Applied Mathematics for Modern Challenges

doi:10.3934/ammc.2023005

LEVERAGING JOINT SPARSITY IN 3D SYNTHETIC APERTURE RADAR IMAGING

Dylan Green[*1], JR Jamora[2] and Anne Gelb[1]

[1]Department of Mathematics, Dartmouth College, Hanover, NH, USA

[2]The Air Force Research Laboratory, Wright-Patterson AFB, OH, USA

(Communicated by Jennifer Mueller)

Abstract. Three-dimensional (3D) synthetic aperture radar (SAR) imaging is an active and growing field of research with various applications in both military and civilian domains. Sparsity promoting computational inverse methods have proven to be effective in providing point estimates for the volumetric image. Such techniques have been enhanced by leveraging sequential joint sparsity information from nearby aperture windows. This investigation extends these ideas by introducing a Bayesian volumetric approach that leverages the assumption of sequential joint sparsity. In addition to obtaining a point estimate, our new approach also enables uncertainty quantification. As demonstrated in simulated experiments, our approach compares favorably to currently used methodology for point estimate approximations, and has the additional advantage of providing uncertainty quantification for two-dimensional projections of the volumetric image.

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 [10, 17, 27, 30]. These techniques all enable reconstructions of large scenes with high resolution and fidelity.

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.

2020 *Mathematics Subject Classification.* 15A29, 62F15, 65F22, 94A12.

Key words and phrases. Synthetic aperture radar, Bayesian inference, inverse problems, regularization, joint sparsity.

This work is partially supported through the Autonomy Technology Research Center by AFRL contracts #FA8650-18-2-1645 (DG & JJ) and # FA8650-22-C-1017 (DG & JJ), and by the NSF grant DMS #1912685 (AG), AFOSR grant #F9550-22-1-0411 (DG & AG), DOE ASCR #DE-AC05-0 0OR22725 (AG), and ONR MURI grant #N00014-20-1-2595 (AG).

*Corresponding author: Dylan Green.

These approaches hold the promise of enhanced target recognition and identification, detailed topographic maps, and improved change detection, [1, 2, 7, 33].

In this work we build on two particular methods of 3D SAR image reconstruction, namely backprojection and sparsity-promoting recovery, [19, 32]. 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, [5, 22]. 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, [32], 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 paper is organized as follows. In Section 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 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 4. Section 5 contains some numerical experiments, and we provide concluding remarks in 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, [25]. In a SAR system that emits linear frequency modulated chirps, the frequency function $\omega(t)$ is given by

$$\omega(t) = \frac{2}{c}[\omega_0 + 2\alpha(t - \tau_0)], \quad |\omega(t)| \leq 2/c(\omega_0 + \alpha T_t), \tag{1}$$

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 (1) by

$$\vec{k} = [k_x, k_y, k_z]^T = \omega(t)[\cos(\theta(\tau))\cos(\varphi(\tau)), \sin(\theta(\tau))\cos(\varphi(\tau)), \sin(\varphi(\tau))]^T. \tag{2}$$

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 of the true underlying scene, $g : \mathbb{R}^3 \to \mathbb{C}$, and is given by [18]

$$\hat{g}(\vec{k}) = \mathcal{F}^3[g](\vec{k}) = \int_{\mathbb{R}^3} g(\vec{x}) \exp\left\{-i\left[\vec{x}^T\vec{k}\right]\right\} \mathrm{d}\vec{x}, \quad \vec{x} = [x, y, z]^T. \tag{3}$$

A more thorough introduction to SAR from a mathematical perspective can be found in [8, 9].

2.1. The Fourier Slice Theorem.

The Fourier Slice Theorem, [26], provides a convenient way to model the collected data. Let $f : \mathbb{R}^3 \to \mathbb{C}$. We are interested in a specific 2D slice of f, where the slice operator is defined by

$$S_2^3[h](\mu, \nu) = h(\mu, \nu, 0). \tag{4}$$

Observe that S_2^3 reduces the dimensionality of f to $\mathbb{R}^2$.

Remark 2.1. Our use of generic variables in (4) is intentional. Theorem 2.3 (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 (4) in the physical domain for the volumetric SAR reconstruction using 2D filtered backprojection (see Section 3.2.2).

We will also make use of the following definition:

Definition 2.2. 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 $\mathcal{S}_2^3[g] = \mathcal{S}_2^3[f_{\mathcal{B}}]$ 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

$$P_2^3[h](\mu,\nu) = \int_{-\infty}^{\infty} h(\mu,\nu,\xi)\mathrm{d}\xi. \tag{5}$$

Theorem 2.3 relates (3), (5), and (4) in the context of the rotational change of basis $\mathcal{B}$ and provides the foundation for the filtered backprojection algorithm.

Theorem 2.3 (The Fourier Slice Theorem)**.** *Suppose we are given $\vec{k}$ and $\vec{x}$ in (2), $\hat{g}(\vec{k})$ in (3), P_2^3 in (5), and S_2^3 in (4). Let $\mathcal{B}$ be an arbitrary rotational change of basis given in Definition 2.2. 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.*

$$S_2^3[\hat{g}_{\mathcal{B}}] = \mathcal{F}^2 \circ P_2^3[g_{\mathcal{B}}]. \tag{6}$$

Proof. Due to the rotation property of Fourier transforms, [18, 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 (3) we then have

$$\begin{aligned}
S_2^3[\hat{g}](k_x,k_y) &= \hat{g}(k_x,k_y,0) \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y,z)\exp\{-i(xk_x+yk_y)\}\,\mathrm{d}x\mathrm{d}y\mathrm{d}z \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left[\int_{-\infty}^{\infty} g(x,y,z)\mathrm{d}z\right]\exp\{-i(xk_x+yk_y)\}\,\mathrm{d}x\mathrm{d}y \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} P_2^3[g](x,y)\exp\{-i(xk_x+yk_y)\}\,\mathrm{d}x\mathrm{d}y \\
&= \left(\mathcal{F}^2 \circ P_2^3[g]\right)(k_x,k_y)
\end{aligned}$$

□

Figure 1 summarizes the implications of Theorem 2.3, which is well-known and is used in many SAR image recovery algorithms, [17, 19, 32]. 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.

2.2. **Data used in experiments.** We use one synthetic data set (see Figure 3) and one measured SAR data collect (see Figure 5) to evaluate our new methods and compare their performance to 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 [19, 32].

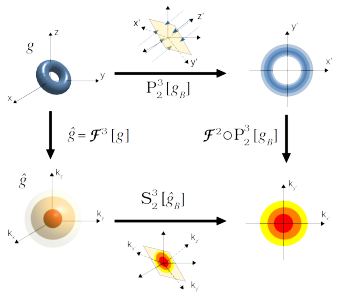


FIGURE 1. A graphical depiction of Theorem 2.3.

Parameter	Dataset Value
Elevation Range	$[-3^o, 3^o]$
Elevation Sampling	0.5^o
Frequency Range	$[27, 39]$GHz
Frequency Sampling	50MHz
Bandwidth	12GHz
Center Frequency	33GHz
Azimuth Range	$[0^o, 359.9^o]$
Azimuth Sampling	0.1^o

TABLE 1. Parameters of data sets used for experimentation.

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 1 summarizes these parameters.

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 (3) and then modeled as

$$\hat{\boldsymbol{g}} = \mathcal{F}\boldsymbol{g} + \boldsymbol{\varepsilon}, \tag{7}$$

where $\hat{\boldsymbol{g}} \in \mathbb{C}^M$ is the SAR phase history data (PHD), $\boldsymbol{\varepsilon} \in \mathbb{C}^M$ is complex-valued circularly symmetric Gaussian noise with covariance matrix $\sigma^2\mathbb{I}_M$, $\boldsymbol{g} \in \mathbb{C}^N$ 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

$$[\mathcal{F}^d]_{m,n} = \exp(-i\boldsymbol{p}_m \cdot \boldsymbol{q}_n), \quad \boldsymbol{p}_m, \boldsymbol{q}_n \in \mathbb{R}^d,$$

where $\boldsymbol{p}_m$ are the nonuniform frequencies and $\boldsymbol{q}_n$ are the nonuniform sources. Note that $\boldsymbol{g}$ and $\hat{\boldsymbol{g}}$ may be 2D or 3D, but we are able to vectorize $\boldsymbol{g}$ and $\hat{\boldsymbol{g}}$ and then formulate $\mathcal{F}$ accordingly. All numerical experiments in this investigation are implemented using the nonuniform fast Fourier transform library FINUFFT developed in [3]. In (7) we assume errors corresponding to aliasing and gridding are insignificant compared to inherent system noise.

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 (7), the SAR PHD measurements $\hat{\boldsymbol{g}}$ in 3D k-space at the respective equispaced azimuth angles, elevation angles, and frequencies, are given by

$$(f_m \cos\theta_p \cos\varphi_r, f_m \sin\theta_p \cos\varphi_r, f_m \sin\varphi_r), \quad \{\theta_p\}_{p=1}^{P}, \quad \{\varphi_r\}_{r=1}^{R}, \quad \{f_m\}_{m=1}^{N_k}. \tag{8}$$

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 $\tau_{p,r}$, $p = 1, \ldots, P$, $r = 1, \ldots, R$. 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 $\theta_1, \ldots, \theta_P$ to be partitioned into N_θ sets $\Theta_1, \ldots, \Theta_{N_\theta}$ as

$$\Theta_n = \left\{ \theta_p : \frac{P}{N_\theta}(n-1) + 1 \leq p \leq \frac{P}{N_\theta} n \right\}, \quad n = 1, \ldots, N_\theta. \tag{9}$$

An example of SAR PHD collected at three elevation angles over all azimuth angles is given in Figure 2.

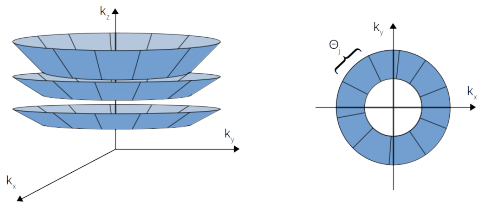


FIGURE 2. A graphical depiction of SAR PHD $\hat{\boldsymbol{g}}$ (3) in k-space as well as the partitioning of the data into N_θ partitions according to the azimuthal angle sets Θ_n given by (9).

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* isotropic as suggested in the model given by (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. [12, 29, 31].

In Section 3.1 we review a composite 2D SAR imaging technique which leverages the assumption that neighboring subapertures have similar support in their sparse domain, [31]. In Section 3.2 we review a volumetric approach using a three-dimensional analogue of the classic backprojection algorithm, [32]. These two ideas are then fused in Section 3.3 to construct an algorithm that compares the spatial information in neighboring imaging planes and penalizes the differences.

[1] Consistent with standard SAR imaging practices, it is always assumed that the data acquisition is sufficient for such partitions.

3.1. **2D SAR imaging leveraging sequential information.** We first explore the sequential imaging approach introduced in [31]. 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 $\hat{\boldsymbol{g}}^{(r)}$ gathered at elevation angle φ_r to reconstruct the 2D scene of interest $\tilde{\boldsymbol{g}} = \tilde{\boldsymbol{g}}^{(r)}$ for any $r \in \{1, \dots, R\}$.

To begin the image formation process, we first partition the SAR PHD data (3) into N_θ sequential subaperture bins $\hat{\boldsymbol{g}}_1^{(r)}, \dots, \hat{\boldsymbol{g}}_{N_\theta}^{(r)}$ so that the forward model (7) becomes

$$\hat{\boldsymbol{g}}_n^{(r)} = \mathcal{F}_n \tilde{\boldsymbol{g}}_n + \boldsymbol{\varepsilon}_n, \quad n = 1, \dots, N_\theta,$$

where $\tilde{\boldsymbol{g}}_n$ 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 (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 $\tilde{\boldsymbol{g}}$ is complex-valued. In this regard, when we are discussing sparsity, it is the magnitude of $\tilde{\boldsymbol{g}}$ 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 $\tilde{\boldsymbol{g}}_n = |\tilde{\boldsymbol{g}}_n| \Psi_n$. We then approximate the phase of the pixels of each image using $\Psi_n = D\left(\exp(i \operatorname{ang}(\tilde{\boldsymbol{g}}_n))\right)$, where $D(\cdot) = \operatorname{diag}(\cdot)$, yielding the estimate $|\tilde{\boldsymbol{g}}_n| \approx \Psi_n^H \tilde{\boldsymbol{g}}_n$. Using this reasoning as well as the assumption that the images themselves are sparse, we obtain the objective function, [31],

$$\begin{aligned} \widetilde{\boldsymbol{G}} = \underset{(\boldsymbol{q}_1, \dots, \boldsymbol{q}_{N_\theta})}{\arg\min} \Bigg\{ & \sum_{n=1}^{N_\theta} \left(\frac{\mu}{2} \left\| \mathcal{F}_n \boldsymbol{q}_n - \hat{\boldsymbol{g}}_n^{(r)} \right\|_2^2 + \|\boldsymbol{q}_n\|_1 \right) \\ & + \gamma \sum_{n=1}^{N_\theta - 1} \left\| \Psi_{n+1}^H \boldsymbol{q}_{n+1} - \Psi_n^H \boldsymbol{q}_n \right\|_2^2 \Bigg\}. \end{aligned} \tag{10}$$

Here $\widetilde{\boldsymbol{G}} = (\tilde{\boldsymbol{g}}_1, \dots, \tilde{\boldsymbol{g}}_{N_\theta})^T$, μ and γ are regularization parameters and Ψ_n are diagonal matrices containing the estimated phase at each pixel of $\boldsymbol{q}_n$. Observe that (10) describes a compressive sensing (CS) approach, [6], 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, $\tilde{\boldsymbol{g}}_n$, $\tilde{\boldsymbol{g}}_{n+1}$, should have the same sparse magnitude structure. We solve (10) using the well-known alternating direction method of multipliers (ADMM) algorithm, [4]. Once all $\tilde{\boldsymbol{g}}_1, \dots, \tilde{\boldsymbol{g}}_{N_\theta}$ are recovered, a final composite vectorized image $\tilde{\boldsymbol{g}}$ is formed where

$$\tilde{\boldsymbol{g}}^i = \underset{n \in [1, N_\theta]}{\arg\max} \left| \tilde{\boldsymbol{g}}_n^i \right| \tag{11}$$

for each pixel $i = 1, \dots, N$.

Remark 3.1. We note that leveraging sequential information in the sparse domain in (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.

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

$$S(f_m, \tau_{p,r}), \quad m = 1, \dots, N_k, \ p = 1, \dots, P, \ r = 1, \dots, R, \tag{12}$$

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 $\boldsymbol{x}$ denotes the vector of pixel locations in the recovered image. Here L is inversely proportional to the range resolution δ, [17], which is given by

$$\delta = \frac{c}{2(N_k - 1)\Delta f},$$

and c is again the speed of light.

3.2.1. *2D filtered backprojection.* Filtered backprojection (FBP) is a common technique used to recover 2D SAR images, [17]. 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 1 summarizes the standard FBP algorithm for 2D SAR. A thorough analysis of FBP for SAR imaging can be found in [14].

Algorithm 1 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 (12), and L range bins.
Output Image $\tilde{\boldsymbol{g}}$.
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 $s_{int}(\boldsymbol{x}, \tau_{p,r})$.[2]
end for
$\tilde{\boldsymbol{g}} = \sum_{r=1}^{R} \sum_{p=1}^{P} s_{int}(\boldsymbol{x}, \tau_{p,r})$.

[2]Since the range bins indexed by $l = 1, \dots, L$ typically do not align with Cartesian grid points $\boldsymbol{x}$, interpolation is needed to form a final pixelated image. A review of the interpolation methods commonly used in signal processing can be found in [13].

3.2.2. *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 1 to construct 2D slices of the 3D scene of interest $\boldsymbol{g}$.

Let $\mathcal{B}_s$, $s = 1, \dots, S$, be the 3D rotational change of basis operator (see Definition 2.2) 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 $\boldsymbol{g}_s = S_2^3[\boldsymbol{g}_{\mathcal{B}_s}]$, which is accomplished by performing Algorithm 1 for each imaging plane using the *full* data set $\hat{\boldsymbol{g}}$ in (3), [32]. 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 [19, 32], this technique can consider any combination of azimuth and elevation angles. Algorithm 2 summarizes this process.

Algorithm 2 Volumetric SAR using 2D FBP

Input SAR PHD $\hat{\boldsymbol{g}}$ (3) and threshhold value c_{thresh}.
Output Binary volumetric image $\boldsymbol{g}$.
Generate azimuth and elevation angle pairs (θ_s, φ_s) for $s = 1, \dots, S$.
Initialize volumetric image $\boldsymbol{g}$.
for $s = 1 : S$ **do**
Derive filtered backprojection $\boldsymbol{g}_s$ on imaging plane with angles θ_s and φ_s using Algorithm 1.
Create $\bar{\boldsymbol{g}}_s$ by thresholding $\boldsymbol{g}_s$ with $\bar{\boldsymbol{g}}_s = \begin{cases} 1 & \text{if } |\boldsymbol{g}_s| \geq c_{thresh} \\ 0 & \text{else} \end{cases}$.
For each $i = 1, \dots, N$ such that pixel $\left[(\bar{\boldsymbol{g}}_s)_{\mathcal{B}_s^{-1}}\right]^i = 1$, set $\boldsymbol{g}^i = 1$.
end for

Remark 3.2. 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 data given in (8), however. For the B747 data set (see Table 1), replaced with .1 degree spacing in elevation, the experiment used in [19] 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.

3.3. **Volumetric SAR leveraging sequential information.** The objective function (10) considers inter- and intra-image information, but only for data collected at a single elevation angle. By contrast, Algorithm 2 considers the full SAR PHD $\hat{\boldsymbol{g}}$ for azimuth angles $\theta_1 < \cdots < \theta_P$ and elevation angles $\varphi_1 < \cdots < \varphi_R$, as given in (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 [19, 32] extends (10) to three dimensions to include this *a priori* information. Since this leveraging of *a priori* intra-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 $\phi_1, \dots, \phi_V$, is established. The goal is then to recover *slices*

of the volumetric image $\boldsymbol{g}$ defined by every possible combination of these azimuth and elevation angles, i.e.

$$\boldsymbol{g}_{u,v} = S_2^3[\boldsymbol{g}_{\mathcal{B}_{u,v}}], \quad u = 1, \dots, U, \quad v = 1, \dots, V, \tag{13}$$

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 $\hat{\boldsymbol{g}}$ in (3), the objective function solved in [19, 32] is then given by

$$\begin{aligned} \boldsymbol{g}_{u,v} = \underset{\boldsymbol{q}_{u,v}}{\arg\min} \Bigg\{ & \|A_{u,v}\boldsymbol{q}_{u,v} - \hat{\boldsymbol{g}}\|_2^2 + \lambda_1 \|\boldsymbol{q}_{u,v}\|_1 \\ & + \frac{\lambda_2}{2} \sum_{i=u-1}^{u+1} \sum_{k=v-1}^{v+1} \left\| \Psi_{u,v}^H \boldsymbol{q}_{u,v} - \Psi_{i,k}^H \boldsymbol{q}_{i,k} \right\|_2^2 \Bigg\}, \end{aligned} \tag{14}$$

where λ_1 and λ_2 are regularization parameters and $\Psi_{u,v}$ are diagonal matrices containing the estimated phase at each pixel of $\boldsymbol{q}_{u,v}$. As is the case in (10), the first term in (14) enforces data fidelity, the second term promotes intra-image sparsity, and the third term encourages inter-image sparsity. The minimization problem given by (14) is then solved in [19, 32] using ADMM. Once the set $\{\boldsymbol{g}_{u,v}\}_{u=1,v=1}^{U,V}$ is recovered, the resulting slices are fused together to form the volumetric image $\boldsymbol{g}$.[3]

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

$$\hat{\mathcal{G}}_j = \mathcal{F}_j \mathcal{G}_j + \hat{\mathcal{E}}_j, \quad j = 1, \dots, J, \tag{15}$$

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 $\hat{\mathcal{E}}_j \sim \mathbb{CN}(0, [D(\boldsymbol{\alpha})]^{-1})$ with some noise precision vector $\boldsymbol{\alpha}$, where again we have $D(\cdot) = \text{diag}(\cdot)$. In our technique, $\mathcal{F}_j$ takes the form of the 2D or 3D NUFT.

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 3.2 and 3.3 to a Bayesian framework. Algorithms that utilize joint sparsity given multiple measurements in an empirical Bayesian setting were introduced in [36] and [35] for real-valued images. In [36] multiple data acquisitions at a single time are assumed, while [35] considered a *temporal* sequence of data acquisitions. Here we modify the technique coined the Joint Hierarchical Bayesian Learning (JHBL) method in [35] to consider a *spatial* sequence of image reconstructions. Analogous to the volumetric point estimate in (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

[3] We note that (14) is modified from the point estimate derived in [19, 32] in two ways: (1) Here we include phase extraction in the sequential difference regularization term, and (2) there $\hat{\boldsymbol{g}}$ was rotated and projected (according to u, v) before incorporating it into the objective function.

likelihood, enabling a closed form for the posterior from which we can efficiently sample.

We derive the general formulation for our approach in Section 4.1. In Section 4.2 we consider sequential data acquisitions along azimuth angles θ_p, $p = 1, \dots, P$, while in Section 4.3 we incorporate the idea that in the volumetric image reconstruction, data from neighboring subapertures should contain similar information.

4.1. **Hierarchical Bayesian model.** Following the Bayesian model in (15), let $\mathscr{G} = \{\mathcal{G}_j \in \mathbb{C}^N : j = 1, \dots, J\}$ be the collection of signals we seek to recover, where N is the number of pixels in each of the sequential images, $\hat{\mathscr{G}} = \{\hat{\mathcal{G}}_j \in \mathbb{C}^M : j = 1, \dots, J\}$ 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.

$$\pi(\varepsilon) = \frac{1}{\pi^N |[D(\boldsymbol{\alpha})]^{-1}|} \exp\big(-\varepsilon^H D(\boldsymbol{\alpha})\varepsilon\big),$$

where ε is a realization of $\hat{\mathcal{E}}$. Samples of signals $\mathscr{G}$ and data $\hat{\mathscr{G}}$ are correspondingly denoted as $G = \{\boldsymbol{g}_j \in \mathbb{C}^N : j = 1, \dots, J\}$ and $\hat{G} = \{\hat{\boldsymbol{g}}_j \in \mathbb{C}^M : j = 1, \dots, J\}$. While we have not yet specified how the set G relates to the volumetric image $\boldsymbol{g}$ 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, and $\pi(\boldsymbol{\alpha})$, $\pi(\boldsymbol{\beta})$, $\pi(\boldsymbol{\gamma})$ 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 (15). It is, however, typically unknown so instead it is standard to employ the relationship

$$\pi(G, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma} | \hat{G}) \propto \pi(\hat{G}|G, \boldsymbol{\alpha})\pi(G|\boldsymbol{\beta}, \boldsymbol{\gamma})\pi(\boldsymbol{\alpha})\pi(\boldsymbol{\beta})\pi(\boldsymbol{\gamma}), \tag{16}$$

from which we compute the right hand side, i.e. an un-normalized version of the posterior $\pi(G, \alpha, \boldsymbol{\beta}, \boldsymbol{\gamma}|\hat{G})$. The task is then to determine each of the five terms on the right hand side of (16), which we now describe.

The likelihood $\pi(\hat{G}|G, \boldsymbol{\alpha})$. From the noise model in (15) we have

$$\pi(\hat{G}|G, \boldsymbol{\alpha}) \propto \prod_{j=1}^{J} \exp\left\{-\alpha_j \|\mathcal{F}_j \boldsymbol{g}_j - \hat{\boldsymbol{g}}_j\|_2^2\right\}. \tag{17}$$

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

$$\pi(\boldsymbol{\alpha}) \propto \prod_{j=1}^{J} \alpha_j^{\eta_\alpha - 1} \exp\{-\nu_\alpha \alpha_j\}, \tag{18}$$

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 (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, [34].

The joint prior $\pi(G|\boldsymbol{\beta},\boldsymbol{\gamma})$. To leverage both the sparsity assumption in the image magnitude and sequential information, 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

$$\pi(G|\boldsymbol{\beta}) \propto \prod_{j=1}^{J}\prod_{i=1}^{N} \beta_{j,i} \exp\left\{-\beta_{j,i}\left|g_{j,i}\right|^2\right\}, \quad j = 1,\dots,J, \quad i = 1,\dots,N. \tag{19}$$

Here each precision $\beta_{j,i}$ is a random variable. The prior in (19) is commonly employed to promote sparsity because it is conjugate for the likelihood density function (17) and therefore results in a closed form posterior, [34]. 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

$$\pi(\boldsymbol{\beta}) \propto \prod_{j=1}^{J}\prod_{i=1}^{N} \beta_{j,i}^{\eta_\beta - 1} \exp\{-\nu_\beta \beta_{j,i}\}, \tag{20}$$

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, [32], we assume that the difference in magnitude of $\boldsymbol{g}_j$ compared with $\boldsymbol{g}_{j-1}$ is small for $j = 1,\dots,J$. We can therefore employ the conditionally inter-image complex-valued Gaussian prior

$$\pi(G|\boldsymbol{\gamma}) \propto \prod_{j=1}^{J}\prod_{i=1}^{N} \gamma_{j,i} \exp\left\{-\gamma_{j,i}\left|\Psi_{j-1,i}^H g_{j-1,i} - \Psi_{j,i}^H g_{j,i}\right|^2\right\}. \tag{21}$$

Since the azimuth angle θ_j, $j = 1,\dots,J$, is subdivided on $[0, 2\pi)$, we also assume periodicity and impose $\boldsymbol{g}_J = \boldsymbol{g}_0$ in (21).

The inter-image hyper-prior $\pi(\gamma)$. Similar to the intra-image prior (19), we use gamma distributed hyper-priors for each $\gamma_{j,i}$ with hyper-parameters η_γ and ν_γ akin to those in (20), i.e.

$$\pi(\boldsymbol{\gamma}) \propto \prod_{j=1}^{J}\prod_{i=1}^{N} \gamma_{j,i}^{\eta_\gamma - 1} \exp\{-\nu_\gamma \gamma_{j,i}\}. \tag{22}$$

The posterior. Combining (17), (19), (20), (21) and (22), we are now ready to calculate the posterior density function, (16) as

$$\begin{aligned}\pi(G,\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\gamma}|\hat{G}) \propto \prod_{j=1}^{J}\Bigg[&\left(\alpha_j^M \exp\left\{-\alpha_j\|\mathcal{F}_j\boldsymbol{g}_j-\hat{\boldsymbol{g}}_j\|_2^2\right\}\right)\left(\prod_{i=1}^{N}\beta_{j,i}\exp\left\{-\beta_{j,i}|g_{j,i}|^2\right\}\right)\\ &\times\left(\prod_{i=1}^{N}\gamma_{j,i}\exp\left\{-\gamma_{j,i}\left|\Psi_{j-1,i}^H g_{j-1,i}-\Psi_{j,i}^H g_{j,i}\right|^2\right\}\right)\left(\alpha_j^{\eta_\alpha-1}\exp\{-\nu_\alpha\alpha_j\}\right)\\ &\times\left(\prod_{i=1}^{N}\beta_{j,i}^{\eta_\beta-1}\exp\{-\nu_\beta\beta_{j,i}\}\right)\times\left(\prod_{i=1}^{N}\gamma_{j,i}^{\eta_\gamma-1}\exp\{-\nu_\gamma\gamma_{j,i}\}\right)\Bigg].\end{aligned} \tag{23}$$

Due to the structure of (23), we can decompose $\pi(G,\alpha,\boldsymbol{\beta},\boldsymbol{\gamma}|\hat{G})$ 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 $\boldsymbol{g}_j$, α_j, $\beta_{j,i}$, and $\gamma_{j,i}$ for $j=1,\dots,J$, $i=1,\dots,N$, as

$$\begin{aligned}\boldsymbol{g}_j^{(\ell+1)} = &\left(\alpha_j^{(\ell)}\mathcal{F}_j^H\mathcal{F}_j + D\left(\boldsymbol{\beta}_j^{(\ell)}\right)+D\left(\boldsymbol{\gamma}_j^{(\ell)}\right)+D\left(\boldsymbol{\gamma}_{j+1}^{(\ell)}\right)\right)^{-1}\times\\ &\left(\alpha_j^{(\ell)}\mathcal{F}_j^H\hat{\boldsymbol{g}}_j+\Psi_j^{(\ell)}D\left(\boldsymbol{\gamma}_j^{(\ell)}\right)\Psi_{j-1}^{(\ell)H}\boldsymbol{g}_{j-1}^{(\ell)}+\Psi_j^{(\ell)}D\left(\boldsymbol{\gamma}_{j+1}^{(\ell)}\right)\Psi_{j+1}^{(\ell)H}\boldsymbol{g}_{j+1}^{(\ell)}\right)\end{aligned} \tag{24}$$

$$\alpha_j^{(\ell+1)} = \frac{\eta_\alpha+M-1}{\nu_\alpha+\left\|\mathcal{F}_j\boldsymbol{g}_j^{(\ell+1)}-\hat{\boldsymbol{g}}_j\right\|_2^2} \tag{25}$$

$$\beta_{j,i}^{(\ell+1)} = \frac{\eta_\beta}{\nu_\beta+\left|g_{j,i}^{(\ell+1)}\right|^2} \tag{26}$$

$$\gamma_{j,i}^{(\ell+1)} = \frac{\eta_\gamma}{\nu_\gamma+\left|\Psi_{j-1,i}^{(\ell+1)H}g_{j-1,i}^{(\ell+1)}-\Psi_{j,i}^{(\ell+1)H}g_{j,i}^{(\ell+1)}\right|^2}, \tag{27}$$

where $\gamma_{J+1,i}=\gamma_{1,i}$ and $\gamma_{0,i}=\gamma_{J,i}$ for $i=1,\dots,N$. The derivations of (24), (25), (26), and (27) are provided in Appendix A. We note that (24) does not rely on any $\boldsymbol{g}_{j'}$, $j\neq j'$, from the $\ell+1$ step, which increases opportunities for parallelization.

Remark 4.1. 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 [11] 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 $\boldsymbol{g}$ update to avoid inverting large matrices, such as the gradient descent method, [16].

Algorithm 3 summarizes how the MAP estimate of (23) for each $\boldsymbol{g}_j$, $j=1,\dots,J$, is obtained based on the update steps (24), (25), (26) and (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 3 in two different contexts for 3D SAR image reconstruction. As described in Section 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 [19, 32]. In this case Algorithm 3 is used over sequential azimuthal angles, $\{\theta_p\}_{p=1}^P$ (Algorithm 4). By contrast, Algorithm 5 in Section 4.3 is performed over volumetric images formed by 3D subapertures, i.e. where the sequenced information

Algorithm 3 Joint Hierarchical Bayesian Learning for J sequential data acquisitions (JHBL)

Input SAR PHD $\hat{\boldsymbol{g}}$ from (3) and hyperparameters η_α, η_β, η_γ, ν_α, ν_β, ν_γ. Define ℓ_{max} as the maximum number of iterations and tol as the threshold determining convergence.
Output Collection of reconstructions $G = \{\boldsymbol{g}_j\}_{j=1}^J$, corresponding phase Ψ, and hierarchical parameters $\boldsymbol{\alpha}, \boldsymbol{\beta}$, and $\boldsymbol{\gamma}$.
if 2D IRB (Algorithm 4) **then**
  $\hat{G} = \{\hat{\boldsymbol{g}}_p\}_{p=1}^{P/2}$ given by (28) and $J = P/2$.
else if 3D SRCI (Algorithm 5) **then**
  $\hat{G} = \{\hat{\boldsymbol{g}}_n\}_{n=1}^{N_\theta}$ according to the azimuth partitions given in (9) and $J = N_\theta$.
end if
Initialize $G^{(0)}$ using the NUFT, set $\Psi^{(0)} = \text{ang}\left(G^{(0)}\right)$, and $\boldsymbol{\alpha}^{(0)} = \boldsymbol{\beta}^{(0)} = \boldsymbol{\gamma}^{(0)} = 1$.
while $\ell < \ell_{max}$ or $\frac{1}{J}\sum_{j=1}^J \left\| \left|\boldsymbol{g}_j^{(\ell)}\right| - \left|\boldsymbol{g}_j^{(\ell-1)}\right| \right\|_1 < tol$ **do**
  $G^{(\ell+1)} = \arg\max_G \pi\left(G | G^{(\ell)}, \boldsymbol{\alpha}^{(\ell)}, \boldsymbol{\beta}^{(\ell)}, \boldsymbol{\gamma}^{(\ell)}, \hat{G}\right)$
  $\Psi^{(\ell+1)} = \text{ang}\left(G^{(\ell+1)}\right)$
  $\boldsymbol{\alpha}^{(\ell+1)} = \arg\max_{\boldsymbol{\alpha}} \pi\left(\boldsymbol{\alpha} | G^{(\ell+1)}\right)$
  $\boldsymbol{\beta}^{(\ell+1)} = \arg\max_{\boldsymbol{\beta}} \pi\left(\boldsymbol{\beta} | G^{(\ell+1)}\right)$
  $\boldsymbol{\gamma}^{(\ell+1)} = \arg\max_{\boldsymbol{\gamma}} \pi\left(\boldsymbol{\gamma} | G^{(\ell+1)}\right)$
end while

is over partitions of the azimuthal angles $\{\Theta_j\}_{j=1}^J$, (9). A composite image is then created from these subaperture reconstructions using (30).

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 $\boldsymbol{g}_p$ to be the unknown image and $\hat{\boldsymbol{g}}_p$ 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 $\hat{\boldsymbol{g}}$ into $P/2$ slices,[4] which we denote as the $z\theta_p$-plane for each $p = 1, \ldots, P/2$, so that

$$\hat{\boldsymbol{g}}_p = \mathcal{S}_2^3[\hat{\boldsymbol{g}}_{\mathcal{B}_p}], \tag{28}$$

where $\mathcal{S}_2^3$ is the slice operator defined in (4) and $\mathcal{B}_p$ is the rotational change of basis that rotates θ_p degrees about the z-axis followed by a rotation of $\frac{\pi}{2}$ radians about the x-axis. The set $\hat{G} = \{\hat{\boldsymbol{g}}_p\}_{p=1}^{P/2}$ is then used as input to Algorithm 3, where the 2D NUFT is utilized. Given that $\boldsymbol{g}_1$ and $\boldsymbol{g}_{P/2}$ are centered on the $z\theta_1$-plane and $z\theta_{P/2}$-plane, we define $\boldsymbol{g}_0$ and $\boldsymbol{g}_{J+1}$ in Algorithm 3 to respectively be $\boldsymbol{g}_{P/2}$ and $\boldsymbol{g}_1$ reflected over the z-axis. This satisfies the expectation that $|\boldsymbol{g}_1|$ should be similar to a *mirrored* version of $|\boldsymbol{g}_{P/2}|$. Then, the volumetric image $\boldsymbol{g}$ is formed from the

[4] We use $P/2$ slices instead of P slices since, in each $\hat{\boldsymbol{g}}_p$, the data along both azimuth θ_p and $\theta_{p+P/2} = \theta_p + \pi$ are included.

output $G = \{\boldsymbol{g}_p\}_{p=1}^{P/2}$ as

$$\boldsymbol{g} = \sum_{p=1}^{P/2} (\mathcal{P}_3^2[|\boldsymbol{g}_p|])_{\mathcal{B}_p^{-1}}, \tag{29}$$

i.e. backprojecting each $|\boldsymbol{g}_p|$ in physical 3D, performing the inverse rotation to the one that was originally performed on the data, and then summing the resulting backprojections to form the volumetric image. This reconstruction process relies on the assumption from Theorem 2.3 that each $\boldsymbol{g}_p$ 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, [31]. 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 $\boldsymbol{g}_p$, $p = 1, \ldots, P/2$, still reconstructs the scene in a robust and predictable manner. Lastly, the resulting image $\boldsymbol{g}$ can be thresholded according to a user-defined threshold value c_{IRB} to form a volumetric point cloud $\boldsymbol{g}_c$. 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 $\boldsymbol{g}$ with relatively little additional computational cost. The 2D IRB method is summarized in Algorithm 4.

Algorithm 4 2D Image Reconstruction with Backprojection (2D IRB)

Input SAR PHD $\hat{\boldsymbol{g}}$ ((3) for (8)) and threshold value c_{IRB}.
Output Volumetric reconstruction $\boldsymbol{g}$ and point cloud $\boldsymbol{g}_c$.
Partition $\hat{\boldsymbol{g}}$ into $P/2$ slices $\hat{\boldsymbol{g}}_p$, $p = 1, \ldots, P/2$ according to (28).
Obtain 2D projections $G = \{\boldsymbol{g}_p\}_{p=1}^{P/2}$ by using $\hat{G} = \{\hat{\boldsymbol{g}}_p\}_{p=1}^{P/2}$ as input to Algorithm 3.
Backproject $\boldsymbol{g}_p$ according to (29) to form $\boldsymbol{g}$.
For each $i = 1, \ldots, N$, set $\boldsymbol{g}_c^i = \begin{cases} 1 & \text{if } \boldsymbol{g}^i > c_{IRB} \\ 0 & \text{else} \end{cases}$.

Remark 4.2. The 2D projections $\{\boldsymbol{g}_p\}_{j=1}^{P/2}$, obtained during the 2D IRB method in Algorithm 4 can be used as input to other volumetric image recovery methods, see e.g. [20].

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 (9). The sequential sparsity assumption holds for the J volumetric images $\boldsymbol{g}_j$ formed using the partitions of the data $\hat{G} = \{\hat{\boldsymbol{g}}_j\}_{j=1}^{J}$.

With the data partitioned, we then perform Algorithm 3 using the 3D NUFT, which results in the collection of J volumetric images $G = \{\boldsymbol{g}_j\}_{j=1}^{J}$. 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 (11), we mitigate this issue by constructing a composite

image, [29, 31],

$$\boldsymbol{g}^i = \arg\max_j \left|\boldsymbol{g}_j^i\right| \tag{30}$$

for each pixel $i = 1, \ldots, N$. As with the 2D IRB method in Algorithm 4, the final point cloud $\boldsymbol{g}_c$ is formed by thresholding $\boldsymbol{g}$ in (30), in this case to some threshold value c_{SRCI}. Multiple threshold values may again be tested for the same volumetric image $\boldsymbol{g}$. The 3D SRCI process is summarized in Algorithm 5.

Algorithm 5 3D Subaperture Reconstruction with Composite Imaging (3D SRCI)

Input SAR PHD $\hat{\boldsymbol{g}}$ ((3) for (8)) and threshold value c_{SRCI}.
Output Volumetric reconstruction $\boldsymbol{g}$ and point cloud $\boldsymbol{g}_c$.
Partition $\hat{\boldsymbol{g}}$ into J subapertures $\hat{\boldsymbol{g}}_j$, $j = 1, \ldots, J$, according to the azimuthal sets Θ_j in (9).
Obtain 3D reconstruction $G = \{\boldsymbol{g}_j\}_{j=1}^J$ by using $\hat{G} = \{\hat{\boldsymbol{g}}_j\}_{j=1}^J$ as input to Algorithm 3.
Perform the composite imaging step according to (30) to form $\boldsymbol{g}$.
For each $i = 1, \ldots, N$, set $\boldsymbol{g}_c^i = \begin{cases} 1 & \text{if } \boldsymbol{g}^i > c_{SRCI} \\ 0 & \text{else} \end{cases}$.

A couple of comments are in order:

- In addition to a point estimate for each $\boldsymbol{g}_j$, $j = 1, \ldots, J$, our method recovers a MAP estimate for the precisions of the likelihood and both priors. This information may be used in lieu of (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 [12] may be useful.
- Following the discussion regarding composite imaging (30) used in the 3D SRCI (Algorithm 5), one might wonder why composite imaging is not incorporated into the 2D IRB (Algorithm 4). This is because the set of images $\{\boldsymbol{g}_p\}_1^{P/2}$ reconstructed using JHBL (Algorithm 3) are *projections* of the volumetric image onto different imaging planes. Thus if $\boldsymbol{g}$ were to be calculated via composite imaging, i.e.

$$\boldsymbol{g}^i = \arg\max_p \left[(\mathcal{P}_3^2[|\boldsymbol{g}_p|])_{\mathcal{B}_p^{-1}} \right]^i, \quad i = 1, \ldots, N,$$

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 (29).

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), [15], for comparable 2D and 3D SAR image formation and data fusion, [21, 24, 32, 37]. Hence we use that here and note that other error metrics may also be useful.

The MHD between two point clouds $\mathcal{S}$ and $\mathcal{T}$ is given by

$$\mathrm{MHD}(\mathcal{S}, \mathcal{T}) = \max(d(\mathcal{S}, \mathcal{T}), d(\mathcal{T}, \mathcal{S})), \tag{31}$$

where

$$d(\mathcal{S},\mathcal{T}) = \frac{1}{N_s}\sum_{s\in\mathcal{S}} d(s,\mathcal{T}), \qquad d(s,\mathcal{T}) = \min_{t\in\mathcal{T}} \|s-t\|_2,$$

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 $\boldsymbol{f}$ are

$$\boldsymbol{f}_{\text{dB}} = 20\log_{10}\left(\frac{1}{\|\boldsymbol{f}\|_\infty}\boldsymbol{f}\right). \tag{32}$$

We also restrict the domain according to the alias-free extents given in [32].

The rest of this section is organized as follows: Section 5.1 establishes the various parameters used for each numerical experiment. Sections 5.2 and 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 (32). Section 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 5.5 evaluates both techniques using a sub-sampled data set with no additional noise, and we compare these results to those obtained in [32].

5.1. **Selection of parameters.** With respect to parameters, for our experiments using the 2D IRB approach (Algorithm 4) 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 5) 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 2.

Recalling the discussion following (18) regarding the selection of hyperparameters, when using Algorithm 3, we set the hyperparameters to be

$$\eta_\alpha = 1.5, \quad \eta_\beta = \eta_\gamma = 0.5, \quad \text{and} \quad \nu_\alpha = \nu_\beta = \nu_\gamma = 10^{-3}.$$

We also set the maximum number of iterations, ℓ_{max}, and threshold that determines convergence, tol, in Algorithm 3 to be

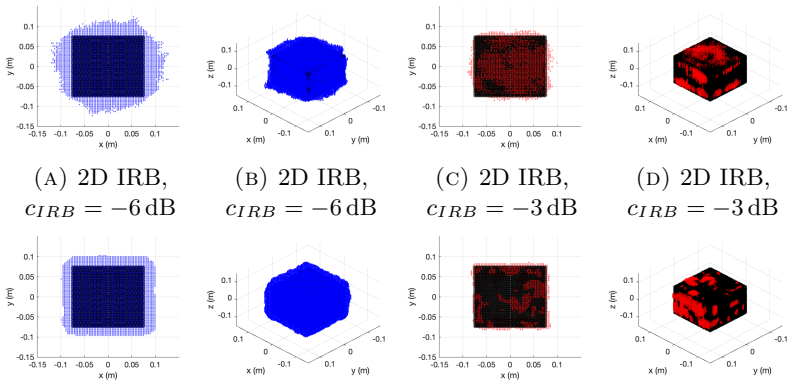
$$\ell_{max} = 10 \quad \text{and} \quad tol = 10^{-5}.$$

We evaluate the utility of the JHBL approach in the noisy regime by also considering reconstructions formed using the MAP of (17), which we call the maximum-likelihood estimate (MLE), in lieu of the MAP of (23) that is calculated using Algorithm 3. 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, (20) and (22), in the presence of noise. When reconstructing the volumetric images in this way, we replace Algorithm 3 with the MLE in Algorithm 4 or Algorithm 5, in which case we refer to these modified techniques as MLE (in contrast to JHBL) methods.

Parameter	**2D IRB** (Algorithm 4)	**3D SRCI** (Algorithm 5)
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 2. Sizes of the inputs and outputs of Algorithm 3 for our numerical experiments.

Finally, we note that the computational cost of each method heavily depends on the NUFT algorithm used. In particular the 3D SRCI technique requires a much larger NUFT transform than the 2D IRB method. The value of P in (29) also contributes significantly to the computational requirements of the 2D IRB method, as the backprojection step of Algorithm 4 involves rotating a 3D image and interpolating the result to uniform grid points.



(A) 2D IRB, $c_{IRB} = -6\,\text{dB}$ (B) 2D IRB, $c_{IRB} = -6\,\text{dB}$ (C) 2D IRB, $c_{IRB} = -3\,\text{dB}$ (D) 2D IRB, $c_{IRB} = -3\,\text{dB}$

(E) 3D SRCI, $c_{SRCI} = -18\,\text{dB}$ (F) 3D SRCI, $c_{SRCI} = -18\,\text{dB}$ (G) 3D SRCI, $c_{SRCI} = -6\,\text{dB}$ (H) 3D SRCI, $c_{SRCI} = -6\,\text{dB}$

FIGURE 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.

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, [3]) 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 on a voxelized grid with grid size $301 \times 301 \times 301$. All other parameters are consistent with those in Table 1.

Figure 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 3 that the reconstruction from the 3D SRCI technique (Algorithm 5) tends to more sharply

[5]Due to how the data are generated the units in this example are arbitrary and are chosen to remain comparable with the B747 data set.

define edges of the cube, while the 2D IRB method (Algorithm 4) 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 the first two panels of Figure 4, which show cross-sections of the 2D IRB and 3D SRCI reconstructions of the cube before any thresholding occurs. From the last panel of Figure 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.

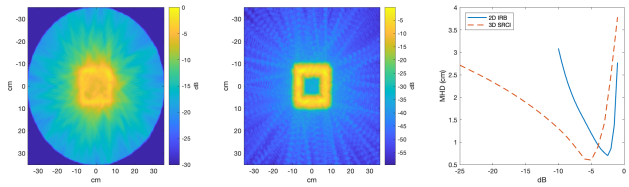
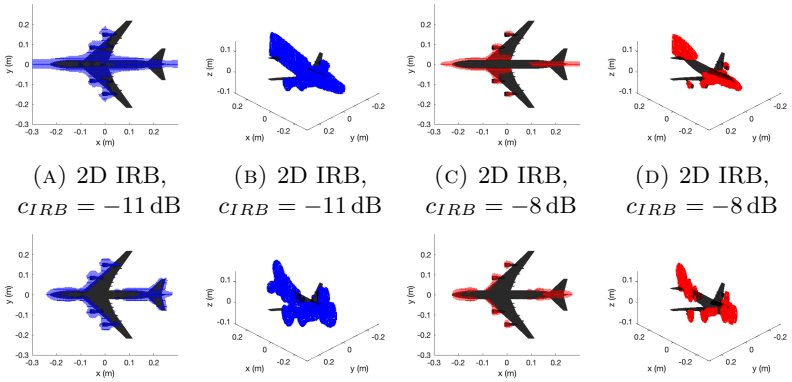


FIGURE 4. Cross-sections of the (left) 2D IRB and (middle) 3D SRCI reconstructions of the cube data set with no additional noise. (right) MHD values at various dB threshold values when either technique is used on the cube data set with no additional noise; the minimum MHD value calculated for the 2D IRB method is 0.7017cm, and for the 3D SRCI, the minimum MHD is 0.6014cm.



(A) 2D IRB, $c_{IRB} = -11\,\text{dB}$ (B) 2D IRB, $c_{IRB} = -11\,\text{dB}$ (C) 2D IRB, $c_{IRB} = -8\,\text{dB}$ (D) 2D IRB, $c_{IRB} = -8\,\text{dB}$

(E) 3D SRCI, $c_{SRCI} = -32\,\text{dB}$ (F) 3D SRCI, $c_{SRCI} = -32\,\text{dB}$ (G) 3D SRCI, $c_{SRCI} = -22\,\text{dB}$ (H) 3D SRCI, $c_{SRCI} = -22\,\text{dB}$

FIGURE 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.

5.3. **Reconstructions from measured data.** Figure 5 displays reconstructions of the B747 with no additional noise using the 2D IRB and 3D SRCI techniques. Observe in the 2D IRB (Algorithm 4) reconstructions (A)-(D) that the tail of the plane is poorly resolved. The tail is particularly 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 5 (E)-(H) show that using the 3D SRCI (Algorithm 5) approach significantly helps to mitigate this issue.

Figure 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 displays cross sections of the reconstructed B747 data set in the left and middle panels using the 2D IRB and 3D SRCI techniques, respectively, where the structures of the wings and the horizontal stabilizers are much clearer in the middle as opposed to the left panel. Observe in the right panel of Figure 6 that, as in Figure 4, the 3D SRCI approach achieves a smaller global MHD value as well as a larger dynamic range.

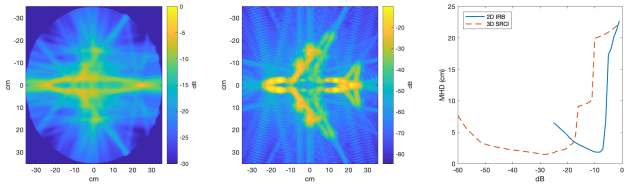


FIGURE 6. Cross-sections of the (left) 2D IRB and (middle) 3D SRCI reconstructions of the B747 data set with no additional noise. (right) MHD values at various dB threshold values when either technique is used on the B747 data set with no additional noise; the minimum MHD value calculated for the 2D IRB method is 1.832cm, and for the 3D SRCI, the minimum MHD is 1.440cm.

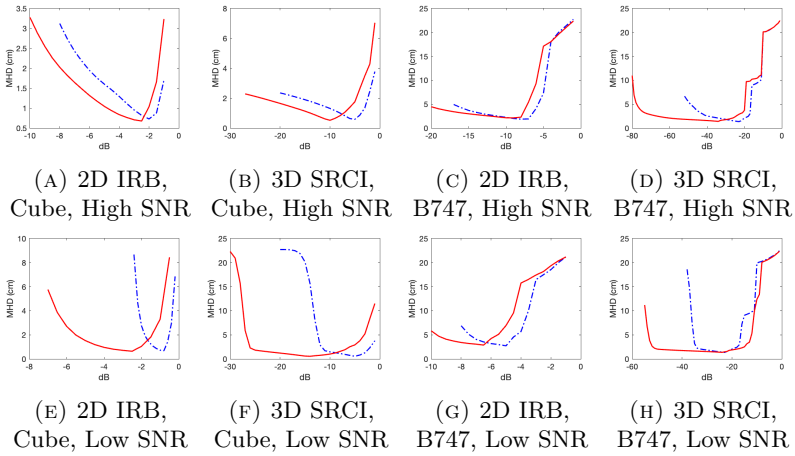
5.4. **Noise study.** For given SAR PHD with additive complex-valued circularly symmetric Gaussian noise, we calculate the signal-to-noise ratio (SNR) as

$$SNR = 10\log_{10}\left(\frac{\mu^2}{\sigma^2}\right)\text{dB},$$

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 (32). We now evaluate our techniques in both high ($\approx 0\,\text{dB}$) and low ($\approx -24\,\text{dB}$) SNR environments. This compares to the B747 data set used in [19, 32], 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 4)		**3D SRCI** (Algorithm 5)	
	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 3. Minimum MHD (cm) achieved across tested dB thresholds.



(A) 2D IRB, Cube, High SNR (B) 3D SRCI, Cube, High SNR (C) 2D IRB, B747, High SNR (D) 3D SRCI, B747, High SNR

(E) 2D IRB, Cube, Low SNR (F) 3D SRCI, Cube, Low SNR (G) 2D IRB, B747, Low SNR (H) 3D SRCI, B747, Low SNR

FIGURE 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 $\approx$ 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).

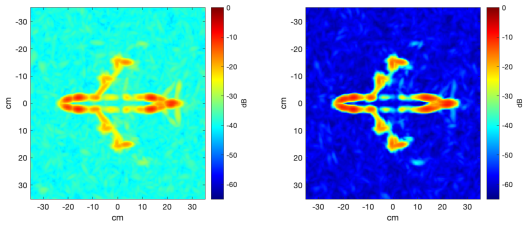


FIGURE 8. Slices of the (left) MLE and (right) JHBL reconstructions of the B747 with SNR of -30 dB using the 3D SRCI approach.

Table 3 displays the minimum MHD values achieved for the 2D IRB and 3D SRCI algorithms using either JHBL (Algorithm 3) 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 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 separation between distinct scatterers, may lead more accurate target classification, [23, 28, 38]. Figure 8 shows a cross section of the MLE and JHBL reconstructions of the B747 at low SNR, highlighting this distinction.

Using the 3D SRCI approach, we are able to maintain data fidelity at even lower SNR values. Figure 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

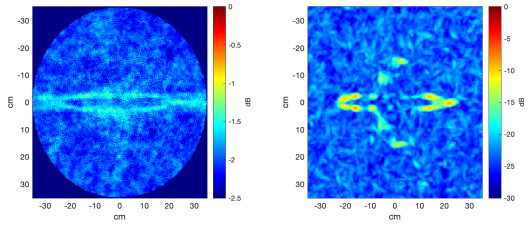


FIGURE 9. Slices of the (left) 2D IRB and (right) 3D SRCI reconstructions of the B747 with SNR of approximately -34 dB. Note that for interpretability, the threshold dB scale is different for each figure.

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.

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 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 5. The same numerical experiment was performed in [32] using (14) (also see footnote there), where a minimum MHD value of 3.78cm was reported.

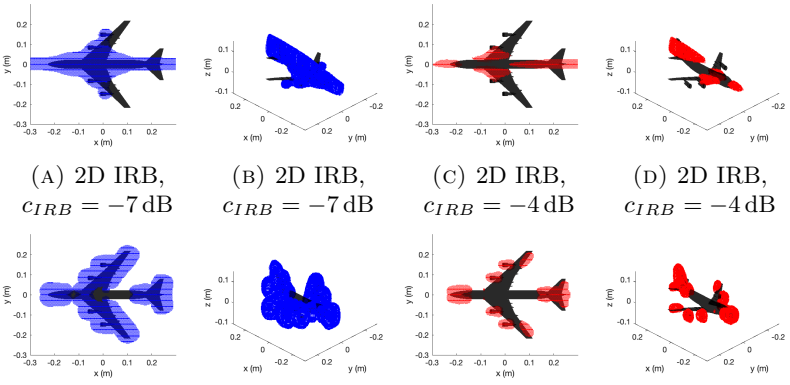
Parameter	**Sub-sampled Value**
Elevation Range	$[-3^o, 3^o]$
Elevation Sampling	0.5^o
Frequency Range	$[31, 35]$GHz
Frequency Sampling	150MHz
Bandwidth	4GHz
Center Frequency	33GHz
Azimuth Range	$[0^o, 359.9^o]$
Azimuth Sampling	0.3^o

TABLE 4. Parameters of sub-sampled data set used for experimentation.

Parameter	**2D IRB** (Algorithm 4)	**3D SRCI** (Algorithm 5)
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 5. Sizes of the parameter inputs and outputs for Algorithm 3 for the sub-sampled data experiments.

Figure 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



(A) 2D IRB, $c_{IRB} = -7\,\text{dB}$ (B) 2D IRB, $c_{IRB} = -7\,\text{dB}$ (C) 2D IRB, $c_{IRB} = -4\,\text{dB}$ (D) 2D IRB, $c_{IRB} = -4\,\text{dB}$

(E) 3D SRCI, $c_{SRCI} = -30\,\text{dB}$ (F) 3D SRCI, $c_{SRCI} = -30\,\text{dB}$ (G) 3D SRCI, $c_{SRCI} = -15\,\text{dB}$ (H) 3D SRCI, $c_{SRCI} = -15\,\text{dB}$

FIGURE 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.

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).

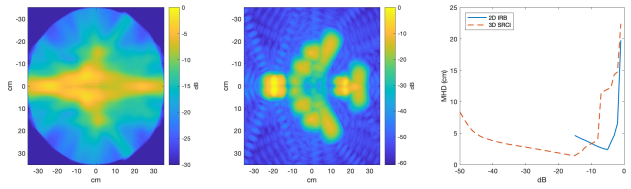


FIGURE 11. Cross-sections of the (left) 2D IRB and (middle) 3D SRCI reconstructions of the sub-sampled B747 data set using the parameters in Tables 4 and 5. (right) MHD values at various dB threshold values.

Figure 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 [32] in this sub-sampling experiment. Indeed, it does not appear that the MHD values differ significantly in the sub-sampling case.

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 reconstruction. 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. 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 subsampled regime and again found that both methods outperform previously designed algorithms.

Future work will focus on evaluating our approach for various levels of undersampling in the azimuth, elevation, and frequency domains. 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.

Appendix A. **Derivation of the JHBL updates.** Here we derive several of the update steps given by (24), (25), (26), and (27) that are used in Algorithm 3. Since the updates for $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, and $\boldsymbol{\gamma}$ are similarly obtained, we only include the derivation for the $\boldsymbol{\alpha}$ update step.

A.1. **The G update.** Let $G = \{\boldsymbol{g}_j\}_{j=1}^J$ in Algorithm 3. To make parallelization possible, we choose each $\boldsymbol{g}_j$ update to depend on the $\boldsymbol{g}_{-j}$ vectors from the previous update step,[6]

$$\boldsymbol{g}_j^{(\ell+1)} = \arg\max_{\boldsymbol{g}_j} \pi\left(\boldsymbol{g}_j | \boldsymbol{g}_{-j}^{(\ell)}, \boldsymbol{\alpha}^{(\ell)}, \boldsymbol{\beta}^{(\ell)}, \boldsymbol{\gamma}^{(\ell)}\right), \quad j = 1, \dots, J. \tag{33}$$

From (23) the probability density function in (33) yields

$$\begin{aligned}
\pi\left(\boldsymbol{g}_j | \boldsymbol{g}_{-j}^{(\ell)}, \boldsymbol{\alpha}^{(\ell)}, \boldsymbol{\beta}^{(\ell)}, \boldsymbol{\gamma}^{(\ell)}\right) &\propto \exp\left\{-\alpha_j^{(\ell)} \|\mathcal{F}_j \boldsymbol{g}_j - \hat{\boldsymbol{g}}_j\|_2^2\right\} \times \left(\prod_{i=1}^N \exp\left\{-\beta_{j,i}^{(\ell)} |g_{j,i}|^2\right\}\right) \\
&\quad \times \left(\prod_{i=1}^N \exp\left\{-\gamma_{j,i}^{(\ell)} \left|\Psi_{j-1,i}^{(\ell)H} g_{j-1,i}^{(\ell)} - \Psi_{j,i}^{(\ell)H} g_{j,i}\right|^2\right\}\right) \\
&\quad \times \left(\prod_{i=1}^N \exp\left\{-\gamma_{j+1,i}^{(\ell)} \left|\Psi_{j,i}^{(\ell)H} g_{j,i} - \Psi_{j+1,i}^{(\ell)H} g_{j+1,i}^{(\ell)}\right|^2\right\}\right) \\
&\propto \exp\Bigg(- \boldsymbol{g}_j^H \Sigma \boldsymbol{g}_j + \alpha_j^{(\ell)} \hat{\boldsymbol{g}}_j^H \mathcal{F}_j \boldsymbol{g}_j + \alpha_j^{(\ell)} \boldsymbol{g}_j^H \mathcal{F}_j^H \hat{\boldsymbol{g}}_j \\
&\quad + \boldsymbol{g}_j^H \Psi_j^{(\ell)} D\left(\boldsymbol{\gamma}_j^{(\ell)}\right) \Psi_{j-1}^{(\ell)H} \boldsymbol{g}_{j-1}^{(\ell)} + \boldsymbol{g}_{j-1}^{(\ell)} H \Psi_{j-1}^{(\ell)} D\left(\boldsymbol{\gamma}_j^{(\ell)}\right) \Psi_j^{(\ell)H} \boldsymbol{g}_j
\end{aligned}$$

[6]The negative subscript refers to all similarly-named variables with subscripts other than the one indicated, in this case $\boldsymbol{g}_{-j} = \{\boldsymbol{g}_{j'} : j' = 1, \dots, J,\ j' \neq j\}$.

$$+ \boldsymbol{g}_j^H \Psi_j^{(\ell)} D\left(\boldsymbol{\gamma}_{j+1}^{(\ell)}\right) \Psi_{j+1}^{(\ell)H} \boldsymbol{g}_{j+1}^{(\ell)} + \boldsymbol{g}_{j+1}^{(\ell)H} \Psi_{j+1}^{(\ell)} D\left(\boldsymbol{\gamma}_{j+1}^{(\ell)}\right) \Psi_j^{(\ell)H} \boldsymbol{g}_j \Bigg)$$

$$\propto \exp\big((\boldsymbol{g}_j - \bar{\boldsymbol{g}})^H \Sigma (\boldsymbol{g}_j - \bar{\boldsymbol{g}})\big),$$

where $D(\cdot) = \mathrm{diag}(\cdot)$ and

$$\Sigma = \alpha_j^{(\ell)} \mathcal{F}_j \mathcal{F}_j + D\left(\boldsymbol{\beta}_j^{(\ell)}\right) + D\left(\boldsymbol{\gamma}_j^{(\ell)}\right) + D\left(\boldsymbol{\gamma}_{j+1}^{(\ell)}\right),$$

$$\bar{\boldsymbol{g}} = \Sigma^{-1}\left(\alpha_j^{(\ell)} \mathcal{F}_j^H \hat{\boldsymbol{g}}_j + \Psi_j^{(\ell)} D\left(\boldsymbol{\gamma}_j^{(\ell)}\right) \Psi_{j-1}^{(\ell)H} \boldsymbol{g}_{j-1}^{(\ell)} + \Psi_j^{(\ell)} D\left(\boldsymbol{\gamma}_{j+1}^{(\ell)}\right) \Psi_{j+1}^{(\ell)H} \boldsymbol{g}_{j+1}^{(\ell)}\right).$$

Equivalently, we have

$$\pi\left(\boldsymbol{g}_j | \boldsymbol{g}_{-j}^{(\ell)}, \boldsymbol{\alpha}^{(\ell)}, \boldsymbol{\beta}^{(\ell)}, \boldsymbol{\gamma}^{(\ell)}\right) \sim \mathbb{C}\mathcal{N}(\bar{\boldsymbol{g}}, \Sigma^{-1}), \tag{34}$$

and since (34) is complex Gaussian, the mode is also given by $\bar{\boldsymbol{g}}$.

A.2. **The $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, and $\boldsymbol{\gamma}$ updates.** Since $\boldsymbol{\alpha}$ is only present in (17) and (18) and each α_j is conditionally independent from α_{-j}, for the $\boldsymbol{\alpha}$ update in (25) we have

$$\pi\left(\boldsymbol{\alpha} | G^{(\ell+1)}, \boldsymbol{\beta}^{(\ell)}, \boldsymbol{\gamma}^{(\ell)}\right) \propto \prod_{j=1}^{J} \pi\left(\alpha_j | \boldsymbol{g}_j^{(\ell+1)}\right),$$

which is maximized by maximizing each $\pi\left(\alpha_j | \boldsymbol{g}_j^{(\ell+1)}\right)$, $j = 1, \ldots, J$. In this case, from (23) we have

$$\pi\left(\alpha_j | \boldsymbol{g}_j^{(\ell+1)}\right) \propto \alpha_j^M \exp\left(-\alpha_j \left\| \mathcal{F}_j \boldsymbol{g}_j^{(\ell+1)} - \hat{\boldsymbol{g}}_j \right\|_2^2\right) \alpha_j^{\eta_\alpha - 1} \exp(-\nu_\alpha \alpha_j). \tag{35}$$

Since the distribution in (35) is continuous, the maximum is obtained simply by solving $\frac{\mathrm{d}}{\mathrm{d}\alpha_j} \pi\left(\alpha_j | \boldsymbol{g}_j^{(\ell+1)}\right) = 0$. In this regard, observe that

$$\frac{\mathrm{d}}{\mathrm{d}\alpha_j} \pi\left(\alpha_j | \boldsymbol{g}_j^{(\ell+1)}\right) \propto \left(\frac{M + \eta_\alpha - 1}{\alpha_j} - \left\| \mathcal{F}_j \boldsymbol{g}_j^{(\ell+1)} - \hat{\boldsymbol{g}}_j \right\|_2^2 - \nu_\alpha\right)$$

$$\times \alpha_j^{M+\eta_\alpha - 1} \exp\left(-\alpha_j \left(\left\| \mathcal{F}_j \boldsymbol{g}_j^{(\ell+1)} - \hat{\boldsymbol{g}}_j \right\|_2^2 + \nu_\alpha\right)\right).$$

Setting the right hand side to zero yields

$$\alpha_j = \frac{\eta_\alpha + M - 1}{\nu_\alpha + \left\| \mathcal{F}_j \boldsymbol{g}_j^{(\ell+1)} - \hat{\boldsymbol{g}}_j \right\|_2^2}.$$

The $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ updates in (26) and (27) are similarly derived.

REFERENCES

[1] G. H. Aranda-Bojorges, B. P. Garcia-Salgado, V. I. Ponomaryov, O. Lopez-García and R. Reyes-Reyes, Despeckling of SAR images using GPU based on 3D-MAP estimation, *Real-Time Image Processing and Deep Learning*, **12102** (2022), 181-192.

[2] C. D. Austin, E. Ertin and R. L. Moses, Sparse multipass 3D SAR imaging: Applications to the GOTCHA data set, *Algorithms for Synthetic Aperture Radar Imagery XVI*, **7337** (2009), 19-30.

[3] A. H. Barnett, J. Magland and L. AF Klinteberg, A parallel nonuniform fast Fourier transform library based on an "exponential of semicircle" kernel, *SIAM Journal on Scientific Computing*, **41** (2019), C479-C504.
[4] S. Boyd, N. Parikh, E. Chu, B. Peleato and J. Eckstein, Distributed optimization and statistical learning via the alternating direction method of multipliers, *Foundations and Trends®in Machine Learning*, **3** (2011), 1-122.
[5] D. Calvetti and E. Somersalo, Inverse problems: From regularization to Bayesian inference, *Wiley Interdisciplinary Reviews: Computational Statistics*, **10** (2018), e1427.
[6] E. J. Candès and M. B. Wakin, An introduction to compressive sampling, *IEEE Signal Processing Magazine*, **25** (2008), 21-30.
[7] L. Chen, D. An, X. Huang and Z. Zhou A 3D reconstruction strategy of vehicle outline based on single-pass single-polarization CSAR data, *IEEE Transactions on Image Processing*, **26** (2017), 5545-5554.
[8] M. Cheney, A mathematical tutorial on synthetic aperture radar, *SIAM Review*, **43** (2001), 301-312.
[9] M. Cheney and B. Borden, Synthetic aperture radar imaging, in *Handbook of Mathematical Methods in Imaging*, Springer New York, 2015, 763-799.
[10] V. Churchill, *Synthetic Aperture Radar Image Formation with Uncertainty Quantification*, Ph.D thesis, Dartmouth College, 2020.
[11] V. Churchill and A. 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), 1225-1249.
[12] V. Churchill and A. Gelb, Sub-aperture SAR imaging with uncertainty quantification, *Inverse Problems*, **39** (2023), 054004.
[13] R. E. Crochiere and L. R. Rabiner, Interpolation and decimation of digital signals-A tutorial review, *Proceedings of the IEEE*, **69** (1981), 300-331.
[14] A. W. Doerry, E. E. Bishop and J. A. Miller, Basics of backprojection algorithm for processing synthetic aperture radar images, *Sandia Report SAND2016-1682*, Unlimited Release, (2016), 59.
[15] M.-P. Dubuisson and A. K. Jain, A modified Hausdorff distance for object matching, *Proceedings of the 12th International Conference on Pattern Recognition*, Jerusalem, Israel, 1994, 566-568.
[16] J. Glaubitz, A. Gelb and G. Song, Generalized sparse Bayesian learning and application to image reconstruction, *SIAM/ASA Journal on Uncertainty Quantification*, **11** (2023), 262-284.
[17] L. A. Gorham and L. J. Moore, SAR image formation toolbox for MATLAB, *Algorithms for Synthetic Aperture Radar Imagery XVII*, **7699** (2010), 46-58.
[18] C. V. Jakowatz, D. E. Wahl, P. H. Eichel, D. C. Ghiglia and P. A. Thompson, *Spotlight-Mode Synthetic Aperture Radar: A Signal Processing Approach*, Springer Science & Business Media, 2012.
[19] J. R. Jamora, *Angular-Dependent Three-Dimensional Imaging Techniques in Multi-Pass Synthetic Aperture Radar*, Master's thesis, Mississippi State University, 2021.
[20] J. R. Jamora, D. Green, A. Talley and T. Curry, Utilizing SAR imagery in three-dimensional neural radiance fields-based applications, *Algorithms for Synthetic Aperture Radar Imagery XXX*, **12520** (2023), 1252002.
[21] J. R. Jamora, P. Sotirelis, A. Nolan, J. Walrath, R. Hubbard, R. Weerasinghe, E. Young and S. 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*, **12095** (2022), 32-43.
[22] S. Ji, Y. Xue and L. Carin Bayesian compressive sensing, *IEEE Transactions on Signal Processing*, **56** (2008), 2346-2356.
[23] T. Li and L. Du, Target discrimination for SAR ATR based on scattering center feature and K-center one-class classification, *IEEE Sensors Journal*, **18** (2018), 2453-2461.
[24] H. Mao, Q. Yu and T. Zhang, Matching SAR image to optical image using modified Hausdorff distance and genetic algorithms, *MIPPR 2007: Pattern Recognition and Computer Vision*, SPIE, **6788** (2007), 532-537.
[25] D. C. Munson, J. D. O'Brien and W. K. Jenkins, A tomographic formulation of spotlight-mode synthetic aperture radar, *Proceedings of the IEEE*, **71** (1983), 917-925.
[26] R. Ng, Fourier slice photography, *ACM SIGGRAPH 2005 Papers*, (2005), 735-744.

[27] J.-C. Ni, Q. Zhang, Y. Luo and L. Sun, Compressed sensing SAR imaging based on centralized sparse representation, *IEEE Sensors Journal*, **18** (2018), 4920-4932.
[28] J.-I. Park, S.-H. Park and K.-T. Kim, New discrimination features for SAR automatic target recognition, *IEEE Geoscience and Remote Sensing Letters*, **10** (2012), 476-480.
[29] C. R. Paulson, *Utilizing Glint Phenomenology to Perform Classification of Civilian Vehicles Using Synthetic Aperture Radar*, Ph.D thesis, University of Florida, 2013.
[30] B. D. Rigling and R. L. Moses, Polar format algorithm for bistatic SAR, *IEEE Transactions on Aerospace and Electronic Systems*, **40** (2004), 1147-1159.
[31] T. Sanders, A. Gelb and R. B. Platte, Composite SAR imaging using sequential joint sparsity, *Journal of Computational Physics*, **338** (2017), 357-370.
[32] T. Scarnati and J. R. Jamora, Three-dimensional object reconstruction from sparse multi-pass SAR data, *Algorithms for Synthetic Aperture Radar Imagery XXVIII*, SPIE, **11728** (2021), 143-155.
[33] Y. Shi, X. X. Zhu and R. Bamler, Nonlocal compressive sensing-based SAR tomography, *IEEE Transactions on Geoscience and Remote Sensing*, **57** (2019), 3015-3024.
[34] M. E. Tipping, Sparse Bayesian learning and the relevance vector machine, *Journal of Machine Learning Research*, **1** (2001), 211-244.
[35] Y. Xiao and J. Glaubitz, Sequential image recovery using joint hierarchical Bayesian learning, *Journal of Scientific Computing*, **96** (2023), 4.
[36] J. Zhang, A. Gelb and T. Scarnati, Empirical Bayesian inference using a support informed prior, *SIAM/ASA Journal on Uncertainty Quantification*, **10** (2022), 745-774.
[37] Z. Zhang, H. Lei and Z. Lv, Vehicle layover removal in circular SAR images via ROSL, *IEEE Geoscience and Remote Sensing Letters*, **12** (2015), 2413-2417.
[38] Q. Zhao and J. C. Principe, Support vector machines for SAR automatic target recognition, *IEEE Transactions on Aerospace and Electronic Systems*, **37** (2001), 643-654.

Received July 2023; revised August 2023; early access September 2023.