Appendix for 'Toward Efficient and Accurate Covariance Matrix Estimation on Compressed Data'

This appendix is organized as follows. In Section 1, we state all theoretical results, including our proposed Lemma 1 and Lemma 2 whose details are not presented in the main text of the paper. In Section 2, we provide detailed proofs for all of the results. In Section 3, we reformulate and discuss the current theoretical results of the counterparts: *Gauss-Inverse* and *UniSample-HD*. In Section 4, we give a detailed analysis of the computational complexity. Finally, in Section 5, we study the impact of different α on the estimation accuracy.

Before proceeding, we first show the notations used in this appendix.

Notation. Let $[k]$ denote a set of integers $\{1, 2, \ldots, k\}$. Given a matrix $\mathbf{X} \in \mathbb{R}^{d\times n}$, for $j \in [d]$, $i \in [n]$, we let $\mathbf{x}_i \in \mathbb{R}^d$ denote the i-th column of $\mathbf{X}$, and x_{ji} denote the (j, i)-th element of $\mathbf{X}$ or j-th element of $\mathbf{x}_i$. Let $\{\mathbf{X}_t\}_{t=1}^k$ denote the set of matrices $\{\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_k\}$, and $x_{ji,t}$ denote the (j, i)-th element of $\mathbf{X}_t$. Let $\mathbf{X}^T$ denote the transpose of $\mathbf{X}$, and $\mathrm{Tr}(\mathbf{X})$ denote its trace. Let $|x|$ denote the absolute value of x. Let $\|\mathbf{X}\|_2$ and $\|\mathbf{X}\|_F$ denote the spectral norm and Frobenius norm of $\mathbf{X}$, respectively. Let $\|\mathbf{x}\|_q = (\sum_{j=1}^d |x_j|^q)^{1/q}$ for $q \geq 1$ be the ℓ_q-norm of $\mathbf{x} \in \mathbb{R}^d$. Let $\mathbb{D}(\mathbf{x})$ or $\mathbb{D}(\{x_j\})$ be a square diagonal matrix with the elements of vector $\mathbf{x}$ on the main diagonal, and $\mathbb{D}(\mathbf{X})$ also be a square diagonal matrix whose main diagonal has only the main diagonal elements of $\mathbf{X}$. Finally, $\mathbf{X} \preceq \mathbf{Y}$ means that $\mathbf{Y} - \mathbf{X}$ is positive semidefinite.

1. Provable Results

For convenience, we first restate the theorems and their corollaries in the following.

Theorem 1. *Assume $\mathbf{X} \in \mathbb{R}^{d\times n}$ and the sampling size $2 \leq m < d$. Sample m entries from each $\mathbf{x}_i \in \mathbb{R}^d$ with replacement by running Algorithm 1. Let $\{p_{ki}\}_{k=1}^d$ and $\mathbf{S}_i \in \mathbb{R}^{d\times m}$ denote the sampling probabilities and sampling matrix, respectively. Then, the unbiased estimator for the target covariance matrix $\mathbf{C} = \frac{1}{n}\sum_{i=1}^n \mathbf{x}_i\mathbf{x}_i^T = \frac{1}{n}\mathbf{X}\mathbf{X}^T$ can be recovered as*

$$\mathbf{C}_e = \widehat{\mathbf{C}}_1 - \widehat{\mathbf{C}}_2, \tag{1}$$

where $\widehat{\mathbf{C}}_1 = \frac{m}{nm-n}\sum_{i=1}^n \mathbf{S}_i\mathbf{S}_i^T\mathbf{x}_i\mathbf{x}_i^T\mathbf{S}_i\mathbf{S}_i^T$, $\widehat{\mathbf{C}}_2 = \frac{m}{nm-n}\sum_{i=1}^n \mathbb{D}(\mathbf{S}_i\mathbf{S}_i^T\mathbf{x}_i\mathbf{x}_i^T\mathbf{S}_i\mathbf{S}_i^T)\mathbb{D}(\mathbf{b}_i)$ with $b_{ki} = \frac{1}{1+(m-1)p_{ki}}$, and $\mathbb{E}[\mathbf{C}_e] = \mathbf{C}$.

Theorem 2. *Given $\mathbf{X} \in \mathbb{R}^{d\times n}$ and the sampling size $2 \leq m < d$, let $\mathbf{C}$ and $\mathbf{C}_e$ be defined as in Theorem 1. If the sampling probabilities satisfy $p_{ki} = \alpha\frac{|x_{ki}|}{\|\mathbf{x}_i\|_1} + (1-\alpha)\frac{x_{ki}^2}{\|\mathbf{x}_i\|_2^2}$ with $0 < \alpha < 1$ for all $k \in [d]$ and $i \in [n]$, then with probability at least $1 - \eta - \delta$,*

$$\|\mathbf{C}_e - \mathbf{C}\|_2 \leq \log(\frac{2d}{\delta})\frac{2R}{3} + \sqrt{2\sigma^2\log(\frac{2d}{\delta})}, \tag{2}$$

where we define that $R = \max_{i\in[n]}\left[\frac{7\|\mathbf{x}_i\|_2^2}{n} + \log^2(\frac{2nd}{\eta})\frac{14\|\mathbf{x}_i\|_1^2}{nm\alpha^2}\right]$, and $\sigma^2 = \sum_{i=1}^n\left[\frac{8\|\mathbf{x}_i\|_2^4}{n^2m^2(1-\alpha)^2} + \frac{4\|\mathbf{x}_i\|_1^2\|\mathbf{x}_i\|_2^2}{n^2m^3\alpha^2(1-\alpha)}\right.$ $\left. + \frac{9\|\mathbf{x}_i\|_2^4}{n^2m(1-\alpha)} + \frac{2\|\mathbf{x}_i\|_2^2\|\mathbf{x}_i\|_1^2}{n^2m^2\alpha(1-\alpha)}\right] + \|\sum_{i=1}^n\frac{\|\mathbf{x}_i\|_1^2\mathbf{x}_i\mathbf{x}_i^2}{n^2m\alpha}\|_2$.

Corollary 1. *Given $\mathbf{X} \in \mathbb{R}^{d\times n}$ and sampling size $2 \leq m < d$, let $\mathbf{C}$ and $\mathbf{C}_e$ be constructed by Algorithm 1. Define $\frac{\|\mathbf{x}_i\|_1}{\|\mathbf{x}_i\|_2} \leq \varphi$ with $1 \leq \varphi \leq \sqrt{d}$, and $\|\mathbf{x}_i\|_2 \leq \tau$ for all $i \in [n]$. Then, with probability at least $1 - \eta - \delta$ we have*

$$\|\mathbf{C}_e - \mathbf{C}\|_2 \leq \min\{\widetilde{O}\Big(f + \frac{\tau^2\varphi}{m}\sqrt{\frac{1}{n}} + \tau^2\sqrt{\frac{1}{nm}}\Big), \widetilde{O}\Big(f + \frac{\tau\varphi}{m}\sqrt{\frac{d\|\mathbf{C}\|_2}{n}} + \tau\sqrt{\frac{d\|\mathbf{C}\|_2}{nm}}\Big)\}, \tag{3}$$

where $f = \frac{\tau^2}{n} + \frac{\tau^2\varphi^2}{nm} + \tau\varphi\sqrt{\frac{\|\mathbf{C}\|_2}{nm}}$, and $\widetilde{O}(\cdot)$ hides the logarithmic factors on η, δ, m, n, d, and α.

Corollary 2. *Given $\mathbf{X} \in \mathbb{R}^{d\times n}$ $(2 \leq d)$ and an unknown population covariance matrix $\mathbf{C}_p \in \mathbb{R}^{d\times d}$ with each column vector $\mathbf{x}_i \in \mathbb{R}^d$ i.i.d. generated from the Gaussian distribution $\mathcal{N}(\mathbf{0}, \mathbf{C}_p)$. Let $\mathbf{C}_e$ be constructed by Algorithm 1 with sampling size $2 \leq m < d$. Then, with probability at least $1-\eta-\delta-\zeta$,*

$$\frac{\|\mathbf{C}_e - \mathbf{C}_p\|_2}{\|\mathbf{C}_p\|_2} \leq \widetilde{O}\Big(\frac{d^2}{nm} + \frac{d}{m}\sqrt{\frac{d}{n}}\Big); \tag{4}$$

Additionally, assuming rank$(\mathbf{C}_p) \leq r$, with probability at least $1-\eta-\delta-\zeta$ we have

$$\frac{\|[\mathbf{C}_e]_r - \mathbf{C}_p\|_2}{\|\mathbf{C}_p\|_2} \leq \widetilde{O}\Big(\frac{rd}{nm} + \frac{r}{m}\sqrt{\frac{d}{n}} + \sqrt{\frac{rd}{nm}}\Big), \tag{5}$$

where $[\mathbf{C}_e]_r$ is the solution to $\min_{rank(A)\leq r} \|\mathbf{A} - \mathbf{C}_e\|_2$, and $\widetilde{O}(\cdot)$ hides the logarithmic factors on η, δ, ζ, m, n, d, and α.

Corollary 3. *Given $\mathbf{X}$, d, m, $\mathbf{C}_p$ and $\mathbf{C}_e$ as in Corollary 2. Let $\prod_k = \sum_{i=1}^k \mathbf{u}_i\mathbf{u}_i^T$ and $\widehat{\prod}_k = \sum_{i=1}^k \hat{\mathbf{u}}_i\hat{\mathbf{u}}_i^T$ with $\{\mathbf{u}_i\}_{i=1}^k$ and $\{\hat{\mathbf{u}}_i\}_{i=1}^k$ being the leading k eigenvectors of $\mathbf{C}_p$ and $\mathbf{C}_e$, respectively. Denote by λ_k the k-th largest eigenvalue of $\mathbf{C}_p$. Then, with probability at least $1-\eta-\delta-\zeta$,*

$$\frac{\|\widehat{\prod}_k - \prod_k\|_2}{\|\mathbf{C}_p\|_2} \leq \frac{1}{\lambda_k - \lambda_{k+1}}\widetilde{O}\Big(\frac{d^2}{nm} + \frac{d}{m}\sqrt{\frac{d}{n}}\Big), \tag{6}$$

where the eigengap $\lambda_k - \lambda_{k+1} > 0$ and $\widetilde{O}(\cdot)$ hides the logarithmic factors on η, δ, ζ, m, n, d, and α.

Next, we present two lemmas: Lemma 1 and Lemma 2, which are used to prove the foregoing theorems. The detailed statements of the two lemmas are **omitted in the main text of the paper** owing to limited space, and now they are described below.

Lemma 1. *Given any vector $\mathbf{x} \in \mathbb{R}^d$, and $m < d$, sample m entries from $\mathbf{x}$ with replacement by running Algorithm 1 with the inputs $\mathbf{x}$ and m. Let $\{p_k\}_{k=1}^d$ denote the corresponding sampling probabilities, $\mathbf{S} \in \mathbb{R}^{d\times m}$ denote the corresponding rescaled sampling matrix, and $\{\mathbf{e}_k\}_{k=1}^d$ denote the standard basis vectors for $\mathbb{R}^d$. Then, we have*

$$\mathbb{E}\left[\mathbf{S}\mathbf{S}^T\mathbf{x}\mathbf{x}^T\mathbf{S}\mathbf{S}^T\right] = \sum_{k=1}^d \frac{x_k^2}{mp_k}\mathbf{e}_k\mathbf{e}_k^T + \frac{m-1}{m}\mathbf{x}\mathbf{x}^T; \tag{7}$$

$$\mathbb{E}\left[\mathbb{D}(\mathbf{S}\mathbf{S}^T\mathbf{x}\mathbf{x}^T\mathbf{S}\mathbf{S}^T)\right] = \sum_{k=1}^d (\frac{1}{mp_k} + \frac{m-1}{m})x_k^2\mathbf{e}_k\mathbf{e}_k^T; \tag{8}$$

$$\begin{aligned}\mathbb{E}\left[(\mathbb{D}(\mathbf{S}\mathbf{S}^T\mathbf{x}\mathbf{x}^T\mathbf{S}\mathbf{S}^T))^2\right] &= \sum_{k=1}^d \Bigg[\frac{1}{m^3p_k^3} + \frac{7(m-1)}{m^3p_k^2} + \frac{6(m^2-3m+2)}{m^3p_k} \\ &+ \frac{m^3-6m^2+11m-6}{m^3}\Bigg] x_k^4\mathbf{e}_k\mathbf{e}_k^T;\end{aligned} \tag{9}$$

$$\begin{aligned}&\mathbb{E}\left[\mathbf{S}\mathbf{S}^T\mathbf{x}\mathbf{x}^T\mathbf{S}\mathbf{S}^T\mathbb{D}(\mathbf{S}\mathbf{S}^T\mathbf{x}\mathbf{x}^T\mathbf{S}\mathbf{S}^T)\right] = (\mathbb{E}\left[\mathbb{D}(\mathbf{S}\mathbf{S}^T\mathbf{x}\mathbf{x}^T\mathbf{S}\mathbf{S}^T)\mathbf{S}\mathbf{S}^T\mathbf{x}\mathbf{x}^T\mathbf{S}\mathbf{S}^T\right])^T \\ &= \sum_{k=1}^d \left[\frac{1}{m^3p_k^3} + \frac{6(m-1)}{m^3p_k^2} + \frac{3(m^2-3m+2)}{m^3p_k}\right] x_k^4\mathbf{e}_k\mathbf{e}_k^T + \frac{m-1}{m^3}\mathbf{x}\mathbf{x}^T\mathbb{D}(\{\frac{x_k^2}{p_k^2}\}) \\ &+ \frac{3(m^2-3m+2)}{m^3}\mathbf{x}\mathbf{x}^T\left[\mathbb{D}(\{\frac{x_k^2}{p_k}\}) + \frac{m-3}{3}\mathbb{D}(\{x_k^2\})\right];\end{aligned} \tag{10}$$

$$\begin{aligned}\mathbb{E}\left[(\mathbf{S}\mathbf{S}^T\mathbf{x}\mathbf{x}^T\mathbf{S}\mathbf{S}^T)^2\right] &= \sum_{k=1}^d \left[\frac{4(m-1)}{m^3p_k^2} + \frac{1}{m^3p_k^3}\right] x_k^4\mathbf{e}_k\mathbf{e}_k^T \\ &+ \sum_{k=1}^d \left[\frac{\|\mathbf{x}\|_2^2(m^2-3m+2)}{m^3} + \frac{m-1}{m^3}\sum_{k=1}^d \frac{x_k^2}{p_k}\right]\frac{x_k^2}{p_k}\mathbf{e}_k\mathbf{e}_k^T \\ &+ \left[\frac{\|\mathbf{x}\|_2^2(m^3-6m^2+11m-6)}{m^3} + \frac{m^2-3m+2}{m^3}\sum_{k=1}^d \frac{x_k^2}{p_k}\right]\mathbf{x}\mathbf{x}^T\end{aligned}$$

$$+ \mathbf{x}\mathbf{x}^T \left[\frac{2(m^2-3m+2)}{m^3} \mathbb{D}(\{\frac{x_k^2}{p_k}\}) + \frac{m-1}{m^3} \mathbb{D}(\{\frac{x_k^2}{p_k^2}\}) \right]$$
$$+ \left[\frac{2(m^2-3m+2)}{m^3} \mathbb{D}(\{\frac{x_k^2}{p_k}\}) + \frac{m-1}{m^3} \mathbb{D}(\{\frac{x_k^2}{p_k^2}\}) \right] \mathbf{x}\mathbf{x}^T, \tag{11}$$

where the expectation is w.r.t. $\mathbf{S}$*, and* $\mathbb{D}(\{x_k^2\})$ *denotes a square diagonal matrix with* $\{x_k^2\}_{k=1}^d$ *on its diagonal that can be extended to other similar notations.*

Lemma 2. *Given the definitions in Lemma 1. Then, with probability at least* $1 - \sum_{k=1}^d \eta_k$*, we have*

$$\|\mathbf{S}\mathbf{S}^T\mathbf{x}\mathbf{x}\mathbf{S}\mathbf{S}^T\|_2 \leq \sum_{k\in\Gamma} f^2(x_k, \eta_k, m), \tag{12}$$

where Γ *is a set containing at most* m *different elements of* $[d]$ *with its cardinality* $|\Gamma| \leq m$*, and* $f(x_k,\eta_k,m) = |x_k| + \log(\frac{2}{\eta_k})\left[\frac{|x_k|}{3mp_k} + |x_k|\sqrt{\frac{1}{9m^2p_k^2} + \frac{2}{\log(2/\eta_k)}(\frac{1}{mp_k} - \frac{1}{m})}\right]$.

Remark 1. For the expressions in Lemma 1 and Lemma 2, the sampling probability p_k appears in the denominator, which indicates that the derived bound may be sensitive to a highly small $p_k \neq 0$. However, in terms of any $p_k = 0$, we can define $\frac{|x_k|^a}{p_k^b} = 0$ for $a, b > 0$, because we follow the rule that $p_k = 0$ only when $x_k = 0$ and $x_k = 0$ can never be sampled. Thus, the aforementioned two lemmas and other derived results are applicable to the case where there exists $p_k = 0$.

2. Analysis

2.1. Technical Theorems

Below, we first show the Matrix Bernstein inequality employed for characterizing the sums of independent random variables/matrices, and then present a matrix perturbation result for eigenvalues.

Theorem 3 (Tropp 2015, p. 76)**.** *Let* $\{\mathbf{A}_i\}_{i=1}^L \in \mathbb{R}^{d\times n}$ *be independent random matrices with* $\mathbb{E}[\mathbf{A}_i] = 0$ *and* $\|\mathbf{A}_i\|_2 \leq R$*. Define the variance* $\sigma^2 = \max\{\|\sum_{i=1}^L \mathbb{E}[\mathbf{A}_i\mathbf{A}_i^T]\|_2, \|\sum_{i=1}^L \mathbb{E}[\mathbf{A}_i^T\mathbf{A}_i]\|_2\}$*. Then,* $\mathbb{P}(\|\sum_{i=1}^L \mathbf{A}_i\|_2 \geq \epsilon) \leq (d+n)\exp(\frac{-\epsilon^2/2}{\sigma^2 + R\epsilon/3})$ *for all* $\epsilon \geq 0$*.*

Theorem 4 (Golub & Van Loan 1996, p. 396)**.** *If* $\mathbf{A} \in \mathbb{R}^{d\times d}$ *and* $\mathbf{A}+\mathbf{E} \in \mathbb{R}^{d\times d}$ *are symmetric matrices, then*

$$\lambda_k(\mathbf{A}) + \lambda_d(\mathbf{E}) \leq \lambda_k(\mathbf{A}+\mathbf{E}) \leq \lambda_k(\mathbf{A}) + \lambda_1(\mathbf{E}) \tag{13}$$

for $k \in [d]$*, where* $\lambda_k(\mathbf{A}+\mathbf{E})$ *and* $\lambda_k(\mathbf{A})$ *designate the* k*-th largest eigenvalues.*

2.2. Proof of Lemma 1

Proof. According to Algorithm 1 in the main text of the paper, each column vector in the rescaled sampling matrix $\mathbf{S} \in \mathbb{R}^{d\times m}$ is sampled with replacement from $\{\mathbf{r}_k = \frac{1}{\sqrt{mp_k}}\mathbf{e}_k\}_{k=1}^d$ with corresponding probabilities $\{p_k\}_{k=1}^d$, where $\{\mathbf{e}_k\}_{k=1}^d$ are the standard basis vectors for $\mathbb{R}^d$.

Firstly, we prove Eq. (7). By the definition, we expand

$$\mathbf{S}\mathbf{S}^T\mathbf{x}\mathbf{x}^T\mathbf{S}\mathbf{S}^T = \sum_{j=1}^m \mathbf{s}_{t_j}\mathbf{s}_{t_j}^T\mathbf{x}\sum_{j=1}^m \mathbf{x}^T\mathbf{s}_{t_j}\mathbf{s}_{t_j}^T \tag{14}$$
$$= \sum_{j=1}^m \mathbf{s}_{t_j}\mathbf{s}_{t_j}^T\mathbf{x}\mathbf{x}^T\mathbf{s}_{t_j}\mathbf{s}_{t_j}^T + \sum_{i\neq j\in[m]} \mathbf{s}_{t_i}\mathbf{s}_{t_i}^T\mathbf{x}\mathbf{x}^T\mathbf{s}_{t_j}\mathbf{s}_{t_j}^T, \tag{15}$$

where the random variable t_j is in $[d]$.

Passing the expectation over $\mathbf{S}$ through the sum in Eq. (15), we have

$$\mathbb{E}\sum_{j=1}^m \mathbf{s}_{t_j}\mathbf{s}_{t_j}^T\mathbf{x}\mathbf{x}^T\mathbf{s}_{t_j}\mathbf{s}_{t_j}^T = \sum_{j=1}^m\sum_{k=1}^d \mathbb{P}(t_j = k)\mathbf{r}_k\mathbf{r}_k^T\mathbf{x}\mathbf{x}^T\mathbf{r}_k\mathbf{r}_k^T$$
$$= \sum_{j=1}^m\sum_{k=1}^d p_k\frac{1}{m^2p_k^2}\mathbf{e}_k\mathbf{e}_k^T\mathbf{x}\mathbf{x}^T\mathbf{e}_k\mathbf{e}_k^T = \sum_{k=1}^d \frac{x_k^2}{mp_k}\mathbf{e}_k\mathbf{e}_k^T, \tag{16}$$

and similarly

$$\mathbb{E}\sum_{i\neq j\in[m]}\mathbf{s}_{t_i}\mathbf{s}_{t_i}^T\mathbf{x}\mathbf{x}^T\mathbf{s}_{t_j}\mathbf{s}_{t_j}^T = \sum_{i\neq j\in[m]}\sum_{k=1}^{d}\sum_{q=1}^{d}\mathbb{P}(t_i=k)\mathbb{P}(t_j=q)\mathbf{r}_k\mathbf{r}_k^T\mathbf{x}\mathbf{x}^T\mathbf{r}_q\mathbf{r}_q^T \tag{17}$$

$$=\sum_{k=1}^{d}\sum_{q=1}^{d}x_kx_q\frac{m-1}{m}\mathbf{e}_k\mathbf{e}_q^T=\frac{m-1}{m}\mathbf{x}\mathbf{x}^T. \tag{18}$$

Now, combing Eq. (16) with Eq. (18) immediately proves Eq. (7).

Then, Eq. (8) can be proved based on Eq. (7) by

$$\mathbb{E}\left[\mathbb{D}(\mathbf{S}\mathbf{S}^T\mathbf{x}\mathbf{x}^T\mathbf{S}\mathbf{S}^T)\right]=\mathbb{D}(\mathbb{E}\left[\mathbf{S}\mathbf{S}^T\mathbf{x}\mathbf{x}^T\mathbf{S}\mathbf{S}^T\right])=\sum_{k=1}^{d}(\frac{1}{mp_k}+\frac{m-1}{m})x_k^2\mathbf{e}_k\mathbf{e}_k^T. \tag{19}$$

Alternatively, $\mathbb{D}(\mathbf{S}\mathbf{S}^T\mathbf{x}\mathbf{x}^T\mathbf{S}\mathbf{S}^T)$ can be explicitly expanded by

$$\mathbb{D}(\mathbf{S}\mathbf{S}^T\mathbf{x}\mathbf{x}^T\mathbf{S}\mathbf{S}^T)=\sum_{j=1}^{m}\mathbf{s}_{t_j}\mathbf{s}_{t_j}^T\sum_{k=1}^{d}x_k^2\mathbf{e}_k\mathbf{e}_k^T\sum_{j=1}^{m}\mathbf{s}_{t_j}\mathbf{s}_{t_j}^T. \tag{20}$$

Thus, the whole target expectations in Eq. (9), Eq. (10) and Eq. (11) can be explicitly expanded, and we can use similar ways of proving Eq. (7) to prove the remainder of the lemma.

To prove Eq. (9), we expand

$$\mathbb{E}\left[(\mathbb{D}(\mathbf{S}\mathbf{S}^T\mathbf{x}\mathbf{x}^T\mathbf{S}\mathbf{S}^T))^2\right]=\mathbb{E}\left[(\sum_{j=1}^{m}\mathbf{s}_{t_j}\mathbf{s}_{t_j}^T\sum_{k=1}^{d}x_k^2\mathbf{e}_k\mathbf{e}_k^T\sum_{j=1}^{m}\mathbf{s}_{t_j}\mathbf{s}_{t_j}^T)^2\right] \tag{21}$$

$$=\mathbb{E}\left[\sum_{j=1}^{m}\mathbf{s}_{t_j}\mathbf{s}_{t_j}^T\sum_{k=1}^{d}x_k^2\mathbf{e}_k\mathbf{e}_k^T\sum_{j=1}^{m}\mathbf{s}_{t_j}\mathbf{s}_{t_j}^T\sum_{j=1}^{m}\mathbf{s}_{t_j}\mathbf{s}_{t_j}^T\sum_{k=1}^{d}x_k^2\mathbf{e}_k\mathbf{e}_k^T\sum_{j=1}^{m}\mathbf{s}_{t_j}\mathbf{s}_{t_j}^T\right] \tag{22}$$

$$=\mathbb{E}\sum_{j=1}^{m}\mathbf{s}_{t_j}\mathbf{s}_{t_j}^T\sum_{k=1}^{d}x_k^2\mathbf{e}_k\mathbf{e}_k^T\mathbf{s}_{t_j}\mathbf{s}_{t_j}^T\sum_{j=1}^{m}\mathbf{s}_{t_j}\mathbf{s}_{t_j}^T\sum_{k=1}^{d}x_k^2\mathbf{e}_k\mathbf{e}_k^T\mathbf{s}_{t_j}\mathbf{s}_{t_j}^T \tag{23}$$

$$+\mathbb{E}\sum_{j=1}^{m}\mathbf{s}_{t_j}\mathbf{s}_{t_j}^T\sum_{k=1}^{d}x_k^2\mathbf{e}_k\mathbf{e}_k^T\mathbf{s}_{t_j}\mathbf{s}_{t_j}^T\sum_{i\neq j\in[m]}\mathbf{s}_{t_i}\mathbf{s}_{t_i}^T\sum_{k=1}^{d}x_k^2\mathbf{e}_k\mathbf{e}_k^T\mathbf{s}_{t_j}\mathbf{s}_{t_j}^T \tag{24}$$

$$+\mathbb{E}\sum_{i\neq j\in[m]}\mathbf{s}_{t_i}\mathbf{s}_{t_i}^T\sum_{k=1}^{d}x_k^2\mathbf{e}_k\mathbf{e}_k^T\mathbf{s}_{t_j}\mathbf{s}_{t_j}^T\sum_{j=1}^{m}\mathbf{s}_{t_j}\mathbf{s}_{t_j}^T\sum_{k=1}^{d}x_k^2\mathbf{e}_k\mathbf{e}_k^T\mathbf{s}_{t_j}\mathbf{s}_{t_j}^T \tag{25}$$

$$+\mathbb{E}\sum_{i\neq j\in[m]}\mathbf{s}_{t_i}\mathbf{s}_{t_i}^T\sum_{k=1}^{d}x_k^2\mathbf{e}_k\mathbf{e}_k^T\mathbf{s}_{t_j}\mathbf{s}_{t_j}^T\sum_{i\neq j\in[m]}\mathbf{s}_{t_i}\mathbf{s}_{t_i}^T\sum_{k=1}^{d}x_k^2\mathbf{e}_k\mathbf{e}_k^T\mathbf{s}_{t_j}\mathbf{s}_{t_j}^T, \tag{26}$$

where the four terms in the last equations are calculated as:

$$(23)=\mathbb{E}\sum_{j=1}^{m}\mathbf{s}_{t_j}\mathbf{s}_{t_j}^T\sum_{k=1}^{d}x_k^2\mathbf{e}_k\mathbf{e}_k^T\mathbf{s}_{t_j}\mathbf{s}_{t_j}^T\sum_{j=1}^{m}\mathbf{s}_{t_j}\mathbf{s}_{t_j}^T\sum_{k=1}^{d}x_k^2\mathbf{e}_k\mathbf{e}_k^T\mathbf{s}_{t_j}\mathbf{s}_{t_j}^T$$

$$=\mathbb{E}\sum_{j=1}^{m}\mathbf{s}_{t_j}\mathbf{s}_{t_j}^T\sum_{k=1}^{d}x_k^2\mathbf{e}_k\mathbf{e}_k^T\mathbf{s}_{t_j}\mathbf{s}_{t_j}^T\mathbf{s}_{t_j}\mathbf{s}_{t_j}^T\sum_{k=1}^{d}x_k^2\mathbf{e}_k\mathbf{e}_k^T\mathbf{s}_{t_j}\mathbf{s}_{t_j}^T$$

$$
\begin{aligned}
&+\mathbb{E}\sum_{i\neq j\in[m]}\mathbf{s}_{t_i}\mathbf{s}_{t_i}^T\sum_{k=1}^d x_k^2\mathbf{e}_k\mathbf{e}_k^T\mathbf{s}_{t_i}\mathbf{s}_{t_i}^T\mathbf{s}_{t_j}\mathbf{s}_{t_j}^T\sum_{k=1}^d x_k^2\mathbf{e}_k\mathbf{e}_k^T\mathbf{s}_{t_j}\mathbf{s}_{t_j}^T\\
&=\sum_{k_1=1}^d\sum_{j=1}^m p_{k_1}\frac{1}{m^4p_{k_1}^4}\mathbf{e}_{k_1}\mathbf{e}_{k_1}^T\sum_{k=1}^d x_k^2\mathbf{e}_k\mathbf{e}_k^T\mathbf{e}_{k_1}\mathbf{e}_{k_1}^T\mathbf{e}_{k_1}\mathbf{e}_{k_1}^T\sum_{k=1}^d x_k^2\mathbf{e}_k\mathbf{e}_k^T\mathbf{e}_{k_1}\mathbf{e}_{k_1}^T\\
&+\mathbb{E}\sum_{i\neq j\in[m]}\sum_{k_1=1}^d\sum_{q=1}^d p_{k_1}p_q\frac{1}{m^4p_{k_1}^2p_q^2}\mathbf{e}_{k_1}\mathbf{e}_{k_1}^T\sum_{k=1}^d x_k^2\mathbf{e}_k\mathbf{e}_k^T\mathbf{e}_{k_1}\mathbf{e}_{k_1}^T\mathbf{e}_q\mathbf{e}_q^T\sum_{k=1}^d x_k^2\mathbf{e}_k\mathbf{e}_k^T\mathbf{e}_q\mathbf{e}_q^T\\
&=\sum_{k=1}^d\frac{x_k^4}{m^3p_k^3}\mathbf{e}_k\mathbf{e}_k^T+\sum_{k=1}^d\frac{(m^2-m)x_k^4}{m^4p_k^2}\mathbf{e}_k\mathbf{e}_k^T\\
&=\sum_{k=1}^d\Big(\frac{1}{m^3p_k^3}+\frac{m-1}{m^3p_k^2}\Big)x_k^4\mathbf{e}_k\mathbf{e}_k^T; \qquad (27)
\end{aligned}
$$

$$
\begin{aligned}
(24)&=\mathbb{E}\sum_{j=1}^m\mathbf{s}_{t_j}\mathbf{s}_{t_j}^T\sum_{k=1}^d x_k^2\mathbf{e}_k\mathbf{e}_k^T\mathbf{s}_{t_j}\mathbf{s}_{t_j}^T\sum_{i\neq j\in[m]}\mathbf{s}_{t_i}\mathbf{s}_{t_i}^T\sum_{k=1}^d x_k^2\mathbf{e}_k\mathbf{e}_k^T\mathbf{s}_{t_j}\mathbf{s}_{t_j}^T\\
&=\mathbb{E}\sum_{g\neq i\neq j\in[m]}\mathbf{s}_{t_g}\mathbf{s}_{t_g}^T\sum_{k=1}^d x_k^2\mathbf{e}_k\mathbf{e}_k^T\mathbf{s}_{t_g}\mathbf{s}_{t_g}^T\mathbf{s}_{t_i}\mathbf{s}_{t_i}^T\sum_{k=1}^d x_k^2\mathbf{e}_k\mathbf{e}_k^T\mathbf{s}_{t_j}\mathbf{s}_{t_j}^T\\
&+\mathbb{E}\sum_{g=i\neq j\in[m]}\mathbf{s}_{t_g}\mathbf{s}_{t_g}^T\sum_{k=1}^d x_k^2\mathbf{e}_k\mathbf{e}_k^T\mathbf{s}_{t_g}\mathbf{s}_{t_g}^T\mathbf{s}_{t_i}\mathbf{s}_{t_i}^T\sum_{k=1}^d x_k^2\mathbf{e}_k\mathbf{e}_k^T\mathbf{s}_{t_j}\mathbf{s}_{t_j}^T\\
&+\mathbb{E}\sum_{g=j\neq i\in[m]}\mathbf{s}_{t_g}\mathbf{s}_{t_g}^T\sum_{k=1}^d x_k^2\mathbf{e}_k\mathbf{e}_k^T\mathbf{s}_{t_g}\mathbf{s}_{t_g}^T\mathbf{s}_{t_i}\mathbf{s}_{t_i}^T\sum_{k=1}^d x_k^2\mathbf{e}_k\mathbf{e}_k^T\mathbf{s}_{t_j}\mathbf{s}_{t_j}^T\\
&=\sum_{k_1,k_2,k_3=1}^d\frac{m(m-1)(m-2)}{m^4p_{k_1}}\mathbf{e}_{k_1}\mathbf{e}_{k_1}^T\sum_{k=1}^d x_k^2\mathbf{e}_k\mathbf{e}_k^T\mathbf{e}_{k_1}\mathbf{e}_{k_1}^T\mathbf{e}_{k_2}\mathbf{e}_{k_2}^T\sum_{k=1}^d x_k^2\mathbf{e}_k\mathbf{e}_k^T\mathbf{e}_{k_3}\mathbf{e}_{k_3}^T\\
&+\sum_{k_1,k_3=1}^d\frac{m(m-1)}{m^4p_{k_1}^2}\mathbf{e}_{k_1}\mathbf{e}_{k_1}^T\sum_{k=1}^d x_k^2\mathbf{e}_k\mathbf{e}_k^T\mathbf{e}_{k_1}\mathbf{e}_{k_1}^T\mathbf{e}_{k_1}\mathbf{e}_{k_1}^T\sum_{k=1}^d x_k^2\mathbf{e}_k\mathbf{e}_k^T\mathbf{e}_{k_3}\mathbf{e}_{k_3}^T\\
&+\sum_{k_1,k_2=1}^d\frac{m(m-1)}{m^4p_{k_1}^2}\mathbf{e}_{k_1}\mathbf{e}_{k_1}^T\sum_{k=1}^d x_k^2\mathbf{e}_k\mathbf{e}_k^T\mathbf{e}_{k_1}\mathbf{e}_{k_1}^T\mathbf{e}_{k_2}\mathbf{e}_{k_2}^T\sum_{k=1}^d x_k^2\mathbf{e}_k\mathbf{e}_k^T\mathbf{e}_{k_1}\mathbf{e}_{k_1}^T\\
&=\sum_{k_1=1}^d\frac{m(m-1)(m-2)}{m^4p_{k_1}}x_{k_1}^4\mathbf{e}_{k_1}\mathbf{e}_{k_1}^T+\sum_{k_1=1}^d\frac{m(m-1)}{m^4p_{k_1}^2}x_{k_1}^4\mathbf{e}_{k_1}\mathbf{e}_{k_1}^T+\sum_{k_1=1}^d\frac{m(m-1)}{m^4p_{k_1}^2}x_{k_1}^4\mathbf{e}_{k_1}\mathbf{e}_{k_1}^T\\
&=\sum_{k=1}^d\left[\frac{m(m-1)(m-2)}{m^4p_k}x_k^4+\frac{2m(m-1)x_k^4}{m^4p_k^2}\right]\mathbf{e}_k\mathbf{e}_k^T; \qquad (28)
\end{aligned}
$$

$$
\begin{aligned}
(25)&=\mathbb{E}\sum_{i\neq j\in[m]}\mathbf{s}_{t_i}\mathbf{s}_{t_i}^T\sum_{k=1}^d x_k^2\mathbf{e}_k\mathbf{e}_k^T\mathbf{s}_{t_j}\mathbf{s}_{t_j}^T\sum_{j=1}^m\mathbf{s}_{t_j}\mathbf{s}_{t_j}^T\sum_{k=1}^d x_k^2\mathbf{e}_k\mathbf{e}_k^T\mathbf{s}_{t_j}\mathbf{s}_{t_j}^T\\
&=\sum_{k=1}^d\left[\frac{m(m-1)(m-2)}{m^4p_k}x_k^4+\frac{2m(m-1)x_k^4}{m^4p_k^2}\right]\mathbf{e}_k\mathbf{e}_k^T; \qquad (29)
\end{aligned}
$$

$$
(26)=\mathbb{E}\sum_{i\neq j\in[m]}\mathbf{s}_{t_i}\mathbf{s}_{t_i}^T\sum_{k=1}^d x_k^2\mathbf{e}_k\mathbf{e}_k^T\mathbf{s}_{t_j}\mathbf{s}_{t_j}^T\sum_{i\neq j\in[m]}\mathbf{s}_{t_i}\mathbf{s}_{t_i}^T\sum_{k=1}^d x_k^2\mathbf{e}_k\mathbf{e}_k^T\mathbf{s}_{t_j}\mathbf{s}_{t_j}^T
$$

$$
\begin{aligned}
&= \mathbb{E} \sum_{i\neq j\neq g\neq h\in[m]} \mathbf{s}_{t_i}\mathbf{s}_{t_i}^T \sum_{k=1}^{d} x_k^2 \mathbf{e}_k\mathbf{e}_k^T \mathbf{s}_{t_j}\mathbf{s}_{t_j}^T \mathbf{s}_{t_g}\mathbf{s}_{t_g}^T \sum_{k=1}^{d} x_k^2 \mathbf{e}_k\mathbf{e}_k^T \mathbf{s}_{t_h}\mathbf{s}_{t_h}^T \\
&+ \mathbb{E} \sum_{i\neq j, i=g, j\neq h, g\neq h\in[m]} \mathbf{s}_{t_i}\mathbf{s}_{t_i}^T \sum_{k=1}^{d} x_k^2 \mathbf{e}_k\mathbf{e}_k^T \mathbf{s}_{t_j}\mathbf{s}_{t_j}^T \mathbf{s}_{t_g}\mathbf{s}_{t_g}^T \sum_{k=1}^{d} x_k^2 \mathbf{e}_k\mathbf{e}_k^T \mathbf{s}_{t_h}\mathbf{s}_{t_h}^T \\
&+ \mathbb{E} \sum_{i\neq j, i=h, j\neq g, g\neq h\in[m]} \mathbf{s}_{t_i}\mathbf{s}_{t_i}^T \sum_{k=1}^{d} x_k^2 \mathbf{e}_k\mathbf{e}_k^T \mathbf{s}_{t_j}\mathbf{s}_{t_j}^T \mathbf{s}_{t_g}\mathbf{s}_{t_g}^T \sum_{k=1}^{d} x_k^2 \mathbf{e}_k\mathbf{e}_k^T \mathbf{s}_{t_h}\mathbf{s}_{t_h}^T \\
&+ \mathbb{E} \sum_{i\neq j, i\neq g, j=h, g\neq h\in[m]} \mathbf{s}_{t_i}\mathbf{s}_{t_i}^T \sum_{k=1}^{d} x_k^2 \mathbf{e}_k\mathbf{e}_k^T \mathbf{s}_{t_j}\mathbf{s}_{t_j}^T \mathbf{s}_{t_g}\mathbf{s}_{t_g}^T \sum_{k=1}^{d} x_k^2 \mathbf{e}_k\mathbf{e}_k^T \mathbf{s}_{t_h}\mathbf{s}_{t_h}^T \\
&+ \mathbb{E} \sum_{i\neq j, i\neq h, j=g, g\neq h\in[m]} \mathbf{s}_{t_i}\mathbf{s}_{t_i}^T \sum_{k=1}^{d} x_k^2 \mathbf{e}_k\mathbf{e}_k^T \mathbf{s}_{t_j}\mathbf{s}_{t_j}^T \mathbf{s}_{t_g}\mathbf{s}_{t_g}^T \sum_{k=1}^{d} x_k^2 \mathbf{e}_k\mathbf{e}_k^T \mathbf{s}_{t_h}\mathbf{s}_{t_h}^T \\
&+ \mathbb{E} \sum_{i\neq j, i=g, j=h, g\neq h\in[m]} \mathbf{s}_{t_i}\mathbf{s}_{t_i}^T \sum_{k=1}^{d} x_k^2 \mathbf{e}_k\mathbf{e}_k^T \mathbf{s}_{t_j}\mathbf{s}_{t_j}^T \mathbf{s}_{t_g}\mathbf{s}_{t_g}^T \sum_{k=1}^{d} x_k^2 \mathbf{e}_k\mathbf{e}_k^T \mathbf{s}_{t_h}\mathbf{s}_{t_h}^T \\
&+ \mathbb{E} \sum_{i\neq j, i=h, j=g, g\neq h\in[m]} \mathbf{s}_{t_i}\mathbf{s}_{t_i}^T \sum_{k=1}^{d} x_k^2 \mathbf{e}_k\mathbf{e}_k^T \mathbf{s}_{t_j}\mathbf{s}_{t_j}^T \mathbf{s}_{t_g}\mathbf{s}_{t_g}^T \sum_{k=1}^{d} x_k^2 \mathbf{e}_k\mathbf{e}_k^T \mathbf{s}_{t_h}\mathbf{s}_{t_h}^T \\
&= \sum_{k=1}^{d} \left[\frac{m(m-1)(m-2)(m-3)}{m^4} x_k^4 + \frac{4m(m-1)(m-2)}{m^4 p_k} x_k^4 + \frac{2m(m-1)}{m^4 p_k^2} x_k^4 \right] \mathbf{e}_k\mathbf{e}_k^T. \qquad (30)
\end{aligned}
$$

$$(31)$$

Combing the above terms with simplification and reformulation completes the proof of Eq. (9).

Now, we continue to prove Eq. (10).

$$
\begin{aligned}
&\mathbb{E}\left[\mathbf{S}\mathbf{S}^T\mathbf{x}\mathbf{x}^T\mathbf{S}\mathbf{S}^T\mathbb{D}(\mathbf{S}\mathbf{S}^T\mathbf{x}\mathbf{x}^T\mathbf{S}\mathbf{S}^T)\right] \\
&= \mathbb{E}\left[\sum_{j=1}^{m}\mathbf{s}_{t_j}\mathbf{s}_{t_j}^T\mathbf{x}\sum_{j=1}^{m}\mathbf{x}^T\mathbf{s}_{t_j}\mathbf{s}_{t_j}^T\sum_{j=1}^{m}\mathbf{s}_{t_j}\mathbf{s}_{t_j}^T\sum_{k=1}^{d}x_k^2\mathbf{e}_k\mathbf{e}_k^T\sum_{j=1}^{m}\mathbf{s}_{t_j}\mathbf{s}_{t_j}^T\right] \\
&= \mathbb{E}\sum_{j=1}^{m}\mathbf{s}_{t_j}\mathbf{s}_{t_j}^T\mathbf{x}\mathbf{x}^T\mathbf{s}_{t_j}\mathbf{s}_{t_j}^T\sum_{j=1}^{m}\mathbf{s}_{t_j}\mathbf{s}_{t_j}^T\sum_{k=1}^{d}x_k^2\mathbf{e}_k\mathbf{e}_k^T\mathbf{s}_{t_j}\mathbf{s}_{t_j}^T \qquad (32)\\
&+ \mathbb{E}\sum_{j=1}^{m}\mathbf{s}_{t_j}\mathbf{s}_{t_j}^T\mathbf{x}\mathbf{x}^T\mathbf{s}_{t_j}\mathbf{s}_{t_j}^T\sum_{i\neq j\in[m]}\mathbf{s}_{t_i}\mathbf{s}_{t_i}^T\sum_{k=1}^{d}x_k^2\mathbf{e}_k\mathbf{e}_k^T\mathbf{s}_{t_j}\mathbf{s}_{t_j}^T \qquad (33)\\
&+ \mathbb{E}\sum_{i\neq j\in[m]}\mathbf{s}_{t_i}\mathbf{s}_{t_i}^T\mathbf{x}\mathbf{x}^T\mathbf{s}_{t_j}\mathbf{s}_{t_j}^T\sum_{j=1}^{m}\mathbf{s}_{t_j}\mathbf{s}_{t_j}^T\sum_{k=1}^{d}x_k^2\mathbf{e}_k\mathbf{e}_k^T\mathbf{s}_{t_j}\mathbf{s}_{t_j}^T \qquad (34)\\
&+ \mathbb{E}\sum_{i\neq j\in[m]}\mathbf{s}_{t_i}\mathbf{s}_{t_i}^T\mathbf{x}\mathbf{x}^T\mathbf{s}_{t_j}\mathbf{s}_{t_j}^T\sum_{i\neq j\in[m]}\mathbf{s}_{t_i}\mathbf{s}_{t_i}^T\sum_{k=1}^{d}x_k^2\mathbf{e}_k\mathbf{e}_k^T\mathbf{s}_{t_j}\mathbf{s}_{t_j}^T, \qquad (35)
\end{aligned}
$$

where we calculate the four terms in the last equation as shown in below:

$$
\begin{aligned}
(32) &= \mathbb{E}\sum_{j=1}^{m}\mathbf{s}_{t_j}\mathbf{s}_{t_j}^T\mathbf{x}\mathbf{x}^T\mathbf{s}_{t_j}\mathbf{s}_{t_j}^T\sum_{j=1}^{m}\mathbf{s}_{t_j}\mathbf{s}_{t_j}^T\sum_{k=1}^{d}x_k^2\mathbf{e}_k\mathbf{e}_k^T\mathbf{s}_{t_j}\mathbf{s}_{t_j}^T \\
&= \mathbb{E}\sum_{j=1}^{m}\mathbf{s}_{t_j}\mathbf{s}_{t_j}^T\mathbf{x}\mathbf{x}^T\mathbf{s}_{t_j}\mathbf{s}_{t_j}^T\mathbf{s}_{t_j}\mathbf{s}_{t_j}^T\sum_{k=1}^{d}x_k^2\mathbf{e}_k\mathbf{e}_k^T\mathbf{s}_{t_j}\mathbf{s}_{t_j}^T + \mathbb{E}\sum_{i\neq j\in[m]}\mathbf{s}_{t_i}\mathbf{s}_{t_i}^T\mathbf{x}\mathbf{x}^T\mathbf{s}_{t_i}\mathbf{s}_{t_i}^T\mathbf{s}_{t_j}\mathbf{s}_{t_j}^T\sum_{k=1}^{d}x_k^2\mathbf{e}_k\mathbf{e}_k^T\mathbf{s}_{t_j}\mathbf{s}_{t_j}^T
\end{aligned}
$$

$$= \sum_{k_1=1}^{d}\sum_{j=1}^{m} p_{k_1}\frac{1}{m^4 p_{k_1}^4}\mathbf{e}_{k_1}\mathbf{e}_{k_1}^T\mathbf{x}\mathbf{x}^T\mathbf{e}_{k_1}\mathbf{e}_{k_1}^T\mathbf{e}_{k_1}\mathbf{e}_{k_1}^T\sum_{k=1}^{d} x_k^2\mathbf{e}_k\mathbf{e}_k^T\mathbf{e}_{k_1}\mathbf{e}_{k_1}^T$$

$$+ \mathbb{E}\sum_{i\neq j\in[m]}\sum_{k_1=1}^{d}\sum_{q=1}^{d} p_{k_1}p_q\frac{1}{m^4 p_{k_1}^2 p_q^2}\mathbf{e}_{k_1}\mathbf{e}_{k_1}^T\mathbf{x}\mathbf{x}^T\mathbf{e}_{k_1}\mathbf{e}_{k_1}^T\mathbf{e}_q\mathbf{e}_q^T\sum_{k=1}^{d} x_k^2\mathbf{e}_k\mathbf{e}_k^T\mathbf{e}_q\mathbf{e}_q^T$$

$$= \sum_{k=1}^{d}\frac{x_k^4}{m^3 p_k^3}\mathbf{e}_k\mathbf{e}_k^T + \sum_{k=1}^{d}\frac{(m^2-m)x_k^4}{m^4 p_k^2}\mathbf{e}_k\mathbf{e}_k^T$$

$$= \sum_{k=1}^{d}(\frac{1}{m^3 p_k^3} + \frac{m-1}{m^3 p_k^2})x_k^4\mathbf{e}_k\mathbf{e}_k^T; \tag{36}$$

$$(33) = \mathbb{E}\sum_{j=1}^{m}\mathbf{s}_{t_j}\mathbf{s}_{t_j}^T\mathbf{x}\mathbf{x}^T\mathbf{s}_{t_j}\mathbf{s}_{t_j}^T\sum_{i\neq j\in[m]}\mathbf{s}_{t_i}\mathbf{s}_{t_i}^T\sum_{k=1}^{d} x_k^2\mathbf{e}_k\mathbf{e}_k^T\mathbf{s}_{t_j}\mathbf{s}_{t_j}^T$$

$$= \mathbb{E}\sum_{g\neq i\neq j\in[m]}\mathbf{s}_{t_g}\mathbf{s}_{t_g}^T\mathbf{x}\mathbf{x}^T\mathbf{s}_{t_g}\mathbf{s}_{t_g}^T\mathbf{s}_{t_i}\mathbf{s}_{t_i}^T\sum_{k=1}^{d} x_k^2\mathbf{e}_k\mathbf{e}_k^T\mathbf{s}_{t_j}\mathbf{s}_{t_j}^T$$

$$+ \mathbb{E}\sum_{g=i\neq j\in[m]}\mathbf{s}_{t_g}\mathbf{s}_{t_g}^T\mathbf{x}\mathbf{x}^T\mathbf{s}_{t_g}\mathbf{s}_{t_g}^T\mathbf{s}_{t_i}\mathbf{s}_{t_i}^T\sum_{k=1}^{d} x_k^2\mathbf{e}_k\mathbf{e}_k^T\mathbf{s}_{t_j}\mathbf{s}_{t_j}^T$$

$$+ \mathbb{E}\sum_{g=j\neq i\in[m]}\mathbf{s}_{t_g}\mathbf{s}_{t_g}^T\mathbf{x}\mathbf{x}^T\mathbf{s}_{t_g}\mathbf{s}_{t_g}^T\mathbf{s}_{t_i}\mathbf{s}_{t_i}^T\sum_{k=1}^{d} x_k^2\mathbf{e}_k\mathbf{e}_k^T\mathbf{s}_{t_j}\mathbf{s}_{t_j}^T$$

$$= \sum_{k_1,k_2,k_3=1}^{d}\frac{m(m-1)(m-2)}{m^4 p_{k_1}}\mathbf{e}_{k_1}\mathbf{e}_{k_1}^T\mathbf{x}\mathbf{x}^T\mathbf{e}_{k_1}\mathbf{e}_{k_1}^T\mathbf{e}_{k_2}\mathbf{e}_{k_2}^T\sum_{k=1}^{d} x_k^2\mathbf{e}_k\mathbf{e}_k^T\mathbf{e}_{k_3}\mathbf{e}_{k_3}^T$$

$$+ \sum_{k_1,k_3=1}^{d}\frac{m(m-1)}{m^4 p_{k_1}^2}\mathbf{e}_{k_1}\mathbf{e}_{k_1}^T\mathbf{x}\mathbf{x}^T\mathbf{e}_{k_1}\mathbf{e}_{k_1}^T\mathbf{e}_{k_1}\mathbf{e}_{k_1}^T\sum_{k=1}^{d} x_k^2\mathbf{e}_k\mathbf{e}_k^T\mathbf{e}_{k_3}\mathbf{e}_{k_3}^T$$

$$+ \sum_{k_1,k_2=1}^{d}\frac{m(m-1)}{m^4 p_{k_1}^2}\mathbf{e}_{k_1}\mathbf{e}_{k_1}^T\mathbf{x}\mathbf{x}^T\mathbf{e}_{k_1}\mathbf{e}_{k_1}^T\mathbf{e}_{k_2}\mathbf{e}_{k_2}^T\sum_{k=1}^{d} x_k^2\mathbf{e}_k\mathbf{e}_k^T\mathbf{e}_{k_1}\mathbf{e}_{k_1}^T$$

$$= \sum_{k_1=1}^{d}\frac{m(m-1)(m-2)}{m^4 p_{k_1}}x_{k_1}^4\mathbf{e}_{k_1}\mathbf{e}_{k_1}^T + \sum_{k_1=1}^{d}\frac{m(m-1)}{m^4 p_{k_1}^2}x_{k_1}^4\mathbf{e}_{k_1}\mathbf{e}_{k_1}^T + \sum_{k_1=1}^{d}\frac{m(m-1)}{m^4 p_{k_1}^2}x_{k_1}^4\mathbf{e}_{k_1}\mathbf{e}_{k_1}^T$$

$$= \sum_{k=1}^{d}\left[\frac{m(m-1)(m-2)}{m^4 p_k}x_k^4 + \frac{2m(m-1)}{m^4 p_k^2}x_k^4\right]\mathbf{e}_k\mathbf{e}_k^T; \tag{37}$$

$$(34) = \mathbb{E}\sum_{i\neq j\in[m]}\mathbf{s}_{t_i}\mathbf{s}_{t_i}^T\mathbf{x}\mathbf{x}^T\mathbf{s}_{t_j}\mathbf{s}_{t_j}^T\sum_{j=1}^{m}\mathbf{s}_{t_j}\mathbf{s}_{t_j}^T\sum_{k=1}^{d} x_k^2\mathbf{e}_k\mathbf{e}_k^T\mathbf{s}_{t_j}\mathbf{s}_{t_j}^T$$

$$= \mathbb{E}\sum_{i\neq j\neq g\in[m]}\mathbf{s}_{t_i}\mathbf{s}_{t_i}^T\mathbf{x}\mathbf{x}^T\mathbf{s}_{t_j}\mathbf{s}_{t_j}^T\mathbf{s}_{t_g}\mathbf{s}_{t_g}^T\sum_{k=1}^{d} x_k^2\mathbf{e}_k\mathbf{e}_k^T\mathbf{s}_{t_g}\mathbf{s}_{t_g}^T$$

$$+ \mathbb{E}\sum_{i\neq j=g\in[m]}\mathbf{s}_{t_i}\mathbf{s}_{t_i}^T\mathbf{x}\mathbf{x}^T\mathbf{s}_{t_j}\mathbf{s}_{t_j}^T\mathbf{s}_{t_g}\mathbf{s}_{t_g}^T\sum_{k=1}^{d} x_k^2\mathbf{e}_k\mathbf{e}_k^T\mathbf{s}_{t_g}\mathbf{s}_{t_g}^T$$

$$+ \mathbb{E}\sum_{i=g\neq j\in[m]}\mathbf{s}_{t_i}\mathbf{s}_{t_i}^T\mathbf{x}\mathbf{x}^T\mathbf{s}_{t_j}\mathbf{s}_{t_j}^T\mathbf{s}_{t_g}\mathbf{s}_{t_g}^T\sum_{k=1}^{d} x_k^2\mathbf{e}_k\mathbf{e}_k^T\mathbf{s}_{t_g}\mathbf{s}_{t_g}^T$$

$$
\begin{aligned}
&= \sum_{k_1,k_2,k_3=1}^{d} \frac{m(m-1)(m-2)}{m^4 p_{k_3}} \mathbf{e}_{k_1}\mathbf{e}_{k_1}^T\mathbf{x}\mathbf{x}^T\mathbf{e}_{k_2}\mathbf{e}_{k_2}^T\mathbf{e}_{k_3}\mathbf{e}_{k_3}^T \sum_{k=1}^{d} x_k^2\mathbf{e}_k\mathbf{e}_k^T\mathbf{e}_{k_3}\mathbf{e}_{k_3}^T \\
&+ \sum_{k_1,k_3=1}^{d} \frac{m(m-1)}{m^4 p_{k_3}^2} \mathbf{e}_{k_1}\mathbf{e}_{k_1}^T\mathbf{x}\mathbf{x}^T\mathbf{e}_{k_3}\mathbf{e}_{k_3}^T\mathbf{e}_{k_3}\mathbf{e}_{k_3}^T \sum_{k=1}^{d} x_k^2\mathbf{e}_k\mathbf{e}_k^T\mathbf{e}_{k_3}\mathbf{e}_{k_3}^T \\
&+ \sum_{k_2,k_3=1}^{d} \frac{m(m-1)}{m^4 p_{k_3}^2} \mathbf{e}_{k_3}\mathbf{e}_{k_3}^T\mathbf{x}\mathbf{x}^T\mathbf{e}_{k_2}\mathbf{e}_{k_2}^T\mathbf{e}_{k_3}\mathbf{e}_{k_3}^T \sum_{k=1}^{d} x_k^2\mathbf{e}_k\mathbf{e}_k^T\mathbf{e}_{k_3}\mathbf{e}_{k_3}^T \\
&= \sum_{k_1,k_3=1}^{d} \frac{m(m-1)(m-2)}{m^4 p_{k_3}} x_{k_1}x_{k_3}^3\mathbf{e}_{k_1}\mathbf{e}_{k_3}^T + \sum_{k_1,k_3=1}^{d} \frac{m(m-1)}{m^4 p_{k_3}^2} x_{k_1}x_{k_3}^3\mathbf{e}_{k_1}\mathbf{e}_{k_3}^T + \sum_{k_3=1}^{d} \frac{m(m-1)}{m^4 p_{k_3}^2} x_{k_3}^4\mathbf{e}_{k_3}\mathbf{e}_{k_3}^T \\
&= \frac{m(m-1)(m-2)}{m^4}\mathbf{x}\mathbf{x}^T\mathbb{D}(\{\frac{x_k^2}{p_k}\}) + \frac{m(m-1)}{m^4}\mathbf{x}\mathbf{x}^T\mathbb{D}(\{\frac{x_k^2}{p_k^2}\}) + \frac{m(m-1)}{m^4}\sum_{k=1}^{d}\frac{x_k^4}{p_k^2}\mathbf{e}_k\mathbf{e}_k^T; \qquad (38)
\end{aligned}
$$

$$
\begin{aligned}
(35) &= \mathbb{E}\sum_{i\neq j\in[m]} \mathbf{s}_{t_i}\mathbf{s}_{t_i}^T\mathbf{x}\mathbf{x}^T\mathbf{s}_{t_j}\mathbf{s}_{t_j}^T \sum_{i\neq j\in[m]} \mathbf{s}_{t_i}\mathbf{s}_{t_i}^T \sum_{k=1}^{d} x_k^2\mathbf{e}_k\mathbf{e}_k^T\mathbf{s}_{t_j}\mathbf{s}_{t_j}^T \\
&= \mathbb{E}\sum_{i\neq j\neq g\neq h\in[m]} \mathbf{s}_{t_i}\mathbf{s}_{t_i}^T\mathbf{x}\mathbf{x}^T\mathbf{s}_{t_j}\mathbf{s}_{t_j}^T\mathbf{s}_{t_g}\mathbf{s}_{t_g}^T \sum_{k=1}^{d} x_k^2\mathbf{e}_k\mathbf{e}_k^T\mathbf{s}_{t_h}\mathbf{s}_{t_h}^T \\
&+ \mathbb{E}\sum_{i\neq j, i=g, j\neq h, g\neq h\in[m]} \mathbf{s}_{t_i}\mathbf{s}_{t_i}^T\mathbf{x}\mathbf{x}^T\mathbf{s}_{t_j}\mathbf{s}_{t_j}^T\mathbf{s}_{t_g}\mathbf{s}_{t_g}^T \sum_{k=1}^{d} x_k^2\mathbf{e}_k\mathbf{e}_k^T\mathbf{s}_{t_h}\mathbf{s}_{t_h}^T \\
&+ \mathbb{E}\sum_{i\neq j, i=h, j\neq g, g\neq h\in[m]} \mathbf{s}_{t_i}\mathbf{s}_{t_i}^T\mathbf{x}\mathbf{x}^T\mathbf{s}_{t_j}\mathbf{s}_{t_j}^T\mathbf{s}_{t_g}\mathbf{s}_{t_g}^T \sum_{k=1}^{d} x_k^2\mathbf{e}_k\mathbf{e}_k^T\mathbf{s}_{t_h}\mathbf{s}_{t_h}^T \\
&+ \mathbb{E}\sum_{i\neq j, i\neq g, j=h, g\neq h\in[m]} \mathbf{s}_{t_i}\mathbf{s}_{t_i}^T\mathbf{x}\mathbf{x}^T\mathbf{s}_{t_j}\mathbf{s}_{t_j}^T\mathbf{s}_{t_g}\mathbf{s}_{t_g}^T \sum_{k=1}^{d} x_k^2\mathbf{e}_k\mathbf{e}_k^T\mathbf{s}_{t_h}\mathbf{s}_{t_h}^T \\
&+ \mathbb{E}\sum_{i\neq j, i\neq h, j=g, g\neq h\in[m]} \mathbf{s}_{t_i}\mathbf{s}_{t_i}^T\mathbf{x}\mathbf{x}^T\mathbf{s}_{t_j}\mathbf{s}_{t_j}^T\mathbf{s}_{t_g}\mathbf{s}_{t_g}^T \sum_{k=1}^{d} x_k^2\mathbf{e}_k\mathbf{e}_k^T\mathbf{s}_{t_h}\mathbf{s}_{t_h}^T \\
&+ \mathbb{E}\sum_{i\neq j, i=g, j=h, g\neq h\in[m]} \mathbf{s}_{t_i}\mathbf{s}_{t_i}^T\mathbf{x}\mathbf{x}^T\mathbf{s}_{t_j}\mathbf{s}_{t_j}^T\mathbf{s}_{t_g}\mathbf{s}_{t_g}^T \sum_{k=1}^{d} x_k^2\mathbf{e}_k\mathbf{e}_k^T\mathbf{s}_{t_h}\mathbf{s}_{t_h}^T \\
&+ \mathbb{E}\sum_{i\neq j, i=h, j=g, g\neq h\in[m]} \mathbf{s}_{t_i}\mathbf{s}_{t_i}^T\mathbf{x}\mathbf{x}^T\mathbf{s}_{t_j}\mathbf{s}_{t_j}^T\mathbf{s}_{t_g}\mathbf{s}_{t_g}^T \sum_{k=1}^{d} x_k^2\mathbf{e}_k\mathbf{e}_k^T\mathbf{s}_{t_h}\mathbf{s}_{t_h}^T \\
&= \sum_{k_1,k_2,k_3,k_4=1}^{d} \frac{m(m-1)(m-2)(m-3)}{m^4} x_{k_1}x_{k_2}\mathbf{e}_{k_1}\mathbf{e}_{k_2}^T\mathbf{e}_{k_3}\mathbf{e}_{k_3}^T \sum_{k=1}^{d} x_k^2\mathbf{e}_k\mathbf{e}_k^T\mathbf{e}_{k_4}\mathbf{e}_{k_4}^T \\
&+ \sum_{k_1,k_2,k_4=1}^{d} \frac{m(m-1)(m-2)}{m^4 p_{k_1}} x_{k_1}x_{k_2}\mathbf{e}_{k_1}\mathbf{e}_{k_2}^T\mathbf{e}_{k_1}\mathbf{e}_{k_1}^T \sum_{k=1}^{d} x_k^2\mathbf{e}_k\mathbf{e}_k^T\mathbf{e}_{k_4}\mathbf{e}_{k_4}^T \\
&+ \sum_{k_1,k_2,k_3=1}^{d} \frac{m(m-1)(m-2)}{m^4 p_{k_1}} x_{k_1}x_{k_2}\mathbf{e}_{k_1}\mathbf{e}_{k_2}^T\mathbf{e}_{k_3}\mathbf{e}_{k_3}^T \sum_{k=1}^{d} x_k^2\mathbf{e}_k\mathbf{e}_k^T\mathbf{e}_{k_1}\mathbf{e}_{k_1}^T \\
&+ \sum_{k_1,k_2,k_3=1}^{d} \frac{m(m-1)(m-2)}{m^4 p_{k_2}} x_{k_1}x_{k_2}\mathbf{e}_{k_1}\mathbf{e}_{k_2}^T\mathbf{e}_{k_3}\mathbf{e}_{k_3}^T \sum_{k=1}^{d} x_k^2\mathbf{e}_k\mathbf{e}_k^T\mathbf{e}_{k_2}\mathbf{e}_{k_2}^T
\end{aligned}
$$

$$
\begin{aligned}
&+\sum_{k_1,k_2,k_4=1}^{d}\frac{m(m-1)(m-2)}{m^4p_{k_2}}x_{k_1}x_{k_2}\mathbf{e}_{k_1}\mathbf{e}_{k_2}^T\mathbf{e}_{k_2}\mathbf{e}_{k_2}^T\sum_{k=1}^{d}x_k^2\mathbf{e}_k\mathbf{e}_k^T\mathbf{e}_{k_4}\mathbf{e}_{k_4}^T\\
&+\sum_{k_1,k_2}^{d}\frac{m(m-1)}{m^4p_{k_1}p_{k_2}}x_{k_1}x_{k_2}\mathbf{e}_{k_1}\mathbf{e}_{k_2}^T\mathbf{e}_{k_1}\mathbf{e}_{k_1}^T\sum_{k=1}^{d}x_k^2\mathbf{e}_k\mathbf{e}_k^T\mathbf{e}_{k_2}\mathbf{e}_{k_2}^T\\
&+\sum_{k_1,k_2=1}^{d}\frac{m(m-1)}{m^4p_{k_1}p_{k_2}}x_{k_1}x_{k_2}\mathbf{e}_{k_1}\mathbf{e}_{k_2}^T\mathbf{e}_{k_2}\mathbf{e}_{k_2}^T\sum_{k=1}^{d}x_k^2\mathbf{e}_k\mathbf{e}_k^T\mathbf{e}_{k_1}\mathbf{e}_{k_1}^T\\
&=\sum_{k_1,k_2=1}^{d}\frac{m(m-1)(m-2)(m-3)}{m^4}x_{k_1}x_{k_2}^3\mathbf{e}_{k_1}\mathbf{e}_{k_2}^T+\sum_{k_1=1}^{d}\frac{m(m-1)(m-2)}{m^4p_{k_1}}x_{k_1}^4\mathbf{e}_{k_1}\mathbf{e}_{k_1}^T\\
&+\sum_{k_1=1}^{d}\frac{m(m-1)(m-2)}{m^4p_{k_1}}x_{k_1}^4\mathbf{e}_{k_1}\mathbf{e}_{k_1}^T+\sum_{k_1,k_2=1}^{d}\frac{m(m-1)(m-2)}{m^4p_{k_2}}x_{k_1}x_{k_2}^3\mathbf{e}_{k_1}\mathbf{e}_{k_2}^T\\
&+\sum_{k_1,k_2=1}^{d}\frac{m(m-1)(m-2)}{m^4p_{k_2}}x_{k_1}x_{k_2}^3\mathbf{e}_{k_1}\mathbf{e}_{k_2}^T+\sum_{k_1=1}^{d}\frac{m(m-1)}{m^4p_{k_1}^2}x_{k_1}^4\mathbf{e}_{k_1}\mathbf{e}_{k_1}^T+\sum_{k_1=1}^{d}\frac{m(m-1)}{m^4p_{k_1}^2}x_{k_1}^4\mathbf{e}_{k_1}\mathbf{e}_{k_1}^T\\
&=\frac{m(m-1)(m-2)(m-3)}{m^4}\mathbf{x}\mathbf{x}^T\mathbb{D}(\{x_k^2\})+\frac{m(m-1)(m-2)}{m^4}\sum_{k=1}^{d}\frac{x_k^4}{p_k}\mathbf{e}_k\mathbf{e}_k^T\\
&+\frac{m(m-1)(m-2)}{m^4}\sum_{k=1}^{d}\frac{x_k^4}{p_k}\mathbf{e}_k\mathbf{e}_k^T+\frac{m(m-1)(m-2)}{m^4}\mathbf{x}\mathbf{x}^T\mathbb{D}(\{\frac{x_k^2}{p_k}\})\\
&+\frac{m(m-1)(m-2)}{m^4}\mathbf{x}\mathbf{x}^T\mathbb{D}(\{\frac{x_k^2}{p_k}\})+\frac{2m(m-1)}{m^4}\sum_{k=1}^{d}\frac{x_k^4}{p_k^2}\mathbf{e}_k\mathbf{e}_k^T.
\end{aligned}\tag{39}
$$

Combing the above terms with simplification and reformulation completes the proof of Eq. (10).

Finally, we have to prove Eq. (11).

$$
\begin{aligned}
\mathbb{E}\left[(\mathbf{S}\mathbf{S}^T\mathbf{x}\mathbf{x}^T\mathbf{S}\mathbf{S}^T)^2\right]&=\mathbb{E}\left[(\sum_{j=1}^{m}\mathbf{s}_{t_j}\mathbf{s}_{t_j}^T\mathbf{x}\sum_{j=1}^{m}\mathbf{x}^T\mathbf{s}_{t_j}\mathbf{s}_{t_j}^T)^2\right]\\
&=\mathbb{E}(\sum_{j=1}^{m}\mathbf{s}_{t_j}\mathbf{s}_{t_j}^T\mathbf{x}\mathbf{x}^T\mathbf{s}_{t_j}\mathbf{s}_{t_j}^T+\sum_{i\neq j\in[m]}\mathbf{s}_{t_i}\mathbf{s}_{t_i}^T\mathbf{x}\mathbf{x}^T\mathbf{s}_{t_j}\mathbf{s}_{t_j}^T)^2
\end{aligned}
$$

$$
=\mathbb{E}(\sum_{j=1}^{m}\mathbf{s}_{t_j}\mathbf{s}_{t_j}^T\mathbf{x}\mathbf{x}^T\mathbf{s}_{t_j}\mathbf{s}_{t_j}^T)^2\tag{40}
$$

$$
+\mathbb{E}(\sum_{i\neq j\in[m]}\mathbf{s}_{t_i}\mathbf{s}_{t_i}^T\mathbf{x}\mathbf{x}^T\mathbf{s}_{t_j}\mathbf{s}_{t_j}^T)^2\tag{41}
$$

$$
+\mathbb{E}\sum_{j=1}^{m}\mathbf{s}_{t_j}\mathbf{s}_{t_j}^T\mathbf{x}\mathbf{x}^T\mathbf{s}_{t_j}\mathbf{s}_{t_j}^T\sum_{i\neq j\in[m]}\mathbf{s}_{t_i}\mathbf{s}_{t_i}^T\mathbf{x}\mathbf{x}^T\mathbf{s}_{t_j}\mathbf{s}_{t_j}^T\tag{42}
$$

$$
+\mathbb{E}\sum_{i\neq j\in[m]}\mathbf{s}_{t_i}\mathbf{s}_{t_i}^T\mathbf{x}\mathbf{x}^T\mathbf{s}_{t_j}\mathbf{s}_{t_j}^T\sum_{j=1}^{m}\mathbf{s}_{t_j}\mathbf{s}_{t_j}^T\mathbf{x}\mathbf{x}^T\mathbf{s}_{t_j}\mathbf{s}_{t_j}^T,\tag{43}
$$

where we calculate the four terms in the last equation as shown in below:

$$
(40)=\mathbb{E}\sum_{j=1}^{m}\mathbf{s}_{t_j}\mathbf{s}_{t_j}^T\mathbf{x}\mathbf{x}^T\mathbf{s}_{t_j}\mathbf{s}_{t_j}^T\mathbf{s}_{t_j}\mathbf{s}_{t_j}^T\mathbf{x}\mathbf{x}^T\mathbf{s}_{t_j}\mathbf{s}_{t_j}^T+\mathbb{E}\sum_{i\neq j\in[m]}\mathbf{s}_{t_i}\mathbf{s}_{t_i}^T\mathbf{x}\mathbf{x}^T\mathbf{s}_{t_i}\mathbf{s}_{t_i}^T\mathbf{s}_{t_j}\mathbf{s}_{t_j}^T\mathbf{x}\mathbf{x}^T\mathbf{s}_{t_j}\mathbf{s}_{t_j}^T
$$

$$= \sum_{k=1}^{d}\sum_{j=1}^{m} p_k \frac{1}{m^4 p_k^4}\mathbf{e}_k\mathbf{e}_k^T\mathbf{x}\mathbf{x}^T\mathbf{e}_k\mathbf{e}_k^T\mathbf{e}_k\mathbf{e}_k^T\mathbf{x}\mathbf{x}^T\mathbf{e}_k\mathbf{e}_k^T$$

$$+ \mathbb{E}\sum_{i\neq j\in[m]}\sum_{k=1}^{d}\sum_{q=1}^{d} p_k p_q \frac{1}{m^4 p_k^2 p_q^2}\mathbf{e}_k\mathbf{e}_k^T\mathbf{x}\mathbf{x}^T\mathbf{e}_k\mathbf{e}_k^T\mathbf{e}_q\mathbf{e}_q^T\mathbf{x}\mathbf{x}^T\mathbf{e}_q\mathbf{e}_q^T$$

$$= \sum_{k=1}^{d}\frac{x_k^4}{m^3 p_k^3}\mathbf{e}_k\mathbf{e}_k^T + \sum_{k=1}^{d}\frac{(m^2-m)x_k^4}{m^4 p_k^2}\mathbf{e}_k\mathbf{e}_k^T$$

$$= \sum_{k=1}^{d}(\frac{1}{m^3 p_k^3} + \frac{m-1}{m^3 p_k^2})x_k^4\mathbf{e}_k\mathbf{e}_k^T; \tag{44}$$

$$(41) = \mathbb{E}\left[\sum_{i\neq j\in[m]}\mathbf{s}_{t_i}\mathbf{s}_{t_i}^T\mathbf{x}\mathbf{x}^T\mathbf{s}_{t_j}\mathbf{s}_{t_j}^T \sum_{i\neq j\in[m]}\mathbf{s}_{t_i}\mathbf{s}_{t_i}^T\mathbf{x}\mathbf{x}^T\mathbf{s}_{t_j}\mathbf{s}_{t_j}^T\right]$$

$$= \mathbb{E}\sum_{i\neq j\neq g\neq h\in[m]}\mathbf{s}_{t_i}\mathbf{s}_{t_i}^T\mathbf{x}\mathbf{x}^T\mathbf{s}_{t_j}\mathbf{s}_{t_j}^T\mathbf{s}_{t_g}\mathbf{s}_{t_g}^T\mathbf{x}\mathbf{x}^T\mathbf{s}_{t_h}\mathbf{s}_{t_h}^T + \mathbb{E}\sum_{i\neq j, i=g, j\neq h, g\neq h\in[m]}\mathbf{s}_{t_i}\mathbf{s}_{t_i}^T\mathbf{x}\mathbf{x}^T\mathbf{s}_{t_j}\mathbf{s}_{t_j}^T\mathbf{s}_{t_g}\mathbf{s}_{t_g}^T\mathbf{x}\mathbf{x}^T\mathbf{s}_{t_h}\mathbf{s}_{t_h}^T$$

$$+ \mathbb{E}\sum_{i\neq j, i=h, j\neq g, g\neq h\in[m]}\mathbf{s}_{t_i}\mathbf{s}_{t_i}^T\mathbf{x}\mathbf{x}^T\mathbf{s}_{t_j}\mathbf{s}_{t_j}^T\mathbf{s}_{t_g}\mathbf{s}_{t_g}^T\mathbf{x}\mathbf{x}^T\mathbf{s}_{t_h}\mathbf{s}_{t_h}^T + \mathbb{E}\sum_{i\neq j, i\neq g, j=h, g\neq h\in[m]}\mathbf{s}_{t_i}\mathbf{s}_{t_i}^T\mathbf{x}\mathbf{x}^T\mathbf{s}_{t_j}\mathbf{s}_{t_j}^T\mathbf{s}_{t_g}\mathbf{s}_{t_g}^T\mathbf{x}\mathbf{x}^T\mathbf{s}_{t_h}\mathbf{s}_{t_h}^T$$

$$+ \mathbb{E}\sum_{i\neq j, i\neq h, j=g, g\neq h\in[m]}\mathbf{s}_{t_i}\mathbf{s}_{t_i}^T\mathbf{x}\mathbf{x}^T\mathbf{s}_{t_j}\mathbf{s}_{t_j}^T\mathbf{s}_{t_g}\mathbf{s}_{t_g}^T\mathbf{x}\mathbf{x}^T\mathbf{s}_{t_h}\mathbf{s}_{t_h}^T + \mathbb{E}\sum_{i\neq j, i=g, j=h, g\neq h\in[m]}\mathbf{s}_{t_i}\mathbf{s}_{t_i}^T\mathbf{x}\mathbf{x}^T\mathbf{s}_{t_j}\mathbf{s}_{t_j}^T\mathbf{s}_{t_g}\mathbf{s}_{t_g}^T\mathbf{x}\mathbf{x}^T\mathbf{s}_{t_h}\mathbf{s}_{t_h}^T$$

$$+ \mathbb{E}\sum_{i\neq j, i=h, j=g, g\neq h\in[m]}\mathbf{s}_{t_i}\mathbf{s}_{t_i}^T\mathbf{x}\mathbf{x}^T\mathbf{s}_{t_j}\mathbf{s}_{t_j}^T\mathbf{s}_{t_g}\mathbf{s}_{t_g}^T\mathbf{x}\mathbf{x}^T\mathbf{s}_{t_h}\mathbf{s}_{t_h}^T$$

$$= \sum_{k_1,k_2,k_3,k_4=1}^{d}\frac{m(m-1)(m-2)(m-3)}{m^4}x_{k_1}x_{k_2}x_{k_3}x_{k_4}\mathbf{e}_{k_1}\mathbf{e}_{k_2}^T\mathbf{e}_{k_3}\mathbf{e}_{k_4}^T$$

$$+ \sum_{k_1,k_2,k_4=1}^{d}\frac{m(m-1)(m-2)}{m^4 p_{k_1}}x_{k_1}^2 x_{k_2}x_{k_4}\mathbf{e}_{k_1}\mathbf{e}_{k_2}^T\mathbf{e}_{k_1}\mathbf{e}_{k_4}^T + \sum_{k_1,k_2,k_3=1}^{d}\frac{m(m-1)(m-2)}{m^4 p_{k_1}}x_{k_1}^2 x_{k_2}x_{k_3}\mathbf{e}_{k_1}\mathbf{e}_{k_2}^T\mathbf{e}_{k_3}\mathbf{e}_{k_1}^T$$

$$+ \sum_{k_1,k_2,k_3=1}^{d}\frac{m(m-1)(m-2)}{m^4 p_{k_2}}x_{k_1}x_{k_2}^2 x_{k_3}\mathbf{e}_{k_1}\mathbf{e}_{k_2}^T\mathbf{e}_{k_3}\mathbf{e}_{k_2}^T + \sum_{k_1,k_2,k_4=1}^{d}\frac{m(m-1)(m-2)}{m^4 p_{k_2}}x_{k_1}x_{k_2}^2 x_{k_4}\mathbf{e}_{k_1}\mathbf{e}_{k_2}^T\mathbf{e}_{k_2}\mathbf{e}_{k_4}^T$$

$$+ \sum_{k_1,k_2}^{d}\frac{m(m-1)}{m^4 p_{k_1}p_{k_2}}x_{k_1}^2 x_{k_2}^2\mathbf{e}_{k_1}\mathbf{e}_{k_2}^T\mathbf{e}_{k_1}\mathbf{e}_{k_2}^T + \sum_{k_1,k_2=1}^{d}\frac{m(m-1)}{m^4 p_{k_1}p_{k_2}}x_{k_1}^2 x_{k_2}^2\mathbf{e}_{k_1}\mathbf{e}_{k_2}^T\mathbf{e}_{k_2}\mathbf{e}_{k_1}^T$$

$$= \sum_{k_2=1}^{d}x_{k_2}^2\sum_{k_1,k_4=1}^{d}\frac{m(m-1)(m-2)(m-3)}{m^4}x_{k_1}x_{k_4}\mathbf{e}_{k_1}\mathbf{e}_{k_4}^T$$

$$+ \sum_{k_1,k_4=1}^{d}\frac{m(m-1)(m-2)}{m^4 p_{k_1}}x_{k_1}^3 x_{k_4}\mathbf{e}_{k_1}\mathbf{e}_{k_4}^T + \sum_{k_2=1}^{d}x_{k_2}^2\sum_{k_1=1}^{d}\frac{m(m-1)(m-2)}{m^4 p_{k_1}}x_{k_1}^2\mathbf{e}_{k_1}\mathbf{e}_{k_1}^T$$

$$+ \sum_{k_1,k_2=1}^{d}\frac{m(m-1)(m-2)}{m^4 p_{k_2}}x_{k_1}x_{k_2}^3\mathbf{e}_{k_1}\mathbf{e}_{k_2}^T + \sum_{k_2=1}^{d}\frac{x_{k_2}^2}{p_{k_2}}\sum_{k_1,k_4=1}^{d}\frac{m(m-1)(m-2)}{m^4}x_{k_1}x_{k_4}\mathbf{e}_{k_1}\mathbf{e}_{k_4}^T$$

$$+ \sum_{k_1=1}^{d}\frac{m(m-1)}{m^4 p_{k_1}^2}x_{k_1}^4\mathbf{e}_{k_1}\mathbf{e}_{k_1}^T + \sum_{k_2=1}^{d}\frac{x_{k_2}^2}{p_{k_2}}\sum_{k_1=1}^{d}\frac{m(m-1)}{m^4 p_{k_1}}x_{k_1}^2\mathbf{e}_{k_1}\mathbf{e}_{k_1}^T$$

$$= \frac{\|\mathbf{x}\|_2^2 m(m-1)(m-2)(m-3)}{m^4}\mathbf{x}\mathbf{x}^T + \frac{m(m-1)(m-2)}{m^4}\mathbb{D}(\{\frac{x_k^2}{p_k}\})\mathbf{x}\mathbf{x}^T$$

$$
+\frac{\|\mathbf{x}\|_2^2 m(m-1)(m-2)}{m^4}\sum_{k=1}^{d}\frac{x_k^2}{p_k}\mathbf{e}_k\mathbf{e}_k^T+\frac{m(m-1)(m-2)}{m^4}\mathbf{x}\mathbf{x}^T\mathbb{D}(\{\frac{x_k^2}{p_k}\})
$$

$$
+\frac{m(m-1)(m-2)}{m^4}\sum_{k=1}^{d}\frac{x_k^2}{p_k}\mathbf{x}\mathbf{x}^T+\frac{m(m-1)}{m^4}\sum_{k=1}^{d}\frac{x_k^4}{p_k^2}\mathbf{e}_k\mathbf{e}_k^T+\frac{m(m-1)}{m^4}\sum_{k=1}^{d}\frac{x_k^2}{p_k}\sum_{k=1}^{d}\frac{x_k^2}{p_k}\mathbf{e}_k\mathbf{e}_k^T; \tag{45}
$$

$$
\begin{aligned}
(42) &= \mathbb{E}\sum_{j=1}^{m}\mathbf{s}_{t_j}\mathbf{s}_{t_j}^T\mathbf{x}\mathbf{x}^T\mathbf{s}_{t_j}\mathbf{s}_{t_j}^T\sum_{i\neq j\in[m]}\mathbf{s}_{t_i}\mathbf{s}_{t_i}^T\mathbf{x}\mathbf{x}^T\mathbf{s}_{t_j}\mathbf{s}_{t_j}^T\\
&= \mathbb{E}\sum_{g\neq i\neq j\in[m]}\mathbf{s}_{t_g}\mathbf{s}_{t_g}^T\mathbf{x}\mathbf{x}^T\mathbf{s}_{t_g}\mathbf{s}_{t_g}^T\mathbf{s}_{t_i}\mathbf{s}_{t_i}^T\mathbf{x}\mathbf{x}^T\mathbf{s}_{t_j}\mathbf{s}_{t_j}^T+\mathbb{E}\sum_{g=i\neq j\in[m]}\mathbf{s}_{t_g}\mathbf{s}_{t_g}^T\mathbf{x}\mathbf{x}^T\mathbf{s}_{t_g}\mathbf{s}_{t_g}^T\mathbf{s}_{t_i}\mathbf{s}_{t_i}^T\mathbf{x}\mathbf{x}^T\mathbf{s}_{t_j}\mathbf{s}_{t_j}^T\\
&+\mathbb{E}\sum_{g=j\neq i\in[m]}\mathbf{s}_{t_g}\mathbf{s}_{t_g}^T\mathbf{x}\mathbf{x}^T\mathbf{s}_{t_g}\mathbf{s}_{t_g}^T\mathbf{s}_{t_i}\mathbf{s}_{t_i}^T\mathbf{x}\mathbf{x}^T\mathbf{s}_{t_j}\mathbf{s}_{t_j}^T\\
&= \sum_{k_1,k_2,k_3=1}^{d}\frac{m(m-1)(m-2)}{m^4p_{k_1}}\mathbf{e}_{k_1}\mathbf{e}_{k_1}^T\mathbf{x}\mathbf{x}^T\mathbf{e}_{k_1}\mathbf{e}_{k_1}^T\mathbf{e}_{k_2}\mathbf{e}_{k_2}^T\mathbf{x}\mathbf{x}^T\mathbf{e}_{k_3}\mathbf{e}_{k_3}^T\\
&+\sum_{k_1,k_3=1}^{d}\frac{m(m-1)}{m^4p_{k_1}^2}\mathbf{e}_{k_1}\mathbf{e}_{k_1}^T\mathbf{x}\mathbf{x}^T\mathbf{e}_{k_1}\mathbf{e}_{k_1}^T\mathbf{e}_{k_1}\mathbf{e}_{k_1}^T\mathbf{x}\mathbf{x}^T\mathbf{e}_{k_3}\mathbf{e}_{k_3}^T+\sum_{k_1,k_2=1}^{d}\frac{m(m-1)}{m^4p_{k_1}^2}\mathbf{e}_{k_1}\mathbf{e}_{k_1}^T\mathbf{x}\mathbf{x}^T\mathbf{e}_{k_1}\mathbf{e}_{k_1}^T\mathbf{e}_{k_2}\mathbf{e}_{k_2}^T\mathbf{x}\mathbf{x}^T\mathbf{e}_{k_1}\mathbf{e}_{k_1}^T\\
&= \sum_{k_1,k_3=1}^{d}\frac{m(m-1)(m-2)}{m^4p_{k_1}}x_{k_1}^3x_{k_3}\mathbf{e}_{k_1}\mathbf{e}_{k_3}^T+\sum_{k_1,k_3=1}^{d}\frac{m(m-1)}{m^4p_{k_1}^2}x_{k_1}^3x_{k_3}\mathbf{e}_{k_1}\mathbf{e}_{k_3}^T+\sum_{k_1=1}^{d}\frac{m(m-1)}{m^4p_{k_1}^2}x_{k_1}^4\mathbf{e}_{k_1}\mathbf{e}_{k_1}^T\\
&= \frac{m(m-1)(m-2)}{m^4}\mathbb{D}(\{\frac{x_k^2}{p_k}\})\mathbf{x}\mathbf{x}^T+\frac{m(m-1)}{m^4}\mathbb{D}(\{\frac{x_k^2}{p_k^2}\})\mathbf{x}\mathbf{x}^T+\frac{m(m-1)}{m^4}\sum_{k=1}^{d}\frac{x_k^4}{p_k^2}\mathbf{e}_k\mathbf{e}_k^T;
\end{aligned} \tag{46}
$$

$$
(43) = \frac{m(m-1)(m-2)}{m^4}\mathbf{x}\mathbf{x}^T\mathbb{D}(\{\frac{x_k^2}{p_k}\})+\frac{m(m-1)}{m^4}\mathbf{x}\mathbf{x}^T\mathbb{D}(\{\frac{x_k^2}{p_k^2}\})+\frac{m(m-1)}{m^4}\sum_{k=1}^{d}\frac{x_k^4}{p_k^2}\mathbf{e}_k\mathbf{e}_k^T. \tag{47}
$$

Combing the above terms with simplification and reformulation completes the proof of Eq. (11). To this end, we complete the whole proof. $\square$

2.3. Proof of Lemma 2

Proof. According to the setting, we have that

$$
\begin{aligned}
\|\mathbf{S}\mathbf{S}^T\mathbf{x}\mathbf{x}^T\mathbf{S}\mathbf{S}^T\|_2 &\overset{(a)}{=} \|\mathbf{S}\mathbf{S}^T\mathbf{x}\|_2^2=\|\sum_{j=1}^{m}\mathbf{s}_{t_j}\mathbf{s}_{t_j}^T\mathbf{x}\|_2^2=\|\sum_{j=1}^{m}\frac{1}{mp_{t_j}}x_{t_j}\mathbf{e}_{t_j}\|_2^2\\
&=\|\sum_{j=1}^{m}\sum_{k=1}^{d}\frac{\delta_{t_jk}}{mp_k}x_k\mathbf{e}_k\|_2^2=\sum_{k=1}^{d}(\sum_{j=1}^{m}\frac{\delta_{t_jk}x_k}{mp_k})^2\overset{(b)}{=}\sum_{k\in\Gamma}(\sum_{j=1}^{m}\frac{\delta_{t_jk}x_k}{mp_k})^2,
\end{aligned} \tag{48}
$$

where we let $\Gamma=\{\gamma_t\}_{t=1}^{|\Gamma|}$ be a set containing at most m different elements of $[d]$ with its cardinality $|\Gamma|\leq m$.

In Eq. (48), (a) follows because $\mathbf{S}\mathbf{S}^T\mathbf{x}\mathbf{x}^T\mathbf{S}\mathbf{S}^T$ is a positive semidefinite matrix of rank 1, δ_{t_jk} returns 1 only when $t_j=k$ and 0 otherwise, and $\mathbb{P}(\delta_{t_jk}=1)=\mathbb{P}(t_j=k)=p_k$. (b) holds due to that we perform random sampling with replacement m times on the d entries of $\mathbf{x}\in\mathbb{R}^d$, and consequently at most m certain different entries from $\mathbf{x}$ are sampled.

Let $k=\gamma_1$ with $\gamma_1\in\Gamma$, and we first bound $|\sum_{j=1}^{m}\frac{\delta_{t_j\gamma_1}x_{\gamma_1}}{mp_{\gamma_1}}|$. Define a random variable $a_j=\frac{\delta_{t_j\gamma_1}x_{\gamma_1}}{mp_{\gamma_1}}-\frac{x_{\gamma_1}}{m}$ for all $j\in[m]$. We can easily check that $\{a_j\}_{j=1}^m$ are independent with $\mathbb{E}[a_j]=0$, so that we can leverage Theorem 3 to continue our following analysis. We see that

$$
\max_{j\in[m]}|a_j|=\max\{\frac{|x_{\gamma_1}|}{m}(\frac{1}{p_{\gamma_1}}-1),\frac{|x_{\gamma_1}|}{m}\}\leq\frac{|x_{\gamma_1}|}{mp_{\gamma_1}}, \tag{49}
$$

and

$$\sum_{j=1}^{m}\mathbb{E}\left[a_j^2\right]=\frac{x_{\gamma_1}^2}{mp_{\gamma_1}}-\frac{x_{\gamma_1}^2}{m}. \tag{50}$$

Thus, applying Theorem 3 with $R=\frac{|x_{\gamma_1}|}{mp_{\gamma_1}}$ and $\sigma^2=\frac{x_{\gamma_1}^2}{mp_{\gamma_1}}-\frac{x_{\gamma_1}^2}{m}$ obtains that

$$\mathbb{P}(|\sum_{j=1}^{m}a_j|\geq\epsilon)\leq 2\exp(\frac{-\epsilon^2/2}{x_{\gamma_1}^2/(mp_{\gamma_1})-x_{\gamma_1}^2/m+|x_{\gamma_1}|\epsilon/(3mp_{\gamma_1})}), \tag{51}$$

whose RHS is denoted by η_{γ_1}. Then, with probability at least $1-\eta_{\gamma_1}$ we have $|\sum_{j=1}^{m}a_j|\leq\epsilon$, i.e., $|\sum_{j=1}^{m}\frac{\delta_{t_j\gamma_1}x_{\gamma_1}}{mp_1}|\leq|x_{\gamma_1}|+\epsilon$. We then replace ϵ by other variables to obtain that

$$|x_{\gamma_1}|+\epsilon=|x_{\gamma_1}|+\log(\frac{2}{\eta_{\gamma_1}})\left[\frac{|x_{\gamma_1}|}{3mp_{\gamma_1}}+|x_{\gamma_1}|\sqrt{\frac{1}{9m^2p_{\gamma_1}^2}+\frac{2}{\log(2/\eta_{\gamma_1})}(\frac{1}{mp_{\gamma_1}}-\frac{1}{m})}\right], \tag{52}$$

which is denoted by $f(x_{\gamma_1},\eta_{\gamma_1},m)$.

In a similar way, we can bound $|\sum_{j=1}^{m}\frac{\delta_{t_jk}x_k}{mp_k}|$ for any other $k\in[d]$. The lemma then follows by using the union bound over cases for all $k\in[d]$. □

2.4. Proof of Theorem 1

Proof. We have to prove that the unbiased estimator for original covariance matrix $\mathbf{C}$ is Eq. (1), i.e., $\mathbf{C}_e=\widehat{\mathbf{C}}_1-\widehat{\mathbf{C}}_2$, where $\widehat{\mathbf{C}}_1=\frac{m}{mn-n}\sum_{i=1}^{n}\mathbf{S}_i\mathbf{S}_i^T\mathbf{x}_i\mathbf{x}_i^T\mathbf{S}_i\mathbf{S}_i^T$, and $\widehat{\mathbf{C}}_2=\frac{m}{mn-n}\sum_{i=1}^{n}\mathbb{D}(\mathbf{S}_i\mathbf{S}_i^T\mathbf{x}_i\mathbf{x}_i^T\mathbf{S}_i\mathbf{S}_i^T)\mathbb{D}(\mathbf{b}_i)$ with $b_{ki}=\frac{1}{1+(m-1)p_{ki}}$. Note that each $\mathbf{S}_i$ is created by running Algorithm 1, and $\{\mathbf{S}_i\}_{i=1}^{n}$ are independent matrices. Thus, taking all summands $\mathbb{E}\left[\mathbf{S}_i\mathbf{S}_i^T\mathbf{x}_i\mathbf{x}_i^T\mathbf{S}_i\mathbf{S}_i^T\right]$ together and leveraging Eq. (7) in Lemma 1 achieves the expectation of $\widehat{\mathbf{C}}_1$,

$$\begin{aligned}\mathbb{E}[\widehat{\mathbf{C}}_1]&=\frac{m}{nm-n}\mathbb{E}\sum_{i=1}^{n}\mathbf{S}_i\mathbf{S}_i^T\mathbf{x}_i\mathbf{x}_i^T\mathbf{S}_i\mathbf{S}_i^T=\frac{m}{nm-n}\sum_{i=1}^{n}\left[\sum_{k=1}^{d}\frac{x_{ki}^2}{mp_{ki}}\mathbf{e}_k\mathbf{e}_k^T+\frac{m-1}{m}\mathbf{x}_i\mathbf{x}_i^T\right]\\&=\frac{1}{nm-n}\sum_{i=1}^{n}\sum_{k=1}^{d}\frac{x_{ki}^2}{p_{ki}}\mathbf{e}_k\mathbf{e}_k^T+\frac{1}{n}\mathbf{X}\mathbf{X}^T.\end{aligned} \tag{53}$$

Eq. (53) indicates that $\widehat{\mathbf{C}}_1$ is a biased estimator for the original covariance matrix $\mathbf{C}=\frac{1}{n}\mathbf{X}\mathbf{X}^T=\frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i\mathbf{x}_i^T$. We still need to apply a debiasing procedure to $\widehat{\mathbf{C}}_1$ to get an unbiased estimator. By Eq. (8) in Lemma 1, it can be shown that

$$\mathbb{E}[\widehat{\mathbf{C}}_2]=\frac{m}{nm-n}\sum_{i=1}^{n}\mathbb{E}\left[\mathbb{D}(\mathbf{S}_i\mathbf{S}_i^T\mathbf{x}_i\mathbf{x}_i^T\mathbf{S}_i\mathbf{S}_i^T)\right]\mathbb{D}(\mathbf{b}_i)=\frac{1}{nm-n}\sum_{i=1}^{n}\sum_{k=1}^{d}\frac{x_{ki}^2}{p_{ki}}\mathbf{e}_k\mathbf{e}_k^T. \tag{54}$$

Combing Eq. (53) with Eq. (54), we immediately see that $\mathbf{C}_e=\widehat{\mathbf{C}}_1-\widehat{\mathbf{C}}_2$ is unbiased for $\mathbf{C}$. □

2.5. Proof of Theorem 2

Proof. Here, we have to bound the error $\|\mathbf{C}_e-\mathbf{C}\|_2$. To make the representation compact, we define $\mathbf{A}_i=\mathbf{A}_{i_1}-\mathbf{A}_{i_2}-\mathbf{A}_{i_3}$ with $\mathbf{A}_{i_1}=\frac{m\mathbf{S}_i\mathbf{S}_i^T\mathbf{x}_i\mathbf{x}_i^T\mathbf{S}_i\mathbf{S}_i^T}{nm-n}$, $\mathbf{A}_{i_2}=\frac{m\mathbb{D}(\mathbf{S}_i\mathbf{S}_i^T\mathbf{x}_i\mathbf{x}_i^T\mathbf{S}_i\mathbf{S}_i^T)\mathbb{D}(\mathbf{b}_i)}{nm-n}$, $\mathbf{A}_{i_3}=\frac{\mathbf{x}_i\mathbf{x}_i^T}{n}$. Then, $\sum_{i=1}^{n}\mathbf{A}_i=\mathbf{C}_e-\mathbf{C}$ holds. It is straightforward to see that $\{\mathbf{A}_i\}_{i=1}^{n}$ are independent zero-mean random matrices, which are exactly the setting of the Matrix Bernstein inequality, as shown in Theorem 3. To bound $\|\mathbf{C}_e-\mathbf{C}\|_2$ via Theorem 3, we need to calculate the relevant parameters R and σ^2 that characterize the range and variance of $\mathbf{A}_i$ respectively.

We first derive R by bounding $\|\mathbf{A}_i\|_2$ so that $\|\mathbf{A}_i\|_2\leq R$ for all $i\in[n]$. Expanding $\|\mathbf{A}_i\|_2$ gets that

$$\|\mathbf{A}_i\|_2=\|\mathbf{A}_{i_1}-\mathbf{A}_{i_2}-\mathbf{A}_{i_3}\|_2\leq\|\mathbf{A}_{i_1}-\mathbf{A}_{i_2}\|_2+\|\mathbf{A}_{i_3}\|_2$$

$$\leq \|\mathbf{A}_{i_1}\|_2 + \|\mathbf{A}_{i_3}\|_2. \tag{55}$$

The last inequality in Eq. (55) results from

$$\begin{aligned}
\|\mathbf{A}_{i_1} - \mathbf{A}_{i_2}\|_2 &= \max_{k\in[d]} |\lambda_k(\mathbf{A}_{i_1} - \mathbf{A}_{i_2})| \\
&\overset{(a)}{\leq} \max\{|\lambda_d(\mathbf{A}_{i_1}) - \lambda_1(\mathbf{A}_{i_2})|, |\lambda_1(\mathbf{A}_{i_1}) - \lambda_d(\mathbf{A}_{i_2})|\} && (56)\\
&\overset{(b)}{=} \max\{\lambda_1(\mathbf{A}_{i_2}), |\lambda_1(\mathbf{A}_{i_1}) - \lambda_d(\mathbf{A}_{i_2})|\} && (57)\\
&\overset{(c)}{=} \max\{\lambda_1(\mathbf{A}_{i_2}), \lambda_1(\mathbf{A}_{i_1}) - \lambda_d(\mathbf{A}_{i_2})\} && (58)\\
&\overset{(d)}{\leq} \lambda_1(\mathbf{A}_{i_1}) && (59)\\
&\overset{(e)}{=} \|\mathbf{A}_{i_1}\|_2, && (60)
\end{aligned}$$

where $\lambda_k(\cdot)$ is the k-th largest eigenvalue.

(a) follows from that $\lambda_k(\mathbf{A}_{i_1}) - \lambda_1(\mathbf{A}_{i_2}) \leq \lambda_k(\mathbf{A}_{i_1} - \mathbf{A}_{i_2}) \leq \lambda_k(\mathbf{A}_{i_1}) - \lambda_d(\mathbf{A}_{i_2})$ for any $k \in [d]$, which can be proved by combining Theorem 4 with the fact that $\lambda_d(-\mathbf{A}_{i_2}) = -\lambda_1(\mathbf{A}_{i_2})$ and $\lambda_1(-\mathbf{A}_{i_2}) = -\lambda_d(\mathbf{A}_{i_2})$ for $\mathbf{A}_{i_2} \in \mathbb{R}^{d\times d}$.

(b) holds because of that $\lambda_{k\geq 2}(\mathbf{A}_{i_1}) = 0$ since $\mathbf{A}_{i_1}$ is a positive semidefinite matrix of rank 1, and $\lambda_{k\in[d]}(\mathbf{A}_{i_2}) \geq 0$ since $\mathbf{A}_{i_2}$ is positive semidefinite.

(c) follows owing to that $\lambda_1(\mathbf{A}_{i_1}) = \mathrm{Tr}(\mathbf{A}_{i_1}) \geq \mathrm{Tr}(\mathbf{A}_{i_2}) = \sum_{k=1}^d \lambda_k(\mathbf{A}_{i_2}) \geq \lambda_d(\mathbf{A}_{i_2}) \geq 0$, where the first equality holds because $\lambda_{k\geq 2}(\mathbf{A}_{i_1}) = 0$, the first inequality results from the fact that the diagonal matrix $\mathbf{A}_{i_2}$ is constructed by the diagonal elements of $\mathbf{A}_{i_1}$ multiplied by positive scalars not bigger than 1, and the second inequality is the consequence of $\lambda_{k\in[d]}(\mathbf{A}_{i_2}) \geq 0$.

(d) results from that $\lambda_{k\in[d]}(\mathbf{A}_{i_2}) \geq 0$.

(e) follows owing to that $\mathbf{A}_{i_1}$ is positive semidefinite.

Now, we only need to bound $\|\mathbf{A}_{i_1}\|_2$ and $\|\mathbf{A}_{i_3}\|_2$. We have that

$$\|\mathbf{A}_{i_3}\|_2 = \|\frac{\mathbf{x}_i\mathbf{x}_i^T}{n}\|_2 = \frac{\|\mathbf{x}_i\|_2^2}{n}. \tag{61}$$

Then, Lemma 2 reveals that with probability at least $1 - \sum_{k=1}^d \eta_{ki}$,

$$\|\mathbf{A}_{i_1}\|_2 \leq \frac{m}{nm-n} \sum_{k\in\Gamma_i} f^2(x_{ki}, \eta_{ki}, m), \tag{62}$$

where $\Gamma_i = \{\gamma_{ti}\}_{t=1}^{|\Gamma_i|}$ is a set occupying at most m different elements of $[d]$ with its cardinality $|\Gamma_i| \leq m$, and $f(x_{ki}, \eta_{ki}, m) = |x_{ki}| + \log(\frac{2}{\eta_{ki}}) \left[\frac{|x_{ki}|}{3mp_{ki}} + |x_{ki}|\sqrt{\frac{1}{9m^2p_{ki}^2} + \frac{2}{\log(2/\eta_{ki})}(\frac{1}{mp_{ki}} - \frac{1}{m})}\right]$.

We derive the similar results for all $\{\mathbf{x}_i\}_{i=1}^n$. Then, by union bound, with probability at least $1 - \sum_{i=1}^n \sum_{k=1}^d \eta_{ki}$, we have

$$R = \max_{i\in[n]} \left[\frac{m}{nm-n} \sum_{k\in\Gamma_i} f^2(x_{ki}, \eta_{ki}, m) + \frac{\|\mathbf{x}_i\|_2^2}{n}\right]. \tag{63}$$

Applying the well known inequality $(\sum_{t=1}^n a_t)^2 \leq n\sum_{t=1}^n a_t^2$, we have

$$\begin{aligned}
f^2(x_{ki}, \eta_{ki}, m) &\leq 3x_{ki}^2 + 3\log^2(\frac{2}{\eta_{ki}})\frac{x_{ki}^2}{9m^2p_{ki}^2} + 3\log^2(\frac{2}{\eta_{ki}})\frac{x_{ki}^2}{9m^2p_{ki}^2} + 6\log(\frac{2}{\eta_{ki}})(\frac{x_{ki}^2}{mp_{ki}} - \frac{x_{ki}^2}{m}) \\
&\leq 3x_{ki}^2 + \log^2(\frac{2}{\eta_{ki}})\frac{2x_{ki}^2}{3m^2p_{ki}^2} + \log(\frac{2}{\eta_{ki}})\frac{6x_{ki}^2}{mp_{ki}}. && (64)
\end{aligned}$$

Before continuing characterizing R in Eq. (63), we set the sampling probabilities as $p_{ki} = \alpha\frac{|x_{ki}|}{\|\mathbf{x}_i\|_1} + (1-\alpha)\frac{x_{ki}^2}{\|\mathbf{x}_i\|_2^2}$. It is easy to check that $\sum_{k=1}^d p_{ki} = 1$. For $0<\alpha<1$, we also have $p_{ki} \geq \alpha\frac{|x_{ki}|}{\|\mathbf{x}_i\|_1}$, then plugging it in the second and third term of Eq. (64) respectively getting that

$$f^2(x_{ki},\eta_{ki},m) \leq 3x_{ki}^2 + \log^2(\frac{2}{\eta_{ki}})\frac{2\|\mathbf{x}_i\|_1^2}{3m^2\alpha^2} + \log(\frac{2}{\eta_{ki}})\frac{6|\mathbf{x}_{ki}|\|\mathbf{x}_i\|_1}{m\alpha}. \tag{65}$$

Equipped with Eq. (63) and setting $\eta_{ki} = \frac{\eta}{nd}$ for all $i\in[n]$ and $k\in[d]$, we bound R with probability at least $1-\sum_{i=1}^n\sum_{k=1}^d \eta_{ki} = 1-\eta$ by

$$\begin{aligned} R &\leq \max_{i\in[n]}\left[\frac{m}{nm-n}\sum_{k\in\Gamma_i}\Big(3x_{ki}^2 + \log^2(\frac{2nd}{\eta})\frac{2\|\mathbf{x}_i\|_1^2}{3m^2\alpha^2} + \log(\frac{2nd}{\eta})\frac{6|\mathbf{x}_{ki}|\|\mathbf{x}_i\|_1}{m\alpha}\Big) + \frac{\|\mathbf{x}_i\|_2^2}{n}\right] \\ &\leq \max_{i\in[n]}\left[\frac{2}{n}\Big(3\|\mathbf{x}_i\|_2^2 + \log^2(\frac{2nd}{\eta})\frac{2\|\mathbf{x}_i\|_1^2}{3m\alpha^2} + \log(\frac{2nd}{\eta})\frac{6\|\mathbf{x}_i\|_1^2}{m\alpha}\Big) + \frac{\|\mathbf{x}_i\|_2^2}{n}\right] \\ &\leq \max_{i\in[n]}\left[\frac{7\|\mathbf{x}_i\|_2^2}{n} + \log^2(\frac{2nd}{\eta})\frac{14\|\mathbf{x}_i\|_1^2}{nm\alpha^2}\right], \end{aligned} \tag{66}$$

where the second inequality follows from that $\frac{m}{m-1}\leq 2$ for $m\geq 2$ and $|\Gamma_i|\leq m$, and the last inequality results from that $\alpha\leq 1$ and $\log(\frac{2nd}{\eta})\geq 1$ for $n\geq 1$, $d\geq 2$, and $\eta\leq 1$.

At this stage, we have to derive σ^2 by only bounding for $\|\sum_{i=1}^n \mathbb{E}[\mathbf{A}_i\mathbf{A}_i]\|_2$ since $\mathbf{A}_i$ is symmetric. Expanding $\mathbb{E}[\mathbf{A}_i\mathbf{A}_i]$ obtains that

$$\begin{aligned} \mathbf{0}\preceq\mathbb{E}[\mathbf{A}_i\mathbf{A}_i] = \mathbb{E}[&\mathbf{A}_{i_1}\mathbf{A}_{i_1} + \mathbf{A}_{i_2}\mathbf{A}_{i_2} + \mathbf{A}_{i_3}\mathbf{A}_{i_3} - \mathbf{A}_{i_1}\mathbf{A}_{i_2} - \mathbf{A}_{i_2}\mathbf{A}_{i_1} \\ &-\mathbf{A}_{i_1}\mathbf{A}_{i_3} - \mathbf{A}_{i_3}\mathbf{A}_{i_1} + \mathbf{A}_{i_2}\mathbf{A}_{i_3} + \mathbf{A}_{i_3}\mathbf{A}_{i_2}], \end{aligned}$$

in RHS of which, we bound the expectation of each term. Specifically, invoking Lemma 1, we have that

$$\begin{aligned} n^2\mathbb{E}[\mathbf{A}_i\mathbf{A}_i] = &\underbrace{\sum_{k=1}^d\left[\frac{4}{m(m-1)p_{ki}^2} + \frac{1}{(m-1)^2mp_{ki}^3}\right]x_{ki}^4\mathbf{e}_k\mathbf{e}_k^T}_{①} \\ &+\underbrace{\sum_{k=1}^d\left[\frac{\|\mathbf{x}_i\|_2^2(m-2)}{m(m-1)} + \frac{1}{m(m-1)}\sum_{k=1}^d\frac{x_{ki}^2}{p_{ki}}\right]\frac{x_{ki}^2}{p_{ki}}\mathbf{e}_k\mathbf{e}_k^T}_{②} + \underbrace{\left[\frac{\|\mathbf{x}_i\|_2^2(m^2-5m+6)}{m(m-1)} + \frac{m-2}{m(m-1)}\sum_{k=1}^d\frac{x_{ki}^2}{p_{ki}}\right]\mathbf{x}_i\mathbf{x}_i^T}_{③} \\ &+\underbrace{\frac{2(m-2)}{m(m-1)}\mathbf{x}_i\mathbf{x}_i^T\mathbb{D}(\{\frac{x_{ki}^2}{p_{ki}}\})}_{④} + \underbrace{\frac{1}{m(m-1)}\mathbf{x}_i\mathbf{x}_i^T\mathbb{D}(\{\frac{x_{ki}^2}{p_{ki}^2}\})}_{⑤} + \underbrace{\frac{2(m-2)}{m(m-1)}\mathbb{D}(\{\frac{x_{ki}^2}{p_{ki}}\})\mathbf{x}_i\mathbf{x}_i^T}_{⑥} + \underbrace{\frac{1}{m(m-1)}\mathbb{D}(\{\frac{x_{ki}^2}{p_{ki}^2}\})\mathbf{x}_i\mathbf{x}_i^T}_{⑦} \\ &+\underbrace{\mathbb{D}(\mathbf{b}_i)\mathbb{D}(\mathbf{b}_i)\sum_{k=1}^d\left[\frac{1}{m(m-1)^2p_{ki}^3} + \frac{7}{m(m-1)p_{ki}^2} + \frac{6(m-2)}{m(m-1)p_{ki}} + \frac{(m-2)(m-3)}{m(m-1)}\right]x_{ki}^4\mathbf{e}_k\mathbf{e}_k^T}_{⑧} \\ &+\underbrace{\|\mathbf{x}_i\|_2^2\mathbf{x}_i\mathbf{x}_i^T}_{⑨} + \underbrace{\sum_{k=1}^d(\frac{1}{(m-1)p_{ki}}+1)x_{ki}^2\mathbf{e}_k\mathbf{e}_k^T\mathbb{D}(\mathbf{b}_i)\mathbf{x}_i\mathbf{x}_i^T}_{⑩} + \underbrace{\mathbf{x}_i\mathbf{x}_i^T\sum_{k=1}^d(\frac{1}{(m-1)p_{ki}}+1)x_{ki}^2\mathbf{e}_k\mathbf{e}_k^T\mathbb{D}(\mathbf{b}_i)}_{⑪} \\ &\underbrace{-2\sum_{k=1}^d\left[\frac{1}{m(m-1)^2p_{ki}^3} + \frac{6}{m(m-1)p_{ki}^2} + \frac{3(m-2)}{m(m-1)p_{ki}}\right]x_{ki}^4\mathbf{e}_k\mathbf{e}_k^T\mathbb{D}(\mathbf{b}_i)}_{⑫} - \underbrace{\frac{3(m-2)}{m(m-1)}\mathbf{x}_i\mathbf{x}_i^T\mathbb{D}(\{\frac{x_{ki}^2}{p_{ki}}\})\mathbb{D}(\mathbf{b}_i)}_{⑬} \end{aligned}$$

$$
\begin{aligned}
&\underbrace{-\frac{(m-2)(m-3)}{m(m-1)}\mathbf{x}_i\mathbf{x}_i^T\mathbb{D}(\{x_{ki}^2\})\mathbb{D}(\mathbf{b}_i)}_{⑭}\underbrace{-\frac{3(m-2)}{m(m-1)}\mathbb{D}(\mathbf{b}_i)\mathbb{D}(\{\frac{x_{ki}^2}{p_{ki}}\})\mathbf{x}_i\mathbf{x}_i^T}_{⑮}\underbrace{-\frac{(m-2)(m-3)}{m(m-1)}\mathbb{D}(\mathbf{b}_i)\mathbb{D}(\{x_{ki}^2\})\mathbf{x}_i\mathbf{x}_i^T}_{⑯} \\
&\underbrace{-\sum_{k=1}^d \frac{x_{ki}^2}{(m-1)p_{ki}}\mathbf{e}_k\mathbf{e}_k^T\mathbf{x}_i\mathbf{x}_i^T}_{⑰}\underbrace{-\|\mathbf{x}_i\|_2^2\mathbf{x}_i\mathbf{x}_i^T}_{⑱}\underbrace{-\sum_{k=1}^d \frac{x_{ki}^2}{(m-1)p_{ki}}\mathbf{x}_i\mathbf{x}_i^T\mathbf{e}_k\mathbf{e}_k^T}_{⑲}\underbrace{-\|\mathbf{x}_i\|_2^2\mathbf{x}_i\mathbf{x}_i^T}_{⑳}\underbrace{-\frac{1}{m(m-1)}\mathbf{x}_i\mathbf{x}_i^T\mathbb{D}(\{\frac{x_{ki}^2}{p_{ki}^2}\})\mathbb{D}(\mathbf{b}_i)}_{㉑} \\
&\underbrace{-\frac{1}{m(m-1)}\mathbb{D}(\mathbf{b}_i)\mathbb{D}(\{\frac{x_{ki}^2}{p_{ki}^2}\})\mathbf{x}_i\mathbf{x}_i^T}_{㉒}. \qquad (67)
\end{aligned}
$$

Because of the limited space, $\mathbb{D}(\{\frac{x_{ki}^2}{p_{ki}}\})$ is to denote a square diagonal matrix in $\mathbb{R}^{d\times d}$ with $\{\frac{x_{ki}^2}{p_{ki}}\}_{k=1}^d$ on its diagonal, which is also extended to other similar notations.

In Eq. (67), it can be checked that for $m \geq 2$, we have

$$
\begin{aligned}
⑩ - ⑰ &= \mathbf{0}; \\
⑪ - ⑲ &= \mathbf{0}; \\
④ - ⑬ + ⑤ - ⑭ - ㉑ &= \frac{\mathbf{x}_i\mathbf{x}_i^T}{m(m-1)}\mathbb{D}(\{\frac{((m-1)/p_{ki})x_{ki}^2}{1+(m-1)p_{ki}} + \frac{(m-2)(m+1-1/p_{ki})x_{ki}^2}{1+(m-1)p_{ki}}\}); \\
⑥ - ⑮ + ⑦ - ⑯ - ㉒ &= \mathbb{D}(\{\frac{((m-1)/p_{ki})x_{ki}^2}{1+(m-1)p_{ki}} + \frac{(m-2)(m+1-1/p_{ki})x_{ki}^2}{1+(m-1)p_{ki}}\})\frac{\mathbf{x}_i\mathbf{x}_i^T}{m(m-1)}; \\
③ + ⑨ - ⑱ - ⑳ &= \left[\frac{(6-4m)\|\mathbf{x}_i\|_2^2}{m(m-1)} + \frac{m-2}{m(m-1)}\sum_{k=1}^d \frac{x_{ki}^2}{p_{ki}}\right]\mathbf{x}_i\mathbf{x}_i^T \\
&\preceq \frac{1}{m}\sum_{k=1}^d \frac{x_{ki}^2}{p_{ki}}\mathbf{x}_i\mathbf{x}_i^T; \\
⑧ - ⑫ &\preceq \mathbf{0}; \\
① &\preceq \sum_{k=1}^d \left[\frac{8x_{ki}^4}{m^2p_{ki}^2} + \frac{4x_{ki}^4}{m^3p_{ki}^3}\right]\mathbf{e}_k\mathbf{e}_k^T; \\
② &\preceq \sum_{k=1}^d \left[\frac{\|\mathbf{x}_i\|_2^2x_{ki}^2}{mp_{ki}} + \frac{2x_{ki}^2}{m^2p_{ki}}\sum_{k=1}^d \frac{x_{ki}^2}{p_{ki}}\right]\mathbf{e}_k\mathbf{e}_k^T. \qquad (68)
\end{aligned}
$$

Then, applying Eq. (67) and Eq. (68) obtains that

$$
\begin{aligned}
\mathbf{0} \preceq \mathbb{E}\left[\mathbf{A}_i\mathbf{A}_i\right] \preceq & \frac{1}{n^2}\sum_{k=1}^d \left[\frac{8x_{ki}^4}{m^2p_{ki}^2} + \frac{4x_{ki}^4}{m^3p_{ki}^3} + \frac{\|\mathbf{x}_i\|_2^2x_{ki}^2}{mp_{ki}} + \frac{2x_{ki}^2}{m^2p_{ki}}\sum_{k=1}^d \frac{x_{ki}^2}{p_{ki}}\right]\mathbf{e}_k\mathbf{e}_k^T \\
&+ \frac{\mathbf{x}_i\mathbf{x}_i^T}{n^2m(m-1)}\mathbb{D}(\{\frac{((m-1)/p_{ki})x_{ki}^2}{1+(m-1)p_{ki}} + \frac{(m-2)(m+1-1/p_{ki})x_{ki}^2}{1+(m-1)p_{ki}}\}) \\
&+ \mathbb{D}(\{\frac{((m-1)/p_{ki})x_{ki}^2}{1+(m-1)p_{ki}} + \frac{(m-2)(m+1-1/p_{ki})x_{ki}^2}{1+(m-1)p_{ki}}\})\frac{\mathbf{x}_i\mathbf{x}_i^T}{n^2m(m-1)} \\
&+ \frac{1}{n^2m}\sum_{k=1}^d \frac{x_{ki}^2}{p_{ki}}\mathbf{x}_i\mathbf{x}_i^T. \qquad (69)
\end{aligned}
$$

With Eq. (69) in hand, we can formulate σ^2 as

$$
\sigma^2 = \|\sum_{i=1}^n \mathbb{E}\left[\mathbf{A}_i\mathbf{A}_i\right]\|_2 \leq \sum_{i=1}^n \max_{k\in[d]} \frac{1}{n^2}\left[\frac{8x_{ki}^4}{m^2p_{ki}^2} + \frac{4x_{ki}^4}{m^3p_{ki}^3} + \frac{\|\mathbf{x}_i\|_2^2x_{ki}^2}{mp_{ki}} + \frac{2x_{ki}^2}{m^2p_{ki}}\sum_{k=1}^d \frac{x_{ki}^2}{p_{ki}}\right]
$$

$$
\begin{aligned}
&+\sum_{i=1}^{n}\max_{k\in[d]}\frac{1}{n^2}\left[\frac{2\|\mathbf{x}_i\|_2^2}{m(m-1)}(\frac{((m-1)/p_{ki})x_{ki}^2}{1+(m-1)p_{ki}}+\frac{(m-2)(m+1+1/p_{ki})x_{ki}^2}{1+(m-1)p_{ki}})\right]+\frac{1}{n^2m}\|\sum_{i=1}^{n}\sum_{k=1}^{d}\frac{x_{ki}^2}{p_{ki}}\mathbf{x}_i\mathbf{x}_i^T\|_2\\
&\le\sum_{i=1}^{n}\max_{k\in[d]}\frac{1}{n^2}\left[\frac{8x_{ki}^4}{m^2p_{ki}^2}+\frac{4x_{ki}^4}{m^3p_{ki}^3}+\frac{\|\mathbf{x}_i\|_2^2x_{ki}^2}{mp_{ki}}+\frac{2x_{ki}^2}{m^2p_{ki}}\sum_{k=1}^{d}\frac{x_{ki}^2}{p_{ki}}\right]\\
&+\sum_{i=1}^{n}\max_{k\in[d]}\frac{1}{n^2}\left[\frac{8\|\mathbf{x}_i\|_2^2x_{ki}^2}{mp_{ki}}\right]+\frac{1}{n^2m}\|\sum_{i=1}^{n}\sum_{k=1}^{d}\frac{x_{ki}^2}{p_{ki}}\mathbf{x}_i\mathbf{x}_i^T\|_2.
\end{aligned}
\tag{70}
$$

Again, we have to consider the sampling distributions $p_{ki}=\alpha\frac{|x_{ki}|}{\|\mathbf{x}_i\|_1}+(1-\alpha)\frac{x_{ki}^2}{\|\mathbf{x}_i\|_2^2}$ with $0<\alpha<1$. Plugging $p_{ki}\ge\alpha\frac{|x_{ki}|}{\|\mathbf{x}_i\|_1}$ and $p_{ki}\ge(1-\alpha)\frac{x_{ki}^2}{\|\mathbf{x}_i\|_2^2}$ in Eq. (70), we have

$$
\begin{aligned}
\sigma^2&\le\sum_{i=1}^{n}\max_{k\in[d]}\frac{1}{n^2}\left[\frac{8\|\mathbf{x}_i\|_2^4}{m^2(1-\alpha)^2}+\frac{4\|\mathbf{x}_i\|_1^2\|\mathbf{x}_i\|_2^2}{m^3\alpha^2(1-\alpha)}+\frac{\|\mathbf{x}_i\|_2^4}{m(1-\alpha)}+\frac{2\|\mathbf{x}_i\|_2^2}{m^2(1-\alpha)}\sum_{k=1}^{d}\frac{|x_{ki}|\|\mathbf{x}_i\|_1}{\alpha}\right]\\
&+\sum_{i=1}^{n}\max_{k\in[d]}\frac{1}{n^2}\left[\frac{8\|\mathbf{x}_i\|_2^4}{m(1-\alpha)}\right]+\frac{1}{n^2m}\|\sum_{i=1}^{n}\sum_{k=1}^{d}\frac{|x_{ki}|\|\mathbf{x}_i\|_1}{\alpha}\mathbf{x}_i\mathbf{x}_i^T\|_2\\
&=\sum_{i=1}^{n}\left[\frac{8\|\mathbf{x}_i\|_2^4}{n^2m^2(1-\alpha)^2}+\frac{4\|\mathbf{x}_i\|_1^2\|\mathbf{x}_i\|_2^2}{n^2m^3\alpha^2(1-\alpha)}+\frac{9\|\mathbf{x}_i\|_2^4}{n^2m(1-\alpha)}+\frac{2\|\mathbf{x}_i\|_2^2\|\mathbf{x}_i\|_1^2}{n^2m^2\alpha(1-\alpha)}\right]\\
&+\|\sum_{i=1}^{n}\frac{\|\mathbf{x}_i\|_1^2\mathbf{x}_i\mathbf{x}_i^2}{n^2m\alpha}\|_2.
\end{aligned}
\tag{71}
$$

Note that employing $p_{ki}=\Omega(\frac{|x_{ki}|^{4/3}}{\sum_{k=1}^{d}|x_{ki}|^{4/3}})$ for the term $\frac{4x_{ki}^4}{m^3p_{ki}^3}$ in Eq. (70) can produce a result tighter than that in Eq. (71), which is because of the fact that $(\sum_{k=1}^{d}|x_{ki}|^{4/3})^3\le\|\mathbf{x}_i\|_1^2\|\mathbf{x}_i\|_2^2$ always holds owing to the Holder's inequality. However, it is not necessary to apply $p_{ki}=\Omega(\frac{|x_{ki}|^{4/3}}{\sum_{k=1}^{d}|x_{ki}|^{4/3}})$ to the term $\frac{4x_{ki}^4}{m^3p_{ki}^3}$ in Eq. (70), because the term $\frac{4\|\mathbf{x}_i\|_1^2\|\mathbf{x}_i\|_2^2}{n^2m^3\alpha^2(1-\alpha)}=O(\frac{\|\mathbf{x}_i\|_1^2\|\mathbf{x}_i\|_2^2}{n^2m^3})$ in Eq. (71) obtained by applying $p_{ki}=\alpha\frac{|x_{ki}|}{\|\mathbf{x}_i\|_1}+(1-\alpha)\frac{x_{ki}^2}{\|\mathbf{x}_i\|_2^2}=\Omega(\frac{|x_{ki}|}{\|\mathbf{x}_i\|_1}+\frac{x_{ki}^2}{\|\mathbf{x}_i\|_2^2})$ to the term $\frac{4x_{ki}^4}{m^3p_{ki}^3}$ in Eq. (70) has already been small enough, which can be smaller than other terms in Eq. (71) like $\frac{2\|\mathbf{x}_i\|_2^2\|\mathbf{x}_i\|_1^2}{n^2m^2\alpha(1-\alpha)}=O(\frac{\|\mathbf{x}_i\|_1^2\|\mathbf{x}_i\|_2^2}{n^2m^2})$. Similarly, applying other sampling probabilities $p_{ki}=\Omega(\frac{|x_{ki}|^q}{\sum_{k=1}^{d}|x_{ki}|^q})$ with $q\neq1,\frac{4}{3},2$ to Eq. (70) will produce a result larger than Eq. (71), which may not be bounded. This is also why we only use $p_{ki}=\alpha\frac{|x_{ki}|}{\|\mathbf{x}_i\|_1}+(1-\alpha)\frac{x_{ki}^2}{\|\mathbf{x}_i\|_2^2}=\Omega(\frac{|x_{ki}|}{\|\mathbf{x}_i\|_1})$ to tighten R in Eq. (66). This derivation justifies our selection of $q=1,2$ in $p_{ki}=\Omega(\frac{|x_{ki}|^q}{\sum_{k=1}^{d}|x_{ki}|^q})$ used for constructing the sampling probability $p_{ki}=\alpha\frac{|x_{ki}|}{\|\mathbf{x}_i\|_1}+(1-\alpha)\frac{x_{ki}^2}{\|\mathbf{x}_i\|_2^2}$.

We then invoke Theorem 3 to obtain that for $\epsilon\ge0$,

$$
\mathbb{P}(\|\mathbf{C}_e-\mathbf{C}\|_2\ge\epsilon)\le2d\exp(\frac{-\epsilon^2/2}{\sigma^2+R\epsilon/3}).
\tag{72}
$$

Denote the RHS of Eq. (72) by $\delta=2d\exp(\frac{-\epsilon^2/2}{\sigma^2+R\epsilon/3})$ and consider the failure probability η in Eq. (66), then by union bound we have $\|\mathbf{C}_e-\mathbf{C}\|_2\le\epsilon$ holds with probability at least $1-\eta-\delta$. Furthermore, $\delta=2d\exp(\frac{-\epsilon^2/2}{\sigma^2+R\epsilon/3})$ yields the following quadratic equation in ϵ

$$
\frac{\epsilon^2}{2\log(2d/\delta)}-\frac{R\epsilon}{3}-\sigma^2=0.
\tag{73}
$$

Solving Eq. (73) gets only one positive root

$$
\epsilon=\log(\frac{2d}{\delta})\left[\frac{R}{3}+\sqrt{(\frac{R}{3})^2+\frac{2\sigma^2}{\log(2d/\delta)}}\right]
$$

$$\leq \log(\frac{2d}{\delta})\frac{2R}{3} + \sqrt{2\sigma^2 \log(\frac{2d}{\delta})}. \tag{74}$$

Thus, immediately we have $\|\mathbf{C}_e - \mathbf{C}\|_2 \leq \log(\frac{2d}{\delta})\frac{2R}{3} + \sqrt{2\sigma^2 \log(\frac{2d}{\delta})}$ holds with probability at least $1-\eta-\delta$, which completes the whole proof. □

2.6. Proof of Corollary 1

Proof. According to the setting, substituting that $\|\mathbf{x}_i\|_2 \leq \tau$ for all $i \in [n]$, $\frac{\|\mathbf{x}_i\|_1}{\|\mathbf{x}_i\|_2} \leq \varphi$ with $1 \leq \varphi \leq \sqrt{d}$, and $m < d$ into Theorem 2 establishes that

$$\begin{aligned}\|\mathbf{C}_e - \mathbf{C}\|_2 &\leq \widetilde{O}\Big(\frac{\tau^2}{n} + \frac{\tau^2\varphi^2}{nm} + \sqrt{\frac{\tau^4}{nm^2} + \frac{\tau^4\varphi^2}{nm^3} + \frac{\tau^4}{nm} + \frac{\tau^4\varphi^2}{nm^2} + \frac{\|\mathbf{C}\|_2\tau^2\varphi^2}{nm}}\Big) \\ &\leq \widetilde{O}\Big(\frac{\tau^2}{n} + \frac{\tau^2\varphi^2}{nm} + \frac{\tau^2\varphi}{m}\sqrt{\frac{1}{n}} + \tau^2\sqrt{\frac{1}{nm}} + \tau\varphi\sqrt{\frac{\|\mathbf{C}\|_2}{nm}}\Big),\end{aligned} \tag{75}$$

where the first inequality invokes $\sum_{i=n}^n \|\mathbf{x}_i\|_2^4 \leq n\tau^4$, and $\mathbf{C} = \sum_{i=1}^n \frac{\mathbf{x}_i\mathbf{x}_i^T}{n}$ is the original covariance matrix.

Also, we can adopt $\sum_{i=1}^n \|\mathbf{x}_i\|_2^4 \leq nd\tau^2\|\mathbf{C}\|_2$, which holds because $\sum_{i=1}^n \|\mathbf{x}_i\|_2^4 \leq \tau^2 \sum_{i=1}^n \|\mathbf{x}_i\|_2^2$ and $\sum_{i=1}^n \|\mathbf{x}_i\|_2^2 = n\mathrm{Tr}(\mathbf{C}) \leq nd\|\mathbf{C}\|_2$. □

Hence, we have

$$\begin{aligned}\|\mathbf{C}_e - \mathbf{C}\|_2 &\leq \widetilde{O}\Big(\frac{\tau^2}{n} + \frac{\tau^2\varphi^2}{nm} + \tau\sqrt{\|\mathbf{C}\|_2}\sqrt{\frac{d}{nm^2} + \frac{d\varphi^2}{nm^3} + \frac{d}{nm} + \frac{d\varphi^2}{nm^2} + \frac{\varphi^2}{nm}}\Big) \\ &\leq \widetilde{O}\Big(\frac{\tau^2}{n} + \frac{\tau^2\varphi^2}{nm} + \frac{\tau\varphi}{m}\sqrt{\frac{d\|\mathbf{C}\|_2}{n}} + \tau\sqrt{\frac{d\|\mathbf{C}\|_2}{nm}} + \tau\varphi\sqrt{\frac{\|\mathbf{C}\|_2}{nm}}\Big).\end{aligned} \tag{76}$$

Finally, assigning $\|\mathbf{C}_e - \mathbf{C}\|_2$ by the smaller one of Eq. (75) and Eq. (76) completes the proof.

2.7. Proof of Corollaries 2 and 3

Proof. The proof follows (Azizyan et al., 2015, Corollaries 4-6), where the key component $\|\mathbf{C}_e - \mathbf{C}_p\|_2$ is upper bounded by $\|\mathbf{C}_e - \frac{1}{n}\sum_{i=1}^n \mathbf{x}_i\mathbf{x}_i^T\|_2 + \|\frac{1}{n}\sum_{i=1}^n \mathbf{x}_i\mathbf{x}_i^T - \mathbf{C}_p\|_2$. Then, the derivation results from Theorem 2 in our paper and the Gaussian tail bounds in (Azizyan et al., 2015, Proposition 14).

(Azizyan et al., 2015, Proposition 14) shows that with probability at least $1-\zeta$ for $d \geq 2$,

$$\begin{aligned}\max_{i\in[n]} \|\mathbf{x}_i\|_2 &\leq \sqrt{2\mathrm{Tr}(\mathbf{C}_p)\log(nd/\zeta)}; \\ \|\frac{1}{n}\sum_{i=1}^n \mathbf{x}_i\mathbf{x}_i^T - \mathbf{C}_p\|_2 &\leq O\big(\|\mathbf{C}_p\|_2\sqrt{\log(2/\zeta)/n}\big).\end{aligned} \tag{77}$$

Then, applying them and Corollary 1 along with the fact that $\|\mathbf{x}_i\|_1 \leq \sqrt{d}\|\mathbf{x}_i\|_2$ and $\mathrm{Tr}(\mathbf{C}_p) \leq d\|\mathbf{C}_p\|_2$ establishes

$$\begin{aligned}&\|\mathbf{C}_e - \mathbf{C}_p\|_2 \leq \|\mathbf{C}_e - \frac{1}{n}\sum_{i=1}^n \mathbf{x}_i\mathbf{x}_i^T\|_2 + \|\frac{1}{n}\sum_{i=1}^n \mathbf{x}_i\mathbf{x}_i^T - \mathbf{C}_p\|_2 \\ &\leq \widetilde{O}\Big(\frac{\tau^2}{n} + \frac{\tau^2\varphi^2}{nm} + \frac{\tau^2\varphi}{m}\sqrt{\frac{1}{n}} + \tau^2\sqrt{\frac{1}{nm}} + \tau\varphi\sqrt{\frac{\|\frac{1}{n}\sum_{i=1}^n \mathbf{x}_i\mathbf{x}_i^T\|_2}{nm}}\Big) + \widetilde{O}\Big(\|\mathbf{C}_p\|_2\sqrt{\frac{1}{n}}\Big) \\ &\leq \widetilde{O}\Big(\frac{\tau^2}{n} + \frac{\tau^2\varphi^2}{nm} + \frac{\tau^2\varphi}{m}\sqrt{\frac{1}{n}} + \tau^2\sqrt{\frac{1}{nm}} + \tau\varphi\sqrt{\frac{\|\mathbf{C}_p\|_2}{nm}}\Big) + \widetilde{O}\Big(\|\mathbf{C}_p\|_2\sqrt{\frac{1}{n}}\Big) \\ &\leq \widetilde{O}\Big(\frac{d^2\|\mathbf{C}_p\|_2}{nm} + \frac{d\|\mathbf{C}_p\|_2}{m}\sqrt{\frac{d}{n}} + d\|\mathbf{C}_p\|_2\sqrt{\frac{1}{nm}} + d\|\mathbf{C}_p\|_2\sqrt{\frac{1}{nm}} + \|\mathbf{C}_p\|_2\sqrt{\frac{1}{n}}\Big)\end{aligned} \tag{78}$$

$$\leq \widetilde{O}\Big(\frac{d^2\|\mathbf{C}_p\|_2}{nm}+\frac{d\|\mathbf{C}_p\|_2}{m}\sqrt{\frac{d}{n}}\Big) \tag{79}$$

with probability at least $1-\eta-\delta-\zeta$, where Eq. (78) results from that we invoke Eq. (77) to get $\|\frac{1}{n}\sum_{i=1}^n \mathbf{x}_i\mathbf{x}_i^T\|_2 \leq \|\frac{1}{n}\sum_{i=1}^n \mathbf{x}_i\mathbf{x}_i^T-\mathbf{C}_p\|_2+\|\mathbf{C}_p\|_2 \leq \widetilde{O}(\|\mathbf{C}_p\|_2)$.

The proof for the low-rank case where rank($\mathbf{C}_p$)$\leq r$ additionally adopts

$$\begin{aligned}\|[\mathbf{C}_e]_r-\mathbf{C}_p\|_2 &\leq \|[\mathbf{C}_e]_r-\mathbf{C}_e\|_2+\|\mathbf{C}_e-\mathbf{C}_p\|_2\\ &\leq \|[\mathbf{C}_p]_r-\mathbf{C}_e\|_2+\|\mathbf{C}_e-\mathbf{C}_p\|_2\\ &\leq \|[\mathbf{C}_p]_r-\mathbf{C}_p\|_2+\|\mathbf{C}_p-\mathbf{C}_e\|_2+\|\mathbf{C}_e-\mathbf{C}_p\|_2\\ &= 2\|\mathbf{C}_e-\mathbf{C}_p\|_2,\end{aligned} \tag{80}$$

where the last equality holds because rank$(\mathbf{C}_p)\leq r$. Then, armed with $\mathrm{Tr}(\mathbf{C}_p)\leq \mathrm{rank}(\mathbf{C}_p)\|\mathbf{C}_p\|_2 \leq r\|\mathbf{C}_p\|_2$, we have

$$\begin{aligned}&\|[\mathbf{C}_e]_r-\mathbf{C}_p\|_2 \leq O(\|\mathbf{C}_e-\mathbf{C}_p\|_2) \leq O(\|\mathbf{C}_e-\frac{1}{n}\sum_{i=1}^n \mathbf{x}_i\mathbf{x}_i^T\|_2+\|\frac{1}{n}\sum_{i=1}^n \mathbf{x}_i\mathbf{x}_i^T-\mathbf{C}_p\|_2)\\ &\leq \widetilde{O}\Big(\frac{rd\|\mathbf{C}_p\|_2}{nm}+\frac{r\|\mathbf{C}_p\|_2}{m}\sqrt{\frac{d}{n}}+r\|\mathbf{C}_p\|_2\sqrt{\frac{1}{nm}}+\|\mathbf{C}_p\|_2\sqrt{\frac{rd}{nm}}+\|\mathbf{C}_p\|_2\sqrt{\frac{1}{n}}\Big)\\ &\leq \widetilde{O}\Big(\frac{rd\|\mathbf{C}_p\|_2}{nm}+\frac{r\|\mathbf{C}_p\|_2}{m}\sqrt{\frac{d}{n}}+\|\mathbf{C}_p\|_2\sqrt{\frac{rd}{nm}}\Big)\end{aligned} \tag{81}$$

with probability at least $1-\eta-\delta-\zeta$.

The given definitions also implicitly indicate that $\mathbf{C}_p$ and $\mathbf{C}_e$ are symmetric. Then, following (Azizyan et al., 2015), the desired bound in Corollary 3 immediately results from Corollary 2 combined with the Davis-Kahan Theorem (Davis & Kahan, 1970) that shows $\|\widehat{\prod}_k-\prod_k\|_2 \leq \frac{1}{\lambda_k-\lambda_{k+1}}\|\mathbf{C}_e-\mathbf{C}_p\|_2$. □

3. Discussion for Counterparts

3.1. Theorems for Gauss-Inverse and UniSample-HD

We first use our notations to rephrase current theoretical results provided in (Azizyan et al., 2015, Theorem 3) and (Anaraki & Becker, 2017, Theorem 6), which correspond to *Gauss-Inverse* and *UniSample-HD*, respectively.

Theorem 5 (Azizyan et al. 2015, Theorem 3)**.** *Let $d\geq 2$ and define,*

$$S_1=\|\frac{1}{n}\sum_{i=1}^n\|\mathbf{x}_i\|_2^2\mathbf{x}_i\mathbf{x}_i^T\|_2,\, S_2=\frac{1}{n}\sum_{i=1}^n\|\mathbf{x}_i\|_2^4.$$

There exists universal constants $\kappa_1,\,\kappa_2>0$ such that for any $0<\delta<1$, with probability at least $1-\delta$,

$$\|\mathbf{C}_e-\mathbf{C}\|_2 \leq \kappa_1\Big(\sqrt{\frac{d}{m}S_1}+\sqrt{\frac{d}{m^2}S_2}\Big)\sqrt{\frac{\log(d/\delta)}{n}}+\kappa_2\frac{d\max_{i\in[n]}\|\mathbf{x}_i\|_2^2}{nm}\log(d/\delta). \tag{82}$$

Theorem 6 (Anaraki & Becker 2017, Theorem 6)**.** *Let each column of $\mathbf{S}_i\in\mathbb{R}^{d\times m}$ be chosen uniformly at random from the set of all canonical basis vectors without replacement. Let $\rho>0$ be a bound such that $\|\mathbf{S}_i\mathbf{S}_i^T\mathbf{x}_i\|_2^2\leq\rho\|\mathbf{x}_i\|_2^2$ for all $i\in[n]$. Then, with probability at least $1-\delta$*

$$\|\mathbf{C}_e-\mathbf{C}\|_2\leq\epsilon, \tag{83}$$

where $\delta = d\exp\Big(\frac{-\epsilon^2/2}{\sigma^2+R\epsilon/3}\Big)$, $R=\frac{1}{n}\Big[\Big(\frac{d(d-1)}{m(m-1)}\rho+1\Big)\max_{i\in[n]}\|\mathbf{x}_i\|_2^2+\frac{d(d-m)}{m(m-1)}\max_{k\in[d],i\in[n]}x_{ki}^2\Big]$, *and* $\sigma^2 = \frac{d(d-1)}{nm(m-1)}\Big[(\rho-\frac{m(m-1)}{d(d-1)})\max_{i\in[n]}\|\mathbf{x}_i\|_2^2\|\mathbf{C}\|_2+\frac{d-m}{m-1}\rho\max_{i\in[n]}\|\mathbf{x}_i\|_2^2\|\mathbb{D}(\mathbf{C})\|_2$
$+\frac{2(d-m)\|\mathbf{X}\|_F^2}{n(m-1)}\max_{k\in[d],i\in[n]}x_{ki}^2+\frac{(d-m)^2\max_{k\in[d]}\sum_{i=1}^n x_{ki}^4}{n(d-1)(m-1)}\Big]$.

3.2. Discussion

In this subsection, we will simplify the foregoing two theorems by making Eq. (82) and Eq. (83) explicitly dependent on n, m and d. Our derivations are natural and straightforward, and we will not deliberately loose Eq. (82) and Eq. (83) in order to demonstrate the superiority of the theoretical results gained by our weighted sampling method. We define that $\max_{i\in n}\|\mathbf{x}_i\|_2 \leq \tau$.

In terms of Eq. (82) in Theorem 5, $S_1 \leq \max_{i\in[n]}\|\mathbf{x}_i\|_2^2\|\mathbf{C}\|_2$ and $S_2 \leq \max_{i\in[n]}\|\mathbf{x}_i\|_2^4$. Note that $\frac{1}{nd}\|\mathbf{X}\|_F^2 \leq \|\mathbf{C}\|_2 \leq \max_{i\in[n]}\|\mathbf{x}_i\|_2^2$. Then, Eq. (82) can be simplified and reformulated as

$$\begin{aligned}\|\mathbf{C}_e - \mathbf{C}\|_2 &\leq \widetilde{O}\Big(\sqrt{\frac{d\|\mathbf{C}\|_2\max_{i\in[n]}\|\mathbf{x}_i\|_2^2}{nm}} + \sqrt{\frac{d\max_{i\in[n]}\|\mathbf{x}_i\|_2^4}{nm^2}} + \frac{d\max_{i\in[n]}\|\mathbf{x}_i\|_2^2}{nm}\Big)\\ &\leq \widetilde{O}\Big(\tau\sqrt{\frac{d\|\mathbf{C}\|_2}{nm}} + \frac{\tau^2}{m}\sqrt{\frac{d}{n}} + \frac{\tau^2 d}{nm}\Big).\end{aligned} \tag{84}$$

If applying $S_2 \leq d\max_{i\in[n]}\|\mathbf{x}_i\|_2^2\|\mathbf{C}\|_2$ in the original paper (Azizyan et al., 2015), we will get that

$$\begin{aligned}\|\mathbf{C}_e - \mathbf{C}\|_2 &\leq \widetilde{O}\Big(\sqrt{\frac{d\|\mathbf{C}\|_2\max_{i\in[n]}\|\mathbf{x}_i\|_2^2}{nm}} + \sqrt{\frac{d^2\max_{i\in[n]}\|\mathbf{x}_i\|_2^2\|\mathbf{C}\|_2}{nm^2}} + \frac{d\max_{i\in[n]}\|\mathbf{x}_i\|_2^2}{nm}\Big)\\ &\leq \widetilde{O}\Big(\frac{\tau d}{m}\sqrt{\frac{\|\mathbf{C}\|_2}{n}} + \frac{\tau^2 d}{nm}\Big).\end{aligned} \tag{85}$$

In summary,

$$\|\mathbf{C}_e - \mathbf{C}\|_2 \leq \min\{\widetilde{O}\Big(\tau\sqrt{\frac{d\|\mathbf{C}\|_2}{nm}} + \frac{\tau^2}{m}\sqrt{\frac{d}{n}} + \frac{\tau^2 d}{nm}\Big), \widetilde{O}\Big(\frac{\tau d}{m}\sqrt{\frac{\|\mathbf{C}\|_2}{n}} + \frac{\tau^2 d}{nm}\Big)\}. \tag{86}$$

For Eq. (83) in Theorem 6, we first simplify its R and σ^2. According to (Anaraki & Becker, 2017), to obtain a more accurate estimation, each $\mathbf{x}_i$ is required to be multiplied by $\mathbf{HD}$ to flatten its large entries before being sampled uniformly without replacement, where $\mathbf{H}$ is a Hadamard matrix with its dimension being 2^l (l is a certain positive integer), and $\mathbf{D}$ is a diagonal matrix with its diagonal elements being i.i.d. Rademacher random variables. Note that $\mathbf{HDD}^T\mathbf{H}^T = \mathbf{D}^T\mathbf{H}^T\mathbf{HD}$ is an identity matrix.

Suppose that we do not have to pad $\mathbf{X}$ with zeros until its dimension $d = 2^l$ holds. Hence, assuming that $d = 2^l$ for $\mathbf{X} \in \mathbb{R}^{d\times n}$ without loss of generality, we define $\mathbf{Y} = \mathbf{HDX} \in \mathbb{R}^{d\times n}$ below.

Corollary 2 of (Anaraki & Becker, 2017) indicates that with probability at least $1-\beta$, we have

$$\max_{k\in[d],i\in[n]}|y_{ki}| \leq \sqrt{\frac{1}{d}}\sqrt{2\log(\frac{2nd}{\beta})}\max_{i\in[n]}\|\mathbf{x}_i\|_2 \tag{87}$$

and

$$\max_{i\in[n]}\|\mathbf{y}_i\|_2 \leq \sqrt{2\log(\frac{2nd}{\beta})}\max_{i\in[n]}\|\mathbf{x}_i\|_2. \tag{88}$$

Corollary 3 of (Anaraki & Becker, 2017) indicates that with probability at least $1-\beta$, we have

$$\|\mathbf{S}_i\mathbf{S}_i^T\mathbf{y}_i\|_2 \leq \sqrt{\frac{m}{d}}\sqrt{2\log(\frac{2nd}{\beta})}\|\mathbf{x}_i\|_2. \tag{89}$$

To make a compact representation, we define $\theta = \sqrt{2\log(\frac{2nd}{\beta})}$. Obviously, $\theta > 1$.

Then, in Theorem 6, we can replace the input data $\mathbf{X}$ by $\mathbf{Y}$. Combing Eq. (89) with the fact that $\|\mathbf{y}_i\|_2 = \|\mathbf{HDx}_i\|_2 = \|\mathbf{x}_i\|_2$ getting $\rho = ((\sqrt{\frac{m}{d}}\theta)^2) = \frac{m\theta^2}{d}$ for the setting of Theorem 6. Along with $\theta > 1$ and $m \leq d$, we have

$$
\begin{aligned}
R &= \frac{1}{n}O\Big((\frac{d^2}{m^2}\frac{m\theta^2}{d}+1)\theta^2\max_{i\in[n]}\|\mathbf{x}_i\|_2^2+\frac{d(d-m)}{m^2}(\sqrt{\frac{1}{d}}\theta)^2\max_{i\in[n]}\|\mathbf{x}_i\|_2^2\Big)\\
&= O\Big((\frac{d\theta^2}{nm})\theta^2\max_{i\in[n]}\|\mathbf{x}_i\|_2^2\Big)\\
&= \widetilde{O}\Big(\frac{d}{nm}\max_{i\in[n]}\|\mathbf{x}_i\|_2^2\Big)\\
&= \widetilde{O}\Big(\frac{\tau^2 d}{nm}\Big),
\end{aligned}
\tag{90}
$$

and

$$
\begin{aligned}
\sigma^2 \le{}& \frac{d^2}{nm^2}O\Big((\frac{m\theta^2}{d}-\frac{m(m-1)}{d(d-1)})\theta^2\max_{i\in[n]}\|\mathbf{x}_i\|_2^2\|\frac{\mathbf{HDXX}^T\mathbf{D}^T\mathbf{H}^T}{n}\|_2 \qquad (91)\\
&+\frac{(d-m)}{m}\frac{m\theta^2}{d}\theta^2\max_{i\in[n]}\|\mathbf{x}_i\|_2^2\|\mathbb{D}(\frac{\mathbf{HDXX}^T\mathbf{D}^T\mathbf{H}^T}{n})\|_2\\
&+\frac{d-m}{nm}\frac{\theta^2}{d}\max_{i\in[n]}\|\mathbf{x}_i\|_2^2\|\mathbf{HDX}\|_F^2+\frac{(d-m)^2}{ndm}n\frac{\theta^4}{d^2}\max_{i\in[n]}\|\mathbf{x}_i\|_2^4\Big)\\
={}& \frac{d^2}{nm^2}O\Big(\frac{m}{d}(\theta^2-\frac{m-1}{d-1})\theta^2\max_{i\in[n]}\|\mathbf{x}_i\|_2^2\|\mathbf{C}\|_2+\frac{(d-m)\theta^4}{d}\max_{i\in[n]}\|\mathbf{x}_i\|_2^2\frac{n(\sqrt{1/d}\theta)^2\max_{i\in[n]}\|\mathbf{x}_i\|_2^2}{n}\\
&+\frac{(d-m)\theta^2}{nmd}\max_{i\in[n]}\|\mathbf{x}_i\|_2^2 nd\frac{\theta^2}{d}\max_{i\in[n]}\|\mathbf{x}_i\|_2^2+\frac{(d-m)^2\theta^4}{d^3m}\max_{i\in[n]}\|\mathbf{x}_i\|_2^4\Big)\\
={}& \widetilde{O}\Big(\frac{d}{nm}\max_{i\in[n]}\|\mathbf{x}_i\|_2^2\|\mathbf{C}\|_2+\frac{d-m}{nm^2}\max_{i\in[n]}\|\mathbf{x}_i\|_2^4\\
&+\frac{d(d-m)}{nm^3}\max_{i\in[n]}\|\mathbf{x}_i\|_2^4+\frac{(d-m)^2}{nm^3d}\max_{i\in[n]}\|\mathbf{x}_i\|_2^4\Big)\\
={}& \widetilde{O}\Big(\frac{d}{nm}\max_{i\in[n]}\|\mathbf{x}_i\|_2^2\|\mathbf{C}\|_2+\frac{d(d-m)}{nm^3}\max_{i\in[n]}\|\mathbf{x}_i\|_2^4+\frac{(d-m)^2}{nm^3d}\max_{i\in[n]}\|\mathbf{x}_i\|_2^4\Big)\\
={}& \widetilde{O}\Big(\frac{\tau^2 d\|\mathbf{C}\|_2}{nm}+\frac{\tau^4 d(d-m)}{nm^3}+\frac{\tau^4(d-m)^2}{nm^3d}\Big).
\end{aligned}
\tag{92}
$$

Note that Eq. (92) for simplifying σ^2 in Eq. (83) is tighter than the simplification result in the original paper (Anaraki & Becker, 2017) that scales with $\frac{d^2}{nm^2}$. Recalling Eq. (83), and replacing its ϵ by R and σ^2 to get that with probability at least $1-\delta-\beta$, we have

$$
\|\mathbf{C}_e-\mathbf{C}\|_2 \le \widetilde{O}\Big(\tau\sqrt{\frac{d\|\mathbf{C}\|_2}{nm}}+\frac{\tau^2}{m}\sqrt{\frac{d(d-m)}{nm}}+\frac{\tau^2(d-m)}{m}\sqrt{\frac{1}{nmd}}+\frac{\tau^2 d}{nm}\Big).
\tag{93}
$$

If $m=d$, then

$$
\|\mathbf{C}_e-\mathbf{C}\|_2 \le \widetilde{O}\Big(\tau\sqrt{\frac{d\|\mathbf{C}\|_2}{nm}}+\frac{\tau^2 d}{nm}\Big).
\tag{94}
$$

Although pure sampling without replacement makes no estimation error when $m=d$, processing the data by a Hadamard matrix before sampling can result in the error as shown in Eq. (94).

If $m<d$ with m being close to d, then $d-m=O(1)$, and thus we have

$$
\|\mathbf{C}_e-\mathbf{C}\|_2 \le \widetilde{O}\Big(\tau\sqrt{\frac{d\|\mathbf{C}\|_2}{nm}}+\frac{\tau^2}{m}\sqrt{\frac{d}{nm}}+\frac{\tau^2 d}{nm}\Big).
\tag{95}
$$

If $m\ll d$ or there exists a certain constant $\kappa<1$ with $m<\kappa d$, then $O(d-m)=O(d)$. In addition to considering that $\frac{1}{nd}\|\mathbf{X}\|_F^2 \le \|\mathbf{C}\|_2 \le \max_{i\in[n]}\|\mathbf{x}_i\|_2^2=\tau^2$, then we have

$$
\|\mathbf{C}_e-\mathbf{C}\|_2 \le \widetilde{O}\Big(\tau\sqrt{\frac{d\|\mathbf{C}\|_2}{nm}}+\frac{\tau^2 d}{m}\sqrt{\frac{1}{nm}}+\frac{\tau^2 d}{nm}\Big).
\tag{96}
$$

4. Computational Complexity

Recall that we have n data samples in the d-dimensional space, and let m be the target compressed dimension. The computational comparisons between our proposed method and the other approaches are presented in Table 1, in which *Standard* method means computing $\mathbf{C}$ directly without data compression. We should explain some terms in the table before proceeding.

Storage: storing data and random projection matrices (if any) in the remote sites and the fusion center, and storing the covariance matrix in the fusion center.

Communication: shipping the data and random projection matrices (if any) from remote sites to the fusion center (*high communication cost requires tremendous bandwidth and power consumption*).

Time (FLOPS): compressing the data in the remote sites, and calculating the covariance matrix in the fusion center (*a low time complexity means a low power cost and high efficiency for the data processing*).

Note that, instead of only using the fusion center, data have to be first collected from many remote sites like a network of $g \ll n$ sensors. Then, they are transmitted to the fusion center to estimate the covariance matrix. This procedure shows why communication cost is required. In the table, except for the communication, the two other compared terms have contained the total costs in both the remote sites and fusion center.

For a covariance matrix defined as $\mathbf{C} = \frac{1}{n}\mathbf{X}\mathbf{X}^T - \bar{\mathbf{x}}\bar{\mathbf{x}}^T$, we can exactly calculate $\bar{\mathbf{x}} = \frac{1}{n}\sum_{i=1}^n \mathbf{x}_i$ in the fusion center by $\bar{\mathbf{x}} = \frac{1}{n}\sum_{j=1}^g \mathbf{u}_j$, where $\{\mathbf{x}_i\}_{i=1}^n$ are distributed in $g \ll n$ remote sites, and $\mathbf{u}_j \in \mathbb{R}^d$ is the summation of all data vectors in the j-th remote site before being compressed. Hence, about $O(gd)$ storage, $O(gd)$ communication cost, and $O(nd)$ time have to be added to the last four methods in Table 1, with $g \ll n$.

Table 1. Computational costs in terms of storage, communication, and time.

Method	Storage	Communication	Time
Standard	$O(nd+d^2)$	$O(nd)$	$O(nd^2)$
Gauss-Inverse	$O(nm+d^2)$	$O(nm)$	$O(nmd+nm^2d+nd^2)+T_G$
Sparse	$O(nm+d^2)$	$O(nm)$	$O(d+nm^2)+T_S$
UniSample-HD	$O(nm+d^2)$	$O(nm)$	$O(nd\log d+nm^2)$
Our method	$O(nm+d^2)$	$O(nm)$	$O(nd+nm\log d+nm^2)$

From now on, we can focus on the covariance matrix defined as $\mathbf{C} = \frac{1}{n}\mathbf{X}\mathbf{X}^T$.

First, we derive the computational costs in our propose algorithm. Computing $\{p_{ki}\}_{k\in[d],i\in[n]}$ takes $O(nd)$ time. Then, sampling nm entries from all data vectors to get $\mathbf{Y} \in \mathbb{R}^{m\times n}$ takes time that is scaled on $nm\log d$ up to a certain small constant. In Eq. (1), each $\mathbf{S}_i$, $\mathbf{S}_i^T\mathbf{x}_i$, $\mathbf{S}_i\mathbf{S}_i^T\mathbf{x}_i$, and $\mathbf{S}_i\mathbf{S}_i^T$ (squared diagonal), has at most m non-zero entries. Hence, recovering $\{\mathbf{S}_i\}_{i=1}^n$ via the sampled nm entries in $\mathbf{Y}$ and the sampling indices in $\mathbf{T} \in \mathbb{R}^{m\times n}$ incurs $O(nm)$ time. With $\mathbf{Y}$ and $\mathbf{T}$ in hand, $\{\mathbf{S}_i\mathbf{S}_i^T\mathbf{x}_i\}_{i=1}^n$ can be accurately computed in $O(nm)$ time. Equipped with $\{\mathbf{S}_i\mathbf{S}_i^T\mathbf{x}_i\}_{i=1}^n$, computing $\widehat{\mathbf{C}}_1 = \frac{m}{nm-n}\sum_{i=1}^n \mathbf{S}_i\mathbf{S}_i^T\mathbf{x}_i\mathbf{x}_i^T\mathbf{S}_i\mathbf{S}_i^T$ additionally takes only $O(nm^2)$ time, this is due to that each $\mathbf{S}_i\mathbf{S}_i^T\mathbf{x}_i \in \mathbb{R}^d$ and $\mathbf{S}_i\mathbf{S}_i^T\mathbf{x}_i\mathbf{x}_i^T\mathbf{S}_i\mathbf{S}_i^T \in \mathbb{R}^{d\times d}$ has at most m and m^2 non-zero entries respectively. Based on the obtained $\widehat{\mathbf{C}}_1$, computing the square diagonal matrix $\widehat{\mathbf{C}}_2 = \frac{m}{nm-n}\sum_{i=1}^n \mathbb{D}(\mathbf{S}_i\mathbf{S}_i^T\mathbf{x}_i\mathbf{x}_i^T\mathbf{S}_i\mathbf{S}_i^T)\mathbb{D}(\mathbf{b}_i)$ takes $O(nm)$ time since each $\mathbf{S}_i\mathbf{S}_i^T\mathbf{x}_i\mathbf{x}_i^T\mathbf{S}_i\mathbf{S}_i^T$ has at most m non-zero entries in its diagonal. Finally, obtaining $\mathbf{C} = \widehat{\mathbf{C}}_1 - \widehat{\mathbf{C}}_2$ incurs $O(d)$ extra time. The total running time is about $O(nd + nm\log d + nm + nm + nm + nm^2 + nm + d) = O(nd + nm\log d + nm^2)$. In the remote sites, data are compressed into m dimensional space. Computing b_{ki} only corresponding to the sampled entries is enough to exactly calculate the $\widehat{\mathbf{C}}_2 = \frac{m}{nm-n}\sum_{i=1}^n \mathbb{D}(\mathbf{S}_i\mathbf{S}_i^T\mathbf{x}_i\mathbf{x}_i^T\mathbf{S}_i\mathbf{S}_i^T)\mathbb{D}(\mathbf{b}_i)$ in Eq. (1), so that at most nm entries from $\{p_{ki}\}_{k\in[d],i\in[n]}$ have to be retained to obtain $\{b_{ki}\}$, since $b_{ki} = \frac{1}{1+(m-1)p_{ki}}$. Thus, in the remote sites, $\mathbf{Y} \in \mathbb{R}^{m\times n}$ and $\mathbf{T} \in \mathbb{R}^{m\times n}$ dominate the storage cost, taking about $O(nm)$ space in total. In the fusion center, $O(d^2)$ storage is additionally used to store the estimated covariance $\mathbf{C}_e \in \mathbb{R}^{d\times d}$. Similarly, about $O(nm)$ communication cost is required because of transmitting $\mathbf{Y} \in \mathbb{R}^{m\times n}$, $\mathbf{T} \in \mathbb{R}^{m\times n}$, $\mathbf{v} \in \mathbb{R}^n$, $\mathbf{w} \in \mathbb{R}^n$ and α.

Then, for *Standard* in Table 1 that means directly calculating covariance matrix through the observed data samples without compression, it is straightforward to check its computational complexity. $\mathbf{X} \in \mathbb{R}^{d\times n}$ and $\mathbf{C} \in \mathbb{R}^{d\times d}$ takes about $O(nd+d^2)$ storage in total, and $\mathbf{X} \in \mathbb{R}^{d\times n}$ leads to about $O(nd)$ communication burden. Calculating the covariance matrix $\mathbf{C} =$

$\frac{1}{n}\mathbf{X}\mathbf{X}^T$ costs $O(nd^2)$ time.

For *Gauss-Inverse*, $\sum_{i=1}^n \mathbf{S}_i(\mathbf{S}_i^T\mathbf{S}_i)^{-1}\mathbf{S}_i^T\mathbf{x}_i\mathbf{x}_i^T\mathbf{S}_i(\mathbf{S}_i^T\mathbf{S}_i)^{-1}\mathbf{S}_i^T$, which is the main part of its unbiased estimator, dominates the computational cost. Generating n different Gaussian matrices $\{\mathbf{S}_i \in \mathbb{R}^{d\times m}\}_{i=1}^n$ by the pseudorandom number generator like Mersenne twister (Matsumoto & Nishimura, 1998), which is by far the most widely used, takes considerably large amount of time in practice. The time cost can be denoted by T_G. As $\mathbf{S}_i$ is dense, computing $\{\mathbf{S}_i^T\mathbf{x}_i\}_{i=1}^n$ takes $O(nmd)$ time. Calculating $\{(\mathbf{S}_i^T\mathbf{S}_i)^{-1}\}_{i=1}^n$ requires $O(nm^2d+nm^3)$, which involves matrix multiplications and inversions. Subsequently, we repeat the matrix-vector multiplications in $\{\mathbf{S}_i(\mathbf{S}_i^T\mathbf{S}_i)^{-1}\mathbf{S}_i^T\mathbf{x}_i \in \mathbb{R}^d\}_{i=1}^n$ from the left to right, based on which we get the target covariance matrix. Finally, it takes at least $O(nmd+nm^2d+nm^3+nm^2+ndm+nd^2)+T_G = O(nmd+nm^2d+nd^2)+T_G$ time for *Gauss-Inverse*. In the remote sites, we compress data by $\mathbf{S}_i^T\mathbf{x}_i \in \mathbb{R}^m$ before sending them to the fusion center. Along with $O(d^2)$ storage for the derived covariance matrix, about $O(nm+d^2)$ storage space is required in total. Also, sending $\{\mathbf{S}_i^T\mathbf{x}_i \in \mathbb{R}^m\}_{i=1}^n$ requires about a $O(nm)$ computational burden.

Note that we have not listed the synchronization cost of *Gauss-Inverse* in Table 1. In practice, a pseudo-random number generator is applied to the program in both the remote sites and the fusion center to generate/reconstruct n Gaussian random matrices $\{\mathbf{S}_i \in \mathbb{R}^{d\times m}\}_{i=1}^n$, and only n seeds are required to be transmitted from remote sites to the fusion center to recover the Gaussian random matrices. Therefore, only about $O(n)$ storage and communication cost have to be added in Table 1. Also, calculating each $(\mathbf{S}_i^T\mathbf{S}_i)^{-1}$ has to load each $\mathbf{S}_i^T\mathbf{S}_i \in \mathbb{R}^{m\times m}$ into memory, hence at least $O(m^2)$ memory is required.

For *Sparse*, calculating $\sum_{i=1}^n \mathbf{S}_i\mathbf{S}_i^T\mathbf{x}_i\mathbf{x}_i^T\mathbf{S}_i\mathbf{S}_i^T$ and subtracting its rescaled diagonal entries dominate the computational cost (Anaraki, 2016). Generating sparse projection matrices $\{\mathbf{S}_i \in \mathbb{R}^{d\times q}\}_{i=1}^n$ is also expensive (Anaraki & Becker, 2017), whose time cost is denoted by T_S. The entries of each $\mathbf{S}_i$ are distributed on $\{-1,0,1\}$ with probabilities $\{\frac{1}{2s}, 1-\frac{1}{s}, \frac{1}{2s}\}$. Then, each column of $\mathbf{S}_i$ has $\frac{d}{s}$ non-zero entries in expectation. Empirically, we can fix that $q/d = 0.2$ or 0.4 according to (Anaraki & Hughes, 2014; Anaraki, 2016). The number of non-zero entries of $\mathbf{S}_i\mathbf{S}_i^T\mathbf{x}_i \in \mathbb{R}^d$ is at least $d(1-(1-\frac{1}{s})^q)$ in expectation, which ranges from $\frac{dq}{s}(1-\frac{q}{2s})$ to $\frac{dq}{s}$. Define $d(1-(1-\frac{1}{s})^q) = m < d$, thus we can solve s with $q/d = 0.2$ or 0.4 fixed to obtain that $s = O(\frac{d^2}{m})$. Then computing $\{\mathbf{S}_i^T\mathbf{x}_i \in \mathbb{R}^q\}_{i=1}^n$ takes $O(\frac{ndq}{s}) = O(nm)$ time in expectation. Based on it, computing $\{\mathbf{S}_i\mathbf{S}_i^T\mathbf{x}_i \in \mathbb{R}^d\}_{i=1}^n$ additionally costs $O(\frac{ndq}{s}) = O(nm)$ time in expectation. Since each $\mathbf{S}_i\mathbf{S}_i^T\mathbf{x}_i \in \mathbb{R}^d$ contains only m non-zeros entries in expectation, thus obtaining $\sum_{i=1}^n \mathbf{S}_i\mathbf{S}_i^T\mathbf{x}_i\mathbf{x}_i^T\mathbf{S}_i\mathbf{S}_i^T$ and subtracting its rescaled diagonal entries requires $O(nm+nm+nm^2+d)+T_S = O(nm^2+d)+T_S$ time in total. Storing $\{\mathbf{S}_i\mathbf{S}_i^T\mathbf{x}_i \in \mathbb{R}^d\}_{i=1}^n$ and the estimated covariance matrix requires $O(nm+d^2)$ storage in expectation, where a $O(nm)$ cost results from $O(nm)$ non-zero entries in $\{\mathbf{S}_i\mathbf{S}_i^T\mathbf{x}_i \in \mathbb{R}^d\}_{i=1}^n$ along with $O(nm)$ corresponding indices. Similarly, sending $\{\mathbf{S}_i\mathbf{S}_i^T\mathbf{x}_i \in \mathbb{R}^d\}_{i=1}^n$ from remote sites to the fusion center takes at most $O(nm)$ communication cost in expectation.

For *UniSample-HD*, processing data by a Hadamard matrix by $\mathbf{HDX} \in \mathbb{R}^{d\times n}$ requires $O(nd\log d)$ time, where $\mathbf{H} \in \mathbb{R}^{d\times d}$ can be a Hadamard matrix, $\mathbf{D} \in \mathbb{R}^{d\times d}$ is a diagonal matrix with diagonal elements being i.i.d. Rademacher random variables, and we suppose that $d = 2^l$ holds (l is a certain positive integer). Then, sampling m entries uniformly without replacement on each data vector by $\{\mathbf{S}_i^T\mathbf{HD}\mathbf{x}_i \in \mathbb{R}^d\}_{i=1}^n$ takes $O(nm)$ time. Hence, it is straightforward to check that $\sum_{i=1}^n \mathbf{HDS}_i\mathbf{S}_i^T\mathbf{D}^T\mathbf{H}^T\mathbf{x}_i\mathbf{x}_i^T\mathbf{D}^T\mathbf{H}^T\mathbf{S}_i\mathbf{S}_i^T\mathbf{HD} \in \mathbb{R}^{d\times d}$ requires $O(nd\log d+nm+nm^2+d^2\log d) = O(nd\log d+nm^2)$ time in total. $\mathbf{HD} \in \mathbb{R}^{d\times d}$ can be generated on the fly when we process the data. About $O(nm+d^2)$ storage has to be used for the compressed data and estimated covariance matrix. Obviously, about $O(nm)$ communication cost is required.

5. Impact of the Parameter α

5.1. Discussion

To determine if the k-th entry of the data vector $\mathbf{x}_i \in \mathbb{R}^d$ should be retained or not, the sampling probability applied in our method is

$$p_{ki} = \alpha\frac{|x_{ki}|}{\|\mathbf{x}_i\|_1} + (1-\alpha)\frac{x_{ki}^2}{\|\mathbf{x}_i\|_2^2}. \tag{97}$$

Achieving our theoretical bound of Theorem 2 requires $0 < \alpha < 1$. However, The case $\alpha = 1$ and $\alpha = 0$ can also obtain weaker error bounds, which can be straightforwardly derived from Eqs. (64)(65) and Eqs. (70)(71). The following illustration reveals the connection between α and error bounds on data owning different properties.

1. Only using $\alpha = 0$, i.e., ℓ_2-norm based sampling $p_{ki} = \frac{x_{ki}^2}{\|\mathbf{x}_i\|_2^2}$ can yield a *very* weak bound if there exist some very

small entries $|x_{ki}|$ in $\mathbf{x}_i \in \mathbb{R}^d$. E.g., substituting $p_{ki} = \frac{x_{ki}^2}{\|\mathbf{x}_i\|_2^2}$ into the term $\max_{k\in[d]} \frac{x_{ki}^2}{p_{ki}^2}$ of Eq. (64) or Eq. (70) results in $\max_{k\in[d]} \frac{\|\mathbf{x}_i\|_2^4}{x_{ki}^2}$ in the final error bound, which becomes infinite if the positive entry $|x_{ki}|$ gets close to 0;

2. Only using $\alpha = 1$, i.e., ℓ_1-norm based sampling $p_{ki} = \frac{|x_{ki}|}{\|\mathbf{x}_i\|_1}$ yields a *slightly* weak bound if there exist some very large entries $|x_{ki}|$ in $\mathbf{x}_i \in \mathbb{R}^d$. E.g., substituting $p_{ki} = \frac{|x_{ki}|}{\|\mathbf{x}_i\|_1}$ into the term $\max_{k\in[d]} \frac{x_{ki}^4}{p_{ki}^2}$ of Eq. (70) results in $\max_{k\in[d]} x_{ki}^2\|\mathbf{x}_i\|_1^2$ in the final error bound, which is always greater than or equal to $\max_{k\in[d]} \|\mathbf{x}_i\|_2^4 = \|\mathbf{x}_i\|_2^4$ derived by employing $p_{ki} = \frac{x_{ki}^2}{\|\mathbf{x}_i\|_2^2}$ to bound $\max_{k\in[d]} \frac{x_{ki}^4}{p_{ki}^2}$. Specifically, assume $\|\mathbf{x}_i\|_2^4 = 1$ without loss of generality, then it is possible that $\max_{\mathbf{x}_i\subset\mathbb{R}^d, \|\mathbf{x}_i\|_2^4=1} \max_{k\in[d]} x_{ki}^2\|\mathbf{x}_i\|_1^2 = \frac{d+2\sqrt{d}+1}{4} \gg 1$ if when $x_{ji} = \sqrt{\frac{\sqrt{d}+1}{2\sqrt{d}}}$ and $x_{ki,k\neq j} = \sqrt{\frac{1}{2d+2\sqrt{d}}}$ for all $k \in [d]$ with $k \neq j$. Also, $\min_{\mathbf{x}_i\subset\mathbb{R}^d, \|\mathbf{x}_i\|_2^4=1} \max_{k\in[d]} x_{ki}^2\|\mathbf{x}_i\|_1^2 = 1$ if we have $x_{ki} = \sqrt{\frac{1}{d}}$ for all $k \in [d]$ or we have $x_{ji} = 1$ and $x_{ki,k\neq j} = 0$ for all $k \in [d]$ with $k \neq j$. Note $\mathbf{x}_i \subset \mathbb{R}^d$ in the above optimizations means that $\mathbf{x}_i$ is a vector variable in the d-dimensional space, and j is an arbitrary integer in the set $[d]$.

3. Therefore, α balances the performance by ℓ_1-norm based sampling and ℓ_2-norm based sampling. ℓ_2 sampling penalizes small entries more than ℓ_1 sampling, hence ℓ_2 sampling is more likely to select larger entries to decrease error (e.g., case 2). However, different from ℓ_1 sampling, ℓ_2 sampling is unstable and sensitive to small entries, and it can make estimation error incredibly high if extremely small entries are picked (e.g., case 1). Then $0 < \alpha < 1$ is applied to achieve the desired tight bound with $p_{ki} \geq (1-\alpha)\frac{x_{ki}^2}{\|\mathbf{x}_i\|_2^2}$ to tackle the extreme situation in the case 2 that cannot be well handled purely by $p_{ki} \geq \alpha\frac{|x_{ki}|}{\|\mathbf{x}_i\|_1}$. When α turns from 1 to 0, the estimation error is likely to first decrease and then increase.

5.2. Experiments

Accordingly, we create four different synthetic datasets: $\{\mathbf{A}_i\}_{i=1}^4 \in \mathbb{R}^{1000\times 10000}$ (i.e., $d = 1000$ and $n = 10000$). All entries in $\mathbf{A}_1$ and $\mathbf{A}_2$ are i.i.d. generated from the Gaussian distributions $\mathcal{N}(\sqrt{\frac{1}{2d+2\sqrt{d}}}, \frac{1}{1000})$ and $\mathcal{N}(\sqrt{\frac{1}{2d+2\sqrt{d}}}, \frac{1}{100})$, respectively. For $\mathbf{A}_3$, the entries of its one row are i.i.d. generated from $\mathcal{N}(\sqrt{\frac{\sqrt{d}+1}{2\sqrt{d}}}, \frac{1}{100})$, and the other entries follow $\mathcal{N}(\sqrt{\frac{1}{2d+2\sqrt{d}}}, \frac{1}{100})$. For $\mathbf{A}_4$, its generation follows the way of $\mathbf{X}_1$ in the main text of the paper.

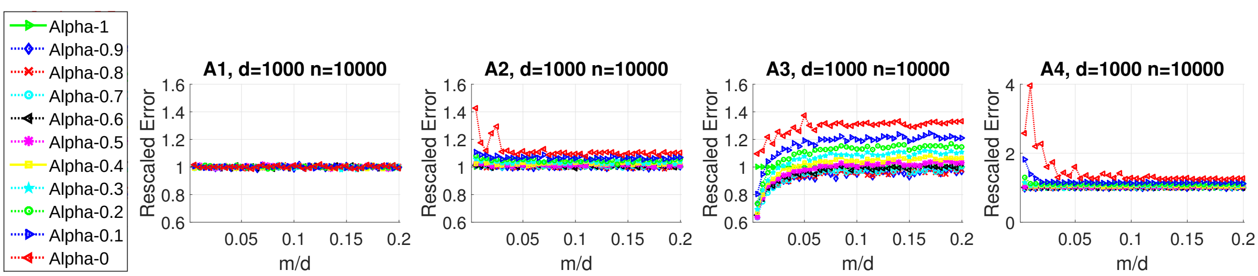

Figure 1. Accuracy comparison by decreasing α from 1 to 0 with a step size of 0.1. The error at each α is normalized by that at $\alpha = 1$ on y-axis, and m/d varies from 0.005 to 0.2 with a step size of 0.005 on x-axis. Roughly, $\alpha = 0.9$ is a good choice, and the smaller parameter like $\alpha = 0$ usually leads to a poorer accuracy and higher variance compared with the other α values.

In Figure 1, the y-axis reports the errors that are normalized by the error incurred at $\alpha = 1$. For $\mathbf{A}_1$, the magnitudes of the data entries tend to be *highly* uniformly distributed. Thus, nearly the same results are returned over all α. For $\mathbf{A}_2$, its entries are *slightly* uniformly distributed with some entries having extremely small magnitudes. Hence, $\alpha = 0$ has a poorer performance compared with the others, which is consistent with the case 1 in Section 5.1. $\mathbf{A}_3$ contains some entries larger than the others, and neither $\alpha = 0$ nor $\alpha = 1$ achieves the best performance obtained roughly at $\alpha = 0.9$. Also, the estimation error first decreases and then increases when α turns from 1 to 0. All such simulation results conform to the case 2 and case 3 in Section 5.1. Considering $\mathbf{A}_4$ that is not likely to contain the extreme situation as mentioned in the case 2 of Section 5.1, we see that best performance is roughly achieved when α gets close to 1.

References

Anaraki, F. Estimation of the sample covariance matrix from compressive measurements. *IET Signal Processing*, 2016.

Anaraki, F. and Becker, S. Preconditioned data sparsification for big data with applications to pca and k-means. *IEEE Transactions on Information Theory*, 2017.

Anaraki, F. and Hughes, S. Memory and computation efficient pca via very sparse random projections. In *Proceedings of the 31st International Conference on Machine Learning (ICML-14)*, pp. 1341–1349, 2014.

Azizyan, M., Krishnamurthy, A., and Singh, A. Extreme compressive sampling for covariance estimation. *arXiv preprint arXiv:1506.00898*, 2015.

Davis, C. and Kahan, W. M. The rotation of eigenvectors by a perturbation. iii. *SIAM Journal on Numerical Analysis*, 7(1):1–46, 1970.

Golub, G. H. and Van Loan, C. F. Matrix computations. 1996.

Matsumoto, M. and Nishimura, T. Mersenne twister: a 623-dimensionally equidistributed uniform pseudo-random number generator. *ACM Transactions on Modeling and Computer Simulation*, 1998.

Tropp, J. A. An introduction to matrix concentration inequalities. *Foundations and Trends in Machine Learning*, 8(1-2):1–230, 2015.