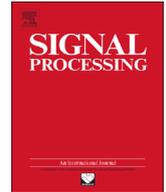




ELSEVIER

Contents lists available at ScienceDirect

Signal Processing

journal homepage: www.elsevier.com/locate/sigproAdaptive one-bit quantization for compressed sensing[☆]Jun Fang^a, Yanning Shen^b, Linxiao Yang^a, Hongbin Li^{c,*}^a National Key Laboratory of Science and Technology on Communications, University of Electronic Science and Technology of China, Chengdu 611731, China^b Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis, MN 55455, USA^c Department of Electrical and Computer Engineering, Stevens Institute of Technology, Hoboken, NJ 07030, USA

ARTICLE INFO

Article history:

Received 25 March 2015

Received in revised form

8 January 2016

Accepted 25 January 2016

Available online 5 February 2016

Keywords:

One-bit compressed sensing

Adaptive quantization

Quantization design

ABSTRACT

There have been a number of studies on sparse signal recovery from one-bit quantized measurements. Nevertheless, less attention has been paid to the choice of the quantization thresholds and its impact on the signal recovery performance. In this paper, we examine the problem of quantization in a general framework of one-bit compressed sensing with non-zero quantization thresholds. Our analysis shows that when the number of one-bit measurements is sufficiently large, with a high probability the sparse signal can be recovered with an error decaying linearly with the ℓ_2 -norm of the difference between the quantization thresholds and the original unquantized measurements. Specifically, by setting the thresholds sufficiently close to the original unquantized measurements, sparse signals can be recovered with an arbitrarily small error. By borrowing an idea from the Delta modulation, we propose an adaptive quantization scheme where the quantization thresholds are iteratively adjusted based on previous encoded bits such that they eventually oscillate around the original unquantized measurements with decreasing granular noise. Numerical results are provided to collaborate our theoretical results and to illustrate the effectiveness of the proposed scheme.

© 2016 Elsevier B.V. All rights reserved.

1. Introduction

Compressive sensing is a recently emerged paradigm of signal sampling and reconstruction, the main purpose of which is to recover sparse signals from much fewer linear measurements [1,2]

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

where $\mathbf{A} \in \mathbb{R}^{m \times n}$ is the sampling matrix with $m \ll n$, and \mathbf{x} denotes the n -dimensional sparse signal with only K non-zero coefficients. Such a problem has been extensively studied and a variety of algorithms that provide consi-

stent recovery performance guarantee were proposed, e.g. [1–12]. Conventional compressed sensing assumes infinite precision of the acquired measurements. In practice, however, signals need to be quantized before further processing, that is, the real-valued measurements need to be mapped to discrete values over some finite range. Besides, in some sensing systems (e.g. distributed sensor networks), data acquisition is expensive due to limited bandwidth and energy constraints [13]. Aggressive quantization strategies which compress real-valued measurements into one or only a few bits of data are preferred in such scenarios. Another benefit brought by low-rate quantization is that it can significantly reduce the hardware complexity and cost of the analog-to-digital converter (ADC).

Inspired by practical necessity and potential benefits, compressed sensing based on quantized measurements has attracted considerable attention recently. A common way to handle quantized measurements is to treat quantization errors as Gaussian noise, in which case the

[☆] This work was supported in part by the National Science Foundation of China under Grant 61172114, 61428103, U1530154, and the National Science Foundation under Grant ECCS-1408182.

* Corresponding author. Tel.: +1 201 216 5604; fax: +1 201 216 8264.

E-mail addresses: JunFang@uestc.edu.cn (J. Fang), shenx513@umn.edu (Y. Shen), Hongbin.Li@stevens.edu (H. Li).

classical basis pursuit denoising (BPDN) method [1,14] can be used to reconstruct sparse signals. Nevertheless, it has been theoretically proven [15] and numerically demonstrated [16,17] that the reconstruction accuracy can be considerably improved by taking into account the quantization consistency constraint or choosing a more suitable data-fidelity constraint. The saturation issue arising in quantization was studied in [18], where by exploiting the democracy property, it was shown that approaches that merely reject saturated measurements or consider the consistency constraint are capable of providing a stable signal recovery. The performance of sparse signal recovery from quantized measurements was also studied in [19,20].

Besides the above studies, a large body of quantized compressed sensing work has focused on the extreme case, i.e. one-bit compressed sensing which was originally introduced in [21] and aims at recovering a sparse or compressible signal from one-bit measurements

$$\mathbf{b} = \text{sign}(\mathbf{y}) = \text{sign}(\mathbf{A}\mathbf{x}) \quad (2)$$

where “sign” denotes an operator that performs the sign function element-wise on the vector, the sign function returns 1 for positive numbers and -1 otherwise. The reconstruction performance from one-bit measurements was later studied in [22,23]. Specifically, in [22], a lower bound on the reconstruction error was provided along with a binary iterative hard thresholding (BIHT) algorithm for recovery of sparse signals. The lower bound holds regardless of the reconstruction method used, and it was shown that the reconstruction error decays at best linearly with the number of measurements. In [23], a first computationally tractable method with provable error guarantees for one-bit compressed sensing was proposed. Inspired by the theoretical advancement [22–24], a variety of efficient one-bit compressed sensing algorithms such as adaptive outlier pursuit method [25], noise-adaptive renormalized fixed point iteration method [26], robust one-bit Bayesian compressed sensing [27], and many others (e.g. [28,29]) were developed. Most of the above works, however, assume a zero quantization threshold, in which case only the sign of the measurement is retained while the information about the magnitude of the signal is lost. The aforementioned bound on the reconstruction error therefore is only concerned about the direction of \mathbf{x} , i.e. $\mathbf{x}/\|\mathbf{x}\|_2$.

In [30], a more general measurement model with a non-zero threshold vector $\boldsymbol{\tau}$ was considered:

$$\mathbf{b} = \text{sign}(\mathbf{y} - \boldsymbol{\tau}) = \text{sign}(\mathbf{A}\mathbf{x} - \boldsymbol{\tau}) \quad (3)$$

It was shown [30] that with appropriately chosen thresholds, both the magnitude and the direction of \mathbf{x} can be estimated from one-bit measurements. In addition, in recent work [28,31,32], it was numerically demonstrated [28] and theoretically proved [31,32] that adaptively choosing the quantization thresholds (in a manner similar to the Sigma-Delta quantization [33]) is able to achieve an exponential decay of the reconstruction error as a function of the number of measurements. To achieve this exponential error decay, the quantization threshold τ_i of the i th entry needs to be adaptively adjusted based on the estimate of \mathbf{x} obtained from previous quantized

measurements. This iterative process therefore requires a feedback from the decoder to the encoder and the decoder involves a computationally intensive task of calculation of the quantization threshold.

In this paper, we investigate the problem of one-bit quantization design for sparse signal recovery. We provide a quantitative analysis which examines the choice of the quantization thresholds and quantifies its impact on the reconstruction performance. Our analysis shows that when the number of one-bit measurements is sufficiently large, with a high probability the sparse signal can be recovered with an error decaying linearly with the ℓ_2 -norm of the difference between the quantization thresholds and the original unquantized measurements. Thus the reconstruction error can be made arbitrarily small by setting the quantization thresholds sufficiently close to the original unquantized measurements \mathbf{y} . Since the original unquantized data samples \mathbf{y} are inaccessible to the decoder, we propose an adaptive quantization approach which iteratively adjusts the quantization thresholds such that the thresholds eventually come close and oscillate around the original data samples. Note that our adaptive quantization approach is very different from the adaptive scheme developed in [28,31]. The quantization thresholds τ are updated in parallel for our approach, while for [28,31], the thresholds are adjusted in a sequential manner, with its i th quantization threshold τ_i depending on the previous $(i-1)$ quantized measurements. In contrast to [28,31], our adaptive quantization approach does not require a feedback from the decoder to the encoder. Besides, the computation of the thresholds at both the encoder and the decoder involves very simple algebraic operations that can be easily implemented.

There have been some other interesting work [34–36] on quantizer design for compressed sensing. Specifically, the work [34] utilizes the high-resolution distributed functional scalar quantization theory for quantizer design, where the quantization error is modeled as random noise following a certain distribution. Nevertheless, such modeling holds valid only for high-resolution quantization, and may bring limited benefits when a low-rate quantization strategy is adopted. In [35], authors proposed a generalized approximate message passing (GAMP) algorithm for quantized compressed sensing, and studied the quantizer design under the GAMP reconstruction. The problem of reconstructing image from one-bit quantized data was considered in [36], where the quantization thresholds are set such that the plus and minus binary measurements are acquired with equal probabilities. Albeit empirically effective, no rigorous theoretical guarantee was available to justify the proposed scheme.

The rest of the paper is organized as follows. In Section 2, we introduce the one-bit compressed sensing problem. The main results of this paper are presented in Section 3, where we show that, by properly selecting the quantization thresholds, sparse signals can be recovered from one-bit measurements with an arbitrarily small error. A rigorous proof of our main result is provided in Section 4. An adaptive quantization method is developed in 5 to iteratively refine the thresholds based on previous encoded bits. In Section 6, numerical results are presented to corroborate our theoretical analysis and illustrate the

effectiveness of the proposed algorithm, followed by concluding remarks in Section 7.

2. Problem formulation

We consider a coarse quantization-based signal acquisition model in which each real-valued sample is encoded into one-bit of information

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

where $\mathbf{y} = [y_1 \ y_2 \ \dots \ y_m]^T$ denotes the unquantized original measurements, $\mathbf{b} = [b_1 \ b_2 \ \dots \ b_m]^T$ are the binary observations, and $\boldsymbol{\tau} = [\tau_1 \ \tau_2 \ \dots \ \tau_m]^T$ denotes the quantization threshold vector. The problem of interest is to recover the sparse signal \mathbf{x} from the one-bit quantized data \mathbf{b} and examine the choice of the quantization threshold $\boldsymbol{\tau}$. Specifically, we consider the following canonical form for sparse signal reconstruction

$$\min_{\mathbf{z}} \|\mathbf{z}\|_0 \quad \text{s.t.} \quad \text{sign}(\mathbf{A}\mathbf{z} - \boldsymbol{\tau}) = \mathbf{b} \quad (5)$$

Such an optimization problem, albeit non-convex, is more amenable for theoretical analysis than its convex counterpart (42).

As mentioned earlier, in most previous one-bit compressed sensing studies, the quantization thresholds are set equal to zero, i.e. $\boldsymbol{\tau} = \mathbf{0}$. In this case, we have to constrain the sparse signal on the unit-sphere, otherwise solving the optimization Eq. (5) yields a trivial all-zero solution because multiplying \mathbf{x} by an arbitrarily small positive scaling factor will result in the same quantized data \mathbf{b} . To circumvent the above issue, a unit-norm constraint is included in Eq. (5) to restrict the sparse signal on the unit-sphere, i.e.

$$\min_{\mathbf{z}} \|\mathbf{z}\|_0 \quad \text{s.t.} \quad \text{sign}(\mathbf{A}\mathbf{z}) = \mathbf{b}, \quad \|\mathbf{z}\|_2 = 1 \quad (6)$$

The recovery performance, under this circumstance, is evaluated by $\|\mathbf{x}/\|\mathbf{x}\|_2 - \hat{\mathbf{x}}\|_2$, where \mathbf{x} is the sparse signal and $\hat{\mathbf{x}}$ denotes the estimated signal obtained via solving the optimization (6). It was shown in [22] that the above defined reconstruction error can be bounded with an error order decaying linearly with the number of measurements.

In this paper, we consider the optimization Eq. (5) where a general non-zero threshold vector is employed for sparse signal recovery. Note that for a non-zero threshold vector, the unit-norm constraint is no longer necessary since the trivial all-zero solution can usually be avoided. Moreover, imposing the unit-norm constraint results in recovery failure as the magnitude of the sparse signal, $\|\mathbf{x}\|_2$, is now one of the factors that determine the signs of the quantized data. That is, even if the thresholds $\boldsymbol{\tau}$ remain unchanged, multiplying \mathbf{x} by different scaling factors leads to different quantized data \mathbf{b} . For notational convenience, we allow a slight abuse of notation and let $\hat{\mathbf{x}}$ denote a solution of the optimization Eq. (5). In the following, we examine the choice of $\boldsymbol{\tau}$ and its impact on the reconstruction error $\|\mathbf{x} - \hat{\mathbf{x}}\|_2$.

3. One-bit quantization design: analysis

To facilitate our analysis, we decompose the quantization threshold vector $\boldsymbol{\tau}$ into a sum of two terms:

$$\boldsymbol{\tau} = \mathbf{y} - \boldsymbol{\delta} \quad (7)$$

where $\boldsymbol{\delta} = [\delta_1 \ \delta_2 \ \dots \ \delta_m]^T$ is the difference vector between the quantization thresholds $\boldsymbol{\tau}$ and the original unquantized measurements \mathbf{y} . Substituting Eq. (7) into Eq. (4), we get

$$\mathbf{b} = \text{sign}(\mathbf{A}\mathbf{x} - \boldsymbol{\tau}) = \text{sign}(\boldsymbol{\delta}) \quad (8)$$

Suppose $\hat{\mathbf{x}}$ is the solution of Eq. (5). Let $\mathbf{h} \triangleq \hat{\mathbf{x}} - \mathbf{x}$ be the residual (reconstruction error) vector. Clearly, \mathbf{h} is a $2K$ -sparse vector which has at most $2K$ non-zero entries since $\hat{\mathbf{x}}$ has at most K non-zero coefficients. Also, the solution $\hat{\mathbf{x}}$ yields estimated measurements that are consistent with the observed binary data, i.e.

$$\mathbf{b} = \text{sign}(\mathbf{A}\hat{\mathbf{x}} - \boldsymbol{\tau}) = \text{sign}(\mathbf{A}\mathbf{h} + \boldsymbol{\delta}) \quad (9)$$

Combining Eqs. (8) and (9), the residual vector \mathbf{h} has to satisfy the following constraint:

$$\text{sign}(\mathbf{A}\mathbf{h} + \boldsymbol{\delta}) = \text{sign}(\boldsymbol{\delta}) \quad (10)$$

In the following, we will show that the residual vector \mathbf{h} can be bounded by the difference vector $\boldsymbol{\delta}$ if the number of measurements are sufficiently large and \mathbf{A} satisfies a certain condition. Thus the reconstruction error \mathbf{h} can be made arbitrarily small by setting $\boldsymbol{\delta}$ close to zero. Our main results are summarized as follows.

Theorem 1. *Let $\mathbf{x} \in \mathbb{R}^n$ be a K -sparse vector. $\mathbf{A} \in \mathbb{R}^{m \times n}$ is the sampling matrix. Suppose there exist an integer $\kappa \geq 2K$ and a positive parameter μ such that any $\kappa \times n$ submatrix $\bar{\mathbf{A}}$ constructed by selecting certain rows of \mathbf{A} satisfies*

$$\|\bar{\mathbf{A}}\mathbf{u}\|_2 \geq \mu \|\mathbf{u}\|_2 \quad (11)$$

for any $2K$ -sparse vector \mathbf{u} . Also, assume that each entry of $\boldsymbol{\delta}$ is independently generated from a certain distribution with equal probabilities being positive or negative. Let $\hat{\mathbf{x}}$ denote the solution of the optimization problem Eq. (5). For any arbitrarily small value $\eta > 0$, we can ensure that the following statement is true with probability exceeding $1 - \eta$: if the number of measurements, m , is sufficiently large and satisfies

$$m > 2K \log(mn/K^2) + \kappa \log m + \log(1/\eta) + c \quad (12)$$

then the sparse signal can be recovered from Eq. (5) with the reconstruction error bounded by

$$\|\hat{\mathbf{x}} - \mathbf{x}\|_2 \leq \frac{\epsilon}{\sqrt{\mu}} \triangleq \lambda \epsilon \quad (13)$$

where $\epsilon \triangleq \|\boldsymbol{\delta}\|_2$, $\lambda \triangleq 1/\sqrt{\mu}$, and c in (12) is defined as

$$c \triangleq (\kappa - 1)(\log(e/(\kappa - 1)) + 1) + 4K \log(e/2) + 2K \quad (14)$$

which is a constant only dependent on κ and K , where e represents the base of the natural logarithm.

From Eq. (13), we see that the sparse signal can be recovered with an error decaying linearly with the ℓ_2 -norm of the difference between the quantization thresholds and the original unquantized measurements. This result suggests that the quantization thresholds should be set as close to the original data samples \mathbf{y} as possible. In particular, sparse signals can be reconstructed with

negligible errors even from one-bit measurements if we set the threshold vector τ sufficiently close to the unquantized data \mathbf{y} (Setting thresholds arbitrarily close to unquantized measurements, of course, is impossible in practice).

Note that for a randomly generated measurement matrix \mathbf{A} whose entries are independently drawn from a normal distribution, since any $\kappa \times 2K$ sub-matrix of \mathbf{A} is full rank with probability one for $\kappa \geq 2K$, we can always find an integer κ and a sufficiently small positive μ to satisfy Eq. (11). The problem is that μ that satisfies Eq. (11) could be very small¹, which makes the error bound Eq. (13) loose or even meaningless. Nevertheless, note that the bound Eq. (13) is given by $\epsilon/\sqrt{\mu}$. Hence even for a very small μ , if the quantization threshold is sufficiently close to the unquantized measurement such that ϵ is as small as μ , a meaningful bound can still be obtained. Of course, assuming a quantization threshold arbitrarily close to the unquantized measurement may be impractical. In practice, we may obtain a meaningful bound by increasing the value of κ . It has been proved [37] that for any $p \times q$ matrix \mathbf{A} with independent and identically distributed sub-Gaussian entries, the smallest singular value of \mathbf{A} is at least of order $(\sqrt{p} - \sqrt{q} - 1)$ with high probability, that is, with probability exponentially (in dimension) close to one. This result implies that the smallest singular value improves as the number of rows grows. Therefore as κ grows, μ may be large enough to provide a meaningful bound for the reconstruction error. It is, however, very difficult to analyze the behavior of the smallest singular value of all possible sub-matrices because we have to take a union bound over all sub-matrices and the sub-matrices are not mutually independent due to overlapping. Therefore we cannot answer how large κ needs to be in order to avoid a too small μ and provide a meaningful bound for the reconstruction error. On the other hand, the key point of our theorem is to show that the reconstruction error can be reduced by setting the quantization thresholds close to the original unquantized measurements. The exact value of κ is not our concern as long as there exists such a reasonable value of κ to provide a meaningful μ .

We also noticed that random Gaussian matrices possess a “democratic” property [38] which allows a meaningful μ to satisfy the condition (11). The democratic property means that after removing a certain number of rows from a random Gaussian matrix, the resulting sub-matrix still preserves some restricted isometry property (RIP). Although the required number of rows (i.e., κ) to satisfy the democratic property is very large, considering the fact that a RIP condition is more restrictive than the condition (11), we believe a moderately large κ should be able to avoid a too small μ and ensure a meaningful error bound.

4. Proof of Theorem 1

Let \mathbf{a}_i^T denote the i th row of the sampling matrix \mathbf{A} . Clearly, to ensure the sign consistency (10) for each component, we should either have

$$\text{sign}(\mathbf{a}_i^T \mathbf{h}) = \text{sign}(\delta_i) \quad (15)$$

or

$$|\mathbf{a}_i^T \mathbf{h}| < |\delta_i| \quad (16)$$

Note that in the latter case the sign of $(\mathbf{a}_i^T \mathbf{h} + \delta_i)$ does not flip regardless of the sign of $\mathbf{a}_i^T \mathbf{h}$. Therefore if there exists a $2K$ -sparse residual vector \mathbf{h} such that each component of $\mathbf{A}\mathbf{h}$ satisfies either Eqs. (15) or (16), then $\hat{\mathbf{x}} = \mathbf{x} + \mathbf{h}$ is the solution of Eq. (5). Our objective in the following is to show that such a residual vector \mathbf{h} is bounded by the deviation vector δ with an arbitrarily high probability.

Without loss of generality, we decompose \mathbf{A} and δ into two parts: $\mathbf{A} = [\mathbf{A}_1^T \ \mathbf{A}_2^T]^T$, $\delta = [\delta_1^T \ \delta_2^T]^T$, according to the two possible relationships between $\mathbf{a}_i^T \mathbf{h}$ and δ_i :

$$\begin{aligned} \text{sign}(\mathbf{A}_1 \mathbf{h}) &= \text{sign}(\delta_1) \\ |\mathbf{A}_2 \mathbf{h}| &< |\delta_2| \end{aligned} \quad (17)$$

where $\mathbf{A}_1 \in \mathbb{R}^{m_1 \times n}$, $\mathbf{A}_2 \in \mathbb{R}^{m_2 \times n}$, $m_1 + m_2 = m$, and in the second equation, both the absolute value operation $|\cdot|$ and the inequality symbol “ $<$ ” applies entrywise to vectors.

We now analyze the probability of the residual vector \mathbf{h} being greater than $\lambda\epsilon$, given that the condition (17) is satisfied. This conditional probability can be denoted as $P(\|\mathbf{h}\|_2 > \lambda\epsilon | \text{E})$, where we use “E” to denote the event (17). Clearly, $P(\|\mathbf{h}\|_2 > \lambda\epsilon | \text{E})$ can also be explained as the probability that the event $\|\mathbf{h}\|_2 > \lambda\epsilon$ will occur, when the event E has occurred. To facilitate our analysis, we divide the event (17) into two disjoint sub-events which are defined as

E1: The event (17) holds true for $m_2 \geq \kappa$

E2: The event (17) holds true for $m_2 < \kappa$

Clearly, the union of these two sub-events is equal to the event (17). Utilizing Bayes' Theorem, the probability $P(\|\mathbf{h}\|_2 > \lambda\epsilon | \text{E})$ can be expressed as

$$\begin{aligned} P(\|\mathbf{h}\|_2 > \lambda\epsilon | \text{E}) &\stackrel{(a)}{=} P(\|\mathbf{h}\|_2 > \lambda\epsilon, \text{E}) \\ &= P(\|\mathbf{h}\|_2 > \lambda\epsilon, \text{E1} \cup \text{E2}) \\ &\stackrel{(b)}{=} P(\|\mathbf{h}\|_2 > \lambda\epsilon, \text{E1}) + P(\|\mathbf{h}\|_2 > \lambda\epsilon, \text{E2}) \end{aligned} \quad (18)$$

where (a) holds because the events (17) is a prerequisite condition that is always met to ensure the sign consistency, i.e. $P(\text{E}) = 1$, and (b) follows from the fact that the probability of the union of two disjoint events is equal to the sum of their respective probabilities.

Let us first examine the probability $P(\|\mathbf{h}\|_2 > \lambda\epsilon, \text{E1})$. We show that the events $\|\mathbf{h}\|_2 > \lambda\epsilon$ and E1 are two mutually exclusive events which cannot occur at the same time. To see this, note that when the event E1 occurs, we should have

$$|\mathbf{A}_2 \mathbf{h}| < |\delta_2| \Rightarrow \|\mathbf{A}_2 \mathbf{h}\|_2^2 < \epsilon^2 \quad (19)$$

in which $\mathbf{A}_2 \in \mathbb{R}^{m_2 \times n}$ and $m_2 \geq \kappa$. On the other hand, recalling that any $\kappa \times n$ sub-matrix formed by selecting

¹ To meet Eq. (11), $\sqrt{\mu}$ has to be equal to or smaller than the minimum singular value of all possible $\kappa \times 2K$ sub-matrices of \mathbf{A} .

certain rows of \mathbf{A} satisfies the condition (11), we have

$$\|\mathbf{A}_2 \mathbf{h}\|_2^2 \geq \mu \|\mathbf{h}\|_2^2 \quad (20)$$

Combining Eqs. (19) and (20), we arrive at

$$\|\mathbf{h}\|_2 < \frac{\epsilon}{\sqrt{\mu}} \triangleq \lambda\epsilon \quad (21)$$

which is contradictory to the event $\|\mathbf{h}\|_2 > \lambda\epsilon$. Hence the events E1 and $\|\mathbf{h}\|_2 > \lambda\epsilon$ cannot occur simultaneously, which implies

$$P(\|\mathbf{h}\|_2 > \lambda\epsilon, E1) = 0 \quad (22)$$

Substituting Eq. (22) into (18), the probability $P(\|\mathbf{h}\|_2 > \lambda\epsilon | E)$ is simplified as

$$P(\|\mathbf{h}\|_2 > \lambda\epsilon | E) = P(\|\mathbf{h}\|_2 > \lambda\epsilon, E2) \quad (23)$$

The probability $P(\|\mathbf{h}\|_2 > \lambda\epsilon, E2)$, however, is still difficult to analyze. To circumvent this difficulty, we, instead, derive an upper bound on $P(\|\mathbf{h}\|_2 > \lambda\epsilon, E2)$:

$$\begin{aligned} & P(\|\mathbf{h}\|_2 > \lambda\epsilon, E2) \\ &= P(\|\mathbf{h}\|_2 > \lambda\epsilon | E2)P(E2) \leq P(E2) \\ &\stackrel{(a)}{=} P(\text{sign}(\mathbf{A}_1 \mathbf{h}) = \text{sign}(\delta_1), |\mathbf{A}_2 \mathbf{h}| < |\delta_2|, m_2 < \kappa) \\ &\leq P(\text{sign}(\mathbf{A}_1 \mathbf{h}) = \text{sign}(\delta_1), m_2 < \kappa) \\ &= P(\text{sign}(\mathbf{A}_1 \mathbf{h}) = \text{sign}(\delta_1), m_1 > m - \kappa) \\ &\stackrel{(b)}{=} P\left(\bigcup_{i=0}^{\kappa-1} \Omega_i\right) \end{aligned} \quad (24)$$

where (a) comes from the definition of the event E2, and in (b), the event Ω_i is defined as

$$\begin{aligned} \Omega_i: \text{sign}(\mathbf{A}_1 \mathbf{h}) &= \text{sign}(\delta_1) \\ \text{where } \mathbf{A}_1 &\in \mathbb{R}^{m_1 \times n}, \text{ and } m_1 = m - i \end{aligned} \quad (25)$$

which means that there exist at least $m - i$ components in $\mathbf{A}\mathbf{h}$ whose signs are consistent with the corresponding entries in δ . Note that since the rest i components are not explicitly specified in Ω_i , the event Ω_i include all possibilities for the rest i components. Based on the definition of Ω_i , we can infer the following relationship: $\Omega_{i_1} \supset \Omega_{i_2}$ for $i_1 > i_2$. This is because Ω_{i_2} can be regarded as a special case of the event Ω_{i_1} with some of the unspecified i_1 components also meeting the sign consistency requirement. With this relation, the upper bound derived in Eq. (24) can be simplified as

$$P(\|\mathbf{h}\|_2 > \lambda\epsilon, E2) \leq P\left(\bigcup_{i=0}^{\kappa-1} \Omega_i\right) = P(\Omega_{\kappa-1}) \quad (26)$$

We note that for the event $\Omega_{\kappa-1}$, \mathbf{A}_1 is not specified and can be any sub-matrix of \mathbf{A} . Considering selection of $m - \kappa + 1$ rows (out of m rows of \mathbf{A}) to construct \mathbf{A}_1 , the event $\Omega_{\kappa-1}$ can be expressed as a union of a set of sub-events

$$\Omega_{\kappa-1} = \bigcup_{j=1}^J \Omega_{\kappa-1}^j \quad (27)$$

where $J \triangleq C(m, m - \kappa + 1)$, $C(m, k)$ denotes the number of k combinations from a given set of m -elements, and each sub-event $\Omega_{\kappa-1}^j$ is defined as

$$\Omega_{\kappa-1}^j: \text{sign}(\mathbf{A}_1 \mathbf{h}) = \text{sign}(\delta_1) \quad \text{where } \mathbf{A}_1 = \mathbf{A}[I_j, :] \quad (28)$$

in which I_j is an unique index set which consists of

$m - \kappa + 1$ non-identical indices selected from $\{1, 2, \dots, m\}$, $\mathbf{A}[I_j, :]$ denotes a sub-matrix of \mathbf{A} constructed by certain rows from \mathbf{A} , and the indices of the selected rows are specified by I_j . From Eq. (27), we have

$$P(\Omega_{\kappa-1}) = P\left(\bigcup_{j=1}^J \Omega_{\kappa-1}^j\right) \stackrel{(a)}{\leq} \sum_{j=1}^J P(\Omega_{\kappa-1}^j) \quad (29)$$

where the inequality (a) follows from the fact that the probability of a union of events is no greater than the sum of probabilities of respective events. The inequality becomes an equality if the events are disjoint. Nevertheless, the sub-events $\{\Omega_{\kappa-1}^j\}$ are not necessarily disjoint and may have overlappings due to the $\kappa - 1$ unspecified components.

We now analyze the probability $P(\Omega_{\kappa-1}^j)$. To begin with our analysis, we introduce the concept of orthant originally proposed in [22] for analysis of one-bit compressed sensing. An orthant in \mathbb{R}^m is a set of vectors that share the same sign pattern, i.e.

$$\mathcal{O}_{\mathbf{u}} = \{\mathbf{u} \in \mathbb{R}^m | \text{sign}(\mathbf{u}) = \mathbf{u}\} \quad (30)$$

A useful result concerning intersections of orthants by subspaces is summarized as follows.

Lemma 1. *Let S be an arbitrary k -dimensional subspace in an m -dimensional space. Then the number of orthants intersected by S can be upper bounded by*

$$l(m, k) \leq 2^k C(m, k) \quad (31)$$

where $C(m, k)$ denotes the number of k -combinations from a set of n -elements.

Proof. See [22, Lemma 1].

The probability $P(\Omega_{\kappa-1}^j)$ of our interest can be interpreted as, the probability of the vector $\mathbf{A}_1 \mathbf{h}$ lying in the same orthant as δ_1 for a given \mathbf{A}_1 . We first examine the number of sign patterns the vector $\mathbf{A}_1 \mathbf{h}$ could possibly have. Let S denote the set of all possible sign patterns for $\mathbf{A}_1 \mathbf{h}$, i.e.

$$\begin{aligned} S = \{\mathbf{u} = \text{sign}(\mathbf{A}_1 \mathbf{h}) | \mathbf{h} \in \mathbb{R}^n \text{ is a } 2K\text{-sparse vector,} \\ \mathbf{A}_1 = \mathbf{A}[I_j, :]\} \end{aligned} \quad (32)$$

Also, let T denote the support of the sparse residual vector \mathbf{h} , we can write

$$\mathbf{g} \triangleq \mathbf{A}_1 \mathbf{h} = \mathbf{A}_1[:, T] \mathbf{h}_T \quad (33)$$

where $\mathbf{g} \in \mathbb{R}^{m - \kappa + 1}$, and $\mathbf{A}_1[:, T]$ denotes a submatrix of \mathbf{A}_1 obtained by concatenating columns whose indices are specified by T . We see that \mathbf{g} is a linear combination of $2K$ columns of \mathbf{A}_1 , and thus \mathbf{g} lies in an $2K$ -dimensional subspace spanned by the columns of $\mathbf{A}_1[:, T]$. Recalling Lemma 1, we know that the number of orthants intersected by this subspace is upper bounded by $2^{2K} C(m - \kappa + 1, 2K)$. Therefore the vector \mathbf{g} which lies in this subspace has at most $2^{2K} C(m - \kappa + 1, 2K)$ possible sign patterns. Note that this result is for a specific choice of the index set T . The selection of the support T from n entries has at most $C(n, 2K)$ combinations. Therefore, in summary, the number of sign patterns in the set S is upper bounded by

$$N_{SP} \leq 2^{2K} C(m - \kappa + 1, 2K) C(n, 2K) \quad (34)$$

The probability $P(\Omega_{\kappa-1}^j)$ can be calculated as

$$P(\Omega_{\kappa-1}^j) = P(\text{sign}(\delta_1) \in S) \stackrel{(a)}{=} \frac{N_{\text{SP}}}{2^{(m-\kappa+1)}} \leq \frac{2^{2K} C(m-\kappa+1, 2K) C(n, 2K)}{2^{(m-\kappa+1)}} \quad (35)$$

where (a) comes from the fact that δ is a vector whose entries are independently generated from a certain distribution with equal probabilities being positive and negative, and δ_1 has $2^{(m-\kappa+1)}$ possible sign patterns.

Combining Eqs. (23), (26), (29), (35), we arrive at

$$P(\|\mathbf{h}\|_2 > \lambda\epsilon|\mathbf{E}) = P(\|\mathbf{h}\|_2 > \lambda\epsilon|\mathbf{E}2) \leq P(\Omega_{\kappa-1}) \leq \frac{2^{2K} C(m-\kappa+1, 2K) C(n, 2K) C(m, \kappa-1)}{2^{(m-\kappa+1)}} \quad (36)$$

where the last inequality comes from $J = C(m, m-\kappa+1) = C(m, \kappa-1)$. Utilizing the following inequality [22]

$$C(a, b) \leq \left(\frac{ae}{b}\right)^b \quad (37)$$

in which $e \approx 2.718$ denotes the base of the natural logarithm, the probability $P(\|\mathbf{h}\|_2 > \lambda\epsilon|\mathbf{E})$ can be further bounded by

$$P(\|\mathbf{h}\|_2 > \lambda\epsilon|\mathbf{E}) \leq \frac{a}{b} \quad (38)$$

where

$$a \triangleq \left(\frac{n(m-\kappa+1)e^2}{(2K)^2}\right)^{2K} \left(\frac{me}{\kappa-1}\right)^{(\kappa-1)} \\ b \triangleq 2^{(m-\kappa+1-2K)}$$

Examine the condition which ensures that a/b is less than a specified value η , where $0 < \eta < 1$. Taking the base-2 logarithm on both sides of $(a/b) \leq \eta$ and slightly relaxing the inequality, we obtain

$$m > 2K \log(mn/K^2) + \kappa \log m + \log(1/\eta) + c \quad (39)$$

where

$$c \triangleq (\kappa-1)(\log(e/(\kappa-1)) + 1) + 4K \log(e/2) + 2K$$

is a constant only dependent on κ and K . In summary, for a specified η , if the condition (39) is satisfied, then we can ensure that the probability of the residual vector \mathbf{h} being greater than $\lambda\epsilon$ is smaller than η , i.e.

$$P(\|\mathbf{h}\|_2 > \lambda\epsilon|\mathbf{E}) \leq \eta \quad (40)$$

or

$$P(\|\mathbf{h}\|_2 \leq \lambda\epsilon|\mathbf{E}) \geq 1 - \eta \quad (41)$$

The proof is completed here.

5. Quantization design: adaptive methods

Previous analyses show that a reliable and accurate recovery of sparse signals is possible even from one-bit measurements. This is a very encouraging result. In this section, we aim to develop a practical algorithm for one-bit compressed sensing based on the above theoretical

analysis. Firstly, we note that the optimization (5) is a non-convex and NP hard problem that has computational complexity growing exponentially with the signal dimension n . To circumvent this issue, we replace the ℓ_0 -norm with the ℓ_1 -norm, the most popular alternative sparsity-promoting functional, which leads to the following optimization

$$\min_{\mathbf{z}} \|\mathbf{z}\|_1 \quad \text{s.t.} \quad \text{sign}(\mathbf{A}\mathbf{z} - \mathbf{r}) = \mathbf{b} \quad (42)$$

which is convex and can be recast as a linear programming problem that can be solved efficiently. Although a rigorous theoretical justification for Eq. (42) is still unavailable, our simulation results indeed suggest that Eq. (42) is an effective alternative to ℓ_0 -minimization and is able to yield a reliable and accurate reconstruction of sparse signals.

The other difficulty we face in developing a practical algorithm is that the suggested quantization thresholds are dependent on the original unquantized data samples \mathbf{y} which are inaccessible to the decoder where the reconstruction of the signal is conducted. This makes the choice of the quantization thresholds a tricky issue. On one hand, a pre-specified threshold vector can be used at the encoder and the decoder, but there is no guarantee that this common threshold will be close to the original unquantized measurements. On the other hand, if the encoder chooses thresholds close to the original measurements, it has to share the knowledge of the quantization thresholds with the decoder, in which case one-bit quantization becomes meaningless since the encoder still requires a high-resolution analog-to-digital converter (ADC) to quantize and transmit the threshold information to the decoder. To deal with this issue, we, in the following, propose a one-bit adaptive quantization (AQ) scheme where the decoder can compute the thresholds based on one-bit quantized data received from the encoder. The adaptive scheme is reminiscent of the Delta modulation with variable step-size.

5.1. Adaptive quantization

Specifically, the proposed AQ scheme involves an encoding and decoding process. At the encoder side, each measurement, say y_i , is encoded into a sequence of one-bit quantized data $\{b_i^{(t)}\}_{t=1}^q$ that are sequentially transmitted to the decoder. We, firstly, use two pre-specified parameters: a randomly generated initial threshold vector $\boldsymbol{\tau}^{(1)} = [\tau_1^{(1)} \ \tau_2^{(1)} \ \dots \ \tau_m^{(1)}]^T$, and an initial quantization step-size Δ , to generate the first two one-bit encoded data:

$$b_i^{(1)} = \text{sgn}(y_i - \tau_i^{(1)}) \\ b_i^{(2)} = \text{sgn}(y_i - \tau_i^{(2)}) \quad (43)$$

where $\text{sgn}\{x\} = -1$ if $x \leq 0$, otherwise $\text{sgn}\{x\} = 1$, $\tau_i^{(2)} = \tau_i^{(1)} + b_i^{(1)}\Delta$, $b_i^{(t)}$ denotes the encoded data associated with the i th element at iteration t , and $\tau_i^{(t)}$ denotes the quantization threshold used to yield $b_i^{(t)}$. At iteration $t \geq 2$, the encoder computes its threshold by performing accumulation of the previous bits, weighted by a variable step-size $\Delta_i^{(t)}$:

$$\tau_i^{(t+1)} = \tau_i^{(t)} + b_i^{(t)} \Delta_i^{(t)} \quad (44)$$

where $\Delta_i^{(t)}$ evolves using the following dynamic model

$$\Delta_i^{(t)} = \begin{cases} \Delta_i^{(t-1)}K_1 & \text{if } b_i^{(t)}b_i^{(t-1)} = 1 \\ \Delta_i^{(t-1)}/K_2 & \text{if } b_i^{(t)}b_i^{(t-1)} = -1 \end{cases} \quad (45)$$

where $K_1 > 1$ and $K_2 > 1$ are constants, and $\Delta_i^{(1)} = \Delta$. The encoder then uses $\tau_i^{(t+1)}$ as a threshold to generate its encoded data at iteration $t + 1$:

$$b_i^{(t+1)} = \text{sgn}(y_i - \tau_i^{(t+1)}) \quad (46)$$

We can immediately recognize that the above process is reminiscent of the Delta modulation with variable step-size. The key idea is to adjust the step-size based on two successive encoded bits. When successive encoded bits have identical signs, it means that we are still in the catch-up phase and the step-size is increased to speed up the process (c.f. Eq. (45)). On the other hand, alternating signs between successive bits indicate that the quantized data are oscillating around the waveform, in which case the step-size is decreased to provide a finer quantization. As shown in Fig. 1, the threshold sequence $\{\tau_i^{(t)}\}$ converges fast and eventually oscillates around the true measurement y_i with decreasing granular noise, which is exactly the property we desire.

Based on the one-bit quantized data $\{b_i^{(t)}\}_{t=1}^q$ received from the encoder, the decoder can recover the quantization thresholds $\{\tau_i^{(t)}\}_{t=1}^q$ that are used at the encoder. The reconstruction of $\tau_i^{(t)}$ can be easily inferred from the received encoded data in a recursive manner by using Eq.

(44)–(45). Finally, using the current and all previous thresholds and binary data $\{\tau^{(t)}, \mathbf{b}^{(t)}\}, \forall t = 1, \dots, q$, we compute an estimate of the sparse signal via solving the optimization Eq. (42) at the decoder. Note that multiple sets of thresholds and binary observations can be readily casted into the form Eq. (42) by replacing \mathbf{A} with a stacked measurement matrix $\tilde{\mathbf{A}} \triangleq [\mathbf{A}^T \dots \mathbf{A}^T]^T$. A schematic of the proposed adaptive quantization scheme is shown in Fig. 2. Note that throughout this iterative process, the unquantized measurements \mathbf{y} remain unchanged. For clarity, the one-bit adaptive quantization scheme is summarized as follows.

One-bit adaptive quantization scheme:

1. Input to the encoder and the decoder: An initial threshold vector $\tau^{(1)}$, an initial quantization step-size Δ , and the parameters K_1, K_2 that are used in the step-size update Eq. (45).
2. Encoding: Generate the one-bit encoded data $\{b_i^{(t)}\}$ and the quantization thresholds $\{\tau_i^{(t)}\}$ according to Eqs. (43)–(46).
3. Decoding: Given the received one-bit quantized data $\{b_i(t)\}$, recover the quantization thresholds $\{\tau_i(t)\}$ in a recursive manner by using Eqs. (43)–(45). Based on the received binary data and the reconstructed thresholds $\{\tau^{(t)}, \mathbf{b}^{(t)}\}, \forall t = 1, \dots, q$, compute an estimate of the sparse signal via solving the optimization Eq. (42).

We note that a similar AQ approach has been proposed and studied in the framework of distributed estimation for sensor networks [39], where sensors collaborate to estimate a common parameter or a random field through exchanging quantized data with their respective neighbors. Nevertheless, this paper presents the first attempt to apply AQ to the one-bit compressed sensing framework.

5.2. Discussions

We see that for the AQ scheme, the encoder only requires a simple comparator to quantize the real-valued measurement, and the computation of the thresholds at both the encoder and the decoder involves very simple algebraic operations (addition, multiplication, division) that can be easily implemented using analog addition/multiplier circuits. One-bit quantization enables significant reduction of the hardware complexity. As indicated in [36], one-bit quantizer which takes the form of a simple comparator is particularly appealing in hardware implementations, and can operate at a much higher sampling

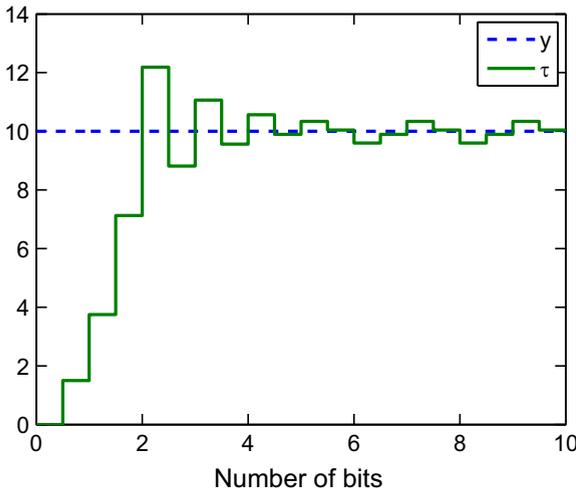


Fig. 1. An example of the adaptive quantization process.

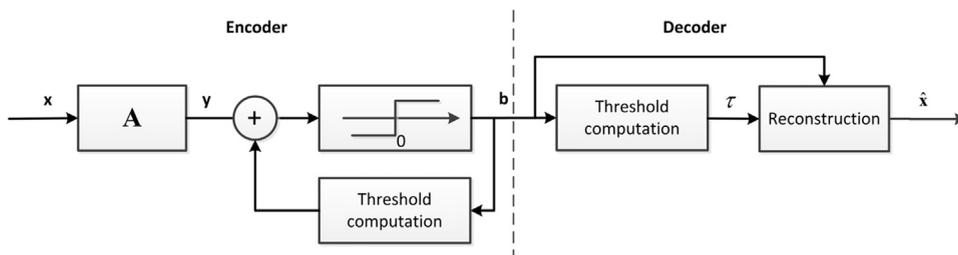


Fig. 2. Schematic of one-bit adaptive quantization for compressed sensing.

rate than the high-resolution quantizer. Besides, one-bit measurements are more amiable for large-scale parallel processing than high-resolution data. With these merits, the proposed adaptive architecture allows us to develop data acquisition devices with lower-cost and faster speed, meanwhile achieving reconstruction performance similar to that of using multiple-bit quantizer.

We discuss the choice of the pre-specified parameters Δ , K_1 and K_2 that are used at the encoder and the decoder. To speed up the catch-up process, the initial quantization step-size Δ should be set a value that is of the same order of the mean magnitude of \mathbf{y} . A small $K_1 \in (1, 2]$ is preferable in most cases, and generally K_2 is chosen equal to or slightly greater than K_1^2 to ensure that the step-size will decrease rapidly after the threshold oscillates around the waveform. To introduce randomness to the difference vector $\delta = \mathbf{y} - \tau$, we randomly generate the initial threshold $\tau^{(1)}$.

Note that the adaptive process can roughly be divided into a transient catch-up phase and a stable oscillating phase. For the transient catch-up phase, the entries of the difference vector δ may not have equal probabilities of being positive or negative because the quantization threshold τ_i is always larger or smaller than the measurement y_i during the catch-up phase. Nevertheless, when the adaptive process reaches the oscillating phase, the quantization threshold τ_i oscillates around the measurement y_i . Thus the entries of the difference vector will generally have equal probabilities of being positive or negative. Moreover, entries of the difference vector are mutually independent because the evolution of each threshold sequence $\{\tau_i^{(t)}\}_{t=1}^q$ is independent of other sequences $\{\tau_j^{(t)}\}_{t=1}^q, \forall j \neq i$. Therefore the independence and equal-probability assumption for entries of the difference vector generally holds for the oscillating phase. To improve the recovery performance, we can choose adaptive quantization parameters to speed up the catch-up phase so that the process is dominated by the oscillating phase.

We notice that an adaptive quantization architecture was also proposed in [28,31]. Nevertheless, the architecture in [28,31] requires a feedback of the estimate of the sparse signal from the decoder to the encoder for the threshold adaptation, which may not be feasible in practice. Our scheme, in contrast, does not involve any feedback from the decoder. Also, the scheme [28] updates the thresholds in a sequential manner, which is computationally expensive since the adaptive scheme requires to compute an estimate of the sparse signal every iteration.

6. Numerical results

We now carry out experiments to corroborate our previous analysis and to illustrate the performance of the proposed adaptive quantization scheme. In our simulations, the K -sparse signal is randomly generated with the support set of the sparse signal randomly chosen according to a uniform distribution. The signals on the support set are independent and identically distributed (i.i.d.) Gaussian random variables with zero mean and unit variance. The measurement matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$ is randomly

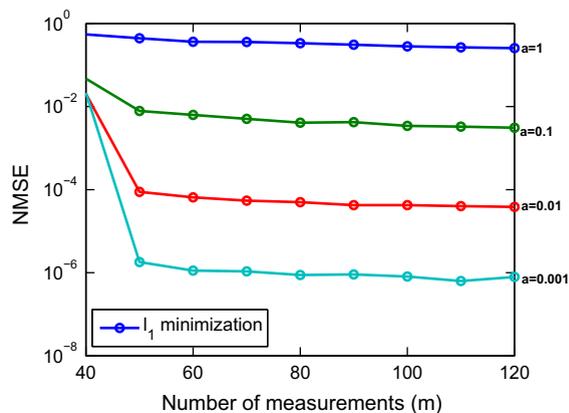


Fig. 3. Reconstruction mean-squared error versus the number of measurements for different choices of a .

generated with each entry independently drawn from Gaussian distribution with zero mean and unit variance.

6.1. Performance under different threshold choices

We first examine the impact of the quantization design on the reconstruction performance. The knowledge of the original unquantized measurements \mathbf{y} is assumed available in order to validate our theoretical results. The thresholds are chosen to be the sum of the unquantized measurements \mathbf{y} and a deviation term δ , i.e. $\tau = \mathbf{y} + \delta$, where δ is a vector with its entries being independent discrete random variables with $P(\delta_i = -a) = 0.5$ and $P(\delta_i = a) = 0.5$, in which the parameter $a > 0$ controls the deviation of τ from \mathbf{y} . Fig. 3 depicts the reconstruction normalized mean squared error (NMSE), $E[\|\mathbf{x} - \hat{\mathbf{x}}\|^2 / \|\mathbf{x}\|^2]$, vs. the number of measurements m for different choices of a , where we set $n=50$, and $K=3$. Results are averaged over 10^4 independent runs. From Fig. 3, we see that the reconstruction accuracy can be significantly improved by reducing the deviation parameter a . In particular, a NMSE as small as 10^{-6} can be achieved when a is set 0.001. This corroborates our theoretical analysis that sparse signals can be recovered from one-bit quantized data with an arbitrarily small error by letting $\delta \rightarrow \mathbf{0}$. Also, as expected, the reconstruction error decreases with an increasing number of measurements m . Nevertheless, the performance improvement due to an increasing m is mild when m is large. This fact suggests that the choice of quantization thresholds is a more critical factor than the number of measurements in achieving an accurate reconstruction. In Fig. 4, we plot the root mean squared error (RMSE), $E[\|\mathbf{x} - \hat{\mathbf{x}}\|_2]$, as a function of the deviation magnitude $\epsilon = \|\delta\|_2 = \sqrt{ma}$, where we set $m=100$, $n=120$, $K=2$ and a varies from 10^{-3} to 1. It can be observed that the RMSE decreases proportionally with the value ϵ , which coincides with our theoretical analysis Eq. (13). To further corroborate our analysis, we consider a different way to generate the deviation vector δ , with its entries randomly generated according to a Gaussian distribution with zero mean and variance σ^2 . Fig. 5 depicts the NMSE vs. the number of measurements m for different values of σ , where $n=50$, and $K=3$. Again, we observe that a more accurate estimate

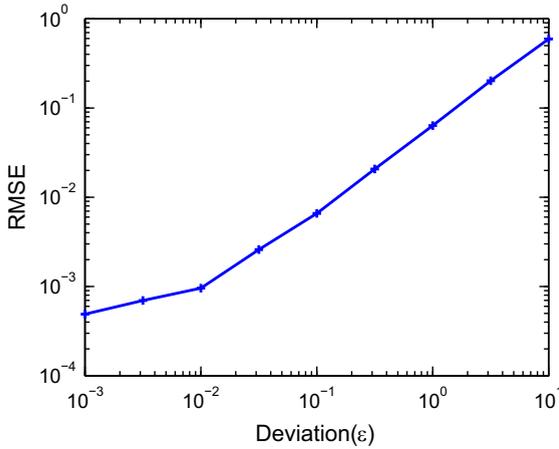


Fig. 4. Root mean-squared error versus ϵ .

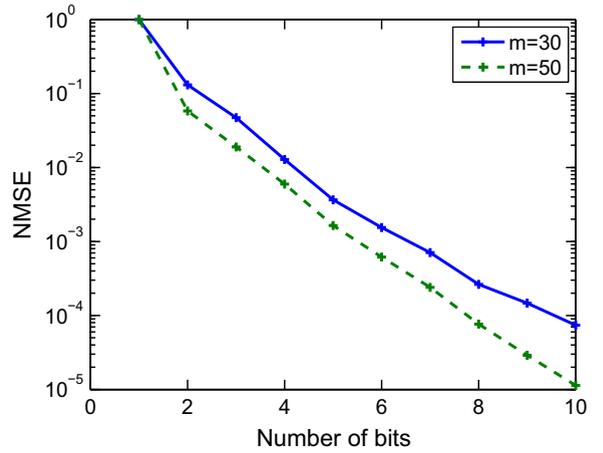


Fig. 6. Normalized mean-squared error versus number of iterations.

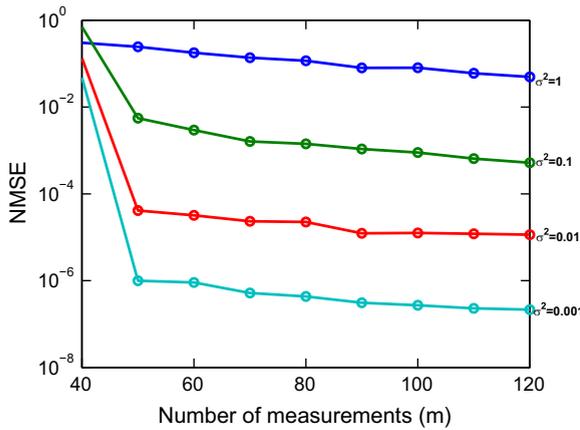


Fig. 5. Reconstruction mean-squared error versus the number of measurements for different choices of σ .

is achieved when the thresholds get closer to the unquantized measurements \mathbf{y} .

6.2. Performance of adaptive quantization scheme

We now carry out experiments to illustrate the performance of the proposed adaptive quantization (AQ) algorithm. In our experiments, we set $n=50$, and $K=2$. The NMSE vs. the number of iterations (i.e. number of bits, q , per measurement) is plotted in Fig. 6, where we set m to 30, and 50, respectively. Results are averaged over 10^3 independent runs, with the sampling matrix and the sparse signal randomly generated for each run. From Fig. 6, we see that the AQ algorithm provides a consistent performance improvement through iteratively refining the quantization thresholds, and usually provides a reasonable reconstruction performance within only a few iterations.

To further show the effectiveness of the AQ scheme, we compare with a uniform quantization (UQ) scheme which employs a q -bit uniform quantizer at the encoder. A uniform q -bit quantizer uniformly divide the observation dynamic range $[-\eta, \eta]$ into intervals of length $\Delta = 2\eta / (2^q - 1)$ and round the message y_i to the nearest

neighboring endpoints of these intervals. The quantized data $\hat{\mathbf{y}} \triangleq [\hat{y}_1 \dots \hat{y}_m]^T$ are then transmitted to the decoder, where multi-bit quantized compressed sensing techniques [15,17] can be employed to reconstruct the sparse signal. Specifically, in [17], the reconstruction problem was casted as the following convex optimization with a quantization consistency constraint

$$\min_{\mathbf{z}} \|\mathbf{z}\|_1 \quad \text{s.t.} \quad \mathbf{Az} \in \mathcal{R}_{\hat{\mathbf{y}}} \quad (47)$$

where $\mathcal{R}_{\hat{\mathbf{y}}}$ denotes the quantization region of $\hat{\mathbf{y}}$. This reconstruction scheme is referred to as compressed sensing with quantization consistency (CS-QC) method in our paper. Another class of decoders called as Basis Pursuit DeQuantizer of moment p (BPDQ) ($\infty \geq p \geq 2$) [15] formulate the recovery problem as

$$\min_{\mathbf{z}} \|\mathbf{z}\|_1 \quad \text{s.t.} \quad \|\hat{\mathbf{y}} - \mathbf{Az}\|_p \leq \epsilon \quad (48)$$

where we choose $p=2$ in our simulations, in which case BPDQ becomes the BPDN method. Note that when $p = \infty$ and $\epsilon = \Delta/2$, the two formulations Eqs. (47) and (48) become identical. We also consider a one-bit random quantization scheme (referred to as the RQ scheme) whose thresholds are randomly generated according to a uniform distribution over a range $[-\eta, \eta]$. To make a fair comparison, the RQ scheme samples each measurement q times using uniformly generated q thresholds, and yields q bits of information which are reported to the decoder. For the RQ scheme, suppose the knowledge of the thresholds are also available at the decoder, the sparse recovery problem can also be formulated as Eq. (42) and efficiently solved. Figs. 7 and 8 depict the NMSEs of respective schemes as a function of the ratio m/n for different choices of q , where the measurement matrix is chosen to be a random matrix in Fig. 7 and a deterministic discrete cosine transform (DCT) matrix in Fig. 8. We set $\Delta = 0.2$, $K_1 = 1.3$, and $K_2 = 1.5$, for the AQ scheme, and $\eta = 2$ (such a choice guarantees that the measurements \mathbf{y} fall into the region $[-\eta, \eta]$ with a high probability) for the CS-QC, BPDQ and RQ schemes. Note that all these schemes send the same number of bits, i.e. mq bits, to the decoder. From Figs. 7 and 8, we see that our proposed AQ scheme presents a significant

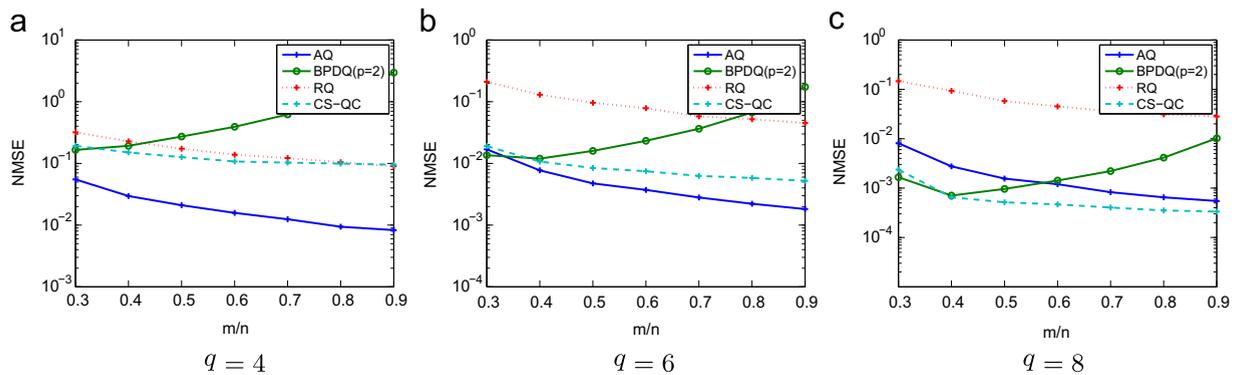


Fig. 7. (a) $q=4$, (b) $q=6$ (c), $q=8$. NMSEs of respective schemes vs. the ratio m/n with randomly generated measurement matrices.

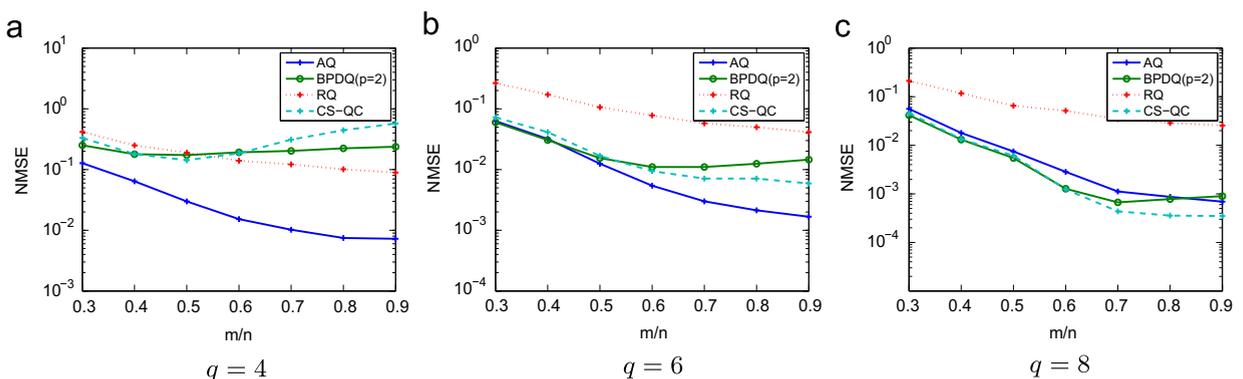


Fig. 8. (a) $q=4$, (b) $q=6$, (c) $q=8$. NMSEs of respective schemes vs. the ratio m/n with a deterministic measurement matrix.

performance advantage over the RQ scheme. This is not surprising since iteratively refined thresholds are generally closer to the original measurements \mathbf{y} than randomly generated thresholds. We also observe that the AQ scheme outperforms the CS-QC and BPDQ methods by a big margin when q is small. The AQ scheme is surpassed by the CS-QC method as q increases. This performance advantage, however, is achieved with the cost of employing a high-resolution ADC at the encoder.

7. Conclusion

We studied the problem of quantization design in a general framework of one-bit compressed sensing with non-zero quantization thresholds. Our theoretical analysis reveals that when the number of one-bit measurements is sufficiently large, with a high probability the sparse signal can be recovered with a bounded error. The error bound is linearly proportional to the ℓ_2 -norm of the difference between the thresholds and the original unquantized measurements. By setting the thresholds sufficiently close to the original unquantized measurements, sparse signals can be recovered with an arbitrarily small error. The unquantized measurements, unfortunately, are inaccessible to the decoder. To address this issue, we proposed an adaptive quantization (AQ) scheme where the encoder iteratively adjusts the quantization thresholds based on

previous encoded bits such that the quantization thresholds eventually oscillate around the true measurements. Based on the one-bit quantized data received from the encoder, the decoder can calculate the quantization thresholds that are used at the encoder, and compute an estimate of the sparse signal based on the thresholds and the binary data. Simulation results showed that the proposed AQ scheme presents superiority over other one-bit compressed sensing schemes and achieves recovery performance similar to that of using a multi-bit quantizer.

References

- [1] E. Candès, T. Tao, Decoding by linear programming, *IEEE Trans. Inf. Theory* 12 (December) (2005) 4203–4215.
- [2] D.L. Donoho, Compressive sensing, *IEEE Trans. Inf. Theory* 52 (2006) 1289–1306.
- [3] J.A. Tropp, A.C. Gilbert, Signal recovery from random measurements via orthogonal matching pursuit, *IEEE Trans. Inf. Theory* 53 (December (12)) (2007) 4655–4666.
- [4] M.J. Wainwright, Information-theoretic limits on sparsity recovery in the high-dimensional and noisy setting, *IEEE Trans. Inf. Theory* 55 (December (12)) (2009) 5728–5741.
- [5] S. Ji, Y. Xue, L. Carin, Bayesian compressive sensing, *IEEE Trans. Signal Process.* 56 (June (6)) (2008) 2346–2356.
- [6] D.P. Wipf, B.D. Rao, An empirical Bayesian strategy for solving the simultaneous sparse approximation problem, *IEEE Trans. Signal Process.* 55 (July (7)) (2007) 3704–3716.
- [7] Z. Zhang, B.D. Rao, Sparse signal recovery with temporally correlated source vectors using sparse Bayesian learning, *IEEE J. Sel. Topics Signal Process.* 5 (September (5)) (2011) 912–926.

- [8] R. Chartrand, W. Yin, Iterative reweighted algorithm for compressive sensing, in: IEEE International Conference on Acoustics, Speech, and Signal Processing, Las Vegas, Nevada, USA, 2008.
- [9] W. Dai, O. Milenkovic, Subspace pursuit for compressive sensing signal reconstruction, *IEEE Trans. Inf. Theory* 5 (May) (2009) 2230–2249.
- [10] J. Huang, T. Zhang, D. Metaxas, Learning with structured sparsity, *J. Mach. Learn. Res.* 12 (2011) 3371–3412.
- [11] L. Hu, Z. Shi, J. Zhou, Q. Fu, Compressed sensing of complex sinusoids: an approach based on dictionary refinement, *IEEE Trans. Signal Process.* 60 (7) (2012) 3809–3822.
- [12] Z. Yang, L. Xie, C. Zhang, Off-grid direction of arrival estimation using sparse Bayesian inference, *IEEE Trans. Signal Process.* 61 (January (1)) (2013) 38–42.
- [13] I.F. Akyildiz, W. Su, Y. Sankarasubramaniam, E. Cayirci, A survey on sensor networks, *IEEE Commun. Mag.* (August) (2002) 102–114.
- [14] S.S. Chen, D.L. Donoho, M.A. Saunders, Atomic decomposition by basis pursuit, *SIAM J. Sci. Comput.* 20 (1) (1998) 33–61.
- [15] L. Jacques, D.K. Hammond, J.M. Fadili, Dequantizing compressed sensing: when oversampling and non-Gaussian constraints combine, *IEEE Trans. Inf. Theory* 57 (January (1)) (2011) 559–571.
- [16] A. Zymnis, S. Boyd, E. Candès, Compressed sensing with quantized measurements, *IEEE Signal Process. Lett.* 17 (February (2)) (2010) 149–152.
- [17] W. Dai, O. Milenkovic, Information theoretical and algorithmic approaches to quantized compressive sensing, *IEEE Trans. Commun.* 59 (July (7)) (2011) 1857–1866.
- [18] J. Laska, P. Boufounos, M. Davenport, R. Baraniuk, Democracy in action: quantization, saturation, and compressive sensing, *Appl. Comput. Harmon. Anal.* 31 (November (3)) (2011) 429–443.
- [19] T. Wimalajeewa, P.K. Varshney, Performance bounds for sparsity pattern recovery with quantized noisy random projections, *IEEE J. Sel. Topics Signal Process.* 6 (February (1)) (2012) 43–57.
- [20] L. Jacques, Error decay of (almost) consistent signal estimations from quantized Gaussian random projections, Available: [arXiv:1406.0022](https://arxiv.org/abs/1406.0022), 2014.
- [21] P.T. Boufounos, R.G. Baraniuk, One-bit compressive sensing, in: Proceedings of the 42nd Annual Conference on Information Sciences and Systems, Princeton, NJ, 2008.
- [22] L. Jacques, J.N. Laska, P.T. Boufounos, R.G. Baraniuk, Robust 1-bit compressive sensing via binary stable embeddings of sparse vectors, *IEEE Trans. Inf. Theory* 59 (April (4)) (2013) 2082–2102.
- [23] Y. Plan, R. Vershynin, One-bit compressed sensing by linear programming, *Commun. Pure Appl. Math.* 66 (August (8)) (2013) 1275–1297.
- [24] J.N. Laska, R.G. Baraniuk, Regime change: bit-depth versus measurement-rate in compressive sensing, *IEEE Trans. Signal Process.* 60 (July (7)) (2012) 3496–3505.
- [25] M. Yan, Y. Yang, S. Osher, Robust 1-bit compressive sensing using adaptive outlier pursuit, *IEEE Trans. Signal Process.* 60 (July (7)) (2012) 3868–3875.
- [26] A. Movahed, A. Panahi, G. Durisi, A robust RFPI-based 1-bit compressive sensing reconstruction algorithm, in: IEEE Information Theory Workshop (ITW), Lausanne, September 3–7, 2012.
- [27] F. Li, J. Fang, H. Li, L. Huang, Robust one-bit Bayesian compressed sensing with sign-flip errors, *IEEE Signal Process. Lett.* 22 (July (7)) (2015) 857–861.
- [28] U.S. Kamilov, A. Bourguard, A. Amini, M. Unser, One-bit measurements with adaptive thresholds, *IEEE Signal Process. Lett.* 19 (October (10)) (2012) 607–610.
- [29] Y. Shen, J. Fang, H. Li, Z. Chen, A one-bit reweighted iterative algorithm for sparse signal recovery, in: IEEE International Conference on Acoustics, Speech, and Signal Processing, Vancouver, Canada, May 26–31, 2013.
- [30] K. Knudson, R. Saab, R. Ward, One-bit compressive sensing with norm estimation, Available: [arXiv:1404.6853](https://arxiv.org/abs/1404.6853), 2014.
- [31] R. Baraniuk, S. Foucart, D. Needell, Y. Plan, M. Wootters, Exponential decay of reconstruction error from binary measurements of sparse signals, Available: [arXiv:1407.8246v1](https://arxiv.org/abs/1407.8246v1), 2014.
- [32] P.T. Boufounos, L. Jacques, F. Krahmer, R. Saab, Quantization and compressive sensing, Available: [arXiv:1405.1194](https://arxiv.org/abs/1405.1194), 2014.
- [33] C.S. Gunturk, M. Lammers, A.M. Powell, R. Saab, O. Yilmaz, Sigma-Delta quantization for compressed sensing, in: Proceedings of the 44th Annual Conference on Information Sciences and Systems (CISS), 2010.
- [34] J.Z. Sun, V.K. Goyal, Optimal quantization of random measurements in compressed sensing, in: Proceedings of IEEE International Symposium on Information Theory, Seoul, Korea, June 28–July 3, 2009.
- [35] U.S. Kamilov, V.K. Goyal, S. Rangan, Optimal quantization for compressed sensing under message passing reconstruction, in: Proceedings of IEEE International Symposium on Information Theory, Saint Petersburg, Russia, July 31–August 5, 2011.
- [36] A. Bourguard, F. Aguet, M. Unser, Optical imaging using binary sensors, *Optics Express* 18 (March (5)) (2010) 4876–4888.
- [37] M. Rudelson, R. Vershynin, Smallest singular value of a random rectangular matrix, *Commun. Pure Appl. Math.* 62 (12) (2009) 1707–1739.
- [38] M.A. Davenport, J.N. Laska, P.T. Boufounos, R.G. Baraniuk, A simple proof that random matrices are democratic, Available: [arXiv:0911.0736](https://arxiv.org/abs/0911.0736), 2009.
- [39] J. Fang, H. Li, Distributed estimation of Gauss–Markov random fields with one-bit quantized data, *IEEE Signal Process. Lett.* 17 (May) (2010) 449–452.