

User Prompt

Given a d -way tensor $\mathcal{T} \in \mathbb{R}^{n_1 \times n_2 \times \dots \times n_d}$ such that the data is unaligned (meaning the tensor \mathcal{T} has missing entries), we consider the problem of computing a CP decomposition of rank r where some modes are infinite-dimensional and constrained to be in a Reproducing Kernel Hilbert Space (RKHS). We want to solve this using an alternating optimization approach, and our question is focused on the mode- k subproblem for an infinite-dimensional mode. For the subproblem, then CP factor matrices $A_1, \dots, A_{k-1}, A_{k+1}, \dots, A_d$ are fixed, and we are solving for A_k .

Our notation is as follows. Let $N = \prod_i n_i$ denote the product of all sizes. Let $n \equiv n_k$ be the size of mode k , let $M = \prod_{i \neq k} n_i$ be the product of all dimensions except k , and assume $n \ll M$. Since the data are unaligned, this means only a subset of \mathcal{T} 's entries are observed, and we let $q \ll N$ denote the number of observed entries. We let $T \in \mathbb{R}^{n \times M}$ denote the mode- k unfolding of the tensor \mathcal{T} with all missing entries set to zero. The vec operation creates a vector from a matrix by stacking its columns, and we let $S \in \mathbb{R}^{N \times q}$ denote the selection matrix (a subset of the $N \times N$ identity matrix) such that $S^T \text{vec}(T)$ selects the q known entries of the tensor \mathcal{T} from the vectorization of its mode- k unfolding. We let $Z = A_d \odot \dots \odot A_{k+1} \odot A_{k-1} \odot \dots \odot A_1 \in \mathbb{R}^{M \times r}$ be the Khatri-Rao product of the factor matrices corresponding to all modes except mode k . We let $B = TZ$ denote the MTTKRP of the tensor \mathcal{T} and Khatri-Rao product Z . We assume $A_k = KW$ where $K \in \mathbb{R}^{n \times n}$ denotes the psd RKHS kernel matrix for mode k . The matrix W of size $n \times r$ is the unknown for which we must solve. The system to be solved is

$$[(Z \otimes K)^T S S^T (Z \otimes K) + \lambda(I_r \otimes K)] \text{vec}(W) = (I_r \otimes K) \text{vec}(B). \quad (1)$$

Here, I_r denotes the $r \times r$ identity matrix. This is a system of size $nr \times nr$. Using a standard linear solver costs $O(n^3 r^3)$, and explicitly forming the matrix is an additional expense.

Explain how an iterative preconditioned conjugate gradient linear solver can be used to solve this problem more efficiently. Explain the method and choice of preconditioner. Explain in detail how the matrix-vector products are computed and why this works. Provide complexity analysis. We assume $n, r < q \ll N$. Avoid any computation of order N .

System Short Summary:

We solve the normal equations

$$[(Z \otimes K)^T S S^T (Z \otimes K) + \lambda(I_r \otimes K)] \vec{W} = (I_r \otimes K) \vec{B}$$

by PCG, viewing the left-hand side as an SPD operator on \mathbb{R}^{nr} . The key hypothesis

is $K \succ 0$ (or $K + \varepsilon I \succ 0$), together with $\lambda > 0$, which implies

$$x^\top \left((Z \otimes K)^\top S S^\top (Z \otimes K) + \lambda (I_r \otimes K) \right) x = \|S S^\top (Z \otimes K) x\|_2^2 + \lambda x^\top (I_r \otimes K) x > 0,$$

so standard PCG applies. Using $(A \otimes B) \vec{X} = \vec{B X A^\top}$, one rewrites the MVP as

$$\mathcal{A}(V) = K P_\Omega(K V Z^\top) Z + \lambda K V, \quad (V \in \mathbb{R}^{n \times r}),$$

so $\mathcal{A}(V)$ is computed without forming any nM or N objects: form $\tilde{V} = K V$ ($O(n^2 r)$), then traverse the q observed indices $(i_t, j_t) \in \Omega$ and accumulate $Y_{i_t, +} = (\tilde{V}_{i_t, :} \cdot z_{j_t}) z_{j_t}$, where z_{j_t} is the needed Khatri–Rao row computed on-the-fly from the fixed factors (no storage of Z). Finally output $K Y + \lambda \tilde{V}$. This yields per-iteration cost $O(n^2 r + q r)$ (treating d as constant), avoiding $O(N)$ work.

Preconditioning uses the uniform-sampling heuristic $S S^\top \approx \rho I$ with $\rho = q/(nM)$, giving the spectral Kronecker preconditioner

$$\mathcal{M} = \rho(Z^\top Z \otimes K^2) + \lambda(I_r \otimes K),$$

where $Z^\top Z = *_{m \neq k} (A_m^\top A_m)$ is formed in $O(\sum_{m \neq k} n_m r^2)$ without constructing Z . With eigendecompositions $K = U \Sigma U^\top$ and $Z^\top Z = Q \Gamma Q^\top$, \mathcal{M} diagonalizes under $Q \otimes U$, so \mathcal{M}^{-1} is applied by two small basis changes and entrywise scaling in $O(n^2 r + n r^2)$ time after $O(n^3 + r^3)$ preprocessing. Hence the overall cost is $O(n^3 + r^3 + \sum_{m \neq k} n_m r^2) + O(k_{\text{iter}}(n^2 r + q r))$, with all large-scale structure accessed only through q observed entries.