A Appendix: Proof of Differential Privacy

In this appendix we aim to prove Theorem 4.1. Algorithm 1, though being an iterative process, can be decomposed into two steps.

1. Central server draws T adaptive queries from the jth data provider,
$$H_j(S_j) = \big(H_j\big(S_j, Q^{(1)}\big), \dots H_j\big(S_j, Q^{(T)}\big)\big),$$
and $H_j(S_j) \in \otimes_{i=1}^{\top} \mathbb{R}^{d*k}$.
2. The central server calculates the output Q^*_{dp} by post-processing of $H_j(S_j)$.

Theorem 4.1 states that the output Q^*_{dp} and $H_j(S_j)$is (ϵ, δ)-differentially private with respect to the presence of any sample. For any two sets of samples differing in only one data item: $S = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_i, \dots, \mathbf{x}_n\}$ and $S' = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}'_i, \dots, \mathbf{x}_n\}$, we aim to prove

$$\Pr\big(Q^*_{dp}(S') \subseteq R_1\big) \le e^{\epsilon} \Pr\big(Q^*_{dp}(S) \subseteq R_1\big) + \delta,$$
$$\Pr\big(H_j(S') \subseteq R_2\big) \le e^{\epsilon} \Pr\big(H_j(S) \subseteq R_2\big) + \delta,$$

for all $R_1 \in \mathbb{R}^{d\times k}$ and $R_2 \in \otimes_{j=1}^{\top} \mathbb{R}^{d*k}$. Suppose sample $\mathbf{x}_i$ is from the jth data provider, i.e. $\mathbf{x}_i \in S_j$. Our proof consists of three parts. In Lemma A.1, we prove that every query $H_j\big(S_j, Q^{(t)}\big)$ is differentially private. In Lemma A.2, we prove that adaptive composition of T queries $H_j(S_j)$ introduces more privacy error, but is still differentially private. In Lemma A.3, we argue that post-processing preserves privacy and thus Q^*_{dp} is differentially private.

Lemma A.1 (Gaussian Mechanism)**.** For S_j and S'_j differing in one data item, i.e. $S_j = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_i, \dots, \mathbf{x}_{n_j}\}$, and $S'_j = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}'_i, \dots, \mathbf{x}_{n_j}\}$, and any measurable set $R \subseteq \mathbb{R}^{d\times k}$, we have

$$\Pr(H_j\big(S'_j, Q^{(t)}\big) \in R) \le e^{\epsilon} \Pr(H_j\big(S_j, Q^{(t)}\big) \in R) + \delta,$$

where $\sigma = 2\epsilon^{-1}\sqrt{2\ln(2/\delta)}$ with $0 < \epsilon < 1$ and $0 < \delta < 1/2$.

Proof. For simplicity of notation, we omit the subscript j in the proof. In each iteration of Algorithm 1, $H(S, Q^{(t)}) = \widehat{\Sigma}(S)Q^{(t)} + G^{(t)}/n_j$, where $\widehat{\Sigma}$ is the sample covariance of the jth data provider, and $G^{(t)}$ is a $d \times k$ matrix with entry-wise i.i.d. normal random variable $N(0, \sigma^2)$. Note that

$$\begin{aligned}&\Pr(H_j\big(S'_j, Q^{(t)}\big) \in R)\\ \le& e^{\epsilon} \Pr(H_j\big(S_j, Q^{(t)}\big) \in R) + \delta,\end{aligned}$$

for all $R \subseteq \mathbb{R}^{d\times k}$ is equivalent to

$$\begin{aligned}&\Pr(n_j H_j\big(S'_j, Q^{(t)}\big) \in R)\\ \le& e^{\epsilon} \Pr(n_j H_j\big(S_j, Q^{(t)}\big) \in R) + \delta,\end{aligned}$$

for all $R \subseteq \mathbb{R}^{d\times k}$, where n_j is the number of samples. Without any lose of generality, we can let $n_j = 1$ and $||\mathbf{x}_i||_2 \le 1$ (because of the boundedness assumption), we have

$$\big\|\mathbf{x}_i\mathbf{x}_i^{\top}\big\|_F \le \big\|\mathbf{x}_i\big\|_F \big\|\mathbf{x}_i^{\top}\big\|_F = \big\|\mathbf{x}_i\big\|_2 \big\|\mathbf{x}_i\big\|_2 \le 1,$$

and since $Q^{(t)} \in \mathbb{R}^{d\times k}$ is an orthonormal matrix, according to the rotation invariant property of Frobenius norm, we have

$$\begin{aligned}&\big\|\widehat{\Sigma}(S)Q^{(t)} - \widehat{\Sigma}(S')Q^{(t)}\big\|_F\\ \le&\big\|\widehat{\Sigma}(S) - \widehat{\Sigma}(S')\big\|_F = \frac{1}{n_j}||\mathbf{x}_i\mathbf{x}_i^{\top} - \mathbf{x}'_i\mathbf{x}_i'^{\top}||_F\\ \le&\frac{1}{n_j}(||\mathbf{x}_i\mathbf{x}_i^{\top}||_F + ||\mathbf{x}'_i\mathbf{x}_i'^{\top}||_F) \le \frac{2}{n_j} = 2.\end{aligned} \tag{A.1}$$

Denote $W = \widehat{\Sigma}(S)Q^{(t)}$ and $W' = \widehat{\Sigma}(S')Q^{(t)}$. Let $\Delta W = W - W' \in \mathbb{R}^{d\times k}$, and $R_W = \{X + W | X \in R\}$. From (A.1) we know $||\Delta W||_F \le 2$. Now we have

$$\begin{aligned}&\Pr\big(H_j(S_j, Q^{(t)}) \in R\big) = \Pr\big(W + G_k^{(t)} \in R\big)\\ =&\frac{1}{(\sqrt{2\pi}\sigma)^{dk}} \int_{R_W} e^{-\frac{1}{2\sigma^2}||X||_F^2} \mathrm{d}\mu(X),\end{aligned}$$

and

$$\begin{aligned}&\Pr\big(H_j(S'_j, Q^{(t)}) \in R\big) = \Pr(W' + G_k^{(t)} \in R)\\ =&\frac{1}{(\sqrt{2\pi}\sigma)^{dk}} \int_{R_W} e^{-\frac{1}{2\sigma^2}||X - \Delta W||_F^2} \mathrm{d}\mu(X).\end{aligned}$$

Let's take X and ΔW as $d \times k$ dimensional vectors. Since the Gaussian distribution is spherical and symmetric, we can assume that $\Delta W = (w_1, 0, 0, \dots)$ (only nonzero in the first entry) by change of basis. Denote w_1 as the first entry of W and x_1 as the first entry of X. Because of the symmetry of W and W', we assume $w_1 > 0$. Then we have

$$\begin{aligned}&\Pr\big(H_j(S'_j, Q^{(t)}) \in R\big)\\ =&\frac{1}{\sqrt{2\pi}\sigma} \int_{R_W} e^{-\frac{1}{2\sigma^2}(x_1 - w_1)^2} \mathrm{d}\mu(x_1)\\ =&\frac{1}{\sqrt{2\pi}\sigma} \int_{R_W} e^{-\frac{x_1^2}{2\sigma^2} + \frac{2x_1w_1 - w_1^2}{2\sigma^2}} \mathrm{d}\mu(x_1).\end{aligned}$$

Note that $x_1 < \sigma^2\epsilon/w_1 - w_1/2$ implies

$$\frac{2x_1w_1 - w_1^2}{2\sigma^2} < \epsilon.$$

Let $R_W^- = R_W \cap (-\infty, \sigma^2\epsilon/w_1 - w_1/2)$, and $R_W^+ =$

$R_W \cap (\sigma^2\epsilon/w_1 - w_1/2, +\infty)$. Now we have

$$
\begin{aligned}
&\Pr\big(H_j(S_j', Q^{(t)}) \in R\big)\\
=&\frac{1}{\sqrt{2\pi}\sigma}\int_{R_W^-} e^{-\frac{x_1^2}{2\sigma^2}+\frac{2x_1w_1-w_1^2}{2\sigma^2}}\mathrm{d}\mu(x_1)\\
&+\frac{1}{\sqrt{2\pi}\sigma}\int_{R_W^+} e^{-\frac{(x_1-w_1)^2}{2\sigma^2}}\mathrm{d}\mu(x_1)\\
\le&\frac{1}{\sqrt{2\pi}\sigma}e^{\epsilon}\int_{R_W^-} e^{-\frac{x_1^2}{2\sigma^2}}\mathrm{d}\mu(x_1) + \Pr_{x\sim N(0,\sigma^2)}\Big(x > \frac{\sigma^2\epsilon}{w_1} - \frac{w_1}{2}\Big)\\
\le&e^{\epsilon}\Pr(H_j(S_j, Q^{(t)}) \in R) + \Pr_{x\sim N(0,\sigma^2)}\Big(x > \frac{\sigma^2\epsilon}{w_1} - \frac{w_1}{2}\Big).
\end{aligned}
\tag{A.2}
$$

Based on (A.1), we know $0 < w_1 = ||\Delta W|| \le 2$. Next we prove that with $\sigma = 2\epsilon^{-1}\sqrt{2\ln(2/\delta)}$, the second term in (A.2) is bounded by δ.
Let $t = \sigma^2\epsilon/w_1 - w_1/2$, we first verify the following inequality

$$\ln(t/\sigma) + t^2/(2\sigma^2) > \ln\Big(\frac{1}{\sqrt{2\pi}\delta}\Big). \tag{A.3}$$

Denote $\sigma = c/\epsilon$, where $c = 2\sqrt{2\ln(2/\delta)}$. We want the first term in the left-hand-side of (A.3) to be non-negative, equivalently,

$$t/\sigma = \frac{c}{w_1} - \frac{\epsilon w_1}{2c} \ge \frac{c}{2} - \frac{\epsilon}{c} \ge 1, \tag{A.4}$$

where the second to last inequality is due to the fact that $w_1 \le 2$ and the function $a/w_1 - w_1/b$ is a decreasing function of $w_1 > 0$, $a, b > 0$.
Since $0 < \delta < 1/2$, $c = 2\sqrt{2\ln(2/\delta)} \ge 3.3$ and $\epsilon \le 1$, we know (A.4) holds because

$$
\begin{aligned}
\frac{t^2}{2\sigma^2} &\ge (c/2 - \epsilon/c)^2/2 = (c^2/4 - \epsilon + \epsilon^2/c^2)/2\\
&\ge (c^2/4 - 1 + 1/c^2)/2\\
&\ge \ln(2/\delta) - 1/2 \ge 1/2,
\end{aligned}
\tag{A.5}
$$

which also implies that the second term of the left-hand-side of (A.3) is bounded, i.e.

$$\frac{t^2}{2\sigma^2} \ge \ln(2/\delta) - 1/2 \ge \ln\Big(\frac{1}{\sqrt{2\pi}\delta}\Big). \tag{A.6}$$

Based on (A.4) and (A.6), we know that (A.3) holds.Taking exponential on both sides of (A.3) gives

$$\frac{\sigma}{\sqrt{2\pi}}\frac{1}{t}e^{-t^2/(2\sigma^2)} < \delta,$$

which implies the tail bound

$$\Pr_{x\sim N(0,\sigma^2)}\big(x > t\big) < \frac{\sigma}{\sqrt{2\pi}}\frac{1}{t}e^{-t^2/(2\sigma^2)} < \delta,$$

where the first inequality is Chernoff bound for normal random variables. Plug in this to (A.2), we concluded that

$$\Pr\big(H_j(S_j', Q^{(t)}) \in R\big) \le e^{\epsilon}\Pr(H_j(S_j, Q^{(t)}) \in R) + \delta.$$

□

Lemma A.2 (Adaptive Composition)**.** Let $S = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_i, \ldots, \mathbf{x}_n\}$ is the set of samples and $S' = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_i', \ldots, \mathbf{x}_n\}$ is a neighboring set of S. Let $\mathcal{M}_1 : S \to \mathcal{M}_1(S) \in \mathcal{C}_1$ is (ϵ_1, δ_1)-differentially private, and for $2 \le k \le m$, $s_{k-1} \in \mathcal{C}_{k-1}$, $T_k : (S, s_{k-1}) \to T_k(S, s_{k-1}) \in \mathcal{C}_k$ is (ϵ_k, δ_k)-differentially private. For each query T_k, its parameter s_{k-1} can be chosen adaptively according to previous queries $T_1, T_2, \ldots, T_{k-1}$. The adaptive composition $(T_1, T_2, \ldots, T_m)$ is $(\sum_{j=1}^m \epsilon_j, \sum_{j=1}^m \delta_j)$-differentially private. For all $R \subseteq \otimes_{j=1}^m \mathcal{C}_j$, we have

$$
\begin{aligned}
&\Pr((\mathcal{M}_1, \mathcal{M}_2, \ldots, \mathcal{M}_m) \in R|S)\\
\le&e^{\epsilon}\Pr((\mathcal{M}_1, \mathcal{M}_2, \ldots, \mathcal{M}_m) \in R|S') + \delta,
\end{aligned}
$$

where $\epsilon = \sum_{j=1}^m \epsilon_j$, and $\delta = \sum_{j=1}^m \delta_j$.

Proof. Let's prove by induction. First we consider the case for $m = 2$. Let μ be the probability measure. For any measurable set $R_1 \in \mathcal{C}_1$, we have

$$
\begin{aligned}
\Pr(\mathcal{M}_1 \in R_1|S) &= \int_{R_1} p(x)\mathrm{d}\mu(x),\\
\Pr(\mathcal{M}_1 \in R_1|S') &= \int_{R_1} p'(x)\mathrm{d}\mu(x),
\end{aligned}
$$

where p, p' are the Random-Nikodym derivatives of the left-hand-side with respect to the probability measure. Bases on the definition of differential privacy, we also have

$$\Pr(\mathcal{M}_1 \in R_1|S) \le e^{\epsilon_1}\Pr(\mathcal{M}_1 \in R_1|S') + \delta_1,$$

which is equivalent to

$$\int_{R_1}(p(x) - e^{\epsilon_1}p'(x))\mathrm{d}\mu(x) \le \delta_1, \tag{A.7}$$

for any measurable $R_1 \in \mathcal{C}_1$.
Let measurable set $R \in \mathcal{C}_1 \otimes \mathcal{C}_2$. Fix $s_1 \in \mathcal{C}_1$. Denote $R(s_1) = \{s_2|(s_1, s_2) \in R\} \in \mathcal{C}_2$. Since R is measurable, $R(s_1)$ is also measurable. Now we have

$$
\begin{aligned}
&\Pr((s_1, \mathcal{M}_2) \in R|S) = \Pr(\mathcal{M}_2 \in R(s_1)|S)\\
&\le e^{\epsilon_2}\min\big\{\Pr\big(\mathcal{M}_2 \in R(s_1)|S'\big), 1\big\} + \delta_2\\
&= e^{\epsilon_2}\min\big\{\Pr\big((s_1, \mathcal{M}_2) \in R|S'\big), 1\big\} + \delta_2.
\end{aligned}
$$

Now we can prove the lemma with $m=2$ as follows

$$
\begin{aligned}
&\Pr((\mathcal{M}_1,\mathcal{M}_2)\in R|S)\\
=&\int_{\mathcal{C}_1}\Pr((s_1,\mathcal{M}_2)\in R|S)p(s_1)\mathrm{d}\mu(s_1)\\
\le&\int_{\mathcal{C}_1}\big(e^{\epsilon_2}\min\big\{\Pr\big((s_1,\mathcal{M}_2)\in R|S'\big),1\big\}+\delta_2\big)p(s_1)\mathrm{d}\mu(s_1)\\
\le&\int_{\mathcal{C}_1}e^{\epsilon_2}\min\big\{\Pr\big((s_1,\mathcal{M}_2)\in R|S'\big),1\big\}p(s_1)\mathrm{d}\mu(s_1)+\delta_2\\
=&\int_{\mathcal{C}_1}e^{\epsilon_2}\min\big\{\Pr\big((s_1,\mathcal{M}_2)\in R|S'\big),1\big\}(e^{\epsilon_1}p'(s_1)+\\
&p(s_1)-e^{\epsilon_1}p'(s_1))\mathrm{d}\mu(s_1)+\delta_2\\
\le&\int_{\mathcal{C}_1}e^{\epsilon_2}\min\big\{\Pr\big((s_1,\mathcal{M}_2)\in R|S'\big),1\big\}e^{\epsilon_1}p'(s_1)\mathrm{d}\mu(s_1)\\
&+\int_{\mathcal{C}_1}(p(s_1)-e^{\epsilon_1}p'(s_1))\mathrm{d}\mu(s_1)+\delta_2\\
\le&e^{\epsilon_1+\epsilon_2}\int_{\mathcal{C}_1}\Pr((s_1,\mathcal{M}_2)\in R|S')p'(s_1)\mathrm{d}\mu(s_1)+\delta_1+\delta_2\\
=&e^{\epsilon_1+\epsilon_2}\Pr((\mathcal{M}_1,\mathcal{M}_2)\in R|S')+\delta_1+\delta_2.
\end{aligned}
$$

Assume that the lemma holds for m. Now in the case of $m+1$, denote

$$\mathcal{M}=(\mathcal{M}_1,\mathcal{M}_2,\ldots,\mathcal{M}_m)\in R_m\subseteq\otimes_{j=1}^m\mathcal{C}_j.$$

From induction hypothesis, we know that $\mathcal{M}$ is (ϵ,δ)-differentially private, where $\epsilon=\sum_{j=1}^m\epsilon_j$ and $\delta=\sum_{j=1}^m\delta_j$. Based on the lemma for $m=2$, we have

$$
\begin{aligned}
&\Pr((\mathcal{M},\mathcal{M}_{m+1})\in R_{m+1}|S)\\
\le&e^{\epsilon+\epsilon_{m+1}}\Pr((\mathcal{M},\mathcal{M}_{m+1})\in R_{m+1}|S')+\delta+\delta_{m+1},
\end{aligned}
$$

which means that the lemma holds for $m+1$. By mathematical induction, we conclude the proof. ☐

Lemma A.3 (Post-processing)**.** Denote the set of samples as $S=\{\mathbf{x}_1,\mathbf{x}_2,\ldots,\mathbf{x}_i,\ldots,\mathbf{x}_n\}$. Let randomized algorithm $\mathcal{M}(S)\in\mathcal{C}_1$ be (ϵ,δ)-differentially private and an arbitrary mapping $f:\mathcal{C}_1\to\mathcal{C}_2$. Then $f\circ\mathcal{M}$ is (ϵ,δ)-differentially private.

Proof. Let sample set $S'=\{\mathbf{x}_1,\mathbf{x}_2,\ldots,\mathbf{x}'_i,\ldots,\mathbf{x}_n\}=S\backslash\{\mathbf{x}_i\}\cup\{\mathbf{x}'_i\}$ as a neighboring dataset of S. The (ϵ,δ)-differential privacy of $\mathcal{M}$ means that for any $R\subseteq\mathcal{C}_1$,

$$\Pr(\mathcal{M}(S')\in R)\le e^{\epsilon}\Pr(\mathcal{M}(S)\in R)+\delta.$$

For any $T\subseteq\mathcal{C}_2$, let $R(T)=\{x|f(x)\in T\}$. Now we have

$$
\begin{aligned}
&\Pr(f\circ\mathcal{M}(S')\in T)\\
=&\Pr(f(\mathcal{M}(S'))\in T)=\Pr(\mathcal{M}(S')\in R(T))\\
\le&e^{\epsilon}\Pr(\mathcal{M}(S)\in R(T))+\delta\\
=&e^{\epsilon}\Pr(f\circ\mathcal{M}(S)\in T)+\delta.
\end{aligned}
$$

By definition we know that $f\circ\mathcal{M}$ is also (ϵ,δ)-differentially private. ☐

Proof of Theorem 4.1. Based on Lemma A.1, we know that $H_j(S_j,Q)$ is $(\epsilon/T,\delta/T)$-differentially private given fixed Q. Then we can deduce from Lemma A.2 that the adaptive composition

$$H_j(S_j)=\big(H_j\big(S_j,Q^{(1)}\big),\ldots H_j\big(S_j,Q^{(T)}\big)\big)$$

is (ϵ,δ)-differentially private. In Algorithm 1, the output Q^*_{dp} is obtained by post-processing of $H_j(S_j)$ and thus Q^*_{dp} is (ϵ,δ)-differentially private. ☐

B Appendix: Proof for the Main Theorem

B.1 Notations

Let $\mathbf{x}_1,\mathbf{x}_2,\ldots,\mathbf{x}_n\in\mathbb{R}^d$ be samples drawn from underlying distribution, where d is the dimension and n is the number of samples. Denote $S=\{\mathbf{x}_1,\mathbf{x}_2,\ldots,\mathbf{x}_n\}$ as the set of all the samples, and S_i as the sample set of ith data provider. Let $\Sigma\in\mathbb{R}^{d\times d}$ be the population covariance matrix of the generating distribution, and let $\widehat{\Sigma}\in\mathbb{R}^{d\times d}$ be sample covariance matrix, i.e. $\widehat{\Sigma}=(\sum_{i=1}^n\mathbf{x}_i\mathbf{x}_i^\top)/n$. Let the $d\times k$ matrix $Q^*\in\mathbb{R}^{d\times k}$ with orthogonal columns spans the k leading eigenspace of Σ.

Let $\mathcal{I}\subseteq\{1,\ldots d\}$ be an index set. For any matrix $M\in\mathbb{R}^{n\times m}$, denote $M_{\mathcal{I}}\in\mathbb{R}^{d\times d}$ as the restriction of M onto the rows and columns indexed by $\mathcal{I}$. Denote $M_{\mathcal{I},*}\in\mathbb{R}^{n\times m}$ as the matrix M restricted on rows indexed by $\mathcal{I}$, with value 0 on rows not indexed by $\mathcal{I}$. Denote $\lambda_k(M)$ as the k-th leading eigenvalue of any matrix M, and we simplify the notation $\lambda_k(\Sigma)$ as λ_k. Define $\widehat{Q}(\mathcal{I})\in\mathbb{R}^{d\times k}$ as a matrix with orthogonal columns which span the top k leading eigenspace of $\widehat{\Sigma}_{\mathcal{I}}$. Let $\widehat{Q}(\mathcal{I})^\perp$ be the matrix with orthogonal columns which span the subspace corresponding to $\lambda_{k+1}(\widehat{\Sigma}_{\mathcal{I}}),\ldots\lambda_d(\widehat{\Sigma}_{\mathcal{I}})$.

We use the notation $||Q^*||_{2,0}$ to describe the number of non-zero rows of Q^* under row-wise ℓ_2-norm. Let $\big\|Q^*_{i,*}\big\|_2$ be the ℓ_2-norm of the ith row of Q^*, then we define

$$||Q^*||_{2,0}=\big\|\big(\big\|Q^*_{1,*}\big\|_2,\big\|Q^*_{2,*}\big\|_2,\ldots,\big\|Q^*_{d,*}\big\|_2\big)\big\|_0,\tag{B.1}$$

which is the sparsity of row-support of Q^*.

B.2 Assumptions

Assumption B.1. Assume that there exists $\alpha\in(0,1)$ such that $0<\alpha<1-\gamma$ and $\tau>0$ such that

$$\frac{2kN\big(\sqrt{s}+\sqrt{k}+\tau\big)}{n\alpha}\le\frac{\lambda_k}{\sigma(\epsilon,\delta)}.\tag{B.2}$$

Here the noise-to-signal ratio parameter α is upper-bounded by

$$\frac{\alpha}{1-\alpha} \leq \frac{1}{1+2\sqrt{k}} \min\left\{\frac{1}{3}(1-\rho^2)^{3/2}, \frac{1}{6}(1-\rho^{1/4})\right\}. \tag{B.3}$$

We also suppose the choice of thresholding parameter $\widehat{s}$ satisfies the following condition

$$\widehat{s} = C_1 \max\left\{\left\lceil \frac{4k}{(\rho^{-1/2}-1)^2} \right\rceil, 1\right\} \cdot s^*,\ C_1 > 1. \tag{B.4}$$

And the sample size n is large enough such that for a positive constant $C_2 > 0$

$$\begin{aligned}\Psi(2\widehat{s}) &= C_2 \cdot \frac{\sqrt{\lambda_1 \lambda_{k+1}}}{\lambda_k - \lambda_{k+1}} \cdot \sqrt{\frac{\widehat{s} \cdot (k + \log d)}{n}} \\ &\leq \min\left\{\frac{1}{24}(1-\rho^2)^{3/2}, \frac{1}{6}(1-\rho^{1/4})\right\}. \end{aligned}\tag{B.5}$$

Initialization of the algorithm $Q^{(0)}$ satisfies

$$\|Q^{(0)\top} Q^{*\perp}\|_{\mathrm{F}} \leq \min\left\{\frac{\rho^{1/2}\sqrt{(1-\rho^{1/2})}}{2}, \frac{\rho^{1/2}}{4}\right\}. \tag{B.6}$$

Assumption Justifications.

- Let us remark that equation (B.2) essentially uses the parameter α, τ to bound the noise-to-signal ratio $\sigma(\epsilon,\delta)/\lambda_k$. Under mild privacy constraint when the signal-to-noise ratio $\lambda_k/\sigma(\epsilon,\delta)$ is sufficiently large or the sample number n is sufficiently large, we can find $\tau > 0$ and $\alpha \in (0, 1-\gamma)$ small enough to satisfy the constraints (B.2) (B.3). Note that (B.2) does not involve dimension d and thus this constraint on sample number n does not scale with d.
- The effective sample size is n/N in (B.2), which characterizes the effect of having samples stored in a distributed system with N data providers.
- The condition (B.4) on choice of sparsity parameter $\widehat{s}$ ensures that we would not lose too much information in the thresholding procedure where only $\widehat{s}$ rows are preserved and others are set to zero. When the effective eigengap $\rho = \gamma/(1-\alpha)$ is close to one, the parameter $\widehat{s}$ has to be comparatively large because the problem becomes ill-conditioned and we cannot afford to lose accuracy in the thresholding procedure.
- The assumption on good initial value for nonconvex optimization (B.6) is a common practice in the literature of sparse PCA and privacy-preserving PCA such as [47] and [18].

B.3 Sketch Of Proof

Now we present a sketch of proof for Theorem 4.3 together with three lemmas supporting the proof. Please see §B.1 for the meaning of notation $\widehat{Q}(\mathcal{I})^\perp$ and $Q^{*\perp}$ in the following lemmas.

Lemma B.2 analyzes the privacy preservation step in Algorithm 1 by presenting a contractive relationship between $\|V^{(t)\top}\widehat{Q}(\mathcal{I})^\perp\|_{\mathrm{F}}$ and $\|Q^{(t)\top}\widehat{Q}(\mathcal{I})^\perp\|_{\mathrm{F}}$. It is shown in (B.15) in Appendix §B.5 that

$$V^{(t)} = \mathrm{Orthogonalize}\big(\widehat{\Sigma} Q^{(t)} + \frac{N}{n} G^{(t)}\big).$$

The contractive relationship means that although Gaussian noise matrices are added, the noise is only effective in $\widehat{s}$ rows thanks to the thresholding procedure and can be controlled in high dimensional setting.

Lemma B.2. Let $\mathcal{I}$ be the index of the row-support of $Q^{(t)}$. If we have

$$\big\|Q^{(t)\top}\widehat{Q}(\mathcal{I})^\perp\big\|_{\mathrm{F}} < 1/2, \tag{B.7}$$

for $t = 1, 2, \ldots, T$ and assume that there exists $\alpha \in (0,1)$ such that $0 < \alpha < 1-\gamma$, and $\tau > 0$ such that (B.2) in Assumption B.1 is satisfied. Denote $\rho = \gamma/(1-\alpha) \in (0,1)$. We have the following result

$$\big\|V^{(t)\top}\widehat{Q}(\mathcal{I})^\perp\big\|_{\mathrm{F}} \leq \frac{\big\|Q^{(t)\top}\widehat{Q}(\mathcal{I})^\perp\big\|_{\mathrm{F}}}{\sqrt{1-\big\|Q^{(t)\top}\widehat{Q}(\mathcal{I})^\perp\big\|_{\mathrm{F}}^2}} \cdot \rho + \frac{\alpha}{2(1-\alpha)}, \tag{B.8}$$

holds for $t = 1, 2, \ldots, T$ with probability at least $1 - 2Te^{-2\tau^2}$.

Proof. Please see Appendix §B.5 for a detailed proof. □

The thresholding procedure in Algorithm 1 aims to impose sparsity in every iteration. However, as we do not know the true support of the k leading eigenspace, thresholding could bring in some extra error. The following lemma analyzes the thresholding error.

Lemma B.3. Remember that the true model sparsity s^* is defined as the sparsity level of the row support of Q^* whose columns span the leading k-dimensional principal subspace of Σ. Given the sparsity parameter $\widehat{s}$ in Algorithm 1, if $\sqrt{s^*/\widehat{s}} \leq 1$ and $\|V^{(t)\top}Q^{*\perp}\|_{\mathrm{F}} \leq 1/2$. We have

$$\big\|Q^{(t+1)\top}Q^{*\perp}\big\|_{\mathrm{F}} \leq \left(1 + 2\sqrt{\frac{ks^*}{\widehat{s}}}\right)\big\|V^{(t)\top}Q^{*\perp}\big\|_{\mathrm{F}}. \tag{B.9}$$

Proof. Please see Appendix §B.6 for a detailed proof. □

The above two lemmas analyze the privacy-preserving and thresholding steps in each iteration, based on which we can now proceed to prove the contractive property of each iteration in Algorithm 1. The function Ψ in (B.10) is defined in (B.5).

Lemma B.4. Under Assumption B.1, suppose we have

$$\left\|Q^{(t)\top}Q^{*\perp}\right\|_{\mathrm{F}} \le \min\Big\{\sqrt{(1-\rho^{1/2})}, 1/2\Big\},$$

where $C_1 \ge 1$ is a constant. We can show that

$$\begin{aligned}\left\|Q^{(t+1)\top}\widehat{Q}^{*\perp}\right\|_{\mathrm{F}} \le& \rho^{1/4}\cdot\left\|Q^{(t)\top}\widehat{Q}^{*\perp}\right\|_{\mathrm{F}} + 3\rho^{1/2}\cdot\Psi(\widehat{s})\\ &+\Big(1+2\sqrt{\frac{ks^*}{\widehat{s}}}\Big)\frac{\alpha}{2(1-\alpha)} \end{aligned}\tag{B.10}$$

with probability at least

$$1-2e^{-\tau^2/2}-4/(n-1)-1/d-6\log n/n-1/n. \tag{B.11}$$

Proof. The proof is based on Lemma B.2 and Lemma B.3. Please see Appendix §B.7 for details of proof. □

From the contractive property (B.10) of each iteration in Lemma B.4, we can derive the estimation error in Theorem 4.3. Please see Appendix §B.8 for a detailed proof.

B.4 Technical Preliminaries

We will first present some auxiliary lemmas here before going into the proof of Lemma B.2, Lemma B.3, Lemma B.4 and Theorem 4.3 in §B.5 §B.6, §B.7 and §B.8 respectively.

The estimation error of the eigenspace will be analyzed in terms of subspace distance between two k dimensional linear subspace $\mathcal{U}, \mathcal{V}$ in $\mathbb{R}^d$. Denote U and V as two $d\times k$ matrices whose orthonormal columns span the linear subspace $\mathcal{U}$ and $\mathcal{V}$. Denote $U^\perp$ and $V^\perp$ as two $d\times(d-k)$ matrices whose orthonormal columns span the linear subspace $\mathcal{U}^\perp$ and $\mathcal{V}^\perp$ orthogonal to $\mathcal{U}$ and $\mathcal{V}$ respectively. The orthogonal projection matrices for $\mathcal{U}, \mathcal{V}$ are denoted as Π_u, Π_v.

Lemma B.5. Singular values of $\Pi_u\Pi_v^\perp$ are zeros after the top k entries

$$s_1, s_2, \ldots, s_k, 0, 0, \cdots, 0.$$

Canonical angles between $\mathcal{U}$ and $\mathcal{V}$ are defined as

$$\theta_i(U,V) = \arcsin(s_i),\ i=1,2,\ldots,k.$$

Let $\Theta(U,V) = \mathrm{diag}(\theta_1,\theta_2,\ldots,\theta_k)$. The distance between $\mathcal{U}$ and $\mathcal{V}$ can be characterized as

$$\begin{aligned}\mathcal{D}[\mathcal{U},\mathcal{V}] &= \|\sin\Theta(U,V)\|_{\mathrm{F}} = \left\|U^\top V^\perp\right\|_{\mathrm{F}} = \left\|V^\top U^\perp\right\|_{\mathrm{F}}\\ &= \frac{1}{\sqrt{2}}\|\Pi_u-\Pi_v\|_{\mathrm{F}} = \|\Pi_u\Pi_v^\perp\|_{\mathrm{F}} = \|\Pi_u^\perp\Pi_v\|_{\mathrm{F}}.\end{aligned}$$

Besides we have the property similar to Pythagorean theorem

$$\|U^\top V^\perp\|_{\mathrm{F}}^2 + \|U^\top V\|_{\mathrm{F}}^2 = k. \tag{B.12}$$

Proof. Please see Theorem I.5.5 in [39] and Theorem 2.5.1 in [2] for details of proof. □

Before the proof we present a lemma here which will be useful in the proof of Lemma B.7.

Lemma B.6. For any $\mathcal{I}\subseteq\{1,2,\ldots,d\}$ with $|\mathcal{I}|\le d/2$, the condition

$$||\widehat{\Sigma}-\Sigma||_{2,|\mathcal{I}|} \le C_1\lambda_1\sqrt{\frac{|\mathcal{I}|\log d}{n}}$$

holds with probability at least $1-1/n$.

Proof. The proof follows from Lemma 3.2.4 in [40]. □

Lemma B.7. Assume that there exists $\alpha\in(0,1)$ such that $0<\alpha<1-\gamma$ and $\tau>0$ sufficiently large such that

$$\sigma \le \frac{\alpha\lambda_{k+1}}{2k\big(\sqrt{\widehat{s}}+\sqrt{k}+\tau\big)}. \tag{B.13}$$

Let matrix G be a $\widehat{s}\times k$ matrix with i.i.d. Gaussian $N(0,\sigma^2)$ entries. Let $\mathcal{I}$ be the row support in thresholding step in Algorithm 1. We have

$$\left\|G^{(t)}\right\|_{\mathrm{F}} \le \alpha\lambda_k(\widehat{\Sigma}_{\mathcal{I}})/2 \tag{B.14}$$

holds for all $t=1,2,\ldots,T$ with probability at least $1-2Te^{-\tau^2/2}$.

Proof. Based on the Tracy-Widom fluctuations in [35], we have

$$\mathbb{P}\Big\{\left\|G^{(t)}\right\|_2 > \sigma(\sqrt{\widehat{s}}+\sqrt{k}+\tau)\Big\} \le 2e^{-\tau^2/2}.$$

Since we have assumption (B.13), it follows that

$$\begin{aligned}&\mathbb{P}\Big\{\left\|G^{(t)}\right\|_2 > \frac{\alpha\lambda_{k+1}}{2k}\Big\}\\ \le&\mathbb{P}\Big\{\left\|G^{(t)}\right\|_2 > \sigma(\sqrt{\widehat{s}}+\sqrt{k}+\tau)\Big\} \le 2e^{-\tau^2/2}.\end{aligned}$$

Besides we have $\left\|G^{(t)}\right\|_{\mathrm{F}} \le k\left\|G^{(t)}\right\|_2$, it follows that

$$\begin{aligned}&\mathbb{P}\Big\{\left\|G^{(t)}\right\|_{\mathrm{F}} > \frac{\alpha\lambda_{k+1}}{2}\Big\}\\ \le&\mathbb{P}\Big\{k\left\|G^{(t)}\right\|_2 > \frac{\alpha\lambda_{k+1}}{2}\Big\} \le 2e^{-\tau^2/2}.\end{aligned}$$

From Lemma B.6 we have

$$\lambda_k(\widehat{\Sigma}_{\mathcal{I}}) \ge \lambda_k - ||\widehat{\Sigma}-\Sigma||_{2,|\mathcal{I}|} \ge \lambda_k - \frac{\lambda_k-\lambda_{k+1}}{4} \ge \lambda_{k+1},$$

which implies

$$\begin{aligned}&\mathbb{P}\Big\{\left\|G^{(t)}\right\|_{\mathrm{F}} > \frac{\alpha\lambda_k(\widehat{\Sigma}_{\mathcal{I}})}{2}\Big\}\\ \le&\mathbb{P}\Big\{\left\|G^{(t)}\right\|_{\mathrm{F}} > \frac{\alpha\lambda_{k+1}}{2}\Big\} \le 2e^{-\tau^2/2}.\end{aligned}$$

Hence the probability that (B.14) holds for $t=1,2,\ldots,T$ is $1-2Te^{-\tau^2/2}$. □

B.5 Proof of Lemma B.2

Proof of Lemma B.2. For the Gaussian noise matrix $G_i^{(t)}$ used by data owner i in iteration t, we simply denote it as $G^{(t)}$. Remind that the notation $\widehat{Q}(\mathcal{I})$ is a row-sparse matrix with row support in $\mathcal{I}$. We use $V_{\mathcal{I},*} \in \mathbb{R}^{d\times k}$ to denote the matrix whose rows are restricted on the index set $\mathcal{I}$, and it is set to zeros for the rows not indexed by $\mathcal{I}$. From (B.12) we have

$$\begin{aligned}&\left\|V^{(t)\top}\widehat{Q}(\mathcal{I})^{\perp}\right\|_{\mathrm{F}}\\ =&\left(k-\left\|V^{(t)\top}\widehat{Q}(\mathcal{I})\right\|_{\mathrm{F}}^2\right)^{1/2}=\left(k-\left\|V_{\mathcal{I},*}^{(t)\top}\widehat{Q}(\mathcal{I})\right\|_{\mathrm{F}}^2\right)^{1/2}.\end{aligned}$$

With the goal to analyze the subspace distance $\left\|V^{(t)\top}\widehat{Q}(\mathcal{I})^{\perp}\right\|_{\mathrm{F}}$, we can focus on $V_{\mathcal{I},*}^{(t)}$ instead of $V^{(t)}$ in our analysis. Note that

$$\begin{aligned}H_i^{(t)}&=\widehat{\Sigma}_i Q^{(t)}+\frac{1}{n_i}G^{(t)},\\ V^{(t)}R_1=K^{(t)}&=\frac{\sum_{i=1}^N n_i H_i^{(t)}}{\sum_{i=1}^N n_i}=\widehat{\Sigma}\cdot Q^{(t)}+\frac{N}{n}G^{(t)},\end{aligned}\tag{B.15}$$

where $\widehat{\Sigma}$ is the sample covariance for the samples of all the N data owners. Besides, since $G^{(t)}$ is a random matrix with i.i.d entries $N(0,\sigma(\epsilon,\delta)^2)$. We will denote $\sigma=N\sigma(\epsilon,\delta)/n$ and assume that $G^{(t)}$ is entrywise i.i.d with $N(0,\sigma^2)$. We now have

$$V^{(t)}R_1=K^{(t)}=\widehat{\Sigma}\cdot Q^{(t)}+G^{(t)},$$

and the standard deviation $\sigma=N\sigma(\epsilon,\delta)/n$ satisfies the requirement (B.13) because of the (B.2) in Assumption B.1.

By taking restriction on the rows indexed by $\mathcal{I}$, we obtain

$$V_{\mathcal{I},*}^{(t)}\cdot R_1=\widehat{\Sigma}_{\mathcal{I}}\cdot Q^{(t)}+G_{\mathcal{I},*}^{(t)}.\tag{B.16}$$

Let the eigen-decomposition of $\widehat{\Sigma}_{\mathcal{I}}$ be

$$\widehat{\Sigma}_{\mathcal{I}}=\widehat{Q}(\mathcal{I})\Lambda_0\widehat{Q}(\mathcal{I})^\top+\widehat{Q}(\mathcal{I})^{\perp}\Lambda_1\left[\widehat{Q}(\mathcal{I})^{\perp}\right]^\top,$$

where Λ_0 is the diagonal matrix of the top k eigenvalues of $\widehat{\Sigma}_{\mathcal{I}}$, and Λ_1 is the diagonal matrix of the rest of eigenvalues in decreasing order.

Equation (B.16) can be written as

$$\begin{aligned}V_{\mathcal{I},*}^{(t)}\cdot R_1=&\widehat{Q}(\mathcal{I})\Lambda_0\widehat{Q}(\mathcal{I})^\top\cdot Q^{(t)}\\ &+\widehat{Q}(\mathcal{I})^{\perp}\Lambda_1\left[\widehat{Q}(\mathcal{I})^{\perp}\right]^\top\cdot Q^{(t)}+G_{\mathcal{I},*}^{(t)}.\end{aligned}\tag{B.17}$$

If $\widehat{Q}(\mathcal{I})^\top$ and $[\widehat{Q}(\mathcal{I})^{\perp}]^\top$ are used to multiply both sides of (B.17) respectively, we can have the following two equations

$$\widehat{Q}(\mathcal{I})^\top V_{\mathcal{I},*}^{(t)}R_1=\Lambda_0\widehat{Q}(\mathcal{I})^\top Q^{(t)}+\widehat{Q}(\mathcal{I})^\top G_{\mathcal{I},*}^{(t)},\tag{B.18}$$

$$[\widehat{Q}(\mathcal{I})^{\perp}]^\top V_{\mathcal{I},*}^{(t)}R_1=\Lambda_1[\widehat{Q}(\mathcal{I})^{\perp}]^\top Q^{(t)}+[\widehat{Q}(\mathcal{I})^{\perp}]^\top G_{\mathcal{I},*}^{(t)}.\tag{B.19}$$

Note that $\widehat{Q}(\mathcal{I})^\top Q^{(t)}$ and $\widehat{Q}(\mathcal{I})^\top G_{\mathcal{I},*}^{(t)}$ are both $k\times k$ square matrices. We prove that the right-hand-side of equation (B.18) is non-singular by showing that the smallest singular value of the right-hand-side is positive. From the perturbation theory for singular value decomposition in [38], we have

$$\begin{aligned}&\sigma_k\left[\Lambda_0\widehat{Q}(\mathcal{I})^\top Q^{(t)}+\widehat{Q}(\mathcal{I})^\top G_{\mathcal{I},*}^{(t)}\right]\\ \geq&\sigma_k\left[\Lambda_0\widehat{Q}(\mathcal{I})^\top Q^{(t)}\right]-\left\|\widehat{Q}(\mathcal{I})^\top G_{\mathcal{I},*}^{(t)}\right\|_2\\ \geq&\lambda_k(\widehat{\Sigma}_{\mathcal{I}})\cdot\sigma_k\left[\widehat{Q}(\mathcal{I})^\top Q^{(t)}\right]-\left\|\widehat{Q}(\mathcal{I})^\top G_{\mathcal{I},*}^{(t)}\right\|_2.\end{aligned}\tag{B.20}$$

According to the Pythagorean relation (B.12), we can derive that

$$\begin{aligned}\sigma_k\left[\widehat{Q}(\mathcal{I})^T Q^{(t)}\right]&=\sqrt{1-\sigma_1\left[(\widehat{Q}(\mathcal{I})^{\perp})^\top Q^{(t)}\right]^2}\\ &\geq\sqrt{1-\left\|(\widehat{Q}(\mathcal{I})^{\perp})^\top Q^{(t)}\right\|_{\mathrm{F}}^2}>0.\end{aligned}$$

Since $\widehat{Q}(\mathcal{I})\in\mathbb{R}^{d\times k}$ has orthonormal columns, the following equation holds with probability at least $1-2Te^{-\tau^2/2}$

$$\begin{aligned}&\left\|\widehat{Q}(\mathcal{I})^T\cdot G_{\mathcal{I},*}^{(t)})\right\|_2\leq\left\|G_{\mathcal{I},*}^{(t)}\right\|_2\\ \leq&\alpha\lambda_k(\widehat{\Sigma}_{\mathcal{I}})\sqrt{1-\left\|(\widehat{Q}(\mathcal{I})^{\perp})^\top Q^{(t)}\right\|_{\mathrm{F}}^2}<\alpha\lambda_k(\widehat{\Sigma}_{\mathcal{I}})/2.\end{aligned}$$

The last inequality follows from the fact that $\left\|(\widehat{Q}(\mathcal{I})^{\perp})^\top Q^{(t)}\right\|_{\mathrm{F}}<1/2$, and the high probability claim follows from Lemma B.7.

Putting the above two equations into equation (B.20) gives

$$\begin{aligned}&\sigma_k\left[\Lambda_0\widehat{Q}(\mathcal{I})^\top Q^{(t)}+\widehat{Q}(\mathcal{I})^\top G_{\mathcal{I},*}^{(t)}\right]\\ \geq&(1-\alpha)\lambda_k(\widehat{\boldsymbol{\Sigma}}_{\mathcal{I}})\sqrt{1-\left\|(\widehat{Q}(\mathcal{I})^{\perp})^\top Q^{(t)}\right\|_{\mathrm{F}}^2}>0,\end{aligned}\tag{B.21}$$

and thus $\Lambda_0\widehat{Q}(\mathcal{I})^\top Q^{(t)}+\widehat{Q}(\mathcal{I})^\top G_{\mathcal{I},*}^{(t)}$ is non-singular. From equation (B.21) and (B.18) we have

$$R_1^{-1}=\left[\Lambda_0\widehat{Q}(\mathcal{I})^\top Q^{(t)}+\widehat{Q}(\mathcal{I})^\top G_{\mathcal{I},*}^{(t)}\right]^{-1}\cdot\widehat{Q}(\mathcal{I})^\top V_{\mathcal{I},*}^{(t)}.$$

Combining the above equation with equation (B.19), we have

$$\begin{aligned}&\left[\widehat{Q}(\mathcal{I})^{\perp}\right]^\top V_{\mathcal{I},*}^{(t)}=\left(\Lambda_1[\widehat{Q}(\mathcal{I})^{\perp}]^\top Q^{(t)}+[\widehat{Q}(\mathcal{I})^{\perp}]^\top G_{\mathcal{I},*}^{(t)}\right)\\ \times&\left[\Lambda_0\widehat{Q}(\mathcal{I})^\top Q^{(t)}+\widehat{Q}(\mathcal{I})^\top G_{\mathcal{I},*}^{(t)}\right]^{-1}\cdot\widehat{Q}(\mathcal{I})^\top V_{\mathcal{I},*}^{(t)}.\end{aligned}\tag{B.22}$$

The Frobenius norm of the right-hand-side of equation (B.22) can be upper bounded by

$$\begin{aligned}&\big\|\big(\Lambda_1[\widehat{Q}(\mathcal{I})^{\perp}]^{\top}Q^{(t)}+[\widehat{Q}(\mathcal{I})^{\perp}]^{\top}G_{\mathcal{I},*}^{(t)}\big)\big\|_{\mathrm{F}}\\ \times&\big\|\big[\Lambda_0\widehat{Q}(\mathcal{I})^{\top}Q^{(t)}+\widehat{Q}(\mathcal{I})^{\top}G_{\mathcal{I},*}^{(t)}\big]^{-1}\big\|_2\big\|\widehat{Q}(\mathcal{I})^{\top}V_{\mathcal{I},*}^{(t)}\big\|_2.\end{aligned} \tag{B.23}$$

The first part can be further controlled in the following way

$$\begin{aligned}&\big\|\Lambda_1[\widehat{Q}(\mathcal{I})^{\perp}]^{\top}Q^{(t)}+[\widehat{Q}(\mathcal{I})^{\perp}]^{\top}G_{\mathcal{I},*}^{(t)}\big\|_{\mathrm{F}}\\ \leq&\big\|\Lambda_1[\widehat{Q}(\mathcal{I})^{\perp}]^{\top}Q^{(t)}\big\|_{\mathrm{F}}+\big\|[\widehat{Q}(\mathcal{I})^{\perp}]^{\top}G_{\mathcal{I},*}^{(t)}\big\|_{\mathrm{F}}\\ \leq&\lambda_{k+1}(\widehat{\Sigma}_{\mathcal{I}})\big\|[\widehat{Q}(\mathcal{I})^{\perp}]^{\top}Q^{(t)}\big\|_{\mathrm{F}}+\big\|G_{\mathcal{I},*}^{(t)}\big\|_{\mathrm{F}}.\end{aligned} \tag{B.24}$$

The second part can also be upper bounded as

$$\begin{aligned}&\big\|\big[\Lambda_0\widehat{Q}(\mathcal{I})^{\top}Q^{(t)}+\widehat{Q}(\mathcal{I})^{\top}G_{\mathcal{I},*}^{(t)}\big]^{-1}\big\|_2\\ =&\frac{1}{\sigma_k\big[\Lambda_0\widehat{Q}(\mathcal{I})^{\top}Q^{(t)}+\widehat{Q}(\mathcal{I})^{\top}G_{\mathcal{I},*}^{(t)}\big]}\\ \leq&\frac{1}{(1-\alpha)\lambda_k(\widehat{\Sigma}_{\mathcal{I}})\sqrt{1-\big\|(\widehat{Q}(\mathcal{I})^{\perp})^{\top}Q^{(t)}\big\|_{\mathrm{F}}^2}},\end{aligned} \tag{B.25}$$

where the second inequality follows from equation (B.21). The third part is obviously upper bounded

$$\big\|\widehat{Q}(\mathcal{I})^{\top}V_{\mathcal{I},*}^{(t)}\big\|_2\leq\big\|\widehat{Q}(\mathcal{I})^{\top}\big\|_2\big\|V_{\mathcal{I},*}^{(t)}\big\|_2\leq 1. \tag{B.26}$$

After we put equation (B.24), (B.25), (B.26) into (B.23), equation (B.22) can be upper-bounded by

$$\begin{aligned}&\frac{\lambda_{k+1}(\widehat{\Sigma}_{\mathcal{I}})\big\|\big[\widehat{Q}(\mathcal{I})^{\perp}\big]^{\top}Q^{(t)}\big\|_{\mathrm{F}}+\big\|G_{\mathcal{I},*}^{(t)}\big\|_{\mathrm{F}}}{(1-\alpha)\lambda_k(\widehat{\Sigma}_{\mathcal{I}})\sqrt{1-\big\|\big[\widehat{Q}(\mathcal{I})^{\perp}\big]^{\top}Q^{(t)}\big\|_{\mathrm{F}}^2}}\\ \leq&\frac{\lambda_{k+1}(\widehat{\Sigma}_{\mathcal{I}})\big\|\big[\widehat{Q}(\mathcal{I})^{\perp}\big]^{\top}Q^{(t)}\big\|_{\mathrm{F}}+\alpha\lambda(\widehat{\Sigma}_{\mathcal{I}})/2}{(1-\alpha)\lambda_k(\widehat{\Sigma}_{\mathcal{I}})\sqrt{1-\big\|\big[\widehat{Q}(\mathcal{I})^{\perp}\big]^{\top}Q^{(t)}\big\|_{\mathrm{F}}^2}}\\ \leq&\frac{1}{1-\alpha}\frac{\big\|\big[\widehat{Q}(\mathcal{I})^{\perp}\big]^{\top}Q^{(t)}\big\|_{\mathrm{F}}}{\sqrt{1-\big\|\big[\widehat{Q}(\mathcal{I})^{\perp}\big]^{\top}Q^{(t)}\big\|_{\mathrm{F}}^2}}\cdot\frac{\lambda_{k+1}(\widehat{\Sigma}_{\mathcal{I}})}{\lambda_k(\widehat{\Sigma}_{\mathcal{I}})}\\ &+\frac{\alpha}{2(1-\alpha)\sqrt{1-\big\|\big[\widehat{Q}(\mathcal{I})^{\perp}\big]^{\top}Q^{(t)}\big\|_{\mathrm{F}}^2}}\\ \leq&\frac{1}{1-\alpha}\frac{\big\|\big[\widehat{Q}(\mathcal{I})^{\perp}\big]^{\top}Q^{(t)}\big\|_{\mathrm{F}}}{\sqrt{1-\big\|\big[\widehat{Q}(\mathcal{I})^{\perp}\big]^{\top}Q^{(t)}\big\|_{\mathrm{F}}^2}}\cdot\frac{\lambda_{k+1}(\Sigma)+||\widehat{\Sigma}-\Sigma||_{2,|\mathcal{I}|}}{\lambda_k(\Sigma)-||\widehat{\Sigma}-\Sigma||_{2,|\mathcal{I}|}}\\ &+\frac{\alpha}{\sqrt{3}(1-\alpha)}\\ \leq&\frac{1}{1-\alpha}\frac{\big\|\big[\widehat{Q}(\mathcal{I})^{\perp}\big]^{\top}Q^{(t)}\big\|_{\mathrm{F}}}{\sqrt{1-\big\|\big[\widehat{Q}(\mathcal{I})^{\perp}\big]^{\top}Q^{(t)}\big\|_{\mathrm{F}}^2}}\frac{3\lambda_{k+1}(\Sigma)+\lambda_k(\Sigma)}{\lambda_{k+1}(\Sigma)+3\lambda_k(\Sigma)}\\ &+\frac{\alpha}{\sqrt{3}(1-\alpha)}\\ =&\frac{\big\|\big[\widehat{Q}(\mathcal{I})^{\perp}\big]^{\top}Q^{(t)}\big\|_{\mathrm{F}}}{\sqrt{1-\big\|\big[\widehat{Q}(\mathcal{I})^{\perp}\big]^{\top}Q^{(t)}\big\|_{\mathrm{F}}^2}}\cdot\rho+\frac{\alpha}{\sqrt{3}(1-\alpha)}.\end{aligned} \tag{B.27}$$

Since the columns of $\widehat{Q}(\mathcal{I})^{\perp}$ are eigenvectors of the row sparse matrix $\widehat{\Sigma}_{\mathcal{I}}$, the rows of $\widehat{Q}(\mathcal{I})^{\perp}$ must also be restricted on the set of index $\mathcal{I}$, and thus

$$\big\|\widehat{Q}(\mathcal{I})^{\perp}V^{(t)}\big\|_{\mathrm{F}}=\big\|\widehat{Q}(\mathcal{I})^{\perp}V_{\mathcal{I},*}^{(t)}\big\|_{\mathrm{F}}.$$

Based on the above equation, (B.27) and (B.22) we have

$$\big\|\widehat{Q}(\mathcal{I})^{\perp}V^{(t)}\big\|_{\mathrm{F}}\leq\frac{\big\|\big[\widehat{Q}(\mathcal{I})^{\perp}\big]^{\top}Q^{(t)}\big\|_{\mathrm{F}}}{\sqrt{1-\big\|\big[\widehat{Q}(\mathcal{I})^{\perp}\big]^{\top}Q^{(t)}\big\|_{\mathrm{F}}^2}}\cdot\rho+\frac{\alpha}{\sqrt{3}(1-\alpha)},$$

which is equivalent to

$$\big\|V^{(t)\top}\widehat{Q}(\mathcal{I})^{\perp}\big\|_{\mathrm{F}}\leq\frac{\big\|Q^{(t)\top}\widehat{Q}(\mathcal{I})^{\perp}\big\|_{\mathrm{F}}}{\sqrt{1-\big\|Q^{(t)\top}\widehat{Q}(\mathcal{I})^{\perp}\big\|_{\mathrm{F}}^2}}\cdot\rho+\frac{\alpha}{\sqrt{3}(1-\alpha)}.$$

The entire proof holds with probability at least $1-Te^{-\tau^2/2}$. □

B.6 Proof of Lemma B.3

Proof. In this proof we will analyze the error in each iteration induced by the thresholding procedure. Denote $\mathcal{I}$ as the set of index used for thresholding in the current iteration. Also let $\mathcal{I}^*$ to be ground truth of the row support of Q^* whose columns span the leading k eigenspace of the population covariance matrix Σ in our model. Under this notation, we have $\widehat{s} = |\mathcal{I}|$ and $s = |S|$. Since the thresholding comes from the difference between $\mathcal{I}$ and $\mathcal{I}^*$, our analysis relies on these critical quantities

$$\mathcal{I}_1 = \mathcal{I}^*\backslash\mathcal{I}, \quad \mathcal{I}_2 = \mathcal{I}^* \cap \mathcal{I}, \quad \mathcal{I}_3 = \mathcal{I}\backslash\mathcal{I}^*.$$

Now we compare the approximation accuracy between $\widetilde{Q}^{(t)}$ and $V^{(t)}$

$$\begin{aligned}&\big\|Q^{*\top}\widetilde{Q}^{(t)}\big\|_{\mathrm{F}} = \big\|Q^{*\top}V^{(t)} + Q^{*\top}\big(\widetilde{Q}^{(t)} - V^{(t)}\big)\big\|_{\mathrm{F}}\\ \leq&\big\|Q^{*\top}V^{(t)}\big\|_{\mathrm{F}} - \|Q^{*\top}\big(\widetilde{Q}^{(t)} - V^{(t)}\big)\|_{\mathrm{F}}\\ =&\big\|Q^{*\top}V^{(t)}\big\|_{\mathrm{F}} - \big\|(Q^*_{\mathcal{I}_1,*})^\top V^{(t)}_{\mathcal{I}_1,*}\big\|_{\mathrm{F}}. \end{aligned} \tag{B.28}$$

On the other hand, we can easily get the upper bound

$$\begin{aligned}&\big\|Q^{*\top}\widetilde{Q}^{(t)}\big\|_{\mathrm{F}} = \big\|Q^{*\top}Q^{(t)}R_2\big\|_{\mathrm{F}} \leq \big\|Q^{*\top}Q^{(t)}\big\|_{\mathrm{F}} \cdot \big\|R_2\big\|_2\\ =&\big\|Q^{*\top}Q^{(t)}\big\|_{\mathrm{F}} \cdot \big\|Q^{(t+1)}R_2\big\|_2 = \big\|Q^{*\top}Q^{(t)}\big\|_{\mathrm{F}} \cdot \big\|\widetilde{Q}^{(t)}\big\|_2\\ =&\big\|Q^{*\top}Q^{(t)}\big\|_{\mathrm{F}} \cdot \big\|V^{(t)}_{\mathcal{I},*}\big\|_2 \leq \big\|Q^{*\top}Q^{(t)}\big\|_{\mathrm{F}} \cdot \big\|V^{(t)}\big\|_2\\ =&\big\|Q^{*\top}Q^{(t)}\big\|_{\mathrm{F}}, \end{aligned} \tag{B.29}$$

where we use the facts that $Q^{(t+1)}$ has orthogonal columns and

$$\begin{aligned}\big\|V^{(t)}_{\mathcal{I},*}\big\|_2 =& \max_{||x||_2=1}\big\|x^\top V^{(t)}_{\mathcal{I},*}\big\|_2\\ \leq& \max_{||x||_2=1}\big\|x^\top V^{(t)}\big\|_2 = \big\|V^{(t)}_2\big\|_2.\end{aligned}$$

Based on (B.28) and (B.29) we have

$$\big\|Q^{*\top}Q^{(t)}\big\|_{\mathrm{F}} \geq \big\|Q^{*\top}V^{(t)}\big\|_{\mathrm{F}} - \big\|Q^*_{\mathcal{I}_1,*}\big\|_2\big\|V^{(t)}_{\mathcal{I}_1,*}\big\|_{\mathrm{F}}. \tag{B.30}$$

First of all we want to show the following

$$\big\|Q^*_{\mathcal{I}_1,*}\big\|_2 \leq \big\|Q^{(t+1)\top}Q^{*\perp}\big\|_{\mathrm{F}}, \tag{B.31}$$

which is very intuitive because when $Q^{(t+1)}$ is very close in subspace distance to Q^*, the index set $\mathcal{I}$ selected in the thresholding procedure would cover the true row support $\mathcal{I}^*$ effectively and the thus index set $\mathcal{I}_1 = \mathcal{I}^*\backslash\mathcal{I}$ must only over rows with very small norm. Hence the norm of $Q^*_{\mathcal{I}_1,*}$ can be controlled by the subspace distance. To be precise with our reasoning, first notice that

$$\begin{aligned}&\big\|Q^*_{\mathcal{I}_1,*}\big\|_{\mathrm{F}}^2 + \big\|Q^*_{\mathcal{I}_2,*}\big\|_{\mathrm{F}}^2 = \big\|Q^*_{\mathcal{I}^*,*}\big\|_{\mathrm{F}}^2\\ =&\big\|Q^*\big\|_{\mathrm{F}}^2 \leq k \cdot \big\|Q^*\big\|_2^2 = k.\end{aligned} \tag{B.32}$$

We also have

$$\begin{aligned}&\big\|(Q^*)^\top Q^{(t+1)}\big\|_{\mathrm{F}} = \big\|(Q^*_{\mathcal{I}_2,*})^\top Q^{(t+1)}\big\|_{\mathrm{F}}\\ \leq&\big\|Q^*_{\mathcal{I}_2,*}\big\|_{\mathrm{F}} \cdot \big\|Q^{(t+1)}\big\|_2 = \big\|Q^*_{\mathcal{I}_2,*}\big\|_{\mathrm{F}}.\end{aligned} \tag{B.33}$$

If we combine the two equations (B.32) and (B.33), we have

$$\begin{aligned}&\big\|Q^*_{\mathcal{I}_1,*}\big\|_2 \leq \big\|Q^*_{\mathcal{I}_1,*}\big\|_{\mathrm{F}} \leq \sqrt{k - \big\|Q^*_{\mathcal{I}_2,*}\big\|_{\mathrm{F}}^2}\\ \leq&\sqrt{k - \big\|(Q^*)^\top Q^{(t+1)}\big\|_{\mathrm{F}}^2} = \big\|Q^{(t+1)\top}Q^{*\perp}\big\|_{\mathrm{F}},\end{aligned}$$

which is exactly what we want to show in (B.31). Secondly we also have to give an upper bound for $\big\|V^{(t)}_{\mathcal{I}_1,*}\big\|_{\mathrm{F}}$. Following similar arguments in (B.32), we have

$$\begin{aligned}&\big\|V^{(t)}_{\mathcal{I}_1\cup\mathcal{I}_2,*}\big\|_{\mathrm{F}}^2 + \big\|V^{(t)}_{\mathcal{I}_3,*}\big\|_{\mathrm{F}}^2 = \big\|V^{(t)}_{\mathcal{I}_1\cup\mathcal{I}_2\cup\mathcal{I}_3}\big\|_{\mathrm{F}}\\ \leq&\big\|V^{(t)}\big\|_{\mathrm{F}}^2 \leq k \cdot \big\|V^{(t)}\big\|_2^2 = k,\end{aligned}$$

based on which we can show that

$$\begin{aligned}&k - \big\|V^{(t)}_{\mathcal{I}_3,*}\big\|_{\mathrm{F}}^2 \geq \big\|V^{(t)}_{\mathcal{I}_1\cup\mathcal{I}_2,*}\big\|_{\mathrm{F}}^2 \geq \big\|Q^{*\top}V^{(t)}_{\mathcal{I}_1\cup\mathcal{I}_2,*}\big\|_{\mathrm{F}}^2\\ =&\big\|Q^{*\top}V^{(t)}_{\mathcal{I}^*,*}\big\|_{\mathrm{F}}^2 = \big\|Q^{*\top}V^{(t)}\big\|_{\mathrm{F}}^2.\end{aligned} \tag{B.34}$$

Besides, we also need the following inequality based on the definition of the thresholding procedure

$$\begin{aligned}&\frac{1}{|\mathcal{I}_1|}\big\|V^{(t)}_{\mathcal{I}_1,*}\big\|_{\mathrm{F}}^2 = \frac{1}{|\mathcal{I}_1|}\sum_{i\in\mathcal{I}_1}\big\|V^{(t)}_{i,*}\big\|_2^2\\ \leq&\frac{1}{|\mathcal{I}_3|}\sum_{i\in\mathcal{I}_3}\big\|V^{(t)}_{i,*}\big\|_2^2 = \frac{1}{|\mathcal{I}_3|}\big\|V^{(t)}_{\mathcal{I}_3,*}\big\|_{\mathrm{F}}^2.\end{aligned} \tag{B.35}$$

Combining (B.34) and (B.35), and note the fact that $a/b \leq (a+c)/(b+c)$ for $0 < a \leq b, c > 0$, we have

$$\begin{aligned}&\big\|V^{(t)}_{\mathcal{I}_1,*}\big\|_{\mathrm{F}}^2 \leq \frac{|\mathcal{I}_1|}{|\mathcal{I}_3|}\big\|V^{(t)}_{\mathcal{I}_3,*}\big\|_{\mathrm{F}}^2 \leq \frac{|\mathcal{I}_1| + |\mathcal{I}_2|}{|\mathcal{I}_3| + |\mathcal{I}_2|}\big\|V^{(t)}_{\mathcal{I}_3,*}\big\|_{\mathrm{F}}^2\\ =&\frac{s^*}{\widehat{s}}\big\|V^{(t)}_{\mathcal{I}_3,*}\big\|_{\mathrm{F}}^2 \leq \frac{s^*}{\widehat{s}}\big(k - \big\|Q^{*\top}V^{(t)}\big\|_{\mathrm{F}}^2\big).\end{aligned} \tag{B.36}$$

By the Pythagorean property (B.12), we can derive the upper bound as

$$\big\|V^{(t)}_{\mathcal{I}_1,*}\big\|_{\mathrm{F}} \leq \sqrt{\frac{s^*}{\widehat{s}}}\big\|V^{(t)\top}Q^{*\perp}\big\|_{\mathrm{F}}. \tag{B.37}$$

Finally we can prove the lemma with (B.30), (B.31) and (B.37). Simply putting (B.31) and (B.37) into (B.30), we have

$$\begin{aligned}\big\|Q^{*\top}Q^{(t+1)}\big\|_{\mathrm{F}} \geq& \big\|Q^{*\top}V^{(t)}\big\|_{\mathrm{F}}\\ &- \sqrt{\frac{s^*}{\widehat{s}}}\big\|V^{(t)\top}Q^{*\perp}\big\|_{\mathrm{F}}\big\|Q^{(t+1)\top}Q^{*\perp}\big\|_{\mathrm{F}}.\end{aligned} \tag{B.38}$$

By our assumption $\big\|V^{(t)\top}Q^{*\perp}\big\|_{\mathrm{F}} < 1/2$ and (B.12), we have

$$\begin{aligned}&\big\|Q^{*\top}V^{(t)}\big\|_{\mathrm{F}} = \big\|V^{(t)\top}Q^*\big\|_{\mathrm{F}}\\ =&\sqrt{k - \big\|V^{(t)\top}Q^{*\perp}\big\|_{\mathrm{F}}^2} > \sqrt{k - 1/2}\\ \geq&\sqrt{k/2} \geq \big\|Q^{(t+1)\top}Q^{*\perp}\big\|_{\mathrm{F}}/\sqrt{2}.\end{aligned} \tag{B.39}$$

Based on the above equation and $s^* < \widehat{s}$, it can be shown that the right-hand side of (B.38) is non-negative

$$\begin{aligned}&\big\|Q^{*\top}V^{(t)}\big\|_{\mathrm{F}} - \sqrt{\frac{s^*}{\widehat{s}}}\big\|V^{(t)\top}Q^{*\perp}\big\|_{\mathrm{F}}\big\|Q^{(t+1)\top}Q^{*\perp}\big\|_{\mathrm{F}}\\ &> \big\|Q^{*\top}V^{(t)}\big\|_{\mathrm{F}} - \frac{1}{\sqrt{2}}\big\|Q^{(t+1)\top}Q^{*\perp}\big\|_{\mathrm{F}} \geq 0,\end{aligned}$$

where the last inequality follows from (B.39). If we take square of both sides of (B.38), we have

$$\begin{aligned}&\big\|Q^{*\top}Q^{(t+1)}\big\|_{\mathrm{F}}^2 \geq \big\|Q^{*\top}V^{(t)}\big\|_{\mathrm{F}}^2\\ &- 2\big\|Q^{*\top}V^{(t)}\big\|_{\mathrm{F}}\sqrt{\frac{s^*}{\widehat{s}}}\big\|V^{(t)\top}Q^{*\perp}\big\|_{\mathrm{F}}\big\|Q^{(t+1)\top}Q^{*\perp}\big\|_{\mathrm{F}}\\ &+ \Big(\sqrt{\frac{s^*}{\widehat{s}}}\big\|V^{(t)\top}Q^{*\perp}\big\|_{\mathrm{F}}\big\|Q^{(t+1)\top}Q^{*\perp}\big\|_{\mathrm{F}}\Big)^2\\ &\geq \big\|Q^{*\top}V^{(t)}\big\|_{\mathrm{F}}^2\\ &- 2\big\|Q^{*\top}V^{(t)}\big\|_{\mathrm{F}}\sqrt{\frac{s^*}{\widehat{s}}}\big\|V^{(t)\top}Q^{*\perp}\big\|_{\mathrm{F}}\big\|Q^{(t+1)\top}Q^{*\perp}\big\|_{\mathrm{F}}.\end{aligned}$$

Again we use the Pythagorean property (B.12) and obtain

$$\begin{aligned}&\big\|Q^{(t+1)\top}Q^{*\perp}\big\|_{\mathrm{F}}^2 = k - \big\|Q^{*\top}Q^{(t)}\big\|_{\mathrm{F}}^2\\ \leq& k - \big\|Q^{*\top}V^{(t)}\big\|_{\mathrm{F}}^2\\ &+ 2\big\|Q^{*\top}V^{(t)}\big\|_{\mathrm{F}}\sqrt{\frac{s^*}{\widehat{s}}}\big\|V^{(t)\top}Q^{*\perp}\big\|_{\mathrm{F}}\big\|Q^{(t+1)\top}Q^{*\perp}\big\|_{\mathrm{F}}\\ \leq& \big\|V^{(t)\top}Q^{*\perp}\big\|_{\mathrm{F}}^2\\ &+ 2\sqrt{k}\sqrt{\frac{s^*}{\widehat{s}}}\big\|V^{(t)\top}Q^{*\perp}\big\|_{\mathrm{F}}\big\|Q^{(t+1)\top}Q^{*\perp}\big\|_{\mathrm{F}}.\end{aligned}$$

This quadratic inequality implies

$$\big\|Q^{(t+1)\top}Q^{*\perp}\big\|_{\mathrm{F}} \leq \left(1 + 2\sqrt{\frac{k\cdot s^*}{\widehat{s}}}\right)\big\|V^{(t)\top}Q^{*\perp}\big\|_{\mathrm{F}}.$$

□

B.7 Proof of Lemma B.4

First of all we will have to prove that this result holds.

Lemma B.8. For n sufficiently large, and $S^* \subseteq \mathcal{I}$, we have

$$\|Q^{*\top}\widehat{Q}(\mathcal{I})^{\perp}\|_{\mathrm{F}} \leq C_1 \cdot \frac{\sqrt{\lambda_1\lambda_{k+1}}}{\lambda_k - \lambda_{k+1}} \cdot \sqrt{\frac{|\mathcal{I}|\cdot(k+\log d)}{n}},$$

which holds with probability at least $1 - 4/(n-1) - 1/d - 6\log n/n$.

Proof. The proof is an extension from the deviation for main upper bound analysis [41]. The details of the proof can be found in Appendix B of [43]. □

Now we give the proof of Lemma B.4.

Proof. In this proof we denote $\mathcal{I}_1 = \mathcal{I} \cup \mathcal{I}^*$ as the union of the row-support of Q^* and $Q^{(t)}$. Under this notation we have $|\mathcal{I}_1| \leq s^* + \widehat{s}$. Besides for sufficiently large n Lemma B.8 implies that we can assume that $\big\|Q^{*\top}\widehat{Q}(\mathcal{I}_1)^{\perp}\big\|_{\mathrm{F}} \leq 1/2$.
Since the assumptions of Lemma B.2 are all satisfied here, we have the result of Lemma B.2

$$\begin{aligned}\|V^{(t)\top}\widehat{Q}(\mathcal{I}_1)^{\perp}\|_{\mathrm{F}} \leq& \frac{\|Q^{(t)\top}\widehat{Q}(\mathcal{I}_1)^{\perp}\|_{\mathrm{F}}}{\sqrt{1-\|Q^{(t)\top}\widehat{Q}(\mathcal{I}_1)^{\perp}\|_{\mathrm{F}}^2}}\cdot\rho\\ &+ \frac{\alpha}{\sqrt{3}(1-\alpha)},\end{aligned} \tag{B.40}$$

where $\rho = \gamma/(1-\alpha)$ follows the same notation in Lemma B.2.
Triangle inequality of subspace distance and Lemma B.8 implies that

$$\begin{aligned}&\Big|\big\|Q^{(t+1)\top}Q^{*\perp}\big\|_{\mathrm{F}} - \big\|Q^{(t+1)\top}\widehat{Q}(\mathcal{I}_1)^{\perp}\big\|_{\mathrm{F}}\Big|\\ &\leq \big\|Q^{*\top}\widehat{Q}(\mathcal{I}_1)^{\perp}\big\|_{\mathrm{F}} \leq \Psi(|\mathcal{I}_1|),\end{aligned} \tag{B.41}$$

$$\begin{aligned}&\Big|\big\|V^{(t)\top}Q^{*\perp}\big\|_{\mathrm{F}} - \big\|V^{(t)\top}\widehat{Q}(\mathcal{I}_1)^{\perp}\big\|_{\mathrm{F}}\Big|\\ &\leq \big\|Q^{*\top}\widehat{Q}(\mathcal{I}_1)^{\perp}\big\|_{\mathrm{F}} \leq \Psi(|\mathcal{I}_1|).\end{aligned} \tag{B.42}$$

Combining equation (B.40) and (B.42) we obtain

$$\begin{aligned}&\big\|V^{(t)\top}Q^{*\perp}\big\|_{\mathrm{F}} \leq \big\|V^{(t)\top}\widehat{Q}(\mathcal{I}_1)^{\perp}\big\|_{\mathrm{F}} + \Psi(|\mathcal{I}_1|)\\ &\leq \frac{\|Q^{(t)\top}\widehat{Q}(\mathcal{I}_1)^{\perp}\|_{\mathrm{F}}}{\sqrt{1-\|Q^{(t)\top}\widehat{Q}(\mathcal{I}_1)^{\perp}\|_{\mathrm{F}}^2}}\cdot\rho + \frac{\alpha}{\sqrt{3}(1-\alpha)} + \Psi(|\mathcal{I}_1|).\end{aligned} \tag{B.43}$$

Here we define an auxiliary function $f(z) = z/\sqrt{1-z^2}$. Since $f(z)$ is convex for $-1 < z < 1$, we have

$$f(z_2) - f(z_1) \leq f'(z_2)(z_2 - z_1) = (1 - z_2^2)^{-\frac{3}{2}}(z_2 - z_1).$$

If we have $z_1 = \big\|Q^{(t)\top}Q^{*\perp}\big\|_{\mathrm{F}}$ and $z_2 = \big\|Q^{(t)\top}\widehat{Q}(\mathcal{I}_1)^{\perp}\big\|_{\mathrm{F}}$, it follows that

$$\begin{aligned}&\frac{\|Q^{(t)\top}\widehat{Q}(\mathcal{I}_1)^{\perp}\|_{\mathrm{F}}}{\sqrt{1-\|Q^{(t)\top}\widehat{Q}(\mathcal{I}_1)^{\perp}\|_{\mathrm{F}}^2}} \leq \frac{\big\|Q^{(t)\top}Q^{*\perp}\big\|_{\mathrm{F}}}{\sqrt{1-\big\|Q^{(t)\top}Q^{*\perp}\big\|_{\mathrm{F}}^2}}\\ &+ \Psi(|\mathcal{I}_1|)\Big(1 - \big\|Q^{(t)\top}\widehat{Q}(\mathcal{I}_1)^{\perp}\big\|_{\mathrm{F}}^{-\frac{3}{2}}\Big).\end{aligned} \tag{B.44}$$

Since we have assumed that $\Psi(2\widehat{s}) \leq 1/24$, equation (B.41) implies that

$$\begin{aligned}\big\|Q^{(t)\top}\widehat{Q}(\mathcal{I}_1)^{\perp}\big\|_{\mathrm{F}} &\leq \big\|Q^{(t)\top}\widehat{Q}(\mathcal{I}_1)^{\perp}\big\|_{\mathrm{F}} + \Psi(|\mathcal{I}_1|)\\ &\leq 1/2 + 1/24 \leq \sqrt{2}/2,\end{aligned}$$

where the last inequality follows from the fact that $|\mathcal{I}_1| \le \widehat{s} + s^* \le 2\widehat{s}$. Due to the fact that the function $f(z)$ is increasing, we have

$$\Big(1 - \big\|Q^{(t)\top}\widehat{Q}(\mathcal{I}_1)^{\perp}\big\|_{\mathrm{F}}^{-\frac{3}{2}}\Big) \le (1-1/2)^{-\frac{3}{2}} < 3. \quad \text{(B.45)}$$

And thus (B.44) and (B.45) implies an upper bound for the right-hand side of (B.43)

$$\big\|V^{(t)\top}Q^{*\perp}\big\|_{\mathrm{F}} \le \frac{\big\|Q^{(t)\top}Q^{*\perp}\big\|_{\mathrm{F}}}{\sqrt{1-\big\|Q^{(t)\top}Q^{*\perp}\big\|_{\mathrm{F}}^2}} \cdot \rho + 3\rho\Psi(|\mathcal{I}_1|) + \Psi(|\mathcal{I}_1|) + \frac{\alpha}{\sqrt{3}(1-\alpha)}. \quad \text{(B.46)}$$

Remind that Lemma B.3 has the following result

$$\big\|Q^{(t+1)\top}Q^{*\perp}\big\|_{\mathrm{F}} \le \Big(1 + 2\sqrt{\frac{ks^*}{\widehat{s}}}\Big)\big\|V^{(t)\top}Q^{*\perp}\big\|_{\mathrm{F}}. \quad \text{(B.47)}$$

Combining (B.46) and (B.47), we obtain

$$\begin{aligned}&\big\|Q^{(t+1)\top}Q^{*\perp}\big\|_{\mathrm{F}} \\ &\le \Big(1 + 2\sqrt{\frac{ks^*}{\widehat{s}}}\Big)\frac{\big\|Q^{(t)\top}Q^{*\perp}\big\|_{\mathrm{F}}}{\sqrt{1-\big\|Q^{(t)\top}Q^{*\perp}\big\|_{\mathrm{F}}^2}} \cdot \rho \\ &+3\rho\Big(1 + 2\sqrt{\frac{ks^*}{\widehat{s}}}\Big)\Psi(|\mathcal{I}_1|) + \Big(1 + 2\sqrt{\frac{ks^*}{\widehat{s}}}\Big)\Psi(|\mathcal{I}_1|) \\ &+\Big(1 + 2\sqrt{\frac{ks^*}{\widehat{s}}}\Big)\frac{\alpha}{\sqrt{3}(1-\alpha)}. \end{aligned} \quad \text{(B.48)}$$

Under our assumption

$$\widehat{s} = C_1 \max\Big\{\Big\lceil \frac{44}{(\rho^{-1/2}-1)^2}\Big\rceil, 1\Big\} \cdot s^*,$$

$$\text{and } \|Q^{(t)\top}Q^{*\perp}\|_{\mathrm{F}} \le \min\Big\{\sqrt{(1-\rho^{1/2})}, 1/2\Big\},$$

with $C_1 > 1$, we have

$$1 + 2\sqrt{\frac{ks^*}{\widehat{s}}} \le \rho^{-1/2}, \quad \frac{1}{\sqrt{1-\big\|Q^{(t)\top}Q^{*\perp}\big\|_{\mathrm{F}}^2}} \le \rho^{-1/4}. \quad \text{(B.49)}$$

Plugging (B.49) into (B.48) and note that $|\mathcal{I}_1| \le 2\widehat{s}$, we obtain

$$\begin{aligned}\big\|Q^{(t+1)\top}Q^{*\perp}\big\|_{\mathrm{F}} &\le \rho^{1/4}\big\|Q^{(t)\top}Q^{*\perp}\big\|_{\mathrm{F}} + 3\rho^{1/2}\Psi(2\widehat{s}) \\ &+\Big(1 + 2\sqrt{\frac{ks^*}{\widehat{s}}}\Big)\Big(\Psi(2\widehat{s}) + \frac{\alpha}{\sqrt{3}(1-\alpha)}\Big).\end{aligned}$$

Since our proof depends on high probability result Lemma B.6 and Lemma B.7 in (B.58) and (B.59). The entire proof holds with probability at least

$$1 - 2e^{-\tau^2/2} - 4/(n-1) - 1/d - 6\log n/n - 1/n.$$

□

B.8 Proof of Main Theorem

Proof. To simplify notation, we introduce

$$\omega = \Big(1 + 2\sqrt{\frac{ks^*}{\widehat{s}}}\Big)\Big(\Psi(2\widehat{s}) + \frac{\alpha}{\sqrt{3}(1-\alpha)}\Big).$$

We will prove by mathematical induction that for $t = 2, \ldots, T$,

$$\begin{aligned}\big\|Q^{(t)\top}Q^{*\perp}\big\|_{\mathrm{F}} &\le \rho^{(t-1)/4}\big\|Q^{(1)\top}Q^{*\perp}\big\|_{\mathrm{F}} \\ &+ \frac{3\rho^{1/2}}{1-\rho^{1/4}}\Psi(2\widehat{s}) + \frac{\omega}{1-\rho^{1/4}}.\end{aligned} \quad \text{(B.50)}$$

First of all we have the assumption for the initial value $Q^{(0)}$

$$\big\|Q^{(0)\top}Q^{*\perp}\big\|_{\mathrm{F}} \le \min\Big\{\sqrt{\frac{\rho(1-\rho^{1/2})}{2}}, \frac{\sqrt{2\rho}}{4}\Big\} < 1. \quad \text{(B.51)}$$

Based on Lemma B.3, the Initialization step of Algorithm 1 implies that

$$\big\|Q^{(1)\top}Q^{*\perp}\big\|_{\mathrm{F}} \le \Big(1 + 2\sqrt{\frac{ks^*}{\widehat{s}}}\Big)\big\|Q^{(0)\top}Q^{*\perp}\big\|_{\mathrm{F}}. \quad \text{(B.52)}$$

By our assumption on $\widehat{s}$ (B.4), we have

$$1 + 2\sqrt{\frac{ks^*}{\widehat{s}}} \le \frac{1}{\sqrt{\rho}}. \quad \text{(B.53)}$$

Combining (B.51), (B.52) and (B.53), we have

$$\big\|Q^{(1)\top}Q^{*\perp}\big\|_{\mathrm{F}} \le \min\Big\{\frac{\sqrt{1-\rho^{1/2}}}{2}, \frac{1}{4}\Big\},$$

which means the condition of Lemma B.4 is satisfied and hence

$$\begin{aligned}\big\|Q^{(2)\top}Q^{*\perp}\big\|_{\mathrm{F}} &\le \rho^{1/4}\big\|Q^{(1)\top}Q^{*\perp}\big\|_{\mathrm{F}} + 3\rho^{1/2}\Psi(2\widehat{2}) + \omega \\ &\le \rho^{1/4}\big\|Q^{(1)\top}Q^{*\perp}\big\|_{\mathrm{F}} + \frac{3\rho^{1/2}}{1-\rho^{1/4}}\Psi(2\widehat{s}) + \frac{\omega}{1-\rho^{1/4}}.\end{aligned}$$

Thus we have proved (B.50) holds for $t = 2$. Now suppose (B.50) holds for $t \ge 2$, we want to prove that it also holds for the case for $t+1$. We notice that

$$\begin{aligned}\rho^{(t-1)/4}\big\|Q^{(t)T}Q^{*\perp}\big\|_{\mathrm{F}} &\le \big\|Q^{(t)T}Q^{*\perp}\big\|_{\mathrm{F}} \\ &\le \min\Big\{\sqrt{\frac{(1-\rho^{1/2})}{2}}, \frac{\sqrt{2}}{4}\Big\}.\end{aligned} \quad \text{(B.54)}$$

Under the assumption (B.3) and (B.5), we also have

$$\begin{aligned}&\frac{1}{1-\rho^{1/4}}\Big(1 + 2\sqrt{\frac{ks^*}{\widehat{s}}}\Big)\Big(\Psi(2\widehat{s}) + \frac{\alpha}{\sqrt{3}(1-\alpha)}\Big) \\ &\le \frac{1}{4}\min\Big\{\sqrt{1-\rho^{1/2}}, \frac{1}{2}\Big\},\end{aligned} \quad \text{(B.55)}$$

$$\frac{3\rho^{1/2}}{1-\rho^{1/4}}\Psi(2\widehat{s}) \le \frac{1}{4}\min\Big\{\sqrt{1-\rho^{1/2}}, \frac{1}{2}\Big\}. \quad \text{(B.56)}$$

Plugging (B.54), (B.56) and (B.55) into the right-hand side of (B.50), we find out that $\left\|Q^{(t)\top}Q^{*\perp}\right\|_{\mathrm{F}}$ satisfies the condition of Lemma B.4

$$\left\|Q^{(t)\top}Q^{*\perp}\right\|_{\mathrm{F}} \le \min\left\{\sqrt{1-\rho^{1/2}}, \frac{1}{2}\right\}.$$

And thus Lemma B.4 implies that

$$\left\|Q^{(t+1)\top}Q^{*\perp}\right\|_{\mathrm{F}} \le \rho^{1/4}\left\|Q^{(t)\top}Q^{*\perp}\right\|_{\mathrm{F}} + 3\rho^{1/2}\Psi(2\widehat{s}) + \omega. \tag{B.57}$$

If we plug (B.50) into the right-hand side of (B.57), we have

$$\begin{aligned}
&\left\|Q^{(t+1)\top}Q^{*\perp}\right\|_{\mathrm{F}} \\
\le&\rho^{1/4}\left(\rho^{(t-1)/4}\left\|Q^{(1)\top}Q^{*\perp}\right\|_{\mathrm{F}} + \frac{3\rho^{1/2}}{1-\rho^{1/4}}\Psi(2\widehat{s}) + \frac{\omega}{1-\rho^{1/4}}\right) \\
&+3\rho^{1/2}\Psi(2\widehat{s}) + \omega \\
\le&\rho^{t/4}\left\|Q^{(1)\top}Q^{*\perp}\right\|_{\mathrm{F}} + \frac{3\rho^{1/2}}{1-\rho^{1/4}}\Psi(2\widehat{s}) + \frac{\omega}{1-\rho^{1/4}}.
\end{aligned}$$

By mathematical induction, we know that (B.50) holds for $t = 1, 2, \ldots, T$. By simply replacing $\Psi(2\widehat{s})$ by the definition of Ψ, and also noticing the assumption B.4, we have our result

$$\begin{aligned}
\left\|Q^{(t)\top}Q^{*\perp}\right\|_{\mathrm{F}} \le& \rho^{t/4}\left\|Q^{(0)\top}Q^{*\perp}\right\|_{\mathrm{F}} \\
&+\frac{C_1}{1-\rho^{1/4}}\frac{\sqrt{\lambda_1\lambda_{k+1}}}{\lambda_k-\lambda_{k+1}}\sqrt{\frac{ks^*(k+\log d)}{n}} \\
&+ \frac{\alpha}{\sqrt{3}(1-\rho^{1/4})(1-\alpha)}\left(1+2\sqrt{\frac{ks^*}{\widehat{s}}}\right) \\
&\le \rho^{t/4}\min\left\{\sqrt{1-\rho^{1/2}}, \frac{1}{2}\right\} \\
&+ \frac{C_1}{1-\rho^{1/4}}\frac{\sqrt{\lambda_1\lambda_{k+1}}}{\lambda_k-\lambda_{k+1}}\sqrt{\frac{ks^*(k+\log d)}{n}} \\
&+ \frac{\alpha}{\sqrt{3}(1-\rho^{1/4})(1-\alpha)}\left(1+2\sqrt{\frac{ks^*}{\widehat{s}}}\right).
\end{aligned}$$

Since the proof depends on high probability result Lemma B.3 and Lemma B.5, our result holds for $t = 1, 2, \ldots, T$ with probability at least

$$1 - 2Te^{-\tau^2/2} - 4/(n-1) - 1/d - 6\log n/n - 1/n.$$

$\square$