Supplementary Material

Carla Schenker, Xiulin Wang, Evrim Acar

April 5, 2025

1 Additional ADMM updates

1.1 ADMM for varying mode B of PARAFAC2

As first proposed in [1], the subproblem for the (regularized) varying mode $\mathbf{B}$ of the PARAFAC2 model,

$$\begin{aligned} \underset{\{\mathbf{B}_k\}_{k\leq K}}{\operatorname{argmin}} \quad & \sum_{k=1}^{K} w_1 \left\| \mathbf{X}_k - \mathbf{A}\mathbf{D}_k\mathbf{B}_k^T \right\|_F^2 + g_B(\mathbf{B}_k) \\ \text{s.t.} \quad & \{\mathbf{B}_k\}_{k\leq K} \in \mathcal{P} \end{aligned} \tag{1}$$

is solved using ADMM with the following splitting scheme:

$$\begin{aligned} \underset{\{\mathbf{B}_k\}_{k\leq K}}{\operatorname{argmin}} \quad & \sum_{k=1}^{K} w_1 \left\| \mathbf{X}_k - \mathbf{A}\mathbf{D}_k\mathbf{B}_k^T \right\|_F^2 + g_B(\mathbf{Z}_{B_k}) + \iota_{\mathcal{P}}(\mathbf{W}_{B_k}) \\ \text{s.t.} \quad & \mathbf{B}_k = \mathbf{Z}_{B_k}, \quad \forall k \leq K, \\ & \mathbf{B}_k = \mathbf{W}_{B_k}, \quad \forall k \leq K. \end{aligned} \tag{2}$$

This is in the standard form of problems that are solvable with ADMM, except the set $\mathcal{P}$, which describes the PARAFAC2 constraint, is not convex and the computation of the corresponding proximal operator is not straightforward. The ADMM algorithm is given in Algorithm 1.

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

1: **while** convergence criterion is not met **do**
2:   **for** $k = 1, ..., K$ **do**
3: $$\begin{aligned} \mathbf{B}_k^{(n+1)} &= \underset{\mathbf{B}_k}{\operatorname{argmin}}\; w_1 \left\| \mathbf{X}_k - \mathbf{A}\mathbf{D}_k\mathbf{B}_k^T \right\|_F^2 + \frac{\rho_{B_k}}{2}\left(\left\| \mathbf{B}_k - \mathbf{Z}_{B_k}^{(n)} + \boldsymbol{\mu}_{Z_{B_k}}^{(n)} \right\|_F^2 + \left\| \mathbf{B}_k - \mathbf{W}^{(n)} + \boldsymbol{\mu}_{\Delta_{B_k}}^{(n)} \right\|_F^2 \right) \\ &= \left[w_1 \mathbf{X}_k^T \mathbf{A}\mathbf{D}_k + \frac{\rho_{B_k}}{2}\left(\mathbf{Z}_{B_k}^{(n)} - \boldsymbol{\mu}_{Z_{B_k}}^{(n)} + \mathbf{W}^{(n)} - \boldsymbol{\mu}_{\Delta_{B_k}}^{(n)} \right) \right] \left[w_1 \mathbf{D}_k \mathbf{A}^T \mathbf{A}\mathbf{D}_k + \rho_{B_k}\mathbf{I} \right]^{-1} \end{aligned}$$
4:     $\mathbf{Z}_{B_k}^{(n+1)} = \operatorname{prox}_{\frac{1}{\rho_{B_k}} g_B}\left(\mathbf{B}_k^{(n+1)} + \boldsymbol{\mu}_{Z_{B_k}}^{(n)} \right)$
5:   **end for**
6:   $\left\{ \mathbf{W}_{B_k}^{(n+1)} \right\}_{k\leq K} = \operatorname{prox}_{\frac{1}{\rho_{B_k}} \iota_{\mathcal{P}}}\left(\left\{ \mathbf{B}_k^{(n+1)} + \boldsymbol{\mu}_{\boldsymbol{\Delta}_{B_k}}^{(n)} \right\}_{k\leq K} \right) \longleftarrow$ Algorithm 2
7:   **for** $k = 1, ..., K$ **do**
8:     $\boldsymbol{\mu}_{Z_{B_k}}^{(n+1)} = \boldsymbol{\mu}_{Z_{B_k}}^{(n)} + \mathbf{B}_k^{(n+1)} - \mathbf{Z}_{B_k}^{(n+1)}$
9:     $\boldsymbol{\mu}_{\Delta_{B_k}}^{(n+1)} = \boldsymbol{\mu}_{\Delta_{B_k}}^{(n)} + \mathbf{B}_k^{(n+1)} - \mathbf{W}_{B_k}^{(n+1)}$
10:   **end for**
11:   $n = n + 1$
12: **end while**

K Cholesky decompositions of the matrices $\left[w_1 \mathbf{D}_k \mathbf{A}^T \mathbf{A}\mathbf{D}_k + \rho_{B_k}\mathbf{I} \right]$ are precomputed outside the ADMM loop, which has a complexity of $\mathcal{O}(KR^3)$. Solving for B_k, $k = 1, ..., K$ at each ADMM update then reduces to one forward- and one backward-substitution with total complexity KJR^2. Note that both FOR-loops can be computed in parallel. For the evaluation of the proximal operator in line 6, we use the (standard) parametrization of the set $\mathcal{P}$, see [1], to obtain the following equivalent formulation:

$$\underset{\boldsymbol{\Delta}_B, \{\mathbf{P}_k\}_{k\leq K}}{\operatorname{argmin}} \sum_{k=1}^{K} \left\| \left(\mathbf{B}_k + \boldsymbol{\mu}_{\Delta_{B_k}} \right) - \mathbf{P}_k \boldsymbol{\Delta}_B \right\|_F^2 + \iota_{orth}(\mathbf{P}_k). \tag{3}$$

In order to efficiently approximate this, we employ an AO scheme, where the orthogonal Procustes problem is solved independently for each $\mathbf{P}_k$, see Algorithm 2. We only use one iteration of Algorithm 2 for each ADMM iteration, as this is sufficient in our experience.

Algorithm 2 Approximate projection onto $\mathcal{P}$

1: **while** convergence criterion is not met **do**
2:   **for** $k = 1, ..., K$ **do**
3:     Compute SVD: $\mathbf{U}^{(k)}\mathbf{\Sigma}^{(k)}\mathbf{V}^{(k)^T} = \left(\mathbf{B}_k + \boldsymbol{\mu}_{\Delta_{B_k}}\right)\mathbf{\Delta}_B^T$
4:     $\mathbf{P}_k = \mathbf{U}^{(k)}\mathbf{V}^{(k)^T}$
5:   **end for**
6:   $\mathbf{\Delta}_B = \frac{1}{\sum_{k=1}^{K}\rho_{B_k}}\sum_{k=1}^{K}\rho_{B_k}\mathbf{P}_k^T\left(\mathbf{B}_k + \boldsymbol{\mu}_{\Delta_{B_k}}\right)$
7: **end while**

1.2 ADMM for other modes of model (4) in main paper

These ADMM updates can also be found in our previous paper [2] and its supplementary material.

1.2.1 ADMM for mode C

The subproblem for regularized $\mathbf{C}$ with split variable $\mathbf{Z}_C$,

$$\begin{aligned} \underset{\mathbf{C},\mathbf{Z}_C}{\operatorname{argmin}} \quad & \sum_{k=1}^{K}\left[w_1 \left\| \mathbf{X}_k - \mathbf{A}\mathrm{Diag}\left(\mathbf{C}(k,:)\right)\mathbf{B}_k^T \right\|_F^2 + g_C(\mathbf{Z}_C)\right] \\ \text{s.t.} \quad & \mathbf{C} = \mathbf{Z}_C \end{aligned} \tag{4}$$

is vectorized as follows:

$$\begin{aligned} \underset{\mathbf{C},\mathbf{Z}_C}{\operatorname{argmin}} \quad & \sum_{k=1}^{K}\left[w_1 \left\| \mathrm{vec}(\mathbf{X}_k) - (\mathbf{B}_k \odot \mathbf{A})\mathbf{C}(k,:)^T \right\|_2^2 + g_C(\mathbf{Z}_C)\right]. \\ \text{s.t.} \quad & \mathbf{C} = \mathbf{Z}_C \end{aligned} \tag{5}$$

Applying row-wise ADMM and the following transformations

$$\begin{aligned} (\mathbf{B}_k \odot \mathbf{A})^T(\mathbf{B}_k \odot \mathbf{A}) &= \mathbf{A}^T\mathbf{A} * \mathbf{B}_k^T\mathbf{B}_k, \\ (\mathbf{B}_k \odot \mathbf{A})^T\mathrm{vec}(\mathbf{X}_k) &= \mathrm{Diag}\left(\mathbf{A}^T\mathbf{X}_k\mathbf{B}_k\right), \end{aligned} \tag{6}$$

results in Algorithm 3, see also [1], where all rows k can be updated in parallel.

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:     $$\begin{aligned} \mathbf{C}^{(n+1)}(k,:)^T &= \underset{\mathbf{c}}{\operatorname{argmin}}\ w_1 \left\| \mathrm{vec}(\mathbf{X}_k) - (\mathbf{B}_k \odot \mathbf{A})\mathbf{c} \right\|_2^2 + \frac{\rho_{C_k}}{2}\left\| \mathbf{c} - \mathbf{Z}_C^{(n)}(k,:) + \boldsymbol{\mu}_{Z_C}^{(n)}(k,:) \right\|_2^2 \\ &= \left[w_1(\mathbf{A}^T\mathbf{A} * \mathbf{B}_k^T\mathbf{B}_k) + \frac{\rho_{C_k}}{2}\mathbf{I}_R\right]^{-1}\left[w_1\mathrm{Diag}\left(\mathbf{A}^T\mathbf{X}_k\mathbf{B}_k\right) + \frac{\rho_{C_k}}{2}\left(\mathbf{Z}_C^{(n)}(k,:) - \boldsymbol{\mu}_{Z_C}^{(n)}(k,:)\right)\right] \end{aligned}$$
4:   **end for**
5:   $\mathbf{Z}_C^{(n+1)} = \mathrm{prox}_{\frac{1}{\max \rho_{C_k}} g_C}\left(\mathbf{C}^{(n+1)} + \boldsymbol{\mu}_{Z_C}^{(n)}\right)$
6:   $\boldsymbol{\mu}_{Z_C}^{(n+1)} = \boldsymbol{\mu}_{Z_C}^{(n)} + \mathbf{C}^{(n+1)} - \mathbf{Z}_C^{(n+1)}$
7:   $n = n + 1$
8: **end while**

1.2.2 ADMM for mode F

The regularized, but uncoupled mode of the matrix decomposition $\mathbf{F}$ is updated using standard ADMM, see also [3], as follows:

Algorithm 4 ADMM for subproblem w.r.t. $\mathbf{F}$

while convergence criterion is not met **do**
$\quad \mathbf{F}^{(n+1)} = \left[w_2 \mathbf{Y}^T \mathbf{E} + \frac{\rho_F}{2}\left(\mathbf{Z}_F^{(n)} - \boldsymbol{\mu}_{Z_F}^{(n)}\right)\right]\left[w_2 \mathbf{E}^T \mathbf{E} + \frac{\rho_F}{2}\mathbf{I}_{R_2}\right]^{-1}$
$\quad \mathbf{Z}_F^{(n+1)} = \mathrm{prox}_{\frac{1}{\rho_F} g_F}\left(\mathbf{F}^{(n+1)} + \boldsymbol{\mu}_{Z_F}^{(n)}\right)$
$\quad \boldsymbol{\mu}_{Z_F}^{(n+1)} = \boldsymbol{\mu}_{Z_F}^{(n)} + \mathbf{F}^{(n+1)} - \mathbf{Z}_F^{(n+1)}$
$\quad n = n + 1$
end while

1.2.3 ADMM for coupled modes A and E

Defining split variables $\mathbf{Z}_A$ and $\mathbf{Z}_E$, the subproblem for regularized $\mathbf{A}$ and $\mathbf{E}$ is written as:

$$\begin{aligned}
\underset{\substack{\mathbf{E},\mathbf{A} \\ \mathbf{Z}_E,\mathbf{Z}_A,\boldsymbol{\Delta}}}{\operatorname{argmin}} \quad & w_1 \sum_{k=1}^{K} \left\| \mathbf{X}_k - \mathbf{A}\mathbf{D}_k \mathbf{B}_k^T \right\|_F^2 + w_2 \left\| \mathbf{Y} - \mathbf{E}\mathbf{F}^T \right\|_F^2 \\
& \qquad + g_A(\mathbf{Z}_A) + g_E(\mathbf{Z}_E) \\
\text{s.t.} \quad & \mathbf{A} = \mathbf{Z}_A, \quad \mathbf{E} = \mathbf{Z}_E \\
& \mathbf{H}_{\mathbf{A}} \mathrm{vec}(\mathbf{A}) = \mathbf{H}_{\mathbf{A}}^{\boldsymbol{\Delta}} \mathrm{vec}(\boldsymbol{\Delta}), \mathbf{H}_{\mathbf{E}} \mathrm{vec}(\mathbf{E}) = \mathbf{H}_{\mathbf{E}}^{\boldsymbol{\Delta}} \mathrm{vec}(\boldsymbol{\Delta}).
\end{aligned} \tag{7}$$

Case 1 Exact coupling: $\mathbf{A} = \boldsymbol{\Delta}$. In this case, $\mathbf{A}$ and $\mathbf{E}$ are updated by solving the following linear systems,

$$\begin{aligned}
&\mathbf{A}^{(n+1)} \left[w_1 \sum_{k=1}^{K} \mathbf{D}_k \mathbf{B}_k^T \mathbf{B}_k \mathbf{D}_k + \frac{\rho_A}{2}\left(\mathbf{I}_{R_1} + \mathbf{I}_{R_1}\right) \right] = \\
&\left[w_1 \sum_{k=1}^{K} \mathbf{X}_k \mathbf{B}_k \mathbf{D}_k + \frac{\rho_A}{2}\left(\mathbf{Z}_A^{(n)} - \boldsymbol{\mu}_{Z_A}^{(n)} + \boldsymbol{\Delta}^{(n)} - \boldsymbol{\mu}_{\boldsymbol{\Delta}_A}^{(n)}\right) \right],
\end{aligned} \tag{8}$$

$$\begin{aligned}
&\mathbf{E}^{(n+1)} \left[w_2 \mathbf{F}^T \mathbf{F} + \frac{\rho_E}{2}\left(\mathbf{I}_{R_2} + \mathbf{I}_{R_2}\right) \right] = \\
&\left[w_2 \mathbf{Y}\mathbf{F} + \frac{\rho_E}{2}\left(\mathbf{Z}_E^{(n)} - \boldsymbol{\mu}_{Z_E}^{(n)} + \boldsymbol{\Delta}^{(n)} - \boldsymbol{\mu}_{\boldsymbol{\Delta}_E}^{(n)}\right) \right],
\end{aligned} \tag{9}$$

The update for $\boldsymbol{\Delta}$ is in this case as follows:

$$\boldsymbol{\Delta}^{(n+1)} = \frac{1}{\rho_A + \rho_E}\left[\rho_A\left(\mathbf{A}^{(n+1)} + \boldsymbol{\mu}_{\boldsymbol{\Delta}_A}^{(n)}\right) + \rho_E\left(\mathbf{E}^{(n+1)} + \boldsymbol{\mu}_{\boldsymbol{\Delta}_E}^{(n)}\right)\right]. \tag{10}$$

The whole ADMM algorithm for this subproblem is given in Alg. 5.

Algorithm 5 ADMM for subproblem w.r.t. $\mathbf{A}$ and $\mathbf{E}$

while convergence criterion is not met **do**
$\quad \mathbf{A}^{(n+1)} \longleftarrow$ solve linear system (8)
$\quad \mathbf{E}^{(n+1)} \longleftarrow$ solve linear system (9)
$\quad \boldsymbol{\Delta}^{(n+1)} \longleftarrow$ (10)
$\quad \mathbf{Z}_A^{(n+1)} = \mathrm{prox}_{\frac{1}{\rho_A} g_A}\left(\mathbf{A}^{(n+1)} + \boldsymbol{\mu}_{Z_A}^{(n)}\right)$
$\quad \mathbf{Z}_E^{(n+1)} = \mathrm{prox}_{\frac{1}{\rho_E} g_E}\left(\mathbf{E}^{(n+1)} + \boldsymbol{\mu}_{Z_E}^{(n)}\right)$
$\quad \boldsymbol{\mu}_{Z_A}^{(n+1)} = \boldsymbol{\mu}_{Z_A}^{(n)} + \mathbf{A}^{(n+1)} - \mathbf{Z}_A^{(n+1)}$
$\quad \boldsymbol{\mu}_{Z_E}^{(n+1)} = \boldsymbol{\mu}_{Z_E}^{(n)} + \mathbf{E}^{(n+1)} - \mathbf{Z}_E^{(n+1)}$
$\quad \boldsymbol{\mu}_{\boldsymbol{\Delta}_A}^{(n+1)} = \boldsymbol{\mu}_{\boldsymbol{\Delta}_A}^{(n)} + \mathbf{A}^{(n+1)} - \boldsymbol{\Delta}^{(n+1)}$
$\quad \boldsymbol{\mu}_{\boldsymbol{\Delta}_E}^{(n+1)} = \boldsymbol{\mu}_{\boldsymbol{\Delta}_E}^{(n)} + \mathbf{E}^{(n+1)} - \boldsymbol{\Delta}^{(n+1)}$
$\quad n = n + 1$
end while

For the other four different types of linear couplings described in [3], the updates of $\mathbf{A}, \mathbf{E}, \boldsymbol{\Delta}, \boldsymbol{\mu}_{\boldsymbol{\Delta}_A}$ and $\boldsymbol{\mu}_{\boldsymbol{\Delta}_E}$ in Algorithm 5 have to be adapted. We give the specific updates for each case in the following. For restrictions on the transformation matrices $\tilde{\mathbf{H}}_{\mathbf{A}}, \tilde{\mathbf{H}}_{\mathbf{E}}, \tilde{\mathbf{H}}_{\mathbf{A}}^{\boldsymbol{\Delta}}, \tilde{\mathbf{H}}_{\mathbf{E}}^{\boldsymbol{\Delta}}$,..., we refer to the supplementary material of [3].

Case $2a$ Linear couplings of type:

$$\tilde{\mathbf{H}}_{\mathbf{A}}\mathbf{A} = \boldsymbol{\Delta}, \quad \tilde{\mathbf{H}}_{\mathbf{E}}\mathbf{E} = \boldsymbol{\Delta} \tag{11}$$

For the update of $\mathbf{A}$, the following Sylvester equation has to be solved:

$$\begin{aligned} \frac{\rho_A}{2}\left(\mathbf{I}_{R_1} + \tilde{\mathbf{H}}_{\mathbf{A}}^T\tilde{\mathbf{H}}_{\mathbf{A}}\right)\mathbf{A}^{(n+1)} + \mathbf{A}^{(n+1)} w_1 \sum_{k=1}^{K} \mathbf{D}_k\mathbf{B}_k^T\mathbf{B}_k\mathbf{D}_k = \\ w_1 \sum_{k=1}^{K} \mathbf{X}_k\mathbf{B}_k\mathbf{D}_k + \frac{\rho_A}{2}\left[\mathbf{Z}_A^{(n)} - \boldsymbol{\mu}_{Z_A}^{(n)} + \tilde{\mathbf{H}}_{\mathbf{A}}^T\left(\boldsymbol{\Delta}^{(n)} - \boldsymbol{\mu}_{\boldsymbol{\Delta}_A}^{(n)}\right)\right] \end{aligned} \tag{12}$$

For the update of $\mathbf{E}$, the following Sylvester equation has to be solved:

$$\frac{\rho_E}{2}\left(\mathbf{I}_{R_2} + \tilde{\mathbf{H}}_{\mathbf{E}}^T\tilde{\mathbf{H}}_{\mathbf{E}}\right)\mathbf{E}^{(n+1)} + \mathbf{E}^{(n+1)} w_2\mathbf{F}^T\mathbf{F} = w_2\mathbf{Y}\mathbf{F} + \frac{\rho_E}{2}\left[\mathbf{Z}_E^{(n)} - \boldsymbol{\mu}_{Z_E}^{(n)} + \tilde{\mathbf{H}}_{\mathbf{E}}^T\left(\boldsymbol{\Delta}^{(n)} - \boldsymbol{\mu}_{\boldsymbol{\Delta}_E}^{(n)}\right)\right] \tag{13}$$

The update of $\boldsymbol{\Delta}$ is given by an average:

$$\boldsymbol{\Delta}^{(n+1)} = \tfrac{1}{\rho_A+\rho_E}\left[\rho_A\left(\tilde{\mathbf{H}}_{\mathbf{A}}\mathbf{A}^{(n+1)} + \boldsymbol{\mu}_{\boldsymbol{\Delta}_A}^{(n)}\right) + \rho_E\left(\tilde{\mathbf{H}}_{\mathbf{E}}\mathbf{E}^{(n+1)} + \boldsymbol{\mu}_{\boldsymbol{\Delta}_E}^{(n)}\right)\right] \tag{14}$$

And finally,

$$\boldsymbol{\mu}_{\boldsymbol{\Delta}_A}^{(n+1)} = \boldsymbol{\mu}_{\boldsymbol{\Delta}_A}^{(n)} + \tilde{\mathbf{H}}_{\mathbf{A}}\mathbf{A}^{(n+1)} - \boldsymbol{\Delta}^{(n+1)}, \tag{15}$$

$$\boldsymbol{\mu}_{\boldsymbol{\Delta}_E}^{(n+1)} = \boldsymbol{\mu}_{\boldsymbol{\Delta}_E}^{(n)} + \tilde{\mathbf{H}}_{\mathbf{E}}\mathbf{E}^{(n+1)} - \boldsymbol{\Delta}^{(n+1)}. \tag{16}$$

Case $2b$ Linear couplings of type:

$$\mathbf{A} = \tilde{\mathbf{H}}_{\mathbf{A}}^{\boldsymbol{\Delta}}\boldsymbol{\Delta}, \quad \mathbf{E} = \tilde{\mathbf{H}}_{\mathbf{E}}^{\boldsymbol{\Delta}}\boldsymbol{\Delta} \tag{17}$$

For the update of $\mathbf{A}$, the following linear system has to be solved:

$$\begin{aligned} \mathbf{A}^{(n+1)}\left[w_1 \sum_{k=1}^{K} \mathbf{D}_k\mathbf{B}_k^T\mathbf{B}_k\mathbf{D}_k + \frac{\rho_A}{2}\ (\mathbf{I}_{R_1} + \mathbf{I}_{R_1})\right] = \\ \left[w_1 \sum_{k=1}^{K} \mathbf{X}_k\mathbf{B}_k\mathbf{D}_k + \frac{\rho_A}{2}\left(\mathbf{Z}_A^{(n)} - \boldsymbol{\mu}_{Z_A}^{(n)} + \tilde{\mathbf{H}}_{\mathbf{A}}^{\boldsymbol{\Delta}}\boldsymbol{\Delta}^{(n)} - \boldsymbol{\mu}_{\boldsymbol{\Delta}_A}^{(n)}\right)\right] \end{aligned} \tag{18}$$

For the update of $\mathbf{E}$, the following linear system has to be solved:

$$\mathbf{E}^{(n+1)}\left[w_2\mathbf{F}^T\mathbf{F} + \frac{\rho_E}{2}\left(\mathbf{I}_{R_2} + \mathbf{I}_{R_2}\right)\right] = \left[w_2\mathbf{Y}\mathbf{F} + \frac{\rho_E}{2}\left(\mathbf{Z}_E^{(n)} - \boldsymbol{\mu}_{Z_E}^{(n)} + \tilde{\mathbf{H}}_{\mathbf{E}}^{\boldsymbol{\Delta}}\boldsymbol{\Delta}^{(n)} - \boldsymbol{\mu}_{\boldsymbol{\Delta}_E}^{(n)}\right)\right] \tag{19}$$

For the update of $\boldsymbol{\Delta}$, the following linear system has to be solved:

$$\left(\rho_A\tilde{\mathbf{H}}_{\mathbf{A}}^{\boldsymbol{\Delta}^T}\tilde{\mathbf{H}}_{\mathbf{A}}^{\boldsymbol{\Delta}} + \rho_E\tilde{\mathbf{H}}_{\mathbf{E}}^{\boldsymbol{\Delta}^T}\tilde{\mathbf{H}}_{\mathbf{E}}^{\boldsymbol{\Delta}}\right)\boldsymbol{\Delta}^{(n+1)} = \rho_A\tilde{\mathbf{H}}_{\mathbf{A}}^{\boldsymbol{\Delta}^T}\left(\mathbf{A}^{(n+1)} + \boldsymbol{\mu}_{\boldsymbol{\Delta}_A}^{(n)}\right) + \rho_E\tilde{\mathbf{H}}_{\mathbf{E}}^{\boldsymbol{\Delta}^T}\left(\mathbf{E}^{(n+1)} + \boldsymbol{\mu}_{\boldsymbol{\Delta}_E}^{(n)}\right) \tag{20}$$

And finally,

$$\boldsymbol{\mu}_{\boldsymbol{\Delta}_A}^{(n+1)} = \boldsymbol{\mu}_{\boldsymbol{\Delta}_A}^{(n)} + \mathbf{A}^{(n+1)} - \tilde{\mathbf{H}}_{\mathbf{A}}^{\boldsymbol{\Delta}}\boldsymbol{\Delta}^{(n+1)}, \tag{21}$$

$$\boldsymbol{\mu}_{\boldsymbol{\Delta}_E}^{(n+1)} = \boldsymbol{\mu}_{\boldsymbol{\Delta}_E}^{(n)} + \mathbf{E}^{(n+1)} - \tilde{\mathbf{H}}_{\mathbf{E}}^{\boldsymbol{\Delta}}\boldsymbol{\Delta}^{(n+1)}. \tag{22}$$

Case $3a$ Linear couplings of type:

$$\mathbf{A}\hat{\mathbf{H}}_{\mathbf{A}} = \boldsymbol{\Delta}, \quad \mathbf{E}\hat{\mathbf{H}}_{\mathbf{E}} = \boldsymbol{\Delta} \tag{23}$$

For the update of $\mathbf{A}$, the following linear system has to be solved:

$$\begin{aligned} \mathbf{A}^{(n+1)}\left[w_1 \sum_{k=1}^{K} \mathbf{D}_k\mathbf{B}_k^T\mathbf{B}_k\mathbf{D}_k + \frac{\rho_A}{2}\left(\mathbf{I}_{R_1} + \hat{\mathbf{H}}_{\mathbf{A}}\hat{\mathbf{H}}_{\mathbf{A}}^T\right)\right] = \\ \left[w_1 \sum_{k=1}^{K} \mathbf{X}_k\mathbf{B}_k\mathbf{D}_k + \frac{\rho_A}{2}\left(\mathbf{Z}_A^{(n)} - \boldsymbol{\mu}_{Z_A}^{(n)} + \left(\boldsymbol{\Delta}^{(n)} - \boldsymbol{\mu}_{\boldsymbol{\Delta}_A}^{(n)}\right)\hat{\mathbf{H}}_{\mathbf{A}}^T\right)\right] \end{aligned} \tag{24}$$

For the update of $\mathbf{E}$, the following linear system has to be solved:

$$\mathbf{E}^{(n+1)}\left[w_2\mathbf{F}^T\mathbf{F}+\frac{\rho_E}{2}\left(\mathbf{I}_{R_1}+\hat{\mathbf{H}}_{\mathbf{E}}\hat{\mathbf{H}}_{\mathbf{E}}^T\right)\right]=\left[w_2\mathbf{Y}\mathbf{F}+\frac{\rho_E}{2}\left(\mathbf{Z}_E^{(n)}-\boldsymbol{\mu}_{Z_E}^{(n)}+\left(\boldsymbol{\Delta}^{(n)}-\boldsymbol{\mu}_{\boldsymbol{\Delta}_E}^{(n)}\right)\hat{\mathbf{H}}_{\mathbf{E}}^T\right)\right] \tag{25}$$

The update of $\boldsymbol{\Delta}$ is given by an average:

$$\boldsymbol{\Delta}^{(n+1)}=\tfrac{1}{\rho_A+\rho_E}\left[\rho_A\left(\mathbf{A}^{(n+1)}\hat{\mathbf{H}}_{\mathbf{A}}+\boldsymbol{\mu}_{\boldsymbol{\Delta}_A}^{(n)}\right)+\rho_E\left(\mathbf{E}^{(n+1)}\hat{\mathbf{H}}_{\mathbf{E}}+\boldsymbol{\mu}_{\boldsymbol{\Delta}_E}^{(n)}\right)\right] \tag{26}$$

And finally,

$$\boldsymbol{\mu}_{\boldsymbol{\Delta}_A}^{(n+1)}=\boldsymbol{\mu}_{\boldsymbol{\Delta}_A}^{(n)}+\mathbf{A}^{(n+1)}\hat{\mathbf{H}}_{\mathbf{A}}-\boldsymbol{\Delta}^{(n+1)}, \tag{27}$$

$$\boldsymbol{\mu}_{\boldsymbol{\Delta}_E}^{(n+1)}=\boldsymbol{\mu}_{\boldsymbol{\Delta}_E}^{(n)}+\mathbf{E}^{(n+1)}\hat{\mathbf{H}}_{\mathbf{E}}-\boldsymbol{\Delta}^{(n+1)}. \tag{28}$$

Case $3b$ Linear couplings of type:

$$\mathbf{A}=\boldsymbol{\Delta}\hat{\mathbf{H}}_{\mathbf{A}}^{\boldsymbol{\Delta}},\quad \mathbf{E}=\boldsymbol{\Delta}\hat{\mathbf{H}}_{\mathbf{E}}^{\boldsymbol{\Delta}} \tag{29}$$

For the update of $\mathbf{A}$, the following linear system has to be solved:

$$\begin{aligned}\mathbf{A}^{(n+1)}&\left[w_1\sum_{k=1}^{K}\mathbf{D}_k\mathbf{B}_k^T\mathbf{B}_k\mathbf{D}_k+\frac{\rho_A}{2}\ (\mathbf{I}_{R_1}+\mathbf{I}_{R_1})\right]=\\ &\left[w_1\sum_{k=1}^{K}\mathbf{X}_k\mathbf{B}_k\mathbf{D}_k+\frac{\rho_A}{2}\left(\mathbf{Z}_A^{(n)}-\boldsymbol{\mu}_{Z_A}^{(n)}+\boldsymbol{\Delta}^{(n)}\hat{\mathbf{H}}_{\mathbf{A}}^{\boldsymbol{\Delta}}-\boldsymbol{\mu}_{\boldsymbol{\Delta}_A}^{(n)}\right)\right]\end{aligned} \tag{30}$$

For the update of $\mathbf{E}$, the following linear system has to be solved:

$$\mathbf{E}^{(n+1)}\left[w_2\mathbf{F}^T\mathbf{F}+\frac{\rho_E}{2}\left(\mathbf{I}_{R_2}+\mathbf{I}_{R_2}\right)\right]=\left[w_2\mathbf{Y}\mathbf{F}+\frac{\rho_E}{2}\left(\mathbf{Z}_E^{(n)}-\boldsymbol{\mu}_{Z_E}^{(n)}+\boldsymbol{\Delta}^{(n)}\hat{\mathbf{H}}_{\mathbf{E}}^{\boldsymbol{\Delta}}-\boldsymbol{\mu}_{\boldsymbol{\Delta}_E}^{(n)}\right)\right] \tag{31}$$

For the update of $\boldsymbol{\Delta}$, the following linear system has to be solved:

$$\boldsymbol{\Delta}^{(n+1)}\left(\rho_A\hat{\mathbf{H}}_{\mathbf{A}}^{\boldsymbol{\Delta}}\hat{\mathbf{H}}_{\mathbf{A}}^{\boldsymbol{\Delta}^T}+\rho_E\hat{\mathbf{H}}_{\mathbf{E}}^{\boldsymbol{\Delta}}\hat{\mathbf{H}}_{\mathbf{E}}^{\boldsymbol{\Delta}^T}\right)=\rho_A\left(\mathbf{A}^{(n+1)}+\boldsymbol{\mu}_{\boldsymbol{\Delta}_A}^{(n)}\right)\hat{\mathbf{H}}_{\mathbf{A}}^{\boldsymbol{\Delta}^T}+\rho_E\left(\mathbf{E}^{(n+1)}+\boldsymbol{\mu}_{\boldsymbol{\Delta}_E}^{(n)}\right)\hat{\mathbf{H}}_{\mathbf{E}}^{\boldsymbol{\Delta}^T} \tag{32}$$

And finally,

$$\boldsymbol{\mu}_{\boldsymbol{\Delta}_A}^{(n+1)}=\boldsymbol{\mu}_{\boldsymbol{\Delta}_A}^{(n)}+\mathbf{A}^{(n+1)}-\boldsymbol{\Delta}^{(n+1)}\hat{\mathbf{H}}_{\mathbf{A}}^{\boldsymbol{\Delta}}, \tag{33}$$

$$\boldsymbol{\mu}_{\boldsymbol{\Delta}_E}^{(n+1)}=\boldsymbol{\mu}_{\boldsymbol{\Delta}_E}^{(n)}+\mathbf{E}^{(n+1)}-\boldsymbol{\Delta}^{(n+1)}\hat{\mathbf{H}}_{\mathbf{E}}^{\boldsymbol{\Delta}}. \tag{34}$$

1.3 ADMM for other modes of model (5) in main paper

1.3.1 ADMM for mode A

This is just as regularized PARAFAC2 and can also be found in [1].
Subproblem for (regularized) $\mathbf{A}$:

$$\begin{aligned}\underset{\mathbf{A},\mathbf{Z}_A}{\operatorname{argmin}}\quad &\sum_{k=1}^{K}\left[\|\,\mathbf{X}_k-\mathbf{A}\mathbf{D}_k\mathbf{B}_k^T\,\|_F^2\right]+g_A(\mathbf{Z}_A)\\ \text{s.t.}\quad &\mathbf{A}=\mathbf{Z}_A\end{aligned} \tag{35}$$

Use ADMM:

Algorithm 6 ADMM for subproblem w.r.t. mode $\mathbf{A}$ of regularized PARAFAC2

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

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

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

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**

Again, we precompute the Cholesky decomposition of the matrix $\left[\sum_{k=1}^{K} \mathbf{D}_k \mathbf{B}_k^T \mathbf{B}_k \mathbf{D}_k + \frac{\rho_A}{2}\mathbf{I}\right]$ outside the ADMM loop, which has a complexity of $\mathcal{O}(R^3)$. The forward- and backward-substition at each ADMM iteration then has a complexity of $\mathcal{O}(IR^2)$.

1.3.2 ADMM updates for E, $\boldsymbol{\mu}_{\boldsymbol{\Delta}_E}$ and $\boldsymbol{\mu}_{\boldsymbol{\Delta}_C}$ in case of linear couplings

The updates for $\boldsymbol{\mu}_{\boldsymbol{\Delta}_E}$ and $\boldsymbol{\mu}_{\boldsymbol{\Delta}_C}$ in case of linear couplings is equivalent to the updates for $\boldsymbol{\mu}_{\boldsymbol{\Delta}_E}$ and $\boldsymbol{\mu}_{\boldsymbol{\Delta}_A}$ given above, when all **A**s are replaced with **C**s. The updates fro **E** are given in the following.

Case $2a$ Linear couplings of type:

$$\tilde{\mathbf{H}}_{\mathbf{C}}\mathbf{C} = \boldsymbol{\Delta}, \quad \tilde{\mathbf{H}}_{\mathbf{E}}\mathbf{E} = \boldsymbol{\Delta} \tag{36}$$

For the update of **E**, the following Sylvester equation has to be solved:

$$\begin{aligned} &\frac{\rho_E}{2}\left(\mathbf{I}_{R_2} + \tilde{\mathbf{H}}_{\mathbf{E}}^T\tilde{\mathbf{H}}_{\mathbf{E}}\right)\mathbf{E}^{(n+1)} + \mathbf{E}^{(n+1)}w_2(\mathbf{G}^T\mathbf{G} * \mathbf{F}^T\mathbf{F}) = \\ &w_2\mathbf{Y}_{[1]}(\mathbf{G}\odot\mathbf{F}) + \frac{\rho_E}{2}\left[\mathbf{Z}_E^{(n)} - \boldsymbol{\mu}_{Z_E}^{(n)} + \tilde{\mathbf{H}}_{\mathbf{E}}^T\left(\boldsymbol{\Delta}^{(n)} - \boldsymbol{\mu}_{\boldsymbol{\Delta}_E}^{(n)}\right)\right] \end{aligned} \tag{37}$$

Case $2b$ Linear couplings of type:

$$\mathbf{C} = \tilde{\mathbf{H}}_{\mathbf{C}}^{\boldsymbol{\Delta}}\boldsymbol{\Delta}, \quad \mathbf{E} = \tilde{\mathbf{H}}_{\mathbf{E}}^{\boldsymbol{\Delta}}\boldsymbol{\Delta} \tag{38}$$

For the update of **E**, the following linear system has to be solved:

$$\mathbf{E}^{(n+1)}\left[w_2(\mathbf{G}^T\mathbf{G} * \mathbf{F}^T\mathbf{F}) + \frac{\rho_E}{2}\left(\mathbf{I}_{R_2} + \mathbf{I}_{R_2}\right)\right] = \left[w_2\mathbf{Y}_{[1]}(\mathbf{G}\odot\mathbf{F}) + \frac{\rho_E}{2}\left(\mathbf{Z}_E^{(n)} - \boldsymbol{\mu}_{Z_E}^{(n)} + \tilde{\mathbf{H}}_{\mathbf{E}}^{\boldsymbol{\Delta}}\boldsymbol{\Delta}^{(n)} - \boldsymbol{\mu}_{\boldsymbol{\Delta}_E}^{(n)}\right)\right] \tag{39}$$

Case $3a$ Linear couplings of type:

$$\mathbf{C}\hat{\mathbf{H}}_{\mathbf{C}} = \boldsymbol{\Delta}, \quad \mathbf{E}\hat{\mathbf{H}}_{\mathbf{E}} = \boldsymbol{\Delta} \tag{40}$$

For the update of **E**, the following linear system has to be solved:

$$\begin{aligned} &\mathbf{E}^{(n+1)}\left[w_2(\mathbf{G}^T\mathbf{G} * \mathbf{F}^T\mathbf{F}) + \frac{\rho_E}{2}\left(\mathbf{I}_{R_1} + \hat{\mathbf{H}}_{\mathbf{E}}\hat{\mathbf{H}}_{\mathbf{E}}^T\right)\right] = \\ &\left[w_2\mathbf{Y}_{[1]}(\mathbf{G}\odot\mathbf{F}) + \frac{\rho_E}{2}\left(\mathbf{Z}_E^{(n)} - \boldsymbol{\mu}_{Z_E}^{(n)} + \left(\boldsymbol{\Delta}^{(n)} - \boldsymbol{\mu}_{\boldsymbol{\Delta}_E}^{(n)}\right)\hat{\mathbf{H}}_{\mathbf{E}}^T\right)\right] \end{aligned} \tag{41}$$

Case $3b$ Linear couplings of type:

$$\mathbf{C} = \boldsymbol{\Delta}\hat{\mathbf{H}}_{\mathbf{C}}^{\boldsymbol{\Delta}}, \quad \mathbf{E} = \boldsymbol{\Delta}\hat{\mathbf{H}}_{\mathbf{E}}^{\boldsymbol{\Delta}} \tag{42}$$

For the update of **E**, the following linear system has to be solved:

$$\mathbf{E}^{(n+1)}\left[w_2(\mathbf{G}^T\mathbf{G} * \mathbf{F}^T\mathbf{F}) + \frac{\rho_E}{2}\left(\mathbf{I}_{R_2} + \mathbf{I}_{R_2}\right)\right] = \left[w_2\mathbf{Y}_{[1]}(\mathbf{G}\odot\mathbf{F}) + \frac{\rho_E}{2}\left(\mathbf{Z}_E^{(n)} - \boldsymbol{\mu}_{Z_E}^{(n)} + \boldsymbol{\Delta}^{(n)}\hat{\mathbf{H}}_{\mathbf{E}}^{\boldsymbol{\Delta}} - \boldsymbol{\mu}_{\boldsymbol{\Delta}_E}^{(n)}\right)\right] \tag{43}$$

1.3.3 ADMM for mode F

This is equivalent to algorithm 4 above, except for the update for matrix **F** which is given by:

$$\mathbf{F}^{(n+1)} = \left[w_2\mathbf{Y}_{[2]}(\mathbf{G}\odot\mathbf{E}) + \frac{\rho_F}{2}\left(\mathbf{Z}_F^{(n)} - \boldsymbol{\mu}_{Z_F}^{(n)}\right)\right]\left[w_2(\mathbf{G}^T\mathbf{G} * \mathbf{E}^T\mathbf{E}) + \frac{\rho_F}{2}\mathbf{I}_{R_2}\right]^{-1} \tag{44}$$

1.3.4 ADMM for mode G

This is also equivalent to algorithm 4 for **G** instead of **F**, and the update for matrix **G** which is given by:

$$\mathbf{G}^{(n+1)} = \left[w_2\mathbf{Y}_{[3]}(\mathbf{F}\odot\mathbf{E}) + \frac{\rho_G}{2}\left(\mathbf{Z}_G^{(n)} - \boldsymbol{\mu}_{Z_G}^{(n)}\right)\right]\left[w_2(\mathbf{F}^T\mathbf{F} * \mathbf{E}^T\mathbf{E}) + \frac{\rho_G}{2}\mathbf{I}_{R_2}\right]^{-1} \tag{45}$$

2 Algorithm details

2.1 Stopping conditions

2.1.1 Inner ADMM loops

The varying mode $\mathbf{B}_k$ of PARAFAC2 The ADMM loop for $\mathbf{B}_k$ is stopped when a maximum number of iterations is reached, here 5, or when all of the following conditions are satisfied, (similar to [1]),

$$\begin{aligned}
&\frac{1}{K}\sum_k \left(\|\mathbf{B}_k^{(n)} - \mathbf{Z}_{B_k}^{(n)}\|_F / \|\mathbf{B}_k^{(n)}\|_F\right) \le \epsilon^{\mathrm{p,constr}} \\
&\frac{1}{K}\sum_k \left(\|\mathbf{B}_k^{(n)} - \mathbf{P}_k^{(n)}\boldsymbol{\Delta}_B^{(n)}\|_F / \|\mathbf{B}_k^{(n)}\|_F\right) \le \epsilon^{\mathrm{p,coupl}} \\
&\frac{1}{K}\sum_k \left(\|\mathbf{Z}_{B_k}^{(n+1)} - \mathbf{Z}_{B_k}^{(n)}\|_F / \|\boldsymbol{\mu}_{Z_{B_k}}^{(n)}\|_F\right) \le \epsilon^{\mathrm{d,constr}} \\
&\frac{1}{K}\sum_k \left(\|\mathbf{P}_k^{(n+1)}\boldsymbol{\Delta}_B^{(n+1)} - \mathbf{P}_k^{(n)}\boldsymbol{\Delta}_B^{(n)}\|_F / \|\boldsymbol{\mu}_{\Delta_{B_k}}^{(n)}\|_F\right) \le \epsilon^{\mathrm{d,coupl}}
\end{aligned} \tag{46}$$

where we set all tolerances to 10^{-5}.

Other modes The other ADMM loops are stopped when a maximum number of iterations is reached, here 5, or when all of the following conditions are satisfied, (here given for the coupled modes $\mathbf{C}$ and $\mathbf{E}$ of model (8) in the main paper, but equivalently for uncoupled but constrained modes, same as in [3]),

$$\begin{aligned}
&\frac{1}{2}\left(\|\mathbf{E}^{(n)} - \mathbf{Z}_E^{(n)}\|_F / \|\mathbf{E}^{(n)}\|_F + \|\mathbf{C}^{(n)} - \mathbf{Z}_A^{(n)}\|_F / \|\mathbf{C}^{(n)}\|_F\right) \le \epsilon^{\mathrm{p,constr}} \\
&\frac{1}{2}\Big(\|\mathbf{H}_E \operatorname{vec}(\mathbf{E}^{(n)}) - \mathbf{H}_E^{\Delta} \operatorname{vec}(\boldsymbol{\Delta}^{(n)})\|_2 / \|\mathbf{H}_E \operatorname{vec}(\mathbf{E}^{(n)})\|_2 \\
&\qquad + \|\mathbf{H}_C \operatorname{vec}(\mathbf{C}^{(n)}) - \mathbf{H}_C^{\Delta} \operatorname{vec}(\boldsymbol{\Delta}^{(n)})\|_2 / \|\mathbf{H}_C \operatorname{vec}(\mathbf{C}^{(n)})\|_2\Big) \le \epsilon^{\mathrm{p,coupl}} \\
&\frac{1}{2}\left(\|\mathbf{Z}_E^{(n+1)} - \mathbf{Z}_E^{(n)}\|_F / \|\boldsymbol{\mu}_{Z_E}^{(n)}\|_F + \|\mathbf{Z}_C^{(n+1)} - \mathbf{Z}_C^{(n)}\|_F / \|\boldsymbol{\mu}_{Z_C}^{(n)}\|_F\right) \le \epsilon^{\mathrm{d,constr}} \\
&\frac{1}{2}\Big(\|\mathbf{H}_E^{\Delta}\left(\operatorname{vec}(\boldsymbol{\Delta}^{(n+1)}) - \operatorname{vec}(\boldsymbol{\Delta}^{(n)})\right)\|_2 / \|\boldsymbol{\mu}_{\boldsymbol{\Delta}_E}^{(n)}\|_2 \\
&\qquad + \|\mathbf{H}_C^{\Delta}\left(\operatorname{vec}(\boldsymbol{\Delta}^{(n+1)}) - \operatorname{vec}(\boldsymbol{\Delta}^{(n)})\right)\|_2 / \|\boldsymbol{\mu}_{\boldsymbol{\Delta}_C}^{(n)}\|_2\Big) \le \epsilon^{\mathrm{d,coupl}}
\end{aligned} \tag{47}$$

where we set all tolerances to 10^{-5}.

2.1.2 Outer AO loop

The whole algorithm is terminated, when each of the following residuals $f_\star^{(n)}$, here given for the model (8) in the main paper,

$$\begin{aligned}
f_{\mathrm{tensors}}^{(n)} =& w_1 \sum_{k=1}^{K} \left\| \mathbf{X}_k - \mathbf{A}\mathbf{D}_k\mathbf{B}_k^T \right\|_F^2 + w_2 \, \| \boldsymbol{\mathcal{Y}} - [\![\mathbf{E},\mathbf{F},\mathbf{G}]\!] \, \|_F^2 \\
f_{\mathrm{couplings}}^{(n)} =& \frac{1}{2}\Big(\|\mathbf{H}_C \operatorname{vec}(\mathbf{C}^{(n)}) - \mathbf{H}_C^{\Delta} \operatorname{vec}(\boldsymbol{\Delta}^{(n)})\|_2 / \|\mathbf{H}_C \operatorname{vec}(\mathbf{C}^{(n)})\|_2 \\
& + \|\mathbf{H}_E \operatorname{vec}(\mathbf{E}^{(n)}) - \mathbf{H}_E^{\Delta} \operatorname{vec}(\boldsymbol{\Delta}^{(n)})\|_2 / \|\mathbf{H}_E \operatorname{vec}(\mathbf{E}^{(n)})\|_2\Big) \\
f_{\mathrm{PAR2constraint}}^{(n)} =& \frac{1}{K}\sum_{k=1}^{K} \left(\|\mathbf{B}_k^{(n)} - \mathbf{P}_k^{(n)}\boldsymbol{\Delta}_B^{(n)}\|_F / \|\mathbf{B}_k^{(n)}\|_F\right) \\
f_{\mathrm{constraints}}^{(n)} =& \frac{1}{6}\Big(\|\mathbf{A}^{(n)} - \mathbf{Z}_A^{(n)}\|_F / \|\mathbf{A}^{(n)}\|_F + \|\mathbf{C}^{(n)} - \mathbf{Z}_C^{(n)}\|_F / \|\mathbf{C}^{(n)}\|_F \\
& + \frac{1}{K}\sum_{k=1}^{K} \left(\|\mathbf{B}_k^{(n)} - \mathbf{Z}_{B_k}^{(n)}\|_F / \|\mathbf{B}_k^{(n)}\|_F\right) \\
& + \|\mathbf{E}^{(n)} - \mathbf{Z}_E^{(n)}\|_F / \|\mathbf{E}^{(n)}\|_F + \|\mathbf{F}^{(n)} - \mathbf{Z}_F^{(n)}\|_F / \|\mathbf{F}^{(n)}\|_F + \|\mathbf{G}^{(n)} - \mathbf{Z}_G^{(n)}\|_F / \|\mathbf{G}^{(n)}\|_F\Big)
\end{aligned} \tag{48}$$

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

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

or a predefined number of maximal outer iterations is reached. Here, we set the maximum number of outer iterations to 8000 and the outer absolute and relative tolerances to be 10^{-7} and 10^{-8}, respectively.

2.1.3 Practical recommendations for choosing stopping conditions

We usually set the maximum number of inner ADMM iterations to a reasonably small number between 5 and 10, since we do not want to solve the subproblems exactly in the beginning of the algorithm, when the initializations are probably far away from the optimal solution. As the algorithm proceeds, the inner ADMM algorithm typically terminates due to the relative tolerances ($10^{-3} - 10^{-5}$ being sufficient in our experience) and the number of iterations goes down. If this is not the case, which can sometimes happen for more difficult constraints on the factor matrices, it can help to increase the maximum number of inner iterations. When it comes to determining the stopping tolerances and maximum number of iterations for the outer AO loop, it is worth to examine convergence plots of all residuals in (48), as shown in experiments 1 and 2 of the main paper, and only stop when all residuals have flattened out or are very small. This can vary depending on the data and the model.

2.2 Choice of ρ

The step-size parameter ρ is set differently for each mode, and as described in [3]. Roughly speaking, for any factor matrix $\mathbf{M}$, we set $\rho_M = \text{trace}(\mathbf{N}^T\mathbf{N})/R$, where $\mathbf{N}$ denotes the matrix that needs to be inverted for the ADMM update of factor matrix $\mathbf{M}$ (not including the weight w) and R is the corresponding rank of the decomposition.

2.2.1 Efficient computation of the PARAFAC2 residual

As described in [3] for the residual in CP decomposition, also the PARAFAC2 residual can be computed efficiently, via the equivalent formulation

$$\sum_{k=1}^{K} \| \mathbf{X}_k - \mathbf{M}_k \|_F^2 = \sum_{k=1}^{K} \| \mathbf{X}_k \|_F^2 + \sum_{k=1}^{K} \| \mathbf{M}_k \|_F^2 - 2 \sum_{k=1}^{K} \langle \mathbf{X}_k, \mathbf{M}_k \rangle \,, \tag{49}$$

where $\mathbf{M}_k = \mathbf{A}\mathbf{D}_k\mathbf{B}_k^T$. The term $\sum_{k=1}^{K} \| \mathbf{X}_k \|_F^2$ is constant and can be precomputed. Furthermore, it holds

$$\sum_{k=1}^{K} \| \mathbf{M}_k \|_F^2 = \sum_{k=1}^{K} \mathbf{e}^T \left[\mathbf{A}^T\mathbf{A} * (\mathbf{B}_k\mathbf{D}_k)^T(\mathbf{B}_k\mathbf{D}_k)\right] \mathbf{e} = \mathbf{e}^T \left[\mathbf{A}^T\mathbf{A} * \sum_{k=1}^{K} \mathbf{D}_k\mathbf{B}_k^T\mathbf{B}_k\mathbf{D}_k\right] \mathbf{e}, \tag{50}$$

where $\mathbf{e}$ is a vector of all ones of matching length. In our code, we make sure that the mode $\mathbf{A}$ is updated last after $\mathbf{B}_k$ and $\mathbf{D}_k$. Then the term $\sum_{k=1}^{K} \mathbf{D}_k\mathbf{B}_k^T\mathbf{B}_k\mathbf{D}_k$ has already been computed in the update of $\mathbf{A}$ and can be used for residual computation. The same holds for the term

$$\sum_{k=1}^{K} \langle \mathbf{X}_k, \mathbf{M}_k \rangle = \mathbf{e}^T \left[\sum_{k=1}^{K} \mathbf{X}_k\mathbf{B}_k\mathbf{D}_k * \mathbf{A}\right] \mathbf{e}, \tag{51}$$

where $\sum_{k=1}^{K} \mathbf{X}_k\mathbf{B}_k\mathbf{D}_k$ is already computed in the last update of $\mathbf{A}$.

3 Additional Plots for Experiment 3

Here, we give an illustration about the clustering structures in the original and recovered $\mathbf{A}$ under different noise levels. Our algorithm can perform clustering effectively (see Figure 1b), and adding a ridge regularization can further improve the clustering performance even when the noise is equal to 1 (see Figure 1c).

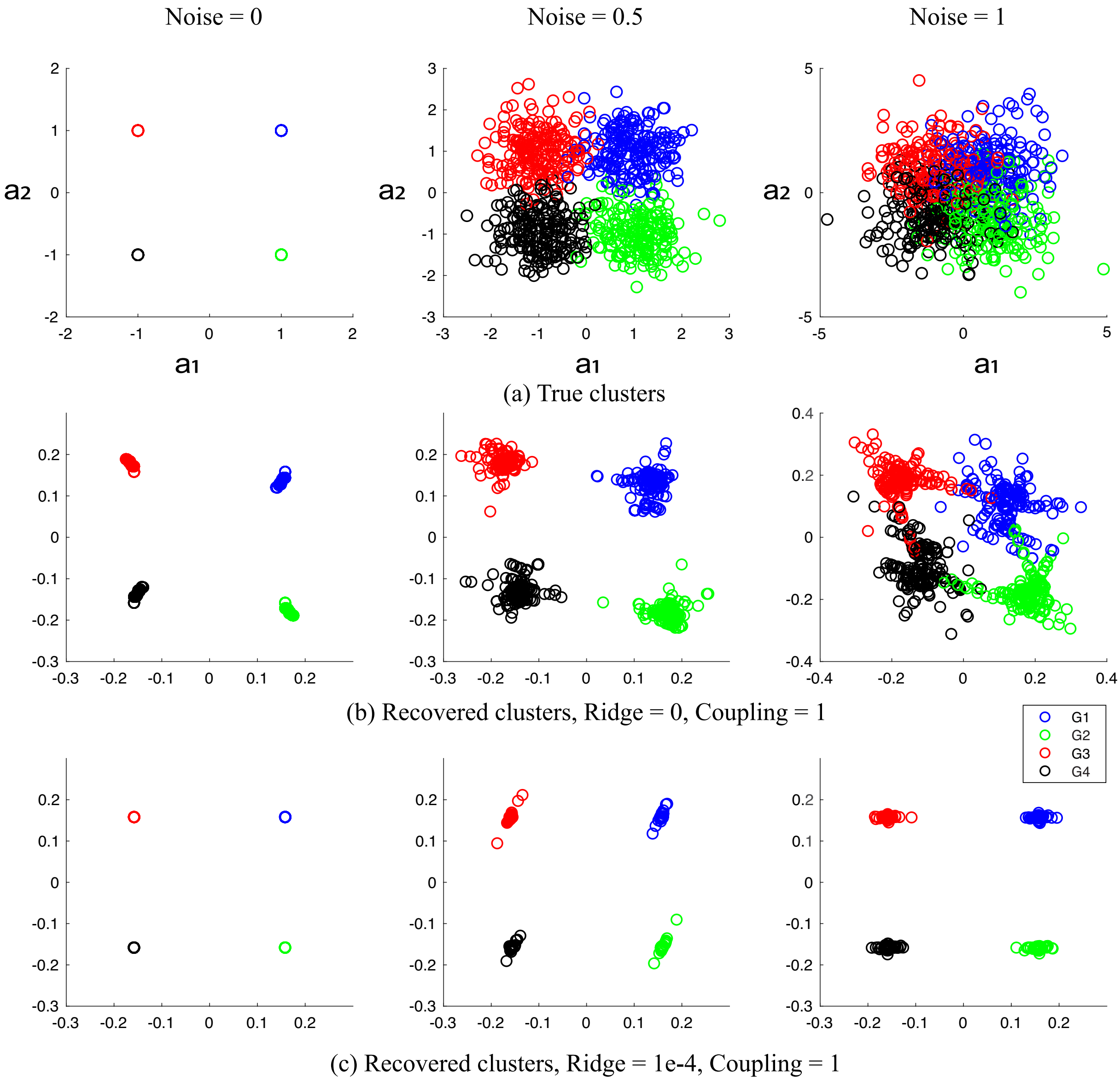

Figure 1: Exp. 3: Example of clustering structures from 20 different runs when noise = 0, 0.5 and 1 in (a) ground-truth $\mathbf{A}$, (b) recovered $\mathbf{A}$ using coupling constraint and ridge regularization, (c) recovered $\mathbf{A}$ using coupling constraint and ridge regularization $(1e-4)$.

4 Additional Plots for the Metabolomics Application

4.1 Patterns in both components

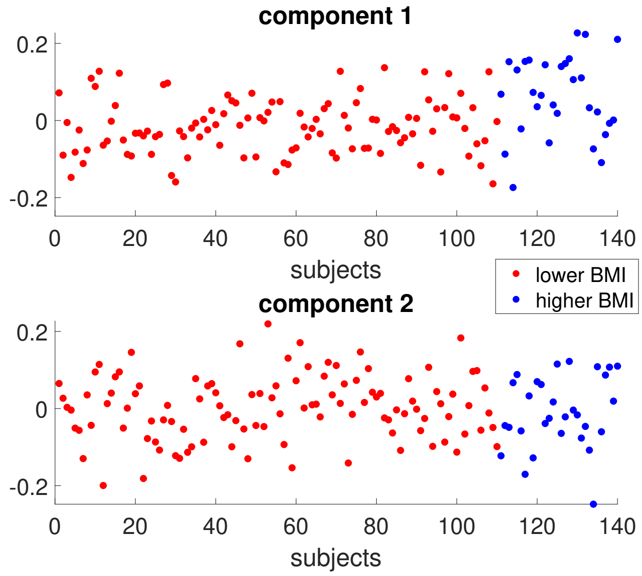

Figure 2: Subject components.

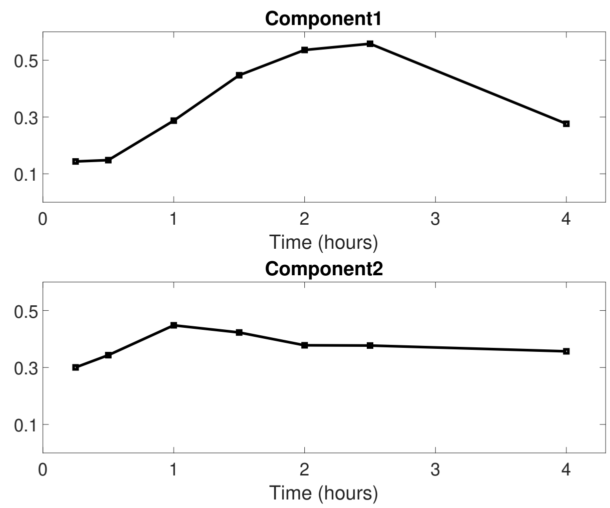

Figure 3: Time components.

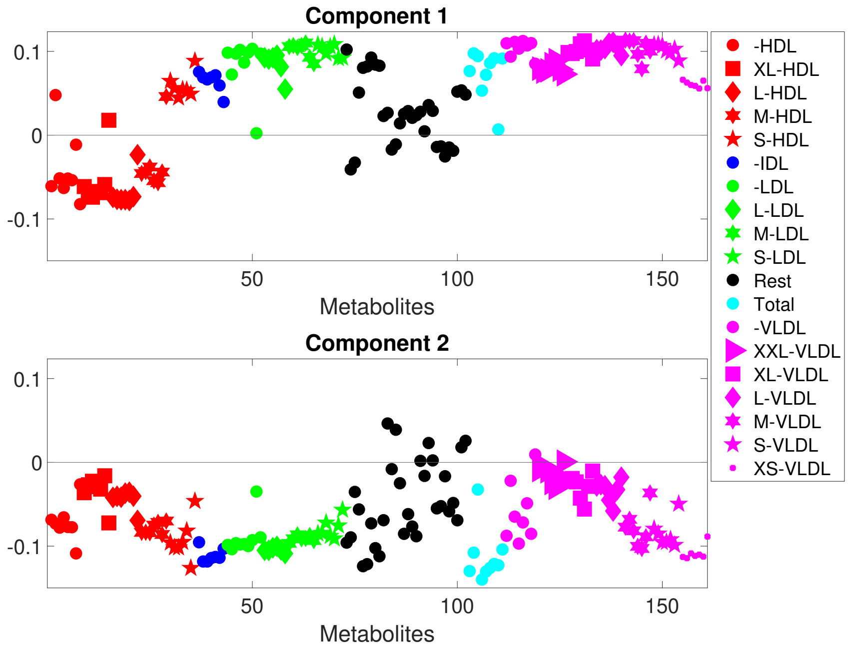

Figure 4: Components in the metabolites mode from the fasting state (static) data.

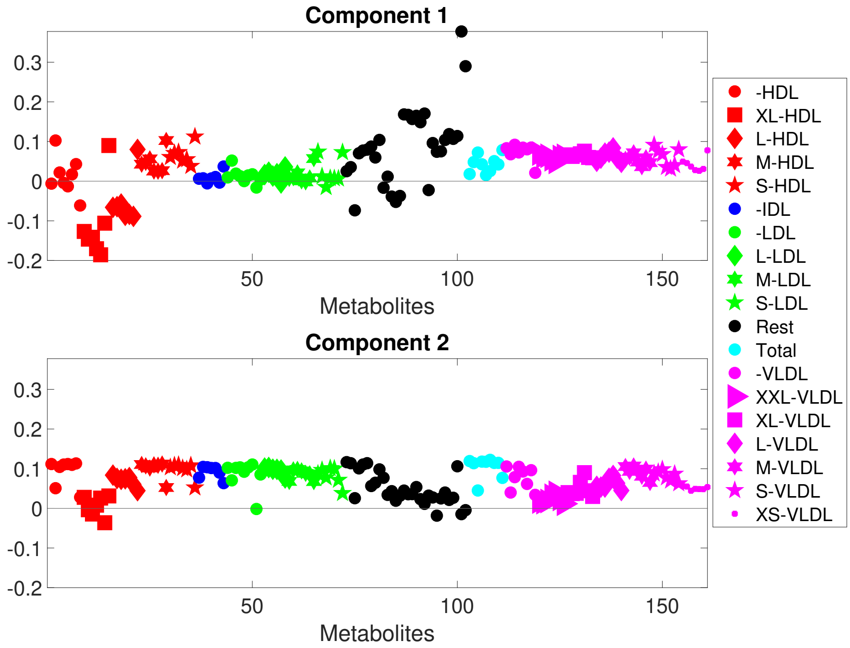

Figure 5: Components in the metabolites mode from the dynamic data - here shown for the first time point.

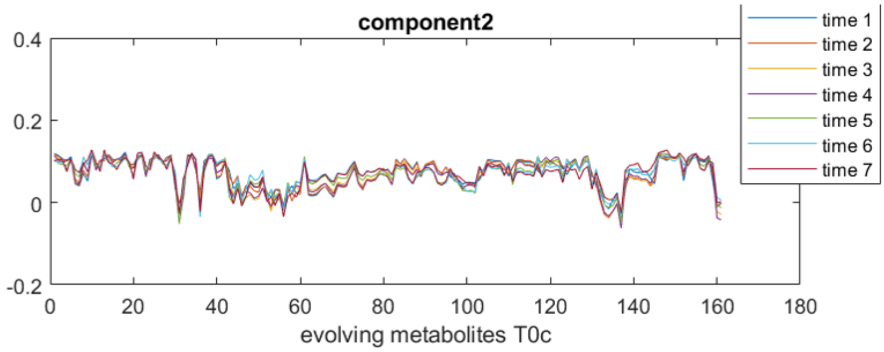

Figure 6: The time evolution of the second component in the metabolites mode from the dynamic data: almost constant. Note that the metabolites are in a different order than in Figure 5.

4.2 Model selection

We select the number of components ($R = 2$) based on the replicability of the extracted components, *i.e.,* the ability to extract similar patterns from random subsamples of the data. This approach has been previously used for the selection of number of components in CP models [4], and is essentially an extension of split-half analysis. The procedure is as follows:

1. Randomly split all subjects into ten parts, such that the proportion of subjects with lower BMI to subjects with higher BMI stays the same as in the original data.
2. Form ten different subsets by leaving out one part at a time.
3. Fit an R-component CMTF model to each subset.
4. Calculate the similarity of the extracted factors in the metabolites and time modes (using FMS) between every pair of CMTF models, *i.e.* 45 similarity scores.
5. Repeat steps $1-4$ ten times using different random splittings in step 1.

An R-component model is considered to be replicable if 95% of the total 450 computed FMS values are higher the 0.9. We then choose the highest number of components that produces a replicable model.

Here, we show the replicability results for $R = 2$ and $R = 3$ components for the PARAFAC2-based CMTF model. The horizontal line indicates the 95% highest FMS values. For $R = 2$, we can clearly see that this line is above 0.9 for both the tensor and the matrix, meaning that this model is replicable. For $R = 3$, on the other hand, the line is far below 0.9 for the matrix, indicting a non-replicable model.

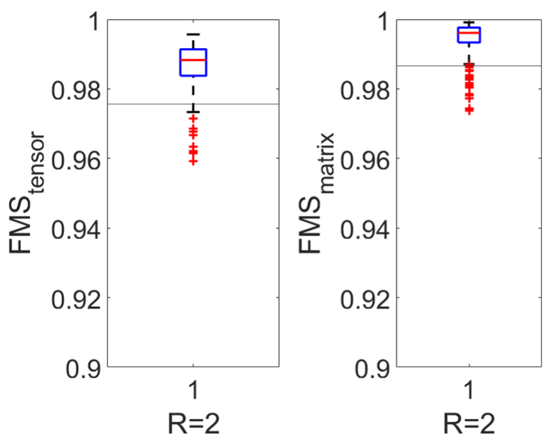

Figure 7: Replicability of a 2-component PARAFAC2-CMTF model.

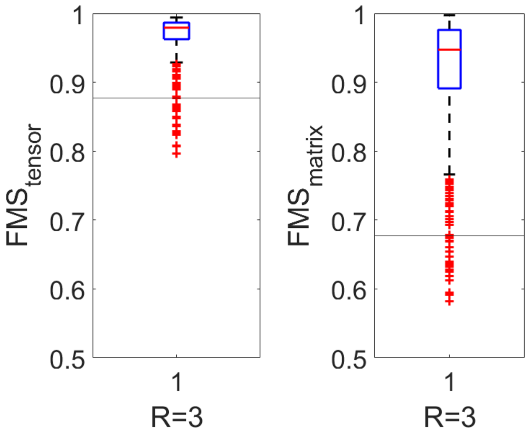

Figure 8: Replicability of a 3-component PARAFAC2-CMTF model.

5 Additional Experiment: Partial coupling in mode A and smoothness

Here, we show a version of Experiment 4 from the main paper, but with coupling in mode $\mathbf{A}$ of PARAFAC2 instead of mode $\mathbf{C}$. We construct a tensor $\boldsymbol{\mathcal{X}}$ of size $30 \times 200 \times 30$ following a PARAFAC2 model, and a tensor $\boldsymbol{\mathcal{Y}}$ of size $30 \times 20 \times 50$ following a CP model. Both have three components, but only two components are shared in the first mode (mode $\mathbf{A}$ of PARAFAC2). The components in the $\mathbf{B}_k$-mode are smooth, generated as in [1]. $\mathbf{C}$ is generated as in Exp. 1 and other factor matrices are drawn from the standard normal distribution. The noise level is set to 0.5. We then solve the following problem,

$$
\begin{aligned}
\underset{\substack{\{\mathbf{D}_k, \mathbf{B}_k\}_{k \leq K} \\ \mathbf{A}, \mathbf{E}, \mathbf{F}, \mathbf{G}, \boldsymbol{\Delta},}}{\operatorname{argmin}} \quad & \frac{1}{2} \| \mathbf{Y} - [\![\mathbf{E}, \mathbf{F}, \mathbf{G}]\!] \|_F^2 + \iota_{\mathcal{B}_1^2}(\mathbf{A}) + \iota_{\mathcal{B}_1^2}(\mathbf{E}) \\
& + \sum_{k=1}^{K} \left[\frac{1}{2} \left\| \mathbf{X}_k - \mathbf{A} \mathbf{D}_k \mathbf{B}_k^T \right\|_F^2 + \iota_{\mathcal{B}_{1,+}^2}(\mathbf{D}_k) + g_{\mathtt{GL}}(\mathbf{B}_k) \right] \\
\text{s.t.} \quad & \{\mathbf{B}_k\}_{k \leq K} \in \mathcal{P}, \quad \mathbf{A} = \boldsymbol{\Delta} \hat{\mathbf{H}}_{\mathbf{A}}^{\Delta}, \mathbf{E} = \boldsymbol{\Delta} \hat{\mathbf{H}}_{\mathbf{E}}^{\Delta},
\end{aligned}
$$

where $g_{\mathtt{GL}}$ denotes a columnwise graph laplacian regularization to promote smooth components, as in [1]. The factor vectors of modes $\mathbf{A}, \mathbf{C}$ and $\mathbf{E}$ are constrained to be inside the unit $\ell 2$-ball $\mathcal{B}_1^2$ in order to make the smoothness regularization effective. Furthermore, mode $\mathbf{C}$ is constrained to be non-negative. We use linear coupling constraints to account for the partial coupling, *i.e.*, $\hat{\mathbf{H}}_{\mathbf{A}}^{\Delta}$ indicates which columns from the "dictionary" $\boldsymbol{\Delta}$ are present in $\mathbf{A}$. The proposed algorithm recovers the true factors with FMS 0.99 for mode $\mathbf{B}$ and 1 for all other modes, yielding smooth $\mathbf{B}_k$ components (Fig. 9).

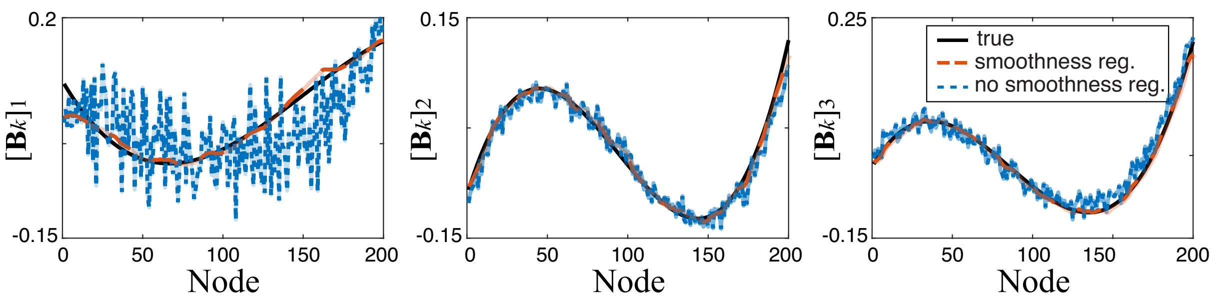

Figure 9: Exp. 4: Example of recovery of the three components of $\mathbf{B}_k$, with and without smoothness regularization.

References

[1] M. Roald, C. Schenker, V. D. Calhoun, T. Adali, R. Bro, J. E. Cohen, and E. Acar, "An AO-ADMM approach to constraining PARAFAC2 on all modes," *SIAM Journal on Mathematics of Data Science*, vol. 4, no. 3, pp. 1191–1222, 2022.

[2] C. Schenker, X. Wang, and E. Acar, "PARAFAC2-based coupled matrix and tensor factorizations," in *ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2023, pp. 1–5.

[3] C. Schenker, J. E. Cohen, and E. Acar, "A flexible optimization framework for regularized matrix-tensor factorizations with linear couplings," *IEEE Journal of Selected Topics in Signal Processing*, vol. 15, no. 3, pp. 506–521, 2021.

[4] S. Yan, L. Li, D. Horner, P. Ebrahimi, B. Chawes, L. Dragsted, M. Rasmussen, A. Smilde, and E. Acar, "Characterizing human postprandial metabolic response using multiway data analysis," *Metabolomics*, 2024.