A Theoretical results for LDA

A.1 Coefficient Setting for Theorem 2.4

Bound of $\sigma_1(M_2)$

We have that with probability greater than

$$1 - Ke^{-\frac{c_2^2}{2}} - KVe^{-\frac{c_1}{2}\min\{\frac{c_1}{2},\sqrt{\beta}\}} - K[\frac{e^{\delta'}}{(1+\delta')^{1+\delta'}}]^{\frac{V\beta}{K(\beta+c_1\beta^{1/2})^2}},$$

we have

$$\sigma_1(M_2) \le \frac{1}{K(K\alpha+1)}\frac{(1+\delta')V(\beta+K\beta^2)}{(V\beta - c_2\sqrt{V\beta})^2}.$$

We can choose c_1, c_2 and δ' as follows to simplify the formula of the bound

- Choose $c_2 = \sqrt{2\log(K/\delta_1)}$, first probability term is less than δ_1.
- Choose $c_1 = \frac{2}{\sqrt{\beta}}\log(KV/\delta_2)$, third probability term is less than δ_2.
- Choose δ' as
$$\delta' = \left(\frac{\log(K/\delta_3)K\left(\beta + 2\log\left(K/\delta_2\right)\right)^2}{V\beta}\right)^{\frac{1}{2}},$$
second probability term is less than δ_3.

As a result, with probability greater than $1-\delta_1-\delta_2-\delta_3$, we have

$$\sigma_1(M_2) \le \frac{1}{K(K\alpha+1)}\frac{(1+\delta')V(\beta+K\beta^2)}{(V\beta - \sqrt{2V\beta\log(K/\delta_1)})^2}.$$

As an alternative, we can choose c_1, c_2 and δ_1 as follows to simplify the formula of the bound

- Choose $c_2 = \sqrt{2\log(K/\delta)}$, first probability term is less than δ.
- Choose $c_1 = \frac{4}{\sqrt{\beta}}\log(KV)$, third probability term is less than $\frac{1}{KV}$.
- Choose $\delta' = 0.1$, second probability term is less than $K(0.995)^{\frac{V(\beta+K\beta^2)}{K(\beta+c_1\beta^{1/2})^2}}$.

As a result, with probability greater than

$$1 - \delta - \frac{1}{KV} - K(0.995)^{\frac{V\beta}{K(\beta+2\log(KV))^2}},$$

we have

$$\sigma_1(M_2) \le \frac{1.1}{K(K\alpha+1)}\frac{V(\beta+K\beta^2)}{(V\beta - \sqrt{2V\beta\log(K/\delta)})^2}.$$

Bound of $\sigma_K(M_2)$

We have that with probability greater than

$$1 - Ke^{-\frac{c_2}{2}\min\{\frac{c_2}{2},V\beta\}} - KVe^{-\frac{c_1}{2}\min\{\frac{c_1}{2},\sqrt{\beta}\}} - K[\frac{e^{-\delta'}}{(1-\delta')^{1-\delta'}}]^{\frac{V\beta}{K(\beta+c_1\beta^{1/2})^2}},$$

we have

$$\sigma_K(M_2) \ge \frac{1}{K(K\alpha+1)}\frac{(1-\delta')V\beta}{(V\beta + c_2\sqrt{V\beta})^2}$$

We can choose c_1, c_2 and δ' as follows to simplify the formula of the bound

- Choose $c_2 = 2\sqrt{\log(K/\delta_1)}$, first probability term is less than δ_1.
- Choose $c_1 = \frac{2}{\sqrt{\beta}}\log(KV/\delta_2)$, third probability term is less than δ_2.
- Choose δ' as
$$\delta' = \left(\frac{\log(K/\delta_3)K\left(\beta + 2\log\left(K/\delta_2\right)\right)^2}{V\beta}\right)^{\frac{1}{2}},$$
second probability term is less than δ_3.

As a result, with probability greater than $1-\delta_1-\delta_2-\delta_3$, we have

$$\sigma_K(M_2) \ge \frac{1}{K(K\alpha+1)}\frac{(1-\delta')V\beta}{(V\beta + 2\sqrt{V\beta}\log(K/\delta_1))^2}$$

As an alternative, we can choose c_1, c_2 and δ_1 as follows to simplify the formula of the bound

- Choose $c_1 = \frac{4}{\sqrt{\beta}}\log(KV)$, third probability term is less than $\frac{1}{KV}$.
- Choose $c_2 = 2\sqrt{\log(K/\delta)}$, first probability term is less than δ.
- Choose $\delta' = 0.1$, second probability term is less than $K(0.995)^{\frac{V(\beta+K\beta^2)}{K(\beta+c_1\beta^{1/2})^2}}$.

As a result, with probability greater than

$$1 - \delta - \frac{1}{KV} - K(0.995)^{\frac{V\beta}{K(\beta+2\log(KV))^2}},$$

we have

$$\sigma_K(M_2) \ge \frac{0.9}{K(K\alpha+1)}\frac{V\beta}{(V\beta + 2\sqrt{V\beta\log K/\delta})^2}.$$

A.2 Lemma for Theorem 2.1

Lemma A.1. *With* $\hat{\mathbf{M}}_2$ *and* $\mathbf{M}_2$ *previously defined, we have that*

$$\max_i |\sigma_i(\hat{\mathbf{M}}_2) - \sigma_i(\mathbf{M}_2)| \leq \max_i |\lambda_i(\hat{\mathbf{M}}_2) - \lambda_i(\mathbf{M}_2)|$$

Proof. Because $\mathbf{M}_2$ is a symmetric semidefinite matrix, so we have

$$\sigma_i(\mathbf{M}_2) = \lambda_i(\mathbf{M}_2), \quad \forall i,$$

And because $\hat{\mathbf{M}}_2$ is a symmetric matrix, we have

$$\sigma_i(\hat{\mathbf{M}}_2) = |\lambda_{s(i)}(\hat{\mathbf{M}}_2)|, \quad \forall i,$$

for some permutation s.

Because we have $\lambda_i(\hat{\mathbf{M}}_2) \leq |\lambda_i(\hat{\mathbf{M}}_2)| = \sigma_j(\hat{\mathbf{M}}_2)$, so we have $\lambda_i(\hat{\mathbf{M}}_2) \leq \sigma_i(\hat{\mathbf{M}}_2)$.

Let j be the smallest index that $|\lambda_j(\hat{\mathbf{M}}_2)| \neq \sigma_j(\hat{\mathbf{M}}_2)$, for $i < j$, we have

$$\begin{aligned}|\sigma_i(\hat{\mathbf{M}}_2) - &\sigma_i(\mathbf{M}_2)| \\ &= |\lambda_i(\hat{\mathbf{M}}_2) - \lambda_i(\mathbf{M}_2)| \\ &\leq \max_i |\lambda_i(\hat{\mathbf{M}}_2) - \lambda_i(\mathbf{M}_2)|\end{aligned}$$

By the fact that $\lambda_i(\mathbf{M}_2) \geq 0$, we have that for $\forall i \geq j$,

$$\sigma_i(\hat{\mathbf{M}}_2) \leq \max_k |\lambda_k(\hat{\mathbf{M}}_2) - \lambda_k(\mathbf{M}_2)|$$

We also have

$$\sigma_i(\hat{\mathbf{M}}_2) \geq \lambda_i(\hat{\mathbf{M}}_2)$$

Because

$$|\lambda_i(\hat{\mathbf{M}}_2) - \sigma_i(\mathbf{M}_2)| \leq \max_k |\lambda_k(\hat{\mathbf{M}}_2) - \lambda_k(\mathbf{M}_2)|$$

We can prove that

$$|\sigma_i(\hat{\mathbf{M}}_2) - \sigma_i(\mathbf{M}_2)| \leq \max_k |\lambda_k(\hat{\mathbf{M}}_2) - \lambda_k(\mathbf{M}_2)|$$

Therefore,

$$\max_i |\sigma_i(\hat{\mathbf{M}}_2) - \sigma_i(\mathbf{M}_2)| \leq \max_i |\lambda_i(\hat{\mathbf{M}}_2) - \lambda_i(\mathbf{M}_2)|$$

□

B Theoretical results for GMM

The proof of Theorem 4.1 is achieved by analyzing the concentration result $\delta_{\mathbf{R}}$ of empirical second order moments and also upper bound for the first singular value of the true moment $\mathbf{M}_2$. Thresholding with $\delta_{\mathbf{R}}$ leads to the first claim, while solving the inequality on the $\sigma_1(\hat{\mathbf{M}}_2)$ provides the second claim.

B.1 Relation Between $\mathbf{M}_2$ and $\hat{\mathbf{M}}_2$

We bound the different between singular values of $\mathbf{M}_2$ through the following Theorem.

Theorem B.1. *For* spherical Gaussian mixtures *with probability at least* $1-\delta$, $\forall i \in \{1, 2, \ldots, m\}$,*we have*

$$|\sigma_i(\hat{\mathbf{M}}_2) - \sigma_i(\mathbf{M}_2)| \leq \frac{\sigma m}{\sqrt{N\delta}}\sqrt{2\sigma_\mu^2 + \frac{m+1}{m}\sigma^2} = \delta_{\mathbf{R}}$$

Especially, when $i \leq K+1$, *we have*

$$\sigma_i(\hat{\mathbf{M}}_2) \leq \frac{\sigma m}{\sqrt{N\delta}}\sqrt{2\sigma_\mu^2 + \frac{m+1}{m}\sigma^2}. \tag{4}$$

Proof. We establish the result by bounding the Frobenius of matrix $\mathbf{R}$ as we do for LDA model. The square of Frobenius norm is $||\mathbf{R}||_{\mathrm{F}}^2 = \sum_{i,j}\mathbf{R}_{ij}^2$. Since we have $\mathbb{E}[\mathbf{R}_{ij}|\mu] = 0$, thus

$$Var[\mathbf{R}_{ij}|\mu] = \mathbb{E}[\mathbf{R}_{ij}^2|\mu] - \mathbb{E}^2[\mathbf{R}_{ij}|\mu] = \mathbb{E}[\mathbf{R}_{ij}^2|\mu],$$

and

$$\begin{aligned}\mathbb{E}[||\mathbf{R}||_{\mathrm{F}}^2] =& \mathbb{E}[\mathbb{E}[||\mathbf{R}||_{\mathrm{F}}^2|\mu]] \\ =& \mathbb{E}[\sum_{i,j} Var[\mathbf{R}_{ij}|\mu]|\mu] \\ =& \mathbb{E}[\sum_{i\neq j} Var[\mathbf{R}_{ij}|\mu] + \sum_i Var[\mathbf{R}_{ii}|\mu]|\mu] \\ =& \frac{m(m-1)}{N}\sigma^2(2\sigma_\mu^2 + \sigma^2) + \frac{m}{N}\sigma^2(2\sigma_\mu^2 + 2\sigma^2) \\ =& \frac{m^2\sigma^2}{N}(2\sigma_\mu^2 + \frac{m+1}{m}\sigma^2).\end{aligned}$$

Then by Markov inequality, we have

$$\Pr(||\mathbf{R}||_{\mathrm{F}}^2 \geq k \times \mathbb{E}[||\mathbf{R}||_{\mathrm{F}}^2]) \leq 1/k.$$

By setting $k = 1/\delta$, we have that with at least probability $1-\delta$,

$$\|\mathbf{R}\|_{\mathrm{F}} \leq \frac{\sigma m}{\sqrt{N\delta}}\sqrt{2\sigma_\mu^2 + \frac{m+1}{m}\sigma^2}$$

□

B.2 Spectral Structure of $\mathbf{M}_2$

We use following theorem to characterize the spectral structure of $\mathbf{M}_2$.

Theorem B.2. *Assume that* $\alpha_i = \alpha$ *in the* spherical Gaussian mixtures, *we have*

(1) With probability at least $1-\delta_1-\delta_2-2\exp(-t^2/2)$, *we have*

$$\sigma_1(\mathbf{M}_2) \leq \frac{\sigma_\mu^2}{K}\frac{\alpha + 2\log(K/\delta_1)}{\alpha - \sqrt{2\alpha\log(1/\delta_2)/K}}(\sqrt{m} + \sqrt{K} + t)^2 \tag{5}$$

(2) Further assume that and $w_i \geq w_{\min}, \forall i$, then with probability at least $1 - 2\exp(-t^2/2)$, we have

$$\sigma_K(\mathbf{M}_2) \geq w_{\min}\sigma_\mu^2(\sqrt{m} - \sqrt{K} - t)^2 \qquad (6)$$

Proof. We have $\mathbf{M}_2 = \sum_{k=1}^K w_k \mu_k \otimes \mu_k = \mathbf{O}\mathbf{A}\mathbf{O}^\top$, where $\mathbf{O} = (\mu_1, \mu_2, \dots, \mu_K)$ is a $m \times K$ matrix and $\mathbf{A} = diag(w_1, w_2, \dots, w_K)$ is a diagonal matrix. Because $\mathbf{M}_2 = \mathbf{O}\mathbf{A}\mathbf{O}^\top = \mathbf{O}\mathbf{A}^{1/2}\mathbf{A}^{1/2}\mathbf{O}^\top$, we have that $\sigma_i(\mathbf{M}_2) = \sigma_i(\mathbf{A}^{1/2}\mathbf{O}^\top\mathbf{O}\mathbf{A}^{1/2}), \forall i = 1, 2, \dots, K$. Therefore, we have the following inequalities [HJ]:

$$\sigma_1(\mathbf{M}_2) \leq \sigma_1(\mathbf{O}^\top\mathbf{O})\sigma_1(\mathbf{A}), \qquad (7)$$
$$\sigma_K(\mathbf{M}_2) \geq \sigma_K(\mathbf{O}^\top\mathbf{O})\sigma_K(\mathbf{A}). \qquad (8)$$

Note that the elements of $\mathbf{O}$ are i.i.d. Gaussian random variables, i.e., $\mathbf{O}_{ij} \sim \mathcal{N}(0, \sigma_\mu^2)$. The distribution of $\sigma_i(\mathbf{O}^\top\mathbf{O})$ has been well-studied in random matrix theory [Ver10]. With probability at least $1 - 2\exp(-t^2/2)$, we have

$$\sigma_1(\mathbf{O}^\top\mathbf{O}) \leq \sigma_\mu^2(\sqrt{m} + \sqrt{K} + t)^2,$$
$$\sigma_K(\mathbf{O}^\top\mathbf{O}) \geq \sigma_\mu^2(\sqrt{m} - \sqrt{K} - t)^2.$$

And since $\sigma_1(\mathbf{A}) = \max_i\{w_i\}$, we can prove that with probability at least $1 - \delta_1 - \delta_2$, we have (see appendix C.3 for proof)

$$\max_i\{w_i\} \leq \frac{1}{K}\frac{\alpha + 2\log(K/\delta_1)}{\alpha - \sqrt{2\alpha\log(1/\delta_2)/K}}$$

We also have $\sigma_K(\mathbf{A}) = \min_i\{w_i\} \geq w_{\min}$. We complete the proof by substituting the above formulas into inequalities (7). ☐

C Tail bound for Gamma distribution

In this section, we proof some tail bound related to the Gamma distribution. Our main tool is the following Lemma.

Lemma C.1. ***[Massart and Laurent] Tail Bound for Chi-square distribution*** *Let U be a χ_D^2 random variable with D degree of freedom, then for any positive x, the following holds*

$$Pr(U \geq D + 2\sqrt{Dx} + 2x) \leq e^{-x},$$
$$Pr(U \leq D - 2\sqrt{Dx}) \leq e^{-x}.$$

Proof. See [LM00] for proof. ☐

C.1 Tail Bound for a Single Gamma Distribution

In this section, we provide tail bound for a single Gamma random variable (R. V.).

Lemma C.2. ***Tail Bound for Gamma R.V.*** *Let $X \sim Gamma(\alpha, 1)$ be a Gamma R.V. with shape parameter α, and scale parameter 1, then for any positive c, the following holds*

$$Pr(X \geq \alpha + c\sqrt{\alpha}) \leq e^{-\frac{c}{2}\min\{\frac{c}{2}, \sqrt{\alpha}\}},$$
$$Pr(X \leq \alpha - c\sqrt{\alpha}) \leq e^{-\frac{c^2}{2}}.$$

Proof. By relationship between Gamma R.V. and chi-square R.V., we have that $2X \sim \chi_{2\alpha}^2$. Apply Lemma C.1 directly, we have

$$\Pr(X \geq \alpha + c\sqrt{\alpha}) \leq e^{-c\sqrt{\alpha} + \alpha(\sqrt{1 + 2c\alpha^{-1/2}} - 1)},$$
$$\Pr(X \leq \alpha - c\sqrt{\alpha}) \leq e^{-\frac{c^2}{2}}.$$

To get the same formula as in the lemma, we can easily prove that $c\sqrt{\alpha} - \alpha(\sqrt{1 + 2c\alpha^{-1/2}} - 1) > \frac{c}{2}\min\{\frac{c}{2}, \sqrt{\alpha}\}, \quad \forall c, \alpha > 0$.

☐

Corollary C.3. ***Tail Bound for Sum of Square of Gamma R.V.*** *If we have n i.i.d Gamma R.V. $X_i \sim Gamma(\alpha, 1), i = 1, \dots, n$, then for any positive c, the following holds*

$$Pr(\sum_i X_i^2 \geq n(\alpha + c\sqrt{\alpha})^2) \leq ne^{-\frac{c}{2}\min\{\frac{c}{2}, \sqrt{\alpha}\}}.$$

C.2 Tail Bound for Maximum/Minimum of Gamma Random Variables

Lemma C.4. *If we have n i.i.d Gamma R.V. $X_i \sim Gamma(\alpha, 1), i = 1, \dots, n$, we have that*

$$Pr(\max_i\{X_i\} \geq \alpha + c\sqrt{\alpha}) \leq ne^{-\frac{c}{2}\min\{\frac{c}{2}, \sqrt{\alpha}\}},$$
$$Pr(\min_i\{X_i\} \leq \alpha - c\sqrt{\alpha}) \leq ne^{-\frac{c^2}{2}}.$$

Proof. It can be proved by applying union bound directly. ☐

C.3 Tail Bound for Maximum/Minimum Element of Dirichlet Distribution

It is well known that a random vector $(x_1, x_2, \dots, x_n) \sim \text{Dir}(\alpha_1, \alpha_2, \dots, \alpha_n)$ is equivalent to a random vector $(y_1, y_2, \dots, y_n)/\sum_i y_i$, where $y_i \sim \text{Gamma}(\alpha_i, 1)$ independently. And we have $\max_i\{x_i\} = \max_i\{y_i\}/\sum_i y_i$.

Assume $\alpha_i = \alpha$, so we have

$$\Pr(\max_i\{y_i\} \geq \alpha + c_1\sqrt{\alpha}) \leq ne^{-\frac{c_1}{2}\min\{\frac{c_1}{2}, \sqrt{\alpha}\}}.$$

And since $\sum_i y_i \sim \text{Gamma}(n\alpha, 1)$, we have

$$\Pr(\sum_i y_i \leq n\alpha - c_2\sqrt{n\alpha}) \leq e^{-\frac{c_2^2}{2}}$$

By setting $c_1 = 2\log(n/\delta_1)/\sqrt{\alpha}$ (when $n > \delta_1 e^{\alpha}$) and $c_2 = \sqrt{2\log(1/\delta_2)}$, we have that with probability at least $1 - \delta_1 - \delta_2$,

$$\max_i\{x_i\} \leq \frac{1}{n}\frac{\alpha + \log(n/\delta_1)}{\alpha - \sqrt{2\alpha\log(1/\delta_2)/n}}$$

Similarity, $\min_i\{x_i\} = \min_i\{y_i\}/\sum_i y_i$. And

$$\begin{aligned}\Pr(\min_i\{x_i\} \leq \alpha - c_1\sqrt{\alpha}) \leq& ne^{-\frac{c_1^2}{2}},\\ \Pr(\sum_i y_i \geq n\alpha + c_2\sqrt{n\alpha}) \leq& e^{-\frac{c_2}{2}\min\{\frac{c_2}{2},\sqrt{n\alpha}\}}.\end{aligned}$$

By setting $c_1 = \sqrt{2\log(n/\delta_1)}$ and $c_2 = \sqrt{2\log(1/\delta_2)}$ (when $\delta_2 > e^(-2\alpha)$), we have that with probability at least $1 - \delta_1 - \delta_2$,

$$\min_i\{x_i\} \geq \frac{1}{n}\frac{\alpha - \sqrt{2\log(n\alpha/\delta_1)}}{\alpha + \sqrt{2\alpha\log(1/\delta_2)/n}}$$

which is nontrivial only when α is large enough.

D Variance Calculation for LDA

In this section, we presents the overall procedure and some important intermediate results of the variance calculation for LDA. Note that we have the following assumptions on the scale of each statistics or parameters: $L = \mathcal{O}(D)$, $V = \mathcal{O}(D)$, $L = \mathcal{O}(V)$, $K = \mathcal{O}(L)$, $1/K = \mathcal{O}(1)$, $\alpha = \Theta(1)$, and $\beta = \Theta(1)$.

First, we have

$$\begin{aligned}R =& \frac{1}{D}\sum_d \frac{1}{L(L-1)}\sum_{l\neq s} x_{d,l}x_{d,s}^\top\\ &- \frac{\alpha_0}{\alpha_0+1}[\frac{1}{D}\sum_d\frac{1}{L}\sum_l x_{d,l}][\frac{1}{D}\sum_d\frac{1}{L}\sum_l x_{d,l}]^\top\\ &- M_2.\end{aligned}$$

We represent each term by

$$\begin{aligned}R^{(1)} =& \frac{1}{D}\sum_d \frac{1}{L(L-1)}\sum_{l\neq s} x_{d,l}x_{d,s}^\top,\\ R^{(2)} =& \frac{\alpha_0}{\alpha_0+1}[\frac{1}{D}\sum_d\frac{1}{L}\sum_l x_{d,l}][\frac{1}{D}\sum_d\frac{1}{L}\sum_l x_{d,l}]^\top,\\ R^{(3)} =& \frac{1}{D}\sum_d\frac{1}{L}\sum_l x_{d,l}.\end{aligned}$$

And we have the following identity:

$$\begin{aligned}E_\mu Var_X[R_{ij}] =& E_\mu Var_X[R_{ij}^{(1)}] + E_\mu Var_X[R_{ij}^{(2)}]\\ &- 2E_\mu Cov_X[R_{ij}^{(1)}, R_{ij}^{(2)}],\end{aligned}$$

with $H = \{\mu, h\}$, $X = \{h, x\}$.

$$R_{ij}^{(2)} = \frac{\alpha_0}{\alpha_0+1}R_i^{(3)}R_j^{(3)}.$$

For simplicity of representation, we assume the following,

$$\begin{aligned}f_d^{(ij)} =& \frac{1}{L(L-1)}\sum_{l\neq s}^{L} x_{d,l}^{(i)}x_{d,s}^{(j)},\\ g_d^{(i)} =& \frac{1}{L}\sum_{l=1}^{L} x_{d,l}^{(i)}.\end{aligned}$$

and the superscript (ij) or (i) will be omitted if there is no ambiguity. By this representation, we have

$$\begin{aligned}R^{(1)} =& \frac{1}{D}\sum_d f_d,\\ R^{(3)} =& \frac{1}{D}\sum_d g_d.\end{aligned}$$

We also assume the representation $z_d^{(i)} = \sum_k \mu_k^{(i)}h_d^{(k)}$, which is the probability of e_i in the d-th documents conditioned on $H = \{\mu, h\}$. And $\delta_{ij} = 1$ if and only if $i = j$.

The intermediate results for diagonal and off-diagonal variance are different, so we provide them separately in the following sections.

D.1 Calculate Off-diagonal Variance

In this section, we assume that $i \neq j$. And we have the following results:

$$\begin{aligned}E_\mu Var_X[R_{ij}^{(1)}] \leq& \frac{1}{DL^2V^2} + \frac{2}{DLV^3} + \frac{1}{DV^4} + O(\epsilon)\\ E_\mu Var_X[R_{ij}^{(2)}] \leq& \frac{2}{DLV^3} + \frac{1}{DV^4} + O(\epsilon)\\ E_\mu Cov_X(R^{(1)}, R^{(2)}) \geq& \frac{2}{DLV^3} + O(\epsilon)\end{aligned}$$

Therefore, we have that

$$\begin{aligned}E_\mu Var_X[R_{ij}] \leq& \frac{1}{DL^2V^2} + \frac{2}{DLV^3} + \frac{1}{DV^4} + \frac{2}{DLV^3}\\ &+ \frac{1}{DV^4} - \frac{4}{DLV^3} + O(\epsilon)\\ =& \frac{1}{DL^2V^2} + \frac{2}{DV^4} + O(\epsilon).\end{aligned}$$

D.2 Calculate Diagonal Variance

In this section, we assume that $i \neq j$. And we have the following results:

$$
\begin{aligned}
E_\mu Var_X[R_{ij}^{(1)}] \leq & \frac{1}{DL^2V} + \frac{4}{DLV^3} + \frac{1}{DV^4} + O(\epsilon), \\
E_\mu Var_X[R_{ij}^{(2)}] \leq & \frac{2}{DLV^3} + \frac{1}{DV^4} + O(\epsilon), \\
E_\mu Cov_X(R^{(1)}, R^{(2)}) \geq & \frac{3}{DLV^3} + O(\epsilon).
\end{aligned}
$$

Therefore, we have that

$$
\begin{aligned}
E_\mu Var_X[R_{ij}] \leq & \frac{1}{DL^2V} + \frac{4}{DLV^3} + \frac{1}{DV^4} + \frac{2}{DLV^3} \\
& + \frac{1}{DV^4} - \frac{6}{DLV^3} + O(\epsilon) \\
= & \frac{1}{DL^2V} + \frac{2}{DV^4} + O(\epsilon).
\end{aligned}
$$