

ACMTF-R gradient derivation

G.R. van der Ploeg

March 2025

1 Introduction

In this text, the derivative of the ACMTF-R loss function is reported. This is done by first repeating the loss function of ACMTF-R and subsequently writing out some of the terms for convenience. Then the gradient is derived term by term.

The loss function of ACMTF-R is repeated here for convenience to the reader, where the assumptions that every term is multiplied by $1/2$ and that the first mode loadings A are shared between all blocks are applied. In addition, each of the six terms are defined as separate functions f_1, f_2, \dots, f_6 such that later derivations can be done per term.

$$\begin{aligned}
f(\alpha, \beta, \epsilon, \pi, \Lambda, \rho, \mathbf{A}, \mathbf{B}^{(1)}, \mathbf{C}^{(1)}, \dots, \mathbf{B}^{(P)}, \mathbf{C}^{(P)}) &= \underbrace{\frac{1}{2} \pi \sum_{p=1}^P \|\underline{\mathbf{X}}^{(p)} - \hat{\underline{\mathbf{X}}}^{(p)}\|^2}_{f_1} \\
&+ \underbrace{\frac{1}{2} (1 - \pi) \|\mathbf{y} - \mathbf{A}\rho\|^2}_{f_2} \\
&+ \underbrace{\frac{1}{2} \beta \sum_{p=1}^P \sum_{f=1}^F \sqrt{\lambda_{fp}^2 + \epsilon}}_{f_3} \\
&+ \underbrace{\frac{1}{2} \alpha \sum_{p=1}^P \sum_{f=1}^F (\|\mathbf{a}_f\| - 1)^2}_{f_4} \\
&+ \underbrace{\frac{1}{2} \alpha \sum_{p=1}^P \sum_{f=1}^F (\|\mathbf{b}_f^{(p)}\| - 1)^2}_{f_5} \\
&+ \underbrace{\frac{1}{2} \alpha \sum_{p=1}^P \sum_{f=1}^F (\|\mathbf{c}_f^{(p)}\| - 1)^2}_{f_6}
\end{aligned} \tag{1}$$

where the $F \times P$ matrix $\Lambda = [\lambda_1 \ \lambda_2 \ \dots \ \lambda_p]$ specifies the norms per component and per data block, $\alpha > 0$ is a penalty setting for norm-1 components, $\beta > 0$ is a sparsity penalty setting for Λ , $\epsilon > 0$ is a small term added to Λ elements to make the loss function differentiable, π is the tuning parameter, ρ is the vector of the regression coefficients of the shared subject loadings in \mathbf{A} onto \mathbf{y} , F is the number of components, and P is the number of blocks.

The aim of this section is to compute the gradient of the loss function

$$\begin{aligned}
\nabla f &= [\text{vec}(\frac{\partial f}{\partial \mathbf{A}})^T \text{vec}(\frac{\partial f}{\partial \mathbf{B}^{(1)}})^T \text{vec}(\frac{\partial f}{\partial \mathbf{C}^{(1)}})^T \dots \text{vec}(\frac{\partial f}{\partial \mathbf{B}^{(P)}})^T \text{vec}(\frac{\partial f}{\partial \mathbf{C}^{(P)}})^T \\
&\quad (\frac{\partial f}{\partial \lambda_1})^T (\frac{\partial f}{\partial \lambda_2})^T \dots (\frac{\partial f}{\partial \lambda_p})^T]^T.
\end{aligned} \tag{2}$$

which will be done by computing this derivative per term.

We briefly note here that the partial derivative of f with respect to $\lambda_1, \lambda_2, \dots, \lambda_p$ is the same as originally derived for ACMTF [1] and will not be repeated here.

2 Useful equations

The first term of the loss function f_1 can be expressed element-wise as

$$\begin{aligned}
f_1 &= \frac{1}{2} \pi \sum_{p=1}^P \|\underline{\mathbf{X}}^{(p)} - \hat{\underline{\mathbf{X}}}^{(p)}\|^2 \\
&= \frac{1}{2} \pi \sum_{p=1}^P \left(\sqrt{\sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^K (x_{ijk}^{(p)} - \hat{x}_{ijk}^{(p)})^2} \right)^2 \\
&= \frac{1}{2} \pi \sum_{p=1}^P \sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^K (x_{ijk}^{(p)} - \hat{x}_{ijk}^{(p)})^2 \\
&= \frac{1}{2} \pi \sum_{p=1}^P \sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^K \left(x_{ijk}^{(p)} - \sum_{f=1}^F \lambda_{f,p} a_{if} b_{jf}^{(p)} c_{kf}^{(p)} \right)^2 \\
&= \frac{1}{2} \pi \sum_{p=1}^P \sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^K \left[(x_{ijk}^{(p)})^2 - 2x_{ijk}^{(p)} \sum_{f=1}^F \lambda_{f,p} a_{if} b_{jf}^{(p)} c_{kf}^{(p)} \right. \\
&\quad \left. + \left(\sum_{f=1}^F \lambda_{f,p} a_{if} b_{jf}^{(p)} c_{kf}^{(p)} \right)^2 \right].
\end{aligned} \tag{3}$$

The second term of the loss function f_2 can be expressed element-wise as

$$\begin{aligned}
f_2 &= \frac{1}{2} (1 - \pi) \|\mathbf{y} - \mathbf{A}\boldsymbol{\rho}\|^2 \\
&= \frac{1}{2} (1 - \pi) \sqrt{\sum_{i=1}^I \left(y_i - \sum_{f=1}^F a_{if} \rho_f \right)^2}^2 \\
&= \frac{1}{2} (1 - \pi) \sum_{i=1}^I \left(y_i - \sum_{f=1}^F a_{if} \rho_f \right)^2 \\
&= \frac{1}{2} (1 - \pi) \sum_{i=1}^I \left[y_i^2 - 2y_i \sum_{f=1}^F a_{if} \rho_f + \left(\sum_{f=1}^F a_{if} \rho_f \right)^2 \right].
\end{aligned} \tag{4}$$

3 Derivative of \mathbf{A}

In this section the partial derivative of the loss function with respect to \mathbf{A} is reported through

$$\begin{aligned}
\frac{\partial f}{\partial A} &= \frac{\partial f_1}{\partial A} + \frac{\partial f_2}{\partial A} + \frac{\partial f_3}{\partial A} + \frac{\partial f_4}{\partial A} + \frac{\partial f_5}{\partial A} + \frac{\partial f_6}{\partial A} \\
&= \frac{\partial f_1}{\partial A} + \frac{\partial f_2}{\partial A} + \frac{\partial f_4}{\partial A}
\end{aligned} \tag{5}$$

where $\frac{\partial f_3}{\partial A} = \frac{\partial f_5}{\partial A} = \frac{\partial f_6}{\partial A} = 0$ since those terms of the loss function do not contain relationships with A .

The element-wise definition of f_1 in Equation 3 can be used to express the partial derivative with respect to A as

$$\begin{aligned}
\frac{\partial f_1}{\partial a_{if}} &= \frac{1}{2}\pi \sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^K \left[-2x_{ijk}^{(p)} \sum_{f=1}^F \lambda_{f,p} b_{jf}^{(p)} c_{kf}^{(p)} \right. \\
&\quad \left. + 2 \left(\sum_{f=1}^F \lambda_{f,p} a_{if} b_{jf}^{(p)} c_{kf}^{(p)} \right) \left(\sum_{f=1}^F \lambda_{f,p} b_{jf}^{(p)} c_{kf}^{(p)} \right) \right] \\
&= \pi \sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^K \left[\left(-x_{ijk}^{(p)} + \sum_{f=1}^F \lambda_{f,p} a_{if} b_{jf}^{(p)} c_{kf}^{(p)} \right) \left(\sum_{f=1}^F \lambda_{f,p} b_{jf}^{(p)} c_{kf}^{(p)} \right) \right] \\
&= \pi \sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^K \left[\left(\sum_{f=1}^F \lambda_{f,p} a_{if} b_{jf}^{(p)} c_{kf}^{(p)} - x_{ijk}^{(p)} \right) \left(\sum_{f=1}^F \lambda_{f,p} b_{jf}^{(p)} c_{kf}^{(p)} \right) \right]
\end{aligned} \tag{6}$$

which can then be written in matrix notation as

$$\frac{\partial f_1}{\partial A} = \pi \sum_{p=1}^P (\hat{\mathbf{X}}_a^{(p)} - \mathbf{X}_a^{(p)}) (\boldsymbol{\lambda}_p \odot \mathbf{C}^{(p)} \odot \mathbf{B}^{(p)}). \tag{7}$$

The element-wise definition of f_2 in Equation 4 can be used to express the partial derivative with respect to A as

$$\begin{aligned}
\frac{\partial f_2}{\partial a_{if}} &= \frac{1}{2}(1 - \pi) \sum_{i=1}^I \left[-2y_i \sum_{f=1}^F \rho_f + 2 \left(\sum_{f=1}^F a_{if} \rho_f \right) \left(\sum_{f=1}^F \rho_f \right) \right] \\
&= (1 - \pi) \sum_{i=1}^I \left[-y_i \sum_{f=1}^F \rho_f + \left(\sum_{f=1}^F a_{if} \rho_f \right) \left(\sum_{f=1}^F \rho_f \right) \right] \\
&= (1 - \pi) \sum_{i=1}^I \left[\left(-y_i + \sum_{f=1}^F a_{if} \rho_f \right) \left(\sum_{f=1}^F \rho_f \right) \right] \\
&= (1 - \pi) \sum_{i=1}^I \left[\left(\sum_{f=1}^F a_{if} \rho_f - y_i \right) \left(\sum_{f=1}^F \rho_f \right) \right]
\end{aligned} \tag{8}$$

which can then be written in matrix notation as

$$\frac{\partial f_2}{\partial \mathbf{A}} = (1 - \pi)(\mathbf{A}\boldsymbol{\rho} - \mathbf{y})\boldsymbol{\rho}^T. \quad (9)$$

The partial derivative of f_4 with respect to \mathbf{A} is the same as originally defined [1] for ACMTF

$$\frac{\partial f_4}{\partial \mathbf{A}} = \alpha(\mathbf{A} - \bar{\mathbf{A}}). \quad (10)$$

Combining the terms as defined in Equations 7, 9, and 10, this yields the full partial derivative of the loss function f with respect to \mathbf{A}

$$\begin{aligned} \frac{\partial f}{\partial \mathbf{A}} &= \frac{\partial f_1}{\partial \mathbf{A}} + \frac{\partial f_2}{\partial \mathbf{A}} + \frac{\partial f_4}{\partial \mathbf{A}} \\ &= \pi \sum_{p=1}^P (\hat{\mathbf{X}}_a^{(p)} - \mathbf{X}_a^{(p)}) (\boldsymbol{\lambda}_p \odot \mathbf{C}^{(p)} \odot \mathbf{B}^{(p)}) \\ &\quad + (1 - \pi)(\mathbf{A}\boldsymbol{\rho} - \mathbf{y})\boldsymbol{\rho}^T + \alpha(\mathbf{A} - \bar{\mathbf{A}}). \end{aligned} \quad (11)$$

4 Derivative of \mathbf{B}

In this section the partial derivative of the loss function with respect to $\mathbf{B}^{(p)}$ is reported through

$$\begin{aligned} \frac{\partial f}{\partial \mathbf{B}^{(p)}} &= \frac{\partial f_1}{\partial \mathbf{B}^{(p)}} + \frac{\partial f_2}{\partial \mathbf{B}^{(p)}} + \frac{\partial f_3}{\partial \mathbf{B}^{(p)}} + \frac{\partial f_4}{\partial \mathbf{B}^{(p)}} + \frac{\partial f_5}{\partial \mathbf{B}^{(p)}} + \frac{\partial f_6}{\partial \mathbf{B}^{(p)}} \\ &= \frac{\partial f_1}{\partial \mathbf{B}^{(p)}} + \frac{\partial f_5}{\partial \mathbf{B}^{(p)}} \end{aligned} \quad (12)$$

where $\frac{\partial f_2}{\partial \mathbf{B}^{(p)}} = \frac{\partial f_3}{\partial \mathbf{B}^{(p)}} = \frac{\partial f_4}{\partial \mathbf{B}^{(p)}} = \frac{\partial f_6}{\partial \mathbf{B}^{(p)}} = 0$ since those terms of the loss function do not contain relationships with $\mathbf{B}^{(p)}$.

The element-wise definition of f_1 in Equation 3 can be used to express the partial derivative with respect to $\mathbf{B}^{(p)}$ as

$$\begin{aligned}
\frac{\partial f_1}{\partial b_{jf}^{(p)}} &= \frac{1}{2} \pi \sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^K \left[-2x_{ijk}^{(p)} \sum_{f=1}^F \lambda_{f,p} a_{if} c_{kf}^{(p)} \right. \\
&\quad \left. + 2 \left(\sum_{f=1}^F \lambda_{f,p} a_{if} b_{jf}^{(p)} c_{kf}^{(p)} \right) \left(\sum_{f=1}^F \lambda_{f,p} a_{if} c_{kf}^{(p)} \right) \right] \\
&= \pi \sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^K \left[\left(-x_{ijk}^{(p)} + \sum_{f=1}^F \lambda_{f,p} a_{if} b_{jf}^{(p)} c_{kf}^{(p)} \right) \left(\sum_{f=1}^F \lambda_{f,p} a_{if} c_{kf}^{(p)} \right) \right] \\
&= \pi \sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^K \left[\left(\sum_{f=1}^F \lambda_{f,p} a_{if} b_{jf}^{(p)} c_{kf}^{(p)} - x_{ijk}^{(p)} \right) \left(\sum_{f=1}^F \lambda_{f,p} a_{if} c_{kf}^{(p)} \right) \right]
\end{aligned} \tag{13}$$

which can then be written in matrix notation as

$$\frac{\partial f_1}{\partial \mathbf{B}^{(p)}} = \pi \sum_{p=1}^P (\hat{\mathbf{X}}_a^{(p)} - \mathbf{X}_a^{(p)}) (\boldsymbol{\lambda}_p \odot \mathbf{C}^{(p)} \odot \mathbf{A}). \tag{14}$$

The partial derivative of f_5 with respect to $\mathbf{B}^{(p)}$ is the same as originally defined [1] for ACMTF

$$\frac{\partial f_5}{\partial \mathbf{B}^{(p)}} = \alpha (\mathbf{B}^{(p)} - \bar{\mathbf{B}}^{(p)}). \tag{15}$$

Combining the terms as defined in Equations 14 and 15, this yields the full partial derivative of the loss function f with respect to $\mathbf{B}^{(p)}$

$$\begin{aligned}
\frac{\partial f}{\partial \mathbf{B}^{(p)}} &= \frac{\partial f_1}{\partial \mathbf{B}^{(p)}} + \frac{\partial f_5}{\partial \mathbf{B}^{(p)}} \\
&= \pi \sum_{p=1}^P (\hat{\mathbf{X}}_a^{(p)} - \mathbf{X}_a^{(p)}) (\boldsymbol{\lambda}_p \odot \mathbf{C}^{(p)} \odot \mathbf{A}) + \alpha (\mathbf{B}^{(p)} - \bar{\mathbf{B}}^{(p)}).
\end{aligned} \tag{16}$$

5 Derivative of \mathbf{C}

In this section the partial derivative of the loss function with respect to $\mathbf{C}^{(p)}$ is reported through

$$\begin{aligned}
\frac{\partial f}{\partial \mathbf{C}^{(p)}} &= \frac{\partial f_1}{\partial \mathbf{C}^{(p)}} + \frac{\partial f_2}{\partial \mathbf{C}^{(p)}} + \frac{\partial f_3}{\partial \mathbf{C}^{(p)}} + \frac{\partial f_4}{\partial \mathbf{C}^{(p)}} + \frac{\partial f_5}{\partial \mathbf{C}^{(p)}} + \frac{\partial f_6}{\partial \mathbf{C}^{(p)}} \\
&= \frac{\partial f_1}{\partial \mathbf{C}^{(p)}} + \frac{\partial f_6}{\partial \mathbf{C}^{(p)}}
\end{aligned} \tag{17}$$

where $\frac{\partial f_2}{\partial \mathbf{C}^{(p)}} = \frac{\partial f_3}{\partial \mathbf{C}^{(p)}} = \frac{\partial f_4}{\partial \mathbf{C}^{(p)}} = \frac{\partial f_5}{\partial \mathbf{C}^{(p)}} = 0$ since those terms of the loss function do not contain relationships with $\mathbf{C}^{(p)}$.

The element-wise definition of f_1 in Equation 3 can be used to express the partial derivative with respect to $\mathbf{C}^{(p)}$ as

$$\begin{aligned}
\frac{\partial f_1}{\partial c_{kf}^{(p)}} &= \frac{1}{2} \pi \sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^K \left[-2x_{ijk}^{(p)} \sum_{f=1}^F \lambda_{f,p} b_{jf}^{(p)} c_{kf}^{(p)} \right. \\
&\quad \left. + 2 \left(\sum_{f=1}^F \lambda_{f,p} a_{if} b_{jf}^{(p)} c_{kf}^{(p)} \right) \left(\sum_{f=1}^F \lambda_{f,p} b_{jf}^{(p)} c_{kf}^{(p)} \right) \right] \\
&= \pi \sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^K \left[\left(-x_{ijk}^{(p)} + \sum_{f=1}^F \lambda_{f,p} a_{if} b_{jf}^{(p)} c_{kf}^{(p)} \right) \left(\sum_{f=1}^F \lambda_{f,p} b_{jf}^{(p)} c_{kf}^{(p)} \right) \right] \\
&= \pi \sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^K \left[\left(\sum_{f=1}^F \lambda_{f,p} a_{if} b_{jf}^{(p)} c_{kf}^{(p)} - x_{ijk}^{(p)} \right) \left(\sum_{f=1}^F \lambda_{f,p} b_{jf}^{(p)} c_{kf}^{(p)} \right) \right]
\end{aligned} \tag{18}$$

which can then be written in matrix notation as

$$\frac{\partial f_1}{\partial \mathbf{C}^{(p)}} = \pi \sum_{p=1}^P (\hat{\mathbf{X}}_a^{(p)} - \mathbf{X}_a^{(p)}) (\boldsymbol{\lambda}_p \odot \mathbf{C}^{(p)} \odot \mathbf{B}^{(p)}). \tag{19}$$

The partial derivative of f_5 with respect to $\mathbf{C}^{(p)}$ is the same as originally defined [1] for ACMTF

$$\frac{\partial f_6}{\partial \mathbf{C}^{(p)}} = \alpha (\mathbf{C}^{(p)} - \bar{\mathbf{C}}^{(p)}). \tag{20}$$

Combining the terms as defined in Equations 19 and 20, this yields the full partial derivative of the loss function f with respect to $\mathbf{C}^{(p)}$

$$\begin{aligned}
\frac{\partial f}{\partial \mathbf{C}^{(p)}} &= \frac{\partial f_1}{\partial \mathbf{C}^{(p)}} + \frac{\partial f_6}{\partial \mathbf{C}^{(p)}} \\
&= \pi \sum_{p=1}^P (\hat{\mathbf{X}}_a^{(p)} - \mathbf{X}_a^{(p)}) (\boldsymbol{\lambda}_p \odot \mathbf{C}^{(p)} \odot \mathbf{B}^{(p)}) + \alpha (\mathbf{C}^{(p)} - \bar{\mathbf{C}}^{(p)}).
\end{aligned} \tag{21}$$

References

1. Acar, E. *et al.* Structure-revealing data fusion. en. *BMC Bioinformatics* **15**, 239. ISSN: 1471-2105. <https://doi.org/10.1186/1471-2105-15-239> (2024) (July 2014).