A Appendix

Derivation of the data-fitting term of (12). We make use of the multi-linearity property of the CPD and rely on re-ordering the summations:

\[ \langle W, \mathcal{Z}(x) \rangle_F = \left( \sum_{r=1}^{R} w_r^{(1)} \otimes w_r^{(2)} \otimes \ldots \otimes w_r^{(D)}, z^{(1)} \otimes z^{(2)} \otimes \ldots \otimes z^{(D)} \right)_F \]

\[ = \sum_{i_1=1}^{\hat{M}} \ldots \sum_{i_D=1}^{\hat{M}} \sum_{r=1}^{R} w_{i_1}^{(1)} z_{i_1} \ldots w_{i_D}^{(D)} z_{i_D} \]

\[ = \sum_{i_1=1}^{\hat{M}} \ldots \sum_{i_D=1}^{\hat{M}} R \sum_{r=1}^{R} w_{i_1}^{(1)} \ldots w_{i_D}^{(D)} z_{i_1}^{(1)} \ldots z_{i_D}^{(D)} \]

\[ = \text{vec} \left( W^{(d)} \right)^T \left( z^{(d)} \otimes (z^{(1)})^T W^{(1)} \otimes \ldots \otimes (z^{(D)})^T W^{(D)} \right) \]

The derivation of the regularization term of (13) follows a similar reasoning as for the data-fitting term:

\[ \langle W, W \rangle_F = \left( \sum_{r=1}^{R} w_r^{(1)} \otimes w_r^{(2)} \otimes \ldots \otimes w_r^{(D)}, \sum_{r=1}^{R} w_r^{(1)} \otimes w_r^{(2)} \otimes \ldots \otimes w_r^{(D)} \right)_F \]

\[ = \sum_{i_1=1}^{\hat{M}} \ldots \sum_{i_D=1}^{\hat{M}} \sum_{r=1}^{R} \sum_{r=1}^{R} w_{i_1}^{(1)} \ldots w_{i_D}^{(D)} \]

\[ = \sum_{i_1=1}^{\hat{M}} \ldots \sum_{i_D=1}^{\hat{M}} R \sum_{r=1}^{R} \sum_{r=1}^{R} w_{i_1}^{(1)} \ldots w_{i_D}^{(D)} \]

\[ = \text{vec} \left( W^{(d)} \right)^T \left( \text{vec} \left( W^{(1)} \right) \otimes \ldots \otimes \text{vec} \left( W^{(D)} \right) \right) \]

\[ \left( \text{vec} \left( W^{(d)} \right), \text{vec} \left( W^{(d)} \right) \right) \]

The derivation of the regularization term of (13) follows a similar reasoning as for the data-fitting term: