Adaptive Dictionary Sparse Signal Recovery using Binary Measurements

Hossein Beheshti, Farzan Haddadi

Abstract—One-bit compressive sensing (CS) is an advanced version of sparse recovery in which the sparse signal of interest can be recovered from extremely quantized measurements. Namely, only the sign of each measurement is available to us. In many applications, the ground truth signal is not sparse itself, but it can be represented in a redundant dictionary. A strong line of research has addressed conventional CS in this signal model including its extension to one-bit measurements. However, one-bit CS suffers from the extremely large number of required measurements to achieve a predefined reconstruction error level. A common alternative to resolve this issue is exploiting adaptive schemes. Adaptive sampling acts on acquired samples to trace the signal in an efficient way. In this work, we utilize adaptive sampling strategy to recover dictionary sparse signals from binary measurements. For this task, a multidimensional threshold is proposed to incorporate the previous signal estimates into the current sampling procedure. This strategy substantially reduces the required number of measurements for exact recovery. Our proof approach is based on the recent tools in high-dimensional geometry like random hyperplane tessellation and Gaussian width. We show through rigorous and numerical analysis that the proposed algorithm considerably outperforms state of the art approaches. Further, our algorithm reaches an exponential error decay in terms of the number of quantized measurements.

Index Terms—One-bit, Dictionary sparse signals, Adaptive measurement, High dimensional geometry.

I. INTRODUCTION

SAMPLING a signal heavily depends on the prior information about the signal structure. For example, if one knows the signal of interest is band-limited, the Nyquist sampling rate is sufficient for exact recovery. Signals with most of its coefficient zero are called sparse. It turns out sparsity is such a powerful assumption that results in a significant reduction of required number of measurements. The science of methods and conditions of signal recovery using small amounts of measurements is called compressed sensing. In compressed sensing, measurement vector is assumed to be a linear combination of the form

$$y = Ax$$

where $A \in \mathbb{R}^{m \times N}$ is measurement matrix, with random elements drawn from a normal distribution and $x \in \mathbb{R}^N$ is an unknown $s$-sparse signal i.e. it has at most $s$ nonzero entries:

$$\|x\|_0 \leq s.$$  \hspace{1cm} (2)

where $\|\cdot\|_0$ is the $\ell_0$-norm which determines the number of non-zero elements. It is shown that $m \sim s \log(N/s)$ measurements is sufficient to guarantee accurate recovery of the signal, by solving the convex program:

$$P_1: \min_{x \in \mathbb{R}^N} \|x\|_1 \text{ s.t. } y = Ax,$$  \hspace{1cm} (3)

with high probability (see [1], [2]).

Practical limitations enforce us to quantize the measurements in (1) as $y = Q(Ax)$ where $Q: \mathbb{R}^m \rightarrow \mathbb{A}^m$ is a non-linear operator that maps the measurements into a finite symbol alphabet $\mathbb{A}$. It is an interesting question what that is the result of extreme quantization? This question was addressed in [3] which states that signal reconstruction is still feasible using only one-bit quantized measurements.

Binary measurement scheme is inspired by the well-known analog comparator as a one-bit analog to digital converter (ADC). In ADC design, increasing the resolution by one more bit exponentially decreases the most achievable sampling rate. Less number of bits also results in lower power consumption of the ADC. Overall, in such applications as ultra-wideband digital receivers or massive multiple-input multiple-output (MIMO) systems, due to bandwidth, power, or operational cost, one-bit ADC is an attractive option. Binary signal model is also utilized in various fields such as binary regression, broadcasting, statistical modeling, ... [4]–[7].

In one-bit compressed sensing, samples are taken as the sign of a linear transform of the signal:

$$y = \text{sign}(Ax).$$  \hspace{1cm} (4)

This sampling scheme discards magnitude information. Therefore, we can only recover signal direction in space. Fortunately, changing the threshold randomly in each measurement as $\tau_i \sim \mathcal{N}(0, 1)$ will conserve the amplitude information. The new sampling scheme is:

$$y = \text{sign}(Ax - \tau).$$  \hspace{1cm} (5)

A great part of compressed sensing literature discusses sparse signals. Nevertheless, most of the natural signals are dictionary sparse i.e. sparse in a transform domain. For instance, sinusoidal signals and natural image are sparse in Fourier and wavelet domains, respectively [8]. This means that our signal of interest $f \in \mathbb{R}^n$ can be expressed as $f = Dx$ where $D \in \mathbb{R}^{n \times N}$ is a redundant dictionary. It is common to use the optimization problem

$$P_{1,D}: \min_{\z \in \mathbb{R}^N} \|D^\mathcal{H}z\|_1 \text{ s.t. } y = Az,$$  \hspace{1cm} (6)

to recover a dictionary sparse signal which is called $\ell_1$ analysis problem.

In this work, we investigate the recovery of dictionary sparse signals from binary measurements. To be more clear, our goal is to solve

$$\min_{\z \in \mathbb{R}^N} \|D^\mathcal{H}z\|_1 \text{ s.t. } y = \text{sign}(Az - \tau),$$  \hspace{1cm} (7)

where $\tau \in \mathbb{R}^m$ is chosen adaptively based on previous estimations (see Figure 3 for more illustration).
A. Contributions

In this section, we state our novelties in compared to previous works. To highlight the contributions, we summarize them below.

1) Proposing a novel algorithm for dictionary sparse signals: We introduce an adaptive thresholding algorithm for reconstructing dictionary sparse signals in case of binary measurements. The proposed algorithm provides accurate signal estimation even in case of redundant and coherent dictionaries. The required number of one-bit measurements considerably outperforms the one in [9].

2) Exponential decay of reconstruction error: The error of our algorithm exhibits exponential decaying behavior as long as the number of adaptive iterations sufficiently grows. To be more precise, we obtain a near-optimal relation between the reconstruction error and the required number of adaptive iterations. Mathematically stated, if one takes the output of our reconstruction algorithm by \( \hat{f} \), then, we show that

\[
\| \hat{f} - f \|_2 \approx \mathcal{O}\left( \frac{1}{2^T} \right),
\]

where \( f \) is the ground truth signal and \( T \) is the number of iterations in our adaptive algorithm (See [2] for more details).

3) High dimensional threshold selection: We propose an adaptive high-dimensional threshold to extract most information from each sample, which substantially improves performance and reduces reconstruction error. Put differently, our proposed threshold has both deterministic and random part; the former shifts the origin to the previous estimate while the latter creates random dithers around the estimate (See [IV-A] for more explanations).

4) Hemisphere projection: We propose a generalized RHT that applies to the sampling scheme [5]. To be more clear, we present a geometric tool (which we call Hemisphere projection) that relates the sampling scheme [5] to [4]. Our generalized RHT to the sampling scheme with non-zero threshold acts as conventional RHT to the simple scheme [4]. This projection also has a second benefit. The touch point of hemisphere and signal subspace (the plane in Figure [4a]) equals the previous estimate and the radius of hemisphere is described as the standard deviation of dithers (random part of high dimensional threshold). The last and important point about hemisphere projection is that the estimation error using generalized RHT on the hemisphere is more magnified when deprojected to the signal subspace. This is since generalized RHT creates a uniform tessellation of hemisphere. This property is due to the fact that hemisphere projection does not preserve size.

B. Prior Work and Key Differences

In this section, we review prior works about applying quantized measurements to CS framework. The authors of [3] propose a heuristic algorithm to reconstruct the ground truth sparse signal from extreme quantized measurements i.e. one bit measurements. In [10], it has been shown that conventional CS algorithms also works well when the measurements are quantized. In [11], the effect of bit depth (number of bits for each measurement) on mean square error (MSE) is studied. They show that, in low signal to noise ratio (SNR) regimes, it is better to use fewer bits per each measurements than taking less measurements with high resolution. There also exist several one-bit algorithms that amount to reconstructing sparse signals such as matching sign pursuit (MSP) [12], restricted-step shrinkage (RSS) [13] and binary iterative hard thresholding (BIHT) [6]. In [6] an algorithm with simple implementation is proposed. This algorithm posses less error in terms of hamming distance compared with the existing algorithms. They show that \( \mathcal{O}(s \log(N)) \) is required for exact recovery. Investigated from a geometric view, the authors of [14], exploit functional analysis tools to provide an almost optimal solution to the problem of one-bit CS. Their approach is based on a new high dimensional concept called random hyperplane tessellation (RHT) theorem. This theorem specifies the minimum number of hyperplanes needed to divide a large set into smaller equal size regions. They show that the number of required one-bit measurements (alternatively the necessary number of hyperplanes in RHT) is \( \mathcal{O}(s \log^2(\frac{s}{\epsilon})) \). The work of [15] presented two algorithms for consummate (direction and norm) reconstruction with provable guarantees. The first approach takes advantage of the random thresholds and the other method, separately predicts the direction and magnitude. The authors in [16] introduce an adaptive thresholding scheme which utilizes a generalized approximate message passing algorithm (GAMP) [16] for recovery and thresholds update throughout sampling. In a different approach, the work [7] proposes an adaptive quantization and recovery scenario making an exponential error decay in one-bit CS frameworks. Many of the techniques mentioned for adaptive sparse signal recovery do not generalize (at least not in an obvious strategy) to dictionary sparse signal. Recently the work [9] shows, both direction and magnitude of a dictionary sparse signal can be recovered by a convex program with strong guarantees. The work [9] has inspired our work for recovering dictionary sparse signal in an adaptive manner. But, there are substantial differences between our work and [9]. In non-adaptive work [9], the error rate is poorly large while in our work, the error rate exponentially decays with increasing the number of adaptive steps. As a rule of thumb our adaptive strategy, enhances the recovery performance of [9] around 50dB in our numerical results.

C. Outline

The paper is organized as follows: in Section [II], first, we provide a short description of geometric tools that used throughout the paper. Then, we study some basic concepts in dictionary sparsity. In Section [III], our system model with some dedicated definition has been explained. In Section [IV], we present our algorithm for the recovery of the dictionary-sparse signal from binary measurements. In Section [V], we investigate the performance of our work and compare it with existing algorithms. Finally, in Appendix, we provide the proof of our main result.
from line to line. We use $\delta$ to denote positive absolute constants which can be different from line to line. We use \[ \|v\|_2 = \sqrt{\Sigma_i v_i^2} \] for the $l_2$-norm of vector $v$ in $\mathbb{R}^n$, $\|v\|_1 = \Sigma_i |v_i|$ for the $l_1$-norm and $\|v\|_\infty = \max_i |v_i|$ for the $l_\infty$-norm. $\|v\|_0$ indicating number of non-zero elements of vector and if $\|v\|_0 \leq s$ the vector $v$ is called $s$-sparse. We also write $B^n_0 := \{v \in \mathbb{R}^n : \|v\|_1 \leq 1\}$ for $l_1$-ball, $B^n_2 := \{v \in \mathbb{R}^n : \|v\|_2 \leq 1\}$ for $l_2$-ball and $S^{n-1} := \{v \in \mathbb{R}^n : \|v\|_2 = 1\}$ for unit Euclidean sphere in $\mathbb{R}^n$. Through the paper, $d_G(v, u) := 1/\pi \arccos(v, u)$ denote the normalized geodesic distance between $u$ and $v$ on $S^{n-1}$ and normalized Hamming distance between $a, b \in \{-1, 1\}^m$ define as $d_H(a, b) := 1/m\langle a, b \rangle$. For a non-invertible matrix $D$ and a set $C$, the inverse operator $D^{-1}(\cdot)$ returns the preimage of $C$ with respect to $D$ i.e. $D^{-1}(C)$.

II. Preliminaries

Since geometric intuition plays an important role in this work, we provide a short description of Gaussian Width and random hyperplane tessellation (RHT). Then, we review the basic concept of dictionary-sparse signals in the CS literature.

A. Gaussian Width

Generally, we need geometric tools to measure the size of a set e.g. $K$. Here, we explain three concepts width, spherical mean width and Gaussian mean width. The width of $K$ in the direction of $\hat{n}$ is equal to the minimum distance of two parallel hyperplanes with normals $\hat{n}$ that include $K$. Written in a mathematical form, the width of $K$ is defined as:

$$\sup_{u, v \in K} \langle \hat{n}, u - v \rangle$$

For example, as shown in Figure 1, the width of $K$ in the direction of $\hat{n}$ is equal to $d$. Considering the direction vector $\hat{n}$ uniformly distributed on $S^{n-1}$ and taking expectation, we reach the new concept spherical mean width:

$$\hat{w}(K) := \mathbb{E} \sup_{u, v \in K} \langle \hat{n}, u - v \rangle$$

In some certain settings, it is more convenient to use standard Gaussian random vectors (instead of spherical random vector $\hat{n}$) in which the elements are chosen independent from each other.

Definition 1 (Gaussian mean width). Let $g \sim N(0, I_n)$ be a standard Gaussian random vector in $\mathbb{R}^n$. The Gaussian mean width of a bounded subset $K$ of $\mathbb{R}^n$ is defined as

$$w(K) := \mathbb{E} \sup_{u, v \in K} \langle g, u - v \rangle$$

In the sequel of the paper, for simplicity, we call Gaussian mean width as mean width. By considering $\hat{n} = g/\|g\|_2$, we may write spherical mean width as

$$\hat{w}(K) := \mathbb{E} \sup_{u, v \in K} \langle g/\|g\|_2, u - v \rangle.$$ 

There is also a relation between spherical mean width and mean width which is given by

$$w(K) = \mathbb{E} \|g\|_2 \hat{w}(K).$$

B. Random Hyperplane Tessellation

Assume that the signal of interest $x$ lies in a subset $K$ of unit Euclidean sphere in $\mathbb{R}^n$ given $m$ measurements of the form $y_i = A_i x + \epsilon_i$. Geometrically, each row of $A$ represents the normal vector of some hyperplanes that tessellate $\mathbb{R}^n$ by some finite cells (see Figure 2 for an interpretation of problem in $\mathbb{R}^3$). Further, each measurement i.e. $y_i$ specifies the side of hyperplane in which $x$ lies. The number of required measurements to achieve a fixed reconstruction error can be cast as the number of hyperplanes encapsulating $x$. The work [17], shows this number of hyperplanes is the squared mean width of $K$ up to a constant multiplicative factor.

Before expressing the main result of [17], it is necessary to define the $\delta$-uniform tessellation which is stated as follows.

Definition 2 ($\delta$-uniform tessellation). [17] Definition 1.1 Let $K$ be a subset of $S^{n-1}$. Consider $m$ hyperplanes in $\mathbb{R}^n$ and let $d(u, v)$ denote the fraction of hyperplanes that separate points $u, v \in K$. We say the set $K$ has been $\delta$-uniform tessellated if

$$|d(u, v) - d_G(v, u)| \leq \delta,$$

for given $\delta > 0$.

The next theorem specifies the number of measurements (alternately hyperplanes), that provides a uniform tessellation of $K$. 

Fig. 1. Width of $K$ in the direction of $\hat{n}$

Fig. 2. The geometric interpretation of hyperplane tessellation
Theorem 1 (Random uniform tessellation). Consider a subset $K \subset S^{n-1}$ and let $\delta > 0$. Let

$$m \geq C\delta^{-6}w(K)^2$$

and consider an arrangement of $m$ independent random hyperplanes in $\mathbb{R}^n$ uniformly distributed according to the Haar measure. Then with probability at least $1 - 2\exp(-c\delta^2m)$, these hyperplanes provide a $\delta$-uniform tessellation of $K$.

Qualitatively spoken, consider Figure 2 and let $u$ and $v$ be two points on the $S^{n-1}$. Assume the normal vectors of hyperplanes be chosen from i.i.d. Gaussian ensemble. If one takes $m = O(w(K)^2)$ measurements and the hypothesized points ($u$ and $v$) satisfy the conditions in Definition 2 then $u$ and $v$ are in the same cell with high probability.

C. Dictionary Sparsity

The basic concepts of dictionary sparsity is explained in [18], [8], and [19]. Here we only provide a brief review of the main ideas. In many practical applications the signal of interest $f \in \mathbb{R}^n$ is not sparse itself, but rather has a sparse representation in an overcomplete dictionary $D \in \mathbb{R}^{n \times N}$ in which there are more columns than rows ($N >> n$). In fact, in the research community, two types of models are investigated confronting with such signals: synthesis sparse signals, analysis sparse signals. In the former, the signal of interest can be represented as $f = Dx = D_s x_s$ while in the latter the sparsity feature of $\Omega f = \Omega \Delta f$ is analyzed. $\Omega \in \mathbb{R}^{p \times n}$ is the analysis operator. $D_s$ and $\Omega \Delta$ is the submatrix of $D$ and $\Omega$ restricted by the columns indexed by $S$ and rows indexed by $\Delta$, respectively. Stated in a different form, the ground truth signal $f \in \mathbb{R}^n$ can be seen as union of some subspaces. To be more clear, $f$ can be stated in two different viewpoints:

Sparse synthesis model: $f \in \cup_{S:|S|=k} \text{span}(D_S)$, \hspace{1cm} (10)

Cosparse analysis model: $f \in \cup_{\Delta:|\Delta|=k} \text{null}(\Omega \Delta)$, \hspace{1cm} (11)

Note that we use cosparsity\footnote{The cosparsity of a signal $f$ with respect to $\Omega \in \mathbb{R}^{N \times n}$ equal to number of zeros in $\Omega f$} to explain the analysis-sparse investigation as a dual of the synthesis-sparse. For these two models, different optimization problems are also investigated in the literature which are called $\ell_1$ synthesis and $\ell_1$ analysis problems as follows.

$$\ell_1 \text{ synthesis: } \min_{z \in \mathbb{R}^N} \|z\|_1 \hspace{1cm} s.t. \hspace{1cm} y = Az$$

$$\ell_1 \text{ analysis: } \min_{z \in \mathbb{R}^N} \|\Omega z\|_1 \hspace{1cm} s.t. \hspace{1cm} y = \Omega z$$

While $\ell_1$ synthesis seeks to find the sparse coefficient which is not unique, $\ell_1$ analysis seeks to find unique ground truth signal $f$. In Figure 3 the process of constructing a dictionary sparse signal is depicted for two models synthesis and analysis.

In the reminder of this work, we consider $\ell_1$ analysis with a tight frame $\Omega = D^H \in \mathbb{R}^{N \times n}$. Put formally, we have $DD^H = I_n$. Different from the conventions in $\ell_1$ synthesis or analysis problems, we consider one bit measurements as shown in Figure 3.

III. System Model

Our system model is build upon the optimization problem \footnote{A major part of this problem is to choose an efficient threshold $\tau \in \mathbb{R}^m$. To this end, we propose a closed-loop feedback system (see Figure 5) which exploits prior information from previous stages. Our adaptive approach consists of two main parts. The former includes measurement acquisition and signal estimation while the latter involves adaptive parameter selection. In the first part, we take a bunch of samples and implement a recovery scheme to exploit a signal estimate. In the second part, the required parameters for the first part are updated based on the estimated signal. Before stating our main results, we fix our definitions which used in the reminder of the work.}

• Hemisphere projection. Our main analysis in this work is based on RHT. RHT in one-bit CS is only used on the Euclidean sphere and the feasible set of our proposed problem (see (7)) does not lie on the sphere. Therefore, we need a tool to transfer any point in $\mathbb{R}^n$ to the sphere. We use the Hemisphere Projection to meet this demand.

Definition 3 (Hemisphere projection). Consider a hemisphere $\mathcal{H}$ with radius $\sigma > 0$. Let $f \in \mathbb{R}^N$ be an arbitrary signal. Define the lifted signal $\tilde{f} := [f^T|\sigma]^T \in \mathbb{R}^{N+1}$. We define lifted $f$ as $\tilde{f} := [f^T|\sigma]^T \in \mathbb{R}^{N+1}$ and it projection to hemisphere as

$$P_{\mathcal{H}}(\tilde{f}) = \frac{\tilde{f}}{\|\tilde{f}\|_2}.$$ \hspace{1cm} (14)

Although might be trivial, nevertheless, geometric interpretation of the hemisphere projection in the definition 3 has special benefits in our work. To more clarify this fact, let us give an example. Let $f_1 = [0,1,0,1]^T$, $f_2 = [0.2,0,1]^T$, $f_3 = [1,0,1]^T$, and $f_4 = [1.1,0.1]^T$ be 4 points on the plane $z = 1$. It is obvious that the distance between $f_1$ and $f_2$ is equal to the distance between $f_3$ and $f_4$ ($d_1 = \|f_1 - f_2\|_2 = d_2 = \|f_3 - f_4\|_2 = 0.1$). We consider $\sigma = 1$ and put the center of our hemisphere on $[0,0,0]^T$ (See Fig 4a). Next, we compute the normalized geodesic distance of the projected points (see Fig 4b) as follows.

$$d_{G1} = d_G(P_{\mathcal{H}}(f_1), P_{\mathcal{H}}(f_2)) = 0.0311$$

$$d_{G2} = d_G(P_{\mathcal{H}}(f_3), P_{\mathcal{H}}(f_4)) = 0.0103.$$
Fig. 3. The process of constructing a dictionary sparse signal and taking measurements in the binary forms.

Fig. 4. The geometric interpretation of hemisphere projection example, typical images are not sparse after the application of wavelet transform. Due to this fact, we introduce a concept that includes this issue. First, we denote two sets $\Sigma^N_s$ for exact $s$-sparse vectors and $\Sigma^N_{s,\text{eff}}$ for effectively $s$-sparse vectors. Note that we say a coefficient vector $\mathbf{x} \in \mathbb{R}^N$ is called effectively $s$-sparse if $\|\mathbf{x}\|_1 \leq \sqrt{s} \|\mathbf{x}\|_2$.

Correspondingly, if the signal of interest is sparse in a dictionary, we say that $f$ is effectively $s$-synthesis sparse if $f = \mathbf{D} \mathbf{x}$ for some $\mathbf{x} \in \mathbb{R}^N$ and effectively $s$-analysis sparse if $\mathbf{D}^* f \in \mathbb{R}^N$.

In what follows in this section, we propose our system model and its main components. Suppose that $f \in \mathbb{R}^n$ is an effective $s$-analysis sparse or $s$-synthesis sparse signal. Let $\mathbf{A}$ be the measurement matrix. In contrast to the existing method [9] for binary dictionary sparse signal recovery which takes all of the measurements in one step with fixed settings, we solve the problem in an adaptive multi-stage manner. In each iteration (which we call it single step recovery (SSR)), regarding the estimate from previous stage, our algorithm is propelled to the desired signal. It is very important to remark that the term adaptivity is related to threshold updating and the measurement matrix $\mathbf{A}$ is fixed. The block diagram of this procedure is depicted in Figure 5.

IV. MAIN RESULT

Our goal in this paper is to design an adaptive thresholding algorithm (based on convex programming) to achieve exponential decay in reconstruction error. We have provided a mathematical framework to guarantee our algorithm results. The following theorem briefly states our purpose.

**Theorem 2** (Main theorem). Consider $f \in (\mathbf{D}^*)^{-1} \Sigma^N_{s,\text{eff}}$ be the desired signal with $\|f\|_2 \leq r$ which is effectively $s$-sparse after the application of the analysis operator $\mathbf{D}^H$. Let $\mathbf{A} \in \mathbb{R}^{m \times n}$ be the measurement matrix with standard normal entries where $m$ is the total number of measurements spread through $T$ stages. Let

$$[y, \varphi] = \text{AS}(\mathbf{A}, f, \mathbf{D}, r, T),$$

$$\hat{f} = \text{AR}(\mathbf{A}, \mathbf{D}, y, \varphi, r, T)$$

be the sampling and recovery algorithms introduced later in Algorithms 2 and 3 respectively where $\varphi$ is determined by Algorithm HDTG. If

then, with probability at least $\tau$ over the choice of $\varphi$ and $\mathbf{A}$, the output of Algorithm AR satisfies

$$\|f - \hat{f}\|_2 \leq \epsilon r 2^{1-T}. \quad (15)$$

A remarkable note is that if we only consider one iteration i.e. $T = 1$, the exponential behavior of our error bound disappears and reaches to state of the art error bound [9, Theorem 8]. In fact, the results of [9] is a special case of our work when the thresholds are non-adaptive.

In what follows, we provide rigorous explanations about proposed algorithms [1] and [2] in three subsections.
A. High Dimensional Threshold

As mentioned in Section II-B, each measurement can be considered as a hyperplane in $\mathbb{R}^n$. Our algorithm for high dimensional threshold selection is given in Algorithm 1. The algorithm output consists of two parts: random ($\tau \in \mathbb{R}^q$) and deterministic ($\Psi \tilde{u}$). The former specifies hyperplanes distance (dither) from the origin (This dither is controlled by the variance parameter $\sigma^2$). While the latter, transfer the origin to the previous signal estimate ($\tilde{u}$). Note that $\Psi$ represents the measurement matrix in each step.

**Algorithm 1 HDTG: High dimension threshold generator**

**Input:** Mapping matrix $A$, number of measurements $q$, dithers variance $\sigma^2$, signal estimation $\hat{f}$.

**Output:** High dimension threshold vector $\varphi \in \mathbb{R}^b$.

1: $\tau \sim N(0, \sigma^2 I_q)$
2: $\varphi = Af + \tau$
3: return $\varphi$

B. Adaptive Sampling

Our adaptive sampling algorithm is given in Algorithm 2. To implement this algorithm, we need the dictionary $D \in \mathbb{R}^{n \times N}$, the measurement matrix $A \in \mathbb{R}^{m \times n}$, linear measurement $Af$ and an over estimation of signal power $r (\|f\|_2 \leq r)$. At the first iteration, we initialize signal estimation to zero vector. By choosing the number of iterations $T$ and dividing the measurement matrix $A$ and the linear measurement vector $Af$ to $T$ blocks of size $q := \frac{m}{T}$, we can start the adaptive sampling process. The adaptive sampling process consists of three essential steps. First, in step 2 of pseudo code, we generate the high dimensional thresholds using Algorithm 1 by variance parameter $\sigma^2 = 2^{1-t}r$ and $\hat{f}_i$. Second, we compare linear measurement block $A^{(i)}f$ with the threshold vector $\varphi^{(i)}$ and obtain the sample vector $y^{(i)}$ (step 3). Third, we implement a second order cone problem which we call single step recovery (SSR). The output of SSR builds the deterministic part of our high dimensional threshold in step 2. Finally, the Algorithm 2 returns binary vectors $\{y^{(i)}\}_{i=1}^T$ and threshold vectors $\{\varphi^{(i)}\}_{i=1}^T$ to the output.

C. Adaptive Recovery

In the Recovery procedure (Algorithm 3), we need the dictionary $D$, the measurement matrix $A$, binary measurements vector $y$, high dimension threshold vector $\varphi$ and an upper norm estimation of signal $r$. In the adaptive recovery algorithm, we first divide the inputs to the $L$ blocks. Then, we simply run the single-step recovery on each block. A notable point in the adaptive recovery is the duplication of operation. If we take a closer look at the adaptive quantization, we will see that Algorithm 3 is a part of Algorithm 2. We can use the adaptive recovery algorithms when linear measurements are not available and we have access to the binary measurements vector. This algorithm is beneficial as long as the linear measurements vector $Af$ is inaccessible. Since in this case, binary samples are stored, it is affordable in terms of memory. This comes to work for example when the sampling and recovery procedure are implemented in two separated systems.

**Algorithm 2 AR: Adaptive Sampling**

**Input:** Dictionary $D \in \mathbb{R}^{n \times N}$, measurement matrix $A \in \mathbb{R}^{m \times n}$, linear measurement $Af \in \mathbb{R}^m$, norm estimation $\|f\|_2 \leq r$, number of blocks $L$.

**Output:** Quantized measurements $y \in \{\pm 1\}^m$, high dimension thresholds $\varphi \in \mathbb{R}^m$

1: Initialization: $f_0 \leftarrow 0$, $q = \left\lceil \frac{m}{T} \right\rceil$, $A^{(i)} \in \mathbb{R}^{q \times m}$
2: for $i = 1, \ldots, T$ do
3: $\varphi^{(i)} \leftarrow \Psi(A^{(i)}, q, 2^{1-t}r, f_i)$
4: $y^{(i)} = \text{sign}(A^{(i)}f - \varphi^{(i)})$
5: $f_i \leftarrow \min_{h \in \mathbb{R}^n} \|D^H h\|_1$
6: return $y^{(i)}, \varphi^{(i)}$ for $i = 1, \ldots, T$

D. Intuitive proof of main theorem

Here, we provide an intuitive proof of Theorem 2. Let’s start with reviewing the expression. Briefly, in this theorem, we run
the single-step recovery for different dithers and center in $T$ stage. We use results from conventional one-bit compressed sensing to prove this theorem. The hemisphere projection acts as a bridge between the conventional one-bit compressed sensing and our work.

Consider a hemisphere with radius $\sigma$ and center the origin. This Hemisphere touch signal space of $f$ at $[0^n | \sigma]^T \in \mathbb{R}^{n+1}$. Remember that each row of the matrix $A$ specifies the normal vector of hyperplane in $\mathbb{R}^n$. This hyperplane can be considered as another hyperplane in $\mathbb{R}^{n+1}$ which passes through the origin. Nevertheless, the new signal space $\sigma$ after projection is the upper half of $\sigma S^n$. Now we can use random hyperplane tessellation to limit the error. As mentioned above we initialized the algorithm with an upper estimation of $\|f\|_2$ and this leads to creating large cells on the hemisphere’s shell. In each step, the algorithm moves the hemisphere to the estimated pint and the size of the hemisphere reduced as regards reduction of $\sigma$. This reduction improves the resolution of mapping and leads to error decrement.

V. NUMERICAL EXPERIMENTS

In this section, we explore the performance our algorithm and compare it to the performance of previous one-bit dictionary sparse recovery given by [9]. The computations, performed in MATLAB using the CVX package, are reproducible and can be downloaded from [Here].

The experimental setup is as follows. The coordinate of random measurements matrix $(a_{i,j})$ were generated as a standard normal distribution. The $s$-sparse coordinate of $x \in \mathbb{R}^N$ for $N = 1000$ selected from uniform random distribution and the values of magnitude are chosen from standard normal distribution. We use redundant dictionary $D \in \mathbb{R}^{n \times N}$ ($N = 1000$, $n = 50$) which forms as follow. First, we construct a matrix where it columns are drawn randomly and independently from $S^{n-1}$. Then, we define our thigh dictionary as an orthonormal basis of the column space of this matrix.

We compare our algorithms result with two convex programming based algorithms, demonstrated in [9]. The first algorithm solves linear programming optimization (LP) [9 Subsection 4.1] and the second algorithm solves second-order cone programming (CP) optimization [9 Subsection 4.2]. These algorithms use the measurement model (5). We generate the threshold coordinates ($\tau$) from an independent normal distribution with mean zero and variance $\sigma^2$. In our algorithm, we set the underestimate of the norm of $f$ to $r = 2 * \|f\|_2$. We also set $\sigma = r$ for LP and CP algorithms. We define the normalized reconstruction error as $\|f - \hat{f}\|_2 / \|f\|_2$. In the our algorithm we set number of blocks $T = 10$.

For each sample of figures we run the simulation 50 times and plot the average of results.

Fig. 6 shows the experiment result for sparsity level 10. As it is clear from the Fig. LP algorithm behaves similar to CP (LP algorithm outperforms CP by 2dB on average). Our algorithm, in fewer samples, behaves slightly weaker than others. But, it seems to be a phase transition when the number of measurements increased and as we see in Fig. our proposed algorithm substantially works better than LP and CP (over 50dB in steady-state condition!).

In the second experiment, we examine the performance of our algorithm for multiple degrees of sparsity. Fig. shows the result for $s = 10, 20, 30, 40, 50$ (we assumed other parameters as the first experiment). As it is clear from the figure, by increasing the sparsity level, our accuracy decreased.

VI. CONCLUSIONS AND FUTURE WORK

In this paper, we proposed an algorithm which exploits the inherent information in the sampling procedure adaptively. This scheme helps to substantially reduce the number of needed measurements. In addition, our algorithm exhibits an exponential decaying behavior in reconstruction error. The proof approach is based on geometric theories in the high dimensional estimation. In this work, we used geometric intuition to explain our result which can be used in other areas of signal processing. We believe our analysis can be extended to the multi-bit setting. Throughout this work, we used a fixed reduction pattern in the thresholds dithers. We believe this

---

2https://gitlab.com/HosseinBeheshti/AdaptiveBDS
reduction can be smart by extracting the geometric features in each step of the algorithm.

REFERENCES

[1] D. L. Donoho, “Compressed sensing,” IEEE Transactions on Information Theory, vol. 52, no. 4, pp. 1289–1306, 2006.
[2] E. J. Candès et al., “Compressive sampling,” in Proceedings of the international congress of mathematicians, vol. 3, pp. 1433–1452, Madrid, Spain, 2006.
[3] P. T. Boufounos and R. G. Baraniuk, “1-bit compressive sensing,” in Information Sciences and Systems, 2008. CISS 2008. 42nd Annual Conference on, pp. 16–21, IEEE, 2008.
[4] H. Yin, Z. Wang, L. Ke, and J. Wang, “Monobit digital receivers: design, performance, and application to impulse radio,” IEEE Transactions on Communications, vol. 58, pp. 1695–1704, Jun 2010.
[5] C. Risi, D. Persson, and E. G. Larsson, “Massive mimo with 1-bit adc,” arXiv preprint arXiv:1404.7736, 2014.
[6] 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 Transactions on Information Theory, vol. 59, no. 4, pp. 2082–2102, 2013.
[7] R. G. Baraniuk, S. Foucart, D. Needell, Y. Plan, and M. Wootters, “Exponential decay of reconstruction error from binary measurements of sparse signals,” IEEE Transactions on Information Theory, vol. 63, no. 6, pp. 3368–3385, 2017.
[8] E. J. Candès, Y. C. Eldar, D. Needell, and P. Randall, “Compressed sensing with coherent and redundant dictionaries,” Applied and Computational Harmonic Analysis, vol. 31, pp. 59–73, Jul 2011.
[9] R. Baraniuk, S. Foucart, D. Needell, Y. Plan, and M. Wootters, “One-bit compressive sensing of dictionary-sparse signals,” Information and Inference: A Journal of the IMA, vol. 7, pp. 83–104, aug 2017.
[10] A. Zymnis, S. Boyd, and E. Candès, “Compressed sensing with quantized measurements,” IEEE Signal Processing Letters, vol. 17, no. 2, pp. 149–152, 2010.
[11] J. N. Laska and R. G. Baraniuk, “Regime change: Bit-depth versus measurement-rate in compressive sensing,” IEEE Transactions on Signal Processing, vol. 60, no. 7, pp. 3496–3505, 2012.
[12] P. T. Boufounos, “Greedy sparse signal reconstruction from sign measurements,” in 2009 Conference Record of the Forty-Third Asilomar Conference on Signals, Systems and Computers, pp. 1305–1309, IEEE, 2009.
[13] J. N. Laska, Z. Wen, W. Yin, and R. G. Baraniuk, “Trust, but verify: Fast and accurate signal recovery from 1-bit compressive measurements,” IEEE Transactions on Signal Processing, vol. 59, no. 11, pp. 5289–5301, 2011.
[14] Y. Plan and R. Vershynin, “One-bit compressed sensing by linear programming,” Communications on Pure and Applied Mathematics, vol. 66, no. 8, pp. 1275–1297, 2013.
[15] K. Knudson, R. Saab, and R. Ward, “One-bit compressive sensing with norm estimation,” IEEE Transactions on Information Theory, vol. 62, no. 5, pp. 2748–2758, 2016.
[16] U. S. Kamilov, A. Bourquard, A. Amini, and M. Unser, “One-bit measurements with adaptive thresholds,” IEEE Signal Processing Letters, vol. 19, no. 10, pp. 607–610, 2012.
[17] Y. Plan and R. Vershynin, “Dimension reduction by random hyperplane tessellations,” Discrete & Computational Geometry, vol. 51, no. 2, pp. 438–461, 2014.
[18] M. Elad, P. Milanfar, and R. Rubinstein, “Analysis versus synthesis in signal priors,” Inverse problems, vol. 23, no. 3, p. 947, 2007.
[19] S. Nam, M. E. Davies, M. Elad, and R. Gribonval, “The cosparse analysis model and algorithms,” Applied and Computational Harmonic Analysis, vol. 34, no. 1, pp. 30–56, 2013.