An autofocus technique for high-resolution inverse synthetic aperture radar imagery

Zhao, Lifan; Wang, Lu; Yang, Lei; Bi, Guoan

2014

Zhao, L., Wang, L., Bi, G., & Yang, L. (2014). An Autofocus Technique for High-resolution Inverse Synthetic Aperture Radar Imagery. IEEE Transactions on Geoscience and Remote Sensing, 52(10), 6392 - 6403.

https://hdl.handle.net/10356/104850

https://doi.org/10.1109/TGRS.2013.2296497

An Autofocus Technique for High-resolution Inverse Synthetic Aperture Radar Imagery

Lifan Zhao, Lu Wang, Guoan Bi, *Senior Member, IEEE,* Lei Yang

***Abstract*—For inverse synthetic aperture radar imagery, the inherent sparsity of the scatterers in range-Doppler domain has been exploited to achieve high-resolution range profile or Doppler spectrum. Prior to applying the sparse recovery technique, preprocessing procedures are performed for minimization of the translational motion induced Doppler effects. Due to the imperfection of coarse motion compensation, autofocus technique is further required to eliminate the residual phase errors. This paper considers the phase error correction problem in the context of sparse signal recovery technique. In order to encode sparsity, a multi-task Bayesian model is utilized to probabilistically formulate this problem in a hierarchical manner. In this novel method, focused high-resolution radar image is obtained by estimating the sparse scattering coefficients and phase errors in individual and global stages, respectively, to statistically make use of the sparsity. The superiority of this algorithm is that the uncertainty information of the estimation can be properly incorporated to obtain enhanced estimation accuracy. Moreover, the proposed algorithm achieves guaranteed convergence and avoids tedious parameter tuning procedure. Experimental results based on synthetic and practical data have demonstrated that our method has a desirable de-noising capability and can produce a relatively well-focused image of the target, particularly in low signal-to-noise ratio (SNR) and high under-sampling ratio scenarios, compared to other recently reported methods.**

***Index Terms*—Inverse synthetic aperture radar imagery, compressive sensing, high-resolution, sparse Bayesian learning, autofocus technique**

I. Introduction

INVERSE synthetic aperture radar (ISAR) has been widely used for imaging moving targets in both civilian and military applications due to its superiority of operating in all-weather and dim light conditions. Many techniques have been developed to obtain high-resolution ISAR image [1]–[5]. Range resolution is generally determined by the bandwidth of the emitted signal, while cross-range resolution depends on the coherent processing time of radar echoes and the motion characteristics of the target. To achieve a desirable radar image, linear frequency modulated (LFM) signal is often used for high range resolution and a long coherent processing interval (CPI) is required for high cross-range resolution. However, long CPI will inevitably introduce undesirable higher order Doppler effects, which would in turn, lead to smeared Doppler spectrum.

To ameliorate this issue, compressive sensing (CS) [6]–[9] has emerged as a promising technique. The theory of CS indicates that a high dimensional signal can be recovered from its low dimensional projection if the signal is parsimonious. This technique has been successfully applied to SAR/ISAR imagery to achieve high cross-range resolution with a limited number of pulses [2], [4], [10]–[13]. The superiority of the CS based approach is that the target motion is not complicated within the short CPI, where range alignment can be done more easily. However, these algorithms generally rely on the assumption that the preprocessing procedures have been perfectly conducted and no residual phase error remains in the processed data, which unfortunately is not true in practical scenarios. Because the motion of the target cannot be precisely compensated by coarse pre-processing in practical applications, phase errors induced by the translational motion of the target often inevitably exist in the processed data [14], [15]. If these errors are not properly corrected or compensated, the target image obtained by the CS based algorithms would be substantially blurred.

Conventional techniques, including phase gradient autofocus (PGA) [16], [17] and minimum entropy method (MEM) [18]–[20], have been developed to compensate for the remaining phase errors. The PGA method generally assumes the existence of prominent scatterers in a range cell and requires tuning of several parameters such as the width of the data window and the number of range cells for phase error compensation. In MEM, a minimum entropy based metric is defined, where the focused image can be obtained by iteratively evaluating the image quality until a minimum entropy criteria is achieved. Empirical results have demonstrated the effectiveness of these algorithms in ISAR imagery applications. Unfortunately, in the context of obtaining high cross-range resolution, these conventional approaches cannot be properly integrated into a compressive sensing framework to recover the high-resolution target image and estimate phase error simultaneously.

Phase error correction has been also considered by the recently reported sparse recovery technique. Alternating regularized approaches [15], [21], [22] are proposed to obtain focused images. A sparsity based autofocus technique has been reported in [21], where a sparse metric is defined to iteratively estimate the sparse scatterer coefficients and phase errors. However, this approach is developed to cope with autofocus problem only and cannot be adjusted to obtain high-resolution image simultaneously. In [15], [22], a similar idea exploiting sparsity is introduced by a regularized l_1 alternating scheme, which could also be used for simultaneously achieving high resolution and autofocus applications. Although these algorithms vary in formulations, the main idea is to obtain

The corresponding author is Dr Guoan Bi. The authors are with the School of Electrical and Electronic Engineering, Nanyang Technological University, 50 Nanyang Avenue, Singapore 639798 (email: zhao0145@e.ntu.edu.sg; wa0001lu@e.ntu.edu.sg; egbi@ntu.edu.sg; yanglei@ntu.edu.sg)

sparse scattering coefficients and phase errors via a convex optimization based framework, which will be summarized as l_1 based alternating method in the latter sections of this paper. Despite the empirical success of these methods, these convex optimization based methods generally suffer from a paradigm that the algorithm might be stuck in a shallow local minimum of the solution during the alternating process. Inevitably, the estimation error propagation phenomenon exists, since these methods are used alternately between the point estimates of the sparse scattering coefficient and phase error. To be more concrete, the estimation error of the sparse signal would degrade the estimation accuracy of phase error during iterations. This error propagation phenomenon is particularly non-negligible with under-sampled data and in low SNR conditions. It is also noted that the parameter tuning process is vital to the robust performance of these methods, however, optimal regularization parameter selection is still an open problem.

To tackle these problems, both high-resolution imagery and phase error correction are considered based on a sparse Bayesian model in this paper. In our method, a hierarchical probabilistic model is imposed on the signal to facilitate convenient inference. Subsequent parameter estimation is conducted within a multi-task learning framework [23], where variational Bayesian expectation maximization technique is used. The sparse signal as well as its hyper-parameters are estimated in the individual task level, while the phase error and noise precision are updated in the global task level. In this multi-task learning framework, phase error and noise level can be more accurately estimated in global learning stage, which also leads to better estimation of the sparse coefficients. A remarkable advantage of the proposed algorithm is that it can properly utilize uncertainty information during iterations to ameliorate the error propagation problem. Due to the utilization of Bayesian inference technique, the possibility of converging to a shallow local minimum is reduced while the convergence of the method is also guaranteed. The proposed algorithm does not require the time-consuming parameter tuning procedures to obtain the improved performances as those done in l_1 based alternating methods.

The rest of this paper is organized as follows. In Section II, the CS based ISAR imagery model is introduced. In Section III, the high-resolution and autofocus problem is formulated by a Bayesian model, where the subsequent Bayesian inference technique is derived based on the multi-task learning framework. Further analysis and discussion of the proposed algorithm are given in the latter of this section. Synthetic and practical data experimental results are given in Section IV to demonstrate the effectiveness of the proposed algorithms. Finally, conclusions are presented in Section V.

The following mathematical notations are used throughout the paper. Scalars, vectors, and matrices are denoted by lower-case letter, bold lower-case letter and bold upper-case letter, respectively. For a given matrix $\mathbf{A}$, $\mathbf{A}^{-1}$, $\mathbf{A}^T$ and $\mathbf{A}^H$ denote the inverse, transpose and conjugate transpose of $\mathbf{A}$. The (i,j)-th entry of a matrix $\mathbf{A}$ is represented by $\mathbf{A}_{ij}$. $\|\cdot\|_p$ is the l_p norm of the vector or matrix.

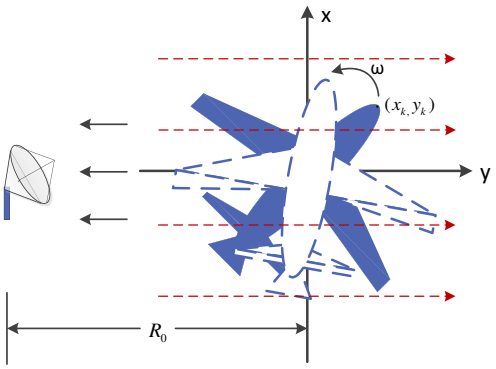

Fig. 1. ISAR geometry for data collection

II. Background

The emitted LFM signal is defined as

$$s(t) = \text{rect}\left(\frac{t}{T}\right) \cdot \exp\left[j2\pi\left(f_c t + \frac{\gamma}{2}t^2\right)\right] \tag{1}$$

where $\text{rect}(\frac{t}{T})$ represents the rectangular function of width T, f_c is the centroid frequency and γ is the chirp rate. The LFM signal is utilized to achieve high range resolution after range compression. To obtain high cross-range resolution, multiple pulses are to be emitted with interval time T_r.

Assuming there exist K scatterers centers in an imagery scene, the received radar echo can be expressed as

$$s_r(t, t_n) = \sum_{k=1}^{K} \sigma_k \cdot s\left[t - \frac{2R_k(t_n)}{c}\right]$$

where σ_k is the amplitude of the k-th scatterer, c is the speed of light, t is fast time, $t_n = nT_r$ is slow time of pulse n and $R_k(t_n)$ represents the range from scatterer k to radar in slow time t_n.

Let us define $\Delta\theta(t_n)$ as the angle variation between radar line of sight and target position. Consider an appropriate dwelling time with a small $\Delta\theta(t_n) = \omega t_n$, the range $R_k(t_n)$ can be approximated as

$$\begin{aligned} R_k(t_n) &= R_0 + y_k \cos\Delta\theta(t_n) + x_k sin\Delta\theta(t_n) \\ &\approx R_0 + y_k + x_k\Delta\theta(t_n). \end{aligned} \tag{2}$$

Figure 1 shows the ISAR geometry for data collection, where the target rotates with angular velocity ω. It is observed from (2) that only the rotational motion of the target contributes to the formation of Doppler spectrum and the undesirable translational motion is required to be compensated [24]. In the dwelling time, it is assumed that the target rotates uniformly. Under this assumption, only translational motion requires to be compensated. However, due to the imprecision of the coarse compensation for the translational motion, image obtained either by range Doppler (RD) or CS based algorithm would degrade. In the following, the mathematical model for high-resolution ISAR imagery with phase error is formulated.

The received data $\mathbf{S}_r(t, t_n)$ is arranged according to Fig. 2, where t axis and t_n axis represent fast time and slow time, respectively. In this paper, $\mathbf{S}_r$ is denoted by $\mathbf{Y}$ for notational brevity. Since the phase error in ISAR imagery often exhibits

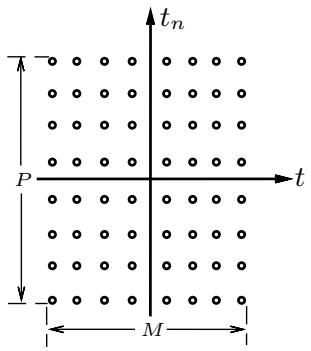

Fig. 2. The obtained data arrangement

range invariant property [15], [18], the mathematical model can therefore be given as

$$\mathbf{Y} = \mathbf{EAX} + \mathbf{N} \tag{3}$$

where $\mathbf{Y} \in \mathbb{C}^{P\times M}$ is the received data, $\mathbf{X} \in \mathbb{C}^{N\times M}$ is the sparse scattering coefficient and $\mathbf{N} \in \mathbb{C}^{P\times M}$ is zero mean Gaussian noise. The phase error matrix is denoted by $\mathbf{E} = \mathrm{diag}(e^{j\varphi_1}, ..., e^{j\varphi_P})$, which is a diagonal matrix representing slow time variant errors. Furthermore, $\mathbf{A} = [\mathbf{a}(f_1), ..., \mathbf{a}(f_N)]$ is a partial Fourier dictionary, where each atom $\mathbf{a}(f_i)$ is defined as $[e^{-j2\pi f_i \tilde{t}_1}, ..., e^{-j2\pi f_i \tilde{t}_P}]^T$. In particular, $\{\tilde{t}_1, ..., \tilde{t}_P\}$ is sampled from $nT_r, n \in \mathbb{N}$. In CS based high-resolution ISAR imagery, the number of pulses P is set to be less than the number of reconstructed Doppler cell N, i.e. $P \ll N$, to achieve high cross-range resolution.

In general, the noise $\mathbf{N}$ is modeled as an independent circularly-symmetric complex Gaussian distribution. It is obvious to see that the received signal $\mathbf{Y}$ obeys a complex Gaussian distribution[1] and the likelihood function of the observation can be formulated as

$$p(\mathbf{Y}|\mathbf{X}; \mathbf{E}) = \prod_{\mathrm{i}=1}^{\mathrm{M}} \mathcal{CN}(\mathbf{Y}_{\cdot i}|\mathbf{EAX}_{\cdot i}, \alpha_0^{-1}\mathbf{I}) \tag{4}$$

where α_0 is the noise precision or reciprocal of the variance.

III. Multi-task Sparse Bayesian Learning Autofocus Technique

A. Mathematical Model

Laplace distribution is a popular choice as a sparse prior [8], where the maximum *a posterior* (MAP) technique is utilized for parameter estimation. It can be shown that the basis pursuit de-noising (BPDN) method corresponds to the MAP estimation with a Gaussian likelihood and Laplace prior. Due to the non-conjugacy of the likelihood and prior, however, this strategy can only provide point estimation without any higher order statistical information. In order to obtain the uncertainty information during estimation, sparse Bayesian method has been introduced and developed in [25]–[28], where the signal is hierarchically modeled to impose a prior that promotes sparsity.

[1]The complex Gaussian distribution is defined as $p(\boldsymbol{\xi}|\boldsymbol{\mu}_\xi, \boldsymbol{\Sigma}_\xi) = \frac{1}{\pi^N|\Sigma_\xi|}\exp\left[-(\boldsymbol{\xi}-\boldsymbol{\mu}_\xi)^H {\boldsymbol{\Sigma}_\xi}^{-1}(\boldsymbol{\xi}-\boldsymbol{\mu}_\xi)\right]$, where $\boldsymbol{\xi} \in \mathbb{C}^N$.

In the ISAR imagery problem, $\mathbf{X}$ is hierarchically modeled to achieve sparse prior as well as convenient inference, since the number of scattering centers exhibits sparsity with respect to the complete imaging scene. Let us start by modeling each entry in $\mathbf{X}$ by following a complex Gaussian distribution,

$$p(\mathbf{X}|\boldsymbol{\alpha}) = \prod_{\mathrm{i}=1}^{\mathrm{M}} \mathcal{CN}(\mathbf{X}_{\cdot i}|\mathbf{0}, \boldsymbol{\Lambda}_i) \tag{5}$$

where $\boldsymbol{\Lambda}_i$ denotes $\mathrm{diag}(\boldsymbol{\alpha}_{\cdot i})$ and $\boldsymbol{\alpha}_{ji}$ is the variance of $\mathbf{X}_{ji}$. It is noted that when $\boldsymbol{\alpha}_{ji}$ approaches zero, its corresponding element in $\mathbf{X}$ will approach zero and be pruned away from the model.

Furthermore, the variance $\boldsymbol{\alpha}$ of the scatterer coefficient $\mathbf{X}$, also known as hyper-parameter, obeys an independent Gamma distribution[2] for convenient inference since it is the conjugate prior of Gaussian distribution [29],

$$\mathrm{p}(\boldsymbol{\alpha}_{\cdot i}|\boldsymbol{\lambda}) = \prod_{k=1}^{N} \Gamma(\alpha_{ki}|\eta, \lambda_i), \quad i = 1, ..., M \tag{6}$$

where $\boldsymbol{\alpha}_{ki}$ is the k-th element in $\boldsymbol{\alpha}_i$. It can be proved that given $\eta = 3/2$, the marginalized distribution of $\mathbf{X}_{\cdot i}$ is a complex Laplace distribution, where the parameter $\boldsymbol{\lambda}$ determines the sparsity of the distribution.

In order to automatically infer $\boldsymbol{\lambda}$ controlling the sparsity of the prior during the learning process, a Gamma distribution is imposed,

$$p(\boldsymbol{\lambda}|v_1, v_2) = \prod_{i=1}^{M} \Gamma(\lambda_i|v_1, v_2). \tag{7}$$

Finally, the noise precision is modeled as a Gamma distribution,

$$p(\alpha_0|v_3, v_4) = \Gamma(\alpha_0|v_3, v_4). \tag{8}$$

In this way, the noise level estimation can be incorporated into the learning procedures.

The main difference between the above-mentioned modeling and the ones for ISAR imagery previously reported in [15], [22] is the hierarchical modeling procedure to encode signal sparsity. In [15], a sparsity-inducing Laplace prior is directly imposed on the signal, where the sparse solution corresponds to the MAP estimation. Rather than merely seeking the mode of posterior, the approximate posterior distribution is obtained in the above hierarchical modeling. Because the probabilistic distribution can be obtained, it is regarded as a full Bayesian method. It is proved that this hierarchical modeling can be used to achieve better sparse solutions [26]. More importantly, higher order statistical information, including the estimation covariance matrix, can be naturally obtained with such a full Bayesian framework.

Thus, the posterior distribution can be expressed as

$$p(\mathbf{X}, \boldsymbol{\alpha}, \boldsymbol{\lambda}, \alpha_0|\mathbf{Y}) = \frac{p(\mathbf{Y}|\mathbf{X}, \alpha_0)p(\mathbf{X}|\boldsymbol{\alpha})p(\boldsymbol{\alpha}|\boldsymbol{\lambda})p(\boldsymbol{\lambda})p(\alpha_0)}{p(\mathbf{Y})}. \tag{9}$$

However, the calculation of $p(\mathbf{Y})$ demands a multi-dimensional integral, which is often intractable. In this scenario, one needs to perform either Monte Carlo Markov Chain

[2]The Gamma distribution is defined as $p(\xi|a, b) = \frac{b^a}{\Gamma(a)}\xi^{a-1}e^{-b\xi}$.

(MCMC) [30] or variational Bayesian (VB) approximation method in [31], [32] to obtain the sampled or approximated posterior. Because the expensive sampling procedure in MCMC often requires an inhibitive computational complexity. The inference is carried out based on the VB method in this paper due to its computational efficiency.

B. Multi-task Variational Sparse Bayesian Learning

In this framework, the estimation of $\mathbf{X}$, $\boldsymbol{\alpha}$ and $\boldsymbol{\lambda}$ is obtained individually since they are treated as task-specific parameters. The estimation of α_0 and $\mathbf{E}$ is performed in a global manner due to the task-invariant property. The key idea of the proposed method is statistically learning of the parameters to jointly achieve sparsity within a multi-task learning framework. According to the graphic model, $\mathbf{X}$, $\boldsymbol{\alpha}$, $\boldsymbol{\lambda}$ and α_0 can be treated as latent variables and $\mathbf{E}$ is the parameter. Since the marginalized distribution, $p(\mathbf{Y})$, is intractable, the direct expectation maximization technique is not applicable, whereby either sampling or approximation method is required for inference. In order to obtain efficient inference, a variational Bayesian method is utilized.

In this method, to calculate the intractable posterior $p(\mathbf{X},\boldsymbol{\alpha},\boldsymbol{\lambda},\alpha_0|\mathbf{Y})$, a distribution is defined to approximated this posterior. With the mean-field assumption that the approximated posterior is factorisable, we have

$$p(\mathbf{X},\boldsymbol{\alpha},\boldsymbol{\lambda},\alpha_0|\mathbf{Y}) \approx q(\mathbf{X})q(\boldsymbol{\alpha})q(\boldsymbol{\lambda})q(\alpha_0). \tag{10}$$

The VB procedure can be obtained by minimizing the KL divergence[3] between the true posterior and the approximated one, expressed as

$$q^*(\boldsymbol{\Theta}) \quad = \quad \arg\min_{q(\boldsymbol{\Theta})} D_{\mathcal{KL}}(q(\boldsymbol{\Theta})||p(\boldsymbol{\Theta}|\mathbf{Y}))$$

where $\boldsymbol{\Theta}$ is $\{\mathbf{X},\boldsymbol{\alpha},\boldsymbol{\lambda},\alpha_0\}$. With some derivations, it is shown that the posterior for the variables can be updated in a round-robin manner [29], [31],

$$q^*(\boldsymbol{\Theta}_i) = \exp\left\{\langle \ln p(\boldsymbol{\Theta},\mathbf{Y})\rangle_{q(\boldsymbol{\Theta}\setminus\boldsymbol{\Theta}_i)}\right\} \tag{11}$$

where $\langle\cdot\rangle_{q(\cdot)}$ represents the expectation with respect to $q(\cdot)$. Subsequently, since no prior for phase error matrix $\mathbf{E}$ is available, the inference of $\mathbf{E}$ can be obtained by maximizing the expected log likelihood function.

1) Individual Learning Stage: In this stage, the variables are updated individually since they are specific in each task and do not share any common information.

i). Updating rule for $\mathbf{X}$: The approximated posterior can be expressed as

$$q^*(\mathbf{X}) = \exp\left\{\left\langle \ln\prod_{i=1}^{M} p(\mathbf{Y}_{\cdot i}|\mathbf{X}_{\cdot i};\mathbf{E})p(\mathbf{X}_{\cdot i}|\boldsymbol{\alpha}_{\cdot i})\right\rangle_{q(\boldsymbol{\alpha})} + c_0\right\} \tag{12}$$

[3]The KullbackLeibler (KL) divergence between distribution $p(x)$ and $q(x)$ is defined as $\int p(x)\ln\frac{p(x)}{q(x)}d\mathbf{x}$

where c_0 is a constant with respect to $\mathbf{X}$. Let us substitute (4) and (5) into (12). It can be obtained that each $\mathbf{X}_{\cdot i}$ obeys a complex Gaussian distribution as

$$q^*(\mathbf{X}_{\cdot i}) \propto \exp\left[-(\mathbf{X}_{\cdot i}-\boldsymbol{\mu}_{\cdot i})^H\boldsymbol{\Sigma}_i^{-1}(\mathbf{X}_{\cdot i}-\boldsymbol{\mu}_{\cdot i})\right] \tag{13}$$

where $\boldsymbol{\mu}_{\cdot i}$ and $\boldsymbol{\Sigma}_i$ are given by

$$\boldsymbol{\mu}_{\cdot i} = \langle\alpha_0\rangle\,\boldsymbol{\Sigma}_i\mathbf{A}^H\hat{\mathbf{E}}^H\mathbf{Y}_{\cdot i} \tag{14}$$

$$\boldsymbol{\Sigma}_i = (\,\langle\alpha_0\rangle\,\mathbf{A}^H\hat{\mathbf{E}}^H\hat{\mathbf{E}}\mathbf{A}+\mathrm{diag}\,\langle 1./\boldsymbol{\alpha}_{\cdot i}\rangle\,)^{-1}. \tag{15}$$

From the matrix inverse lemma [26], the above covariance matrix could be expressed as $\boldsymbol{\Sigma}_i = \mathrm{diag}\,\langle\boldsymbol{\alpha}_{\cdot i}\rangle - \mathrm{diag}\,\langle\boldsymbol{\alpha}_{\cdot i}\rangle\,\mathbf{A}^H\hat{\mathbf{E}}^H[1/\langle\boldsymbol{\alpha}_0\rangle\,\mathbf{I}+\hat{\mathbf{E}}\mathbf{A}\mathrm{diag}\,\langle\boldsymbol{\alpha}_{\cdot i}\rangle\,\mathbf{A}^H\hat{\mathbf{E}}^H]^{-1}\hat{\mathbf{E}}\mathbf{A}\cdot\mathrm{diag}\,\langle 1./\boldsymbol{\alpha}_{\cdot i}\rangle$. Therefore, computational complexity can be decreased. It is noted that when $\boldsymbol{\alpha}_{ij}$ becomes very small, its corresponding entry in $\mathbf{X}$ approaches zero and will be pruned away.

ii). Updating rule for $\boldsymbol{\alpha}$: The approximated posterior is given by

$$q^*(\boldsymbol{\alpha}) = \tag{16}$$
$$\exp\left\{\left\langle \ln\prod_{i=1}^{M} p(\mathbf{X}_{\cdot i}|\boldsymbol{\alpha}_{\cdot i})p(\boldsymbol{\alpha}_{\cdot i}|\lambda_i)\right\rangle_{\prod_{i=1}^{M} q(\mathbf{X}_{\cdot i})q(\lambda_i)} + c_0\right\}.$$

Substituting (5) and (6) into (16), it is seen that each $\boldsymbol{\alpha}_{ji}$ obeys a generalized inverse Gaussian (GIG) distribution as,

$$q^*(\alpha_{ji}) \propto {\boldsymbol{\alpha}_{ji}}^{\eta-1}\exp\left[-\langle 2\lambda_i\rangle\,\boldsymbol{\alpha}_{ji} - \left\langle 2|\mathbf{x}_{ji}|^2\right\rangle\alpha_{ji}^{-1}\right]. \tag{17}$$

Therefore, the k-th moment of the GIG distribution [33] is given as,

$$\left\langle\alpha_{ji}^k\right\rangle = \left(\frac{\left\langle 2|\mathbf{X}_{ji}|^2\right\rangle}{\langle 2\lambda_i\rangle}\right)^{\frac{k}{2}}\cdot\frac{\kappa_{\eta-1+k}(\sqrt{\langle 2\lambda_i\rangle\left\langle 2|\mathbf{X}_{ji}|^2\right\rangle})}{\kappa_{\eta-1}(\sqrt{\langle 2\lambda_i\rangle\left\langle 2|\mathbf{X}_{ji}|^2\right\rangle})} \tag{18}$$

where κ_a is the modified Bessel function of the second kind. The updating rule for $\langle\boldsymbol{\alpha}_{ji}\rangle$ and $\left\langle\boldsymbol{\alpha}_{ji}^{-1}\right\rangle$ is given in (18) with $k=1$ and $k=-1$, which will be further used to update $\mathbf{X}$ and $\boldsymbol{\lambda}$, respectively.

iii). Updating rule for $\boldsymbol{\lambda}$: The approximated posterior of $\boldsymbol{\lambda}$ can be obtained by

$$q^*(\boldsymbol{\lambda}) = \exp\left\{\left\langle \ln\prod_{i=1}^{M} p(\boldsymbol{\alpha}_{\cdot i}|\lambda_i)p(\lambda_i|v_1,v_2)\right\rangle_{\prod_{i=1}^{M} q(\boldsymbol{\alpha}_{\cdot i})} + c_0\right\}. \tag{19}$$

Substituting (6) and (7) into (19), it is seen that the approximated posterior for λ_i obeys a Gamma distribution due to the prior conjugacy,

$$q(\lambda_i) \propto \lambda_i^{\eta N+v_1-1}\exp\left[-(\sum_{k=1}^{N}\boldsymbol{\alpha}_{ki}+v_2)\lambda_i\right]. \tag{20}$$

The mean of the $\boldsymbol{\lambda}_i, i\in\{1,\cdots,M\}$ is therefore given by

$$\langle\boldsymbol{\lambda}_i\rangle = \frac{\eta N+v_1}{\sum_{k=1}^{N}\boldsymbol{\alpha}_{ki}+v_2}. \tag{21}$$

2) Global Learning Stage: In the global learning stage, the noise precision parameter, α_0, and phase error matrix, $\mathbf{E}$, are derived since they are shared among tasks. Now let us derive the updating rules for α_0 and $\mathbf{E}$, respectively.

i). Updating rule for noise precision α_0 : The approximated posterior is

$$q^*(\alpha_0) = \exp\left\{\langle \ln p(\mathbf{Y}|\mathbf{X},\alpha_0)p(\alpha_0|v_3,v_4)\rangle_{q(\mathbf{X})} + c_0\right\}. \tag{22}$$

The update of α_0 is thus given by the mean of Gamma distribution

$$q^*(\alpha_0) \propto \alpha_0^{MP+v_3-1}\exp \left\{-\left[\left\|\mathbf{Y}-\hat{\mathbf{E}}\mathbf{A}\boldsymbol{\mu}\right\|_F^2 + \sum_{i=1}^{N}\mathrm{trace}(\mathbf{A}^H\hat{\mathbf{E}}^H\hat{\mathbf{E}}\mathbf{A}\boldsymbol{\Sigma}_i) + v_4\right]\alpha_0\right\}. \tag{23}$$

From (23), we can obtain the estimate for α_0 as,

$$\langle\alpha_0\rangle = \frac{MP+v_3}{\left\|\mathbf{Y}-\hat{\mathbf{E}}\mathbf{A}\boldsymbol{\mu}\right\|_F^2 + \sum_{i=1}^{M}\mathrm{trace}(\mathbf{A}^H\hat{\mathbf{E}}^H\hat{\mathbf{E}}\mathbf{A}\boldsymbol{\Sigma}_i) + v_4}. \tag{24}$$

ii). Updating rule for $\mathbf{E}$: The solution to $\mathbf{E}$ can be formulated by minimizing the negative expected log-likelihood function as,

$$\begin{aligned}\hat{\mathbf{E}} &= \arg\min_{\mathbf{E}} \langle -\ln p(\mathbf{Y},\mathbf{X},\boldsymbol{\alpha},\boldsymbol{\lambda};\mathbf{E})\rangle_{q(\mathbf{X})q(\alpha)q(\lambda)} \\ &= \arg\min_{\mathbf{E}} \left\langle \|\mathbf{Y}-\mathbf{E}\mathbf{A}\mathbf{X}\|_2^2\right\rangle_{q(\mathbf{X})}.\end{aligned} \tag{25}$$

It is noted that the above problem is a convex optimization having a closed-form solution [34]. By solving $\frac{\partial\langle\ln p(\mathbf{Y},\mathbf{X},\alpha,\lambda;\mathbf{E})\rangle_{q(\mathbf{X})q(\alpha)q(\lambda)}}{\partial\mathbf{E}} = 0$, as shown in Appendix A.1, we can obtain

$$\hat{\varphi}_i = \arctan\left\{-\frac{\mathrm{Re}\left[\mathbf{Y}_{i\cdot}(\mathbf{A}_{i\cdot}\boldsymbol{\mu})^H\right]}{\mathrm{Im}\left[\mathbf{Y}_{i\cdot}(\mathbf{A}_{i\cdot}\boldsymbol{\mu})^H\right]}\right\}. \tag{26}$$

where $i = 1,...,P$. The corresponding $\mathbf{E}(i,i)$ can be updated by $\mathbf{E}_{i,i} = \exp(j\varphi_i)$ accordingly. This rule is denoted as updating **rule I**, which is also used in the regularized based methods. However, it should be noted that the obtained uncertainty information Σ of estimation $\mathbf{X}$ in (15) does not appear in this updating rule. More concretely, updating rule I only exploits the point estimation of $\mathbf{X}$ to estimate $\mathbf{E}$. In order to properly utilize the uncertainty information, let us propose another method incorporating $\boldsymbol{\Sigma}$ to enhance estimation accuracy. Towards this end, rather than explicitly modeling the phase error as $e^{j\varphi_i}$, the real and imaginary parts of the error are modeled as a_i and b_i, respectively. By introducing these two parameters instead of the angle φ_i, the uncertainty information can be naturally incorporated into the estimation process to achieve enhanced estimation of $\mathbf{E}$ in each iteration. The detailed derivations, shown in Appendix A.2, lead to

$$\hat{\mathbf{E}}_{i,i} = \frac{[\mathbf{Y}_{i\cdot}(\mathbf{A}_{i\cdot}\boldsymbol{\mu})^H]}{\mathrm{trace}(\boldsymbol{\mu}^{\mathrm{H}}\mathbf{A}_{\mathrm{i}\cdot}^{\mathrm{H}}\mathbf{A}_{\mathrm{i}\cdot}\boldsymbol{\mu}) + \sum_{\mathrm{k}=1}^{\mathrm{M}}\mathrm{trace}(\mathbf{A}_{\mathrm{i}\cdot}^{\mathrm{H}}\mathbf{A}_{\mathrm{i}\cdot}\boldsymbol{\Sigma}_{\mathrm{k}})} \tag{27}$$

where $i = 1,...,P$. Thus, (27) is denoted as updating **rule II**. The corresponding algorithms are further denoted by AFSBL1 and AFSBL2, respectively.

Remark. Although the normalization information seems to be lost with this modification, the empirical results validate that this scarification does not degrade the performance. Our experimental results also validate that AFSBL2, which exploits the uncertainty information of the estimation, can achieve better recovery results. Further results and discussions can be found in Section IV. It is worthwhile to point out that AFSBL2 algorithm is quite different from the l_1 alternating method. In addition to performing alternately point estimation, the proposed method uses the statistical information to enhance the estimation performance and avoid converging to a shallow local minimum.

In summary, the proposed autofocus sparse Bayesian learning method is given in Algorithm 1.

C. Algorithm Initialization

Proper initialization is required for this algorithm to obtain desirable results. A reasonable initialization of Algorithm 1 can be set as follows.

1). Each column in $\boldsymbol{\alpha}_{\cdot i}$ is initiated by $1/\mathrm{abs}(\mathbf{A}^H\mathbf{Y}_{\cdot i})$, which is known as projection of $\mathbf{Y}$ onto the space spanned by the columns of $\mathbf{A}$.

2). The phase error matrix $\mathbf{E}$ is initiated as $\mathbf{I}$ since no prior information is available. It is possible to be initialized with other value when the prior distribution is available.

3). The noise precision α_0 is initiated as $1/\mathrm{var}(\mathbf{Y})$.

4). The hyper-parameters are set to $v1 = v2 = v3 = v4 = 10^{-6}$ as suggested in [35].

Algorithm 1 Autofocus Sparse Bayesian Learning

```
Input: Y, A, E, α, v1, v2, v3, v4.
while ∼ Converge do
    I. Individual Learning Stage
    for i = 1 : M do
        Update μ_{·i} and Σ_i by (14) and (15).
        Update α_{·i} by (18)
        Update λ_i by (21)
    end for
    II. Global Learning Stage
    Update noise precision α_0 by (24)
    for i = 1 : P do
        Update E(i,i) by rule I (26) or rule II (27).
    end for
end while
Output: X, E.
```

D. Convergence Analysis

The hidden variables are defined as $\Theta = \{\mathbf{X},\boldsymbol{\alpha},\lambda\}$. We prove that under certain conditions $\mathbf{X}$ and $\mathbf{E}$ will monotonically decrease the KL divergence and the negative expected log likelihood function, respectively, until reaching to a convergence point.

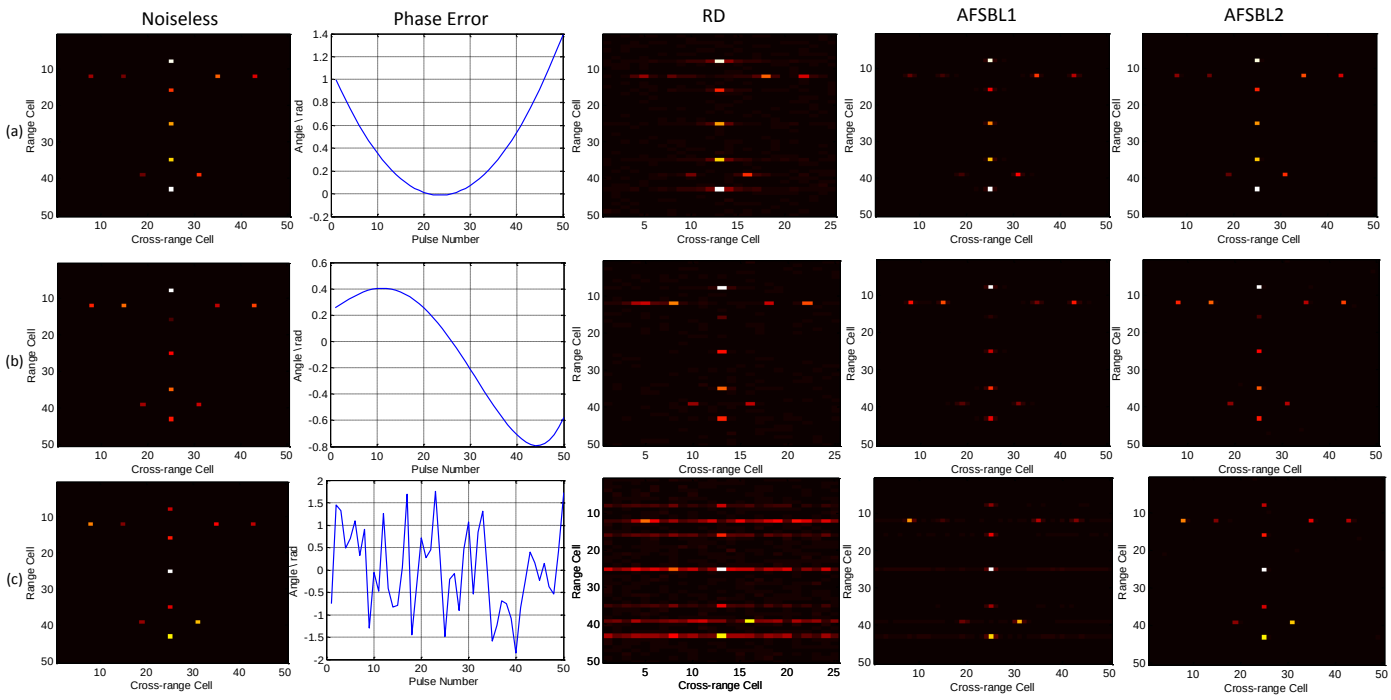

Fig. 3. ISAR imaging with 25 pulses (50% of the measurement and SNR=20dB). (a) Low-order phase error, AFSBL1 ($\mathrm{NMSE}_{\mathbf{X}} = -8.0730$dB, $\mathrm{MSE}_{\varphi} = 0.1042$), AFSBL2 ($\mathrm{NMSE}_{\mathbf{X}} = -14.5165$dB, $\mathrm{MSE}_{\varphi} = 0.0301$), (b) High-order phase error, AFSBL1 ($\mathrm{NMSE}_{\mathbf{X}} = -7.4258$dB, $\mathrm{MSE}_{\varphi} = -0.0952$), AFSBL2 ($\mathrm{NMSE}_{\mathbf{X}} = -12.9717$dB, $\mathrm{MSE}_{\varphi} = -0.0848$), (c) random phase error, AFSBL1 ($\mathrm{NMSE}_{\mathbf{X}} = -4.0294$dB, $\mathrm{MSE}_{\varphi} = 0.4019$), AFSBL2 ($\mathrm{NMSE}_{\mathbf{X}} = -12.4399$dB, $\mathrm{MSE}_{\varphi} = 0.0059$).

Theorem 1: Let $\{\mathbf{X}_{\cdot i}^{n}\}$ and $\{\mathbf{E}_{ii}^{n}\}$ denote the estimation sequence for the sparse scatterer coefficient and phase error, respectively. The sequence $\mathbf{X}_{\cdot i}^{n}$ and $\mathbf{E}_{ii}^{n}$ are guaranteed to converge to a convergence point.

Proof: The proof is given in Appendix B. ■

As indicated in *Theorem* 1, the convergence of the proposed algorithm can be guaranteed. The experimental results also validate that the algorithm will converge within tens of iterations in Section IV. Compared to the convex based alternating strategy, our alternating scheme exploits Bayesian estimation rather than point estimation. It is noteworthy that our proposed algorithm has a smaller probability that the convergence point is a shallow minimum due to the utilization of higher order estimation information compared to other ones.

E. Discussion

Autofocus technique based on sparse signal processing has been also considered in the literature. In [15], [21], the reported algorithm is shown to achieve desirable improvements on auto-focusing. However, it cannot be properly modified to achieve high-resolution imagery. The proposed method, however, is formulated in the sparse Bayesian learning framework to specify the corresponding parameters for both high-resolution imagery scene and phase errors, respectively. Furthermore, all the required parameters in this method are learned from the data directly, avoiding the time-consuming tuning procedures for validating parameters. At the same time, more information including estimation variance is exploited to achieve better estimation.

TABLE I
ISAR SYSTEM PARAMETERS FOR SYNTHETIC DATA

Centroid frequency f_c	10 GHz
Chirp rate γ	3 Hz/s
Repetition frequency f_r	25 Hz
Rotational angular velocity ω	0.1 rad/s
Scatterer center range R_0	5 km
Number of Range cell M	50
Number of pulse N	50

Although this paper only considers the cross-range variant phase error in the context of ISAR imagery, the framework presented above can also be flexibly modified to other types of phase errors encountered in SAR imagery [22], such as range variant phase error. In this case, phase error estimation is required to be updated in the individual learning stage.

IV. Experimental Results

In this section, experimental results by using synthetic and practical data are presented, respectively, to evaluate the performances achieved by the proposed algorithm. Comparisons with other popular methods are also presented.

A. Synthetic Data Experiments

To qualitatively and quantitatively evaluate the performance, the synthetic data are generated with a radar system whose parameters are given in Table I. The imagery size is 50×50 and the number of scatterers is 11. The magnitude of each scatterer obeys $\mathcal{CN}(0, 1)$. The under-sampled data is obtained

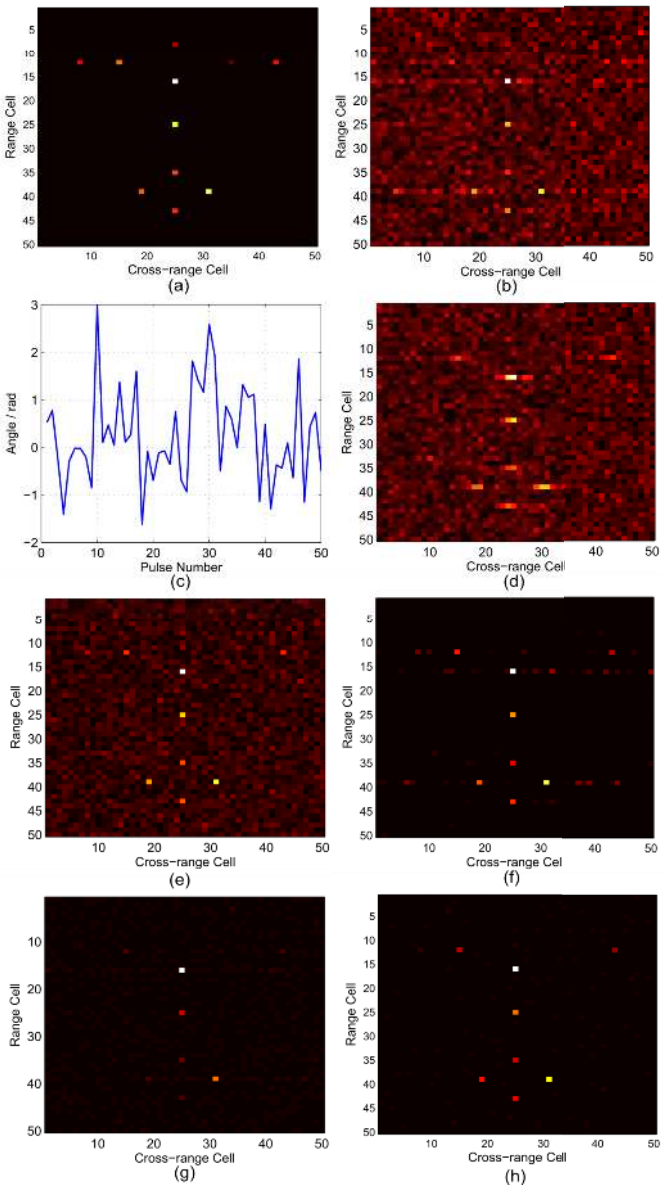

Fig. 4. ISAR imaging with full measurement and SNR=0dB. (a) The true target scene, (b) Image obtained by RD, (c) True phase error, (d) Image obtained by PGA, (e) Image obtained by MEM, (f) Image obtained by $l1$ based method, (g) Image obtained by AFSBL1, (h) Image obtained by AFSBL2.

by randomly selecting the number of pulses from the full aperture data. The noiseless RD imagery with full measurement data is shown in Fig. 3.

The normalized mean square error (NMSE) of the scatterer coefficient estimation is defined as,

$$\mathrm{NMSE}_{\mathbf{X}} = 10\log_{10}\left\|\hat{\mathbf{X}}/|\hat{\mathbf{X}}|_{\max} - \mathbf{X}/|\mathbf{X}|_{\max}\right\|_F^2 \qquad (28)$$

and the mean square error (MSE) of phase error estimation is defined as,

$$\mathrm{MSE}_{\varphi} = \left\|angle(\hat{\mathbf{E}}) - angle(\mathbf{E})\right\|_F^2 \Big/ P. \qquad (29)$$

In the following experiments, 50 iterations are used by MEM [18], l_1 [22], AFSBL1 and AFSBL2 for performance evaluation, where the l_1 based method is implemented with CVX toolbox [36].

TABLE II
$\mathrm{NMSE}_{\mathbf{X}}$ AND MSE_{φ}

SNR		PGA	MEM	l_1	AFSBL1	AFSBL2
10db	$\mathrm{NMSE}_{\mathbf{X}}$	-2.8294	-4.1886	-4.6554	-4.2565	$\mathbf{-10.7487}$
	MSE_{φ}	1.1290	0.0380	0.7855	0.6925	**0.0365**
5db	$\mathrm{NMSE}_{\mathbf{X}}$	4.6718	0.3128	-4.0954	-4.1024	$\mathbf{-9.9212}$
	MSE_{φ}	2.1928	0.0710	0.8398	1.1825	**0.0590**
0db	$\mathrm{NMSE}_{\mathbf{X}}$	8.8631	5.5049	-1.3158	1.0843	$\mathbf{-5.1497}$
	MSE_{φ}	2.8042	0.0894	0.9149	1.2149	**0.0665**

In Fig. 3, an illustrative example is given to evaluate the high-resolution and autofocus performances of updating **rule I** and **rule II** as discussed in Section III-B in terms of different types of phase errors. In Fig. 3(a)-(c), low-order, high-order and random phase errors are tested by the proposed algorithms, AFSBL1 and AFSBL2, respectively. From these figures, it is seen that AFSBL2 can obtain a well focused image, while AFSBL1 achieves less focused image due to the undesirable side-lobe effects. It is also observed from Fig. 3(a) to (c) that the imagery results obtained by both AFSBL1 and RD degrade. However, the images obtained by AFSBL2 are well focused in all these scenarios. It is particularly interesting to observe that in the random phase error scenario, AFSBL2 can give more concentrated image by exploiting estimation covariance. Therefore, despite the form of the phase errors, the AFSBL2 can provide lower $\mathrm{NMSE}_{\mathbf{X}}$ as well as MSE_{φ} due to its inherent ability of utilizing the uncertainty information of estimation $\mathbf{X}$. Since the random phase errors representing the most general case with the worst imaging quality, let us consider the following experiments with the random phase errors.

The proposed algorithm is compared with other conventional algorithms with full measurement data. Figure 4 shows the performance comparison of the proposed algorithm and other popular ones with SNR = 0dB. In this experiment, the SNR is so low that some of the scatterers are covered by the noise, which can be observed from Fig. 4(b). In Fig. 4(d) and (e), it is observed that the image cannot be properly focused by PGA and MEM due to the low SNR. In contrast, as demonstrated in Fig. 4 (f), (g) and (h), the sparsity based methods can obtain the focused image and exhibit a desirable de-noising effect. In Fig. 4 (f), some of the undesirable artifacts are not properly removed by the l_1 based method and almost one half of true scatterers are lost in Fig. 4 (g) by AFSBL1. Among all the imagery results, it is shown that the proposed AFSBL2 can obtain the most focused and concentrated image, though two scatterers are lost.

In Table II, quantitative results, including $\mathrm{NMSE}_{\mathbf{X}}$ and MSE_{φ}, are given to evaluate the performance when SNR = 10, 5 and 0 dB, respectively. We conduct 50 Monte Carlo trails to test the algorithms. The PGA algorithm gives the worst $\mathrm{NMSE}_{\mathbf{X}}$ as well as MSE_{φ}, particularly in low SNR conditions. Although the MEM algorithm gives lower MSE_{φ} than those obtained by l_1 and AFSBL1 methods, the obtained $\mathrm{NMSE}_{\mathbf{X}}$ is much higher than those achieved by these

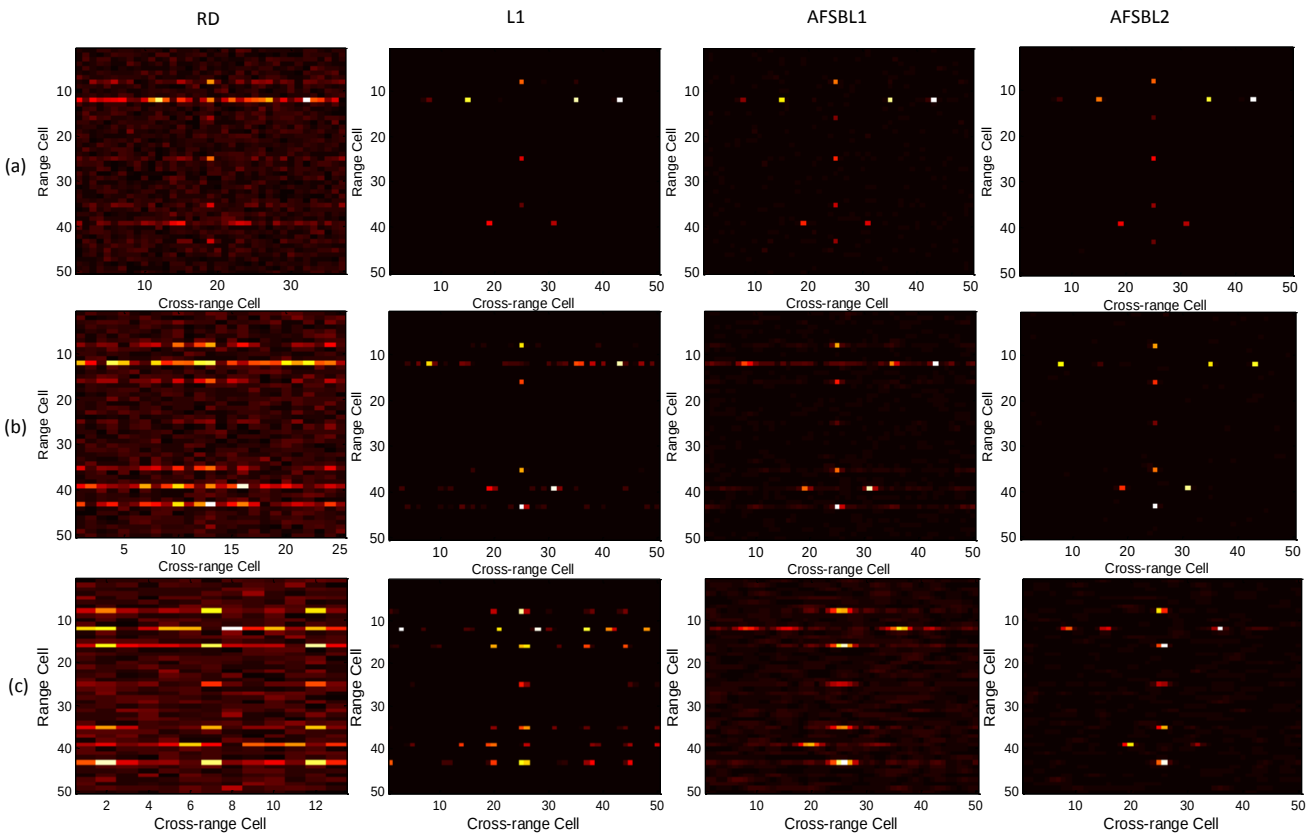

Fig. 5. ISAR imaging results in SNR=10dB. (a) 75% of the measurement, (b) 50% of the measurement, (c) 25% of the measurement.

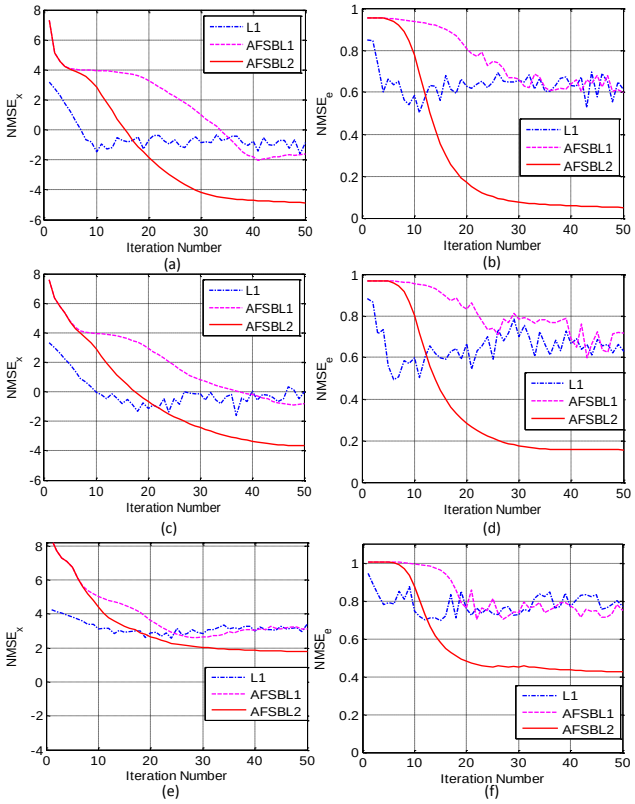

Fig. 6. The convergence of the algorithms in terms of under-sampling ratio. The convergence of the sparse coefficient $\mathbf{X}$ and the corresponding phase error $\mathbf{E}$ with 75%, 50% and 25% are in (a), (c), (e) and (b), (d), (f), respectively.

two algorithms due to the inability of de-noising. For all the tested SNRs, it is seen that the proposed AFSBL2 can obtain lowest $\mathrm{NMSE}_{\mathbf{X}}$ and MSE_{φ}. The difference between the results obtained by AFSBL1 and AFSBL2 also validates that improved estimation has been obtained by exploiting uncertainty information. It is concluded that the proposed AFSBL2 algorithm can obtain more focused image with full measurement compared to other algorithms in various scenarios.

Let us consider how the proposed algorithm deals with under-sampled data for high-resolution and autofocus. In Fig. 5, the proposed algorithm is compared to other algorithms in terms of number of measurements. PGA and MEM are not included in the comparison because they cannot be properly modified to achieve high-resolution imagery. From Fig. 5 (a) to (c), we can observe that the performances degrade as the number of measurements decreases. When 75% or 50% of the measurements are used, all these algorithms can obtain reasonable results although undesirable points are still visible in the image obtained by l_1 and AFSBL1. However, when 25% of the measurement is used, Fig. 5 (c) shows that the image is not properly focused no matter which method is used. However, AFSBL2 can achieve relatively better image with fewer artifact points. In summary, the proposed AFSBL2 algorithm can outperform other methods in various under-sampling scenarios.

Finally, the convergence of the $l1$, AFSBL1 and AFSBL2 is compared with SNR=5dB in Fig. 6. As seen from these figures,

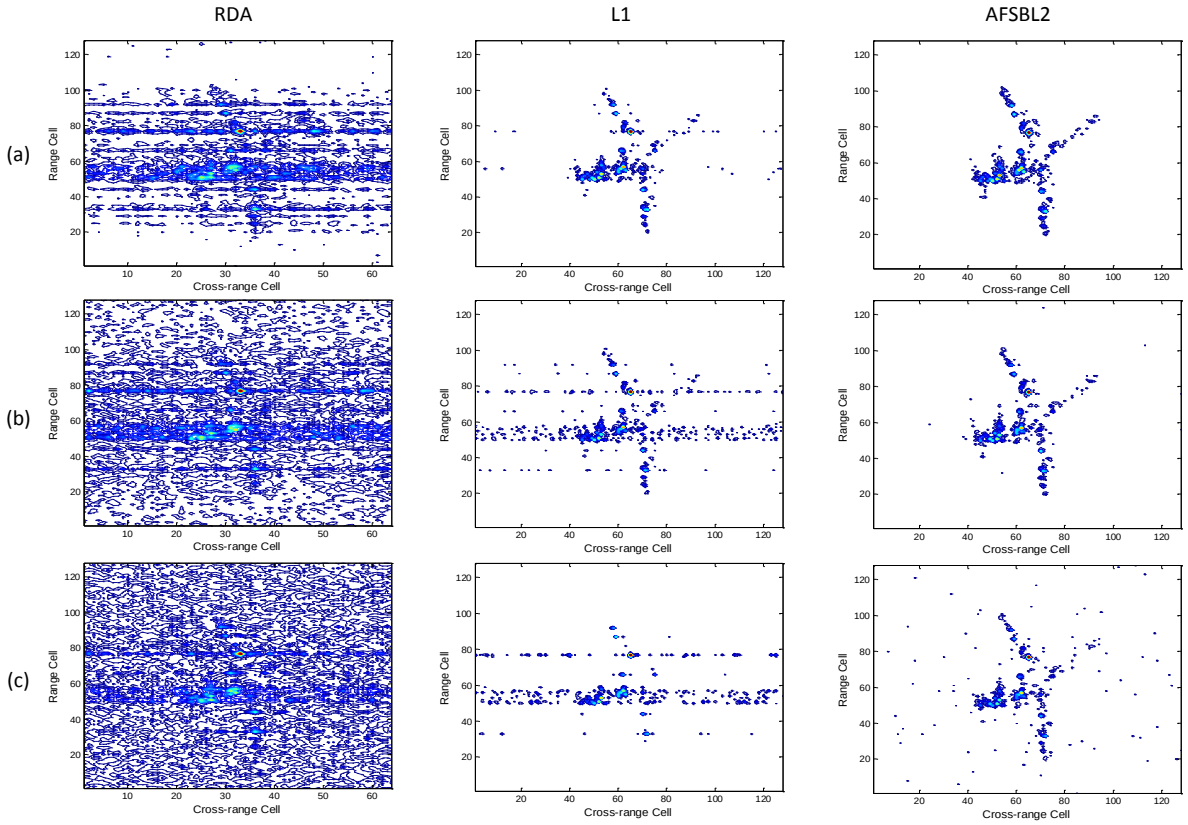

Fig. 8. ISAR imagery results with one half of the measurement with (a) SNR = 15dB, (b) SNR = 10dB, and (c) SNR = 5dB.

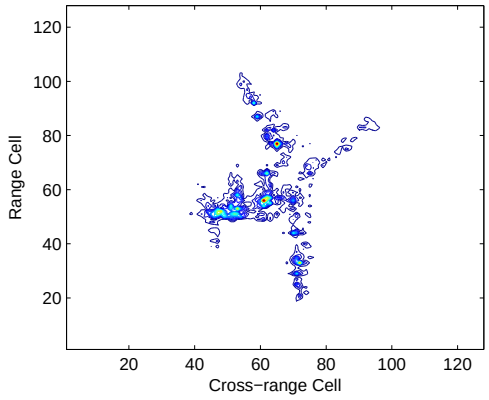

Fig. 7. The noiseless RD imagery of the Yak-42 dataset

the $\text{NMSE}_{\mathbf{X}}$ of the sparse coefficient $\mathbf{X}$ and MSE_{φ} of phase error show that all the algorithms have degraded performances with less measurement. In particular, the convergence of the $l1$ regularized method and AFSBL1 suffer from oscillations, which is also indicated in [21]. In contrast, AFSBL2 can lead to smooth convergence without obvious oscillations, which is obtained by utilizing the uncertainty information. In this experiment, it not only demonstrates the convergence of the algorithm, but also empirically shows that the AFSBL2 can obtain enhanced estimation when the algorithm converges. Therefore, it is concluded that the utilization of the estimation uncertainty information by our particular modeling can greatly benefit the estimation accuracy.

TABLE III
ISAR SYSTEM PARAMETERS FOR YAK-42 DATASET

Centroid frequency f_c	10 GHz
Band width B	400 MHz
Repetition frequency f_r	25 Hz
Number of Range cell M	128
Number of Pulse N	128

B. Practical Data Experiments

The Yak-42 dataset is tested with the proposed and other popular algorithms in various scenarios. The radar system parameters of this dataset are given in Table III. In these experiments, we particularly demonstrate both high-resolution and autofocus results achieved by the proposed algorithm compared with those obtained by l_1 based method. The image size is 128×128 and the noiseless RD imagery result with full measurement is shown in Fig. 7. In this subsection, we will highlight the performances achieved by the proposed AFSBL2 and l_1 methods.

In Fig. 8, the performances are evaluated in terms of various SNR levels. In these figures, the phase error is generated according to a complex Gaussian distribution $\mathcal{CN}(0, 1)$ and the under-sampling ratio is 50%. As observed from Fig. 8(a) and (b), the images obtained by l_1 method show a

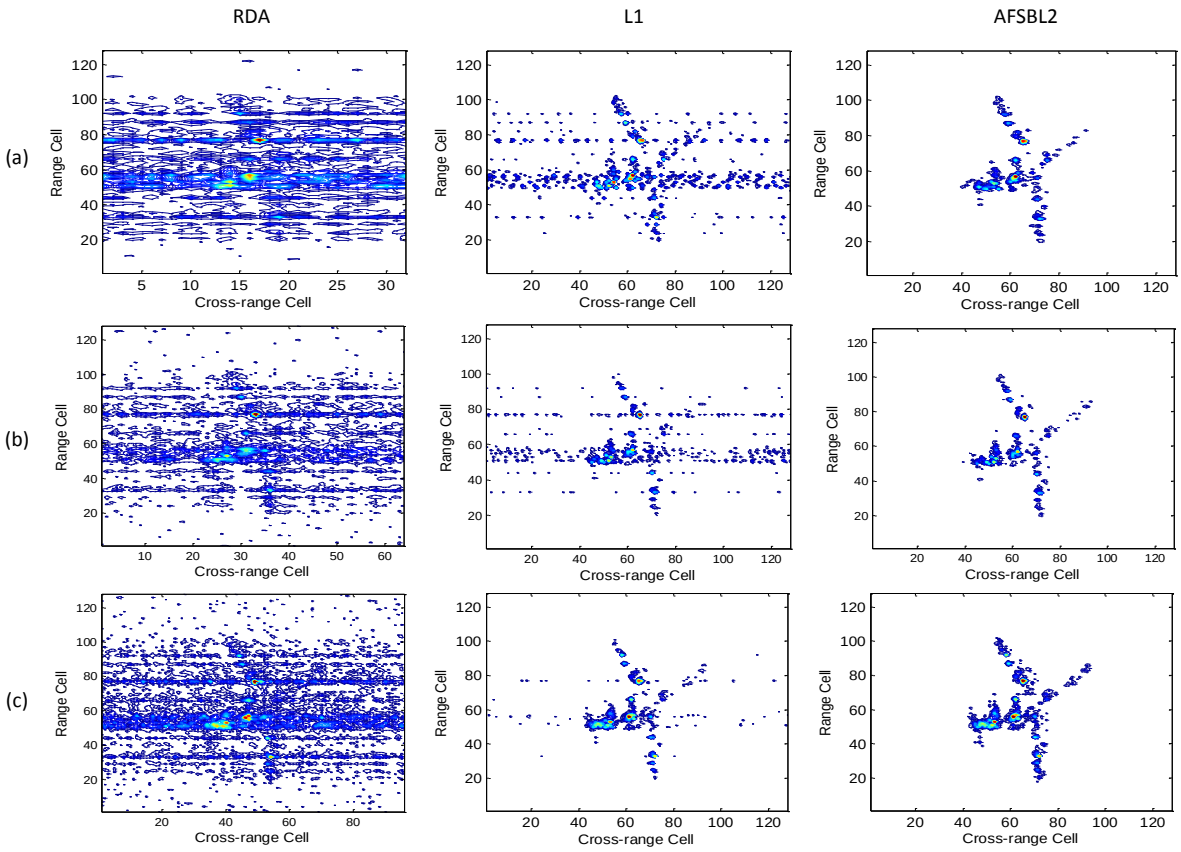

Fig. 9. ISAR imagery results with SNR=8 dB using (a) 25%, (b) 50%, (c) 75%, of the measurements.

relatively reasonable profile of the airplane while the image blurring effects still exist in these images and some of the true scattering points are not recovered. In contrast, AFSBL2 removes most of the undesirable artifacts and obtains better concentration results. When heavier noise is added, the RD image in Fig. 8(c) is almost covered by the noise and the l_1 method can hardly obtain an airplane profile with a limited number of true scatterers. In contrast, the AFSBL2 algorithm can achieve better results by recovering more true scattering points. With these comparisons, it can be concluded that the proposed AFSBL2 algorithm is able to obtain superior imagery results from different SNR scenarios.

In Fig. 9, the performances are compared in terms of different under-sampling ratios. It is quite obvious that more measurement data generally lead to better imagery results. It is reasonable to conclude that both l_1 and AFSBL2 algorithms can effectively estimate the phase errors with large number of measurements. When the number of measurements decreases, however, the l_1 method cannot obtain a good concentration of the target images since they are blurred and out-focused. More importantly, it can be observed that the images obtained by AFSBL2 are much more focused and concentrated in the target region. Remarkably, the results obtained by AFSBL2 do not require any tedious tuning of parameters.

Finally, the computational time is compared. When the number of iterations is set to T, the computational complexities of the proposed and the l_1 based algorithms are in the orders of $O(TMP^3)$ and $O(TMP^2N)$ [37], respectively. In the application of high-resolution imagery, where $P < N$, the proposed method has lower computational complexity. The computational time is measured based on the Matlab code (unoptimized) run on Intel 3.40GHz CPU. In this experiment, the l_1 regularized based method is implemented with the CVX toolbox. However, there are other more efficient ways of handling l_1 optimization such as proximal operator to reduce the computational costs, which is not discussed further. The computational time is compared with different number of measurements. In Fig. 10, it is easy to observe that the proposed algorithm can achieve desirable results with relatively less computational time.

V. Conclusion

In this paper, an autofocus technique for high-resolution ISAR imagery is proposed based on sparse Bayesian method. In the proposed algorithm, a multi-task learning framework is formulated, where sparse scattering coefficients and phase error are updated iteratively in a variational Bayesian inference framework. Benefiting from such hierarchical modeling, the proposed ASFBL2 is able to incorporate uncertainty information in parameter learning. The ASFBL2 approach can not only decrease the possibility of converging to shallow local minimal, but also learn the parameters automatically without labor intensive parameter selection procedures.

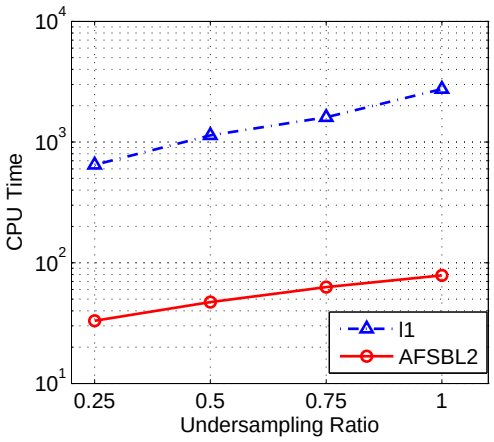

Fig. 10. The CPU time of l_1 and AFSBL2

APPENDIX

A. Derivation of Updating Rule I and II

The derivations of (26) and (27) are presented in this appendix.

1) Proof of (26): The i-th diagonal entry in $\mathbf{E}$ is defined as $e^{j\varphi_i}$. Substituting it into (25), the minimization problem is equivalent to

$$\begin{aligned}\varphi_i &= \arg\min_{\varphi_i} \left\langle \left\| \mathbf{Y}_{i\cdot} - e^{j\varphi_i}\mathbf{A}_{i\cdot}\mathbf{X} \right\|_2^2 \right\rangle_{q(\mathbf{X})} \\ &= \arg\min_{\varphi_i} \left[-e^{-j\varphi_i}\mathbf{Y}_{i\cdot}\boldsymbol{\mu}^H\mathbf{A}_{i\cdot}^H - e^{j\varphi_i}\mathbf{A}_{i\cdot}\boldsymbol{\mu}\mathbf{Y}_{i\cdot}^H \right].\end{aligned} \tag{30}$$

Taking the derivative with respect to φ_i and setting it to zero, we can obtain,

$$\varphi_i = \arctan\left\{ -\frac{\mathrm{Re}\left[\mathbf{Y}_{i\cdot}(\mathbf{A}_{i\cdot}\boldsymbol{\mu})^H\right]}{\mathrm{Im}\left[\mathbf{Y}_{i\cdot}(\mathbf{A}_{i\cdot}\boldsymbol{\mu})^H\right]} \right\}. \tag{31}$$

The updating **rule I** is obtained by $\mathbf{E}(i,i) = \exp(j\varphi_i)$.

2) Proof of (27): Let us now define the i-th diagonal entry in $\mathbf{E}$ as $a_i + jb_i$. Substituting it into (25), the minimization problem is equivalent to

$$\begin{aligned}\mathbf{E}(i,i) &= \arg\min_{a_i+jb_i} \left\langle \left\| \mathbf{Y}_{i\cdot} - (a_i+jb_i)\mathbf{A}_{i\cdot}\mathbf{X} \right\|_2^2 \right\rangle_{q(\mathbf{X})} \\ &= \arg\min_{\mathbf{E}(i,i)} -(a_i+jb_i)\mathbf{Y}_{i\cdot}\boldsymbol{\mu}^H\mathbf{A}_{i\cdot}^H - (a_i-jb_i) \\ &\quad \cdot\mathbf{A}_{i\cdot}\boldsymbol{\mu}\mathbf{Y}_{i\cdot}^H + (a_i^2+b_i^2)\cdot\mathrm{trace}(\boldsymbol{\mu}^H\mathbf{A}_{i\cdot}^H\mathbf{A}_{i\cdot}\boldsymbol{\mu}) \\ &\quad +(a_i^2+b_i^2)\cdot\sum_{k=1}^{M}\mathrm{trace}(\mathbf{A}_{i\cdot}^H\mathbf{A}_{i\cdot}\boldsymbol{\Sigma}_k).\end{aligned} \tag{32}$$

Taking the derivative with respect to a_i and b_i and setting them to zeros, we can obtain

$$a_i = \frac{\mathrm{Re}[\mathbf{Y}_{i\cdot}(\mathbf{A}_{i\cdot}\boldsymbol{\mu})^H]}{\mathrm{trace}(\boldsymbol{\mu}^{\mathrm{H}}\mathbf{A}_{\mathrm{i}\cdot}^{\mathrm{H}}\mathbf{A}_{\mathrm{i}\cdot}\boldsymbol{\mu}) + \sum_{\mathrm{k}=1}^{\mathrm{M}}\mathrm{trace}(\mathbf{A}_{\mathrm{i}\cdot}^{\mathrm{H}}\mathbf{A}_{\mathrm{i}\cdot}\boldsymbol{\Sigma}_{\mathrm{k}})} \tag{33}$$

$$b_i = \frac{\mathrm{Im}[\mathbf{Y}_{i\cdot}(\mathbf{A}_{i\cdot}\boldsymbol{\mu})^H]}{\mathrm{trace}(\boldsymbol{\mu}^{\mathrm{H}}\mathbf{A}_{\mathrm{i}\cdot}^{\mathrm{H}}\mathbf{A}_{\mathrm{i}\cdot}\boldsymbol{\mu}) + \sum_{\mathrm{k}=1}^{\mathrm{M}}\mathrm{trace}(\mathbf{A}_{\mathrm{i}\cdot}^{\mathrm{H}}\mathbf{A}_{\mathrm{i}\cdot}\boldsymbol{\Sigma}_{\mathrm{k}})}. \tag{34}$$

The updating **rule II** is obtained by $\mathbf{E}(i,i) = a_i + jb_i$.

B. Proof of Theorem 1

Firstly, the convergence of $\{\mathbf{X}_{\cdot i}^n\}$ is proved. It can monotonically decrease the KL divergence of the approximate and the true joint distribution. It can be proved that the new approximated posterior $q(\mathbf{X}_{\cdot i})$ will necessarily decrease the KL divergence of $D_{KL}\left[q(\mathbf{X}_{\cdot i})||p(\mathbf{Y},\mathbf{X}_{\cdot i},\alpha_{\cdot i},\hat{\mathbf{E}})\right]$ [29].

Secondly, let us prove that the phase error sequence $\{\mathbf{E}^n\}$ increases the lower bound of log-likelihood function. Define the log-likelihood function as $L(\mathbf{E}^n) = \ln \mathrm{p}(\mathbf{Y};\mathbf{E}^n)$, where $\mathbf{E}$ is the phase error matrix. The convergence of the sequence $\{\mathbf{E}^n\}$ is proved as follows. The expected log-likelihood function of $\mathbf{E}^{n+1}$ can be expressed as

$$\mathbf{E}^{n+1} = \arg\max_{\mathbf{E}} \int q(\mathbf{X})\ln\left(p(\mathbf{Y}|\mathbf{X};\mathbf{E})\right)d\mathbf{X}. \tag{35}$$

Maximizing (35) is equivalent to maximize the following one by introducing some constant terms with respect to $\mathbf{E}^{n+1}$,

$$\int q(\mathbf{X})\ln\left(p(\mathbf{Y}|\mathbf{X};\mathbf{E})\right)d\mathbf{X} - \ln p(\mathbf{Y};\mathbf{E}^n). \tag{36}$$

Due to the non-negativity of the KL divergence, the above equation is upper bounded by

$$\begin{aligned}&\ln \int p(\mathbf{Y},\mathbf{X};\mathbf{E}^{n+1})p(\mathbf{X})d\mathbf{X} - \ln p(\mathbf{Y};\mathbf{E}^n) \\ &= \ln p(\mathbf{Y};\mathbf{E}^{n+1}) - \ln p(\mathbf{Y};\mathbf{E}^n) \\ &= L(\mathbf{E}^{n+1}) - L(\mathbf{E}^n).\end{aligned} \tag{37}$$

From (1), it is seen that the maximization of the expected log-likelihood function in the global stage does not decrease the lower bound of the expected log-likelihood function. In other words, the sequence $\mathbf{E}_{ii}^n$ is guaranteed to increase the lower bound of log-likelihood, thus resulting in non-decreasing log-likelihood.

In summary, the sequence $\{\mathbf{X}_{\cdot i}^n\}$ decreases the KL divergence and $\{\mathbf{E}_{ii}^n\}$ decreases the negative expected likelihood until convergence.

REFERENCES

[1] C. Chen and H.C. Andrews, "Target-motion-induced radar imaging," *IEEE Transactions on Aerospace and Electronic Systems*, , no. 1, pp. 2–14, 1980.

[2] L. Zhang, M. Xing, C. Qiu, J. Li, J. Sheng, Y. Li, and Z. Bao, "Resolution enhancement for inversed synthetic aperture radar imaging under low SNR via improved compressive sensing," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 48, no. 10, pp. 3824–3838, 2010.

[3] F. Berizzi, E. Mese, M. Diani, and M. Martorella, "High-resolution ISAR imaging of maneuvering targets by means of the range instantaneous doppler technique: modeling and performance analysis," *IEEE Transactions on Image Processing*, vol. 10, no. 12, pp. 1880–1890, 2001.

[4] M.A. Herman and T. Strohmer, "High-resolution radar via compressed sensing," *IEEE Transactions on Signal Processing*, vol. 57, no. 6, pp. 2275–2284, Jun. 2009.

[5] L. Wang, L. Zhao, G. Bi, C. Wan, and L. Yang, "Enhanced ISAR imaging by exploiting the continuity of the target scene," *to appear in IEEE Transactions on Geoscience and Remote Sensing*.

[6] S.S. Chen, D.L. Donoho, and M.A. Saunders, "Atomic decomposition by basis pursuit," *SIAM Review*, vol. 43, no. 1, pp. 129–159, Jan. 2001.

[7] D.L. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289 –1306, Apr. 2006.

[8] E.J. Candes and M.B. Wakin, "An introduction to compressive sampling," *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 21–30, Mar. 2008.
[9] L. Zhao, G. Bi, L. Wang, and H. Zhang, "An improved auto-calibration algorithm based on sparse bayesian learning framework," *IEEE Signal Processing Letters*, vol. 20, no. 9, pp. 889–892, 2013.
[10] M. Çetin and W. C. Karl, "Feature-enhanced synthetic aperture radar image formation based on nonquadratic regularization," *IEEE Transactions on Image Processing*, vol. 10, no. 4, pp. 623–631, 2001.
[11] M. Tello Alonso, P. Lopez-Dekker, and J.J. Mallorqui, "A novel strategy for radar imaging based on compressive sensing," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 48, no. 12, pp. 4285–4295, 2010.
[12] L. Zhang, Z. Qiao, M. Xing, J. Sheng, R. Guo, and Z. Bao, "High-resolution ISAR imaging by exploiting sparse apertures," *IEEE Transactions on Antennas and Propagation*, vol. 60, no. 2, pp. 997–1008, Feb. 2012.
[13] J. Xu, Y. Pi, and Z. Cao, "Bayesian compressive sensing in synthetic aperture radar imaging," *IET Radar, Sonar Navigation*, vol. 6, no. 1, pp. 2–8, 2012.
[14] W. Ye, T. Yeo, and Z. Bao, "Weighted least-squares estimation of phase errors for SAR/ISAR autofocus," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 37, no. 5, pp. 2487–2494, 1999.
[15] G. Xu, M. Xing, L. Zhang, Y. Liu, and Y. Li, "Bayesian inverse synthetic aperture radar imaging," *IEEE Geoscience and Remote Sensing Letters*, vol. 8, no. 6, pp. 1150–1154, 2011.
[16] D. Wahl, P. Eichel, D. Ghiglia, and C. Jakowatz, "Phase gradient autofocus-a robust tool for high resolution sar phase correction," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 30, no. 3, pp. 827–835, 1994.
[17] L. Yang, M. Xing, Y. Wang, L. Zhang, and Z. Bao, "Compensation for the NsRCM and phase error after polar format resampling for airborne spotlight sar raw data of high resolution," *IEEE Geoscience and Remote Sensing Letters*, vol. 10, no. 1, pp. 165–169, 2013.
[18] X. Li, G. Liu, and J. Ni, "Autofocusing of ISAR images based on entropy minimization," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 35, no. 4, pp. 1240–1252, 1999.
[19] J. Wang, X. Liu, and Z. Zhou, "Minimum-entropy phase adjustment for ISAR," *IEE Proceedings- Radar Sonar and Navigation*, vol. 151, no. 4, pp. 203–209, 2004.
[20] L. Yang, M. Xing, L. Zhang, J. Sheng, and Z. Bao, "Entropy-based motion error correction for high-resolution spotlight sar imagery," *IET Radar, Sonar Navigation*, vol. 6, no. 7, pp. 627–637, 2012.
[21] X. Du, C. Duan, and W. Hu, "Sparse representation based autofocusing technique for ISAR images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 51, no. 3, pp. 1826–1835, 2013.
[22] N. Onhon and M. Cetin, "A sparsity-driven approach for joint sar imaging and phase error correction," *IEEE Transactions on Image Processing*, vol. 21, no. 4, pp. 2075–2088, 2012.
[23] S. Ji, D. Dunson, and L. Carin, "Multitask compressive sensing," *IEEE Transactions on Signal Processing*, vol. 57, no. 1, pp. 92–106, 2009.
[24] H. Wu, D. Grenier, G. Y Delisle, and D. Fang, "Translational motion compensation in ISAR image processing," *IEEE Transactions on Image Processing*, vol. 4, no. 11, pp. 1561–1571, 1995.
[25] M. Tipping, "Sparse bayesian learning and the relevance vector machine," *Journal of Machine Learning Research*, vol. 1, pp. 211–244, Sep. 2001.
[26] D. Wipf and B.D. Rao, "Sparse bayesian learning for basis selection," *IEEE Transactions on Signal Processing*, vol. 52, no. 8, pp. 2153–2164, Aug. 2004.
[27] S. Ji, Y. Xue, and L. Carin, "Bayesian compressive sensing," *IEEE Transactions on Signal Processing*, vol. 56, no. 6, pp. 2346–2356, Jun. 2008.
[28] N. L. Pedersen, D. Shutin, C. N. Manchón, and B. H. Fleury, "Sparse estimation using bayesian hierarchical prior modeling for real and complex models," *arXiv preprint arXiv:1108.4324*, 2011.
[29] C.M. Bishop et al., *Pattern recognition and machine learning*, vol. 1, springer New York, 2006.
[30] C. P. Robert and G. Casella, *Monte Carlo statistical methods*, vol. 319, Citeseer, 2004.
[31] D.G. Tzikas, A.C. Likas, and N.P. Galatsanos, "The variational approximation for bayesian inference," *IEEE Signal Processing Magazine*, vol. 25, no. 6, pp. 131–146, Nov. 2008.
[32] M. J. Beal, *Variational algorithms for approximate Bayesian inference*, Ph.D. thesis, University of London, 2003.
[33] B. Jørgensen, *Statistical properties of the generalized inverse Gaussian distribution*, Springer-Verlag, New York, 1982.
[34] S. Boyd and L. Vandenberghe, *Convex optimization*, Cambridge university press, 2004.
[35] S.D. Babacan, R. Molina, and A.K. Katsaggelos, "Bayesian compressive sensing using laplace priors," *IEEE Transaction on Image Processing*, vol. 19, no. 1, pp. 53–63, Jan. 2010.
[36] M. Grant, S. Boyd, and Y. Ye, "CVX: Matlab software for disciplined convex programming," 2008.
[37] P. Bühlmann and L. Meier, "Discussion: One-step sparse estimates in nonconcave penalized likelihood models," *The Annals of Statistics*, vol. 36, no. 4, pp. 1534–1541, 2008.

Lifan Zhao received the B.S. degree in electronic engineering from Xidian University, Xi'an, China, in 2010. He is currently working towards the Ph.D. degree in the School of Electrical and Electronic Engineering from Nanyang Technological University, Singapore.

His research interests include sparse signal recovery techniques and their applications in source localization and radar imagery.

Lu Wang received the B.S. in electrical and electronic engineering from Xidian University in 2007 and M.Eng. degree in signal processing from National Key Lab of Radar Signal Processing (RSP), Xidian University, China, in 2010. She is currently pursuing the Ph.D. degree in the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore. Her major research interests include sparse signal processing, radar imaging (SAR/ISAR), sonar signal processing.

Guoan Bi received the B.Sc. degree in radio communications from Dalian University of Technology, P. R. China, 1982, and the M.Sc. degree in telecommunication systems, and the Ph.D. degree in electronics systems, both from Essex University, U.K., in 1985 and 1988, respectively. Since 1991, he has been with the school of Electrical and Electronic Engineering, Nanyang Technological University, Singapore. His current research interests include DSP algorithms and hardware structures, and signal processing for various applications including sonar, radar, and communications.

Lei Yang was born in Tianjing City, China, 1984. He received the B.S. degree in electrical engineering from Xidian University in 2007 and the Ph. D degree in Signal Processing at National Key Lab of Radar Signal Processing of Xidian University in 2012. His major research interests are SAR/ISAR imaging and motion compensation. From 2012, he was with the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore