

RESEARCH ARTICLE

1-Bit Compressed Sensing via an L1-TV Regularization Method

JINGYAO HOU¹ AND XINLING LIU^{1,2}¹Key Laboratory of Optimization Theory and Applications, School of Mathematics and Information, China West Normal University, Nanchong, Sichuan 637002, China²School of Mathematics and Statistics, Southwest University, Chongqing 400715, China

Corresponding author: Jingyao Hou (hgy17623226280@163.com)

This work was supported in part by the National Natural Science Foundation of China under Grant 12201505, and in part by the Fundamental Research Funds of China West Normal University under Grant 20B001.

ABSTRACT This paper considers the 1-bit compressed sensing (1-bit CS) of signals that are simultaneously sparse and gradient sparse. Since L1-norm and total variation (TV) penalties are beneficial for reconstructing element-sparse and gradient-sparse signals, respectively, we combine both to propose the L1-TV regularization model for 1-bit CS. We show that the proposed model has a closed-form solution so that one can easily calculate. Despite the apparent simplicity, our theoretical analysis reveals that the solution provides an approximate estimate for the underlying signal. Besides, in the case of introducing a dither vector, we develop an adaptive algorithm to accelerate the decay rate of recovery error. The key idea is that generating the dither for each iteration relying on the last estimate. In addition to theoretical analysis, we conduct a series of experiments on both synthetic and real-world data to show the superiority of our algorithms.

INDEX TERMS 1-bit compressed sensing, L1-norm, total variation, adaptivity.

I. INTRODUCTION

In the past decade, 1-bit CS [1], [2], [3], [4], [5], [6], [7] has drawn extensive attraction as a new paradigm in classic compressed sensing [8]. Instead of assuming infinite-precision real-valued measurements as in classic compressed sensing, 1-bit CS only retains the signs of real-valued measurements, leading to low-cost implementation in hardware and fast sampling speed. For the above reasons, it has a broad application prospect in massive multiple-input multiple-output systems [9], wireless sensor networks [10], synthetic aperture radar systems [11], and so on, where large-scale sparse data is usually involved.

Mathematically, 1-bit CS aims to recover the direction of the underlying sparse signal $\mathbf{x} \in \mathbb{R}^n$ (i.e., $\mathbf{x}/\|\mathbf{x}\|_2$) from binary measurements

$$\mathbf{y} = \text{sign}(\mathbf{A}\mathbf{x}), \quad (1)$$

The associate editor coordinating the review of this manuscript and approving it for publication was Frederico Guimarães ^{1D}.

where $\mathbf{A} \in \mathbb{R}^{m \times n}$ denotes the sensing matrix and $\text{sign}(\cdot)$ applies componentwise on $\mathbf{A}\mathbf{x}$, satisfying

$$\text{sign}(v) := \begin{cases} 1, & \text{if } v \geq 0, \\ -1, & \text{if } v < 0. \end{cases}$$

To this end, Jacques et al. [2] show that the solution to an L0-minimization problem provides a robust estimate of the original signal, where the L0-norm, $\|\mathbf{u}\|_0$, counts the number of non-zero entries in $\mathbf{u} \in \mathbb{R}^n$. They also design a binary iterative hard threshold (BIHT) algorithm to solve the problem approximately. Plan and Vershynin [3] employ the L1-norm as sparsity promotion to propose a convex programming approach and provide a theoretical guarantee for it. After that, in an early work [7], we propose an Lp ($0 < p < 1$)-minimization method and prove that this method requires fewer measurements than the L1-norm based counterpart. Unfortunately, the above-mentioned methods are computationally complex. They either involve iterative procedures or rely on some toolboxes for convex problems, e.g., CVX [12], resulting in low efficiency when processing high-dimensional

signals. To enhance the solving efficiency, Zhang et al. [4] suggest the following optimization model

$$\hat{\mathbf{u}} := \arg \min_{\|\mathbf{u}\|_2 \leq 1} -\frac{1}{m} \langle \mathbf{A}^T \mathbf{y}, \mathbf{u} \rangle + \lambda \|\mathbf{u}\|_1. \quad (2)$$

It has the closed-form solution $\mathcal{P}_\lambda(\mathbf{A}^T \mathbf{y}) / \|\mathcal{P}_\lambda(\mathbf{A}^T \mathbf{y})\|_2$ where

$$\mathcal{P}_\lambda(\mathbf{t}) := \text{sign}(\mathbf{t}) \odot \max(|\mathbf{t}| - \lambda, \mathbf{0}) \quad (3)$$

is the soft-thresholding operator with \odot denoting a Hadamard product.

On the other hand, 1-bit measurements (1) lose all information on the magnitude of \mathbf{x} (i.e., $\|\mathbf{x}\|_2$). For full recovery (both direction and magnitude), we shall consider the thresholded binary measurements [5]

$$\mathbf{y} = \text{sign}(\mathbf{A}\mathbf{x} + \boldsymbol{\tau}), \quad (4)$$

where $\boldsymbol{\tau} \in \mathbb{R}^m$ denotes a known dither vector. With this setting, References [5], [6], and [13] provide different selection strategies for $\boldsymbol{\tau}$. Specifically, Baraniuk et al. [6] show that an adaptive choice method of the thresholds used for quantization can significantly lower the error decay rate.

The above results rely on the sparsity prior assumption of the underlying signal, and they can be extended to signals synthesized by a linear combination of some atoms in a (redundant) dictionary with incoherent atoms [14]. However, there are numerous signals possessing other structure information that can not fall into the category in the aforementioned theoretical work. One representative example is the signal consisting of piecewise constants (i.e., its gradient is sparse), which arises widely in image restoration [15] and image reconstruction [16], [17]. Let $\mathbf{D}\mathbf{x} \in \mathbb{R}^{n-1}$ denote the finite difference of $\mathbf{x} \in \mathbb{R}^n$ defined by $[\mathbf{D}\mathbf{x}]_i = x_{i+1} - x_i$ for $i = 1, \dots, n-1$ where $\mathbf{D} \in \mathbb{R}^{(n-1) \times n}$ is a differential matrix. In fact, \mathbf{D} is extracted from the first $n-1$ rows of the row circulant matrix of $(-1, 1, 0, \dots, 0) \in \mathbb{R}^n$. The piecewise constant property of \mathbf{x} means that $\mathbf{D}\mathbf{x}$ is sparse. In this situation, to reconstruct \mathbf{x} , one usually borrows the idea of the element-sparse case to solve

$$\min_{\mathbf{x}} \|\mathbf{D}\mathbf{x}\|_1 \quad \text{subject to} \quad \mathbf{y} = \mathbf{A}\mathbf{x},$$

where $\mathbf{A} \in \mathbb{R}^{m \times n}$ is a sensing matrix and $\|\mathbf{D}\mathbf{x}\|_1$ is called the total variation of \mathbf{x} , denoted as $\|\mathbf{x}\|_{TV}$. The method of TV-minimization has been analyzed in recent years [18], [19], [20] for compressed sensing of gradient-sparse signals and some optimization algorithms are designed to solve it, such as proximal gradient descent algorithm and projected subgradient algorithm [21].

Besides separately sparse or gradient sparse signals, there also exists a type of signals, involving both of the two properties simultaneously, such as Electrocardiogram (ECG) signals [22] and some slow time-varying signals [23]. In this situation, neither the L1-term penalty nor the TV-term penalty can comprehensively character both properties, which inspires us to combine them. An intuitive combination

method is to use addition, namely solving the following problem

$$\min_{\mathbf{x}} \frac{1}{2} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 + \lambda_1 \|\mathbf{x}\|_1 + \lambda_2 \|\mathbf{D}\mathbf{x}\|_1, \quad (5)$$

where $\lambda_1, \lambda_2 > 0$ denote penalty parameters. The model relates to the fused LASSO model in statistics and has obtained some applications in biological research, such as gene classification [24] and coefficient selection for the colon tumor data set [25]. It has also been used for the compressed sensing of some slow time-varying signals [23], exhibiting a great potential.

In this paper, we aim to introduce the L1-TV penalty model in 1-bit CS framework to solve the reconstruction problem of signals possessing both element sparse and gradient sparse properties. We employ $-\frac{1}{m} \langle \mathbf{A}^T \mathbf{y}, \mathbf{x} \rangle$ as the loss function following the suggestion by [3] and combine it with the L1-penalty term and the TV-penalty term simultaneously as the regularizer, leading to the following optimization problem

$$\min_{\|\mathbf{x}\|_2 \leq 1} -\frac{1}{m} \langle \mathbf{A}^T \mathbf{y}, \mathbf{x} \rangle + \lambda_1 \|\mathbf{x}\|_1 + \lambda_2 \|\mathbf{D}\mathbf{x}\|_1, \quad (6)$$

where λ_1 and λ_2 are regularization parameters, whose value will be discussed later. We will provide the closed-form solution to (6) and prove that it is an approximate direction estimate of the underlying signal of both sparse and gradient sparse.

To our knowledge, in the literature, there are few related works in the field of 1-bit CS. Though References [26] and [27] have employed the L1-TV term penalty in their methods, they aim at the compressed sensing of block-sparse signals, leading their models to distinguish from ours. In addition, their methods are intuition-driven without providing theoretical analysis. The main contributions of this paper are listed below

- This paper first proposes an L1-TV method for 1-bit CS and shows that the signal simultaneously being element sparse and gradient sparse can be approximately estimated by solving the proposed optimization problem.
- The proposed model has a closed-form solution, ensuring it can be easily obtained. Furthermore, we introduce a quantization scheme that chooses the dither vector adaptively, thus exponentially accelerating the recovery error decay rate.
- We conduct experiments on both synthetic and real-world datasets where the experimental results show that the proposed algorithms perform better than their counterparts.

The rest of this paper is organized as follows. In Section II, we define some notations and give some important lemmas that will be utilized in the proof process of our main results. In Section III, we present the proposed methods with corresponding theoretical guarantees. Comparisons between different algorithms on synthetic data, ECG signals, and gray images are conducted in Section IV. Section V concludes the paper and presents some future research problems.

Appendix provides detailed proofs for the proposed lemmas and theorems.

II. NOTATIONS AND PRELIMINARY

A. NOTATIONS

We begin with some necessary notations used throughout the paper. Scalars are denoted with lowercase or capital letters (e.g., x or X), vectors with boldface lowercase letters (e.g., \mathbf{x}), and matrices with boldface capital letters (e.g., \mathbf{X}). We denote x_i or $(\mathbf{x})_i$ as the i -th element of the vector \mathbf{x} . The field of real numbers is denoted as \mathbb{R} . $\mathbb{S}^{n-1} = \{\mathbf{u} \in \mathbb{R}^n : \|\mathbf{u}\|_2 = 1\}$ denotes the unit hypersphere in \mathbb{R}^n and $\mathbb{B}^n = \{\mathbf{u} \in \mathbb{R}^n : \|\mathbf{u}\|_2 \leq 1\}$ denotes the unit ball in \mathbb{R}^n . $[n]$ denotes the index set $\{1, 2, \dots, n\}$. Throughout the paper, we denote $\mathbf{A} \in \mathbb{R}^{m \times n}$ as the sensing matrix and \mathbf{a}_i as its i -th row where $i \in [m]$. We define the Lp norm of $\mathbf{x} \in \mathbb{R}^n$ as $\|\mathbf{x}\|_p = (\sum_{i=1}^n |x_i|^p)^{1/p}$ where $p \geq 1$.

B. PRELIMINARIES

In this part, we define some operations and provide some lemmas which may be used in the following sections. First, we consider the optimization problem

$$\hat{\mathbf{x}} := \arg \min_{\mathbf{u} \in \mathbb{R}^n} \frac{1}{2} \|\mathbf{x} - \mathbf{u}\|_2^2 + \lambda \|\mathbf{u}\|_1. \quad (7)$$

As is known, (7) has the closed-form solution $\hat{\mathbf{x}} = \mathcal{P}_\lambda(\mathbf{x})$ where $\mathcal{P}_\lambda(\cdot)$ is defined in (3). Besides, for the TV-term penalty based optimization problem

$$\mathcal{T}_\lambda(\mathbf{x}) := \arg \min_{\mathbf{u} \in \mathbb{R}^n} \frac{1}{2} \|\mathbf{x} - \mathbf{u}\|_2^2 + \lambda \|\mathbf{u}\|_{TV}, \quad (8)$$

Condat [28] shows that $\mathcal{T}_\lambda(\mathbf{x})$ can be exactly (numerically) computed by some dynamic programming approaches, such as the taut-string algorithm [28]. Throughout the paper, we keep that $\mathcal{P}_\lambda(\cdot)$ is defined as in (3) and $\mathcal{T}_\lambda(\cdot)$ is defined as in (8).

When considering both penalties, Liu et al. [29] show that the L1-TV regularized problem

$$\hat{\mathbf{x}} := \arg \min_{\mathbf{u} \in \mathbb{R}^n} \frac{1}{2} \|\mathbf{x} - \mathbf{u}\|_2^2 + \lambda_1 \|\mathbf{u}\|_1 + \lambda_2 \|\mathbf{D}\mathbf{u}\|_1 \quad (9)$$

also has a closed-form solution described in the following lemma.

Lemma 1 (See [29]): One has that $\hat{\mathbf{x}} = \mathcal{P}_{\lambda_1}(\mathcal{T}_{\lambda_2}(\mathbf{x}))$ is the solution to (9) where $\lambda_1 > 0$ and $\lambda_2 > 0$ are parameters balancing two penalties.

The operator $\mathcal{P}_{\lambda_1}(\mathcal{T}_{\lambda_2}(\cdot))$ has the following property, which is proved in Appendix A.

Lemma 2: For any $\mathbf{y} \in \mathbb{R}^n$, we have

$$\langle \mathbf{y}, \mathbf{x}^* \rangle = \|\mathbf{x}^*\|_2^2 + \lambda_1 \|\mathbf{x}^*\|_1 + \lambda_2 \|\mathbf{D}\mathbf{x}^*\|_1,$$

where $\mathbf{x}^* := \mathcal{P}_{\lambda_1}(\mathcal{T}_{\lambda_2}(\mathbf{y}))$.

In this paper, we need to solve the optimization problem of the form

$$\arg \min_{\mathbf{u} \in \mathbb{R}^n} -\langle \mathbf{v}, \mathbf{u} \rangle + \lambda_1 \|\mathbf{u}\|_1 + \lambda_2 \|\mathbf{D}\mathbf{u}\|_1 \quad \text{s.t.} \quad \|\mathbf{u}\|_2 \leq 1, \quad (10)$$

where $\mathbf{v} \in \mathbb{R}^n$ is a known vector. Note that the following lemma shows that (10) has a closed-form solution, which is proved in Appendix A.

Lemma 3: For any $\lambda_1, \lambda_2 \geq 0$, the solution \mathbf{u}^ to (10) is given by*

$$\mathbf{u}^* = \frac{\mathcal{P}_{\lambda_1}(\mathcal{T}_{\lambda_2}(\mathbf{v}))}{\|\mathcal{P}_{\lambda_1}(\mathcal{T}_{\lambda_2}(\mathbf{v}))\|_2}.$$

For the underlying signal $\mathbf{x} \in \mathbb{R}^n$ and each measurement \mathbf{a}_i ($i \in [m]$), following [3], we assume each sign measurement $y_i \in \{-1, 1\}$ is drawn independently at random satisfying

$$\mathbb{E}(y_i | \mathbf{a}_i) = \theta(\mathbf{x}^T \mathbf{a}_i), \quad i \in [m], \quad (11)$$

where the function $\theta(\cdot)$ may be unknown or unspecified, which automatically must satisfy $0 \leq \theta(z) \leq 1$. Suggested by [3], we define

$$\mathbb{E}_{g \sim N(0,1)} (\theta(g)g) := \gamma > 0 \quad (12)$$

to capture the relation between $\mathbf{x}^T \mathbf{a}_i$ and y_i . Standard 1-bit measurement becomes a special case of the model (11) with $\theta(\cdot) = \text{sign}(\cdot)$. In this case, we have $\gamma = \mathbb{E}(|g|) = \sqrt{2/\pi}$, achieving the maximum relation.

III. MAIN RESULTS

Since the method for direction recovery is the foundation of the method for full recovery, we first consider direction recovery from standard binary measurement given by model (1) and then extend it to full recovery from thresholded binary measurement given by model (4). Furthermore, we introduce an adaptive quantization scheme to improve recovery accuracy and reduce the number of 1-bit measurements.

A. DIRECTION RECOVERY

As indicated by Lemma 3, the problem in (6) has the closed-form solution

$$\hat{\mathbf{x}} = \frac{\mathcal{P}_{\lambda_1}(\mathcal{T}_{\lambda_2}(\mathbf{A}^T \mathbf{y}/m))}{\|\mathcal{P}_{\lambda_1}(\mathcal{T}_{\lambda_2}(\mathbf{A}^T \mathbf{y}/m))\|_2}.$$

Define $\mathbb{K}_{s_1, s_2} := \{\mathbf{u} \in \mathbb{R}^n : \|\mathbf{u}\|_1 \leq \sqrt{s_1}, \|\mathbf{D}\mathbf{u}\|_1 \leq 2\sqrt{s_2}\}$. The following result implies that any signals in the set $\mathbb{K}_{s_1, s_2} \cap \mathbb{S}^{n-1}$ can be approximately recovered by solving the optimization problem (6).

Theorem 4: Assume

$$\lambda_1 = 2c \sqrt{\frac{t + \log(n)}{m}} \quad (13)$$

for some constant $c > 0$, $\lambda_2 = k\lambda_1$ where $k > 0$ is also a constant, $t > 0$ is a pre-chosen parameter. If $\mathbf{x} \in \mathbb{K}_{s_1, s_2} \cap \mathbb{S}^{n-1}$, with a probability at least $1 - e^{-t}$, the solution $\hat{\mathbf{x}}$ to the optimization problem (6) satisfies

$$\begin{aligned} \|\mathbf{x} - \hat{\mathbf{x}}\|_2 &\leq \sqrt{\frac{3\lambda_1 \sqrt{s_1} + 4\lambda_2 \sqrt{s_2}}{\gamma}} \\ &= \mathcal{O}\left(\sqrt{\frac{(s_1 + s_2)(\log(n) + t)}{m}}\right). \end{aligned}$$

The proof of Theorem 4 is provided in Appendix B.

Remark 5: For fixed $t = \log(n)$, Theorem 4 suggests a recovery accuracy $\varepsilon \in (0, 1)$ provided $m = \mathcal{O}((s_1 + s_2)\log(n)/\varepsilon^4)$. This is the first theoretical result for the L1-TV solver in 1-bit CS. When $\lambda_2 = 0$, the optimization problem (6) reduces to the L1-norm based model proposed in [4] and Theorem 4 is also consistent with the theory result in [4]. When $\lambda_1 = 0$, the optimization problem (6) reduces to a TV-seminorm based model fitting to signals of gradient sparsity.

B. FULL RECOVERY

In this part, we consider full recovery based on the established recovery strategy of direction. To facilitate discussion, we first define the sets

$$\mathbb{U}_{s_1, s_2} := \{\mathbf{u} \in \mathbb{R}^n : \|\mathbf{u}\|_1 / \|\mathbf{u}\|_2 \leq \sqrt{s_1}, \|\mathbf{D}\mathbf{u}\|_1 / \|\mathbf{u}\|_2 \leq 2\sqrt{s_2}\}$$

and

$$\mathbb{V}_{s_1, s_2}(\ell) = \mathbb{U}_{s_1, s_2} \cap \ell\mathbb{B}^n$$

and always assume $\mathbf{x} \in \mathbb{V}_{s_1, s_2}(\ell)$, i.e., the underlying signal satisfies the prior $\|\mathbf{x}\|_2 \leq \ell$. To achieve full recovery of \mathbf{x} , we have to allow a well-chosen threshold vector in 1-bit measurements, i.e.,

$$\mathbf{y} = \text{sign}(\mathbf{A}\mathbf{x} + \boldsymbol{\tau}), \tag{14}$$

where $\boldsymbol{\tau} \in \mathbb{R}^m$ denotes the dither vector whose entries are generated from $N(0, \ell^2)$. Define $\tilde{\mathbf{A}} \in \mathbb{R}^{m \times (n+1)}$ as $\tilde{\mathbf{A}}_i := [\mathbf{A}_i; \tau_i/\ell]$ and $\tilde{\mathbf{D}} \in \mathbb{R}^{n \times (n+1)}$ as the differential matrix. Let $\tilde{\mathbf{x}} = [\mathbf{x}; \ell] \in \mathbb{R}^n$. Then the measurements (14) can be reshaped as

$$\mathbf{y} = \text{sign}(\tilde{\mathbf{A}}\tilde{\mathbf{x}}). \tag{15}$$

Intuitively, both $\|\tilde{\mathbf{x}}\|_1 / \|\tilde{\mathbf{x}}\|_2$ and $\|\tilde{\mathbf{D}}\tilde{\mathbf{x}}\|_1 / \|\tilde{\mathbf{x}}\|_2$ are bounded, and thus we can recover $\tilde{\mathbf{x}}$ from (15) by solving the established optimization problem (6). Indeed, the desired upper bounds on them are guaranteed by the following lemma proved in Appendix A.

Lemma 6: One has $\|\tilde{\mathbf{x}}\|_1 / \|\tilde{\mathbf{x}}\|_2 \leq \sqrt{s_1 + 1}$ and $\|\tilde{\mathbf{D}}\tilde{\mathbf{x}}\|_1 / \|\tilde{\mathbf{x}}\|_2 \leq 2\sqrt{s_2 + 1}$.

Thus, we can solve

$$\tilde{\mathbf{x}}^\# = \arg \min_{\mathbf{u} \in \mathbb{R}^{n+1}} -\frac{1}{m} \tilde{\mathbf{A}}^T \mathbf{y}, \mathbf{u} + \lambda_1 \|\mathbf{u}\|_1 + \lambda_2 \|\tilde{\mathbf{D}}\mathbf{u}\|_1 \tag{16}$$

to reconstruct the underlying signal $\tilde{\mathbf{x}}$. Full recovery of \mathbf{x} relies on the following lemma which combines direction recovery error and full recovery error.

Lemma 7 (See [14]): For $\mathbf{f}, \mathbf{g} \in \mathbb{R}^n$, $\hat{\mathbf{f}}, \hat{\mathbf{g}}$ denote the vectors that are lifted by one dimension, taking the form of

$$\hat{\mathbf{f}} := (\mathbf{f}; f_{n+1}) \in \mathbb{R}^{n+1} \quad \text{and} \quad \hat{\mathbf{g}} := (\mathbf{g}; g_{n+1}) \in \mathbb{R}^{n+1}$$

with $f_{n+1} \neq 0, g_{n+1} \neq 0$. One has

$$\left\| \frac{\mathbf{f}}{f_{n+1}} - \frac{\mathbf{g}}{g_{n+1}} \right\|_2 \leq \frac{\|\hat{\mathbf{f}}\|_2 \|\hat{\mathbf{g}}\|_2}{|f_{n+1}| |g_{n+1}|} \left\| \frac{\hat{\mathbf{f}}}{\|\hat{\mathbf{f}}\|_2} - \frac{\hat{\mathbf{g}}}{\|\hat{\mathbf{g}}\|_2} \right\|_2.$$

Lemma 7 indicates that $f_{n+1}\mathbf{g}/g_{n+1}$ is an approximate estimate of \mathbf{f} provided the direction recovery error $\|\hat{\mathbf{f}}/\|\hat{\mathbf{f}}\|_2 - \hat{\mathbf{g}}/\|\hat{\mathbf{g}}\|_2\|_2$ is small enough and $\|\hat{\mathbf{f}}\|_2, \|\hat{\mathbf{g}}\|_2/|g_{n+1}|$ are above bounded. Inspired by this fact, the following theorem provides a method for the full recovery of signals with a theoretical guarantee.

Theorem 8: Consider $\ell > 0$ and $\mathbf{x} \in \mathbb{V}_{s_1, s_2}(\ell)$. Let λ_1 and λ_2 be defined the same as Theorem 4. If $m \geq C\epsilon^{-4}\gamma^{-2}(s_1 + k^2s_2)(t + \log(n))$ for some constant $C > 0$, then with a probability at least $1 - e^{1-t}$, the solution $\hat{\mathbf{x}}^\# = (\mathbf{x}^\#; x^\#)$ to the optimization problem (16) satisfies

$$\left\| \frac{\ell\mathbf{x}^\#}{x^\#} - \mathbf{x} \right\|_2 \leq R\epsilon. \tag{17}$$

The proof of Theorem 8 is provided in Appendix B.

Remark 9: Theorem 8 implies that $\ell\mathbf{x}^\#/x^\#$ is an approximate estimate of the underlying signal \mathbf{x} . For fixed $t = \log(n)$, the method also inherits the sample complexity $\mathcal{O}((s_1 + s_2)\log(n)/\epsilon^4)$ with full recovery error relating to the prior magnitude upper ℓ . Via dividing Eq.(17) by ℓ , we have that

$$\left\| \frac{\ell\mathbf{x}^\#}{x^\#} - \mathbf{x} \right\|_2 / \ell \leq \epsilon.$$

Since ℓ is usually in the same order of $\|\mathbf{x}\|_2$, we conclude that the relative recovery error $\|\ell\mathbf{x}^\#/x^\# - \mathbf{x}\|_2 / \|\mathbf{x}\|_2$ turns to be above bounded by a quantity independent of ℓ .

From Theorem 8, the proposed method exhibits polynomial error decay in the oversampling factor $\mu = m/[(s_1 + s_2)\log(n)]$, which sometimes may be unsatisfactory. Next, we aim to obtain an exponential decay. To this end, borrowing the idea from [6], we provide a new 1-bit quantization scheme that adaptively chooses the dither vector at each iteration. Before explaining this idea, we clarify some notations. We define the projection operator $\mathcal{P}_{\mathbb{K}}(\cdot)$ as

$$\mathcal{P}_{\mathbb{K}}(\mathbf{x}) := \arg \min_{\mathbf{u}} \|\mathbf{x} - \mathbf{u}\|_2 \quad \text{s.t.} \quad \mathbf{u} \in \mathbb{K},$$

and let $\Delta(\mathbf{y})$ denote the reconstruction operator that recovers the underlying signal from thresholded 1-bit measurements \mathbf{y} by solving the problem (16). Moreover, we set $\mathbf{x}_0 = \mathbf{0}$,

$$\mathbf{x}_k := \mathcal{P}_{\mathbb{V}_{s_1, s_2}(\ell)}(\mathbf{x}_{k-1} + \widehat{\mathbf{x} - \mathbf{x}_{k-1}})$$

for each $k = \{1, \dots, K\}$ where K denotes the maximum number of iterations, and define

$$\widehat{\mathbf{x} - \mathbf{x}_{k-1}} := \Delta(\text{sign}(\mathbf{A}^{(k)}(\mathbf{x} - \mathbf{x}_{k-1}) + \boldsymbol{\tau}^{(k)})).$$

For any $\mathbf{x} \in \mathbb{V}_{s_1, s_2}(\ell)$ and $0 < \eta \leq \frac{1}{4}$, we may choose

$$q_1 = \mathcal{O}(\eta^{-4}\gamma^{-2}(s_1 + k^2s_2)(t + \log(n))) \tag{18}$$

to generate standard Gaussian measurement matrix $\mathbf{A}^{(1)} \in \mathbb{R}^{q \times m}$ and observe $\mathbf{y}^{(1)}$ via $\mathbf{y}^{(1)} = \text{sign}(\mathbf{A}^{(1)}\mathbf{x} + \boldsymbol{\tau}^{(1)})$ where $(\boldsymbol{\tau}^{(1)})_i \sim \mathcal{LN}(0, 1)$ for each $i \in [q_1]$. Then Theorem 8 implies that

$$\|\widehat{\mathbf{x} - \mathbf{x}_0} - \mathbf{x}\|_2 \leq \eta\ell$$

holds with a probability at least $1 - e^{1-t}$. Since $\mathbf{x} \in \mathbb{V}_{s_1, s_2}(\ell)$ and \mathbf{x}_1 is the best approximation to $\widehat{\mathbf{x}} - \mathbf{x}_0$ in the set $\mathbb{V}_{s_1, s_2}(\ell)$, using triangle inequality indicates that

$$\|\mathbf{x}_1 - \mathbf{x}\|_2 \leq 2\|\widehat{\mathbf{x}} - \mathbf{x}_0 - \mathbf{x}\|_2 \leq 2\eta\ell. \quad (19)$$

Then we regard the signal $\mathbf{x} - \mathbf{x}_1 \in \mathbb{U}_{s'_1, s'_2} \cap 2\eta\ell\mathbb{B}^n := \mathbb{V}_{s'_1, s'_2}(2\eta\ell)$ as the target signal to be reconstructed and conduct another q_2 measurements via

$$\mathbf{y}^{(2)} = \text{sign}(\mathbf{A}^{(2)}(\mathbf{x} - \mathbf{x}_1) + \boldsymbol{\tau}^{(2)})$$

where $(\boldsymbol{\tau}^{(2)})_i \sim 2\eta\ell\mathcal{N}(0, 1)$ for each $i \in [q_2]$. Utilizing Theorem 8 again indicates that

$$\|\widehat{\mathbf{x}} - \mathbf{x}_1 - (\mathbf{x} - \mathbf{x}_1)\|_2 \leq 2\eta^2\ell$$

holds with a high probability, provided $q_2 \geq C\eta^{-4}\gamma^{-2}(s'_1 + k^2s'_2)(t + \log(n))$. A similar discussion to the proof of Eq.(19) shows that

$$\|\mathbf{x}_2 - \mathbf{x}\|_2 \leq (2\eta)^2\ell. \quad (20)$$

Repeating this process, in the l -th iteration, we have that

$$\|\mathbf{x}_l - \mathbf{x}\|_2 \leq (2\eta)^l\ell$$

holds with a high probability. Through the above discussion, it is not hard to find that the recovery error converges to 0 at an exponential rate with the increase in the number of iterations provided

$$q_l \geq C\eta^{-4}\gamma^{-2}(1 + k^2)s'_{l-1}\log(n). \quad (21)$$

We summarize the quantization process in Algorithm 1 and the recovery process in Algorithm 2 to fit the common paradigm of compressed sensing.

It should be noted that achieving the projection operation $\mathcal{P}_{\mathbb{V}_{s_1, s_2}(\ell)}(\cdot)$ is not easy. In practice, we may consider the convex set $\ell\mathbb{K}_{s_1, s_2}$ and thus replace $\mathcal{P}_{\mathbb{V}_{s_1, s_2}(\ell)}(\cdot)$ by $\mathcal{P}_{\ell\mathbb{K}_{s_1, s_2}}(\cdot)$ so that it can be solved by existing convex optimization software, such as the CVX toolbox. Since R is in the same order of $\|\mathbf{x}\|_2$, this modification is acceptable though it relaxes the approximately sparse constraint to some extent. Moreover, for ease of applying, we usually fix $q_i = q, (i \in [L])$ with a well-chosen q and thus $m = Lq$ in practice.

The proposed adaptive quantization/recovery scheme is guaranteed by the following theorem which is proved in Appendix B.

Theorem 10: For any $\mathbf{x} \in \mathbb{V}_{s_1, s_2}(\ell)$ and $0 < \eta \leq \frac{1}{4}$, let $\alpha = \min\{|x_i| : i \in \text{supp}(\mathbf{x})\}$ and $\{q_1, q_2, \dots, q_L\}$ is defined as in Algorithm 1, where L denotes the maximum number of iterations. Suppose we have $m := \sum_{i=1}^L q^{(i)}$ independent measurement vectors $\mathbf{a}_i \in \mathbb{R}^n (i \in [m])$ which are populated by independent random standard Gaussian entries. Then with a probability at least $1 - Le/n$, the output \mathbf{x}_L of Algorithm 2 satisfies

$$\|\mathbf{x} - \mathbf{x}_L\|_2 \leq (2\eta)^L\ell.$$

Algorithm 1 Adaptive Quantization Process

Input: $\mathbf{x} \in \mathbb{R}^n, \ell, s_1, s_2, \eta, L, \alpha;$

Output: $\mathbf{A}, \mathbf{y}, \boldsymbol{\tau}.$

- 1: Initialize: Choose q_1 according to (18), generate $\mathbf{A}^{(1)} \in \mathbb{R}^{q_1 \times n}$ such that $(\mathbf{A}^{(1)})_{ij} \sim \mathcal{N}(0, 1)$ for any $(i, j) \in [n_1] \times [n_2]$, generate $\boldsymbol{\tau}^{(1)} \in \mathbb{R}^{q_1}$ such that $(\boldsymbol{\tau}^{(1)})_i \sim \ell\mathcal{N}(0, 1)$ for all $i \in [q_1]$.
 - 2: $\mathbf{y}^{(1)} = \text{sign}(\mathbf{A}^{(1)}\mathbf{x} + \boldsymbol{\tau}^{(1)});$
 - 3: $\widehat{\mathbf{x}} - \mathbf{x}_0 = \Delta(\mathbf{y}^{(1)});$
 - 4: $\mathbf{x}_1 = \mathcal{P}_{\mathbb{V}_{s_1, s_2}(\ell)}(\widehat{\mathbf{x}} - \mathbf{x}_0);$
 - 5: **for** $l = 2 : L$ **do**
 - 6: Calculate $\alpha_{l-1} = \min\{|\mathbf{x}_{l-1}|_i : i \in \text{supp}(\mathbf{x}_{l-1})\}$ and $s'_{l-1} = \frac{\ell(\alpha_{l-1}/\alpha + 1)\sqrt{s_1}}{\alpha_{l-1}}$, and set q_l according to (21);
 - 7: Generate $\mathbf{A}^{(l)} \in \mathbb{R}^{q_l \times n}$ such that $(\mathbf{A}^{(l)})_{ij} \sim \mathcal{N}(0, 1)$ for any $(i, j) \in [n_1] \times [n_2]$, generate $\boldsymbol{\tau}^{(l)} \in \mathbb{R}^{q_l}$ such that $(\boldsymbol{\tau}^{(l)})_i \sim (2\eta)^l\ell\mathcal{N}(0, 1)$ for all $i \in [q_l]$;
 - 8: $\boldsymbol{\delta}^{(l)} = \mathbf{A}^{(l)}\mathbf{x}_{l-1};$
 - 9: $\mathbf{y}^{(l)} = \text{sign}(\mathbf{A}^{(l)}\mathbf{x} - \boldsymbol{\delta}^{(l)} + \boldsymbol{\tau}^{(l)});$
 - 10: $\widehat{\mathbf{x}} - \mathbf{x}_{l-1} = \Delta(\mathbf{y}^{(l)});$
 - 11: $\mathbf{x}_l = \mathcal{P}_{\mathbb{V}_{s_1, s_2}(\ell)}(\widehat{\mathbf{x}} - \mathbf{x}_{l-1});$
 - 12: **end for**
 - 13: **return** the sensing matrix $\mathbf{A} = (\mathbf{A}^{(1)}; \dots; \mathbf{A}^{(L)})$, the sign observations $\mathbf{y} = (\mathbf{y}^{(1)}; \dots; \mathbf{y}^{(L)})$, and the dither vector $\boldsymbol{\tau} = (\boldsymbol{\tau}^{(1)}; \dots; \boldsymbol{\tau}^{(L)})$.
-

Algorithm 2 Adaptive Recovery Process

Input: $\mathbf{A}, \mathbf{y}, \boldsymbol{\tau}, s_1, s_2, L;$

Output: $\hat{\mathbf{x}};$

- 1: $\mathbf{x}^{(0)} = \mathbf{0};$
 - 2: **for** $l = 1 : L$ **do**
 - 3: Update \mathbf{x}_l via $\mathbf{x}_l = \mathcal{P}_{\mathbb{V}_{s_1, s_2}(\ell)}(\mathbf{x}_{l-1} + \Delta(\mathbf{y}^{(l)}))$
 - 4: **end for**
 - 5: **return** $\hat{\mathbf{x}} := \mathbf{x}_L$.
-

IV. EXPERIMENTS

In this section, we conduct a series of experiments on synthetic data and real-world data to test the validity of the proposed algorithms and the correctness of the theoretical results. For simplicity, we use the same notations between models and algorithms, calling the proposed nonadaptive algorithm *L1-TV* and the adaptive algorithm *Adap L1-TV* (i.e., Algorithm 1, 2). To show the superiority of the proposed model, we compare L1-TV with the algorithm for solving only L1-penalty based optimization model and the algorithm for solving only TV-penalty based optimization model, respectively. Specifically, we employ the algorithm proposed in [4] to solve the L1-penalty based model and call the algorithm *L1* for simplicity. For the TV-penalty based model, there is no existing related research. Thus we zero the penalty parameter for the L1-term in L1-TV to get the corresponding algorithm, naming it *TV*.

In Section IV-A, we conduct some simulations to explain how to select proper parameters for L1-TV and Adap L1-TV. In Section IV-B and Section IV-C, we apply the proposed

algorithms to ECG signal recovery and gray image recovery, respectively. For the underlying signal \mathbf{x} and its estimate $\hat{\mathbf{x}}$, we define the direction error as $\text{DirError} := \|\hat{\mathbf{x}}/\|\hat{\mathbf{x}}\|_2 - \mathbf{x}/\|\mathbf{x}\|_2\|_2$ and the relative error as $\text{RelError} = \|\hat{\mathbf{x}} - \mathbf{x}\|_2/\|\mathbf{x}\|_2$. Note that in 1-bit quantization, it is meaningful to acquire sign measurements at high, super-Nyquist rates since the measurements are cheap and fast. Thus, we conduct experiments at the oversampling rate m/n more than 1. We perform all experiments on a PC with 32 GB of RAM and Intel Core i7-8700M(3.2GHz) and always fix $t = 10$ in assumption (13).

A. SYNTHETIC EXPERIMENTS

This part conducts some synthetic experiments to help clarify the parameter setting and verify our theoretical results. In the first experiment, three signals are generated at random as shown in FIGURE 1 (a),(b),(c). With fixed $m = 20n$ and $\lambda_2 = 0.15\sqrt{(t + \log(n))/m}$, the constant c in Eq.(13) values in the range $[0, 1.5]$, and thus the parameter λ_1 also varies. The average direction errors for different values of c are reported in FIGURE 1 (d) over 50 repeated trials. From the experimental results, we may choose $c = 0.075$ in different cases since it almost always achieves relatively good performance.

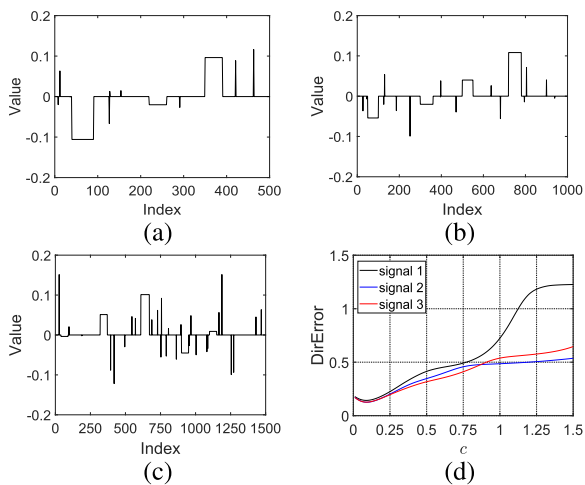


FIGURE 1. (a)-(c) are signal 1, signal 2, and signal 3, respectively; (d) shows direction errors for different values of c .

Next, we select the value of parameter λ_2 over synthetic experiments. First, we generate four signals of different lengths ($n = 500, n = 1000, n = 1500$, and $n = 2000$) at random, termed as signal 1, signal 2, signal 3, and signal 4, respectively (They are generated in a similar way to FIGURE 1 (a),(b),(c). Due to limited space, we do not plot them here). We fix the constant $c = 0.075$, parameters $\lambda_1 = 2c\sqrt{(t + \log(n))/m}$, $\lambda_2 = k\lambda_1$ and test the performance of the proposed method for different values of k . FIGURE 2 plots the direction errors, and TABLE 1 lists the values of $\sqrt[4]{s_1/s_2}$ for different tested signals \mathbf{x} where $s_1 := \|\mathbf{x}\|_0$ and $s_2 := \|\mathbf{D}\mathbf{x}\|_0$. Observed from the numerical results, we may approximately set $k = \sqrt[4]{s_1/s_2}$ in general cases.

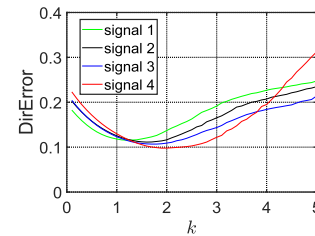


FIGURE 2. Direction errors obtained by L1-TV with different values of λ_2 .

TABLE 1. Values of $\sqrt[4]{s_1/s_2}$ of different signals.

Tested data	signal 1	signal 2	signal 3	signal 4
$\sqrt[4]{s_1/s_2}$	1.4218	1.7924	2.0205	2.3377

With the above parameter setting, we conduct an experiment to compare the proposed L1-TV algorithm with the methods of L1 and TV by calculating the direction recovery errors for different sampling rates. FIGURE 3 (a) plots the original signal and FIGURE 3 (b) reports the average direction errors over 50 trials. As expected, the experimental results show that L1-TV performs better than the compared algorithms since the former penalizes both sparsity and gradient sparsity.

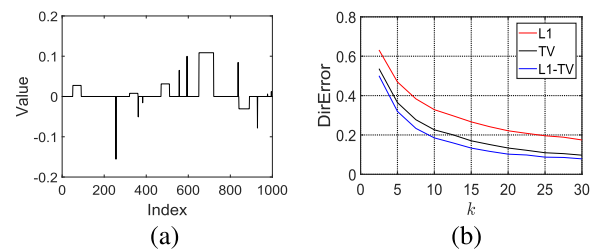


FIGURE 3. (a) is the original signal; (b) plots direction errors obtained by different algorithms at varying sampling rates.

The above experiments only consider the recovery of directions of signals. In the following, we further test the performance of the proposed algorithms for recovering the full signal (including both direction and magnitude). Firstly, we give a convenient choice for the standard deviation σ of the dither vector $\boldsymbol{\tau}$ in the measurement model (14). Though Theorem 8 provides a theoretical guarantee for the case that σ equals to the known upper bound ℓ of $\|\mathbf{x}\|_2$, the setting may not be optimal in practice. In fact, the standard deviation σ should be smaller than the real value of $\|\mathbf{x}\|_2$. To see it, we generate three signals of different lengths ($n = 500, n = 1000, n = 1500$) at random, termed signal 1, signal 2, and signal 3, respectively (we do not plot them due to limited space). Then we fix $m = 20n$ and test the performance of L1-TV for different values of $\sigma/\|\mathbf{x}\|_2$. The relative errors are plotted in FIGURE 4 over 100 trials. It can be seen that too large or too small values of σ all tend to result in poor performance. We find that L1-TV performs the best when σ is about $0.2\|\mathbf{x}\|_2$ and behaves acceptably in the

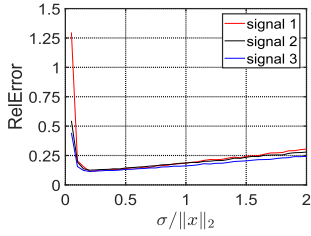


FIGURE 4. Relative errors obtained by L1-TV with different values of $\sigma/\|x\|_2$.

range $[0.1\|x\|_2, \|x\|_2]$. The observation inspires us to choose the value of σ in the following way: (i) For synthetic experiments, we first generate 100 vectors with the same scale and sparsity as the underlying signals at random and obtain the average value $\hat{\ell}$ of their magnitude. Then we empirically set $\sigma = 0.2\hat{\ell}$. (ii) For experiments on image processing, since the maximum of the underlying signal is known, e.g., 255, we may assume $\hat{\ell} = 255\sqrt{n}/2$ and easily set $\sigma = 255\sqrt{n}/10$ where n denotes the number of pixels. Besides, some other methods, such as cross-validation, can also be employed to tune σ , which may also obtain good performance.

With the above setting on parameters, we experiment to compare the proposed algorithms, L1-TV and Adap L1-TV, with the competitive algorithms, L1 and TV. Note that for Adap L1-TV, we easily set $q_1 = q_2 = \dots = q_{10} = 3n$ for the sampling complexity of each iteration. For a randomly generated signal x of dimension 1000, we calculate the average relative errors over 100 random trials at each sampling rate and plot the results in FIGURE 5. One can easily observe that L1-TV outperforms both L1 and TV at all sampling rates. Besides, the introduction of an adaptive quantization/recovery scheme (i.e. Adap L1-TV) significantly improves the recovery performance of L1-TV, which coincides to the conclusion reflected by Theorem 10.

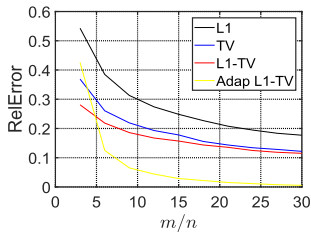


FIGURE 5. Relative errors obtained by different algorithms at varying sampling rates.

B. APPLICATION TO ECG SIGNALS RECOVERY

In this part, we employ the proposed method for ECG signals recovery. The ECG signals were collected from the MIT-BIH Arrhythmia Database [30] and are windowed to 1000 samples in the experiment. FIGURE 6 (a) displays a segment of such ECG signals which is tested as the underlying signal.

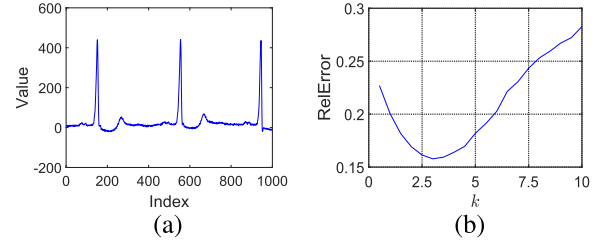


FIGURE 6. (a) is a segment from ECG signals; (b) shows relative errors for different values of k .

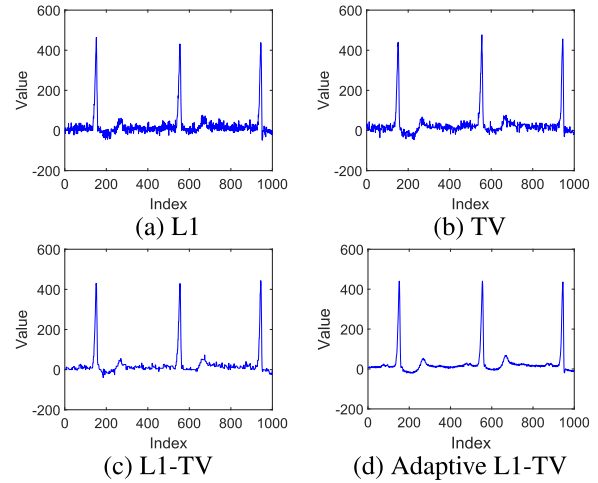


FIGURE 7. Reconstructed results by different methods. (a) shows the result obtained by L1 (RelError=0.1909); (b) shows the result obtained by TV (RelError=0.1638); (c) shows the result obtained by L1-TV (RelError=0.1192); (d) shows the result obtained by Adap L1-TV (RelError=0.0053).

The upper bound R on the magnitude is selected using the cross-validation on the tested set. The parameter λ_1 is set as $\lambda_1 = 0.15\sqrt{(t + \log(n))/m}$, suggested by the experimental results in Section IV-A. It remains to determine the ratio $k = \lambda_2/\lambda_1$. To this end, we fix $m = 20n$ and calculate the average relative errors for different values of k . The results are plotted in FIGURE 6 (b), which indicates that $k = 3$ may be an appropriate choice. For TV, a similar exhaustive test suggests us to choose the regularization parameter $0.30\sqrt{(t + \log(n))/m}$. For L1, we follow the suggestion by [4] that is $\lambda = \sqrt{(t + \log(n))/m}$. Under such parameter setting, we compare the proposed L1-TV and Adap L1-TV with L1 and TV at the sampling rate $m/n = 30$ and report the experimental results in FIGURE 7. We observe that Adap L1-TV performs the best, behaving with great advantages. L1-TV also achieves better performance than either L1 or TV, attributed to considering both signal sparsity and sparsity in the difference between consecutive components of the signal.

C. APPLICATION TO GRAY IMAGES RECOVERY

Next, we test the recovery performance of the proposed algorithms on gray images. Five images of size 256×256 are considered as examples, named as Barbara, Lenna, Pepper, and House, respectively (see FIGURE 8 (a)). Note that the

TABLE 2. Comparison of the PSNR and SSIM at different sampling rates.

$\frac{m}{n}$	Data	L1		TV		L1-TV		Adap L1-TV	
		PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
10	Barbara	20.86	0.43	20.75	0.43	22.43	0.47	27.77	0.79
	Lenna	20.36	0.43	21.74	0.47	22.32	0.52	27.96	0.80
	Pepper	20.06	0.44	21.20	0.47	21.54	0.53	28.38	0.80
	House	21.39	0.48	21.53	0.48	22.51	0.49	28.80	0.81
	Tree	21.53	0.48	22.14	0.43	22.62	0.46	29.01	0.85
20	Barbara	22.44	0.46	22.63	0.53	24.06	0.61	32.04	0.89
	Lenna	22.02	0.46	22.34	0.54	23.60	0.57	32.96	0.86
	Pepper	22.64	0.47	21.88	0.55	24.02	0.61	33.23	0.88
	House	23.97	0.50	23.24	0.57	24.47	0.64	34.70	0.92
	Tree	23.48	0.40	23.12	0.61	24.30	0.61	33.27	0.92
30	Barbara	24.03	0.53	23.79	0.58	25.06	0.66	37.86	0.96
	Lenna	23.63	0.49	23.57	0.60	24.85	0.67	38.13	0.96
	Pepper	24.10	0.54	23.17	0.61	25.06	0.66	38.72	0.96
	House	24.42	0.56	24.62	0.65	25.68	0.71	39.02	0.97
	Tree	24.94	0.46	24.73	0.63	25.91	0.75	37.87	0.96

**FIGURE 8.** The original image and the corresponding recovered results by different algorithms with the sampling rate $m/n = 30$. (a) is the original image; (b) shows the result obtained by L1; (c) shows the result obtained by TV; (d) shows the result obtained by L1-TV; (e) shows the result obtained by Adap L1-TV.

recovery process is not performed directly on the whole image but small disjoint segments. Specifically, we first decompose the original image into 64 parts of size 32×32 and then restore these segments one by one. These segments are sparse in some bases, such as discrete cosine transform (DCT) or wavelet basis. In this experiment, we exploit the MATLAB command, dct2, to generate the desired sparse basis and then use different algorithms to reconstruct the induced sparse coefficients from thresholded binary measurements at sampling rates $m/n = 10$, $m/n = 20$, and $m/n = 30$, respectively. We set the maximum number of iterations $K = 5$ for Adap L1-TV. TABLE 2 exhibits the PSNR and SSIM values of the tested images obtained by different algorithms. From FIGURE 8 (b) to FIGURE 8 (e), we plot the recovered visual results by different algorithms respectively at the sampling rate $m/n = 30$. Whether from a visual or quantitative perspective, our algorithms equip more powerful recovery ability than classic sparse and gradient sparse based methods. Indeed, the superiority of L1-TV benefits from the combined utilization of sparse and local-smooth properties of image data while the outstanding performance of Adap L1-TV comes from the integrated use of an L1-TV regularizer and an adaptive measurement scheme.

V. CONCLUSION AND FUTURE WORK

This paper proposed a recovery method for 1-bit compressed sensing based on the L1-TV regularizer with an efficient solving algorithm. We also provide theoretical guarantees for the proposed method by detailed analysis. Compared with existing methods, our algorithms can better process signals that are element sparse and gradient sparse. In addition, the proposed adaptive algorithm achieves a significantly faster recovery error decay rate.

It is worth mentioning that some signals are sparse under particular linear transformations rather than themselves. We may consider linear transforms in the L1-norm penalty to fit more data in applications in future work. It is interesting to provide a corresponding theoretical analysis and consider how to design an efficient algorithm in this case.

APPENDIX

In this part, we present proofs for the proposed lemmas and theorems.

A. PROOFS OF THE PROPOSED LEMMAS

Proof of Lemma 1: We first define $\text{SGN}(\cdot)$ such that: if $t > 0$, $\text{SGN}(t) = 1$; if $t < 0$, $\text{SGN}(t) = -1$; and if $t = 0$, $\text{SGN}(t) = [-1, 1]$ that is projecting a element onto an interval. Define

$$g(\mathbf{u}) := \frac{1}{2} \|\mathbf{u} - \mathbf{y}\|_2^2 + \lambda_1 \|\mathbf{u}\|_1 + \lambda_2 \|D\mathbf{u}\|_1$$

and let $\partial g(\mathbf{u}^*)$ denote the subdifferential of $g(\cdot)$ at \mathbf{u}^* (refer to e.g. [31] for details of subdifferential). Noting that $\text{SGN}(\cdot)$ is

the subgradient of $\|\cdot\|_1$, we then have

$$\partial g(\mathbf{u}^*) = \mathbf{u}^* - \mathbf{y} + \lambda_1 \text{SGN}(\mathbf{u}^*) + \lambda_2 D^T \text{SGN}(D\mathbf{u}^*).$$

For any $\mathbf{u} \in \partial g(\mathbf{u}^*)$, using the definition of $\text{SGN}(\cdot)$, we can easily check that

$$\langle \mathbf{u}, \mathbf{u}^* \rangle = \|\mathbf{u}^*\|_2 - \langle \mathbf{y}, \mathbf{u}^* \rangle + \lambda_1 \|\mathbf{u}^*\|_1 + \lambda_2 \|D\mathbf{u}^*\|_1. \quad (22)$$

From Lemma 1, we know $\mathbf{u}^* = \arg \min_{\mathbf{u} \in \mathbb{R}^n} g(\mathbf{u})$. Since the objective function $g(\mathbf{u})$ is strictly convex, it admits a unique minimizer by \mathbf{u}^* . The optimality condition implies that

$$\mathbf{0} \in \partial g(\mathbf{u}^*).$$

Thus, Eq.(22) holds for $\mathbf{u} = \mathbf{0}$, indicating that

$$\langle \mathbf{y}, \mathbf{u}^* \rangle = \|\mathbf{u}^*\|_2 + \lambda_1 \|\mathbf{u}^*\|_1 + \lambda_2 \|D\mathbf{u}^*\|_1.$$

We have completed the proof. \square

Proof of Lemma 3: Following the standard analysis of convex optimization [32], the Lagrange dual function $L(\mu)$ of the optimization problem (10) is given by

$$\begin{aligned} L(\mu) &= \min_{\mathbf{x} \in \mathbb{R}^n} -\langle \mathbf{x}, \mathbf{v} \rangle + \lambda_1 \|\mathbf{x}\|_1 + \lambda_2 \|D\mathbf{x}\|_1 + \mu(\|\mathbf{x}\|_2^2 - 1) \\ &= \min_{\mathbf{x} \in \mathbb{R}^n} 2\mu \left(\frac{1}{2} \left\| \mathbf{x} - \frac{\mathbf{v}}{2\mu} \right\|_2^2 + \frac{\lambda_1}{2\mu} \|\mathbf{x}\|_1 + \frac{\lambda_2}{2\mu} \|D\mathbf{x}\|_1 \right) \\ &\quad - \frac{\|\mathbf{v}\|_2^2}{4\mu} - \mu \\ &= 2\mu \left(\min_{\mathbf{x} \in \mathbb{R}^n} \frac{1}{2} \left\| \mathbf{x} - \frac{\mathbf{v}}{2\mu} \right\|_2^2 + \frac{\lambda_1}{2\mu} \|\mathbf{x}\|_1 + \frac{\lambda_2}{2\mu} \|D\mathbf{x}\|_1 \right) \\ &\quad - \frac{\|\mathbf{v}\|_2^2}{4\mu} - \mu \\ &= 2\mu p(\mu) - \frac{\|\mathbf{v}\|_2^2}{4\mu} - \mu, \end{aligned}$$

where

$$\begin{aligned} p(\mu) &:= \min_{\mathbf{x} \in \mathbb{R}^n} \frac{1}{2} \left\| \mathbf{x} - \frac{\mathbf{v}}{2\mu} \right\|_2^2 + \frac{\lambda_1}{2\mu} \|\mathbf{x}\|_1 + \frac{\lambda_2}{2\mu} \|D\mathbf{x}\|_1 \\ &= \min_{\mathbf{x} \in \mathbb{R}^n} \frac{1}{4\mu^2} \left(\frac{1}{2} \|2\mu\mathbf{x} - \mathbf{v}\|_2^2 + \lambda_1 \|2\mu\mathbf{x}\|_1 \right. \\ &\quad \left. + \lambda_2 \|D(2\mu\mathbf{x})\|_1 \right). \end{aligned}$$

Using Lemma 1, we have

$$p(\mu) = \frac{1}{8\mu^2} \|\mathbf{u}^* - \mathbf{v}\|_2^2 + \frac{\lambda_1}{4\mu^2} \|\mathbf{u}^*\|_1 + \frac{\lambda_2}{4\mu^2} \|D(\mathbf{u}^*)\|_1,$$

where

$$\mathbf{u}^* = \mathcal{P}_{\lambda_1}(\mathcal{T}_{\lambda_2}(\mathbf{v})).$$

The Lagrange dual problem of (10) is given by

$$\begin{aligned} \max_{\mu \geq 0} & -\mu - \frac{1}{4\mu} (\|\mathbf{v}\|_2^2 - \|\mathbf{u}^* - \mathbf{v}\|_2^2 - 2\lambda_1 \|\mathbf{u}^*\|_1 \\ & - 2\lambda_2 \|\mathbf{D}\mathbf{u}^*\|_1) \\ = & \max_{\mu \geq 0} -\mu - \frac{1}{4\mu} (2\langle \mathbf{u}^*, \mathbf{v} \rangle - \|\mathbf{u}^*\|_2^2 - 2\lambda_1 \|\mathbf{u}^*\|_1 \\ & - 2\lambda_2 \|\mathbf{D}\mathbf{u}^*\|_1) \\ = & \max_{\mu \geq 0} -\mu - \frac{\|\mathbf{u}^*\|_2^2}{4\mu}, \end{aligned}$$

where the final equality utilizes Lemma 2. Obviously, the optimal dual solution is

$$\mu^* = \frac{\|\mathbf{u}^*\|_2}{2} = \frac{\|\mathcal{P}_{\lambda_1}(\mathcal{T}_{\lambda_2}(\mathbf{v}))\|_2}{2}.$$

Therefore, using Lemma 1, the optimal solution is

$$\begin{aligned} \mathbf{x}^* &= \arg \min_{\mathbf{x} \in \mathbb{R}^n} \frac{1}{2} \left\| \mathbf{x} - \frac{\mathbf{v}}{2\mu^*} \right\|_2^2 + \frac{\lambda_1}{2\mu^*} \|\mathbf{x}\|_1 + \frac{\lambda_2}{2\mu^*} \|\mathbf{D}\mathbf{x}\|_1 \\ &= \arg \min_{\mathbf{x} \in \mathbb{R}^n} \frac{1}{2} \|2\mu^* \mathbf{x} - \mathbf{v}\|_2^2 + \lambda_1 \|2\mu^* \mathbf{x}\|_1 + \lambda_2 \|2\mu^* \mathbf{D}\mathbf{x}\|_1 \\ &= \frac{\mathcal{P}_{\lambda_1}(\mathcal{T}_{\lambda_2}(\mathbf{v}))}{2\mu^*} = \frac{\mathcal{P}_{\lambda_1}(\mathcal{T}_{\lambda_2}(\mathbf{v}))}{\|\mathcal{P}_{\lambda_1}(\mathcal{T}_{\lambda_2}(\mathbf{v}))\|_2}. \end{aligned}$$

The proof is completed. \square

Proof of Lemma 6: Since $\|\mathbf{x}\|_1 / \|\mathbf{x}\|_2 \leq \sqrt{s_1}$, we have

$$\begin{aligned} \|\tilde{\mathbf{x}}\|_1 &= \|\mathbf{x}\|_1 + \ell \\ &\leq \sqrt{s_1} \|\mathbf{x}\|_2 + \ell \\ &= \sqrt{s_1 \|\mathbf{x}\|_2^2 + \ell^2 + 2\sqrt{s_1} \ell \|\mathbf{x}\|_2} \\ &\leq \sqrt{(s_1 + 1) \|\mathbf{x}\|_2^2 + (s_1 + 1) \ell^2} \\ &= \sqrt{s_1 + 1} \|\tilde{\mathbf{x}}\|_2. \end{aligned}$$

Besides, since $x_n \leq \|\mathbf{x}\|_2 \leq \ell$, we have

$$\begin{aligned} \|\tilde{\mathbf{D}}\tilde{\mathbf{x}}\|_1 &= \|\mathbf{D}\mathbf{x}\|_1 + |x_n - \ell| \\ &\leq 2(\sqrt{s_2} \|\mathbf{x}\|_2 + \ell) \\ &= 2\sqrt{s_2 + 1} \|\tilde{\mathbf{x}}\|_2. \end{aligned}$$

The proof is completed. \square

B. PROOFS OF THE PROPOSED THEOREMS

Proof of Theorem 4: The proof is fundamentally built upon the following observation between $\frac{1}{m} \mathbf{A}^T \mathbf{y}$ and $\gamma \mathbf{x}$, which is shown in [29].

Lemma 11 (See [29]): Let $\mathbf{A} \in \mathbb{R}^{m \times n}$ be a random matrix whose entries are independently generated from standard Gaussian distribution, $\mathbf{x} \in \mathbb{R}^n$, and $\mathbf{y} = \text{sign}(\mathbf{A}\mathbf{x})$. With a probability at least $1 - e^{1-t}$, we have

$$\left\| \frac{1}{m} \mathbf{A}^T \mathbf{y} - \gamma \mathbf{x} \right\|_\infty \leq c \sqrt{\frac{t + \log(n)}{m}}$$

for some constant $c > 0$.

Since $\hat{\mathbf{x}}$ is the optimal solution, we have

$$\begin{aligned} & -\frac{1}{m} \langle \mathbf{A}^T \mathbf{y}, \hat{\mathbf{x}} \rangle + \lambda_1 \|\hat{\mathbf{x}}\|_1 + \lambda_2 \|\mathbf{D}\hat{\mathbf{x}}\|_1 \\ & \leq -\frac{1}{m} \langle \mathbf{A}^T \mathbf{y}, \mathbf{x} \rangle + \lambda_1 \|\mathbf{x}\|_1 + \lambda_2 \|\mathbf{D}\mathbf{x}\|_1. \end{aligned}$$

Thus, one has

$$\begin{aligned} & \lambda_1 \|\mathbf{x}\|_1 + \lambda_2 \|\mathbf{D}\mathbf{x}\|_1 \\ & \geq \langle \mathbf{x} - \hat{\mathbf{x}}, \frac{\mathbf{A}^T \mathbf{y}}{m} \rangle + \lambda_1 \|\hat{\mathbf{x}}\|_1 + \lambda_2 \|\mathbf{D}\hat{\mathbf{x}}\|_1 \\ & = \langle \mathbf{x} - \hat{\mathbf{x}}, \gamma \mathbf{x} \rangle + \langle \mathbf{x} - \hat{\mathbf{x}}, \frac{\mathbf{A}^T \mathbf{y}}{m} - \gamma \mathbf{x} \rangle + \lambda_1 \|\hat{\mathbf{x}}\|_1 \\ & \quad + \lambda_2 \|\mathbf{D}\hat{\mathbf{x}}\|_1 \\ & \geq \gamma(1 - \langle \hat{\mathbf{x}}, \mathbf{x} \rangle) - c \sqrt{\frac{t + \log(n)}{m}} \|\hat{\mathbf{x}} - \mathbf{x}\|_1 + \lambda_1 \|\hat{\mathbf{x}}\|_1 \\ & \quad + \lambda_2 \|\mathbf{D}\hat{\mathbf{x}}\|_1, \end{aligned}$$

which implies

$$\begin{aligned} \gamma(1 - \langle \hat{\mathbf{x}}, \mathbf{x} \rangle) & \leq \lambda_1 (\|\mathbf{x}\|_1 - \|\hat{\mathbf{x}}\|_1) + \lambda_2 (\|\mathbf{D}\mathbf{x}\|_1 - \|\mathbf{D}\hat{\mathbf{x}}\|_1) \\ & \quad + c \sqrt{\frac{t + \log(n)}{m}} \|\hat{\mathbf{x}} - \mathbf{x}\|_1. \end{aligned}$$

Recalling $\mathbf{x} \in \mathbb{K}_{s_1, s_2} \cap \mathbb{S}^n$ and $\lambda_1 = 2c \sqrt{\frac{t + \log(n)}{m}}$, one has

$$\begin{aligned} \gamma(1 - \langle \hat{\mathbf{x}}, \mathbf{x} \rangle) & \leq \frac{3}{2} \lambda_1 \|\mathbf{x}\|_1 - \frac{1}{2} \lambda_1 \|\hat{\mathbf{x}}\|_1 + \lambda_2 \|\mathbf{D}\mathbf{x}\|_1 \\ & \leq \frac{3}{2} \lambda_1 \sqrt{s_1} + 2\lambda_2 \sqrt{s_2}. \end{aligned}$$

Thus we have

$$\begin{aligned} \|\hat{\mathbf{x}} - \mathbf{x}\|_2^2 & \leq 2(1 - \langle \hat{\mathbf{x}}, \mathbf{x} \rangle) \\ & \leq \frac{1}{\gamma} (3\lambda_1 \sqrt{s_1} + 4\lambda_2 \sqrt{s_2}). \end{aligned}$$

So, we conclude that

$$\begin{aligned} \|\hat{\mathbf{x}} - \mathbf{x}\|_2 & \leq \sqrt{\frac{3\lambda_1 \sqrt{s_1} + 4\lambda_2 \sqrt{s_2}}{\gamma}} \\ & = \mathcal{O} \left(\sqrt[4]{\frac{(s_1 + s_2)(\log(n) + t)}{m}} \right). \end{aligned}$$

The proof is thus completed. \square

Proof of Theorem 8: Using Lemma 6 indicates that $\|\tilde{\mathbf{x}}\|_1 / \|\tilde{\mathbf{x}}\|_2 \leq \sqrt{s_1 + 1}$ and $\|\tilde{\mathbf{D}}\tilde{\mathbf{x}}\|_1 / \|\tilde{\mathbf{x}}\|_2 \leq 2\sqrt{s_2 + 1}$. Thus, applying Theorem 4 implies

$$\left\| \frac{\tilde{\mathbf{x}}^\#}{\|\tilde{\mathbf{x}}^\#\|_2} - \frac{\tilde{\mathbf{x}}}{\|\tilde{\mathbf{x}}\|_2} \right\|_2 \leq \sqrt{\frac{3\lambda_1 \sqrt{s_1 + 1} + 4\lambda_2 \sqrt{s_2 + 1}}{\gamma}} \leq \epsilon',$$

provided $m \geq C' \epsilon'^{-4} \gamma^{-2} (s_1 + k^2 s_2) (t + \log(n)) = C \epsilon^{-4} \gamma^{-2} (s_1 + k^2 s_2) (t + \log(n))$, where $\epsilon' = \frac{1}{2\sqrt{2}} \epsilon$ and $C = 64C'$.

In particular, looking at the last coordinate, this inequality yields

$$\left\| \frac{\tilde{x}_{n+1}^\#}{\|\tilde{\mathbf{x}}^\#\|_2} - \frac{\ell}{\|\tilde{\mathbf{x}}\|_2} \right\|_2 \leq \epsilon'.$$

Since $\|\tilde{\mathbf{x}}\|_2 \leq \sqrt{2}\ell$, we thus have

$$\frac{|\tilde{x}_{n+1}^\#|}{\|\tilde{\mathbf{x}}^\#\|_2} \geq \frac{\ell}{\|\tilde{\mathbf{x}}\|_2} - \epsilon' \geq \frac{1}{\sqrt{2}} - \epsilon \geq \frac{1}{2},$$

provided ϵ' is small enough.

In turn, applying Lemma 7 while taking $\tilde{\mathbf{f}} := \tilde{\mathbf{x}} = (\mathbf{x}; R)$ and $\tilde{\mathbf{g}} := \tilde{\mathbf{x}}^\# = (\mathbf{x}^\#; x^\#)$ into consideration gives

$$\left\| \frac{\mathbf{x}^\#}{x^\#} - \frac{\mathbf{x}}{\ell} \right\|_2 \leq 2\sqrt{2}\epsilon' = \epsilon.$$

Multiplying on both sides of this inequality by ℓ completes the proof. \square

Proof of Theorem 10: For any $\mathbf{x} \in \mathbb{V}_{s_1, s_2}(\ell)$, we shall prove by induction on $l \in \llbracket 0 : L \rrbracket$ that

$$\|\mathbf{x} - \mathbf{x}_l\|_F \leq (2\eta)^l \ell \quad (23)$$

holds with a high probability. When $l = 0$, Eq.(23) holds since $\mathbf{x}_l = \mathbf{0}$ and $\|\mathbf{x}\|_F \leq \ell$. When $l = 1$, triangle inequality indicates

$$\|\mathbf{x} - \mathbf{x}_1\|_2 \leq \|\mathbf{x} - \hat{\mathbf{x}}\|_2 + \|\hat{\mathbf{x}} - \mathbf{x}_1\|_2.$$

Since $\mathbf{x}_1 = \mathcal{P}_{\mathbb{V}_{s_1, s_2}(\ell)}(\hat{\mathbf{x}})$ is the best approximation to $\hat{\mathbf{x}}$ in $\mathbb{V}_{s_1, s_2}(\ell)$ and $\mathbf{x} \in \mathbb{V}_{s_1, s_2}(\ell)$, we have

$$\|\mathbf{x} - \mathbf{x}_1\|_2 \leq 2\|\mathbf{x} - \hat{\mathbf{x}}\|_2,$$

which combining Theorem 8 implies that

$$\|\mathbf{x} - \mathbf{x}_1\|_2 \leq 2\eta\ell$$

hold with a probability at least $1 - e^{-(1-t)}$, provided $q_1 \geq C\eta^{-4}\gamma^{-2}(s_1 + k^2s_2)(t + \log(n))$.

Now assume that Eq.(23) holds for $l - 1, l \in \{2, \dots, L\}$, we next prove that Eq.(23) holds for l . Define $\alpha_{l-1} = \min\{|\langle \mathbf{x}_{l-1}, \mathbf{i} \rangle| : \mathbf{i} \in \text{supp}(\mathbf{x}_{l-1})\}$. Noting that $\mathbf{x}, \mathbf{x}_{l-1} \in \mathbb{V}_{s_1, s_2}(\ell)$, we have

$$\begin{aligned} \|\mathbf{x} - \mathbf{x}_{l-1}\|_1 &\leq \sqrt{\frac{\|\mathbf{x}\|_1}{\alpha} + \frac{\|\mathbf{x}_{l-1}\|_1}{\alpha_{l-1}}} \|\mathbf{x} - \mathbf{x}_{l-1}\|_2 \\ &\leq \sqrt{\frac{\ell(\alpha_{l-1} + \alpha)\sqrt{s_1}}{\alpha\alpha_{l-1}}} \|\mathbf{x} - \mathbf{x}_{l-1}\|_2 \end{aligned}$$

and

$$\begin{aligned} \|D(\mathbf{x} - \mathbf{x}_{l-1})\|_1 &\leq 2\|\mathbf{x} - \mathbf{x}_{l-1}\|_1 \\ &\leq 2\sqrt{\frac{\ell(\alpha_{l-1} + \alpha)\sqrt{s_1}}{\alpha\alpha_{l-1}}} \|\mathbf{x} - \mathbf{x}_{l-1}\|_2. \end{aligned}$$

Thus, we get $\mathbf{x} - \mathbf{x}_{l-1} \in \mathbb{U}_{s'_{l-1}, s'_{l-1}} \cap (2\eta)^{l-1} \ell \mathbb{B}^n := \mathbb{V}_{s'_{l-1}, s'_{l-1}}((2\eta)^{l-1} \ell)$, where

$$s'_{l-1} = \frac{\ell(\alpha_{l-1} + \alpha)\sqrt{s_1}}{\alpha\alpha_{l-1}}.$$

Using Theorem 8 with $q^{(l)} \geq C\eta^{-4}\gamma^{-2}(1 + k^2)s'_{l-1}(t + \log(n))$ implies

$$\|\mathbf{x} - \mathbf{x}_{l-1} - \widehat{\mathbf{x}}_{l-1}\|_2 \leq 2^{l-1}\eta^l \ell. \quad (24)$$

Combining $\mathbf{x}_l = \mathcal{P}_{\mathbb{V}_{s_1, s_2}(\ell)}(\mathbf{x}_{l-1} + \widehat{\mathbf{x}}_{l-1})$ with Eq.(24), triangle inequality indicates that

$$\|\mathbf{x} - \mathbf{x}_l\|_2 \leq (2\eta)^l \ell$$

holds with a probability at least $1 - e^{(1-t)}$. This shows that the induction hypothesis for the iteration l . Fixing $t = \log(n)$, a union bound over all K iterations completes this proof. \square

REFERENCES

- [1] P. T. Boufounos and R. G. Baraniuk, "1-Bit compressive sensing," in *Proc. 43rd Asilomar Conf. Signals Syst. Comput.*, Mar. 2008, pp. 16–21.
- [2] 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.
- [3] Y. Plan and R. Vershynin, "Robust 1-bit compressed sensing and sparse logistic regression: A convex programming approach," *IEEE Trans. Inf. Theory*, vol. 59, no. 1, pp. 482–494, Jan. 2013.
- [4] L. J. Zhang, J. F. Yi, and R. Jin, "Efficient algorithms for robust one-bit compressive sensing," in *Proc. Int. Conf. Mach. Learn.*, Dec. 2014, pp. 820–828.
- [5] K. Knudson, R. Saab, and R. Ward, "One-bit compressive sensing with norm estimation," *IEEE Trans. Inf. Theory*, vol. 62, no. 2, pp. 2748–2758, May 2016.
- [6] R. G. Baraniuk, S. Foucart, D. Needell, Y. Plan, and M. Wootters, "Exponential decay of reconstruction error from binary measurements of sparse signals," *IEEE Trans. Inf. Theory*, vol. 63, no. 6, pp. 3368–3385, Jun. 2017.
- [7] J. Y. Hou, J. J. Wang, F. Zhang, and J. W. Huang, "One-bit compressed sensing via $\ell_p(0 < p < 1)$ -minimization method," *Inverse Probl.*, vol. 36, no. 5, 2020, Art. no. 055005.
- [8] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.
- [9] J. Mo, P. Schniter, N. G. Prelicic, and R. W. Heath, Jr., "Channel estimation in millimeter wave MIMO systems with one-bit quantization," in *Proc. 48th Asilomar Conf. Signals, Syst. Comput.*, Sep. 2014, pp. 957–961.
- [10] J. P. Xiong and Q. H. Tang, "1-bit compressive data gathering for wireless sensor networks," *J. Sensors*, vol. 2014, no. 7, pp. 177–183, 2014.
- [11] X. Dong and Y. Zhang, "A MAP approach for 1-Bit compressive sensing in synthetic aperture radar imaging," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 6, pp. 1237–1241, Jun. 2015.
- [12] CVX Research. (2014). *CVX: MATLAB Software for Disciplined Convex Programming, Version 2.1*. [Online]. Available: <http://cvxr.com/cvx>
- [13] S. Dirksen and S. Mendelson, "Non-Gaussian hyperplane tessellations and robust one-bit compressed sensing," *J. Eur. Math. Soc.*, vol. 23, no. 9, pp. 2913–2947, 2021.
- [14] R. Baraniuk, S. Foucart, D. Needell, Y. Plan, and M. Wootters, "One-bit compressive sensing of dictionary-sparse signals," *Inf. Inference, A J. IMA*, vol. 7, no. 1, pp. 83–104, Mar. 2018.
- [15] J.-F. Cai, B. Dong, S. Osher, and Z. Shen, "Image restoration: Total variation, wavelet frames, and beyond," *J. Amer. Math. Soc.*, vol. 25, no. 4, pp. 1033–1089, Oct. 2012.
- [16] Y. Wang, J. Yang, W. Yin, and Y. Zhang, "A new alternating minimization algorithm for total variation image reconstruction," *SIAM J. Imag. Sci.*, vol. 1, no. 3, pp. 248–272, Aug. 2008.
- [17] W. Wang and J. Wang, "Enhancing matrix completion using a modified second-order total variation," *Discrete Dyn. Nature Soc.*, vol. 2018, pp. 1–12, Sep. 2018.
- [18] D. Needell and R. Ward, "Near-optimal compressed sensing guarantees for total variation minimization," *IEEE Trans. Image Process.*, vol. 22, no. 10, pp. 3941–3949, Oct. 2013.
- [19] D. Needell and R. Ward, "Stable image reconstruction using total variation minimization," *SIAM J. Imag. Sci.*, vol. 6, no. 2, pp. 1035–1058, Mar. 2013.
- [20] J.-F. Cai and W. Xu, "Guarantees of total variation minimization for signal recovery," *Inf. Inference A, J. IMA*, vol. 4, no. 4, pp. 328–353, 2015.
- [21] A. Beck, *First-Order Methods in Optimization*. Philadelphia, PA, USA: Society for Industrial and Applied Mathematics (SIAM), 2017.
- [22] H. Garudadri, P. K. Baheti, S. Majumdar, C. Lauer, F. Masse, J. V. D. Molengraft, and J. Penders, "Artifacts mitigation in ambulatory ECG telemetry," in *Proc. 12th IEEE Int. Conf. e-Health Netw., Appl. Services*, Jul. 2010, pp. 338–344.

- [23] D. Angelosante, G. B. Giannakis, and E. Grossi, "Compressed sensing of time-varying signals," in *Proc. 16th Int. Conf. Digit. Signal Process.*, Jul. 2009, pp. 1–8.
- [24] R. Tibshirani, M. Saunders, S. Rosset, J. Zhu, and K. Knight, "Sparsity and smoothness via the fused lasso," *J. Roy. Statist. Soc. B, Statist. Methodol.*, vol. 67, no. 1, pp. 91–108, 2005.
- [25] X. X. Li, L. L. Mo, X. M. Yuan, and J. Z. Zhang, "Linearized alternating direction method of multipliers for sparse group and fused LASSO models," *Comput. Stat. Data Anal.*, vol. 79, pp. 203–221, Nov. 2014.
- [26] X. Zeng and M. A. T. Figueiredo, "Binary fused compressive sensing: 1-bit compressive sensing meets group sparsity," 2014, *arXiv:1402.5074*.
- [27] N. Koep and R. Mathar, "Block-sparse signal recovery from binary measurements," in *Proc. IEEE Stat. Signal Process. Workshop (SSP)*, Jun. 2018, pp. 293–297.
- [28] L. Condat, "A direct algorithm for 1-D total variation denoising," *IEEE Signal Process. Lett.*, vol. 20, no. 11, pp. 1054–1057, Nov. 2013.
- [29] J. Liu, L. Yuan, and J. Ye, "An efficient algorithm for a class of fused lasso problems," in *Proc. 16th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining (KDD)*, 2010, pp. 323–332.
- [30] G. B. Moody and R. G. Mark, "The impact of the MIT-BIH arrhythmia database," *IEEE Eng. Med. Biol. Mag.*, vol. 20, no. 3, pp. 45–50, May 2001.
- [31] S. Boyd, L. Xiao, and A. Mutapcic, "Subgradient methods," Autumn Quarter, Stanford Univ., Stanford, CA, USA, Lecture Notes EE392o, 2003, pp. 2004–2005, vol. 2004.
- [32] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge, U.K.: Cambridge Univ. Press, 2004.



JINGYAO HOU received the B.Sc. degree in mathematics and applied mathematics from the School of Mathematics and Statistics, Chongqing Three Gorges University, Chongqing, China, in 2014, and the M.Sc. and Ph.D. degrees in statistics from the School of Mathematics and Statistics, Southwest University, Chongqing, in 2017 and 2021, respectively. He is currently a Research Associate with the College of Mathematics and Information, China West Normal University, Nanchong, China. His work has been funded by the National Natural Science Foundation of China. His research interests include one-bit compressed sensing and tensor sparsity.



XINLING LIU received the B.S. degree in mathematics and applied mathematics and the M.S. degree in probability theory and mathematical statistics from Southwest University, Chongqing, China, in 2012 and 2015, respectively, where he is currently pursuing the Ph.D. degree with the School of Mathematics and Statistics. His current research interests include high-dimensional data analysis, machine learning, compressive sensing, and approximation theory.

• • •