A APPENDIX - SUPPLEMENTARY MATERIAL

A.1 Additional Technical Tools

We state below two results that we use frequently in our proofs. The first is well-known consequence of the CS decomposition. It relates the canonical angles between subspaces to the singular values of products and differences of their corresponding projection matrices.

Lemma A.1.1 (Stewart and Sun [22, Theorem I.5.5]). *Let $\mathcal{X}$ and $\mathcal{Y}$ be k-dimensional subspaces of $\mathbb{R}^p$ with orthogonal projections $\Pi_{\mathcal{X}}$ and $\Pi_{\mathcal{Y}}$. Let $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_k$ be the sines of the canonical angles between $\mathcal{X}$ and $\mathcal{Y}$. Then*

1. *The singular values of $\Pi_{\mathcal{X}}(I_p - \Pi_{\mathcal{Y}})$ are*

$$\sigma_1, \sigma_2, \ldots, \sigma_k, 0, \ldots, 0 .$$

2. *The singular values of $\Pi_{\mathcal{X}} - \Pi_{\mathcal{Y}}$ are*

$$\sigma_1, \sigma_1, \sigma_2, \sigma_2, \ldots, \sigma_k, \sigma_k, 0, \ldots, 0 .$$

Lemma A.1.2. *Let $x, y \in \mathbb{S}_2^{p-1}$. Then*

$$\|xx^T - yy^T\|_F^2 \leq 2\|x - y\|_2^2$$

If in addition $\|x - y\|_2 \leq \sqrt{2}$, then

$$\|xx^T - yy^T\|_F^2 \geq \|x - y\|_2^2$$

Proof. By Lemma A.1.1 and the polarization identity

$$\begin{aligned}\frac{1}{2}\|xx^T - yy^T\|_F^2 &= 1 - (x^T y)^2 \\ &= 1 - \left(\frac{2 - \|x - y\|^2}{2}\right)^2 \\ &= \|x - y\|_2^2 - \|x - y\|_2^4/4 \\ &= \|x - y\|_2^2(1 - \|x - y\|_2^2/4) .\end{aligned}$$

The upper bound follows immediately. Now if $\|x - y\|_2^2 \leq 2$, then the above right-hand side is bounded from below by $\|x - y\|_2^2/2$. ∎

A.2 Proofs for Theorem 2.1

Proof of Lemma 3.1.2 Our construction is based on a hypercube argument. We require a variation of the Varshamov-Gilbert bound due to Birgé and Massart [4]. We use a specialization of the version that appears in [15, Lemma 4.10].

Lemma. *Let d be an integer satisfying $1 \leq d \leq (p-1)/4$. There exists a subset $\Omega_d \subset \{0,1\}^{p-1}$ that satisfies the following properties:*

1. *$\|\omega\|_0 = d$ for all $\omega \in \Omega_d$,*
2. *$\|\omega - \omega'\|_0 > d/2$ for all distinct pairs $\omega, \omega' \in \Omega_d$, and*
3. *$\log|\Omega_d| \geq cd\log((p-1)/d)$, where $c \geq 0.233$.*

Let $d \in [1, (p-1)/4]$ be an integer, Ω_d be the corresponding subset of $\{0,1\}^{p-1}$ given by preceding lemma,

$$x(\omega) = \left((1 - \epsilon^2)^{\frac{1}{2}}, \epsilon\omega d^{-\frac{1}{2}}\right) \in \mathbb{R}^p ,$$

and

$$\Theta = \{x(\omega) : \omega \in \Omega_d\} .$$

Clearly, Θ satisfies the following properties:

1. $\Theta \subseteq \mathbb{S}_2^{p-1}$,
2. $\epsilon/\sqrt{2} < \|\theta_1 - \theta_2\|_2 \leq \sqrt{2}\epsilon$ for all distinct pairs $\theta_1, \theta_2 \in \Theta_d$,
3. $\|\theta\|_q^q \leq 1 + \epsilon^q d^{(2-q)/2}$ for all $\theta \in \Theta$, and
4. $\log|\Theta| \geq cd[\log(p-1) - \log d]$, where $c \geq 0.233$.

To ensure that Θ is also contained in $\mathbb{B}_q^p(R_q)$, we will choose d so that the right side of the upper bound in item 3 is smaller than R_q. Choose

$$d = \left\lfloor \min\left\{(p-1)/4, \left(\bar{R}_q/\epsilon^q\right)^{\frac{2}{2-q}}\right\}\right\rfloor .$$

The assumptions that $p \geq 5$, $\epsilon \leq 1$, and $\bar{R}_q \geq 1$ guarantee that this is a valid choice satisfying $d \in [1, (p-1)/4]$. The choice also guarantees that $\Theta \subset \mathbb{B}_q^p(R_q)$, because

$$\begin{aligned}\|\theta\|_q^q &\leq 1 + \epsilon^q d^{(2-q)/2} \\ &\leq 1 + \epsilon^q \left(\bar{R}_q/\epsilon^q\right) = R_q\end{aligned}$$

for all $\theta \in \Theta$. To complete the proof we will show that $\log|\Theta|$ satisfies the lower bound claimed by the lemma. Note that the function $a \mapsto a\log[(p-1)/a]$ is increasing on $[0, (p-1)/e]$ and decreasing on $[(p-1)/e, \infty)$. So if

$$a := \left(\frac{\bar{R}_q}{\epsilon^q}\right)^{\frac{2}{2-q}} \leq \frac{p-1}{4} ,$$

then

$$\begin{aligned}\log|\Theta| &\geq cd\left[\log(p-1) - \log d\right] \\ &\geq (c/2)a\left[\log(p-1) - \log a\right] ,\end{aligned}$$

because $d = \lfloor a \rfloor \geq a/2$. Moreover, since $d \leq (p-1)/4$ and the above right hand side is maximized when $a =$

$(p-1)/e$, the inequality remains valid for all $a \geq 0$ if we replace the constant $(c/2)$ with the constant

$$\begin{aligned} c' &= (c/2)\frac{\frac{p-1}{4}[\log(p-1) - \log\frac{p-1}{4}]}{\frac{p-1}{e}[\log(p-1) - \log\frac{p-1}{e}]} \\ &= (c/2)\frac{e\log 4}{4} \geq 0.109\,. \end{aligned}$$

■

Proof of Lemma 3.1.3 Let $A_i = x_i x_i^T$ for $i = 1, 2$. Then $\Sigma_i = \lambda_1 A_i + \lambda_2(I_p - A_i)$. Since Σ_1 and Σ_2 have the same eigenvalues and hence the same determinant,

$$\begin{aligned} D(\mathbb{P}_1\|\mathbb{P}_2) &= \frac{n}{2}\left[\mathrm{Tr}(\Sigma_2^{-1}\Sigma_1) - p - \log\det(\Sigma_2^{-1}\Sigma_1)\right] \\ &= \frac{n}{2}\left[\mathrm{Tr}(\Sigma_2^{-1}\Sigma_1) - p\right] \\ &= \frac{n}{2}\,\mathrm{Tr}(\Sigma_2^{-1}(\Sigma_1 - \Sigma_2))\,. \end{aligned}$$

The spectral decomposition $\Sigma_2 = \lambda_1 A_2 + \lambda_2(I_p - A_2)$ allows us to easily calculate that

$$\Sigma_2^{-1} = \lambda_2^{-1}(I_p - A_2) + \lambda_1^{-1}A_2\,.$$

Since orthogonal projections are idempotent, i.e. $A_i A_i = A_i$,

$$\begin{aligned} &\Sigma_2^{-1}(\Sigma_1 - \Sigma_2) \\ &= \frac{\lambda_1 - \lambda_2}{\lambda_1}[(\lambda_1/\lambda_2)(I_p - A_2) + A_2](A_1 - A_2) \\ &= \frac{\lambda_1 - \lambda_2}{\lambda_1}[(\lambda_1/\lambda_2)(I_p - A_2)A_1 - A_2(A_2 - A_1)] \\ &= \frac{\lambda_1 - \lambda_2}{\lambda_1}[(\lambda_1/\lambda_2)(I_p - A_2)A_1 - A_2(I_p - A_1)]\,. \end{aligned}$$

Using again the idempotent property and symmetry of projection matrices,

$$\begin{aligned} &\mathrm{Tr}((I_p - A_2)A_1) \\ &= \mathrm{Tr}((I_p - A_2)(I_p - A_2)A_1A_1) \\ &= \mathrm{Tr}(A_1(I_p - A_2)(I_p - A_2)A_1) \\ &= \|A_1(I_p - A_2)\|_F^2 \end{aligned}$$

and similarly,

$$\mathrm{Tr}(A_2(I_p - A_1)) = \|A_2(I_p - A_1)\|_F^2\,.$$

By Lemma A.1.1,

$$\|A_1(I_p - A_2)\|_F^2 = \|A_2(I_p - A_1)\|_F^2 = \frac{1}{2}\|A_1 - A_2\|_F^2\,.$$

Thus,

$$\mathrm{Tr}(\Sigma_2^{-1}(\Sigma_1 - \Sigma_2)) = \frac{(\lambda_1 - \lambda_2)^2}{2\lambda_1\lambda_2}\|A_1 - A_2\|_F^2$$

and the result follows. ■

A.3 Proofs for Theorem 2.2

Proof of Lemma 3.2.1 We begin with the expansion,

$$\begin{aligned} &\langle\Sigma, \theta_1\theta_1^T - \theta\theta^T\rangle \\ &= \mathrm{Tr}\{\Sigma\theta_1\theta_1^T\} - \mathrm{Tr}\{\Sigma\theta\theta^T\} \\ &= \mathrm{Tr}\{\Sigma(I_p - \theta\theta^T)\theta_1\theta_1^T\} - \mathrm{Tr}\{\Sigma\theta\theta^T(I_p - \theta_1\theta_1^T)\}\,. \end{aligned}$$

Since θ_1 is an eigenvector of Σ corresponding to the eigenvalue λ_1,

$$\begin{aligned} &\mathrm{Tr}\{\Sigma(I_p - \theta\theta^T)\theta_1\theta_1^T\} \\ &= \mathrm{Tr}\{\theta_1\theta_1^T\Sigma(I_p - \theta\theta^T)\theta_1\theta_1^T\} \\ &= \lambda_1\,\mathrm{Tr}\{\theta_1\theta_1^T(I_p - \theta\theta^T)\theta_1\theta_1^T\} \\ &= \lambda_1\,\mathrm{Tr}\{\theta_1\theta_1^T(I_p - \theta\theta^T)^2\theta_1\theta_1^T\} \\ &= \lambda_1\|\theta_1\theta_1^T(I_p - \theta\theta^T)\|_F^2\,. \end{aligned}$$

Similarly, we have

$$\begin{aligned} &\mathrm{Tr}\{\Sigma\theta\theta^T(I_p - \theta_1\theta_1^T)\} \\ &= \mathrm{Tr}\{(I_p - \theta_1\theta_1^T)\Sigma\theta\theta^T(I_p - \theta_1\theta_1^T)\} \\ &= \mathrm{Tr}\{\theta^T(I_p - \theta_1\theta_1^T)\Sigma(I_p - \theta_1\theta_1^T)\theta\} \\ &\leq \lambda_2\,\mathrm{Tr}\{\theta^T(I_p - \theta_1\theta_1^T)^2\theta\} \\ &= \lambda_2\|\theta\theta^T(I_p - \theta_1\theta_1^T)\|_F^2\,. \end{aligned}$$

Thus,

$$\begin{aligned} \langle\Sigma, \theta_1\theta_1^T - \theta\theta^T\rangle &\geq (\lambda_1 - \lambda_2)\|\theta\theta^T(I_p - \theta_1\theta_1^T)\|_F^2 \\ &= \frac{1}{2}(\lambda_1 - \lambda_2)\|\theta\theta^T - \theta_1\theta_1^T\|_F^2\,. \end{aligned}$$

The last inequality follows from Lemma A.1.1. ■

Proof of Lemma 3.2.2 Since the distribution of $S - \Sigma$ does not depend on $\mu = \mathbb{E}X_i$, we assume without loss of generality that $\mu = 0$. Let $a, b \in \{1, \ldots, p\}$ and

$$\begin{aligned} D_{ab} &= \frac{1}{n}\sum_{i=1}^n (X_m)_a(X_m)_b - \Sigma_{ab} \\ &=: \frac{1}{n}\sum_{i=1}^n \zeta_i - \mathbb{E}\zeta_i\,. \end{aligned}$$

Then

$$(S - \Sigma)_{ab} = D_{ab} - \bar{X}_a\bar{X}_b\,.$$

Using the elementary inequality $2|ab| \leq a^2 + b^2$, we have by Assumption 2.2 that

$$\begin{aligned} \|\zeta_i\|_{\psi_1} &= \|\langle X_i, 1_a\rangle\langle X_i, 1_b\rangle\|_{\psi_1} \\ &\leq \max_a \||\langle X_i, 1_a\rangle|^2\|_{\psi_1} \\ &\leq 2\max_a \|\langle\Sigma^{1/2}Z_i, 1_a\rangle\|_{\psi_2}^2 \\ &\leq 2\lambda_1 K^2\,. \end{aligned}$$

In the third line, we used the fact that the ψ_1-norm is bounded above by a constant times the ψ_2-norm [see 24, p. 95]. By a generalization of Bernstein's Inequality for the ψ_1-norm [see 24, Section 2.2] , for all $t > 0$

$$\begin{aligned}\mathbb{P}(|D_{ab}| > 8t\lambda_1 K^2) &\leq \mathbb{P}(|(D_{ab}| > 4t\|\zeta_i\|_{\psi_1}) \\ &\leq 2\exp(-n\min\{t,t^2\}/2)\,.\end{aligned}$$

This implies [24, Lemma 2.2.10] the bound

$$\begin{aligned}&\big\|\max_{ab}|D_{ab}|\big\|_{\psi_1} \\ &\leq cK^2\lambda_1\max\left\{\sqrt{\frac{\log p}{n}},\frac{\log p}{n}\right\}.\end{aligned} \tag{22}$$

Similarly,

$$\begin{aligned}2\|\bar{X}_a\bar{X}_b\|_{\psi_1} &\leq \||\langle\bar{X},1_a\rangle|^2\|_{\psi_1} + \||\langle\bar{X},1_b\rangle|^2\|_{\psi_1} \\ &\leq \|\langle\bar{X},1_a\rangle\|_{\psi_2}^2 + \|\langle\bar{X},1_b\rangle\|_{\psi_2}^2 \\ &\leq \frac{2}{n^2}\sum_{i=1}^n\|\langle X_i,1_a\rangle\|_{\psi_2}^2 + \|\langle X_i,1_b\rangle\|_{\psi_2}^2 \\ &\leq \frac{4}{n}\lambda_1 K^2\,.\end{aligned}$$

So by a union bound [24, Lemma 2.2.2],

$$\big\|\max_{ab}|\bar{X}_a\bar{X}_b|\big\|_{\psi_1} \leq cK^2\lambda_1\frac{\log p}{n}\,. \tag{23}$$

Adding eqs. (22) and (23) and then adjusting the constant c gives the desired result, because

$$\begin{aligned}&\big\|\|\mathrm{vec}(S-\Sigma)\|_\infty\big\|_{\psi_1} \\ &\leq \big\|\max_{ab}|D_{ab}|\big\|_{\psi_1} + \big\|\max_{ab}|\bar{X}_a\bar{X}_b|\big\|_{\psi_1}\,.\end{aligned}$$

■

Proof of Lemma 3.2.3 Let $B = \mathbb{S}_2^{p-1}\cap\mathbb{B}_1^p(R_1)$. We will use a recent result in empirical process theory due to Mendelson [16] to bound

$$\sup_{b\in B} b^T(S-\Sigma)b\,.$$

The result uses Talagrand's generic chaining method, and allows us to reduce the problem to bounding the supremum of a Gaussian process. The statement of the result involves the generic chaining complexity, $\gamma_2(B,d)$, of a set B equipped with the metric d. We only use a special case, $\gamma_2(B,\|\cdot\|_2)$, where the complexity measure is equivalent to the expectation of the supremum of a Gaussian process on B. We refer the reader to [23] for a complete introduction.

Lemma A.3.1 (Mendelson [16])**.** *Let Z_i, $i = 1,\ldots,n$ be i.i.d. random variables. There exists an absolute constant c for which the following holds. If $\mathcal{F}$ is a symmetric class of mean-zero functions then*

$$\begin{aligned}&\mathbb{E}\sup_{f\in\mathcal{F}}\left|\frac{1}{n}\sum_{i=1}^n f^2(Z_i) - \mathbb{E}f^2(Z_i)\right| \\ &\quad\leq c\max\left\{d_{\psi_1}\frac{\gamma_2(\mathcal{F},\psi_2)}{\sqrt{n}},\frac{\gamma_2^2(\mathcal{F},\psi_2)}{n}\right\},\end{aligned}$$

where $d_{\psi_1} = \sup_{f\in\mathcal{F}}\|f\|_{\psi_1}$.

Since the distribution of $S-\Sigma$ does not depend on $\mu = \mathbb{E}X_i$, we assume without loss of generality that $\mu = 0$. Then $|b^T(S-\Sigma)b|$ is bounded from above by a sum of two terms,

$$\left|b^T\left(\frac{1}{n}\sum_{i=1}^n X_iX_i^T - \Sigma\right)b\right| + b^T\bar{X}\bar{X}^T b\,,$$

which can be rewritten as

$$D_1(b) := \left|\frac{1}{n}\sum_{i=1}^n\langle Z_i,\Sigma^{1/2}b\rangle^2 - \mathbb{E}\langle Z_i,\Sigma^{1/2}b\rangle^2\right|$$

and $D_2(b) := \langle\bar{Z},\Sigma^{1/2}b\rangle^2$, respectively. To apply Lemma A.3.1 to D_1, define the class of linear functionals

$$\mathcal{F} := \{\langle\cdot,\Sigma^{1/2}b\rangle : b\in B\}\,.$$

Then

$$\sup_{b\in B} D_1(b) = \sup_{f\in\mathcal{F}}\left|\frac{1}{n}\sum_{i=1}^n f^2(Z_i) - \mathbb{E}f^2(Z_i)\right|,$$

and we are in the setting of Lemma A.3.1.

First, we bound the ψ_1-diameter of $\mathcal{F}$.

$$\begin{aligned}d_{\psi_1} &= \sup_{b\in B}\|\langle Z_i,\Sigma^{1/2}b\rangle\|_{\psi_1} \\ &\leq c\sup_{b\in B}\|\langle Z_i,\Sigma^{1/2}b\rangle\|_{\psi_2}\,.\end{aligned}$$

By Assumption 2.2,

$$\|\langle Z_i,\Sigma^{1/2}b\rangle\|_{\psi_2} \leq K\|\Sigma^{1/2}b\|_2 \leq K\lambda_1^{1/2}$$

and so

$$d_{\psi_1} \leq cK\lambda_1^{1/2}\,. \tag{24}$$

Next, we bound $\gamma_2(\mathcal{F},\psi_2)$ by showing that the metric induced by the ψ_2-norm on $\mathcal{F}$ is equivalent to the Euclidean metric on B. This will allow us to reduce the problem to bounding the supremum of a Gaussian process. For any $f,g\in\mathcal{F}$, by Assumption 2.2,

$$\begin{aligned}\|(f-g)(Z_i)\|_{\psi_2} &= \|\langle Z_i,\Sigma^{1/2}(b_f-b_g)\rangle\|_{\psi_2} \\ &\leq K\|\Sigma^{1/2}(b_f-b_g)\|_2 \\ &\leq K\lambda_1^{1/2}\|b_f-b_g\|_2\,,\end{aligned} \tag{25}$$

where $b_f, b_g \in B$. Thus, by [23, Theorem 1.3.6],

$$\gamma_2(\mathcal{F}, \psi_2) \le cK\lambda_1^{1/2}\gamma_2(B, \|\cdot\|_2)\,.$$

Then applying Talagrand's Majorizing Measure Theorem [23, Theorem 2.1.1] yields

$$\gamma_2(\mathcal{F}, \psi_2) \le cK\lambda_1^{1/2}\mathbb{E}\sup_{b\in B}\langle Y, b\rangle\,, \tag{26}$$

where Y is a p-dimensional standard Gaussian random vector. Recall that $B = \mathbb{B}_1^p(R_1) \cap \mathbb{S}_2^{p-1}$. So

$$\mathbb{E}\sup_{b\in B}\langle Y, b\rangle \le \mathbb{E}\sup_{b\in\mathbb{B}_1^p(R_1)\cap\mathbb{B}_2^p(1)}\langle Y, b\rangle\,.$$

Here, we could easily upper bound the above quantity by the supremum over $\mathbb{B}_1^p(R_q)$ alone. Instead, we use a sharper upper bound due to Gordon et al. [8, Theorem 5.1]:

$$\begin{aligned}\mathbb{E}\sup_{b\in\mathbb{B}_1^p(R_1)\cap\mathbb{B}_2^p(1)}\langle Y, b\rangle &\le R_1\sqrt{2+\log(2p/R_1^2)}\\ &\le 2R_1\sqrt{\log(p/R_1^2)}\,,\end{aligned}$$

where we used the assumption that $R_1^2 \le p/e$ in the last inequality. Now we apply Lemma A.3.1 to get

$$\begin{aligned}&\mathbb{E}\sup_{b\in B} D_1(B)\\ &\le cK^2\lambda_1\max\left\{R_1\sqrt{\frac{\log(p/R_1^2)}{n}}, R_1^2\frac{\log(p/R_1^2)}{n}\right\}\,.\end{aligned}$$

Turning to $D_2(b)$, we can take $n = 1$ in Lemma A.3.1 and use a similar argument as above, because

$$D_2(b) \le |\langle\bar{Z}, \Sigma^{1/2}b\rangle^2 - \mathbb{E}\langle\bar{Z}, \Sigma^{1/2}b\rangle^2| + \mathbb{E}\langle\bar{Z}, \Sigma^{1/2}b\rangle^2\,.$$

We just need to bound the ψ_2-norms of $f(\bar{Z})$ and $(f-g)(\bar{Z})$ to get bounds that are analogous to eqs. (24) and (25). Since $\bar{Z}$ is the sum of the independent random variables Z_i/n,

$$\begin{aligned}\sup_{b\in B}\|f(\bar{Z})\|_{\psi_2}^2 &= \sup_{b\in B}\|\langle\bar{Z}, \Sigma^{1/2}b_f\rangle\|_{\psi_2}^2\\ &\le \sup_{b\in B} c\sum_{i=1}^n\|\langle Z_i, \Sigma^{1/2}b_f\rangle\|_{\psi_2}^2/n^2\\ &\le \sup_{b\in B} cK^2\lambda_1\|b_f\|_2^2/n\\ &\le cK^2\lambda_1/n\,,\end{aligned}$$

and similarly,

$$\|(f-g)(\bar{Z})\|_{\psi_2} \le cK\lambda_1\|b_f - b_g\|_2^2/n\,.$$

So repeating the same arguments as for D_1, we get a similar bound for D_2. Finally, we bound $\mathbb{E}D_2(b)$ by

$$\begin{aligned}\mathbb{E}\langle\bar{X}, b\rangle^2 &= b^T\Big(\sum_{i=1}^n\sum_{j=1}^n\mathbb{E}X_iX_j^T/n^2\Big)b\\ &= b^T\Big(\sum_{i=1}^n\mathbb{E}X_iX_i^T/n^2\Big)b\\ &= \|\Sigma^{1/2}b\|_2^2/n\\ &\le \lambda_1/n\,.\end{aligned}$$

Putting together the bounds for D_1 and D_2 and then adjusting constants completes the proof. ■

Proof of Lemma 3.2.4 Using a similar argument as in the proof of Lemma 3.2.3 we can show that

$$\mathbb{E}\sup_{b\in\mathbb{S}_2^{p-1}\cap\mathbb{B}_0^p(d)}|b^T(S-\Sigma)b| \le cK^2\lambda_1\max\left\{\frac{A}{\sqrt{n}}, \frac{A^2}{n}\right\}\,,$$

where

$$A = \mathbb{E}\sup_{\mathbb{S}_2^{p-1}\cap\mathbb{B}_0^p(d)}\langle Y, b\rangle$$

and Y is a p-dimensional standard Gaussian Y. Thus we can reduce the problem to bounding the supremum of a Gaussian process.

Let $\mathcal{N} \subset \mathbb{S}_2^{p-1}\cap\mathbb{B}_0^p(d)$ be a minimal δ-covering of $\mathbb{S}_2^{p-1}\cap\mathbb{B}_0^p(d)$ in the Euclidean metric with the property that for each $x \in \mathbb{S}_2^{p-1}\cap\mathbb{B}_0^p(d)$ there exists $y \in \mathcal{N}$ satisfying $\|x-y\|_2 \le \delta$ and $x - y \in \mathbb{B}_0^p(d)$. (We will show later that such a covering exists.)

Let $b^* \in \mathbb{S}_2^{p-1}\cap\mathbb{B}_0^p(d)$ satisfy

$$\sup_{\mathbb{S}_2^{p-1}\cap\mathbb{B}_0^p(d)}\langle Y, b\rangle = \langle Y, b^*\rangle\,.$$

Then there is $\tilde{b} \in \mathcal{N}$ such that $\|b^* - \tilde{b}\|_2 \le \delta$ and $b^* - \tilde{b} \in \mathbb{B}_0^p(d)$. Since $(b^*-\tilde{b})/\|b^*-\tilde{b}\|_2 \in \mathbb{S}_2^{p-1}\cap\mathbb{B}_0^p(d)$,

$$\begin{aligned}\langle Y, b^*\rangle &= \langle Y, b^*-\tilde{b}\rangle + \langle Y, \tilde{b}\rangle\\ &\le \delta\sup_{u\in\mathbb{S}_2^{p-1}\cap\mathbb{B}_0^p(d)}\langle Y, u\rangle + \langle Y, \tilde{b}\rangle\\ &\le \delta\langle Y, b^*\rangle + \max_{b\in\mathcal{N}}\langle Y, b\rangle\,.\end{aligned}$$

Thus,

$$\sup_{b\in\mathbb{S}_2^{p-1}\cap\mathbb{B}_0^p(d)}\langle Y, b\rangle \le (1-\delta)^{-1}\max_{b\in\mathcal{N}}\langle Y, b\rangle\,.$$

Since $\langle Y, b\rangle$ is a standard Gaussian for every $b \in \mathcal{N}$, a union bound [24, Lemma 2.2.2] implies

$$\mathbb{E}\max_{b\in\mathcal{N}}\langle Y, b\rangle \le c\sqrt{\log|\mathcal{N}|}$$

for an absolute constant $c > 0$. Thus,

$$\mathbb{E} \sup_{b \in \mathbb{S}_2^{p-1} \cap \mathbb{B}_0^p(d)} \langle Y, b \rangle \leq c(1-\delta)^{-1}\sqrt{\log|\mathcal{N}|}$$

Finally, we will bound $\log|\mathcal{N}|$ by constructing a δ-covering set and then choosing δ. It is well known that the minimal δ-covering of $\mathbb{S}_2^{d-1}$ in the Euclidean metric has cardinality at most $(1 + 2/\delta)^d$. Associate with each subset $I \subseteq \{1, \ldots, p\}$ of size d, a minimal δ-covering of the corresponding isometric copy of $\mathbb{S}_2^{d-1}$. This set covers every possible subset of size d, so for each $x \in \mathbb{S}_2^{p-1} \cap \mathbb{B}_0(d)$ there is $y \in \mathcal{N}$ satisfying $\|x - y\|_2 \leq \delta$ and $x - y \in \mathbb{B}_0(d)$. Since there are ($p$ choose d) possible subsets,

$$\begin{aligned}
\log|\mathcal{N}| &\leq \log \binom{p}{d} + d\log(1 + 2/\delta) \\
&\leq \log\left(\frac{pe}{d}\right)^d + d\log(1 + 2/\delta) \\
&= d + d\log(p/d) + d\log(1 + 2/\delta)\,.
\end{aligned}$$

In the second line, we used the binomial coefficient bound $\binom{p}{d} \leq (ep/d)^d$. If we take $\delta = 1/4$, then

$$\begin{aligned}
\log|\mathcal{N}| &\leq d + d\log(p/d) + d\log 9 \\
&\leq cd\log(p/d)\,,
\end{aligned}$$

where we used the assumption that $d < p/2$. Thus,

$$A = \mathbb{E} \sup_{\mathbb{S}_2^{p-1} \cap \mathbb{B}_0^p(d)} \langle Y, b \rangle \leq cd\log(p/d)$$

for all $d \in [1, p/2)$. ■