APPENDIX

In this section we discuss in detail how the Tensor Decomposition problem presented in Eq. (9) is solved by the CP-OPT algorithm of [1].

In Section 4, we have seen that an N-D tensor can be represented as the sum of the outer products of 1-D filters:

$$\mathcal{F} \approx \sum_{k=1}^{K} \mathbf{a}^{k,1} \circ \mathbf{a}^{k,2} \circ \cdots \mathbf{a}^{k,N}.$$

The sum of the outer products of 1-D filters can be written using a shorthand *Kruskal Operator* notation:

$$[[\mathbf{A}^{(1)} \dots \mathbf{A}^{(N)}]] = \sum_{k=1}^{K} \mathbf{a}^{k,1} \circ \mathbf{a}^{k,2} \circ \cdots \circ \mathbf{a}^{k,N},$$

where $\mathbf{A}^{(n)}$ is the matrix having vectors $\mathbf{a}^{k,n}$ as columns, for $n = 1, \dots, N$ and $k = 1, \dots, K$.

Using this notation, Tensor Decomposition problem can be formulated as,

$$\min_{\{\mathbf{A}^{(n)}\}_n} f(\mathbf{A}^{(1)}, \dots, \mathbf{A}^{(N)}) = \min_{\{\mathbf{A}^{(n)}\}_n} \frac{1}{2} \left|\left|\mathcal{F} - [[\mathbf{A}^{(1)}, \dots, \mathbf{A}^{(N)}]]\right|\right|^2. \tag{1}$$

In this equation, f is expressed as a function of matrices. It can also be written as a scalar-valued function, where the parameters are vectorized and stacked:

$$\mathbf{x} = \begin{bmatrix} \mathbf{a}^{1,1} \\ \vdots \\ \mathbf{a}^{K,1} \\ \vdots \\ \mathbf{a}^{1,N} \\ \vdots \\ \mathbf{a}^{K,N} \end{bmatrix}.$$

Using this approach and knowing the derivatives of the objective function with respect to $\mathbf{x}$, a first-order optimization method can be applied. In the CP-OPT algorithm [1], a nonlinear conjugate gradient method is employed in the optimization and the partial derivatives of the objective function are computed using the following theorem:

Theorem A.1. *The partial derivatives of the objective function f are given by*

$$\frac{\partial f}{\partial \mathbf{a}^{r,n}} = -\left(\mathcal{F} \bigtimes_{\substack{m=1 \\ m \neq n}}^{N} \mathbf{a}^{r,m}\right) + \sum_{\ell=1}^{K} \gamma_{r\ell}^{(n)} \mathbf{a}^{\ell,n}, \tag{2}$$

where $r = 1, \dots, K$ and $n = 1, \dots, N$ with $\gamma_{r\ell}^{(n)}$ defined as

$$\gamma_{r\ell}^{(n)} = \prod_{\substack{m=1 \\ m \neq n}}^{N} (\mathbf{a}^{r,m})^T \mathbf{a}^{\ell,m}$$

and $\bigtimes$ is used to denote the following multiplication:

$$\mathcal{F} \bigtimes_{m=1}^{N} \mathbf{a}^{r,m} = \sum_{i_1=1}^{I_1} \sum_{i_2=1}^{I_2} \cdots \sum_{i_N=1}^{I_N} f_{i_1 i_2 \dots i_N} a^{i_1,1} a^{i_2,2} \dots a^{i_N,N} = \left(\mathcal{F} \bigtimes_{\substack{m=1 \\ m \neq n}}^{N} \mathbf{a}^{r,m}\right)^T \mathbf{a}^{r,n}.$$

Proof: Rewriting the objective function given in Eq. (1) as three summands, the following is obtained:

$$f(\mathbf{x}) = \frac{1}{2}||\mathcal{F}||^2 - \langle \mathcal{F}, [[\mathbf{A}^{(1)}, \dots, \mathbf{A}^{(N)}]] \rangle + \frac{1}{2} \left|\left|[[\mathbf{A}^{(1)}, \dots, \mathbf{A}^{(N)}]]\right|\right|^2,$$

where the first summand is $f_1(\mathbf{x})$, the second summand is $f_2(\mathbf{x})$ and the third summand is $f_3(\mathbf{x})$.

There is no variable in the first summand, thus it does not contribute to the derivative, i.e., $\frac{\partial f_1}{\partial \mathbf{a}^{r,n}} = \mathbf{0}$.

$$\begin{aligned} f_2(\mathbf{x}) &= \langle \mathcal{F}, [[\mathbf{A}^{(1)}, \dots, \mathbf{A}^{(N)}]] \rangle \\ &= \langle \mathcal{F}, \sum_{r=1}^{K} \mathbf{a}^{r,1} \circ \cdots \circ \mathbf{a}^{r,N} \rangle \\ &= \sum_{r=1}^{K} \sum_{i_1=1}^{I_1} \sum_{i_2=1}^{I_2} \cdots \sum_{i_N=1}^{I_N} z_{i_1 i_2 \dots i_N} a_{i_1}^{r,1} a_{i_2}^{r,2} \dots a_{i_N}^{r,N} \\ &= \sum_{r=1}^{K} \left(\mathcal{F} \bigtimes_{m=1}^{N} \mathbf{a}^{r,m}\right) \\ &= \sum_{r=1}^{K} \left(\mathcal{F} \bigtimes_{m=1}^{N} \mathbf{a}^{r,m}\right)^T \mathbf{a}^{r,n}. \end{aligned}$$

As a result the partial derivative of f_2 is found as

$$\frac{\partial f_2}{\partial \mathbf{a}_r^{(n)}} = \left(\mathcal{F} \bigtimes_{\substack{m=1 \\ m \neq n}} \mathbf{a}^{r,m}\right). \tag{3}$$

The third summand is

$$\begin{aligned} f_3(\mathbf{x}) &= \left|\left|[[\mathbf{A}^{(1)}, \dots, \mathbf{A}^{(N)}]]\right|\right|^2 \\ &= \langle \sum_{r=1}^{K} \mathbf{a}^{r,1} \circ \cdots \circ \mathbf{a}^{r,N}, \sum_{r=1}^{R} \mathbf{a}^{r,1} \circ \dots \mathbf{a}^{r,N} \rangle \\ &= \sum_{k=1}^{R} \sum_{\ell=1}^{R} \prod_{m=1}^{N} (\mathbf{a}^{k,m})^T \mathbf{a}^{\ell,m} \\ &= \prod_{m=1}^{N} (\mathbf{a}^{r,m})^T \mathbf{a}^{r,m} + 2 \sum_{\substack{\ell=1 \\ \ell \neq r}}^{R} \prod_{m=1}^{N} (\mathbf{a}^{r,m})^T \mathbf{a}^{\ell,m} \\ &\quad + \sum_{\substack{k=1 \\ k \neq r}}^{R} \sum_{\substack{\ell=1 \\ \ell \neq r}}^{N} (\mathbf{a}^{k,m})^T \mathbf{a}^{\ell,m}. \end{aligned}$$

Thus,

$$\begin{aligned}\frac{\partial f_3}{\partial \mathbf{a}^{r,n}} &= 2\left(\prod_{\substack{m=1\\m\neq n}}^{N}(\mathbf{a}^{r,m})^T\mathbf{a}^{r,m}\right)\mathbf{a}^{r,n}\\ &\quad + 2\sum_{\substack{\ell=1\\\ell\neq r}}^{R}\left(\prod_{\substack{m=1\\m\neq n}}^{N}(\mathbf{a}^{r,m})^T\mathbf{a}^{\ell,m}\right)\mathbf{a}^{\ell,n}\\ &= 2\sum_{\ell=1}^{R}\left(\prod_{\substack{m=1\\m\neq n}}^{N}(\mathbf{a}^{r,m})^T\mathbf{a}^{\ell,m}\right)\mathbf{a}^{\ell,n}. \qquad (4)\end{aligned}$$

Using the partial derivatives of f_2 and f_3, we obtain the results given in Eq. (2). □

References

[1] E. Acar, D. M. Dunlavy, and T. G. Kolda. A Scalable Optimization Approach for Fitting Canonical Tensor Decompositions. *Journal of Chemometrics*, 2011.