Comprehensive Solution to Question 10: Efficient RKHS Tensor Completion via Matrix-Free PCG

Dietmar Wolz, Ingo Althöfer

February 10, 2026

Abstract

We present a complete solution for the mode-k subproblem of CP tensor decomposition with missing data and Reproducing Kernel Hilbert Space (RKHS) constraints. We derive the normal equations for the regularized least squares problem and prove that the system matrix is symmetric positive definite (SPD). Crucially, we develop a *matrix-free* matrix-vector product (MVP) that operates in $O(qnr)$ time—scaling only with the number of observed entries q, not the total tensor size N. We further introduce a Kronecker-structured preconditioner that is efficiently invertible in $O(n^2 r)$ time, ensuring rapid convergence of the Preconditioned Conjugate Gradient (PCG) solver.

1 Problem Statement and Notation

Let $\boldsymbol{\mathcal{T}} \in \mathbb{R}^{n_1 \times \cdots \times n_d}$ be a d-way tensor with observed entries index by Ω. We focus on the update for the k-th mode factor matrix $\mathbf{A}_k \in \mathbb{R}^{n_k \times r}$, which is constrained to lie in an RKHS defined by a kernel $\mathbf{K} \in \mathbb{R}^{n_k \times n_k}$.

1.1 Notation

- $n := n_k$: Dimension of the active mode.
- $M := \prod_{j \neq k} n_j$: Product of dimensions of all other modes.
- $N := nM$: Total number of entries in the tensor.
- $q := |\Omega|$: Number of observed entries. We assume $n, r < q \ll N$.
- $\bar{\mathbf{T}} \in \mathbb{R}^{n \times M}$: The mode-$k$ unfolding of $\boldsymbol{\mathcal{T}}$ with unobserved entries set to 0.
- $\mathcal{P}_\Omega$: The projection operator onto the observation set Ω. For a matrix $\mathbf{X}$, $(\mathcal{P}_\Omega(\mathbf{X}))_{i,m} = X_{i,m}$ if $(i, m) \in \Omega$, and 0 otherwise.
- $\mathbf{Z} \in \mathbb{R}^{M \times r}$: The Khatri-Rao product of all other factor matrices:

$$\mathbf{Z} = \mathbf{A}_d \odot \ldots \odot \mathbf{A}_{k+1} \odot \mathbf{A}_{k-1} \odot \ldots \odot \mathbf{A}_1.$$

- $\mathbf{W} \in \mathbb{R}^{n \times r}$: The unknown weight matrix such that $\mathbf{A}_k = \mathbf{K}\mathbf{W}$.

1.2 The Optimization Problem

We minimize the squared error over observed entries plus an RKHS norm penalty:

$$\min_{\mathbf{W}\in\mathbb{R}^{n\times r}} \frac{1}{2}\left\|\mathcal{P}_\Omega(\mathbf{K}\mathbf{W}\mathbf{Z}^\top - \bar{\mathbf{T}})\right\|_F^2 + \frac{\lambda}{2}\operatorname{tr}(\mathbf{W}^\top\mathbf{K}\mathbf{W}). \tag{1}$$

Setting the gradient with respect to $\mathbf{W}$ to zero yields the normal equations. Using the identity $\operatorname{vec}(\mathbf{ABC}) = (\mathbf{C}^\top \otimes \mathbf{A})\operatorname{vec}(\mathbf{B})$, the system for $\mathbf{w} = \operatorname{vec}(\mathbf{W})$ is:

$$\underbrace{\left[(\mathbf{Z}\otimes\mathbf{K})^\top\mathcal{P}_\Omega(\mathbf{Z}\otimes\mathbf{K}) + \lambda(\mathbf{I}_r\otimes\mathbf{K})\right]}_{\mathbf{A}_{sys}}\mathbf{w} = (\mathbf{I}_r\otimes\mathbf{K})\operatorname{vec}(\bar{\mathbf{T}}\mathbf{Z}). \tag{2}$$

Lemma 1 (SPD Property)**.** *If $\mathbf{K}\succ 0$ and $\lambda > 0$, the system matrix $\mathbf{A}_{sys}$ is symmetric positive definite.*

Proof. For any $\mathbf{v}\neq 0$, $\mathbf{v}^\top\mathbf{A}_{sys}\mathbf{v} = \|\mathcal{P}_\Omega(\mathbf{K}\mathbf{V}\mathbf{Z}^\top)\|_F^2 + \lambda\operatorname{tr}(\mathbf{V}^\top\mathbf{K}\mathbf{V})$. Since $\mathbf{K}\succ 0$, the second term is strictly positive. □

2 The Matrix-Free MVP

Explicitly forming $\mathbf{A}_{sys}$ would involve the $N\times N$ operator $\mathcal{P}_\Omega$, costing $O(N)$ or more. We derive a matrix-free approach that scales only with q.

Theorem 1 (Efficient MVP)**.** *The product $\mathbf{A}_{sys}\mathbf{w}$ can be computed in $O(qnr+n^2r)$ time without forming any $O(N)$ object.*

Proof. Let $\mathbf{w} = \operatorname{vec}(\mathbf{W})$. The regularization term $\lambda(\mathbf{I}_r\otimes\mathbf{K})\mathbf{w} = \operatorname{vec}(\lambda\mathbf{K}\mathbf{W})$ costs $O(n^2r)$. For the data term, let $\mathcal{X} = \mathbf{K}\mathbf{W}\mathbf{Z}^\top$. We proceed in two steps:

1. Forward Pass (Gather): We need $\mathbf{u} = \mathcal{P}_\Omega(\mathbf{Z}\otimes\mathbf{K})\mathbf{w}$. This vector contains the values of $\mathcal{X}$ at observed indices. For each $\ell\in\{1,\ldots,q\}$ with indices (i_ℓ, m_ℓ):

$$u_\ell = (\mathbf{K}\mathbf{W}\mathbf{Z}^\top)_{i_\ell,m_\ell} = \mathbf{K}(i_\ell,:)\cdot\mathbf{W}\cdot\mathbf{Z}(m_\ell,:)^\top.$$

Implementation:

- Precompute $\mathbf{H} = \mathbf{K}_\Omega\mathbf{W}\in\mathbb{R}^{q\times r}$, where $\mathbf{K}_\Omega$ contains rows of $\mathbf{K}$ for observed indices. Cost: $O(qnr)$.
- Compute dot products: $u_\ell = \sum_{j=1}^r H_{\ell,j}Z_{m_\ell,j}$. Cost: $O(qr)$.

2. Backward Pass (Scatter): We need $\mathbf{Y} = (\mathbf{Z}\otimes\mathbf{K})^\top\mathbf{u}$. This is the adjoint operation, mapping the sparse observation vector $\mathbf{u}$ back to the parameter space:

$$\mathbf{Y} = \sum_{\ell=1}^q u_\ell\left(\mathbf{K}(i_\ell,:)^\top\otimes\mathbf{Z}(m_\ell,:)^\top\right) = \mathbf{K}_\Omega^\top\left(\operatorname{diag}(\mathbf{u})\mathbf{Z}_\Omega\right).$$

Implementation:

- Scale rows of $\mathbf{Z}_\Omega$ by $\mathbf{u}$: $O(qr)$.
- Multiply by $\mathbf{K}_\Omega^\top$: $O(qnr)$.

Total complexity is dominated by the two $O(qnr)$ matrix multiplications. □

3 Kronecker-Structured Preconditioner

To accelerate PCG, we approximate $\mathcal{P}_\Omega \approx \alpha\mathbf{I}$ (assuming uniform sampling), leading to the preconditioner:

$$\mathbf{P} = (\mathbf{Z} \otimes \mathbf{K})^\top (\mathbf{Z} \otimes \mathbf{K}) + \lambda(\mathbf{I}_r \otimes \mathbf{K}) = (\mathbf{Z}^\top \mathbf{Z}) \otimes \mathbf{K}^2 + \lambda(\mathbf{I}_r \otimes \mathbf{K}). \tag{3}$$

Theorem 2 (Fast Inversion)**.** $\mathbf{P}^{-1}$ *can be applied in* $O(n^2 r + nr^2)$ *time.*

Proof. Let $\mathbf{K} = \mathbf{V}_K \mathbf{\Lambda} \mathbf{V}_K^\top$ and $\mathbf{Z}^\top\mathbf{Z} = \mathbf{V}_Z \mathbf{\Sigma} \mathbf{V}_Z^\top$. Substituting these into (3) and using the mixed-product property, $\mathbf{P}$ diagonalizes in the basis $\mathbf{V}_Z \otimes \mathbf{V}_K$:

$$\mathbf{P} = (\mathbf{V}_Z \otimes \mathbf{V}_K) \left[\mathbf{\Sigma} \otimes \mathbf{\Lambda}^2 + \lambda(\mathbf{I} \otimes \mathbf{\Lambda})\right] (\mathbf{V}_Z \otimes \mathbf{V}_K)^\top.$$

The eigenvalues are diagonal entries $\gamma_{ij} = \sigma_j \mu_i^2 + \lambda\mu_i$. Inversion requires: 1. Transforming the residual to the eigenbasis: $\tilde{\mathbf{R}} = \mathbf{V}_K^\top \mathbf{R} \mathbf{V}_Z$. 2. Scaling by $1/\gamma_{ij}$. 3. Transforming back: $\mathbf{W}_{new} = \mathbf{V}_K \tilde{\mathbf{R}} \mathbf{V}_Z^\top$. All steps involve dense matrix multiplications of size $n \times r$ or $n \times n$. □

4 Algorithm and Complexity

Algorithm 1 Matrix-Free PCG for Mode-k RKHS Update

1: **Input:** $\mathbf{K}$, $\mathbf{Z}$, Ω, values $\mathbf{t}_\Omega$, λ.
2: **Setup:** $\mathbf{V}_K, \mu \leftarrow \text{eig}(\mathbf{K})$; $\mathbf{V}_Z, \sigma \leftarrow \text{eig}(\mathbf{Z}^\top\mathbf{Z})$.
3: **RHS:** $\mathbf{B} \leftarrow \text{MTTKRP}(\Omega, \mathbf{t}_\Omega, \mathbf{Z})$ (sparse); $\mathbf{b} \leftarrow \text{vec}(\mathbf{KB})$.
4: **Loop:** Run PCG.
5:    **MVP:** $\mathbf{v} \mapsto \mathbf{A}_{sys}\mathbf{v}$ using Forward/Backward passes (Sec 3).
6:    **Precond:** $\mathbf{r} \mapsto \mathbf{P}^{-1}\mathbf{r}$ using spectral scaling (Sec 4).
7: **Output:** $\mathbf{W}$.

Table 1: Complexity Analysis (Dominant Terms)

Operation	Complexity	Frequency
Eigendecomposition of $\mathbf{K}$	$O(n^3)$	Once
Formation of $\mathbf{Z}^\top\mathbf{Z}$	$O(Mr^2)$	Once
Matrix-Vector Product (MVP)	$\mathbf{O(qnr)}$	Per Iteration
Preconditioner Apply	$O(n^2 r + nr^2)$	Per Iteration

5 Conclusion

We have derived a solver that strictly avoids $O(N)$ computations. By exploiting the sparse structure of observations in the MVP and the dense Kronecker structure in the preconditioner, we achieve a highly efficient iteration cost of $O(qnr)$, making this approach suitable for large-scale tensor completion.