Robust One-Bit Bayesian Compressed Sensing with Sign-Flip Errors

Fuwei Li, Jun Fang, *Member, IEEE*, Hongbin Li, *Senior Member, IEEE*, and Lei Huang, *Member, IEEE*

***Abstract*—We consider the problem of sparse signal recovery from one-bit measurements. Due to the noise present in the acquisition and transmission process, some quantized bits may be flipped to their opposite states. These bit-flip errors, also referred to as the sign-flip errors, may result in severe performance degradation. To address this issue, we introduce a robust Bayesian compressed sensing framework to account for sign flip errors. Specifically, sign-flip errors are considered as a result of a sparse noise-corrupted model in which original (unquantized) observations are corrupted by sparse (impulse) noise. A Gaussian-inverse Gamma hierarchical prior is assigned to the noise vector to promote sparsity. Based on the modified hierarchical model, we develop a variational expectation-maximization (EM) algorithm to identify the sign-flip errors and recover the sparse signal simultaneously. Numerical results are provided to illustrate the effectiveness and superiority of the proposed method.**

***Index Terms*—One-bit Bayesian compressed sensing, sign-flip errors, variational expectation-maximization.**

I. Introduction

COMPRESSED sensing with one-bit quantized measurements has attracted considerable attention recently due to its substantial potential benefits in data acquisition. In particular, one-bit quantization can significantly reduce the hardware complexity and cost, which makes large-scale data acquisition more tractable [1]. One-bit compressed sensing was firstly introduced by Boufounos and Baraniuk [2], with the objective to recover a sparse or compressible signal from one-bit measurements

$$\boldsymbol{t} = \text{sign}(\boldsymbol{y}) = \text{sign}(\boldsymbol{Ax}) \tag{1}$$

where "sign" denotes an operator that performs the sign function element-wise on the vector, the sign function returns 1 for positive numbers and 0 otherwise. Following [2], a variety of one-bit compressed sensing algorithms such as matching sign pursuit (MSP) [3], binary iterative hard thresholding (BIHT) [4], and many others [1], [5]–[7] were proposed. All these studies assume that one-bit measurements are error-free. In practice, however, due to the noise in signal acquisition and transmission, some of the signs may be flipped to their opposite states, in which case the above algorithms may suffer from considerable performance loss. To address this issue, an adaptive outlier pursuit method [8] and a noise-adaptive renormalized fixed point iteration method [9] were developed to automatically find the outlier (sign-flip) errors. These methods, however, require the knowledge of the number of sign-flip errors (which is usually unknown in advance) in their formulations, and their performance deteriorates if the number is set different from the true value.

In this paper, we introduce a robust Bayesian compressed sensing framework to address sign flip errors. Specifically, the sign-flip errors are modeled as a result of corrupting unquantized observations by a sparse (impulse) noise vector. A Gaussian-inverse Gamma hierarchical prior is assigned to the noise vector to encourage sparsity. Based on this model, we develop a variational expectation-maximization (EM) algorithm which simultaneously identifies the bit-flip errors and recovers the sparse signal. The proposed algorithm does not require the *a priori* knowledge of the number of bit flips and the sparsity level of the sparse signal. Besides, no additional tuning parameters are needed for the proposed algorithm.

II. Hierarchical Model

We aim to recover a sparse signal from one-bit quantized measurements, where some measurements' signs may be flipped to their opposite states due to observation or transmission noise. The sign-flip errors can naturally be modeled as a result of corrupting the original observations by a sparse (impulse) noise vector. Specifically, we have

$$\boldsymbol{t} = \text{sign}(\boldsymbol{y}) = \text{sign}(\boldsymbol{Ax} + \boldsymbol{w}) \tag{2}$$

where $\boldsymbol{t} \triangleq [t_1\, t_2 \dots t_m]^T$ are the binary observations, $\boldsymbol{y} \triangleq [y_1\, y_2 \dots y_m]^T$ denote the unquantized original measurements, and $\boldsymbol{w} \triangleq [w_1\, w_2 \dots w_m]^T$ is a sparse (impulse) noise vector with only a few nonzero coefficients. Note that although the true observation noise vector does not have a sparse structure, only those elements which cause bit-flip errors should be retained, other elements which have no impact on the signs can be ignored. Our objective is to jointly estimate the sparse signal $\boldsymbol{x}$ and detect the sign-flip errors based on one-bit observations $\boldsymbol{t}$.

Manuscript received October 04, 2014; revised November 12, 2014; accepted November 18, 2014. Date of publication November 20, 2014; date of current version December 03, 2014. This work was supported in part by the National Science Foundation of China under Grants 61172114, 61201274 and 61428103, and by the National Science Foundation under Grant ECCS-1408182. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Michael Wakin.

F. Li and J. Fang are with the National Key Laboratory of Science and Technology on Communications, University of Electronic Science and Technology of China, Chengdu 611731, China (e-mail: JunFang@uestc.edu.cn).

H. Li is with the Department of Electrical and Computer Engineering, Stevens Institute of Technology, Hoboken, NJ 07030 USA (e-mail: Hongbin.Li@stevens.edu).

L. Huang is with the Department of Electronic and Information Engineering, Harbin Institute of Technology Shenzhen Graduate School, Shenzhen, China (e-mail: dr.lei.huang@ieee.org).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/LSP.2014.2373380

To develop a Bayesian framework for one-bit compressed sensing, we first need to introduce a probabilistic model to quantify the probability of $\boldsymbol{t}$ given the input $\boldsymbol{y}$. Following [10], the likelihood of $\boldsymbol{t}$ given $\boldsymbol{y}$ can be expressed as

$$p(\boldsymbol{t}|\boldsymbol{y}) = \prod_{i=1}^{m} \sigma(y_i)^{t_i}[1-\sigma(y_i)]^{1-t_i} \tag{3}$$

where $\sigma(y) \triangleq 1/(1+\exp(-y))$ is the logistic function. Note that the logistic function, with an 'S' shape, is differentiable, and thus is a good substitute for the sign function. Meanwhile, to encourage sparsity, hierarchical Gaussian-inverse-Gamma priors [10], [11] are assigned to both $\boldsymbol{x}$ and $\boldsymbol{w}$, i.e.

$$p(\boldsymbol{x}|\alpha) = \prod_{i=1}^{n} p(x_i|\alpha_i) = \prod_{i=1}^{n} \mathcal{N}(x_i|0,\alpha_i^{-1})$$
$$p(\alpha) = \prod_{i=1}^{n} \text{Gamma}(\alpha_i|a,b) = \prod_{i=1}^{n} \Gamma(a)^{-1} b^a \alpha_i^{a-1} e^{-b\alpha_i}$$

and

$$p(\boldsymbol{w}|\beta) = \prod_{i=1}^{m} p(w_i|\beta_i) = \prod_{i=1}^{m} \mathcal{N}(w_i|0,\beta_i^{-1})$$
$$p(\beta) = \prod_{i=1}^{m} \text{Gamma}(\beta_i|c,d) = \prod_{i=1}^{m} \Gamma(c)^{-1} d^c \beta_i^{c-1} e^{-d\beta_i}$$

where $\mathcal{N}(x|\mu,\sigma^2)$ represents a Gaussian distribution with mean μ and variance σ^2. The parameters a, b, c, and d that are used to characterize the Gamma distributions are chosen to be very small values, e.g. 10^{-4}, in order to provide non-informative (over a logarithmic scale) hyperpriors over $\boldsymbol{\alpha} \triangleq \{\alpha_i\}$ and $\boldsymbol{\beta} \triangleq \{\beta_i\}$. As discussed in [10], a broad hyperprior allows the posterior means of α_i and β_i to become arbitrarily large. As a consequence, the associated coefficients x_i and w_i will be driven to zero, thus yielding a sparse solution.

III. Variational Inference

A. Review of The Variational Bayesian Methodology

Before proceeding, we firstly provide a brief review of the variational Bayesian methodology. In a probabilistic model, let $\boldsymbol{t}$ and $\boldsymbol{\theta}$ denote the observed data and the hidden variables, respectively. It is straightforward to show that the marginal probability of the observed data can be decomposed into two terms

$$\ln p(\boldsymbol{t}) = L(q) + \text{KL}(q||p) \tag{4}$$

where

$$L(q) = \int q(\boldsymbol{\theta}) \ln \frac{p(\boldsymbol{t},\boldsymbol{\theta})}{q(\boldsymbol{\theta})} d\boldsymbol{\theta} \tag{5}$$

and

$$\text{KL}(q||p) = -\int q(\boldsymbol{\theta}) \ln \frac{p(\boldsymbol{\theta}|\boldsymbol{t})}{q(\boldsymbol{\theta})} d\boldsymbol{\theta} \tag{6}$$

where $q(\boldsymbol{\theta})$ is any probability density function, $\text{KL}(q||p)$ is the Kullback-Leibler divergence between $p(\boldsymbol{\theta}|\boldsymbol{t})$ and $q(\boldsymbol{\theta})$. Since $\text{KL}(q||p) \geq 0$, it follows that $L(q)$ is a rigorous lower bound on $\ln p(\boldsymbol{t})$. Moreover, notice that the left hand side of (4) is independent of $q(\boldsymbol{\theta})$. Therefore maximizing $L(q)$ is equivalent to minimizing $\text{KL}(q||p)$, and thus the posterior distribution $p(\boldsymbol{\theta}|\boldsymbol{t})$ can be approximated by $q(\boldsymbol{\theta})$ through maximizing $L(q)$.

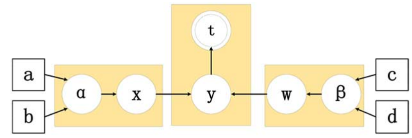


Fig. 1. Hierarchical model for joint sign-flip detection and sparse signal recovery.

The significance of the above transformation is that it circumvents the difficulty of computing the posterior probability $p(\boldsymbol{\theta}|\boldsymbol{t})$ (which is usually computationally intractable). For a suitable choice for the distribution $q(\boldsymbol{\theta})$, the quantity $L(q)$ may be more amiable to compute. 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 $\{\theta_i\}$ in $\boldsymbol{\theta}$ [12], i.e. $q(\boldsymbol{\theta}) = \prod_i q_i(\theta_i)$. We therefore 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 [12]

$$q_i(\theta_i) = \frac{\exp(\langle \ln p(\boldsymbol{t},\boldsymbol{\theta})\rangle_{k\neq i})}{\int \exp(\langle \ln p(\boldsymbol{t},\boldsymbol{\theta})\rangle_{k\neq i}) d\theta_i} \tag{7}$$

where $\langle\cdot\rangle_{k\neq i}$ denotes an expectation with respect to the distributions $q_i(\theta_i)$ for all $k \neq i$.

B. Proposed Bayesian Inference Algorithm

Let $\boldsymbol{\theta} \triangleq \{\boldsymbol{x},\boldsymbol{\alpha},\boldsymbol{w},\boldsymbol{\beta}\}$ denote all hidden variables appeared in our hierarchical model. We find an approximation of the posterior distribution $p(\boldsymbol{\theta}|\boldsymbol{t})$ through maximizing

$$L(q) = \int q(\boldsymbol{\theta}) \ln \frac{p(\boldsymbol{t},\boldsymbol{\theta})}{q(\boldsymbol{\theta})} d\boldsymbol{\theta} \tag{8}$$

where $p(\boldsymbol{t},\boldsymbol{\theta})$ is given by

$$p(\boldsymbol{t},\boldsymbol{\theta}) = p(\boldsymbol{t}|\boldsymbol{x},\boldsymbol{w})p(\boldsymbol{x}|\boldsymbol{\alpha})p(\boldsymbol{\alpha})p(\boldsymbol{w}|\boldsymbol{\beta})p(\boldsymbol{\beta}) \tag{9}$$

However, the integral in (8) is difficult to compute due to the sigmoid function $p(\boldsymbol{t}|\boldsymbol{x},\boldsymbol{w})$, which makes maximizing $L(q)$ a tricky problem. To circumvent this difficulty, we search for a tractable lower bound on $L(q)$. Recalling the Jaakkola-Jordon inequality [13]

$$\begin{aligned}\sigma(y)^t[1-\sigma(y)]^{1-t} &= \sigma(z)\\ &\geq \sigma(\delta)\exp\left(\frac{z-\delta}{2} - \lambda(\delta)(z^2-\delta^2)\right)\end{aligned} \tag{10}$$

where $z = (2t-1)y$, $\lambda(\delta) = (1/4\delta)\tanh(\delta/2)$, $\tanh(x) \triangleq (\exp(x)-\exp(-x))/(\exp(x)+\exp(-x))$ is the hyperbolic tangent function, and the inequality becomes equality when $\delta = z$. By utilizing (10), we have

$$\begin{aligned}p(\boldsymbol{t}|\boldsymbol{x},\boldsymbol{w}) &\geq F(\boldsymbol{t},\boldsymbol{x},\boldsymbol{w},\boldsymbol{\delta})\\ &\triangleq \prod_{i=1}^{m} \sigma(\delta_i)\exp\left(\frac{z_i-\delta_i}{2} - \lambda(\delta_i)(z_i^2-\delta_i^2)\right)\end{aligned} \tag{11}$$

where $z_i \triangleq (2t_i - 1)(\boldsymbol{a}_i^T\boldsymbol{x} + w_i)$, $\boldsymbol{a}_i$ denotes the transpose of the ith row of the sampling matrix $\boldsymbol{A}$, and $\boldsymbol{\delta} \triangleq \{\delta_i\}$. Combining (8) and (11), a tractable lower bound on $L(q)$ can eventually be obtained as

$$L(q) \geq \tilde{L}(q, \boldsymbol{\delta}) = \int q(\boldsymbol{\theta}) \ln \frac{G(\boldsymbol{t}, \boldsymbol{\theta}, \boldsymbol{\delta})}{q(\boldsymbol{\theta})} d\boldsymbol{\theta} \tag{12}$$

where

$$G(\boldsymbol{t}, \boldsymbol{\theta}, \boldsymbol{\delta}) \triangleq F(\boldsymbol{t}, \boldsymbol{x}, \boldsymbol{w}, \boldsymbol{\delta}) p(\boldsymbol{x}|\boldsymbol{\alpha}) p(\boldsymbol{\alpha}) p(\boldsymbol{w}|\boldsymbol{\beta}) p(\boldsymbol{\beta})$$

Let $q(\boldsymbol{\theta}) = q_x(\boldsymbol{x}) q_\alpha(\boldsymbol{\alpha}) q_w(\boldsymbol{w}) q_\beta(\boldsymbol{\beta})$ denote the factorized form of $q(\boldsymbol{\theta})$. Our objective is to maximize $\tilde{L}(q, \boldsymbol{\delta})$ with respect to the functions $q_x(\boldsymbol{x})$, $q_\alpha(\boldsymbol{\alpha})$, $q_w(\boldsymbol{w})$, and $q_\beta(\boldsymbol{\beta})$ as well as with respect to the parameters $\boldsymbol{\delta}$. This naturally leads to a variational expectation-maximization (EM) algorithm. In the E-step, the posterior distribution approximations are computed in an alternating fashion for each hidden variable, with other variables fixed. In the M-step, $\tilde{L}(q, \boldsymbol{\delta})$ is maximized with respect to $\boldsymbol{\delta}$, given the posterior distribution $q(\boldsymbol{\theta})$ fixed. Details of this Bayesian inference scheme are provided below.

E-Step:

1. ***Update of*** $q_x(\boldsymbol{x})$: Recalling (7), the approximate posterior distribution $q_x(\boldsymbol{x})$ can be computed by

$$\begin{aligned}
\ln q_x(\boldsymbol{x}) &\propto \langle \ln G(\boldsymbol{t}, \boldsymbol{\theta}, \boldsymbol{\delta}) \rangle_{q_\alpha(\boldsymbol{\alpha}) q_w(\boldsymbol{w}) q_\beta(\boldsymbol{\beta})} \\
&\propto \langle \ln F(\boldsymbol{t}, \boldsymbol{x}, \boldsymbol{w}, \boldsymbol{\delta}) + \ln p(\boldsymbol{x}|\boldsymbol{\alpha}) \rangle_{q_\alpha(\boldsymbol{\alpha}) q_w(\boldsymbol{w})} \\
&\propto \left\langle \sum_{i=1}^{m} \left(\frac{z_i}{2} - \lambda(\delta_i) z_i^2 \right) - \frac{1}{2} \boldsymbol{x}^T \boldsymbol{\Lambda}_\alpha \boldsymbol{x} \right\rangle_{q_\alpha(\boldsymbol{\alpha}) q_w(\boldsymbol{w})} \\
&\propto -\boldsymbol{x}^T \boldsymbol{A}^T \boldsymbol{\Lambda}_\delta \boldsymbol{A} \boldsymbol{x} + \frac{1}{2} (2\boldsymbol{t} - \boldsymbol{1})^T \boldsymbol{A} \boldsymbol{x} \\
&\quad - 2 \langle \boldsymbol{x}^T \boldsymbol{A}^T \boldsymbol{\Lambda}_\delta \boldsymbol{w} \rangle_{q_w(\boldsymbol{w})} - \frac{1}{2} \langle \boldsymbol{x}^T \boldsymbol{\Lambda}_\alpha \boldsymbol{x} \rangle_{q_\alpha(\boldsymbol{\alpha})}
\end{aligned} \tag{13}$$

where $\boldsymbol{\Lambda}_\alpha \triangleq \text{diag}(\alpha_1, \ldots, \alpha_n)$ and $\boldsymbol{\Lambda}_\delta \triangleq \text{diag}(\lambda(\delta_1), \ldots, \lambda(\delta_m))$. It can be readily observed that $q_x(\boldsymbol{x})$ follows a Gaussian distribution with its mean $\boldsymbol{\mu}_x$ and covariance matrix $\boldsymbol{\Phi}_x$ given respectively as

$$\begin{aligned}
\boldsymbol{\mu}_x &= \boldsymbol{\Phi}_x \boldsymbol{A}^T \left(\frac{1}{2} (2\boldsymbol{t} - \boldsymbol{1}) - 2 \boldsymbol{\Lambda}_\delta \boldsymbol{\mu}_w \right) \\
\boldsymbol{\Phi}_x &= (\boldsymbol{\Lambda}_{\langle \alpha \rangle} + 2 \boldsymbol{A}^T \boldsymbol{\Lambda}_\delta \boldsymbol{A})^{-1}
\end{aligned} \tag{14}$$

in which $\boldsymbol{\Lambda}_{\langle \alpha \rangle} \triangleq \text{diag}(\langle \alpha_1 \rangle, \ldots, \langle \alpha_n \rangle)$, $\langle \alpha_i \rangle$ denotes the expectation of α_i with respect to the distribution of $q_\alpha(\boldsymbol{\alpha})$, and $\boldsymbol{\mu}_w$ denotes the mean of the posterior distribution $q_w(\boldsymbol{w})$.

2. ***Update of*** $q_w(\boldsymbol{w})$: The posterior approximation $q_w(\boldsymbol{w})$ can be obtained as

$$\begin{aligned}
\ln q_w(\boldsymbol{w}) &\propto \langle \ln G(\boldsymbol{t}, \boldsymbol{\theta}, \boldsymbol{\delta}) \rangle_{q_x(\boldsymbol{x}) q_\alpha(\boldsymbol{\alpha}) q_\beta(\boldsymbol{\beta})} \\
&\propto \langle \ln F(\boldsymbol{t}, \boldsymbol{x}, \boldsymbol{w}, \boldsymbol{\delta}) + \ln p(\boldsymbol{w}|\boldsymbol{\beta}) \rangle_{q_x(\boldsymbol{x}) q_\beta(\boldsymbol{\beta})} \\
&\propto \frac{1}{2} (2\boldsymbol{t} - \boldsymbol{1})^T \boldsymbol{w} - \boldsymbol{w}^T \boldsymbol{\Lambda}_\delta \boldsymbol{w} \\
&\quad - 2 \boldsymbol{\mu}_x^T \boldsymbol{A}^T \boldsymbol{\Lambda}_\delta \boldsymbol{w} - \frac{1}{2} \boldsymbol{w}^T \boldsymbol{\Lambda}_{\langle \beta \rangle} \boldsymbol{w}
\end{aligned} \tag{15}$$

where $\boldsymbol{\Lambda}_{\langle \beta \rangle} \triangleq \text{diag}(\langle \beta_1 \rangle, \ldots, \langle \beta_n \rangle)$, $\langle \beta_i \rangle$ denotes the expectation of β_i with respect to the distribution of $q_\beta(\boldsymbol{\beta})$. Again, we can easily verify that $q_w(\boldsymbol{w})$ follows a Gaussian distribution with its mean and covariance matrix given by

$$\begin{aligned}
\boldsymbol{\mu}_w &= \boldsymbol{\Phi}_w \left(\frac{1}{2} (2\boldsymbol{t} - \boldsymbol{1}) - 2 \boldsymbol{\Lambda}_\delta \boldsymbol{A} \mu_x \right) \\
\boldsymbol{\Phi}_w &= (\boldsymbol{\Lambda}_{\langle \beta \rangle} + 2 \boldsymbol{\Lambda}_\delta)^{-1}
\end{aligned} \tag{16}$$

3. ***Update of*** $q_\alpha(\boldsymbol{\alpha})$: The posterior $q_\alpha(\boldsymbol{\alpha})$ can be obtained by computing

$$\begin{aligned}
\ln q_\alpha(\boldsymbol{\alpha}) &\propto \langle \ln G(\boldsymbol{t}, \boldsymbol{\theta}, \boldsymbol{\delta}) \rangle_{q_x(\boldsymbol{x}) q_w(\boldsymbol{w}) q_\beta(\boldsymbol{\beta})} \\
&\propto \langle \ln p(\boldsymbol{x}|\boldsymbol{\alpha}) + \ln p(\boldsymbol{\alpha}) \rangle_{q_x(\boldsymbol{x})} \\
&\propto \sum_{i=1}^{n} \left\{ \left(a - \frac{1}{2} \right) \ln \alpha_i - \left(\frac{\langle x_i^2 \rangle}{2} + b \right) \alpha_i \right\}
\end{aligned} \tag{17}$$

where $\langle x_i^2 \rangle$ denotes the expectation of x_i^2 with respect to $q_x(\boldsymbol{x})$. Hence $\boldsymbol{\alpha}$ has a form of a product of Gamma distributions

$$q_\alpha(\boldsymbol{\alpha}) = \prod_{i=1}^{n} \text{Gamma}(\alpha_i; \tilde{a}, \tilde{b}_i) \tag{18}$$

in which the parameters $\tilde{a}$ and $\tilde{b}_i$ are respectively given as

$$\tilde{a} = a + \frac{1}{2} \qquad \tilde{b}_i = b + \frac{1}{2} \langle x_i^2 \rangle \tag{19}$$

4. ***Update of*** $q_\beta(\boldsymbol{\beta})$: Similarly, variational optimization of $q_\beta(\boldsymbol{\beta})$ yields

$$\begin{aligned}
\ln q_\beta(\boldsymbol{\beta}) &\propto \langle \ln G(\boldsymbol{t}, \boldsymbol{\theta}, \boldsymbol{\delta}) \rangle_{q_x(\boldsymbol{x}) q_w(\boldsymbol{w}) q_\alpha(\boldsymbol{\alpha})} \\
&\propto \langle \ln p(\boldsymbol{w}|\boldsymbol{\beta}) + \ln p(\boldsymbol{\beta}) \rangle_{q_w(\boldsymbol{w})} \\
&\propto \sum_{i=1}^{m} \left\{ \left(c - \frac{1}{2} \right) \ln \beta_i - \left(\frac{\langle w_i^2 \rangle}{2} + d \right) \beta_i \right\}
\end{aligned} \tag{20}$$

Thus $\boldsymbol{\beta}$ has a form of a product of Gamma distributions as well

$$q_\beta(\boldsymbol{\beta}) = \prod_{i=1}^{m} \text{Gamma}(\beta_i; \tilde{c}, \tilde{d}_i) \tag{21}$$

with the parameters $\tilde{c}$ and $\tilde{d}_i$ given as

$$\tilde{c} = c + \frac{1}{2} \qquad \tilde{d}_i = b + \frac{1}{2} \langle w_i^2 \rangle \tag{22}$$

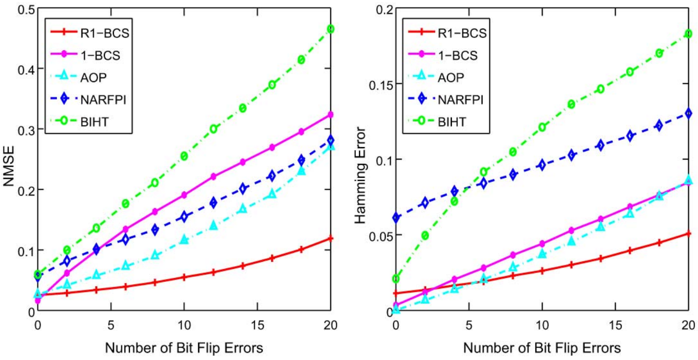
In summary, the E-step involves update of the posterior approximations for hidden variables $\boldsymbol{x}, \boldsymbol{w}, \boldsymbol{\alpha}$, and $\boldsymbol{\beta}$. Some of the expectations and moments used during the update are summarized as

$$\begin{aligned}
&\langle \alpha_i \rangle = \frac{\tilde{a}}{\tilde{b}_i} \quad \langle \beta_i \rangle = \frac{\tilde{c}}{\tilde{d}_i} \\
&\langle x_i^2 \rangle = \mu_{x,i}^2 + \phi_x^{(i,i)} \quad \langle w_i^2 \rangle = \mu_{w,i}^2 + \phi_w^{(i,i)}
\end{aligned} \tag{23}$$

where $\mu_{x,i}$ and $\mu_{w,i}$ denote the ith element of $\boldsymbol{\mu}_x$ and $\boldsymbol{\mu}_w$, respectively; $\phi_x^{(i,i)}$ and $\phi_w^{(i,i)}$ represent the ith diagonal element of $\boldsymbol{\Phi}_x$ and $\boldsymbol{\Phi}_w$, respectively.

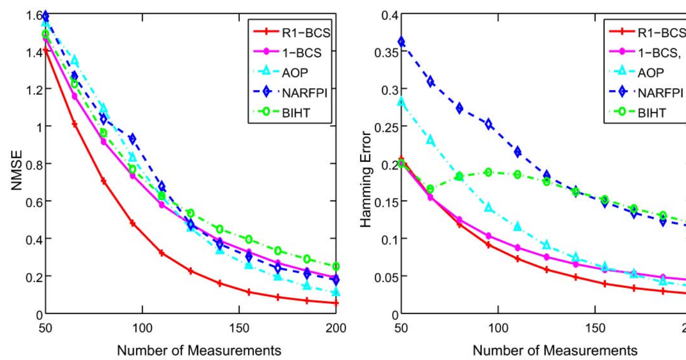
M-Step: By substituting $q(\boldsymbol{\theta}; \boldsymbol{\delta}^{\text{old}})$ into $\tilde{L}(q, \boldsymbol{\delta})$, the estimate of $\boldsymbol{\delta}$ can be found via the following optimization

$$\boldsymbol{\delta}^{\text{new}} = \arg\max_{\boldsymbol{\delta}} \quad \langle \ln G(\boldsymbol{t}, \boldsymbol{\theta}, \boldsymbol{\delta}) \rangle_{q(\boldsymbol{\theta}; \boldsymbol{\delta}^{\text{old}})} \triangleq \tilde{Q}(\boldsymbol{\delta}|\boldsymbol{\delta}^{\text{old}}) \tag{24}$$



(a) (b)

Fig. 2. (a) NMSEs of respective algorithms vs. L; (b) Hamming errors of respective algorithms vs. L.



(a) (b)

Fig. 3. (a) NMSEs of respective algorithms vs. m; (b) Hamming errors of respective algorithms vs. m.

Taking the derivative of $\tilde{Q}(\boldsymbol{\delta}|\boldsymbol{\delta}^{\text{old}})$ with respect to each variable δ_i yields

$$\frac{\partial \tilde{Q}(\boldsymbol{\delta}|\boldsymbol{\delta}^{\text{old}})}{\partial \delta_i} = \frac{\partial \lambda(\delta_i)}{\partial \delta_i}(\delta_i^2 - \langle z_i^2 \rangle_{q(\boldsymbol{\theta};\boldsymbol{\delta}^{\text{old}})}) \tag{25}$$

Setting the derivative to zero gives the solution

$$\begin{aligned}\delta_i^2 &= \langle z_i^2 \rangle_{q(\boldsymbol{\theta};\boldsymbol{\delta}^{\text{old}})} \\ &= \boldsymbol{a}_i^T \langle \boldsymbol{x}\boldsymbol{x}^T \rangle_{q_x(\boldsymbol{x})} \boldsymbol{a}_i + \mu_{w,i}^2 + \phi_w^{(i,i)} + 2\boldsymbol{a}_i^T \boldsymbol{\mu}_x \mu_{w,i}\end{aligned} \tag{26}$$

where $\langle \boldsymbol{x}\boldsymbol{x}^T \rangle = \boldsymbol{\mu}_x \boldsymbol{\mu}_x^T + \boldsymbol{\Phi}_x$.

For clarity, we summarize the variational Bayesian algorithm as follows.

1) Given the current estimate of $\boldsymbol{\delta}$, update the posterior approximations $q_x(\boldsymbol{x})$, $q_w(\boldsymbol{w})$, $q_\alpha(\boldsymbol{\alpha})$, and $q_\beta(\boldsymbol{\beta})$ according to (14), (16), (19), and (22).
2) Update the parameter $\boldsymbol{\delta}$ according to (26).
3) Continue the above iteration until $\|\boldsymbol{\mu}_x^{(t)} - \boldsymbol{\mu}_x^{(t-1)}\|_2 \leq \varepsilon$, where ε is a prescribed tolerance value.

IV. Simulation Results

We now carry out experiments to illustrate the performance of our proposed robust one-bit Bayesian compressed sensing algorithm (referred to as R1-BCS)[1]. In our simulations, the K-sparse signal is randomly generated with its support set randomly chosen according to a uniform distribution. The measurement matrix $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ is randomly generated with each entry independently drawn from Gaussian distribution with zero mean and unit variance, and then each column of $\boldsymbol{A}$ is normalized to unit norm. The sign-flip errors are also randomly generated according to a uniform distribution. We compare our proposed method with the BIHT method [4], the one-bit Bayesian compressed sensing algorithm (1-BCS) which is a simplified version of the proposed R1-BCS method developed without considering the sign-flip errors (i.e. the sparse noise is removed from our proposed hierarchical model), the adaptive outlier pursuit method (AOP) [8], and the noise-adaptive renormalized fixed point iteration (NARFPI) method [9]. Note that the AOP and the BIHT require the knowledge of the sparsity level of the signal, which is assumed perfectly known to these methods.

Two metrics are used to evaluate the recovery performance, namely, the normalized mean squared error (NMSE) and the Hamming error. Since the information about the magnitude of the signal is lost due to one-bit quantization, the norm of the original signal and the estimated signal are normalized to unity in computing the NMSEs. The Hamming error is defined as $(1/m)\|\text{sign}(\boldsymbol{A}\hat{\boldsymbol{x}}) - \text{sign}(\boldsymbol{A}\boldsymbol{x})\|_0$. Clearly, the Hamming error will become zero if the estimated sign measurements are consistent with the original sign measurements, which means that all bit-flip errors are detected and corrected. We first examine the robustness of respective algorithms against the sign-flip errors. Note that the AOP and NARFPI methods require to pre-specify the total number of sign-flip errors, L, in their formulations. We assume that L is perfectly known by them. Fig. 2 depicts the NMSEs and the Hamming errors as a function of L, where we set $m = 200$, $n = 100$, and $K = 10$. Results are averaged over 10^3 independent runs. From Fig. 2, we see that the methods which take the flip errors into account generally outperform those methods (BIHT and 1-BCS) ignoring the flip errors. Particularly, the recovery performance of the 1-BCS method deteriorates significantly as the number of flip errors increases, while the proposed R1-BCS method is robust against the flip errors and incurs only mild performance loss with a growing L. It can also be observed that the proposed algorithm achieves the best performance in terms of both the NMSE and the Hamming error. Fig. 3 plots the NMSEs and the Hamming errors vs. the number of measurements m, where we set $n = 100$, $K = 10$, and $L = 10$. This result again demonstrates the superiority of the proposed algorithm over other existing methods.

V. Conclusions

We developed a robust one-bit Bayesian compressed sensing algorithm for joint sign-flip error detection and sparse signal recovery. A noteworthy merit of the proposed algorithm is that it does not need any tuning parameters (such as the sparsity level and the number of sign-flip errors). Simulation results show that the proposed algorithm is robust against the sign-flip errors and provides superior recovery performance as compared with other methods.

[1]Codes are available at http://www.junfang-uestc.net/codes/R1-BCS.rar

References

[1] J. N. Laska, Z. Wen, W. Yin, and R. G. Baraniuk, "Trust, but verify: Fast and accurate signal recovery from 1-bit compressive measurements," *IEEE Trans. Signal Process.*, vol. 59, no. 11, pp. 5289–5301, Nov. 2011.

[2] P. T. Boufounos and R. G. Baraniuk, "One-bit compressive sensing," in *Proc. the 42nd Annu. Conf. Information Sciences and Systems*, Princeton, NJ, USA, 2008.

[3] P. T. Boufounos, "Greedy sparse signal reconstruction from sign measurements," in *Proc. 43rd Asilomar Conf. Signals, Systems, and Computers*, Pacific Grove, CA, USA, 2009.

[4] L. Jacques, J. N. Laska, P. T. Boufounos, and R. G. Baraniuk, "Robust 1-bit compressive sensing via binary stable embeddings of sparse vectors," *IEEE Trans. Inf. Theory*, vol. 59, no. 4, pp. 2082–2102, Apr. 2013.

[5] A. Bourquard, F. Aguet, and M. Unser, "Optical imaging using binary sensors," *Opt. Exp.*, vol. 18, no. 5, pp. 4876–4888, Mar. 2010.

[6] Y. Plan and R. Vershynin, "One-bit compressed sensing by linear programming," Sep. 2011 [Online]. Available: http://arxiv.org/abs/1109.4299

[7] Y. Shen, J. Fang, H. Li, and Z. Chen, "A one-bit reweighted iterative algorithm for sparse signal recovery," in *IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, Vancouver, BC, Canada, May 26–31, 2013.

[8] M. Yan, Y. Yang, and S. Osher, "Robust 1-bit compressive sensing using adaptive outlier pursuit," *IEEE Trans. Signal Process.*, vol. 60, no. 7, pp. 3868–3875, Jul. 2012.

[9] A. Movahed, A. Panahi, and G. Durisi, "A robust RFPI-based 1-bit compressive sensing reconstruction algorithm," in *IEEE Information Theory Workshop (ITW)*, Lausanne, Sep. 3–7, 2012.

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

[11] S. Ji, Y. Xue, and L. Carin, "Bayesian compressive sensing," *IEEE Trans. Signal Process.*, vol. 56, no. 6, pp. 2346–2356, Jun. 2008.

[12] D. G. Tzikas, A. C. Likas, and N. P. Galatsanos, "The variational approximation for bayesian inference," *IEEE Signal Process. Mag.*, pp. 131–146, Nov. 2008.

[13] C. M. Bishop and M. E. Tipping, "Variational relevance vector machine," in *Proc. the 16th Conf. Uncertainty in Artificial Intelligence*, 2000, pp. 46–53.