Digital Signal Processing 50 (2016) 84–92

Contents lists available at ScienceDirect

Digital Signal Processing

www.elsevier.com/locate/dsp

Robust 1-bit compressive sensing via variational Bayesian algorithm

Chongbin Zhou [a,b,*], Zhida Zhang [a,b], Falin Liu [a,b,*]

[a] Department of EEIS, University of Science and Technology of China, Hefei 230027, China
[b] Key Laboratory of Electromagnetic Space Information, Chinese Academy of Sciences, Hefei 230027, China

ARTICLE INFO

Article history:
Available online 29 December 2015

Keywords:
1-Bit quantization
Compressive sensing
Sparse Bayesian learning
Variational message passing

ABSTRACT

In a compressive sensing (CS) framework, a sparse signal can be stably reconstructed at a reduced sampling rate. Quantization and noise corruption are inevitable in practical applications. Recent studies have shown that using only the sign information of measurements can achieve accurate signal reconstruction in a CS framework. We consider the problem of reconstructing a sparse signal from 1-bit quantized, Gaussian noise corrupted measurements. In this paper, we present a variational Bayesian inference based 1-bit compressive sensing algorithm, which essentially models the effect of quantization as well as the Gaussian noise. A variational message passing method is adopted to achieve the inference. Through numerical experiments, we demonstrate that our algorithm outperforms state-of-the-art 1-bit compressive sensing algorithms in the presence of Gaussian noise corruption.

1. Introduction

In a compressive sensing (CS) framework, the goal is to recover sparse signals from a small number of measurement samples [1,2]. The measurement of a signal $\boldsymbol{x} \in R^N$ is obtained via

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} \tag{1}$$

where $\boldsymbol{A} \in R^{M\times N}$ is the measurement matrix with $M < N$, and $\boldsymbol{y} \in R^M$ is the linear measurement of the signal. The recovery of $\boldsymbol{x}$ from $\boldsymbol{y}$ is generally ill-posed; however, it was demonstrated that if $\boldsymbol{x}$ is K-sparse, i.e., no more than $K \ll N$ entries of $\boldsymbol{x}$ are non-zero, the signal can be reconstructed exactly and effectively if $\boldsymbol{A}$ satisfies the restricted isometry property (RIP) [3].

The classical CS framework always assumes that the measurement vector $\boldsymbol{y}$ is real-valued and has infinite bit precision. For practical considerations, however, the measurements must be quantized to finite bit depth, i.e., each continuous-valued measurement must be mapped to a discrete value in a finite set. The effect of quantization has been studied [4–6].

Recent studies have shown that stable signal reconstruction can be achieved even if each measurement is quantized to a single bit. In many applications, the 1-bit quantization has significant benefits. For example, in analog-to-digital conversion (ADC), the quantizer for 1-bit measurement is a simple comparator, which is fast, inexpensive and robust to amplification distortion. In this case, we have

$$\boldsymbol{y} = sign(\boldsymbol{z}) = sign(\boldsymbol{A}\boldsymbol{x}) \tag{2}$$

where $\boldsymbol{z} \in R^M$ stands for the signal before quantization, and the $sign(\cdot)$ operator performs the sign function element-wise on the vector $\boldsymbol{z}$, which returns $+1$ for positive numbers and -1 otherwise. In this case, the scaling information of the measurements is wholly lost. Then, the goal is to recover the signal on the unit hyper-sphere.

The 1-bit CS framework was first studied by Boufounos and Baraniuk, and in [7], they proposed an algorithm named renormalized fixed point iteration (RFPI). Since then, many studies have been performed, and many algorithms have been developed, including matching sign pursuit (MSP) [8], restricted-step shrinkage (RSS) [9], and binary iterative hard thresholding (BIHT) [10]. BIHT has been shown to perform better than the previous algorithms and is robust to sign flips. According to the experiments in [10], the one-sided l_1 objective (BIHT) performs better when the noise level is low, while the one-sided l_2 objective (BIHT-l_2) is more suitable when more measurements flip their signs with increased noise. In [11], an adaptive outlier pursuit (AOP) method was introduced to handle sign flips by adaptively detecting all the measurements with sign flips. In [12], a linear program was proposed by Plan and Vershynin to address the noiseless 1-bit CS problem, and in [13], they introduced another convex program for the noisy case. In [14], a variational Bayesian algorithm was proposed to handle the sign flips caused by Gaussian noise, and this Bayesian method outperforms both BIHT and the convex program in [13]. In this Bayesian method, the effect of quantization is modeled as additive noise that is independent of the signal. However, there is an inherent relevance between the quantization noise and the signal,

* Corresponding authors.
E-mail addresses: zhouzcb@mail.ustc.edu.cn (C. Zhou), liufl@ustc.edu.cn (F. Liu).

http://dx.doi.org/10.1016/j.dsp.2015.12.006
1051-2004/

causing inaccuracy of the model. More recently, Fang et al. [15] introduced a sigmoid function to enforce the consistency and use a majorization-minimization (MM) [16] based method to handle the noise-free case.

In this paper, we propose a robust variational Bayesian framework to address the 1-bit CS problem with Gaussian noise corruption. Unlike the method in [14], we essentially model the effect of quantization as well as the Gaussian noise. A two-layer hierarchical prior is adopted to encourage the sparsity of the signal, and a variational message passing method [17] is performed to complete the Bayesian inference.

This paper is organized as follows. Section 2 introduces the new framework for 1-bit CS. Section 3 introduces the binary variational message passing (B-VMP) algorithm. The performance of the algorithm is illustrated in Section 4 through comparison to the state-of-the-art algorithms. We conclude this paper in Section 5.

2. A new framework for 1-bit CS

In this paper, we consider the case in which the linear measurements are corrupted with Gaussian noise before quantization:

$$\boldsymbol{y} = sign(\boldsymbol{z}), \quad \boldsymbol{z} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{n} \tag{3}$$

where $\boldsymbol{x}$ is the sparse signal of interest, $\boldsymbol{y}$ is the 1-bit quantized measurement vector, $\boldsymbol{z}$ is the noisy measurement vector before quantization, and $\boldsymbol{n}$ denotes the Gaussian noise. As the magnitude information of $\boldsymbol{x}$ is lost during the quantization, we restrict the solution to the unit hyper-sphere

$$D_{\boldsymbol{x}} = \left\{\boldsymbol{x} \middle| \|\boldsymbol{x}\|_2 = 1\right\} \tag{4}$$

For the sparse prior of signal $\boldsymbol{x}$, we adopt a two-layer hierarchical prior

$$p(\boldsymbol{x}; a, b) = \int p(\boldsymbol{x}|\boldsymbol{\alpha}) p(\boldsymbol{\alpha}|a, b) d\boldsymbol{\alpha} \tag{5}$$

where

$$p(\boldsymbol{x}|\boldsymbol{\alpha}) \propto \mathcal{N}(\boldsymbol{x}|\boldsymbol{0}, \boldsymbol{\Lambda}) \cdot I_D(\boldsymbol{x}) \tag{6}$$

$$p(\boldsymbol{\alpha}|a, b) = \prod_{i=1}^{N} \Gamma(a_i|a, b) \tag{7}$$

with $\boldsymbol{\Lambda} = diag(\boldsymbol{\alpha}^{-1})$ and an indicator function $I_D(\boldsymbol{x})$ that is equal to 1 if $\boldsymbol{x} \in D$ and 0 otherwise. $\mathcal{N}(\boldsymbol{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma})$ is a Gaussian distribution with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$. For a variable $u \sim \Gamma(a, b)$, the probability density function (PDF) is $\Gamma(u|a, b) = \Gamma(a)^{-1} b^a u^{a-1} \exp(-bu)$, where $\Gamma(a)$ denotes the Gamma function. In this paper, we follow the studies in [18] that argued that $a = 1$, $b = 0$ encourage sparser solutions than $a = 0$, $b = 0$, although the latter are commonly used in sparse Bayesian learning frameworks [19]. Readers can refer to [18] for more information.

We assume that the noise variance σ^2 is known, as in [14], because the estimate of σ^2 can be inaccurate, as addressed in [20]. Recalling (3), the conditional likelihood of $\boldsymbol{y}$ given $\boldsymbol{x}$ and σ can be written as

$$p(\boldsymbol{y}|\boldsymbol{x}, \sigma) = \prod_{i=1}^{M} \Phi\left(\sigma^{-1}(\boldsymbol{B}\boldsymbol{x})_i\right) \tag{8}$$

where

$$\boldsymbol{B} = diag(\boldsymbol{y}) \cdot \boldsymbol{A} \tag{9}$$

and

$$\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} \exp\left(-\frac{t^2}{2}\right) dt \tag{10}$$

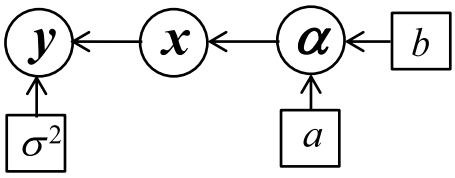

Fig. 1. Graphical hierarchical model of the proposed Bayesian model.

is the cumulative distribution function of the standard normal distribution. Finally, we can write the joint PDF of the observation model

$$p(\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{\alpha}) = p(\boldsymbol{y}|\boldsymbol{x}) p(\boldsymbol{x}|\boldsymbol{\alpha}) p(\boldsymbol{\alpha}) \tag{11}$$

The formulation of this problem is very similar to the sparse probit regression [5] or sparse logistic regression [13]. The hierarchical model is shown in Fig. 1.

3. B-VMP: binary variational message passing for CS

3.1. Bayesian formulation of B-VMP

In a Bayesian framework, the goal is to maximize the posterior distribution given the observed value $p(\boldsymbol{x}, \boldsymbol{\alpha}|\boldsymbol{y}) = p(\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{\alpha})/p(\boldsymbol{y})$. However, this posterior distribution is intractable for exact calculation in practice because $p(\boldsymbol{y}) = \int\int p(\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{\alpha}) d\boldsymbol{x} d\boldsymbol{\alpha}$ cannot be integrated in closed form.

In this paper, a variational inference approach [17] is introduced to achieve Bayesian inference. We denote by $V = \{\boldsymbol{y}\}$ and $H = \{\boldsymbol{x}, \boldsymbol{\alpha}\}$ the visible data and the hidden variables in the Bayesian network, respectively. Then, the goal is to find a tractable variational distribution $Q(H)$ that closely approximates the true posterior distribution $p(H|V)$. The log marginal probability of the observed data can be split into two terms:

$$\ln p(V) = L(Q) + KL(Q||P) \tag{12}$$

where

$$L(Q) = \int Q(H) \ln \frac{p(H, V)}{Q(H)} dH \tag{13}$$

$$KL(Q||P) = -\int Q(H) \ln \frac{p(H|V)}{Q(H)} dH \tag{14}$$

Here, $KL(Q||P)$ is the Kullback–Leibler divergence between $p(H|V)$ and the approximation $Q(H)$. Because $KL(Q||P) \geq 0$ holds, $L(Q)$ forms a lower bound on $p(V)$, and maximizing $L(Q)$ is equivalent to minimizing $KL(Q||P)$. If $Q(H)$ can have complete flexibility, maximizing $L(Q)$ leads to the true posterior $Q(H) = p(H|V)$. However, this approach always leads to computational intractability. A commonly used variational distribution $Q(H)$ has a factorized form $Q(H) = Q(\boldsymbol{x})Q(\boldsymbol{\alpha})$ in which disjoint groups of variables are independent. In variational message passing, $Q(\boldsymbol{x})$ and $Q(\boldsymbol{\alpha})$ are updated iteratively to monotonically decrease the KL divergence.

1) Update of $Q(\boldsymbol{\alpha})$: According to [17], we have

$$\ln Q_i(\alpha_i) = \left\langle \ln p(x_i|\alpha_i) \right\rangle_{Q(x_i)} + \ln p(\alpha_i|a, b) + const \tag{15}$$

where the subscript $Q(x_i)$ denotes an expectation with respect to $Q(x_i)$. Substituting (6) and (7) into (15), we have

$$\ln Q_i(\alpha_i) = \begin{pmatrix} -\frac{1}{2}\langle x_i^2 \rangle_{Q(x_i)} - b \\ a - \frac{1}{2} \end{pmatrix}^T \cdot \begin{pmatrix} \alpha_i \\ \ln \alpha_i \end{pmatrix} + const \tag{16}$$

The superscript T denotes the transpose operator. Thus, we can obtain