

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:

$$\begin{aligned}\nabla \varepsilon_{u,i}(p_{u,\cdot}) &= -\frac{1}{\alpha} \left(r_{u,i}^\alpha (p_{u,\cdot} q_{i,\cdot}^\top)^{\beta-1} q_{i,\cdot} - (p_{u,\cdot} q_{i,\cdot}^\top)^{\alpha+\beta-1} q_{i,\cdot} \right) + \lambda p_{u,\cdot} \Rightarrow \\ \|\nabla \varepsilon_{u,i}(p_{u,\cdot})\|_2 &= \left\| -\frac{1}{\alpha} \left(r_{u,i}^\alpha (p_{u,\cdot} q_{i,\cdot}^\top)^{\beta-1} - (p_{u,\cdot} q_{i,\cdot}^\top)^{\alpha+\beta-1} \right) q_{i,\cdot} + \lambda p_{u,\cdot} \right\|_2\end{aligned}\quad (S1)$$

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

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

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:

$$\begin{aligned}0 \leq \|p_{u,\cdot}\|_2 < \sqrt{f}, \quad 0 \leq \|q_{i,\cdot}\|_2 < \sqrt{f}, \\ \Rightarrow 0 \leq M = p_{u,\cdot} q_{i,\cdot}^\top < f.\end{aligned}\quad (S3)$$

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

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

Hence, **Lemma 1** holds. \square

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:

$$\nabla \varepsilon_{u,i}(p_{g,\cdot}) - \nabla \varepsilon_{u,i}(p_{u,\cdot}) = \nabla^2 \varepsilon_{u,i}(p_{u,\cdot}) (p_{g,\cdot} - p_{u,\cdot})^\top. \quad (S5)$$

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

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

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

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

Based on (S3), the last two steps of (S7) are obtained, and N is computed as:

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

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

$$\left\| \nabla \varepsilon_{u,i}(p_{g,\cdot}) - \nabla \varepsilon_{u,i}(p_{u,\cdot}) \right\|_2 \leq \|N q_{i,\cdot} q_{i,\cdot}^\top + \lambda E_f\|_2 \left\| (p_{g,\cdot} - p_{u,\cdot}) \right\|_2, \quad (S9)$$

where $\|N q_{i,\cdot} q_{i,\cdot}^\top + \lambda E_f\|_2$ denotes the largest singular value of $(N q_{i,\cdot} q_{i,\cdot}^\top + \lambda E_f)$. Note that we have $L = \|N q_{i,\cdot} q_{i,\cdot}^\top + \lambda E_f\|_2$. Therefore, **Lemma 2** holds. \square

3) Proof of Lemma 3

Based on the principle of Taylor-series, we achieve the following equation:

$$\begin{aligned}\varepsilon_{u,i}(p_{g,\cdot}) - \varepsilon_{u,i}(p_{u,\cdot}) &\approx \nabla \varepsilon_{u,i}(p_{u,\cdot}) (p_{g,\cdot} - p_{u,\cdot})^\top \\ &\quad + \frac{1}{2} (p_{g,\cdot} - p_{u,\cdot}) \nabla^2 \varepsilon_{u,i}(p_{u,\cdot}) (p_{g,\cdot} - p_{u,\cdot})^\top.\end{aligned}\quad (S10)$$

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

$$\varepsilon_{u,i}(p_{g,\cdot}) - \varepsilon_{u,i}(p_{u,\cdot}) \geq \nabla \varepsilon_{u,i}(p_{u,\cdot}) (p_{g,\cdot} - p_{u,\cdot})^\top + \frac{1}{2} \rho \|p_{g,\cdot} - p_{u,\cdot}\|_2^2. \quad (S11)$$

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

$$\begin{aligned}(p_{g,\cdot} - p_{u,\cdot}) \nabla^2 \varepsilon_{u,i}(p_{u,\cdot}) (p_{g,\cdot} - p_{u,\cdot})^\top &\geq \rho \|p_{g,\cdot} - p_{u,\cdot}\|_2^2 \Rightarrow \\ (p_{g,\cdot} - p_{u,\cdot}) (N q_{i,\cdot} q_{i,\cdot}^\top + \lambda E_f - \rho E_f) (p_{g,\cdot} - p_{u,\cdot})^\top &\geq 0.\end{aligned}\quad (S12)$$

where the last inference is achieved based on (S7). Note that (S12) holds when the term $(N q_{i,\cdot} q_{i,\cdot}^\top + \lambda E_f - \rho E_f)$ is positive semi-definite. As discussed in prior studies [65, 66], $(N q_{i,\cdot} q_{i,\cdot}^\top + \lambda E_f)$

$q_{i..} + \lambda E_f - \rho E_f$ is a positive semi-definite matrix when ρ is the minimum singular value of the matrix $(Nq_{i..}^\top q_{i..} + \lambda E_f)$. Hence, **Lemma 3** holds. \square

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

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

$$p_{u..}^\tau \leftarrow p_{u..}^{\tau-1} - \eta^{t-1} \nabla \varepsilon_{u,i}(p_{u..}^{\tau-1}). \quad (\text{S13})$$

Let $p_{u..}^*$ be the optimal state of $p_{u..}$, then we achieve:

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

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

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

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

$$\begin{aligned} \varepsilon_{u,i}(p_{u..}^*) &\geq \varepsilon_{u,i}(p_{u..}^{\tau-1}) + \nabla \varepsilon_{u,i}(p_{u..}^{\tau-1})(p_{u..}^* - p_{u..}^{\tau-1})^\top \\ &\quad + \frac{1}{2} \rho \|p_{u..}^* - p_{u..}^{\tau-1}\|_2^2. \end{aligned} \quad (\text{S16})$$

Since $p_{u..}^*$ stands for the optimal state of $p_{u..}$, we naturally obtain:

$$\varepsilon_{u,i}(p_{u..}^*) < \varepsilon_{u,i}(p_{u..}^{\tau-1}). \quad (\text{S17})$$

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

$$\nabla \varepsilon_{u,i}(p_{u..}^{\tau-1})(p_{u..}^{\tau-1} - p_{u..}^*)^\top \geq \frac{1}{2} \rho \|p_{u..}^* - p_{u..}^{\tau-1}\|_2^2. \quad (\text{S18})$$

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

$$\|p_{u..}^\tau - p_{u..}^*\|_2^2 \leq (1 - \eta^{t-1} \rho) \|p_{u..}^{\tau-1} - p_{u..}^*\|_2^2 + (\eta_j^{t-1})^2 \|\nabla \varepsilon_{u,i}(p_{u..}^{\tau-1})\|_2^2. \quad (\text{S19})$$

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

$$\begin{aligned} \mathbb{E}[\|p_{u..}^\tau - p_{u..}^*\|_2^2] &\leq (1 - \eta^{t-1} \rho) \mathbb{E}[\|p_{u..}^{\tau-1} - p_{u..}^*\|_2^2] \\ &\quad + (\eta_j^{t-1})^2 \mathbb{E}[\|\nabla \varepsilon_{u,i}(p_{u..}^{\tau-1})\|_2^2]. \end{aligned} \quad (\text{S20})$$

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

$$\mathbb{E}[\|\nabla \varepsilon_{u,i}(p_{u..}^{\tau-1})\|_2^2] \leq c^2. \quad (\text{S21})$$

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

$$\mathbb{E}[\|p_{u..}^\tau - p_{u..}^*\|_2^2] \leq (1 - \eta^{t-1} \rho) \mathbb{E}[\|p_{u..}^{\tau-1} - p_{u..}^*\|_2^2] + (\eta^{t-1})^2 c^2. \quad (\text{S22})$$

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

$$\mathbb{E}[\|p_{u..}^\tau - p_{u..}^*\|_2^2] \leq \left(1 - \frac{\gamma}{t}\right) \mathbb{E}[\|p_{u..}^{\tau-1} - p_{u..}^*\|_2^2] + \frac{1}{t^2} \left(\frac{c\gamma}{\rho}\right)^2. \quad (\text{S23})$$

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

$$\mathbb{E}[\|p_{u..}^\tau - p_{u..}^*\|_2^2] \leq \frac{1}{t} \max \left\{ \|p_{u..}^1 - p_{u..}^*\|_2^2, \frac{\gamma^2 c^2}{\rho\gamma - 1} \right\}, \quad (\text{S24})$$

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

$$\varepsilon_{u,i}(p_{u..}^\tau) - \varepsilon_{u,i}(p_{u..}^*) - \nabla \varepsilon_{u,i}(p_{u..}^*)(p_{u..}^\tau - p_{u..}^*) \leq \frac{L}{2} \|p_{u..}^* - p_{u..}^\tau\|_2^2. \quad (\text{S25})$$

Since $p_{u..}^*$ denotes the optimal state of $p_{u..}$, it has the following property:

$$\nabla \varepsilon_{u,i}(p_{u..}^*) = 0, \quad (\text{S26})$$

which enables the following reformulation of (S25):

$$\varepsilon_{u,i}(p_{u..}^\tau) - \varepsilon_{u,i}(p_{u..}^*) \leq \frac{L}{2} \|p_{u..}^* - p_{u..}^\tau\|_2^2. \quad (\text{S27})$$

Considering the expectation of (S27), we have:

$$\mathbb{E}[\varepsilon_{u,i}(p_{u..}^\tau) - \varepsilon_{u,i}(p_{u..}^*)] \leq \frac{L}{2} \mathbb{E}[\|p_{u..}^* - p_{u..}^\tau\|_2^2]. \quad (\text{S28})$$

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

$$\mathbb{E}[\varepsilon_{u,i}(p_{u..}^\tau) - \varepsilon_{u,i}(p_{u..}^*)] \leq \frac{L}{2t} W(\gamma), \quad (\text{S29})$$

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

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

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

$$\mathbb{E} \left[\sum_{r_{u,i} \in \Lambda(u)} (\varepsilon_{u,i}(p_{u..}^\tau) - \varepsilon_{u,i}(p_{u..}^*)) \right] \leq |\Lambda(u)| \frac{L}{2t} W(\gamma), \quad (\text{S31})$$

where $\Lambda(u)$ is 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:

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

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

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

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