Digital Signal Processing 71 (2017) 95–107

Contents lists available at ScienceDirect

Digital Signal Processing

www.elsevier.com/locate/dsp

Structured Bayesian compressive sensing with spatial location dependence via variational Bayesian inference ☆

Qisong Wu *, Shiliang Fang

Key Laboratory of Underwater Acoustic Signal Processing of Ministry of Education, Southeast University, Nanjing 210096, China

ARTICLE INFO

Article history:
Available online 1 September 2017

Keywords:
Compressive sensing
Gaussian kernel
Logistic model
DOA estimation
ISAR imaging

ABSTRACT

In this paper, a novel non-parametric Bayesian compressive sensing algorithm is proposed to enhance reconstruction performance of sparse entries with a continuous structure by exploiting the location dependence of entries. An approach is proposed which involves the logistic model and location-dependent Gaussian kernel. The variational Bayesian inference scheme is used to perform the posterior distributions and acquire an approximately analytical solution. Compared to the conventional clustered based methods, which only exploit the information of neighboring pixels, the proposed approach takes the relationship between the pixels of the entire image into account to enable the utilization of the underlying sparse signal structure. It significantly reduces the required number of observations for sparse reconstruction. Both real-valued signal applications, including one-dimension signal and two-dimension image, and complex-valued signal applications, including single-snapshot direction-of-arrival (DOA) estimation of distributed sources and inverse synthetic aperture radar (ISAR) imaging with a limited number of pluses, demonstrate the superiority of the proposed algorithm.

1. Introduction

Sparse signal recovery and the associated compressive sensing (CS) approaches have attracted significant attention in recent years [1,2]. The CS techniques have the capability of recovering signals from a small number of measurement samples with a high probability, provided that the signals are sparse or can be sparsely represented in some known domains.

A typical sparse reconstruction problem in the CS model is given by,

$$\mathbf{y} = \mathbf{\Phi}\mathbf{w} + \boldsymbol{\varepsilon}, \tag{1}$$

where $\mathbf{y} \in \mathcal{R}^N$ is the observation vector, and $\boldsymbol{\varepsilon} \in \mathcal{R}^N$ is an unknown zero-mean Gaussian noise vector. $\mathbf{\Phi} \in \mathcal{R}^{N\times M}$ is a known and wide dictionary matrix with $N \ll M$, and $\mathbf{w} \in \mathcal{R}^{M\times 1}$ is a sparsity vector to be estimated. It can be shown that vector $\mathbf{y}$ preserves the information of $\mathbf{w}$ if $\mathbf{w}$ is sparse and the so-called restricted isometry property (RIP) is satisfied [2]. Define δ_A as the constant of RIP for a sensing matrix $\mathbf{\Phi} \in \mathcal{R}^{N\times M}$. If

$$N > \frac{2}{c\delta_A}\left(\ln(2m_K) + K\ln\frac{12}{\delta_A} + t\right), \tag{2}$$

where $m_K = \binom{M}{K}$ for a K-sparse signal, and $c > 0$ is a constant, then the RIP is satisfied for all elements in the subspace with the probability $1-e^{-t}$ [3]. Using the approximation $\binom{M}{K} \approx (N/K)^K$, the required number of observations is given by [1,2]

$$N = \mathcal{O}(K\log(N/K)). \tag{3}$$

To invert Eq. (1), i.e., to reconstruct the original sparse signal vector $\mathbf{w}$ from $\mathbf{y}$, a sparsity-promoting scheme is often used, such as

$$\hat{\mathbf{w}} = \arg\min_{\mathbf{w}} \frac{1}{2}\|\mathbf{y} - \mathbf{\Phi}\mathbf{w}\|_2^2 + \lambda\|\mathbf{w}\|_p, \tag{4}$$

where $\|\cdot\|_p$ is the l_p-norm with $p \in [0,1]$, to encourage signal sparsity. In the above expression, λ is an elastic parameter to balance the observation fitness and the sparse prior. For $p = 0$, the above expression corresponds to the iterative hard thresholding (IHT) algorithm [4]; for $p \in (0,1)$, it becomes the iterative reweighed algorithm [5]; when $p = 1$, it becomes a typical formula of basis pursuit denoising (BPDN) or Lasso problem [6,7].

A number of algorithms have been proposed to recover sparse signals. Commonly used algorithms include greedy algorithms, such as orthogonal matching pursuit (OMP) [8], and dynamic programming algorithms, such as basis pursuit (BP) [9], its extended version for denoising [6], and Lasso algorithm [7]. While all these

☆ The work was supported in part by Contract BK20160701 with National Natural Science Foundation of Jiangsu Province, in part Contacts 61701109, 11674057 and 11574048 with National Natural Science Foundation of China, in part Contract 3204007404 with the Fundamental Research Funds for the Central Universities.

* Corresponding author.
E-mail address: qisong.wu@seu.edu.cn (Q. Wu).

http://dx.doi.org/10.1016/j.dsp.2017.08.007

algorithms have the capabilities of reconstructing sparse signals, they require information about the number of non-zero elements either explicitly or through setting of the regularization parameter, which, in practice, may not be easily obtained. Bayesian CS (BCS) or sparse Bayesian learning (SBL) approaches form a different class of sparse signal reconstruction algorithms, which generally yield improved performance [10–12]. Sparse Bayesian learning algorithms provide effective solutions to a large class of problems based on non-parametric BCS framework, and thus have the capabilities of inferring the sparsity parameters and avoiding the nuisance parameters.

In addition to the signal sparsity, the coefficient vector $\mathbf{w}$ often exhibits special structures. For example, in direction-of-arrival (DOA) estimation, spatially spreading sources exhibit continuous angular occupancies [13–15]. In the inverse synthetic aperture radar (ISAR) imaging, the illuminated targets usually exhibit a cluster structure in the high-resolution radar system [16,17]. In through-the-wall radar imaging, most human and structural targets of interest have extended occupancies that are clustered in the image domain [18,19]. In the time-frequency analysis, frequency modulated (FM) signals have a sparse and continuous signatures in the time-frequency domain [20]. These structure characteristics can be exploited as a known information, and heuristically enhance the reconstruction performance.

The existing recovery algorithms for various structure models include block sparsity [21–28], tree structure [29], and cluster structure [16,17,30–34]. By exploiting signal structures, those methods significantly improve the sparse signal reconstruction performance in two aspects, namely increasing the recovery robustness and reducing the required number of measurements. Compared with the required number in Eq. (3) in the conventional CS, N reduces to $\mathcal{O}(K)$. Although all these algorithms take the known structure into account to enhance the reconstruction performance, they also introduce some new unknown parameters, such as the size and the number of block sparsity, which, in practice, may be not easily obtained. In the reconstruction of the block sparse signals, the locations and the size of each block are required in the recovery procedure, and the number of block sparsity is also necessary in [21]. Non-parametric clustered BCS approaches, which impose the cluster structure prior by exploiting the information of neighboring pixels, automatically infer the sparsity number, locations and sizes of the clusters. However, each local pattern has to be appropriately assigned to corresponding sets of hyper-parameters [16,31,33], in which only first-order neighboring pixels are taken into account. The extended approach in including the second-order neighboring pixels is considered by exploiting the Markov random fields (MRF) to avoid the selection of nuisance hyper-parameters [17]. In addition, when the patterns in the real-world applications do not exactly match those assigned patterns, the recovery performance would degrade.

In this paper, we propose a novel logistic Gaussian kernel based algorithm to improve sparse signal reconstruction performance by exploiting the location features. A spike-and-slab prior is first employed to impose the signal sparsity. Unlike the previous approaches developed in [16,30,31,33], which assign specific local patterns to set priors for the exploitation of clustered structures, the proposed logistic Gaussian kernel model, which combines the logistic model with the location-dependent Gaussian kernel, takes the relationship between a pixel and the entire signal entries into account. It enhances the sparse reconstruction performance by properly representing the underlying contiguous structure. The proposed approach is outlined in [35] where a Markov Chain Monte Carlo (MCMC) sampler scheme was used to perform posterior inference. As an application example, the multi-static passive radar imaging problem is considered in [36]. In this paper, we provide a comprehensive description, derivation, and analysis of this approach. In addition, to overcome the issues of expensive sampling and inefficient convergence diagnosis as observed in the MCMC sampler scheme, we further implement the variational Bayesian (VB) inference scheme to perform the derivation of posterior distributions. Due to the exploitation of the signal structure, the required number of measurements for robust reconstruction is significantly reduced. The effectiveness of the proposed method is verified with one-dimension signal example, two-dimension image example and two complex-valued applications, i.e., single-snapshot DOA estimation of distributed sources and ISAR imaging with a limited number of pluses.

Notations: We use lower-case (upper-case) bold characters to denote vectors (matrices). $(\cdot)^T$ denotes the transpose a matrix or vector. $\mathrm{diag}(\mathbf{x})$ represents a diagonal matrix that uses the elements of $\mathbf{x}$ as its diagonal elements. "$\circ$" denotes an element-wise multiplication. $\mathrm{p}(\cdot)$ denotes the probability density function (pdf). $\mathcal{N}(x|a,b)$ and $\mathcal{CN}(x|a,b)$, respectively, denote that random variable x follow a real and complex Gaussian distribution with mean a and variance b. $\mathrm{Gamma}(x|a,b)$ represents that random variable x follows a Gamma distribution with parameters a and b, and $\mathrm{Bern}(x|\pi)$ implies that random variable x follows a Bernoulli distribution with weight π. $\|\cdot\|_p$ is the l_p norm. $\mathbb{E}(x)$ denotes the expectation of the random variable x. $\mathbf{I}_N$ denotes the $N \times N$ identity matrix.

2. Data-augmentation approach for logistic mode

Bayesian inference for the logistic model is considered a difficult problem, owing to the analytically inconvenient form of the model's likelihood function. However, in [37], a novel data-augmentation strategy was proposed for Bayesian inference of the logistic model by introducing latent variables following the Pólya-Gamma (PG) distribution. This strategy leads to a simple, effective method for Bayesian inference [37]. To briefly introduce the data-augmentation approach, we focus on a fundamental integral identity at the core of the method,

$$\frac{(e^x)^a}{(1+e^x)^b} = 2^{-b} e^{\kappa x} \int_0^\infty e^{-\omega x^2/2} f(\omega) d\omega, \tag{5}$$

where $\kappa = a - (b/2)$, and $f(\omega) = \mathrm{PG}(\omega|b,0)$ is a PG distribution with parameters $(b,0)$. The function $\mathrm{PG}(\omega|b,c)$ is defined as

$$\mathrm{PG}(\omega\,|b,c) = \tau(b,c) \sum_{n=0}^{\infty} (-1)^n \frac{\Gamma(n+b)\Gamma(2n+b)}{\Gamma(n+1)\sqrt{2\pi\omega^3}} e^{-\frac{(2n+b)^2}{8\omega} - \frac{c^2\omega}{2}}, \tag{6}$$

where $\tau(b,c) = \cosh^b(c/2)2^{b-1}/\Gamma(b)$ and $\Gamma(\cdot)$ is a Gamma function. According to the property of PG distribution, the expectation of ω is given as [37],

$$\mathbb{E}(\omega) = \frac{b}{2c} \cdot \frac{e^c - 1}{1+e^c}. \tag{7}$$

Moreover, the conditional distribution

$$p(\omega\,|x) = \frac{e^{-\omega x^2/2} f(\omega)}{\int_0^\infty e^{-\omega x^2/2} f(\omega) d\omega}, \tag{8}$$

which arises in treating the intergrand in Eq. (5) as an un-normalized joint density in (x,ω), is also in the PG class: $p(\omega|x) \sim \mathrm{PG}(\omega\,|b,x)$ [37].

By introducing the latent PG variable ω, the logistic model in Eq. (5) can be achieved by a hierarchical sampling structure, where the main parameter x follows the Gaussian distribution with parameter ω, whereas ω in a lower layer follows the PG distribution.