1-bit Compressed Quantization

Tianyi Zhou
QCIS, University of Technology, Sydney
NSW 2007, Australia
tianyi.zhou@student.uts.edu.au

Dacheng Tao
QCIS, University of Technology, Sydney
NSW 2007, Australia
dacheng.tao@uts.edu.au

Abstract

Compressed sensing (CS) and 1-bit CS cannot directly recover quantized signals preferred in digital systems and require time consuming recovery. In this paper, we introduce *1-bit compressed quantization* (1-bit CQ) that directly recovers the k-bit quantization of a signal of dimensional n from its 1-bit measurements via invoking n times of nearest neighbor search. Compared to CS and 1-bit CS, 1-bit CQ allows the signal to be dense, takes considerably less (linear) recovery time and requires substantially less measurements ($\mathcal{O}(\log n)$), with the cost of quantization error. Extensive numerical simulations verify the appealing accuracy, robustness and efficiency of 1-bit CQ.

1 Introduction

Recent results in compressed sensing (CS) [1][2] prove that a sparse or compressible signal can be exactly recovered from its linear measurements, rather than uniform samplings, at a rate significantly lower than the Nyquist rate. The measurement matrix is required to have the restricted isometry property (RIP) [3][4] for the purpose of ensuring the exact reconstruction via an ℓ_p ($0 \leq p \leq 2$) penalized/constrained minimization of the measurement error.

However, CS [5] encounters several problems when applied to practical digital systems, where analog-to-digital converters (ADCs) not only sample, but also quantize each measurement to a finite number of bits. One key problem is that CS cannot explicitly handle the quantized measurements. Thus 1-bit CS [6][7] is developed to reconstruct sparse signals from 1-bit measurements, which capture the signs of the CS measurements. The 1-bit measurements significantly reduce the costs and strengthen the robustness of hardware implementation. Although the 1-bit measurements lead to the loss of scale information, 1-bit CS ensures consistent reconstructions of signals on the unit ℓ_2 sphere [8].

Another important problem is that digital systems prefer to use the quantized recovery of the original signal, which they can directly process, but the recoveries of both CS and 1-bit CS are continuous. In order to apply them to digital systems, additional quantization is required. Moreover, the time consuming optimization based and iterative recovery in CS and 1-bit CS limits their applications in practical systems, especially when signals are of high-dimension. In addition, CS or 1-bit CS achieves exact recovery under the Nyquist rate due to replacement of the previous uniform sampling with random linear measurements or their signs. However, this reduction of sampling rate signal relies on the sparsity of the signal. Quantization is an irreversible description of the original signal and introduces quantization error. This information loss implies the possibility to recover the quantization of a dense signal from a small number of measurements.

The primary contribution of this paper is developing *1-bit compressed quantization* (1-bit CQ) to recover the quantized signal from its quantized measurements with extremely small time cost and without signal sparsity constraint. In compression, we adopt the 1-bit measurements [7] as in 1-bit CS. In particular, we introduce a bijection between each dimension of the signal and a Bernoulli

distribution. The underlying idea of 1-bit CQ is to estimate the Bernoulli distribution for each dimension from the 1-bit measurements, and thus each dimension of the signal can be recovered from the corresponding Bernoulli distribution. In recovery, we propose a k-bit quantizer for the signal domain, whose intervals are the mappings of the uniform linear quantization boundaries for the Bernoulli distribution domain. 1-bit CQ searches the nearest neighbor of the estimated Bernoulli distribution among the boundaries and recovers the quantization of the corresponding dimension as the quantizer interval associated with the nearest neighbor. The main significance of 1-bit CQ is as follows: 1) it provides a direct and simple recovery of quantized signal for digital systems; 2) it only requires to compute nk pairwise distances for obtaining k-bit recovery of an n-dimensional signal, and is therefore considerably more efficient than CS and 1-bit CS; 3) successful recovery can be obtained from only $\mathcal{O}(\log n)$ measurements. Thus 1-bit CQ can be applied to general signals without sparse assumption.

2 1-bit Measurements

1-bit CQ recovers the quantized signal directly from its quantized measurements. We consider the extreme case of 1-bit measurements of a signal $x \in \mathbb{R}^n$, which are given by

$$y = A(x) = \operatorname{sign}(\Phi x), \tag{1}$$

where $\operatorname{sign}(\cdot)$ is an element-wise sign operator and $A(\cdot)$ maps x from $\mathbb{R}^n$ to the Boolean cube $\mathbb{B}^M := \{-1, 1\}^M$. Since the scale of the signal is lost in 1-bit measurements y (multiplying x with a positive scalar will not change the signs of the measurements), the consistent reconstruction can be obtained by enforcing the signal $x \in \Sigma_K^* := \{x \in S^{n-1} : \|x\|_0 \leq K\}$ where $S^{n-1} := \{x \in \mathbb{R}^n : \|x\|_2 = 1\}$ is the n-dimensional unit hyper-sphere.

2.1 Bijection

In contrast to CS and 1-bit CS, 1-bit CQ does not recover the original signal, but reconstructs the quantized signal by recovering each dimension in isolation. In particular, according to Lemma 3.2 in [9], we show that there exists a bijection (cf. Theorem 1) between each dimension of the signal x and a Bernoulli distribution, which can be uniquely estimated from the 1-bit measurements. The underlying idea of 1-bit CQ is to recover the quantization of the corresponding dimension as the interval where the estimated Bernoulli distribution's mapping lies in.

Theorem 1. (**Bijection**) *For a normalized signal $x \in \mathbb{R}^n$ with $\|x\|_2 = 1$ and a normalized Gaussian random vector ϕ that is drawn uniformly from the unit ℓ_2 sphere in $\mathbb{R}^n$ (i.e., each element of ϕ is firstly drawn i.i.d. from the standard Gaussian distribution $\mathcal{N}(0, 1)$ and then ϕ is normalized as $\phi/\|\phi\|_2$), given the i^{th} dimension of the signal x_i and the corresponding coordinate unit vector $e_i = \{0, \cdots, 0, 1, 0, \cdots, 0\}$, where 1 appears in the i^{th} dimension, there exists a bijection $P : \mathbb{R} \to \mathbb{P}$ from x_i to the Bernoulli distribution of the binary random variable $s_i = \operatorname{sign}(\langle x, \phi\rangle) \cdot \operatorname{sign}(\langle e_i, \phi\rangle)$:*

$$P(x_i) = \begin{cases} \Pr(s_i = -1) = \frac{1}{\pi} \arccos(x_i), \\ \Pr(s_i = 1) = 1 - \frac{1}{\pi} \arccos(x_i). \end{cases} \tag{2}$$

Since the mapping between x_i and $P(x_i)$ is bijective, given $P(x_i)$, the i^{th} dimension of x can be uniquely identified. According to the definition of s_i, $P(x_i)$ can be estimated from the instances of the random variable $\operatorname{sign}(\langle x, \phi\rangle)$, which are exactly the 1-bit measurements y defined in (1). Therefore, the 1-bit measurements y include sufficient information to reconstruct x_i from the estimation of $P(x_i)$, and the recovery accuracy of x_i depends on the accuracy of the estimation to $P(x_i)$.

3 k-bit Reconstruction

The primary contribution of this paper is the quantized recovery in 1-bit CQ, which reconstructs the quantized signal from its 1-bit measurements (1). Figure 1 illustrates 1-bit CQ quantized recovery. To define the k-bit quantizer used in 1-bit CQ, we firstly find k boundaries $P_j (j = 0, \cdots, k-1)$ (4) in Bernoulli distribution domain by imposing the uniform linear quantizer to the range of P_j^-. Given an arbitrary x_i, the nearest neighbor of $P(x_i)$ among the k boundaries $P_j (j = 0, \cdots, k-1)$

indicates the interval q_i that x_i lies in the signal domain. The $k+1$ boundaries $S_j(j=0,\cdots,k)$ associated with the k intervals $q_j(j=0,\cdots,k)$ are calculated from the k boundaries $P_j(j=0,\cdots,k-1)$ according to the bijection defined in Theorem 1. In 1-bit CQ recovery, $P(x_i)$ is estimated as $\hat{P}(x_i)$ from the 1-bit measurements y. Then the nearest neighbor of $\hat{P}(x_i)$ among the k boundaries $P_j(j=0,\cdots,k-1)$ is determined by comparing the ℓ_1 distances between $\hat{P}(x_i)^-$ and P_j^-. The quantization of x_i is recovered as the interval q_i corresponding to the nearest neighbor. We study the upper bound of the quantized recovery error err_H.

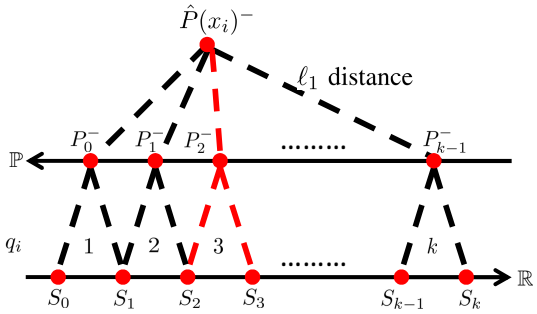

Figure 1: **Quantized recovery in 1-bit CQ**. $P(x_i)$ in Theorem 1 has estimate $\hat{P}(x_i)$ (8) from $y=A(x)$. 1-bit CQ searches the nearest neighbor of $\hat{P}(x_i)^-$ among the k boundaries $P_j^-(j=0,\cdots,k-1)$ (4). The quantization of x_i, i.e., q_i is recovered as the interval between the two boundaries S_{i-1} and S_i (6) corresponding to the nearest neighbor.

3.1 1-bit CQ quantizer

We introduce the 1-bit CQ quantizer $Q(\cdot)$ by defining a bijective mapping from the boundaries of the Bernoulli distribution domain to the intervals of the signal domain according to Theorem 1. Assume the range of a signal x is given by:

$$-1 \leq x_{inf} \leq x_i \leq x_{sup} \leq 1, \forall i, \cdots, n. \tag{3}$$

By applying the uniform linear quantizer with the quantization interval Δ to the Bernoulli distribution domain, we get the corresponding boundaries

$$P_i = \begin{cases} P_i^- = \Pr(-1) = \frac{1}{\pi}\arccos(x_{inf}) - i\Delta, \\ P_i^+ = \Pr(1) = 1 - \Pr(-1). \end{cases}, \quad i = 0, \cdots, k-1. \tag{4}$$

The interval Δ is

$$\Delta = \frac{1}{k-1}\left(\frac{1}{\pi}\arccos(x_{inf}) - \frac{1}{\pi}\arccos(x_{sup})\right). \tag{5}$$

We define the 1-bit CQ quantizer in the signal domain by computing its $k+1$ boundaries as a mapping from the k boundaries $P_i(i=0,\cdots,k-1)$ to $\mathbb{R}$ in the Bernoulli domain:

$$S_i = \begin{cases} x_{inf}, & i=0; \\ \cos\left(\pi\left(P_i^- + \frac{\Delta}{2}\right)\right), & i=1,\cdots,k-1; \\ x_{sup}, & i=k. \end{cases} \tag{6}$$

According to (6), 1-bit CQ quantizer performs closely to the uniform linear quantizer when x_i is not very close to -1 or 1.

Given a signal x and the boundaries defined in (6), its k-bit quantization q is:

$$Q(x) = q, q_i = \{j : S_{j-1} \leq x_i \leq S_j\}. \tag{7}$$

3.2 ℓ_1 nearest neighbor search

The $k+1$ boundaries of the 1-bit CQ quantizer in (6) define k intervals in $\mathbb{R}$. Quantized recovery in 1-bit CQ reconstructs a quantized signal by estimating which interval each dimension of the signal x

lies in. The estimation is obtained by a nearest neighbor search in the Bernoulli distribution domain. To be specific, an estimation of $P(x_i)^-$ given in (2) can be derived from the 1-bit measurements y. For each $P(x_i)^-$, we find its nearest neighbor among the k boundaries $P_j^-(j = 0, \cdots, k-1)$ (4) in the Bernoulli distribution domain. The interval that x_i lies in is then estimated as the quantizer's interval corresponding to the nearest neighbor.

According to Theorem 1, the bijection from x_i to a particular Bernoulli distribution, i.e., $P(x_i)$ given in (2), has an unbiased estimation from the 1-bit measurements y

$$\hat{P}(x_i) = \begin{cases} \hat{P}(x_i)^- & = \left| j : [y \cdot \operatorname{sign}(\Phi_i')]_j = -1 \right| / m, \\ \hat{P}(x_i)^+ & = 1 - \hat{P}(x_i)^-, \end{cases} \tag{8}$$

where Φ_i is the i^{th} column of the measurement matrix Φ.

The quantization of x_i can then be recovered by searching the nearest neighbor of $\hat{P}(x_i)^-$ among the k boundaries $P_j^-(j = 0, \cdots, k-1)$ in (4). Specifically, the interval that x_i lies in among the k intervals defined by the boundaries $S_j(j = 0, \cdots, k)$ in (6) is identified as the one whose corresponding boundary P_j^- is the nearest neighbor of $\hat{P}(x_i)^-$. In this paper, the distance between P_j^- and $\hat{P}(x_i)^-$ is measured by ℓ_1 distance. Therefore, the quantized recovery of x, i.e., q^*, is given by

$$R(y) = q^*, q_i^* = 1 + \arg\min_j \left\| P_j^- - \hat{P}(x_i)^- \right\|_1, \\ \forall i = 1, \cdots, n, \forall j = 0, \cdots, k-1. \tag{9}$$

Thus the interval that x_i lies in can be recovered as

$$S_{q_i^*-1} \leq x_i \leq S_{q_i^*}. \tag{10}$$

The 1-bit CQ recovery algorithm is fully summarized in (9), which only includes simple computations without iteration and thus can be easily implemented in real systems. According to (9), the quantized recovery in 1-bit CQ requires nk computations of absolute values. This indicates the high efficiency of 1-bit CQ (linear recovery time), and the trade-off between resolution (k) and time cost (nk).

Theorem 2. (**Amount of measurements**) *HCS successfully reconstructs the signal x with probability exceeding $1 - \eta$ if the number of measurements $m \geq C \log \frac{n}{2\eta}$, wherein C is a constant.*

References

[1] David L. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[2] Emmanuel J. Candès and Terence Tao, "Near-optimal signal recovery from random projections: Universal encoding strategies?," *IEEE Transactions on Information Theory*, vol. 52, no. 12, pp. 5406–5425, 2006.

[3] Emmanuel J. Candès, Justin K. Romberg, and Terence Tao, "Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information," *IEEE Transactions on Information Theory*, vol. 52, no. 2, pp. 489–509, 2006.

[4] Emmanuel J. Candès, Justin K. Romberg, and Terence Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Communications on Pure and Applied Mathematics*, vol. 59, no. 8, pp. 1207–1223, 2006.

[5] Alfred M. Bruckstein, David L. Donoho, and Michael Elad, "From sparse solutions of systems of equations to sparse modeling of signals and images," *SIAM Review*, vol. 51, no. 1, pp. 34–81, 2009.

[6] Petros T. Boufounos and Richard G. Baraniuk, "One-bit compressive sensing," in *Conference on Information Sciences and Systems (CISS)*, 2008.

[7] Laurent Jacques, Jason N. Laska, Petros T. Boufounos, and Richard G. Baraniuk, "Robust 1-bit compressive sensing via binary stable embeddings of sparse vectors," *arXiv:1104.3160*, 2011.

[8] Petros T. Boufounos, "Greedy sparse signal reconstruction from sign measurements," in *Proc. Asilomar Conference on Signals Systems and Computers*, 2009.

[9] Michel X. Goemans and David P. Williamson, "Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming," *Journal of the ACM*, vol. 42, no. 6, pp. 1115–1145, 1995.

Appendix: Numerical Results

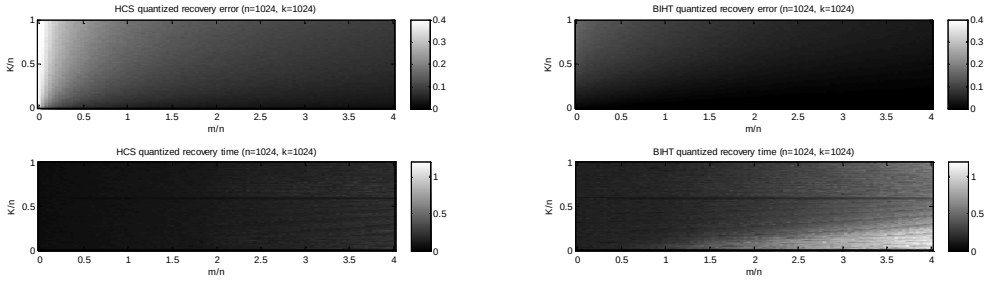

Figure 2: Phase plots of 1-bit CQ and "1-bit CS+1-bit CQ quantizer" in the noiseless case.

This section evaluates 1-bit CQ and compares it with BIHT [7] for 1-bit CS on two groups of numerical experiments. We use average quantized recovery error $\sum_{i=1}^{n} |q_i - q_i^*| / nk$ to measure the quantization error err_H. In each trial, we draw a normalized Gaussian random matrix $\Phi \in \mathbb{R}^{m \times n}$ given in Theorem 1 and a signal of length n and cardinality K, whose K nonzero entries drawn uniformly at random on the unit ℓ_2 sphere.

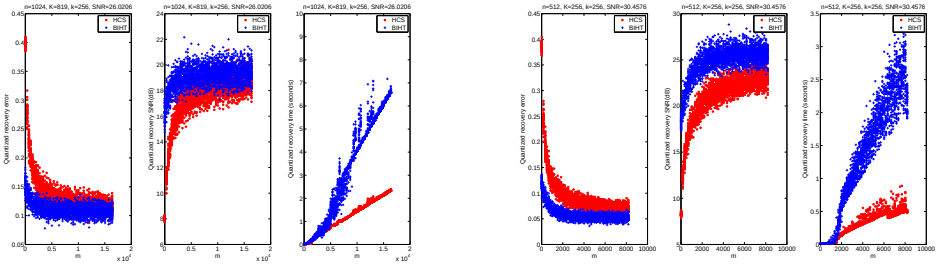

Figure 3: Quantized recovery error vs. number of measurements of 1-bit CQ and "1-bit CS+1-bit CQ quantizer" in the noisy case.

Phase transition in the noiseless case

We first study the phase transition properties of 1-bit CQ and 1-bit CS on quantized recovery error and on recovery time in the noiseless case. We conduct 1-bit CQ and "BIHT+1-bit CQ quantizer" for 10^5 trials. In particular, given fixed n and k, we uniformly choose 100 different K/n values between 0 and 1, and 100 different m/n values between 0 and 4. For each $\{K/n, m/n\}$ pair, we conduct 10 trials, i.e., 1-bit CQ recovery and "1-bit CS+1-bit CQ quantizer" of 10 n-dimensional signals with cardinality K from their m 1-bit measurements. The average quantized recovery errors and average time costs of the two methods on overall 10^4 $\{K/n, m/n\}$ pairs are shown in Figure 2.

In Figure 2, the phase plots of quantized recovery error show the quantized recovery of 1-bit CQ is accurate if the the 1-bit measurements are sufficient. Compared to "1-bit CS+1-bit CQ quantizer", 1-bit CQ needs slightly more measurements to reach the same recovery precision, because 1-bit CS recovers the exact signal, while 1-bit CQ recovers its quantization. However, the phase plots of quantized recovery time shows that 1-bit CQ takes substantially less time than "1-bit CS+1-bit CQ quantizer". Thus 1-bit CQ can significantly improve the efficiency of practical digital systems and eliminate the hardware cost for additional quantization.

Quantized recovery error vs. number of measurements in the noisy case

We show the trade-off between quantized recovery error and the amount of measurements on 2500 trials for noisy signals of different n, K, k and signal-to-noise ratio (SNR). Given fixed n, K, k and SNR, we uniformly choose 50 values of m between 0 and $16n$. For each m value, we conduct 50 trials of 1-bit CQ recovery and "1-bit CS+1-bit CQ quantizer" by recovering the quantizations of 50 noisy signals from their m 1-bit measurements. The quantized recovery error and time cost of each trial are shown in Figure 3.

Figure 3 shows that the quantized recovery error of both 1-bit CQ and "1-bit CS+1-bit CQ quantizer" drops drastically with an increase in the number of measurements. For dense signals with large noise, the two methods perform nearly the same on the recovery accuracy. This phenomenon indicates that 1-bit CQ works well on dense signals and is robust to noise compared to CS and 1-bit CS. In addition, the time taken for 1-bit CQ increases substantially slower than that of "1-bit CS+1-bit CQ quantizer" with an increase in the number of measurements.