Regularized PARAFAC2 with AO-ADMM

Carla Schenker

May 11, 2021

1 PARAFAC2 Notation

Given: K matrices $\mathbf{X}_k \in \mathbb{R}^{I \times J_k}$.
PARAFAC2 model:

$$\mathbf{X}_k \approx \mathbf{A}\mathbf{D}_k\mathbf{B}_k^T, \quad \text{where } \mathbf{B}_k = \mathbf{P}_k\mathbf{B} \text{ and } \mathbf{P}_k^T\mathbf{P}_k = I, \quad \text{for } k = 1, ..., K, \tag{1}$$

with $\mathbf{A} \in \mathbb{R}^{I \times R}$, $\mathbf{B}_k \in \mathbb{R}^{J_k \times R}$, $\mathbf{P}_k \in \mathbb{R}^{J_k \times R}$, $\mathbf{B} \in \mathbb{R}^{R \times R}$, $\mathbf{D}_k = \mathrm{Diag}\left(\mathbf{c}_{k,:}\right) \in \mathbb{R}^{R \times R}$ and $\mathbf{C} \in \mathbb{R}^{K \times R}$.

Often, regularizations or constraints need to be imposed on the factor matrices. For this, we introduce (convex) regularization functions g_A, g_B and g_D. Hard constraints are realized when the regularization functions are given as indicator functions of a (convex) set, e.g. the nonnegative orthant. Thus, we want to solve optimization problems of the form:

$$\begin{aligned} \underset{\mathbf{A},\mathbf{B},\{\mathbf{D}_k,\mathbf{B}_k,\mathbf{P}_k\}_{k\leq K}}{\operatorname{argmin}} \quad & \sum_{k=1}^{K}\left[\left\| \mathbf{X}_k - \mathbf{A}\mathbf{D}_k\mathbf{B}_k^T \right\|_F^2 + g_D(\mathbf{D}_k) + g_B(\mathbf{B}_k) + \iota_{orth}(\mathbf{P}_k)\right] + g_A(\mathbf{A}) \\ \text{s.t.} \quad & \mathbf{B}_k = \mathbf{P}_k\mathbf{B}, \quad k = 1, ..., K \end{aligned} \tag{2}$$

2 ADMM for the varying mode $\mathbf{B}_k$

We use ADMM to solve the following subproblem, where $\mathbf{A}$ and $\mathbf{C}$ are fixed:

$$\begin{aligned} \underset{\mathbf{B},\{\mathbf{B}_k,\mathbf{P}_k\}_{k\leq K}}{\operatorname{argmin}} \quad & \sum_{k=1}^{K}\left[\left\| \mathbf{X}_k - \mathbf{A}\mathbf{D}_k\mathbf{B}_k^T \right\|_F^2 + g_B(\mathbf{B}_k) + \iota_{orth}(\mathbf{P}_k)\right] \\ \text{s.t.} \quad & \mathbf{B}_k = \mathbf{P}_k\mathbf{B}, \quad k = 1, ..., K \end{aligned} \tag{3}$$

For this, we introduce auxiliary variables $\{\mathbf{Z}_{B_k}\}_{k\leq K}$, in order to split the regularization from the factorization:

$$\begin{aligned} \underset{\mathbf{B},\left\{\mathbf{B}_k,\mathbf{P}_k,\mathbf{Z}_{B_k}\right\}_{k\leq K}}{\operatorname{argmin}} \quad & \sum_{k=1}^{K}\left[\left\| \mathbf{X}_k - \mathbf{A}\mathbf{D}_k\mathbf{B}_k^T \right\|_F^2 + g_B(\mathbf{Z}_{B_k}) + \iota_{orth}(\mathbf{P}_k)\right] \\ \text{s.t.} \quad & \mathbf{B}_k = \mathbf{P}_k\mathbf{B}, \quad k = 1, ..., K \\ & \mathbf{B}_k = \mathbf{Z}_{B_k}, \quad k = 1, ..., K \end{aligned} \tag{4}$$

Introducing dual variables $\{\boldsymbol{\mu}_{Z_{B_k}}\}_{k\leq K}$ for the regularization and $\{\boldsymbol{\mu}_{\Delta_k}\}_{k\leq K}$ for the coupling, the augmented Lagrangian can be written as:

$$L = \sum_{k=1}^{K}\left[\left\| \mathbf{X}_k - \mathbf{A}\mathbf{D}_k\mathbf{B}_k^T \right\|_F^2 + g_B(\mathbf{Z}_{B_k}) + \iota_{orth}(\mathbf{P}_k) + \frac{\rho_k}{2}\left(\left\| \mathbf{B}_k - \mathbf{Z}_{B_k} + \boldsymbol{\mu}_{Z_{B_k}} \right\|_F^2 + \left\| \mathbf{B}_k - \mathbf{P}_k\mathbf{B} + \boldsymbol{\mu}_{\Delta_k} \right\|_F^2\right)\right] \tag{5}$$

ADMM alternatingly minimizes L with respect to the blocks $\{\mathbf{B}_k\}$, $\{\mathbf{Z}_{B_k}\}$, $\mathbf{B}$ and $\{\mathbf{P}_k\}$ and performs a gradient ascent step for the dual variables $\{\boldsymbol{\mu}_{Z_{B_k}}\}_{k\leq K}$ and $\{\boldsymbol{\mu}_{\Delta_k}\}_{k\leq K}$.

Algorithm 1 ADMM for subproblem w.r.t. mode $\mathbf{B}$ of regularized PARAFAC2

while convergence criterion is not met **do**

  for $k = 1, ..., K$ **do**

$$\mathbf{B}_k^{(n+1)} = \underset{\mathbf{Y}}{\operatorname{argmin}} \left\| \mathbf{X}_k - \mathbf{A}\mathbf{D}_k\mathbf{Y}^T \right\|_F^2 + \frac{\rho_k}{2}\left(\left\| \mathbf{Y} - \mathbf{Z}_{B_k}^{(n)} + \boldsymbol{\mu}_{Z_{B_k}}^{(n)} \right\|_F^2 + \left\| \mathbf{Y} - \mathbf{P}_k^{(n)}\mathbf{B}^{(n)} + \boldsymbol{\mu}_{\Delta_k}^{(n)} \right\|_F^2 \right)$$
$$= \left[\mathbf{X}_k^T\mathbf{A}\mathbf{D}_k + \frac{\rho_k}{2}\left(\mathbf{Z}_{B_k}^{(n)} - \boldsymbol{\mu}_{Z_{B_k}}^{(n)} + \mathbf{P}_k^{(n)}\mathbf{B}^{(n)} - \boldsymbol{\mu}_{\Delta_k}^{(n)}\right)\right]\left[\mathbf{D}_k\mathbf{A}^T\mathbf{A}\mathbf{D}_k + \rho_k\mathbf{I}\right]^{-1}$$

$$\mathbf{P}_k^{(n+1)} = \underset{\mathbf{Y}}{\operatorname{argmin}} \left\| \left(\mathbf{B}_k^{(n+1)} + \boldsymbol{\mu}_{\Delta_k}^{(n)}\right) - \mathbf{Y}\mathbf{B}^{(n)} \right\|_F^2 + \iota_{orth}(\mathbf{Y})$$

  end for

$$\mathbf{B}^{(n+1)} = \underset{\mathbf{Y}}{\operatorname{argmin}} \sum_{k=1}^{K} \frac{\rho_k}{2} \left\| \mathbf{B}_k^{(n+1)} - \mathbf{P}_k^{(n+1)}\mathbf{Y} + \boldsymbol{\mu}_{\Delta_k}^{(n)} \right\|_F^2 = \frac{1}{\sum_{k=1}^{K}\rho_k} \sum_{k=1}^{K} \rho_k \mathbf{P}_k^{(n+1)^T}\left(\mathbf{B}_k^{(n+1)} + \boldsymbol{\mu}_{\Delta_k}^{(n)}\right)$$

  for $k = 1, ..., K$ **do**

$$\mathbf{Z}_{B_k}^{(n+1)} = \underset{\mathbf{Y}}{\operatorname{argmin}}\ g_B(\mathbf{Y}) + \frac{\rho_k}{2}\left\| \mathbf{B}_k^{(n+1)} - \mathbf{Y} + \boldsymbol{\mu}_{Z_{B_k}}^{(n)} \right\|_F^2 = \operatorname{prox}_{\frac{1}{\rho_k} g_B}\left(\mathbf{B}_k^{(n+1)} + \boldsymbol{\mu}_{Z_{B_k}}^{(n)}\right)$$

$$\boldsymbol{\mu}_{Z_{B_k}}^{(n+1)} = \boldsymbol{\mu}_{Z_{B_k}}^{(n)} + \mathbf{B}_k^{(n+1)} - \mathbf{Z}_{B_k}^{(n+1)}$$

$$\boldsymbol{\mu}_{\Delta_k}^{(n+1)} = \boldsymbol{\mu}_{\Delta_k}^{(n)} + \mathbf{B}_k^{(n+1)} - \mathbf{P}_k^{(n+1)}\mathbf{B}^{(n+1)}$$

  end for

  $n = n + 1$

end while

The update of $\mathbf{P}_k$ is the solution of an orthogonal procustes problem and can be computed via the singular value decomposition as $\mathbf{P}_k^{(n+1)} = \mathbf{U}(:, 1 : R)\mathbf{V}(:, 1 : R)^T$ where $[U, S, V] = \operatorname{svd}\left(\left(\mathbf{B}_k^{(n+1)} + \boldsymbol{\mu}_{\Delta_k}^{(n)}\right)\mathbf{B}^{(n)^T}\right)$.

For the Update of $\mathbf{B}_k$, a Cholesky decomposition of the matrix $\left[\mathbf{D}_k\mathbf{A}^T\mathbf{A}\mathbf{D}_k + \rho_k\mathbf{I}\right]$ can be precomputed outside the ADMM loop.

Both FOR loops can be computed as parallel FOR loops.

2.1 Stopping conditions

All of the following conditions should be satisfied:

$$\begin{aligned}
&\frac{1}{K}\sum_k \left(\|\mathbf{B}_k^{(n)} - \mathbf{Z}_{B_k}^{(n)}\|_F / \|\mathbf{B}_k^{(n)}\|_F\right) \leq \epsilon^{\text{p,constr}} \\
&\frac{1}{K}\sum_k \left(\|\mathbf{B}_k^{(n)} - \mathbf{P}_k^{(n)}\mathbf{B}^{(n)}\|_F / \|\mathbf{B}_k^{(n)}\|_F\right) \leq \epsilon^{\text{p,coupl}} \\
&\frac{1}{K}\sum_k \left(\|\mathbf{Z}_{B_k}^{(n+1)} - \mathbf{Z}_{B_k}^{(n)}\|_F / \|\boldsymbol{\mu}_{Z_{B_k}}^{(n)}\|_F\right) \leq \epsilon^{\text{d,constr}} \\
&\frac{1}{K}\sum_k \left(\|\mathbf{P}_k^{(n+1)}\mathbf{B}^{(n+1)} - \mathbf{P}_k^{(n)}\mathbf{B}^{(n)}\|_F / \|\boldsymbol{\mu}_{\Delta_k}^{(n)}\|_F\right) \leq \epsilon^{\text{d,coupl}}
\end{aligned} \tag{6}$$

or when a maximum number of iterations is reached (e.g. 5).

3 ADMM for the other modes

3.1 Update for A

Subproblem for (regularized) $\mathbf{A}$:

$$\begin{aligned} \underset{\mathbf{A},\mathbf{Z}_A}{\operatorname{argmin}} \quad & \sum_{k=1}^{K} \left[\| \mathbf{X}_k - \mathbf{A}\mathbf{D}_k\mathbf{B}_k^T \|_F^2 \right] + g_A(\mathbf{Z}_A) \\ \text{s.t.} \quad & \mathbf{A} = \mathbf{Z}_A \end{aligned} \tag{7}$$

Use ADMM:

Algorithm 2 ADMM for subproblem w.r.t. mode $\mathbf{A}$ of regularized PARAFAC2

1: **while** convergence criterion is not met **do**

2: $$\begin{aligned} \mathbf{A}^{(n+1)} &= \underset{\mathbf{Y}}{\operatorname{argmin}} \sum_{k=1}^{K} \| \mathbf{X}_k - \mathbf{Y}\mathbf{D}_k\mathbf{B}_k^T \|_F^2 + \frac{\rho}{2} \left\| \mathbf{Y} - \mathbf{Z}_A^{(n)} + \boldsymbol{\mu}_{Z_A}^{(n)} \right\|_F^2 \\ &= \left[\sum_{k=1}^{K} \mathbf{X}_k\mathbf{B}_k\mathbf{D}_k + \frac{\rho}{2}\left(\mathbf{Z}_A^{(n)} - \boldsymbol{\mu}_{Z_A}^{(n)}\right) \right] \left[\sum_{k=1}^{K} \mathbf{D}_k\mathbf{B}_k^T\mathbf{B}_k\mathbf{D}_k + \frac{\rho}{2}\mathbf{I} \right]^{-1} \end{aligned}$$

3: $\mathbf{Z}_A^{(n+1)} = \operatorname{prox}_{\frac{1}{\rho}g_A}\left(\mathbf{A}^{(n+1)} + \boldsymbol{\mu}_{Z_A}^{(n)}\right)$

4: $\boldsymbol{\mu}_{Z_A}^{(n+1)} = \boldsymbol{\mu}_{Z_A}^{(n)} + \mathbf{A}^{(n+1)} - \mathbf{Z}_A^{(n+1)}$

5: $n = n + 1$

6: **end while**

Precompute the Cholesky decomposition of the matrix $\left[\sum_{k=1}^{K} \mathbf{D}_k\mathbf{B}_k^T\mathbf{B}_k\mathbf{D}_k + \frac{\rho_k}{2}\mathbf{I}\right]$ outside the ADMM loop!
Stopping condition like for the funny mode (conditions 1 and 3), without summation over k.

3.2 Update for C

Subproblem for (regularized) $\mathbf{C}$:

$$\begin{aligned} \underset{\mathbf{C},\mathbf{Z}_C}{\operatorname{argmin}} \quad & \sum_{k=1}^{K} \left[\| \mathbf{X}_k - \mathbf{A}\operatorname{Diag}(\mathbf{c}_{k,:})\,\mathbf{B}_k^T \|_F^2 + g_C(\mathbf{z}_{C_k}) \right] \\ \text{s.t.} \quad & \mathbf{c}_{k,:}^T = \mathbf{z}_{C_k}, \quad k = 1, ..., K \end{aligned} \tag{8}$$

After vectorization:

$$\begin{aligned} \underset{\mathbf{C},\mathbf{Z}_C}{\operatorname{argmin}} \quad & \sum_{k=1}^{K} \left[\| \operatorname{vec}(\mathbf{X}_k) - (\mathbf{B}_k \odot \mathbf{A})\mathbf{c}_{k,:}^T \|_2^2 + g_C(\mathbf{z}_{C_k}) \right] \\ \text{s.t.} \quad & \mathbf{c}_{k,:}^T = \mathbf{z}_{C_k}, \quad k = 1, ..., K \end{aligned} \tag{9}$$

Use ADMM and the fact that $(\mathbf{B}_k \odot \mathbf{A})^T(\mathbf{B}_k \odot \mathbf{A}) = \mathbf{A}^T\mathbf{A} * \mathbf{B}_k^T\mathbf{B}_k$ and $(\mathbf{B}_k \odot \mathbf{A})^T\operatorname{vec}(\mathbf{X}_k) = \operatorname{Diag}\left(\mathbf{A}^T\mathbf{X}_k\mathbf{B}_k\right)$.

Algorithm 3 ADMM for subproblem w.r.t. mode $\mathbf{C}$ of regularized PARAFAC2

1: **while** convergence criterion is not met **do**

2: **for** $k = 1, ..., K$ **do**

3:
$$\begin{aligned}
\mathbf{c}_{k,:}^{(n+1)^T} &= \underset{\mathbf{y}}{\operatorname{argmin}} \ \| \operatorname{vec}(\mathbf{X}_k) - (\mathbf{B}_k \odot \mathbf{A})\mathbf{y} \|_2^2 + \frac{\rho_k}{2} \left\| \mathbf{y} - \mathbf{z}_{C_k}^{(n)} + \boldsymbol{\mu}_{z_{C_k}}^{(n)} \right\|_2^2 \\
&= \left[(\mathbf{A}^T\mathbf{A} * \mathbf{B}_k^T \mathbf{B}_k) + \frac{\rho_k}{2} \mathbf{I}_R \right]^{-1} \left[\operatorname{Diag}\left(\mathbf{A}^T \mathbf{X}_k \mathbf{B}_k\right) + \frac{\rho_k}{2} \left(\mathbf{z}_{C_k}^{(n)} - \boldsymbol{\mu}_{z_{C_k}}^{(n)} \right) \right]
\end{aligned}$$

4: $\mathbf{z}_{C_k}^{(n+1)} = \operatorname{prox}_{\frac{1}{\rho_k} g_C} \left(\mathbf{c}_{k,:}^{(n+1)^T} + \boldsymbol{\mu}_{z_k}^{(n)} \right)$

5: $\boldsymbol{\mu}_{z_k}^{(n+1)} = \boldsymbol{\mu}_{z_k}^{(n)} + \mathbf{c}_{k,:}^{(n+1)^T} - \mathbf{z}_{C_k}^{(n+1)}$

6: **end for**

7: $n = n + 1$

8: **end while**

Precompute the Cholesky decomposition of the matrices $\left[(\mathbf{A}^T\mathbf{A} * \mathbf{B}_k^T\mathbf{B}_k) + \frac{\rho_k}{2}\mathbf{I}_R \right]$ outside the ADMM loop!

4 AO-ADMM for regularized PARAFAC2

Initialize everything (like Jeremy?) and alternatingly optimize over $\mathbf{A}$, $\mathbf{C}$ and $\{\mathbf{B}_k, \mathbf{P}_k, \mathbf{B}\}$ (not sure about the best order) using a few steps of ADMM as described above.

4.1 Overall stopping conditions

The whole algorithm is terminated, when each of the following residuals $f_\star^{(n)}$,

$$\begin{aligned}
f_{\text{tensors}}^{(n)} &= \sum_{k=1}^{K} \left(\| \mathbf{X}_k - \mathbf{A}^{(n)} \mathbf{D}_k^{(n)} \mathbf{B}_k^{(n)^T} \|_F^2 / \| \mathbf{X}_k \|_F^2 \right) \\
f_{\text{couplings}}^{(n)} &= \sum_{k=1}^{K} \left(\| \mathbf{B}_k^{(n)} - \mathbf{P}_k^{(n)} \mathbf{B}^{(n)} \|_F / \| \mathbf{B}_k^{(n)} \|_F \right) \\
f_{\text{constraints}}^{(n)} &= \sum_{k=1}^{K} \left(\| \mathbf{B}_k^{(n)} - \mathbf{Z}_{B_k}^{(n)} \|_F / \| \mathbf{B}_k^{(n)} \|_F \right) + \| \mathbf{A}^{(n)} - \mathbf{Z}_A^{(n)} \|_F / \| \mathbf{A}^{(n)} \|_F + \| \mathbf{C}^{(n)} - \mathbf{Z}_C^{(n)} \|_F / \| \mathbf{C}^{(n)} \|_F
\end{aligned} \tag{10}$$

has either reached a small absolute tolerance $\epsilon^{\text{abs,outer}}$, or has not changed more than some small relative tolerance $\epsilon^{\text{rel,outer}}$ compared to the previous iteration,

$$f_\star^{(n)} < \epsilon^{\text{abs,outer}}, \qquad | \, f_\star^{(n)} - f_\star^{(n-1)} \, | \, / \, | \, f_\star^{(n)} \, | < \epsilon^{\text{rel,outer}}$$

or a predefined number of maximal outer iterations is reached.