Signal Processing 140 (2017) 104–109

Contents lists available at ScienceDirect

Signal Processing

journal homepage: www.elsevier.com/locate/sigpro

Short communication

Robust Bayesian compressed sensing with outliers☆

Qian Wan[a], Huiping Duan[b], Jun Fang[a,c,*], Hongbin Li[c], Zhengli Xing[d]

[a] National Key Laboratory 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 Electrical and Computer Engineering, Stevens Institute of Technology, Hoboken, NJ 07030, USA
[d] China Academy of Engineering Physics, Mianyang 621900, China

ARTICLE INFO

Article history:
Received 18 February 2017
Revised 27 April 2017
Accepted 13 May 2017
Available online 15 May 2017

Keywords:
Robust Bayesian compressed sensing
Variational Bayesian inference
Outlier detection

ABSTRACT

We consider the problem of robust compressed sensing where the objective is to recover a high-dimensional sparse signal from compressed measurements partially corrupted by outliers. A new sparse Bayesian learning method is developed for this purpose. The basic idea of the proposed method is to identify the outliers and exclude them from sparse signal recovery. To automatically identify the outliers, we employ a set of binary indicator variables to indicate which observations are outliers. These indicator variables are assigned a beta-Bernoulli hierarchical prior such that their values are confined to be binary. In addition, a Gaussian-inverse Gamma prior is imposed on the sparse signal to promote sparsity. Based on this hierarchical prior model, we develop a variational Bayesian method to estimate the indicator variables as well as the sparse signal. Simulation results show that the proposed method achieves a substantial performance improvement over existing robust compressed sensing techniques.

© 2017 Elsevier B.V. All rights reserved.

1. Introduction

Compressed sensing, a new paradigm for data acquisition and reconstruction, has drawn much attention over the past few years [1–3]. The main purpose of compressed sensing is to recover a high-dimensional sparse signal from a low-dimensional linear measurement vector. In practice, measurements are inevitably contaminated by noise due to hardware imperfections, quantization errors, or transmission errors. Most existing studies (e.g. [4–6]) assume that measurements are corrupted with noise that is evenly distributed across the observations, such as independent and identically distributed (i.i.d.) Gaussian, thermal, or quantization noise. This assumption is valid for many cases. Nevertheless, for some scenarios, measurements may be corrupted by outliers that are significantly different from their nominal values. For example, during the data acquisition process, outliers can be caused by sensor failures or calibration errors [7,8], and it is usually unknown which measurements have been corrupted. Outliers can also arise as a result of signal clipping/saturation or impulse noise [9,10]. Conventional compressed sensing techniques may incur severe performance degradation in the presence of outliers. To address this issue, in previous works (e.g. [7–10]), outliers are modeled as a sparse error vector, and the observed data are expressed as

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{s} + \boldsymbol{w} \tag{1}$$

where $\boldsymbol{A} \in \mathbb{R}^{M\times N}$ is the sampling matrix with $M \ll N$, $\boldsymbol{x}$ denotes an N-dimensional sparse vector with only K nonzero coefficients, $\boldsymbol{s} \in \mathbb{R}^M$ denotes the outlier vector consisting of $T \ll M$ nonzero entries with arbitrary amplitudes, and $\boldsymbol{w}$ denotes the additive multivariate Gaussian noise with zero mean and covariance matrix $(1/\gamma)\boldsymbol{I}$. The above model can be formulated as a conventional compressed sensing problem as

$$\boldsymbol{y} = \begin{bmatrix}\boldsymbol{A} & \boldsymbol{I}\end{bmatrix}\begin{bmatrix}\boldsymbol{x}\\ \boldsymbol{s}\end{bmatrix} + \boldsymbol{w} \triangleq \boldsymbol{B}\boldsymbol{u} + \boldsymbol{w} \tag{2}$$

Efficient compressed sensing algorithms can then be employed to estimate the sparse signal as well as the outliers. Recovery guarantees of $\boldsymbol{x}$ and $\boldsymbol{e}$ were also analyzed in [7–10].

The rationale behind the above approach is to detect and compensate for these outliers simultaneously. In this paper, we develop a new approach which automatically identifies and excludes the outliers from sparse signal recovery. To our best knowledge, this identify-and-reject approach is originally introduced for robust compressed sensing. It was brought to our attention that a similar idea of excluding the impulsive samples from the adaptive filter is used in [11]. Although it may seem preferable to com-

☆ This work was supported in part by the National Science Foundation of China under Grant 61522104 and Grant U1530154, and the National Science Foundation under Grant ECCS-1408182 and Grant ECCS-1609393.

* Corresponding author.
E-mail addresses: 201611260117@std.uestc.edu.cn (Q. Wan), huipingduan@uestc.edu.cn (H. Duan), JunFang@uestc.edu.cn, aquarius.fang@gmail.com (J. Fang), Hongbin.Li@stevens.edu (H. Li).

http://dx.doi.org/10.1016/j.sigpro.2017.05.017
0165-1684/© 2017 Elsevier B.V. All rights reserved.

pensate rather than simply reject outliers, inaccurate estimation of the compensation (i.e. outlier vector) could result in a destructive effect on sparse signal recovery, particularly when the number of measurements is limited. In this case, identifying and rejecting outliers could be a more sensible strategy. Also, to see the potential of the identify-and-reject approach, consider an ideal case where all outlier-corrupted measurements are correctly identified and removed, in which case the problem is simplified as

$$\bar{\boldsymbol{y}} = \bar{\boldsymbol{A}}\boldsymbol{x} + \bar{\boldsymbol{w}} \tag{3}$$

where $\bar{\boldsymbol{A}} \in \mathbb{R}^{(M-T)\times N}$ is obtained by removing the outlier-corrupted rows from $\boldsymbol{A}$. Based on compressed sensing theories [12], it can be verified that, in order to attain a same recovery probability, the data model (3) requires a smaller number of measurements than that needed by (2). This result shows that the identify-and-reject approach has the potential to outperform the compensation approach, particularly when outliers can be perfectly or near-perfectly identified.

Motivated by the identify-and-reject idea, we develop a Bayesian framework for robust compressed sensing, in which a set of binary indicator variables are employed to indicate which observations are outliers. These variables are assigned a beta-Bernoulli hierarchical prior such that their values are confined to be binary. Also, a Gaussian inverse-Gamma prior is placed on the sparse signal to promote sparsity. A variational Bayesian method is developed to find the approximate posterior distributions of the indicators, the sparse signal and other latent variables. Note that Bayesian methods, as an important class of compressed sensing techniques, have received significant attention over the past few years, e.g. [13–15].

The rest of this paper is organized as follows. In Section 2, we introduce our proposed hierarchical prior model for robust compressed sensing. An variational Bayesian method is developed in Section 3 to learn the indicator hyper-parameters as well as the sparse signal. Simulation results are provided in Section 4, followed by the concluding remarks in Section 5.

2. Hierarchical prior model

We develop a Bayesian framework which employs a set of indicator variables $\boldsymbol{z} \triangleq \{z_m\}$ to indicate which observation is an outlier, i.e. $z_m = 1$ indicates that y_m is a normal observation; otherwise y_m is an outlier. More precisely, we can write

$$y_m = \begin{cases} \boldsymbol{a}_m^r \boldsymbol{x} + w_m & z_m = 1 \\ \boldsymbol{a}_m^r \boldsymbol{x} + w_m + s_m & z_m = 0 \end{cases} \tag{4}$$

where $\boldsymbol{a}_m^r$ denotes the mth row of $\boldsymbol{A}$, s_m and w_m are the mth entry of $\boldsymbol{s}$ and $\boldsymbol{w}$, respectively. The probability of the observed data conditional on these indicator variables can be expressed as

$$p(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{z}, \gamma) = \prod_{m=1}^{M} (\mathcal{N}(y_m|\boldsymbol{a}_m^r \boldsymbol{x}, 1/\gamma))^{z_m} \tag{5}$$

From (5), we can see that if the indicator variable z_m is set to zero, then the factorization term $(\mathcal{N}(y_m|\boldsymbol{a}_m^r \boldsymbol{x}, 1/\gamma))^{z_m}$ is equal to one, which implies that the observation y_m is excluded from the probability calculation. Eq. (5) can also be deemed as a likelihood function of $\boldsymbol{x}$. It is therefore natural to discard those outliers and only keep the normal observations. Here $\mathcal{N}(x|\mu, \sigma^2)$ denotes a Gaussian distribution with mean μ and variance σ^2. To infer the indicator variables, a beta-Bernoulli hierarchical prior [16,17] is placed on $\boldsymbol{z}$, i.e. each component of $\boldsymbol{z}$ is assumed to be drawn from a Bernoulli distribution parameterized by π_m

$$p(z_m|\pi_m) = \text{Bernoulli}(z_m|\pi_m) = \pi_m^{z_m}(1-\pi_m)^{1-z_m} \quad \forall m \tag{6}$$

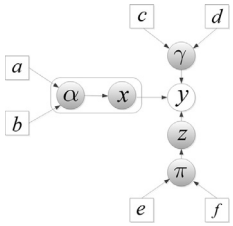

Fig. 1. Graphical model for robust Bayesian compressed sensing.

and π_m follows a beta distribution

$$p(\pi_m) = \text{Beta}(e, f) \quad \forall m \tag{7}$$

where Beta(e, f) denotes the beta distribution, and e and f are parameters characterizing the beta distribution. Note that the beta-Bernoulli prior assumes the random variables $\{z_m\}$ are mutually independent, and so are the random variables $\{\pi_m\}$.

To encourage a sparse solution, a Gaussian-inverse Gamma hierarchical prior, which has been widely used in sparse Bayesian learning (e.g. [18–21]), is employed. Specifically, in the first layer, $\boldsymbol{x}$ is assigned a Gaussian prior distribution

$$p(\boldsymbol{x}|\boldsymbol{\alpha}) = \prod_{n=1}^{N} p(x_n|\alpha_n) \tag{8}$$

where $p(x_n|\alpha_n) = \mathcal{N}(x_n|0, \alpha_n^{-1})$, and $\boldsymbol{\alpha} \triangleq \{\alpha_n\}$ are non-negative hyperparameters controlling the sparsity of the signal $\boldsymbol{x}$. The second layer specifies Gamma distributions as hyperpriors over the precision parameters $\{\alpha_n\}$, i.e.

$$p(\boldsymbol{\alpha}) = \prod_{n=1}^{N} \text{Gamma}(\alpha_n|a, b) = \prod_{n=1}^{N} \Gamma(a)^{-1} b^a \alpha_n^{a-1} e^{-b\alpha_n} \tag{9}$$

where Gamma$(\alpha_n|a, b)$ denotes the Gamma distribution, $\Gamma(a) = \int_0^\infty t^{a-1}e^{-t}dt$ is the Gamma function, and the parameters a and b are set to small values (e.g. $a = b = 10^{-10}$) in order to provide non-informative (over a logarithmic scale) hyperpriors over $\{\alpha_n\}$. Also, to estimate the noise variance, we place a Gamma hyperprior over γ, i.e.

$$p(\gamma) = \text{Gamma}(\gamma|c, d) = \Gamma(c)^{-1} d^c \gamma^{c-1} e^{-d\gamma} \tag{10}$$

where the parameters c and d are set to be small, e.g. $c = d = 10^{-10}$. The graphical model of the proposed hierarchical prior is shown in Fig. 1.

3. Variational Bayesian inference

Based on the hierarchical prior model, we now develop a variational Bayesian method [22] for robust compressed sensing. Let $\boldsymbol{\theta} \triangleq \{\boldsymbol{z}, \boldsymbol{x}, \boldsymbol{\pi}, \boldsymbol{\alpha}, \gamma\}$ denote the hidden variables in our hierarchical model. Our objective is to find the posterior distribution $p(\boldsymbol{\theta}|\boldsymbol{y})$, which is usually computationally intractable. To circumvent this difficulty, observe that the marginal probability of the observed data can be decomposed into two terms

$$\ln p(\boldsymbol{y}) = L(q) + \text{KL}(q||p) \tag{11}$$

where

$$L(q) = \int q(\boldsymbol{\theta}) \ln \frac{p(\boldsymbol{y}, \boldsymbol{\theta})}{q(\boldsymbol{\theta})} d\boldsymbol{\theta} \tag{12}$$

and

$$\text{KL}(q||p) = -\int q(\boldsymbol{\theta}) \ln \frac{p(\boldsymbol{\theta}|\boldsymbol{y})}{q(\boldsymbol{\theta})} d\boldsymbol{\theta} \tag{13}$$

where $q(\boldsymbol{\theta})$ is any probability density function, KL$(q||p)$ is the Kullback–Leibler divergence between $p(\boldsymbol{\theta}|\boldsymbol{y})$ and $q(\boldsymbol{\theta})$. Since KL$(q||p)$

≥ 0, it follows that $L(q)$ is a rigorous lower bound on $\ln p(\boldsymbol{y})$. Moreover, notice that the left hand side of (11) is independent of $q(\boldsymbol{\theta})$. Therefore maximizing $L(q)$ is equivalent to minimizing $\mathrm{KL}(q||p)$, and thus the posterior distribution $p(\boldsymbol{\theta}|\boldsymbol{y})$ can be approximated by $q(\boldsymbol{\theta})$ through maximizing $L(q)$. Specifically, we could assume some specific parameterized functional form for $q(\boldsymbol{\theta})$ and then maximize $L(q)$ with respect to the parameters of the distribution. A particular form of $q(\boldsymbol{\theta})$ that has been widely used with great success is the factorized form over the component variables in $\boldsymbol{\theta}$ [22]. For our case, the factorized form of $q(\boldsymbol{\theta})$ can be written as

$$q(\boldsymbol{\theta}) = q_z(\boldsymbol{z})q_x(\boldsymbol{x})q_\alpha(\boldsymbol{\alpha})q_\pi(\boldsymbol{\pi})q_\gamma(\gamma) \tag{14}$$

We can compute the posterior distribution approximation by finding $q(\boldsymbol{\theta})$ of the factorized form that maximizes the lower bound $L(q)$. The maximization can be conducted in an alternating fashion for each latent variable, which leads to [22]

$$\begin{aligned}
\ln q_x(\boldsymbol{x}) &= \langle \ln p(\boldsymbol{y},\boldsymbol{\theta})\rangle_{q_\alpha(\boldsymbol{\alpha})q_\gamma(\gamma)q_z(\boldsymbol{z})q_\pi(\boldsymbol{\pi})} + \text{constant}\\
\ln q_\alpha(\boldsymbol{\alpha}) &= \langle \ln p(\boldsymbol{y},\boldsymbol{\theta})\rangle_{q_x(\boldsymbol{x})q_\gamma(\gamma)q_z(\boldsymbol{z})q_\pi(\boldsymbol{\pi})} + \text{constant}\\
\ln q_\gamma(\gamma) &= \langle \ln p(\boldsymbol{y},\boldsymbol{\theta})\rangle_{q_x(\boldsymbol{x})q_\alpha(\boldsymbol{\alpha})q_z(\boldsymbol{z})q_\pi(\boldsymbol{\pi})} + \text{constant}\\
\ln q_z(\boldsymbol{z}) &= \langle \ln p(\boldsymbol{y},\boldsymbol{\theta})\rangle_{q_x(\boldsymbol{x})q_\alpha(\boldsymbol{\alpha})q_\gamma(\gamma)q_\pi(\boldsymbol{\pi})} + \text{constant}\\
\ln q_\pi(\boldsymbol{\pi}) &= \langle \ln p(\boldsymbol{y},\boldsymbol{\theta})\rangle_{q_x(\boldsymbol{x})q_\alpha(\boldsymbol{\alpha})q_\gamma(\gamma)q_z(\boldsymbol{z})} + \text{constant}
\end{aligned} \tag{15}$$

where $\langle\cdot\rangle_{\cdot}$ denotes an expectation with respect to the distributions specified in the subscript. More details of the Bayesian inference are provided below.

1) **Update of** $q_x(\boldsymbol{x})$: We first consider the calculation of $q_x(\boldsymbol{x})$. Keeping those terms that are dependent on $\boldsymbol{x}$, we have

$$\begin{aligned}
\ln q_x(\boldsymbol{x}) &\propto \langle \ln p(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{z},\gamma) + \ln p(\boldsymbol{x}|\boldsymbol{\alpha})\rangle_{q_\alpha(\boldsymbol{\alpha})q_\gamma(\gamma)q_z(\boldsymbol{z})}\\
&\propto -\sum_{m=1}^{M}\frac{\langle \gamma z_m(y_m-\boldsymbol{a}_m^r\boldsymbol{x})^2\rangle}{2} - \frac{1}{2}\sum_{n}^{N}\langle\alpha_n x_n^2\rangle\\
&= -\frac{\langle\gamma\rangle(\boldsymbol{y}-\boldsymbol{A}\boldsymbol{x})^T\boldsymbol{D}_z(\boldsymbol{y}-\boldsymbol{A}\boldsymbol{x})}{2} - \frac{1}{2}\boldsymbol{x}^T\boldsymbol{D}_\alpha\boldsymbol{x}
\end{aligned} \tag{16}$$

where

$$\boldsymbol{D}_z \triangleq \mathrm{diag}(\langle\boldsymbol{z}\rangle),\quad \boldsymbol{D}_\alpha \triangleq \mathrm{diag}(\langle\boldsymbol{\alpha}\rangle) \tag{17}$$

$\langle\boldsymbol{z}\rangle$ and $\langle\boldsymbol{\alpha}\rangle$ denote the expectation of $\boldsymbol{z}$ and $\boldsymbol{\alpha}$, respectively. It is easy to show that $q(\boldsymbol{x})$ follows a Gaussian distribution with its mean and covariance matrix given respectively by

$$\boldsymbol{\mu}_x = \langle\gamma\rangle\boldsymbol{\Phi}_x\boldsymbol{A}^T\boldsymbol{D}_z\boldsymbol{y} \tag{18}$$

$$\boldsymbol{\Phi}_x = \left(\langle\gamma\rangle\boldsymbol{A}^T\boldsymbol{D}_z\boldsymbol{A} + \boldsymbol{D}_\alpha\right)^{-1} \tag{19}$$

2) **Update of** $q_\alpha(\boldsymbol{\alpha})$: Keeping only the terms that depend on $\boldsymbol{\alpha}$, the variational optimization of $q_\alpha(\boldsymbol{\alpha})$ yields

$$\begin{aligned}
\ln q_\alpha(\boldsymbol{\alpha}) &\propto \langle \ln p(\boldsymbol{x}|\boldsymbol{\alpha}) + \ln p(\boldsymbol{\alpha}|a,b)\rangle_{q_x(\boldsymbol{x})}\\
&= \sum_{n=1}^{N}(a+0.5)\ln\alpha_n - (0.5\langle x_n^2\rangle + b)\alpha_n
\end{aligned} \tag{20}$$

The posterior $q_\alpha(\boldsymbol{\alpha})$ therefore follows a Gamma distribution

$$q_\alpha(\boldsymbol{\alpha}) = \prod_{n=1}^{N}\mathrm{Gamma}(\alpha_n|\tilde{a},\tilde{b}_n) \tag{21}$$

in which $\tilde{a}$ and $\tilde{b}_n$ are given respectively as

$$\begin{aligned}
\tilde{a} &= a + 0.5\\
\tilde{b}_n &= b + 0.5\langle x_n^2\rangle
\end{aligned}$$

3) **Update of** $q_\gamma(\gamma)$: The variational approximation of $q_\gamma(\gamma)$ can be obtained as:

$$\begin{aligned}
\ln q_\gamma(\gamma) &\propto \langle \ln p(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{z},\gamma) + \ln p(\gamma|c,d)\rangle_{q_x(\boldsymbol{x})q_z(\boldsymbol{z})}\\
&\propto \sum_{m=1}^{M}\left(0.5\langle z_m\rangle\ln\gamma - 0.5\gamma\langle z_m\rangle\langle(y_m-\boldsymbol{a}_m^r\boldsymbol{x})^2\rangle\right)\\
&\quad + (c-1)\ln\gamma - d\gamma\\
&= (c+0.5\sum_{m=1}^{M}\langle z_m\rangle - 1)\ln\gamma - (d + 0.5\langle(\boldsymbol{y}-\boldsymbol{A}\boldsymbol{x})^T\\
&\quad \boldsymbol{D}_z(\boldsymbol{y}-\boldsymbol{A}\boldsymbol{x})\rangle)\gamma
\end{aligned} \tag{22}$$

Clearly, the posterior $q_\gamma(\gamma)$ obeys a Gamma distribution

$$q_\gamma(\gamma) = \mathrm{Gamma}(\gamma|\tilde{c},\tilde{d}) \tag{23}$$

where $\tilde{c}$ and $\tilde{d}$ are given respectively as

$$\tilde{c} = c + 0.5\sum_{m=1}^{M}\langle z_m\rangle \tag{24}$$

$$\tilde{d} = d + 0.5\langle(\boldsymbol{y}-\boldsymbol{A}\boldsymbol{x})^T\boldsymbol{D}_z(\boldsymbol{y}-\boldsymbol{A}\boldsymbol{x})\rangle_{q_x(\boldsymbol{x})} \tag{25}$$

in which

$$\begin{aligned}
&\langle(\boldsymbol{y}-\boldsymbol{A}\boldsymbol{x})^T\boldsymbol{D}_z(\boldsymbol{y}-\boldsymbol{A}\boldsymbol{x})\rangle_{q_x(\boldsymbol{x})}\\
&= (\boldsymbol{y}-\boldsymbol{A}\boldsymbol{\mu}_x)^T\boldsymbol{D}_z(\boldsymbol{y}-\boldsymbol{A}\boldsymbol{\mu}_x) + \mathrm{trace}(\boldsymbol{A}^T\boldsymbol{D}_z\boldsymbol{A}\boldsymbol{\Phi}_x)
\end{aligned}$$

4) **Update of** $q_z(\boldsymbol{z})$: The posterior approximation of $q_z(\boldsymbol{z})$ yields

$$\begin{aligned}
\ln q_z(\boldsymbol{z}) &\propto \langle \ln p(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{z},\gamma) + \ln p(\boldsymbol{z}|\boldsymbol{\pi})\rangle_{q_x(\boldsymbol{x})q_\gamma(\gamma)q_\pi(\boldsymbol{\pi})}\\
&\propto \sum_{m=1}^{M}\langle z_m\left(-0.5\gamma(y_m-\boldsymbol{a}_m^r\boldsymbol{x})^2 + \ln\pi_m\right)\\
&\quad + (1-z_m)\ln(1-\pi_m)\rangle
\end{aligned} \tag{26}$$

Clearly, z_m still follows a Bernoulli distribution with its probability given by

$$P(z_m=1) = Ce^{\langle\ln\pi_m\rangle}e^{-\frac{\gamma\langle(y_m-\boldsymbol{a}_m^r\boldsymbol{x})^2\rangle}{2}} \tag{27}$$

$$P(z_m=0) = Ce^{\langle\ln(1-\pi_m)\rangle} \tag{28}$$

where C is a normalizing constant such that $P(z_m=1)+P(z_m=0)=1$, and

$$\begin{aligned}
\langle(y_m-\boldsymbol{a}_m^r\boldsymbol{x})^2\rangle &= (y_m-\boldsymbol{a}_m^r\boldsymbol{\mu}_{\boldsymbol{x}})^2 + \boldsymbol{a}_m^r\boldsymbol{\Phi}_x\boldsymbol{a}_m^{r\,T}\\
\langle\ln\pi_m\rangle &= \Psi(e+\langle z_m\rangle) - \Psi(e+f+1)\\
\langle\ln(1-\pi_m)\rangle &= \Psi(1+f-\langle z_m\rangle) - \Psi(e+f+1)
\end{aligned} \tag{29}$$

The last two equalities can also be found in [17], in which $\Psi(\cdot)$ represents the digamma function.

5) **Update of** $q_\pi(\boldsymbol{\pi})$: The posterior approximation of $q_\pi(\boldsymbol{\pi})$ can be calculated as

$$\begin{aligned}
\ln q_\pi(\boldsymbol{\pi}) &\propto \langle \ln p(\boldsymbol{z}|\boldsymbol{\pi}) + \ln p(\boldsymbol{\pi}|e,f)\rangle_{q_z(\boldsymbol{z})}\\
&\propto \sum_{m=1}^{M}\langle z_m\ln\pi_m + (1-z_m)\ln(1-\pi_m) + (e-1)\ln\pi_m\\
&\quad + (f-1)\ln(1-\pi_m)\rangle\\
&= \sum_{m=1}^{M}\langle(z_m+e-1)\ln\pi_m + (f-z_m)\ln(1-\pi_m)\rangle
\end{aligned} \tag{30}$$

It can be easily verified that $q_\pi(\boldsymbol{\pi})$ follows a Beta distribution, i.e.

$$q_\pi(\boldsymbol{\pi}) = \prod_m p(\pi_m) = \prod_m \mathrm{Beta}(\langle z_m\rangle + e, 1+f-\langle z_m\rangle) \tag{31}$$

In summary, the variational Bayesian inference involves updates of the approximate posterior distributions for hidden variables $\boldsymbol{x}$, $\boldsymbol{\alpha}$, $\boldsymbol{z}$, $\boldsymbol{\pi}$, and γ in an alternating fashion. Some of the expectations and moments used during the update are summarized as

$$\langle \alpha_n \rangle = \frac{\tilde{a}}{\tilde{b}_n} \tag{32}$$

$$\langle \gamma \rangle = \frac{\tilde{c}}{\tilde{d}} \tag{33}$$

$$\langle x_n^2 \rangle = \langle x_n \rangle^2 + \boldsymbol{\Phi}_x(n,n) \tag{34}$$

$$\langle z_m \rangle = \frac{P(z_m = 1)}{P(z_m = 1) + P(z_m = 0)} \tag{35}$$

where $\boldsymbol{\Phi}_x(n, n)$ denotes the nth diagonal element of $\boldsymbol{\Phi}_x$.

Remark. We discuss the computational complexity of our proposed method and the compensation approach. The main computational complexity of our proposed method involves computing an inverse of an $N \times N$ matrix (cf. (19)), which has the computational complexity of order $O(N^3)$. For the compensation-based method (2), it has a computational complexity of order $O((M+N)^3)$. Therefore, our proposed approach has an advantage over the compensation-based method in terms of computational complexity.

For clarity, we summarize our algorithm as follows.

BP-RBCS Algorithm

Input: Observed data $\boldsymbol{y}$, measurement matrix $\boldsymbol{A}$
1. Given initial estimates $\boldsymbol{\alpha}$, γ and parameter a, b, c, d, e and f
2. At iteration $t = 0, 1, \ldots$:
Update the estimate $\hat{\boldsymbol{x}}^{(t)}$ according to (18);
Update the hyperparameters $\boldsymbol{\alpha}$ and γ according to (32) and (33), respectively;
Update the indicators $\boldsymbol{z}$ according to (35).
3. Continue the above iteration until $\|\hat{\boldsymbol{x}}^{(t+1)} - \hat{\boldsymbol{x}}^{(t)}\|_2 \le \epsilon$, where ϵ is a prescribed tolerance value.
Output: Recovered signal $\hat{\boldsymbol{x}}^{(t)}$.

4. Simulation results

We now carry out experiments to illustrate the performance of our proposed method which is referred to as the beta-Bernoulli prior model-based robust Bayesian compressed sensing method (BP-RBCS)[1]. As discussed earlier, another robust compressed sensing approach is compensation-based and can be formulated as a conventional compressed sensing problem (2). For comparison, the sparse Bayesian learning method [18,23] is employed to solve (2), and this method is referred to as the compensation-based robust Bayesian compressed sensing method (C-RBCS). The SL0[2] algorithm which has demonstrated superior performance is also included and employed to solve (2) in our experiments. Also, we consider an "ideal" method which assumes the knowledge of the locations of the outliers. The outliers are then removed and the sparse Bayeisan learning method is employed to recover the sparse signal. This ideal method is referred to as RBCS-ideal, and serves as a benchmark for the performance of the BP-RBCS and C-RBCS. Note that both C-RBCS and RBCS-ideal use the sparse Bayesian learning method for sparse signal recovery. The parameters $\{a, b, c, d\}$ of the sparse Bayesian learning method are set to $a = b = c = d = 10^{-10}$. Our proposed method involves the parameters $\{a, b, c, d, e, f\}$. The first four are also set to $a = b = c = d = 10^{-10}$. The beta-Bernoulli parameters $\{e, f\}$ are set to $e = 0.7$ and $f = 1 - e = 0.3$ since we expect that the number of outliers is usually small relative to the total number of measurements. Our simulation results suggest that stable recovery is ensured as long as e is set to a value in the range $[0.5, 1]$.

We first examine the performance of respective algorithms for random measurement matrices $\boldsymbol{A} \in \mathbb{R}^{M\times N}$ with i.i.d Gaussian entries. The K nonzero entries of the sparse signal $\boldsymbol{x} \in \mathbb{R}^N$ are drawn from a Gaussian distribution $\mathcal{N}(0, 1)$. We consider a noiseless case. Suppose that T out of M measurements are corrupted by outliers. For those corrupted measurements $\{y_m\}$, their values are chosen uniformly from $[-10, 10]$. Fig. 2 depicts the success rates of different methods vs. the number of measurements and the number of outliers, respectively, where we set $N = 64$, $K = 2$, $T = 10$ in Fig. 2(a), and $M = 25$, $K = 2$, $N = 64$ in Fig. 2(b). The success rate is computed as the ratio of the number of successful trials to the total number of independent runs. A trial is considered successful if the normalized reconstruction error of the sparse signal $\boldsymbol{x}$ is no greater than 10^{-6}. Results are averaged over 10^3 independent runs. From Fig. 2, we see that our proposed BP-RBCS presents a clear performance advantage over the C-RBCS and the SL0.

Next, we consider the problem of direction-of-arrival (DOA) estimation where K narrowband far-field sources impinge on a uniform linear array of M sensors from different directions. The received signal can be expressed as

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

where $\boldsymbol{w}$ denotes i.i.d. Gaussian observation noise with zero mean and variance $1/\gamma$, $\boldsymbol{A} \in \mathbb{C}^{M\times N}$ is an overcomplete dictionary constructed by evenly-spaced angular points $\{\theta_n\}$, with the (m, n)th entry of $\boldsymbol{A}$ given by $a_{m,n} = \exp\{-2j\pi(m-1)\sin(\theta_n)D/\lambda\}$, in which D denotes the distance between two adjacent sensors, λ represents the wavelength of the source signal, and $\{\theta_n\}$ are evenly-spaced grid points in the interval $[-\pi/2, \pi/2]$. The signal $\boldsymbol{x}$ contains K nonzero entries that are independently drawn from a unit circle. The values of measurements corrupted with outliers are chosen uniformly from $[-10, 10]$. The SL0 method is not included in order to better show the visual difference between the proposed BP-RBCS and the C-RBCS. We first consider a noiseless case, i.e. $1/\gamma = 0$. Fig. 3 depicts the success rates of different methods vs. the number of measurements and the number of outliers, respectively, where we set $N = 64$, $K = 3$, $T = 7$ in Fig. 3(a), and $M = 25$, $K = 3$, $N = 64$ in Fig. 3(b). From Fig. 3, we see that our proposed BP-RBCS achieves a substantial performance improvement over the C-RBCS. This result corroborates our claim that rejecting outliers is a better strategy than compensating for outliers, particularly when the number of measurements is limited. Next, we consider a noisy case with $1/\gamma = 0.01$. Fig. 4 plots the normalized mean square errors (NMSEs) of different methods vs. the number of measurements and the number of outliers, respectively, we set $N = 64$, $K = 3$, $T = 7$ in Fig. 4(a), and $M = 25$, $K = 3$, $N = 64$ in Fig. 4(b). The 95% confidence intervals for the NMSEs are also shown in Fig. 4, where the vertical line segments represent the confidence intervals surrounding the means. This result, again, demonstrates the superiority of our proposed method over the C-RBCS. We show in Table 1 the average computing time of the BP-RBCS and the C-RBCS, respectively, where we set $K = 3$, $T = 7$. We see that our proposed BP-RBCS is more computationally efficient than the compensation-based method.

5. Conclusions

We proposed a new Bayesian method for robust compressed sensing. The rationale behind the proposed method is to identify the outliers and exclude them from sparse signal recovery. To

[1] Codes are available at http://www.junfang-uestc.net/codes/RBCS.rar.
[2] Codes are available at http://ee.sharif.edu/~SLzero/.

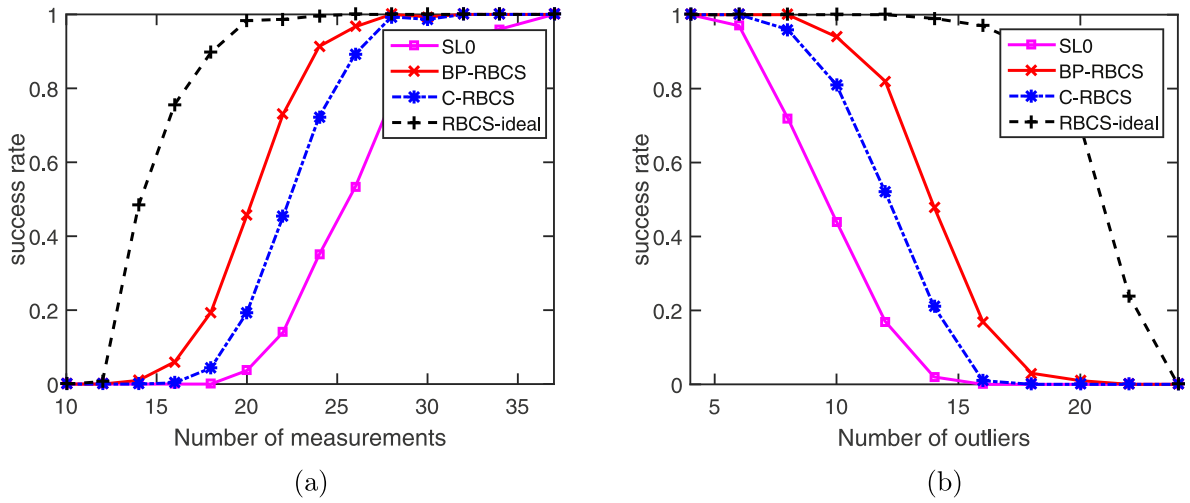

Fig. 2. (a) Success rates of respective algorithms vs. M; (b). Success rates of respective algorithms vs. T.

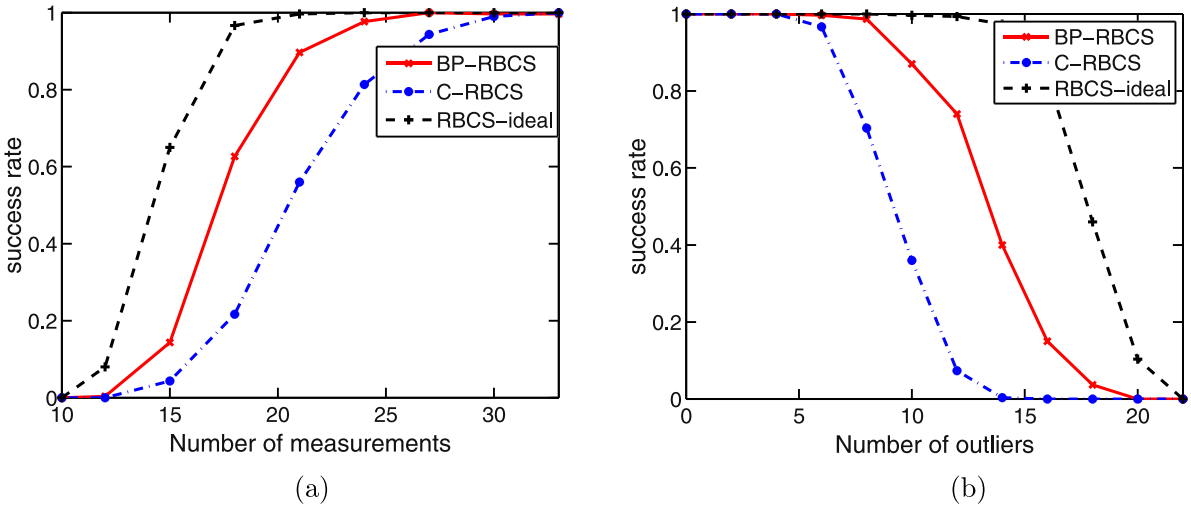

Fig. 3. DOA estimation (a) Success rates of respective algorithms vs. M; (b). Success rates of respective algorithms vs. T.

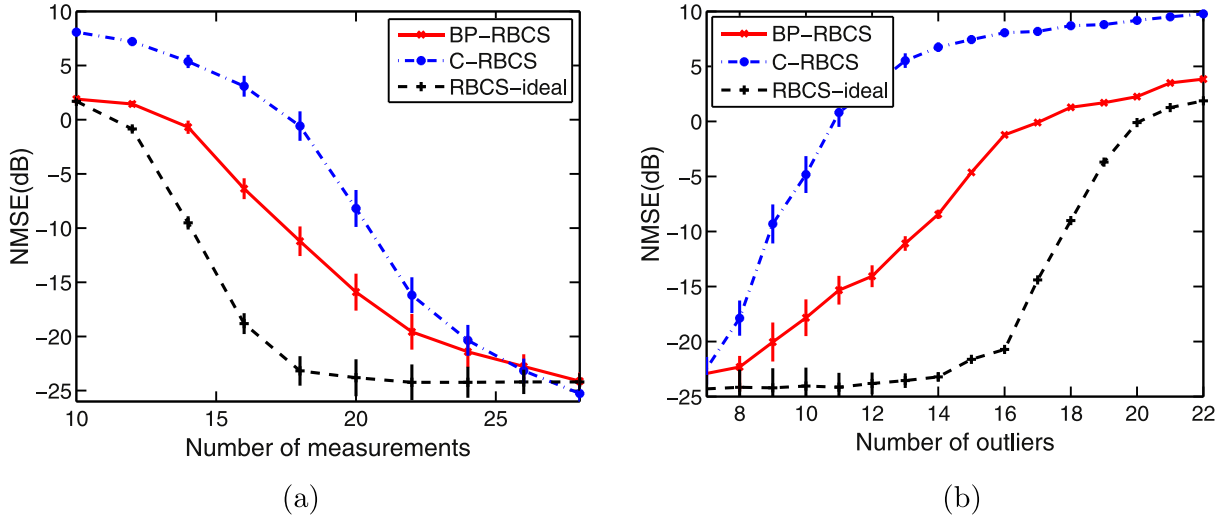

Fig. 4. DOA estimation (a) NMSEs of respective algorithms vs. M; (b). NMSEs of respective algorithms vs. T.

Table 1
Average computing time of algorithms BP-RBCS and C-RBCS.

Standard deviation	Algorithm	$M \times N$	Runtime(s)
$\sigma = 0$	C-RBCS	25×64	0.39
	C-RBCS	45×64	0.21
	BP-RBCS	25×64	**0.18**
	BP-RBCS	45×64	**0.09**
$\sigma = 0.1$	C-RBCS	25×64	1.56
	C-RBCS	45×64	2.32
	BP-RBCS	25×64	**1.23**
	BP-RBCS	45×64	**1.39**

this objective, a set of indicator variables were employed to indicate which observations are outliers. A beta-Bernoulli prior is assigned to these indicator variables. A variational Bayesian inference method was developed to find the approximate posterior distributions of the latent variables. Simulation results show that our proposed method achieves a substantial performance improvement over the compensation-based robust compressed sensing method.

References

[1] S.S. Chen, D.L. Donoho, M.A. Saunders, Atomic decomposition by basis pursuit, SIAM J. Sci. Comput. 20 (1) (1998) 33–61.
[2] E. Candés, T. Tao, Decoding by linear programming, IEEE Trans. Inf. Theory (12) (2005) 4203–4215.
[3] D.L. Donoho, Compressive sensing, IEEE Trans. Inf. Theory 52 (2006) 1289–1306.
[4] E. Candes, The restricted isometry property and its implications for compressive sensing, Compte Rendus de l'Academie des Sciences, Paris, Serie I 346 (2008) 589–592.
[5] M.J. Wainwright, Information-theoretic limits on sparsity recovery in the high-dimensional and noisy setting, IEEE Trans. Inf. Theory 55 (12) (2009) 5728–5741.
[6] T. Wimalajeewa, P.K. Varshney, Performance bounds for sparsity pattern recovery with quantized noisy random projections, IEEE J. Sel. Topics Signal Process. 6 (1) (2012) 43–57.
[7] R.G.B. Jason, N. Laska, M.A. Davenport, Exact signal recovery from sparsely corrupted measurements through the pursuit of justice, in: The 43rd Asilomar Conference on Signals, Systems and Computers, 2009. Pacific Grove, California, USA
[8] R.C.K. Mitra, A. Veeraraghavan, Analysis of sparse regularization based robust regression approaches, IEEE Trans. Signal Process. (5) (2013) 1249–1257.
[9] R.E. Carrillo, K.E. Barner, T.C. Aysal, Robust sampling and reconstruction methods for sparse signals in the presence of impulsive noise, IEEE J. Sel. Topics Signal Process. (2) (2010) 392–408.
[10] C. Studer, P. Kuppinger, G. Pope, H. Bolcskei, Recovery of sparsely corrupted signals, IEEE Trans. Inf. Theory (5) (2012) 3115–3130.
[11] H. Zayyani, Y. Attar, An oracle normalized least mean square (nlms) and a simple Bayesian detection nlms algorithm robust to impulse noise, Majlesi J. Electr. Eng. 10 (3) (2016) 27.
[12] E. Candès, M.B. Wakin, An introduction to compressive sensing, IEEE Signal Process. Mag. 25 (2) (2008) 21–30.
[13] H. Zayyani, M. Babaie-Zadeh, C. Jutten, An iterative bayesian algorithm for sparse component analysis in presence of noise, IEEE Trans. Signal Process. 57 (11) (2009) 4378–4390.
[14] D.P. Wipf, B.D. Rao, Sparse Bayesian learning for basis selection, IEEE Trans. Signal Process. 52 (8) (2004) 2153–2164.
[15] H. Zayyani, M. Babaie-Zadeh, C. Jutten, Bayesian pursuit algorithm for sparse representation, in: Acoustics, Speech and Signal Processing, 2009. ICASSP 2009. IEEE International Conference on, IEEE, 2009, pp. 1549–1552.
[16] L. He, L. Carin, Exploiting structure in wavelet-based Bayesian compressive sensing, IEEE Trans. Signal Process. 57 (9) (2009) 3488–3497.
[17] J. Paisley, L. Carin, Nonparametric factor analysis with Beta process priors, in: 26th Annual International Conference on Machine Learning, Montreal, Canada, 2009, pp. 14–18.
[18] S. Ji, Y. Xue, L. Carin, Bayesian compressive sensing, IEEE Trans. Signal Process. 56 (6) (2008) 2346–2356.
[19] Z. Zhang, B.D. Rao, Extension of SBL algorithms for the recovery of block sparse signals with intra-block correlation, IEEE Trans. Signal Process. 61 (8) (2013) 2009–2015.
[20] Z. Yang, L. Xie, C. Zhang, Off-grid direction of arrival estimation using sparse Bayesian inference, IEEE Trans. Signal Process. 61 (1) (2013) 38–42.
[21] J. Fang, Y. Shen, H. Li, P. Wang, Pattern-coupled sparse Bayesian learning for recovery of block-sparse signals, IEEE Trans. Signal Process. (2) (2015) 360–372.
[22] D.G. Tzikas, A.C. Likas, N.P. Galatsanos, The variational approximation for Bayesian inference, IEEE Signal Process. Mag. (2008) 131–146.
[23] M. Tipping, Sparse Bayesian learning and the relevance vector machine, J. Mach. Learn. Res. 1 (2001) 211–244.