Streaming Sparse Principal Component Analysis

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. Preliminaries

Theorem A-1. *(Theorem 3.1, (Chang, 2012)) Let* $\mathbf{A} \in \mathbb{R}^{m \times n}$ *be of full column rank with QR factorization* $\mathbf{A} = \mathbf{QR}$, $\Delta\mathbf{A}$ *be a perturbation in* $\mathbf{A}$*, and* $\mathbf{A} + \Delta\mathbf{A} = (\mathbf{Q} + \Delta\mathbf{Q})(\mathbf{R} + \Delta\mathbf{R})$ *be the QR-factorization of* $\mathbf{A} + \Delta\mathbf{A}$. *Let* $\mathbf{P}_{\mathbf{A}}$ *and* $\mathbf{P}_{\mathbf{A}^\perp}$ *be the orthogonal projectors onto the range of* $\mathbf{A}$ *and the orthogonal complement of the range of* $\mathbf{A}$*, respectively. Let* $\mathbf{Q}_\perp$ *be an orthonormal matrix such that matrix* $[\mathbf{Q}, \mathbf{Q}_\perp]$ *is orthogonal. Define* $\kappa_2(\mathbf{A}) = \|\mathbf{A}\|_2 \|\mathbf{A}^\dagger\|_2$*, where* $\mathbf{A}^\dagger$ *is the Moore-Penrose pseudo-inverse of* $\mathbf{A}$*. If*

$$(1+\sqrt{2})\kappa_2(\mathbf{A})\frac{\|\Delta\mathbf{A}\|_F}{\|\mathbf{A}\|_2} < 1,$$

then

$$\|\mathbf{P}_{\mathbf{A}^\perp}\Delta\mathbf{Q}\|_F \le \frac{\kappa_2(\mathbf{A})\frac{\|\mathbf{Q}_\perp^\top \Delta\mathbf{A}\|_F}{\|\mathbf{A}\|_2}}{1-(1+\sqrt{2})\kappa_2(\mathbf{A})\frac{\|\Delta\mathbf{A}\|_F}{\|\mathbf{A}\|_2}}.$$

Lemma A-1. *(Lemma 14, (Loh & Wainwright, 2012)) 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*

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

for constant c *that depends on* σ*.*

Theorem A-2. *Let* $\mathcal{F}$ *be a subset of* $\{1, 2, \cdots, p\}$ *with cardinality* $|\mathcal{F}| = s$*. For matrix* $\mathbf{X}$, $\mathbf{X}_\mathcal{F}$ *denotes the submatrix of* $\mathbf{X}$ *determined by the set* $\mathcal{F}$*, i.e., extracting the entries of* $\mathbf{X}$ *whose row and column indices are both in* $\mathcal{F}$*. Suppose that there are* n *samples* $\{\mathbf{x}_1, \cdots, \mathbf{x}_n\}$ *drawn from the spike model we have described in Section 2, then the following holds with probability at least* $1 - \frac{s^{-10}}{T}$*,*

$$\sup_{\mathcal{F}\subseteq[n], |\mathcal{F}|=s} \left\|(\frac{1}{n}\sum_{i=1}^n \mathbf{x}_i\mathbf{x}_i^\top - \mathbf{\Sigma})_\mathcal{F}\right\|_2 \le c\|\mathbf{\Sigma}\|_2\sqrt{\frac{s\log p + \log T}{n}},$$

for parameter T *and constant* c*.*

Proof. In the proof, the constants may vary from line to line. Notice that if we can show that the inequality

$$\mathbb{P}\left[\left\|(\frac{1}{n}\sum_{i=1}^n \mathbf{x}_i\mathbf{x}_i^\top - \mathbf{\Sigma})_\mathcal{F}\right\|_2 \ge t\right] \le 2\exp\left(-c_1 n\min\left(\frac{t^2}{\sigma^4}, \frac{t}{\sigma^2}\right) + c_2 s\right) \quad \text{(A-1)}$$

holds for all $t > 0$ and any fixed $\mathcal{F}$, where c_1 and c_2 are universal constants, then by the union bound, we have

$$\begin{aligned}\mathbb{P}\left[\sup_{\mathcal{F}\subseteq[n],|\mathcal{F}|=s}\left\|(\frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i\mathbf{x}_i^\top-\boldsymbol{\Sigma})_{\mathcal{F}}\right\|_2\geq t\right]\leq&2p^s\exp\left(-c_1 n\min\left(\frac{t^2}{\sigma^4},\frac{t}{\sigma^2}\right)+c_2 s\right)\\=&2\exp\left(-c_1 n\min\left(\frac{t^2}{\sigma^4},\frac{t}{\sigma^2}\right)+c_2 s\log p\right).\end{aligned}\tag{A-2}$$

The right hand side of (A-2) can be equal to $\frac{s^{-10}}{T}$ when $t=c\sigma^2\sqrt{\frac{s\log p+\log T}{n}}$ for large enough n and constant c. Notice that $\sigma^2 = c_3\sup_{\mathcal{F}\subseteq[n],|\mathcal{F}|=s}\|\boldsymbol{\Sigma}_{\mathcal{F}}\|_2\leq c_3\|\boldsymbol{\Sigma}\|_2$ for a certain constant c_3, we obtain Theorem A-2. To prove Inequality (A-1), we apply Lemma A-1. Let $\mathbf{x}_{i,\mathcal{F}}\in\mathbb{R}^s$ be the vector constructed by selecting the entries of $\mathbf{x}_i$ whose indices are in $\mathcal{F}$, then we have $\mathbb{E}[\mathbf{x}_{i,\mathcal{F}}\mathbf{x}_{i,\mathcal{F}}^\top]=\boldsymbol{\Sigma}_{\mathcal{F}}$. Let $\mathbf{X}_{\mathcal{F}}\in\mathbb{R}^{n\times s}$ be the matrix whose i^{th} row is $\mathbf{x}_{i,\mathcal{F}}^\top$. From Lemma A-1, for fixed unit vector $\mathbf{v}\in\mathbb{R}^s$ and $t>0$, there exists a universal constant c such that

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

or equivalently,

$$\mathbb{P}\left[\left|\frac{1}{n}\mathbf{v}^\top\mathbf{X}_{\mathcal{F}}^\top\mathbf{X}_{\mathcal{F}}\mathbf{v}-\mathbf{v}^\top\boldsymbol{\Sigma}_{\mathcal{F}}\mathbf{v}\right|>t\right]\leq 2\exp\left(-cn\min\left(\frac{t^2}{\sigma^4},\frac{t}{\sigma^2}\right)\right).$$

Suppose that $\mathcal{A}$ is a $1/3$ cover of $\mathcal{S}_s\triangleq\{\mathbf{v}\in\mathbb{R}^s:\|\mathbf{v}\|_2=1\}$, then for any $\mathbf{v}\in\mathcal{S}_s$, there exists a certain $\mathbf{u}\in\mathcal{A}$ so that $\|\mathbf{u}-\mathbf{v}\|_2\leq 1/3$. We define $\psi(\mathbf{v}_1,\mathbf{v}_2)\triangleq\left|\mathbf{v}_1^\top\left(\frac{\mathbf{X}_{\mathcal{F}}^\top\mathbf{X}_{\mathcal{F}}}{n}-\boldsymbol{\Sigma}_{\mathcal{F}}\right)\mathbf{v}_2\right|$, then

$$\sup_{\mathbf{v}\in\mathcal{S}_s}\psi(\mathbf{v},\mathbf{v})\leq\sup_{\mathbf{u}\in\mathcal{A}}\psi(\mathbf{u},\mathbf{u})+\sup_{\mathbf{v}\in\mathcal{S}_s,\mathbf{u}\in\mathcal{A},\|\mathbf{u}-\mathbf{v}\|_2\leq\frac{1}{3}}[2\psi(\mathbf{v}-\mathbf{u},\mathbf{u})+\psi(\mathbf{v}-\mathbf{u},\mathbf{v}-\mathbf{u})].$$

Since $\|\mathbf{u}-\mathbf{v}\|_2\leq\frac{1}{3}$, we have $\sup_{\mathbf{v}\in\mathcal{S}_s}\psi(\mathbf{v},\mathbf{v})\leq\sup_{\mathbf{u}\in\mathcal{A}}\psi(\mathbf{u},\mathbf{u})+(\frac{2}{3}+\frac{1}{9})\sup_{\mathbf{v}\in\mathcal{S}_s}\psi(\mathbf{v},\mathbf{v})$. Hence $\sup_{\mathbf{v}\in\mathcal{S}_s}\psi(\mathbf{v},\mathbf{v})\leq\frac{9}{2}\sup_{\mathbf{u}\in\mathcal{A}}\psi(\mathbf{u},\mathbf{u})$. Since $|\mathcal{A}|\leq 9^s$, the follows follow from the union bound,

$$\mathbb{P}\left[\sup_{\mathbf{v}\in\mathcal{S}_s}\psi(\mathbf{v},\mathbf{v})>t\right]\leq\mathbb{P}\left[\frac{9}{2}\sup_{\mathbf{u}\in\mathcal{A}}\psi(\mathbf{u},\mathbf{u})>t\right]\leq 9^s\cdot 2\exp\left(-cn\min\left(\frac{t^2}{\sigma^4},\frac{t}{\sigma^2}\right)\right),$$

which implies that

$$\mathbb{P}\left[\left\|\frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_{i,\mathcal{F}}\mathbf{x}_{i,\mathcal{F}}^\top-\boldsymbol{\Sigma}_{\mathcal{F}}\right\|_2>t\right]\leq 2\exp\left(-c_1 n\min\left(\frac{t^2}{\sigma^4},\frac{t}{\sigma^2}\right)+c_2 s\right).$$

Hence we obtain Inequality (A-1). □

Lemma A-2. *(Lemma 11, (Chen & Caramanis, 2013)) If samples* $\{\mathbf{y}_1,\cdots,\mathbf{y}_n\}$ *are independently drawn from* $\mathcal{N}(0,\mathbf{I}_p)$ *and samples* $\{\mathbf{z}_1,\cdots,\mathbf{z}_n\}$ *are independently drawn from* $\mathcal{N}(0,\mathbf{I}_d)$*, then*

$$\mathbb{P}\left[\left\|\frac{1}{n}\sum_{i=1}^{n}\mathbf{z}_i\mathbf{y}_i^\top\right\|_2>t\right]\leq 2\exp\left(-c_1 n\min\left(t^2,t\right)+c_2\max\{p,d\}\right).$$

Theorem A-3. *Suppose that there are* n *samples* $\{\mathbf{x}_1,\cdots,\mathbf{x}_n\}$ *drawn from the spike model we have described in Section 2, then the following holds with probability at least* $1-d^{-10}$,

$$\left\|\frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i\mathbf{x}_i^\top-\boldsymbol{\Sigma}\right\|_2\leq c\|\mathbf{A}\|_2^2\sqrt{\frac{d}{n}}+c(2\sigma\|\mathbf{A}\|_2+\sigma^2)\sqrt{\frac{p}{n}},$$

where c *is a universal constant.*

Proof. Since $\mathbf{x}_i = \mathbf{A}\mathbf{z}_i + \mathbf{w}_i$, we have

$$\begin{aligned}\left\|\frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i\mathbf{x}_i^\top - \mathbf{\Sigma}\right\|_2 &= \left\|\frac{1}{n}\sum_{i=1}^{n}(\mathbf{A}\mathbf{z}_i\mathbf{z}_i\mathbf{A}^\top + \mathbf{w}_i\mathbf{z}_i^\top\mathbf{A}^\top + \mathbf{A}\mathbf{z}_i\mathbf{w}_i^\top + \mathbf{w}_i\mathbf{w}_i^\top) - (\mathbf{A}\mathbf{A}^\top + \sigma^2\mathbf{I}_p)\right\|_2 \\ &\leq \left\|\mathbf{A}\left(\frac{1}{n}\sum_{i=1}^{n}\mathbf{z}_i\mathbf{z}_i^\top - \mathbf{I}_d\right)\mathbf{A}^\top\right\|_2 + 2\left\|\mathbf{A}\left(\frac{1}{n}\sum_{i=1}^{n}\mathbf{z}_i\mathbf{w}_i^\top\right)\right\|_2 + \left\|\frac{1}{n}\sum_{i=1}^{n}\mathbf{w}_i\mathbf{w}_i^\top - \sigma^2\mathbf{I}_p\right\|_2.\end{aligned}$$

From Lemma A-2 and the proof of Theorem A-2, the following inequalities hold with probability at least $1 - d^{-10}$:

$$\left\|\frac{1}{n}\sum_{i=1}^{n}\mathbf{z}_i\mathbf{z}_i^\top - \mathbf{I}_d\right\|_2 \leq c\sqrt{\frac{d}{n}},\ \left\|\frac{1}{n}\sum_{i=1}^{n}\mathbf{z}_i\mathbf{w}_i^\top\right\|_2 \leq c\sigma\sqrt{\frac{\max\{p,d\}}{n}},\ \left\|\frac{1}{n}\sum_{i=1}^{n}\mathbf{w}_i\mathbf{w}_i^\top - \sigma^2\mathbf{I}_p\right\|_2 \leq c\sigma^2\sqrt{\frac{p}{n}},$$

where c is a universal constant. Hence this theorem is obtained. ☐

Theorem A-4. *Let $k(\cdot) : \mathbb{R}^p \times \mathbb{R}^p \to \mathbb{R}^{p\times p}$ be a matrix value function defined as $k(\mathbf{x},\mathbf{y}) = \frac{(\mathbf{x}-\mathbf{y})(\mathbf{x}-\mathbf{y})^\top}{\|\mathbf{x}-\mathbf{y}\|_2^2}$. Let $\mathbf{x}_1, \cdots, \mathbf{x}_n$ be n independent observations of random vector $\mathbf{x}$. Suppose that n is even and $\mathbf{K} \triangleq \mathbb{E}[k(\mathbf{x},\tilde{\mathbf{x}})]$ exists where $\tilde{\mathbf{x}}$ is an independent copy of $\mathbf{x}$, then we have*

$$\mathbb{P}\left[\left\|\frac{2}{n}\sum_{i=1}^{n/2}k(\mathbf{x}_{2i-1},\mathbf{x}_{2i}) - \mathbf{K}\right\|_2 \geq t\right] \leq \begin{cases} p\cdot\exp\left(-\frac{3nt^2}{16(\|\mathbf{K}\|_2+\|\mathbf{K}\|_2^2)}\right), & t \leq \|\mathbf{K}\|_2 \\ p\cdot\exp\left(-\frac{3nt}{16(1+\|\mathbf{K}\|_2)}\right), & t \geq \|\mathbf{K}\|_2. \end{cases}$$

Proof. We let $m = n/2$ and $\mathbf{S}_i = \frac{1}{m}k(\mathbf{x}_{2i-1},\mathbf{x}_{2i}) - \frac{1}{m}\mathbf{K}$, then we have $\frac{2}{n}\sum_{i=1}^{n/2}k(\mathbf{x}_{2i-1},\mathbf{x}_{2i}) - \mathbf{K} = \sum_{i=1}^{m}\mathbf{S}_i$ and $\mathbb{E}[\mathbf{S}_i] = 0$. Note that for $i = 1,\cdots,m$,

$$\|\mathbf{S}_i - \mathbb{E}[\mathbf{S}_i]\|_2 = \frac{1}{m}\|k(\mathbf{x}_{2i-1},\mathbf{x}_{2i}) - \mathbf{K}\|_2 \leq \frac{1}{m}(\|k(\mathbf{x}_{2i-1},\mathbf{x}_{2i})\|_2 + \|\mathbf{K}\|_2) = \frac{1}{m}(1 + \|\mathbf{K}\|_2),$$

and

$$\begin{aligned}\left\|\mathbb{E}\left[\left(\sum_{i=1}^{m}\mathbf{S}_i\right)^2\right]\right\|_2 =& m\left\|\mathbb{E}\left[\mathbf{S}_i^2\right]\right\|_2 = m\left\|\mathbb{E}\left[\left(\frac{1}{m}k(\mathbf{x}_{2i-1},\mathbf{x}_{2i}) - \frac{1}{m}\mathbf{K}\right)^2\right]\right\|_2 = \frac{1}{m}\left\|\mathbb{E}\left[k(\mathbf{x}_{2i-1},\mathbf{x}_{2i})^2\right] - \mathbf{K}^2\right\|_2 \\ \leq& \frac{1}{m}\left(\left\|\mathbb{E}\left[k(\mathbf{x}_{2i-1},\mathbf{x}_{2i})^2\right]\right\|_2 + \|\mathbf{K}\|_2^2\right) = \frac{1}{m}\left(\|\mathbf{K}\|_2 + \|\mathbf{K}\|_2^2\right).\end{aligned}$$

From the Matrix Bernstein inequality (Tropp, 2012),

$$\mathbb{P}\left[\left\|\frac{1}{m}\sum_{i=1}^{m}k(\mathbf{x}_{2i-1},\mathbf{x}_{2i}) - \mathbf{K}\right\|_2 \geq t\right] \leq \begin{cases} p\cdot\exp\left(-\frac{3mt^2}{8(\|\mathbf{K}\|_2+\|\mathbf{K}\|_2^2)}\right), & t \leq \frac{\frac{1}{m}\left(\|\mathbf{K}\|_2+\|\mathbf{K}\|_2^2\right)}{\frac{1}{m}(1+\|\mathbf{K}\|_2)} = \|\mathbf{K}\|_2 \\ p\cdot\exp\left(-\frac{3mt}{8(1+\|\mathbf{K}\|_2)}\right), & \text{otherwise.} \end{cases}$$

Hence this theorem holds. ☐

Theorem A-5. *Define $S(\mathbf{x}) \triangleq \frac{\mathbf{x}-\tilde{\mathbf{x}}}{\|\mathbf{x}-\tilde{\mathbf{x}}\|_2}$ for random vector $\mathbf{x}$ and its independent copy $\tilde{\mathbf{x}}$. Let $k(\cdot) : \mathbb{R}^p \times \mathbb{R}^p \to \mathbb{R}^{p\times p}$ be a matrix value function defined as $k(\mathbf{x},\mathbf{y}) = \frac{(\mathbf{x}-\mathbf{y})(\mathbf{x}-\mathbf{y})^\top}{\|\mathbf{x}-\mathbf{y}\|_2^2}$. Let $\mathbf{x}_1,\cdots,\mathbf{x}_n$ be n independent observations of random vector $\mathbf{x} \sim EC_p(\mu,\mathbf{\Sigma},\xi)$. Suppose that n is even and let $\mathbf{K} \triangleq \mathbb{E}[S(\mathbf{x})S(\mathbf{x})^\top]$ and $\hat{\mathbf{K}} \triangleq \frac{2}{n}\sum_{i=1}^{n/2}k(\mathbf{x}_{2i-1},\mathbf{x}_{2i})$. For any $\mathbf{v}$ such that $\|\mathbf{v}\|_2 = 1$, if the following inequality holds,*

$$\mathbb{E}[\exp(t[(\mathbf{v}^\top S(\mathbf{x}))^2 - \mathbf{v}^\top\mathbf{K}\mathbf{v}])] \leq \exp(\eta t^2),\ \textit{for}\ t \leq \frac{c_0}{\sqrt{\eta}}, \tag{A-3}$$

where c_0 is a constant and $\eta > 0$ only depends on the eigenvalues of $\mathbf{\Sigma}$, then the following holds with probability at least $1 - \frac{s^{-10}}{T}$,

$$\sup_{\|\mathbf{v}\|_2=1,\|\mathbf{v}\|_0=s}|\mathbf{v}^\top(\hat{\mathbf{K}} - \mathbf{K})\mathbf{v}| \leq c\eta^{\frac{1}{2}}\sqrt{\frac{s\log p + \log T}{n}},$$

for parameter T and constant c.

Proof. The proof is similar to the proof of Lemma 8.2 in (Han & Liu, 2013). For simplicity, the constants may vary from line to line. Suppose that $\mathcal{A}$ is a $1/3$ cover of $\mathcal{S}_s \triangleq \{\mathbf{v} \in \mathbb{R}^s : \|\mathbf{v}\|_2 = 1\}$. As shown in the proof of Theorem A-2, we know that for any symmetric matrix $\mathbf{X}$, $\sup_{\mathbf{v}\in\mathcal{S}_s} |\mathbf{v}^\top \mathbf{X}\mathbf{v}| \le \frac{9}{2}\sup_{\mathbf{v}\in\mathcal{A}} |\mathbf{v}^\top \mathbf{X}\mathbf{v}|$. Let $\mathcal{F}_s$ be a subset of $\{1, 2, \cdots, p\}$ with cardinality $|\mathcal{F}_s| = s$. For square matrix $\mathbf{X}$, $\mathbf{X}_{\mathcal{F}_s}$ denotes the submatrix of $\mathbf{X}$ determined by the set $\mathcal{F}_s$, i.e., extracting the entries of $\mathbf{X}$ whose row and column indices are in $\mathcal{F}_s$. For vector $\mathbf{x}$, $\mathbf{x}_{\mathcal{F}_s}$ selects the entries of $\mathbf{x}$ indexed by $\mathcal{F}_s$.

Let $\beta = c\eta^{\frac{1}{2}}\sqrt{\frac{s\log p+\log T}{n}}$ for constant c. By the union bound, we have

$$\mathbb{P}\left[\sup_{\mathbf{v}\in\mathcal{S}_s}\sup_{\mathcal{F}_s\subseteq[p]} |\mathbf{v}^\top(\hat{\mathbf{K}}-\mathbf{K})_{\mathcal{F}_s}\mathbf{v}| > \frac{9}{2}\beta\right] \le \mathbb{P}\left[\sup_{\mathbf{v}\in\mathcal{A}}\sup_{\mathcal{F}_s\subseteq[p]} |\mathbf{v}^\top(\hat{\mathbf{K}}-\mathbf{K})_{\mathcal{F}_s}\mathbf{v}| > \beta\right]$$
$$\le 9^s p^s \mathbb{P}\left[|\mathbf{v}^\top(\hat{\mathbf{K}}-\mathbf{K})_{\mathcal{F}_s}\mathbf{v}| > c\eta^{\frac{1}{2}}\sqrt{\frac{s\log p+\log T}{n}},\ \text{for fixed } \mathbf{v}, \mathcal{F}_s\right].$$

Therefore, if we can show that for fixed $\mathbf{v}$ and $\mathcal{F}_s$,

$$\mathbb{P}\left[|\mathbf{v}^\top(\hat{\mathbf{K}}-\mathbf{K})_{\mathcal{F}_s}\mathbf{v}| > \beta\right] \le 2\exp(-\frac{cn\beta^2}{\eta}), \tag{A-4}$$

then we have

$$9^s p^s \mathbb{P}\left[|\mathbf{v}^\top(\hat{\mathbf{K}}-\mathbf{K})_{\mathcal{F}_s}\mathbf{v}| > c\eta^{\frac{1}{2}}\sqrt{\frac{s\log p+\log T}{n}},\ \text{for fixed } \mathbf{v}, \mathcal{F}_s\right] \le 2\exp(-c(s\log p+\log T)) \le \frac{s^{-10}}{T}$$

holds for a certain constant c, which implies this theorem. We now prove Inequality (A-4). Let $m = \frac{n}{2}$, for any $0 < t < \frac{c_0 m}{\sqrt{\eta}}$,

$$\mathbb{E}\left[\exp(t\mathbf{v}^\top(\hat{\mathbf{K}}-\mathbf{K})_{\mathcal{F}_s}\mathbf{v})\right] = \mathbb{E}\left[\exp\left(t\mathbf{v}^\top\left(\frac{1}{m}\sum_{i=1}^m k(\mathbf{x}_{2i-1},\mathbf{x}_{2i}) - \mathbf{K}\right)_{\mathcal{F}_s}\mathbf{v}\right)\right]$$
$$= \mathbb{E}\left[\exp\left(\frac{t}{m}\sum_{i=1}^m \mathbf{v}^\top\left(k(\mathbf{x}_{2i-1},\mathbf{x}_{2i}) - \mathbf{K}\right)_{\mathcal{F}_s}\mathbf{v}\right)\right] = \left(\mathbb{E}\left[\exp\left(\frac{t}{m}\cdot\mathbf{v}^\top\left(k(\mathbf{x}_{2i-1},\mathbf{x}_{2i}) - \mathbf{K}\right)_{\mathcal{F}_s}\mathbf{v}\right)\right]\right)^m$$
$$= \left(\mathbb{E}\left[\exp\left(\frac{t}{m}((\mathbf{v}^\top S(\mathbf{x})_{\mathcal{F}_s})^2 - \mathbf{v}^\top\mathbf{K}_{\mathcal{F}_s}\mathbf{v})\right)\right]\right)^m \le \exp(\frac{\eta t^2}{m}),$$

where the last inequality follows from (A-3). Suppose that $\beta \le 2c_0\sqrt{\eta}$ (this can be satisfied when n is large enough, i.e., $c\sqrt{\frac{s\log p+\log T}{n}} \le c_0$) and let $t = \frac{\beta m}{2\eta}$, then by the Markov inequality, we have

$$\mathbb{P}\left[\mathbf{v}^\top(\hat{\mathbf{K}}-\mathbf{K})_{\mathcal{F}_s}\mathbf{v} > \beta\right] \le \exp(-\frac{cn\beta^2}{\eta}).$$

By symmetry, we have the same bound for $\mathbb{P}\left[\mathbf{v}^\top(\hat{\mathbf{K}}-\mathbf{K})_{\mathcal{F}_s}\mathbf{v} < -\beta\right]$. Hence we obtain Inequality (A-4). □

Theorem A-6. *Let $k(\cdot) : \mathbb{R}^p \times \mathbb{R}^p \to \mathbb{R}^{p\times p}$ be a matrix value function defined as $k(\mathbf{x},\mathbf{y}) = \frac{(\mathbf{x}-\mathbf{y})(\mathbf{x}-\mathbf{y})^\top}{\|\mathbf{x}-\mathbf{y}\|_2^2}$. Let $\mathbf{x}_1, \cdots, \mathbf{x}_n$ be n independent observations of random vector $\mathbf{x} \sim EC_p(\mu, \mathbf{\Sigma}, \xi)$. Suppose that n is even. Let $\mathbf{K} \triangleq \mathbb{E}[S(\mathbf{x})S(\mathbf{x})^\top]$, $\hat{\mathbf{K}} \triangleq \frac{2}{n}\sum_{i=1}^{n/2} k(\mathbf{x}_{2i-1},\mathbf{x}_{2i})$ and $q = rank(\mathbf{K})$, then there exists a universal constant c such that the following holds with probability at least $1 - \frac{s^{-10}}{T}$,*

$$\sup_{\|\mathbf{v}\|_2=1,\|\mathbf{v}\|_0=s} |\mathbf{v}^\top(\hat{\mathbf{K}}-\mathbf{K})\mathbf{v}| \le c\left(\min\left\{\frac{4\lambda_1(\mathbf{K})}{q\lambda_q(\mathbf{K})}, 1\right\} + \|\mathbf{K}\|_2\right)\sqrt{\frac{s\log p+\log T}{n}},$$

for parameter T.

Proof. Apply Theorem A-5 and follow the proofs of Lemma 8.3 and Theorem 4.2 in (Han & Liu, 2013). □

2. Proofs in Section 4

In this section, we show the theoretical performance guarantees of our algorithms regardless of the requirement that the initial solution $\mathbf{Q}_0$ satisfies $\|\mathbf{U}_{k,\perp}^\top \mathbf{Q}_0\| > \epsilon$. The results in the paper can be easily obtained from the following theorems.

2.1. Streaming Sparse PCA

Recall that each sample is independently drawn from the spike model $\mathbf{x}_t = \mathbf{A}\mathbf{z}_t + \mathbf{w}_t$ where $\mathbf{z}_t \sim \mathcal{N}(0, \mathbf{I}_d)$ and $\mathbf{w}_t \sim \mathcal{N}(0, \sigma^2 \mathbf{I}_p)$. The covariance matrix of $\mathbf{x}_t$ is denoted by $\mathbf{\Sigma}$, i.e., $\mathbf{\Sigma} = \mathbb{E}[\mathbf{x}_t\mathbf{x}_t^\top] = \mathbf{A}\mathbf{A}^\top + \sigma^2\mathbf{I}_p$. We let $\mathbf{\Sigma} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^\top$ be the singular value decomposition of $\mathbf{\Sigma}$, and let λ_k be the k^{th} largest eigenvalue of $\mathbf{\Sigma}$ which equals Λ_{kk}, and $\mathbf{\Lambda}_k$ be the diagonal matrix such that the first k diagonal entries of $\mathbf{\Lambda}_k$ are equal to $\lambda_1, \cdots, \lambda_k$, respectively, while the rest entries are zero. Let $\mathbf{U}_k \in \mathbb{R}^{p\times k}$ be the matrix consisting of the leading k eigenvectors of $\mathbf{\Sigma}$, i.e, the first k columns of $\mathbf{U}$, and $\mathbf{U}_{k,\perp}$ be the matrix consisting of the last $p-k$ columns of $\mathbf{U}$. By the assumption, we know that $\mathbf{U}_k$ is row sparse, i.e., $\|\mathbf{U}_k\|_{2,0} \le s$ where $\|\mathbf{U}_k\|_{2,0}$ is the number of non-zero rows in $\mathbf{U}_k$. In Algorithm 2, $\mathbf{Q}_\tau$ is the solution generated on the $\tau-1$ iteration, $\tilde{\mathbf{S}}_{\tau+1} = \frac{1}{B}\sum_{t=B\tau+1}^{B(\tau+1)} \mathbf{x}_t\mathbf{x}_t^\top \mathbf{Q}_\tau$, $\mathbf{S}_{\tau+1} = \text{Truncate}(\tilde{\mathbf{S}}_{\tau+1}, \gamma)$ and $\mathbf{S}_{\tau+1} = \mathbf{Q}_{\tau+1}\mathbf{R}_{\tau+1}$ where parameter γ controls the desired sparsity.

We denote the row support of $\mathbf{U}_k$, $\mathbf{Q}_\tau$ and $\mathbf{Q}_{\tau+1}$ by $\mathcal{S}$, $\mathcal{F}_\tau$ and $\mathcal{F}_{\tau+1}$, respectively, and denote $\frac{1}{B}\sum_{t=B\tau+1}^{B(\tau+1)} \mathbf{x}_t\mathbf{x}_t^\top$ by $\hat{\mathbf{\Sigma}}_\tau$. Let $\mathcal{F} = \mathcal{S} \cup \mathcal{F}_\tau \cup \mathcal{F}_{\tau+1}$, and $\mathbf{X}(i,j)$ be the $(i,j)^{th}$ entry of matrix $\mathbf{X}$. For a $p\times p$ squared matrix, e.g., $\hat{\mathbf{\Sigma}}_\tau$, we let $\hat{\mathbf{\Sigma}}_{\tau,\mathcal{F}}$ be the matrix whose $(i,j)^{th}$ entry equals $\hat{\mathbf{\Sigma}}_\tau(i,j)$ if $i,j \in \mathcal{F}$ or 0 otherwise. For a $p\times k$ matrix, e.g., $\tilde{\mathbf{S}}_{\tau+1}$, we let $\tilde{\mathbf{S}}_{\tau+1,\mathcal{F}}$ be the matrix satisfying that $\tilde{\mathbf{S}}_{\tau+1,\mathcal{F}}(i,j) = \tilde{\mathbf{S}}_{\tau+1}(i,j)$ if $i \in \mathcal{F}$ or $\tilde{\mathbf{S}}_{\tau+1,\mathcal{F}}(i,j) = 0$ otherwise. In other words, $\hat{\mathbf{\Sigma}}_{\tau,\mathcal{F}}$ selects the entries whose row index and column index are both in $\mathcal{F}$ while $\tilde{\mathbf{S}}_{\tau+1,\mathcal{F}}$ selects the entries whose row index is in $\mathcal{F}$. One can easily verify that $\tilde{\mathbf{S}}_{\tau+1,\mathcal{F}} = \hat{\mathbf{\Sigma}}_{\tau,\mathcal{F}}\mathbf{Q}_\tau$ and $\mathbf{S}_{\tau+1} = \text{Truncate}(\tilde{\mathbf{S}}_{\tau+1,\mathcal{F}}, \gamma)$. Let the QR decomposition of $\tilde{\mathbf{S}}_{\tau+1,\mathcal{F}}$ be $\tilde{\mathbf{S}}_{\tau+1,\mathcal{F}} = \mathbf{Q}_{\tau+1,\mathcal{F}}\mathbf{R}_{\tau+1,\mathcal{F}}$.

2.1.1. Proof of Theorem 1

Lemma A-3. *Let $\mathbf{W}_\tau = \hat{\mathbf{\Sigma}}_\tau - \mathbf{\Sigma}_\tau$ and $\xi_{s+2\gamma} = \sup_{\mathcal{F}:|\mathcal{F}|\le s+2\gamma} \|\mathbf{W}_{\tau,\mathcal{F}}\|_2$, then if $\tilde{\mathbf{S}}_{\tau+1,\mathcal{F}}$ has full column rank, we have*

$$\|\mathbf{U}_{k,\perp}^\top \mathbf{Q}_{\tau+1,\mathcal{F}}\|_2^2 \le \frac{[\lambda_{k+1}\|\mathbf{U}_{k,\perp}^\top\mathbf{Q}_\tau\|_2 + \xi_{s+2\gamma}]^2}{[\lambda_{k+1}\|\mathbf{U}_{k,\perp}^\top\mathbf{Q}_\tau\|_2 + \xi_{s+2\gamma}]^2 + [\lambda_k\sqrt{1-\|\mathbf{U}_{k,\perp}^\top\mathbf{Q}_\tau\|_2^2} - \xi_{s+2\gamma}]^2}.$$

Proof. By the definition of $\|\cdot\|_2$, there exists vector $\mathbf{v}$ so that the following holds,

$$\begin{aligned}\|\mathbf{U}_{k,\perp}^\top \mathbf{Q}_{\tau+1,\mathcal{F}}\|_2^2 =& \frac{\|\mathbf{U}_{k,\perp}^\top \mathbf{Q}_{\tau+1,\mathcal{F}}\mathbf{v}\|_2^2}{\|\mathbf{v}\|_2^2} = \frac{\|\mathbf{U}_{k,\perp}^\top \mathbf{Q}_{\tau+1,\mathcal{F}}\mathbf{R}_{\tau+1,\mathcal{F}}\tilde{\mathbf{v}}\|_2^2}{\|\mathbf{R}_{\tau+1,\mathcal{F}}\tilde{\mathbf{v}}\|_2^2}\\ =& \frac{\|\mathbf{U}_{k,\perp}^\top \tilde{\mathbf{S}}_{\tau+1,\mathcal{F}}\tilde{\mathbf{v}}\|_2^2}{\|\tilde{\mathbf{S}}_{\tau+1,\mathcal{F}}\tilde{\mathbf{v}}\|_2^2} = \frac{\|\mathbf{U}_{k,\perp}^\top \tilde{\mathbf{S}}_{\tau+1,\mathcal{F}}\tilde{\mathbf{v}}\|_2^2}{\|\mathbf{U}_k^\top \tilde{\mathbf{S}}_{\tau+1,\mathcal{F}}\tilde{\mathbf{v}}\|_2^2 + \|\mathbf{U}_{k,\perp}^\top \tilde{\mathbf{S}}_{\tau+1,\mathcal{F}}\tilde{\mathbf{v}}\|_2^2},\end{aligned}$$

where $\tilde{\mathbf{v}} = \frac{\mathbf{R}_{\tau+1,\mathcal{F}}^{-1}\mathbf{v}}{\|\mathbf{R}_{\tau+1,\mathcal{F}}^{-1}\mathbf{v}\|_2}$ ($\mathbf{R}_{\tau+1,\mathcal{F}}^{-1}$ exists since $\tilde{\mathbf{S}}_{\tau+1,\mathcal{F}}$ has full column rank).

We now bound $\|\mathbf{U}_{k,\perp}^\top \tilde{\mathbf{S}}_{\tau+1,\mathcal{F}}\tilde{\mathbf{v}}\|_2$ and $\|\mathbf{U}_k^\top \tilde{\mathbf{S}}_{\tau+1,\mathcal{F}}\tilde{\mathbf{v}}\|_2$. Since $\mathcal{S} \subseteq \mathcal{F}$, we have $\mathbf{\Sigma}_\mathcal{F} = \mathbf{\Sigma}$ and

$$\begin{aligned}\|\mathbf{U}_k^\top \tilde{\mathbf{S}}_{\tau+1,\mathcal{F}}\tilde{\mathbf{v}}\|_2 =& \|\mathbf{U}_k^\top \hat{\mathbf{\Sigma}}_{\tau,\mathcal{F}}\mathbf{Q}_\tau\tilde{\mathbf{v}}\|_2\\ =& \|\mathbf{U}_k^\top(\mathbf{\Sigma}_\mathcal{F} + \mathbf{W}_{\tau,\mathcal{F}})\mathbf{Q}_\tau\tilde{\mathbf{v}}\|_2\\ \ge& \|\mathbf{U}_k^\top\mathbf{\Sigma}_\mathcal{F}\mathbf{Q}_\tau\tilde{\mathbf{v}}\|_2 - \|\mathbf{U}_k^\top\mathbf{W}_{\tau,\mathcal{F}}\mathbf{Q}_\tau\tilde{\mathbf{v}}\|_2\\ \ge& \|\mathbf{\Lambda}_k\mathbf{U}_k^\top\mathbf{Q}_\tau\tilde{\mathbf{v}}\|_2 - \sup_{\mathcal{F}:|\mathcal{F}|\le s+2\gamma}\|\mathbf{W}_{\tau,\mathcal{F}}\|_2\\ \ge& \lambda_k\|\mathbf{U}_k^\top\mathbf{Q}_\tau\tilde{\mathbf{v}}\|_2 - \sup_{\mathcal{F}:|\mathcal{F}|\le s+2\gamma}\|\mathbf{W}_{\tau,\mathcal{F}}\|_2\\ \ge& \lambda_k\sqrt{1-\|\mathbf{U}_{k,\perp}^\top\mathbf{Q}_\tau\|_2^2} - \sup_{\mathcal{F}:|\mathcal{F}|\le s+2\gamma}\|\mathbf{W}_{\tau,\mathcal{F}}\|_2.\end{aligned}$$

Similarly,

$$\begin{aligned}\|\mathbf{U}_{k,\perp}^\top \tilde{\mathbf{S}}_{\tau+1,\mathcal{F}}\tilde{\mathbf{v}}\|_2 =&\|\mathbf{U}_{k,\perp}^\top \hat{\boldsymbol{\Sigma}}_{\tau,\mathcal{F}}\mathbf{Q}_\tau\tilde{\mathbf{v}}\|_2\\ =&\|\mathbf{U}_{k,\perp}^\top (\boldsymbol{\Sigma}_{\mathcal{F}}+\mathbf{W}_{\tau,\mathcal{F}})\mathbf{Q}_\tau\tilde{\mathbf{v}}\|_2\\ \le&\|\mathbf{U}_{k,\perp}^\top \boldsymbol{\Sigma}_{\mathcal{F}}\mathbf{Q}_\tau\tilde{\mathbf{v}}\|_2+\|\mathbf{U}_{k,\perp}^\top \mathbf{W}_{\tau,\mathcal{F}}\mathbf{Q}_\tau\tilde{\mathbf{v}}\|_2\\ \le&\|(\boldsymbol{\Lambda}-\boldsymbol{\Lambda}_k)\mathbf{U}_{k,\perp}^\top\mathbf{Q}_\tau\tilde{\mathbf{v}}\|_2+\sup_{\mathcal{F}:|\mathcal{F}|\le s+2\gamma}\|\mathbf{W}_{\tau,\mathcal{F}}\|_2\\ \le&\lambda_{k+1}\|\mathbf{U}_{k,\perp}^\top\mathbf{Q}_\tau\|_2+\sup_{\mathcal{F}:|\mathcal{F}|\le s+2\gamma}\|\mathbf{W}_{\tau,\mathcal{F}}\|_2,\end{aligned}$$

Hence we have

$$\begin{aligned}&\|\mathbf{U}_{k,\perp}^\top\mathbf{Q}_{\tau+1,\mathcal{F}}\|_2^2\\ \le&\frac{[\lambda_{k+1}\|\mathbf{U}_{k,\perp}^\top\mathbf{Q}_\tau\|_2+\sup_{\mathcal{F}:|\mathcal{F}|\le s+2\gamma}\|\mathbf{W}_{\tau,\mathcal{F}}\|_2]^2}{[\lambda_{k+1}\|\mathbf{U}_{k,\perp}^\top\mathbf{Q}_\tau\|_2+\sup_{\mathcal{F}:|\mathcal{F}|\le s+2\gamma}\|\mathbf{W}_{\tau,\mathcal{F}}\|_2]^2+[\lambda_k\sqrt{1-\|\mathbf{U}_{k,\perp}^\top\mathbf{Q}_\tau\|_2^2}-\sup_{\mathcal{F}:|\mathcal{F}|\le s+2\gamma}\|\mathbf{W}_{\tau,\mathcal{F}}\|_2]^2}.\end{aligned}$$

Let $\xi_{s+2\gamma}=\sup_{\mathcal{F}:|\mathcal{F}|\le s+2\gamma}\|\mathbf{W}_{\tau,\mathcal{F}}\|_2$, then

$$\|\mathbf{U}_{k,\perp}^\top\mathbf{Q}_{\tau+1,\mathcal{F}}\|_2^2\le\frac{[\lambda_{k+1}\|\mathbf{U}_{k,\perp}^\top\mathbf{Q}_\tau\|_2+\xi_{s+2\gamma}]^2}{[\lambda_{k+1}\|\mathbf{U}_{k,\perp}^\top\mathbf{Q}_\tau\|_2+\xi_{s+2\gamma}]^2+[\lambda_k\sqrt{1-\|\mathbf{U}_{k,\perp}^\top\mathbf{Q}_\tau\|_2^2}-\xi_{s+2\gamma}]^2}.$$

Hence we obtain this lemma. □

Lemma A-4. *Let $\mathbf{Q}_{\tau+1,\mathcal{F},\perp}$ be an orthonormal matrix such that matrix $[\mathbf{Q}_{\tau+1,\mathcal{F}},\mathbf{Q}_{\tau+1,\mathcal{F},\perp}]$ is orthogonal. Let $\mathbf{W}_\tau=\hat{\boldsymbol{\Sigma}}_\tau-\boldsymbol{\Sigma}_\tau$ and $\xi_{s+2\gamma}=\sup_{\mathcal{F}:|\mathcal{F}|\le s+2\gamma}\|\mathbf{W}_{\tau,\mathcal{F}}\|_2$, then if $\gamma\ge s$ and*

$$\frac{\sqrt{k}[\lambda_{k+1}\|\mathbf{U}_{k,\perp}^\top\mathbf{Q}_\tau\|_2+\xi_{s+2\gamma}]}{\lambda_k\sqrt{1-\|\mathbf{U}_{k,\perp}^\top\mathbf{Q}_\tau\|_2^2}}<\frac{1}{2+\sqrt{2}},\tag{A-5}$$

we have

$$\|\mathbf{Q}_{\tau+1,\mathcal{F},\perp}^\top(\mathbf{Q}_{\tau+1}-\mathbf{Q}_{\tau+1,\mathcal{F}})\|_2\le\frac{k[\lambda_{k+1}\|\mathbf{U}_{k,\perp}^\top\mathbf{Q}_\tau\|_2+\xi_{s+2\gamma}]}{\lambda_k\sqrt{1-\|\mathbf{U}_{k,\perp}^\top\mathbf{Q}_\tau\|_2^2}-(2+\sqrt{2})\sqrt{k}[\lambda_{k+1}\|\mathbf{U}_{k,\perp}^\top\mathbf{Q}_\tau\|_2+\xi_{s+2\gamma}]}.$$

Proof. Suppose that $\mathbf{Q}_\tau=\mathbf{U}_k\mathbf{A}+\mathbf{U}_{k,\perp}\mathbf{B}$, then

$$\begin{aligned}\tilde{\mathbf{S}}_{\tau+1,\mathcal{F}}=&\hat{\boldsymbol{\Sigma}}_{\tau,\mathcal{F}}\mathbf{Q}_\tau=\boldsymbol{\Sigma}_{\mathcal{F}}\mathbf{Q}_\tau+\mathbf{W}_{\tau,\mathcal{F}}\mathbf{Q}_\tau\\ =&\mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^\top(\mathbf{U}_k\mathbf{A}+\mathbf{U}_{k,\perp}\mathbf{B})+\mathbf{W}_{\tau,\mathcal{F}}\mathbf{Q}_\tau\\ =&\mathbf{U}_k\boldsymbol{\Lambda}_k\mathbf{A}+\mathbf{U}(\boldsymbol{\Lambda}-\boldsymbol{\Lambda}_k)\mathbf{B}+\mathbf{W}_{\tau,\mathcal{F}}\mathbf{Q}_\tau.\end{aligned}$$

Denote $\mathcal{F}\setminus\mathcal{S}$ by $\mathcal{S}^c$. Let $\Delta\tilde{\mathbf{S}}_{\tau+1,\mathcal{F}}$ be the matrix whose the i^{th} row equals the one of $\tilde{\mathbf{S}}_{\tau+1,\mathcal{F}}$ if $i\in\mathcal{S}^c$ or 0 otherwise, i.e., $\Delta\tilde{\mathbf{S}}_{\tau+1,\mathcal{F}}=\tilde{\mathbf{S}}_{\tau+1,\mathcal{F}}-\tilde{\mathbf{S}}_{\tau+1,\mathcal{S}}$. Since $\Delta\tilde{\mathbf{S}}_{\tau+1,\mathcal{F}}\in\mathbb{R}^{p\times k}$, the row support of $\mathbf{U}_k$ is $\mathcal{S}$ and $\|\mathbf{U}_{k,\perp}^\top\mathbf{Q}_\tau\|_2=\|\mathbf{U}_{k,\perp}^\top\mathbf{U}_{k,\perp}\mathbf{B}\|_2=\|\mathbf{B}\|_2$, we have

$$\begin{aligned}\|\Delta\tilde{\mathbf{S}}_{\tau+1,\mathcal{F}}\|_F\le&\sqrt{k}\|\Delta\tilde{\mathbf{S}}_{\tau+1,\mathcal{F}}\|_2\\ \le&\sqrt{k}\|\mathbf{U}(\boldsymbol{\Lambda}-\boldsymbol{\Lambda}_k)\mathbf{B}+\mathbf{W}_{\tau,\mathcal{F}}\mathbf{Q}_\tau\|_2\\ \le&\sqrt{k}[\lambda_{k+1}\|\mathbf{B}\|_2+\sup_{\mathcal{F}:|\mathcal{F}|\le s+2\gamma}\|\mathbf{W}_{\tau,\mathcal{F}}\|_2]\\ =&\sqrt{k}[\lambda_{k+1}\|\mathbf{U}_{k,\perp}^\top\mathbf{Q}_\tau\|_2+\sup_{\mathcal{F}:|\mathcal{F}|\le s+2\gamma}\|\mathbf{W}_{\tau,\mathcal{F}}\|_2].\end{aligned}$$

Recall that the row support of $\mathbf{S}_{\tau+1}$ is $\mathcal{F}_{\tau+1}$, and $\mathbf{S}_{\tau+1} = \text{Truncate}(\tilde{\mathbf{S}}_{\tau+1,\mathcal{F}}, \gamma)$ where the truncation operator sorts the l_2-norms of the row vectors of $\tilde{\mathbf{S}}_{\tau+1,\mathcal{F}}$ and then selects γ rows with the largest l_2-norms. Since $\gamma \geq s$, or equivalently, $|\mathcal{F}_{\tau+1}| \geq |\mathcal{S}|$, we have

$$\begin{aligned}
\|\mathbf{S}_{\tau+1} - \tilde{\mathbf{S}}_{\tau+1,\mathcal{F}}\|_F^2 =& \|\tilde{\mathbf{S}}_{\tau+1,\mathcal{F}_{\tau+1}} - \tilde{\mathbf{S}}_{\tau+1,\mathcal{F}}\|_F^2 = \|\tilde{\mathbf{S}}_{\tau+1,\mathcal{F}_{\tau+1}} - \tilde{\mathbf{S}}_{\tau+1,\mathcal{S}\cup\mathcal{S}^c}\|_F^2 \\
=& \|\tilde{\mathbf{S}}_{\tau+1,(\mathcal{S}\setminus\mathcal{F}_{\tau+1})\cup(\mathcal{S}^c\setminus\mathcal{F}_{\tau+1})}\|_F^2 = \|\tilde{\mathbf{S}}_{\tau+1,\mathcal{S}\setminus\mathcal{F}_{\tau+1}}\|_F^2 + \|\tilde{\mathbf{S}}_{\tau+1,\mathcal{S}^c\setminus\mathcal{F}_{\tau+1}}\|_F^2 \\
\leq& \|\tilde{\mathbf{S}}_{\tau+1,\mathcal{F}_{\tau+1}\setminus\mathcal{S}}\|_F^2 + \|\tilde{\mathbf{S}}_{\tau+1,\mathcal{S}^c\setminus\mathcal{F}_{\tau+1}}\|_F^2 \leq \|\tilde{\mathbf{S}}_{\tau+1,\mathcal{S}^c}\|_F^2 = \|\Delta\tilde{\mathbf{S}}_{\tau+1,\mathcal{F}}\|_F^2,
\end{aligned}$$

where the first inequality follows from the definition of the truncation operator and the fact that $|\mathcal{S} \setminus \mathcal{F}_{\tau+1}| \leq |\mathcal{F}_{\tau+1} \setminus \mathcal{S}|$. Hence

$$\|\mathbf{S}_{\tau+1} - \tilde{\mathbf{S}}_{\tau+1,\mathcal{F}}\|_F \leq \sqrt{k}[\lambda_{k+1}\|\mathbf{U}_{k,\perp}^\top \mathbf{Q}_\tau\|_2 + \sup_{\mathcal{F}:|\mathcal{F}|\leq s+2\gamma} \|\mathbf{W}_{\tau,\mathcal{F}}\|_2]. \tag{A-6}$$

We now bound the minimal singular value of $\tilde{\mathbf{S}}_{\tau+1,\mathcal{F}}$, which is denoted by $\delta_{\min}(\tilde{\mathbf{S}}_{\tau+1,\mathcal{F}})$. By applying the Weyl's Inequalities,

$$\begin{aligned}
\delta_{\min}(\tilde{\mathbf{S}}_{\tau+1,\mathcal{F}}) \geq& \delta_{\min}(\mathbf{U}_k\mathbf{\Lambda}_k\mathbf{A}) - \|\mathbf{U}(\mathbf{\Lambda} - \mathbf{\Lambda}_k)\mathbf{B} + \mathbf{W}_{\tau,\mathcal{F}}\mathbf{Q}_\tau\|_2 \\
\geq& \sqrt{\delta_{\min}(\mathbf{A}^\top\mathbf{\Lambda}_k^2\mathbf{A})} - \lambda_{k+1}\|\mathbf{B}\|_2 - \sup_{\mathcal{F}:|\mathcal{F}|\leq s+2\gamma} \|\mathbf{W}_{\tau,\mathcal{F}}\|_2.
\end{aligned}$$

Note that $\sqrt{\delta_{\min}(\mathbf{A}^\top\mathbf{\Lambda}_k^2\mathbf{A})} \geq \lambda_k\sqrt{\delta_{\min}(\mathbf{A}^\top\mathbf{A})} \geq \lambda_k\sqrt{1 - \|\mathbf{B}\|_2^2}$ where the last inequality follows from that $\mathbf{A}^\top\mathbf{A} + \mathbf{B}^\top\mathbf{B} = \mathbf{I}_k$ which implies that $\delta_{\min}(\mathbf{A}^\top\mathbf{A}) \geq 1 - \|\mathbf{B}^\top\mathbf{B}\|_2$. Thus,

$$\frac{1}{\|\tilde{\mathbf{S}}_{\tau+1,\mathcal{F}}^\dagger\|_2} = \delta_{\min}(\tilde{\mathbf{S}}_{\tau+1,\mathcal{F}}) \geq \lambda_k\sqrt{1 - \|\mathbf{U}_{k,\perp}^\top\mathbf{Q}_\tau\|_2^2} - \lambda_{k+1}\|\mathbf{U}_{k,\perp}^\top\mathbf{Q}_\tau\|_2 - \sup_{\mathcal{F}:|\mathcal{F}|\leq s+2\gamma} \|\mathbf{W}_{\tau,\mathcal{F}}\|_2, \tag{A-7}$$

Note that when Inequality (A-5) holds, the right hand side of (A-7) must be positive. One can also verify that when Inequalities (A-5), (A-7) and (A-6) hold, $\|\tilde{\mathbf{S}}_{\tau+1,\mathcal{F}}^\dagger\|_2\|\mathbf{S}_{\tau+1} - \tilde{\mathbf{S}}_{\tau+1,\mathcal{F}}\|_F < \sqrt{2} - 1$. Then from Theorem A-1, we have

$$\begin{aligned}
\|\mathbf{Q}_{\tau+1,\mathcal{F},\perp}^\top(\mathbf{Q}_{\tau+1} - \mathbf{Q}_{\tau+1,\mathcal{F}})\|_F \leq& \frac{\kappa_2(\tilde{\mathbf{S}}_{\tau+1,\mathcal{F}})\frac{\|\mathbf{Q}_{\tau+1,\mathcal{F},\perp}^\top(\mathbf{S}_{\tau+1}-\tilde{\mathbf{S}}_{\tau+1,\mathcal{F}})\|_F}{\|\tilde{\mathbf{S}}_{\tau+1,\mathcal{F}}\|_2}}{1 - (1+\sqrt{2})\kappa_2(\tilde{\mathbf{S}}_{\tau+1,\mathcal{F}})\frac{\|\mathbf{S}_{\tau+1}-\tilde{\mathbf{S}}_{\tau+1,\mathcal{F}}\|_F}{\|\tilde{\mathbf{S}}_{\tau+1,\mathcal{F}}\|_2}} \\
=& \frac{\|\mathbf{Q}_{\tau+1,\mathcal{F},\perp}^\top(\mathbf{S}_{\tau+1} - \tilde{\mathbf{S}}_{\tau+1,\mathcal{F}})\|_F}{\frac{1}{\|\tilde{\mathbf{S}}_{\tau+1,\mathcal{F}}^\dagger\|_2} - (1+\sqrt{2})\|\mathbf{S}_{\tau+1} - \tilde{\mathbf{S}}_{\tau+1,\mathcal{F}}\|_F} \\
\leq& \frac{\sqrt{k}\|\mathbf{Q}_{\tau+1,\mathcal{F},\perp}^\top(\mathbf{S}_{\tau+1} - \tilde{\mathbf{S}}_{\tau+1,\mathcal{F}})\|_2}{\frac{1}{\|\tilde{\mathbf{S}}_{\tau+1,\mathcal{F}}^\dagger\|_2} - (1+\sqrt{2})\|\mathbf{S}_{\tau+1} - \tilde{\mathbf{S}}_{\tau+1,\mathcal{F}}\|_F} \\
\leq& \frac{\sqrt{k}\|\tilde{\mathbf{S}}_{\tau+1,\mathcal{F}}^\dagger\|_2\|\mathbf{S}_{\tau+1} - \tilde{\mathbf{S}}_{\tau+1,\mathcal{F}}\|_F}{1 - (1+\sqrt{2})\|\tilde{\mathbf{S}}_{\tau+1,\mathcal{F}}^\dagger\|_2\|\mathbf{S}_{\tau+1} - \tilde{\mathbf{S}}_{\tau+1,\mathcal{F}}\|_F}.
\end{aligned}$$

Let $\xi_{s+2\gamma} = \sup_{\mathcal{F}:|\mathcal{F}|\leq s+2\gamma} \|\mathbf{W}_{\tau,\mathcal{F}}\|_2$, from Inequalities (A-6) and (A-7), we have

$$\begin{aligned}
&\|\mathbf{Q}_{\tau+1,\mathcal{F},\perp}^\top(\mathbf{Q}_{\tau+1} - \mathbf{Q}_{\tau+1,\mathcal{F}})\|_2 \\
\leq& \|\mathbf{Q}_{\tau+1,\mathcal{F},\perp}^\top(\mathbf{Q}_{\tau+1} - \mathbf{Q}_{\tau+1,\mathcal{F}})\|_F \\
\leq& \frac{k[\lambda_{k+1}\|\mathbf{U}_{k,\perp}^\top\mathbf{Q}_\tau\|_2 + \xi_{s+2\gamma}]}{\lambda_k\sqrt{1 - \|\mathbf{U}_{k,\perp}^\top\mathbf{Q}_\tau\|_2^2} - \lambda_{k+1}\|\mathbf{U}_{k,\perp}^\top\mathbf{Q}_\tau\|_2 - \xi_{s+2\gamma} - (1+\sqrt{2})\sqrt{k}[\lambda_{k+1}\|\mathbf{U}_{k,\perp}^\top\mathbf{Q}_\tau\|_2 + \xi_{s+2\gamma}]} \\
\leq& \frac{k[\lambda_{k+1}\|\mathbf{U}_{k,\perp}^\top\mathbf{Q}_\tau\|_2 + \xi_{s+2\gamma}]}{\lambda_k\sqrt{1 - \|\mathbf{U}_{k,\perp}^\top\mathbf{Q}_\tau\|_2^2} - (2+\sqrt{2})\sqrt{k}[\lambda_{k+1}\|\mathbf{U}_{k,\perp}^\top\mathbf{Q}_\tau\|_2 + \xi_{s+2\gamma}]}.
\end{aligned}$$

Hence we obtain this lemma. □

Lemma A-5. *For $\eta > 0$, $0 < \epsilon < 1$, $T > 0$ and $\gamma \geq s$, let $\mu \triangleq \frac{(k+1)\lambda_{k+1}+2\eta\lambda_k}{\lambda_k}$ and $f(\mu,\eta,k) \triangleq \max\{\frac{(2+\sqrt{2})\mu}{\sqrt{k}}, \frac{\eta}{k}\}$, then if the block size B satisfies*

$$B \geq \frac{ck^2\lambda_1^2[(s+2\gamma)\log p + \log T]}{\epsilon^2\eta^2\lambda_k^2} \tag{A-8}$$

and the initial solution $\mathbf{Q}_0$ satisfies

$$\max\{\|\mathbf{U}_{k,\perp}^\top \mathbf{Q}_0\|_2, \epsilon\} < \frac{-\mu f(\mu,\eta,k) + \sqrt{1 + f(\mu,\eta,k)^2 - \mu^2}}{1 + f(\mu,\eta,k)^2}, \tag{A-9}$$

the following inequality holds with probability at least $1 - \frac{s^{-10}}{T}$:

$$\|\mathbf{U}_{k,\perp}^\top \mathbf{Q}_{\tau+1}\|_2 \leq \frac{\mu}{\sqrt{1 - \max\{\|\mathbf{U}_{k,\perp}^\top \mathbf{Q}_0\|_2, \epsilon\}^2} - f(\mu,\eta,k)\max\{\|\mathbf{U}_{k,\perp}^\top \mathbf{Q}_0\|_2, \epsilon\}} \max\{\|\mathbf{U}_{k,\perp}^\top \mathbf{Q}_\tau\|_2, \epsilon\}. \tag{A-10}$$

Proof. We use the same notation as Lemma A-4. We let $\theta_\tau \triangleq |\mathbf{U}_{k,\perp}^\top \mathbf{Q}_\tau\|_2$ and let $\phi_0 \triangleq \max\{\theta_0, \epsilon\}$. Suppose that $\mathbf{Q}_{\tau+1} = \mathbf{Q}_{\tau+1,\mathcal{F}}\mathbf{A} + \mathbf{Q}_{\tau+1,\mathcal{F},\perp}\mathbf{B}$, then

$$\begin{aligned}
\|\mathbf{U}_{k,\perp}^\top \mathbf{Q}_{\tau+1}\|_2 &\leq \|\mathbf{U}_{k,\perp}^\top \mathbf{Q}_{\tau+1,\mathcal{F}}\mathbf{A}\|_2 + \|\mathbf{U}_{k,\perp}^\top \mathbf{Q}_{\tau+1,\mathcal{F},\perp}\mathbf{B}\|_2 \\
&\leq \|\mathbf{U}_{k,\perp}^\top \mathbf{Q}_{\tau+1,\mathcal{F}}\|_2 + \|\mathbf{B}\|_2 \\
&= \|\mathbf{U}_{k,\perp}^\top \mathbf{Q}_{\tau+1,\mathcal{F}}\|_2 + \|\mathbf{Q}_{\tau+1,\mathcal{F},\perp}^\top (\mathbf{Q}_{\tau+1} - \mathbf{Q}_{\tau+1,\mathcal{F}})\|_2.
\end{aligned}$$

Since B satisfies (A-8), Theorem A-2 implies that $\xi_{s+2\gamma} \leq \frac{\eta\epsilon\lambda_k}{k}$ holds with probability at least $1 - \frac{s^{-10}}{T}$. One can easily verify that Inequality (A-9) implies that there exist α and β such that $\alpha < 1$, $\eta\beta < 1$, $\frac{1}{k}\phi_0 \leq \beta\sqrt{1-\phi_0^2}$ and $\frac{2+\sqrt{2}}{\sqrt{k}}[(k+1)\lambda_{k+1} + 2\eta\lambda_k]\phi_0 \leq \alpha\lambda_k\sqrt{1-\phi_0^2}$, e.g.,

$$\alpha = \frac{(2+\sqrt{2})[(k+1)\lambda_{k+1} + 2\eta\lambda_k]}{\sqrt{k}\lambda_k} \cdot \frac{\phi_0}{\sqrt{1-\phi_0^2}} \text{ and } \beta = \frac{\phi_0}{k\sqrt{1-\phi_0^2}}.$$

This theorem can be proved by mathematical induction. We suppose $\theta_\tau \leq \phi_0$ and then prove $\theta_{\tau+1} \leq \phi_0$ by showing that Inequality (A-10) holds. Since $\theta_\tau, \epsilon \leq \phi_0$, we have

$$\frac{\sqrt{k}[\lambda_{k+1}\theta_\tau + \xi_{s+2\gamma}]}{\lambda_k\sqrt{1-\theta_\tau^2}} \leq \frac{\sqrt{k}[\lambda_{k+1}\phi_0 + \frac{\eta\phi_0\lambda_k}{k}]}{\lambda_k\sqrt{1-\phi_0^2}} \leq \frac{\alpha}{2+\sqrt{2}} < \frac{1}{2+\sqrt{2}}.$$

As shown in Lemma A-4, this inequality implies that $\tilde{\mathbf{S}}_{\tau+1,\mathcal{F}}$ has full column rank. Hence we can apply Lemma A-3 and Lemma A-4,

$$\|\mathbf{U}_{k,\perp}^\top \mathbf{Q}_{\tau+1,\mathcal{F}}\|_2^2 \leq \frac{[\lambda_{k+1}\theta_\tau + \xi_{s+2\gamma}]^2}{[\lambda_{k+1}\theta_\tau + \xi_{s+2\gamma}]^2 + [\lambda_k\sqrt{1-\theta_\tau^2} - \xi_{s+2\gamma}]^2},$$

$$\|\mathbf{Q}_{\tau+1,\mathcal{F},\perp}^\top (\mathbf{Q}_{\tau+1} - \mathbf{Q}_{\tau+1,\mathcal{F}})\|_2 \leq \frac{k[\lambda_{k+1}\theta_\tau + \xi_{s+2\gamma}]}{\lambda_k\sqrt{1-\theta_\tau^2} - (2+\sqrt{2})\sqrt{k}[\lambda_{k+1}\theta_\tau + \xi_{s+2\gamma}]}.$$

We first consider the case where $\theta_\tau \geq \epsilon$. From the definition of α and β above, we have

$$\|\mathbf{U}_{k,\perp}^\top \mathbf{Q}_{\tau+1,\mathcal{F}}\|_2^2 \leq \frac{(\lambda_{k+1} + \frac{\eta\lambda_k}{k})^2\theta_\tau^2}{(\lambda_{k+1} + \frac{\eta\lambda_k}{k})^2\theta_\tau^2 + [\lambda_k\sqrt{1-\theta_\tau^2} - \frac{\eta\lambda_k}{k}\theta_\tau]^2} \leq \left[\frac{(\lambda_{k+1} + \frac{\eta\lambda_k}{k})\theta_\tau}{(1-\eta\beta)\lambda_k\sqrt{1-\theta_\tau^2}}\right]^2,$$

$$\|\mathbf{Q}_{\tau+1,\mathcal{F},\perp}^\top (\mathbf{Q}_{\tau+1} - \mathbf{Q}_{\tau+1,\mathcal{F}})\|_2 \leq \frac{(k\lambda_{k+1} + \eta\lambda_k)\theta_\tau}{\lambda_k\sqrt{1-\theta_\tau^2} - \frac{2+\sqrt{2}}{\sqrt{k}}(k\lambda_{k+1} + \eta\lambda_k)\theta_\tau} \leq \frac{(k\lambda_{k+1} + \eta\lambda_k)\theta_\tau}{(1-\alpha)\lambda_k\sqrt{1-\theta_\tau^2}},$$

which implies that

$$\theta_{\tau+1} \triangleq \|\mathbf{U}_{k,\perp}^\top \mathbf{Q}_{\tau+1}\|_2 \leq \frac{(\lambda_{k+1} + \frac{\eta\lambda_k}{k})\theta_\tau}{(1-\eta\beta)\lambda_k\sqrt{1-\theta_\tau^2}} + \frac{(k\lambda_{k+1} + \eta\lambda_k)\theta_\tau}{(1-\alpha)\lambda_k\sqrt{1-\theta_\tau^2}} \leq \frac{(k+1)\lambda_{k+1} + 2\eta\lambda_k}{(1-\max\{\alpha,\eta\beta\})\lambda_k\sqrt{1-\phi_0^2}}\theta_\tau,$$

where the last inequality holds because $\theta_\tau \le \phi_0$. Denote $\frac{(k+1)\lambda_{k+1}+2\eta\lambda_k}{\lambda_k}$ by μ, then we have

$$\theta_{\tau+1} \le \frac{\mu}{\sqrt{1-\phi_0^2} - \max\{\frac{(2+\sqrt{2})\mu}{\sqrt{k}}, \frac{\eta}{k}\}\phi_0}\theta_\tau.$$

Clearly, when $0 < \frac{\mu}{\sqrt{1-\phi_0^2}-\max\{\frac{(2+\sqrt{2})\mu}{\sqrt{k}},\frac{\eta}{k}\}\phi_0} < 1$, or equivalently,

$$\phi_0 < \frac{-\mu\max\{\frac{(2+\sqrt{2})\mu}{\sqrt{k}}, \frac{\eta}{k}\} + \sqrt{1+\max\{\frac{(2+\sqrt{2})\mu}{\sqrt{k}}, \frac{\eta}{k}\}^2 - \mu^2}}{1+\max\{\frac{(2+\sqrt{2})\mu}{\sqrt{k}}, \frac{\eta}{k}\}^2},$$

θ_τ will decrease in each iteration, i.e., $\theta_{\tau+1} < \theta_\tau \le \phi_0$. For the case where $\theta_\tau < \epsilon \le \phi_0$, one can follow the same proof above and obtain

$$\theta_{\tau+1} \le \frac{\mu}{\sqrt{1-\phi_0^2} - \max\{\frac{(2+\sqrt{2})\mu}{\sqrt{k}}, \frac{\eta}{k}\}\phi_0}\epsilon.$$

Let $f(\mu,\eta,k) \triangleq \max\{\frac{(2+\sqrt{2})\mu}{\sqrt{k}}, \frac{\eta}{k}\}$, we conclude that

$$\theta_{\tau+1} \le \frac{\mu}{\sqrt{1-\phi_0^2} - f(\mu,\eta,k)\phi_0}\max\{\theta_\tau, \epsilon\}$$

when $\phi_0 < \frac{-\mu f(\mu,\eta,k)+\sqrt{1+f(\mu,\eta,k)^2-\mu^2}}{1+f(\mu,\eta,k)^2}$. Hence this lemma holds. □

Theorem A-7. *For $\eta > 0$, $0 < \epsilon < 1$, and $\gamma \ge s$, let $\mu \triangleq \frac{(k+1)\lambda_{k+1}+2\eta\lambda_k}{\lambda_k}$, $f(\mu,\eta,k) \triangleq \max\{\frac{(2+\sqrt{2})\mu}{\sqrt{k}}, \frac{\eta}{k}\}$ and $g(\mathbf{Q},\epsilon) \triangleq \max\{\|\mathbf{U}_{k,\perp}^\top\mathbf{Q}\|_2, \epsilon\}$. If the initial solution $\mathbf{Q}_0$ satisfies that*

$$g(\mathbf{Q}_0,\epsilon) < \frac{1-\mu^2}{\sqrt{1-\mu^2}+(\mu+1)f(\mu,\eta,k)}, \tag{A-11}$$

and the following two inequalities hold

$$T \ge \frac{\log(\epsilon/g(\mathbf{Q}_0,\epsilon))}{\log\left[\mu/\left(\sqrt{1-g(\mathbf{Q}_0,\epsilon)^2} - f(\mu,\eta,k)g(\mathbf{Q}_0,\epsilon)\right)\right]}, \tag{A-12}$$

$$B \ge \frac{ck^2\lambda_1^2[(s+2\gamma)\log p + \log T]}{\epsilon^2\eta^2\lambda_k^2},$$

where c is a universal constant, then $\|\mathbf{U}_{k,\perp}^\top\mathbf{Q}_T\|_2 \le \epsilon$ holds with probability at least $1-s^{-10}$.

Proof. This theorem can be easily derived from Lemma A-5. Notice that

$$\frac{-\mu f(\mu,\eta,k)+\sqrt{1+f(\mu,\eta,k)^2-\mu^2}}{1+f(\mu,\eta,k)^2} = \frac{1-\mu^2}{\mu f(\mu,\eta,k)+\sqrt{1+f(\mu,\eta,k)^2-\mu^2}} \ge \frac{1-\mu^2}{\sqrt{1-\mu^2}+(\mu+1)f(\mu,\eta,k)},$$

which means that (A-11) implies (A-9). In order to make $\|\mathbf{U}_{k,\perp}^\top\mathbf{Q}_T\|_2 \le \epsilon$, one should ensure that the number of iterations T satisfies that

$$\left[\frac{\mu}{\sqrt{1-g(\mathbf{Q}_0,\epsilon)^2} - f(\mu,\eta,k)g(\mathbf{Q}_0,\epsilon)}\right]^T g(\mathbf{Q}_0,\epsilon) \le \epsilon,$$

which implies (A-12). □

Corollary 1. *When $k = d$, one can set η in Theorem A-7 to $\frac{\lambda_d-(d+1)\sigma^2}{4\lambda_d}$ so that $\mu = \frac{1}{2}\left[1+\frac{(d+1)\sigma^2}{\lambda_d}\right]$ and $f(\mu,\eta,k) = \frac{2+\sqrt{2}}{\sqrt{d}}\mu$.*

Proof. It can be easily verified that $\mu = \frac{(d+1)\lambda_{d+1}+2\eta\lambda_d}{\lambda_d}$ and $\frac{(2+\sqrt{2})\mu}{\sqrt{k}} \geq \frac{\eta}{k}$. □

Corollary 2. *When $k = 1$, one can set η in Theorem 1 to $\frac{\lambda_1 - 2\lambda_2}{4\lambda_1}$ so that $\mu = \frac{1}{2} + \frac{\lambda_2}{\lambda_1}$ and $f(\mu, \eta, k) = (2 + \sqrt{2})\mu$.*

Proof. The proof is the same as that of Corollary 1. □

2.1.2. Proof of Theorem 2

Lemma A-6. *Let $\hat{\mathbf{u}}_t \triangleq \mathbf{q}^{(t)}$ – the estimation of the t^{th} leading eigenvector $\mathbf{u}_t$ of $\mathbf{\Sigma}$, and $\hat{\mathbf{\Sigma}}_0 \triangleq \frac{1}{B}\sum_{i=1}^{B} \mathbf{x}_i\mathbf{x}_i^\top$ be the empirical covariance matrix of these B samples. Define $\hat{\mathbf{\Sigma}}_t \triangleq \prod_{j=1}^{t-1}(\mathbf{I} - \hat{\mathbf{u}}_j\hat{\mathbf{u}}_j^\top) \cdot \hat{\mathbf{\Sigma}}_0 \cdot \prod_{j=1}^{t-1}(\mathbf{I} - \hat{\mathbf{u}}_j\hat{\mathbf{u}}_j^\top)$ and $\mathbf{\Sigma}_t \triangleq \prod_{i=1}^{t-1}(\mathbf{I} - \mathbf{u}_i\mathbf{u}_i^\top) \cdot \mathbf{\Sigma} \cdot \prod_{i=1}^{t-1}(\mathbf{I} - \mathbf{u}_i\mathbf{u}_i^\top)$. If $\|\mathbf{u}_{i,\perp}^\top \hat{\mathbf{u}}_i\|_2 \leq \epsilon$ for $i = 1, \cdots, t-1$ and $\epsilon \leq \frac{\sqrt{2}}{2}$, where $\mathbf{u}_{i,\perp}$ is an orthogonal matrix such that $[\mathbf{u}_i, \mathbf{u}_{i,\perp}]$ is orthogonal, then for any fixed $\mathcal{F}$, we have*

$$\|(\hat{\mathbf{\Sigma}}_t - \mathbf{\Sigma}_t)_{\mathcal{F}}\|_2 \leq (1 + 10t\epsilon)\|(\hat{\mathbf{\Sigma}}_0 - \mathbf{\Sigma})_{\mathcal{F}}\|_2 + 10t\epsilon\|\mathbf{\Sigma}\|_2. \tag{A-13}$$

Proof. Notice that

$$\begin{aligned}
\|(\hat{\mathbf{\Sigma}}_t - \mathbf{\Sigma}_t)_{\mathcal{F}}\|_2 =& \|(\prod_{j=1}^{t-1}(\mathbf{I} - \hat{\mathbf{u}}_j\hat{\mathbf{u}}_j^\top) \cdot \hat{\mathbf{\Sigma}}_0 \cdot \prod_{j=1}^{t-1}(\mathbf{I} - \hat{\mathbf{u}}_j\hat{\mathbf{u}}_j^\top) - \prod_{i=1}^{t-1}(\mathbf{I} - \mathbf{u}_i\mathbf{u}_i^\top) \cdot \mathbf{\Sigma} \cdot \prod_{i=1}^{t-1}(\mathbf{I} - \mathbf{u}_i\mathbf{u}_i^\top))_{\mathcal{F}}\|_2 \\
=& \|((\mathbf{I} - \hat{\mathbf{u}}_{t-1}\hat{\mathbf{u}}_{t-1}^\top)\hat{\mathbf{\Sigma}}_{t-1}(\mathbf{I} - \hat{\mathbf{u}}_j\hat{\mathbf{u}}_j^\top) - (\mathbf{I} - \mathbf{u}_{t-1}\mathbf{u}_{t-1}^\top)\mathbf{\Sigma}_{t-1}(\mathbf{I} - \mathbf{u}_{t-1}\mathbf{u}_{t-1}^\top))_{\mathcal{F}}\|_2 \\
\leq& \|((\mathbf{I} - \mathbf{u}_{t-1}\mathbf{u}_{t-1}^\top)(\hat{\mathbf{\Sigma}}_{t-1} - \mathbf{\Sigma}_{t-1})(\mathbf{I} - \mathbf{u}_{t-1}\mathbf{u}_{t-1}^\top))_{\mathcal{F}}\|_2 + \\
& \|((\mathbf{u}_{t-1}\mathbf{u}_{t-1}^\top - \hat{\mathbf{u}}_{t-1}\hat{\mathbf{u}}_{t-1}^\top)\hat{\mathbf{\Sigma}}_{t-1}(\mathbf{I} - \hat{\mathbf{u}}_{t-1}\hat{\mathbf{u}}_{t-1}^\top) + (\mathbf{I} - \mathbf{u}_{t-1}\mathbf{u}_{t-1}^\top)\hat{\mathbf{\Sigma}}_{t-1}(\mathbf{u}_{t-1}\mathbf{u}_{t-1}^\top - \hat{\mathbf{u}}_{t-1}\hat{\mathbf{u}}_{t-1}^\top))_{\mathcal{F}}\|_2 \\
\leq& \|(\hat{\mathbf{\Sigma}}_{t-1} - \mathbf{\Sigma}_{t-1})_{\mathcal{F}}\|_2 + 2\|\hat{\mathbf{u}}_{t-1}\hat{\mathbf{u}}_{t-1}^\top - \mathbf{u}_{t-1}\mathbf{u}_{t-1}^\top\|_2\|\hat{\mathbf{\Sigma}}_{t-1}\|_2 \\
\leq& \|(\hat{\mathbf{\Sigma}}_{t-1} - \mathbf{\Sigma}_{t-1})_{\mathcal{F}}\|_2 + 2\|\hat{\mathbf{u}}_{t-1}\hat{\mathbf{u}}_{t-1}^\top - \mathbf{u}_{t-1}\mathbf{u}_{t-1}^\top\|_2\|\hat{\mathbf{\Sigma}}_0\|_2
\end{aligned}$$

Let $\mathbf{v}_{t-1} = \hat{\mathbf{u}}_{t-1} - \mathbf{u}_{t-1}$, since $\|\hat{\mathbf{u}}_{t-1}\|_2 = 1$ and $\|\mathbf{u}_{t-1,\perp}^\top \hat{\mathbf{u}}_{t-1}\|_2 \leq \epsilon$, we have

$$\mathbf{u}_{t-1}^\top \mathbf{v}_{t-1} = -\frac{1}{2}\|\mathbf{v}_{t-1}\|_2^2, \text{ and } |1 + \mathbf{u}_{t-1}^\top \mathbf{v}_{t-1}| \geq \sqrt{1 - \epsilon^2},$$

which implies that $\|\mathbf{v}_{t-1}\|_2^2 \leq 2(1 - \sqrt{1 - \epsilon^2}) = \frac{2\epsilon^2}{1 + \sqrt{1 - \epsilon^2}} \leq 2\epsilon^2 \leq 1$. Hence

$$\begin{aligned}
\|\hat{\mathbf{u}}_{t-1}\hat{\mathbf{u}}_{t-1}^\top - \mathbf{u}_{t-1}\mathbf{u}_{t-1}^\top\|_2 =& \|\mathbf{v}_{t-1}\mathbf{u}_{t-1}^\top + \mathbf{u}_{t-1}\mathbf{v}_{t-1}^\top + \mathbf{v}_{t-1}\mathbf{v}_{t-1}^\top\|_2 \\
\leq& (2 + \|\mathbf{v}_{t-1}\|_2)\|\mathbf{v}_{t-1}\|_2 \leq 5\epsilon.
\end{aligned}$$

Thus, $\|\hat{\mathbf{\Sigma}}_t - \mathbf{\Sigma}_t\|_2 \leq \|\hat{\mathbf{\Sigma}}_{t-1} - \mathbf{\Sigma}_{t-1}\|_2 + 10\epsilon\|\hat{\mathbf{\Sigma}}_0\|_2$, which implies (A-13). □

Theorem A-8. *Let $\eta > 0$, $0 < \epsilon < \frac{\sqrt{2}}{2}$, $\gamma_i \geq s$ for $i = 1, \cdots, k$. Let $\{\epsilon_1, \epsilon_2, \cdots, \epsilon_k\}$ be such that $\epsilon_k = \epsilon, \epsilon_{k-1} = \frac{\eta\lambda_k\epsilon_k}{20\lambda_1 k}, \epsilon_{k-2} = \frac{\eta\lambda_{k-1}\epsilon_{k-1}}{20\lambda_1(k-1)}, \cdots, \epsilon_1 = \frac{\eta\lambda_2\epsilon_2}{40\lambda_1}$. For the i^{th} iteration, let $\mu_i \triangleq \frac{2\lambda_{i+1}+2\eta\lambda_i}{\lambda_i}$, $f(\mu_i, \eta) \triangleq \max\{(2 + \sqrt{2})\mu_i, \eta\}$ and $g(\mathbf{q}_0^{(i)}, \epsilon_i) = \max\{\sqrt{1 - |\mathbf{u}_i^\top \mathbf{q}_0^{(i)}|^2}, \epsilon_i\}$. If the initial solution $\mathbf{q}_0^{(i)}$ satisfies that*

$$g(\mathbf{q}_0^{(i)}, \epsilon_i) < \frac{1 - \mu_i^2}{\sqrt{1 - \mu_i^2} + (\mu_i + 1)f(\mu_i, \eta)},$$

and the following two inequalities hold

$$T_i \geq \frac{\log(\epsilon_i / g(\mathbf{q}_0^{(i)}, \epsilon_i))}{\log\left[\mu_i / \left(\sqrt{1 - g(\mathbf{q}_0^{(i)}, \epsilon_i)^2} - f(\mu_i, \eta)g(\mathbf{q}_0^{(i)}, \epsilon_i)\right)\right]}, \text{ and } B_i \geq \frac{c\lambda_1^2[(s + 2\gamma_i)\log p + \log(kT_i)]}{\epsilon_i^2\eta^2\lambda_i^2},$$

where c is a universal constant, then $|\mathbf{u}_i^\top \mathbf{q}^{(i)}| \geq \sqrt{1 - \epsilon_i^2}$ holds with probability at least $1 - s^{-10}$.

Proof. We use the same notation as Lemma A-6. By following the proof of Lemma A-5, we only need to select B_i on the i^{th} iteration so that $\sup_{\mathcal{F}:|\mathcal{F}|\le s+2\gamma_i} \|(\hat{\mathbf{\Sigma}}_i - \mathbf{\Sigma}_i)_{\mathcal{F}}\|_2 \le \eta\lambda_i\epsilon_i$. By Lemma A-6 and Theorem A-2, we have

$$\begin{aligned}\sup_{\mathcal{F}:|\mathcal{F}|\le s+2\gamma_i} \|(\hat{\mathbf{\Sigma}}_i - \mathbf{\Sigma}_i)_{\mathcal{F}}\|_2 \le &(1+10i\epsilon_{i-1}) \sup_{\mathcal{F}:|\mathcal{F}|\le s+2\gamma_i} \|(\hat{\mathbf{\Sigma}}_0 - \mathbf{\Sigma})_{\mathcal{F}}\|_2 + 10i\epsilon_{i-1}\|\mathbf{\Sigma}\|_2 \\ =&(1+\frac{\eta\lambda_i\epsilon_i}{2\lambda_1}) \sup_{\mathcal{F}:|\mathcal{F}|\le s+2\gamma_i} \|(\hat{\mathbf{\Sigma}}_0 - \mathbf{\Sigma})_{\mathcal{F}}\|_2 + \frac{\eta\lambda_i\epsilon_i}{2\lambda_1}\|\mathbf{\Sigma}\|_2 \\ \le&2 \sup_{\mathcal{F}:|\mathcal{F}|\le s+2\gamma_i} \|(\hat{\mathbf{\Sigma}}_0 - \mathbf{\Sigma})_{\mathcal{F}}\|_2 + \frac{\eta\lambda_i\epsilon_i}{2} \le \eta\lambda_i\epsilon_i\end{aligned}$$

holds with probability at least $1-\frac{s^{-10}}{kT_i}$ when $B_i \ge \frac{c\lambda_1^2[(s+2\gamma_i)\log p + \log(kT_i)]}{\epsilon_i^2\eta^2\lambda_i^2}$. □

2.2. Streaming Sparse ECA

As we have discussed in the paper, the difference between ECA and PCA is that ECA considers the multivariate Kendall's tau estimator instead of the empirical covariance matrix as its input. The multivariate Kendall's tau matrix is defined as follows:

$$\mathbf{K} \triangleq \mathbb{E}\left[\frac{(\mathbf{x}-\tilde{\mathbf{x}})(\mathbf{x}-\tilde{\mathbf{x}})^\top}{\|\mathbf{x}-\tilde{\mathbf{x}}\|_2^2}\right],$$

where $\mathbf{x}$ is a random vector and $\tilde{\mathbf{x}}$ is an independent copy of $\mathbf{x}$. Let $\mathbf{x}_1, \cdots, \mathbf{x}_n$ be n independent realizations of a random vector $\mathbf{x} \sim EC_p(\mu, \mathbf{\Sigma}, \xi)$. Han & Liu (2013) considered a second-order U-statistic as the estimator of $\mathbf{K}$, which averages $\frac{(\mathbf{x}_i-\mathbf{x}_j)(\mathbf{x}_i-\mathbf{x}_j)^\top}{\|\mathbf{x}_i-\mathbf{x}_j\|_2^2}$ for all i, j such that $i \neq j$. This estimator has high computational cost and is hard to be extended to the streaming data model. Without loss of generality, we assume that n is an even number. If n is odd, one can just use $n-1$ samples. We consider another estimator of $\mathbf{K}$:

$$\hat{\mathbf{K}} \triangleq \frac{2}{n}\sum_{i=1}^{n/2} \frac{(\mathbf{x}_{2i-1}-\mathbf{x}_{2i})(\mathbf{x}_{2i-1}-\mathbf{x}_{2i})^\top}{\|\mathbf{x}_{2i-1}-\mathbf{x}_{2i}\|_2^2}.$$

In other words, $\hat{\mathbf{K}}$ is the empirical covariance matrix of $\{\frac{\mathbf{x}_1-\mathbf{x}_2}{\|\mathbf{x}_1-\mathbf{x}_2\|_2}, \cdots, \frac{\mathbf{x}_{n-1}-\mathbf{x}_n}{\|\mathbf{x}_{n-1}-\mathbf{x}_n\|_2}\}$. Recall that the k^{th} largest eigenvalue of $\mathbf{K}$ is denoted by $\lambda_k(\mathbf{K})$, $\mathbf{U}_k$ are the leading k eigenvectors of $\mathbf{K}$ and $\mathbf{U}_{k,\perp}$ is an orthogonal basis of the perpendicular subspace to the one spanned by $\mathbf{U}_k$. Then we have the following theorem.

Theorem A-9. *For $\eta > 0$, $0 < \epsilon < 1$, and $\gamma \ge s$, let $\mu \triangleq \frac{(k+1)\lambda_{k+1}+2\eta\lambda_k}{\lambda_k}$, $f(\mu,\eta,k) \triangleq \max\{\frac{(2+\sqrt{2})\mu}{\sqrt{k}}, \frac{\eta}{k}\}$ and $g(\mathbf{Q},\epsilon) \triangleq \max\{\|\mathbf{U}_{k,\perp}^\top\mathbf{Q}\|_2, \epsilon\}$. If the initial solution $\mathbf{Q}_0$ satisfies that*

$$g(\mathbf{Q}_0,\epsilon) < \frac{1-\mu^2}{\sqrt{1-\mu^2}+(\mu+1)f(\mu,\eta,k)},$$

and the following two inequalities hold

$$T \ge \frac{\log(\epsilon/g(\mathbf{Q}_0,\epsilon))}{\log\left[\mu/\left(\sqrt{1-g(\mathbf{Q}_0,\epsilon)^2}-f(\mu,\eta,k)g(\mathbf{Q}_0,\epsilon)\right)\right]},$$

$$B \ge \frac{ck^2\left(1+\lambda_1(\mathbf{K})\right)^2[(s+2\gamma)\log p + \log T]}{\epsilon^2\eta^2\lambda_k(\mathbf{K})^2},$$

where c is a universal constant, then $\|\mathbf{U}_{k,\perp}^\top\mathbf{Q}_T\|_2 \le \epsilon$ holds with probability at least $1-s^{-10}$.

Proof. From Theorem A-6, when $B \ge \frac{ck^2(1+\lambda_1(\mathbf{K}))^2[(s+2\gamma)\log p+\log T]}{\epsilon^2\eta^2\lambda_k(\mathbf{K})^2}$, we know that $\sup_{\|\mathbf{v}\|_2=1,\|\mathbf{v}\|_0=s} |\mathbf{v}^\top(\hat{\mathbf{K}}-\mathbf{K})\mathbf{v}| \le \frac{\eta\epsilon\lambda_k(\mathbf{K})}{k}$. holds with probability at least $1-\frac{s^{-10}}{T}$. Then follow the proof of Theorem 1, we can obtain this theorem. □

3. Proofs in Section 5

Proposition A-1. *For $0 < \epsilon < 1$, let $\bar{\theta} = \sqrt{\frac{\epsilon^2}{\epsilon^2 + 8(3+2\sqrt{2})(1+\frac{k\lambda_{k+1}}{\lambda_k})^2}}$, then $\|\mathbf{U}_{k,\perp}^\top \mathbf{Q}_0\|_2 \le \epsilon$ holds with probability at least $1 - d^{-10}$ if $\gamma \ge s$, $\|\mathbf{U}_{k,\perp}^\top \hat{\mathbf{Q}}_0\|_2 \le \bar{\theta}$ and $B \ge \frac{ck^2[\|\mathbf{A}\|_2^4 d + (2\|\mathbf{A}\|_2+\sigma)^2\sigma^2 p]}{\lambda_k^2 \bar{\theta}^2}$, where c is a universal constant.*

Proof. We follow the proof of Lemma A-4. Let $\mathbf{W} = \hat{\mathbf{\Sigma}} - \mathbf{\Sigma}$, then $\|\mathbf{W}\|_2 \le \frac{\lambda_k \bar{\theta}}{k}$ with probability at least $1 - d^{-10}$ when $B \ge \frac{ck^2[\|\mathbf{A}\|_2^4 d + (2\|\mathbf{A}\|_2+\sigma)^2\sigma^2 p]}{\lambda_k^2 \bar{\theta}^2}$ (by Theorem A-3). Notice that $\bar{\theta}$ satisfies

$$\frac{1}{2}\lambda_k\sqrt{1-\bar{\theta}^2} \ge (2+\sqrt{2})k(\lambda_{k+1}\bar{\theta} + \|\mathbf{W}\|_2) \ge (2+\sqrt{2})\sqrt{k}(\lambda_{k+1}\bar{\theta} + \|\mathbf{W}\|_2). \tag{A-14}$$

Suppose that $\hat{\mathbf{Q}}_0 = \mathbf{U}_k\mathbf{A} + \mathbf{U}_{k,\perp}\mathbf{B}$, then we have

$$\begin{aligned}
\tilde{\mathbf{S}}_0 =& \hat{\mathbf{\Sigma}}\hat{\mathbf{Q}}_0 = \mathbf{\Sigma}\hat{\mathbf{Q}}_0 + \mathbf{W}\hat{\mathbf{Q}}_0 \\
=& \mathbf{U}\mathbf{\Lambda}\mathbf{U}^\top(\mathbf{U}_k\mathbf{A} + \mathbf{U}_{k,\perp}\mathbf{B}) + \mathbf{W}\hat{\mathbf{Q}}_0 \\
=& \mathbf{U}_k\mathbf{\Lambda}_k\mathbf{A} + \mathbf{U}(\mathbf{\Lambda} - \mathbf{\Lambda}_k)\mathbf{B} + \mathbf{W}\hat{\mathbf{Q}}_0,
\end{aligned}$$

which implies that

$$\begin{aligned}
\|\tilde{\mathbf{S}}_{0,\mathcal{S}} - \tilde{\mathbf{S}}_0\|_F \le& \sqrt{k}\|\tilde{\mathbf{S}}_{0,\mathcal{S}} - \tilde{\mathbf{S}}_0\|_2 \\
\le& \sqrt{k}\|\mathbf{U}(\mathbf{\Lambda} - \mathbf{\Lambda}_k)\mathbf{B} + \mathbf{W}\hat{\mathbf{Q}}_0\|_2 \\
\le& \sqrt{k}[\lambda_{k+1}\|\mathbf{B}\|_2 + \|\mathbf{W}\|_2] \\
=& \sqrt{k}[\lambda_{k+1}\|\mathbf{U}_{k,\perp}^\top\hat{\mathbf{Q}}_0\|_2 + \|\mathbf{W}\|_2].
\end{aligned}$$

Since $\mathbf{S}_0 = \text{Truncate}(\tilde{\mathbf{S}}_0, \gamma)$ and $\gamma \ge s$, we have

$$\|\mathbf{S}_0 - \tilde{\mathbf{S}}_0\|_F \le \|\tilde{\mathbf{S}}_{0,\mathcal{S}} - \tilde{\mathbf{S}}_0\|_F \le \sqrt{k}[\lambda_{k+1}\|\mathbf{U}_{k,\perp}^\top\hat{\mathbf{Q}}_0\|_2 + \|\mathbf{W}\|_2].$$

By applying the Weyl's Inequalities,

$$\begin{aligned}
\frac{1}{\|\tilde{\mathbf{S}}_0^\dagger\|_2} = \delta_{\min}(\tilde{\mathbf{S}}_0) \ge& \delta_{\min}(\mathbf{U}_k\mathbf{\Lambda}_k\mathbf{A}) - \|\mathbf{U}_{k,\perp}(\mathbf{\Lambda} - \mathbf{\Lambda}_k)\mathbf{B} + \mathbf{W}\hat{\mathbf{Q}}_0\|_2 \\
\ge& \sqrt{\delta_{\min}(\mathbf{A}^\top\mathbf{\Lambda}_k^2\mathbf{A})} - \lambda_{k+1}\|\mathbf{B}\|_2 - \|\mathbf{W}\|_2 \\
\ge& \lambda_k\sqrt{1 - \|\mathbf{U}_{k,\perp}^\top\hat{\mathbf{Q}}_0\|_2^2} - \lambda_{k+1}\|\mathbf{U}_{k,\perp}^\top\hat{\mathbf{Q}}_0\|_2 - \|\mathbf{W}\|_2 > 0
\end{aligned}$$

where the last inequality (> 0) follows from $\|\mathbf{U}_{k,\perp}^\top\hat{\mathbf{Q}}_0\|_2 \le \bar{\theta}$ and Inequality (A-14). Notice that $\|\mathbf{U}_{k,\perp}^\top\hat{\mathbf{Q}}_0\|_2 \le \bar{\theta}$ and Inequality (A-14) also imply $\lambda_k\sqrt{1 - \|\mathbf{U}_{k,\perp}^\top\hat{\mathbf{Q}}_0\|_2^2} > (2+\sqrt{2})\sqrt{k}[\lambda_{k+1}\|\mathbf{U}_{k,\perp}^\top\hat{\mathbf{Q}}_0\|_2 + \|\mathbf{W}\|_2]$. Then from Theorem A-1, we have

$$\begin{aligned}
\|\hat{\mathbf{Q}}_{0,\perp}^\top(\mathbf{Q}_0 - \hat{\mathbf{Q}}_0)\|_F \le& \frac{\sqrt{k}\|\tilde{\mathbf{S}}_0^\dagger\|_2\|\mathbf{S}_0 - \tilde{\mathbf{S}}_0\|_F}{1 - (1+\sqrt{2})\|\tilde{\mathbf{S}}_0^\dagger\|_2\|\mathbf{S}_0 - \tilde{\mathbf{S}}_0\|_F} \\
\le& \frac{k[\lambda_{k+1}\|\mathbf{U}_{k,\perp}^\top\hat{\mathbf{Q}}_0\|_2 + \|\mathbf{W}\|_2]}{\lambda_k\sqrt{1 - \|\mathbf{U}_{k,\perp}^\top\hat{\mathbf{Q}}_0\|_2^2} - \lambda_{k+1}\|\mathbf{U}_{k,\perp}^\top\hat{\mathbf{Q}}_0\|_2 - \|\mathbf{W}\|_2 - (1+\sqrt{2})\sqrt{k}[\lambda_{k+1}\|\mathbf{U}_{k,\perp}^\top\hat{\mathbf{Q}}_0\|_2 + \|\mathbf{W}\|_2]} \\
\le& \frac{k[\lambda_{k+1}\|\mathbf{U}_{k,\perp}^\top\hat{\mathbf{Q}}_0\|_2 + \|\mathbf{W}\|_2]}{\lambda_k\sqrt{1 - \|\mathbf{U}_{k,\perp}^\top\hat{\mathbf{Q}}_0\|_2^2} - (2+\sqrt{2})\sqrt{k}[\lambda_{k+1}\|\mathbf{U}_{k,\perp}^\top\hat{\mathbf{Q}}_0\|_2 + \|\mathbf{W}\|_2]}.
\end{aligned}$$

Thus,

$$\begin{aligned}
\|\mathbf{U}_{k,\perp}^\top\mathbf{Q}_0\|_2 \le& \|\mathbf{U}_{k,\perp}^\top\hat{\mathbf{Q}}_0\|_2 + \|\hat{\mathbf{Q}}_{0,\perp}^\top(\mathbf{Q}_0 - \hat{\mathbf{Q}}_0)\|_2 \\
\le& \|\mathbf{U}_{k,\perp}^\top\hat{\mathbf{Q}}_0\|_2 + \frac{k[\lambda_{k+1}\|\mathbf{U}_{k,\perp}^\top\hat{\mathbf{Q}}_0\|_2 + \|\mathbf{W}\|_2]}{\lambda_k\sqrt{1 - \|\mathbf{U}_{k,\perp}^\top\hat{\mathbf{Q}}_0\|_2^2} - (2+\sqrt{2})\sqrt{k}[\lambda_{k+1}\|\mathbf{U}_{k,\perp}^\top\hat{\mathbf{Q}}_0\|_2 + \|\mathbf{W}\|_2]}.
\end{aligned}$$

For simplicity, let $\theta = \|\mathbf{U}_{k,\perp}^\top \hat{\mathbf{Q}}_0\|_2$ and $\xi = \|\mathbf{W}\|_2$. In order to make the right hand side of the equation above less than or equal to ϵ, we only need to ensure that

$$\theta \leq \frac{1}{2}\epsilon, \text{ and } \frac{k[\lambda_{k+1}\theta + \xi]}{\lambda_k\sqrt{1-\theta^2} - (2+\sqrt{2})\sqrt{k}[\lambda_{k+1}\theta + \xi]} \leq \frac{1}{2}\epsilon.$$

By the assumptions, we know that $\theta \leq \bar{\theta} \leq \frac{1}{2}\epsilon$. Then by Inequality (A-14), we have

$$\frac{k[\lambda_{k+1}\theta + \xi]}{\lambda_k\sqrt{1-\theta^2} - (2+\sqrt{2})\sqrt{k}[\lambda_{k+1}\theta + \xi]} \leq \frac{k[\lambda_{k+1}\bar{\theta} + \xi]}{\lambda_k\sqrt{1-\bar{\theta}^2} - (2+\sqrt{2})\sqrt{k}[\lambda_{k+1}\bar{\theta} + \xi]} \leq \frac{2(k\lambda_{k+1} + \lambda_k)\bar{\theta}}{\lambda_k\sqrt{1-\bar{\theta}^2}} \leq \frac{\epsilon}{2+\sqrt{2}} < \frac{\epsilon}{2}.$$

Thus, $\|\mathbf{U}_{k,\perp}^\top \mathbf{Q}_0\|_2 \leq \epsilon$. □

References

Chang, X.-W. On the perturbation of the Q-factor of the QR factorization. *Numerical Linear Algebra with Applications*, 19(3):607–619, 2012.

Chen, Y. and Caramanis, C. Noisy and missing data regression: Distribution-oblivious support recovery. In *ICML*, 2013.

Han, F. and Liu, H. ECA: High dimensional elliptical component analysis in non-Gaussian distributions. *arXiv:1310.3561*, 2013.

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.

Tropp, J. A. User-friendly tail bounds for sums of random matrices. *Foundations of Computational Mathematics*, 12(4): 389–434, 2012.