A new sparse Bayesian learning method for inverse synthetic aperture radar imaging via exploiting cluster patterns

Jun Fang, Lizao Zhang, Huiping Duan, Lei Huang, Hongbin Li

Jun Fang, Lizao Zhang, Huiping Duan, Lei Huang, Hongbin Li, "A new sparse Bayesian learning method for inverse synthetic aperture radar imaging via exploiting cluster patterns," Proc. SPIE 9857, Compressive Sensing V: From Diverse Modalities to Big Data Analytics, 98570D (4 May 2016); doi: 10.1117/12.2225157

SPIE.

Event: SPIE Commercial + Scientific Sensing and Imaging, 2016, Baltimore, Maryland, United States

A New Sparse Bayesian Learning Method for Inverse Synthetic Aperture Radar Imaging via Exploiting Cluster Patterns

Jun Fang[a], Lizao Zhang[a], Huiping Duan[b], Lei Huang[c] and Hongbin Li[d]

[a]National Key Laboratory of Science and Technology on Communications, University of Electronic Science and Technology of China, Chengdu 611731, China; [b]School of Electronic Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China; [c]Department of Electronic and Information Engineering, Shenzhen University, Shenzhen, China; [d]Department of Electrical and Computer Engineering, Stevens Institute of Technology, Hoboken, NJ 07030 USA

ABSTRACT

The application of sparse representation to SAR/ISAR imaging has attracted much attention over the past few years. This new class of sparse representation based imaging methods present a number of unique advantages over conventional range-Doppler methods, the basic idea behind these works is to formulate SAR/ISAR imaging as a sparse signal recovery problem. In this paper, we propose a new two-dimensional pattern-coupled sparse Bayesian learning(SBL) method to capture the underlying cluster patterns of the ISAR target images. Based on this model, an expectation-maximization (EM) algorithm is developed to infer the maximum a posterior (MAP) estimate of the hyperparameters, along with the posterior distribution of the sparse signal. Experimental results demonstrate that the proposed method is able to achieve a substantial performance improvement over existing algorithms, including the conventional SBL method.

Keywords: Pattern-coupled sparse bayesian learning, Cluster patterns, ISAR

1. INTRODUCTION

Sparse representation provides a new perspective and methodology for advanced image formation. The application of this new methodology to SAR/ISAR imaging has attracted much attention over the past few years, e.g.[1–6] This new class of advanced imaging methods present a number of unique advantages over conventional range-Doppler methods,[7] including improved resolvability of point scatterers, reduced speckle, and recoverability from fewer data samples. The basic idea behind these works is to formulate SAR/ISAR imaging as a sparse signal recovery problem by exploiting the sparsity in the reflectivity of objects of interest. In addition to the sparse structure, real-world SAR/ISAR images often have additional structures that can be utilized. For example, in,[4] by exploiting the continuity structure of the target image, the proposed method demonstrates substantial improvements compared with other reported algorithms for ISAR imaging. Also, in practice, the target of interest usually exhibits block-sparse structures in which nonzero large scatterers occur in clusters. This block-sparse pattern can be considered as a special form of the continuity structure where nonzero scatterers demonstrate continuity in both the range and cross-range domains.

This paper aims at leveraging the block-sparse structure of the target image to enhance ISAR imaging performance. To this objective, we introduce a two-dimensional pattern-coupled hierarchical Gaussian prior to model the pattern dependencies among neighboring scatterers on the target scene. Based on this hierarchical

Further author information:
Jun Fang: E-mail: JunFang@uestc.edu.cn
Lizao Zhang: E-mail: 201321260134@std.uestc.edu.cn
Huiping Duan: E-mail: huipingduan@uestc.edu.cn
Lei Huang: E-mail: dr.lei.huang@ieee.org
Hongbin Li: E-mail: Hongbin.Li@stevens.edu

Compressive Sensing V: From Diverse Modalities to Big Data Analytics, edited by Fauzia Ahmad,
Proc. of SPIE Vol. 9857, 98570D · © 2016 SPIE · CCC code: 0277-786X/16/$18 · doi: 10.1117/12.2225157

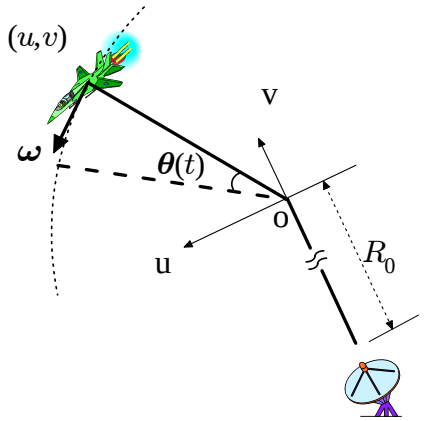

Figure 1. ISAR model.

model, an expectation-maximization (EM) strategy is employed to obtain the maximum a posterior (MAP) estimate of the hyperparameters, along with the posterior distribution of the sparse target signal. Experimental results demonstrate that the proposed algorithm achieves a considerable performance improvement over other existing algorithms.

2. ISAR IMAGING MODEL

ISAR imaging can be viewed as a linear inverse problem in which the unknown reflectivity field is reconstructed from noisy measurements of waves backscattered from a target scene. Suppose the radar transmits a linear frequency-modulated signal

$$s(\tau) = \text{rect}(\tau/T_p) \exp\left\{j2\pi\left(f_c\tau + \frac{\gamma}{2}\tau^2\right)\right\} \tag{1}$$

where τ is the fast time, T_p is the pulse duration, and $\text{rect}(\cdot)$ represents the unit rectangular function, f_c is the carrier frequency, and γ is the chirp rate. Consider the signal reflected from a single scatterer located at position (u, v) in the cross-range and range domains. After range compression, the echoed signal is given as

$$r(\tau, t) = x \cdot \text{rect}(t/T_a)\text{sinc}\left(T_p\gamma[\tau - 2R(t)/c]\right) \exp\left(-j4\pi R(t)/\lambda\right) \tag{2}$$

where x denotes the scattering amplitude, t is the slow time, T_a is the coherent processing interval, $\text{sinc}(\cdot)$ is the sinc function, c denotes the speed of light, λ is the wavelength, and $R(t)$ is the instantaneous distance between the scatterer and the radar which can be approximated by $R(t) \approx R_0 + v + u\omega t$, where R_0 is the target coordinate origin, and ω denotes the rotational angular velocity (see Fig. 1). Substituting $R(t)$ into (2), we arrive at

$$r(\tau, t) \approx x \cdot \text{rect}(t/T_a)\text{sinc}\left(T_p\gamma[\tau - 2(R_0 + v)/c]\right) \exp\left(-j4\pi(R_0 + v)/\lambda\right) \exp(-j2\pi f t) \tag{3}$$

where $f \triangleq 2u\omega/\lambda$ is the Doppler frequency.

To model the signal backscattered from a target scene, we discretize the target scene into an $M \times N$ grid in the cross-range and range domains, with each scatterer located at grid (m, n), and its position and scattering amplitude given respectively by (u_m, v_n) and $x_{m,n}$. In other words, we assume that each range cell consists of M scatterers with different cross-range locations. Hence the signal in the range cell corresponding to $\tau_n = 2(R_0 + v_n)/c$ is a superposition of the echoed signals from the M scatterers

$$y_n(t) = \sum_{m=1}^{M} x_{m,n}\text{rect}(t/T_a) \exp(-j2\pi f_m t) + w_n(t) \quad \forall n \tag{4}$$

where $f_m = 2u_m\omega/\lambda$, $w_n(t)$ denotes the additive noise, and the constant phase term $\exp(-j4\pi(R_0+v_n)/\lambda)$ is neglected since it can be absorbed into the scattering amplitude. By sampling the time series: $t_l = l\Delta t, \forall l = 1,\ldots,L$, where Δt is the pulse repetition period, and $L \triangleq T_a/\Delta t$ is the total number of pulses, the sampled signal $y_n(t_l)$ can be expressed compactly in a vector form as follows

$$\boldsymbol{y}_n = \boldsymbol{F}\boldsymbol{x}_n + \boldsymbol{w}_n \quad \forall n \tag{5}$$

where $\boldsymbol{y}_n \triangleq [y_n(t_1) \ \ldots \ y_n(t_L)]^T$, $\boldsymbol{x}_n \triangleq [x_{1,n} \ \ldots \ x_{M,n}]^T$, $\boldsymbol{w}_n \triangleq [w_n(t_1) \ \ldots \ w_n(t_L)]^T$, and $\boldsymbol{F}$ is an $L \times M$ $(L < M)$ matrix with its (l,m)th entry given by $\exp(-j2\pi f_m t_l)$. Putting together signals in different range cells, we obtain

$$\boldsymbol{Y} = \boldsymbol{F}\boldsymbol{X} + \boldsymbol{W} \tag{6}$$

in which $\boldsymbol{Y} \triangleq [\boldsymbol{y}_1 \ \ldots \ \boldsymbol{y}_N]$, $\boldsymbol{X} \triangleq [\boldsymbol{x}_1 \ \ldots \ \boldsymbol{x}_N]$, and $\boldsymbol{W} \triangleq [\boldsymbol{w}_1 \ \ldots \ \boldsymbol{w}_N]$. We assume that $\{\boldsymbol{w}_n\}$ is the additive multivariate Gaussian noise with zero mean and covariance matrix $(1/\gamma)\boldsymbol{I}$. The problem of interest here is recover the unknown reflectivity field $\boldsymbol{X} = \{x_{m,n}\}$ from the noise-corrupted observed data $\boldsymbol{Y}$. Since the underlying reflectivity field of the target contains a very few strong scatterers, reconstructing the image is in fact a sparse signal recovery problem.

3. BAYESIAN MODEL

In addition to the sparsity, the underlying reflectivity field $\boldsymbol{X}$ usually exhibits two-dimensional block-sparse structures in which nonzero large scatterers occur in clusters. To exploit this cluster pattern, we introduce a pattern-coupled sparse Bayesian learning (SBL) model which is a generalization of our previous model[8] to the two-dimensional case. Before proceeding, we provide a brief review of the conventional SBL model.[9]

3.1 Review of Hierarchical Models for SBL

In the conventional SBL framework,[9] a two-layer hierarchical Gaussian prior was introduced to promote the sparsity of the solution. In the first layer, coefficients $\{x_{m,n}\}$ are assigned a Gaussian prior distribution

$$p(\boldsymbol{X}|\boldsymbol{\alpha}) = \prod_{m=1}^{M}\prod_{n=1}^{N}\mathcal{N}(x_{m,n}|0,\alpha_{m,n}^{-1}) \tag{7}$$

where $\alpha_{m,n}$ is a non-negative hyperparameter controlling the sparsity of the coefficient $x_{m,n}$. The second layer specifies Gamma distributions as hyperpriors over the hyperparameters $\boldsymbol{\alpha} \triangleq \{\alpha_{m,n}\}$, i.e.

$$p(\boldsymbol{\alpha}) = \prod_{m=1}^{M}\prod_{n=1}^{N}\text{Gamma}(\alpha_{m,n}|a,b) \tag{8}$$

As discussed in,[9] for properly chosen a and b, this hyperprior allows the posterior mean of $\alpha_{m,n}$ to become arbitrarily large. As a consequence, the associated coefficient $x_{m,n}$ will be driven to zero, thus yielding a sparse solution. This conventional hierarchical model, however, has no potential to encourage structured-sparse solutions since the sparsity of each coefficient is determined by its own hyperparameter. In,[10,11] the above hierarchical model was generalized to deal with block-sparse signals, in which a group of coefficients sharing the same sparsity pattern are assigned a multivariate Gaussian prior parameterized by a common hyperparameter. Nevertheless, this model requires the knowledge of the block partition to determine which coefficients should be grouped and assigned a common hyperparameter.

3.2 Proposed Pattern-Coupled Hierarchical Model

To exploit the block-sparse structure inherent in the target image, we utilize the fact that the sparsity patterns of neighboring coefficients are statistically dependent. To capture the pattern dependencies among neighboring

coefficients, the Gaussian prior for each coefficient $x_{m,n}$ not only involves its own hyperparameter $\alpha_{m,n}$, but also its immediate neighbor hyperparameters. Specifically, a prior over $\boldsymbol{X}$ is given by

$$p(\boldsymbol{X}|\boldsymbol{\alpha}) = \prod_{m=1}^{M}\prod_{n=1}^{N} \mathcal{CN}(x_{m,n}|0, \delta_{m,n}^{-1}) \tag{9}$$

where $\mathcal{CN}(\cdot)$ denotes the complex Gaussian distribution, and

$$\delta_{m,n} \triangleq \alpha_{m,n} + \beta \sum_{(i,j)\in N_{(m,n)}} \alpha_{i,j} \tag{10}$$

in which $N_{(m,n)}$ denotes the neighborhood of the grid point (m,n), i.e. $N_{(m,n)} \triangleq \{(m,n-1),(m,n+1),(m-1,n),(m+1,n)\}^*$, and $\beta \in [0,1]$ is a parameter indicating the pattern relevance between the coefficient $\alpha_{m,n}$ and its neighboring coefficients. When $\beta = 0$, the prior model (9) reduces to the conventional sparse Bayesian learning model. When $\beta > 0$, we see that the sparsity of each coefficient $x_{m,n}$ is not only controlled by the hyperparameter $\alpha_{m,n}$, but also by the neighboring hyperparameters $S_{\alpha_{m,n}} \triangleq \{\alpha_{i,j}|(i,j) \in N_{(m,n)}\}$. The coefficient $x_{m,n}$ will be driven to zero if $\alpha_{m,n}$ or any of its neighboring hyperparameters goes to infinity. In other words, suppose $\alpha_{m,n}$ approaches infinity, then not only its corresponding coefficient $x_{m,n}$ will be driven to zero, the neighboring coefficients $S_{x_{m,n}} \triangleq \{x_{i,j}|(i,j) \in N_{(m,n)}\}$ will decrease to zero as well. We see that the sparsity patterns of neighboring coefficients are related with each other through their shared hyperparameters. On the other hand, for any pair of neighboring coefficients, each of them has its own hyperparameters that are not shared by the other coefficient. It means that no coefficients are pre-specified to share a common sparsity pattern, which enables the prior to provide flexibility to model any block-sparse structures.

Following,[9] we use Gamma distributions as hyperpriors over the hyperparameters $\{\alpha_{m,n}\}$, i.e.

$$p(\boldsymbol{\alpha}) = \prod_{m=1}^{M}\prod_{n=1}^{N} \text{Gamma}(\alpha_{m,n}|a,b) \tag{11}$$

where we set $a > 1$, and $b = 10^{-6}$. Also, the noise variance $1/\gamma$ is assumed unknown, and to estimate this parameter, we place a Gamma hyperprior over γ, i.e.

$$p(\gamma) = \text{Gamma}(\gamma|c,d) \tag{12}$$

where we set $c = 1$ and $d = 10^{-6}$.

4. BAYESIAN INFERENCE

We now proceed to perform Bayesian inference for the proposed pattern-coupled hierarchical model. To facilitate our algorithm development, we first convert (6) into a single measurement vector model by vectorizing the observation matrix $\boldsymbol{Y}$:

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{w} \tag{13}$$

where $\boldsymbol{y} \triangleq \text{vec}(\boldsymbol{Y})$, $\boldsymbol{x} \triangleq \text{vec}(\boldsymbol{X})$, $\boldsymbol{w} \triangleq \text{vec}(\boldsymbol{W})$, and $\boldsymbol{A} \triangleq \boldsymbol{I} \otimes \boldsymbol{F}$, $\otimes$ stands for the Kronecker product. Clearly, we have $\boldsymbol{y} \in \mathbb{C}^{LN}$, $\boldsymbol{x} \in \mathbb{C}^{MN}$, and $\boldsymbol{A} \in \mathbb{C}^{LN\times MN}$. Based on the above hierarchical model, the posterior distribution of $\boldsymbol{x}$ can be computed as

$$p(\boldsymbol{x}|\boldsymbol{\alpha},\gamma,\boldsymbol{y}) \propto p(\boldsymbol{x}|\boldsymbol{\alpha})p(\boldsymbol{y}|\boldsymbol{x},\gamma) \tag{14}$$

*Note that for the edge grid points, they only have two or three immediate neighboring points, in which case the definition of $N_{(m,n)}$ changes accordingly.

where according to (9), $p(\boldsymbol{x}|\boldsymbol{\alpha})$ can be expressed as

$$p(\boldsymbol{x}|\boldsymbol{\alpha}) = \prod_{m=1}^{M}\prod_{n=1}^{N} \mathcal{CN}(x_{(n-1)M+m}|0, \delta_{m,n}^{-1}) \tag{15}$$

where $x_{(n-1)M+m} = x_{m,n}$ denotes the $((n-1)M+m)$th entry of $\boldsymbol{x}$. For notational simplicity, we use $x_{m,n}$ to represent $x_{(n-1)M+m}$. It can be readily verified that the posterior $p(\boldsymbol{x}|\boldsymbol{\alpha}, \gamma, \boldsymbol{y})$ follows a Gaussian distribution with its mean and covariance matrix given respectively by

$$\begin{aligned} \boldsymbol{\mu} =& \gamma \boldsymbol{\Sigma} \boldsymbol{A}^H \boldsymbol{y} \\ \boldsymbol{\Sigma} =& (\gamma \boldsymbol{A}^H \boldsymbol{A} + \boldsymbol{D})^{-1} \end{aligned} \tag{16}$$

where $\boldsymbol{D}$ is a diagonal matrix with its $((n-1)M+m)$th diagonal element equal to $\delta_{m,n}$, i.e.

$$\boldsymbol{D} \triangleq \mathrm{diag}(\delta_{1,1}, \ldots, \delta_{M,1}, \delta_{1,2}, \ldots, \delta_{M,2}, \ldots, \delta_{1,N}, \ldots, \delta_{M,N})$$

Given the estimated hyperparameters $\{\alpha_{m,n}\}$ and γ, the maximum a posterior (MAP) estimate of $\boldsymbol{x}$ is the mean of its posterior distribution, i.e. $\hat{\boldsymbol{x}}_{\mathrm{MAP}} = \boldsymbol{\mu}$.

Our problem therefore reduces to estimating the set of sparsity-controlling hyperparameters $\{\alpha_{m,n}\}$ and the noise-related hyperparameter γ. With hyperpriors placed over $\{\alpha_{m,n}\}$ and γ, learning the hyperparameters becomes a search for their posterior mode, i.e. maximization of the posterior probability $p(\boldsymbol{\alpha}, \gamma|\boldsymbol{y})$. A strategy to maximize the posterior probability is to exploit the expectation-maximization (EM) formulation,[9] treating the signal $\boldsymbol{x}$ as hidden variables. Specifically, the algorithm produces a sequence of estimates $\boldsymbol{\alpha}^{(t)}, \gamma^{(t)}$, $t = 1, 2, 3, \ldots$, by applying two alternating steps, namely, the E-step and the M-step,[12] which is elaborated next.

E-Step: Given the current estimates of the hyperparameters $\{\boldsymbol{\alpha}^{(t)}, \gamma^{(t)}\}$ and the observed data $\boldsymbol{y}$, the E-step requires computing the expected value (with respect to the hidden variables $\boldsymbol{x}$) of the complete log-posterior of $\{\boldsymbol{\alpha}, \gamma\}$, i.e. $E_{\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\alpha}^{(t)},\gamma^{(t)}}[\log p(\boldsymbol{\alpha}, \gamma|\boldsymbol{x}, \boldsymbol{y})]$, where the operator $E_{\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\alpha}^{(t)},\gamma^{(t)}}[\cdot]$ denotes the expectation with respect to the distribution $p(\boldsymbol{x}|\boldsymbol{y}, \boldsymbol{\alpha}^{(t)}, \gamma^{(t)})$. This complete log-posterior is also referred to as the Q-function. Since

$$p(\boldsymbol{\alpha}, \gamma|\boldsymbol{x}, \boldsymbol{y}) \propto p(\boldsymbol{\alpha}) p(\boldsymbol{x}|\boldsymbol{\alpha}) p(\gamma) p(\boldsymbol{y}|\boldsymbol{x}, \gamma) \tag{17}$$

the Q-function can be decomposed into a summation of two terms

$$\begin{aligned} Q(\boldsymbol{\alpha}, \gamma|\boldsymbol{\alpha}^{(t)}, \gamma^{(t)}) &= E_{\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\alpha}^{(t)},\gamma^{(t)}}[\log p(\boldsymbol{\alpha}) p(\boldsymbol{x}|\boldsymbol{\alpha})] + E_{\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\alpha}^{(t)},\gamma^{(t)}}[\log p(\gamma) p(\boldsymbol{y}|\boldsymbol{x}, \gamma)] \\ &\triangleq Q(\boldsymbol{\alpha}|\boldsymbol{\alpha}^{(t)}, \gamma^{(t)}) + Q(\gamma|\boldsymbol{\alpha}^{(t)}, \gamma^{(t)}) \end{aligned} \tag{18}$$

Recalling (15), it can be easily derived

$$Q(\boldsymbol{\alpha}|\boldsymbol{\alpha}^{(t)}, \gamma^{(t)}) \propto \log p(\boldsymbol{\alpha}) + \sum_{m=1}^{M}\sum_{n=1}^{N} \left(\log \delta_{m,n} - \delta_{m,n} \int p(\boldsymbol{x}|\boldsymbol{y}, \boldsymbol{\alpha}^{(t)}, \gamma^{(t)}) |x_{m,n}|^2 d\boldsymbol{x} \right)$$

Since the posterior $p(\boldsymbol{x}|\boldsymbol{y}, \boldsymbol{\alpha}^{(t)}, \gamma^{(t)})$ follows a multivariate Gaussian distribution with its mean and covariance matrix given by (16), we have

$$\int p(\boldsymbol{x}|\boldsymbol{y}, \boldsymbol{\alpha}^{(t)}, \gamma^{(t)}) |x_{m,n}|^2 d\boldsymbol{x} = |\hat{\mu}_{m,n}|^2 + \hat{\sigma}_{m,n} \triangleq \omega_{m,n} \tag{19}$$

where $\hat{\mu}_{m,n}$ denotes the $((n-1)M+m)$th entry of $\hat{\boldsymbol{\mu}}$, $\hat{\sigma}_{m,n}$ denotes the $((n-1)M+m)$th diagonal element of the covariance matrix $\hat{\boldsymbol{\Sigma}}$, $\hat{\boldsymbol{\mu}}$ and $\hat{\boldsymbol{\Sigma}}$ are computed according to (16), with $\boldsymbol{\alpha}$ and γ replaced by the current estimate $\boldsymbol{\alpha}^{(t)}$ and $\gamma^{(t)}$. Hence the Q-function of $\boldsymbol{\alpha}$ can eventually be written as

$$Q(\boldsymbol{\alpha}|\boldsymbol{\alpha}^{(t)}, \gamma^{(t)}) = \sum_{m=1}^{M}\sum_{n=1}^{N} \left((a-1)\log \alpha_{m,n} - b\alpha_{m,n} + \log \delta_{m,n} - \delta_{m,n}\omega_{m,n} \right) \tag{20}$$

The Q-function of γ can be easily obtained as follows

$$Q(\gamma|\boldsymbol{\alpha}^{(t)},\gamma^{(t)}) = (LN+c-1)\log\gamma - \gamma\left(E_{\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\alpha}^{(t)},\gamma^{(t)}}\left[\|\boldsymbol{y}-\boldsymbol{A}\boldsymbol{x}\|_2^2\right]+d\right) \tag{21}$$

M-Step: In the M-step, a new estimate of $\{\boldsymbol{\alpha},\gamma\}$ is obtained by maximizing the Q-function, i.e.

$$\{\boldsymbol{\alpha}^{(t+1)},\gamma^{(t+1)}\} = \arg\max_{\boldsymbol{\alpha},\gamma} Q(\boldsymbol{\alpha},\gamma|\boldsymbol{\alpha}^{(t)},\gamma^{(t)}) \tag{22}$$

We observe that in the Q-function (18), the hyperparameters $\boldsymbol{\alpha}$ and γ are separated from each other. This allows the estimation of $\boldsymbol{\alpha}$ and γ to be decoupled into two independent optimization problems. We first examine the optimization of $\boldsymbol{\alpha}$, i.e.

$$\boldsymbol{\alpha}^{(t+1)} = \arg\max_{\boldsymbol{\alpha}} Q(\boldsymbol{\alpha}|\boldsymbol{\alpha}^{(t)},\gamma^{(t)}) \tag{23}$$

We see that in the Q-function (20), hyperparameters are entangled with each other due to the logarithm term $\log\delta_{m,n}$ (note that $\delta_{m,n}$, defined in (10), is a function of $\boldsymbol{\alpha}$). In this case, an analytical solution to the optimization (23) is difficult to obtain. Gradient descent methods can certainly be used to search for the optimal solution. Nevertheless, for gradient descent methods, there is no explicit formula for the hyperparameter update. Here we consider an alternative strategy which aims at finding a simple, analytical sub-optimal solution of (23). Such an analytical sub-optimal solution can be obtained by examining the optimality condition of (23). Suppose $\boldsymbol{\alpha}^* = \{\alpha^*_{m,n}\}$ is the optimal solution of (23), then the first derivative of the Q-function with respect to $\boldsymbol{\alpha}$ equals to zero at the optimal point, i.e.

$$\left.\frac{\partial Q(\boldsymbol{\alpha}|\boldsymbol{\alpha}^{(t)},\gamma^{(t)})}{\partial\alpha_{m,n}}\right|_{\boldsymbol{\alpha}=\boldsymbol{\alpha}^*} = \frac{a-1}{\alpha^*_{m,n}} - b + \nu^*_{m,n} - \chi_{m,n} = 0 \tag{24}$$

where

$$\nu^*_{m,n} \triangleq \frac{1}{\delta^*_{m,n}} + \beta \sum_{(i,j)\in N_{(m,n)}} \frac{1}{\delta^*_{i,j}} \tag{25}$$

$$\chi_{m,n} \triangleq \omega_{m,n} + \beta \sum_{(i,j)\in N_{(m,n)}} \omega_{i,j} \tag{26}$$

Since all hyperparameters $\{\alpha_i\}$ and β are non-negative, it can be easily verified $(1/\alpha^*_{m,n}) > (1/\delta^*_{m,n}) > 0$, and $(1/\beta\alpha^*_{m,n}) > (1/\delta^*_{i,j}) > 0$ for $(i,j)\in N_{(m,n)}$. Therefore $\nu^*_{m,n}$ is bounded by

$$\frac{5}{\alpha^*_{m,n}} > \nu^*_{m,n} > 0 \tag{27}$$

Consequently we have

$$\frac{a+4}{\alpha^*_{m,n}} > \frac{a-1}{\alpha^*_{m,n}} + \nu^*_{m,n} > \frac{a-1}{\alpha^*_{m,n}} \tag{28}$$

Combining (24) and (28), we reach that $\alpha^*_{m,n}$ is within the range

$$\alpha^*_{m,n} \in \left[\frac{a-1}{\chi_{m,n}+b}, \frac{a+4}{\chi_{m,n}+b}\right] \quad \forall m,n \tag{29}$$

A sub-optimal solution to (23) can therefore be simply chosen as

$$\alpha^{(t+1)}_{m,n} = \frac{a-1}{\chi_{m,n}+b} \quad \forall m,n \tag{30}$$

We see that the solution (30) provides a simple rule for the update of $\boldsymbol{\alpha}$. Also, notice that the update rule (30) resembles that of the conventional sparse Bayesian learning work[9] except that $\chi_{m,n}$ is equal to $\omega_{m,n}$ for the conventional sparse Bayesian learning method, while for our case, $\chi_{m,n}$ is a weighted summation of $\omega_{m,n}$ and $\omega_{i,j}$ for $(i,j) \in N_{(m,n)}$. Numerical results show that this sub-optimal update rule is quite effective and presents recovery performance similar to using a gradient-based search method.

We now discuss the update of γ. Computing the first derivative of (21) with respect to γ and setting it equal to zero, we obtain

$$\frac{1}{\gamma^{(t+1)}} = \frac{\eta + d}{LN + c - 1} \tag{31}$$

where

$$\eta \triangleq E_{\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\alpha}^{(t)},\gamma^{(t)}} \left[\|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|_2^2 \right] = \|\boldsymbol{y} - \boldsymbol{A}\hat{\boldsymbol{\mu}}\|_2^2 + (\gamma^{(t)})^{-1} \sum_{m=1}^{M} \sum_{n=1}^{N} \rho_{m,n} \tag{32}$$

in which $\rho_{m,n} \triangleq 1 - \hat{\sigma}_{m,n}\hat{\delta}_{m,n}$, $\hat{\delta}_{m,n}$ is given by (10), with $\boldsymbol{\alpha}$ replaced by $\boldsymbol{\alpha}^{(t)}$.

For clarity, we summarize our proposed algorithm as follows.

Algorithm 1 Pattern-Coupled Sparse Bayesian Learning

1. At iteration t $(t = 0, 1, \ldots)$: given the current estimates of $\boldsymbol{\alpha}^{(t)}$ and $\gamma^{(t)}$, compute the mean $\hat{\boldsymbol{\mu}}$ and the covariance matrix $\hat{\boldsymbol{\Sigma}}$ of the posterior distribution $p(\boldsymbol{x}|\boldsymbol{\alpha}^{(t)}, \gamma^{(t)}, \boldsymbol{y})$ via (16).
2. Compute a new estimate $\boldsymbol{\alpha}^{(t+1)}$ and $\gamma^{(t+1)}$, according to (30) and (31), respectively;
3. Continue the above iteration until $\|\hat{\boldsymbol{x}}^{(t+1)} - \hat{\boldsymbol{x}}^{(t)}\|_2 \leq \epsilon$, where $\hat{\boldsymbol{x}}^{(t)}$ is chosen as the mean of the posterior distribution $p(\boldsymbol{x}|\boldsymbol{\alpha}^{(t)}, \gamma^{(t)}, \boldsymbol{y})$, and ϵ is a prescribed tolerance value.

5. SIMULATION RESULTS

We now carry out experiments to illustrate the performance of our proposed pattern-coupled sparse Bayesian learning (PC-SBL) method. We compare our algorithm with the conventional range-Doppler (RD) algorithm,[7] the SBL method,[9] the basis pursuit (BP) method,[13] and the MCMC method[4] which exploits the continuity structure of the image. The "Yak-42" plane data set is used in our experiments. The radar system parameters for this data set are the same as that in.[5] The number of range cells is 256, and the total number of pulses is 256 within dwell time of 2.56s. As shown in,[4] the underlying reflectivity field of the data set exhibits a block-sparse structure, where strong scatterers are clustered and appear in a plane shape. The parameters a, b and β for our proposed algorithm are set equal to $a = 2$, $b = 10^{-6}$ and $\beta = 1$ throughout our experiments. Also, for PC-SBL and SBL, a pruning operation is adopted, that is, at each iteration, we prune those small coefficients whose associated hyperparameters are greater than $\tau = 10^2$. The corresponding atoms in $\boldsymbol{A}$ are removed from the matrix accordingly.

We first consider a noiseless case. Fig. 2 shows ISAR images recovered by respective algorithms using $L = 32$ number of pulses. We see that when only a small amount of data are available, the conventional RD method provides a defocused and blurred image which is hardly recognized. Also, our proposed PC-SBL method renders the best imaging quality among all methods. It preserves most of the strong scatterers, whereas a considerable portion of scatterers are missing in the image recovered by the SBL method. The BP method yields an obscure image with artifacts outside the target region. We also consider a noisy case where the data are corrupted by an additive white Gaussian noise with noise variance $(1/\gamma) = 0.0009$ (corresponding to SNR = 6dB). Fig. 3 depicts the images reconstructed by respective algorithms, where $L = 32$. Note that the BP method is now replaced by its noisy version method, namely, BPDN. It can be observed that the PC-SBL method still retains a decent imaging performance, whereas other compressed sensing methods degrade significantly in this case.

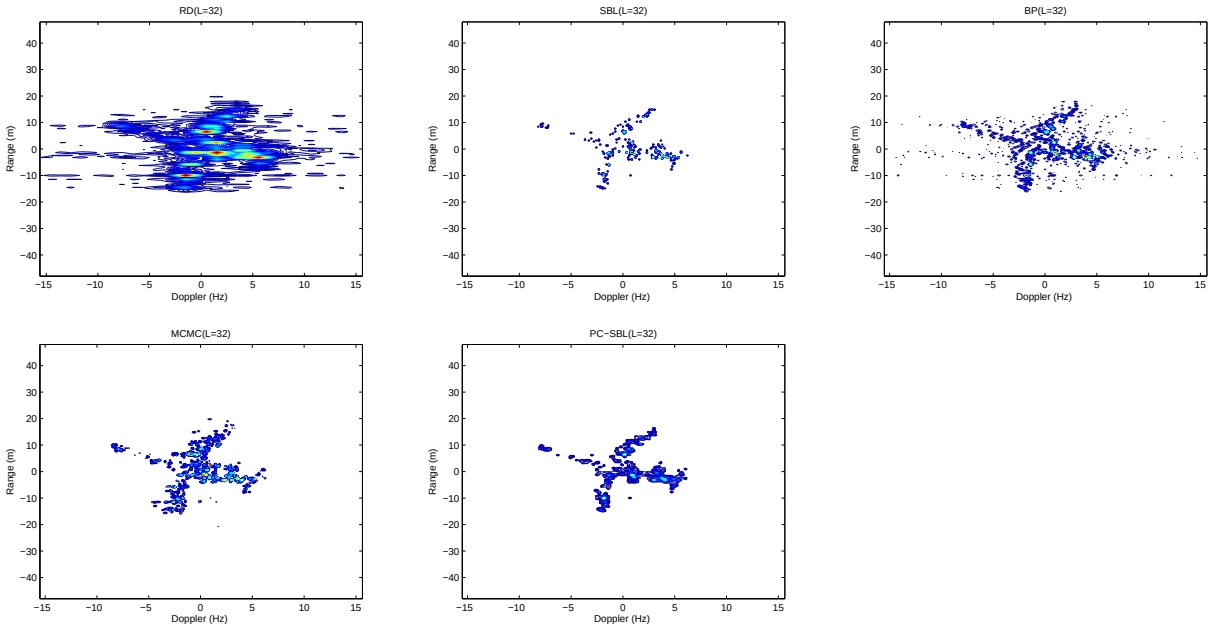

Figure 2. Images reconstructed by respective algorithms in a noiseless case, where $L = 32$.

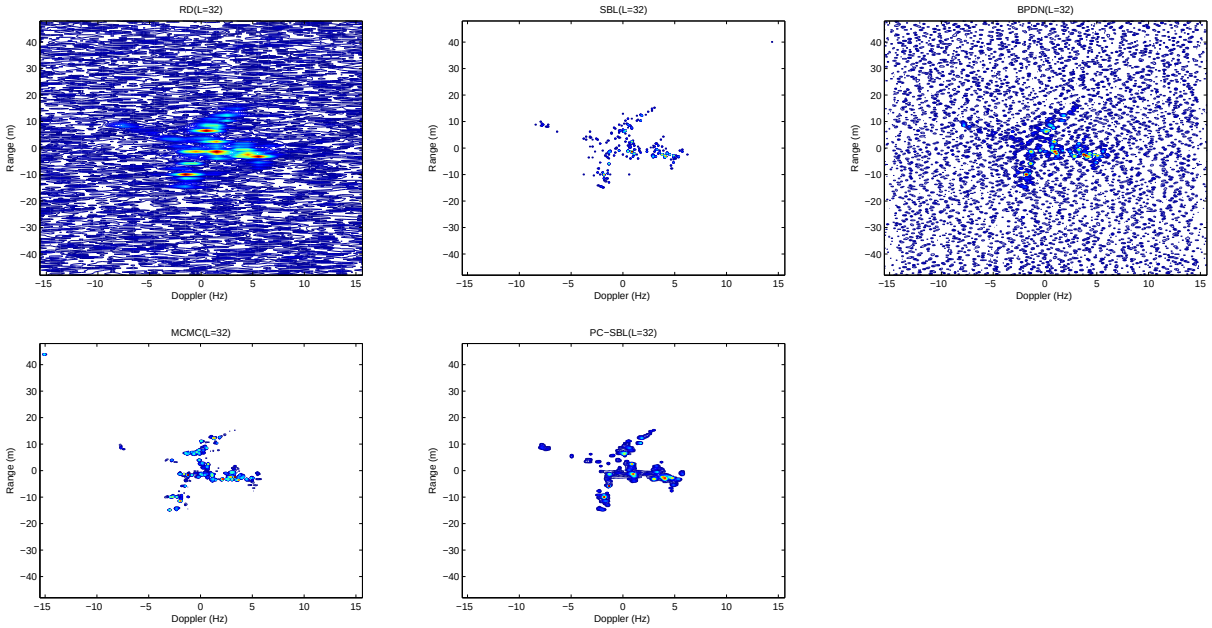

Figure 3. Images reconstructed by respective algorithms, where $L = 32$ and SNR $= 5$dB.

6. CONCLUSIONS

We developed a pattern-coupled sparse Bayesian learning method for ISAR imaging by exploiting the cluster pattern inherent in the ISAR targets. A pattern-coupled hierarchical Gaussian prior is introduced to model the pattern dependencies among neighboring scatterers on the target scene. Numerical results show that the proposed method achieves a significant image recovery improvement as compared with existing methods through exploiting the underlying block-sparse structure.

REFERENCES

[1] Cetin, M. and Karl, W. C., "Feature-enhanced synthetic aperture radar image formation based on non-quadratic regularization," *Image Processing, IEEE Transactions on* **10**(4), 623–631 (2001).

[2] Zhang, L., Xing, M., Qiu, C.-W., Li, J., and Bao, Z., "Achieving higher resolution ISAR imaging with limited pulses via compressed sampling," *IEEE Geoscience and Remote Sensing Letters* **6**, 567–571 (July 2009).

[3] Samadi, S., Cetin, M., and Masnadi-Shirazi, M. A., "Sparse representation-based synthetic aperture radar imaging," *IET Radar, Sonar and Navigation* **5**(2), 182–193 (2011).

[4] Wang, L., Zhao, L., Bi, G., Wan, C., and Yang, L., "Enhanced ISAR imaging by exploiting the continuity of the target scene," *IEEE Trans. Geoscience and Remote Sensing* **52**, 5736–5750 (Sept. 2014).

[5] Lv, J., Huang, L., Shi, Y., and Fu, X., "Inverse synthetic aperture radar imaging via modified smooth L_0 norm," *IEEE Antennas and Wireless Propogation Letters* **13**, 1235–1238 (2014).

[6] Cetin, M., Stojanovic, I., Onhon, N. O., Varshney, K. R., Samadi, S., Karl, W. C., and Willsky, A. S., "Sparsity-driven synthetic aperture radar imaging," *IEEE Signal Processing Magazine* **31**, 27–40 (July 2014).

[7] Walker, J. L., "Range-Doppler imaging of rotating objects," *IEEE Trans. Aerosp. Electron. Syst.* **AES-16**, 23–52 (Jan. 1980).

[8] Fang, J., Shen, Y., Li, H., and Wang, P., "Pattern-coupled sparse Bayesian learning for recovery of block-sparse signals," *IEEE Trans. Signal Processing* **63**, 360–372 (Jan. 2015).

[9] Tipping, M., "Sparse Bayesian learning and the relevance vector machine," *Journal of Machine Learning Research* **1**, 211–244 (2001).

[10] Wipf, D. P. and Rao, B. D., "An empirical Bayesian strategy for solving the simultaneous sparse approximation problem," *IEEE Trans. Signal Processing* **55**, 3704–3716 (July 2007).

[11] Zhang, Z. and Rao, B. D., "Sparse signal recovery with temporally correlated source vectors using sparse Bayesian learning," *IEEE Journal of Selected Topics in Signal Processing* **5**, 912–926 (Sept. 2011).

[12] Bishop, C. M., [*Pattern Recognition and Machine Learning*], Springer (2007).

[13] Chen, S. S., Donoho, D. L., and Saunders, M. A., "Atomic decomposition by basis pursuit," *SIAM J. Sci. Comput.* **20**(1), 33–61 (1998).