A. Proof of Theorem 3.2

We begin by defining the discretized sparse point set for $\ell = O(k/\varepsilon^4)$:

Definition A.1 (Discretized sparse vectors)**.**

$$\widehat{\mathsf{Sp}}_\ell^m = \left\{\mathbf{x} : \mathbf{x} \in \left\{0, \pm\frac{1}{\sqrt{\ell}}\right\}^m, \|\mathbf{x}\|_0 = \ell\right\}$$

The intent here is to get the sparse point set $\widehat{\mathsf{Sp}}_\ell^m$ *distorted* on projection, so that it forms an ε-cover of the unit sphere on the smaller dimension. However, doing so rules out proofs that rely on simple union bound arguments. For instance, on allowing the projections to become distorted, we run into the risk of lots of points collapsing together into a small fraction of $\mathcal{S}^{n-1}$. As a result, the set $\Phi^{\mathrm{norm}}(\widehat{\mathsf{Sp}}_\ell^m)$ could turn out to be insufficient for forming a cover of the unit sphere. These issues are avoided by carefully relating the gaussian width of $\Phi^{\mathrm{norm}}(\widehat{\mathsf{Sp}}_\ell^m)$ to that of $\mathcal{S}^{n-1}$, followed by a partitioning argument. The partitioning crucially uses the block structure of elements in $\widehat{\mathsf{Sp}}_\ell^m$, which results in independent and distributionally identical blocks, allowing us to take union bounds effectively.

Proof. The proof of this Theorem proceeds in steps. We first partition Φ into L blocks of Nn columns each, for some appropriately chosen N. So, $\Phi = [\phi_1 \cdots \phi_L]$. Note that ϕ_i is a $n \times Nn$ submatrix of $\Phi \in \mathbb{R}^{n\times m}$. Write $\phi_i^{\mathrm{norm}}(\widehat{\mathsf{Sp}}_\ell)$ to denote $\phi_i^{\mathrm{norm}}(\widehat{\mathsf{Sp}}_\ell^{Nn})$. Also, note that $\phi_i^{\mathrm{norm}}(\widehat{\mathsf{Sp}}_\ell)$ is equal to the set obtained by applying Φ to vectors in $\widehat{\mathsf{Sp}}_\ell$ whose support is contained in P, where $P \subset [m]$ is the set of Nn columns of Φ that are present in ϕ_i.

For any such fixed partition ϕ_i, we show that the restriction $\phi_i^{\mathrm{norm}}(\widehat{\mathsf{Sp}}_\ell)$ has large expected gaussian width (Lemma A.2), where again, $\phi_i^{\mathrm{norm}}(\mathbf{x}) = \phi_i(\mathbf{x})/\|\phi_i(\mathbf{x})\|$. Furthermore, using Lemma A.3 we argue that any fixed point on $\mathcal{S}^{n-1}$ has distance $O(\varepsilon^{1/4})$ to $\phi_i(\widehat{\mathsf{Sp}}_\ell)$ with large probability. Now, we use the independence of the ϕ_i's to argue that the probability of $\mathbf{x} \in \mathcal{S}^{n-1}$ being simultaneously far away from all $\phi_i^{\mathrm{norm}}(\widehat{\mathsf{Sp}}_\ell)$ is exponentially small. Finally, taking a union bound over the ε-net of $\mathcal{S}^{n-1}$ completes the proof.

We fix the parameter N to be $N = \sqrt{\frac{m}{k}}\left(\log\frac{m}{k}\right)^{-1}$. Set the number of blocks $L = \frac{m}{Nn} = O(\varepsilon^2\sqrt{m/k})$ blocks of Nn coordinates each. By construction, for any fixed $i \in [L]$, $\phi_i \sim \frac{1}{\sqrt{n}}N(0,1)^{n\times Nn}$. The following lemma allows gives us a lower bound on the gaussian width of each projection:

Lemma A.2. *Let $\phi \sim \frac{1}{\sqrt{n}}\mathcal{N}(0,1)^{n\times Nn}$. Then for $\widehat{\mathsf{Sp}}_\ell \subset \mathcal{S}^{Nn-1}$,*

$$\mathop{\mathbf{E}}_{\phi}\left[\omega(\phi^{\mathrm{norm}}(\widehat{\mathsf{Sp}}_\ell))\right] \geqslant (1-4\varepsilon)\sqrt{n} \tag{6}$$

From Lemma A.2, we show the following lower bound on the gaussian width of the projections of $\widehat{\mathsf{Sp}}_\ell^{Nn-1}$: $\mathbb{E}_{\phi_i}\left[\omega(\phi_i^{\mathrm{norm}}(\widehat{\mathsf{Sp}}_\ell))\right] \geqslant (1-4\varepsilon)\sqrt{n}$.

Now, we argue that because $\phi_i^{\mathrm{norm}}(\widehat{\mathsf{Sp}}_\ell)$ has gaussian width close to $\sqrt{n}$, it in fact covers any fixed point on $\mathcal{S}^{n-1}$ with high probability. We begin by stating the following lemma on concentration of minimum distance with respect to large gaussian width sets.

Lemma A.3. *Let $T \subset \mathcal{S}^{n-1}$ be such that $\omega(T) \geqslant \sqrt{n}(1-\varepsilon)$, where $n = O\left(\frac{k}{\varepsilon^2}\log\frac{m}{k}\right)$. Let $R : \mathbb{R}^n \mapsto \mathbb{R}^n$ be a uniform random rotation operator. Then for any $\mathbf{x} \in \mathcal{S}^{n-1}$*

$$\mathop{\mathbf{Pr}}_{R}\left[\min_{\mathbf{y}\in R(T)}\|\mathbf{x}-\mathbf{y}\|_2 \geqslant 2\varepsilon^{1/4}\right] \leqslant \exp\left(-O\Big(\frac{k}{\varepsilon}\log\frac{d}{k}\Big)\right)$$

Observe that the distribution of ϕ_i^{norm} is rotation-invariant; see B.4 for a formal proof. Fixing $\mathbf{x} \in \mathcal{S}^{n-1}$, we invoke Lemma A.3 in our setting:

$$\mathop{\mathbf{Pr}}_{\phi_i}\left[\min_{\mathbf{y}\in\phi_i^{\mathrm{norm}}(\widehat{\mathsf{Sp}}_\ell)}\|\mathbf{y}-\mathbf{x}\|_2 > 16\varepsilon^{\frac{1}{4}}\right] \leqslant \exp\left(-O\Big(\frac{k}{\varepsilon}\log\frac{m}{k}\Big)\right) \tag{7}$$

Let P_ε be an ε-cover of $\mathcal{S}^{n-1}$ such that $|P_\varepsilon| = O(1/\varepsilon)^n$. Then:

$$\begin{aligned}
&\Pr_{\Phi}\left[\exists \mathbf{x} \in P_\varepsilon : \min_{\mathbf{y}\in\Phi^{\mathrm{norm}}(\widehat{\mathsf{Sp}}_\ell)} \|\mathbf{y}-\mathbf{x}\|_2 \geqslant 16\varepsilon^{\frac{1}{4}}\right] \\
&\leqslant |P_\varepsilon| \Pr_{\Phi}\left[\forall \quad i \in [L] \min_{\mathbf{y}\in\phi_i^{\mathrm{norm}}(\widehat{\mathsf{Sp}}_\ell)} \|\mathbf{y}-\mathbf{x}\|_2 \geqslant 16\varepsilon^{\frac{1}{4}}\right] \\
&\overset{1}{=} |P_\varepsilon| \prod_{i=1}^{L} \Pr_{\phi_i}\left[\min_{\mathbf{y}\in\phi_i^{\mathrm{norm}}(\widehat{\mathsf{Sp}}_\ell)} \|\mathbf{y}-\mathbf{x}\|_2 \geqslant 16\varepsilon^{\frac{1}{4}}\right] \\
&= |P_\varepsilon| \left(\Pr_{\phi_i}\left[\min_{\mathbf{y}\in\phi_i^{\mathrm{norm}}(\widehat{\mathsf{Sp}}_\ell)} \|\mathbf{y}-\mathbf{x}\|_2 \geqslant 16\varepsilon^{\frac{1}{4}}\right]\right)^L \\
&\leqslant |P_\varepsilon| \exp\Big(-O\Big(\frac{k}{\varepsilon}\log\frac{m}{k}\Big)\Big)^{\varepsilon^2 O(\sqrt{m/k})} \\
&\leqslant \exp\Big(\Big(\log\frac{C'}{\varepsilon}\Big)\Big(\frac{k}{\varepsilon^2}\log\frac{m}{k}\Big) - k\varepsilon\sqrt{\frac{m}{k}}\log\frac{m}{k}\Big) \\
&\leqslant \exp\Big(-O\Big(k\log\frac{m}{k}\Big)\Big)
\end{aligned}$$

where step 1 uses the independence of the ϕ_i's and the last step uses the fact that $\varepsilon \geqslant \left(\frac{k}{m}\right)^{\frac{1}{8}}$. The above inequality states that $\Phi^{\mathrm{norm}}(\widehat{\mathsf{Sp}}_\ell)$ is an $\big(16\varepsilon^{1/4}+\varepsilon\big)$-cover of $\mathcal{S}^{n-1}$ with positive probability, which completes the proof.

$\square$

B. Auxiliary Lemmas for the proof of Theorem 3.2

In this section, we prove additional Lemmas used in the proof of Theorem 3.2.

B.1. Proof of Lemma A.2

The proof of the lemma proceeds in two steps: we first restrict to the case where the maximum $\|\cdot\|_2$ length of the projected vectors is not much larger than the expected value. This is done by using Gordon's theorem and Lipschitz concentration for gaussians (see Theorem 2.3 and Corollary 2.5). Following that, we observe that conditioning the expectation by this high probability event on the length of the projected vectors does not affect the expectation by much (Lemma B.1). The rest of the proof follows using standard estimates of gaussian widths of $\widehat{\mathsf{Sp}}_\ell$ (see Lemma B.2).

Upper Bound on the $\|\cdot\|_2$ length: By setting $D = Nn$ in Theorem 2.3, we upper bound the expected maximum $\|\cdot\|_2$ length :

$$\mathop{\mathbf{E}}_{\phi}\left[\max_{\mathbf{x}\in\widehat{\mathsf{Sp}}_\ell} \|\phi(\mathbf{x})\|_2\right] \leqslant 1 + \frac{\omega(\widehat{\mathsf{Sp}}_\ell)}{\sqrt{n}} \tag{8}$$

Furthermore, from Lemma B.2, we have $\omega(\widehat{\mathsf{Sp}}_\ell) = \sqrt{C_0 \ell \log\frac{Nn}{\ell}}$ for some constant $C_0 > 0$. Therefore by our choice of parameters we have:

$$\frac{\omega(\widehat{\mathsf{Sp}}_\ell)}{\sqrt{n}} = \Big(\frac{Ck}{\varepsilon^2}\log\frac{m}{k}\Big)^{-\frac{1}{2}}\sqrt{C_0\ell\log\frac{Nn}{\ell}} = \frac{C'}{\varepsilon} \tag{9}$$

Note that Eq. 9 holds with an equality for some constant C'. Now consider the event $\mathcal{E}$ where the maximum $\|\cdot\|_2$ length is at most $1 + \frac{C'}{\varepsilon}(1+\varepsilon)$. Then by using gaussian concentration for Lipschitz functions (Corollary 2.5), we upper bound probability of the event $\neg\mathcal{E}$:

$$\begin{aligned}\Pr_{\phi\sim\frac{1}{\sqrt{n}}N(0,1)^{n\times Nn}}\left(\max_{\mathbf{x}\in\widehat{\mathsf{Sp}}_\ell}\|\phi(\mathbf{x})\|_2\geqslant 1+\frac{C'}{\varepsilon}(1+\varepsilon)\right) &\leqslant \exp\Big(-c\varepsilon^2\omega(\widehat{\mathsf{Sp}}_\ell)^2\Big) \qquad \text{(Using Corollary2.5)}\\ &\leqslant \exp\Big(-O\Big(\frac{k}{\varepsilon^2}\log\frac{m}{k}\Big)\Big)=\varepsilon_{k,m} \qquad (10)\end{aligned}$$

where $\varepsilon_{k,m}$ can be made arbitrarily small by choosing k large enough.

Lower Bounding the gaussian width : Recall that the operator ϕ^{norm} is defined as $\phi^{\text{norm}}(\mathbf{x})\overset{\text{def}}{=}\phi(\mathbf{x})/\|\phi(\mathbf{x})\|_{\mathbf{2}}$. The operational expression for the gaussian width of the projected set restricted to coordinates in $[Nn]$ is given by :

$$\omega(\phi^{\text{norm}}(\widehat{\mathsf{Sp}}_\ell))=\mathop{\mathbf{E}}_{\mathbf{g}\sim N(0,1)^n}\left[\max_{\mathbf{x}\in\widehat{\mathsf{Sp}}_\ell}\mathbf{g}^\top\phi^{\text{norm}}(\mathbf{x})\right] \qquad (11)$$

We shall also need the following lemma which states that conditioning by the large probability event $\mathcal{E}$ does not reduce the expectation by much.

Lemma B.1. *There exists universal constants ℓ_0, m_0 such that for all $m\geqslant m_0$ and $d\geqslant k\geqslant \ell_0$ and the event $\mathcal{E}$ defined as above, we have*

$$\mathop{\mathbf{E}}_{\mathbf{g}\sim N(0,1)^n,\phi}\left[\max_{\mathbf{x}\in\widehat{\mathsf{Sp}}_\ell}\mathbf{g}^\top\phi(\mathbf{x})\Big|\mathcal{E}\right]\geqslant\mathop{\mathbf{E}}_{\mathbf{g}\sim N(0,1)^n,\phi}\left[\max_{\mathbf{x}\in\widehat{\mathsf{Sp}}_\ell}\mathbf{g}^\top\phi(\mathbf{x})\right]-\gamma_{k,m}$$

where $\gamma_{k,m}$ decays exponentially in k.

We defer the proof of the Lemma to Section B.6. Equipped with the above, we proceed to lower bound the expected Gaussian width:

$$\begin{aligned}&\mathop{\mathbf{E}}_{\phi}\mathop{\mathbf{E}}_{\mathbf{g}\sim N(0,1)^n}\left[\max_{\mathbf{x}\in\widehat{\mathsf{Sp}}_\ell}\mathbf{g}^\top\phi^{\text{norm}}(\mathbf{x})\right]\\ \geqslant&(1-\varepsilon_{k,m})\mathop{\mathbf{E}}_{\phi}\mathop{\mathbf{E}}_{\mathbf{g}\sim N(0,1)^n}\left[\max_{\mathbf{x}\in\widehat{\mathsf{Sp}}_\ell}\mathbf{g}^\top\phi^{\text{norm}}(\mathbf{x})\ \Big|\ \mathcal{E}\right]\\ \geqslant&\frac{(1-\varepsilon_{k,m})}{1+\frac{C'}{\varepsilon}(1+\varepsilon)}\mathop{\mathbf{E}}_{\phi}\mathop{\mathbf{E}}_{\mathbf{g}\sim N(0,1)^n}\left[\max_{\mathbf{x}\in\widehat{\mathsf{Sp}}_\ell}\mathbf{g}^\top\phi(\mathbf{x})\ \Big|\ \mathcal{E}\right]\\ \overset{1}{\simeq}&\frac{(1-\varepsilon_{k,m})}{1+\frac{C'}{\varepsilon}(1+\varepsilon)}\mathop{\mathbf{E}}_{\phi}\mathop{\mathbf{E}}_{\mathbf{g}\sim N(0,1)^n}\left[\max_{\mathbf{x}\in\widehat{\mathsf{Sp}}_\ell}\mathbf{g}^\top\phi(\mathbf{x})\right]\end{aligned}$$

where the first inequality follows from the fact that $\mathcal{E}$ is a large probability event, and step 1 follows from Lemma B.1. Removing the conditioning allows us to relate the expectation term to the gaussian width of $\widehat{\mathsf{Sp}}_\ell$. Let B denote the event that $\|\mathbf{g}\|_2\in[\sqrt{n}(1-\varepsilon/4),\sqrt{n}(1+\varepsilon/4)]$. Using concentration of χ^2-variables, we get $\mathbf{Pr}(B)\geqslant 1-\exp\Big(-4\varepsilon^2 n\Big)\geqslant 1-\varepsilon$ since $k\geqslant C\log\frac{1}{\varepsilon}$. Then,

$$
\begin{aligned}
\mathop{\mathbf{E}}_{\phi}\mathop{\mathbf{E}}_{\mathbf{g}\sim N(0,1)^n}\left[\max_{\mathbf{x}\in\widehat{\mathsf{Sp}}_\ell}\mathbf{g}^\top\phi(\mathbf{x})\right] &\geqslant (1-\varepsilon)\mathop{\mathbf{E}}_{\phi}\mathop{\mathbf{E}}_{\mathbf{g}\sim N(0,1)^n}\left[\max_{\mathbf{x}\in\widehat{\mathsf{Sp}}_\ell}\mathbf{g}^\top\phi(\mathbf{x})\middle| B\right] \\
&\overset{2}{=}(1-\varepsilon)\mathop{\mathbf{E}}_{\mathbf{g}\sim N(0,1)^n}\mathop{\mathbf{E}}_{\tilde{\mathbf{g}}\sim N\left(0,\frac{\|\mathbf{g}\|^2}{n}\right)^{Nn}}\left[\max_{\mathbf{x}\in\widehat{\mathsf{Sp}}_\ell}\tilde{\mathbf{g}}^\top\mathbf{x}\middle| B\right] \qquad (12)\\
&\overset{3}{\geqslant}(1-\varepsilon)\mathop{\mathbf{E}}_{\mathbf{g}\sim N(0,1)^n}\mathop{\mathbf{E}}_{\tilde{\mathbf{g}}\sim N\left(0,(1-\varepsilon/4)^2\right)^{Nn}}\left[\max_{\mathbf{x}\in\widehat{\mathsf{Sp}}_\ell}\tilde{\mathbf{g}}^\top\mathbf{x}\middle| B\right] \\
&=(1-\varepsilon)\mathop{\mathbf{E}}_{\tilde{\mathbf{g}}\sim N\left(0,(1-\varepsilon/4)^2\right)^{Nn}}\left[\max_{\mathbf{x}\in\widehat{\mathsf{Sp}}_\ell}\tilde{\mathbf{g}}^\top\mathbf{x}\right] \\
&\geqslant(1-\varepsilon)^2\mathop{\mathbf{E}}_{\tilde{\mathbf{g}}\sim N(0,1)^{Nn}}\left[\max_{\mathbf{x}\in\widehat{\mathsf{Sp}}_\ell}\tilde{\mathbf{g}}^\top\mathbf{x}\right] \qquad (13)\\
&\geqslant(1-\varepsilon)^2\omega(\widehat{\mathsf{Sp}}_\ell)
\end{aligned}
$$

In step 2, in the inner expectation $\mathbf{g}\in\mathbb{R}^n$ is a fixed vector, and therefore $\mathbf{g}^T\phi$ is distributionally equivalent to a gaussian vector in $\mathbb{R}^{Nn}$, scaled by $\frac{\|\mathbf{g}\|_2}{\sqrt{n}}$ (since the columns of ϕ are independent $N(0,1/n)^{Nn}$-gaussian vectors). Step 3 follows from the lower bound on the $\|\cdot\|_2$-length of $\mathbf{g}$. Plugging in the lower bound on the expectation term, we get :

$$
\begin{aligned}
\mathop{\mathbf{E}}_{\phi}\mathop{\mathbf{E}}_{\mathbf{g}\sim N(0,1)^n}\left[\max_{\mathbf{x}\in\widehat{\mathsf{Sp}}_\ell}\mathbf{g}^\top\phi^{\mathrm{norm}}(\mathbf{x})\right] &\geqslant (1-\varepsilon)^2\frac{(1-\varepsilon_{k,m})}{1+\frac{C'}{\varepsilon}(1+\varepsilon)}\omega(\widehat{\mathsf{Sp}}_\ell) \\
&= (1-\varepsilon)^2\frac{(1-\varepsilon_{k,m})}{1+\frac{C'}{\varepsilon}(1+\varepsilon)}\Big(\frac{C'}{\varepsilon}\sqrt{n}\Big) \\
&\geqslant \sqrt{n}(1-4\varepsilon)
\end{aligned}
$$

for sufficiently small ε and large k.

B.2. Proof of Lemma A.3

We begin by looking at the expression of the square of the $\|\cdot\|_2$ distance. For any fixed $\mathbf{x},\mathbf{y}\in\mathcal{S}^{n-1}$, we have

$$\|\mathbf{y}-\mathbf{x}\|_2^2 = 2-2\mathbf{x}^\top\mathbf{y} \qquad (14)$$

Therefore, minimizing the $\|\cdot\|_2$ norm would be equivalent to maximizing the dot product term. Furthermore, it is known that a random gaussian vector $\mathbf{g}\sim N(0,1)^n$ can be rewritten as $Z\mathbf{r}$ where $\mathbf{r}\overset{\text{unif}}{\sim}\mathcal{S}^{n-1}$ and Z^2 is a χ^2-random variable with n degrees of freedom. Using this decomposition, we get :

$$\begin{aligned}
&\Pr_{\mathbf{g}\sim N(0,1)^n}\left(\max_{\mathbf{y}\in T}\mathbf{g}^\top\mathbf{y}\leqslant\sqrt{n}(1-\varepsilon-\sqrt{\varepsilon})\right)\\
&=\Pr_{\mathbf{r},Z}\left(Z\max_{\mathbf{y}\in T}\mathbf{r}^\top\mathbf{y}\leqslant\sqrt{n}(1-\varepsilon-\sqrt{\varepsilon})\right)\\
&\geqslant\Pr_{\mathbf{r},Z}\left(Z\max_{\mathbf{y}\in T}\mathbf{r}^\top\mathbf{y}\leqslant\sqrt{n}(1-\varepsilon-\sqrt{\varepsilon})\Big|Z\leqslant\sqrt{n}(1+\varepsilon)\right)\Pr\left(Z\leqslant\sqrt{n}(1+\varepsilon)\right)\\
&\overset{1}{\geqslant}\Pr_{\mathbf{r},Z}\left(Z\max_{\mathbf{y}\in T}\mathbf{r}^\top\mathbf{y}\leqslant\sqrt{n}(1-\varepsilon-\sqrt{\varepsilon})\Big|Z\leqslant\sqrt{n}(1+\varepsilon)\right)(1-\varepsilon)\\
&\geqslant\frac{1}{2}\Pr_{\mathbf{r}}\left(\max_{\mathbf{y}\in T}\mathbf{r}^\top\mathbf{y}\leqslant\frac{1-\varepsilon-\sqrt{\varepsilon}}{1+\varepsilon}\right)\\
&\geqslant\frac{1}{2}\Pr_{\mathbf{r}}\left(\max_{\mathbf{y}\in T}\mathbf{r}^\top\mathbf{y}\leqslant(1-2\sqrt{\varepsilon})\right)
\end{aligned}$$

where in step 1, we used concentration for χ^2-random variables and use the fact that $k\geqslant C\log\frac{1}{\varepsilon}$. We now relate the behavior of the maximum dot product of a set with respect to a random vector to the maximum dot product of a fixed vector with respect to a randomly rotated set.

$$\begin{aligned}
\Pr_{\mathbf{r}}\left(\max_{\mathbf{y}\in T}\mathbf{r}^\top\mathbf{y}\leqslant(1-2\sqrt{\varepsilon})\right)&\overset{2}{=}\Pr_{R}\left(\max_{\mathbf{y}\in T}R(\mathbf{x})^\top\mathbf{y}\leqslant(1-2\sqrt{\varepsilon})\right)\\
&=\Pr_{R}\left(\max_{\mathbf{y}\in T}\mathbf{x}^\top R^{-1}(\mathbf{y})\leqslant(1-2\sqrt{\varepsilon})\right)\\
&\overset{3}{=}\Pr_{R}\left(\max_{\mathbf{y}\in T}\mathbf{x}^\top R(\mathbf{y})\leqslant(1-2\sqrt{\varepsilon})\right)\\
&=\Pr_{R}\left(\max_{\mathbf{y}\in R(T)}\mathbf{x}^\top\mathbf{y}\leqslant(1-2\sqrt{\varepsilon})\right)
\end{aligned}\tag{15}$$

Step 2 follows from the fact that applying a uniformly random rotation on a unit vector is equivalent to sampling uniformly from the unit sphere $\mathcal{S}^{n-1}$, and step 3 follows from the fact that if R is uniformly random rotation, then R^{-1} is also a uniformly random rotation. Furthermore, using gaussian concentration for Lipschitz functions :

$$\Pr_{\mathbf{g}\sim N(0,1)^n}\left(\max_{\mathbf{y}\in T}\mathbf{g}^\top\mathbf{y}\leqslant\sqrt{n}(1-\varepsilon-\sqrt{\varepsilon})\right)\leqslant\exp\left(-O\Big(\frac{k}{\varepsilon}\log\frac{m}{k}\Big)\right)\tag{16}$$

and the l.h.s is an upper bound on $\frac{1}{2}\Pr_R\left(\max_{\mathbf{y}\in R(T)}\mathbf{x}^\top\mathbf{y}\leqslant(1-2\sqrt{\varepsilon})\right)$ (Eq. 15). Therefore rearranging the equations, we have:

$$\Pr_R\left(\min_{\mathbf{y}\in R(T)}\|\mathbf{x}-\mathbf{y}\|\geqslant 2\varepsilon^{1/4}\right)=\Pr_R\left(\max_{\mathbf{y}\in R(T)}\mathbf{x}^\top\mathbf{y}\leqslant(1-2\sqrt{\varepsilon})\right)\leqslant 2\exp\left(-O\Big(\frac{k}{\varepsilon}\log\frac{m}{k}\Big)\right)\tag{17}$$

B.3. Gaussian Width of the discretized sparse set $\widehat{\mathsf{Sp}}_k$

Lemma B.2. *Let $\widehat{\mathsf{Sp}}_\ell\subset\mathcal{S}^{m-1}$ be the discrete k-sparse set on the unit sphere. Then,*

$$\omega(\widehat{\mathsf{Sp}}_\ell)=\Theta\Big(\sqrt{\ell\log\frac{m}{\ell}}\Big)\tag{18}$$

Proof. For the upper bound, observe that $\widehat{\mathsf{Sp}}_\ell\subset\mathsf{Sp}_\ell^m$ and gaussian width is monotonic. Therefore, from Lemma 1.9, we have $\omega(\widehat{\mathsf{Sp}}_\ell)=O\Big(\sqrt{\ell\log\frac{m}{\ell}}\Big)$. Towards proving the asymptotic lower bound: Given independent gaussians $g_1,\ldots,g_n\sim$

$N(0,1)$, it is known that

$$\mathop{\mathbf{E}}_{g_1,\cdots,g_n}\left[\max_{g_i} g_i\right] \geqslant C_0\sqrt{\log n} \tag{19}$$

for some constant $C_0 > 0$ independent of the number of gaussians. Now, without loss of generality let m be divisible by ℓ. We partition the m coordinates into ℓ blocks $B_1, \ldots, B_\ell$ of $\frac{m}{\ell}$ coordinates each. Then,

$$\mathop{\mathbf{E}}_{\mathbf{g}}\left[\max_{\mathbf{x}\in\widehat{\mathsf{Sp}}_\ell} \mathbf{g}^\top\mathbf{x}\right] \geqslant \frac{1}{\sqrt{\ell}}\mathbf{E}\left[\sum_{j\in[\ell]}\max_{g_{i_j}\in B_j} g_{i_j}\right] \tag{20}$$

The inequality follows from the following observation: For any fixed realization of $\mathbf{g}\sim N(0,1)^m$, let i_j be the index of the maximum in the j^{th} block. Then there exists a vector in $\widehat{\mathsf{Sp}}_\ell$ which is supported on $i_1, \ldots, i_\ell$. Therefore, the dot product would be at least the sum of maximum Gaussians from each of the blocks scaled by $\frac{1}{\sqrt{\ell}}$. The lemma now follows from applying the lower bound from Eq. 19.

□

B.4. Proof of Lemma 3.1

The proof of this Lemma uses the following comparison inequality for suprema of gaussian processes:

Lemma B.3 (Slepian's lemma (Slepian, 1962))**.** *Let $\{X_u\}_{u\in U}$ and $\{Y_u\}_{u\in U}$ be two almost surely bounded centered Gaussian processes, indexed by the same compact set U. If for every $u_1, u_2 \in U$:*

$$\mathbf{E}\left[|X_{u_1} - X_{u_2}|^2\right] \leqslant \mathbf{E}\left[|Y_{u_1} - Y_{u_2}|^2\right]$$

then we have

$$\mathbf{E}\left[\sup_{u\in U} X_u\right] \leqslant \mathbf{E}\left[\sup_{u\in U} Y_u\right]$$

Equipped with the above Lemma, we now prove Lemma 3.1:

Proof. First we prove the upper bound. Let $\Psi : X \mapsto S$ be the ε-isometric embedding map. Given $\mathbf{g}\sim N(0,\sqrt{1+\varepsilon})^m$ and $\mathbf{h}\sim N(0,1)^d$, we define the gaussian processes $\{G_{\mathbf{x}}\}_{\mathbf{x}\in X}$ and $\{H_{\mathbf{x}}\}_{\mathbf{x}\in X}$ as follows

$$\begin{aligned} G_{\mathbf{x}} &\stackrel{\text{def}}{=} \mathbf{g}^\top\mathbf{x} \\ H_{\mathbf{x}} &\stackrel{\text{def}}{=} \mathbf{h}^\top\Psi(\mathbf{x}) \end{aligned}$$

Fix $\mathbf{x}, \mathbf{y} \in X$. Using the ε-isometry of Ψ we get:

$$\begin{aligned} \mathop{\mathbf{E}}_{\mathbf{h}\sim N(0,1)^d}\left[\left|H_{\mathbf{x}} - H_{\mathbf{y}}\right|^2\right] &= \mathop{\mathbf{E}}_{\mathbf{h}\sim N(0,1)^d}\left[\left|\mathbf{h}^\top\Psi(\mathbf{x}) - \mathbf{h}^\top\Psi(\mathbf{y})\right|^2\right] \\ &\stackrel{1}{=} \|\Psi(\mathbf{x}) - \Psi(\mathbf{y})\|^2 \\ &\stackrel{2}{\leqslant} (1+\varepsilon)\|\mathbf{x}-\mathbf{y}\|^2 \\ &= \mathop{\mathbf{E}}_{\mathbf{g}\sim N(0,1+\varepsilon)^m}\left[\left|G_{\mathbf{x}} - G_{\mathbf{y}}\right|^2\right] \end{aligned}$$

where step 1 follows from the fact that the variance h in the direction of a vector $\mathbf{v}$ is $\|\mathbf{v}\|_2^2$, and inequality 2 follows from the isometric property. Therefore, using Lemma B.3, we have

$$\underset{\mathbf{h}\sim N(0,1)^d}{\mathbf{E}}\left[\max_{\mathbf{x}\in S}\mathbf{g}^\top\Psi(\mathbf{x})\right] \leqslant \underset{\mathbf{g}\sim\sqrt{1+\varepsilon}N(0,1)^m}{\mathbf{E}}\left[\max_{\mathbf{x}\in S}\mathbf{g}^\top\mathbf{x}\right] \tag{21}$$

which directly gives us $\omega(S) \leqslant \sqrt{1+\varepsilon}\cdot\omega(X) \leqslant (1+\varepsilon)\omega(X)$. The other direction follows by using the lower bound given by isometry.

$\square$

B.5. Rotational Invariance of ϕ^{norm}

Lemma B.4. *For any finite set $T \subset \mathcal{S}^{Nn-1}$, the distribution of $\Phi^{\mathrm{norm}}(S)$ is rotation invariant.*

Proof. Fix any $N = |T|$ vectors $\{\mathbf{z}'_1,\ldots,\mathbf{z}'_N\}$. Let $R : \mathbb{R}^{Nn} \mapsto \mathbb{R}^{Nn}$ be a fixed rotation. With a slight abuse of notation, we shall use $\mathbf{Pr}_\phi(\cdot)$ to denote the pdf of the distribution here. Then,

$$\begin{aligned}
\underset{\phi}{\mathbf{Pr}}\left(\phi^{\mathrm{norm}}(T) = \left\{\frac{\mathbf{z}'_1}{\|\mathbf{z}'_1\|_2},\ldots,\frac{\mathbf{z}'_N}{\|\mathbf{z}'_N\|_2}\right\}\right) &= \underset{\phi}{\mathbf{Pr}}\left(\phi(T) = \{\mathbf{z}'_1,\ldots,\mathbf{z}'_N\}\right)\\
&\overset{1}{=} \underset{\phi}{\mathbf{Pr}}\left(\phi(T) = \{R(\mathbf{z}'_1),\ldots,R(\mathbf{z}'_N)\}\right)\\
&= \underset{\phi}{\mathbf{Pr}}\left(\phi^{\mathrm{norm}}(T) = \left\{\frac{R(\mathbf{z}'_1)}{\|R(\mathbf{z}'_1)\|_2},\ldots,\frac{R(\mathbf{z}'_N)}{\|R(\mathbf{z}'_N)\|_2}\right\}\right)\\
&\overset{2}{=} \underset{\phi}{\mathbf{Pr}}\left(\phi^{\mathrm{norm}}(T) = \left\{\frac{R(\mathbf{z}'_1)}{\|(\mathbf{z}'_1)\|_2},\ldots,\frac{R(\mathbf{z}'_N)}{\|(\mathbf{z}'_N)\|_2}\right\}\right)\\
&= \underset{\phi}{\mathbf{Pr}}\left(\phi^{\mathrm{norm}}(T) = R\left(\left\{\frac{\mathbf{z}'_1}{\|\mathbf{z}'_1\|_2},\ldots,\frac{\mathbf{z}'_N}{\|\mathbf{z}'_N\|_2}\right\}\right)\right)
\end{aligned}$$

where step 1 follows from the rotational invariance of ϕ, and step 2 uses the observation that rotating a vector does not change it's ℓ_2-length. Since the equality holds for any rotation, the statement follows. $\square$

B.6. Proof of Lemma B.1

Proof. The proof uses the more general observation relating conditional expectations to their unconditioned counterparts :

Proposition B.5. *Let $\mathcal{E}$ be an event such that $\mathbf{Pr}(\mathcal{E}) \geqslant 1-\eta$. Let $Z : \mathbb{R}^n \mapsto \mathbb{R}_+$ be a non-negative random variable. Let $t_0 > 0$ be such that*

$$\int_{t_0}^{\infty}\mathbf{Pr}(Z \geqslant t)dt \leqslant \alpha$$

Then, the following holds true : $\mathbf{E}[Z|\mathcal{E}] \geqslant \mathbf{E}[Z] - \eta t_0 - \alpha$.

Proof. We begin by observing that for any event B,

$$\mathbf{Pr}(B|\mathcal{E}) \geqslant \mathbf{Pr}(B) - \eta$$

which follows from the definition of conditional expectation. Therefore,

$$\begin{aligned}\mathbf{E}[Z|\mathcal{E}] = \int_0^\infty \mathbf{Pr}(Z \geqslant t|\mathcal{E})dt &\geqslant \int_0^{t_0} \mathbf{Pr}(Z \geqslant t|\mathcal{E})dt \\ &\geqslant \int_0^{t_0} \big(\mathbf{Pr}(Z \geqslant t) - \eta\big)dt \\ &\geqslant \int_0^{t_0} \mathbf{Pr}(Z \geqslant t)dt - \int_0^{t_0} \eta dt \\ &= \mathbf{E}[Z] - \alpha - \eta t_0\end{aligned}$$

□

We apply the above lemma to our setting: let $\mathcal{E}$ be the event as described in the proof of Lemma A.2 and let Z be the random variable :

$$Z := \max_{\mathbf{x} \in \widehat{\mathsf{Sp}}_\ell} \mathbf{g}^\top \phi(\mathbf{x})$$

From Eq. 10, we know that $\mathbf{Pr}(\neg\mathcal{E}) = \varepsilon_{k,m} = \eta_{k,m}$. Abusing notation, we denote $\omega^* = \omega(\widehat{\mathsf{Sp}}_\ell)$. Let $t_0 = 4\omega(\widehat{\mathsf{Sp}}_\ell) = 4\omega^*$. For a fixed choice of $\delta > 0$, let B_δ be the event that $\|\mathbf{g}\|_2 \leqslant \sqrt{n}(2+\delta)$. Then,

$$\begin{aligned}&\mathop{\mathbf{Pr}}_{\mathbf{g}\sim N(0,1)^n,\phi}\Big(\max_{\mathbf{x}\in\widehat{\mathsf{Sp}}_\ell} \mathbf{g}^\top\phi(\mathbf{x}) \geqslant (2+\delta)^2\omega^*\Big) \\ &\leqslant \mathbf{Pr}(\neg B_\delta) + \mathop{\mathbf{Pr}}_{\mathbf{g}\sim N(0,1)^n,\phi}\Big(\max_{\mathbf{x}\in\widehat{\mathsf{Sp}}_\ell} \mathbf{g}^\top\phi(\mathbf{x}) \geqslant (2+\delta)^2\omega^* \Big| B_\delta\Big) \\ &\stackrel{1}{\leqslant} \exp\Big(-O((1+\delta)^2 n)\Big) + \mathop{\mathbf{Pr}}_{\mathbf{g}\sim N(0,1)^n,\phi}\Big(\max_{\mathbf{x}\in\widehat{\mathsf{Sp}}_\ell} \mathbf{g}^\top\phi(\mathbf{x}) \geqslant (2+\delta)^2\omega^*|B_\delta\Big)\end{aligned} \tag{22}$$

where inequality 1 follows by concentration on χ^2 variables. We upper bound the remaining probability term as :

$$\begin{aligned}&\mathop{\mathbf{Pr}}_{\mathbf{g}\sim N(0,1)^n,\phi}\Big(\max_{\mathbf{x}\in\widehat{\mathsf{Sp}}_\ell} \mathbf{g}^\top\phi(\mathbf{x}) \geqslant (2+\delta)^2\omega^*|B_\delta\Big) \\ &\stackrel{2}{\leqslant} \mathop{\mathbf{Pr}}_{\tilde{\mathbf{g}}\sim N(0,1)^{Nn}}\Big(\max_{\mathbf{x}\in\widehat{\mathsf{Sp}}_\ell} \tilde{\mathbf{g}}^\top\mathbf{x} \geqslant \frac{(2+\delta)^2}{2+\delta}\omega^*|B_\delta\Big) \\ &\leqslant \mathop{\mathbf{Pr}}_{\tilde{\mathbf{g}}\sim N(0,1)^{Nn}}\Big(\max_{\mathbf{x}\in\widehat{\mathsf{Sp}}_\ell} \tilde{\mathbf{g}}^\top\mathbf{x} \geqslant (2+\delta)\omega^*\Big) \qquad (23) \\ &\stackrel{3}{\leqslant} \exp\Big(-O((1+\delta)\omega^*)^2\Big) \\ &\leqslant \exp\Big(-O\Big((1+\delta)^2 k\log\frac{m}{k}\Big)\Big) \qquad (24)\end{aligned}$$

Step 2 can be shown using arguments identical to the ones used in steps 12-13 in the proof of Lemma 3.2, and step 3 follows from gaussian concentration. Now we proceed to upper bound the quantity α (as in Proposition B.5):

$$
\begin{aligned}
\int_{t_0}^{\infty} \mathbf{Pr}(Z \geqslant t)dt &= \int_0^{\infty} \mathop{\mathbf{Pr}}_{\mathbf{g}\sim N(0,1)^n,\phi}\Big(\max_{\mathbf{x}\in\widehat{\mathsf{Sp}}_\ell} \mathbf{g}^\top \phi(\mathbf{x}) \geqslant 4\omega^* + t\Big)dt \\
&\overset{1}{=} \int_0^{\infty} 2(2+\delta)\omega^* \mathop{\mathbf{Pr}}_{\mathbf{g}\sim N(0,1)^n,\phi}\Big(\max_{\mathbf{x}\in\widehat{\mathsf{Sp}}_\ell} \mathbf{g}^\top \phi(\mathbf{x}) \geqslant (2+\delta)^2\omega^*\Big)d\delta \\
&\overset{2}{\leqslant} \int_0^{\infty} (2+\delta)\omega^* \exp\Big(-O((1+\delta)^2 k \log\frac{m}{k})\Big)d\delta \\
&\leqslant \int_0^{\infty} \exp\Big(-O((1+\delta)^2 k \log\frac{m}{k})\Big)d\delta = \alpha_{k,m}
\end{aligned} \tag{25}
$$

where step 1 is a change of variables argument where we set $t = (2+\delta)^2\omega^* - 4\omega^*$, and the second step follows from by combining upper bounds from 22 and 24.

Plugging in the upper bounds from Equations 22,24 and 25, we get

$$
\mathbf{E}[Z|\mathcal{E}] \geqslant \mathbf{E}[Z] - \alpha_{k,m} - 4(\omega^*)\eta_{k,m} = \mathbf{E}[Z] - \gamma_{k,m}
$$

where $\gamma_{k,m}$ decays exponentially in k.

□

B.7. Proof of Lemma 3.3

The proof uses the observation that for $\mathbf{R} \sim \mathbb{O}_d$, for any $i \in [d]$ marginal distribution of the vector $\mathbf{r}_i$ is that of a uniformly random vector drawn from $\mathcal{S}^{d-1}$ (c.f., Exercise 5 (Vershynin, 2011)). Therefore, it suffices to show large probability upper bounds for a single random vector $\mathbf{r} \sim \mathcal{S}^{d-1}$, which can then be used to complete the proof by a union bound argument.

.

Concentration for random unit vectors: Let $C > 0$ be a constant which is fixed later. The first step follows from replacing the unit vector by a normalized gaussian vector:

$$
\begin{aligned}
\mathop{\mathbf{Pr}}_{\mathbf{r}\sim\mathcal{S}^{d-1}}\left[\max_{\mathbf{x}\in S} \mathbf{r}^\top\mathbf{x} \geqslant C\omega(S)/\sqrt{d}\right] &= \mathop{\mathbf{Pr}}_{\mathbf{g}\sim N(0,1)^d}\left[\max_{\mathbf{x}\in S} \mathbf{g}^\top\mathbf{x} \geqslant C\omega(S)\frac{\|\mathbf{g}\|}{\sqrt{d}}\right] \\
&\leqslant \mathop{\mathbf{Pr}}_{\mathbf{g}\sim N(0,1)^d}\left[\mathbf{g}^\top\mathbf{x} \geqslant C\omega(S)\frac{\|\mathbf{g}\|}{\sqrt{d}}\Big|\|\mathbf{g}\| \geqslant \sqrt{d}/2\right] + \mathop{\mathbf{Pr}}_{\mathbf{g}\sim N(0,1)^d}\left[\|\mathbf{g}\| \leqslant \sqrt{d}/2\right] \\
&\leqslant 2\mathop{\mathbf{Pr}}_{\mathbf{g}\sim N(0,1)^d}\left[\mathbf{g}^\top\mathbf{x} \geqslant C\omega(S)/2\right] + \mathop{\mathbf{Pr}}_{\mathbf{g}\sim N(0,1)^d}\left[\|\mathbf{g}\| \leqslant \sqrt{d}/2\right] \\
&\leqslant 4\max\Big(\exp(-C'\omega^2(S)), \exp(-C'd)\Big)
\end{aligned}
$$

where the first term is upper bounded using Lemma 2.6, and the second term is bounded using χ^2-concentration.

Concentration for random rotations: We now extend the above concentration bound to an expectation bound for random rotations. Let $\mathcal{E}$ denote the event for $\mathbf{R} \sim \mathbb{O}_d$, there exists $\mathbf{x} \in S$ such that $\|\mathbf{R}\mathbf{x}\|_\infty \geqslant C\omega(S)/\sqrt{d}$.

$$\begin{aligned}\Pr_{\mathbf{R}\sim\mathbb{O}_d}\left[\max_{\mathbf{x}\in S}\|\mathbf{Rx}\|_\infty > C\omega(S)/\sqrt{d}\right] &= \Pr_{\mathbf{R}\sim\mathbb{O}_d}\left[\max_{i\in[d]}\max_{\mathbf{x}\in S}\mathbf{r}_i^\top\mathbf{x} > C\omega(S)/\sqrt{d}\right] \\ &\leqslant \sum_{i\in[d]}\Pr_{\mathbf{R}\sim\mathbb{O}_d}\left[\max_{\mathbf{x}\in S}\mathbf{r}_i^\top\mathbf{x} > C\omega(S)/\sqrt{d}\right] \\ &\leqslant 4\max\left(\exp(-C'\omega^2(S)), \exp(-C'd)\right) \leqslant \frac{1}{2}\end{aligned}$$

where the last step follows from the fact that $d >> \log d$ and by choice $C'\omega(S) \geqslant 10\log d$, when C is chosen to be large enough.

C. Intractability Results for Testing Sparsity

In this section, we show that the additive error/sparsity tradeoff achieved by Theorem 1.3 is asymptotically the best possible for efficient algorithms. In fact, we shall prove a stronger statement, which rules out algorithms with multiplicative sparsity of the order $1/\varepsilon^{\log^{1-\delta} m}$.

For ease of notation, we say that an algorithm is a η-bicriteria approximation algorithm for testing sparsity, if it has the following guarantees. On inputs matrix $\mathbf{A} \in \mathbb{R}^{d\times m}$ (where $\|\mathbf{a}_i\| = 1$ for every $i \in [m]$),vector $\mathbf{y} \in \mathbb{R}^d$, parameters $k \leqslant m, \varepsilon > 0$, it satisfies the following:

– **Completeness**: If $\mathbf{y} = \mathbf{Ax}$ for some $\mathbf{x} \in \mathbb{R}^m$ such that $\|\mathbf{x}\|_0 \leqslant k$, the tester accepts.

– **Soundness**: If for all choices of $\mathbf{x} \in \mathbb{R}^m$ such that $\|\mathbf{x}\|_0 \leqslant k/\varepsilon^{\log^{1-\delta} m}$, it satisfies $\|\mathbf{Ax} - \mathbf{y}\| \geqslant \varepsilon\|\mathbf{x}\|^\eta$, then the tester rejects.

The following Theorem shows the tightness of error/sparsity tradeoff.

Theorem C.1. *Assuming* SAT *does not have $n^{O(\log\log n)}$-time algorithm, there does not exist polynomial*[7] *time η-bicriteria approximation algorithm for testing sparsity, for any $\eta < 1$.*

Proof. The proof uses the construction of hard sparse recovery instances from (Foster et al., 2015). It reduces Feige's hard instances for SET COVER to hard instances of Sparse Recovery with the following guarantees:

Theorem C.2. *Given $\mathbf{B} \in \{0, 1\}^{d\times m}$ and a parameter $k \leqslant m$, there does not exist polynomial time algorithm which can distinguish between the following cases:*

– **YES***: There exist k-support disjoint columns $\mathbf{b}_{i_1}, \ldots, \mathbf{b}_{i_k}$ such that $\sum_{j\in[k]}\mathbf{b}_{i_j} = \mathbf{e}$*

– **NO***: There does not exist $i_1, \ldots, i_{k'} \in [m]$ and coefficients $c_{i_1}, \ldots, c_{i'_k} \in \mathbb{R}$ such that*

$$\Big\|\sum_{j\in[k']} c_{i_j}\mathbf{b}_{i_j} - \mathbf{e}\Big\|^2 \leqslant 1$$

where $k' = C2^{\log^{1-\delta} m}$, for all $\delta \in (0, 1)$ and some constant $C > 0$, assuming SAT *has no $n^{O(\log\log n)}$-time algorithm.*

As in (Foster et al., 2015), we now construct a new sparse recovery instance as follows. Let $r = d^{\lceil\frac{1}{\delta}\rceil - 1}$. Let $\mathbf{A} \in \mathbb{R}^{rd\times m}$ be the matrix constructed by stacking r-copies of $\mathbf{B}$. We set $\mathbf{y} \in \mathbb{R}^{rd}$ to be the vector of all ones. Finally, we construct $\mathbf{A}^{\text{unit}}$ from $\mathbf{A}$ by normalizing all the columns to have unit length in ℓ_2-norm. By construction, the following properties hold:

– If $(\mathbf{B}, k)$ is a YES instance, then there exist k-support disjoint columns $\mathbf{a}_{i_1}, \ldots, \mathbf{a}_{i_k}$ such that $\sum_{j\in[k]}\mathbf{a}_{i_j} = \mathbf{y}$

[7] Since we are dealing with bicriteria algorithms, this is polynomial in the input size and ε^{-1}

– If $(\mathbf{B}, k)$ is a NO instance, there does not exist $i_1, \ldots, i_{k'} \in [m]$ and coefficients $c_{i_1}, \ldots, c_{i_{k'}} \in \mathbb{R}$ such that

$$\Big\| \sum_{j \in [k']} c_{i_j} \mathbf{a}_{i_j} - \mathbf{e} \Big\|^2 \leqslant r$$

Furthermore, since $\mathbf{A}$ is a $\{0,1\}$-matrix, we have $\|\mathbf{a}_i\| \leqslant \sqrt{rd}$. Therefore, in the YES case, there exists $\mathbf{x}^* \in \mathbb{R}^q$ such that $\|\mathbf{x}^*\|_0 = k$, $\|\mathbf{x}^*\|_2 \leqslant \sqrt{krd}$ and $\mathbf{A}^{\text{unit}}\mathbf{x}^* = \mathbf{y}$. Additionally, since the columns chosen in the YES case are support disjoint, we have $k \leqslant d$.

Now consider a η-bicriteria approximation algorithm for sparse recovery. For $\varepsilon^{-\log^{1-\delta} m} = C2^{\log^{1-\delta} m}$, we must have $\varepsilon^2 \|\mathbf{x}^*\|^{2\eta} \geqslant r = d^{1/\delta - 1}$ (otherwise the approximation algorithm can be used to distinguish between the YES and NO cases). On the other hand,

$$\varepsilon^2 \|\mathbf{x}^*\|^{2\eta} \leqslant \frac{(rkd)^\eta}{\left(C2^{\log^{1-\delta} m}\right)^{2/\log^{1-\delta} m}} \leqslant (d^{1/\delta} k)^\eta \leqslant d^{\eta(1/\delta + 1)}$$

Comparing the lower and upper bound for $\varepsilon^2 \|\mathbf{x}^*\|^2$ we get $\eta \geqslant \frac{1/\delta - 1}{1/\delta + 1}$ which goes to 1 as δ goes to 0. This completes the proof.

□

D. Sparse Recovery via Sketching

The techniques used to prove Theorem 1.3 can also be used to give a sketching algorithm for sparse recovery, as stated by the following Theorem:

Theorem D.1. *Given matrix $\mathbf{A} \in \mathbb{R}^{d \times m}$ (where $\|\mathbf{a}_i\|_2 = 1$ for every $i \in [m]$) and vector $\mathbf{y} = \mathbf{A}\mathbf{x}$ (for some unknown $\mathbf{x} \in \mathbb{R}^m$ such that $\|\mathbf{x}\|_0 \leqslant k$). Then for every choice of $\varepsilon > 0$ and $\delta \in (0,1)$, there exists a randomized sketching algorithm which returns $\hat{\mathbf{x}} \in \mathbb{R}^m$ such that $\|\hat{\mathbf{x}}\|_0 \leqslant k/\varepsilon^2$ and $\|\mathbf{A}\hat{\mathbf{x}} - \mathbf{y}\| \leqslant C\varepsilon\|\mathbf{x}\|$, with probability at least $1 - \delta$. The number of linear queries made by the algorithm is $O\Big(\frac{k}{\varepsilon^2} \log \frac{m}{\delta}\Big)$.*

Proof. The algorithm follows from the soundness analysis for Theorem 1.3. Consider the setting where $\|\mathbf{x}\| \leqslant 1$ (the bound as stated in the Theorem can be recovered by rescaling) and let $\mathbf{y} = \mathbf{A}\mathbf{x}$. Let $\Phi, A_{\varepsilon/\sqrt{k}}$ and $\tilde{A}_{\varepsilon/\sqrt{k}}$ be as in the proof of Theorem 1.3. By identical arguments, it follows that with probability $1 - \delta$, we have $\tilde{\mathbf{y}} \in \sqrt{k} \cdot \mathrm{conv}\Big(\tilde{A}_{\pm}\Big)$. By construction, there exist $\mathbf{z} \in A_{\varepsilon/\sqrt{k}}$ be such that $\hat{\mathbf{z}} = \Phi(\mathbf{z})$. Since the choice of the projection dimension ensures that with probability at least $1 - \delta$, the set $A_{\varepsilon/\sqrt{k}} \cup \{\mathbf{y}\}$ is ε-isometric to $\tilde{A}_{\varepsilon/\sqrt{k}} \cup \{\tilde{\mathbf{y}}\}$ it follows that $\|\mathbf{z} - \mathbf{y}\| \leqslant (1 - \varepsilon^{-1})\|\hat{\mathbf{z}} - \tilde{\mathbf{y}}\| \leqslant 2\varepsilon$. But $\mathbf{z} = \mathbf{A}\hat{\mathbf{x}}$ for some $\hat{\mathbf{x}} \in \mathbb{R}^m$ such that $\|\hat{\mathbf{x}}\|_0 \leqslant k/\varepsilon^2$.

Considering the above arguments, the algorithm for sparse recovery is the following. On input $\mathbf{y}$, it computes linear sketch $\tilde{\mathbf{y}} = \Phi(\mathbf{y})$, and using the constructive variant of Approximate Caratheodory's Theorem (Blum et al., 2016), it outputs $\hat{\mathbf{x}} \in \mathbb{R}^m$ such that $\|\tilde{\mathbf{A}}\hat{\mathbf{x}} - \tilde{\mathbf{y}}\| \leqslant C.\varepsilon$ and $\|\hat{\mathbf{x}}\|_0 \leqslant k/\varepsilon^2$. The guarantees follow using the above arguments. □

E. Analysis for the Dimensionality Tester

We state the definition of ε-approximate rank of a matrix, as defined in (Alon et al., 2013):

Definition E.1 (Approximate Rank)**.** *Given a matrix $\mathbf{Y} \in \mathbb{R}^{m \times n}$ and an $\varepsilon > 0$, the ε-approximate rank of the matrix* (denoted by $\mathrm{rank}_\varepsilon(\mathbf{Y})$) *is defined as follows:*

$$\mathrm{rank}_\varepsilon(\mathbf{Y}) = \min\left(\left\{\mathrm{rank}(\hat{\mathbf{Y}}) : \hat{\mathbf{Y}} \in \mathbb{R}^{m \times n}, \|\mathbf{Y} - \hat{\mathbf{Y}}\|_{\infty,\infty} \leqslant \varepsilon\right\}\right) \tag{26}$$

where $\|\cdot\|_{\infty,\infty}$ is norm defined as the largest absolute value of an entry in the matrix.

We first prove Lemma E.2 which relates the approximate rank of a matrix in terms of the gaussian width, and use that to analyze the tester.

Lemma E.2. *For a matrix $\mathbf{Y} \in \mathbb{R}^{d \times n}$, where $\|\mathbf{y}_i\| = 1 \;\forall\; i \in [n]$, the following holds:*

$$\mathrm{rank}_{\varepsilon}(\mathbf{Y}) \leqslant O\left(\frac{1}{\varepsilon^2} \max\left(\omega^2(\mathbf{Y}), \log d\right)\right) \tag{27}$$

for any $\varepsilon \geqslant O(1/\sqrt{d})$.

Proof. Let $Y = \{\mathbf{y}_1, \ldots, \mathbf{y}_n\}$ be the set of columns from the matrix $\mathbf{Y}$. Let $Y_0 = Y \cup \mathrm{I}_d$ where I_d is the set of standard basis vectors $\{\mathbf{e}_i\}_{i \in [d]}$. It is known that gaussian width is subadditive, and therefore

$$\omega(Y_0) \leqslant \omega(Y) + \omega(\mathrm{I}_d) \leqslant 2 \max\left(\omega(Y), 2\sqrt{\log d}\right) \tag{28}$$

Let $d' = \frac{16C}{\varepsilon^2} \max\left(\log d, \omega^2(Y)\right)$ where C is the constant given by the generalized JL-lemma and let $\mathbf{G} \sim \frac{1}{\sqrt{d'}} N(0,1)^{d' \times d}$. Then with high probability, $\mathbf{G}(Y_0)$ is ε-isometric to Y_0. For every $i \in [d]$ and $j \in [n]$, we observe that:

1. $1 - \varepsilon \geqslant \|\mathbf{G}\mathbf{e}_i\|^2, \|\mathbf{G}\mathbf{y}_j\|^2 \leqslant 1 + \varepsilon$
2. $(1-\varepsilon)\|\mathbf{e}_i - \mathbf{y}_j\|^2 \leqslant \|\mathbf{G}\mathbf{e}_i - \mathbf{G}\mathbf{y}_j\|^2 \leqslant (1+\varepsilon)\|\mathbf{e}_i - \mathbf{y}_j\|^2$ which in turn implies that $|\langle \mathbf{G}\mathbf{e}_i, \mathbf{G}\mathbf{y}_j\rangle - \langle \mathbf{e}_i, \mathbf{y}_j\rangle| \leqslant O(\varepsilon)$.

Let $\mathbf{Y}' = \mathbf{G}^\top \mathbf{G}\mathbf{Y}$. Since the above observation is true for any $i \in [d], j \in [n]$, it follows that $\mathbf{Y}'$ is entry wise $O(\varepsilon)$-close to $\mathbf{Y}$, and by construction $\mathrm{rank}(\mathbf{Y}') \leqslant d'$. Hence, the claim follows.

□

The algorithm for testing dimensionality is the following:

Algorithm 3 TestDimension

1: Use Lemma 2.8 to obtain $\hat{\omega}$, an estimate of $\omega(S)$ within additive error $\sqrt{k}/2$ with probability at least $1 - \delta$.
2: Accept if $\hat{\omega} \leqslant 2\sqrt{k}$, else reject.

We now show completeness and soundness for the tester:

Proof of Theorem 1.5. Let S denote the set $\{\mathbf{y}_1, \ldots, \mathbf{y}_p\}$. The tester obtains $\hat{\omega}$ that approximates $\omega(S)$ to an additive error of $\sqrt{k}$ and accepts iff $\hat{\omega} \leqslant 2\sqrt{k}$. By Lemma 2.8, the tester requires $O(p \log \delta^{-1})$ linear queries to obtain $\hat{\omega}$.

If $\dim(S) \leqslant k$, then by Lemma 1.9, $\hat{\omega} \leqslant 2\sqrt{k}$ with probability at least $1 - \delta$, so that the tester accepts with the same probability.

If the tester accepts, then with probability at least $1 - \delta$, $\omega(S) \leqslant 3\sqrt{k}$. Therefore, from Lemma E.2, we have $\mathrm{rank}_{\varepsilon}(Y) \leqslant O(k/\varepsilon^2)$ which completes the proof. □

F. On the relationship between RIP and Incoherence

Even though our results are stated in terms of dictionaries which satisfy RIP, they can be stated in terms of incoherence as well. This is because the incoherence[8] and RIP constants of the dictionary matrix are roughly equivalent. We formalize this observation in the following lemma:

Proposition F.1. *Let $\mathbf{A} \in \mathbb{R}^{d \times m}$ be a matrix with $\|\mathbf{a}\|_i = 1$ for every $i \in [m]$. Then,*

– *If $\mathbf{A}$ is $(2k, \zeta)$-RIP then it is ζ-incoherent.*

[8] Here incoherence is stated in dimension free terms i.e., $|\langle \mathbf{a}_i, \mathbf{a}_j\rangle| \leqslant \mu$ for every $i \neq j$

– *If* $\mathbf{A}$ *is* μ*-incoherent, then it is* $(2k, 4k\mu)$*-RIP*

Proof. Suppose $\mathbf{A}$ is $(2k, \zeta)$-RIP. Then for any $i, j \in [m]$

$$|\langle \mathbf{a}_i, \mathbf{a}_j \rangle| = 1 - \frac{\|\mathbf{a}_i - \mathbf{a}_j\|^2}{2} \overset{1}{\in} 1 - (1 \pm \zeta) = \pm\zeta \tag{29}$$

where 1 follows using the RIP guarantee. On the other hand, let $\mathbf{A}$ be μ-incoherent. Then for any $S \subset [m]$ of size $2k$ let $\mathbf{M} = \mathbf{A}_S^\top \mathbf{A}_S$ where $\mathbf{A}_S$ is the submatrix induced by columns in S. Then we observe that $M_{ii} = \|\mathbf{a}_i\|^2 = 1$ for every $i \in [2k]$ and off-diagonal entries satisfy $|M_{ij}| \leqslant \mu$. Therefore, using the Gershgorin's disk theorem $\lambda(\mathbf{M}) \in [1 \pm 2\mu k]$. Therefore for every $\mathbf{x}$ supported on S we have $\|\mathbf{A}\mathbf{x}\|^2 \in (1 \pm 2\mu k)^2\|\mathbf{x}\|^2 \in (1 \pm 4\mu k)\|\mathbf{x}\|^2$. Since this is true for any arbitrary $2k$-sized subset S, the result follows.

□

Note that quantitatively, incoherence is a stronger property than RIP since μ-incoherence implies $(2k, 4k\mu)$-RIP but $(2k, 4k\mu)$-RIP only implies $4k\mu$-incoherence. Naturally, Theorem 1.2 can be restated in terms of incoherent linear transformations as well.

G. Implicit results for Known Design setting

In this section, we discuss results that follow implicitly from previous work.

Sketching in the Streaming Model. In the streaming model, one has a series of updates (i, v) where each $i \in [n]$ and $v \in \{-T, \ldots, T\}$. Each update modifies a vector $\mathbf{x}$, initialized at $\mathbf{0}$, to $\mathbf{x} + v\mathbf{e}_i$. The L_0-estimation problem in streaming is to estimate the sparsity of $\mathbf{x}$ upto a multiplicative $(1 \pm \varepsilon)$ factor. A linear sketch algorithm maintains $\mathbf{M}\mathbf{x}$ during the stream, where $\mathbf{M} \in \mathbb{R}^{s \times n}$ is a randomized matrix.

A linear sketch algorithm for the L_0-estimation problem directly yields a tester in the setting where the design matrix is known to be the identity matrix. By invoking the space-optimal L_0-estimation result from (Kane et al., 2010), we obtain:

Theorem G.1 (Implicit in (Kane et al., 2010))**.** *Fix* $\varepsilon \in (0, 1)$*, positive integers* m, k *and an* invertible *matrix* $\mathbf{A} \in \mathbb{R}^{m \times m}$*. Then, there is a tester with query complexity* $O(\varepsilon^{-2}\log(m))$ *that, for an input* $\mathbf{y} \in \mathbb{R}^m$*, accepts with probability at least* $2/3$ *if* $\mathbf{y} = \mathbf{A}\mathbf{x}$ *for some* k*-sparse* $\mathbf{x} \in \mathbb{Z}^m$*, and rejects with probability* $2/3$ *if* $\mathbf{y} = \mathbf{A}\mathbf{x}$ *for some* $(1+\varepsilon)k$*-sparse* $\mathbf{x} \in \mathbb{Z}^m$*. The running time of the algorithm is* $\mathrm{poly}(m, 1/\varepsilon)$*.*

We believe that the theorem should also extend (albeit with a mild change in parameters) to the setting where $\mathbf{x}$ is an arbitrary real vector (not necessarily discrete), but the assumption that $\mathbf{A}$ is invertible seems hard to circumvent.

Sketching in Numerical Linear Algebra. Low-dimensional sketches used in numerical linear algebra can also yield testers in the known design matrix case of our model. For a matrix $\mathbf{A} \in \mathbb{R}^{d \times m}$, suppose we want a property tester that, for input $\mathbf{y} \in \mathbb{R}^d$, distinguishes between the case $\mathbf{y} = \mathbf{A}\mathbf{x}^*$ for some k-sparse $\mathbf{x}^*$, and the case $\min_{k\text{-sparse } \mathbf{x}} \|\mathbf{A}\mathbf{x} - \mathbf{y}\| > \varepsilon$.

In the sketching approach, one looks to solve the optimization problem $\min_{k\text{-sparse } \mathbf{x}} \|\mathbf{A}\mathbf{x} - \mathbf{y}\|$ in a smaller dimension i.e., one looks at:

$$\widehat{\mathbf{x}} = \arg\min_{\mathbf{x}' \in K} \|\mathbf{S}\mathbf{A}\mathbf{x}' - \mathbf{S}\mathbf{y}\| = \arg\min_{\mathbf{x}' \in K} \|\mathbf{S}(\mathbf{A}\mathbf{x}' - \mathbf{y})\| \tag{30}$$

where $\mathbf{S} \in \mathbb{R}^{q \times d}$ is a sketch matrix (where $q \ll d$) and $K = \{\mathbf{x} : \|\mathbf{x}\|_0 \leqslant k\}$. The intent here is that the vector $\hat{\mathbf{x}}$ would also be an approximate minimizer to the original optimization problem.

An *oblivious* ℓ_2*-subspace embedding* with parameters $(d, m, \varepsilon, \delta)$ is a distribution on $q \times d$ matrices $\mathbf{M}$ such that with probability at least $1 - \delta$, for any fixed $d \times m$ matrix $\mathbf{A}$, $(1 - \varepsilon)\|\mathbf{A}\mathbf{x}\| \leqslant \|\mathbf{M}\mathbf{A}\mathbf{x}\| \leqslant (1 + \varepsilon)\|\mathbf{A}\mathbf{x}\|$ for all $\mathbf{x} \in \mathbb{R}^m$. For our application, suppose we draw $\mathbf{S}$ from an oblivious subspace embedding with parameters[9] $\left(d, k+1, \varepsilon, \delta/\binom{m}{k}\right)$. Then, we get a valid property tester with query complexity q if we accept when $\|\mathbf{S}\mathbf{A}\widehat{\mathbf{x}} - \mathbf{S}\mathbf{y}\| = 0$ and reject when it is at least $\varepsilon(1 - \varepsilon)$.

[9] That is, consider all possible choices of supports $\Omega \in \binom{[m]}{\leqslant k}$ and let $\mathbf{A}_\Omega$ be the submatrix corresponding to the columns of Ω. For the given choice of parameters, it follows that with probability $\geqslant 1 - \delta$, every $\mathbf{x} \in \mathbb{R}^m$ will satisfy $\|\mathbf{S}(\mathbf{A}_\Omega\mathbf{x} - \mathbf{y})\| \in (1 \pm \varepsilon)\|\mathbf{A}_\Omega\mathbf{x} - \mathbf{y}\|$

Algorithm 4 SparseTestKnown-Noisy

1: Set $n = \frac{200k}{\varepsilon^2} \log \frac{m}{\delta}$, sample projection matrix $\Phi \sim \frac{1}{\sqrt{n}} N(0,1)^{n \times d}$
2: Observe linear sketch $\tilde{\mathbf{y}} = \Phi(\mathbf{y})$
3: Let $A_\pm = A \cup -A$
4: Accept iff $\operatorname{dist}\Big(\tilde{\mathbf{y}}, \sqrt{k}.\operatorname{conv}\big(\Phi(A_\pm)\big)\Big) \leqslant 2\varepsilon$

Using the oblivious subspace embedding from Theorem 2.3 in (Woodruff, 2014), we get the following theorem:

Theorem G.2 (Implicit in prior work)**.** *Fix $\varepsilon, \delta \in (0,1)$ and positive integers d, k, m and a matrix $\mathbf{A} \in \mathbb{R}^{d \times m}$. Then, there is a tester with query complexity $O(k\varepsilon^{-2} \log(m/\delta))$ that, for an input vector $\mathbf{y} \in \mathbb{R}^d$, accepts with probability 1 if $\mathbf{y} = \mathbf{A}\mathbf{x}$ for some k-sparse $\mathbf{x}$ and rejects with probability at least $1 - \delta$ if $\|\mathbf{y} - \mathbf{A}\mathbf{x}\| > \varepsilon$ for all k-sparse $\mathbf{x}$. The running time of the tester is the time required to solve Equation (30).*

Unfortunately, for general design matrices $\mathbf{A}$, solving the optimization problem in Equation (30) is NP-hard. When $\mathbf{A}$ satisfies (ε, k)-RIP, then it is easy to verify that with probability at least $1 - \delta$, $\mathbf{SA}$ is also $(O(\varepsilon), k)$-RIP. In such cases, which as we explain next, Equation (30) can be solved efficiently, which in turn implies that the above property tester has polynomial running time.

H. Tolerant testers for Known and Unknown Designs

The simplicity of the testers for the known and unknown design settings directly translates to their robustness to noise. In this section, we state and prove our results for the tolerant variants of these problems.

Theorem H.1. *Fix $\varepsilon \in (0,1)$ and positive integers d, k, m and a matrix $\mathbf{A} \in \mathbb{R}^{d \times m}$ such that $\|\mathbf{a}_i\| = 1$ for every $i \in [m]$. There exists a randomized testing algorithm which makes linear queries to the input vector $\mathbf{y} \in \mathbb{R}^d$ and has the following properties:*

- **Completeness***: If $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{e}$ for some $\mathbf{x} \in \mathsf{Sp}_k^m$ such that $\|\mathbf{e}\| \leqslant \varepsilon$, then the tester accepts with probability $\geqslant 1 - \delta$.*
- **Soundness***: If $\|\mathbf{A}\mathbf{x} - \mathbf{y}\| > \varepsilon$ for every $\mathbf{x} : \|\mathbf{x}\|_0 \leqslant K$, then the tester rejects with probability $\geqslant 1 - \delta$. Here, $K = O(k/\varepsilon^2)$.*

The query complexity of the tester is $O(k\varepsilon^{-2} \log \frac{m}{\delta})$.

The tolerant testing algorithm is the following:

The difference here is in the final step, where instead of checking exact membership of the point $\tilde{\mathbf{y}}$ inside the convex hull, we check if the point is close enough to it. We now prove Theorem H.1:

Proof. We again consider the set $A_{\varepsilon/\sqrt{k}}$ from the soundness analysis of Theorem 1.3. As before, by our choice of n, with probability at least $1 - \delta/2$, the set $\Phi\Big(\{\mathbf{y}\} \cup A_{\varepsilon/\sqrt{k}}\Big)$ is ε-isometric to $\{\mathbf{y}\} \cup A_{\varepsilon/\sqrt{k}}$. Given this observation, for completeness we observe that

$$\begin{aligned} \mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{e} \quad &\Rightarrow \quad \operatorname{dist}\Big(\mathbf{y}, \sqrt{k} \cdot \operatorname{conv}\big(A_\pm\big)\Big) \leqslant \varepsilon \\ &\Rightarrow \quad \operatorname{dist}\Big(\mathbf{y}, A_{\varepsilon/\sqrt{k}}\Big) \leqslant 2\varepsilon \\ &\overset{1}{\Rightarrow} \quad \operatorname{dist}\Big(\tilde{\mathbf{y}}, \sqrt{k}.\operatorname{conv}\big(\Phi(A_\pm)\big)\Big) \leqslant \varepsilon(1+\varepsilon) \leqslant 2\varepsilon \end{aligned}$$

where 1 follows from the ε-isometry guarantee, and hence the tester accepts. The arguments for the soundness direction are identical to the ones used in Theorem 1.3, and hence the claim follows. □

The noise-tolerant algorithm for testing sparsity in the unknown design setting is the same as the one for Theorem 1.2. Hence, we just state and prove the guarantees in the noisy setting:

Theorem H.2 (Testing Noisy Sparse representations). *Fix $\varepsilon, \eta, \delta \in (0,1)$ and positive integers d, k, m and p, such that $(k/d)^{1/8} < \varepsilon < \frac{1}{100}, k \geqslant C' \log \frac{1}{\varepsilon}$ and $m \geqslant 20k\varepsilon^{-4}, \eta \leqslant (\sqrt{2}-1)\frac{\omega(\mathsf{Sp}_k^m)}{\sqrt{\log p}}$. There exists a randomized testing algorithm which makes linear queries to input vectors $\hat{\mathbf{y}}_1, \hat{\mathbf{y}}_2, \dots, \hat{\mathbf{y}}_p \in \mathbb{R}^d$ and has the following properties (where $\widehat{\mathbf{Y}}$ is the matrix having $\hat{\mathbf{y}}_1, \hat{\mathbf{y}}_2, \dots, \hat{\mathbf{y}}_p$ as columns):*

- **Completeness***: If there exists Y with $\|\mathbf{y}_i - \hat{\mathbf{y}}_i\| \leqslant \eta$ for every $i \in [p]$ and $Y = AX$ such for some (ε, k)-RIP matrix $A \in \mathbb{R}^{d \times m}$ and $X \in \mathbb{R}^{m \times p}$ with each column of X in Sp_k^m, then the tester accepts with probability $\geqslant 1 - \delta$.*
- **Soundness***: If $\mathbf{Y}$ does not admit factorization $\mathbf{Y} = \mathbf{A}(\mathbf{X} + \mathbf{Z}) + \mathbf{W}$ with*
  1. *The design matrix $\mathbf{A} \in \mathbb{R}^{d \times m}$ being (ε, k)-RIP, with $\|\mathbf{a}_i\| = 1$ for every $i \in [m]$*
  2. *The coefficient matrix $\mathbf{X} \in \mathbb{R}^{m \times p}$ being column wise ℓ-sparse, where $\ell = O(k/\varepsilon^4)$.*
  3. *The error matrices $\mathbf{Z} \in \mathbb{R}^{m \times p}$ and $\mathbf{W} \in \mathbb{R}^{d \times p}$ satisfying*

  $$\|\mathbf{z}_i\|_\infty \leqslant \varepsilon^2, \qquad \|\mathbf{w}_i\|_2 \leqslant O(\varepsilon^{1/4}) \qquad \text{for all } i \in [p].$$

  Then the tester rejects with probability $\geqslant 1 - \delta$.

The query complexity of the tester is $O(\varepsilon^{-2} \log(p/\delta))$.

Proof. For the completeness, let there be a $Y \in \mathbb{R}^{d \times n}$ s.t. $d_{\mathcal{H}}(Y, \hat{Y}) \leqslant \eta$ i.e., Y is column wise η-close to $\hat{Y}$ in the ℓ_2-norm, as in the completeness criteria. We can then upper bound the gaussian width of the perturbed set as :

$$\begin{aligned}
\omega(\widehat{Y}) = \mathop{\mathbf{E}}_{\mathbf{g}}\left[\max_{\hat{\mathbf{y}} \in \hat{Y}} \mathbf{g}^\top \hat{\mathbf{y}}\right] &= \mathop{\mathbf{E}}_{\mathbf{g}}\left[\max_{\hat{\mathbf{y}} \in \hat{Y}} \mathbf{g}^\top \mathbf{y} + \mathbf{g}^\top(\hat{\mathbf{y}} - \mathbf{y})\right] \\
&\leqslant \mathop{\mathbf{E}}_{\mathbf{g}}\left[\max_{\hat{\mathbf{y}} \in \hat{Y}} \mathbf{g}^\top \mathbf{y}\right] + \mathop{\mathbf{E}}_{\mathbf{g}}\left[\mathbf{g}^\top(\hat{\mathbf{y}} - \mathbf{y})\right] \\
&\overset{1}{\leqslant} \mathop{\mathbf{E}}_{\mathbf{g}}\left[\max_{\mathbf{y} \in Y} \mathbf{g}^\top \mathbf{y}\right] + C\eta\sqrt{\log p} \\
&\leqslant \omega(\mathsf{Sp}_{2k})
\end{aligned}$$

where step 1 follows from the observation that the maximum of p (not-necessarily i.i.d) gaussians (with variance at most η^2) is upper bounded by $O(\eta\sqrt{\log p})$, and the last step follows from our choice of η. Now as in Theorem 1.2, with high probability the gaussian width estimated by the tester is at most $\omega(\mathsf{Sp}_{4k}^m)$ and therefore the tester accepts. For the soundness, if the tester accepts with high probability then $\omega(\widehat{Y}) \leqslant \omega(\mathsf{Sp}_{6k}^m)$, and therefore soundness follows using arguments identical to the main theorem. □

Remark H.3. *Note that the noise model being considered here is adversarial as opposed to the standard gaussian noise. This is a relatively stronger assumption in the sense that an adversary can perturb the vectors depending on the instance i.e., the noise here can be worst case.*