

# Supplementary Material

## PARAFAC2-based Coupled Matrix and Tensor Factorizations

Carla Schenker, Xiulin Wang, Evrim Acar

March 23, 2023

### 1 ADMM for uncoupled modes of PARAFAC2

#### 1.1 ADMM for $\mathbf{B}_k$

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

$$\begin{aligned} \underset{\{\mathbf{B}_k\}_{k \leq K}}{\operatorname{argmin}} \quad & \sum_{k=1}^K w_1 \left\| \mathbf{X}_k - \mathbf{A} \mathbf{D}_k \mathbf{B}_k^T \right\|_F^2 + g_B(\mathbf{B}_k) \\ \text{s.t.} \quad & \{\mathbf{B}_k\}_{k \leq K} \in \mathcal{P} \end{aligned} \quad (5)$$

is solved using ADMM with the following splitting scheme:

$$\begin{aligned} \underset{\{\mathbf{B}_k\}_{k \leq K}}{\operatorname{argmin}} \quad & \sum_{k=1}^K \left[ w_1 \left\| \mathbf{X}_k - \mathbf{A} \mathbf{D}_k \mathbf{B}_k^T \right\|_F^2 + g_B(\mathbf{Z}_{B_k}) \right] + \iota_{\mathcal{P}}(\{\mathbf{W}_{B_k}\}_{k=1}^K) \\ \text{s.t.} \quad & \mathbf{B}_k = \mathbf{Z}_{B_k}, \quad \forall k \leq K, \\ & \mathbf{B}_k = \mathbf{W}_{B_k}, \quad \forall k \leq K. \end{aligned} \quad (6)$$

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 2.

---

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

---

```

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

---

Note that both FOR-loops can be computed in parallel and a Cholesky decomposition of  $[w_1 \mathbf{D}_k \mathbf{A}^T \mathbf{A} \mathbf{D}_k + \rho_{B_k} \mathbf{I}]$  is precomputed outside of the loop. 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:

$$\begin{aligned}
\left\{ \mathbf{W}_{B_k}^{(n+1)} \right\}_{k \leq K} &= \text{prox}_{\frac{1}{\rho_{B_k}} \iota_{\mathcal{P}}} \left( \left\{ \mathbf{B}_k^{(n+1)} + \boldsymbol{\mu}_{\Delta_{B_k}}^{(n)} \right\}_{k \leq K} \right) \\
&= \underset{\left\{ \mathbf{W}_{B_k} \right\}_{k \leq K}}{\text{argmin}} \quad \iota_{\mathcal{P}} \left( \left\{ \mathbf{W}_{B_k} \right\}_{k \leq K} \right) + \sum_{k=1}^K \frac{\rho_{B_k}}{2} \left\| \left( \mathbf{B}_k^{(n+1)} + \boldsymbol{\mu}_{\Delta_{B_k}}^{(n)} \right) - \mathbf{W}_{B_k} \right\|_F^2 \quad (7) \\
&= \underset{\Delta_B, \left\{ \mathbf{P}_k \right\}_{k \leq K}}{\text{argmin}} \quad \iota_{\text{orth}}(\mathbf{P}_k) + \sum_{k=1}^K \frac{\rho_{B_k}}{2} \left\| \left( \mathbf{B}_k^{(n+1)} + \boldsymbol{\mu}_{\Delta_{B_k}}^{(n)} \right) - \mathbf{P}_k \Delta_B \right\|_F^2
\end{aligned}$$

In order to efficiently approximate this, we employ an AO scheme, where the orthogonal Procrustes problem is solved independently for each  $\mathbf{P}_k$ , see Algorithm 3. We only use one iteration of Algorithm 3 for each ADMM iteration, as this is sufficient in our experience.

---

**Algorithm 3** Approximate projection onto  $\mathcal{P}$

---

```

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

```

---

## 1.2 ADMM for $\mathbf{C}$

The subproblem for regularized  $\mathbf{C}$  ( $\mathbf{D}_k = \text{Diag}(\mathbf{C}_{k,:})$ ) with split variable  $\mathbf{Z}_C$  ( $\mathbf{Z}_{C_{k,:}} = \mathbf{z}_{C_k}$ ),

$$\begin{aligned}
&\underset{\mathbf{C}, \mathbf{Z}_C}{\text{argmin}} \quad \sum_{k=1}^K \left[ w_1 \left\| \mathbf{X}_k - \mathbf{A} \text{Diag}(\mathbf{C}_{k,:}) \mathbf{B}_k^T \right\|_F^2 + g_C(\mathbf{Z}_C) \right] \\
&\text{s.t.} \quad \mathbf{C} = \mathbf{Z}_C
\end{aligned} \quad (8)$$

is vectorized as follows:

$$\begin{aligned}
&\underset{\mathbf{C}, \mathbf{Z}_C}{\text{argmin}} \quad \sum_{k=1}^K \left[ w_1 \left\| \text{vec}(\mathbf{X}_k) - (\mathbf{B}_k \odot \mathbf{A}) \mathbf{C}_{k,:}^T \right\|_2^2 + g_C(\mathbf{Z}_C) \right] \\
&\text{s.t.} \quad \mathbf{C} = \mathbf{Z}_C
\end{aligned} \quad (9)$$

Applying row-wise ADMM and the following transformations

$$\begin{aligned}
(\mathbf{B}_k \odot \mathbf{A})^T (\mathbf{B}_k \odot \mathbf{A}) &= \mathbf{A}^T \mathbf{A} * \mathbf{B}_k^T \mathbf{B}_k, \\
(\mathbf{B}_k \odot \mathbf{A})^T \text{vec}(\mathbf{X}_k) &= \text{Diag}(\mathbf{A}^T \mathbf{X}_k \mathbf{B}_k),
\end{aligned} \quad (10)$$

results in Algorithm 4, see also [1], where all rows  $k$  can be updated in parallel and a Cholesky decomposition of  $\left[ w_1 (\mathbf{A}^T \mathbf{A} * \mathbf{B}_k^T \mathbf{B}_k) + \frac{\rho_{C_k}}{2} \mathbf{I}_R \right]$  is precomputed.

---

**Algorithm 4** ADMM for subproblem w.r.t. mode  $\mathbf{C}$  of regularized PARAFAC2

---

```

1: while convergence criterion is not met do
2:   for  $k = 1, \dots, K$  do
3:      $\mathbf{C}_{k,:}^{(n+1)^T} = \underset{\mathbf{c}}{\operatorname{argmin}} w_1 \|\operatorname{vec}(\mathbf{X}_k) - (\mathbf{B}_k \odot \mathbf{A})\mathbf{c}\|_2^2 + \frac{\rho_{C_k}}{2} \left\| \mathbf{c} - \mathbf{z}_{C_k}^{(n)} + \boldsymbol{\mu}_{z_{C_k}}^{(n)} \right\|_2^2$ 
4:      $= \left[ w_1 (\mathbf{A}^T \mathbf{A} * \mathbf{B}_k^T \mathbf{B}_k) + \frac{\rho_{C_k}}{2} \mathbf{I}_R \right]^{-1} \left[ w_1 \operatorname{Diag}(\mathbf{A}^T \mathbf{X}_k \mathbf{B}_k) + \frac{\rho_{C_k}}{2} (\mathbf{z}_{C_k}^{(n)} - \boldsymbol{\mu}_{z_{C_k}}^{(n)}) \right]$ 
5:    $\mathbf{Z}_C^{(n+1)} = \operatorname{prox}_{\frac{1}{\max \rho_{C_k}} g_C} \left( \mathbf{C}^{(n+1)} + \boldsymbol{\mu}_{Z_C}^{(n)} \right)$ 
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

```

---

## 2 ADMM for $\mathbf{F}$

The regularized, but uncoupled mode of the matrix decomposition  $\mathbf{F}$  is updated using standard ADMM, see also [2], as follows:

---

**Algorithm 5** ADMM for subproblem w.r.t.  $\mathbf{F}$ 


---

```

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

```

---

## 3 Update for coupled modes for different linear couplings

For the four different types of linear couplings described in [2], the updates of  $\mathbf{A}, \mathbf{E}, \boldsymbol{\Delta}, \boldsymbol{\mu}_{\boldsymbol{\Delta}_A}$  and  $\boldsymbol{\mu}_{\boldsymbol{\Delta}_E}$  in Algorithm 1 have to be adapted. We give the specific updates for each case in the following. For restrictions on the transformation matrices  $\tilde{\mathbf{H}}_{\mathbf{A}}, \tilde{\mathbf{H}}_{\mathbf{E}}, \tilde{\mathbf{H}}_{\mathbf{A}}^{\boldsymbol{\Delta}}, \tilde{\mathbf{H}}_{\mathbf{E}}^{\boldsymbol{\Delta}}, \dots$ , we refer to the supplementary material of [2].

### 3.1 Case 2a

Linear couplings of type:

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

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

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

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

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

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

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

And finally,

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

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

### 3.2 Case 2b

Linear couplings of type:

$$\mathbf{A} = \tilde{\mathbf{H}}_A^\Delta \Delta, \quad \mathbf{E} = \tilde{\mathbf{H}}_E^\Delta \Delta \quad (17)$$

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

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

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

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

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

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

And finally,

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

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

### 3.3 Case 3a

Linear couplings of type:

$$\mathbf{A} \hat{\mathbf{H}}_A = \Delta, \quad \mathbf{E} \hat{\mathbf{H}}_E = \Delta \quad (23)$$

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

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

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

$$\mathbf{E}^{(n+1)} \left[ w_2 \mathbf{F}^T \mathbf{F} + \frac{\rho_E}{2} (\mathbf{I}_{R_2} + \hat{\mathbf{H}}_E \hat{\mathbf{H}}_E^T) \right] = \left[ w_2 \mathbf{Y} \mathbf{F} + \frac{\rho_E}{2} \left( \mathbf{Z}_E^{(n)} - \boldsymbol{\mu}_{Z_E}^{(n)} + (\Delta^{(n)} - \boldsymbol{\mu}_{\Delta_E}^{(n)}) \hat{\mathbf{H}}_E^T \right) \right] \quad (25)$$

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

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

And finally,

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

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

### 3.4 Case 3b

Linear couplings of type:

$$\mathbf{A} = \Delta \hat{\mathbf{H}}_A^\Delta, \quad \mathbf{E} = \Delta \hat{\mathbf{H}}_E^\Delta \quad (29)$$

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

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

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

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

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

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

And finally,

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

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

## 4 Stopping conditions

### 4.1 Inner ADMM loops

#### 4.1.1 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]),

$$\begin{aligned} \frac{1}{K} \sum_k \left( \|\mathbf{B}_k^{(n)} - \mathbf{Z}_{B_k}^{(n)}\|_F / \|\mathbf{B}_k^{(n)}\|_F \right) &\leq \epsilon^{\text{p,constr}} \\ \frac{1}{K} \sum_k \left( \|\mathbf{B}_k^{(n)} - \mathbf{P}_k^{(n)} \Delta_B^{(n)}\|_F / \|\mathbf{B}_k^{(n)}\|_F \right) &\leq \epsilon^{\text{p,coupl}} \\ \frac{1}{K} \sum_k \left( \|\mathbf{Z}_{B_k}^{(n+1)} - \mathbf{Z}_{B_k}^{(n)}\|_F / \|\boldsymbol{\mu}_{Z_{B_k}}^{(n)}\|_F \right) &\leq \epsilon^{\text{d,constr}} \\ \frac{1}{K} \sum_k \left( \|\mathbf{P}_k^{(n+1)} \Delta_B^{(n+1)} - \mathbf{P}_k^{(n)} \Delta_B^{(n)}\|_F / \|\boldsymbol{\mu}_{\Delta_{B_k}}^{(n)}\|_F \right) &\leq \epsilon^{\text{d,coupl}} \end{aligned} \quad (35)$$

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

### 4.1.2 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  $\mathbf{A}$  and  $\mathbf{E}$ , but equivalently for uncoupled but constrained modes, same as in [2]),

$$\begin{aligned}
& \frac{1}{2} \left( \|\mathbf{E}^{(n)} - \mathbf{Z}_E^{(n)}\|_F / \|\mathbf{E}^{(n)}\|_F + \|\mathbf{A}^{(n)} - \mathbf{Z}_A^{(n)}\|_F / \|\mathbf{A}^{(n)}\|_F \right) \leq \epsilon^{\text{p,constr}} \\
& \frac{1}{2} \left( \|\mathbf{H}_E \text{vec}(\mathbf{E}^{(n)}) - \mathbf{H}_E^\Delta \text{vec}(\Delta^{(n)})\|_2 / \|\mathbf{H}_E \text{vec}(\mathbf{E}^{(n)})\|_2 \right. \\
& \quad \left. + \|\mathbf{H}_A \text{vec}(\mathbf{A}^{(n)}) - \mathbf{H}_A^\Delta \text{vec}(\Delta^{(n)})\|_2 / \|\mathbf{H}_A \text{vec}(\mathbf{A}^{(n)})\|_2 \right) \leq \epsilon^{\text{p,coupl}} \\
& \frac{1}{2} \left( \|\mathbf{Z}_E^{(n+1)} - \mathbf{Z}_E^{(n)}\|_F / \|\boldsymbol{\mu}_{Z_E}^{(n)}\|_F + \|\mathbf{Z}_A^{(n+1)} - \mathbf{Z}_A^{(n)}\|_F / \|\boldsymbol{\mu}_{Z_A}^{(n)}\|_F \right) \leq \epsilon^{\text{d,constr}} \\
& \frac{1}{2} \left( \|\mathbf{H}_E^\Delta \left( \text{vec}(\Delta^{(n+1)}) - \text{vec}(\Delta^{(n)}) \right)\|_2 / \|\boldsymbol{\mu}_{\Delta_E}^{(n)}\|_2 \right. \\
& \quad \left. + \|\mathbf{H}_A^\Delta \left( \text{vec}(\Delta^{(n+1)}) - \text{vec}(\Delta^{(n)}) \right)\|_2 / \|\boldsymbol{\mu}_{\Delta_A}^{(n)}\|_2 \right) \leq \epsilon^{\text{d,coupl}}
\end{aligned} \tag{36}$$

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

## 4.2 Outer AO loop

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

$$\begin{aligned}
f_{\text{tensors}}^{(n)} &= w_1 \sum_{k=1}^K \|\mathbf{X}_k - \mathbf{A} \mathbf{D}_k \mathbf{B}_k^T\|_F^2 + w_2 \|\mathbf{Y} - \mathbf{E} \mathbf{F}^T\|_F^2 \\
f_{\text{couplings}}^{(n)} &= \frac{1}{2} \left( \|\mathbf{H}_A \text{vec}(\mathbf{A}^{(n)}) - \mathbf{H}_A^\Delta \text{vec}(\Delta^{(n)})\|_2 / \|\mathbf{H}_A \text{vec}(\mathbf{A}^{(n)})\|_2 \right. \\
& \quad \left. + \|\mathbf{H}_E \text{vec}(\mathbf{E}^{(n)}) - \mathbf{H}_E^\Delta \text{vec}(\Delta^{(n)})\|_2 / \|\mathbf{H}_E \text{vec}(\mathbf{E}^{(n)})\|_2 \right) \\
f_{\text{PAR2constraint}}^{(n)} &= \frac{1}{K} \sum_{k=1}^K \left( \|\mathbf{B}_k^{(n)} - \mathbf{P}_k^{(n)} \Delta_B^{(n)}\|_F / \|\mathbf{B}_k^{(n)}\|_F \right) \\
f_{\text{constraints}}^{(n)} &= \frac{1}{5} \left( \|\mathbf{A}^{(n)} - \mathbf{Z}_A^{(n)}\|_F / \|\mathbf{A}^{(n)}\|_F + \|\mathbf{C}^{(n)} - \mathbf{Z}_C^{(n)}\|_F / \|\mathbf{C}^{(n)}\|_F \right. \\
& \quad + \frac{1}{K} \sum_{k=1}^K \left( \|\mathbf{B}_k^{(n)} - \mathbf{Z}_{B_k}^{(n)}\|_F / \|\mathbf{B}_k^{(n)}\|_F \right) \\
& \quad \left. + \|\mathbf{E}^{(n)} - \mathbf{Z}_E^{(n)}\|_F / \|\mathbf{E}^{(n)}\|_F + \|\mathbf{F}^{(n)} - \mathbf{Z}_F^{(n)}\|_F / \|\mathbf{F}^{(n)}\|_F \right)
\end{aligned} \tag{37}$$

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,

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

or a predefined number of maximal outer iterations is reached. Here, we set the maximum number of outer iterations to 4000 and the outer absolute and relative tolerances to be  $10^{-7}$  and  $10^{-8}$ , respectively.

## 5 Efficient computation of the PARAFAC2 residual

As described in [2] for the residual in CP decomposition, also the PARAFAC2 residual can be computed efficiently, via the equivalent formulation

$$\sum_{k=1}^K \|\mathbf{X}_k - \mathbf{M}_k\|_F^2 = \sum_{k=1}^K \|\mathbf{X}_k\|_F^2 + \sum_{k=1}^K \|\mathbf{M}_k\|_F^2 - 2 \sum_{k=1}^K \langle \mathbf{X}_k, \mathbf{M}_k \rangle, \tag{38}$$

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

$$\sum_{k=1}^K \|\mathbf{M}_k\|_F^2 = \sum_{k=1}^K \mathbf{e}^T [\mathbf{A}^T \mathbf{A} * (\mathbf{B}_k \mathbf{D}_k)^T (\mathbf{B}_k \mathbf{D}_k)] \mathbf{e} = \mathbf{e}^T \left[ \mathbf{A}^T \mathbf{A} * \sum_{k=1}^K \mathbf{D}_k \mathbf{B}_k^T \mathbf{B}_k \mathbf{D}_k \right] \mathbf{e}, \quad (39)$$

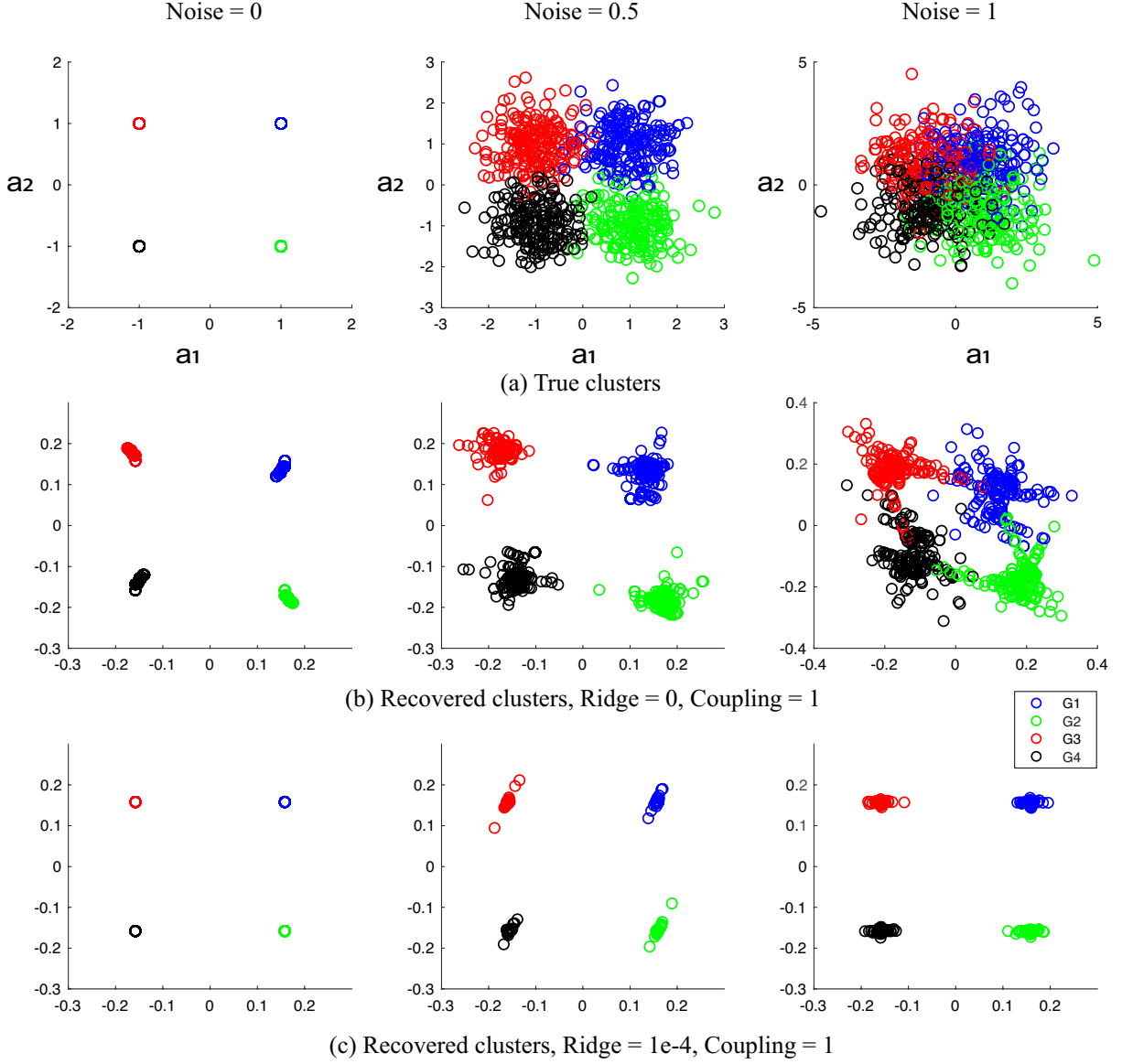
where  $\mathbf{e}$  is a vector of all ones of matching length. When the mode  $\mathbf{A}$  is updated last (after  $\mathbf{B}_k$  and  $\mathbf{D}_k$ ), then the term  $\sum_{k=1}^K \mathbf{D}_k \mathbf{B}_k^T \mathbf{B}_k \mathbf{D}_k$  has already been computed in the update of  $\mathbf{A}$  and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

$$\sum_{k=1}^K \langle \mathbf{X}_k, \mathbf{M}_k \rangle = \mathbf{e}^T \left[ \sum_{k=1}^K \mathbf{X}_k \mathbf{B}_k \mathbf{D}_k * \mathbf{A} \right] \mathbf{e}, \quad (40)$$

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

## 6 Details on Experiments

### 6.1 Experiment 2



### 6.2 Experiment 3

Table 3: Average model fit and FMS values for experiment 3.

Fit (%)		FMS					
PAR2	CP	A	B	C	E	F	G
96.16	96.19	1	0.99	1	1	1	1



## 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, 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, 2020.