Supplemental Material: Unsupervised Feature Extraction Inspired by Latent Low-Rank Representation

Yaming Wang, Vlad I. Morariu, Larry S. Davis
University of Maryland, College Park, MD, 20742, USA
{wym, morariu, lsd}@umiacs.umd.edu

1. Proof of Proposition 1

Proposition 1. *During the iteration procedure described in Algorithm 1, Z_k and L_k always keep the form*

$$Z_k = V_X W_{Zk} V_X^{\mathrm{T}}, \quad L_k = U_X W_{Lk} U_X^{\mathrm{T}}, \tag{1}$$

$\forall$ positive integer k, where W_{Zk} and W_{Lk} are $r \times r$ diagonal matrices containing the singular values.

Furthermore, the orthogonal matrices involved in the SVD decomposition of any matrix in Algorithm 1 must be U_X or V_X, depending on its shape.

Proof. Proof by induction.

X has skinny svd decomposition

$$X = U_X \Sigma_X V_X^T, \tag{2}$$

where $\Sigma_X = \mathrm{diag}\{\sigma_{Xi}\}$ is an $r \times r$ diagonal matrix containing nonzero singular values. In the rest of the proof, the size of all the diagonal matrices is $r \times r$.

(i) When $n = 1$, it follows from Algorithm 1 that

$$\begin{aligned}
J_1 &= 0 = V_X \mathrm{diag}\left\{0\right\} V_X^{\mathrm{T}}, \\
S_1 &= 0 = U_X \mathrm{diag}\left\{0\right\} U_X^{\mathrm{T}}, \\
Z_1 &= V_X \mathrm{diag}\left\{\frac{\sigma_{Xi}^2}{1+\sigma_{Xi}^2}\right\} V_X^{\mathrm{T}}, \\
L_1 &= U_X \mathrm{diag}\left\{\frac{\sigma_{Xi}^2}{(1+\sigma_{Xi}^2)^2}\right\} U_X^{\mathrm{T}}, \\
Y_{11} &= U_X \mathrm{diag}\left\{\frac{\mu_0 \sigma_{Xi}}{(1+\sigma_{Xi}^2)^2}\right\} V_X^{\mathrm{T}}, \\
Y_{21} &= \mu_0 Z_1 = V_X \mathrm{diag}\left\{\frac{\mu_0 \sigma_{Xi}^2}{1+\sigma_{Xi}^2}\right\} V_X^{\mathrm{T}}, \\
Y_{31} &= \mu_0 L_1 = U_X \mathrm{diag}\left\{\frac{\mu_0 \sigma_{Xi}^2}{(1+\sigma_{Xi}^2)^2}\right\} U_X^{\mathrm{T}}.
\end{aligned}$$

Therefore the conclusion holds for $n = 1$.

(ii) Assume the conclusion holds for $n = k$, *i.e.*

$$\begin{aligned}
J_k &= V_X \mathrm{diag}\left\{j_{ki}\right\} V_X^{\mathrm{T}}, \\
S_k &= U_X \mathrm{diag}\left\{s_{ki}\right\} U_X^{\mathrm{T}}, \\
Z_k &= V_X \mathrm{diag}\left\{z_{ki}\right\} V_X^{\mathrm{T}}, \\
L_k &= U_X \mathrm{diag}\left\{l_{ki}\right\} U_X^{\mathrm{T}}, \\
Y_{1k} &= U_X \mathrm{diag}\left\{y_{1ki}\right\} V_X^{\mathrm{T}}, \\
Y_{2k} &= V_X \mathrm{diag}\left\{y_{2ki}\right\} V_X^{\mathrm{T}}, \\
Y_{3k} &= U_X \mathrm{diag}\left\{y_{3ki}\right\} U_X^{\mathrm{T}}.
\end{aligned}$$

When $n = k + 1$, it follows that

$$\begin{aligned}
J_{k+1} &= \mathrm{argmin}_J \frac{1}{\mu_k} \|J\|_* + \frac{1}{2} \|J - (Z_k + Y_{2k}/\mu_k)\|_{\mathrm{F}}^2 \\
&= V_X \mathrm{diag}\left\{j_{(k+1)i}\right\} V_X^{\mathrm{T}} \\
S_{k+1} &= \mathrm{argmin}_S \frac{1}{\mu_k} \|S\|_* + \frac{1}{2} \|S - (L_k + Y_{3k}/\mu_k)\|_{\mathrm{F}}^2 \\
&= U_X \mathrm{diag}\left\{s_{(k+1)i}\right\} U_X^{\mathrm{T}} \\
Z_{k+1} &= (I + X^{\mathrm{T}} X)^{-1} (X^{\mathrm{T}} (X - L_k X) + J_{k+1} \\
&\quad + (X^{\mathrm{T}} Y_{1k} - Y_{2k})/\mu_k) \\
&= V_X \mathrm{diag}\left\{z_{(k+1)i}\right\} V_X^{\mathrm{T}}, \\
L_{k+1} &= ((X - X Z_{k+1}) X^{\mathrm{T}} + S_{k+1} \\
&\quad + (Y_{1k} X^{\mathrm{T}} - Y_{3k})/\mu_k)(I + X X^{\mathrm{T}})^{-1} \\
&= U_X \mathrm{diag}\left\{l_{(k+1)i}\right\} U_X^{\mathrm{T}}, \\
Y_{1(k+1)} &= Y_{1k} + \mu_k (X - X Z_{k+1} - L_{k+1} X) \\
&= U_X \mathrm{diag}\left\{y_{1(k+1)i}\right\} V_X^{\mathrm{T}}, \\
Y_{2(k+1)} &= Y_{2k} + \mu_k (Z_{k+1} - J_{k+1}) \\
&= V_X \mathrm{diag}\left\{y_{2(k+1)i}\right\} V_X^{\mathrm{T}}, \\
Y_{3(k+1)} &= Y_{3k} + \mu_k (L_{k+1} - S_{k+1}) \\
&= U_X \mathrm{diag}\left\{y_{3(k+1)i}\right\} U_X^{\mathrm{T}},
\end{aligned}$$

where

$$
\begin{aligned}
j_{(k+1)i} &= \max\left\{0, z_{ki} + \frac{y_{2ki}}{\mu_k} - \frac{1}{\mu_k}\right\} \\
s_{(k+1)i} &= \max\left\{0, l_{ki} + \frac{y_{3ki}}{\mu_k} - \frac{1}{\mu_k}\right\} \\
z_{(k+1)i} &= \frac{1}{1+\sigma_{Xi}^2}(\sigma_{Xi}^2(1-l_{ki}) + j_{(k+1)i}) \\
&\quad + \frac{\sigma_{Xi}y_{1ki} - y_{2ki}}{(1+\sigma_{Xi}^2)\mu_k} \\
l_{(k+1)i} &= \frac{1}{1+\sigma_{Xi}^2}(\sigma_{Xi}^2(1-z_{(k+1)i}) + s_{(k+1)i}) \\
&\quad + \frac{\sigma_{Xi}y_{1ki} - y_{3ki}}{(1+\sigma_{Xi}^2)\mu_k} \\
y_{1(k+1)i} &= y_{1ki} + \mu_k\sigma_{Xi}(1 - z_{(k+1)i} - l_{(k+1)i}) \\
y_{2(k+1)i} &= y_{2ki} + \mu_k(z_{(k+1)i} - j_{(k+1)i}) \\
y_{3(k+1)i} &= y_{3ki} + \mu_k(l_{(k+1)i} - s_{(k+1)i})
\end{aligned}
$$

From (i) and (ii), we know that the conclusion holds for arbitrary positive integer k. □

2. Proof of Proposition 2

Proposition 2. *Assuming ρ is relatively large, when μ_k is small, the iteration procedure is approximately equivalent to*

$$z_{k+1} = \alpha(1-l_k), \quad l_{k+1} = \alpha(1-z_{k+1}), \tag{3}$$

and the solution after the k^{th} iteration is as follows.

$$z_k = \frac{\alpha+\alpha^{2k}}{1+\alpha}, \quad l_k = \alpha(1-z_k) = \frac{\alpha^3+\alpha^{2k+1}}{1+\alpha}, \tag{4}$$

where α is defined by

$$\alpha = \frac{\sigma_X^2}{1+\sigma_X^2} = 1 - \frac{1}{1+\sigma_X^2}, \tag{5}$$

Proof. The approximation process is included in the paper. This proof solves recurrence sequence (3).

From (3), we have

$$z_1 = \alpha(1-l_0) = \alpha, \tag{6}$$

and

$$
\begin{aligned}
z_{k+1} &= \alpha(1-l_k) \\
&= \alpha\left[1-\alpha(1-z_k)\right] \\
&= \alpha(1-\alpha) + \alpha^2 z_k,
\end{aligned} \tag{7}
$$

for all $k \geq 1$, which is equivalent to

$$z_{k+1} - \frac{\alpha}{1+\alpha} = \alpha^2\left(z_k - \frac{\alpha}{1+\alpha}\right).$$

Therefore,

$$
\begin{aligned}
z_k &= \frac{\alpha}{1+\alpha} + \alpha^{2(k-1)}\left(z_1 - \frac{\alpha}{1+\alpha}\right) \\
&= \frac{\alpha+\alpha^{2k}}{1+\alpha}, \\
l_k &= \alpha(1-z_k) = \frac{\alpha^3+\alpha^{2k+1}}{1+\alpha}.
\end{aligned}
$$

□

3. Proof of Proposition 3

Proposition 3. *Assuming ρ is relatively large, when μ_k is large, the iteration procedure is approximately equivalent to*

$$
\begin{aligned}
z_{k+1} &= \alpha(1-l_k) + (1-\alpha)z_k, \\
l_{k+1} &= \alpha(1-z_{k+1}) + (1-\alpha)l_k,
\end{aligned} \tag{8}
$$

When the iteration terminates, the final results z and l satisfy

$$z \to \frac{(1-\alpha)z_{k_0} - l_{k_0} + 1}{2-\alpha}, \quad l \to 1 - z. \tag{9}$$

where α is defined by (5), and k_0 is some starting point from which the large-μ_k condition holds.

Proof. The approximation process is included in the paper. This proof solves (8).

(8) is equivalent to

$$z_{k+1} + l_k - 1 = (1-\alpha)(l_k + z_k - 1) \tag{10}$$

$$l_{k+1} + z_{k+1} - 1 = (1-\alpha)(z_{k+1} + l_k - 1) \tag{11}$$

Then

$$l_{k+1} + z_{k+1} - 1 = (1-\alpha)^2(l_k + z_k - 1) \tag{12}$$

for all $k \geq k_0$, where k_0 is some point of the large-μ_k case.

Define $\beta = l_{k_0} + z_{k_0} - 1$, and (12) (10) gives

$$l_k + z_k - 1 = (1-\alpha)^{2(k-k_0)}\beta, \tag{13}$$

$$l_k + z_{k+1} - 1 = (1-\alpha)^{2(k-k_0)+1}\beta. \tag{14}$$

Subtract (13) from (14) and it follows that

$$z_{k+1} - z_k = (1-\alpha)^{2(k-k_0)+1} - (1-\alpha)^{2(k-k_0)}, \tag{15}$$

which is equivalent to

$$
\begin{aligned}
&z_{k+1} - \frac{\beta}{2-\alpha}(1-\alpha)^{2(k+1-k_0)} \\
=&z_k - \frac{\beta}{2-\alpha}(1-\alpha)^{2(k-k_0)}
\end{aligned} \tag{16}
$$

for all $k \geq k_0$.

From (16), we can obtain the expression of z_k as

$$z_k = z_{k_0} - \frac{\beta}{2-\alpha} + \frac{\beta}{2-\alpha}(1-\alpha)^{2(k-k_0)}. \tag{17}$$

Plug (17) into (13) and it follows that

$$l_k = 1 - z_k + (1-\alpha)^{2(k-k_0)}\beta \tag{18}$$

Since $|1-\alpha| = 1/(1+\sigma_X^2) < 1$ in (17) and (18), let $k \to +\infty$ and we conclude that

$$z \to z_{k0} - \frac{\beta}{2-\alpha} = \frac{(1-\alpha)z_{k_0} - l_{k_0} + 1}{2-\alpha}, \tag{19}$$

$$l \to 1 - z. \tag{20}$$

When $k \to +\infty$, we have $1 - l_k - z_k \to 1 - l - z = 0$, $z_k - j_k = z_k - z_{k-1} \to 0$, and $l_k - s_k = l_k - l_{k-1} \to 0$. Then the stopping criterion in Step 7 of Algorithm 1 will be satisfied, which terminates the iteration. □