remote sensing

MDPI

Article

Sparse Bayesian Learning Based Three-Dimensional Imaging Algorithm for Off-Grid Air Targets in MIMO Radar Array

Zekun Jiao [1,2,3], Chibiao Ding [1,2,*], Xingdong Liang [1,2], Longyong Chen [1,2] and Fubo Zhang [1,2]

[1] National Key Laboratory of Science and Technology on Microwave Imaging, Beijing 100190, China; ustcjzk@gmail.com (Z.J.); xdliang@mail.ie.ac.cn (X.L.); lychen@mail.ie.ac.cn (L.C.); zhangfubo8866@126.com (F.Z.)
[2] Institute of Electronics, Chinese Academy of Sciences, Beijing 100190, China
[3] University of the Chinese Academy of Sciences, Beijing 100049, China
* Correspondence: cbding@mail.ie.ac.cn; Tel.: +86-138-0107-7025

Received: 20 December 2017; Accepted: 26 February 2018; Published: 27 February 2018

Abstract: In recent years, the development of compressed sensing (CS) and array signal processing provides us with a broader perspective of 3D imaging. The CS-based imaging algorithms have a better performance than traditional methods. In addition, the sparse array can overcome the limitation of aperture size and number of antennas. Since the signal to be reconstructed is sparse for air targets, many CS-based imaging algorithms using a sparse array are proposed. However, most of those algorithms assume that the scatterers are exactly located at the pre-discretized grids, which will not hold in real scene. Aiming at finding an accurate solution to off-grid target imaging, we propose an off-grid 3D imaging method based on improved sparse Bayesian learning (SBL). Besides, the Bayesian Cramér-Rao Bound (BCRB) for off-grid bias estimator is provided. Different from previous algorithms, the proposed algorithm adopts a three-stage hierarchical sparse prior to introduce more degrees of freedom. Then variational expectation maximization method is applied to solve the sparse recovery problem through iteration, during each iteration joint sparsity is used to improve efficiency. Experimental results not only validate that the proposed method outperforms the existing off-grid imaging methods in terms of accuracy and resolution, but have compared the root mean square error with corresponding BCRB, proving effectiveness of the proposed method.

Keywords: three-dimensional imaging; off-grid model; sparse Bayesian learning; variational expectation maximization; Bayesian Cramér-Rao bound

1. Introduction

In the last few years, three-dimensional (3D) radar imaging systems and algorithms have received significant attention among researchers worldwide. Different from the 2D radar image, which is a projection of the 3D model, 3D radar image can provide shape and energy distribution of dominated scatterers of the target [1,2]. This information is of great significance to target identification and diagnostic analysis. Multiple-input–multiple-output (MIMO) radar system has drawn much attention and shows its potential in 3D imaging in the last decade [3]. For one thing, it can overcome the problem of limited number of antennas and limited aperture size. For another, the coherent processing interval (CPI) is reduced in MIMO radar by using space sampling instead of time sampling [4]. In traditional interferometric inverse synthetic aperture radar (InISAR), the imaging performance is severely degraded for high maneuvering target because the rotation axis of the target relative to the radar is time varying during the CPI [5,6]. Compared with the InISAR technique, the 3D image can be reconstructed in one snapshot [7] with a wideband monostatic MIMO radar system, thus free

from the target's high mobility. In the last few years, compressed sensing (CS) has been widely used in sparse signal recovery with fewer samples [8]. It has shown great advantage in radar imaging due to its super-resolution ability. It is proven that CS-based imaging algorithms provide a better resolution enhancement effect than the RELAX algorithm [9]. Besides, it can also exploit the sparsity of signal and reconstruct the signal from limited samples with high probability. Thus, the combination of MIMO radar and CS technique becomes a hot topic, and much existing research focus on CS recovery techniques for MIMO radar imaging [4,10–12]. In Ding et al. [4], a CS-based 2D imaging method based on sparse array is proposed. In Hu et al. [12], a narrowband 3D imaging method based on Kronecker CS is discussed. As a matter of fact, the sparse recovery algorithms severely affect the imaging performances. Hence, sparse recovery algorithm with high accuracy is of fundamental importance in MIMO 3D imaging.

The CS-based imaging methods, such as orthogonal matching pursuit (OMP) or basis pursuit (BP), can achieve a better imaging performance in the cases of low signal-to-noise-ratio (SNR) and limited snapshots. However, most of the existing CS-based imaging methods require that the scatterers of target to be exactly on the discretized sampling grid [13]. Otherwise, the off-grid problem will introduce basis mismatch and leakage of energy over all grids, finally lead to the deterioration of imaging performance. In practical MIMO radar imaging, the targets are distributed in the continuous space, and the off-grid problem always exists. There are mainly two ways to solve the problem. One is to compensate for the mismatch, the other is to use the gridless sparse recovery [14]. The latter is more accurate but the guaranteed theoretical resolution of this kind of method is a few Rayleigh limits [15]. So in practical CS imaging, we still use the former method. A denser grid may mitigate the effect of mismatch to some extent, but it will increase the mutual coherence of the sensing matrix [16]. This may cause violation of the restricted isometry property (RIP) property, which is the guarantee for reliable recovery. The computation cost will also increase sharply with a denser grid. Hence, CS recovery with a perturbed sensing matrix has been a hot topic. In Chi et al. [17], the performance of CS recovery is analyzed when the mismatch problem exists, nevertheless, it does not provide any algorithm for off-grid CS recovery. Sparsity-cognizant total least-squares (S-TLS) is the first recovery method for perturbed compressed sensing problem [18]. The S-TLS algorithm supposes that the off-grid bias obeys the Gaussian distribution. However, without any prior information, the uniform distribution is more suitable for the off-grid bias (the distance between the true scatterer and its nearest grid). A perturbation approach is established for compressive radar imaging based on OMP [13], but it does not use the statistical information and the performance is not satisfying. An off-grid direction of arrival (DOA) estimator based on sparse Bayesian inference (OGSBI) is proposed in Yang et al. [19], which adopts the uniform distribution. However, research on MIMO radar CS imaging are mainly restricted to one dimensional(1D) resolution enhancement [12]. Study on problems with high dimensionality has just begun. Hence, this study sets out to find an algorithm for high dimensionality MIMO imaging for off-grid air target.

Starting from the purpose of finding a 3D imaging method with high accuracy for maneuvering off-grid air target, we propose a novel algorithm based on improved sparse Bayesian learning to overcome the aforementioned problems. Under this framework, we first derive the on-grid and off-grid 3D imaging model using a sparse array. Then the three-stage hierarchical sparse prior is adopted, which introduces more degrees of freedom. Finally, this algorithm can get the approximated analytic expressions of the unknown parameters by using the variational expectation maximization (EM) method. Besides, both the imaging result and estimations of unknown parameters such as off-grid biases can be obtained at the same time. Simulation results prove the superiority of the proposed algorithm over other existing off-grid sparse recovery algorithms such as S-TLS and OGSBI. In addition, the BCRB and the Mean Square Error (MSE) of off-grid bias estimator are compared to verify effectiveness of the proposed method.

Notations used in this paper are as follows. Bold-case letters are reserved for vectors and matrices. $\hat{a}$ is the expectation of a. $diag(\boldsymbol{x})$ is a matrix with its main diagonal being $\boldsymbol{x}$. $\odot$ denotes the Hardmard

product and $\Re\{\boldsymbol{x}\}$ means the real part of complex $\boldsymbol{x}$. $\mathcal{K}_p(\cdot)$ is the modified Bessel function of the 3rd kind. $(\cdot)^{\boldsymbol{T}}$, $(\cdot)^{\boldsymbol{H}}$ are the transpose and conjugate transpose operator, respectively. $\overline{\boldsymbol{A}}$ is the conjugate of $\boldsymbol{A}$. $E[\cdot]_p$ is the expectation operator which is taken with respect to the probability density function p. $x^{(i)}$ is the update of x in the ith iteration. $\boldsymbol{A}_i, \boldsymbol{A}^j, \boldsymbol{A}_{i,j}$ are the ith column, the jth row and the (i,j)th element of $\boldsymbol{A}$, respectively.

The remaining parts of this paper proceed as follows: the off-grid target 3D imaging problem based on sparse array is formulated in Section 2. In Section 3, an algorithm for 3D imaging based on improved sparse Bayesian learning is proposed and the BCRB for off-grid bias is also derived. In Section 4, experimental results and analyses of the imaging performance are presented. Finally, conclusions are drawn in Section 5.

2. Problem Formulation of 3D Imaging

2.1. Ideal Imaging Model Based on Sparse Antenna Array

The geometry of the 3D imaging based on sparse antenna array is depicted in Figure 1. In this paper, we consider a wideband MIMO radar system with M transmitters and N receivers. Considering the need for a good orthogonal characteristic of transmitted signals, we adopt a group of M orthogonal phase-code modulation (PCM) signals with the same carrier frequency and bandwidth in this system.

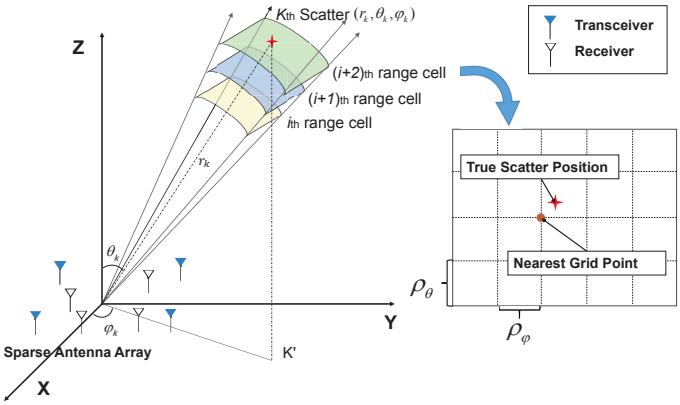

Figure 1. The imaging geometry of the proposed method.

Transmitted signal from the mth transmitter can be expressed as follows:

$$s_m(t) = rect(\frac{t - T_r}{T_r})\exp(\mathrm{j}\varphi_m(t))\exp(\mathrm{j}2\pi f_c t) \quad (m = 1, 2, \ldots, M) \tag{1}$$

In which T_r is the pulse width, f_c is the carrier frequency and $\varphi_m(t)$ is the phase-code function. $rect(\frac{t}{T})$ is defined as follows:

$$rect(\frac{t}{T}) = \begin{cases} 0 & |t| > \frac{T}{2} \\ 1 & |t| \leq \frac{T}{2} \end{cases} \tag{2}$$

This group of transmitted signals are orthogonal to each other.

$$\int_0^{T_r} s_p(t)s_q^*(t) = \begin{cases} 0 & \text{if } p \neq q \\ Constant & \text{if } p = q \end{cases} \tag{3}$$

Assuming that the transmitters and receivers are in the XOY plane, (x_m, y_m) and (x_n, y_n) are the coordinates of mth transmitter and nth receiver, respectively. Suppose that there are K dominant scatterers in the imaging scene. The coordinate of the kth scatterer is $(r_k, \theta_k, \varphi_k)$, and its RCS is σ_k.

According to the sparse radar array configuration in Figure 1, the transmit and receive steering vector can be described:

$$\boldsymbol{t}(k) = [1 \quad \ldots \quad \exp(\mathrm{j}\frac{2\pi}{\lambda}\gamma_t(m,k)) \quad \ldots \quad \exp(\mathrm{j}\frac{2\pi}{\lambda}\gamma_t(M,k))] \tag{4}$$

$$\boldsymbol{r}(k) = [1 \quad \ldots \quad \exp(\mathrm{j}\frac{2\pi}{\lambda}\gamma_r(n,k)) \quad \ldots \quad \exp(\mathrm{j}\frac{2\pi}{\lambda}\gamma_r(N,k))] \tag{5}$$

$$\gamma_t(m,k) = (x_m \sin\theta_k \cos\varphi_k + y_m \sin\theta_k \sin\varphi_k) \tag{6}$$

$$\gamma_r(n,k) = (x_n \sin\theta_k \cos\varphi_k + y_n \sin\theta_k \sin\varphi_k) \tag{7}$$

In which $\lambda = \frac{c}{f_c}$ is the carrier wavelength, and c is the speed of light. Define the transmitting vector as $\boldsymbol{s}(t) = [s_1(t), s_2(t), \ldots, s_M(t)]^T$, then the received signal can be expressed as:

$$\boldsymbol{g}(t) = \sum_{k=1}^{K} \sigma_k \boldsymbol{r}^T(k)\boldsymbol{t}(k)\boldsymbol{s}(t - \frac{2r_k}{c}) \tag{8}$$

After the matched filters are adopted, we can get the received signal from the mth transmitter and the nth receiver

$$g_{mn}(t) = \sum_{k=1}^{K} \sigma_k r_n(k) t_m(k) \psi_m(t - \frac{2r_k}{c}) \tag{9}$$

In Equation (9), $r_n(k)$ is the nth element of $\boldsymbol{r}(k)$, and $t_m(k)$ is similar. $\psi_m(t)$ is the autocorrelation function of $s_m(t)$, which is the 1D range image. Since we adopt a wideband sparse antenna array, the range resolution can be achieved through matched filtering. After above operations, we can get MN one-dimensional range images from the N receivers. Because the range resolution provided by the bandwidth is usually sufficient, we only use CS technique to enhance resolutions of the other two dimensions. After range alignment, we transfer the 3D imaging problem into a corresponding lower dimensional one. For the ith range cell, assuming that there are I out of K dominant scatterers falling in this range cell, echo return of the ith range cell can be formulated as follows:

$$y_{mn} = \sum_{i=1}^{I} \sigma_i r_n(i) t_m(i) \tag{10}$$

For the application of air targets imaging, the true scatterers are always sparse compared with the background. Thus, the sparsity of those targets are exploited and CS-based methods show its advantages. According to traditional CS-based imaging method, we then discretize the imaging scene into uniform grids. Assuming there are P grids for θ and Q grids for φ, i.e., $Z(= PQ)$ grids in total. Then the observed formula with noise can be described as:

$$y_{mn} = \sum_{p=1}^{P}\sum_{q=1}^{Q} \sigma_{p,q} \exp\{\mathrm{j}\frac{2\pi}{\lambda}[(x_m + x_n)\sin\theta_p \cos\varphi_q + [y_m + y_n]\sin\theta_p \sin\varphi_q]\} + e_{mn} \tag{11}$$

After data rearrangement and vectorization, i.e., $\sigma_i = \sigma_{p,q}, i = (q-1) \times P + p$, the signal model for sparse recovery is as follows:

$$y_{mn} = \sum_{i}^{Z} \sigma_i \exp\{\mathrm{j}\frac{2\pi}{\lambda}[(x_m + x_n)\sin\theta_i \cos\varphi_i + [y_m + y_n]\sin\theta_i \sin\varphi_i]\} + e_{mn} \tag{12}$$

$$\boldsymbol{y} = \boldsymbol{\Phi}_0 \sigma + \boldsymbol{e} \tag{13}$$

In Equation (13) $\boldsymbol{y}$ is a $MN \times 1$ vector standing for the echo return after range alignment. Φ_0 is a $MN \times Z$ matrix with its (i,j)th element being $\exp\{\mathrm{j}\frac{2\pi}{\lambda}[(x_m + x_n)\sin\theta_j \cos\varphi_j + [y_m + y_n]\sin\theta_j \sin\varphi_j]\}$,

in which $i = (n-1) \times M + m$. $\boldsymbol{\sigma}$ for different range cells can be extracted from Equation (13) via sparse recovery, which finally construct the 3D image.

2.2. Off-Grid Imaging Model Using Taylor Expansion

The imaging model in Equation (12) is under the assumption that all the dominant scatterers are exactly located on the pre-discretized grid. In practical radar imaging this assumption does not hold since the scatterers of real target are distributed in a continuous space. This means that the off-grid problem always exists. There are two off-grid models to solve the problem, one is based on the Taylor series expansion and the other based on linear interpolation. The performances of these two models are compared in Das [20]. The result is that the former has a better accuracy, so we choose the Taylor series expansion to formulate the off-grid imaging model. Considering the mismatch between scatterer's true position and the grids, the true position of a scatterer can be expressed as:

$$\begin{cases} \theta_{i0} = \theta_i + \delta_{\theta_i} \\ \varphi_{i0} = \varphi_i + \delta_{\varphi_i} \end{cases} \tag{14}$$

In which θ_{i0} and φ_{i0} are the true coordinates. θ_i and φ_i stand for the nearest grid point while δ_{θ_i} and δ_{φ_i} stand for the off-grid bias. Let $\boldsymbol{\Psi} = \{(\theta_1, \varphi_1), (\theta_2, \varphi_2), \ldots, (\theta_Z, \varphi_Z)\}$ be the uniform discretization of the 2D scene after vectorization and let the matrix $\boldsymbol{\Phi}_0$ in Equation (13) expressed as $\boldsymbol{\Phi}_0 = [\boldsymbol{a}(\theta_1, \varphi_1), \boldsymbol{a}(\theta_2, \varphi_2), \ldots, \boldsymbol{a}(\theta_Z, \varphi_Z)]$. The true position set of the scatterers is $\{(\theta_{10}, \varphi_{10}), \ldots, (\theta_{k0}, \varphi_{k0}), \ldots, (\theta_{K0}, \varphi_{K0})\}$. Supposing for a certain $k \in \{1, 2, \ldots, K\}$, its corresponding coordinate $(\theta_{k0}, \varphi_{k0}) \notin \boldsymbol{\Psi}$. The nearest grid point is $(\theta_k, \varphi_k), k \in \{1, 2, \ldots, Z\}$. We adopt the first order Taylor series expansion approximation of $\boldsymbol{a}(\theta_{k0}, \varphi_{k0})$ with respect to $(\delta_{\theta_k}, \delta_{\varphi_k})$ as:

$$\boldsymbol{a}(\theta_{k0}, \varphi_{k0}) \approx \boldsymbol{a}(\theta_k, \varphi_k) + \boldsymbol{b}(\theta_k, \varphi_k)(\theta_{k0} - \theta_k) + \boldsymbol{c}(\theta_k, \varphi_k)(\varphi_{k0} - \varphi_k) \tag{15}$$

In which $\boldsymbol{b}(\theta_k, \varphi_k)$ and $\boldsymbol{c}(\theta_k, \varphi_k)$ are the partial derivatives of $\boldsymbol{a}(\theta_k, \varphi_k)$ with respect to θ_k and φ_k.

Denote $\boldsymbol{\Phi}_\theta = [\boldsymbol{b}(\theta_1, \varphi_1), \boldsymbol{b}(\theta_2, \varphi_2), \ldots, \boldsymbol{b}(\theta_Z, \varphi_Z)]$, $\boldsymbol{\Phi}_\varphi = [\boldsymbol{c}(\theta_1, \varphi_1), \boldsymbol{c}(\theta_2, \varphi_2), \ldots, \boldsymbol{c}(\theta_Z, \varphi_Z)]$, $\boldsymbol{\delta}_\theta = [\delta_{\theta_1}, \delta_{\theta_2}, \ldots, \delta_{\theta_Z}]$ and $\boldsymbol{\delta}_\varphi = [\delta_{\varphi_1}, \delta_{\varphi_2}, \ldots, \delta_{\varphi_Z}]$. The off-grid bias vector $\boldsymbol{\delta}_\theta$ satisfies

$$\delta_{\theta_k} = \begin{cases} \theta_{k0} - \theta_k & \text{if } \sigma_k \neq 0 \\ 0 & \text{otherwise} \end{cases} \tag{16}$$

where $k = 1, 2, \ldots, Z$. It is the same for $\boldsymbol{\delta}_\varphi$. Hence, the off-grid imaging model corresponding to Equation (13) can be expressed as:

$$\boldsymbol{y} = (\boldsymbol{\Phi}_0 + \boldsymbol{\Phi}_\theta \boldsymbol{\Delta}_\theta + \boldsymbol{\Phi}_\varphi \boldsymbol{\Delta}_\varphi)\boldsymbol{\sigma} + \boldsymbol{e} \tag{17}$$

with $\boldsymbol{\Delta}_\theta = diag(\boldsymbol{\delta}_\theta)$ and $\boldsymbol{\Delta}_\varphi = diag(\boldsymbol{\delta}_\varphi)$. With this off-grid imaging model, our goal is to find the optimal $\boldsymbol{\sigma}, \boldsymbol{\delta}_\theta, \boldsymbol{\delta}_\varphi$ simultaneously, after which the accurate 3D image can be reconstructed.

3. The Proposed Off-Grid Imaging Algorithm

According to Equation (17), the off-grid 3D imaging problem is presented. The off-grid bias can be seen as a kind of multiplicative noise, thus the traditional sparse recovery algorithms, such as OMP and BP, fail to solve this problem. In Yang et al. [19], OGSBI is proposed and proves its potential to the off-grid DOA estimation problem. Inspired by the idea in Bishop [21], we propose an algorithm based on sparse Bayesian learning using a three-stage sparse prior to solve the off-grid imaging problem. In this section, we first describe the SBL-based off-grid imaging algorithm. Then the Bayesian Cramér-Rao bounds for off-grid bias estimator are presented.

3.1. Algorithm Description

3.1.1. The Three-Stage Sparse Prior Model

The overall graphical model of the three-stage sparse prior is shown in Figure 2.

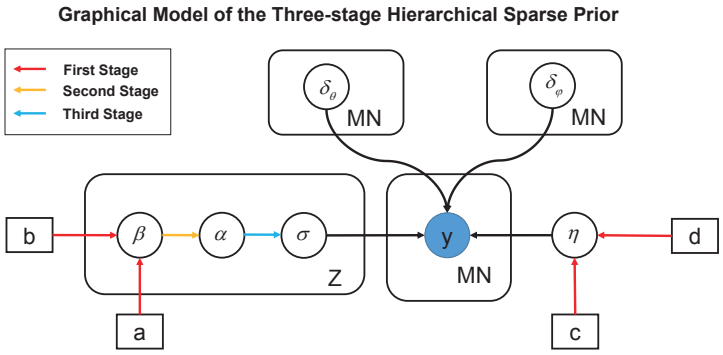

Figure 2. The three-stage sparse prior model.

Starting from a Bayesian perspective, all the knowns in Equation (17) are assigned with probability distributions. Then we can estimate $\boldsymbol{\sigma}, \boldsymbol{\delta}_\theta, \boldsymbol{\delta}_\varphi$ based on the maximum a posteriori (MAP) criterion. Note that the probability density function of a complex Gaussian distributed vector $\boldsymbol{x} \sim \mathcal{CN}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with its mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$ is:

$$\mathcal{CN}(\boldsymbol{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{\pi^Z |\boldsymbol{\Sigma}|} \exp(-(\boldsymbol{x} - \boldsymbol{\mu})^H \boldsymbol{\Sigma}^{-1} (\boldsymbol{x} - \boldsymbol{\mu})) \tag{18}$$

First, we start from the sparse prior model for the radar cross section (RCS) vector $\boldsymbol{\sigma}$. Here we adopt a three-stage sparse prior for $\boldsymbol{\sigma}$. Supposing that the RCS vector $\boldsymbol{\sigma}$ has a complex Gaussian distribution, $\boldsymbol{\sigma} \sim \mathcal{CN}(0, \boldsymbol{\Sigma})$, which corresponds to the well known Swerling-1 model. In addition, the covariance matrix $\boldsymbol{\Sigma} = diag(\boldsymbol{\alpha})$, in which $\boldsymbol{\alpha} = [\alpha_1, \alpha_2, \ldots, \alpha_Z]^T$. The reason why we choose this probability distribution function is that a sparse prior is needed under the Bayesian framework. Since $\boldsymbol{\sigma}$ is sparse, that is to say, most of the elements in $\boldsymbol{\sigma}$ are equal or close to zero, we choose the Gaussian zero-mean model as the sparse prior. Along with the three-stage sparse prior model in the following, $\boldsymbol{\sigma}$ is strongly peaked at the origin. As a result, this Gaussian zero-mean prior favors that most elements of $\boldsymbol{\sigma}$ being zero. It is shown in Bishop [21] that both Gaussian distribution and Gamma distribution belong to the exponential distribution family and they are a pair of conjugate priors. This property leads to the posterior function having the same functional form as the prior and to a simplified Bayesian analysis. So we assume a Gamma distribution for $\boldsymbol{\alpha}$:

$$p(\boldsymbol{\alpha}|\boldsymbol{\beta}) = \prod_{i=1}^{Z} \Gamma(\alpha_i|1, \beta_i) \tag{19}$$

with $\Gamma(x|a, b) = [\Gamma(a)]^{-1} b^a x^{a-1} e^{-bx}$ and $\Gamma(\cdot)$ being the Gamma function. The first stage of the sparse prior model is that $\boldsymbol{\beta}$ also follows a Gamma distribution with $a, b > 0$:

$$p(\boldsymbol{\beta}|a, b) = \prod_{i=1}^{Z} \Gamma(\beta_i|a, b) \tag{20}$$

As is shown in Figure 2, this three-stage prior is a sparse prior of $\boldsymbol{\sigma}$. Compared with the method in Yang et al. [19], this model introduces Z more hyperparameters into the sparse recovery model to offer more degrees of freedom.

Secondly, assuming that the noise is white complex Gaussian, we have:

$$p(\boldsymbol{e}|\eta) = \mathcal{CN}(\boldsymbol{e}|0, \eta \boldsymbol{I}) \tag{21}$$

where η being the noise variance. Similarly, a conjugate prior is assigned that $p(\eta|c,d) = \Gamma(\eta|c,d)$. It is worth mentioning here why the sensing noise $\boldsymbol{e}$ is assumed zero-mean. In Equation (15), a first order Taylor series expansion approximation is adopted and higher order items are neglected. So sensing noise in Equation (17) only contains the measurement noise which is assumed white Gaussian distributed with zero mean.

Finally, for the off-grid bias δ_θ and δ_ϕ, the uniform distribution is more suitable than the Gaussian distribution since we do not have any information about them. Supposing that the pre-discretized grid intervals are ρ_θ and ρ_φ, we get:

$$\delta_\theta \sim U\Big([-\frac{\rho_\theta}{2}, \frac{\rho_\theta}{2}]^Z\Big) \qquad \delta_\varphi \sim U\Big([-\frac{\rho_\varphi}{2}, \frac{\rho_\varphi}{2}]^Z\Big) \tag{22}$$

3.1.2. Variational Inference EM Based Sparse Recovery Algorithm

Under the framework of sparse Bayesian learning, we always seek to find the optimal estimations from the MAP criterion. The posterior distribution can be expressed by the following equations:

$$p(\boldsymbol{\sigma}, \boldsymbol{\delta}_\theta, \boldsymbol{\delta}_\varphi, \boldsymbol{\alpha}, \boldsymbol{\beta}, \eta|\boldsymbol{y}) = p(\boldsymbol{y}, \boldsymbol{\sigma}, \boldsymbol{\delta}_\theta, \boldsymbol{\delta}_\varphi, \boldsymbol{\alpha}, \boldsymbol{\beta}, \eta)/p(\boldsymbol{y}) \tag{23}$$

$$p(\boldsymbol{y}, \boldsymbol{\sigma}, \boldsymbol{\delta}_\theta, \boldsymbol{\delta}_\varphi, \boldsymbol{\alpha}, \boldsymbol{\beta}, \eta) = p(\boldsymbol{y}|\boldsymbol{\sigma}, \boldsymbol{\delta}_\theta, \boldsymbol{\delta}_\varphi, \eta)p(\boldsymbol{\sigma}|\boldsymbol{\alpha})p(\boldsymbol{\alpha}|\boldsymbol{\beta})p(\boldsymbol{\beta})p(\boldsymbol{\delta}_\theta)p(\boldsymbol{\delta}_\varphi)p(\eta) \tag{24}$$

$$p(\boldsymbol{y}|\boldsymbol{\sigma}, \boldsymbol{\delta}_\theta, \boldsymbol{\delta}_\varphi, \eta) = \mathcal{CN}(\boldsymbol{y}|(\boldsymbol{\Phi}_0 + \boldsymbol{\Phi}_\theta \boldsymbol{\Delta}_\theta + \boldsymbol{\Phi}_\varphi \boldsymbol{\Delta}_\varphi)\boldsymbol{\sigma}, \eta \boldsymbol{I}) \tag{25}$$

However, according to Equation (24) the marginalized likelihood function

$$p(\boldsymbol{y}) = \oint p(\boldsymbol{y}|\boldsymbol{\sigma}, \boldsymbol{\delta}_\theta, \boldsymbol{\delta}_\varphi, \eta)p(\boldsymbol{\sigma}|\boldsymbol{\alpha})p(\boldsymbol{\alpha}|\boldsymbol{\beta})p(\boldsymbol{\beta})p(\boldsymbol{\delta}_\theta)p(\boldsymbol{\delta}_\varphi)p(\eta)d\boldsymbol{\sigma}d\boldsymbol{\delta}_\theta d\boldsymbol{\delta}_\varphi d\boldsymbol{\alpha}d\boldsymbol{\beta}d\eta \tag{26}$$

which cannot be calculated analytically, thus the normalized constant in Equation (23) cannot be computed. In Blei et al. [22], the author gives a solution to this kind of problem using variational inference EM algorithm. Hence, we adopt the variational EM algorithm in this sparse recovery problem. According to the jargon in the EM algorithm, we treat $\boldsymbol{\sigma}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \eta$ as the hidden variables while $\boldsymbol{\delta}_\theta, \boldsymbol{\delta}_\varphi$ being the parameters. Since the variational inference EM algorithm is a two-stage iterative optimization algorithm, assuming that $\boldsymbol{\sigma}^{(i)}, \boldsymbol{\alpha}^{(i)}, \boldsymbol{\beta}^{(i)}, \eta^{(i)}, \boldsymbol{\delta}_\theta^{(i)}, \boldsymbol{\delta}_\varphi^{(i)}$ are obtained in the ith iteration, we now seek for their updates in the $(i+1)$th iteration.

(1) The E-step

Since the posterior distribution $p(\boldsymbol{\sigma}, \boldsymbol{\delta}_\theta, \boldsymbol{\delta}_\varphi, \boldsymbol{\alpha}, \boldsymbol{\beta}, \eta|\boldsymbol{y})$ cannot be calculated analytically, we use the variational inference technique in Tzikas et al. [23] to find an approximation for the posterior distribution:

$$p(\boldsymbol{\sigma}, \boldsymbol{\delta}_\theta, \boldsymbol{\delta}_\varphi, \boldsymbol{\alpha}, \boldsymbol{\beta}, \eta|\boldsymbol{y}) \approx q(\boldsymbol{\sigma}, \boldsymbol{\delta}_\theta, \boldsymbol{\delta}_\varphi, \boldsymbol{\alpha}, \boldsymbol{\beta}, \eta|\boldsymbol{y}) := q(\boldsymbol{\sigma})q(\boldsymbol{\delta}_\theta)q(\boldsymbol{\delta}_\varphi)q(\boldsymbol{\alpha})q(\boldsymbol{\beta})q(\eta) \tag{27}$$

which minimizes the Kullback-Leibler divergence. This approximation comes from the variational methods used in Bayesian inference. Among these applications, a particular form that has been used with great success is the factorized one which is used in this paper. In a broad sense, this approximation is always valid, but the performance is effected by the form of the prior distributions. In our method, we use the conjugate priors in order to get an analytical expression, and the performance of the approximation in Equation (27) is satisfactory. Based on the above equation, we can get the estimation of the hidden variables. The detailed procedures are as follows.

For $\boldsymbol{\sigma}$, we have:

$$\begin{aligned} lnq(\boldsymbol{\sigma}) &= E[lnp(\boldsymbol{y}|\boldsymbol{\sigma}, \boldsymbol{\delta}_\theta, \boldsymbol{\delta}_\varphi, \eta) + lnp(\boldsymbol{\sigma}|\boldsymbol{\alpha})]_{q(\boldsymbol{\delta}_\theta)q(\boldsymbol{\delta}_\varphi)q(\boldsymbol{\alpha})q(\eta)} + C \\ &\propto -\boldsymbol{\sigma}^H[\hat{\eta}\boldsymbol{\Phi}^H\boldsymbol{\Phi} + diag(1/\hat{\boldsymbol{\alpha}})]\boldsymbol{\sigma} - 2\hat{\eta}\boldsymbol{y}^H\boldsymbol{\Phi}\boldsymbol{\sigma} \end{aligned} \tag{28}$$

where $\boldsymbol{\Phi} = \boldsymbol{\Phi}_0 + \boldsymbol{\Phi}_\theta diag(\hat{\boldsymbol{\delta}}_\theta) + \boldsymbol{\Phi}_\varphi diag(\hat{\boldsymbol{\delta}}_\varphi)$. According to the above equation, $q(\boldsymbol{\sigma})$ is a complex Gaussian distribution, i.e., $\boldsymbol{\sigma} \sim \mathcal{CN}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with

$$\begin{aligned} \boldsymbol{\mu} &= \hat{\eta}\boldsymbol{\Sigma}\boldsymbol{\Phi}^H\boldsymbol{y} \\ \boldsymbol{\Sigma} &= [\hat{\eta}\boldsymbol{\Phi}^H\boldsymbol{\Phi} + diag(1/\hat{\boldsymbol{\alpha}})]^{-1} \end{aligned} \tag{29}$$

Since we have already obtained both hidden variables and parameters in the ith iteration, the expectations in Equation (29) are all replaced by the ith updates, i.e.,

$$\begin{aligned} \boldsymbol{\mu}^{(i+1)}(\boldsymbol{\sigma}^{(i+1)}) &= \eta^{(i)}\boldsymbol{\Sigma}^{(i+1)}\boldsymbol{\Phi}^H\boldsymbol{y} \\ \boldsymbol{\Sigma}^{(i+1)} &= [\eta^{(i)}\boldsymbol{\Phi}^H\boldsymbol{\Phi} + diag(1/\boldsymbol{\alpha}^{(i)})]^{-1} \\ \boldsymbol{\Phi} &= \boldsymbol{\Phi}_0 + \boldsymbol{\Phi}_\theta diag(\boldsymbol{\delta}_\theta^{(i)}) + \boldsymbol{\Phi}_\varphi diag(\boldsymbol{\delta}_\varphi^{(i)}) \end{aligned} \tag{30}$$

In the following part, the expectations of these unknowns are replaced by their corresponding latest updates.

For $\boldsymbol{\alpha}$, we have:

$$\begin{aligned} lnq(\boldsymbol{\alpha}) &= E[lnp(\boldsymbol{\sigma}|\boldsymbol{\alpha}) + lnp(\boldsymbol{\alpha}|\boldsymbol{\beta})]_{q(\boldsymbol{\sigma})q(\boldsymbol{\beta})} + C \\ lnq(\boldsymbol{\alpha}_j) &\propto -\hat{\sigma_j^2}\alpha_j^{-1} - ln\alpha_j - \hat{\beta}_j\alpha_j \end{aligned} \tag{31}$$

Thus, the elements in $\boldsymbol{\alpha}$ are independent and their distribution is a generalized inverse Gaussian distribution [24]. Here we use the expectation of $\boldsymbol{\alpha}_j$ to get the $(i+1)$th update for $\boldsymbol{\alpha}^{(i+1)}$. For the jth element in $\boldsymbol{\alpha}$:

$$\alpha_j^{(i+1)} = (\frac{\xi_j^{(i+1)}}{\beta_j^{(i)}})^{1/2}\frac{\kappa_1(2\sqrt{\beta_j^{(i)}\xi_j^{(i+1)}})}{\kappa_0(2\sqrt{\beta_j^{(i)}\xi_j^{(i+1)}})} \tag{32}$$

In which $\xi_j^{(i+1)} = (\sigma_j^2)^{(i+1)} = |\sigma_j^{(i+1)}|^2 + \Sigma_{j,j}^{(i+1)}$.

For $\boldsymbol{\beta}$, we have:

$$\begin{aligned} lnq(\boldsymbol{\beta}) &= E[lnp(\boldsymbol{\alpha}|\boldsymbol{\beta}) + lnp(\boldsymbol{\beta})]_{q(\boldsymbol{\alpha})} + C \\ lnq(\boldsymbol{\beta}_j) &\propto aln\beta_j - (b + \hat{\alpha_j})\beta_j \end{aligned} \tag{33}$$

Similarly, elements in $\boldsymbol{\beta}$ are independent and $p(\beta_j) = \Gamma(\beta_j|a+1, b+\alpha_j)$. The $(i+1)$th update of $\boldsymbol{\beta}$ are as follows:

$$\beta_j^{(i+1)} = \frac{a+1}{b + \alpha_j^{(i+1)}} \tag{34}$$

For η, we have:

$$\begin{aligned} lnq(\eta) &= E[lnp(\boldsymbol{y}|\boldsymbol{\sigma}, \boldsymbol{\delta}_\theta, \boldsymbol{\delta}_\varphi, \eta) + lnp(\eta)]_{q(\boldsymbol{\alpha})q(\boldsymbol{\delta}_\theta)q(\boldsymbol{\delta}_\varphi)} + C \\ &\propto (Z + c - 1)ln\eta - \Big(\|\boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{\mu}\|_F^2 + tr(\boldsymbol{\Phi}^H\boldsymbol{\Phi}\boldsymbol{\Sigma}) + d\Big)\eta \end{aligned} \tag{35}$$

where $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are defined in Equation (29). Similar to $\boldsymbol{\beta}$, $q(\eta)$ is also a Gamma distribution. Combining the expression for $\boldsymbol{\Sigma}$ in Equation (29) and the expectation for a Gamma distribution, we can get:

$$\eta^{(i+1)} = \frac{Z + c}{\|\boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{\mu}^{(i+1)}\|_F^2 + d + \frac{1}{\eta^{(i)}} \sum_{n=1}^{Z} (1 - \frac{1}{\alpha_n^{(i+1)}} \Sigma_{n,n}^{(i+1)})} \tag{36}$$

(2) The M-step

For $\boldsymbol{\delta}_\theta$, its estimate maximize $E[lnp(\boldsymbol{y}|\boldsymbol{\sigma}, \boldsymbol{\delta}_\theta, \boldsymbol{\delta}_\varphi, \eta) + lnp(\boldsymbol{\delta}_\theta)]$, which equivalent to minimizing

$$\begin{aligned} & E\left\{ \|\boldsymbol{y} - \left[\boldsymbol{\Phi}_0 + \boldsymbol{\Phi}_\theta diag(\boldsymbol{\delta}_\theta) + \boldsymbol{\Phi}_\varphi diag(\boldsymbol{\delta}_\varphi)\right]\boldsymbol{\sigma}\|_2^2 \right\} \\ & = \boldsymbol{\delta}_\theta^T \boldsymbol{A}_1 \boldsymbol{\delta}_\theta + 2\boldsymbol{a}_1^T \boldsymbol{\delta}_\theta + C \end{aligned} \tag{37}$$

in which

$$\begin{aligned} \boldsymbol{A}_1 &= \Re\{\overline{\boldsymbol{\Phi}_1^H \boldsymbol{\Phi}_1} \odot (\boldsymbol{\Sigma}^{(i+1)} + \boldsymbol{\mu}^{(i+1)} (\boldsymbol{\mu}^{(i+1)})^H)\} \\ \boldsymbol{a}_1 &= \Re\{diag(\boldsymbol{\Phi}_\theta^H \boldsymbol{\Phi}_1 \boldsymbol{\Sigma}^{(i+1)}) - diag(\boldsymbol{\mu}^{(i+1)}) \boldsymbol{\Phi}_\theta^H (\boldsymbol{y} - \boldsymbol{\Phi}_1 \boldsymbol{\mu}^{(i+1)})\} \\ \boldsymbol{\Phi}_1 &= \boldsymbol{\Phi}_0 + \boldsymbol{\Phi}_\varphi diag(\boldsymbol{\delta}_\varphi^{(i)}) \end{aligned} \tag{38}$$

Hence, the update of $\boldsymbol{\delta}_\theta$ in the $(i+1)$th iteration can be calculated by minimizing the expression in Equation (37).

For $\boldsymbol{\delta}_\varphi$, it is the same as $\boldsymbol{\delta}_\theta$, so we directly give the result:

$$\boldsymbol{\delta}_\varphi^{(i+1)} = \arg\min_{\boldsymbol{\delta}_\varphi} \left\{ \boldsymbol{\delta}_\varphi^T \boldsymbol{A}_2 \boldsymbol{\delta}_\varphi + 2\boldsymbol{a}_2^T \boldsymbol{\delta}_\varphi \right\} \tag{39}$$

where $\boldsymbol{A}_2$, $\boldsymbol{a}_2$ and $\boldsymbol{\Phi}_2$ are similar to those for $\boldsymbol{\delta}_\theta$. Notice that in the maximization step, both $\boldsymbol{\delta}_\theta$ and $\boldsymbol{\delta}_\varphi$ are jointly sparse with $\boldsymbol{\sigma}$, leading to the dimensionality reduction and significant decrease of the computation load. Based on the above analysis, the variational EM-based off-grid imaging algorithm can be described in Table 1. It is worth mentioning that this algorithm does not need the sparsity K as prior information. This property makes it different from traditional super-resolution methods such as MUSIC or ESPRIT. Since the true sparsity of the imaging scene is often not available, this algorithm has vast application prospect.

Table 1. Main steps of the proposed imaging method

Processing Steps of the Off-Grid Imaging Method.
Input: $\boldsymbol{y}, \boldsymbol{\Phi}_0, \boldsymbol{\Phi}_\theta, \boldsymbol{\Phi}_\varphi$
Initialization:
$\sigma_i^{(0)} = \boldsymbol{\Phi}_i^H \boldsymbol{y} / \boldsymbol{\Phi}_i^H \boldsymbol{\Phi}_i$, $\eta^0 = \|\boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{\sigma}^{(0)}\|^2 / MN$, $\boldsymbol{\delta}_\theta = \boldsymbol{\rho}_\theta / 2$, $\boldsymbol{\delta}_\phi = \boldsymbol{\rho}_\varphi / 2$, $a = b = c = d = 10^{-5}$
Iteration: let i denotes the iteration counter
(1) update $\boldsymbol{\sigma}^{(i+1)}$ by Equation (30)
(2) update $\boldsymbol{\alpha}^{(i+1)}$ by Equation (32)
(3) update $\boldsymbol{\beta}^{(i+1)}$ by Equation (34)
(4) update $\eta^{(i+1)}$ by Equation (36)
(5) update $\boldsymbol{\delta}_\theta^{(i+1)}$ and $\boldsymbol{\delta}_\varphi^{(i+1)}$ by Equations (37) and (39)
Termination condition:
The iteration ends when $\|\boldsymbol{\sigma}^{(i+1)} - \boldsymbol{\sigma}^{(i)}\|^2 / \|\boldsymbol{\sigma}^{(i)}\|^2 < 10^{-5}$.
Output:
The imaging result $\boldsymbol{\sigma}^{(i+1)}$, the off-gird bias estimation $\boldsymbol{\delta}_\theta^{(i+1)}$ and $\boldsymbol{\delta}_\varphi^{(i+1)}$

The overall 3D imaging method is shown in Figure 3. First, the channel separation and range compression are conducted. In practical applications, the range resolution provided by the wide bandwidth is sufficient. After we get the echo signal, the range compression technique is used to get the one dimensional range profile first. For range cells where scatterers exist, the algorithm will further solve the two dimensional imaging problem in the elevation and azimuth direction. Thus, the 3D imaging problem is reduced to a series of off-grid imaging problems in the elevation and azimuth direction. By using the proposed method, these problems can be solved and the 3D image is finally reconstructed. The proposed 3D imaging algorithm adopts the uniform prior, which is more suitable than the Gaussian distribution in S-TLS. This algorithm seeks to find the optimal estimation according to the MAP criterion and it can be seen as an extension to the imaging problem with higher dimensionality. Moreover, more degrees of freedom are introduced into the model via a three-stage sparse prior, which will improve the imaging performance.

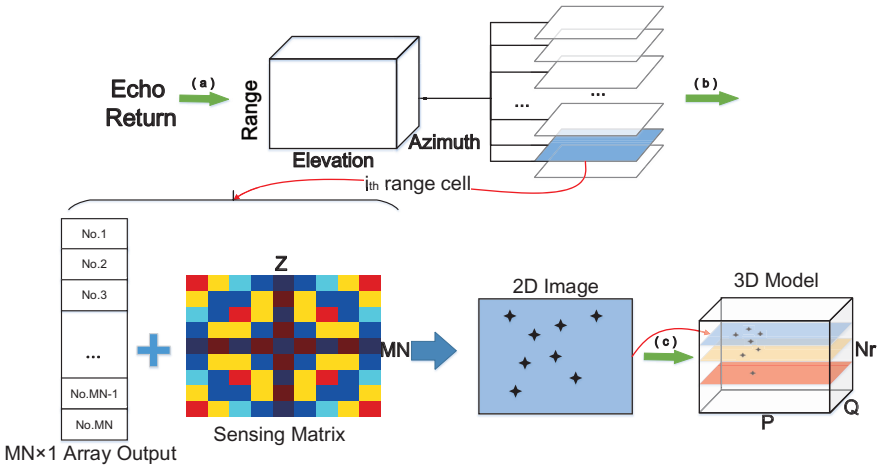

Figure 3. Workflow of the proposed 3D imaging algorithm. (**a**) Range compression and channel separation; (**b**) 2D off-grid imaging in elevation and azimuth direction; (**c**) Formation of the 3D image.

3.2. Bayesian Cramér-Rao Bounds For Off-Grid Biases

It is well known that the Cramér-Rao Lower Bound(CRLB) is an effective indicator for the MSE performance of unbiased estimators. In Prasad and Murthy [25], the author provides an analogous bound called Bayesian Cramér-Rao Bound to provide a lower bounds for estimation problems in sparse Bayesian learning. Different from the traditional CRLB, the prior distribution of the unknowns are considered. The MSEs of δ_θ and δ_φ are compared with their corresponding BCRBs. This estimation algorithm is seen as effective if their MSEs approach the BCRB with the increase of signal-to-noise ratio (SNR).

In this off-grid imaging algorithm, σ, δ_θ and δ_φ are the unknowns to be estimated. We denote a new vector $\boldsymbol{\Theta} = [\sigma; \delta_\theta; \delta_\varphi]$ containing the all the unknowns to be estimated and the MSE matrix is defined as $\boldsymbol{E}^{\boldsymbol{\Theta}} \triangleq E_{\boldsymbol{y},\boldsymbol{\Theta}}[(\boldsymbol{\Theta} - \hat{\boldsymbol{\Theta}}(\boldsymbol{y}))(\boldsymbol{\Theta} - \hat{\boldsymbol{\Theta}}(\boldsymbol{y}))^T]$. The first step to calculate the Cramér-Rao Lower Bound is to derive the Fisher Information Matrix(FIM) $\boldsymbol{I}^{\boldsymbol{\Theta}}$. Usually it is convenient to express $\boldsymbol{I}^{\boldsymbol{\Theta}}$ in terms of submatrices, in which the (i,j)th block is as follows [25]:

$$\boldsymbol{I}_{ij}^{\boldsymbol{\Theta}} \triangleq -E_{\boldsymbol{y},\boldsymbol{\Theta}}\left[\nabla_{\boldsymbol{\Theta}_i} \nabla_{\boldsymbol{\Theta}_i}^T \log p(\boldsymbol{y}, \boldsymbol{\Theta}; \eta, \boldsymbol{\alpha})\right] \tag{40}$$

Thus, the Fisher information matrix can be written as:

$$\boldsymbol{I}^{\boldsymbol{\Theta}} = \begin{pmatrix} \boldsymbol{I}_{11} & \boldsymbol{I}_{12} & \boldsymbol{I}_{13} \\ \boldsymbol{I}_{21} & \boldsymbol{I}_{22} & \boldsymbol{I}_{23} \\ \boldsymbol{I}_{31} & \boldsymbol{I}_{32} & \boldsymbol{I}_{33} \end{pmatrix} \tag{41}$$

Only consider the items that are relevant to $\boldsymbol{\sigma}, \boldsymbol{\delta}_\theta, \boldsymbol{\delta}_\varphi$, we have:

$$-\log p(\boldsymbol{y}, \boldsymbol{\Theta}; \eta, \boldsymbol{\alpha}) = \frac{1}{\eta}\|\boldsymbol{y} - (\boldsymbol{\Phi}_0 + \boldsymbol{\Phi}_\theta \boldsymbol{\Delta}_\theta + \boldsymbol{\Phi}_\varphi \boldsymbol{\Delta}_\varphi)\boldsymbol{\sigma}\|_2^2 + \boldsymbol{\sigma}^H \boldsymbol{\Sigma}^{-1} \boldsymbol{\sigma} \tag{42}$$

Substituting Equation (42) into Equation (40), we can get the following results. For $\boldsymbol{I}_{11}$, noting that only quadratic items of $\boldsymbol{\sigma}$ are useful and the expectations of $\boldsymbol{\delta}_\theta, \boldsymbol{\delta}_\varphi$ are $\mathbf{0}$, we can get the equation:

$$\begin{aligned} \boldsymbol{I}_{11} &= -E_{\boldsymbol{y},\boldsymbol{\Theta}}\left[\nabla_{\boldsymbol{\sigma}} \nabla_{\boldsymbol{\sigma}}^T \log p(\boldsymbol{y}, \boldsymbol{\Theta}; \eta, \boldsymbol{\alpha})\right] \\ &= E_{\boldsymbol{y},\boldsymbol{\Theta}}\left[\frac{2}{\eta}(\boldsymbol{\Phi}_0^H \boldsymbol{\Phi}_0 + \boldsymbol{\Delta}_\theta \boldsymbol{\Phi}_\theta^H \boldsymbol{\Phi}_\theta \boldsymbol{\Delta}_\theta + \boldsymbol{\Delta}_\varphi \boldsymbol{\Phi}_\varphi^H \boldsymbol{\Phi}_\varphi \boldsymbol{\Delta}_\varphi + \eta \boldsymbol{\Sigma}^{-1})\right] \\ &= \frac{2}{\eta}[\boldsymbol{\Phi}_0^H \boldsymbol{\Phi}_0 + \frac{\rho_\theta^2}{12}(\boldsymbol{I} \odot \boldsymbol{\Phi}_\theta^H \boldsymbol{\Phi}_\theta) + \frac{\rho_\varphi^2}{12}(\boldsymbol{I} \odot \boldsymbol{\Phi}_\varphi^H \boldsymbol{\Phi}_\varphi) + \eta \boldsymbol{\Sigma}^{-1}] \end{aligned} \tag{43}$$

As for $\boldsymbol{I}_{12}, \boldsymbol{I}_{13}, \boldsymbol{I}_{21}, \boldsymbol{I}_{31}$, since $\boldsymbol{\sigma}, \boldsymbol{\delta}_\theta, \boldsymbol{\delta}_\varphi$ are independent, it is easy to find that they are all zero matrices. For $\boldsymbol{I}_{22}$ and $\boldsymbol{I}_{33}$, the derivations are similar. Here we take $\boldsymbol{I}_{22}$ as an example. Similar to that of $\boldsymbol{\sigma}$, only quadratic items of $\boldsymbol{\delta}_\theta$ contribute to the final result.

$$\begin{aligned} \boldsymbol{I}_{22} &= E_{\boldsymbol{y},\boldsymbol{\Theta}}\left[\nabla_{\boldsymbol{\delta}_\theta} \nabla_{\boldsymbol{\delta}_\theta}^T \frac{1}{\eta} \boldsymbol{\sigma}^H (\boldsymbol{\Delta}_\theta \boldsymbol{\Phi}_\theta^H \boldsymbol{\Phi}_\theta \boldsymbol{\Delta}_\theta) \boldsymbol{\sigma}\right] \\ &= \frac{1}{\eta} \nabla_{\boldsymbol{\delta}_\theta} \nabla_{\boldsymbol{\delta}_\theta}^T E_{\boldsymbol{y},\boldsymbol{\Theta}}\left[tr(\boldsymbol{\sigma}^H (\boldsymbol{\Delta}_\theta \boldsymbol{\Phi}_\theta^H \boldsymbol{\Phi}_\theta \boldsymbol{\Delta}_\theta) \boldsymbol{\sigma})\right] \end{aligned} \tag{44}$$

By exchanging the expectation and the trace operator, the above equation reduces to $\frac{1}{\eta} \nabla_{\boldsymbol{\delta}_\theta} \nabla_{\boldsymbol{\delta}_\theta}^T tr(\boldsymbol{\Sigma} \boldsymbol{\Delta}_\theta \boldsymbol{\Phi}_\theta^H \boldsymbol{\Phi}_\theta \boldsymbol{\Delta}_\theta)$. Notice that $tr(\boldsymbol{\Sigma} diag(\boldsymbol{\delta}_\theta) \boldsymbol{\Phi}_\theta^H \boldsymbol{\Phi}_\theta diag(\boldsymbol{\delta}_\theta)) = \boldsymbol{\delta}_\theta^T (\Sigma \odot (\boldsymbol{\Phi}_\theta^H \boldsymbol{\Phi}_\theta)^T) \boldsymbol{\delta}_\theta$. Besides, by observing that $\Sigma \odot (\boldsymbol{\Phi}_\theta^H \boldsymbol{\Phi}_\theta)^T$ is positive semi-definite and that $\boldsymbol{\delta}_\theta$ is real-valued, we can get $\boldsymbol{\delta}_\theta^T (\Sigma \odot (\boldsymbol{\Phi}_\theta^H \boldsymbol{\Phi}_\theta)^T) \boldsymbol{\delta}_\theta = \boldsymbol{\delta}_\theta^T \Re(\Sigma \odot (\boldsymbol{\Phi}_\theta^H \boldsymbol{\Phi}_\theta)^T) \boldsymbol{\delta}_\theta$. This leads to the result that $\boldsymbol{I}_{22} = \frac{2}{\eta} \Re(\Sigma \odot \boldsymbol{\Phi}_\theta^H \boldsymbol{\Phi}_\theta)$. The derivation process of $\boldsymbol{I}_{33}$ is the same.

Similarly, we only present the derivation process of $\boldsymbol{I}_{23}$ since that of $\boldsymbol{I}_{32}$ is the same. For $\boldsymbol{I}_{23}$:

$$\begin{aligned} \boldsymbol{I}_{23} &= \frac{1}{\eta} E_{\boldsymbol{y},\boldsymbol{\Theta}}\left[\nabla_{\boldsymbol{\delta}_\theta} \nabla_{\boldsymbol{\delta}_\varphi}^T \boldsymbol{\sigma}^H (\boldsymbol{\Delta}_\theta \boldsymbol{\Phi}_\theta^H \boldsymbol{\Phi}_\varphi \boldsymbol{\Delta}_\varphi + \boldsymbol{\Delta}_\varphi \boldsymbol{\Phi}_\varphi^H \boldsymbol{\Phi}_\theta \boldsymbol{\Delta}_\theta) \boldsymbol{\sigma}\right] \\ &= \frac{1}{\eta} (\boldsymbol{\Phi}_\theta^H \boldsymbol{\Phi}_\varphi + \boldsymbol{\Phi}_\varphi^H \boldsymbol{\Phi}_\theta) \odot \boldsymbol{\Sigma} \\ &= \frac{2}{\eta} \Sigma \odot \Re(\boldsymbol{\Phi}_\theta^H \boldsymbol{\Phi}_\varphi) \end{aligned} \tag{45}$$

The submatrices of the FIM are listed here:

$$\begin{aligned}
&\boldsymbol{I}_{11} = \frac{2}{\eta}[\boldsymbol{\Phi}_0^H\boldsymbol{\Phi}_0 + \frac{\rho_\theta^2}{12}(\boldsymbol{I}\odot\boldsymbol{\Phi}_\theta^H\boldsymbol{\Phi}_\theta) + \frac{\rho_\varphi^2}{12}(\boldsymbol{I}\odot\boldsymbol{\Phi}_\varphi^H\boldsymbol{\Phi}_\varphi) + \eta\boldsymbol{\Sigma}^{-1}]\\
&\boldsymbol{I}_{12}, \boldsymbol{I}_{13}, \boldsymbol{I}_{21}, \boldsymbol{I}_{31} = \boldsymbol{0}\\
&\boldsymbol{I}_{22} = \frac{2}{\eta}\Re(\Sigma\odot\boldsymbol{\Phi}_\theta^H\boldsymbol{\Phi}_\theta)\\
&\boldsymbol{I}_{23} = \boldsymbol{I}_{32} = \frac{2}{\eta}\Sigma\odot\Re(\boldsymbol{\Phi}_\theta^H\boldsymbol{\Phi}_\varphi)\\
&\boldsymbol{I}_{33} = \frac{2}{\eta}\Re(\Sigma\odot\boldsymbol{\Phi}_\varphi^H\boldsymbol{\Phi}_\varphi)
\end{aligned} \tag{46}$$

Then a lower bound on the MSE matrix $\boldsymbol{E}^{\boldsymbol{\Theta}}$ is presented by the inversion of the FIM $\boldsymbol{I}^{\boldsymbol{\Theta}}$:

$$\boldsymbol{E}^{\boldsymbol{\Theta}} - (\boldsymbol{I}^{\boldsymbol{\Theta}})^{-1} \succeq \boldsymbol{0} \tag{47}$$

where $\succeq \boldsymbol{0}$ is interpreted as meaning that the matrix is positive semidefinite. Noting that $\boldsymbol{I}_{22}$, $\boldsymbol{I}_{23}$, $\boldsymbol{I}_{32}$ and $\boldsymbol{I}_{33}$ are diagonal matrices, $(\boldsymbol{I}^{\boldsymbol{\Theta}})^{-1}$ can be expressed as follows:

$$\boldsymbol{J}^{\boldsymbol{\Theta}} = (\boldsymbol{I}^{\boldsymbol{\Theta}})^{-1} = \begin{pmatrix} \boldsymbol{J}_{11} & \boldsymbol{J}_{12} & \boldsymbol{J}_{13} \\ \boldsymbol{J}_{21} & \boldsymbol{J}_{22} & \boldsymbol{J}_{23} \\ \boldsymbol{J}_{31} & \boldsymbol{J}_{32} & \boldsymbol{J}_{33} \end{pmatrix} \tag{48}$$

where $\boldsymbol{J}_{11} = \boldsymbol{I}_{11}^{-1}$ and $\boldsymbol{J}_{12}, \boldsymbol{J}_{13}, \boldsymbol{J}_{21}, \boldsymbol{J}_{31} = \boldsymbol{0}$ according to Equation (46). Since $\boldsymbol{I}_{22}$, $\boldsymbol{I}_{23}$, $\boldsymbol{I}_{32}$ and $\boldsymbol{I}_{33}$ are all diagonal matrices, we find that $\boldsymbol{J}_{22}$, $\boldsymbol{J}_{33}$ and $\boldsymbol{J}_{23}$ are also diagonal matrices and their elements can be easily calculated by solving a set of linear equations. For a positive semidefinite matrix, its diagonal elements are nonnegative. Thus, the BCRBs of the off-grid bias δ_θ and δ_φ are obtained. For the ith element of δ_θ and δ_φ, the expressions are:

$$\begin{aligned}
E(|\hat{\delta}_{\theta_i} - \delta_{\theta_i}|^2) &\geq \frac{\Re[(\boldsymbol{\Phi}_\varphi^H\boldsymbol{\Phi}_\varphi)_{i,i}]}{2\{\Re[(\boldsymbol{\Phi}_\varphi^H\boldsymbol{\Phi}_\varphi)_{i,i}]\Re[(\boldsymbol{\Phi}_\theta^H\boldsymbol{\Phi}_\theta)_{i,i}] - \Re^2[(\boldsymbol{\Phi}_\theta^H\boldsymbol{\Phi}_\varphi)_{i,i}]\}\mathrm{SNR}}\\
E(|\hat{\delta}_{\varphi_i} - \delta_{\varphi_i}|^2) &\geq \frac{\Re[(\boldsymbol{\Phi}_\theta^H\boldsymbol{\Phi}_\theta)_{i,i}]}{2\{\Re[(\boldsymbol{\Phi}_\varphi^H\boldsymbol{\Phi}_\varphi)_{i,i}]\Re[(\boldsymbol{\Phi}_\theta^H\boldsymbol{\Phi}_\theta)_{i,i}] - \Re^2[(\boldsymbol{\Phi}_\theta^H\boldsymbol{\Phi}_\varphi)_{i,i}]\}\mathrm{SNR}}
\end{aligned} \tag{49}$$

where the SNR is defined as $\frac{\alpha_i}{\eta}$.

4. Experimental Results

In this section, we present some simulation results to analyze performance of the proposed algorithm. In Section 4.1, a few sparse recovery algorithms are applied to the 3D imaging problem and the results are compared. The 3D imaging performance, i.e., the normalized mean square error(NMSE) of target image recovery and off-grid bias estimation are presented. Besides, the imaging results of a complex model is provided to further validate the feasibility of proposed method. Then super-resolution ability is analyzed with respect to SNR in Section 4.2. Finally the off-grid bias estimations are compared with corresponding BCRBs in Section 4.3. The simulation parameters in the sparse array 3D imaging system are listed in Table 2.

Table 2. Simulation Conditions

Parameter	Symbol	Value
Bandwidth	B	500 MHz
Carrier frequency	f_c	10 GHz
Baseline in the X-direction	L_x	6 m
Baseline in the Y-direction	L_y	6 m
Target range	R	2500 m
Pulse width	T_r	10 μs
Number of transmitters	M	4
Number of receivers	N	225

4.1. Validation of The Proposed Algorithm

A simulated target with 24 scatterers is considered in this part. Since the range resolution is guaranteed through range compression, for simplicity, the scatterers of the target are located at three different range cells. The original 3D model is shown in Figure 4a, and both on-grid and off-grid scatterers are included. The sparse antenna array is distributed on a circular area with its diameter being 6 m and center being the origin. Four transceivers are located at the intersections of the boundary and the axes. The other receivers are first uniformly distributed on this area, then random disturbances are added to their positions in order to alleviate the ambiguity problem which emerges in sparse array. The center of target is located at $(2500\text{ m}, \frac{\pi}{4}, \frac{\pi}{4})$.

3D imaging results using different sparse recovery algorithms with SNR equals 20 dB are presented in Figure 4, with blue circle and red circle representing the true scatterer and the reconstructed scatterer, respectively. The size of the circle is decided by the scatterer's RCS. As discussed before, the classic sparse recovery algorithms including OMP and BP fail to reconstruct the 3D imaging because of the mismatch. Compared with the classic algorithms, S-TLS takes the off-grid bias into consideration and gets a better imaging performance. However, there exist some false scatterers and the off-grid bias estimator does not perform well. This can be explained by the Gaussian prior used for the off-grid biases while their true distribution being uniform. Different from S-TLS, OGSBI adopts a uniform prior and its imaging performance is better. Yet there still exists mismatch between the scatterers' true position and reconstructed position. In contrast, the imaging performance of the proposed algorithm is the best among all methods because of the application of the three-stage sparse prior.

The NMSEs using these different algorithms are summarized in Table 3. NMSE is defined as follows and I is the number of Monte Carlo trails:

$$NMSE(\sigma) = \frac{1}{I}\sum \frac{||\hat{\sigma} - \sigma||^2}{||\sigma||^2} \quad NMSE(\delta_{\theta,\varphi}) = \frac{1}{I}\sum \frac{||\hat{\delta}_{\theta,\varphi} - \delta_{\theta,\varphi}||^2}{||\delta_{\theta,\varphi}||^2} \tag{50}$$

Table 3. NMSEs of 3D image and off-grid bias estimation.

Algorithm	OMP	BP	S-TLS	OGSBI	Proposed Method
NMSE of σ	1.3371	1.1129	0.6921	0.5397	0.3016
NMSE of δ_θ	*	*	0.6048	0.2452	0.1333
NMSE of δ_φ	*	*	0.6321	0.2283	0.1266

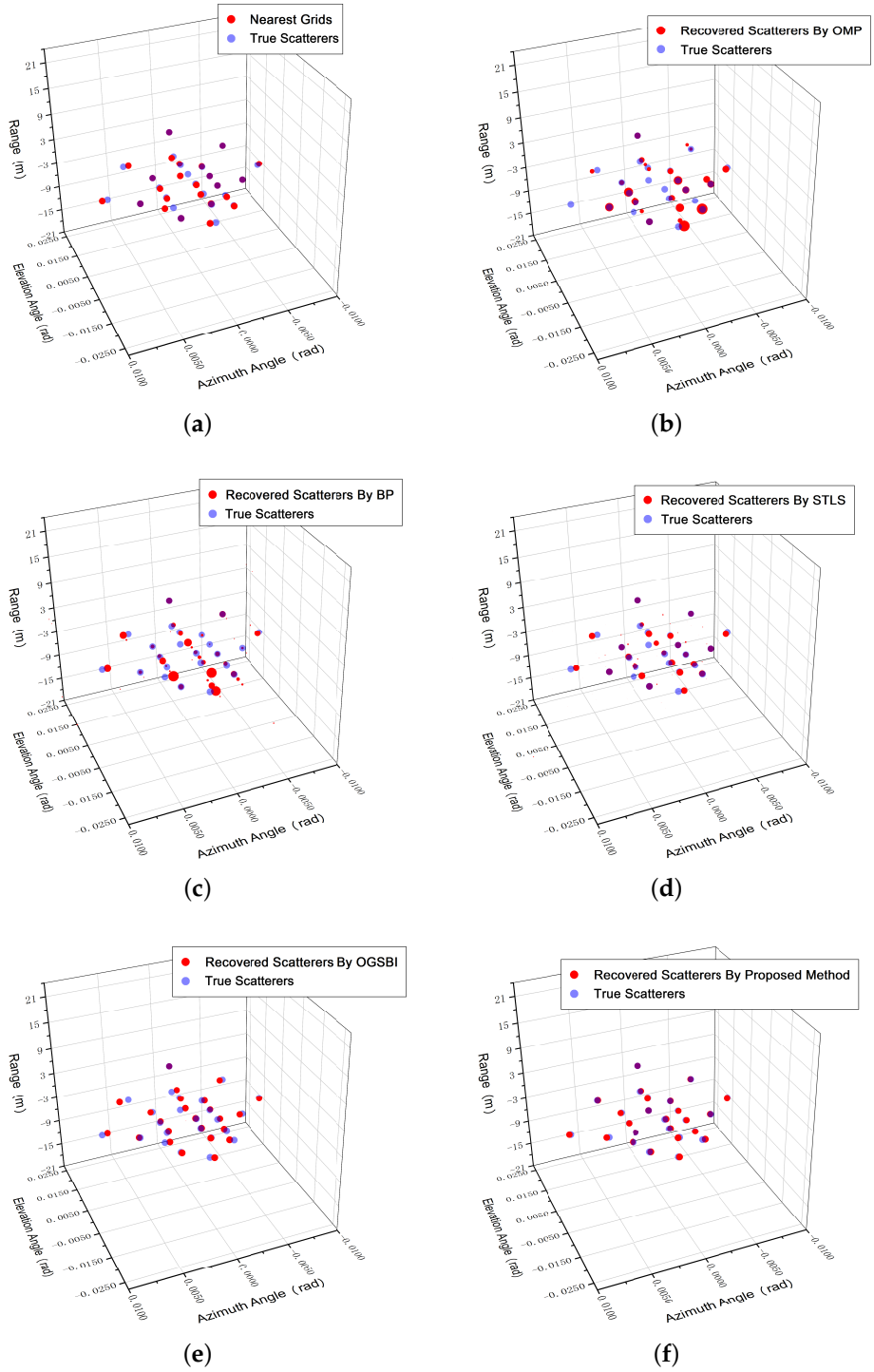

Figure 4. The original target model and reconstructed ones by different algorithms: (**a**) The original target model and the corresponding nearest grids; (**b**) Reconstructed target scatterers by OMP; (**c**) Reconstructed target scatterers by BP; (**d**) Reconstructed target scatterers by S-TLS; (**e**) Reconstructed target scatterers by OGSBI; (**f**) Reconstructed target scatterers by proposed method.

In Table 3, the asterisks mean that these algorithms cannot estimate the off-grid biases. From this table, it is clear that the proposed method obtains the best performance among all the algorithms. However, we also find that the NMSEs of 3D image is larger than those of off-grid bias estimation, which implies that current algorithms fail to estimate the RCS as accurate as position. It is also shown

that the differences between θ and φ are small. So in the following part we only present the NMSE of δ_θ for brevity.

NMSEs of sparse recovery with different SNRs and grid intervals are presented. Figure 5a,b shows the NMSEs under different SNR levels. The grid interval is set to half the Rayleigh resolution. The target model is the same and the SNR increases from 0dB to 45 dB with an interval of 5 dB. Number of the Monte-Carlo trails is set to 100. Since the imaging procedure can be seen as a joint estimation problem, the correlation between θ and φ in Equation (17) makes the performances more sensitive to noise. As we can see, the NMSEs stay at a high level for all algorithms when the SNR is low. However, the image recovery error of both OMP and BP are always high even if the SNR increases. The reason is that these two algorithms fail when the mismatch problem emerges. The performance of S-TLS is better than the former two algorithms. However, it is not satisfying because of the mismatch between Gaussian distribution and the true model. OGSBI can achieve a relatively good performance. The proposed algorithm can get the best performance among all the algorithms when the SNR is larger than 20 dB because the three-stage hierarchical model introduces more degrees of freedom and is a better approximation of the l_0-norm optimization.

In our study, we find that the imaging performances vary with different grid sizes. So we analyze the performances versus different grid intervals in Figure 5c,d. The simulation parameters are all the same but the grid interval varies. $\chi = \frac{\rho}{R_{Rayleigh}}$ is the ratio of grid interval to the Rayleigh limit. It is shown in Figure 5 that the NMSEs of all algorithms decrease first and then increase. The reason for this phenomenon is that denser grids can alleviate the mismatch problem to some extent, however, much too dense grids will lead to a false recovery result due to violation of RIP. It is interesting that the NMSEs of OMP and BP get the minimum when χ equals 0.5 while for the rest 0.4. This is caused by the fact that these algorithms incorporate the mismatch factor into sparse recovery model. So the off-grid biases can be estimated and compensated, leading to a better performance when small grid interval is used. However, the NMSEs increase with χ increasing because the approximation error caused by the first order Taylor expansion also increases.

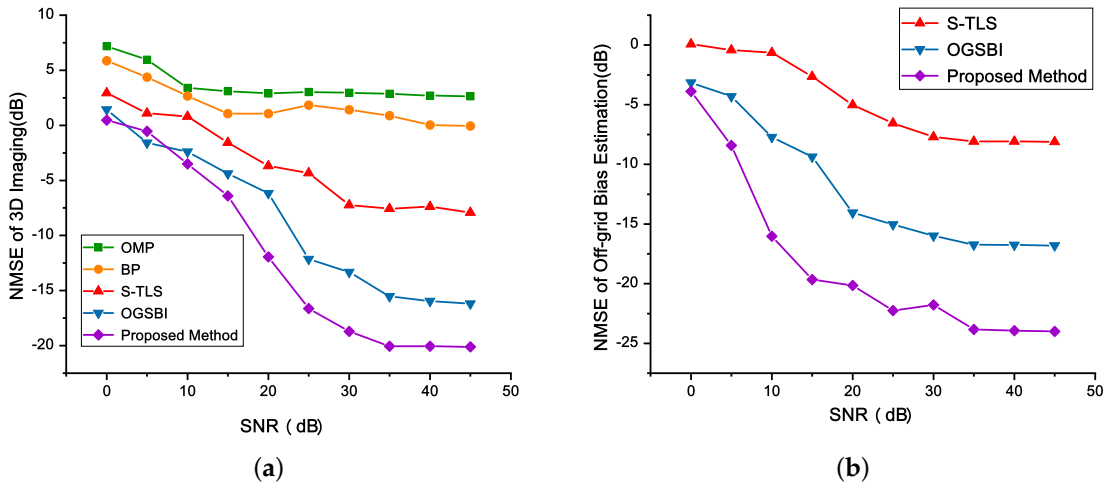

Figure 5. *Cont.*

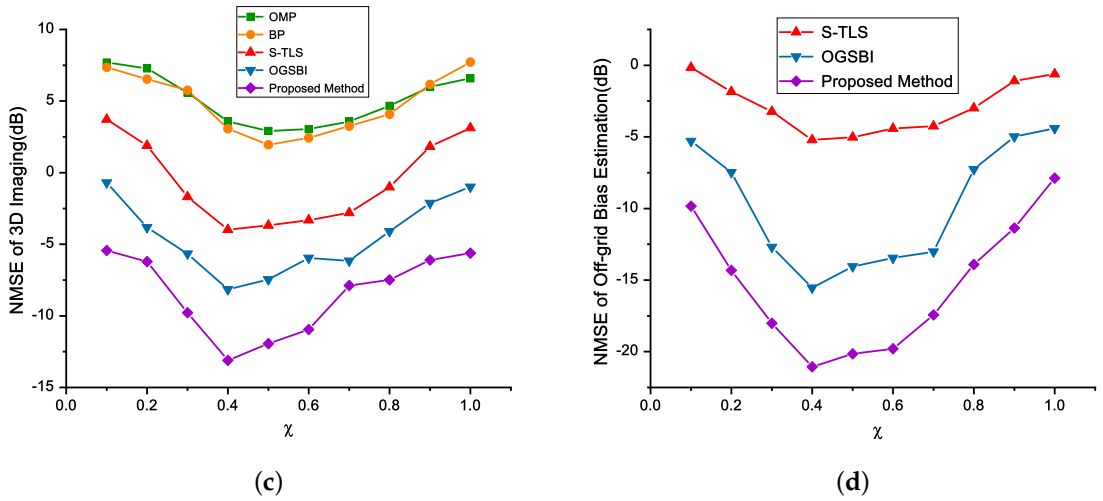

Figure 5. NMSEs versus different SNRs and grid invervals: (**a**) NMSEs of 3D reconstruction versus different SNRs; (**b**) NMSEs of off-grid bias estimations versus different SNRs; (**c**) NMSEs of 3D reconstruction versus different grid intervals; (**d**) NMSEs of off-grid bias estimations versus different grid intervals.

To further validate the practicality of proposed method for complex target 3D reconstruction, the imaging results based on a complex plane model is presented. This plane model is based on the RCS reconstruction result of a real aircraft at the airport. We use the model to generate echo signal and then use the proposed method to reconstruct the 3D plane model. In the future, we would further validate the proposed method using real-world radar data. In this simulation, the target center is located at $(2400, \frac{\pi}{4}, \frac{\pi}{4})$ and the system configuration is the same as before. The reconstructed 3D image of the plane is presented in Figure 6. Figure 6a is an overall view of the 3D image while Figure 6b–d are the three views of the reconstructed target model. The size of the target can be calculated using these figures. Furthermore, features of the plane, such as the nose, engines, vertical and horizontal stabilizer, can be recognized from the reconstructed 3D image as is shown in Figure 6a,c. These results show its potential for target identification.

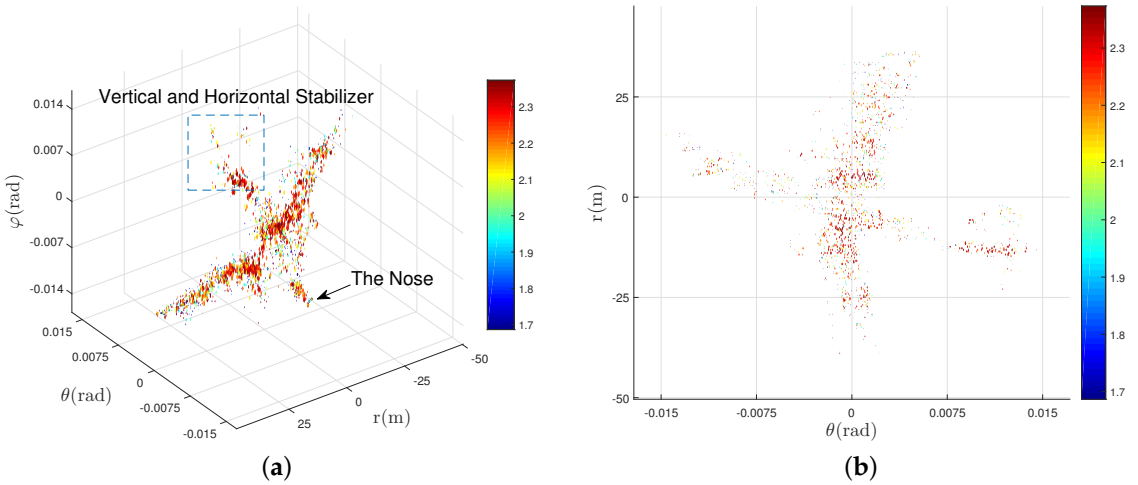

Figure 6. *Cont.*

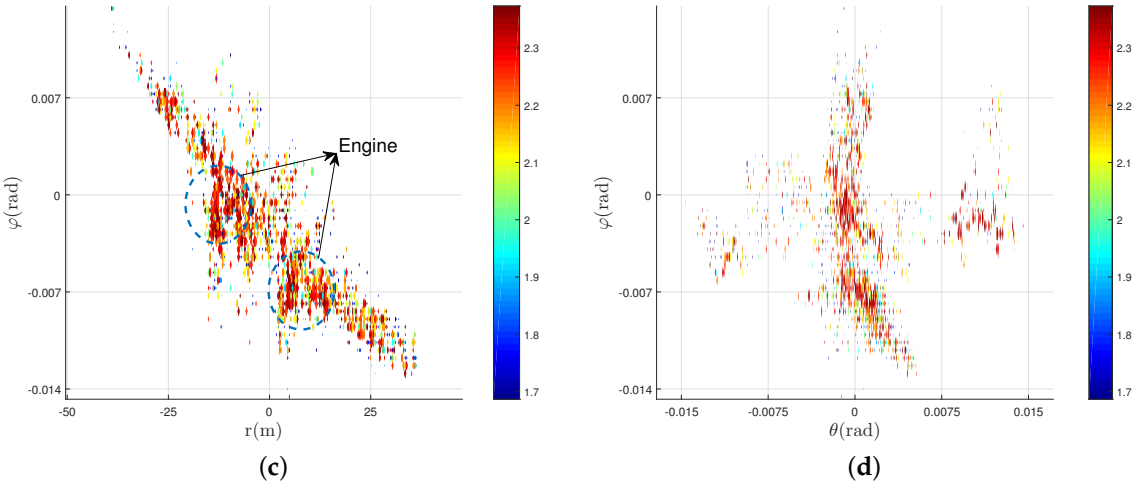

Figure 6. Reconstructed 3D image of the plane by proposed method. (**a**) Overall view of reconstructed result; (**b**–**d**) 2D projection of the reconstructed 3D image.

4.2. Super-Resolution Performance Versus SNR

When it comes to imaging algorithm, one key problem is its resolution. Since the range resolution is obtained through pulse compression, we consider the resolutions of the other two dimensions. In this part, the grid interval is set to the best in Section 4.1 and only two scatterers are considered. One experiment is defined as successful if the two scatterers are separated. Then the success rate is calculated through 100 Monte Carlo simulations.

Since the resolutions in elevation and azimuth directions are similar, only the success rate for θ are presented for brevity. In Figure 7, $\omega = \frac{d}{R_{Rayleigh}}$ is the normalized distance between the two scatterers. Simulations are conducted with different SNRs and scatterer distances. This plot gives us some perspectives about the problem. First, the super-resolution ability of proposed method is better than that of OGSBI. The reason is that the off-grid bias estimation is more accurate, making the separation of two closely spaced off-grid scatterers possible. Second, with the increasing of SNR, super-resolution ability of the proposed algorithm is increasing. Here we define the resolution as the distance between two scatterers when the success rate is more than 50%. The proposed algorithm can realize super-resolution with its resolutions of elevation being a quarter of the Rayleigh limit when the SNR is 20 dB. Considering the shape of scatter plot, we adopt the Boltzmann fit. The red line and black line are fitted curves for proposed method and OGSBI. The model of Boltzmann fit is:

$$y = A_2 + \frac{A_1 - A_2}{1 + exp(\frac{x - x_0}{dx})} \tag{51}$$

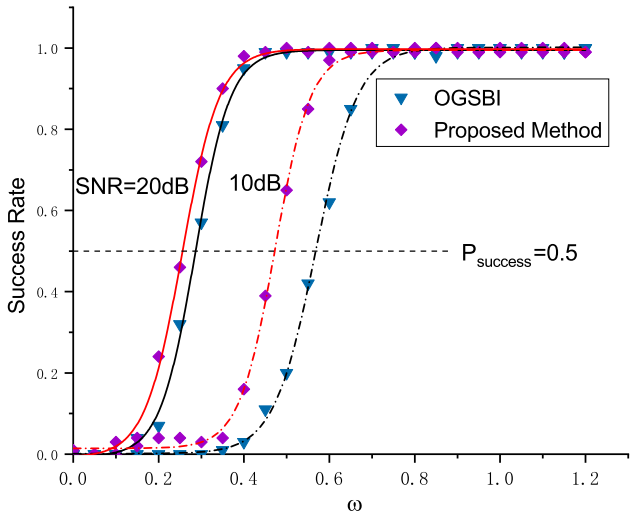

Figure 7. Success rate of separation using different algorithms and under different SNRs.

The fitting parameters for the two algorithms under different SNRs are listed in Table 4. According to Equation (51), both A_1 and A_2 are the normalization factors and A_2 can be regarded as the max value of the curve. dx is a constant which influences the sharpness of the curve. It is insensitive to the SNR level according to Table 3. The success rate is around 0.5 when x equals to x_0, which can be seen as the reciprocal of the super-resolution factor. From Table 3 we can see it is inversely proportion to the SNR.

Table 4. Fitting Parameters

	A_1	A_2	x_0	dx
OGSBI, SNR = 10 dB	0.0064	0.9932	0.5682	0.0490
OGSBI, SNR = 20 dB	−0.0064	0.9949	0.2857	0.0412
Proposed method, SNR = 10 dB	0.0014	0.9948	0.4726	0.0418
Proposed method, SNR = 20 dB	−0.0080	0.9975	0.2553	0.0435

4.3. BCRB for Off-Grid Biases

In this subsection, the Bayesian Cramér-Rao Bound for off-grid bias is presented, then the root mean square error(RMSE) of the proposed algorithm is compared with it. The expression of RMSE is:

$$RMSE(\delta_\theta) = \sqrt{\frac{1}{I}\sum_{i=1}^{I}(\hat{\delta}_\theta - \delta_\theta)^2} \tag{52}$$

where I is the number of Monte Carlo trails. The BCRB for either δ_θ or δ_φ is a function of two variables, i.e., θ and φ. Since $\theta \in [0, \frac{\pi}{2}]$ and $\varphi \in [0, \pi]$, the overall BCRB for them is shown in Figure 8 with SNR = 20 dB. From these two figures we can find that the two BCRBs are insensitive to φ. However, for θ, the corresponding BCRB increases with θ increasing. When $\theta = \frac{\pi}{2}$, the BCRB is approaching infinity, which means that the imaging performance is bad when θ tends to $\frac{\pi}{2}$. As for BCRB of δ_φ, it approaches infinity when θ tends to zero. The reason is that the information of φ is lost since $\sin\theta \to 0$ according to Equation (12).

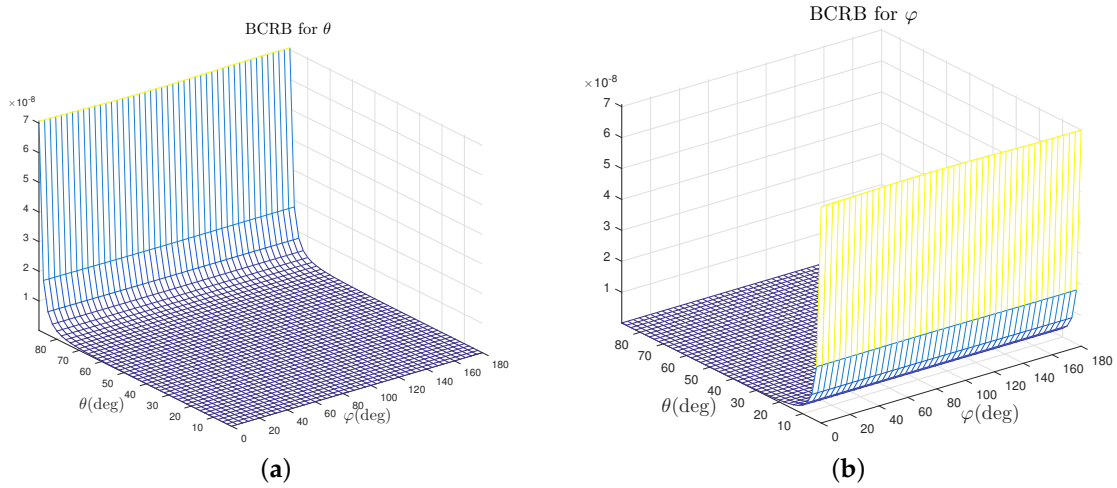

Figure 8. Bayesian Cramér-Rao Bound (BCRB) for off-grid bias estimation: (**a**) BCRB for θ. (**b**) BCRB for φ.

To verify the theory in Section 3.2, the RMSE of proposed algorithm and square root of the BCRB are compared in this part. Only one point scatterer located at $(2500\text{ m}, \frac{\pi}{4}, \frac{\pi}{4})$ is considered. The SNR increases from 0 dB to 40 dB with a stepsize of 5dB. The BCRB is calculated according to Equation (49). The RMSE is calculated using 100 Monte Carlo trails. The results are presented in Figure 9. The green stars in the plots mean that the sparse recovery failed because of the noise. With the increase of SNR, the estimation accuracy increases and the RMSE of proposed algorithm is approaching the BCRB. Thus, the proposed off-grid bias estimator is proven to be effective.

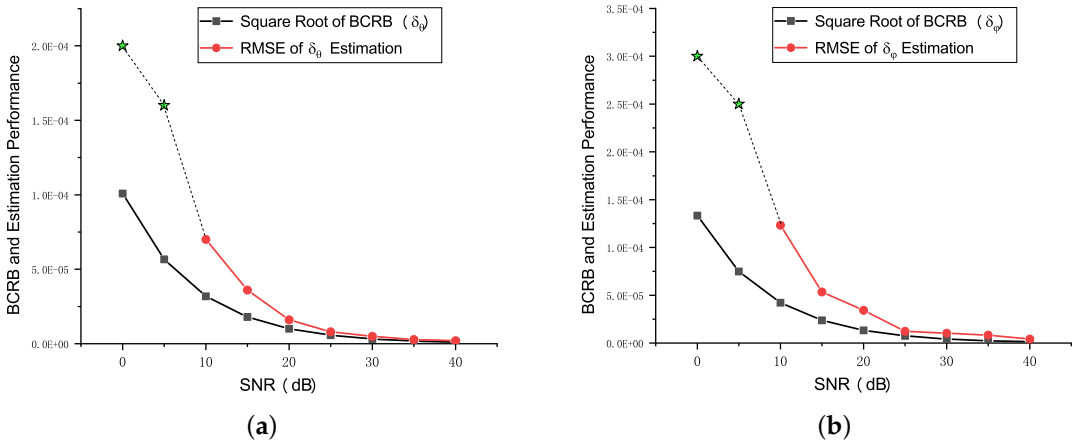

Figure 9. Comparison between square root of BCRB and the RMSE of off-grid bias estimation: (**a**) Off-grid bias for θ; (**b**) Off-grid bias for φ.

5. Conclusions

In order to solve the problems of off-grid target 3D imaging, a novel sparse Bayesian learning-based algorithm using sparse antenna array is proposed in this paper. The main impact of the off-grid problem is that it will lead to energy leakage, which will finally spoil the reconstruction results. In consideration of the characteristics of the off-grid target, a sparse Bayesian learning-based imaging algorithm is proposed to estimate not only the RCS but also the off-grid biases simultaneously. A three-stage hierarchical sparse prior is introduced and the BCRB for off-grid bias is presented,

providing a lower bound for methods based on the MAP criterion. Quantitative analyses are provided to compare the 3D imaging performances of the proposed method with other state of art methods. The results show that the imaging performance of proposed method is better than other popular methods,i.e., a higher reconstruction precision and a better resolution. The effectiveness of the algorithm is verified by comparing the RMSE with its corresponding BCRB. These results all directly or indirectly validate the feasibility of proposed method. The proposed method can reconstruct the 3D model accurately and different components of the target can be recognized by the result. These findings contribute in several ways to our understanding of off-grid air target 3D reconstruction and identification, and also provide a reference for further applications of aircraft 3D imaging based on sparse antenna array.

Author Contributions: Zekun Jiao, Chibiao Ding, Xingdong Liang, Longyong Chen and Fubo Zhang initiated the research; Chibiao Ding proposed the main idea and gave guidance on the work; Under the supervision of Fubo Zhang, Zekun Jiao performed the experiments and wrote the manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Wang, G.; Xia, X.G.; Chen, V.C. Three-dimensional ISAR imaging of maneuvering targets using three receivers. *IEEE Trans. Image Process.* **2001**, *10*, 436–447.
2. Duan, G.Q.; Wang, D.W.; Ma, X.Y.; Su, Y. Three-dimensional imaging via wideband MIMO radar system. *IEEE Geosci. Remote Sens. Lett.* **2010**, *7*, 445–449.
3. Ma, C.; Yeo, T.S.; Tan, C.S.; Li, J.Y.; Shang, Y. Three-dimensional imaging using colocated MIMO radar and ISAR technique. *IEEE Trans. Geosci. Remote Sens.* **2012**, *50*, 3189–3201.
4. Ding, S.; Tong, N.; Zhang, Y.; Hu, X. Super-resolution 3D imaging in MIMO radar using spectrum estimation theory. *IET Radar Sonar Navig.* **2016**, *11*, 304–312.
5. Ma, C.; Yeo, T.S.; Tan, H.S.; Wang, J.; Chen, B. Three-dimensional ISAR imaging using a two-dimensional sparse antenna array. *IEEE Geosci. Remote Sens. Lett.* **2008**, *5*, 378–382.
6. Ma, C.; Yeo, T.S.; Tan, C.S.; Liu, Z. Three-dimensional imaging of targets using colocated MIMO radar. *IEEE Trans. Geosci. Remote Sens.* **2011**, *49*, 3009–3021.
7. Gu, F.; Chi, L.; Zhang, Q.; Zhu, F. Single snapshot imaging method in multiple-input multiple-output radar with sparse antenna array. *IET Radar Sonar Navig.* **2013**, *7*, 535–543.
8. Donoho, D.L. Compressed sensing. *IEEE Trans. Inf. Theory* **2006**, *52*, 1289–1306.
9. Zhang, L.; Xing, M.; Qiu, C.W.; Li, J.; Sheng, J.; Li, Y.; Bao, Z. Resolution enhancement for inversed synthetic aperture radar imaging under low SNR via improved compressive sensing. *IEEE Trans. Geosci. Remote Sens.* **2010**, *48*, 3824–3838.
10. Liu, H.; Jiu, B.; Liu, H.; Bao, Z. Superresolution ISAR imaging based on sparse Bayesian learning. *IEEE Trans. Geosci. Remote Sens.* **2014**, *52*, 5005–5013.
11. Xu, G.; Xing, M.; Zhang, L.; Liu, Y.; Li, Y. Bayesian inverse synthetic aperture radar imaging. *IEEE Geosci. Remote Sens. Lett.* **2011**, *8*, 1150–1154.
12. Hu, X.; Tong, N.; Zhang, Y.; Hu, G.; He, X. Multiple-input–multiple-output radar super-resolution three-dimensional imaging based on a dimension-reduction compressive sensing. *IET Radar Sonar Navig.* **2016**, *10*, 757–764.
13. Fannjiang, A.; Tseng, H.C. Compressive radar with off-grid targets: A perturbation approach. *Inverse Probl.* **2013**, *29*, doi:10.1088/0266-5611/29/5/054008.
14. Yang, Z.; Xie, L. On gridless sparse methods for line spectral estimation from complete and incomplete data. *IEEE Trans. Signal Process.* **2015**, *63*, 3139–3153.
15. Boche, H.; Calderbank, R.; Kutyniok, G.; Vybiral, J. *Compressed Sensing and Its Applications*; Springer: Berlin, Germany, 2015; pp. 75–93
16. He, X.; Liu, C.; Liu, B.; Wang, D. Sparse frequency diverse MIMO radar imaging for off-grid target based on adaptive iterative MAP. *Remote Sens.* **2013**, *5*, 631–647.
17. Chi, Y.; Scharf, L.L.; Pezeshki, A.; Calderbank, A.R. Sensitivity to basis mismatch in compressed sensing. *IEEE Trans. Signal Process.* **2011**, *59*, 2182–2195.

18. Zhu, H.; Leus, G.; Giannakis, G.B. Sparsity-cognizant total least-squares for perturbed compressive sampling. *IEEE Trans. Signal Process.* **2011**, *59*, 2002–2016.
19. Yang, Z.; Xie, L.; Zhang, C. Off-grid direction of arrival estimation using sparse Bayesian inference. *IEEE Trans. Signal Process.* **2013**, *61*, 38–43.
20. Das, A. Theoretical and Experimental Comparison of Off-Grid Sparse Bayesian Direction-of-Arrival Estimation Algorithms. *IEEE Access.* **2017**, *5*, 18075–18087.
21. Bishop, C.M. *Pattern Recognition and Machine Learning*; Springer: Berlin, Germany, 2006.
22. Blei, D.M.; Kucukelbir, A.; McAuliffe, J.D. Variational inference: A review for statisticians. *J. Am. Stat. Assoc.* **2017**, *112*, doi:10.1080/01621459.2017.1285773.
23. Tzikas, D.G.; Likas, A.C.; Galatsanos, N.P. The variational approximation for Bayesian inference. *IEEE Signal Process. Mag.* **2008**, *25*, 131–146.
24. Lemonte, A.J.; Cordeiro, G.M. The exponentiated generalized inverse Gaussian distribution. *Stat. Probab. Lett.* **2011**, *81*, 506–517.
25. Prasad, R.; Murthy, C.R. Cramér-Rao-type bounds for sparse Bayesian learning. *IEEE Trans. Signal Process.* **2013**, *61*, 622–632.