Understanding the Sparsity: Augmented Matrix Factorization with Sampled Constraints on Unobservables
(Supplementary Material)

CIKM 2014, Nov. 3-7, Shanghai, China

Yongfeng Zhang†, Min Zhang†, Yi Zhang†, Yiqun Liu†, Shaoping Ma†
†Department of Computer Science & Technology, Tsinghua University, Beijing, 100084, China
‡School of Engineering, University of California, Santa Cruz, CA 95060, USA
zhangyf07@gmail.com, \{z-m,yiqunliu,msp\}@tsinghua.edu.cn, yiz@soe.ucsc.edu

I. Computation of the Gradients of the Lagrangian Function

The Lagrangian function is:
\[
L(U, V) = \|U\|_F^2 + \|V\|_F^2 + \Lambda \|A(X) - b\|^2_2
\] (1)

where \(X = UV'\). Considering the fact that \(A = \{A_1, A_2, \cdots, A_p\}\), the Lagrangian function could be reformulated as:
\[
L(U, V) = \|U\|_F^2 + \|V\|_F^2 + \Lambda \sum_{i=1}^{p} ((X, A_i) - b_i)^2
\] (2)

As a result, we have the following:
\[
\frac{\partial L}{\partial U} = 2U + \Lambda \sum_{i=1}^{p} 2(\text{tr}(A_i'X) - b_i) \frac{\partial \text{tr}(A_i'X)}{\partial U}
\]
\[
= 2U + \Lambda \sum_{i=1}^{p} 2(\text{tr}(A_i'X) - b_i) \frac{\partial \text{tr}(A_i'X)}{\partial X} \frac{\partial X}{\partial U}
\]
\[
= 2U + \Lambda \sum_{i=1}^{p} 2(\text{tr}(A_i'X) - b_i) A_i V
\]
\[
= 2 \left( U + \Lambda \left( \sum_{i=1}^{p} (\text{tr}(A_i'X) - b_i) A_i \right) \right) V
\] (3)
Similarly:

\[
\frac{\partial \mathcal{L}}{\partial V} = 2V + \Lambda \sum_{i=1}^{p} 2\left( \text{tr}(A'_iX) - b_i \right) \frac{\partial \text{tr}(A'_iX)}{\partial V} \\
= 2V + \Lambda \sum_{i=1}^{p} 2\left( \text{tr}(A'_iX) - b_i \right) \frac{\partial \text{tr}(A'_iX)}{\partial X} \frac{\partial X}{\partial V} \\
= 2V + \Lambda \sum_{i=1}^{p} 2\left( \text{tr}(A'_iX) - b_i \right) A'_iU \\
= 2 \left( V + \Lambda \sum_{i=1}^{p} \left( \text{tr}(A'_iX) - b_i \right) A'_i \right) U \\
\]

(4)

Note that the Chain Rule is applicable to Eq.(3) and Eq.(4) because of the linear relationship between \( A_i \) and \( X \). More rigorously, we give following derivation.

Let \( A \in \mathbb{R}^{m \times n}, U \in \mathbb{R}^{m \times r}, V \in \mathbb{R}^{n \times r} \), and let \( y = \langle A, UV' \rangle \), then:

\[
\frac{\partial \langle A, UV' \rangle}{\partial U} = \frac{\partial y}{\partial U} = \left[ \frac{\partial y}{\partial u_{ij}} \right]_{m \times r} 
\]

(5)

Let \( U = [U'_1 U'_2 \cdots U'_m]' \) and \( V = [V'_1 V'_2 \cdots V'_n]' \), where \( U_i \) and \( V_i \) are the i-th row vectors of \( U \) and \( V \), respectively. Then the function \( y = \langle A, UV' \rangle \) can be expanded in the following way:

\[
y = \langle A, UV' \rangle = \sum_{i,j} a_{ij} U_i V'_j = \sum_{i,j} a_{ij} \sum_{k} u_{ik} v_{jk} = \sum_{i,j} \sum_{k} a_{ij} u_{ik} v_{jk} 
\]

(6)

As a result:

\[
\frac{\partial y}{\partial u_{ik}} = \sum_{j} a_{ij} v_{jk} = A_i \tilde{V}_k
\]

(7)

where \( A_i \) is the i-th row vector of \( A \), and \( \tilde{V}_k \) is the k-th column vector of \( V \). As a result, the scalar-to-matrix partial deviation can be derived in the following way:

\[
\frac{\partial y}{\partial U} = \left[ \frac{\partial y}{\partial u_{ij}} \right]_{m \times r} = \left[ A_i \tilde{V}_j \right]_{m \times r} = AV
\]

(8)

which gives the same result as that of Eq.(3), and Eq.(4) can be derived in a similar way.

II. Derivation of the Updating Rules of the Optimization Problem

According to the above section, we have the following gradients of \( U \) and \( V \):

\[
\nabla_U = U + \Lambda \left( \sum_{i=1}^{p} (\text{tr}(A'_i UV') - b_i) A_i \right) V \\
\nabla_V = V + \Lambda \left( \sum_{i=1}^{p} (\text{tr}(A'_i UV') - b_i) A'_i \right) U 
\]

(9)

Now we conduct linear search for \( U \) on the direction given by \( \nabla_U \), which means that \( U \) could
be updated as $U \leftarrow U + \gamma \nabla_U$, and the Lagrangian function is reformulated as:

$$
\varphi(\gamma) = \|U + \gamma \nabla_U\|_F^2 + \|V\|_F^2 + \Lambda \sum_{i=1}^p \left( \langle (U + \gamma \nabla_U) V', A_i \rangle - b_i \right)^2 
$$

$$
= \text{tr}((U + \gamma \nabla_U)'(U + \gamma \nabla_U)) + \|V\|_F^2 + \Lambda \sum_{i=1}^p \left( \text{tr}(A_i'(U + \gamma \nabla_U)V') - b_i \right)^2 
$$

$$
= \text{tr}(U'U) + 2\gamma \text{tr}(\nabla_U U) + \gamma^2 \text{tr}(\nabla_U \nabla_U) + \|V\|_F^2 + \Lambda \sum_{i=1}^p \left( \text{tr}(A_i'UV') + \gamma \text{tr}(A_i'\nabla_U V') - b_i \right)^2 
$$

As a result, the derivative in terms of $\gamma$ is:

$$
\varphi'(\gamma) = 2 \text{tr}(\nabla_U U) + 2\gamma \text{tr}(\nabla_U \nabla_U) + \Lambda \sum_{i=1}^p 2 \left( \text{tr}(A_i'UV') + \gamma \text{tr}(A_i'\nabla_U V') - b_i \right) \text{tr}(A_i'\nabla_U V') 
$$

$$
= 2 \left\{ \text{tr}(\nabla_U U) + \gamma \text{tr}(\nabla_U \nabla_U) + \Lambda \sum_{i=1}^p \text{tr}(A_i'\nabla_U V') \left( \text{tr}(A_i'UV') + \gamma \text{tr}(A_i'\nabla_U V') - b_i \right) \right\} 
$$

$$
= 2 \left\{ \left( \text{tr}(\nabla_U U) + \Lambda \sum_{i=1}^p \text{tr}(A_i'\nabla_U V')(\text{tr}(A_i'UV') - b_i) \right) + \gamma \left( \text{tr}(\nabla_U \nabla_U) + \Lambda \sum_{i=1}^p \text{tr}^2(A_i'\nabla_U V') \right) \right\} 
$$

Let $\varphi'(\gamma) = 0$, we then have the step size $\gamma_U$ for $U$ as:

$$
\gamma_U = -\frac{\text{tr}(\nabla_U U) + \Lambda \sum_{i=1}^p \text{tr}(A_i'\nabla_U V')(\text{tr}(A_i'UV') - b_i)}{\text{tr}(\nabla_U U) + \Lambda \sum_{i=1}^p \text{tr}^2(A_i'\nabla_U V')} 
$$

and the corresponding updating rule for $U$ is:

$$
U \leftarrow U + \gamma_U \nabla_U 
$$

Similarly, the step size for updating $V$ is:

$$
\gamma_V = -\frac{\text{tr}(\nabla_V V) + \Lambda \sum_{i=1}^p \text{tr}(A_i'U\nabla_V')(\text{tr}(A_i'UV') - b_i)}{\text{tr}(\nabla_V V) + \Lambda \sum_{i=1}^p \text{tr}^2(A_i'U\nabla_V')} 
$$

and the corresponding updating rule for $V$ is:

$$
V \leftarrow V + \gamma_V \nabla_V 
$$

III. Reference