

Supplementary Material for NIPS Submission 1169

Structured Local Optima in Sparse Blind Deconvolution

Yuqian Zhang, Han-Wen Kuo, John Wright

January 8, 2019

Section 1 contains additional experiment results of recovery accuracy of the ground truth kernel (rather than a shift truncation). Section 2 contains some basic lemmas for quantities used repeatedly; Section 6 presents the proofs of the main theorem and corollary of this paper. Section 4 and Section 5 proves for lemmas around the initialization point \mathbf{q}_{init} and the preconditioning term $\mathbf{Y}^T \mathbf{Y}$ (or $\mathbf{A}_0^T \mathbf{A}_0$) respectively. Finite sample concentration for the Riemannian gradient and Hessian are presented in Section 7 and Section 8 respectively.

1 Experiments

Recovery Accuracy of the Ground Truth Kernel In this section, we provide experiment results for the recovery of the ground truth kernel obtained by the annealing algorithm proposed in [ZLK⁺17]. The annealing algorithm recovers the ground truth kernel by minimizing the Lasso cost, initialized at the zero-padded shift truncated kernel rendered from Algorithm 1. The recovery accuracy presented in Figure 2 is measured as $\text{err} = \min_{\tau} \|\bar{\mathbf{a}}^{(+)} \pm s_{\tau}[\tilde{\mathbf{a}}_0]\|_2$. Here, $\bar{\mathbf{a}}^{(+)}$ denote the local minimum in the lifted optimization space.

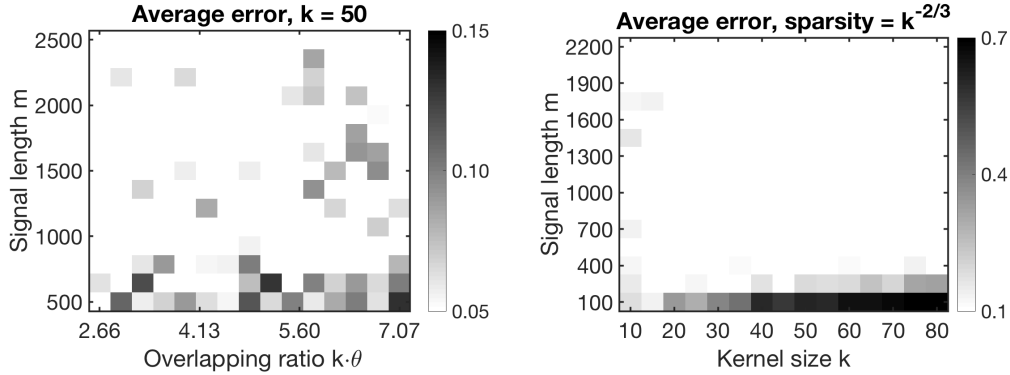


Figure 1: Recovery Error of the Ground Truth Kernel with Algorithm 1 finding a shift truncated kernel and the annealing Lasso problem recovering the ground truth kernel.

For comparison, we also present experiment results of the algorithm proposed by [ZLK⁺17], which is composed of solving two Lasso minimization problems over the original kernel sphere and lifted kernel sphere respectively.

In terms of the recovery accuracy of the ground truth kernel, Algorithm 1 proposed in this paper achieves better recovery for sparser and longer observations, while the [ZLK⁺17] manifests slight advantages when the observations is limited. As the optimization landscape studied in [ZLK⁺17] varies with different choice

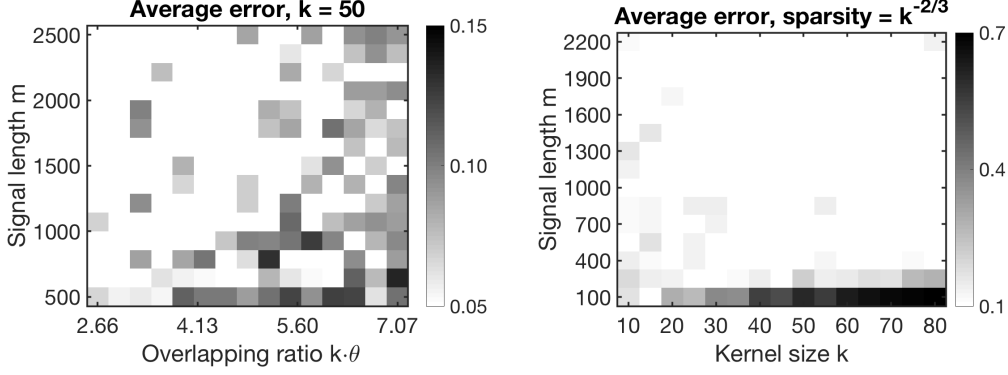


Figure 2: Recovery Error of the Ground Truth Kernel by minimizing the Lasso objective function recovering both the shift truncated kernel as well as the ground truth kernel.

of sparsity parameter λ , it is possible that experiment results for [ZLK⁺17] could be improved. On the other hand, only empirical knowledge about the choice of λ is available while there is little disciplined understanding. In contrast, Algorithm 1 does not depend on any parameter tuning and guarantees recovery once the working conditions are met.

2 Basics

Lemma 2.1 (Expectation of the Approximate Objective Function) *Assuming $\mathbf{x}_0 \sim \text{i.i.d.}$ $\text{BG}(\theta) \in \mathbb{R}^m$, then*

$$\begin{aligned} \mathbb{E}_{\mathbf{x}_0} \left[\frac{1}{m} \left\| \mathbf{Y}^T (\mathbf{A}_0 \mathbf{A}_0^T)^{-1/2} \mathbf{q} \right\|_4^4 \right] \\ = 3\theta (1 - \theta) \left\| \mathbf{A}^T \mathbf{q} \right\|_4^4 + 3\theta^2 \left\| \mathbf{A}^T \mathbf{q} \right\|_2^4. \end{aligned} \quad (2.1)$$

Proof Let $\mathbf{g} \in \mathbb{R}^{2k-1}$ be a standard random Gaussian vector and \mathbf{P}_I be the projection operator onto Bernoulli vector $I \sim \text{Ber}(\theta)$. Then any column $\mathbf{x}_i \in \mathbb{R}^{2k-1}$ of \mathbf{X}_0 is equal in distribution to $\mathbf{x}_i = \mathbf{P}_I \mathbf{g}$ with $\mathbf{g} \sim \text{i.i.d. } \mathcal{N}(0, 1)$.

$$\begin{aligned} \mathbb{E}_{\mathbf{x}_0} \left[\frac{1}{m} \left\| \mathbf{Y}^T (\mathbf{A}_0 \mathbf{A}_0^T)^{-1/2} \mathbf{q} \right\|_4^4 \right] \\ = \frac{1}{m} \mathbb{E}_I \mathbb{E}_{\mathbf{g}} \left\| \mathbf{q}^T \mathbf{A} \mathbf{X}_0 \right\|_4^4 \end{aligned} \quad (2.2)$$

$$= \mathbb{E}_I \mathbb{E}_{\mathbf{g}} \left\| \mathbf{q}^T \mathbf{A} \mathbf{x}_i \right\|_4^4 \quad (2.3)$$

$$= \mathbb{E}_I \mathbb{E}_{\mathbf{g}} \left(\mathbf{q}^T \mathbf{A} \mathbf{P}_I \mathbf{g} \right)^4 \quad (2.4)$$

$$= 3 \mathbb{E}_I \left(\mathbf{q}^T \mathbf{A} \mathbf{P}_I \mathbf{A}^T \mathbf{q} \right)^2 \quad (2.5)$$

$$= 3 \mathbb{E}_I \left(\sum_{i \in I} \langle \mathbf{a}_i, \mathbf{q} \rangle^4 + \sum_{\{i \neq j\} \in I} \langle \mathbf{a}_i, \mathbf{q} \rangle^2 \langle \mathbf{a}_j, \mathbf{q} \rangle^2 \right) \quad (2.6)$$

$$= 3\theta (1 - \theta) \left\| \mathbf{A}^T \mathbf{q} \right\|_4^4 + 3\theta^2 \left\| \mathbf{A}^T \mathbf{q} \right\|_2^4 \quad (2.7)$$

■

Lemma 2.2 (Root Estimation for Cubic Gradient Function) Consider an equation of the form

$$f(x) = x(\alpha - x^2) - \beta = 0, \quad (2.8)$$

with $\alpha > 0$. Suppose that $\beta < \frac{1}{4}\alpha^{3/2}$. Then $f(x) = 0$ has three solutions, x_1, x_2, x_3 satisfying

$$\max\{|x_1 - \sqrt{\alpha}|, |x_2 + \sqrt{\alpha}|, |x_3|\} \leq \frac{2\beta}{\alpha}. \quad (2.9)$$

Proof Suppose first that $\beta > 0$. Then $f(0) < 0$. Moreover,

$$f\left(\frac{2\beta}{\alpha}\right) = 2\beta - 8\beta^3/\alpha^3 - \beta \quad (2.10)$$

$$= \beta(1 - 8\beta^2/\alpha^3) \quad (2.11)$$

$$> 0. \quad (2.12)$$

Hence, f has at least one root in the interval $[0, \frac{2\beta}{\alpha}]$. Similarly, notice that $f(\sqrt{\alpha}) < 0$ and that

$$\begin{aligned} f\left(\sqrt{\alpha} - \frac{2\beta}{\alpha}\right) &= \alpha^{3/2} - 2\beta - (\sqrt{\alpha} - 2\beta/\alpha)^3 - \beta \\ &= \alpha^{3/2} - 3\beta - \alpha^{3/2} + 6\beta - 12\beta^2/\alpha^{3/2} + 8\beta^3/\alpha^3 \end{aligned} \quad (2.13)$$

$$= \alpha^{3/2} - 3\beta - \alpha^{3/2} + 6\beta - 12\beta^2/\alpha^{3/2} + 8\beta^3/\alpha^3 \quad (2.14)$$

$$= \beta\left(3 - \frac{12\beta}{\alpha^{3/2}} + \frac{8\beta^2}{\alpha^3}\right) \quad (2.15)$$

$$> 0. \quad (2.16)$$

Thus, there is at least one root in the interval $[\sqrt{\alpha} - \frac{2\beta}{\alpha}, \sqrt{\alpha}]$. Finally, note that $f(-\sqrt{\alpha}) < 0$, $\frac{df}{dx}(-\sqrt{\alpha}) = -2\alpha$, and $\frac{d^2f}{dx^2}(x') = -3x'$ is positive for $x' \leq -\sqrt{\alpha}$. Hence, convexity gives that

$$\begin{aligned} f\left(-\sqrt{\alpha} - \frac{2\beta}{\alpha}\right) &\geq f(-\sqrt{\alpha}) + \frac{df}{dx}(-\sqrt{\alpha}) \times (-2\beta/\alpha) \\ &= -\beta + (-2\alpha) \times (-2\beta/\alpha) \end{aligned} \quad (2.17)$$

$$= 3\beta > 0. \quad (2.18)$$

$$= 3\beta > 0. \quad (2.19)$$

Under this condition, there is at least one root in the interval, $[-\sqrt{\alpha} - 2\beta/\alpha, -\sqrt{\alpha}]$. These three intervals do not overlap, as long as $\frac{4\beta}{\alpha} < \sqrt{\alpha}$, or $\beta < \frac{1}{4}\alpha^{3/2}$.

In the case that $\beta \leq 0$, a symmetric argument applies. Thus there are exactly three solutions to equation (2.8) in the specified intervals. ■

Lemma 2.3 Let \mathbf{a}_l and $\mathbf{a}_{l'}$ be two nonzero vectors with inner product $\mu_{l,l'} \doteq \langle \mathbf{a}_l, \mathbf{a}_{l'} \rangle$. Then for any unit vector $\mathbf{v} \in \text{span}(\mathbf{a}_l, \mathbf{a}_{l'})$,

$$\left| \left\langle \frac{\mathbf{a}_l}{\|\mathbf{a}_l\|_2}, \mathbf{v} \right\rangle \right|^2 + \left| \left\langle \frac{\mathbf{a}_{l'}}{\|\mathbf{a}_{l'}\|_2}, \mathbf{v} \right\rangle \right|^2 \geq 1 - \frac{|\mu_{l,l'}|}{\|\mathbf{a}_l\|_2 \|\mathbf{a}_{l'}\|_2}. \quad (2.20)$$

Proof Let \mathbf{u} and \mathbf{u}^\perp be two orthogonal unit vectors, such that

$$\mathbf{a}_l = \|\mathbf{a}_l\|_2 \mathbf{u}, \quad (2.21)$$

$$\mathbf{a}_{l'} = \frac{\mu_{l,l'}}{\|\mathbf{a}_l\|_2} \mathbf{u} + \sqrt{\|\mathbf{a}_{l'}\|_2^2 - \frac{\mu_{l,l'}^2}{\|\mathbf{a}_l\|_2^2}} \mathbf{u}^\perp. \quad (2.22)$$

Suppose $\mathbf{v} = a\mathbf{u} + b\mathbf{u}^\perp$ with $a^2 + b^2 = 1$. Let $\mu_{\text{rel}} = \frac{\mu_{l,l'}}{\|\mathbf{a}_l\|_2 \|\mathbf{a}_{l'}\|_2}$, then we can expand the quantity of interests as

$$\begin{aligned} & \left| \left\langle \frac{\mathbf{a}_l}{\|\mathbf{a}_l\|_2}, \mathbf{v} \right\rangle \right|^2 + \left| \left\langle \frac{\mathbf{a}_{l'}}{\|\mathbf{a}_{l'}\|_2}, \mathbf{v} \right\rangle \right|^2 \\ &= \left| \langle \mathbf{u}, a\mathbf{u} + b\mathbf{u}^\perp \rangle \right|^2 \\ & \quad + \left| \left\langle \mu_{\text{rel}}\mathbf{u} + \sqrt{1 - \mu_{\text{rel}}^2}\mathbf{u}^\perp, a\mathbf{u} + b\mathbf{u}^\perp \right\rangle \right|^2 \end{aligned} \quad (2.23)$$

$$= a^2 + \left(a\mu_{\text{rel}} + b\sqrt{1 - \mu_{\text{rel}}^2} \right)^2 \quad (2.24)$$

$$= a^2 + b^2 + (a^2 - b^2)\mu_{\text{rel}}^2 + 2ab\mu_{\text{rel}}\sqrt{1 - \mu_{\text{rel}}^2} \quad (2.25)$$

$$= 1 + [a^2 - b^2, 2ab] \left[\mu_{\text{rel}}^2, \mu_{\text{rel}}\sqrt{1 - \mu_{\text{rel}}^2} \right]^T \quad (2.26)$$

Since $[a^2 - b^2, 2ab]$ is a unit vector, then above equation is lower bounded by

$$\begin{aligned} & 1 - \left\| \left[\mu_{\text{rel}}^2, \mu_{\text{rel}}\sqrt{1 - \mu_{\text{rel}}^2} \right] \right\|_2 \\ &= 1 - |\mu_{\text{rel}}| \end{aligned} \quad (2.27)$$

$$= 1 - \frac{|\mu_{l,l'}|}{\|\mathbf{a}_l\|_2 \|\mathbf{a}_{l'}\|_2} \quad (2.28)$$

as claimed. ■

Lemma 2.4 (Nonzeros in a Bernoulli Vector) *Let $\mathbf{v} \sim_{\text{i.i.d.}} \text{Ber}(\theta) \in \mathbb{R}^n$, then*

$$\mathbb{P}[\|\mathbf{v}\|_0 \geq (1+t)\theta n] \leq 2 \exp\left(-\frac{3t^2}{2t+6}\theta n\right). \quad (2.29)$$

Proof As $\|\mathbf{v}\|_0 = v_0 + \dots + v_{n-1}$, and

$$|v_i - \theta| \leq 1, \quad \mathbb{E}[(v_i - \theta)^2] = \theta(1 - \theta) \leq \theta \quad (2.30)$$

with Bernstein's inequality, we obtain that

$$\begin{aligned} & \mathbb{P}[\|\mathbf{v}\|_0 \geq (1+t)\theta n] \\ & \leq 2 \exp\left(-\frac{t^2\theta^2 n^2}{2(\theta - \theta^2)n + \frac{2}{3}t\theta n}\right) \end{aligned} \quad (2.31)$$

$$\leq 2 \exp\left(-\frac{3t^2}{2t+6}\theta n\right), \quad (2.32)$$

as claimed. ■

Lemma 2.5 (Entry-wise Truncation of a Bernoulli Gaussian Vector) *Suppose $\mathbf{x}_0 \sim_{\text{i.i.d.}} \text{BG}(\theta) \in \mathbb{R}^m$, then*

$$\mathbb{P}[\|\mathbf{x}_0\|_\infty > t] \leq 2\theta m e^{-t^2/2}. \quad (2.33)$$

Proof A Bernoulli-Gaussian variable $x = \omega \cdot g$ satisfies

$$\mathbb{P}[|x| \geq t] = \theta \cdot \mathbb{P}[|g| \geq t] \leq 2\theta e^{-t^2/2}, \quad (2.34)$$

Taking a union bound over the m entries of \mathbf{x}_0 , we obtain

$$\mathbb{P}[\|\mathbf{x}_0\|_\infty > t] \leq m\mathbb{P}[|x| > t] \quad (2.35)$$

$$\leq 2\theta m e^{-t^2/2}, \quad (2.36)$$

as claimed. \blacksquare

Lemma 2.6 (Operator Norm of a Bernoulli Gaussian Circulant Matrix) Let $\mathbf{C}_{\mathbf{x}_0} \in \mathbb{R}^{m \times m}$ be the circulant matrix generated from $\mathbf{x}_0 \sim_{\text{i.i.d. BG}}(\theta) \in \mathbb{R}^m$, then

$$\mathbb{P}[\|\mathbf{C}_{\mathbf{x}_0}\|_2 \geq t] \leq 2m \exp\left(-\frac{t^2}{2\theta m + 2t}\right). \quad (2.37)$$

Proof The operator norm of a circulant matrix is

$$\|\mathbf{C}_{\mathbf{x}_0}\|_2 = \max_l |\langle \mathbf{x}_0, \mathbf{w}_l \rangle|, \quad (2.38)$$

where \mathbf{w}_l is the l -th (discrete) Fourier basis vector

$$\mathbf{w}_l = \left[1, e^{l\frac{2\pi j}{m}}, \dots, e^{l(m-1)\frac{2\pi j}{m}}\right]^T, \quad l = 0, \dots, m-1, \quad (2.39)$$

and j is the imaginary unit. With moment control Bernstein inequality, we obtain

$$\begin{aligned} \mathbb{P}[|\langle \mathbf{x}_0, \mathbf{w}_l \rangle| \geq t] &\leq 2 \exp\left(-\frac{t^2}{2\theta \|\mathbf{w}_l\|_2^2 + 2\|\mathbf{w}_l\|_\infty t}\right) \\ &\leq 2 \exp\left(-\frac{t^2}{2\theta m + 2t}\right) \end{aligned} \quad (2.40)$$

together with the union bound,

$$\mathbb{P}[\|\mathbf{C}_{\mathbf{x}_0}\|_2 \geq t] \leq m\mathbb{P}[|\langle \mathbf{x}_0, \mathbf{w}_l \rangle| \geq t] \quad (2.41)$$

$$\leq 2m \exp\left(-\frac{t^2}{2\theta m + 2t}\right), \quad (2.42)$$

as claimed. \blacksquare

Lemma 2.7 (Norms of $\boldsymbol{\eta}$ and $\bar{\boldsymbol{\eta}}$) Suppose $\delta = \left\|\frac{1}{\theta m} \mathbf{X}_0 \mathbf{X}_0^T - \mathbf{I}\right\|_2 \leq 1/(2\kappa^2)$, then vectors $\boldsymbol{\eta} = \mathbf{Y}^T (\mathbf{Y} \mathbf{Y}^T)^{-1/2} \mathbf{q}$ and $\bar{\boldsymbol{\eta}} = \mathbf{Y}^T (\theta m \mathbf{A}_0 \mathbf{A}_0^T)^{-1/2} \mathbf{q}$ satisfy

$$\|\boldsymbol{\eta}\|_\infty \leq \left(1 + \frac{4\kappa^3 \delta}{\sigma_{\min}}\right) \left(\frac{2k}{\theta m}\right)^{1/2} \|\mathbf{x}_0\|_\infty, \quad (2.43)$$

$$\|\bar{\boldsymbol{\eta}}\|_\infty \leq \left(\frac{2k}{\theta m}\right)^{1/2} \|\mathbf{x}_0\|_\infty, \quad (2.44)$$

$$\|\boldsymbol{\eta}\|_6^6 \leq \left(1 + \frac{4\kappa^3 \delta}{\sigma_{\min}}\right)^4 \frac{4k^2}{\theta^2 m^2} \|\mathbf{x}_0\|_\infty^4, \quad (2.45)$$

$$\|\bar{\boldsymbol{\eta}}\|_2 \leq 1 + \delta/2, \quad (2.46)$$

$$\|\boldsymbol{\eta} - \bar{\boldsymbol{\eta}}\|_\infty \leq \frac{4\kappa^3 \delta}{\sigma_{\min}} \left(\frac{2k}{\theta m}\right)^{1/2} \|\mathbf{x}_0\|_\infty, \quad (2.47)$$

$$\|\boldsymbol{\eta} - \bar{\boldsymbol{\eta}}\|_2 \leq (1 + \delta/2) \frac{4\kappa^3 \delta}{\sigma_{\min}}. \quad (2.48)$$

Proof Since $\delta = \left\| \frac{1}{\theta m} \mathbf{X}_0 \mathbf{X}_0^T - \mathbf{I} \right\|_2$, then

$$\|\mathbf{X}_0\|_2 \leq (\theta m)^{1/2} \sqrt{1 + \delta} \quad (2.49)$$

$$\leq (\theta m)^{1/2} (1 + \delta/2). \quad (2.50)$$

As $\boldsymbol{\eta} = \mathbf{Y}^T (\mathbf{Y} \mathbf{Y}^T)^{-1/2} \mathbf{q} = \mathbf{X}_0^T \mathbf{A}_0^T (\mathbf{Y} \mathbf{Y}^T)^{-1/2} \mathbf{q}$, together with Lemma 5.3:

$$\begin{aligned} & \left\| \mathbf{A}_0^T (\mathbf{Y} \mathbf{Y}^T)^{-1/2} \mathbf{q} \right\|_\infty \\ & \leq \left\| \mathbf{A}_0^T (\mathbf{Y} \mathbf{Y}^T)^{-1/2} \mathbf{q} \right\|_2 \end{aligned} \quad (2.51)$$

$$\begin{aligned} & \leq \left\| \mathbf{A}_0^T \left((\mathbf{Y} \mathbf{Y}^T)^{-1/2} - (\theta m \mathbf{A}_0 \mathbf{A}_0^T)^{-1/2} \right) \mathbf{q} \right\|_2 \\ & \quad + \left\| \mathbf{A}_0^T (\theta m \mathbf{A}_0 \mathbf{A}_0^T)^{-1/2} \mathbf{q} \right\|_2 \end{aligned} \quad (2.52)$$

$$\leq (\theta m)^{-1/2} \frac{4\kappa^3 \delta}{\sigma_{\min}} \|\mathbf{q}\|_2 + (\theta m)^{-1/2} \|\mathbf{A}^T \mathbf{q}\|_2 \quad (2.53)$$

$$\leq (\theta m)^{-1/2} \left(1 + \frac{4\kappa^3 \delta}{\sigma_{\min}} \right) \quad (2.54)$$

Norms of $\boldsymbol{\eta}$. Since $\|\mathbf{X}_0 \mathbf{e}_l\|_2 \leq \sqrt{2k-1} \|\mathbf{X}_0 \mathbf{e}_l\|_\infty$, we have

$$\|\boldsymbol{\eta}\|_\infty = \max_{l \in [1, \dots, m]} \left\langle \mathbf{X}_0 \mathbf{e}_l, \mathbf{A}_0^T (\mathbf{Y} \mathbf{Y}^T)^{-1/2} \mathbf{q} \right\rangle \quad (2.55)$$

$$\leq \max_l \|\mathbf{X}_0 \mathbf{e}_l\|_2 \left\| \mathbf{A}_0^T (\mathbf{Y} \mathbf{Y}^T)^{-1/2} \mathbf{q} \right\|_2 \quad (2.56)$$

$$\leq \sqrt{2k} \|\mathbf{x}_0\|_\infty \cdot (\theta m)^{-1/2} \left(1 + \frac{4\kappa^3 \delta}{\sigma_{\min}} \right). \quad (2.57)$$

At the same time, plugging in $\|\boldsymbol{\eta}\|_2 = 1$, we have

$$\|\boldsymbol{\eta}\|_6^6 \leq \|\boldsymbol{\eta}\|_2^2 \|\boldsymbol{\eta}\|_\infty^4 \leq \left(1 + \frac{4\kappa^3 \delta}{\sigma_{\min}} \right)^4 \frac{4k^2}{\theta^2 m^2} \|\mathbf{x}_0\|_\infty^4. \quad (2.58)$$

Norms of $\bar{\boldsymbol{\eta}}$. Here, $\bar{\boldsymbol{\eta}} = \mathbf{Y}^T (\theta m \mathbf{A}_0 \mathbf{A}_0^T)^{-1/2} \mathbf{q} = \mathbf{X}_0^T \mathbf{A}_0^T (\theta m \mathbf{A}_0 \mathbf{A}_0^T)^{-1/2} \mathbf{q}$ with

$$\begin{aligned} & \left\| \mathbf{A}_0^T (\theta m \mathbf{A}_0 \mathbf{A}_0^T)^{-1/2} \mathbf{q} \right\|_\infty \\ & \leq \left\| \mathbf{A}_0^T (\theta m \mathbf{A}_0 \mathbf{A}_0^T)^{-1/2} \mathbf{q} \right\|_2 \end{aligned} \quad (2.59)$$

$$= (\theta m)^{-1/2}, \quad (2.60)$$

therefore

$$\begin{aligned} \|\bar{\boldsymbol{\eta}}\|_\infty & \leq \max_l \|\mathbf{X}_0 \mathbf{e}_l\|_2 \left\| \mathbf{A}_0 (\theta m \mathbf{A}_0 \mathbf{A}_0^T)^{-1/2} \mathbf{q} \right\|_2 \\ & \leq \left(\frac{2k}{\theta m} \right)^{1/2} \|\mathbf{x}_0\|_\infty, \end{aligned} \quad (2.61)$$

$$\begin{aligned} \|\bar{\boldsymbol{\eta}}\|_2 & \leq \|\mathbf{X}_0^T\|_2 \left\| \mathbf{A}_0 (\theta m \mathbf{A}_0 \mathbf{A}_0^T)^{-1/2} \mathbf{q} \right\|_2 \\ & \leq 1 + \delta/2. \end{aligned} \quad (2.62)$$

Norms of $\boldsymbol{\eta} - \bar{\boldsymbol{\eta}}$. With similar reasoning, we can obtain

$$\begin{aligned}
& \|\boldsymbol{\eta} - \bar{\boldsymbol{\eta}}\|_\infty \\
&= \left\| \mathbf{Y}^T (\mathbf{Y} \mathbf{Y}^T)^{-1/2} \mathbf{q} - \mathbf{Y}^T (\theta m \mathbf{A}_0 \mathbf{A}_0^T)^{-1/2} \mathbf{q} \right\|_\infty \\
&\leq \max_{l \in [1, \dots, m]} \|\mathbf{X}_0 \mathbf{e}_l\|_2 (\theta m)^{-1/2} \times \\
&\quad \left\| \mathbf{A}_0^T \left(\frac{1}{\theta m} \mathbf{Y} \mathbf{Y}^T \right)^{-1/2} - \mathbf{A}_0^T (\mathbf{A}_0 \mathbf{A}_0^T)^{-1/2} \right\|_2
\end{aligned} \tag{2.63}$$

$$\leq \frac{4\kappa^3 \delta}{\sigma_{\min}} \left(\frac{2k}{\theta m} \right)^{1/2} \|\mathbf{x}_0\|_\infty, \tag{2.64}$$

and

$$\begin{aligned}
& \|\boldsymbol{\eta} - \bar{\boldsymbol{\eta}}\|_2 \\
&\leq \|\mathbf{X}_0\|_2 (\theta m)^{-1/2} \|\mathbf{q}\|_2 \times \\
&\quad \left\| \mathbf{A}_0^T \left(\frac{1}{\theta m} \mathbf{Y} \mathbf{Y}^T \right)^{-1/2} - \mathbf{A}_0^T (\mathbf{A}_0 \mathbf{A}_0^T)^{-1/2} \right\|_2
\end{aligned} \tag{2.65}$$

$$\leq (\theta m)^{-1/2} \frac{4\kappa^3 \delta}{\sigma_{\min}} \|\mathbf{X}_0\|_2 \tag{2.66}$$

$$\leq (1 + \delta/2) \frac{4\kappa^3 \delta}{\sigma_{\min}}, \tag{2.67}$$

completing the proof. ■

3 Proof for Geometry around the Stationary Points

3.1 Local Minima

Lemma 3.1 Suppose $\bar{\mathbf{q}}$ is a stationary point in \mathcal{R}_{C_\star} ($C_\star \geq 10$), and its corresponding $\boldsymbol{\zeta}$ has only one entry ζ_l with nontrivial magnitude, then this stationary point is a local minimum near \mathbf{a}_l that $|\langle \bar{\mathbf{q}}, \mathcal{P}_{\mathbb{S}}[\mathbf{a}_l] \rangle| > 1 - 2c_\star \kappa^{-2}$ with $c_\star = 1/C_\star$.

Proof Suppose $\boldsymbol{\zeta}$ has only one big entry ζ_l , and other entries are bounded by $2\beta/\alpha$

$$\|\boldsymbol{\zeta}\|_4^4 = \zeta_l^4 + \sum_{j \neq l} \zeta_j^4 \leq \zeta_l^4 + \max_{j \neq l} \zeta_j^2 \cdot \sum_{j \neq l} \zeta_j^2 \leq \zeta_l^4 + \frac{4\mu^2 \|\boldsymbol{\zeta}\|_3^6}{\|\boldsymbol{\zeta}\|_4^8}, \tag{3.1}$$

with $\|\boldsymbol{\zeta}\|_4^6 \geq C_\star \mu \kappa^2 \|\boldsymbol{\zeta}\|_3^3$, and for simplicity let $c_\star = 1/C_\star$, we have

$$\zeta_l^4 \geq \|\boldsymbol{\zeta}\|_4^4 - \frac{4\mu^2 \|\boldsymbol{\zeta}\|_3^6}{\|\boldsymbol{\zeta}\|_4^8} \geq (1 - 4c_\star^2 \kappa^{-4}) \|\boldsymbol{\zeta}\|_4^4. \tag{3.2}$$

On the other hand, we also have

$$\zeta_l^2 \leq (\sqrt{\alpha_l} + \frac{2\beta_l}{\alpha_l})^2 \leq \frac{\|\boldsymbol{\zeta}\|_4^4}{\|\mathbf{a}_i\|_2^2} + \frac{4\mu \|\boldsymbol{\zeta}\|_3^3}{\|\mathbf{a}_i\|_2 \|\boldsymbol{\zeta}\|_4^2} + \frac{4\mu^2 \|\boldsymbol{\zeta}\|_3^6}{\|\boldsymbol{\zeta}\|_4^8} \leq \frac{\|\boldsymbol{\zeta}\|_4^4}{\|\mathbf{a}_i\|_2^2} (1 + 4c_\star \kappa^{-2} + 4c_\star^2 \kappa^{-4}). \tag{3.3}$$

Combining above two inequalities, we have

$$\zeta_l^2 \leq \frac{1 + 4c_\star \kappa^{-2} + 4c_\star^2 \kappa^{-4}}{1 - 4c_\star^2 \kappa^{-4}} \frac{\zeta_l^4}{\|\mathbf{a}_l\|_2^2}, \quad (3.4)$$

thus the local minimum $\bar{\mathbf{q}}$ is close to \mathbf{a}_l :

$$\frac{|\langle \bar{\mathbf{q}}, \mathbf{a}_l \rangle|}{\|\mathbf{a}_l\|_2} \geq \frac{\sqrt{1 - 4c_\star^2 \kappa^{-4}}}{1 + 2c_\star \kappa^{-2}} \geq 1 - 2c_\star \kappa^{-2}. \quad (3.5)$$

Next, we need to verify that the Riemannian Hessian at $\bar{\mathbf{q}}$ is definite positive, recall that

$$\text{Hess } \varphi(\bar{\mathbf{q}}) = -\mathbf{P}_{\bar{\mathbf{q}}^\perp} \left[3\mathbf{A} \text{diag}(\zeta^{\circ 2}) \mathbf{A}^T - \|\zeta\|_4^4 \mathbf{I} \right] \mathbf{P}_{\bar{\mathbf{q}}^\perp}. \quad (3.6)$$

Let \mathbf{v} be a unit vector such that $\mathbf{v} \perp \bar{\mathbf{q}}$, then

$$\mathbf{v}^T \text{Hess } \varphi(\bar{\mathbf{q}}) \mathbf{v} = -\mathbf{v}^T \left(3\mathbf{A} \text{diag}(\zeta^{\circ 2}) \mathbf{A}^T - \|\zeta\|_4^4 \mathbf{I} \right) \mathbf{v} \quad (3.7)$$

$$= \|\zeta\|_4^4 - 3\mathbf{v}^T \mathbf{A} \text{diag}(\zeta^{\circ 2}) \mathbf{A}^T \mathbf{v} \quad (3.8)$$

$$= \|\zeta\|_4^4 - 3 \langle \mathbf{a}_l, \mathbf{v} \rangle^2 \zeta_l^2 - 3 \sum_{i \neq l} \langle \mathbf{a}_i, \mathbf{v} \rangle^2 \zeta_i^2 \quad (3.9)$$

$$\geq \|\zeta\|_4^4 - 3 \langle \mathbf{a}_l, \mathbf{v} \rangle^2 \zeta_l^2 - 3 \max_{i \neq l} \zeta_i^2. \quad (3.10)$$

The last inequality is due to $\sum_{i \neq l} \langle \mathbf{a}_i, \mathbf{v} \rangle^2 \leq \|\mathbf{A}^T \mathbf{v}\|_2^2 = 1$. Since $\mathbf{v} \perp \bar{\mathbf{q}}$ and ζ_l is the only entry with nontrivial magnitude, then derive from eq. (3.5):

$$\langle \mathbf{a}_l, \mathbf{v} \rangle^2 \zeta_l^2 \leq 2c_\star \|\mathbf{a}_l\|_2^2 \left(\sqrt{\alpha_l} + \frac{2\beta_l}{\alpha_l} \right)^2 \leq 2c_\star \|\mathbf{a}_l\|_2^2 \cdot (1 + 2c_\star)^2 \alpha_l \leq 2c_\star (1 + 2c_\star^2)^2 \|\zeta\|_4^4, \quad (3.11)$$

$$\max_{i \neq l} \zeta_i^2 \leq \frac{4\beta^2}{\alpha^2} \leq \frac{4\mu^2 \|\zeta\|_3^6}{\|\zeta\|_4^8} \leq \frac{4c_\star^2 \|\zeta\|_4^{12}}{\|\zeta\|_4^8} \leq 4c_\star^2 \|\zeta\|_4^4. \quad (3.12)$$

Hence, the inequality $\mathbf{v}^T \text{Hess } \varphi(\bar{\mathbf{q}}) \mathbf{v} \geq (1 - 6c_\star - 36c_\star^2 - 24c_\star^3) \|\zeta\|_4^4$ holds for any choice of $\mathbf{v} \perp \bar{\mathbf{q}}$, thus when $C_\star \geq 10$ this implies positive curvature along any tangent direction at such stationary point $\bar{\mathbf{q}}$. \blacksquare

3.2 Saddle Points

Lemma 3.2 Suppose $\bar{\mathbf{q}}$ is a stationary point in \mathcal{R}_{C_\star} , and its corresponding ζ has only at least two entries ζ_l and $\zeta_{l'}$ with nontrivial magnitude, then the Riemannian Hessian at $\bar{\mathbf{q}}$ has negative eigenvalues.

Proof Suppose ζ has at least two big entries ζ_l and $\zeta_{l'}$ that

$$\zeta_l^2 \geq \left(\sqrt{\alpha_l} - \frac{2\beta_l}{\alpha_l} \right)^2 \quad (3.13)$$

$$\geq \frac{\|\zeta\|_4^4}{\|\mathbf{a}_l\|_2^2} - \frac{4\mu \|\zeta\|_3^3}{\|\zeta\|_4^2 \|\mathbf{a}_l\|_2} + \frac{4\mu^2 \|\zeta\|_3^6}{\|\zeta\|_4^8} \quad (3.14)$$

$$> \frac{\|\zeta\|_4^4}{\|\mathbf{a}_l\|_2^2} - \frac{4\mu \|\zeta\|_3^3}{\|\zeta\|_4^2 \|\mathbf{a}_l\|_2}. \quad (3.15)$$

Since the nontrivial entry $\zeta_l = \langle \mathbf{a}_l, \mathbf{q} \rangle$, and again let $c_\star = 1/C_\star$, it is easy to show that the norm of \mathbf{a}_l is sufficiently large:

$$\|\mathbf{a}_l\|_2^2 \geq \zeta_l^2 \geq \left(\sqrt{\alpha_l} - \frac{2\beta_l}{\alpha_l} \right)^2 \geq (1 - 2c_\star)^2 \frac{\|\zeta\|_4^4}{\|\mathbf{a}_l\|_2^2} \geq (1 - c_\star)^2 C_\star^{2/3} \frac{\mu^{2/3} \|\zeta\|_3^2}{\|\mathbf{a}_l\|_2^2}, \quad (3.16)$$

or

$$\|\mathbf{a}_l\|_2 \geq (1 - c_\star)^{1/2} C_\star^{1/6} \mu^{1/6} \|\zeta\|_3^{1/2}. \quad (3.17)$$

Similar result holds for $\|\mathbf{a}_{l'}\|_2$, therefore

$$\frac{\mu}{\|\mathbf{a}_l\|_2 \|\mathbf{a}_{l'}\|_2} \leq \frac{\mu^{2/3}}{C_\star^{1/3} \|\zeta\|_3} \leq \frac{C_\star^{-2/3} \|\zeta\|_4^4}{C_\star^{1/3} \|\zeta\|_3^3} \leq c_\star. \quad (3.18)$$

Now we are ready to show there exists a unit vector \mathbf{v} such that $\mathbf{v} \in \text{span}(\mathbf{a}_l, \mathbf{a}_{l'})$ and $\mathbf{v} \perp \mathbf{q}$, and the Hessian has negative curvature along such \mathbf{v} :

$$\mathbf{v}^T \text{Hess } \varphi(\mathbf{q}) \mathbf{v} = -3\mathbf{v}^T \mathbf{A} \text{diag}(\zeta^2) \mathbf{A}^T \mathbf{v} + \|\zeta\|_4^4 \quad (3.19)$$

$$\leq -3\mathbf{v}^T (\mathbf{a}_l \zeta_l^2 \mathbf{a}_l^T + \mathbf{a}_{l'} \zeta_{l'}^2 \mathbf{a}_{l'}^T) \mathbf{v} + \|\zeta\|_4^4 \quad (3.20)$$

$$< -3 \left(\left| \left\langle \frac{\mathbf{a}_l}{\|\mathbf{a}_l\|_2}, \mathbf{v} \right\rangle \right|^2 + \left| \left\langle \frac{\mathbf{a}_{l'}}{\|\mathbf{a}_{l'}\|_2}, \mathbf{v} \right\rangle \right|^2 \right) \|\zeta\|_4^4 \quad (3.21)$$

$$+ \frac{4\mu \|\zeta\|_3^3}{\|\zeta\|_4^2} (\|\mathbf{a}_l\|_2 + \|\mathbf{a}_{l'}\|_2) + \|\zeta\|_4^4 \quad (3.22)$$

$$< -3 \left(1 - \frac{\mu}{\|\mathbf{a}_l\|_2 \|\mathbf{a}_{l'}\|_2} \right) \|\zeta\|_4^4 + \frac{4\mu \|\zeta\|_3^3}{\|\zeta\|_4^2} (\|\mathbf{a}_l\|_2 + \|\mathbf{a}_{l'}\|_2) + \|\zeta\|_4^4 \quad (3.23)$$

$$\leq (-2 + 11c_\star) \|\zeta\|_4^4. \quad (3.24)$$

The last inequality is implied by Lemma 2.3 and is negative when $C_\star \geq 10$. ■

4 Initialization

Lemma 4.1 Suppose $\mathbf{x}_0 \sim_{\text{i.i.d.}} \text{BG}(\theta) \in \mathbb{R}^m$. There exists a positive constant $C > 2560^4$ such that whenever

$$m \geq C \theta^2 \sigma_{\min}^{-2} \kappa^6 k^3 (1 + 36\mu^2 k \log k)^4 \log(\kappa k / \sigma_{\min}) \quad (4.1)$$

and the sparsity rate

$$64k^{-1} \log k \leq \theta \leq \min \left\{ \frac{1}{48^2} \mu^{-2} k^{-1} \log^{-2} k, \left(\frac{1}{4} - \frac{640}{C^{1/4}} \right) (3C_\star \mu \kappa^2)^{-2/3} k^{-1} (1 + 36\mu^2 k \log k)^{-2} \right\}, \quad (4.2)$$

Then the initialization $\mathbf{q}_{\text{init}} = \mathcal{P}_{\mathbb{S}} \left[(\mathbf{Y} \mathbf{Y}^T)^{-1/2} \mathbf{y}_i \right]$ satisfies

$$\|\mathbf{A}^T \mathbf{q}_{\text{init}}\|_4^6 \geq 3C_\star \mu \kappa^2, \quad (4.3)$$

namely $\mathbf{q}_{\text{init}} \in \hat{\mathcal{R}}_{3C_\star}$, with probability no smaller than $1 - k^{-1} - 8k^{-2} - 2 \exp(-\theta k) - 48k^{-7} - 48m^{-5} - 24k \exp\left(-\frac{1}{144} \min\left\{k, 3\sqrt{\theta m}\right\}\right)$.

Proof

Since

$$m \geq C \frac{\theta^2}{\sigma_{\min}^2} \kappa^6 k^3 (1 + 36\mu^2 k \log k)^4 \log(\kappa k / \sigma_{\min}) \quad (4.4)$$

with $C \geq 2560^4$, then from Lemma 5.1, then with probability no smaller than $1 - 2 \exp(-\theta k) - 24k \exp(-\frac{1}{144} \min\{k, 3\sqrt{\theta m}\}) - 48k^{-7} - 48m^{-5}$, we have

$$\delta \doteq \left\| \frac{1}{\theta m} \mathbf{X}_0 \mathbf{X}_0^T - \mathbf{I} \right\|_2 \quad (4.5)$$

$$\leq 10\sqrt{k \log m / m} \quad (4.6)$$

$$\begin{aligned} &\leq \frac{10\sigma_{\min}}{\theta \sigma_{\min}^{-1} \kappa^3 k (1 + 36\mu^2 k \log k)^2} \times \\ &\quad \sqrt{\frac{\log\left(\frac{C \kappa^6 k^3 (1 + 36\mu^2 k \log k)^4}{\sigma_{\min}^2} \log\left(\frac{\kappa k}{\sigma_{\min}}\right)\right)}{C \log(\kappa k / \sigma_{\min})}} \end{aligned} \quad (4.7)$$

$$\leq \frac{20\sigma_{\min}}{C^{1/4} \theta \kappa^3 k (1 + 36\mu^2 k \log k)^2} \quad (4.8)$$

obtains, and the last inequality holds when $C \geq 1000$ that

$$\log(37^4 C) \leq \log 2\sqrt{C}. \quad (4.9)$$

Therefore

$$\begin{aligned} &C \sigma_{\min}^{-2} \kappa^6 k^3 (1 + 36\mu^2 k \log k)^4 \log(\kappa k / \sigma_{\min}) \\ &\leq 37^4 C (\kappa k / \sigma_{\min})^7 \log^5(\kappa k / \sigma_{\min}) \end{aligned} \quad (4.10)$$

$$\leq 37^4 C (\kappa k / \sigma_{\min})^{12} \quad (4.11)$$

or

$$\begin{aligned} &\sqrt{\frac{\log\left(\frac{C \kappa^6 k^3 (1 + 36\mu^2 k \log k)^4}{\sigma_{\min}^2} \log\left(\frac{\kappa k}{\sigma_{\min}}\right)\right)}{C \log(\kappa k / \sigma_{\min})}} \\ &\leq \sqrt{\frac{\log(37^4 C) + 12 \log(\kappa k / \sigma_{\min})}{C \log(\kappa k / \sigma_{\min})}} \end{aligned} \quad (4.12)$$

$$\leq \sqrt{\frac{\log 2}{\sqrt{C} \log(\kappa k / \sigma_{\min})}} + \frac{12}{C} \quad (4.13)$$

$$\leq \frac{2}{C^{1/4}} \quad (k \geq 2, C \geq 16) \quad (4.14)$$

Moreover, $\kappa^2 \delta \leq 1/2$ always holds provided

$$C \geq \left(\frac{40}{\theta k (1 + 36\mu^2 k \log k)^2} \right)^4. \quad (4.15)$$

Notice that because θ is lower bounded by $c \log k / k$, the right hand side is indeed bounded by an absolute constant.

Set $\zeta_{\text{init}} = \mathbf{A}^T \mathbf{q}_{\text{init}}$ and $\hat{\zeta}_{\text{init}} = \mathcal{P}_{\mathbb{S}} [\mathbf{A}^T \mathbf{A} \mathbf{x}_i]$. Then using for any nonzero vectors \mathbf{u} and \mathbf{v} ,

$$\left\| \frac{\mathbf{u}}{\|\mathbf{u}\|_2} - \frac{\mathbf{v}}{\|\mathbf{v}\|_2} \right\|_2 \leq \frac{2}{\|\mathbf{v}\|_2} \|\mathbf{u} - \mathbf{v}\|_2, \quad (4.16)$$

we have that

$$\begin{aligned} & \left\| \zeta_{\text{init}} - \hat{\zeta}_{\text{init}} \right\|_2 \\ &= \left\| \mathbf{A}^T \mathcal{P}_{\mathbb{S}} \left[(\mathbf{Y} \mathbf{Y}^T)^{-1/2} \mathbf{A}_0 \mathbf{x}_i \right] - \mathcal{P}_{\mathbb{S}} [\mathbf{A}^T \mathbf{A} \mathbf{x}_i] \right\|_2 \end{aligned} \quad (4.17)$$

$$= \left\| \frac{\mathbf{A}^T \left(\frac{1}{\theta m} \mathbf{Y} \mathbf{Y}^T \right)^{-1/2} \mathbf{A}_0 \mathbf{x}_i}{\left\| \left(\frac{1}{\theta m} \mathbf{Y} \mathbf{Y}^T \right)^{-1/2} \mathbf{A}_0 \mathbf{x}_i \right\|_2} - \frac{\mathbf{A}^T \mathbf{A} \mathbf{x}_i}{\|\mathbf{A}^T \mathbf{A} \mathbf{x}_i\|_2} \right\|_2 \quad (4.18)$$

$$\leq \frac{2}{\|\mathbf{A} \mathbf{x}_i\|_2} \left\| \left(\frac{1}{\theta m} \mathbf{Y} \mathbf{Y}^T \right)^{-1/2} \mathbf{A}_0 \mathbf{x}_i - \mathbf{A} \mathbf{x}_i \right\|_2 \quad (4.19)$$

$$\leq 2 \|\mathbf{A}_0\|_2 \left\| \left(\frac{1}{\theta m} \mathbf{Y} \mathbf{Y}^T \right)^{-1/2} - (\mathbf{A}_0 \mathbf{A}_0^T)^{-1/2} \right\|_2 \quad (4.20)$$

$$\leq \frac{8\kappa^3 \delta}{\sigma_{\min}}, \quad (4.21)$$

where we have used Lemma 5.3 in the final bound.

Since $\|\cdot\|_4^4$ is convex, $\|\zeta_{\text{init}}\|_4^4$ can be lower bounded via

$$\|\zeta_{\text{init}}\|_4^4 \geq \|\hat{\zeta}_{\text{init}}\|_4^4 + 4 \left\langle \hat{\zeta}_{\text{init}}^{\circ 3}, \zeta_{\text{init}} - \hat{\zeta}_{\text{init}} \right\rangle \quad (4.22)$$

$$\geq \|\hat{\zeta}_{\text{init}}\|_4^4 - 4 \|\zeta_{\text{init}} - \hat{\zeta}_{\text{init}}\|_2 \quad (4.23)$$

$$\geq \|\hat{\zeta}_{\text{init}}\|_4^4 - \frac{32\kappa^3 \delta}{\sigma_{\min}}. \quad (4.24)$$

Let $I = \text{supp}(\mathbf{x}_i)$, then the vector $\hat{\zeta}_{\text{init}} = \mathcal{P}_{\mathbb{S}} [\mathbf{A}^T \mathbf{A} \mathbf{x}_i]$ is composed of $|I|$ large components and small components on the off-support I^c of \mathbf{x}_i .

Dense Component of $\hat{\zeta}_{\text{init}}$. Note that $\left\| (\mathbf{A}^T \mathbf{A})_{I^c, I} \mathbf{x}_i \right\|_2 \leq \|\text{offdiag}(\mathbf{A}^T \mathbf{A}) \mathbf{x}_i\|_2$ with $\|\text{offdiag}(\mathbf{A}^T \mathbf{A})\|_{\infty} \leq \mu$. We have

$$\mathbb{E} [\text{offdiag}(\mathbf{A}^T \mathbf{A}) \mathbf{x}_i] = \mathbf{0} \quad (4.25)$$

$$\begin{aligned} \mathbb{E} \left[|e_j^T \text{offdiag}(\mathbf{A}^T \mathbf{A}) \mathbf{x}_i|^2 \right] &= \theta \|e_j^T \text{offdiag}(\mathbf{A}^T \mathbf{A})\|_2^2 \\ &\leq \mu^2 \theta k \end{aligned} \quad (4.26)$$

With Bernstein's Inequality, the summation of moment-bounded independent random variables can be controlled via

$$\mathbb{P} \left[|e_j^T \text{offdiag}(\mathbf{A}^T \mathbf{A}) \mathbf{x}_i| \geq \mu t \right] \leq 2 \exp \left(-\frac{t^2}{2\theta k + 2t} \right) \quad (4.27)$$

and via union bound

$$\mathbb{P} \left[\|\text{offdiag}(\mathbf{A}^T \mathbf{A}) \mathbf{x}_i\|_2^2 \geq 2k(\mu t)^2 \right] \leq 4k \exp \left(-\frac{t^2}{2\theta k + 2t} \right) \quad (4.28)$$

Therefore, setting $t^2 = 9\theta k \log k$, we obtain

$$\|\text{offdiag}(\mathbf{A}^T \mathbf{A}) \mathbf{x}_i\|_2^2 \leq 18\mu^2 \theta k^2 \log k \quad (4.29)$$

with failure probability bounded by

$$\begin{aligned} & 4k \exp\left(-\frac{9\theta k \log k}{2\theta k + 2\sqrt{9\theta k \log k}}\right) \\ &= 4k \exp\left(-\frac{9 \log k}{2 + 6\sqrt{(\theta k)^{-1} \log k}}\right) \end{aligned} \quad (4.30)$$

$$\leq 4k^{-2} \quad (4.31)$$

The last inequality is derived under the assumption $(\theta k)^{-1} \log k \leq \frac{1}{64}$.

Spiky Component of $\hat{\zeta}_{\text{init}}$. On the other hand,

$$\mathbb{E} \left[\|\text{diag}(\mathbf{A}^T \mathbf{A}) \mathbf{x}_i\|_2^2 \right] = \theta \|\text{diag}(\mathbf{A}^T \mathbf{A})\|_F^2 \quad (4.32)$$

$$= \theta k. \quad (4.33)$$

For $\text{diag}(\mathbf{A}^T \mathbf{A}) \mathbf{x}_i$, applying the moment control Bernstein Inequality, we have

$$\mathbb{P} \left[\left| \|\text{diag}(\mathbf{A}^T \mathbf{A}) \mathbf{x}_i\|_2^2 - \mathbb{E}[\cdot] \right| \geq t \right] \leq 2 \exp \left(-\frac{t^2}{2\theta k + 2t} \right). \quad (4.34)$$

By setting $t = 2\sqrt{\theta k \log k}$, we obtain that with probability no smaller than $1 - k^{-1}$,

$$\|\text{diag}(\mathbf{A}^T \mathbf{A}) \mathbf{x}_i\|_2^2 \geq \theta k - 2\sqrt{\theta k \log k}. \quad (4.35)$$

Denote the following events for the entry-wise magnitude

$$\mathcal{E}_j = \{|e_j^T \text{offdiag}(\mathbf{A}^T \mathbf{A}) \mathbf{x}_i| \leq \mu t\}, \quad (4.36)$$

and for the support size

$$\mathcal{E}_{\text{supp}} = \{\|\mathbf{x}_i\|_0 \leq 4\theta k\}. \quad (4.37)$$

On their intersection $\mathcal{E}_{\text{supp}} \cap \bigcap_{j=1}^{2k} \mathcal{E}_j$, we have

$$\|\text{offdiag}(\mathbf{A}^T \mathbf{A})_{I,I} \mathbf{x}_i\|_2^2 \leq 4\theta k(\mu t)^2. \quad (4.38)$$

The the failure probability can be bounded from the union bound as

$$\begin{aligned} & \mathbb{P} \left[\|\text{offdiag}(\mathbf{A}^T \mathbf{A})_{I,I} \mathbf{x}_i\|_2^2 \geq 4\theta k(\mu t)^2 \right] \\ & \leq \mathbb{P} \left[\left(\mathcal{E}_{\text{supp}} \cap \bigcap_j \mathcal{E}_j \right)^c \right] \end{aligned} \quad (4.39)$$

$$= \mathbb{P} \left[\mathcal{E}_{\text{supp}}^c \cup \bigcup_j \mathcal{E}_j^c \right] \quad (4.40)$$

$$\leq \mathbb{P}[\mathcal{E}_{\text{supp}}^c] + \sum_j \mathbb{P}[\mathcal{E}_j^c] \quad (4.41)$$

$$\leq \exp(-\theta k) + 4k \exp\left(-\frac{t^2}{2\theta k + 2t}\right). \quad (4.42)$$

Therefore, by setting $t^2 = 9\theta k \log k$, we obtain

$$\left\| \text{offdiag}(\mathbf{A}^T \mathbf{A})_{I,I} \mathbf{x}_i \right\|_2^2 \leq 36\mu^2 \theta^2 k^2 \log k \quad (4.43)$$

with probability no smaller than $1 - \exp(-\theta k) - 8k^{-2}$.

Therefore, with probability no smaller than $1 - k^{-1} - 8k^{-2} - \exp(-\theta k)$,

$$\left\| \text{diag}(\mathbf{A}^T \mathbf{A}) \mathbf{x}_i \right\|_2^2 \geq \theta k - 2\sqrt{\theta k \log k} \quad (4.44)$$

$$\left\| \text{offdiag}(\mathbf{A}^T \mathbf{A})_{I,I} \mathbf{x}_i \right\|_2^2 \leq 36\mu^2 \theta^2 k^2 \log k \quad (4.45)$$

and via Cauchy-Schwartz inequality, we obtain

$$\left\| (\mathbf{A}^T \mathbf{A})_{I,I} \mathbf{x}_i \right\|_2^2 \quad (4.46)$$

$$= \left\| \text{diag}(\mathbf{A}^T \mathbf{A}) \mathbf{x}_i + \text{offdiag}(\mathbf{A}^T \mathbf{A})_{I,I} \mathbf{x}_i \right\|_2^2 \quad (4.47)$$

$$= \left\| \text{diag}(\mathbf{A}^T \mathbf{A}) \mathbf{x}_i \right\|_2^2 + \left\| \text{offdiag}(\mathbf{A}^T \mathbf{A})_{I,I} \mathbf{x}_i \right\|_2^2 + 2 \left\langle \text{diag}(\mathbf{A}^T \mathbf{A}) \mathbf{x}_i, \text{offdiag}(\mathbf{A}^T \mathbf{A})_{I,I} \mathbf{x}_i \right\rangle \quad (4.48)$$

$$\geq \left\| \text{diag}(\mathbf{A}^T \mathbf{A}) \mathbf{x}_i \right\|_2^2 - 2 \left\| \text{diag}(\mathbf{A}^T \mathbf{A}) \mathbf{x}_i \right\|_2 \left\| \text{offdiag}(\mathbf{A}^T \mathbf{A})_{I,I} \mathbf{x}_i \right\|_2 \quad (4.49)$$

$$\geq \theta k \left(1 - 2\sqrt{(\theta k)^{-1} \log k} - 12\mu\sqrt{\theta k \log k} \right) \quad (4.50)$$

$$\geq \theta k / 2. \quad (4.51)$$

The last equation is derived by plugging in

$$(\theta k)^{-1} \log k \leq \frac{1}{64}, \quad \mu^2 \theta k \log k \leq \frac{1}{48^2} \quad (4.52)$$

under the assumption

$$64k^{-1} \log k \leq \theta \leq \frac{1}{48^2} \mu^{-2} k^{-1} \log^{-1} k. \quad (4.53)$$

Lower Bound of $\|\cdot\|_4^4$. Since with probability no smaller than $1 - 4k^{-2}$, $\left\| \text{offdiag}(\mathbf{A}^T \mathbf{A}) \mathbf{x}_i \right\|_2^2 \leq 36\mu^2 \theta k^2 \log k$ obtains and the relative $\|\cdot\|_2^2$ norm between the flat entries to the spiky entries in $\mathbf{A}^T \mathbf{A} \mathbf{x}_i$ can be bounded as

$$\frac{\left\| (\mathbf{A}^T \mathbf{A})_{I^c,I} \mathbf{x}_i \right\|_2^2}{\left\| (\mathbf{A}^T \mathbf{A})_{I,I} \mathbf{x}_i \right\|_2^2} \leq \frac{\left\| \text{offdiag}(\mathbf{A}^T \mathbf{A}) \mathbf{x}_i \right\|_2^2}{\left\| (\mathbf{A}^T \mathbf{A})_{I,I} \mathbf{x}_i \right\|_2^2} \quad (4.54)$$

$$\leq 36\mu^2 k \log k \doteq r. \quad (4.55)$$

Since

$$\left\| \hat{\boldsymbol{\zeta}}_{\text{init}} \right\|_4^4 = \left\| \mathcal{P}_{\mathbb{S}} [\mathbf{A}^T \mathbf{A} \mathbf{x}_i] \right\|_4^4 \quad (4.56)$$

$$\begin{aligned}
&= \frac{1}{\|\mathbf{A}^T \mathbf{A} \mathbf{x}_i\|_2^4} \left\| (\mathbf{A}^T \mathbf{A})_{I^c, I} \mathbf{x}_i \right\|_4^4 \\
&\quad + \frac{1}{\|\mathbf{A}^T \mathbf{A} \mathbf{x}_i\|_2^4} \left\| (\mathbf{A}^T \mathbf{A})_{I, I} \mathbf{x}_i \right\|_4^4
\end{aligned} \tag{4.57}$$

$$\geq \frac{1}{\|\mathbf{A}^T \mathbf{A} \mathbf{x}_i\|_2^4} \left\| (\mathbf{A}^T \mathbf{A})_{I, I} \mathbf{x}_i \right\|_4^4 \tag{4.58}$$

$$\begin{aligned}
&= \frac{\left\| (\mathbf{A}^T \mathbf{A})_{I, I} \mathbf{x}_i \right\|_2^4 \left\| \mathcal{P}_{\mathbb{S}} \left[(\mathbf{A}^T \mathbf{A})_{I, I} \mathbf{x}_i \right] \right\|_4^4}{\left\| (\mathbf{A}^T \mathbf{A})_{I, I} \mathbf{x}_i + (\mathbf{A}^T \mathbf{A})_{I^c, I} \mathbf{x}_i \right\|_2^4} \\
&\geq \frac{1}{(1+r)^2} \left\| \mathcal{P}_{\mathbb{S}} \left[(\mathbf{A}^T \mathbf{A})_{I, I} \mathbf{x}_i \right] \right\|_4^4
\end{aligned} \tag{4.59}$$

$$\geq \frac{1}{(1+r)^2} \left\| \mathcal{P}_{\mathbb{S}} \left[(\mathbf{A}^T \mathbf{A})_{I, I} \mathbf{x}_i \right] \right\|_4^4 \tag{4.60}$$

and with high probability $1 - \exp(-\theta k)$ according to Lemma 2.4, $\mathcal{P}_{\mathbb{S}} \left[(\mathbf{A}^T \mathbf{A})_{I, I} \mathbf{x}_i \right]$ satisfies

$$\left\| \mathcal{P}_{\mathbb{S}} \left[(\mathbf{A}^T \mathbf{A})_{I, I} \mathbf{x}_i \right] \right\|_4^4 \geq \frac{1}{\|\mathbf{x}_i\|_0} \geq \frac{1}{2\theta(2k-1)}, \tag{4.61}$$

Together, we have

$$\|\zeta_{\text{init}}\|_4^4 \geq \left\| \hat{\zeta}_{\text{init}} \right\|_4^4 - \frac{32\kappa^3\delta}{\sigma_{\min}} \tag{4.62}$$

$$\begin{aligned}
&\geq \frac{1}{(1+r)^2} \left\| \mathcal{P}_{\mathbb{S}} \left[(\mathbf{A}^T \mathbf{A})_{I, I} \mathbf{x}_i \right] \right\|_4^4 \\
&\quad - \frac{640C^{-1/4}}{\theta k (1 + 36\mu^2 k \log k)^2}
\end{aligned} \tag{4.63}$$

$$\geq \left(\frac{1}{4} - \frac{640}{C^{1/4}} \right) \frac{1}{\theta k (1 + 36\mu^2 k \log k)^2} \tag{4.64}$$

holds with probability no smaller than $1 - k^{-1} - 8k^{-2} - 2\exp(-\theta k) - 24k \exp\left(-\frac{1}{144} \min\left\{k, 3\sqrt{\theta m}\right\}\right) - 48k^{-7} - 48m^{-5}$. To make sure $\|\zeta_{\text{init}}\|_4^6 \geq 3C_*\mu\kappa^2$ as desired, we require the sparsity to satisfy

$$\theta \leq \left(\frac{1}{4} - \frac{640}{C^{1/4}} \right) (3C_*\mu\kappa^2)^{-2/3} k^{-1} (1 + 36\mu^2 k \log k)^{-2}, \tag{4.65}$$

then the initialization $\mathbf{q}_{\text{init}} \in \hat{\mathcal{R}}_{3C_*}$ follows by ??.

5 Preconditioning

Lemma 5.1 Suppose $\mathbf{x}_0 \sim_{\text{i.i.d.}} \text{BG}(\theta) \in \mathbb{R}^m$, then following inequality holds

$$\left\| \frac{1}{\theta m} \mathbf{X}_0 \mathbf{X}_0^T - \mathbf{I} \right\|_2 \leq 10\sqrt{k \log m/m}, \tag{5.1}$$

with probability no smaller than $1 - 2\exp(-\theta k) - 24k \exp\left(-\frac{1}{144} \min\left\{k, 3\sqrt{\theta m}\right\}\right) - 48k^{-7} - 48m^{-5}$.

Proof Since

$$\left\| \frac{1}{\theta m} \mathbf{X}_0 \mathbf{X}_0^T - \mathbf{I} \right\|_2$$

$$\begin{aligned} &\leq \left\| \text{diag} \left(\frac{1}{\theta m} \mathbf{X}_0 \mathbf{X}_0^T \right) - \mathbf{I} \right\|_2 \\ &\quad + \left\| \text{offdiag} \left(\frac{1}{\theta m} \mathbf{X}_0 \mathbf{X}_0^T \right) \right\|_2, \end{aligned} \quad (5.2)$$

which is bounded by δ with probability no smaller than $1 - \varepsilon_d - \varepsilon_o$ whenever the probability that each of the terms is upper bounded by $\delta/2$ satisfies

$$\mathbb{P} \left[\left\| \text{diag} \left(\frac{1}{\theta m} \mathbf{X}_0 \mathbf{X}_0^T \right) - \mathbf{I} \right\|_2 \geq \delta/2 \right] \leq \varepsilon_d, \quad (5.3)$$

$$\mathbb{P} \left[\left\| \text{offdiag} \left(\frac{1}{\theta m} \mathbf{X}_0 \mathbf{X}_0^T \right) \right\|_2 \geq \delta/2 \right] \leq \varepsilon_o. \quad (5.4)$$

Diagonal of $\frac{1}{\theta m} \mathbf{X}_0 \mathbf{X}_0^T$. Note that $\text{diag}(\mathbf{X}_0 \mathbf{X}_0^T) = \|\mathbf{x}_0\|_2^2 \mathbf{I}$, so

$$\left\| \text{diag} \left(\frac{1}{\theta m} \mathbf{X}_0 \mathbf{X}_0^T \right) - \mathbf{I} \right\|_2 = \left| \frac{1}{\theta m} \|\mathbf{x}_0\|_2^2 - 1 \right|. \quad (5.5)$$

We calculate the moment for each summand of $\|\mathbf{x}_0\|_2^2$. The summands can be seen as a χ_1^2 random variable but populated with probability θ , whence

$$\mathbb{E}_{x_i \sim \text{BG}(\theta)} \left[(x_i^2)^p \right] = \theta \mathbb{E}_{X_i \sim \chi_1^2} [X_i^p] \quad (5.6)$$

$$= \theta \frac{\Gamma(p + \frac{1}{2})}{\Gamma(\frac{1}{2})} \quad (5.7)$$

$$\leq \frac{\theta p! (2)^p}{2} \quad (5.8)$$

$$= \frac{p!}{2} \sigma^2 R^{p-2}. \quad (5.9)$$

Apply Bernstein's inequality for moment bounded random variables (9.4) with $R = 2$, $\sigma^2 = 4\theta$, then

$$\mathbb{P} \left[\left| \frac{1}{m} \|\mathbf{x}_0\|_2^2 - \theta \right| \geq t \right] \leq 2 \exp \left(-\frac{mt^2}{8\theta + 4t} \right). \quad (5.10)$$

By taking $t = \frac{1}{2}\theta\delta$, we obtain

$$\begin{aligned} &\mathbb{P} \left[\left\| \text{diag} \left(\frac{1}{\theta m} \mathbf{X}_0 \mathbf{X}_0^T \right) - \mathbf{I} \right\|_2 \geq \delta/2 \right] \\ &\leq 2 \exp \left(-\frac{\theta m \delta^2}{32 + 8\delta} \right) \end{aligned} \quad (5.11)$$

$$\leq 2 \exp \left(-\frac{100\theta k \log m}{32 + 80\sqrt{k \log m/m}} \right) \quad (5.12)$$

$$\leq 2 \exp(-\theta k). \quad (5.13)$$

Off-diagonal of $\frac{1}{\theta m} \mathbf{X}_0 \mathbf{X}_0^T$. Note that $\text{offdiag}(\mathbf{X}_0 \mathbf{X}_0^T)$ is a sub-circulant matrix generated by

$$\mathbf{r}_{\mathbf{x}_0} = [r_{\mathbf{x}_0}(2k-2), \dots, 0, \dots, r_{\mathbf{x}_0}(2k-2)]^T \quad (5.14)$$

with $r_{\mathbf{x}_0}(\tau) = \langle \mathbf{x}_0, s_\tau[\mathbf{x}_0] \rangle$ for $\tau = 1, \dots, 2k-2$. Equivalently, we can write

$$\mathbf{r}_{\mathbf{x}_0} = \mathbf{R}_{\mathbf{x}_0}^T \mathbf{x}_0, \quad (5.15)$$

with

$$\mathbf{R}_{\mathbf{x}_0} = [s_{2k-2}[\mathbf{x}_0], \dots, \mathbf{0}, \dots, s_{2k-2}[\mathbf{x}_0]] \in \mathbb{R}^{m \times (4k-3)}. \quad (5.16)$$

Operator norm of a circulant matrix is defined as the following

$$\left\| \text{offdiag} \left(\frac{1}{\theta m} \mathbf{X}_0 \mathbf{X}_0^T \right) \right\|_2 = \max_{l=0, \dots, 4k-4} \left| \left\langle \mathbf{v}_l, \frac{1}{\theta m} \mathbf{r}_{\mathbf{x}_0} \right\rangle \right|, \quad (5.17)$$

where \mathbf{v}_l is the l -th (discrete) Fourier basis vector

$$\mathbf{v}_l = \left[1, e^{l \frac{2\pi j}{4k-3}}, \dots, e^{l(4k-4) \frac{2\pi j}{4k-3}} \right]^T, \quad (5.18)$$

and j is the imaginary unit. Let $v_{l,\tau} = \mathbf{v}_l(2k-2-\tau) + \mathbf{v}_l(2k-2+\tau)$, then

$$\langle \mathbf{v}_l, \mathbf{r}_{\mathbf{x}_0} \rangle = \sum_{\tau=1}^{2k-2} v_{l,\tau} \langle \mathbf{x}_0, s_\tau[\mathbf{x}_0] \rangle \quad (5.19)$$

$$= \sum_{\tau=1}^{2k-2} v_{l,\tau} \sum_{i=0}^{m-1} \mathbf{x}_0(i) \mathbf{x}_0([i+\tau]_m). \quad (5.20)$$

By decoupling (Theorem 3.4.1 of [DIPG99]), the tail probability of the weighted autocorrelation $\langle \mathbf{v}_l, \mathbf{r}_{\mathbf{x}_0} \rangle$ can be upper bounded via

$$\begin{aligned} & \mathbb{P} [|\langle \mathbf{v}_l, \mathbf{r}_{\mathbf{x}_0} \rangle| > t] \\ &= \mathbb{P} \left[\left| \sum_{\tau=1}^{2k-2} v_{l,\tau} \langle \mathbf{x}_0, s_\tau[\mathbf{x}_0] \rangle \right| > t \right] \end{aligned} \quad (5.21)$$

$$\leq 6 \mathbb{P} \left[\left| \sum_{\tau=1}^{2k-2} v_{l,\tau} \langle \mathbf{x}_0, s_\tau[\mathbf{x}'_0] \rangle \right| > \frac{t}{6} \right], \quad (5.22)$$

where $\mathbf{x}'_0 \sim_{\text{i.i.d.}} \text{BG}(\theta)$ is an independent copy of the random vector \mathbf{x}_0 , we have Plugging in $\langle \mathbf{v}_l, \mathbf{r}_{\mathbf{x}_0} \rangle = \langle \mathbf{v}_l, \mathbf{R}_{\mathbf{x}_0}^T \mathbf{x}_0 \rangle = \langle \mathbf{R}_{\mathbf{x}_0} \mathbf{v}_l, \mathbf{x}_0 \rangle$.

$$\mathbb{P} \left[\left| \left\langle \mathbf{v}_l, \frac{1}{\theta m} \mathbf{r}_{\mathbf{x}_0} \right\rangle \right| > t \right] \leq 6 \mathbb{P} \left[\left| \frac{1}{\theta m} \langle \mathbf{R}_{\mathbf{x}'_0} \mathbf{v}_l, \mathbf{x}_0 \rangle \right| > \frac{t}{6} \right]. \quad (5.23)$$

Again with Bernstein's inequality for moment bounded random variable, we have

$$\begin{aligned} & \mathbb{P} \left[\left| \frac{1}{\theta m} \langle \mathbf{R}_{\mathbf{x}'_0} \mathbf{v}_l, \mathbf{x}_0 \rangle \right| \geq t \right] \\ & \leq 2 \exp \left(- \frac{\theta m^2 t^2}{2 \|\mathbf{R}_{\mathbf{x}'_0} \mathbf{v}_l\|_2^2 + 2 \|\mathbf{R}_{\mathbf{x}'_0} \mathbf{v}_l\|_\infty m t} \right) \end{aligned} \quad (5.24)$$

Control $\|\mathbf{R}_{\mathbf{x}'_0} \mathbf{v}_l\|_2$.

$$\|\mathbf{R}_{\mathbf{x}_0} \mathbf{v}_l\|_2^2 \leq \|\mathbf{R}_{\mathbf{x}_0}\|_2^2 \|\mathbf{v}_l\|_2^2 = k \|\mathbf{R}_{\mathbf{x}_0}\|_2^2 \quad (5.25)$$

With tail bound of the operator norm of a circulant matrix in Lemma 2.6, we have

$$\mathbb{P} [\|\mathbf{R}_{\mathbf{x}_0}\|_2 \geq t] \leq 4m \exp \left(- \frac{t^2}{2\theta m + 2t} \right) \quad (5.26)$$

Control $\|\mathbf{R}_{\mathbf{x}'_0} \mathbf{v}_l\|_\infty$. For a discrete Fourier basis \mathbf{v}_l as defined, we have

$$\|\mathbf{v}_l\|_2^2 = \|\mathbf{v}_l\|_0 = 4k-3, \quad \|\mathbf{v}_l\|_\infty = 1 \quad (5.27)$$

Note that

$$\|\mathbf{R}_{\mathbf{x}_0} \mathbf{v}_l\|_\infty = \max_{\tau=1, \dots, 2k-2} |\langle s_\tau[\mathbf{x}_0], \mathbf{v}_l \rangle| \quad (5.28)$$

and moment control Bernstein inequality implies that

$$\mathbb{P} [|\langle s_\tau[\mathbf{x}_0], \mathbf{v}_l \rangle| \geq t] \leq 2 \exp \left(-\frac{t^2}{2\theta \|\mathbf{v}_l\|_2^2 + 2\|\mathbf{v}_l\|_\infty t} \right). \quad (5.29)$$

with union bound, we obtain

$$\mathbb{P} [\|\mathbf{R}_{\mathbf{x}_0} \mathbf{v}_l\|_\infty \geq t] \leq \sum_{\tau=1}^{2k-2} \mathbb{P} [|\langle s_\tau[\mathbf{x}_0], \mathbf{v}_l \rangle| \geq t] \quad (5.30)$$

$$\leq 4k \exp \left(-\frac{t^2}{8\theta k + 2t} \right) \quad (5.31)$$

Therefore, by plugging in

$$\|\mathbf{R}_{\mathbf{x}'_0} \mathbf{v}_l\|_\infty \leq t_1 = 10\sqrt{\theta k \log k}, \quad (5.32)$$

$$\|\mathbf{R}_{\mathbf{x}'_0} \mathbf{v}_l\|_2 \leq t_2 = 5\sqrt{\theta m \log m}, \quad (5.33)$$

we obtain the following probabilities

$$\begin{aligned} \mathbb{P} [\|\mathbf{R}_{\mathbf{x}'_0} \mathbf{v}_l\|_\infty \geq t_1] &\leq 4k \exp \left(-\frac{t_1^2}{8\theta k + 2t_1} \right) \\ &\leq 4k^{-8}, \end{aligned} \quad (5.34)$$

$$\begin{aligned} \mathbb{P} [\|\mathbf{R}_{\mathbf{x}'_0} \mathbf{v}_l\|_2 \geq t_2] &\leq 4m \exp \left(-\frac{t_2^2}{2\theta m + 2t_2} \right) \\ &\leq 4m^{-6}. \end{aligned} \quad (5.35)$$

Denoting event

$$\mathbf{E} = \left\{ \|\mathbf{R}_{\mathbf{x}'_0} \mathbf{v}_l\|_\infty \leq t_1, \|\mathbf{R}_{\mathbf{x}'_0} \mathbf{v}_l\|_2 \leq t_2 \right\}, \quad (5.36)$$

and combining these bounds with (5.23), we obtain

$$\begin{aligned} &\mathbb{P} \left[\left\| \text{offdiag} \left(\frac{1}{\theta m} \mathbf{X}_0 \mathbf{X}_0^T \right) \right\|_2 \geq \delta/2 \right] \\ &\leq 6 \mathbb{P} \left[\max_l \left| \frac{1}{\theta m} \langle \mathbf{R}_{\mathbf{x}'_0} \mathbf{v}_l, \mathbf{x}_0 \rangle \right| \geq \frac{\delta}{12} \right] \end{aligned} \quad (5.37)$$

$$\leq 12k \mathbb{P} \left[\left| \frac{1}{\theta m} \langle \mathbf{R}_{\mathbf{x}'_0} \mathbf{v}_l, \mathbf{x}_0 \rangle \right| \geq \frac{\delta}{12} \right] \quad (5.38)$$

$$\begin{aligned} &\leq 12k \mathbb{P} [\|\mathbf{R}_{\mathbf{x}'_0} \mathbf{v}_l\|_\infty > t_1] + 12k \mathbb{P} [\|\mathbf{R}_{\mathbf{x}'_0} \mathbf{v}_l\|_2 > t_2] \\ &\quad + 12k \mathbb{P} \left[\left| \frac{1}{\theta m} \langle \mathbf{R}_{\mathbf{x}'_0} \mathbf{v}_l, \mathbf{x}_0 \rangle \right| \geq \frac{\delta}{12} \mid \mathbf{E} \right] \end{aligned} \quad (5.39)$$

$$\begin{aligned} &\leq 24k \exp \left(-\frac{100\theta k m \log m / 144}{50\theta m \log m + \frac{200}{12} k \sqrt{\theta m \log k \log m}} \right) \\ &\quad + 12k (4k^{-8} + 4m^{-6}) \\ &\quad \left(t_1 = 10\sqrt{\theta k \log k}, t_2 = 5\sqrt{\theta m \log m} \right) \end{aligned} \quad (5.40)$$

$$\leq 24k \exp\left(-\frac{1}{144} \min\left\{k, 3\sqrt{\theta m}\right\}\right) + 48k^{-7} + 48m^{-5} \quad (5.41)$$

At last, by combining the control for both the diagonal and off-diagonal term, we obtain that with probability no smaller than $1 - 2 \exp(-\theta k) - 24k \exp\left(-\frac{1}{144} \min\left\{k, 3\sqrt{\theta m}\right\}\right) - 48k^{-7} - 48m^{-5}$,

$$\left\|\frac{1}{\theta m} \mathbf{X}_0 \mathbf{X}_0^T - \mathbf{I}\right\|_2 \leq 10\sqrt{k \log m/m}, \quad (5.42)$$

holds and completes the proof. \blacksquare

Lemma 5.2 Suppose $\delta = \left\|\frac{1}{\theta m} \mathbf{X}_0 \mathbf{X}_0^T - \mathbf{I}\right\|_2 \leq 1/(2\kappa^2)$, then

$$\left\|\left(\frac{1}{\theta m} \mathbf{Y} \mathbf{Y}^T\right)^{1/2} (\mathbf{A}_0 \mathbf{A}_0^T)^{-1/2} - \mathbf{I}\right\|_2 \leq \kappa^2 \delta / \sigma_{\min}. \quad (5.43)$$

Proof As in by [?], we denote the directional derivative of f at direction Δ with

$$Df(\mathbf{M})(\Delta) = \left.\frac{d}{dt}\right|_{t=0} f(\mathbf{M} + t\Delta), \quad (5.44)$$

Denote symmetric matrix $\mathbf{M} = \mathbf{A}_0 \mathbf{A}_0^T = \mathbf{U} \mathbf{\Lambda} \mathbf{U}^T$, with λ_{\max} and λ_{\min} being its maximum and minimum eigenvalue. Then we have

$$\frac{1}{\theta m} \mathbf{Y} \mathbf{Y}^T = \mathbf{M} + \Delta, \quad \|\Delta\|_2 \leq \lambda_{\max} \delta. \quad (5.45)$$

Then derivative of f with $Df(\mathbf{M})$. By differential calculus, we can obtain that

$$\begin{aligned} & \left\|\left(\frac{1}{\theta m} \mathbf{Y} \mathbf{Y}^T\right)^{1/2} (\mathbf{A}_0 \mathbf{A}_0^T)^{-1/2} - \mathbf{I}\right\|_2 \\ &= \left\|(\mathbf{A}_0 \mathbf{A}_0^T + \Delta)^{1/2} (\mathbf{A}_0 \mathbf{A}_0^T)^{-1/2} - \mathbf{I}\right\|_2 \end{aligned} \quad (5.46)$$

$$= \left\|(\mathbf{A}_0 \mathbf{A}_0^T)^{-1/2} \int_{t=0}^1 Df(\mathbf{A}_0 \mathbf{A}_0^T + t\Delta)(\Delta) dt\right\|_2 \quad (5.47)$$

$$\leq \sup_{t \in [0,1]} \|Df(\mathbf{A}_0 \mathbf{A}_0^T + t\Delta)\|_2 \|\Delta\|_2 \left\|(\mathbf{A}_0 \mathbf{A}_0^T)^{-1/2}\right\|_2 \quad (5.48)$$

$$\leq \sup_{t \in [0,1]} \|Df(\mathbf{A}_0 \mathbf{A}_0^T + t\Delta)\|_2 \lambda_{\max} \delta / \sigma_{\min} \quad (5.49)$$

Moreover, we denote $f(t) = t^{1/2}$ and $g(t) = t^2$, then $f = g^{-1}$. The directional derivative of g has following form

$$Dg(\mathbf{M})(\mathbf{X}) = \mathbf{M}\mathbf{X} + \mathbf{X}\mathbf{M}, \quad (5.50)$$

and directional derivative $\mathbf{Z} = Df(\mathbf{M})(\mathbf{X})$ satisfies

$$\mathbf{M}\mathbf{Z} + \mathbf{Z}\mathbf{M} = \mathbf{X}. \quad (5.51)$$

Denote $\mathbf{M} = \mathbf{U} \mathbf{\Lambda} \mathbf{U}^T$ with \mathbf{U} orthogonal, without loss of generality,

$$\mathbf{\Lambda} \mathbf{Z} + \mathbf{Z} \mathbf{\Lambda} = \mathbf{X}. \quad (5.52)$$

Applying Theorem VII.2.3 of [?], we have

$$\|Df(\mathbf{M})(\mathbf{X})\|_2 = \sup_{\|\mathbf{X}\|_2 \leq 1} \|\mathbf{Z}\|_2 \quad (5.53)$$

$$\leq \int_{t=0}^{\infty} \|e^{-\Lambda t} \mathbf{X} e^{-\Lambda t}\|_2 dt \quad (5.54)$$

$$\leq \int_{t=0}^{\infty} e^{-2\lambda_{\min} t} \|\mathbf{X}\|_2 dt \quad (5.55)$$

and

$$\begin{aligned} & \sup_{t \in [0,1]} \|Df(\mathbf{A}_0 \mathbf{A}_0^T + t\mathbf{\Delta})\|_2 \\ & \leq \frac{\|\mathbf{X}\|_2}{2(\lambda_{\min} - \lambda_{\max}\delta)} \end{aligned} \quad (5.56)$$

$$\leq 1/\lambda_{\min}. \quad (5.57)$$

Therefore,

$$\left\| \left(\frac{1}{\theta m} \mathbf{Y} \mathbf{Y}^T \right)^{1/2} (\mathbf{A}_0 \mathbf{A}_0^T)^{-1/2} - \mathbf{I} \right\|_2 \leq \kappa^2 \delta / \sigma_{\min}. \quad (5.58)$$

■

Lemma 5.3 Suppose \mathbf{A}_0 has condition number κ and

$$\delta = \left\| \frac{1}{\theta m} \mathbf{X}_0 \mathbf{X}_0^T - \mathbf{I} \right\|_2 \leq 1/(2\kappa^2) \quad (5.59)$$

then

$$\left\| \left(\frac{1}{\theta m} \mathbf{Y} \mathbf{Y}^T \right)^{-1/2} - (\mathbf{A}_0 \mathbf{A}_0^T)^{-1/2} \right\|_2 \leq 4\kappa^2 \delta / \sigma_{\min}^2. \quad (5.60)$$

Proof Denote symmetric matrix

$$\mathbf{M} = \mathbf{A}_0 \mathbf{A}_0^T = \mathbf{U} \mathbf{\Lambda} \mathbf{U}^T, \quad (5.61)$$

with λ_{\max} and λ_{\min} being its maximum and minimum eigenvalue. Then we have

$$\frac{1}{\theta m} \mathbf{Y} \mathbf{Y}^T = \mathbf{M} + \mathbf{\Delta}, \quad \|\mathbf{\Delta}\|_2 \leq \lambda_{\max} \delta. \quad (5.62)$$

Then

$$\begin{aligned} & \left\| \left(\frac{1}{\theta m} \mathbf{Y} \mathbf{Y}^T \right)^{-1/2} - (\mathbf{A}_0 \mathbf{A}_0^T)^{-1/2} \right\|_2 \\ & = \left\| (\mathbf{M} + \mathbf{\Delta})^{-1/2} - \mathbf{M}^{-1/2} \right\|_2 \end{aligned} \quad (5.63)$$

$$\leq \|\mathbf{\Delta}\|_2 \cdot \sup_{0 \leq t \leq 1} \|Df(\mathbf{M} + t\mathbf{\Delta})\|_2. \quad (5.64)$$

Here, $f(t) = t^{-1/2}$ and Df is the derivative of function f . In addition, we define function $g(t) = t^{-2}$, $h(t) = t^{-1}$, $w(t) = t^2$, and following function compositions hold

$$f = g^{-1}, \quad g = h \circ w. \quad (5.65)$$

For differential function g and if $Dg(f(\mathbf{M})) \neq 0$, we have

$$Df(\mathbf{M}) = [Dg(f(\mathbf{M}))]^{-1}. \quad (5.66)$$

The derivative of function g satisfies the chain rule that

$$Dg(\mathbf{M}) = Dh(w(\mathbf{M}))(Dw(\mathbf{M})). \quad (5.67)$$

Plug in

$$Dh(\mathbf{M})(\mathbf{X}) = -\mathbf{M}^{-1}\mathbf{X}\mathbf{M}^{-1}, \quad (5.68)$$

$$Dw(\mathbf{M})(\mathbf{X}) = \mathbf{M}\mathbf{X} + \mathbf{X}\mathbf{M}, \quad (5.69)$$

we obtain that

$$Dg(\mathbf{M})(\mathbf{X}) \quad (5.70)$$

$$= Dh(w(\mathbf{M}))(Dw(\mathbf{M})(\mathbf{X})) \quad (5.71)$$

$$= Dh(w(\mathbf{M}))[\mathbf{M}\mathbf{X} + \mathbf{X}\mathbf{M}] \quad (5.72)$$

$$= Dh(\mathbf{M}^2)[\mathbf{M}\mathbf{X} + \mathbf{X}\mathbf{M}] \quad (5.73)$$

$$= -\mathbf{M}^{-2}[\mathbf{M}\mathbf{X} + \mathbf{X}\mathbf{M}]\mathbf{M}^{-2} \quad (5.74)$$

$$= -[\mathbf{M}^{-1}\mathbf{X}\mathbf{M}^{-2} + \mathbf{M}^{-2}\mathbf{X}\mathbf{M}^{-1}]. \quad (5.75)$$

Since the function g is differentiable and $Dg(\mathbf{M}) \neq \mathbf{0}$, then

$$Df(\mathbf{M}) = [Dg(f(\mathbf{M}))]^{-1} \quad (5.76)$$

$$= [Dg(\mathbf{M}^{-1/2})]^{-1}. \quad (5.77)$$

Hence, directional derivative $\mathbf{Z} \doteq Df(\mathbf{M})(\mathbf{X})$ satisfies

$$\mathbf{M}^{1/2}\mathbf{Z}\mathbf{M} + \mathbf{M}\mathbf{Z}\mathbf{M}^{1/2} = -\mathbf{X}. \quad (5.78)$$

Denote $\mathbf{M} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^T$ with $\mathbf{\Lambda} \succ 0$ and \mathbf{U} orthogonal, without loss of generality

$$\mathbf{\Lambda}\mathbf{Z}\mathbf{\Lambda}^{1/2} + \mathbf{\Lambda}^{1/2}\mathbf{Z}\mathbf{\Lambda} = -\mathbf{X}. \quad (5.79)$$

Above equation can be reformulated as a Sylvester equation as following

$$\mathbf{\Lambda}^{1/2}\mathbf{Z} - \mathbf{Z}(-\mathbf{\Lambda}^{1/2}) = -\mathbf{\Lambda}^{-1/2}\mathbf{X}\mathbf{\Lambda}^{-1/2}. \quad (5.80)$$

From Theorem VII.2.3 of [?], when there are no common eigenvalues of $\mathbf{\Lambda}^{1/2}$ and $-\mathbf{\Lambda}^{1/2}$, then there exists a closed form solution for matrix \mathbf{Z} that

$$\mathbf{Z} = \int_{t=0}^{\infty} e^{-\mathbf{\Lambda}^{1/2}t} (-\mathbf{\Lambda}^{-1/2}\mathbf{X}\mathbf{\Lambda}^{-1/2}) e^{-\mathbf{\Lambda}^{1/2}t} dt \quad (5.81)$$

Therefore, the operator norm of $Df(\mathbf{M})$ can be obtained as

$$\|Df(\mathbf{M})(\mathbf{X})\|_2 = \sup_{\|\mathbf{X}\|_2 \leq 1} \|\mathbf{Z}\|_2 \quad (5.82)$$

$$\leq \int_{t=0}^{\infty} \left\| e^{-\mathbf{\Lambda}^{1/2}t} (\mathbf{\Lambda}^{-1/2}\mathbf{X}\mathbf{\Lambda}^{-1/2}) e^{-\mathbf{\Lambda}^{1/2}t} \right\| dt \quad (5.83)$$

$$\leq \int_{t=0}^{\infty} e^{-\lambda_{\min}t} \left\| \mathbf{\Lambda}^{-1/2}\mathbf{X}\mathbf{\Lambda}^{-1/2} \right\| dt \quad (5.84)$$

$$\leq \frac{\|\mathbf{X}\|}{\lambda_{\min}^2}. \quad (5.85)$$

Therefore

$$\begin{aligned} & \left\| (\mathbf{M} + \mathbf{\Delta})^{-1/2} - \mathbf{M}^{-1/2} \right\|_2 \\ & \leq \frac{\|\mathbf{\Delta}\|_2}{(\lambda_{\min} - \|\mathbf{\Delta}\|_2)^2} \end{aligned} \quad (5.86)$$

$$\leq \frac{4\|\mathbf{\Delta}\|_2}{\lambda_{\min}^2} \quad (\delta \leq 1/(2\kappa^2)) \quad (5.87)$$

$$\leq \frac{4\lambda_{\max}\delta}{\lambda_{\min}^2} \quad (5.88)$$

$$= \frac{4\kappa^2\delta}{\sigma_{\min}^2} \quad (5.89)$$

■

6 Proof of the Main Theorem and Corollary

6.1 Proof of the Main Theorem

Lemma 6.1 *If following inequalities hold*

$$\begin{aligned} & \left\| \text{grad}[\psi](\mathbf{q}) - \frac{3(1-\theta)}{\theta m^2} \text{grad}[\varphi](\mathbf{q}) \right\|_2 \\ & \leq \frac{3c_\star}{2\kappa^2} \frac{1-\theta}{\theta m^2} \|\mathbf{A}^T \mathbf{q}\|_4^6, \end{aligned} \quad (6.1)$$

$$\begin{aligned} & \left\| \text{Hess}[\psi](\mathbf{q}) - \frac{3(1-\theta)}{\theta m^2} \text{Hess}[\varphi](\mathbf{q}) \right\|_2 \\ & \leq 3(1-6c_\star-36c_\star^2-24c_\star^3) \frac{1-\theta}{\theta m^2} \|\mathbf{A}^T \mathbf{q}\|_4^4. \end{aligned} \quad (6.2)$$

for all $\mathbf{q} \in \mathcal{R}_{2C_\star}$ with $C_\star \geq 10$ and $c_\star = 1/C_\star$, then any local minimum $\bar{\mathbf{q}}$ of $\psi(\mathbf{q})$ in \mathcal{R}_{2C_\star} satisfies $|\langle \bar{\mathbf{q}}, \mathcal{P}_{\mathbb{S}}[\mathbf{a}_l] \rangle| \geq 1 - 2c_\star \kappa^{-2}$ for some index l .

Proof Let

$$\delta_{\text{grad}} = \text{grad}[\psi](\mathbf{q}) - \frac{3(1-\theta)}{\theta m^2} \text{grad}[\varphi](\mathbf{q}), \quad (6.3)$$

and let

$$\bar{\delta}_{\text{grad}} = \frac{\theta m^2}{3(1-\theta)} \delta_{\text{grad}}. \quad (6.4)$$

Then at any stationary point of $\psi(\mathbf{q})$, we have

$$\mathbf{0} = \mathbf{A}^T \text{grad}[\psi](\mathbf{q}) \quad (6.5)$$

$$= \frac{3(1-\theta)}{\theta m^2} \mathbf{A}^T \text{grad}[\varphi](\mathbf{q}) + \mathbf{A}^T \delta_{\text{grad}}. \quad (6.6)$$

Hence for any index i , following equality always holds

$$0 = \|\mathbf{a}_i\|_2^2 \zeta_i^3 + \sum_{j \neq i} \langle \mathbf{a}_i, \mathbf{a}_j \rangle \zeta_j^3 - \zeta_i \|\boldsymbol{\zeta}\|_4^4 + \langle \mathbf{a}_i, \bar{\delta}_{\text{grad}} \rangle$$

$$= \zeta_i^3 - \underbrace{\zeta_i \frac{\|\zeta\|_4^4}{\|\mathbf{a}_i\|_2^2}}_{\alpha_i} + \underbrace{\frac{\sum_{j \neq i} \langle \mathbf{a}_i, \mathbf{a}_j \rangle \zeta_j^3 + \langle \mathbf{a}_i, \bar{\boldsymbol{\delta}}_{\text{grad}} \rangle}{\|\mathbf{a}_i\|_2^2}}_{\beta'_i} \quad (6.7)$$

with $\zeta = \mathbf{A}^T \mathbf{q}$. Under the assumption that

$$\left\| \text{grad}[\psi](\mathbf{q}) - \frac{3(1-\theta)}{\theta m^2} \text{grad}[\varphi](\mathbf{q}) \right\|_2 \leq \frac{3c_\star}{2\kappa^2} \frac{1-\theta}{\theta m^2} \|\zeta\|_4^6, \quad (6.8)$$

the perturbed part can be bounded via

$$|\langle \mathbf{a}_i, \bar{\boldsymbol{\delta}}_{\text{grad}} \rangle| \leq \|\mathbf{a}_i\|_2 \|\bar{\boldsymbol{\delta}}_{\text{grad}}\|_2 \leq \frac{c_\star}{2\kappa^2} \|\mathbf{a}_i\|_2 \|\zeta\|_4^6, \quad (6.9)$$

and also

$$\frac{\beta'_i}{\alpha_i^{3/2}} \leq \frac{\mu \|\zeta\|_3^3 + \frac{1}{2} c_\star \kappa^{-2} \|\zeta\|_4^6}{\|\zeta\|_4^6} \leq c_\star \kappa^{-2} \leq \frac{1}{4}. \quad (6.10)$$

Then by Lemma 2.2, at every stationary point $\bar{\mathbf{q}}$, the i -th entry of ζ resides in the set $\bigcup_{x \in \{0, \pm\sqrt{\alpha_i}\}} [x - \frac{2\beta'_i}{\alpha_i}, x + \frac{2\beta'_i}{\alpha_i}]$ – i.e., ζ is nearly a trinary vector.

Moreover, we can characterize the curvature of critical points in terms of the number of large entries of ζ . Indeed, whenever ζ has at least two entries in

$$\bigcup_{x \in \{\pm\sqrt{\alpha_i}\}} \left[x - \frac{2\beta'_i}{\alpha_i}, x + \frac{2\beta'_i}{\alpha_i} \right],$$

using (3.24), there exists a direction of strict negative curvature, provided

$$\begin{aligned} \text{Hess}[\psi](\mathbf{q}) &\prec \frac{3(1-\theta)}{\theta m^2} \text{Hess}[\varphi](\mathbf{q}) \\ &\quad + 3(2-11c_\star) \frac{1-\theta}{\theta m^2} \|\zeta\|_4^4 \mathbf{I}. \end{aligned} \quad (6.11)$$

Similarly, whenever ζ has only one entry in

$$\bigcup_{x \in \{\pm\sqrt{\alpha_i}\}} \left[x - \frac{2\beta'_i}{\alpha_i}, x + \frac{2\beta'_i}{\alpha_i} \right],$$

using (3.10), we have that $\text{Hess}[\psi](\mathbf{q}) \succ \mathbf{0}$, provided

$$\begin{aligned} \text{Hess}[\psi](\mathbf{q}) &\succ \frac{3(1-\theta)}{\theta m^2} \text{Hess}[\varphi](\mathbf{q}) \\ &\quad - 3(1-6c_\star-36c_\star^2-24c_\star^3) \frac{1-\theta}{\theta m^2} \|\zeta\|_4^4 \mathbf{I}. \end{aligned} \quad (6.12)$$

When $C_\star \geq 10$ and $c_\star \leq 0.1$, we have $2-11c_\star > 1-6c_\star-36c_\star^2-24c_\star^3 \geq 0.016$, and so above characterization obtains. \blacksquare

Theorem 6.2 (Main Result) Assume the observation $\mathbf{y} \in \mathbb{R}^m$ is the cyclic convolution of $\mathbf{a}_0 \in \mathbb{R}^k$ and $\mathbf{x}_0 \sim_{\text{i.i.d.}}$ BG(θ) $\in \mathbb{R}^m$, where the convolution matrix $\mathbf{A}_0 \in \mathbb{R}^{k \times (2k-1)}$ has minimum singular value $\sigma_{\min} > 0$ and condition number $\kappa \geq 1$, and \mathbf{A} has column incoherence μ . If

$$m \geq C \frac{\min \left\{ (2C_\star \mu)^{-1}, \kappa^2 k^2 \right\}}{(1-\theta)^2 \sigma_{\min}^2} \kappa^8 k^4 \log^3(\kappa k) \quad (6.13)$$

and $\theta \geq \log k/k$, then with probability no smaller than $1 - \exp(-k) - \theta^2(1-\theta)^2 k^{-4} - 2\exp(-\theta k) - 48k^{-7} - 48m^{-5} - 24k \exp\left(-\frac{1}{144} \min\{k, 3\sqrt{\theta m}\}\right)$, any local minimum $\bar{\mathbf{q}}$ of ψ in $\hat{\mathcal{R}}_{2C_\star}$ satisfies $|\langle \bar{\mathbf{q}}, \mathcal{P}_{\mathbb{S}}[\mathbf{a}_\tau] \rangle| \geq 1 - c_\star \kappa^{-2}$ for some integer τ .

Proof From the concentration analysis for the Riemannian gradient (Lemma 7.1) and Hessian (Lemma 8.1), if

$$m \geq C \frac{\min\{(2C_\star \mu)^{-1}, \kappa^2 k^2\}}{(1-\theta)^2 \sigma_{\min}^2} \kappa^8 k^4 \log^3(\kappa k), \quad (6.14)$$

then with probability no smaller than $1 - \exp(-k) - \theta^2(1-\theta)^2 k^{-4} - 2\exp(-\theta k) - 24k \exp\left(-\frac{1}{144} \min\{k, 3\sqrt{\theta m}\}\right) - 48k^{-7} - 48m^{-5}$,

$$\begin{aligned} & \left\| \text{grad}[\psi](\mathbf{q}) - \frac{3(1-\theta)}{\theta m^2} \text{grad}[\varphi](\mathbf{q}) \right\|_2 \\ & \leq \frac{3c_\star}{2\kappa^2} \frac{1-\theta}{\theta m^2} \|\mathbf{A}^T \mathbf{q}\|_4^6, \end{aligned} \quad (6.15)$$

$$\begin{aligned} & \left\| \text{Hess}[\psi](\mathbf{q}) - \frac{3(1-\theta)}{\theta m^2} \text{Hess}[\varphi](\mathbf{q}) \right\|_2 \\ & \leq 3(1-6c_\star - 36c_\star^2 - 24c_\star^3) \frac{1-\theta}{\theta m^2} \|\mathbf{A}^T \mathbf{q}\|_4^4. \end{aligned} \quad (6.16)$$

hold for all $\mathbf{q} \in \hat{\mathcal{R}}_{2C_\star}$ with $C_\star \geq 10$ and $c_\star = 1/C_\star$. Therefore, by Lemma 6.1 any local minimum $\bar{\mathbf{q}}$ of $\psi(\mathbf{q})$ in \mathcal{R}_{2C_\star} satisfies $|\langle \bar{\mathbf{q}}, \mathcal{P}_{\mathbb{S}}[\mathbf{a}_l] \rangle| \geq 1 - 2c_\star \kappa^{-2}$ for some index l . ■

6.2 Proof of the Main Corollary

Corollary 6.3 Suppose the ground truth kernel \mathbf{a}_0 has induces coherence $0 \leq \mu \leq \frac{1}{8 \times 48} \log^{-3/2}(k)$ and sparse coefficient $\mathbf{x}_0 \sim_{\text{i.i.d.}} \text{BG}(\theta) \in \mathbb{R}^m$. there exist positive constants $C \geq 2560^4$ and C' such that whenever the sparsity level

$$\begin{aligned} 64k^{-1} \log k \leq \theta \leq \min\left\{\frac{1}{48^2} \mu^{-2} k^{-1} \log^{-2} k, \right. \\ \left. \left(\frac{1}{4} - \frac{640}{C^{1/4}}\right) (3C_\star \mu \kappa^2)^{-2/3} k^{-1} (1 + 36\mu^2 k \log k)^{-2}\right\}, \end{aligned} \quad (6.17)$$

and signal length

$$\begin{aligned} m \geq \max\left\{C\theta^2 \sigma_{\min}^{-2} \kappa^6 k^3 (1 + 36\mu^2 k \log k)^4 \log(\kappa k), \right. \\ \left. C' (1-\theta)^{-2} \sigma_{\min}^{-2} \min\{\mu^{-1}, \kappa^2 k^2\} \kappa^8 k^4 \log^3(\kappa k)\right\}, \end{aligned} \quad (6.18)$$

then Algorithm 1 recovers $\bar{\mathbf{a}}$ such that

$$\|\bar{\mathbf{a}} \pm \mathcal{P}_{\mathbb{S}}[\boldsymbol{\nu}_k s_\tau [\bar{\mathbf{a}}_0]]\|_2 \leq 4\sqrt{c_\star} + ck^{-1} \quad (6.19)$$

for some integer shift $\tau \in [-(k-1), k-1]$ with probability no smaller than $1 - k^{-1} - 8k^{-2} - \exp(-k) - \theta^2(1-\theta)^2 k^{-4} - 2\exp(-\theta k) - 24k \exp\left(-\frac{1}{144} \min\{k, 3\sqrt{\theta m}\}\right) - 48k^{-7} - 48m^{-5}$.

Proof From the concentration results for the Riemannian gradient, at every point $\mathbf{q} \in \hat{\mathcal{R}}_{2C_\star}$, the objective value of $\psi(\mathbf{q})$ satisfies

$$\left| \psi(\mathbf{q}) - \frac{3(1-\theta)}{\theta m^2} \varphi(\mathbf{q}) + \frac{3}{4m^2} \right|$$

$$\leq \left| \frac{\left\| \mathbf{Y}^T (\mathbf{Y} \mathbf{Y}^T)^{-1/2} \mathbf{q} \right\|_4^4}{4m} - \frac{3(1-\theta) \|\zeta\|_4^4}{4\theta m^2} - \frac{3}{4m^2} \right| \quad (6.20)$$

$$\leq \left| \left\langle \mathbf{q}, \frac{(\mathbf{Y} \mathbf{Y}^T)^{-1/2} \mathbf{Y} \boldsymbol{\eta}^{\circ 3}}{4m} - \frac{3(1-\theta)}{4\theta m^2} \mathbf{A} \boldsymbol{\zeta}^{\circ 3} - \frac{3}{4m^2} \mathbf{q} \right\rangle \right| \quad (6.21)$$

$$\leq \left\| \frac{(\mathbf{Y} \mathbf{Y}^T)^{-1/2} \mathbf{Y} \boldsymbol{\eta}^{\circ 3}}{4m} - \frac{3(1-\theta)}{4\theta m^2} \mathbf{A} \boldsymbol{\zeta}^{\circ 3} - \frac{3}{4m^2} \mathbf{q} \right\|_2 \quad (6.22)$$

$$\begin{aligned} &\leq \frac{1}{4m} \left\| (\mathbf{Y} \mathbf{Y}^T)^{-1/2} \mathbf{Y} \boldsymbol{\eta}^{\circ 3} - (\theta m \mathbf{A}_0 \mathbf{A}_0^T)^{-1/2} \mathbf{Y} \boldsymbol{\eta}^{\circ 3} \right\|_2 \\ &\quad + \frac{1}{4\theta^{1/2} m^{3/2}} \left\| (\mathbf{A}_0 \mathbf{A}_0^T)^{-1/2} \mathbf{Y} (\boldsymbol{\eta}^{\circ 3} - \bar{\boldsymbol{\eta}}^{\circ 3}) \right\|_2 \\ &\quad + \left\| \frac{1}{4\theta^{1/2} m^{3/2}} (\mathbf{A}_0 \mathbf{A}_0^T)^{-1/2} \mathbf{Y} \bar{\boldsymbol{\eta}}^{\circ 3} \right. \\ &\quad \left. - \frac{3(1-\theta)}{4\theta m^2} \mathbf{A} \boldsymbol{\zeta}^{\circ 3} - \frac{3}{4m^2} \mathbf{q} \right\|_2 \end{aligned} \quad (6.23)$$

$$\leq \frac{3c_\star}{8\kappa^2} \frac{1-\theta}{\theta m^2} \min_{\mathbf{q} \in \mathcal{R}_{2C_\star}} \|\mathbf{A}^T \mathbf{q}\|_4^6 \quad (6.24)$$

with probability no smaller than $1 - 2 \exp(-\theta k) - 24k \exp\left(-\frac{1}{144} \min\{k, 3\sqrt{\theta m}\}\right) - 48k^{-7} - 48m^{-5}$. The last inequality is derived with similar arguments in Lemma 7.1, for simplicity, we do not present them here. Moreover, with Lemma 4.1, we can obtain an initialization point \mathbf{q}_{init} such that

$$\|\mathbf{A}^T \mathbf{q}_{\text{init}}\|_4^4 \geq (3C_\star \mu \kappa^2)^{2/3} \quad (6.25)$$

$$\geq (2C_\star \mu \kappa^2)^{2/3} + \mu/2. \quad (6.26)$$

Consider any descent method for ψ , which generates a sequence of iterates $\mathbf{q}^{(0)} = \mathbf{q}_{\text{init}}, \mathbf{q}^{(1)}, \dots, \mathbf{q}^{(k)}, \dots$ such that $\psi(\mathbf{q}^{(k)})$ is non-increasing with k . Then

$$\psi(\mathbf{q}^{(k)}) \leq \psi(\mathbf{q}_{\text{init}}) \quad (6.27)$$

$$\begin{aligned} &\leq \frac{3(1-\theta)}{\theta m^2} \varphi(\mathbf{q}_{\text{init}}) + \frac{3}{4m^2} \\ &\quad + \frac{3c_\star}{8\kappa^2} \frac{1-\theta}{\theta m^2} \min_{\mathbf{q} \in \mathcal{R}_{2C_\star}} \|\mathbf{A}^T \mathbf{q}\|_4^6. \end{aligned} \quad (6.28)$$

On the other hand, the finite sample objective function value ψ is close to that of $\frac{3(1-\theta)}{\theta m^2} \varphi(\mathbf{q}) - \frac{3}{4m^2}$,

$$\begin{aligned} &\frac{3(1-\theta)}{\theta m^2} \varphi(\mathbf{q}^{(k)}) \\ &\leq \psi(\mathbf{q}^{(k)}) + \frac{3}{4m^2} + \frac{3c_\star}{8\kappa^2} \frac{1-\theta}{\theta m^2} \min_{\mathbf{q} \in \mathcal{R}_{2C_\star}} \|\mathbf{A}^T \mathbf{q}\|_4^6 \end{aligned} \quad (6.29)$$

$$\leq \frac{3(1-\theta)}{\theta m^2} \varphi(\mathbf{q}_{\text{init}}) + \frac{3c_\star}{8\kappa^2} \frac{1-\theta}{\theta m^2} \min_{\mathbf{q} \in \mathcal{R}_{2C_\star}} \|\mathbf{A}^T \mathbf{q}\|_4^6, \quad (6.30)$$

Therefore, we obtain that

$$\varphi(\mathbf{q}^{(k)}) \leq \varphi(\mathbf{q}_{\text{init}}) + \frac{\mu}{2} \quad (6.31)$$

$$\leq \varphi(\mathbf{q}_{\text{init}}) + \frac{c_\star}{4\kappa^2} \min_{\mathbf{q} \in \hat{\mathcal{R}}_{2C_\star}} \|\mathbf{A}^T \mathbf{q}\|_4^6, \quad (6.32)$$

which implies that $\mathbf{q}^{(k)} \in \hat{\mathcal{R}}_{2C_\star}$ always holds. At last, Theorem 6.2 says that any local minimum $\bar{\mathbf{q}}$ is close to $\pm \mathbf{a}_i$ for some i , in the sense that

$$|\langle \bar{\mathbf{q}}, \mathcal{P}_{\mathbb{S}}[\mathbf{a}_i] \rangle| \geq 1 - c_\star \kappa^{-2}. \quad (6.33)$$

Write $\frac{1}{\theta m} \mathbf{Y} \mathbf{Y}^T = \mathbf{A}_0 (\mathbf{I} + \Delta) \mathbf{A}_0^T$ with $\|\Delta\|_2 \leq \delta$, and let

$$\bar{\mathbf{q}} = \pm \frac{\mathbf{a}_i}{\|\mathbf{a}_i\|_2} + \sqrt{2 \left(1 - \left| \left\langle \bar{\mathbf{q}}, \frac{\mathbf{a}_i}{\|\mathbf{a}_i\|_2} \right\rangle \right| \right)} \delta, \quad (6.34)$$

with $\|\delta\|_2 = 1$. Since

$$\mathbf{a}_i = (\mathbf{A}_0 \mathbf{A}_0^T)^{-1/2} \boldsymbol{\iota}_k^* s_{-(k-i)}[\widetilde{\mathbf{a}}_0], \quad (6.35)$$

we have

$$\begin{aligned} & \left(\frac{\mathbf{Y} \mathbf{Y}^T}{\theta m} \right)^{1/2} \bar{\mathbf{q}} \\ &= \pm \left(\frac{\mathbf{Y} \mathbf{Y}^T}{\theta m} \right)^{1/2} \left[\frac{\mathbf{a}_i}{\|\mathbf{a}_i\|_2} + \sqrt{2 \left(1 - \left| \left\langle \bar{\mathbf{q}}, \frac{\mathbf{a}_i}{\|\mathbf{a}_i\|_2} \right\rangle \right| \right)} \delta \right] \end{aligned} \quad (6.36)$$

$$\begin{aligned} &= \pm \left(\frac{\mathbf{Y} \mathbf{Y}^T}{\theta m} \right)^{1/2} (\mathbf{A}_0 \mathbf{A}_0^T)^{-1/2} \frac{\boldsymbol{\iota}_k^* s_{-(k-i)}[\widetilde{\mathbf{a}}_0]}{\|\mathbf{a}_i\|_2} \\ &\quad + \sqrt{2 \left(1 - \left| \left\langle \bar{\mathbf{q}}, \frac{\mathbf{a}_i}{\|\mathbf{a}_i\|_2} \right\rangle \right| \right)} \left(\frac{\mathbf{Y} \mathbf{Y}^T}{\theta m} \right)^{1/2} \delta \end{aligned} \quad (6.37)$$

therefore the error can be bounded as

$$\begin{aligned} & \left\| \left(\frac{\mathbf{Y} \mathbf{Y}^T}{\theta m} \right)^{1/2} \bar{\mathbf{q}} \pm \frac{\boldsymbol{\iota}_k^* s_{-(k-i)}[\widetilde{\mathbf{a}}_0]}{\|\mathbf{a}_i\|_2} \right\|_2 \\ &\leq \left\| \left(\frac{\mathbf{Y} \mathbf{Y}^T}{\theta m} \right)^{1/2} (\mathbf{A}_0 \mathbf{A}_0^T)^{-1/2} \mathbf{I} \right\|_2 \left\| \frac{\boldsymbol{\iota}_k^* s_{-(k-i)}[\widetilde{\mathbf{a}}_0]}{\|\mathbf{a}_i\|_2} \right\|_2 \\ &\quad + \sqrt{2 \left(1 - \left| \left\langle \bar{\mathbf{q}}, \frac{\mathbf{a}_i}{\|\mathbf{a}_i\|_2} \right\rangle \right| \right)} \left\| \left(\frac{\mathbf{Y} \mathbf{Y}^T}{\theta m} \right)^{1/2} \right\|_2. \end{aligned} \quad (6.38)$$

Finally, using the fact that for any nonzero vectors \mathbf{u} and \mathbf{v} that $\langle \mathbf{u}, \mathbf{v} \rangle \geq 0$,

$$\left\| \frac{\mathbf{u}}{\|\mathbf{u}\|_2} - \frac{\mathbf{v}}{\|\mathbf{v}\|_2} \right\|_2 \leq \frac{\sqrt{2}}{\|\mathbf{v}\|_2} \|\mathbf{u} - \mathbf{v}\|_2 \quad (6.39)$$

always holds. Therefore,

$$\begin{aligned} & \|\bar{\mathbf{a}} \pm \mathcal{P}_{\mathbb{S}}[\boldsymbol{\iota}_k s_i[\widetilde{\mathbf{a}}_0]]\|_2 \\ &= \left\| \mathcal{P}_{\mathbb{S}} \left[\left(\frac{\mathbf{Y} \mathbf{Y}^T}{\theta m} \right)^{1/2} \bar{\mathbf{q}} \right] \pm \mathcal{P}_{\mathbb{S}}[\boldsymbol{\iota}_k s_i[\widetilde{\mathbf{a}}_0]] \right\|_2 \end{aligned} \quad (6.40)$$

$$\leq \frac{\sqrt{2} \|\mathbf{a}_i\|_2}{\|\boldsymbol{\iota}_k^* s_{i-k}[\widetilde{\mathbf{a}}_0]\|_2} \left\| \left(\frac{\mathbf{Y} \mathbf{Y}^T}{\theta m} \right)^{1/2} \bar{\mathbf{q}} \pm \frac{\boldsymbol{\iota}_k^* s_{i-k}[\widetilde{\mathbf{a}}_0]}{\|\mathbf{a}_i\|_2} \right\|_2 \quad (6.41)$$

$$\leq \kappa \sqrt{2(1+\delta) \left(1 - \left| \left\langle \bar{\mathbf{q}}, \frac{\mathbf{a}_i}{\|\mathbf{a}_i\|_2} \right\rangle \right| \right)} \quad (6.42)$$

$$+ \sqrt{2\kappa} \left\| \left(\frac{\mathbf{Y}\mathbf{Y}^T}{\theta m} \right)^{1/2} (\mathbf{A}_0 \mathbf{A}_0^T)^{-1/2} - \mathbf{I} \right\|_2$$

$$\leq 2\kappa \sqrt{2 \left(1 - \left| \left\langle \bar{\mathbf{q}}, \frac{\mathbf{a}_i}{\|\mathbf{a}_i\|_2} \right\rangle \right| \right)} + \sqrt{2\kappa^3 \delta / \sigma_{\min}} \quad (6.43)$$

(Lemma 5.2)

$$\leq 4\sqrt{c_\star} + 10\sqrt{2\kappa^3 \sigma_{\min}^{-1}} \sqrt{k \log m / m} \quad (6.44)$$

$$\leq 4\sqrt{c_\star} + ck^{-1}, \quad (6.45)$$

completing the proof. \blacksquare

7 Concentration for Gradient (Lemma 7.1)

Lemma 7.1 Suppose $\mathbf{x}_0 \sim_{\text{i.i.d.}} \text{BG}(\theta)$. There exists a positive constant C such that whenever

$$m \geq C \frac{\min \left\{ (2C_\star \mu)^{-1}, \kappa^2 k^2 \right\}}{(1-\theta)^2 \sigma_{\min}^2} \kappa^8 k^4 \log^3 \left(\frac{\kappa k}{(1-\theta) \sigma_{\min}} \right) \quad (7.1)$$

and $\theta > \log k/k$, then with probability no smaller than $1 - c_1 \exp(-k) - c_2 k^{-4} - 2 \exp(-\theta k) - 24k \exp\left(-\frac{1}{144} \min\{k, 3\sqrt{\theta m}\}\right) - 48k^{-7} - 48m^{-5}$,

$$\left\| \text{grad}[\psi](\mathbf{q}) - \frac{3(1-\theta)}{\theta m^2} \text{grad}[\varphi](\mathbf{q}) \right\|_2 \leq c \frac{1-\theta}{\theta m^2} \frac{\|\mathbf{A}^T \mathbf{q}\|_4^6}{\kappa^2}, \quad (7.2)$$

holds for all $\mathbf{q} \in \hat{\mathcal{R}}_{2C_\star}$ with positive constant $c \leq 3/(2C_\star)$.

Proof Denote $\boldsymbol{\eta} = \mathbf{Y}^T (\mathbf{Y}\mathbf{Y}^T)^{-1/2} \mathbf{q}$ and $\bar{\boldsymbol{\eta}} = \mathbf{Y}^T (\theta m \mathbf{A}_0 \mathbf{A}_0^T)^{-1/2} \mathbf{q} = (\theta m)^{-1/2} \mathbf{X}_0^T \boldsymbol{\zeta}$, then

$$\begin{aligned} & \left\| \text{grad}[\psi](\mathbf{q}) - \frac{3(1-\theta)}{\theta m^2} \text{grad}[\varphi](\mathbf{q}) \right\|_2 \\ &= \left\| \mathbf{P}_{\mathbf{q}^\perp} \left[\frac{1}{m} (\mathbf{Y}\mathbf{Y}^T)^{-1/2} \mathbf{Y} \boldsymbol{\eta}^{\circ 3} - \frac{3(1-\theta)}{\theta m^2} \mathbf{A} \boldsymbol{\zeta}^{\circ 3} \right] \right\|_2 \\ &\leq \underbrace{\frac{1}{m} \left\| (\mathbf{Y}\mathbf{Y}^T)^{-1/2} \mathbf{Y} \boldsymbol{\eta}^{\circ 3} - (\theta m)^{-1/2} \mathbf{A} \mathbf{X}_0 \boldsymbol{\eta}^{\circ 3} \right\|_2}_{\Delta_1^g} \\ &\quad + \underbrace{\frac{1}{\theta^{1/2} m^{3/2}} \left\| \mathbf{A} \mathbf{X}_0 \boldsymbol{\eta}^{\circ 3} - \mathbf{A} \mathbf{X}_0 \bar{\boldsymbol{\eta}}^{\circ 3} \right\|_2}_{\Delta_2^g} \\ &\quad + \underbrace{\left\| \mathbf{P}_{\mathbf{q}^\perp} \left[\frac{1}{\theta^{1/2} m^{3/2}} \mathbf{A} \mathbf{X}_0 \bar{\boldsymbol{\eta}}^{\circ 3} - \frac{3(1-\theta)}{\theta m^2} \mathbf{A} \boldsymbol{\zeta}^{\circ 3} \right] \right\|_2}_{\Delta_3^g}. \end{aligned}$$

First, let us note that

$$C(1-\theta)^{-2} \sigma_{\min}^{-2} \kappa^{10} k^6 \log^3 \left(\frac{\kappa k}{(1-\theta) \sigma_{\min}} \right)$$

$$\leq C \left(\frac{\kappa k}{\sigma_{\min} (1 - \theta)} \right)^{10} \log^3 \left(\frac{\kappa k}{(1 - \theta) \sigma_{\min}} \right) \quad (7.3)$$

$$\leq C \left(\frac{\kappa k}{(1 - \theta) \sigma_{\min}} \right)^{13}, \quad (7.4)$$

hence

$$\frac{\log^3 \left(C (1 - \theta)^{-2} \sigma_{\min}^{-2} \kappa^{10} k^6 \log^3 \left((1 - \theta)^{-1} \sigma_{\min}^{-1} \kappa k \right) \right)}{C \log^3 \left((1 - \theta)^{-1} \sigma_{\min}^{-1} \kappa k \right)}$$

$$\leq \left(\frac{\log C + 13 \log \left((1 - \theta)^{-1} \sigma_{\min}^{-1} \kappa k \right)}{C^{1/3} \log \left((1 - \theta)^{-1} \sigma_{\min}^{-1} \kappa k \right)} \right)^3 \quad (7.5)$$

$$\leq \left(\frac{\log C}{C^{1/3} \log \left((1 - \theta)^{-1} \sigma_{\min}^{-1} \kappa k \right)} + \frac{13}{C^{1/3}} \right)^3 \quad (7.6)$$

$$\leq \left(\frac{1}{C^{1/6}} + \frac{1}{2} \frac{1}{C^{1/6}} \right)^3 \quad (C \geq 10^8) \quad (7.7)$$

$$\leq \frac{4}{C^{1/2}}. \quad (7.8)$$

Given

$$m \geq C \frac{\min \left\{ (2C_* \mu)^{-1}, \kappa^2 k^2 \right\}}{(1 - \theta)^2 \sigma_{\min}^2} \kappa^8 k^4 \log^3 \left(\frac{\kappa k}{\sigma_{\min} (1 - \theta)} \right), \quad (7.9)$$

as the ratio $\log^3 m / m$ decreases with increasing m , then

$$\begin{aligned} \frac{\log^3 m}{m} &\leq \frac{\log^3 \left(\frac{C \kappa^{10} k^6}{(1 - \theta)^2 \sigma_{\min}^2} \log^3 \left(\frac{\kappa k}{(1 - \theta) \sigma_{\min}} \right) \right)}{C \log^3 \left(\frac{\kappa k}{(1 - \theta) \sigma_{\min}} \right)} \\ &\quad \times \frac{(1 - \theta)^2 \sigma_{\min}^2}{\min \left\{ (2C_* \mu)^{-1}, \kappa^2 k^2 \right\} \kappa^8 k^4} \end{aligned} \quad (7.10)$$

$$\leq \frac{4}{C^{1/2}} \frac{(1 - \theta)^2 \sigma_{\min}^2}{\min \left\{ (2C_* \mu)^{-1}, \kappa^2 k^2 \right\} \kappa^8 k^4} \quad (7.11)$$

According to Lemma 5.1, following inequality always holds

$$\left\| \frac{1}{\theta m} \mathbf{X}_0 \mathbf{X}_0^T - \mathbf{I} \right\|_2 \leq \delta \quad (7.12)$$

$$\leq 10 \sqrt{k \log m / m} \quad (7.13)$$

$$\leq \frac{20 (1 - \theta) \sigma_{\min} \max \left\{ (2C_* \mu)^{1/2}, (\kappa k)^{-1} \right\}}{C^{1/4} \kappa^4 k^{3/2} \log m} \quad (7.14)$$

$$\leq \frac{20 \sigma_{\min} (1 - \theta) \|\mathbf{A}^T \mathbf{q}\|_4^6}{C^{1/4} \kappa^3 \kappa^2 k \log m}, \quad \forall \mathbf{q} \in \hat{\mathcal{R}}_{2C_*}. \quad (7.15)$$

with probability no smaller than $1 - \varepsilon_0$ with $\varepsilon_0 = 2 \exp(-\theta k) + 24k \exp\left(-\frac{1}{144} \min\{k, 3\sqrt{\theta m}\}\right) + 48k^{-7} + 48m^{-5}$.

Moreover, $4\kappa^3\delta/\sigma_{\min} \leq 1/2$ whenever

$$C \geq \left(\frac{160(1-\theta)}{k \log m} \right)^4, \quad (7.16)$$

whence $\delta \leq 1/(8\kappa^2)$, and Lemma 5.3 implies that

$$\begin{aligned} & \left\| \left(\frac{1}{\theta m} \mathbf{Y} \mathbf{Y}^T \right)^{-1/2} \mathbf{A}_0 - (\mathbf{A}_0 \mathbf{A}_0^T)^{-1/2} \mathbf{A}_0 \right\|_2 \\ & \leq 4\kappa^3\delta/\sigma_{\min} \end{aligned} \quad (7.17)$$

$$\leq \frac{80(1-\theta)}{C^{1/4}k \log m} \frac{\|\mathbf{A}^T \mathbf{q}\|_4^6}{\kappa^2}, \quad \forall \mathbf{q} \in \hat{\mathcal{R}}_{2C_*}. \quad (7.18)$$

At the same time,

$$\|\mathbf{X}_0\|_2 \leq (\theta m)^{1/2} \sqrt{1+\delta} \leq (\theta m)^{1/2} (1+\delta/2). \quad (7.19)$$

Moreover, Lemma 2.5 implies that with probability no smaller than $1 - \varepsilon_B$, we have

$$\|\mathbf{x}_0\|_\infty \leq \sqrt{2} \log^{1/2} \left(\frac{2\theta m}{\varepsilon_B} \right). \quad (7.20)$$

Upper Bound for Δ_1^g . Using Lemma 2.7, on the an event of probability at least $1 - \varepsilon_0 - \varepsilon_B$,

$$\|\boldsymbol{\eta}^{\circ 3}\|_2 = \|\boldsymbol{\eta}\|_6^3 \quad (7.21)$$

$$\leq \left(1 + \frac{4\kappa^3\delta}{\sigma_{\min}} \right)^2 \frac{2k}{\theta m} \|\mathbf{x}_0\|_\infty^2 \quad (7.22)$$

$$\leq \frac{9k}{\theta m} \log \left(\frac{2\theta m}{\varepsilon_B} \right). \quad (7.23)$$

Therefore, we can obtain following upper bound

$$\Delta_1^g = \frac{1}{m} \left\| (\mathbf{Y} \mathbf{Y}^T)^{-1/2} \mathbf{Y} \boldsymbol{\eta}^{\circ 3} - (\theta m)^{-1/2} \mathbf{A} \mathbf{X}_0 \boldsymbol{\eta}^{\circ 3} \right\|_2 \quad (7.24)$$

$$\begin{aligned} & \leq \frac{1}{\theta^{1/2} m^{3/2}} \|\mathbf{X}_0\|_2 \|\boldsymbol{\eta}^{\circ 3}\|_2 \times \\ & \quad \left\| \left(\frac{1}{\theta m} \mathbf{Y} \mathbf{Y}^T \right)^{-1/2} \mathbf{A}_0 - \mathbf{A} \right\|_2 \end{aligned} \quad (7.25)$$

$$\leq \frac{5}{4m} \cdot \frac{4\kappa^3\delta}{\sigma_{\min}} \cdot \frac{9k}{\theta m} \log \left(\frac{2\theta m}{\varepsilon_B} \right) \quad (7.26)$$

$$\leq \frac{900(1-\theta) \log(2\theta m/\varepsilon_B)}{C^{1/4}\theta m^2 \log m} \frac{\|\mathbf{A}^T \mathbf{q}\|_4^6}{\kappa^2} \quad \forall \mathbf{q} \in \hat{\mathcal{R}}_{2C_*}.$$

Upper Bound for Δ_2^g . Similarly, with probability no smaller than $1 - \varepsilon_0 - \varepsilon_B$, together with Lemma 2.7, following upper bound can be obtained

$$\begin{aligned} & \|\boldsymbol{\eta}^{\circ 3} - \bar{\boldsymbol{\eta}}^{\circ 3}\|_2 \\ & = \|\boldsymbol{\eta}^{\circ 3} - \text{diag}(\boldsymbol{\eta}^{\circ 2}) \bar{\boldsymbol{\eta}} + \text{diag}(\boldsymbol{\eta}^{\circ 2}) \bar{\boldsymbol{\eta}} - \bar{\boldsymbol{\eta}}^{\circ 3}\|_2 \end{aligned} \quad (7.27)$$

$$\leq \|\boldsymbol{\eta} - \bar{\boldsymbol{\eta}}\|_2 \|\text{diag}(\boldsymbol{\eta}^{\circ 2})\|_2 + \|\bar{\boldsymbol{\eta}}\|_2 \|\text{diag}(\boldsymbol{\eta}^{\circ 2} - \bar{\boldsymbol{\eta}}^{\circ 2})\|_2 \quad (7.28)$$

$$= \|\boldsymbol{\eta} - \bar{\boldsymbol{\eta}}\|_2 \|\boldsymbol{\eta}\|_\infty^2 + \|\bar{\boldsymbol{\eta}}\|_2 \|\boldsymbol{\eta}^{\circ 2} - \bar{\boldsymbol{\eta}}^{\circ 2}\|_\infty \quad (7.29)$$

$$\leq \|\boldsymbol{\eta} - \bar{\boldsymbol{\eta}}\|_2 \|\boldsymbol{\eta}\|_\infty^2 + \|\bar{\boldsymbol{\eta}}\|_2 \|\boldsymbol{\eta} - \bar{\boldsymbol{\eta}}\|_\infty \|\boldsymbol{\eta} + \bar{\boldsymbol{\eta}}\|_\infty \quad (7.30)$$

$$\leq 4(1 + \delta/2) \frac{4\kappa^3 \delta}{\sigma_{\min}} \frac{k}{\theta m} \log(2\theta m/\varepsilon_B) \times \left[\left(1 + \frac{4\kappa^3 \delta}{\sigma_{\min}}\right)^2 + \left(2 + \frac{4\kappa^3 \delta}{\sigma_{\min}}\right) \right] \quad (7.31)$$

$$\leq \frac{24k}{\theta m} \log(2\theta m/\varepsilon_B) \cdot \frac{4\kappa^3 \delta}{\sigma_{\min}}. \quad (7.32)$$

Therefore, we can obtain following upper bound

$$\Delta_2^g = \frac{1}{\theta^{1/2} m^{3/2}} \|\mathbf{A} \mathbf{X}_0^T \boldsymbol{\eta}^{\circ 3} - \mathbf{A} \mathbf{X}_0^T \bar{\boldsymbol{\eta}}^{\circ 3}\|_2 \quad (7.33)$$

$$\leq \frac{1}{\theta^{1/2} m^{3/2}} \|\mathbf{A}\|_2 \|\mathbf{X}_0\|_2 \|\boldsymbol{\eta}^{\circ 3} - \bar{\boldsymbol{\eta}}^{\circ 3}\|_2 \quad (7.34)$$

$$\leq \frac{5}{4m} \cdot \frac{24k}{\theta m} \log(2\theta m/\varepsilon_B) \cdot \frac{4\kappa^3 \delta}{\sigma_{\min}} \quad (7.35)$$

$$\leq \frac{2400}{C^{1/4}} \frac{1 - \theta}{\theta m^2} \frac{\|\mathbf{A}^T \mathbf{q}\|_4^6}{\kappa^2} \cdot \frac{\log(2\theta m/\varepsilon_B)}{\log m}. \quad (7.36)$$

For both Δ_1^g and Δ_2^g to be bounded by $\frac{1}{2C_\star} \frac{1-\theta}{\theta m^2} \frac{\|\mathbf{A}^T \mathbf{q}\|_4^6}{\kappa^2}$, we set

$$C \geq \left(4800 C_\star \frac{\log(2\theta m/\varepsilon_B)}{\log m} \right)^4. \quad (7.37)$$

Notice that the right hand side is indeed bounded by a numerical constant for all m .

Tail Bound for Δ_3^g . Note that

$$\begin{aligned} & (\mathbf{A}_0 \mathbf{A}_0^T)^{-1/2} \mathbf{Y} \bar{\boldsymbol{\eta}}^{\circ 3} \\ &= (\mathbf{A}_0 \mathbf{A}_0^T)^{-1/2} \mathbf{A}_0 \mathbf{X}_0 \left(\mathbf{Y}^T (\theta m \mathbf{A}_0 \mathbf{A}_0^T)^{-1/2} \mathbf{q} \right)^{\circ 3} \end{aligned} \quad (7.38)$$

$$= (\theta m)^{-3/2} \mathbf{A} \mathbf{X}_0 (\mathbf{X}_0^T \mathbf{A}^T \mathbf{q})^{\circ 3}, \quad (7.39)$$

and its expectation with respect to \mathbf{x}_0

$$\begin{aligned} & \mathbb{E} \left[\frac{1}{m} \mathbf{A} \mathbf{X}_0 (\mathbf{X}_0^T \mathbf{A}^T \mathbf{q})^{\circ 3} \right] \\ &= \mathbb{E} \left[\mathbf{A} \mathbf{x}_i (\mathbf{x}_i^T \mathbf{A}^T \mathbf{q})^3 \right] \end{aligned} \quad (7.40)$$

$$= 3\theta(1 - \theta) \mathbf{A} \boldsymbol{\zeta}^{\circ 3} + 3\theta^2 \|\mathbf{A}^T \mathbf{q}\|_2^2 \mathbf{A} \mathbf{A}^T \mathbf{q} \quad (7.41)$$

$$= 3\theta(1 - \theta) \mathbf{A} \boldsymbol{\zeta}^{\circ 3} + 3\theta^2 \mathbf{q}, \quad (7.42)$$

hence

$$\begin{aligned} & \mathbf{P}_{\mathbf{q}^\perp} \left[\mathbb{E} \left[\frac{1}{m} \mathbf{A} \mathbf{X}_0 (\mathbf{X}_0^T \mathbf{A}^T \mathbf{q})^{\circ 3} \right] \right] \\ &= \mathbf{P}_{\mathbf{q}^\perp} [3\theta(1 - \theta) \mathbf{A} \boldsymbol{\zeta}^{\circ 3}]. \end{aligned} \quad (7.43)$$

Therefore, the Δ_3^g term can be simplified as

$$\Delta_3^g = \left\| \mathbf{P}_{\mathbf{q}^\perp} \left[\frac{1}{\theta^{1/2} m^{3/2}} \mathbf{A} \mathbf{X}_0^T \bar{\boldsymbol{\eta}}^{\circ 3} - \frac{3(1 - \theta)}{\theta m^2} \mathbf{A} \boldsymbol{\zeta}^{\circ 3} \right] \right\|_2 \quad (7.44)$$

$$= \frac{1}{\theta^2 m^2} \left\| \mathbf{P}_{\mathbf{q}^\perp} \left[\frac{\mathbf{A} \mathbf{X}_0 (\mathbf{X}_0^T \boldsymbol{\zeta})^{\circ 3}}{m} - 3\theta (1-\theta) \mathbf{A} \boldsymbol{\zeta}^{\circ 3} \right] \right\|_2 \quad (7.45)$$

$$\leq \frac{1}{\theta^2 m^2} \left\| \mathbf{P}_{\mathbf{q}^\perp} \left[\frac{\mathbf{A} \mathbf{X}_0 (\mathbf{X}_0^T \boldsymbol{\zeta})^{\circ 3}}{m} - \mathbb{E}[\cdot] \right] \right\|_2 + \frac{1}{\theta^2 m^2} \left\| \mathbf{P}_{\mathbf{q}^\perp} [3\theta^2 \mathbf{q}] \right\|_2 \quad (7.46)$$

$$\leq \frac{1}{\theta^2 m^2} \left\| \frac{1}{m} \mathbf{X}_0 (\mathbf{X}_0^T \boldsymbol{\zeta})^{\circ 3} - \mathbb{E}[\cdot] \right\|_2. \quad (7.47)$$

Under the assumption that

$$m \geq \frac{C}{(1-\theta)^2} \min \{ \mu^{-1}, \kappa^2 k^2 \} \kappa^2 k^4 \log^3(\kappa k), \quad (7.48)$$

applying Lemma 7.2, we have

$$\left\| \frac{1}{m} \mathbf{X}_0 (\mathbf{X}_0^T \mathbf{A}^T \mathbf{q})^{\circ 3} - \mathbb{E}[\cdot] \right\|_2 \leq c\theta (1-\theta) \frac{\|\mathbf{A}^T \mathbf{q}\|_4^6}{\kappa^2}. \quad (7.49)$$

with probability larger than $1 - c_2 \exp(-k) - c_2 k^{-4}$. At last, taking $\varepsilon_B = \theta^2 k^{-4}$, we obtain that

$$\begin{aligned} & \left\| \text{grad}[\psi](\mathbf{q}) - \frac{3(1-\theta)}{\theta m^2} \text{grad}[\varphi](\mathbf{q}) \right\|_2 \\ & \leq c \frac{1-\theta}{\theta m^2} \frac{\|\mathbf{A}^T \mathbf{q}\|_4^6}{\kappa^2}, \quad \forall \mathbf{q} \in \hat{\mathcal{R}}_{2C_*} \end{aligned} \quad (7.50)$$

with probability larger than $1 - c_2 \exp(-k) - c_2 k^{-4} - \varepsilon_B - \varepsilon_0$ as desired. \blacksquare

7.1 Proof of Lemma 7.2

Lemma 7.2 Suppose $\mathbf{x}_0 \sim_{\text{i.i.d.}} \text{BG}(\theta) \in \mathbb{R}^m$. There exist positive constant C such that whenever

$$m \geq \frac{C}{(1-\theta)^2} \min \{ (2C_* \mu)^{-1}, \kappa^2 k^2 \} \kappa^2 k^4 \log^3(\kappa k) \quad (7.51)$$

and $\theta k \geq 1$, then with probability no smaller than $1 - c_1 \exp(-k) - c_2 k^{-4}$,

$$\left\| \frac{1}{m} \mathbf{X}_0 (\mathbf{X}_0^T \mathbf{A}^T \mathbf{q})^{\circ 3} - \mathbb{E}[\cdot] \right\|_2 \leq c\theta (1-\theta) \frac{\|\mathbf{A}^T \mathbf{q}\|_4^6}{\kappa^2} \quad (7.52)$$

holds for all $\mathbf{q} \in \hat{\mathcal{R}}_{2C_*}$ with positive constant $c \leq 1/(2C_*)$.

Proof Let $\bar{\mathbf{x}}_i \in \mathbb{R}^{2k-1}$ be generated via

$$\bar{\mathbf{x}}_i = \begin{cases} \mathbf{x}_i & \|\mathbf{x}_i\|_\infty \leq B \text{ and } \|\mathbf{x}_i\|_0 \leq 4\theta k \log m \\ \mathbf{0} & \text{else} \end{cases} \quad (7.53)$$

Let $\bar{\mathbf{X}}_0 \in \mathbb{R}^{(2k-1) \times m}$ denote the circulant submatrix generated by $\bar{\mathbf{x}}_0$. Then $\bar{\mathbf{X}}_0 = \mathbf{X}_0$ obtains whenever

1. $\|\mathbf{x}_0\|_\infty \leq B$, which happens with probability no smaller than $1 - 2\theta m e^{-B^2/2}$ according to Lemma 2.5;

2. $\|\mathbf{x}_i\|_0 \leq 4\theta k \log m$ holds for any index i , applying Lemma 2.4 and Boole's inequality we have

$$\begin{aligned} & \mathbb{E} [\mathbf{1}_{\cup_i \|\mathbf{x}_i\|_0 > 4\theta k \log m}] \\ & \leq m \mathbb{P} [\|\mathbf{x}_i\|_0 > 4\theta k \log m] \end{aligned} \quad (7.54)$$

$$\leq 2m \exp\left(-\frac{3}{4}\theta k \log m\right). \quad (7.55)$$

Denote $\zeta = \mathbf{A}^T \mathbf{q}$ and

$$\mathbf{g}_E = \mathbb{E} \left[\frac{1}{m} \mathbf{X}_0 (\mathbf{X}_0^T \mathbf{A}^T \mathbf{q})^{\circ 3} \right], \quad (7.56)$$

$$\bar{\mathbf{g}}_E = \mathbb{E} \left[\frac{1}{m} \bar{\mathbf{X}}_0 (\bar{\mathbf{X}}_0^T \mathbf{A}^T \mathbf{q})^{\circ 3} \right], \quad (7.57)$$

then,

$$\begin{aligned} & \mathbb{P} \left[\left\| \frac{1}{m} \mathbf{X}_0 (\mathbf{X}_0^T \zeta)^{\circ 3} - \mathbf{g}_E \right\|_2 \geq c\theta(1-\theta) \frac{\|\zeta\|_4^6}{\kappa^2} \right] \\ & \leq \mathbb{P} \left[\left\| \frac{1}{m} \bar{\mathbf{X}}_0 (\bar{\mathbf{X}}_0^T \zeta)^{\circ 3} - \mathbf{g}_E \right\|_2 \geq c\theta(1-\theta) \frac{\|\zeta\|_4^6}{\kappa^2} \right] \\ & \quad + 2\theta m e^{-B^2/2} + 2m \exp\left(-\frac{3}{4}\theta k \log m\right) \end{aligned} \quad (7.58)$$

With triangle inequality, we have

$$\begin{aligned} & \left\| \frac{1}{m} \bar{\mathbf{X}}_0 (\bar{\mathbf{X}}_0^T \zeta)^{\circ 3} - \mathbf{g}_E \right\|_2 \\ & \leq \left\| \mathbb{E} \left[\frac{1}{m} \bar{\mathbf{X}}_0 (\bar{\mathbf{X}}_0^T \zeta)^{\circ 3} \right] - \bar{\mathbf{g}}_E \right\|_2 + \|\bar{\mathbf{g}}_E - \mathbf{g}_E\|_2. \end{aligned} \quad (7.59)$$

Hence, provided

$$\|\bar{\mathbf{g}}_E - \mathbf{g}_E\|_2 \leq \frac{c}{2}\theta(1-\theta) \frac{\|\zeta\|_4^6}{\kappa^2}, \quad (7.60)$$

we have

$$\begin{aligned} & \mathbb{P} \left[\left\| \frac{1}{m} \bar{\mathbf{X}}_0 (\bar{\mathbf{X}}_0^T \zeta)^{\circ 3} - \mathbf{g}_E \right\|_2 \geq c\theta(1-\theta) \frac{\|\zeta\|_4^6}{\kappa^2} \right] \\ & \leq \mathbb{P} \left[\left\| \frac{1}{m} \bar{\mathbf{X}}_0 (\bar{\mathbf{X}}_0^T \zeta)^{\circ 3} - \bar{\mathbf{g}}_E \right\|_2 \geq \frac{c}{2}\theta(1-\theta) \frac{\|\zeta\|_4^6}{\kappa^2} \right]. \end{aligned} \quad (7.61)$$

Truncation Level Next, we choose a large enough entry-wise truncation level B such that the expectation of the gradient $\mathbb{E} \left[\frac{1}{m} \mathbf{X}_0 (\mathbf{X}_0^T \zeta)^{\circ 3} \right]$ is close to that of its truncation $\mathbb{E} \left[\frac{1}{m} \bar{\mathbf{X}}_0 (\bar{\mathbf{X}}_0^T \zeta)^{\circ 3} \right]$. Moreover, we introduce following events notation

$$\mathcal{E}_i \doteq \{\|\mathbf{x}_i\|_\infty > B \cup \|\mathbf{x}_i\|_0 > 4\theta k \log m\}, \quad (7.62)$$

then

$$\begin{aligned} & \|\bar{\mathbf{g}}_E - \mathbf{g}_E\|_2 \\ & = \left\| \mathbb{E} \left[\frac{1}{m} \sum_i \mathbf{x}_i \langle \mathbf{x}_i, \zeta \rangle^3 \cdot \mathbf{1}_{\mathcal{E}_i} \right] \right\|_2 \end{aligned} \quad (7.63)$$

$$\leq \frac{1}{m} \sum_i \left\| \mathbb{E} \left[\mathbf{x}_i \langle \mathbf{x}_i, \boldsymbol{\zeta} \rangle^3 \cdot \mathbf{1}_{\mathcal{E}_i} \right] \right\|_2 \quad (7.64)$$

$$\leq \frac{1}{m} \sum_i \left(\mathbb{E} \left[\left\| \mathbf{x}_i (\mathbf{x}_i^T \boldsymbol{\zeta})^{\circ 3} \right\|_2^2 \right] \cdot \mathbb{E} [\mathbf{1}_{\mathcal{E}_i}] \right)^{1/2} \quad (7.65)$$

$$\leq \left(\mathbb{E} \left[\left\| \mathbf{x}_i \right\|_2^8 \right] \right)^{1/2} \times \sqrt{\mathbb{E} [\mathbf{1}_{\|\mathbf{x}_i\|_\infty > B}] + \mathbb{E} [\mathbf{1}_{\|\mathbf{x}_i\|_0 > 4\theta k \log m}]} \quad (7.66)$$

$$\leq 50k^2 \sqrt{4\theta k e^{-B^2/2} + \exp(-\frac{3}{4}\theta k \log m)} \quad (7.67)$$

By setting

$$B \geq C' \log^{1/2} \left(\frac{\kappa^4 k^8}{\theta (1-\theta)^2} \right), \quad (7.68)$$

we have

$$\theta k e^{-B^2/2} \leq \frac{1}{2} \left(\frac{c}{100} \right)^2 \theta^2 (1-\theta)^2 \frac{\|\boldsymbol{\zeta}\|_4^{12}}{\kappa^4 k^4}. \quad (7.69)$$

In addition, whenever

$$\theta k \geq \frac{4}{3 \log m} \log \left(\frac{400^2 \kappa^4 k^4}{c^2 \theta^2 (1-\theta)^2 \|\boldsymbol{\zeta}\|_4^{12}} \right), \quad (7.70)$$

we have

$$\exp(-\frac{3}{4}\theta k \log m) \leq \frac{1}{2} \left(\frac{c}{100} \right)^2 \theta^2 (1-\theta)^2 \frac{\|\boldsymbol{\zeta}\|_4^{12}}{\kappa^4 k^4}. \quad (7.71)$$

Therefore,

$$\sqrt{4\theta k e^{-B^2/2} + \exp(-\frac{3}{4}\theta k \log m)} \leq \frac{c}{2} \theta (1-\theta) \frac{\|\boldsymbol{\zeta}\|_4^6}{50\kappa^2 k^2}. \quad (7.72)$$

In addition,

$$\left(\mathbb{E} \left[\left\| \mathbf{x}_i \right\|_2^8 \right] \right)^{1/2} \leq (7!! \cdot 2^4 k^4)^{1/2} < 50k^2. \quad (7.73)$$

Plugging in Eq (7.73) and (7.72) back to (7.67), we obtain that

$$\|\bar{\mathbf{g}}_E - \mathbf{g}_E\|_2 \leq \frac{c}{2} \theta (1-\theta) \frac{\|\mathbf{A}^T \mathbf{q}\|_4^6}{\kappa^2}, \quad (7.74)$$

and hence

$$\begin{aligned} & \mathbb{P} \left[\left\| \frac{1}{m} \bar{\mathbf{X}}_0 (\bar{\mathbf{X}}_0^T \boldsymbol{\zeta})^{\circ 3} - \mathbf{g}_E \right\|_2 \geq c\theta (1-\theta) \frac{\|\boldsymbol{\zeta}\|_4^6}{\kappa^2} \right] \\ & \leq \mathbb{P} \left[\left\| \frac{1}{m} \bar{\mathbf{X}}_0 (\bar{\mathbf{X}}_0^T \boldsymbol{\zeta})^{\circ 3} - \bar{\mathbf{g}}_E \right\|_2 \geq \frac{c}{2} \theta (1-\theta) \frac{\|\boldsymbol{\zeta}\|_4^6}{\kappa^2} \right]. \end{aligned} \quad (7.75)$$

Independent Submatrices. To deal with the complicated dependence within the random circulant matrix \mathbf{X}_0 , we break \mathbf{X}_0 into submatrices $\mathbf{X}_1, \dots, \mathbf{X}_{2k-1}$, each of which is (marginally) distributed as a $(2k-1) \times \frac{m}{2k-1}$ i.i.d. BG(θ) random matrix. Indeed, there exists a permutation $\boldsymbol{\Pi}$ such that

$$\mathbf{X}_0 \boldsymbol{\Pi} = [\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_{2k-1}], \quad (7.76)$$

with

$$\mathbf{X}_i = [\mathbf{x}_i, \mathbf{x}_{i+(2k-1)}, \dots, \mathbf{x}_{i+(m-2k-1)}]. \quad (7.77)$$

We apply similar matrix breaking approach for the truncated matrix $\bar{\mathbf{X}}$. The summands within each term $\bar{\mathbf{X}}_i (\bar{\mathbf{X}}_i^T \boldsymbol{\zeta})^{\circ 3}$ are mutually independent and hence is amenable to classical concentration results.

$$\frac{1}{m} \bar{\mathbf{X}}_0 (\bar{\mathbf{X}}_0^T \boldsymbol{\zeta})^{\circ 3} = \frac{1}{m} \sum_{l=1}^m \langle \bar{\mathbf{x}}_l, \boldsymbol{\zeta} \rangle^3 \bar{\mathbf{x}}_l \quad (7.78)$$

$$= \sum_{i=1}^{2k-1} \frac{1}{m} \left(\sum_{j=0}^{\frac{m}{2k-1}-1} \langle \bar{\mathbf{x}}_{i+(2k-1)j}, \boldsymbol{\zeta} \rangle^3 \bar{\mathbf{x}}_{i+(2k-1)j} \right) \quad (7.79)$$

$$= \sum_{i=1}^{2k-1} \frac{1}{m} \bar{\mathbf{X}}_i (\bar{\mathbf{X}}_i^T \boldsymbol{\zeta})^{\circ 3}. \quad (7.80)$$

We conservatively bound the quantity of interest, $\frac{1}{m} \bar{\mathbf{X}}_0 (\bar{\mathbf{X}}_0^T \boldsymbol{\zeta})^{\circ 3}$, by ensuring that for each k , $\bar{\mathbf{X}}_k (\bar{\mathbf{X}}_k^T \boldsymbol{\zeta})^{\circ 3}$ be close to its expectation.

$$\begin{aligned} & \mathbb{P} \left[\left\| \frac{1}{m} \bar{\mathbf{X}}_0 (\bar{\mathbf{X}}_0^T \boldsymbol{\zeta})^{\circ 3} - \bar{\mathbf{g}}_E \right\|_2 \geq \frac{c}{2} \theta (1 - \theta) \frac{\|\boldsymbol{\zeta}\|_4^6}{\kappa^2} \right] \\ & \leq \sum_{i=1}^{2k-1} \mathbb{P} \left[\left\| \frac{1}{m} \bar{\mathbf{X}}_i (\bar{\mathbf{X}}_i^T \boldsymbol{\zeta})^{\circ 3} - \frac{\bar{\mathbf{g}}_E}{2k-1} \right\|_2 \geq \frac{c}{2} \frac{\theta (1 - \theta) \|\boldsymbol{\zeta}\|_4^6}{\kappa^2 (2k-1)} \right] \\ & = \sum_{i=1}^{2k-1} \mathbb{P} \left[\left\| \frac{1}{m} \bar{\mathbf{X}}_i (\bar{\mathbf{X}}_i^T \boldsymbol{\zeta})^{\circ 3} - \bar{\mathbf{g}}_E \right\|_2 \geq \frac{c}{2} \frac{\theta (1 - \theta) \|\boldsymbol{\zeta}\|_4^6}{\kappa^2 (2k-1)} \right] \end{aligned}$$

Applying Bernstein inequality for matrix variables as in Lemma 9.7, with $d_1 = 2k - 1$, $d_2 = 1$, we can obtain that for independent random vectors $\mathbf{v}_1, \dots, \mathbf{v}_n$ with

$$\sigma^2 = \sum_{i=1}^n \mathbb{E}[\|\mathbf{v}_i\|_2^2] \quad (7.81)$$

and ensuring that

$$\|\mathbf{v}_i\|_2 \leq R \quad a.s. \quad (7.82)$$

we obtain that

$$\mathbb{P} \left[\left\| \sum_i \mathbf{v}_i - \mathbb{E}[\cdot] \right\| > t \right] \leq 4k \exp \left(\frac{-t^2/2}{\sigma^2 + 2Rt/3} \right) \quad (7.83)$$

Here, we have used that

$$\left\| \sum_{i=1}^n \mathbb{E}[\mathbf{v}_i \mathbf{v}_i^*] \right\| \leq \text{tr} \sum_{i=1}^n \mathbb{E}[\mathbf{v}_i \mathbf{v}_i^*] \quad (7.84)$$

$$= \sum_{i=1}^n \mathbb{E}[\|\mathbf{v}_i\|_2^2]. \quad (7.85)$$

and

$$\mathbf{w}_i = \bar{\mathbf{x}}_i \langle \bar{\mathbf{x}}_i, \boldsymbol{\zeta} \rangle^3. \quad (7.86)$$

Notice that

$$\|\mathbf{w}_i\|_2 \leq \|\bar{\mathbf{x}}_i\|_2^4 \quad (7.87)$$

$$\leq (4B^2 \theta k \log m)^2 \quad (7.88)$$

$$= 16B^4\theta^2k^2\log m. \quad (7.89)$$

Let us further note that

$$\begin{aligned} & \sum_{\substack{j_1, \\ j_2 \neq j_3 \neq j_4}} \mathbb{E} [\bar{x}_i(j_1)^2 \bar{x}_i(j_2)^2 \zeta_{j_2}^2 \bar{x}_i(j_3)^2 \zeta_{j_3}^2 \bar{x}_i(j_4)^2 \zeta_{j_4}^2] \\ &= 3 \sum_{j_1 \neq j_2 \neq j_3} \mathbb{E} [\bar{x}_i(j_1)^4 \zeta_{j_1}^2 \bar{x}_i(j_2)^2 \zeta_{j_2}^2 \bar{x}_i(j_3)^2 \zeta_{j_3}^2] \\ &+ \sum_{j_1=1}^{2k-1} \mathbb{E} [\bar{x}_i(j_1)^2] \times \\ & \sum_{j_1 \neq j_2 \neq j_3 \neq j_4} \mathbb{E} [\bar{x}_i(j_2)^2 \zeta_{j_2}^2 \bar{x}_i(j_3)^2 \zeta_{j_3}^2 \bar{x}_i(j_4)^2 \zeta_{j_4}^2] \end{aligned} \quad (7.90)$$

$$\leq 2\theta k \times \theta^3 \|\zeta\|_2^6 + 3 \times 3\theta^3 \|\zeta\|_2^6 \quad (7.91)$$

In similar vein, we can obtain that

$$\begin{aligned} & \sum_{j_1, j_2 \neq j_3} \mathbb{E} [\bar{x}_i(j_1)^2 \bar{x}_i(j_2)^2 \zeta_{j_2}^2 \bar{x}_i(j_3)^4 \zeta_{j_3}^4] \\ &= \sum_{j_1} \mathbb{E} [\bar{x}_i(j_1)^2] \sum_{j_2 \neq j_3 \neq j_1} \mathbb{E} [\bar{x}_i(j_2)^2 \zeta_{j_2}^2 \bar{x}_i(j_3)^4 \zeta_{j_3}^4] \\ &+ \sum_{j_1 \neq j_2} \mathbb{E} [\bar{x}_i(j_1)^4 \zeta_{j_1}^2 \bar{x}_i(j_2)^4 \zeta_{j_2}^4] \\ &+ \sum_{j_1 \neq j_2} \mathbb{E} [\bar{x}_i(j_1)^2 \zeta_{j_1}^2 \bar{x}_i(j_2)^6 \zeta_{j_2}^4] \end{aligned} \quad (7.92)$$

$$\leq 2\theta k \times 3\theta^2 \|\zeta\|_2^2 \|\zeta\|_4^4 + (9 + 15) \theta^2 \|\zeta\|_2^2 \|\zeta\|_4^4 \quad (7.93)$$

and

$$\begin{aligned} & \sum_{j_1, j_2} \mathbb{E} [\bar{x}_i(j_1)^2 \bar{x}_i(j_2)^6 \zeta_{j_2}^6] \\ &= \sum_{j_1} \mathbb{E} [\bar{x}_i(j_1)^2] \sum_{j_2 \neq j_1} \mathbb{E} [\bar{x}_i(j_2)^6 \zeta_{j_2}^6] \\ &+ \sum_{j_1} \mathbb{E} [\bar{x}_i(j_1)^8 \zeta_{j_1}^6] \end{aligned} \quad (7.94)$$

$$\leq 2\theta k \times 15\theta \|\zeta\|_6^6 + 105\theta \|\zeta\|_6^6 \quad (7.95)$$

Now we calculate

$$\mathbb{E} [\|\mathbf{w}_i\|_2^2] = \mathbb{E} [\|\bar{x}_i\|_2^2 \langle \bar{x}_i, \zeta \rangle^6] \quad (7.96)$$

$$\begin{aligned} &= \mathbb{E} \left[\sum_{j_1, \dots, j_7} \bar{x}_i(j_1)^2 \prod_{\ell=2}^7 \bar{x}_i(j_\ell) \zeta_{j_\ell} \right] \\ &= 15 \sum_{\substack{j_1, \\ j_2 \neq j_3 \neq j_4}} \mathbb{E} [\bar{x}_i(j_1)^2 \bar{x}_i(j_2)^2 \zeta_{j_2}^2 \bar{x}_i(j_3)^2 \zeta_{j_3}^2 \bar{x}_i(j_4)^2 \zeta_{j_4}^2] \\ &+ 15 \sum_{j_1, j_2 \neq j_3} \mathbb{E} [\bar{x}_i(j_1)^2 \bar{x}_i(j_2)^2 \zeta_{j_2}^2 \bar{x}_i(j_3)^4 \zeta_{j_3}^4] \end{aligned} \quad (7.97)$$

$$+ \sum_{j_1, j_2} \mathbb{E} [\bar{x}_i(j_1)^2 \bar{x}_i(j_2)^6 \zeta_{j_2}^6] \quad (7.98)$$

$$\begin{aligned} &\leq 15\theta^3 \|\zeta\|_2^6 (2\theta k + 9) \\ &\quad + 15\theta^2 \|\zeta\|_4^4 (6 + 24) \\ &\quad + \theta \|\zeta\|_6^6 (30\theta k + 105) \end{aligned} \quad (7.99)$$

$$\leq 150\theta^2 k + 600\theta \quad (7.100)$$

whence for $\theta > 1/k$,

$$\mathbb{E} [\|\mathbf{w}_i\|_2^2] \leq C\theta^2 k, \quad (7.101)$$

and hence

$$\sigma^2 \leq C'\theta^2 m. \quad (7.102)$$

Matrix Bernstein gives that

$$\begin{aligned} &\mathbb{P} [\|\bar{\mathbf{X}}_i (\bar{\mathbf{X}}_i^T \zeta)^{\circ 3} - \mathbb{E}[\cdot]\|_2 \geq t] \\ &\leq 4k \exp \left(\frac{-t^2/2}{C\theta^2 m + C' B^4 \theta^2 k^2 \log^2 kt} \right). \end{aligned} \quad (7.103)$$

Setting $t = \frac{c}{4} \frac{m\theta(1-\theta)\|\zeta\|_4^6}{\kappa^2(2k-1)}$, we obtain that

$$\begin{aligned} &\mathbb{P} \left[\left\| \frac{1}{m} \bar{\mathbf{X}}_i (\bar{\mathbf{X}}_i^T \zeta)^{\circ 3} - \mathbb{E}[\cdot] \right\|_2 \geq \frac{c}{4} \frac{\theta(1-\theta)\|\zeta\|_4^6}{\kappa^2(2k-1)} \right] \\ &\leq 4k \exp \left(- \frac{c'' m (1-\theta)^2 \|\zeta\|_4^{12}}{\kappa^4 k^2 + \theta(1-\theta) B^4 \kappa^2 k^3 \|\zeta\|_4^6} \right) \end{aligned} \quad (7.104)$$

ε -Net Covering To obtain a probability bound for all $\mathbf{q} \in \mathbb{S}^{k-1}$, we choose a set of $\zeta_n = \mathbf{A}^T \mathbf{q}_n$ with $n = 1, \dots, N$. Suppose for any $\mathbf{q} \in \mathbb{S}^{k-1}$, there exists \mathbf{q}_n such that $\|\mathbf{q} - \mathbf{q}_n\|_2 \leq \varepsilon$, then

$$\left\| \frac{1}{m} \bar{\mathbf{X}}_i (\bar{\mathbf{X}}_i^T \zeta)^{\circ 3} - \frac{1}{m} \bar{\mathbf{X}}_i (\bar{\mathbf{X}}_i^T \zeta_n)^{\circ 3} \right\|_2 \leq L \|\mathbf{q} - \mathbf{q}_n\|_2. \quad (7.105)$$

For entry wise bounded $\bar{\mathbf{X}}_i \in \mathbb{R}^{(2k-1) \times \frac{m}{2k-1}}$, we have

$$\|\bar{\mathbf{X}}_i\|_2 \leq \sqrt{2\theta m} B, \quad \|\bar{\mathbf{X}}_i \mathbf{e}_j\|_2 \leq \sqrt{4\theta k} B, \quad (7.106)$$

then the Lipschitz constant L can be bounded as

$$L \leq \frac{1}{m} \|\bar{\mathbf{X}}_i\|_2 \left\| \text{diag}(\bar{\mathbf{X}}_i^T \zeta)^{\circ 2} \right\|_2 \|\bar{\mathbf{X}}_i^T \mathbf{A}^T\|_2 \quad (7.107)$$

$$\leq 8\theta^2 k B^4. \quad (7.108)$$

With triangle inequality, we have

$$\begin{aligned} &\left\| \frac{1}{m} \bar{\mathbf{X}}_i (\bar{\mathbf{X}}_i^T \zeta)^{\circ 3} - \mathbb{E} \left[\frac{1}{m} \bar{\mathbf{X}}_i (\bar{\mathbf{X}}_i^T \zeta)^{\circ 3} \right] \right\|_2 \\ &\leq \left\| \mathbb{E} \left[\frac{1}{m} \bar{\mathbf{X}}_i (\bar{\mathbf{X}}_i^T \zeta)^{\circ 3} \right] - \mathbb{E} \left[\frac{1}{m} \bar{\mathbf{X}}_i (\bar{\mathbf{X}}_i^T \zeta_n)^{\circ 3} \right] \right\|_2 \\ &\quad + \left\| \frac{1}{m} \bar{\mathbf{X}}_i (\bar{\mathbf{X}}_i^T \zeta_n)^{\circ 3} - \mathbb{E} \left[\frac{1}{m} \bar{\mathbf{X}}_i (\bar{\mathbf{X}}_i^T \zeta_n)^{\circ 3} \right] \right\|_2 \end{aligned}$$

$$+ \left\| \frac{1}{m} \bar{\mathbf{X}}_i (\bar{\mathbf{X}}_i^T \boldsymbol{\zeta})^{\circ 3} - \frac{1}{m} \bar{\mathbf{X}}_i (\bar{\mathbf{X}}_i^T \boldsymbol{\zeta}_n)^{\circ 3} \right\|_2 \quad (7.109)$$

$$\leq \left\| \frac{1}{m} \bar{\mathbf{X}}_i (\bar{\mathbf{X}}_i^T \boldsymbol{\zeta}_n)^{\circ 3} - \mathbb{E} \left[\frac{1}{m} \bar{\mathbf{X}}_i (\bar{\mathbf{X}}_i^T \boldsymbol{\zeta}_n)^{\circ 3} \right] \right\|_2 + 2L\varepsilon. \quad (7.110)$$

Hence we need to choose the ε -net to cover the sphere of \mathbf{q} with

$$\varepsilon = \frac{c}{4} \frac{\theta(1-\theta)}{\kappa^2(2k-1)} \min_{\mathbf{q} \in \mathbb{S}^{k-1}} \|\boldsymbol{\zeta}\|_4^6, \quad (7.111)$$

plug in $L \leq 4\theta^2 k B^4$ and number of sample N suffice

$$N \leq \left(\frac{3}{\varepsilon} \right)^k \quad (7.112)$$

$$\leq \exp \left(k \ln \left(\frac{3}{\varepsilon} \right) \right) \quad (7.113)$$

$$\leq \exp \left[k \ln \left(C \frac{\theta^2 \kappa^2 k^4 B^4}{\theta(1-\theta)} \right) \right] \quad (7.114)$$

For $n = 1, \dots, N$, denote

$$P_i(\mathbf{q}_n) = \mathbb{P} \left[\left\| \frac{\bar{\mathbf{X}}_i (\bar{\mathbf{X}}_i^T \boldsymbol{\zeta}_n)^{\circ 3}}{m} - \mathbb{E}[\cdot] \right\|_2 \geq \frac{c\theta(1-\theta)\|\boldsymbol{\zeta}_n\|_4^6}{4\kappa^2(2k-1)} \right], \quad (7.115)$$

then together with union bound over all \mathbf{q}_n , we obtain that,

$$\begin{aligned} & \mathbb{P} \left[\sup_{\mathbf{q} \in \hat{\mathcal{R}}_{2C_*}} \frac{\left\| \frac{1}{m} \bar{\mathbf{X}}_i (\bar{\mathbf{X}}_i^T \boldsymbol{\zeta})^{\circ 3} - \mathbb{E}[\cdot] \right\|_2}{\|\boldsymbol{\zeta}\|_4^6} \geq \frac{c}{2} \frac{\theta(1-\theta)}{\kappa^2(2k-1)} \right] \\ & \leq \sum_{\mathbf{q}_n \in \hat{\mathcal{R}}_{2C_*}} P_i(\mathbf{q}_n) \end{aligned} \quad (7.116)$$

$$\leq N \max_{\mathbf{q}_n \in \hat{\mathcal{R}}_{2C_*}} P_i(\mathbf{q}_n) \quad (7.117)$$

$$\begin{aligned} & \leq 4k \sup_{\mathbf{q} \in \hat{\mathcal{R}}_{2C_*}} \exp \left(- \frac{cm(1-\theta)^2 \|\boldsymbol{\zeta}\|_4^{12}}{\kappa^4 k^2 + \theta(1-\theta) B^4 \kappa^2 k^3 \|\boldsymbol{\zeta}\|_4^6} \right) \times \\ & \exp \left(k \ln \left(\frac{3}{\varepsilon} \right) \right). \end{aligned} \quad (7.118)$$

Hence

$$\begin{aligned} & \mathbb{P} \left[\sup_{\mathbf{q} \in \hat{\mathcal{R}}_{2C_*}} \frac{\left\| \frac{1}{m} \bar{\mathbf{X}}_0 (\bar{\mathbf{X}}_0^T \boldsymbol{\zeta})^{\circ 3} - \mathbb{E}[\cdot] \right\|_2}{\|\boldsymbol{\zeta}\|_4^6} \geq \frac{c}{2} \frac{\theta(1-\theta)}{\kappa^2} \right] \\ & \leq \sum_i \mathbb{P} \left[\sup_{\mathbf{q} \in \hat{\mathcal{R}}_{2C_*}} \frac{\left\| \frac{1}{m} \bar{\mathbf{X}}_i (\bar{\mathbf{X}}_i^T \boldsymbol{\zeta})^{\circ 3} - \mathbb{E}[\cdot] \right\|_2}{\|\boldsymbol{\zeta}\|_4^6} \geq \frac{c\theta(1-\theta)}{2\kappa^2(2k-1)} \right] \\ & \leq (2k-1) \max_i \end{aligned} \quad (7.119)$$

$$\mathbb{P} \left[\sup_{\mathbf{q} \in \hat{\mathcal{R}}_{2C_*}} \frac{\left\| \frac{1}{m} \bar{\mathbf{X}}_i (\bar{\mathbf{X}}_i^T \boldsymbol{\zeta})^{\circ 3} - \mathbb{E}[\cdot] \right\|_2}{\|\boldsymbol{\zeta}\|_4^6} \geq \frac{c\theta(1-\theta)}{2\kappa^2(2k-1)} \right] \quad (7.120)$$

$$\leq 8k^2 \sup_{\mathbf{q} \in \hat{\mathcal{R}}_{2C_*}} \exp \left(- \frac{cm(1-\theta)^2 \|\boldsymbol{\zeta}\|_4^{12}}{\kappa^4 k^2 + \theta(1-\theta) B^4 \kappa^2 k^3 \|\boldsymbol{\zeta}\|_4^6} \right) \times \exp \left(k \ln \left(\frac{3}{\varepsilon} \right) \right), \quad (7.121)$$

which is bounded by $\exp(-k)$ as long as

$$m \geq C \frac{\min \left\{ (2C_* \mu)^{-2}, \kappa^2 k^2 \right\}}{(1-\theta)^2} \kappa^2 k^4 \log^3(\kappa k) \quad (7.122)$$

$$\geq C' k \log \left(\frac{\theta \kappa^2 k^2 B^4}{(1-\theta) \|\boldsymbol{\zeta}\|_4^6} \right) \times \max \left\{ \frac{\kappa^4 k^2}{(1-\theta)^2 \|\boldsymbol{\zeta}\|_4^{12}}, \frac{\theta B^4 \kappa^2 k^3}{(1-\theta) \|\boldsymbol{\zeta}\|_4^6} \right\}. \quad (7.123)$$

To sum up, we obtain that for all $\mathbf{q} \in \hat{\mathcal{R}}_{2C_*}$, inequality

$$\left\| \frac{1}{m} \mathbf{X}_0 (\mathbf{X}_0^T \mathbf{A}^T \mathbf{q})^{\circ 3} - \mathbb{E}[\cdot] \right\|_2 \leq c\theta(1-\theta) \frac{\|\mathbf{A}^T \mathbf{q}\|_4^6}{\kappa^2} \quad (7.124)$$

holds with probability no smaller than $1 - c_1 \exp(-k) - c_2 k^{-4} - c_3 \exp(-\theta k)$. \blacksquare

8 Concentration for Hessian (Lemma 8.1)

Lemma 8.1 Suppose $\mathbf{x}_0 \sim_{\text{i.i.d.}} \text{BG}(\theta)$. There exists positive constant C that whenever

$$m \geq C\theta \frac{\min \left\{ (2C_* \mu \kappa^2)^{-4/3}, k^2 \right\}}{(1-\theta)^2 \sigma_{\min}^2} \kappa^6 k^4 \log^3 \left(\frac{\kappa k}{(1-\theta) \sigma_{\min}} \right) \quad (8.1)$$

and $\theta \geq \log k/k$, then with probability no smaller than $1 - c_1 \exp(-k) - c_2 k^{-4} - 48k^{-7} - 48m^{-5} - 24k \exp\left(-\frac{1}{144} \min\{k, 3\sqrt{\theta m}\}\right)$,

$$\left\| \text{Hess}[\psi](\mathbf{q}) - \frac{3(1-\theta)}{\theta m^2} \text{Hess}[\varphi](\mathbf{q}) \right\|_2 \leq c \frac{1-\theta}{\theta m^2} \|\mathbf{A}^T \mathbf{q}\|_4^4, \quad (8.2)$$

holds for all $\mathbf{q} \in \hat{\mathcal{R}}_{2C_*}$ with positive constant $c \leq 0.048 \leq 3(1 - 6c_* - 36c_*^2 - 24c_*^3)$.

Proof Denote $\boldsymbol{\eta} = \mathbf{Y}^T (\mathbf{Y} \mathbf{Y}^T)^{-1/2} \mathbf{q}$ and $\bar{\boldsymbol{\eta}} = \mathbf{Y}^T (\theta m \mathbf{A}_0 \mathbf{A}_0^T)^{-1/2} \mathbf{q} = (\theta m)^{-1/2} \mathbf{X}_0^T \boldsymbol{\zeta}$, and

$$\mathbf{W} = \left(\frac{1}{\theta m} \mathbf{Y} \mathbf{Y}^T \right)^{-1/2} - (\mathbf{A}_0 \mathbf{A}_0^T)^{-1/2}, \quad (8.3)$$

$$\hat{\mathbf{Y}} = (\mathbf{Y} \mathbf{Y}^T)^{-1/2} \mathbf{Y}. \quad (8.4)$$

Then we have

$$\left\| \text{Hess}[\psi](\mathbf{q}) - \frac{3(1-\theta)}{\theta m^2} \text{Hess}[\varphi](\mathbf{q}) \right\|_2$$

$$\begin{aligned}
&= \left\| \mathbf{P}_{q^\perp} \left[\frac{3}{m} \widehat{\mathbf{Y}} \text{diag}(\boldsymbol{\eta}^{\circ 2}) \widehat{\mathbf{Y}}^T - \langle \mathbf{q}, \nabla \psi(\mathbf{q}) \rangle \mathbf{I} \right] \mathbf{P}_{q^\perp} \right. \\
&\quad \left. - \frac{3(1-\theta)}{\theta m^2} \mathbf{P}_{q^\perp} \left[3\mathbf{A} \text{diag}(\boldsymbol{\zeta}^{\circ 2}) \mathbf{A}^T - \|\boldsymbol{\zeta}\|_4^4 \mathbf{I} \right] \mathbf{P}_{q^\perp} \right\|_2
\end{aligned} \tag{8.5}$$

$$\begin{aligned}
&\leq \left\| \mathbf{P}_{q^\perp} \left[\frac{3}{m} \widehat{\mathbf{Y}} \text{diag}(\boldsymbol{\eta}^{\circ 2}) \widehat{\mathbf{Y}}^T \right] \mathbf{P}_{q^\perp} \right. \\
&\quad \left. - \mathbf{P}_{q^\perp} \left[\frac{9(1-\theta)}{\theta m^2} \mathbf{A} \text{diag}(\boldsymbol{\zeta}^{\circ 2}) \mathbf{A}^T - \frac{3}{m^2} \mathbf{I} \right] \mathbf{P}_{q^\perp} \right\|_2 \\
&\quad + \left\| \left[\langle \mathbf{q}, \nabla \psi(\mathbf{q}) \rangle - \frac{3(1-\theta)}{\theta m^2} \|\boldsymbol{\zeta}\|_4^4 - \frac{3}{m^2} \right] \mathbf{P}_{q^\perp} \right\|_2
\end{aligned} \tag{8.6}$$

$$\begin{aligned}
&\leq \underbrace{\frac{3}{\theta m^2} \left\| \mathbf{W} \mathbf{Y} \text{diag}(\boldsymbol{\eta}^{\circ 2}) \mathbf{Y}^T \left(\frac{1}{\theta m} \mathbf{Y} \mathbf{Y}^T \right)^{-1/2} \right\|_2}_{\Delta_1^H} \\
&\quad + \underbrace{\frac{3}{\theta m^2} \left\| \mathbf{A} \mathbf{X}_0 \text{diag}(\boldsymbol{\eta}^{\circ 2}) \mathbf{Y}^T \mathbf{W} \right\|_2}_{\Delta_2^H} \\
&\quad + \underbrace{\frac{3}{\theta m^2} \left\| \mathbf{A} \mathbf{X}_0 \text{diag}(\boldsymbol{\eta}^{\circ 2} - \bar{\boldsymbol{\eta}}^{\circ 2}) \mathbf{X}_0^T \mathbf{A}^T \right\|_2}_{\Delta_3^H} \\
&\quad + \underbrace{\frac{3}{\theta m^2} \left\| \mathbf{P}_{q^\perp} \left[\mathbf{A} \mathbf{X}_0 \text{diag}(\bar{\boldsymbol{\eta}}^{\circ 2}) \mathbf{X}_0^T \mathbf{A}^T \right] \mathbf{P}_{q^\perp} \right.}_{\Delta_4^H} \\
&\quad \left. - \mathbf{P}_{q^\perp} \left[3(1-\theta) \mathbf{A} \text{diag}(\boldsymbol{\zeta}^{\circ 2}) \mathbf{A}^T + \theta \mathbf{I} \right] \mathbf{P}_{q^\perp} \right\|_2 \\
&\quad + \underbrace{\left\| \left[\langle \mathbf{q}, \nabla \psi(\mathbf{q}) \rangle - \frac{3(1-\theta)}{\theta m^2} \|\boldsymbol{\zeta}\|_4^4 - \frac{3}{m^2} \right] \mathbf{P}_{q^\perp} \right\|_2}_{\Delta_5^H}
\end{aligned} \tag{8.7}$$

In the rest of the proof, we prove that

$$\Delta_i^H \leq \frac{c}{9} \frac{1-\theta}{\theta m^2} \|\boldsymbol{\zeta}\|_4^4, \quad i = 1, 2, 3. \tag{8.8}$$

and

$$\Delta_i^H \leq \frac{c}{3} \frac{1-\theta}{\theta m^2} \|\boldsymbol{\zeta}\|_4^4, \quad i = 4, 5. \tag{8.9}$$

First, let us note that

$$C(1-\theta)^{-2} \sigma_{\min}^{-2} \kappa^6 k^5 \log^3 \left(\frac{\kappa k}{(1-\theta) \sigma_{\min}} \right) \tag{8.10}$$

$$\leq C \left(\frac{\kappa k}{(1-\theta) \sigma_{\min}} \right)^6 \log^3 \left(\frac{\kappa k}{(1-\theta) \sigma_{\min}} \right) \tag{8.11}$$

$$\leq C \left(\frac{\kappa k}{(1-\theta) \sigma_{\min}} \right)^9 \tag{8.12}$$

or

$$\begin{aligned} & \frac{\log^3 \left(C (1-\theta)^{-2} \sigma_{\min}^{-2} \kappa^6 k^5 \log^3 \left(\frac{\kappa k}{(1-\theta) \sigma_{\min}} \right) \right)}{C \log^3 \left(\frac{\kappa k}{(1-\theta) \sigma_{\min}} \right)} \\ & \leq \left(\frac{\log C + 9 \log \left(\frac{\kappa k}{(1-\theta) \sigma_{\min}} \right)}{C^{1/3} \log \left(\frac{\kappa k}{(1-\theta) \sigma_{\min}} \right)} \right)^3 \end{aligned} \quad (8.13)$$

$$\leq \left(\frac{\log C}{C^{1/3} \log \left(\frac{\kappa k}{(1-\theta) \sigma_{\min}} \right)} + \frac{9}{C^{1/3}} \right)^3 \quad (8.14)$$

$$\leq \left(\frac{1}{C^{1/6}} + \frac{1}{2} \frac{1}{C^{1/6}} \right)^3 \quad (C \geq 10^8) \quad (8.15)$$

$$\leq \frac{4}{C^{1/2}}. \quad (8.16)$$

Since

$$m \geq C \frac{\min \left\{ (2C_* \mu \kappa^2)^{-4/3}, k^2 \right\}}{(1-\theta)^2 \sigma_{\min}^2} \kappa^6 k^4 \log^3 \left(\frac{\kappa k}{(1-\theta) \sigma_{\min}} \right), \quad (8.17)$$

as the ratio $\log^3 m/m$ decreases with increasing m , then

$$\begin{aligned} & \frac{\log^3 m}{m} \\ & \leq \frac{\log^3 \left(C \frac{\kappa^6 k^5}{(1-\theta)^2 \sigma_{\min}^2} \log^3 \left(\frac{\kappa k}{\sigma_{\min} (1-\theta)} \right) \right)}{C \log^3 \left(\frac{\kappa k}{\sigma_{\min} (1-\theta)} \right)} \\ & \quad \times \frac{(1-\theta)^2 \sigma_{\min}^2}{\min \left\{ (2C_* \mu \kappa^2)^{-2/3}, k \right\} \kappa^6 k^4} \end{aligned} \quad (8.18)$$

$$\leq \frac{4}{C^{1/2}} \frac{(1-\theta)^2 \sigma_{\min}^2}{\min \left\{ (2C_* \mu \kappa^2)^{-2/3}, k \right\} \kappa^6 k^4} \quad (8.19)$$

According to Lemma 5.1, following inequality obtains

$$\left\| \frac{1}{\theta m} \mathbf{X}_0 \mathbf{X}_0^T - \mathbf{I} \right\|_2 \leq \delta \quad (8.20)$$

$$\leq 10 \sqrt{k \log m/m} \quad (8.21)$$

$$\leq \frac{20 (1-\theta) \sigma_{\min} \max \left\{ (2C_* \mu \kappa^2)^{2/3}, k^{-1} \right\}}{C^{1/4} \kappa^3 k^{3/2} \log m} \quad (8.22)$$

$$\leq \frac{20 \sigma_{\min}}{C^{1/4} \kappa^3} \cdot \frac{(1-\theta) \|\mathbf{A}^T \mathbf{q}\|_4^4}{k^{3/2} \log m}, \quad \forall \mathbf{q} \in \hat{\mathcal{R}}_{2C_*} \quad (8.23)$$

with probability no smaller than $1 - \varepsilon_0$ with $\varepsilon_0 = 2 \exp(-\theta k) + 24k \exp\left(-\frac{1}{144} \min \left\{ k, 3\sqrt{\theta m} \right\}\right) + 48k^{-7} + 48m^{-5}$.

We have $4\kappa^3\delta/\sigma_{\min} \leq 1/2$ whenever

$$C \geq \left(\frac{160(1-\theta)}{k^{3/2} \log m} \right)^4 \quad (8.24)$$

whence $\delta \leq 1/(8\kappa^2)$, and Lemma 5.3 implies that

$$\begin{aligned} & \left\| \left(\frac{1}{\theta m} \mathbf{Y} \mathbf{Y}^T \right)^{-1/2} \mathbf{A}_0 - (\mathbf{A}_0 \mathbf{A}_0^T)^{-1/2} \mathbf{A}_0 \right\|_2 \\ & \leq 4\kappa^3\delta/\sigma_{\min} \end{aligned} \quad (8.25)$$

$$\leq \frac{80(1-\theta) \|\mathbf{A}^T \mathbf{q}\|_4^4}{C^{1/4} k^{3/2} \log m}, \quad \forall \mathbf{q} \in \hat{\mathcal{R}}_{2C_*}. \quad (8.26)$$

Moreover,

$$\|\mathbf{X}_0\|_2 \leq (\theta m)^{1/2} \sqrt{1+\delta} \quad (8.27)$$

$$\leq (\theta m)^{1/2} (1+\delta/2) \quad (8.28)$$

$$\leq \frac{17}{16} (\theta m)^{1/2}. \quad (8.29)$$

Finally, Lemma 2.5 implies that with probability no smaller than $1 - \varepsilon_B$, we have

$$\|\mathbf{x}_0\|_\infty \leq \sqrt{2} \log^{1/2} \left(\frac{2\theta m}{\varepsilon_B} \right). \quad (8.30)$$

Upper Bound for Δ_1^H and Δ_2^H . With probability no smaller than $1 - \varepsilon_0 - \varepsilon_B$, the norms of $\boldsymbol{\eta}$ are upper bounded as in Lemma 2.7,

$$\begin{aligned} \Delta_1^H & \leq \frac{3}{\theta m^2} \left\| \left(\frac{1}{\theta m} \mathbf{Y} \mathbf{Y}^T \right)^{-1/2} \mathbf{A}_0 - \mathbf{A} \right\|_2 \times \\ & \quad \|\mathbf{X}_0\|_2^2 \|\boldsymbol{\eta}\|_\infty^2 \left\| \mathbf{A}_0^T \left(\frac{1}{\theta m} \mathbf{Y} \mathbf{Y}^T \right)^{-1/2} \right\|_2 \end{aligned} \quad (8.31)$$

$$\begin{aligned} & \leq \frac{3}{\theta m^2} \cdot \frac{4\kappa^3\delta}{\sigma_{\min}} \cdot (1+\delta/2)^2 \theta m \cdot \\ & \quad \left(1 + \frac{4\kappa^3\delta}{\sigma_{\min}} \right)^3 \frac{4k}{\theta m} \log(2\theta m/\varepsilon_B) \end{aligned} \quad (8.32)$$

$$\leq \frac{3660}{C^{1/4}} \frac{1-\theta}{\theta m^2} \|\boldsymbol{\zeta}\|_4^4 \cdot \frac{\log(2\theta m/\varepsilon_B)}{k^{1/2} \log m}. \quad (8.33)$$

A similar result holds for

$$\begin{aligned} \Delta_2^H & \leq \frac{3}{\theta m^2} \|\mathbf{X}_0\|_2^2 \|\text{diag}(\boldsymbol{\eta}^{\circ 2})\|_2 \times \\ & \quad \left\| \left(\frac{1}{\theta m} \mathbf{Y} \mathbf{Y}^T \right)^{-1/2} \mathbf{A}_0 - \mathbf{A} \right\|_2 \end{aligned} \quad (8.34)$$

$$\leq \frac{2440}{C^{1/4}} \frac{1-\theta}{\theta m^2} \|\boldsymbol{\zeta}\|_4^4 \cdot \frac{\log(2\theta m/\varepsilon_B)}{k^{1/2} \log m}. \quad (8.35)$$

To make $\Delta_1^H \leq \frac{c}{9} \frac{1-\theta}{\theta m^2} \|\zeta\|_4^4$ and $\Delta_2^H \leq \frac{c}{9} \frac{1-\theta}{\theta m^2} \|\zeta\|_4^4$, we require

$$C \geq \left(9 \times 3660c^{-1} \frac{\log(2\theta m/\varepsilon_B)}{k^{1/2} \log m} \right)^4. \quad (8.36)$$

The right hand side is bounded by an absolute constant for all m .

Upper Bound for Δ_3^H . With probability no smaller than $1 - \varepsilon_0 - \varepsilon_B$, the difference between $\bar{\eta}^{\circ 2}$ and $\eta^{\circ 2}$ is upper bounded as in Lemma 2.7,

$$\begin{aligned} & \|\eta^{\circ 2} - \bar{\eta}^{\circ 2}\|_\infty \\ & \leq \|\eta - \bar{\eta}\|_\infty \|\eta + \bar{\eta}\|_\infty \end{aligned} \quad (8.37)$$

$$\leq \frac{4\kappa^3\delta}{\sigma_{\min}} \left(2 + \frac{4\kappa^3\delta}{\sigma_{\min}} \right) \frac{2k}{\theta m} \log(2\theta m/\varepsilon_B) \quad (8.38)$$

$$\leq \frac{5k}{\theta m} \log(2\theta m/\varepsilon_B) \cdot \frac{4\kappa^3\delta}{\sigma_{\min}}. \quad (8.39)$$

Therefore

$$\Delta_3^H = \frac{3}{\theta m^2} \|\mathbf{A} \mathbf{X}_0 \text{diag}(\eta^{\circ 2} - \bar{\eta}^{\circ 2}) \mathbf{X}_0^T \mathbf{A}^T\|_2 \quad (8.40)$$

$$\leq \frac{3}{\theta m^2} \|\mathbf{A}\|_2^2 \|\mathbf{X}_0\|_2^2 \|\text{diag}(\eta^{\circ 2} - \bar{\eta}^{\circ 2})\|_2 \quad (8.41)$$

$$\leq \frac{15k}{\theta m^2} (1 + \delta/2)^2 \log(2\theta m/\varepsilon_B) \cdot \frac{4\kappa^3\delta}{\sigma_{\min}} \quad (8.42)$$

$$\leq \frac{1400(1-\theta) \log(2\theta m/\varepsilon_B)}{C^{1/4} \theta k^{1/2} m^2 \log m} \|\zeta\|_4^4. \quad (8.43)$$

Again, Δ_3^H is bounded by $\frac{c}{9} \frac{1-\theta}{\theta m^2} \|\zeta\|_4^4$ whenever

$$C \geq \left(9 \times 1400c^{-1} \frac{\log(2\theta m/\varepsilon_B)}{k^{1/2} \log m} \right)^4 \quad (8.44)$$

Upper Bound for Δ_4^H . Recall that

$$\bar{\eta} = \mathbf{Y}^T (\theta m \mathbf{A}_0 \mathbf{A}_0^T)^{-1/2} \mathbf{q}, \quad (8.45)$$

then

$$\begin{aligned} & \mathbb{E} [\mathbf{X}_0 \text{diag}(\bar{\eta}^{\circ 2}) \mathbf{X}_0^T] \\ & = \mathbb{E} \left[\frac{1}{\theta m} \mathbf{X}_0 \text{diag}(\mathbf{X}_0^T \mathbf{A}^T \mathbf{q})^{\circ 2} \mathbf{X}_0^T \right] \end{aligned} \quad (8.46)$$

$$\begin{aligned} & = 3(1-\theta) \text{diag}(\mathbf{A}^T \mathbf{q})^{\circ 2} + 2\theta \mathbf{A}^T \mathbf{q} \mathbf{q}^T \mathbf{A} \\ & \quad + \theta \|\mathbf{A}^T \mathbf{q}\|_2^2 \mathbf{I}, \end{aligned} \quad (8.47)$$

once including the projection $\mathbf{P}_{\mathbf{q}^\perp}$, we have

$$\begin{aligned} & \mathbf{P}_{\mathbf{q}^\perp} \mathbb{E} [\mathbf{A} \mathbf{X}_0 \text{diag}(\bar{\eta}^{\circ 2}) \mathbf{X}_0^T \mathbf{A}^T] \mathbf{P}_{\mathbf{q}^\perp} \\ & = \mathbf{P}_{\mathbf{q}^\perp} [3(1-\theta) \mathbf{A} \text{diag}(\zeta^{\circ 2}) \mathbf{A}^T + \theta \mathbf{I}] \mathbf{P}_{\mathbf{q}^\perp}. \end{aligned} \quad (8.48)$$

Therefore

$$\Delta_4^H = \frac{3}{\theta m^2} \left\| \mathbf{P}_{\mathbf{q}^\perp} [\mathbf{A} \mathbf{X}_0 \text{diag}(\bar{\eta}^{\circ 2}) \mathbf{X}_0^T \mathbf{A}^T] \mathbf{P}_{\mathbf{q}^\perp} \right\|$$

$$- \mathbf{P}_{\mathbf{q}^\perp} \left[3(1-\theta) \mathbf{A} \text{diag}(\boldsymbol{\zeta}^{\circ 2}) \mathbf{A}^T + \theta \mathbf{I} \right] \mathbf{P}_{\mathbf{q}^\perp} \Big\|_2 \quad (8.49)$$

$$\leq \frac{3}{\theta^2 m^2} \left\| \frac{1}{m} \mathbf{X}_0 \text{diag}(\mathbf{X}_0^T \boldsymbol{\zeta})^{\circ 2} \mathbf{X}_0^T - \mathbb{E}[\cdot] \right\|_2 \quad (8.50)$$

Under the assumption for sample size that $m \geq C(1-\theta)^{-2} \kappa^4 \min \left\{ (2C_\star \mu)^{-2/3}, k \right\} k^3 \log^5(\kappa k)$, applying Lemma 8.2, we have

$$\left\| \frac{1}{m} \mathbf{X}_0 \text{diag}(\mathbf{X}_0^T \boldsymbol{\zeta})^{\circ 2} \mathbf{X}_0^T - \mathbb{E}[\cdot] \right\|_2 \leq \frac{c}{9} \theta (1-\theta) \|\boldsymbol{\zeta}\|_4^4. \quad (8.51)$$

simultaneously at every $\mathbf{q} \in \hat{\mathcal{R}}_{2C_\star}$ with probability no smaller than $1 - c_1 \exp(-k) - c_2 k^{-4}$.

Upper Bound for Δ_5^H . Note that this term is essentially the difference between

$$\Delta_5^H = \left\| \left[\langle \mathbf{q}, \nabla \psi(\mathbf{q}) \rangle - \frac{3(1-\theta)}{\theta m^2} \|\boldsymbol{\zeta}\|_4^4 - \frac{3}{m^2} \right] \mathbf{P}_{\mathbf{q}^\perp} \right\|_2 \quad (8.52)$$

$$\leq \left| \langle \mathbf{q}, \nabla \psi(\mathbf{q}) \rangle - \frac{3(1-\theta)}{\theta m^2} \|\boldsymbol{\zeta}\|_4^4 - \frac{3}{m^2} \right| \quad (8.53)$$

$$\begin{aligned} &\leq \frac{1}{\theta^2 m^2} \left| \frac{1}{m} \|\mathbf{X}_0^T \boldsymbol{\zeta}\|_4^4 - 3\theta(1-\theta) \|\boldsymbol{\zeta}\|_4^4 - 3\theta^2 \right| \\ &\quad + \left| \langle \mathbf{q}, \nabla \psi(\mathbf{q}) \rangle - \frac{1}{\theta^2 m^2} \|\mathbf{X}_0^T \boldsymbol{\zeta}\|_4^4 \right| \end{aligned} \quad (8.54)$$

$$\begin{aligned} &\leq \frac{1}{\theta^2 m^2} \left\| \frac{\mathbf{A} \mathbf{X}_0 (\mathbf{X}_0^T \boldsymbol{\zeta})^{\circ 3}}{m} - 3\theta(1-\theta) \mathbf{A}^T \boldsymbol{\zeta}^{\circ 3} - 3\theta^2 \mathbf{q} \right\|_2 \\ &\quad + \frac{1}{m} \left| \|\boldsymbol{\eta}\|_4^4 - \|\bar{\boldsymbol{\eta}}\|_4^4 \right| \end{aligned} \quad (8.55)$$

Recall that

$$\begin{aligned} &\mathbb{E} \left[\frac{1}{m} \mathbf{A} \mathbf{X}_0 (\mathbf{X}_0^T \mathbf{A}^T \mathbf{q})^{\circ 3} \right] \\ &= \mathbb{E} \left[\mathbf{A} \mathbf{x}_i (\mathbf{x}_i^T \mathbf{A}^T \mathbf{q})^3 \right] \end{aligned} \quad (8.56)$$

$$= 3\theta(1-\theta) \mathbf{A} \boldsymbol{\zeta}^{\circ 3} + 3\theta^2 \mathbf{q}, \quad (8.57)$$

With similar argument as in Lemma 7.1, we can show that this term can be bounded by $\frac{c}{6} \frac{1-\theta}{\theta m^2} \|\boldsymbol{\eta}\|_4^4$ whenever

$$m \geq C' \frac{\min \left\{ (\mu \kappa^2)^{-4/3}, k^2 \right\}}{(1-\theta)^2 \sigma_{\min}^2} \kappa^6 k^4 \log^3 \left(\frac{\kappa k}{(1-\theta) \sigma_{\min}} \right). \quad (8.58)$$

Moreover, with probability $1 - \varepsilon_0 - \varepsilon_B$

$$\begin{aligned} &\frac{1}{m} \left| \|\boldsymbol{\eta}\|_4^4 - \|\bar{\boldsymbol{\eta}}\|_4^4 \right| \\ &\leq \frac{1}{m} |\langle \boldsymbol{\eta} - \bar{\boldsymbol{\eta}}, 4\boldsymbol{\eta}^{\circ 3} \rangle| \end{aligned} \quad (8.59)$$

$$\leq \frac{4}{m} \|\boldsymbol{\eta} - \bar{\boldsymbol{\eta}}\|_2 \|\boldsymbol{\eta}\|_6^3 \quad (8.60)$$

$$\leq \frac{16\kappa^3 \delta}{\sigma_{\min} m} (1 + \delta/2) \left(1 + \frac{4\kappa^3 \delta}{\sigma_{\min}} \right)^2 \frac{4k}{\theta m} \log(2\theta m / \varepsilon_B) \quad (8.61)$$

$$\leq \frac{153k}{\theta m^2} \log(2\theta m/\varepsilon_B) \cdot \frac{\kappa^3 \delta}{\sigma_{\min}} \quad (8.62)$$

$$\leq \frac{3060}{C^{1/4}} \frac{(1-\theta)}{\theta m^2} \|\zeta\|_4^4 \cdot \frac{\log(2\theta m/\varepsilon_B)}{k^{1/2} \log m}, \quad (8.63)$$

which is bounded by $\frac{c}{6} \frac{1-\theta}{\theta m^2} \|\zeta\|_4^4$ whenever

$$C \geq \left(6 \times 3060 c^{-1} \frac{(1-\theta) \log(2\theta m/\varepsilon_B)}{k^{1/2} \log m} \right)^4. \quad (8.64)$$

The right hand side is bounded by an absolute constant for all m .

Adding up failure probabilities, we have that with probability larger than $1 - c_2 \exp(-k) - c_2 k^{-4} - \varepsilon_0$,

$$\left\| \text{Hess}[\psi](\mathbf{q}) - \frac{3(1-\theta)}{\theta m^2} \text{Hess}[\varphi](\mathbf{q}) \right\|_2 \leq c \frac{1-\theta}{\theta m^2} \|\mathbf{A}^T \mathbf{q}\|_4^4 \quad (8.65)$$

holds as desired for all $\mathbf{q} \in \hat{\mathcal{R}}_{2C_*}$, where $\varepsilon_0 = 2 \exp(-\theta k) + 24k \exp\left(-\frac{1}{144} \min\left\{k, 3\sqrt{\theta m}\right\}\right) + 48k^{-7} + 48m^{-5}$. \blacksquare

8.1 Proof of Lemma 8.2

Lemma 8.2 Suppose $\mathbf{x}_0 \sim_{\text{i.i.d.}} \text{BG}(\theta)$. There exist constants $C > 0$ that whenever

$$m \geq C \frac{\min\left\{(2C_* \mu \kappa^2)^{-4/3}, k^2\right\}}{(1-\theta)^2} k^4 \log^3(\kappa k), \quad (8.66)$$

and $\theta k > 1$, then with probability no smaller than $1 - c_1 \exp(-k) - c_2 k^{-4}$,

$$\begin{aligned} & \left\| \frac{1}{m} \mathbf{X}_0 \text{diag}(\mathbf{X}_0^T \mathbf{A}^T \mathbf{q})^{\circ 2} \mathbf{X}_0^T - \mathbb{E}[\cdot] \right\|_2 \\ & \leq c\theta(1-\theta) \|\mathbf{A}^T \mathbf{q}\|_4^4, \end{aligned} \quad (8.67)$$

holds for all $\mathbf{q} \in \hat{\mathcal{R}}_{2C_*}$ with positive constant $c \leq 0.005 \leq (1 - 6c_* - 36c_*^2 - 24c_*^3)/3$.

Proof The proof strategy for the finite sample concentration of the Hessian is similar to that of the gradient as presented in Lemma 7.2. For simplicity, we will only demonstrate some key steps here, please refer to Lemma 7.2 for detailed arguments.

Again, from Lemma 2.5, the coefficient satisfies $\|\mathbf{x}_0\|_\infty \leq B$ with probability no smaller than $1 - 2\theta m e^{-B^2/2}$. We write $\bar{\mathbf{x}}_0(i) = \mathbf{x}_0(i) \mathbb{1}_{|\mathbf{x}_0(i)| \leq B}$, and let $\bar{\mathbf{X}}_0$ denote the circulant matrix generated by the truncated vector $\bar{\mathbf{x}}_0$. Denote

$$\mathbf{H}_E = \mathbb{E} \left[\frac{1}{m} \mathbf{X}_0 \text{diag}(\mathbf{X}_0^T \mathbf{A}^T \mathbf{q})^{\circ 2} \mathbf{X}_0^T \right], \quad (8.68)$$

$$\bar{\mathbf{H}}_E = \mathbb{E} \left[\frac{1}{m} \bar{\mathbf{X}}_0 \text{diag}(\bar{\mathbf{X}}_0^T \mathbf{A}^T \mathbf{q})^{\circ 2} \bar{\mathbf{X}}_0^T \right], \quad (8.69)$$

then

$$\mathbb{P} \left[\left\| \frac{\mathbf{X}_0 \text{diag}(\mathbf{X}_0^T \zeta)^{\circ 2} \mathbf{X}_0^T}{m} - \mathbf{H}_E \right\|_2 \geq c\theta(1-\theta) \|\zeta\|_4^4 \right]$$

$$\begin{aligned}
&\leq \mathbb{P} \left[\left\| \frac{\bar{\mathbf{X}}_0 \text{diag} (\bar{\mathbf{X}}_0^T \boldsymbol{\zeta})^{\circ 2} \bar{\mathbf{X}}_0^T}{m} - \mathbf{H}_E \right\|_2 \geq c\theta(1-\theta) \|\boldsymbol{\zeta}\|_4^4 \right] \\
&\quad + 2\theta m e^{-B^2/2} + 2m \exp \left(-\frac{3}{4} \theta k \log m \right)
\end{aligned} \tag{8.70}$$

while via triangle inequality,

$$\begin{aligned}
&\left\| \frac{1}{m} \bar{\mathbf{X}}_0 \text{diag} (\bar{\mathbf{X}}_0^T \mathbf{A}^T \mathbf{q})^{\circ 2} \bar{\mathbf{X}}_0^T - \mathbf{H}_E \right\|_2 \\
&\leq \left\| \frac{1}{m} \bar{\mathbf{X}}_0 \text{diag} (\bar{\mathbf{X}}_0^T \mathbf{A}^T \mathbf{q})^{\circ 2} \bar{\mathbf{X}}_0^T - \bar{\mathbf{H}}_E \right\|_2 \\
&\quad + \|\bar{\mathbf{H}}_E - \mathbf{H}_E\|_2.
\end{aligned} \tag{8.71}$$

Truncation Level. Next, we choose a large enough entry-wise truncation level B such that the expectation of the Hessian $\mathbb{E} [\mathbf{X}_0 \text{diag} (\mathbf{X}_0^T \mathbf{A}^T \mathbf{q})^{\circ 2} \mathbf{X}_0^T]$ is close to that of its truncation $\mathbb{E} [\bar{\mathbf{X}}_0 \text{diag} (\bar{\mathbf{X}}_0^T \mathbf{A}^T \mathbf{q})^{\circ 2} \bar{\mathbf{X}}_0^T]$. Moreover, we introduce following events notation

$$\mathcal{E}_i \doteq \{\|\mathbf{x}_i\|_\infty > B \cup \|\mathbf{x}_i\|_0 > 4\theta k \log m\}, \tag{8.72}$$

then

$$\begin{aligned}
&\|\bar{\mathbf{H}}_E - \mathbf{H}_E\|_2 \\
&= \left\| \mathbb{E} \left[\frac{1}{m} \sum_i \langle \mathbf{x}_i, \boldsymbol{\zeta} \rangle^2 \mathbf{x}_i \mathbf{x}_i^T \cdot \mathbf{1}_{\mathbf{E}_i} \right] \right\|_F
\end{aligned} \tag{8.73}$$

$$\leq \frac{1}{m} \sum_i \left\| \mathbb{E} [\langle \mathbf{x}_i, \boldsymbol{\zeta} \rangle^2 \mathbf{x}_i \mathbf{x}_i^T \cdot \mathbf{1}_{\mathbf{E}_i}] \right\|_F \tag{8.74}$$

$$\leq \frac{1}{m} \sum_i \left(\mathbb{E} \left[\left\| \langle \mathbf{x}_i, \boldsymbol{\zeta} \rangle^2 \mathbf{x}_i \mathbf{x}_i^T \right\|_F^2 \right] \cdot \mathbb{E} [\mathbf{1}_{\mathbf{E}_i}] \right)^{1/2} \tag{8.75}$$

$$\begin{aligned}
&\leq \left(\mathbb{E} [\|\mathbf{x}_i\|_2^8] \right)^{1/2} \times \\
&\quad \sqrt{\mathbb{E} [\mathbf{1}_{\|\mathbf{x}_i\|_\infty > B}] + \mathbb{E} [\mathbf{1}_{\|\mathbf{x}_i\|_0 > 4\theta k \log m}]}
\end{aligned} \tag{8.76}$$

$$\leq 50k^2 \sqrt{4\theta k e^{-B^2/2} + \exp(-\frac{3}{4} \theta k \log m)} \tag{8.77}$$

By setting

$$B \geq C' \log^{1/2} \left(\frac{k^7}{\theta(1-\theta)^2} \right) \tag{8.78}$$

we have

$$\theta k e^{-B^2/2} \leq c' \theta^2 (1-\theta)^2 \frac{\|\boldsymbol{\zeta}\|_4^8}{k^4} \tag{8.79}$$

In addition, whenever

$$\theta k \geq \frac{4}{3 \log m} \log \left(\frac{400^2 k^4}{c^2 \theta^2 (1-\theta)^2 \|\boldsymbol{\zeta}\|_4^8} \right), \tag{8.80}$$

we have

$$\exp\left(-\frac{3}{4}\theta k \log m\right) \leq \frac{1}{2} \left(\frac{c\theta(1-\theta)}{100}\right)^2 \frac{\|\zeta\|_4^8}{k^4}. \quad (8.81)$$

Hence,

$$\sqrt{4\theta k e^{-B^2/2} + \exp\left(-\frac{3}{4}\theta k \log m\right)} \leq \frac{c\theta(1-\theta)}{100k^2} \|\zeta\|_4^4. \quad (8.82)$$

Therefore, we can obtain that

$$\|\bar{\mathbf{H}}_E - \mathbf{H}_E\|_2 \leq \frac{c}{2} \theta (1-\theta) \|\zeta\|_4^4 \quad (8.83)$$

always holds, hence

$$\begin{aligned} & \mathbb{P} \left[\left\| \frac{\bar{\mathbf{X}}_0 \text{diag}(\bar{\mathbf{X}}_0^T \zeta)^{\circ 2} \bar{\mathbf{X}}_0^T}{m} - \mathbf{H}_E \right\|_2 \geq c\theta(1-\theta) \|\zeta\|_4^4 \right] \\ & \leq \mathbb{P} \left[\left\| \frac{\bar{\mathbf{X}}_0 \text{diag}(\bar{\mathbf{X}}_0^T \zeta)^{\circ 2} \bar{\mathbf{X}}_0^T}{m} - \bar{\mathbf{H}}_E \right\|_2 \geq \frac{c}{2} \theta (1-\theta) \|\zeta\|_4^4 \right]. \end{aligned} \quad (8.84)$$

Independent Sub-matrices. As we did in Lemma 7.2, we remove the dependence in \mathbf{X}_0 by sampling every $2k-1$ column such that

$$\mathbf{X}_0 \mathbf{\Pi} = [\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_{2k-1}], \quad (8.85)$$

where

$$\mathbf{X}_i = [\mathbf{x}_i, \mathbf{x}_{i+(2k-1)}, \dots, \mathbf{x}_{i+(m-2k-1)}], \quad (8.86)$$

and $\mathbf{\Pi}$ is a certain permutation of the columns of \mathbf{X}_0 .

Applying Bernstein inequality for matrix variables as in Lemma 9.7, with $\mathbf{M}_i = \langle \bar{\mathbf{x}}_i, \mathbf{A}^T \mathbf{q} \rangle^2 \bar{\mathbf{x}}_i \bar{\mathbf{x}}_i^T \in \mathbb{R}^{(2k-1) \times (2k-1)}$. Since

$$\|\mathbf{M}_i\|_2 = \left\| \langle \bar{\mathbf{x}}_i, \mathbf{A}^T \mathbf{q} \rangle^2 \bar{\mathbf{x}}_i \bar{\mathbf{x}}_i^T \right\|_2 \quad (8.87)$$

$$\leq \|\bar{\mathbf{x}}_i\|_2^4 \quad (8.88)$$

$$\leq 4B^4 k^2 \quad (8.89)$$

and

$$\|\mathbb{E}[\mathbf{M}_i \mathbf{M}_i^*]\| = \|\mathbb{E}[\mathbf{M}_i^* \mathbf{M}_i]\| \quad (8.90)$$

$$= \left\| \mathbb{E} \left[\langle \bar{\mathbf{x}}_i, \mathbf{A}^T \mathbf{q} \rangle^4 \bar{\mathbf{x}}_i \bar{\mathbf{x}}_i^T \bar{\mathbf{x}}_i \bar{\mathbf{x}}_i^T \right] \right\| \quad (8.91)$$

$$= \left\| \mathbb{E} \left[\langle \bar{\mathbf{x}}_i, \zeta \rangle^4 \|\bar{\mathbf{x}}_i\|_2^2 \bar{\mathbf{x}}_i \bar{\mathbf{x}}_i^T \right] \right\| \quad (8.92)$$

$$\leq \mathbb{E} \left[\langle \bar{\mathbf{x}}_i, \zeta \rangle^4 \|\bar{\mathbf{x}}_i\|_2^4 \right], \quad (8.93)$$

we obtain the following upper bound:

$$\mathbb{E} \left[\langle \bar{\mathbf{x}}_i, \zeta \rangle^4 \|\bar{\mathbf{x}}_i\|_2^4 \right]$$

$$= \mathbb{E} \left[\sum_{j_1, j_2}^{2k-1} \bar{\mathbf{x}}_i(j_1)^2 \bar{\mathbf{x}}_i(j_2)^2 \sum_{j_3, \dots, j_6} \prod_{\ell=3}^6 \bar{\mathbf{x}}_i(j_\ell) \zeta_{j_\ell} \right] \quad (8.94)$$

$$= 3\mathbb{E} \left[\sum_{j_1, j_2}^{2k-1} \bar{\mathbf{x}}_i(j_1)^2 \bar{\mathbf{x}}_i(j_2)^2 \sum_{j_3 \neq j_4} \bar{\mathbf{x}}_i(j_3)^2 \zeta_{j_3}^2 \bar{\mathbf{x}}_i(j_4)^2 \zeta_{j_4}^2 \right] \\ + \mathbb{E} \left[\sum_{j_1, j_2}^{2k-1} \bar{\mathbf{x}}_i(j_1)^2 \bar{\mathbf{x}}_i(j_2)^2 \cdot \sum_{j_3} \bar{\mathbf{x}}_i(j_3)^4 \zeta_{j_3}^4 \right] \quad (8.95)$$

$$= 3\mathbb{E} \left[\sum_{\substack{j_1 \neq j_2 \\ \neq j_3 \neq j_4}} \bar{\mathbf{x}}_i(j_1)^2 \bar{\mathbf{x}}_i(j_2)^2 \bar{\mathbf{x}}_i(j_3)^2 \zeta_{j_3}^2 \bar{\mathbf{x}}_i(j_4)^2 \zeta_{j_4}^2 \right] \\ + 3\mathbb{E} \left[\sum_{j_1 \neq j_2 \neq j_3} \bar{\mathbf{x}}_i(j_1)^4 \bar{\mathbf{x}}_i(j_2)^2 \zeta_{j_2}^2 \bar{\mathbf{x}}_i(j_3)^2 \zeta_{j_3}^2 \right] \\ + 6\mathbb{E} \left[\sum_{j_1 \neq j_2} \bar{\mathbf{x}}_i(j_1)^6 \zeta_{j_1}^2 \bar{\mathbf{x}}_i(j_2)^2 \zeta_{j_2}^2 \right] \\ + 6\mathbb{E} \left[\sum_{j_1 \neq j_2 \neq j_3} \bar{\mathbf{x}}_i(j_1)^2 \bar{\mathbf{x}}_i(j_2)^4 \zeta_{j_2}^2 \bar{\mathbf{x}}_i(j_3)^2 \zeta_{j_3}^2 \right] \\ + 6\mathbb{E} \left[\sum_{j_1 \neq j_2} \bar{\mathbf{x}}_i(j_1)^4 \zeta_{j_1}^2 \bar{\mathbf{x}}_i(j_2)^4 \zeta_{j_2}^2 \right] \\ + 2\mathbb{E} \left[\sum_{j_1 \neq j_2} \bar{\mathbf{x}}_i(j_1)^2 \bar{\mathbf{x}}_i(j_2)^6 \zeta_{j_2}^4 \right] \\ + \mathbb{E} \left[\sum_{j_1 \neq j_2 \neq j_3} \bar{\mathbf{x}}_i(j_1)^2 \bar{\mathbf{x}}_i(j_2)^2 \bar{\mathbf{x}}_i(j_3)^4 \zeta_{j_3}^4 \right] \\ + \mathbb{E} \left[\sum_{j_1 \neq j_2} \bar{\mathbf{x}}_i(j_1)^4 \bar{\mathbf{x}}_i(j_2)^4 \zeta_{j_2}^4 \right] \\ + \mathbb{E} \left[\sum_j \bar{\mathbf{x}}_i(j)^8 \zeta_j^4 \right] \quad (8.96)$$

$$\leq (105\theta + 18\theta^2 k + 60\theta^2 k + 12\theta^3 k^2) \|\zeta\|_4^4 \\ + 3(21\theta^2 + 30\theta^2 + 4\theta^4 k^2 + 12\theta^2 k) \|\zeta\|_2^4 \quad (8.97)$$

$$\leq C\theta^3 k^2 \quad (8.98)$$

Assuming $\theta m \geq 1$, hence

$$\sigma^2 = C\theta^3 km. \quad (8.99)$$

Setting $t = \frac{c}{2} \frac{\theta(1-\theta)m \|\zeta\|_4^4}{2k-1}$ in Matrix Bernstein gives that

$$\mathbb{P} \left[\left\| \bar{\mathbf{X}}_i (\bar{\mathbf{X}}_i^T \zeta)^{\circ 3} - \mathbb{E}[\cdot] \right\|_2 > t \right]$$

$$\leq 8k \exp \left(\frac{-t^2/2}{C\theta^3 km + C' B^4 \theta^2 k^2 t} \right), \quad (8.100)$$

we obtain that

$$\begin{aligned} & \mathbb{P} \left[\left\| \frac{\bar{\mathbf{X}}_i \text{diag} (\bar{\mathbf{X}}_i^T \boldsymbol{\zeta})^{\circ 2} \bar{\mathbf{X}}_i^T}{m} - \mathbb{E} [\cdot] \right\|_2 > c \frac{\theta (1-\theta) \|\boldsymbol{\zeta}\|_4^6}{2k-1} \right] \\ & \leq 8k \exp \left(- \frac{cm (1-\theta)^2 \|\boldsymbol{\zeta}\|_4^8}{\theta k^3 + \theta (1-\theta) B^4 k^3 \|\boldsymbol{\zeta}\|_4^4} \right) \end{aligned} \quad (8.101)$$

ε -Net Covering To obtain a probability bound for all $\mathbf{q} \in \mathbb{S}^{k-1}$, we choose a set of $\boldsymbol{\zeta}_n = \mathbf{A}^T \mathbf{q}_n$ with $n = 1, \dots, N$. Since for any $\mathbf{q}, \mathbf{q}' \in \mathbb{S}^{k-1}$ and $\boldsymbol{\zeta}' = \mathbf{A}^T \mathbf{q}'$, we have

$$\begin{aligned} & \left\| \frac{\bar{\mathbf{X}}_i \text{diag} (\bar{\mathbf{X}}_i^T \boldsymbol{\zeta})^{\circ 2} \bar{\mathbf{X}}_i^T}{m} - \frac{\bar{\mathbf{X}}_i \text{diag} (\bar{\mathbf{X}}_i^T \boldsymbol{\zeta}')^{\circ 2} \bar{\mathbf{X}}_i^T}{m} \right\|_2 \\ & = \frac{1}{m} \left\| \bar{\mathbf{X}}_i \text{diag} \left[(\bar{\mathbf{X}}_i^T \boldsymbol{\zeta})^{\circ 2} - (\bar{\mathbf{X}}_i^T \boldsymbol{\zeta}')^{\circ 2} \right] \bar{\mathbf{X}}_i^T \right\|_2 \end{aligned} \quad (8.102)$$

$$\leq \frac{\|\bar{\mathbf{X}}_i\|_2^2}{m} \left\| \text{diag} \left[(\bar{\mathbf{X}}_i^T \boldsymbol{\zeta})^{\circ 2} - (\bar{\mathbf{X}}_i^T \boldsymbol{\zeta}')^{\circ 2} \right] \right\|_2 \quad (8.103)$$

$$\leq \frac{\|\bar{\mathbf{X}}_i\|_2^2}{m} \|\bar{\mathbf{X}}_i^T \boldsymbol{\zeta} + \bar{\mathbf{X}}_i^T \boldsymbol{\zeta}'\|_\infty \|\bar{\mathbf{X}}_i^T \boldsymbol{\zeta} - \bar{\mathbf{X}}_i^T \boldsymbol{\zeta}'\|_\infty \quad (8.104)$$

$$\leq L \|\mathbf{q} - \mathbf{q}'\|_2 \quad (8.105)$$

Then the Lipschitz constant L is upper bounded by

$$L \leq \frac{\|\bar{\mathbf{X}}_i\|_2^2}{m} \|\bar{\mathbf{X}}_i^T \mathbf{A}^T\|_2 (\|\bar{\mathbf{X}}_i^T \boldsymbol{\zeta}\|_\infty + \|\bar{\mathbf{X}}_i^T \boldsymbol{\zeta}'\|_\infty) \quad (8.106)$$

$$\leq \frac{2}{m} \|\bar{\mathbf{X}}_i\|_2^4 \quad (8.107)$$

$$\leq 8\theta^2 m B^4. \quad (8.108)$$

With triangle inequality, we have

$$\begin{aligned} & \left\| \frac{\bar{\mathbf{X}}_i \text{diag} (\bar{\mathbf{X}}_i^T \boldsymbol{\zeta})^{\circ 2} \bar{\mathbf{X}}_i^T}{m} - \mathbb{E} [\cdot] \right\|_2 \\ & \leq \left\| \frac{\bar{\mathbf{X}}_i \text{diag} (\bar{\mathbf{X}}_i^T \boldsymbol{\zeta})^{\circ 2} \bar{\mathbf{X}}_i^T}{m} - \frac{\bar{\mathbf{X}}_i \text{diag} (\bar{\mathbf{X}}_i^T \boldsymbol{\zeta}_n)^{\circ 2} \bar{\mathbf{X}}_i^T}{m} \right\|_2 \\ & \quad + \left\| \mathbb{E} \left[\frac{\bar{\mathbf{X}}_i \text{diag} (\bar{\mathbf{X}}_i^T \boldsymbol{\zeta})^{\circ 2} \bar{\mathbf{X}}_i^T}{m} \right] - \mathbb{E} \left[\frac{\bar{\mathbf{X}}_i \text{diag} (\bar{\mathbf{X}}_i^T \boldsymbol{\zeta}_n)^{\circ 2} \bar{\mathbf{X}}_i^T}{m} \right] \right\|_2 \\ & \quad + \left\| \frac{\bar{\mathbf{X}}_i \text{diag} (\bar{\mathbf{X}}_i^T \boldsymbol{\zeta}_n)^{\circ 2} \bar{\mathbf{X}}_i^T}{m} - \mathbb{E} \left[\frac{\bar{\mathbf{X}}_i \text{diag} (\bar{\mathbf{X}}_i^T \boldsymbol{\zeta}_n)^{\circ 2} \bar{\mathbf{X}}_i^T}{m} \right] \right\|_2 \end{aligned} \quad (8.109)$$

$$\begin{aligned} & \leq \left\| \frac{\bar{\mathbf{X}}_i (\bar{\mathbf{X}}_i^T \boldsymbol{\zeta}_n)^{\circ 2} \bar{\mathbf{X}}_i^T}{m} - \mathbb{E} \left[\frac{\bar{\mathbf{X}}_i (\bar{\mathbf{X}}_i^T \boldsymbol{\zeta}_n)^{\circ 2} \bar{\mathbf{X}}_i^T}{m} \right] \right\|_2 \\ & \quad + 2L\varepsilon \end{aligned} \quad (8.110)$$

Next, we are going to choose the ε -net to cover the sphere of \mathbf{q} with

$$\varepsilon = \frac{c}{4} \frac{\theta(1-\theta)}{(2k-1)L} \min_{\mathbf{q} \in \mathbb{S}^{k-1}} \|\zeta\|_4^4, \quad (8.111)$$

hence the number of samples N is bounded by

$$N = \left(\frac{3}{\varepsilon}\right)^k \quad (8.112)$$

$$\leq \exp(-k \ln \varepsilon) \quad (8.113)$$

$$\leq C \exp \left[k \log \left(\frac{\theta B^4 k^2 m}{1-\theta} \right) \right]. \quad (8.114)$$

For $n = 1, \dots, N$, denote

$$P_i(\mathbf{q}_n) = \mathbb{P} \left[\left\| \frac{\bar{\mathbf{X}}_i \text{diag}(\bar{\mathbf{X}}_i^T \zeta_n)^{\circ 2} \bar{\mathbf{X}}_i^T}{m} - \mathbb{E}[\cdot] \right\|_2 \geq \frac{c\theta(1-\theta)\|\zeta_n\|_4^4}{4(2k-1)} \right] \quad (8.115)$$

together with union bound over all \mathbf{q}_n , we obtain that,

$$\begin{aligned} & \mathbb{P} \left[\sup_{\mathbf{q} \in \hat{\mathcal{R}}_{2C_*}} \frac{\left\| \frac{\bar{\mathbf{X}}_i \text{diag}(\bar{\mathbf{X}}_i^T \zeta)^{\circ 2} \bar{\mathbf{X}}_i^T}{m} - \mathbb{E}[\cdot] \right\|_2}{\|\zeta\|_4^4} \geq \frac{c\theta(1-\theta)}{2(2k-1)} \right] \\ & \leq \sum_{\mathbf{q}_n \in \hat{\mathcal{R}}_{2C_*}} P_i(\mathbf{q}_n) \end{aligned} \quad (8.116)$$

$$\leq N \max_{\mathbf{q}_n \in \hat{\mathcal{R}}_{2C_*}} P_i(\mathbf{q}_n) \quad (8.117)$$

$$\begin{aligned} & \leq 8k \sup_{\mathbf{q} \in \hat{\mathcal{R}}_{2C_*}} \exp \left(-\frac{cm(1-\theta)^2 \|\zeta\|_4^8}{\theta k^3 + \theta(1-\theta) B^4 k^3 \|\zeta\|_4^4} \right) \times \\ & \exp \left(k \ln \left(\frac{3}{\varepsilon} \right) \right). \end{aligned} \quad (8.118)$$

Hence

$$\begin{aligned} & \mathbb{P} \left[\sup_{\mathbf{q} \in \hat{\mathcal{R}}_{2C_*}} \frac{\left\| \frac{\bar{\mathbf{X}}_0 \text{diag}(\bar{\mathbf{X}}_0^T \zeta)^{\circ 2} \bar{\mathbf{X}}_0^T}{m} - \mathbb{E}[\cdot] \right\|_2}{\|\zeta\|_4^4} \geq \frac{c}{2} \theta(1-\theta) \right] \\ & \leq \sum_i \mathbb{P} \left[\sup_{\mathbf{q} \in \hat{\mathcal{R}}_{2C_*}} \frac{\left\| \frac{\bar{\mathbf{X}}_i \text{diag}(\bar{\mathbf{X}}_i^T \zeta)^{\circ 2} \bar{\mathbf{X}}_i^T}{m} - \mathbb{E}[\cdot] \right\|_2}{\|\zeta\|_4^4} \geq \frac{c\theta(1-\theta)}{2(2k-1)} \right] \end{aligned} \quad (8.119)$$

$$\begin{aligned} & \leq (2k-1) \max_i \mathbb{P} \left[\sup_{\mathbf{q} \in \hat{\mathcal{R}}_{2C_*}} \frac{\left\| \frac{\bar{\mathbf{X}}_i \text{diag}(\bar{\mathbf{X}}_i^T \zeta)^{\circ 2} \bar{\mathbf{X}}_i^T}{m} - \mathbb{E}[\cdot] \right\|_2}{\|\zeta\|_4^4} \geq \frac{c\theta(1-\theta)}{2(2k-1)} \right] \end{aligned} \quad (8.120)$$

$$\leq 16k^2 \sup_{\mathbf{q} \in \hat{\mathcal{R}}_{2C_\star}} \exp \left(-\frac{c'm(1-\theta)^2 \|\zeta\|_4^8}{\theta k^3 + \theta(1-\theta)B^4 k^3 \|\zeta\|_4^4} \right) \times \exp \left(k \ln \left(\frac{3}{\varepsilon} \right) \right) \quad (8.121)$$

Therefore, by taking

$$m \geq \frac{C\theta}{(1-\theta)^2} \min \left\{ (2C_\star \mu \kappa^2)^{-4/3}, k^2 \right\} k^4 \log^3 k \quad (8.122)$$

$$\geq C'\theta k \log \left(\frac{\theta k m B^4}{(1-\theta) \|\zeta\|_4^4} \right) \frac{k^3 + (1-\theta) B^4 k^3 \|\zeta\|_4^4}{(1-\theta)^2 \|\zeta\|_4^8} \quad (8.123)$$

and adding up failure probability, we obtain

$$\left\| \frac{1}{m} \mathbf{X}_0 \text{diag}(\mathbf{X}_0^T \mathbf{A}^T \mathbf{q})^{\circ 2} \mathbf{X}_0^T - \mathbb{E}[\cdot] \right\|_2 \leq c\theta(1-\theta) \|\mathbf{A}^T \mathbf{q}\|_4^4 \quad (8.124)$$

with probability no smaller than $1 - c_1 \exp(-k) - c_2 \theta(1-\theta)^2 k^{-4} - c_3 \exp(-\theta k)$. ■

9 Tools

Lemma 9.1 (Moments of the Gaussian Random Variables) *If $X \sim \mathcal{N}(0, \sigma^2)$, then it holds for all integer $p \geq 1$ that*

$$\mathbb{E}[|X|^p] = \sigma^p (p-1)!! \left[\sqrt{\frac{2}{\pi}} \mathbb{1}_{p \text{ odd}} + \mathbb{1}_{p \text{ even}} \right] \quad (9.1)$$

$$\leq \sigma^p (p-1)!! \quad (9.2)$$

Lemma 9.2 (Moments of the χ^2 Random Variables) *If $X \sim \chi^2(n)$, then it holds for all integer $p \geq 1$,*

$$\mathbb{E}[X^p] = 2^p \frac{\Gamma(p + n/2)}{\Gamma(n/2)} \quad (9.3)$$

$$= \prod_{k=1}^p (n + 2k - 2) \leq p!(2n)^{p/2} \quad (9.4)$$

Lemma 9.3 (Moments of the χ Random Variables) *If $X \sim \chi(n)$, then it holds for all integer $p \geq 1$,*

$$\mathbb{E}[X^p] = 2^{p/2} \frac{\Gamma(p/2 + n/2)}{\Gamma(n/2)} \leq p! n^{p/2}. \quad (9.5)$$

Lemma 9.4 (Moment-Control Bernstein's Inequality for Scalar RVs, Theorem 2.10 of [FR13]) *Let X_1, \dots, X_p be i.i.d. real-valued random variables. Suppose that there exist some positive number R and σ^2 such that*

$$\mathbb{E}[|X_k|^m] \leq \frac{m!}{2} \sigma^2 R^{m-2}, \text{ for all integers } m \geq 2.$$

Let $S \doteq \frac{1}{p} \sum_{k=1}^p X_k$, then for all $t > 0$, it holds that

$$\mathbb{P}[|S - \mathbb{E}[S]| \geq t] \leq 2 \exp \left(-\frac{pt^2}{2\sigma^2 + 2Rt} \right). \quad (9.6)$$

Corollary 9.5 (Moment-Control Bernstein's Inequality for Vector RVs, Corollary A.10 of [SQW15]) Let $\mathbf{x}_1, \dots, \mathbf{x}_p \in \mathbb{R}^d$ be i.i.d. random vectors. Suppose there exist some positive number R and σ^2 such that

$$\mathbb{E} [\|\mathbf{x}_k\|^m] \leq \frac{m!}{2} \sigma^2 R^{m-2}, \quad \text{for all integers } m \geq 2.$$

Let $\mathbf{s} = \frac{1}{p} \sum_{k=1}^p \mathbf{x}_k$, then for any $t > 0$, it holds that

$$\mathbb{P} [\|\mathbf{s} - \mathbb{E}[\mathbf{s}]\| \geq t] \leq 2(d+1) \exp \left(-\frac{pt^2}{2\sigma^2 + 2Rt} \right). \quad (9.7)$$

Lemma 9.6 (Moment-Control Bernstein's Inequality for Matrix RVs, Theorem 6.2 of [Tro12]) Let $\mathbf{X}_1, \dots, \mathbf{X}_p \in \mathbb{R}^{d \times d}$ be i.i.d. random, symmetric matrices. Suppose there exist some positive number R and σ^2 such that

$$\mathbb{E} [\mathbf{X}_k^m] \preceq \frac{m!}{2} \sigma^2 R^{m-2} \mathbf{I}, \quad (9.8)$$

$$-\mathbb{E} [\mathbf{X}_k^m] \preceq \frac{m!}{2} \sigma^2 R^{m-2} \mathbf{I}. \quad (9.9)$$

for all integers $m \geq 2$. Let $\mathbf{S} \doteq \frac{1}{p} \sum_{k=1}^p \mathbf{X}_k$, then for all $t > 0$, it holds that

$$\mathbb{P} [\|\mathbf{S} - \mathbb{E}[\mathbf{S}]\| \geq t] \leq 2d \exp \left(-\frac{pt^2}{2\sigma^2 + 2Rt} \right). \quad (9.10)$$

Lemma 9.7 (Bernstein's Inequality for Uncentered Matrix RVs) The matrix Bernstein inequality states that for independent random matrices $\mathbf{M}_1, \dots, \mathbf{M}_n \in \mathbb{R}^{d_1 \times d_2}$, if

$$\sigma^2 = \max \left\{ \left\| \sum_{i=1}^n \mathbb{E}[\mathbf{M}_i \mathbf{M}_i^*] \right\|, \left\| \sum_{i=1}^n \mathbb{E}[\mathbf{M}_i^* \mathbf{M}_i] \right\| \right\}, \quad (9.11)$$

and

$$\|\mathbf{M}_i\|_2 \leq R \quad \text{a.s.}, \quad (9.12)$$

then

$$\mathbb{P} \left[\left\| \sum_i \mathbf{M}_i - \mathbb{E}[\cdot] \right\| > t \right] \leq (d_1 + d_2) \exp \left(\frac{-t^2/2}{\sigma^2 + 2Rt/3} \right). \quad (9.13)$$

Proof For zero mean random matrices

$$\mathbf{M}_1 - \mathbb{E}\mathbf{M}_1, \dots, \mathbf{M}_n - \mathbb{E}\mathbf{M}_n \in \mathbb{R}^{d_1 \times d_2}, \quad (9.14)$$

we have that

$$\|\mathbf{M}_i - \mathbb{E}\mathbf{M}_i\|_2 \leq 2R, \quad (9.15)$$

and

$$\mathbf{0} \preceq \sum_{i=1}^n \mathbb{E}[(\mathbf{M}_i - \mathbb{E}\mathbf{M}_i)(\mathbf{M}_i - \mathbb{E}\mathbf{M}_i)^*] \quad (9.16)$$

$$\preceq \sum_{i=1}^n \mathbb{E}[\mathbf{M}_i \mathbf{M}_i^*], \quad (9.17)$$

$$\mathbf{0} \preceq \sum_{i=1}^n \mathbb{E}[(\mathbf{M}_i - \mathbb{E}\mathbf{M}_i)^*(\mathbf{M}_i - \mathbb{E}\mathbf{M}_i)] \quad (9.18)$$

$$\preceq \sum_{i=1}^n \mathbb{E}[\mathbf{M}_i^* \mathbf{M}_i]. \quad (9.19)$$

Plugging corresponding quantities back to Theorem 1.6 of [Tro12], we obtain that

$$\mathbb{P} \left[\left\| \sum_i \mathbf{M}_i - \mathbb{E}[\cdot] \right\| > t \right] \leq (d_1 + d_2) \exp \left(\frac{-t^2/2}{\sigma^2 + 2Rt/3} \right). \quad (9.20)$$

■

References

- [DIPG99] Victor De la Pena and Evarist Giné. *Decoupling: from dependence to independence*. Springer, 1999.
- [FR13] Simon Foucart and Holger Rauhut. *A Mathematical Introduction to Compressive Sensing*. Springer, 2013.
- [SQW15] Ju Sun, Qing Qu, and John Wright. Complete dictionary recovery over the sphere. preprint, 2015.
- [Tro12] Joel A. Tropp. User-friendly tail bounds for sums of random matrices. *Foundations of Computational Mathematics*, 12(4):389–434, 2012.
- [ZLK⁺17] Yuqian Zhang, Yenson Lau, Han-wen Kuo, Sky Cheung, Abhay Pasupathy, and John Wright. On the global geometry of sphere-constrained sparse blind deconvolution. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017.