

The prompts and responses below have been edited only for formatting compatibility (to compile in a L<sup>A</sup>T<sub>E</sub>X document).

### Fullproof # 10

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

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

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  $N$ .

### Model Response (Agent A)

#### 1. Algebraic Factorization of the Linear System

The alternating optimization subproblem for the infinite-dimensional factor matrix  $W \in \mathbb{R}^{n \times r}$  associated with a Reproducing Kernel Hilbert Space (RKHS) governed by the symmetric positive semi-definite (PSD) kernel matrix  $K \in \mathbb{R}^{n \times n}$  requires solving



the normal equations

$$\mathcal{H} \text{vec}(W) = \mathbf{c}, \quad (1)$$

where the global system matrix  $\mathcal{H} \in \mathbb{R}^{nr \times nr}$  and the right-hand side objective vector  $\mathbf{c} \in \mathbb{R}^{nr}$  are defined as

$$\mathcal{H} := (Z \otimes K)^T \Omega (Z \otimes K) + \lambda (I_r \otimes K), \quad (2)$$

$$\mathbf{c} := (I_r \otimes K) \text{vec}(B). \quad (3)$$

Here,  $Z \in \mathbb{R}^{M \times r}$  denotes the Khatri-Rao product of the fixed factor matrices from the remaining  $d - 1$  modes. To represent the unaligned missing data topology, we define the diagonal masking matrix  $\Omega := SS^T \in \mathbb{R}^{N \times N}$ , where  $S \in \mathbb{R}^{N \times q}$  is the selection matrix projecting onto the  $q$  observed tensor entries. Explicitly forming the global operator  $\mathcal{H}$  or executing computations within the ambient tensor dimensions  $M$  and  $N$  is computationally intractable.

Because  $K$  is a valid PSD reproducing kernel, the linear operator  $\mathcal{H}$  is inherently symmetric and PSD. For any Tikhonov regularization parameter  $\lambda > 0$ ,  $\mathcal{H}$  is strictly positive definite if and only if  $K$  is strictly positive definite ( $K \succ 0$ ). In the broader regime where the kernel  $K$  may be rank-deficient, the global operator  $\mathcal{H}$  inherently shares a null space with the operator  $(I_r \otimes K)$ . Specifically, for any test vector  $\mathbf{x} \in \mathbb{R}^{nr}$ , the condition  $\mathbf{x}^T \mathcal{H} \mathbf{x} = 0$  intrinsically requires both  $(Z \otimes I_n)(I_r \otimes K)\mathbf{x} = \mathbf{0}$  and  $(I_r \otimes K)\mathbf{x} = \mathbf{0}$  due to the strict positive semi-definiteness of the individual summands. Thus,  $\text{Null}(\mathcal{H}) = \text{Null}(I_r \otimes K)$ . Because both operators are symmetric, their column spaces are identical:  $\text{Col}(\mathcal{H}) = \text{Col}(I_r \otimes K)$ . Since the objective vector  $\mathbf{c} = (I_r \otimes K) \text{vec}(B)$  resides natively in the column space of  $(I_r \otimes K)$ , it necessarily lies precisely in the column space of  $\mathcal{H}$ . This intrinsic algebraic consistency rigorously guarantees that the Preconditioned Conjugate Gradient (PCG) method will safely converge to a valid minimizer without requiring strict positive definiteness of the global system.

To construct an optimal matrix-vector product (MVP) for the PCG solver, we invoke the standard Kronecker mixed-product property,  $(A_1 \otimes B_1)(A_2 \otimes B_2) = (A_1 A_2 \otimes B_1 B_2)$ , to decouple the spatial kernel  $K$  from the discrete fixed Khatri-Rao product  $Z$ :

$$(Z \otimes K) = (Z \otimes I_n)(I_r \otimes K). \quad (4)$$

Leveraging the symmetry of the kernel,  $(I_r \otimes K)^T = I_r \otimes K$ . Substituting this exact factorization into the definition of  $\mathcal{H}$  cleanly isolates the kernel operators:

$$\mathcal{H} = (I_r \otimes K) \Phi (I_r \otimes K) + \lambda (I_r \otimes K), \quad (5)$$

where  $\Phi := (Z \otimes I_n)^T \Omega (Z \otimes I_n) \in \mathbb{R}^{nr \times nr}$  acts as the core structural matrix encapsulating the unaligned data topology.

## 2. Spatial Decoupling and Matrix-Free Matrix-Vector Product

Standard sparse MVP algorithms strictly demand evaluating the missing data mask  $\Omega$  at every iteration, instantiating a prohibitive  $\mathcal{O}(qr)$  computational bottleneck. We



systematically bypass this limitation by proving that  $\Phi$  operates completely independently across the  $n$  spatial coordinates.

**Theorem 1 (Spatial Block-Diagonal Structure of  $\Phi$ ).** *Let  $\mathcal{E}$  denote the set of  $q$  observed tensor indices. For each observation  $e \in \mathcal{E}$ , let  $i_e \in \{1, \dots, n\}$  be its mode- $k$  spatial index, and let  $z_e \in \mathbb{R}^{1 \times r}$  denote the corresponding row extracted from the Khatri-Rao product  $Z$ . Let  $\Psi_i \in \mathbb{R}^{r \times r}$  denote the local uncentered rank-covariance matrix for spatial slice  $i$ :*

$$\Psi_i = \sum_{e \in \mathcal{E}: i_e = i} z_e^T z_e \quad \text{for } i = 1, \dots, n. \quad (6)$$

Then applying the core structural operator  $\Phi$  to the vectorization of any matrix  $U \in \mathbb{R}^{n \times r}$  strictly decouples row-by-row in the unvectorized spatial domain. Specifically, if we unambiguously define the intermediate mapped matrix  $V = \text{unvec}(\Phi \text{vec}(U)) \in \mathbb{R}^{n \times r}$ , its  $i$ -th row precisely evaluates to  $V_{i,:} = U_{i,:} \Psi_i$ .

*Proof.* We formally expand the diagonal unaligned masking matrix as  $\Omega = \sum_{e \in \mathcal{E}} \mathbf{e}_{j_e} \mathbf{e}_{j_e}^T \otimes \mathbf{e}_{i_e} \mathbf{e}_{i_e}^T$ , where  $j_e$  represents the complementary geometric multi-index in the  $M$ -dimensional ambient space. Evaluating the inner projection for  $\Phi$  yields:

$$\begin{aligned} \Phi &= \sum_{e \in \mathcal{E}} (Z \otimes I_n)^T (\mathbf{e}_{j_e} \mathbf{e}_{j_e}^T \otimes \mathbf{e}_{i_e} \mathbf{e}_{i_e}^T) (Z \otimes I_n) \\ &= \sum_{e \in \mathcal{E}} (Z^T \mathbf{e}_{j_e} \mathbf{e}_{j_e}^T Z) \otimes (I_n \mathbf{e}_{i_e} \mathbf{e}_{i_e}^T I_n) \\ &= \sum_{e \in \mathcal{E}} (z_e^T z_e) \otimes (\mathbf{e}_{i_e} \mathbf{e}_{i_e}^T) = \sum_{a=1}^n \Psi_a \otimes (\mathbf{e}_a \mathbf{e}_a^T). \end{aligned} \quad (7)$$

Applying this decoupled operator directly to  $\text{vec}(U)$ , we utilize the standard vectorization identity  $(B^T \otimes A) \text{vec}(X) = \text{vec}(AXB)$ . Noting that the localized empirical covariance matrix is strictly symmetric ( $\Psi_a^T = \Psi_a$ ), we obtain:

$$\Phi \text{vec}(U) = \sum_{a=1}^n \text{vec}(\mathbf{e}_a \mathbf{e}_a^T U \Psi_a). \quad (8)$$

Unvectorizing this summation naturally maps the operator back to the dense matrix domain as  $V = \sum_{a=1}^n \mathbf{e}_a \mathbf{e}_a^T U \Psi_a$ . Extracting the targeted  $i$ -th row of  $V$  mathematically establishes  $V_{i,:} = \mathbf{e}_i^T V = \sum_{a=1}^n \mathbf{e}_i^T \mathbf{e}_a \mathbf{e}_a^T U \Psi_a$ . Because the canonical basis vectors are orthonormal ( $\mathbf{e}_i^T \mathbf{e}_a = \delta_{ia}$ ), the spatial summation collapses perfectly to  $V_{i,:} = \mathbf{e}_i^T U \Psi_i = U_{i,:} \Psi_i$ , concluding the proof.  $\blacksquare$

**Accelerated Matrix-Free MVP Algorithm.** By precomputing the  $n$  compact spatial covariance matrices  $\Psi_i$ , evaluating the complete action of the system  $Y = \text{unvec}(\mathcal{H} \text{vec}(W))$  strictly bypasses the explicit necessity to traverse the observation mask  $\Omega$  inside the iterative solver:

1. Compute  $U = KW \in \mathbb{R}^{n \times r}$ . (Cost:  $\mathcal{O}(n^2 r)$ )
2. Evaluate  $V_{i,:} = U_{i,:} \Psi_i$  for  $i = 1, \dots, n$ . (Cost:  $\mathcal{O}(nr^2)$ )



3. Compute  $\tilde{U} = KV \in \mathbb{R}^{n \times r}$ . (Cost:  $\mathcal{O}(n^2r)$ )

4. Formulate  $Y = \tilde{U} + \lambda U$ . (Cost:  $\mathcal{O}(nr)$ )

This analytical factorization cleanly bounds the strict per-iteration MVP computational overhead to merely  $\mathcal{O}(n^2r + nr^2)$ .

### 3. Exact Spatial Block-Jacobi Preconditioner

CP tensor decompositions naturally suffer from severe structural collinearity among the latent rank components  $r$ , and the unaligned missing data topology  $\Omega$  invariably leaves specific spatial slices sparsely observed. To robustly resolve this severe geometric ill-conditioning, we formally derive the exact spatial Block-Jacobi preconditioner  $\mathcal{M}$ , intrinsically extracted from the  $n$  strictly diagonal  $r \times r$  blocks of  $\mathcal{H}$ .

**Theorem 2 (Diagonal Blocks of  $\mathcal{H}$ ).** *Let the continuous variables of  $\text{vec}(W)$  be sequentially partitioned by their spatial rows. The  $i$ -th diagonal block  $B_i \in \mathbb{R}^{r \times r}$  of  $\mathcal{H}$ , uniquely operating over the localized spatial variables  $W_{i,:}$ , is exactly given by:*

$$B_i = \sum_{a=1}^n K_{i,a}^2 \Psi_a + \lambda K_{i,i} I_r. \quad (9)$$

*Proof.* The continuous decision variables encapsulating the  $i$ -th row of  $W$  are selected via the orthogonal identity projection  $(I_r \otimes \mathbf{e}_i^T) \text{vec}(W) = W_{i,:}^T$ . Thus, the exact  $r \times r$  principal submatrix of  $\mathcal{H}$  fundamentally mapping row  $i$  back to itself is isolated strictly by the submatrix projection  $B_i = (I_r \otimes \mathbf{e}_i^T) \mathcal{H} (I_r \otimes \mathbf{e}_i)$ . Expanding  $\mathcal{H}$  using the core structural identity established in Theorem 1:

$$\begin{aligned} \mathcal{H} &= (I_r \otimes K) \left( \sum_{a=1}^n \Psi_a \otimes \mathbf{e}_a \mathbf{e}_a^T \right) (I_r \otimes K) + \lambda (I_r \otimes K) \\ &= \sum_{a=1}^n \Psi_a \otimes (K \mathbf{e}_a \mathbf{e}_a^T K) + \lambda (I_r \otimes K). \end{aligned} \quad (10)$$

Projecting this expanded summation sequentially with  $(I_r \otimes \mathbf{e}_i^T)$  on the left and  $(I_r \otimes \mathbf{e}_i)$  on the right yields:

$$B_i = \sum_{a=1}^n \Psi_a \otimes (\mathbf{e}_i^T K \mathbf{e}_a \mathbf{e}_a^T K \mathbf{e}_i) + \lambda (I_r \otimes \mathbf{e}_i^T K \mathbf{e}_i). \quad (11)$$

By functional kernel symmetry,  $\mathbf{e}_i^T K \mathbf{e}_a = K_{i,a}$  and  $\mathbf{e}_a^T K \mathbf{e}_i = K_{a,i} = K_{i,a}$ . Thus, the central quadratic form perfectly simplifies to  $K_{i,a}^2$ . Incorporating the local diagonal evaluation  $\mathbf{e}_i^T K \mathbf{e}_i = K_{i,i}$ , we obtain:

$$B_i = \sum_{a=1}^n \Psi_a \otimes (K_{i,a}^2) + \lambda (I_r \otimes K_{i,i}) = \sum_{a=1}^n K_{i,a}^2 \Psi_a + \lambda K_{i,i} I_r, \quad (12)$$



which analytically isolates the target dense preconditioner sub-block.  $\blacksquare$

Because the empirical covariance inherently satisfies  $\Psi_a \succeq 0$ , the squared kernel interactions naturally enforce  $K_{i,a}^2 \geq 0$ , and valid spatial coordinate maps within a non-trivial RKHS evaluate to  $K_{i,i} > 0$ . Therefore, every local block  $B_i$  safely absorbs the strictly positive scalar shift  $\lambda K_{i,i} I_r$ . This mathematically guarantees that for any regularization parameter  $\lambda > 0$ , every block  $B_i$  strictly maps as symmetric positive definite ( $B_i \succ 0$ ), rigorously safeguarding the preconditioning sub-blocks against local rank degeneracy. We can independently precompute their  $n$  spatial Cholesky factorizations  $L_i L_i^T = B_i$ . During the active PCG descent phase, preconditioning an arbitrary residual matrix  $R \in \mathbb{R}^{n \times r}$  strictly reduces to solving  $n$  decoupled dense triangular systems,  $B_i \tilde{R}_{i,:}^T = R_{i,:}^T$ , which costs purely  $\mathcal{O}(nr^2)$  computational operations.

## 4. Direct Right-Hand Side Construction

The exact evaluation of the initial right-hand side vector  $\mathbf{c} = \text{vec}(KB)$  depends purely on the sparse Matricized Tensor Times Khatri-Rao Product (MTTKRP) array  $B = TZ$ . Because the unaligned tensor unfolding  $T \in \mathbb{R}^{n \times M}$  operates structurally as a complete zero-filled tensor aside from the natively defined  $q$  observed continuous entries  $\mathcal{T}_e$ , we mathematically formulate the condensed objective matrix  $B \in \mathbb{R}^{n \times r}$  directly without explicitly instantiating the massive ambient unfolding space:

$$B_{i,:} = \sum_{e \in \mathcal{E}: i_e = i} \mathcal{T}_e z_e. \quad (13)$$

Aggregating  $B$  implicitly requires strictly  $\mathcal{O}(qr)$  dense operations. A final dense matrix multiplication  $KB$  generates the fully shifted unvectorized objective vector  $\mathbf{c}$  efficiently in  $\mathcal{O}(n^2 r)$  sequential time.

## 5. Computational Complexity Analysis

The unified algorithmic topography seamlessly circumvents any explicit calculations scaling with the native global geometric volume  $\mathcal{O}(N)$  and strictly isolates all heavy  $\mathcal{O}(q)$  calculations directly to an isolated, one-time precomputational phase.

### I. One-Time Setup Precomputations:

1. Extract the Khatri-Rao rows  $z_e$ :  $\mathcal{O}(qdr)$  via sequential Hadamard products across the fixed unrolled factor matrices.
2. Accumulate the spatial uncentered covariances  $\Psi_i$  and evaluate the unaligned MTTKRP  $B$ :  $\mathcal{O}(qr^2 + qr)$ .
3. Formulate the explicit dense initial objective matrix condition  $KB$ :  $\mathcal{O}(n^2 r)$ .
4. Assemble the exact block-Jacobi preconditioner geometries  $B_i$  ( $\mathcal{O}(n^2 r^2)$ ) and solve their spatial Cholesky factors ( $\mathcal{O}(nr^3)$ ):  $\mathcal{O}(n^2 r^2 + nr^3)$ .



*Total Setup Complexity:*  $\mathcal{O}(qdr + qr^2 + n^2r^2 + nr^3)$ .

## **II. Per-Iteration PCG Phase:**

1. Accelerated matrix-free iteration MVP evaluation via  $\text{unvec}(\mathcal{H} \text{vec}(W))$ :  $\mathcal{O}(n^2r + nr^2)$ .
2. Fast mathematical application of the decoupled spatial block-Jacobi preconditioner:  $\mathcal{O}(nr^2)$ .

*Total Iteration Complexity:*  $\mathcal{O}(n^2r + nr^2)$ .

The overarching spatial memory footprint rigidly restricts scaling to  $\mathcal{O}(qr + n^2 + nr^2)$ , definitively respecting the unaligned data constraints dictating  $n, r < q \ll N$ .