tPARAFAC2: Tracking evolving patterns in (incomplete) temporal data
Supplementary material

Christos Chatzis[1,2], Carla Schenker[1], Max Pfeffer[4], and Evrim Acar[1]

[1]Simula Metropolitan Center for Digital Engineering
[2]Oslo Metropolitan University, Faculty of Art, Design and Technology
[4]Institute for Numerical and Applied Mathematics Georg-August-Universität Göttingen

1 Derivation of Row-Wise update rules

To form the update rules for each factor row, we need to solve Problems (9), (10) and (11). To simplify the notation, we have set all missing entries of input tensor $\mathcal{X}$ to zero. For (9), we need to:

$$
\begin{aligned}
\mathbf{A}^*(i,:) &= \arg\min_{\mathbf{A}(i,:)} \mathcal{L} \\
&= \arg\min_{\mathbf{A}(i,:)} \sum_{k=1}^{K} \left\| \mathbf{W}_k(i,:) * \mathbf{X}_k(i,:) - \mathbf{W}_k(i,:) * (\mathbf{A}(i,:)\mathbf{D}_k\mathbf{B}_k^\mathsf{T}) \right\|_F^2 + \frac{\rho_{A_i}}{2} \| \mathbf{A}(i,:) - \mathbf{Z}_A(i,:) + \boldsymbol{\mu}_A(i,:) \|_F^2 \\
&= \arg\min_{\mathbf{A}(i,:)} \sum_{k=1}^{K} \left\| diag(\mathbf{W}_k(i,:))\mathbf{X}_k(i,:)^\mathsf{T} - diag(\mathbf{W}_k(i,:))\mathbf{B}_k\mathbf{D}_k\mathbf{A}(i,:)^\mathsf{T} \right\|_F^2 + \frac{\rho_{A_i}}{2} \| \mathbf{A}(i,:) - \mathbf{Z}_A(i,:) + \boldsymbol{\mu}_A(i,:) \|_F^2
\end{aligned}
$$

Setting the partial derivative equal to zero:

$$
\frac{\partial \mathcal{L}}{\partial \mathbf{A}(i,:)} = 0
$$

$$
2\sum_{k=1}^{K} \left[\left(diag(\mathbf{W}_k(i,:))\mathbf{X}_k(i,:)^\mathsf{T} - diag(\mathbf{W}_k(i,:))\mathbf{B}_k\mathbf{D}_k\mathbf{A}(i,:)^\mathsf{T} \right) \left(diag(\mathbf{W}_k(i,:))\mathbf{B}_k\mathbf{D}_k \right) \right] + \rho_{A_i} \left(\mathbf{A}(i,:) - \mathbf{Z}_A(i,:) + \boldsymbol{\mu}_A(i,:) \right) = 0
$$

$$
\Big(\sum_{k=1}^{K} \mathbf{D}_k\mathbf{B}_k diag(\mathbf{W}_k(i,:)\mathbf{B}_k\mathbf{D}_k + \rho_{A_i}\mathbf{I}\Big)\mathbf{A}(i,:) - \sum_{k=1}^{K} \mathbf{X}_K\mathbf{B}_k\mathbf{D}_k - \frac{\rho_{A_i}}{2}\left(\mathbf{Z}_A(i,:) - \boldsymbol{\mu}_A(i,:) \right) = 0
$$

$$
\Leftrightarrow \mathbf{A}(i,:) = \Big(\sum_{k=1}^{K} \mathbf{X}_K\mathbf{B}_k\mathbf{D}_k + \frac{\rho_{A_i}}{2}\left(\mathbf{Z}_A(i,:) - \boldsymbol{\mu}_A(i,:) \right) \Big) \Big(\sum_{k=1}^{K} \mathbf{D}_k\mathbf{B}_k diag(\mathbf{W}_k(i,:)\mathbf{B}_k\mathbf{D}_k + \rho_{A_i}\mathbf{I} \Big)^{-1} \tag{S.1}
$$

which is the update rule shown in Equation (12). For problem (10) we work accordingly:

$$
\frac{\partial \mathcal{L}}{\partial \mathbf{B}_k(j,:)} = 0
$$

$$
\begin{aligned}
2\big(diag(\mathbf{W}_k(:,j))(\mathbf{X}_k(:,j) - (\mathbf{A}\mathbf{D}_k\mathbf{B}_k(j,:)^\mathsf{T})) \big) \big(diag(\mathbf{W}_k(:,j))\mathbf{A}\mathbf{D}_k \big) + \rho_{B_{k,j}} \big(\mathbf{B}_k(j,:) - \mathbf{Y}_{B_k}(j,:) + \boldsymbol{\mu}_{Z_{\Delta_{B_k}}}(j,:) \big) \\
+ \rho_{B_{k,j}} \big(\mathbf{B}_k(j,:)) - \mathbf{Z}_{B_k}(j,:) + \boldsymbol{\mu}_{Z_{B_k}}(j,:) \big) = 0
\end{aligned}
$$

$$
\big(\mathbf{D}_k\mathbf{A}^\mathsf{T} diag(\mathbf{W}_k(:,j))\mathbf{A}\mathbf{D}_k + \rho_{B_{k,j}}\mathbf{I} \big)\mathbf{B}_k(j,:)) - \mathbf{X}_k(:,j)^\mathsf{T}\mathbf{A}\mathbf{D}_k - \rho_{B_{k,j}} \big(\mathbf{Z}_{B_k}(j,:) - \boldsymbol{\mu}_{Z_{B_k}}(j,:) + \mathbf{Y}_{B_k}(j,:) - \boldsymbol{\mu}_{Z_{\Delta_{B_k}}}(j,:) \big) = 0
$$

$$
\Leftrightarrow \mathbf{B}_k(j,:) = \big(\rho_{B_{k,j}} \big(\mathbf{Z}_{B_k}(j,:) + \boldsymbol{\mu}_{Z_{B_k}}(j,:) + \mathbf{Y}_{B_k}(j,:) - \boldsymbol{\mu}_{Z_{\Delta_{B_k}}}(j,:) \big) \big) \big(\mathbf{D}_k\mathbf{A}^\mathsf{T} diag(\mathbf{W}_k(:,j))\mathbf{A}\mathbf{D}_k + \rho_{B_{k,j}}\mathbf{I} \big)^{-1} \tag{S.2}
$$

which is shown in Equation (13). For the last factor matrix $\mathbf{C}$:

$$
\frac{\partial \mathcal{L}}{\partial \mathbf{C}(k,:)} = 0
$$

$$\Big(\sum_{(i,j)\in\mathbf{W}_k} (\mathbf{X}_k(i,j) - (\mathbf{B}_k(j,:)\odot\mathbf{A}(i,:)) diag(\mathbf{C}(k,:))\big(\mathbf{B}_k(j,:)\odot\mathbf{A}(i,:)\big)\big) + \frac{\rho_{D_k}}{2}\big(\mathbf{D}_k - \mathbf{Z}_{D_k} + \boldsymbol{\mu}_{D_k}\big) = 0$$

$$diag(\mathbf{A}^\top\mathbf{X}_k\mathbf{B}_k) - \Big(\sum_{(i,j)\in\mathbf{W}_k} (\mathbf{A}(i,:)^\top\mathbf{A}(i,:)) * \mathbf{B}_k(j,:)^\top\mathbf{B}_k(j,:)\big) diag(\mathbf{C}(k,:) + \frac{\rho_{D_k}}{2}\big(\mathbf{D}_k - \mathbf{Z}_{D_k} + \boldsymbol{\mu}_{D_k}\big) = 0$$

$$\Leftrightarrow \mathbf{C}(k,:) = \Big(\sum_{(i,j)\in\mathbf{W}_k} (\mathbf{A}(i,:)^\top\mathbf{A}(i,:)) * \mathbf{B}_k(j,:)^\top\mathbf{B}_k(j,:)) + \frac{\rho_{D_k}}{2}\mathbf{I}\big)^{-1}\big(diag(\mathbf{A}^\top\mathbf{X}_k\mathbf{B}_k) + \frac{\rho_{D_k}}{2}(\mathbf{Z}_{D_k} - \boldsymbol{\mu}_{D_k})\big) \tag{S.3}$$

where we have utilized:

$$(\mathbf{B}_k\odot\mathbf{A})^\top(\mathbf{B}_k\odot\mathbf{A}) = \mathbf{A}^\top\mathbf{A} * \mathbf{B}_k^\top\mathbf{B}_k$$
$$(\mathbf{B}_k\odot\mathbf{A})^\top vec(\mathbf{X}_k) = diag(\mathbf{A}^\top\mathbf{X}_k\mathbf{B}_k)$$

Pseudocode for the factor updates can be found on Algorithms 1, 2 and 3.

Algorithm 1 ADMM for subproblem w.r.t. $\mathbf{A}$ of regularized PARAFAC2 with missing entries

while convergence criterion is not met **do**
  for $i = 1 : I$ **do**
    $\mathbf{A}^{(n+1)}(i,:) = \left[\sum_{k=1}^{K}\mathbf{X}_k(i,:)\mathbf{B}_k\mathbf{D}_k + \frac{\rho_{A_i}}{2}\left(\mathbf{Z}_A^{(n)}(i,:) - \boldsymbol{\mu}_{Z_A}^{(n)}(i,:)\right)\right]\left[\sum_{k=1}^{K}\mathbf{D}_k\mathbf{B}_k^\top diag(\mathbf{W}_k(i,:))\mathbf{B}_k\mathbf{D}_k + \frac{\rho_{A_i}}{2}\mathbf{I}\right]^{-1}$
  end for
  $\mathbf{Z}_A^{(n+1)} = prox_{\frac{1}{\max_i \rho_{A_i}} g_A}\left(\mathbf{A}^{(n+1)} + \boldsymbol{\mu}_{Z_A}^{(n)}\right)$
  $\boldsymbol{\mu}_{Z_A}^{(n+1)} = \boldsymbol{\mu}_{Z_A}^{(n)} + \mathbf{A}^{(n+1)} - \mathbf{Z}_A^{(n+1)}$
  $n = n + 1$
end while =0

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

while convergence criterion is not met **do**
  for $k = 1, ..., K$ **do**
    for $j = 1, ..., J_k$ **do**
      $\mathbf{B}_k^{(n+1)}(j,:) = \left[\mathbf{X}_k(:,j)^\top\mathbf{A}\mathbf{D}_k + \frac{\rho_{B_{k,j}}}{2}\left(\mathbf{Z}_{B_k}^{(n)}(j,:) - \boldsymbol{\mu}_{Z_{B_k}}^{(n)}(j,:) + \mathbf{Y}_{B_k}^{(n)}(j,:) - \boldsymbol{\mu}_{\boldsymbol{\Delta}_{B_k}}^{(n)}(j,:)\right)\right]$
      $\left[\mathbf{D}_k\mathbf{A}^T diag(\mathbf{W}_k(:,j))\mathbf{A}\mathbf{D}_k + \rho_{B_{k,j}}\mathbf{I}\right]^{-1}$
    end for
    $\mathbf{Z}_{B_k}^{(n+1)} = prox_{\frac{1}{\max_j \rho_{B_{k,j}}} g_B}\left(\mathbf{B}_k^{(n+1)} + \boldsymbol{\mu}_{Z_{B_k}}^{(n)}\right)$
  end for
  $\left\{\mathbf{Y}_{B_k}{}^{(n+1)}\right\}_{k\leq K} = prox_{\frac{1}{\max_j \rho_{B_{k,j}}}\iota_{\mathcal{P}}}\left(\left\{\mathbf{B}_k^{(n+1)} + \boldsymbol{\mu}_{\boldsymbol{\Delta}_{B_k}}^{(n)}\right\}_{k\leq K}\right) \longleftarrow$ [1, Algorithm 4]
  for $k = 1, ..., K$ **do**
    $\boldsymbol{\mu}_{Z_{B_k}}^{(n+1)} = \boldsymbol{\mu}_{Z_{B_k}}^{(n)} + \mathbf{B}_k^{(n+1)} - \mathbf{Z}_{B_k}^{(n+1)}$
    $\boldsymbol{\mu}_{\Delta_{B_k}}^{(n+1)} = \boldsymbol{\mu}_{\Delta_{B_k}}^{(n)} + \mathbf{B}_k^{(n+1)} - \mathbf{Y}_{B_k}^{(n+1)}$
  end for
  $n = n + 1$
end while =0

Algorithm 3 ADMM for subproblem w.r.t. mode **C** of regularized PARAFAC2 with missing entries

while convergence criterion is not met **do**
  for $k = 1, ..., K$ **do**
    $\mathbf{C}_{k,:}^{(n+1)^T} = \left[\sum_{(i,j)\in \mathbf{W}_k} (\mathbf{A}(i,:)^T \mathbf{A}(i,:) * \mathbf{B}_k(j,:)^T \mathbf{B}_k(j,:)) + \frac{\rho_{d_k}}{2}\mathbf{I} \right]^{-1} \left[diag(\mathbf{A}^\top \mathbf{X}_k \mathbf{B}_k) + \frac{\rho_{D_k}}{2} \left(\mathbf{Z}_{D_k}^{(n)} - \boldsymbol{\mu}_{D_k}^{(n)} \right) \right]$
  end for
  $\mathbf{Z}_{D_k}^{(n+1)} = prox_{\frac{1}{\max_k \rho_{D_k}} g_D} \left(\mathbf{D}_K^{(n+1)} + \boldsymbol{\mu}_{D_k}^{(n)} \right)$
  $\boldsymbol{\mu}_{D_k}^{(n+1)} = \boldsymbol{\mu}_{D_k}^{(n+1)} + \mathbf{D}^{(n+1)} - \mathbf{Z}_{D_k}^{(n+1)}$
  $n = n + 1$
end while $=0$

2 AO-ADMM exit conditions

As also mentioned in the main text, outer AO-ADMM iterations stop once either the absolute or the relative change in function value is smaller than the pre-defined tolerance:

$$\left\| f^{(n+1)} - f^{(n)} \right\| < \epsilon_{abs} \tag{S.4}$$

$$\frac{\left\| f^{(n+1)} - f^{(n)} \right\|}{\| f^{(n)} \|} < \epsilon_{rel} \tag{S.5}$$

where the function value at the n-th iteration is given by:

$$f^{(n)} = \left\{ \sum_{k=1}^{K} \left\| \mathbf{X}_k - \mathbf{A}^{(n)} \mathbf{D}_k^{(n)} \mathbf{B}_k^{(n)T} \right\|^2 + \lambda_A \left\| \mathbf{A}^{(n)} \right\|^2 + \sum_{k=1}^{K} (\lambda_{\mathbf{D}} \left\| \mathbf{D}_k^{(n)} \right\|^2) \right\} \tag{S.6}$$

with $\mathbf{A}^{(n)}$, $\mathbf{B}_k^{(n)}$ and $\mathbf{D}_k^{(n)}$ denote the computed factor at the n-th iteration. A maximum number of iterations is pre-set as well. Additionally, the solution has to have adequately small feasibility gaps:

$$\frac{\| \mathbf{M} - \mathbf{Z}_M \|}{\| \mathbf{M} \|} < \epsilon_{feasibility} \tag{S.7}$$

Furthermore, for the inner AO-ADMM loops (e.g. Algorithm 1 in the main text), the stopping conditions are:

$$\frac{\left\| \mathbf{M}^{(l)} - \mathbf{Z}_M^{(l)} \right\|}{\| \mathbf{M}^{(l)} \|} < \epsilon_{inner} \tag{S.8}$$

$$\frac{\left\| \mathbf{M}^{(l)} - \mathbf{Z}_M^{(l-1)} \right\|}{\left\| \mathbf{Z}_M^{(l)} \right\|} < \epsilon_{inner} \tag{S.9}$$

where l denotes the inner iteration number and $\mathbf{M}$ and $\mathbf{Z}_M$ denote any of the factor matrices (that involve regularization) and the respective auxiliary variable introduced. The maximum number of inner iterations is set to 5 in all of our experiments for the inner loops.

3 Additional information about the metabolomics application

3.1 Selection of the number of components

When fitting the PARAFAC2 model to NMR data, we select the number of components based on the replicability of the extracted components, *i.e.,* the ability to extract similar patterns (in terms of their FMS) from random subsamples of the data. This approach has been previously used for the selection of number of components in CP models [2], and is essentially an extension of split-half analysis. An R-component model is considered to be replicable if 95% of all computed FMS values are higher the 0.9. We then choose the highest number of components that produces a replicable model.

Figure S.1 shows the replicability results for $R = 2$ and $R = 3$ components for the PARAFAC2 model. The horizontal line indicates the 95% highest FMS values of $R = 3$. For $R = 2$, we can clearly see that all FMS values are above 0.9, meaning that this model is replicable. For $R = 3$, on the other hand, the line is below 0.9, indicting a non-replicable model. Therefore, we choose a 2-component PARAFAC2 model.

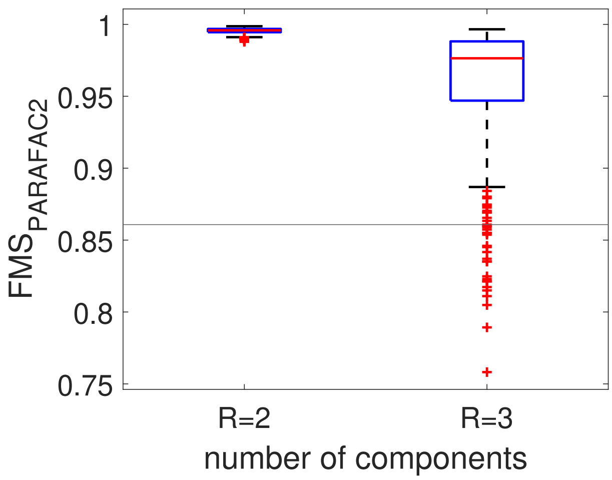

Figure S.1: Result of the replicability check for $R = 2$ and $R = 3$ components.

3.2 Correlations with meta variables

Figure S.2 shows the correlations between the subject scores from the component of interest, i.e., the one that shows a statistically significant group difference in terms of BMI, and meta variables, for the CP model [2], the PARAFAC2 model, the PARAFAC2 model with ridge ($\lambda = 1$) and tPARAFAC2 with $\lambda_B = 10$. Descriptions of meta variables are as follows: HOMAIR: Homeostatic model assessment for Insulin Resistance; MuscleFatRatio: Muscle to fat ratio; FatPercent: Body fat percentage; MuscleMass: Amount of muscle in the body (kg); Weight (kg); BMI: Body Mass Index; Waist: Waist circumferance (cm); WaistHeightRatio: Waist measurement divided by height (cm); FatMass: Amount of body fat (kg); FatMassIndex; FFMI: Fat Free Mass Index.

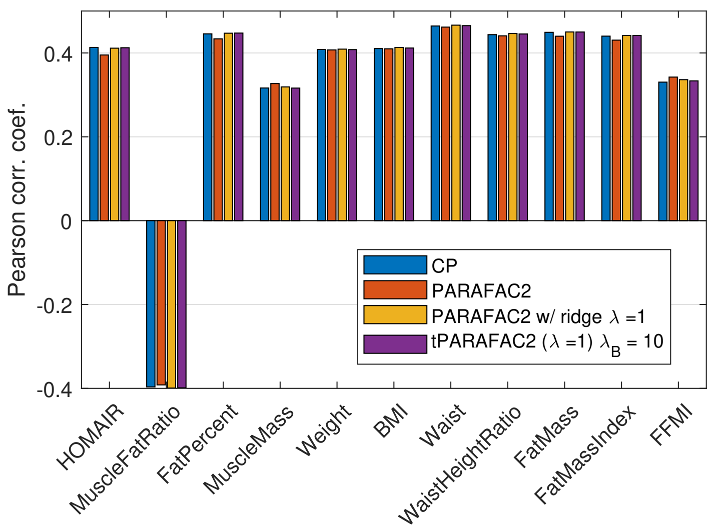

Figure S.2: Correlation coefficients of subject scores with meta variables for different models.

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] S. Yan, L. Li, D. Horner, P. Ebrahimi, B. Chawes, L. O. Dragsted, M. A. Rasmussen, A. K. Smilde, and E. Acar, "Characterizing human postprandial metabolic response using multiway data analysis," *Metabolomics*, vol. 20, no. 3, p. 50, 2024.