Supplementary Material for "Limits on Sparse Support Recovery via Linear Sketching with Random Expander Matrices"

(AISTATS 2016, Jonathan Scarlett and Volkan Cevher)

Note that all citations here are to the bibliography in the main document, and similarly for many of the cross-references.

A Proof of Lemma 1

In the notation of Definition 1, let $\mathcal{E}_\ell$ ($\ell = 1, \dots, k$) be the event that some set S of cardinality ℓ fails to satisfy the expansion property, i.e., $|\mathcal{N}_{\mathbf{X}}(S)| < (1-\epsilon)d|S|$. We start with the following non-asymptotic bound given in [8]:

$$\mathbb{P}[\mathcal{E}_\ell] \le \binom{p}{\ell}\binom{d\ell}{\epsilon d\ell}\Big(\frac{d\ell}{n}\Big)^{\epsilon d\ell}. \tag{44}$$

Applying the bounds $\log\binom{p}{\ell} \le \ell \log p$ and $\log\binom{d\ell}{\epsilon d\ell} \le d\ell H_2(\epsilon)$, we obtain

$$\log \mathbb{P}[\mathcal{E}_\ell] \le \ell \log p + d\ell H_2(\epsilon) + \epsilon d\ell \log \frac{d\ell}{n} \tag{45}$$

$$= \ell \log p - d\ell\Big(\epsilon \log \frac{n}{d\ell} - H_2(\epsilon)\Big). \tag{46}$$

Since $k = \Theta(1)$, we obtain from the union bound that $\mathbb{P}\big[\cup_{\ell=1,\dots,k} \mathcal{E}_s\big] \to 0$ provided that (46) tends to $-\infty$ for all ℓ. This is true provided that in (2) holds; the dominant condition is the one with $\ell = k$.

B Proof of Theorem 3

Recall the definitions of the random variables in (10)–(11), and the information densities in (25)–(27). We fix the constants $\gamma_1, \dots, \gamma_k$ arbitrarily, and consider a decoder that searches for the unique set $s \in \mathcal{S}$ such that

$$\tilde{\imath}(\mathbf{x}_{s_{\rm dif}}; \mathbf{y}|\mathbf{x}_{s_{\rm eq}}) > \gamma_{|s_{\rm dif}|} \tag{47}$$

for all $2^k - 1$ partitions $(s_{\rm dif}, s_{\rm eq})$ of s with $s_{\rm dif} \neq \emptyset$. If no such s exists, or if multiple exist, then an error is declared.

Since the joint distribution of $(\beta_s, \mathbf{X}_s, \mathbf{Y}_s \,|\, S = s)$ is the same for all s in our setup (cf., Section 1.2), and the decoder that we have chosen exhibits a similar symmetry, we can condition on $S = s = \{1, \dots, k\}$. By the union bound, the error probability is upper bounded by

$$P_{\rm e} \le \mathbb{P}\Big[\bigcup_{(s_{\rm dif}, s_{\rm eq})} \Big\{\tilde{\imath}(\mathbf{X}_{s_{\rm dif}}; \mathbf{Y}|\mathbf{X}_{s_{\rm eq}}) \le \gamma_{|s_{\rm dif}|}\Big\}\Big] + \sum_{\bar{s} \in \mathcal{S}\setminus\{s\}} \mathbb{P}\Big[\tilde{\imath}(\mathbf{X}_{\bar{s}\setminus s}; \mathbf{Y}|\mathbf{X}_{\bar{s}\cap s}) > \gamma_{|s_{\rm dif}|}\Big], \tag{48}$$

where here and subsequently we let the condition $s_{\rm dif} \neq \emptyset$ remain implicit. In the summand of the second term, we have upper bounded the probability of an intersection of $2^k - 1$ events by just one such event, namely, the one with the information density corresponding to $s_{\rm dif} = \bar{s}\setminus s$ and $s_{\rm eq} = s \cap \bar{s}$.

As mentioned previously, a key tool in the proof is the following change of measure (with $\ell := |s_{\rm dif}|$):

$$P_{\mathbf{Y}|\mathbf{X}_{s_{\rm eq}}}(\mathbf{y}|\mathbf{x}_{s_{\rm eq}}) = \sum_{\mathbf{x}_{s_{\rm dif}}} \Big(\prod_{i \in s_{\rm dif}} P_{\mathbf{X}_0}(\mathbf{x}_i)\Big) P_{\mathbf{Y}|\mathbf{X}_{s_{\rm dif}}\mathbf{X}_{s_{\rm eq}}}(\mathbf{y}|\mathbf{x}_{s_{\rm dif}}, \mathbf{x}_{s_{\rm eq}}) \tag{49}$$

$$\le (n+1)^\ell \sum_{\mathbf{x}_{s_{\rm dif}}} \Big(\prod_{i \in s_{\rm dif}} P_X^n(\mathbf{x}_i)\Big) P_{\mathbf{Y}|\mathbf{X}_{s_{\rm dif}}\mathbf{X}_{s_{\rm eq}}}(\mathbf{y}|\mathbf{x}_{s_{\rm dif}}, \mathbf{x}_{s_{\rm eq}}) \tag{50}$$

$$= (n+1)^\ell \widetilde{P}_{\mathbf{Y}|\mathbf{X}_{s_{\rm eq}}}(\mathbf{y}|\mathbf{x}_{s_{\rm eq}}), \tag{51}$$

where we have used the definitions in (23)–(24), and (50) follows from (12). By an identical argument, we have

$$P_{\mathbf{Y}|\mathbf{X}_{s_{\rm eq}}\beta_s}(\mathbf{y}|\mathbf{x}_{s_{\rm eq}}, b_s) \le (n+1)^\ell \widetilde{P}_{\mathbf{Y}|\mathbf{X}_{s_{\rm eq}}\beta_s}(\mathbf{y}|\mathbf{x}_{s_{\rm eq}}, b_s), \tag{52}$$

where $\widetilde{P}_{\mathbf{Y}|\mathbf{X}_{s_{\rm eq}}\beta_s} := P^n_{Y|X_{s_{\rm eq}}\beta_s}$ has an i.i.d. law.

We can weaken the second probability in (48) as follows (with $\ell := |\bar{s}\backslash s|$):

$$\mathbb{P}\Big[\tilde{\imath}(\mathbf{X}_{\bar{s}\backslash s};\mathbf{Y}|\mathbf{X}_{\bar{s}\cap s}) > \gamma_\ell\Big]$$

$$= \sum_{\mathbf{x}_{\bar{s}\cap s},\mathbf{x}_{\bar{s}\backslash s}} P^{k-\ell}_{\mathbf{X}_0}(\mathbf{x}_{\bar{s}\cap s}) P^{\ell}_{\mathbf{X}_0}(\mathbf{x}_{\bar{s}\backslash s}) \int_{\mathbb{R}^n} d\mathbf{y}\, P_{\mathbf{Y}|\mathbf{X}_{s_{\rm eq}}}(\mathbf{y}|\mathbf{x}_{\bar{s}\cap s}) \mathbb{1}\bigg\{\log\frac{P_{\mathbf{Y}|\mathbf{X}_{s_{\rm dif}}\mathbf{X}_{s_{\rm eq}}}(\mathbf{y}|\mathbf{x}_{\bar{s}\backslash s},\mathbf{x}_{\bar{s}\cap s})}{\widetilde{P}_{\mathbf{Y}|\mathbf{X}_{s_{\rm eq}}}(\mathbf{y}|\mathbf{x}_{\bar{s}\cap s})} > \gamma_\ell\bigg\} \tag{53}$$

$$\le (n+1)^\ell \sum_{\mathbf{x}_{\bar{s}\cap s},\mathbf{x}_{\bar{s}\backslash s}} P^{k-\ell}_{\mathbf{X}_0}(\mathbf{x}_{\bar{s}\cap s}) P^{\ell}_{\mathbf{X}_0}(\mathbf{x}_{\bar{s}\backslash s}) \int_{\mathbb{R}^n} d\mathbf{y}\, \widetilde{P}_{\mathbf{Y}|\mathbf{X}_{s_{\rm eq}}}(\mathbf{y}|\mathbf{x}_{\bar{s}\cap s}) \mathbb{1}\bigg\{\log\frac{P_{\mathbf{Y}|\mathbf{X}_{s_{\rm dif}}\mathbf{X}_{s_{\rm eq}}}(\mathbf{y}|\mathbf{x}_{\bar{s}\backslash s},\mathbf{x}_{\bar{s}\cap s})}{\widetilde{P}_{\mathbf{Y}|\mathbf{X}_{s_{\rm eq}}}(\mathbf{y}|\mathbf{x}_{\bar{s}\cap s})} > \gamma_\ell\bigg\} \tag{54}$$

$$\le (n+1)^\ell \sum_{\mathbf{x}_{\bar{s}\cap s},\mathbf{x}_{\bar{s}\backslash s}} P^{k-\ell}_{\mathbf{X}_0}(\mathbf{x}_{\bar{s}\cap s}) P^{\ell}_{\mathbf{X}_0}(\mathbf{x}_{\bar{s}\backslash s}) \int_{\mathbb{R}^n} d\mathbf{y}\, P_{\mathbf{Y}|\mathbf{X}_{s_{\rm dif}}\mathbf{X}_{s_{\rm eq}}}(\mathbf{y}|\mathbf{x}_{\bar{s}\backslash s},\mathbf{x}_{\bar{s}\cap s}) e^{-\gamma_\ell} \tag{55}$$

$$= (n+1)^\ell e^{-\gamma_\ell}, \tag{56}$$

where in (53) we used the fact that the output vector depends only on the columns of $\mathbf{x}_{\bar{s}}$ corresponding to entries of $\bar{s}$ that are also in s, (54) follows from (51), and (55) follows by bounding $\widetilde{P}_{\mathbf{Y}|\mathbf{X}_{s_{\rm eq}}}$ using the event within the indicator function, and then upper bounding the indicator function by one. Substituting (56) into (48) gives

$$P_{\rm e} \le \mathbb{P}\bigg[\bigcup_{(s_{\rm dif},s_{\rm eq})} \Big\{\tilde{\imath}(\mathbf{X}_{s_{\rm dif}};\mathbf{Y}|\mathbf{X}_{s_{\rm eq}}) \le \gamma_\ell\Big\}\bigg] + \sum_{\ell=1}^{k}\binom{p-k}{\ell}\binom{k}{\ell}(n+1)^\ell e^{-\gamma_\ell}, \tag{57}$$

where the combinatorial terms arise from a standard counting argument [7].

We now fix the constants $\gamma'_1,\dots,\gamma'_k$ arbitrarily, and recall the following steps from [17] (again writing $\ell := |s_{\rm dif}|$):

$$\mathbb{P}\bigg[\bigcup_{(s_{\rm dif},s_{\rm eq})} \Big\{\tilde{\imath}(\mathbf{X}_{s_{\rm dif}};\mathbf{Y}|\mathbf{X}_{s_{\rm eq}}) \le \gamma_\ell\Big\}\bigg]$$

$$= \mathbb{P}\bigg[\bigcup_{(s_{\rm dif},s_{\rm eq})} \bigg\{\log\frac{P_{\mathbf{Y}|\mathbf{X}_{s_{\rm dif}}\mathbf{X}_{s_{\rm eq}}}(\mathbf{Y}|\mathbf{X}_{s_{\rm dif}},\mathbf{X}_{s_{\rm eq}})}{\widetilde{P}_{\mathbf{Y}|\mathbf{X}_{s_{\rm eq}}}(\mathbf{Y}|\mathbf{X}_{s_{\rm eq}})} \le \gamma_\ell\bigg\}\bigg] \tag{58}$$

$$\le \mathbb{P}\bigg[\bigcup_{(s_{\rm dif},s_{\rm eq})} \bigg\{\log\frac{P_{\mathbf{Y}|\mathbf{X}_{s_{\rm dif}}\mathbf{X}_{s_{\rm eq}}}(\mathbf{Y}|\mathbf{X}_{s_{\rm dif}},\mathbf{X}_{s_{\rm eq}})}{\widetilde{P}_{\mathbf{Y}|\mathbf{X}_{s_{\rm eq}}}(\mathbf{Y}|\mathbf{X}_{s_{\rm eq}})} \le \gamma_\ell \cap \log\frac{\widetilde{P}_{\mathbf{Y}|\mathbf{X}_{s_{\rm eq}}}(\mathbf{Y}|\mathbf{X}_{s_{\rm eq}})}{\widetilde{P}_{\mathbf{Y}|\mathbf{X}_{s_{\rm eq}}\beta_s}(\mathbf{Y}|\mathbf{X}_{s_{\rm eq}},\beta_s)} \le \gamma'_\ell\bigg\}\bigg]$$
$$+ \mathbb{P}\bigg[\bigcup_{(s_{\rm dif},s_{\rm eq})} \bigg\{\log\frac{\widetilde{P}_{\mathbf{Y}|\mathbf{X}_{s_{\rm eq}}}(\mathbf{Y}|\mathbf{X}_{s_{\rm eq}})}{\widetilde{P}_{\mathbf{Y}|\mathbf{X}_{s_{\rm eq}}\beta_s}(\mathbf{Y}|\mathbf{X}_{s_{\rm eq}},\beta_s)} > \gamma'_\ell\bigg\}\bigg] \tag{59}$$

$$\le \mathbb{P}\bigg[\bigcup_{(s_{\rm dif},s_{\rm eq})} \bigg\{\log\frac{P_{\mathbf{Y}|\mathbf{X}_{s_{\rm dif}}\mathbf{X}_{s_{\rm eq}}}(\mathbf{Y}|\mathbf{X}_{s_{\rm dif}},\mathbf{X}_{s_{\rm eq}})}{\widetilde{P}_{\mathbf{Y}|\mathbf{X}_{s_{\rm eq}}\beta_s}(\mathbf{Y}|\mathbf{X}_{s_{\rm eq}},\beta_s)} \le \gamma_\ell + \gamma'_\ell\bigg\}\bigg]$$
$$+ \mathbb{P}\bigg[\bigcup_{(s_{\rm dif},s_{\rm eq})} \bigg\{\log\frac{\widetilde{P}_{\mathbf{Y}|\mathbf{X}_{s_{\rm eq}}}(\mathbf{Y}|\mathbf{X}_{s_{\rm eq}})}{\widetilde{P}_{\mathbf{Y}|\mathbf{X}_{s_{\rm eq}}\beta_s}(\mathbf{Y}|\mathbf{X}_{s_{\rm eq}},\beta_s)} > \gamma'_\ell\bigg\}\bigg]. \tag{60}$$

The second term in (60) is upper bounded as

$$\mathbb{P}\bigg[\bigcup_{(s_{\rm dif},s_{\rm eq})} \bigg\{\log\frac{\widetilde{P}_{\mathbf{Y}|\mathbf{X}_{s_{\rm eq}}}(\mathbf{Y}|\mathbf{X}_{s_{\rm eq}})}{\widetilde{P}_{\mathbf{Y}|\mathbf{X}_{s_{\rm eq}}\beta_s}(\mathbf{Y}|\mathbf{X}_{s_{\rm eq}},\beta_s)} > \gamma'_\ell\bigg\}\bigg]$$

$$\le \sum_{(s_{\rm dif},s_{\rm eq})} \mathbb{P}\bigg[\log\frac{\widetilde{P}_{\mathbf{Y}|\mathbf{X}_{s_{\rm eq}}}(\mathbf{Y}|\mathbf{X}_{s_{\rm eq}})}{\widetilde{P}_{\mathbf{Y}|\mathbf{X}_{s_{\rm eq}}\beta_s}(\mathbf{Y}|\mathbf{X}_{s_{\rm eq}},\beta_s)} > \gamma'_\ell\bigg] \tag{61}$$

$$= \sum_{(s_{\rm dif},s_{\rm eq})} \sum_{b_s,\mathbf{x}_{s_{\rm eq}}} P_{\beta_s}(b_s) P^{k-\ell}_{\mathbf{X}_0}(\mathbf{x}_{s_{\rm eq}}) \int_{\mathbb{R}^n} d\mathbf{y}\, P_{\mathbf{Y}|\mathbf{X}_{s_{\rm eq}}\beta_s}(\mathbf{y}|\mathbf{x}_{s_{\rm eq}},b_s) \mathbb{1}\bigg\{\log\frac{\widetilde{P}_{\mathbf{Y}|\mathbf{X}_{s_{\rm eq}}}(\mathbf{y}|\mathbf{x}_{s_{\rm eq}})}{\widetilde{P}_{\mathbf{Y}|\mathbf{X}_{s_{\rm eq}}\beta_s}(\mathbf{y}|\mathbf{x}_{s_{\rm eq}},b_s)} > \gamma'_\ell\bigg\} \tag{62}$$

$$\leq (n+1)^\ell \sum_{(s_{\rm dif},s_{\rm eq})} \sum_{b_s,\mathbf{x}_{s_{\rm eq}}} P_{\beta_s}(b_s) P_{\mathbf{X}_0}^{k-\ell}(\mathbf{x}_{s_{\rm eq}}) \int_{\mathbb{R}^n} d\mathbf{y}\, \widetilde{P}_{\mathbf{Y}|\mathbf{X}_{s_{\rm eq}}\beta_s}(\mathbf{y}|\mathbf{x}_{s_{\rm eq}},b_s) \mathbb{1}\Big\{ \log \frac{\widetilde{P}_{\mathbf{Y}|\mathbf{X}_{s_{\rm eq}}}(\mathbf{y}|\mathbf{x}_{s_{\rm eq}})}{\widetilde{P}_{\mathbf{Y}|\mathbf{X}_{s_{\rm eq}}\beta_s}(\mathbf{y}|\mathbf{x}_{s_{\rm eq}},b_s)} > \gamma_\ell' \Big\} \tag{63}$$

$$\leq (n+1)^\ell \sum_{(s_{\rm dif},s_{\rm eq})} \sum_{b_s,\mathbf{x}_{s_{\rm eq}}} P_{\beta_s}(b_s) P_{\mathbf{X}_0}^{k-\ell}(\mathbf{x}_{s_{\rm eq}}) \int_{\mathbb{R}^n} d\mathbf{y}\, \widetilde{P}_{\mathbf{Y}|\mathbf{X}_{s_{\rm eq}}}(\mathbf{y}|\mathbf{x}_{s_{\rm eq}}) e^{-\gamma_\ell'} \tag{64}$$

$$= (n+1)^\ell \sum_{\ell=1}^k \binom{k}{\ell} e^{-\gamma_\ell'}, \tag{65}$$

where (61) follows from the union bound, and the remaining steps follow the arguments used in (53)–(56) (with (52) used in place of (51)).

We now upper bound the first term in (60), again following [17]. The numerator in the first term in (60) equals $P_{\mathbf{Y}|\mathbf{X}_s}(\mathbf{Y}|\mathbf{X}_s)$ for all $(s_{\rm dif}, s_{\rm eq})$ (recall the definition in (22)), and we can thus write the overall term as

$$\mathbb{P}\Big[\log P_{\mathbf{Y}|\mathbf{X}_s}(\mathbf{Y}|\mathbf{X}_s) \leq \max_{(s_{\rm dif},s_{\rm eq})} \big\{ \log \widetilde{P}_{\mathbf{Y}|\mathbf{X}_{s_{\rm eq}}\beta_s}(\mathbf{Y}|\mathbf{X}_{s_{\rm eq}},\beta_s) + \gamma_\ell + \gamma_\ell' \big\} \Big]. \tag{66}$$

Using the same steps as those used in (58)–(60), we can upper bound this by

$$\mathbb{P}\Big[\log P_{\mathbf{Y}|\mathbf{X}_s\beta_s}(\mathbf{Y}|\mathbf{X}_s,\beta_s) \leq \max_{(s_{\rm dif},s_{\rm eq})} \big\{ \log \widetilde{P}_{\mathbf{Y}|\mathbf{X}_{s_{\rm eq}}\beta_s}(\mathbf{Y}|\mathbf{X}_{s_{\rm eq}},\beta_s) + \gamma_\ell + \gamma_\ell' + \gamma \big\} \Big] + \mathbb{P}\Big[\log \frac{P_{\mathbf{Y}|\mathbf{X}_s,\beta_s}(\mathbf{Y}|\mathbf{X}_s,\beta_s)}{P_{\mathbf{Y}|\mathbf{X}_s}(\mathbf{Y}|\mathbf{X}_s)} > \gamma \Big] \tag{67}$$

for any constant γ. Reversing the step in (66), this can equivalently be written as

$$\mathbb{P}\Big[\bigcup_{(s_{\rm dif},s_{\rm eq})} \Big\{ \log \frac{P_{\mathbf{Y}|\mathbf{X}_{s_{\rm dif}}\mathbf{X}_{s_{\rm eq}}\beta_s}(\mathbf{Y}|\mathbf{X}_{s_{\rm dif}},\mathbf{X}_{s_{\rm eq}},\beta_s)}{\widetilde{P}_{\mathbf{Y}|\mathbf{X}_{s_{\rm eq}}\beta_s}(\mathbf{Y}|\mathbf{X}_{s_{\rm eq}},\beta_s)} \leq \gamma_\ell + \gamma_\ell' + \gamma \Big\} \Big] + \mathbb{P}\Big[\log \frac{P_{\mathbf{Y}|\mathbf{X}_s,\beta_s}(\mathbf{Y}|\mathbf{X}_s,\beta_s)}{P_{\mathbf{Y}|\mathbf{X}_s}(\mathbf{Y}|\mathbf{X}_s)} > \gamma \Big]. \tag{68}$$

The first logarithm in the first term is the information density in (26). Moreover, the choices

$$\gamma_\ell = \log\Big(\frac{k}{\delta_1} \binom{p-k}{\ell} \binom{k}{\ell} (n+1)^\ell \Big) \tag{69}$$

$$\gamma_\ell' = \log\Big(\frac{k}{\delta_1} \binom{k}{\ell} (n+1)^\ell \Big) \tag{70}$$

make (65) and the second term in (57) be upper bounded by δ_1 each. Hence, and combining (60) with (65) and (68), and recalling that $\ell = |s_{\rm dif}|$, we obtain (28).

C Proof of Theorem 2

Fix $0 < b_{\min} < b_{\max} < \infty$, and let $\mathcal{B}_0 := \{b_s : \min_i |b_i| \geq b_{\min} \cap \max_i |b_i| \leq b_{\max}\}$. The main step in proving Theorem 2 is in extending the arguments of Section 4.5 to show that

$$P_e \leq \mathbb{P}\Big[n \leq \max_{(s_{\rm dif},s_{\rm eq}) \,:\, s_{\rm dif} \neq \emptyset} \frac{|s_{\rm dif}| \log p}{I_{s_{\rm dif},s_{\rm eq}}(\beta_s)} (1+\eta) \cap \beta_s \in \mathcal{B}_0 \Big] + \mathbb{P}\big[\beta_s \notin \mathcal{B}_0\big] + o(1), \tag{71}$$

and

$$P_e \geq \mathbb{P}\Big[n \leq \max_{(s_{\rm dif},s_{\rm eq}) \,:\, s_{\rm dif} \neq \emptyset} \frac{|s_{\rm dif}| \log p}{I_{s_{\rm dif},s_{\rm eq}}(\beta_s)} (1-\eta) \cap \beta_s \in \mathcal{B}_0 \Big] + o(1), \tag{72}$$

Before proving these, we show how they yield the theorem. Using (16), it is readily verified that each $I_{s_{\rm dif},s_{\rm eq}}(\beta_s)$, with an i.i.d. Gaussian vector β_s, is a continuous random variable having no mass points. By taking $\eta \to 0$ sufficiently slowly and noting that we have restricted β_s to the set $\mathcal{B}_0$ (within which all of the $I_{s_{\rm dif},s_{\rm eq}}(\beta_s)$ are

bounded away from zero and infinity), we conclude that (71)–(72) remain true when η is replaced by zero, and its contribution is factored into the $o(1)$ terms. Hence, we obtain Theorem 2 by (i) dropping the condition $\beta_s \in \mathcal{B}_0$ from the first probability in (71); (ii) using the identity $\mathbb{P}[\mathcal{A}_1 \cap \mathcal{A}_2] \geq \mathbb{P}[\mathcal{A}_1] - \mathbb{P}[\mathcal{A}_2]$ to remove the same condition from the first probability in (72); (iii) noting that the remainder term $\mathbb{P}\big[\beta_s \notin \mathcal{B}_0\big]$ can be made arbitrarily small by choosing $b_{\min}$ sufficiently small and $b_{\max}$ sufficiently large.

It remains to establish (71)–(72). Recall the value of κ given following Lemma 3. The above choice of $\mathcal{B}_0$ ensures that all of the non-zero entries are bounded away from 0 and ∞, so that the mutual informations $I_{s_{\text{dif}},s_{\text{eq}}}(\beta_s)$ and variances $V_{s_{\text{dif}},s_{\text{eq}}}(\beta_s)$ are bounded away from zero and infinity, and hence $\kappa = \Theta(1)$.

Since P_{β_s} is continuous, we must choose γ and handle P_0 in (29) differently to the above. Similarly to the analysis of Gaussian measurements in [17], we fix $\delta_0 > 0$ and note that Chebyshev's inequality implies

$$\gamma = I_0 + \sqrt{\frac{V_0}{\delta_0}} \implies P_0(\gamma) \leq \delta_0, \tag{73}$$

where

$$I_0 := I(\beta_s; \mathbf{Y}|\mathbf{X}_s) \tag{74}$$

$$V_0 := \text{Var}\bigg[\log \frac{P_{\mathbf{Y}|\mathbf{X}_s,\beta_s}(\mathbf{Y}|\mathbf{X}_s,\beta_s)}{P_{\mathbf{Y}|\mathbf{X}_s}(\mathbf{Y}|\mathbf{X}_s)}\bigg]. \tag{75}$$

The following is a straightforward extension of [17, Prop. 4] to expander-based measurements.

Proposition 1. *The quantities I_0 and V_0 defined in (74)–(75) satisfy*

$$I_0 \leq \frac{k}{2}\log\bigg(1 + \frac{d\sigma_\beta^2}{\sigma^2}\bigg) \tag{76}$$

$$V_0 \leq 2n. \tag{77}$$

Proof. See Appendix E. □

We can now obtain (71)–(72) using the steps of the previous subsection; the condition $\mathbb{P}[\beta_s \in \mathcal{B}_0]$ arises in (35) and (39) due to the fact that this condition was used to obtain a bounded variance in (32), and the first two probabilities in (71) arise from the identity $\mathbb{P}[\mathcal{A}_1 \cup \mathcal{A}_2] \leq \mathbb{P}[\mathcal{A}_1 \cup \mathcal{A}_2^c] + \mathbb{P}[\mathcal{A}_2]$. The only additional step is in showing that we can simultaneously achieve $\gamma = o(\log p)$ and $P_0(\gamma) = o(1)$ in the achievability part whenever $n = \Theta(\log p)$, in the same way that we showed $2|s_{\text{dif}}|\log n = o(\log p)$ in the previous subsection. This immediately follows by substituting (76)–(77) into (73) (along with $d = O(n) = O(\log p)$) to obtain $\gamma = O(\log\log p) + \sqrt{\log p} = o(\log p)$ for any $\delta_0 > 0$, and noting that δ_0 (and hence $P_0(\gamma)$) in (73) can be arbitrarily small.

D Proof of Lemma 3

We prove the lemma by characterizing the variance of a general function of $(\mathbf{X}_s, \mathbf{Y})$ of the form $f^n(\mathbf{X}_s, \mathbf{Y}) := \sum_{i=1}^n f(X_s^{(i)}, Y^{(i)})$. Clearly all of the quantities $\imath^n$ for the various $(s_{\text{dif}}, s_{\text{eq}})$ can be written in this general form. We have

$$\text{Var}\big[f^n(\mathbf{X}_s, \mathbf{Y})\big] = \text{Var}\bigg[\sum_{i=1}^n f(X_s^{(i)}, Y^{(i)})\bigg] \tag{78}$$

$$= \sum_{i=1}^n \sum_{j=1}^n \text{Cov}\Big[f(X_s^{(i)}, Y^{(i)}), f(X_s^{(j)}, Y^{(j)})\Big] \tag{79}$$

$$= n\text{Var}\Big[f(X_s, Y)\Big] + (n^2 - n)\text{Cov}\Big[f(X_s, Y), f(X_s', Y')\Big], \tag{80}$$

where (X_s, Y) and (X_s', Y') correspond to two different indices in $\{1, \cdots, n\}$; here (80) follows by simple symmetry considerations for the cases $i = j$ and $i \neq j$.

To compute the covariance term in (80), we first find the joint distribution of (X_s, Y) and (X'_s, Y'). As noted in [29, Sec. IV-B], a uniform permutation of a vector with d ones and $n-d$ zeros can be interpreted as successively performing uniform sampling from a collection of symbols without replacement (n times in total), where the initial collection contains d ones and $n-d$ zeros. By considering the first two steps of this procedure, we obtain

$$\mathbb{P}[X_i = x_i] = P_X(x_i) \tag{81}$$

$$\mathbb{P}[X'_i = x'_i | X_i = x_i] = \frac{nP_X(x'_i) - \mathbb{1}\{x_i = x'_i\}}{n-1} \tag{82}$$

for $\nu = 1, 2$, where $P_X(1) = 1 - P_X(0) = \frac{d}{n}$. Denoting the right-hand side of (82) by $P'_X(x'_i|x_i)$, and writing $\mu_f := \mathbb{E}[f(X_s, Y)]$, the covariance in (80) is given by

$$\begin{aligned}
&\mathrm{Cov}\Big[f(X_s, Y), f(X'_s, Y')\Big] \\
&\quad = \mathbb{E}\Big[\big(f(X_s, Y) - \mu_f\big)\big(f(X'_s, Y') - \mu_f\big)\Big] \qquad (83) \\
&\quad = \sum_{x_s} P_X^k(x_s) \sum_{x'_s} \bigg(\prod_{i \in s} P'_X(x'_i|x_i)\bigg) \mathbb{E}\Big[\big(f(x_s, Y) - \mu_f\big)\big(f(x'_s, Y') - \mu_f\big) \,\Big|\, X_s = x_s, X'_s = x'_s\Big]. \qquad (84)
\end{aligned}$$

We now consider the various terms arising by substituting (82) into (84) and performing a binomial-type expansion of the product:

- There is a single term of the form (84) with each $P'_x(x'_i|x_i)$ replaced by $\frac{nP_X(x'_i)}{n-1}$. This yields an average of $\big(f(X_s, Y) - \mu_f\big)\big(f(X'_s, Y') - \mu_f\big)$ over *independent* random variables X_s and X'_s, and therefore evaluates to zero.
- There are k terms in which one value $P'_x(x'_i|x_i)$ in (84) is replaced by $\frac{-\mathbb{1}\{x_i = x'_i\}}{n-1}$ and the other $k-1$ are replaced by $\frac{nP_X(x'_i)}{n-1}$. Each such term can be written as $-\frac{n}{(n-1)^2}\mathrm{Var}\big[\mathbb{E}[f(X_s, Y) \,|\, X_{s\setminus\{i\}}]\big]$, which in turn behaves as $-\frac{1}{n}\mathrm{Var}\big[\mathbb{E}[f(X_s, Y) \,|\, X_{s\setminus\{i\}}]\big] + O(1)$.
- All of the remaining terms replace $P'_x(x'_i|x_i)$ in (84) by $\frac{-\mathbb{1}\{x_i = x'_i\}}{n-1}$ for at least two values of i. All such terms are easily verified to behave as $O\big(\frac{1}{n^2}\big)$, and the number of such terms is finite and does not scale with n (recall that k is fixed by assumption).

Substituting these cases into (84) and recalling that $k = \Theta(1)$ and $\frac{d}{n} = \Theta(1)$, we obtain (40).

E Proof of Proposition 1

Here we characterize I_0 and V_0, defined in (74)–(75), via an extension of the analysis given in [17, App. B]. Since $\mathbf{Y} = \mathbf{X}_s\beta_s + \mathbf{Z}$, we have

$$\begin{aligned}
I_0 = I(\beta_s; \mathbf{Y}|\mathbf{X}_s) &= H(\mathbf{Y}|\mathbf{X}_s) - H(\mathbf{Y}|\mathbf{X}_s, \beta_s) \qquad (85) \\
&= H(\mathbf{X}_s\beta_s + \mathbf{Z}|\mathbf{X}_s) - H(\mathbf{Z}). \qquad (86)
\end{aligned}$$

From [25, Ch. 9], we have $H(\mathbf{Z}) = \frac{n}{2}\log(2\pi e\sigma^2)$ and $H(\mathbf{X}_s\beta_s + \mathbf{Z}|\mathbf{X}_s = \mathbf{x}_s) = \frac{1}{2}\log\big((2\pi e)^n \det(\sigma^2\mathbf{I}_n + \sigma_\beta^2\mathbf{x}_s\mathbf{x}_s^T)\big)$, where $\mathbf{I}_n$ is the $n \times n$ identity matrix. Averaging the latter over $\mathbf{X}_s$ and substituting these into (86) gives

$$\begin{aligned}
I_0 &= \frac{1}{2}\mathbb{E}\Big[\log\det\Big(\mathbf{I}_n + \frac{\sigma_\beta^2}{\sigma^2}\mathbf{X}_s\mathbf{X}_s^T\Big)\Big] \qquad (87) \\
&= \frac{1}{2}\mathbb{E}\Big[\log\det\Big(\mathbf{I}_k + \frac{\sigma_\beta^2}{\sigma^2}\mathbf{X}_s^T\mathbf{X}_s\Big)\Big] \qquad (88) \\
&= \frac{1}{2}\sum_{i=1}^{k}\mathbb{E}\Big[\log\Big(1 + \frac{\sigma_\beta^2}{\sigma^2}\lambda_i(\mathbf{X}_s^T\mathbf{X}_s)\Big)\Big] \qquad (89) \\
&\le \frac{k}{2}\log\Big(1 + \frac{d\sigma_\beta^2}{\sigma^2}\Big), \qquad (90)
\end{aligned}$$

where (88) follows from the identity $\det(\mathbf{I}+\mathbf{AB}) = \det(\mathbf{I}+\mathbf{BA})$, (89) follows by writing the determinant as a product of eigenvalues (denoted by $\lambda_i(\cdot)$), and (90) follows from Jensen's inequality and the following calculation:

$$\frac{1}{k}\mathbb{E}\Big[\sum_{i=1}^{k}\lambda_i(\mathbf{X}_s^T\mathbf{X}_s)\Big] = \frac{1}{k}\mathbb{E}[\mathrm{Tr}(\mathbf{X}_s^T\mathbf{X}_s)] = \mathbb{E}[\mathbf{X}_1^T\mathbf{X}_1] = d, \tag{91}$$

since the squared norm of $\mathbf{X}_1$ is d almost surely. This concludes the proof of (76).

We now turn to the bounding of the variance. Again using the fact that $\mathbf{Y} = \mathbf{X}_s\beta_s + \mathbf{Z}$, we have

$$\log\frac{P_{\mathbf{Y}|\mathbf{X}_s,\beta_s}(\mathbf{Y}|\mathbf{X}_s,\beta_s)}{P_{\mathbf{Y}|\mathbf{X}_s}(\mathbf{Y}|\mathbf{X}_s)} = \log\frac{P_{\mathbf{Z}}(\mathbf{Z})}{P_{\mathbf{Y}|\mathbf{X}_s}(\mathbf{X}_s\beta_s+\mathbf{Z}|\mathbf{X}_s)} \tag{92}$$

$$= I_0 - \frac{1}{2\sigma^2}\mathbf{Z}^T\mathbf{Z} + \frac{1}{2}(\mathbf{X}_s\beta_s+\mathbf{Z})^T\big(\sigma^2\mathbf{I}+\sigma_\beta^2\mathbf{X}_s\mathbf{X}_s^T\big)^{-1}(\mathbf{X}_s\beta_s+\mathbf{Z}), \tag{93}$$

where $P_{\mathbf{Z}}$ is the density of $\mathbf{Z}$, and (93) follows by a direct substitution of the densities $P_{\mathbf{Z}} \sim N(\mathbf{0},\sigma^2\mathbf{I})$ and $P_{\mathbf{Y}|\mathbf{X}_S}(\cdot|\mathbf{x}_s) \sim N(\mathbf{0},\sigma^2\mathbf{I}+\sigma_\beta^2\mathbf{x}_s\mathbf{x}_s^T)$. Observe now that $\frac{1}{\sigma^2}\mathbf{Z}^T\mathbf{Z}$ is a sum of n independent χ^2 random variables with one degree of freedom (each having a variance of 2), and hence the second term in (93) has a variance of $\frac{n}{2}$. Moreover, by writing $\mathbf{M}^{-1} = (\mathbf{M}^{-\frac{1}{2}})^T\mathbf{M}^{-\frac{1}{2}}$ for the symmetric positive definite matrix $\mathbf{M} = \sigma^2\mathbf{I}+\sigma_\beta^2\mathbf{X}_s\mathbf{X}_s^T$, we similarly observe that the final term in (93) is a sum of χ^2 variables (this is true conditioned on any $\mathbf{X}_s = \mathbf{x}_s$, and hence also true unconditionally), again yielding a variance of $\frac{n}{2}$. We thus obtain (77) using the identity $\mathrm{Var}[A+B] \le \mathrm{Var}[A] + \mathrm{Var}[B] + 2\max\{\mathrm{Var}[A],\mathrm{Var}[B]\}$.