MIMO Radar Imaging With Multiple Probing Pulses for 2D Off-Grid Targets via Variational Sparse Bayesian Learning

CHAO WEN[ID][1], **LU CHEN**[ID][1], **PENGTING DUAN**[ID][2], **AND XUEFENG CUI**[ID][2]
[1]Institute of Big Data Science and Industry, Shanxi University, Taiyuan 030006, China
[2]North Automatic Control Technology Research Institute, Taiyuan 030006, China

Corresponding author: Chao Wen (cwen@sxu.edu.cn)

ABSTRACT Spatial sparsity of the target space has been successfully exploited to provide accurate range-angle images by the methods based on sparse signal reconstruction (SSR) in Multiple-Input Multiple-Output (MIMO) radar imaging applications. The SSR based method discretizes the continuous target space into finite grid points and generates an observation model utilized in image reconstruction. However, inaccuracies in the observation model may cause various degradations and spurious peaks in the reconstructed images. In the process of the image formation, the off-grid problem frequently occurs that the true locations of targets that do not coincide with the computation grid. In this article, we consider the case that the true location of a target has both range and angle-varying two-dimensional (2D) off-grid errors with a noninformative prior. From a variational Bayesian perspective, an iterative algorithm is developed for joint MIMO radar imaging with orthogonal frequency division multiplexing (OFDM) linear frequency modulated (LFM) waveforms and 2D off-grid error estimation of off-grid targets. The targets during multiple probing pulses are modeled as Swerling II case and a unified generalized inverse Gaussian (GIG) prior is adopted for the target reflection coefficient variance at all snapshots. Furthermore, an approach to reducing the computational workload of the signal recovery process is proposed by using singular value decomposition. Experimental results show that the proposed algorithm is insensitive to noise and has improved accuracy in terms of mean squared estimation error under different computation grid interval.

INDEX TERMS Sparse Bayesian learning, multiple-input multiple-output (MIMO) radar imaging, multiple probing pulses, orthogonal frequency division multiplexing (OFDM), two-dimensional (2D) off-grid error.

I. INTRODUCTION

Multiple-Input Multiple-Output (MIMO) radar is an emerging field of radar research which has attracted intensive researches. MIMO radars with waveform diversity offer unique advantages over conventional phased-array counterparts, such as extra degrees of freedom, higher resolution [1], [2] and better parameter identifiability [3]. Besides, MIMO radar which employs transmit Linear frequency modulated (LFM) waveforms can achieve a better range resolution [4]. Different conventional phased-array, MIMO radar adopts orthogonal waveforms transmitted from different antenna elements, which avoids transmitting waveform interference

The associate editor coordinating the review of this manuscript and approving it for publication was Zaharias D. Zaharis[ID].

from each other and enables each received signal separated by using a bank of matched filters [5]. Recently, an orthogonal frequency division multiplexing (OFDM) LFM waveform has been applied to MIMO radar for high-resolution remote sensing [6]. The OFDM scheme is introduced in [7], which is one of the ways to accomplish simultaneous use of several carriers. With each transmitter of the MIMO radar operating at one of the OFDM carrier frequencies, OFDM-LFM MIMO radar employing OFDM LFM waveforms can generate range-angle images of targets. This article aims at colocated OFDM-LFM MIMO radar imaging.

A. REVIEW OF EXISTING LITERATURE AND MOTIVATION

The goal of the MIMO radar imaging is to provide an estimate of the radar cross sections (RCS) of targets at precise

IEEE *Access*

C. Wen *et al.*: MIMO Radar Imaging With Multiple Probing Pulses for 2D Off-Grid Targets via Variational Sparse Bayesian Learning

angular locations and distances relative to the radar. One of the most important problems in MIMO radar imaging is to obtain the two-dimensional (2D) range-angle images. In most radar imaging applications, the number of targets in the region of interest (ROI) is sufficiently less than the number of potential source locations [8]. The sparsity of the target scene has been successfully exploited to provide accurate range-angle images in MIMO radar imaging applications, owing to the development of methods based on sparse signal reconstruction (SSR) or compressive sensing (CS) [9]. The problem of synthetic aperture radar (SAR) image degradation at low signal-to-noise ratio (SNR) has been addressed in [10] where spatial sparsity has been utilized. Sparsity in transform domain has also been exploited for image reconstruction in [11] where smart sensing technique is proposed for online and automatic evaluation of carbon fiber reinforced polymer structure integrity. In these methods, e.g., [8], [10]–[12], the angle or the range domain which is continuous space in practice has to be discretized into a finite set, or equivalently, the angles or the ranges of interest are constrained on a fixed grid. The SSR problem can be formulated, where target locations of interest constitute the support of the sparse signal to be reconstructed. By assuming that all true target locations are exactly on the selected grid, an observation model is established for image reconstruction by the SSR based method.

Under the assumption that the reflection coefficients of the targets are constant varying from pulse to pulse, which is also referred to as the single-measurement-vector (SMV) case, ℓ_1 optimization is a favourable approach in MIMO radar imaging [13] due to its guaranteed recovery accuracy [9]. Further, ℓ_q optimization based on maximum a posteriori (MAP) approach is proposed in [8] with the small values of q promoting sparser solutions. However, these methods are based on the conventional CS model which adopts the fixed dictionary. Recently, the models combined with parametric dictionary are proposed in [10], [14]–[17]. They have achieved improvements in SAR imaging [10], SAR autofocusing [14], sparse array imaging [15], moving target imaging [16] and time-frequency analysis [17], by optimizing both the off-grid parameters and the sparse solution. Motived by this idea, in this work, we focus on the off-grid problem of MIMO radar imaging during multiple probing pulses, with the reflection coefficients of the targets varying independently from pulse to pulse where Swerling II model is assumed [18]. It should be noted that the off-grid problem is in multiple-measurement-vector (MMV) case. The MMV case is often encountered in practical applications, such as source localization and direction-of-arrival estimation (DOA) [19]. In the MMV case, the sparse signals at all snapshots share the same support [20]. Such joint sparsity has been exploited to improve the probability of successful recovery [21]. Furthermore, sparse Bayesian learning (SBL) algorithm [22], [23] is another popular method for the SSR problems, which induces less structural error and convergence error [24]

and flexibly models the possible structures of signals in order to obtain improved performance [25], [26].

The MIMO radar imaging can be regarded as the estimation of the target amplitude response on a two-dimensional (2D) grid, with one dimension being the angular bin and the other dimension being the range bin. It is worth noting, however, that the SSR algorithms aforementioned involve division of the 2D range-angle space of the potential tagets into a mass of grids, and they are based on the on-grid assumption that all targets should fall right on the predefined computation grid so that reliable estimations can be guaranteed [12]. Unfortunately, such on-grid scenario can rarely be met in reality, where there is a gap between the true location and its nearest computation grid which the estimated range-angle points are constrained on. The employment of a coarse grid set often leads to unsatisfactory estimation accuracy because of the off-grid error. Although accurate solution can be achieved by denser grid search following the coarse grid search, this would bring about much more computation burden. Furthermore, too dense grid set may result in unacceptable degradation of computational efficiency as well as highly coherent matrix that violates the condition for the SSR. Another solution is to establish an off-grid model and the improved estimation performance can be expected. With the basis mismatch of the measurement matrix in the observation model caused by the off-grid error taken into account, an off-grid model with Gaussian prior on off-grid error has been successfully applied to DOA estimation [27]. In fact, the noninformative off-grid error is more reasonable. The uniformly distributed off-grid angluar error has been well addressed in [28]. Furthermore, we focus on the off-grid target imaging with the 2D noninformative range-angle error in MMV case from variational Bayesian perspective.

B. CONTRIBUTIONS OF PRESENT WORK

A variational SBL algorithm and an acceleration scheme are proposed for joint OFDM-LFM MIMO radar imaging and 2D off-grid error estimation of targets. The off-grid errors are considered as model errors that are estimated and compensated during image formation, which enables our technique to alleviate degradations and spurious peaks in the reconstructed image. At Bayesian formulation stage, a hierarchical model is developed with a unified generalized inverse Gaussian (GIG) prior adopted for all snapshots. The GIG is referred to as a family of distributions including Gaussian, Laplacian and many other heavy-tailed distributions, which provides more flexibility to model the prior information for target. This is distinguished from simply assigning Laplacian prior for coefficient variance [29], [30]. At Bayesian learning stage, a variational Bayesian technique is incorporated in hyperparameters inference, where the sparsity regularizing parameter is learned rather than fixed. Since the reconstruction process with measurement matrix based on the 2D imaging model is usually a large computational burden. An acceleration scheme is also incorporated in our algorithm by using

C. Wen *et al.*: MIMO Radar Imaging With Multiple Probing Pulses for 2D Off-Grid Targets via Variational Sparse Bayesian Learning

IEEE *Access*

sequential singular value decomposition (SVD) to reduce the rows and the columns of measurement matrix rather than only the columns [29]. The signal and the noise subspace induced by SVD can be exploited to improve the image quality at low SNR [10].

The proposed algorithm is named as 2D off-grid sparse Bayesian learning with two SVDs (2DOGSBL-2SVDs). We will compare our method with the widely used on-grid MMV model based approaches, such as ℓ_1 optimization based algorithms [12], [31] and ℓ_q optimization based algorithms [32], [33], and the similar work of off-grid model based method OGSBI [29] which has been extended to 2D parameter estimation [30]. Note that it is difficult for these typical methods to solve the problem to achieve good performance, so they are combined with the proposed acceleration scheme in order to more fairly compete with the proposed method. It is shown that the proposed method is more accurate in terms of mean squared error (MSE) under different computation grid interval as well as more insensitive to Gaussian noise.

The remaining sections are organized as follows. Section 2 presents the OFDM-LFM MIMO radar signal model. Next, off-grid OFDM-LFM MIMO sparse imaging model is studied in Section 3. Section 4 details the proposed 2DOGSBL-2SVDs algorithm with 2SVDs strategy. Our simulation results are shown in Section 5. Finally, the conclusions on this article are provided in Section 6.

Notation: Bold-face letters are reserved for vectors and matrices. $(\cdot)^*$, $(\cdot)^T$ and $(\cdot)^H$ denote conjugate, transpose and conjugate transpose operations, respectively. $\|\cdot\|_2$ and $\|\cdot\|_F$ denote the ℓ_2 norm and Frobenious norm, respectively. $\boldsymbol{I}_N$ represents a $N \times N$ identity matrix. $\boldsymbol{A} \in \boldsymbol{R}^{M \times N}$ and $\boldsymbol{R} \in \mathbb{C}^{M \times N}$ denote a real and complex-valued $M \times N$ matrix $\boldsymbol{A}$, respectively. $diag(\boldsymbol{A})$ denotes a column vector composed of the diagonal elements of a matrix $\boldsymbol{A}$, and $diag(\boldsymbol{x})$ is a diagonal matrix with $\boldsymbol{x}$ being its diagonal elements. x_i denotes the ith entry of $\boldsymbol{x}$. $(\boldsymbol{A})_i$, $(\boldsymbol{A})^j$ and $A_{i,j}$ are the ith column, jth row and (i, j)th entry of a matrix $\boldsymbol{A}$. $\Re(\cdot)$ and $\Im(\cdot)$ take the real and imaginary parts of a complex variable respectively.

II. OFDM-LFM MIMO RADAR SIGNAL MODEL
A. BASIC ASSUMPTION
Consider the scenario of the targets in the far field and a narrowband MIMO radar system composed by N transmitters and M receivers. Due to the attractive ambiguity characteristics, such as range resolution, adjacent band interference and matched filtering sidelobe performance [6], OFDM LFM waveforms are utilized to the MIMO radar. In fact, the OFDM LFM waveform employed here can be regarded as a kind of frequency stepped LFM waveform.

Assuming each transmitter transmit LFM signal with the equal chirp rate γ_c and operates at one of the OFDM carrier frequency f_n. Frequency intervals between the arbitrary two adjacent transmitters Δf are assumed to be same. The transmitted LFM waveform for the n-th transmitter can be

expressed as

$$s_n(t) = \begin{cases} e^{-j2\pi\left(f_n t + \frac{1}{2}\gamma_c t^2\right)}, & -\frac{T_P}{2} \leq t \leq \frac{T_P}{2} \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

where $f_n = f_c + (n-1)\Delta f$, T_P denotes the pulse width and f_c is the carrier frequency of the signal on the first transmit channel. It can be easily shown that when $\Delta f = \kappa_0/T_s$ with $\kappa_0 \in \mathbb{N}^\dagger$ holds, the requirement of orthogonal multitransmission is satisfied with

$$\int_{-T_P/2}^{T_P/2} s_n(t)s_m^*(t)dt = \begin{cases} constant, & n = m \\ 0, & n \neq m \end{cases} \quad (2)$$

For simplicity's sake, a uniform linear array is applied to both at the transmitter and receiver. The steering vectors can be written as

$$\boldsymbol{a}(\theta) = \begin{bmatrix} 1 & e^{-j2\pi d_t \sin\theta/\lambda} & \cdots & e^{-j2\pi(N-1)d_t \sin\theta/\lambda} \end{bmatrix}^T \quad (3)$$

and

$$\boldsymbol{b}(\theta) = \begin{bmatrix} 1 & e^{-j2\pi d_r \sin\theta/\lambda} & \cdots & e^{-j2\pi(M-1)d_r \sin\theta/\lambda} \end{bmatrix}^T \quad (4)$$

where d_t and d_r denote the inter-element spacing of the transmit and receive arrays, respectively, λ is the wavelength. In this article, the targets in our scenario are modeled according to Swerling II case [18], and are assumed to have no translation relative to the radar so that a target can be modeled as a peak in final range-angle image.. However, the model allows for target spinning relative to the radar, with the RCS fluctuations of targets fixing during a pulse and varying independently pulse to pulse. The received signal at the m-th receive element can be expressed as

$$D(m, \tau, t) = \sigma(\tau)b_m(\theta)\sum_{n=1}^{N} a_n(\theta)s_n\left(t - \frac{2R(\tau)}{c}\right) \quad (5)$$

where c is the light velocity, $b_m(\theta)$ denotes the nth element of $\boldsymbol{b}(\theta)$, $a_n(\theta)$ denotes the nth element of $\boldsymbol{a}(\theta)$, $R(\tau)$ denotes the instantaneous distance from the scattering center to the radar, $\sigma(\tau)$ represents the complex reflection coefficients of the targets which fix during a pulse and vary independently pulse to pulse, t represents the fast time, $t_s = (\tau - 1)T_s$ represents the slow time, T_s denotes the pulse repetition interval, $\tau = 1, \ldots, T$ denotes pulse index and T denotes the number of pulses.

B. MATCH FILTERING
For the τ th pulse, $D(m, \tau, t)$ can be filtered out as the mth signal by the matched filter set [6] with multiplying by a phase factor $e^{j2\pi f_n t_s}$. The matched filtering of each receive channel with respect to the (n, m)th channel can be represented as

$$D_{MF}(n, m, t, \tau) = \sigma(\tau)b_m(\theta)a_n(\theta)\frac{\sin[\pi B(t - 2R(\tau)/c)]}{\pi B(t - 2R(\tau)/c)}$$

$$\times exp\{j2\pi f_n[t_s - 2R(\tau)/c]\}$$

IEEE *Access*

C. Wen *et al.*: MIMO Radar Imaging With Multiple Probing Pulses for 2D Off-Grid Targets via Variational Sparse Bayesian Learning

$$= \sigma(\tau)b_m(\theta)\, a_n(\theta)\frac{\sin[\pi B(t - 2R/c)]}{\pi B(t - 2R/c)}$$

$$\times exp\{j2\pi f_n[(\tau - 1)T_s - 2R/c]\} \quad (6)$$

where $B = \gamma_c T_P$, $t_s = (\tau - 1)T_s$ represents the slow time and the received signal $D_{MF}(n, m, t, \tau)$ corresponding to each target has the same time delay $2R/c$ with the instantaneous distance fixing from pulse to pulse, for the assumption that target has no translation relative to the radar during T pulses.

We consider P range bins and K angular bins relative to the radar, so that the grids $\{R_p\}_{p=1}^{P}$ and $\{\theta_k\}_{k=1}^{K}$ divide the ROI. Let $\sigma_{p,k}(\tau)$ represent the complex reflection coefficients for the targets in the ROI. Assuming the pth range cell contains K scattering centers with K angle locations, the received signal in the pth range cell corresponding to $t = 2R_p/c$ during the τth pulse can be derived from (6) as follows:

$$D_{MF}(p, n, m, \tau) = \sum_{k=1}^{K} \sigma_{p,k}(\tau)b_m(\theta_k)a_n(\theta_k)$$

$$\times exp\{j2\pi f_n[(\tau - 1)T_s - 2R_p/c]\} \quad (7)$$

by inserting $B = 0$. Thus, (7) can be rewritten as

$$D_{out}(p, n, m, \tau)$$

$$= \sum_{k=1}^{K} \sigma_{p,k}(\tau)e^{j2\pi\left[f_n((\tau-1)T_s - \frac{2R_p}{c}) - \frac{(md_r + nd_t)\sin\theta_k}{\lambda}\right]} \quad (8)$$

C. ON-GRID MODEL FOR IMAGING
With respect to (n, m)th channel, the received signal from the ROI can be formed by a superposition of the received signal from the (p, k)th position.

$$y(\tau)_{n,m,\tau}$$

$$= \sum_{k=1}^{K}\sum_{p=1}^{P} \sigma_{p,k}(\tau)e^{j2\pi\left[f_n\left((\tau-1)T_s - \frac{2R_p}{c}\right) - \frac{(md_r + nd_t)\sin\theta_k}{\lambda}\right]} \quad (9)$$

The received signal can be represented in vector form

$$y(\tau) = \big[y(\tau)_{1,1,1} \cdots y(\tau)_{N,1,1}\, y(\tau)_{1,2,1}$$

$$\cdots y(\tau)_{N-1,M,T}\, y(\tau)_{N,M,T}\big]^T \quad (10)$$

which leads to the on-grid observation model

$$y(\tau) = G\sigma(\tau) + e(\tau) \quad (11)$$

where $e(\tau)$ denotes the measurement noise vector, $\sigma(\tau) = [\sigma_{1,1}(\tau), \cdots, \sigma_{K,1}(\tau), \sigma_{1,2}(\tau), \cdots, \sigma_{K,P}(\tau)]^T$ is a vector representing the sampled and column-stacked version of the range-angle image during the τth pulse, $\tau = 1, \cdots, T$ and $G \in \mathbb{C}^{NMT \times PK}$ is the true observation matrix composed of the element in (9), which can be expressed as

$$g(R_p, \theta_k)_{n,m,\tau}$$

$$= exp\{j2\pi[f_n((\tau - 1)T_s - 2R_p/c) - (md_r + nd_t)\sin\theta_k]\} \quad (12)$$

which is the $((\tau - 1)NM + (m - 1)N + n, (p - 1)K + k)$th element of G. It is evident that the observation element can

map any target location in terms of the range-angle plane to the measurement, and the target is continously distributed at the 2D space. As a result, the goal of the radar imaging is to find $\sigma_{p,k}(\tau)$ given $y(\tau)$, where the key point is the accurate modeling of the observation matrix. For many radar imaging applications like air surveillance, the target of ROI is spatially sparse.

The on-grid assumption based SSR algorithms like ℓ_1-SVD and M-FOCUSS has successfully produce ange-angle image by exploiting the sparsity in $\sigma(\tau)$. However, these algorithms obtain the accurate results only if the targets exactly fall on the fixed computation grid. Any grid mismatch leads to ambiguities in the estimation due to the leakage of energy over all the grid cells which depends on the kernel used for the construction of the observation matrix. To overcome this problem, we develop the off-grid model for MIMO radar imaging and show its relationship with the on-grid one.
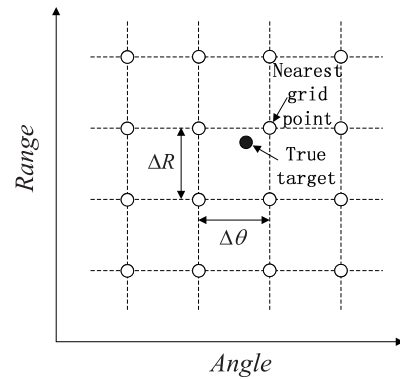

FIGURE 1. Off-grid target in range-angle plane.

III. OFF-GRID OFDM-LFM MIMO SPARSE IMAGING MODEL
The on-grid model adopts the discretization of the continous range-angle plane of potential targets into PK computation grids. P and K denote the number of points in the range and the angle dimension, respectively. Let $(\tilde{R}, \tilde{\theta}) = \{(\tilde{R}_1, \tilde{\theta}_1), \ldots, (\tilde{R}_P, \tilde{\theta}_K)\}$ be the set of the fixed computation grid points in range-angle plane. Without loss of generality, let $\tilde{R}$ and $\tilde{\theta}$ be the uniform grid with the grid interval of range cell and angle cell being ΔR and $\Delta\theta$, respectively, as is shown in Fig.1. By contrast, off-grid assumption is adopted in this section that the true targets are distributed in a continuous manner and off the fixed computation grid points. Suppose the lth location of the true target $(R_l, \theta_l) \notin \{(\tilde{R}_1, \tilde{\theta}_1), \ldots, (\tilde{R}_P, \tilde{\theta}_K)\}$, $l \in \{1, \ldots, L\}$ with the lth corresponding nearest grid $(\tilde{R}_{p_l}, \tilde{\theta}_{k_l})$, $(p_l, k_l) \in \{(1, 1), \ldots, (P, K)\}$. Note that the grid point number should be far greater than the true target number L, which is a kind of sparse description of ROI. However, the off-grid error of the true target leads to the fact that $y(\tau)$ is not truly sparse with respect to the observation matrix. To alleviate this problem, the observation element $g(R_p, \theta_k)_{n,m,\tau}$ can be approximated

C. Wen *et al.*: MIMO Radar Imaging With Multiple Probing Pulses for 2D Off-Grid Targets via Variational Sparse Bayesian Learning

IEEE *Access*

via first-order partial derivative:

$$g(R_l, \theta_l)_{n,m,\tau}$$

$$\approx g(\tilde{R}_{p_l}, \tilde{\theta}_{k_l})_{n,m,\tau} + \frac{\partial g(\tilde{R}_{p_l}, \tilde{\theta}_{k_l})_{n,m,\tau}}{\partial R_{p_l}}(R_l - \tilde{R}_{p_l}) \quad (13)$$

where $\frac{\partial g(\tilde{R}_{p_l}, \tilde{\theta}_{k_l})_{n,m,\tau}}{\partial R_{p_l}} = -j4\pi f_n/c(R_l - \tilde{R}_{p_l})$ and $\frac{\partial g(\tilde{R}_{p_l}, \tilde{\theta}_{k_l})_{n,m,\tau}}{\partial \theta_{k_l}} = -j2\pi(md_r + nd_t)\cos(\tilde{\theta}_{k_l})/\lambda(\theta_l - \tilde{\theta}_{k_l})$ denote the two fixed kernels, and the true off-grid location is modeled with the fixed grid (the first term) and the 2D off-grid distance (the last two terms). It is evident that, the true off-grid location in terms of the observation during the τth pulse is regarded as a two-variable function.

Substitute (13) into (11), the off-grid observation model can be obtained as

$$y(\tau) = G(\alpha, \beta)\sigma(\tau) + e(\tau), \quad \tau = 1, \ldots, T. \quad (14)$$

where

$$G(\alpha, \beta) = G_0 + B\Delta_\alpha + C\Delta_\beta$$
$$= G[I_{PK} \; diag(\alpha) \; diag(\beta)]^T \quad (15)$$

where $G = [G_0 \; B \; C]$, B and $C \in \mathbb{C}^{NMT \times PK}$ are the matrices with the fixed kernels, and $G(\alpha, \beta) \in \mathbb{C}^{NMT \times PK}$ is parameterized by the off-grid range dependent vector

$$\alpha = [\underbrace{\alpha_1, \cdots, \alpha_1}_{K}, \alpha_2, \cdots, \alpha_P]^T \in [-\Delta R/2, \Delta R/2]^{PK} \quad (16)$$

and the off-grid anlge dependent vector

$$\beta = [\underbrace{\beta_1, \cdots, \beta_1}_{K}, \beta_2, \cdots, \beta_P]^T \in [-\Delta\theta/2, \Delta\theta/2]^{PK} \quad (17)$$

For $(p, k) \in \{(1, 1), \ldots, (P, K)\}$, we have

$$(\alpha_p, \beta_k) = \begin{cases} (R_l - \tilde{R}_{p_l}, \theta_l - \tilde{\theta}_{k_l}), & (p, k) = (p_l, k_l) \\ (0, 0), & otherwise \end{cases} \quad (18)$$

which ensures the joint sparsity between the $\sigma(\tau)$ and the two off-grid distance dependent vectors.

It should be noted that the off-grid observation model is closely related to the on-grid one. By setting $(\alpha, \beta) = (0, 0)$ and $G(0, 0) = G_0$, the on-grid observation model can be obtained. The on-grid observation model only exploits the observation matrix with the fixed kernel, which suffers from the quantization error caused by the 2D discretization of potential space. By contrast, the off-grid model incorporates the 2D parameterized observation matrix with first-order approximation of the true kernel. It is observed that off-grid model has a smaller modeling error and higher accuracy than the on-grid one under the same computation grid. Note that higher order approximation of the kernel can be adopted in the off-grid model to achieve less modeling error, but this would lead to heavier computational burden and is out of the scope of this research.

In order to form the image of MIMO radar, we need to obtain not only the the support of coefficient vector, but also the 2D off-grid error (α, β). Aimed at the joint estimation of them, an iterative algorithm is developed from a Bayesian perspective in the following section.

IV. 2DOGSBL AND 2DOGSBL-2SVDs

Since the complex valued reflection coefficients varies from pulse to pluse, our variational SBL algorithm is derived in the MMV case. For multiple snapshots, the off-grid OFDM-LFM MIMO imaging model is given as follows:

$$Y = G(\alpha, \beta)\Omega + E \quad (19)$$

where $Y = [y(1), \ldots, y(T)] \in \mathbb{C}^{NMT \times T}$ denotes the spatial-time measurement, $\Omega = [\sigma(1), \ldots, \sigma(T)] \in \mathbb{C}^{PK \times T}$ represents the complex valued coefficient matrix and $E = [e(1), \ldots, e(T)] \in \mathbb{C}^{NMT \times T}$ stands for noise matrix.

A. SPARSE BAYESIAN FORMULATION

Some sparse prior has been frequently utilized for SSR based parameter estimation. The variational SBL method is introduced and developed to obtain the uncertainty information during estimation, where the signal is hierarchically modeled to impose a prior that promotes sparsity as well as learning. Instead of using too dense grid, we adopt a set of coarse grids to reduce the cost of computation, which is more practical.

1) GAMMA PRIOR FOR NOISE MODELING

From pulse to pulse, multiple reflection coefficient vectors $\sigma(\tau)$, $\tau = 1, \ldots, T$ and noise vectors $e(\tau)$, $\tau = 1, \ldots, T$, are assumed to be independently and identically complex Gaussian distributed, i.e.,

$$p(Y|\Omega, \gamma_0, \alpha, \beta) = \prod_{\tau=1}^{T} \mathcal{CN}\left(y(\tau)|G(\alpha, \beta)\sigma(\tau), \gamma_0^{-1}I\right) \quad (20)$$

where γ_0^{-1} denotes the noise variance with γ_0 being the noise precision, I is an identity matrix with a proper dimension and $\mathcal{CN}(\mu, \Sigma)$ denotes the complex Gaussian probability density function (PDF) of a random variable with mean μ and covariance Σ [34].

Since γ_0 is unknown, and we embed this unknown parameter in the hierarchical framework. For Gamma prior is a conjugate prior of the Gaussian distribution, the perturbation precision is modeled as a Gamma distribution

$$p(\gamma_0; c, d) = \mathcal{G}(\gamma_0; c, d), \quad (21)$$

where $\mathcal{G}(\gamma_0; c, d) = [\Gamma(c)]^{-1}d^c\gamma_0^{c-1}\exp(-d\gamma_0)$ and $\Gamma(\cdot)$ denotes the Gamma function.

2) UNIFIED GIG FOR REFLECTION COEFFICIENT MODELING

Each column of Ω is jointly sparse. In this case, the sparse reflection coefficient vectors $\sigma(\tau)$ share the same support. The spatial distribution of the nonzero coefficients is substantially sparse, so the hierarchical prior is adopted which favors

IEEE *Access*

C. Wen *et al.*: MIMO Radar Imaging With Multiple Probing Pulses for 2D Off-Grid Targets via Variational Sparse Bayesian Learning

most elements of being zeroes. First, we append a complex Gaussian prior to $\boldsymbol{\Omega}$:

$$p\left(\boldsymbol{\Omega}|\boldsymbol{\gamma}\right) = \prod_{\tau=1}^{T} \mathcal{CN}\left(\boldsymbol{\sigma}\left(\tau\right)|\mathbf{0}, \boldsymbol{\Lambda}\right) \quad (22)$$

where $\boldsymbol{\Lambda} = diag\left(\boldsymbol{\gamma}\right)$ and $\boldsymbol{\gamma} = [\gamma_1, \ldots, \gamma_{PK}]^T \in \boldsymbol{R}^{PK}$ with γ_η being the variance of $\sigma_\eta\left(\tau\right)$, $\eta = 1, 2, \ldots, PK$. Depending on the choice of the prior distribution for the variance parameters in $\boldsymbol{\gamma}$ at the second level of hierarchy, various sparsity promoting distributions may arise for $\boldsymbol{\sigma}\left(\tau\right)$. To unify these distributions in a single model and take the nonnegative value of $\boldsymbol{\gamma}$ into consideration, the variance parameters in $\boldsymbol{\gamma}$ are assumed to follow a GIG distribution [35]

$$p\left(\gamma_\eta|\rho, \kappa, \lambda\right) = \frac{(\rho/\kappa)^{\lambda/2}}{2\mathcal{K}_\lambda\left(\sqrt{\rho\kappa}\right)} \gamma_k^{\lambda-1} \exp\left[-\frac{1}{2}\left(\rho\gamma_k + \kappa\gamma_k^{-1}\right)\right] \quad (23)$$

where $\mathcal{K}_\lambda(.)$ is the modified Bessel function of the second kind, the hyperparameters λ, κ and ρ can be selected so as to formulate the widely used heavy-tailed priors. According to (23), the marginal distribution of $\boldsymbol{\sigma}\left(\tau\right)$ can be obtained by integrating out the latent variables as the generalized hyperbolic distribution:

$$p\left(\sigma_\eta(t)|\rho, \kappa, \lambda\right)$$
$$= \frac{\sqrt{\rho}}{2\pi} \frac{(\kappa)^{-\lambda/2}}{\mathcal{K}_\lambda\left(\sqrt{\rho\kappa}\right)} \frac{\mathcal{K}_{\lambda-1}\left(\sqrt{\rho\left(\kappa + 2|\sigma_\eta(t)|^2\right)}\right)}{\left(\kappa + 2|\sigma_\eta(t)|^2\right)^{1/2-\lambda/2}}. \quad (24)$$

And the resulting marginal distribution, the generalized hyperbolic distribution, covers a large number of distributions as special cases. Heavy-tailed prior distributions have been exploited for sparsity inducing, due to the fact that their heavy tails favor a few large coefficients and severely attenuate the substantial noise. Instead of fixing the sparsity regularizing parameter ρ, we infer it from the data and assume it to follow a Gamma distribution with parameters a and b at the third level, i.e.

$$p(\rho; a, b) = \mathcal{G}(\rho; a, b) \quad (25)$$

which provides more adaptiveness regarding ρ during the learning process. Thus, the proposed modeling constitutes a the three-level hierarchical prior that is equivalent to assigning an almost Jeffrey's prior distribution over $\boldsymbol{\gamma}$. The first two levels of this hierarchical prior (22) and (23), and the last level (25) is embedded to calculate ρ. Note that the hyperparameters a, b, c and d at the highest level are set to be trivial values to induce non-informative prior on ρ and γ, respectively, which means their posterior depends only on the data and not the prior knowledge. In this article, $a = b = c = d = 10^{-6}$. Although various other heavy-tailed distributions can be generated by the unified GIG function, the related deep study is not covered in this work.

3) NONINFORMATIVE PRIOR FOR 2D OFF-GRID ERROR MODELING

Since there is usually no additional a priori knowledge of $(\boldsymbol{\alpha}, \boldsymbol{\beta})$ with the two independent variables $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$. We use its boundedness as the prior information. A 2D uniform distribution is defined on the 2D off-grid error

$$(\boldsymbol{\alpha}, \boldsymbol{\beta}) \sim \{(U[-\Delta R/2, \Delta R/2], U[-\Delta/2, \Delta\theta/2])\}^{PK} \quad (26)$$

By combining the stages of the hierarchical model, the posterior distribution is given

$$(\boldsymbol{\Omega}, \gamma_0, \boldsymbol{\gamma}, \boldsymbol{\alpha}, \boldsymbol{\beta}|\boldsymbol{Y})$$
$$= \frac{p(\boldsymbol{Y}|\boldsymbol{\Omega}, \gamma_0, \boldsymbol{\alpha}, \boldsymbol{\beta})p(\boldsymbol{\Omega}|\boldsymbol{\gamma})p(\boldsymbol{\gamma}|\rho)p(\rho)p(\gamma_0)p(\boldsymbol{\alpha})p(\boldsymbol{\beta})}{p(\boldsymbol{Y})} \quad (27)$$

with $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ being independent of each other. The problem in (27) is often intractable to calculate explicitly, since a multidimensional integral is required. In most cases, it is either impossible or very difficult to compute the integral in closed form. Instead of solving it directly, a deterministic approximation method is utilized through variational Bayeisan approximation.

B. BAYESIAN LEARNING

According to the variational Bayeisan technique [36], the posterior in (28) is calculated with $\boldsymbol{\Omega}$, $\boldsymbol{\gamma}$, ρ and γ_0 treated as latent variables, and $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ being the parameters. The approximated posterior is factorizable under the mean-field assumption. Furthermore, by minimizing the KL divergence between the true posterior and the approximated one, the approximated posterior w.r.t. the individual variable can be updated via alternate operation. The piteratively updating rocedures of the individual variables and the parameter as follows.

1) UPDATING FOR THE MOMENTS OF COEFFICIENT

The approximated posterior w.r.t. $\boldsymbol{\sigma}\left(\tau\right)$ can be given by

$$q(\boldsymbol{\sigma}(\tau)) \propto e^{\left\langle \ln \mathcal{CN}\left(\boldsymbol{y}(\tau)|\boldsymbol{G}(\boldsymbol{\alpha}, \boldsymbol{\beta})\boldsymbol{\sigma}(\tau), \gamma_0^{-1}\boldsymbol{I}\right)\mathcal{CN}(\boldsymbol{\sigma}(\tau)|\mathbf{0}, \boldsymbol{\Lambda})\right\rangle} \quad (28)$$

where $\langle.\rangle$ denotes the expectation operation. The approximated posterior distribution of $\boldsymbol{\sigma}\left(\tau\right)$ is a complex Gaussian distribution:

$$q(\boldsymbol{\sigma}(\tau)) \propto \exp\left\{-(\boldsymbol{y}(\tau) - \boldsymbol{\mu}(\tau))^H \boldsymbol{\Sigma}^{-1}(\boldsymbol{y}(\tau) - \boldsymbol{\mu}(\tau))\right\} \quad (29)$$

where

$$\boldsymbol{\Sigma} = \left(\gamma_0 \boldsymbol{G}^H\left(\boldsymbol{\alpha}, \boldsymbol{\beta}\right)\boldsymbol{G}\left(\boldsymbol{\alpha}, \boldsymbol{\beta}\right) + \boldsymbol{\Lambda}^{-1}\right)^{-1} \quad (30)$$

and

$$\boldsymbol{\mu}(\tau) = \gamma_0 \boldsymbol{\Sigma} \boldsymbol{G}^H\left(\boldsymbol{\alpha}, \boldsymbol{\beta}\right)\boldsymbol{y}\left(\tau\right), \tau = 1, \ldots, T \quad (31)$$

where $\boldsymbol{\Sigma} = \boldsymbol{\Lambda} - \boldsymbol{\Lambda}\boldsymbol{G}^H\left(\boldsymbol{\alpha}, \boldsymbol{\beta}\right)\boldsymbol{\Theta}^{-1}\boldsymbol{G}\left(\boldsymbol{\alpha}, \boldsymbol{\beta}\right)\boldsymbol{\Lambda}$ and $\boldsymbol{\Theta} = \gamma_0^{-1}\boldsymbol{I} + \boldsymbol{G}\left(\boldsymbol{\alpha}, \boldsymbol{\beta}\right)\boldsymbol{\Lambda}\boldsymbol{G}^H\left(\boldsymbol{\alpha}, \boldsymbol{\beta}\right)$, according to the Woodbury matrix identity. The updating $\boldsymbol{\Sigma}$ requires a $PK \times PK$ matrix inversion.

C. Wen *et al.*: MIMO Radar Imaging With Multiple Probing Pulses for 2D Off-Grid Targets via Variational Sparse Bayesian Learning

IEEE *Access*

2) UPDATING FOR COEFFICIENT VARIANCE

The approximated posterior w.r.t. $\boldsymbol{\gamma}$ is given by

$$q(\boldsymbol{\gamma}) \propto \exp\left\{\langle \ln p(\boldsymbol{\Omega}|\boldsymbol{\gamma})\,p(\boldsymbol{\gamma}|\rho)\rangle\right\}. \tag{32}$$

Substituting (22) and (23) into (32), the approximated posterior for each γ_η obeys a GIG distribution [19] and the hth moment of the GIG distribution is obtained as

$$\left\langle (\gamma_\eta)^h \right\rangle$$

$$= \left(\frac{\langle 2\|\underline{\boldsymbol{\Omega}}^\eta\|_2^2 + \kappa\rangle}{\langle \rho \rangle}\right)^{h/2} \frac{\mathcal{K}_{\lambda-1+h}\left(\sqrt{\langle \rho \rangle \langle 2\|\underline{\boldsymbol{\Omega}}^\eta\|_2^2 + \kappa\rangle}\right)}{\mathcal{K}_{\lambda-1}\left(\sqrt{\langle \rho \rangle \langle 2\|\underline{\boldsymbol{\Omega}}^\eta\|_2^2 + \kappa\rangle}\right)}, \tag{33}$$

where $\langle \|\underline{\boldsymbol{\Omega}}^\eta\|_2^2 \rangle = \|\underline{\mathcal{U}}^\eta\|_2^2 + \Sigma_{\eta,\eta}$, $\underline{\mathcal{U}} = \mathcal{U}/\sqrt{T}$ with $\underline{\mathcal{U}}^\eta$ is the ηth row of $\mathcal{U} = [\boldsymbol{\mu}(1), \ldots, \boldsymbol{\mu}(T)] = \gamma_0 \boldsymbol{\Sigma} \boldsymbol{G}^H (\boldsymbol{\alpha}, \boldsymbol{\beta}) \boldsymbol{Y}$.

3) UPDATING FOR ρ

The approximated posterior w.r.t. ρ is given by

$$q(\rho) = \exp\left\{\langle \ln p(\boldsymbol{\gamma}|\rho)\,p(\rho)\rangle\right\}. \tag{34}$$

Substituting (23) and (25) into (34), the approximated posterior for ρ obeys a Gamma distribution and the mean of ρ is given by

$$\langle \rho \rangle = \frac{\lambda PK + a}{\displaystyle\sum_{\eta=1}^{PK} \langle \gamma_\eta \rangle / 2 + b}. \tag{35}$$

4) UPDATING FOR NOISE VARIANCE

The approximated posterior w.r.t. γ_0 is given by

$$q(\gamma_0) \propto \exp\left\{\langle p(\boldsymbol{Y}|\boldsymbol{\Omega}, \gamma_0, \boldsymbol{\alpha}, \boldsymbol{\beta})\,p(\gamma_0; c, d)\rangle\right\}. \tag{36}$$

Substituting (20) and (21) into (36), the approximated posterior distribution of γ_0 is given by

$$\langle \gamma_0 \rangle = \frac{MT + c}{T\|\underline{\boldsymbol{Y}} - \boldsymbol{G}(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}})\underline{\boldsymbol{\Omega}}\|_F^2 + Tr(GSG) + d}. \tag{37}$$

where $GSG = \boldsymbol{G}(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}})\boldsymbol{\Sigma}\boldsymbol{G}(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}})$, $\underline{\boldsymbol{\Omega}} = \boldsymbol{\Omega}/\sqrt{T}$ and $\underline{\boldsymbol{Y}} = \boldsymbol{Y}/\sqrt{T}$.

5) Updating FOR 2D OFF-GRID PARAMETERS

Since no informative prior for the off-grid difference $(\boldsymbol{\alpha}, \boldsymbol{\beta})$ is available, the inference of $(\boldsymbol{\alpha}, \boldsymbol{\beta})$ can be obtained by minimizing the negative expected log-likelihood function as

$$\boldsymbol{\varphi} = \arg\min_{(\boldsymbol{\alpha}, \boldsymbol{\beta})} \langle -\ln p(\boldsymbol{Y}, \boldsymbol{\Omega}, \boldsymbol{\gamma}; \boldsymbol{\alpha}, \boldsymbol{\beta})\rangle$$

$$= \arg\min_{(\boldsymbol{\alpha}, \boldsymbol{\beta})} \left\langle \frac{1}{T}\sum_{t=1}^T \|\boldsymbol{y}(\tau) - \boldsymbol{G}(\boldsymbol{\alpha}, \boldsymbol{\beta})\boldsymbol{\sigma}(\tau)\|_2^2 \right\rangle$$

$$= \arg\min_{(\boldsymbol{\alpha}, \boldsymbol{\beta})\in\Theta} \{\boldsymbol{\varphi}^T \boldsymbol{P}\boldsymbol{\varphi} - 2\boldsymbol{v}^T\boldsymbol{\varphi} + C\} \tag{38}$$

where $\boldsymbol{\varphi} = \left[\boldsymbol{\alpha}^T, \boldsymbol{\beta}^T\right]^T$, C is a constant term independent of $\boldsymbol{\varphi}$, $\Theta = \{([-\Delta\frac{R}{2}, \Delta\frac{R}{2}], [-\Delta\frac{\theta}{2}, \Delta\frac{\theta}{2}])\}^{PK}$, $\boldsymbol{P}$ is a positive semidefinite matrix and

$$\boldsymbol{P} = \Re\left\{\left(\boldsymbol{G}'^H\boldsymbol{G}'\right)^* \odot \left(\begin{bmatrix}\underline{\mathcal{U}}\\\underline{\mathcal{U}}\end{bmatrix}\begin{bmatrix}\underline{\mathcal{U}}\\\underline{\mathcal{U}}\end{bmatrix}^H + \begin{bmatrix}\boldsymbol{\Sigma} & \boldsymbol{\Sigma}\\\boldsymbol{\Sigma} & \boldsymbol{\Sigma}\end{bmatrix}\right)\right\} \tag{39}$$

$$\boldsymbol{v} = \Re\left\{\frac{1}{T}\sum_{\tau=1}^T diag\left(\begin{bmatrix}\boldsymbol{\mu}(\tau)\\\boldsymbol{\mu}(\tau)\end{bmatrix}^*\right)\boldsymbol{G}'^H\,(\boldsymbol{y}(\tau) - \boldsymbol{G}_0\boldsymbol{\mu}(\tau))\right\}$$

$$\quad - \Re\left\{diag\left(\begin{bmatrix}\boldsymbol{B}^H\boldsymbol{G}_0\boldsymbol{\Sigma} & \\ & \boldsymbol{C}^H\boldsymbol{G}_0\boldsymbol{\Sigma}\end{bmatrix}\right)\right\} \tag{40}$$

where $\boldsymbol{G}' = [\boldsymbol{B}, \boldsymbol{C}]$. Next, the joint sparsity of $\boldsymbol{\varphi}$ with $\boldsymbol{\sigma}(\tau)$ is exploited so that the dimension of $\boldsymbol{\varphi}$ can be reduced according to the number of true targets L for efficiently computing $\boldsymbol{\varphi}$. The sparse prior information is exploited, only L entries of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ that correspond to locations of the L maximum entries of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are calculated. It should be noted that we estimate the 2D off-grid error denoted as $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ simultaneously rather than separately in our frame. Because $\boldsymbol{\varphi}$ is obtained by stacking $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, $2L$ entries of $\boldsymbol{\varphi}$ that correspond to the maximum $2L$ entries of $\boldsymbol{\gamma}$ should be preserved and others be set to zeroes. As a result, $\boldsymbol{\varphi}$, $\boldsymbol{v}$ and $\boldsymbol{P}$ can be truncated into dimension of $2L \times 1$ or $2L \times 2L$. Their truncated versions are still denoted by $\boldsymbol{\varphi}$, $\boldsymbol{v}$ and $\boldsymbol{P}$ hereafter just for simplicity. Specifically, if $\boldsymbol{P}$ is invertible and $\hat{\boldsymbol{\varphi}} = \boldsymbol{P}^{-1}\boldsymbol{v} \in \boldsymbol{R}^{2L \times 1}$. Otherwise, $\hat{\boldsymbol{\varphi}}$ is updated element wise, i.e., at each step we update one φ_i by fixing up the other entries of $\boldsymbol{\varphi}$. For $i = 1, \ldots, 2L$,

$$\varphi_i = \frac{v_i - ((\boldsymbol{P})_i)_{-i}^T\boldsymbol{\varphi}_{-i}}{P_{i,i}} = \begin{cases} \alpha_i, & i \le L \\ \beta_{i-L}, & i > L \end{cases} \tag{41}$$

where $\left(\hat{\alpha}_l, \hat{\beta}_l\right) \in ([-\Delta R/2, \Delta R/2], [-\Delta\theta/2, \Delta\theta/2])$ should be satisfied, and the subscript $-i$ of the vector stands for the vector with the ith entry being zero.

Thus, the proposed 2DOGSBL algorithm is implemented as follows. First, the hyperparameters $\boldsymbol{\gamma}$, γ_0, $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are initialized, and $\boldsymbol{\Sigma}$ and $\boldsymbol{\mu}(\tau)$, $\tau = 1, \ldots, T$, are calculated by (30) and (31) respectively. Then $\boldsymbol{\gamma}$, ρ, γ_0, $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are updated by (33), (35), (37) and (38), respectively. The process is repeated until the convergence criterion that $\frac{\|\boldsymbol{\gamma}^{iter+1} - \boldsymbol{\gamma}^{iter}\|_2}{\|\boldsymbol{\gamma}^{iter}\|_2} < tol$ or the maximum number of iterations is reached is satisfied, where tol is a user-defined tolerance and the superscript $iter$ refers to the iteration. According to the property of expectation-maximization algorithm [37], the convergence of the algorithm is guaranteed since the function $p(\boldsymbol{\gamma}, \gamma_0, \boldsymbol{\alpha}, \boldsymbol{\beta}|\boldsymbol{Y})$ increases with iteration.

C. 2DOGSBL-2SVDs

In the practical implement of the 2DOGSBL algorithm, the computation burden is often huge, which is mainly caused by the large number of pulses T. In [29], SVD is used to reduce the columns of the measurement matrix $\boldsymbol{Y}$. In this section, another SVD is used to reduce its rows. We refer

to this scheme of use of SVD twice as 2SVDs. The dimensionality reduction of the measurement matrix $Y \in \mathbb{C}^{NMT \times T}$ preserves the most significant components while it reduces the computation of the signal reconstruction process and the sensitivity to the measurement noise.

Consider the imaging model in (14) where $G(\alpha, \beta) = G[I_{PK} \; \Delta_\alpha \; \Delta_\beta]^T$. There is a correlation among the row elements used for constructing G with respect to τ and that $Rank(G) \leq NM < NMT$ can be satisfied when $T > 1$. To some extent, the information in G does not significantly increase with a larger T. It calls for the dimensionality reduction of G. Performing the SVD of the observation matrix $G = [G_0 \; B \; C] \in \mathbb{C}^{NMT \times 3PK}$ without unknown α and β, we have $G = U_G S_G V_G^H$, where $S_G = diag(\lambda_1, \ldots, \lambda_{\min\{MNT, 3PK\}})$ with $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_{\min\{MNT, 3PK\}} > 0$. The optimal rank-$k_r$ approximation of G, denoted as G_r, is computed by $G_r = U_{k_r} S_{k_r} V_{k_r}^H$, where U_{k_r} consists of the leftmost k_r columns of U_G, $V_{k_r}^H$ consists of the rightmost k_r rows of V_G^H and $S_{k_r} = diag(\lambda_1, \ldots, \lambda_{k_r})$ is a diagonal matrix.

According to the theory of principal component analysis (PCA) [38], G_r can be obtained by preserving most energy of G in [39]. In our work, we consider that the leakage of the energy of G is so small that it can be negligible. Given $\lambda_i > 0$, k_r can be obtained by approximating $\delta = \sum_{i=1}^{k_r} \lambda_i / \sum_{i=1}^{\min\{MNT, 3PK\}} \lambda_i$ to 1, where k_r represents the number of the largest k_r singular values of G. Thus, the left singular values can be discarded as noise components. Then, with the projection matrix U_G^H, the dimensionality reduction version of G can be obtained as

$$G_{S_r} = U_G^H G = S_{k_r} V_{k_r}^H = [U_G^H G_0 \; U_G^H B \; U_G^H C] \quad (42)$$

Most of the energy of Y can be preserved as a good representation with fewer rows, and we have

$$Y_r = U_G^H Y = G_{S_r}[I_{PK} \; \Delta_\alpha \; \Delta_\beta]^T \Omega + E_L \quad (43)$$

where $Y_r \in \mathbb{C}^{k_r \times T}$ and $E_L = U_G^H E \in \mathbb{C}^{k_r \times T}$.

Another dimensionality reduction is implemented on the number of columns of Y_r. A similar subspace-based analysis in [29] is adopted. In the noise-free case of (41), we have $Y_r = G_{S_r}[I_{PK} \; \Delta_\alpha \; \Delta_\beta]^T \Omega$ with $L \leq T$ and $Rank(Y_r) \leq Rank(\Omega) \leq L$. We use the SVD of the measurement matrix $Y_r = U_c S_c V_c^H$. Let $V_c = [V_L \; V_{TL}]$, where V_L and V_{TL} are matrices that consist of the first L and the other $T - L$ columns of V_c, respectively. Then we have that $Y_{rc} = Y_r V_L \in \mathbb{C}^{k_r \times L}$ preserves all signal information. Let $\Omega_c = \Omega V_L$ and $E_{rc} = E_L V_L$. From (41), we have totally new version of (19) which can be expressed as

$$Y_{rc} = G_{S_r}[I_{PK} \; \Delta_\alpha \; \Delta_\beta]^T \Omega_c + E_{rc} \quad (44)$$

where Y_{rc}, G_{S_r}, Ω_c and E_{rc} are all new. The joint sparsity still holds in Ω_c and the proposed 2DOGSBL algorithm can be applied. As the resulting algorithm, 2DOGSBL-2SVDs is used to solve (42) to estimate Ω_c with α and β.

The complexity for the computation of the proposed method is analyzed. It can be shown that 2DOGSBL-2SVDs has a computational complexity in order of $O(k_r P^2 K^2)$ per iteration while that for 2DOGSBL is $O(\max(NMTP^2K^2, N^2M^2T^2PK))$ per iteration. The additional computational workloads of $O(\max(k_r^2 T, k_r T^2))$ and $O(\max(9NMTP^2K^2, 3N^2M^2T^2PK))$ are for the SVDs of Y_r and G in 2DOGSBL-2SVDs, respectively. Given $PK > NM \geq k_r$, it can be shown that the whole computational workload of 2DOGSBL-2SVDs is less than that of 2DOGSBL.

D. RANGE-ANGLE IMAGE FORMATION

Since the reflection coefficients of targets Ω are assumed to be complex-valued, the power estimate of Ω in range-angle plane is used to form the range-angle image. The locations of the highest peaks in the range-angle image with off-grid error $(\hat{\alpha}, \hat{\beta})$ are used to estimate the target locations. Let $\hat{\Omega} = \Omega_c V_L^H$, $\hat{\mathcal{U}}$ and $\hat{\Sigma}$ be the estimates of the reflection coefficient matrix Ω, and the mean $\mathcal{U}$ and the covariance $\hat{\Sigma}$ of Ω_c, respectively. Then the PDF of the reconstructed reflection coefficients $p(\hat{\Omega}) = \prod_{t=1}^{T} \mathcal{CN}(\hat{\sigma}(\tau) | \hat{\mu}(t), \Sigma)$. Consider $\hat{\Omega}$ row by row, We have $(\hat{\Omega})^\eta \sim \mathcal{CN}((\hat{\mathcal{U}})^\eta V_L^H, \hat{\Sigma}_{\eta, \eta} V_L V_L^H)$, $\eta = 1, \ldots, PK$. Let ψ be the vector representation of the gridded map $(\tilde{R}, \tilde{\theta})$ of the range-angle plane. ψ_η denotes the power of the ηth location which corresponds to the $(\hat{p}, \hat{k}) = (ceil(\eta/K), \mod(\eta, K))$th element of $(\tilde{R}, \tilde{\theta})$, where

$$(\hat{p}, \hat{k}) = \begin{cases} (ceil(\eta/K), \mod(\eta, K)), & \mod(\eta, K) \neq 0 \\ (ceil(\eta/K), K), & others \end{cases} \quad (45)$$

where $ceil(\cdot)$ and $\mod(\cdot)$, which are matlab functions, denote the operations of rounding toward positive infinity and returning modulus after division, respectively.

The estimate of the power ψ_η can be obtained by computing the expectation

$$\begin{aligned} \hat{\psi}_\eta &= \frac{1}{T}\left\langle \|\hat{\Omega}^\eta\|_2^2 \right\rangle \\ &= \frac{1}{T}\left(\|\hat{\mathcal{U}}^\eta V_L^H\|_2^2 + Tr\left\{ \hat{\Sigma}_{\eta, \eta} V_L V_L^H \right\} \right) \\ &= \frac{1}{T}\left(\|\hat{\mathcal{U}}^\eta\|_2^2 + L\hat{\Sigma}_{\eta, \eta} \right). \end{aligned} \quad (46)$$

Thus, the locations of the nearest grid points are estimated using the locations of the highest peaks of the spectrum. Suppose that the grid point indices of the highest L peaks of $\psi = [\psi_1, \ldots, \psi_{PK}]^T$ are $(\hat{p}_l, \hat{k}_l)$, $l = 1, \ldots, L$, and the estimated locations are $(\hat{R}_p, \hat{\theta}_k) = (\tilde{R}_{\hat{p}_l}, \tilde{\theta}_{\hat{k}_l}) + (\hat{\alpha}_{\hat{p}_l}, \hat{\beta}_{\hat{k}_l})$.

V. EXPERIMENT RESULTS

The MIMO radar under consideration contains $N = 5$ transmit antennas and $M = 5$ receive antennas. The antennas are arranged to form a uniform linear array (ULA), with $d_t = d_r = 0.5\lambda$. Carrier frequency f_c is $10GHz$. Transmitted pulse width T_s is $2\mu s$. Bandwidth B is $15MHz$. Frequency interval Δf between two adjacent channels is

C. Wen *et al.*: MIMO Radar Imaging With Multiple Probing Pulses for 2D Off-Grid Targets via Variational Sparse Bayesian Learning

IEEE *Access*

15*MHz* with $\kappa_0 = 30$ which meets the requirement of orthogonal multitransmission. The angle grid interval is $\vartheta \cdot \Delta\theta$, where $\Delta\theta = 0.1°$ denotes a constant and ϑ is the factor. A uniform sampling is adopted. The uniform computation grids in angle and range dimension can be written as $\{-1.6° \cdot \vartheta, (-1.6° + \Delta\theta) \cdot \vartheta, \ldots, 1.6° \cdot \vartheta\}$ and $\{998.6m, 998.6m + \Delta R, \ldots, 1001.5m\}$ with $\Delta R = 0.1m$. Besides, we set the reference range and frequency to $1000m$ and f_c, respectively, in order to reduce the approximation error in (13). The number of snapshots is set to $T = 100$ in the MMV case. The targets are modeled as Swerling II case. We assume that the coefficients of the targets are standard complex normally distributed. For the proposed 2DOGSBL-2SVDs, we initialize $\rho = 10^{-4}$, $\lambda = \kappa = 0$, $\boldsymbol{\alpha} = \boldsymbol{\beta} = \mathbf{0}$, $\gamma_0 = 1$ and $\boldsymbol{\gamma}$ is initialized as a vector with all entries being 1. In our method, the large energy loss of $\boldsymbol{G}$ is not expected and $\delta > 0.9999$ with $\lambda_i > 0.1$, $i = 1, \ldots, k_r$ is approximately required. For the sake of simplicity, $k_r = MN$ is set in our experiment.

We take L1 and Lq as the representatives of ℓ_1 and the more general ℓ_q $(0 < q \le 1)$ optimization methods based on on-grid MMV model, respectively. The value of $0 < q \le 1$ gives sparser solution and it can be chosen according to [32]. For small values of q have been found to have a higher likelihood of getting trapped in local minima. It should be noted that the large dimensionality of $\boldsymbol{G}_0$ in our on-grid model leads to a huge burden during signal reconstruction process. Besides, the employment of subspace-idea based SVD [12] can alleviate the sensitivity to the measurement noise in the MMV case. For fair comparison, $\boldsymbol{G}_0$ used in on-grid model based methods in our experiment is replaced by $\boldsymbol{G}_{S_{r0}} = \boldsymbol{U}_G^H \boldsymbol{G}_0$ obtained by 2SVDs. With the off-grid error $(\boldsymbol{\alpha}, \boldsymbol{\beta})$ being zero, the on-grid version of the model in (42) solved by both of L1 and Lq can be written as

$$\boldsymbol{Y}_{rc} = \boldsymbol{G}_{S_{r0}} \boldsymbol{\Omega} + \boldsymbol{E}_{rc}. \tag{47}$$

Then, L1 and Lq are used to solve (45) estimate $\boldsymbol{\Omega}$, and they are referred to as L1-2SVDs and Lq-2SVDs in our experiment.

In addition, we also compare our method with the off-grid MMV model based method, i.e. OGSBI [29]. It should be noted that the large dimensionality of $\boldsymbol{G}(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \boldsymbol{G}[\boldsymbol{I}_{PK} \ \boldsymbol{\Delta}_\alpha \ \boldsymbol{\Delta}_\beta]^T$ in our off-grid model of (19) leads to a great complexity for OGSBI which only models the angle error and $\boldsymbol{\Delta}_\alpha = diag(\mathbf{0})$ is assumed in problem (19). Although the measurement matrix can be thinner via the employment of subspace-idea based SVD in [29] which to some extent reduces the computation, it is still difficult to solve the problem in (19). For fair comparison, the dimensionality reduction has been preprocessed via the employment of 2SVDs in Section IV. Then, a similar model with (42) solved by OGSBI can be expressed as

$$\boldsymbol{Y}_{rc} = \boldsymbol{G}_{S_r}[\boldsymbol{I}_{PK} \ \boldsymbol{\Delta}_\alpha \ \boldsymbol{\Delta}_\beta]^T \boldsymbol{\Omega}_c + \boldsymbol{E}_{rc} \tag{48}$$

where $\boldsymbol{\Delta}_\alpha = diag(\mathbf{0})$. As the result, OGSBI is used to solve (46) to estimate $\boldsymbol{\Omega}_c$ with only angle error $\boldsymbol{\beta}$ and hence it is referred to as OGSBI-2SVDs in our experiment.
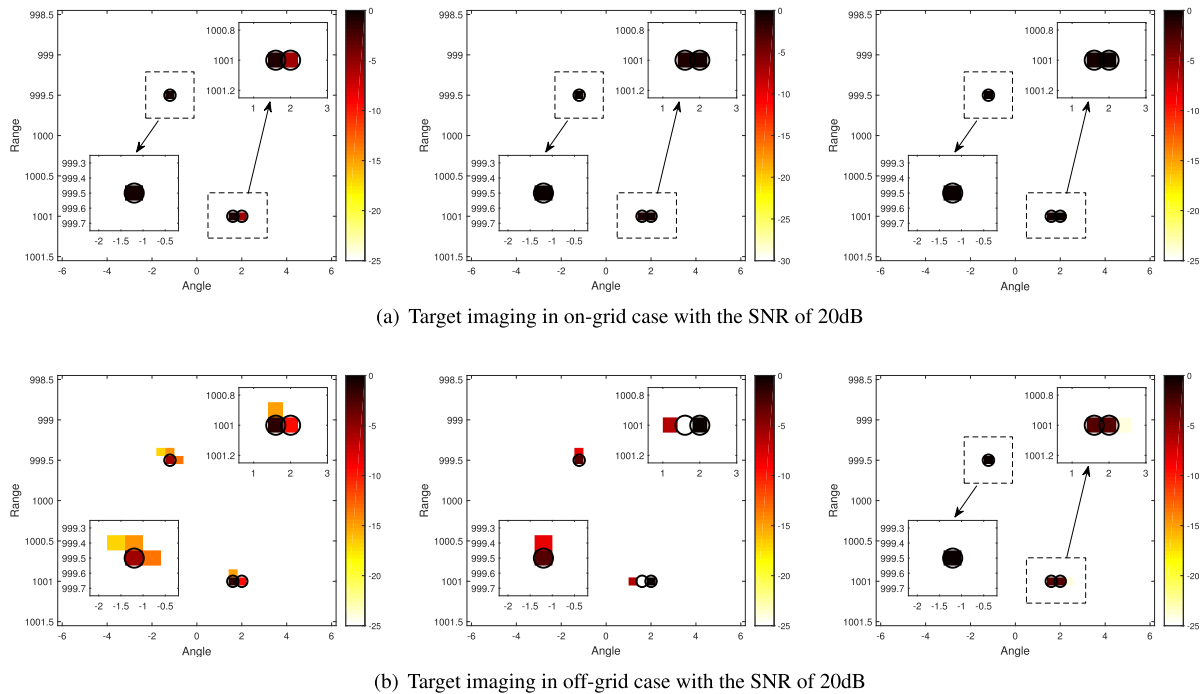
In subsection 5.1, we present the imaging results on three targets in both on-grid and off-grid cases with different SNRs and demonstrate the improvements in visual image quality, as compared with the on-grid model based methods. In subsection 5.2, we provide a quantitative comparison of our approach with on-grid model based methods.

A. VISUALLY QUALITATIVE RESULTS

In the range-angle image, the locations of the targets are usually the nearest grid points. In this subsection, we compare ℓ_1-2SVDs, M-FOCUSS-2SVDs and 2DOGSBL-2SVDs on the imaging results with the estimated nearest grid points in both the on-grid and the off-grid cases. The factor of the angle grid interval $\vartheta = 4$ is set. We consider three on-grid targets at the locations $(\tilde{R}_1, \tilde{\theta}_1) = (999.5m, -1.2°)$, $(\tilde{R}_2, \tilde{\theta}_2) = (1001.0m, 1.6°)$ and $(\tilde{R}_3, \tilde{\theta}_3) = (1001.0m, 2.0°)$. Correspondingly, the three targets in the off-grid case with 2D off-grid errors are considered with the locations $(R_1, \theta_1) = (999.47m, -1.18°)$, $(R_2, \theta_2) = (1000.98m, 1.55°)$ and $(R_3, \theta_3) = (1001.01m, 2.09°)$. Two SNR cases of $SNR = 20dB$ and $SNR = 0dB$ are considered. In this subsection, the user-defined tolerance $tol = 10^{-3}$ with the maximum number of iterations being 500 is set.

Figs. 2(a) and (b) show the comparative results for the on-grid and the off-grid targets respectively with the SNR of 20dB. We observe that all the methods perform successful in on-grid case at the relatively high SNR of 20dB in Fig. 2(a). However, in off-grid case, spurious peaks appear in the results of L1-2SVDs and Lq-2SVDs in Fig. 2(b). Lq-2SVDs shows its good property of sparsity promoting but sensitivity to the off-grid error, and the off-grid error even causes its misestimation of one target location parameter. By contrast, 2DOGSBL-2SVDs method provides high resolution as well as spurious peak suppression in off-grid case. It is shown that the superior results of 2DOGSBL-2SVDs method benefits from the accurate estimation and correction of the 2D off-grid error.

Figs. 3(a) and (b) show the comparative results for the on-grid and the off-grid targets respectively with the SNR of 0dB. It is shown from Figs. 2(a) and Figs. 3(a) that almost all the methods benefit from their denoising effects, which are closely related to their sparsity promoting and the common 2SVDs. As is shown in Figs. 3(a) and (b), under the relatively low SNR of 0dB, Lq-2SVDs shows its good property of sparsity promoting in low SNR case but sensitivity to the off-grid error. As is shown in the middle column of Figs. 3(b), Lq-2SVDs misestimates the location and amplitude of one target in off-grid case. On the other hand, both in off-grid case and at relatively low SNR, 2DOGSBL-2SVDs method provides high resolution as well as relatively well spurious peak suppression. In fact, both of the noise and the off-grid error may lead to degraded quality of imaging and spurious peaks, which also degrades the accuracy of parameter

IEEE *Access*

C. Wen *et al.*: MIMO Radar Imaging With Multiple Probing Pulses for 2D Off-Grid Targets via Variational Sparse Bayesian Learning

(a) Target imaging in on-grid case with the SNR of 20dB

(b) Target imaging in off-grid case with the SNR of 20dB

FIGURE 2. Left-Images reconstructed by L1-2SVDs. Middle-Images reconstructed by Lq-2SVDs. Right-Images reconstructed by 2DOGSBL-2SVDs. The circle represents the on-grid location of a true target, and the filled square represents a target estimate. The darkness represents the normalized amplitude of the target coefficients, and the observed amplitude range is 25dB. The darker the filled square is, the larger the amplitude is.

estimation. The results demonstrate that the proposed method is more robust to Gaussian measurement noise and 2D off-grid errors than L1-2SVDs and Lq-2SVDs. All the three sparse reconstruction based methods may utilize the SVD to improve the performance at low SNR, but it is difficult for only SVD to handle well with the observation model inaccuracies caused by off-grid errors which may lead to degradations and spurious peaks in the reconstructed images. The main reason for the superior performance of the proposed method is that the 2D off-grid errors are carefully modeled with a noninformative prior and well estimated within our framework. It can be seen from Fig. 3 that 2DOGSBL-2SVDs method can handle with both of the 2D off-grid error and the low SNR.

B. QUANTITATIVE COMPARISON

In this subsection, we compare 2DOGSBL-2SVDs with the on-grid model based methods L1-2SVDs and Lq-2SVDs and the off-grid model based method, i.e., OGSBI-2SVDs, in terms of MSE with respect to the grid interval and SNR. We assume there are $L = 2$ targets. The true range coordinates R_1 and R_2 are uniformly generated with intervals [$999.8m - \Delta R/2, 999.8m + \Delta R/2$] and [$1000.2m - \Delta R/2, 1000.2m + \Delta R/2$] respectively, and the true angle coordinates θ_1, θ_2 are uniformly generated with intervals [$-2.4° - \vartheta \cdot \Delta\theta/2, -2.4° + \vartheta \cdot \Delta\theta/2$] and [$2.4° - \vartheta \cdot \Delta\theta/2, 2.4° + \vartheta \cdot \Delta\theta/2$] respectively, where the factor of the angle grid interval $\vartheta = 2, 4, 6, 8$. Two SNR cases of $SNR =$

$10dB$ and $SNR = 0dB$ are considered. We set $tol = 10^{-3}$ and the maximum number of iterations to 1000.

For each case of SNR and ($\vartheta \cdot \Delta\theta, \Delta R$) with $\vartheta = 2, 4, 6, 8$, the MSEs of the angle and the range estimations are averaged respectively over 200 trials:
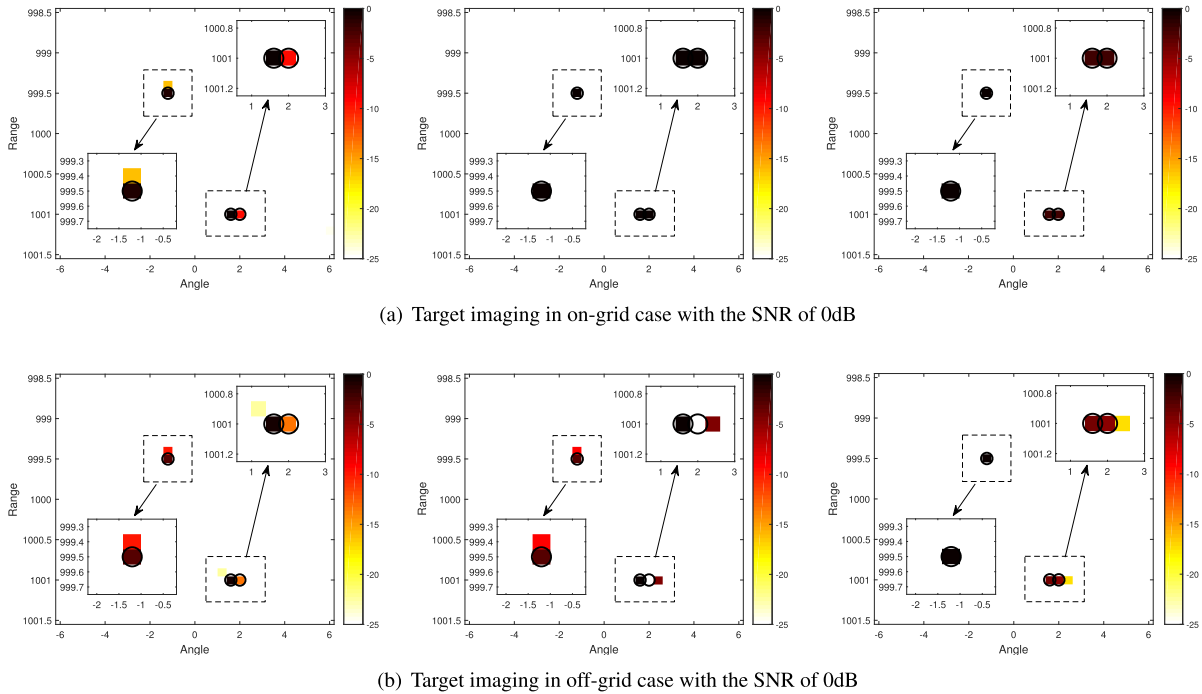
$$MSE\_angle = \frac{1}{200L} \sum_{i=1}^{200} \sum_{l=1}^{L} (\hat{\theta}_l^i - \theta_l)^2 \qquad (49)$$

and

$$MSE\_range = \frac{1}{200L} \sum_{i=1}^{200} \sum_{l=1}^{L} (\hat{R}_l^i - R_l)^2 \qquad (50)$$

where the superscript i refers to the ith trial. It should be noted that there exists a lower bound for the MSE of on-grid MMV model based methods like L1-2SVDs and Lq-2SVDs regardless of the SNR since the best locations estimate that they can obtain is the grid point nearest to the true location. Since the angle and the range of true targets are uniformly distributed, the lower bounds can be calculated as $LB\_angle = (\vartheta \cdot \Delta\theta)^2/12$ and $LB\_range = (\Delta R)^2/12$ which are the expectation in the case of limited trials. However, the lower bounds here are computed independently without the coupling effect between the 2D off-grid errors taken into account. This means that the lower bounds may be lower than the actual ones.

Fig. 4 presents our experimental results of comparison with both on-grid and off-grid model based methods. In almost

C. Wen *et al.*: MIMO Radar Imaging With Multiple Probing Pulses for 2D Off-Grid Targets via Variational Sparse Bayesian Learning

IEEE*Access*

(a) Target imaging in on-grid case with the SNR of 0dB

(b) Target imaging in off-grid case with the SNR of 0dB

FIGURE 3. Left-Images reconstructed by L1-2SVDs. Middle-Images reconstructed by Lq-2SVDs. Right-Images reconstructed by 2DOGSBL-2SVDs. The circle represents the on-grid location of a true target, and the filled square represents a target estimate. The darkness represents the normalized amplitude of the target coefficients, and the observed amplitude range is 25dB. The darker the filled square is, the larger the amplitude is.
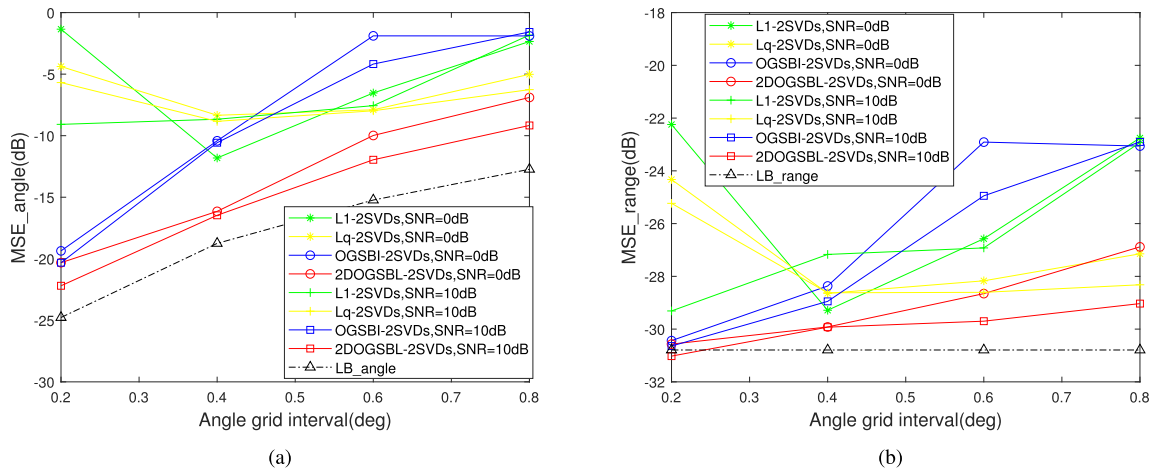
FIGURE 4. MSEs of L1-2SVDs, Lq-2SVDs, OGSBI-2SVDs and 2DOGSBI-2SVDs. (a) The MSEs of the angle estimations. (b) The MSEs of the range estimations.

all scenarios under consideration, 2DOGSBL-2SVDs has more accurate angle and range estimations than L1-2SVDs, Lq-2SVDs and OGSBI-2SVDs. This demonstrates that the proposed 2D off-grid model based method is superior to the on-grid model based methods; what's more, the off-grid model with 2D off-grid errors taken into account is more accurate than the one with only one-dimensional error modeled like OGSBI-2SVDs. It is seen that the estimation errors of the off-grid model based methods become smaller with

decreasing grid interval, while the estimation performance of on-grid model based methods like L1-2SVDs and Lq-2SVDs degrades with the grid interval being smaller than $0.2°$. This demonstrates that although a dense computation grid is necessary for on-grid model based methods like L1-2SVDs and Lq-2SVDs to obtain an accurate estimation of the target location since the gap between the true location and its nearest grid point can be narrowed, this would lead to the sensitivity to noise of the observation matrix used for

IEEE *Access*

C. Wen *et al.*: MIMO Radar Imaging With Multiple Probing Pulses for 2D Off-Grid Targets via Variational Sparse Bayesian Learning

the SSR. However, this can be largely overcome through the off-grid model which has a good performance on describing the true observation model.

It is shown in Fig. 4 that 2DOGSBL-2SVDs has better precision than OGSBI-2SVDs in different grid intervals and SNRs. The main reason is that 2DOGSBL-2SVDs has modeled the 2D off-grid errors as the range-angle dependent variables of the parameterized observation matrix, rather than OGSBI-2SVDs with range dependent variables being zeros. The off-grid modeling error may result in the degraded performance of the method without range dependent off-grid error incorporated. Besides, the proposed method adopts variational Bayesian learning framework with a unified GIG prior for coefficient variance, which has more degrees of freedom to mimic a kind of heavy-tailed distribution so that better adaptive modeling of the sparsity. The lower bound can be viewed as the expectation of the off-grid error between the true location and the computation grid under the assumption that the off-grid errors are not modeled in the observation process and they are uniformly distributed. The off-grid error can be compensated with high probability using our method, because the proposed off-grid model can be viewed as the first order approximation to the true observation model with the two variables. However, there is still coupling between the two dimensional off-grid errors that is difficult to model and compensate, so the MSEs of the proposed algorithm is still difficult to exceed the lower bound.

All experiments are carried out in Matlab on a PC with 3.6 GHz CPUs. For the three off-grid target case with 500 iterations, the average computation time of L1-2SVDs, Lq-2SVDs, OGSBI-2SVDs and 2DOGSBL-2SVDs are 31.19s, 0.04s, 33.13s, 53.64s, excluding the 2SVDs preprocessing that takes about 13.51s in our case. The computational load of 2DOGSBL-2SVDs is relatively more than the on-grid model and one-dimensional off-grid based methods, but this can be justified through the benefits provided by our imaging framework, as demonstrated in our experiments.

VI. CONCLUSION
In this article, an off-grid imaging framework of OFDM-LFM MIMO radar with multiple probing pulses is presented. The orthogonal transmission property of the OFDM LFM waveform is employed in order to obtain the image in range dimension. The targets are modeled as Swerling II case that the reflection coefficients of the targets is fixed during a pulse while varying independently from pulse to pulse. Different from the mostly proposed on-grid model for SSR, we studied the off-grid sparse imaging model for OFDM-LFM MIMO radar. The true off-grid location is modeled with the first-order partial derivative of a two-variable function, which is exploited to reduce the modeling error due to discretization of a continuous range-angle plane. From a variational Bayesian perspective, an iterative algorithm is developed for joint MIMO radar imaging and 2D position error estimation and autocorrection of off-grid targets. Besides, a subspace-like method referred to 2SVDs is used to reduce the com-

putational complexity of the signal recovery process and the sensitivity to noise. Simulation results demonstrate that the proposed approach outperforms existing on-grid and off-grid model based methods.

In future, further experimental validation, including using dedicated samples can be investigated on not only imaging the static point targets with only frequency domain features, but also more complex and dynamic targets using time-frequency representations and their micro doppler characteristics [16].

REFERENCES
[1] K. W. Forsythe, D. W. Bliss, and G. S. Fawcett, "Multiple-input multiple-output (MIMO) radar: Performance issues," in *Proc. 38th Asilomar Conf. Signals, Syst., Comput.*, Pacific Grove, CA, USA, vol. 1, Nov. 2004, pp. 305–309.

[2] D. W. Bliss and K. W. Forsythe, "Multiple-input multiple-output (MIMO) radar and imaging: Degrees of freedom and resolution," in *Proc. 37th Asilomar Conf. Signals, Syst., Comput.*, Pacific Grove, CA, USA, vol. 1, Nov. 2003, pp. 54–59.

[3] J. Li, P. Stoica, L. Xu, and W. Roberts, "On parameter identifiability of MIMO radar," *IEEE Signal Process. Lett.*, vol. 14, no. 12, pp. 968–971, Dec. 2007.

[4] R. Entezari and A. Rashidi, "Incoherent waveform design for compressed sensing radar based on pulse-train scenario," *IET Commun.*, vol. 12, no. 17, pp. 2132–2136, Oct. 2018.

[5] Y. Zhu, Y. Su, and W. Yu, "An ISAR imaging method based on MIMO technique," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 8, pp. 3290–3299, Aug. 2010.

[6] W.-Q. Wang, "Space–time coding MIMO-OFDM SAR for high-resolution imaging," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 8, pp. 3094–3104, Aug. 2011.

[7] A. Pandharipande, "Principles of OFDM," *IEEE Potentials*, vol. 21, no. 2, pp. 16–19, Apr. 2002.

[8] X. Tan, W. Roberts, J. Li, and P. Stoica, "Sparse learning via iterative minimization with application to MIMO radar imaging," *IEEE Trans. Signal Process.*, vol. 59, no. 3, pp. 1088–1101, Mar. 2011.

[9] E. J. Candès, J. K. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Commun. Pure Appl. Math.*, vol. 59, no. 8, pp. 1207–1223, 2006.

[10] M. D. Buhari, G. Y. Tian, R. Tiwari, and A. H. Muqaibel, "Multicarrier SAR image reconstruction using integrated MUSIC-LSE algorithm," *IEEE Access*, vol. 6, pp. 22827–22838, 2018.

[11] C. Tang, G. Y. Tian, K. Li, R. Sutthaweekul, and J. Wu, "Smart compressed sensing for online evaluation of CFRP structure integrity," *IEEE Trans. Ind. Electron.*, vol. 64, no. 12, pp. 9608–9617, Dec. 2017.

[12] D. Malioutov, M. Cetin, and A. S. Willsky, "A sparse signal reconstruction perspective for source localization with sensor arrays," *IEEE Trans. Signal Process.*, vol. 53, no. 8, pp. 3010–3022, Aug. 2005.

[13] Y. Yu, A. P. Petropulu, and H. V. Poor, "MIMO radar using compressive sampling," *IEEE J. Sel. Topics Signal Process.*, vol. 4, no. 1, pp. 146–163, Feb. 2010.

[14] N. Ö. Onhon and M. Çetin, "A sparsity-driven approach for joint SAR imaging and phase error correction," *IEEE Trans. Image Process.*, vol. 21, no. 4, pp. 2075–2088, Apr. 2012.

[15] H. Zamani and M. Fakharzadeh, "1.5-D sparse array for millimeter-wave imaging based on compressive sensing techniques," *IEEE Trans. Antennas Propag.*, vol. 66, no. 4, pp. 2008–2015, Apr. 2018.

[16] Y. Zhao and Y. Su, "Sparse recovery on intrinsic mode functions for the micro-Doppler parameters estimation of small UAVs," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 9, pp. 7182–7193, Sep. 2019.

[17] W. He, G. Wang, J. Hu, C. Li, B. Guo, and F. Li, "Simultaneous human health monitoring and time-frequency sparse representation using EEG and ECG signals," *IEEE Access*, vol. 7, pp. 85985–85994, 2019.

[18] I. S. Merrill, *Introduction to Radar Systems.* New York, NY, USA: McGraw-Hill, 2001.

[19] M. M. Hyder and K. Mahata, "A joint sparse signal representation perspective for target detection using bistatic MIMO radar system," in *Proc. 17th Int. Conf. (DSP)*, Jul. 2011, pp. 1–5.

C. Wen *et al.*: MIMO Radar Imaging With Multiple Probing Pulses for 2D Off-Grid Targets via Variational Sparse Bayesian Learning

IEEE *Access*

[20] S. M. Hosseini, R. A. Sadeghzadeh, and B. S. Virdee, "DOA estimation using multiple measurement vector model with sparse solutions in linear array scenarios," *EURASIP J. Wireless Commun. Netw.*, vol. 2017, no. 1, p. 58, Dec. 2017.

[21] Y. C. Eldar and H. Rauhut, "Average case analysis of multichannel sparse recovery using convex relaxation," *IEEE Trans. Inf. Theory*, vol. 56, no. 1, pp. 505–519, Jan. 2010.

[22] S. D. Babacan, R. Molina, and A. K. Katsaggelos, "Bayesian compressive sensing using Laplace priors," *IEEE Trans. Image Process.*, vol. 19, no. 1, pp. 53–63, Jan. 2010.

[23] M. E. Tipping, "Sparse Bayesian learning and the relevance vector machine," *J. Mach. Learn. Res.*, vol. 1, pp. 211–244, Sep. 2001.

[24] Z.-M. Liu, Z.-T. Huang, and Y.-Y. Zhou, "An efficient maximum likelihood method for Direction-of-Arrival estimation via sparse Bayesian learning," *IEEE Trans. Wireless Commun.*, vol. 11, no. 10, pp. 1–11, Oct. 2012.

[25] L. He and L. Carin, "Exploiting structure in wavelet-based Bayesian compressive sensing," *IEEE Trans. Signal Process.*, vol. 57, no. 9, pp. 3488–3497, Sep. 2009.

[26] L. Qin, J. Tan, Z. Wang, G. Wang, and X. Guo, "Exploiting the tree-structured compressive sensing of wavelet coefficients via block sparse Bayesian learning," *Electron. Lett.*, vol. 54, no. 16, pp. 975–976, Aug. 2018.

[27] H. Zhu, G. Leus, and G. B. Giannakis, "Sparsity-cognizant total least-squares for perturbed compressive sampling," *IEEE Trans. Signal Process.*, vol. 59, no. 5, pp. 2002–2016, May 2011.

[28] A. Abtahi, S. Gazor, and F. Marvasti, "Off-grid localization in MIMO radars using sparsity," *IEEE Signal Process. Lett.*, vol. 25, no. 2, pp. 313–317, Feb. 2018.

[29] Z. Yang, L. Xie, and C. Zhang, "Off-grid direction of arrival estimation using sparse Bayesian inference," *IEEE Trans. Signal Process.*, vol. 61, no. 1, pp. 38–43, Jan. 2013.

[30] Z. Cao, L. Zhou, and J. Dai, "Sparse Bayesian approach for DOD and DOA estimation with bistatic MIMO radar," *IEEE Access*, vol. 7, pp. 155335–155346, 2019.

[31] J. Shi, G. Hu, X. Zhang, and F. Sun, "Sparsity-based DOA estimation of coherent and uncorrelated targets with flexible MIMO radar," *IEEE Trans. Veh. Technol.*, vol. 68, no. 6, pp. 5835–5848, Jun. 2019.

[32] S. F. Cotter, B. D. Rao, K. Engan, and K. Kreutz-Delgado, "Sparse solutions to linear inverse problems with multiple measurement vectors," *IEEE Trans. Signal Process.*, vol. 53, no. 7, pp. 2477–2488, Jul. 2005.

[33] P. Gong and Z. Shao, "Target estimation by iterative reweighted LQ minimization for MIMO radar," *Signal Process.*, vol. 101, pp. 35–41, Aug. 2014.

[34] N. R. Goodman, "Statistical analysis based on a certain multivariate complex Gaussian distribution (An introduction)," *Ann. Math. Statist.*, vol. 34, no. 1, pp. 152–177, Mar. 1963.

[35] Z. Zhang, S. Wang, D. Liu, and M. I. Jordan, "EP-GIG priors and applications in Bayesian sparse learning," *J. Mach. Learn. Res.*, vol. 13, no. 1, pp. 2031–2061, 2012.

[36] D. G. Tzikas, A. C. Likas, and N. P. Galatsanos, "The variational approximation for Bayesian inference," *IEEE Signal Process. Mag.*, vol. 25, no. 6, pp. 131–146, Nov. 2008.

[37] G. McLachlan and T. Krishnan, *The EM Algorithm and Extensions*. New York, NY, USA: Wiley, 1997.

[38] I. T. Jolliffe, *Principal Component Analysis*, 2nd ed. New York, NY, USA: Springer-Verlag, 2002.

[39] G. Shi, C. Chen, J. Lin, X. Xie, and X. Chen, "Narrowband ultrasonic detection with high range resolution: Separating echoes via compressed sensing and singular value decomposition," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 59, no. 10, pp. 2237–2253, Oct. 2012.
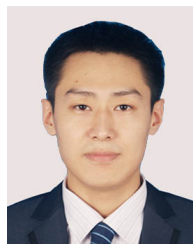
CHAO WEN received the Ph.D. degree from Xidian University, China, in 2017. He is currently a Lecturer with the Institute of Big Data Science and Industry, Shanxi University, China. His research interests include array signal processing, passive sensing, and artificial intelligence.

LU CHEN received the M.S. and Ph.D. degrees from Northwestern Polytechnical University, in 2015 and 2019, respectively. He is currently a Lecturer with the Institute of Big Data Science and Industry, Shanxi University. His research interests include machine learning, robotic vision, and low-light image enhancement.

PENGTING DUAN received the M.S. degree from Xidian University, China, in 2014. She is currently an Engineer with the North Automatic Control Technology Research Institute, China. Her research interests include parameter estimation and multi-station communication.

XUEFENG CUI received the M.S. degree from Southeast University, China, in 2012. He is currently a Senior Engineer with the North Automatic Control Technology Research Institute, China. His research interests include electromagnetic field and microwave technology, and network communication.

• • •