SYNTHETIC APERTURE RADAR IMAGE FORMATION WITH UNCERTAINTY QUANTIFICATION

A Thesis

Submitted to the Faculty

in partial fulfillment of the requirements for the

degree of

Doctor of Philosophy

in

Mathematics

by

Victor Churchill

DARTMOUTH COLLEGE

Hanover, New Hampshire

May 2020

Examining Committee:

Dr. Anne Gelb, Chair

Dr. Yoonsang Lee

Dr. Geoffrey Luke

Dr. Douglas Cochran

F. Jon Kull, Ph.D.
Dean of the Guarini School of Graduate and Advanced Studies

Abstract

Synthetic aperture radar (SAR) is a day or night any-weather imaging modality that is an important tool in remote sensing. Most existing SAR image formation methods result in a maximum a posteriori image estimate which approximates the reflectivity function of an unknown ground scene. This single image provides no quantification of the certainty with which the features in the estimate should be trusted. In addition, finding the mode is generally not the best way to interrogate a posterior. Therefore, we introduce a sampling approach to SAR image formation. A hierarchical Bayesian model is constructed using conjugate priors that directly incorporate coherent imaging and the problematic speckle phenomenon which is known to degrade image quality. A Gibbs sampler is used to sample the resulting posterior, obtaining samples of the image as well as the parameters governing speckle and noise. Utilizing a non-uniform fast Fourier transform, $\sim 5 \times 10^5$ latent variables are efficiently sampled. The resulting samples are used to compute estimates, as well as to derive other statistics such as confidence images which aid in uncertainty quantification. The latter information is particularly important in SAR, where ground truth images even for synthetically-created examples are typically unknown. Examples are performed using the GOTCHA Volumetric SAR data set from the Air Force Research Laboratory. Our results and analysis show that a sampling-based approach to SAR image formation provides parameter-free estimates with improved contrast and significantly reduced speckle, as well as unprecedented uncertainty quantification information.

Preface

The idea for this project came in January during an AFOSR meeting. I gave a talk on how to apply a Bayesian learning method to image processing tasks using a high order total variation prior. In what I believe was a kind effort to make my work seem more relevant to the audience, Doug Cochran asked if the methods I used would translate to the problem of SAR image formation. During the next few talks I began to write out the likelihood and prior structure for SAR imaging, and the rest is history. So thank you, Doug.

Thank you to my other committee members Geoff Luke, Yoonsang Lee, and especially my advisor Anne Gelb. I sincerely appreciate your guidance and advocacy over the past four years. Thank you also to Matt Parno for supporting me while at CRREL, and thank you to Theresa Scarnati for supporting me while at the ATR Center. Thank you to my fellow graduate students, especially my close friends.

Thank you to my supportive family: Olivia, Mom, Dad, and Emma. Olivia, thanks for sharing distracting content. Mom, thanks for letting us stay at your house for the last two months. Dad, thanks for inspiring me to be academic from a very young age. Emma, thanks for being silly and reminding me not to take everything so seriously. Finally, thank you to Mary. I would have quit school long ago if you weren't here pushing me. You have been by my side in this process from the applications through this thesis. I am grateful for everything you do to support me. Also thank you to Luke, our dog, for being a good boy.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Synthetic aperture radar (SAR) is a widely-used imaging technology for surveillance and mapping. Because SAR is capable of all-weather day-or-night imaging, it overcomes several challenges faced by optical imaging technologies and is an important tool in modern remote sensing, [42]. Applications for SAR include areal mapping and analysis of ground scenes in environmental monitoring, remote mapping, and military surveillance, [2]. Figure 1.1 shows a basic geometry for spotlight mode SAR data collection. An airborne sensor traverses a flight path, periodically transmitting an interrogating waveform toward an illuminated region of interest. The microwave energy pulses emitted impinge on targets in the illuminated region that scatter electromagnetic energy back to the sensor. The sensor measures the reflected signal and processes it. The demodulated data, called a phase history, is collected and passed on to an image formation processor. This paper mainly concerns the function of the image formation processor: to produce a reconstruction of the two-dimensional electromagnetic reflectivity function of the illuminated ground scene from SAR phase history data. For a thorough overview of SAR and basic image formation techniques, see e.g. [34, 39, 42, 52]. Traditionally, SAR images are formed using back projection, see e.g. [40]. Figure 1.2 shows a SAR image of a parking lot formed using back projec-

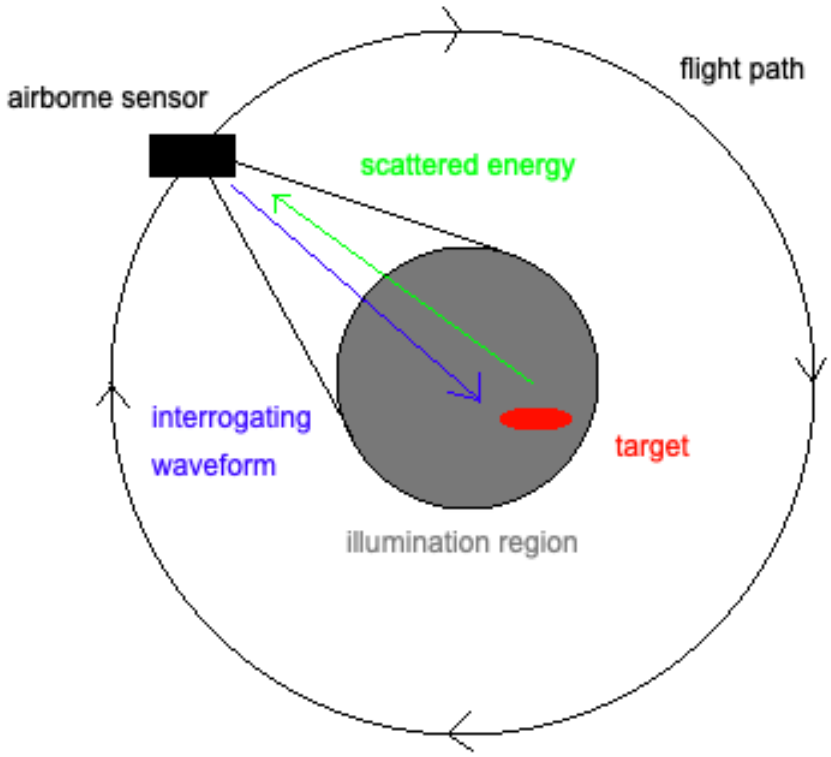

Figure 1.1: Basic diagram of circular spotlight mode SAR data collection.

tion, [20]. While this SAR image looks much different than an optical image (optical images of this parking lot are shown in Chapter 3), key features from the parking lot can be recognized such as the roads, medians, and cars. Only the magnitude of this complex-valued reflectivity image is viewed here, and we notice the very grainy appearance due to the speckle phenomenon, which we later discuss at length. Back projection can also produce streaking and sidelobe artifacts. Several issues make it difficult to form artifact-free denoised SAR images from phase history data. These issues and our approach to addressing them are given below.

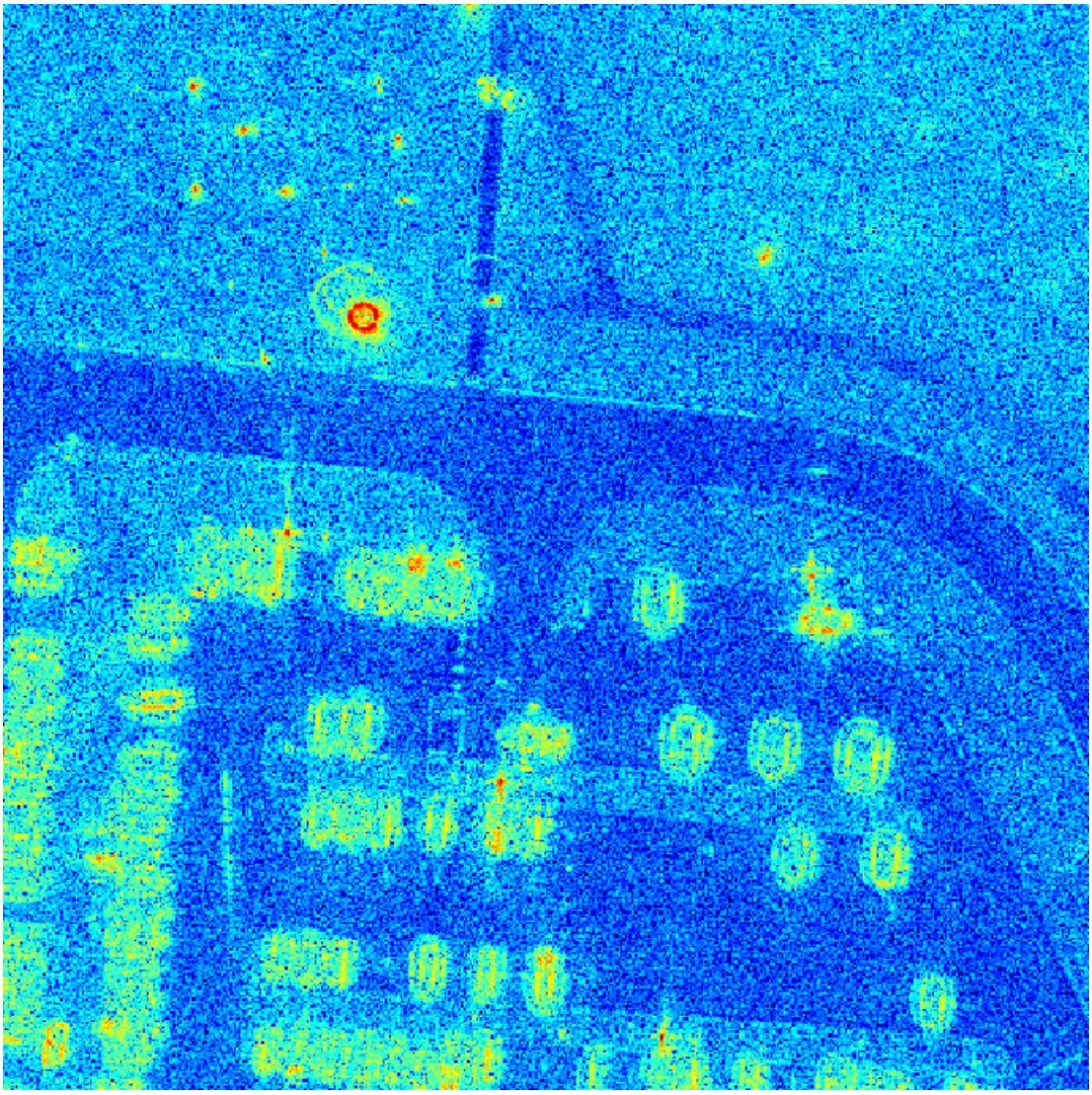

Figure 1.2: A SAR image of a parking lot formed using back projection.

Section 1.1

Addressing SAR image formation challenges

In this paper, we develop a sampling-based SAR image formation method which directly addresses problem size, speckle, and uncertainty quantification.

1.1.1. Problem size

SAR image formation is a large problem, requiring efficient storage and methods. Our examples will have images with $\sim 3 \times 10^5$ pixels and data with $\sim 2 \times 10^7$ elements. This large scale prohibits use of traditional matrix-based methods for linear inverse problems, as even storing dense matrices of this size is problematic. Hence function handles are used to apply the action of such matrices. The SAR forward model can be viewed as a non-uniform Fourier transform. Hence, we use the NUFFT to quickly apply its approximate forward and inverse action, [33]. More information on the NUFFT can be found in [29, 30, 33, 41, 44]. Such a high-dimensional problem is rarely addressed by sampling methods. Our algorithm employs an NUFFT to efficiently sample each of $\sim 5 \times 10^5$ latent variables in the posterior.

1.1.2. Coherent imaging and speckle

SAR is a coherent imaging system, meaning that both the data and the reflectivity image are complex-valued. While typically only the magnitude is viewed, the phase information cannot be neglected in the image formation process, and is important for downstream tasks like interferometry, [46]. All coherent imaging modalities are contaminated with speckle, a multiplicative-noise-like phenomenon which causes images to have a grainy appearance, [4]. A hierarchical Bayesian network structure, [8, 36, 55], allows our model to directly incorporate fully-developed speckle into the prior probability distribution. In addition, the parameter governing the distribution of the speckle is also estimated for each image pixel, spatially specifying the distribution of speckle. This hierarchical structure also reduces the importance of parameter selection. In what follows we need only choose hyper-hyper-parameters, i.e. parameters on the parameters of the prior distribution on the reflectivity image. In general these hyper-hyper-parameters can be chosen to make the resulting prior uninformative, but we focus on exploring how these parameters can be chosen to encourage

sparsity.

1.1.3. Uncertainty quantification

The deterministic approach to SAR image formation typically only obtains a single image that estimates the ground truth reflectivity of the scene. However, in a Bayesian formulation an entire probability density function for the unknown variables is returned. Hence, in an effort to better describe the unknown reflectivity, we take an approach to compute an entire posterior density from which samples can be drawn and statistics can be computed. In the Bayesian approach to estimation, all quantities of interest are viewed as random variables, with probability distributions describing their behavior. Known quantities, e.g. SAR phase history data, are called observable variables, and unknown quantities, e.g. the reflectivity image, are called latent variables. The goal is to infer the latent variables from the observable variables. The encoding of these quantities as random variables does not contradict that they are defined quantities, but expresses our lack of certainty about them. Recall that Bayes' theorem tells us that

$$\text{Posterior probability} \propto \text{Likelihood function} \times \text{Prior probability}.$$

The posterior is built from the prior distribution on the latent variable, i.e. our belief about it before data has been considered, and the likelihood function, which governs how well the data fits the model. In this way the posterior is a synthesis of prior belief and information carried by the data, [16, 17, 43].

There are two main point estimates used in SAR image formation. Maximum likelihood estimates, e.g. NUFFT image formation, that minimize a least-squares cost function typically result in noisy reconstructions. Hence, usually maximum *a posteriori* (MAP) estimates are sought. By definition a MAP estimate uses the max-

imum of the posterior probability distribution to estimate a latent variable. However, MAP estimates in SAR are usually applied within the traditional deterministic approach, where ill-posedness resulting in artifacts and noise like in Figures 1.2 and 2.1 is addressed using regularization, meaning that a penalty term is added to the least squares cost function seen in maximum likelihood estimates. The regularity imposed is often thought of as including *a priori* information, e.g. that the image should be smoothly varying. Usually a convex optimization method is required to compute the maximum.

There are a few issues with MAP estimates. The maximum is not a categorically strong representative of the posterior probability distribution. In general, sampling is a better way to interrogate a probability distribution than finding its maximum. A consequence of only estimating with the maximum is that we do not know the certainty with which we can trust the estimate. Hence we have no way of knowing which structures in the reflectivity estimate are truly there and which are noise or artifacts. Hence, in this paper we develop a sampling-based SAR image formation method with uncertainty quantification. The resulting posterior density is frequently not from a known family of distributions. Therefore, samples cannot easily be drawn from it and a sampling scheme is required. In particular, we use a Markov Chain Monte Carlo (MCMC) based Gibbs sampler to sample the posterior. While sampling-based methods have been developed to quantify the uncertainty in basic linear image reconstruction, e.g. [8], they have not been used in SAR image formation, perhaps due to the aforementioned circumstances of problem size and speckle which make this more than a straightforward application of established methods.

The extra information obtained by sampling the posterior density can be used in a variety of ways. It can be used to compute sample mean or median estimates of the unknown reflectivity image as well as the speckle parameter. In addition, sam-

ple variance statistics can be computed to aid in uncertainty quantification. The examples presented in this paper suggest that our sampling-based method provides estimates with significantly better contrast and drastically reduced speckle compared with other SAR image formation methods, in addition to the unprecedented uncertainty quantification information.

Section 1.2

Outline

This paper is organized as follows.

- Chapter 2 realizes a method for SAR image formation that incorporates coherent imaging and the corresponding fully-developed speckle phenomenon into a hierarchical Bayesian model, using an MCMC-based Gibbs sampler to sample from the posterior in order to quantify uncertainty in the data and compute superior estimates.

- Chapter 3 demonstrates the method of Chapter 2 with examples of SAR image formation from real-world data. While the methods presented are independent of the flightpath, we focus on two-dimensional image formation from single-pass circular spotlight mode SAR phase history data. In particular, the method is tested on real-world phase history data from the Air Force Research Laboratory's GOTCHA Volumetric Dataset Version 1.0, [20].

- Finally, Chapter 4 contains perspectives on this work as it relates to existing sparsity-encouraging image reconstruction methods. A summary of the work and an outline for future research directions are also provided.

Chapter 2

Methods

SAR image formation is the process of transforming measured electromagnetic scattering data into an image. This is a difficult problem for several reasons. SAR image formation is a large-scale problem with high-resolution images requiring more than 10^5 pixels and phase history data in some cases containing more than 10^7 elements. SAR is a coherent imaging system, meaning that SAR images are complex-valued, containing both magnitude and phase information. Hence they are affected by speckle, a phenomenon similar to multiplicative noise, which degrades the image quality. The data are also complex-valued, and are typically modeled as the two-dimensional discrete Fourier transform of the desired reflectivity image with Fourier modes located on a non-uniform grid.

The result of most existing SAR image formation methods is a single image that represents an estimate of the unknown ground truth image, typically a maximum *a posteriori* estimate. However, the maximum is not categorically representative of the posterior probability distribution. In addition, this prediction is not probabilistic. Hence, in this paper a new approach is formed. Considering each element of the model for SAR image formation, a hierarchical Bayesian model is constructed using conjugate priors, and a Gibbs sampler is used to sample the resulting posterior density.

The result is a set of samples from the posterior density, from which a variety of statistics (including the sample mean and sample variance, etc.) can be computed and used not only to estimate the image but the speckle itself and in general to quantify uncertainty. In this chapter, this new model and algorithm for sampling its posterior are presented.

Section 2.1

Model and posterior specification

This section begins by specifying the linear system used to model the relationship between SAR images and phase history data. Next, a hierarchical prior structure is specified in order to account for the speckle phenomenon. We work with the fully developed speckle model, [28], to form a posterior density for the latent variables, which is analytically computed.

2.1.1. Continuous model

This subsection closely follows [51]. SAR data are acquired by measuring reflections from interrogating waveforms sent along a flight path. We focus on spotlight SAR, where waveforms are high bandwidth pulses transmitted at equally-spaced azimuth angles θ. The illuminated circular region of interest is $\mathcal{D} = \{(x, y)|x^2 + y^2 \leq R^2\}$. For each angle θ, the emitted pulses are linear FM chirp signals given by the real part of

$$s(t) = \begin{cases} e^{i(\omega_0 t + \alpha t^2)}, & |t| \leq \frac{T}{2} \\ 0, & \text{otherwise} \end{cases}, \tag{2.1}$$

where ω_0 is the center frequency, 2α is the chirp rate, and $T > 0$ is the chirp pulse duration. The chirp signal is convolved with the reflectivity function $f(x, y)$ and the reflected waveforms are recorded at equally spaced discrete time steps. What follows

relies on five assumptions:

(a) The chirp rate α is small enough so that for each recorded time t, the demodulated reflected waveform is mixed with the wave having instantaneous frequency $\omega(t) = \omega_0 + 2\alpha(t - 2d_\theta/c)$, [42], where d_θ is the distance from the sensor at angle θ to the scene center, and c is the speed of light.

(b) The distance from the flight path to the receiver is assumed sufficiently larger than the illuminated region radius R, such that the range error due to wavefront curvature is negligible.

(c) The scene contains only isotropic scatterers. This assumption implies that the response amplitude $f(x, y)$ is independent of the azimuth angle θ.

(d) There are no scatterers outside $\mathcal{D}$.

(e) The Born approximation applies. Essentially this means that the returned signal does not include any scattered energy that has impinged on a target and then subsequently impinged on other targets before returning to the sensor.

While assumptions (a), (b), (d), and (e), are commonly used in SAR and introduce little error, assumption (c) is only even approximately true for narrow-angle synthetic apertures, [51]. In future work, we will discuss composite imaging. However, in this paper we focus on direct imaging using full 360^o synthetic apertures.

Given a constant elevation angle ϕ between the flight path and $\mathcal{D}$, the reflected waveforms are of the form

$$\hat{f}(\omega(t), \theta) = \int\int_{\mathcal{D}} f(x, y) \exp\left(-i\frac{4\pi\omega(t)\cos\phi}{c}(x, y) \cdot (\cos\theta, \sin\theta)\right) dxdy \qquad (2.2)$$

where c is the speed of light, [51]. Hence the phase history data $\hat{f}(\omega(t), \theta)$ are the two-dimensional Fourier transform of the reflectivity function $f(x, y)$.

2.1.2. Discrete model

To discretize (2.2), consider $\hat{f}(\omega(t), \theta)$ for a discrete set of azimuth angles $\{\theta_j\}$, and a set of time steps corresponding to a discrete set of frequency values $\{\omega_k\}$, [51]. The SAR phase history data provided by AFRL come in a matrix, so we can think of $\hat{\mathbf{f}}$ describing the grid of vectors containing all frequency samples at angle θ_j, and $\mathbf{f}$ the complex reflectivity image to be recovered on a uniform grid. For simplicity, we use the concatenation of both $\hat{\mathbf{f}} \in \mathbb{C}^M$ and $\mathbf{f} \in \mathbb{C}^N$ where M and N are defined as the number of data samples and pixels, respectively. Then we have discretized (2.2) as a complex-valued linear system. To summarize, we have

$$\hat{\mathbf{f}} = \mathbf{F}\mathbf{f} + \mathbf{n} \tag{2.3}$$

where $\hat{\mathbf{f}} \in \mathbb{C}^M$ is the vertically concatenated phase history data. The matrix $\mathbf{F} \in \mathbb{C}^{M \times N}$ is the two-dimensional discrete non-uniform Fourier transform matrix. The vector $\mathbf{f} \in \mathbb{C}^N$ is the unknown reflectivity image matrix having been vertically concatenated. Finally, the vector $\mathbf{n} \in \mathbb{C}^M$ represents model and measurement error. The objective is to infer $\mathbf{f}$ from $\hat{\mathbf{f}}$. Note that while SAR images are complex-valued, usually only the magnitude is viewed while the phase is omitted. However, recovering the phase is also important, [46], and therefore cannot be neglected. We note that (2.3) is a fairly simple model for the relationship between image and data in SAR, and modifications to it are common in order to deal with issues such as autofocusing to reduce phase error, [53]. These additional modifications are not a main concern in this paper, and indeed are able to be incorporated into the method that follows.

Throughout this paper we assume $\mathbf{n}$ is complex circularly-symmetric white Gaussian noise. That is, $\mathbf{n} \sim \mathcal{CN}(\mathbf{0}_M, \beta^{-1}\mathbf{I}_M)$ where $\mathbf{0}_M$ is a M-length vector of zeros, $\beta^{-1} > 0$ is the noise variance, and $\mathbf{I}_M$ is the $M \times M$ identity matrix. The assumption

of white Gaussian noise is often made in SAR, but a diagonal covariance matrix implying independently distributed noise across each pixel with a potentially different variance for each element is easily accommodated in the proposed sampling procedure. Nevertheless, we focus on the white Gaussian noise assumption. Therefore we have a Gaussian likelihood function

$$p(\hat{\mathbf{f}}|\mathbf{f},\beta) \propto \beta^M \exp\left(-\frac{\beta}{2}||\hat{\mathbf{f}} - \mathbf{F}\mathbf{f}||^2\right). \tag{2.4}$$

Read 'the probability of $\tilde{\mathbf{f}}$ given $\mathbf{f}$ and β', (2.4) measures the goodness of fit of the model (2.3). Note $||\mathbf{g}||^2 = \mathbf{g}^H\mathbf{g}$ where $\mathbf{g}^H$ is the conjugate transpose of $\mathbf{g}$.

2.1.3. Estimation techniques

In the Bayesian approach to linear inverse problems, the data and unknown image are encoded as observable variables and latent variables (i.e. variables that need to be inferred). Recall that the likelihood function is defined as the probability distribution of the observed variables conditional on the other variables. In this subsection, we'll use the likelihood function to derive a few different inversion techniques to find an estimate for the unknown quantity $\mathbf{f}$. As of now, the only model parameter is the noise variance β^{-1}, which for now we will consider a known quantity (hence observable).

Maximum likelihood estimation. Perhaps the most straightforward way to estimate $\mathbf{f}$ from $\hat{\mathbf{f}}$ is to maximize this likelihood function. From the Gaussian likelihood defined above by (2.4), this estimate is

$$\begin{aligned}\mathbf{f}^*_{ML} &= \arg\max_{\mathbf{f}} \left\{p(\hat{\mathbf{f}}|\mathbf{f},\beta)\right\} \\ &= \arg\min_{\mathbf{f}} \left\{||\hat{\mathbf{f}} - \mathbf{F}\mathbf{f}||^2\right\}.\end{aligned} \tag{2.5}$$

Note this estimate is referred to as the NUFFT reconstruction, as $\mathbf{f}^*_{ML} = \mathbf{F}^H\hat{\mathbf{f}}$ only requires an inverse NUFFT application in order to invert the data, [33]. Specifically, the reflectivity image can be found by interpolating the typically polar grid of measured samples in frequency space to an equally spaced rectangular grid, then computing an inverse uniform fast Fourier transform, [2]. However, noisy data and model error can cause artifacts or a noisy image using this traditional least-squares-type NUFFT inversion. An example is shown in Figure 2.1, [20]. As shown in the figure, this estimate, depending on the sampling, can have an even grainier appearance than the traditional back projection reconstruction. While it may result in lower image quality, the NUFFT reconstruction has the advantage of speed and is very quickly reconstructed. Hence in general an improvement is needed.

Priors, MAP estimation, and regularization. In SAR, to improve on the maximum likelihood estimate given by (2.5), the cost function is frequently regularized by adding a penalty term on the ℓ_1 norm of $|\mathbf{f}|$ or a transform $\mathbf{T}|\mathbf{f}|$

$$\mathbf{f}^* = \arg\min_{\mathbf{f}} \left\{ \frac{\beta}{2}||\hat{\mathbf{f}} - \mathbf{F}\mathbf{f}||^2 + \lambda||\mathbf{T}|\mathbf{f}|||_1 \right\}. \tag{2.6}$$

In addition to regularizing the ill-posed problem, where more than one $\mathbf{f}$ may satisfy the model equation, this formulation encourages sparsity in the magnitude $|\mathbf{f}|$. The phase, which is not modeled as sparse, [42], is governed only by the least squares fit term. Equation (2.6) in general has no direct solution and must be minimized using a convex optimization method like the alternating direction method of multipliers (ADMM), [13]. The ℓ_1 regularization term in (2.6) imposes the sparsity penalty on $\mathbf{f}$. In the field of compressive sensing the sparsity prior parameter λ and noise variance β^{-1} are often combined and relabeled as the regularization parameter, which balances the fidelity term, the sparsity penalty, and noise reduction. Even though the

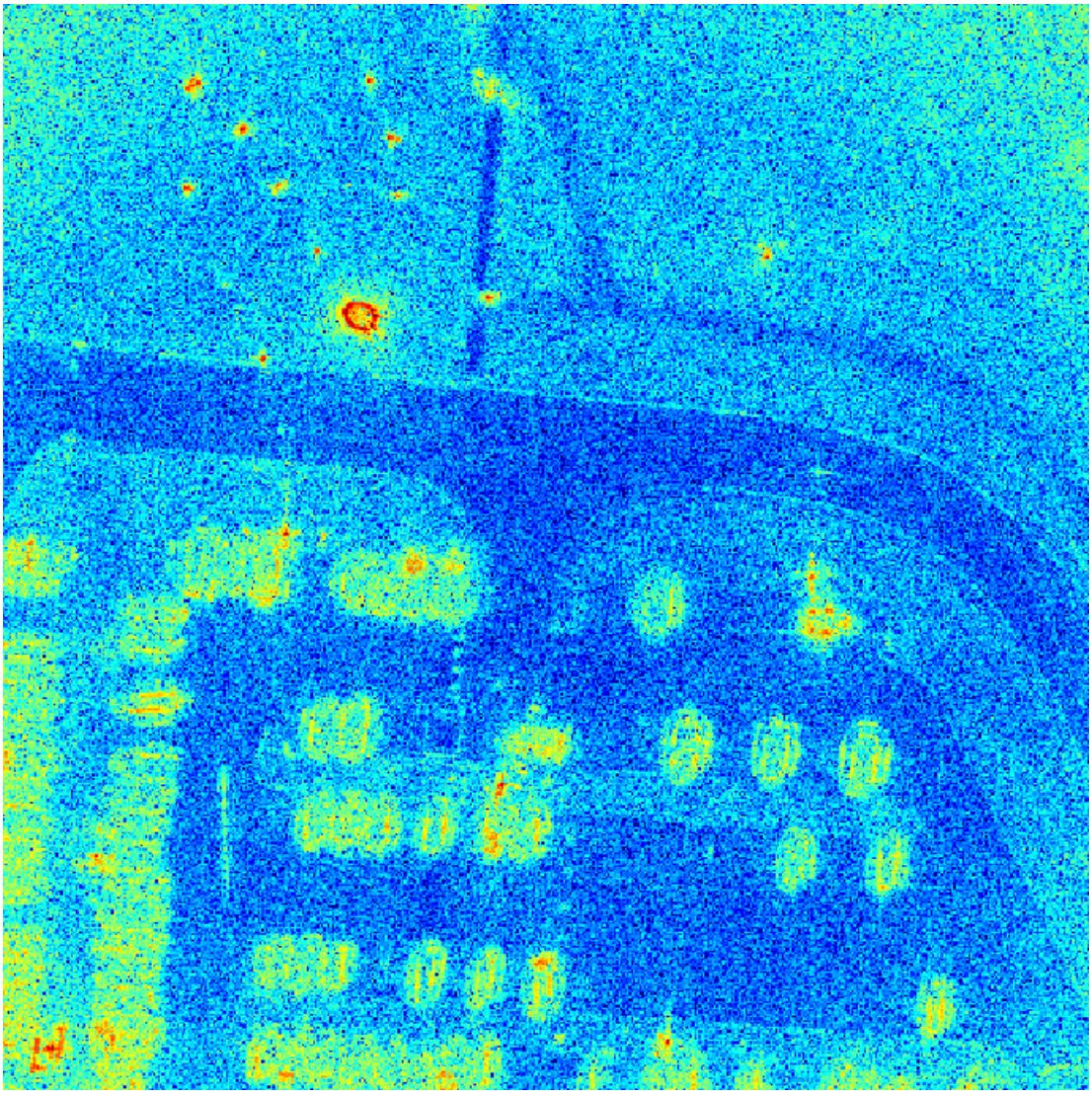

Figure 2.1: A SAR image of a parking lot formed using the NUFFT.

inversion can be ill-posed, if certain conditions are met, then with high probability $\mathbf{x}$ can be exactly recovered from many fewer than N measurements using the noise-free version of this method known as basis pursuit, [18]. Figures 2.2 and 2.3 show two reconstructed images using (2.6), the first with $\mathbf{T} = \mathbf{I}$ and the second with $\mathbf{T}$ the total variation (approximate gradient) operator. We see that the ℓ_1 regularization reconstruction does a decent job at sparsifying the image, i.e. drawing values to zero, especially along the road area which is fairly smooth and hence won't scatter much

electromagnetic energy back to the sensor. This has the effect of increased contrast making targets like the cars more clearly visible. However, there is still significant speckle causing an overall grainy appearance in the rough grassy areas. The TV regularization reconstruction on the other hand handles the grainy speckle issue very well, smoothing out much of the image. The issue here is the block-like appearance, which is an artifact known to occur in TV regularization. These blocky regions make it more difficult to determine if there is actually signal there, particularly without any uncertainty quantification information.

From the cost function used in (2.6), however, the observation can be made that had the prior probability distribution

$$p(\mathbf{f}|\lambda) \propto \exp\left(-\lambda||\mathbf{T}|\mathbf{f}|||_1\right), \tag{2.7}$$

been invoked, the resulting posterior would be

$$\begin{aligned} p(\mathbf{f}|\hat{\mathbf{f}},\beta,\lambda) &\propto p(\hat{\mathbf{f}}|\mathbf{f},\beta)p(\mathbf{f}|\lambda) \\ &\propto \exp\left(-\frac{\beta}{2}||\hat{\mathbf{f}}-\mathbf{F}\mathbf{f}||^2 - \lambda||\mathbf{T}|\mathbf{f}|||_1\right). \end{aligned} \tag{2.8}$$

It is clear that maximizing (2.8) would yield (2.6), hence (2.6) is known as a maximum a posteriori estimate. This ℓ_1 prior is often chosen in the field of compressed sensing, where limited data is collected and sparsity is suspected in $|\mathbf{f}|$. Of course this is not the only prior distribution that can be used and others would invoke other *a priori* beliefs. From this discussion above, it is clear that the regularization penalty term within the cost function imposes the *a priori* belief specified in the prior probability distribution, [54]. It is also evident that without prior information of λ or β, they will be difficult to choose. Hence in what follows we take the view that they should also be estimated.

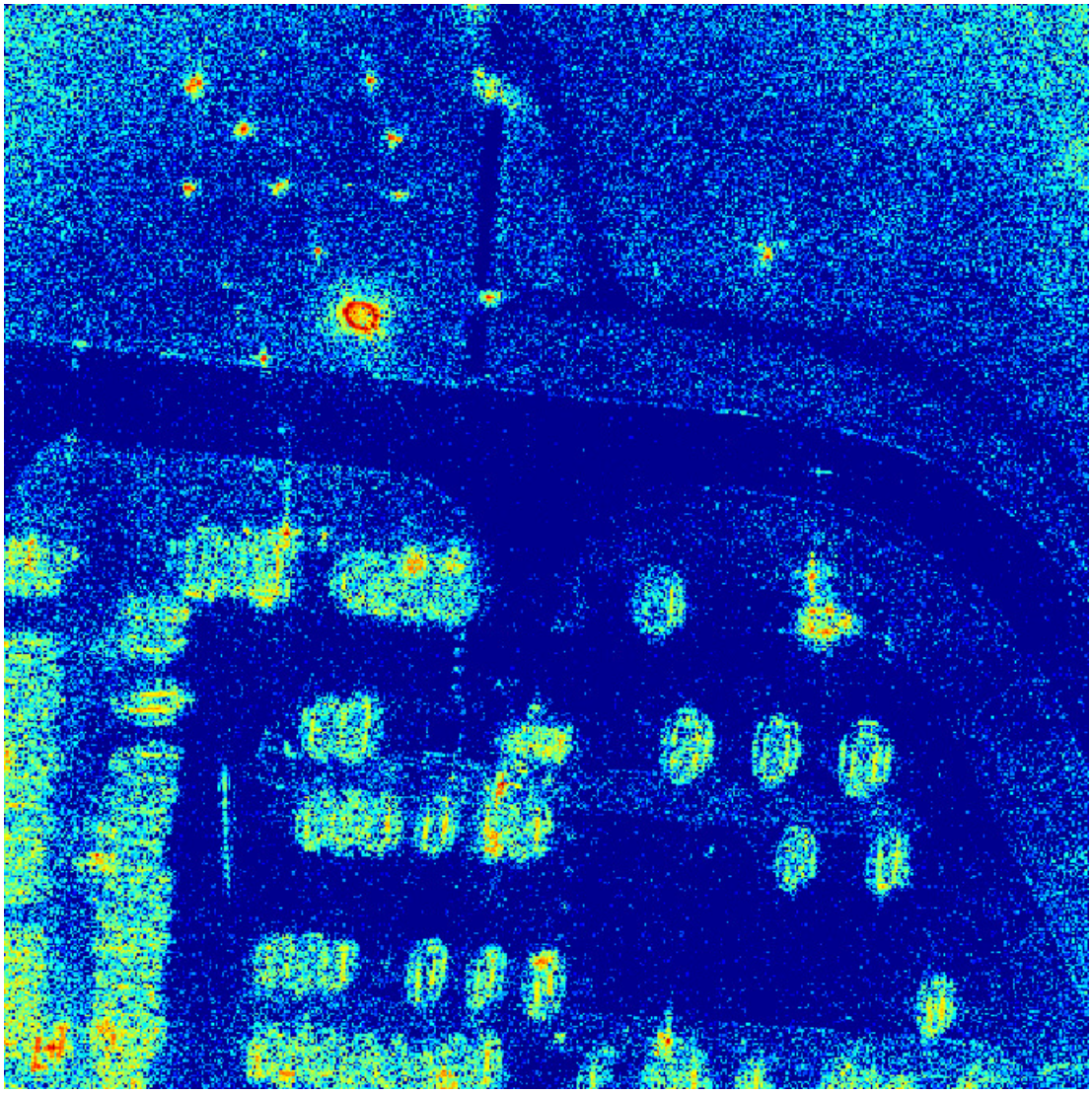

Figure 2.2: A SAR image of a parking lot formed using ℓ_1 regularization.

We note that due to the difficulty of minimizing the ℓ_1 norm of the magnitude of a complex vector, in SAR, typically an approximation $|\mathbf{f}| = \Theta^* \mathbf{f}$ is made, where $\Theta_{j,j} = \text{angle}(\tilde{\mathbf{f}}_j)$ and $\tilde{\mathbf{f}}$ is an approximate cheaply-computed solution like e.g. the NUFFT image, [51]. This has two consequences. First, this no longer regularizes the sparsity of the magnitude of $|\mathbf{f}|$, but an approximation to the magnitude. In addition, the regularization term no longer corresponds to any prior distribution as the data was considered in order to form the initial estimate.

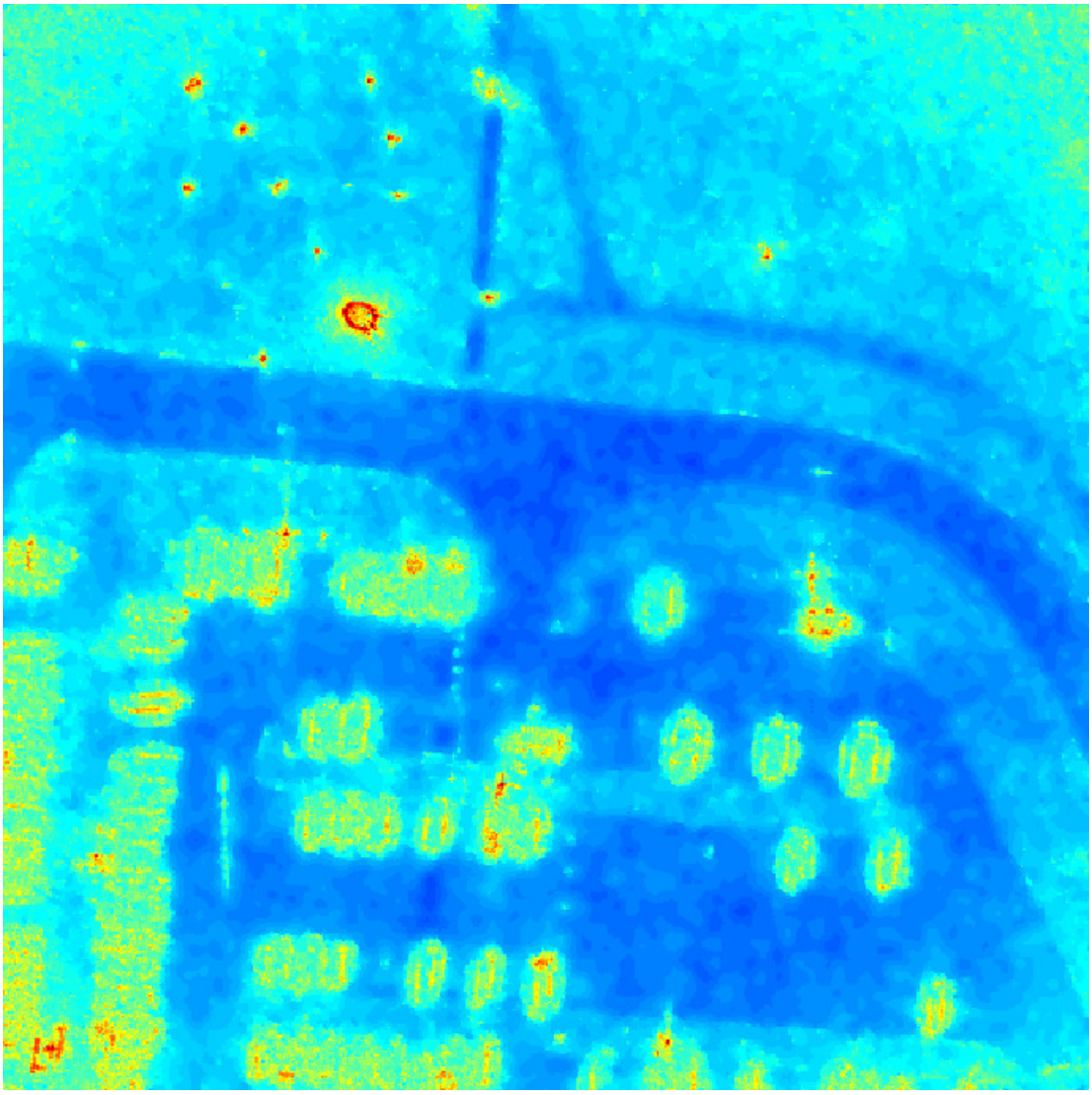

Figure 2.3: A SAR image of a parking lot formed using TV regularization.

2.1.4. Efficient application of large matrices

Forming, storing, and applying the dense matrices $\mathbf{F}$ and $\mathbf{F}^H$ matrices in real-world problems is not practical. Hence function handles will be used to apply the action of these matrices. In particular, let the application of the discrete nonuniform Fourier transform matrix $\mathbf{F}$ be approximated by the forward NUFFT. Similarly let the application of its Hermitian transpose $\mathbf{F}^H$ be approximated by the inverse NUFFT. We make this change in order to utilize libraries to quickly apply NUFFTs. The NUFFT

operation uses discrete convolution to regrid the non-uniform Fourier mode locations onto a uniform grid so that a uniform FFT can be applied, [51]. For the specific implementation of the NUFFT used in the examples, see [33]. General information on NUFFTs can be found in [29, 30, 33, 41, 44]. Therefore, using existing libraries for the NUFFT, [33], application of $\mathbf{F}$, $\mathbf{F}^H$, and their respective products can be efficiently applied without requiring the storage of any prohibitively large matrices. For efficient use of space, we continue to use $\mathbf{F}$ and $\mathbf{F}^H$ notation. For more research regarding using function handle as forward operator, see e.g. [12, 57]. This is commonly seen in nonlinear inverse problems and PDEs, [7, 10, 11].

Note at least three sources of error in this model thus far: the approximation of a continuous Fourier transform with a discrete Fourier transform which includes Gibbs phenomenon, the gridding of non-uniform to uniform Fourier modes taking place in the NUFFT which makes the operation only approximately unitary, and model and measurement error. Further work is needed to meaningfully quantify this error.

2.1.5. Posterior computation

With the likelihood given by (2.4) and the NUFFT in hand, the next step in computing the posterior distribution is to specify prior probability distributions for the latent variable $\mathbf{f}$. The prior expresses a belief about a quantity before observation, which may support or contradict the prior. We use the fact that SAR images are affected by the speckle phenomenon as a prior. This phenomenon, often misidentified as "noise", causes a complicated granular pattern of bright and dark spots throughout an image, [42]. While speckle is in fact signal (as opposed to noise) it degrades the image quality by lowering the contrast, and hence we would like to remove it. Speckle occurs in all coherent imaging, that is imaging where one tries to retrieve both magnitude and phase information, [42]. Our focus in this paper is not on reducing speckle through denoising or other methods, see e.g. [4], but simply by incorporating it into the image

formation model. While acknowledging that it may not be appropriate for all image pixels, we uniformly employ the fully-developed speckle model, [28, 49], (a.k.a. Gaussian scattering), based on the assumption that the spatial resolution dimension is considerably larger than the wavelength and the illuminated surface is rough enough, [28]. This model is widely used to describe speckle in SAR, [42, 52]. In this model, assume elements of the real and imaginary parts of the image, $\mathrm{Re}(\mathbf{f}_i)$ and $\mathrm{Im}(\mathbf{f}_i)$, respectively, are independently and identically Gaussian distributed with precision (inverse variance) $\boldsymbol{\alpha}_i$ to be determined. That is, $\mathrm{Re}(\mathbf{f}_i), \mathrm{Im}(\mathbf{f}_i) \sim \mathcal{N}(0, 2\boldsymbol{\alpha}_i^{-1})$. By independence, the distribution of the vector quantities is also zero-mean Gaussian with covariance $\mathrm{diag}(2\boldsymbol{\alpha}^{-1})$. That is, $\mathrm{Re}(\mathbf{f}), \mathrm{Im}(\mathbf{f}) \sim \mathcal{N}(\mathbf{0}_N, \mathrm{diag}(2\boldsymbol{\alpha}^{-1}))$ with precision $\boldsymbol{\alpha} \in \mathbb{R}^N$ sometimes called the radar cross section, [28]. This is conveniently encoded by $\mathbf{f} \sim \mathcal{CN}(\mathbf{0}_N, \mathrm{diag}(\boldsymbol{\alpha}^{-1}))$. That is, $\mathbf{f}$ is distributed circularly-symmetric complex Gaussian. The prior density is

$$p(\mathbf{f}|\boldsymbol{\alpha}) \propto \prod_{i=1}^{N} \boldsymbol{\alpha}_i \exp\left(-\frac{1}{2}||\sqrt{\boldsymbol{\alpha}} \odot \mathbf{f}||^2\right), \tag{2.9}$$

where $\odot$ is elementwise multiplication. Therefore, the prior on the magnitude $|\mathbf{f}_i| = \sqrt{\mathrm{Re}(\mathbf{f}_i)^2 + \mathrm{Im}(\mathbf{f}_i)^2}$ is a Rayleigh probability distribution with mean proportional to $\boldsymbol{\alpha}_i^{-1}$, [49]. This is the standard specification for fully-developed speckle. Because the changes in the magnitude of each pixel $|\mathbf{f}_i|$ is proportional to $\boldsymbol{\alpha}_i^{-1}$, the speckle phenomenon has also been modeled as a multiplicative noise, [28]. While efforts to reduce speckle abound, [4], here we address the speckle directly by including it in our model with the joint prior probability distribution (2.9), and later estimating the associated speckle parameter $\boldsymbol{\alpha}^{-1}$ through sampling rather than post-image-formation techniques. By parametrizing $\mathbf{f}$ with $\boldsymbol{\alpha}$ we are doubling the number of parameters in this model! This is of course a *computational* challenge, but not a methodological one, [48].

Digression on MAP estimates. At this point, we have a likelihood (2.4) and a prior (2.9), so a posterior can be computed for $\mathbf{f}$ if a noise variance β and a prior precision $\boldsymbol{\alpha}$ are specified. Recall that the posterior is the probability of the unknown latent variables conditional on the observable variables. By Bayes' theorem,

$$\begin{aligned} p(\mathbf{f}|\hat{\mathbf{f}}, \boldsymbol{\alpha}, \beta) &\propto p(\hat{\mathbf{f}}|\mathbf{f}, \beta) p(\mathbf{f}|\boldsymbol{\alpha}, \beta) \\ &\propto \beta^M \prod_{i=1}^{N} \boldsymbol{\alpha}_i \exp\left(-\frac{\beta}{2}||\hat{\mathbf{f}} - \mathbf{F}\mathbf{f}||^2 - \frac{1}{2}||\sqrt{\boldsymbol{\alpha}} \odot \mathbf{f}||^2\right). \end{aligned} \tag{2.10}$$

This would allow us to take the typical approach of obtaining a maximum a posteriori (MAP) estimate by solving the optimization problem

$$\begin{aligned} \mathbf{f}^*_{MAP} &= \arg\max_{\mathbf{f}} p(\mathbf{f}|\hat{\mathbf{f}}, \boldsymbol{\alpha}, \beta) \\ &= \arg\min_{\mathbf{f}} \left\{ \frac{\beta}{2}||\hat{\mathbf{f}} - \mathbf{F}\mathbf{f}||^2 + \frac{1}{2}||\sqrt{\boldsymbol{\alpha}} \odot \mathbf{f}||^2 \right\}. \end{aligned} \tag{2.11}$$

As explained at length above, however, a MAP estimate is not necessarily representative of the posterior, and does not provide an estimate of certainty. Nevertheless, we present the optimization problem as it may be useful for reference or for those more experienced with regularization to analyze the form of this cost function. In (2.11) we have a least-squares fidelity term coming from the likelihood function which measures the fit of the data to the proposed $\mathbf{f}$, followed by a regularization term which penalizes the ℓ_2 norm of $\mathbf{f}$ after being transformed by $\sqrt{\boldsymbol{\alpha}}$. Resulting from the Gaussian prior (2.9), regularization with the ℓ_2 norm, known as Tikhonov regularization or ridge regression, can be used to encourage smoothness in the solution. The solution requires an convex minimization with ℓ_2-regularized solutions at each step. See, e.g. [28], for a detailed example of how one could handle a minimization of this type in the context of SAR image formation. This model is fairly flexible, in that modifications can be

made to the effective regularization term to include an analysis operator, or this type of regularization can be encouraged in a synthesis manner, [31]. Our previous work discussed a synthesis approach for the total variation operator, [24]. If an analysis operator is invoked, then attention must be paid to its rank, see [8, 9].

Hierarchical model. As mentioned, we do not wish to limit ourselves to a MAP estimate first because it may not be representative of the posterior. Second, although an estimated value for each pixel is given, there is no estimate of the statistical confidence we have in those pixel values or in the resulting features formed in the image, [47]. That is, we not only wish to obtain parameter estimates, but also to quantify the uncertainty in those estimates. Third, (2.11) is lacking in that the noise variance β^{-1} and prior precision $\boldsymbol{\alpha}$ are simply specified or asserted, similar to the regularization parameter and regularization matrix in a standard regularized cost function common in the deterministic approach. In reality, β and $\boldsymbol{\alpha}$ are unknown and therefore should also be considered latent variables which need to be inferred. Hence the posterior we seek is $p(\mathbf{f}, \boldsymbol{\alpha}, \beta|\hat{\mathbf{f}})$, not $p(\mathbf{f}|\hat{\mathbf{f}}, \boldsymbol{\alpha}, \beta)$ as in (2.10). It is important to note that we will thus be estimating an entire density for the speckle parameter $\boldsymbol{\alpha}$ for every pixel contemporaneously with the image. This is important because it will be more clear when speckle reduction techniques (if incorporated into this model) are actually working. Typically, without a reference ground truth image, speckle statistics are only estimated from small regions of already-formed images, [4]. In order to calculate this posterior, we need to define prior probability distributions on β and $\boldsymbol{\alpha}$.

In general, we have no intuition for the values of β and $\boldsymbol{\alpha}$, and we can encode that uncertainty by choosing uninformative priors to allow as much variation as possible and let the data choose. In order to compute an analytical form of the posterior, a conjugate Gamma prior is invoked for β. That is, $\beta \sim \Gamma(c, d)$ with probability density

function

$$p(\beta|c,d) \propto \beta^{c-1}\exp(-d\beta). \tag{2.12}$$

Similarly a conjugate Gamma prior is invoked on each element of the speckle parameter $\boldsymbol{\alpha}$, i.e. $\boldsymbol{\alpha}_i \sim \Gamma(a,b)$ with probability density function

$$p(\boldsymbol{\alpha}_i|a,b) \propto \boldsymbol{\alpha}_i^{a-1}\exp(-b\boldsymbol{\alpha}_i). \tag{2.13}$$

By independence, $\boldsymbol{\alpha} \sim \Gamma(a,b)$ with

$$p(\boldsymbol{\alpha}|a,b) \propto \prod_{i=1}^{N} \boldsymbol{\alpha}_i^{a-1}\exp\left(-b\sum_{i=1}^{N}\boldsymbol{\alpha}_i\right). \tag{2.14}$$

Note the dependence of (2.12) and (2.14) on parameters a, b, c, and d, which as in [8, 55] are simply chosen rather than inferred themselves. In [8], the parameters are chosen to reflect the uncertainty in the latent variable, making the prior uninformative. Recall that the mean of this Gamma distribution is $\frac{c}{d}$ and the variance is $\frac{c}{d^2}$. Hence if c is chosen to be 1 and d is chosen to be 10^{-4}, then the mean is 10^4 and the variance is 10^8. By giving this distribution such a large variance, we impose little bias on the value of β. Similar choices can be made for a and b to make this prior uninformative. However, in [55] the same prior structure is used with $a,b,c,d = 0$. This improper prior makes the marginalized priors $p(\mathbf{f}|a,b)$ Student-t distributions, which are peaked at zero for appropriate a and b, and hence encourage sparsity. This provides an interesting prospect as there has been significant research of sparsity-promoting image formation methods for SAR. Our previous work used this prior to perform edge detection from data similar to that seen in SAR, [25]. Hence our examples in Chapter 3 include both the uninformative and sparsity-encouraging

parameters. Nevertheless, the derivation below is done for general a, b, c, and d. Importantly, a, b, c, and d, are the only parameters required to be defined in this model, and require no tuning beyond the two potential schemes mentioned above. The choice of Gamma priors for β and $\boldsymbol{\alpha}_i$ is deliberate. Because the Gamma prior is conjugate to the Gaussian in this case, the prior and the posterior are from the same distribution family. That is, the individual posterior densities for β or $\boldsymbol{\alpha}$ will be Gamma.

The form of the posterior is achieved through the hierarchical Bayesian model described above, where the model parameters (here $\mathbf{f}$) are given priors with prior parameters (here $\boldsymbol{\alpha}$, and β), referred to as hyperparameters. Moving up to the final level of hierarchy in this model, the hyperparameters $\boldsymbol{\alpha}$ and β also have priors (called hyperpriors) with hyperhyperparameters a, b, c, and d. Conjugate priors are used so that all posteriors can be solved for analytically and easily sampled. Specifically, the noise in the system is assumed Gaussian, the noise precision is assumed Gamma, the priors on the real and imaginary parts of the image are Gaussian, and the prior precision at each pixel is Gamma. Finally, again using Bayes' theorem, the joint posterior distribution for $\mathbf{f}$, β, and $\boldsymbol{\alpha}$ is

$$\begin{aligned} p(\mathbf{f}, \boldsymbol{\alpha}, \beta | \hat{\mathbf{f}}, a, b, c, d) &\propto p(\hat{\mathbf{f}}|\mathbf{f}, \beta) p(\beta|c, d) p(\mathbf{f}|\boldsymbol{\alpha}) p(\boldsymbol{\alpha}|a, b) \\ &\propto \frac{\prod_{i=1}^{N} \boldsymbol{\alpha}_i^a}{\beta^{-M-c+1}} \exp(-\frac{\beta}{2}||\hat{\mathbf{f}} - \mathbf{F}\mathbf{f}||^2 - \frac{1}{2}||\sqrt{\boldsymbol{\alpha}} \odot \mathbf{f}||^2 - d\beta - b\sum_{i=1}^{N} \boldsymbol{\alpha}_i). \end{aligned} \tag{2.15}$$

The algorithm below will require the individual posteriors for each latent variable.

Individual posteriors. Because of the conjugate priors used, the individual posterior densities for each latent variable can be found analytically. The posterior for $\mathbf{f}$ is Gaussian, for $\boldsymbol{\alpha}$ is a product of independent Gammas, and for β is Gamma. We

have

$$p(\mathbf{f}|\hat{\mathbf{f}},\boldsymbol{\alpha},\beta) \propto \exp\left(-\frac{\beta}{2}||\hat{\mathbf{f}}-\mathbf{F}\mathbf{f}||^2-\frac{1}{2}||\sqrt{\boldsymbol{\alpha}}\odot\mathbf{f}||^2\right) \tag{2.16}$$

$$p(\boldsymbol{\alpha}|\hat{\mathbf{f}},\mathbf{f},\beta,a,b) \propto \prod_{i=1}^{N}\boldsymbol{\alpha}_i^a \exp\left(-\frac{1}{2}||\sqrt{\boldsymbol{\alpha}}\odot\mathbf{f}||^2-b\sum_{i=1}^{N}\boldsymbol{\alpha}_i\right) \tag{2.17}$$

$$p(\beta|\hat{\mathbf{f}},\mathbf{f},\boldsymbol{\alpha},c,d) \propto \beta^{M+c-1}\exp\left(-\beta\left[\frac{1}{2}||\hat{\mathbf{f}}-\mathbf{F}\mathbf{f}||^2+d\right]\right). \tag{2.18}$$

Therefore each latent variable can be sampled from the following distributions

$$\mathbf{f}|\hat{\mathbf{f}},\boldsymbol{\alpha},\beta \sim \mathcal{CN}\left((\beta\mathbf{F}^H\mathbf{F}+\mathrm{diag}(\boldsymbol{\alpha}))^{-1}\beta\mathbf{F}^H\hat{\mathbf{f}},(\beta\mathbf{F}^H\mathbf{F}+\mathrm{diag}(\boldsymbol{\alpha}))^{-1}\right) \tag{2.19}$$

$$\boldsymbol{\alpha}|\hat{\mathbf{f}},\mathbf{f},\beta,a,b \sim \Gamma\left(1+a,\frac{1}{2}\mathbf{f}\odot\bar{\mathbf{f}}+b\right) \tag{2.20}$$

$$\beta|\hat{\mathbf{f}},\mathbf{f},\boldsymbol{\alpha},c,d \sim \Gamma\left(M+c,\frac{1}{2}||\hat{\mathbf{f}}-\mathbf{F}\mathbf{f}||^2+d\right). \tag{2.21}$$

Section 2.2

Sampling-based SAR image formation

In this section, a sampling-based image formation procedure based on that of [8] is used to obtain approximate samples from each latent variable in (2.15). From these samples, various estimates and confidence statistics can be retrieved.

2.2.1. Efficient sampling method

Now that the joint posterior has been specified (2.15), it remains to be defined how to learn information about and interrogate it by efficiently gathering samples and later developing statistics. Clearly, the density (2.15) is of a complicated form that is not described by a known family of probability distributions. It is essentially the product of two Gaussian and two Gamma distributions. Therefore, (2.15) cannot be efficiently sampled directly. While a standard MCMC implementation like the

Metropolis-Hastings algorithm could be used to obtain approximate samples, we will use the following Gibbs sampler scheme to obtain approximate samples from (2.15), [8]. Therefore we can forego the discussion of MCMC algorithms like Metropolis-Hastings and focus on the Gibbs sampler, [37]. Using a Gibbs sampler, we are able to sample a large set of variables by sampling each variable in succession, [8]. That is, we obtain approximate samples of (2.15) by successively sampling the above individual posteriors. As with other MCMC methods, Gibbs sampling creates a Markov chain of samples, each of which is correlated with the other samples.

Efficiently sampling $\mathbf{f}$**.** In terms of computational efficiency, a potential issue occurs in sampling the individual posterior for $\mathbf{f}$ given by (2.19), where the following linear system must be solved for $\mathbf{f}$

$$\beta \mathbf{F}^H \mathbf{F}\mathbf{f} + \boldsymbol{\alpha} \odot \mathbf{f} = \beta \mathbf{F}^H (\hat{\mathbf{f}} + \mathbf{v}_1) + \mathbf{v}_2, \tag{2.22}$$

with $\mathbf{v}_1, \mathbf{v}_3 \sim \mathcal{N}(\mathbf{0}_M, \mathbf{I}_M/\sqrt{\beta})$ and $\mathbf{v}_2, \mathbf{v}_4 \sim \mathcal{N}(\mathbf{0}_{N/2}, \mathrm{diag}(\sqrt{\boldsymbol{\alpha}}))$, [8]. The way Fessler's NUFFT, [33], works is by interpolating non-uniform Fourier mode quantities to a uniform grid such that the uniform FFT can be used with zero-padding. This is called "gridding." The NUFFT is not without error of course, the main part of which comes from the error accumulated when interpolating non-uniform to uniform Fourier modes. Hence, the end result is actually the application of a uniform FFT, such that

$$\beta \mathbf{F}^H \mathbf{F}\mathbf{f} + \boldsymbol{\alpha} \odot \mathbf{f} \approx (\beta + \boldsymbol{\alpha}) \odot \mathbf{f} \tag{2.23}$$

That is, the LHS of (2.22) can be approximately diagonalized. Therefore approximately inversion only requires an elementwise division by $\beta + \boldsymbol{\alpha}$ which can be efficiently applied. Note this does not mean that this solve comes without error. There is still

error from modifying the non-uniform modes in order to make them conform with a uniform grid. A significantly less efficient but potentially more accurate method would be to use elementwise division by $\beta + \boldsymbol{\alpha}$ as a preconditioner in a conjugate gradient descent scheme. PCG, [56], is used to solve a similar linear system in [8]. The detailed algorithm is below.

An efficient MCMC method for sampling from $p(\mathbf{f}, \boldsymbol{\alpha}, \beta | \hat{\mathbf{f}}, a, b, c, d)$**.**

(a) Initiate $\mathbf{f}^0$, $\boldsymbol{\alpha}^0$, β^0. Choose a, b, c, d. Let $k = 0$;

(b) Compute $\mathbf{f}^{k+1} = (\beta^k \mathbf{F}^H(\hat{\mathbf{f}} + \mathbf{v}_1) + \mathbf{v}_2) \oslash (\beta^k + \boldsymbol{\alpha}^k)$;

(c) Compute $\boldsymbol{\alpha}^{k+1} \sim \Gamma\left(a + 1, \frac{1}{2}\mathbf{f}^{k+1} \odot \overline{\mathbf{f}^{k+1}} + b\right)$;

(d) Compute $\beta^{k+1} \sim \Gamma\left(M/2 + c, \frac{1}{2}||\hat{\mathbf{f}} - \mathbf{F}\mathbf{f}^{k+1}||^2 + d\right)$;

(e) Set $k = k + 1$ and return to Step (b).

The result of this algorithm is a group of k samples for each of the latent variables $\mathbf{f}$, β, and $\boldsymbol{\alpha}$, that are approximately drawn from the joint posterior (2.15). Each sample k requires just two NUFFT applications. This speed is of course relative, as NUFFT image formation takes only a single NUFFT application.

2.2.2. Chain convergence

It has been proven that the Gibbs sampling method above converges, [8], such that samples of each latent variable are approximately from the posterior (2.15). However, it is unknown how quickly it will converge. The following describes how we determine chain convergence.

A trace plot is often used to display the history of a parameter's samples, showing where the chain has been exploring. These time series of the individually sampled

parameters have been used to gauge chain convergence, [16]. In particular, the average value of a converged chain should have no long term trend, and samples should look like random noise. Colloquially this is referred to as "mixing well." However, displaying trace plots for $\sim 5 \times 10^5$ latent variables is not practical. Hence we use the following statistic from [8, 35] to determine chain convergence.

Computing the convergence statistic $\hat{R}$. In this approach, multiple chains are computed using randomly chosen starting points based on the observation that the variance within a single chain will converge faster than the variance between chains. A statistic is computed for each element of each latent variable, the value of which is a measure of convergence of that individual parameter. This statistic is computed as follows, and the derivation closely follows that in [8]. Compute n_r chains (in our implementation this is done in parallel) each of length $2n_s$, keeping only the latter n_s samples. Let ψ_{ij} denote the ith sample from the jth chain for a single parameter, and define

$$B = \frac{n_s}{n_r - 1} \sum_{j=1}^{n_r} \left(\bar{\psi}_{\cdot j} - \bar{\psi}_{\cdot\cdot}\right)^2, \quad \text{with} \quad \bar{\psi}_{\cdot j} = \frac{1}{n_s} \sum_{i=1}^{n_s} \psi_{ij}, \quad \text{and} \quad \bar{\psi}_{\cdot\cdot} = \frac{1}{n_r} \sum_{j=1}^{n_r} \bar{\psi}_{\cdot j};$$

and

$$W = \frac{1}{n_r} \sum_{j=1}^{n_r} s_j^2, \quad \text{with} \quad s_j^2 = \frac{1}{n_s - 1} \sum_{i=1}^{n_s} \left(\psi_{ij} - \bar{\psi}_{\cdot j}\right)^2$$

Notice that $\bar{\psi}_{\cdot j}$ is the mean of the samples in the chain j, and $\bar{\psi}_{\cdot\cdot}$ is the mean of the samples in every chain. Hence B is a measure of the variance between the chains while W is a measure of the variance within each individual chain. The marginal

posterior variance $\text{var}(\psi|\hat{\mathbf{f}})$ is then estimated by

$$\widehat{\text{var}}^{+}(\psi|\hat{\mathbf{f}}) = \frac{n_s - 1}{n_s} W + \frac{1}{n_s} B, \tag{2.24}$$

which is an unbiased estimate under stationarity, [35]. From this variance estimate, we compute the aforementioned desired statistic

$$\hat{R} = \sqrt{\frac{\widehat{\text{var}}^{+}(\psi|\hat{\mathbf{f}})}{W}}, \tag{2.25}$$

which tends to 1 from above as $n_s \to \infty$. Once $\hat{R}$ dips below 1.1 for all sampled parameters, the $n_s n_r$ samples can together be considered samples from the posterior (2.15), [35]. Other values can also be chosen as a tolerance for $\hat{R}$, [8]. We note that this is not the only statistic used to determine chain convergence. See, e.g. [1] for theoretical analysis of the Gibbs sampler for hierarchical Bayesian models.

2.2.3. Computing statistics from the samples

At this point, we have samples of $\mathbf{f}$, β, and $\boldsymbol{\alpha}$, from the posterior (2.15). Now we form statistics with these samples to summarize that complicated density. Perhaps the most obvious statistic to compute from the samples is the mean of $\mathbf{f}$, the noise precision β, and the speckle parameter $\boldsymbol{\alpha}$. For $\mathbf{f}$ and $\boldsymbol{\alpha}$, these will be images that can give information about objects and features and their locations within the image. For β, this statistic will be scalar. In addition, sorting the samples by pixel value and looking at the image formed by the median pixel value can also provide an estimate. To quantify uncertainty, viewing the variance or standard deviation can be useful in determining the range of possible values for each pixels. In addition to the mean and variance, computing Bayesian confidence intervals for each sampled parameters. In this uncertainty quantification method, the samples are again sorted by pixel values

and a confidence intervals is formed for each pixel. That is, sort the samples of each parameters from lowest to highest. Take the the 0.025 percentile pixel value and the 0.975 percentile pixel value. The interval between these two values represents a 95% confidence interval for the value of that parameter. In order to display this, samples are drawn uniformly from this interval for each pixel and displayed in a GIF, called a confidence image, [47]. While not thoroughly explored in this paper, it is suspected that these samples and their confidence images, explained later, can offer more information and answer downstream questions, e.g. about the support of the scene, [8]. With that being said, a few observations of this type are in fact made in the next chapters.

Chapter 3

Results

In this chapter, we provide an example that demonstrates the accuracy, efficiency, and robustness of the preceding sampling-based SAR image formation algorithm. Note that the ground truth reflectivity images is unknown, preventing the computation of standard error statistics such as the relative error. This is common even in synthetically-created SAR examples, where the true reflectivity is still unknown. Therefore, the unprecedented uncertainty quantification information this method provides is all the more valuable, as it is able to quantify how much we should trust pixel values and structures in the image even in the absence of ground truth. Throughout, reflectivity images $\mathbf{f}$ are displayed in decibels (dB):

$$20 \log_{10} \left(\frac{|\mathbf{f}|}{\max |\mathbf{f}|} \right). \tag{3.1}$$

To set a standard for images and also minimize clutter in the figures all images will be displayed with a minimum of -60 dB and maximum of 0 dB. Lesser or greater values are assigned the minimum or maximum. We begin with a specification of the data used in the image formation examples that follow.

Figure 3.1: Optical images of parking lot being imaged in GOTCHA dataset. Note scene contains a variety of calibration targets, such as primitive reflectors like the tophat shown, a Toyota Camry, forklift, and tractor.

Section 3.1

Data

The GOTCHA Volumetric Data Set 1.0 consists of SAR phase history data collected at X-band with a 640 MHz bandwidth with full azimuth coverage at 8 different elevation angles with full polarization, [20]. This is a real-world SAR dataset captured by the Air Force Research Laboratory. A plane carrying a sensor flew a roughly circular measurement flight around a parking lot near the Sensors Directorate Building at Wright-Patterson Air Force Base in Dayton, Ohio, and collected SAR phase history data each time. The parking lot contains various vehicle targets including civilian vehicles, construction vehicles, calibration targets, primitive reflectors, and military vehicles. See Figure 3.1 for optical images of the targets. Note that because this is real-world data, the elevation angle is not perfectly constant, and path is not perfectly circular. The center frequency is 9.6GHz and bandwidth is 640MHz. This public release data has been used extensively for testing new SAR image formation methods,

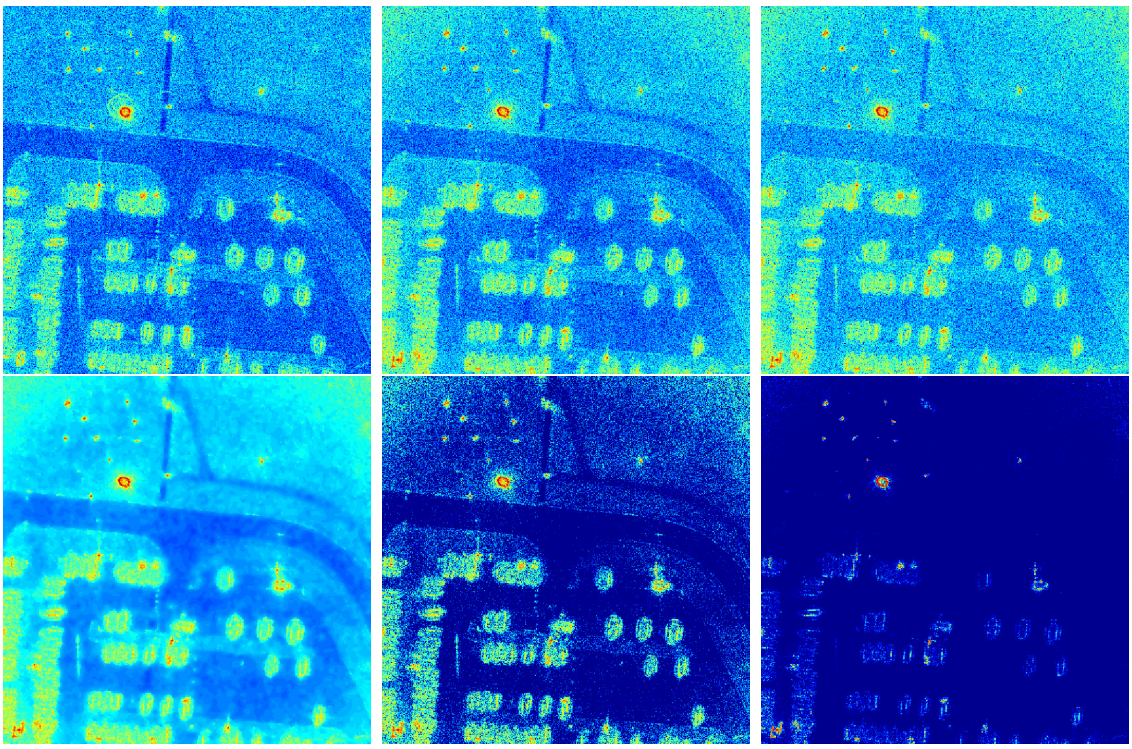

Figure 3.2: Full images formed with (clockwise from top left) BP; NUFFT; uninformative sample mean; sparse sample mean; ℓ_1 regularization; TV regularization.

[5, 6, 32].

Section 3.2

Example image estimates

The paper accompanying the GOTCHA data set, [20], indicates that research interest is focused on mitigating the large side lobes in the point spread function, since due to the sparse nature of the elevation aperture, traditional imaging techniques introduce excessive artifacts in the processed images. Hence, we have reason to believe that setting the hyperhyperparameters a, b c, and d, to be sparsity-encouraging may be useful. For this reason, in the following figures for comparison's sake we show images formed using two aforementioned popular sparsity-encouraging image formation methods: ℓ_1 regularization and total variation (TV) regularization. In ℓ_1

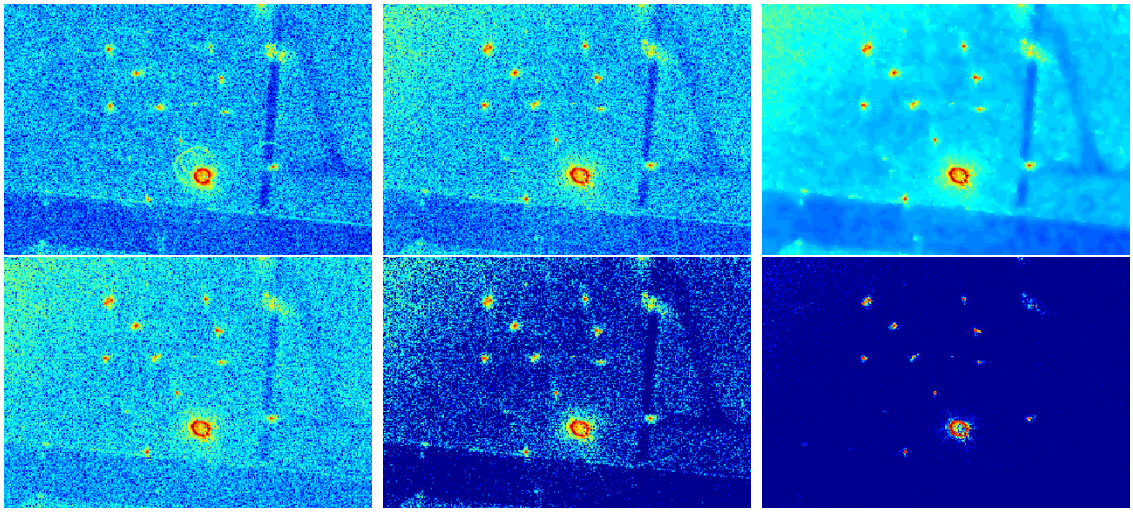

Figure 3.3: Subregion of images from Fig. 3.2 formed with (clockwise from top left) BP; NUFFT; uninformative sample mean; sparse sample mean; ℓ_1 regularization; TV regularization.

regularization, sparsity is promoted in the image itself, while in TV regularization, sparsity is promoted in the approximate edges in the image. These methods have been extensively applied in SAR. See, e.g., [3, 21, 28, 51, 53].

In Figure 3.2, a full image comparison of the parking lot scene is shown using back projection, NUFFT, ℓ_1 regularization, total variation (TV) regularization, and the novel sampling-based method with both uninformative and sparsity-encouraging parameters. The full images shown are square with $N = 512^2$. Code from [50] is used to perform image formation for the comparison methods, as well as to wrangle the GOTCHA data. Figures 3.3 and 3.4 zoom in on smaller subregions of the illuminated scene in order to see how each image formation method compares when localizing particular targets. The sampling-based method using uninformative parameters seems to provide perhaps the worst image quality, appearing much grainier than even the back projection and NUFFT reconstructions. While this is indeed considered a poor estimate, it may be worth considering that setting the parameters to be uninformative has simply brought out more signal altogether including speckle, which, despite our

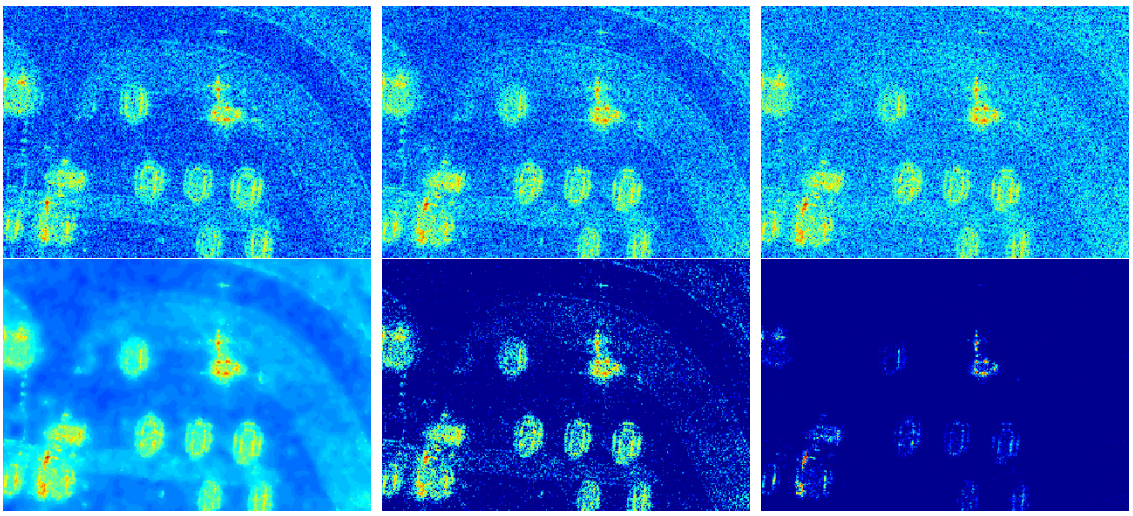

Figure 3.4: Another subregion of images from Fig. 3.2 formed with (clockwise from top left) BP; NUFFT; uninformative sample mean; sparse sample mean; ℓ_1 regularization; TV regularization.

NUFFT	**ℓ_1 regularization**	**uninformative sampling**	**sparse sampling**
.03s	5.8s	16s	526s

Table 3.1: Runtimes for each algorithm with $N = 512^2$.

wish to remove it, is indeed signal. On the other hand, the sampling-based method using sparsity-encouraging parameters appears to retrieve a significantly better estimate than the other methods in terms of sidelobe reduction, noise reduction, speckle reduction, and contrast improvement. In addition, there are no artifacts from the image formation method as seen, e.g., in the block-like TV regularization formation.

N	**uninformative sampling**	**sparse sampling**
128^2	33	517
256^2	35	896
512^2	41	1322

Table 3.2: Required chain length n_s for various N.

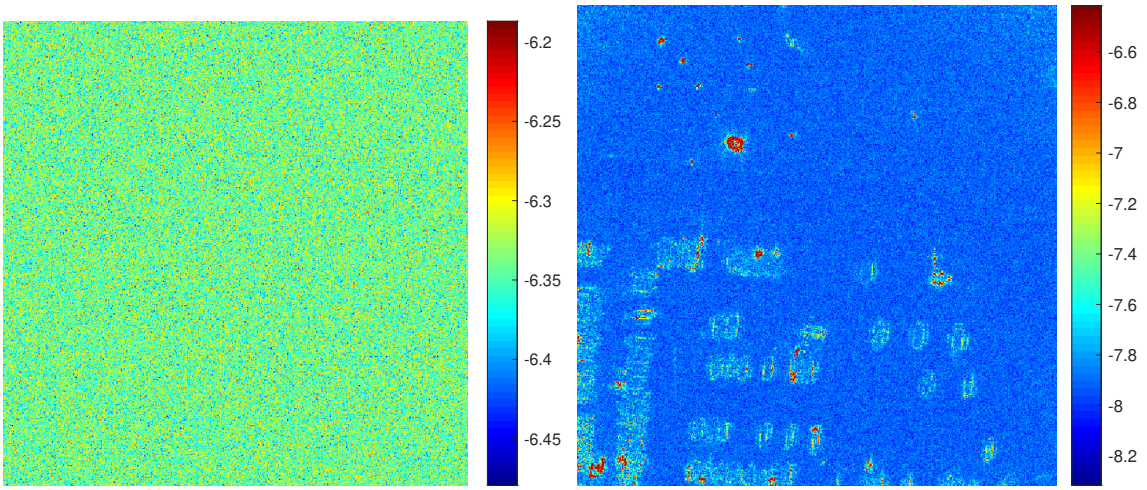

Figure 3.5: Logarithm base 10 of the sample variance of $\mathbf{f}$ using (left) uninformative parameters; (right) sparsity-encouraging parameters.

Table 3.1 gives the runtime for each algorithm. Each method was performed on Polaris, a shared memory computer operated by Dartmouth Research Computing with 40 cores, 64-bit Intel processors, and 1 TB of memory. Note that the back projection reconstruction has not been optimized for runtime. While only images with $N = 512^2$ are shown throughout this paper, converged chains were computed for other values and the required chain lengths are shown in Table 3.2. In particular, Tables 3.1 and 3.2 show that while parameter choice does not affect the time per sample, convergence takes significantly more samples, and hence significantly more time, when using the sparsity-encouraging parameters. In addition, Table 3.2 begins to demonstrate that the required chain length appears roughly linear, although more examination is needed. Hence this method requires large storage and memory.

Section 3.3

Visualizing uncertainty quantification

With the samples having been drawn, and image estimates computed, a way to display confidence information is required to provide valuable information about the certainty

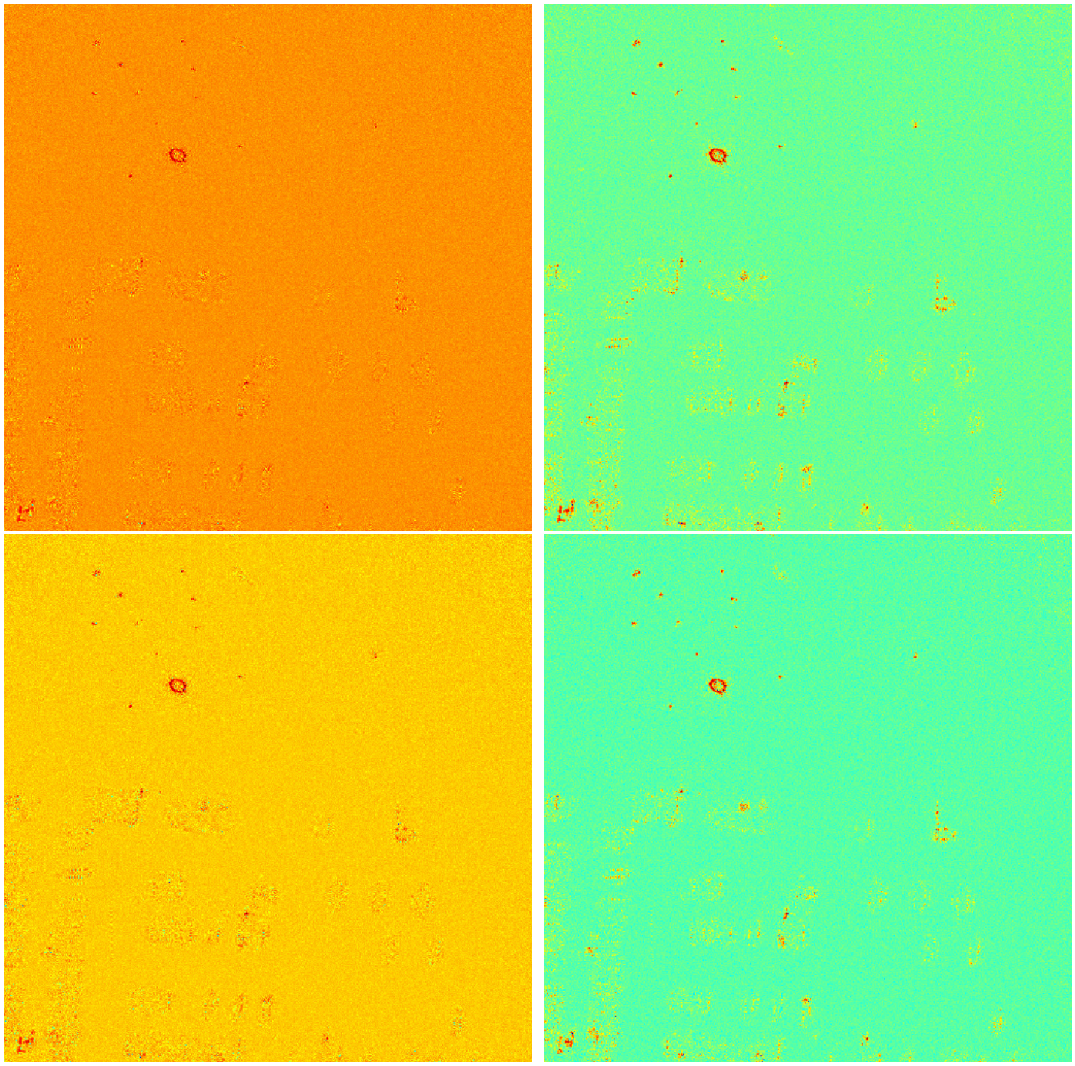

Figure 3.6: Comparison of the 95% confidence images. Top: 0.025 percentile image; bottom: 0.975 percentile image. Left: uninformative; right: sparsity.

of these estimates. Because this is an imaging application, this useful information needs to be displayed in a visibly tractable way. A confidence image, defined above in Section 2.2.3, gives a viewer visual information and insight into the uncertainties in the estimates, i.e. which features in the image can be trusted.

3.3.1. Variance images

A simple way to quantify uncertainty in this problem using the collected samples is to look at the variance of the samples (or standard deviation) at each pixel. This can

be helpful in forming a confidence estimate by acknowledging that roughly 2 standard deviations from the mean contains 95% of samples in a Gaussian distribution. Figure 3.5 shows the logarithm base 10 of the variance of the samples of $\mathbf{f}$ for the example from Section 3.2. Notice that for the uninformative parameters, the variance is roughly constant across all pixels, while for the sparsity-encouraging parameters, the variance away from scatterers is significantly lower for pixels with low-magnitude. Hence, the uninformative parameter plot says that high-magnitude pixels and low-magnitude pixels have the same spread, which is one reason why the corresponding image estimate appears so grainy. Therefore, this is further evidence beyond the estimate that the uninformative parameter sampling-based method is not terribly useful in this scenario. Meanwhile, the sparsity-encouraging parameter plot shows exactly what we would expect with multiple scales in a scene, that high-magnitude pixels tend to vary more than low-magnitude pixels.

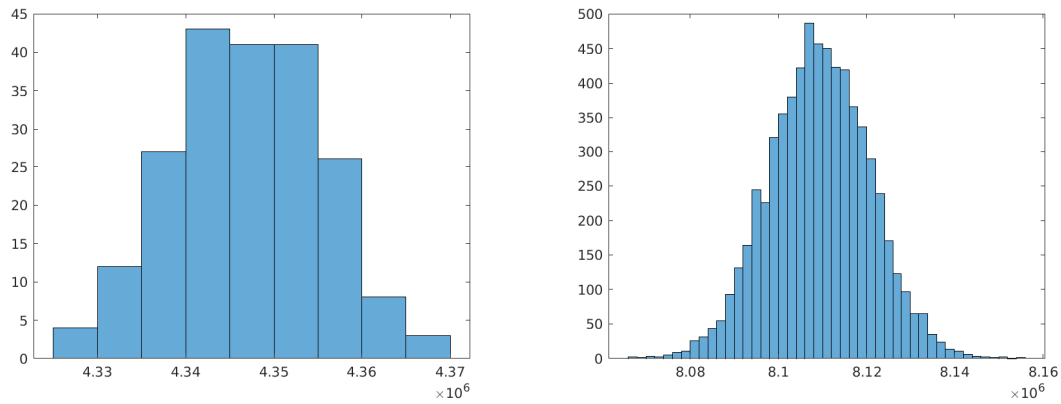

Figure 3.7: Histogram of samples of the noise precision β using (left) uninformative; (right) sparsity-encouraging parameters.

3.3.2. Confidence images

Another method to visualize uncertainty in $\mathbf{f}$ is to compute a confidence image, the process for which is given below. Visualizing samples of one-dimensional signal can

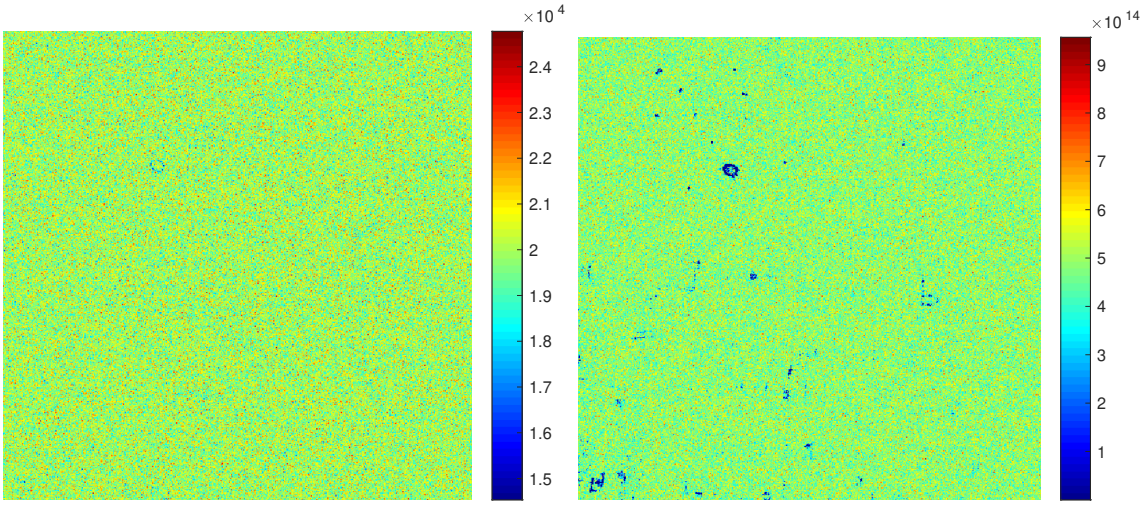

Figure 3.8: Comparison of the sample mean of $\boldsymbol{\alpha}$ for (left) uninformative parameters; (right) sparsity-encouraging parameters.

be done using, e.g., confidence intervals with error bars on the mean estimate, trace plots represented as error bars at each point of the signal, or histograms. E.g., a histogram for samples of the one-dimensional noise precision β is shown in Figure 3.7. For reference, the mean of the uninformative β samples is 2.3005×10^{-7}, and the mean of the sparsity β samples is 1.2331×10^{-7}. In many applications, a trace plot of the sample chain is shown to show a cursory level of convergence. However, for two-dimensional images the visualization of the chain variance of an image is less obvious. In [47], the authors develop Twinkle, a tool to visualize 2D confidence images. In Twinkle, the samples are sorted in increasing order and the 0.025 percentile value and the 0.975 percentile value are chosen as the lower and upper bounds for a 95% confidence interval at a particular pixel. Such an interval is computed for every pixel. Figure 3.6 shows the lower and upper bounds for the confidence images using both uninformative and sparsity-encouraging parameters. We see that the lower and upper confidence bounds for the sparsity-encouraging parameters appear quite similar in contrast and background noise level, indicating in general a tighter confidence interval than that of the uninformative parameter image which has a larger

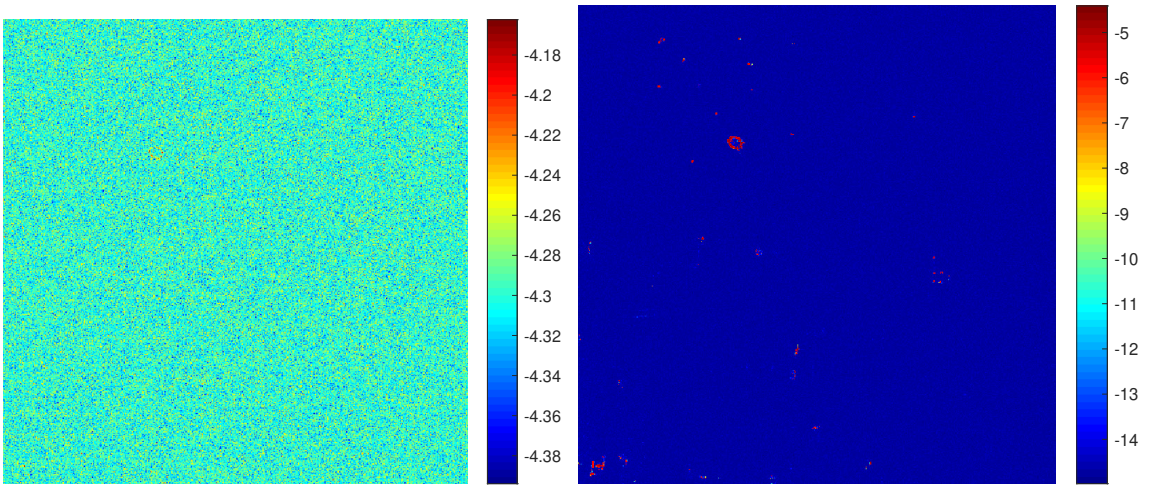

Figure 3.9: Comparison of the logarithm base 10 of the sample mean speckle parameter $\boldsymbol{\alpha}^{-1}$ for (left) uninformative parameters; (right) sparsity-encouraging parameters.

variance between the lower and upper confidence bounds. Then, image samples are formed by drawing pixel values uniformly at random from each of these confidence intervals. A GIF or short movie is then created from the image samples, showing them in quick succession for a fraction of a second each. The heuristic is that we can be more confident in features that persist in the image throughout the video, and less confident in features or pixel values in the image that flicker or twinkle. The latter could be real or attributable to an artifact or noise. As of now, GIFs are not easily embedded in LaTeX, so a GIF showing the twinkling effect for the example above is available at my website. Besides Twinkle, another reasonable way to view this type of information is to simply display the posterior samples themselves in a GIF or short movie.

Section 3.4

Estimating the speckle parameter

Similar to the image estimates obtained in Section 3.2, the sampling-based image formation method described in Chapter 2 produces samples of $\boldsymbol{\alpha}$, the parameter

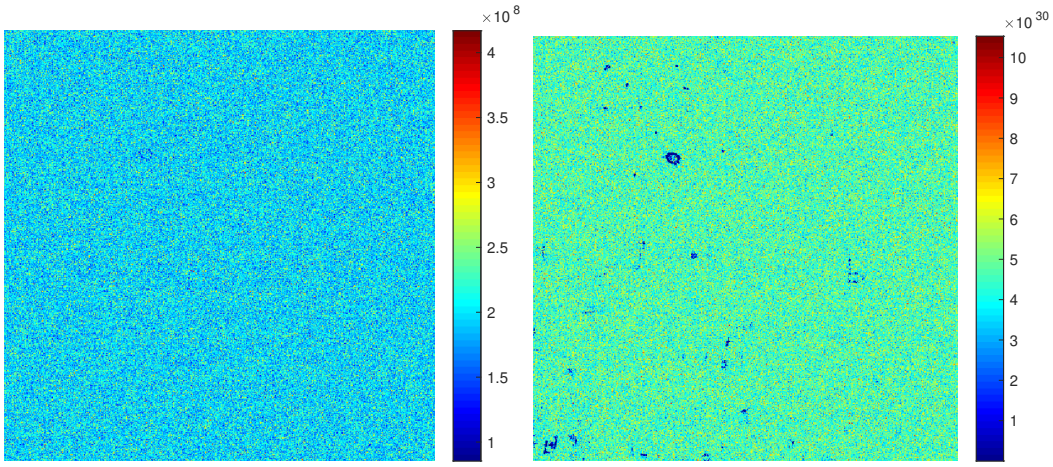

Figure 3.10: Sample variance of $\boldsymbol{\alpha}$ using (left) uninformative parameters; (right) sparsity-encouraging parameters.

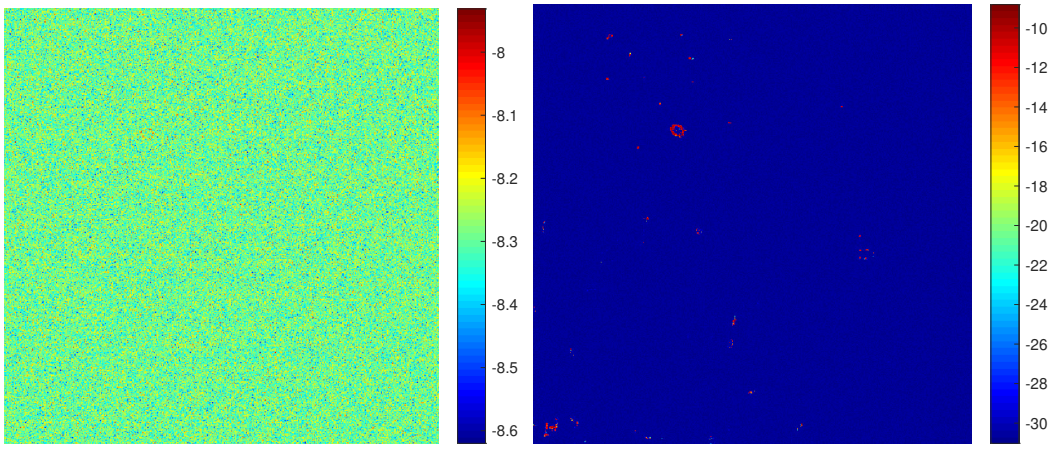

Figure 3.11: Logarithm base 10 of the sample variance of $\boldsymbol{\alpha}^{-1}$ using (left) uninformative parameters; (right) sparsity-encouraging parameters.

governing speckle. The mean of these samples is shown in Figure 3.8 for both the uninformative and sparsity-encouraging parameters. There are several observations that can be made from these images. Notice how in the uninformative parameter image even the perhaps most-prominent feature of the tophat is barely visible in the top left of the image, while in the sparsity-encouraging parameter image, many features in the reflectivity image are visible. Therefore, by looking at (2.11), we see that in a sense the uninformative parameters provide little spatially varying regularization while the

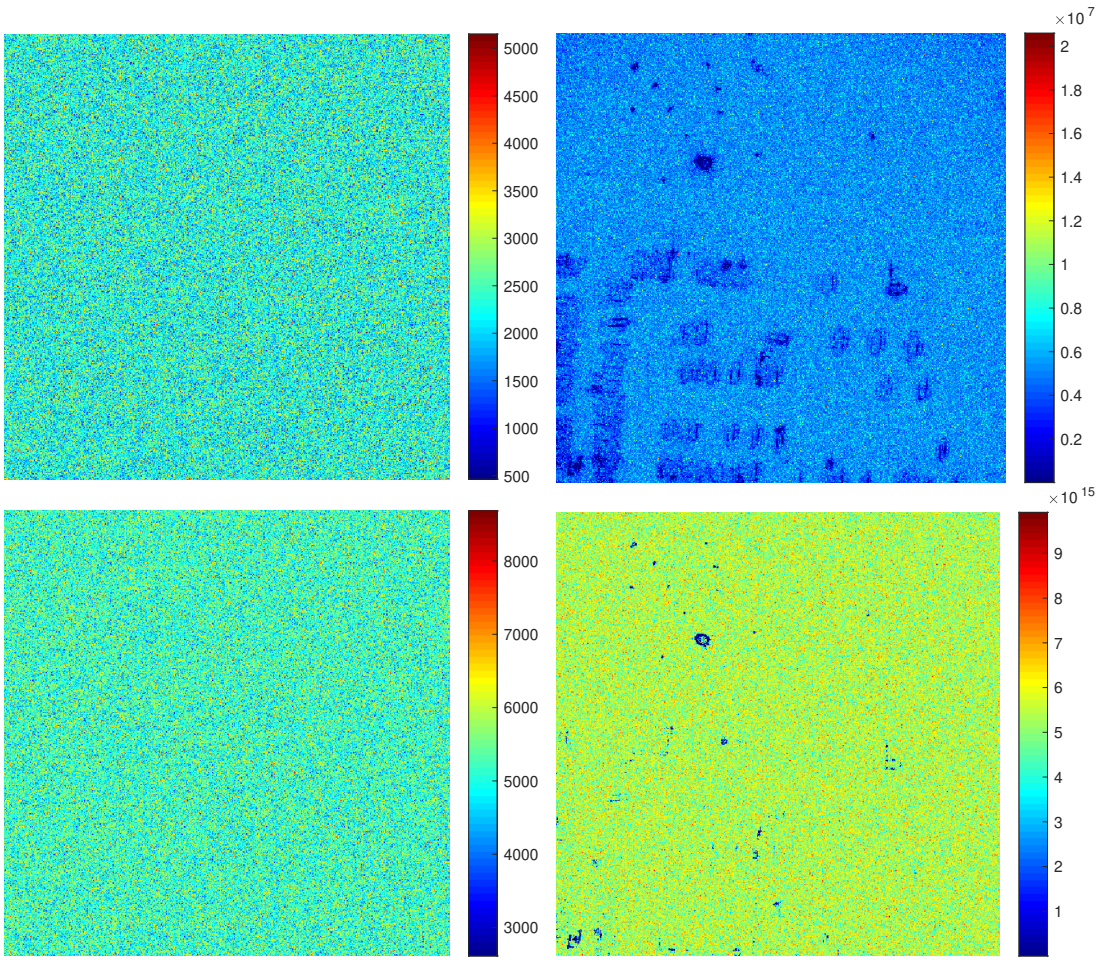

Figure 3.12: Comparison of the 95% confidence images for $\boldsymbol{\alpha}$. Top: 0.025 percentile image, bottom: 0.975 percentile image. Left: uninformative; right: sparsity.

sparsity-encouraging parameters provide quite sharp regularization only away from high-magnitude image features. In addition to providing heuristics about the success of this algorithm through the lens of deterministic regularization, we also have that the reciprocal values of this image provide an estimate for the mean speckle parameter. Recall that the magnitude of each pixel $|\mathbf{f}_i|$ is Rayleigh distributed with mean proportional to $\boldsymbol{\alpha}_i^{-1}$, hence changes in the magnitude of each pixel $|\mathbf{f}_i|$ are proportional to $\boldsymbol{\alpha}_i^{-1}$. Hence, since roughly all pixels in the uninformative parameter image are on the order of 10^{-4}, the speckle phenomenon is consistently high throughout the entire image. This is confirmed by checking the noisy image estimate in Figure

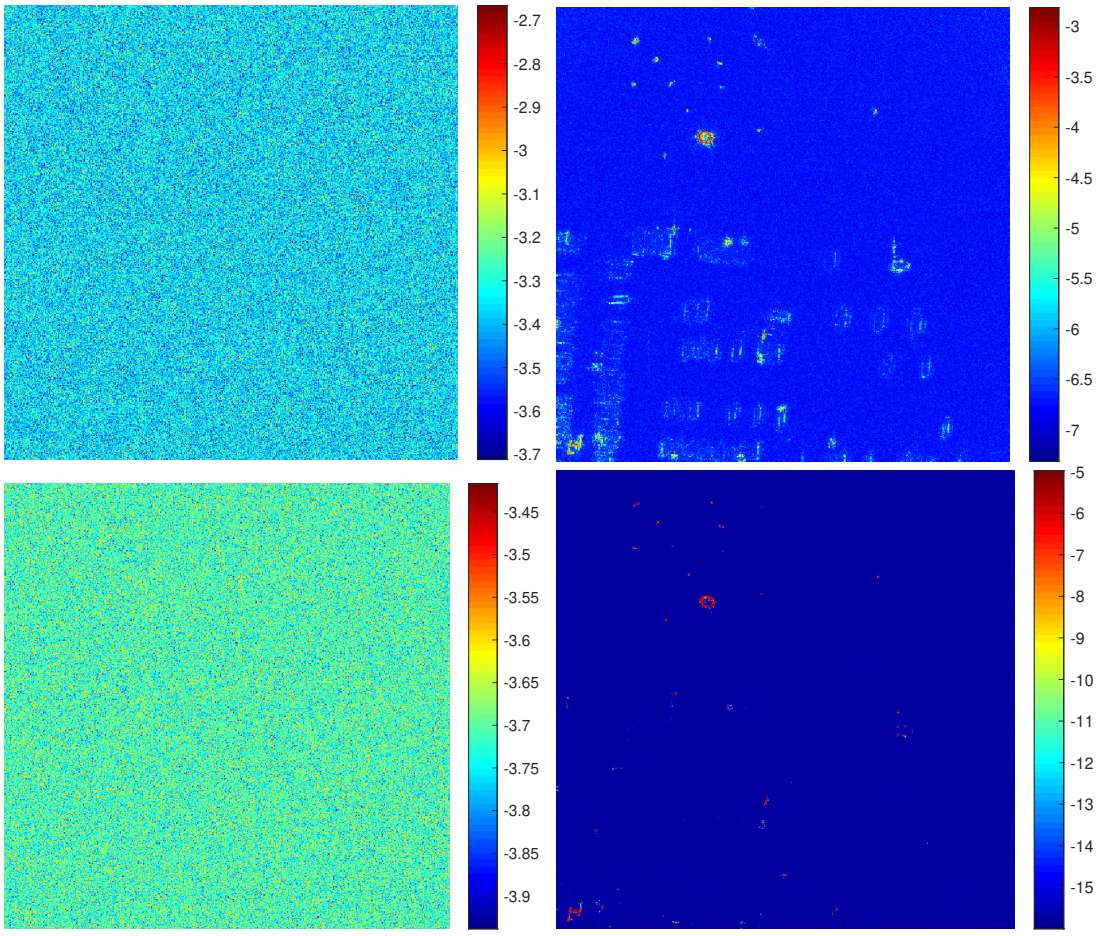

Figure 3.13: Comparison of the 95% confidence images for $\boldsymbol{\alpha}$. Top: 0.025 percentile image; bottom: 0.975 percentile image. Left: uninformative; right: sparsity.

3.2. Meanwhile for the sparsity-encouraging parameter image, there is practically no speckle (on the order 10^{-14}) except at the various high-magnitude target reflectivities, matching the speckle reduction we saw in Figures 3.2, 3.3, and 3.4. This confirms that the sparsity-encouraging measures taken effectively reduced speckle. The absent median images for $\boldsymbol{\alpha}$ and $\boldsymbol{\alpha}^{-1}$ support these conclusions as well.

In addition to these estimates, we can also perform similar uncertainty quantification analysis for the $\boldsymbol{\alpha}$ as in Sections 3.3.1 and 3.3.2 by looking at the variance and confidence images of the samples. Figure 3.10 shows the sample variance of the prior precision (or regularization matrix) $\boldsymbol{\alpha}$ for both sampling schemes, while Figure

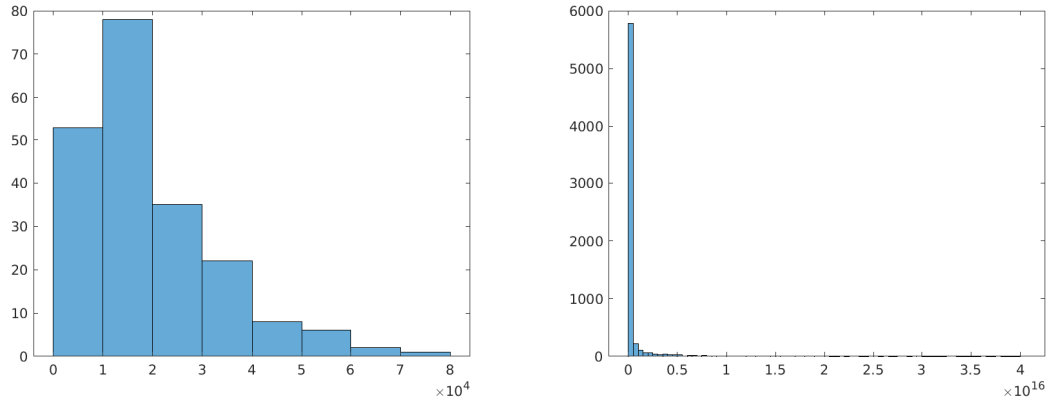

Figure 3.14: Histogram of samples for $\boldsymbol{\alpha}$ using (left) uninformative; (right) sparse sampling.

3.11 shows the same for log of the speckle parameter $\boldsymbol{\alpha}^{-1}$. In both cases, we see that the plots indicate a tighter confidence range, particularly in regions away from high-magnitude points, for the sparsity-encouraging parameters over the uninformative parameters. Similar conclusions can be drawn from the lower and upper confidence interval images shown for $\boldsymbol{\alpha}$ in Figure 3.12, and $\log_{10}(\boldsymbol{\alpha}^{-1})$ in Figure 3.13. Finally, Figure 3.14 shows a histogram plot of the first pixel of each $\boldsymbol{\alpha}$ sample. We see that the maximum of the uninformative sampling strategy is greater than 1, while for the sparse sampling strategy is very nearly 0.

Chapter 4

Discussion

In this final chapter, we begin by discussing and expanding upon the image estimate results of Chapter 3, and move on to summarize this paper and outline future research directions.

Section 4.1

Sparsity perspectives

The use of sparsity-encouraging methods in SAR image formation is not new. See, e.g., [3, 21, 28, 51, 53]. Nor is use of a hierarchical Bayesian model to encourage such sparsity, [15, 14]. These approaches are limited to MAP estimates, however, and limited to viewing sparsity in the magnitude of the image rather than in the real and imaginary parts as is done above.

4.1.1. Draws from individual posteriors

Of significance in discussing sparsity is Figure 4.1, which shows a histograms of values for the real and imaginary parts of the first pixel of each sample. Notice how the uninformative parameter histograms on the left look Gaussian (albeit there aren't many samples), while on the right the sparsity-encouraging parameter histograms

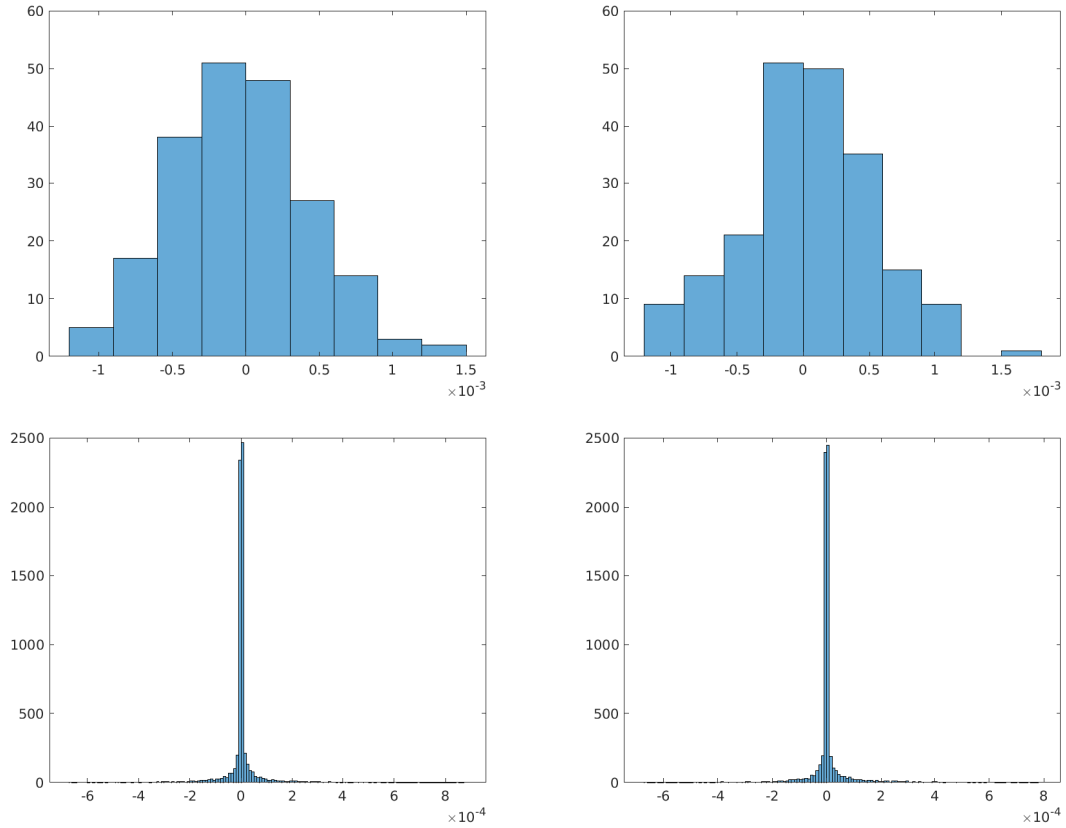

Figure 4.1: Histogram comparison of samples for the first pixel of the image $\mathbf{f}$. Top is uninformative parameters, bottom is sparsity parameters. Left is real part samples, right is imaginary part samples.

are sharply peaked at zero. This is the manifestation of the sparsity encouraging improper prior $p(\mathbf{f}_i) \approx 1/|\mathbf{f}_i|$ when $a, b, c, d = 0$.

4.1.2. Relation to sparsity-encouraging regularization schemes; IRLS

In our scheme above, $\boldsymbol{\alpha}$ is effectively a regularization parameter that is chosen by the data in order to encourage sparsity. This idea is not new, and is common is iterative reweighting schemes. See, e.g. [19, 22]. In iteratively reweighted least squares, e.g., the goal is to encourage sparsity in an unknown $\mathbf{x}$ by solving the ℓ_p minimization

problem

$$\arg\min_{\mathbf{x}} \left\{ ||\mathbf{x}||_p^p \right\} \quad \text{subject to} \quad \mathbf{A}\mathbf{x} = \mathbf{b}. \tag{4.1}$$

This is achieved by replacing the unweighted ℓ_p norm with a weighted ℓ_2 norm

$$\arg\min_{\mathbf{x}} \left\{ ||\mathbf{w} \odot \mathbf{x}||_2^2 \right\} \quad \text{subject to} \quad \mathbf{A}\mathbf{x} = \mathbf{b}, \tag{4.2}$$

where $\mathbf{w}$ is a weighting vector and $\odot$ is the elementwise product. This is similar to in our scheme where we effectively have an individual regularization parameter for each element of $\mathbf{x}$. The weighting vector is set with the update rule

$$\mathbf{w}_i = \left(\left(\mathbf{x}_i^{k-1} \right)^2 + \epsilon^k \right)^{p/2-1} \tag{4.3}$$

for $0 \leq p \leq 1$. Note that $p = 0$ corresponds to a cost function $\sum_i \log(|\mathbf{x}_i|^2)$, which is actually the same regularization that occurs when using the improper prior $1/|\mathbf{f}_i|$ above. The regularization with ϵ^k keeps the algorithm from dividing by zero if the previous iterate of $\mathbf{x}$ is zero. This scheme has been used in ill-posed inverse problems typically in a manner where this problem is repeatedly solved and weights are updated at each step in order to encourage sparsity, [22, 27, 38]. We believe that the advantage of our algorithm is the uncertainty quantification information and provided, as well as the theoretical underpinning of Bayesian inference. Indeed, as above with ℓ_1 and TV regularization used in SAR with an approximation to $|\mathbf{f}|$, because data information is considered in developing $\mathbf{w}$, this cost function does not correspond to putting a prior distribution on $\mathbf{x}$. In our scheme, we have more appropriately put a prior on $\boldsymbol{\alpha}$ (similar to $\mathbf{w}$ here), as it is unknown and needs to be included in the inference problem.

For further insight, we show below that when $\mathbf{w}$ is updated at each iteration it is essentially treated $\mathbf{w}$ as another variable that can be minimized over. So we solve

$$\mathbf{x}^*, \mathbf{w}^* = \arg\min_{\mathbf{x},\mathbf{w}} L(\mathbf{x}, \mathbf{w}) = \arg\min_{\mathbf{x},\mathbf{w}} \left\{ ||\mathbf{A}\mathbf{x} - \mathbf{b}||_2^2 + ||\mathbf{w} \odot \mathbf{x}||_2^2 \right\}. \tag{4.4}$$

That is, we minimize the cost function over $\mathbf{x}$ and $\mathbf{w}$. Taking an alternating minimization approach to this problem, after initializing $\mathbf{x}^0$ and $\mathbf{w}^0$, we minimize with respect to $\mathbf{x}$ and then $\mathbf{w}$. The cost function $L(\mathbf{x}, \mathbf{w})$ is differentiable, so

$$\nabla_{\mathbf{x}} L(\mathbf{x}, \mathbf{w}) = \mathbf{A}^T(\mathbf{A}\mathbf{x} - \mathbf{b}) + \mathbf{w} \odot \mathbf{w} \odot \mathbf{x} \tag{4.5}$$

$$\nabla_{\mathbf{w}} L(\mathbf{x}, \mathbf{w}) = \mathbf{x} \odot \mathbf{x} \odot \mathbf{w}. \tag{4.6}$$

Setting $\nabla_{\mathbf{x}} L(\mathbf{x}, \mathbf{w})$ equal to zero, the update for $\mathbf{x}$ is

$$\mathbf{x}^k = (\mathbf{A}^T\mathbf{A} + \mathbf{w}^k \odot \mathbf{w}^k)^{-1}\mathbf{A}^T\mathbf{b}. \tag{4.7}$$

Setting $\nabla_{\mathbf{w}} L(\mathbf{x}, \mathbf{w})$ to zero, however, results in $\mathbf{w} = \mathbf{0}$. So it is clear that this alternating minimization approach yields a trivial solution for $\mathbf{w}$, such that no regularization is applied in the original inverse problem.

We see at this point the modifications from IRLS, [22, 27], make this problem solvable. E.g., rather than setting $\nabla_{\mathbf{w}} L(\mathbf{x}, \mathbf{w})$ to zero, we can set it equal to 1 and retrieve the update for IRLS with $p = 0$, albeit without ϵ-regularization,

$$\mathbf{w}_i^k = \left(\left(\mathbf{x}_i^{k-1} \right)^2 \right)^{-1}. \tag{4.8}$$

Other modifications would retrieve the updates for $0 < p \leq 1$. An alternative approach is to use binary weighting: $\mathbf{w}_i^k = 0$ if $(\mathbf{x}_i^{k-1})^2 > \epsilon$, else $\mathbf{w}_i = 1$. This was the approach of our previous work, [23, 26].

Section 4.2

Conclusions

In this paper, we developed a new framework for coherent SAR image formation. This is a challenging task due to the problem size, the speckle phenomenon, and the lack of uncertainty quantification in current methods. This framework uses a hierarchical Bayesian model with conjugate priors to directly incorporate fully-developed speckle. A parameter-free sparsity-encouraging sampling method is introduced to provide estimates of the image, the speckle, and the noise. Our example from the GOTCHA data set shows that in real-world applications our method reduces speckle and noise significantly more than other commonly used methods. Uncertainty quantification information unprecedented in SAR is also provided in the form of confidence images that indicate if the pixel values and features shown in an estimate can be trusted. Uncertainty quantification is also provided for the speckle and noise. This information is of particular importance in SAR, where ground truth images even for synthetically-created examples are unknown.

Future work in this direction will focus on further accelerating the sampling method, as well as decreasing storage and memory requirements. This will enable image formation with more pixels, as well as multi-pass and three-dimensional imaging. In addition, this will allow composite image formation for wide angle SAR to complement the strong direct imaging results of this paper. In composite image formation, the 360^o aperture is broken up into small overlapping chunks. An image is formed for each chunk, and those chunk images are then combined by taking the maximum pixel value of the chunks. Hence if there are 60 chunks, the current speed of this method this would be very costly. Finally, comparisons with deep learning based SAR image formation (still in its nascent stages, [45]) will be necessary.

Bibliography

[1] Sergios Agapiou, Johnathan M Bardsley, Omiros Papaspiliopoulos, and Andrew M Stuart, *Analysis of the Gibbs sampler for hierarchical inverse problems*, SIAM/ASA Journal on Uncertainty Quantification **2** (2014), no. 1, 511–544.

[2] Fredrik Andersson, Randolph Moses, and Frank Natterer, *Fast Fourier methods for synthetic aperture radar imaging*, IEEE Transactions on Aerospace and Electronic Systems **48** (2012), no. 1, 215–229.

[3] 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.

[4] Fabrizio Argenti, Alessandro Lapini, Tiziano Bianchi, and Luciano Alparone, *A tutorial on speckle reduction in synthetic aperture radar images*, IEEE Geoscience and Remote Sensing Magazine **1** (2013), no. 3, 6–35.

[5] 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.

[6] Christian David Austin, *Sparse methods for model estimation with applications to radar imaging*, Ph.D. thesis, The Ohio State University, 2012.

[7] Johnathan Bardsley and Tiangang Cui, *Optimization-based MCMC methods for nonlinear hierarchical statistical inverse problems*, arXiv preprint arXiv:2002.06358 (2020).

[8] Johnathan M Bardsley, *MCMC-based image reconstruction with uncertainty quantification*, SIAM Journal on Scientific Computing **34** (2012), no. 3, A1316–A1332.

[9] ———, *Computational uncertainty quantification for inverse problems*, vol. 19, SIAM, 2018.

[10] Johnathan M Bardsley and Tiangang Cui, *A Metropolis-Hastings-within-Gibbs sampler for nonlinear hierarchical-Bayesian inverse problems*, 2017 MATRIX Annals, Springer, 2019, pp. 3–12.

[11] Johnathan M Bardsley, Aku Seppanen, Antti Solonen, Heikki Haario, and Jari Kaipio, *Randomize-then-optimize for sampling and uncertainty quantification in electrical impedance tomography*, SIAM/ASA Journal on Uncertainty Quantification **3** (2015), no. 1, 1136–1158.

[12] Johnathan M Bardsley, Antti Solonen, Heikki Haario, and Marko Laine, *Randomize-then-optimize: A method for sampling from posterior distributions in nonlinear inverse problems*, SIAM Journal on Scientific Computing **36** (2014), no. 4, A1895–A1910.

[13] 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.

[14] 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 (2019).

[15] Daniela Calvetti, E Somersalo, and A Strang, *Hierachical Bayesian models and sparsity: ℓ_2-magic*, Inverse Problems **35** (2019), no. 3, 035003.

[16] Daniela Calvetti and Erkki Somersalo, *An introduction to Bayesian scientific computing: Ten lectures on subjective computing*, vol. 2, Springer Science & Business Media, 2007.

[17] ———, *Hypermodels in the Bayesian imaging framework*, Inverse Problems **24** (2008), no. 3, 034013.

[18] 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.

[19] 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.

[20] Curtis H Casteel Jr, LeRoy A Gorham, Michael J Minardi, Steven M Scarborough, Kiranmai D Naidu, and Uttam K Majumder, *A challenge problem for 2D/3D imaging of targets from a volumetric data set in an urban environment*, Algorithms for Synthetic Aperture Radar Imagery XIV, vol. 6568, International Society for Optics and Photonics, 2007, p. 65680D.

[21] Müjdat Çetin and William Clement Karl, *Feature-enhanced synthetic aperture radar image formation based on nonquadratic regularization*, IEEE Transactions on Image Processing **10** (2001), no. 4, 623–631.

[22] Rick Chartrand and Wotao Yin, *Iteratively reweighted algorithms for compressive sensing*, IEEE International Conference on Acoustics, Speech and Signal Processing, 2008., IEEE, 2008, pp. 3869–3872.

[23] Victor Churchill, Rick Archibald, and Anne Gelb, *Edge-adaptive ℓ_2 regularization image reconstruction from non-uniform Fourier data*, Inverse Problems and Imaging **13** (2019), no. 5, 931–958.

[24] Victor Churchill and Anne Gelb, *Estimation and uncertainty quantification for piecewise smooth signal recovery*, Submitted to Journal of Scientific Computing.

[25] ———, *Detecting edges from non-uniform Fourier data via sparse Bayesian learning*, Journal of Scientific Computing (2019), 1–22.

[26] ———, *Edge-masked CT image reconstruction from limited data*, 15th International Meeting on Fully Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine, vol. 11072, International Society for Optics and Photonics, 2019, p. 110721V.

[27] Ingrid Daubechies, Ronald DeVore, Massimo Fornasier, and C Sinan Güntürk, *Iteratively reweighted least squares minimization for sparse recovery*, Communications on Pure and Applied Mathematics: A Journal Issued by the Courant Institute of Mathematical Sciences **63** (2010), no. 1, 1–38.

[28] Xiao Dong and Yunhua Zhang, *SAR image reconstruction from undersampled raw data using maximum a posteriori estimation*, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing **8** (2014), no. 4, 1651–1664.

[29] Alok Dutt and Vladimir Rokhlin, *Fast Fourier transforms for nonequispaced data*, SIAM Journal on Scientific Computing **14** (1993), no. 6, 1368–1393.

[30] ———, *Fast Fourier transforms for nonequispaced data, II*, Applied and Computational Harmonic Analysis **2** (1995), no. 1, 85–100.

[31] Michael Elad, Peyman Milanfar, and Ron Rubinstein, *Analysis versus synthesis in signal priors*, Inverse Problems **23** (2007), no. 3, 947.

[32] Emre Ertin, Christian D Austin, Samir Sharma, Randolph L Moses, and Lee C Potter, *GOTCHA experience report: Three-dimensional SAR imaging with complete circular apertures*, Algorithms for Synthetic Aperture Radar Imagery XIV, vol. 6568, International Society for Optics and Photonics, 2007, p. 656802.

[33] Jeffrey A Fessler and Bradley P Sutton, *Nonuniform fast Fourier transforms using min-max interpolation*, IEEE Transactions on Signal Processing **51** (2003), no. 2, 560–574.

[34] Giorgio Franceschetti and Riccardo Lanari, *Synthetic aperture radar processing*, CRC press, 1999.

[35] Andrew Gelman, John B Carlin, Hal S Stern, David B Dunson, Aki Vehtari, and Donald B Rubin, *Bayesian data analysis*, Chapman and Hall/CRC, 2013.

[36] ———, *Bayesian data analysis*, vol. 2, CRC press Boca Raton, FL, 2014.

[37] Stuart Geman and Donald Geman, *Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images*, IEEE Transactions on Pattern Analysis and Machine Intelligence (1984), no. 6, 721–741.

[38] James E Gentle, *Matrix algebra*, Springer Texts in Statistics, Springer, New York, NY **10** (2007), 978–0.

[39] Walter G Goodman, Ron S Carrara, and Ronald M Majewski, *Spotlight synthetic aperture radar signal processing algorithms*, Artech House (1995), 245–285.

[40] LeRoy A Gorham and Linda J Moore, *SAR image formation toolbox for MATLAB*, Algorithms for Synthetic Aperture Radar Imagery XVII, vol. 7699, International Society for Optics and Photonics, 2010, p. 769906.

[41] Leslie Greengard and June-Yub Lee, *Accelerating the nonuniform fast Fourier transform*, SIAM Review **46** (2004), no. 3, 443–454.

[42] 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*, Springer Science & Business Media, 2012.

[43] Jari Kaipio and Erkki Somersalo, *Statistical and computational inverse problems*, vol. 160, Springer Science & Business Media, 2006.

[44] June-Yub Lee and Leslie Greengard, *The type 3 nonuniform FFT and its applications*, Journal of Computational Physics **206** (2005), no. 1, 1–5.

[45] Eric Mason, Bariscan Yonel, and Birsen Yazici, *Deep learning for SAR image formation*, Algorithms for Synthetic Aperture Radar Imagery XXIV, vol. 10201, International Society for Optics and Photonics, 2017, p. 1020104.

[46] 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.

[47] James G Nagy and Dianne P O'Leary, *Image restoration through subimages and confidence images*, Electronic Transactions on Numerical Analysis **13** (2002), 22–37.

[48] Radford M Neal, *Bayesian learning for neural networks*, vol. 118, Springer Science & Business Media, 2012.

[49] Chris Oliver and Shaun Quegan, *Understanding synthetic aperture radar images*, SciTech Publishing, 2004.

[50] Toby Sanders, *MATLAB imaging algorithms: Image reconstruction, restoration, and alignment, with a focus in tomography*, http://www.toby-sanders.com/software, https://doi.org/10.13140/RG.2.2.33492.60801.

[51] Toby Sanders, Anne Gelb, and Rodrigo B Platte, *Composite SAR imaging using sequential joint sparsity*, Journal of Computational Physics **338** (2017), 357–370.

[52] Theresa Scarnati, *Recent techniques for regularization in partial differential equations and imaging*, Arizona State University, 2018.

[53] Theresa Scarnati and Anne Gelb, *Joint image formation and two-dimensional autofocusing for synthetic aperture radar data*, Journal of Computational Physics **374** (2018), 803–821.

[54] Andrew M Stuart, *Inverse problems: a Bayesian perspective*, Acta numerica **19** (2010), 451–559.

[55] Michael E Tipping, *Sparse Bayesian learning and the relevance vector machine*, Journal of Machine Learning Research **1** (2001), no. Jun, 211–244.

[56] Curtis R Vogel, *Computational methods for inverse problems*, vol. 23, SIAM, 2002.

[57] Zheng Wang, Tiangang Cui, Johnathan Bardsley, and Youssef Marzouk, *Scalable optimization-based sampling on function space*, arXiv preprint arXiv:1903.00870 (2019).