

**ADVANCES IN COMPUTATIONAL AND STATISTICAL INVERSE
PROBLEMS**

A Thesis
Submitted to the Faculty
in partial fulfillment of the requirements for the
degree of

Doctor of Philosophy

in

Mathematics

by Dylan Patrick Green

Guarini School of Graduate and Advanced Studies
Dartmouth College
Hanover, New Hampshire

June 2024

Examining Committee:

Anne Gelb, Chair

Geoffrey P. Luke

Yoonsang Lee

Theresa Scarnati

F. Jon Kull, Ph.D.

Dean of the Guarini School of Graduate and Advanced Studies

Abstract

Inverse problems are prevalent in many fields of science and engineering, such as signal processing and medical imaging. In such problems, indirect data are used to recover information regarding some unknown parameters of interest. When these problems fail to be well-posed, the original problems must be modified to include additional constraints or optimization terms, giving rise to so-called *regularization* techniques. Classical methods for solving inverse problems are often deterministic and focus on finding point estimates for the unknowns. Some newer methods approach the solving of inverse problems by instead casting them in a statistical framework, allowing for novel point estimate approaches and for the recovery of uncertainty information. In this dissertation, we first use a deterministic approach in the context of a medical imaging application to reconstruct volumetric images of blood vessels while enforcing sparsity in the edge domain. We then propose and investigate methods for the statistical inference of complex-valued signals as well as techniques for volumetric reconstruction using complex-valued synthetic aperture radar data.

This work is partially supported by NSF DMS grant #1912685, AFOSR grant #F9550-22-1-0411, DOE ASCR grant #DE-ACO5-000R22725, DoD ONR MURI grant #N00014-20-1-2595, and through Autonomy Technology Research Center by AFRL contract #FA 8650-18-2-1645.

Preface

I want to begin by thanking my advisor, Professor Anne Gelb. Without her, none of this work would have been possible, and I would not be the mathematician and scholar that I am today. Her guidance has been invaluable, and I am incredibly thankful to call her my mentor and friend. I would also like to thank the other three members of my committee, Professor Geoffrey P. Luke, Professor Yoonsang Lee, and Dr. Theresa Scarnati, for their insightful comments and support.

I want share my appreciation for my office-mate Ben and for the rest of our cohort, Kameron, Kathy, Melanie, and Ryan. I want to especially thank Kameron McCombs, who has been there for me through thick and thin over the past five years and is one of my best friends in life. I also want to thank Jonathan Lindbloom for the research conversations we've had. I look forward to seeing where the years will take all of us.

I want to thank my friends who kept me sane during this process. Jordan, Evan, Ryan, Logan, Jeremy, and Jacob, our regular virtual hangouts got me through the cold months. José and Emily, our calls always brighten my day and warm my heart.

Thank you to Mom, Dad, Mema, Papaw, and Shamaw for always loving me, guiding me, and believing in me every step of the way for as long as I can remember. Finally, I am ever thankful to my husband and the love of my life, Paolo. You inspire me daily, and being with you makes me happier than I ever thought possible. I am grateful for every moment I get to share with you, and I can't wait to see what the future holds. And thank you to our dog, Piper, for being the best pickle.

Joint and Published Work

Chapter 3 describes work done in collaboration with Professors Anne Gelb (Dartmouth College) and Geoffrey P. Luke (Dartmouth College) and was previously published in [55], which may be reused without permission under the Creative Commons CC BY license. The contribution of each author is summarized as follows: Conceptualization, G.P.L.; methodology, G.P.L.; software, D.G.; validation, all; formal analysis, all; investigation, D.G.; writing–original draft preparation, D.G.; writing–review and editing, all; visualization, D.G. and G.P.L.

Chapters 4 and 5 describe work done with Professor Anne Gelb and with Jonathan Lindbloom (Dartmouth College). The contribution of each author can be summarized as: Conceptualization, D.G. and A.G.; methodology, D.G.; software, D.G. and J.L.; validation, all; formal analysis, D.G.; investigation, D.G.; writing–original draft preparation, D.G.; writing–review and editing, all; visualization, D.G. and J.L.

Chapter 6 details work published in [56] by AIMS in collaboration with Professor Anne Gelb and with JR Jamora (Air Force Research Laboratory). The journal allows for the reuse of material so long as proper acknowledgments are made to the publisher. The author contributions can be summarized as: Conceptualization, D.G. and J.J.; methodology, D.G.; software, all; validation, all; formal analysis, D.G.; investigation, D.G. and J.J.; writing–original draft preparation, D.G.; writing–review and editing, all; visualization, D.G.

Contents

Abstract	ii
Preface	iii
Joint and Published Work	iv
1 Introduction	1
1.1 Contributions	2
1.2 Outline of Dissertation	2
2 Preliminaries	3
2.1 Regularization Techniques	3
2.1.1 Singular Value Decomposition	4
2.1.2 Tikhonov Regularization	8
2.1.3 ℓ_1 -Regularization	10
2.1.4 Regularization in Other Domains	11
2.2 Statistical Inversion Theory	13
2.2.1 Probability	13
2.2.2 Bayesian Inference for Complex-Valued Variables	16
2.2.3 Markov Chain Monte Carlo Methods	18
2.3 SAR Signal Processing	25
3 Sparsity-Based Reconstruction of 3D Photoacoustic Images	30

3.1	Introduction	30
3.2	Methods	32
3.2.1	Optical Setup	32
3.2.2	Continuous Model	33
3.2.3	Discrete Model	35
3.2.4	Image Reconstruction Via Compressed Sensing	37
3.2.5	Simulation Setup and Analysis	39
3.3	Numerical Results	41
3.3.1	Baseline Test	41
3.3.2	Simulated Experiments	41
3.4	Discussion and Conclusions	48
4	Complex-Valued Bayesian LASSO	50
4.1	Introduction	50
4.2	Bayesian Formulation	54
4.2.1	The likelihood	54
4.2.2	The prior density function	55
4.2.3	The sparsifying transform operator	58
4.3	Real-Valued Bayesian LASSO	59
4.4	Complex-valued Bayesian LASSO	64
4.4.1	CVBL for the Sparse Magnitude Case	66
4.4.2	Sparsity in a Transform Domain of $ \mathcal{Z} $	70
4.5	Numerical Results	81
4.5.1	Numerical Efficiency	84
4.5.2	Sparsity in Magnitude	85
4.5.3	Sparsity in Transform of Magnitude	86
4.5.4	Noise Study	90

4.5.5	2D Experiments	92
4.6	Concluding remarks and future work	95
5	Empirical Bayesian Inference for Complex-Valued Signals	97
5.1	Introduction	97
5.2	A Bayesian Approach	102
5.2.1	The likelihood	104
5.2.2	The prior density function	104
5.2.3	Weighted mask construction for the prior	109
5.3	Sampling the Support-Informed Posterior for Complex-Valued Signals	116
5.3.1	Sparse magnitude prior	116
5.3.2	Sparsity in magnitude gradient domain	117
5.4	Numerical Results	122
5.5	Concluding remarks	134
6	Leveraging Joint Sparsity in 3D SAR Imaging	136
6.1	Introduction	136
6.2	SAR data collection	138
6.2.1	The Fourier Slice Theorem	139
6.2.2	Data used in experiments	140
6.2.3	SAR image recovery	141
6.3	SAR imaging leveraging sequential information	143
6.3.1	2D SAR imaging leveraging sequential information	144
6.3.2	Volumetric SAR using backprojection	146
6.3.3	Volumetric SAR leveraging sequential information	149
6.4	Bayesian approach to 3D SAR image formation	150
6.4.1	Hierarchical Bayesian model	151

6.4.2	2D image reconstruction with backprojection (2D IRB)	157
6.4.3	3D subaperture reconstruction with composite imaging (3D SRCI)	159
6.5	Numerical experiments	161
6.5.1	Selection of parameters	162
6.5.2	Reconstruction from synthetic data	163
6.5.3	Reconstructions from measured data	165
6.5.4	Noise study	167
6.5.5	Sub-sampled data	169
6.6	Conclusion	171
6.7	Derivation of the JHBL updates	173
6.7.1	The G update	173
6.7.2	The α , β , and γ updates	175
7	Concluding Remarks	176
	References	179

List of Tables

3.1	The parameters used in our simulations and their associated values.	40
6.1	Parameters of synthetic and measured data sets used for experimentation.	142
6.2	Sizes of the inputs and outputs of Algorithm 12 and Algorithm 13 for our numerical experiments.	163
6.3	Minimum MHD (cm) achieved across tested dB thresholds.	167
6.4	Parameters of sub-sampled data set used for experimentation.	170
6.5	Sizes of the parameter inputs and outputs for Algorithm 11 for the sub-sampled data experiments.	170

List of Figures

2.1	Example of regularization using both the 1- and 2-norms. (a) details the ground truth and blurred noisy data. (b) shows solutions obtained using (2.12) with L as the TV operator.	12
2.2	Example of regularization using both the 1- and 2-norms using the data in Figure 2.1. Different values of λ are tested in the objective function given by (2.12) for (a) $q = 2$ and (b) $q = 1$	13
2.3	The geometry of 3D SAR data collection.	26
2.4	Linear FM chirp.	26
3.1	The proposed optical system. SMF = single-mode fiber, OI = optical isolator, COL = collimator, PBS = polarized beam splitter, $\lambda/4$ = quarter wave plate, L = lens, CAM = camera, DMD = digital micromirror device, LP = linear polarizer, FPE = Fabry–Pérot etalon.	33
3.2	A pictorial representation of the imaging process as individual forward operations.	37
3.3	(a) A cross-section of the ground truth impulse IPD and (b–d) cross-sections orthogonal to the y -, x -, and z -axes, respectively, of a noise reconstruction of the impulse IPD. Here we use regularization parameter $\lambda = 2.5 \times 10^{-4}$ in (3.6).	43

3.4	(a) The base case cylindrical IPD and (b) an example of a vessel-like IPD.	43
3.5	Average MS-SSIM (a) and average MSE (b) of a cylinder IPD with a radius of six perpendicular to the direction of shearing. Average MS-SSIM (c) and average MSE (d) of a cylinder IPD with a radius of 6 h parallel to the direction of shearing. Average MS-SSIM (e) and average MSE (f) of a vessel IPD with ten vessels present. Each point is averaged over five trials for each SNR value considered and is plotted against the SNR value used to calculate the additive Gaussian noise.	44
3.6	Average MSE (a) and average MS-SSIM (b) of the reconstructed cylinder IPD with varying radius, averaged over five trials for each radius value considered. A cross section of the ground truth cylinder with radius of five (c), ten (d) and fifteen (e) voxel widths and a cross section of the reconstruction of the same cylinder, respectively, (f–h). Each cylinder considered is orthogonal to the xz -plane.	46
3.7	Average MSE (a) and average MS-SSIM (b) of the reconstructed vessel IPDs with a varying number of vessels present, averaged over twenty IPDs for each number of vessels considered. Ground truth projection onto the xy -plane of a four vessel IPD (c), eight vessel IPD (d), and twelve vessel IPD (e), and the reconstruction of the same IPDs, respectively, (f–h). In images (c–h), hue represents depth in the z -dimension, with the colorbar indicating pixel lengths away from the FPE, while intensity is proportional the value of the voxels after being thresholded at 0.15.	47

3.8	Average MSE (a) and average MS-SSIM (b) of the reconstructed cylinder IPD as well as average MSE (c) and average MS-SSIM (d) of the reconstructed vessel-like IPD, averaged over five trials for each value of the regularization parameter considered. Low, medium, and high values of SNR are considered for comparison.	48
4.1	The RVBL method applied to a signal deblurring problem. The posterior mean estimate for the parameter η^{-2} is $\bar{\eta}^{-2} \approx 401$ and SNR = 30. in (4.52).	64
4.2	Recovered magnitude values for complex-valued signals with sparse magnitude. (left) The magnitude of the LASSO solution and the mean of the CVBL output; (right) 90% CI for the magnitude reconstruction. Here SNR = 20 dB.	86
4.3	Means of (left) real and (right) imaginary parts of the signal in the sparse magnitude case. Shaded regions indicate 90% CIs for the CVBL recovery. SNR = 20 dB.	87
4.4	Phase estimate in the sparse magnitude case at a randomly selected “on” pixel ($g_j = 1$). True phase, LASSO phase point estimate, and CVBL marginal posterior for the phase. SNR = 20 dB.	88
4.5	Magnitude recovery for \mathbf{g} given by (4.56). (left) The CVBL mean along with the magnitude of LASSO solutions using different regularization parameters; (right) 90% CI for the magnitude reconstruction. Here SNR = 20 dB.	89
4.6	Phase estimate in the sparse transform case of a random pixel. True phase (black dash), LASSO phase point estimate (red dash-dot), and CVBL marginal posterior for the phase. SNR = 20 dB.	90

4.7	Mean of τ^2 (right axis) compared with the true magnitude (left axis). SNR = 20 dB.	90
4.8	Average error in the sample mean generated using Algorithm 5 for increasing SNR (dB). The vertical axes are in a natural logarithmic scale.	90
4.9	Comparison of the average phase error in mean of samples generated using CVBL with the LASSO solution (4.55) for the test function (4.56) at (a)-(c) increasing SNR and (d) increasing sampling rate $0.5 \leq \nu \leq 1$ with SNR = 20 dB. The vertical axes are in a natural logarithmic scale.	91
4.10	Maximum likelihood and mean posterior point estimates of a sparse magnitude complex-valued signal for forward operator F_F in (4.48). SNR = 0 dB.	92
4.11	Maximum likelihood and mean posterior point estimates recoveries of a sparse magnitude complex-valued signal. Forward operator (a-b) F_B in (4.49) (c-d) F_U $\nu = 0.8$ in (4.50). SNR = 0 dB.	92
4.12	Magnitude recovery of the complex-valued Shepp Logan phantom using F_F in (4.48). (a) Shepp-Logan phantom; (d) horizontal differences (computed using $L \mathbf{z}^{exact} $) of the ground truth. (b)-(c) and (e)-(f) correspond to SNR values of 15 dB and 25 dB, respectively. $\hat{\eta} = 10^{-2}$.	93
4.13	Magnitude recovery of the complex-valued Shepp Logan phantom using F_F in (4.48). (a)-(c) and (d)-(f) correspond to SNR values of 15 dB and 25 dB, respectively. $\hat{\eta} = 10^{-2}$	94
4.14	Magnitude recovery of the complex-valued Shepp Logan phantom using F_B in (4.49). (a)-(c) and (d)-(f) correspond to SNR values of 15 dB and 25 dB, respectively. $\hat{\eta} = 10^{-2}$	94

4.15	Recovered phase at a randomly chosen pixel of the complex-valued Shepp Logan image using F_F in (4.48) and F_B in (4.49) with SNR of 15 dB and 25 dB. $\hat{\eta} = 10^{-2}$	95
5.1	Sparse signal magnitude. (a) Ground truth; (b) \mathbf{z}^{CS} with $\alpha = 1$; (c) \mathbf{z}^{VBJs} computed via (5.46); and (d) Mean $\boldsymbol{\mu}$ in (5.34). Here SNR = -20 dB.	125
5.2	Sparse signal recovery. Here SNR = -20 dB.	125
5.3	Sparse signal phase at three randomly chosen “on” pixels. Here SNR = -20 dB.	126
5.4	Ground truth of (a) magnitude; (b) phase; and (c) magnitude gradient (vertical components).	127
5.5	(a)-(b) Average of (vertical) magnitude gradient, $\frac{1}{J} \sum_{j=1}^J \mathbf{s}^j$, computed via (5.30); (c)-(d) vertical components of $\log_{10}(W)$	127
5.6	Magnitude recovery: (top) SNR = 15 dB; (bottom) SNR = 10 dB.	128
5.7	Cross-sections of the magnitude of the true image (black), mean of Algorithm 8 (red), \mathbf{z}^{VBJs} (yellow), and \mathbf{z}^{CS} with $\alpha = 1$ (purple).	128
5.8	Phase recovery: (top) SNR = 15 dB; (bottom) SNR = 10 dB.	129
5.9	(top) Magnitude pointwise error in (a) \mathbf{z}^{CS} with $\alpha = 1$ computed via (5.45), (b) mean from Algorithm 8, (c) \mathbf{z}^{VBJs} computed via (5.47); (bottom) Pointwise variance computed from Algorithm 8 of (c) magnitude and (d) phase (in logarithmic scale). Here SNR = 10 dB.	130
5.10	Mean squared error of \mathbf{z}^{CS} in (5.45) for $\alpha = 0.1, 1, 10$ and the mean of the samples generated using Algorithm 8 plotted against SNR values. The vertical axis is in logarithmic scale.	130
5.11	Reflectivity of full SLC image. Scale is -65 dB to 0 dB.	132

5.12	(a) Ground truth reflectivity; (b) \mathbf{z}^{CS} with $\alpha = 1$ computed via (5.44); (c) $\mathbf{z}^{VBJ S}$ computed via (5.46); (d) mean μ computed using (5.34); and (e) pointwise variance computed using (5.34). Here SNR = 0 dB.	133
5.13	(top) Pointwise (circular) phase variance, (bottom) phase results at a randomly chosen pixel. (left) SNR = -5 dB, (middle) SNR = 0 dB, (right) SNR = 5 dB.	133
5.14	(top) SNR = 5 dB; (bottom) SNR = -5 dB.	134
6.1	A graphical depiction of Theorem 6.1.	141
6.2	A graphical depiction of SAR PHD $\hat{\mathbf{g}}$ (6.3) in k -space as well as the partitioning of the data into N_θ partitions according to the azimuthal angle sets Θ_n given by (6.9).	144
6.3	Different views at various dB thresholds of the 3D reconstruction of the synthetic cube data set in the ideal case; ground truth point cloud is displayed in black. The threshold values chosen to best demonstrate reconstruction quality.	164
6.4	Centered cross-sections of the (left) 2D IRB and (middle) 3D SRCI reconstructions of the cube data set with no additional noise and no thresholding. (right) MHD values at various c_{IRB} or c_{SRCI} threshold values (dB) when either technique is used on the cube data set with no additional noise; the minimum MHD value calculated for the 2D IRB method is 0.7017cm, and for the 3D SRCI, the minimum MHD is 0.6014cm.	165
6.5	Different views at various dB thresholds of the 3D reconstruction of the B747 data set with no added noise; ground truth CAD model is displayed in black.	165

6.6	Centered cross-sections of the (left) 2D IRB and (middle) 3D SRCI reconstructions of the B747 data set with no additional noise and no thresholding. (right) MHD values at various c_{IRB} or c_{SRCI} threshold values (dB) when either technique is used on the B747 data set with no additional noise; the minimum MHD value calculated for the 2D IRB method is 1.832cm, and for the 3D SRCI, the minimum MHD is 1.440cm.	166
6.7	Threshold value vs. MHD for the cube (left) and B747 (right) data sets comparing the 2D IRB and 3D SRCI for both the JHBL and MLE approximations. (top) SNR ≈ 0 dB; (bottom) SNR ≈ -24 dB. In all plots, the dashed blue lines are the MLE MHD values, while the solid red lines are the JHBL MHD values. In all cases, it is straightforward to infer the rest of the characterization of the MHD values by continuing the trends in (a)-(h).	168
6.8	Slices of the (left) MLE and (right) JHBL reconstructions of the B747 with SNR of -30 dB using the 3D SRCI approach.	168
6.9	Slices of the (left) 2D IRB and (right) 3D SRCI reconstructions of the B747 with SNR of approximately -34 dB. Note that for interpretability, the threshold dB scale is different for each figure.	169
6.10	Different views at various dB thresholds of the 3D reconstruction of the B747 sub-sampled data set using our reconstruction techniques with no additional noise added; ground truth CAD model is displayed in black.	170
6.11	Cross-sections of the (left) 2D IRB and (middle) 3D SRCI reconstructions of the sub-sampled B747 data set using the parameters in Tables 6.4 and 6.5. (right) MHD values at various dB threshold values. . . .	171

Chapter 1

Introduction

This dissertation is primarily concerned with deriving meaningful information from so-called *inverse problems*. An inverse problem provides data regarding an unknown quantity; the goal is then to solve for the unknown. In this work we are interested in solutions to discrete computational and statistical inverse problems and their applications in science and engineering.

Let \mathbb{F} be the field of either the real or complex numbers. In its most general form, an inverse problem can be described as finding \mathbf{x} given

$$\mathbf{y} = f(\mathbf{x}, \boldsymbol{\varepsilon}), \tag{1.1}$$

where $\mathbf{y} \in \mathbb{F}^m$ is the known data, $\mathbf{x} \in \mathbb{F}^n$ is the unknown of interest, $\boldsymbol{\varepsilon} \in \mathbb{F}^k$ encompasses any unknown nuisance parameters, and $f : \mathbb{F}^{n \times k} \rightarrow \mathbb{F}^m$ is the relationship between the data and the unknown.

We will be considering problems where the forward operator f is a known linear operator $A \in \mathbb{F}^{m \times n}$. Hence we reformulate (1.1) as

$$\mathbf{y} = A\mathbf{x} + \boldsymbol{\varepsilon}, \tag{1.2}$$

We will also assume that \mathbf{x} has some known structure, e.g. \mathbf{x} may have relatively few large elements or edges. For additional background regarding discrete inverse problems, see [60].

Section 1.1

Contributions

This dissertation details a new optical system for photoacoustic imaging and methods for solving the resulting linear inverse problem. The work then expands on existing real-valued Bayesian inference techniques by adapting them to complex-valued unknowns where sparsity is assumed in some transform domain of the magnitude. Additionally, a new Bayesian method for 3D SAR imaging is implemented.

Section 1.2

Outline of Dissertation

In Chapter 2 we cover several preliminary ideas utilized in our work, including regularization techniques, synthetic aperture radar (SAR) image formation, and Bayesian inference. Chapter 3 details an optical system for capturing photoacoustic data and a compressed sensing approach to reconstructing an initial 3D pressure distribution. Chapter 4 and Chapter 5 present two approaches for generating samples of a complex-valued posterior distribution given various prior information regarding sparsity. In Chapter 4 we discuss an approach for dealing with one measurement vector, while Chapter 5 considers multiple observations of the same scene of interest. Chapter 6 explores a technique for reconstructing 3D SAR images using a Bayesian *maximum a posteriori* estimate that enforces two kinds of sparsity. We conclude in Chapter 7 by discussing the work accomplished as well as future work.

Chapter 2

Preliminaries

The research in this dissertation is motivated by solving inverse problems for synthetic aperture radar (SAR). This chapter discusses the necessary preliminary information needed to provide context for our contributions. We begin by discussing several classical regularization techniques in Section 2.1, followed by an overview of statistical inversion theory in Section 2.2. We conclude this chapter with an examination of SAR image formation in Section 2.3. For our discussion of discrete inverse problems and SAR, we follow the conventions of [60] and [66], respectively. For statistical inversion, we draw primarily from [73].

Section 2.1

Regularization Techniques

The classic inverse problem (1.1) must meet three criteria to be considered well-posed [60]. These conditions are:

- A solution to the problem *exists*.
- The solution is *unique*.
- The problem is *stable*.

By *stability*, we mean that any perturbation in the data \mathbf{y} leads to a comparable change in the solution \mathbf{x} . When an inverse problem fails to be well-posed, it is often useful to alter the problem in some way to make it well-posed while still providing an adequate answer to the original problem. These alterations are called *regularization* methods.

2.1.1. Singular Value Decomposition

Consider the forward model given in (1.2). Singular value decomposition (SVD) is a powerful tool used to extract meaningful information from the matrix A and provides the foundation for several inversion and regularization techniques [60, 73].

Theorem 2.1. *Let $A \in \mathbb{C}^{m \times n}$ be of rank ℓ . It is possible to decompose A into*

$$A = USV^*,$$

where $U \in \mathbb{C}^{m \times m}$ is unitary, $S \in \mathbb{R}^{m \times n}$ is rectangular diagonal with nonnegative entries, and $V \in \mathbb{C}^{n \times n}$ is unitary.

Proof. By the spectral theorem [9], there exists unitary $V \in \mathbb{C}^{n \times n}$ such that

$$V^* A^* A V = \Lambda \tag{2.1}$$

with

$$\Lambda = \begin{bmatrix} D & 0 \\ 0 & 0 \end{bmatrix},$$

where $D \in \mathbb{R}^{\ell \times \ell}$ is a diagonal matrix with positive entries. Clearly, $\ell \leq \min(m, n)$. Consider the block partition $V = [V_1 \ V_2]$, where the columns of $V_1 \in \mathbb{C}^{n \times \ell}$ are the ℓ eigenvectors with positive eigenvalues and the columns of V_2 are the eigenvectors

with eigenvalue 0. Then we have from (2.1) that

$$\begin{bmatrix} V_1^* A^* A V_1 & V_1^* A^* A V_2 \\ V_2^* A^* A V_1 & V_2^* A^* A V_2 \end{bmatrix} = \begin{bmatrix} D & 0 \\ 0 & 0 \end{bmatrix}.$$

Now, consider $U_1 = A V_1 D^{-1/2}$. We have that

$$U_1^* U_1 = D^{-1/2} V_1^* A^* A V_1 D^{-1/2} = D^{-1/2} D D^{-1/2} = \mathbb{I}_\ell.$$

Thus, the columns of U_1 are orthonormal. Hence, if $m = \ell$, then $U = U_1$ is unitary.

Otherwise, $m > \ell$, and there exists $U_2 \in \mathbb{C}^{m \times (m-\ell)}$ such that $U = [U_1 \ U_2]$ is unitary.

Let $S \in \mathbb{R}^{m \times n}$ be such that

$$S = \begin{bmatrix} D^{1/2} & 0 \\ 0 & 0 \end{bmatrix}.$$

Therefore,

$$U S V^* = [U_1 \ U_2] \begin{bmatrix} D^{1/2} & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} V_1^* \\ V_2^* \end{bmatrix} = U_1 D^{1/2} V_1^* = A.$$

□

The columns of U are called the *left singular vectors*, the columns of V are called the *right singular vectors*, and the diagonal elements of S are the *singular values*, which we denote σ_i , $i = 1, 2, \dots, \min(m, n)$.

Several comments regarding U , S , and V are in order. First, we order σ_i such that $\sigma_i \geq \sigma_j$ when $i \leq j$, and $\sigma_i > 0$ for $i = 1, 2, \dots, \ell$ where again $\ell \leq \min(m, n)$. We then have $\sigma_{\ell+1}, \dots, \sigma_{\min(m, n)} = 0$. The columns of U and V are the orthonormal eigenvectors of AA^* and A^*A , respectively. Lastly, σ_i^2 , $i = 1, 2, \dots, \ell$, are the first ℓ

eigenvalues of AA^* and A^*A .

Denote by \mathbf{u}_i and \mathbf{v}_i the i th column of U and V , respectively. Let us consider the least squares solution to (1.2) with $\mathbb{F} = \mathbb{C}$, which is given by

$$\mathbf{x}_{LS} = \arg \min_{\mathbf{x}} \|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2. \quad (2.2)$$

Using the SVD and the fact that multiplying by unitary matrices preserves norms, we find that

$$\begin{aligned} \mathbf{x}_{LS} &= \arg \min_{\mathbf{x}} \|U^*(USV^*\mathbf{x} - \mathbf{y})\|_2^2 \\ &= \arg \min_{\mathbf{x}} \|SV^*\mathbf{x} - U^*\mathbf{y}\|_2^2 \\ &= \arg \min_{\mathbf{x}} \left\| \begin{bmatrix} D^{1/2} & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} V_1^* \\ V_2^* \end{bmatrix} \mathbf{x} - \begin{bmatrix} U_1^* \\ U_2^* \end{bmatrix} \mathbf{y} \right\|_2^2 \\ &= \arg \min_{\mathbf{x}} \|D^{1/2}V_1^*\mathbf{x} - U_1^*\mathbf{y}\|_2^2 \end{aligned} \quad (2.3)$$

Thus we seek \mathbf{x} such that (2.3) is minimized, which is true when

$$\begin{bmatrix} \mathbf{v}_1^*\mathbf{x} \\ \vdots \\ \mathbf{v}_\ell^*\mathbf{x} \end{bmatrix} = \begin{bmatrix} \mathbf{u}_1^*\mathbf{y}/\sigma_1 \\ \vdots \\ \mathbf{u}_\ell^*\mathbf{y}/\sigma_\ell \end{bmatrix}$$

For each $j = 1, \dots, \ell$, we then have $\mathbf{v}_j^*\mathbf{x} = \mathbf{u}_j^*\mathbf{y}/\sigma_j$. Since the columns of V_1 are orthonormal, this implies that the least squares solution is given by

$$\mathbf{x}_{LS} = \sum_{j=1}^{\ell} \frac{\mathbf{u}_j^*\mathbf{y}}{\sigma_j} \mathbf{v}_j. \quad (2.4)$$

The formulation in (2.4) offers several key insights. First, \mathbf{x} is a linear combination

of the vectors \mathbf{v}_j , $i = 1, \dots, \ell$. We also observe that the least squares solution is not unique when A has a nontrivial null space. This can be seen by adding any multiple of \mathbf{v}_j , $j = \ell + 1, \dots, n$, to \mathbf{x}_{LS} . Lastly, the weight of each vector \mathbf{v}_j is inversely proportional to the corresponding singular value. This means that in general the terms of the summation in (2.4) become more sensitive to noise in \mathbf{y} as the index j increase.

This insight then leads to what are called *spectral filtering methods*, where (2.4) is reformulated as

$$\mathbf{x} = \sum_{j=1}^{\ell} \varphi_j \frac{\mathbf{u}_j^T \mathbf{y}}{\sigma_j} \mathbf{v}_j. \quad (2.5)$$

Here, $\varphi_i \in [0, 1]$ are called the *filter factors*. One such method is the truncated singular value decomposition (TSVD). In this technique, for some user-specified integer $k < \ell$, $\varphi_i = 1$ for $i \leq k$ and $\varphi_i = 0$ otherwise. This gives the TSVD solution \mathbf{x}_{TSVD} as

$$\mathbf{x}_{TSVD} = \sum_{j=1}^k \frac{\mathbf{u}_j^T \mathbf{y}}{\sigma_j} \mathbf{v}_j.$$

Another regularization method based off of the SVD is the selective singular value decomposition (SSVD). Similar to TSVD, the SSVD approach discards a portion of the terms in the summation in (2.4). In SSVD, however, a threshold τ is set for $|\mathbf{u}_i^T \mathbf{y}|$ where $\varphi_i = 1$ if $|\mathbf{u}_i^T \mathbf{y}| > \tau$ and $\varphi_i = 0$ otherwise. Hence, the solution \mathbf{x}_{SSVD} is given as

$$\mathbf{x}_{SSVD} = \sum_{|\mathbf{u}_j^T \mathbf{y}| > \tau} \frac{\mathbf{u}_j^T \mathbf{y}}{\sigma_j} \mathbf{v}_j.$$

Additional information regarding TSVD and SSVD can be found in [60, Chapter 4].

2.1.2. Tikhonov Regularization

The condition number of a matrix A is defined as

$$\text{cond}(A) = \|A\| \|A^\dagger\|,$$

where A^\dagger is the Moore-Penrose pseudoinverse of A . When the condition number of A is large, the least squares solution \mathbf{x}_{LS} in (2.2) may not be stable, i.e. small perturbations may have an outsized impact on the solution. One method to mitigate the effect of noise while still making full use of all singular vectors associated with nonzero singular values is by using Tikhonov regularization. This approach recasts the problem in (2.2) as

$$\mathbf{x}_\lambda = \arg \min_{\mathbf{x}} \|A\mathbf{x} - \mathbf{y}\|_2^2 + \lambda^2 \|\mathbf{x}\|_2^2, \quad (2.6)$$

where $\lambda > 0$ is called the *regularization parameter*.

Theorem 2.2. *The Tikhonov regularization problem (2.6) has a unique solution \mathbf{x}_λ given as*

$$\mathbf{x}_\lambda = (A^*A + \lambda^2 \mathbb{I}_n)^{-1} A^* \mathbf{y}. \quad (2.7)$$

Proof. This proof is adapted from the real-valued analogue in [60, Chapter 4]. Let $\mathbf{z} \in \mathbb{C}^n$, $\mathbf{z} \neq \mathbf{0}$. Because A^*A is semi-positive definite, we see that

$$\mathbf{z}^*(A^*A + \lambda^2 \mathbb{I}_n)\mathbf{z} = \mathbf{z}^*A^*A\mathbf{z} + \lambda^2 \mathbf{z}^*\mathbf{z} \geq \lambda^2 \mathbf{z}^*\mathbf{z} > 0.$$

Thus, $(A^*A + \lambda^2 \mathbb{I}_n)$ is positive definite and its inverse exists. Then \mathbf{x}_λ is well-defined.

Consider the functional

$$F_\lambda(\mathbf{x}) = \|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2 + \lambda^2 \|\mathbf{x}\|_2^2.$$

Let $\mathbf{w} \in \mathbb{C}^n$. Using (2.7), we have then that

$$\begin{aligned} F_\lambda(\mathbf{x}_\lambda + \mathbf{w}) &= \|\mathbf{A}\mathbf{x}_\lambda + \mathbf{A}\mathbf{w} - \mathbf{y}\|_2^2 + \lambda^2 \|\mathbf{x}_\lambda + \mathbf{w}\|_2^2 \\ &= F_\lambda(\mathbf{x}_\lambda) + \mathbf{w}^*((\mathbf{A}^*\mathbf{A} + \lambda^2\mathbb{I}_n)\mathbf{x}_\lambda - \mathbf{A}^*\mathbf{y}) \\ &\quad + ((\mathbf{A}^*\mathbf{A} + \lambda^2\mathbb{I}_n)\mathbf{x}_\lambda - \mathbf{A}^*\mathbf{y})^*\mathbf{w} + \mathbf{w}^*(\mathbf{A}^*\mathbf{A} + \lambda^2\mathbb{I})\mathbf{w} \\ &= F_\lambda(\mathbf{x}_\lambda) + \mathbf{w}^*(\mathbf{A}^*\mathbf{A} + \lambda^2\mathbb{I})\mathbf{w} \\ &\geq F_\lambda(\mathbf{x}_\lambda), \end{aligned}$$

with equality being achieved if and only if $\mathbf{w} = \mathbf{0}$. Thus, \mathbf{x}_λ given by (2.7) exists and is the unique solution to (2.6). \square

Now, using the SVD of A in (2.7), we have

$$\begin{aligned} \mathbf{x}_\lambda &= (\mathbf{V}\mathbf{S}^2\mathbf{V}^* + \lambda^2\mathbb{I}_n)^{-1}\mathbf{V}\mathbf{S}\mathbf{U}^*\mathbf{y} \\ &= \mathbf{V}(\mathbf{S}^2 + \lambda^2\mathbb{I}_n)^{-1}\mathbf{S}\mathbf{U}^*\mathbf{y} \\ &= \sum_{j=1}^{\ell} \frac{\sigma_j^2}{\sigma_j^2 + \lambda^2} \frac{\mathbf{u}_j^*\mathbf{y}}{\sigma_j} \mathbf{v}_j. \end{aligned} \tag{2.8}$$

Comparing (2.8) with (2.5), we see the solution \mathbf{x}_λ given by (2.7) is a spectral filtering method with filter factor

$$\varphi_j^{[\lambda]} = \frac{\sigma_j^2}{\sigma_j^2 + \lambda^2}.$$

2.1.3. ℓ_1 -Regularization

Lastly, we turn our attention to ℓ_1 -regularization problems. Like Tikhonov regularization, ℓ_1 -regularization takes the least squares problem in (2.2) and adds an additional term weighted by a regularization parameter δ . Now, however, the extra term includes an ℓ_1 -norm instead of an ℓ_2 -norm. Thus we have the ℓ_1 -regularization problem as

$$\mathbf{x}_\delta = \arg \min_{\mathbf{x}} \|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2 + \delta \|\mathbf{x}\|_1, \quad (2.9)$$

where $\delta > 0$. This use of the ℓ_1 -norm encourages the solution \mathbf{x}_δ to be *sparse*.

Definition 2.1. When a signal $\mathbf{v} \in \mathbb{F}^n$ contains at most s nonzero values, we say that \mathbf{v} is *s-sparse*.

To get an intuitive sense of why (2.9) promotes sparsity, let us instead consider the so-called ℓ_0 -“norm,”¹ which for arbitrary vector $\mathbf{v} \in \mathbb{F}^n$ is given as

$$\|\mathbf{v}\|_0 = |\{v_j : v_j \neq 0, 1 \leq j \leq n\}|. \quad (2.10)$$

If we were to regularize (2.2) using $\|\cdot\|_0$, it is clear that the resulting problem would promote sparsity due to the number of nonzero elements in the solution being penalized. The objective function, however, would be intractable in high dimensions due to the nonlinearity of (2.10) and is indeed an NP-hard problem. Thus we choose to use a surrogate in the form of the ℓ_1 -norm, which has been shown to promote sparsity similarly to regularization using $\|\cdot\|_0$. Note that finding the solution of (2.9) often requires the use of iterative optimization algorithms, and the technique is not a spectral filtering method since its solution does not follow the form of (2.5).

¹We note that the ℓ_0 -“norm” (2.10) does not meet the definition of a norm and is commonly referred to as a *pseudo*-norm.

Regularization with the ℓ_1 -norm is used extensively in compressive sensing, an area of study which assumes \mathbf{x} is sparse and $m \ll n$. It has been shown that when these assumptions hold and A obeys what is called the *restricted isometry property*, \mathbf{x} can be recovered with certain theoretical guarantees [24, 26, 41]. The reader is referred to [23] for a thorough discussion of this topic.

2.1.4. Regularization in Other Domains

The objective functions given by (2.6) and (2.9) both bias the solution \mathbf{x} in (1.2) to have a small q -norm, where $q = 2, 1$ respectively. There are situations where this form of regularization may be undesirable but where \mathbf{x} has other structural information that is known *a priori*. For example, the gradient of \mathbf{x} may be sparse, or \mathbf{x} may have few values that deviate from some known value \mathbf{x}_0 . In these cases, it is often beneficial to employ a generalized regularization approach where the objective function is

$$\mathbf{x}_{sol} = \arg \min_{\mathbf{x}} \|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2 + \Phi(\mathbf{x}). \quad (2.11)$$

In (2.11), $\Phi : \mathbb{F}^n \rightarrow \mathbb{R}$ is called the *regularizer*, and it is assumed that $\Phi(\mathbf{x})$ should be small. This work deals exclusively with linear regularizers enclosed within the 1- or 2-norm, i.e.

$$\mathbf{x}_{sol} = \arg \min_{\mathbf{x}} \|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2 + \frac{\lambda}{q} \|\mathbf{L}\mathbf{x}\|_q^q, \quad q = 1, 2, \quad (2.12)$$

where $L \in \mathbb{F}^{\ell \times n}$ and $\lambda > 0$. Such methods promote small values in $L\mathbf{x}$, and the level of regularization is determined by the user-specified value λ . When $L\mathbf{x}$ is expected to be sparse, the 1-norm is often used.

As an example, consider data $\mathbf{y} \in \mathbb{R}^n$ formed using

$$\mathbf{y} = A_{blur}\mathbf{x} + \boldsymbol{\varepsilon},$$

where $\mathbf{x} \in \mathbb{R}^n$ is piecewise-constant, $A_{blur} \in \mathbb{R}^{n \times n}$ is a full-rank Gaussian blur matrix, and $\boldsymbol{\varepsilon} \in \mathbb{R}^n$ is Gaussian white noise. Figure 2.1 (a) shows the ground truth and resulting data associated with this process. Deblurring is a notoriously ill-posed inverse problem, so the least squares objective function (2.2) causes high frequency terms to blow up in the solution. Taking L to be the first order finite difference operator (also known as discrete total variation, or TV), we can use the objective function (2.12) with either ℓ_1 or ℓ_2 norms to solve a well-posed version of the problem.

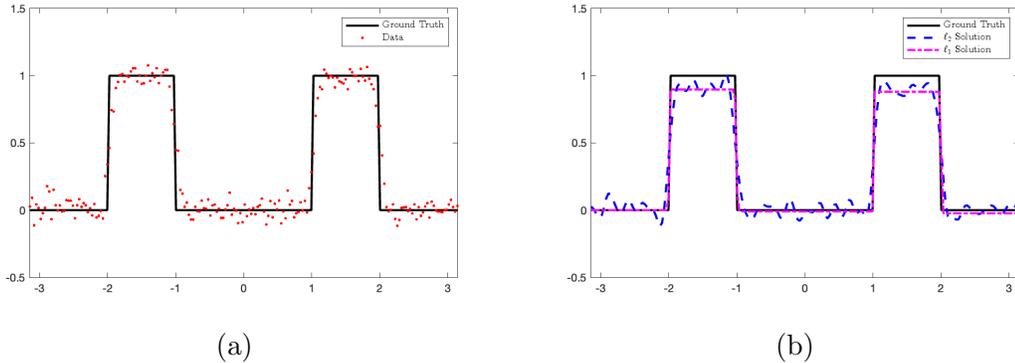


Figure 2.1: Example of regularization using both the 1- and 2-norms. (a) details the ground truth and blurred noisy data. (b) shows solutions obtained using (2.12) with L as the TV operator.

Figure 2.1 (b) shows that sensible solutions can be recovered using (2.12) with $q = 1$ or $q = 2$. The ℓ_2 form of (2.12) recovers a smoother solution, while ℓ_1 regularization finds a solution that is truly sparse in the gradient domain. These phenomena are commonly observed when using these two kinds of regularization, and the choice of which regularization to use is heavily problem-dependent and relies on *a priori* information of the unknown signal.

It is important to note that the choice of λ in (2.12) has a significant impact

on the fidelity of the resulting reconstruction. We see in Figure 2.2 that changing the choice of λ by an order of magnitude can have a drastic effect on the recovered signal. There are several methods for choosing λ , such as using the L-curve criterion and generalized cross-validation [60, Ch. 5]. Such methods are still prone to over- or under-regularizing the solution, which is in part what motivates our research.

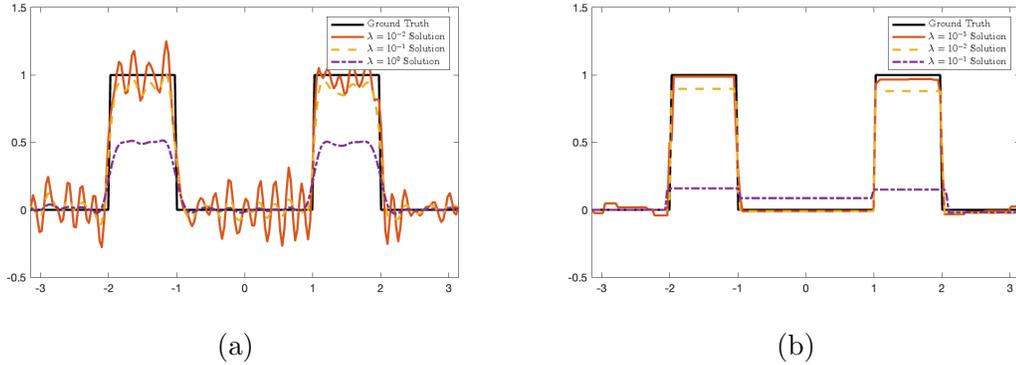


Figure 2.2: Example of regularization using both the 1- and 2-norms using the data in Figure 2.1. Different values of λ are tested in the objective function given by (2.12) for (a) $q = 2$ and (b) $q = 1$.

Section 2.2

Statistical Inversion Theory

The following discussions on probability and Bayesian inference are adapted from [21, 45, 73].

2.2.1. Probability

Define the *probability space* as the triplet (Ω, E, \mathbb{P}) . The abstract set Ω is referred to as the *sample space*. The σ -algebra E , called the *event space*, is a collection of subsets of Ω with properties

- (a) $\Omega \in E$,

- (b) If $A \in E$, then $A^c \in E$,
- (c) If $A_1, A_2, \dots \in E$, then $\cup_{i=1}^{\infty} A_i \in E$.

Lastly, \mathbb{P} is a *probability measure* satisfying

- (a) $0 \leq \mathbb{P}(A) \leq 1$ for $A \in E$,
- (b) $\mathbb{P}(\Omega) = 1$,
- (c) If $A_1, A_2, \dots \in E$ are disjoint, then $\mathbb{P}(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mathbb{P}(A_i)$.

We say that two events $A, B \in E$ are *independent* if $p(A \cap B) = p(A)p(B)$; that is, the probability of A and B is the product of the probabilities of A and B . The *conditional probability* of A given B is $p(A|B) = p(A \cap B)/p(B)$. In what follows, we assume that E is the Borel algebra on \mathbb{R}^n , $n \in \mathbb{N}$.

A *random variable* is a function $X : \Omega \rightarrow \mathbb{R}^n$ such that for every open set $A \subset \mathbb{R}^n$, $X^{-1}(A) \in E$. This condition is expressed by saying that X is a *measurable function*. A *realization* of X is $\mathbf{x} = X(\omega)$ for some $\omega \in \Omega$. Note that although we define random variables to be real-valued, complex-valued random variables can be defined such that the real and imaginary components are both real-valued random variables.

Consider the function $\mu_X : \mathbb{R}^n \rightarrow [0, 1]$ given as

$$\mu_X(A) = \mathbb{P}(X^{-1}(A)).$$

Every random variable induces such a function, which is called the *probability distribution* of X . The *distribution function* F of a random variable X is

$$F(\mathbf{x}) = \mathbb{P}(X_1 \leq x_1, X_2 \leq x_2, \dots, X_n \leq x_n), \quad \mathbf{x} = (x_1, x_2, \dots, x_n).$$

With every probability distribution μ_X , there exists *probability density* π such that²

$$\mu_X(A) = \int_A \pi_X(\mathbf{x})d\mathbf{x}, \quad A \subset \mathbb{R}^n.$$

Given two random variables $X \in \mathbb{R}^n$ and $Y \in \mathbb{R}^m$, the *joint probability density* $\pi_{X,Y}$ is

$$\mu_{X,Y}(A \times B) = \int \int_{A \times B} \pi_{X,Y}(\mathbf{x}, \mathbf{y})d\mathbf{x}d\mathbf{y}, \quad A \subset \mathbb{R}^n, B \subset \mathbb{R}^m,$$

and the *marginal density* of X is the probability of X when Y may have any value, i.e.

$$\pi_X(\mathbf{x}) = \int_{\mathbb{R}^m} \pi_{X,Y}(\mathbf{x}, \mathbf{y})d\mathbf{y}.$$

Similarly to how conditional probability was defined for sets in E , the conditional probability of X given $Y = \mathbf{y}$ is

$$\pi_{X|Y}(\mathbf{x}|\mathbf{y}) = \frac{\pi_{X,Y}(\mathbf{x}, \mathbf{y})}{\pi_Y(\mathbf{y})}. \quad (2.13)$$

This leads us then to the statement of Bayes' Theorem, which is the foundation for our approach to statistical inverse problems.

Theorem 2.3 (Bayes' Theorem). *Let $X \in \mathbb{R}^n$ and $Y \in \mathbb{R}^m$ be random variables with respective realizations \mathbf{x} and \mathbf{y} . Then, the following statement is true:*

$$\pi_{X|Y}(\mathbf{x}|\mathbf{y}) = \frac{\pi_{Y|X}(\mathbf{y}|\mathbf{x})\pi_X(\mathbf{x})}{\pi_Y(\mathbf{y})}. \quad (2.14)$$

Note that Theorem 2.3 follows directly from the definition of conditional proba-

²Throughout this text, both π and f will be used interchangeably to refer to probability density functions.

bility (2.13). In (2.14), $\pi_{X|Y}$ is the *posterior*, $\pi_{Y|X}$ is the *likelihood*, π_X is the *prior*, and π_Y is the *evidence*. In practice, the evidence term π_Y is commonly ignored, as it acts as a normalizing constant and is presumed to be nonzero. Thus, we have the proportional form of Bayes' Theorem

$$\pi_{X|Y}(\mathbf{x}|\mathbf{y}) \propto \pi_{Y|X}(\mathbf{y}|\mathbf{x})\pi_X(\mathbf{x}), \quad (2.15)$$

which states that the probability of \mathbf{x} given \mathbf{y} is proportional to the probability of \mathbf{y} given \mathbf{x} times the probability of \mathbf{x} .

2.2.2. Bayesian Inference for Complex-Valued Variables

Consider the statistical linear forward problem given by

$$Y = AZ + E \quad (2.16)$$

with random variables $Y, E \in \mathbb{C}^m$ and $Z \in \mathbb{C}^n$, linear operator $A \in \mathbb{C}^{m \times n}$, and where Y, E , and Z have respective realizations $\mathbf{y}, \boldsymbol{\varepsilon}$, and \mathbf{z} . In (2.16), Y is the observable data, Z is the unknown of interest, A defines the relationship between Z and Y , and E is noise. We will assume that Z and E are mutually independent.

We are interested in finding the probability density of the random variable of interest Z given the data Y . Using the proportional form of Bayes' Theorem (2.15), we have

$$\pi_{Z|Y}(\mathbf{z}|\mathbf{y}) \propto \pi_{Y|Z}(\mathbf{y}|\mathbf{z})\pi_Z(\mathbf{z}). \quad (2.17)$$

Thus to find the posterior density $\pi_{Z|Y}$, we must first discuss the likelihood $\pi_{Y|Z}$ and the prior π_Z .

Given the relation in (2.16), we have that $\pi_{Y|Z,E}(\mathbf{y}|\mathbf{z}, \boldsymbol{\varepsilon}) = \delta(\mathbf{y} - A\mathbf{z} - \boldsymbol{\varepsilon})$, where

the Dirac delta function δ is defined such that for any test function $g : \mathbb{C}^n \rightarrow \mathbb{R}$,

$$\int_{\mathbb{C}^n} \delta(\mathbf{x})g(\mathbf{x})d\mathbf{x} = g(\mathbf{0}).$$

The Dirac delta function can be considered as a probability density that equals zero everywhere other than at the origin. In other words, if $Z = \mathbf{z}$ and $E = \boldsymbol{\varepsilon}$, then $Y = \mathbf{y} = A\mathbf{z} + \boldsymbol{\varepsilon}$ with probability 1.

Suppose that the probability density function π_E for E is known. Then using the law of total probability, we can rewrite the likelihood function $\pi_{Y|Z}(\mathbf{y}|\mathbf{z})$ as

$$\begin{aligned} \pi_{Y|Z}(\mathbf{y}|\mathbf{z}) &= \int_{\mathbb{R}^m} \pi_{Y|Z,E}(\mathbf{y}|\mathbf{z}, \boldsymbol{\varepsilon})\pi_{E|Z}(\boldsymbol{\varepsilon}|\mathbf{z})d\boldsymbol{\varepsilon} \\ &= \int_{\mathbb{R}^m} \delta(\mathbf{y} - A\mathbf{z} - \boldsymbol{\varepsilon})\pi_E(\boldsymbol{\varepsilon})d\boldsymbol{\varepsilon} \\ &= \pi_E(A\mathbf{z} - \mathbf{y}). \end{aligned}$$

From (2.17), we then have

$$\pi_{Z|Y}(\mathbf{z}|\mathbf{y}) \propto \pi_E(A\mathbf{z} - \mathbf{y})\pi_Z(\mathbf{z}). \quad (2.18)$$

All that we need now is the prior density π_Z . This probability distribution encodes *a priori* information regarding the unknown Z , which may take many forms. In some cases, we may wish to promote sparsity in the unknown Z or in some transform of the unknown $\psi(Z)$. The use of a Laplace prior or a spike-and-slab prior would then be beneficial [65]. We may also have prior knowledge of the structure of Z , in which case we may wish to utilize a Gaussian smoothness prior.

2.2.3. Markov Chain Monte Carlo Methods

When $\pi_{Z|Y}$ in (2.18) is not a distribution with a known sampling scheme, such as a Gaussian or Gamma distribution, the question becomes how to explore the posterior density to obtain useful information, such as point estimates and uncertainty quantification. Markov Chain Monte Carlo (MCMC) methods are a well-known class of methods for accomplishing this task.

Monte Carlo Integration. Let \mathbb{P} denote a probability measure over \mathbb{R}^n , and let f be a vector-valued measurable function over \mathbb{R}^n with respect to \mathbb{P} , i.e. $f \in L^1(\mathbb{P}(d\mathbf{x}))$. Now suppose that the objective is to calculate the integral of f with respect to the measure \mathbb{P} . Using numerical quadrature methods, we define a set of support points \mathbf{x}_j and corresponding weights w_j , $j = 1, \dots, N$, to get

$$\int_{\mathbb{R}^n} f(\mathbf{x})\mathbb{P}(d\mathbf{x}) \approx \sum_{j=1}^N f(\mathbf{x}_j)w_j.$$

In Monte Carlo integration, the support points \mathbf{x}_j are determined by drawing from some probability density, and then the weights w_j are calculated using that same distribution.

Suppose \mathbf{x}_j are drawn from the distribution determined by the measure \mathbb{P} itself. So if X is a \mathbb{P} -distributed random variable with random samples $\mathbf{x}_1, \dots, \mathbf{x}_N$, we could approximate the integral of f by the *ergodic average*

$$\int_{\mathbb{R}^n} f(\mathbf{x})\mathbb{P}(d\mathbf{x}) = E[f(X)] \approx \frac{1}{N} \sum_{j=1}^N f(\mathbf{x}_j). \quad (2.19)$$

An MCMC method generates a set of samples such that (2.19) holds.

Probability Transition Kernels. Let \mathcal{B} be the Borel sets over \mathbb{R}^n . Recall that these are any sets in \mathbb{R}^n that can be formed using the operations of countable union, countable intersection, and complement. A mapping $P : \mathbb{R}^n \times \mathcal{B} \rightarrow [0, 1]$ is called a *probability transition kernel* if

- for each $B \in \mathcal{B}$, the mapping $\mathbf{x} \mapsto P(\mathbf{x}, B)$ is a measurable function, and
- for each $\mathbf{x} \in \mathbb{R}^n$, the mapping $B \mapsto P(\mathbf{x}, B)$ is a probability distribution.

An ordered set $\{X_j\}_{j=1}^\infty$ of random variables $X_j \in \mathbb{R}^n$ is a *discrete time stochastic process*. A stochastic process $\{X_j\}_{j=1}^\infty$ with the properties

$$\mu_{X_{j+1}}(B_{j+1} | \mathbf{x}_1, \dots, \mathbf{x}_j) = \mu_{X_{j+1}}(B_{j+1} | \mathbf{x}_j) = P(\mathbf{x}_j, B_{j+1}) \quad (2.20)$$

is a *time-homogenous Markov chain*. The first equality in (2.20) can be understood as saying that, given all previous states of the Markov chain $X_1 = \mathbf{x}_1, \dots, X_j = \mathbf{x}_j$, the probability that $X_{j+1} \in B_{j+1}$ is the same as if we conditioned only on the immediate previous state $X_j = \mathbf{x}_j$. A common way of stating this property is that “the future depends on the past only through the present,” [73]. The second equality in (2.20) is regarding time homogeneity in the sense that the probability distribution of the next state based on the current state does not vary in time. To this end, notice that P does not have a dependence on $j + 1$.

The transition kernel that propagates k steps forward in time is defined inductively for $k \geq 2$ as

$$P^{(k)}(\mathbf{x}_j, B_{j+k}) = \mu_{X_{j+k}}(B_{j+k} | \mathbf{x}_j) = \int_{\mathbb{R}^n} P(\mathbf{x}_{j+k-1}, B_{j+k}) P^{(k-1)}(\mathbf{x}_j, d\mathbf{x}_{j+k-1}).$$

Introducing the operator $\mu_{X_j} P$, where μ_{X_j} denote the probability distribution of X_j ,

the probability distribution of X_{j+1} is given as

$$\mu_{X_{j+1}}(B_{j+1}) = \mu_{X_j}P(B_{j+1}) = \int_{\mathbb{R}^n} P(\mathbf{x}_j, B_{j+1})\mu_{X_j}(d\mathbf{x}_j).$$

A measure μ is an *invariant measure* of $P(\mathbf{x}_j, B_{j+1})$ if

$$\mu P = \mu.$$

A transition kernel P is *irreducible* if for each $\mathbf{x} \in \mathbb{R}^n$ and $B \in \mathcal{B}$ with $\mu(B) > 0$, there exists $k \in \mathbb{Z}_+$ such that $P^{(k)}(\mathbf{x}, B) > 0$. In other words, a transition kernel P is irreducible if every set of positive measure B is visited with positive probability by the Markov chain generated by P , regardless of starting point \mathbf{x} .

An irreducible kernel P is *periodic* if there is a set of disjoint nonempty sets $\{E_1, \dots, E_m\} \subset \mathbb{R}^n$, $m \geq 2$, such that for all $j = 1, \dots, m$ and $\mathbf{x} \in E_j$,

$$P(\mathbf{x}, E_{(j+1) \bmod m}) = 1.$$

This means that a periodic irreducible kernel is one that generates a Markov chain that will remain in an infinitely repeating loop. An *aperiodic* kernel is one that is not periodic.

The following theorem is directly stated from [73]:

Theorem 2.4. *Let μ be a probability measure in \mathbb{R}^n and $\{X_j\}$ a time-homogenous Markov chain with a transition kernel P . Assume further that μ is an invariant measure of the transition kernel P , and that P is irreducible and aperiodic. Then for all $\mathbf{x} \in \mathbb{R}^n$,*

$$\lim_{N \rightarrow \infty} P^{(N)}(\mathbf{x}, B) = \mu(B) \text{ for all } B \in \mathcal{B},$$

and for all $f \in L^1(\mu(d\mathbf{x}))$,

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{j=1}^N f(X_j) = \int_{\mathbb{R}^n} f(\mathbf{x}) \mu(d\mathbf{x})$$

almost certainly.

Thus to explore a given probability distribution, we need to (1) construct an aperiodic, invariant, and irreducible transition kernel P , and (2) draw a sequence of sample points $\mathbf{x}_1, \mathbf{x}_2, \dots$ using P .

In other words, we must derive a realization of the Markov chain. Two common MCMC methods are the Metropolis-Hastings algorithm and the Gibbs sampler. As Gibbs sampling is used in this dissertation, we briefly review it below. The reader is referred to [21, 45, 73] for more information regarding the Metropolis-Hastings algorithm.

The Gibbs Sampler. Let μ denote the target probability distribution in \mathbb{R}^n , and assume μ is absolutely continuous with respect to the Lebesgue measure such that $\mu(d\mathbf{x}) = \pi(\mathbf{x})d\mathbf{x}$. As a consequence of Theorem 2.4, we require a transition kernel $P(\mathbf{x}, B)$ such that μ is its invariant measure and that the resulting Gibbs Markov chain is irreducible and aperiodic.

Let P denote a transition kernel given by

$$P(\mathbf{x}, B) = \int_B K(\mathbf{x}, \mathbf{y}) d\mathbf{y}. \quad (2.21)$$

We can think of $K(\mathbf{x}, \mathbf{y})d\mathbf{y}$ as the probability of the move from \mathbf{x} to the infinitesimal set $d\mathbf{y}$ at \mathbf{y} . The condition $P(\mathbf{x}, \mathbb{R}^n) = 1$ provides

$$\int_{\mathbb{R}^n} K(\mathbf{x}, \mathbf{y}) d\mathbf{y} = 1,$$

which then implies

$$\int_{\mathbb{R}^n} K(\mathbf{y}, \mathbf{x})\pi(\mathbf{y})d\mathbf{x} = \pi(\mathbf{y}). \quad (2.22)$$

In order for $\pi(\mathbf{x})d\mathbf{x}$ to be an invariant measure of P , we must have

$$\mu P(B) = \mu(B) = \int_B \pi(\mathbf{y})d\mathbf{y}.$$

Since

$$\mu P(B) = \int_{\mathbb{R}^n} \left(\int_B K(\mathbf{x}, \mathbf{y})d\mathbf{y} \right) \pi(\mathbf{x})d\mathbf{x} = \int_B \left(\int_{\mathbb{R}^n} K(\mathbf{x}, \mathbf{y})\pi(\mathbf{x})d\mathbf{x} \right) d\mathbf{y}$$

for all B , it must be that

$$\int_{\mathbb{R}^n} K(\mathbf{x}, \mathbf{y})\pi(\mathbf{x})d\mathbf{x} = \pi(\mathbf{y}). \quad (2.23)$$

Combining (2.22) with (2.23), we have the sufficient condition

$$\int_{\mathbb{R}^n} K(\mathbf{x}, \mathbf{y})\pi(\mathbf{x})d\mathbf{x} = \int_{\mathbb{R}^n} K(\mathbf{y}, \mathbf{x})\pi(\mathbf{y})d\mathbf{x}.$$

This condition is called the *balance equation*, and it is defined as follows:

Definition 2.2. Given densities $K(\mathbf{x}, \mathbf{y})$ and $\pi(\mathbf{x})$, the *balance equation* is the condition that

$$\int_{\mathbb{R}^n} K(\mathbf{x}, \mathbf{y})\pi(\mathbf{x})d\mathbf{x} = \int_{\mathbb{R}^n} K(\mathbf{y}, \mathbf{x})\pi(\mathbf{y})d\mathbf{x}. \quad (2.24)$$

We now introduce some notation needed for our discussion of the Gibbs sampler. Let $\mathcal{I} = \{1, \dots, n\}$ be the index set in \mathbb{R}^n , and let $\mathcal{I}_j, j = 1, \dots, m$ be a partitioning

of the index set into disjoint nonempty subsets with $k_j = |\mathcal{I}_j|$ such that $\cup_{j=1}^m \mathcal{I}_j = \mathcal{I}$.

We then partition \mathbb{R}^n as

$$\mathbb{R}^n = \mathbb{R}^{k_1} \times \cdots \times \mathbb{R}^{k_m}.$$

For any vector $\mathbf{x} \in \mathbb{R}^n$, we then rearrange the elements of \mathbf{x} such that

$$\mathbf{x} = \begin{bmatrix} \mathbf{x}_{\mathcal{I}_1} \\ \vdots \\ \mathbf{x}_{\mathcal{I}_m} \end{bmatrix} \in \mathbb{R}^n, \quad \text{where } \mathbf{x}_{\mathcal{I}_j} \in \mathbb{R}^{k_j}, \quad j = 1, \dots, m.$$

Hence $x_i \in \mathbb{R}$ is a component of $\mathbf{x}_{\mathcal{I}_j}$ if and only if $i \in \mathcal{I}_j$. Note that if $k_j = 1$ for all $j = 1, \dots, m$, then $m = n$ and $\mathbf{x}_{\mathcal{I}_j} = x_j \in \mathbb{R}$.

We use the negative subindex to indicate that the corresponding elements are deleted from the vector, i.e.

$$\mathbf{x}_{-\mathcal{I}_j} = \left[\mathbf{x}_{\mathcal{I}_1}^T \quad \cdots \quad \mathbf{x}_{\mathcal{I}_{j-1}}^T \quad \mathbf{x}_{\mathcal{I}_{j+1}}^T \quad \cdots \quad \mathbf{x}_{\mathcal{I}_m}^T \right]^T.$$

With this notation in place, suppose we have some partitioning of the index set \mathcal{I} of \mathbb{R}^n , and let $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$. We define a kernel K such that

$$K(\mathbf{x}, \mathbf{y}) = \prod_{j=1}^m \pi(\mathbf{y}_{\mathcal{I}_j} | \mathbf{y}_{\mathcal{I}_1}, \dots, \mathbf{y}_{\mathcal{I}_{j-1}}, \mathbf{x}_{\mathcal{I}_{j+1}}, \dots, \mathbf{x}_{\mathcal{I}_m}). \quad (2.25)$$

This leads us to the following theorem:

Theorem 2.5. *Consider the probability distribution μ in \mathbb{R}^n with probability density function $\pi(\mathbf{x})$. If P is the transition kernel in (2.21), where K is given by (2.25), then P satisfies the balance equation (2.24).*

The reader is referred to [73, Chapter 3] for a proof of Theorem 2.5. Since K

satisfies the balance equation, the transition kernel $P(\mathbf{x}, B)$ is an invariant measure of μ .

To satisfy Theorem 2.4, we still must show that the resulting Gibbs Markov chain is both irreducible and aperiodic. For the chain to be irreducible, it is sufficient for the associated probability density π to satisfy the *positivity condition*.

Definition 2.3. Let (X_1, \dots, X_n) be random variables with joint probability density $\pi(x_1, \dots, x_n)$, and let $\pi^{(i)}$ denote the marginal density of X_i . If $\pi^{(i)}(x_i) > 0$ for every $i = 1, \dots, n$ implies that $\pi(x_1, \dots, x_n) > 0$, then π is said to satisfy the *positivity condition*.

In other words, the positivity condition implies that the support of π is the same as the Cartesian product of the support of each $\pi^{(i)}$, $i = 1, \dots, n$. If $\pi(\mathbf{x})$ satisfies the positivity condition, then the resulting Gibbs Markov chain is irreducible [106, Chapters 9 and 10].

To see how the Gibbs Markov chain is aperiodic, consider the state $\mathbf{x}^{(t)}$ of the Markov chain. Since $\pi(\mathbf{x}^{(t)}) > 0$, each $\pi(\mathbf{x}_{\mathcal{I}_j}^{(t)} | \mathbf{x}_{\mathcal{I}_1}^{(t)}, \dots, \mathbf{x}_{\mathcal{I}_{j-1}}^{(t)}, \mathbf{x}_{\mathcal{I}_{j+1}}^{(t)}, \dots, \mathbf{x}_{\mathcal{I}_m}^{(t)}) > 0$ for $j = 1, \dots, m$ as well. Thus, the next state $\mathbf{x}^{(t+1)}$ may be the same as the current state $\mathbf{x}^{(t)}$ with positive probability, implying that the chain is aperiodic.

A summary of how to perform the Gibbs sampler is given in Algorithm 1.

Algorithm 1 The Gibbs Sampler

Input probability density $\pi(\mathbf{x})$, index sets \mathcal{I}_j for $j = 1, \dots, m$, number of iterations K , and initial value $\mathbf{x}^{(0)}$.

Output Markov chain $\{\mathbf{x}^{(k)}\}_{k=1}^K$.

for $k = 1 : K$ **do**

 Set $\mathbf{x} = \mathbf{x}^{(k)}$.

for $j = 1 : m$ **do**

 Draw $\mathbf{y}_{\mathcal{I}_j} \in \mathbb{R}^{k_j}$ from the density $\pi(\mathbf{y}_{\mathcal{I}_j} | \mathbf{y}_{\mathcal{I}_1}, \dots, \mathbf{y}_{\mathcal{I}_{j-1}}, \mathbf{x}_{\mathcal{I}_{j+1}}, \dots, \mathbf{x}_{\mathcal{I}_m})$.

end for

 Set $\mathbf{x}^{(k+1)} = \mathbf{y}$.

end for

Section 2.3

Synthetic Aperture Radar Signal Processing

Synthetic aperture radar is a day and night, all-weather imaging modality capable of providing high resolution images over large areas of interest. This dissertation is concerned with *spotlight-mode* synthetic aperture radar, henceforth referred to as *SAR*. The following description of the SAR signal processing problem is adapted from [34, 66, 91, 109].

In SAR imaging, a platform takes a full or partial circular path around a scene of interest, transmitting and receiving electromagnetic pulses as it does so. Denote by θ and φ the respective azimuth and depression viewing angles of the imaging platform. Let the volumetric scene be described with view-independent x, y, z coordinates. Define the view-dependent coordinate system u, v, w to be such that the u -axis contains the imaging platform and the scene center, the v -axis is orthogonal to the u -axis and is contained in the xy -plane, and the w -axis is normal to the uv -plane. A graphical depiction of this setup can be seen in Figure 2.3.

Suppose that the radar attached to the imaging platform launches a linear FM chirp waveform described by $\text{Re}(s(t))$ [66], where

$$s(t) = \begin{cases} \exp(i(\omega_0 t + \alpha t^2)) & |t| \leq \tau_c/2 \\ 0 & \text{else} \end{cases}, \quad (2.26)$$

τ_c is the pulse duration, 2α is the FM chirp rate, and ω_0 is the center frequency. Figure 2.4 demonstrates an example of a linear FM chirp.

Let $g(\mathbf{x}) = g(x, y, z)$ be the scene reflectivity function. Consider the interval $[-u_1, u_1]$ along the u -axis, where $u_1 \in \mathbb{R}_+$ is the maximum *slant range* of any target illuminated in the scene. Let τ_p be the *patch propagation time*, which is the differ-

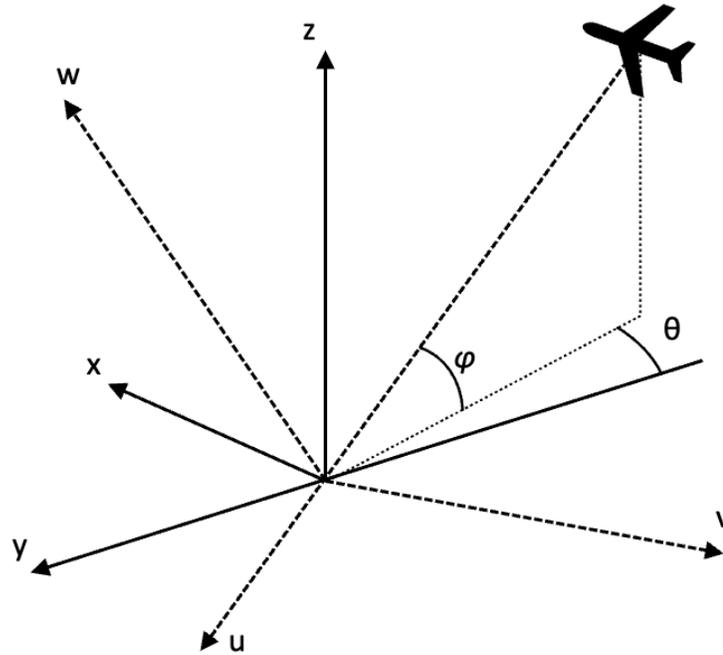


Figure 2.3: The geometry of 3D SAR data collection.

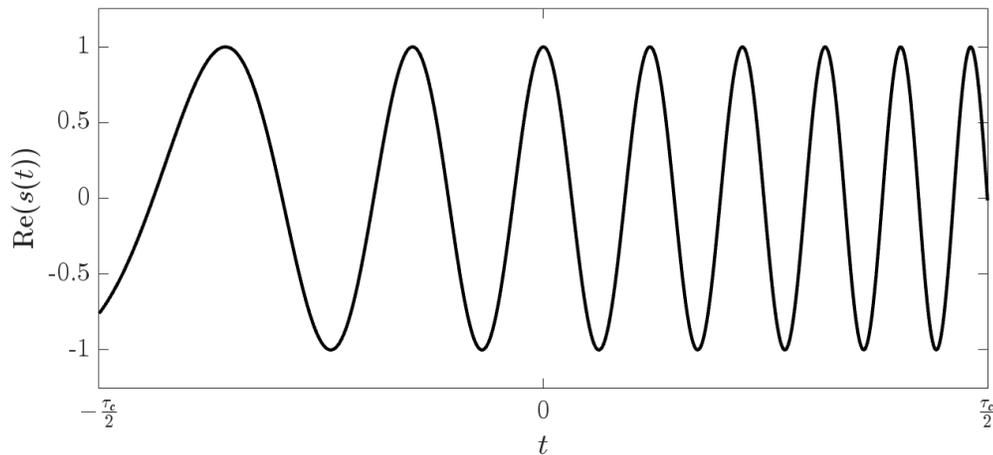


Figure 2.4: Linear FM chirp.

ence in two-way propagation delay between a target closest to and farthest from the imaging platform in the illuminated scene.

We approximate the signal propagation using a planar wavefront. At any given time, what is transduced by a single received pulse is an integration of the scene

reflectivity function g over the plane given by slant range u . Thus we define $p_{\theta,\varphi}$ as the view-dependent projection function given as

$$p_{\theta,\varphi}(u) = \int \int g(x(u, v, w), y(u, v, w), z(u, v, w)) dv dw.$$

The chirp return echo $r_{\theta,\varphi}$ can be seen as a superposition of multiple delayed versions of the chirp waveform, where the scaling of each version is proportional to the view-dependent projection function $p_{\theta,\varphi}$, i.e.

$$r_{\theta,\varphi}(t) = A \operatorname{Re} \left(\int_{-u_1}^{u_1} p_{\theta,\varphi}(u) s \left(t - \frac{2(R+u)}{c} \right) du \right), \quad (2.27)$$

where R is the distance from the origin to the imaging platform, c is the speed of light, and A is a scaling factor. Note that (2.27) is only valid when

$$\tau_0 - \frac{\tau_c}{2} + \frac{\tau_p}{2} \leq t \leq \tau_0 + \frac{\tau_c}{2} - \frac{\tau_p}{2}. \quad (2.28)$$

The inequalities in (2.28) indicate the time segment during which chirp returns from all targets in the illumination patch exist simultaneously.

Rewriting (2.27) using (2.26) along with the relations

$$\tau_0 = \frac{2R}{c} \quad \text{and} \quad \tau(u) = \frac{2u}{c} \quad (2.29)$$

yields

$$r_{\theta,\varphi}(t) = A \operatorname{Re} \left(\int_{-u_1}^{u_1} p_{\theta,\varphi}(u) \exp(i(\omega_0(t - \tau_0 - \tau(u)) + \alpha(t - \tau_0 - \tau(u))^2)) du \right). \quad (2.30)$$

Deramp processing is used to deconvolve the chirp signal $s(t)$ from the chirp

return echo $r_{\theta,\varphi}(t)$. This process begins with *quadrature demodulation*. The first step in quadrature demodulation is to mix (multiply) $r_{\theta,\varphi}(t)$ with delayed in-phase (real) and quadrature (imaginary) versions of $s(t)$. Multiplying $r_{\theta,\varphi}(t)$ with the in-phase version of $s(t - \tau_0)$, that is

$$\operatorname{Re}(s(t - \tau_0)) = \operatorname{Re}(\exp(i(\omega_0(t - \tau_0) + \alpha(t - \tau_0)^2))) = \cos(\omega_0(t - \tau_0) + \alpha(t - \tau_0)^2),$$

produces the *in-phase* term of the mixer output as

$$\begin{aligned} \tilde{r}_{\theta,\varphi I} = & \frac{A}{2} \operatorname{Re} \left(\int_{-u_1}^{u_1} p_{\theta,\varphi}(u) \exp \left(i \left(\omega_0(2t - \tau(u) - 2\tau_0) \right. \right. \right. \\ & \left. \left. \left. + \alpha((t - \tau_0)^2 + (t - \tau(u) - \tau_0)^2) \right) \right) du \right) \\ & + \frac{A}{2} \operatorname{Re} \left(\int_{-u_1}^{u_1} p_{\theta,\varphi}(u) \exp \left(i \left(\alpha\tau(u)^2 - \tau(u)(\omega_0 + 2\alpha(t - \tau_0)) \right) \right) du \right). \end{aligned} \quad (2.31)$$

Observe that the first term in (2.31) is centered on the frequency $2\omega_0$. This leads to the second step of quadrature demodulation, which is to apply a low-pass filter to $\tilde{r}_{\theta,\varphi I}$ to remove terms with large frequencies. The output of this filter is then

$$\bar{r}_{\theta,\varphi I}(t) = \frac{A}{2} \operatorname{Re} \left(\int_{-u_1}^{u_1} p_{\theta,\varphi}(u) \exp \left(i \left(\alpha\tau(u)^2 - \tau(u)(\omega_0 + 2\alpha(t - \tau_0)) \right) \right) du \right). \quad (2.32)$$

We similarly obtain the *quadrature term* $\bar{r}_{\theta,\varphi Q}$ by mixing $r_{\theta,\varphi}$ with the quadrature

component of $s(t - \tau_0)$ and filtering with a low-pass filter:

$$\bar{r}_{\theta,\varphi Q}(t) = \frac{A}{2} \operatorname{Im} \left(\int_{-u_1}^{u_1} p_{\theta,\varphi}(u) \exp \left(i \left(\alpha \tau(u)^2 - \tau(u)(\omega_0 + 2\alpha(t - \tau_0)) \right) \right) du \right). \quad (2.33)$$

Adding together the right-hand-sides of (2.32) and (2.33) yields the resulting signal

$$C_{\theta,\varphi} = \frac{A}{2} \int_{-u_1}^{u_1} p_{\theta,\varphi}(u) \exp \left(i \left(\alpha \tau(u)^2 - \tau(u)(\omega_0 + 2\alpha(t - \tau_0)) \right) \right) du. \quad (2.34)$$

If we ignore the second-order term $\alpha \tau(u)^2$ in (2.34) and apply the definition of $\tau(u)$ (2.29), we obtain

$$\bar{C}_{\theta,\varphi} = \frac{A}{2} \int_{-u_1}^{u_1} p_{\theta,\varphi}(u) \exp \left(-iu \frac{2(\omega_0 + 2\alpha(t - \tau_0))}{c} \right) du. \quad (2.35)$$

Notice the presence of the Fourier transform kernel $\exp(-iuU)$ in the integrand in (2.35), where

$$U = \frac{2(\omega_0 + 2\alpha(t - \tau_0))}{c}.$$

Therefore, $\bar{C}_{\theta,\varphi}$ is the Fourier transform of the view-dependent projection function $p_{\theta,\varphi}$ evaluated over the limited range of spatial frequencies

$$\frac{2(\omega_0 - \alpha(\tau_c - \tau_p))}{c} \leq U \leq \frac{2(\omega_0 + \alpha(\tau_c - \tau_p))}{c}$$

according to (2.28). We will make extensive use of this equivalence when we discuss 3D SAR imaging techniques in Chapter 6.

Chapter 3

Sparsity-Based Reconstruction of 3D Photoacoustic Images

This chapter describes work that was published in [55] in the *Journal of Imaging*.

Section 3.1

Introduction

Photoacoustic (PA) imaging provides a method of in vivo non-invasive and high-resolution molecular imaging at centimeter depth scales [84, 126, 129, 133]. In PA imaging, biological tissue is irradiated by a pulsed laser. The absorption of the laser by endogenous or exogenous chromophores induces a local increase in temperature, which in turn causes a pressure rise through thermoelastic expansion of the tissue. Ultrasound receivers placed at the surface of the tissue detect the resulting acoustic waves. Images of the optical absorption can be reconstructed by solving acoustic and optical inverse problems [37]. While high-speed data acquisition is possible with PA imaging, the multichannel data acquisition systems that are available to record this data are expensive [127]. The detection of PA signals is most commonly accom-

plished using digital-to-analog converters recording the voltage connected to piezoelectric transducers, which drastically increase the complexity of the imaging system [71]. A well-established alternative to this is a method for optical interferometric detection which utilizes a Fabry–Pérot etalon (FPE) [12, 63, 137]. The FPE exploits the resonant interference of an interrogating continuous wave laser between two reflecting surfaces to provide both sensitive detection and large acoustic bandwidth.

A major limitation to using a FPE for PA signal detection, however, is that detection is performed only at a single point on the etalon. In practice, the detection laser must be raster scanned along the surface of the etalon to acquire a volumetric image of the tissue. Thus, the PA signal generating laser must be shot for each position in the raster scan to produce the photoacoustic signal, making the temporal resolution highly dependent on the pulse repetition rate of the laser. This dependence makes the imaging system prone to motion artifacts and unable to capture fast dynamic processes.

In order to improve the temporal resolution of the imaging system, recent work has been done in which a FPE was imaged onto a scrambled Hadamard pattern over multiple sequential measurements [62, 77]. Though the main features of imaging phantoms used were successfully recovered at compression rates as low as 10%, multiple acquisitions of data were necessary, and thus the total data acquisition time was still much longer than pulse repetition rate of the laser.

Here, FPE-based PA image detection is combined with compressed ultrafast photography to further improve the temporal resolution. Compressed ultrafast photography applies core compressed sensing principles to acquire a sequence of images at a high rate using a single exposure of a camera [103]. In the original work, a random binary mask is applied to the image, and a streak camera is then used to scan the image quickly across a sensor [46]. The mask in conjunction with the streak camera

provide enough structure such that the original time-resolved image sequence can be reconstructed. This has enabled reconstruction of 980-frame videos with framerates achieving 7×10^{13} frames per second [128]. In applications that do not require such extreme framerates, the streak camera can be replaced with a low-cost galvanometer to achieve the spatial shifting of the image [78].

In this work, we design and simulate an optical system that is capable of acquiring the interference pattern of the entire FPE with a sampling rate of greater than 12 MHz. This optical system utilizes a digital micromirror device to apply a binary mask to the interference pattern of the FPE. The resulting masked image is then rapidly swept across an imaging sensor with a galvanometer, thus encoding time information in the spatial domain. Finally, the PA image is reconstructed using a compressive sensing approach that iteratively solves a convex optimization problem specifically designed for problems where data are under-sampled and the true solution has sparse representation in some related domain (e.g., the gradient domain).

The rest of this chapter is organized as follows: In Section 3.2, we describe the problem and provide the background in compressed sensing needed in our new approach. The results of the simulations are presented in Section 3.3, with a discussion of these findings and some concluding remarks in Section 3.4.

Section 3.2

Methods

3.2.1. Optical Setup

The proposed optical system is depicted in Figure 3.1. A pulsed laser is applied to the tissue of interest. Optical absorbers in the tissue then convert the optical energy to heat, inducing a thermoelastic expansion of the surrounding tissue. The expansion generates broadband ultrasound waves, which are detected at the surface of the tissue

with a FPE. This interaction of the PA waves with the surface of the FPE results in modulation of the reflected interrogating continuous-wave (CW) laser beam on the opposite side of the FPE. The linear polarizer and quarter wave plates allow for the directed flow of the light through the polarized beam splitters.

The modulated CW laser beam is then imaged onto a digital micromirror device (DMD) via a 4-f optical system, which consists of two lenses spaced by twice their focal length. The DMD consists of a two-dimensional array of mirrors that will either reflect the light along the optical path or deflect it away from it, resulting in a binary mask on the FPE. A second 4-f optical system then images the DMD directly onto an imaging sensor array. A galvanometer is placed in the Fourier plane of the second 4-f system, which rapidly sweeps the image of the masked FPE across the camera during a single exposure.

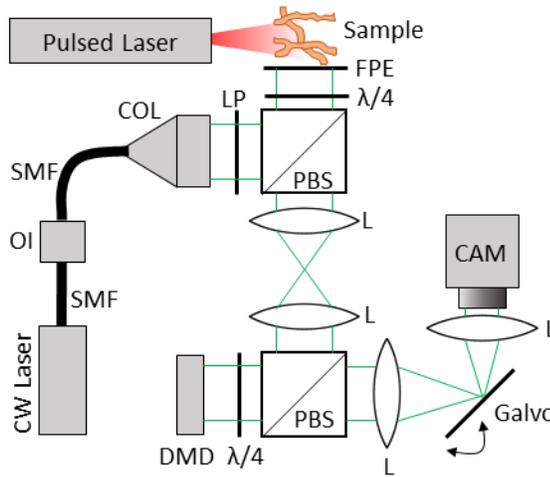


Figure 3.1: The proposed optical system. SMF = single-mode fiber, OI = optical isolator, COL = collimator, PBS = polarized beam splitter, $\lambda/4$ = quarter wave plate, L = lens, CAM = camera, DMD = digital micromirror device, LP = linear polarizer, FPE = Fabry-Pérot etalon.

3.2.2. Continuous Model

We now seek to build a forward model that transforms a given initial pressure distribution (IPD) to a camera image based on the proposed optical system. Let $d, \tau \in \mathbb{R}^+$.

The function $P(x, y, z, t)$ is defined on $[0, d] \times [0, d] \times [0, d] \times [0, \tau]$ to be the pressure distribution at (x, y, z) at time t . The IPD is then $P_0 := P(x, y, z, 0)$. Acoustic waves then propagate outwards in a manner determined by the governing equations

$$\begin{aligned}\frac{\partial \vec{u}}{\partial t} &= -\frac{1}{\rho_0} \nabla P \\ \frac{\partial \rho}{\partial t} &= -\rho_0 \nabla \cdot \vec{u} \\ P &= c_0^2 \rho.\end{aligned}$$

Here \vec{u} is the acoustic particle velocity, ρ is the density, ρ_0 is the density in the absence of acoustic waves, and c_0 is the isentropic sound speed [101].

The FPE is placed on the xy -plane and encodes the pressure data $P(x, y, 0, t)$. We next define the binary mask $M \subset \mathbb{R}^2$. The interaction of the light from the FPE with the DMD can be characterized as

$$P_M(x, y, t) = \begin{cases} P(x, y, 0, t) & \text{if } (x, y) \in M \\ 0 & \text{else.} \end{cases} \quad (3.1)$$

The data then undergoes a shearing operation from the motion of the galvanometer, leading to

$$P_S(x, y, t) = P_M(x - \alpha t, y, t), \quad (3.2)$$

where it is assumed that the galvanometer sweeps with a constant speed $\alpha > 0$, where α is determined by the physical limitations of the galvanometer and the focal length of the lenses in the second 4-f system.

Lastly, P_S undergoes a temporal integration operation as it is swept across the camera sensor for an exposure time equal to the acoustic wave propagation time τ ,

yielding the camera image

$$E(x, y) = \int_0^\tau P_S(x, y, t) dt = \int_0^\tau P_M(x - \alpha t, y, t) dt, \quad (3.3)$$

where $P_M(x, y, t)$ is given in (3.1). This describes the full continuous forward model.

3.2.3. Discrete Model

We now move to discretize (3.3) to enable solving the inverse problem. First, the IPD is discretized into a uniform three-dimensional computational grid of size $N \times N \times N$ for a given choice of $N \in \mathbb{Z}$. The dimension of each voxel is then $h \times h \times h$ where $h = \frac{d}{N}$. The grid elements are then rearranged to form a single vector u of length N^3 . Next, we model the propagation of the acoustic waves through body tissue over time using the k-Wave simulation toolbox in MATLAB [122]. The propagation is also temporally discretized with time steps Δt determined by the Courant–Friedrichs–Lewy condition, which is dependent on c_0 and h . The total number of time steps $T \in \mathbb{Z}$ is then calculated as $T = \lfloor \frac{\tau}{\Delta t} \rfloor$. The acoustic waves are observed by the FPE located at the base of the computational grid at each time step. This transformation from the IPD to the sequence of T images of size $N \times N$ detected by the FPE can be modeled by the $N^2 T \times N^3$ matrix \mathbf{K} , which is constructed by simulating the FPE output for each of the standard basis vectors in \mathbb{R}^{N^3} .

The binary mask M used in (3.1) is now discretized to form \mathbf{M}' , and is defined as

$$\mathbf{M}'_{i,j} = \begin{cases} 1 & \text{if } (ih, jh) \in M \\ 0 & \text{else,} \end{cases}, \quad i, j = 1, \dots, N. \quad (3.4)$$

The diagonal matrix $\mathbf{M} \in \mathbb{R}^{N^2 T \times N^2 T}$ is subsequently formed by reshaping \mathbf{M}' into an $N^2 \times 1$ vector and inserting it into \mathbf{M} such that $\mathbf{M}_{j+iN^2, j+iN^2} = \mathbf{M}'_j$ for $j = 1, \dots, N^2$

and $i = 1, \dots, (T - 1)$.

The shearing operation is then applied, which we write as matrix \mathbf{S} , where the sequence of images is shifted (spatially) along the x -axis as a function of time. Since the shearing speed has, in practice, a significantly greater magnitude than Δt , down-sampling is also performed during this step. This is accomplished by calculating the downsize factor $s = \lceil \frac{\alpha}{\Delta t} \rceil$. For every s entries, all but one entry is discarded so that \mathbf{S} is a matrix of size $(\lfloor \frac{T}{s} \rfloor + N - 1) N \lfloor \frac{T}{s} \rfloor \times N^2 T$.

Lastly, the light intensity incident to each pixel is summed over time using a left Riemann sum by the $(\lfloor \frac{T}{s} \rfloor + N - 1) N \times (\lfloor \frac{T}{s} \rfloor + N - 1) N \lfloor \frac{T}{s} \rfloor$ matrix \mathbf{I} , resulting in the camera image v . Since \mathbf{I} , \mathbf{S} , \mathbf{M} , and \mathbf{K} are matrices, the full forward model can thus be represented as

$$w = \mathbf{A}v + e \tag{3.5}$$

where $\mathbf{A} = \mathbf{ISMK}$, w is a vector of length NL , and e is a vector of additive white Gaussian noise with mean zero and covariance matrix $\mathbb{I}_{NL}\sigma^2$. Although the matrices \mathbf{I} , \mathbf{S} , and \mathbf{K} are not analytically constructed, one can explicitly form \mathbf{A} , as will be described in Section 3.2.5. The parameter L is determined by the angular velocity of the galvanometer, and with a constant angular velocity, we have from above that $L = \lfloor \frac{T}{s} \rfloor + N - 1$. Figure 3.2 indicates how each component of \mathbf{A} affects the image.

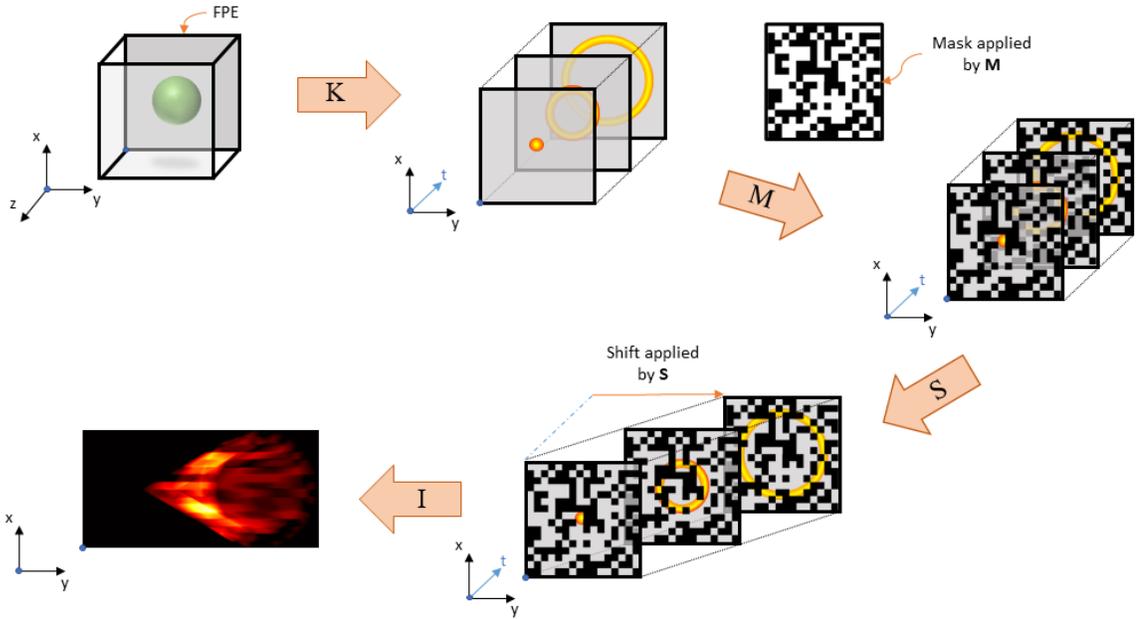


Figure 3.2: A pictorial representation of the imaging process as individual forward operations.

3.2.4. Image Reconstruction Via Compressed Sensing

To reconstruct the IPD, we utilize ideas and algorithms from compressed sensing, which is based on the key notion that a sparse signal can be reconstructed with relatively few measurements. Following the seminal work in [24, 41], many investigations have centered around compressed sensing algorithms – sometimes with the goal of generally improving the methodology, e.g., its efficiency, robustness, and accuracy, and in other cases to use the method for a particular application of interest. For example, compressed sensing has been extensively used in the area of PA image reconstruction [7, 14, 58, 59].

Compressed sensing requires that the image satisfy certain sparsity and incoherence constraints [75]. They are (i) the image should contain only a few nonzero values in some domain, known as the sparse domain, and (ii) the image acquisition domain should not be coherent with the sparse domain, i.e. the rows of the measurement matrix are not correlated with the columns of the sparsifying matrix. The origi-

nal image can then be accurately reconstructed from under-sampled data using an iterative method that utilizes a data fidelity term and includes a sparsity constraint.

As we are attempting to reconstruct a vector of length N^3 with one of length NL , where $L < N^2$, this inverse problem is under-determined. Note that in practice, $L \ll N^2$, so the compression ratio is quite high. In this work, we arrive at the L/N^2 ratio of $163/4096 \approx 25$. We will address this issue using a compressive sensing approach, which, as noted above, requires that the IPD is sparse in some domain. Since it is anticipated that real-world applications will consider IPDs that are approximately piecewise-constant, the sparsity-enforcing regularization term Φ is chosen as the isotropic discrete total variation (TV) operator. This leads us to the convex optimization problem

$$v = \arg \min_{\tilde{v} \geq 0} \frac{1}{2} \|\mathbf{A}\tilde{v} - w\|_2^2 + \lambda\Phi(\tilde{v}), \quad (3.6)$$

where \mathbf{A} is described in (3.5), λ is the regularization parameter, and the problem is augmented by the physical constraint $v \geq 0$.

Many algorithms have been developed to numerically solve (3.6), and there have also been numerous investigations into parameter selection [42]. For the simulations we employ the Two-Step Iterative Shrinkage/Thresholding (TwIST) algorithm to recover v in (3.6) [15]. This method brings together the high denoising capabilities of iterative shrinkage/thresholding (IST) and the efficiency for dealing with ill-posed problems of iterative reweighted shrinkage (IRS) algorithms. IST has good denoising properties, while IRS is good at handling ill-posed problems, and TwIST aims at keeping both these advantages. Since the IPD is expected to be nonnegative, we have modified TwIST such that after each iteration, all negative values are set to zero. This modification is not trivial, and further investigation is needed to quantify the stability and accuracy of this revised method. The unmodified TwIST algorithm is

commonly used in related applications [77, 114].

To compensate for the finite resolution effects and aliasing arising from discretization of the k-Wave simulation, reconstructed IPD is normalized prior to quantitative comparison. We acknowledge that this is not necessarily the best way to treat the error, but the approach is effective in our experiments. More extensive study is required to identify a better mitigating technique. The two main criteria used to quantify the success of the reconstruction are mean square error (MSE) and multi-scale structural similarity (MS-SSIM) index. The MS-SSIM index incorporates image details at different resolutions to provide an image quality assessment based on the human visual system [130]. For MSE, a smaller number indicates less error, while MS-SSIM is between -1 and 1 , with 1 indicating a perfect reconstruction.

3.2.5. Simulation Setup and Analysis

In addition to the parameters defined previously, there are several other important parameters to discuss for the simulations. The size of the computational grid for the k-Wave simulation is $N_c > N$, where we apply a perfectly matched layer absorbing boundary condition to the edges of the computational grid. The layer occupies a strip of size N_L grid points around the outer perimeter of the computational domain. The speed of sound in the medium containing the IPD is c_s , the length of each voxel is h , and the center frequency is f .

We now describe how the resulting image is computed from the optical setup given an IPD. An N_c^3 computational grid is created on which to run the k-Wave simulation. The FPE is incorporated as a N^2 sensor placed parallel to the xy-plane in the extended computational grid with corner at $(\frac{N_c-N}{2} + 1, \frac{N_c-N}{2} + 1, \frac{N_c-N}{2} + 1)$. The speed of sound is defined corresponding to the average speed of sound in human tissue [83]. The sensor data is stored in a $N \times N \times T$ array, and we proceed as described in 3.2.4. The parameters chosen for the simulations are defined in Table

3.1.

Parameter	Description	Value
N	length of IPD grid [pixels]	64
N_c	length of computational grid [pixels]	96
N_L	width of boundary condition layer [pixels]	15
c_s	speed of sound in the medium [m/s]	1540
h	pixel width [μm]	122
f	center frequency [MHz]	5
T	total time steps	400
Δt	time step size [ns]	24
α	[pixels-widths/s]	12
s	downsize factor	4

Table 3.1: The parameters used in our simulations and their associated values.

To test our new method we will consider (1) the base case cylinder IPD, (2) the base case cylinder IPD rotated so that its axis is parallel to the x -axis, and (3) a vessel-like IPD with ten total vessels. The rotated cylinder is included to analyze how the orientation of the cylinder relative to the direction of shearing affects reconstruction. In our simulations, the direction of shearing is parallel to the x -axis. In this analysis, the probability that a given pixel, m , in the mask is set to one is varied. For this purpose we define

$$p = \text{Prob}(m = 1). \quad (3.7)$$

The cylinder is passed through the forward model and then added noise to the final image before attempting reconstruction. The regularization parameter in (3.6) was chosen as $\lambda = 2.5 \times 10^{-3}$, which was optimized heuristically for the no noise case. The

reconstruction is performed for various levels of signal-to-noise ratio (SNR), which is defined as

$$\text{SNR} = 20 \log_{10} \left(\frac{\mu}{\sigma} \right) \text{ dB}, \quad (3.8)$$

where μ is the mean of the signal strength in its area of support and σ is the standard deviation of the noise present. Note that μ is approximated over the inferred region of support of the image.

Section 3.3

Numerical Results

3.3.1. Baseline Test

As a baseline test, and to demonstrate that our methods perform as expected, Figure 3.3 displays the reconstruction of an IPD of a single impulse; that is, it contains all zeros except for a single voxel that is set to a value of one. Since the matrix \mathbf{A} is constructed by passing each basis vector through the forward model, this is a good test to see if the reconstruction algorithm is working as expected. We see in Figure 3.3 that, as predicted, the reconstruction of the single impulse is spread across neighboring pixels with a peak at the true impulse pixel.

3.3.2. Simulated Experiments

Having established our method performs as expected in the simple impulse case, we now consider two primary types of more realistic IPDs, namely cylindrical IPDs and vessel-like IPDs, each of which has binary-valued voxels representing either the presence or lack thereof of an initial acoustic impulse response.

For the base case of the cylindrical IPDs, shown in Figure 3.4a, a solid cylinder with a radius of six voxels, including the center voxel, and with axis parallel to the y-

axis and centered on the xz -plane is used. The vessel-like IPDs, an example of which is displayed in Figure 3.4b, are constructed with a more random behavior meant to simulate the structure of blood vessels in tissue, including growing and shrinking as well as branching vessels.

In our experiments, we used Gaussian noise with mean zero and variance dependent on the desired SNR parameter given by (3.8). For each value of SNR examined, the image reconstruction was performed five times (after adding noise as described in Section 3.2). The average value of MS-SSIM and MSE over those five trials was then computed. Shown in Figure 3.5 are the MSE and MS-SSIM results for the reconstruction of the cylinder IPD parallel to the y -axis, the cylinder IPD parallel to the x -axis, and the vessel IPD, all for different values of p in (3.7), as well as one using the normalized time reversal reconstruction included with the k-Wave toolbox. The k-Wave reconstruction is performed without compression on the data acquired from the FPE with the appropriate noise added. For SNR greater than 0 dB, we observe that our method performs consistently better than the k-Wave reconstruction for p values between 0.2 and 0.9, and the accuracy of the reconstruction tends to increase faster with our method than with the k-Wave reconstruction as the SNR increases.

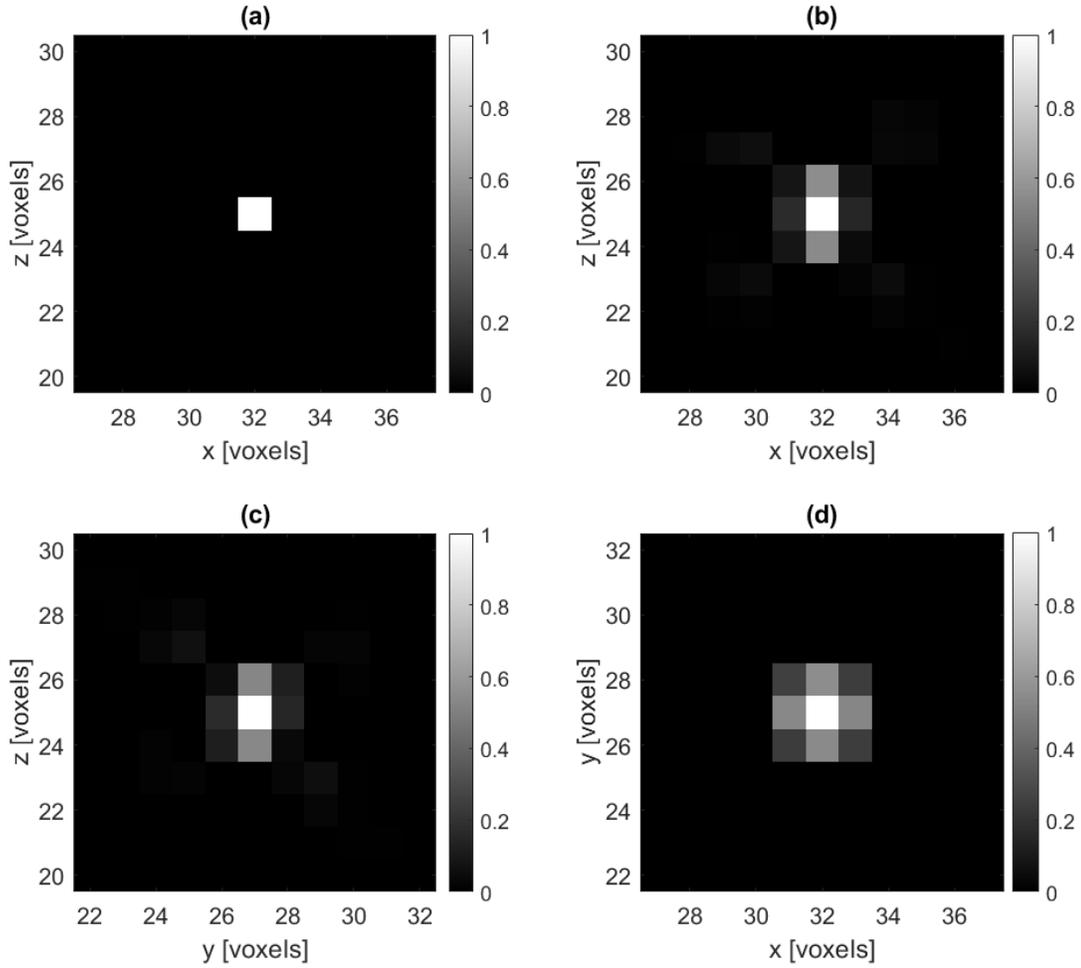


Figure 3.3: (a) A cross-section of the ground truth impulse IPD and (b–d) cross-sections orthogonal to the y -, x -, and z -axes, respectively, of a no-noise reconstruction of the impulse IPD. Here we use regularization parameter $\lambda = 2.5 \times 10^{-4}$ in (3.6).

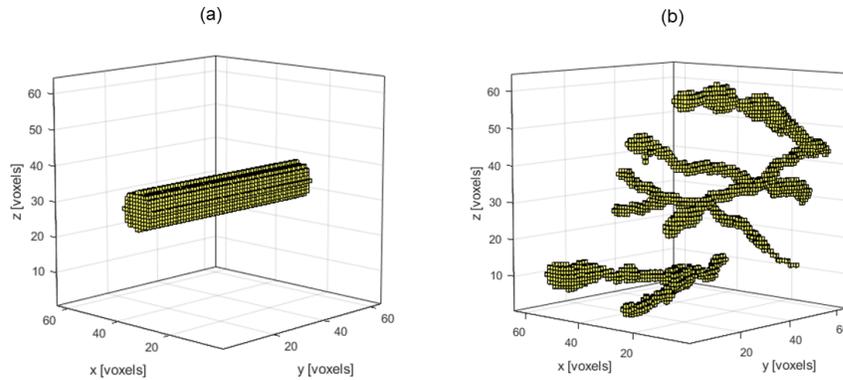


Figure 3.4: (a) The base case cylindrical IPD and (b) an example of a vessel-like IPD.

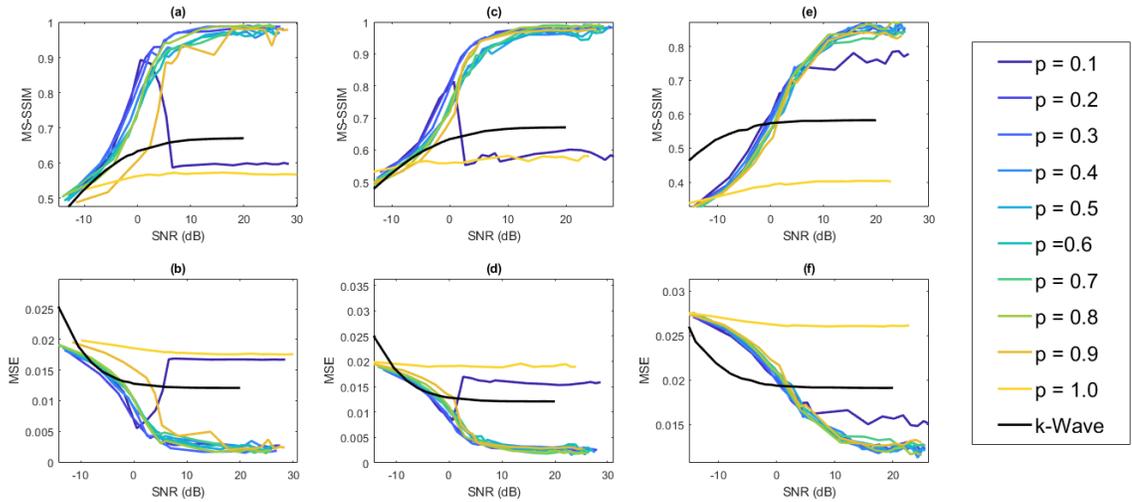


Figure 3.5: Average MS-SSIM (a) and average MSE (b) of a cylinder IPD with a radius of six perpendicular to the direction of shearing. Average MS-SSIM (c) and average MSE (d) of a cylinder IPD with a radius of 6 h parallel to the direction of shearing. Average MS-SSIM (e) and average MSE (f) of a vessel IPD with ten vessels present. Each point is averaged over five trials for each SNR value considered and is plotted against the SNR value used to calculate the additive Gaussian noise.

We now examine the effects of varying cylinder size on the reconstruction using the base case cylindrical IPD. We use $\text{SNR} \approx 27$ dB and examine the reconstruction using different values for the radius of the cylinder, including the center voxel, in the cylindrical IPD. The results displayed in Figure 3.5 show comparable performance for p values between 0.2 and 0.8 in (3.7). The value of $p = 0.3$ was selected for all subsequent analyses. Figure 3.6 displays the MSE and MS-SSIM results.

We next examine the effect of increased complexity on the reconstruction. The amount of noise is fixed so that $\text{SNR} \approx 25$ dB. Using the vessel-like IPDs, the reconstruction is attempted for an increasing number of vessels. These results are displayed in Figure 3.7.

An important consideration in these reconstructions is the choice of regularization parameter, λ in (3.6). While the optimal regularization parameter is a function of the noise present in the system, it is desirable for the method to be robust in terms of

choice of regularization parameter. Figure 3.8 displays results for the reconstruction of a cylinder orthogonal to the direction of shearing and of a vessel-like IPD with ten vessels present for a range of regularization parameters λ in (3.6). Observe that our method performs consistently for the same choice of regularization parameter across various noise levels for both the MS-SSIM and MSE metrics. Figure 3.8 (right) also demonstrates that the method is robust with respect to the choice of regularization parameter for the vessel-like IPD reconstruction. On the other hand, the large jump displayed in Figure 3.8 (left) shows that the method is not as robust with respect to the choice of the regularization parameter for the single-cylinder case. We speculate that this might be due to the fact that most of the true underlying image has zero value, making it difficult to tune the regularization parameter. We do not see this lack of robustness as a practical issue, however, since real-world applications more closely resemble the multiple vessel case. This issue will be investigated in future work.

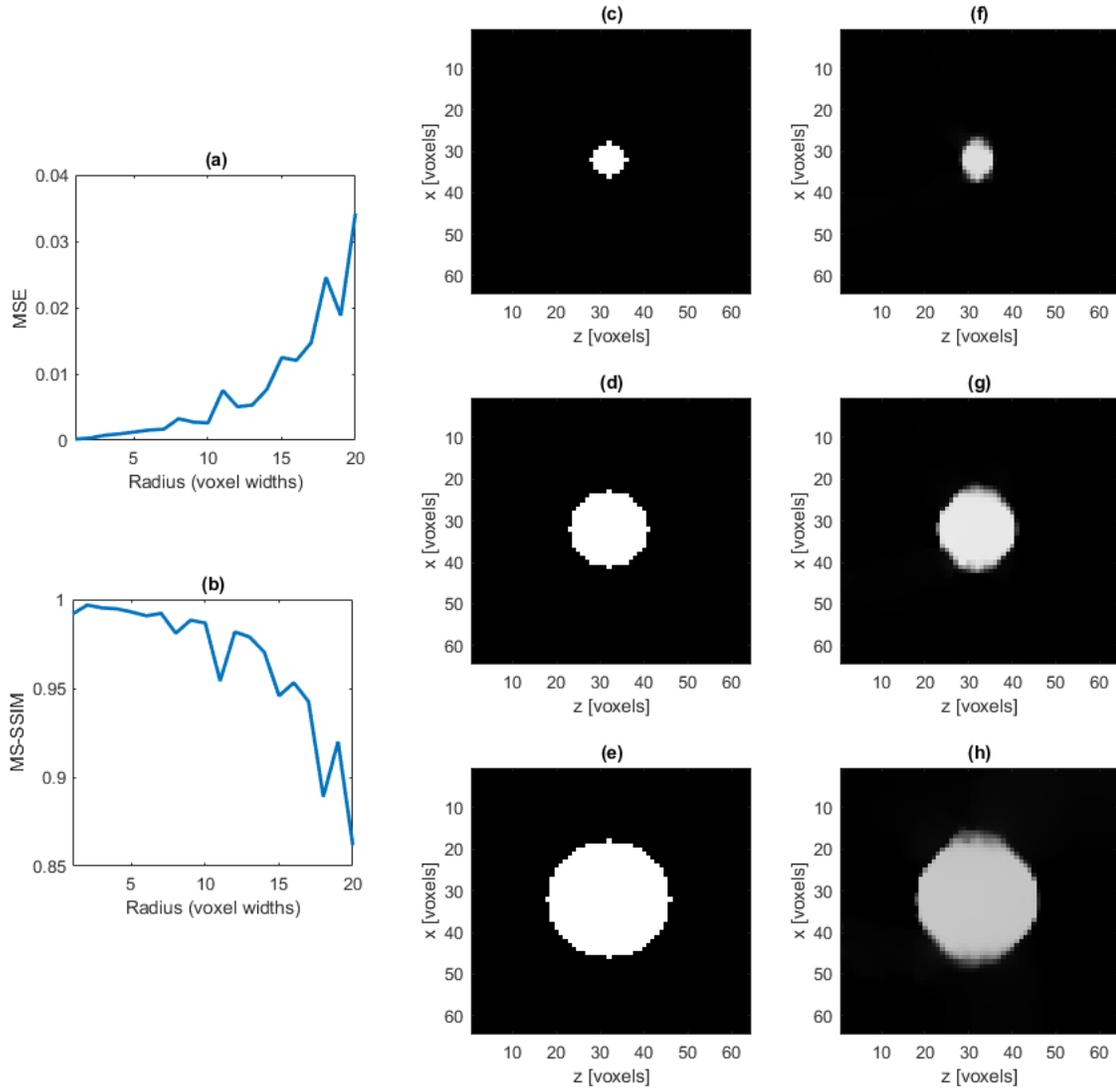


Figure 3.6: Average MSE (a) and average MS-SSIM (b) of the reconstructed cylinder IPD with varying radius, averaged over five trials for each radius value considered. A cross section of the ground truth cylinder with radius of five (c), ten (d) and fifteen (e) voxel widths and a cross section of the reconstruction of the same cylinder, respectively, (f–h). Each cylinder considered is orthogonal to the xz -plane.

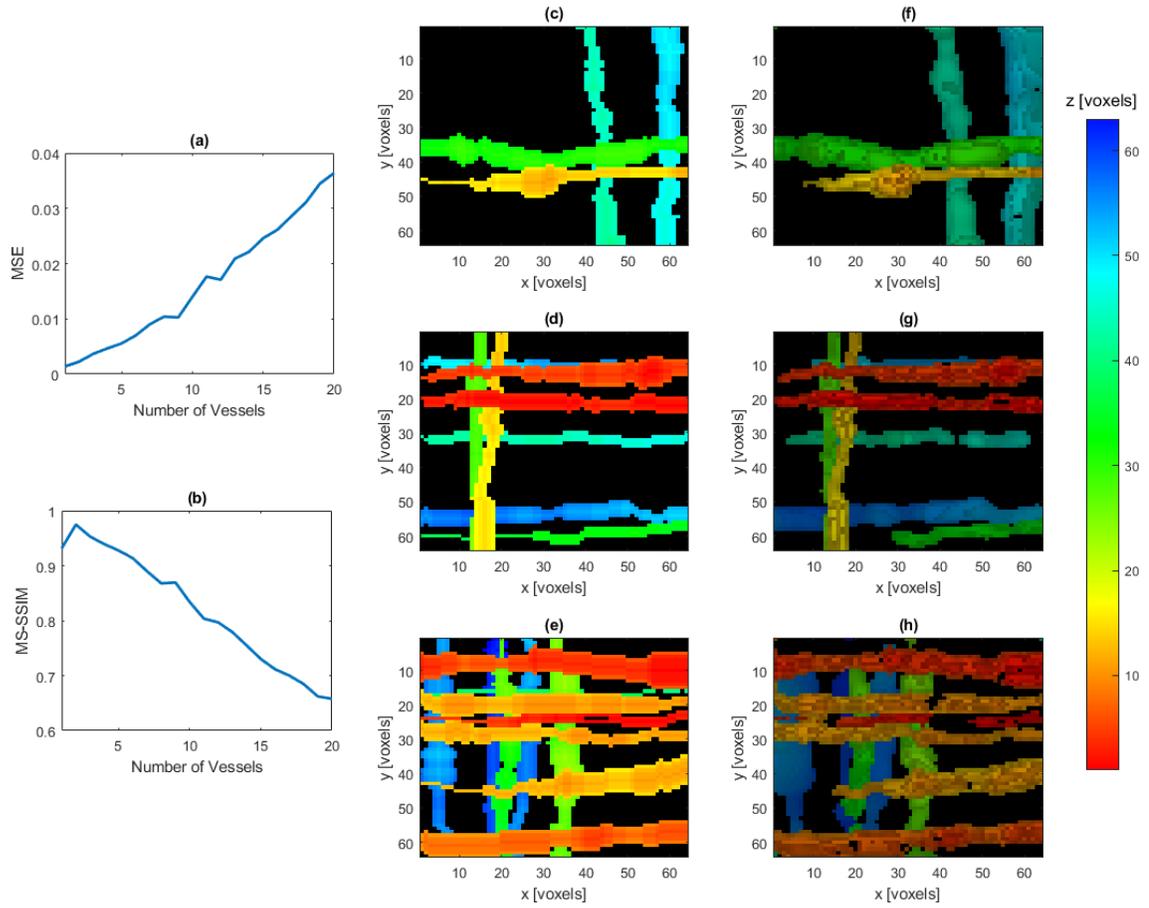


Figure 3.7: Average MSE (a) and average MS-SSIM (b) of the reconstructed vessel IPDs with a varying number of vessels present, averaged over twenty IPDs for each number of vessels considered. Ground truth projection onto the xy -plane of a four vessel IPD (c), eight vessel IPD (d), and twelve vessel IPD (e), and the reconstruction of the same IPDs, respectively, (f–h). In images (c–h), hue represents depth in the z -dimension, with the colorbar indicating pixel lengths away from the FPE, while intensity is proportional the value of the voxels after being thresholded at 0.15.

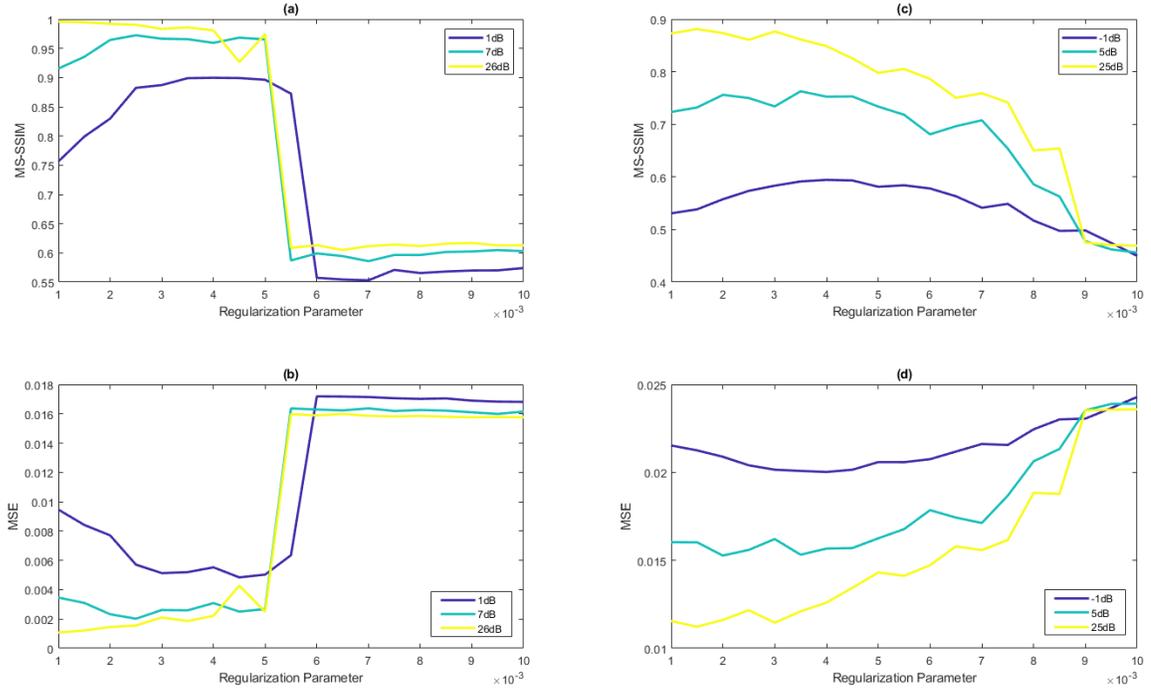


Figure 3.8: Average MSE (a) and average MS-SSIM (b) of the reconstructed cylinder IPD as well as average MSE (c) and average MS-SSIM (d) of the reconstructed vessel-like IPD, averaged over five trials for each value of the regularization parameter considered. Low, medium, and high values of SNR are considered for comparison.

Section 3.4

Discussion and Conclusions

In this paper, we modeled a new method for compressed single-shot PA image reconstruction using various types of DMDs to encode temporal information, and then demonstrated through simulated experiments that our approach is capable of accurately reconstructing a variety of IPDs. Moreover, it is robust in the presence of additive Gaussian white noise. We note that while the IPDs modeled here are piecewise constant, the k-Wave toolbox uses methods best-suited for smooth IPDs. We do not anticipate this presenting issues in real-world applications, since the physical process will not experience the aliasing that is observed with the k-Wave simulations.

Figure 3.5 demonstrates that for a range of probabilities that a given pixel in the

binary mask is turned on, i.e., $0.2 \leq p \leq 0.8$, the performance of our method is consistently better than the k-Wave reconstruction whenever $\text{SNR} \geq 0$ dB. This is a significant improvement given that the k-Wave reconstruction is done with the full time-series data and that our method experiences approximately 25-fold compression. While the random construction of the mask was effective in the simulations, there may be other ways to construct the mask leading to more accurate reconstructions in some cases. This will be the subject of future work.

Figures 3.6 and 3.7 demonstrate the effectiveness of the reconstruction as the number of nonzero values increases in the system, and we note that we are able to achieve accurate reconstructions in the presence of both large and complex vessel systems.

In future investigations we will assemble the optical system and employ the methods and techniques discussed here to reconstruct phantoms using images generated by the physical forward model. In the construction and implementation of the physical optical system, there are several considerations regarding the continuous wave laser for which to account. In the simulations considered here, the sample was assumed to be under uniform illumination by the laser, while the real system will likely experience a more Gaussian-type illumination. A correction to the forward model would then be needed for spatially varying beam intensity. In addition, the FPE cavity must be tuned to match the wavelength of the continuous wave laser. Nanoscopic variations in cavity thickness could prove detrimental to the sensitivity of the system. Finally, the imaging sensor must be selected to have adequate sensitivity to the continuous wave laser wavelength. As the sweeping speed is increased to reach adequate sampling in time, fewer photons are incident on each camera pixel. Thus, MHz sampling rates may require a sensitive camera and relatively high-power laser.

Chapter 4

Complex-Valued Bayesian LASSO

Section 4.1

Introduction

Recovering complex-valued images or signals¹ from noisy and/or under-sampled data is important in coherent imaging applications such as synthetic aperture radar (SAR) [32, 66], ultrasound [135], and digital holography [134]. The problem is often modeled as

$$\mathbf{y} = F\mathbf{z} + \boldsymbol{\varepsilon}. \quad (4.1)$$

Here $\mathbf{z} = \mathbf{g} \odot e^{i\boldsymbol{\phi}} \in \mathbb{C}^n$ with \odot indicating componentwise multiplication is the unknown signal of interest decomposed into magnitude $\mathbf{g} \in (\mathbb{R}^+)^n$ and phase $\boldsymbol{\phi} \in [-\pi, \pi)^n$, $\mathbf{y} \in \mathbb{C}^m$ represents the observable data, $F \in \mathbb{C}^{m \times n}$ is the known forward linear operator, and $\boldsymbol{\varepsilon} \in \mathbb{C}^m$ is centered Gaussian noise with covariance $\sigma^2 \mathbb{I}_m$ and probability

¹We use the terms image and signal interchangeably throughout this chapter.

density

$$f_{\mathcal{E}}(\boldsymbol{\varepsilon}) = \frac{1}{\pi^m \sigma^{2m}} \exp\left(-\frac{1}{\sigma^2} \|\boldsymbol{\varepsilon}\|_2^2\right).$$

Many techniques seek to exclusively recover the magnitude \mathbf{g} , that is, without considering the phase $\boldsymbol{\phi}$, and as such are designed to promote sparsity in \mathbf{g} or in some known transformation of \mathbf{g} (e.g., its gradient) [81, 107]. However phase information is also useful for some applications, including coherent change detection [10, 66, 95, 125], high-resolution imaging [121], and interferometry [51, 66]. In this regard, some methods have been developed to recover point estimates of complex-valued signals to promote sparsity of \mathbf{g} or of some transformation of \mathbf{g} to a sparse domain, see e.g. [30, 56, 104, 110, 136].

More recent approaches have incorporated uncertainty quantification (UQ) into complex-valued signal recovery methods (see e.g. [35, 36, 43]), thereby providing useful information for real-time decision making. These methods are generally designed to incorporate prior sparse knowledge of the magnitude or some linear transform of the magnitude of the signal into the recovery of the posterior. In the case where sparsity is in some linear transform of the magnitude, however, these methods do not *infer* information regarding the phase. Instead the phase is approximated as part of an optimization step, limiting the uncertainty information available.

The method developed in this investigation addresses some of these issues and brings a more comprehensive approach to recovering complex-valued images. Specifically, we build on the *Bayesian LASSO* (least absolute shrinkage and selection operator) technique [99], which was originally designed to recover sparse *real*-valued signals. While other inference methods, such as generalized sparse Bayesian learning [53], recover uncertainty information for the signal itself, the Bayesian LASSO method also provides UQ for hyperparameters that describe the structure and overall sparsity of

the image. The idea in Bayesian LASSO is to recast the ℓ_1 -regularized optimization problem (commonly employed in compressed sensing applications [28, 41]) in a Bayesian framework by treating both the data and the unknown signal as random variables, yielding the linear system

$$\mathcal{Y} = F\mathcal{X} + \mathcal{E}. \quad (4.2)$$

The forward operator $F \in \mathbb{R}^{m \times n}$ is known, and $\mathcal{X} \in \mathbb{R}^n$, $\mathcal{E}, \mathcal{Y} \in \mathbb{R}^m$ are random variables with respective realizations \mathbf{x} , $\boldsymbol{\varepsilon}$, $\mathbf{y} = F\mathbf{x} + \boldsymbol{\varepsilon}$. An estimate of the full posterior density function of the real-valued signal can then be recovered using Bayes' theorem, expressed as

$$f_{\mathcal{X}|\mathcal{Y}}(\mathbf{x}|\mathbf{y}) \propto f_{\mathcal{Y}|\mathcal{X}}(\mathbf{y}|\mathbf{x})f_{\mathcal{X}|\eta}(\mathbf{x}|\eta)f_{\eta}(\eta). \quad (4.3)$$

Here $f_{\mathcal{Y}|\mathcal{X}}(\mathbf{y}|\mathbf{x})$ is the likelihood density function determined by F and assumptions on \mathcal{E} , $f_{\mathcal{X}|\eta}(\mathbf{x}|\eta)$ is the prior density function that encodes a priori assumptions about the unknown, and $f_{\eta}(\eta)$ is the hyperprior density function on the scale parameter η of the prior density. For example, for sparsifying transform matrix $L \in \mathbb{R}^{k \times n}$, $f_{\mathcal{X}|\eta}(\mathbf{x}|\eta)$ may be defined as the product of Laplace probability densities yielding

$$f_{\mathcal{X}|\eta}(\mathbf{x}|\eta) \propto \prod_{i=1}^k \frac{1}{2\eta} \exp\left(-\frac{1}{\eta}|L_i\mathbf{x}|\right), \quad (4.4)$$

where L_i is the i th row of L for $i = 1, \dots, k$.

Our new method extends the Bayesian LASSO technique to *complex*-valued signals whose magnitude is sparse in some domain by treating \mathbf{y} , \mathbf{z} , and $\boldsymbol{\varepsilon}$ in (4.1) as realizations of respective random variables \mathcal{Y} , \mathcal{Z} , and \mathcal{E} , leading to the probabilistic

forward model

$$\mathcal{Y} = F\mathcal{Z} + \mathcal{E}. \quad (4.5)$$

The corresponding estimate of the full posterior density function of the complex-valued signal can then be recovered according to

$$f_{\mathcal{Z}|\mathcal{Y}}(\mathbf{z}|\mathbf{y}) \propto f_{\mathcal{Y}|\mathcal{Z}}(\mathbf{y}|\mathbf{z})f_{\mathcal{Z}|\eta}(\mathbf{z}|\eta)f_{\eta}(\eta), \quad (4.6)$$

where similarly $f_{\mathcal{Y}|\mathcal{Z}}(\mathbf{y}|\mathbf{z})$ is the likelihood density function determined by F and assumptions on \mathcal{E} , $f_{\mathcal{Z}|\eta}(\mathbf{z}|\eta)$ is the prior density function encoding a priori assumptions about the unknown, and $f_{\eta}(\eta)$ is the hyperprior density function on η , the scale parameter of the prior density. We call our resulting method the complex-valued Bayesian LASSO (CVBL). As a primary benefit, the CVBL allows us to fully exploit the sparsity of the underlying complex-valued signal in the recovery without sacrificing the phase, all while quantifying the uncertainty regarding the entire complex-valued signal and the hyperparameters that describe its structure and sparsity.

Our contribution

Given noisy indirect observable data, we introduce a new Bayesian model that uses *a priori* assumptions regarding the magnitude of the underlying complex-valued image of interest to recover its posterior distribution as well as to quantify the uncertainty for both the magnitude and the phase of the unknown signal. Adapting the Bayesian LASSO approach allows us to develop efficient sampling techniques to simulate random draws from these resulting posterior distributions.

Chapter organization

The rest of this paper is organized as follows. Section 4.2 details the construction of the likelihood, prior, and hyperprior densities for our method. Section 4.3 discusses the real-valued Bayesian LASSO approach for sparse signal recovery, which we then expand to include recovery of signals that are sparse in some transform domain. The corresponding *complex-valued* Bayesian LASSO is then proposed in Section 4.4. Numerical experiments in Section 4.5 consider different forward operators F in (4.1) as well as various signal to noise (SNR) values, demonstrating our method’s utility and robustness to noise. We also compare our results to those obtained using the more classical LASSO maximum a posteriori (MAP) estimate approach [119]. We provide some concluding remarks and ideas for future work in Section 4.6.

Section 4.2

Bayesian Formulation

Since (4.1) is easily understood for one-dimensional problems, we develop our method for $\mathbf{z} \in \mathbb{C}^n$ and note that higher-dimensional signals can be readily vectorized to fit this form. Section 4.5 includes both one and two-dimensional examples. Below we collect the ingredients needed for our new method, including a description of the likelihood density function in Section 4.2.1, the construction of sparsity-promoting priors in Section 4.2.2, and a review of commonly used sparsifying transform operators in Section 4.2.3.

4.2.1. The likelihood

The likelihood density function $f_{\mathbf{y}|\mathbf{z}}(\mathbf{y}|\mathbf{z})$ in (4.6) is determined from the density function of the noise present in the system. Following [99] and adjusting accordingly for the complex-valued signal case, here we assume that \mathcal{E} follows a central complex

normal distribution with covariance $\sigma^2\mathbb{I}_m$, yielding

$$f_{\boldsymbol{\varepsilon}}(\boldsymbol{\varepsilon}) = \mathcal{CN}(\mathbf{0}, \sigma^2\mathbb{I}_m) = \frac{1}{\pi^m \sigma^{2m}} \exp\left(-\frac{1}{\sigma^2} \|\boldsymbol{\varepsilon}\|_2^2\right).$$

The law of total probability provides that

$$f_{\mathcal{Y}|\mathcal{Z}}(\mathbf{y}|\mathbf{z}) = \int_{\mathbb{C}^m} f_{\mathcal{Y}|\mathcal{Z},\boldsymbol{\varepsilon}}(\mathbf{y}|\mathbf{z}, \boldsymbol{\varepsilon}) f_{\boldsymbol{\varepsilon}}(\boldsymbol{\varepsilon}) d\boldsymbol{\varepsilon}. \quad (4.7)$$

When conditioned on $\mathbf{z} = \mathcal{Z}$ and $\boldsymbol{\varepsilon} = \mathcal{E}$, (4.5) implies that \mathcal{Y} is entirely determined according to the density function

$$f_{\mathcal{Y}|\mathcal{Z},\boldsymbol{\varepsilon}}(\mathbf{y}|\mathbf{z}, \boldsymbol{\varepsilon}) = \delta(\mathbf{y} - F\mathbf{z} - \boldsymbol{\varepsilon}). \quad (4.8)$$

Combining (4.7) and (4.8), the likelihood density function is therefore

$$f_{\mathcal{Y}|\mathcal{Z}}(\mathbf{y}|\mathbf{z}) = \int_{\mathbb{C}^m} \delta(\mathbf{y} - F\mathbf{z} - \boldsymbol{\varepsilon}) f_{\boldsymbol{\varepsilon}}(\boldsymbol{\varepsilon}) d\boldsymbol{\varepsilon} = f_{\boldsymbol{\varepsilon}}(\mathbf{y} - F\mathbf{z}) = \frac{1}{\pi^m \sigma^{2m}} \exp\left(-\frac{1}{\sigma^2} \|\mathbf{y} - F\mathbf{z}\|_2^2\right). \quad (4.9)$$

Remark 4.1. We note that an improper hyperprior was placed on σ^2 in [99]. For simplicity here we assume that σ^2 is either explicitly known or easily estimated. In general, σ^2 may be provided *a priori* when the errors present in the forward model are well understood. In other situations, an estimate of σ^2 may be acquired when an area of low intensity in \mathcal{Z} is known or when multiple measurement vectors (MMV) are available, see e.g. [64, 108, 138].

4.2.2. The prior density function

A primary goal for our new method is to ensure that the prior density function used in (4.6) enables efficient computation for the complex-valued Bayesian LASSO (CVBL). In this regard we assume that either the magnitude or some linear transform of the

magnitude of the unknown signal is sparse.

Sparse magnitude. Let \mathcal{Z} be decomposed as $\mathcal{Z} = \mathcal{A} + i\mathcal{B}$, where $\mathcal{A} \in \mathbb{R}^n$ and $\mathcal{B} \in \mathbb{R}^n$ are mutually independent with respective realizations \mathbf{a} and \mathbf{b} . When $|\mathcal{Z}|$ is sparse we can simply adapt the approach from [138] used for real-valued signals and define the conditional prior density function as

$$f_{\mathcal{Z}|\eta}(\mathbf{z}|\eta) = f_{\mathcal{A},\mathcal{B}|\eta}(\mathbf{a}, \mathbf{b}|\eta) = \frac{1}{(2\eta)^n} \exp\left(-\frac{1}{\eta} \sum_{j=1}^n \sqrt{a_j^2 + b_j^2}\right), \quad (4.10)$$

where $\eta \in \mathbb{R}^+$ is a random variable with realization η .

Sparse transform of magnitude. When the magnitude is presumed sparse in a domain other than the imaging domain, we first reformulate the model as

$$\mathcal{Y} = F\mathcal{Z} + \mathcal{E} = F(\mathcal{G} \odot e^{i\Phi}) + \mathcal{E}. \quad (4.11)$$

Here $\mathcal{Z} = \mathcal{G} \odot e^{i\Phi}$, where $\mathcal{G} \in \mathbb{R}_+^n$ and $\Phi \in [-\pi, \pi]^n$, with \mathcal{G} and Φ assumed to be mutually independent. The componentwise decomposition of \mathcal{Z} allows us to rewrite the likelihood and prior density functions in (4.9) respectively as

$$f_{\mathcal{Y}|\mathcal{G},\Phi}(\mathbf{y}|\mathbf{g}, \boldsymbol{\phi}) = \frac{1}{\pi^n \sigma^{2n}} \exp\left(-\frac{1}{\sigma^2} \|\mathbf{y} - F(\mathbf{g} \odot e^{i\boldsymbol{\phi}})\|_2^2\right), \quad f_{\mathcal{Z}|\eta}(\mathbf{z}|\eta) = f_{\mathcal{G}|\eta}(\mathbf{g}|\eta) f_{\Phi}(\boldsymbol{\phi}).$$

Let $L \in \mathbb{R}^{k \times n}$ be the rank n operator that transforms \mathcal{G} to the sparse domain.²

Denoting $\mathbf{1}_{\mathbb{R}^+}$ as the indicator function for positive real vectors, the conditional prior

²The rank n requirement is needed for the particular computational implementation in Algorithm 2 and Algorithm 5. A rank deficient sparse transform operator is commonly augmented by imposing boundary constraints [74].

probability density for \mathcal{G} is then

$$f_{\mathcal{G}|\eta}(\mathbf{g}|\eta) \propto \frac{1}{(2\eta)^k} \exp\left(-\frac{1}{\eta}\|L\mathbf{g}\|_1\right) \mathbf{1}_{\mathbb{R}_+^n}(\mathbf{g}), \quad (4.12)$$

where again $\eta \in \mathbb{R}_+$ is a random variable with realization η .

As there is often no prior information regarding the phase Φ in coherent imaging systems, it is reasonable to impose the uniform prior as

$$f_{\Phi}(\boldsymbol{\phi}) = \frac{1}{(2\pi)^n} \mathbf{1}_{[-\pi, \pi]^n}(\boldsymbol{\phi}), \quad (4.13)$$

where $\mathbf{1}_{[-\pi, \pi]^n}$ denotes the indicator function for vectors whose elements are in $[-\pi, \pi]$.

The hyperprior. There are, of course, various options for the hyperprior on η in (4.10) and (4.12). In some cases it may be appropriate to simply use the delta density function

$$f_{\eta}(\eta) = \delta(\eta - \hat{\eta}), \quad (4.14)$$

with point estimate $\hat{\eta}$ for η . Such an estimate may be available when considering signals whose sparsity is known or well-approximated, see e.g. [108, 138]. Here we follow the original Bayesian LASSO method [99] and place a gamma hyperprior on η^{-2} in order to maintain conjugacy, yielding the probability density function³

$$f_{\eta^{-2}}(\eta^{-2}) = \frac{\delta^r}{\Gamma(r)} (\eta^{-2})^{r-1} \exp(-\delta\eta^{-2}), \quad r > 0, \delta > 0. \quad (4.15)$$

Note that choosing *shape* parameter $r \leq 1$ ensures that the mode of $f_{\eta^{-2}}(\eta^{-2})$ is zero while also encouraging sparsity in the solution. Furthermore, having *rate* parameter

³See discussion surrounding (4.21) for explanation regarding using η^{-2} in place of η in (4.3).

$\delta \ll 1$ gives $f_{\eta^{-2}}(\eta^{-2})$ a fatter tail, making the hyperprior relatively uninformative. To encourage sparsity while still allowing for small values of η^{-2} , unless otherwise noted our numerical experiments use hyperparameters

$$r = 1, \quad \delta = 10^{-3}, \quad (4.16)$$

with no additional tuning.

4.2.3. The sparsifying transform operator

This investigation assumes that the underlying signal is either sparse or that its magnitude is piecewise constant, in which case the first order differencing (TV) operator is well suited to suppress variation and noise in smooth regions. The numerical experiment in Subsection 4.5.3 verifies that this assumption is reasonable even when the true signal has piecewise smooth (but not constant) magnitude. We note, however, that such an assumption is not an inherent limitation to our new method, which can be straightforwardly adapted to other sparsifying transform operators, such as higher order TV (HOTV) [5] or wavelets [3, 79] as appropriate.

Remark 4.2. When using $L \in \mathbb{R}^{k \times n}$ with $k > n$ as the sparsifying transform, for instance when $L = [L_1^T \ L_2^T]^T$ is the 2D gradient operator with $L_1 \in \mathbb{R}^{n \times n}$ computing vertical differences and $L_2 \in \mathbb{R}^{n \times n}$ computing horizontal differences, the posterior magnitude may become overly smoothed. We conjecture that when $k \gg n$, the resulting posterior density is multimodal and has significant mass concentrated where η^{-2} is large and the magnitude is relatively flat. Future investigations will explore methods to overcome this undesirable result.

Before introducing our approach for complex-valued signal recovery in Section 4.4, Section 4.3 first discusses its real-valued analog. We use what we will call the *real-valued* Bayesian LASSO (RVBL) technique [99], which was initially developed to

sample a posterior consisting of a Gaussian likelihood and Laplace prior on the signal itself, that is, where L is the identity matrix. The technique was extended to include the anisotropic TV operator in [70]. We now adapt these ideas to accommodate *any* sparsifying rank n linear operator L for real-valued signal recovery, which to our knowledge has not been previously done.

Section 4.3

Bayesian LASSO for a Real-Valued Signal

The RVBL method is a blocked Gibbs sampling technique. Critical to the approach is the equivalent representation of each element of the product in (4.4) as a scale mixture of Gaussians with an exponential mixing density written for each component $L_i \mathbf{x}$, $i = 1, \dots, k$. In particular,

$$\frac{1}{2\eta} \exp\left(-\frac{1}{\eta} |L_i \mathbf{x}|\right) = \int_0^\infty \frac{1}{\sqrt{2\pi s_i}} \exp\left(-\frac{(L_i \mathbf{x})^2}{2s_i}\right) \frac{1}{2\eta^2} \exp\left(-\frac{s_i}{2\eta^2}\right) ds_i. \quad (4.17)$$

Substituting (4.17) into (4.4) and denoting $D(\mathbf{s}) = \text{diag}(\mathbf{s})$ yields

$$\begin{aligned} f_{\mathcal{X}|\mathbf{y}}(\mathbf{x}|\eta) &= \prod_{i=1}^k \int_0^\infty \frac{1}{\sqrt{2\pi s_i}} \exp\left(-\frac{(L_i \mathbf{x})^2}{2s_i}\right) \frac{1}{2\eta^2} \exp\left(-\frac{s_i}{2\eta^2}\right) ds_i \\ &= \int_{(\mathbb{R}^+)^k} \prod_{i=1}^k \frac{1}{\sqrt{2\pi s_i}} \exp\left(-\frac{(L_i \mathbf{x})^2}{2s_i}\right) \frac{1}{2\eta^2} \exp\left(-\frac{s_i}{2\eta^2}\right) d\mathbf{s} \\ &= \int_{(\mathbb{R}^+)^k} \frac{1}{(2\pi)^{\frac{k}{2}} |D(\mathbf{s})|^{\frac{1}{2}}} \exp\left(-\frac{1}{2} (L\mathbf{x})^T [D(\mathbf{s})]^{-1} (L\mathbf{x})\right) \\ &\quad \times \frac{1}{2^k \eta^{2k}} \exp\left(-\frac{\|\mathbf{s}\|_1}{2\eta^2}\right) d\mathbf{s}, \end{aligned} \quad (4.18)$$

where we have used Fubini's theorem [45] to obtain the second equation.

Now consider the marginal density of \mathcal{X} over the joint distribution of \mathcal{X} and some

random variable $\boldsymbol{\tau}^2 \in \mathbb{R}^n$, which we refer to as the scale mixture parameter, given by

$$f_{\mathcal{X}|\eta}(\mathbf{x}|\eta) = \int_{(\mathbb{R}^+)^k} f_{\mathcal{X},\boldsymbol{\tau}^2|\eta}(\mathbf{x}, \boldsymbol{\tau}^2|\eta) d\boldsymbol{\tau}^2 = \int_{(\mathbb{R}^+)^k} f_{\mathcal{X}|\boldsymbol{\tau}^2,\eta}(\mathbf{x}|\boldsymbol{\tau}^2, \eta) f_{\boldsymbol{\tau}^2|\eta}(\boldsymbol{\tau}^2|\eta) d\boldsymbol{\tau}^2. \quad (4.19)$$

Note that in (4.19) $\boldsymbol{\tau}^2 \in \mathbb{R}^k$ is defined componentwise, with

$$\boldsymbol{\tau}^2 = \begin{bmatrix} \tau_1^2 & \tau_2^2 & \cdots & \tau_k^2 \end{bmatrix}^T. \quad (4.20)$$

By making the substitution $\boldsymbol{\tau}^2$ for \mathbf{s} and directly comparing the integrand terms in (4.18) to those in (4.19), we obtain the Gaussian density function

$$f_{\mathcal{X}|\boldsymbol{\tau}^2,\eta^{-2}}(\mathbf{x}|\boldsymbol{\tau}^2, \eta^{-2}) = f_{\mathcal{X}|\boldsymbol{\tau}^2}(\mathbf{x}|\boldsymbol{\tau}^2) \propto \frac{1}{|D(\boldsymbol{\tau}^2)|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(L\mathbf{x})^T [D(\boldsymbol{\tau}^2)]^{-1} (L\mathbf{x})\right) \quad (4.21)$$

as the prior along with the product of exponential densities

$$f_{\boldsymbol{\tau}^2|\eta^{-2}}(\boldsymbol{\tau}^2|\eta^{-2}) = \prod_{i=1}^k \frac{\eta^{-2}}{2} \exp\left(-\frac{\tau_i^2 \eta^{-2}}{2}\right) \quad (4.22)$$

as the hyperprior in (4.3). Finally with $f_{\eta^{-2}}(\eta^{-2})$ given in (4.15),⁴ we are now able to completely characterize the hierarchical model from (4.9), (4.15), (4.21), and (4.22)

⁴For ease of presentation the rest of this manuscript uses the hyperparameter η^{-2} in place of the original hyperparameter η .

as

$$\mathbf{y}|\mathbf{x} \sim \mathcal{N}(F\mathbf{x}, \sigma^2\mathbb{I}) \quad (4.23a)$$

$$\mathbf{x}|\boldsymbol{\tau}^2 \sim \mathcal{N}(0, (L^T[D(\boldsymbol{\tau}^2)]^{-1}L)^{-1}) \quad (4.23b)$$

$$\boldsymbol{\tau}^2|\eta^{-2} \sim \prod_{j=1}^k \frac{\eta^{-2}}{2} \exp\left(-\frac{\tau_j^2 \eta^{-2}}{2}\right) d\tau_j^2 \quad (4.23c)$$

$$\eta^{-2} \sim \Gamma(r, \delta), \quad (4.23d)$$

which yields the corresponding posterior density function

$$f_{\mathcal{X}, \mathcal{T}^2, \eta^{-2}|\mathcal{Y}}(\mathbf{x}, \boldsymbol{\tau}^2, \eta^{-2}|\mathbf{y}) \propto f_{\mathcal{Y}|\mathcal{X}}(\mathbf{y}|\mathbf{x}) f_{\mathcal{X}|\mathcal{T}^2}(\mathbf{x}|\boldsymbol{\tau}^2) f_{\mathcal{T}^2|\eta}(\boldsymbol{\tau}^2|\eta) f_{\eta^{-2}}(\eta^{-2}), \quad (4.24)$$

where

$$\begin{aligned} f_{\mathcal{Y}|\mathcal{X}}(\mathbf{y}|\mathbf{x}) &\propto \frac{1}{(\sigma^2)^{\frac{n}{2}}} \exp\left(-\frac{1}{2\sigma^2} \|\mathbf{y} - F\mathbf{x}\|_2^2\right) \\ f_{\mathcal{X}|\mathcal{T}^2}(\mathbf{x}|\boldsymbol{\tau}^2) &\propto \frac{1}{\sqrt{|D(\boldsymbol{\tau}^2)|}} \exp\left(-\frac{1}{2}(L\mathbf{x})^T [D(\boldsymbol{\tau}^2)]^{-1} (L\mathbf{x})\right) \\ f_{\mathcal{T}^2|\eta}(\boldsymbol{\tau}^2|\eta) &\propto \prod_{j=1}^k \eta^{-2} \exp\left(-\frac{\tau_j^2 \eta^{-2}}{2}\right) \\ f_{\eta^{-2}}(\eta^{-2}) &\propto (\eta^{-2})^{r-1} \exp(-\delta \eta^{-2}). \end{aligned}$$

By isolating the parts of (4.24) that depend on \mathbf{x} and defining

$$\begin{aligned} G &:= \frac{1}{\sigma^2} F^T F + L^T [D(\boldsymbol{\tau}^2)]^{-1} L, \\ \bar{\mathbf{x}} &:= \frac{1}{\sigma^2} G^{-1} F^T \mathbf{y}, \\ c(\mathbf{y}) &:= \exp\left(-\frac{1}{2\sigma^2} \mathbf{y}^T \mathbf{y} + \frac{1}{2} \bar{\mathbf{x}}^T G \bar{\mathbf{x}}\right), \end{aligned}$$

we obtain

$$\begin{aligned}
f_{\mathcal{X}|\mathcal{Y},\tau^2}(\mathbf{x}|\mathbf{y},\tau^2) &\propto \exp\left(-\frac{1}{2\sigma^2}\|\mathbf{y}-F\mathbf{x}\|_2^2-\frac{1}{2}\mathbf{x}^T L^T[D(\tau_1^2)]^{-1}L\mathbf{x}\right) \\
&= \exp\left(-\frac{1}{2\sigma^2}(\mathbf{y}^T\mathbf{y}-\mathbf{x}^T F^T\mathbf{y}-\mathbf{y}^T F\mathbf{x})-\frac{1}{2}\mathbf{x}^T G\mathbf{x}\right) \\
&= c(\mathbf{y})\exp\left(-\frac{1}{2}(\mathbf{x}-\bar{\mathbf{x}})^T G(\mathbf{x}-\bar{\mathbf{x}})\right) \\
&\propto \exp\left(-\frac{1}{2}(\mathbf{x}-\bar{\mathbf{x}})^T G(\mathbf{x}-\bar{\mathbf{x}})\right). \tag{4.25}
\end{aligned}$$

Conditionals on \mathcal{T}_j^2 . When conditioned on all other variables *including* $\mathcal{T}_{-j}^2 = \{\mathcal{T}_{j'}^2 : j' = 1, \dots, k, j' \neq j\}$, (4.24) gives

$$f_{\mathcal{T}_j^2|\mathcal{T}_{-j}^2,\mathcal{X},\mathcal{Y},\eta^{-2}}(\tau_j^2|\tau_{-j}^2,\mathbf{x},\mathbf{y},\eta^{-2}) \propto (\tau_j^2)^{-\frac{1}{2}} \exp\left(-\frac{[L\mathbf{x}]_j^2}{2\tau_j^2}-\frac{\tau_j^2\eta^{-2}}{2}\right), \quad j = 1, \dots, k. \tag{4.26}$$

Making the change of variables $\nu_j^2 = 1/\tau_j^2$, the conditional can alternatively be expressed as

$$f_{\nu_j^2|\mathcal{T}_{-j}^2,\mathcal{X},\eta^{-2}}(\nu_j^2|\tau_{-j}^2,\mathbf{x},\eta^{-2}) \propto (\nu_j^2)^{-\frac{3}{2}} \exp\left(-\frac{[L\mathbf{x}]_j^2\nu_j^2}{2}-\frac{\eta^{-2}}{2\nu_j^2}\right). \tag{4.27}$$

By defining mean parameter $\mu' = \sqrt{\eta^{-2}/[L\mathbf{x}]_j^2}$ and shape parameter $\lambda' = \eta^{-2}$, we observe that (4.27) fits the form of the density function of inverse Gaussian distribution given by

$$f_{\mu',\lambda'}(y) \propto y^{-3/2} \exp\left(-\frac{\lambda'y}{2(\mu')^2}-\frac{\lambda'}{2y}\right) \propto y^{-3/2} \exp\left(-\frac{\lambda'(y-\mu')^2}{2(\mu')^2y}\right) \quad \mu', \lambda' > 0.$$

Thus (4.27) can be directly sampled using a standard routine for sampling the inverse Gaussian or Wald distribution (e.g., see [86]).

Remark 4.3. In the case where $[L\mathbf{x}]_j \approx 0$ (where L could be the identity matrix \mathbb{I}_n), we can refrain from performing the change of variables to obtain (4.27) and note that the original conditional (4.26) simplifies to

$$f_{\mathcal{T}_j^2|\tau_{-j}^2, \mathcal{X}, \mathcal{Y}, \eta^{-2}}(\tau_j^2|\tau_{-j}^2, \mathbf{x}, \mathbf{y}, \eta^{-2}) \propto (\tau_j^2)^{-\frac{1}{2}} \exp\left(-\frac{\tau_j^2 \eta^{-2}}{2}\right), \quad (4.28)$$

corresponding to the Gamma distribution $\Gamma(1/2, \eta^{-2}/2)$. Thus whenever $|[L\mathbf{x}]_j| < 10^{-8}$ in our numerical experiments we instead sample the conditional density of each \mathcal{T}_j^2 according to (4.28).⁵

Conditional on η^{-2} . Lastly, to sample the conditional on η^{-2} , we note that

$$f_{\eta^{-2}|\mathcal{X}, \mathcal{Y}, \tau^2}(\eta^{-2}|\mathbf{x}, \mathbf{y}, \tau^2) \propto (\eta^{-2})^{r+k-1} \exp\left(-\eta^{-2} \left(\delta + \frac{1}{2} \sum_{j=1}^k \tau_j^2\right)\right), \quad (4.29)$$

which due to conjugacy is the density for the Gamma distribution $\Gamma(r+k, \delta + \frac{1}{2} \sum_{j=1}^k \tau_j^2)$. Algorithm 2 outlines the full Gibbs sampling approach.

Algorithm 2 Real-valued Bayesian LASSO (RVBL)

Input Data vector \mathbf{y} , parameters σ and η , sparsifying operator $L \in \mathbb{R}^{k \times n}$, chain length N_M , and burn-in length B .

Output Samples $\mathbf{x}^{(s-B+1)}$ for $s = B, \dots, N_M$.

1 Set $\mathbf{x}^{(0)} = F^T \mathbf{y}$ and $(\tau_j^2)^{(0)} = (\eta^{-2})^{(0)} = 1$ for $j = 1, \dots, k$.

2 **For** $l = 1, \dots, N_M$ **do**

i. Sample $\mathbf{x}^{(l)}$ from $\mathcal{X} | \mathcal{T}^2 = (\tau^2)^{(l-1)}$

ii. Sample each $(\tau_j^2)^{(l)}$ using $\mathcal{T}_j^{-2} | \mathcal{X} = \mathbf{x}^{(l)}, \eta^{-2} = (\eta^{-2})^{(l-1)}$.

iii. Sample $(\eta^{-2})^{(l)}$ using $\eta^{-2} | \mathcal{T}^2 = (\tau^2)^{(l-1)}$.

⁵Due to system noise, nearly all values $|[L\mathbf{x}]_j|$ are greater than 10^{-8} with typical values in $[\cdot 001, \cdot 1]$.

We conclude this section with illustrative numerical example in Figure 4.1 of the RVBL method applied to a signal deblurring problem, where F is a Gaussian blurring kernel with standard deviation $\sigma_{\text{blur}} = 5.0$ and with a zero boundary condition. We assume that the data \mathbf{y} are generated by $\mathbf{y} = F\mathbf{x} + \boldsymbol{\varepsilon}$, where $\boldsymbol{\varepsilon}$ is Gaussian white noise with $\text{SNR} = 30$ dB, where SNR is defined in (4.52) for complex-valued signals. The problem data and ground truth signal are shown in Figure 4.1 (a). To apply the RVBL method, we take the sparsifying transformation L to be the discrete gradient operator (4.51) and adopt the uninformative hyperprior given by (4.15). Figure 4.1 (b)-(c) show the resulting posterior mean estimates for \mathbf{x} and $\boldsymbol{\tau}^2$, as well as their 95% credible intervals.

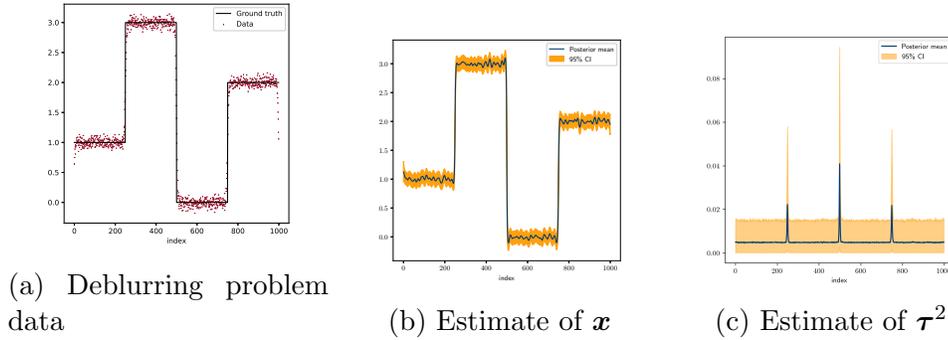


Figure 4.1: The RVBL method applied to a signal deblurring problem. The posterior mean estimate for the parameter η^{-2} is $\bar{\eta}^{-2} \approx 401$ and $\text{SNR} = 30$. in (4.52).

Section 4.4

Complex-valued Bayesian LASSO

We now have all of the ingredients needed to compute and sample from the complex-valued posterior density functions using the likelihood given in (4.9), the priors described in (4.10) and (4.12), and the hyperprior (4.15). We will describe the CVBL for two distinct cases: Subsection 4.4.1 considers the magnitude itself to be sparse, while Subsection 4.4.2 examines the case for which the magnitude is sparse in the

some transform domain.

The scale mixture (4.17) used in the RVBL relies on the likelihood being a real-valued probability density function. As such, for our algorithmic development we will make use of the equivalency that for given $\mathbf{y} \in \mathbb{C}^m$, $F \in \mathbb{C}^{m \times n}$, and $\mathbf{x} \in \mathbb{R}^n$, we have

$$\|\mathbf{y} - F\mathbf{x}\|_2^2 = \|\tilde{\mathbf{y}} - \tilde{F}\mathbf{x}\|_2^2, \quad (4.30)$$

where $\tilde{\mathbf{y}} = [\operatorname{Re}(\mathbf{y})^T \operatorname{Im}(\mathbf{y})^T]^T$ and $\tilde{F} = [\operatorname{Re}(F)^T \operatorname{Im}(F)^T]^T$.

To show (4.30), observe that

$$\begin{aligned} \|\tilde{\mathbf{y}} - \tilde{F}\mathbf{x}\|_2^2 &= (\tilde{\mathbf{y}}^T - \mathbf{x}^T \tilde{F}^T)(\tilde{\mathbf{y}} - \tilde{F}\mathbf{x}) \\ &= \tilde{\mathbf{y}}^T \tilde{\mathbf{y}} - 2\mathbf{x}^T \tilde{F}^T \tilde{\mathbf{y}} + \mathbf{x}^T \tilde{F}^T \tilde{F}\mathbf{x} \\ &= \operatorname{Re}(\mathbf{y})^T \operatorname{Re}(\mathbf{y}) + \operatorname{Im}(\mathbf{y})^T \operatorname{Im}(\mathbf{y}) \\ &\quad - 2(\mathbf{x}^T \operatorname{Re}(F)^T \operatorname{Re}(\mathbf{y}) + \mathbf{x}^T \operatorname{Im}(F)^T \operatorname{Im}(\mathbf{y})) \\ &\quad + \mathbf{x}^T \operatorname{Re}(F)^T \operatorname{Re}(F)\mathbf{x} + \mathbf{x}^T \operatorname{Im}(F)^T \operatorname{Im}(F)\mathbf{x} \\ &= \mathbf{y}^H \mathbf{y} - 2\mathbf{x}^T (\operatorname{Re}(F)^T \operatorname{Re}(\mathbf{y}) + \operatorname{Im}(F)^T \operatorname{Im}(\mathbf{y})) + \mathbf{x}^T F^H F \mathbf{x}, \end{aligned} \quad (4.31)$$

and that furthermore

$$\begin{aligned} \mathbf{x}^T F^H \mathbf{y} + \mathbf{y}^H F \mathbf{x} &= 2 \operatorname{Re}(\mathbf{x}^T F^H \mathbf{y}) \\ &= 2\mathbf{x}^T \operatorname{Re}(F^H \mathbf{y}) \\ &= 2\mathbf{x}^T \operatorname{Re} \left(((\operatorname{Re}(F) - i \operatorname{Im}(F))^T (\operatorname{Re}(\mathbf{y}) + i \operatorname{Im}(\mathbf{y}))) \right) \\ &= 2\mathbf{x}^T \operatorname{Re} \left(\operatorname{Re}(F)^T \operatorname{Re}(\mathbf{y}) - i \operatorname{Im}(F)^T \operatorname{Re}(\mathbf{y}) \right. \\ &\quad \left. + i \operatorname{Re}(F)^T \operatorname{Im}(\mathbf{y}) + \operatorname{Im}(F)^T \operatorname{Im}(\mathbf{y}) \right) \\ &= 2\mathbf{x}^T (\operatorname{Re}(F)^T \operatorname{Re}(\mathbf{y}) + \operatorname{Im}(F)^T \operatorname{Im}(\mathbf{y})). \end{aligned}$$

Combining the above result with (4.31) yields (4.30) since

$$\left\| \tilde{\mathbf{y}} - \tilde{F}\mathbf{x} \right\|_2^2 = \mathbf{y}^H \mathbf{y} - \mathbf{x}^T F^H \mathbf{y} - \mathbf{y}^H F \mathbf{x} + \mathbf{x}^T F^H F \mathbf{x} = \|\mathbf{y} - F\mathbf{x}\|_2^2.$$

4.4.1. CVBL for the Sparse Magnitude Case

We begin by writing \mathcal{Z} as

$$\mathcal{Z} = \mathcal{A} + i\mathcal{B},$$

where $\mathcal{Z} \in \mathbb{C}^n$ and $\mathcal{A}, \mathcal{B} \in \mathbb{R}^n$, with realizations $\mathbf{z} = \mathbf{a} + i\mathbf{b}$. When $|\mathcal{Z}| = |\mathcal{A} + i\mathcal{B}|$ is assumed to be sparse, we choose to impose a 1-norm prior on $\sqrt{\mathcal{A}^2 + \mathcal{B}^2}$ using (4.10). We now provide the details for how this is accomplished.

For observations given in (4.1), an analogous derivation that yielded (4.23) results in the hierarchical model for the *complex-valued* sparse signal recovery

$$\begin{aligned} \mathbf{y} | \mathbf{a}, \mathbf{b} &\sim \mathcal{CN}(F(\mathbf{a} + i\mathbf{b}), \sigma^2 \mathbb{I}), \\ \mathbf{a} | \boldsymbol{\tau}^2 &\sim \mathcal{N}(0, D(\boldsymbol{\tau}^{-2})), \\ \mathbf{b} | \boldsymbol{\tau}^2 &\sim \mathcal{N}(0, D(\boldsymbol{\tau}^{-2})), \\ \boldsymbol{\tau}^2 | \eta^{-2} &\sim \prod_{j=1}^k \frac{\eta^{-2}}{2} \exp\left(-\frac{\tau_j^2 \eta^{-2}}{2}\right) d\tau_j^2, \\ \eta^{-2} &\sim \Gamma(r, \delta), \end{aligned}$$

where $\boldsymbol{\tau}^2$ (defined componentwise as in (4.20)) is a realization of the scale mixture parameter \mathcal{T}^2 introduced in (4.19). We can choose hyperparameters r and δ as in (4.16) to promote sparsity. The posterior is then written as

$$f_{\mathcal{Z}, \mathcal{T}^2, \eta^{-2} | \mathcal{Y}}(\mathbf{z}, \boldsymbol{\tau}^2, \eta^{-2} | \mathbf{y}) = f_{\mathcal{A}, \mathcal{B}, \mathcal{T}^2, \eta^{-2} | \mathcal{Y}}(\mathbf{a}, \mathbf{b}, \boldsymbol{\tau}^2, \eta^{-2} | \mathbf{y}),$$

with

$$f_{\mathcal{A}, \mathcal{B}, \tau^2, \eta^{-2} | \mathcal{Y}}(\mathbf{a}, \mathbf{b}, \boldsymbol{\tau}^2, \eta^{-2} | \mathbf{y}) \propto f_{\mathcal{Y} | \mathcal{A}, \mathcal{B}}(\mathbf{y} | \mathbf{a}, \mathbf{b}) f_{\eta^{-2}}(\eta^{-2}) \\ \times \prod_{j=1}^k f_{\mathcal{A}_j, \mathcal{B}_j | \tau_j^2}(a_j, b_j | \tau_j^2) f_{\tau_j^2 | \eta^{-2}}(\tau_j^2 | \eta^{-2}). \quad (4.32)$$

Here

$$f_{\mathcal{Y} | \mathcal{A}, \mathcal{B}}(\mathbf{y} | \mathbf{a}, \mathbf{b}) \propto \frac{1}{(\sigma^2)^n} \exp\left(-\frac{1}{\sigma^2} \|\mathbf{y} - F\mathbf{a} - iF\mathbf{b}\|_2^2\right), \\ f_{\eta^{-2}}(\eta^{-2}) \propto (\eta^{-2})^{r-1} \exp(-\delta\eta^{-2}), \\ f_{\mathcal{A}_j, \mathcal{B}_j | \tau_j^2}(a_j, b_j | \tau_j^2) \propto \frac{1}{\sqrt{\tau_j^2}} \exp\left(-\frac{a_j^2 + b_j^2}{2\tau_j^2}\right), \\ f_{\tau_j^2 | \eta^{-2}}(\tau_j^2 | \eta^{-2}) \propto \eta^{-2} \exp\left(-\frac{\tau_j^2 \eta^{-2}}{2}\right).$$

For ease of notation we define

$$\tilde{F} = [\operatorname{Re}(F)^T \operatorname{Im}(F)^T]^T, \quad \tilde{F}^* = [-\operatorname{Im}(F)^T \operatorname{Re}(F)^T]^T, \quad \tilde{\mathbf{y}} = [\operatorname{Re}(\mathbf{y})^T \operatorname{Im}(\mathbf{y})^T]^T, \quad (4.33)$$

so that the real and imaginary parts of the observations are respectively

$$\mathbf{y}_1 = \tilde{\mathbf{y}} - \tilde{F}^* \mathbf{b} \quad \text{and} \quad \mathbf{y}_2 = \tilde{\mathbf{y}} - \tilde{F} \mathbf{a}. \quad (4.34)$$

Sampling on \mathcal{A} and \mathcal{B} . Analogous to the procedure resulting in (4.25), we now sample from the conditionals on \mathcal{A} and \mathcal{B} by first isolating the parts of (4.32) that depend on \mathbf{a} and \mathbf{b} respectively, yielding

$$f_{\mathcal{A} | \mathcal{Y}, \mathcal{B}, \tau^2}(\mathbf{a} | \mathbf{y}, \mathbf{b}, \boldsymbol{\tau}^2) \propto \exp\left(-\frac{1}{2}(\mathbf{a} - \bar{\mathbf{a}})^T G(\mathbf{a} - \bar{\mathbf{a}})\right), \quad (4.35a)$$

$$f_{\mathcal{B}|\mathcal{Y},\mathcal{A},\tau^2}(\mathbf{b}|\mathbf{y}, \mathbf{a}, \tau^2) \propto \exp\left(-\frac{1}{2}(\mathbf{b} - \bar{\mathbf{b}})^T G(\mathbf{b} - \bar{\mathbf{b}})\right), \quad (4.35b)$$

where $G = \frac{2}{\sigma^2}\tilde{F}^T\tilde{F} + [D(\tau^2)]^{-1}$, $\bar{\mathbf{a}} = \frac{2}{\sigma^2}G^{-1}\tilde{F}^T\mathbf{y}_1$ and $\bar{\mathbf{b}} = \frac{2}{\sigma^2}G^{-1}\tilde{F}^{*T}\tilde{\mathbf{y}}_2$. Observe that both \mathcal{A} and \mathcal{B} are conditionally Gaussian.

Remark 4.4. If F is unitary, for example when it represents the $n \times n$ normalized discrete Fourier transform, then (4.35a) and (4.35b) can be simplified to multivariate Gaussian densities with respective means

$$\mathbf{a} = \frac{2}{\sigma^2}G^{-1} \operatorname{Re}(F^H(\mathbf{y} - iF\mathbf{b})), \quad \mathbf{b} = \frac{2}{\sigma^2}G^{-1} \operatorname{Im}(F^H(\mathbf{y} - F\mathbf{a})),$$

where $G = \frac{2}{\sigma^2}\mathbb{I} + [D(\tau^2)]^{-1}$. The splitting in (4.33) is needed in applications where the input data are under-sampled, or when portions of the data must be discarded.

Sampling on scale mixing parameter \mathcal{T}_j^2 . Obtaining the conditional density of $\{\mathcal{T}_j^2\}_{j=1}^k$ from (4.32) yields

$$f_{\mathcal{T}_j^2|\mathcal{A},\mathcal{B},\tau_{-j}^2,\eta^{-2}}(\tau_j^2|\mathbf{a}, \mathbf{b}, \tau_{-j}^2, \eta^{-2}) \propto (\tau_j^2)^{-\frac{1}{2}} \exp\left(-\frac{a_j^2 + b_j^2}{2\tau_j^2} - \frac{\tau_j^2\eta^{-2}}{2}\right), \quad j = 1, \dots, k. \quad (4.36)$$

By change of variable $\nu_j^2 = (\mathcal{T}_j^2)^{-1}$, with realization $\nu_j^2 = (\tau_j^2)^{-1}$, we obtain

$$f_{\nu_j^2|\mathcal{A},\mathcal{B},\tau_{-j}^2,\eta^{-2}}(\nu_j^2|\mathbf{a}, \mathbf{b}, \tau_{-j}^2, \eta^{-2}) \propto (\nu_j^2)^{-\frac{3}{2}} \exp\left(-\frac{1}{2\nu_j^2} (a_j^2 + b_j^2) \left(\nu_j^2 - \sqrt{\frac{\eta^{-2}}{(a_j^2 + b_j^2)}}\right)^2\right). \quad (4.37)$$

Observe that comparable to (4.27), (4.37) is the probability density function for an inverse Gaussian distribution with mean parameter $\mu' = \sqrt{\eta^{-2}/(a_j^2 + b_j^2)}$ and shape parameter $\lambda' = \eta^{-2}$. Moreover, since each $(\mathcal{T}_j^2)^{-1}$ is mutually independent from

$(\mathcal{T}_{-j}^2)^{-1}$, $j = 1, \dots, k$, we can efficiently sample the conditionals on each $(\mathcal{T}_j^2)^{-1}$ in parallel.

Remark 4.5. Similar to the discussion in Remark 4.3, we observe that the density in (4.37) is not well-defined when $a_j = b_j = 0$. Although the probability of sampling such a_j and b_j is zero, the probability of sampling a_j and b_j within machine precision is not. Thus in the case where a_j and b_j are sufficiently small, we instead use $\mathcal{T}_j^2 | \mathcal{A}, \mathcal{B}, \mathcal{T}_{-j}^2, \eta^{-2} \sim \Gamma(1/2, \eta^{-2}/2)$ following (4.36).

Sampling on hyperparameter η^{-2} . The conditional posterior of η^{-2} depends exclusively on \mathcal{T}^2 , hence $f_{\eta^{-2} | \mathcal{A}, \mathcal{B}, \mathcal{Y}, \mathcal{T}^2} = f_{\eta^{-2} | \mathcal{T}^2}$. This density then has the form of (4.29), from which we are able to sample directly.

CVBL algorithm for sparse magnitude signals

The conditional distributions in (4.35a), (4.35b), and (4.37) are now combined to form the CVBL Gibbs sampling method provided in Algorithm 3. Methods to efficiently sample \mathbf{a} and \mathbf{b} in Algorithm 3 are discussed in Subsection 4.5.1.

Algorithm 3 CVBL for complex-valued signal with sparse magnitude

Input Data vector \mathbf{y} , noise variance σ^2 , hyperparameters $r = 1$ and $\delta = 10^{-3}$, chain length N_M , and burn-in length B .

Output samples $\mathbf{z}^{(s-B+1)} = \mathbf{a}^{(s)} + i\mathbf{b}^{(s)}$ for $s = B, \dots, N_M$.

1 Set $\mathbf{a}^{(0)} = \text{Re}(A^H \mathbf{y})$, $\mathbf{b}^{(0)} = \text{Im}(A^H \mathbf{y})$, and $(\tau_j^2)^{(0)} = (\eta^{-2})^{(0)} = 1$ for $j = 1, \dots, k$.

2 **For** $\ell = 1, \dots, N_M$ **do**

i. Sample $\mathbf{a}^{(\ell)}$ from $\mathcal{A}|\mathcal{Y} = \mathbf{y}, \mathcal{B} = \mathbf{b}^{(\ell-1)}, \mathcal{T}^2 = (\boldsymbol{\tau}^2)^{(\ell-1)}$ (4.35a).

ii. Sample $\mathbf{b}^{(\ell)}$ from $\mathcal{B}|\mathcal{Y} = \mathbf{y}, \mathcal{A} = \mathbf{a}^{(\ell)}, \mathcal{T}^2 = (\boldsymbol{\tau}^2)^{(\ell-1)}$ (4.35b).

iii. Sample $(\tau_j^{-2})^{(\ell)}$ from $\mathcal{T}_j^{-2}|\mathcal{A} = \mathbf{a}^{(\ell)}, \mathcal{B} = \mathbf{b}^{(\ell)}, \eta^{-2} = (\eta^{-2})^{(\ell-1)}$ (4.37) for $j = 1, \dots, n$.

iv. Sample $(\eta^{-2})^{(\ell)}$ using $\eta^{-2}|\mathcal{T}^2 = (\boldsymbol{\tau}^2)^{(\ell)}$ (4.23d).

4.4.2. Sparsity in a Transform Domain of $|\mathcal{Z}|$

We now turn our attention to the case where sparsity is expected in some transform domain of $|\mathcal{Z}|$. For simplicity we assume the signal magnitude is piecewise constant, so that there is sparsity in the corresponding gradient domain. Hence we define L in (4.12) to be the TV operator (see Section 4.2.3 for more discussion).

Starting from (4.11) and assuming $f_{\mathcal{Y}}(\mathbf{y}) > 0$, Bayes' Theorem yields

$$\begin{aligned} f_{\mathcal{G}, \Phi, \eta|\mathcal{Y}}(\mathbf{g}, \boldsymbol{\phi}, \eta|\mathbf{y}) &= \frac{f_{\mathcal{Y}|\mathcal{G}, \Phi, \eta}(\mathbf{y}|\mathbf{g}, \boldsymbol{\phi}, \eta) f_{\mathcal{G}, \Phi, \eta}(\mathbf{g}, \boldsymbol{\phi}, \eta)}{f_{\mathcal{Y}}(\mathbf{y})} \\ &\propto \exp\left(-\frac{1}{\sigma^2} \|\mathbf{y} - A(\mathbf{g} \odot e^{i\boldsymbol{\phi}})\|_2^2 - \frac{1}{\eta} \|L\mathbf{g}\|_1\right) \mathbf{1}_{\mathbb{R}_+^n}(\mathbf{g}) \mathbf{1}_{[-\pi, \pi]^n}(\boldsymbol{\phi}). \end{aligned} \quad (4.38)$$

Analogously rewriting the prior as a scale mixture of normals as in (4.19) provides

the posterior density function

$$\begin{aligned}
f_{\mathcal{G}, \Phi, \mathcal{T}^2, \eta^{-2} | \mathcal{Y}}(\mathbf{g}, \boldsymbol{\phi}, \boldsymbol{\tau}^2, \eta^{-2} | \mathbf{y}) &\propto f_{\mathcal{Y} | \mathcal{G}, \Phi}(\mathbf{y} | \mathbf{g}, \boldsymbol{\phi}) f_{\Phi}(\boldsymbol{\phi}) f_{\mathcal{G} | \mathcal{T}^2}(\mathbf{g} | \boldsymbol{\tau}^2) \\
&\times f_{\eta^{-2}}(\eta^{-2}) \prod_{j=1}^k f_{\mathcal{T}_j^2 | \eta^{-2}}(\boldsymbol{\tau}_j^2 | \eta^{-2}), \tag{4.39}
\end{aligned}$$

where

$$\begin{aligned}
f_{\mathcal{Y} | \mathcal{G}, \Phi}(\mathbf{y} | \mathbf{g}, \boldsymbol{\phi}) &\propto \exp\left(-\frac{1}{\sigma^2} \|\mathbf{y} - F(\mathbf{g} \odot e^{i\boldsymbol{\phi}})\|_2^2\right) \mathbf{1}_{\mathbb{R}_+^n}(\mathbf{g}), \\
f_{\Phi}(\boldsymbol{\phi}) &\propto \mathbf{1}_{[-\pi, \pi]^n}(\boldsymbol{\phi}), \\
f_{\mathcal{G} | \mathcal{T}^2}(\mathbf{g} | \boldsymbol{\tau}^2) &\propto \frac{1}{\sqrt{|D(\boldsymbol{\tau}^2)|}} \exp\left(-\frac{1}{2} (L\mathbf{g})^T [D(\boldsymbol{\tau}^2)]^{-1} (L\mathbf{g})\right), \\
f_{\eta^{-2}}(\eta^{-2}) &\propto (\eta^{-2})^{r-1} \exp(-\delta\eta^{-2}), \\
f_{\mathcal{T}_j^2 | \eta^{-2}}(\boldsymbol{\tau}_j^2 | \eta^{-2}) &\propto \eta^{-2} \exp\left(-\frac{\boldsymbol{\tau}_j^2 \eta^{-2}}{2}\right).
\end{aligned}$$

Below we introduce a four-step Gibbs sampling process to sample from (4.39) that combines (1) the use of Metropolis-within-Gibbs to sample from $\mathcal{G} | \mathcal{Y}, \Phi, \mathcal{T}^2, \eta^{-2}$; (2) Gibbs steps to sample from $\mathcal{T}^2 | \mathcal{Y}, \mathcal{G}, \Phi, \eta^{-2}$ and $\eta^{-2} | \mathcal{Y}, \mathcal{G}, \Phi, \mathcal{T}^2$; and (3) rejection sampling to sample from $\Phi | \mathcal{Y}, \mathcal{G}, \mathcal{T}^2, \eta^{-2}$.

Sampling on magnitude \mathcal{G} . The conditional distribution on \mathcal{G} is a nonnegatively-constrained Gaussian density. For computational efficiency in sampling the magnitude, we use as the posterior density the untruncated density $\tilde{f}_{\mathcal{G} | \mathcal{Y}, \mathcal{T}^2, \Phi}(\mathbf{g} | \mathbf{y}, \boldsymbol{\tau}^2, \boldsymbol{\phi})$, where

$$f_{\mathcal{G} | \mathcal{Y}, \Phi, \mathcal{T}^2}(\mathbf{g} | \mathbf{y}, \boldsymbol{\phi}, \boldsymbol{\tau}^2) = \tilde{f}_{\mathcal{G} | \mathcal{Y}, \Phi, \mathcal{T}^2}(\mathbf{g} | \mathbf{y}, \boldsymbol{\phi}, \boldsymbol{\tau}^2) \mathbf{1}_{\mathbb{R}_+^n}(\mathbf{g}).$$

Thus

$$\tilde{f}_{\mathcal{G}|\mathcal{Y},\Phi,\tau^2}(\mathbf{g}|\mathbf{y},\boldsymbol{\phi},\boldsymbol{\tau}^2) \propto \exp\left(-\frac{\|\mathbf{y} - F_1\mathbf{g}\|_2^2}{\sigma^2} - \frac{1}{2}(\mathbf{L}\mathbf{g})^T[D(\boldsymbol{\tau}^2)]^{-1}(\mathbf{L}\mathbf{g})\right), \quad (4.40)$$

where $F_1 = FD(e^{i\phi})$. For consistency with (4.39), we simply reject any proposed samples of \mathcal{G} that contain a negative entry. This technique can be categorized as acceptance-rejection sampling, and to this end, Theorem 4.1 shows that (4.40) is a conditional Gaussian distribution, implying that we can directly draw its samples.

Theorem 4.1. *Let $F_1 = FD(e^{i\phi})$, $\tilde{F}_1 = [\text{Re}(F_1)^T \text{Im}(F_1)^T]^T$, and $\tilde{\mathbf{y}} = [\text{Re}(\mathbf{y})^T \text{Im}(\mathbf{y})^T]^T$. Assume that L has rank n . The function $\tilde{f}_{\mathcal{G}|\mathcal{Y},\Phi,\tau^2}(\mathbf{g}|\mathbf{y},\boldsymbol{\phi},\boldsymbol{\tau}^2)$ in (4.40) defines a Gaussian density over \mathbb{R}^n with center and precision given by*

$$\bar{\mathbf{g}} = \Gamma^{-1} \left(\frac{2}{\sigma^2} \tilde{F}_1^T \tilde{\mathbf{y}} \right), \quad (4.41a)$$

$$\Gamma = L^T[D(\boldsymbol{\tau}^2)]^{-1}L + \frac{2}{\sigma^2} \tilde{F}_1^T \tilde{F}_1. \quad (4.41b)$$

Proof. Let $\Gamma = L^T[D(\boldsymbol{\tau}^2)]^{-1}L + \frac{2}{\sigma^2} \tilde{F}_1^T \tilde{F}_1$, and suppose $\mathbf{x} \in \text{Ker}(\Gamma)$ so that $\mathbf{x}^T \Gamma \mathbf{x} = 0$.

By (4.41b) we then have

$$\mathbf{x}^T \Gamma \mathbf{x} = \mathbf{x}^T L^T [D(\boldsymbol{\tau}^2)]^{-1} L \mathbf{x} + \frac{2}{\sigma^2} \mathbf{x}^T \tilde{F}_1^T \tilde{F}_1 \mathbf{x} = 0,$$

that is $\mathbf{x} \in \text{Ker}(L) \cap \text{Ker}(\tilde{F}_1)$. Since $\text{Ker}(L) = \{0\}$, we have immediately that $\text{Ker}(\Gamma) = \{0\}$ so that Γ is invertible. We furthermore observe that Γ is also symmetric positive definite. To explicitly determine the center and precision as $\bar{\mathbf{g}}$ and Γ in (4.41a)

and (4.41b), consider the following from (4.40):

$$\begin{aligned}
-\frac{\|\mathbf{y} - F_1 \mathbf{g}\|_2^2}{\sigma^2} - \frac{1}{2}(\mathbf{L}\mathbf{g})^T [D(\boldsymbol{\tau}^2)]^{-1}(\mathbf{L}\mathbf{g}) &= -\frac{\|\tilde{\mathbf{y}} - \tilde{F}_1 \mathbf{g}\|_2^2}{\sigma^2} - \frac{1}{2}(\mathbf{L}\mathbf{g})^T [D(\boldsymbol{\tau}^2)]^{-1}(\mathbf{L}\mathbf{g}) \\
&= -\frac{1}{2}(\mathbf{L}\mathbf{g})^T [D(\boldsymbol{\tau}^2)]^{-1}(\mathbf{L}\mathbf{g}) - \frac{1}{\sigma^2}(\tilde{\mathbf{y}} - \tilde{F}_1 \mathbf{g})^T (\tilde{\mathbf{y}} - \tilde{F}_1 \mathbf{g}) \\
&= -\frac{1}{2}\mathbf{g}^T \mathbf{L}^T [D(\boldsymbol{\tau}^2)]^{-1} \mathbf{L}\mathbf{g} - \frac{1}{\sigma^2}\tilde{\mathbf{y}}^T \tilde{\mathbf{y}} + \frac{1}{\sigma^2}\mathbf{g}^T \tilde{F}_1^T \tilde{\mathbf{y}} \\
&\quad + \frac{1}{\sigma^2}\tilde{\mathbf{y}}^T \tilde{F}_1 \mathbf{g} - \frac{1}{\sigma^2}\mathbf{g}^T \tilde{F}_1^T \tilde{F}_1 \mathbf{g} \\
&= -\frac{1}{2}\mathbf{g}^T \Gamma \mathbf{g} + \frac{1}{\sigma^2}\mathbf{g}^T \tilde{F}_1^T \tilde{\mathbf{y}} + \frac{1}{\sigma^2}\tilde{\mathbf{y}}^T \tilde{F}_1 \mathbf{g} - \frac{1}{\sigma^2}\tilde{\mathbf{y}}^T \tilde{\mathbf{y}} \\
&= -\frac{1}{2}\mathbf{g}^T \Gamma \mathbf{g} + \frac{2}{\sigma^2}\mathbf{g}^T \tilde{F}_1^T \tilde{\mathbf{y}} - \frac{1}{\sigma^2}\tilde{\mathbf{y}}^T \tilde{\mathbf{y}} \\
&= -\frac{1}{2}\mathbf{g}^T \Gamma \mathbf{g} + \mathbf{g}^T \Gamma \bar{\mathbf{g}} - \frac{1}{2}\bar{\mathbf{g}}^T \Gamma \bar{\mathbf{g}} + \frac{1}{2}\bar{\mathbf{g}}^T \Gamma \bar{\mathbf{g}} - \frac{1}{\sigma^2}\tilde{\mathbf{y}}^T \tilde{\mathbf{y}} \\
&= -\frac{1}{2}(\mathbf{g} - \bar{\mathbf{g}})^T \Gamma (\mathbf{g} - \bar{\mathbf{g}}) + c(\tilde{\mathbf{y}}),
\end{aligned}$$

where $c(\tilde{\mathbf{y}}) = \frac{1}{2}\bar{\mathbf{g}}^T \Gamma \bar{\mathbf{g}} - \frac{1}{\sigma^2}\tilde{\mathbf{y}}^T \tilde{\mathbf{y}}$ is mutually independent of \mathbf{g} . Thus we have mean $\bar{\mathbf{g}}$ and precision Γ , yielding the desired result. \square

When F is unitary, the resulting Gaussian density attains a simpler form provided by Corollary 4.2 which enables faster sampling.

Corollary 4.2. *Let $F_1 = FD(e^{i\phi})$ and suppose F is unitary and L has rank n . For $F_1^H \mathbf{y} = \mathbf{q} \odot e^{i\varphi}$ where $\mathbf{q} \in (\mathbb{R}^+)^n$ and $\varphi \in [-\pi, \pi]^n$, the function $\tilde{f}_{\mathcal{G}|\mathcal{Y}, \mathcal{T}^2, \Phi}(\mathbf{g}|\mathbf{y}, \boldsymbol{\tau}^2, \boldsymbol{\phi})$ in (4.40) defines a Gaussian density over \mathbb{R}^n with center and precision given by*

$$\bar{\mathbf{g}} = \Gamma^{-1} \left(\frac{2}{\sigma^2} \mathbf{q} \odot \cos(\boldsymbol{\varphi}) \right), \quad (4.42a)$$

$$\Gamma = L^T [D(\boldsymbol{\tau}^2)]^{-1} L + \frac{2}{\sigma^2} \mathbb{I}_n. \quad (4.42b)$$

Proof. Since F is unitary, we have in (4.40)

$$\begin{aligned}
-\frac{\|\mathbf{y} - F_1 \mathbf{g}\|_2^2}{\sigma^2} - \frac{1}{2} (L\mathbf{g})^T [D(\boldsymbol{\tau}^2)]^{-1} (L\mathbf{g}) &= -\frac{1}{2} \mathbf{g}^T \Gamma \mathbf{g} + \frac{1}{\sigma^2} \mathbf{g}^T (\mathbf{q} \odot e^{i\varphi}) \\
&\quad + \frac{1}{\sigma^2} (\mathbf{q} \odot e^{-i\varphi})^T \mathbf{g} - \frac{1}{\sigma^2} \mathbf{y}^H \mathbf{y} \\
&= -\frac{1}{2} \mathbf{g}^T \Gamma \mathbf{g} + \frac{1}{\sigma^2} \sum_{j=1}^n g_j q_j (e^{i\varphi_j} + e^{-i\varphi_j}) - \frac{1}{\sigma^2} \mathbf{y}^H \mathbf{y} \\
&= -\frac{1}{2} \mathbf{g}^T \Gamma \mathbf{g} + \frac{2}{\sigma^2} \sum_{j=1}^n g_j q_j \cos(\varphi_j) - \frac{1}{\sigma^2} \mathbf{y}^H \mathbf{y} \\
&= -\frac{1}{2} \mathbf{g}^T \Gamma \mathbf{g} + \mathbf{g}^T \Gamma \bar{\mathbf{g}} - \frac{1}{2} \bar{\mathbf{g}}^T \Gamma \bar{\mathbf{g}} + \frac{1}{2} \bar{\mathbf{g}}^T \Gamma \bar{\mathbf{g}} - \frac{1}{\sigma^2} \mathbf{y}^H \mathbf{y} \\
&= -\frac{1}{2} (\mathbf{g} - \bar{\mathbf{g}})^T \Gamma (\mathbf{g} - \bar{\mathbf{g}}) + c(\mathbf{y})
\end{aligned}$$

where $c(\mathbf{y}) = \frac{1}{2} \bar{\mathbf{g}}^T \Gamma \bar{\mathbf{g}} - \frac{1}{\sigma^2} \mathbf{y}^H \mathbf{y}$ is mutually independent of \mathbf{g} . Thus we have mean $\bar{\mathbf{g}}$ and precision Γ , yielding the desired result. \square

A couple of remarks are in order.

Remark 4.6. If the density (4.40) has little mass in the region where $\mathbf{g} \geq 0$, this rejection method may become computationally infeasible. Other techniques exist for sampling from truncated multivariate normal distributions, such as the exact Hamiltonian Monte Carlo method in [96] or a Gibbs sampling technique that updates each g_i separately for $i = 1, \dots, n$. These methods are often more computationally expensive and may become cost-prohibitive for high dimensional problems, however.

Remark 4.7. When $\bar{\mathbf{g}}$ in (4.41a) and (4.42a) has mostly nonnegative entries with a few negative elements, sampling $f_{\mathcal{G}|\mathcal{Y},\Phi,\mathcal{T}^2}$ using the rejection technique involving $\tilde{f}_{\mathcal{G}|\mathcal{Y},\Phi,\mathcal{T}^2}$ may become computationally inefficient. In this case the rejection sampling from the mode (RSM) [82] provides a possible alternative. In short, RSM generates an exact sample of $f_{\mathcal{G}|\mathcal{Y},\Phi,\mathcal{T}^2}$ by sampling a shifted truncated Gaussian density followed by an acceptance-rejection step. In our numerical experiments RSM is implemented when

the rejection technique involving (4.40) fails to generate a sample after N_s attempts (in our experiments $N_s = 10$).

Updating the scale mixture parameter \mathcal{T}^2 . From (4.39) we have the conditional posterior

$$f_{\mathcal{T}_j^2|\mathcal{Y},\mathcal{G},\Phi,\mathcal{T}_{-j}^2,\eta^{-2}}(\tau_j^2|\mathbf{y},\mathbf{g},\phi,\tau_{-j}^2,\eta^{-2}) \propto (\tau_j^2)^{-\frac{1}{2}} \exp\left(-\frac{[L\mathbf{g}]_j^2}{2\tau_j^2} - \frac{\tau_j^2\eta^{-2}}{2}\right), \quad j = 1, \dots, k, \quad (4.43)$$

which is equivalent to the density function in the real-valued signal case (4.26) with $\mathcal{X} = \mathcal{G}$. Analogously to what followed there, we recognize (4.43) as an inverse Gaussian distribution with mean parameter $\sqrt{\eta^{-2}/[L\mathbf{g}]_j^2}$ and shape parameter η^{-2} . The conditional posterior on η^{-2} is given by (4.29).

Updating the phase Φ . Lastly from (4.39) we have the phase posterior distribution

$$f_{\Phi|\mathcal{Y},\mathcal{G},\mathcal{T}^2}(\phi|\mathbf{y},\mathbf{g},\boldsymbol{\tau}^2) \propto \exp\left(-\frac{1}{\sigma^2}\|\mathbf{y} - F_2e^{i\phi}\|_2^2\right) \mathbf{1}_{[-\pi,\pi]^n}(\phi), \quad (4.44)$$

where $F_2 = FD(\mathbf{g})$. The nature of (4.44) can be better understood by defining random variable $\Theta = e^{i\Phi}$ and corresponding realization $\boldsymbol{\theta} = e^{i\phi}$. The posterior is then expressed as

$$f_{\Theta|\mathcal{Y},\mathcal{G}}(\boldsymbol{\theta}|\mathbf{y},\mathbf{g}) \propto \exp\left(-\frac{1}{\sigma^2}\|\mathbf{y} - F_2\boldsymbol{\theta}\|_2^2\right) \mathbf{1}_{CS_1}(\boldsymbol{\theta}), \quad (4.45)$$

where CS_1 is the unit circle in the complex plane. Clearly (4.45), and by extension (4.44), are probability density functions of complex Gaussian distributions restricted to the unit circle. For a general forward operator F , each $f_{\Phi_i|\Phi_{-i},\mathcal{Y},\mathcal{G},\mathcal{T}^2}(\phi_i|\phi_{-i},\mathbf{y},\mathbf{g},\boldsymbol{\tau}^2)$ is conditionally von Mises (4.46) for $i = 1, \dots, n$, which we now state in Theorem 4.3.

Theorem 4.3. Let $F_2 = FD(\mathbf{g})$ as in (4.44) and (4.45) and define $A = F_2^H F_2$ with elements $[A]_{j,k} = a_{j,k} e^{i\alpha_{j,k}}$, where $a_{j,k} \in \mathbb{R}$ and $\alpha_{j,k} \in [-\pi, \pi)$. Further denote $F_2^H \mathbf{y} = \mathbf{q} \odot e^{i\boldsymbol{\varphi}}$, where $\mathbf{q} \in \mathbb{R}_+^n$ and $\boldsymbol{\varphi} \in [-\pi, \pi)^n$, with

$$u_i = \frac{2q_i}{\sigma^2} \cos \varphi_i - \sum_{\substack{k=1 \\ k \neq i}}^n \frac{2a_{i,k}}{\sigma^2} \cos(\tilde{\alpha}_{i,k} + \phi_k), \quad v_i = \frac{2q_i}{\sigma^2} \sin \varphi_i - \sum_{\substack{k=1 \\ k \neq i}}^n \frac{2a_{i,k}}{\sigma^2} \sin(\tilde{\alpha}_{i,k} + \phi_k),$$

and $\tilde{\alpha}_{i,k} = \text{sgn}(k - i)\alpha_{i,k}$ for $k = 1, \dots, n$. Then

$$f_{\Phi_i | \Phi_{-i}, \mathcal{Y}, \mathcal{G}, \mathcal{T}^2}(\phi_i | \phi_{-i}, \mathbf{y}, \mathbf{g}, \boldsymbol{\tau}^2) \propto f_{vM}(\phi_i | \mu_i, \kappa_i), \quad i = 1, \dots, n,$$

where $f_{vM}(x | \mu, \kappa)$ is the von Mises probability density function, given as

$$f_{vM}(x | \mu, \kappa) = \frac{\exp(\kappa \cos(x - \mu))}{2\pi I_0(\kappa)}, \quad (4.46)$$

with location μ and concentration κ . Here

$$\kappa_i = \sqrt{u_i^2 + v_i^2}, \quad \mu_i = \begin{cases} \arctan\left(-\frac{v_i}{u_i}\right) & \text{if } u_i > 0 \\ \pi/2 & \text{if } u_i = 0 \\ \arctan\left(-\frac{v_i}{u_i}\right) + \pi & \text{if } u_i < 0. \end{cases} \quad (4.47)$$

Note that in (4.46), I_0 is the zeroth order modified Bessel function of the first kind.

Proof. By direct calculation using (4.44), we have

$$\begin{aligned}
f_{\Phi_i|\Phi_{-i},\mathcal{Y},\mathcal{G},\mathcal{T}^2}(\phi_i|\phi_{-i},\mathbf{y},\mathbf{g},\boldsymbol{\tau}^2) &\propto \exp\left(-\frac{1}{\sigma^2}\|\mathbf{y}-F_2e^{i\phi}\|_2^2\right) \\
&\propto \exp\left(-\frac{1}{\sigma^2}\left(e^{-i\phi^T}Ae^{i\phi}-\mathbf{y}^HF_2e^{i\phi}-e^{-i\phi^T}F_2^H\mathbf{y}\right)\right) \\
&= \exp\left(-\frac{1}{\sigma^2}\left(\sum_{\substack{j,k=1 \\ k>j}}^n(a_{j,k}e^{i(\alpha_{j,k}-\phi_j+\phi_k)}+a_{j,k}e^{-i(\alpha_{j,k}-\phi_j+\phi_k)})-2\mathbf{q}^T\cos(\phi-\boldsymbol{\varphi})\right)\right) \\
&= \exp\left(-\frac{1}{\sigma^2}\left(\sum_{\substack{j,k=1 \\ k>j}}^n2a_{j,k}\cos(\alpha_{j,k}-\phi_j+\phi_k)-2\mathbf{q}^T\cos(\phi-\boldsymbol{\varphi})\right)\right) \\
&\propto \exp\left(-\frac{1}{\sigma^2}\left(\sum_{\substack{k=1 \\ k\neq i}}^n2a_{i,k}\left(\cos\phi_i\cos(\phi_k+\tilde{\alpha}_{i,k})\right.\right.\right. \\
&\quad \left.\left.\left.+\sin\phi_i\sin(\phi_k+\tilde{\alpha}_{i,k})\right)-2q_i\cos(\phi_i-\varphi_i)\right)\right) \\
&= \exp\left(\left(\frac{2q_i}{\sigma^2}\cos\varphi_i-\sum_{\substack{k=1 \\ k\neq i}}^n\frac{2a_{i,k}}{\sigma^2}\cos(\phi_k+\tilde{\alpha}_{j,k})\right)\cos\phi_i\right. \\
&\quad \left.+\left(\frac{2q_i}{\sigma^2}\sin\varphi_i-\sum_{\substack{k=1 \\ k\neq i}}^n\frac{2a_{i,k}}{\sigma^2}\sin(\phi_k+\tilde{\alpha}_{i,k})\right)\sin\phi_i\right) \\
&= \exp(u_i\cos\phi_i+v_i\sin\phi_i) \\
&= \exp(\kappa_i\cos(\phi_i-\mu_i)).
\end{aligned}$$

□

Since the calculation of $\{\kappa_i\}_{i=1}^n$ and $\{\mu_i\}_{i=1}^n$ in (4.47) require only scalar operations, our method does not suffer from the curse of dimensionality as n grows large. Furthermore, when F is unitary, $f_{\Phi|\mathcal{Y},\mathcal{G},\mathcal{T}^2}(\boldsymbol{\phi}|\mathbf{y},\mathbf{g},\boldsymbol{\tau}^2)$ is precisely a product of von Mises density functions, as is told in Theorem 4.4.

Theorem 4.4. *Suppose $F_2 = FD(\mathbf{g})$, where F is unitary. Let $F_2^H\mathbf{y} = \mathbf{q} \odot e^{i\boldsymbol{\varphi}}$ where*

$\mathbf{q} \in \mathbb{R}_+^n$ and $\boldsymbol{\varphi} \in [-\pi, \pi]^n$. Then

$$f_{\Phi|\mathcal{Y},\mathcal{G}}(\boldsymbol{\phi}|\mathbf{y}, \mathbf{g}) = \prod_{i=1}^n f_{vM} \left(\phi_i \middle| \varphi_i, \frac{2}{\sigma^2} q_i \right).$$

Proof. Consider the following:

$$\begin{aligned} f_{\Phi|\mathcal{Y},\mathcal{G}}(\boldsymbol{\phi}|\mathbf{y}, \mathbf{g}) &= C \exp \left(-\frac{1}{\sigma^2} \|\mathbf{y} - F_2 e^{i\boldsymbol{\phi}}\|_2^2 \right) \mathbf{1}_{[-\pi, \pi]^n}(\boldsymbol{\phi}) \\ &= C \exp \left(-\frac{1}{\sigma^2} (e^{-i\boldsymbol{\phi}} F_2^H F_2 e^{i\boldsymbol{\phi}} - \mathbf{y}^H F_2 e^{i\boldsymbol{\phi}} - e^{-i\boldsymbol{\phi}} F_2^H \mathbf{y} + \mathbf{y}^H \mathbf{y}) \right) \mathbf{1}_{[-\pi, \pi]^n}(\boldsymbol{\phi}) \\ &= C \exp \left(-\frac{1}{\sigma^2} (\mathbf{g}^T \mathbf{g} - 2\mathbf{q}^T \cos(\boldsymbol{\phi} - \boldsymbol{\varphi}) + \mathbf{y}^H \mathbf{y}) \right) \mathbf{1}_{[-\pi, \pi]^n}(\boldsymbol{\phi}), \end{aligned}$$

where C is a normalization constant. To find C , we integrate the above expression over \mathbb{R}^n as follows:

$$\begin{aligned} \frac{1}{C} &= \int_{\mathbb{R}^n} \exp \left(-\frac{1}{\sigma^2} (\mathbf{g}^T \mathbf{g} - 2\mathbf{q}^T \cos(\boldsymbol{\phi} - \boldsymbol{\varphi}) + \mathbf{y}^H \mathbf{y}) \right) \mathbf{1}_{[-\pi, \pi]^n}(\boldsymbol{\phi}) d\boldsymbol{\phi} \\ &= \int_{[-\pi, \pi]^n} \exp \left(-\frac{1}{\sigma^2} (\mathbf{g}^T \mathbf{g} - 2\mathbf{q}^T \cos(\boldsymbol{\phi} - \boldsymbol{\varphi}) + \mathbf{y}^H \mathbf{y}) \right) d\boldsymbol{\phi} \\ &= \exp \left(-\frac{1}{\sigma^2} (\mathbf{g}^T \mathbf{g} + \mathbf{y}^H \mathbf{y}) \right) \int_{[-\pi, \pi]^n} \exp \left(\frac{2}{\sigma^2} \mathbf{q}^T \cos(\boldsymbol{\phi} - \boldsymbol{\varphi}) \right) d\boldsymbol{\phi} \\ &= \exp \left(-\frac{1}{\sigma^2} (\mathbf{g}^T \mathbf{g} + \mathbf{y}^H \mathbf{y}) \right) \prod_{i=1}^n 2\pi I_0 \left(\frac{2}{\sigma^2} q_i \right). \end{aligned}$$

Thus, we have

$$\begin{aligned} f_{\Phi|\mathcal{Y},\mathcal{G},\mathcal{T}^2}(\boldsymbol{\phi}|\mathbf{y}, \mathbf{g}, \boldsymbol{\tau}^2) &= \frac{\exp \left(-\frac{1}{\sigma^2} (\mathbf{g}^T \mathbf{g} - 2\mathbf{q}^T \cos(\boldsymbol{\phi} - \boldsymbol{\varphi}) + \mathbf{y}^H \mathbf{y}) \right) \mathbf{1}_{[-\pi, \pi]^n}(\boldsymbol{\phi})}{\exp \left(-\frac{1}{\sigma^2} (\mathbf{g}^T \mathbf{g} + \mathbf{y}^H \mathbf{y}) \right) \prod_{i=1}^n 2\pi I_0 \left(\frac{2}{\sigma^2} q_i \right)} \\ &= \prod_{j=1}^n \frac{\exp \left(\frac{2}{\sigma^2} q_j \cos(\phi_j - \varphi_j) \right)}{2\pi I_0 \left(\frac{2}{\sigma^2} q_j \right)} \mathbf{1}_{[-\pi, \pi]^n}(\boldsymbol{\phi}_j) \\ &= \prod_{j=1}^n \pi_{vM} \left(\phi_j \middle| \varphi_j, \frac{2}{\sigma^2} q_j \right), \end{aligned}$$

which completes the proof. \square

When F is unitary, the conditional independence shown in Theorem 4.4 allows us to update each Φ_j independently of Φ_{-j} , increasing the opportunity for parallelization in our technique. Even when F is not unitary, Theorem 4.3 allows us to update each Φ_i sequentially using a Gibbs sampling scheme, although we do not benefit from the same parallelization opportunities as when F is unitary.

To sample the von Mises distribution, we use the wrapped Cauchy distribution, which has probability density function

$$f_{WC}(\theta|\mu, \gamma) = \sum_{k=-\infty}^{\infty} \frac{\gamma}{\pi(\gamma^2 + (\theta - \mu + 2\pi k)^2)}, \quad -\pi < \theta < \pi.$$

Here γ is the scale factor and μ is the mode of the unwrapped distribution. The acceptance-rejection method introduced in [13] utilizes a wrapped Cauchy density as an envelope for sampling from the von Mises distribution (4.46).

Algorithm 4 Sampling the von Mises density using the wrapped Cauchy

Input Location μ and concentration κ of von Mises distribution.

Output Sample θ .

- 1 Set $\tau = 1 + (1 + 4\kappa^2)^{\frac{1}{2}}$, $\rho = (\tau - (2\tau)^{\frac{1}{2}})/(2\kappa)$, and $r = (1 + \rho^2)/(2\rho)$.
 - 2 Generate $u_1 \sim U(0, 1)$, then set $z = \cos(\pi u_1)$, $f = (1 + rz)/(r + z)$, $c = \kappa(r - f)$.
 - 3 Generate $u_2 \sim U(0, 1)$, then if $c(2 - c) - u_2 > 0$, go to step 5.
 - 4 If $\ln(c/u_2) + 1 - c < 0$, return to step 2.
 - 5 Generate $u_3 \sim U(0, 1)$, then set $\theta = [\text{sgn}(u_3 - 0.5)] \cos^{-1}(f)$.
-

Algorithm 5 CVBL for a signal with sparsity in the transformed magnitude

Input data \mathbf{y} , noise variance σ^2 , hyperparameters $r = 1$ and $\delta = 10^{-3}$, sparse transform operator L , chain length N_M , and burn-in length B .

Output samples $\mathbf{z}^{(s-B+1)} = \mathbf{g}^{(s)} \odot \exp(i\phi^{(s)})$ for $s = B, \dots, N_M$.

1 Set $\mathbf{g}^{(0)} = |A^H \mathbf{y}|$, $(\tau^2)^{(0)} = 1$, and $\phi^{(0)} = \arg(A^H \mathbf{y})$

2 **For** $\ell = 1, \dots, N_M$ **do**

i. Draw a sample \mathbf{g}^* from $\tilde{f}_{\mathcal{G}|\mathcal{Y},\Phi,\tau^2}(\mathbf{g}|\bar{\mathbf{y}}, \phi^{(\ell-1)}, (\tau^2)^{(\ell-1)})$ using Theorem 4.1.

ii. If \mathbf{g}^* contains a negative element, return to (i). Otherwise, set $\mathbf{g}^{(\ell)} = \mathbf{g}^*$.

iii. Sample each $(\tau_j^{-2})^{(\ell)}$ from $\mathcal{T}_j^{-2}|\mathcal{G} = \mathbf{g}^{(\ell)}, \eta^{-2} = (\eta^{-2})^{(\ell-1)}$ (4.43) for $j = 1, \dots, k$.

iv. Sample $(\eta^{-2})^{(\ell)}$ from $\mathcal{T}_j^{-2} = (\tau^2)^{(\ell)}$ (4.23d).

v. For each $i = 1, \dots, n$, draw a sample ϕ_i^* from $\Phi_i|\mathcal{Y} = \mathbf{y}, \mathcal{G} = \mathbf{g}^{(\ell)}$ (4.44) using Algorithm 4 and set $\phi_i^{(\ell)} = \phi_i^*$.

We are now ready to sample from the joint distribution $f_{\mathcal{G},\Phi|\mathcal{Y},\eta}(\mathbf{g}, \phi|\mathbf{y}, \eta)$ in (4.39). After initializing our chain, the three-stage Gibbs sampler is implemented, where \mathbf{g} is updated first, followed by the τ^2 update, and concluded by the ϕ update. This is done for some predetermined number of iterations N_M , after which the output chain is formed using all the samples generated after the burn-in period B . This method is summarized in Algorithm 5.⁶

⁶ Both Algorithm 3 and Algorithm 5 may be easily modified to use hyperprior (4.14) instead of (4.15) by fixing the value of $(\eta^{-2})^{(\ell)} = \hat{\eta}^{-2}$ for all $\ell = 1, \dots, N_M$.

Section 4.5

Numerical Results

We now demonstrate the efficacy of the CVBL algorithm given in Algorithm 3 for the sparse magnitude case and Algorithm 5 for the sparse transform of the magnitude case by performing experiments in 1D and 2D using three different forward operators for F in (4.1):

- $F_F \in \mathbb{C}^{n \times n}$: The discrete Fourier transform operator with entries (in 1D) given by

$$[F_F]_{j,k} = \frac{1}{\sqrt{n}} \exp\left(-2\pi i \frac{jk}{n}\right), \quad j, k = 0, \dots, n-1. \quad (4.48)$$

Observe that F_F is unitary, so Theorem 4.4 applies.

- $F_B \in \mathbb{R}^{n \times n}$: A blurring operator with entries (in 1D) given by

$$[F_B]_{j,k} = \begin{cases} \frac{1}{\sqrt{26}}(2^{2-|j-k|}) & \text{if } |j-k| \leq 2 \\ 0 & \text{else} \end{cases}, \quad j, k = 0, \dots, n-1. \quad (4.49)$$

Observe that F_B is a banded Toeplitz matrix, and although non-singular, it is notoriously ill-conditioned.

- $F_U \in \mathbb{C}^{m \times n}$: A random under-sampled Fourier transform matrix which has entries (in 1D) given by (4.48) but with randomly zeroed-out rows so that $m = \lceil \nu n \rceil$, where $0 < \nu \leq 1$. For $\mathcal{M}_\nu \subseteq \{2, 3, \dots, n\}$ such that $|\mathcal{M}_\nu| = m$, F_U

is defined as

$$[F_U]_{j,k} = \begin{cases} [F_F]_{j,k} & \text{if } j \in \mathcal{M}_\nu \\ 0 & \text{else} \end{cases}, \quad j, k = 1, \dots, n. \quad (4.50)$$

Note that the zeroth frequency term is never zeroed-out in the construction of \mathcal{M}_ν .

Remark 4.8. The extension to 2D for F_F and F_U is straightforward. For F_B , the 2D operator is defined such that $F_B \mathbf{z}$ convolves the image \mathbf{z} with the kernel K given by

$$K = \frac{1}{2\sqrt{70}} \begin{bmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 2 & 1 & 0 \\ 1 & 2 & 16 & 2 & 1 \\ 0 & 1 & 2 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix},$$

causing the blurring effect to occur in both dimensions.⁷

In the numerical examples that follow, $L = \mathbb{I}_n$ in the sparse signal case or the first order differencing operator in the case of the magnitude having a sparse gradient, where we enforce zero boundary conditions to ensure that it is of rank n . For 1D signals this amounts to

$$[L]_{i,j} = \begin{cases} 1 & \text{if } i = j, \\ -1 & \text{if } i = j + 1, \\ 0 & \text{else.} \end{cases} \quad (4.51)$$

⁷The 1D and 2D forward operators used in our experiments are explicitly stated for reproducibility purposes.

Finally we assume that the variance σ^2 in our observation model (4.1) is known, with corresponding signal-to-noise (SNR) given by

$$\text{SNR} = 10 \log_{10} \left(\frac{\|F \mathbf{z}^{exact}\|_2^2}{m \sigma^2} \right), \quad (4.52)$$

where $\mathbf{z}^{exact} \in \mathbb{C}^m$ is the exact solution to (4.1). The SNR values chosen in our experiments highlight the effectiveness of the CVBL method in noisy environments.

In all experiments, a total of 5000 samples are drawn with a burn-in period of $B = 200$. To encourage the sampler to move towards high-mass areas of the probability density, for $\ell < B$ we set $\mathbf{g}^{(\ell)} = |\mathbf{g}^*|$, regardless of whether or not \mathbf{g}^* contains negative elements. Our 1D analysis includes figures showing the means and 90% credibility intervals (CI) for the marginal magnitude of the posterior distribution. We also approximate the marginal density function of the phase of random individual pixels using a kernel density estimation technique [16]. In 2D we provide the magnitude means and the size of the 90% CIs. Finally, for each choice of F we also compare the MAP estimate of our CVBL posterior to the corresponding classical LASSO solution \mathbf{z}^{LASSO} .

- **sparse signal case:** We compute the solution to the objective function

$$\mathbf{z}^{LASSO} = \arg \min_{\mathbf{z}} \|\mathbf{F} \mathbf{z} - \mathbf{y}\|_2^2 + \lambda \|\mathbf{z}\|_1 \quad (4.53)$$

by considering the real and imaginary parts of \mathbf{z} , namely \mathbf{a} and \mathbf{b} , respectively, giving the objective function as

$$\begin{bmatrix} \mathbf{a}^{LASSO} \\ \mathbf{b}^{LASSO} \end{bmatrix} = \arg \min_{\mathbf{a}, \mathbf{b}} \left(\left\| \begin{bmatrix} F_R & -F_I \\ F_I & F_R \end{bmatrix} \begin{bmatrix} \mathbf{a} \\ \mathbf{b} \end{bmatrix} - \begin{bmatrix} \mathbf{y}_R \\ \mathbf{y}_I \end{bmatrix} \right\|_2^2 + \lambda \sum_{j=1}^n \sqrt{a_j + b_j} \right), \quad (4.54)$$

which we solve using the alternating direction method of multipliers (ADMM) [17].

- **sparse transform case:** We again employ ADMM to compute the solution in the sparse transform case to the generalized LASSO problem. Here, however, since $|\mathbf{z}|$ is not differentiable, in order to solve

$$\mathbf{z}^{LASSO} = \arg \min_{\mathbf{z}} \|F\mathbf{z} - \mathbf{y}\|_2^2 + \lambda \|L|\mathbf{z}|\|_1, \quad (4.55)$$

we follow what was done in [36, 107] and instead use the diagonal matrix $\Theta^{(k)}$ with non-zero entries $\Theta_{jj}^{(k)} = z_j^{(k)} / |z_j^{(k)}|$ at each iteration k of the algorithm. The objective function then becomes

$$\mathbf{z}^{LASSO} = \arg \min_{\mathbf{z}} \|F\mathbf{z} - \mathbf{y}\|_2^2 + \lambda \left\| L (\Theta^{(k)})^H \mathbf{z} \right\|_1.$$

In both cases we test $\lambda = \alpha\sigma^2/\bar{\eta}$, where $\bar{\eta}$ is the mean of the samples of η generated by the CVBL method. In some sense $\alpha = 1$ represents the “best case” scenario for selecting suitable parameters, while other values of α allow us to test for robustness. We use $\alpha = 1$ in each experiment unless otherwise specified.

4.5.1. Numerical Efficiency

Sampling Gaussians using matrix (e.g. Cholesky) factorization can be prohibitively expensive in high dimensions. This is because new matrices must be factorized for each re-sampling of the hyper-parameters τ^2 , yielding a general cost $\mathcal{O}(n^3)$ flops per sample. The approach detailed in Algorithm 6 provides an efficient way to generate samples of multivariate Gaussian distributions [97, 124], which we use to sample $\mathcal{A}|\mathcal{Y}, \mathcal{B}, \mathcal{T}^2$ and $\mathcal{B}|\mathcal{Y}, \mathcal{A}, \mathcal{T}^2$ in Algorithm 3 and $\mathcal{G}|\mathcal{Y}, \Phi, \mathcal{T}^2$ in Algorithm 5. Moreover, we can efficiently solve the system in Step 3 of Algorithm 6 using the conjugate

gradient method [61].

Algorithm 6 Gaussian sampling by perturbation-optimization

Input Mean $\boldsymbol{\mu}_p \in \mathbb{R}^n$, measurement $\mathbf{y} \in \mathbb{R}^m$, variances Σ_p, Σ_ℓ , sparsifying transform $L \in \mathbb{R}^{k \times n}$, and forward model F corresponding to Gaussian prior $f_{\mathcal{X}} = \mathcal{N}(L\boldsymbol{\mu}_p, \Sigma_p)$ and likelihood $f_{\mathcal{Y}|\mathcal{X}} = \mathcal{N}(F\mathbf{x}, \Sigma_\ell)$.

Output Sample \mathbf{x}_s of the posterior $f_{\mathcal{X}|\mathcal{Y}}$.

- 1 Perturb the prior mean $\tilde{\boldsymbol{\mu}}_p \sim \mathcal{N}(\boldsymbol{\mu}_p, \Sigma_p)$.
 - 2 Perturb the data $\tilde{\mathbf{y}} \sim \mathcal{N}(\mathbf{y}, \Sigma_\ell)$.
 - 3 Solve $(L^T \Sigma_p^{-1} L + F^T \Sigma_\ell^{-1} F) \mathbf{x}_s = (L^T \Sigma_p^{-1} \tilde{\boldsymbol{\mu}}_p + F^T \Sigma_\ell^{-1} \tilde{\mathbf{y}})$ with respect to \mathbf{x}_s .
-

Finally we point out that when the forward operator is given by F_B (4.49) and the magnitude sparsity is in the gradient domain, we use a block sampling approach to generate samples of the phase (4.44), allowing us to take advantage of the inherent sparsity in F_B and to update portions of the phase in parallel.

4.5.2. Sparsity in Magnitude

For the sparse magnitude case, we perform experiments with Algorithm 3 using 5000 samples and the forward operators in (4.48), (4.49), and (4.50). Figure 4.2 shows the magnitude means and 90% CIs for the 1D noisy experiments for SNR = 20 dB, (4.52).

Although the recovered magnitude means are not sparse, we see in Figure 4.3 that the real and imaginary components of the signal are close to zero outside of the signal support. Combined with Figure 4.2 our results show that, as expected, the CVBL provides mean information similar to that of the classical LASSO point-estimate technique (4.53).

Figure 4.4 displays the phase results for each forward operator at a randomly selected “on” pixel (where $\mathbf{g}_j > 0$).⁸ In all three experiments, the CVBL method recovers the support of the signal as well as uncertainty information for both the

⁸The probability density plots in Figure 4.4, Figure 4.6, and Figure 4.15 are formed using kernel density estimation techniques [16].

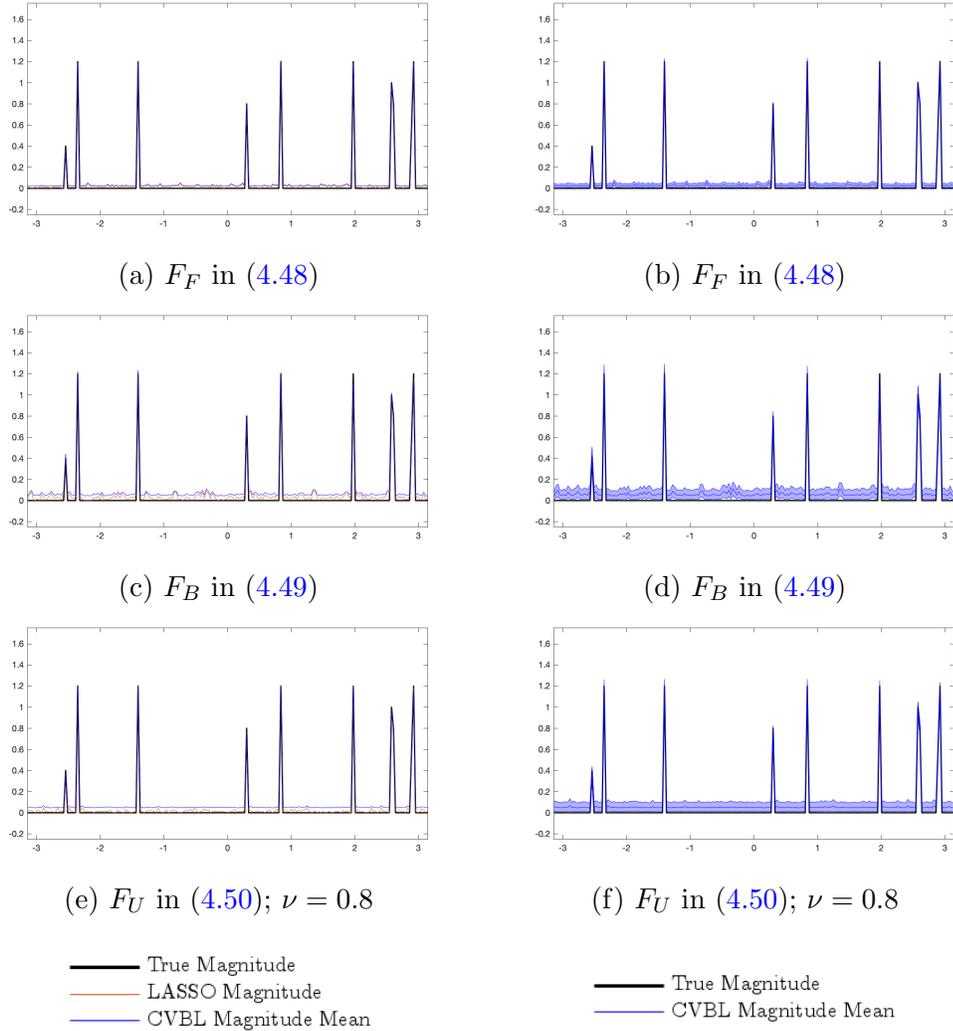


Figure 4.2: Recovered magnitude values for complex-valued signals with sparse magnitude. (left) The magnitude of the LASSO solution and the mean of the CVBL output; (right) 90% CI for the magnitude reconstruction. Here SNR = 20 dB.

magnitude and the phase. We emphasize that this phase information is not recoverable when using RVBL and is one of the primary advantages of using CVBL.

4.5.3. Sparsity in Transform of Magnitude

We now consider the case where sparsity is expected in some linear transform domain of the magnitude. Here the magnitude $\mathbf{g} = \{g_j\}_{j=1}^{200}$ is given by $\{f(t_j)\}_{j=1}^{200}$ for the

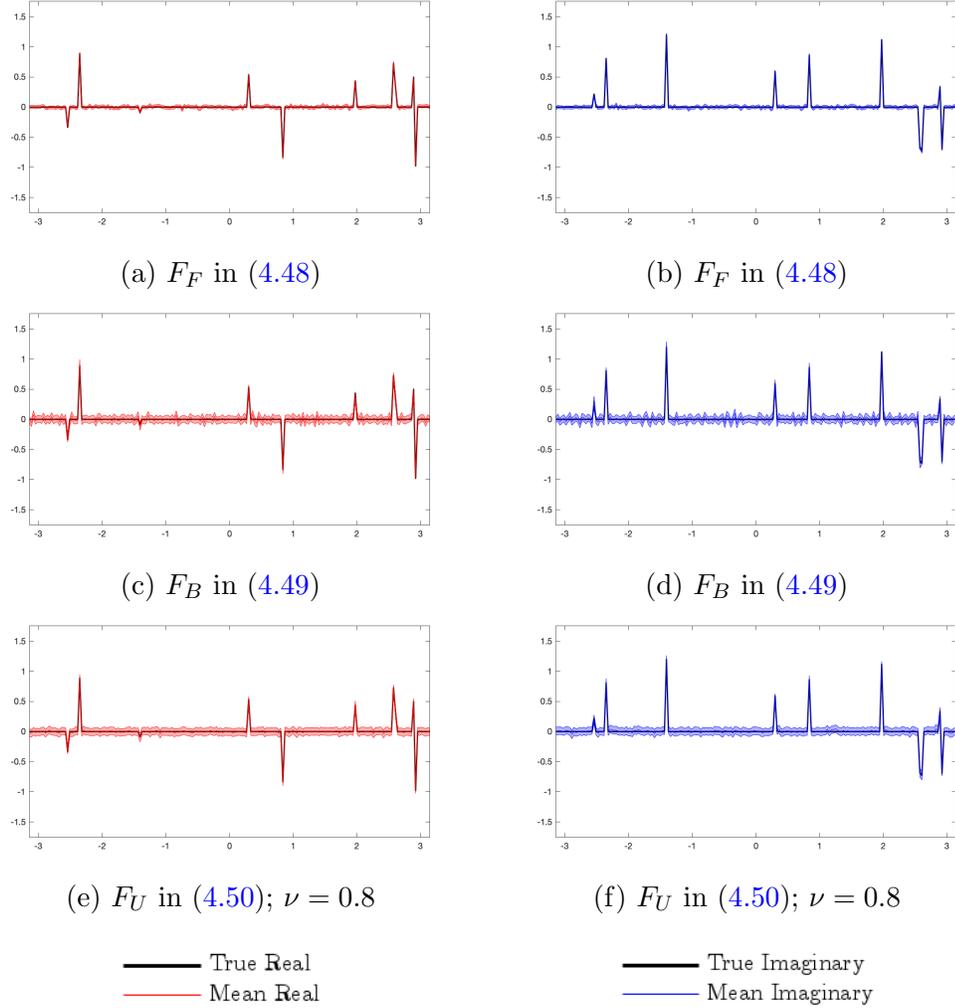


Figure 4.3: Means of (left) real and (right) imaginary parts of the signal in the sparse magnitude case. Shaded regions indicate 90% CIs for the CVBL recovery. SNR = 20 dB.

function $f : [-\pi, \pi) \rightarrow \mathbb{R}_+$ defined as

$$f(t) = \begin{cases} 2, & -2.8 \leq t \leq -2.1 \\ 1.5 & -1.6 \leq t \leq -1.3 \\ 1 + \frac{3}{2} \exp\left(\left(\frac{t-\pi/2}{2/3}\right)^2\right) & t > 0 \\ 1 & \text{else} \end{cases} \quad (4.56)$$

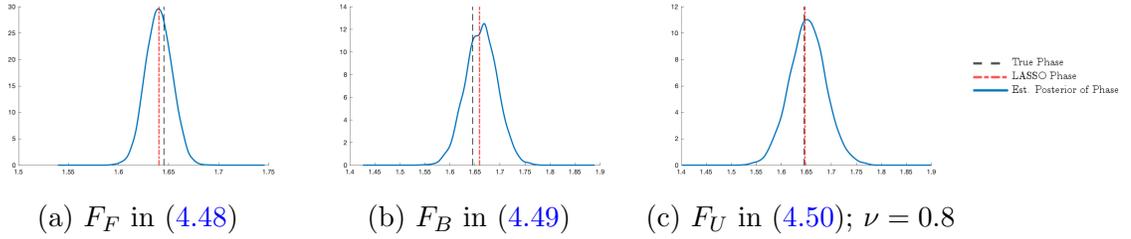


Figure 4.4: Phase estimate in the sparse magnitude case at a randomly selected “on” pixel ($\mathbf{g}_j = 1$). True phase, LASSO phase point estimate, and CVBL marginal posterior for the phase. SNR = 20 dB.

and $t_j = -\pi + \frac{j\pi}{100}$. Each pixel of the corresponding phase $\Phi = \{\phi_j\}_{j=1}^{200}$ is randomly chosen uniformly from $[-\pi, \pi)$. Observe that while $|f|(t)$ is sparse in the gradient domain when $t \leq 0$, this is not the case when $t > 0$. Since our prior is based on the assumption that the gradient domain is sparse, using (4.56) allows us to test the effectiveness of the CVBL method in regions where the gradient is nonzero. As already noted in Section 4.2.3, other choices of sparsifying transform operators such as HOTV may be more suitable and our method is not inherently limited to using the identity or TV operators. For simplicity, as well as to emphasize robustness of our approach, we use the differencing operator in (4.51) and leave other operators for future investigations.

Figure 4.5 compares results for recovering (4.56) from Algorithm 5 where 5000 samples were generated to the LASSO solution in (4.55) with $\lambda = \alpha\sigma^2/\bar{\eta}$, where $\bar{\eta}$ is again the mean of the samples of η generated by the CVBL method and $\alpha = .1, 1, 10$. Each of the three transforms, F_F (4.48), F_B (4.49), and F_U (4.50) with $\nu = 0.8$ were considered for SNR = 20 dB. The LASSO method is clearly sensitive to the choice of regularization parameter, and is most accurate for $\alpha = 1$.

Figure 4.6 demonstrates that the CVBL method recovers marginal posterior density functions for the phase that are consistent with the estimates calculated by (4.55). The true phase is also located in the regions of large mass generated by the corre-

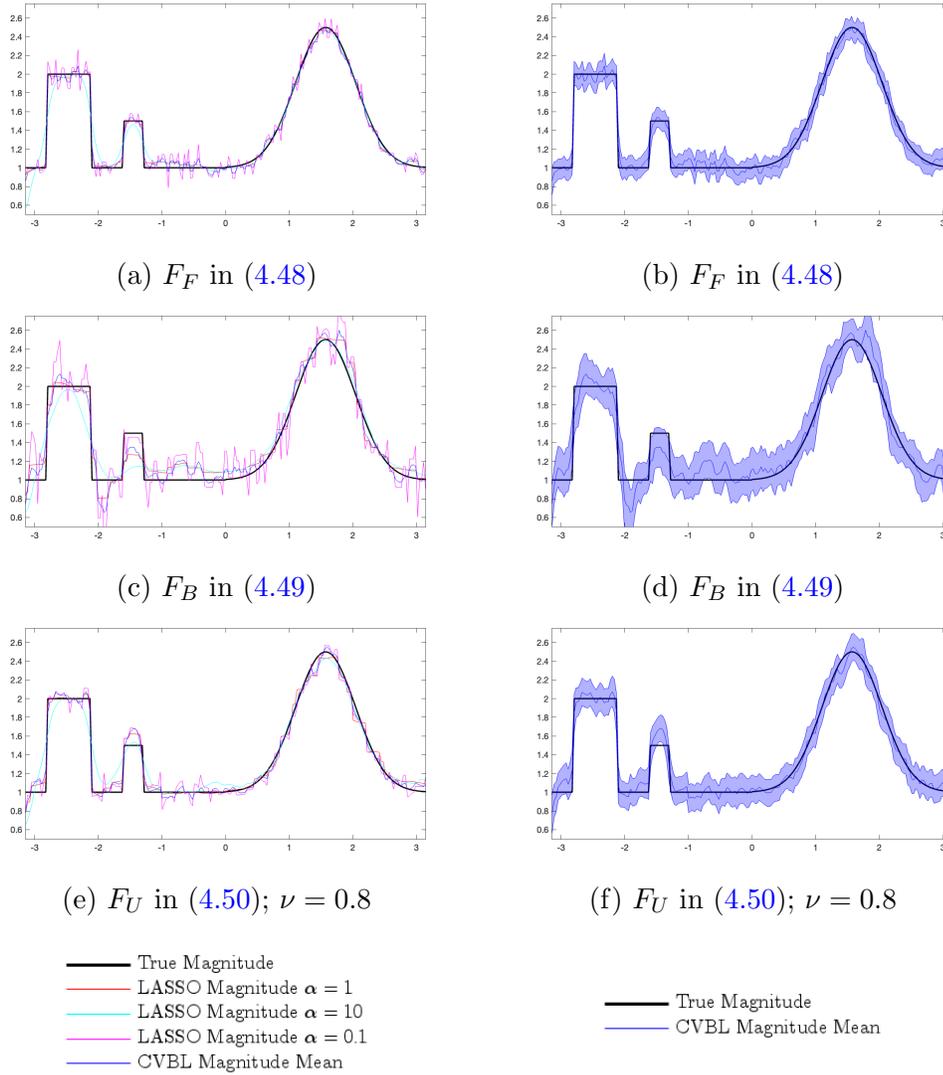


Figure 4.5: Magnitude recovery for \mathbf{g} given by (4.56). (left) The CVBL mean along with the magnitude of LASSO solutions using different regularization parameters; (right) 90% CI for the magnitude reconstruction. Here $\text{SNR} = 20$ dB.

sponding kernel density estimation. Figure 4.7 shows the mean of the τ^2 samples (4.43) for each experiment, which as expected is largest in support regions of the sparse domain.

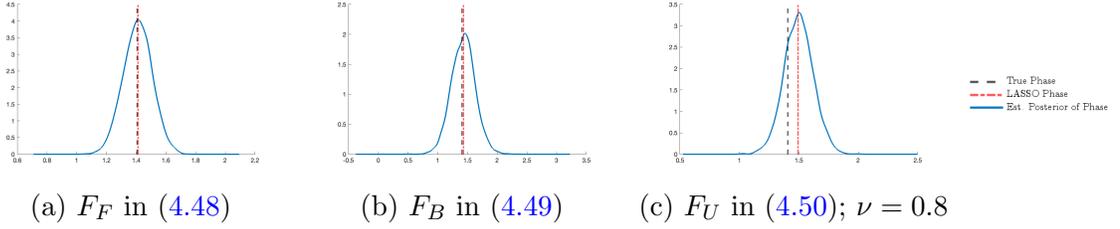


Figure 4.6: Phase estimate in the sparse transform case of a random pixel. True phase (black dash), LASSO phase point estimate (red dash-dot), and CVBL marginal posterior for the phase. SNR = 20 dB.

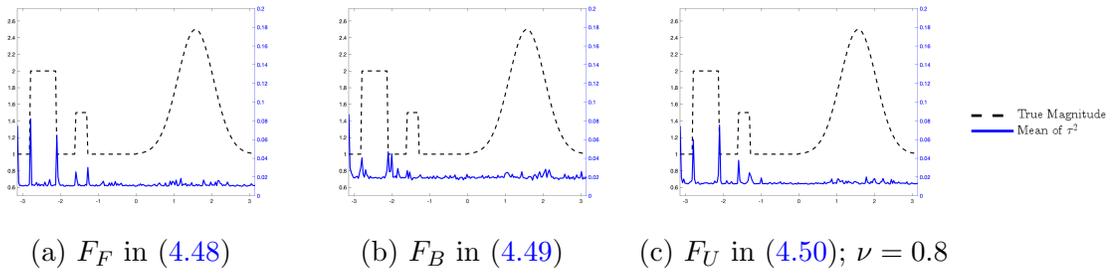


Figure 4.7: Mean of τ^2 (right axis) compared with the true magnitude (left axis). SNR = 20 dB.

4.5.4. Noise Study

We now analyze the effect of noise on Algorithm 5. To this end for the sparse magnitude case and the signal corresponding to (4.56) we consider $10 \text{ dB} \leq \text{SNR} \leq 30 \text{ dB}$.

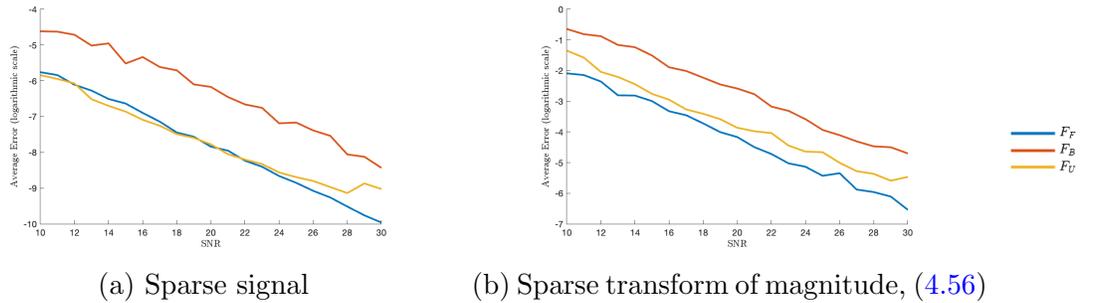


Figure 4.8: Average error in the sample mean generated using Algorithm 5 for increasing SNR (dB). The vertical axes are in a natural logarithmic scale.

Figure 4.8 demonstrates that the average error is similar for all three forward operators across the range of SNR values, with a greater overall range of errors with

respect to the forward operator in the sparse transform case. These results are consistent with what is observed in Figure 4.5 as well as what is apparent in Figure 4.5 (b). In particular the approximation in the region $[-1.7, -1]$ is not well resolved. More insight is provided in Figure 4.7 (b), where it is evident that the support in the sparse domain in that range is not clearly identified.

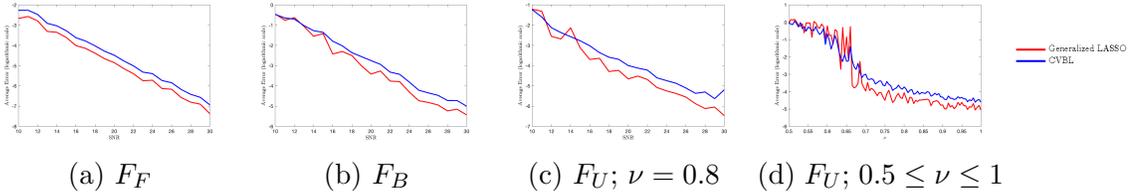


Figure 4.9: Comparison of the average phase error in mean of samples generated using CVBL with the LASSO solution (4.55) for the test function (4.56) at (a)-(c) increasing SNR and (d) increasing sampling rate $0.5 \leq \nu \leq 1$ with SNR = 20 dB. The vertical axes are in a natural logarithmic scale.

Figure 4.9 (a)-(c) compares the average phase error over a range of SNR values for the CVBL and generalized LASSO point estimates. While the LASSO method outperforms Algorithm 5 for each choice of F with increasing SNR, the error difference is negligible for low SNR values for F_B and F_U . Hence we see that the optimal *point estimate* recovery algorithm essentially depends on the SNR and sampling rate of the observable data. *Uncertainty information*, however, is only acquired when using CVBL, as the generalized LASSO technique does not infer the phase information.

Finally, Figure 4.9 (d) compares the average phase error in the sample mean for the function with magnitude given by (4.56) using CVBL with the LASSO solution for the undersampled observable data case. Specifically, the data are obtained using F_U in (4.50) for a range of sampling rates ν while the SNR is held constant at 20 dB. While consistent with Figure 4.9 (c), this result also suggests that for smaller values of ν (more undersampling) the CVBL method provides on average a lower average phase error than the generalized LASSO.

4.5.5. 2D Experiments

Our 2D experiments consider sparsity in the signal magnitude with $\text{SNR} = 0$ dB and sparsity in the signal magnitude gradient with $\text{SNR} = 15$ dB and $\text{SNR} = 25$ dB.

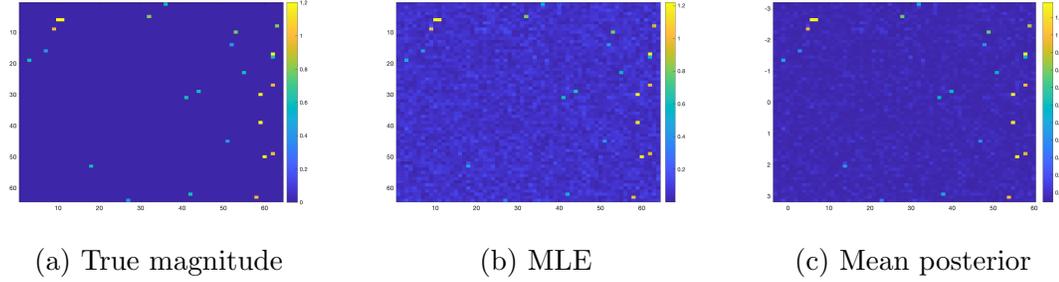


Figure 4.10: Maximum likelihood and mean posterior point estimates of a sparse magnitude complex-valued signal for forward operator F_F in (4.48). $\text{SNR} = 0$ dB.

Figure 4.10 displays point estimates for the magnitude using F_F while Figure 4.11 shows the corresponding results for F_B and F_U . In each case the same 64×64 sparse images is used as input and we simulate 5000 samples of the posterior. We observe that much of the background noise is suppressed while the fidelity of the support of the signal is maintained when compared to the maximum likelihood point estimate.

We now consider the case of promoting sparsity in the magnitude gradient for the Shepp-Logan phantom image [115] depicted in Figure 4.12 (a) on a 256×256 grid. A random phase is then added to each pixel, so that the resulting image \mathbf{z} is modeled by (4.1). Figure 4.12 and Figure 4.14 compare the posterior means of the

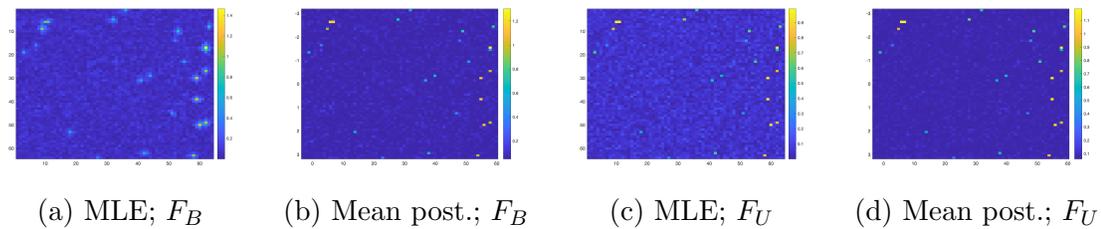


Figure 4.11: Maximum likelihood and mean posterior point estimates recoveries of a sparse magnitude complex-valued signal. Forward operator (a-b) F_B in (4.49) (c-d) F_U $\nu = 0.8$ in (4.50). $\text{SNR} = 0$ dB.

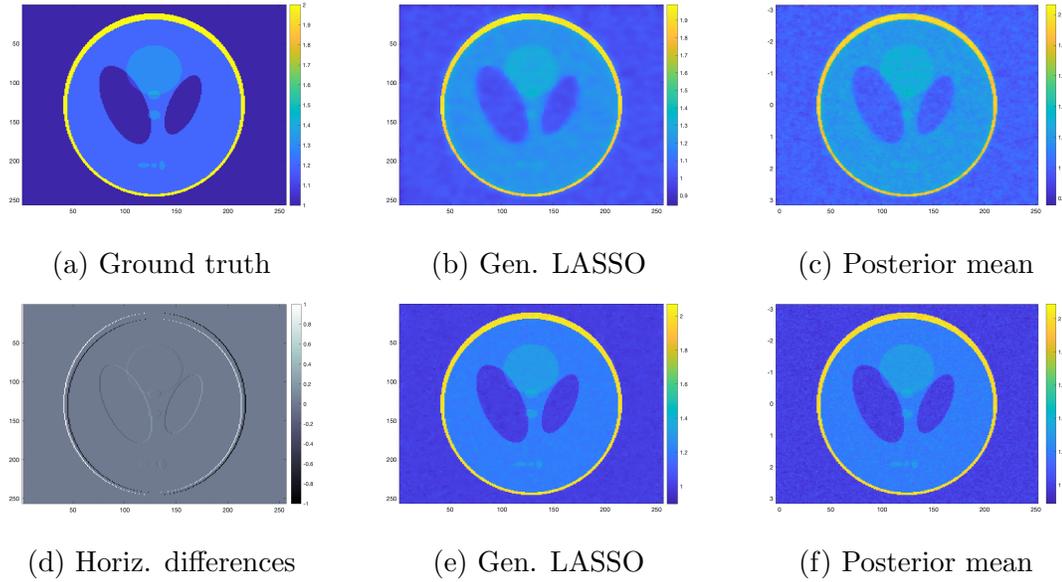


Figure 4.12: Magnitude recovery of the complex-valued Shepp Logan phantom using F_F in (4.48). (a) Shepp-Logan phantom; (d) horizontal differences (computed using $L|z^{exact}|$) of the ground truth. (b)-(c) and (e)-(f) correspond to SNR values of 15 dB and 25 dB, respectively. $\hat{\eta} = 10^{-2}$.

CVBL to the generalized LASSO estimates for F_F in (4.48) and F_B in (4.49).⁹ To avoid the issue discussed in Remark 4.2, in the Shepp Logan experiments we fix η to be $\hat{\eta} = 10^{-2}$, which was chosen heuristically, with $\lambda = \sigma^2/\hat{\eta}$ used for the generalized LASSO regularization parameter. The hyperprior in turn is given by (4.14) rather than (4.15).

Figure 4.12, Figure 4.13, and Figure 4.14 indicate that the generalized LASSO estimate (4.55) is smoother when compared to the CVBL posterior mean. In particular we observe that the three lower circles of the generalized LASSO estimate of the phantom are “blurred” together but remain distinct for the CVBL. These results seem to indicate that when using the CVBL method, a smaller value of $\hat{\eta}$, that is, increasing the reliance on the prior, is needed to match the amount of regularization

⁹The dense nature of F_U in (4.50) makes exact sampling of the phase ϕ impractical for large signals. Approximate or stochastic techniques may expedite sampling, but is beyond the scope of the current investigation.

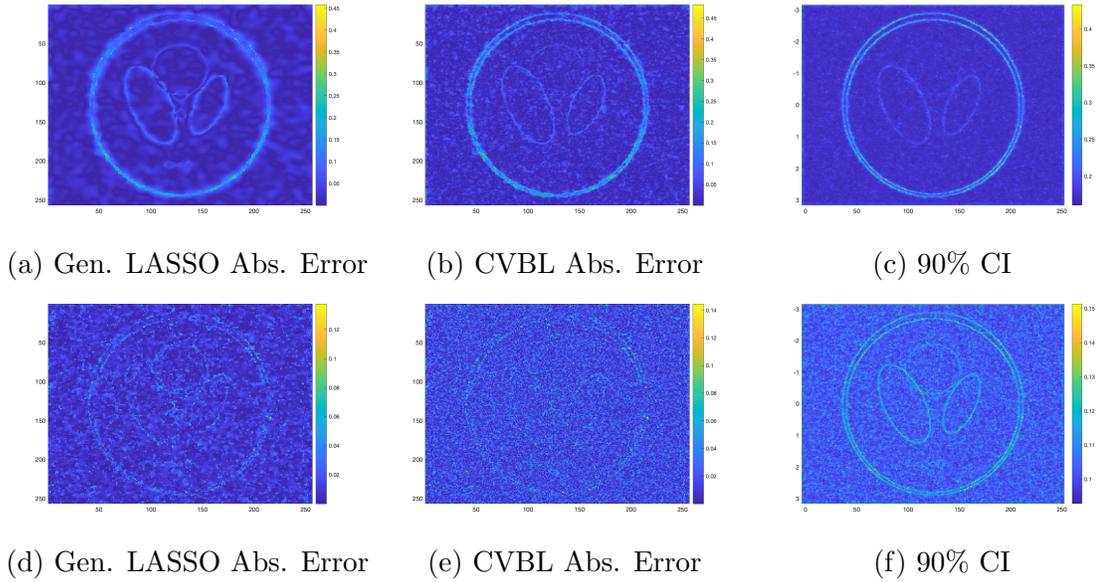


Figure 4.13: Magnitude recovery of the complex-valued Shepp Logan phantom using F_F in (4.48). (a)-(c) and (d)-(f) correspond to SNR values of 15 dB and 25 dB, respectively. $\hat{\eta} = 10^{-2}$.

apparent the generalized LASSO approach.

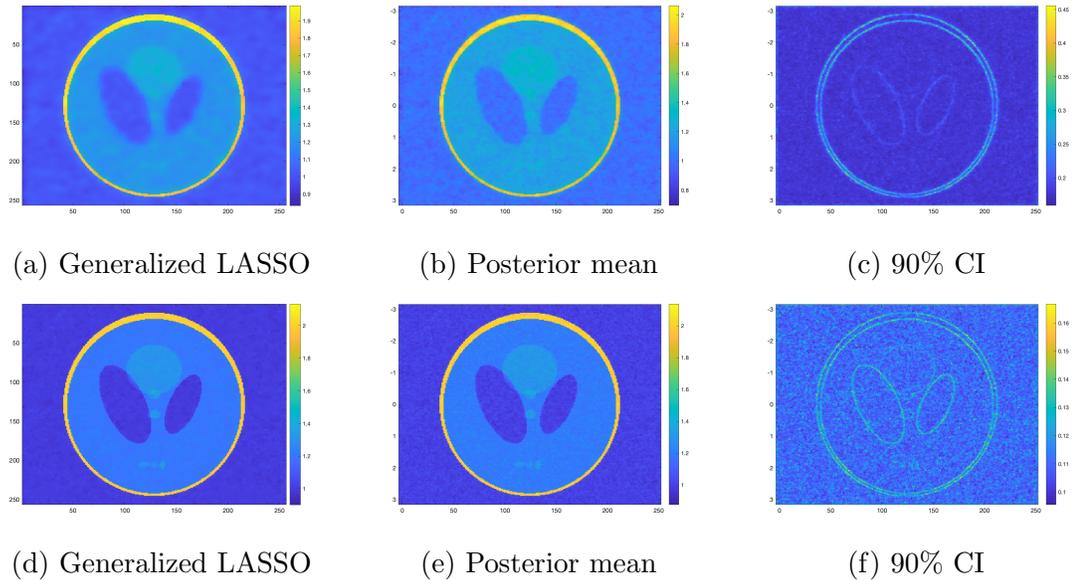


Figure 4.14: Magnitude recovery of the complex-valued Shepp Logan phantom using F_B in (4.49). (a)-(c) and (d)-(f) correspond to SNR values of 15 dB and 25 dB, respectively. $\hat{\eta} = 10^{-2}$.

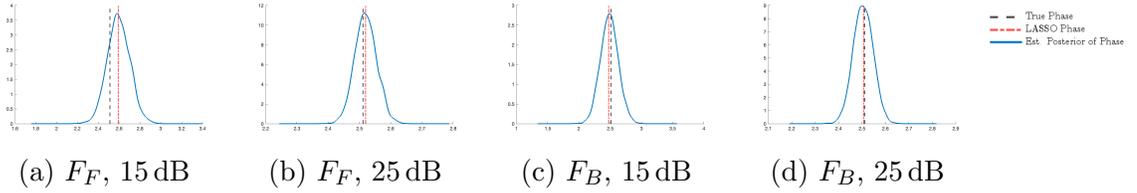


Figure 4.15: Recovered phase at a randomly chosen pixel of the complex-valued Shepp Logan image using F_F in (4.48) and F_B in (4.49) with SNR of 15 dB and 25 dB. $\hat{\eta} = 10^{-2}$.

Finally, Figure 4.15 demonstrates that the true phase of a randomly selected pixel is within an area of high density of the kernel density estimation of the marginal posterior density of the phase, and that the kernel density estimation mode is consistent with those obtained using the generalized LASSO technique.

Section 4.6

Concluding remarks and future work

This investigation extends the real-valued Bayesian LASSO (RVBL), which was originally designed to promote sparsity in a sparse signal, in two ways. We first show that it can be modified to promote sparsity in a chosen transform (here the gradient) domain. We then demonstrate that the RVBL can be further expanded to include *complex*-valued signals and images. We call our method the complex-valued Bayesian LASSO (CVBL). Our numerical experiments show that the CVBL can efficiently recover samples from the entire complex-valued posterior density function, enabling uncertainty quantification of both the magnitude and phase of the true signal.

The CVBL is practical for coherent imaging problems with unitary or sparse forward operators, since it is easily parallelizable. Developing surrogates for dense forward operators will be necessary to efficiently sample large problems, and will be the focus of future investigations. We will also consider different sparse transform operators (such as HOTV) along with adaptive empirical hyperparameters to avoid

the pitfalls of over-regularization. Recent work in [\[138\]](#) may be useful in this regard.

Chapter 5

Empirical Bayesian Inference for Complex-Valued Signals Using Support-Informed Priors

Section 5.1

Introduction

In spotlight synthetic aperture radar (SAR), a moving radar platform is used to send and receive a complex-valued signal, with multiple data acquisitions often made over the same scene. The measured SAR phase history data can be modeled as a continuous (non-uniform) Fourier transform of the complex-valued reflectivity function, leading to the corresponding discretized SAR recovery problem

$$\hat{\mathbf{y}}^j = F\mathbf{z} + \hat{\boldsymbol{\varepsilon}}^j, \quad j = 1, \dots, J, \quad (5.1)$$

where $\hat{\mathbf{y}}^j \in \mathbb{C}^m$, $j = 1, \dots, J$, are J noisy independent measurements, $\mathbf{z} \in \mathbb{C}^n$ is the unobserved variable of interest, $F \in \mathbb{C}^{m \times n}$ is the (non-uniform) discrete Fourier

transform matrix, and each $\hat{\boldsymbol{\epsilon}}^j \sim \mathcal{CN}(0, \Sigma_j)$ is independent, circularly symmetric complex Gaussian noise in \mathbb{C}^m with a symmetric positive definite covariance matrix Σ_j . A comprehensive discussion of spotlight SAR from a signal processing perspective may be found in [66]. Other coherent imaging modalities modeled similarly to (5.1) include ultrasound [135] and digital holography [134].

In this investigation we develop a new algorithm to solve (5.1) that encodes prior information regarding \mathbf{z} , or more specifically its magnitude, since prior phase information is typically unknown. We seek to recover a point estimate solution and to quantify its uncertainty. To this end we adopt a Bayesian approach by letting the data $\hat{\mathbf{y}}^j$, noise $\hat{\boldsymbol{\epsilon}}^j$, and signal \mathbf{z} describe samples of random variables $\hat{\mathcal{Y}}$, $\hat{\mathcal{E}}$, and \mathcal{Z} , respectively, defined over a common probability space, with each variable admitting a well-defined probability density. We also assume that \mathcal{Z} and $\hat{\mathcal{E}}$ are mutually independent, yielding the analogous formulation to (5.1)

$$\hat{\mathcal{Y}} = F\mathcal{Z} + \hat{\mathcal{E}}. \quad (5.2)$$

A particular observation $\hat{\mathbf{y}}^j$ for $j = 1, \dots, J$ then gives (5.1). The posterior distribution $f_{\mathcal{Z}|\hat{\mathcal{Y}}}(\mathbf{z}|\hat{\mathbf{y}})$ of the unknown random variable \mathcal{Z} given data $\hat{\mathcal{Y}}$ and prior assumptions on \mathcal{Z} has the corresponding density function characterized by Bayes' theorem as¹

$$f_{\mathcal{Z}|\hat{\mathcal{Y}}}(\mathbf{z}|\hat{\mathbf{y}}) \propto f_{\hat{\mathcal{Y}}|\mathcal{Z}}(\hat{\mathbf{y}}|\mathbf{z})f_{\mathcal{Z}}(\mathbf{z}). \quad (5.3)$$

Here $f_{\hat{\mathcal{Y}}|\mathcal{Z}}(\hat{\mathbf{y}}|\mathbf{z})$ is the likelihood density function determined by our forward operator F and assumptions on $\hat{\boldsymbol{\epsilon}}$, and $f_{\mathcal{Z}}(\mathbf{z})$ is the density of the prior distribution encoding *a priori* assumptions about the solution. We note that (5.3) inherently contains the likelihood and prior conditionals on hyperpriors, along with their respective density

¹To avoid cumbersome notation, we drop the superscript j when the dependence on j is not relevant.

functions, which will be included in our construction.

In SAR and other coherent imaging problems, accurately recovering both the magnitude and phase is important for downstream processing tasks such as coherent change detection and target location [87, 88]. Providing uncertainty quantification (UQ) is also valuable as it gives practitioners information regarding the reliability of the recovery.

Current methodology

SAR image recovery from phase history data is a highly studied area of research. Techniques such as matched-filter and backprojection are capable of creating high-fidelity reconstructions of the magnitude, and the nonuniform discrete fast Fourier transform (NUFFT) can map the measurement data directly to the complex-valued spatial domain [54, 118]. Regularization methods have also been used to leverage prior information. In particular, since the intensity of the underlying image is presumably sparse in some (e.g. gradient) domain, the point estimate solution can be obtained by solving

$$\mathbf{z}^* = \arg \min_{\mathbf{z}} \left(\frac{1}{2} \|F\mathbf{z} - \bar{\mathbf{y}}\|_2^2 - \frac{\lambda}{p} \|L\mathbf{z}\|_p^p \right), \quad p = 1, 2, \quad (5.4)$$

where L transforms \mathbf{z} to the sparse domain, $\bar{\mathbf{y}}$ is the mean of the data observations, and $\lambda > 0$ is a regularization parameter. We note the extensive use of compressive sensing (CS) algorithms ($p = 1$) [24, 28, 41] for recovering SAR intensity values [89, 102, 117].

SAR data can be collected at various elevation and azimuth angles, and here we refer to these distinct data collections over the same scene as *multiple measurement vectors* (MMVs). The *joint sparsity* assumption, that is, the assumption that the sparse domain of the underlying signal is similar across all collected measurements of

that same signal, has been exploited to produce high quality SAR signal recovery in noisy environments [56, 107].

A *spatially-weighted diagonal* regularization matrix, $W = \text{diag}(w_1, \dots, w_n)$, replaced the constant parameter λ in (5.4) in the so-called *Variance Based Joint Sparsity* (VBJS) method developed in [2, 47, 112]. The essential idea behind VBJS is to construct W so that the solution of (5.4) is more heavily penalized in true sparse regions than in those of support in the sparse domain, i.e. W is *support-informed*. Moreover, each $w_i > 0$, $i = 1, \dots, n$, reflects a relative confidence regarding that support, which is determined by the *variance* at each pixel in the sparse domain. Importantly, the results in [2, 112] demonstrated that W can be *directly* constructed from the MMV data, reducing potential information loss due to pre-processing. VBJS is also more computationally efficient since applying W means that $p = 2$ is appropriate in (5.4), as opposed to standard CS algorithms that require $p = 1$. Finally, VBJS was shown to effectively suppress SAR speckle in [111].

Using a Bayesian framework for SAR image recovery is also becoming more widespread. For example, the technique introduced in [40] uses a Bayesian approach for the full azimuth speckle model to obtain a *maximum a posteriori* (MAP) estimate of the SAR image (although it does not provide any UQ). The method in [35] expands on this approach to recover uncertainty information, but is computationally expensive due to inefficient sampling. It also does not address the anisotropic nature of scatterers. A follow-up paper [36] uses multiple sub-apertures to help mitigate cost and alleviate the faulty assumptions about the anisotropic scatterers, but does not fully exploit the joint MMV information, and in particular does not track the support locations in the sparse domain. Moreover, obtaining the distribution requires some strong assumptions regarding the independence of the random variables for the sub-apertures. These limitations have motivated us to develop a more efficient sampling

approach that also directly uses the MMV data to provide support information in the signal magnitude sparse domain, consequently reducing the uncertainty in the complex-valued image recovery.

Our contribution

We introduce an empirical Bayesian method that uses the given MMV Fourier data and *a priori* assumptions regarding the magnitude of the variable of interest to recover the posterior distribution of a complex-valued signal. Expanding on the support-informed empirical Bayesian approach developed for real-valued signals from MMV data in [138], our new method adapts the VBS approach in [2] to construct a weighted mask *directly* from the given data that encodes the relative confidence regarding the support locations in the signal magnitude gradient domain, which is presumably sparse.² This support-informed mask is then incorporated into the prior, thereby reducing the uncertainty in the posterior recovery. We also develop sampling techniques to simulate random draws from these complex-valued posterior distributions. We test our method on a 2D sparse signal and piecewise constant images with varying signal-to-noise ratio (SNR) to demonstrate both reduced uncertainty and efficiency.

We emphasize that while SAR imaging serves as a useful prototype, our goal is to provide a fundamental framework for recovering the posterior distribution of a complex-valued signal from MMV data. While for ease of presentation we assume we are given continuous uniform Fourier measurements, our approach can be readily adapted to consider modalities leading to other forward operators. The support-informed mask could still be obtained using the methods described in Section 5.2 (specifically Section 5.2.3) via discrete Fourier transform (DFT). Other methods to determine local support, such as the method developed in [6], may also be used. For similar reasons we assume that the magnitude is sparse in the signal or gradi-

²By contrast, the method in [138] applied a binary mask.

ent domain and again note that our method is generalizable to other sparse prior assumptions. Finally we point out that since the forward operators in (5.1) and (5.2) are DFTs, an additional error is incurred in the recovery due to aliasing artifacts. This error is typically expected to be smaller in magnitude than measurement noise. Nevertheless, to avoid any potential inverse crime, in our numerical experiments we simulate highly resolved approximations of the (noisy) *continuous* Fourier measurements.

Paper organization

The rest of this paper is organized as follows. Section 5.2 provides all of the necessary ingredients for our new method, which is proposed in Section 5.3. Numerical experiments are provided in Section 5.4, with some concluding remarks in Section 5.5.

Section 5.2

A Bayesian Approach

Since our study is motivated by SAR imaging, we consider F in (5.1) to be the discrete Fourier transform and \mathbf{z} to be a pixelated SAR image. For ease of presentation we consider only the case $m = n$ in (5.1), although this is not an inherent limitation of our approach. Also, since (5.1) is readily understood for one-dimensional problems, we develop our method for $\mathbf{z} \in \mathbb{C}^n$, and note that higher-dimensional signals can be easily vectorized to fit this form. In this regard, our examples in Section 5.4 are for 2D images. Finally, the terms signal and image are used interchangeably throughout our manuscript.

We follow the framework developed in [138] for recovering posterior distributions of *real*-valued signals from (5.2). The components of the measured data $\hat{\mathbf{y}}^j \in \mathbb{C}^n$,

$j = 1, \dots, J$, in (5.1) are modeled by

$$\hat{\mathbf{y}}_k = (F\mathbf{z})_k + \hat{\boldsymbol{\epsilon}}_k, \quad k = -\frac{n}{2}, \dots, \frac{n}{2} - 1, \quad (5.5)$$

where $F \in \mathbb{C}^{n \times n}$ is the discrete Fourier transform operator, $\mathbf{z} \in \mathbb{C}^n$ is the unknown variable of interest, and the noise $\hat{\boldsymbol{\epsilon}} \in \mathbb{C}^n$ is circularly-symmetric complex Gaussian with covariance matrix $\sigma^2 \mathbb{I}_n$. The corresponding probability density function for $\hat{\boldsymbol{\epsilon}}$ is then

$$f_{\hat{\boldsymbol{\epsilon}}}(\hat{\boldsymbol{\epsilon}}) \propto \exp\left(-\frac{\|\hat{\boldsymbol{\epsilon}}\|_2^2}{\sigma^2}\right). \quad (5.6)$$

Since the magnitude and phase components of the signal can require different prior densities in (5.3), we also formulate (5.3) by decomposing the underlying signal as $\mathbf{z} = \mathbf{g} \odot e^{i\boldsymbol{\Phi}} \in \mathbb{C}^n$, where $\mathbf{g} \in \mathbb{R}_+^n$ is the magnitude, $\boldsymbol{\Phi} \in [-\pi, \pi]^n$ is the phase, and \odot indicates component-wise multiplication. Hence we can rewrite (5.2) as

$$\hat{\mathcal{Y}} = F\mathcal{Z} + \hat{\boldsymbol{\epsilon}} = F(\mathcal{G} \odot e^{i\boldsymbol{\Phi}}) + \hat{\boldsymbol{\epsilon}}, \quad (5.7)$$

where $\mathcal{Z} = \mathcal{G} \odot e^{i\boldsymbol{\Phi}}$, with $\mathcal{G} \in \mathbb{R}_+^n$ and $\boldsymbol{\Phi} \in [-\pi, \pi]^n$ denoting random variables with respective realizations \mathbf{g} and $\boldsymbol{\phi}$. As is standard in coherent imaging problems, we also assume that \mathcal{G} and $\boldsymbol{\Phi}$ are mutually independent, yielding the posterior density function characterized by Bayes' theorem as

$$f_{\mathcal{G}, \boldsymbol{\Phi} | \hat{\mathcal{Y}}}(\mathbf{g}, \boldsymbol{\phi} | \hat{\mathcal{Y}}) \propto f_{\hat{\mathcal{Y}} | \mathcal{G}, \boldsymbol{\Phi}}(\hat{\mathcal{Y}} | \mathbf{g}, \boldsymbol{\phi}) f_{\mathcal{G}, \boldsymbol{\Phi}}(\mathbf{g}, \boldsymbol{\phi}) = f_{\hat{\mathcal{Y}} | \mathcal{G}, \boldsymbol{\Phi}}(\hat{\mathcal{Y}} | \mathbf{g}, \boldsymbol{\phi}) f_{\mathcal{G}}(\mathbf{g}) f_{\boldsymbol{\Phi}}(\boldsymbol{\phi}), \quad (5.8)$$

where once again the likelihood and prior density functions are conditioned on hyper-priors, each with its own respective density function. We now discuss how each term in (5.8) is determined.

5.2.1. The likelihood

The likelihood density function is determined by assumptions regarding the noise, here given by (5.6). Specifically the law of total probability yields the likelihood density function

$$f_{\hat{\mathbf{y}}|\mathbf{z}}(\hat{\mathbf{y}}|\mathbf{z}) \propto \prod_{i=1}^J \exp\left(-\frac{1}{\sigma^2} \|\hat{\mathbf{y}}^{(j)} - F\mathbf{z}\|_2^2\right) \propto \exp\left(-\frac{J}{\sigma^2} \|\bar{\hat{\mathbf{y}}} - F\mathbf{z}\|_2^2\right), \quad (5.9)$$

or equivalently

$$f_{\hat{\mathbf{y}}|\mathcal{G},\Phi}(\hat{\mathbf{y}}|\mathbf{g},\phi) \propto \exp\left(-\frac{J}{\sigma^2} \|\bar{\hat{\mathbf{y}}} - F(\mathbf{g} \odot e^{i\phi})\|_2^2\right), \quad (5.10)$$

if the decomposition in (5.7) is used. Following [138], in our numerical experiments we approximate σ^2 as

$$\sigma^2 = \frac{1}{n(J-1)} \sum_{j=1}^J \|\hat{\mathbf{y}}^j - \bar{\hat{\mathbf{y}}}\|_2^2, \quad (5.11)$$

where again $\bar{\hat{\mathbf{y}}}$ is the mean of $\hat{\mathbf{y}}^1, \dots, \hat{\mathbf{y}}^J$.

5.2.2. The prior density function

The primary contribution of this work is in the construction of prior and hyperprior density functions. In our development we assume that the magnitude of the true image is sparse in the gradient domain, although other sparse domains are easily accommodated. Because the signal itself being sparse is also a common assumption in SAR, we include its corresponding prior construction in Section 5.2.2. Importantly, we seek to recover a posterior density on the entire *complex*-valued signal \mathcal{Z} so that phase information is preserved.

Sparse magnitude prior. Assume $|\mathcal{Z}|$ in (5.2) is sparse. What distinguishes our prior from other standard sparsity-promoting priors, such as those typically used in (empirical) sparse Bayesian learning (SBL) techniques [120, 131], is that rather than learn sparsity-promoting hyper-parameter distributions, here instead we construct a weighted mask $W = \text{diag}(w_1, \dots, w_n) \in \mathbb{R}_+^{n \times n}$ that encodes the relative confidence regarding the support of the magnitude of \mathbf{z} . This mask is formed by taking advantage of the support information in the sparse domain empirically provided by the MMV data, and its construction is discussed in Section 5.2.3.

We adapt the approach from [138] for real-valued signals and define the conditional prior probability density function

$$f_{\mathcal{Z}|\eta}(\mathbf{z}|\eta) = \frac{1}{(\pi\eta^2)^n} \exp\left(-\frac{1}{\eta^2} \|W\mathbf{z}\|_2^2\right), \quad (5.12)$$

where $\eta \in \mathbb{R}^+$ is the prior variance with realization η .

Remark 5.1. We note that using the 2–norm in (5.12) is only reasonable *because* of the mask, that is, since the sparsity is promoted *locally* in regions of smoothness. On the other hand, if W is the identity matrix (i.e. not support informed), we must use $p = 1$ to obtain an accurate MAP estimate (see e.g. [25]), resulting in a more costly computation of the point estimate solution, and without the benefit of localization. More details can be found in [47].

Finally, observe that the prior (5.12) does not require decomposing the signal into its magnitude and phase components. Hence we can use (5.3) directly, which means we can also use (5.9) as the likelihood function.

Sparsity in the magnitude gradient domain. When the signal magnitude is assumed to be piecewise constant, the total variation (TV) operator effectively suppresses variation and noise in smooth regions. While it is well known that the noise

in coherent imaging is due to a multitude of factors, including speckle, successful extraction of important features in SAR imaging using TV has been demonstrated in [35, 107, 110]. Thus we use it here. To this end, we explicitly define $L_{1D} \in \mathbb{R}^{n-1 \times n}$ as the difference operator, which for 1D signals is

$$[L_{1D}]_{i,j} = \begin{cases} 1 & \text{if } i = j - 1, \\ -1 & \text{if } i = j, \\ 0 & \text{else.} \end{cases} \quad (5.13)$$

For a 2D $N \times N$ image³ we extend (5.13) to construct a composite difference operator that computes both vertical and horizontal differences. Specifically, we define $L_{2D} \in \mathbb{R}^{2(N^2-N) \times N^2}$ as the block matrix $L_{2D} = [L_v^T \ L_h^T]^T$, where $L_v, L_h \in \mathbb{R}^{N^2-N \times N^2}$ denote vertical and horizontal difference matrices, respectively. We have

$$L_v = \begin{bmatrix} L_{1D} & 0 & \cdots & 0 \\ 0 & L_{1D} & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & L_{1D} \end{bmatrix}$$

where $L_{1D} \in \mathbb{R}^{N-1 \times N}$ is defined in (5.13), and

$$[L_h]_{i,j} = \begin{cases} 1 & \text{if } i = j + N, \\ -1 & \text{if } i = j, \\ 0 & \text{else.} \end{cases}$$

³For simplicity, we define the 2D difference operator only on square images and note that similar operators can be constructed for non-square images.

For ease of notation we simply denote the difference operator as $L \in \mathbb{R}^{\tilde{n} \times n}$ for $\tilde{n} \in \mathbb{Z}_+$ for both 1D and 2D signals.

The conditional prior probability density for \mathcal{G} is then

$$f_{\mathcal{G}|\eta}(\mathbf{g}|\eta) \propto \frac{1}{(2\eta^2\pi)^{\frac{n}{2}}} \exp\left(-\frac{1}{2\eta^2}\|WL\mathbf{g}\|_2^2\right) \mathbf{1}_{\mathbb{R}_+^n}(\mathbf{g}), \quad (5.14)$$

where $\mathbf{1}_{\mathbb{R}_+^n}$ is the indicator function for positive real vectors. Similar to (5.12), the weighted mask W uses MMV data to encode the relative confidence regarding the support locations in the gradient domain.

Remark 5.2 (Higher order TV). Higher order TV (HOTV) priors have been used effectively to recover SAR images from given MMV phase history data [107, 110], and although beyond the scope of the current investigation, our method can be straightforwardly adapted to use HOTV or other sparsifying transform operators as appropriate. Since TV is still the predominant choice, we use it here even when the underlying image is piecewise-smooth (rather than piecewise-constant). In this way we are also able to test the effectiveness of our method even when the prior assumptions are not met.

Prior for the phase. Coherent imaging systems often have no prior information regarding the phase. It is therefore reasonable to impose the uniform prior as

$$f_{\Phi}(\boldsymbol{\phi}) = \frac{1}{(2\pi)^n} \mathbf{1}_{[-\pi, \pi]^n}(\boldsymbol{\phi}), \quad (5.15)$$

where $\mathbf{1}_{[-\pi, \pi]^n}$ denotes the indicator function for vectors in $[-\pi, \pi]^n$.

Finally, since in this case the prior is decomposed into its magnitude and phase components, we employ likelihood function (5.10) to be used in (5.8).

The hyperprior. There are a variety of ways to choose the hyperprior on η in (5.12) and (5.14). By effective application of the MMV data, we observe that it is both reasonable and simple to adopt the delta distribution $f_\eta(\eta) = \delta(\eta - \hat{\eta})$ and then use the respective point estimate for $\hat{\eta}$. As a preprocessing step, we first compute the solution to the compressed sensing problem

$$\mathbf{z}^{CS} = \arg \min_{\mathbf{z}} \left\| F\mathbf{z} - \bar{\hat{\mathbf{y}}} \right\|_2^2 + \frac{1}{\tilde{\eta}} \|R|\mathbf{z}|\|_1, \quad (5.16)$$

where $R = \mathbb{I}_n$ in the sparse signal case (5.12) and $R = L$ defined by (5.13) in the sparse magnitude gradient case (5.14). Following a similar construction in [108], we calculate the parameter $\tilde{\eta}$ using

$$\tilde{\eta}^2 := \frac{1}{\tilde{n} - 1} \mathbb{E} \left[\|P\mathbf{g}\|_2^2 | \hat{\mathcal{Y}} \right] \approx \frac{1}{\tilde{n} - 1} \left\| P \left| F^H \left(\frac{1}{J} \sum_{j=1}^J \hat{\mathbf{y}}^j \right) \right\|_2 \right\|_2^2. \quad (5.17)$$

The hyperprior approximation $\hat{\eta}$ used in (5.12) and (5.14) is then obtained as

$$\hat{\eta}^2 := \frac{1}{\tilde{n} - 1} \left\| P |\mathbf{z}^{CS}| \right\|_2^2, \quad (5.18)$$

where $P = W$ in the sparse signal case (5.12) and $P = WL$ in the sparse magnitude gradient case (5.14). By first computing the ℓ_1 approximation \mathbf{z}^{CS} , the parameter $\hat{\eta}$ is less sensitive to noise than the one computed in (5.17). It then follows from (5.12) and (5.14) that

$$f_{\mathbf{z}}(\mathbf{z}) \propto \exp \left(-\frac{1}{\hat{\eta}^2} \|W\mathbf{z}\|_2^2 \right), \quad (5.19a)$$

$$f_{\mathbf{g}}(\mathbf{g}) \propto \exp \left(-\frac{1}{2\hat{\eta}^2} \|WL\mathbf{g}\|_2^2 \right) \mathbf{1}_{\mathbb{R}_+^n}(\mathbf{g}). \quad (5.19b)$$

Remark 5.3 (The common kernel condition). It is possible that (5.19b) describes an improper probability density function. We note, however, that when combined with

their respective likelihood functions, the common kernel condition [73]

$$\text{Ker}(F) \cap \text{Ker}(P) = \{\mathbf{0}\}, \quad (5.20)$$

for $P = WL$, is satisfied since F is full rank. This guarantees that the resulting posterior distribution is a proper density function. As a result, we do not need to impose any artificial boundary conditions on the sparsifying transform L .

5.2.3. Weighted mask construction for the prior

We now describe how to construct the empirically-learned weighted mask W used in each of the prior density functions (5.19a) and (5.19b). As already discussed, the mask encodes relative confidence regarding the support locations in the signal magnitude gradient domain, which is expected to be sparse. To this end we note that in [138] a *binary* mask was similarly used, but required thresholding to determine support locations. Importantly, to prevent data loss we will employ the *concentration factor edge detection* method [48, 50] as it can be applied *directly* to the given Fourier data *without* requiring an initial approximation for the magnitude of the underlying signal.

Remark 5.4 (Edge domain vs gradient domain). The concentration factor edge detection method was designed to determine the corresponding *edge function* of a piecewise analytic function [48, 49]. Because here we assume that the underlying signals have piecewise constant magnitude, the magnitude edge and gradient domains are equivalent. Hence the concentration factor method, which is directly applied to the observed Fourier data, is consistent with Lg , where L is defined in (5.13). We note that it was shown in [50] that concentration factors can be specifically designed to be consistent with HOTV operators. In this context (also see Remark 5.2), although beyond the scope of this current investigation, it is straightforward to adapt the weighted matrix W to accommodate other sparse domains as appropriate.

Edge detection from Fourier data. We first review the concentration factor edge detection method for real-valued signals on $[-\pi, \pi]$. In Section 5.2.3 we describe how it can be effectively used when the sparsity corresponds to the magnitude of a complex-valued signal. We also note that the method can be easily modified to any arbitrary one-dimensional finite interval. Finally, the concentration factor method (as given in (5.25)) is similarly extended to 2D as the difference operator (5.13), that is, by using separate constructions in the horizontal and vertical directions.

Let $h(x)$ be a real-valued periodic piecewise-smooth function on $[-\pi, \pi]$ for which we are given Fourier coefficients

$$\hat{h}(k) = \frac{1}{2\pi} \int_{-\pi}^{\pi} h(x) e^{-ikx} dx, \quad k = -\frac{n}{2} + 1, \dots, \frac{n}{2}. \quad (5.21)$$

We seek to approximate the corresponding jump function defined as

$$[h](x) := h(x^+) - h(x^-), \quad (5.22)$$

where $h(x^+)$ and $h(x^-)$ denote right-hand and left-hand side limit of h at x , respectively. The concentration factor edge detection method computes the partial sum (in sequel referred to as either the edge or jump function approximation)

$$S_n^\sigma[h](x) = i \sum_{\substack{k=-\frac{n}{2}+1 \\ k \neq 0}}^{\frac{n}{2}} \hat{h}(k) \operatorname{sgn}(k) \sigma\left(\frac{2|k|}{n}\right) e^{ikx} \approx [h](x). \quad (5.23)$$

The function $\sigma(\nu)$, $\nu \in [0, 1]$, discretized in (5.23) as $\sigma\left(\frac{2|k|}{n}\right)$, is known as the *concentration factor*. It can be described as a band pass filter that amplifies and represses contributions from different bands of Fourier data. Satisfaction of the so-called admissibility conditions [49] guarantees the convergence of (5.23).

If the domain is discretized as $y_j = -\pi + \frac{2\pi j}{n}$, $j = 0, \dots, n$, then we can approximate $[\mathbf{h}] = \{[h](x_\ell)\}_{\ell=1}^n$ by choosing $x = x_\ell = \frac{y_\ell + y_{\ell-1}}{2}$ in (5.23). This means that an edge location is resolved up to the distance $\Delta x = \frac{2\pi}{n}$. Since the inverse problem (5.1) is in the discrete setting, we analogously define the matrix $S^\sigma \in \mathbb{C}^{n \times n}$ with entries

$$S^\sigma(\ell, k) = i \operatorname{sgn}(k) \sigma\left(\frac{2|k|}{n}\right) e^{ikx_\ell}, \quad (5.24)$$

where $\ell = 1, \dots, n$, $k = -\frac{n}{2} + 1, \dots, \frac{n}{2}$, and $S^\sigma(\ell, 0) = 0$ by construction. It follows that

$$S^\sigma \hat{\mathbf{h}} \approx [\mathbf{h}], \quad (5.25)$$

where $\hat{\mathbf{h}} = \{\hat{h}_k\}_{k=-\frac{n}{2}+1}^{\frac{n}{2}}$ in (5.21).

In our investigation we use the admissible trigonometric concentration factor, defined as

$$\sigma(t) = \frac{\pi \sin(\pi t)}{\operatorname{Si}(\pi)}, \quad \operatorname{Si}(\pi) = \int_0^\pi \frac{\sin(x)}{x} dx. \quad (5.26)$$

Remark 5.5. The trigonometric concentration factor (5.26) is particularly well-suited for (real-valued) piecewise constant functions. This is because for equally spaced points $\{x_\mu\}_{\mu=1}^n$, we have [50]

$$S_n^\sigma[h](x_\mu) \approx \frac{2\pi}{n} \sum_{\ell=0}^{n-1} (h(x_{\ell+1}) - h(x_\ell)) \sum_{k=1}^{\frac{n}{2}} \frac{\sigma(2k/n)}{2k\pi/n} \cos k(x_\mu - x_{\ell+1/2}) \approx (L\mathbf{h})_\mu, \quad (5.27)$$

where L is given in (5.13). Other concentration factors are similarly shown to be consistent with HOTV operators.⁴ To this end we note that in environments where some observations are missing or less reliable, such as those considered in [112, 123,

⁴We note that in [50, Section 3], (5.27) was given as an equality, and was based on *discrete* Fourier data, $(F\mathbf{h})_k$, $k = -\frac{n}{2}, \dots, \frac{n}{2}$, where F is the discrete Fourier transform. Similar equivalencies for other concentration factors were also derived. The approximation for *continuous* Fourier data, $\{\hat{h}_k\}_{k=-\frac{n}{2}+1}^{\frac{n}{2}}$, still holds.

132], the concentration factors designed in [123] may provide more accurate and robust recovery. Finally, we observe that by construction any admissible concentration factor can be used for sparse signal recovery. Hence we also use (5.26) for our sparse signal recovery experiments.

2D edge maps. For 2D image $\mathbf{H} \in \mathbb{R}^{N \times M}$ with given Fourier data $\hat{\mathbf{H}} \in \mathbb{C}^{N \times M}$, we adapt the concentration factor method to compute both horizontal and vertical edges, that is, we use a fixed value in one direction and calculate the edges in the opposite direction at that value. Following the jump function definition in (5.22), vertical edges $[\mathbf{H}]_v$ are approximated as

$$S^\sigma \hat{\mathbf{H}} \bar{F} \approx [\mathbf{H}]_v, \quad (5.28)$$

where \bar{F} is the complex conjugate of the $M \times M$ discrete Fourier transform matrix F and S^σ is the $N \times N$ concentration factor matrix with corresponding entries given by (5.24). The horizontal edges $[\mathbf{H}]_h$ are similarly extracted and are then used in tandem with the vertical edges to form a composite edge map similar to L_{2D} following (5.13). We do not explicitly derive this operator since we will see in Section 5.2.3 that for complex-valued images, edge maps are formed for the *squared* magnitude in the sparse domain.

In what follows we adapt (5.23) and (5.28) to determine the support locations in the complex-valued signal magnitude gradient domain, which will in turn be used to construct the weighted mask W in prior density functions (5.19a) and (5.19b).

The concentration factor edge detection method for complex-valued signals. We apply (5.23) on $|\mathbf{z}|^2$ rather than on magnitude \mathbf{g} to remove the added computational complexity of having to extract \mathbf{g} , specifically to avoid having to approxi-

mate the phase ϕ in $\mathbf{z} = \mathbf{g} \odot e^{i\phi}$. To this end, let $f(x) = g(x)e^{i\phi(x)}$, where $g : \mathbb{R} \rightarrow \mathbb{R}_+$ and $\phi : \mathbb{R} \rightarrow [-\pi, \pi)$, for which the first n Fourier coefficients $\{\hat{f}(k)\}_{k=-\frac{n}{2}+1}^{\frac{n}{2}}$ are known (as in (5.21), but for complex-valued function f). If we denote the corresponding Fourier transform operator as \mathcal{F} , i.e. $\mathcal{F}(f(x)) = \hat{f}(k)$, then the convolution theorem yields

$$\mathcal{F}(g^2) = \mathcal{F}(|f|^2) = \mathcal{F}(f^* f) = \mathcal{F}(f^*) * \mathcal{F}(f) = \hat{f} \star \hat{f}, \quad (5.29)$$

where f^* is the complex conjugate of f and $*$ and \star are the convolution and cross-correlation operations, respectively. Therefore when considering the discrete model (5.5), (5.29) suggests employing (5.25) on $\hat{\mathbf{y}}^j \star \hat{\mathbf{y}}^j$, $j = 1, \dots, J$, to recover J approximations of $L\mathbf{g}$. Hence we define $\mathbf{s} \in \mathbb{R}_+^n$ as

$$\mathbf{s}^j := S^\sigma(\hat{\mathbf{y}}^j \star \hat{\mathbf{y}}^j) \approx [(\mathbf{g}^j)^2], \quad j = 1, \dots, J. \quad (5.30)$$

Following (5.28), in the 2D case we recover \mathbf{s}_h^j , the horizontal edge map for the sparse domain magnitude squared, as

$$\mathbf{s}_h^j := S^\sigma(\hat{\mathbf{Y}}^j \star \hat{\mathbf{Y}}^j) \bar{F} \approx [(\mathbf{G}^j)^2], \quad j = 1, \dots, J.$$

Here we have defined $\mathbf{Z} \in \mathbb{C}^{N \times M}$ as the 2D image with magnitude-phase decomposition $\mathbf{Z} = \mathbf{G} \odot e^{i\Phi}$ and corresponding Fourier data $\hat{\mathbf{Y}} \in \mathbb{C}^{N \times M}$ as given by (5.1) with DFT matrix F . The vertical edge map for the sparse domain magnitude squared, \mathbf{s}_v^j , is similarly defined.

Sparse signal prior. When using the sparse signal prior (5.19a), the composite edge map approximation is calculated as

$$\mathbf{s}^j := \max(|\mathbf{s}_h^j|, |\mathbf{s}_v^j|), \quad j = 1, \dots, J,$$

so that the \mathbf{s}_i^j indicates whether an edge exists in *either* the vertical or horizontal direction at pixel i for image j .

Sparse magnitude gradient prior

For the sparse magnitude gradient prior (5.19b), we instead construct each vector \mathbf{s}^j using \mathbf{s}_h^j and \mathbf{s}_v^j such that \mathbf{s}_i^j corresponds to the composite form of $[L\mathbf{g}]_i$, $i = 1, \dots, n$, for L defined in (5.13). The vertical and horizontal edge approximations in \mathbf{s}^j therefore correspond pixel-wise to the vertical and horizontal differences computed using L , i.e. $\mathbf{s}^j = [(s_v^j)^T (s_h^j)^T]^T$ for $j = 1, \dots, J$.

Remark 5.6. For highly-resolved SAR images, the phase is approximately smooth except in locations where the scene height is also discontinuous [66]. In this case, the concentration factor method can be applied *directly* to the complex-valued data to approximate edges in the real and imaginary parts of the image *instead of* in the magnitude squared. A composite edge map is then formed using the magnitude of the complex-valued edge approximation. That is, we compute

$$\mathbf{s}^j = |S^\sigma \hat{\mathbf{y}}^j| \approx |[\mathbf{z}_R^j] + i[\mathbf{z}_I^j]|, \quad j = 1, \dots, J,$$

where the subscripts R and I indicate real and imaginary components, respectively.

Construction of the weighting matrix. As already discussed in Section 5.1, the VBJS method [2, 47, 112] constructs a *spatially-weighted diagonal* regularization matrix, $W = \text{diag}(w_1, \dots, w_n)$, to replace the constant penalty parameter λ in (5.4).

In particular, given multiple measurements of a real-valued signal, W is designed to reflect the relative confidence regarding the support in the sparse domain, with entries determined using the variance of the transformed solution in the sparse domain at each pixel. In [138] the VBJS idea inspired the construction an empirical support-informed prior that used a binary matrix to “mask out” regions of support in the sparse domain so that the sparse prior was only employed in regions that were assumed sparse in the recovery of the posterior distribution. Below we show how VBJS can be adapted to provide the weighting matrix used in the empirical support-informed prior for complex-valued signals given by either (5.19a) or (5.19b).

Our construction of W is analogous to what is described (for real-valued signals) in [2, Section 5]. Its diagonal entries are

$$w_i = \frac{\frac{1}{J} \sum_{j=1}^J |\mathbf{s}_i^j| + \epsilon}{v_i + \epsilon}, \quad i = 1, \dots, n, \quad (5.31)$$

where \mathbf{s}^j are the sparse domain vectors defined in (5.30) and v_i is the corresponding component-wise second moment

$$v_i = \frac{1}{J} \sum_{j=1}^J (\mathbf{s}_i^j)^2, \quad i = 1, \dots, n. \quad (5.32)$$

We choose $\epsilon = 10^{-4}$ to ensure both that $w_i > 0$ and that w_i does not become too large when $v_i \ll 1$. It then follows that $w_i \ll 1$ corresponds to support in the signal magnitude sparse domain at pixel x_i (or $[L\mathbf{g}]_i$ if a sparse magnitude gradient prior is employed), which in turn is used to calibrate the sparse prior via (5.19a) or (5.19b).

Remark 5.7. We use (5.32) in (5.31) instead of the variance, as was originally done in [2], because our numerical experiments indicate that (5.32) scales more weakly with \mathbf{g} than the variance of (5.30) does. Importantly, since the magnitude \mathbf{g} has only positive entries, bias can be potentially introduced in a weight mask that employs

the variance, a problem that is further exacerbated since (5.30) is applied to $|\mathbf{g}|^2$. By contrast, VBJS was used for *real*-valued signals in [2], with the non-zero values in the sparse domain of either sign. We also observe that since $v_i \propto \bar{s}_i^2$ in (5.32), where \bar{s}_i is the average of (5.30) across the J measurements, then $w_i \propto 1/\bar{s}_i$, which is an intuitive choice for determining the relative confidence of the support. In this regard, numerical experiments for recovering point estimates in [2] indicate that directly defining $w_i = 1/(\bar{s}_i + \epsilon)$ is less effective when the sparse domain has non-zero entries that span multiple scales. In particular, the large magnitudes dominate the small ones, and as a consequence, defining the weight in this direct way may be less useful in providing *relative information* in the sparse domain.

Section 5.3

Sampling the Support-Informed Posterior for Complex-Valued Signals

We now have all of the ingredients needed to compute and sample from the posterior distributions using the likelihoods given in (5.9) and (5.10), and the priors provided in (5.19a) and (5.19b).

5.3.1. Sparse magnitude prior

As in Section 5.2.2, we first briefly consider the case where $|\mathcal{Z}|$ is sparse, with likelihood given by (5.9) and support-informed prior by (5.19a). The posterior is then

$$f_{\mathcal{Z}|\hat{\mathbf{y}},\hat{\eta}}(\mathbf{z}|\hat{\mathbf{y}},\hat{\eta}) \propto \exp\left(-\frac{J}{\sigma^2}\|\bar{\hat{\mathbf{y}}}-F\mathbf{z}\|_2^2-\frac{1}{\hat{\eta}^2}\|W\mathbf{z}\|_2^2\right). \quad (5.33)$$

We can directly sample from (5.33) since it defines a complex Gaussian density over \mathbb{C}^n with mean $\boldsymbol{\mu}$ and covariance Γ respectively given by

$$\boldsymbol{\mu} = \Gamma \left(\frac{J}{\sigma^2} F^H \hat{\boldsymbol{y}} \right), \quad \Gamma = \left(\frac{J}{\sigma^2} \mathbb{I}_n + \frac{1}{\hat{\eta}^2} W^T W \right)^{-1}. \quad (5.34)$$

5.3.2. Sparsity in magnitude gradient domain

Following Section 5.2.2, we now assume magnitude sparsity in the gradient domain. In this case we decompose the signal into its magnitude and phase components to accommodate their different prior density functions.

Since our goal is to generate samples from a posterior distribution and provide meaningful results such as mean estimates and credibility intervals, we first want to characterize the hierarchical model in terms of its distributions. This can be directly accomplished by combining (5.10), (5.11), (5.15), (5.18), (5.19b), and (5.31), resulting in

$$\hat{\mathcal{Y}} | \mathcal{G}, \Phi \sim \mathcal{CN} \left(F(\mathbf{g} \odot e^{i\phi}), \frac{\sigma^2}{J} \mathbb{I}_n \right) \quad (5.35a)$$

$$\mathcal{G} | \eta \sim \mathcal{N}_+ \left(0, \eta^2 ((WL)^T (WL))^{-1} \right) \quad (5.35b)$$

$$\Phi \sim \mathcal{U}[-\pi, \pi]^n \quad (5.35c)$$

$$\eta \sim \delta(\eta - \hat{\eta}). \quad (5.35d)$$

Here \mathcal{N}_+ indicates the nonnegative real-valued multivariate Gaussian distribution. The probability density for the posterior distribution $\mathcal{G}, \Phi | \hat{\mathcal{Y}}, \eta$ is then

$$f_{\mathcal{G}, \Phi | \hat{\mathcal{Y}}, \eta}(\mathbf{g}, \phi | \hat{\boldsymbol{y}}, \hat{\eta}) \propto \exp \left(-\frac{J}{\sigma^2} \|\hat{\boldsymbol{y}} - F(\mathbf{g} \odot e^{i\phi})\|_2^2 - \frac{1}{2\hat{\eta}^2} \|P\mathbf{g}\|_2^2 \right) \mathbf{1}_{\mathbb{R}_+^n}(\mathbf{g}) \mathbf{1}_{[-\pi, \pi]^n}(\phi), \quad (5.36)$$

where for ease of notation, we have replaced WL with P .

Upon inspection of (5.36), it is apparent that the corresponding posterior distribution is not Gaussian. Indeed we are not aware of any studied probability distribution whose density has the form of (5.36), effectively prohibiting us from directly sampling the posterior in a computationally efficient manner. We thus turn to the class of Markov Chain Monte Carlo (MCMC) techniques to generate its approximate samples [18, 80]. We choose to use Gibbs sampling [29, 73] since it has the benefit of being efficient in high dimensions when large groups (blocks) of random variables can be updated in parallel.

Sampling the magnitude

When (5.36) is conditioned on Φ , the resulting distribution $\mathcal{G}|\hat{\mathcal{Y}}, \Phi, \eta$ has probability density

$$f_{\mathcal{G}|\hat{\mathcal{Y}}, \Phi, \eta}(\mathbf{g}|\hat{\mathbf{y}}, \boldsymbol{\phi}, \hat{\eta}) \propto \exp\left(-\frac{J}{\sigma^2}\|\bar{\hat{\mathbf{y}}} - F_1\mathbf{g}\|_2^2 - \frac{1}{2\hat{\eta}^2}\|P\mathbf{g}\|_2^2\right) \mathbf{1}_{\mathbb{R}_+^n}(\mathbf{g}), \quad (5.37)$$

where $F_1 = FD(e^{i\boldsymbol{\phi}})$ and $D(e^{i\boldsymbol{\phi}}) = \text{diag}(e^{i\phi_1}, \dots, e^{i\phi_n})$. Notice that (5.37) is the probability density of a multivariate Gaussian distribution truncated to \mathbb{R}_+^n , and as such sampling from its corresponding distribution can be computationally prohibitive. To mitigate this expense we instead use the untruncated version of (5.37)

$$\tilde{f}_{\mathcal{G}|\hat{\mathcal{Y}}, \Phi, \eta}(\mathbf{g}|\hat{\mathbf{y}}, \boldsymbol{\phi}, \eta) \propto \exp\left(-\frac{J}{\sigma^2}\|\bar{\hat{\mathbf{y}}} - F_1\mathbf{g}\|_2^2 - \frac{1}{2\hat{\eta}^2}\|P\mathbf{g}\|_2^2\right). \quad (5.38)$$

Observe that (5.38) corresponds to multivariate Gaussian distribution over \mathbb{R}^n with center and covariance given by

$$\bar{\mathbf{g}} = \Gamma \left(\frac{2J}{\sigma^2} \mathbf{q} \odot \cos(\boldsymbol{\varphi}) \right), \quad \Gamma = \left(\frac{1}{\hat{\eta}^2} P^T P + \frac{2J}{\sigma^2} \mathbb{I}_n \right)^{-1}, \quad (5.39)$$

where $\mathbf{q} \in \mathbb{R}_+^n$ and $\boldsymbol{\varphi} \in [-\pi, \pi)^n$ such that $\mathbf{q} \odot e^{i\boldsymbol{\varphi}} = F_1^H \bar{\hat{\mathbf{y}}}$. Importantly, we are now able to sample directly from the distribution characterized by the (modified) density in (5.38), from which samples with negative elements are rejected to remain consistent with (5.37).

Remark 5.8. When the mode $\bar{\mathbf{g}}$ in (5.39) contains mostly nonnegative entries with few negative ones, we instead implement the *rejection sampling from the mode* technique to more efficiently generate samples for the distribution corresponding to $f_{\mathcal{G}|\hat{\mathbf{y}},\Phi}$ in (5.37) [82]. Methods such as the exact Hamiltonian Monte Carlo method [96] can also be used to generate distribution samples corresponding to (5.37) when (5.38) has little mass in the region where $\mathbf{g} \geq 0$, but may become cost-prohibitive for high-dimensional problems.

Sampling the phase: the von Mises distribution

Conditioning (5.36) on \mathcal{G} , the density of the conditional distribution $\Phi|\hat{\mathbf{y}}, \mathcal{G}$ is

$$f_{\Phi|\hat{\mathbf{y}},\mathcal{G}}(\boldsymbol{\phi}|\hat{\mathbf{y}}, \mathbf{g}) \propto \exp\left(-\frac{J}{\sigma^2} \|\bar{\hat{\mathbf{y}}} - F_2 e^{i\boldsymbol{\phi}}\|_2^2\right) \mathbf{1}_{[-\pi,\pi)^n}(\boldsymbol{\phi}), \quad (5.40)$$

where $F_2 = FD(\mathbf{g})$ and $D(\mathbf{g}) = \text{diag}(g_1, \dots, g_n)$. In [57] it was demonstrated that (5.40) is equivalent to

$$f_{\Phi|\hat{\mathbf{y}},\mathcal{G}}(\boldsymbol{\phi}|\hat{\mathbf{y}}, \mathbf{g}) = \prod_{i=1}^n f_{vM}\left(\phi_i \middle| \varphi_i, \frac{2J}{\sigma^2} q_i\right). \quad (5.41)$$

Here $f_{vM}(x|\boldsymbol{\mu}, \kappa)$ is the probability density of a von Mises distribution with location $\boldsymbol{\mu}$ and concentration κ , and we have defined $F_2^H \bar{\hat{\mathbf{y}}} = \mathbf{q} \odot e^{i\boldsymbol{\varphi}}$, for $\mathbf{q} \in \mathbb{R}_+^n$ and $\boldsymbol{\varphi} \in [-\pi, \pi)^n$. We also observe that that the concentration of the distribution described by (5.41) is proportional to the magnitude \mathbf{g} through \mathbf{q} , while the location is independent

of \mathbf{g} . Furthermore, since

$$f_{\Phi|\hat{\mathcal{Y}},\mathcal{G}}(\boldsymbol{\phi}|\hat{\mathbf{y}},\mathbf{g}) = f_{\Phi_1|\hat{\mathcal{Y}},\mathcal{G}}(\phi_1|y,\mathbf{g}) \cdots f_{\Phi_n|\hat{\mathcal{Y}},\mathcal{G}}(\phi_n|y,\mathbf{g}),$$

then each Φ_i is conditionally independent from all other Φ_{-i} , $i = 1, \dots, n$. Hence we can update each ϕ_i independently of ϕ_{-i} , thereby increasing the sampling efficiency.

To generate samples from the von Mises density, we use the density function of the wrapped Cauchy distribution given as

$$f_{WC}(\theta|\mu,\gamma) = \sum_{k=-\infty}^{\infty} \frac{\gamma}{\pi(\gamma^2 + (\theta - \mu + 2\pi k)^2)}, \quad -\pi < \theta < \pi, \quad (5.42)$$

where μ is the mode of the unwrapped density and γ is the scale factor. We then sample each Φ_i , $i = 1, \dots, n$, using an acceptance-rejection method for which the envelope is given by (5.42). Algorithm 7 [13] describes the implementation process.

Algorithm 7 Sampling the von Mises distribution using the Wrapped Cauchy

Input Location μ and concentration κ of von Mises distribution as defined in (5.41).

Output Sample θ .

1. Set $\tau = 1 + (1 + 4\kappa^2)^{\frac{1}{2}}$, $\rho = (\tau - (2\tau)^{\frac{1}{2}})/(2\kappa)$, and $r = (1 + \rho^2)/(2\rho)$.
 2. Generate $u_1 \sim U(0, 1)$, then set $z = \cos(\pi u_1)$, $f = (1 + rz)/(r + z)$, $c = \kappa(r - f)$.
 3. Generate $u_2 \sim U(0, 1)$, then if $c(2 - c) - u_2 > 0$, go to step 5.
 4. If $\ln(c/u_2) + 1 - c < 0$, return to step 2.
 5. Generate $u_3 \sim U(0, 1)$, then set $\theta = [\text{sgn}(u_3 - 0.5)] \cos^{-1}(f)$.
-

A block-Gibbs sampling approach

We will use a block-Gibbs approach to generate approximate samples from the joint distribution $\mathcal{G}, \Phi|\hat{\mathcal{Y}}, \eta$ corresponding to probability density function (5.36). This is accomplished by drawing exact samples from the conditional distributions $\mathcal{G}|\hat{\mathcal{Y}}, \Phi, \eta$ and $\Phi|\hat{\mathcal{Y}}, \mathcal{G}$, with respective probability density functions given by (5.37) and (5.40).

We begin by using the measurements $\hat{\mathbf{y}}^j$, $j = 1, \dots, J$, to construct the weighted

mask W in (5.31), likelihood variance σ^2/J in (5.11), and the hyperprior point estimate $\hat{\eta}^2$ in (5.18). The chain is initialized using the sample mean $\bar{\mathbf{y}}$. To implement the two-stage Gibbs sampler, we first update \mathbf{g} and then ϕ . The process is repeated for some predetermined number of iterations N_M . The output chain consists of all samples generated after prescribed burn-in period B . Algorithm 8 summarizes the procedure.

Algorithm 8 Two-Step Gibbs Sampling Method for Sparsity in Magnitude Gradient Domain

Input Fourier MMV data $\hat{\mathbf{y}}^j$, $j = 1, \dots, J$, chain length N_M , and burn-in length B .

Output samples $\mathbf{z}^{(\ell-B+1)} = \mathbf{g}^{(\ell)} \odot \exp(i\phi^{(\ell)})$ for $\ell = B, \dots, N_M$.

1. Construct empirical mask W (5.31) using the concentration factor method (5.30).
 2. Estimate σ^2 and $\hat{\eta}^2$ from (5.11) and (5.18), respectively.
 3. Set $\bar{\mathbf{y}} = \frac{1}{J} \sum_{j=1}^J \hat{\mathbf{y}}^j$, $\mathbf{g}^{(0)} = |F^H \bar{\mathbf{y}}|$, and $\phi^{(0)} = \arg(F^H \bar{\mathbf{y}})$.
 4. **For** $k = 1, \dots, N_M$ **do**
 - (a). Draw a sample \mathbf{g}^* from $\tilde{f}_{\mathcal{G}|\hat{\mathbf{y}},\Phi}(\mathbf{g}|\bar{\mathbf{y}}, \phi^{(k-1)})$ in (5.38).
 - (b). If \mathbf{g}^* contains a negative element, return to (a). Otherwise, set $\mathbf{g}^{(k)} = \mathbf{g}^*$.
 - (c). For each $i = 1, \dots, n$, draw a sample ϕ_i^* from $f_{\Phi_i|\hat{\mathbf{y}},\mathcal{G}}(\phi_i|\bar{\mathbf{y}}, \mathbf{g}^{(k)})$ using (5.40) with Algorithm 7. Set $\phi_i^{(k)} = \phi_i^*$.
-

Remark 5.9. The weighted mask W can also be generated using the iterative alternating sequential (IAS) algorithm, which computes a MAP estimate for posterior densities similar to (5.33) and (5.36), but instead with a hyperprior placed on each element of W [19, 22].⁵ A byproduct of IAS is a diagonal matrix of variances, or weights, that promote sparsity in the solution. There are two important distinctions between the IAS approach and the method introduced here. First, IAS requires the repeated solution of large linear systems to jointly compute both the solution and the mask, while our construction of the weighted mask W using (5.30) is a one-time

⁵Although to our knowledge IAS has not been adapted for complex-valued signals, the MAP estimate for the phase is the same as the MLE when using a unitary forward operator. Thus the phase can first be extracted as a preprocessing step followed by use of the IAS algorithm for real-valued signal recovery.

direct calculation that requires only element-wise multiplication in tandem with the FFT algorithm, after which the solution is computed with a single linear system solve. Second, it has been shown that sparsity-promoting priors formed using MMV information can outperform IAS and other sparsity-promoting methods that do not take advantage of joint information [52].

Section 5.4

Numerical Results

We now provide some computational experiments to demonstrate the efficacy of our new method. In each experiment using synthetic data we generate $J = 8$ multiple “continuous” Fourier data measurements, which we simulate using a highly resolved DFT. We then add centered complex-valued Gaussian white noise to each of the generated data sets, and finally downsample the result to obtain the data modeled by (5.1).⁶ The signal-to-noise ratio (SNR) is computed as

$$\text{SNR} = 10 \log_{10} \left(\frac{\|F \mathbf{z}^{exact}\|_2^2}{m\sigma^2} \right), \quad (5.43)$$

where \mathbf{z}^{exact} is the ground truth solution and SNR is given in decibels (dB). We compare our results with those obtained using the classical CS method in (5.4) with $p = 1$, which we rewrite as follows:

⁶Specifically, we use a 900×900 complex-valued image to generate the “continuous” Fourier measurements and then downsample by a factor of 3.

- (a) **sparse signal case:** We split $\mathbf{z} = \mathbf{a} + i\mathbf{b}$ into its real and imaginary parts to obtain the objective function

$$\begin{bmatrix} \mathbf{a}^{CS} \\ \mathbf{b}^{CS} \end{bmatrix} = \arg \min_{\mathbf{a}, \mathbf{b}} \left(\frac{1}{2} \left\| \begin{bmatrix} F_R & -F_I \\ F_I & F_R \end{bmatrix} \begin{bmatrix} \mathbf{a} \\ \mathbf{b} \end{bmatrix} - \begin{bmatrix} \bar{\mathbf{y}}_R \\ \bar{\mathbf{y}}_I \end{bmatrix} \right\|_2^2 + \lambda \sum_{j=1}^n \sqrt{a_j + b_j} \right). \quad (5.44)$$

Here F_R and F_I respectively refer to the real and imaginary parts of F , and $\bar{\mathbf{y}}_R$ and $\bar{\mathbf{y}}_I$ respectively refer to the real and imaginary part of $\bar{\mathbf{y}} = \frac{1}{J} \sum_{j=1}^J \hat{\mathbf{y}}^j$ in (5.1).

- (b) **sparse magnitude gradient case:** Since $|\mathbf{z}|$ is not differentiable, in order to solve (5.4) we follow what was done in [36, 94, 107], which involves first estimating the phase angles of \mathbf{z} and then building a diagonal unitary matrix Θ , so that $\Theta^* \mathbf{z} \approx |\mathbf{z}|$. Here $\Theta_{j,j} = \text{angle}(\tilde{z}_j)$ and $\tilde{\mathbf{z}}$ is some initial approximate solution, e.g. the MLE solution obtained directly from (5.9). The objective function in (5.4) then becomes

$$\mathbf{z}^{CS} = \arg \min_{\mathbf{z}} \left(\frac{1}{2} \|F\mathbf{z} - \bar{\mathbf{y}}\|_2^2 + \lambda \|L\Theta^* \mathbf{z}\|_1 \right). \quad (5.45)$$

We use the alternating direction method of multipliers (ADMM) [17] to solve (5.44) and (5.45), and in both cases we choose $\lambda = \alpha\sigma^2/J\tilde{\eta}$, where $\tilde{\eta}$ is provided in (5.17). In some sense $\alpha = 1$ represents the “best case” scenario for selecting suitable parameters, while other values of α allow us to test for robustness.

We also compare our results to those obtained using a similar deterministic ap-

proach adapted from the VBJS method [110]. We use either the objective function

$$\begin{bmatrix} \mathbf{a}^{VBJS} \\ \mathbf{b}^{VBJS} \end{bmatrix} = \arg \min_{\mathbf{a}, \mathbf{b}} \left(\frac{1}{2} \left\| \begin{bmatrix} F_R & -F_I \\ F_I & F_R \end{bmatrix} \begin{bmatrix} \mathbf{a} \\ \mathbf{b} \end{bmatrix} - \begin{bmatrix} \bar{\mathbf{y}}_R \\ \bar{\mathbf{y}}_I \end{bmatrix} \right\|_2^2 + \lambda \sum_{j=1}^n w_j \sqrt{a_j + b_j} \right) \quad (5.46)$$

in the sparse signal case or

$$\mathbf{z}^{VBJS} = \arg \min_{\mathbf{z}} \left(\frac{1}{2} \|F\mathbf{z} - \bar{\mathbf{y}}\|_2^2 + \lambda \|WL\Theta^* \mathbf{z}\|_1 \right), \quad (5.47)$$

in the sparse magnitude edge case, where W is the same as in (5.31), Θ is the same as in (5.45), and either $L = \mathbb{I}_n$ in the sparse signal case or L is the difference operator (5.13) in the sparse magnitude edge case. In this case we choose $\lambda = \alpha\sigma^2/J\hat{\eta}$, where $\hat{\eta}$ is provided in (5.18).⁷ We further observe that the ℓ_1 -norm is used for the regularization term in (5.46) and (5.47), and note that in our experiments we employ the ADMM algorithm to determine \mathbf{z}^{VBJS} . By contrast, our approach implicitly uses an ℓ_2 -norm regularization term through the Gaussian prior. To this end we note that the VBJS method enables either the ℓ_1 - or ℓ_2 -norm to be used in the regularization term [47]. The use of the ℓ_2 -norm in VBJS results in a generalized Tikhonov problem. Moreover, the solution using the ℓ_2 -norm is identical to the MAP estimate of our method when using the sparse signal assumption and nearly the same as the MAP estimate of our approach when promoting sparsity in the magnitude gradient domain. In the latter case, the primary difference between VBJS using an ℓ_2 -norm regularization term and the MAP estimate of our approach is that VBJS uses a precomputed MLE estimate of the signal phase while our method learns the phase through Algorithm 8.

⁷In the original VBJS method, W is calculated using the variance and $\lambda = 1$. We define $\lambda = \alpha\sigma^2/J\hat{\eta}$ to better reflect the weighting between the data fidelity and regularization terms and to be consistent with the approach developed in this investigation.

Sparse signal

We first test our method on a sparse complex-valued signal where $\mathbf{g}_i = 0$ or $\mathbf{g}_i = 1$, $i = 1, \dots, n$, with SNR = -20 dB. Small regions of pixel size 3×3 (prior to downsampling) containing “on” pixels are randomly distributed to form the ground truth image, where each “on” region has approximately smooth phase. Since (5.33) is a multivariate complex-valued Gaussian density, we immediately have the mean $\boldsymbol{\mu}$ and diagonal covariance Γ through (5.34).

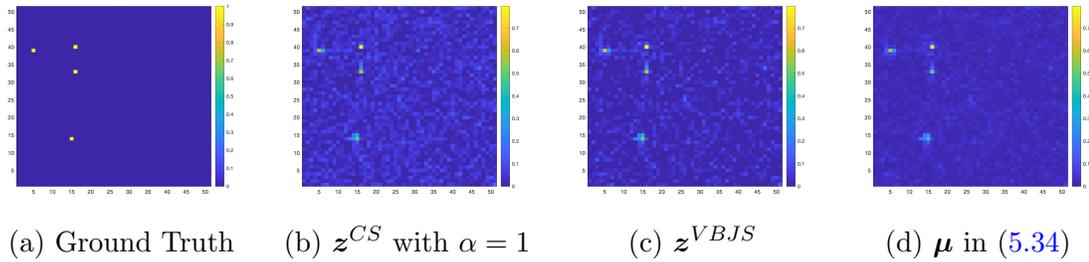


Figure 5.1: Sparse signal magnitude. (a) Ground truth; (b) \mathbf{z}^{CS} with $\alpha = 1$; (c) \mathbf{z}^{VBJs} computed via (5.46); and (d) Mean $\boldsymbol{\mu}$ in (5.34). Here SNR = -20 dB.

Figure 5.1 displays point estimate solutions of the magnitude. For better visualization we show only a small region of the image. Observe that our method is able to suppress noise while also while maintaining image fidelity, and is similar to \mathbf{z}^{CS} computed by (5.44) with $\alpha = 1$ and \mathbf{z}^{VBJs} computed using (5.46).

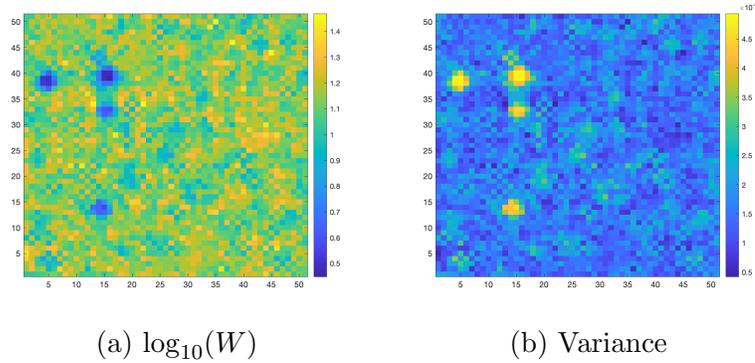


Figure 5.2: Sparse signal recovery. Here SNR = -20 dB.

Figure 5.2 shows the weighted mask W and the pointwise marginal variance ob-

tained through (5.34). As an expected consequence of employing W in (5.31), we observe larger variance in regions of support. In this regard we also note that for this experiment $\frac{\sigma^2}{J} \approx 5.5 \times 10^{-3}$ for σ^2 computed via (5.11). That is, our method indeed decreases the overall recovery uncertainty, especially in sparse regions.

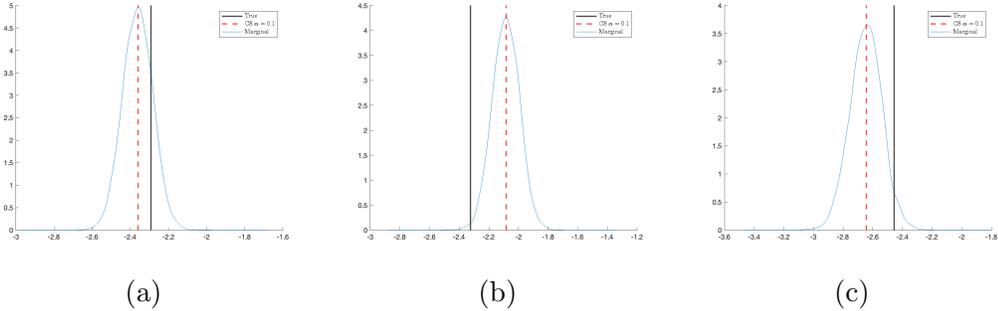


Figure 5.3: Sparse signal phase at three randomly chosen “on” pixels. Here SNR = -20 dB.

Figure 5.3 displays the phase results for three randomly selected “on” pixels, that is, for $\mathbf{g}_i = 1$. The marginal densities for each phase variable are generated using a kernel density estimation technique [16]. Observe that the phase of \mathbf{z}^{CS} computed by (5.44) is near the mode of the computed marginal density computed in all three cases, while the true phase value is located in a region of significant probability density. Since the phase results for \mathbf{z}^{VBJS} are nearly identical to \mathbf{z}^{CS} in our experiments, we do not include them here.

Sparse magnitude gradient

In this case we generate $N_M = 5000$ total samples using Algorithm 8 with a burn-in period of $B = 200$. We use a smooth, highly varying phase to ensure an accurate approximation of the continuous Fourier measurements from discrete data (to avoid the inverse crime). Figure 5.4 displays the (downsampled) magnitude and phase components along with the true vertical edges (5.28) in the magnitude.

Remark 5.10. We emphasize that Algorithm 8 is not changed when given multiple

observations of complex-valued *discretized* image data. Indeed using the DFT in this case yields less model discrepancy, since the error in (5.1) is strictly due to noise. The concentration factor edge detection method in Subsection 5.2.3 is also calculated in exactly the same way. Importantly, the concentration factor edge detection method filters the noisy Fourier data and enhances the resulting edge maps [123], which is critical at low SNR values. We demonstrate this at the end of this section where a complex-valued SAR image is used to generate MMV data.

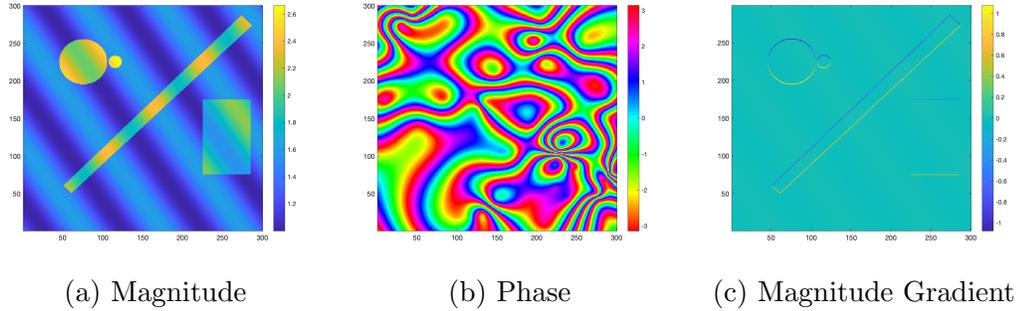


Figure 5.4: Ground truth of (a) magnitude; (b) phase; and (c) magnitude gradient (vertical components).

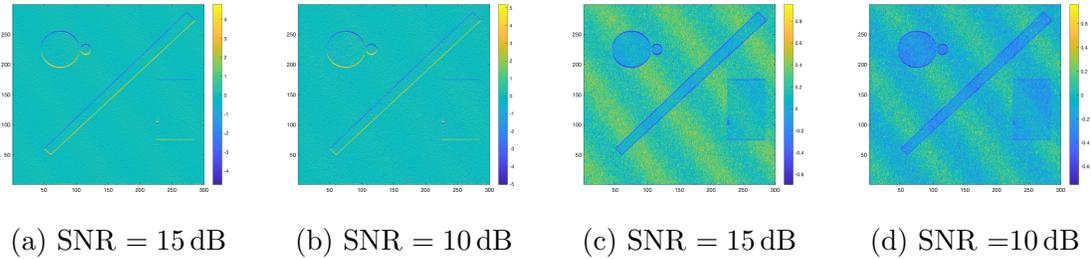


Figure 5.5: (a)-(b) Average of (vertical) magnitude gradient, $\frac{1}{J} \sum_{j=1}^J \mathbf{s}^j$, computed via (5.30); (c)-(d) vertical components of $\log_{10}(W)$.

Figures 5.5 – 5.9 illustrate the efficacy of our method for SNR values of 15 dB and 10 dB, and its ability to reduce the uncertainty in the magnitude recovery. First, Figure 5.5 displays the (vertical component) average of (5.30), $\frac{1}{J} \sum_{j=1}^J \mathbf{s}_v^j$, demonstrating the accurate approximation for the magnitude gradient sparsity needed to construct the components of W in (5.31). The vertical component of the mask W

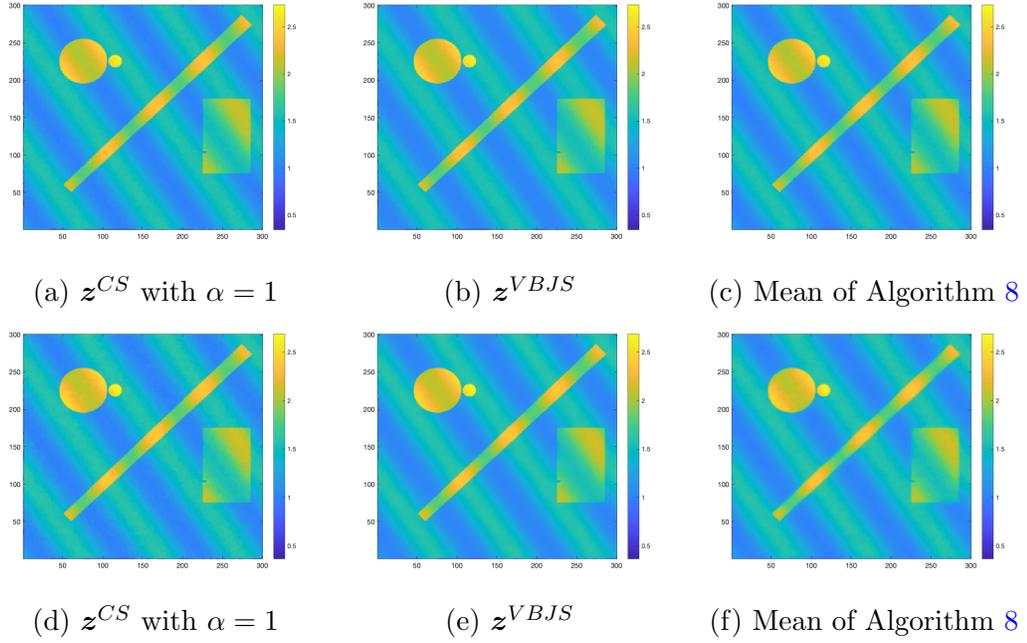
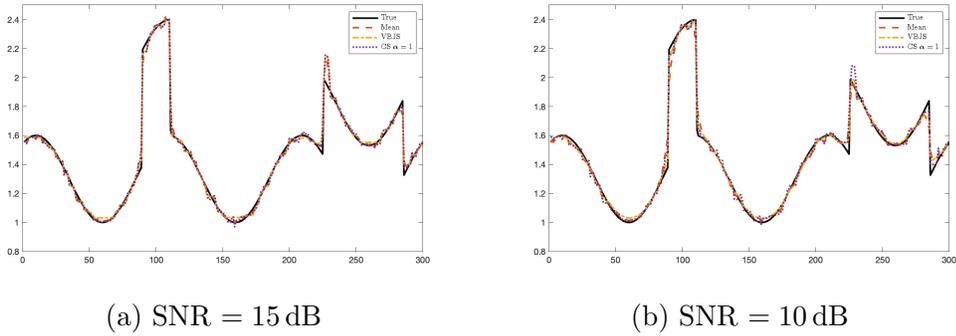


Figure 5.6: Magnitude recovery: (top) SNR = 15 dB; (bottom) SNR = 10 dB.

Figure 5.7: Cross-sections of the magnitude of the true image (black), mean of Algorithm 8 (red), z^{VBJS} (yellow), and z^{CS} with $\alpha = 1$ (purple).

is also provided to demonstrate the impact of noise on determining gradient domain support. As is evident in Figure 5.6 and Figure 5.7, our algorithm yields comparable solutions to z^{CS} calculated via (5.45) with $\alpha = 1$ and z^{VBJS} computed using (5.47) in magnitude recovery. Figure 5.8 compares their corresponding phase estimates. As expected there is little difference in the recovered phase estimates, since no joint phase information is used in Algorithm 8.

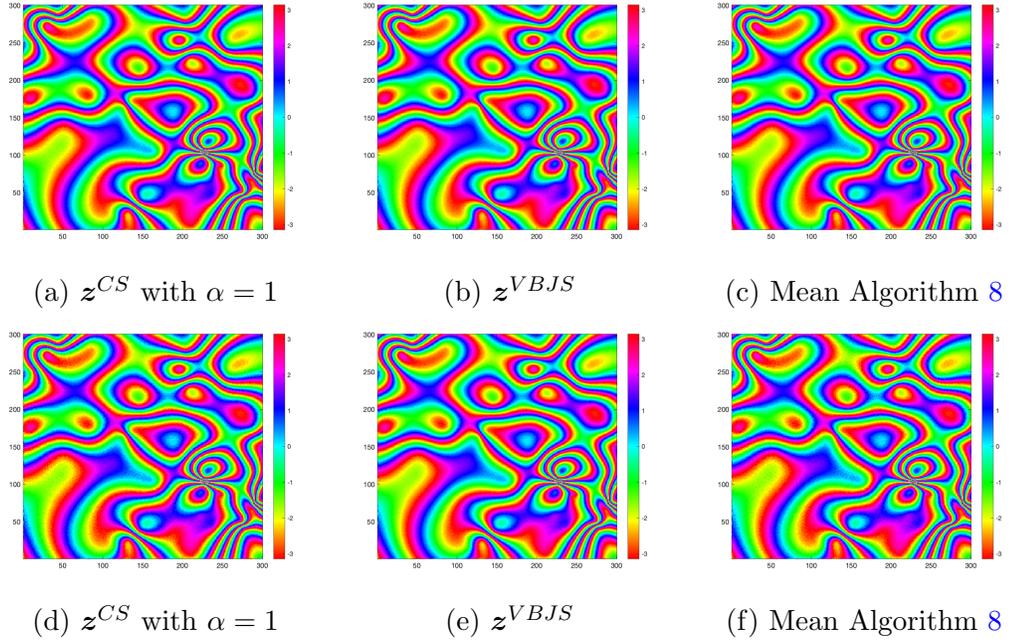


Figure 5.8: Phase recovery: (top) SNR = 15 dB; (bottom) SNR = 10 dB.

Figure 5.9 provides a more illuminating comparison. Here we show the pointwise error in the magnitude along with the pointwise variances of the magnitude and phase obtained via Algorithm 8 for SNR = 10 dB. While the overall point estimate solutions are comparable, it is also clear that the weighted mask works as expected. That is, the error is smaller in the true sparse gradient regions. The pointwise magnitude variance, in addition to being larger near the magnitude edges as one would expect, also scales proportionally with the magnitude (see Remark 5.7). On the other hand, the pointwise phase variance scales inversely proportional to the magnitude. This effect is explained by the concentration of the conditional von Mises density in (5.41) being proportional to the magnitude. The results corresponding to SNR = 15 dB demonstrate similar behavior (not shown).

Noise study

We now examine robustness of our method with respect to noise, using the previous example on a range of SNR values from 0 dB to 30 dB. Once again we generate

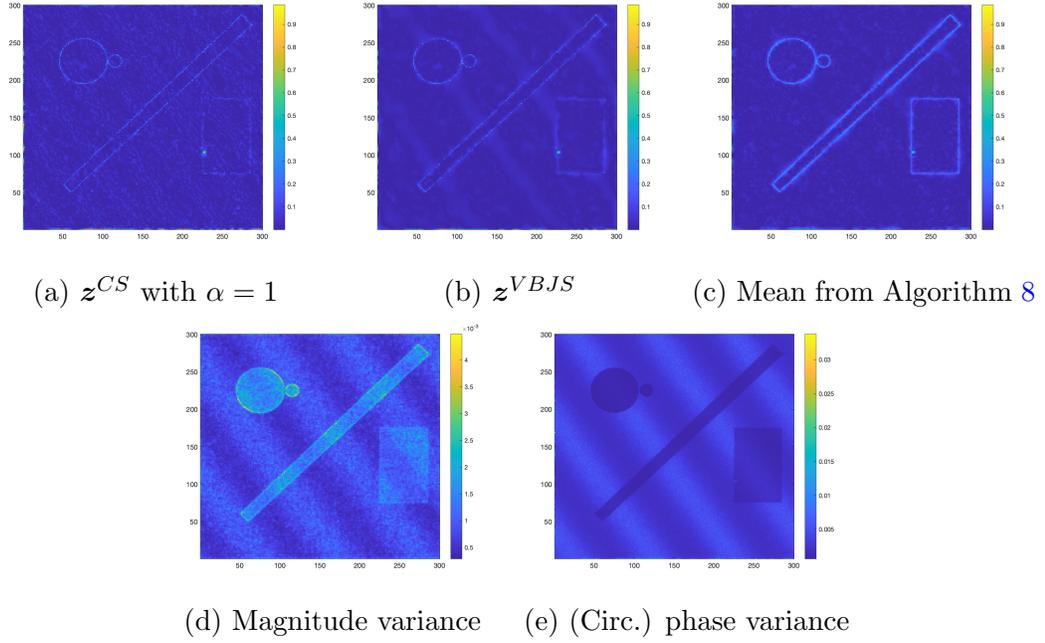


Figure 5.9: (top) Magnitude pointwise error in (a) z^{CS} with $\alpha = 1$ computed via (5.45), (b) mean from Algorithm 8, (c) z^{VBJS} computed via (5.47); (bottom) Pointwise variance computed from Algorithm 8 of (c) magnitude and (d) phase (in logarithmic scale). Here SNR = 10 dB.

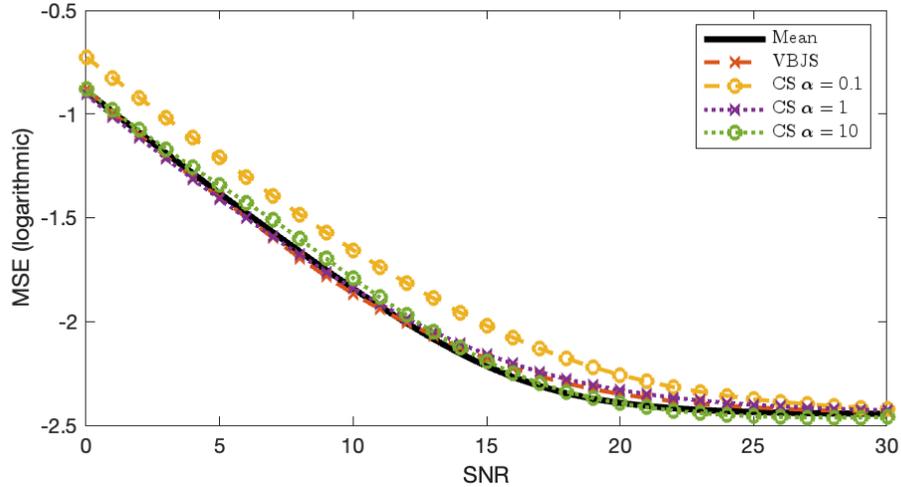


Figure 5.10: Mean squared error of z^{CS} in (5.45) for $\alpha = 0.1, 1, 10$ and the mean of the samples generated using Algorithm 8 plotted against SNR values. The vertical axis is in logarithmic scale.

$N_M = 5000$ samples with a burn-in period of $B = 200$. Figure 5.10 demonstrates that our new method yields solutions that are comparable to \mathbf{z}^{VBJS} and to \mathbf{z}^{CS} when $\alpha = 1, 10$. Our method outperforms (5.45) for $\alpha = 0.1$ at low SNR. Importantly, while the CS solutions require tuning of the α parameter, with the optimal value of α changing at different SNR, neither the VBJS nor the empirical Bayes method require any parameter tuning.

Remark 5.11. We note that Algorithm 8 cannot efficiently generate samples of the magnitude for $\text{SNR} < 0$ dB, (see Remark 5.8). It is not clear that SNR is the sole contributing factor to this failure, however. Indeed it could be that the sampling algorithm would not encounter such difficulties for signals with overall larger signal strength (not near zero). This will be the subject of future investigations.

Real-world example

As a final example, we consider SAR single-look complex (SLC) data from [1] of part of a traffic circle in Paris, France. Figure 5.11 shows the reflectivity of the full SLC image as well as the portion of the image used to generate data for our experiments. Specifically, we generate $J = 8$ measurement vectors following (5.1) where F is the 2D discrete Fourier transform matrix and complex-valued Gaussian white noise is added to the Fourier data. The edges are computed via the concentration factor method (5.25) from the formed Fourier measurements in the real and imaginary components of the image (see Remark 5.6). We display each reflectivity image on a dB scale according to

$$20 \log_{10} \left(\frac{|z|}{\max |z|} \right)$$

with a minimum of -60 dB and maximum of 0 dB. Lesser or greater values are assigned the minimum or maximum accordingly.

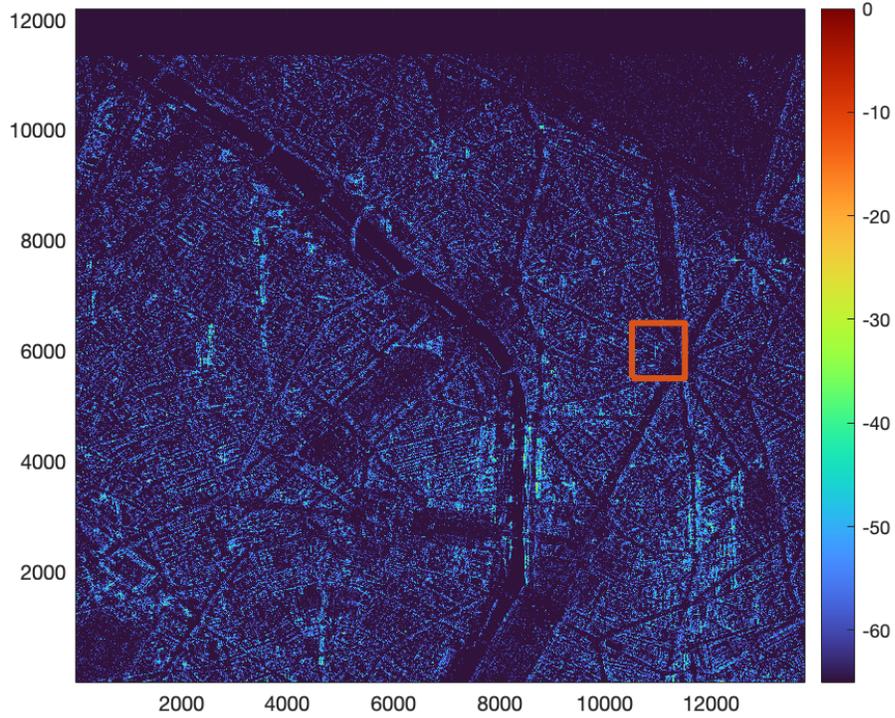


Figure 5.11: Reflectivity of full SLC image. Scale is -65 dB to 0 dB.

Figure 5.12 shows the results obtained from (5.34) and (5.44) using the sparse magnitude assumption with $\text{SNR} = 0$ dB. In addition to recovering pointwise variance estimates, our new method yields a solution that better suppresses the noise outside of the signal support when compared with the z^{CS} solution. This is particularly apparent when considering the region around the traffic circle. Figure 5.13 displays the phase variance estimated using 1000 samples of (5.33). As expected, the phase variance is inversely proportional to the image magnitude and decreases as SNR increases. We also observe that the marginal density of the phase (calculated using a circular kernel density estimation technique [90]) concentrates closer to the true value with increasing SNR. Moreover, when either z^{CS} or z^{VBJs} contains a 0 entry, as is common in ℓ_1 -regularized solutions, the phase information for the corresponding

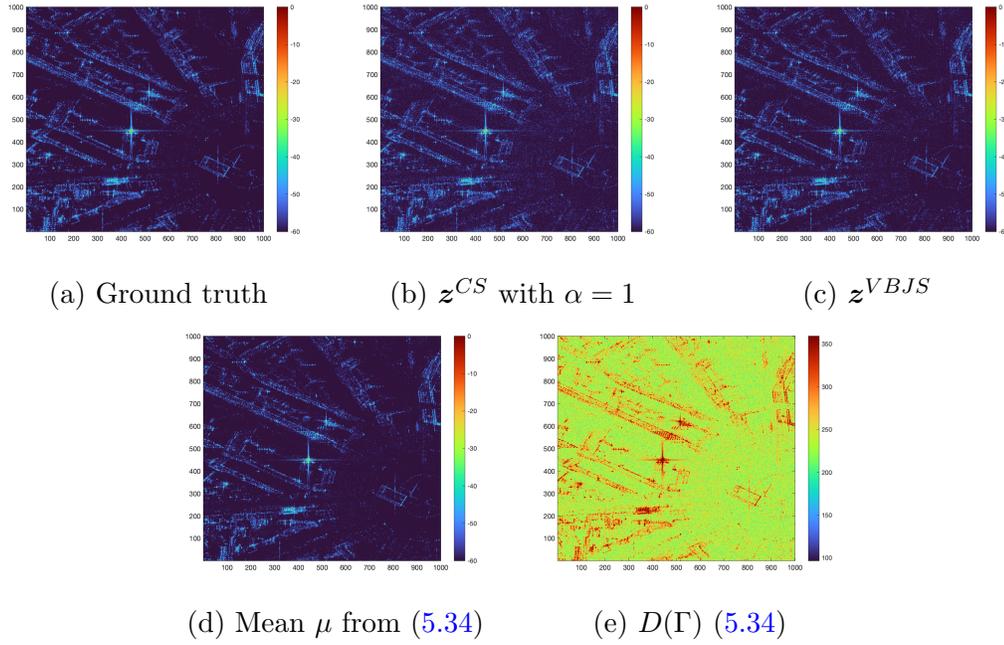


Figure 5.12: (a) Ground truth reflectivity; (b) z^{CS} with $\alpha = 1$ computed via (5.44); (c) z^{VBJS} computed via (5.46); (d) mean μ computed using (5.34); and (e) pointwise variance computed using (5.34). Here SNR = 0 dB.

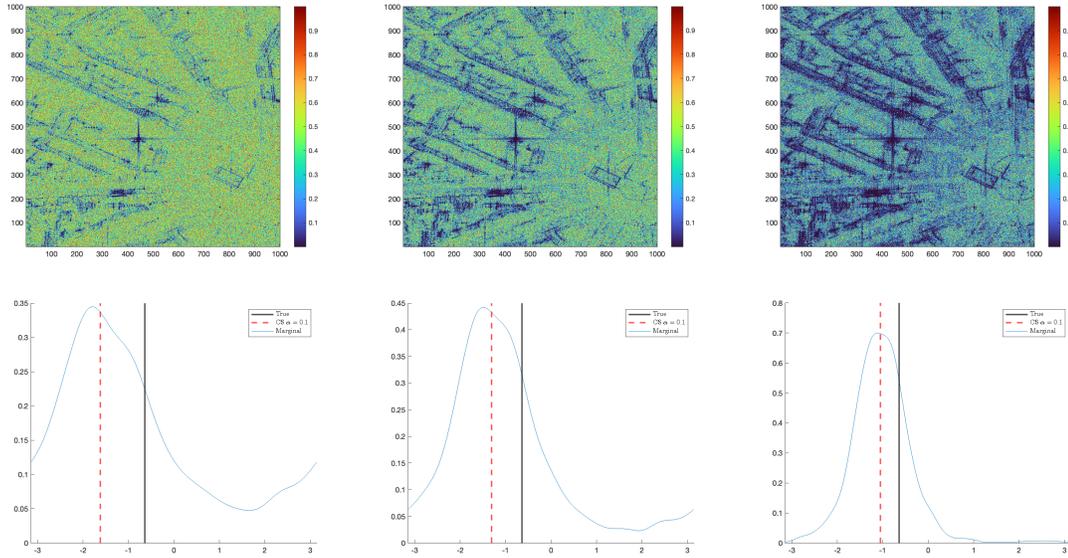


Figure 5.13: (top) Pointwise (circular) phase variance, (bottom) phase results at a randomly chosen pixel. (left) SNR = -5 dB, (middle) SNR = 0 dB, (right) SNR = 5 dB.

pixel is lost. By contrast, our method does not cause this loss of phase information. Finally, Figure 5.14 provides magnitude results for $\text{SNR} = 5$ and -5 . Observe our

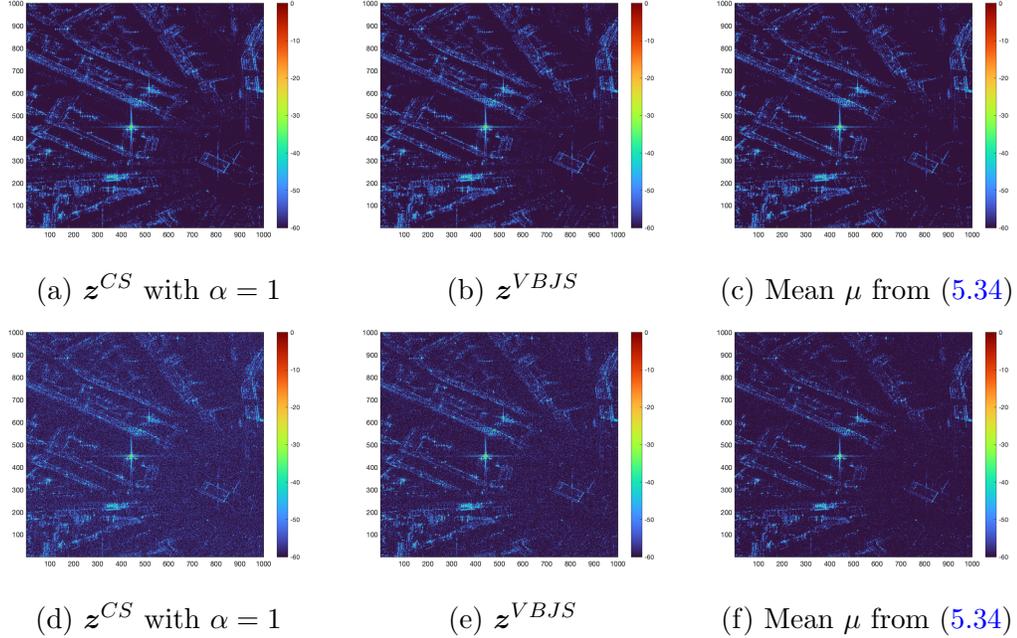


Figure 5.14: (top) $\text{SNR} = 5$ dB; (bottom) $\text{SNR} = -5$ dB.

method continues to outperform the compressed sensing and VBS approaches by suppressing noise in regions with true reflectivity at or below -60 dB.

Section 5.5

Concluding remarks

We introduced a new sampling algorithm to recover complex-valued images from multiple Fourier observations. In addition to point estimates, our method provides uncertainty information. Our approach utilizes a support-informed prior by constructing a weighted mask directly from the measurement vectors, allowing for efficient sampling without losing valuable phase information. By design, and as our results demonstrate, uncertainty is reduced in sparse regions of the magnitude gradient domain. Moreover, while optimization algorithms such as ADMM are required to compute z^{CS} in (5.44)

and $\mathbf{z}^{VBJ S}$ in (5.46) in the sparse signal case, our approach only requires the computation of W in (5.31) and of $\boldsymbol{\mu}$ and Γ in (5.34). Although our investigation focused on sparsity in either the signal or magnitude gradient domain, this is not an inherent limitation of our method. Indeed the concentration factor edge detection method can readily accommodate sparse transform domains constructed using HOTV operators.

While SAR imaging provided our initial motivation, our new methodology is applicable to other coherent imaging modalities, such as ultrasound and digital holography. In some cases prior phase information may be available, and this can also be incorporated into our sampling method. This will be the focus of future work. Finally, other edge detection methods, such as the polynomial annihilation method in [6], can be used to construct the weight mask when the data are provided in image domain. These types of techniques methods will also be included into future adaptations of our approach.

Chapter 6

Leveraging Joint Sparsity in 3D Synthetic Aperture Radar Imaging

This chapter describes work published in [56] in *Applied Mathematics for Modern Challenges*.

Section 6.1

Introduction

Spotlight-mode synthetic aperture radar (SAR) is an all-weather sensing modality capable of imaging through all illumination conditions. SAR data are ubiquitous in several applications, such as sea ice monitoring and military reconnaissance. The current state of the art for the reconstruction of two-dimensional (2D) SAR images includes the matched filter, filtered backprojection, polar format, and compressive sensing methods, along with sampling-based SAR image formation [34, 54, 93, 105]. These techniques all enable reconstructions of large scenes with high resolution and fidelity.

DISTRIBUTION A. Approved for public release; distribution unlimited.
PA Approval #:[ARFL-2023-3703]

While 2D SAR imaging is already a mature field of study, 3D SAR reconstruction is becoming increasingly important. Specifically, 3D SAR is capable of providing practitioners with a more complete representation of a target of interest by providing height resolution as well as relative scaling of the object’s three spatial dimensions. Novel techniques and powerful computational capabilities are driving the push for efficient and accurate reconstructions of 3D landscapes and objects. These approaches hold the promise of enhanced target recognition and identification, detailed topographic maps, and improved change detection [4, 8, 31, 116].

In this work we build on two particular methods of 3D SAR image reconstruction, namely backprojection and sparsity-promoting recovery [68, 113]. In both of these, 2D images centered at the origin are recovered and then used to form the volumetric image. Our methods seek to extend the sparsity-promoting inversion approach, first by leveraging sequential information from neighboring apertures, and then by utilizing techniques from hierarchical Bayesian modeling [20, 72]. This serves not only to extend the dynamic range of the recovered volumetric image and to lower the dependency on user input when compared to the previous sparsity-based method [113], but also has the potential to quantify the uncertainty regarding the reconstruction, which is not generally possible with point-estimate methods.

The rest of this chapter is organized as follows. In Section 6.2 we provide a brief overview of the data collection process as well as details of the data set used in our experiments. In Section 6.3 we discuss how to leverage sequential joint sparsity information in point estimate image recovery. We extend these ideas to a Bayesian framework in Section 6.4. Section 6.5 contains some numerical experiments, and we provide concluding remarks in Section 6.6.

Section 6.2

SAR data collection

Spotlight SAR data are acquired as the imaging platform revolves circularly around a scene of interest. A chirp signal is transmitted by the radar, and backscattered signal is then detected by the antenna [91]. In a SAR system that emits linear frequency modulated chirps (2.26), the frequency function $\omega(t)$ is given by

$$\omega(t) = \frac{2}{c}[\omega_0 + 2\alpha(t - \tau_0)], \quad |\omega(t)| \leq \frac{2}{c}(\omega_0 + \alpha T_t), \quad (6.1)$$

where c is the speed of light, ω_0 is the carrier frequency of the chirp, 2α is the chirp rate, T_t is the pulse duration, τ_0 is the round trip time of the chirp to the scene center, and t is the fast time variable. Other measurement parameters include azimuth angle $\theta(\tau)$ and elevation angle $\varphi(\tau)$, where τ is the slow time parameter. The spatial frequency locations \vec{k} are related to the temporal frequency function (6.1) by

$$\vec{k} = [k_x, k_y, k_z]^T = \omega(t)[\cos(\theta(\tau)) \cos(\varphi(\tau)), \sin(\theta(\tau)) \cos(\varphi(\tau)), \sin(\varphi(\tau))]^T. \quad (6.2)$$

More specifically, each value \vec{k} where data may be acquired is governed by the flight path of the imaging platform, which provides $\theta(\tau)$ and $\varphi(\tau)$, along with the band where the radar operates, which determines the range of $\omega(t)$.

Following demodulation, the spotlight SAR data collection process can be modeled as the 3D Fourier transform (2.35) of the true underlying scene, $g : \mathbb{R}^3 \rightarrow \mathbb{C}$, and is given by

$$\hat{g}(\vec{k}) = \mathcal{F}^3[g](\vec{k}) = \int_{\mathbb{R}^3} g(\vec{x}) \exp \left\{ -i \left[\vec{x}^T \vec{k} \right] \right\} d\vec{x}, \quad \vec{x} = [x, y, z]^T. \quad (6.3)$$

A more thorough introduction to SAR from a mathematical perspective can be found in [32, 33].

6.2.1. The Fourier Slice Theorem

The Fourier Slice Theorem [92] provides a convenient way to model the collected data. Let $f : \mathbb{R}^3 \rightarrow \mathbb{C}$. We are interested in a specific 2D slice of f , where the slice operator is defined by

$$S_2^3[h](\mu, \nu) = h(\mu, \nu, 0). \quad (6.4)$$

Observe that S_2^3 reduces the dimensionality of f to \mathbb{R}^2 .

Remark 6.1. Our use of generic variables in (6.4) is intentional. Theorem 6.1 (Fourier Slice Theorem), which is foundational to many SAR image recovery algorithms, considers the slice operator in the Fourier domain (with $h := \hat{g}$ in the theorem). We later apply (6.4) in the physical domain for the volumetric SAR reconstruction using 2D filtered backprojection (see Section 6.3.2).

We will also make use of the following definition:

Definition 6.1. Let \mathcal{B} be a rotational change of basis. The function $g = f_{\mathcal{B}}$ is a rotation of f about the origin so that $S_2^3[g] = S_2^3[f_{\mathcal{B}}]$ is the 2D slice of interest. The corresponding inverse rotation \mathcal{B}^{-1} is defined such that if $g = f_{\mathcal{B}}$, then $g_{\mathcal{B}^{-1}} = f$.

Finally, we define the integral projection operator that projects f onto \mathbb{R}^2 by integrating out the third dimension as

$$P_2^3[h](\mu, \nu) = \int_{-\infty}^{\infty} h(\mu, \nu, \xi) d\xi. \quad (6.5)$$

Theorem 6.1 relates (6.3), (6.5), and (6.4) in the context of the rotational change of basis \mathcal{B} and provides the foundation for the filtered backprojection algorithm.

Theorem 6.1 (The Fourier Slice Theorem). *Suppose we are given \vec{k} and \vec{x} in (6.2), $\hat{g}(\vec{k})$ in (6.3), P_2^3 in (6.5), and S_2^3 in (6.4). Let \mathcal{B} be an arbitrary rotational change of basis given in Definition 6.1. Then the slice S_2^3 of $\hat{g}_{\mathcal{B}}$ is the 2D Fourier transform, \mathcal{F}^2 , of the projection P_2^3 of $g_{\mathcal{B}}$, i.e.*

$$S_2^3[\hat{g}_{\mathcal{B}}] = \mathcal{F}^2 \circ P_2^3[g_{\mathcal{B}}]. \quad (6.6)$$

Proof. Due to the rotation property of Fourier transforms [66, Appendix A], without loss of generality we can let \mathcal{B} be the identity. In this case $g = g_{\mathcal{B}}$ and $\hat{g} = \hat{g}_{\mathcal{B}}$. From (6.3) we then have

$$\begin{aligned} S_2^3[\hat{g}](k_x, k_y) &= \hat{g}(k_x, k_y, 0) \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y, z) \exp\{-i(xk_x + yk_y)\} dx dy dz \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} g(x, y, z) dz \right] \exp\{-i(xk_x + yk_y)\} dx dy \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} P_2^3[g](x, y) \exp\{-i(xk_x + yk_y)\} dx dy \\ &= (\mathcal{F}^2 \circ P_2^3[g])(k_x, k_y) \end{aligned}$$

□

Figure 6.1 summarizes the implications of Theorem 6.1, which is well-known and is used in many SAR image recovery algorithms [54, 68, 113]. We use it here in the development of our own algorithm by treating 2D slices of frequency domain data as the Fourier transform of the 2D projections of the 3D spatial scene.

6.2.2. Data used in experiments

We use one synthetic data set (see Figure 6.3) and one measured SAR data collect (see Figure 6.5) to evaluate our new methods and compare their performance to

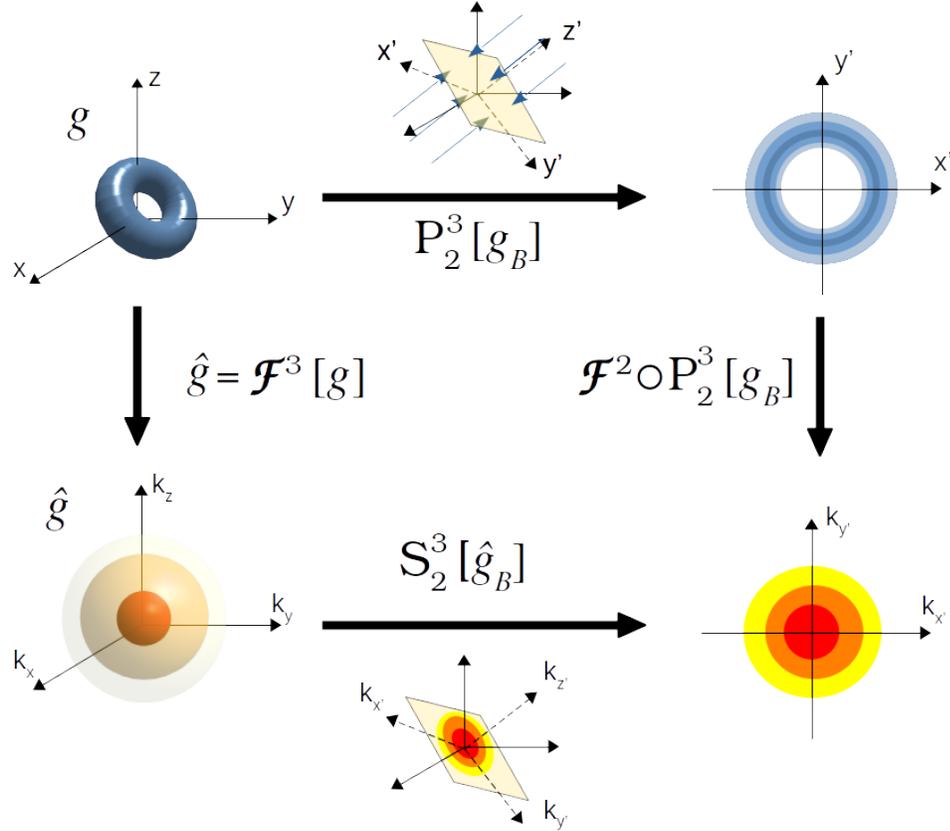


Figure 6.1: A graphical depiction of Theorem 6.1.

established techniques. The synthetic data set consists of a cube centered at the origin, while the measured data set is the same model B747 data set used in [68, 113].

The B747 data set is created with an asymptotic prediction code that simulates data collections taken over multiple passes around a target of interest in a circular flight path. All data were collected in the Ka band where a total of 13 passes are made over the full azimuth range with elevation -3° to 3° with 0.5° spacing. Table 6.1 summarizes these parameters.

6.2.3. SAR image recovery

Our technique builds on methodology used in SAR image formation and hierarchical Bayesian inference. In particular, our forward problem is discretized from (6.3) and

Parameter	Dataset Value
Elevation Range	$[-3^\circ, 3^\circ]$
Elevation Sampling	0.5°
Frequency Range	$[27, 39]$ GHz
Frequency Sampling	50MHz
Bandwidth	12GHz
Center Frequency	33GHz
Azimuth Range	$[0^\circ, 359.9^\circ]$
Azimuth Sampling	0.1°

Table 6.1: Parameters of synthetic and measured data sets used for experimentation.

then modeled as

$$\hat{\mathbf{g}} = \mathcal{F}\mathbf{g} + \boldsymbol{\varepsilon}, \quad (6.7)$$

where $\hat{\mathbf{g}} \in \mathbb{C}^M$ is the SAR phase history data (PHD), $\boldsymbol{\varepsilon} \in \mathbb{C}^M$ is complex-valued circularly symmetric Gaussian noise with covariance matrix $\sigma^2\mathbb{I}_M$, $\mathbf{g} \in \mathbb{C}^N$ is the image we seek to recover, and \mathcal{F} is the (discrete) nonuniform Fourier transform (NUFT) matrix. The d -dimensional NUFT matrix $\mathcal{F}^d \in \mathbb{C}^{M \times N}$ is given by

$$[\mathcal{F}^d]_{m,n} = \exp(-i\mathbf{p}_m \cdot \mathbf{q}_n), \quad \mathbf{p}_m, \mathbf{q}_n \in \mathbb{R}^d,$$

where \mathbf{p}_m are the nonuniform frequencies and \mathbf{q}_n are the nonuniform sources. Note that \mathbf{g} and $\hat{\mathbf{g}}$ may be 2D or 3D, but we are able to vectorize \mathbf{g} and $\hat{\mathbf{g}}$ and then formulate \mathcal{F} accordingly. All numerical experiments in this investigation are implemented using the nonuniform fast Fourier transform library FINUFFT developed in [11]. In (6.7) we assume errors corresponding to aliasing and gridding are insignificant compared to inherent system noise.

Section 6.3

SAR imaging leveraging sequential information

The data considered in the SAR image formation process are in general determined by the flight path of the SAR imaging platform along with the frequency band over which data are collected. From (6.7), the SAR PHD measurements $\hat{\mathbf{g}}$ in 3D k -space at the respective equispaced azimuth angles, elevation angles, and frequencies, are given by

$$(f_m \cos \theta_p \cos \varphi_r, f_m \sin \theta_p \cos \varphi_r, f_m \sin \varphi_r), \quad \{\theta_p\}_{p=1}^P, \quad \{\varphi_r\}_{r=1}^R, \quad \{f_m\}_{m=1}^{N_k}. \quad (6.8)$$

We will denote the spacing between azimuth angles, elevation angles, and frequencies as $\Delta\theta$, $\Delta\varphi$, and Δf . We also have that the data are collected by a total of $N_p = PR$ pulses and that each of these pulses occurs at slow time $\tau_{p,r}$, $p = 1, \dots, P$, $r = 1, \dots, R$. In several of the techniques that follow, it may become necessary to further partition the azimuth angles.¹ In such cases we consider the azimuth angles $\theta_1, \dots, \theta_P$ to be partitioned into N_θ sets $\Theta_1, \dots, \Theta_{N_\theta}$

$$\Theta_n = \left\{ \theta_p : \frac{P}{N_\theta}(n-1) + 1 \leq p \leq \frac{P}{N_\theta}n \right\}, \quad n = 1, \dots, N_\theta. \quad (6.9)$$

An example of SAR PHD collected at three elevation angles over all azimuth angles is given in Figure 6.2.

Due to SAR's specular scattering physics, the level of backscatter detected is not only dependent on the imaging platform position, but also on the geometry of the imaging scene. Hence the measured return from a point in the scene may have a strong dependence on the angle from which it is viewed, i.e. the scattering is *not*

¹Consistent with standard SAR imaging practices, it is always assumed that the data acquisition is sufficient for such partitions.

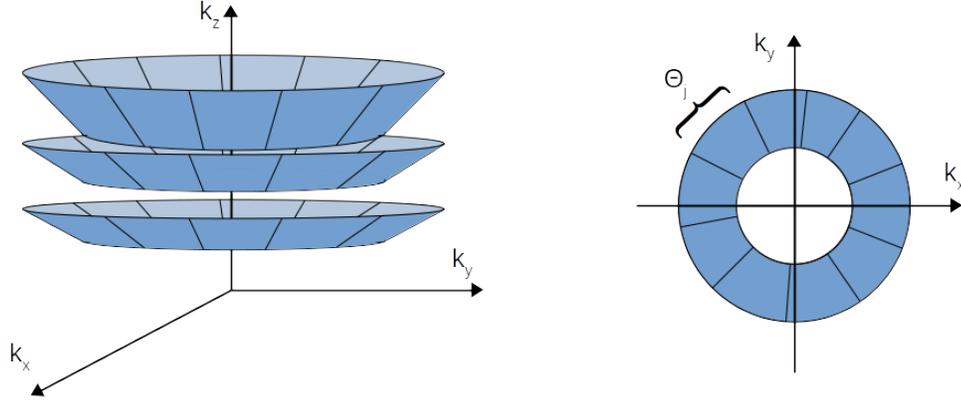


Figure 6.2: A graphical depiction of SAR PHD $\hat{\mathbf{g}}$ (6.3) in k -space as well as the partitioning of the data into N_θ partitions according to the azimuthal angle sets Θ_n given by (6.9).

isotropic as suggested in the model given by (6.7). This issue has been addressed using *composite imaging*, which mitigates the effects of this incorrect assumption by using some weighted average, or maximum, of image approximations recovered from different subapertures, see e.g. [36, 100, 107].

In Section 6.3.1 we review a composite 2D SAR imaging technique which leverages the assumption that neighboring subapertures have similar support in their sparse domain [107]. In Section 6.3.2 we review a volumetric approach using a three-dimensional analogue of the classic backprojection algorithm [113]. These two ideas are then fused in Section 6.3.3 to construct an algorithm that compares the spatial information in neighboring imaging planes and penalizes the differences.

6.3.1. 2D SAR imaging leveraging sequential information

We first explore the sequential imaging approach introduced in [107]. This method seeks to recover a 2D scene from SAR PHD gathered at a single (fixed) elevation angle φ_r . We will use all data $\hat{\mathbf{g}}^{(r)}$, gathered at elevation angle φ_r to reconstruct the 2D scene of interest $\tilde{\mathbf{g}} = \tilde{\mathbf{g}}^{(r)}$, for any $r \in [1, R]$.

To begin the image formation process, we first partition the SAR PHD data (6.3)

into N_θ sequential subaperture bins $\hat{\mathbf{g}}_1^{(r)}, \dots, \hat{\mathbf{g}}_{N_\theta}^{(r)}$ so that the forward model (6.7) becomes

$$\hat{\mathbf{g}}_n^{(r)} = \mathcal{F}_n \tilde{\mathbf{g}}_n + \boldsymbol{\varepsilon}_n, \quad n = 1, \dots, N_\theta,$$

where $\tilde{\mathbf{g}}_n$ are the optimized 2D images we seek to recover, \mathcal{F}_n is the corresponding NUFT for subaperture Θ_n , and $\boldsymbol{\varepsilon}_n$ is complex-valued circularly-symmetric Gaussian noise with covariance matrix $\sigma^2 \mathbb{I}$. We note that the partitioning of the SAR PHD is based on (6.9).

Though the strength of the backscatter may change as the azimuth changes, it tends to do so continuously as a function of the viewing angle. We can therefore expect images formed from sequential subapertures to contain similarities. It is important to keep in mind that $\tilde{\mathbf{g}}$ is complex-valued. In this regard, when we are discussing sparsity, it is the magnitude of $\tilde{\mathbf{g}}$ that is sparse, not the signal itself. Hence when computing terms involving sparsity, the magnitude must first be extracted. This can be accomplished by decomposing $\tilde{\mathbf{g}}_n = |\tilde{\mathbf{g}}_n| \Psi_n$. We then approximate the phase of the pixels of each image using $\Psi_n = D(\exp(i \text{ang}(\tilde{\mathbf{g}}_n)))$, where $D(\cdot) = \text{diag}(\cdot)$, yielding the estimate $|\tilde{\mathbf{g}}_n| \approx \Psi_n^H \tilde{\mathbf{g}}_n$. Using this reasoning as well as the assumption that the images themselves are sparse, we obtain the objective function ([107])

$$\tilde{\mathbf{G}} = \arg \min_{(\mathbf{q}_1, \dots, \mathbf{q}_{N_\theta})} \left\{ \sum_{n=1}^{N_\theta} \left(\frac{\mu}{2} \|\mathcal{F}_n \mathbf{q}_n - \hat{\mathbf{g}}_n^{(r)}\|_2^2 + \|\mathbf{q}_n\|_1 \right) + \gamma \sum_{n=1}^{N_\theta-1} \|\Psi_{n+1}^H \mathbf{q}_{n+1} - \Psi_n^H \mathbf{q}_n\|_2^2 \right\}. \quad (6.10)$$

Here $\tilde{\mathbf{G}} = (\tilde{\mathbf{g}}_1, \dots, \tilde{\mathbf{g}}_{N_\theta})^T$, μ and γ are regularization parameters and Ψ_n are diagonal matrices containing the estimated phase at each pixel of \mathbf{q}_n . Observe that (6.10) describes a compressive sensing (CS) approach [27] coupled with a term that leverages neighboring (sequential) information, which is designed to promote solutions that are

sequentially structurally similar. In particular, the assumption here is that two sequential images, $\tilde{\mathbf{g}}_n, \tilde{\mathbf{g}}_{n+1}$, should have the same sparse magnitude structure. We solve (6.10) using the well-known alternating direction method of multipliers (ADMM) algorithm [17]. Once all $\tilde{\mathbf{g}}_1, \dots, \tilde{\mathbf{g}}_{N_\theta}$ are recovered, a final composite vectorized image $\tilde{\mathbf{g}}$ is formed where

$$\tilde{\mathbf{g}}^i = \arg \max_{n \in [1, N_\theta]} |\tilde{\mathbf{g}}_n^i| \quad (6.11)$$

for each pixel $i = 1, \dots, N$.

Remark 6.2. We note that leveraging sequential information in the sparse domain in (6.10) does not require use of the ℓ_1 norm. This is because it is not the sparsity property that is being utilized in this term, but rather the idea that the difference should be small. This is in contrast to the ℓ_1 norm used in the first term, which is designed to promote sparsity in the underlying scene. There is an advantage to considering sequential sparse domain information, however, as opposed to sequential measurement information. This is because most entries in the sparse domain will contribute (close to) zero value, and the overall difference between neighboring apertures can be better captured this way.

6.3.2. Volumetric SAR using backprojection

Now denote the SAR PHD collected by $N_p = PR$ pulses over a range of N_k frequencies as

$$S(f_m, \tau_{p,r}), \quad m = 1, \dots, N_k, \quad p = 1, \dots, P, \quad r = 1, \dots, R, \quad (6.12)$$

where $\tau_{p,r}$ represent slow time pulses along azimuth θ_p and elevation φ_r .

In SAR imaging, the signal returns are sorted into bins corresponding to different

ranges in the spatial domain based on time of arrival. In what follows, each discrete range bin is indexed by $\ell = 1, \dots, L$, while \mathbf{x} denotes the vector of pixel locations in the recovered image. Here L is inversely proportional to the range resolution δ [54], which is given by

$$\delta = \frac{c}{2(N_k - 1)\Delta f},$$

and c is again the speed of light.

2D filtered backprojection. Filtered backprojection (FBP) is a common technique used to recover 2D SAR images [54]. Due to the polar formatting of the data, lower frequencies are sampled at a greater rate than higher frequencies. To account for this, the FBP method applies a filter (e.g. a ramp function) to the PHD before the data are processed to accentuate the higher frequency terms. The inverse Fourier transform is then performed on the (filtered) data collected at each time $\tau_{p,r}$. The resulting 1D function is then backprojected over the whole 2D domain. In the last step these 2D images are summed together to form the final image. Algorithm 9 summarizes the standard FBP algorithm for 2D SAR. A thorough analysis of FBP for SAR imaging can be found in [39].

Algorithm 9 FBP for 2D SAR PHD

Input SAR PHD $S(f_m, \tau_{p,r})$, $m = 1, \dots, N_k$, $p = 1, \dots, P$, and $r = 1, \dots, R$ in (6.12), and L range bins.

Output Image $\tilde{\mathbf{g}}$.

Apply filter to PHD.

for $n = 1 : P$ **do**

Set $s(\ell, \tau_{p,r}) = \sum_{m=1}^{N_k} S(f_m, \tau_{p,r}) \exp\left(\frac{2\pi i f_m (\ell-1)}{N_k \Delta f}\right)$, $\ell = 1, \dots, L$.

Interpolate values of $s(\ell, \tau_{p,r})$ to a rectangular grid as $s_{int}(\mathbf{x}, \tau_{p,r})$.²

end for

$\tilde{\mathbf{g}} = \sum_{r=1}^R \sum_{p=1}^P s_{int}(\mathbf{x}, \tau_{p,r})$.

Volumetric SAR. With the 2D FBP now in hand, we proceed to incorporate information from multiple elevation angles to obtain a volumetric SAR image. We begin by considering an arbitrary set of azimuth and elevation angle pairs $\{(\theta_s, \varphi_s)\}_{s=1}^S$. As we will see, these pairs of angles each define a plane onto which we will use Algorithm 9 to construct 2D slices of the 3D scene of interest \mathbf{g} .

Let \mathcal{B}_s , $s = 1, \dots, S$, be the 3D rotational change of basis operator (see Definition 6.1) from (x, y, z) to (x'_s, y'_s, z'_s) such that the z'_s -axis has azimuth angle $\theta_s \in [0, 2\pi)$ and elevation angle $\varphi_s \in (-\frac{\pi}{2}, \frac{\pi}{2}]$ from the (x, y, z) coordinate system. Our goal is to recover $\mathbf{g}_s = S_2^3[\mathbf{g}_{\mathcal{B}_s}]$, which is accomplished by performing Algorithm 9 for each imaging plane using the *full* data set $\hat{\mathbf{g}}$ in (6.3) [113]. The resulting S images are then fused together to form the volumetric image by constructing a radial point cloud, with the ability to interpolate this point cloud onto a Cartesian grid as desired. As shown previously in [68, 113], this technique can consider any combination of azimuth and elevation angles. Algorithm 10 summarizes this process.

Algorithm 10 Volumetric SAR using 2D FBP

Input SAR PHD $\hat{\mathbf{g}}$ (6.3) and threshold value c_{thresh} .

Output Binary volumetric image \mathbf{g} .

Generate azimuth and elevation angle pairs θ_s, φ_s for $s = 1, \dots, S$.

Initialize volumetric image \mathbf{g} .

for $s = 1 : S$ **do**

Derive filtered backprojection \mathbf{g}_s on imaging plane with angles θ_s and φ_s using Algorithm 9.

Create $\bar{\mathbf{g}}_s$ by thresholding \mathbf{g}_s with $\bar{\mathbf{g}}_s = \begin{cases} 1 & \text{if } |\mathbf{g}_s| \geq c_{thresh} \\ 0 & \text{else} \end{cases}$.

For each $i = 1, \dots, N$ such that pixel $[(\bar{\mathbf{g}}_s)_{\mathcal{B}_s^{-1}}]^i = 1$, set $\mathbf{g}^i = 1$.

end for

Remark 6.3. Since the data can be backprojected onto as many planes we choose, S is in some sense arbitrary. The resolution of the image is tied to the amount of

²Since the range bins indexed by $l = 1, \dots, L$ typically do not align with Cartesian grid points \mathbf{x} , interpolation is needed to form a final pixelated image. A review of the interpolation methods commonly used in signal processing can be found in [38].

data given in (6.8), however. For the B747 data set (see Table 6.1), replaced with .1 degree spacing in elevation, the experiment used in [68] set $S = 180$ (corresponding to $R = 61$ and $P = 3600$). The emphasis there was to fix the elevation to be 90° for the purpose of height extraction.

6.3.3. Volumetric SAR leveraging sequential information

The objective function (6.10) considers inter- and intra-image information, but only for data collected at a single elevation angle. By contrast, Algorithm 10 considers the full SAR PHD $\hat{\mathbf{g}}$ for azimuth angles $\theta_1 < \dots < \theta_P$ and elevation angles $\varphi_1 < \dots < \varphi_R$, as given in (6.8), but depends exclusively on information given by the acquired data. That is, it does not infer any assumptions regarding the underlying volumetric image, such as intra-image and sequential sparsity. The method in [68, 113] extends (6.10) to three dimensions to include this *a priori* information. Since this leveraging of *a priori* inter-image and sequential information in the the 3D setting inspires our hierarchical Bayesian approach to the 3D SAR image formation problem, we include it here.

In the first step, an arbitrary set of azimuth and elevation angles, given respectively as $\vartheta_1, \dots, \vartheta_U$ and ϕ_1, \dots, ϕ_V , is established. The goal is then to recover *slices* of the volumetric image \mathbf{g} defined by every possible combination of these azimuth and elevation angles, i.e.

$$\mathbf{g}_{u,v} = S_2^3[\mathbf{g}_{\mathcal{B}_{u,v}}], \quad u = 1, \dots, U, \quad v = 1, \dots, V, \quad (6.13)$$

where $\mathcal{B}_{u,v}$ is the rotational change of basis corresponding to azimuth rotation ϑ_u and elevation rotation ϕ_v .

Letting $A_{u,v}$ be the forward operator from each corresponding slice to the full SAR

PHD $\hat{\mathbf{g}}$ in (6.3), the objective function solved in [68, 113] is then given by

$$\mathbf{g}_{u,v} = \arg \min_{\mathbf{q}_{u,v}} \left\{ \|A_{u,v}\mathbf{q}_{u,v} - \hat{\mathbf{g}}\|_2^2 + \lambda_1 \|\mathbf{q}_{u,v}\|_1 + \frac{\lambda_2}{2} \sum_{i=u-1}^{u+1} \sum_{k=v-1}^{v+1} \|\Psi_{u,v}^H \mathbf{q}_{u,v} - \Psi_{i,k}^H \mathbf{q}_{i,k}\|_2^2 \right\}, \quad (6.14)$$

where λ_1 and λ_2 are regularization parameters and $\Psi_{u,v}$ are diagonal matrices containing the estimated phase at each pixel of $\mathbf{q}_{u,v}$. As is the case in (6.10), the first term in (6.14) enforces data fidelity, the second term promotes intra-image sparsity, and the third term encourages inter-image sparsity. The minimization problem given by (6.14) is then solved in [68, 113] using ADMM. Once the set $\{\mathbf{g}_{u,v}\}_{u=1,v=1}^{U,V}$ is recovered, the resulting slices are fused together to form the volumetric image \mathbf{g} .³

Section 6.4

Bayesian approach to 3D SAR image formation

To make use of statistical inversion methods as well as to quantify the uncertainty of the signal recovery, we now cast the inverse imaging problem in a Bayesian setting. Consider the linear inverse problem

$$\hat{\mathcal{G}}_j = \mathcal{F}_j \mathcal{G}_j + \hat{\mathcal{E}}_j, \quad j = 1, \dots, J, \quad (6.15)$$

where \mathcal{G}_j , $\hat{\mathcal{G}}_j$, and $\hat{\mathcal{E}}_j$ are random variables defined over a common probability space, and \mathcal{G}_j and $\hat{\mathcal{E}}_j$ are assumed to be independent. In this framework \mathcal{G}_j represents the unknown we seek to recover, $\hat{\mathcal{G}}_j$ are the data, \mathcal{F}_j is a known linear operator, and $\hat{\mathcal{E}}_j \sim \mathcal{CN}(0, [D(\boldsymbol{\alpha})]^{-1})$ where again we have $D(\cdot) = \text{diag}(\cdot)$. with some noise precision vector $\boldsymbol{\alpha}$. In our technique, \mathcal{F}_j takes the form of the 2D or 3D NUFT.

³We note that (6.14) is modified from the point estimate derived in [68, 113] in two ways: (1) Here we include phase extraction in the sequential difference regularization term, and (2) there $\hat{\mathbf{g}}$ was rotated and projected (according to u, v) before incorporating it into the objective function.

By treating the data and the unknown image as random variables, we are able to leverage hierarchical Bayesian learning methods by creating appropriate likelihood and prior distributions to describe our data and assumptions. As such, we extend the techniques used in Sections 6.3.2 and 6.3.3 to a Bayesian framework. Algorithms that utilize joint sparsity given multiple measurements in an empirical Bayesian setting were introduced in [138] and [132] for real-valued images. In [138] multiple data acquisitions at a single time are assumed, while [132] considered a *temporal* sequence of data acquisitions. Here we modify the technique coined the Joint Hierarchical Bayesian Learning (JHBL) method in [132] to consider a *spatial* sequence of image reconstructions. Analogous to the volumetric point estimate in (6.14), in which the second term incorporates intra-signal information and the final term leverages the sequential inter-signal similarities, in our JHBL approach we construct the priors to leverage intra- and inter-image information for a more accurate point estimate SAR image recovery. The priors are furthermore designed to be conjugate to the likelihood, enabling a closed form for the posterior from which we can efficiently sample.

We derive the general formulation for our approach in Section 6.4.1. In Section 6.4.2 we consider sequential data acquisitions along azimuth angles θ_p , $p = 1, \dots, P$, while in Section 6.4.3 we incorporate the idea that in the volumetric image reconstruction, data from neighboring subapertures should contain similar information.

6.4.1. Hierarchical Bayesian model

Following the Bayesian model in (6.15), let $\mathcal{G} = \{\mathcal{G}_j \in \mathbb{C}^N : j = 1, \dots, J\}$ be the collection of signals we seek to recover, where N is the number of pixels in each of the sequential images, $\hat{\mathcal{G}} = \{\hat{\mathcal{G}}_j \in \mathbb{C}^M : j = 1, \dots, J\}$ is the collection of J observable measurements in the frequency (PHD) domain, and $\hat{\mathcal{E}}$ is circularly symmetric additive

Gaussian white noise, i.e.

$$\pi(\varepsilon) = \frac{1}{\pi^N |[D(\boldsymbol{\alpha})]^{-1}|} \exp(-\varepsilon^H D(\boldsymbol{\alpha}) \varepsilon),$$

where ε is a realization of $\hat{\mathcal{E}}$. Samples of signals \mathcal{G} and data $\hat{\mathcal{G}}$ are correspondingly denoted as $G = \{\mathbf{g}_j \in \mathbb{C}^N : j = 1, \dots, J\}$ and $\hat{G} = \{\hat{\mathbf{g}}_j \in \mathbb{C}^M : j = 1, \dots, J\}$. While we have not yet specified how the set G relates to the volumetric image \mathbf{g} we are seeking to recover, we assume for now that the sequential sparsity assumption holds for the elements of G .

We proceed by recalling that Bayes' Theorem yields

$$\pi(G, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma} | \hat{G}) = \frac{\pi(\hat{G} | G, \boldsymbol{\alpha}) \pi(G | \boldsymbol{\beta}, \boldsymbol{\gamma}) \pi(\boldsymbol{\alpha}) \pi(\boldsymbol{\beta}) \pi(\boldsymbol{\gamma})}{\pi(\hat{G})},$$

where $\pi(G, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma} | \hat{G})$ is the posterior density function, $\pi(\hat{G} | G, \boldsymbol{\alpha})$ is the likelihood, $\pi(G | \boldsymbol{\beta}, \boldsymbol{\gamma})$ is the prior, $\pi(\boldsymbol{\alpha})$, $\pi(\boldsymbol{\beta})$, and $\pi(\boldsymbol{\gamma})$ are the hyper-priors. In this context we also define the random variables $A \in (\mathbb{R}^+)^J$, $B \in (\mathbb{R}^+)^{J \times N}$, and $C \in (\mathbb{R}^+)^{J \times N}$ as the noise precision, the precision of the intra-image prior, and the precision of the sequential sparsity-promoting prior, with realizations $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, and $\boldsymbol{\gamma}$, respectively.

Sometimes called the evidence, $\pi(\hat{G}) \neq 0$ since otherwise there would be no observations in (6.15). It is, however, typically unknown so instead it is standard to employ the relationship

$$\pi(G, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma} | \hat{G}) \propto \pi(\hat{G} | G, \boldsymbol{\alpha}) \pi(G | \boldsymbol{\beta}, \boldsymbol{\gamma}) \pi(\boldsymbol{\alpha}) \pi(\boldsymbol{\beta}) \pi(\boldsymbol{\gamma}), \quad (6.16)$$

from which we compute the right hand side, i.e. an un-normalized version of $\pi(G, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma} | \hat{G})$. The task is then to determine each of the five terms on the right hand side of (6.16), which we now describe.

The likelihood $\pi(\hat{G}|G, \boldsymbol{\alpha})$. From the noise model in (6.15) we have

$$\pi(\hat{G}|G, \boldsymbol{\alpha}) \propto \prod_{j=1}^J \exp \{-\alpha_j \|\mathcal{F}_j \mathbf{g}_j - \hat{\mathbf{g}}_j\|_2^2\}. \quad (6.17)$$

The hyper-prior for the likelihood $\pi(\boldsymbol{\alpha})$. While there are many choices for the hyper-prior on $\boldsymbol{\alpha}$, we choose to use an uninformative gamma prior on each α_j , $j = 1, \dots, J$, to maintain conjugacy and allow for flexibility regarding whether or not prior knowledge of the noise precision is known. Thus we have

$$\pi(\boldsymbol{\alpha}) \propto \prod_{j=1}^J \alpha_j^{\eta_\alpha - 1} \exp\{-\nu_\alpha \alpha_j\}, \quad (6.18)$$

where η_α and ν_α are chosen either to be in accordance with *a priori* knowledge of the noise in the images or to be uninformative. Since the mode of (6.18) is zero when $\eta_\alpha \leq 1$, values in this range promote sparsity in $\boldsymbol{\alpha}$, while smaller ν_α result in more uninformative hyperpriors [120].

The joint prior $\pi(G|\boldsymbol{\beta}, \boldsymbol{\gamma})$. To leverage both the sparsity assumption in the image magnitude and sequential information [132], we define the joint prior as

$$\pi(G|\boldsymbol{\beta}, \boldsymbol{\gamma}) := \pi(G|\boldsymbol{\beta})\pi(G|\boldsymbol{\gamma}),$$

where $\pi(G|\boldsymbol{\beta})$ and $\pi(G|\boldsymbol{\gamma})$ are the intra- and inter-image priors, respectively.

The intra-image prior $\pi(G|\boldsymbol{\beta})$. Sparsity is encouraged in the SAR image magnitude by imposing a conditional complex-valued Gaussian intra-image prior on each

image pixel as

$$\pi(G|\boldsymbol{\beta}) \propto \prod_{j=1}^J \prod_{i=1}^N \beta_{j,i} \exp \left\{ -\beta_{j,i} |g_{j,i}|^2 \right\}, \quad j = 1, \dots, J, \quad i = 1, \dots, N. \quad (6.19)$$

Here each precision $\beta_{j,i}$ is a random variable. The prior in (6.19) is commonly employed to promote sparsity because it is conjugate for the likelihood density function (6.17) and therefore results in a closed form posterior [120]. Other sparsity promoting priors may also be used.

The intra-image hyper-prior $\pi(\boldsymbol{\beta})$. Since each image is expected to have a number of relatively small-magnitude pixels, we allow the precision $\beta_{j,i}$ to vary, specifically by using a gamma distribution

$$\pi(\boldsymbol{\beta}) \propto \prod_{j=1}^J \prod_{i=1}^N \beta_{j,i}^{\eta_\beta - 1} \exp \{ -\nu_\beta \beta_{j,i} \}, \quad (6.20)$$

where η_β and ν_β are predetermined shape and rate parameters that are the same for all $\beta_{j,i}$.

The inter-image prior $\pi(G|\boldsymbol{\gamma})$. As in previous work [113], we assume that the difference in magnitude of \mathbf{g}_j compared with \mathbf{g}_{j-1} is small for $j = 1, \dots, J$. We can therefore employ the conditionally inter-image complex-valued Gaussian prior

$$\pi(G|\boldsymbol{\gamma}) \propto \prod_{j=1}^J \prod_{i=1}^N \gamma_{j,i} \exp \left\{ -\gamma_{j,i} \left| \Psi_{j-1,i}^H g_{j-1,i} - \Psi_{j,i}^H g_{j,i} \right|^2 \right\}. \quad (6.21)$$

Since the azimuth angle θ_j , $j = 1, \dots, J$, is subdivided on $[0, 2\pi)$, we also assume periodicity and impose $\mathbf{g}_J = \mathbf{g}_0$ in (6.21).

The inter-image hyper-prior $\pi(\boldsymbol{\gamma})$. Similar to the intra-image prior (6.19), we use gamma distributed hyper-priors for each $\gamma_{j,i}$ with hyper-parameters η_γ and ν_γ akin to those in (6.20), i.e.

$$\pi(\boldsymbol{\gamma}) \propto \prod_{j=1}^J \prod_{i=1}^N \gamma_{j,i}^{\eta_\gamma - 1} \exp\{-\nu_\gamma \gamma_{j,i}\}. \quad (6.22)$$

The posterior. Combining (6.17), (6.19), (6.20), (6.21) and (6.22), we are now ready to calculate the posterior density function, (6.16) as

$$\begin{aligned} \pi(G, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma} | \hat{G}) \propto & \prod_{j=1}^J \left[(\alpha_j^M \exp\{-\alpha_j \|\mathcal{F}_j \mathbf{g}_j - \hat{\mathbf{g}}_j\|_2^2\}) \times \left(\prod_{i=1}^N \beta_{j,i} \exp\{-\beta_{j,i} |g_{j,i}|^2\} \right) \times \right. \\ & \left. \left(\prod_{i=1}^N \gamma_{j,i} \exp\{-\gamma_{j,i} |\Psi_{j-1,i}^H \mathbf{g}_{j-1,i} - \Psi_{j,i}^H \mathbf{g}_{j,i}|^2\} \right) \times (\alpha_j^{\eta_\alpha - 1} \exp\{-\nu_\alpha \alpha_j\}) \times \right. \\ & \left. \left(\prod_{i=1}^N \beta_{j,i}^{\eta_\beta - 1} \exp\{-\nu_\beta \beta_{j,i}\} \right) \times \left(\prod_{i=1}^N \gamma_{j,i}^{\eta_\gamma - 1} \exp\{-\nu_\gamma \gamma_{j,i}\} \right) \right]. \quad (6.23) \end{aligned}$$

Due to the structure of (6.23), we can decompose $\pi(G, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma} | \hat{G})$ into conditional distributions whose modes we are able to analytically derive. Specifically, we can update from iteration step ℓ to iteration step $\ell + 1$ for each \mathbf{g}_j , α_j , $\beta_{j,i}$, and $\gamma_{j,i}$ for $j = 1, \dots, J$, $i = 1, \dots, N$, as

$$\begin{aligned} \mathbf{g}_j^{(\ell+1)} = & \left(\alpha_j^{(\ell)} \mathcal{F}_j^H \mathcal{F}_j + D(\boldsymbol{\beta}_j^{(\ell)}) + D(\boldsymbol{\gamma}_j^{(\ell)}) + D(\boldsymbol{\gamma}_{j+1}^{(\ell)}) \right)^{-1} \\ & \times \left(\alpha_j^{(\ell)} \mathcal{F}_j^H \hat{\mathbf{g}}_j + \Psi_j^{(\ell)} D(\boldsymbol{\gamma}_j^{(\ell)}) \Psi_{j-1}^{(\ell)H} \mathbf{g}_{j-1}^{(\ell)} + \Psi_j^{(\ell)} D(\boldsymbol{\gamma}_{j+1}^{(\ell)}) \Psi_{j+1}^{(\ell)H} \mathbf{g}_{j+1}^{(\ell)} \right) \end{aligned} \quad (6.24)$$

$$\alpha_j^{(\ell+1)} = \frac{\eta_\alpha + M - 1}{\nu_\alpha + \left\| \mathcal{F}_j \mathbf{g}_j^{(\ell+1)} - \hat{\mathbf{g}}_j \right\|_2^2} \quad (6.25)$$

$$\beta_{j,i}^{(\ell+1)} = \frac{\eta_\beta}{\nu_\beta + \left| \mathbf{g}_{j,i}^{(\ell+1)} \right|^2} \quad (6.26)$$

$$\gamma_{j,i}^{(\ell+1)} = \frac{\eta_\gamma}{\nu_\gamma + \left| \Psi_{j-1,i}^{(\ell+1)H} \mathbf{g}_{j-1,i}^{(\ell+1)} - \Psi_{j,i}^{(\ell+1)H} \mathbf{g}_{j,i}^{(\ell+1)} \right|^2}, \quad (6.27)$$

where $\gamma_{J+1,i} = \gamma_{1,i}$ and $\gamma_{0,i} = \gamma_{J,i}$ for $i = 1, \dots, N$. The derivations of (6.24), (6.25), (6.26), and (6.27) are provided in Appendix 6.7. We note that (6.24) does not rely on any \mathbf{g}_j from the $\ell + 1$ step, which increases opportunities for parallelization.

Remark 6.4. Since \mathcal{F}_j is a non-uniform Fourier transform matrix, it is not unitary. However most of the mass in $\mathcal{F}_j^H \mathcal{F}_j$ is concentrated near the diagonal, so following [35] we choose to approximate $\mathcal{F}_j^H \mathcal{F}_j \approx \mathbb{I}$ for computational simplicity. When using this approximation is not desirable, other techniques may be employed in the \mathbf{g} update to avoid inverting large matrices, such as the gradient descent method [53].

Algorithm 11 summarizes how the MAP estimate of (6.23) for each \mathbf{g}_j , $j = 1, \dots, J$, is obtained based on the update steps (6.24), (6.25), (6.26) and (6.27). Observe that each of these parameters is updated based on the mode of its conditional distribution, and is then fixed as updates are made on subsequent parameters.

We will employ Algorithm 11 in two different contexts for 3D SAR image reconstruction. As described in Section 6.4.2, our first approach considers 2D slices of the frequency domain data whose Fourier transforms can be interpreted as projections of the 3D scene of interest onto the corresponding 2D plane, similar to what was done in [68, 113]. In this case Algorithm 11 is used over sequential azimuthal angles, $\{\theta_p\}_{p=1}^P$ (Algorithm 12). By contrast, Algorithm 13 in Section 6.4.3 is performed over volumetric images formed by 3D subapertures, i.e. where the sequenced information is over partitions of the azimuthal angles $\{\Theta_j\}_{j=1}^J$, (6.9). A composite image is then created from these subaperture reconstructions using (6.30).

Algorithm 11 Joint Hierarchical Bayesian Learning for J sequential data acquisitions (JHBL)

Input SAR PHD $\hat{\mathbf{g}}$ from (6.3) and hyperparameters $\eta_\alpha, \eta_\beta, \eta_\gamma, \nu_\alpha, \nu_\beta, \nu_\gamma$. Define ℓ_{max} as the maximum number of iterations and tol as the threshold determining convergence.

Output Collection of reconstructions $G = \{\mathbf{g}_j\}_{j=1}^J$, corresponding phase Ψ , and hierarchical parameters $\boldsymbol{\alpha}, \boldsymbol{\beta}$, and $\boldsymbol{\gamma}$.

if 2D IRB (Algorithm 12) **then**

$\hat{G} = \{\hat{\mathbf{g}}_p\}_{p=1}^{P/2}$ given by (6.28) and $J = P/2$.

else if 3D SRCI (Algorithm 13) **then**

$\hat{G} = \{\hat{\mathbf{g}}_n\}_{n=1}^{N_\theta}$ according to the azimuth partitions given in (6.9) and $J = N_\theta$.

end if

Initialize $G^{(0)}$ using the NUFT, set $\Psi^{(0)} = \text{ang}(G^{(0)})$, and $\boldsymbol{\alpha}^{(0)} = \boldsymbol{\beta}^{(0)} = \boldsymbol{\gamma}^{(0)} = 1$.

while $\ell < \ell_{max}$ or $\frac{1}{J} \sum_{j=1}^J \left\| \left\| \mathbf{g}_j^{(\ell)} \right\| - \left\| \mathbf{g}_j^{(\ell-1)} \right\| \right\|_1 < tol$ **do**

$G^{(\ell+1)} = \arg \max_G \pi(G | G^{(\ell)}, \boldsymbol{\alpha}^{(\ell)}, \boldsymbol{\beta}^{(\ell)}, \boldsymbol{\gamma}^{(\ell)}, \hat{G})$

$\Psi^{(\ell+1)} = \text{ang}(G^{(\ell+1)})$

$\boldsymbol{\alpha}^{(\ell+1)} = \arg \max_{\boldsymbol{\alpha}} \pi(\boldsymbol{\alpha} | G^{(\ell+1)})$

$\boldsymbol{\beta}^{(\ell+1)} = \arg \max_{\boldsymbol{\beta}} \pi(\boldsymbol{\beta} | G^{(\ell+1)})$

$\boldsymbol{\gamma}^{(\ell+1)} = \arg \max_{\boldsymbol{\gamma}} \pi(\boldsymbol{\gamma} | G^{(\ell+1)})$

end while

6.4.2. 2D image reconstruction with backprojection (2D IRB)

In our first approach using the JHBL method, which we denote as 2D image reconstruction with backprojection (2D IRB), we consider \mathbf{g}_p to be the unknown image and $\hat{\mathbf{g}}_p$ the corresponding PHD on the imaging plane with azimuth angle θ_p and elevation angle $\varphi = \frac{\pi}{2}$.

This technique for reconstruction begins by partitioning the data $\hat{\mathbf{g}}$ into $P/2$ slices,⁴ which we denote as the $z\theta_p$ -plane for each $p = 1, \dots, P/2$, so that

$$\hat{\mathbf{g}}_p = \mathcal{S}_2^3[\hat{\mathbf{g}}_{\mathcal{B}_p}], \quad (6.28)$$

where \mathcal{S}_2^3 is the slice operator defined in (6.4) and \mathcal{B}_p is the rotational change of basis

⁴We use $P/2$ slices instead of P slices since, in each $\hat{\mathbf{g}}_p$, the data along both azimuth θ_p and $\theta_{p+P/2} = \theta_p + \pi$ are included.

that rotates θ_p degrees about the z -axis followed by a rotation of $\frac{\pi}{2}$ radians about the x -axis. The set $\hat{G} = \{\hat{\mathbf{g}}_p\}_{p=1}^{P/2}$ is then used as input to Algorithm 11, where the 2D NUFT is utilized. Given that \mathbf{g}_1 and $\mathbf{g}_{P/2}$ are centered on the $z\theta_1$ -plane and $z\theta_{P/2}$ -plane, we define \mathbf{g}_0 and \mathbf{g}_{J+1} in Algorithm 11 to respectively be $\mathbf{g}_{P/2}$ and \mathbf{g}_1 reflected over the z -axis. This satisfies the expectation that $|\mathbf{g}_1|$ should be similar to a *mirrored* version of $|\mathbf{g}_{P/2}|$. Then, the volumetric image \mathbf{g} is formed from the output $G = \{\mathbf{g}_p\}_{p=1}^{P/2}$ as

$$\mathbf{g} = \sum_{p=1}^{P/2} (\mathcal{P}_3^2[|\mathbf{g}_p|])_{\mathcal{B}_p^{-1}}, \quad (6.29)$$

i.e. backprojecting each \mathbf{g}_p in physical 3D, performing the inverse rotation to the one that was originally performed on the data, and then summing the resulting back-projections to form the volumetric image. This reconstruction process relies on the assumption from Theorem 6.1 that each \mathbf{g}_p is a projection of the spatial domain onto the vertical plane intersecting the origin with normal in the azimuth direction θ_p . We note that this assumption is not accurate when the scatterers in the scene are not isotropic [107]. We are able to demonstrate, however, that even when anisotropic scatterers are found in the volumetric image, backprojecting the magnitude of the two-dimensional images \mathbf{g}_p , $p = 1, \dots, P/2$, still reconstructs the scene in a robust and predictable manner. Lastly, the resulting image \mathbf{g} can be thresholded according to a user-defined threshold value c_{IRB} to form a volumetric point cloud \mathbf{g}_c . This point cloud can then be used for 3D visualization and error estimation. It is important to note that the thresholding is done as a post-processing step, so multiple threshold values may be tested on the same volumetric image \mathbf{g} with relatively little additional computational cost. The 2D IRB method is summarized in Algorithm 12.

Remark 6.5. The 2D projections $\{\mathbf{g}_p\}_{j=1}^{P/2}$, obtained during the 2D IRB method in

Algorithm 12 2D Image Reconstruction with Backprojection (2D IRB)

Input SAR PHD $\hat{\mathbf{g}}$ ((6.3) for (6.8)) and threshold value c_{IRB} .

Output Volumetric reconstruction \mathbf{g} and point cloud \mathbf{g}_c .

Partition $\hat{\mathbf{g}}$ into $P/2$ slices $\hat{\mathbf{g}}_p$, $p = 1, \dots, P/2$ according to (6.28).

Obtain 2D projections $G = \{\mathbf{g}_p\}_{p=1}^{P/2}$ by using $\hat{G} = \{\hat{\mathbf{g}}_p\}_{p=1}^{P/2}$ as input to Algorithm 11.

Backproject \mathbf{g}_p according to (6.29) to form \mathbf{g} .

For each $i = 1, \dots, N$, set $\mathbf{g}_c^i = \begin{cases} 1 & \text{if } \mathbf{g}^i > c_{IRB} \\ 0 & \text{else} \end{cases}$.

Algorithm 12 can be used as input to other volumetric image recovery methods, see e.g. [67].

6.4.3. 3D subaperture reconstruction with composite imaging (3D SRCI)

Our second approach utilizing the JHBL method, which we call 3D subaperture reconstruction with composite imaging (3D SRCI), partitions the frequency domain data into J subapertures according to the sets Θ_j in (6.9). The sequential sparsity assumption holds for the J volumetric images \mathbf{g}_j formed using the partitions of the data $\hat{G} = \{\hat{\mathbf{g}}_j\}_{j=1}^J$.

With the data partitioned, we then perform Algorithm 11 using the 3D NUFT, which results in the collection of J volumetric images $G = \{\mathbf{g}_j\}_{j=1}^J$. Each of these 3D images describes the entire scene of interest, but the strength of anisotropic scatterers in each image is affected by the azimuth angles in Θ_j . Analogous to the 2D approach given in (6.11), we mitigate this issue by constructing a composite image [100, 107]

$$\mathbf{g}^i = \arg \max_j |\mathbf{g}_j^i| \quad (6.30)$$

for each pixel $i = 1, \dots, N$. As with the 2D IRB method in Algorithm 12, the final point cloud \mathbf{g}_c is formed by thresholding \mathbf{g} in (6.30), in this case to some threshold value c_{SRCI} . Multiple threshold values may again be tested for the same volumetric

image \mathbf{g} . The 3D SRCI process is summarized in Algorithm 13.

Algorithm 13 3D Subaperture Reconstruction with Composite Imaging (3D SRCI)

Input SAR PHD $\hat{\mathbf{g}}$ ((6.3) for (6.8)) and threshold value c_{SRCI} .

Output Volumetric reconstruction \mathbf{g} and point cloud \mathbf{g}_c .

Partition $\hat{\mathbf{g}}$ into J subapertures $\hat{\mathbf{g}}_j$, $j = 1, \dots, J$, according to the azimuthal sets Θ_j in (6.9).

Obtain 3D reconstruction $G = \{\mathbf{g}_j\}_{j=1}^J$ by using $\hat{G} = \{\hat{\mathbf{g}}_j\}_{j=1}^J$ as input to Algorithm 11.

Perform the composite imaging step according to (6.30) to form \mathbf{g} .

For each $i = 1, \dots, N$, set $\mathbf{g}_c^i = \begin{cases} 1 & \text{if } \mathbf{g}^i > c_{SRCI} \\ 0 & \text{else} \end{cases}$.

A couple of comments are in order:

- In addition to a point estimate for each \mathbf{g}_j , $j = 1, \dots, J$, our method recovers a MAP estimate for the precisions of the likelihood and both priors. This information may be used in lieu of (6.30) to form the final volumetric image, and will be considered in future investigations. In this regard, the technique developed to despeckle SAR from composite sub-aperture data in [36] may be useful.
- Following the discussion regarding composite imaging (6.30) used in the 3D SRCI (Algorithm 13), one might wonder why composite imaging is not incorporated into the 2D IRB (Algorithm 12). This is because the set of images $\{\mathbf{g}_p\}_1^{P/2}$ reconstructed using JHBL (Algorithm 11) are *projections* of the volumetric image onto different imaging planes. Thus if \mathbf{g} were to be calculated via composite imaging, i.e.

$$\mathbf{g}^i = \arg \max_p \left[(\mathcal{P}_3^2[\|\mathbf{g}_p\|])_{\mathcal{B}_p^{-1}} \right]^i, \quad i = 1, \dots, N,$$

the volumetric image would be incorrectly comprised of high-valued intensity

streaks rather than the properly formed local regions of high-valued intensity resulting from (6.29).

Section 6.5

Numerical experiments

We now provide some numerical examples to evaluate our methods. While there is no consensus on which error metric best captures the efficacy of 3D SAR imaging, there is precedent for using the modified Hausdorff distance (MHD) [44] for comparable 2D and 3D SAR image formation and data fusion [69, 85, 113, 139]. Hence we use that here and note that other error metrics may also be useful and provide better comparisons at various noise levels.

The MHD between two point clouds \mathcal{S} and \mathcal{T} is given by

$$\text{MHD}(\mathcal{S}, \mathcal{T}) = \max(d(\mathcal{S}, \mathcal{T}), d(\mathcal{T}, \mathcal{S})), \quad (6.31)$$

where

$$d(\mathcal{S}, \mathcal{T}) = \frac{1}{N_s} \sum_{s \in \mathcal{S}} d(s, \mathcal{T}), \quad d(s, \mathcal{T}) = \min_{t \in \mathcal{T}} \|s - t\|_2,$$

and N_s and N_t represent the number of points in \mathcal{S} and \mathcal{T} , respectively. When thresholding the reconstructed images, for ease of interpretability we choose to use a dB scale, where dB is the *decibel* unit of measurement. In this scaling the dB-scaled values of a given vector \mathbf{f} are

$$\mathbf{f}_{\text{dB}} = 20 \log_{10} \left(\frac{1}{\|\mathbf{f}\|_{\infty}} \mathbf{f} \right). \quad (6.32)$$

We also restrict the domain according to the alias-free extents given in [113].

The rest of this section is organized as follows: Section 6.5.1 establishes the various parameters used for each numerical experiment. Sections 6.5.2 and 6.5.3 analyze the ideal case, respectively for synthetic and measured data, for which there is no added noise. In each experiment we calculate and compare the MHD based on various dB threshold values determined by (6.32). Section 6.5.4 considers the more realistic scenario where the PHD is affected by additive white Gaussian noise. We compare results using the 2D IRB and 3D SRCI techniques for different signal-to-noise ratio (SNR) levels. Section 6.5.5 evaluates both techniques using a sub-sampled data set with no additional noise, and we compare these results to those obtained in [113].

6.5.1. Selection of parameters

With respect to parameters, for our experiments using the 2D IRB approach (Algorithm 12) we have $N = 40401$, $M = 6266$, and $J = 1800$. We also use a volumetric imaging cube with 201 equispaced gridpoints in each dimension, and 1800 imaging planes. By contrast, for our experiments using the 3D SRCI approach (Algorithm 13) we have $N = 8120601$, $M = 313300$, and $J = 36$. In this case the data are partitioned according to 36 equally-sized azimuth sets, and the spatial 3D imaging cube contains 201 equispaced gridpoints in each dimension. The image reconstruction sizes N for both the 2D IRB and 3D SRCI methods are chosen based on computational feasibility, as are the subaperture sizes in the 3D SCRI approach. The parameters for both types of experiments are summarized in Table 6.2.

Recalling the discussion following (6.18) regarding the selection of hyperparameters, when using Algorithm 11, we set the hyperparameters to be

$$\eta_\alpha = 1.5, \quad \eta_\beta = \eta_\gamma = 0.5, \quad \text{and} \quad \nu_\alpha = \nu_\beta = \nu_\gamma = 10^{-3}.$$

Our experiments indicate that Algorithm 11 is robust to choice of hyperparameters.

We also set the maximum number of iterations, ℓ_{max} , and threshold that determines convergence, tol , in Algorithm 11 to be

$$\ell_{max} = 10 \quad \text{and} \quad tol = 10^{-5}.$$

We evaluate the utility of the JHBL approach in the noisy regime by also considering reconstructions formed using the MAP of (6.17), which we call the maximum-likelihood estimate (MLE), in lieu of the MAP of (6.23) that is calculated using Algorithm 11. That is, we compare using just the likelihood density function to using the full posterior density function. This allows us to analyze the effects of the intra- and inter-image priors, (6.20) and (6.22), in the presence of noise. When reconstructing the volumetric images in this way, we replace Algorithm 11 with the MLE in Algorithm 12 or Algorithm 13, in which case we refer to these modified techniques as MLE (in contrast to JHBL) methods.

Parameter	2D IRB (Algorithm 12)	3D SRCI (Algorithm 13)
Image Size	201×201	$201 \times 201 \times 201$
Data Size	$241 \times 13 \times 2$	$241 \times 13 \times 100$
Data Partitions	1800	36

Table 6.2: Sizes of the inputs and outputs of Algorithm 12 and Algorithm 13 for our numerical experiments.

6.5.2. Reconstruction from synthetic data

We first evaluate the 2D IRB and 3D SRCI techniques on the synthetically generated Fourier data (using the NUFT [11]) from a hollow cube centered at the origin. The cube is 15 cm in length, the sides of the cube are 1 cm thick, and the length of the imaging domain is 70 cm in each direction.⁵ To accurately approximate the integral transform and avoid the inverse crime, we generate the data with the cube centered

⁵Due to how the data are generated the units in this example are arbitrary and are chosen to remain comparable with the B747 data set.

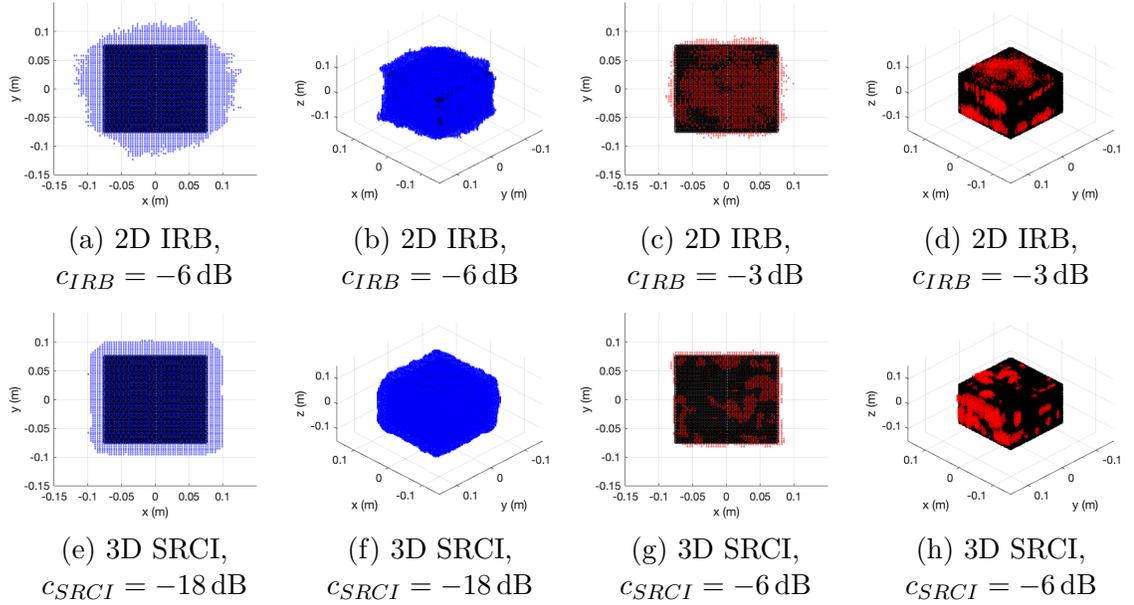


Figure 6.3: Different views at various dB thresholds of the 3D reconstruction of the synthetic cube data set in the ideal case; ground truth point cloud is displayed in black. The threshold values chosen to best demonstrate reconstruction quality.

on a voxelized grid with grid size $301 \times 301 \times 301$. All other parameters are consistent with those in Table 6.1.

Figure 6.3 displays the 2D IRB and 3D SRCI reconstructions of the cube in (a)-(d) and (e)-(h), respectively. We see from (a)-(b) and (e)-(f) of Figure 6.3 that the reconstruction from the 3D SRCI technique (Algorithm 13) tends to more sharply define edges of the cube, while the 2D IRB method (Algorithm 12) neither captures the edges nor the corners of the cube as well thus supporting the hypothesis that composite imaging for anisotropic scatterers yields better resolution. This is also observed in Figure 6.4 (a)-(b), which shows centered cross-sections of the 2D IRB and 3D SRCI reconstructions of the cube before any thresholding occurs. From (c) of Figure 6.4, it also appears that the 3D SRCI technique results both in a lower global MHD value as well as a wider dynamic range in this ideal case of no added noise.

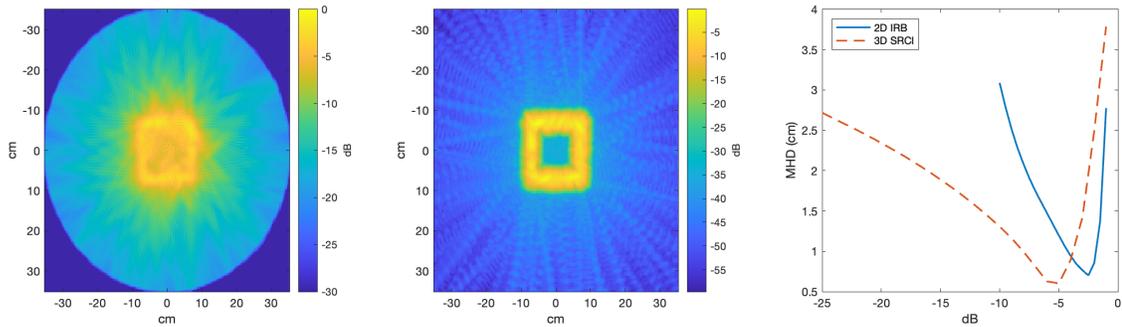


Figure 6.4: Centered cross-sections of the (left) 2D IRB and (middle) 3D SRCI reconstructions of the cube data set with no additional noise and no thresholding. (right) MHD values at various c_{IRB} or c_{SRCI} threshold values (dB) when either technique is used on the cube data set with no additional noise; the minimum MHD value calculated for the 2D IRB method is 0.7017cm, and for the 3D SRCI, the minimum MHD is 0.6014cm.

6.5.3. Reconstructions from measured data

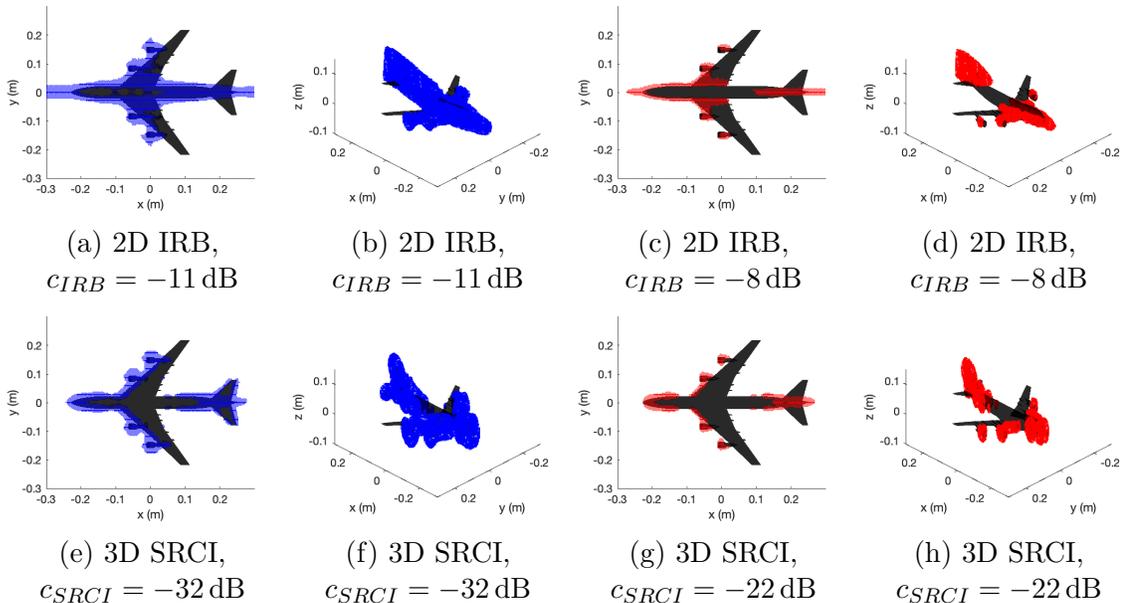


Figure 6.5: Different views at various dB thresholds of the 3D reconstruction of the B747 data set with no added noise; ground truth CAD model is displayed in black.

Figure 6.5 displays reconstructions of the B747 with no additional noise using the 2D IRB and 3D SRCI techniques. Observe in the 2D IRB (Algorithm 12) reconstructions (a)-(d) that the tail of the plane is poorly resolved. The tail is particularly dif-

difficult to resolve without the benefit of composite imaging because it is an anisotropic scatterer, and the signal does not persist across all viewing angles. Consistent with what we already observed in the synthetic data case, Figure 6.5(e)-(h) show that using the 3D SRCI (Algorithm 13) approach significantly helps to mitigate this issue.

Figure 6.5 shows that the both the 2D IRB and 3D SRCI techniques can identify and isolate persistent scatterers in the original scene, such as the engines of the plane, while the 3D SRCI is able to detect weaker scatterers, such as horizontal stabilizers in (e)-(f). Figure 6.6 displays cross sections of the reconstructed B747 data set using both the (a) 2D IRB and (b) 3D SRCI techniques, where the structure of the wings and the horizontal stabilizers are much clearer (b) as opposed to (a). Note that the sidelobe artifacts result from the use of subapertures but are tempered in the 3D SRCI solution by the use of the composite imaging step (6.30). Observe in Figure 6.6(c) that, as in Figure 6.4, the 3D SRCI approach achieves a smaller global MHD value as well as a larger dynamic range.

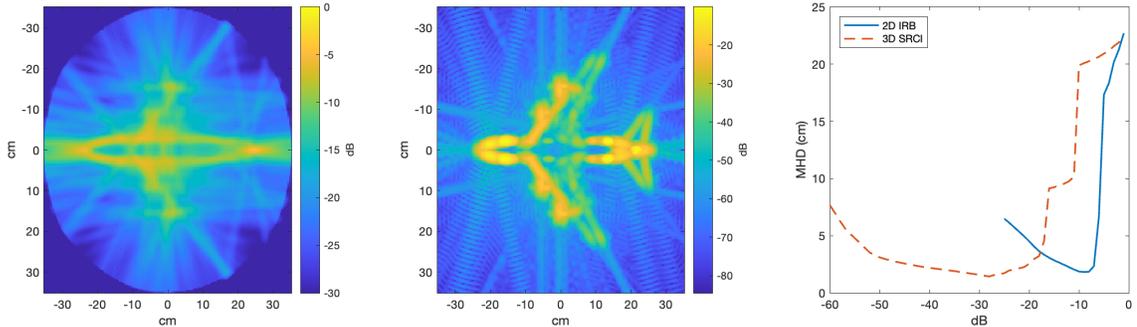


Figure 6.6: Centered cross-sections of the (left) 2D IRB and (middle) 3D SRCI reconstructions of the B747 data set with no additional noise and no thresholding. (right) MHD values at various c_{IRB} or c_{SRCI} threshold values (dB) when either technique is used on the B747 data set with no additional noise; the minimum MHD value calculated for the 2D IRB method is 1.832cm, and for the 3D SRCI, the minimum MHD is 1.440cm.

6.5.4. Noise study

For given SAR PHD with additive complex-valued circularly symmetric Gaussian noise, we calculate the signal-to-noise ratio (SNR) as

$$SNR = 10 \log_{10} \left(\frac{\mu^2}{\sigma^2} \right) \text{ dB},$$

where μ is the mean of the magnitude of the data, and σ is the noise standard deviation. Note that the SNR is also written in terms of decibels, and this is not to be confused with the decibel scaling adopted in (6.32). We now evaluate our techniques in both high (≈ 0 dB) and low (≈ -24 dB) SNR environments. This compares to the B747 data set used in [68, 113], where the noise was given as standard deviation values of 0.1 and 1, respectively. We evaluate the MHD at various threshold values for each noise value tested.

	2D IRB (Algorithm 12)		3D SRCI (Algorithm 13)	
	MLE	JHBL	MLE	JHBL
Cube, High SNR	0.7281	0.6774	0.5925	0.5222
Cube, Low SNR	0.6820	0.6408	0.5542	0.5570
B747, High SNR	1.868	2.150	1.354	1.401
B747, Low SNR	2.728	2.882	1.357	1.369

Table 6.3: Minimum MHD (cm) achieved across tested dB thresholds.

Table 6.3 displays the minimum MHD values achieved for the 2D IRB and 3D SRCI algorithms using either JHBL (Algorithm 11) or replacing it by the MLE. While comparable minimum MHD values are obtained regardless of whether the MLE or JHBL is used, the 3D SRCI method tends to outperform the 2D IRB technique at both noise levels. Figure 6.7 provides further insight. Here we observe that JHBL typically yields a larger dynamic range than MLE reconstruction does. A larger dynamic range is beneficial in applications such as automatic target recognition (ATR), since a wider distinction between target scatterers and background noise, and clear

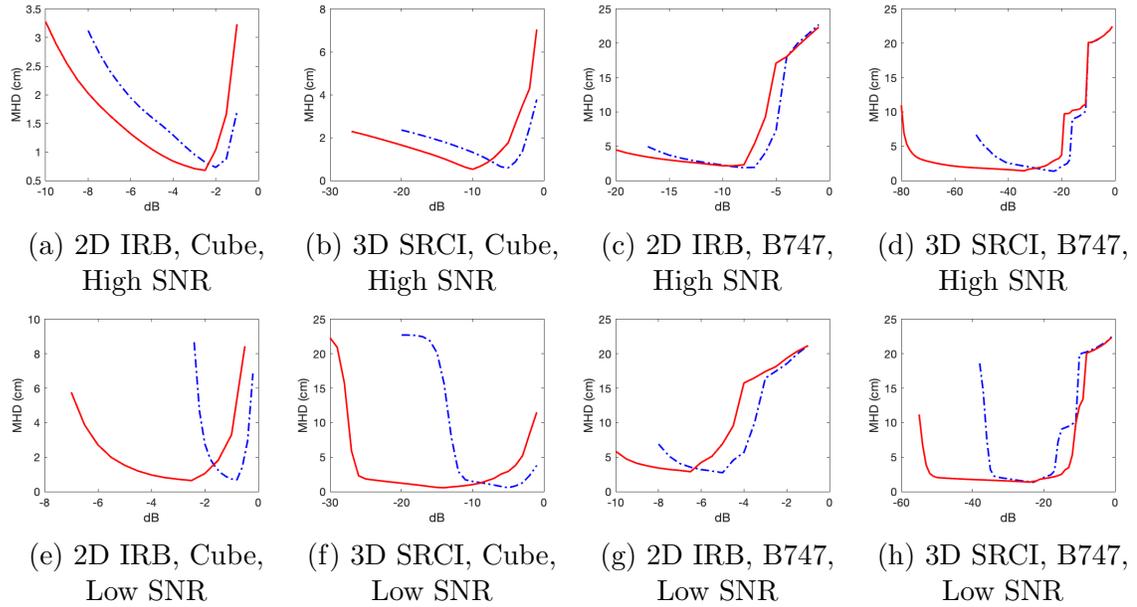


Figure 6.7: Threshold value vs. MHD for the cube (left) and B747 (right) data sets comparing the 2D IRB and 3D SRCI for both the JHBL and MLE approximations. (top) $\text{SNR} \approx 0$ dB; (bottom) $\text{SNR} \approx -24$ dB. In all plots, the dashed blue lines are the MLE MHD values, while the solid red lines are the JHBL MHD values. In all cases, it is straightforward to infer the rest of the characterization of the MHD values by continuing the trends in (a)-(h).

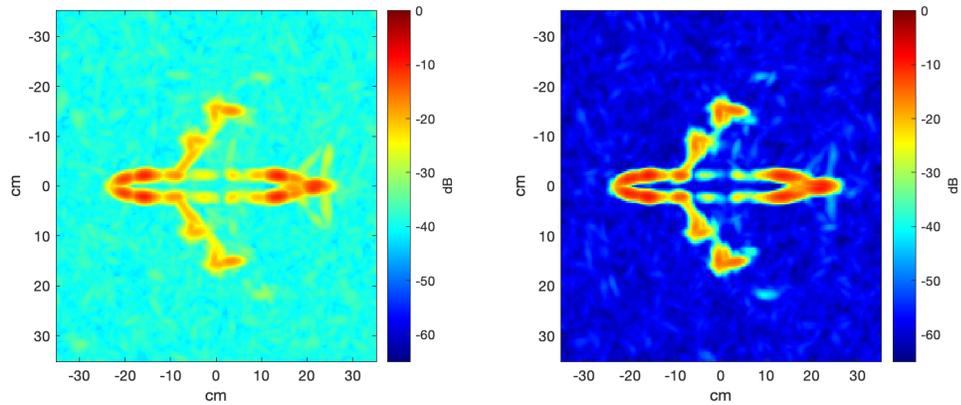


Figure 6.8: Slices of the (left) MLE and (right) JHBL reconstructions of the B747 with SNR of -30 dB using the 3D SRCI approach.

separation between distinct scatterers, may lead more accurate target classification [76, 98, 140]. Figure 6.8 shows a cross section of the MLE and JHBL reconstructions of the B747 at low SNR, highlighting this distinction. Observe that the background

in the MLE image is approximately -35 dB while the background in the 3D SRCI image is approximately -60 dB.

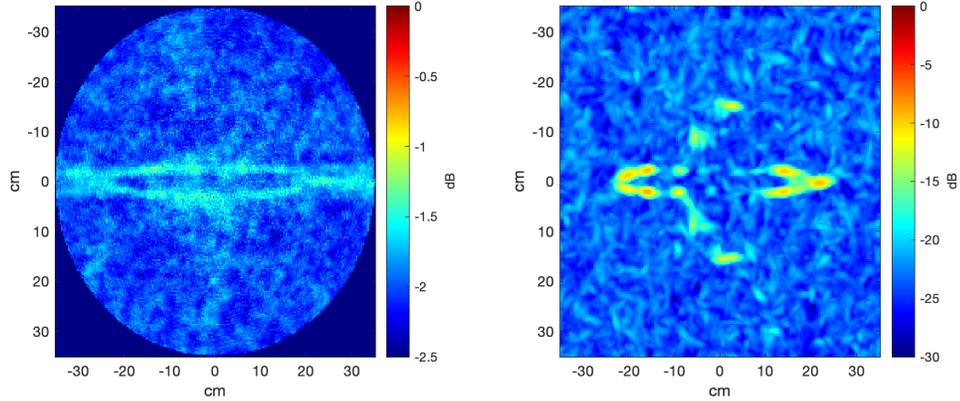


Figure 6.9: Slices of the (left) 2D IRB and (right) 3D SRCI reconstructions of the B747 with SNR of approximately -34 dB. Note that for interpretability, the threshold dB scale is different for each figure.

Using the 3D SRCI approach, we are able to maintain data fidelity at even lower SNR values. Figure 6.9 shows slices of the 2D IRB and 3D SRCI reconstructions (both using JHBL) of the B747 with an SNR of approximately -34 dB. Observe that while the outline of the hull of the plane is barely visible in the 2D IRB reconstruction, the front and back of the plane as well as the engines are clearly visible when using the 3D SRCI approach.

6.5.5. Sub-sampled data

As our final experiment we consider the case where the acquired data are noise-free but sub-sampled. In particular, data are given in a smaller bandwidth with lower frequency and azimuth sampling rates (see Table 6.4). The data sizes and partitions used in the sub-sampled case are adjusted accordingly (as well as for computational feasibility), with the resulting parameters displayed in Table 6.5. The same numerical experiment was performed in [113] using (6.14) (also see footnote there), where a minimum MHD value of 3.78cm was reported.

Parameter	Sub-sampled Value
Elevation Range	$[-3^\circ, 3^\circ]$
Elevation Sampling	0.5°
Frequency Range	$[31, 35]$ GHz
Frequency Sampling	150MHz
Bandwidth	4GHz
Center Frequency	33GHz
Azimuth Range	$[0^\circ, 359.9^\circ]$
Azimuth Sampling	0.3°

Table 6.4: Parameters of sub-sampled data set used for experimentation.

Parameter	2D IRB (Algorithm 12)	3D SRCI (Algorithm 13)
Image Size	201×201	$201 \times 201 \times 201$
Data Size	$27 \times 13 \times 2$	$27 \times 13 \times 30$
Data Partitions	600	40

Table 6.5: Sizes of the parameter inputs and outputs for Algorithm 11 for the sub-sampled data experiments.

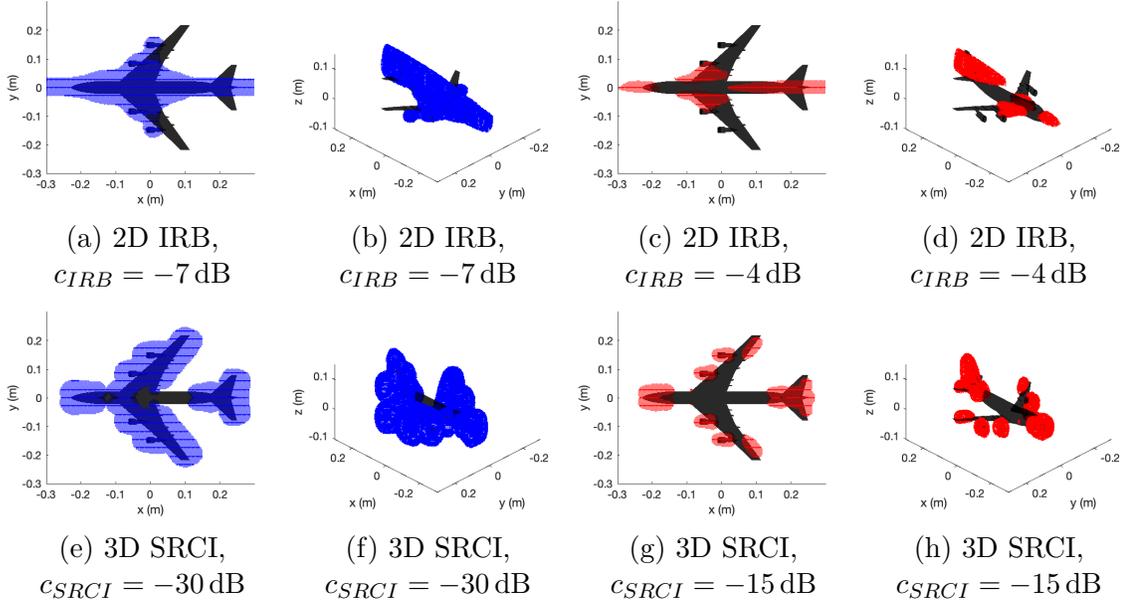


Figure 6.10: Different views at various dB thresholds of the 3D reconstruction of the B747 sub-sampled data set using our reconstruction techniques with no additional noise added; ground truth CAD model is displayed in black.

Figure 6.10 shows the results of using the 2D IRB and 3D SRCI techniques in the sub-sampled data case. Both the 2D IRB and 3D SRCI techniques are still able to

recover key features of the B747. The 3D SRCI method in particular is still able to separate scattering from the nose, tail, and engines when the threshold is set high enough. As expected, we observe that sub-sampling causes a loss of fine feature information in the reconstruction, such as the tail in (c)-(d) and the nose and engines in (e)-(h).

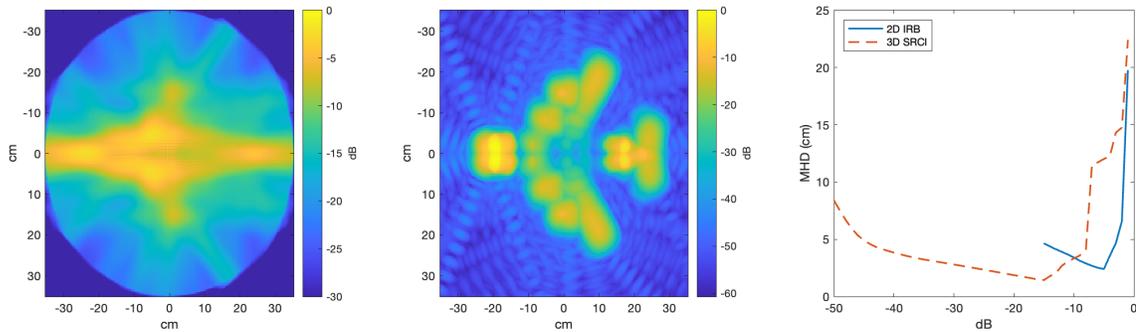


Figure 6.11: Cross-sections of the (left) 2D IRB and (middle) 3D SRCI reconstructions of the sub-sampled B747 data set using the parameters in Tables 6.4 and 6.5. (right) MHD values at various dB threshold values.

Figure 6.11 shows cross-sections of reconstructions of the B747 using the sub-sampled data set (no additive noise), as well as MHD values plotted at different dB thresholds. The minimum MHD value calculated in the 2D IRB case is 2.405cm, while the minimum MHD using the 3D SRCI technique is 1.442cm. Thus we see that in the MHD metric, the methods developed in this investigation outperform the technique in [113] in this sub-sampling experiment. Indeed, it does not appear that the MHD values differ significantly in the sub-sampling case.

Section 6.6

Conclusion

This investigation develops a new 3D SAR imaging technique that leverages joint sparsity using hierarchical Bayesian modeling. The method has the advantage of both enabling the learning of hyper-parameters as well as efficient composite recon-

struction. By using standard SAR imaging assumptions and employing appropriate conjugate priors, we are able to build a posterior from which we can analytically derive modes of the conditional distributions.

The 2D IRB and 3D SRCI techniques both enable high-fidelity reconstructions of synthetic and measured data in noisy environments, while also yielding a higher dynamic range when compared to methods that employ an MLE estimate, that is, those that do *not* leverage the sequential joint sparsity. The difference in computational complexity between the two approaches depends heavily on the NUFT implementation used, since the 3D SRCI technique requires a much larger transform than the 2D IRB method, as well as the value of P in (6.29) that is used with 2D IRB, as the backprojection step of Algorithm 12 involves rotating a 3D image and interpolating the result to uniform grid points. While both approaches perform well in low SNR environments, the 3D SRCI technique yields a lower MHD when compared with the 2D IRB approach at all noise levels tested. The 3D SRCI method is also able to maintain data fidelity at a lower SNR, and it tends to produce a larger dynamic range when compared with the 2D IRB method. We hypothesize that this is mainly because the composite imaging helps to further mitigate the effects of the faulty assumption regarding isotropic scatterers beyond using sequential joint sparsity. Finally, we evaluated our new techniques in the sub-sampled regime and again found that both methods outperform previously designed algorithms.

Future work will focus on evaluating our approach for various levels of under-sampling in the azimuth, elevation, and frequency domains. Additional insight may be gained from the MAP estimates for α , β , and γ , and these may prove useful in the post-processing used to form the final volumetric image. We will also leverage the Bayesian nature of our method to quantify the uncertainty we hold in our reconstruction and explore how this information can be used in the volumetric reconstruction

process as well as downstream processing tasks, such as coherent change detection and interferometry.

Lastly, we note that the minimum MHD achieved is sometimes smaller for low values of SNR than for high values in Table 6.3. This indicates that the MHD metric may not be best suited for comparison purposes using various noise levels. Other metrics that may provide more insight include the mean squared error and the structural similarity index measure.

Section 6.7

Derivation of the JHBL updates

Here we derive several of the update steps given by (6.24), (6.25), (6.26), and (6.27) that are used in Algorithm 11. Since the updates for α , β , and γ are similarly obtained, we only include the derivation for the α update step.

6.7.1. The G update

Let $G = \{\mathbf{g}_j\}_{j=1}^J$ in Algorithm 11. To make parallelization possible, we choose each \mathbf{g}_j update to depend on the \mathbf{g}_{-j} vectors from the previous update step,⁶

$$\mathbf{g}_j^{(\ell+1)} = \arg \max_{\mathbf{g}_j} \pi \left(\mathbf{g}_j | \mathbf{g}_{-j}^{(\ell)}, \boldsymbol{\alpha}^{(\ell)}, \boldsymbol{\beta}^{(\ell)}, \boldsymbol{\gamma}^{(\ell)} \right), \quad j = 1, \dots, J. \quad (6.33)$$

⁶The negative subscript refers to all similarly-named variables with subscripts other than the one indicated, in this case $\mathbf{g}_{-j} = \{\mathbf{g}_i : i = 1, \dots, J, i \neq j\}$.

From (6.23) the probability density function in (6.33) yields

$$\begin{aligned}
\pi(\mathbf{g}_j | \mathbf{g}_{-j}^{(\ell)}, \boldsymbol{\alpha}^{(\ell)}, \boldsymbol{\beta}^{(\ell)}, \boldsymbol{\gamma}^{(\ell)}) &\propto \exp\left\{-\alpha_j^{(\ell)} \|\mathcal{F}_j \mathbf{g}_j - \hat{\mathbf{g}}_j\|_2^2\right\} \times \left(\prod_{i=1}^N \exp\left\{-\beta_{j,i}^{(\ell)} |g_{j,i}|^2\right\}\right) \times \\
&\quad \left(\prod_{i=1}^N \exp\left\{-\gamma_{j,i}^{(\ell)} |\Psi_{j-1,i}^{(\ell)H} g_{j-1,i} - \Psi_{j,i}^{(\ell)H} g_{j,i}|^2\right\}\right) \times \\
&\quad \left(\prod_{i=1}^N \exp\left\{-\gamma_{j+1,i}^{(\ell)} |\Psi_{j,i}^{(\ell)H} g_{j,i} - \Psi_{j+1,i}^{(\ell)H} g_{j+1,i}|^2\right\}\right) \\
&\propto \exp\left(-\mathbf{g}_j^H \Sigma \mathbf{g}_j + \alpha_j^{(\ell)} \hat{\mathbf{g}}_j^H \mathcal{F}_j \mathbf{g}_j + \alpha_j^{(\ell)} \mathbf{g}_j^H \mathcal{F}_j^H \hat{\mathbf{g}}_j\right. \\
&\quad \left.+ \mathbf{g}_j^H \Psi_j^{(\ell)} D\left(\boldsymbol{\gamma}_j^{(\ell)}\right) \Psi_{j-1}^{(\ell)H} \mathbf{g}_{j-1}^{(\ell)} + \mathbf{g}_{j-1}^{(\ell)H} \Psi_{j-1}^{(\ell)} D\left(\boldsymbol{\gamma}_j^{(\ell)}\right) \Psi_j^{(\ell)H} \mathbf{g}_j\right. \\
&\quad \left.+ \mathbf{g}_j^H \Psi_j^{(\ell)} D\left(\boldsymbol{\gamma}_{j+1}^{(\ell)}\right) \Psi_{j+1}^{(\ell)H} \mathbf{g}_{j+1}^{(\ell)} + \mathbf{g}_{j+1}^{(\ell)H} \Psi_{j+1}^{(\ell)} D\left(\boldsymbol{\gamma}_{j+1}^{(\ell)}\right) \Psi_j^{(\ell)H} \mathbf{g}_j\right) \\
&\propto \exp((\mathbf{g}_j - \bar{\mathbf{g}})^H \Sigma (\mathbf{g}_j - \bar{\mathbf{g}})),
\end{aligned}$$

where $D(\cdot) = \text{diag}(\cdot)$ and

$$\begin{aligned}
\Sigma &= \alpha_j^{(\ell)} \mathcal{F}_j \mathcal{F}_j + D\left(\boldsymbol{\beta}_j^{(\ell)}\right) + D\left(\boldsymbol{\gamma}_j^{(\ell)}\right) + D\left(\boldsymbol{\gamma}_{j+1}^{(\ell)}\right), \\
\bar{\mathbf{g}} &= \Sigma^{-1} \left(\alpha_j^{(\ell)} \mathcal{F}_j^H \hat{\mathbf{g}}_j + \Psi_j^{(\ell)} D\left(\boldsymbol{\gamma}_j^{(\ell)}\right) \Psi_{j-1}^{(\ell)H} \mathbf{g}_{j-1}^{(\ell)} + \Psi_j^{(\ell)} D\left(\boldsymbol{\gamma}_{j+1}^{(\ell)}\right) \Psi_{j+1}^{(\ell)H} \mathbf{g}_{j+1}^{(\ell)}\right).
\end{aligned}$$

Equivalently, we have

$$\pi\left(\mathbf{g}_j | \mathbf{g}_{-j}^{(\ell)}, \boldsymbol{\alpha}^{(\ell)}, \boldsymbol{\beta}^{(\ell)}, \boldsymbol{\gamma}^{(\ell)}\right) \sim \mathbb{CN}(\bar{\mathbf{g}}, \Sigma^{-1}), \tag{6.34}$$

and since (6.34) is complex Gaussian, the mode is also given by $\bar{\mathbf{g}}$.

6.7.2. The α , β , and γ updates

Since α is only present in (6.17) and (6.18) and each α_j is conditionally independent from α_{-j} , for the α update in (6.25) we have

$$\pi(\alpha | G^{(\ell+1)}, \beta^{(\ell)}, \gamma^{(\ell)}) \propto \prod_{j=1}^J \pi(\alpha_j | \mathbf{g}_j^{(\ell+1)}),$$

which is maximized by maximizing each $\pi(\alpha_j | \mathbf{g}_j^{(\ell+1)})$, $j = 1, \dots, J$. In this case, from (6.23) we have

$$\pi(\alpha_j | \mathbf{g}_j^{(\ell+1)}) \propto \alpha_j^M \exp\left(-\alpha_j \left\| \mathcal{F}_j \mathbf{g}_j^{(\ell+1)} - \hat{\mathbf{g}}_j \right\|_2^2\right) \alpha_j^{\eta_\alpha - 1} \exp(-\nu_\alpha \alpha_j). \quad (6.35)$$

Since the distribution in (6.35) is continuous, the maximum is obtained simply by solving $\frac{d}{d\alpha_j} \pi(\alpha_j | \mathbf{g}_j^{(\ell+1)}) = 0$. In this regard, observe that

$$\begin{aligned} \frac{d}{d\alpha_j} \pi(\alpha_j | \mathbf{g}_j^{(\ell+1)}) &\propto \left(\frac{M + \eta_\alpha - 1}{\alpha_j} - \left\| \mathcal{F}_j \mathbf{g}_j^{(\ell+1)} - \hat{\mathbf{g}}_j \right\|_2^2 - \nu_\alpha \right) \\ &\times \alpha_j^{M + \eta_\alpha - 1} \exp\left(-\alpha_j \left(\left\| \mathcal{F}_j \mathbf{g}_j^{(\ell+1)} - \hat{\mathbf{g}}_j \right\|_2^2 + \nu_\alpha \right)\right). \end{aligned}$$

Setting the right hand side to zero yields

$$\alpha_j = \frac{\eta_\alpha + M - 1}{\nu_\alpha + \left\| \mathcal{F}_j \mathbf{g}_j^{(\ell+1)} - \hat{\mathbf{g}}_j \right\|_2^2}.$$

The β and γ updates in (6.26) and (6.27) are similarly derived.

Chapter 7

Concluding Remarks

The goal of this dissertation was to explore and develop new methods of solving computational and statistical inverse problems with applications in medical imaging and remote sensing. Four distinct approaches were considered, each of which relies on some *a priori* knowledge regarding the unknown signal of interest. While a classical regularization technique was implemented in Chapter 3, Chapters 4, 5, and 6 all utilize inversion approaches rooted in Bayesian statistics. These methods enable uncertainty quantification in addition to generating point estimate solutions, and they allow for the incorporation of uncertainty regarding a practitioner's prior knowledge. Several preliminary ideas used in these methods were discussed in Chapter 2, such as classical regularization techniques, Bayesian probability fundamentals, and signal processing for synthetic aperture radar (SAR), giving a broader context to our work.

In Chapter 3, we detailed the mathematics behind a proposed volumetric photoacoustic imaging system. In this novel approach, FPE-based PA image detection is combined with compressed ultrafast photography to further improve the temporal resolution of the data. The associated inverse problem was then solved using an ℓ_1 -regularization approach. Several of the statistical techniques explored in later chapters, particularly the joint sparsity approach detailed in Chapter 6, offer alternate

ways of solving for the initial pressure distribution (IPD), i.e. the unknown of interest. A statistical approach would have the added benefit of recovering uncertainty information regarding the IPD as well.

Chapters 4 and 5 discussed new ways of applying statistical inversion techniques to complex-valued signals. The focus of these chapters was developing techniques that recovered not only magnitude information but phase information as well. These problems are particularly relevant when using coherent imaging systems such as SAR, ultrasound, and digital holography. The techniques discussed in both chapters are used to enforce sparsity in some domain of the magnitude, with Chapter 4 focusing on the single measurement vector case with a scale mixture prior and Chapter 5 discussing the multiple measurement vector regime with a support-informed empirical prior. Both methods decompose the complex-valued signal of interest into its magnitude and phase components, enabling distinct prior distributions to be placed on each. This decomposition is novel to both works, and we prove that the resulting conditional distribution on the phase component is a von Mises distribution. This enables sampling of the complex-valued signal when an sparsity-promoting prior is placed *exclusively* on the magnitude, which is not possible when using previous techniques. To our knowledge, our work is also the first to extend the real-valued Bayesian LASSO to consider any full-rank sparsifying linear operator.

While much care is placed on the derivation of the magnitude prior, a simple uniform prior is used on the phase components to reflect no *a priori* information regarding the phase. Future investigations will explore the use of structural priors on the phase in addition to the magnitude. These sampling methods could also potentially be used as the first step in Bayesian interferometric or coherent change detection techniques, which would require further investigation.

Finally in Chapter 6, we implemented a joint hierarchical Bayesian learning ap-

proach to recover *maximum a posteriori* (MAP) estimates regarding a volumetric scene of interest from multi-pass SAR data. These MAP estimates were then used to form a composite image of the 3D scene. Our work is the first to apply the joint hierarchical Bayesian learning approach to 3D SAR imaging, and we do so using two distinct approaches based on backprojection and composite imaging. This approach also provided estimates for the covariance terms in the likelihood and joint prior, which may be used to further enhance the reconstruction. Notably, only data with uniform sampling in azimuth, elevation, and frequency were considered in our experiments. Future work will prioritize extending these ideas for use with nonuniformly-sampled data that may also have additional uncertainty in the data acquisition process.

Bibliography

- [1] *Capella space synthetic aperture radar (SAR) open dataset*, https://registry.opendata.aws/capella_opendata, Accessed: 2024-04-25.
- [2] Ben Adcock, Anne Gelb, Guohui Song, and Yi Sui, *Joint sparse recovery based on variances*, SIAM Journal on Scientific Computing **41** (2019), no. 1, A246–A268.
- [3] Esteban Aguilera, Matteo Nannini, and Andreas Reigber, *Wavelet-based compressed sensing for SAR tomography of forested areas*, IEEE transactions on geoscience and remote sensing **51** (2013), no. 12, 5283–5295.
- [4] Gibran H Aranda-Bojorges, Beatriz P Garcia-Salgado, Volodymyr I Ponomaryov, Osvaldo Lopez-García, and Rogelio Reyes-Reyes, *Parallel implementation of a despeckling in SAR images using gpu based segmentation and 3d-map estimation*, Real-Time Image Processing and Deep Learning 2022, vol. 12102, SPIE, 2022, pp. 181–192.
- [5] Rick Archibald, Anne Gelb, and Rodrigo B Platte, *Image reconstruction from undersampled Fourier data using the polynomial annihilation transform*, Journal of Scientific Computing **67** (2016), no. 2, 432–452.

- [6] Rick Archibald, Anne Gelb, and Jungho Yoon, *Polynomial fitting for edge detection in irregularly sampled signals and images*, SIAM Journal on Numerical Analysis **43** (2005), no. 1, 259–279.
- [7] Simon Arridge, Paul Beard, Marta Betcke, Ben Cox, Nam Huynh, Felix Lucka, Olumide Ogunlade, and Edward Zhang, *Accelerated high-resolution photoacoustic tomography via compressed sensing*, Physics in Medicine and Biology **61** (2016), no. 24, 8908–8940.
- [8] Christian D Austin, Emre Ertin, and Randolph L Moses, *Sparse multipass 3d SAR imaging: Applications to the gotcha data set*, Algorithms for Synthetic Aperture Radar Imagery XVI, vol. 7337, International Society for Optics and Photonics, 2009, p. 733703.
- [9] Sheldon Axler, *Linear algebra done right*, Springer Science & Business Media, 1997.
- [10] Jarred Barber, *A generalized likelihood ratio test for coherent change detection in polarimetric SAR*, IEEE Geoscience and Remote Sensing Letters **12** (2015), no. 9, 1873–1877.
- [11] Alexander H Barnett, Jeremy Magland, and Ludvig af Klinteberg, *A parallel nonuniform fast fourier transform library based on an “exponential of semicircle” kernel*, SIAM Journal on Scientific Computing **41** (2019), no. 5, C479–C504.
- [12] Paul C Beard, Frederic Perennes, and Tim N Mills, *Transduction mechanisms of the fabry-perot polymer film sensing concept for wideband ultrasound detection*, IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control **46** (1999), no. 6, 1575–1582.

- [13] DJ Best and Nicholas I Fisher, *Efficient simulation of the von Mises distribution*, Journal of the Royal Statistical Society: Series C (Applied Statistics) **28** (1979), no. 2, 152–157.
- [14] Marta M Betcke, Ben T Cox, Nam Huynh, Edward Z Zhang, Paul C Beard, and Simon R Arridge, *Acoustic wave field reconstruction from compressed measurements with application in photoacoustic tomography*, IEEE Transactions on Computational Imaging **3** (2017), no. 4, 710–721.
- [15] José M Bioucas-Dias and Mário AT Figueiredo, *A new twist: Two-step iterative shrinkage/thresholding algorithms for image restoration*, IEEE Transactions on Image processing **16** (2007), no. 12, 2992–3004.
- [16] Zdravko I Botev, Joseph F Grotowski, and Dirk P Kroese, *Kernel density estimation via diffusion*, (2010).
- [17] Stephen Boyd, Neal Parikh, Eric Chu, Borja Peleato, Jonathan Eckstein, et al., *Distributed optimization and statistical learning via the alternating direction method of multipliers*, Foundations and Trends® in Machine learning **3** (2011), no. 1, 1–122.
- [18] Steve Brooks, Andrew Gelman, Galin Jones, and Xiao-Li Meng, *Handbook of markov chain monte carlo*, CRC press, 2011.
- [19] Daniela Calvetti, Monica Pragliola, Erkki Somersalo, and Alexander Strang, *Sparse reconstructions from few noisy data: analysis of hierarchical Bayesian models with generalized gamma hyperpriors*, Inverse Problems **36** (2020), no. 2, 025010.

- [20] Daniela Calvetti and Erkki Somersalo, *Inverse problems: From regularization to bayesian inference*, Wiley Interdisciplinary Reviews: Computational Statistics **10** (2018), no. 3, e1427.
- [21] ———, *Bayesian scientific computing*, vol. 215, Springer Nature, 2023.
- [22] Daniela Calvetti, Erkki Somersalo, and A Strang, *Hierarchical bayesian models and sparsity: ℓ_2 -magic*, Inverse Problems **35** (2019), no. 3, 035003.
- [23] Emmanuel J Candès et al., *Compressive sampling*, Proceedings of the international congress of mathematicians, vol. 3, Madrid, Spain, 2006, pp. 1433–1452.
- [24] Emmanuel J Candès, Justin Romberg, and Terence Tao, *Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information*, IEEE Transactions on information theory **52** (2006), no. 2, 489–509.
- [25] Emmanuel J Candes, Justin K Romberg, and Terence Tao, *Stable signal recovery from incomplete and inaccurate measurements*, Communications on Pure and Applied Mathematics: A Journal Issued by the Courant Institute of Mathematical Sciences **59** (2006), no. 8, 1207–1223.
- [26] Emmanuel J Candes and Terence Tao, *Decoding by linear programming*, IEEE transactions on information theory **51** (2005), no. 12, 4203–4215.
- [27] Emmanuel J Candès and Michael B Wakin, *An introduction to compressive sampling*, IEEE signal processing magazine **25** (2008), no. 2, 21–30.
- [28] Emmanuel J Candes, Michael B Wakin, and Stephen P Boyd, *Enhancing sparsity by reweighted ℓ_1 minimization*, Journal of Fourier Analysis and Applications **14** (2008), no. 5, 877–905.

- [29] George Casella and Edward I George, *Explaining the Gibbs sampler*, *The American Statistician* **46** (1992), no. 3, 167–174.
- [30] Müjdat Cetin, William C Karl, and Alan S Willsky, *Feature-preserving regularization method for complex-valued inverse problems with application to coherent imaging*, *Optical Engineering* **45** (2006), no. 1, 017003–017003.
- [31] Leping Chen, Daoxiang An, Xiaotao Huang, and Zhimin Zhou, *A 3d reconstruction strategy of vehicle outline based on single-pass single-polarization CSAR data*, *IEEE Transactions on Image Processing* **26** (2017), no. 11, 5545–5554.
- [32] Margaret Cheney, *A mathematical tutorial on synthetic aperture radar*, *SIAM review* **43** (2001), no. 2, 301–312.
- [33] Margaret Cheney and Brett Borden, *Synthetic aperture radar imaging*, *Handbook of Mathematical Methods in Imaging*, 2015.
- [34] Victor Churchill, *Synthetic aperture radar image formation with uncertainty quantification*, Ph.D. thesis, Dartmouth College, 2020.
- [35] Victor Churchill and Anne Gelb, *Sampling-based spotlight sar image reconstruction from phase history data for speckle reduction and uncertainty quantification*, *SIAM/ASA Journal on Uncertainty Quantification* **10** (2022), no. 3, 1225–1249.
- [36] ———, *Sub-aperture sar imaging with uncertainty quantification*, *Inverse Problems* **39** (2023), no. 5, 054004.
- [37] Benjamin T Cox, Jan G Laufer, Paul C Beard, and Simon R Arridge, *Quantitative spectroscopic photoacoustic imaging: a review*, *Journal of biomedical optics* **17** (2012), no. 6, 061202.

- [38] Ronald E Crochiere and Lawrence R Rabiner, *Interpolation and decimation of digital signals—a tutorial review*, Proceedings of the IEEE **69** (1981), no. 3, 300–331.
- [39] Armin W Doerry, Edward E Bishop, and John A Miller, *Basics of backprojection algorithm for processing synthetic aperture radar images*, Sandia Report SAND2016-1682, Unlimited Release (2016).
- [40] Xiao Dong and Yunhua Zhang, *A map approach for 1-bit compressive sensing in synthetic aperture radar imaging*, IEEE Geoscience and Remote Sensing Letters **12** (2015), no. 6, 1237–1241.
- [41] David L Donoho, *Compressed sensing*, IEEE Transactions on information theory **52** (2006), no. 4, 1289–1306.
- [42] Andjela Draganic, Irena Orovic, and Srdjan Stankovic, *On some common compressive sensing recovery algorithms and applications-review paper*, arXiv preprint arXiv:1705.05216 (2017).
- [43] Huiping Duan, Lizao Zhang, Jun Fang, Lei Huang, and Hongbin Li, *Pattern-coupled sparse bayesian learning for inverse synthetic aperture radar imaging*, IEEE Signal Processing Letters **22** (2015), no. 11, 1995–1999.
- [44] M-P Dubuisson and Anil K Jain, *A modified hausdorff distance for object matching*, Proceedings of 12th international conference on pattern recognition, vol. 1, IEEE, 1994, pp. 566–568.
- [45] Gerald B Folland, *Real analysis: modern techniques and their applications*, vol. 40, John Wiley & Sons, 1999.

- [46] Liang Gao, Jinyang Liang, Chiye Li, and Lihong V Wang, *Single-shot compressed ultrafast photography at one hundred billion frames per second*, *Nature* **516** (2014), no. 7529, 74–77.
- [47] Anne Gelb and Theresa Scarnati, *Reducing effects of bad data using variance based joint sparsity recovery*, *Journal of Scientific Computing* **78** (2019), no. 1, 94–120.
- [48] Anne Gelb and Eitan Tadmor, *Detection of edges in spectral data*, *Applied and Computational Harmonic Analysis* **7** (1999), no. 1, 101–135.
- [49] ———, *Detection of edges in spectral data ii. nonlinear enhancement*, *SIAM Journal on Numerical Analysis* **38** (2000), no. 4, 1389–1408.
- [50] ———, *Adaptive edge detectors for piecewise smooth data based on the minmod limiter*, *Journal of Scientific Computing* **28** (2006), no. 2, 279–306.
- [51] Rudiger Gens and John L Van Genderen, *Review article SAR interferometry—issues, techniques, applications*, *International journal of remote sensing* **17** (1996), no. 10, 1803–1835.
- [52] Jan Glaubitz and Anne Gelb, *Leveraging joint sparsity in hierarchical Bayesian learning*, arXiv preprint arXiv:2303.16954 (2023).
- [53] Jan Glaubitz, Anne Gelb, and Guohui Song, *Generalized sparse bayesian learning and application to image reconstruction*, *SIAM/ASA Journal on Uncertainty Quantification* **11** (2023), no. 1, 262–284.
- [54] LeRoy A Gorham and Linda J Moore, *SAR image formation toolbox for matlab*, *Algorithms for Synthetic Aperture Radar Imagery XVII*, vol. 7699, SPIE, 2010, pp. 46–58.

- [55] Dylan Green, Anne Gelb, and Geoffrey P Luke, *Sparsity-based recovery of three-dimensional photoacoustic images from compressed single-shot optical detection*, Journal of Imaging **7** (2021), no. 10, 201.
- [56] Dylan Green, JR Jamora, and Anne Gelb, *Leveraging joint sparsity in 3d synthetic aperture radar imaging*, Applied Mathematics for Modern Challenges **1** (2023), no. 1, 61–86.
- [57] Dylan Green, Jonathan Lindbloom, and Anne Gelb, *Complex-valued signal recovery using the bayesian lasso*, 2024.
- [58] Zijian Guo, Changhui Li, Liang Song, and Lihong V Wang, *Compressed sensing in photoacoustic tomography in vivo*, Journal of biomedical optics **15** (2010), no. 2, 021311.
- [59] Markus Haltmeier, Michael Sandbichler, Thomas Berer, Johannes Bauer-Marschallinger, Peter Burgholzer, and Linh Nguyen, *A sparsification and reconstruction strategy for compressed sensing photoacoustic tomography*, The Journal of the Acoustical Society of America **143** (2018), no. 6, 3838–3848.
- [60] Per Christian Hansen, *Discrete inverse problems: Insight and algorithms*, SIAM, 2010.
- [61] Magnus R Hestenes, Eduard Stiefel, et al., *Methods of conjugate gradients for solving linear systems*, Journal of research of the National Bureau of Standards **49** (1952), no. 6, 409–436.
- [62] Nam Huynh, Felix Lucka, Edward Z Zhang, Marta M Betcke, Simon R Arridge, Paul C Beard, and Benjamin T Cox, *Single-pixel camera photoacoustic tomography*, Journal of Biomedical Optics **24** (2019), no. 12, 121907.

- [63] Nam Huynh, Edward Zhang, Marta Betcke, Simon Arridge, Paul Beard, and Ben Cox, *Single-pixel optical camera for video rate ultrasonic imaging*, *Optica* **3** (2016), no. 1, 26–29.
- [64] John Immerkaer, *Fast noise variance estimation*, *Computer vision and image understanding* **64** (1996), no. 2, 300–302.
- [65] Hemant Ishwaran and J Sunil Rao, *Spike and slab variable selection: frequentist and bayesian strategies*, (2005).
- [66] Charles VJ Jakowatz, Daniel E Wahl, Paul H Eichel, Dennis C Ghiglia, and Paul A Thompson, *Spotlight-mode synthetic aperture radar: a signal processing approach: a signal processing approach*, Springer Science & Business Media, 2012.
- [67] J. R. Jamora, Dylan Green, Ander Talley, and Thomas Curry, *Utilizing SAR imagery in three-dimensional neural radiance fields-based applications*, *Algorithms for Synthetic Aperture Radar Imagery XXX* (Edmund Zelnio and Frederick D. Garber, eds.), vol. 12520, International Society for Optics and Photonics, SPIE, 2023, p. 1252002.
- [68] Jan Rainer Jamora, *Angular-dependent three-dimensional imaging techniques in multi-pass synthetic aperture radar*, Master’s thesis, Mississippi State University, 2021.
- [69] JR Jamora, Paul Sotirelis, Adam Nolan, Jeff Walrath, Rick Hubbard, Rajith Weerasinghe, Eric Young, and Seth Young, *Multiple modality sensor fusion from synthetic aperture radar, lidar, and electro-optical systems using three-dimensional data representations*, *Algorithms for Synthetic Aperture Radar Imagery XXIX*, vol. 12095, SPIE, 2022, pp. 32–43.

- [70] Marko Järvenpää and Robert Piché, *Bayesian hierarchical model of total variation regularisation for image deblurring*, arXiv preprint arXiv:1412.4384 (2014).
- [71] Seungwan Jeon, Jongbeom Kim, Donghyun Lee, Jin Woo Baik, and Chulhong Kim, *Review on practical photoacoustic microscopy*, *Photoacoustics* **15** (2019), 100141.
- [72] Shihao Ji, Ya Xue, and Lawrence Carin, *Bayesian compressive sensing*, *IEEE Transactions on signal processing* **56** (2008), no. 6, 2346–2356.
- [73] Jari Kaipio and Erkki Somersalo, *Statistical and computational inverse problems*, vol. 160, Springer Science & Business Media, 2006.
- [74] Jari P Kaipio, Ville Kolehmainen, Erkki Somersalo, and Marko Vauhkonen, *Statistical inversion and Monte Carlo sampling methods in electrical impedance tomography*, *Inverse Problems* **16** (2000), no. 5, 1487.
- [75] Gitta Kutyniok, *Theory and applications of compressed sensing*, *GAMM-Mitteilungen* **36** (2013), no. 1, 79–101.
- [76] Tingli Li and Lan Du, *Target discrimination for SAR ATR based on scattering center feature and k-center one-class classification*, *IEEE Sensors Journal* **18** (2018), no. 6, 2453–2461.
- [77] Yang Li, Lei Li, Liren Zhu, Konstantin Maslov, Junhui Shi, Peng Hu, En Bo, Junjie Yao, Jinyang Liang, Lidai Wang, et al., *Snapshot photoacoustic topography through an ergodic relay for high-throughput imaging of optical absorption*, *Nature Photonics* **14** (2020), no. 3, 164–170.
- [78] Xianglei Liu, Jingdan Liu, Cheng Jiang, Fiorenzo Vetrone, and Jinyang Liang, *Single-shot compressed optical-streaking ultra-high-speed photography*, *Optics letters* **44** (2019), no. 6, 1387–1390.

- [79] Ignace Loris, Guust Nolet, Ingrid Daubechies, and FA Dahlen, *Tomographic inversion using ℓ_1 -norm regularization of wavelet coefficients*, Geophysical Journal International **170** (2007), no. 1, 359–370.
- [80] David Luengo, Luca Martino, Mónica Bugallo, Víctor Elvira, and Simo Särkkä, *A survey of monte carlo methods for parameter estimation*, EURASIP Journal on Advances in Signal Processing **2020** (2020), 1–62.
- [81] Michael Lustig, David Donoho, and John M Pauly, *Sparse MRI: The application of compressed sensing for rapid MR imaging*, Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine **58** (2007), no. 6, 1182–1195.
- [82] Hassan Maatouk and Xavier Bay, *A new rejection sampling method for truncated multivariate Gaussian random variables restricted to convex sets*, Monte Carlo and Quasi-Monte Carlo Methods: MCQMC, Leuven, Belgium, April 2014, Springer, 2016, pp. 521–530.
- [83] Ernest L Madsen, James A Zagzebski, Richard A Banjavie, and Ronald E Jutila, *Tissue mimicking materials for ultrasound phantoms*, Medical physics **5** (1978), no. 5, 391–394.
- [84] Srivalleesha Mallidi, Geoffrey P Luke, and Stanislav Emelianov, *Photoacoustic imaging in cancer detection, diagnosis, and treatment guidance*, Trends in biotechnology **29** (2011), no. 5, 213–221.
- [85] Haicen Mao, Qiuzhe Yu, and Tianxu Zhang, *Matching SAR image to optical image using modified hausdorff distance and genetic algorithms*, MIPPR 2007: Pattern Recognition and Computer Vision, vol. 6788, SPIE, 2007, pp. 532–537.

- [86] John R Michael, William R Schucany, and Roy W Haas, *Generating random variates using transformations with multiple roots*, The American Statistician **30** (1976), no. 2, 88–90.
- [87] Linda J. Moore, Brian D. Rigling, and Robert P. Penno, *Characterization of information in phase of radar range profiles*, 2014 48th Asilomar Conference on Signals, Systems and Computers, 2014, pp. 2027–2031.
- [88] Linda J Moore, Brian D Rigling, and Robert P Penno, *Characterization of phase information of synthetic aperture radar imagery*, IEEE Transactions on Aerospace and Electronic Systems **55** (2018), no. 2, 676–688.
- [89] Randolph L Moses, Lee C Potter, and Mujdat Cetin, *Wide-angle SAR imaging*, Algorithms for Synthetic Aperture Radar Imagery XI, vol. 5427, SPIE, 2004, pp. 164–175.
- [90] Dylan Muir, *Kernel density estimation for circular functions*, 2014.
- [91] David C Munson, James D O’Brien, and W Kenneth Jenkins, *A tomographic formulation of spotlight-mode synthetic aperture radar*, Proceedings of the IEEE **71** (1983), no. 8, 917–925.
- [92] Ren Ng, *Fourier slice photography*, ACM Siggraph 2005 Papers, 2005, pp. 735–744.
- [93] Jia-Cheng Ni, Qun Zhang, Ying Luo, and Li Sun, *Compressed sensing SAR imaging based on centralized sparse representation*, IEEE Sensors Journal **18** (2018), no. 12, 4920–4932.
- [94] Kang-Yu Ni and Shankar Rao, *SAR moving target imaging using sparse and low-rank decomposition*, SPIE Defense+ Security, International Society for Optics and Photonics, 2014, pp. 90771D–90771D.

- [95] Leslie M Novak, *Coherent change detection for multi-polarization SAR*, Conference Record of the Thirty-Ninth Asilomar Conference on Signals, Systems and Computers, 2005., IEEE, 2005, pp. 568–573.
- [96] Ari Pakman and Liam Paninski, *Exact Hamiltonian Monte Carlo for truncated multivariate Gaussians*, Journal of Computational and Graphical Statistics **23** (2014), no. 2, 518–542.
- [97] George Papandreou and Alan L Yuille, *Gaussian sampling by local perturbations*, Advances in Neural Information Processing Systems **23** (2010).
- [98] Jong-Il Park, Sang-Hong Park, and Kyung-Tae Kim, *New discrimination features for SAR automatic target recognition*, IEEE Geoscience and Remote Sensing Letters **10** (2012), no. 3, 476–480.
- [99] Trevor Park and George Casella, *The Bayesian Lasso*, Journal of the American Statistical Association **103** (2008), no. 482, 681–686.
- [100] Christopher R Paulson, *Utilizing glint phenomenology to perform classification of civilian vehicles using synthetic aperture radar*, Ph.D. thesis, University of Florida, 2013.
- [101] Allan D Pierce, *Acoustics: an introduction to its physical principles and applications*, 3 ed., Springer, 2019.
- [102] Lee C Potter, Emre Ertin, Jason T Parker, and Müjdat Cetin, *Sparsity and compressed sensing in radar imaging*, Proceedings of the IEEE **98** (2010), no. 6, 1006–1020.
- [103] Dalong Qi, Shian Zhang, Chengshuai Yang, Yilin He, Fengyan Cao, Jiali Yao, Pengpeng Ding, Liang Gao, Tianqing Jia, Jinyang Liang, et al., *Single-shot*

- compressed ultrafast photography: a review*, *Advanced Photonics* **2** (2020), no. 1, 014003.
- [104] Abhejit Rajagopal, Jason Hilton, David Boutte, Andrew P Brown, and Jan R Jamora, *Enhanced compressed sensing 3D SAR imaging via cross-modality EO-SAR joint-sparsity priors*, *Algorithms for Synthetic Aperture Radar Imagery XXX*, vol. 12520, SPIE, 2023, pp. 13–23.
- [105] Brian D Rigling and RANDOLPH L Moses, *Polar format algorithm for bistatic SAR*, *IEEE Transactions on Aerospace and Electronic Systems* **40** (2004), no. 4, 1147–1159.
- [106] Christian P Robert and George Casella, *Monte carlo statistical methods*, vol. 2, Springer, 1999.
- [107] Toby Sanders, Anne Gelb, and Rodrigo B Platte, *Composite SAR imaging using sequential joint sparsity*, *Journal of Computational Physics* **338** (2017), 357–370.
- [108] Toby Sanders, Rodrigo B Platte, and Robert D Skeel, *Effective new methods for automated parameter selection in regularized inverse problems*, *Applied Numerical Mathematics* **152** (2020), 29–48.
- [109] Theresa Scarnati, *Recent techniques for regularization in partial differential equations and imaging*, Arizona State University, 2018.
- [110] Theresa Scarnati and Anne Gelb, *Joint image formation and two-dimensional autofocus for synthetic aperture radar data*, *Journal of Computational Physics* **374** (2018), 803–821.

- [111] ———, *Variance based joint sparsity reconstruction of synthetic aperture radar data for speckle reduction*, Algorithms for Synthetic Aperture Radar Imagery XXV, vol. 10647, SPIE, 2018, pp. 199–212.
- [112] ———, *Accurate and efficient image reconstruction from multiple measurements of Fourier samples*, Journal of Computational Mathematics **38** (2020), no. 5, 797.
- [113] Theresa Scarnati and Jan Ranier Jamora, *Three-dimensional object reconstruction from sparse multi-pass SAR data*, Algorithms for Synthetic Aperture Radar Imagery XXVIII, vol. 11728, International Society for Optics and Photonics, 2021, p. 117280I.
- [114] Ruibo Shang, Richard Archibald, Anne Gelb, and Geoffrey P Luke, *Sparsity-based photoacoustic image reconstruction with a linear array transducer and direct measurement of the forward model*, Journal of biomedical optics **24** (2018), no. 3, 031015.
- [115] Lawrence A Shepp and Benjamin F Logan, *The Fourier reconstruction of a head section*, IEEE Transactions on Nuclear Science **21** (1974), no. 3, 21–43.
- [116] Yilei Shi, Xiao xiang Zhu, and Richard Bamler, *Nonlocal compressive sensing-based SAR tomography*, IEEE Transactions on Geoscience and Remote Sensing **57** (2019), no. 5, 3015–3024.
- [117] Ivana Stojanovic, Müjdat Çetin, and W Clem Karl, *Compressed sensing of monostatic and multistatic SAR*, IEEE Geoscience and Remote Sensing Letters **10** (2013), no. 6, 1444–1448.
- [118] Begoña Subiza, Encarna Gimeno-Nieves, Juan M Lopez-Sanchez, and Joaquim Fortuny-Guasch, *An approach to SAR imaging by means of non-uniform ffts*,

- IGARSS 2003. 2003 IEEE International Geoscience and Remote Sensing Symposium. Proceedings (IEEE Cat. No. 03CH37477), vol. 6, IEEE, 2003, pp. 4089–4091.
- [119] Robert Tibshirani, *Regression shrinkage and selection via the Lasso*, Journal of the Royal Statistical Society Series B: Statistical Methodology **58** (1996), no. 1, 267–288.
- [120] Michael E Tipping, *Sparse bayesian learning and the relevance vector machine*, Journal of machine learning research **1** (2001), no. Jun, 211–244.
- [121] Ridha Touzi, Armand Lopes, Jérôme Bruniquel, and Paris W Vachon, *Coherence estimation for SAR imagery*, IEEE transactions on geoscience and remote sensing **37** (1999), no. 1, 135–149.
- [122] Bradley E Treeby and Benjamin T Cox, *k-wave: Matlab toolbox for the simulation and reconstruction of photoacoustic wave fields*, Journal of biomedical optics **15** (2010), no. 2, 021314.
- [123] Adityavikram Viswanathan, Anne Gelb, and Douglas Cochran, *Iterative design of concentration factors for jump detection*, Journal of Scientific Computing **51** (2012), no. 3, 631–649.
- [124] Maxime Vono, Nicolas Dobigeon, and Pierre Chainais, *High-dimensional Gaussian sampling: a review and a unifying approach based on a stochastic proximal point algorithm*, SIAM Review **64** (2022), no. 1, 3–56.
- [125] Daniel E Wahl, David A Yocky, Charles V Jakowatz, and Katherine M Simonson, *A new maximum-likelihood change estimator for two-pass SAR coherent change detection*, IEEE Transactions on Geoscience and Remote Sensing **54** (2016), no. 4, 2460–2469.

- [126] Lihong V Wang and Song Hu, *Photoacoustic tomography: in vivo imaging from organelles to organs*, *science* **335** (2012), no. 6075, 1458–1462.
- [127] Lihong V Wang and Junjie Yao, *A practical guide to photoacoustic tomography in the life sciences*, *Nature methods* **13** (2016), no. 8, 627–638.
- [128] Peng Wang, Jinyang Liang, and Lihong V Wang, *Single-shot ultrafast imaging attaining 70 trillion frames per second*, *Nature communications* **11** (2020), no. 1, 1–9.
- [129] Xueding Wang, Yongjiang Pang, Geng Ku, Xueyi Xie, George Stoica, and Lihong V Wang, *Noninvasive laser-induced photoacoustic tomography for structural and functional in vivo imaging of the brain*, *Nature biotechnology* **21** (2003), no. 7, 803–806.
- [130] Zhou Wang, Eero P Simoncelli, and Alan C Bovik, *Multiscale structural similarity for image quality assessment*, *The Thirty-Seventh Asilomar Conference on Signals, Systems & Computers*, 2003, vol. 2, Ieee, 2003, pp. 1398–1402.
- [131] David P Wipf and Bhaskar D Rao, *Sparse Bayesian learning for basis selection*, *IEEE Transactions on Signal processing* **52** (2004), no. 8, 2153–2164.
- [132] Yao Xiao and Jan Glaubitz, *Sequential image recovery using joint hierarchical bayesian learning*, *arXiv preprint arXiv:2206.12745* **91** (2022), no. 3, 79.
- [133] Minghua Xu and Lihong V Wang, *Photoacoustic imaging in biomedicine*, *Review of scientific instruments* **77** (2006), no. 4, 041101.
- [134] Leonid Pinkhusovich Yaroslavskii and Nikolai S Merzlyakov, *Methods of digital holography*, Springer, 1980.

- [135] Juha Tapani Ylitalo and Helmut Ermert, *Ultrasound synthetic aperture imaging: monostatic approach*, IEEE transactions on ultrasonics, ferroelectrics, and frequency control **41** (1994), no. 3, 333–339.
- [136] Siwei Yu, A Shaharyar Khwaja, and Jianwei Ma, *Compressed sensing of complex-valued data*, Signal Processing **92** (2012), no. 2, 357–362.
- [137] Edward Zhang, Jan Laufer, and Paul Beard, *Backward-mode multiwavelength photoacoustic scanner using a planar fabry-perot polymer film ultrasound sensor for high-resolution three-dimensional imaging of biological tissues*, Applied optics **47** (2008), no. 4, 561–577.
- [138] Jiahui Zhang, Anne Gelb, and Theresa Scarnati, *Empirical Bayesian inference using a support informed prior*, SIAM/ASA Journal on Uncertainty Quantification **10** (2022), no. 2, 745–774.
- [139] Zhiguang Zhang, Hong Lei, and Zhifeng Lv, *Vehicle layover removal in circular SAR images via rosl*, IEEE Geoscience and Remote Sensing Letters **12** (2015), no. 12, 2413–2417.
- [140] Qun Zhao and Jose C Principe, *Support vector machines for SAR automatic target recognition*, IEEE Transactions on Aerospace and Electronic Systems **37** (2001), no. 2, 643–654.