Correction of Theorem

Chanwoo Lee, March 4, 2020

Using the notations in Correction,

$$\mathcal{L}_{\mathcal{Y}}(\hat{\mathcal{Z}}) = \mathcal{L}_{\mathcal{Y}}(\mathcal{Z}^{\text{true}}) + (\mathcal{Z}^{\text{true}} - \hat{\mathcal{Z}})^T \nabla_{\mathcal{Z}} \mathcal{L}_{\mathcal{Y}}(\mathcal{Z}^{\text{true}}) + \frac{1}{2}(\mathcal{Z}^{\text{true}} - \hat{\mathcal{Z}})^T \nabla^2_{\mathcal{Z}} \mathcal{L}_{\mathcal{Y}}(\check{\mathcal{Z}})(\mathcal{Z}^{\text{true}} - \hat{\mathcal{Z}})$$

where, $\check{\mathcal{Z}} = \gamma \mathcal{Z}^{\text{true}} + (1-\gamma)\mathcal{Z}$.

- Score part,

$$\frac{\partial \mathcal{L}_{\mathcal{Y}}}{\partial z_{\omega,\ell}} = \mathbb{1}_{\{y_\omega = \ell\}} \frac{\dot{f}(z)}{f(z) - f(z')}\Bigg|_{(z,\ z')=(z_{\omega,\ell},\ z_{\omega,\ell-1})} - \mathbb{1}_{\{y_\omega = \ell+1\}} \frac{\dot{f}(z)}{f(z'') - f(z)}\Bigg|_{(z'',\ z)=(z_{\omega,\ell+1},\ z_{\omega,\ell})},$$

  This implies,

$$\left|\frac{\partial \mathcal{L}_{\mathcal{Y}}}{\partial z_{\omega,\ell}}\right| \leq C_{\alpha,\beta,\Delta}$$

  Let us define $s_{\omega,\ell} = \frac{\partial \mathcal{L}_{\mathcal{Y}}}{\partial z_{\omega,\ell}}\Big|_{\mathcal{Z}=\mathcal{Z}^{\text{true}}}$ for all (ω,ℓ). Based on the definition of $C_{\alpha,\beta,\Delta}$, $\nabla_{\mathcal{Z}} \mathcal{L}_{\mathcal{Y}}(\mathcal{Z}^{\text{true}}) = [\![s_{\omega,\ell}]\!]$ is a random tensor whose entries are independently distributed satisfying

$$\mathbb{E}(s_{\omega,\ell}) = 0, \quad |s_\omega| \leq C_{\alpha,\beta,Delta} \quad \text{for all } (\omega,\ell) \in [d_1] \times \cdots \times [d_K] \times (L-1). \tag{1}$$

  By lemma 6, with probability at least $1 - \exp(-C_1(\sum_k d_k + L - 1))$, we have

$$\|\nabla_{\mathcal{Z}} \mathcal{L}_{\mathcal{Y}}(\mathcal{Z}^{\text{true}})\|_\sigma \leq C_2 C_{\alpha,\beta,\Delta} \sqrt{\sum_k d_k + L - 1}, \tag{2}$$

  where C_1, C_2 are two positive constants that depend only on K. Furthermore, note that $\text{rank}(\mathcal{Z}) \leq (\boldsymbol{r}+1, 2)$, $\text{rank}(\mathcal{Z}^{\text{true}}) \leq (\boldsymbol{r}+1, 2)$, so $\text{rank}(\mathcal{Z} - \mathcal{Z}^{\text{true}}) \leq 2(\boldsymbol{r}+1, 2)$. By lemma 3, $\|\mathcal{Z} - \mathcal{Z}^{\text{true}}\|_* \leq (2(r_{\max}+1))^{\frac{K}{2}} \|\mathcal{Z} - \mathcal{Z}^{\text{true}}\|_F$. Combining (15), (1) and (2), we have that, with probability at least $1 - \exp(-C_1(\sum_k d_k + L - 1))$,

$$|\langle \text{Vec}(\nabla_{\mathcal{Z}} \mathcal{L}_{\mathcal{Y}}(\mathcal{Z}^{\text{true}})), \text{Vec}(\mathcal{Z} - \mathcal{Z}^{\text{true}}) \rangle| \leq C_2 C_{\alpha,\beta,\Delta} \sqrt{(r_{\max}+1)^K (\sum_k d_k + L - 1)} \|\mathcal{Z} - \mathcal{Z}^{\text{true}}\|_F.$$

- Hessian part,
  For this part, I fail to find good conditions that make the following hold.

$$\sum_\omega \sum_\ell z^2_{\omega,\ell} \mathbb{1}_{\{y_\omega = \ell \text{ or } \ell+1\}} \rightarrow \sum_\omega \sum_\ell z^2_{\omega,\ell} \mathbb{P}(y_\omega = \ell \text{ or } \ell+1).$$

  (If we use CLT, we get trivial lower bound for $\sum_\omega \sum_\ell z^2_{\omega,\ell} \mathbb{1}_{\{y_\omega = \ell \text{ or } \ell+1\}}$).
  Instead, since we know $\nabla^2_{\mathcal{Z}} \mathcal{L}_{\mathcal{Y}}(\check{\mathcal{Z}})$ is negative definite,

$$\lambda = \max_{\gamma \in [0,1]} \lambda_1(\nabla^2_{\mathcal{Z}} \mathcal{L}_{\mathcal{Y}}(\gamma \mathcal{Z}^{\text{true}} + (1-\gamma)\mathcal{Z}))$$

  where, λ_1 is the largest eigen-value. The strict log concavity of $\mathcal{L}_{\mathcal{Y}}(\mathcal{Z})$ implies $\lambda < 0$ ($\because [0,1]$ is compact). This implies that for any $\boldsymbol{z} \in \mathbb{R}^{d_1 \times \cdots d_K \times (L-1)}$, we have

$$\boldsymbol{z}^T \nabla^2_{\mathcal{Z}} \mathcal{L}_{\mathcal{Y}}(\check{\mathcal{Z}}) \boldsymbol{z} \leq \lambda \|\boldsymbol{z}\|_F.$$

Therefore, we can bound the quadratic term.

$$\mathrm{Vec}(\mathcal{Z} - \mathcal{Z}^{\mathrm{true}})^T \nabla^2_{\mathcal{Z}} \mathcal{L}_{\mathcal{Y}}(\check{\mathcal{Z}})\, \mathrm{Vec}(\mathcal{Z} - \mathcal{Z}^{\mathrm{true}}) = \lambda \|\mathcal{Z} - \mathcal{Z}^{\mathrm{true}}\|_F^2.$$

Based on the above arguments, we can have the following bound with high probability.

$$\|\mathcal{Z} - \mathcal{Z}^{\mathrm{true}}\|_F \leq \frac{c C^2_{\alpha,\beta,\Delta} (r_{\max} + 1)^K}{\lambda^2} \left(L - 1 + \sum_k d_k \right).$$