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 $B_K$ be obtained accordingly in the main manuscript, where $v_0 \sim \mathcal{N}(0, I)$. Then the expectation of $B_K$ is given by

$$E(B_K) = U \text{diag}([\sigma_1(1 - \epsilon \alpha_1), \cdots, \sigma_D(1 - \epsilon \alpha_D)]) U^T,$$

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

**Proof.** Recall that $B_K = B - \epsilon R_K$, where

$$R_K = Bv_K v_K^T / \|v_K\|^2_2.$$  \hfill (2)

Using SVD, we factorize

$$B = U \Sigma U^T,$$  \hfill (3)

where $U$ is orthonormal containing the singular vectors and $\Sigma = \text{diag}(\sigma_1, \cdots, \sigma_D)$ is a diagonal matrix containing singular values. Based on the iteration $v_k = Bv_{k-1}$, we have

$$v_K = B^K v_0 = U \Sigma^K v_0 = U \Sigma^K a,$$  \hfill (4)

where $a = U^T v_0$. Plugging Eq. (3) and Eq. (4) into Eq. (2), we have

$$R_K = \frac{U \Sigma^{K+1} a a^T \Sigma^K U^T}{a^T \Sigma^{2K} a} = U H U^T,$$  \hfill (5)

where

$$H = (\Sigma^{K+1} a a^T \Sigma^K) / (a^T \Sigma^{2K} a).$$  \hfill (6)

As $v_0 \sim \mathcal{N}(0, I)$ and $UU^T = I$, thus $a \sim \mathcal{N}(0, I)$. That is, $a$'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 $H$ is 0. That is, $\mathbb{E}(H)$ is a diagonal matrix. We define $H_{i,j}$ as the entry in $i$-th row and $j$-th column of the matrix $H$ where $i \neq j$. According to the definition of $H$ in Eq. (6),

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

(7)

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

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

(8)

Since

$$H_{i,j}(a_1, \cdots, a_i, \cdots, a_D) = -H_{i,j}(a_1, \cdots, -a_i, \cdots, a_D),$$

$$f(a_1, \cdots, a_i, \cdots, a_D) = f(a_1, \cdots, -a_i, \cdots, a_D),$$

(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 $\mathbb{E}(H)$ is a diagonal matrix, we rewrite $\mathbb{E}(H) = \text{diag}(h_1, \cdots, h_D)$ and

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

(10)

where

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

(11)

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

As we know

$$\alpha_s - \alpha_t = \mathbb{E} \left( \frac{a^2 \sigma_s^{2k} - a^2 \sigma_t^{2k}}{\sum_{i=1}^{D} a_i^2 \sigma_i^{2k}} \right).$$  

(12)

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

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

(13)

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

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

(14)

Thus,

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

(15)
supplementary material 3

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

\[
E\left( \frac{b_s - b_t}{\sum_{i=1}^D b_i} \right) = E\left( \frac{b_s}{\sum_{i=1}^D b_i} \right) - E\left( \frac{b_t}{\sum_{i=1}^D b_i} \right) = 0. \tag{16}
\]

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

\[
E\left( \frac{b_s y_s - b_t y_t}{\sum_{i=1}^D b_i y_i} \right) \geq 0, \tag{17}
\]

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 \( F_K \) based on \( F_K = F - \eta Fv_K v_K^T / \|v_K\|^2 \):

\[
dF_K = dF - \eta \frac{(dF)v_K v_K^T + F(dv_K)v_K^T + Fv_K(dv_K^T)}{v_K^T v_K} v_K
+ \eta \frac{(dv_K^T) v_K + v_K^T dv_K}{(v_K^T v_K)^2} Fv_K v_K^T. \tag{18}
\]

Meanwhile, the iteration \( v_k = F^T F v_{k-1} \) leads to

\[
v_K = (F^T F)^K v_0. \tag{19}
\]

Since \( v_0 \) is a constant vector, based on Eq. (19), we obtain

\[
dv_K \equiv K(F^T F)^K \left[ (dF^T)F + F^T dF \right] v_0, \\
dv_K^T \equiv K v_0^T \left[ (dF^T)F + F^T dF \right] (F^T F)^K \tag{20}
\]

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

\[
dF_K = \sum_{i=0}^4 l_i^1(F) dF r_i^1(F) + \sum_{j=1}^4 l_j^2(F) (dF)^T r_j^2(F), \tag{21}
\]
where \( \{ l_1^i(F), r_1^i(F) \}_{i=0}^4 \) and \( \{ l_2^i(F), r_2^i(F) \}_{i=1}^4 \) are
\[
\begin{align*}
l_0^0(F) &= I, \quad r_0^0(F) = I - \eta K v_v v_v^T / (v_v v_v^T), \\
l_1^1(F) &= \frac{-\eta K F (F^T F)^{-1} F^T}{v_v v_v^T}, \quad r_1^1(F) = v_0 v_v^T, \\
l_2^2(F) &= \frac{-\eta K F v_v v_v^T}{v_v^T}, \quad r_2^2(F) = (F^T F)^{-1}, \\
l_3^3(F) &= \frac{\eta K v_0 F^T}{(v_v v_v^T)^2}, \quad r_3^3(F) = (F^T F)^{-1} v_v F v_v^T, \\
l_4^4(F) &= \frac{\eta K v_0^2 F}{(v_v v_v^T)^2}, \quad r_4^4(F) = v_0 F v_v v_v^T.
\end{align*}
\]

According to the definition,
\[
dL \equiv \text{vec}(\frac{\partial L}{\partial F})^T \text{vec}(dF) \equiv \text{vec}(\frac{\partial L}{\partial F_K})^T \text{vec}(dF_K).
\]  \hfill (23)

Since \( \text{trace}(A B^T) \equiv \text{vec}(A)^T \text{vec}(B) \), we further obtain
\[
\text{trace}(dF_k^T \frac{\partial L}{\partial F_k}) \equiv \text{trace}[dF_k^T \frac{\partial L}{\partial F_k}]
\]  \hfill (24)

Plugging Eq. (21) into Eq. (24), we obtain
\[
\text{trace}(dF_k^T \frac{\partial L}{\partial F_k}) \equiv \text{trace}\left\{ \left[ \sum_{i=1}^5 l_1^i(F) dF r_1^i(F) + \sum_{j=1}^4 l_2^j(F) (dF_k)^T r_2^j(F) \right] \frac{\partial L}{\partial F_k} \right\}
\]  \hfill (25)

Compare the LHS and RHS of Eq. (25), we obtain
\[
\frac{\partial L}{\partial F} = \left[ \sum_{i=1}^5 l_1^i(F)^T \frac{\partial L}{\partial F_K} r_1^i(F)^T + \sum_{j=1}^4 r_2^j(F) (\frac{\partial L}{\partial F_K})^T l_2^j(F) \right].
\]  \hfill (26)

Eq. (26) gives the backward path which takes \( \partial L / \partial F_K \) as input and outputs \( \partial L / \partial F \).