A Unified Framework for Outlier-Robust PCA-like Algorithms

Wenzhuo Yang A0096049@NUS.EDU.SG

Department of Mechanical Engineering, National University of Singapore, Singapore 117576

Huan Xu MPEXUH@NUS.EDU.SG

Department of Mechanical Engineering, National University of Singapore, Singapore 117576

1. Restatement of Main Results

Suppose that $\mathbf{x} \in \mathbb{R}^d$ is a random vector with mean 0 and variance $\mathbf{I}_d$. The distribution of $\mathbf{x}$ is denoted by ν which is unknown. We denote its one-dimensional marginal along direction $\mathbf{v} \in \mathcal{S}_d$ by $\bar{\nu}_{\mathbf{v}}$, and assume that $\bar{\nu}_{\mathbf{v}}(\{0\}) < 0.5$ for all $\mathbf{v} \in \mathcal{S}_d$ and it is sub-Gaussian, i.e. there exists $\theta > 0$ such that $\bar{\nu}_{\mathbf{v}}((-\infty, x] \cup [x, +\infty)) \leq \exp(1 - x^2/\theta)$ for all $x > 0$. We define $\mathcal{V}_{\mathbf{v}}(\gamma)$ as follows,

Definition 1. *For any $\gamma \in [0,1]$, let $\delta_\gamma = \min\{\delta \geq 0 | \bar{\nu}_{\mathbf{v}}([-\delta, \delta]) \geq \gamma\}$ and $\gamma_\nu^- = \bar{\nu}_{\mathbf{v}}((-\delta, \delta))$. Then the "tail weight" functions $\mathcal{V}_{\mathbf{v}}$ is defined as follows:*

$$\mathcal{V}_{\mathbf{v}}(\gamma) \triangleq \lim_{\epsilon \downarrow 0} \int_{-\delta_\gamma + \epsilon}^{\delta_\gamma - \epsilon} x^2 \bar{\nu}_{\mathbf{v}}(dx) + (\gamma - \gamma_\nu^-)\delta_r^2.$$

We also define $\mathcal{V}^+(\gamma) \triangleq \sup_{\mathbf{v} \in \mathcal{S}_d} \mathcal{V}_{\mathbf{v}}(\gamma)$ and $\mathcal{V}^-(\gamma) \triangleq \inf_{\mathbf{v} \in \mathcal{S}_d} \mathcal{V}_{\mathbf{v}}(\gamma)$. Note that if ν is spherically symmetric, then $\mathcal{V}^+(\gamma) = \mathcal{V}^-(\gamma)$. In this case, we denote $\mathcal{V}_{\mathbf{v}}(\gamma)$ by $\mathcal{V}(\gamma)$.

Recall that the authentic samples $\mathbf{z}_i$ are generated by the equation $\mathbf{z}_i = \mathbf{A}\mathbf{x}_i + \mathbf{n}_i$ where $\mathbf{x}_i \in \mathbb{R}^d$ are i.i.d. samples of random variable $\mathbf{x}$ and $\mathbf{n}_i$ are independent realizations of $\mathbf{n} \sim \mathcal{N}(0, \mathbf{I}_p)$. For convenience, we let $\zeta \triangleq \max\{\theta, 2\}$.

Let $\mathcal{Z}$ and $\mathcal{O}$ be the index sets of the authentic samples and the outliers, respectively. For fixed $\kappa > 0$, let $\mathbf{X}_s$ be the optimal solution of the PCA-like algorithm in the s^{th} stage. The event "good output is generated at step s" is denoted by $\mathcal{E}(s)$:

$$\mathcal{E}(s) \triangleq \left\{ \sum_{i \in \mathcal{Z}} \alpha_i(s) \langle \mathbf{y}_i \mathbf{y}_i^\top, \mathbf{X}_s \rangle \geq \frac{1}{\kappa} \sum_{i \in \mathcal{O}} \alpha_i(s) \langle \mathbf{y}_i \mathbf{y}_i^\top, \mathbf{X}_s \rangle \right\}.$$

Lemma 1. *The event $\mathcal{E}(s)$ is true for some $1 \leq s \leq s_0$ where $s_0 = \frac{\rho n(1+\kappa)}{\kappa}$.*

Proof. See Section 3. □

1.1. Upper Bound of Subspace Distance

Let $f(B) = \min\left\{2B\|\mathbf{A}\|_2^2 + c\tau, \gamma B\|\mathbf{A}\|_2^2 + c\gamma(d\|\mathbf{A}\|_2 + 1)\right\}$ where $\tau = \max\{p/n, 1\}$ and c is a universal constant and let $\delta_k(\mathbf{A}\mathbf{A}^\top) = \lambda_k(\mathbf{A})^2 - \lambda_{k+1}(\mathbf{A})^2$.

Theorem 1. *Suppose that $\rho < 0.5$ and $\log p \leq n$, then there exists a finite number $s \leq n$ such that the output $\mathbf{X}_s$ of the PCA-like algorithm in the s^{th} stage satisfies the following inequality with high probability,*

$$\|\mathbf{X}_s - \mathbf{\Pi}\|_F \leq R(\mu) + \sqrt{k} \min_{1 \geq \kappa > 2\rho} \sqrt{\frac{f(B_1) + \eta\beta B_0}{\delta_k(\mathbf{A}\mathbf{A}^\top)}}, \tag{1}$$

where

$$R(\mu)=\begin{cases}\frac{8(\gamma[\epsilon_0(\|\mathbf{A}\|_2^2+1)-\mu]_+ +\mu\beta)}{\delta_k(\mathbf{A}\mathbf{A}^\top)}, & \mu\neq 0\\ \min\left\{\frac{8\epsilon_0\gamma(\|\mathbf{A}\|_2^2+1)}{\delta_k(\mathbf{A}\mathbf{A}^\top)}, 2\sqrt{\frac{\epsilon_1 k(\|\mathbf{A}\|_2^2+1)}{\delta_k(\mathbf{A}\mathbf{A}^\top)}}\right\}, & \mu=0,\end{cases}$$

$\epsilon_0=c_0\sqrt{\frac{\log p}{n}}$, $\epsilon_1=c_1\sqrt{\frac{p}{n}}$, $B_0=c_2(\|\mathbf{A}\|_2^2+1)$, $B_1=\kappa+1-\mathcal{V}^-(1-\frac{\rho}{\kappa(1-\rho)})+\epsilon_0+c_3\left(\frac{d\log^3 n}{n}\right)^{\frac{1}{4}}$, *and* c_0,c_1,c_2,c_3 *are universal constants.*

Proof. See Section 3. □

Theorem 2. *Suppose that* $\rho<0.5$ *and* $\log p\leq n$*, the following holds with high probability,*

$$\|\mathbf{X}^*-\mathbf{\Pi}\|_F\leq\sqrt{\frac{2\left[(dB_2+kB_4)\lambda_1(\mathbf{A})^2+kf(B_3)\right]}{\delta_k(\mathbf{A}\mathbf{A}^\top)}}, \tag{2}$$

where

$$B_3=2-\mathcal{V}^-(\frac{\hat{t}}{t})-\mathcal{V}^-(\frac{\hat{t}-\rho n}{t})+\epsilon_1,\ B_4=\min\left\{c_1\sqrt{\frac{p}{n}}, c_2\gamma\sqrt{\frac{\log p}{n}}\right\},$$

B_2 *is the right hand side of* (1), $\epsilon_1=c_3\left(\frac{d\log^3 n}{n}\right)^{\frac{1}{4}}$, *and* c_1,c_2,c_3 *are universal constants.*

Proof. See Section 3. □

1.2. Lower Bound of Expressed Variance

Let $H^*\triangleq\langle\mathbf{A}\mathbf{A}^\top,\mathbf{X}^*\rangle$, $H_s\triangleq\langle\mathbf{A}\mathbf{A}^\top,\mathbf{X}_s\rangle$ and $\overline{H}\triangleq\langle\mathbf{A}\mathbf{A}^\top,\mathbf{\Pi}\rangle$. The lower bound of E.V is shown in the following theorem.

Theorem 3. *Suppose that* $\rho<0.5$*. For any* κ*, there exists a constant* c *such that the following inequalities hold w.h.p,*

$$\begin{aligned}E.V\geq&\frac{(1-\eta)\mathcal{V}^-\left(\frac{\hat{t}}{t}-\frac{\rho}{1-\rho}\right)\mathcal{V}^-\left(1-\frac{\rho}{\kappa(1-\rho)}\right)}{(1+\kappa)\mathcal{V}^+\left(\frac{\hat{t}}{t}\right)}-\frac{10}{\mathcal{V}^+(0.5)}\left(\frac{ck\min\{\tau,\gamma\varsigma\}}{\overline{H}}\right)^{1/2}\\&-\frac{c\{\theta^{\frac{1}{2}}d^{\frac{1}{4}}(\log^{\frac{3}{4}}n)n^{-\frac{1}{4}}\vee\theta[(1+\kappa)/\kappa]^{\frac{3}{2}}(\log^{\frac{3}{2}}n)n^{-\frac{1}{2}}\}}{\mathcal{V}^+(0.5)}\\&-\frac{2(1-\eta)\mu\beta\sqrt{k}}{\mathcal{V}^+(\frac{\hat{t}}{t})\overline{H}}-\max\{1-\lambda_k(\mathbf{X}^*),\lambda_{k+1}(\mathbf{X}^*)\},\end{aligned}$$

where $\mathbf{w}_1^*,\cdots,\mathbf{w}_k^*$ *are the top* k *eigenvectors of* $\mathbf{X}^*$, $\tau=\max\{\frac{p}{n},1\}$ *and* $\varsigma=\max\{\frac{\log p}{n},1\}$.

Proof. See Section 4. □

Lemma 2. *Suppose that* $\mathcal{S}$ *is a sequence of matrices such that for any* $\mathbf{S}_n\in\mathcal{S}$, $\mathbf{S}_n\in\mathbb{S}_+^{p\times p}$ *and* $\lambda_d(\mathbf{S}_n)-\lambda_{d+1}(\mathbf{S}_n)\geq\delta>0$*. Let*

$$\mathbf{A}_n\triangleq\arg\max_{\mathbf{X}\in\mathcal{F}(d)}\langle\mathbf{S}_n,\mathbf{X}\rangle-\mu_n\|\mathbf{X}\|_1,\ \ \mathbf{B}_n\triangleq\arg\max_{\mathbf{X}\in\mathcal{F}(d)}\langle\mathbf{S}_n,\mathbf{X}\rangle.$$

Then if $pd^{3/2}=o(\frac{1}{\mu_n})$ *and* $\mu_n\to 0$ *as* $n\to+\infty$*, we have*

$$\|\mathbf{A}_n-\mathbf{B}_n\|_F\to 0 \text{ as } n\to+\infty,$$

which implies that $\lambda_d(\mathbf{A}_n)\to 1$ *and* $\lambda_{d+1}(\mathbf{A}_n)\to 0$ *as* $n\to+\infty$.

Proof. See Section 4. □

Theorem 4. *(Asymptotic Bound): Consider a sequence of $\{\mathcal{Y}_i, \lambda_i, \beta_i, \gamma_i\}$, where the asymptotic scaling satisfies*

$$n_i \uparrow +\infty, \ \lim_{i \uparrow +\infty} \frac{\log p_i}{n_i} \leq +\infty, \ \lim_{i \uparrow +\infty} \frac{\min\{\frac{p_i}{n_i}, \gamma_i\}}{\sum_{j=1}^k \lambda_j(\mathbf{A}_i)^2} \downarrow 0, \ \frac{n_i}{d_i \log^3 n_i} \uparrow +\infty, \ \frac{d_i}{\sum_{j=1}^k \lambda_j(\mathbf{A}_i)^2} \downarrow 0, \ \mu_i \beta_i \downarrow 0.$$

Let $\rho^* = \limsup \rho_i \leq 0.5$ and suppose $\hat{t} > 0.5n$, then if $\lambda_k(\mathbf{X}^*) \to 1$ and $\lambda_{k+1}(\mathbf{X}^*) \to 0$ as $n_i \uparrow +\infty$, the following holds in probability when $i \uparrow +\infty$,

$$\liminf_i E.V \geq (1-\eta) \max_\kappa \frac{\mathcal{V}^-\left(1 - \frac{\rho^*}{(1-\rho^*)\kappa}\right) \mathcal{V}^-\left(\frac{\hat{t}}{t} - \frac{\rho^*}{1-\rho^*}\right)}{(1+\kappa)\mathcal{V}^+\left(\frac{\hat{t}}{t}\right)}.$$

Furthermore, if $\bar{\mu}_{\mathbf{v}}(\{0\}) = 0$ for all $\mathbf{v} \in \mathcal{S}_d$, then the breakdown point is $\rho^* = 0.5$.

Proof. Notice that k is a constant and $\overline{H}_i = \sum_{j=1}^k \lambda_j(\mathbf{A}_i^\top \mathbf{A}_i)$. Since $\lambda_k(\mathbf{X}^*) \to 1$ and $\lambda_{k+1}(\mathbf{X}^*) \to 0$ as $n_i \uparrow +\infty$, we know that $\max\{1 - \lambda_k(\mathbf{X}^*), \lambda_{k+1}(\mathbf{X}^*)\} \to 0$. Since $\lim_{i \uparrow +\infty} \frac{\log p_i}{n_i} \leq +\infty$, $\lim_{i \uparrow +\infty} \frac{\min\{\frac{p_i}{n_i}, \gamma_i\}}{\overline{H}_i} \downarrow 0$, $\mu_i \beta_i \downarrow 0$ and $\frac{d_i}{\overline{H}_i} \downarrow 0$, we have $\frac{\mu_i \beta_i \sqrt{k}}{\overline{H}_i} \downarrow 0$ and $\frac{k \min\{\tau_i, \gamma_i \varsigma_i\}}{\overline{H}_i} \downarrow 0$. Since $\bar{\mu}_{\mathbf{v}}(\{0\}) =$ for all $\mathbf{v} \in \mathcal{S}_d$, we have $\mathcal{V}^+(c), \mathcal{V}^-(c) > 0$ for any $c > 0$. Hence we obtain this lower bound of E.V. To show that the breakdown point is $\rho^* = 0.5$, one only needs to show that $\liminf_i$ E.V > 0 when $\rho^* < 0.5$. Since $\rho^* < 0.5$, we have $\frac{\rho^*}{(1-\rho)\kappa} < 1$ for $\kappa = 1$, then we only need to show that $\mathcal{V}^-\left(\frac{\hat{t}}{t} - \frac{\rho^*}{1-\rho^*}\right) > 0$. Since $\hat{t} > 0.5n$, we have

$$\mathcal{V}^-\left(\frac{\hat{t}}{t} - \frac{\rho^*}{1-\rho^*}\right) \geq \mathcal{V}^-\left(\frac{0.5n}{(1-\rho^*)n} - \frac{\rho^*}{1-\rho^*}\right) = \mathcal{V}^-\left(\frac{0.5-\rho^*}{1-\rho^*}\right) > 0.$$

□

Corollary 1. *Under the settings of the above theorem, the following holds in probability for some constant C when $i \uparrow +\infty$,*

$$\liminf_i E.V \geq (1-\eta) \left[\frac{\mathcal{V}^-(\frac{\hat{t}}{t}) - C\sqrt{\theta \rho^* \log(1/2\rho^*)}}{\mathcal{V}^+(\frac{\hat{t}}{t})}\right].$$

Proof. See Section 4. □

2. Useful Concentration Results

2.1. Concentration Results for Isotropic Random Vectors

Lemma 3. *(Lemma 2, (Xu et al., 2013)) For any $0 \leq a_1 < a_2 < a_3 \leq 1$ and $\mathbf{v} \in \mathcal{S}_d$, we have*

$$\frac{\mathcal{V}_{\mathbf{v}}(a_2) - \mathcal{V}_{\mathbf{v}}(a_1)}{a_2 - a_1} \leq \frac{\mathcal{V}_{\mathbf{v}}(a_3) - \mathcal{V}_{\mathbf{v}}(a_2)}{a_3 - a_2}.$$

Lemma 4. *(Lemma 3, (Xu et al., 2013)) 1) For any $a \in [0,1]$ and $\mathbf{v} \in \mathcal{S}_d$, we have $\mathcal{V}_{\mathbf{v}}(a) \leq a$. 2) For any $0 \leq a_1 < a_2 \leq 1$ and $\mathbf{v} \in \mathcal{S}_d$, we have*

$$\mathcal{V}_{\mathbf{v}}(a_2) - \mathcal{V}_{\mathbf{v}}(a_1) \leq \frac{a_2 - a_1}{1 - a_1}.$$

Lemma 5. *For any $1 > \epsilon > 0$ and $\kappa \in [\epsilon, 1]$ and $\mathbf{v} \in \mathcal{S}_d$, we have $\mathcal{V}_{\mathbf{v}}(\kappa) - \mathcal{V}_{\mathbf{v}}(\kappa - \epsilon) \leq C\theta\epsilon\log(1/\epsilon)$.*

Proof. By monotonicity, it suffices to prove the result for $\kappa = 1$. Notice that for $K \geq 2\theta$,

$$\begin{aligned} & \mathcal{V}_{\mathbf{v}}(1) - \mathcal{V}_{\mathbf{v}}(1-\epsilon) \\ \leq & \epsilon K^2 + \mathbb{E}_{x \sim \bar{\mu}_{\mathbf{v}}}[x^2 \cdot \mathbf{1}(x > K)] \\ = & \epsilon K^2 + \int_{K^2}^{\infty} \mathbb{P}_{x \sim \bar{\mu}_{\mathbf{v}}}[x^2 > z] dz \\ \leq & \epsilon K^2 + \int_{K^2}^{\infty} \exp(1 - z/\theta) dz \\ = & \epsilon K^2 + e_0 \theta \exp(-K^2/\theta) \end{aligned}$$

Let $K^2 = \theta \log(1/\epsilon)$, then we have $\mathcal{V}_{\mathbf{v}}(1) - \mathcal{V}_{\mathbf{v}}(1-\epsilon) \leq C\theta\epsilon \log(1/\epsilon)$. □

Theorem 5. *(Theorem 7(I), (Xu et al., 2013)) Suppose random vector $\mathbf{n}_i \sim \mathcal{N}(0, I_p)$. Let $\tau \triangleq \max\{p/n, 1\}$. There exist a universal constant $c > 0$ such that with high probability,*

$$\sup_{\mathbf{w} \in \mathcal{S}_p} \frac{1}{t} \sum_{i=1}^{t} (\mathbf{w}^\top \mathbf{n}_i)^2 \leq c\tau.$$

Theorem 6. *There exists an absolute constant $C > 0$, such that with high probability,*

$$\sup_{\mathbf{v} \in \mathcal{S}_d} |\frac{1}{n} \sum_{i=1}^{n} (\mathbf{v}^\top \mathbf{x}_i)^2 - 1| \leq C\theta \sqrt{\frac{d}{n}}.$$

Proof. The proof depends on the following Lemma (Lemma 14 in (Loh & Wainwright, 2012)).

Lemma 6. *If $\mathbf{X} \in \mathbb{R}^{n \times d}$ is a zero-mean sub-Gaussian matrix with parameters $(\mathbf{\Sigma}, \sigma^2)$, then for any fixed (unit) vector $\mathbf{v} \in \mathbb{R}^d$ and any $t > 0$, we have*

$$\mathbb{P}[|\|\mathbf{X}\mathbf{v}\|_2^2 - \mathbb{E}[\|\mathbf{X}\mathbf{v}\|_2^2]| > nt] \leq 2 \exp\left(-cn \min\left(\frac{t^2}{\sigma^4}, \frac{t}{\sigma^2}\right)\right),$$

for a universal constant c.

Consider matrix $\mathbf{Z} \in \mathbb{R}^{n \times d}$ where the i^{th} row is $\mathbf{x}_i^\top$, then for any fixed (unit) vector $\mathbf{v} \in \mathbb{R}^d$ and any $t > 0$, there exists a universal constant c such that

$$\mathbb{P}[|\|\mathbf{Z}\mathbf{v}\|_2^2 - \mathbb{E}[\|\mathbf{Z}\mathbf{v}\|_2^2]| > nt] \leq 2 \exp\left(-cn \min\left(\frac{t^2}{\theta^2}, \frac{t}{\theta}\right)\right).$$

Let $\mathcal{A}$ be a $1/3$ cover of $\mathcal{S}_d$, then for any $\mathbf{v} \in \mathcal{S}_d$, there is some $\mathbf{u} \in \mathcal{A}$ such that $\|\mathbf{u} - \mathbf{v}\|_2 \leq 1/3$. It is known that $|\mathcal{A}| \leq 9^d$. Define $\psi(\mathbf{v}_1, \mathbf{v}_2) = |\mathbf{v}_1^\top \left(\frac{\mathbf{Z}^\top \mathbf{Z}}{n} - \frac{\mathbb{E}[\mathbf{Z}^\top \mathbf{Z}]}{n}\right) \mathbf{v}_2|$, then we have

$$\sup_{\mathbf{v} \in \mathcal{S}_d} \psi(\mathbf{v}, \mathbf{v}) \leq \max_{\mathbf{u} \in \mathcal{A}} \psi(\mathbf{u}, \mathbf{u}) + 2 \sup_{\mathbf{v} \in \mathcal{S}_d} \psi(\mathbf{v} - \mathbf{u}, \mathbf{u}) + \sup_{\mathbf{v} \in \mathcal{S}_d} \psi(\mathbf{v} - \mathbf{u}, \mathbf{v} - \mathbf{u}).$$

Since $\|\mathbf{u} - \mathbf{v}\|_2 \leq \frac{1}{3}$, we have

$$\sup_{\mathbf{v} \in \mathcal{S}_d} \psi(\mathbf{v}, \mathbf{v}) \leq \max_{\mathbf{u} \in \mathcal{A}} \psi(\mathbf{u}, \mathbf{u}) + (\frac{2}{3} + \frac{1}{9}) \sup_{\mathbf{v} \in \mathcal{S}_d} \psi(\mathbf{v}, \mathbf{v}).$$

Hence $\sup_{\mathbf{v} \in \mathcal{S}_d} \psi(\mathbf{v}, \mathbf{v}) \leq \frac{9}{2} \max_{\mathbf{u} \in \mathcal{A}} \psi(\mathbf{u}, \mathbf{u})$. By the lemma above and the union bound,

$$\mathbb{P}[\sup_{\mathbf{v} \in \mathcal{S}_d} \psi(\mathbf{v}, \mathbf{v}) > t] \leq \mathbb{P}[\frac{9}{2} \max_{\mathbf{u} \in \mathcal{A}} \psi(\mathbf{u}, \mathbf{u}) > t] \leq 9^d \cdot 2 \exp\left(-cn \min\left(\frac{t^2}{\theta^2}, \frac{t}{\theta}\right)\right).$$

Thus, we have

$$\mathbb{P}[\sup_{\mathbf{v} \in \mathcal{S}_d} |\frac{1}{n} \sum_{i=1}^{n} (\mathbf{v}^\top \mathbf{x}_i)^2 - 1| > t] \leq 2 \exp\left(-c_1 n \min\left(\frac{t^2}{\theta^2}, \frac{t}{\theta}\right) + c_2 d\right).$$

Let the right hand side be d^{-10}, then $t = C\theta\sqrt{\frac{d}{n}}$ for constant C and large enough n. □

Lemma 7. *With high probability, the following holds uniformly over $\bar{n} \leq n$ and $\mathbf{v} \in \mathcal{S}_d$,*

$$\left| \frac{1}{n} \sum_{i=1}^{\bar{n}} [\mathbf{v}^\top \mathbf{x}]_{(i)}^2 - \mathcal{V}_{\mathbf{v}}(\bar{n}/n) \right| \leq \frac{Cn\sqrt{d \log n/n}}{n - \bar{n}},$$

for a universal constant C.

Proof. The proof is similar to the proof of Theorem 11 (Xu et al., 2013). We just need to replace $\mathcal{V}$ with $\mathcal{V}_{\mathbf{v}}$. □

Theorem 7. *With high probability, the following holds uniformly over $\bar{n} \leq n$ and $\mathbf{v} \in \mathcal{S}_d$,*

$$\left| \frac{1}{n} \sum_{i=1}^{\bar{n}} [\mathbf{v}^\top \mathbf{x}]_{(i)}^2 - \mathcal{V}_{\mathbf{v}}(\bar{n}/n) \right| \leq C \max\left\{ \theta \sqrt{\frac{d}{n}}, \theta^{1/2} d^{1/4} (\log n)^{3/4} n^{-1/4} \right\},$$

for a universal constant C.

Proof. Follow the proof of Corollary 5 in (Xu et al., 2013). As shown above, Theorem 6 and Lemma 7 hold w.h.p. Under the condition of Theorem 6 and Lemma 7, we define n_0

$$n_0 = (1 - \Theta(\theta^{-1/2} d^{1/4} n^{-1/4} \log^{-1/4} n))n.$$

If $\bar{n} \leq n_0$, then Lemma 7 leads to

$$\left| \frac{1}{n} \sum_{i=1}^{\bar{n}} [\mathbf{v}^\top \mathbf{x}]_{(i)}^2 - \mathcal{V}_{\mathbf{v}}(\bar{n}/n) \right| \leq C\theta^{1/2} d^{1/4} (\log n)^{3/4} n^{-1/4}.$$

If $\bar{n} > n_0$, we have

$$\begin{aligned}
&\frac{1}{n} \sum_{i=1}^{\bar{n}} [\mathbf{v}^\top \mathbf{x}]_{(i)}^2 - \mathcal{V}_{\mathbf{v}}(\bar{n}/n) \\
\leq &\left| \frac{1}{n} \sum_{i=1}^{n} [\mathbf{v}^\top \mathbf{x}]_{(i)}^2 - 1 \right| + |1 - \mathcal{V}_{\mathbf{v}}(\bar{n}/n)| \\
\leq &C_1 \theta \sqrt{\frac{d}{n}} + C_2 \theta \frac{n - n_0}{n} \log \frac{n}{n - n_0} \\
\leq &C \max\left\{ \theta \sqrt{\frac{d}{n}}, \theta^{1/2} d^{1/4} (\log n)^{3/4} n^{-1/4} \right\}.
\end{aligned}$$

On the other hand,

$$\begin{aligned}
&\mathcal{V}_{\mathbf{v}}(\bar{n}/n) - \frac{1}{n} \sum_{i=1}^{\bar{n}} [\mathbf{v}^\top \mathbf{x}]_{(i)}^2 \\
\leq &\mathcal{V}_{\mathbf{v}}(\bar{n}/n) - \frac{1}{n} \sum_{i=1}^{n_0} [\mathbf{v}^\top \mathbf{x}]_{(i)}^2 \\
\leq &\left| \frac{1}{n} \sum_{i=1}^{n_0} [\mathbf{v}^\top \mathbf{x}]_{(i)}^2 - \mathcal{V}_{\mathbf{v}}(n_0/n) \right| + |\mathcal{V}_{\mathbf{v}}(n_0/n) - \mathcal{V}_{\mathbf{v}}(\bar{n}/n)| \\
\leq &C_1 \frac{n\sqrt{d \log n/n}}{n - n_0} + C_2 \theta \frac{n - n_0}{n} \log \frac{n}{n - n_0} \\
\leq &C\theta^{1/2} d^{1/4} (\log n)^{3/4} n^{-1/4}.
\end{aligned}$$

Hence this theorem holds. □

2.2. Concentration Results for Non-isotropic Random Vectors

Lemma 8. *There exists a constant $c > 0$ such that with high probability,*

$$\sup_{\mathbf{v}\in\mathcal{S}_p}\left|\frac{1}{t}\sum_{i=1}^{t}(\mathbf{v}^\top\mathbf{z}_i)^2-\mathbf{v}^\top\mathbf{\Sigma}\mathbf{v}\right|\le c\zeta\|\mathbf{\Sigma}\|_2\sqrt{\frac{p}{n}}.$$

Proof. Recall that $\mathbf{z}=[\mathbf{A},\mathbf{I}_p]\mathbf{u}$ where $\mathbf{u}$ is a sub-gaussian random variable with mean zero and variance $\mathbf{I}_{p+d}$. Denote $(\mathbf{A},\mathbf{I}_p)^\top$ by $\bar{\mathbf{A}}$, since $\bar{\mathbf{A}}^\top\bar{\mathbf{A}}=\mathbf{\Sigma}$ and $d\le p$, then with high probability

$$\begin{aligned}
&\sup_{\mathbf{v}\in\mathcal{S}_p}\left|\frac{1}{t}\sum_{i=1}^{n}(\mathbf{v}^\top\mathbf{z}_i)^2-\mathbf{v}^\top\mathbf{\Sigma}\mathbf{v}\right|=\sup_{\mathbf{v}\in\mathcal{S}_p}\left|\frac{1}{t}\sum_{i=1}^{n}((\bar{\mathbf{A}}\mathbf{v})^\top\mathbf{u})^2-\mathbf{v}^\top\bar{\mathbf{A}}^\top\bar{\mathbf{A}}\mathbf{v}\right|\\
\le&\sup_{\mathbf{v}\in\mathcal{S}_p\cap\{\mathbf{v}:\bar{\mathbf{A}}\mathbf{v}\neq 0\}}\|\bar{\mathbf{A}}^\top\bar{\mathbf{A}}\|_2\cdot\left|\frac{1}{n}\sum_{i=1}^{n}\frac{((\bar{\mathbf{A}}\mathbf{v})^\top\mathbf{u})^2}{\|\bar{\mathbf{A}}\mathbf{v}\|^2}-1\right|\le\|\mathbf{\Sigma}\|_2\cdot\sup_{\mathbf{q}\in\mathcal{S}_{p+d}}\left|\frac{1}{n}\sum_{i=1}^{n}(\mathbf{q}^\top\mathbf{u})^2-1\right|\\
\le& c\zeta\|\mathbf{\Sigma}\|_2\sqrt{\frac{p}{n}}.
\end{aligned}$$

□

Lemma 9. *Let $\tau\triangleq\max\{p/n,1\}$. There exists a constant $C>0$ such that with high probability,*

$$\sup_{\mathbf{v}\in\mathcal{S}_p}\frac{1}{t}\sum_{i=1}^{t}(\mathbf{v}^\top\mathbf{z}_i)^2\le 2\|\mathbf{A}\|_2^2(1+c_1\theta\sqrt{\frac{d}{n}})+c_2\tau.$$

Proof. Consider the following inequalities

$$\begin{aligned}
&\sup_{\mathbf{v}\in\mathcal{S}_p}\frac{1}{t}\sum_{i=1}^{t}(\mathbf{v}^\top\mathbf{z}_i)^2=\sup_{\mathbf{v}\in\mathcal{S}_p}\frac{1}{t}\sum_{i=1}^{t}(\mathbf{v}^\top(\mathbf{A}\mathbf{x}_i+\mathbf{n}_i))^2\\
\le&\sup_{\mathbf{v}\in\mathcal{S}_p}\frac{1}{t}\sum_{i=1}^{t}(\mathbf{v}^\top\mathbf{A}\mathbf{x}_i)^2+\sup_{\mathbf{v}\in\mathcal{S}_p}\frac{1}{t}\sum_{i=1}^{t}(\mathbf{v}^\top\mathbf{n}_i)^2+\sup_{\mathbf{v}\in\mathcal{S}_p}\frac{2}{t}\sum_{i=1}^{t}(\mathbf{v}^\top\mathbf{A}\mathbf{x}_i\mathbf{n}_i^\top\mathbf{v})\\
\le&2\left(\sup_{\mathbf{v}\in\mathcal{S}_p}\frac{1}{t}\sum_{i=1}^{t}(\mathbf{v}^\top\mathbf{A}\mathbf{x}_i)^2+\sup_{\mathbf{v}\in\mathcal{S}_p}\frac{1}{t}\sum_{i=1}^{t}(\mathbf{v}^\top\mathbf{n}_i)^2\right)\\
=&2\left(\sup_{\mathbf{v}\in\mathcal{S}_p\cap\{\mathbf{v}:\mathbf{A}^\top\mathbf{v}\neq 0\}}\|\mathbf{v}^\top\mathbf{A}\|_2^2\cdot\frac{1}{t}\sum_{i=1}^{t}(\frac{\mathbf{v}^\top\mathbf{A}\mathbf{x}_i}{\|\mathbf{v}^\top\mathbf{A}\|_2})^2+\sup_{\mathbf{v}\in\mathcal{S}_p}\frac{1}{t}\sum_{i=1}^{t}(\mathbf{v}^\top\mathbf{n}_i)^2\right)\\
\le&2\|\mathbf{A}\|_2^2(1+c_1\theta\sqrt{\frac{d}{n}})+c_2\tau
\end{aligned}$$

where the last inequality follows from Theorem 5 and Theorem 6. □

Lemma 10. *Let $\tau\triangleq\max\{p/n,1\}$. There exists a universal constant c such that with high probability the following holds uniformly over $\bar{n}\le n$,*

$$\sup_{\mathbf{v}\in\mathcal{S}_p}\frac{1}{n}\sum_{i=1}^{n-\bar{n}}|\mathbf{v}^\top\mathbf{z}|_{[i]}^2\le 2\|\mathbf{A}\|_2^2\left(1-\mathcal{V}^-(\bar{n}/n)+c_1\theta\sqrt{\frac{d}{n}}+c_2\theta^{1/2}d^{1/4}(\log n)^{3/4}n^{-1/4}\right)+c\tau.$$

Proof. Consider the following inequalities

$$\begin{aligned}
&\sup_{\mathbf{v}\in\mathcal{S}_p}\frac{1}{n}\sum_{i=1}^{n-\bar{n}}|\mathbf{v}^\top\mathbf{z}|_{[i]}^2=\sup_{\mathbf{v}\in\mathcal{S}_p}\frac{1}{n}\sum_{i=1}^{n-\bar{n}}|\mathbf{v}^\top(\mathbf{A}\mathbf{x}+\mathbf{n})|_{[i]}^2\\
\le&2\left(\sup_{\mathbf{v}\in\mathcal{S}_p}\frac{1}{n}\sum_{i=1}^{n-\bar{n}}|\mathbf{v}^\top\mathbf{A}\mathbf{x}|_{[i]}^2+\sup_{\mathbf{v}\in\mathcal{S}_p}\frac{1}{n}\sum_{i=1}^{n-\bar{n}}|\mathbf{v}^\top\mathbf{n}|_{[i]}^2\right).
\end{aligned}$$

Note that

$$\sup_{\mathbf{v}\in\mathcal{S}_p}\frac{1}{n}\sum_{i=1}^{n-\bar{n}}|\mathbf{v}^\top\mathbf{A}\mathbf{x}|^2_{[i]}\le\sup_{\mathbf{v}\in\mathcal{S}_p\cap\{\mathbf{v}:\mathbf{A}^\top\mathbf{v}\neq 0\}}\|\mathbf{A}\mathbf{A}^\top\|_2\cdot\frac{1}{n}\sum_{i=1}^{n-\bar{n}}\left|\frac{\mathbf{v}^\top\mathbf{A}\mathbf{x}}{\|\mathbf{v}^\top\mathbf{A}\|_2}\right|^2_{[i]}$$
$$=\|\mathbf{A}\mathbf{A}^\top\|_2\cdot\sup_{\mathbf{v}\in\mathcal{S}_d}\left|\frac{1}{n}\sum_{i=1}^{n}|\mathbf{v}^\top\mathbf{x}|^2-\frac{1}{n}\sum_{i=1}^{\bar{n}}|\mathbf{v}^\top\mathbf{x}|^2_{(i)}\right|$$

From Theorem 6 and Theorem 7, we know that

$$\sup_{\mathbf{v}\in\mathcal{S}_d}\left|\frac{1}{n}\sum_{i=1}^{n}|\mathbf{v}^\top\mathbf{x}|^2-\frac{1}{n}\sum_{i=1}^{\bar{n}}|\mathbf{v}^\top\mathbf{x}|^2_{(i)}-(1-\mathcal{V}_{\mathbf{v}}(\bar{n}/n))\right|$$
$$\le\sup_{\mathbf{v}\in\mathcal{S}_d}\left|\frac{1}{n}\sum_{i=1}^{n}|\mathbf{v}^\top\mathbf{x}|^2-1\right|+\sup_{\mathbf{v}\in\mathcal{S}_d}\left|\frac{1}{n}\sum_{i=1}^{\bar{n}}|\mathbf{v}^\top\mathbf{x}|^2_{(i)}-\mathcal{V}_{\mathbf{v}}(\bar{n}/n)\right|$$
$$\le c_1\theta\sqrt{\frac{d}{n}}+c_2\max\left\{\theta\sqrt{\frac{d}{n}},\theta^{1/2}d^{1/4}(\log n)^{3/4}n^{-1/4}\right\},$$

which implies that

$$\sup_{\mathbf{v}\in\mathcal{S}_p}\frac{1}{n}\sum_{i=1}^{n-\bar{n}}|\mathbf{v}^\top\mathbf{A}\mathbf{x}|^2_{[i]}\le\|\mathbf{A}\|_2^2\cdot\left(1-\mathcal{V}^-(\bar{n}/n)+c_1\theta\sqrt{\frac{d}{n}}+c_2\theta^{1/2}d^{1/4}(\log n)^{3/4}n^{-1/4}\right).\tag{3}$$

Similarly, for the term $\sup_{\mathbf{v}\in\mathcal{S}_p}\frac{1}{n}\sum_{i=1}^{n-\bar{n}}|\mathbf{v}^\top\mathbf{n}|^2_{[i]}$, since $n\sim\mathcal{N}(0,\mathbf{I}_p)$, from Theorem 5 we have

$$\sup_{\mathbf{v}\in\mathcal{S}_p}\frac{1}{n}\sum_{i=1}^{n-\bar{n}}|\mathbf{v}^\top\mathbf{n}|^2_{[i]}\le\sup_{\mathbf{v}\in\mathcal{S}_p}\frac{1}{n}\sum_{i=1}^{n}|\mathbf{v}^\top\mathbf{n}|^2\le c\tau.$$

Hence we obtain this theorem. □

Lemma 11. *With high probability the following holds uniformly over $\bar{n}\le n$ for every matrix $\mathbf{X}\in\mathbb{R}^{p\times p}$,*

$$\frac{1}{n}\sum_{i=1}^{n-\bar{n}}\langle\mathbf{z}\mathbf{z}^\top,\mathbf{X}\rangle_{[i]}\le\min\left\{\left[2\left(1-\mathcal{V}^-(\bar{n}/n)+\epsilon(d)\right)\|\mathbf{A}\|_2^2+c\tau\right]\|\mathbf{X}\|_*\right.,$$
$$\left.\left[\left(1-\mathcal{V}^-(\bar{n}/n)+\epsilon(d)\right)\|\mathbf{A}\|_2^2+c\phi(1+\zeta d\|\mathbf{A}\|_2)\right]\|\mathbf{X}\|_1\right\},$$

where $\epsilon(d)=c_1\theta\sqrt{\frac{d}{n}}+c_2\theta^{1/2}d^{1/4}(\log n)^{3/4}n^{-1/4}$, $\tau=\max\{1,\frac{p}{n}\}$, $\phi=\max\{1,\sqrt{\frac{\log p}{n}}\}$ and c,c_1,c_2 are universal constants.

Proof. Let $\{k(i)\}$ be the indices of the largest $n-\bar{n}$ values of $\langle\mathbf{z}\mathbf{z}^\top,\mathbf{X}\rangle$, then

$$\frac{1}{n}\sum_{i=1}^{n-\bar{n}}\langle\mathbf{z}\mathbf{z}^\top,\mathbf{X}\rangle_{[i]}=\langle\frac{1}{n}\sum_{i=1}^{n-\bar{n}}\mathbf{z}_{k(i)}\mathbf{z}_{k(i)}^\top,\mathbf{X}\rangle$$
$$\le\min\left\{\|\frac{1}{n}\sum_{i=1}^{n-\bar{n}}\mathbf{z}_{k(i)}\mathbf{z}_{k(i)}^\top\|_\infty\|\mathbf{X}\|_1,\|\frac{1}{n}\sum_{i=1}^{n-\bar{n}}\mathbf{z}_{k(i)}\mathbf{z}_{k(i)}^\top\|_2\|\mathbf{X}\|_*\right\}.$$

Notice that $\|\frac{1}{n}\sum_{i=1}^{n-\bar{n}}\mathbf{z}_{k(i)}\mathbf{z}_{k(i)}^\top\|_2$ can be bounded by Lemma 10, so we only need to bound $\|\frac{1}{n}\sum_{i=1}^{n-\bar{n}}\mathbf{z}_{k(i)}\mathbf{z}_{k(i)}^\top\|_\infty$. We have

$$\|\frac{1}{n}\sum_{i=1}^{n-\bar{n}}\mathbf{z}_{k(i)}\mathbf{z}_{k(i)}^\top\|_\infty$$
$$\le\|\frac{1}{n}\sum_{i=1}^{n-\bar{n}}\mathbf{A}\mathbf{x}_{k(i)}\mathbf{x}_{k(i)}^\top\mathbf{A}^\top\|_\infty+\|\frac{1}{n}\sum_{i=1}^{n-\bar{n}}\mathbf{n}_{k(i)}\mathbf{n}_{k(i)}^\top\|_\infty+\|\frac{2}{n}\sum_{i=1}^{n-\bar{n}}\mathbf{A}\mathbf{x}_{k(i)}\mathbf{n}_{k(i)}^\top\|_\infty$$
$$\le\|\frac{1}{n}\sum_{i=1}^{n-\bar{n}}\mathbf{A}\mathbf{x}_{k(i)}\mathbf{x}_{k(i)}^\top\mathbf{A}^\top\|_2+\|\frac{1}{n}\sum_{i=1}^{n}|\mathbf{n}_i||\mathbf{n}_i|^\top\|_\infty+2d\|\mathbf{A}\|_\infty\|\frac{1}{n}\sum_{i=1}^{n}|\mathbf{x}_i||\mathbf{n}_i|^\top\|_\infty.$$

Let $\mathbf{P} \triangleq \frac{1}{n}\sum_{i=1}^{n} |\mathbf{n}_i||\mathbf{n}_i|^\top$ and $\mathbf{Q} \triangleq \frac{1}{n}\sum_{i=1}^{n} |\mathbf{x}_i||\mathbf{n}_i|^\top$, then

$$\mathbf{P}_{ij} = \frac{1}{n}\sum_{k=1}^{n} |\mathbf{n}_{ki}\mathbf{n}_{kj}| \leq \frac{2}{n}\sum_{k=1}^{n}(\mathbf{n}_{ki}^2 + \mathbf{n}_{kj}^2),\ \mathbf{Q}_{ij} = \frac{1}{n}\sum_{k=1}^{n} |\mathbf{x}_{ki}\mathbf{n}_{kj}| \leq \frac{2}{n}\sum_{k=1}^{n}(\mathbf{x}_{ki}^2 + \mathbf{n}_{kj}^2).$$

Since $\mathbf{x}_{ki}$ is a zero-mean sub-Gaussian random variable and $\mathbf{n}_{ki}$ is a standard Gaussian random variable, from Proposition 5.16 (Vershynin, 2012)

$$\mathbb{P}\left[\left|\frac{1}{n}\sum_{k=1}^{n}\mathbf{n}_{ki}^2 - 1\right| > t\right] \leq 2\exp(-c_1\min\{nt^2, nt\}),\ \text{and}$$
$$\mathbb{P}\left[\left|\frac{1}{n}\sum_{k=1}^{n}\mathbf{x}_{ki}^2 - 1\right| > t\right] \leq 2\exp(-c_2\min\{\frac{nt^2}{\theta^2}, \frac{nt}{\theta}\})$$

for universal constant c_1, c_2. There exists c (may change from line to line) so that when $t = c\sqrt{\frac{\log p}{n}}$, by the union bound we have

$$\mathbb{P}\left[\|\mathbf{P}\|_\infty > 1 + c\sqrt{\frac{\log p}{n}}\right] \leq p^{-10},\ \text{and}\ \mathbb{P}\left[\|\mathbf{Q}\|_\infty > \zeta(1 + c\sqrt{\frac{\log p}{n}})\right] \leq p^{-10}$$

where $\zeta = \max\{\theta, 2\}$. Let $\phi = \max\{1, \frac{\log p}{n}\}$, then with high probability

$$\|\mathbf{P}\|_\infty \leq c\phi,\ \text{and}\ \|\mathbf{Q}\|_\infty \leq c\zeta\phi.$$

Thus,

$$\|\frac{1}{n}\sum_{i=1}^{n-\bar{n}}\mathbf{z}_{k(i)}\mathbf{z}_{k(i)}^\top\|_\infty \leq \|\frac{1}{n}\sum_{i=1}^{n-\bar{n}}\mathbf{A}\mathbf{x}_{k(i)}\mathbf{x}_{k(i)}^\top\mathbf{A}^\top\|_2 + c\phi(1 + \zeta d\|\mathbf{A}\|_2).$$

The first term on the right hand side can be bound by Equation (3). Hence we obtain this lemma. □

Lemma 12. *(Corollary 3.3, (Vu et al., 2013)) There exists a universal constant c such that with probability at least* $1-p^{-10}$,

$$\|\frac{1}{n}\sum_{i=1}^{n}\mathbf{z}_i\mathbf{z}_i^\top - \mathbf{\Sigma}\|_\infty \leq c\zeta\|\mathbf{\Sigma}\|_2\sqrt{\frac{\log p}{n}}.$$

3. Proofs of Section 1.1

Lemma 13. *(Lemma 3.1, (Vu et al., 2013)) Let* $\mathbf{\Sigma}$ *be a symmetric matrix and* $\mathbf{\Pi}$ *be the projection onto the subspace spanned by the eigenvectors of* $\mathbf{\Sigma}$ *corresponding to its* k *largest eigenvalues* $\lambda_1 \geq \lambda_2 \geq \cdots$. *If* $\delta = \lambda_k(\mathbf{\Sigma}) - \lambda_{k+1}(\mathbf{\Sigma}) > 0$, *then*

$$\frac{\delta}{2}\|\mathbf{\Pi} - \mathbf{X}\|_F^2 \leq \langle\mathbf{\Sigma}, \mathbf{\Pi} - \mathbf{X}\rangle$$

for all $\mathbf{X}$ *satisfying* $0 \preceq \mathbf{X} \preceq \mathbf{I}$ *and* $\mathrm{tr}(\mathbf{X}) = k$.

3.1. Proof of Lemma 1

Proof. If $\mathcal{E}(s)$ is false, then

$$\sum_{i\in\mathcal{Z}}\alpha_i(s)\langle\mathbf{y}_i\mathbf{y}_i^\top, \mathbf{X}_s\rangle < \frac{1}{\kappa}\sum_{i\in\mathcal{O}}\alpha_i(s)\langle\mathbf{y}_i\mathbf{y}_i^\top, \mathbf{X}_s\rangle.$$

Let $\Delta\alpha_i \triangleq \frac{\langle\mathbf{y}_i\mathbf{y}_i^\top, \hat{\mathbf{X}}\rangle}{\max_{\{i|\alpha_i\neq 0\}}\langle\mathbf{y}_i\mathbf{y}_i^\top, \hat{\mathbf{X}}\rangle}\alpha_i$, if $\bigcup_{s=1}^{s_0}\mathcal{E}(s)$ is false, we have

$$\sum_{s=1}^{s_0}\sum_{i\in\mathcal{Z}}\Delta\alpha_i(s) < \frac{1}{\kappa}\sum_{s=1}^{s_0}\sum_{i\in\mathcal{O}}\Delta\alpha_i(s).$$

From the algorithm, at least one α is eliminated in each iteration. Thus, we have $\sum_{s=1}^{s_0}\sum_{i=1}^{n}\Delta\alpha_i(s)\geq s_0$, which implies that

$$\sum_{s=1}^{s_0}\sum_{i\in\mathcal{Z}}\Delta\alpha_i(s)+\sum_{s=1}^{s_0}\sum_{i\in\mathcal{O}}\Delta\alpha_i(s)\geq s_0.$$

Hence

$$\frac{1}{\kappa}\sum_{s=1}^{s_0}\sum_{i\in\mathcal{O}}\Delta\alpha_i(s)+\sum_{s=1}^{s_0}\sum_{i\in\mathcal{O}}\Delta\alpha_i(s)\geq s_0.$$

Note that $\rho n\geq\sum_{s=1}^{s_0}\sum_{i\in\mathcal{O}}\Delta\alpha_i(s)$, then $\rho n\geq\frac{\kappa s_0}{1+\kappa}$, so $s_0\leq\frac{\rho n(1+\kappa)}{\kappa}$. □

3.2. Proof of Theorem 1

Proof. Note that for any $1\leq\bar{s}\leq s$, the event $\mathcal{E}(\bar{s})$ is false, which implies that

$$\sum_{i\in\mathcal{Z}}\alpha_i(\bar{s})\langle\mathbf{y}_i\mathbf{y}_i^\top,\mathbf{X}_{\bar{s}}\rangle<\frac{1}{\kappa}\sum_{i\in\mathcal{O}}\alpha_i(\bar{s})\langle\mathbf{y}_i\mathbf{y}_i^\top,\mathbf{X}_{\bar{s}}\rangle,$$

Thus, we have

$$\sum_{i\in\mathcal{Z}}\Delta\alpha_i(\bar{s})<\frac{1}{\kappa}\sum_{i\in\mathcal{O}}\Delta\alpha_i(\bar{s}).$$

Since $\alpha_i(s)=1-\sum_{k=1}^{s-1}\Delta\alpha_i(k)$,

$$\sum_{i\in\mathcal{Z}}\alpha_i(s)=t-\sum_{i\in\mathcal{Z}}\sum_{k=1}^{s-1}\Delta\alpha_i(k)>t-\frac{1}{\kappa}\sum_{k=1}^{s-1}\sum_{i\in\mathcal{O}}\Delta\alpha_i(k)\geq t-\frac{\rho n}{\kappa}.$$

Hence for any $\mathbf{X}\in\mathcal{F}(k)$, we have

$$\begin{aligned}
&\sum_{i\in\mathcal{Z}}\alpha_i(s)\langle\mathbf{y}_i\mathbf{y}_i^\top,\mathbf{X}\rangle-\sum_{i=1}^{t-\rho n/\kappa}\langle\mathbf{z}\mathbf{z}^\top,\mathbf{X}\rangle_{(i)}\\
=&\sum_{i=1}^{t-\rho n/\kappa}(\alpha_{j(i)}(s)-1)\langle\mathbf{z}_{j(i)}\mathbf{z}_{j(i)}^\top,\mathbf{X}\rangle+\sum_{i=t-\rho n/\kappa+1}^{t}\alpha_{j(i)}(s)\langle\mathbf{z}_{j(i)}\mathbf{z}_{j(i)}^\top,\mathbf{X}\rangle\\
\geq&\sum_{i=1}^{t-\rho n/\kappa}(\alpha_{j(i)}(s)-1)\langle\mathbf{z}\mathbf{z}^\top,\mathbf{X}\rangle_{(t-\rho n/\kappa)}+\sum_{i=t-\rho n/\kappa+1}^{t}\alpha_{j(i)}(s)\langle\mathbf{z}\mathbf{z}^\top,\mathbf{X}\rangle_{(t-\rho n/\kappa)}\\
=&\left(\sum_{i\in\mathcal{Z}}\alpha_i-(t-\frac{\rho n}{\kappa})\right)\langle\mathbf{z}\mathbf{z}^\top,\mathbf{X}\rangle_{(t-\rho n/\kappa)}\geq 0.
\end{aligned}\tag{4}$$

Since $\mathbf{X}_s$ is the optimal solution of the PCA-like algorithm and event $\mathcal{E}(s)$ is true, we have

$$\begin{aligned}
&\frac{1}{n}\langle\sum_{i=1}^{n}\alpha_i(s)\mathbf{y}_i\mathbf{y}_i^\top,\mathbf{X}_s\rangle-\mu\|\mathbf{X}_s\|_1\\
\geq&(1-\eta)\left[\frac{1}{n}\langle\sum_{i=1}^{n}\alpha_i(s)\mathbf{y}_i\mathbf{y}_i^\top,\mathbf{\Pi}\rangle-\mu\|\mathbf{\Pi}\|_1\right]\\
\geq&(1-\eta)\left[\frac{1}{n}\langle\sum_{i\in\mathcal{Z}}\alpha_i(s)\mathbf{y}_i\mathbf{y}_i^\top,\mathbf{\Pi}\rangle-\mu\|\mathbf{\Pi}\|_1\right]\\
\geq&\frac{1}{n}\langle\sum_{i\in\mathcal{Z}}\alpha_i(s)\mathbf{y}_i\mathbf{y}_i^\top,\mathbf{\Pi}\rangle-\mu\|\mathbf{\Pi}\|_1-\eta\left(\mu+\|\frac{1}{n}\sum_{i\in\mathcal{Z}}\alpha_i(s)\mathbf{y}_i\mathbf{y}_i^\top\|_\infty\right)\|\mathbf{\Pi}\|_1
\end{aligned}$$

and

$$\begin{aligned}&\frac{1}{n}\langle\sum_{i=1}^{n}\alpha_i(s)\mathbf{y}_i\mathbf{y}_i^\top,\mathbf{X}_s\rangle-\mu\|\mathbf{X}_s\|_1\\=&\frac{1}{n}\langle\sum_{i\in\mathcal{Z}}\alpha_i(s)\mathbf{y}_i\mathbf{y}_i^\top,\mathbf{X}_s\rangle+\frac{1}{n}\langle\sum_{i\in\mathcal{O}}\alpha_i(s)\mathbf{y}_i\mathbf{y}_i^\top,\mathbf{X}_s\rangle-\mu\|\mathbf{X}_s\|_1\\\le&\frac{1+\kappa}{n}\langle\sum_{i\in\mathcal{Z}}\alpha_i(s)\mathbf{y}_i\mathbf{y}_i^\top,\mathbf{X}_s\rangle-\mu\|\mathbf{X}_s\|_1.\end{aligned}$$

Denote $\left(\mu+\|\frac{1}{n}\sum_{i\in\mathcal{Z}}\alpha_i(s)\mathbf{y}_i\mathbf{y}_i^\top\|_\infty\right)\|\mathbf{\Pi}\|_1$ by B, we have

$$\frac{1}{n}\langle\sum_{i\in\mathcal{Z}}\alpha_i(s)\mathbf{y}_i\mathbf{y}_i^\top,\mathbf{X}_s-\mathbf{\Pi}\rangle-\mu\|\mathbf{X}_s\|_1+\mu\|\mathbf{\Pi}\|_1+\frac{\kappa}{n}\langle\sum_{i\in\mathcal{Z}}\alpha_i(s)\mathbf{y}_i\mathbf{y}_i^\top,\mathbf{X}_s\rangle+\eta B\ge0.$$

Since Inequality 4 holds and $0\le\alpha_i\le1$,

$$\frac{1}{n}\langle\sum_{i\in\mathcal{Z}}\mathbf{y}_i\mathbf{y}_i^\top,\mathbf{X}_s\rangle-\frac{1}{n}\sum_{i=1}^{t-\rho n/\kappa}\langle\mathbf{z}\mathbf{z}^\top,\mathbf{\Pi}\rangle_{(i)}-\mu\|\mathbf{X}_s\|_1+\mu\|\mathbf{\Pi}\|_1+\frac{\kappa}{n}\langle\sum_{i\in\mathcal{Z}}\mathbf{y}_i\mathbf{y}_i^\top,\mathbf{X}_s\rangle+\eta B\ge0$$

or equivalently,

$$\frac{1}{n}\langle\sum_{i=1}^{t}\mathbf{z}_i\mathbf{z}_i^\top,\mathbf{X}_s-\mathbf{\Pi}\rangle-\mu\|\mathbf{X}_s\|_1+\mu\|\mathbf{\Pi}\|_1+\frac{\kappa}{n}\langle\sum_{i=1}^{t}\mathbf{z}_i\mathbf{z}_i^\top,\mathbf{X}_s\rangle+\frac{1}{n}\sum_{i=1}^{\rho n/\kappa}\langle\mathbf{z}\mathbf{z}^\top,\mathbf{\Pi}\rangle_{[i]}+\eta B\ge0$$

Let $\mathbf{\Delta}=\mathbf{X}_s-\mathbf{\Pi}$ and $\mathbf{W}=\frac{1}{t}\sum_{i=1}^{t}\mathbf{z}_i\mathbf{z}_i^\top-\mathbf{\Sigma}$, then

$$\langle\mathbf{W}+\mathbf{\Sigma},\mathbf{\Delta}\rangle-\frac{n\mu}{t}\|\mathbf{\Pi}+\mathbf{\Delta}\|_1+\frac{n\mu}{t}\|\mathbf{\Pi}\|_1+\kappa\langle\mathbf{W}+\mathbf{\Sigma},\mathbf{X}_s\rangle+\frac{1}{t}\sum_{i=1}^{\rho n/\kappa}\langle\mathbf{z}\mathbf{z}^\top,\mathbf{\Pi}\rangle_{[i]}+\frac{n\eta B}{t}\ge0.$$

Since $-\langle\mathbf{\Sigma},\mathbf{\Delta}\rangle\ge\frac{\delta}{2}\|\mathbf{\Delta}\|_F^2$ where $\delta=\lambda_k(\mathbf{\Sigma})-\lambda_{k+1}(\mathbf{\Sigma})$ (Lemma 13), we have

$$\langle\mathbf{W},\mathbf{\Delta}\rangle-\frac{n\mu}{t}\|\mathbf{\Pi}+\mathbf{\Delta}\|_1+\frac{n\mu}{t}\|\mathbf{\Pi}\|_1+\kappa\langle\mathbf{W}+\mathbf{\Sigma},\mathbf{X}_s\rangle+\frac{1}{t}\sum_{i=1}^{\rho n/\kappa}\langle\mathbf{z}\mathbf{z}^\top,\mathbf{\Pi}\rangle_{[i]}+\frac{n\eta B}{t}\ge\frac{\delta}{2}\|\mathbf{\Delta}\|_F^2. \quad (5)$$

For simplicity, we let

$$T=\kappa\langle\mathbf{W}+\mathbf{\Sigma},\mathbf{X}_s\rangle+\frac{1}{t}\sum_{i=1}^{\rho n/\kappa}\langle\mathbf{z}\mathbf{z}^\top,\mathbf{\Pi}\rangle_{[i]}+\frac{n\eta B}{t}.$$

We first consider the case that $\mu\neq0$. Since $\langle\mathbf{W},\mathbf{\Delta}\rangle\le\|\mathbf{W}\|_\infty\|\mathbf{\Delta}\|_1$ and $n\ge t\ge0.5n$,

$$[\|\mathbf{W}\|_\infty-\mu]_+\|\mathbf{\Delta}\|_1+\mu\|\mathbf{\Delta}\|_1-\mu\|\mathbf{\Pi}+\mathbf{\Delta}\|_1+\mu\|\mathbf{\Pi}\|_1+T\ge\frac{\delta}{4}\|\mathbf{\Delta}\|_F^2.$$

Let N be the subset of indices of the nonzero entries of $\mathbf{\Pi}$, since $\|\mathbf{\Pi}\|_0\le\beta^2$ and $\|\mathbf{\Delta}_N\|_1\le\beta\|\mathbf{\Delta}_N\|_F\le\beta\|\mathbf{\Delta}\|_F$,

$$\|\mathbf{\Delta}\|_1-\|\mathbf{\Pi}+\mathbf{\Delta}\|_1+\|\mathbf{\Pi}\|_1=\|\mathbf{\Delta}_N\|_1-\|\mathbf{\Pi}_N+\mathbf{\Delta}_N\|_1+\|\mathbf{\Pi}_N\|_1\le2\|\mathbf{\Delta}_N\|_1.$$

Also note that $\mathbf{\Delta}$ has at most $\gamma^2+\beta^2$ non-zero entries, so $\|\mathbf{\Delta}\|_1\le\sqrt{\gamma^2+\beta^2}\|\mathbf{\Delta}\|_F\le2\gamma\|\mathbf{\Delta}\|_F$. Thus,

$$2(\gamma[\|\mathbf{W}\|_\infty-\mu]_++\mu\beta)\|\mathbf{\Delta}\|_F+T\ge\frac{\delta}{4}\|\mathbf{\Delta}\|_F^2,$$

which implies that

$$\|\mathbf{\Delta}\|_F\le\frac{8(\gamma[\|\mathbf{W}\|_\infty-\mu]_++\mu\beta)}{\delta}+2\sqrt{\frac{T}{\delta}}\le\frac{8(\gamma[c_0\zeta\sqrt{\frac{\log p}{n}}\|\mathbf{\Sigma}\|_2-\mu]_++\mu\beta)}{\delta}+2\sqrt{\frac{T}{\delta}},$$

where the last inequality follows from Lemma 12.

We now consider the case that $\mu = 0$, then (5) becomes $\langle \mathbf{W}, \boldsymbol{\Delta} \rangle + T \geq \frac{\delta}{2}\|\boldsymbol{\Delta}\|_F^2$. Since $\langle \mathbf{W}, \boldsymbol{\Delta} \rangle \leq \min\{\|\mathbf{W}\|_\infty \|\boldsymbol{\Delta}\|_1, \|\mathbf{W}\|_2 \|\boldsymbol{\Delta}\|_*\} \leq 2\min\{\gamma\|\mathbf{W}\|_\infty\|\boldsymbol{\Delta}\|_F, k\|\mathbf{W}\|_2\}$, $\|\boldsymbol{\Delta}\|_F$ should satisfy that

$$2\min\{\gamma\|\mathbf{W}\|_\infty\|\boldsymbol{\Delta}\|_F, k\|\mathbf{W}\|_2\} + T \geq \frac{\delta}{2}\|\boldsymbol{\Delta}\|_F^2.$$

By simple calculation, we have

$$\|\boldsymbol{\Delta}\|_F \leq \min\{\frac{8\gamma\|\mathbf{W}\|_\infty}{\delta}, 2\sqrt{\frac{k\|\mathbf{W}\|_2}{\delta}}\} + 2\sqrt{\frac{T}{\delta}} \leq \min\{\frac{8c_0\gamma\zeta\sqrt{\frac{\log p}{n}}\|\boldsymbol{\Sigma}\|_2}{\delta}, 2\sqrt{\frac{c_1 k\zeta\sqrt{\frac{p}{n}}\|\boldsymbol{\Sigma}\|_2}{\delta}}\} + 2\sqrt{\frac{T}{\delta}},$$

where the last inequality follows from Lemma 8 and Lemma 12. Hence we have $\|\boldsymbol{\Delta}\|_F \leq R(\mu) + 2\sqrt{\frac{T}{\delta}}$ where

$$R(\mu) = \begin{cases} \frac{8(\gamma[c_0\sqrt{\frac{\log p}{n}}(\|\mathbf{A}\|_2^2+1)-\mu]_+ + \mu\beta)}{\delta}, & \mu \neq 0 \\ \min\left\{\frac{8c_0\gamma\sqrt{\frac{\log p}{n}}(\|\mathbf{A}\|_2^2+1)}{\delta}, 2\sqrt{\frac{c_1 k\sqrt{\frac{p}{n}}(\|\mathbf{A}\|_2^2+1)}{\delta}}\right\}, & \mu = 0. \end{cases}$$

We ignore ζ in $R(\mu)$ because it's a constant.

We now bound T. Notice that $\|\boldsymbol{\Pi}\|_* \leq k$ and $\|\boldsymbol{\Pi}\|_1 \leq \beta\|\boldsymbol{\Pi}\|_F \leq \beta k$, from Lemma 11, the following inequality holds with high probability,

$$\frac{1}{n}\sum_{i=1}^{n-\bar{n}}\langle \mathbf{z}\mathbf{z}^\top, \boldsymbol{\Pi}\rangle_{[i]} \leq k\min\left\{2\left(1-\mathcal{V}^-(\bar{n}/n)+\epsilon(d)\right)\|\mathbf{A}\|_2^2 + c\tau,\right.$$
$$\left.\beta\left[\left(1-\mathcal{V}^-(\bar{n}/n)+\epsilon(d)\right)\|\mathbf{A}\|_2^2 + c\phi(1+\zeta d\|\mathbf{A}\|_2)\right]\right\}.$$

Since $t \geq 0.5n$, there exist constants c_1, c_2 such that $\|\mathbf{W}+\boldsymbol{\Sigma}\|_2 \leq 2(1+c_1\theta\sqrt{\frac{d}{n}})\|\mathbf{A}\|_2^2 + c\tau$ and $\|\mathbf{W}\|_\infty \leq c_0\zeta\sqrt{\frac{\log p}{n}}(\|\mathbf{A}\|_2^2+1)$ hold with high probability (Lemma 9 and Lemma 12). Hence

$$\begin{aligned}\langle \mathbf{W}+\boldsymbol{\Sigma}, \mathbf{X}_s\rangle &\leq \min\{\|\mathbf{W}+\boldsymbol{\Sigma}\|_2\|\mathbf{X}_s\|_*, \|\mathbf{W}+\boldsymbol{\Sigma}\|_\infty\|\mathbf{X}_s\|_1\} \\ &\leq k\min\left\{2\left(1+c_1\theta\sqrt{\frac{d}{n}}\right)\|\mathbf{A}\|_2^2 + c\tau, \gamma\left(1+c_0\zeta\sqrt{\frac{\log p}{n}}\right)\left(\|\mathbf{A}\|_2^2+1\right)\right\}\end{aligned}$$

where the last inequality follows from $\|\mathbf{X}_s\|_* \leq k$ and $\|\mathbf{X}_s\|_1 \leq \gamma\|\mathbf{X}_s\|_F \leq \gamma k$. Also notice that $0 \leq \alpha_i(s) \leq 1$ and $\|\boldsymbol{\Pi}\|_1 \leq \beta\|\boldsymbol{\Pi}\|_F = \beta k$, we have

$$\frac{n\eta B}{t} \leq \eta\beta k\left(2\mu + \|\frac{1}{t}\sum_{i\in\mathcal{Z}}\mathbf{y}_i\mathbf{y}_i^\top\|_\infty\right) \leq \eta\beta k\left(2\mu + \left(c_0\zeta\sqrt{\frac{\log p}{n}}+1\right)\|\boldsymbol{\Sigma}\|_2\right).$$

Let $B_0' \triangleq 2\mu + \left(c_0\zeta\sqrt{\frac{\log p}{n}}+1\right)(\|\mathbf{A}\|_2^2+1)$, since μ is less than some universal constant and $\log p \leq n$, there exists constant c_2 such that $B_0 \triangleq c_2(\|\mathbf{A}\|_2^2+1) \geq B_0'$. Let $\epsilon_0 \triangleq c_0\zeta\sqrt{\frac{\log p}{n}}$, $\epsilon_1 \triangleq \epsilon_0 + \epsilon(d) + c_1\theta\sqrt{\frac{d}{n}}$. Since $d < n$, $\epsilon_1 \leq \epsilon_0 + c_1\left(\frac{d\log^3 n}{n}\right)^{\frac{1}{4}}$. Since ζ is a constant, we have $\kappa \leq 1$, $\beta \leq \gamma$ and $\log p \leq n$,

$$T = k\min\left\{2B_1\|\mathbf{A}\|_2^2 + c\tau, \gamma B_1\|\mathbf{A}\|_2^2 + c\gamma(d\|\mathbf{A}\|_2+1)\right\} + \eta\beta k B_0,$$

where $B_1 = \kappa + 1 - \mathcal{V}^-(1-\frac{\rho}{\kappa(1-\rho)}) + \epsilon_1$. By minimizing T over κ, we can obtain this theorem. □

3.3. Proof of Theorem 2

Proof. Under the conditions of Theorem 2, the conditions of Theorem 1 are satisfied, let $\mathbf{\Delta} = \mathbf{X}_s - \mathbf{\Pi}$, $\mathbf{W} = \frac{1}{t}\sum_{i=1}^{t}\mathbf{z}_i\mathbf{z}_i^\top - \mathbf{\Sigma}$, and $f(B) = \min\left\{2B\|\mathbf{A}\|_2^2 + c\tau, \gamma B\|\mathbf{A}\|_2^2 + c\gamma(d\|\mathbf{A}\|_2 + 1)\right\}$ then w.h.p

$$\|\mathbf{\Delta}\|_F \leq R(\mu) + \sqrt{k}\min_{1\geq\kappa>2\rho}\sqrt{\frac{f(B_1) + \eta\beta B_0}{\delta}} \triangleq B_2.$$

From the Algorithm, we know that

$$\frac{1}{\hat{t}}\sum_{i=1}^{\hat{t}}\langle\mathbf{z}\mathbf{z}^\top, \mathbf{X}^*\rangle_{(i)} \geq \frac{1}{\hat{t}}\sum_{i=1}^{\hat{t}}\langle\mathbf{y}\mathbf{y}^\top, \mathbf{X}^*\rangle_{(i)} \geq \frac{1}{\hat{t}}\sum_{i=1}^{\hat{t}}\langle\mathbf{y}\mathbf{y}^\top, \mathbf{X}_s\rangle_{(i)} \geq \frac{1}{\hat{t}}\sum_{i=1}^{\hat{t}-\rho n}\langle\mathbf{z}\mathbf{z}^\top, \mathbf{X}_s\rangle_{(i)}.$$

Hence we have

$$\begin{aligned}
&\frac{1}{t}\sum_{i=1}^{t}\langle\mathbf{z}_i\mathbf{z}_i^\top, \mathbf{X}^*\rangle - \frac{1}{t}\sum_{i=1}^{t-\hat{t}}\langle\mathbf{z}\mathbf{z}^\top, \mathbf{X}^*\rangle_{[i]} + \frac{1}{t}\sum_{i=1}^{t-\hat{t}+\rho n}\langle\mathbf{z}\mathbf{z}^\top, \mathbf{X}_s\rangle_{[i]} \geq \frac{1}{t}\sum_{i=1}^{t}\langle\mathbf{z}_i\mathbf{z}_i^\top, \mathbf{X}_s\rangle \\
\Rightarrow&\langle\mathbf{W} + \mathbf{\Sigma}, \mathbf{X}^*\rangle + \frac{1}{t}\sum_{i=1}^{t-\hat{t}}\langle\mathbf{z}\mathbf{z}^\top, \mathbf{X}^*\rangle_{[i]} + \frac{1}{t}\sum_{i=1}^{t-\hat{t}+\rho n}\langle\mathbf{z}\mathbf{z}^\top, \mathbf{X}_s\rangle_{[i]} \geq \langle\mathbf{W} + \mathbf{\Sigma}, \mathbf{X}_s\rangle.
\end{aligned}$$

let $T \triangleq \frac{1}{t}\sum_{i=1}^{t-\hat{t}}\langle\mathbf{z}\mathbf{z}^\top, \mathbf{X}^*\rangle_{[i]} + \frac{1}{t}\sum_{i=1}^{t-\hat{t}+\rho n}\langle\mathbf{z}\mathbf{z}^\top, \mathbf{X}_s\rangle_{[i]}$ and $\mathbf{\Delta}^* \triangleq \mathbf{X}^* - \mathbf{\Pi}$. Note that $\mathbf{\Sigma} = \mathbf{A}\mathbf{A}^\top + \mathbf{I}_p$, from Lemma 13, we have

$$\langle\mathbf{W}, \mathbf{\Delta}^* - \mathbf{\Delta}\rangle + \|\mathbf{A}\mathbf{A}^\top\|_F\|\mathbf{\Delta}\|_F + \|\mathbf{\Delta}\|_* + T \geq \frac{\delta}{2}\|\mathbf{\Delta}^*\|_F^2,$$

where $\delta = \lambda_k(\mathbf{\Sigma}) - \lambda_{k+1}(\mathbf{\Sigma})$. Since $\|\mathbf{\Delta}^*\|_* \leq \|\mathbf{X}^*\|_* + \|\mathbf{\Pi}\|_* \leq 2k$, $\|\mathbf{A}\mathbf{A}^\top\|_F = \sqrt{\sum_{i=1}^{d}\lambda_i(\mathbf{A}\mathbf{A}^\top)^2} \leq d\|\mathbf{A}\|_2^2$ and $\|\mathbf{\Delta}\|_F \leq B_2$, we have

$$\|\mathbf{\Delta}^*\|_F \leq \sqrt{\frac{2}{\delta}(\langle\mathbf{W}, \mathbf{\Delta}^* - \mathbf{\Delta}\rangle + dB_2\|\mathbf{A}\|_2^2 + T + 2k)}.$$

We first bound the term $\langle\mathbf{W}, \mathbf{\Delta}^* - \mathbf{\Delta}\rangle$. Notice that

$$\langle\mathbf{W}, \mathbf{\Delta}^* - \mathbf{\Delta}\rangle \leq \min\{\|\mathbf{W}\|_2(\|\mathbf{\Delta}^*\|_* + \|\mathbf{\Delta}\|_*), \|\mathbf{W}\|_\infty(\|\mathbf{\Delta}^*\|_1 + \|\mathbf{\Delta}\|_1)\}.$$

Since $\|\mathbf{\Delta}^*\|_* \leq 2k$ and $\|\mathbf{\Delta}^*\|_1 \leq \|\mathbf{X}^*\|_1 + \|\mathbf{\Pi}\|_1 \leq \gamma\|\mathbf{X}^*\|_F + \beta\|\mathbf{\Pi}\|_F \leq k(\gamma + \beta) \leq 2k\gamma$ ($\mathbf{\Delta}$ has similar inequalities), we have

$$\langle\mathbf{W}, \mathbf{\Delta}^* - \mathbf{\Delta}\rangle \leq 4k\min\{\|\mathbf{W}\|_2, \gamma\|\mathbf{W}\|_\infty\}.$$

From Lemma 8, there exists constant c_2 such that $\|\mathbf{W}\|_2 \leq c_1\zeta\sqrt{\frac{p}{n}}\|\mathbf{\Sigma}\|_2 = c_1\zeta\sqrt{\frac{p}{n}}(\|\mathbf{A}\|_2^2 + 1)$ holds with high probability, where $\zeta = \max\{\theta, 2\}$. From Lemma 12, $\|\mathbf{W}\|_\infty \leq c_2\zeta\|\mathbf{\Sigma}\|_2\sqrt{\frac{\log p}{n}} = c_2\zeta\sqrt{\frac{\log p}{n}}(\|\mathbf{A}\|_2^2 + 1)$ holds for constant c_2. Let $B_4 \triangleq 4\zeta\min\{c_1\sqrt{\frac{p}{n}}, c_2\gamma\sqrt{\frac{\log p}{n}}\}$, then

$$\|\mathbf{\Delta}^*\|_F \leq \sqrt{\frac{2}{\delta}[(dB_2 + kB_4)\|\mathbf{A}\|_2^2 + T + 2k + kB_4]}.$$

For term T, we follow the same proof of Theorem 1. The following inequality holds w.h.p,

$$T \leq k\min\left\{2B_3\|\mathbf{A}\|_2^2 + c\tau, \gamma B_3\|\mathbf{A}\|_2^2 + c\gamma(d\|\mathbf{A}\|_2 + 1)\right\},$$

where $B_3 = 2 - \mathcal{V}^-(\frac{\hat{t}}{t}) - \mathcal{V}^-(\frac{\hat{t}-\rho n}{t}) + \epsilon(d)$. Hence we have

$$\|\mathbf{\Delta}^*\|_F \leq \sqrt{\frac{2}{\delta}\left[(dB_2 + kB_4)\|\mathbf{A}\|_2^2 + k\min\left\{2B_3 + c\tau, \gamma B_3 + c\gamma(d\|\mathbf{A}\|_2 + 1)\right\}\right]},$$

which establishes this theorem. ∎

4. Proofs in Section 1.2

Let $H^* \triangleq \langle \mathbf{A}\mathbf{A}^\top, \mathbf{X}^* \rangle$, $H_s \triangleq \langle \mathbf{A}\mathbf{A}^\top, \mathbf{X}_s \rangle$ and $\overline{H} \triangleq \langle \mathbf{A}\mathbf{A}^\top, \mathbf{\Pi} \rangle$. In order to bound E.V, we first bound $|H^* - \sum_{i=1}^k {\mathbf{w}_i^*}^\top \mathbf{A}\mathbf{A}^\top \mathbf{w}_i^*|$, and then bound $H^*/\overline{H}$. This involves the following steps:

1. Bound $|H^* - \sum_{i=1}^k {\mathbf{w}_i^*}^\top \mathbf{A}\mathbf{A}^\top \mathbf{w}_i^*|$.
2. Bound the robust variance estimator of the the authentic samples by applying the concentration inequalities (Theorem 5, Theorem 6 and Theorem 7, i.e. bounding $\frac{1}{t}\sum_{i=1}^{\bar{t}} |\mathbf{w}^\top \mathbf{z}|_{(i)}^2$.
3. Show that with high probability, the algorithm finds a "good" solution within a bounded number of steps and then show that the "good" solution in previous step is close to the optimal solution and the final solution of our algorithm is close to this "good" solution.

4.1. Step 1

Lemma 14. *For any $\mathbf{X} \in \mathbb{R}^{p\times p}$ such that $0 \preceq \mathbf{X} \preceq \mathbf{I}_p$ and $\mathrm{tr}(\mathbf{X}) = k$, let $\mathbf{w}_1, \cdots, \mathbf{w}_k$ be the top k eigenvectors of $\mathbf{X}$, then*

$$\left| \langle \mathbf{A}\mathbf{A}^\top, \mathbf{X} \rangle - \sum_{i=1}^k {\mathbf{w}_i}^\top \mathbf{A}\mathbf{A}^\top \mathbf{w}_i \right| \le \max\{1 - \lambda_k(\mathbf{X}), \lambda_{k+1}(\mathbf{X})\} \cdot \mathrm{tr}(\mathbf{A}\mathbf{A}^\top),$$

where λ_k is the k^{th} largest eigenvalue of $\mathbf{X}$.

From this lemma, we have E.V$\{\mathbf{w}_1^*, \cdots, \mathbf{w}_k^*\} \ge \frac{H^*}{\overline{H}} - \max\{1 - \lambda_k(\mathbf{X}), \lambda_{k+1}(\mathbf{X})\}$.

4.2. Step 2

From Theorem 5, Theorem 6, Theorem 7 and Lemma 12, the following inequalities hold with high probability for constant c, c_1 and c_2:

$$(I)\ \sup_{\mathbf{w}\in\mathcal{S}_p} \frac{1}{t}\sum_{i=1}^t (\mathbf{w}^\top \mathbf{n}_i)^2 \le c\tau,$$

$$(II)\ \sup_{\mathbf{w}\in\mathcal{S}_d} \left| \frac{1}{t}\sum_{i=1}^t (\mathbf{w}^\top \mathbf{x}_i)^2 - 1 \right| \le c_1\theta\sqrt{\frac{d}{n}} \triangleq \epsilon_0,$$

$$(III)\ \sup_{\mathbf{w}\in\mathcal{S}_d} \left| \frac{1}{t}\sum_{i=1}^{\bar{t}} |\mathbf{w}^\top \mathbf{x}|_{(i)}^2 - \mathcal{V}(\frac{\bar{t}}{t}) \right| \le \frac{c_2 t(1+\epsilon_0)\sqrt{d\log n/n}}{t-\bar{t}} \wedge c_2\theta^{\frac{1}{2}} d^{\frac{1}{4}} (\log n)^{\frac{3}{4}} n^{-\frac{1}{4}} \triangleq \epsilon_1(\frac{\bar{t}}{t}),$$

$$(IV)\ \|\frac{1}{t}\sum_{i=1}^t \mathbf{n}_i \mathbf{n}_i^\top\|_\infty \le c\varsigma,$$

where $\tau = \max\{\frac{p}{n}, 1\}$ and $\varsigma = \max\{\sqrt{\frac{\log p}{n}}, 1\}$. When $\bar{t} = t$, we can indeed sharpen the result of (III) by applying (II), so let $\epsilon_1(1) = \epsilon_0$. We have the following theorem:

Theorem 8. *There exists a constant c such that the following inequalities hold w.h.p,*

$$\begin{aligned}
&\|\mathbf{X}^{1/2}\mathbf{A}\|_F^2 \left(\mathcal{V}^-(\frac{\bar{t}}{t}) - \epsilon_1(\frac{\bar{t}}{t}) \right) - 2\|\mathbf{X}^{1/2}\mathbf{A}\|_F \sqrt{(1+\epsilon_0)ck\min\{\tau, \gamma\varsigma\}} \\
\le& \frac{1}{t}\sum_{i=1}^{\bar{t}} \langle \mathbf{z}\mathbf{z}^\top, \mathbf{X} \rangle_{(i)} \\
\le& \|\mathbf{X}^{1/2}\mathbf{A}\|_F^2 \left(\mathcal{V}^+(\frac{\bar{t}}{t}) + \epsilon_1(\frac{\bar{t}}{t}) \right) + 2\|\mathbf{X}^{1/2}\mathbf{A}\|_F \sqrt{(1+\epsilon_0)ck\min\{\tau, \gamma\varsigma\}} + ck\min\{\tau, \gamma\varsigma\},
\end{aligned}$$

for any $\bar{t} \le t$ and $\mathbf{X} \in \mathcal{F}(k)$.

4.3. Step 3

Suppose that a "good" solution $\mathbf{X}_s$ is found at stage s $(0 \le s \le s_0)$, namely event $\mathcal{E}(s)$ is true. We can bound $H^*/\overline{H}$ by leveraging the relationship between $\mathbf{X}_s$ and $\mathbf{\Pi}$ and the connection between $\mathbf{X}^*$ and $\mathbf{X}_s$.

Lemma 15. *If $\|\mathbf{\Pi}\|_0 \le \beta^2$ and $\mathcal{E}(s)$ is true for $s \le s_0$, there exists a constant c such that the following inequalities hold w.h.p,*

$$\begin{aligned}&(1+\epsilon_0)H_s + 2\sqrt{(1+\epsilon_0)ck\min\{\tau,\gamma\varsigma\}H_s} + ck\min\{\tau,\gamma\varsigma\}\\ \ge&\frac{1-\eta}{\kappa+1}\left[\left(\mathcal{V}^-(1-\frac{\rho}{(1-\rho)\kappa}) - \epsilon_1(1-\frac{\rho}{(1-\rho)\kappa})\right)\overline{H} - 2\sqrt{(1+\epsilon_0)ck\min\{\tau,\gamma\varsigma\}\overline{H}} - \frac{1}{1-\rho}\mu\beta\sqrt{k}\right].\end{aligned}$$

Lemma 16. *Fix $\hat{t} \le t$, there exists constant c such that the following inequalities hold w.h.p,*

$$\begin{aligned}&\left(\mathcal{V}^+(\frac{\hat{t}}{t}) + \epsilon_1(\frac{\hat{t}}{t})\right)H^* + 2\sqrt{(1+\epsilon_0)ck\min\{\tau,\gamma\varsigma\}H^*} + ck\min\{\tau,\gamma\varsigma\}\\ \ge&\left(\mathcal{V}^-(\frac{\hat{t}-\rho n}{t}) - \epsilon_1(\frac{\hat{t}-\rho n}{t})\right)H_s - 2\sqrt{(1+\epsilon_0)ck\min\{\tau,\gamma\varsigma\}H_s}.\end{aligned}$$

Theorem 9. *Suppose $\|\mathbf{\Pi}\|_0 \le \beta^2$ and $\rho \le 0.5$. For any κ, there exists a constant c such that the following inequalities hold w.h.p,*

$$\begin{aligned}\frac{H^*}{\overline{H}} \ge& \frac{\mathcal{V}^-\left(\frac{\hat{t}}{t} - \frac{\rho}{1-\rho}\right)\mathcal{V}^-\left(1 - \frac{\rho}{\kappa(1-\rho)}\right)}{(1+\kappa)\mathcal{V}^+\left(\frac{\hat{t}}{t}\right)} - \frac{10}{\mathcal{V}^+(0.5)}\left(\frac{ck\min\{\tau,\gamma\varsigma\}}{\overline{H}}\right)^{1/2}\\ &- \frac{c\{\theta^{\frac{1}{2}}d^{\frac{1}{4}}(\log^{\frac{3}{4}} n)n^{-\frac{1}{4}} \vee \theta[(1+\kappa)/\kappa]^{\frac{3}{2}}(\log^{\frac{3}{2}} n)n^{-\frac{1}{2}}\}}{\mathcal{V}^+(0.5)} - \frac{2\mu\beta\sqrt{k}}{\mathcal{V}^+(\frac{\hat{t}}{t})\overline{H}}\end{aligned}$$

4.4. Details of the Proof

4.4.1. Proof of Lemma 14

Proof. Since $0 \preceq \mathbf{X} \preceq \mathbf{I}_p$, we have

$$\begin{aligned}&\left|\langle\mathbf{A}\mathbf{A}^\top, \mathbf{X}\rangle - \sum_{i=1}^k \mathbf{w}_i{}^\top\mathbf{A}\mathbf{A}^\top\mathbf{w}_i\right| \le \|\mathbf{A}\mathbf{A}^\top\|_* \cdot \|\mathbf{X} - \sum_{i=1}^k \mathbf{w}_i\mathbf{w}_i^\top\|_2\\ =&\mathrm{tr}(\mathbf{A}\mathbf{A}^\top)\cdot\|\sum_{i=1}^k(\lambda_i - 1)\mathbf{w}_i\mathbf{w}_i^\top + \sum_{i=k+1}^p \lambda_i\mathbf{w}_i\mathbf{w}_i^\top\|_2 = \mathrm{tr}(\mathbf{A}\mathbf{A}^\top)\cdot\max\{1-\lambda_k(\mathbf{X}), \lambda_{k+1}(\mathbf{X})\}.\end{aligned}$$

Hence we obtain this lemma. □

4.4.2. Proof of Lemma 2

Proof. Let $\mathbf{S} = \mathbf{S}_n$, $\mu = \mu_n$ and $\mathbf{\Delta} = \mathbf{B}_n - \mathbf{A}_n$, then $\langle\mathbf{S}, \Delta\rangle \ge 0$ and $\langle\mathbf{S}, \Delta\rangle \le \mu\|\mathbf{B}_n\|_1 - \mu\|\mathbf{A}_n\|_1 \le \mu\|\mathbf{B}_n\|_1$. Since $\mathrm{tr}(\mathbf{B}_n) = d$ and $\mathbf{B}_n \succeq 0$, $\|\mathbf{B}_n\|_1 \le p\|\mathbf{B}_n\|_F = p\sqrt{d}$. Then we have $0 \le \langle\mathbf{S}, \mathbf{\Delta}\rangle \le \mu p\sqrt{d}$. Since $\mathbf{A}_n, \mathbf{B}_n \in \mathcal{F}_d$,

$$0 \le \mathrm{tr}(\mathbf{S}\mathbf{\Delta}) \le \mu p\sqrt{d},\ 0 \preceq \mathbf{B}_n - \mathbf{\Delta} \preceq \mathbf{I}_p,\ \mathrm{tr}(\mathbf{\Delta}) = 0.$$

By SVD decomposition, $\mathbf{S} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^\top$ where $\mathbf{Q}$ is an orthogonal matrix and $\mathbf{\Lambda}$ is a diagonal matrix. Let $\bar{\mathbf{\Delta}} = \mathbf{Q}^\top\mathbf{\Delta}\mathbf{Q}$, then,

$$0 \le \mathrm{tr}(\mathbf{\Lambda}\bar{\mathbf{\Delta}}) \le \mu p\sqrt{d},\ 0 \preceq \mathbf{\Sigma} - \bar{\mathbf{\Delta}} \preceq \mathbf{I}_p,\ \mathrm{tr}(\bar{\mathbf{\Delta}}) = 0,$$

where $\mathbf{\Sigma} = \begin{bmatrix} I_d & 0 \\ 0 & 0 \end{bmatrix}$. Thus, $0 \le \sum_{i=1}^p \lambda_i\bar{\Delta}_{ii} \le \mu p\sqrt{d}$ and

$$\begin{aligned}0 \le \bar{\Delta}_{ii} \le 1 \text{ for } 1 \le i \le d,\\ -1 \le \bar{\Delta}_{ii} \le 0 \text{ for } d+1 \le i \le p,\end{aligned}$$

which implies that $\sum_{i=d+1}^p |\bar{\Delta}_{ii}| \leq \frac{\mu p\sqrt{d}}{\delta}$ (otherwise, $\sum_{i=1}^p \lambda_i \bar{\Delta}_{ii} \geq \sum_{i=1}^p \lambda_d \bar{\Delta}_{ii} + |\lambda_d - \lambda_{d+1}| \sum_{i=d+1}^p |\Delta_{ii}| > \mu p\sqrt{d}$). Since $\mathrm{tr}(\bar{\mathbf{\Delta}}) = 0$, we also have $\sum_{i=1}^d \bar{\Delta}_{ii} \leq \frac{\mu p\sqrt{d}}{\delta}$. Let $\bar{\mathbf{\Delta}} = \begin{bmatrix} \mathbf{\Delta}_1 & -\mathbf{D} \\ -\mathbf{D}^\top & -\mathbf{\Delta}_2 \end{bmatrix}$, then $0 \preceq \begin{bmatrix} \mathbf{I}_d - \mathbf{\Delta}_1 & \mathbf{D} \\ \mathbf{D}^\top & \mathbf{\Delta}_2 \end{bmatrix} \preceq \mathbf{I}_p$, which implies that $\mathbf{\Delta}_1 \succeq 0$ and $\mathbf{\Delta}_2 \succeq 0$. Hence

$$\begin{aligned}
\|\bar{\mathbf{\Delta}}\|_F^2 =& \|\mathbf{\Delta}_1\|_F^2 + \|\mathbf{\Delta}_2\|_F^2 + 2\|\mathbf{D}\|_F^2 \\
\leq& \mathrm{tr}(\mathbf{\Delta}_1)^2 + \mathrm{tr}(\mathbf{\Delta}_2)^2 + 2\sum_{i=1}^d \sum_{j=d+1}^p D_{ij}^2 \\
\leq& 2\left(\frac{\mu\beta\sqrt{d}}{\delta}\right)^2 + 2\sum_{i=1}^d \sum_{j=d+1}^p |(1-\bar{\Delta}_{ii})\bar{\Delta}_{jj}| \\
\leq& 2\left(\frac{\mu\beta\sqrt{d}}{\delta}\right)^2 + 2\sum_{i=1}^d \sum_{j=d+1}^p |\bar{\Delta}_{jj}| \\
\leq& 2\left(\frac{\mu\beta\sqrt{d}}{\delta}\right)^2 + 2\frac{\mu p d^{3/2}}{\delta}.
\end{aligned}$$

Thus, $\|\mathbf{A}_n - \mathbf{B}_n\|_F^2 = \|\mathbf{Q}\bar{\mathbf{\Delta}}\mathbf{Q}^\top\|_F^2 = \|\bar{\mathbf{\Delta}}\|_F^2 \leq 2\left(\frac{\mu p\sqrt{d}}{\delta}\right)^2 + 2\frac{\mu p d^{3/2}}{\delta} \to 0$ as $\mu \to 0$ when $pd^{3/2} = o(\frac{1}{\mu})$. $\square$

4.4.3. Proof of Theorem 8

Proof. For an arbitrary $\mathbf{w} \in \mathcal{S}_p$, let $j(i)$ be permutations of $\{1, \cdots, n\}$ such that $(\mathbf{w}^\top \mathbf{x}_{j(i)})^2$ is non-decreasing. Thus,

$$\begin{aligned}
&\frac{1}{t}\sum_{i=1}^{\bar{t}} \langle \mathbf{z}\mathbf{z}^\top, \mathbf{X}\rangle_{(i)} \leq \frac{1}{t}\sum_{i=1}^{\bar{t}} \mathrm{tr}\left((\mathbf{A}\mathbf{x}_{j(i)} + \mathbf{n}_{j(i)})^\top \mathbf{X}(\mathbf{A}\mathbf{x}_{j(i)} + \mathbf{n}_{j(i)})\right) \\
=&\frac{1}{t}\sum_{i=1}^{\bar{t}} \mathrm{tr}\left(\mathbf{x}_{j(i)}^\top \mathbf{A}^\top \mathbf{X}\mathbf{A}\mathbf{x}_{j(i)} + 2\mathbf{n}_{j(i)}^\top \mathbf{X}\mathbf{A}\mathbf{x}_{j(i)} + \mathbf{n}_{j(i)}^\top \mathbf{X}\mathbf{n}_{j(i)}\right) \\
\leq&\frac{1}{t}\sum_{i=1}^{\bar{t}} \mathbf{x}_{j(i)}^\top \mathbf{A}^\top \mathbf{X}\mathbf{A}\mathbf{x}_{j(i)} + \frac{2}{t}\sum_{i=1}^{\bar{t}} \langle \mathbf{X}^{1/2}\mathbf{A}\mathbf{x}_{j(i)}, \mathbf{X}^{1/2}\mathbf{n}_{j(i)}\rangle + \frac{1}{t}\sum_{i=1}^{t} \langle \mathbf{X}^{1/2}\mathbf{n}_i, \mathbf{X}^{1/2}\mathbf{n}_i\rangle \\
\leq&\|\mathbf{A}^\top \mathbf{X}\mathbf{A}\|_* \cdot \|\frac{1}{t}\sum_{i=1}^{\bar{t}} \mathbf{x}_{j(i)}\mathbf{x}_{j(i)}^\top\|_2 + \frac{2}{t}\sum_{i=1}^{t} \|\mathbf{X}^{1/2}\mathbf{A}\mathbf{x}_i\|_2 \cdot \|\mathbf{X}^{1/2}\mathbf{n}_i\|_2 + \frac{1}{t}\sum_{i=1}^{t} \|\mathbf{X}^{1/2}\mathbf{n}_i\|_2^2
\end{aligned}$$

Since $\|\mathbf{A}^\top \mathbf{X}\mathbf{A}\|_* = \|\mathbf{X}^{1/2}\mathbf{A}\|_F^2$ and the Cauchy-Schwarz inequality holds, we have

$$\begin{aligned}
\frac{1}{t}\sum_{i=1}^{\bar{t}} \langle \mathbf{z}\mathbf{z}^\top, \mathbf{X}\rangle_{(i)} \leq& \|\mathbf{X}^{1/2}\mathbf{A}\|_F^2 \cdot \sup_{\mathbf{w}\in\mathcal{S}_d} \frac{1}{t}\sum_{i=1}^{\bar{t}} (\mathbf{w}^\top \mathbf{x})_{(i)}^2 + \\
&2\sqrt{\frac{1}{t}\sum_{i=1}^{t} \|\mathbf{X}^{1/2}\mathbf{A}\mathbf{x}_i\|_2^2} \cdot \sqrt{\frac{1}{t}\sum_{i=1}^{t} \|\mathbf{X}^{1/2}\mathbf{n}_i\|_2^2} + \frac{1}{t}\sum_{i=1}^{t} \|\mathbf{X}^{1/2}\mathbf{n}_i\|_2^2 \\
\leq& \|\mathbf{X}^{1/2}\mathbf{A}\|_F^2 \cdot \sup_{\mathbf{w}\in\mathcal{S}_d} \frac{1}{t}\sum_{i=1}^{\bar{t}} (\mathbf{w}^\top \mathbf{x})_{(i)}^2 + \\
&2\|\mathbf{X}^{1/2}\mathbf{A}\|_F \sqrt{\sup_{\mathbf{w}\in\mathcal{S}_d} \frac{1}{t}\sum_{i=1}^{t} (\mathbf{w}^\top \mathbf{x}_i)^2} \cdot \sqrt{\frac{1}{t}\sum_{i=1}^{t} \|\mathbf{X}^{1/2}\mathbf{n}_i\|_2^2} + \frac{1}{t}\sum_{i=1}^{t} \|\mathbf{X}^{1/2}\mathbf{n}_i\|_2^2
\end{aligned}$$

Note that

$$\frac{1}{t}\sum_{i=1}^{t}\|\mathbf{X}^{1/2}\mathbf{n}_i\|_2^2 = \langle \mathbf{X}, \frac{1}{t}\sum_{i=1}^{t}\mathbf{n}_i\mathbf{n}_i^\top\rangle \leq \min\{\|\mathbf{X}\|_* \cdot \|\frac{1}{t}\sum_{i=1}^{t}\mathbf{n}_i\mathbf{n}_i^\top\|_2, \|\mathbf{X}\|_1 \cdot \|\frac{1}{t}\sum_{i=1}^{t}\mathbf{n}_i\mathbf{n}_i^\top\|_\infty\}$$
$$\leq \min\{k\|\frac{1}{t}\sum_{i=1}^{t}\mathbf{n}_i\mathbf{n}_i^\top\|_2, \gamma k\|\frac{1}{t}\sum_{i=1}^{t}\mathbf{n}_i\mathbf{n}_i^\top\|_\infty\}.$$

Then from (I)(II)(III)(IV), we have w.h.p

$$\frac{1}{t}\sum_{i=1}^{\bar{t}}\langle \mathbf{z}\mathbf{z}^\top, \mathbf{X}\rangle_{(i)} \leq \|\mathbf{X}^{1/2}\mathbf{A}\|_F^2\left(\mathcal{V}^+(\frac{\bar{t}}{t}) + \epsilon_1(\frac{\bar{t}}{t})\right) + 2\|\mathbf{X}^{1/2}\mathbf{A}\|_F\sqrt{(1+\epsilon_0)ck\min\{\tau,\gamma\varsigma\}} + ck\min\{\tau,\gamma\varsigma\}.$$

We now compute the lower bound. For an arbitrary $\mathbf{w} \in \mathcal{S}_p$, let $k(i)$ be permutations of $\{1,\cdots,n\}$ such that $\langle \mathbf{z}_{k(i)}\mathbf{z}_{k(i)}^\top, \mathbf{X}\rangle$ is non-decreasing, then

$$\frac{1}{t}\sum_{i=1}^{\bar{t}}\langle \mathbf{z}\mathbf{z}^\top, \mathbf{X}\rangle_{(i)} = \frac{1}{t}\sum_{i=1}^{\bar{t}}\mathbf{x}_{k(i)}^\top\mathbf{A}^\top\mathbf{X}\mathbf{A}\mathbf{x}_{k(i)} + \frac{2}{t}\sum_{i=1}^{\bar{t}}\mathbf{n}_{k(i)}^\top\mathbf{X}\mathbf{A}\mathbf{x}_{k(i)} + \frac{1}{t}\sum_{i=1}^{\bar{t}}\mathbf{n}_{k(i)}^\top\mathbf{n}_{k(i)}$$
$$\geq \langle \mathbf{X}, \frac{1}{t}\mathbf{A}\sum_{i=1}^{\bar{t}}\mathbf{x}_{k(i)}^\top\mathbf{x}_{k(i)}\mathbf{A}^\top\rangle - \frac{2}{t}\sum_{i=1}^{t}\|\mathbf{X}^{1/2}\mathbf{A}\mathbf{x}_i\|_2 \cdot \|\mathbf{X}^{1/2}\mathbf{n}_i\|_2$$

Perform SVD on $\mathbf{X}$, we have $\mathbf{X} = \sum_{i=1}^{p}\alpha_i\mathbf{v}_i\mathbf{v}_i^\top$, then

$$\langle \mathbf{X}, \frac{1}{t}\mathbf{A}\sum_{i=1}^{\bar{t}}\mathbf{x}_{k(i)}^\top\mathbf{x}_{k(i)}\mathbf{A}^\top\rangle = \sum_{j=1}^{p}\frac{\alpha_j}{t}\sum_{i=1}^{\bar{t}}(\mathbf{v}_j^\top\mathbf{A}\mathbf{x}_{k(i)})^2$$
$$\geq \sum_{j=1}^{p}\frac{\alpha_j}{t}\sum_{i=1}^{\bar{t}}(\mathbf{v}_j^\top\mathbf{A}\mathbf{x})_{(i)}^2 \geq \|\mathbf{v}_j^\top\mathbf{A}\|_2^2 \cdot \sum_{j=1}^{p}\frac{\alpha_j}{t}\sum_{i=1}^{\bar{t}}(\frac{\mathbf{v}_j^\top\mathbf{A}}{\|\mathbf{v}_j^\top\mathbf{A}\|_2}\mathbf{x})_{(i)}^2$$

Then from Lemma 6 and Lemma 7 (Note that we assume $\mathbf{v}_j^\top\mathbf{A} \neq 0$ in the last inequality. We ignore the case that $\mathbf{v}_j^\top\mathbf{A} = 0$ since the bound holds trivially),

$$\frac{1}{t}\sum_{i=1}^{\bar{t}}\langle \mathbf{z}\mathbf{z}^\top, \mathbf{X}\rangle_{(i)}$$
$$\geq \sum_{j=1}^{p}\alpha_j\|\mathbf{v}_j^\top\mathbf{A}\|_2^2\left(\mathcal{V}^-(\frac{\bar{t}}{t}) - \epsilon_1(\frac{\bar{t}}{t})\right) - 2\|\mathbf{X}^{1/2}\mathbf{A}\|_F\sqrt{(1+\epsilon_0)ck\min\{\tau,\gamma\varsigma\}}$$
$$= \mathrm{tr}(\mathbf{A}^\top \cdot \sum_{j=1}^{p}\alpha_j\mathbf{v}_j\mathbf{v}_j^\top \cdot \mathbf{A})\left(\mathcal{V}^-(\frac{\bar{t}}{t}) - \epsilon_1(\frac{\bar{t}}{t})\right) - 2\|\mathbf{X}^{1/2}\mathbf{A}\|_F\sqrt{(1+\epsilon_0)ck\min\{\tau,\gamma\varsigma\}}$$
$$= \|\mathbf{X}^{1/2}\mathbf{A}\|_F^2\left(\mathcal{V}^-(\frac{\bar{t}}{t}) - \epsilon_1(\frac{\bar{t}}{t})\right) - 2\|\mathbf{X}^{1/2}\mathbf{A}\|_F\sqrt{(1+\epsilon_0)ck\min\{\tau,\gamma\varsigma\}}$$

Hence the theorem holds. □

4.5. Proof of Lemma 15

Proof. Since $\mathcal{E}(s)$ is true, we have $\sum_{i\in\mathcal{Z}}\alpha_i(s)\langle \mathbf{y}_i\mathbf{y}_i^\top, \mathbf{X}_s\rangle \geq \frac{1}{\kappa}\sum_{i\in\mathcal{O}}\alpha_i(s)\langle \mathbf{y}_i\mathbf{y}_i^\top, \mathbf{X}_s\rangle$, which implies that

$$(\kappa+1)\sum_{i\in\mathcal{Z}}\alpha_i(s)\langle \mathbf{y}_i\mathbf{y}_i^\top, \mathbf{X}_s\rangle \geq \sum_{i=1}^{n}\alpha_i(s)\langle \mathbf{y}_i\mathbf{y}_i^\top, \mathbf{X}_s\rangle \geq (1-\eta)\left(\sum_{i=1}^{n}\alpha_i(s)\langle \mathbf{y}_i\mathbf{y}_i^\top, \mathbf{\Pi}\rangle - n\mu\|\mathbf{\Pi}\|_1\right),$$

where the last inequality holds because $\mathbf{X}_s$ is the $(1-\eta)$-optimal solution of the PCA-like algorithm at stage s. Note that $\|\mathbf{\Pi}\|_1 \leq \beta\|\mathbf{\Pi}\|_F = \beta\sqrt{\mathrm{tr}(\mathbf{\Pi}^2)} = \beta\sqrt{k}$, then

$$
\begin{aligned}
\frac{1}{t}\sum_{i=1}^{t}\langle \mathbf{z}_i\mathbf{z}_i^\top, \mathbf{X}_s\rangle &\geq \frac{1}{t}\sum_{i\in\mathcal{Z}}\alpha_i(s)\langle \mathbf{y}_i\mathbf{y}_i^\top, \mathbf{X}_s\rangle \\
&\geq \frac{1-\eta}{\kappa+1}\left(\frac{1}{t}\sum_{i=1}^{n}\alpha_i(s)\langle \mathbf{y}_i\mathbf{y}_i^\top, \mathbf{\Pi}\rangle - \frac{n}{t}\mu\beta\sqrt{k}\right) \\
&\geq \frac{1-\eta}{\kappa+1}\left(\frac{1}{t}\sum_{i=1}^{t-\rho n/\kappa}\langle \mathbf{z}_i\mathbf{z}_i^\top, \mathbf{\Pi}\rangle_{(i)} - \frac{n}{t}\mu\beta\sqrt{k}\right),
\end{aligned}
$$

where the last inequality follows from Equation (4). From Theorem 8, the following inequality holds w.h.p,

$$
\begin{aligned}
&(1+\epsilon_0)H_s + 2\sqrt{(1+\epsilon_0)ck\min\{\tau,\gamma\varsigma\}H_s} + ck\min\{\tau,\gamma\varsigma\} \\
\geq&\frac{1-\eta}{\kappa+1}\left[\left(\mathcal{V}^-(\frac{t-\rho n/\kappa}{t}) - \epsilon_1(\frac{t-\rho n/\kappa}{t})\right)\overline{H} - 2\sqrt{(1+\epsilon_0)ck\min\{\tau,\gamma\varsigma\}\overline{H}} - \frac{n}{n-\rho n}\mu\beta\sqrt{k}\right] \\
=&\frac{1-\eta}{\kappa+1}\left[\left(\mathcal{V}^-(1-\frac{\rho}{(1-\rho)\kappa}) - \epsilon_1(1-\frac{\rho}{(1-\rho)\kappa})\right)\overline{H} - 2\sqrt{(1+\epsilon_0)ck\min\{\tau,\gamma\varsigma\}\overline{H}} - \frac{1}{1-\rho}\mu\beta\sqrt{k}\right]
\end{aligned}
$$

□

4.6. Proof of Lemma 16

Proof. Since $|\mathcal{O}| = |\mathcal{Y}\backslash\mathcal{Z}| = \rho n$, we have

$$
\sum_{i=1}^{\hat{t}-\rho n}\langle \mathbf{z}\mathbf{z}^\top, \mathbf{X}^*\rangle_{(i)} \leq \sum_{i=1}^{\hat{t}}\langle \mathbf{y}\mathbf{y}^\top, \mathbf{X}^*\rangle_{(i)} \leq \sum_{i=1}^{\hat{t}}\langle \mathbf{z}\mathbf{z}^\top, \mathbf{X}^*\rangle_{(i)}.
$$

Since $\mathbf{X}^*$ is the final output of this algorithm, $\overline{V}_{\hat{t}}(\mathbf{X}^*) \geq \overline{V}_{\hat{t}}(\mathbf{X}_s)$. Thus,

$$
\frac{1}{\hat{t}}\sum_{i=1}^{\hat{t}}\langle \mathbf{z}\mathbf{z}^\top, \mathbf{X}^*\rangle_{(i)} \geq \overline{V}_{\hat{t}}(\mathbf{X}^*) \geq \overline{V}_{\hat{t}}(\mathbf{X}_s) \geq \frac{1}{\hat{t}}\sum_{i=1}^{\hat{t}-\rho n}\langle \mathbf{z}\mathbf{z}^\top, \mathbf{X}_s\rangle_{(i)}.
$$

Then from Theorem 8, the following inequality holds w.h.p,

$$
\begin{aligned}
&\left(\mathcal{V}^+(\frac{\hat{t}}{t}) + \epsilon_1(\frac{\hat{t}}{t})\right)H^* + 2\sqrt{(1+\epsilon_0)ck\min\{\tau,\gamma\varsigma\}H^*} + ck\min\{\tau,\gamma\varsigma\} \\
\geq&\left(\mathcal{V}^-(\frac{\hat{t}-\rho n}{t}) - \epsilon_1(\frac{\hat{t}-\rho n}{t})\right)H_s - 2\sqrt{(1+\epsilon_0)ck\min\{\tau,\gamma\varsigma\}H_s}.
\end{aligned}
$$

□

4.7. Proof of Theorem 9

Proof. Recall that with high probability $\mathcal{E}(s)$ is true for $s \leq s_0$ and notice that we can assume $\epsilon_0 \leq 1$ for large enough n. From Lemma 15 and Lemma 16, since $\overline{H} \geq H^*$ and $\overline{H} \geq H_s$, the following inequalities hold w.h.p,

$$
\begin{aligned}
&\frac{1-\eta}{\kappa+1}\left[\left(\mathcal{V}^-(1-\frac{\rho}{(1-\rho)\kappa}) - \epsilon_1(1-\frac{\rho}{(1-\rho)\kappa})\right)\overline{H} - 2\sqrt{(1+\epsilon_0)ck\min\{\tau,\gamma\varsigma\}\overline{H}} - \frac{1}{1-\rho}\mu\beta\sqrt{k}\right] \\
\leq&(1+\epsilon_0)H_s + 2\sqrt{(1+\epsilon_0)ck\min\{\tau,\gamma\varsigma\}\overline{H}} + ck\min\{\tau,\gamma\varsigma\},
\end{aligned}
$$

and

$$\left(\mathcal{V}^-(\frac{\hat{t}-\rho n}{t}) - \epsilon_1(\frac{\hat{t}-\rho n}{t})\right) H_s \le \left(\mathcal{V}^+(\frac{\hat{t}}{t}) + \epsilon_1(\frac{\hat{t}}{t})\right) H^* + 4\sqrt{(1+\epsilon_0)ck\min\{\tau,\gamma\varsigma\}\overline{H}} + ck\min\{\tau,\gamma\varsigma\}.$$

By re-organization, we have

$$\left(\mathcal{V}^-(1-\frac{\rho}{(1-\rho)\kappa}) - \epsilon_1(1-\frac{\rho}{(1-\rho)\kappa})\right)\overline{H} - \frac{2\kappa+4}{1-\eta}\sqrt{(1+\epsilon_0)ck\min\{\tau,\gamma\varsigma\}\overline{H}} - \frac{\mu\beta\sqrt{k}}{1-\rho} - \frac{1+\kappa}{1-\eta}ck\min\{\tau,\gamma\varsigma\} \le \frac{1+\kappa}{1-\eta}(1+\epsilon_0)H_s,$$

$$\left(\mathcal{V}^-(\frac{\hat{t}-\rho n}{t}) - \epsilon_1(\frac{\hat{t}-\rho n}{t})\right) H_s \le \left(\mathcal{V}^+(\frac{\hat{t}}{t}) + \epsilon_1(\frac{\hat{t}}{t})\right) H^* + 4\sqrt{(1+\epsilon_0)ck\min\{\tau,\gamma\varsigma\}\overline{H}} + ck\min\{\tau,\gamma\varsigma\}.$$

Let $\epsilon_1 = c_2\theta^{\frac{1}{2}}d^{\frac{1}{4}}(\log n)^{\frac{3}{4}}n^{-\frac{1}{4}}$. Since $\epsilon_1(\frac{\hat{t}-\rho n}{t}) \le \epsilon_1$ and $\epsilon_1(\frac{t-s_0}{t}) \le \epsilon_1$, we have

$$\begin{aligned}\frac{H^*}{\overline{H}} \ge & \frac{(1-\eta)\left(\mathcal{V}^-(1-\frac{\rho}{(1-\rho)\kappa})-\epsilon_1\right)\left(\mathcal{V}^-(\frac{\hat{t}}{t}-\frac{\rho}{1-\rho})-\epsilon_1\right)}{(1+\kappa)(1+\epsilon_0)\left(\mathcal{V}^+(\frac{\hat{t}}{t})+\epsilon_1\right)} - \frac{(2\kappa+4)\left(\mathcal{V}^-(\frac{\hat{t}}{t}-\frac{\rho}{1-\rho})-\epsilon_1\right)\sqrt{(1+\epsilon_0)ck\min\{\tau,\gamma\varsigma\}}}{(1+\kappa)(1+\epsilon_0)\left(\mathcal{V}^+(\frac{\hat{t}}{t})+\epsilon_1\right)}\overline{H}^{-1/2} \\ & - \frac{4(1+\kappa)(1+\epsilon_0)\sqrt{(1+\epsilon_0)ck\min\{\tau,\gamma\varsigma\}}}{(1+\kappa)(1+\epsilon_0)\left(\mathcal{V}^+(\frac{\hat{t}}{t})+\epsilon_1\right)}\overline{H}^{-1/2} - \frac{\left(\mathcal{V}^-(\frac{\hat{t}}{t}-\frac{\rho}{1-\rho})-\epsilon_1+1+\epsilon_0\right)ck\min\{\tau,\gamma\varsigma\}}{(1+\epsilon_0)\left(\mathcal{V}^+(\frac{\hat{t}}{t})+\epsilon_1\right)}\overline{H}^{-1} \\ & - \frac{(1-\eta)\frac{\mu\beta\sqrt{k}}{1-\rho}}{(1+\kappa)(1+\epsilon_0)\left(\mathcal{V}^+(\frac{\hat{t}}{t})+\epsilon_1\right)}\overline{H}^{-1}.\end{aligned}$$

Note that the last term

$$\frac{(1-\eta)\frac{\mu\beta\sqrt{k}}{1-\rho}}{(1+\kappa)(1+\epsilon_0)\left(\mathcal{V}^+(\frac{\hat{t}}{t})+\epsilon_1\right)}\overline{H}^{-1} \le \frac{(1-\eta)\frac{\mu\beta\sqrt{k}}{1-\rho}}{\mathcal{V}^+(\frac{\hat{t}}{t})}\overline{H}^{-1} \le \frac{2(1-\eta)\mu\beta\sqrt{k}}{\mathcal{V}^+(\frac{\hat{t}}{t})}\overline{H}^{-1}.$$

Since $\epsilon_0 = c_1\theta\sqrt{\frac{d}{n}} = \epsilon_0, \epsilon_1(\frac{\bar{t}}{t}) = \frac{c_2 t(1+\epsilon_0)\sqrt{d\log n/n}}{t-t} \wedge c_2\theta^{\frac{1}{2}}d^{\frac{1}{4}}(\log n)^{\frac{3}{4}}n^{-\frac{1}{4}}$, and $\mathcal{V}_{\mathbf{v}}(\kappa) - \mathcal{V}_{\mathbf{v}}(\kappa-\epsilon) \le C\theta\epsilon\log\epsilon$ by Lemma 5, we can follow the proof of Theorem 2 in (Xu et al., 2013) and obtain that the following inequality holds w.h.p,

$$\begin{aligned}\frac{H^*}{\overline{H}} \ge & \frac{(1-\eta)\mathcal{V}^-\left(\frac{\hat{t}}{t}-\frac{\rho}{1-\rho}\right)\mathcal{V}^-\left(1-\frac{\rho}{\kappa(1-\rho)}\right)}{(1+\kappa)\mathcal{V}^+\left(\frac{\hat{t}}{t}\right)} - \frac{10}{\mathcal{V}^+(0.5)}\left(\frac{ck\min\{\tau,\gamma\varsigma\}}{\overline{H}}\right)^{1/2} \\ & - \frac{c\{\theta^{\frac{1}{2}}d^{\frac{1}{4}}(\log^{\frac{3}{4}}n)n^{-\frac{1}{4}} \vee \theta[(1+\kappa)/\kappa]^{\frac{3}{2}}(\log^{\frac{3}{2}}n)n^{-\frac{1}{2}}\}}{\mathcal{V}^+(0.5)} - \frac{2(1-\eta)\mu\beta\sqrt{k}}{\mathcal{V}^+(\frac{\hat{t}}{t})\overline{H}}.\end{aligned}$$

□

4.8. Proof of Corollary 1

Proof. When $\kappa > 1$, the corollary holds trivially. Hence, fix $\kappa \le 1$. From Theorem 4, we have

$$
\begin{aligned}
&\liminf_k \mathrm{E.V}\{\mathbf{w}_1^*, \cdots, \mathbf{w}_k^*\}\\
\ge\ & (1-\eta)\max_\kappa \left[\frac{\mathcal{V}^-\left(1-\frac{\rho^*}{(1-\rho^*)\kappa}\right)}{1+\kappa}\right] \times \left[\frac{\mathcal{V}^-\left(\frac{\hat{t}}{t}-\frac{\rho^*}{1-\rho^*}\right)}{\mathcal{V}^+\left(\frac{\hat{t}}{t}\right)}\right]\\
\ge\ & (1-\eta)\max_\kappa \left[\frac{1}{1+\kappa} - \frac{C\theta\frac{\rho^*}{(1-\rho^*)\kappa}\log\frac{(1-\rho^*)\kappa}{\rho^*}}{1+\kappa}\right] \times \left[\frac{\mathcal{V}^-\left(\frac{\hat{t}}{t}\right)}{\mathcal{V}^+\left(\frac{\hat{t}}{t}\right)} - \frac{C\theta\frac{\rho^*}{1-\rho^*}\log\frac{1-\rho^*}{\rho^*}}{\mathcal{V}^+\left(\frac{\hat{t}}{t}\right)}\right]\\
\ge\ & (1-\eta)\max_\kappa \left[1-\kappa-\frac{C\theta\rho^*}{\kappa}\log\frac{1}{2\rho^*}\right] \times \left[\frac{\mathcal{V}^-\left(\frac{\hat{t}}{t}\right)}{\mathcal{V}^+\left(\frac{\hat{t}}{t}\right)} - \frac{C\theta\rho^*\log\frac{1}{2\rho^*}}{\mathcal{V}^+\left(\frac{\hat{t}}{t}\right)}\right]\\
\ge\ & (1-\eta)\max_\kappa \left[1-\kappa-\frac{C\theta\rho^*}{\kappa}\log\frac{1}{2\rho^*}\right] \times \left[1 - \frac{C\theta\rho^*\log\frac{1}{2\rho^*}}{\mathcal{V}^-\left(\frac{\hat{t}}{t}\right)}\right] \times \frac{\mathcal{V}^-\left(\frac{\hat{t}}{t}\right)}{\mathcal{V}^+\left(\frac{\hat{t}}{t}\right)}\\
\ge\ & (1-\eta)\max_\kappa \left[1-\kappa-\left(\frac{1}{\kappa}+\frac{1}{\mathcal{V}^-\left(\frac{\hat{t}}{t}\right)}\right)C\theta\rho^*\log\frac{1}{2\rho^*}\right] \times \frac{\mathcal{V}^-\left(\frac{\hat{t}}{t}\right)}{\mathcal{V}^+\left(\frac{\hat{t}}{t}\right)}\\
\ge\ & (1-\eta)\max_\kappa \left[1-\kappa-\frac{C\theta\rho^*\log\frac{1}{2\rho^*}}{\kappa\mathcal{V}^-\left(\frac{\hat{t}}{t}\right)}\right] \times \frac{\mathcal{V}^-\left(\frac{\hat{t}}{t}\right)}{\mathcal{V}^+\left(\frac{\hat{t}}{t}\right)}.
\end{aligned}
$$

The second inequality is due to Lemma 5 and $\mathcal{V}^-(1) = 1$. The third inequality is due to $\rho^* < 0.5$ and $\kappa \le 1$. The sixth inequality holds because $\kappa \le 1$ and $\mathcal{V}^-\left(\frac{\hat{t}}{t}\right) \le 1$. Taking $\kappa = \sqrt{\theta\rho^*\log\frac{1}{2\rho^*}}$, we can obtain this corollary. $\square$

References

Loh, P. and Wainwright, M. J. High-dimensional regression with noisy and missing data: Provable guarantees with non-convexity. *The Annals of Statistics*, 40(3):1637–1664, 2012.

Vershynin, R. Introduction to the non-asymptotic analysis of random matrices. In *Compressed Sensing: Theory and Applications*, 2012.

Vu, V. Q., Cho, J., Lei, J., and Robe, K. Fantope projection and selection: A near-optimal convex relaxation of sparse PCA. In *Advances in Neural Information Processing Systems*, NIPS '13, 2013.

Xu, H., Caramanis, C., and Mannor, S. Outlier-robust PCA: the high-dimensional case. *IEEE Transactions on Information Theory*, 59(1):546–572, 2013.