Supplemental Materials: On the Privacy-preserving Generalized Eigenvalue Problem

1. Symbols

Symbol	Meaning
n, d, s	Sample size; dimension of feature vectors; sparsity
A, B	Between-class covariance matrix; within-class covariance matrix
$\lambda_i(A), \lambda_{\max}(A), \lambda_{\min}(A)$	i-th, maximum, minimum eigenvalue of A
$\sigma_i(A), \sigma_{\max}(A), \sigma_{\min}(A)$	i-th, maximum, minimum singular value of A
$\|\cdot\|_2$	Spectral 2-norm of a matrix
$\|\cdot\|_F$	Frobenius norm of a matrix
Λ_A	Eigenvalue matrix of A
Φ_A	Eigenvector matrix of A
C_1, C_2	Constants
ϵ, δ	Differential privacy parameters
η_A, η_B	Step sizes
m	Number of iterations
ξ	Regularization term for whitening to ensure the matrix is full rank
v, v^*	Generalized eigenvector; optimal eigenvector
v_t	output eigenvector after t iterations
e_t	the error between v^* and v_t, which is $v^* - v_t$
ζ, ζ_s	Gaussian noise vector; s-sparse Gaussian noise vector after truncation

Table 1: List of Symbols

2. Experiments

In this section, we present additional experiments to demonstrate that our method is better than the existing method. With step sizes set to 0.5 and 0.25 (as shown in Figures 1 and 2), SR and RF represent the non-DP versions of DPSR (ours) and DPRF, respectively. The results confirm that DPRF performs better without DP, as it uses an optimal vector as the initial vector; however, this advantage disappears when DP is applied. In the RT-IoT2022 dataset, our method outperforms the non-DP algorithm.

3. Matrix Sensitivity

Lemma 1 *(Hu et al. 2023) The sensitivity of the between-class covariance matrix is* $\frac{8}{n}$.

Lemma 2 *(Hu et al. 2023) The sensitivity of the within-class covariance matrix is* $\frac{6}{n}$.

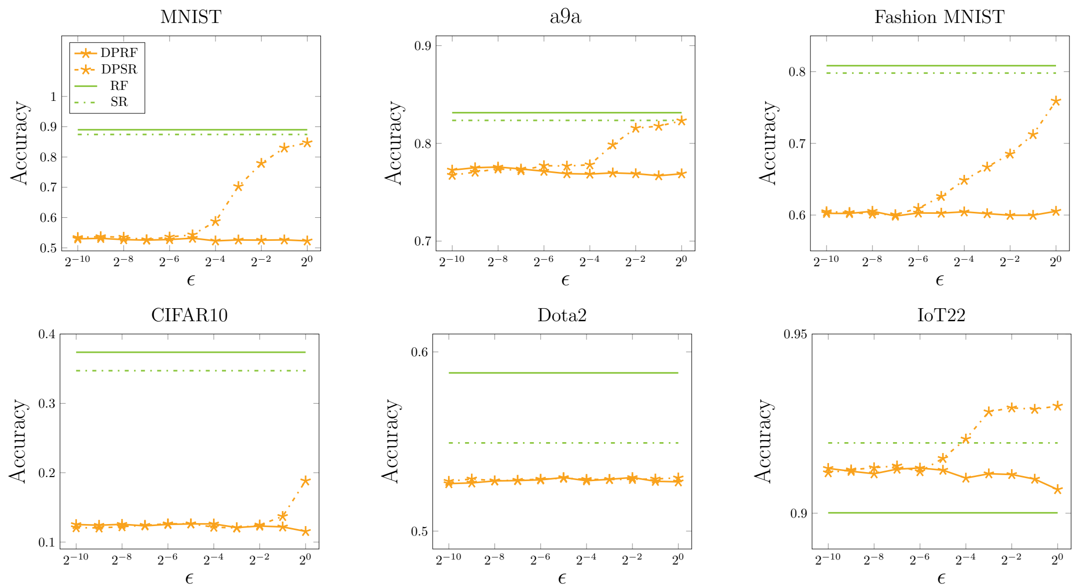

Figure 1: Compare performance between DPSR (ours) and DPRF with step size 0.5.

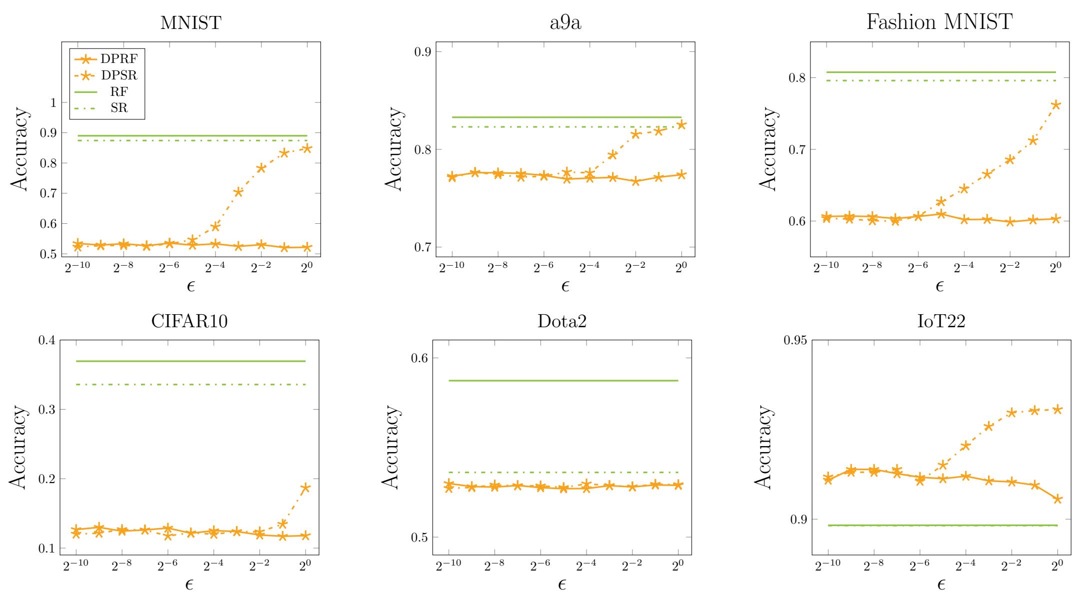

Figure 2: Compare performance between DPSR (ours) and DPRF with step size 0.25.

4. Proof of Theorem 11

Although GEP is applied in various methods, we will focus on FDA. This is because, in other dimension reduction methods, the matrices $\mathbf{A}$ and $\mathbf{B}$ corresponding to GEP are related to the covariance matrix. Therefore, the proof in the context of FDA can be generalized to other methods.

Theorem 11 *(restate) The sensitivities of ∇J_1 and ∇J_2 are bounded by $\frac{C_1}{n}$ and $\frac{C_2}{n\xi}$, respectively, where C_1, C_2 are constants, n is the sample size, and $\xi > 0$ is the regularization ratio, where $\nabla J_1 = 2\mathbf{B}\mathbf{v}$, $\nabla J_2 = 2\tilde{\mathbf{A}}\mathbf{v}$ and $\mathbf{v}$ is a unit vector.*

Proof Given two neighboring samples $\mathbf{X}$ and $\mathbf{X}'$, we assume that the GEP problem we are currently addressing is FDA. For the sensitivity of ∇J_1, we denote that $\mathbf{B}$ and $\mathbf{B}'$ are within-class covariance matrices of $\mathbf{X}$ and $\mathbf{X}'$, respectively. Then the sensitivity of ∇J_1 is

$$\|2\mathbf{B}\mathbf{v} - 2\mathbf{B}'\mathbf{v}\|_2 \leq 2 \cdot \|\mathbf{B} - \mathbf{B}'\|_2 \cdot \|\mathbf{v}\|_2 \leq 2 \cdot \|\mathbf{B} - \mathbf{B}'\|_2 \overset{(1)}{\leq} \frac{12}{n}.$$

By Lemma 2, inequality (1) is true.

For the sensitivity of ∇J_2, we first fix $\mathbf{B}$ and denote that $\mathbf{A}$ and $\mathbf{A}'$ are between-class covariance matrices of $\mathbf{X}$ and $\mathbf{X}'$, respectively. Then the sensitivity of ∇J_2 is

$$\begin{aligned}
\|2\tilde{\mathbf{A}}\mathbf{v} - 2\tilde{\mathbf{A}}'\mathbf{v}\|_2 &\leq 2 \cdot \|\tilde{\mathbf{A}} - \tilde{\mathbf{A}}'\|_2 \cdot \|\mathbf{v}\|_2 \overset{(2)}{\leq} 2 \cdot \|(\tilde{\mathbf{\Phi}}_{\mathbf{B}})^\top (\mathbf{A} - \mathbf{A}')\tilde{\mathbf{\Phi}}_{\mathbf{B}}\|_2 \\
&\overset{(3)}{\leq} 2 \cdot \|((\mathbf{\Lambda}_{\mathbf{B}} + \xi\mathbf{I})^{-\frac{1}{2}})^\top \mathbf{\Phi}_{\mathbf{B}}^\top\|_2 \cdot \|\mathbf{A} - \mathbf{A}'\|_2 \cdot \|\mathbf{\Phi}_{\mathbf{B}}(\mathbf{\Lambda}_{\mathbf{B}} + \xi\mathbf{I})^{-\frac{1}{2}}\|_2 \\
&\overset{(4)}{\leq} \frac{16}{n} \cdot \|(\mathbf{\Lambda}_{\mathbf{B}} + \xi\mathbf{I})^{-\frac{1}{2}}\|_2^2 = \frac{16}{n} \cdot \sigma^2_{\max}((\mathbf{\Lambda}_{\mathbf{B}} + \xi\mathbf{I})^{-\frac{1}{2}}) \\
&\overset{(5)}{=} \frac{1}{\sigma_{\min}((\mathbf{\Lambda}_{\mathbf{B}} + \xi\mathbf{I}))} = \frac{16}{n\xi}
\end{aligned}$$

Since $\tilde{\mathbf{A}} = \tilde{\mathbf{\Phi}}_{\mathbf{B}}^\top \mathbf{A} \tilde{\mathbf{\Phi}}_{\mathbf{B}}$ and $\tilde{\mathbf{\Phi}}_{\mathbf{B}} = \mathbf{\Phi}_{\mathbf{B}}(\mathbf{\Lambda}_{\mathbf{B}} + \xi\mathbf{I})^{-\frac{1}{2}}$, inequality (2) and (3) hold. As $\mathbf{\Phi}_{\mathbf{B}}$ is orthonormal matrix implies $\|\mathbf{\Phi}_{\mathbf{B}}\|_2 = 1$ and by Lemma 1, inequality (4) holds. Since $\mathbf{\Lambda}_{\mathbf{B}}$ is diagonal matrix and each entry is nonzero, we have the maxima value of $(\mathbf{\Lambda}_{\mathbf{B}} + \xi\mathbf{I})^{-1}$ is one over the minima value of $\mathbf{\Lambda}_{\mathbf{B}} + \xi\mathbf{I}$. Hence, inequality (5) holds.

Next, we consider the case where matrix B is not fixed. Then,

$$\begin{aligned}
\|2\tilde{\mathbf{A}}\mathbf{v} - 2\tilde{\mathbf{A}}'\mathbf{v}\|_2 &\leq 2 \cdot \|(\tilde{\mathbf{\Phi}}_{\mathbf{B}})^\top \mathbf{A}\tilde{\mathbf{\Phi}}_{\mathbf{B}} - (\tilde{\mathbf{\Phi}}_{\mathbf{B}'})^\top \mathbf{A}'\tilde{\mathbf{\Phi}}_{\mathbf{B}'}\|_2 \\
&\leq 2 \cdot \|(\tilde{\mathbf{\Phi}}_{\mathbf{B}})^\top \mathbf{A}\tilde{\mathbf{\Phi}}_{\mathbf{B}} - (\tilde{\mathbf{\Phi}}_{\mathbf{B}})^\top \mathbf{A}'\tilde{\mathbf{\Phi}}_{\mathbf{B}} + (\tilde{\mathbf{\Phi}}_{\mathbf{B}})^\top \mathbf{A}'\tilde{\mathbf{\Phi}}_{\mathbf{B}} - (\tilde{\mathbf{\Phi}}_{\mathbf{B}'})^\top \mathbf{A}'\tilde{\mathbf{\Phi}}_{\mathbf{B}'}\|_2 \\
&\leq 2 \cdot \|(\tilde{\mathbf{\Phi}}_{\mathbf{B}})^\top \mathbf{A}\tilde{\mathbf{\Phi}}_{\mathbf{B}} - (\tilde{\mathbf{\Phi}}_{\mathbf{B}})^\top \mathbf{A}'\tilde{\mathbf{\Phi}}_{\mathbf{B}}\|_2 + 2 \cdot \|(\tilde{\mathbf{\Phi}}_{\mathbf{B}})^\top \mathbf{A}'\tilde{\mathbf{\Phi}}_{\mathbf{B}} - (\tilde{\mathbf{\Phi}}_{\mathbf{B}'})^\top \mathbf{A}'\tilde{\mathbf{\Phi}}_{\mathbf{B}'}\|_2 \\
&\leq \frac{16}{n\xi} + 2 \cdot \|(\tilde{\mathbf{\Phi}}_{\mathbf{B}})^\top \mathbf{A}'\tilde{\mathbf{\Phi}}_{\mathbf{B}} - (\tilde{\mathbf{\Phi}}_{\mathbf{B}'})^\top \mathbf{A}'\tilde{\mathbf{\Phi}}_{\mathbf{B}'}\|_2
\end{aligned}$$

Recall the within-class covariance matrices $\mathbf{B}$ and $\mathbf{B}'$, we have $n\mathbf{B} - (n-1)\mathbf{B}' = \sum_{i\in\mathcal{C}_1}(x_i - \mu_1)(x_i - \mu_1)^\top - \sum_{i\in\mathcal{C}_1, i\neq n}(x_i - \mu_1')(x_i - \mu_1')^\top$, where $\mu_1 = \frac{1}{n_1}\sum_{i\in\mathcal{C}_1} x_i$ and $\mu_1' = \frac{1}{n_1-1}\sum_{i\in\mathcal{C}_1, i\neq n} x_i$. It is clear that $\mu_1 \approx \mu_1'$ as n_1 is large enough. Thus, $n\mathbf{B} - (n-1)\mathbf{B}' \approx (x_n - \mu_1)(x_n - \mu_1)^\top$. I.e., $\mathbf{B} \approx \frac{(n-1)}{n}\mathbf{B}' + \frac{1}{n}(x_n - \mu_1)(x_n - \mu_1)^\top$. So, we have $\mathbf{B} \approx \mathbf{B}'$ as n is large enough. Therefore, $\|(\tilde{\mathbf{\Phi}}_{\mathbf{B}})^\top \mathbf{A}'\tilde{\mathbf{\Phi}}_{\mathbf{B}} - (\tilde{\mathbf{\Phi}}_{\mathbf{B}'})^\top \mathbf{A}'\tilde{\mathbf{\Phi}}_{\mathbf{B}'}\|_2 = O(\frac{1}{n\xi})$. To support this approximation, we conduct a series of controlled experiments under various ξ and sample size conditions. When ξ is not too large, the empirical results tend to exhibit behavior consistent with the theoretical bound of $O(\frac{1}{n\xi})$. These numerical experiments are provided in the appendix for reference.

■

(**Remarks:** The above argument requires n and n_1 to be large enough. In the appendix, we estimate the gaps between our analysis and empirical study. It indicates that the solutions converge when n is large enough. For small data set or imbalance data, it needs more rigorous analysis to justify our argument.)

5. Proof of Theorem 14

As shown in Fact 12 and Fact 13, one fundamental property of Gaussian distributions is that the linear combination of two Gaussian distributions is also Gaussian. Another important property is the Gaussian tail bound.

Fact 12 *If X, Y are i.i.d $\mathcal{N}(\mu, \sigma^2)$ and a, b are constants, then $aX+bY \sim \mathcal{N}((a+b)\mu, (a^2+b^2)\sigma^2)$*

Fact 13 *If $X \sim \mathcal{N}(0, \sigma^2)$, then the standard Gaussian tail bound is $\Pr[|X| \geq t] \leq e^{-\frac{t^2}{2\sigma^2}}$, when $t \geq \frac{2\sigma}{\sqrt{2\pi}}$.*

Lemma 14 *Let $\boldsymbol{\zeta} \in \mathbb{R}^d$ be a Gaussian noise vector with each entry $\zeta_i \sim \mathcal{N}(0, \sigma^2)$. Then, we have with probability at least $1-\beta$,*

$$\|\boldsymbol{\zeta}\|_2 \leq O(\sigma\sqrt{d \log \frac{d}{\beta}}).$$

Proof By Fact 13, we have $\Pr[|\zeta_i| \geq t] \leq e^{-\frac{t^2}{2\sigma^2}}$, where ζ_i is sampled from $\mathcal{N}(0, \sigma^2)$. We can use Gaussian distribution tail-bound to bound the maximum norm of the vector $\boldsymbol{\zeta}$.

$$\Pr[\|\boldsymbol{\zeta}\|_\infty \geq t] = \Pr[|\zeta_1| \geq t \cup \dots \cup |\zeta_d| \geq t] \overset{(1)}{\leq} d \cdot \Pr[|\zeta_i| \geq t] \leq d \cdot e^{-\frac{t^2}{2\sigma^2}}.$$

(1) is by the union bound. Let the failure probability $\beta \leq d \cdot e^{-\frac{t^2}{2\sigma^2}}$. Then, we can get $t \leq O(\sigma\sqrt{\log \frac{d}{\beta}})$. Finally, by the relation between ℓ_2-norm and ℓ_∞-norm, we have $\|\boldsymbol{\zeta}\|_2 \leq \sqrt{d} \cdot \|\boldsymbol{\zeta}\|_\infty \leq O(\sigma\sqrt{d \log \frac{d}{\beta}})$. ■

Lemma 15 (Stewart 1979) *Let $\mathbf{M}$ and $\mathbf{N}$ be $d \times d$ symmetric matrices. Then, for all $i \in \{1, ..., d\}$,*

$$\lambda_i(\mathbf{M}) + \lambda_{\min}(\mathbf{N}) \leq \lambda_i(\mathbf{M} + \mathbf{N}) \leq \lambda_i(\mathbf{M}) + \lambda_{\max}(\mathbf{N}).$$

Theorem 14 *(restate) With Algorithm 2 and if we set the step size $\eta_B = \frac{2}{\lambda_{\max}(\hat{\mathbf{B}})+\lambda_{\min}(\hat{\mathbf{B}})}$, then with probability at least $1-\beta$,*

$$\|\mathbf{v}_* - \mathbf{v}_m\|_2 \leq O(\frac{\sqrt{d \log{(\frac{d}{\beta})} \log{(\frac{1}{\delta})}}}{n\epsilon\xi}), \tag{1}$$

where $\mathbf{v}_*$ is the optimal eigenvector of $\hat{\mathbf{B}}$ and $\mathbf{v}_m$ is the output vector after Step 4 of Algorithm 2.

Proof Let $\mathbf{v}_t$ be the t-th iteration vector in Step 3 of Algorithm 2 and $\mathbf{v}_*$ be the optimal eigenvector of the matrix $\hat{\mathbf{B}}$, which means $\hat{\mathbf{B}}\mathbf{v}_* = \lambda_*\mathbf{v}_*$ for some λ_*. Then, the error between $\mathbf{v}_t$ and $\mathbf{v}_*$ is $\mathbf{v}_* - \mathbf{v}_t$, denoted as $\mathbf{e}_t$. We define the error function as $\mathbf{E}(\mathbf{v}) = \frac{1}{2}(\mathbf{v}_* - \mathbf{v})^\top\hat{\mathbf{B}}(\mathbf{v}_* - \mathbf{v})$. If $\mathbf{v} = \mathbf{v}_*$, then $\mathbf{E}(\mathbf{v}_*) = 0$. On the other hand, the error function can be expressed as:

$$
\begin{aligned}
\mathbf{E}(\mathbf{v}) &= \frac{1}{2}\langle \mathbf{v}_* - \mathbf{v}, \hat{\mathbf{B}}(\mathbf{v}_* - \mathbf{v})\rangle = \frac{1}{2}\langle \mathbf{v}_*, \hat{\mathbf{B}}\mathbf{v}_*\rangle - \frac{1}{2}\langle \mathbf{v}_*, \hat{\mathbf{B}}\mathbf{v}\rangle - \frac{1}{2}\langle \mathbf{v}, \hat{\mathbf{B}}\mathbf{v}_*\rangle + \frac{1}{2}\langle \mathbf{v}, \hat{\mathbf{B}}\mathbf{v}\rangle \\
&= \frac{1}{2}\langle \mathbf{v}_*, \hat{\mathbf{B}}\mathbf{v}_*\rangle - \langle \mathbf{v}, \hat{\mathbf{B}}\mathbf{v}_*\rangle + \frac{1}{2}\langle \mathbf{v}, \hat{\mathbf{B}}\mathbf{v}\rangle.
\end{aligned}
$$

Hence, the gradient function of $\mathbf{E}(\mathbf{v})$ with respect to $\mathbf{v}$ is $\nabla\mathbf{E}(\mathbf{v}) = \hat{\mathbf{B}}\mathbf{v} - \hat{\mathbf{B}}\mathbf{v}_*$. By the gradient descent method, we obtain the following:

$$
\begin{aligned}
\mathbf{v}_{t+1} = \mathbf{v}_t - \eta_B\nabla\mathbf{E}(\mathbf{v}_t) = \mathbf{v}_t - \eta_B\hat{\mathbf{B}}(\mathbf{v_t} - \mathbf{v}_*) &\Longrightarrow \mathbf{v}_* - \mathbf{v}_{t+1} = \mathbf{v}_* - \mathbf{v}_t - \eta_B\hat{\mathbf{B}}(\mathbf{v}_* - \mathbf{v}_t) \\
&\Longrightarrow \mathbf{e}_{t+1} = \mathbf{e}_t - \eta_B\hat{\mathbf{B}}\mathbf{e}_t = (\mathbf{I} - \eta_B\hat{\mathbf{B}})\mathbf{e}_t, \qquad (*)
\end{aligned}
$$

where η_B is the step size. Based on the above formula, we can further analyze its ℓ_2-norm.

$$
\begin{aligned}
\|\mathbf{e}_t\|_2 &= \|(\mathbf{I} - \eta_B\hat{\mathbf{B}})\mathbf{e}_{t-1}\|_2 \leq \|\mathbf{I} - \eta_B\hat{\mathbf{B}}\|_2 \cdot \|\mathbf{e}_{t-1}\|_2 = \max_i \{|1 - \eta_B\lambda_i(\hat{\mathbf{B}})|\} \cdot \|\mathbf{e}_{t-1}\|_2 \\
&= \max\{|1 - \eta_B\lambda_{\max}(\hat{\mathbf{B}})|, |1 - \eta_B\lambda_{\min}(\hat{\mathbf{B}})|\} \cdot \|\mathbf{e}_{t-1}\|_2 \\
&\overset{(1)}{=} \left(\frac{\lambda_{\max}(\hat{\mathbf{B}}) - \lambda_{\min}(\hat{\mathbf{B}})}{\lambda_{\max}(\hat{\mathbf{B}}) + \lambda_{\min}(\hat{\mathbf{B}})}\right) \cdot \|\mathbf{e}_{t-1}\|_2 \leq \left(\frac{\lambda_{\max}(\hat{\mathbf{B}}) - \lambda_{\min}(\hat{\mathbf{B}})}{\lambda_{\max}(\hat{\mathbf{B}}) + \lambda_{\min}(\hat{\mathbf{B}})}\right)^t \cdot \|\mathbf{e}_0\|_2,
\end{aligned}
$$

where $\lambda_i(\hat{\mathbf{B}}), \lambda_{\max}(\hat{\mathbf{B}})$ and $\lambda_{\min}(\hat{\mathbf{B}})$ is the i-th eigenvalue, maximum eigenvalue, and minimum eigenvalue of the matrix $\hat{\mathbf{B}}$, respectively. Equality (1) is true by setting $\eta_B = \frac{2}{\lambda_{\max}(\hat{\mathbf{B}}) + \lambda_{\min}(\hat{\mathbf{B}})}$. The previous derivation is only applicable to the noiseless case. Now, we introduce Gaussian noise into the derivation of the error bounds, starting from the gradient function.

$$
\mathbf{e}_{t+1} = \mathbf{e}_t - \eta_B\hat{\mathbf{B}}\mathbf{e}_t + \boldsymbol{\zeta}_t = (\mathbf{I} - \eta_B\hat{\mathbf{B}})\mathbf{e}_t + \boldsymbol{\zeta}_t,
$$

where $\boldsymbol{\zeta}_t$ is i.i.d sampled from $\mathcal{N}(0, \sigma^2)$. Since $\hat{\mathbf{B}} = \mathbf{B} + \xi\mathbf{I}$, by Lemma 15, we have

$$
\begin{aligned}
\lambda_{\max}(\mathbf{B}) + \lambda_{\min}(\xi\mathbf{I}) \leq \lambda_{\max}(\mathbf{B} + \xi\mathbf{I}) \leq \lambda_{\max}(\mathbf{B}) + \lambda_{\max}(\xi\mathbf{I}), \\
\lambda_{\min}(\mathbf{B}) + \lambda_{\min}(\xi\mathbf{I}) \leq \lambda_{\min}(\mathbf{B} + \xi\mathbf{I}) \leq \lambda_{\min}(\mathbf{B}) + \lambda_{\max}(\xi\mathbf{I}).
\end{aligned}
$$

On the other hand, by Lemma 1, we have $\|\mathbf{B}\|_2 \leq 1$, that is, $\sigma_{\max}(\mathbf{B}) \leq 1$. Since $\mathbf{B}$ is symmetric matrix, $\sigma_{\max}(\mathbf{B}) \leq 1$ implies $\lambda_{\max}(\mathbf{B}) \leq 1$. Since $\lambda_i(\xi\mathbf{I}) = \xi$ for any i and $\lambda_i(\mathbf{B}) \geq 0$ ($\mathbf{B}$ is positive semidefinite), we have $\lambda_{\max}(\hat{\mathbf{B}}) \leq 1 + \xi$ and $\xi \leq \lambda_{\min}(\hat{\mathbf{B}})$. Hence, we have the followings:

$$
\begin{aligned}
\left(\frac{\lambda_{\max}(\hat{\mathbf{B}}) - \lambda_{\min}(\hat{\mathbf{B}})}{\lambda_{\max}(\hat{\mathbf{B}}) + \lambda_{\min}(\hat{\mathbf{B}})}\right) &= \left(\frac{2\lambda_{\max}(\hat{\mathbf{B}})}{\lambda_{\max}(\hat{\mathbf{B}}) + \lambda_{\min}(\hat{\mathbf{B}})} - 1\right) \overset{(2)}{=} \left(\frac{2}{1 + \frac{\lambda_{\min}(\hat{\mathbf{B}})}{\lambda_{\max}(\hat{\mathbf{B}})}} - 1\right) \\
&\overset{(3)}{\leq} \frac{2}{1 + \frac{\xi}{1+\xi}} - 1 \leq \frac{2 + 2\xi}{1 + 2\xi} - 1 = \frac{1}{1 + 2\xi}.
\end{aligned}
$$

By dividing both the numerator and the denominator by $\lambda_{\max}(\hat{\mathbf{B}})$, Equality (2) holds. Since the smaller $\frac{\lambda_{\min}(\hat{\mathbf{B}})}{\lambda_{\max}(\hat{\mathbf{B}})}$ is, the larger $\frac{2}{1+\frac{\lambda_{\min}(\hat{\mathbf{B}})}{\lambda_{\max}(\hat{\mathbf{B}})}}$ is, then inequality (3) holds . We have $\left(\frac{\lambda_{\max}(\hat{\mathbf{B}})-\lambda_{\min}(\hat{\mathbf{B}})}{\lambda_{\max}(\hat{\mathbf{B}})+\lambda_{\min}(\hat{\mathbf{B}})}\right) \leq \frac{1}{1+2\xi}$, denoted as ν_B. Hence, the error bound is

$$\|\mathbf{e}_{t+1}\|_2 = \|(\mathbf{I}-\eta_B\hat{\mathbf{B}})\mathbf{e}_t + \boldsymbol{\zeta}_t\|_2 \leq \nu_B \cdot \|\mathbf{e}_t\|_2 + \|\boldsymbol{\zeta}_t\|_2 \overset{(4)}{\leq} \nu_B \cdot \|\mathbf{e}_t\|_2 + O(\sigma\sqrt{d\log\frac{d}{\beta}})$$
$$\leq \nu_B^t \cdot \|\mathbf{e}_0\|_2 + \left[1+\nu_B+\cdots+\nu_B^{t-1}\right] O(\sigma\sqrt{d\log\frac{d}{\beta}}).$$

Since by Lemma 14, we have $\|\boldsymbol{\zeta}\| \leq O(\sigma\sqrt{d\log\frac{d}{\beta}})$ with probability at least $1-\beta$, then inequality (4) holds. Since $\sigma = O(\frac{\sqrt{\log(1/\delta)}}{n\epsilon})$, the error bound is

$$\begin{aligned}\|\mathbf{e}_m\|_2 &\leq \nu_B^m + \frac{1-\nu_B^m}{1-\nu_B} \cdot O(\frac{\sqrt{d\log(d/\beta)\log(1/\delta)}}{n\epsilon}) \\ &\overset{(5)}{\leq} \nu_B^m + \frac{1}{1-\nu_B} \cdot O(\frac{\sqrt{d\log(d/\beta)\log(1/\delta)}}{n\epsilon}) \\ &\overset{(6)}{\leq} \nu_B^m + O(\frac{\sqrt{d\log(d/\beta)\log(1/\delta)}}{n\epsilon\xi}).\end{aligned}$$

Note $\nu_B^m \geq 0$ implies $1-\nu_B^m \leq 1$, so inequality (5) is true. With $\frac{1}{1-\nu_B} = \frac{1}{1-\frac{1}{1+2\xi}} = 1+\frac{1}{2\xi}$, we have $\frac{1}{1-\nu_B} \leq O(\frac{1}{\xi})$ when $\xi \leq \frac{1}{2}$, so inequality (6) holds. Besides, we can set $m = O(\log(n))$ and $\xi \geq \Omega\left((\frac{\sqrt{d\log(d/\beta)\log(1/\delta)}}{n\epsilon})^{-\frac{1}{m}}\right) - 1$, then $\nu_B^m = (\frac{1}{1+2\xi})^m \leq O(\frac{\sqrt{d\log(d/\beta)\log(1/\delta)}}{n\epsilon})$. Hence, we now have a more concise error bound for $\mathbf{v}_m$ and $\mathbf{v}_*$,

$$\|\mathbf{e}_m\|_2 = \|\mathbf{v}_* - \mathbf{v}_m\|_2 \leq O(\frac{\sqrt{d\log(d/\beta)\log(1/\delta)}}{n\epsilon\xi}).$$

■

6. Proof of Theorem 15

Theorem 15 *(restate) With Algorithm 2 and if we set the step size* $\eta_A = \frac{2}{\lambda_{\max}(\tilde{\mathbf{A}})+\lambda_{\min}(\tilde{\mathbf{A}})}$, *then with probability at least* $1-\beta$,

$$\|\mathbf{v}'_* - \mathbf{v}'_m\|_2 \leq O(\frac{\sqrt{d\log(\frac{d}{\beta})\log(\frac{1}{\delta})}}{n\epsilon\xi}), \tag{2}$$

where $\mathbf{v}'_*$ *is the optimal eigenvector of* $\tilde{\mathbf{A}}$ *and* $\mathbf{v}'_m$ *is the output vector after Step 11.*

Proof Since the proof strategy is similar to Theorem 14, similarly, we can derive the following formula:

$$\|\mathbf{e}'_m\|_2 \leq \nu_A^m + O(\frac{\sqrt{d\log{(d/\beta)}\log{(1/\delta)}}}{n\epsilon\xi}),$$

where $\mathbf{e}'_m$ is the error between the optimal eigenvector and the output eigenvector in Algorithm 2 of the matrix $\tilde{\mathbf{A}}$. We present the bound of $\lambda_{\max}(\tilde{\mathbf{A}})$ and $\lambda_{\min}(\tilde{\mathbf{A}})$. We have

$$\lambda_{\max}(\tilde{\mathbf{A}}) = \|\tilde{\mathbf{A}}\|_2 \leq \|\mathbf{\Phi}_{\hat{\mathbf{B}}}^\top \mathbf{\Lambda}_{\hat{\mathbf{B}}}^{-1/2} \mathbf{A} \mathbf{\Lambda}_{\hat{\mathbf{B}}}^{-1/2} \mathbf{\Phi}_{\hat{\mathbf{B}}}\|_2 \leq 1 \cdot \frac{1}{\sqrt{\xi}} \cdot 1 \cdot \frac{1}{\sqrt{\xi}} \cdot 1 = \frac{1}{\xi}$$

$$\lambda_{\min}(\tilde{\mathbf{A}}) \geq \frac{1}{\|\tilde{\mathbf{A}}^{-1}\|_2} = \frac{1}{\|(\mathbf{\Phi}_{\hat{\mathbf{B}}}^\top \mathbf{\Lambda}_{\hat{\mathbf{B}}}^{-1/2} \mathbf{A} \mathbf{\Lambda}_{\hat{\mathbf{B}}}^{-1/2} \mathbf{\Phi}_{\hat{\mathbf{B}}})^{-1}\|_2} \geq \frac{1}{\sqrt{1+\xi} \cdot 1 \cdot \sqrt{1+\xi}} = \frac{1}{1+\xi}.$$

The inequality holds because $\xi \leq \lambda_i(\hat{\mathbf{B}}) \leq 1+\xi$ implies $\frac{1}{\sqrt{\xi+1}} \leq \frac{1}{\sqrt{\lambda_i(\hat{\mathbf{B}})}} = \lambda_i(\mathbf{\Lambda}_{\hat{\mathbf{B}}}^{-1/2}) \leq \frac{1}{\sqrt{\xi}}$. We have $\frac{\lambda_{\max}(\tilde{\mathbf{A}})-\lambda_{\min}(\tilde{\mathbf{A}})}{\lambda_{\max}(\tilde{\mathbf{A}})+\lambda_{\min}(\tilde{\mathbf{A}})} \leq \frac{1}{1+2\xi}$, denoted as ν_A. Setting $\eta_A = \frac{2}{\lambda_{\max}(\tilde{\mathbf{A}})+\lambda_{\min}(\tilde{\mathbf{A}})}$ and $\xi = \Omega\left((\frac{\sqrt{d\log{(d/\beta)}\log{(1/\delta)}}}{n\epsilon})^{-\frac{1}{m}} - 1\right)$, we have the above result. ■

7. Proof of Theorem 16

Theorem 16 *(restate) With Algorithm 2, Theorem 14 and 15, we have,*

$$1 - \langle \boldsymbol{\phi}^*, \boldsymbol{\phi}_m \rangle \leq O(\frac{d\log{(\frac{d}{\beta})}\log{(\frac{1}{\delta})}}{n^2\epsilon^2\xi^2}), \tag{3}$$

where $\boldsymbol{\phi}^*$ is the optimal generalized eigenvector of symmetric-definite pair $\{\mathbf{A}, \mathbf{B}\}$ and $\boldsymbol{\phi}$ is the output vector of Algorithm 2.

Proof According to the half-angle formula, we can derive the following formula:

$$1 - \langle \boldsymbol{\phi}^*, \boldsymbol{\phi}_m \rangle \leq \frac{1}{2}\|\boldsymbol{\phi}^* - \boldsymbol{\phi}_m\|_2^2 = \frac{1}{2}\|\tilde{\mathbf{\Phi}}_{\hat{\mathbf{B}}}^* \mathbf{v}'_* - \tilde{\mathbf{\Phi}}_{\hat{\mathbf{B}}} \mathbf{v}'_m\|_2^2.$$

Through the proofs of Theorem 14 and 15, we can conclude that.

$$\|\mathbf{v}_* - \mathbf{v}_m\|_2 \leq O(\frac{\sqrt{d\log{(d/\beta)}\log{(1/\delta)}}}{n\epsilon\xi}), \quad \|\mathbf{v}'_* - \mathbf{v}'_m\|_2 \leq O(\frac{\sqrt{d\log{(d/\beta)}\log{(1/\delta)}}}{n\epsilon\xi}).$$

Based on this, we need to rearrange the original formula and derive the final result.

$$\begin{aligned}\|\tilde{\mathbf{\Phi}}_{\hat{\mathbf{B}}}^* \mathbf{v}'_* - \tilde{\mathbf{\Phi}}_{\hat{\mathbf{B}}} \mathbf{v}'_m\|_2 &= \|\tilde{\mathbf{\Phi}}_{\hat{\mathbf{B}}}^* \mathbf{v}'_* - \tilde{\mathbf{\Phi}}_{\hat{\mathbf{B}}}^* \mathbf{v}'_m + \tilde{\mathbf{\Phi}}_{\hat{\mathbf{B}}}^* \mathbf{v}'_m - \tilde{\mathbf{\Phi}}_{\hat{\mathbf{B}}} \mathbf{v}'_m\|_2 \\ &\leq \|\tilde{\mathbf{\Phi}}_{\hat{\mathbf{B}}}^* (\mathbf{v}'_* - \mathbf{v}'_m) + (\tilde{\mathbf{\Phi}}_{\hat{\mathbf{B}}}^* - \tilde{\mathbf{\Phi}}_{\hat{\mathbf{B}}}) \mathbf{v}'_m\|_2 \\ &\leq \|\tilde{\mathbf{\Phi}}_{\hat{\mathbf{B}}}^* (\mathbf{v}'_* - \mathbf{v}'_m)\|_2 + \|(\tilde{\mathbf{\Phi}}_{\hat{\mathbf{B}}}^* - \tilde{\mathbf{\Phi}}_{\hat{\mathbf{B}}}) \mathbf{v}'_m\|_2.\end{aligned}$$

For the first part,

$$\|\tilde{\mathbf{\Phi}}^*_{\hat{\mathbf{B}}}(\mathbf{v}'_* - \mathbf{v}'_m)\|_2 \leq \|\tilde{\mathbf{\Phi}}^*_{\hat{\mathbf{B}}}\|_2 \cdot \|\mathbf{v}'_* - \mathbf{v}'_m\|_2 \overset{(1)}{\leq} \frac{1}{\sqrt{\xi}} \cdot O(\frac{\sqrt{d\log{(d/\beta)}\log{(1/\delta)}}}{n\epsilon\xi}).$$

Since $\|\tilde{\mathbf{\Phi}}_{\hat{\mathbf{B}}}\|_2 \leq \|\mathbf{\Phi}_{\hat{\mathbf{B}}}\mathbf{\Lambda}_{\hat{\mathbf{B}}}^{-\frac{1}{2}})\|_2 \leq 1 \cdot \frac{1}{\sqrt{\xi}}$, inequality (1) is true. The second part is

$$\|(\tilde{\mathbf{\Phi}}^*_{\hat{\mathbf{B}}} - \tilde{\mathbf{\Phi}}_{\hat{\mathbf{B}}})\mathbf{v}'_m\|_2 \leq \|\tilde{\mathbf{\Phi}}^*_{\hat{\mathbf{B}}} - \tilde{\mathbf{\Phi}}_{\hat{\mathbf{B}}}\|_2 \cdot \|\mathbf{v}'_m\|_2 \leq \|\tilde{\mathbf{\Phi}}^*_{\hat{\mathbf{B}}} - \tilde{\mathbf{\Phi}}_{\hat{\mathbf{B}}}\|_2.$$

To derive $\|\tilde{\mathbf{\Phi}}^*_{\hat{\mathbf{B}}} - \tilde{\mathbf{\Phi}}_{\hat{\mathbf{B}}}\|_2$, we first need to derive the following formula:

$$\begin{aligned}
\|{\mathbf{\Lambda}^*_{\hat{\mathbf{B}}}}^{-\frac{1}{2}} - \mathbf{\Lambda}_{\hat{\mathbf{B}}}^{-\frac{1}{2}}\|_2 &= \|{\mathbf{\Phi}^*_{\hat{\mathbf{B}}}}^{-1}\hat{\mathbf{B}}^{-\frac{1}{2}}\mathbf{\Phi}^*_{\hat{\mathbf{B}}} - \mathbf{\Phi}_{\hat{\mathbf{B}}}^{-1}\hat{\mathbf{B}}^{-\frac{1}{2}}\mathbf{\Phi}_{\hat{\mathbf{B}}}\|_2 \\
&= \|{\mathbf{\Phi}^*_{\hat{\mathbf{B}}}}^{-1}\hat{\mathbf{B}}^{-\frac{1}{2}}\mathbf{\Phi}^*_{\hat{\mathbf{B}}} - \mathbf{\Phi}_{\hat{\mathbf{B}}}^{-1}\hat{\mathbf{B}}^{-\frac{1}{2}}\mathbf{\Phi}^*_{\hat{\mathbf{B}}} + \mathbf{\Phi}_{\hat{\mathbf{B}}}^{-1}\hat{\mathbf{B}}^{-\frac{1}{2}}\mathbf{\Phi}^*_{\hat{\mathbf{B}}} - \mathbf{\Phi}_{\hat{\mathbf{B}}}^{-1}\hat{\mathbf{B}}^{-\frac{1}{2}}\mathbf{\Phi}_{\hat{\mathbf{B}}}\|_2 \\
&\leq \|({\mathbf{\Phi}^*_{\hat{\mathbf{B}}}}^{-1} - \mathbf{\Phi}_{\hat{\mathbf{B}}}^{-1})\hat{\mathbf{B}}^{-\frac{1}{2}}\mathbf{\Phi}^*_{\hat{\mathbf{B}}}\|_2 + \|\mathbf{\Phi}_{\hat{\mathbf{B}}}^{-1}\hat{\mathbf{B}}^{-\frac{1}{2}}(\mathbf{\Phi}^*_{\hat{\mathbf{B}}} - \mathbf{\Phi}_{\hat{\mathbf{B}}})\|_2 \\
&\leq \|{\mathbf{\Phi}^*_{\hat{\mathbf{B}}}}^T - \mathbf{\Phi}_{\hat{\mathbf{B}}}^T\|_2 \cdot \|\hat{\mathbf{B}}^{-\frac{1}{2}}\mathbf{\Phi}^*_{\hat{\mathbf{B}}}\|_2 + \|\mathbf{\Phi}_{\hat{\mathbf{B}}}^T\hat{\mathbf{B}}^{-\frac{1}{2}}\|_2 \cdot \|\mathbf{\Phi}^*_{\hat{\mathbf{B}}} - \mathbf{\Phi}_{\hat{\mathbf{B}}}\|_2.
\end{aligned}$$

By eigendecomposition, we have $\hat{\mathbf{B}} = \mathbf{\Phi}^*_{\hat{\mathbf{B}}}\mathbf{\Lambda}^*_{\hat{\mathbf{B}}}{\mathbf{\Phi}^*_{\hat{\mathbf{B}}}}^{-1}$, then

$$\|\hat{\mathbf{B}}^{-\frac{1}{2}}\|_2 \overset{(2)}{=} \|\mathbf{\Phi}^*_{\hat{\mathbf{B}}}\left(\mathbf{\Lambda}^*_{\hat{\mathbf{B}}}\right)^{-\frac{1}{2}}{\mathbf{\Phi}^*_{\hat{\mathbf{B}}}}^{-1}\|_2 \leq 1 \cdot \|{\mathbf{\Lambda}^*_{\hat{\mathbf{B}}}}^{-\frac{1}{2}}\|_2 \cdot 1 \leq \frac{1}{\sqrt{\xi}}.$$

Since $\hat{\mathbf{B}}^{-\frac{1}{2}} = \mathbf{\Phi}^*_{\hat{\mathbf{B}}}\left(\mathbf{\Lambda}^*_{\hat{\mathbf{B}}}\right)^{-\frac{1}{2}}{\mathbf{\Phi}^*_{\hat{\mathbf{B}}}}^{-1}$, then

$$\begin{aligned}
\hat{\mathbf{B}}^{-\frac{1}{2}} \cdot \hat{\mathbf{B}}^{-\frac{1}{2}} &= \mathbf{\Phi}^*_{\hat{\mathbf{B}}}\left(\mathbf{\Lambda}^*_{\hat{\mathbf{B}}}\right)^{-\frac{1}{2}}{\mathbf{\Phi}^*_{\hat{\mathbf{B}}}}^{-1} \cdot \mathbf{\Phi}^*_{\hat{\mathbf{B}}}\left(\mathbf{\Lambda}^*_{\hat{\mathbf{B}}}\right)^{-\frac{1}{2}}{\mathbf{\Phi}^*_{\hat{\mathbf{B}}}}^{-1} \\
&= \mathbf{\Phi}^*_{\hat{\mathbf{B}}}\left(\mathbf{\Lambda}^*_{\hat{\mathbf{B}}}\right)^{-1}{\mathbf{\Phi}^*_{\hat{\mathbf{B}}}}^{-1} = \left(\mathbf{\Phi}^*_{\hat{\mathbf{B}}}\mathbf{\Lambda}^*_{\hat{\mathbf{B}}}{\mathbf{\Phi}^*_{\hat{\mathbf{B}}}}^{-1}\right)^{-1} = \hat{\mathbf{B}}^{-1},
\end{aligned}$$

Equation (2) holds. On the other hand, since we have $\|\mathbf{e}_m\|_2 \leq O(\frac{\sqrt{d\log{(d/\beta)}\log{(1/\delta)}}}{n\epsilon\xi})$, and $\mathbf{\Phi}^*_{\hat{\mathbf{B}}} - \mathbf{\Phi}_{\hat{\mathbf{B}}}$ represents the errors between d eigenvectors. Hence, we have $\mathbf{\Phi}^*_{\hat{\mathbf{B}}} - \mathbf{\Phi}_{\hat{\mathbf{B}}} = [\mathbf{e}^1_m, \cdots, \mathbf{e}^d_m]$, where $\mathbf{e}^i_m$ is the error of the i-th eigenvector. In addition, we can normalize and orthogonalize each error vector, that is, $\mathbf{e}^i_m = \mathbf{u}^i \cdot O(\frac{\sqrt{d\log{(d/\beta)}\log{(1/\delta)}}}{n\epsilon\xi})$ with $\|\mathbf{u}^i\|_2 = 1$ for all $i \in \{1, \cdots, d\}$ and $\langle \mathbf{u}^i, \mathbf{u}^j \rangle = 0$ for all $i \neq j$. We can get that

$$\|\mathbf{\Phi}^*_{\hat{\mathbf{B}}} - \mathbf{\Phi}_{\hat{\mathbf{B}}}\|_2 = O(\frac{\sqrt{d\log{(\frac{d}{\beta})}\log{\frac{1}{\delta}}}}{n\epsilon\xi}) \cdot \|[\mathbf{u}^1, \cdots, \mathbf{u}^d]\|_2 \overset{(3)}{=} O(\frac{\sqrt{d\log{(\frac{d}{\beta})}\log{\frac{1}{\delta}}}}{n\epsilon\xi}.$$

Equation (3) holds because the ℓ_2-norm of an orthonormal matrix is 1. Hence, we can get that

$$\|{\mathbf{\Lambda}^*_{\hat{\mathbf{B}}}}^{-\frac{1}{2}} - \mathbf{\Lambda}_{\hat{\mathbf{B}}}^{-\frac{1}{2}}\|_2 \leq \frac{1}{\sqrt{\xi}} \cdot O(\frac{\sqrt{d\log{(d/\beta)}\log{(1/\delta)}}}{n\epsilon\xi}).$$

Next, we discuss the derivation process for

$$\begin{aligned}
\|\tilde{\mathbf{\Phi}}^*_{\hat{\mathbf{B}}} - \tilde{\mathbf{\Phi}}_{\hat{\mathbf{B}}}\|_2 &= \|\mathbf{\Phi}^*_{\hat{\mathbf{B}}}\mathbf{\Lambda}^{*\,-\frac{1}{2}}_{\hat{\mathbf{B}}} - \mathbf{\Phi}_{\hat{\mathbf{B}}}\mathbf{\Lambda}^{-\frac{1}{2}}_{\hat{\mathbf{B}}}\|_2 \leq \|\mathbf{\Phi}^*_{\hat{\mathbf{B}}}\mathbf{\Lambda}^{*\,-\frac{1}{2}}_{\hat{\mathbf{B}}} - \mathbf{\Phi}^*_{\hat{\mathbf{B}}}\mathbf{\Lambda}^{-\frac{1}{2}}_{\hat{\mathbf{B}}} + \mathbf{\Phi}^*_{\hat{\mathbf{B}}}\mathbf{\Lambda}^{-\frac{1}{2}}_{\hat{\mathbf{B}}} - \mathbf{\Phi}_{\hat{\mathbf{B}}}\mathbf{\Lambda}^{-\frac{1}{2}}_{\hat{\mathbf{B}}}\|_2 \\
&\leq \|\mathbf{\Phi}^*_{\hat{\mathbf{B}}}(\mathbf{\Lambda}^{*\,-\frac{1}{2}}_{\hat{\mathbf{B}}} - \mathbf{\Lambda}^{-\frac{1}{2}}_{\hat{\mathbf{B}}})\|_2 + \|(\mathbf{\Phi}^*_{\hat{\mathbf{B}}} - \mathbf{\Phi}_{\hat{\mathbf{B}}})\mathbf{\Lambda}^{-\frac{1}{2}}_{\hat{\mathbf{B}}}\|_2 \\
&\leq \|\mathbf{\Phi}^*_{\hat{\mathbf{B}}}\|_2 \cdot \|\mathbf{\Lambda}^{*\,-\frac{1}{2}}_{\hat{\mathbf{B}}} - \mathbf{\Lambda}^{-\frac{1}{2}}_{\hat{\mathbf{B}}}\|_2 + \|\mathbf{\Phi}^*_{\hat{\mathbf{B}}} - \mathbf{\Phi}_{\hat{\mathbf{B}}}\|_2 \cdot \|\mathbf{\Lambda}^{-\frac{1}{2}}_{\hat{\mathbf{B}}}\|_2 \\
&\leq 1 \cdot \frac{1}{\sqrt{\xi}} \cdot O(\frac{\sqrt{d\log(d/\beta)\log(1/\delta)}}{n\epsilon\xi}) + O(\frac{\sqrt{d\log(d/\beta)\log(1/\delta)}}{n\epsilon\xi}) \cdot \frac{1}{\sqrt{\xi}} \\
&\leq \frac{1}{\sqrt{\xi}} \cdot O(\frac{\sqrt{d\log(d/\beta)\log(1/\delta)}}{n\epsilon\xi}).
\end{aligned}$$

Hence, the second part is

$$\|(\tilde{\mathbf{\Phi}}^*_{\hat{\mathbf{B}}} - \tilde{\mathbf{\Phi}}_{\hat{\mathbf{B}}})\mathbf{v}'_m\|_2 \leq \frac{1}{\sqrt{\xi}} \cdot O(\frac{\sqrt{d\log(d/\beta)\log(1/\delta)}}{n\epsilon\xi}).$$

Finally, through the preceding derivations, we can establish the following:

$$\begin{aligned}
1 - \langle \boldsymbol{\phi}^*, \boldsymbol{\phi}_m \rangle &\leq \frac{1}{2}\|\boldsymbol{\phi}^* - \boldsymbol{\phi}_m\|_2^2 = \frac{1}{2}\|\tilde{\mathbf{\Phi}}^*_{\hat{\mathbf{B}}}\mathbf{v}'_* - \tilde{\mathbf{\Phi}}_{\hat{\mathbf{B}}}\mathbf{v}'_m\|_2^2 \\
&\leq \left(\|\tilde{\mathbf{\Phi}}^*_{\hat{\mathbf{B}}}\|_2 \cdot \|\mathbf{v}'_* - \mathbf{v}'_m\|_2 + \|\tilde{\mathbf{\Phi}}^*_{\hat{\mathbf{B}}} - \tilde{\mathbf{\Phi}}_{\hat{\mathbf{B}}}\|_2 \cdot \|\mathbf{v}'_m\|_2\right)^2 \\
&\leq \left(\frac{1}{\sqrt{\xi}} \cdot O(\frac{\sqrt{d\log(d/\beta)\log(1/\delta)}}{n\epsilon\xi})\right)^2 \\
&\leq O(\frac{d\log(d/\beta)\log(1/\delta)}{n^2\epsilon^2\xi^2}).
\end{aligned}$$

■

8. Proof of Theorem 17

In the high-dimensional case, we assume $d \gg n$ and analyze whether the truncated operation causes significant errors. Specifically, we examine the effects of the truncated operation on Gaussian noise and eigenvectors obtained through gradient descent. We first focus on the error caused by Gaussian noise. Unlike the proof of the error bound in the high-dimensional case (Hu et al. 2023), they add noise to the symmetric matrices, which allows them to introduce the condition of sparse vectors when proving the ℓ_2-norm error bound. In contrast, intuitively, although the initial Gaussian noise we add is d dimensional, only s dimensions will be affected by the noise after performing the truncated operation. After the truncated operation, the following lemma presents the ℓ_2-norm of a Gaussian noise vector.

Lemma 17 *Let $\boldsymbol{\zeta} \in \mathbb{R}^d$ be a Gaussian noise vector with each entry $\zeta_i \sim \mathcal{N}(0, \sigma^2)$. Then, after the truncated operation, we have with probability at least $1 - \beta$,*

$$\|\boldsymbol{\zeta}_s\|_2 \leq O(\sigma\sqrt{s\log\frac{d}{\beta}}),$$

where $\boldsymbol{\zeta}_s$ is the s-sparse Gaussian noise vector after the truncated operation.

Proof By Lemma 14, with probability at least $1-\beta$

$$\|\boldsymbol{\zeta}\|_\infty \leq O(\sigma\sqrt{\log\frac{d}{\beta}}).$$

Since the truncated operation leaves at most s nonzero entries in the vector, it implies that only s indices remain nonzero in the Gaussian noise. Consequently, we derive the following equation:

$$\|\boldsymbol{\zeta}_s\|_2 \leq \sqrt{s}\|\boldsymbol{\zeta}\|_2 \leq O(\sigma\sqrt{s\log\frac{d}{\beta}})$$

■

Theorem 17 *(restate) With Algorithm 4, with probability at least $1-\beta$, the output generalized eigenvector $\boldsymbol{\phi}_s$ of Algorithm 4 satisfies*

$$1-\langle\boldsymbol{\phi}_s^*,\boldsymbol{\phi}_s\rangle \leq \tilde{O}(\frac{s\log d\log(1/\delta)}{n^2\epsilon^2}), \tag{4}$$

where $\boldsymbol{\phi}_s^*$ is the optimal s-sparse generalized eigenvector of symmetric-definite pair $\{\mathbf{A},\mathbf{B}\}$ and s is the sparsity of eigenvectors.

Proof We show how the error caused by the truncated operation (Algorithm 3) does not affect the error bound after adding Gaussian noise in Algorithm 4. Recall that in Theorem 4 (non-truncated), we have the t-th iteration error in Step 3 of Algorithm 2 from Equation (*)):

$$\mathbf{e}_{t+1} = \mathbf{v}_* - \mathbf{v}_{t+1} = (\mathbf{I}-\eta_B\hat{\mathbf{B}})\cdot\mathbf{e}_t = (\mathbf{I}-\eta_B\hat{\mathbf{B}})\cdot(\mathbf{v}_*-\mathbf{v}_t),$$

where $\mathbf{e}_t = \mathbf{v}_*-\mathbf{v}_t$, $\mathbf{e}_0=\mathbf{v}_*-\mathbf{v}_0$, and $\mathbf{v}_t$ is the t-th iteration vector in Step 3 of Algorithm 2, and $\mathbf{v}_0$, $\mathbf{v}_*$ is the initial vector and optimal eigenvector of matrix $\hat{\mathbf{B}}$, respectively.

Using the analysis of Algorithm 2, we derive the error bound for Algorithm 4. If we apply the truncated operation to the t-th iteration vector $\mathbf{v}_t$ in Step 4 of Algorithm 4, we obtain the following inequality from Equation (*).

$$\mathbf{e}_{t+1} = \mathbf{v}_* - \mathbf{v}_{t+1} = (\mathbf{I}-\eta_B\hat{\mathbf{B}})\cdot(\mathbf{v}_*-(\mathbf{v}_t)_s),$$

where $\mathbf{v}_t$ is the t-th iteration vector in Step 3 of Algorithm 4 and $(\mathbf{v}_t)_s$ is obtained from $\mathbf{v}_t$ through a truncated operation and QR decomposition (Step 5 of Algorithm 4). Move $\mathbf{v}_{t+1}$ to the right side of the equation, and then subtract $(\mathbf{v}_{t+1})_s$ from both sides, then we have

$$\mathbf{v}_* - (\mathbf{v}_{t+1})_s = (\mathbf{I}-\eta_B\hat{\mathbf{B}})\cdot(\mathbf{v}_*-(\mathbf{v}_t)_s) + \mathbf{v}_{t+1} - (\mathbf{v}_{t+1})_s.$$

Hence, the ℓ_2-norm error bound is

$$
\begin{aligned}
\|(\mathbf{e}_{t+1})_s\|_2 &= \|(\mathbf{I} - \eta_B \hat{\mathbf{B}}) \cdot (\mathbf{v}_* - (\mathbf{v}_t)_s) + \mathbf{v}_{t+1} - (\mathbf{v}_{t+1})_s\|_2 \\
&\overset{(1)}{\leq} \left(\frac{\lambda_{\max}(\hat{\mathbf{B}}) - \lambda_{\min}(\hat{\mathbf{B}})}{\lambda_{\max}(\hat{\mathbf{B}}) + \lambda_{\min}(\hat{\mathbf{B}})} \right) \cdot \|(\mathbf{e}_t)_s\|_2 + \|\mathbf{v}_{t+1} - (\mathbf{v}_{t+1})_s\|_2 \\
&\leq \nu_B \cdot \|(\mathbf{e}_t)_s\|_2 + 2 \leq \nu_B^{t+1} \cdot \|(\mathbf{e}_0)_s\|_2 + (1 + \nu_B + \cdots + \nu_B^t) \cdot 2 \\
&\overset{(2)}{\leq} \nu_B^{t+1} \cdot \|(\mathbf{e}_0)_s\|_2 + \frac{1}{1-\nu_B} \cdot 2 \overset{(3)}{\leq} \nu_B^{t+1} \cdot \|(\mathbf{e}_0)_s\|_2 + O(\frac{1}{\xi}) \leq O(\nu_B^{t+1}) + O(\frac{1}{\xi}),
\end{aligned}
$$

where $(\mathbf{e}_t)_s = \mathbf{v}_* - (\mathbf{v}_t)_s$. Since $\|\mathbf{I} - \eta_B \hat{\mathbf{B}}\|_2 = \max\{|1 - \eta_B \lambda_{\max}(\hat{\mathbf{B}})|, |1 - \eta_B \lambda_{\min}(\hat{\mathbf{B}})|\} \leq \left(\frac{\lambda_{\max}(\hat{\mathbf{B}}) - \lambda_{\min}(\hat{\mathbf{B}})}{\lambda_{\max}(\hat{\mathbf{B}}) + \lambda_{\min}(\hat{\mathbf{B}})}\right) \leq \frac{1}{1+2\xi} = \nu_B$ (when $\eta_B = \frac{2}{\lambda_{\max}(\hat{\mathbf{B}}) + \lambda_{\min}(\hat{\mathbf{B}})}$), inequality (1) holds. Since $\nu_B \leq \frac{1}{1+2\xi} \leq 1$ and $1 + \cdots + \nu_B^t = \frac{1-\nu_B^{t+1}}{1-\nu_B} \leq \frac{1}{1-\nu_B}$, inequality (2) holds. Since $\frac{1}{1-\nu_B} = 1 + \frac{1}{2\xi} \leq O(\frac{1}{\xi})$ when $\xi \leq \frac{1}{2}$, inequality (3) holds. We obtain the following inequality, incorporating the error bound caused by adding Gaussian noise (Lemma 17).

$$
\|(\mathbf{e}_t)_s\|_2 \leq O(\frac{1}{\xi}) + O(\frac{\sqrt{s \log d \log 1/\delta}}{n \epsilon \xi}) \overset{(4)}{\leq} O(\frac{\sqrt{s \log d \log 1/\delta}}{n \epsilon \xi}).
$$

Since we assume $d \gg n$, inequality (4) holds when $O(\frac{\sqrt{s \log d \log 1/\delta}}{n\epsilon}) \geq O(1)$. Finally, we can similarly prove this theorem by Lemma 17 and Theorem 14, 15, and 16. ■

9. Explanation of the Wilcoxon Test

In this section, we provide a detailed explanation of the Wilcoxon signed-rank test used in our experiments, as well as the interpretation of the results shown in Table 3. The Wilcoxon signed-rank test is a non-parametric statistical test used to compare two related samples, in this case, the performance of DPRF and DPSR.

The results in Table 3 show the p-values associated with the Wilcoxon test. A p-value less than 0.05 suggests that the difference is statistically significant, supporting the claim that our method, DPSR, performs better than DPRF. On the other hand, a p-value greater than 0.05 implies no significant difference between the two methods. The test results in Table 3 confirm the statistical reliability of DPSR's superior performance in most datasets, except where challenges specific to the dataset occur.

Appendix A. Additional Experimental Results

A.1. MNIST

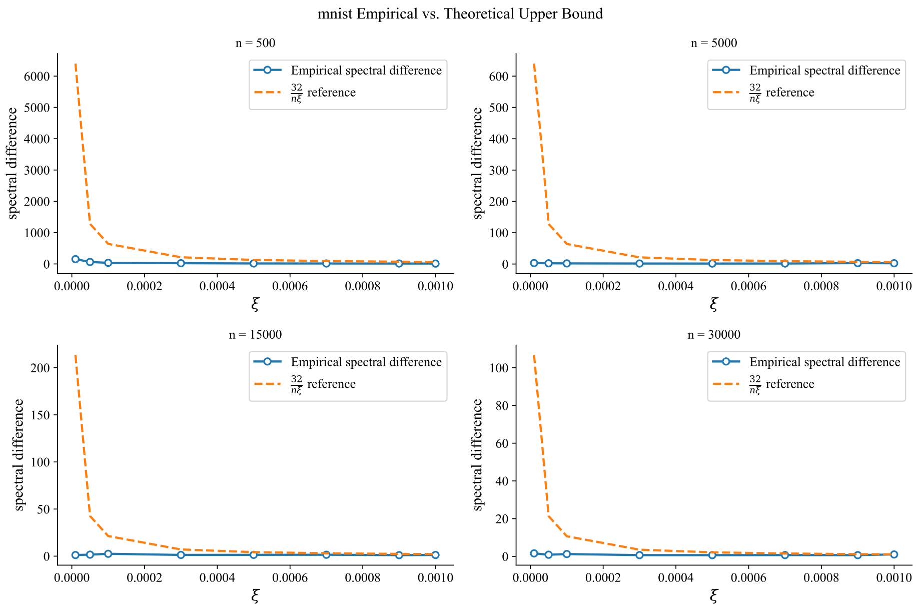

Figure 3: Spectral difference under varying n at fixed ξ.

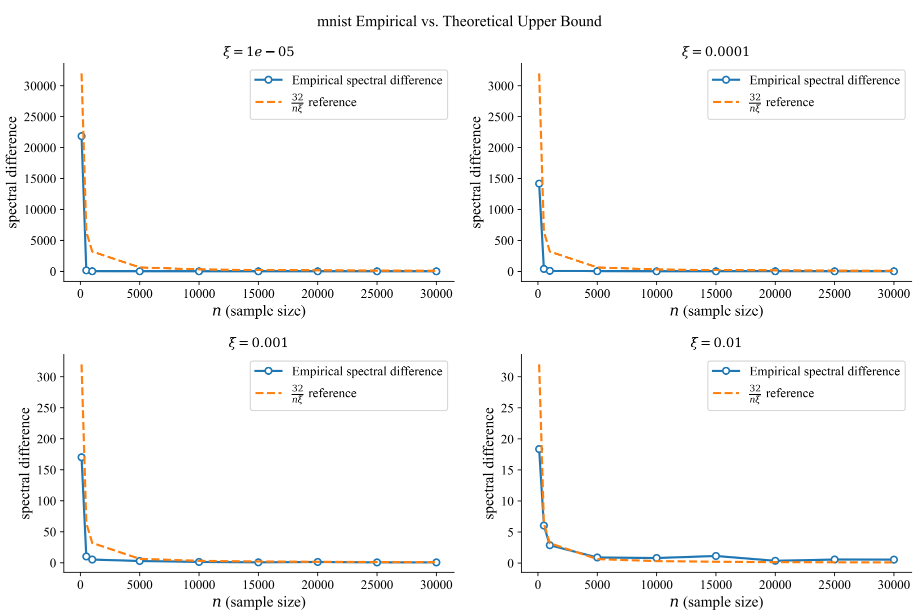

Figure 4: Spectral difference under varying ξ at fixed n.

A.2. a9a

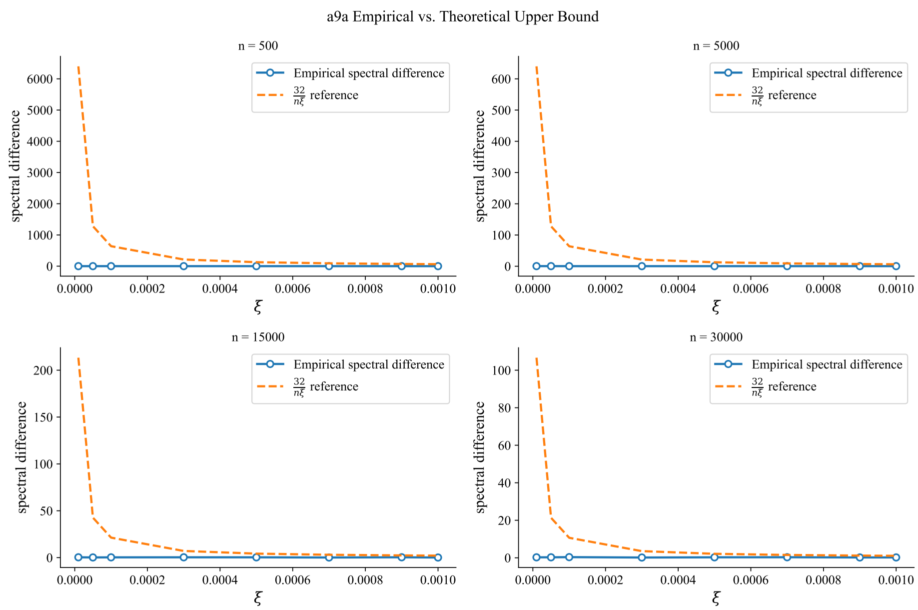

Figure 5: Spectral difference under varying n at fixed ξ.

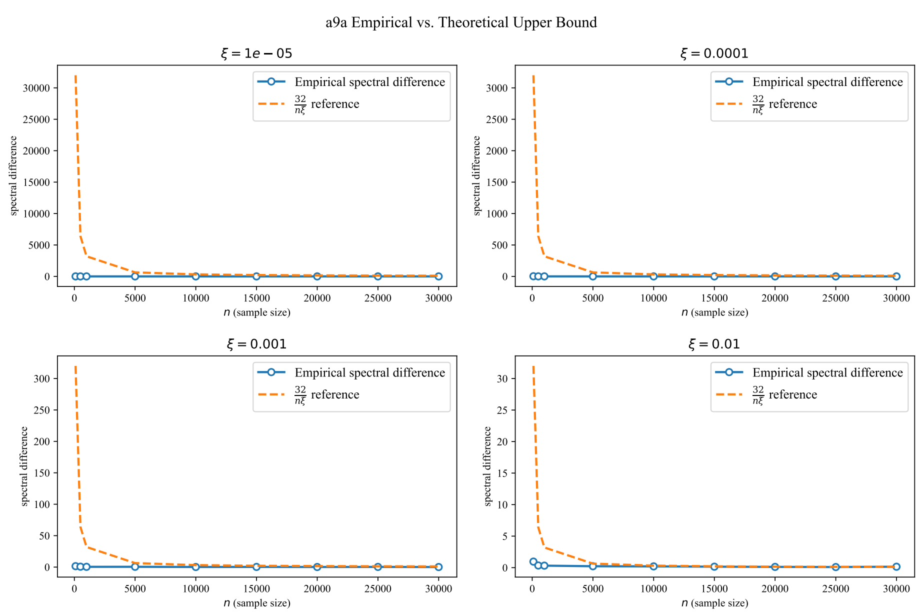

Figure 6: Spectral difference under varying ξ at fixed n.

A.3. IoT22

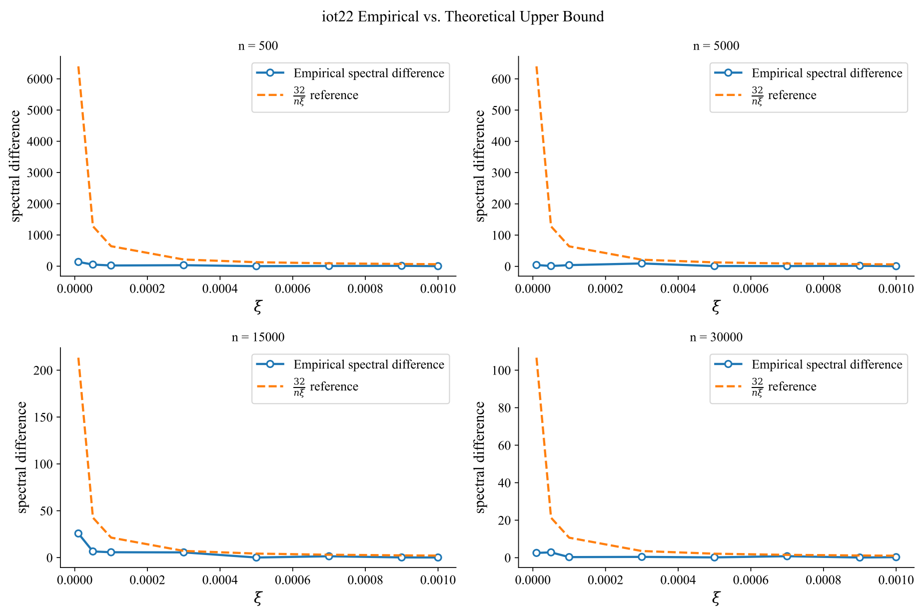

Figure 7: Spectral difference under varying n at fixed ξ.

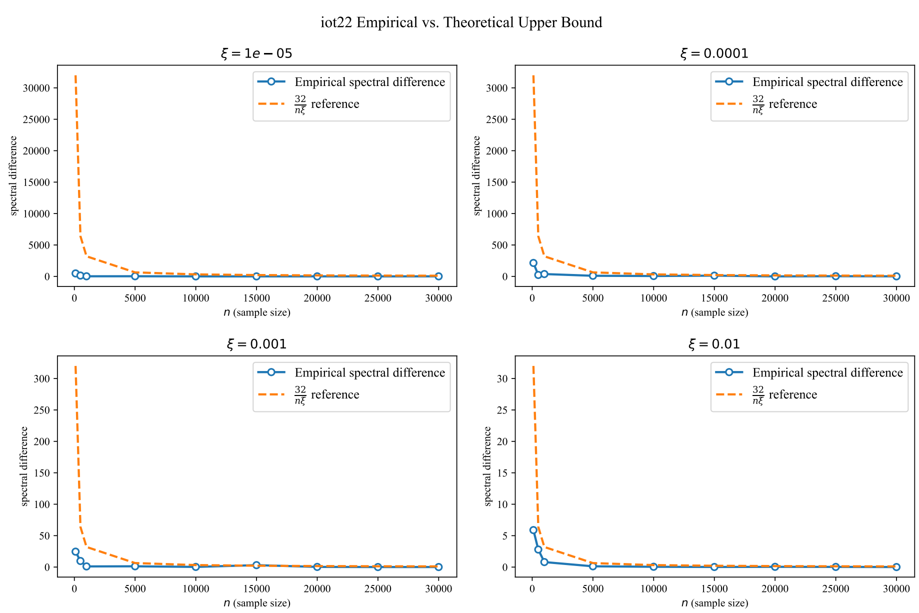

Figure 8: Spectral difference under varying ξ at fixed n.

A.4. Dota2

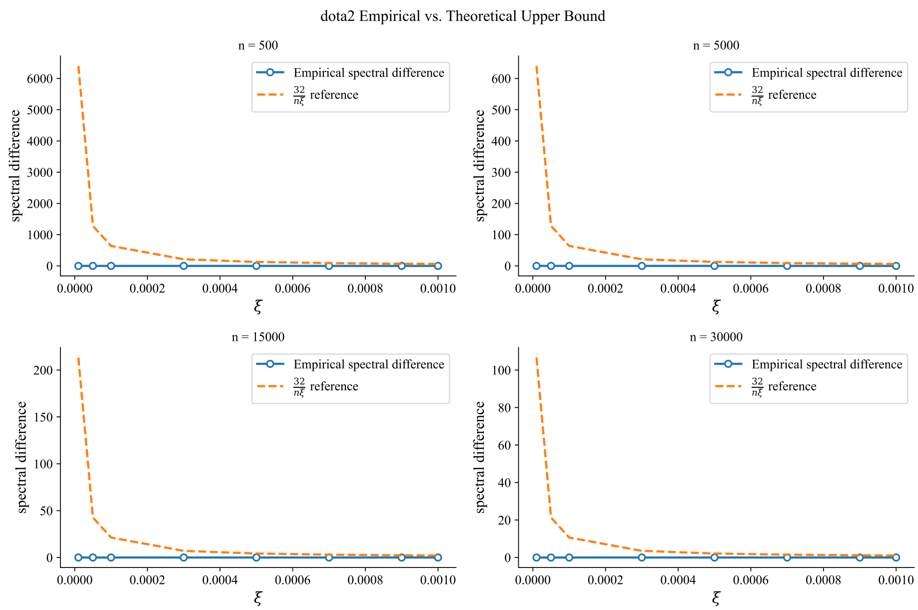

Figure 9: Spectral difference under varying n at fixed ξ.

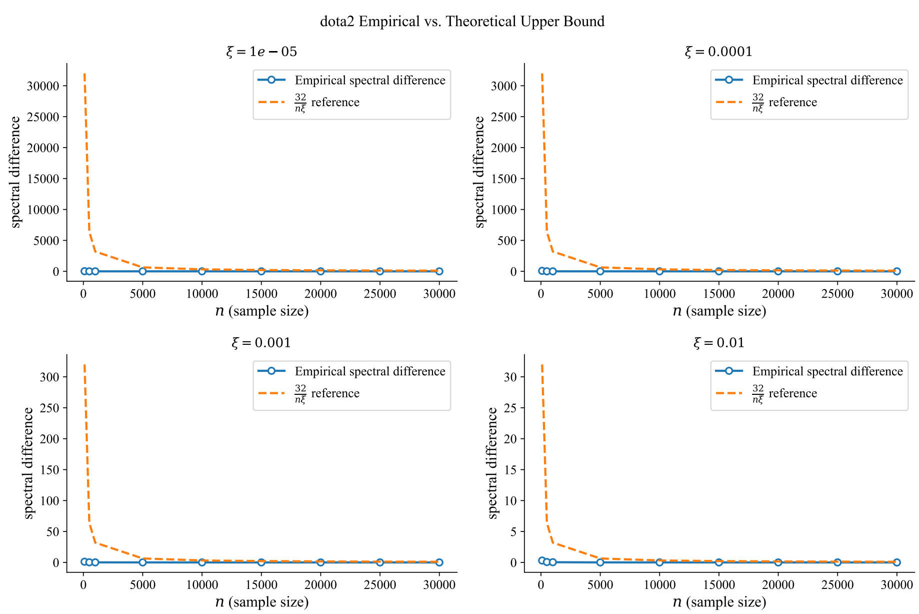

Figure 10: Spectral difference under varying ξ at fixed n.

A.5. Fashion-MNIST

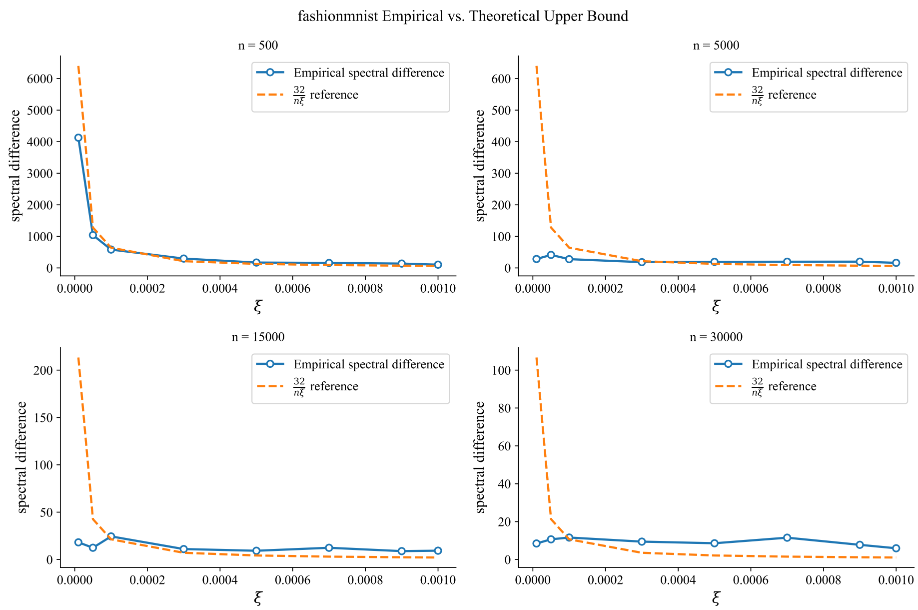

Figure 11: Spectral difference under varying n at fixed ξ.

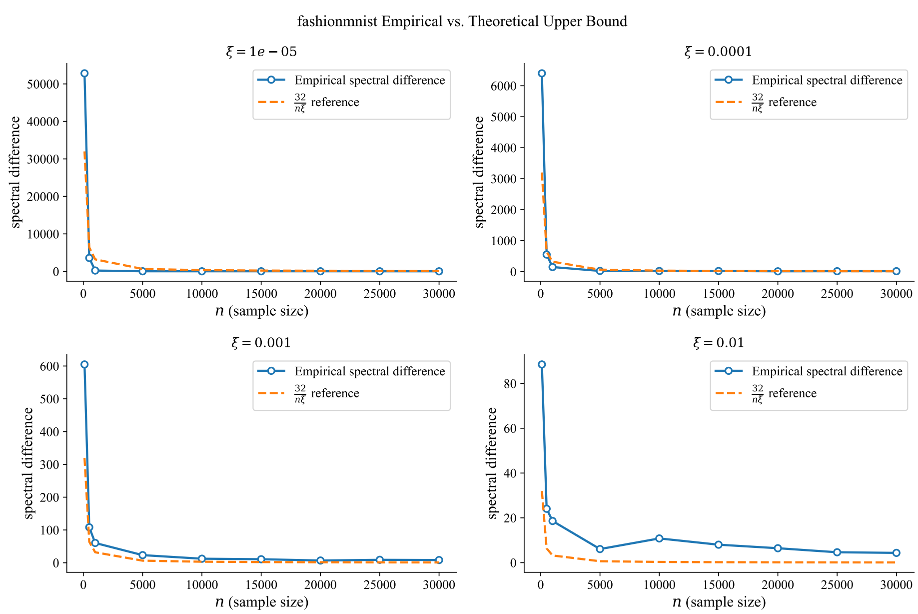

Figure 12: Spectral difference under varying ξ at fixed n.

A.6. CIFAR10

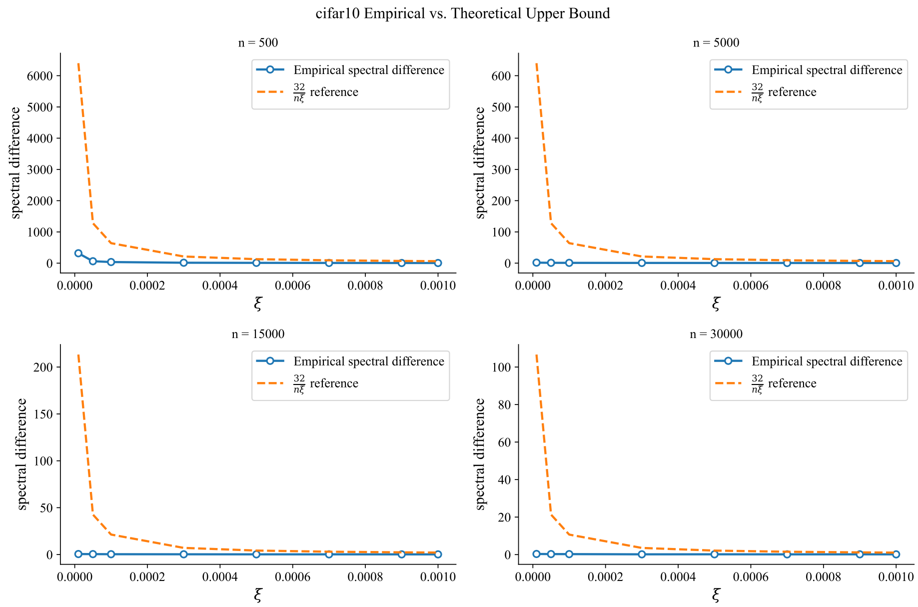

Figure 13: Spectral difference under varying n at fixed ξ.

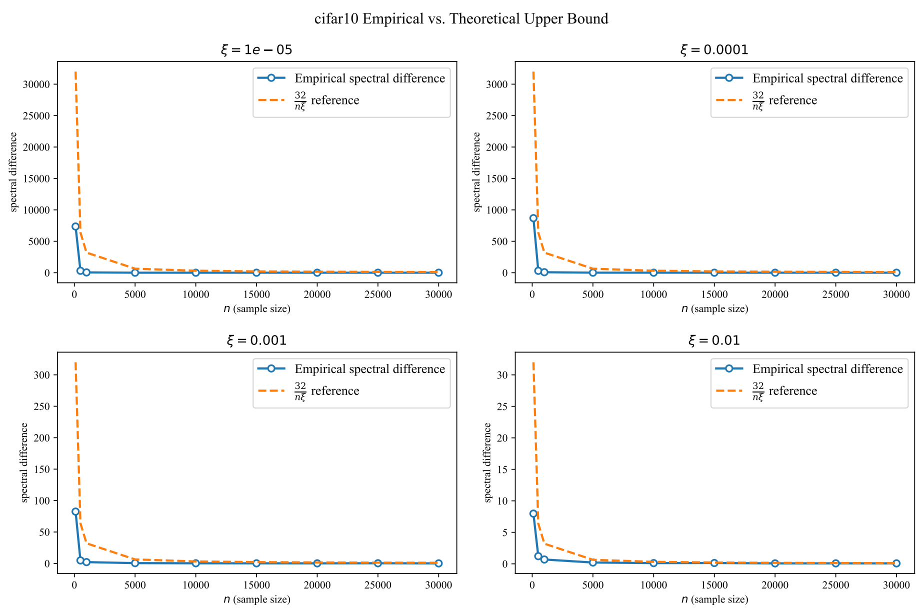

Figure 14: Spectral difference under varying ξ at fixed n.

Additional experiments are conducted on six datasets: MNIST, Fashion-MNIST, a9a, Dota2, CIFAR10, and IoT22. We examine the value of $\|(\tilde{\mathbf{\Phi}}_{\mathbf{B}})^\top \mathbf{A}\tilde{\mathbf{\Phi}}_{\mathbf{B}} - (\tilde{\mathbf{\Phi}}_{\mathbf{B}'})^\top \mathbf{A}'\tilde{\mathbf{\Phi}}_{\mathbf{B}'}\|_2$ under varying conditions of the sample size n and regularization parameter ξ. For each dataset, two sets of plots are provided: one where ξ is fixed and n varies, and one where n is fixed and ξ varies.

Notably, on MNIST and Fashion-MNIST, the empirical spectral difference tends to deviate from the expected $O(1/(n\xi))$ behavior when ξ is not sufficiently small. This suggests that the bound may not hold tightly when the regularization parameter ξ is relatively large.