An Adaptive Divergence-based Non-negative Latent Factor Model: Supplementary File

Ye Yuan, Guangxiao Yuan, Renfang Wang, and Xin Luo, *Senior Member, IEEE*

This is the supplementary file for the paper entitled *An Adaptive Divergence-based Non-negative Latent Factor Model*. It mainly contains the convergence analysis of an ADNLF model.

Supplementary Convergence Analysis

1) *Proof of Lemma* 1

Considering the gradient of (16) on $p_{u,\cdot}$, we have:

$$\nabla\varepsilon_{u,i}\left(p_{u,\cdot}\right)=-\frac{1}{\alpha}\left(r_{u,i}^{\alpha}\left(p_{u,\cdot}q_{i,\cdot}^{\mathrm{T}}\right)^{\beta-1}q_{i,\cdot}-\left(p_{u,\cdot}q_{i,\cdot}^{\mathrm{T}}\right)^{\alpha+\beta-1}q_{i,\cdot}\right)+\lambda p_{u,\cdot}\Rightarrow$$
$$\left\|\nabla\varepsilon_{u,i}\left(p_{u,\cdot}\right)\right\|_2=\left\|-\frac{1}{\alpha}\left(r_{u,i}^{\alpha}\left(p_{u,\cdot}q_{i,\cdot}^{\mathrm{T}}\right)^{\beta-1}-\left(p_{u,\cdot}q_{i,\cdot}^{\mathrm{T}}\right)^{\alpha+\beta-1}\right)q_{i,\cdot}+\lambda p_{u,\cdot}\right\|_2 \tag{S1}$$

According to the L_2-norm's characteristics [65], we have the following inequality based on (S1)

$$\left\|\nabla\varepsilon_{u,i}\left(p_{u,\cdot}\right)\right\|_2\leq\frac{1}{\alpha}\left(r_{u,i}^{\alpha}\left(p_{u,\cdot}q_{i,\cdot}^{\mathrm{T}}\right)^{\beta-1}-\left(p_{u,\cdot}q_{i,\cdot}^{\mathrm{T}}\right)^{\alpha+\beta-1}\right)\left\|q_{i,\cdot}\right\|_2+\lambda\left\|p_{u,\cdot}\right\|_2 \tag{S2}$$

Considering the bridging function given in (5), each element in $p_{u,\cdot}$ and $q_{i,\cdot}$ is constrained in the [0, 1) interval. Thus, the following condition holds:

$$\begin{aligned}&0\leq\left\|p_{u,\cdot}\right\|_2<\sqrt{f},\ 0\leq\left\|q_{i,\cdot}\right\|_2<\sqrt{f},\\&\Rightarrow 0\leq M=p_{u,\cdot}q_{i,\cdot}^{\mathrm{T}}<f.\end{aligned} \tag{S3}$$

Note that in (S3) M denotes the constant achieved by the product $p_{u,\cdot}q_{i,\cdot}^{\mathrm{T}}$. By combining (S2) and (S3), we infer that:

$$\left\|\nabla\varepsilon_{u,i}\left(p_{u,\cdot}\right)\right\|_2<\frac{1}{\alpha}\left(r_{u,i}^{\alpha}f^{\beta-1}-f^{\alpha+\beta-1}\right)\sqrt{f}+\lambda\sqrt{f}. \tag{S4}$$

Hence, ***Lemma* 1** holds.□

2) *Proof of Lemma* 2

Assuming that $p_{g,\cdot}$ and $p_{u,\cdot}$ are two arbitrary row vectors of P, following the principle of a directional gradient, we have:

$$\nabla\varepsilon_{u,i}\left(p_{g,\cdot}\right)-\nabla\varepsilon_{u,i}\left(p_{u,\cdot}\right)=\nabla^2\varepsilon_{u,i}\left(p_{u,\cdot}\right)\left(p_{g,\cdot}-p_{u,\cdot}\right)^{\mathrm{T}}. \tag{S5}$$

Considering the L_2-norm of (S1), we have:

$$\begin{aligned}\left\|\nabla\varepsilon_{u,i}\left(p_{g,\cdot}\right)-\nabla\varepsilon_{u,i}\left(p_{u,\cdot}\right)\right\|_2&=\left\|\nabla^2\varepsilon_{u,i}\left(p_{u,\cdot}\right)\left(p_{g,\cdot}-p_{u,\cdot}\right)^{\mathrm{T}}\right\|_2\\&\leq\left\|\nabla^2\varepsilon_{u,i}\left(p_{u,\cdot}\right)\right\|_2\left\|\left(p_{g,\cdot}-p_{u,\cdot}\right)\right\|_2.\end{aligned} \tag{S6}$$

By combing (16) and (S3), we achieve the following equation:

$$\begin{aligned}&\nabla^2\varepsilon_{u,i}\left(p_{u,\cdot}\right)\\&=\frac{1}{\alpha}\left(\left(\alpha+\beta-1\right)\left(p_{u,\cdot}q_{i,\cdot}^{\mathrm{T}}\right)^{\alpha}-r_{u,i}^{\alpha}\left(\beta-1\right)\right)\left(p_{u,\cdot}q_{i,\cdot}^{\mathrm{T}}\right)^{\beta-2}q_{i,\cdot}^{\mathrm{T}}q_{i,\cdot}+\lambda E_f\\&=\frac{1}{\alpha}\left(\left(\alpha+\beta-1\right)\left(M\right)^{\alpha}-r_{u,i}^{\alpha}\left(\beta-1\right)\right)\left(M\right)^{\beta-2}q_{i,\cdot}^{\mathrm{T}}q_{i,\cdot}+\lambda E_f\\&=Nq_{i,\cdot}^{\mathrm{T}}q_{i,\cdot}+\lambda E_f,\end{aligned} \tag{S7}$$

Note that the last two steps of (S7) are achieved based on (S3), and N is computed as:

$$N=\frac{1}{\alpha}\left(\left(\alpha+\beta-1\right)\left(M\right)^{\alpha}-r_{u,i}^{\alpha}\left(\beta-1\right)\right)\left(M\right)^{\beta-2}. \tag{S8}$$

Based on (S7), we reformulate (S6) as:

$$\left\|\nabla\varepsilon_{u,i}\left(p_{g,\cdot}\right)-\nabla\varepsilon_{u,i}\left(p_{u,\cdot}\right)\right\|_2\leq\left\|Nq_{i,\cdot}^{\mathrm{T}}q_{i,\cdot}+\lambda E_f\right\|_2\left\|\left(p_{g,\cdot}-p_{u,\cdot}\right)\right\|_2, \tag{S9}$$

where $\left\|Nq_{i,\cdot}^{\mathrm{T}}q_{i,\cdot}+\lambda E_f\right\|_2$ is the largest singular value of $\left(Nq_{i,\cdot}^{\mathrm{T}}q_{i,\cdot}+\lambda E_f\right)$. Note that we have $L=\left\|Nq_{i,\cdot}^{\mathrm{T}}q_{i,\cdot}+\lambda E_f\right\|_2$. Therefore, ***Lemma* 2** holds.□

3) *Proof of Lemma* 3

Based on the principle of Taylor-series, we have the following deduction:

$$\begin{aligned}\varepsilon_{u,i}\left(p_{g,\cdot}\right)-\varepsilon_{u,i}\left(p_{u,\cdot}\right)\approx&\nabla\varepsilon_{u,i}\left(p_{u,\cdot}\right)\left(p_{g,\cdot}-p_{u,\cdot}\right)^{\mathrm{T}}\\&+\frac{1}{2}\left(p_{g,\cdot}-p_{u,\cdot}\right)\nabla^2\varepsilon_{u,i}\left(p_{u,\cdot}\right)\left(p_{g,\cdot}-p_{u,\cdot}\right)^{\mathrm{T}}.\end{aligned} \tag{S10}$$

According to ***Definitions* 5**, the strong-convexity of $\varepsilon_{u,i}$ depends on the following condition:

$$\varepsilon_{u,i}\left(p_{g,\cdot}\right)-\varepsilon_{u,i}\left(p_{u,\cdot}\right)\geq\nabla\varepsilon_{u,i}\left(p_{u,\cdot}\right)\left(p_{g,\cdot}-p_{u,\cdot}\right)^{\mathrm{T}}+\frac{1}{2}\rho\left\|p_{g,\cdot}-p_{u,\cdot}\right\|_2^2. \tag{S11}$$

Based on (S10) and (S11), to make *Lemma* 3 hold, ρ should fulfill the following condition:

$$\begin{aligned}&\left(p_{g,\cdot}-p_{u,\cdot}\right)\nabla^2\varepsilon_{u,i}\left(p_{u,\cdot}\right)\left(p_{g,\cdot}-p_{u,\cdot}\right)^{\mathrm{T}}\geq\rho\left\|p_{g,\cdot}-p_{u,\cdot}\right\|_2^2\Rightarrow\\&\left(p_{g,\cdot}-p_{u,\cdot}\right)\left(Nq_{i,\cdot}^{\mathrm{T}}q_{i,\cdot}+\lambda E_f-\rho E_f\right)\left(p_{g,\cdot}-p_{u,\cdot}\right)^{\mathrm{T}}\geq 0.\end{aligned} \tag{S12}$$

where the last inference is achieved based on (S7). Note that (S12) holds when the term ($Nq_{i,\cdot}^{\mathrm{T}}q_{i,\cdot}+\lambda E_f-\rho E_f$) is positive semi-definite. As discussed in prior studies [65, 66], ($Nq_{i,\cdot}^{\mathrm{T}}q_{i,\cdot}+\lambda E_f-\rho E_f$) is a positive semi-definite matrix when ρ is the minimum singular value of the matrix ($Nq_{i,\cdot}^{\mathrm{T}}q_{i,\cdot}+\lambda E_f$). Hence,

Lemma **3** holds.□

4) *Proof of ADNLF's Convergence via Lemmas* **1** ***to*** **3**

Afterwards, we consider the t-th training iteration of an ADNLF model, $p_{u,\cdot}$ is updated with the following learning scheme to minimize the instance loss $\varepsilon_{u,i}$:

$$p_{u,\cdot}^{\tau} \leftarrow p_{u,\cdot}^{\tau-1} - \eta^{t-1} \nabla \varepsilon_{u,i}\left(p_{u,\cdot}^{\tau-1}\right). \tag{S13}$$

Let $p_{u,\cdot}^{*}$ be the optimal state of $p_{u,\cdot}$, then we have:

$$\left\| p_{u,\cdot}^{\tau} - p_{u,\cdot}^{*} \right\|_2^2 = \left\| p_{u,\cdot}^{\tau-1} - \eta^{t-1} \nabla \varepsilon_{u,i}\left(p_{u,\cdot}^{\tau-1}\right) - p_{u,\cdot}^{*} \right\|_2^2. \tag{S14}$$

By expanding (S14), we have the following deduction:

$$\begin{aligned} \left\| p_{u,\cdot}^{\tau} - p_{u,\cdot}^{*} \right\|_2^2 = & \left\| p_{u,\cdot}^{\tau-1} - p_{u,\cdot}^{*} \right\|_2^2 - 2\eta^{t-1} \nabla \varepsilon_{u,i}\left(p_{u,\cdot}^{\tau-1}\right)\left(p_{u,\cdot}^{\tau-1} - p_{u,\cdot}^{*}\right) \\ & + \left(\eta_j^{t-1}\right)^2 \left\| \nabla \varepsilon_{u,i}\left(p_{u,\cdot}^{\tau-1}\right) \right\|_2^2. \end{aligned} \tag{S15}$$

According to ***Lemma*** **3** and ***Definition*** **5,** based on (S15), the following inequality is achieved:

$$\begin{aligned} \varepsilon_{u,i}\left(p_{u,\cdot}^{*}\right) \ge & \varepsilon_{u,i}\left(p_{u,\cdot}^{\tau-1}\right) + \nabla \varepsilon_{u,i}\left(p_{u,\cdot}^{\tau-1}\right)\left(p_{u,\cdot}^{*} - p_{u,\cdot}^{\tau-1}\right)^{\mathrm{T}} \\ & + \frac{1}{2}\rho \left\| p_{u,\cdot}^{*} - p_{u,\cdot}^{\tau-1} \right\|_2^2. \end{aligned} \tag{S16}$$

Since $p_{u,\cdot}^{*}$ is the optimal state of $p_{u,\cdot}$, we naturally have:

$$\varepsilon_{u,i}\left(p_{u,\cdot}^{*}\right) < \varepsilon_{u,i}\left(p_{u,\cdot}^{\tau-1}\right). \tag{S17}$$

By substituting (S17) into (S16), we achieve that:

$$\nabla \varepsilon_{u,i}\left(p_{u,\cdot}^{\tau-1}\right)\left(p_{u,\cdot}^{\tau-1} - p_{u,\cdot}^{*}\right)^{\mathrm{T}} \ge \frac{1}{2}\rho \left\| p_{u,\cdot}^{*} - p_{u,\cdot}^{\tau-1} \right\|_2^2. \tag{S18}$$

Based on (S15) and (S18), we further infer that:

$$\left\| p_{u,\cdot}^{\tau} - p_{u,\cdot}^{*} \right\|_2^2 \le \left(1 - \eta^{t-1}\rho\right) \left\| p_{u,\cdot}^{\tau-1} - p_{u,\cdot}^{*} \right\|_2^2 + \left(\eta_j^{t-1}\right)^2 \left\| \nabla \varepsilon_{u,i}\left(p_{u,\cdot}^{\tau-1}\right) \right\|_2^2. \tag{S19}$$

By taking the expectations on both sides of (S19), we have:

$$\begin{aligned} \mathrm{E}\left[\left\| p_{u,\cdot}^{\tau} - p_{u,\cdot}^{*} \right\|_2^2\right] \le & \left(1 - \eta^{t-1}\rho\right) \mathrm{E}\left[\left\| p_{u,\cdot}^{\tau-1} - p_{u,\cdot}^{*} \right\|_2^2\right] \\ & + \left(\eta_j^{t-1}\right)^2 \mathrm{E}\left[\left\| \nabla \varepsilon_{u,i}\left(p_{u,\cdot}^{\tau-1}\right) \right\|_2^2\right]. \end{aligned} \tag{S20}$$

According to ***Lemma*** **1**, we see that $\left\| \nabla \varepsilon_{u,i}\left(p_{u,\cdot}^{\tau-1}\right) \right\|_2$ has the upper bound. Assume that a positive number c enables the following inequality:

$$\mathrm{E}\left[\left\| \nabla \varepsilon_{u,i}\left(p_{u,\cdot}^{\tau-1}\right) \right\|_2^2\right] \le c^2. \tag{S21}$$

Thus, based on (S21), (S20) is equivalent to:

$$\mathrm{E}\left[\left\| p_{u,\cdot}^{\tau} - p_{u,\cdot}^{*} \right\|_2^2\right] \le \left(1 - \eta^{t-1}\rho\right) \mathrm{E}\left[\left\| p_{u,\cdot}^{\tau-1} - p_{u,\cdot}^{*} \right\|_2^2\right] + \left(\eta^{t-1}\right)^2 c^2. \tag{S22}$$

Let $\eta^{t-1} = \gamma/(\rho t)$ for a constant γ>1, then (S22) is rewritten as:

$$\mathrm{E}\left[\left\| p_{u,\cdot}^{\tau} - p_{u,\cdot}^{*} \right\|_2^2\right] \le \left(1 - \frac{\gamma}{t}\right) \mathrm{E}\left[\left\| p_{u,\cdot}^{\tau-1} - p_{u,\cdot}^{*} \right\|_2^2\right] + \frac{1}{t^2}\left(\frac{c\gamma}{\rho}\right)^2. \tag{S23}$$

By expanding the iterative expression of (S23), we see that it satisfies the following condition:

$$\mathrm{E}\left[\left\| p_{u,\cdot}^{\tau} - p_{u,\cdot}^{*} \right\|_2^2\right] \le \frac{1}{t} \max\left\{ \left\| p_{u,\cdot}^{1} - p_{u,\cdot}^{*} \right\|_2^2, \frac{\gamma^2 c^2}{\rho\gamma - 1} \right\}, \tag{S24}$$

where $p_{u,\cdot}^{1}$ denotes the initial state of $p_{u,\cdot}$ at the t-th iteration. According to ***Lemma*** **2**, with the L-smooth property, we have:

$$\varepsilon_{u,i}\left(p_{u,\cdot}^{\tau}\right) - \varepsilon_{u,i}\left(p_{u,\cdot}^{*}\right) - \nabla \varepsilon_{u,i}\left(p_{u,\cdot}^{*}\right)\left(p_{u,\cdot}^{\tau} - p_{u,\cdot}^{*}\right) \le \frac{L}{2}\left\| p_{u,\cdot}^{*} - p_{u,\cdot}^{\tau} \right\|_2^2. \tag{S25}$$

Since $p_{u,\cdot}^{*}$ is the optimal state of $p_{u,\cdot}$, it has the following property:

$$\nabla \varepsilon_{u,i}\left(p_{u,\cdot}^{*}\right) = 0, \tag{S26}$$

which enables the following reformulation of (S25):

$$\varepsilon_{u,i}\left(p_{u,\cdot}^{\tau}\right) - \varepsilon_{u,i}\left(p_{u,\cdot}^{*}\right) \le \frac{L}{2}\left\| p_{u,\cdot}^{*} - p_{u,\cdot}^{\tau} \right\|_2^2. \tag{S27}$$

Considering the expectation of (S27), we have:

$$E\left[\varepsilon_{u,i}\left(p_{u,\cdot}^{\tau}\right) - \varepsilon_{u,i}\left(p_{u,\cdot}^{*}\right)\right] \le \frac{L}{2} E\left[\left\| p_{u,\cdot}^{*} - p_{u,\cdot}^{\tau} \right\|_2^2\right]. \tag{S28}$$

By substituting (S24) into (S28), we deduce that:

$$E\left[\varepsilon_{u,i}\left(p_{u,\cdot}^{\tau}\right) - \varepsilon_{u,i}\left(p_{u,\cdot}^{*}\right)\right] \le \frac{L}{2t} W(\gamma), \tag{S29}$$

where the term $W(\gamma)$ is given as:

$$W(\gamma) = \max\left\{ \left\| p_{u,\cdot}^{1} - p_{u,\cdot}^{*} \right\|_2^2, \frac{\gamma^2 c^2}{\rho\gamma - 1} \right\}. \tag{S30}$$

Note that $\forall u \in U$, (S29) can be expanded into:

$$\mathrm{E}\left[\sum_{r_{u,i} \in \Lambda(u)} \left(\varepsilon_{u,i}\left(p_{u,\cdot}^{\tau}\right) - \varepsilon_{u,i}\left(p_{u,\cdot}^{*}\right)\right)\right] \le |\Lambda(u)| \frac{L}{2t} W(\gamma), \tag{S31}$$

where $\Lambda(u)$ denotes the subset of Λ related to $u \in U$, and we naturally have $\sum_{u \in U} |\Lambda(u)| = |\Lambda|$. Hence, (S31) can be extended into the following inequality:

$$\mathrm{E}\left[\sum_{u \in U} \sum_{r_{u,i} \in \Lambda(u)} \left(\varepsilon_{u,i}\left(p_{u,\cdot}^{\tau}\right) - \varepsilon_{u,i}\left(p_{u,\cdot}^{*}\right)\right)\right] \le |\Lambda| \frac{L}{2t} W(\gamma). \tag{S32}$$

Note that (S32) indicates that $|\Lambda| \frac{L}{2t} W(\gamma) \to 0$ when $t \to \infty$.

Note that the convergence of ADNLF with the update of $q_{i,\cdot}$ can be achieved in the same way.

Based on the above inferences, we see that the convergence of ADNLF is theoretically guaranteed.■