8), repeated below

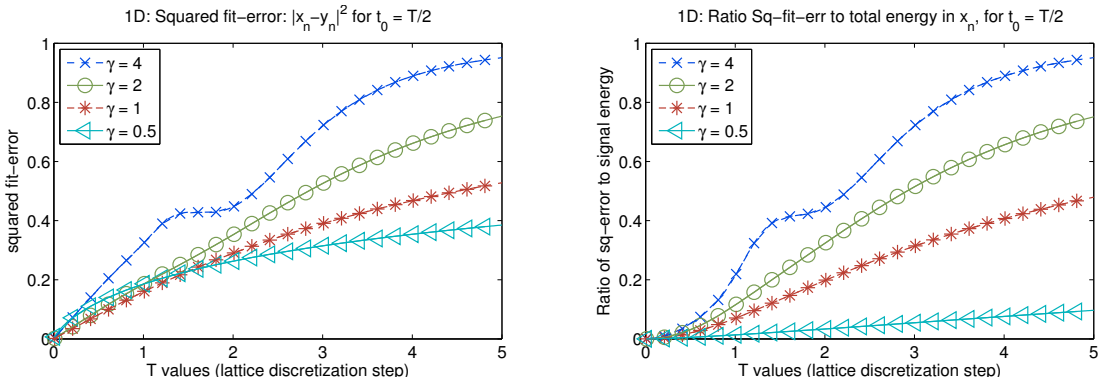
$$\psi_{\mathbf{p}_i}(\mathbf{l}_j) = \frac{1}{(\|\mathbf{l}_j - \mathbf{p}_i\|_2^{\gamma} + 1)} \tag{4.40}$$

The power laws we consider concern γ in the range from 1 to 4, modeling propagation losses for a variety of applications.

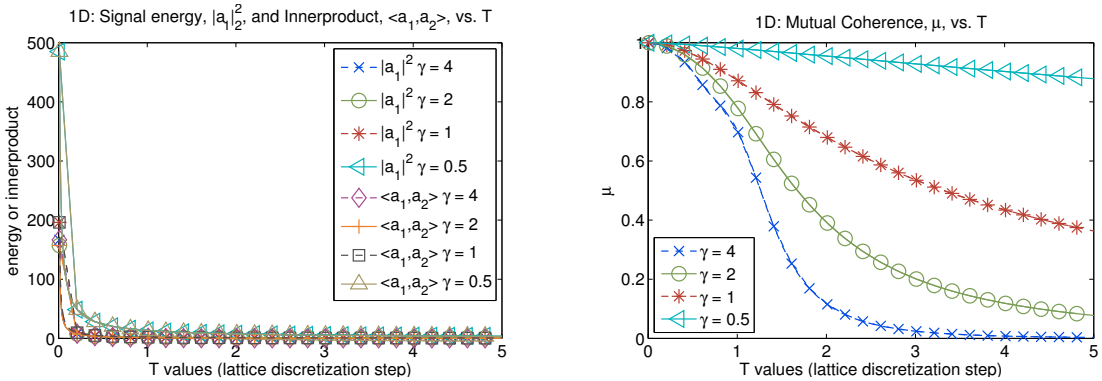
Figure 4.6 shows the measured $g_{\max}$ for the 1D lattice, but with various choices of γ. To simulate this, 10,000 sensors were modeled at each regular spacing and the squared error between the sampled field of a unit-magnitude source and a shifted version was computed. We see a higher sensitivity of model error to spacing T when γ is larger, due the more peaked shape of the loss function. In contrast, the mutual coherence is lower for large γ at a fixed T, suggesting increased distinguishability between sources at neighboring locations.

We perform the same experiment in a 2D lattice, as illustrated in figure 4.7. In this case a 400 by 400 array for sensors was modeled with spacing T in each direction. The maximum offset, $t_0 = T/2$, is assumed in both directions, resulting in an extremal boundary point of the quantized region.

Figure 4.8 shows the resulting modeling errors, $g_{\max}$, in panel (a). The behavior of the absolute error for $\gamma \geq 2$ is less intuitve. The drop in sampled energy with larger spacing makes it appear that model error is falling over some intervals. However, when we compare the ratio of model error to the total sampled energy of the field using the spacing T, we see that the percent error is strictly increasing with T, and for the large γ it increases at a high rate. On the other hand, the lower panel shows that mutual coherence falls quickly with T especially for the $\gamma = 4$ and 2 cases. This may indicate that fine spacings



(a) model error constants



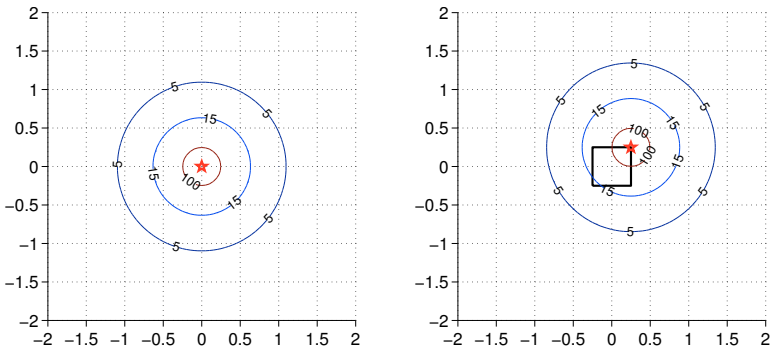
(b) mutual coherence

Figure 4.6: Worst case model error and coherence versus discretization spacing T for a 1D lattice and various inverse power propagation laws. The power γ is varied from 0.5 to 4, and 10,000 sensors are modeled. The left plot of panel (a) shows squared error for the worst case offset, $t_0 = T/2$, versus lattice spacing T, while the right shows the ratio of the model error to the total squared energy $x(nT)$. The left plot of panel (b) shows signal energy and innerproduct of adjacent bases, while the right plot shows the resulting mutual coherence vs. T

can perform well in the propagation environments with larger exponents to propagation loss. We test this with numerical simulations in the next chapter.

4.5 Conclusions

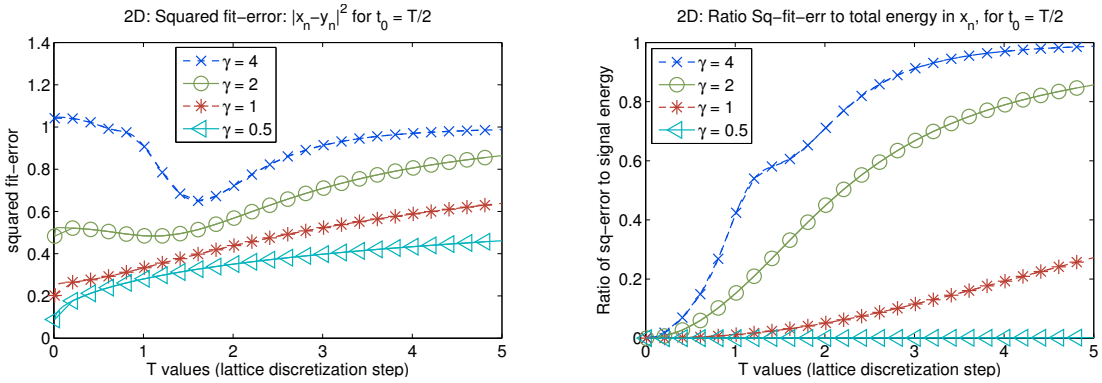
In this chapter, we studied the stuctured model error due to spatial quantization of continuous source locations. We derived bounds for the model errors of single sources and presented model-robust estimators, MRBP and PMRBP, which expand the cone constraint for noise-aware basis pursuit to account for



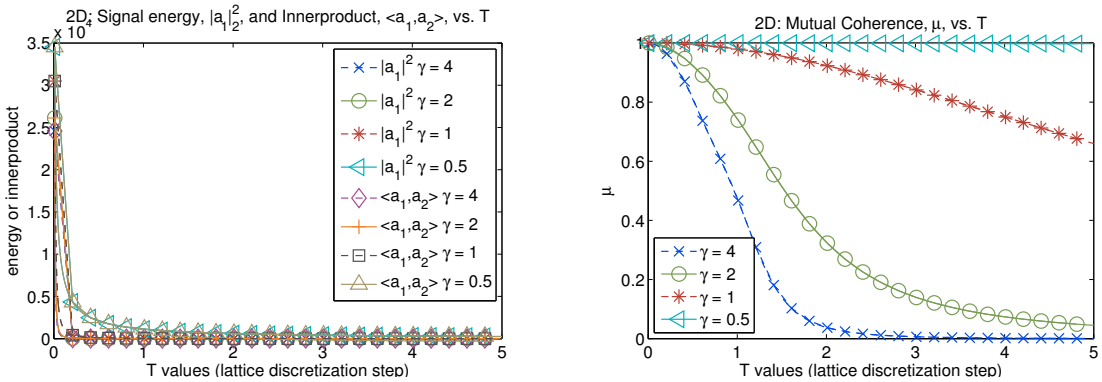
(a) on-grid source (b) off-grid source

Figure 4.7: Illustration of simulation of worst case modeling error in the 2D lattice for $T = 1/2$. Panel (a) shows a source exactly at a gridpoint with contours based in inverse squared distance, Panel (b) shows the source at location $[T/2, T/2]$ away from the grid point. The black box denotes the quantization region corresponding to the origin. Selected contours are labeled with the field level values.

the quantization error. We presented an upperbound on the ℓ_2 error of MRBP estimates and computed lower bounds on the reconstruction error of traditional basis pursuit due to off-grid source location. We analyzed the relationships of model error constants and mutual coherence as a function of lattice spacing in both one and two dimensional systems with power law propagation functions. The next chapter tests the various sparse estimators proposed in off-grid numerical experiments testing the sensitivity of each method to system and chosen tuning parameters.



(a) model error constants



(b) mutual coherence

Figure 4.8: Worst case model error and coherence versus discretization spacing T for a 2D lattice and various inverse power propagation laws. The power γ is varied from 0.5 to 4, using 160,000 sensors. The left plot of panel (a) shows squared error for the worst case offset, $(T/2, T/2)$, versus lattice spacing T, while the right shows the ratio of the model error to the total squared energy $x(nT)$. The left plot of panel (b) shows signal energy and innerproduct of adjacent bases, while the right plot shows the resulting mutual coherence vs. T

Chapter 5

Motivating Applications

In this chapter, we a present number of studies using numerical experiments to simulate FICCS in surveillance applications. We discuss the methodology of applying FICCS to three examples of distributed monitoring systems, each with differing sensing modalities that commonly obey inverse power propagation laws. Our approach is to describe various networked sensing applications and how they may be similarly modeled to cast them as linear discrete models with sparse underlying source parameter vectors. As we discussed in chapter 2, a network of omnidirectional microphones measuring the received signal strength (RSS) of an acoustic field can be adapted to the FICCS distributed inversion approach by quantizing spatial dimensions of the monitored area. We also present alternative sensing modalities that arise in radar and underwater sonar applications, and how these distributed sensing problems may also be cast in terms of FICCS. We recognize, through considering these various applications, that this approach for distributed field inversion depends only on the construction of a linear superposition model whose continuous parameter spaces may be quantized to a discrete sparse vector describing the sensed field.

5.1 Acoustic Sources Localization and Infrared Fields, $\gamma = 2$

This section discusses the application of FICCS to passive sensing regimes, using acoustic and infrared sensors. In chapter 2, we present distributed acoustic source localization as the motivating example application for our approach to distributed field estimation. We model a network of omnidirectional mi-

crophones sampling received signal strength (RSS) measurements of an acoustic field. We assume free-space propagation proportional to the inverse square distance between sources and sensors.By quantizing the spatial dimensions of the monitored area, we adapt the field inverison problem to FICCS. The result is the finite linearized model for sensor measurements $\mathbf{x}$, recapped below.

$$\mathbf{x} \quad = \quad \Psi\mathbf{v} + \mathbf{n}, \qquad \text{for} \quad \Psi \in \mathbb{R}^{N\times L} : \Psi = \left[\frac{1}{d_{j,i}^2+1}\right]_{1\leq i\leq N, 1\leq j\leq L} \tag{5.1}$$

where $d_{j,i}$ represents the distance between sensor j and the potential source location at index i of the quantized source lattice.

In the acoustic source regime and free-space propagation environment, it is typical to use the parameter $\gamma = 2$ in the inverse power law function (2.8). However, this parameter can be adjusted based on the experiment of interest. In addition, directional gain patterns of microphone receivers may be accounted for during the construction of the linear model (5.1) using calibation information and locations of the microphones. By assuming no directional preferences in receivers, our analysis treats the challenging case when no directional information is available, motivating intersensor cooperation.

This set-up can also be applied to a distributed monitoring network using low resolution infrared (IR) cameras. In the long-wave IR regime, the radiation collected by most objects is dominated by emissive components and, at long distances, unresolved objects can behave as sources with magnitudes based on emissivity-area and temperature. A collection of networked IR cameras can be deployed in the field, each integrating received energy over pixels and orientation. Using spatial quantization of the location dimensions of the monitored area, these sensors can apply FICCS to detect and resolve an unknown number targets with unknown magnitudes. A number of practical issues need to be addressed, such as accounting for time-varying radiation from the surrounding background. It is notable, however, that distributed field inversion may enable the use of low resolution passive sensors to accomplish monitoring tasks that are not be possible with a single, or a small number of sensors. Many of the studies in chapter 6 use this passive sensing model, corresponding to $\gamma = 2$, to test the effects of various system parameters in FICCS, and demonstrate the performance of this approach to field reconstruction in these applications.

5.2 Radar Surveillance Network, $\gamma = 4$

The distributed monitoring approach of FICCS may also be applied to networks of active sensors. This represents a significant difference in sensing modality, in which propagation losses must be accounted in two-way paths. Certain radar and active sonar sensors can be modeled in this $\gamma = 4$ power law regime. We refer readers to the introduction of [50] for more on modeling of range propagation losses in radar systems.

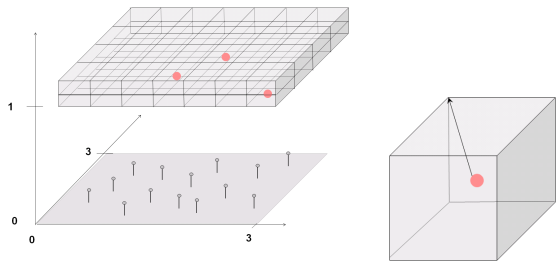
As an illustration, we focus on the application of multiple target detection in an area monitored by N distributed networked radars. Our set-up uses a common model for adapting compressed sensing to radar applications, similar to [51]. Each radar is assumed to have an isotropic antenna and emits narrowband pulses. We assume the networked radars are placed either regularly or randomly at the same altitude. The objective is to monitor a volume of air space above the radars. For example, we may place the radars in a 3 km by 3 km area to monitor the space above at altitudes between 1 and 1.5 km. Figure 5.1 shows this set-up in panel (a), with radars at the bottom and a 3D volume above them where targets may lie. Three targets are shown in as red circles.

An additional issue raised in active sensor networks is that the group of radars must coordinate to actively interrogate the region without interfering with the pulses of other radars. This maybe accomplished with time-division multiplexing as well as the use of orthogonal waveforms. In the simplest model, we may assume that each radar collects the returns of only its own pulses, as opposed to a full MIMO radar system, in which radars also process the bistatic returns from pulses generated by the other sensors.

Targets are modeled as deterministic point-targets, meaning their radar cross section (RCS) is constant from all aspects and does not vary probabilistically from one pulse to the next. In this analysis, we consider targets with zero Doppler shift, meaning they are stationary. The narrowband pulses transmitted by each radar are given by

$$p(t) \quad = \quad \text{Re}\left\{A\, w(t) e^{i2\pi f_c t}\right\} \tag{5.2}$$

where t represents time, A is a constant representing the gain of the amplifier, f_c is the carrier frequency of 3 MHz, and $w(t)$ is the square window function with a pulse width of 1 ms. Since we are considering scenarios with multiple targets, received radar returns of sensor i will contain the superposition of the



(a) sensors and target space (b) location bin

Figure 5.1: Depiction of the distributed radars and monitored discretized space with 3 targets, red circles, in (a), and, in (b), a close-up showing the potential location variation within a voxel.

returns from each of S targets.

$$x_i(t) \quad = \quad \sum_{k=1}^{S} \text{Re}\left\{A_k\, w(t-\tau_k) e^{i2\pi f_{d,k}(t-\tau_k)}\right\} \tag{5.3}$$

The above, $x_i(t)$, is assumed to be the received pulse after it has been demodulated to baseband frequency. The parameters A_k, τ_k and $f_{d,k}$ are unique to each target, representing respectively the amplitude of the received reflection due to target RCS and distance, the time-delay of the return due to distance, and the Doppler shift in frequency due to target velocity. Specifically, each parameter, is given by

$$A_k = \frac{\sigma_k P}{r_k^4}, \quad \tau_k = \frac{2r_k}{c}, \quad f_{d,k} = \frac{2f_c v_k}{c} \tag{5.4}$$

where σ_k is the target RCS, P is an overall system gain in Joules-square-meters ($J \cdot m^2$), r_k is the range from the radar to target k, v_k is the projection of the target velocity vector onto the radar's line of sight, and c is the speed of light. Since we are assuming stationary targets, returns will all have 0 Doppler; i.e., $f_{d,k} = 0$. We collect superimposed square waveforms containing only amplitude and delay information.

The continuous volume of potential target locations is discretized using a finite $L_x \times L_y \times L_z$ lattice, where the spacing between lattice points is given in each direction by δ_x, δ_y, and δ_z. Panel (a) of fig. 5.1 also shows these discrete bins covering the monitored area. The goal is only to localize detected targets to their bins, rather than estimate continuous target locations. As shown in panel (b), the actual location of a target within a bin may differ by up to half the lattice spacing in each direction, inducing error into the observation model.

We integrate the energy of the received pulses of each radar to obtain the discrete vector of observations at radar i, $\mathbf{x}_i$, proportional to the A_k values. The discrete target vector, $\mathbf{v} \in \mathbb{R}^L$, contains

magnitudes at potential target locations of the spatial lattice and has dimension $L = L_x L_y L_z$. Each entry of $\mathbf{v}$ will be zero, if there is no target in that location bin, or will contain the RCS of the target, if one is present. Since there are many more bins than targets, $\mathbf{v}$ is a sparse vector. We stack the pulse energy samples of each sensor into a single vector of all network sensor observations, $\mathbf{x}$. This vector has dimension NT where N is the number of sensors and T the number of time-samples. Based on the simple model of (5.3) and (5.4), the observations due to a unit-RCS target at the center of each bin are known and collected into the columns of the matrix Ψ. The result is a linear observation model for the collected observations at all sensors due to all discrete target locations,

$$\mathbf{x} = \Psi\mathbf{v} + \mathbf{n}, \;\; \mathbf{n} \in \mathbb{R}^{NT}, \qquad \text{for} \;\; \Psi \in \mathbb{R}^{N\times L} : \Psi = \left[\frac{1}{d_{j,i}^4+1}\right]_{1\le i\le N, 1\le j\le L} \tag{5.5}$$

where $\mathbf{n}$ is additive measurement noise.

We have shown the casting of a distributed narrowband radar surveillance network as a linear field inversion model with a sparse discrete source vector. Alternative formulations could use bistatic returns and directional information in the model. In many of the studies in chapter 6, we present results of numerical tests using the radar model, showing the potential of FICCS in applications of distributed monitoring using active sensors.

5.3 Guided Wave Applications, $\gamma = 1$

When studying, the effects of the exponent γ in power law propagation, in section 6.3.3, we will compare the inversion performance in the previous two regimes with the $\gamma = 1$ case. This corresponds to wave guided propagation, rather than spherical wave propagation, modeled in sections 5.1 and 5.2. This type of propagation can arrise in certain underwater passive sonar applications when the density variation with depth acts as a waveguide, [52]. In structural health monitoring, wave propagation is modeled by guided Lamb waves; see [53, 54]. While we will not elaborate on the modeling details, we recognize the importance of this regime, and include studies with the propagation law corresponding to inverse distance to assess the sensitivity of distributed inversion performance to the exponent γ.

5.4 Conclusions

In this chapter, we have discussed the application of the FICCS approach to distributed field inversion to distributed networks of passive and active sensors with varying sensing modalities. In the next chapter, we use numerical simulation experiments to compare the performance of the alternative sparse estimators to the localizations of targets that are continuously located in the monitored area. We present a number of studies detailing the system tradeoffs on field inversion and target detection performance.

Chapter 6

Numerical Simulation Testing in Surveillance Applications

In this chapter, we a present number of studies using numerical experiments to simulate field inversion experiments in the surveillance applications described in chapter 5. We show average ℓ_2 estimation and detection error computed through numerical simulation tests involving a small number of continuously located sources. We test robustness to the discretization of the source field, comparing the performance of the four estimators, summarized in figure 4.2. We measure field inversion performance by the average ℓ_2 estimation and detection error as they depend on the number of network projections used.

In addition to comparing the performance of the various sparse estimators, we study the sensitivity of FICCS to a wide variety of design parameters, operational choices, and system characteristics that may depend on the application. We organize the study of these considerations into two major categories. The first category involves parameters and choices related to the use of the field reconstruction algorithms, including penalization parameters, positivity constraints, and spatial quantization step-size. The second category contains the parameters that are dictated by the physics of the application. We include under these the signal-to-noise ratio (SNR) of sensor readings, the particular exponent for the inverse power law propagation function, and the locations of deployed sensors.

In 6.1, we give an overview of the methodology of numerical testing, listing the major studies conducted. The following sections then present the relevant studies and discussions regarding estimation

tradeoffs and system tradeoffs.

6.1 Overview of Experiments and Tradeoffs

In order to better understand the tradeoffs between the four estimators of fig. 4.2, we simulate sensor network surveillance experiments to model the FICCS algorithm, as given in chapter 3. For all of these tests, we simulate the placement of either 100 regularly-spaced or 150 randomly-placed sensors in a 20-by-20 area. Measurements are of the total received signal strength (RSS) for a field generated by a small number of passively or actively monitored sources, modeled as given by (2.8), with power law parameter $1 \leq \gamma \leq 4$. Randomly placed sensor locations are chosen sequentially and uniformly at random, rejecting any locations that are closer than the minimum spacing, $\delta/6$, to existing sensors. The sensor locations and communication radii determine the network topology, which allows for the distributed averaging in the sensing phase of FICCS, section 3.1.2. After confirming convergence of the node states to the first m rows of the randomly selected orthonormal matrix Φ using optimal uniform averaging weights, we focus on the field inversion phase of FICCS, in which the full field is estimated using sparse reconstruction.

A discrete $\sqrt{L}$-by-$\sqrt{L}$ lattice with spacing δ is imposed on the square region as the quantized set of potential source locations. In each test, $S = 3$ sources with magnitude 15 are present near distinct lattice points, selected at random. The exact locations of the sources are chosen uniformly in the 2D monitored area, inducing the spatial quantization error discussed in section 4. Sensor readings are corrupted by additive measurement noise, where $\mathbf{n}$ of (2.6) is chosen uniformly from the vectors with norm b_{σ^2}. The norm b is defined by the 95% norm-bound[1] of the Gaussian random vector with variance σ^2. Random network projections used in all trials are the first m rows of a randomly generated orthonormal matrix, determining the Φ of (2.7).

We model sources with positive magnitude, but magnitude information in not used by the estimators, except in the form of optional positivity constraints on source estimates. The sparse estimates are computed using the CVX optimization package, [39, 55]. Unless otherwise stated, the columns of the $A = \Phi\Psi$ matrix of (2.7), are normalized to have unit ℓ_2-norm before computing sparse estimates. The

[1]The 95% norm bound corresponds to the inverse chi-squared cumulative density function with L degrees of freedom.

saved column scale factors are then multiplied with the sparse estimate $\widehat{\mathbf{v}}$ to give the recovery estimate. When normalizing the columns of A for one of the model-robust estimators, the vector of perturbation constants $\mathbf{g}$ must also be re-scaled by the column scale factors. Section 6.2.5 will show the tradeoffs of sparse estimation with normalized versus un-normalized observation systems.

Estimation performance of the discrete target source vector is measured by the average ℓ_2 estimation error from the discretized true source vector, of (2.6); i.e., $\|\widehat{\mathbf{v}} - \mathbf{v}_{\text{true}}\|_2$. This metric depends on both source localization and magnitude estimation accuracy. Since ℓ_1 optimization methods are known to be biased for inferring magnitude, [56], we also study the detection rate of sources at fixed low false alarm rates. These are stringent performance metrics since a source that arrives near the border of two quantized regions could be detected in the neighboring cell and would be counted as both a miss and a false alarm. To show the impact of source location quantization, in each study, we compare the estimation error of each of the sparse reconstruction methods to the error of noise-aware basis pursuit, BP (3.12), for sources that lie exactly at their closest discrete grid-points.

The tests presented illustrate the tradeoffs of using the various sparse estimators in the FICCS algorithm. Section 6.2 presents the studies of sensitivity of inversion performance to choices regarding the estimators while section 6.3 studies the impacts of the parameters of the phyical system.

6.2 Studies of Estimation Tradeoffs

We first examine examples where the expansion of the basis pursuit cone constraint addresses feasibility issues of traditional noise-aware basis pursuit, even for moderate choices of the parameter β. Next, we study the sensitivity of the penalty methods to choices of the penalty parameter λ, which affects ℓ_1 penalized least squares and penalized model-robust basis pursuit (PMRBP).

We study the effect of the lattice spacing δ used for the quantized location grid-points. We present the average error and detection performance as a function of the lattice spacing, noting that, for a finely spaced lattice, the performance metrics and localization accuracy are stricter than in a coarse lattice. We show an example with high model error, in which the penalized model-robust basis pursuit estimator (PMRBP), of (4.10), can be tuned through β to trade off between the performance of traditional basis pursuit and the ℓ_1 Penalty, of (3.13).

We then present comparisons of reconstruction accuracy when using positivity constraints on estimated source magnitudes, as well as normalizing the columns of the system observation matrix A, given by (2.7), showing that while both practices are recommendable there are certain scenarios when using these options can hurt performance.

6.2.1 Cone Expansion Factor β

The model-robust basis pursuit (MRBP) and penalized model-robust basis pursuit (PMRBP) estimators were introduced to account for modeling errors due to spatial quantization, allowing for stable estimation using constraints on the observation fit residual. The estimators, detailed in chapter 4, expand the cone constraint of noise-aware basis pursuit, using bounds on the norms of induced model error for each source location in the lattice. The bounds are given by the vector $\mathbf{g} \in \mathbb{R}_{+}^{\mathbb{L}}$. The cone constraint is expanded to

$$\|\mathbf{y} - A\mathbf{v}\|_2 \quad \leq \quad \epsilon + \beta \mathbf{g}^T \mathbf{v} \tag{6.1}$$

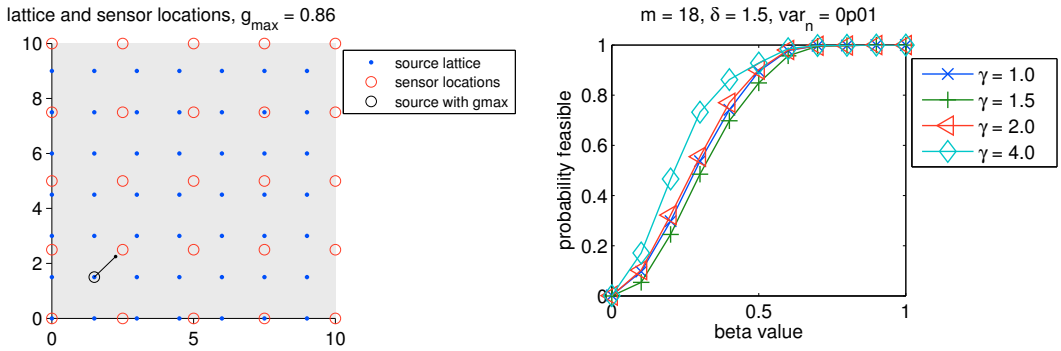
$$\mathbf{v} \quad \geq \quad 0 \tag{6.2}$$

in (4.10), which gives the positive magnitude version of MRBP. The scalar weight, $\beta \in (0, 1)$, is introduced as a tuning parameter of this algorithm. A choice of $\beta = 1$ would guarantee that the true source vector is in the feasible set. But because the vector $\mathbf{g}$ represents worst-case bounds, often the full expansion in not needed.

In this section, we first examine the effect of the β chosen on the probability that the original source vector remains a feasible solution to the model-robust basis pursuit estimators, MRBP and PMRBP. We show by numerical experiment, that moderate choices of β; e.g. 0.5, often covers the vast majority of all cases. Then we perform a sensitivity study, comparing all of the estimators, as β is varied, showing that, in cases with significant quantization error, average estimation and detection performance is sensitivity to β being sufficiently large, and as β becomes too large MRBP performance suffers, while PMRBP is able to maintain performance similar to the ℓ_1 Penalty estimator.

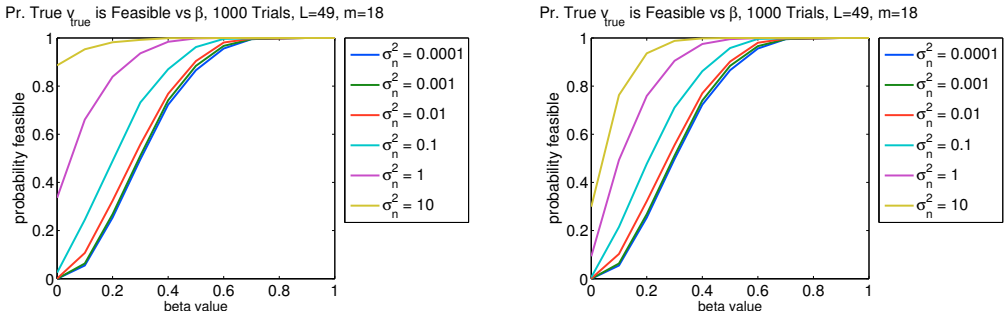
Probability of Feasibility versus Cone Expansion Factor β

We performed numerical simulation to test the effect on the choice of β on the likelihood that the original source vector is a feasible solution to the model-robust estimators. These tests also included the



(a) Set-up

(b) Prob. Feas. vs. β



(c) $m = 18$, $\gamma = 2$, Gaussian Noise

(d) $m = 18$, $\gamma = 2$, Exact-Norm Noise

Figure 6.1: Study of Feasibility of $\mathbf{v}_{\text{true}}$ vs. β: Panel (a) illustrates the set-up of the experiments, which use a regular grid of 25 sensors spaced by distance 2 and source lattice spacing $\delta = 1.5$. The maximum perturbation constant and its source location are also shown. Panels (b-d) show probability of feasibility over 1000 trials of the original source vector for residual cone constraints expanded using varying β, using $m = 18$ random orthonormal network projections. The series in (b) compare various propagation factors, for Gaussian measurement noise with variance 0.01. Panels (c) and (d) show the feasibility for varying measurement noise levels σ_n^2 in the case of Gaussian noise, and random noise with norm b_{σ^2}, respectively.

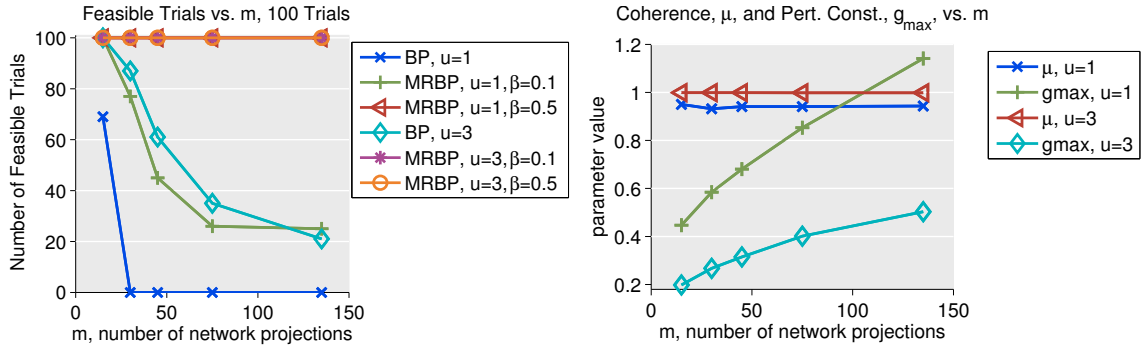
positivity constraint on source magnitudes, and additive measurement noise at various levels, translating to the 95% norm-bound ϵ, in (6.2). A regular 5-by-5 grid of sensors is deployed with spacing 2, and a source discretization lattice with spacing $\delta = 1.5$ is used. The set-up is illustrated in panel (a) of figure 6.1. Source vectors with 3 sources at continuous random locations and magnitudes of 15 were chosen. The network computes 18 random orthonormal projections (about 36% of the dimension L) during the consensus phase of FICCS. The feasibility of the true source vector was averaged over 1000 trials. Panels (b), (c), and (d) show the probabilities that the true vectors are feasible versus β for various propagation laws, additive noise types and levels. From (b), we see that propagation factor γ had little effect on the

feasibility of orignal $\mathbf{v}$. Comparing (c) and (d) we see that using random Gaussian measurement noise, especially with high variance, results in near 100% feasibility at small β. This is due to the worst-case nature of choosing the bound ϵ. When measurement noise is added that has exactly the norm ϵ, the effect is not as dramatic. However, it remained that for most noise levels over 80% of trials maintained feasible truth vectors for $\beta = 0.5$.

Unfortunately, the feasible set containing the true source vector does not guarantee a stable recovery. The next section tests the recovery error as they depend on β.

Estimation Performance versus Cone Expansion Factor β

In this section, we present a study of ℓ_2 estimation error and detection performance, as a function of the chosen cone expansion weight, averaged over 100 recovery trials. The experimental set-up uses 150 randomly-placed sensors, inverse square law propagation; i.e., $\gamma = 2$, and additive measurement noise with norm of b_{σ^2} for $\sigma^2 = 0.001$. Two resolutions of source location lattices were used; one with a coarse $\delta = 2$ spacing, and an upsampled by 3 version with $\delta = 2/3$. Estimation was performed enforcing the positivity constraint on source magnitudes.



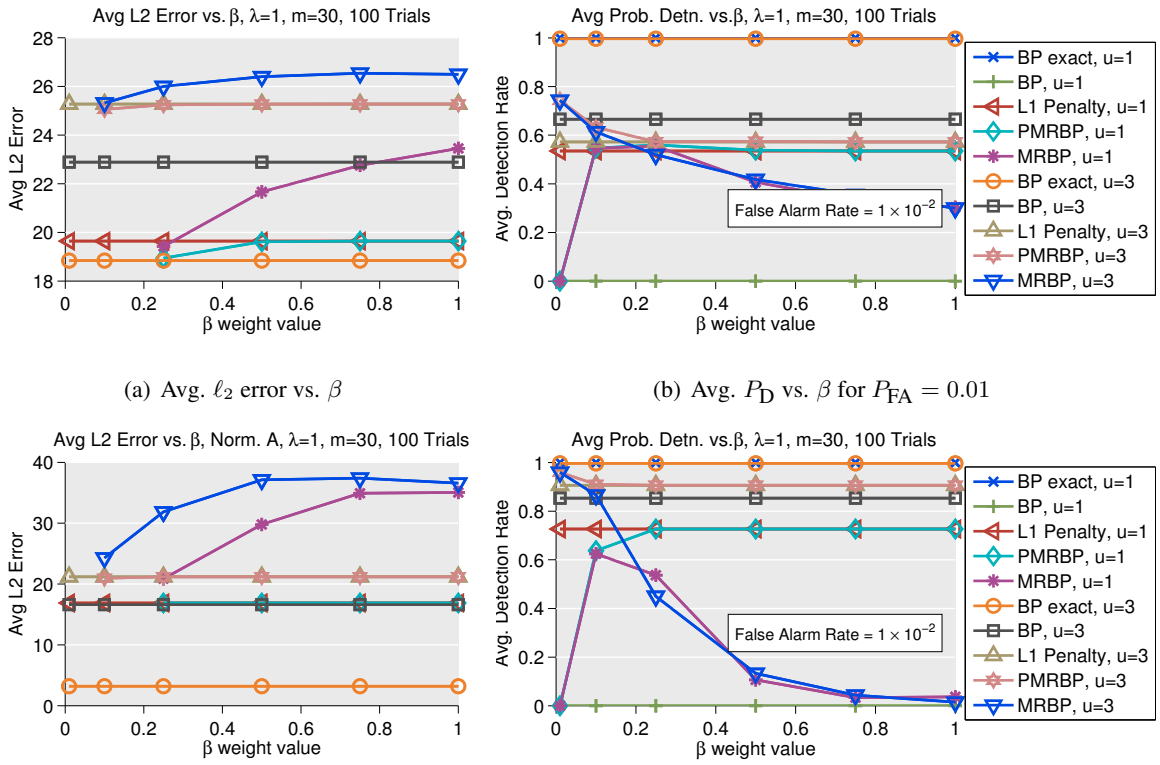
(a) Avg. Feasibility vs. m

(b) μ and $g_{\max}$ vs. m

Figure 6.2: Number of Feasible Trials and System Parameters vs. m, for two source lattice sampling rates, $u = 1, 3$ with $\delta = 2, 2/3$, $N = 150$ randomly placed sensors, random additive noise of norm $b_{0.001}$. Panel (a) shows feasibility of BP and (P)MRBP for various β at each lattice sample rate. Panel (b) compares average coherence μ to model perturbation constants $g_{\max}$ for the two lattice sample rates.

Because of the positivity constraint, it was possible to have infeasible basis pursuit estimates. The first study, shown in figure Figure 6.2 shows a comparison of feasibility, system coherence, and quantization error parameters for the two lattice resolutions, $u = 1$ and $u = 3$, and various values of the cone

expansion factor β. These quantities are shown as a function of the number of network projections m used. We see from panel (a) that basis pursuit, with unmodified cone constraints, suffers quickly from infeasibility with increasing numbers of measurements due to the increase in unaccounted model error. Even in the upsampled version with a finely spaced lattice, more than half of the trials are infeasible. In the model-robust versions, denoted by MRBP, the feasibility issue is addressed by moderate values of β, with all trials feasible for $\beta = 0.5$, even in the coarse lattice. Panel (b) shows the system coherence and maximum perturbation constants versus m, showing that the worst case model error is much lower in the finely sampled lattice than in the coarse one and increases with m. Mutual coherence is high, suggesting that performing field inversion in this scenario may be challenging. Figure 6.2 shows the new cone constraint can address feasibility issues of traditional basis pursuit under model uncertainty.



(a) Avg. ℓ_2 error vs. β

(b) Avg. P_{D} vs. β for $P_{\mathrm{FA}} = 0.01$

(c) normalized A, Avg. ℓ_2 error vs. β

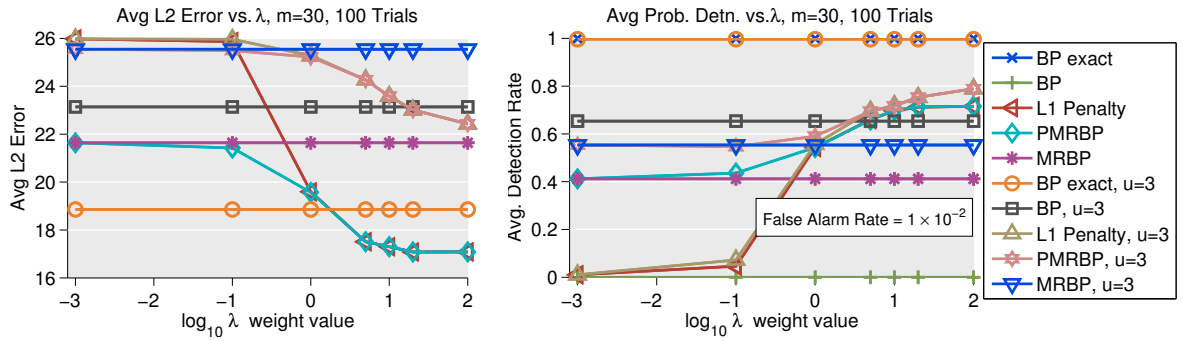
(d) normalized A, Avg. P_{D} vs. β for $P_{\mathrm{FA}} = 0.01$

Figure 6.3: β Study: Recovery error and detection rates of FICCS for two source lattice sampling rates, $u = 1, 3$ with $\delta = 2, 2/3$, over 100 trials, 150 randomly-placed sensors, random additive noise of norm $b_{0.001}$; $\lambda = 1$ and $m = 30$ projections are used. Top panels show shows the average ℓ_2 error and detection rate vs. β for estimators using the un-normalized system matrix A and fixed false alarm rate 0.01. Bottom panels show average ℓ_2 error and detection rates vs. β for estimators using a normalized system matrix A.

Figure 6.3 shows the average estimation error and detection performance for the various estimators, with $\lambda = 1$, $m = 30$ network projections, and both un-normalized and normalized A matrices. For the $u = 1$, BP was infeasible, and so provided no field estimates. Panels (a) and (c) show the average recovery error versus β at the two lattice sample rates and compared with the performance of basis pursuit when sources are actually at grid locations (BP exact). For the $u = 1$ lattice sample rate, the average estimation error of BP exact (not shown in figure) was below 3. We first note that estimation error is reduced for all but the MRBP method when a normalized A was used. In particular the mean squared error of the upsample lattice ($u = 3$) is much reduced by normalizing, even in the BP exact case. Due to large modeling perturbation constants MRBP at both lattice resolutions performed worse that its counterparts, with performance heavily degraded by normalizing the columns of A. The PMRBP estimates, however, did slightly better or the same as ℓ_1 Penalty. In particular, for un-normalized A, and the coarse lattice, PMRBP matches the low error of BP exact for $\beta = 0.25$, and moves up to the error of ℓ_1 Penalty with larger β. Panels (b) and (d) show the average detection rates for a fixed false alarm rate of 0.01. This false alarm rate applies to all the lattice points in the network, so for the 100 dimensional lattice $P_{\text{FA}} = 0.01$ translates to 1 expected false alarm per inversion. To allow for comparison, the detections rates for the finer lattice were converted to coarse detection vectors, where a detection in any of the upsampled lattice bins corresponding to the $u = 1$ lattice counted as a detection. The detection rates in all series except MRBP were again improved by normalizing A. The cases of PMRBP that were feasible matched or slightly improved the detection rate ov ℓ_1 penalty. This shows that the penalty term allows the PMRBP to retain good performance as the cone constraint is expanded large with increasing β, while MRBP suffers as the cone constraint enlarges. Overall, this shows PRMBP to be robust to a wide set β and can trade between the performance of the BP and ℓ_1 Penalty estimators. In contrast, MRBP is very sensitive to the choice of β, and suffers from too quickly expanding residual constraints.

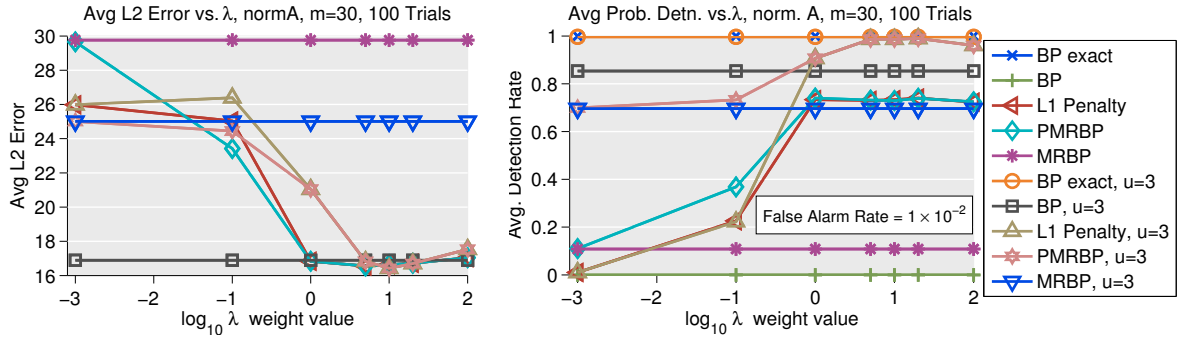
6.2.2 ℓ_1-Norm Penalization Factor λ

We study the impact of choice of the penality parameter λ on the ℓ_1 Penalty and PMRBP estimators. Using the same experimental set-up as in section 6.2.1, but with a β fixed at 0.5, we compare average reconstruction error and detection rates versus λ. Again, detections rates for the finer lattice were con-



(a) Avg. ℓ_2 error vs. λ tradeoff parameter

(b) Avg. P_{D} vs. λ for $P_{\mathrm{FA}} = 0.01$



(c) Nm A, Avg. ℓ_2 error vs. λ tradeoff parameter

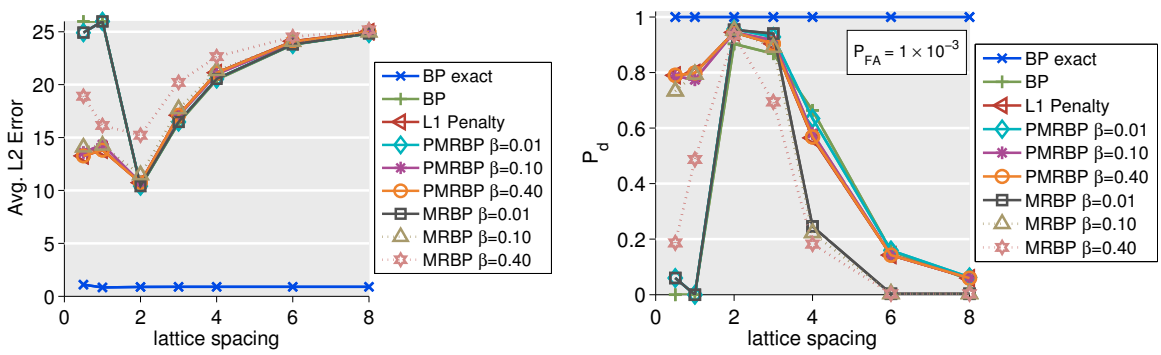
(d) Nm A, Avg. P_{D} vs. λ for $P_{\mathrm{FA}} = 0.01$

Figure 6.4: λ Study : Recovery error and detection rates of FICCS for two source lattice sampling rates, $u = 1, 3$ with $\delta = 2, 2/3$, 100 trials. $N = 150$ randomly-placed sensors, random additive noise of norm $b_{0.001}$; $\beta = 0.5$ and $m = 30$ projections are used. Top panels show shows the average ℓ_2 error and detection rate vs. λ for estimators using the un-normalized system matrix A and fixed false alarm rate 0.01. Bottom panels show average ℓ_2 error and detection rates vs. λ for estimators using a normalized system matrix A.

verted to coarse detection vectors, where a detection in any of the bins corresponding to same location in the coarse lattice counted as a detection.

Figure 6.4 compares not normalizing observation matrix A, in the upper panels, to normalizing A, in the panels, before ℓ_1 optimizations. Traditional basis pursuit in the $u = 1$ lattice does not appear because it was infeasible. In these scenarios, the estimation effect due to source arrivals not on the lattice is considerable. For the $\delta = 2$ spacing, the average estimation error of BP with targets located exactly at grid locations (not shown in figure), is less than 3. The penalty parameter λ is varied from 0.001 to 100, with average error plotted versus the logarithm of the parameter λ.

The error of ℓ_1 Penalty is large for small λ, in both δ cases. When not normalizing A, estimates



(a) Average ℓ_2 Error

(b) Prob. Detection

Figure 6.5: Performance of FICCS, averaged over 100 trials, vs. quantization spacing. In both, $N = 100$ sensors are distributed in a regular grid, spaced by the same δ as the source quantization lattice, using $m = 30$ network projections and additive noise has an exact norm $b_{0.01}$. Columns of A are not normalized, in estimators. Panel (a) shows average ℓ_2 estimation error of the sparse estimators, using $\lambda = 2$, while (b) shows the detection rates for a fixed false alarm rate of 0.001.

PMRBP for the coarse sampling are robust to small λ. For normalized A, PMRBP at the fine lattice performed much better than for the coarse lattice for small λ. In terms of detection rate, panels (b) and (d) show PMRBP to be more robust than ℓ_1 Penalty to too small choices of penalty parameter. As λ increases, average error of both ℓ_1 Penalty and PMRBP with $u = 1$ both drop below that of BP for the fine lattice, suggesting some advantages of performing coarse detection using a model-robust estimator rather than using BP with a finer sample rate. This shows PRMBP displays more robustness to too small choices of penalty parameters, through having also a constraint on fit residual.

6.2.3 Quantization Lattice Spacing δ

The panels of figure 6.5 compare the performance of each estimator versus the spatial quantization parameter δ. In this study, 100 sensors are placed in a regular 10-by-10 grid spanning the region, using $m = 30$ network projections. In this study, the parameter δ refers to both the source quantization lattice as well as the sensor spacing, since the two grids coincide. The propagation law uses $\gamma = 4$, modeling active sensors such as in radar. The additive noise corresponds to a vector with norm $b_{0.01}$, which equals the 95% norm bound on Gaussian noise vectors with variance $\sigma^2 = 0.01$. The estimators do not normalize the columns of the observation system A, before performing optimization. Basis pursuit with no location quantization error (BP Exact) has the lowest error and highest detection rate. As the lattice

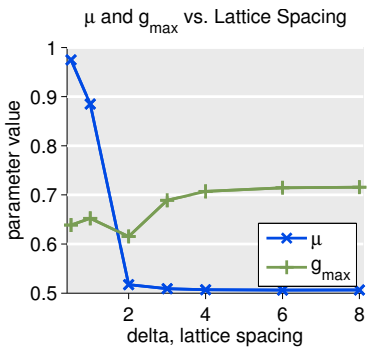
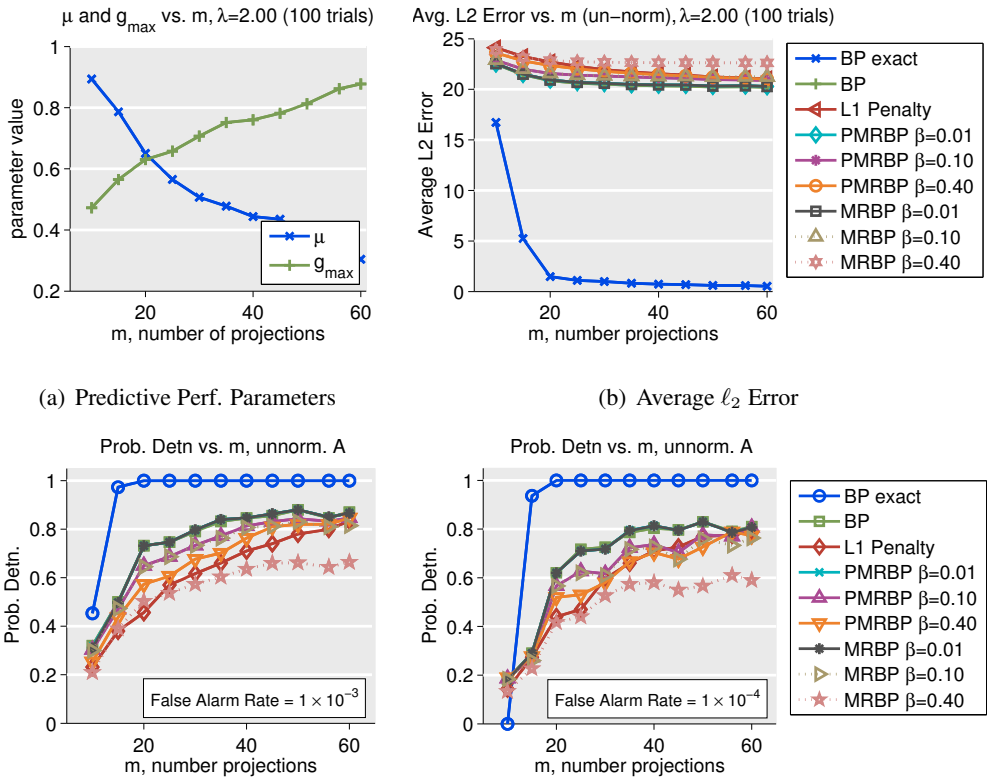


Figure 6.6: μ and $g_{\max}$ versus quantization lattice spacing.

spacing δ increases, the number of source lattice points L decreases, which could reduce reconstruction error simply by reducing possible locations. However, model error for $\gamma = 4$ increases sharply with δ, so we see a sharp optimum spacing for lowest ℓ_2 error at $\delta = 2$. This is not the full story, since, from a detection perspective, sources are localized well using almost all of the estimators for δ between 2 and 4. Since support detection is the most difficult aspect of this problem, and other methods can be used to de-bias the magnitude estimates after localization, these results are promising in showing some robustness of FICCS estimators to quantization level.

Figure 6.6 shows the coherence and model perturbation constants as a function of lattice spacing. We see from this that the region near $\delta = 2$ corresonds to a sweet spot with low coherence but not yet high bounds on model uncertainty.

Selecting the case of $\delta = 4$ from the study of figure 6.5, we now show the performance of the various estimators as a function of varying numbers of network projections, in figure 6.7. This is an example of a challenging case in which PRMBP may have lower error than the unconstrained ℓ_1 Penalty estimator. Model error is high, resulting in poor average ℓ_2 error across the estimators. As the number of observation projections m are varied, there is little change in the average ℓ_2 error, and little difference among estimators, with the exception of BP exact. However, the detection rate of the 3 sources is higher for fixed false alarm rate of 0.001, using the estimators containing the residual cone constraints. This shows the adaptability of the PMRBP method for support detection. In cases where the unconstrained ℓ_1 Penalty performs well, it does similarly. In cases where the cone constraint is needed it can take advantage of this constraint without running as large of a risk for infeasibility as does basis pursuit.



(a) Predictive Perf. Parameters

(b) Average ℓ_2 Error

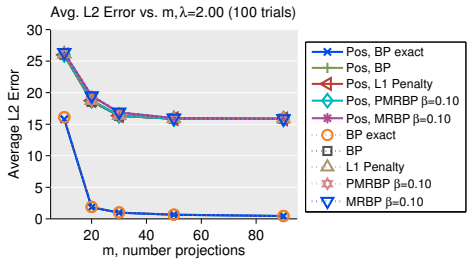
(c) Avg. P_{D} for $P_{\mathrm{FA}} = 0.01$

(d) Avg. P_{D} for $P_{\mathrm{FA}} = 0.001$

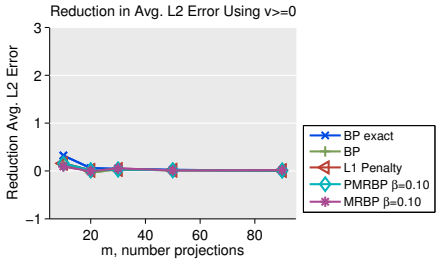
Figure 6.7: Challenging Case: 100 Trial Average FICCS ℓ_2 error and detection rates vs. number of network projections m, for 100 sensors in a 10-by-10 grid, source lattice spacing $\delta = 4$. The propagation law corresponds to a Radar test, with $\gamma = 4$. Additive noise has an exact norm $b_{0.01}$, and penalization parameter is $\lambda = 2$.

6.2.4 Positivity Constraints

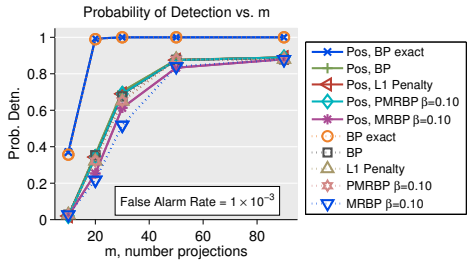
An important consideration to the sparse estimators is whether or not to enforce the positivity constraint on source magnitudes. In general, it helps to impose this constraint when it is known. However, we can show examples where hurts to impose this constraint. The first study of this section is shown in figure 6.8. In this study, the propagation factor is $\gamma = 1$, additive noise has norm $b_{0.01}$, and $\lambda = 2$, and source lattice spacing $\delta = 2.0$. Each plot directly compares the avg. ℓ_2 error or detection rate of the sparse estimators using the positivity constraint versus estimators with unconstrained magnitudes. The left sides show these overlayed, with the unconstrained series in dashed lines, while the right sides show differences in the estimation error or detection rates. Panels (a-d) correspond to 100 sensors placed in a square grid, which (e-h) correspond to 150 randomly-placed sensors.



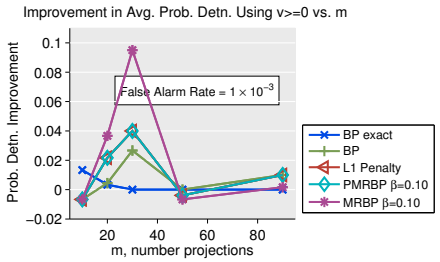
(a) Lattice Sensors, Compare of Avg. ℓ_2 Error



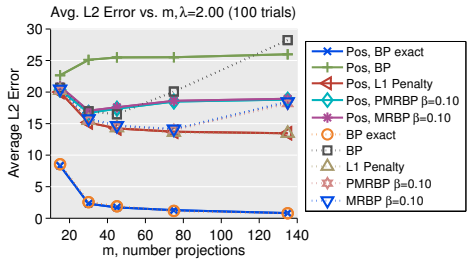
(b) Lattice Sensors, Reduction ℓ_2 Error



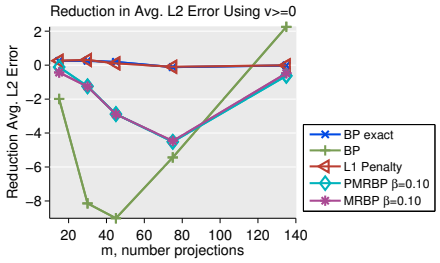
(c) Lattice Sensors, P_{D} for $P_{\mathrm{FA}} = 0.01$



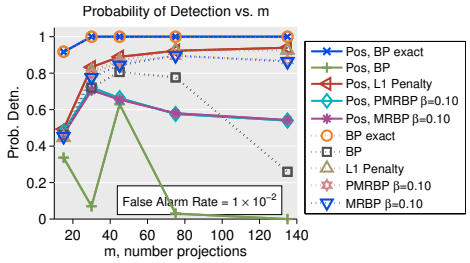
(d) Lattice Sensors, Gain in P_{D}



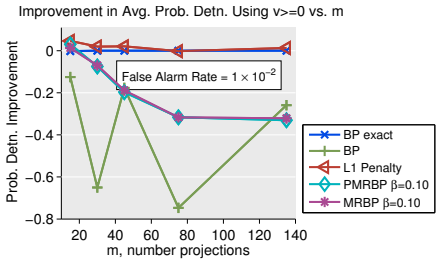
(e) Sensors Random, Compare of Avg. ℓ_2 Error



(f) Sensors Random, Reduction ℓ_2 Error



(g) Sensors Random, P_{D} for $P_{\mathrm{FA}} = 0.01$



(h) Sensors Random, Gain in P_{D}

Figure 6.8: Effects of Positivity Constraints: Propagation factor $\gamma = 1$ (data from study of figures 6.13 and 6.14), source lattice spacing $\delta = 2$, 100 recovery trials, additive noise with norm $b_{0.01}$, and $\lambda = 2$. Left panels comparing avg. ℓ_2 error or detection rate of sparse estimators using the positivity constraint on $\mathbf{v}$ versus none, plot the performance of both with unconstrained in dashed lines. Right panels show direct differences due to requiring $\mathbf{v} \geq 0$. Upper panels (a-d) use a system with 100 sensors in a regular grid, while lower panels (e-h) have 150 randomly placed sensors.

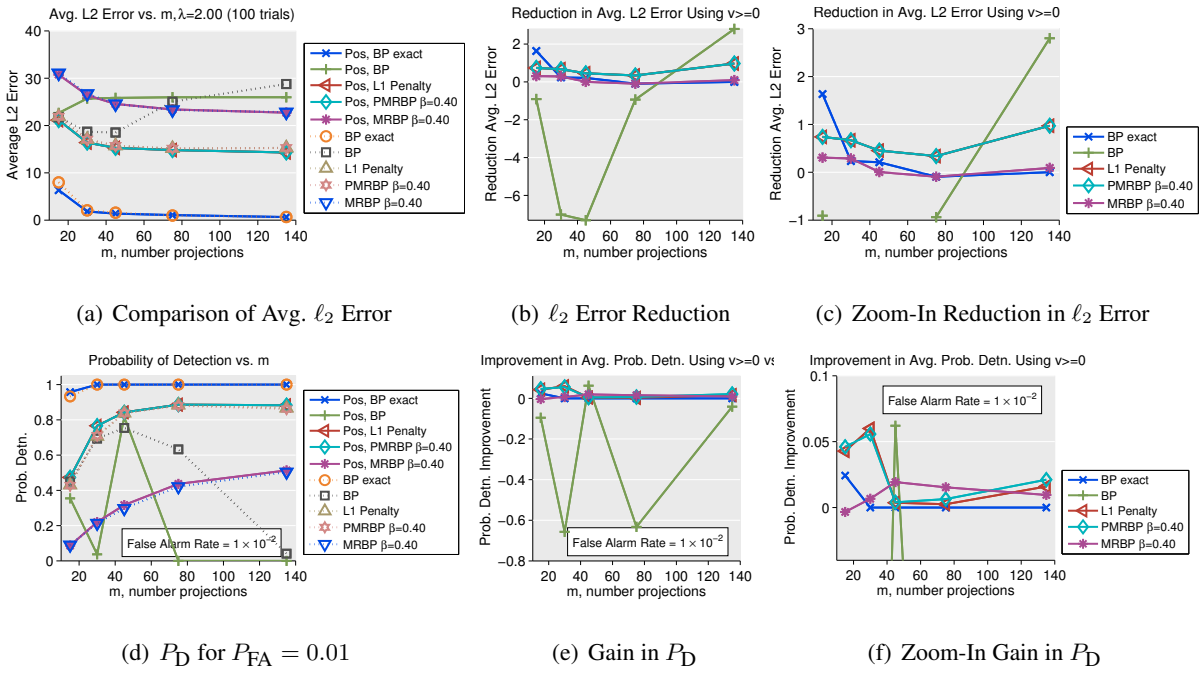
When sensors are in a lattice, the differences due to imposing positivity are small. Mean-squared estimation error is relatively unaffected, and detection rates are generally improved (see panel d). However, for randomly placed sensors, model errors are higher. We see in (e-h) that imposing the constraint $\mathbf{v} \geq 0$ increases mean-squared error and reduces detection performance for the estimators containing residual constraints. In this case, the unexpected results of (e-h) could be related to poor conditioning due to random sensor locations, or to higher quantization errors.

Figure 6.9 also shows the reduction in L2 error, and improvement in probability for using positivity constraints. For this study, $\gamma = 2$, $\delta = 2$, and 150 sensors are placed randomly. Estimators use the normalized A matrix, $\lambda = 2$, and $\beta = 0.40$. Both average ℓ_2 recovery error and detections rates were consistently improved using the positivity constraint for all estimator except traditional basis pursuit. The positivity constraint is generally helpful as long as cone is expanded sufficiently large and model error is moderate.

6.2.5 Normalization of Observation Matrices

When performing the ℓ_1 optimizations, Basis Pursuit and ℓ_1 Penalty, it is often recommended to normalize the columns of the observation matrix, which we denote as A, to have unit norm. An equivalent transformation of the problem, is to use a diagonally weighted ℓ_1 norm of potential source vector estimates, where the weights are norms of the columns of A. Each of the column norms represents the expected received energy due to a single unit-magnitude source, at the corresponding position. Some positions have higher SNR than others due to their being near more sensors, etc.. The weighting of the ℓ_1 norm most likely allows these estimators to hypothesize sources at low SNR positions without having to pay as high of a penalty as with source having high SNR, somewhat equalizing the playing field for detection.

In the previous section, 6.2.4, we already commented on the decision of normalizing the A matrix for ℓ_1 optimizations. We saw that this may help, especially when perturbation constants are low, as fine lattice ($u = 3$) of the β and λ studies, in sections 6.2.1 and 6.2.2. The comparisons of both studies, shown in figure 6.3 and 6.4, showed normalization of the columns of A greatly improved detection rates and the effect was mostly present for all optimizations. However, the reduction of average ℓ_2 estimation



(a) Comparison of Avg. ℓ_2 Error (b) ℓ_2 Error Reduction (c) Zoom-In Reduction in ℓ_2 Error

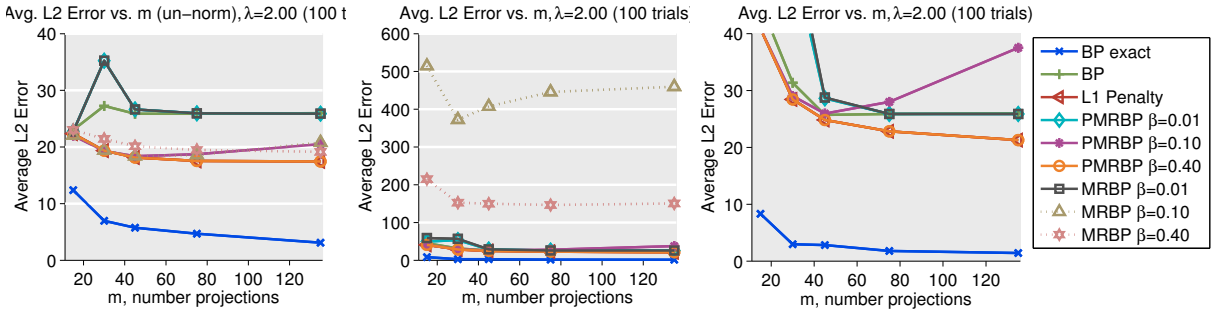
(d) P_{D} for $P_{\mathrm{FA}} = 0.01$ (e) Gain in P_{D} (f) Zoom-In Gain in P_{D}

Figure 6.9: Effects of Positivity Constraints on Estimation Performance in the propagation factor $\gamma = 2$ study. Upper panels show mean ℓ_2 error over 100 trials, while lower panels show detection rates at a fixed false alarm rate. Panels (b) and (c) show the increase of detection rate using positivity constraints over unconstrained optimization. Panels (e) and (f) show the average of reduction of ℓ_2 recovery error of using the positivity constraint. Simulations correspond to those the study of figure 6.16, which uses inverse squared distance propagation loss, 150 randomly placed sensors, source lattice spacing $\delta = 2$, 100 recovery trials, with additive noise with an exact norm b for $\sigma^2 = 0.01$. Sparse estimators use $\lambda = 2$ and $\beta = 0.40$.

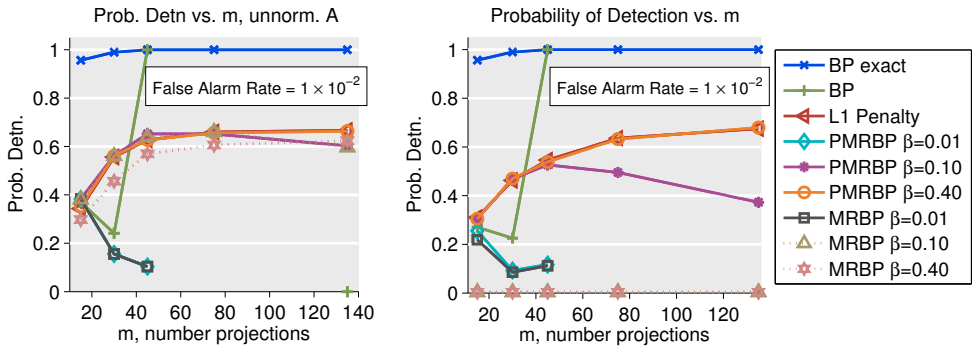
error was smaller and not as consistent.

In this section, we will directly compare mean recovery error and detection rates in challenging scenarios. We will find that although normalizing the observation matrix is generally recommended, in some cases in which quantization model error is quite large and A is poorly conditioned, normalizing A can hurt the performance of estimators that do not sufficiently account for these model errors.

Figure 6.10 shows a comparison of the impact of normalizing the observation matrices, for the $\gamma = 4$ propagation, with 150 randomly placed sensors and positively constrained source vector magnitudes. Recovery error and detection rates are averaged over 100 trials, with additive noise of norm $b_{0.01}$. Sparse estimators used $\lambda = 2$ and $\beta = 0.10$. In this case, normalizing A greatly increased the average estimation error, which is due to the large increase in perturbation constants after rescaling. See figure 6.15 for the normalized perturbation constants. For detection performance, except for PMRBP with $\beta = 0.4$, all of



(a) Avg. ℓ_2 Error with un-norm. A (b) Avg. ℓ_2 error with norm. A (c) Zoom-In Avg. ℓ_2 error with norm. A



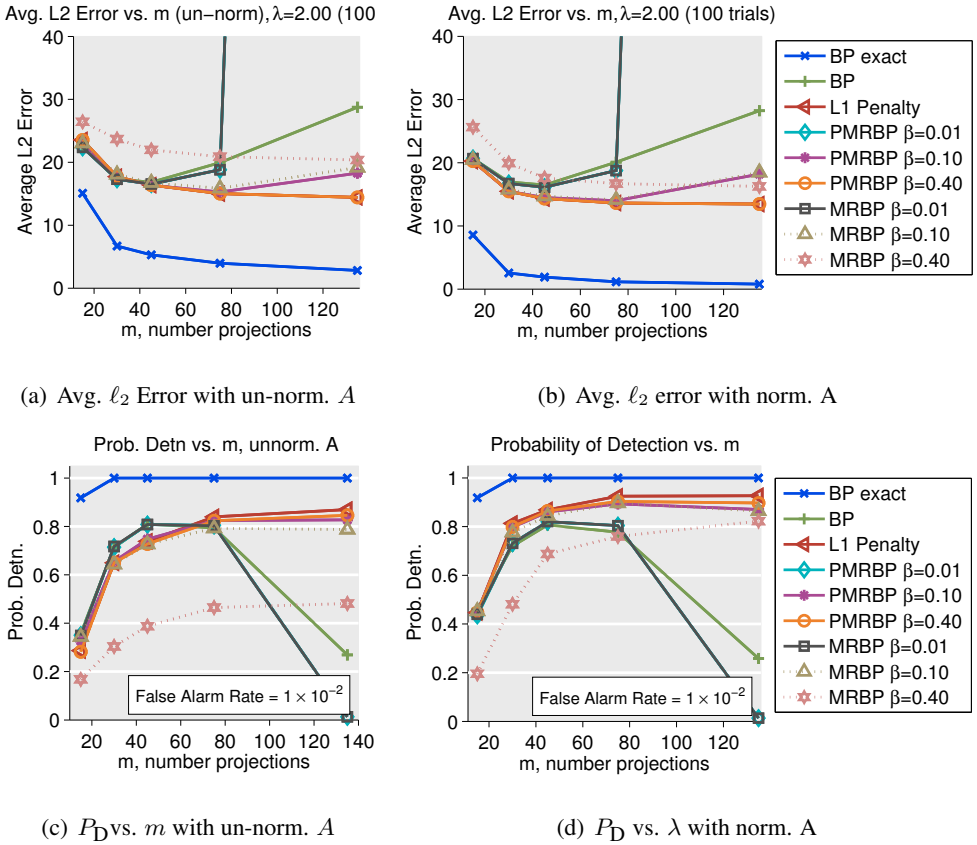
(d) P_{D}vs. m with un-norm. A (e) P_{D} vs. m with norm. A

Figure 6.10: Un-normalized vs. Normalized A in for $\gamma = 4$, 150 randomly-placed sensors, (study also in figure 6.18). Lattice spacing $\delta = 2.0$, 100 recovery trials, with random additive noise with norm $b_{0.01}$. Sparse estimators use $\lambda = 2$ and $\beta = 0.10$, and contrained $\mathbf{v} \geq 0$. Upper panels show average ℓ_2 estimation error and lower panels show detection rates, as a function of the number of network projections m.

the cone constrained estimators either remained infeasible, or fell to 0 detection rate after normalizing A. The performance of ℓ_1 Penalty and PMRBP with $\beta = 0.4$ remained the same. In the $\gamma = 4$ case, where modeling error can be large (and SNR low) renormalization of the observation matrix is better to be avoided.

Figure 6.11 shows a comparison of ℓ_2 error and detection rates for the $\gamma = 1$ study (also shown in figure 6.14). Again sensor are placed randomly, lattice spacing δ is 2. 100 recovery trials are performed with additive noise norm $b_{0.01}$. Sparse estimators use $\lambda = 2$ and $\beta = 0.10$, but this time did not constrain the sign of $\mathbf{v}$. Examining the top panels we see the effect of normalizing to be very minimal on average ℓ_2 recovery error. However, check the detection rates, we see that, for the same false alarm rates of 0.01, detection rates for all but BP and MRBP have improved due to normalizing A. From figure 6.15, we see

that in this case the perturbation constants are much more moderate, suggesting that normalization of the observation matrix is beneficial in such cases.



(a) Avg. ℓ_2 Error with un-norm. A

(b) Avg. ℓ_2 error with norm. A

(c) P_{D}vs. m with un-norm. A

(d) P_{D} vs. λ with norm. A

Figure 6.11: Un-normalized vs. Normalized A for $\gamma = 1$ Study, (from figure 6.14): Recovery error and detection Performance, as a function of the number of network projections m. Lower panels show detection rates at a fixed false alarm rate, upper panels show average ℓ_2 error. 150 Sensors are randomly placed, 100 simulations trials use lattice spacing $\delta = 2.0$, additive noise with norm $b_{0.01}$, $\lambda = 2$, and $\beta = 0.10$, and no constraints on the signs of $\mathbf{v}$.

We will compared ℓ_2 recovery error and detection rates when using un-normalized versus normalized observations matrices in ℓ_1 optimizations. We found that normalized A generally improves detection rates. A few cases in which the system may be poorly conditioned, especially when sensors are placed randomly and $\gamma = 4$ is used., normalizing A can hurt the performance of MRBP and PMRBP estimators that do use large enough cone expansion factors, β.

6.3 Studies of System Tradeoffs

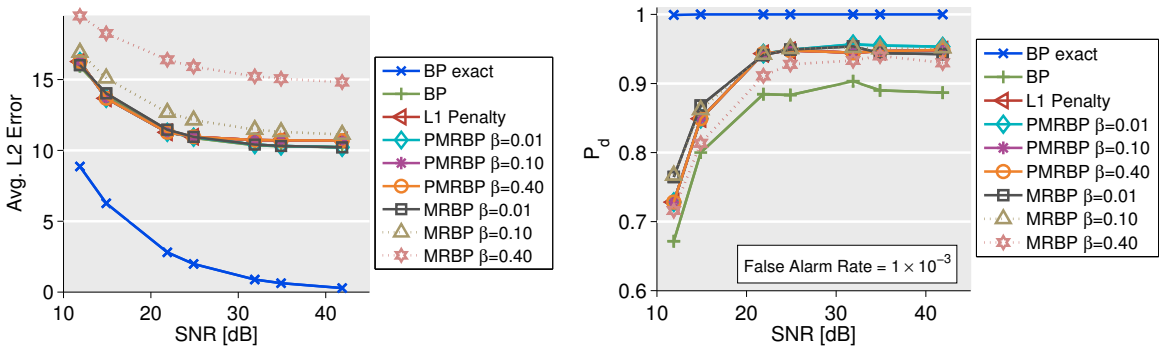
The simulation experiments of section 6.2 studied the effects of a variety of estimation parameters on the performance of field inversion. However, we saw that many system parameters of these tests played a critical role. For example, the most challanging cases generally involved random placements of sensors and $\gamma = 4$ propagation losses.

In this section, we focus on system related parameters and their impact on field inversion performance. We show the sensitivity of sparse estimation to the amount of measurement noise in 6.3.1. Then, in 6.3.2, we discuss the effects of having randomly located sensors versus having sensors on a regular lattice. Embedded in many of our discussions is the impact of the propagation factor γ. Section 6.3.3, presents a full comparison of the inversion performance for each propagation factor for fixed choices of estimation parameters.

6.3.1 Signal-to-noise Ratio

In this section, we show a study over a range measurement noise levels, presenting both average error and detection performance as a function of signal-to-noise ratio (SNR). This study uses 100 regularly spaced sensors. The propagation law uses $\gamma = 4$, modeling active sensors, such as in radar. The source quantization spacing was $\delta = 2$, $m = 30$ random orthogonal network projections are computed, and random additive noise with norm equal to b_{σ^2}. Recall that b_{σ^2} is the norm bound for 95% Gaussian vectors with variance σ^2. Random source vectors have three continuously located sources with magnitudes 15. The choices of variances are adjusted resulting in a change in the ratio of signal power of the true source vector to the total noise power of sensor readings.

The panels of figure 6.12 compare performance of the FICCS estimators as a function of the resulting SNR. Average ℓ_2 error and detection performance are shown for each sparse estimator, and 3 choices of cone expansion weights β. Basis pursuit with performed on source vectors with no location quantization error (BP Exact) has the lowest error. We see that most estimators using off-grid observations perform similarly in terms of the ℓ_2 error metric. However, when we compare detection rates for a fixed false alarm rate of 0.001, the robust estimators as well as ℓ_1 Penalty outperform BP in detecting the source locations.



(a) Average ℓ_2 Error

(b) Prob. Detection

Figure 6.12: SNR Study: Performance of FICCS averaged over 100 trials, with $N = 100$ regularly spaced sensors with $\delta = 2$, using $m = 30$ projections, additive noise has an exact norm b, and propagation constant $\gamma = 4$. The constant b is adjusted to provide the various SNR levels. Columns of A are not normalized in the estimators. Panel (a) shows average ℓ_2 estimation error of the sparse estimators, using $\lambda = 2$, while (b) shows the detection rates for a fixed false alarm rate of 0.001.

While the off-grid sources resulted in the largest impact on estimation error, detection performance falls sharply for SNR's that are too low. The increase in ℓ_2 error with falling SNR is more gradual, and makes sense due to linear dependence of bounds on recovery error to the noise level. Overall, however these sparse estimators require scenarios with high SNR to effectively recovery the support of an unknown source vector as we see in panel (b).

6.3.2 Random versus Regular Sensor Placement

A number of the tests in section 6.2 used regular and random sensor placements. When placing sensors continuously at random in the monitored area, we enforce a distance bound of $\delta/6$ (where δ is given by the source quantization lattice), so that any sensors places closer than this bound to an existing sensor will be thrown out. In each case, if the sensors were placed randomly, and compared to a regular grid of sensors, 50% more sensors would be used. This is due to the lower efficiency of randomly placed sensors at observing an area in which sources arrive uniformly at random. In a random placement we run high risks of some areas not being adequately monitored. We found that using 50% more of them helped in mitigating this effect.

In studying the effects of positivity constraints on the estimators, we compared lattice versus randomly placed sensors in figure 6.8. The scenario used propagation factor $\gamma = 1$, source lattice spacing

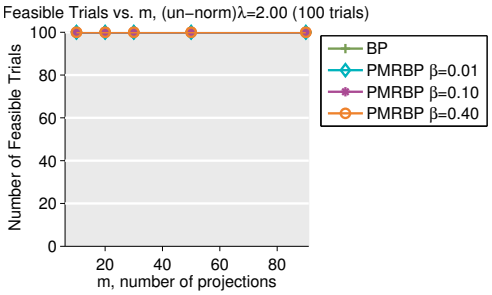
$\delta = 2$, additive noise with norm $b_{0.01}$, and penalty parameter $\lambda = 2$. We found that using positivity constraints improved the performance of most estimators except in the case of random sensor placement with cone-constrained optimizations. In these cases, we suspected that poor conditioning of the observation system increased their sensitivity to the additional constraint. We show, here, the full set of ℓ_2 error and detection performance for this case of $\gamma = 1$, using both lattice and random sensors placement, and for positively constrained and unconstrained $\mathbf{v}$ estimates in figures 6.13 and 6.14. Estimators use $\lambda = 2$, $\beta = 0.10$, and do not normalize columns of A before computation.

From figures 6.13 and 6.14, we notice that estimators are always feasible under lattice placement, whether or not a positivity constraint is used. The effect on feasibility is strongest with the positivity constraint, however. Panel (b) of 6.13 shows that, with the constraint and random placement, only about 60% of trials model-robust trials were feasible for the larger cone factor of 0.40. This helps to explain the increase in recovery error and falls in detection performance of (d) and (f). Even for the ℓ_1 penalty estimator, with no feasibility issues, the detection rates is lower when using the random placement set-up.

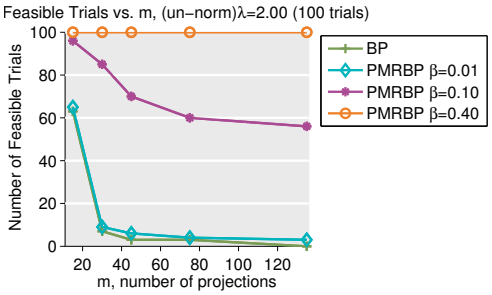
In figure for estimation performance the study of figure 6.14, which does not use positivity constraints, shows more severe jump in ℓ_2 estimation error, most likely due to the cone constraint estimators being near to infeasible. However, detection performance is more robust since more of the trials were feasible.

Overall, this study shows that even spacing is a more efficient use of sensors. The system can use fewer of them, and have less sensitivity to quantization and conditioning issues. By examining 6.15, in the next section, we can see that the coherence of the randomly placed sensor system is much higher than in the regular placement. This indicates poor conditioning of the network observations when not using a regular grid of sensors.

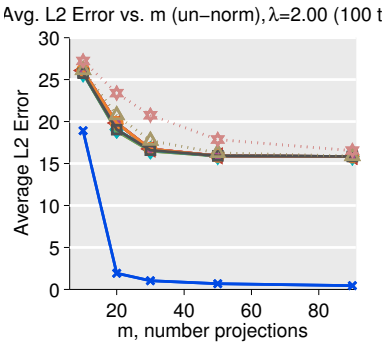
The next section, presents the detailed estimation and detection results, with random and lattice sensor placements, for the other propagation factors, $\gamma = 2$ and 4. These studies also show a strong preference for regularly placed sensors.



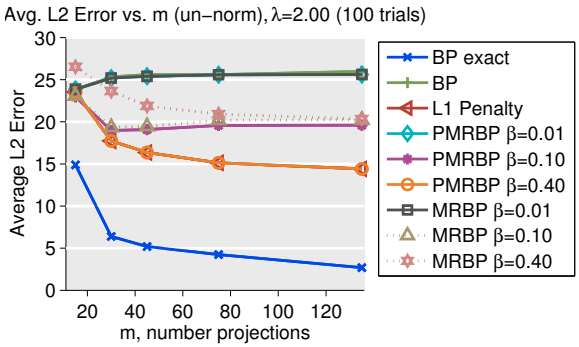
(a) Sensors on Lattice, Num. Feasible



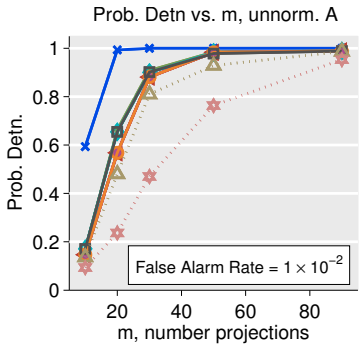
(b) Sensors Random, Num. Feasible



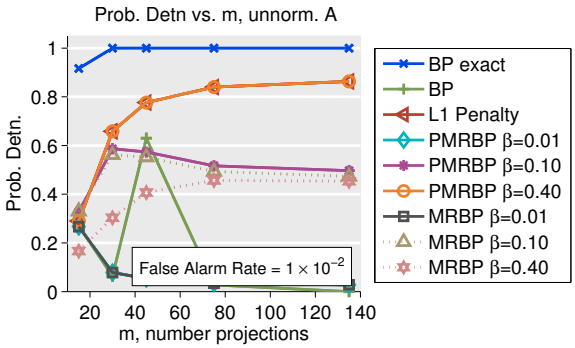
(c) Sensors on Lattice, Avg. ℓ_2 Error



(d) Sensors Random, Avg. ℓ_2 Error

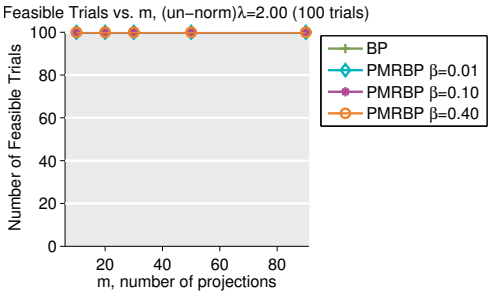


(e) Sensors on Lattice, P_{D}

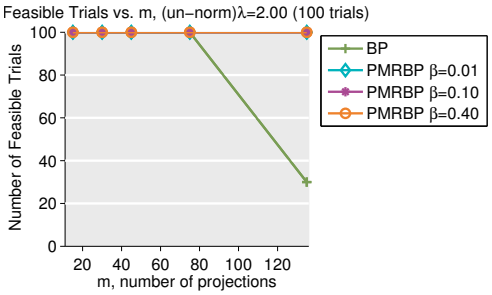


(f) Sensors Random, P_{D}

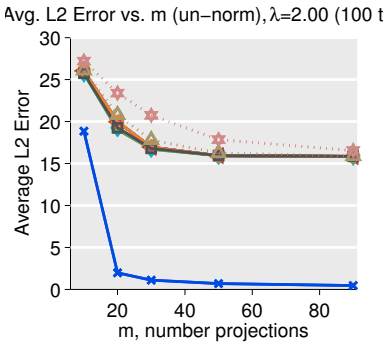
Figure 6.13: Study with $\gamma = 1$, Lattice vs. Random Sensor Placement, Estimators using Positivity constraints. Left panels show detection rates at a fixed false alarm rate, while right panels show the difference of the detection rate using positivity constraints minus the rates when the signs of $\mathbf{v}$ are unconstrained. The upper panels use a system with 100 sensors in a regular grid, while lower panels have 150 randomly placed sensors. Simulations correspond to source lattice spacing $\delta = 2$, 100 recovery trials, with additive noise of norm $b_{0.01}$. Estimators use $\lambda = 2$ and $\beta = 0.10$.



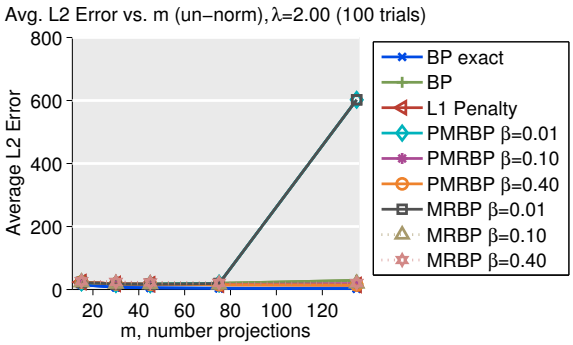
(a) Sensors on Lattice, Num. Feasible



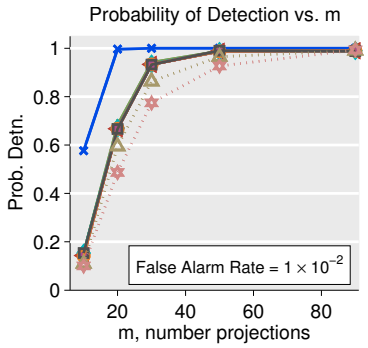
(b) Sensors Random, Num. Feasible



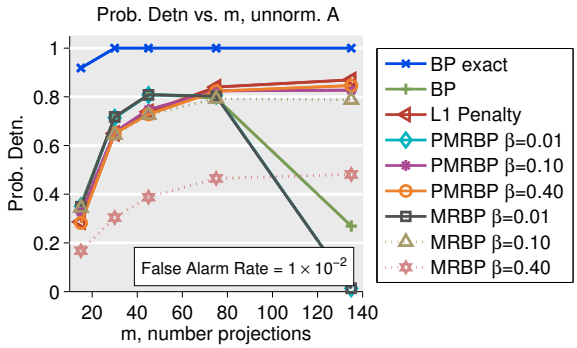
(c) Sensors on Lattice, Avg. ℓ_2 Error



(d) Sensors Random, Avg. ℓ_2 Error



(e) Sensors on Lattice, P_{D}



(f) Sensors Random, P_{D}

Figure 6.14: Study with $\gamma = 1$, Lattice vs. Random Sensor Placement, Estimators not using positivity constraints. Left panels show detection rates at a fixed false alarm rate, while right panels show the difference of the detection rate using positivity constraints minus the rates when the signs of $\mathbf{v}$ are unconstrained. The upper panels use a system with 100 sensors in a regular grid, while lower panels have 150 randomly placed sensors. Simulations correspond to source lattice spacing $\delta = 2$, 100 recovery trials, with additive noise of norm $b_{0.01}$. Estimators use $\lambda = 2$ and $\beta = 0.10$.

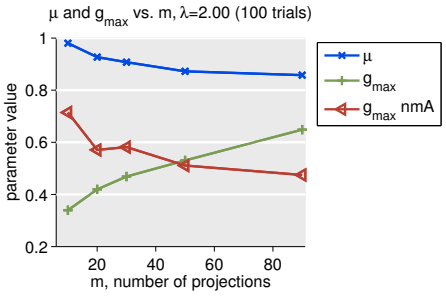
6.3.3 Propagation Power Law

Throughout our numerical testing, we used three main exponents for inverse distance propagation loss functions. The factors $\gamma = 4$, 2, and 1, correspond to radar, acoustic source, and guided wave applications, respectively. We continue the experiments of section 6.3.2 for the additional propagation factors of $\gamma = 2$ and 4, comparing also the effects of lattice versus random sensor placement and use of positivity constrained on estimated magnitudes.

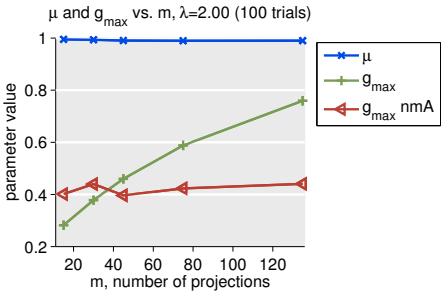
First, figure 6.15 shows the systems performance parameters of coherence μ and maximum model error $g_{\max}$ for each of these propagation factors. First we notice that in all of the random sensor placement experiments, the coherence is much higher than in the regular placement experiments, helping to explain the inversion performance degredation we saw in the previous section. In addition, we see that coherence is lower for higher γ, while model perturbation constants are highest for the larger γ. The factor of $\gamma = 2$ may represent a good tradeoff in these two competing issues.

Figures 6.16, 6.17, 6.18, and 6.19 show the full set of ℓ_2 errror and detection performance for $\gamma = 2, 4$, for both lattice and random sensors placement, and for positively constrained and unconstrained $\mathbf{v}$ estimates. When $\gamma = 2$, we see similar results to the $\gamma = 1$ case, except that the ℓ_2 estimation error is more stable in the case where positivity constraints are not imposed, see panel (d) of figure6.17. In addition the detection performance of the best estimators, ℓ_1 Penalty and PMRBP, is better for $\gamma = 2$ than it was in the $\gamma = 1$ cases, for both lattice and nonlattice sensors.

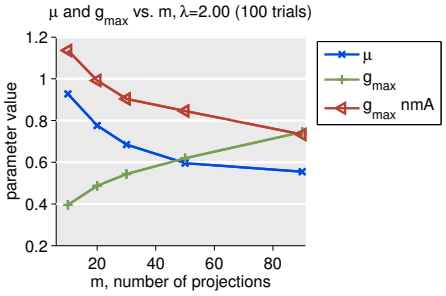
When $\gamma = 4$, feasibility becomes a major issue because of the very high model error in this regime. When using positivity constraints, there are a few basis pursuit trials with lattice placement of sensors that are infeasible, whereas for the other power laws, feasibility never was an issue when using sensors on the grid. In addition, we see a similar phenomenon to the results of the $\gamma = 1$ study with no positivity constraints. When the number of measurement projections is large the system becomes infeasible, and the basis pursuit estimates that are feasible have enormous mean-squared error. However, from the detection standpoint, the performance suffers a moderate amount, with the unconstrained ℓ_1 Penalty estimator performing the best. The radar propagation cases show that, in this regime, the effects of sensor placement are the strongest, with regular placement being more critical to inversion performance. Indeed both figures 6.18 and 6.19 display excellent estimation performance for regularly spaced sensors.



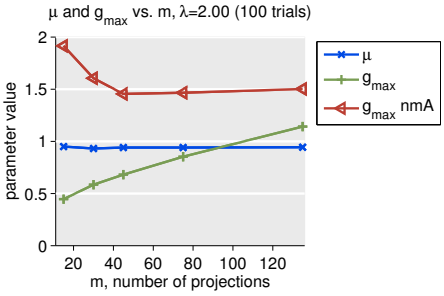
(a) Sensors on Lattice, μ and $g_{\max}$, $\gamma = 1$



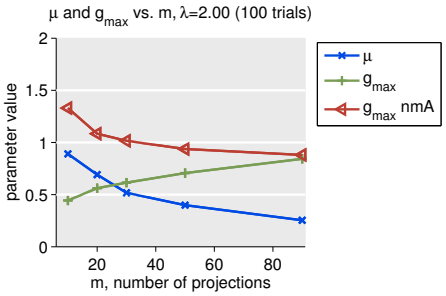
(b) Sensors Random, μ and $g_{\max}$, $\gamma = 1$



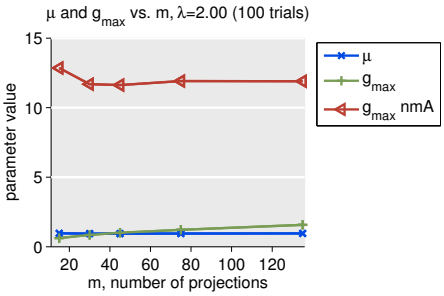
(c) Sensors on Lattice, μ and $g_{\max}$, $\gamma = 2$



(d) Sensors Random, μ and $g_{\max}$, $\gamma = 2$

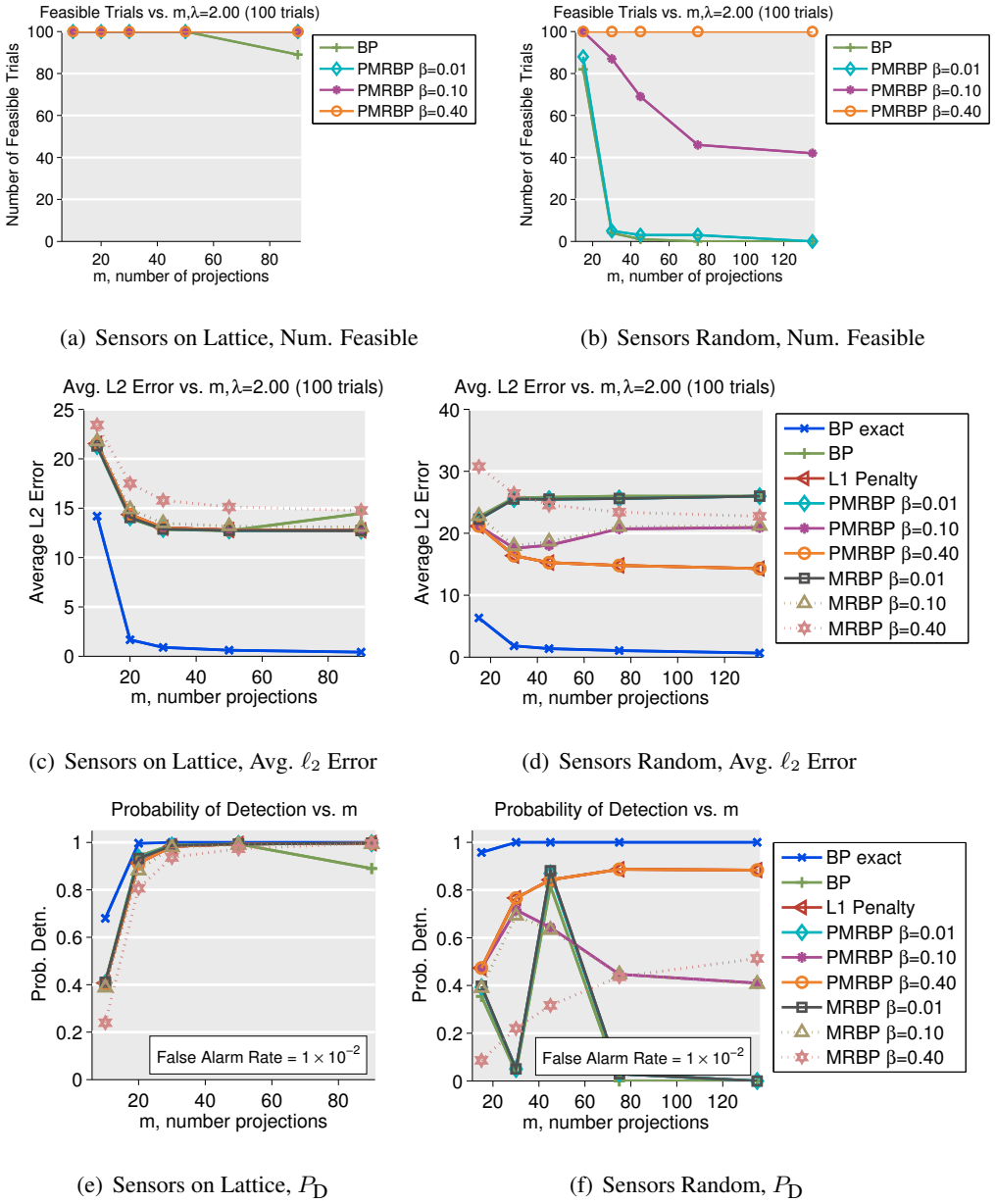


(e) Sensors on Lattice, μ and $g_{\max}$, $\gamma = 4$



(f) Sensors Random, μ and $g_{\max}$, $\gamma = 4$

Figure 6.15: Coherence μ and perturbation constants $g_{\max}$ for each power law propagation exponent $\gamma = 2, 1, 4$, using lattice spacing $\delta = 2$. Left panels show these for a system of 100 sensors in a regular grid. Right panels show these for 150 randomly placed sensors.



(a) Sensors on Lattice, Num. Feasible

(b) Sensors Random, Num. Feasible

(c) Sensors on Lattice, Avg. ℓ_2 Error

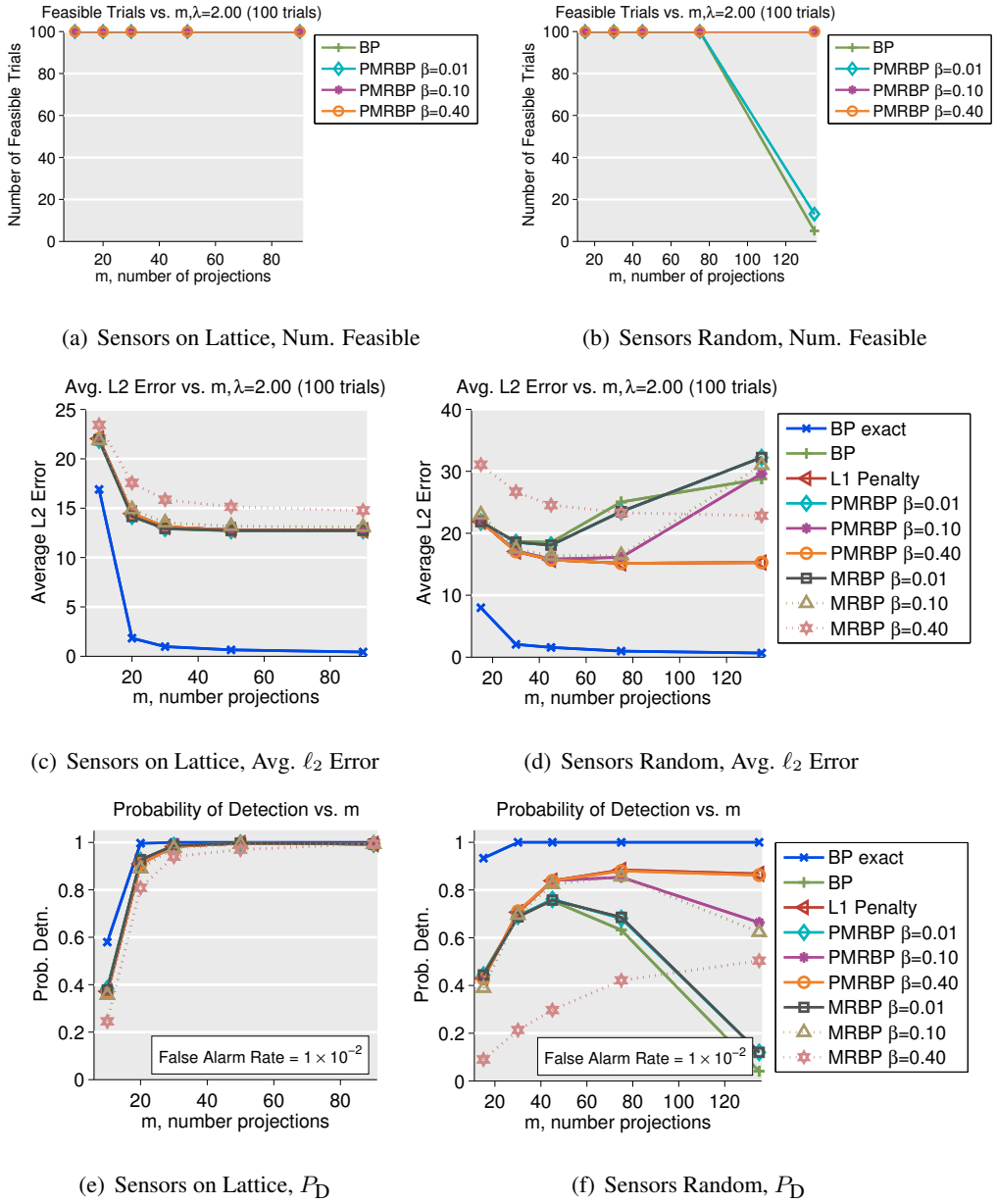
(d) Sensors Random, Avg. ℓ_2 Error

(e) Sensors on Lattice, P_{D}

(f) Sensors Random, P_{D}

Figure 6.16: Study with $\gamma = 2$, Lattice vs. Random Sensor Placement, Estimators Using positivity constraints. Left panels show detection rates at a fixed false alarm rate, while right panels show the difference of the detection rate using positivity constraints minus the rates when the signs of $\mathbf{v}$ are unconstrained. The upper panels use a system with 100 sensors in a regular grid, while lower panels have 150 randomly placed sensors. Simulations correspond to source lattice spacing $\delta = 2$, 100 recovery trials, with additive noise of norm $b_{0.01}$. Estimators use $\lambda = 2$ and $\beta = 0.10$, and normalized A.

Overall, we find that in the various tests performed, the particular propagation law at work does not have as large of an impact as we expected. If using random sensor placements, then the $\gamma = 2$ regime is preferred. However, when using a lattice for sensors performance using each law is roughly equivalent,



(a) Sensors on Lattice, Num. Feasible

(b) Sensors Random, Num. Feasible

(c) Sensors on Lattice, Avg. ℓ_2 Error

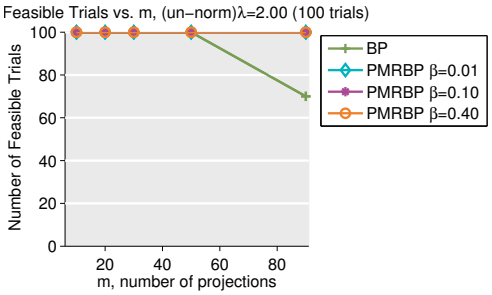
(d) Sensors Random, Avg. ℓ_2 Error

(e) Sensors on Lattice, P_{D}

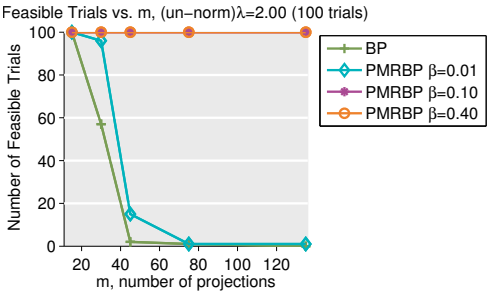
(f) Sensors Random, P_{D}

Figure 6.17: Study with $\gamma = 2$, Lattice vs. Random Sensor Placement, Estimators not using positivity constraints. Left panels show detection rates at a fixed false alarm rate, while right panels show the difference of the detection rate using positivity constraints minus the rates when the signs of $\mathbf{v}$ are unconstrained. The upper panels use a system with 100 sensors in a regular grid, while lower panels have 150 randomly placed sensors. Simulations correspond to source lattice spacing $\delta = 2$, 100 recovery trials, with additive noise of norm $b_{0.01}$. Estimators use $\lambda = 2$ and $\beta = 0.10$, and normalized A.

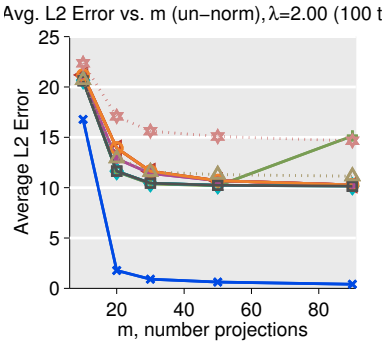
with $\gamma \geq 2$ outperforming $\gamma = 1$ by a little. We trace this effect back to the study of coherence and model error parameters in chapter 4. It seems for the higher propagation exponent, although model error increases, coherence decreases, and for small exponent, the model error is low even though system



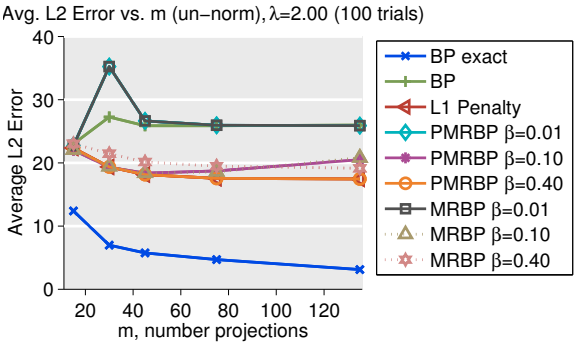
(a) Sensors on Lattice, Num. Feasible



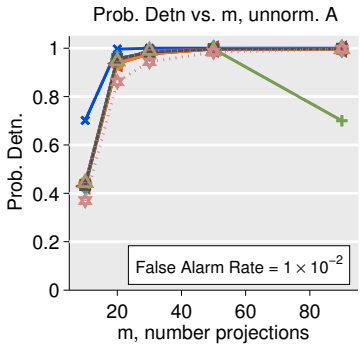
(b) Sensors Random, Num. Feasible



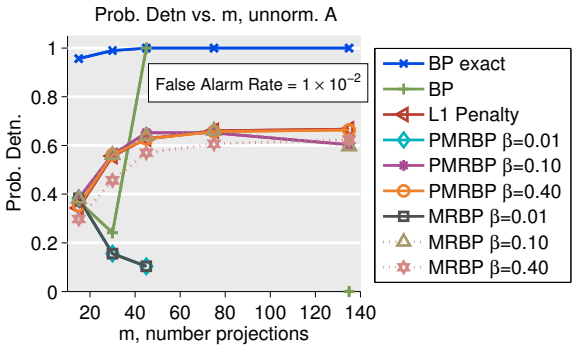
(c) Sensors on Lattice, Avg. ℓ_2 Error



(d) Sensors Random, Avg. ℓ_2 Error



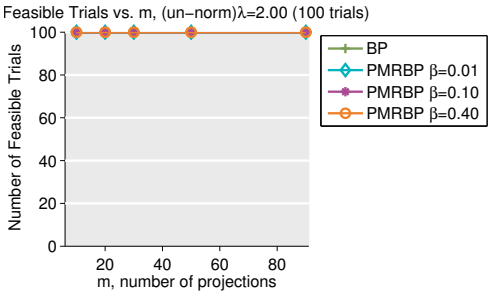
(e) Sensors on Lattice, P_{D}



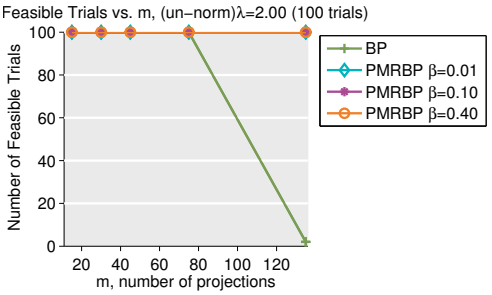
(f) Sensors Random, P_{D}

Figure 6.18: Study with $\gamma = 4$, Lattice vs. Random Sensor Placement, Estimators Using positivity constraints. Left panels show detection rates at a fixed false alarm rate, while right panels show the difference of the detection rate using positivity constraints minus the rates when the signs of $\mathbf{v}$ are unconstrained. The upper panels use a system with 100 sensors in a regular grid, while lower panels have 150 randomly placed sensors. Simulations correspond to source lattice spacing $\delta = 2$, 100 recovery trials, with additive noise of norm $b_{0.01}$. Estimators use $\lambda = 2$ and $\beta = 0.10$, and un-normalized A.

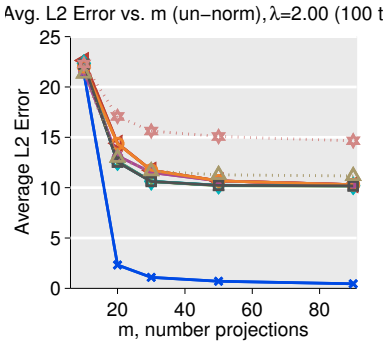
coherence is higher.



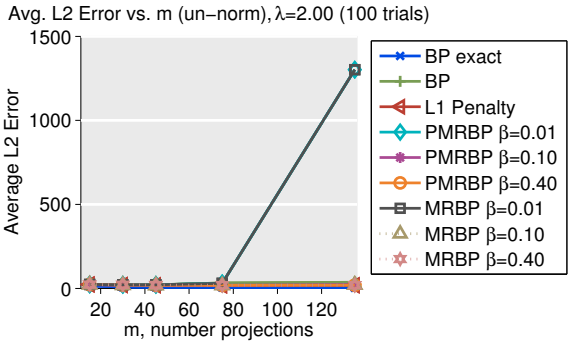
(a) Sensors on Lattice, Num. Feasible



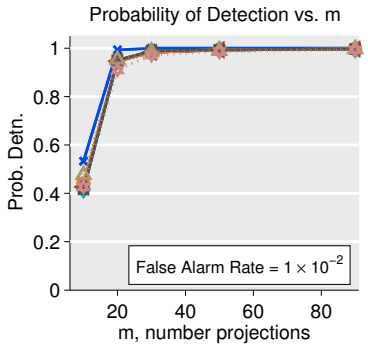
(b) Sensors Random, Num. Feasible



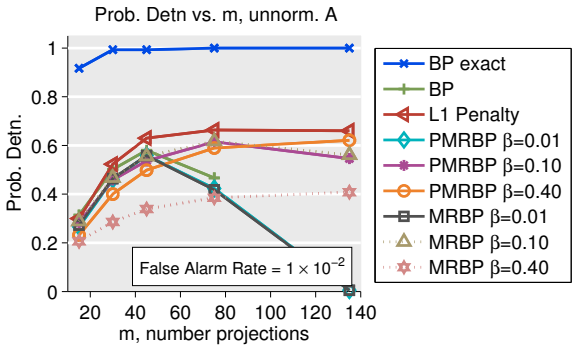
(c) Sensors on Lattice, Avg. ℓ_2 Error



(d) Sensors Random, Avg. ℓ_2 Error



(e) Sensors on Lattice, P_{D}



(f) Sensors Random, P_{D}

Figure 6.19: Study with $\gamma = 4$, Lattice vs. Random Sensor Placement, Estimators not using positivity constraints. Left panels show detection rates at a fixed false alarm rate, while right panels show the difference of the detection rate using positivity constraints minus the rates when the signs of $\mathbf{v}$ are unconstrained. The upper panels use a system with 100 sensors in a regular grid, while lower panels have 150 randomly placed sensors. Simulations correspond to source lattice spacing $\delta = 2$, 100 recovery trials, with additive noise of norm $b_{0.01}$. Estimators use $\lambda = 2$ and $\beta = 0.10$, and un-normalized A.

6.4 Conclusions

We have studied, through numerical experiments, the considerations in parameter choice and system design as they affect the performance of the set of compressed sensing estimators. We found that the

proposed expansion of residual cone constraints, used in the model-robust estimators, successfully addressed feasibility issues, and when combined with an ℓ_1 penalty term in the objective is quite robust to discretization factor, as well as choices of the tuning parameters β and λ. We found that although the traditional ℓ_1 penalty often had the best performance, the PMRBP estimator was able to either match its performance or that of traditional basis pursuit in many challenging off-grid recovery scenarios. We also identified the placements of sensors in the distributed network to have strong implications on the quality of field inversions. Overall, this distributed estimation algorithm proved adaptable to a wide set of propagation regimes and system designs.

Chapter 7

Time-Varying Field Inversion Using Dynamically Re-Weighted ℓ_1 Penalized Least Squares

We extend the FICCS approach, presented in chapter 3 for the static configuration, to a dynamic environment. Like before, we treat the problem of field reconstruction where a field, generated by point-sources with known propagation, is observed by a distributed network of sensors. The sensors are constrained to local communications, leading to a fully distributed architecture that employs consensus message passing to transmit low dimensional projections of the global field. The approach, called field inversion for consensus and compressed sensing (FICCS), relies on sparse optimization methods, e.g. basis pursuit and ℓ_1 penalized least-squares, [24, 11], to recover multiple target locations as well as their unknown magnitudes. We simulate targets within the surveilled area via a random appearance and disappearance model. The physical field generated by the targets is the superposition of a small number of spatially-distributed sources, with known inverse power-law propagation. The stationary sources are assumed to appear and disappear in the monitored area at random, and at a relatively slow rate compared to the rate of communication. This work shows the advantages of the FICCS technique for the tracking of time-varying support using the Adaptive LASSO sparse estimator of Zou, [56]. The re-weighted sparse ℓ_1 optimization provides a robust for applying previous support knowledge to a time-varying source

tracking application, while allowing for the discovery of new sources with a minimum of computed network observations. We compare the performance of adaptively weighted sparse minimization to an unweighted version for tracking the locations and magnitudes of the slowly changing target vectors in both the mean-squared error and detection senses, showing the ease and advantages of the adaptively reweighted sparse estimators.

Section 7.1 formulates the dynamics of the field.

7.1 Dynamic Target Support Model

We describe the discrete time-varying system with a linear additive model. At each large time-step, the support of the source vector may have changed by at most one. There are two time-scales at which the system operates, shown in fig. 7.1. Targets may appear and disappear at the long time-scale, while the network communication occurs in the short time-scale, allowing the sensors sufficient time to spread new global field information.

At each time-step, a new source may appear in the vector $\mathbf{v}$, with probability p_+, an existing source may disappear with probability p_-, or the support of nonzeros in $\mathbf{v}$ may stay the same, with probability $p_= = 1 - p_+ - p_-$. As an example, we may choose $p_+ = p_- = 0.2$, resulting in a $p_=$ of 0.6.

At each time-step we can model the time-varying system for source vector and sensor readings as

$$\begin{cases} \mathbf{v}_{t+1} &= \mathbf{v}_t + \mathbf{d}_t \\ \mathbf{y}_t &= \Phi\left(\Psi\mathbf{v}_t + \mathbf{n}_t\right) \;=\; A\mathbf{v}_t + \mathbf{n}_t \end{cases} \tag{7.1}$$

where $\mathbf{n}_t$ is an independent additive noise vector, and the difference vector, $\mathbf{d}_t$, is constrained to have $\|\mathbf{d}_t\|_0 \leq 1$. The entries of $\mathbf{d}_t$ either contain a source magnitude at a new location, the negative magnitude of a source at an existing location representing the disappearance of that source, or no entries, corresponding to probabilities p_+, p_-, and $p_=$.

For simplicity, in (7.1) we will collect the same network projections at each time-step. While making the new projections adaptive can be more efficient, the protocols for collecting the same projections are simpler, and we may assume a constant observation matrix $A = \Phi\Psi$. In the next section, we describe our approach to leveraging an estimate from the previous time-step, $\hat{\mathbf{v}}_t|\mathbf{y}_{0:t}$ in computing the next estimate $\hat{\mathbf{v}}_{t+1}|\mathbf{y}_{0:t+1}$.

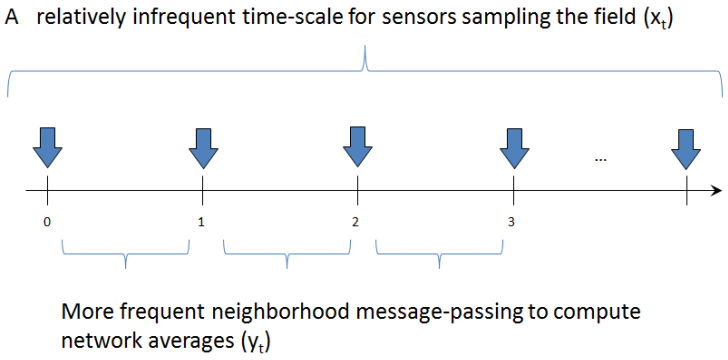


Figure 7.1: Diagram of the slow observation time-scale, in which a single source may disappear or appear, versus the fast network communication time-scale, in which the m distributed averages are computed using distributed consensus

7.2 Dynamically Re-weighted ℓ_1 Penalized Least Squares

In the Adaptive Lasso, [56], it is suggested that given an estimate $\mathbf{v}_t$, we may solve an adaptively re-weighted ℓ_1-penalized least squares with weights given by

$$w_t(i) = \frac{1}{|v_t(i)|}, \quad \mathbf{w}_t = \begin{bmatrix} w_1 & w_2 & \cdots & w_N \end{bmatrix} \tag{7.2}$$

Then the next estimate $\hat{\mathbf{v}}_{t+1}$ is the minimizer of the Weighted ℓ_1 penalized Least Squares (wtdL1LS), below.

$$\begin{array}{lll} \text{(wtd L1LS)} & \text{minimize} & \|\text{diag}(\mathbf{w}_t)\mathbf{v}\|_1 + \frac{\lambda}{2}\|\mathbf{y}_{t+1} - A\mathbf{v}\|_2^2 \\ & \text{subj. to} & \mathbf{v} \geq 0 \end{array} \tag{7.3}$$

The objective of this method is to iteratively learn the support of the sparse $\mathbf{v}$. As the weights on the support get smaller and off the support get larger, the estimator becomes unbiased. Instead of using this to debias the estimates, we will use it to encode our assumptions that most of the support of $\mathbf{v}_{t+1}$ will be the same as the support of $\mathbf{v}_t$. We slightly modify the weighting function, using a floor parameter $\beta = 0.10$,

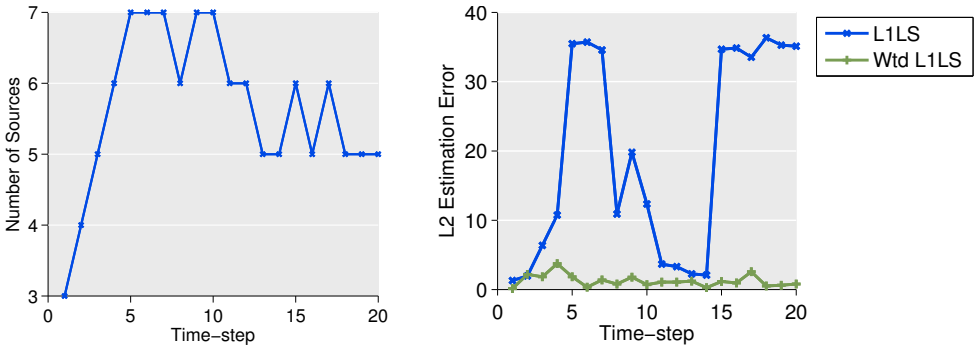
$$w_t(i) = \frac{1}{|\max\{v_t(i), \beta\}|}, \quad \mathbf{w}_t = \begin{bmatrix} w_1 & w_2 & \cdots & w_N \end{bmatrix} \tag{7.4}$$

and then estimate $\hat{\mathbf{v}}_{t+1}$ using the weighted ℓ_1 Penalty, in (7.3).

As we will show with experimental simulations, this method allows the use of fewer network observations in the time-varying dynamic FICCS approach.

7.3 Example Simulation Experiment

To test the approach, we simulate a field monitored by $N = 100$ sensors arranged in a 10-by-10 regular grid. The potential source locations also correspond to this grid, with spacing $\delta = 2$ units. The initial number of sources is $S = 3$. We present field reconstruction estimates for randomly generated scenarios using the model of (7.1), with unknown source magnitudes set to 15, and measurement noise with variance 0.01. At the initial time-step, we used a larger set of measurements, $\mathbf{y}_0 \in \mathbb{R}^M$, with $M = 30$, to compute the initial estimate $\hat{\mathbf{v}}_0$. After, there were 20 time-steps of new observations of the changing $\mathbf{v}_t$, each $\mathbf{y}_t \in \mathbb{R}^m$, for $m = 15$. By leveraging results of previous time-steps estimates, FICCS should be able to withstand fewer new measurements.

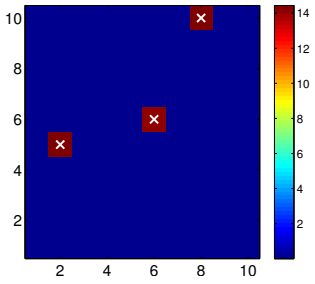


(a) Num. Sources vs. time-step

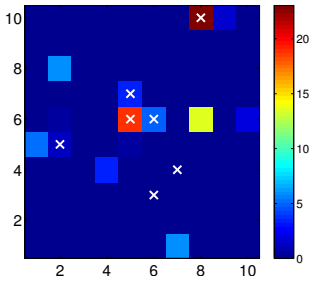
(b) Estimation Error vs. time-step

Figure 7.2: Example Simulation Test: The number of sources, panel a, and ℓ_2 estimation error, panel b, versus time-step of FICCS estimates. Panel (b) compares the mean ℓ_2 estimation error of performing independent L1LS at each new time-step with only $m = 15$ network projections versus the dynamically re-weighted L1LS which uses the estimate of the previous time-step as a prior.

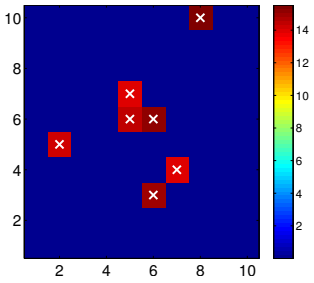
Figure 7.2 shows the ℓ_2 estimation error the equal-weighted ℓ_1-penalized least squares (L1LS) and the WtdL1LS estimators. The upper left panel shows the source support change over time. This is a challenging scenario with frequent changes of where the sources are and has as many as 7 sources. The upper right panel compares the estimates using only the current observation, as a control, to the adaptively reweighted method. We see the reweighted method is better able to handle the support changes over time since it uses the *priors* of the estimate of the previous time-step. The two lower panels show the reconstructed source field with the correct locations overlayed at the last time-step, showing that the reweighted estimator accurately localized and recovered the source field.



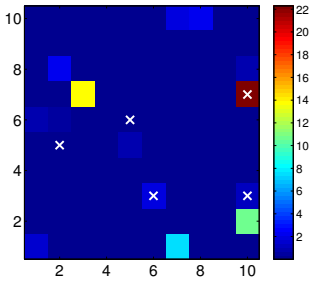
(a) L1LS, $t = 0$



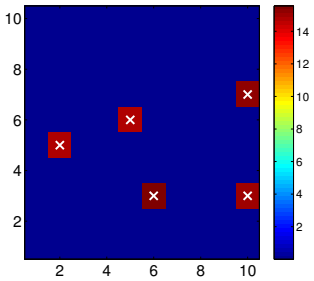
(b) L1LS, $t = 5$



(c) WtD L1LS, $t = 5$



(d) L1LS, $t = 20$



(e) WtD L1LS, $t = 20$

Figure 7.3: Example Simulation Test: Comparison of the source field estimates of L1Ls, panels (b) and (d), vs. the sequential estimator Wtd. L1LS, panels (c) and (e), at time-steps 5 and 20. Both estimators start with the initial L1LS $\widehat{\mathbf{v}}_0$, shown in panel (a), which uses 30 network projections to compute the initial estimate. Correct source locations as white crosses.

The next section, 7.4, shows average estimation error and detection rates as a function of time-step and number of network projections for a wider set of scenarios.

7.4 Simulation Experiments

We performed 100 trials, similar to that of section 7.3, except that now sources can arrive anywhere in the monitored area, and not just at grid-points. The propagation law used is the inverse squared distance; i.e.,

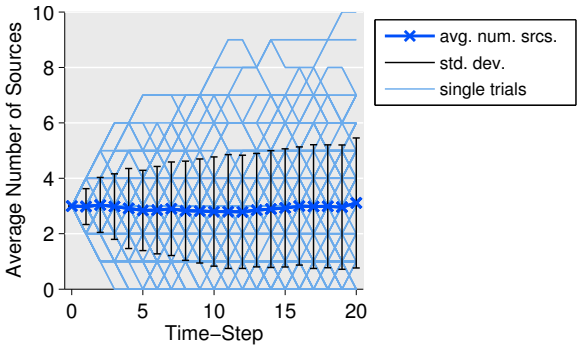
$\gamma = 2$. A network of 100 sensors arranged in a grid observes an area, with discretization lattice spacing of $\delta = 2$. Initially there are 3 sources in unknown locations. The support of $\mathbf{v}_t$ changes according to $p_+ = p_- = 0.2$, and stays the same with probability 0.6. The initial number of consensus network projections at $t = 0$ is $m_0 = 30$, while at each subsequent time-step the network computes only 15 new projections.

Figure 7.4, panel (a) shows the average number of sources over time, for the 20 time-step simulations. Individual traces are shown in light blue and the $\pm$ one standard deviation error bars are also shown. This shows that the variance in the number of sources increases over time, but the expected number of sources stays near 3. The ℓ_2 estimation error of ℓ_1 penalty (L1LS) and dynamically re-weighted ℓ_1 penalty (Wtd L1LS) are shown in green $+$'s and blue $\times$'s, respectively. Since the re-weighted ℓ_1 penalty uses the values of the previous time-step to compute weights for the next time-step, it is better able to track the changing support vector with fewer network observations. Panels (b) and (c) contrast the effects of sources arriving only at grid locations, in (b), and anywhere in the monitored region, in (c). In both cases, dynamically re-weighted ℓ_1 penalty improves on the estimates that do not use previous information. The increases in error due to off-grid source locations in this scenario is considerable.

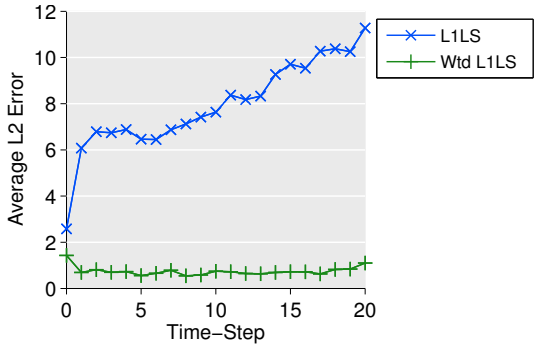
Figure 7.5 performs the dynamic inversion tests in a scenario where the sensor and source lattice have the smaller spacing, $\delta = 1$. The initial number of consensus network projections at $t = 0$ is increased to $m_0 = 40$, and at each subsequent time-step the network computes 20 projections. In this case, the modeling errors are lower, and more information is collected, so the average estimation error of the off-grid estimates is lower. This shows that, by using priors based on previous estimates, the ℓ_1 penalized least squares optimization can be simply adapted to the time-varying setting.

7.5 Conclusions

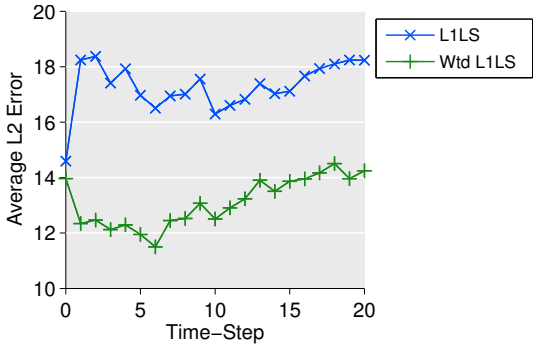
We studied a recursive decentralized sensor network sparse reconstruction algorithm that is adapted to a time-varying appearance and disappearance model for point-sources. After each measurement collection a small number of compressed network observations are computed by distributed averaging message passing. Each node estimates the field using ℓ_1 penalized least squares optimization, with a weighted ℓ_1 norm penalty, roughly corresponding to the inverse magnitudes of the values of the field estimated during



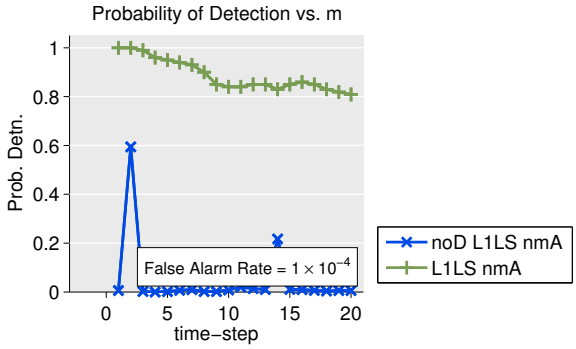
(a) Avg. Number of Sources vs. t



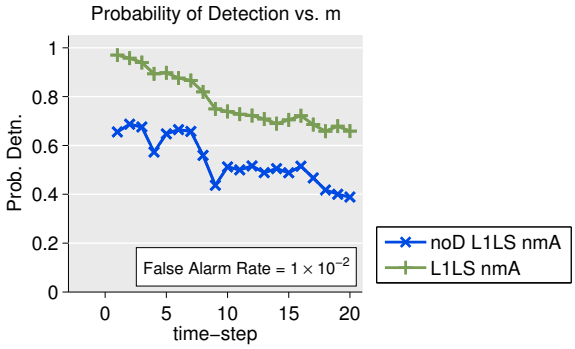
(b) On-grid Avg. Estimation Error vs. t



(c) Off-grid Avg. Estimation Error vs. t

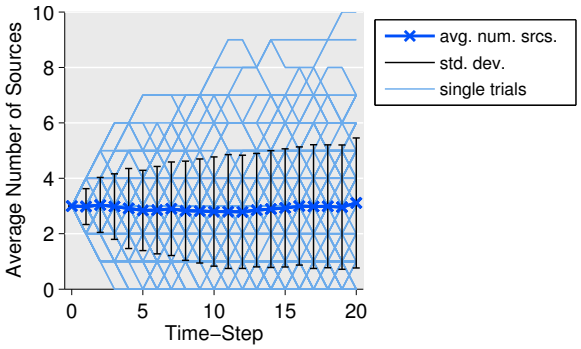


(d) On-grid P_{D} vs. t

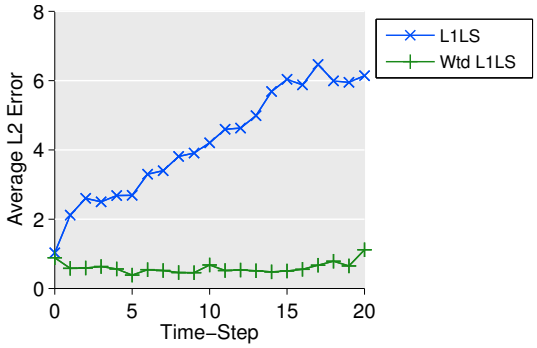


(e) Off-grid P_{D} vs. t

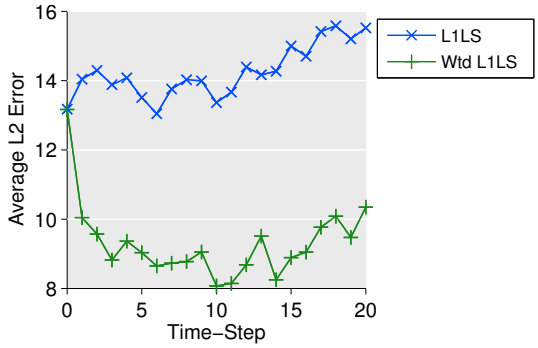
Figure 7.4: Tests of dynamic reweighted ℓ_1 penalty, for acoustic source simulation, source and sensor lattice spacing of $\delta = 2$, averaged over 100 trials. The initial number of network projections at $t = 0$ is $m_0 = 30$, with each subsequent time-step using only 15 network projections. Panel (a) shows the average number of sources versus time-step, in blue $\times$'s, with black error bars marking the $\pm$ one standard deviation and light blue series showing individual traces of number of sources vs. time. Panels (b) and (c) show average estimation error in the case where source fall exactly on grid points, on the left, and uniformly continuously in the space (i.e. off-grid) on the right. The blue $\times$'s show the performance of ℓ_1 penalty with no information from prior estimates, while green $+$'s denote the performance of the dynamically reweighted ℓ_1 penalty estimator. Panels (d) and (e) show detection rates for average false alarm rate of 0.0001 and 0.01, respectively.



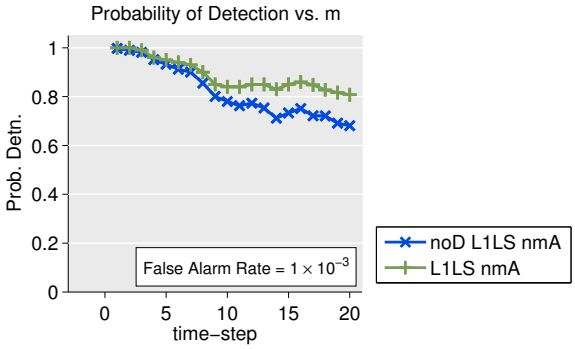
(a) Avg. Number of Sources vs. t



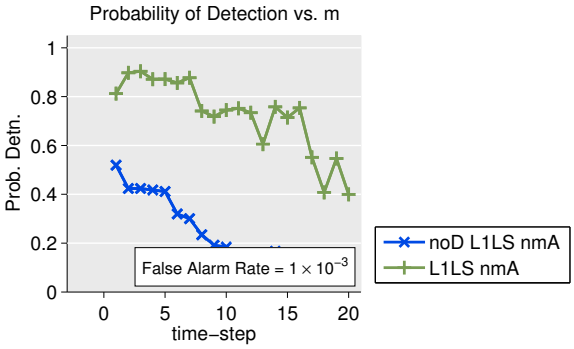
(b) On-grid Avg. Estimation Error vs. t



(c) Off-grid Avg. Estimation Error vs. t



(d) On-grid P_{D} vs. t



(e) Off-grid P_{D} vs. t

Figure 7.5: Tests of dynamic reweighted ℓ_1 penalty, for acoustic source simulation, source and sensor lattice spacing of $\delta = 1$, averaged over 100 trials. The initial number of network projections at $t = 0$ is $m_0 = 40$, with each subsequent time-step using 20 network projections. Panel (a) shows the average number of sources versus time-step, in blue $\times$'s, with black error bars marking the $\pm\sigma$ and light blue series showing individual traces of number of sources vs. time. Panels (b) and (c) show average estimation error in the case where source fall exactly on grid points, on the left, and uniformly continuously in the space (i.e. off-grid) on the right. The blue $\times$'s show the performance of ℓ_1 penalty with no information from prior estimates, while green $+$'s denotes the performance of the dynamically reweighted ℓ_1 penalty estimator using the values of the previous time-step to compute weights. Panels (d) and (e) show detection rates for average false alarm rate of 0.001.

the previous time-step. We found the adaptive re-weighted LASSO method was able to incorporate source estimates from the previous time-steps. This method allows for the computation of fewer network observations, and is simple to implement. We conclude that the dynamically re-weighted ℓ_1 penalized least squares estimator lends itself well to problems with time-varying support change.

Chapter 8

On the Benefits of Measurement Mixing

At the heart of the FICCS algorithm is the idea of using local measurement mixing to create informative projections of global parameters. We first noticed the improvements due to mixing in [19], in which sensor nodes mix their values with random linear combinations of the measurements of neighbors. This initial approach showed the potential for communication-efficient information fusion through distributed averaging. In this chapter, we present an analysis justifying why low dimensional network projections can efficiently capture global information of the full field.

First, we treat the case where sparse system parameters are multivariate normal, but their locations are known. This corresponds to inverting an acoustic field where source locations are known but magnitudes are random. This leads to principle components analysis (PCA), [57]. Previous work on linear experiment design treated the case where additive noise enters the system after measurement compression and obtained the same form of optimal sensing matrices only when imposing the restriction of right-orthonormality, [58]. We show that restriction is unneeded in this particular observation model, with proof of the optimality of PCA when measurement noise that is added prior to mixing. The optimal linear projections correspond in compressed sensing theory to the optimal sparsifying basis.

The second component of the analysis in section 8.3 reintroduces randomness to sparse parameter locations with a Bernoulli model for active parameter locations, leading to a Gaussian Mixture model for the field. By examining a lower bound on the mutual information of mixed measurements, we see a different perspective on the design of the observation projection matrices. The performance of the

basis pursuit estimator is compared to the optimal MMSE and MAP estimators in an acoustic source localization example, showing that in the case of low measurement noise there is little gap between the performance of the estimators, but in the case of high measurement noise, there is a large gap between the mean-squared error of optimal estimators as compared with basis pursuit.

8.1 Problem Formulation

We assume a known linear model for the phenomena to be estimated and that the fundamental parameters may be described as a sparse combination of known basis functions. Each sensor observes a scalar value of the field, corrupted by additive Gaussian measurement noise, $\mathbf{n} \sim N(\mathbf{0}, \sigma_n^2 I)$. The vector $\mathbf{x}$ corresponds to the collection of all sensor measurements prior to the addition of measurement noise.

$$\mathbf{x} \;=\; \sum_{i=1}^{N} \psi_i v_i \;=\; \Psi \mathbf{v} \tag{8.1}$$

The sparse parameter vector $\mathbf{v}$ has the same dimension as $\mathbf{x}$, but has at most s nonzero components. This is sometimes written as $\|\mathbf{v}\|_0 \leq s$.

In the example application, acoustic source localization described in chapter 2, we model the basis functions, ψ_i, by the propagation of acoustic wavefields. We model the measured value at location i due to source j as the source intensity, v_j, divided by 1 plus the square distance between locations i and j. The N sensors are placed on a two-dimensional $\sqrt{N}$-by-$\sqrt{N}$ regular lattice, which also corresponds to the potential locations for acoustic sources. Since the measurements at each sensor are the superposition of all propagating wavefields, we may represent the observations by a linear matrix equation where the entries in Ψ correspond to inverse intersensor square distances.

In the FICCS approach, presented in chapter 3, each sensor may estimate the unknown random vector, given knowledge of the prior distribution of $\mathbf{v}$, the linear propagation matrix of the system Ψ, and the distributed observation projections. The observation equation is a linear mapping represented by selection and mixing matrices, $H \in \mathbb{R}^{m \times N}$ and $\Phi \in \mathbb{R}^{N \times N}$, where the selection matrix is structured to have one 1 in each row.

$$\mathbf{z} \;=\; H(\Psi \mathbf{v} + \mathbf{n}) \tag{8.2}$$

$$\mathbf{y} \;=\; H\Phi(\Psi \mathbf{v} + \mathbf{n}), \quad \Phi \in \mathbb{R}^{N \times N} \tag{8.3}$$

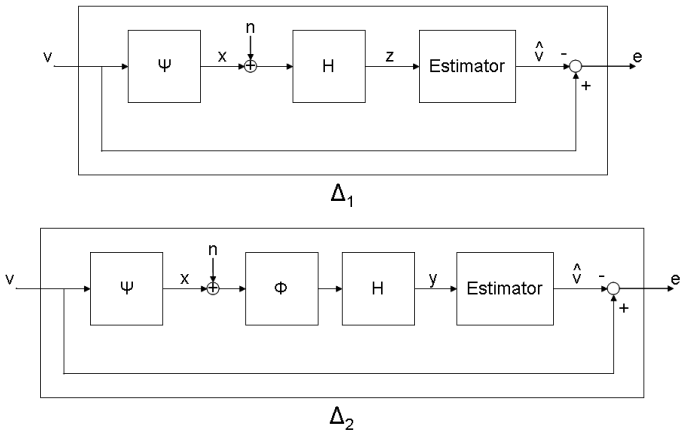


Figure 8.1: The two error systems; in the upper system the observation matrix is restricted to a simple selection matrix H, in the lower system, Φ allows the production of any linear mapping.

In the subset selection case, we refer to the selected observations as $\mathbf{z}$. We compare this with $\mathbf{y}$, created through iterative intersensor communications, to compute a general linear mapping, $H\Phi : \mathbb{R}^N \rightarrow \mathbb{R}^m$.

The observation vector $\mathbf{y}$ is distributed throughout the network using distributed average consensus as detailed in the FICCS algorithm, in chapter 3. Using the FICCS approach to distribute linear projections of the network state, we analyze the problem of estimating the source vector $\mathbf{v}$ from the compressed measurements $\mathbf{z}$ or $\mathbf{y}$. Figure 8.1 shows the alternate error systems. The next section proves the form of the optimal measurement projection matrix $H\Phi$ when the locations of the sparse entries of $\mathbf{v}$ are known.

8.2 Linear Gaussian Model

When acoustic source locations are known in $\mathbf{v}$, and the intensities of the sources have a multivariate normal distribution, $N(\mu_v, \Sigma_v)$, the observation system corresponds to a linear Gaussian model. Conditioned on source locations, the observations $\mathbf{y}$ and $\mathbf{v}$, alternatively $\mathbf{z}$ and $\mathbf{v}$, are jointly multivariate normal, and so the performance of the minimum-mean-squared error (MMSE) estimator for a given observation matrix is well characterized.

The MMSE estimate is the conditional expected value of $\mathbf{v}|\mathbf{y}$. Using the well-known Gauss-Markov

theorem, for $\mathbf{y} = A(\Psi\mathbf{v} + \mathbf{n})$, the conditional mean and variance are given by,

$$\begin{aligned} \text{for } B &= \Sigma_v \Psi^T A^T \left(A\Psi\Sigma_v\Psi^T A^T + \sigma_n^2 AA^T \right)^{-1} && (8.4) \\ E\left[\mathbf{v}|\mathbf{y}\right] &= \mu_v + B(\mathbf{y} - \mu_y) && (8.5) \\ var\left(\mathbf{v}|\mathbf{y}\right) &= \Sigma_v - BA\Psi\Sigma_v && (8.6) \end{aligned}$$

The mutual information follows $I(\mathbf{y};\mathbf{v}) = h(\mathbf{y}) - h(A\mathbf{n})$.

$$I(\mathbf{y};\mathbf{v}) = \frac{1}{2}\log\left(\frac{\left|A(\Psi\Sigma_v\Psi^T + \sigma_n^2 I)A^T\right|}{|\sigma_n^2 AA^T|}\right) \quad (8.7)$$

The source covariance matrix Σ_v is known and has exactly s nonzero eigenvalues, since there are a known set of s source locations. Next we will prove that the sensing matrix, A, that maximizes the mutual information of the measurements corresponds to rows of the inverse eigenvector matrix. In other words, the rows of A are the m eigenvectors of $\Sigma_x = \Psi\Sigma_v\Psi^T$ corresponding to the largest eigenvalues. The proof first restricts A to the set of right-orthonormal matrices, i.e., $AA^T = I$. Then we show that the maximum over all $A \in \mathbb{R}^{m\times N}$ is attained in the set of right-orthonormal A.

Optimal orthonormal sensing matrices

We prove that, in the case of known source locations, performing optimal mixing must result in more informative measurements than simple selection. The observations functions are as defined in (8.2) and (8.3).

$$\textit{Claim:} \quad \max_{H,\Phi} I(\mathbf{y};\mathbf{v}) \geq \max_{H} I(\mathbf{z};\mathbf{v}) \quad (8.8)$$

The H matrix is an m-by-N selection matrix and, in this section we restrict the matrix Φ to be square and orthonormal. We note that the solution to the selection matrix that maximizes $I(\mathbf{z};\mathbf{v})$ is a combinatorial problem, and so computing the optimal H is generally not tractable, although many suboptimal relaxations of the problem exist; e.g. [59, 60].

As an outline of the proof of (8.8), we will first find an upperbound for $I(\mathbf{y};\mathbf{v})$, then show that it also applies to $I(\mathbf{z};\mathbf{v})$, after which we show the upperbound for $I(\mathbf{y};\mathbf{v})$ is achievable using a particular orthonormal Φ. We will first seek to upperbound $I(\mathbf{y};\mathbf{v})$, corresponding to (8.3). We may represent Σ_x

in diagonalized form,

$$\Sigma_x = \Psi\Sigma_v\Psi^T \quad = \quad U\Lambda_x U^T \tag{8.9}$$

where we restrict U to be orthonormal and arrange the columns such that the eigenvalues in the diagonal matrix Λ_x are ordered $\lambda_1 \geq \cdots \geq \lambda_N \geq 0$. Since there are only s sources we know that λ_i for $i > s$ are all zero. The mutual information is,

$$I(\mathbf{y};\mathbf{v}) = \frac{1}{2}\log\left(\frac{\left|H\Phi(\gamma U\Lambda_x U^T + I)\Phi^T H^T\right|}{|H\Phi\Phi^T H^T|}\right) \tag{8.10}$$

where γ is $1/\sigma_n^2$. Without loss of generality, assume that the selection matrix H selects the top m rows. Because we restricted Φ to be orthonormal, the information simplifies to

$$\frac{1}{2}\log\left(\left|H\left(\gamma\Phi U\Lambda_x U^T\Phi^T + I\right)H^T\right|\right) \tag{8.11}$$

Maximizing the mutual information is then equivalent to maximizing the determinant of (8.11).

Now we show $\left|H(M+I)H^T\right|$ is less than or equal to the product of the largest m eigenvalues of the matrix $M + I$, for all selection matrices H and all $M \geq \mathbf{0}$. This is true by the Interlacing Inequality on eigenvalues of principle submatrices, [61]. For an m-by-m principle submatrix, $K = HMH^T$, of the n-by-n Hermitian matrix positive semi-definite, M, the eigenvalues of the submatrix, $\widetilde{\lambda}_1 \geq \cdots \geq \widetilde{\lambda}_m$, satisfy

$$\widetilde{\lambda}_i \leq \lambda_i \tag{8.12}$$

where the λ_i are the eigenvalues of M also arranged in decreasing order. Using the fact that the determinants of symmetric matrices are the products of eigenvalues, we represent the determinants of $K + I$ and $M + I$.

$$|K+I| \quad = \quad \prod_{i=1}^{m}(\widetilde{\lambda}_i + 1) \tag{8.13}$$

$$|M+I| \quad = \quad \prod_{i=1}^{n}(\lambda_i + 1) \tag{8.14}$$

The inequality on each eigenvalue implies that $\widetilde{\lambda}_i + 1 \leq \lambda_i + 1$ for all $i = 1, \ldots, m$, so we have

$$|K+I| \leq \prod_{i=1}^{m}(\lambda_i + 1), \text{ where } K = HMH^T \tag{8.15}$$

Replacing M with $\gamma\Phi U\Lambda U^T\Phi^T$, and noticing that ΦU is orthonormal, we have

$$\left|H(\gamma\Phi U\Lambda U^T\Phi^T + I)H^T\right| \;\leq\; \prod_{i=1}^{m}\left(\frac{\lambda_i}{\sigma_n^2}+1\right) \tag{8.16}$$

where λ_i are the eigenvalues in Λ listed in decreasing order. Similarly we may replace the M matrix with $U\Lambda U^T$ and see that the upperbound applies to the maximization over simple selection matrices as well.

$$\left|H(\gamma U\Lambda U^T + I)H^T\right| \;\leq\; \prod_{i=1}^{m}\left(\frac{\lambda_i}{\sigma_n^2}+1\right) \tag{8.17}$$

We have shown a common upperbound on both maximizations in (8.8). Now we show that the choice of $\Phi = U^T$ in the first maximization achieves the upperbound.

$$\left|\gamma HU^TU\Lambda U^TUH^T + I\right| = \prod_{i=1}^{m}\left(\frac{\lambda_i}{\sigma_n^2}+1\right) \tag{8.18}$$

In the case of maximizing $I(\mathbf{y};\mathbf{v})$, choosing Φ as the inverse eigenvector matrix is optimal. Since H selects the first columns, we see that the optimal mapping $H\Phi$ is to choose, as rows, the eigenvectors having the largest eigenvalues in Σ_x.

Since we have shown an upperbound on both informations and achievability by a particular H and Φ, we have proven the inequality in (8.8). To show this relationship is practically interesting one can take a particular nondiagonal choice Σ_x and show experimentally that all possible selection matrices H cannot achieve this upperbound of $I(\mathbf{y};\mathbf{v})$. In practice, the amount of improvement due to Φ, as compared with the optimal selection matrix, depends on the particular Σ_x involved. However, the process of finding the optimal $H\Phi$ is only as difficult as diagonalizing the signal covariance matrix, whereas computing the optimal H is a combinatorial problem, [59].

Proof of invariance to non-orthonormality

We proved that the optimal orthonormal mixing matrix corresponds to the PCA matrix for the signal covariance. Now we show that relaxing the restriction of orthonormality of Φ does not increase the upperbound of mutual information. To prove this we return to maximizing $I(\mathbf{y};\mathbf{z})$ given in (8.10). Without loss of generality, define $\widehat{\Phi} = \Phi U$, using the same U from (8.9) , and replace Φ with $\widehat{\Phi}U^T$,

yielding

$$I(\mathbf{y};\mathbf{v}) = \frac{1}{2}\log\left(\frac{\left|H\widehat{\Phi}(\gamma\Lambda_x + I)\widehat{\Phi}^T H^T\right|}{\left|H\widehat{\Phi}\widehat{\Phi}^T H^T\right|}\right) \tag{8.19}$$

The argument of the maximization is now $\widehat{\Phi}$. Now we note that $H\widehat{\Phi}$ is a rectangular matrix in $\mathbb{R}^{m\times N}$. We define the left m-by-m block of $H\widehat{\Phi}$ as A and the remaining right block as B. The matrix $\gamma\Lambda_x + I$ is diagonal in structure with the first m values on the diagonal equal to $\gamma\lambda_k + 1$ for $1 \leq k \leq m$, and the remaining values on the diagonal equal to 1 for $k \geq m+1$. We can write this matrix in blocks with the upper left m-by-m block called Σ and the lower right block is the identity. In terms of these block matrices, our objective is to maximize the following.

$$\max_{A,\,B} \ \frac{\left|A\Sigma A^T + BB^T\right|}{|AA^T + BB^T|} \tag{8.20}$$

We note that $\Sigma > I$ in the positive definite sense, since the λ_k of the signal covariance matrix were positive for $1 \leq k \leq m$. This also implies, $A\Sigma A^T > AA^T$ for all A. We can show the ratio of determinants of (8.20) is maximized when B is the zero matrix. For positive definite Q and R, with $Q > R$, and positive semi-definite M, the maximum of $|Q+M|/|R+M|$ is achieved at $M = \mathbf{0}$. First we can show that we may treat M as a diagonal matrix. If it were not diagonal, we could use its diagonalization, $M = VDV^T$, to rewrite the problem into the equivalent one,

$$\max_{D} \ f(D) = \frac{\left|V^T\right|}{|V^T|}\frac{|Q+M|}{|R+M|}\frac{|V|}{|V|} = \frac{\left|V^TQV + D\right|}{|V^TRV + D|} \tag{8.21}$$

where for $Q' = V^TQV$ and $R' = V^TRV$, we again have $Q' > R'$. This function of a diagonal positive M is a strictly decreasing function of diagonal elements of M, which is seen by taking the derivative with respect to the matrix M.

$$\frac{\delta f(M)}{\delta M} = \frac{|Q+M|\left[(Q+M)^{-1} - (R+M)^{-1}\right]}{|R+M|} \tag{8.22}$$

Since $Q > R$, if follows that $Q+M > R+M$ and that $(Q+M)^{-1} < (R+M)^{-1}$, meaning the overall derivative matrix is strictly negative definite. For a strictly decreasing f the maximum, over $M \geq \mathbf{0}$ is achieved at $M = \mathbf{0}$. Applying this, we find the choice of the square A in (8.20) is arbitrary, assuming that it is full rank.

$$\left|A\Sigma A^T\right| / \left|AA^T\right| = |\Sigma| = \prod_{i=1}^{m}\left(\frac{\lambda_i}{\sigma_n^2} + 1\right) \tag{8.23}$$

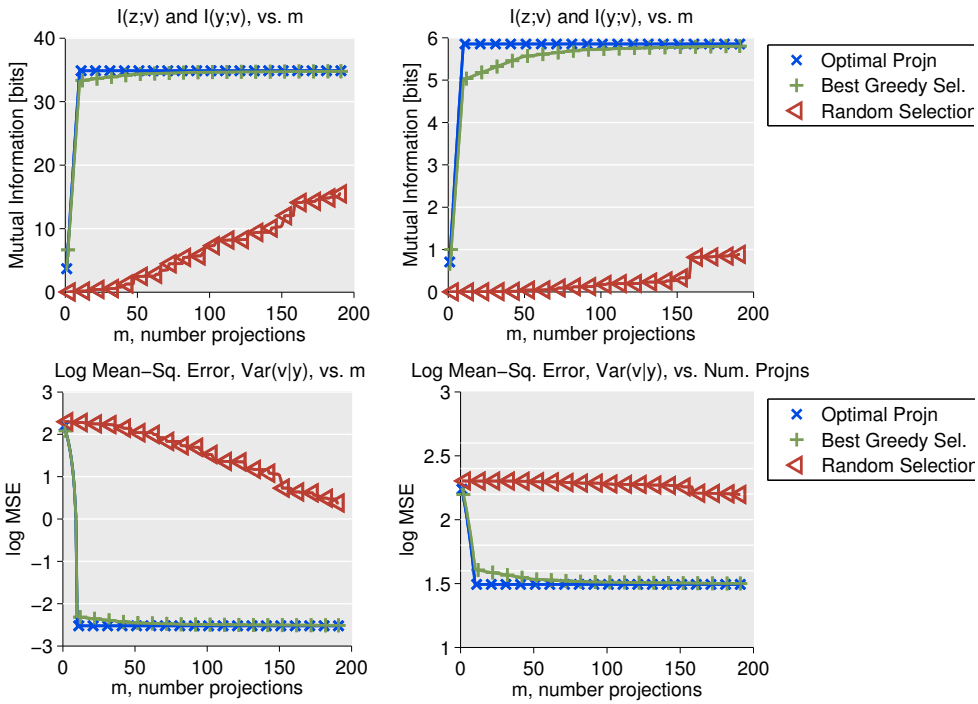


Figure 8.2: Numerical tests for 10 Known Source Locations, 2 noise levels; left plots: $\sigma_n^2 = 0.01$, right plots: $\sigma_n^2 = 1$. $N = 1024$, m is varied, $\mu_s = 20$, $\sigma_s^2 = 1$. Ψ the inverse squared distance matrix. Upper panels compare bounds and estimated mutual information for mixed versus selected measurements. In the selection case, both random and best greedy selection are shown. Lower panels compare squared estimation error for the MMSE estimate.

We have recovered the optimum found in the previous section, where we restricted Φ to the set of orthonormal matrices showing that $\Phi = U^T$, is also optimal over all matrices. This differs from the noisy PCA analyses in [58], since our observation equation $H\Phi$ scales both signal and noise components equally. The interlacing inequalities arguments can also be applied to show that $\Phi = U^T$ minimizes the mean-squared error of the estimate $E[\mathbf{v}|\mathbf{y}]$, recovering the well-known result that optimal linear experimental design by maximizing mutual information is equivalent to the objective of minimizing mean-squared error for Gaussian random vectors.

Figure 8.2 shows the resutling mutual information and variance of the minimum mean-squared error estimates in an experiment with 1024 sensors and 10 sources at known locations. The optimal Φ uses the eigenvectors of the measurement covariance matrix, reaching the upperbound on information after 10 measurements. Both random selection and the best greedy selection are shown. In greedy selection, sensors are selected in sequence according to which measurement lowers the variance of the estimate

the most. This is known to yield a sensor selection within information within $(1 - 1/e) = 0.63$ of the optimal selection set. However, determining this best sequence takes time, with complexity $\mathcal{O}(N^2)$. In this case, since the Ψ has propagation functions with highly localized energy, the greedy selection gets close to the optimal Φ projection measurements. However, choosing the optimal Φ was only as complex as computing the eigenvectors of Σ_x, and so is still a preferable strategy. Compared to random sensor, selection, the gains in mutual information and reduced estimation variance are very large.

8.3 Bernoulli Locations and Gaussian Intensities

We now analyze the benefits of measurement mixing for the problem of sparse sources at unknown locations with random Gaussian intensities. Similar to the objective in the linear Gaussian model, stated in (8.8), we would like to show that mutual information of the mixed measurements is greater than selected sensor measurements.

Unlike the case of known source locations, in which the signal vector $\mathbf{v}$ is multivariate normal, now source presence at each location in $\mathbf{v}$ is modeled as a Bernoulli random variable. Each component, v_i, is independent, identically distributed, with the model,

$$v_i \sim \begin{cases} N\left(\mu_s, \sigma_s^2\right) & \text{w.p. } p \\ 0 & \text{w.p. } 1-p \end{cases} \tag{8.24}$$

where $N\left(\mu_s, \sigma_s^2\right)$ denotes the normal distribution with mean μ_s and variance σ_s^2.

The linear observation equations, (8.2) and (8.3), are unchanged. Since the additive noise is Gaussian, the random observation variable $\mathbf{y}$ follows a Gaussian mixture distribution with $L = 2^N$ components. We can see this by introducing the binary random vector $\mathbf{q} \in \mathbb{Z}^N$, where $q_i = 1$ indicates a source present at location i, while a zero means no source. Indexing each of the L potential source configurations by the index k, we may write the distribution for $\mathbf{y}$ as

$$p(\mathbf{y}) = \sum_{k=1}^{L} p(\mathbf{q}_k) p(\mathbf{y}|\mathbf{q}_k) = \sum_{k=1}^{L} p(\mathbf{q}_k) N\left(\mu_k, \Sigma_k\right) \tag{8.25}$$

The vector $\mathbf{y}|\mathbf{q}$ is multivariate normal with mean $\mu_k = \mu_s A_{T_k} \mathbf{1}$ and variance $\Sigma_k = \sigma_s^2 A_{T_k} A_{T_k}^T + \sigma_n^2 I$, in which A corresponds to the right-orthonormal observation matrix, $H\Psi$ in the selection case and $H\Phi\Psi$ in the mixing case, and A_{T_k} is A restricted to the columns of the indices in the set T_k.

The mutual information is the entropy of the observations less the entropy of the additive noise. To see this,

$$I(\mathbf{y};\mathbf{v}) = h(\mathbf{y}) - h(\mathbf{y}|\mathbf{v}) \tag{8.26}$$

$$\text{Since} \quad h(\mathbf{y}|\mathbf{v}) = h(\mathbf{n}), \tag{8.27}$$

$$I(\mathbf{y};\mathbf{v}) = h(\mathbf{y}) - h(\mathbf{n}) \tag{8.28}$$

However since the observations are Gaussian mixtures, their entropies do not have a closed form expression, [62]. Instead, we may analyze an upperbound to $h(\mathbf{y}) = h(\mathbf{y},\mathbf{q}) - H(\mathbf{q}|\mathbf{y})$. Since $\mathbf{q}|\mathbf{y}$ is a discrete random variable, its entropy is strictly positive, meaning $h(\mathbf{y}) \leq h(\mathbf{y},\mathbf{q})$, where the joint entropy is readily computed yielding the upperbound $I_{UB}(\mathbf{y};\mathbf{v})$ below.

$$I(\mathbf{y};\mathbf{v}) = h(\mathbf{y}) - h(\mathbf{n}) \tag{8.29}$$

$$\leq h(\mathbf{y},\mathbf{q}) - h(\mathbf{n}) \tag{8.30}$$

$$= \sum_{k=1}^{L} p(\mathbf{q}_k) \log \left\{ \frac{(2\pi e\,|\Sigma_k|)^{\frac{1}{2}}}{p(\mathbf{q}_k)} \right\} - \frac{m}{2} \log\left\{2\pi e \sigma_n^2\right\} \quad = \quad I_{UB}(\mathbf{y};\mathbf{v}) \tag{8.31}$$

The upperbound on mutual information involves an exponential number of terms. To test how mixing can increase mutual information, we consider a case where an oracle tells us there is at most one source. In this case the number of mixture components is $N+1$. We compare the upperbound on $I(\mathbf{y},\mathbf{v})$ to the upperbound on $I(\mathbf{z};\mathbf{v})$ for the particular case of $\Psi = I$, meaning no propagation.

Conditioning the mixture components on the event of at most one source, we denote w_i, for $i = 1, ..., N+1$, as the sequence of probabilities. All of these probabilities are equal except for w_{N+1}, which we denote as corresponding to no source present. Upperbounds on $I(\mathbf{y},\mathbf{v})$, where $\mathbf{a}_i$ refer to the columns of the observation matrix, are given by,

$$I(\mathbf{y};\mathbf{v}) \leq \sum_{i=1}^{N} \frac{w_i}{2} \log\left(\frac{\sigma_s^2}{\sigma_n^2}\mathbf{a}_i^T\mathbf{a}_i + 1\right) - \sum_{i=1}^{N+1} w_i \log\left(w_i\right) \tag{8.32}$$

This yields a simple upperbound for selection $I_{UB}(\mathbf{z};\mathbf{v})$ that is linear in m. Since Φ is orthonormal, all $\mathbf{a}_i^T\mathbf{a}_i \leq 1$ and we can set an upperbound on mixing performance.

$$I_{UB}(\mathbf{y};\mathbf{v}) = \frac{N \cdot w_1}{2} \log\left(\frac{\sigma_s^2}{\sigma_n^2} + 1\right) - \sum_{i=1}^{L} w_i \log\left(w_i\right) \tag{8.33}$$

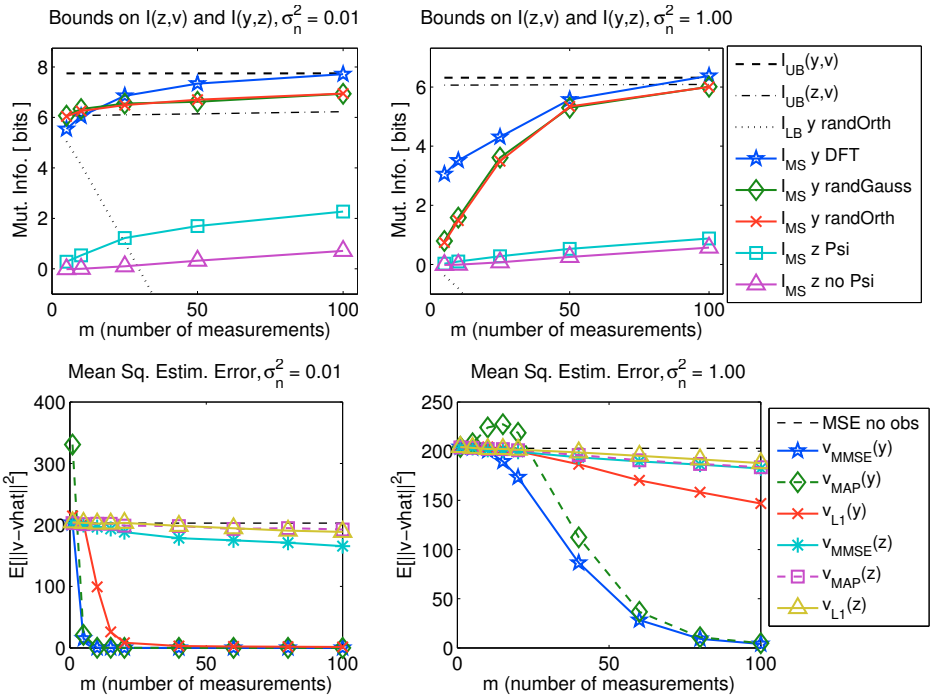


Figure 8.3: Numerical tests for 2 noise levels; left plots: $\sigma_n^2 = 0.01$, right plots: $\sigma_n^2 = 1$. $N = 1024$, m is varied, $\mu_s = 20$, $\sigma_s^2 = 1$. Ψ the inverse squared distance matrix. Upper panels compare bounds and estimated mutual information (5000 samples) for various choices of Φ and Ψ. Lower panels compare average squared estimation error, 1000 trials, for MMSE, MAP, and ℓ_1-minimization estimators, for orthonormal Φ versus random selection.

We can also lower bound this mutual information, using Jensen's inequality on the entropy, as described in [62]. The resulting lowerbound $I_{LB}(\mathbf{y};\mathbf{v})$ is

$$I(\mathbf{y};\mathbf{v}) \quad \geq \quad -\sum_{i=1}^{N+1} w_i \log \left\{ \sum_{j=1}^{N+1} w_j N(\mu_j;\mu_i,\Sigma_i+\Sigma_j) \right\} - \frac{m}{2}\log\left\{2\pi e\sigma_n^2\right\} \tag{8.34}$$

where μ_i and Σ_i are as defined in (8.31), and $N(\cdot;\mu,\Sigma)$ refers to the evaluation of the normal distribution at the argument. From I_{LB}, we see this bound on mutual information is maximized when the modes of $p(\mathbf{y})$, the $N(\mu_i,\Sigma_i)$, have minimal expected overlap. This is like an average-case version of mutual coherence, a measure of the measure of the maximum correlation of the columns of A, which is the key of mutual coherence. Since this is similar to the mutual coherence measure in compressed sensing, we have reason to expect random matrices will provide informative measurement projections.

The upper panels of fig. 8.3 compare the upperbounds on mutual information, the dashed lines, to measured values of $I(\mathbf{y};\mathbf{v})$ for various mixing matrices, applied to the acoustic source localization problem of sect. 8.1. Each information estimate is computed via Monte Carlo sampling each with 5000

sampled observations. The x-axis varies the dimension of $\mathbf{y}$, showing the increase in information with additional measurements. The upper three series represent randomly sampled Φ matrices suggested by compressed sensing literature; Gaussian entries, orthonormal [14], and Φ_{DFT}, which, with the propagation matrix Ψ, yields rows of the discrete Fourier transform [12]. In both noise regimes, $I(\mathbf{y};\mathbf{v})$ quickly approach their upperbounds. The series for $I(\mathbf{z};\mathbf{v})$, computed using random selections, provided very little information both with and without propagation. The dotted line shows the lowerbound on information for the random orthonormal Φ, which was much lower than the experimental estimates.

The lower panels of fig. 8.3 show estimation error averaged over 1000 trials where $\mathbf{y}$ are projections using random orthonormal mixing and $\mathbf{z}$ are selections of the propagated acoustic field. We compare the MMSE, maximum a posteriori (MAP), and sparse recovery ($\hat{\mathbf{v}}_{L1}$) estimators. Sparse recovery estimates minimize the ℓ_1-norm subject to a 99% threshold on observation errors, solved using the `cvx` semidefinite programming package [39]. In both noise regimes the minimum mean-squared error using mixed measurements falls quickly with increasing m, while estimation errors using $\mathbf{z}$ were close that of the prior. The ℓ_1 recovery estimates performed similarly to the MMSE estimates when noise was low, but had very high average error in the noisier case. These experiments support the use of matrices with good compressed sensing properties and sparse recovery in probabilistic estimation with low measurement noise. Although good results were obtained using MMSE and MAP estimates, these would be impractical to compute for many sparse components, since computation complexity grows with the modes of the posterior distribution.

8.4 Conclusions

We have studied the mutual information of projection measurements computed using consensus in a sensor network, for a sparse source estimation problem. Mixing allows for robust estimation of the global random field using a small number of observation projections and estimates may be computed using compressed sensing recovery algorithms. We compared the mutual information of low dimension projections to that of simple sensor selection in the cases of known and unknown source locations. We showed that projections yield more informative measurements; proving this in the former case, and showing experimentally in the latter. The type of projections preferred in each are different with the

PCA matrix being optimal in the linear Gaussian case and random projections performing well in the unknown location case.

Chapter 9

Conclusions and Directions for Future Work

This thesis focuses on the problem of scalable distributed field inversion in sensor networks. We proposed a new approach to field inversion by consensus and compressed sensing (FICCS), particularly suited for reconstructing a monitored field. Our method uses a small number of communications to mix the network observations, and compressed sensing ℓ_1 optimizations to estimate sparse source parameters of the field.

This thesis contributes to several areas of active research. The first is in the modeling of problems of distributed estimation in sensor networks. For continuous fields generated by a small number of point sources, spatial quantization enables the linearization of the sensor observations model. Then distributed inference is cast as a distributed average consensus. Sensors mix their information with neighbors to compute a low dimensional projection of the global field. This low dimensional projection allows for the inversion of the field using sparse estimation techniques such as ℓ_1 penalized least squares and basis pursuit.

Quantization induces a special type of error, namely, model or structural errors. We provided an analysis of sparse estimation in the presence of modeling error induced by the spatial quantization of the source locations. This approach to model uncertainty in basis pursuit bounds on modeling error for single parameters, and incorporates these bounds into an expanded cone constraint for basis pursuit. The new estimator is called model-robust basis pursuit and is proven to be stable under appropriate conditions

on the coherence of the observation system and bounded sparse parameter magnitudes. We developed a methodology for accounting for structured model uncertainty in noise-aware basis pursuit that provides a framework for studying model error and alternative robust estimators.

We study the application of FICCS and robust sparse estimators to a set of motivating surveillance applications using inverse power law propagation. Using simulation testing we found that an estimator combining the cone constraint expansion and the ℓ_1-penalized objective function, which we called penalized model-robust basis pursuit (PMRBP), displayed robust estimation and detection performance in a variety of scenarios. The PMRBP estimator matches the performance of the ℓ_1 Penalty estimator, sometimes outperforming it in cases where using a cone constraint helps to restrict the solution. This combination also proved to have low sensitivity to tuning parameters and discretization size for the spatial lattice.

We extended FICCS to settings in which the sparse field has time-varying support. By using a recursive method of re-weighting the ℓ_1-norm of ℓ_1 penalized least squares, we incorporated information of the field from the previous time-step to allow for tracking of the sources with fewer measurements than would be required using single time samples.

Lastly, we considered the fundamental impact of mixing observations in a sensor network to distribute information. We characterized upper and lower bounds on the information of mixed measurements, showing mixing outperforms the alternative of selection of unmixed sensor information.

Future work in FICCS need not be restricted to using specific consensus and compressed sensing techniques. For example, if rather than a broadcast set-up, the network wanted to forward information to a fusion center, it could use spanning trees or other message passing techniques to efficiently compute low dimensional projections of the global field, which may then be used for inference.

Also, our approach to accounting for model quantization errors shows promise as a means of bridging the gap between compressed sensing theory and many practical sensing and surveillance applications. While we have focused on sensing models based on power-law propagation, there may be other sensing modalities with tradeoffs between coherence and quantization error. Future work will include the expansion of FICCS to additional time-varying detection settings involving updates close to the communication timescale.

Appendices

Appendix A

Scaling of mixing time with N in the 2D 8-Nearest-Neighbor Graph

Proof of Theorem 3.3.1: The steps are 1) compute the 2nd minimum and maximum eigenvalues of the Laplacian, $\lambda_2 = d - \lambda_{N-1}(A)$ and $\lambda_N = d - \lambda_1(A)$, as they depend on $g = \sqrt{N}$; and 2) characterize the behavior of the number of mixings k for large g.

$$k = \frac{\log(c)}{\log\left(\frac{\lambda_N - \lambda_2}{\lambda_N + \lambda_2}\right)} \tag{A.1}$$

First notice that the adjacency matrix A is circulant with rows that are cyclic shifted versions of a discrete function, composed of discrete Dirac δ_n. The rows are given by $x_n =$

$$\begin{aligned} &\delta_{n+\sqrt{N}+1} + \delta_{n+\sqrt{N}} + \delta_{n+\sqrt{N}-1} + \delta_{n+1} + \cdots \\ &\qquad \delta_{n-1} + \delta_{n-\sqrt{N}+1} + \delta_{n-\sqrt{N}} + \delta_{n-\sqrt{N}-1} \end{aligned} \tag{A.2}$$

A well-known property of circulant matrices is that they are diagonalized by the discrete Fourier matrix, and so computing the eigenvalues of A will be equivalent to finding the discrete Fourier transform coefficients of x_n. We denote these as X_m where the indices are taken to be in $\left\{\left\lceil\frac{-N}{2}\right\rceil, \left\lceil\frac{-N}{2}+1\right\rceil, \ldots, \left\lceil\frac{N}{2}\right\rceil\right\}$.

$$\begin{aligned} X_m = &\, 2\cos\left(\frac{2\pi m}{g^2}\right) + 2\cos\left(\frac{2\pi m}{g}\right) + \cdots \\ &+2\cos\left(\frac{2\pi(g-1)m}{g^2}\right) + 2\cos\left(\frac{2\pi(g+1)m}{g^2}\right) \end{aligned} \tag{A.3}$$

To find the second smallest and largest eigenvalues of the Laplacian matrix, we need to find the second largest and smallest eigenvalues of A. Local minima and maxima of X_m are located at $c\frac{\sqrt{N}}{2}$, with the global maximum at $m = 0$, second maximum at $m = \sqrt{N}$ and global minimum at $m = N/2$.[1]

The minimum eigenvalue of A is therefore

$$\begin{aligned} X_{\min} = X_{N/2} = 2\cos(\pi) + 2\cos\left(\pi\sqrt{N}\right) + \cdots \\ +2\cos\left(\pi(\sqrt{N}-1)\right) + 2\cos\left(\pi(\sqrt{N}+1)\right) \end{aligned} \tag{A.4}$$

We then simplify using the trigonometric sum formula $\cos(a) + \cos(b) = 2\cos\left(\frac{a+b}{2}\right)\cos\left(\frac{a-b}{2}\right)$ for the last two terms.

$$X_{\min} = -2 + 2\cos\left(\pi\sqrt{N}\right) + 4\cos\left(\pi\sqrt{N}\right)\cos(\pi) \tag{A.5}$$

Recall the initial assumption that $g = \sqrt{N}$ is an even integer, making $\cos\left(\pi\sqrt{N}\right) = -1$, and $X_{\min} = -4$.

The second largest eigenvalue of A is given by

$$\begin{aligned} X_{\sqrt{N}} = 2\cos\left(\frac{2\pi}{\sqrt{N}}\right) + 2\cos(2\pi) + \cdots \\ +2\cos\left(\frac{2\pi(\sqrt{N}-1)}{\sqrt{N}}\right) + 2\cos\left(\frac{2\pi(\sqrt{N}+1)}{\sqrt{N}}\right) \end{aligned} \tag{A.6}$$

We again simplify using the sum of the last two terms.

$$X_{\sqrt{N}} = 2 + 2\cos\left(\frac{2\pi}{\sqrt{N}}\right) + 4\cos\left(\frac{2\pi}{\sqrt{N}}\right) \tag{A.7}$$

$$= 2 + 6\cos\left(\frac{2\pi}{\sqrt{N}}\right) \tag{A.8}$$

We now conclude the eigenvalues of the Laplacian are $\lambda_2 = 6 - 6\cos\left(\frac{2\pi}{g}\right)$ and $\lambda_N = 12$. From these it is clear that λ_2 converges to 0 as $g \to \infty$. Next, we characterize how the rate of decrease in algebraic connectivity impacts the number of messages needed k.

$$k = \frac{\log(c)}{\log\left(\frac{\lambda_N - \lambda_2}{\lambda_N + \lambda_2}\right)} = \frac{\log(c)}{\log\left(\frac{6+6\cos\left(\frac{2\pi}{g}\right)}{18-6\cos\left(\frac{2\pi}{g}\right)}\right)} \tag{A.9}$$

[1] This remains to be proven, but is verified by extensive experiments.

For large g, the arguments of the cosines are close to 0. Applying a Taylor approximation $\cos(x) = 1 - \frac{1}{2}x^2 + O(x^4)$.

$$k = \frac{\log(c)}{\log\left(\frac{6+6\left(1-\frac{1}{2}\frac{4\pi^2}{g^2}+O(g^{-8})\right)}{18-6\left(1-\frac{1}{2}\frac{4\pi^2}{g^2}+O(g^{-8})\right)}\right)} \tag{A.10}$$

$$\approx \frac{\log(c)}{\log\left(\frac{12-\frac{12\pi^2}{g^2}}{12+\frac{12\pi^2}{g^2}}\right)} = \frac{\log(c)}{\log\left(\frac{1-\frac{\pi^2}{g^2}}{1+\frac{\pi^2}{g^2}}\right)} \tag{A.11}$$

Dividing out the fractions of polynomials $\frac{1-x}{1+x}$ yields $1 - 2x + 2x^2 - 2x^4 + \cdots$, which for $x = \pi^2/g^2$ and g large we further approximate k,

$$k \approx \frac{\log(c)}{\log\left(1 - 2\frac{\pi^2}{g^2} + O(g^{-4})\right)} \approx \frac{\log(c)}{\log\left(1 - 2\frac{\pi^2}{g^2}\right)} \tag{A.12}$$

Using the Mercator series for $\log(1+y) = y - \frac{y^2}{2} + O(y^3)$, for $-1 < y \leq 1$, and $y = -2\pi^2/g^2$,

$$\frac{\log(c)}{\log\left(1 - 2\frac{\pi^2}{g^2}\right)} \approx \frac{\log(c)}{-2\frac{\pi^2}{g^2} + O(g^{-4})} \approx -\frac{g^2}{2\pi^2}\log(c) \tag{A.13}$$

Recalling that $c < 1$, this is a positive scaling of g^2. We have shown the asymptotic growth rate in the number of mixings needed k is on the order of,

$$k \sim C_1 g^2 = C_1 N \tag{A.14}$$

Appendix B

Error Stability of Model Robust Basis Pursuit

This section presents the proof to the stability bound for model-robust basis pursuit (MRBP), Theorem 4.3.1, which follows the approach of [11].

Given the MRBP estimator, $\widehat{\mathbf{v}}$, given in (4.11), we want an upperbound on the mean-squared error. That is, find the maximum, over $\mathbf{v}_0$, $\mathbf{z}$, and $E = \begin{bmatrix} \mathbf{e}_1 & \cdots & \mathbf{e}_N \end{bmatrix}$ of

$$\begin{aligned} \text{maximize} \quad & \|\widehat{\mathbf{v}}_{1,\epsilon,\mathbf{g}}(\mathbf{y}) - \mathbf{v}_0\|_2 \\ \text{subj. to} \quad & \mathbf{y} = \Phi\left(\Psi_0 + E\right)\mathbf{v}_0 + \mathbf{z}, \quad \|\mathbf{z}\|_2 \leq \epsilon, \\ & \|\mathbf{v}_0\|_0 \leq S, |v_0(i)| \leq \bar{v}, \|\Phi\mathbf{e}_i\|_2 \leq \bar{g}\,\forall\, i \end{aligned} \tag{B.1}$$

for $\widehat{\mathbf{v}}(\mathbf{y})$ the solution of $(P_{1,\epsilon,\mathbf{g}})$, given in (4.11) and below.

$$\text{argmin}_{\mathbf{v}} \left\{ \|\mathbf{v}\|_1 : \begin{array}{c} \|\mathbf{y} - \Phi\Psi_0\mathbf{v}\|_2 \leq \epsilon + \mathbf{g}^T\left(\mathbf{v}_+ + \mathbf{v}_-\right) \\ \mathbf{v} = \mathbf{v}_+ - \mathbf{v}_-,\ \mathbf{v}_+ \geq \mathbf{0},\ \mathbf{v}_- \geq \mathbf{0} \end{array} \right\} \tag{B.2}$$

We recall that (4.8) shows $\mathbf{v}_0$ must be in the feasible set of $(P_{1,\epsilon,\mathbf{g}})$, so estimates must satisfy, $\|\widehat{\mathbf{v}}\|_1 \leq \|\mathbf{v}_0\|_1$. Since $\|\mathbf{v}_0\|_1 \leq S\bar{v}$, replacing $\mathbf{g}^T\mathbf{v}$ on the right hand side of the cone constraint with $S\bar{g}\bar{v}$ would expand the feasible set of (B.2), which only increases the maximum of (B.1). We alter the problem to

maximize $\|\widehat{\mathbf{v}} - \mathbf{v}_0\|_2$, over $\mathbf{v}_0$, $\mathbf{z}$ such that

$$\begin{aligned} \max \quad & \|\widehat{\mathbf{v}} - \mathbf{v}_0\|_2 \\ \text{s.t.} \quad & \hat{\mathbf{v}} = \operatorname{argmin} \{\|\mathbf{v}\|_1 : \|\mathbf{y} - \Phi\Psi_0\mathbf{v}\|_2 \leq \epsilon + S\bar{g}\bar{v}\}, \\ & \|\mathbf{z}\|_2 \leq \epsilon, \|\mathbf{v}_0\|_0 \leq S, \|\Phi\mathbf{e}_i\|_2 \leq \bar{g}, |v_0(i)| \leq \bar{v} \end{aligned} \tag{B.3}$$

Define the error vector $\mathbf{w} = \widehat{\mathbf{v}} - \mathbf{v}_0$, and denote $\mathbf{u} = \mathbf{v} - \mathbf{v}_0$, equivalently, we look for an upperbound on $\|\mathbf{w}\|_2^2$ for $\mathbf{w}$ given by $\operatorname{argmin}_{\mathbf{u}} \{\|\mathbf{v}_0 + \mathbf{u}\|_1\}$ over the set

$$\left\{ \begin{array}{c} \|\Phi\Psi_0\mathbf{u} - (\mathbf{z} + \Phi E\mathbf{v}_0)\|_2 \leq \epsilon + S\bar{g}\bar{v}, \\ \|\mathbf{z}\|_2 \leq \epsilon,\ \|\Phi E\mathbf{v}_0\|_2 \leq S\bar{g}\bar{v},\ \ \|\mathbf{v}_0\|_0 \leq S \end{array} \right\} \tag{B.4}$$

Since $\mathbf{w}$ is the minimizer of $\|\mathbf{w} + \mathbf{v}_0\|_1$ over the set of allowed $\mathbf{u}$, we know $\|\mathbf{w} + \mathbf{v}_0\|_1 \leq \|\mathbf{u} + \mathbf{v}_0\|_1$. Arguing that $\mathbf{u} = \mathbf{0}$ is feasible, $\|\mathbf{w} + \mathbf{v}_0\|_1 \leq \|\mathbf{v}_0\|_1$. Now, expand the problem to look for the maximum of $\|\mathbf{w}\|_2^2$ over this larger set satisfying that property,

$$\left\{ \begin{array}{c} \|\mathbf{v}_0 + \mathbf{w}\|_1 \leq \|\mathbf{v}_0\|_1, \\ \|\Phi\Psi_0\mathbf{w} - (\mathbf{z} + \Phi E\mathbf{v}_0)\|_2 \leq \epsilon + S\bar{g}\bar{v}, \\ \|\mathbf{z}\|_2 \leq \epsilon,\ \|\Phi E\mathbf{v}_0\|_2 \leq S\bar{g}\bar{v},\ \ \|\mathbf{v}_0\|_0 \leq S \end{array} \right\} \tag{B.5}$$

Let $\mathbf{f} = \mathbf{z} + \Phi E\mathbf{v}_0$ and $\epsilon + S\bar{g}\bar{v} = \epsilon'$, and denote $\Phi\Psi_0$ as the matrix A_0, making the maximization,

$$\max_{\mathbf{w},\mathbf{v}_0,\mathbf{f}} \left\{ \|\mathbf{w}\|_2^2 : \begin{array}{c} \|\mathbf{v}_0 + \mathbf{w}\|_1 \leq \|\mathbf{v}_0\|_1,\ \ \|\mathbf{v}_0\|_0 \leq S, \\ \|A_0\mathbf{w} - \mathbf{f}\|_2 \leq \epsilon',\ \|\mathbf{f}\|_2 \leq \epsilon' \end{array} \right\} \tag{B.6}$$

now an equivalent problem to that upperbounded in [11], the remaining steps are recapped for completeness. The following are successive expansions of the feasible set for $\mathbf{w}$. From $\|\mathbf{v}_0 + \mathbf{w}\|_1 \leq \|\mathbf{v}_0\|_1$, we can derive what Candès, Romberg, and Tao refer to in [10] as the *Tube Constraint*. We may write the error vector $\mathbf{w}$ in terms of the set of true support indices $\mathcal{T}$, as $\mathbf{w} = \mathbf{w}_{\mathcal{T}} + \mathbf{w}_{\mathcal{T}^c}$, where $\mathcal{T}^c$ is the complement; i.e., indices for which entries of $\mathbf{v}_0$ are zero.

$$\begin{aligned} 0 &\geq \|\mathbf{v}_0 + \mathbf{w}\|_1 - \|\mathbf{v}_0\|_1 && \text{(B.7)} \\ &= \|\mathbf{v}_0 + \mathbf{w}_{\mathcal{T}}\|_1 + \|\mathbf{w}_{\mathcal{T}^c}\|_1 - \|\mathbf{v}_0\|_1 && \text{(B.8)} \\ &\geq \|\mathbf{v}_0\|_1 - \|\mathbf{w}_{\mathcal{T}}\|_1 + \|\mathbf{w}_{\mathcal{T}^c}\|_1 - \|\mathbf{v}_0\|_1 && \text{(B.9)} \end{aligned}$$

where the last inequality follows by reverse triangle inequality, yielding the constraint on the norms of the error vector on versus off of the support indices, $\|\mathbf{w}_{\mathcal{T}}\|_1 \geq \|\mathbf{w}_{\mathcal{T}^c}\|_1$. Next, we remove the dependence on the additive error term, $\mathbf{f}$, by noting

$\{\mathbf{w} : \|A_0\mathbf{w} - \mathbf{f}\|_2 \leq \epsilon', \|\mathbf{f}\|_2 \leq \epsilon'\} \subseteq \{\mathbf{w} : \|A_0\mathbf{w}\|_2 \leq 2\epsilon'\}$

This yields the expanded maximization of $\|\mathbf{w}\|_2^2$ over the set,

$$\left\{ \|\mathbf{w}_{\mathcal{T}^c}\|_1 \leq \|\mathbf{w}_{\mathcal{T}}\|_1, \ \|A_0\mathbf{w}\|_2 \leq 2\epsilon', |\mathcal{T}| \leq S \right\} \tag{B.10}$$

Then, expand the constraint $\|A_0\mathbf{w}\|_2^2 \leq (2\epsilon')^2$ using the coherence of A_0, defined in (3.15), and denoting $A_0^T A_0 = G$. For an A_0 with unit-norm columns, the coherence μ is then equal to $\max_{i \neq j} |G_{i,j}|$.

$$\begin{aligned}
(2\epsilon')^2 &\geq \mathbf{w}^T A_0^T A_0 \mathbf{w} = \mathbf{w}^T\mathbf{w} + \mathbf{w}^T(G - I)\mathbf{w} && \text{(B.11)} \\
&\geq \mathbf{w}^T\mathbf{w} - |\mathbf{w}|^T |G - I| \, |\mathbf{w}| && \text{(B.12)} \\
&\geq \mathbf{w}^T\mathbf{w} - \mu|\mathbf{w}|^T \left(\mathbf{1}\mathbf{1}^T - I\right) |\mathbf{w}| && \text{(B.13)} \\
&= (1+\mu)\|\mathbf{w}\|_2^2 - \mu\|\mathbf{w}\|_1^2 && \text{(B.14)}
\end{aligned}$$

The expanded problem is to maximize $\|\mathbf{w}\|_2^2$ over the set

$$\left\{ \begin{array}{c} \|\mathbf{w}_{\mathcal{T}^c}\|_1 \leq \|\mathbf{w}_{\mathcal{T}}\|_1, \ |\mathcal{T}| \leq S, \\ (1+\mu)\|\mathbf{w}\|_2^2 - \mu\|\mathbf{w}\|_1^2 \leq (2\epsilon')^2 \end{array} \right\} \tag{B.15}$$

Then without loss of generality, assume the nonzeros are collected in the first S indices; i.e., $\mathcal{T} = \{1, 2, \ldots, S\}$. Denote $\mathbf{w} = \mathbf{w}_0 + \mathbf{w}_1$, where $\mathbf{w}_0$ has nonzeros only on $\mathcal{T}$ and $\mathbf{w}_1$ on $\mathcal{T}^c$. The constraints of (B.15) can then be represented by the equivalent constraints $\|\mathbf{w}_1\|_1 \leq \|\mathbf{w}_0\|_1$ and

$(1+\mu)\left(\|\mathbf{w}_0\|_2^2 + \|\mathbf{w}_1\|_2^2\right) - \mu\left(\|\mathbf{w}_0\|_1 + \|\mathbf{w}_1\|_1\right)^2 \leq (2\epsilon')^2$.

Next, define the constants

$$\|\mathbf{w}_0\|_1 = A, \quad \|\mathbf{w}_1\|_1 = B, \tag{B.16}$$

$$c_0 = \left(\frac{\|\mathbf{w}_0\|_2}{\|\mathbf{w}_0\|_1}\right)^2, \quad c_1 = \left(\frac{\|\mathbf{w}_1\|_2}{\|\mathbf{w}_1\|_1}\right)^2 \tag{B.17}$$

We seek to maximize $\|\mathbf{w}\|_2^2 = \|\mathbf{w}_0\|_2^2 + \|\mathbf{w}_1\|_2^2 = c_0 A^2 + c_1 B^1$, in the set

$$\left\{ \begin{array}{c} A \geq B, \ A \geq 0, B \geq 0, \frac{1}{S} \leq c_0 \leq 1, 0 \leq c_1 \leq 1, \\ (1+\mu)\left(c_0 A^2 + c_1 B^2\right) - \mu\left(A + B\right)^2 \leq 4\epsilon'^2 \end{array} \right\} \tag{B.18}$$

where the constraints on c_0 and c_1 are due to equivalence of norms. For $B = \rho A$, with $\rho \leq 1$

$$\begin{aligned} \text{max.} \quad & (c_0 + c_1\rho^2)A^2 \\ \text{s.t.} \quad & A \geq 0, \tfrac{1}{S} \leq c_0 \leq 1, 0 \leq c_1 \leq 1, 0 < \rho \leq 1, \\ & (1+\mu)(c_0 + c_1\rho^2)A^2 - \mu\left(1+\rho\right)^2 A^2 \leq 4\epsilon'^2 \end{aligned} \tag{B.19}$$

Then define $r = \frac{(1+\rho)^2}{c_0+\rho^2 c_1}$, and based on the limits of its components, $1 \leq r \leq 4S$. Also, $V = A^2(c_0 + c_1\rho^2)$, making the last constraint $(1+\mu)V - \mu r V \leq 4\epsilon'^2$ and the problem now

$$\max\left\{ V \,:\, V \leq \frac{4\epsilon'^2}{(1+\mu) - \mu r}, 1 \leq r \leq 4S \right\} \tag{B.20}$$

giving a maximum at $V = \frac{4\epsilon'^2}{1-\mu(4S-1)}$ the upperbound in (4.21).

Appendix C

Model Error for 1D System

This section derives the equation for model fit error of section 4.4.1. The one dimensional lattice of source and sensor locations spaced by T, for the inverse squared-distance propagation law.

We may use the the Fourier Domain representation to compute this,

$$E(j\omega) \;=\; X(j\omega) - Y(j\omega) \;=\; \pi e^{-|\omega|}\left(1 - e^{-j\omega t_0}\right) \tag{C.1}$$

since by Parsevals we know $\int |e(t)|^2 dt = \left(\int |E(j\omega)|^2 d\omega\right)/2\pi$.

Note that after sampling $x(t)$ and $y(t)$ by a pulse train $p(t) = \sum_{n\in\mathbb{Z}} \delta(nT)$, the Fourier transform of x and y now repeat in the frequency domain at every $\omega_s = 2\pi/T$. We restrict our attention to the square error over one period, i.e. from ω between $-\omega_s/2$ to $\omega_s/2$, rescaling by $\frac{1}{T}$.

Compute $\frac{1}{2\pi T}\int_{-\frac{\omega_s}{2}}^{\frac{\omega_s}{2}} |E(j\omega)|^2 d\omega$, where ω_s is the sampling frequency, which is $2\pi/T$.

$$\frac{1}{2\pi T}\int_{-\frac{\omega_s}{2}}^{\frac{\omega_s}{2}} |E(j\omega)|^2 d\omega \;=\; \frac{1}{2\pi T}\int_{-\pi/T}^{\pi/T} |\pi e^{-|\omega|}\left(1 - e^{-j\omega t_0}\right)|^2 d\omega \tag{C.2}$$

$$\;=\; \frac{\pi}{2T}\int_{-\pi/T}^{\pi/T} |e^{-|\omega|}\left(1 - e^{-j\omega t_0}\right)|^2 d\omega \tag{C.3}$$

$$\;=\; \frac{\pi}{2T}\left(\int_{-\pi/T}^{0} e^{2\omega}|\left(1 - e^{-j\omega t_0}\right)|^2 d\omega + \int_{0}^{\pi/T} e^{-2\omega}|\left(1 - e^{-j\omega t_0}\right)|^2 d\omega\right) \tag{C.4}$$

Simplifying the complex exponential,

$$|\left(1-e^{-j\omega t_0}\right)|^2 = |\left(1-\cos(t_0\omega)-j\sin(t_0\omega)\right)|^2 \quad \text{(C.5)}$$

$$= (1-\cos(t_0\omega))^2+(\sin(t_0\omega))^2 \quad \text{(C.6)}$$

$$= 1-2\cos(t_0\omega)+\cos(t_0\omega)^2+\sin(t_0\omega)^2 \quad \text{(C.7)}$$

$$= 2-2\cos(t_0\omega) \quad \text{(C.8)}$$

Splitting the integral, we have $\frac{1}{2\pi T}\int_{-\frac{\omega_s}{2}}^{\frac{\omega_s}{2}}|E(j\omega)|^2d\omega =$

$$= \frac{\pi}{2T}\left(\int_{-\pi/T}^{0}e^{2\omega}\left(2-2\cos(t_0\omega)\right)d\omega+\int_{0}^{\pi/T}e^{-2\omega}\left(2-2\cos(t_0\omega)\right)d\omega\right) \quad \text{(C.9)}$$

$$= \frac{\pi}{T}\int_{-\pi/T}^{0}e^{2\omega}\left(1-\cos(t_0\omega)\right)d\omega+\frac{\pi}{T}\int_{0}^{\pi/T}e^{-2\omega}\left(1-\cos(t_0\omega)\right)d\omega \quad \text{(C.10)}$$

Through a change of variable we can show $\int_{-\pi/T}^{0}e^{2\omega}\left(1-\cos(t_0\omega)\right)d\omega$ is equal to $\int_{0}^{\pi/T}e^{-2\omega}\left(1-\cos(t_0\omega)\right)d\omega$, (noting that cos is a symmetric function). This means this evaluates to...

$$\frac{1}{2\pi T}\int_{-\frac{\omega_s}{2}}^{\frac{\omega_s}{2}}|E(j\omega)|^2d\omega = 2\frac{\pi}{T}\int_{0}^{\pi/T}e^{-2\omega}\left(1-\cos(t_0\omega)\right)d\omega \quad \text{(C.11)}$$

$$= \frac{2\pi}{T}\int_{0}^{\pi/T}e^{-2\omega}d\omega-\frac{2\pi}{T}\int_{0}^{\pi/T}e^{-2\omega}\cos(t_0\omega)d\omega \quad \text{(C.12)}$$

$$2\int_{0}^{\pi/T}e^{-2\omega}d\omega = 2\left(\frac{e^{-2\omega}}{-2}\Big|_{0}^{\pi/T}\right)=\left(1-e^{-2\pi/T}\right) \quad \text{(C.13)}$$

We then evaluate the second term applying integration by parts twice. See section C.1 of this appendix for this. The resulting integral is...

$$2\int_{0}^{\pi/T}e^{-2\omega}\cos(t_0\omega)d\omega=\frac{2t_0}{(4+t_0^2)}e^{-2\pi/T}\sin(\pi t_0/T)+\frac{4}{(4+t_0^2)}-\frac{4}{(4+t_0^2)}\left(e^{-2\pi/T}\cos(\pi t_0/T)\right) \quad \text{(C.14)}$$

And so plugging that into the original expression we have... $\frac{1}{2\pi T}\int_{-\frac{\omega_s}{2}}^{\frac{\omega_s}{2}}|E(j\omega)|^2d\omega =$

$$=\frac{\pi}{T}\left(1-e^{-2\pi/T}\right)-\frac{\pi}{T}\left[\frac{2t_0}{(4+t_0^2)}e^{-2\pi/T}\sin(\pi t_0/T)+\frac{4}{(4+t_0^2)}-\frac{4}{(4+t_0^2)}\left(e^{-2\pi/T}\cos(\pi t_0/T)\right)\right] \quad \text{(C.15)}$$

$$=\frac{\pi}{T}\left[1-e^{-2\pi/T}-\frac{2t_0}{(4+t_0^2)}e^{-2\pi/T}\sin(\pi t_0/T)-\frac{4}{(4+t_0^2)}+\frac{4}{(4+t_0^2)}\left(e^{-2\pi/T}\cos(\pi t_0/T)\right)\right] \quad \text{(C.16)}$$

Simplify: $1 - \frac{4}{(4+t_0^2)} = \frac{t_0^2}{4+t_0^2}$.

$$= \frac{\pi}{T}\left[\left(\frac{t_0^2}{4+t_0^2}\right) + e^{\frac{-2\pi}{T}}\left(\frac{4}{(4+t_0^2)}\cos(\pi t_0/T) - \frac{2t_0}{(4+t_0^2)}\sin(\pi t_0/T) - 1\right)\right] \quad \text{(C.17)}$$

To recap, we have evaluated the total squared error of the sampled error function to be,

$\frac{1}{2\pi T}\int_0^{\pi/T} |E(j\omega)|^2 d\omega =$

$$= \frac{\pi}{T}\left[\left(\frac{t_0^2}{4+t_0^2}\right) + e^{\frac{-2\pi}{T}}\left(\frac{4}{(4+t_0^2)}\cos(\pi t_0/T) - \frac{2t_0}{(4+t_0^2)}\sin(\pi t_0/T) - 1\right)\right] \quad \text{(C.18)}$$

a function of both the sample rate T and the shift amount t_0.

C.1 Evaluating $2\int_0^{\pi/T} e^{-2\omega}\cos(t_0\omega)d\omega$ by parts

Compute $2\int_0^{\pi/T} e^{-2\omega}\cos(t_0\omega)d\omega$ using integration by parts (let $e^{-2\omega} = U$ and $\cos(t_0\omega)d\omega = dV$, so $\int UdV = UV - \int VdU$, and $dU = -2e^{-2\omega}d\omega$ and $V = \sin(t_0\omega)/t_0$).

$$2\int_0^{\pi/T} e^{-2\omega}\cos(t_0\omega)d\omega \quad = \quad 2\left(\frac{e^{-2\omega}\sin(t_0\omega)}{t_0}\Big|_0^{\pi/T} + \frac{2}{t_0}\int_0^{\pi/T} e^{-2\omega}\sin(t_0\omega)d\omega\right) \quad \text{(C.19)}$$

Evaluate $2\left(\frac{e^{-2\omega}\sin(t_0\omega)}{t_0}\Big|_0^{\pi/T}\right) = \frac{2}{t_0}\left(e^{-2\pi/T}\sin(\pi t_0/T)\right)$, so

$$2\int_0^{\pi/T} e^{-2\omega}\cos(t_0\omega)d\omega \quad = \quad \frac{2}{t_0}e^{-2\pi/T}\sin(\pi t_0/T) + \frac{4}{t_0}\int_0^{\pi/T} e^{-2\omega}\sin(t_0\omega)d\omega \quad \text{(C.20)}$$

Now apply integration by parts to $\int_0^{\pi/T} e^{-2\omega}\sin(t_0\omega)d\omega$, where $e^{-2\omega} = U$ and $\sin(t_0\omega)d\omega = dV$, so $dU = -2e^{-2\omega}d\omega$ and $V = -\cos(t_0\omega)/t_0$

$$\int_0^{\pi/T} e^{-2\omega}\sin(t_0\omega)d\omega \quad = \quad \left(\frac{-e^{-2\omega}\cos(t_0\omega)}{t_0}\Big|_0^{\pi/T} - \frac{2}{t_0}\int_0^{\pi/T} e^{-2\omega}\cos(t_0\omega)d\omega\right) \quad \text{(C.21)}$$

Evaluate $\frac{-e^{-2\omega}\cos(t_0\omega)}{t_0}\Big|_0^{\pi/T} = \frac{1}{t_0} - \frac{1}{t_0}\left(e^{-2\pi/T}\cos(\pi t_0/T)\right)$, so

$$\int_0^{\pi/T} e^{-2\omega}\sin(t_0\omega)d\omega \quad = \quad \frac{1}{t_0} - \frac{1}{t_0}\left(e^{-2\pi/T}\cos(\pi t_0/T)\right) - \frac{2}{t_0}\int_0^{\pi/T} e^{-2\omega}\cos(t_0\omega)d\omega \quad \text{(C.22)}$$

Plugging this back into (C.19), $2\int_0^{\pi/T} e^{-2\omega}\cos(t_0\omega)d\omega = \frac{2}{t_0}e^{-2\pi/T}\sin(\pi t_0/T)+\frac{4}{t_0}\int_0^{\pi/T} e^{-2\omega}\sin(t_0\omega)d\omega$,

$$= \frac{2}{t_0}e^{-2\pi/T}\sin(\pi t_0/T) + \frac{4}{t_0}\left[\frac{1}{t_0} - \frac{1}{t_0}\left(e^{-2\pi/T}\cos(\pi t_0/T)\right) - \frac{2}{t_0}\int_0^{\pi/T} e^{-2\omega}\cos(t_0\omega)d\omega\right] \quad \text{(C.23)}$$

$$= \frac{2}{t_0}e^{-2\pi/T}\sin(\pi t_0/T) + \frac{4}{t_0^2} - \frac{4}{t_0^2}\left(e^{-2\pi/T}\cos(\pi t_0/T)\right) - \frac{8}{t_0^2}\int_0^{\pi/T} e^{-2\omega}\cos(t_0\omega)d\omega \quad \text{(C.24)}$$

Bring the integral to the left side...

$$\left(2+\frac{8}{t_0^2}\right)\int_0^{\pi/T} e^{-2\omega}\cos(t_0\omega)d\omega = \frac{2}{t_0}e^{-2\pi/T}\sin(\pi t_0/T) + \frac{4}{t_0^2} - \frac{4}{t_0^2}\left(e^{-2\pi/T}\cos(\pi t_0/T)\right) \quad \text{(C.25)}$$

$$2\int_0^{\pi/T} e^{-2\omega}\cos(t_0\omega)d\omega = \left(\frac{t_0^2}{4+t_0^2}\right)\left[\frac{2}{t_0}e^{-2\pi/T}\sin(\pi t_0/T) + \frac{4}{t_0^2} - \frac{4}{t_0^2}\left(e^{-2\pi/T}\cos(\pi t_0/T)\right)\right] \quad \text{(C.26)}$$

$$= \frac{2t_0}{(4+t_0^2)}e^{-2\pi/T}\sin(\pi t_0/T) + \frac{4}{(4+t_0^2)} - \frac{4}{(4+t_0^2)}\left(e^{-2\pi/T}\cos(\pi t_0/T)\right) \quad \text{(C.27)}$$

Appendix D

Coherence for 1D System

We derive the expression for mutual coherence $\mu(\Psi)$, for matrix Ψ of (2.6), as it depends on lattice spacing parameter δ, herein refered to as T, for the a one dimensional formulation of the source localization problem. We assume inverse square power law propagation, as given in (2.9), and an infinite lattice of sensors sample the source field, regularly spaced in a grid that coincides with the source lattice.

The mutual coherence of a matrix $A \in \mathbb{R}^{m \times L}$ is defined by

$$\mu(A) \quad = \quad \max_{i \neq j} \frac{|\langle \mathbf{a}_i, \mathbf{a}_j \rangle|}{\|\mathbf{a}_i\|_2 \cdot \|\mathbf{a}_j\|_2} \tag{D.1}$$

where $\mathbf{a}_i$ and $\mathbf{a}_j$ denote distinct columns of the matrix A.

For the square propagation system Ψ, each column is the same signal shifted by T. The maximum innerproduct will be attained by any two adjacent columns, so without loss of generality, we will consider $\mathbf{a}_i$ to be the basis centered at the origin and $\mathbf{a}_j$ centered at T.

First we compute the energy of the signal, to evaluate the denominator of (D.1). Then we compute the absolute inner product of the propagation function with a version of itself shifted by T.

D.1 Energy of $x(nT)$ on the Infinite 1D Lattice

The discrete basis functions are defined by an inverse square propagation law,

$$x_n \quad = \quad \frac{1}{d_{0,n}^2 + 1} = \frac{1}{(Tn)^2 + 1}, \qquad \text{for } n = -\infty, \ \ldots, \ -1, \ 0, \ 1, \ \ldots, \ \infty \tag{D.2}$$

The energy of this signal is given by,

$$\|\mathbf{x}\|_2^2 \;=\; \sum_n |x_n|^2 \;=\; \sum_n \frac{1}{(T^2n^2+1)^2} \tag{D.3}$$

Rather than compute this infinite sum, we apply Parsevals Theorem, which relates the energy of the signal to the energy of its Fourier coefficients.

$$\sum_{n=-\infty}^{\infty} |x_n|^2 \;=\; \frac{1}{2\pi}\int_{-\pi}^{\pi} |X(e^{j\omega})|^2 d\omega \;=\; \frac{1}{T}\frac{1}{2\pi}\int_{-\pi/T}^{\pi/T} |X(j\omega)|^2 d\omega \tag{D.4}$$

The r.h.s. of (D.4) relates the discrete time Fourier Transform coefficients to the continuous time Fourier transform of the signal (before regular sampling with period T). The Fourier transform of the continuous signal $x(t) = \frac{1}{t^2+1}$ is a well known form.

$$X(j\omega) \;=\; \pi e^{-|\omega|} \tag{D.5}$$

The squared energy of one basis function is then equal to,

$$\|\mathbf{x}\|_2^2 \;=\; \frac{1}{T}\frac{1}{2\pi}\int_{-\pi/T}^{\pi/T} \left|\pi e^{-|\omega|}\right|^2 d\omega \tag{D.6}$$

$$\;=\; \frac{\pi}{2T}\int_{-\pi/T}^{\pi/T} e^{-2|\omega|} d\omega \tag{D.7}$$

$$\;=\; \frac{\pi}{2T}\int_{-\pi/T}^{0} e^{2\omega} d\omega + \frac{\pi}{2T}\int_{0}^{\pi/T} e^{-2\omega} d\omega \tag{D.8}$$

Using $\int_a^b e^{-2\omega} d\omega = -\frac{1}{2} e^{-2\omega}\big|_a^b$, we see

$$\int_0^{\pi/T} e^{-2\omega} d\omega \;=\; \frac{1}{2} - \frac{1}{2}e^{-2\pi/T} \tag{D.9}$$

$$\text{and } \int_{-\pi/T}^{0} e^{-2\omega} d\omega \;=\; \frac{1}{2} - \frac{1}{2}e^{-2\pi/T} \tag{D.10}$$

The squared energy is therefore,

$$\|\mathbf{x}\|_2^2 \;=\; \frac{\pi}{2T}\left(1 - e^{-2\pi/T}\right) \tag{D.11}$$

We note that this assumes no aliasing due to sampling, which, for this propagation function, effectively means $0 < T < 1/2$.

D.2 Inner product of adjacent bases, $x(nT)$ and $x((n-1)T)$

Since we want to know the maximum inner product, we consider without loss of generality, the adjacent basis functions, centered at 0 and 1; i.e.,

$$\mathbf{a}_1 = \sum_{n\in\mathbb{Z}} \frac{1}{(Tn)^2+1}\delta(n) \tag{D.12}$$

$$\mathbf{a}_2 = \sum_{n\in\mathbb{Z}} \frac{1}{(T(n-1))^2+1}\delta(n) \tag{D.13}$$

$$\tag{D.14}$$

$$\langle \mathbf{a}_1, \mathbf{a}_2\rangle = \sum_{n\in\mathbb{Z}} \frac{1}{(Tn)^2+1}\frac{1}{(T(n-1))^2+1} \tag{D.15}$$

Since these signals are non-negative, their inner product must be positive, so $|\langle \mathbf{a}_1, \mathbf{a}_2\rangle| = \langle \mathbf{a}_1, \mathbf{a}_2\rangle$.

To make this computation easier we note that one signal is a shifted version of the other, $\mathbf{a}_2[n] = \mathbf{a}_1[n-1]$. The inner product of these two is equivalent to the value of the convolution of the two signals at the zero position.

$$\langle \mathbf{a}_1, \mathbf{a}_2\rangle = (\mathbf{a}_1[n] * \mathbf{a}_2[n])\,[0] \tag{D.16}$$

Convolution in time is equivalent to multiplication in the frequency domain.

$$\text{For} \quad \mathbf{n} = x_1(t) * x_2(t), \tag{D.17}$$

$$N(j\omega) = X(j\omega)\cdot X(j\omega)e^{-j\omega T} \tag{D.18}$$

$$= X(j\omega)^2 e^{-j\omega T} \tag{D.19}$$

where $x_1(t)$ and $x_2(t)$ are continuous versions of $\mathbf{a}_1[n]$ and $\mathbf{a}_2[n]$. The value of the convolution at time 0 can then be found via the inverse Fourier Transform,

$$n(0) = \frac{1}{2\pi}\int N(j\omega)e^{j\omega t}d\omega\Big|_{t=0} = \frac{1}{2\pi}\int N(j\omega)d\omega \tag{D.20}$$

$$= \frac{1}{2\pi}\int X(j\omega)^2 e^{-j\omega T}d\omega \tag{D.21}$$

Applying similar relationship between the sampled signal and the continuous signals as in the previous section we now write the inner product as,

$$\langle \mathbf{a}_1, \mathbf{a}_2\rangle = \frac{1}{T}\frac{1}{2\pi}\int_{-\pi/T}^{\pi/T} X(j\omega)^2 e^{-j\omega T}d\omega \tag{D.22}$$

Recall that the Fourier transform of our continuous propagation function is $X(jw) = \pi e^{-|\omega|}$.

$$\begin{aligned}
\langle \mathbf{a}_1, \mathbf{a}_2 \rangle &= \frac{1}{T}\frac{1}{2\pi}\int_{-\pi/T}^{\pi/T}\left(\pi e^{-|\omega|}\right)^2 e^{-j\omega T}d\omega && \text{(D.23)}\\
&= \frac{\pi}{2T}\int_{-\pi/T}^{\pi/T} e^{-2|\omega|}e^{-j\omega T}d\omega && \text{(D.24)}\\
&= \frac{\pi}{2T}\int_{-\pi/T}^{\pi/T} e^{-2|\omega|-j\omega T}d\omega && \text{(D.25)}\\
&= \frac{\pi}{2T}\left[\int_{0}^{\pi/T} e^{-2\omega-j\omega T}d\omega + \int_{-\pi/T}^{0} e^{2\omega-j\omega T}d\omega\right] && \text{(D.26)}\\
&= \frac{\pi}{2T}\left[\int_{0}^{\pi/T} e^{(-2-jT)\omega}d\omega + \int_{-\pi/T}^{0} e^{(2-jT)\omega}d\omega\right] && \text{(D.27)}
\end{aligned}$$

$$\begin{aligned}
\int_{0}^{\pi/T} e^{(-2-jT)\omega}d\omega &= \frac{-1}{(2+jT)}e^{(-2-jT)\omega}\Big|_{0}^{\pi/T} && \text{(D.28)}\\
&= \frac{-1}{(2+jT)}e^{(-2-jT)\pi/T} - \frac{-1}{(2+jT)} && \text{(D.29)}\\
&= \frac{1}{(2+jT)} - \frac{1}{(2+jT)}e^{(-2\pi/T-j\pi)} && \text{(D.30)}
\end{aligned}$$

Since $e^{-j\pi} = -1$,

$$\int_{0}^{\pi/T} e^{(-2-jT)\omega}d\omega = \frac{1}{(2+jT)} + \frac{1}{(2+jT)}e^{-2\pi/T} = \frac{1}{(2+jT)}\left(1+e^{-2\pi/T}\right) \quad \text{(D.31)}$$

Similarly for the second term:

$$\begin{aligned}
\int_{-\pi/T}^{0} e^{(2-jT)\omega}d\omega &= \frac{1}{(2-jT)}e^{(2-jT)\omega}\Big|_{-\pi/T}^{0} && \text{(D.32)}\\
&= \frac{1}{(2-jT)} - \frac{1}{(2-jT)}e^{(2-jT)(-\pi/T)} && \text{(D.33)}\\
&= \frac{1}{(2-jT)} - \frac{1}{(2-jT)}e^{(-2\pi/T+j\pi)} && \text{(D.34)}
\end{aligned}$$

Since $e^{j\pi} = -1$,

$$\int_{-\pi/T}^{0} e^{(2-jT)\omega}d\omega = \frac{1}{(2-jT)} + \frac{1}{(2-jT)}e^{-2\pi/T} = \frac{1}{(2-jT)}\left(1+e^{-2\pi/T}\right) \quad \text{(D.35)}$$

Plugging these back into (D.27), we have...

$$\langle \mathbf{a}_1, \mathbf{a}_2 \rangle = \frac{\pi}{2T}\left[\frac{1}{(2+jT)}\left(1+e^{-2\pi/T}\right) + \frac{1}{(2-jT)}\left(1+e^{-2\pi/T}\right)\right] \quad \text{(D.36)}$$

Simplify $\frac{1}{(2+jT)} + \frac{1}{(2-jT)} = \frac{2-jT+2+jT}{(2+jT)(2-jT)} = \frac{4}{4+T^2}$

$$\begin{aligned}
\langle \mathbf{a}_1, \mathbf{a}_2 \rangle &= \frac{\pi}{2T}\left[\frac{4}{4+T^2} + \frac{e^{-2\pi/T}}{(2+jT)} + \frac{e^{-2\pi/T}}{(2-jT)}\right] && \text{(D.37)}\\
&= \frac{\pi}{2T}\left[\frac{4}{4+T^2} + \left(\frac{1}{(2+jT)} + \frac{1}{(2-jT)}\right)e^{-2\pi/T}\right] && \text{(D.38)}\\
&= \frac{\pi}{2T}\left[\frac{4}{4+T^2} + \frac{4e^{-2\pi/T}}{4+T^2}\right] && \text{(D.39)}\\
&= \frac{\pi}{2T}\left(\frac{4}{4+T^2}\right)\left(1+e^{-2\pi/T}\right) && \text{(D.40)}
\end{aligned}$$

Again this expression is not valid when there is significant aliasing, and applies to the interval $0 < T < 1/2$.

D.3 Expression for μ

Putting the energy and innerproduct expressions together we have,

$$\begin{aligned}
\mu &= \frac{|\langle \mathbf{a}_i, \mathbf{a}_j \rangle|}{\|\mathbf{a}_i\|_2 \cdot \|\mathbf{a}_j\|_2} = \frac{\frac{\pi}{2T}\left(\frac{4}{4+T^2}\right)\left(1+e^{-2\pi/T}\right)}{\frac{\pi}{2T}\left(1-e^{-2\pi/T}\right)} && \text{(D.41)}\\
&= \left(\frac{4}{4+T^2}\right)\frac{\left(1+e^{-2\pi/T}\right)}{\left(1-e^{-2\pi/T}\right)} && \text{(D.42)}
\end{aligned}$$

Bibliography

[1] V. V. Titov, F. I. Gonzalez, E. N. Bernard, M. C. Eble, H. O. Mofjeld, J. C. Newman, and A. J. Venturato, "Real-time tsunami forecasting: Challenges and solutions," *Natural Hazards*, vol. 35, pp. 35–41, 2005, 10.1007/s11069-004-2403-3. [Online]. Available: http://dx.doi.org/10.1007/s11069-004-2403-3

[2] D. Percival, D. Denbo, M. Eble, E. Gica, H. Mofjeld, M. Spillane, L. Tang, and V. Titov, "Extraction of tsunami source parameters via inversion of DART buoy data," *NOAA Tech. Memo. OAR PMEL*, vol. 144, 2010. [Online]. Available: http://www.pmel.noaa.gov/pubs/PDF/perc3134/perc3134.pdf

[3] J. N. Tsitsiklis, *Decentralized Detection*, H. V. Poor and J. B. Thomas, Eds. Cambridge: JAI Press, 1993.

[4] M. Rabbat and R. Nowak, "Decentralized source localization and tracking," *29th IEEE Int'l Conf. on Acoustics, Speech and Sig. Proc. (ICASSP)*, 2004.

[5] M. Rabbat, R. Nowak, and J. Bucklew, "Robust decentralized source localization via averaging," *30th IEEE Int'l Conf. on Acoustics, Speech and Sig. Proc. (ICASSP)*, 2005.

[6] D. Blatt and A. O. Hero, "Energy based sensor network source localization via projection onto convex sets (pocs)," *IEEE Trans. on Signal Processing*, vol. 54, no. 9, pp. 3614–3619, Sep. 2006.

[7] D. O. North, "An analysis of the factors which determine signal/noise discrimination in pulsed carrier systems," RCA Labs., Princeton, NJ, Rep. PTR-6C, Tech. Rep., 1943.

[8] W. Bajwa, R. Calderbank, and S. Jafarpour, "Why Gabor frames? two fundamental measures of coherence and their role in model selection," *Journal of Communications and Networks*, vol. 12, no. 4, pp. 289–307, Aug. 2010.

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

[10] E. J. Candès, J. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Communications on Pure and Applied Mathematics*, vol. 59, no. 8, pp. 1207–1233, Aug. 2006.

[11] D. L. Donoho, M. Elad, and V. Temlyakov, "Stable recovery of sparse overcomplete representations in the presence of noise," *IEEE Trans. Info. Theory*, vol. 52, no. 1, pp. 6–18, Jan. 2006.

[12] A. M. Bruckstein, D. L. Donoho, and M. Elad, "From sparse solutions of systems of equations to sparse modeling of signals and images," *SIAM Review*, vol. 51, no. 1, pp. 34–81, 2009.

[13] E. J. Candès and T. Tao, "Decoding by linear programming," *IEEE Trans. on Info. Theory*, vol. 51, no. 12, pp. 4203–4215, 2005.

[14] E. J. Candès, "Compressive sampling," *Int. Congress of Mathematics*, vol. 3, no. 7, pp. 1433–1452, 2006.

[15] D. L. Donoho and X. Huo, "Uncertainty principles and ideal atomic decomposition," *IEEE Trans. on Info. Theory*, vol. 47, no. 7, pp. 2845–2862, 2001.

[16] L. E. Kinsler, A. R. Frey, A. B. Coppens, and J. V. Sanders, *Fundamentals of Acoustics*, 4th ed. John Wiley & Sons, Inc., 2000.

[17] M. Rabbat, J. Haupt, A. Singh, and R. Nowak, "Decentralized compression and predistribution via randomized gossiping," *5th Int. Conf. on Info. Proc. in Snsr. Net.'s*, pp. 51–59, Apr. 2006.

[18] A. Dimakis, S. Kar, J. M. F. Moura, M. Rabbat, and A. Scaglione, "Gossip algorithms for distributed signal processing," *Proc. of the IEEE*, vol. 98, no. 11, pp. 1847 –1864, Nov. 2010.

[19] A. Schmidt and J. M. F. Moura, "Field inversion by consensus and compressed sensing," in *34th IEEE Int'l Conf. on Acoustics, Speech and Sig. Proc. (ICASSP), Taipei*, Apr. 2009, pp. 2417–2420.

[20] ——, "A distributed sensor fusion algorithm for the inversion of sparse fields," *43rd Asilomar Conf. Signals, Systems and Computers, Pacific Grove, CA*, pp. 1332–1336, Nov. 2009.

[21] U. A. Khan, S. Kar, and J. M. F. Moura, "Distributed sensor localization in random environments using minimal number of anchor nodes," *IEEE Trans. on Signal Processing*, vol. 57, no. 5, pp. 2000–2016, May 2009.

[22] U. Khan, S. Kar, and J. M. F. Moura, "DILAND: An algorithm for distributed sensor localization with noisy distance measurements," *IEEE Trans. Signal Proc.*, vol. 58, no. 3, pp. 1940–1947, Mar. 2010.

[23] S. Boyd, A. Ghosh, B. Prabhakar, and D. Shah, "Gossip algorithms: Design, analysis and applications," *Proc. IEEE Infocomm, Miami*, vol. 3, pp. 1653–1664, Mar. 2005.

[24] J. Tropp, "Just relax: convex programming methods for identifying sparse signals in noise," *IEEE Trans. Info. Theory*, vol. 52, no. 3, pp. 1030–1051, Mar. 2006.

[25] A. K. Fletcher, S. Rangan, and V. K. Goyal, "Necessary and sufficient conditions on sparsity pattern recovery," *IEEE Trans. on Info. Theory*, vol. 55, no. 12, pp. 5758–5772, Dec. 2009.

[26] J. A. Tropp and A. C. Gilbert., "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. Info. Theory*, vol. 53, no. 12, pp. 1030–1051, Dec. 2007.

[27] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.

[28] M. A. T. Figueiredo, R. D. Nowak, and S. J. Wright, "Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems," *IEEE J. of Sel. Topics in Sig. Proc.*, vol. 1, no. 4, pp. 586–597, Dec. 2007.

[29] A. d'Aspremont and L. El Ghaoui, "Testing the nullspace property using semidefinite programming," *Mathematical Programming*, vol. 127, no. 1, pp. 123–144, 2011.

[30] M. Püschel and J. Kovacevic, "Real, tight frames maximally robust to erasures," *Data Compression Conference (DCC)*, pp. 63–72, 2005.

[31] T. Strohmer and R. W. Heath, "Grassmannian frames with applications to coding and communication," *Appl. Comput. Harmon. Anal.*, vol. 14, no. 3, pp. 257–275, May 2003.

[32] W. Xu and B. Hassibi, "Compressed sensing over the grassmann manifold: A unified analytical framework," *46th Allerton Conf. on Comm., Control, and Computing*, pp. 562–567, Sep. 2008.

[33] S. Kar, S. A. Aldosari, and J. M. F. Moura, "Topology for distributed inference on graphs," *IEEE Trans. on Signal Processing*, vol. 56, no. 6, pp. 2609–2613, Jun. 2008.

[34] L. Xiao and S. Boyd, "Fast linear iterations for distributed averaging," *Proceedings of IEEE Conf.*

on Decision and Control, pp. 4997–5002, Dec. 2003.

[35] S. Boyd, P. Diaconis, P. Parrilo, and L. Xiao, "Fastest mixing Markov chain on graphs with symmetries," *SIAM J. Optim.*, vol. 20, no. 2, pp. 792–819, 2009.

[36] A. G. Dimakis, "Geographic gossip: Efficient aggregation for sensor networks," *In ACM/IEEE Symposium on Information Processing in Sensor Networks*, pp. 69–76, 2006.

[37] F. Bénézit, A. G. Dimakis, P. Thiran, and M. Vetterli, "Order-optimal consensus through randomized path averaging," *IEEE Trans. on Info. Theory*, vol. 56, no. 10, pp. 5150–5167, 2010.

[38] S. Kar and J. M. F. Moura, "Sensor networks with random links: Topology design for distributed consensus," *IEEE Transactions on Signal Processing*, vol. 56, no. 7, pp. 3315–3326, Jul. 2008.

[39] CVX Research, Inc., "CVX: Matlab software for disciplined convex programming, version 2.0 beta," http://cvxr.com/cvx, Sep. 2012.

[40] E. van den Berg and M. P. Friedlander, "Probing the Pareto frontier for basis pursuit solutions," *SIAM J. on Scientific Computing*, vol. 31, no. 2, pp. 890–912, 2008.

[41] ——, "SPGL1: A solver for large-scale sparse reconstruction," Jun 2007, http://www.cs.ubc.ca/labs/scl/spgl1.

[42] M. Vetterli, P. Marziliano, and T. Blu, "Sampling signals with finite rate of innovation," *IEEE Transactions on Signal Processing*, vol. 50, no. 6, pp. 1417–1428, Jun. 2002.

[43] G. Tang, B. Bhaskar, P. Shah, and B. Recht, "Compressive sensing off the grid," in *2012 50th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, Oct. 2012, pp. 778–785.

[44] H. Zhu, G. Leus, and G. B. Giannakis, "Sparsity-cognizant total least-squares for perturbed compressive sampling," *IEEE Trans. on Signal Processing*, vol. 59, no. 5, pp. 2002–2016, May 2011.

[45] M. A. Herman and T. Strohmer, "General deviants: An analysis of perturbations in compressed sensing," *J. of Sel. Topics in Sig. Proc.: Special Issue on Compr. Sensing*, vol. 4, no. 2, pp. 342–349, Apr. 2010.

[46] P. Boufounos, V. Cevher, A. Gilbert, Y. Li, and M. Strauss, "Whats the frequency, kenneth?: Sublinear fourier sampling off the grid," in *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*, ser. Lecture Notes in Computer Science, A. Gupta, K. Jansen, J. Rolim, and R. Servedio, Eds. Springer Berlin Heidelberg, 2012, vol. 7408, pp. 61–72.

[47] C. Ekanadham, D. Tranchina, and E. P. Simoncelli, "Recovery of sparse translation-invariant signals with continuous basis pursuit," *IEEE Transactions on Signal Processing*, vol. 59, no. 10, pp. 4735–4744, Oct.

[48] A. Schmidt and J. M. F. Moura, "Scalable sensor network field reconstruction with robust basis pursuit," *submitted to IEEE Transactions on Signal Processing*, Apr. 2013.

[49] ——, "Distributed field reconstruction with model-robust basis pursuit," *37th IEEE Int'l Conf. on Acoustics, Speech and Sig. Proc. (ICASSP), Kyoto*, pp. 2673 – 2676, Mar. 2012.

[50] M. Skolnik, *Introduction to Radar Systems*, 3rd ed. McGraw-Hill Science, 2002.

[51] M. A. Herman and T. Strohmer, "High-resolution radar via compressed sensing," *IEEE Trans. on Signal Processing*, vol. 57, no. 6, pp. 2275–2284, Jun. 2009.

[52] L. M. Brekhovskikh and Y. P. Lysanov, *Fundamentals of Ocean Acoustics (Modern Acoustics and Signal Processing)*, 3rd ed. Springer, 2003.

[53] J. B. Harley and J. M. Moura, "Sparse recovery of the multimodal and dispersive characteristics of lamb waves," *Journal of the Acoustical Society of America*, vol. 133, no. 5, pp. 2732–2745, 2013.

[54] J. B. Harley and J. M. F. Moura, "High resolution localization with Lamb wave sparse wavenumber analysis," *International Workshop on Structural Health Monitoring*, Sep. 2013.

[55] M. Grant and S. Boyd, "Graph implementations for nonsmooth convex programs," in *Recent Advances in Learning and Control*, V. Blondel, S. Boyd, and H. Kimura, Eds. Springer-Verlag Limited, 2008, pp. 95–110, http://stanford.edu/~boyd/graph_dcp.html.

[56] H. Zou, "The adaptive LASSO and its oracle properties," *J. of the American Statistical Assoc.*, vol. 101, no. 476, pp. 1418–1429, 2006.

[57] I. T. Jolliffe, *Principal Component Analysis*, 2nd ed. Springer, 2002.

[58] K. Diamantaras and K. Hornik, "Noisy principal component analysis," in *Inst. of Measurement Science, Slovak Acad. of Sci.*, 1993, pp. 25–33.

[59] A. Krause, A. Singh, and C. Guestrin, "Near-optimal sensor placements in gaussian processes: Theory, efficient algorithms and empirical studies," *Jrnl of Machine Learning Research*, vol. 9, pp. 235–284, Jan. 2008.

[60] S. Joshi and S. Boyd, "Sensor selection via convex optimization," *IEEE Transactions on Signal Processing*, 2008.

[61] C. kwong Li and R. Mathias, "Interlacing inequalities for totally nonnegative matrices," *Linear Algebra Appl.*, vol. 341, pp. 35–44, 2002.

[62] M. F. Huber, T. Bailey, H. Durrant-Whyte, and U. D. Hanebeck, "On entropy approximation for gaussian mixture random vectors," in *Multisensor Fusion and Integration for Intelligent Systems, 2008. MFI 2008. IEEE Int'l Conference on*, 2008, pp. 181–188.