Appendix: An end-to-end Differentially Private Latent Dirichlet Allocation Using a Spectral Algorithm

A. Differential Privacy Review

Differential privacy was developed in (Dwork et al., 2006) and has been increasingly adopted as the *de facto* mathematical definition for privacy in statistical, machine learning and data science applications. We include additional information in this section that is relevant to this paper, but will defer more exposition to recent book (Dwork et al., 2014a) and the references therein.

Definition 21 (Utility Loss & Error). *Let $f : D \to Y$ be a random algorithm and $f^{\mathsf{DP}}(X)$ be the differentially private version of f. For some value $x \in D$, let $y \in Y$ be the ground truth value. Then define $\left\|f(x) - f^{\mathsf{DP}}(X)\right\|_F$ as the **utility loss** for this input. Additionally, define $\left\|y - f^{\mathsf{DP}}(X)\right\|_F$ as the **error** for this input.*

The gaussian mechanism makes a random algorithm differentially private by adding specifically designed Gaussian noise to the output.

Proposition 22. ***[Gaussian mechanism]*** *Let $f : D \to Y$ ($Y \subset \mathbb{R}^k$) be a random algorithm with ℓ_2 sensitivity Δ_f. Let $g \in \mathbb{R}^k$ and each coordinate g_i be sampled i.i.d. from $\mathcal{N}(0, \Delta^2_{f,\epsilon,\delta})$, where $\Delta_{f,\epsilon,\delta} = \Delta_f\ \tau_{\epsilon,\delta} = \frac{\Delta_f\sqrt{2\ln(1.25/\delta)}}{\epsilon}$. Then the output $f_{\mathsf{DP}} = f + g$ is (ϵ, δ) differentially private if $0 < \epsilon \leq 1$.*

The above bound is used for theoretical purposes only, a tighter and more general calibration of the Gaussian mechanism that does not require $\epsilon \leq 1$ is to set

$$\sigma = \frac{\Delta_f}{2\epsilon}\left(\sqrt{\epsilon + \log(1/\delta)} + \sqrt{\log(1/\delta)}\right).$$

Moreover, the optimal calibration (no closed-form formula available) was proposed in (Balle & Wang, 2018) and is available through, e.g., *autodp.calibrator*: `https://github.com/yuxiangw/autodp`.

Differential privacy composes over multiple DP releases.

Proposition 23. ***[Composition theorem]*** *Let $f_1^{\mathsf{DP}}(X), \ldots, f_n^{\mathsf{DP}}(X)$ be n differentially private algorithms with privacy parameters $(\epsilon_1, \delta_1), \ldots, (\epsilon_n, \delta_n)$. Then $g^{\mathsf{DP}}(X) = f(f_1^{\mathsf{DP}}(X), \ldots, f_n^{\mathsf{DP}}(X))$ is $(\epsilon_1 + \ldots + \epsilon_n, \delta_1 + \ldots + \delta_n)$ differentially private.*

This is what we called a simple composition where epsilon increases linearly. There is also an advanced composition where privacy loss for accessing for k times obey that $\sqrt{k}$, see, e.g., Section 3.5 of (Dwork et al., 2014a). Increasingly, the advanced composition and other privacy loss computation has been conducted numerically using modern tools such as Concentrated Differential privacy (Bun & Steinke, 2016) and Renyi Differential Privacy (or equivalently the moments accountant) (Mironov, 2017). We used simple composition in our theoretical analysis and for calibrating noise to privacy so as to be comparable to older literature that does not take advantage of the modern tool (Park et al., 2016; Wang & Anandkumar, 2016).

B. Latent Dirichlet Allocation

LDA, despite being a bag of words model, allows modeling of the mixed topics in a document to account for the more general case in which a document belongs to several different latent classes (topics) simultaneously. Latent Dirichlet Allocations has two major model parameters: topic prior $\boldsymbol{\alpha}$ and topic-word matrix $\boldsymbol{\mu}$. Topic prior $\boldsymbol{\alpha}$ determines the topic proportions and the topic word matrix controls the word distribution per topic.

Topic Proportions The proportion of words in topics, known as *topic proportion* (denoted as θ_n for document n), is drawn from a Dirichlet distribution (topic prior) parameterized by $\alpha = (\alpha_1, \ldots, \alpha_k)$, with density $P_\alpha(\theta = \theta_n) = \frac{\Gamma(\alpha_0)}{\prod_{i=1}^{k}\Gamma(\alpha_i)}\prod_{i=1}^{k}\theta_{n,i}^{\alpha_k - 1}$, where $\alpha_0 = \sum_{i=1}^{k}\alpha_i$.

Topic-Word Matrix Under a topic i, tokens in the documents are assumed to be generated in a conditionally independent manner through μ_i, i.e., token $x_1 \sim \text{Cat}(d, \mu_i)$ where $\text{Cat}(d, \mu_i)$ denotes the categorical distribution. Under different topics, these conditional distributions μ_i are linearly independent, $\forall i \in [k]$.

With the definition of the two major parameters, we now describe the generative model of LDA topic model. The process involves generating topics first, followed by tokens.

Topic Generation LDA remains simple as each token in the corpus belongs to one of the k topics only, although tokens in the same document could belong to different topics. We denote the topic of token j in document n as $z_{n,j}$. Therefore, topics generated are categorical $z_{n,j} \in [k]$ and distributed according to θ_n, i.e., $z_{n,j} \sim \text{Cat}(k, \theta_n)$ where $\text{Cat}(k, \theta_n)$ denotes the categorical distribution.

Word Generation Let x denote the tokens. After determining the topic of the token j, $z_{n,j}$, token j is generated conditionally independently through $\mu_{z_{n,j}}$, i.e., token $\sim \text{Cat}(d, \mu_{z_{n,j}})$. In a document n, if the j'^{th} token $x_{n,j'}$ is the v-th word in the dictionary, then $x_{n,j'} = e_v$ where e_v is a one-hot encoding, i.e., $x_{n,j'}(j) = 0 \; \forall j \neq v$ and $x_{n,j'}(j) = 1$ if $j = v$. Let l_n be the length of document n, random realizations of token x, i.e., $\{x_{n,j'}\}_{j'=1}^{l_n}$, are i.i.d.

Term-Document Matrix The term-document matrix $D \in \mathbb{N}_0^{d \times N}$. The n^{th} column in D is denoted by c_n, where its j^{th} component $c_n(j) =$ number of times word j in the vocabulary appeared in document n. This means that $c_n = \sum_{j'=1}^{l_n} x_{n,j'}$ where l_n is the number of words in document n. Clearly, $l_n = \sum_j^d c_n(j) = \|c_n\|_1$.

C. Method of Moments for Latent Dirichlet Allocation

Empirical Moment Estimators The moments that we obtain are not the population moments but rather empirically estimated moments from the given data set. We list the forms of first, second, and third order empirical moment estimators for the single topic case as shown in (Zou et al., 2013). Given a document n, the following quantities are calculated.

$$\tilde{\hat{M}}_1^n = \frac{c_n}{l_n} \tag{6}$$

$$\tilde{\hat{M}}_2^n = \frac{1}{2\binom{l_n}{2}}(c_n \otimes c_n - \text{diag}(c_n)) \tag{7}$$

$$\begin{aligned}\tilde{\hat{M}}_3^n = \frac{1}{6\binom{l_n}{3}}\Big(&c_n \otimes c_n \otimes c_n + 2\sum_{i=1}^{d} c_n(i)(e_i \otimes e_i \otimes e_i) \\ &- \sum_{i=1}^{d}\sum_{j=1}^{d} c_n(i)c_n(j)(e_i \otimes e_i \otimes e_j + e_i \otimes e_j \otimes e_j + e_j \otimes e_i \otimes e_j)\Big)\end{aligned} \tag{8}$$

The empirically estimated moments are the averages of these quantities over the entire data set. Specifically,

Lemma 24. *Single Topic Empirical Moment Estimators(Propositions 3 and 4 in (Zou et al., 2013))*

$$\hat{\mathbb{E}}[x_1] = \frac{1}{N}\sum_{n=1}^{N} \tilde{\hat{M}}_1^n \tag{9}$$

$$\hat{\mathbb{E}}[x_1 \otimes x_2] = \frac{1}{N}\sum_{n=1}^{N} \tilde{\hat{M}}_2^n \tag{10}$$

$$\hat{\mathbb{E}}[x_1 \otimes x_2 \otimes x_3] = \frac{1}{N}\sum_{n=1}^{N} \tilde{\hat{M}}_3^n \tag{11}$$

$$\tag{12}$$

Further these moments are unbiased, i.e.:

$$\mathbb{E}[\hat{\mathbb{E}}[x_1]] = \mathbb{E}[\frac{1}{N}\sum_{n=1}^{N}\tilde{\tilde{M}}_1^n] = \mathbb{E}[x_1] \tag{13}$$

$$\mathbb{E}[\hat{\mathbb{E}}[x_1 \otimes x_2]] = \mathbb{E}[\frac{1}{N}\sum_{n=1}^{N}\tilde{\tilde{M}}_2^n] = \mathbb{E}[x_1 \otimes x_2] \tag{14}$$

$$\mathbb{E}[\hat{\mathbb{E}}[x_1 \otimes x_2 \otimes x_3]] = \mathbb{E}[\frac{1}{N}\sum_{n=1}^{N}\tilde{\tilde{M}}_3^n] = \mathbb{E}[x_1 \otimes x_2 \otimes x_3] \tag{15}$$

$$\tag{16}$$

Note that this lemma implies that: $\mathbb{E}[\tilde{\tilde{M}}_1^n] = \mathbb{E}[x_1], \mathbb{E}[\tilde{\tilde{M}}_2^n] = \mathbb{E}[x_1 \otimes x_2]$, and that $\mathbb{E}[\tilde{\tilde{M}}_3^n] = \mathbb{E}[x_1 \otimes x_2 \otimes x_3]$ for any sampled document n.

We extend the single topic moment estimators of (Zou et al., 2013) to the LDA case.

Lemma 25. *Empirical Moment estimators for LDA*

$$\hat{M}_1 = \frac{1}{N}\sum_{n=1}^{N}\tilde{\tilde{M}}_1^n \tag{17}$$

$$\hat{M}_2 = \frac{1}{N}\sum_{n=1}^{N}\left[\tilde{\tilde{M}}_2^n\right] - \frac{a}{2\binom{N}{2}}\left[\sum_{m,n=1}^{N}\tilde{\tilde{M}}_1^n \otimes \tilde{\tilde{M}}_1^m - \sum_{n=1}^{N}\tilde{\tilde{M}}_1^n \otimes \tilde{\tilde{M}}_1^n\right] \tag{18}$$

$$\hat{M}_3 = \left[\frac{1}{N}\sum_{n=1}^{N}\tilde{\tilde{M}}_3^n + \mathbf{B}_1 + \mathbf{B}_2 + \mathbf{B}_3 + \mathbf{b}\right] \tag{19}$$

where

$$\mathbf{B}_1 \overset{\text{def}}{=} \frac{b}{2\binom{N}{2}}\left[\Big(\sum_{n=1}^{N}\tilde{\tilde{M}}_2^n\Big) \otimes \Big(\sum_{n=1}^{N}\tilde{\tilde{M}}_1^n\Big)\right], \tag{20}$$

$$\mathbf{b} \overset{\text{def}}{=} c\left[\Big(\sum_{n=1}^{N}\tilde{\tilde{M}}_1^n\Big) \otimes \Big(\sum_{n=1}^{N}\tilde{\tilde{M}}_1^n\Big) \otimes \Big(\sum_{n=1}^{N}\tilde{\tilde{M}}_1^n\Big)\right], \tag{21}$$

$\mathbf{B}_2$ *and* $\mathbf{B}_3$ *are formed from* $\mathbf{B}_1$ *by permuting, i.e.,* $[\mathbf{B}_2]_{ijk} = [\mathbf{B}_1]_{ikj}$ *and* $[\mathbf{B}_3]_{ijk} = [\mathbf{B}_1]_{kij}$*. Further,* $a = \frac{\alpha_0}{\alpha_0+1}, b = \frac{-\alpha_0}{\alpha_0+2}, c = \frac{2\alpha_0^2}{(\alpha_0+1)(\alpha_0+2)}$.

Now we prove that these estimators are unbiased.

Lemma 26 (The LDA Moment Estimators are Unbiased)**.** *The estimators defined in definition 25 are unbiased, i.e.,*

$$\mathbb{E}[\hat{M}_1] = M_1 \tag{22}$$

$$\mathbb{E}[\hat{M}_2] = M_2 \tag{23}$$

$$\mathbb{E}[\hat{M}_3] = M_3 \tag{24}$$

Proof. **First order moment**:

$$\mathbb{E}[\hat{M}_1] = \mathbb{E}[\frac{1}{N}\sum_{n=1}^{N}\tilde{\tilde{M}}_1^n] = \frac{1}{N}\sum_{n=1}^{N}\mathbb{E}[\tilde{\tilde{M}}_1^n] = \frac{1}{N}\sum_{n=1}^{N}\mathbb{E}[\frac{c_n}{l_n}] \tag{25}$$

$$= \frac{1}{N}\sum_{n=1}^{N}\frac{1}{l_n}\mathbb{E}[c_n] = \frac{1}{N}\sum_{n=1}^{N}\frac{1}{l_n}\mathbb{E}[\sum_{i=1}^{l_n}x_{n,i}] = \frac{1}{N}\sum_{n=1}^{N}\frac{1}{l_n}\sum_{i=1}^{l_n}\mathbb{E}[x_{n,i}] \tag{26}$$

$$= \frac{1}{N}\sum_{n=1}^{N}\frac{1}{l_n}\sum_{i=1}^{l_n}\mathbb{E}[x_1] = \frac{1}{N}\sum_{n=1}^{N}\frac{1}{l_n}l_n\mathbb{E}[x_1] = \frac{1}{N}N\mathbb{E}[x_1] = \mathbb{E}[x_1] = M_1 \tag{27}$$

Second order moment: The first term of $\hat{M}_2$ is actually the estimator the single-topic second order moment and $\mathbb{E}[\frac{1}{N}\sum_{n=1}^{N}\tilde{\tilde{M}}_2^n]] = \mathbb{E}[x_1 \otimes x_2]$ see proposition 3 in (Zou et al., 2013) and its appendix for the proof. Now we have:

$$\mathbb{E}\left[\frac{a}{2\binom{N}{2}}\left[\sum_{m,n=1}^{N}\tilde{\tilde{M}}_1^n \otimes \tilde{\tilde{M}}_1^m - \sum_{n=1}^{N}\tilde{\tilde{M}}_1^n \otimes \tilde{\tilde{M}}_1^n\right]\right] \tag{28}$$

$$=\mathbb{E}\left[\frac{a}{2\binom{N}{2}}\left[\sum_{\substack{m=1\\n=1\\m\neq n}}^{N}\tilde{\tilde{M}}_1^n \otimes \tilde{\tilde{M}}_1^m + \sum_{n=1}^{N}\tilde{\tilde{M}}_1^n \otimes \tilde{\tilde{M}}_1^n - \sum_{n=1}^{N}\tilde{\tilde{M}}_1^n \otimes \tilde{\tilde{M}}_1^n\right]\right] \tag{29}$$

$$=\mathbb{E}\left[\frac{a}{2\binom{N}{2}}\sum_{\substack{m=1\\n=1\\m\neq n}}^{N}\tilde{\tilde{M}}_1^n \otimes \tilde{\tilde{M}}_1^m\right] \tag{30}$$

$$=\frac{a}{2\binom{N}{2}}\sum_{\substack{m=1\\n=1\\m\neq n}}^{N}\mathbb{E}[\tilde{\tilde{M}}_1^n] \otimes \mathbb{E}[\tilde{\tilde{M}}_1^n] = \frac{a}{2\binom{N}{2}}\sum_{\substack{m=1\\n=1\\m\neq n}}^{N}\mathbb{E}[x_1] \otimes \mathbb{E}[x_1] \tag{31}$$

$$=\frac{a}{N(N-1)}N(N-1)\mathbb{E}[x_1] \otimes \mathbb{E}[x_1] = a\mathbb{E}[x_1] \otimes \mathbb{E}[x_1] \tag{32}$$

Thus, we have that: $\mathbb{E}[\hat{M}_2] = \mathbb{E}[x_1 \otimes x_2] - \frac{\alpha_0}{\alpha_0+2}\mathbb{E}[x_1] \otimes \mathbb{E}[x_1] = M_2$.

Third order moment: Similar to the second order moment, the first term of $\hat{M}_3$ is the estimator the single-topic second order moment and $\mathbb{E}[\frac{1}{N}\sum_{n=1}^{N}\tilde{\tilde{M}}_3^n]] = \mathbb{E}[x_1 \otimes x_2 \otimes x_3]$ as shown in proposition 4 in (Zou et al., 2013) and proved in its appendix. We need to prove that (1): $\mathbb{E}[\mathbf{B}_1] = b\mathbb{E}[x_1 \otimes x_2 \otimes \mathbb{E}[x_3]]$, note that $\mathbb{E}[x_3] = M_1$ and (2): $\mathbb{E}[\mathbf{b}] = cE[x_1] \otimes E[x_1] \otimes E[x_1] = cM_1 \otimes M_1 \otimes M_1 \otimes M_1$. Since $\mathbf{B}_2$ and $\mathbf{B}_3$ are permuted version of $\mathbf{B}_1$ their proofs follow from the proof of $\mathbf{B}_1$.

For $\mathbf{B}_1$ we simplify the expression and then show that the expectation of the resultant is equal to the desired moment:

$$\mathbb{E}[\mathbf{B}_1] =\frac{b}{2\binom{N}{2}}\mathbb{E}\left[\left(\sum_{n=1}^{N}\tilde{\tilde{M}}_2^n\right) \otimes \left(\sum_{n=1}^{N}\tilde{\tilde{M}}_1^n\right) - \sum_{n=1}^{N}\left(\tilde{\tilde{M}}_2^n \otimes \tilde{\tilde{M}}_1^n\right)\right] \tag{33}$$

$$=\frac{b}{2\binom{N}{2}}\mathbb{E}\left[\sum_{\substack{m=1,n=1\\m\neq n}}^{N}\left(\tilde{\tilde{M}}_2^n \otimes \tilde{\tilde{M}}_1^m\right) + \sum_{n=1}^{N}\left(\tilde{\tilde{M}}_2^n \otimes \tilde{\tilde{M}}_1^n\right) - \sum_{n=1}^{N}\left(\tilde{\tilde{M}}_2^n \otimes \tilde{\tilde{M}}_1^n\right)\right] \tag{34}$$

$$=\frac{b}{2\binom{N}{2}}\mathbb{E}\left[\sum_{\substack{m=1,n=1\\m\neq n}}^{N}\left(\tilde{\tilde{M}}_2^n \otimes \tilde{\tilde{M}}_1^m\right)\right] \tag{35}$$

$$=\frac{b}{2\binom{N}{2}}\sum_{\substack{m=1,n=1\\m\neq n}}^{N}\mathbb{E}\left[\tilde{\tilde{M}}_2^n\right] \otimes \mathbb{E}\left[\tilde{\tilde{M}}_1^m\right] \tag{36}$$

$$=\frac{b}{2\binom{N}{2}}\sum_{\substack{m=1,n=1\\m\neq n}}^{N}\mathbb{E}[x_1 \otimes x_2] \otimes \mathbb{E}[x_3] \tag{37}$$

$$=\frac{b}{N(N-1)}N(N-1)\mathbb{E}[x_1 \otimes x_2] \otimes \mathbb{E}[x_3] \tag{38}$$

$$=b\mathbb{E}[x_1 \otimes x_2] \otimes \mathbb{E}[x_3] \tag{39}$$

$$=b\mathbb{E}\left[x_1 \otimes x_2 \otimes \mathbb{E}[x_3]\right] \tag{40}$$

For $\mathbf{b}$ identity 38 is applied, this leads to the following

$$\mathbb{E}[\mathbf{b}] = \frac{c}{6\binom{N}{3}}\mathbb{E}\Bigg[\Bigg(\sum_{i=1}^{N}(\tilde{\tilde{M}}_1^n)^{\otimes^3} + 3\sum_{\substack{n=1,m=1\\n\neq m}}^{N,N}(\tilde{\tilde{M}}_1^n)^{\otimes^2}\tilde{\tilde{M}}_1^m + \sum_{\substack{n=1,m=1,p=1\\n\neq m,m\neq p,p\neq n}}^{N,N,N}\tilde{\tilde{M}}_1^n\otimes\tilde{\tilde{M}}_1^m\otimes\tilde{\tilde{M}}_1^p \tag{41}$$

$$- 3\sum_{m=1}^{N}\Big(\sum_{n=1}^{N}\Big(\tilde{\tilde{M}}_1^n\Big)^{\otimes^2}\otimes\Big(\tilde{\tilde{M}}_1^m\Big)\Big) + 2\sum_{n=1}^{N}\Big(\tilde{\tilde{M}}_1^n\Big)^{\otimes^3}\Bigg)\Bigg] \tag{42}$$

$$= \frac{c}{6\binom{N}{3}}\mathbb{E}\Bigg[\sum_{\substack{n=1,m=1,p=1\\n\neq m,m\neq p,p\neq n}}^{N,N,N}\tilde{\tilde{M}}_1^n\otimes\tilde{\tilde{M}}_1^m\otimes\tilde{\tilde{M}}_1^p\Bigg] \tag{43}$$

$$= \frac{c}{N(N-1)(N-2)}(N)(N-1)(N-2)\mathbb{E}[\tilde{\tilde{M}}_1^n]\otimes\mathbb{E}[\tilde{\tilde{M}}_1^m]\otimes\mathbb{E}[\tilde{\tilde{M}}_1^p] \tag{44}$$

$$= c\mathbb{E}[x_1]\otimes\mathbb{E}[x_1]\otimes\mathbb{E}[x_1] \tag{45}$$

Combing these results and plugging the values for a, b,and c we get:

$$\mathbb{E}[\hat{M}_3] = \mathbb{E}[x_1\otimes x_2\otimes x_3] - \frac{\alpha_0}{\alpha_0+2}\Bigg(\mathbb{E}[x_1\otimes x_2\otimes\mathbb{E}[x_3]] + \mathbb{E}[x_1\otimes\mathbb{E}[x_2]\otimes x_3] + \mathbb{E}[\mathbb{E}[x_1]\otimes x_2\otimes x_3]\Bigg) \tag{46}$$

$$+\frac{2\alpha_0^2}{(\alpha_0+1)(\alpha_0+2)}\mathbb{E}[x_1]\otimes\mathbb{E}[x_1]\otimes\mathbb{E}[x_1] = M_3 \tag{47}$$

□

D. Lemmas regarding Dirichlet Moments

This section introduces two lemmas regarding the moments of the dirichlet distribution that will be useful for the proof of Lemma 3.

D.1. Dirichlet Moments

Lemma 27. *The first, second and third moments of dirichlet distribution are*

$$\mathbb{E}[\theta] = \frac{1}{\alpha_0}\alpha \tag{48}$$

$$\mathbb{E}[\theta\otimes\theta] = \frac{1}{\alpha_0(\alpha_0+1)}[\alpha\otimes\alpha + \sum_{t=1}^{T}\alpha_t e_t\otimes e_t] \tag{49}$$

$$\mathbb{E}[\theta\otimes\theta\otimes\theta] = \frac{1}{\alpha_0(\alpha_0+1)(\alpha_0+2)}[\alpha\otimes\alpha\otimes\alpha + \sum_{t=1}^{T}\alpha_t e_t\otimes e_t\otimes\alpha$$
$$+\sum_{t=1}^{T}\alpha_t\alpha\otimes e_t\otimes e_t + \sum_{t=1}^{T}\alpha_t e_t\otimes\alpha\otimes e_t + 2\sum_{t=1}^{T}\alpha_t e_t\otimes e_t\otimes e_t] \tag{50}$$

D.2. Raw Moments

Lemma 28.

$$\mathbb{E}[x_1] = \mu\mathbb{E}[\theta] \tag{51}$$

$$\mathbb{E}[x_1\otimes x_2] = \mu\mathbb{E}[\theta\otimes\theta]\mu^\top \tag{52}$$

$$\mathbb{E}[x_1\otimes x_2\otimes x_3] = \mathbb{E}[\theta\otimes\theta\otimes\theta](\mu,\mu,\mu) \tag{53}$$

Proof. **First Order Moments** Let us omit n and use x_1 to denote a token in any document, and we will use x_2 and x_3 to denote other two tokens in the same document. The the expectation of a token is

$$\mathbb{E}[x_1] = \mathbb{E}[x_2] = \mathbb{E}[x_3] = \mathbb{E}[\mathbb{E}[x_1|\theta]] = \mu\mathbb{E}[\theta] \tag{54}$$

This is called the first order moment.

Second Order Moments The second order moment is defined as

$$\mathbb{E}[x_1 \otimes x_2] = \mathbb{E}[\mathbb{E}[x_1 \otimes x_2|\theta]] \tag{55}$$
$$= \sum_{i,i'} \mathbb{E}[x_1 \otimes x_2|z_{n,j} = e_i, z_{n,k} = e_{i'}]P(z_{n,j} = e_i, z_{n,k} = e_{i'}) \tag{56}$$
$$= \sum_{i,i'} \mathbb{E}[x_1|z_{n,j} = e_i] \otimes \mathbb{E}[x_2|z_{n,k} = e_{i'}]P(z_{n,j} = e_i, z_{n,k} = e_{i'}) \tag{57}$$
$$= \sum_{i,i'} \mu e_i \otimes (\mu e_{i'})P(z_{n,j} = e_i, z_{n,k} = e_{i'}) \tag{58}$$
$$= \mu \sum_{i,i'} e_i \otimes e_{i'} P(z_{n,j} = e_i, z_{n,k} = e_{i'})\mu^\top \tag{59}$$
$$= \mu \mathbb{E}[\theta \otimes \theta]\mu^\top \tag{60}$$

Third Order Moments The third order moment is defined as

$$\mathbb{E}[x_1 \otimes x_2 \otimes x_3] = \mathbb{E}[\mathbb{E}[x_1 \otimes x_2 \otimes x_3|\theta]] = \mathbb{E}[\theta \otimes \theta \otimes \theta](\mu, \mu, \mu) \tag{61}$$

To clarify the notations, $x \otimes y$ is a $length(x)$-by-$length(y)$ matrix which has entries $[x \otimes y]_{i,j} = x_i y_j$. And $\mathbb{E}[\theta \otimes \theta \otimes \theta](\mu, \mu, \mu)$ is a tucker with core tensor $\mathbb{E}[\theta \otimes \theta \otimes \theta]$ and projection μ in all three modes. □

E. Proof of Lemma 3

The lemma relates the LDA moments to the model parameters α and μ.

Proof. In order to prove this relation, we combine Lemmas 27 and Lemma 28 to prove the forms of M_1, M_2 and M_3 in Lemma 3 as follows.

$$M_1 = \mathbb{E}[x_1] = \mu \mathbb{E}[\theta] = \sum_{i=1}^{k} \frac{\alpha_i}{\alpha_0} \mu_i \tag{62}$$
$$M_2 = \mathbb{E}[x_1 \otimes x_2] - \frac{\alpha_0}{\alpha_0 + 1} \mathbb{E}[x_1] \otimes \mathbb{E}[x_1] \tag{63}$$
$$= \mathbb{E}[\theta \otimes \theta](\mu, \mu) - \frac{1}{\alpha_0(\alpha_0 + 1)} M_1 \otimes M_1 \tag{64}$$
$$= \sum_{i=1}^{k} \frac{\alpha_i}{\alpha_0(\alpha_0 + 1)} \mu_i \otimes \mu_i \tag{65}$$
$$M_3 = \mathbb{E}[x_1 \otimes x_2 \otimes x_3] - \frac{1}{\alpha_0 + 2}(\mathbb{E}[x_1 \otimes x_2 \otimes \mathbb{E}[x_3]] + \mathbb{E}[x_1 \otimes \mathbb{E}[x_2] \otimes x_3]$$
$$+ \mathbb{E}[\mathbb{E}[x_1] \otimes x_2 \otimes x_3]) + \frac{2}{\alpha_0(\alpha_0 + 1)(\alpha_0 + 2)} \mathbb{E}[x_1] \otimes \mathbb{E}[x_1] \otimes \mathbb{E}[x_1] \tag{66}$$
$$= \mathbb{E}[\theta \otimes \theta \otimes \theta](\mu, \mu, \mu) \tag{67}$$
$$- \frac{1}{\alpha_0 + 2}\{\mathbb{E}[\theta \otimes \theta \otimes \mathbb{E}[\theta]] - \mathbb{E}[\theta \otimes \mathbb{E}[\theta] \otimes \theta] - \mathbb{E}[\mathbb{E}[\theta] \otimes \theta \otimes \theta]\}(\mu, \mu, \mu) \tag{68}$$
$$+ \frac{2}{\alpha_0(\alpha_0 + 1)(\alpha_0 + 2)} M_1 \otimes M_1 \otimes M_1 \tag{69}$$
$$= \sum_{i}^{k} \frac{2\alpha_i}{\alpha_0(\alpha_0 + 1)(\alpha_0 + 2)} \mu_i \otimes \mu_i \otimes \mu_i \tag{70}$$

□

F. Correctness of Method of Moments for Latent Dirichlet Allocation

Lemma 29 (Correctness of Method of Moments in Learning LDA (Anandkumar et al., 2012))**.** *Applying the method of moments over a corpus of N documents sampled iid. There exist universal constants $C_1, C_2 \geq 0$ such that if $N > C_1((\alpha_0+1)/p_{\mathsf{min}}^2\sigma_k(\mu)^2)$, then $\|\mu_i - \hat{\mu_i}\|_2 \leq C_2 \frac{(\alpha_0+1)^2k^3}{p_{\mathsf{min}}^2\sigma_k(\mu)\sqrt{N}}$, where $p_{\mathsf{min}} = \min_i \frac{\alpha_i}{\alpha_0}$, μ is a matrix of stacked word-topic vectors, i.e. $\mu = [\mu_1|\ldots|\mu_k]$.*

G. Sensitivity Proofs

In proving the sensitivities for $\hat{M_2}$ and $\hat{M_3}$ we rely on the fact that frequently in the calculations, we encounter probability vectors, matrices, and tensors where the elements sum to 1. This is identical to the stating that the l_1 norm equals 1. Further, we note the following Lemma which essentially states that taking the outer product of a vector with a probability vector or probability matrix does not increase the l_q norm of the vector and in fact keeps it the same if $q = 1$.

Lemma 30 (Multiplying by probabilities does not change the norm)**.** *Let v_p, M_p be a probability vector, matrix, respectively and let v, u be ordinary vectors, matrices, respectively. Then the following holds:*

$$\left\|uv_p^T\right\|_q \leq \|u\|_q, \text{ which is equal if } q = 1. \tag{71}$$

$$\text{If } T = M_p \otimes u, \text{ then } \|T\|_q = \|M_p \otimes u\|_q \leq \|u\|_q, \text{ which is equal if } q = 1. \tag{72}$$

Proof.

$$\left\|uv_p^T\right\|_q = (\sum_{i,j} |u_i v_{p_j}|^q)^{1/q} = (\sum_i |u_i|^q \sum_j |v_{p_j}|^q)^{1/q} = \|v\|_q \|u\|_q \leq \|u\|_q . \tag{73}$$

Where we used the fact that $\|x\|_1 \geq \|x\|_q$ for any $q \geq 1$ and that $\|v_p\|_1 = 1$. Thus the above inequality is tight if $q = 1$.

$$\|T\|_q = \|M_p \otimes u\|_q = \Big(\sum_{i,j,k} |M_{p_{i,j}} u_k|^q\Big)^{1/q} = \Big(\sum_k |u_k|^q \sum_{i,j} |M_{p_{i,j}}|^q\Big)^{1/q} = \|u\|_q \|M_p\|_q \leq \|u\|_q . \tag{74}$$

Where we used the fact that for any matrix M, $\|M\|_1 \geq \|M\|_q$ for any $q \geq 1$ [*] and that $\|M_p\|_1 = 1$. Thus the above inequality is tight if $q = 1$. □

Proposition 31. *$\tilde{\hat{M}}_1^n$ is a probability vector, $\tilde{\hat{M}}_2^n$ is a probability matrix, and $\tilde{\hat{M}}_3^n$ is a probability tensor.*

Proof. The proof is immediate as these moments correspond to join probability estimates (Zou et al., 2013), specifically:

$$\tilde{\hat{M}}_1^n(i) = \mathbb{P}[x_1 = i] \tag{75}$$

$$\tilde{\hat{M}}_1^n(i,j) = \mathbb{P}[x_1 = i, x_2 = j] \tag{76}$$

$$\tilde{\hat{M}}_1^n(i,j,k) = \mathbb{P}[x_1 = i, x_2 = j, x_3 = k] \tag{77}$$

□

G.1. Proof for Theorem 4 (sensitivity for $\hat{M}_2$)

Let Δ_2 be the l_1 sensitivity for $\hat{M}_2$, then Δ_2 is $\frac{2}{N} + \frac{\alpha_0}{\alpha_0+1}\frac{4}{N}$=$O(\frac{1}{N})$.

[*]These norms are obtained by extending the vector definition to matrices or simply vectorizing the matrix and then calculating the norm.

Proof. Let $\hat{M}_2$ and $\hat{M'}_2$ be two second order LDA moments generated from two neighboring corpora, WLOG assume the difference is in the n^{th} record, i.e. $D = [c_1|\dots|c_{N-1}|c_N]$ and $D' = [c_1|\dots|c_{N-1}|c'_N]$ then:

$$\hat{M}_2 - \hat{M'}_2 = \frac{1}{N}(\tilde{\tilde{M}}_2^N - \tilde{\tilde{M}}_2^{N'}) - \frac{a}{2\binom{N}{2}}\left(\left[\tilde{\tilde{M}}_1^N \otimes \left(\sum_{n=1}^{N-1}\tilde{\tilde{M}}_1^n\right) + \left(\sum_{n=1}^{N-1}\tilde{\tilde{M}}_1^n\right)\otimes \tilde{\tilde{M}}_1^N\right]\right. \tag{78}$$

$$\left. - \left[\tilde{\tilde{M}}_1^{N'} \otimes \left(\sum_{n=1}^{N-1}\tilde{\tilde{M}}_1^n\right) + \left(\sum_{n=1}^{N-1}\tilde{\tilde{M}}_1^n\right)\otimes \tilde{\tilde{M}}_1^{N'}\right]\right) \tag{79}$$

$$= \frac{1}{N}(\tilde{\tilde{M}}_2^N - \tilde{\tilde{M}}_2^{N'}) - \frac{a}{2\binom{N}{2}}\left((\tilde{\tilde{M}}_1^N - \tilde{\tilde{M}}_1^{N'})\otimes \left(\sum_{n=1}^{N-1}\tilde{\tilde{M}}_1^n\right) + \left(\sum_{n=1}^{N-1}\tilde{\tilde{M}}_1^n\right)\otimes (\tilde{\tilde{M}}_1^N - \tilde{\tilde{M}}_1^{N'})\right) \tag{80}$$

$$= \frac{1}{N}(\tilde{\tilde{M}}_2^N - \tilde{\tilde{M}}_2^{N'}) - \frac{a}{N}\left((\tilde{\tilde{M}}_1^N - \tilde{\tilde{M}}_1^{N'})\otimes \left(\frac{1}{N-1}\sum_{n=1}^{N-1}\tilde{\tilde{M}}_1^n\right) + \left(\frac{1}{N-1}\sum_{n=1}^{N-1}\tilde{\tilde{M}}_1^n\right)\otimes (\tilde{\tilde{M}}_1^N - \tilde{\tilde{M}}_1^{N'})\right) \tag{81}$$

Note that according to proposition (31) $\tilde{\tilde{M}}_1^N$ and $\tilde{\tilde{M}}_1^{N'}$ are probability vectors and $\tilde{\tilde{M}}_2^N$ and $\tilde{\tilde{M}}_2^{N'}$ are probability matrices. Further, $\left(\frac{1}{N-1}\sum_{n=1}^{N-1}\tilde{\tilde{M}}_2^n\right)$ is also a probability matrix since it's the normalized sum of probability matrices. We upper bound the l_1 norm of the expression by applying the triangular inequality and using lemma (30) for the terms involving a tensor product. This leads to the following:

$$\left\|\hat{M}_2 - \hat{M'}_2\right\|_1 \leq \frac{2}{N} + \frac{4a}{N} = \frac{2}{N} + \frac{\alpha_0}{\alpha_0+1}\frac{4}{N} = O(\frac{1}{N}) \tag{82}$$

$$\tag{83}$$

a was replaced by its expression as in the above $a = \frac{\alpha_0}{\alpha_0+1}$ in the above. □

G.2. Proof for Theorem 4 (sensitivity for $\hat{M}_3$)

Let Δ_3 be the l_1 sensitivity for $\hat{M}_3$, then Δ_3 is $\frac{2}{N} + \frac{4\alpha_0}{\alpha_0+2}\frac{1}{N} + \frac{12\alpha_0^2}{(\alpha_0+1)(\alpha_0+2)}\frac{(N-1)}{N(N-2)} = O(\frac{1}{N})$.

Proof. Following a similar setting as in G.1 we have the two moments $\hat{M}_3$ and $\hat{M}'_3$ generated from two neighboring corpora. First we note that the expression of $\hat{M}_3$ and $\hat{M'}_3$ have the following form: $\frac{1}{N}\sum_{n=1}^{N}\tilde{\tilde{M}}_3^n + \mathbf{B}_1 + \mathbf{B}_2 + \mathbf{B}_3 + \mathbf{b}$. Effectively there are three kinds of terms: **(a)** $\frac{1}{N}\sum_{n=1}^{N}\tilde{\tilde{M}}$, **(b)** $\mathbf{B}_1$, and **(c)**$\mathbf{b}$. Since $\mathbf{B}_2$ and $\mathbf{B}_3$are permuted versions of $\mathbf{B}_1$ they have a similar behavior

(a) $\frac{1}{N}\sum_{n=1}^{N}\tilde{\tilde{M}}$**:** The first term difference between $\hat{M}_3$ and $\hat{M'}_3$ would result in $\frac{1}{N}(\tilde{\tilde{M}}_3^N - \tilde{\tilde{M}}_3^{N'})$.

$$\frac{1}{N}\left\|\tilde{\tilde{M}}_3^N - \tilde{\tilde{M}}_3^{N'}\right\|_1 \leq \frac{1}{N}\left(\left\|\tilde{\tilde{M}}_3^N\right\|_1 + \left\|\tilde{\tilde{M}}_3^{N'}\right\|_1\right) \leq \frac{2}{N} \tag{84}$$

Note that both $\tilde{\tilde{M}}_3^N$ and $\tilde{\tilde{M}}_3^{N'}$ are probability tensors.

(b) $\mathbf{B}_1$**:** Based on the minimized expression, the $\mathbf{B}_1$ term difference between $\hat{M}_3$ and $\hat{M}'_3$ is equal to:

$$\mathbf{B}_1 - \mathbf{B'}_1 = \frac{b}{2\binom{N}{2}}\left[\tilde{\tilde{M}}_2^N \otimes \left(\sum_{n=1}^{N-1}\tilde{\tilde{M}}_1^n\right) + \left(\sum_{n=1}^{N-1}\tilde{\tilde{M}}_2^n\right)\otimes \tilde{\tilde{M}}_1^N\right. \tag{85}$$

$$\left. - \tilde{\tilde{M}}_2^{N'} \otimes \left(\sum_{n=1}^{N-1}\tilde{\tilde{M}}_1^n\right) - \left(\sum_{n=1}^{N-1}\tilde{\tilde{M}}_2^n\right)\otimes \tilde{\tilde{M}}_1^{N'}\right] \tag{86}$$

$$= \frac{b}{N}\left[\left(\tilde{\tilde{M}}_2^N - \tilde{\tilde{M}}_2^{N'}\right)\otimes \left(\frac{1}{N-1}\sum_{n=1}^{N-1}\tilde{\tilde{M}}_1^n\right) + \left(\frac{1}{N-1}\sum_{n=1}^{N-1}\tilde{\tilde{M}}_2^n\right)\otimes \left(\tilde{\tilde{M}}_1^N - \tilde{\tilde{M}}_2^N\right)'\right] \tag{87}$$

Note that $\frac{1}{N-1}\sum_{n=1}^{N-1}\tilde{\tilde{M}}_1^n$ and $\frac{1}{N-1}\sum_{n=1}^{N-1}\tilde{\tilde{M}}_2^n$ are probability vectors and matrices, respectively. Thus lemma 30 can be used to upper bound the l_1 norm, leading to the following:

$$\|\mathbf{B}_1 - \mathbf{B'}_1\|_1 \leq \frac{|b|}{N}(2+2) = \frac{4|b|}{N} = \frac{4\alpha_0}{\alpha_0+2}\frac{1}{N} \tag{88}$$

(c) b: Based on the minimized expression, the $\mathbf{b}$ term difference between $\hat{M}_3$ and $\hat{M}_3'$ is equal to:

$$\mathbf{b} - \mathbf{b}' = \frac{c}{6\binom{N}{3}}\Bigg[\Bigg(\tilde{\tilde{M}}_1^N \otimes \Big(\sum_{\substack{m=1,p=1\\ \text{distinct}}}^{N-1}\tilde{\tilde{M}}_1^m \otimes \tilde{\tilde{M}}_1^p\Big) + \Big(\sum_{\substack{n=1,p=1\\ \text{distinct}}}^{N-1}\tilde{\tilde{M}}_1^n \otimes \tilde{\tilde{M}}_1^N \otimes \tilde{\tilde{M}}_1^p\Big) \tag{89}$$

$$+ \Big(\sum_{\substack{n=1,m=1\\ \text{distinct}}}^{N-1}\tilde{\tilde{M}}_1^n \otimes \tilde{\tilde{M}}_1^m\Big) \otimes \tilde{\tilde{M}}_1^N\Bigg) - \Bigg(\tilde{\tilde{M}}_1^{N'} \otimes \Big(\sum_{\substack{m=1,p=1\\ \text{distinct}}}^{N-1}\tilde{\tilde{M}}_1^m \otimes \tilde{\tilde{M}}_1^p\Big) \tag{90}$$

$$- \Big(\sum_{\substack{n=1,p=1\\ \text{distinct}}}^{N-1}\tilde{\tilde{M}}_1^n \otimes \tilde{\tilde{M}}_1^{N'} \otimes \tilde{\tilde{M}}_1^p\Big) - \Big(\sum_{\substack{n=1,m=1\\ \text{distinct}}}^{N-1}\tilde{\tilde{M}}_1^n \otimes \tilde{\tilde{M}}_1^m\Big) \otimes \tilde{\tilde{M}}_1^N\Bigg)\Bigg] \tag{91}$$

$$= \frac{c(N-1)}{N(N-2)}\Bigg[\Big(\tilde{\tilde{M}}_1^N - \tilde{\tilde{M}}_1^{N'}\Big) \otimes \Big(\frac{1}{(N-1)^2}\sum_{\substack{m=1,p=1\\ \text{distinct}}}^{N-1}\tilde{\tilde{M}}_1^m \otimes \tilde{\tilde{M}}_1^p\Big) \tag{92}$$

$$+ \Big(\frac{1}{(N-1)^2}\sum_{\substack{n=1,p=1\\ \text{distinct}}}^{N-1}\tilde{\tilde{M}}_1^n \otimes \Big(\tilde{\tilde{M}}_1^N - \tilde{\tilde{M}}_1^{N'}\Big) \otimes \tilde{\tilde{M}}_1^p\Big) \tag{93}$$

$$+ \Big(\frac{1}{(N-1)^2}\sum_{\substack{n=1,m=1\\ \text{distinct}}}^{N-1}\tilde{\tilde{M}}_1^n \otimes \tilde{\tilde{M}}_1^m\Big) \otimes \Big(\tilde{\tilde{M}}_1^N - \tilde{\tilde{M}}_1^{N'}\Big)\Bigg] \tag{94}$$

Similarly, we have probability tensors so we use lemma 30 to bound the l_1 norm. This results in:

$$\|\mathbf{b} - \mathbf{b}'\|_1 \leq \frac{c(N-1)}{N(N-2)}(2+2+2) = \frac{6c(N-1)}{N(N-2)} = \frac{12\alpha_0^2}{(\alpha_0+1)(\alpha_0+2)}\frac{(N-1)}{N(N-2)} \tag{95}$$

Combing the results from **(a), (b)** and **(c)**, we have the following bound:

$$\Delta_3 \leq \frac{2}{N} + \frac{4\alpha_0}{\alpha_0+2}\frac{1}{N} + \frac{12\alpha_0^2}{(\alpha_0+1)(\alpha_0+2)}\frac{(N-1)}{N(N-2)} = O(\frac{1}{N}) \tag{96}$$

□

G.3. Proof for Theorem 5 (sensitivity for $\hat{M}_3(\hat{W},\hat{W},\hat{W})$)

As explained before, the whitened tensor is denoted as $\widehat{\mathcal{T}}$ for simplicity. Therefore we denote the sensitivity of $\hat{M}_3(\hat{W},\hat{W},\hat{W})$ as $\Delta_{\widehat{\mathcal{T}}}(D)$. Theorem 5 states that $\Delta_{\widehat{\mathcal{T}}}(D) = O(\frac{k^{3/2}}{N\sigma_k(\hat{M}_2)^{3/2}})$.

We need the following Lemma to prove Theorem 5.

Lemma 32. $\left\|\hat{W}' - \hat{W}\right\|_F \leq \frac{\sqrt{2k}\Delta_2}{\sigma_k(\hat{M}_2)\sqrt{\sigma_k(\hat{M}_2)-\Delta_2}}$

Proof. We follow an analysis similar to (Anandkumar et al., 2012). Note that the whitening matrix $\hat{W}$ is defined such that:

$$\hat{W}^T\hat{M}_{2,k}\hat{W} = I. \tag{97}$$

Analogously for the neighboring corpus,

$$\hat{W'}^T \hat{M'}_{2,k} \hat{W'} = I. \tag{98}$$

Let E_{M_2} denote the perturbation introduced to $\hat{M}_2$ by changing a single record. Because the spectral gap of the perturbation introduced by modifying a single record is small according to the condition, applying the original whitening matrix to the neighboring data base moment $\hat{M'}_2$ would lead to a rank k matrix of size $k \times k$. Therefore, $\hat{W}^T \hat{M'}_{2,k} \hat{W}$ is a rank k matrix of size $k \times k$, which can be factorized as:

$$\hat{W}^T \hat{M'}_{2,k} \hat{W} = ADA^T \tag{99}$$

where A are the singular vectors of $\hat{W}^T \hat{M'}_{2,k} \hat{W}$, and D is a diagonal matrix of the corresponding singular values of $\hat{W}^T \hat{M'}_{2,k} \hat{W}$. This also leads to $\hat{W'} = \hat{W} A D^{\frac{-1}{2}} A^T$. Using this, we observe:

$$\left\| \hat{W'} - \hat{W} \right\| = \left\| \hat{W'} - \hat{W'} A D^{\frac{1}{2}} A^T \right\| = \left\| \hat{W'} (I - A D^{\frac{1}{2}} A^T) \right\| \leq \left\| \hat{W'} \right\| \left\| I - A D^{\frac{1}{2}} A^T \right\| \tag{100}$$

Now we bound $\left\| I - A D^{\frac{1}{2}} A^T \right\|$:

$$\left\| I - A D^{\frac{1}{2}} A^T \right\| = \left\| A^T A - \hat{W'} A D^{\frac{1}{2}} A^T \right\| = \left\| I - D^{\frac{1}{2}} \right\| \tag{101}$$

$$\leq \left\| (I - D^{\frac{1}{2}})(I + D^{\frac{1}{2}}) \right\| \leq \|(I - D)\| \tag{102}$$

$$= \left\| I - A D A^T \right\| = \left\| \hat{W}^T \hat{M}_{2,k} \hat{W} - \hat{W'}^T \hat{M'}_{2,k} \hat{W'} \right\| \tag{103}$$

$$\leq \left\| \hat{W} \right\|^2 \left\| \hat{M}_{2,k} - \hat{M'}_{2,k} \right\| \leq \left\| \hat{W} \right\|^2 \|E_{M_2}\| \tag{104}$$

We know that

$$\left\| \hat{W} \right\|^2 \leq \frac{1}{\sigma_k(\hat{M}_2)} \tag{105}$$

$$\left\| \hat{W'} \right\| \leq \frac{1}{\sqrt{\sigma_k(\hat{M'}_2)}} \leq \frac{1}{\sqrt{\sigma_k(\hat{M}_2) - \|E_{M_2}\|_2}} \leq \frac{1}{\sqrt{\sigma_k(\hat{M}_2) - \Delta_2}} \tag{106}$$

Weyl's theorem was used in the last bound in Equation (106). Bounding the Frobenius norm, would result in the following:

$$\left\| \hat{W'} - \hat{W} \right\|_F \leq \sqrt{2k} \left\| \hat{W'} - \hat{W} \right\| \leq \frac{\sqrt{2k}\, \|E_{M_2}\|}{\sigma_k(\hat{M}_2) \sqrt{\sigma_k(\hat{M}_2) - \|E_{M_2}\|}} \leq \frac{\sqrt{2k} \Delta_2}{\sigma_k(\hat{M}_2) \sqrt{\sigma_k(\hat{M}_2) - \Delta_2}}, \tag{107}$$

where we have used the fact that the l_1 norm upper bounds the spectral norm of a matrix, since it upper bounds the Frobenius. □

Now we are ready to prove Theorem 5.

Proof. $\hat{M'}_3 = \hat{M}_3 + E_3$.

$$\left\| \hat{M}_3(\hat{W}, \hat{W}, \hat{W}) - \hat{M'}_3(\hat{W'}, \hat{W'}, \hat{W'}) \right\|_F = \|\hat{M}_3(\hat{W}, \hat{W}, \hat{W}) - \hat{M}_3^{LDA}(\hat{W'}, \hat{W'}, \hat{W'}) - E_3(\hat{W'}, \hat{W'}, \hat{W'})\|_F \tag{108}$$

$$\leq \left\| \hat{M}_3^{LDA}(W \hat{-} W', W \hat{-} W', W \hat{-} W') \right\|_F + \left\| E_3(\hat{W'}, \hat{W'}, \hat{W'}) \right\|_F \tag{109}$$

$$\leq \left\| \hat{M}_3 \right\|_F \left\| \hat{W} - \hat{W'} \right\|_F^3 + \|\Delta_3\|_F \left\| \hat{W'} \right\|_F^3 \tag{110}$$

We have used the fact that the Frobenius norm of the difference between the tensors is bounded above by the l_1 norm of the difference Δ_3. To bound the l_1 norm of $\hat{M}_3$ we use an analysis similar to calculating Δ_3. Again we note that the l_1 norm upper bounds the Frobenius norm:

$$\left\| \hat{M}_3 \right\|_F \leq \left\| \hat{M}_2 \right\|_1 = 1 + \frac{6\alpha_0}{\alpha_0 + 2} \frac{N}{N-1} + \frac{6\alpha_0^2}{(\alpha_0 + 1)(\alpha_0 + 2)} \frac{N^3}{N(N-1)(N-2)} \tag{111}$$

Combining all the expressions we get:

$$\Delta_{\widehat{\mathcal{T}}}(D) = \left\| \hat{M}_3(\hat{W},\hat{W},\hat{W}) - \hat{M'}_3(\hat{W'},\hat{W'},\hat{W'}) \right\|_F \tag{112}$$

$$\leq (1 + \frac{6\alpha_0}{\alpha_0+2}\frac{N}{N-1} + \frac{6\alpha_0^2}{(\alpha_0+1)(\alpha_0+2)}\frac{N^3}{N(N-1)(N-2)}) \times \frac{(2k)^{3/2}(\Delta_2)^3}{\sigma_k(\hat{M}_2)^3(\sigma_k(\hat{M}_2)-\Delta_2)^{3/2}} + \frac{\Delta_3 k^{3/2}}{(\sigma_k(\hat{M}_2)-\Delta_2)^{3/2}} \tag{113}$$

$$= O(\frac{k^{3/2}}{N\sigma_k(\hat{M}_2)^{3/2}}) \tag{114}$$

We see that if N is larger than $d^{3/2}$, then $N\sigma_k(\hat{M}_2)^{3/2} \geq 1$ as $\sigma_i(\hat{M}_2)$ is in the order of $1/d$. □

G.4. Proof for Theorem 6 (sensitivity of the output of tensor decomposition $\bar{\mu}_i, \bar{\alpha}_i$)

Let $\bar{\mu}_1, \ldots, \bar{\mu}_k$ and $\bar{\alpha}_1, \ldots, \bar{\alpha}_k$ be the results of tensor decomposition before unwhitening. The sensitivity of $\bar{\mu}_i$, denoted as $\Delta_{\bar{\mu}}(D)$, and the sensitivity of $\bar{\alpha}_i$, denoted as $\Delta_{\bar{\alpha}}(D)$, are both upper bounded by $\Delta_{\bar{\mu}}(D) \leq O(\frac{k^2}{\gamma_s N(\sigma_k(\hat{M}_2))^{3/2}})$, where $\gamma_s = \min_{i\in[k]} \frac{\sigma_i - \sigma_{i+1}}{4}$, σ_i is the i^{th} eigenvalue of $\hat{M}_3(\hat{W},\hat{W},\hat{W})$.

Proof. The proof follows from the result of the simultaneous tensor power method (Theorem 1 in (Wang & Lu, 2017)). Replacing the original eigenvectors with those resulting from database D leads to tensor $\hat{M}_3(\hat{W},\hat{W},\hat{W})$, then the tensor resulting from corpus D' with one record changed yields $\hat{M'}_3(\hat{W'},\hat{W'},\hat{W'})$ where the spectral norm of the error is upper bounded by ϵ, if $\Delta_{\widehat{\mathcal{T}}}(D)$ is sufficiently small $\Delta_{\widehat{\mathcal{T}}}(D) \leq \frac{\gamma_s \epsilon}{2\sqrt{k}}$. Therefore we get $\left\|\bar{\mu}_i - \bar{\mu'_i}\right\|_2 \leq \frac{2\sqrt{k}\Delta_{\widehat{\mathcal{T}}}(D)}{\gamma_s}$ and $|\bar{\alpha}_i - \bar{\alpha'}_i| \leq \frac{2\sqrt{k}\Delta_{\widehat{\mathcal{T}}}(D)}{\gamma_s}$. □

G.5. Proof for Theorem 7 (sensitivity of the final output μ_i, α_i)

We now prove the sensitivity of the final output μ_i, α_i: $\Delta_\mu(D) = O(\frac{k^2\sqrt{\sigma_1(\hat{M}_2)}}{\gamma_s N \sigma_k^{3/2}(\hat{M}_2)})$.

Proof. We point out a number of things. Tensor decomposition outputs are: $\bar{\mu}_i, \bar{\alpha}_i, i \in [k]$, where, $\bar{\alpha}_i = \frac{2\sqrt{(\alpha_0+1)\alpha_0}}{(\alpha_0+2)\sqrt{\alpha_i}}$. In order to recover the desired word topic vector μ, we have to "unwhiten" to get the μ_i and α_i before whitening, i.e. $\mu_i = \frac{1}{\sqrt{\alpha_i^r}}(W^T)^\dagger \bar{\mu}_i$, where $\frac{1}{\sqrt{\alpha_i^r}} = \frac{(\alpha_0+2)}{2\sqrt{(\alpha_0+1)\alpha_0}}\bar{\alpha}_i$. The sensitivity would be:

$$\max_{D,D'} \|\mu_i - \mu\prime_i\| \leq \max_{D,D'} \left\{ \left\| \frac{1}{\sqrt{\alpha_i^r}}(W^T)^\dagger \bar{\mu}_i - \frac{1}{\sqrt{\alpha_i^{r,'}}}(W^{T,'})^\dagger \bar{\mu}'_i \right\|_2 \right\} \tag{115}$$

$$\leq \max_{D,D'} \left\{ \frac{1}{\sqrt{\alpha_i^r}} \left\|(W^T)^\dagger\right\| \|\bar{\mu}_i - \bar{\mu}'_i\| + \frac{1}{\sqrt{\alpha_i^r}} \left\|W^\dagger - (W')^\dagger\right\| + \left\|(W')^\dagger\right\| |\frac{1}{\sqrt{\alpha_i^r}} - \frac{1}{\sqrt{\alpha_{i,\prime}^r}}| \right\} \tag{116}$$

We note the following:

- **(i)** $\max_{D,D'} |\frac{1}{\sqrt{\alpha_i^r}} - \frac{1}{\sqrt{\alpha_{i,\prime}^r}}| = \max_{D,D'} |\frac{(\alpha_0+2)}{2\sqrt{(\alpha_0+1)\alpha_0}}\bar{\alpha}_i - \frac{(\alpha_0+2)}{2\sqrt{(\alpha_0+1)\alpha_0}}\bar{\alpha}'_i| \leq \frac{(\alpha_0+2)}{2\sqrt{(\alpha_0+1)\alpha_0}} \max_{D,D'} |\bar{\alpha}_i - \bar{\alpha}'_i| \leq \frac{(\alpha_0+2)}{2\sqrt{(\alpha_0+1)\alpha_0}} \frac{2\sqrt{k}\Delta_{\widehat{\mathcal{T}}}(D)}{\gamma_s}$, where the above follows from the simultaneous power iteration method.
- **(ii)** $\max_{i\in[k]} \frac{1}{\sqrt{\alpha_i^r}} \leq \frac{(\alpha_0+2)}{2\sqrt{(\alpha_0+1)\alpha_0}} \max_{i\in[k]} \bar{\alpha}_i = \frac{(\alpha_0+2)}{2\sqrt{(\alpha_0+1)\alpha_0}} \sigma_1(\widehat{\mathcal{T}})$
- **(iii)** $\max \left\|((W\prime)^T)\dagger\right\| \leq \sqrt{\sigma_1(\hat{M'}_2)} \leq \sqrt{\sigma_1(\hat{M}_2) + \Delta_2}$
- **(iv)** Following an analysis similar to that in 32, we obtain $\left\|W^\dagger - (W')^\dagger\right\| \leq \frac{\sqrt{\sigma_1(\hat{M}_2)}}{\sigma_k(\hat{M}_2)}\Delta_2$.

Combining all of this together leads to the following

$$\max_{D,D'} \|\mu_i - \mu\prime_i\| \leq \frac{(\alpha_0+2)}{2\sqrt{(\alpha_0+1)\alpha_0}}\sigma_1(\widehat{\mathcal{T}})\sqrt{\sigma_1(\hat{M}_2)}\frac{2\sqrt{k}\Delta_{\widehat{\mathcal{T}}}(D)}{\gamma_s} + \frac{(\alpha_0+2)}{2\sqrt{(\alpha_0+1)\alpha_0}}\sigma_1(\widehat{\mathcal{T}})\frac{\sqrt{\sigma_1(\hat{M}_2)}}{\sigma_k(\hat{M}_2)}\Delta_2$$
$$+ \frac{(\alpha_0+2)}{2\sqrt{(\alpha_0+1)\alpha_0}}\sqrt{\sigma_1(\hat{M}_2)+\Delta_2}\frac{2\sqrt{k}\Delta_{\widehat{\mathcal{T}}}(D)}{\gamma_s} \tag{117}$$
$$= O(\frac{k^2\sqrt{\sigma_1(\hat{M}_2)}}{\gamma_s N\sigma_k^{3/2}(\hat{M}_2)}) \tag{118}$$

□

G.6. Proof for Lemma 8

Let $\tilde{LS}$ denote the local sensitivity. We prove a slightly more general version where the construction of $\tilde{LS}$ is (ϵ_1,δ_1)-DP, and it is a valid upper bound with probability $\geq 1-\delta_3$.

Lemma 33. *Let* LS *be the ℓ_p local sensitivity of a function f on a fixed data set. Let* $\tilde{\text{LS}}$ *obeys (ϵ_1,δ_1)-DP and that $\mathbb{P}[\text{LS} \geq \tilde{\text{LS}}] \leq \delta_3$ (where the probability is only over the randomness in releasing* $\tilde{\text{LS}}$*). Then the algorithm releases $f(DATA) + Z(\epsilon,\delta,\tilde{\text{LS}})$ that is $(\epsilon_1+\epsilon_2, \delta_1+\delta_2+\delta_3)$-DP, where $Z(\epsilon_2,\delta_2,\tilde{\text{LS}})$ is any way of calibrating the noise for privacy which takes the local sensitivity as if it is a global sensitivity.*

Proof. Let x, x' be two adjacent data sets and the overall output be $O := f(\text{DATA}) + Z(\epsilon_2,\delta_2,\tilde{LS})$. Let $S_1 \subset \text{Range}(f), S_2 \subset \mathbb{R}_+$ be any measurable sets.

Let E be the measurable set of $\tilde{\text{LS}}$ that represents the event that $\tilde{\text{LS}} \geq \text{LS}$.

$$\mathbb{P}[(O,\tilde{\text{LS}}) \in S_1 \times S_2 | x] \tag{119}$$
$$= \mathbb{P}[(O,\tilde{\text{LS}}) \in S_1 \times (S_2 \cap E)|x] + \mathbb{P}[(O,\tilde{\text{LS}}) \in S_1 \times S_2 \cap E^c|x] \tag{120}$$
$$\leq \mathbb{P}[(O,\tilde{\text{LS}}) \in S_1 \times (S_2 \cap E)|x] + \delta_3 \tag{121}$$
$$\leq e^{\epsilon_1+\epsilon_2}\mathbb{P}[(O,\tilde{\text{LS}}) \in S_1 \times (S_2 \cap E)|x'] + \delta_1 + \delta_2 + \delta_3 \tag{122}$$
$$\leq e^{\epsilon_1+\epsilon_2}\mathbb{P}[(O,\tilde{\text{LS}}) \in S_1 \times S_2|x'] + \delta_1 + \delta_2 + \delta_3 \tag{123}$$

The fourth line holds due to the fact that under event the E, $\tilde{\text{LS}}$ is always a valid upper bound of the local sensitivity, therefore, conditioning on the σ-field induced by $E \cap S_2$ for any S_2, O is an (ϵ_2,δ_2)-DP release. By the simple composition Theorem of (ϵ,δ)-DP (Dwork et al., 2014a)[Theorem B.1,], by taking the measurable set of interest to be $S_1 \times (S_2 \cap E)$, we have that

$$\mathbb{P}[(O,\tilde{\text{LS}}) \in S_1 \times (S_2 \cap E)|x] \leq e^{\epsilon_1+\epsilon_2}\mathbb{P}[(O,\tilde{\text{LS}}) \in S_1 \times (S_2 \cap E)|x'] + \delta_1 + \delta_2$$

which wraps up the proof. □

The proof of Lemma 8 is a corollary which takes $\delta_1 = 0$.

G.7. Proof for Sensitivity of singular values $\sigma_k(\hat{M}_2)$ (Lemma 9)

Proof. We first prove that the global sensitivity of $\sigma_k(\hat{M}_2)$ is $1/n$. By Weyl's lemma (Stewart, 1998)[Theorem 1], for any matrix X, any i, the singular value $|\sigma_i(X) - \sigma_i(X+E)| \leq \|E\|_2$. In our case, E is coming from adding or removing one data point and we know that $\|E\|_2 \leq \|E\|_F \leq \|E\|_{1,1} \leq 2/n$, hence the bound.

Now we prove that the global sensitivity of $\gamma_s = \min_{i\in[k]} \frac{\sigma_i(\widehat{\mathcal{T}})-\sigma_{i+1}(\widehat{\mathcal{T}})}{4}$. For any tensor $\widehat{\mathcal{T}}$, we consider a polyadic form or the so called tensor decomposition form, and denote the singular values as the amplitude of the components in the polyadic form. As shown in Section G.2, $|\sigma_i(\widehat{\mathcal{T}}) - \sigma_i(\widehat{\mathcal{T}}+\mathcal{E})| \leq \|\mathcal{E}\| \leq 1$, where $\mathcal{E}$ comes from adding or removing one data point.

□

H. Utility Proofs

Before starting the utility proofs, we point out a number of things. Tensor decomposition outputs:$\bar{\mu}_i, \bar{\alpha}_i, i \in [k]$. Where, $\bar{\alpha}_i = \frac{2\sqrt{(\alpha_0+1)\alpha_0}}{(\alpha_0+2)\sqrt{\alpha_i}}$. In order to recover the desired word topic vector μ, we have to 'reverse whiten', i.e. $\mu_i = \frac{1}{\sqrt{\alpha_i^r}}(W^T)^\dagger \bar{\mu}_i$, where $\frac{1}{\sqrt{\alpha_i^r}} = \frac{(\alpha_0+2)}{2\sqrt{(\alpha_0+1)\alpha_0}}\bar{\alpha}_i$. We need to establish the distance between the non-differentially private output and the differentially private output, i.e. $\left\|\mu_i - \mu_i^{DP}\right\|$. This can be upper bounded similar to G.5 by the following:

$$\left\|\mu_i - \mu_i^{DP}\right\| \leq \frac{1}{\sqrt{\alpha_i^r}} \left\|(W^T)^\dagger\right\| \left\|\bar{\mu}_i - \bar{\mu}_i^{DP}\right\| + \frac{1}{\sqrt{\alpha_i^r}} \left\|W^\dagger - (W^{DP})^\dagger\right\| + \left\|(W^{DP})^\dagger\right\| \left|\frac{1}{\sqrt{\alpha_i^r}} - \frac{1}{\sqrt{\alpha_{i,DP}^r}}\right| \tag{124}$$

For this we frequently need to bound the following: $\left\|\bar{\mu}_i - \bar{\mu}_i^{DP}\right\|$, $\left\|W^\dagger - (W^{DP})^\dagger\right\|$, $\left\|(W^{DP})^\dagger\right\|$, $\left|\frac{1}{\sqrt{\alpha_i^r}} - \frac{1}{\sqrt{\alpha_{i,DP}^r}}\right|$, and $|\bar{\alpha}_i - \bar{\alpha}_i^{DP}|$.

We point out the following facts before preceding.

Fact 34. $\left|\frac{1}{\sqrt{\alpha_i^r}} - \frac{1}{\sqrt{\alpha_{i,DP}^r}}\right| \leq \left|\frac{(\alpha_0+2)}{2\sqrt{(\alpha_0+1)\alpha_0}}\bar{\alpha}_i - \frac{(\alpha_0+2)}{2\sqrt{(\alpha_0+1)\alpha_0}}\bar{\alpha}_i^{DP}\right| \leq \frac{(\alpha_0+2)}{2\sqrt{(\alpha_0+1)\alpha_0}}|\bar{\alpha}_i - \bar{\alpha}_i^{DP}|$.

Fact 35. $\left\|(W^T)^\dagger\right\| \leq \sqrt{\sigma_1(\hat{M}_2)}$.

Fact 36. $\frac{1}{\sqrt{\alpha_i^r}} = \frac{(\alpha_0+2)}{2\sqrt{(\alpha_0+1)\alpha_0}}\bar{\alpha}_i \leq \frac{(\alpha_0+2)}{2\sqrt{(\alpha_0+1)\alpha_0}}\sigma_1(\widehat{\mathcal{T}})$.

H.1. Perturbation on $\hat{M}_2$, $\hat{M}_3$ Config. 1 (e_3, e_4, e_8): Proof for Theorem 10

Similar to the perturbation on (e_6, e_8). We have that

$$\left\|W^\dagger - (W^{DP})^\dagger\right\| \leq \frac{\sqrt{\sigma_1(\hat{M}_2)}\left\|E_{8,G}\right\|}{\sigma_k(\hat{M}_2)} \tag{125}$$

$$\left\|(W^{DP})^\dagger\right\| \leq \sqrt{\sigma_1(\hat{M}_2) + \left\|E_{8,G}\right\|} \tag{126}$$

Now the perturbed tensor can be represented as $\hat{M}_3^{DP} = \hat{M}_3 + E_{3,G}$, where $E_{3,G}$ is symmetric Gaussian noise that has been added to the original tensor. Similar to the sensitivity analysis for the whitened tensor, we have that the error Φ can be bounded as follows:

$$\|\Phi\|_2 = \left\|\hat{M}_3(\hat{W}, \hat{W}, \hat{W}) - \hat{M}_3^{DP}(W^{DP}, W^{DP}, W^{DP})\right\|_2 \tag{127}$$

$$\leq \left\|\hat{M}_3\right\| \left\|W - W^{DP}\right\|^3 + \left\|E_{3,G}\right\| \left\|W^{DP}\right\| \tag{128}$$

Following an analysis similar to bounding $\left\|W^\dagger - (W^{DP})^\dagger\right\|$, we get that $\left\|W^\dagger - (W^{DP})^\dagger\right\| \leq \frac{\|E_{8,G}\|}{\sigma_k(\hat{M}_2)\sqrt{\frac{\sigma_k(\hat{M}_2)}{2}}}$. According to 43 we have that with high probability $\|E_{3,G}\| = O(\sqrt{d}\Delta_3\tau_{\epsilon_3,\delta_3})$. We note the following $\left\|\bar{\mu}_i - \bar{\mu}_i^{DP}\right\|_2 \leq \frac{2\sqrt{k}\|\Phi\|}{\gamma_s}$ using the simultaneous power iteration of (Wang & Lu, 2017). Similarly we have $|\bar{\alpha}_i - \bar{\alpha}_i^{DP}| \leq \frac{2\sqrt{k}\|\Phi\|}{\gamma_s}$ and that $\left|\frac{1}{\sqrt{\alpha_i^r}} - \frac{1}{\sqrt{\alpha_{i,DP}^r}}\right| \leq \frac{(\alpha_0+2)}{2\sqrt{(\alpha_0+1)\alpha_0}}\frac{2\sqrt{k}\|\Phi\|}{\gamma_s}$. This leads to $\left\|\mu_i - \mu_i^{DP}\right\|_2 \leq \frac{(\alpha_0+2)}{2\sqrt{(\alpha_0+1)\alpha_0}}\sigma_1(\widehat{\mathcal{T}})\sqrt{\sigma_1(\hat{M}_2)}\frac{2\sqrt{k}\|\Phi\|}{\gamma_s} + \frac{(\alpha_0+2)}{2\sqrt{(\alpha_0+1)\alpha_0}}\sigma_1(\widehat{\mathcal{T}})\frac{\sqrt{\sigma_1(\hat{M}_2)}}{\sigma_k(\hat{M}_2)}\left\|E_{8,G}\right\| + \sqrt{\sigma_1(\hat{M}_2) + \left\|E_{8,G}\right\|}\frac{(\alpha_0+2)}{2\sqrt{(\alpha_0+1)\alpha_0}}\frac{2\sqrt{k}\|\Phi\|}{\gamma_s}$.

Based on the bound on $\|\Phi\|$ we have with high probability $\left\|\mu_i - \mu_i^{DP}\right\|_2 = O(\frac{\sqrt{\sigma_1(\hat{M}_2)k}}{\gamma_s}((\frac{\sqrt{d}}{N\sigma_k(\hat{M}_2)^{3/2}}\tau_{\epsilon_4,\delta_4})^3 + \frac{\sqrt{d}}{N\sigma_k(\hat{M}_2)^{3/2}}\tau_{\epsilon_3,\delta_3}) + \frac{\sqrt{\sigma_1(\hat{M}_2)d}}{\sigma_k(\hat{M}_2)N}\tau_{\epsilon_8,\delta_8} + \sqrt{\sigma_1(\hat{M}_2) + \frac{\sqrt{d}}{N}\tau_{\epsilon_8,\delta_8}}\frac{\sqrt{k}}{\gamma_s}\left[(\frac{\sqrt{d}}{N\sigma_k(\hat{M}_2)}\tau_{\epsilon_4,\delta_4})^3 + \frac{\sqrt{d}}{N\sigma_k(\hat{M}_2)^{3/2}}\tau_{\epsilon_3,\delta_3}\right])$.

H.2. Perturbation on $\widehat{\mathcal{T}}$ and $\hat{M}_2$ Config. 2(e_6, e_8): Proof for Theorem 12

This configuration has two properties: the noise level introduced is low because the whitening step reduces the tensor dimension from $\hat{M}_3 \in \mathbb{R}^{d\times d\times d}$ to $\widehat{\mathcal{T}} = \hat{M}_3(\hat{W}, \hat{W}, \hat{W}) \in \mathbb{R}^{k\times k\times k}$. However, even though the dimension of the tensor is

reduced, unless the whitening tensor (resulting from eigendecomposition over $\hat{M}_2$) is stable, the sensitivity of the whitened tensor is not necessarily low.

Note that the sensitivity of $\hat{M}_2$ falls with $\frac{1}{N}$ (Theorem 4). Therefore, we expect the sensitivity of $\hat{M}_3(\hat{W}, \hat{W}, \hat{W})$ to drop with an increasing number of records. As Theorem 5 states, $\Delta_{\widehat{\mathcal{T}}}(D) = O(\frac{k^{3/2}}{N\sigma_k^{3/2}(\hat{M}_2)})$, if $\Delta_2 \leq \sigma_k(\hat{M}_2) - \sigma_{k+1}(\hat{M}_2)$. Thus, given the spectral gap requirement, the sensitivity of the whitened tensor is $\Delta_{\widehat{\mathcal{T}}}(D)$.

$\hat{M}_2$ is used to generate both the whitening and unwhitening matrix, and unlike input perturbation, the sensitivity over $\hat{M}_2$ and $\hat{M}_3$ falls as the dataset size increases (Theorem 4). However, an issue with this configuration is that adding noise to $\hat{M}_3$ leads to higher noise build up prior to the tensor decomposition. Note that by (43) w.h.p the norm of the error is $O(\sqrt{d}\sigma)$, with σ being the variance of the noise (this bound would be $\sqrt{k}\sigma$ if the noise is added to a symmetric tensor of size k). Tensor decomposition methods, in particular (Wang & Lu, 2017) require the spectral norm of the perturbation to the tensor to be lower than a certain threshold. Following arguments similar to (Wang & Anandkumar, 2016), the spectral norm of the error is $O(\frac{\sqrt{d}}{N\epsilon_3})$ and should be below $\frac{\sqrt{k}}{\gamma_s\sigma_k(\widehat{\mathcal{T}})}$. Thus ϵ_3 should satisfy $\epsilon_3 = \Omega(\frac{\sqrt{kd}}{\gamma_s\sigma_k(\widehat{\mathcal{T}})N})$ to establish utility guarantees for tensor decomposition. Following similar arguments, this time using the bound on the spectral norm of the noisy matrices, to guarantee utility, the differentially private whitening W and pseudo-inverse $W^\dagger$ should be close to their non-differentially private values, which requires both ϵ_4 and ϵ_8 to be $\Omega(\frac{\sqrt{d}}{(\sigma_k(\hat{M}_2)-\sigma_{k+1}(\hat{M}_2)N)})$. Although, the privacy parameters have a lower bound of $\sqrt{d}$, the bound also falls with $\frac{1}{N}$.

The spectral norm of the noise added to $\hat{M}_2$ can be bounded by 42 to be $O(\frac{\sqrt{d}}{N}\tau_{\epsilon_8,\delta_8})$ with high probability. Now, if we have $N = \Omega(\frac{\sqrt{d}\tau_{\epsilon_8,\delta_8}}{\sigma_k(\hat{M}_2)-\sigma_{k+1}(\hat{M}_2)})$, then with w.h.p we have that $\|E_{8,G}\| \leq \frac{\sigma_k(\hat{M}_2)-\sigma_{k+1}(\hat{M}_2)}{2}$, where $\|E_{8,G}\|$ is the spectral norm of the Gaussian matrix. This condition enables us to bound $\left\|W^\dagger - (W^{DP})^\dagger\right\|$, in a manner similar to establishing the bounds between $\|W - W'\|$ in 32. Following a similar analysis, given that

$$W^T(\hat{M}_2)_k W = I, \tag{129}$$

$$W^{T,DP}(\hat{M}_2 + E_{8,G})_k W^{DP} = I, \tag{130}$$

$$W^T(\hat{M}_2 + E_{8,G})_k W = ADA^T, \tag{131}$$

we have that $\left\|W^\dagger - (W^{DP})^\dagger\right\| \leq \left\|W^\dagger\right\| \|I - D\|$. We know that $\left\|W^\dagger\right\| \leq \frac{1}{\sqrt{\sigma_k(\hat{M}_2)}}$ and $\|I - D\|$ can be bounded as follows:

$$\|I - D\| \leq \left\|I - ADA^T\right\| \leq \left\|W^T(\hat{M}_2)_k W - W^T(\hat{M}_2 + E_{8,G})_k W\right\| \tag{132}$$

$$\leq \|W\|^2 \left\|(\hat{M}_2)_k - (\hat{M}_2 + E_{8,G})_k\right\| \leq \|W\|^2 \|E_{8,G}\| \leq \frac{\|E_{8,G}\|}{\sigma_k(\hat{M}_2)} \tag{133}$$

This leads to $\left\|W^\dagger - (W^{DP})^\dagger\right\| \leq \frac{\sqrt{\sigma_1(\hat{M}_2)}\|E_{8,G}\|}{\sigma_k(\hat{M}_2)}$.

Moreover, it is immediate by Weyl's theorem that $\left\|(W^{DP})^\dagger\right\| \leq \sqrt{\sigma_1(\hat{M}_2 + E_{8,G})} \leq \sqrt{\sigma_1(\hat{M}_2) + \|E_{8,G}\|}$.

Finally, by the results of simultaneous power iteration (with an argument similar to Theorem 6), if N is sufficiently large, we have that $\left\|\bar{\mu}_i - \bar{\mu}_i^{DP}\right\| \leq \frac{2\sqrt{k}\|E_{6,G}\|}{\gamma_s}$ where $E_{6,G}$ is the Gaussian tensor added to the whitened tensor $\Delta_{\widehat{\mathcal{T}}}(D)$. An identical bound is established for the eigenvalues, i.e. $|\bar{\alpha}_i - \bar{\alpha}_i^{DP}| \leq \frac{2\sqrt{k}\|E_{6,G}\|}{\gamma_s}$.

Now we can state the utility:

$$\left\|\mu_i - \mu_i^{DP}\right\| \leq \frac{(\alpha_0+2)}{2\sqrt{(\alpha_0+1)\alpha_0}}\sigma_1(\widehat{\mathcal{T}})\sqrt{\sigma_1(\hat{M}_2)}\frac{2\sqrt{k}\,\|E_{6,G}\|}{\gamma_s} + \frac{(\alpha_0+2)}{2\sqrt{(\alpha_0+1)\alpha_0}}\sigma_1(\widehat{\mathcal{T}})\frac{\sqrt{\sigma_1(\hat{M}_2)}}{\sigma_k(\hat{M}_2)}\|E_{8,G}\|$$
$$+ \frac{(\alpha_0+2)}{2\sqrt{(\alpha_0+1)\alpha_0}}\sqrt{\sigma_1(\hat{M}_2) + \|E_{8,G}\|}\frac{2\sqrt{k}\,\|E_{6,G}\|}{\gamma_s} \tag{134}$$

We note that w.h.p we have the following bounds on spectral norms of noisy Gaussian matrix and noisy Gaussian tensor. In particular, $\|E_{6,G}\| = O(\frac{k^2}{N\tilde{\sigma}_k^{3/2}}\tau_{\epsilon_6,\delta_6})$ and $\|E_{8,G}\| = O(\frac{\sqrt{d}}{N}\tau_{\epsilon_8,\delta_8})$. This leads to the following utility

$$\left\|\mu_i - \mu_i^{DP}\right\| = O(\frac{\sqrt{\sigma_1(\hat{M}_2)k^{2.5}}}{\gamma_s N\tilde{\sigma}_k^{3/2}}\tau_{\epsilon_6,\delta_6} + \frac{\sqrt{\sigma_1(\hat{M}_2)d}}{\sigma_k(\hat{M}_2)N}\tau_{\epsilon_8,\delta_8} + \sqrt{\sigma_1(\hat{M}_2) + \frac{\sqrt{d}}{N}\tau_{\epsilon_8,\delta_8}}\frac{k^{2.5}\tau_{\epsilon_6,\delta_6}}{\gamma_s N\tilde{\sigma}_k^{3/2}}). \tag{135}$$

H.3. Perturbation on the output of tensor decomposition $\bar{\mu}_i$,$\bar{\alpha}_i$ and $\hat{M}_2$ Config. 3 (e_7, e_8): Proof for Theorem 14

This configuration shares edge 8 with the previous. This enables us to borrow the same bounds for the pseudo-inverse $W^\dagger$. Specifically, we have:

$$\left\|W^\dagger - (W^{DP})^\dagger\right\| \leq \frac{\sqrt{\sigma_1(\hat{M}_2)}\,\|E_{8,G}\|}{\sigma_k(\hat{M}_2)} \tag{136}$$

$$\left\|(W^{DP})^\dagger\right\| \leq \sqrt{\sigma_1(\hat{M}_2) + \|E_{8,G}\|} \tag{137}$$

In this method, noise is added directly to the eigenvectors and eigenvalues resulting from the tensor decomposition. Therefore, we have:

$$\bar{\mu}_i^{DP} = \bar{\mu}_i + Y, \qquad Y \sim \mathcal{N}(0, \Delta_{\epsilon,\delta}^2 I_k) \tag{138}$$

$$\bar{\alpha}_i^{DP} = \bar{\alpha}_i + n_i, \qquad n_i \sim \mathcal{N}(0, \Delta_{\epsilon,\delta}^2) \tag{139}$$

where $\Delta_{\epsilon,\delta} = \frac{\sqrt{2k}\Delta_{\widehat{T}}(D)}{\gamma_s}\tau_{\epsilon_7,\delta_7}$ with $\tau_{\epsilon_7,\delta_7} = \frac{\sqrt{2ln(1.25/\delta_7)}}{\epsilon_7}$. This leads to the following bound:

$$\begin{aligned}\left\|\mu_i - \mu_i^{DP}\right\| \leq & \frac{(\alpha_0+2)}{2\sqrt{(\alpha_0+1)\alpha_0}}\sigma_1(\widehat{\mathcal{T}})\sqrt{\sigma_1(\hat{M}_2)}\,\|Y\| + \frac{(\alpha_0+2)}{2\sqrt{(\alpha_0+1)\alpha_0}}\sigma_1(\widehat{\mathcal{T}})\frac{\sqrt{\sigma_1(\hat{M}_2)}}{\sigma_k(\hat{M}_2)}\,\|E_{8,G}\| \\ & + \frac{(\alpha_0+2)}{2\sqrt{(\alpha_0+1)\alpha_0}}\sqrt{\sigma_1(\hat{M}_2) + \|E_{8,G}\|}|n_i|\end{aligned} \tag{140}$$

As before w.h.p $\|E_{6,G}\| = O(\frac{\sqrt{d}}{N}\tau_{\epsilon_6,\delta_6})$.The following bounds hold on $\|Y\|$ and $|n_i|$, because they are a Gaussian vector and variable. In particular, w.h.p. $\|Y\| = O(\frac{k^{5/2}}{N\tilde{\sigma}_k^{3/2}\tilde{\gamma}_s}\tau_{\epsilon_7,\delta_7})$ and $|n_i| = O(\frac{k^2}{N\tilde{\sigma}_k^{3/2}\tilde{\gamma}_s}\tau_{\epsilon_7,\delta_7})$. This leads to the following utility: $O(\frac{\sqrt{\sigma_1(\hat{M}_2)k^{2.5}}}{\tilde{\gamma}_s N\tilde{\sigma}_k^{3/2}}\tau_{\epsilon_7,\delta_7} + \frac{\sqrt{\sigma_1(\hat{M}_2)d}}{\sigma_k(\hat{M}_2)N}\tau_{\epsilon_8,\delta_8} + \sqrt{\sigma_1(\hat{M}_2) + \frac{\sqrt{d}}{N}\tau_{\epsilon_8,\delta_8}}\frac{k^2\tau_{\epsilon_7,\delta_7}}{\tilde{\gamma}_s N\tilde{\sigma}_k^{3/2}})$.

H.4. Perturbation on the final output μ_i, α_i Config. 4 ($e9$): Proof for Theorem 16

In this configuration, we add noise proportional to the output's sensitive

$$\mu_i^{DP} = \mu_i + Z, \text{ where } Z \sim \mathcal{N}(0, \Delta_{\epsilon,\delta}^2 I_k) \tag{141}$$

where $\Delta_{\epsilon,\delta} = \Delta_\mu(D)\tau_{\epsilon_9,\delta_9}$, with $\tau_{\epsilon_9,\delta_9} = \frac{\sqrt{2ln(1.25/\delta_9)}}{\epsilon_9}$. Similar to the previous analysis, since Z is Gaussian, then w.h.p. $\|Z\| = O(\frac{\sqrt{d\sigma_1(\hat{M}_2)}k^2}{N\tilde{\gamma}_s\tilde{\sigma}_k^{3/2}})$. We have the utility $O(\frac{\sqrt{\sigma_1(\hat{M}_2)dk^2}}{N\tilde{\gamma}_s\tilde{\sigma}_k^{3/2}}\tau_{\epsilon_9,\delta_9})$.

I. Some Useful Identities and Theorems

Identity 37 (Square of Sum)**.**

$$\Big(\sum_{i=1}^N a_i\Big)^2 = \sum_{i=1}^N a_i^2 + \sum_{\substack{i=1,j=1 \\ i\neq j}}^{N,N} a_i a_j \tag{142}$$

Identity 38 (Cube of Sum)**.**

$$\Big(\sum_{i=1}^{N} a_i\Big)^3 = \sum_{i=1}^{N} a_i^3 + 3 \sum_{\substack{i=1,j=1 \\ i\neq j}}^{N,N} a_i^2 a_j + \sum_{\substack{i=1,j=1,k=1 \\ i\neq j, j\neq k, k\neq i}}^{N,N,N} a_j a_j a_k \tag{143}$$

Theorem 39 (Weyl's theorem; Theorem 4.11, p. 204 in (Stewart, 1990))**.** . *Let A, E be given $m \times n$ matrices with $m \geq n$, then*

$$\max_{i\in[n]} |\sigma_i(A) - \sigma_i(A+E)| \leq \|E\|_2 \tag{144}$$

Theorem 40 (Bound on the norm of a Gaussian Random Variable)**.** *Let Z be a Gaussian $\mathcal{N}(0,\sigma)$. Then $\mathbb{P}[|Z| \leq t] \geq 1 - 2e^{\frac{-t^2}{2\sigma^2}}$ for all $t > 0$' or alternatively, $\mathbb{P}[|Z| > \sigma\sqrt{2\log(1/\delta)}] \leq \delta$ for all $0 < \delta \leq 1$.*

Theorem 41 (Bound on the norm of a Gaussian Vector)**.** *Let $Y \sim \mathcal{N}(0, \sigma I_k)$, then $\mathbb{P}[\|Y\|_2^2 \geq \sigma^2(k + 2\sqrt{kt} + 2t)] \leq e^{-t}$.*

Proof. The proof is immediate from Theorem 2.1 in (Hsu et al., 2012) with $A = I, \mu = 0$. □

Theorem 42 (Bound on the spectral norm of a Gaussian Matrix (Tao, 2012))**.** *Let $E \in \mathbb{R}^{d\times d}$ be a symmetric Gaussian matrix with elements sampled iid from $\mathcal{N}(0,\sigma)$, then $\mathbb{P}[\|E\|_2 = O(\sqrt{d}\sigma)] \geq 1 - negl(d)$.*

Theorem 43 (Bound on the spectral norm of a Gaussian Tensor (Tomioka & Suzuki, 2014))**.** *Let E be a K^{th} order tensor with each $E_{i_1,\ldots,i_K}$ be sampled i.i.d. from a Gaussian $\mathcal{N}(0,\sigma)$, then $\mathbb{P}[\|E\|_2 \leq \sqrt{8\sigma^2(\sum_{i=1}^{K} d_i)\ln(2K/K_0) + \ln(2/\delta)}] \geq 1 - \delta$, where $K_0 = \ln(3/2)$. Note by extension the bound also holds if the tensor is symmetric as well.*

Lemma 44 (Laplace tail bound)**.** *Let Z be drawn from a Laplace distribution with density $\frac{1}{2b}e^{-\frac{|z|}{b}}$, then $\mathbb{P}(Z \geq t) = \frac{1}{2}e^{-\frac{t}{b}}$ for all $t > 0$, or equivalently $Z \leq b\log(1/(2\delta))$ with probability at least $1 - \delta$ for all $0 < \delta \leq 1$.*