Supplementary Material

Sample Complexity Bounds for 1-Bit Compressive Sensing and Binary Stable Embeddings with Generative Priors (ICML 2020)

Zhaoqiang Liu, Selwyn Gomes, Avtansh Tiwari, and Jonathan Scarlett

This document presents the proofs of several results from the main text. Cross-references without the prefix S refer to those in the main text, whereas cross-references with the prefix S refer to those given in this document. Numbered citations refer to the reference list at the end of this document.

S-1 Proof of Theorem 1 (Noiseless Upper Bound)

For fixed $\delta > 0$ and a positive integer l, let $M = M_0 \subseteq M_1 \subseteq \ldots \subseteq M_l$ be a chain of nets of $B_2^k(r)$ such that M_i is a $\frac{\delta_i}{L}$-net with $\delta_i = \frac{\delta}{2^i}$. There exists such a chain of nets with [1, Lemma 5.2]

$$\log |M_i| \le k \log \frac{4Lr}{\delta_i}. \tag{S-1}$$

By the L-Lipschitz assumption on G, we have for any $i \in [l]$ that $G(M_i)$ is a δ_i-net of $G(B_2^k(r))$.

For any pair of points $\mathbf{x}, \mathbf{s} \in G(B_2^k(r))$ with $\|\mathbf{x} - \mathbf{s}\|_2 > \epsilon$, we write

$$\mathbf{x} = (\mathbf{x} - \mathbf{x}_l) + (\mathbf{x}_l - \mathbf{x}_{l-1}) + \ldots + (\mathbf{x}_1 - \mathbf{x}_0) + \mathbf{x}_0, \tag{S-2}$$

$$\mathbf{s} = (\mathbf{s} - \mathbf{s}_l) + (\mathbf{s}_l - \mathbf{s}_{l-1}) + \ldots + (\mathbf{s}_1 - \mathbf{s}_0) + \mathbf{s}_0, \tag{S-3}$$

where $\mathbf{x}_i, \mathbf{s}_i \in G(M_i)$ for all $i \in [l]$, and $\|\mathbf{x} - \mathbf{x}_l\| \le \frac{\delta}{2^l}$, $\|\mathbf{s} - \mathbf{s}_l\| \le \frac{\delta}{2^l}$, $\|\mathbf{x}_i - \mathbf{x}_{i-1}\|_2 \le \frac{\delta}{2^{i-1}}$, and $\|\mathbf{s}_i - \mathbf{s}_{i-1}\|_2 \le \frac{\delta}{2^{i-1}}$ for all $i \in [l]$. Therefore, the triangle inequality gives

$$\|\mathbf{x} - \mathbf{x}_0\|_2 < 2\delta, \quad \|\mathbf{s} - \mathbf{s}_0\|_2 < 2\delta. \tag{S-4}$$

Let $\delta = c_1 \epsilon^2$ with $c_1 > 0$ being a sufficiently small constant. From (S-4), the triangle inequality, and $\|\mathbf{x} - \mathbf{s}\|_2 > \epsilon$, we obtain

$$\|\mathbf{x}_0 - \mathbf{s}_0\|_2 > \frac{\epsilon}{2}. \tag{S-5}$$

This separation between $\mathbf{x}_0$ and $\mathbf{s}_0$ permits the application of Lemma 1. Specifically, letting $\underline{\mathbf{a}}_i \in \mathbb{R}^n$ be the i-th row of $\mathbf{A}$, Lemma 1 (with $\frac{\epsilon}{2}$ in place of ϵ) implies for each $i \in [m]$ that

$$\mathbb{P}\left(\langle \underline{\mathbf{a}}_i, \mathbf{x}_0 \rangle > \frac{\epsilon}{24}, \langle \underline{\mathbf{a}}_i, \mathbf{s}_0 \rangle < -\frac{\epsilon}{24}\right) \ge \frac{\epsilon}{24}. \tag{S-6}$$

In accordance with the event inside the probability, and adopting the generic notation $(\mathbf{x}', \mathbf{s}') \in G(M) \times G(M)$ for an arbitrary pair in the net with $\|\mathbf{x}' - \mathbf{s}'\|_2 > \frac{\epsilon}{2}$, we define

$$\tilde{I}(\mathbf{x}', \mathbf{s}') := \left\{ i \in [m] \,:\, \langle \underline{\mathbf{a}}_i, \mathbf{x}' \rangle > \frac{\epsilon}{24}, \langle \underline{\mathbf{a}}_i, \mathbf{s}' \rangle < -\frac{\epsilon}{24} \right\}. \tag{S-7}$$

By (S-6) and a standard concentration inequality for binomial random variables [2, Theorem. A.1.13], we have

$$\mathbb{P}\left(|\tilde{I}(\mathbf{x}', \mathbf{s}')| < \frac{\epsilon m}{48}\right) \le e^{-\frac{\epsilon m}{192}}. \tag{S-8}$$

Recall from (S-1) that $\log|M| \le k\log\frac{4Lr}{\delta}$. By the union bound, for $m = \Omega\left(\frac{k}{\epsilon}\log\frac{Lr}{\delta}\right)$, we have with probability at least $1-\exp(-\Omega(\epsilon m))$ that for *all* $(\mathbf{x}', \mathbf{s}') \in G(M)\times G(M)$ with $\|\mathbf{x}'-\mathbf{s}'\|_2 > \frac{\epsilon}{2}$, the following holds:

$$|\tilde{I}(\mathbf{x}', \mathbf{s}')| \ge \frac{\epsilon m}{48}. \tag{S-9}$$

We now turn to bounding the following normalized summation:

$$\frac{1}{m}\sum_{i=1}^{m}|\langle\underline{\mathbf{a}}_i, \mathbf{x}-\mathbf{x}_0\rangle| \le \left(\frac{1}{m}\sum_{i=1}^{m}\langle\underline{\mathbf{a}}_i, \mathbf{x}-\mathbf{x}_0\rangle^2\right)^{1/2} \tag{S-10}$$

$$= \left\|\frac{1}{\sqrt{m}}\mathbf{A}(\mathbf{x}-\mathbf{x}_0)\right\|_2 \tag{S-11}$$

$$= \left\|\frac{1}{\sqrt{m}}\mathbf{A}\left(\sum_{i=1}^{l}(\mathbf{x}_i-\mathbf{x}_{i-1})\right) + \frac{1}{\sqrt{m}}\mathbf{A}(\mathbf{x}-\mathbf{x}_l)\right\|_2 \tag{S-12}$$

$$\le \sum_{i=1}^{l}\left\|\frac{1}{\sqrt{m}}\mathbf{A}(\mathbf{x}_i-\mathbf{x}_{i-1})\right\|_2 + \left\|\frac{1}{\sqrt{m}}\mathbf{A}(\mathbf{x}-\mathbf{x}_l)\right\|_2. \tag{S-13}$$

Using $\sqrt{1+\varepsilon} \le 1+\frac{\varepsilon}{2}$ (for $\varepsilon \ge -1$), Lemma 2, and the union bound, we have that for any $\epsilon_1, \dots, \epsilon_l \in (0,1)$, with probability at least $1-\sum_{i=1}^{l}|M_i|\times|M_{i-1}|\times e^{-\Omega(\epsilon_i^2 m)}$, the following holds for all $i \in [l]$:

$$\left\|\frac{1}{\sqrt{m}}\mathbf{A}(\mathbf{x}_i-\mathbf{x}_{i-1})\right\|_2 \le \left(1+\frac{\epsilon_i}{2}\right)\|\mathbf{x}_i-\mathbf{x}_{i-1}\|_2. \tag{S-14}$$

uniformly in $\mathbf{x}_i \in G(M_i)$ and $\mathbf{x}_{i-1} \in G(M_{i-1})$. In addition, (S-1) gives $\log\left(|M_i|\times|M_{i-1}|\right) \le 2ik + 2k\log\frac{4Lr}{\delta}$. As a result, if we choose the ϵ_i to satisfy $\epsilon_i^2 = \Theta(\epsilon + \frac{ik}{m})$, then we have

$$\sum_{i=1}^{l}|M_i|\times|M_{i-1}|\times e^{-\Omega(\epsilon_i^2 m)} \le e^{-\Omega(\epsilon m)}\sum_{i=1}^{l}e^{-c_2 ik} \tag{S-15}$$

$$= e^{-\Omega(\epsilon m)}, \tag{S-16}$$

where c_2 is a positive constant.

Recall that $m = \Omega\left(\frac{k}{\epsilon}\log\frac{Lr}{\delta}\right)$, and that we assume $L = \Omega\left(\frac{1}{r}\right)$ with a sufficiently large implied constant; these together imply $m = \Omega\left(\frac{k}{\epsilon}\right)$. Hence, we have

$$\sum_{i=1}^{l}\left\|\frac{1}{\sqrt{m}}\mathbf{A}(\mathbf{x}_i-\mathbf{x}_{i-1})\right\|_2 \le \sum_{i=1}^{l}\left(1+\frac{\epsilon_i}{2}\right)\|\mathbf{x}_i-\mathbf{x}_{i-1}\|_2 \tag{S-17}$$

$$\le \sum_{i=1}^{l}\left(1+\frac{\epsilon_i}{2}\right)\frac{\delta}{2^{i-1}} \tag{S-18}$$

$$\le \delta\sum_{i=1}^{l}\frac{\sqrt{\epsilon}}{2^{i-1}}\times O\left(\sqrt{1+\frac{ik}{m\epsilon}}\right) \tag{S-19}$$

$$= O(\sqrt{\epsilon}\delta) \tag{S-20}$$

$$= O(\delta), \tag{S-21}$$

where (S-17) follows from (S-14), (S-18) uses the definition of $\mathbf{x}_i$, (S-19) follows from the above choice of ϵ_i, and (S-20) from the above-established fact $m = \Omega\left(\frac{k}{\epsilon}\right)$, and (S-21) since we selected $\delta = c_1\epsilon^2$.

Recall that $\|\cdot\|_{2\to 2}$ is the spectral norm. By [1, Corollary 5.35], we have $\left\|\frac{1}{\sqrt{m}}\mathbf{A}\right\|_{2\to 2} \le 2+\sqrt{\frac{n}{m}}$ with probability at least $1-e^{-m/2}$. Hence, choosing $l = \lceil\log_2 n\rceil$, we have with probability at least $1-e^{-m/2}$ that

$$\left\|\frac{1}{\sqrt{m}}\mathbf{A}(\mathbf{x}-\mathbf{x}_l)\right\|_{2\to 2} \le \left(2+\sqrt{\frac{n}{m}}\right)\frac{\delta}{2^l} = O(\delta), \tag{S-22}$$

where we used the fact that $\|\mathbf{x}-\mathbf{x}_l\| \le \frac{\delta}{2^i}$.

Substituting (S-21) and (S-22) into (S-13), we deduce that with probability at least $1-e^{-\Omega(\epsilon m)}$,

$$\frac{1}{m}\sum_{i=1}^{m}|\langle \underline{\mathbf{a}}_i, \mathbf{x}-\mathbf{x}_0\rangle| \le c_3\delta, \tag{S-23}$$

where $c_3 > 0$ is a constant. Note that this holds *uniformly* in $\mathbf{x} \in G(B_2^k(r))$, since all preceding high-probability events only concerned signals in the chain $M_0, \dots, M_l$ of nets, and were proved uniformly with respect to those nets. Taking the inequality for both $\mathbf{x}$ and $\mathbf{s}$ and adding the two together, we obtain

$$\frac{1}{m}\sum_{i=1}^{m}|\langle \underline{\mathbf{a}}_i, \mathbf{x}-\mathbf{x}_0\rangle| + \frac{1}{m}\sum_{i=1}^{m}|\langle \underline{\mathbf{a}}_i, \mathbf{s}-\mathbf{s}_0\rangle| \le 2c_3\delta. \tag{S-24}$$

To combine the preceding findings, let $I_1 = \tilde{I}(\mathbf{x}_0, \mathbf{s}_0)$ (*cf.*, (S-7)), and

$$I_2 = \left\{ i \in [m] : |\langle \underline{\mathbf{a}}_i, \mathbf{x}-\mathbf{x}_0\rangle| + |\langle \underline{\mathbf{a}}_i, \mathbf{s}-\mathbf{s}_0\rangle| \le \frac{192 c_3 \delta}{\epsilon} \right\}. \tag{S-25}$$

By (S-9) and (S-24), we have that when $m = \Omega\left(\frac{k}{\epsilon}\log\frac{Lr}{\delta}\right) = \Omega\left(\frac{k}{\epsilon}\log\frac{Lr}{\epsilon^2}\right)$ (recalling the choice $\delta = c_1\epsilon^2$), with probability at least $1-\exp(-\Omega(\epsilon m))$,

$$|I_1| \ge \frac{\epsilon m}{48}, \quad |I_2^c| \le \frac{\epsilon m}{96}. \tag{S-26}$$

Defining $I := I_1 \cap I_2$, it follows that

$$|I| \ge |I_1| - |I_2^c| \ge \frac{\epsilon m}{96}. \tag{S-27}$$

In addition, for any $i \in I$, we have

$$\begin{aligned}
\langle \underline{\mathbf{a}}_i, \mathbf{x}\rangle &= \langle \underline{\mathbf{a}}_i, \mathbf{x}_0\rangle + \langle \underline{\mathbf{a}}_i, \mathbf{x}-\mathbf{x}_0\rangle && \text{(S-28)}\\
&> \frac{\epsilon}{24} - \frac{192 c_3\delta}{\epsilon} && \text{(S-29)}\\
&= \frac{\epsilon}{24} - 192 c_1 c_3 \epsilon && \text{(S-30)}\\
&> \frac{\epsilon}{25}, && \text{(S-31)}
\end{aligned}$$

where (S-30) holds because $\delta = c_1\epsilon^2$, and (S-31) follows by choosing c_1 sufficiently small. By a similar argument, we have for $i \in I$ that $\langle \underline{\mathbf{a}}_i, \mathbf{s}\rangle < -\frac{\epsilon}{25}$. Therefore, for any $i \in I$, we have $1 = \mathrm{sign}(\langle \underline{\mathbf{a}}_i, \mathbf{x}\rangle) \ne \mathrm{sign}(\langle \underline{\mathbf{a}}_i, \mathbf{s}\rangle) = -1$, and (S-27) gives

$$\mathrm{d_H}(\Phi(\mathbf{x}), \Phi(\mathbf{s})) \ge \frac{|I|}{m} \ge \frac{\epsilon}{96}, \tag{S-32}$$

which leads to the desired result in Theorem 1.

S-2 Proof of Corollary 2 (Supplementary Guarantee to Theorem 1)

Similar to Lemma 1, we have the following lemma.

Lemma S-1 *Let $\mathbf{x}, \mathbf{s} \in \mathcal{S}^{n-1}$ and assume that $\|\mathbf{x}-\mathbf{s}\|_2 \le \epsilon$ for some $\epsilon > 0$. If $\mathbf{a} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_n)$, then for $\epsilon_0 = \frac{\epsilon}{12}$, we have*

$$\mathbb{P}\Big((\langle \mathbf{a}, \mathbf{x}\rangle > \epsilon_0, \langle \mathbf{a}, \mathbf{s}\rangle > \epsilon_0) \cup (\langle \mathbf{a}, \mathbf{x}\rangle < -\epsilon_0, \langle \mathbf{a}, \mathbf{s}\rangle < -\epsilon_0) \Big) \ge 1 - \frac{2\epsilon}{3}. \tag{S-33}$$

Proof We have

$$\begin{aligned}
&\mathbb{P}\Big((\langle \mathbf{a},\mathbf{x}\rangle > \epsilon_0, \langle \mathbf{a},\mathbf{s}\rangle > \epsilon_0) \cup (\langle \mathbf{a},\mathbf{x}\rangle < -\epsilon_0, \langle \mathbf{a},\mathbf{s}\rangle < -\epsilon_0)\Big) \\
&= \mathbb{P}(\langle \mathbf{a},\mathbf{x}\rangle > \epsilon_0, \langle \mathbf{a},\mathbf{s}\rangle > \epsilon_0) + \mathbb{P}(\langle \mathbf{a},\mathbf{x}\rangle < -\epsilon_0, \langle \mathbf{a},\mathbf{s}\rangle < -\epsilon_0) && \text{(S-34)} \\
&\ge \mathbb{P}(\langle \mathbf{a},\mathbf{x}\rangle > 0, \langle \mathbf{a},\mathbf{s}\rangle > 0) - \mathbb{P}(|\langle \mathbf{a},\mathbf{x}\rangle| \le \epsilon_0) + \mathbb{P}(\langle \mathbf{a},\mathbf{x}\rangle < 0, \langle \mathbf{a},\mathbf{s}\rangle < 0) - \mathbb{P}(|\langle \mathbf{a},\mathbf{s}\rangle| \le \epsilon_0). && \text{(S-35)}
\end{aligned}$$

Note that by successively applying Lemmas 3 and 5, we have $\mathbb{P}(\langle \mathbf{a},\mathbf{x}\rangle > 0, \langle \mathbf{a},\mathbf{s}\rangle > 0) + \mathbb{P}(\langle \mathbf{a},\mathbf{x}\rangle < 0, \langle \mathbf{a},\mathbf{s}\rangle < 0) = 1 - \mathrm{d_S}(\mathbf{x},\mathbf{s}) \ge 1 - \frac{\epsilon}{2}$. In addition, because that $\langle \mathbf{a},\mathbf{x}\rangle \sim \mathcal{N}(0,1)$, we have

$$\mathbb{P}(|\langle \mathbf{a},\mathbf{x}\rangle| \le \epsilon_0) \le \epsilon_0 \sqrt{\frac{2}{\pi}}, \tag{S-36}$$

which is seen by trivially upper bounding the standard Gaussian density by $\frac{1}{\sqrt{2\pi}}$. Substituting $\epsilon_0 = \frac{\epsilon}{12}$, we obtain the desired inequality. ■

Using Lemma S-1 and following similar ideas to those in the proof of Theorem 1, we deduce Corollary 2. To avoid repetition, we provide only an outline below.

We again construct a chain of nets and select $\mathbf{x}_0, \mathbf{x}_1, \dots, \mathbf{x}_l$ and $\mathbf{s}_0, \mathbf{s}_1, \dots, \mathbf{s}_l$ in the nets such that (S-4) is satisfied. Let $\delta = c_1 \epsilon^2$ with $c_1 > 0$ being a sufficiently small constant. From the triangle inequality, we obtain

$$\|\mathbf{x}_0 - \mathbf{s}_0\|_2 \le \frac{3\epsilon}{2}. \tag{S-37}$$

Then, let

$$\tilde{J}(\mathbf{x}',\mathbf{s}') := \Big\{ i \in [m] \,:\, \Big(\langle \underline{\mathbf{a}}_i, \mathbf{x}'\rangle > \frac{\epsilon}{8}, \langle \underline{\mathbf{a}}_i, \mathbf{s}'\rangle > \frac{\epsilon}{8}\Big) \cup \Big(\langle \underline{\mathbf{a}}_i, \mathbf{x}'\rangle < -\frac{\epsilon}{8}, \langle \underline{\mathbf{a}}_i, \mathbf{s}'\rangle < -\frac{\epsilon}{8}\Big) \Big\}. \tag{S-38}$$

Similar to (S-9), we can show that when $m = \Omega\left(\frac{k}{\epsilon}\log\frac{Lr}{\delta}\right)$, with probability at least $1 - e^{-\Omega(\epsilon m)}$, for all $(\mathbf{x}',\mathbf{s}')$ pairs in $G(M) \times G(M)$ with $\|\mathbf{x}' - \mathbf{s}'\|_2 \le \frac{3\epsilon}{2}$, we have

$$|\tilde{J}(\mathbf{x}',\mathbf{s}')| \ge \left(1 - \frac{3\epsilon}{2}\right) m. \tag{S-39}$$

Combining (S-39) with (S-25) and a suitable analog of (S-26), we obtain the desired result.

S-3 Proof of Theorem 2 (Noiseless Lower Bound)

The proof proceeds in several steps, given in the following subsections.

S-3.1 Choice of Generative Model

Recall that Theorem 2 only states the existence of some generative model for which $m = \Omega\left(k\log(Lr) + \frac{k}{\epsilon}\right)$ measurements are necessary. Here we formally introduce the generative model, building on the approach from [3] of generating group-sparse signals. We say that a signal in $\mathbb{R}^n$ is *k-group-sparse* if, when divided into k blocks of size $\frac{n}{k}$,[1] each block contains at most one non-zero entry.[2]

We construct an auxiliary generative model $\tilde{G} : B_2^k(r) \to \mathbb{R}^n$, and then normalize it to obtain the final model $G : B_2^k(r) \to \mathcal{S}^{n-1}$. Noting that $B_\infty^k\big(\frac{r}{\sqrt{k}}\big) \subseteq B_2^k(r) \subseteq B_\infty^k(r)$, we fix $x_{\rm c}, x_{\max} > 0$ and construct $\tilde{G}$ as follows:

- The output vector $\mathbf{x} \in \mathbb{R}^n$ is divided into k blocks of length $\frac{n}{k}$, denoted by $\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(k)} \in \mathbb{R}^{\frac{n}{k}}$.
- A given block $\mathbf{x}^{(i)}$ is only a function of the corresponding input z_i, for $i = 1, \dots, k$.

[1] To simplify the notation, we assume that n is an integer multiple of k. For general values of n, the same analysis goes through by letting the final $n - k\lfloor \frac{n}{k} \rfloor$ entries of $\mathbf{x}$ always equal zero.

[2] More general notions of group sparsity exist, but for compactness we simply refer to this specific notion as k-group-sparse.

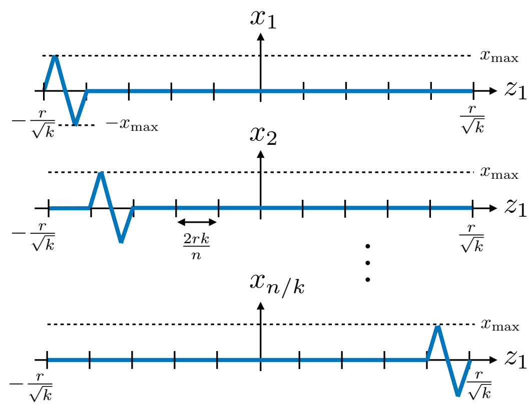

Figure S-1: Generative model that produces sparse signals. This figure shows the mapping from $z_1 \to (x_1, \dots, x_{\frac{n}{k}})$, and the same relation holds for $z_2 \to (x_{\frac{n}{k}+1}, \dots, x_{\frac{2n}{k}})$ and so on up to $z_{k-1} \to (x_{n-k+1-n/k}, \dots, x_{n-n/k})$.

- The mapping from z_i to $\mathbf{x}^{(i)}$, $i \in [k-1]$ is as shown in Figure S-1. The interval $\big[-\frac{r}{\sqrt{k}}, \frac{r}{\sqrt{k}}\big]$ is divided into $\frac{n}{k}$ intervals of length $\frac{2r\sqrt{k}}{n}$, and the j-th entry of $\mathbf{x}^{(i)}$ can only be non-zero if z_i takes a value in the j-th interval. Within that interval, the mapping takes a "double-triangular" shape with extremal values $-x_{\max}$ and $x_{\max}$.
- To handle the values of z_i (with $i \in [k-1]$) outside $\big[-\frac{r}{\sqrt{k}}, \frac{r}{\sqrt{k}}\big]$, we extend the functions in Figure S-1 to take values on the whole real line: For all values outside the indicated interval, each function value simply remains zero.
- Different from [3], we let the map corresponding to z_k always produce $x_{n-n/k+1} = x_{n-n/k+2} = \dots = x_{n-1} = 0$ and $x_n = x_{\mathrm{c}} > 0$, where the subscript 'c' is used to signify "constant". We allow $x_{\mathrm{c}} > x_{\max}$, as $x_{\max}$ only bounds the first $k-1$ non-zero entries.
- The preceding dot point leads to a Lipschitz-continuous function defined on all of $\mathbb{R}^k$, and we simply take $\tilde{G}$ to be that function restricted to $B_2^k(r)$.

To attain the final generative model used to prove Theorem 2, we take the output of $\tilde{G}$ and normalize it to be a unit vector: $G(\mathbf{z}) = \frac{\tilde{G}(\mathbf{z})}{\|\tilde{G}(\mathbf{z})\|_2}$. We define

$$\mathcal{X}_k := \big\{\mathbf{x} \in \mathcal{S}^{n-1} \,:\, \mathbf{x} \text{ is } k\text{-group-sparse}\big\}. \tag{S-40}$$

We observe the range $G(B_2^k(r))$ of G is[3]

$$\tilde{\mathcal{X}}_k := \left\{\mathbf{x} \in \mathcal{X}_k \,:\, x_n \geq \frac{x_{\mathrm{c}}}{\sqrt{(k-1)x_{\max}^2 + x_{\mathrm{c}}^2}}\right\}. \tag{S-41}$$

Furthermore, we have the following lemma regarding the Lipschitz continuity of G.

Lemma S-2 *The generative model $G : B_2^k(r) \to \mathcal{S}^{n-1}$ defined above, with parameters n, k, r, x_{c}, and $x_{\max}$, has a Lipschitz constant given by*

$$L = \frac{2nx_{\max}}{\sqrt{k}rx_{\mathrm{c}}}. \tag{S-42}$$

[3] For the extreme case that $x_{\mathrm{c}} = 0$, it is easy to see that $G(B_2^k(r)) = \mathcal{X}_k$ (ignoring the zero vector generated by $\tilde{G}$). It follows that for general $x_{\mathrm{c}} > 0$, the range of the generative model is as given in (S-41). Indeed, x_{c} gets divided by $\|\tilde{G}(\mathbf{z})\|_2$, which in turn can take any value between x_{c} and $\sqrt{(k-1)x_{\max}^2 + x_{\mathrm{c}}^2}$.

Proof From [3, Lemma 1], we know that $\tilde{G}$ is $\tilde{L}$-Lipschitz with $\tilde{L} = \frac{2nx_{\max}}{\sqrt{kr}}$. It is straightforward to show that for any $\mathbf{x}, \mathbf{x}' \neq \mathbf{0}$, $\left\| \frac{\mathbf{x}}{\|\mathbf{x}\|_2} - \frac{\mathbf{x}'}{\|\mathbf{x}'\|_2} \right\| \leq \max\left\{ \frac{1}{\|\mathbf{x}\|_2}, \frac{1}{\|\mathbf{x}'\|_2} \right\} \|\mathbf{x} - \mathbf{x}'\|_2$. Due to the choice of x_n in our construction, we have $\|\tilde{G}(\mathbf{z})\|_2 \geq x_{\mathrm{c}}$ for any $\mathbf{z} \in B_2^k(r)$; hence, for any $\mathbf{z}_1, \mathbf{z}_2 \in B_2^k(r)$, we have

$$\|G(\mathbf{z}_1) - G(\mathbf{z}_2)\|_2 = \left\| \frac{\tilde{G}(\mathbf{z}_1)}{\|\tilde{G}(\mathbf{z}_1)\|_2} - \frac{\tilde{G}(\mathbf{z}_2)}{\|\tilde{G}(\mathbf{z}_2)\|_2} \right\|_2 \tag{S-43}$$

$$\leq \frac{1}{x_{\mathrm{c}}} \|\tilde{G}(\mathbf{z}_1) - \tilde{G}(\mathbf{z}_2)\|_2 \tag{S-44}$$

$$\leq \frac{\tilde{L}}{x_{\mathrm{c}}} \|\mathbf{z}_1 - \mathbf{z}_2\|_2, \tag{S-45}$$

meaning that G is L-Lipschitz with $L = \frac{2nx_{\max}}{\sqrt{kr}x_{\mathrm{c}}}$. ■

S-3.2 Proof of $\Omega\left(\frac{k}{\epsilon}\right)$ Lower Bound

With the generative model in place that produces group-sparse signals, we proceed by following ideas from the 1-bit sparse recovery literature [4, 5]. The following lemma is a simple modification of a lower bound for the packing number of the unit sphere. The proof is deferred to Section S-3.4.

Lemma S-3 *For $\lambda \in (0,1)$, define*

$$Z_k(\lambda) := \{\mathbf{z} \in \mathcal{S}^{k-1} \,:\, z_k \geq \lambda\}. \tag{S-46}$$

Then, for any k and $\epsilon \in \left(0, \frac{1}{2}\right)$, there exists a subset $\mathcal{C} \subseteq Z_k(\frac{1}{2})$ of size $|\mathcal{C}| \geq \left(\frac{c}{\epsilon}\right)^k$ (with c being an absolute constant) such that for all $\mathbf{z}, \mathbf{z}' \in \mathcal{C}$, it holds that $\|\mathbf{z} - \mathbf{z}'\|_2 > 2\epsilon$.

The following lemma allows us to bound the number of distinct $\mathbf{b}$ vectors (observed vectors) that can be produced by sparse signals.

Lemma S-4 [4, Lemma. 8] *For $m \geq 2k$, the number of orthants intersected by a single k-dimensional subspace in an m-dimensional space is upper bounded by $2^k \binom{m}{k}$.*

With the above lemmas in place, we proceed by deriving a lower bound on the minimal worst-case reconstruction error, defined as follows (and implicitly depending on a fixed but arbitrary measurement matrix $\mathbf{A}$):

$$\epsilon_{\mathrm{opt}} := \inf_{\psi(\cdot)} \sup_{\mathbf{x} \in G(B_2^k(r))} \|\mathbf{x} - \psi(\mathbf{x})\|_2, \tag{S-47}$$

where $\psi(\cdot)$ is the overall mapping from $\mathbf{x}$ to its estimate $\hat{\mathbf{x}}$, and is therefore implicitly constrained to depend only on $(\mathbf{A}, \Phi(\mathbf{x}))$ with $\Phi(\mathbf{x}) = \mathrm{sign}(\mathbf{A}\mathbf{x})$. Note that our definition of ϵ_{opt} differs from that in [4], since we adopt a refined strategy more similar to [5] to arrive at $\epsilon_{\mathrm{opt}} = \Omega\left(\frac{k}{m}\right)$ instead of the weaker $\epsilon_{\mathrm{opt}} = \Omega\left(\frac{k}{m+k^{3/2}}\right)$.

Lemma S-5 *For the generative model G described above with x_{c} and $x_{\max}$ chosen to satisfy $(k-1)x_{\max}^2 = 3x_{\mathrm{c}}^2$, we have*

$$\epsilon_{\mathrm{opt}} = \Omega\left(\frac{k}{m}\right). \tag{S-48}$$

Proof Note that $G(B_2^k(r))$ corresponds to a union of $N_{\mathrm{supp}} = \left(\frac{n}{k}\right)^{k-1}$ subsets $\cup_{i \in [N_{\mathrm{supp}}]} S_i$, with

$$S_i := \left\{ \mathbf{x} \in \mathcal{X}_k \,:\, \mathrm{supp}(\mathbf{x}) \subseteq T_i, x_n \geq \frac{x_{\mathrm{c}}}{\sqrt{(k-1)x_{\max}^2 + x_{\mathrm{c}}^2}} \right\}, \tag{S-49}$$

where the sets $T_i \subseteq [n]$ equal the $N_{\rm supp}$ possible supports of size k for group sparse vectors. Substituting the assumption $(k-1)x_{\max}^2 = 3x_{\rm c}^2$ gives $\frac{x_{\rm c}}{\sqrt{(k-1)x_{\max}^2 + x_{\rm c}^2}} = \frac{1}{2}$, and it follows that for any $i^* \in [N_{\rm supp}]$, we have

$$\epsilon_{\rm opt} = \inf_{\psi(\cdot)} \sup_{\mathbf{x} \in G(B_2^k(r))} \|\mathbf{x} - \psi(\mathbf{x})\|_2 \tag{S-50}$$

$$\geq \inf_{\psi(\cdot)} \sup_{\mathbf{x} \in S_{i^*}(\frac{1}{2})} \|\mathbf{x} - \psi(\mathbf{x})\|_2, \tag{S-51}$$

where we write (S-49) as $S_{i^*}(\frac{1}{2}) := \{\mathbf{x} \in \mathcal{S}^{n-1} \cap \mathcal{X}_k \,:\, \mathrm{supp}(\mathbf{x}) \subseteq T_{i^*}, x_n \geq \frac{1}{2}\}$ to highlight the fact that $\frac{x_{\rm c}}{\sqrt{(k-1)x_{\max}^2 + x_{\rm c}^2}} = \frac{1}{2}$. Hence, it suffices to derive the lower bound for $\epsilon_{\rm opt}^* := \inf_{\psi(\cdot)} \sup_{\mathbf{x} \in S_{i^*}(\frac{1}{2})} \|\mathbf{x} - \psi(\mathbf{x})\|_2$.

To simplify notation, we assume in the following that the preceding infimum over $\psi(\cdot)$ is attained by some $\psi^*(\cdot)$.[4] By Lemma S-3, there exists a set $\mathcal{C} \subseteq S_{i^*}(\frac{1}{2})$, and a constant $c > 0$ such that $|\mathcal{C}| \geq \big(\frac{c}{\epsilon_{\rm opt}^*}\big)^k$, and for all $\mathbf{x}, \mathbf{s} \in \mathcal{C}$, $\|\mathbf{x} - \mathbf{s}\|_2 > 2\epsilon_{\rm opt}^*$. In addition, from Lemma S-4, the cardinality of the set $\widehat{\mathcal{X}}^* := \{\hat{\mathbf{x}} \in \mathbb{R}^n \,:\, \hat{\mathbf{x}} = \psi^*(\mathbf{x}) \text{ for some } \mathbf{x} \in S_{i^*}(\frac{1}{2})\}$ satisfies $|\widehat{\mathcal{X}}^*| \leq 2^k \binom{m}{k}$, since each distinct outcome $\mathbf{b} \in \{-1, 1\}^m$ produces at most one additional estimated vector.

For any $\mathbf{x} \neq \mathbf{s} \in \mathcal{C}$, we must have $\psi^*(\mathbf{x}) \neq \psi^*(\mathbf{s})$. To see this, suppose by contradiction that there exist $\mathbf{x} \neq \mathbf{s} \in \mathcal{C}$ such that $\psi^*(\mathbf{x}) = \psi^*(\mathbf{s})$. Because $\|(\mathbf{x} - \psi^*(\mathbf{x})) - (\mathbf{s} - \psi^*(\mathbf{s}))\|_2 = \|\mathbf{x} - \mathbf{s}\|_2 > 2\epsilon_{\rm opt}^*$, we have that at least one of $\|\mathbf{x} - \psi^*(\mathbf{x})\|_2$ and $\|\mathbf{s} - \psi^*(\mathbf{s})\|_2$ is larger than $\epsilon_{\rm opt}^*$, which contradicts the condition that $\sup_{\mathbf{x} \in S_{i^*}(\frac{1}{2})} \|\mathbf{x} - \psi^*(\mathbf{x})\|_2 \leq \epsilon_{\rm opt}^*$.

Hence, combining the above cardinality bounds, we find

$$2^k \binom{m}{k} \geq |\widehat{\mathcal{X}}^*| \geq |\mathcal{C}| \geq \left(\frac{c}{\epsilon_{\rm opt}^*}\right)^k, \tag{S-52}$$

and applying the inequality $\binom{m}{k} \leq \big(\frac{em}{k}\big)^k$, it follows that $\epsilon_{\rm opt}^* \geq \frac{ck}{2em}$ as desired. ■

Lemma S-5 implies that for any $\epsilon \in \big(0, \frac{1}{2}\big)$, to ensure that there exists a reconstruction function $\psi(\cdot)$ such that $\sup_{\mathbf{x} \in G(B_2^k(r))} \|\mathbf{x} - \psi(\mathbf{x})\|_2 \leq \epsilon$, we require that the number of samples m satisfies $m = \Omega\left(\frac{k}{\epsilon}\right)$.

S-3.3 Proof of $\Omega\left(k \log(Lr)\right)$ Lower Bound

The proof of the $m = \Omega\left(k \log(Lr)\right)$ lower bound follows a similar high-level approach to that of $m = \Omega\left(\frac{k}{\epsilon}\right)$. We first state the lower bound in terms of n as follows.

Lemma S-6 *For any $\epsilon \leq \frac{\sqrt{3}}{4\sqrt{2}}$ and any reconstruction function $\phi(\cdot)$, in order to attain the recovery guarantee $\sup_{\mathbf{x} \in G(B_2^k(r))} \|\mathbf{x} - \phi(\mathbf{x})\|_2 \leq \epsilon$, the number of samples m must satisfy $m = \Omega\left(k \log \frac{n}{k}\right)$.*

Proof Recall from (S-40) that $\mathcal{X}_k$ contains the k-group sparse signals on the unit sphere. For any $\lambda \in (0, 1)$, let

$$S(\lambda) := \{\mathbf{x} \in \mathcal{X}_k \,:\, x_n \geq \lambda\}. \tag{S-53}$$

We claim that for some constant $c > 0$ and any $\epsilon \leq \frac{\sqrt{3}}{4\sqrt{2}}$, there exists a subset $\mathcal{C} \subseteq S(\frac{1}{2})$ such that $\log |\mathcal{C}| \geq ck \log\left(\frac{n}{k}\right)$, and for all $\mathbf{x}, \mathbf{s} \in \mathcal{C}$, it holds that $\|\mathbf{x} - \mathbf{s}\|_2 > 2\epsilon$. To see this, consider the set

$$\mathcal{U} := \left\{\mathbf{x} \in \mathcal{X}_k \,:\, x_n = \frac{1}{2},\, x_i \in \left\{0, \sqrt{\frac{3}{4(k-1)}}\right\} \forall i \leq n-1, \|\mathbf{x}\|_0 = k\right\} \tag{S-54}$$

of group-sparse signals with exactly k non-zero entries, $k-1$ of which take the value $\sqrt{\frac{3}{4(k-1)}}$. By a simple counting argument, we have $|\mathcal{U}| = \left(\frac{n}{k}\right)^{k-1}$.

[4] If not, a similar argument applies with $\psi_\zeta^*(\cdot)$ satisfying $\sup_{\mathbf{x} \in S_{i^*}(\frac{1}{2})} \|\mathbf{x} - \psi^*(\mathbf{x})\|_2 \leq \epsilon_{\rm opt}^* + \zeta$ for an arbitrarily small ζ.

Let $k' = k-1$ for convenience, and for each $\mathbf{x} \in \mathcal{U}$, let $\mathbf{v} \in \{1, \dots, \frac{n}{k}\}^{k'}$ be a length-k' vector indicating which index in each block of the group-sparse signal (except the k-th one) is non-zero. Then, for $\mathbf{x}, \mathbf{x}' \in \mathcal{U}$ and the corresponding $\mathbf{v}, \mathbf{v}'$, we have

$$\|\mathbf{x} - \mathbf{x}'\|_2^2 = \frac{3}{4k'} \mathrm{d}'_{\mathrm{H}}(\mathbf{v}, \mathbf{v}'), \tag{S-55}$$

where $\mathrm{d}'_{\mathrm{H}}(\mathbf{v}, \mathbf{v}') = \sum_{i=1}^{n} \mathbf{1}\{v_i \neq v'_i\}$ is the unnormalized Hamming distance. By the Gilbert-Varshamov bound, we know that there exists a set $\mathcal{V}$ of signals in $\{1, \dots, \frac{n}{k}\}^{k'}$ whose pairwise unnormalized Hamming distance is at least d, and with the number of elements satisfying

$$|\mathcal{V}| \geq \frac{(\frac{n}{k'})^{k'}}{\sum_{j=0}^{d-1} (n/k - 1)^j} \tag{S-56}$$

$$\geq \frac{(\frac{n}{k'})^{k'}}{d (\frac{n}{k'})^d}. \tag{S-57}$$

Setting $d = \frac{k'}{2}$, we find that $\log |\mathcal{V}| = \Omega\big(k \log \frac{n}{k}\big)$, and by (S-55), we have that the corresponding $\mathbf{x}$ sequences are pairwise separated by at least a squared distance of $\frac{3}{8}$. This gives us the desired set $\mathcal{C}$ stated following (S-53).

By the triangle inequality, every $\mathbf{x} \in \mathcal{C}$ must have a different outcome $\Phi(\mathbf{x})$, since if two have the same outcome then their 2ϵ-separation (along with the triangle inequality) implies that the decoder's output cannot be ϵ-close to both. Since m binary measurements can result in 2^m possible outcomes, it follows that $2^m \geq |\mathcal{C}|$, and hence $m \geq \log_2 |\mathcal{C}| = \Omega\left(k \log \frac{n}{k}\right)$. ■

Combining the preceding two lower bounds, we readily deduce Theorem 2: From Lemma S-2, the generative model G that we used above has a Lipschitz constant given by

$$L = \frac{2n x_{\max}}{\sqrt{k} r x_{\mathrm{c}}} = \frac{n}{k} \frac{2\sqrt{k} x_{\max}}{r x_{\mathrm{c}}}, \tag{S-58}$$

which implies that when $(k-1)x_{\max}^2 = 3x_{\mathrm{c}}^2$, the condition $m = \Omega\left(k \log \frac{n}{k}\right)$ is equivalent to $m = \Omega\left(k \log(Lr)\right)$. Combining with the lower bound $\Omega\left(\frac{k}{\epsilon}\right)$ derived in Section S-3.2, we complete the proof of Theorem 2.

S-3.4 Proof of Lemma S-3 (Lower Bound on the Packing Number)

We first recall the following well-known lower bound on the packing number of the unit sphere.

Lemma S-7 [6, Ch. 13] *For any k and $\epsilon \in \left(0, \frac{1}{2}\right)$, there exists a subset $\mathcal{C} \subseteq \mathcal{S}^{k-1}$ of size $|\mathcal{C}| \geq \left(\frac{c}{\epsilon}\right)^k$ (with c being an absolute constant) such that for all $\mathbf{z}, \mathbf{z}' \in \mathcal{C}$, it holds that $\|\mathbf{z} - \mathbf{z}'\|_2 > 2\epsilon$.*

Recall that Lemma S-3 is stated for $\lambda = \frac{1}{2}$. Fix $\tilde{\lambda} \in [\frac{1}{2}, \frac{3}{4}]$, and consider the set $\mathcal{T}(\tilde{\lambda}) := \{\mathbf{z} \in \mathcal{S}^{k-1} \,:\, z_k = \tilde{\lambda}\}$. Applying Lemma S-7 to $\sqrt{1 - \tilde{\lambda}^2} \mathcal{S}^{k-2}$, we obtain that for any $\epsilon > 0$, there exists a subset $\mathcal{C}'(\tilde{\lambda}) \subseteq \mathcal{T}(\tilde{\lambda})$, and a constant $c'(\tilde{\lambda}) > 0$, such that $|\mathcal{C}'(\tilde{\lambda})| \geq \big(\frac{c'(\tilde{\lambda})}{\epsilon}\big)^{k-1}$, and for all $\mathbf{z}, \mathbf{z}' \in \mathcal{C}'(\tilde{\lambda})$, it holds that $\|\mathbf{z} - \mathbf{z}'\|_2 > 2\epsilon$. In addition, since we consider $\tilde{\lambda} \in [\frac{1}{2}, \frac{3}{4}]$, we have $\min_{\tilde{\lambda} \in [\frac{1}{2}, \frac{3}{4}]} c'(\tilde{\lambda}) > 0$.

For the final entry, observe that there exists a set $\mathcal{L} \subseteq [\frac{1}{2}, \frac{3}{4}]$ with $|\mathcal{L}| \geq \frac{1}{8\epsilon}$ such that for all $a, b \in \mathcal{L}$, it holds that $|a - b| > 2\epsilon$. Then, considering $\cup_{l \in [\mathcal{L}]} \mathcal{T}(l)$ and letting $\mathcal{C} := \cup_{l \in [\mathcal{L}]} \mathcal{C}'(l) \subseteq Z_k\big(\frac{1}{2}\big)$ (see (S-46)). We deduce that there exists a constant $c > 0$ such that $|\mathcal{C}| \geq \left(\frac{c}{\epsilon}\right)^k$, and for all $\mathbf{x}, \mathbf{s} \in \mathcal{C}$ it holds that $\|\mathbf{x} - \mathbf{s}\|_2 > 2\epsilon$. ■

References

[1] R. Vershynin, "Introduction to the non-asymptotic analysis of random matrices," *https://arxiv.org/abs/1011.3027*, 2010.

[2] N. Alon and J. H. Spencer, *The probabilistic method.* John Wiley & Sons, 2004.

[3] Z. Liu and J. Scarlett, "Information-theoretic lower bounds for compressive sensing with generative models," *IEEE J. Sel. Areas Inf. Theory*, vol. 1, no. 1, pp. 292–303, 2020.

[4] L. Jacques, J. N. Laska, P. T. Boufounos, and R. G. Baraniuk, "Robust 1-bit compressive sensing via binary stable embeddings of sparse vectors," *IEEE Trans. Inf. Theory*, vol. 59, no. 4, pp. 2082–2102, 2013.

[5] J. Acharya, A. Bhattacharyya, and P. Kamath, "Improved bounds for universal one-bit compressive sensing," in *Int. Symp. Inf. Theory (ISIT)*, pp. 2353–2357, 2017.

[6] G. G. Lorentz, M. von Golitschek, and Y. Makovoz, *Constructive approximation: Advanced problems*, vol. 304. Springer, 1996.