# 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:

$$\mathcal{L}(U,V) = \|U\|_F^2 + \|V\|_F^2 + \Lambda\|\mathcal{A}(X) - b\|_2^2 \tag{1}$$

where $X = UV'$. Considering the fact that $\mathcal{A} = \{A_1, A_2, \cdots, A_p\}$, the Lagrangian function could be reformulated as:

$$\begin{aligned}\mathcal{L}(U,V) &= \|U\|_F^2 + \|V\|_F^2 + \Lambda\sum_{i=1}^{p}\big(\langle X, A_i\rangle - b_i\big)^2 \\ &= \|U\|_F^2 + \|V\|_F^2 + \Lambda\sum_{i=1}^{p}\big(\operatorname{tr}(A_i'X) - b_i\big)^2\end{aligned} \tag{2}$$

As a result, we have the following:

$$\begin{aligned}\frac{\partial\mathcal{L}}{\partial U} &= 2U + \Lambda\sum_{i=1}^{p}2\big(\operatorname{tr}(A_i'X) - b_i\big)\frac{\partial\operatorname{tr}(A_i'X)}{\partial U} \\ &= 2U + \Lambda\sum_{i=1}^{p}2\big(\operatorname{tr}(A_i'X) - b_i\big)\frac{\partial\operatorname{tr}(A_i'X)}{\partial X}\frac{\partial X}{\partial U} \\ &= 2U + \Lambda\sum_{i=1}^{p}2\big(\operatorname{tr}(A_i'X) - b_i\big)A_iV \\ &= 2\left(U + \Lambda\left(\sum_{i=1}^{p}\big(\operatorname{tr}(A_i'X) - b_i\big)A_i\right)V\right)\end{aligned} \tag{3}$$

Similarly:

$$\begin{aligned}\frac{\partial \mathcal{L}}{\partial V} &= 2V + \Lambda \sum_{i=1}^{p} 2\big(\operatorname{tr}(A_i'X) - b_i\big)\frac{\partial \operatorname{tr}(A_i'X)}{\partial V} \\ &= 2V + \Lambda \sum_{i=1}^{p} 2\big(\operatorname{tr}(A_i'X) - b_i\big)\frac{\partial \operatorname{tr}(A_i'X)}{\partial X}\frac{\partial X}{\partial V} \\ &= 2V + \Lambda \sum_{i=1}^{p} 2\big(\operatorname{tr}(A_i'X) - b_i\big)A_i'U \\ &= 2\left(V + \Lambda\left(\sum_{i=1}^{p}\big(\operatorname{tr}(A_i'X) - b_i\big)A_i'\right)U\right)\end{aligned} \tag{4}$$

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:

$$\frac{\partial\langle A, UV'\rangle}{\partial U} = \frac{\partial y}{\partial U} = \left[\frac{\partial y}{\partial u_{ij}}\right]_{m\times r} \tag{5}$$

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:

$$y = \langle A, UV'\rangle = \sum_{i,j} a_{ij}U_iV_j' = \sum_{i,j} a_{ij}\sum_k u_{ik}v_{jk} = \sum_{i,j}\sum_k a_{ij}u_{ik}v_{jk} \tag{6}$$

As a result:

$$\frac{\partial y}{\partial u_{ik}} = \sum_j a_{ij}v_{jk} = A_i\tilde{V}_k \tag{7}$$

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:

$$\frac{\partial y}{\partial U} = \left[\frac{\partial y}{\partial u_{ij}}\right]_{m\times r} = \left[A_i\tilde{V}_j\right]_{m\times r} = AV \tag{8}$$

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 $V$:

$$\begin{aligned}\nabla_U &= U + \Lambda\left(\sum\nolimits_{i=1}^{p}\left(\operatorname{tr}(A_i'UV') - b_i\right)A_i\right)V \\ \nabla_V &= V + \Lambda\left(\sum\nolimits_{i=1}^{p}\left(\operatorname{tr}(A_i'UV') - b_i\right)A_i'\right)U\end{aligned} \tag{9}$$

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:

$$
\begin{aligned}
\varphi(\gamma) &= \|U + \gamma\nabla_U\|_F^2 + \|V\|_F^2 + \Lambda \sum_{i=1}^{p} \Big( \langle (U+\gamma\nabla_U)V', A_i\rangle - b_i \Big)^2 \\
&= \operatorname{tr}\left((U+\gamma\nabla_U)'(U+\gamma\nabla_U)\right) + \|V\|_F^2 + \Lambda \sum_{i=1}^{p} \Big( \operatorname{tr}(A_i'(U+\gamma\nabla_U)V') - b_i \Big)^2 \\
&= \operatorname{tr}(U'U) + 2\gamma \operatorname{tr}(\nabla_U' U) + \gamma^2 \operatorname{tr}(\nabla_U'\nabla_U) + \|V\|_F^2 + \Lambda \sum_{i=1}^{p} \Big( \operatorname{tr}(A_i'UV') + \gamma \operatorname{tr}(A_i'\nabla_U V') - b_i \Big)^2
\end{aligned}
\tag{10}
$$

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

$$
\begin{aligned}
\varphi'(\gamma) &= 2\operatorname{tr}(\nabla_U'U) + 2\gamma\operatorname{tr}(\nabla_U'\nabla_U) + \Lambda \sum_{i=1}^{p} 2\Big( \operatorname{tr}(A_i'UV') + \gamma\operatorname{tr}(A_i'\nabla_U V') - b_i \Big) \operatorname{tr}(A_i'\nabla_U V') \\
&= 2\left\{ \operatorname{tr}(\nabla_U'U) + \gamma\operatorname{tr}(\nabla_U'\nabla_U) + \Lambda \sum_{i=1}^{p} \operatorname{tr}(A_i'\nabla_U V')\Big( \operatorname{tr}(A_i'UV') + \gamma\operatorname{tr}(A_i'\nabla_U V') - b_i \Big) \right\} \\
&= 2\left\{ \operatorname{tr}(\nabla_U'U) + \gamma\operatorname{tr}(\nabla_U'\nabla_U) + \Lambda \sum_{i=1}^{p} \left[ \operatorname{tr}(A_i'\nabla_U V')(\operatorname{tr}(A_i'UV') - b_i) + \gamma\operatorname{tr}^2(A_i'\nabla_U V') \right] \right\} \\
&= 2\left\{ \left( \operatorname{tr}(\nabla_U'U) + \Lambda \sum_{i=1}^{p} \operatorname{tr}(A_i'\nabla_U V')(\operatorname{tr}(A_i'UV') - b_i) \right) + \gamma \left( \operatorname{tr}(\nabla_U'\nabla_U) + \Lambda \sum_{i=1}^{p} \operatorname{tr}^2(A_i'\nabla_U V') \right) \right\}
\end{aligned}
\tag{11}
$$

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

$$
\gamma_U = -\frac{\operatorname{tr}(\nabla_U'U) + \Lambda \sum_{i=1}^{p} \operatorname{tr}(A_i'\nabla_U V')(\operatorname{tr}(A_i'UV') - b_i)}{\operatorname{tr}(\nabla_U'\nabla_U) + \Lambda \sum_{i=1}^{p} \operatorname{tr}^2(A_i'\nabla_U V')} \tag{12}
$$

and the corresponding updating rule for $U$ is:

$$
U \leftarrow U + \gamma_U \nabla_U \tag{13}
$$

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

$$
\gamma_V = -\frac{\operatorname{tr}(\nabla_V'V) + \Lambda \sum_{i=1}^{p} \operatorname{tr}(A_i'U\nabla_V')(\operatorname{tr}(A_i'UV') - b_i)}{\operatorname{tr}(\nabla_V'\nabla_V) + \Lambda \sum_{i=1}^{p} \operatorname{tr}^2(A_i'U\nabla_V')} \tag{14}
$$

and the corresponding updating rule for $V$ is:

$$
V \leftarrow V + \gamma_V \nabla_V \tag{15}
$$

**III. Reference**


[1] Y. Zhang, M. Zhang, Y. Zhang, Y. Liu and S. Ma, Understanding the Sparsity: Augmented Matrix Factorization with Sampled Constraints on Unobservables, *In Proceedings of the 23rd ACM International Conference on Information and Knowledge Management (CIKM 2014), Shanghai, China.*