One-bit Compressed Sensing with the k-Support Norm

Sheng Chen **Arindam Banerjee**
Department of Computer Science and Engineering, University of Minnesota, Twin Cities
{shengc,banerjee}@cs.umn.edu

Abstract

In one-bit compressed sensing (1-bit CS), one attempts to estimate a structured parameter (signal) only using the sign of suitable linear measurements. In this paper, we investigate 1-bit CS problems for sparse signals using the recently proposed k-support norm. We show that the new estimator has a closed-form solution, so no optimization is needed. We establish consistency and recovery guarantees of the estimator for both Gaussian and sub-Gaussian random measurements. For Gaussian measurements, our estimator is comparable to the best known in the literature, along with guarantees on support recovery. For sub-Gaussian measurements, our estimator has an irreducible error which, unlike existing results, can be controlled by scaling the measurement vectors. In both cases, our analysis covers the setting of model misspecification, i.e., when the true sparsity is unknown. Experimental results illustrate several strengths of the new estimator.

1 Introduction

In recent years, one-bit compressed sensing (1-bit CS) (Boufounos & Baraniuk, 2008) for estimation of a sparse or structured parameter (signal) has become increasingly popular due to its low implementation cost and robustness (Boufounos, 2010). Compared with conventional compressed sensing (Donoho, 2006; Candes & Tao, 2006), which tries to recover a signal using real-valued measurements, 1-bit CS quantizes each measurement into $+1$ or -1 instead.

Previous work on 1-bit CS can be categorized based on assumptions on noise and measurements. In the noiseless setting (Jacques et al., 2013; Gopi et al., 2013; Plan & Vershynin, 2013a), 1-bit CS estimates a signal vector $\mathbf{x}^* \in \mathbb{R}^p$ from $\mathbf{y} = \text{sign}(\mathbf{U}\mathbf{x}^*)$, where $\mathbf{y} = \{-1, +1\}^n$ are the measurements, and $\mathbf{U} = [\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_n]^T \in \mathbb{R}^{n \times p}$ is the measurement matrix. In the noisy setting (Plan & Vershynin, 2013b), y_i is a suitable noisy function of the inner product $\langle \mathbf{u}_i, \mathbf{x}^* \rangle$ (see Section 2). In the non-adaptive or passive setting, $\mathbf{u}_i$ are assumed to be random samples from a suitable distribution, e.g., $u_{ij} \sim \mathcal{N}(0, 1)$; in the adaptive or active setting, $\mathbf{u}_i$ are chosen sequentially based on prior measurements. In this paper, we focus on a general noisy setting for 1-bit CS with random measurements where the signal $\mathbf{x}^*$ is assumed to be k-sparse, i.e., has k non-zero entries.

Given an estimate $\hat{\mathbf{x}}$ obtained from $(\mathbf{y}, \mathbf{U})$, consistency analysis for 1-bit CS considers the error $\|\hat{\mathbf{x}} - \mathbf{x}^*\|_2$ and also the support of $\hat{\mathbf{x}}$ and $\mathbf{x}^*$. For the noisy setting, Plan & Vershynin (2013b) proposed a constrained optimization framework with a linear objective. This convex formulation can work with a general notion of noise and achieve $O(\sqrt[4]{\frac{k \log p}{n}})$ error for both exactly and approximately k-sparse signals. Recently Zhang et al. (2014) considered a variant of this formulation and developed an efficient passive algorithm with closed-form solution, which improves the error bound to $O(\sqrt{\frac{k \log p}{n}})$ for exactly k-sparse signal. Gupta et al. (2010) and Haupt & Baraniuk (2011) studied the support recovery of $\mathbf{x}^*$ based on thresholding methods, both of which can achieve support recovery with a sample complexity $O(k \log p)$. Most such existing results rely on using Gaussian measurements. One notable exception is Ai et al. (2014), which extends the work by (Plan & Vershynin, 2013b) to sub-Gaussian measurements. In contrast with the Gaussian case, one gets an irreducible component in the error which depends on $\|\mathbf{x}^*\|_\infty$ and cannot be controlled/reduced by increasing the sample size or otherwise.

In this paper, we focus on recovering exactly k-sparse $\mathbf{x}^*$ in the noisy and non-adaptive measurement setting for 1-bit CS. Building on the work in (Plan & Ver-

Appearing in Proceedings of the 18^{th} International Conference on Artificial Intelligence and Statistics (AISTATS) 2015, San Diego, CA, USA. JMLR: W&CP volume 38.

shynin, 2013b), we propose a simple *closed-form estimator* based on the recently proposed *k-support norm*. In previous studies (Argyriou et al., 2012; Chatterjee et al., 2014; McDonald et al., 2014), the k-support norm has been shown to be an effective alternative to the elastic net (Zou & Hastie, 2005) in estimating correlated parameter with theoretical guarantees. For the 1-bit CS setting, we first establish recovery guarantees of the new closed-form estimator for Gaussian measurement matrices. Our general results and analyses allow for model misspecification, i.e., do not assume knowledge of true sparsity. Such analyses based on model misspecification has not been considered in previous work. With the model correctly specified, our analysis yields a similar error bound compared with the best known result in (Zhang et al., 2014), and also matches the sample complexity for support recovery in (Gupta et al., 2010), which is not available in (Zhang et al., 2014). For sub-Gaussian measurement matrices, the bound we obtain contains an additional *irreducible* error term depending on $\mathbf{x}^*$, which can be related to the one in (Ai et al., 2014). Interestingly, our irreducible error can be controlled under certain situation by using properly scaled sub-Gaussian distributions, which cannot be achieved in (Ai et al., 2014). Through experiments with both Gaussian and sub-Gaussian measurements, we show the effectiveness of our estimator.

The rest of the paper is organized as follows: In Section 2, we introduce our k-support norm estimator for 1-bit CS. In Section 3 and 4, we present the recovery analysis for Gaussian and sub-Gaussian measurement matrix, respectively. In Section 5, we present experimental results, and we conclude in Section 6.

2 1-bit CS with the k-Support Norm

In this section, we introduce the new k-support norm estimator for 1-bit CS. We focus on recovering an *exactly k-sparse* signal $\mathbf{x}^* \in \mathbb{R}^p$ with $\|\mathbf{x}^*\|_2 = 1$ from n 1-bit measurements, $\mathbf{y} = [y_1, y_2, \ldots, y_n]^T \in \{+1, -1\}^n$. Provided with a measurement matrix $\mathbf{U} = [\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n]^T \in \mathbb{R}^{n\times p}$, whose entries are independently drawn from an identical distribution, each measurement y_i is assumed to be generated randomly based on the quantity $\langle \mathbf{u}_i, \mathbf{x}^*\rangle$, satisfying

$$\mathbf{E}[y_i | \langle \mathbf{u}_i, \mathbf{x}^*\rangle] = \theta(\langle \mathbf{u}_i, \mathbf{x}^*\rangle) \ ,$$

where θ is some nonlinear function, representing the noise, with a range $[-1, 1]$. The estimator as well as the analysis does not assume knowledge of θ, allowing for fairly general noise models. Plan & Vershynin (2013b) proposed the following estimator to recover $\mathbf{x}^*$:

$$\max_{\mathbf{x}\in\mathbb{R}^p} \ \langle \mathbf{x}, \mathbf{U}^T\mathbf{y}\rangle \quad \text{s.t.} \ \ \mathbf{x} \in \mathcal{K} \ , \tag{1}$$

where $\mathcal{K}$ is ideally a convex signal set. In particular, the set $\mathcal{K} = \{\mathbf{x} \mid \|\mathbf{x}\|_1 \le \sqrt{k}, \ \|\mathbf{x}\|_2 \le 1\}$ is considered to approximate the k-sparse signal set $\mathcal{S} = \{\mathbf{x} \mid \|\mathbf{x}\|_0 \le k, \ \|\mathbf{x}\|_2 \le 1\}$, which is non-convex. Instead of using such $\mathcal{K}$, we propose using the convex hull $\mathrm{conv}(\mathcal{S})$, the tightest convex relaxation of $\mathcal{S}$, which leads to the following problem,

$$\max_{\mathbf{x}\in\mathbb{R}^p} \ \langle \mathbf{x}, \mathbf{U}^T\mathbf{y}\rangle \quad \text{s.t.} \ \ \mathbf{x} \in \mathrm{conv}(\mathcal{S}) \ . \tag{2}$$

As discussed in (Argyriou et al., 2012), $\mathrm{conv}(\mathcal{S})$ is in fact the unit ball of k-support norm, defined as

$$\|\mathbf{x}\|_k^{sp} = \inf_{\mathbf{v}_I} \Big\{ \sum_{I\in\mathcal{G}_k} \|\mathbf{v}_I\|_2 \ \Big| \ \mathrm{supp}(\mathbf{v}_I) \subseteq I, \sum_{I\in\mathcal{G}_k} \mathbf{v}_I = \mathbf{x} \Big\},$$

where $\mathcal{G}_k$ is the collection of all index sets I with $|I| = k$, and $\mathrm{supp}(\mathbf{v}_I)$ denotes the support of $\mathbf{v}_I$. Hence (2) is equivalent to

$$\max_{\mathbf{x}\in\mathbb{R}^p} \ \langle \mathbf{x}, \mathbf{U}^T\mathbf{y}\rangle \quad \text{s.t.} \ \ \|\mathbf{x}\|_k^{sp} \le 1 \ . \tag{3}$$

We note that this convex program simply computes the dual norm of the k-support norm, which turns out to be *2-k symmetric gauge* $\|\mathbf{z}\|_{(k)} = \||\mathbf{z}|_{1:k}^{\downarrow}\|_2$, where $|\mathbf{z}|^{\downarrow}$ denotes the permuted vector of $|\mathbf{z}|$ with entries sorted in decreasing order. The following lemma characterizes the *closed-form* solution to (3).

Lemma 1 (k-support norm estimator) *Let* $\hat{\mathbf{z}} = \mathbf{U}^T\mathbf{y}$. *The solution* $\hat{\mathbf{x}}$ *to the convex program* (3) *is given by*

$$\hat{x}_i = \begin{cases} \frac{\hat{z}_i}{\||\hat{\mathbf{z}}|_{1:k}^{\downarrow}\|_2}, & \text{if } |\hat{z}_i| \text{ is in the largest} \\ & \quad k \text{ entries of } |\hat{\mathbf{z}}| \\ 0, & \text{if otherwise} \end{cases} \ . \tag{4}$$

Proof: It is easy to verify that $\langle \hat{\mathbf{x}}, \mathbf{U}^T\mathbf{y}\rangle = \|\mathbf{U}^T\mathbf{y}\|_{(k)}$ and $\hat{\mathbf{x}}$ is in $\mathrm{conv}(\mathcal{S})$. Hence $\hat{\mathbf{x}}$ is the solution to (3). ■

Note that the recovered signal $\hat{\mathbf{x}}$ is k-sparse, thus belonging to $\mathcal{S}$, which is the sparse signal set of interest. More generally, if $\mathcal{K}$ is a closed convex set, the convex program (1) is to solve the *polar operator* of the gauge function induced by $\mathcal{K}$, which has efficient algorithms for certain other $\mathcal{K}$ (Zhang et al., 2013).

3 Recovery Analysis: Gaussian Measurements

For the analysis,we assume that the quantity λ defined as

$$\lambda \triangleq \mathbf{E}[\theta(g)g] \ , \tag{5}$$

where g is a standard Gaussian random variable, is strictly positive, i.e., $\lambda > 0$. Next we show that with high probability, the $\hat{\mathbf{x}}$ obtained from Lemma 1 has small ℓ_2 error $\|\hat{\mathbf{x}} - \mathbf{x}^*\|_2$, and nonzero entries in $\hat{\mathbf{x}}$ exactly give the support of $\mathbf{x}^*$. We start by recalling the following lemma from (Zhang et al., 2014).

Lemma 2 *The expectation of* $\mathbf{z} = \frac{\mathbf{U}^T\mathbf{y}}{\lambda n}$ *is* $\mathbf{E}[\mathbf{z}] = \frac{1}{\lambda}\mathbf{E}[\theta(\langle \mathbf{g}, \mathbf{x}^*\rangle)\mathbf{g}] = \mathbf{x}^*$*, in which* $\mathbf{g}$ *is a standard Gaussian random vector. Moreover, with probability at least* $1 - e^{1-t}$*, the following inequality holds,*

$$\|\mathbf{z} - \mathbf{x}^*\|_\infty \le \frac{c}{\lambda}\sqrt{\frac{t + \log p}{n}} \ , \tag{6}$$

where c *is an absolute constant.*

The proof of Lemma 2 can be found in (Zhang et al., 2014). The lemma also provides an intuition for our estimator since the support of $\mathbf{x}^*$ corresponds to the largest k entries of $|\mathbf{z}|$ in expectation. Theorem 3 gives the recovery error bound in a general scenario, allowing for model misspecification, i.e., not assuming knowledge of the true sparsity of $\mathbf{x}^*$.

Theorem 3 *Given a* k_0*-sparse signal* $\mathbf{x}^* \in \mathbb{R}^p$ *with* $\|\mathbf{x}^*\|_2 = 1$ *and a standard Gaussian measurement matrix* $\mathbf{U} \in \mathbb{R}^{n\times p}$*, we choose a specific* k *in the convex program* (3) *to obtain* $\hat{\mathbf{x}}$ *(possibly* $k \neq k_0$*). Assume that* $k_1 (\le k_0)$ *nonzero entries of* $\mathbf{x}^*$ *are recovered. Then the recovered signal* $\hat{\mathbf{x}}$*, with probability at least* $1 - \eta e^{1-t}$*, satisfies*

$$\begin{aligned}\|\hat{\mathbf{x}} - \mathbf{x}^*\|_2 &\le \frac{c_1}{\lambda}\sqrt{\frac{k(t+\log p)}{n}} + \frac{c_2 k(t + \log k)}{\lambda^2 n} \\ &\quad + \frac{c_3 \tau k_1 (t + \log k_1)}{\lambda\sqrt{n}} + \xi^2(k_0 - k_1) + \xi\sqrt{k_0 - k_1} \\ &= O(\sqrt{\frac{k\log p}{n}} + \frac{k_1 \log k_1}{\sqrt{n}}) + \xi^2(k_0 - k_1) + \xi\sqrt{k_0 - k_1}\end{aligned}$$

where η*,* c_1, c_2 *and* c_3 *are absolute constants, and* τ *and* ξ *are defined as*

$$\tau = \max\left\{x_i^* \ \middle|\ \hat{x}_i \neq 0\right\}, \quad \xi = \max\left\{x_i^* \ \middle|\ \hat{x}_i = 0\right\} \ .$$

Proof: With loss of generality, we assume that the entries in $\hat{\mathbf{x}}$ and $\mathbf{x}^*$ are rearranged such that $\hat{\mathbf{x}}_{1:k}$ contains all k nonzero entries in $\hat{\mathbf{x}}$, where $\hat{\mathbf{x}}_{1:k_1}$ correspond to k_1 nonzero entries in $\mathbf{x}^*$, and $\hat{\mathbf{x}}_{k_1+1:k}$ correspond to $k - k_1$ zeros in $\mathbf{x}^*$. Then we establish ℓ_2 error bound

$$\begin{aligned}\|\hat{\mathbf{x}} - \mathbf{x}^*\|_2 &\le \|\hat{\mathbf{x}}_{1:k} - \mathbf{x}^*_{1:k}\|_2 + \|\hat{\mathbf{x}}_{k+1:p} - \mathbf{x}^*_{k+1:p}\|_2 \\ &= \Big\|\frac{\mathbf{z}_{1:k}}{\|\mathbf{z}_{1:k}\|_2} - \mathbf{x}^*_{1:k}\Big\|_2 + \|\mathbf{x}^*_{k+1:p}\|_2 \\ &\le \|\mathbf{z}_{1:k} - \mathbf{x}^*_{1:k}\|_2 + \Big\|\frac{\mathbf{z}_{1:k}}{\|\mathbf{z}_{1:k}\|_2} - \mathbf{z}_{1:k}\Big\|_2 + \|\mathbf{x}^*_{k+1:p}\|_2 \\ &\le \sqrt{k}\|\mathbf{z}_{1:k} - \mathbf{x}^*_{1:k}\|_\infty + \big|\|\mathbf{z}_{1:k}\|_2^2 - 1\big| + \xi\sqrt{k_0 - k_1} \\ &\le \frac{c_1}{\lambda}\sqrt{\frac{k(t+\log p)}{n}} + \Big|\sum_{i=1}^k z_i^2 - 1\Big| + \xi\sqrt{k_0 - k_1} \ ,\end{aligned} \tag{7}$$

in which $c_1 \triangleq c$. To complete the analysis, we only need to bound $|\sum_{i=1}^k z_i^2 - 1|$. We decompose it as

$$\begin{aligned}\Big|\sum_{i=1}^k z_i^2 - 1\Big| &\le \Big|\sum_{i=1}^{k_1} z_i^2 - \sum_{i=1}^{k_1}(x_i^*)^2\Big| + \sum_{i=k_1+1}^{k} z_i^2 \\ &\quad + \Big|1 - \sum_{i=1}^{k_1}(x_i^*)^2\Big| \\ &\le \sum_{i=1}^{k_1}\big|z_i^2 - (x_i^*)^2\big| + \sum_{i=k_1+1}^{k} z_i^2 + \Big|1 - \sum_{i=1}^{k_1}(x_i^*)^2\Big| \\ &\le \sum_{i=1}^{k_1}\big|z_i^2 - (x_i^*)^2\big| + \sum_{i=k_1+1}^{k} z_i^2 + \xi^2(k_0 - k_1) \ ,\end{aligned}$$

Using the results from (Zhang et al., 2014; Vershynin, 2012), we know $u_{ji}y_j$ is a sub-gaussian random variable with sub-Gaussian norm $\|u_{ji}y_j\|_{\psi_2} = \|u_{ji}\|_{\psi_2} \triangleq K$, then the centered sub-gaussian random variable $\frac{u_{ji}y_j}{\lambda} - x_i^*$ has $\|\frac{u_{ji}y_j}{\lambda} - x_i^*\|_{\psi_2} \le \frac{2K}{\lambda}$. Further, $z_i - x_i^* = \frac{1}{n}\sum_{j=1}^n(\frac{u_{ji}y_j}{\lambda} - x_i^*)$ is an average of n independent centered sub-gaussian random variables, whose sub-gaussian norm satisfies $\|z_i - x_i^*\|_{\psi_2}^2 \le \frac{4CK^2}{\lambda^2 n}$, in which C is an absolute constant. Hence $(z_i - x_i^*)^2$ is sub-exponential with its sub-exponential norm satisfying

$$\begin{aligned}\|(z_i - x_i^*)^2\|_{\psi_1} \le 2\|z_i - x_i^*\|_{\psi_2}^2 \le \frac{8CK^2}{\lambda^2 n} \quad &\Longrightarrow \\ \|z_i^2 - 2x_i^* z_i + 2(x_i^*)^2 - (x_i^*)^2\|_{\psi_1} \le \frac{8CK^2}{\lambda^2 n} \ ,\end{aligned}$$

By triangular inequality for sub-exponential norm, we get

$$\begin{aligned}\|z_i^2 - (x_i^*)^2\|_{\psi_1} - \|2x_i^* z_i - 2(x_i^*)^2\|_{\psi_1} &\le \frac{8CK^2}{\lambda^2 n} \\ \|z_i^2 - (x_i^*)^2\|_{\psi_1} &\le \frac{8CK^2}{\lambda^2 n} + 2|x_i^*|\|z_i - x_i^*\|_{\psi_1}\end{aligned}$$

For $1 \le i \le k_1$, using the fact $\|\cdot\|_{\psi_1} \le \|\cdot\|_{\psi_2}$, we have

$$\begin{aligned}\|z_i^2 - (x_i^*)^2\|_{\psi_1} &\le \frac{8CK^2}{\lambda^2 n} + 2\tau\|z_i - x_i^*\|_{\psi_2} \\ &\le \frac{8CK^2}{\lambda^2 n} + \frac{4K\tau}{\lambda}\sqrt{\frac{C}{n}} \ ,\end{aligned}$$

For $k_1 < i \le k$, as $x_i^* = 0$, we have

$$\|z_i^2\|_{\psi_1} = \|z_i^2 - (x_i^*)^2\|_{\psi_1} \le \frac{8CK^2}{\lambda^2 n} \ .$$

Hence we obtain the following concentrations for z_i^2 by the definition of sub-exponential variable,

$$\mathbb{P}\Big\{\big|z_i^2 - (x_i^*)^2\big| > \epsilon_1\Big\} \le \exp\Big(1 - \frac{C'\epsilon_1}{\frac{8CK^2}{\lambda^2 n} + \frac{4K\tau}{\lambda}\sqrt{\frac{C}{n}}}\Big), \quad \text{if } 1 \le i \le k_1 \ ,$$

$$\mathbb{P}\Big\{z_i^2 > \epsilon_2\Big\} \le \exp\Big(1 - \frac{C'\epsilon_2}{\frac{8CK^2}{\lambda^2 n}}\Big), \quad \text{if } k_1 < i \le k \ ,$$

in which C' is also an absolute constant. Taking the union bound over all z_i^2, we get

$$\mathbb{P}\Big\{\sum_{i=1}^{k_1}\big|z_i^2-(x_i^*)^2\big|>k_1\epsilon_1\Big\}\le\sum_{i=1}^{k_1}\mathbb{P}\Big\{\big|z_i^2-(x_i^*)^2\big|>\epsilon_1\Big\}$$
$$\le\exp\Big(1-\frac{C'\epsilon_1}{\frac{8CK^2}{\lambda^2n}+\frac{4K\tau}{\lambda}\sqrt{\frac{C}{n}}}+\log k_1\Big)\ ,$$

$$\mathbb{P}\Big\{\sum_{i=k_1+1}^{k}z_i^2>(k-k_1)\epsilon_2\Big\}\le\sum_{i=k_1+1}^{k}\mathbb{P}\Big\{z_i^2>\epsilon_2\Big\}$$
$$\le\exp\Big(1-\frac{C'\epsilon_2}{\frac{8CK^2}{\lambda^2n}}+\log(k-k_1)\Big)\ ,$$

Let $t=\frac{C'\epsilon_1}{\frac{8CK^2}{\lambda^2n}+\frac{4K\tau}{\lambda}\sqrt{\frac{C}{n}}}-\log k_1=\frac{C'\epsilon_2}{\frac{8CK^2}{\lambda^2n}}-\log(k-k_1)$, then we get the following by using t to represent ϵ_1 and ϵ_2,

$$\begin{aligned}&\sum_{i=1}^{k_1}\big|z_i^2-(x_i^*)^2\big|+\sum_{i=k_1+1}^{k}z_i^2\\
&\le k_1\big(t+\log k_1\big)\Big(\frac{8CK^2}{C'\lambda^2n}+\frac{4K\tau}{C'\lambda}\sqrt{\frac{C}{n}}\Big)\\
&\quad+\big(k-k_1\big)\big(t+\log(k-k_1)\big)\Big(\frac{8CK^2}{C'\lambda^2n}\Big)\\
&\le k(t+\log k)\frac{8CK^2}{C'\lambda^2n}+k_1(t+\log k_1)\frac{4K\tau}{C'\lambda}\sqrt{\frac{C}{n}}\\
&\triangleq\frac{c_2k(t+\log k)}{\lambda^2n}+\frac{c_3\tau k_1(t+\log k_1)}{\lambda\sqrt{n}}\ ,\end{aligned}$$

with probability at least $1-2e^{1-t}$. Combining it with the bound for $\|\mathbf{z}_{1:k}-\mathbf{x}^*_{1:k}\|_2$ and other terms, we obtain

$$\begin{aligned}\|\hat{\mathbf{x}}-\mathbf{x}^*\|_2&\le\frac{c_1}{\lambda}\sqrt{\frac{k(t+\log p)}{n}}+\frac{c_2k(t+\log k)}{\lambda^2n}\\
&\quad+\frac{c_3\tau k_1(t+\log k_1)}{\lambda\sqrt{n}}+\xi^2(k_0-k_1)+\xi\sqrt{k_0-k_1}\\
&=O\Big(\sqrt{\frac{k\log p}{n}}+\frac{k_1\log k_1}{\sqrt{n}}\Big)+\xi^2(k_0-k_1)+\xi\sqrt{k_0-k_1}\end{aligned}$$

with probability at least $1-3e^{1-t}\triangleq1-\eta e^{1-t}$. ■

Remark In (7), the recovery error is decomposed into three parts: the error due to empirical mean $\|\mathbf{z}_{1:k}-\mathbf{x}^*_{1:k}\|_2$, the scale error $\|\frac{\mathbf{z}_{1:k}}{\|\mathbf{z}_{1:k}\|_2}-\mathbf{z}_{1:k}\|_2$, and the error incurred by unrecovered support $\|\mathbf{x}^*_{k+1:p}\|_2$. The first two errors can be reduced by increasing sample size. Interestingly the scale error does not quite depend on the choice of k even if $\mathbf{z}_{1:k}$ is k-dimensional. The third depends explicitly on the choice of k and implicitly on the sample size and $\mathbf{x}^*$ itself.

Based on the general result, we consider some special cases below:

Corollary 4 *Under the setting of Theorem 3, denote the largest and smallest nonzero elements in $|\mathbf{x}^*|$ by $\kappa_{\max}$ and $\kappa_{\min}$ respectively. If $n\ge\frac{c_4(t+\log p)}{\lambda^2\kappa_{\min}^2}$ and $k\ge k_0$ (over-specified sparsity), then with probability at least $1-\eta e^{1-t}$, the following happen:*

- *The support of $\mathbf{x}^*$ is recovered, i.e., $\{i\mid x_i^*\neq0\}\subseteq\{i\mid\hat{x}_i\neq0\}$,*
- $\|\hat{\mathbf{x}}-\mathbf{x}^*\|_2\le\frac{c_1}{\lambda}\sqrt{\frac{k(t+\log p)}{n}}+\frac{c_2k(t+\log k)}{\lambda^2n}+\frac{c_3\kappa_{\max}k_0(t+\log k_0)}{\lambda\sqrt{n}}=O\big(\sqrt{\frac{k\log p}{n}}+\frac{k_0\log k_0}{\sqrt{n}}\big)$,

If $n\ge\frac{c_4(t+\log p)}{\lambda^2\kappa_{\min}^2}$ and $k<k_0$ (under-specified sparsity), with probability at least $1-\eta e^{1-t}$, the following happen:

- *k out of k_0 nonzero entries are recovered, i.e., $\{i\mid\hat{x}_i\neq0\}\subseteq\{i\mid x_i^*\neq0\}$,*
- $\|\hat{\mathbf{x}}-\mathbf{x}^*\|_2\le\frac{c_1}{\lambda}\sqrt{\frac{k(t+\log p)}{n}}+\frac{c_2k(t+\log k)}{\lambda^2n}+\frac{c_3\kappa_{\max}k(t+\log k)}{\lambda\sqrt{n}}+\xi^2(k_0-k)+\xi\sqrt{k_0-k}=O\big(\sqrt{\frac{k\log p}{n}}+\frac{k\log k}{\sqrt{n}}\big)+\xi^2(k_0-k)+\xi\sqrt{k_0-k}$,

Proof: Let $c_4=4c^2$, where c is the constant in Lemma 2. Then we get $\frac{c}{\lambda}\sqrt{\frac{t+\log p}{n}}\le\frac{\kappa_{\min}}{2}$ from $n\ge\frac{c_4(t+\log p)}{\lambda^2\kappa_{\min}^2}$. According to Lemma 2, we know $\|\mathbf{z}-\mathbf{x}^*\|_\infty\le\frac{\kappa_{\min}}{2}$. If $k\ge k_0$, then it is sufficient for all k_0 nonzero entries to be recovered by $\hat{\mathbf{x}}$. Hence $k_1=k_0$ and $\tau=\kappa_{\max}$. Substituting them into Theorem 3, we get the result for $k\ge k_0$. If $k<k_0$, then only k out of k_0 nonzero entries are recovered, i.e., $k_1=k$ and $\tau\le\kappa_{\max}$, which gives the similar result for $k<k_0$. ■

Remark From the results above, if $k=k_0$ (correct sparsity), the signal support can be *exactly recovered* with $n=\frac{c_4(t+\log p)}{\lambda^2\kappa_{\min}^2}=\frac{c_4(t+\log p)}{\lambda^2\kappa_{\max}^2}\frac{\kappa_{\max}^2}{\kappa_{\min}^2}\le\frac{c_4\kappa_{\max}^2}{\lambda^2\kappa_{\min}^2}k_0(t+\log p)=O(k_0\log p)$, which matches the sample complexity in (Gupta et al., 2010), and the error bound $O\big(\sqrt{\frac{k_0\log p}{n}}+\frac{k_0\log k_0}{\sqrt{n}}\big)$ is comparable to the best known result $O\big(\sqrt{\frac{k_0\log p}{n}}\big)$ in (Zhang et al., 2014). The additional term $\frac{k_0\log k_0}{\sqrt{n}}$ stems from the analysis of scale error $\|\frac{\mathbf{z}_{1:k}}{\|\mathbf{z}_{1:k}\|_2}-\mathbf{z}_{1:k}\|_2$. Our analysis also yields results under model misspecification. When the sample size is sufficient large, a slightly larger $k>k_0$ (over-specified sparsity) will not increase the recovery error by much, while a smaller $k<k_0$ (under-specified sparsity) can impact the error adversely due to the presence of irreducible terms, which do not diminish with sample size, due to unrecovered support. Note that the results also explicitly involve $\kappa_{\max}$ and $\kappa_{\min}$, and it is not difficult to see that smaller $\kappa_{\max}$ and larger $\kappa_{\min}$ would yield

better recovery, which is the case that nonzero entries in $\mathbf{x}^*$ are correlated, e.g., their magnitudes are roughly equal (See Section 5 for empirical results).

4 Recovery Analysis: Sub-Gaussian Measurements

In this section, we consider 1-bit CS based on sub-Gaussian measurement matrix $\mathbf{U}$. In certain settings, suitable sub-Gaussian $\mathbf{U}$, e.g., sampled from the Bernoulli distribution, can lead to more efficient implementation compared to the Gaussian case. As shown in (Ai et al., 2014), some signals $\mathbf{x}^*$ cannot be recovered using non-Gaussian $\mathbf{U}$. In particular, the error bound for their estimator includes an irreducible term depending on $\|\mathbf{x}^*\|_\infty$ which does not decay with number of samples n.

We present an analysis of the error bound for our k-support norm estimator and sub-Gaussian $\mathbf{U}$. Our estimator also has an irreducible term, which can be shown to be smaller than twice of that based on $\|\mathbf{x}^*\|_\infty$ in (Ai et al., 2014). Moreover, in sharp contrast to existing sub-Gaussian analysis (Ai et al., 2014), the irreducible term for our estimator can be controlled by choosing an appropriate scaling of the sub-Gaussian measurement vectors, at the cost of increasing constants in terms which are reducible by increasing the number of samples. Thus, our estimator allows for a trade-off, which we illustrate empirically in Section 5.

For sub-Gaussian $\mathbf{U}$, the result in Lemma 2 for the expectation of $\mathbf{z}$ is not valid anymore. As a result, the subsequent analysis in Section 3 breaks down. Hence we start the sub-Gaussian analysis by defining

$$\mathbf{w} \triangleq \mathbf{E}\left[\frac{\mathbf{U}^T\mathbf{y}}{n}\right] = \mathbf{E}[y\mathbf{u}] = \mathbf{E}[\theta(\langle \mathbf{x}^*, \mathbf{u}\rangle)\mathbf{u}] , \qquad (8)$$

where $\mathbf{u}$ is a random vector with i.i.d. centered sub-Gaussian elements and $\mathbf{U}$ consists of n such independent $\mathbf{u}$ vectors. Note that $\mathbf{w}$ is deterministic when $\mathbf{x}^*$ is given. We also redefine the random vector

$$\mathbf{z} = \frac{1}{n}\frac{\mathbf{U}^T\mathbf{y}}{\|\mathbf{w}\|_2} . \qquad (9)$$

It is not difficult to see that if $\mathbf{u}$ is standard Gaussian random vector, then $\mathbf{w} = \lambda\mathbf{x}^*$ and $\mathbf{z}$ is reduced to the original definition in Lemma 2. In order to recover $\mathbf{x}^*$, we need to assume that $\mathbf{w} \neq \mathbf{0}$ for each $\mathbf{x}^*$. The recovery guarantee is provided in Theorem 5.

Theorem 5 *Given a k_0-sparse signal $\mathbf{x}^* \in \mathbb{R}^p$ with $\|\mathbf{x}^*\|_2 = 1$ and a measurement matrix $\mathbf{U}$ containing n i.i.d. samples, where each $\mathbf{u}_i$ consists of i.i.d. centered sub-Gaussian entries with $\|u_{ij}\|_{\psi_2} \leq K$ (See (Vershynin, 2012) for more on $\|\cdot\|_{\psi_2}$) , we choose a specific k in the convex program (3) to obtain $\hat{\mathbf{x}}$ (possibly $k \neq k_0$). Assume that k_1 nonzero entries of $\mathbf{x}^*$ are recovered ($k_1 < k_0$), and define*

$$\beta = \frac{\langle \mathbf{w}, \mathbf{x}^*\rangle}{\|\mathbf{w}\|_2\|\mathbf{x}^*\|_2}, \quad \tau = \max\left\{\frac{w_i}{\|\mathbf{w}\|_2} \;\middle|\; \hat{x}_i \neq 0\right\},$$
$$\xi = \max\left\{\frac{w_i}{\|\mathbf{w}\|_2} \;\middle|\; \hat{x}_i = 0\right\} .$$

Then the recovered signal $\hat{\mathbf{x}}$, with probability at least $1 - \eta e^{1-t}$, satisfies

$$\begin{aligned}\|\hat{\mathbf{x}} - \mathbf{x}^*\|_2 &\leq \frac{c_1 K}{\|\mathbf{w}\|_2}\sqrt{\frac{k(t+\log p)}{n}} + \frac{c_2K^2}{\|\mathbf{w}\|_2^2}\frac{k(t+\log k)}{n} \\ &\quad + \frac{c_3K\tau}{\|\mathbf{w}\|_2}\frac{k_1(t+\log k_1)}{\sqrt{n}} + \xi^2(k_0-k_1) \\ &\quad + \xi\sqrt{k_0-k_1} + \sqrt{2(1-\beta)} \\ &= O\Big(\sqrt{\frac{k\log p}{n}} + \frac{k_1\log k_1}{\sqrt{n}}\Big) + \xi^2(k_0-k_1) \\ &\quad + \xi\sqrt{k_0-k_1} + \sqrt{2(1-\beta)} ,\end{aligned}$$

where η, c_1, c_2 and c_3 are absolute constants.

Proof Sketch: We assume that $\hat{\mathbf{x}}$ has the same structure as in the proof of Theorem 3. Then the error satisfies

$$\begin{aligned}\|\hat{\mathbf{x}} - \mathbf{x}^*\|_2 &\leq \Big\|\hat{\mathbf{x}} - \frac{\mathbf{w}}{\|\mathbf{w}\|_2}\Big\|_2 + \Big\|\frac{\mathbf{w}}{\|\mathbf{w}\|_2} - \mathbf{x}^*\Big\|_2 \\ &\leq \Big\|\hat{\mathbf{x}}_{1:k} - \frac{\mathbf{w}_{1:k}}{\|\mathbf{w}\|_2}\Big\|_2 + \frac{\|\mathbf{w}_{k+1:p}\|_2}{\|\mathbf{w}\|_2} + \Big\|\frac{\mathbf{w}}{\|\mathbf{w}\|_2} - \mathbf{x}^*\Big\|_2 \\ &\leq \Big\|\mathbf{z}_{1:k} - \frac{\mathbf{w}_{1:k}}{\|\mathbf{w}\|_2}\Big\|_2 + \Big\|\frac{\mathbf{z}_{1:k}}{\|\mathbf{z}_{1:k}\|_2} - \mathbf{z}_{1:k}\Big\|_2 \\ &\quad + \frac{\|\mathbf{w}_{k+1:p}\|_2}{\|\mathbf{w}\|_2} + \Big\|\frac{\mathbf{w}}{\|\mathbf{w}\|_2} - \mathbf{x}^*\Big\|_2 \\ &\leq \sqrt{k}\Big\|\mathbf{z}_{1:k} - \frac{\mathbf{w}_{1:k}}{\|\mathbf{w}\|_2}\Big\|_\infty + \big|\|\mathbf{z}_{1:k}\|_2^2 - 1\big| \\ &\quad + \xi\sqrt{k_0-k_1} + \sqrt{2(1-\beta)} ,\end{aligned}$$

Using similar argument in the proof of Theorem 3, we have following inequalities simultaneously hold with probability at least $1 - \eta e^{1-t}$,

$$\begin{aligned}\Big\|\mathbf{z}_{1:k} - \frac{\mathbf{w}_{1:k}}{\|\mathbf{w}\|_2}\Big\|_\infty &\leq \frac{c_1K}{\|\mathbf{w}\|_2}\sqrt{\frac{t+\log p}{n}} , \\ \big|\|\mathbf{z}_{1:k}\|_2^2 - 1\big| &\leq \frac{c_2K^2}{\|\mathbf{w}\|_2^2}\frac{k(t+\log k)}{n} \\ &\quad + \frac{c_3K\tau}{\|\mathbf{w}\|_2}\frac{k_1(t+\log k_1)}{\sqrt{n}} + \xi^2(k_0-k_1) ,\end{aligned}$$

where η, c_1, c_2 and c_3 are absolute constants. Combining all the inequalities, we complete the proof. ■

Remark The new quantity which plays a role in the error bound is $\beta = \frac{\langle \mathbf{w}, \mathbf{x}^*\rangle}{\|\mathbf{w}\|_2\|\mathbf{x}^*\|_2}$. Clearly, $\beta \leq 1$.

Compared to the Gaussian case, the error bound involves an additional *irreducible* term $\sqrt{2(1-\beta)} = \|\frac{\mathbf{w}}{\|\mathbf{w}\|_2} - \mathbf{x}^*\|_2$, which does not decrease with increasing n. Further, $\|\mathbf{w}\|_2$ controls the convergence rate for the reducible terms, which decrease with n. Interestingly, for two special cases, viz $\mathbf{U}$ being standard Gaussian and nonzero entries in $\mathbf{x}^*$ being equal, it is easy to show that $\beta = 1$, thus eliminating the additional error.

The next result shows that the term $(1-\beta)$ is upper bounded by the corresponding irreducible term based on $\|\mathbf{x}^*\|_\infty$ in (Ai et al., 2014, Theorem 1.3).

Proposition 6 *Under the setting of Theorem 5, if u_{ij} is unit-variance and $\|u_{ij}\|_{\psi_2} \le K$ as assumed in (Ai et al., 2014, Theorem 1.3), and the second and third derivatives of θ are bounded by τ_2 and τ_3 respectively, then*

$$1-\beta \le \frac{2CK^4}{\lambda}(\tau_2+\tau_3)\|\mathbf{x}^*\|_\infty \ ,$$

where C is an absolute constant.

Proof: Note that $1-\beta \le (1-\beta)(1+\frac{\|\mathbf{w}\|_2}{\lambda}) \le \frac{1}{\lambda}(|\|\mathbf{w}\|_2 - \beta\lambda| + |\lambda - \beta\|\mathbf{w}\|_2|)$. Using the Berry-Esseen type central limit theorem (i.e. Lemma 3.1 in (Ai et al., 2014)), we obtain

$$\begin{aligned}
&\left|\mathbf{E}[\theta(\langle\mathbf{u},\mathbf{x}^*\rangle)\langle\mathbf{u},\frac{\mathbf{w}}{\|\mathbf{w}\|_2}\rangle] - \mathbf{E}[\theta(\langle\mathbf{g},\mathbf{x}^*\rangle)\langle\mathbf{g},\frac{\mathbf{w}}{\|\mathbf{w}\|_2}\rangle]\right| \\
&= \left|\|\mathbf{w}\|_2 - \beta\lambda\right| \le CK^4(\tau_2+\tau_3)\|\mathbf{x}^*\|_\infty \ , \\
&\left|\mathbf{E}[\theta(\langle\mathbf{u},\mathbf{x}^*\rangle)\langle\mathbf{u},\mathbf{x}^*\rangle] - \mathbf{E}[\theta(\langle\mathbf{g},\mathbf{x}^*\rangle)\langle\mathbf{g},\mathbf{x}^*\rangle]\right| \\
&= \left|\beta\|\mathbf{w}\|_2 - \lambda\right| \le CK^4(\tau_2+\tau_3)\|\mathbf{x}^*\|_\infty \ ,
\end{aligned}$$

in which $\mathbf{g}$ is the random vector with i.i.d standard Gaussian entries and we use the equality $\mathbf{E}[\theta(\langle\mathbf{g},\mathbf{x}^*\rangle)\mathbf{g}] = \lambda\mathbf{x}^*$ from Lemma 2. The result simply follows. ■

Though the error term $\sqrt{2(1-\beta)}$ can be bounded by twice of $\|\mathbf{x}^*\|_\infty$ term in (Ai et al., 2014, Theorem 1.3), the bound can be loose in general. Next we illustrate that under suitable scalings of the sub-Gaussian vectors, β is close to 1. We start with the following lemma.

Lemma 7 *Suppose that $\mathbf{u}$ is a random vector with i.i.d. centered sub-Gaussian entries of variance σ^2, then for any vector $\mathbf{x} \in \mathbb{R}^p$, $\mathbf{E}[\langle\mathbf{u},\mathbf{x}\rangle\mathbf{u}] = \sigma^2\mathbf{x}$.*

Proof: We expand the expectation for each entry, $\mathbf{E}[\langle\mathbf{u},\mathbf{x}\rangle u_i] = \mathbf{E}[u_i^2 x_i] + \sum_{j\ne i}\mathbf{E}[u_i u_j x_j] = \mathbf{E}[u_i^2]x_i + 0 = \sigma^2 x_i$. Hence the result holds for the vector. ■

By this lemma, if we can ensure that $\theta(\langle\mathbf{u}_i,\mathbf{x}^*\rangle)$ looks like a linear function in a suitable neighborhood, say around 0, then with proper scaling of the $\mathbf{u}_i$, we expect $\mathbf{w}$ to be approximately aligned with $\mathbf{x}^*$, thus $\beta \approx 1$. Specifically we have the following result.

Theorem 8 *Assume that θ is twice continuously differentiable, and the second-order derivative θ'' is bounded by ϕ. Let $\nu = \theta'(0)$, and $\mathbf{u}_i$ consists of i.i.d. centered sub-Gaussian of variance σ^2 , the following inequality holds for every unit k_0-sparse vector $\mathbf{x}^*$.*

$$\beta = \frac{\langle\mathbf{w},\mathbf{x}^*\rangle}{\|\mathbf{w}\|_2\|\mathbf{x}^*\|_2} \ge \frac{1-\alpha}{\sqrt{1+2\alpha+\frac{k_0}{C^6}\alpha^2}} \ ,$$

where $\alpha = \frac{3\sqrt{3}\phi C^3K^3}{2\sigma^2\nu}$, C is an absolute constant, and $K = \|u_{ij}\|_{\psi_2}$ is the sub-Gaussian norm.

Proof: We expand $\theta(\langle\mathbf{u}_i,\mathbf{x}^*\rangle)$ at 0 by Taylor expansion, $\theta(\langle\mathbf{u}_i,\mathbf{x}^*\rangle) = \theta(0) + \theta'(0)\langle\mathbf{u}_i,\mathbf{x}^*\rangle + \frac{\theta''(r)}{2}\langle\mathbf{u}_i,\mathbf{x}^*\rangle^2$, where r is between 0 and $\langle\mathbf{u}_i,\mathbf{x}^*\rangle$. Then we have

$$\begin{aligned}
&\mathbf{w} = \mathbf{E}[\theta(\langle\mathbf{u}_i,\mathbf{x}^*\rangle)\mathbf{u}_i] = \mathbf{E}[\theta(0)\mathbf{u}_i] + \mathbf{E}[\theta'(0)\langle\mathbf{u}_i,\mathbf{x}^*\rangle\mathbf{u}_i] \\
&\quad + \mathbf{E}[\frac{\theta''(r)}{2}\langle\mathbf{u}_i,\mathbf{x}^*\rangle^2\mathbf{u}_i] = \sigma^2\nu\mathbf{x}^* + \mathbf{E}[\frac{\theta''(r)}{2}\langle\mathbf{u}_i,\mathbf{x}^*\rangle^2\mathbf{u}_i], \\
&\langle\mathbf{w},\mathbf{x}^*\rangle = \langle\sigma^2\nu\mathbf{x}^* + \mathbf{E}[\frac{\theta''(r)}{2}\langle\mathbf{u}_i,\mathbf{x}^*\rangle^2\mathbf{u}_i],\ \mathbf{x}^*\rangle \\
&\quad = \sigma^2\nu + \mathbf{E}[\frac{\theta''(r)}{2}\langle\mathbf{u}_i,\mathbf{x}^*\rangle^3] \\
&\quad \ge \sigma^2\nu - \frac{\phi}{2}\mathbf{E}[|\langle\mathbf{u}_i,\mathbf{x}^*\rangle|^3] \ge \sigma^2\nu - \frac{3\sqrt{3}\phi}{2}C^3K^3 \ .
\end{aligned}$$

The last inequality follows the definition of sub-Gaussian norm and the fact that $\langle\mathbf{u}_i,\mathbf{x}^*\rangle$ is sub-Gaussian (Vershynin, 2012). Similarly we have

$$\begin{aligned}
&\|\mathbf{w}\|_2^2 = \langle\mathbf{w},\mathbf{w}\rangle = \big\langle\sigma^2\nu\mathbf{x}^* + \mathbf{E}[\frac{\theta''(r)}{2}\langle\mathbf{u}_i,\mathbf{x}^*\rangle^2\mathbf{u}_i], \\
&\quad \sigma^2\nu\mathbf{x}^* + \mathbf{E}[\frac{\theta''(r)}{2}\langle\mathbf{u}_i,\mathbf{x}^*\rangle^2\mathbf{u}_i]\big\rangle \\
&\quad \le \sigma^4\nu^2 + 3\sqrt{3}\phi\sigma^2\nu C^3K^3 + \big\|\mathbf{E}[\frac{\theta''(r)}{2}\langle\mathbf{u}_i,\mathbf{x}^*\rangle^2\mathbf{u}_i]\big\|_2^2 \ .
\end{aligned}$$

To bound the last ℓ_2-norm term, we first try to bound each entry in that vector. For simplicity, we assume that $\mathbf{x}^*_{1:k_0}$ are nonzero. For $j > k_0$, it is easy to see

$$\mathbf{E}[\frac{\theta''(r)}{2}\langle\mathbf{u}_i,\mathbf{x}^*\rangle^2 u_{ij}] = \mathbf{E}[\frac{\theta''(r)}{2}\langle\mathbf{u}_{i1:k_0},\mathbf{x}^*_{1:k_0}\rangle^2]\mathbf{E}[u_{ij}] = 0$$

For $1 \le j \le k_0$, we have

$$\begin{aligned}
&\left|\mathbf{E}[\frac{\theta''(r)}{2}\langle\mathbf{u}_i,\mathbf{x}^*\rangle^2 u_{ij}]\right| \le \mathbf{E}\big[\big|\frac{\theta''(r)}{2}\langle\mathbf{u}_i,\mathbf{x}^*\rangle^2 u_{ij}\big|\big] \\
&\quad \le \frac{\phi}{2}\mathbf{E}[\langle\mathbf{u}_i,\mathbf{x}^*\rangle^2|u_{ij}|] = \frac{\phi}{2}\mathbf{E}[\sum_{k,l=1}^{k_0} x_k x_l u_{ik}u_{il}\cdot|u_{ij}|] \\
&\quad = \frac{\phi}{2}\big(x_j^2\mathbf{E}[|u_{ij}|^3] + \sum_{k\ne j}^{k_0} x_k^2\mathbf{E}[u_{ik}^2]\mathbf{E}[|u_{ij}|]\big) \\
&\quad \le \frac{\phi}{2}\big(3\sqrt{3}x_j^2K^3 + \sum_{k\ne j}^{k_0} 2x_k^2K^3\big) \le \frac{3\sqrt{3}\phi}{2}K^3
\end{aligned}$$

$$\Longrightarrow \quad \|\mathbf{E}[\frac{\theta''(r)}{2}\langle \mathbf{u}_i, \mathbf{x}^*\rangle^2 \mathbf{u}_i]\|_2^2 \leq \frac{27\phi^2 k_0}{4} K^6 \ .$$

Combining the lower bound for $\langle \mathbf{w}, \mathbf{x}^* \rangle$ and the upper bound for $\|\mathbf{w}\|_2^2$, we get the inequality for β. ■

Remark Note that ϕ and ν are constants in this theorem, if we replace $\mathbf{u}_i$ with $\tilde{\mathbf{u}}_i = \frac{\mathbf{u}_i}{\gamma}$ for some large $\gamma > 0$, then the sub-Gaussian norm $\tilde{K} = \frac{K}{\gamma}$, variance $\tilde{\sigma} = \frac{\sigma}{\gamma}$, and thus $\tilde{\alpha} = \frac{\alpha}{\gamma}$ accordingly. As γ increases, $\tilde{\alpha}$ approaches 0, so the corresponding $\tilde{\beta}$ approaches 1.

We highlight the key differences between our bound for the irreducible error and that in (Ai et al., 2014). Essentially our irreducible term characterizes the ℓ_2 norm $\|\frac{\mathbf{w}}{\|\mathbf{w}\|_2} - \mathbf{x}^*\|_2$, while Ai et al. (2014) considers the quantity $|\langle \mathbf{w}, \hat{\mathbf{x}}\rangle - \lambda\langle \mathbf{x}^*, \hat{\mathbf{x}}\rangle|$ instead. For the bound in (Ai et al., 2014) to hold, the sub-Gaussian entry u_{ij} should be unit-variance. Hence it is impermissible to reduce it by scaling down $\mathbf{u}_i$. As for our result, the irreducible term can be made arbitrarily small by using $\frac{\mathbf{u}_i}{\gamma}$ with γ large enough. However, the scaling is not a free lunch since using $\frac{\mathbf{u}_i}{\gamma}$ with large γ decreases $\|\mathbf{w}\|_2$ which adversely affects the constant $\frac{1}{\|\mathbf{w}\|_2}$ for the reducible terms (decreasing with n) in Theorem 5.

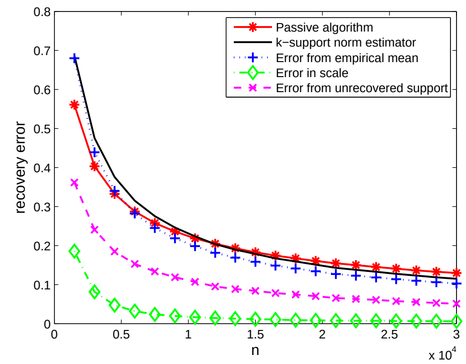
5 Experiments

In recent work, Zhang et al. (2014) illustrated that their passive algorithm outperforms other baselines. Hence, in the experiments, we directly compare our estimator against their passive algorithm. The regularization parameter γ of the passive algorithm is set to $\sqrt{\frac{\log p}{n}}$, which is the optimal choice used in (Zhang et al., 2014). All results are reported based on an average over 100 trials.

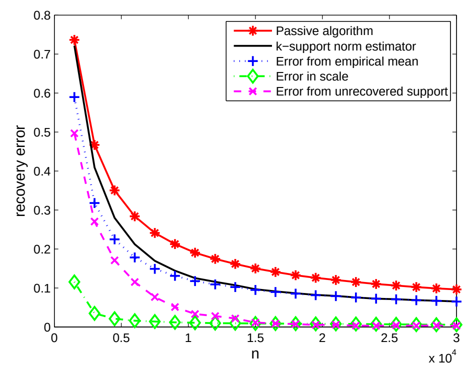
5.1 Gaussian Measurement Matrix

We use standard Gaussian $\mathbf{U}$, i.e., $u_{ij} \sim \mathcal{N}(0,1)$. The noise model is random bit-flip with probability 0.1, i.e., $y_i = \rho\,\text{sign}(\langle \mathbf{u}_i, \mathbf{x}^*\rangle)$, where ρ equals 1 with probability 0.9, -1 with probability 0.1. Apart from total error $\|\hat{\mathbf{x}} - \mathbf{x}^*\|_2$, we also investigate the behavior of the three types of error, i.e., error from empirical mean $\|\mathbf{z}_{1:k} - \mathbf{x}^*_{1:k}\|_2$, error in scale $\|\frac{\mathbf{z}_{1:k}}{\|\mathbf{z}_{1:k}\|_2} - \mathbf{z}_{1:k}\|_2$, and error from unrecovered support $\|\mathbf{x}^*_{k+1:p}\|_2$.

First we study the recovery error at different sample sizes n. In particular, we choose $k = k_0 = 50$, $p = 5000$, and vary n from 1500 to 30000. We focus on two different scenarios, $\mathbf{x}^*$ being uncorrelated or correlated. Correlated $\mathbf{x}^*$ tends to have some nonzero entries with similar magnitude, whereas the uncorrelated does not (see settings in Figure 1(a),1(b)). The error curves are shown in Figure 1. The performance



(a) Uncorrelated signal. Nonzero entries are sampled from $\mathcal{N}(0,1)$. Our estimator performs worse at small n, but quickly catch up when n grows.

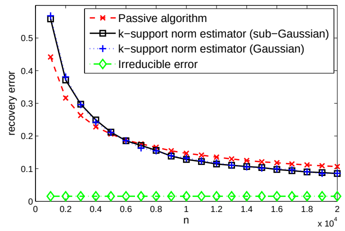


(b) Correlated signal. Half of the nonzero entries are sampled from $\mathcal{N}(5,1)$, half are from $\mathcal{N}(-4,1)$. Our estimator outperforms the passive algorithm.

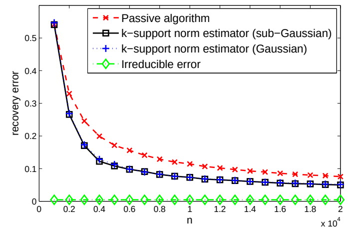
Figure 1: Recovery error vs. sample size n (Gaussian).

of the two estimators are comparable for the uncorrelated case, while the new estimator does better in the correlated case. Further, in the correlated case (Figure 1(b)), the error from unrecovered support decreases nearly to zero, which means that the support of $\mathbf{x}^*$ is well recovered in this situation and matches our analysis. Besides, we can see that the error from empirical mean always dominates the total and the error in scale plays little role in both cases.

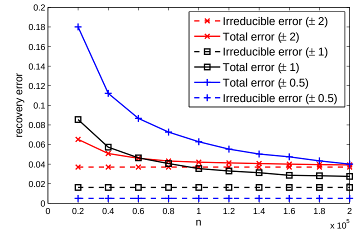
Next we study the error of our estimator under misspecified model, i.e., $k \neq k_0$ (Figure 2). We stay with correlated $\mathbf{x}^*$ and two scenarios, sample size being large or small. For large n (Figure 2(a)), the error sharply drops at the correct $k = k_0$, and our estimator performs better than the passive algorithm in a neighborhood of k_0. Under misspecification with $k < k_0$, the error is large since the error from unrecovered support is large. For $k > k_0$, the support is correctly recovered so that the corresponding error is small, but there is some additional error due to empirical mean. For small n (Figure 2(b)), our estimator outperforms the passive algorithm over a wide range of k, with the best performance being around $k = k_0$. The trend of



(a) Recovery error for uncorrelated signal. Nonzero entries are generated from $\mathcal{N}(0, 1)$.

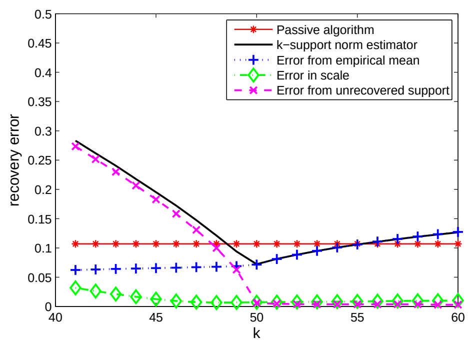


(b) Recovery error for correlated signal. Half of the nonzero entries are from $\mathcal{N}(5, 1)$, half are from $\mathcal{N}(-4, 1)$.

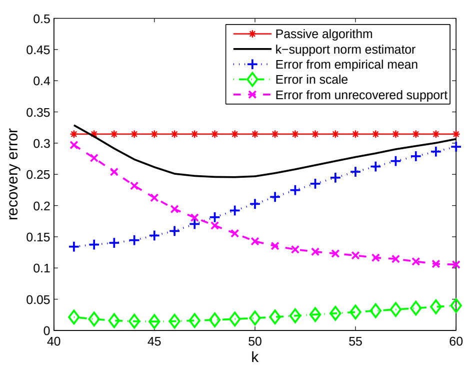


(c) Recovery errors for differently scaled Bernoulli distributions. Nonzero entries are generated from $\mathcal{N}(0, 1)$.

Figure 3: Recovery error for sub-Gaussian.



(a) Large sample size, $n = 25000$. Total error is minimized at $k = k_0$. $k < k_0$ incurs a large error from unrecovered support, and $k > k_0$ only slightly increases the error from empirical mean.



(b) Small sample size, $n = 5000$. The error from empirical mean is comparable to error from unrecovered support, and the curve of total error is smoother than the one for large sample size.

Figure 2: Recovery error vs. parameter k (Gaussian).

three types of error matches our theoretical bounds.

5.2 Sub-Gaussian Measurement Matrix

Here we specifically use a centered Bernoulli distribution to generate $\mathbf{U}$, i.e, each u_{ij} takes -1 or $+1$ with equal probability. We choose $k_0 = 10$, $p = 1000$ and set function $\theta(t) = \frac{e^t-1}{e^t+1}$, which corresponds to logistic regression model. The small k_0 facilitates the calculation of $\mathbf{w}$, whose complexity is basically $O(2^{k_0})$ for Bernoulli distribution. Thus we can compute the irreducible error term $\|\frac{\mathbf{w}}{\|\mathbf{w}\|_2} - \mathbf{x}^*\|_2$ using $\mathbf{w}$.

From Figures 3(a) and 3(b), we note similar error curves against sample size compared to those for Gaussian measurement matrix. In fact, the curve for Bernoulli measurements almost overlaps with the Gaussian one, and the irreducible error is negligible relative to the total. Again, our estimator outperforms the passive algorithm, especially for the correlated $\mathbf{x}^*$.

We also study how the scaling of sub-Gaussian distribution affects the recovery error. We choose three scaled Bernoulli distributions, $u_{ij} = \pm 2, \pm 1$ and ± 0.5. In Figure 3(c), $u_{ij} = \pm 2$ has largest irreducible error, while its reducible error has almost converged, yielding a moderate total error. In contrast, $u_{ij} = \pm 0.5$ has smallest irreducible error, but the reducible error has not yet converged in range of n we consider, resulting in a large total error. $u_{ij} = \pm 1$ balances the two error in a way and gives the smallest total error. These observations confirm our conclusions in Theorem 8, and the optimal scaling depends on the sub-Gaussian distribution, the noise model and the sample size.

6 Conclusions

In this paper, we first introduce the k-support norm estimator for 1-bit CS, which has a closed-form solution. Then we establish its recovery guarantees. For Gaussian measurement matrix, our result considers model misspecification and is comparable to the best known results for both ℓ_2 error and support recovery. In sub-Gaussian case, our bound ends up with an irreducible error similar to previous work. However, we show that this error can be controlled under certain assumptions by properly scaling the sub-Gaussian distribution. Experimental results provide sound support to our theoretical development.

Acknowledgements

The research was supported by NSF grants IIS-1447566, IIS-1422557, CCF-1451986, CNS-1314560, IIS-0953274, IIS-1029711, and by NASA grant NNX12AQ39A.

References

Albert Ai, Alex Lapanowski, Yaniv Plan, and Roman Vershynin. One-bit compressed sensing with non-gaussian measurements. *Linear Algebra and Applications*, 441:222–239, 2014.

Andreas Argyriou, Rina Foygel, and Nathan Srebro. Sparse prediction with the k-support norm. In *NIPS*, pages 1466–1474, 2012.

Petros T. Boufounos. Reconstruction of sparse signals from distorted randomized measurements. In *ICASSP*, 2010.

Petros T. Boufounos and Richard G. Baraniuk. 1-bit compressive sensing. In *CISS*, 2008.

Emmanuel J. Candes and Terrence Tao. Near-optimal signal recovery from random projections: Universal encoding strategies? *IEEE Transactions on Information Theory*, 52(12):5406–5425, December 2006. ISSN 0018-9448.

Soumyadeep Chatterjee, Sheng Chen, and Arindam Banerjee. Generalized dantzig selector: Application to the k-support norm. In *NIPS*, pages 1934–1942. 2014.

David L. Donoho. Compressed sensing. *IEEE Transactions on Information Theory*, 52:1289–1306, 2006.

Sivakant Gopi, Praneeth Netrapalli, Prateek Jain, and Aditya V. Nori. One-bit compressed sensing: Provable support and vector recovery. In *ICML*, pages 154–162, 2013.

Ankit Gupta, Robert Nowak, and Benjamin Recht. Sample complexity for 1-bit compressed sensing and sparse classification. In *ISIT*, pages 1553–1557, 2010.

Jarvis Haupt and Richard Baraniuk. Robust support recovery using sparse compressive sensing matrices. In *CISS*, pages 1–6, 2011.

Laurent Jacques, Jason N. Laska, Petros T. Boufounos, and Richard G. Baraniuk. Robust 1-bit compressive sensing via binary stable embeddings of sparse vectors. *IEEE Transactions on Information Theory*, 59(4):2082–2102, 2013.

Andrew M. McDonald, Massimiliano Pontil, and Dimitris Stamos. Spectral k-support norm regularization. In *NIPS*, pages 3644–3652. 2014.

Yaniv Plan and Roman Vershynin. One-bit compressed sensing by linear programming. *Communications on Pure and Applied Mathematics*, 66:1275–1297, 2013a.

Yaniv Plan and Roman Vershynin. Robust 1-bit Compressed Sensing and Sparse Logistic Regression: A Convex Programming Approach. *IEEE Transactions on Information Theory*, 59(1):482–494, January 2013b. ISSN 0018-9448.

Roman Vershynin. Introduction to the non-asymptotic analysis of random matrices. In *Compressed Sensing*, chapter 5, pages 210–268. Cambridge University Press, 2012.

Lijun Zhang, Jinfeng Yi, and Rong Jin. Efficient algorithms for robust one-bit compressive sensing. In *ICML*, pages 820–828, 2014.

Xinhua Zhang, Yaoliang Yu, and Dale Schuurmans. Polar operators for structured sparse estimation. In *NIPS*, pages 82–90, 2013.

Hui Zou and Trevor Hastie. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society, Series B*, 67:301–320, 2005.