Supplementary Material for ECCV Submission: Toward Faster and Simpler Matrix Normalization via Rank-1 Update

Tan Yu, Yunfeng Cai, and Ping Li

Cognitive Computing Lab, Baidu Research
{tanyu01,caiyunfeng,liping11}@baidu.com

Abstract. In this supplementary material, we provide the proof for Theorem 1 of the main manuscript and derive the backward propagation of the proposed RUN algorithm. Meanwhile, we release the codes of the proposed algorithm in the attachment for for re-implementation.

1 Proof of Theorem 1

Theorem 1. *Let* $\mathbf{B}_K$ *be obtained accordingly in the main manuscript, where* $\mathbf{v}_0 \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$*. Then the expectation of* $\mathbf{B}_K$ *is given by*

$$\mathbb{E}(\mathbf{B}_K) = \mathbf{U}\,\mathrm{diag}([\sigma_1(1-\epsilon\alpha_1), \cdots, \sigma_D(1-\epsilon\alpha_D))\mathbf{U}^\top, \tag{1}$$

where $1 \geq \alpha_1 \geq \alpha_2 \geq \cdots \geq \alpha_D$*.*

Proof. Recall that $\mathbf{B}_K = \mathbf{B} - \epsilon\mathbf{R}_K$, where

$$\mathbf{R}_K = \mathbf{B}\mathbf{v}_K\mathbf{v}_K^\top/\|\mathbf{v}_K\|_2^2, \tag{2}$$

Using SVD, we factorize

$$\mathbf{B} = \mathbf{U}\mathbf{\Sigma}\mathbf{U}^\top, \tag{3}$$

where $\mathbf{U}$ is orthonormal containing the singular vectors and $\mathbf{\Sigma} = \mathrm{diag}(\sigma_1, \cdots, \sigma_D)$ is a diagonal matrix containing singular values. Based on the iteration $\mathbf{v}_k = \mathbf{B}\mathbf{v}_{k-1}$, we have

$$\mathbf{v}_K = \mathbf{B}^K\mathbf{v}_0 = \mathbf{U}\mathbf{\Sigma}^K\mathbf{U}^\top\mathbf{v}_0 = \mathbf{U}\mathbf{\Sigma}^K\mathbf{a}, \tag{4}$$

where $\mathbf{a} = \mathbf{U}^\top\mathbf{v}_0$. Plugging Eq. (3) and Eq. (4) into Eq. (2), we have

$$\mathbf{R}_K = \frac{\mathbf{U}\mathbf{\Sigma}^{K+1}\mathbf{a}\mathbf{a}^\top\mathbf{\Sigma}^K\mathbf{U}^\top}{\mathbf{a}^\top\mathbf{\Sigma}^{2K}\mathbf{a}} = \mathbf{U}\mathbf{H}\mathbf{U}^\top, \tag{5}$$

where

$$\mathbf{H} = (\mathbf{\Sigma}^{K+1}\mathbf{a}\mathbf{a}^\top\mathbf{\Sigma}^K)/(\mathbf{a}^\top\mathbf{\Sigma}^{2K}\mathbf{a}). \tag{6}$$

As $\mathbf{v}_0 \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ and $\mathbf{U}\mathbf{U}^\top = \mathbf{I}$, thus $\mathbf{a} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$. That is, $\mathbf{a}$'s entries $\{a_1, \cdots, a_D\}$ are *i.i.d* random variables with normal distribution.

We first prove that the expectation of each off-diagonal entry of $\mathbf{H}$ is 0. That is, $\mathbb{E}(\mathbf{H})$ is a diagonal matrix. We define $H_{i,j}$ as the entry in i-th row and j-th column of the matrix $\mathbf{H}$ where $i \neq j$. According to the definition of $\mathbf{H}$ in Eq. (6),

$$H_{i,j} = \frac{\sigma_i^{K+1}\sigma_j^K a_i a_j}{\sum_{l=1}^D a_l^2 \sigma_l^{2K}}. \tag{7}$$

We define $f(a_1, \cdots, a_D)$ as the probability density function. Thus

$$\mathbb{E}(H_{i,j}) = \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} f(a_1, \cdots, a_D) H_{i,j} da_1 \cdots da_D, \tag{8}$$

Since

$$\begin{aligned} H_{i,j}(a_1, \cdots, a_i, \cdots, a_D) &= -H_{i,j}(a_1, \cdots, -a_i, \cdots, a_D), \\ f(a_1, \cdots, a_i, \cdots, a_D) &= f(a_1, \cdots, -a_i, \cdots, a_D), \end{aligned} \tag{9}$$

that is, $H_{i,j}$ is an odd function with respect to a_i and $f(a_1, \cdots, a_D)$ is an even function with respect to a_i, it is straightforward to obtain $\mathbb{E}(H_{i,j}) = 0$.

Since we have proved that $\mathbb{E}(\mathbf{H})$ is a diagonal matrix, we rewrite $\mathbb{E}(\mathbf{H}) = \mathrm{diag}(h_1, \cdots, h_D)$ and

$$h_l = \mathbb{E}(\sigma_l (a_l \sigma_l^k)^2 / \sum_{i=1}^D (a_i \sigma_i^k)^2) = \sigma_l \alpha_l, \tag{10}$$

where

$$\alpha_l = \mathbb{E}((a_l \sigma_l^k)^2 / \sum_{i=1}^D (a_i \sigma_i^k)^2). \tag{11}$$

In this case, proving Theorem 1 is equivalent to proving that $\alpha_s \geq \alpha_t$ if $s < t$. As we know

$$\alpha_s - \alpha_t = \mathbb{E}\Big(\frac{a_s^2 \sigma_s^{2k} - a_s^2 \sigma_t^{2k}}{\sum_{i=1}^D a_i^2 \sigma_i^{2k}}\Big). \tag{12}$$

We define $b_i = a_i^2$ and $y_i = \sigma_i^{2k}$, then seek to prove

$$\alpha_s - \alpha_t = \mathbb{E}\Big(\frac{b_s y_s - b_t y_t}{\sum_{i=1}^D b_i y_i}\Big) \geq 0, \quad \text{if } s < t. \tag{13}$$

As $y_s \geq y_t$ and $y_1 \geq y_2 \cdots \geq y_D$, we obtain

$$\frac{b_s y_s - b_t y_t}{\sum_{i=1}^D b_i y_i} \geq \frac{y_t}{y_1} \frac{b_s - b_t}{\sum_{i=1}^D b_i}. \tag{14}$$

Thus,

$$\mathbb{E}\Big(\frac{b_s y_s - b_t y_t}{\sum_{i=1}^D b_i y_i}\Big) \geq \frac{y_t}{y_1} \mathbb{E}\Big(\frac{b_s - b_t}{\sum_{i=1}^D b_i}\Big). \tag{15}$$

Since $\{a_i\}_1^D$ are *i.i.d*, $\{b_i\}_1^D$ are also *i.i.d*. Therefore,

$$\mathbb{E}\Big(\frac{b_s - b_t}{\sum_{i=1}^D b_i}\Big) = \mathbb{E}\Big(\frac{b_s}{\sum_{i=1}^D b_i}\Big) - \mathbb{E}\Big(\frac{b_t}{\sum_{i=1}^D b_i}\Big) = 0. \tag{16}$$

Plugging Eq. (16) into Eq. (15), we obtain

$$\mathbb{E}\Big(\frac{b_s y_s - b_t y_t}{\sum_{i=1}^D b_i y_i}\Big) \geq 0, \tag{17}$$

that is, $\alpha_s \geq \alpha_t$ if $s < t$. This completes the proof.

2 Back-propagation Derivation of RUN

We compute the differentiation of $\mathbf{F}_K$ based on $\mathbf{F}_K = \mathbf{F} - \eta \mathbf{F}\mathbf{v}_K\mathbf{v}_K^\top / \|\mathbf{v}_K\|_2^2$:

$$\begin{aligned} d\mathbf{F}_K = d\mathbf{F} - \eta \frac{(d\mathbf{F})\mathbf{v}_K\mathbf{v}_K^\top + \mathbf{F}(d\mathbf{v}_K)\mathbf{v}_K^\top + \mathbf{F}\mathbf{v}_K(d\mathbf{v}_K^\top)}{\mathbf{v}_K^\top\mathbf{v}_K} \\ + \eta \frac{(d\mathbf{v}_K^\top)\mathbf{v}_K + \mathbf{v}_K^\top d\mathbf{v}_K}{(\mathbf{v}_K^\top\mathbf{v}_K)^2}\mathbf{F}\mathbf{v}_K\mathbf{v}_K^\top. \end{aligned} \tag{18}$$

Meanwhile, the iteration $\mathbf{v}_k = \mathbf{F}^\top\mathbf{F}\mathbf{v}_{k-1}$ leads to

$$\mathbf{v}_K = (\mathbf{F}^\top\mathbf{F})^K\mathbf{v}_0. \tag{19}$$

Since $\mathbf{v}_0$ is a constant vector, based on Eq. (19), we obtain

$$\begin{aligned} d\mathbf{v}_K &\equiv K(\mathbf{F}^\top\mathbf{F})^{K-1}[(d\mathbf{F}^\top)\mathbf{F} + \mathbf{F}^\top d\mathbf{F}]\mathbf{v}_0, \\ d\mathbf{v}_K^\top &\equiv K\mathbf{v}_0^\top[(d\mathbf{F}^\top)\mathbf{F} + \mathbf{F}^\top d\mathbf{F}](\mathbf{F}^\top\mathbf{F})^{K-1}. \end{aligned} \tag{20}$$

Plugging Eq. (20) in Eq. (18), we obtain

$$d\mathbf{F}_K = \sum_{i=0}^{4} l_i^1(\mathbf{F})d\mathbf{F}r_i^1(\mathbf{F}) + \sum_{j=1}^{4} l_j^2(\mathbf{F})(d\mathbf{F})^\top r_j^2(\mathbf{F}), \tag{21}$$

where $\{l_i^1(\mathbf{F}), r_i^1(\mathbf{F})\}_{i=0}^4$ and $\{l_i^2(\mathbf{F}), r_i^2(\mathbf{F})\}_{i=1}^4$ are

$$\begin{aligned}
&l_0^1(\mathbf{F}) = \mathbf{I}, \quad r_0^1(\mathbf{F}) = \mathbf{I} - \eta \mathbf{v}_K \mathbf{v}_K^\top / (\mathbf{v}_K^\top \mathbf{v}_K), \\
&l_1^1(\mathbf{F}) = \frac{-\eta K \mathbf{F}(\mathbf{F}^\top \mathbf{F})^{K-1}\mathbf{F}^\top}{\mathbf{v}_K^\top \mathbf{v}_N}, \quad r_1^2(\mathbf{F}) = \mathbf{v}_0 \mathbf{v}_K^\top, \\
&l_2^1(\mathbf{F}) = \frac{-\eta K \mathbf{F}\mathbf{v}_K \mathbf{v}_0^\top \mathbf{F}^\top}{\mathbf{v}_K^\top \mathbf{v}_N}, \quad r_2^1(\mathbf{F}) = (\mathbf{F}^\top \mathbf{F})^{K-1}, \\
&l_3^1(\mathbf{F}) = \frac{\eta K \mathbf{v}_0^\top \mathbf{F}^\top}{(\mathbf{v}_K^\top \mathbf{v}_K)^2}, \quad r_3^1(\mathbf{F}) = (\mathbf{F}^\top \mathbf{F})^{K-1}\mathbf{v}_K \mathbf{F} \mathbf{v}_K \mathbf{v}_K^\top, \\
&l_4^1(\mathbf{F}) = \frac{\eta K \mathbf{v}_K^\top (\mathbf{F}^\top \mathbf{F})^{K-1}\mathbf{F}^\top}{(\mathbf{v}_K^\top \mathbf{v}_K)^2}, \quad r_4^1(\mathbf{F}) = \mathbf{v}_0 \mathbf{F} \mathbf{v}_K \mathbf{v}_K^\top, \\
&l_1^2(\mathbf{F}) = \frac{-\eta K \mathbf{F}(\mathbf{F}^\top \mathbf{F})^{K-1}}{\mathbf{v}_K^\top \mathbf{v}_K}, \quad r_1^2(\mathbf{F}) = \mathbf{F}\mathbf{v}_K \mathbf{v}_K^\top, \\
&l_2^2(\mathbf{F}) = \frac{-\eta K \mathbf{F}\mathbf{v}_K \mathbf{v}_0^\top}{\mathbf{v}_K^\top \mathbf{v}_K}, \quad r_2^2(\mathbf{F}) = \mathbf{F}(\mathbf{F}^\top \mathbf{F})^{K-1}, \\
&l_3^2(\mathbf{F}) = \frac{\eta K \mathbf{v}_0^\top}{(\mathbf{v}_K^\top \mathbf{v}_K)^2}, \quad r_3^1(\mathbf{F}) = \mathbf{F}(\mathbf{F}^\top \mathbf{F})^{K-1}\mathbf{v}_K \mathbf{F} \mathbf{v}_K \mathbf{v}_K^\top, \\
&l_4^2(\mathbf{F}) = \frac{\eta K \mathbf{v}_K^\top (\mathbf{F}^\top \mathbf{F})^{K-1}}{(\mathbf{v}_K^\top \mathbf{v}_K)^2}, \quad r_4^2(\mathbf{F}) = \mathbf{F}\mathbf{v}_0 \mathbf{F} \mathbf{v}_K \mathbf{v}_K^\top.
\end{aligned} \tag{22}$$

According to the definition,

$$dL \equiv \mathrm{vec}(\frac{\partial L}{\partial \mathbf{F}})^\top \mathrm{vec}(d\mathbf{F}) \equiv \mathrm{vec}(\frac{\partial L}{\partial \bar{\mathbf{F}}_K})^\top \mathrm{vec}(d\bar{\mathbf{F}}_K). \tag{23}$$

Since $\mathrm{trace}(\mathbf{A}\mathbf{B}^\top) \equiv \mathrm{vec}(\mathbf{A})^\top \mathrm{vec}(\mathbf{B})$, we further obtain

$$\mathrm{trace}(d\mathbf{F}^\top \frac{\partial L}{\partial \mathbf{F}}) \equiv \mathrm{trace}[d\bar{\mathbf{F}}_K^\top \frac{\partial L}{\partial \bar{\mathbf{F}}_K}] \tag{24}$$

Plugging Eq. (21) into Eq. (24), we obtain

$$\begin{aligned}
\mathrm{trace}(d\mathbf{F}^\top \frac{\partial L}{\partial \mathbf{F}}) &\equiv \mathrm{trace}\Big\{\big[\sum_{i=1}^{5} l_i^1(\mathbf{F}) d\mathbf{F} r_i^1(\mathbf{F}) + \sum_{j=1}^{4} l_j^2(\mathbf{F})(d\mathbf{F})^\top r_j^2(\mathbf{F})\big]^\top \frac{\partial L}{\partial \bar{\mathbf{F}}_K}\Big\} \\
&\equiv \mathrm{trace}\Big\{d\mathbf{F}^\top\big[\sum_{i=1}^{5} l_i^1(\mathbf{F})^\top \frac{\partial L}{\partial \bar{\mathbf{F}}_K} r_i^1(\mathbf{F})^\top + \sum_{j=1}^{4} r_j^2(\mathbf{F})(\frac{\partial L}{\partial \bar{\mathbf{F}}_K})^\top l_j^2(\mathbf{F})\big]\Big\}.
\end{aligned} \tag{25}$$

Compare the LHS and RHS of Eq. (25), we obtain

$$\frac{\partial L}{\partial \mathbf{F}} = \big[\sum_{i=1}^{5} l_i^1(\mathbf{F})^\top \frac{\partial L}{\partial \bar{\mathbf{F}}_K} r_i^1(\mathbf{F})^\top + \sum_{j=1}^{4} r_j^2(\mathbf{F})(\frac{\partial L}{\partial \bar{\mathbf{F}}_K})^\top l_j^2(\mathbf{F})\big]. \tag{26}$$

Eq. (26) gives the backward path which takes $\partial L / \partial \mathbf{F}_K$ as input and outputs $\partial L / \partial \mathbf{F}$.