Preconditioned Conjugate Gradient for RKHS-Constrained CP Tensor Decomposition

First Proof - Problem 10

AI Mathematician — AI Reviewer
Prompted by
Scott Armstrong, Julia Kempe and Remi Munos

February 10, 2026

Introduction

First Proof [arXiv:2602.05192] is a small, deliberately "contamination-resistant" collection of ten research-level mathematics questions released by the authors as a public stress test of modern AI systems on genuine, non-toy proof tasks. Each question arose naturally in the authors' research, and the corresponding human solutions were initially withheld (via encrypted commitments) to encourage independent attempts.

This document presents a solution write-up for **Problem 10** from [arXiv:2602.05192]. **The write-up was generated entirely by AI.** Problem 10 asks for an efficient way to solve a structured linear system arising in a kernel-parameterized mode-k update for tensor completion with missing data. We give a preconditioned conjugate gradient (PCG) approach, including explicit Kronecker/vec identities, matrix–vector product routines that avoid any $O(N)$-scale objects, a Kronecker preconditioner, and a detailed complexity analysis.

Contents

1 Goal, Scope, and Conventions

1.1 Goal

We consider the structured linear system

$$H\,\mathrm{vec}(W) = b, \tag{1}$$

where

$$H := (Z \otimes K)^\top S S^\top (Z \otimes K) + \lambda\,(I_r \otimes K), \qquad b := (I_r \otimes K)\,\mathrm{vec}(B),$$

and the unknowns, data, and parameters are defined precisely in Section 2. The system arises from a kernel-parameterized mode-k update in tensor completion. Our objectives in this document are:

(i) restate the problem with explicit dimensions (Section 2);

(ii) collect the linear-algebraic identities needed for subsequent analysis (Section 3);

(iii) prove that H is symmetric positive definite under stated assumptions, so that CG/PCG is applicable (Section 4).

The fast matrix–vector product, preconditioner design, and PCG algorithm will be developed in later sections.

1.2 Conventions

Vectorization. For a matrix $X \in \mathbb{R}^{n\times r}$ with columns $x_1, \dots, x_r$, the vectorization operator stacks columns in order:

$$\operatorname{vec}(X) := \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_r \end{pmatrix} \in \mathbb{R}^{nr}.$$

Throughout, $\operatorname{vec}\colon \mathbb{R}^{n\times r} \to \mathbb{R}^{nr}$ denotes this column-major vectorization.

Real matrices. All matrices and vectors in this document are real.

Symmetry and definiteness notation. A symmetric matrix $K \in \mathbb{R}^{n\times n}$ is called *positive semidefinite* (PSD), written $K \succeq 0$, if $x^\top K x \geq 0$ for every $x \in \mathbb{R}^n$. It is called *positive definite* (SPD), written $K \succ 0$, if $x^\top K x > 0$ for every nonzero $x \in \mathbb{R}^n$.

Notation reserved for later use. The letter P is reserved exclusively for the preconditioner (introduced in a later section). We write $U := KW_v$ for the intermediate product appearing in the matrix–vector computation (where W_v is the reshaped input vector; see Section 5) and never denote it by P.

2 Problem Restatement

2.1 Dimensions

Let $\mathcal{T} \in \mathbb{R}^{n_1\times\cdots\times n_d}$ be a dth-order tensor. Write

$$N := \prod_{i=1}^{d} n_i.$$

Fix a mode index $k \in \{1, \dots, d\}$ and set

$$n := n_k, \qquad M := \frac{N}{n} = \prod_{\substack{i=1 \\ i\neq k}}^{d} n_i.$$

Let $r \geq 1$ denote the target rank, and let q denote the number of observed entries.

2.2 Mode-k unfolding

The mode-k matricization of $\mathcal{T}$ is the matrix $T \in \mathbb{R}^{n\times M}$ whose (i,j)-entry is the entry of $\mathcal{T}$ with mode-k index equal to i and the remaining indices encoded in column index j (via the standard lexicographic mapping of the complementary modes). Unobserved entries of T are set to zero.

2.3 Observed index set and selection matrix

The observed entries form a set of q index tuples

$$\Omega = \left\{ (i_1(\ell), \dots, i_d(\ell)) \right\}_{\ell=1}^{q}.$$

For each observation ℓ, the mode-k unfolding maps the tuple to a row–column pair $(i_k(\ell),\, m(\ell))$, where $m(\ell) \in \{1, \dots, M\}$ is the column index determined by the remaining $d-1$ mode indices of observation ℓ.

Assumption 2.1 (No duplicate observations)**.** The pairs $\{(i_k(\ell), m(\ell))\}_{\ell=1}^{q}$ are pairwise distinct.

Under Assumption 2.1, each observed entry corresponds to a unique position in the vectorization $\mathrm{vec}(T) \in \mathbb{R}^N$. Denote this position by $p(\ell) \in \{1, \dots, N\}$, and write

$$\mathcal{P} := \{p(1), \dots, p(q)\} \ \subset \ \{1, \dots, N\} \tag{2}$$

for the set of observed vectorization indices. Define the *selection matrix*

$$S := \left[\, e_{p(1)} \ \ e_{p(2)} \ \ \cdots \ \ e_{p(q)} \,\right] \in \mathbb{R}^{N\times q}, \tag{3}$$

where e_j is the jth standard basis vector of $\mathbb{R}^N$. Since the positions $p(1), \dots, p(q)$ are distinct, S consists of q distinct columns of the $N \times N$ identity. The properties of S are collected in Lemma 3.5 below.

2.4 Factor matrices and Khatri–Rao product

For each mode $i \neq k$, let $A_i \in \mathbb{R}^{n_i \times r}$ be a given factor matrix. Define the *Khatri–Rao product*

$$Z := A_d \odot \cdots \odot A_{k+1} \odot A_{k-1} \odot \cdots \odot A_1 \ \in \ \mathbb{R}^{M\times r}, \tag{4}$$

where $\odot$ denotes the column-wise Kronecker (Khatri–Rao) product: for matrices $C \in \mathbb{R}^{p\times r}$ and $D \in \mathbb{R}^{s\times r}$, $(C \odot D) \in \mathbb{R}^{ps\times r}$ has jth column equal to $c_j \otimes d_j$.

We adopt the convention that the column index $m \in \{1, \dots, M\}$ in the mode-k unfolding $T \in \mathbb{R}^{n\times M}$ (Section 2.2) corresponds bijectively to the complementary multi-index $(i_1, \dots, i_{k-1}, i_{k+1}, \dots, i_d)$ under the same lexicographic ordering used to form the Khatri–Rao product in (4). Under this convention, row m of Z corresponds to that complementary multi-index.

2.5 MTTKRP

Define the *matricized tensor times Khatri–Rao product*

$$B := TZ \ \in \ \mathbb{R}^{n\times r}. \tag{5}$$

2.6 Kernel matrix and unknown

Let $K \in \mathbb{R}^{n\times n}$ be a symmetric kernel Gram matrix ($K = K^\top$). The mode-k factor is parameterized as $A_k = KW$, where $W \in \mathbb{R}^{n\times r}$ is the unknown.

2.7 The linear system

With regularization parameter $\lambda > 0$, the system to be solved is

$$H\ \mathrm{vec}(W) = b, \tag{6}$$

where

$$H := (Z \otimes K)^\top S\, S^\top (Z \otimes K) + \lambda\,(I_r \otimes K)\ \in\ \mathbb{R}^{nr \times nr}, \tag{7}$$

$$b := (I_r \otimes K)\ \mathrm{vec}(B)\ \in\ \mathbb{R}^{nr}. \tag{8}$$

Dimension verification. The matrix $Z \otimes K$ has size $(M \cdot n) \times (r \cdot n) = N \times nr$. Hence $S^\top (Z \otimes K) \in \mathbb{R}^{q \times nr}$, and $(Z \otimes K)^\top S \in \mathbb{R}^{nr \times q}$. Therefore $(Z \otimes K)^\top S\, S^\top (Z \otimes K) \in \mathbb{R}^{nr \times nr}$. Since $I_r \otimes K \in \mathbb{R}^{nr \times nr}$, the matrix H is $nr \times nr$. Likewise, $\mathrm{vec}(B) \in \mathbb{R}^{nr}$, so $b = (I_r \otimes K)\,\mathrm{vec}(B) \in \mathbb{R}^{nr}$.

2.8 Standing assumptions

The problem statement specifies $K \succeq 0$ (PSD). However, CG and PCG require the coefficient matrix to be SPD. Accordingly, our main results are proved under the following two assumptions.

Assumption 2.2 (SPD kernel)**.** $K \succ 0$ (symmetric positive definite).

Assumption 2.3 (Positive regularization)**.** $\lambda > 0$.

The case of singular K is addressed separately in Remark 4.5.

3 Background Lemmas

Lemma 3.1 (Kronecker–vec identity)**.** *Let $A \in \mathbb{R}^{m \times n}$, $B \in \mathbb{R}^{p \times s}$, and $X \in \mathbb{R}^{n \times s}$. Then*

$$(B \otimes A)\ \mathrm{vec}(X) = \mathrm{vec}(AXB^\top).$$

Proof. Write $X = [x_1, \dots, x_s]$ with $x_j \in \mathbb{R}^n$. Set $y := (B \otimes A)\,\mathrm{vec}(X) \in \mathbb{R}^{pm}$ and view y as p blocks $y_\ell \in \mathbb{R}^m$, $\ell = 1, \dots, p$. By the block structure of the Kronecker product, the (ℓ, j)-block of $B \otimes A$ is $B_{\ell j} A \in \mathbb{R}^{m \times n}$, so

$$y_\ell = \sum_{j=1}^{s} B_{\ell j}\, A x_j = A\Big(\sum_{j=1}^{s} B_{\ell j}\, x_j\Big).$$

Now observe that $\sum_{j=1}^{s} B_{\ell j}\, x_j$ is the ℓth column of $XB^\top \in \mathbb{R}^{n \times p}$ (since $(XB^\top)_{:,\ell} = \sum_j B_{\ell j}\, x_j$). Therefore y_ℓ equals the ℓth column of $AXB^\top \in \mathbb{R}^{m \times p}$, and stacking all p blocks gives $y = \mathrm{vec}(AXB^\top)$. $\square$

Lemma 3.2 (Transpose of Kronecker product)**.** *For any $A \in \mathbb{R}^{m \times n}$ and $B \in \mathbb{R}^{p \times s}$,*

$$(A \otimes B)^\top = A^\top \otimes B^\top.$$

Proof. The (i, j)-block of $A \otimes B$ is $A_{ij} B \in \mathbb{R}^{p \times s}$, so the (j, i)-block of $(A \otimes B)^\top$ is $(A_{ij} B)^\top = A_{ij} B^\top$. This equals the (j, i)-block of $A^\top \otimes B^\top$, since the (j, i)-entry of $A^\top$ is A_{ij}. $\square$

Lemma 3.3 (Inverse of Kronecker product)**.** *If $A \in \mathbb{R}^{n \times n}$ and $B \in \mathbb{R}^{m \times m}$ are invertible, then $A \otimes B$ is invertible and*

$$(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}.$$

Proof. We use the *mixed-product property* of the Kronecker product: for matrices of compatible sizes,

$$(A \otimes B)(C \otimes D) = (AC) \otimes (BD). \tag{9}$$

Applying (9),

$$(A \otimes B)(A^{-1} \otimes B^{-1}) = (AA^{-1}) \otimes (BB^{-1}) = I_n \otimes I_m = I_{nm}.$$

Similarly, $(A^{-1} \otimes B^{-1})(A \otimes B) = I_{nm}$. $\square$

Lemma 3.4 (Gram matrix of a Khatri–Rao product)**.** *Let $C_1 \in \mathbb{R}^{p_1 \times r}, \dots, C_L \in \mathbb{R}^{p_L \times r}$ and define $Z := C_L \odot \cdots \odot C_1 \in \mathbb{R}^{(p_1 \cdots p_L) \times r}$. Then*

$$Z^\top Z = (C_L^\top C_L) * (C_{L-1}^\top C_{L-1}) * \cdots * (C_1^\top C_1),$$

where $*$ denotes the Hadamard (entrywise) product.

Proof. The jth column of Z is $z_j = c_j^{(L)} \otimes c_j^{(L-1)} \otimes \cdots \otimes c_j^{(1)}$, where $c_j^{(\alpha)}$ denotes the jth column of C_α. Hence

$$(Z^\top Z)_{ij} = z_i^\top z_j = \big(c_i^{(L)} \otimes \cdots \otimes c_i^{(1)}\big)^\top \big(c_j^{(L)} \otimes \cdots \otimes c_j^{(1)}\big).$$

Using the identity $(u_L \otimes \cdots \otimes u_1)^\top (v_L \otimes \cdots \otimes v_1) = \prod_{\alpha=1}^L u_\alpha^\top v_\alpha$ (which follows from the mixed-product property (9) applied to 1×1 outputs), we obtain

$$(Z^\top Z)_{ij} = \prod_{\alpha=1}^{L} \big(c_i^{(\alpha)}\big)^\top c_j^{(\alpha)} = \prod_{\alpha=1}^{L} (C_\alpha^\top C_\alpha)_{ij}.$$

Since the entrywise product of matrices satisfies $\big(\prod_\alpha^* G_\alpha\big)_{ij} = \prod_\alpha (G_\alpha)_{ij}$, the result follows. $\square$

Lemma 3.5 (Selection matrix identities)**.** *Let $S \in \mathbb{R}^{N \times q}$ be as in (3), with columns $e_{p(1)}, \dots, e_{p(q)}$ drawn from the standard basis of $\mathbb{R}^N$, where the indices $p(1), \dots, p(q)$ are pairwise distinct. Then:*

(i) $S^\top S = I_q$.

(ii) $SS^\top = \operatorname{diag}(\mathbf{1}_\mathcal{P})$, *where $(\mathbf{1}_\mathcal{P})_j = 1$ if $j \in \mathcal{P}$ and 0 otherwise (with $\mathcal{P}$ as in (2)). In particular, $SS^\top$ is an orthogonal projector of rank q.*

(iii) *For every $x \in \mathbb{R}^N$, $SS^\top x = S(S^\top x)$.*

(iv) (Gather.) *For any $x \in \mathbb{R}^N$, $(S^\top x)_\ell = x_{p(\ell)}$ for $\ell = 1, \dots, q$.*

(v) (Scatter.) *For any $y \in \mathbb{R}^q$, $(Sy)_{p(\ell)} = y_\ell$ for $\ell = 1, \dots, q$, and $(Sy)_j = 0$ for $j \notin \mathcal{P}$.*

Proof. Write $S = [e_{p(1)}, \dots, e_{p(q)}]$.

(i). $(S^\top S)_{ij} = e_{p(i)}^\top e_{p(j)} = \delta_{p(i),p(j)}$. Since the indices $p(1), \dots, p(q)$ are distinct, $\delta_{p(i),p(j)} = \delta_{ij}$, giving $S^\top S = I_q$.

(ii). $SS^\top = \sum_{\ell=1}^q e_{p(\ell)}\, e_{p(\ell)}^\top$. Each summand is the $N \times N$ matrix with a single 1 at position $(p(\ell), p(\ell))$ and zeros elsewhere. Since the $p(\ell)$ are distinct, the sum is the diagonal matrix $\operatorname{diag}(\mathbf{1}_\mathcal{P})$. As a diagonal $\{0,1\}$-matrix, it satisfies $(SS^\top)^2 = SS^\top(SS^\top) = S(S^\top S)S^\top = SI_qS^\top = SS^\top$, confirming that $SS^\top$ is an orthogonal projector.

(iii). This is the associativity of matrix multiplication: $SS^\top x = S(S^\top x)$.

(iv). $(S^\top x)_\ell = e_{p(\ell)}^\top x = x_{p(\ell)}$.

(v). $Sy = \sum_{\ell=1}^q y_\ell\, e_{p(\ell)}$. The jth entry of this sum is y_ℓ if $j = p(\ell)$ for some (unique) ℓ, and 0 if $j \notin \mathcal{P}$. $\square$

Lemma 3.6 (Vec–trace identity)**.** *For any* $A, B \in \mathbb{R}^{n\times r}$*,*

$$\mathrm{vec}(A)^\top \mathrm{vec}(B) = \mathrm{tr}(A^\top B) = \langle A, B\rangle_F.$$

In particular, $\mathrm{vec}(A)^\top \mathrm{vec}(A) = \|A\|_F^2$.

Proof. Write $A = [a_1, \dots, a_r]$ and $B = [b_1, \dots, b_r]$ with $a_j, b_j \in \mathbb{R}^n$. Then $\mathrm{vec}(A)^\top \mathrm{vec}(B) = \sum_{j=1}^r a_j^\top b_j$. On the other hand, $(A^\top B)_{jj} = a_j^\top b_j$, so $\mathrm{tr}(A^\top B) = \sum_{j=1}^r a_j^\top b_j$. The Frobenius inner product $\langle A, B\rangle_F$ is defined as $\mathrm{tr}(A^\top B)$. □

Lemma 3.7 (Cyclic property of trace)**.** *For any matrices* $A \in \mathbb{R}^{m\times p}$*,* $B \in \mathbb{R}^{p\times s}$*,* $C \in \mathbb{R}^{s\times m}$*,*

$$\mathrm{tr}(ABC) = \mathrm{tr}(BCA) = \mathrm{tr}(CAB).$$

Proof. By definition, $\mathrm{tr}(ABC) = \sum_{i=1}^m (ABC)_{ii} = \sum_{i=1}^m \sum_{j=1}^p \sum_{\ell=1}^s A_{ij}\, B_{j\ell}\, C_{\ell i}$. Rearranging the summation order as $\sum_j \sum_\ell \sum_i B_{j\ell}\, C_{\ell i}\, A_{ij} = \mathrm{tr}(BCA)$. The second equality follows by applying the first to the triple (B, C, A). □

4 Positive Definiteness of H

Write $H = H_1 + H_2$ with

$$H_1 := (Z \otimes K)^\top S\, S^\top (Z \otimes K), \qquad H_2 := \lambda\, (I_r \otimes K). \tag{10}$$

Lemma 4.1 (Symmetry of H)**.** *If K is symmetric, then H is symmetric.*

Proof. For H_1: by Lemma 3.2, $(Z \otimes K)^\top = Z^\top \otimes K^\top = Z^\top \otimes K$ (using $K^\top = K$). Hence

$$H_1^\top = \left[(Z \otimes K)^\top S S^\top (Z \otimes K)\right]^\top = (Z \otimes K)^\top (S S^\top)^\top (Z \otimes K).$$

Since $SS^\top$ is symmetric (it is a diagonal matrix by Lemma 3.5(ii)), we obtain $H_1^\top = H_1$.

For H_2: since K is symmetric, $H_2^\top = \lambda\, (I_r \otimes K)^\top = \lambda\, (I_r^\top \otimes K^\top) = \lambda\, (I_r \otimes K) = H_2$, by Lemma 3.2.

Hence $H = H_1 + H_2$ is symmetric. □

Lemma 4.2. $H_1 \succeq 0$.

Proof. For any $x \in \mathbb{R}^{nr}$, let $y := (Z \otimes K)\, x \in \mathbb{R}^N$. Then

$$x^\top H_1\, x = y^\top S S^\top y = (S^\top y)^\top (S^\top y) = \left\|S^\top y\right\|_2^2 \;\geq\; 0. \qquad \square$$

Lemma 4.3. *Under Assumptions 2.2–2.3 ($K \succ 0$ and $\lambda > 0$),* $H_2 \succ 0$.

Proof. The eigenvalues of $I_r \otimes K$ are the eigenvalues of K, each repeated r times. Indeed, if $Kv = \mu v$ with $v \neq 0$, then $(I_r \otimes K)(e_j \otimes v) = e_j \otimes (Kv) = \mu\, (e_j \otimes v)$ for each standard basis vector $e_j \in \mathbb{R}^r$, $j = 1, \dots, r$. Since $K \succ 0$, every eigenvalue μ of K is positive, so every eigenvalue of $I_r \otimes K$ is positive. Multiplying by $\lambda > 0$ preserves strict positivity. Hence $H_2 = \lambda\, (I_r \otimes K) \succ 0$. □

Theorem 4.4 (SPD property of H)**.** *Under Assumptions 2.2–2.3, the matrix H defined in (7) is symmetric positive definite.*

Proof. Symmetry is given by Lemma 4.1. For any nonzero $x \in \mathbb{R}^{nr}$,

$$x^\top H\, x = \underbrace{x^\top H_1\, x}_{\geq 0 \text{ (Lemma 4.2)}} + \underbrace{x^\top H_2\, x}_{> 0 \text{ (Lemma 4.3)}} > 0.$$

Hence $H \succ 0$. □

Remark 4.5 (The case of singular K). If K is only positive semidefinite ($K \succeq 0$, $K \not\succ 0$), then H is necessarily singular, regardless of the value of $\lambda > 0$.

Proposition 4.6. *If* $\ker(K) \neq \{0\}$ *and* $\lambda > 0$*, then* H *is singular.*

Proof. Choose any nonzero $u \in \ker(K)$ and any nonzero $v \in \mathbb{R}^r$. Define $W := uv^\top \in \mathbb{R}^{n\times r}$; since $u \neq 0$ and $v \neq 0$, we have $\mathrm{vec}(W) \neq 0$. We show $H\ \mathrm{vec}(W) = 0$ by verifying both terms vanish.

Term H_2. By Lemma 3.1,

$$(I_r \otimes K)\ \mathrm{vec}(uv^\top) = \mathrm{vec}(Kuv^\top I_r) = \mathrm{vec}(Ku \cdot v^\top).$$

Since $u \in \ker(K)$, $Ku = 0$, so $H_2\ \mathrm{vec}(W) = 0$.

Term H_1. Similarly, $(Z \otimes K)\ \mathrm{vec}(uv^\top) = \mathrm{vec}(Ku \cdot v^\top Z^\top)$ by Lemma 3.1 (with $A = K$, $B = Z$, $X = uv^\top$). Since $Ku = 0$, the entire expression vanishes. Thus $SS^\top (Z \otimes K)\,\mathrm{vec}(W) = 0$, and consequently $H_1\ \mathrm{vec}(W) = 0$.

Hence $H\ \mathrm{vec}(W) = H_1\ \mathrm{vec}(W) + H_2\ \mathrm{vec}(W) = 0$ with $\mathrm{vec}(W) \neq 0$, so H is singular. □

Remark 4.7 (Practical remedies for singular K). When K is only PSD, two standard remedies are available.

(i) **Nugget regularization.** Replace K by $K_\varepsilon := K + \varepsilon I_n$ with $\varepsilon > 0$. Then $K_\varepsilon \succ 0$, and Theorem 4.4 applies to the modified system. Note that this changes the linear system (and hence its solution); bounding the perturbation of the solution requires additional analysis that we do not pursue here.

(ii) **Least-squares / minimum-residual solution.** One may seek a least-squares solution of the singular system via an appropriate iterative method (e.g., LSQR; or MINRES-QLP if a minimum-norm solution is desired). We do not develop this approach further.

5 Fast Matrix–Vector Product for H

5.1 Input/output contract

Given an input vector $v \in \mathbb{R}^{nr}$, reshape it as $W_v \in \mathbb{R}^{n\times r}$ so that $v = \mathrm{vec}(W_v)$. The goal is to compute the output

$$y := Hv \ \in\ \mathbb{R}^{nr}$$

using only the following data: the observed index tuples $\Omega = \{(i_1(\ell), \ldots, i_d(\ell))\}_{\ell=1}^{q}$, the factor matrices $A_i \in \mathbb{R}^{n_i\times r}$ for $i \neq k$, the kernel matrix $K \in \mathbb{R}^{n\times n}$, and the parameter $\lambda > 0$. In particular, we never form:

(i) the Khatri–Rao product $Z \in \mathbb{R}^{M\times r}$ as a dense matrix;

(ii) the Kronecker product $Z \otimes K \in \mathbb{R}^{N\times nr}$;

(iii) any dense vector of length N or dense matrix with an N- or M-dimensional axis.

5.2 Decomposition

Using the splitting $H = H_1 + H_2$ from (10), write

$$y = y_1 + y_2, \qquad y_1 := H_1\, v, \quad y_2 := H_2\, v. \tag{11}$$

We compute y_2 and y_1 separately.

5.3 Regularization term y_2

Apply Lemma 3.1 with $A = K \in \mathbb{R}^{n\times n}$, $B = I_r \in \mathbb{R}^{r\times r}$, $X = W_v \in \mathbb{R}^{n\times r}$:

$$(I_r \otimes K)\ \mathrm{vec}(W_v) = \mathrm{vec}(K\, W_v\, I_r^\top) = \mathrm{vec}(KW_v). \tag{12}$$

Define the intermediate matrix

$$U := KW_v \ \in\ \mathbb{R}^{n\times r}. \tag{13}$$

Then

$$y_2 = \lambda\ \mathrm{vec}(U). \tag{14}$$

Cost: computing $U = KW_v$ requires one dense matrix multiply of $K \in \mathbb{R}^{n\times n}$ by $W_v \in \mathbb{R}^{n\times r}$, costing $O(n^2 r)$ operations.

5.4 Data-fitting term y_1

The data-fitting term is

$$y_1 = (Z \otimes K)^\top S\, S^\top (Z \otimes K)\ \mathrm{vec}(W_v).$$

We compute y_1 in three stages, reading right to left.

Stage (a): Compute $f = S^\top (Z \otimes K)\,\mathrm{vec}(W_v) \in \mathbb{R}^q$

Apply Lemma 3.1 with $A = K \in \mathbb{R}^{n\times n}$, $B = Z \in \mathbb{R}^{M\times r}$, $X = W_v \in \mathbb{R}^{n\times r}$:

$$(Z \otimes K)\ \mathrm{vec}(W_v) = \mathrm{vec}(K\, W_v\, Z^\top) = \mathrm{vec}(UZ^\top), \tag{15}$$

where $U = KW_v$ as in (13), and $UZ^\top \in \mathbb{R}^{n\times M}$.

The vector $f := S^\top\ \mathrm{vec}(UZ^\top) \in \mathbb{R}^q$ extracts the q entries of $\mathrm{vec}(UZ^\top)$ at the observed positions (Lemma 3.5(iv)). Since the ℓth observed position in the mode-k unfolding corresponds to row $i_k(\ell)$ and column $m(\ell)$, we have

$$f_\ell = (UZ^\top)_{i_k(\ell),\, m(\ell)} = U_{i_k(\ell),:}\ Z^\top_{m(\ell),:}, \qquad \ell = 1, \ldots, q, \tag{16}$$

where the right-hand side is a dot product in $\mathbb{R}^r$ between the $i_k(\ell)$th row of U and the $m(\ell)$th row of Z.

Computing rows of Z without forming Z. The matrix $Z = A_d \odot \cdots \odot A_{k+1} \odot A_{k-1} \odot \cdots \odot A_1$ has rows indexed by multi-indices $(j_1, \ldots, j_{k-1}, j_{k+1}, \ldots, j_d)$ with $j_i \in \{1, \ldots, n_i\}$. The following observation lets us evaluate any single row on the fly.

Lemma 5.1 (Rows of the Khatri–Rao product)**.** *Let $Z = C_L \odot \cdots \odot C_1$ with $C_\alpha \in \mathbb{R}^{p_\alpha \times r}$, and let a row index $\mu \in \{1, \ldots, p_1 \cdots p_L\}$ correspond to the multi-index $(\mu_1, \ldots, \mu_L)$ under the lexicographic ordering that defines the Khatri–Rao product. Then*

$$Z_{\mu,:} = (C_L)_{\mu_L,:} * (C_{L-1})_{\mu_{L-1},:} * \cdots * (C_1)_{\mu_1,:} \;\in\; \mathbb{R}^{1\times r},$$

where $*$ denotes the Hadamard (entrywise) product of row vectors.

Proof. By definition, the jth column of Z is $c_j^{(L)} \otimes \cdots \otimes c_j^{(1)}$, whose μth entry (corresponding to multi-index $(\mu_1, \ldots, \mu_L)$) is $\prod_{\alpha=1}^{L} (C_\alpha)_{\mu_\alpha, j}$. Hence $Z_{\mu,j} = \prod_\alpha (C_\alpha)_{\mu_\alpha,j}$, which is exactly the jth entry of the Hadamard product of the rows $(C_\alpha)_{\mu_\alpha,:}$. □

Applying Lemma 5.1 to (4), the row of Z corresponding to observation ℓ is

$$z_\ell := Z_{m(\ell),:} = A_d\big(i_d(\ell),:\big) * \cdots * A_{k+1}\big(i_{k+1}(\ell),:\big) * A_{k-1}\big(i_{k-1}(\ell),:\big) * \cdots * A_1\big(i_1(\ell),:\big) \;\in\; \mathbb{R}^{1\times r}, \tag{17}$$

requiring $d-2$ Hadamard products of vectors in $\mathbb{R}^r$, i.e. $(d-2)r$ multiplications per observation. The scalar (16) then becomes

$$f_\ell = U_{i_k(\ell),:}\, z_\ell^\top \qquad \text{(dot product in } \mathbb{R}^r\text{, cost } O(r)\text{)}. \tag{18}$$

Total cost of Stage (a): forming all q vectors z_ℓ costs $O(q(d-2)r)$; the q dot products cost $O(qr)$. Since each observation requires $(d-2)r$ multiplications for z_ℓ plus $O(r)$ for the dot product, the per-observation cost is $O((d-1)r)$. Overall: $O(q(d-2)r + qr) = O(q(d-1)r)$.

Stage (b): Scatter f into the sparse matrix $\widetilde{T}$

Define the sparse matrix

$$\widetilde{T} \in \mathbb{R}^{n\times M}, \qquad \widetilde{T}_{i,m} := \begin{cases} f_\ell & \text{if } (i,m) = (i_k(\ell),\, m(\ell)) \text{ for some } \ell \in \{1,\ldots,q\},\\ 0 & \text{otherwise.}\end{cases} \tag{19}$$

This is well-defined under Assumption 2.1 (no two observations share the same (i_k, m) pair).

Claim 5.2. $\mathrm{vec}(\widetilde{T}) = Sf$.

Proof of Claim 5.2. By Lemma 3.5(v), $Sf = \sum_{\ell=1}^{q} f_\ell\, e_{p(\ell)}$, so $(Sf)_{p(\ell)} = f_\ell$ for each ℓ and $(Sf)_j = 0$ for $j \notin \mathcal{P}$. By definition of $p(\ell)$ (Section 2.3), position $p(\ell)$ in $\mathrm{vec}(T)$ corresponds to row $i_k(\ell)$ and column $m(\ell)$ in the mode-k unfolding. The vector $\mathrm{vec}(\widetilde{T})$ has exactly the same nonzero pattern and values, by (19). □

Combining (15), Claim 5.2, and Lemma 3.5(iii):

$$SS^\top (Z \otimes K)\,\mathrm{vec}(W_v) = S\big(S^\top \mathrm{vec}(UZ^\top)\big) = Sf = \mathrm{vec}(\widetilde{T}). \tag{20}$$

Stage (c): Compute $(Z \otimes K)^\top\,\mathrm{vec}(\widetilde{T})$ without length-N objects

By Lemma 3.2 and symmetry of K,

$$(Z \otimes K)^\top = Z^\top \otimes K^\top = Z^\top \otimes K. \tag{21}$$

Apply Lemma 3.1 with $A = K \in \mathbb{R}^{n\times n}$, $B = Z^\top \in \mathbb{R}^{r\times M}$ (so $B^\top = Z$), $X = \widetilde{T} \in \mathbb{R}^{n\times M}$:

$$(Z^\top \otimes K)\;\mathrm{vec}(\widetilde{T}) = \mathrm{vec}\big(K\,\widetilde{T}\,(Z^\top)^\top\big) = \mathrm{vec}(K\,\widetilde{T}\,Z). \tag{22}$$

Define

$$G := \widetilde{T}\,Z \;\in\; \mathbb{R}^{n\times r}. \tag{23}$$

Then $y_1 = \mathrm{vec}(KG)$.

Computing G by scatter-add. Since $\widetilde{T}$ is sparse with at most q nonzeros (one per observation), the product $G = \widetilde{T}\,Z$ can be accumulated without forming either $\widetilde{T}$ or Z as dense matrices:

$$\text{Initialize } G = 0 \in \mathbb{R}^{n\times r}; \quad \text{for } \ell = 1, \dots, q\text{:}\;\; G_{i_k(\ell),:} \;\mathrel{+}=\; f_\ell\, z_\ell, \tag{24}$$

where $z_\ell = Z_{m(\ell),:} \in \mathbb{R}^{1\times r}$ is the row computed via (17).

Claim 5.3. *The loop* (24) *computes* $G = \widetilde{T}\,Z$.

Proof of Claim 5.3. By (19), the ith row of $\widetilde{T}$ is supported on positions $\{m(\ell) : i_k(\ell) = i\}$, so

$$(\widetilde{T}\,Z)_{i,:} = \sum_{m=1}^{M} \widetilde{T}_{i,m}\, Z_{m,:} = \sum_{\substack{\ell=1\\ i_k(\ell)=i}}^{q} f_\ell\, z_\ell.$$

This is exactly the result of accumulating $f_\ell\, z_\ell$ into row $i_k(\ell)$ for each ℓ. □

After obtaining G, compute

$$Y := KG \;\in\; \mathbb{R}^{n\times r} \tag{25}$$

by a dense matrix multiply, and set

$$y_1 := \mathrm{vec}(Y). \tag{26}$$

Cost of Stage (c): computing all z_ℓ and the scatter-add costs $O(q(d-1)r)$; the dense multiply KG costs $O(n^2 r)$. Overall: $O(q(d-1)r + n^2 r)$.

5.5 Final assembly

Combining (14) and (26):

$$y = y_1 + y_2 = \mathrm{vec}(KG) + \lambda\;\mathrm{vec}(U) = \mathrm{vec}(KG + \lambda\, U). \tag{27}$$

5.6 Correctness

Proposition 5.4 (Matvec correctness)**.** *The procedure described in Sections 5.3–5.5 returns exactly* $y = H\;\mathrm{vec}(W_v)$.

Proof. We trace the equality chain for the data-fitting term. Starting from the right:

$$\begin{aligned}
y_1 &= (Z\otimes K)^\top S\, S^\top (Z\otimes K)\;\mathrm{vec}(W_v) && \\
&= (Z\otimes K)^\top S\, S^\top\,\mathrm{vec}(UZ^\top) && \text{(Lemma 3.1, eq. (15))}\\
&= (Z\otimes K)^\top S\big(S^\top\,\mathrm{vec}(UZ^\top)\big) && \text{(Lemma 3.5(iii))}\\
&= (Z\otimes K)^\top S\, f && \text{(definition } f := S^\top\,\mathrm{vec}(UZ^\top))\\
&= (Z\otimes K)^\top\,\mathrm{vec}(\widetilde{T}) && \text{(Claim 5.2)}\\
&= (Z^\top\otimes K)\;\mathrm{vec}(\widetilde{T}) && \text{(Lemma 3.2, } K = K^\top)\\
&= \mathrm{vec}(K\,\widetilde{T}\,Z) && \text{(Lemma 3.1, eq. (22))}\\
&= \mathrm{vec}(KG). && \text{(Definition of } G\text{, eq. (23))}
\end{aligned}$$

For the regularization term, $y_2 = \lambda\,(I_r\otimes K)\;\mathrm{vec}(W_v) = \lambda\;\mathrm{vec}(U)$ by (12)–(14). Summing gives $y = y_1 + y_2 = H\;\mathrm{vec}(W_v)$. □

5.7 Complexity

Proposition 5.5 (Matvec cost)**.** *Each evaluation of $y = H\ \mathrm{vec}(W_v)$ via the procedure in Sections 5.3–5.5 requires*

$$O\big(n^2 r + q(d-1)r\big)$$

arithmetic operations.

Proof. The costs are:

(i) $U = KW_v$: $O(n^2 r)$.

(ii) Stage (a) (form each z_ℓ and dot product): $O(q(d-1)r)$.

(iii) Stage (c) scatter-add (recompute each z_ℓ and accumulate into G): $O(q(d-1)r)$.

(iv) $Y = KG$: $O(n^2 r)$.

(v) Vector addition $y_1 + y_2$: $O(nr)$.

Summing: $O(n^2 r + q(d-1)r)$, since $nr \le n^2 r$ for $n \ge 1$. □

Remark 5.6 (Caching Khatri–Rao rows). The row vectors $z_\ell = Z_{m(\ell),:}$ are computed identically in Stage (a) (17) and Stage (c) (24). If $O(qr)$ additional memory is acceptable, one may cache all q row vectors after Stage (a) and reuse them in Stage (c), replacing the $O(q(d-1)r)$ cost of recomputation in Stage (c) by $O(qr)$ for the scatter-add alone. The total cost then becomes $O(n^2 r + q(d-1)r)$ with a smaller constant.

6 Fast Right-Hand Side Construction

6.1 Observed values

For each observation $\ell \in \{1, \dots, q\}$, denote the observed scalar entry of the tensor (equivalently, of the mode-k unfolding) by

$$t_\ell := T_{i_k(\ell),\, m(\ell)}. \tag{28}$$

Since unobserved entries of T are set to zero (Section 2.2), the values $\{t_\ell\}_{\ell=1}^{q}$ account for all nonzeros of T.

6.2 Computing $B = TZ$ without forming T or Z

Proposition 6.1 (Sparse MTTKRP)**.** *The matrix $B = TZ \in \mathbb{R}^{n\times r}$ can be computed by:*

$$\text{Initialize } B = 0 \in \mathbb{R}^{n\times r}; \quad \text{for } \ell = 1, \dots, q\colon\ B_{i_k(\ell),:} \mathrel{+}= t_\ell\, z_\ell, \tag{29}$$

where $z_\ell = Z_{m(\ell),:} \in \mathbb{R}^{1\times r}$ is computed via Lemma 5.1 (eq. (17)).

Proof. The ith row of $B = TZ$ is

$$B_{i,:} = \sum_{m=1}^{M} T_{i,m}\, Z_{m,:}.$$

Since $T_{i,m} = 0$ unless $(i, m) = (i_k(\ell), m(\ell))$ for some observation ℓ (and these pairs are distinct by Assumption 2.1), the sum reduces to

$$B_{i,:} = \sum_{\substack{\ell=1 \\ i_k(\ell)=i}}^{q} t_\ell \, z_\ell.$$

This is precisely the result of accumulating $t_\ell \, z_\ell$ into row $i_k(\ell)$ for each ℓ. $\square$

6.3 Computing $b = (I_r \otimes K)\,\mathrm{vec}(B)$

Apply Lemma 3.1 with $A = K$, $B_{\mathrm{right}} = I_r$, $X = B$:

$$b = (I_r \otimes K)\ \mathrm{vec}(B) = \mathrm{vec}(K\,B\,I_r^\top) = \mathrm{vec}(KB). \tag{30}$$

Proposition 6.2 (RHS correctness and cost)**.** *The right-hand side $b \in \mathbb{R}^{nr}$ of (6) is computed correctly by (29) followed by (30). The total cost is $O\big(q(d-1)r + n^2 r\big)$.*

Proof. Correctness follows from Proposition 6.1 and (30). The cost comprises:

(i) Computing all z_ℓ and the scatter-add (29): $O(q(d-1)r)$ (each z_ℓ costs $O((d-2)r)$ via $d-2$ Hadamard products, and each accumulation costs $O(r)$, for $O((d-1)r)$ per observation).

(ii) Dense multiply $KB \in \mathbb{R}^{n\times r}$: $O(n^2 r)$.

Total: $O(q(d-1)r + n^2 r)$. $\square$

Dimension check. $B \in \mathbb{R}^{n\times r}$ by construction, so $\mathrm{vec}(B) \in \mathbb{R}^{nr}$ and $b = \mathrm{vec}(KB) \in \mathbb{R}^{nr}$, consistent with the system (6).

7 Kronecker Preconditioner

7.1 Definition

Define the preconditioner

$$P := (Z^\top Z + \lambda I_r) \otimes K \ \in\ \mathbb{R}^{nr\times nr}. \tag{31}$$

For later convenience, define the $r \times r$ matrix

$$R := Z^\top Z + \lambda I_r, \tag{32}$$

so that $P = R \otimes K$.

Heuristic motivation. In the (hypothetical) fully observed case $SS^\top = I_N$, the data-fitting term of H becomes $(Z \otimes K)^\top (Z \otimes K) = (Z^\top Z) \otimes K^2$ (by the mixed-product property (9) and Lemma 3.2), and the full operator is $(Z^\top Z) \otimes K^2 + \lambda\, I_r \otimes K$. The preconditioner (31) matches the regularization term $\lambda\, I_r \otimes K$ exactly and replaces K^2 by K in the data-fitting part, preserving the Kronecker structure that enables efficient inversion. We do *not* claim any spectral condition-number bound for $P^{-1}H$.

7.2 Positive definiteness of P

Lemma 7.1. *Under Assumption 2.3 ($\lambda > 0$), $R = Z^\top Z + \lambda I_r$ is symmetric positive definite.*

Proof. Since $Z^\top Z$ is a Gram matrix, it is symmetric and PSD: for any $w \in \mathbb{R}^r$, $w^\top (Z^\top Z) w = \|Zw\|_2^2 \geq 0$. Adding λI_r with $\lambda > 0$ gives, for any nonzero w, $w^\top R w = \|Zw\|_2^2 + \lambda \|w\|_2^2 > 0$. □

Proposition 7.2 (SPD property of P)**.** *Under Assumptions 2.2–2.3, the preconditioner $P = R \otimes K$ is symmetric positive definite.*

Proof. Both R and K are symmetric (Lemma 7.1 and Assumption 2.2), so $P = R \otimes K$ is symmetric by Lemma 3.2: $(R \otimes K)^\top = R^\top \otimes K^\top = R \otimes K$.

For positive definiteness, let $x \in \mathbb{R}^{nr}$ be nonzero and write $x = \operatorname{vec}(X)$ with $X \in \mathbb{R}^{n \times r}$. By Lemma 3.1,

$$x^\top (R \otimes K)\, x = \operatorname{vec}(X)^\top \operatorname{vec}(KXR^\top) = \operatorname{vec}(X)^\top \operatorname{vec}(KXR),$$

where we used $R^\top = R$. Since K and R are SPD, they admit Cholesky factorizations $K = L_K L_K^\top$ and $R = L_R L_R^\top$. Substituting and applying Lemma 3.6 ($\operatorname{vec}(A)^\top \operatorname{vec}(B) = \operatorname{tr}(A^\top B)$):

$$\begin{aligned} x^\top P\, x &= \operatorname{vec}(X)^\top \operatorname{vec}(L_K L_K^\top X\, L_R L_R^\top) = \operatorname{tr}\big(X^\top L_K L_K^\top X\, L_R L_R^\top\big) \\ &= \operatorname{tr}\big(L_R^\top X^\top L_K L_K^\top X\, L_R\big) \qquad \text{(Lemma 3.7)} \\ &= \operatorname{tr}\big((L_K^\top X L_R)^\top (L_K^\top X L_R)\big) = \big\|L_K^\top X L_R\big\|_F^2. \end{aligned}$$

Since L_K and L_R are invertible and $X \neq 0$, the matrix $L_K^\top X L_R$ is nonzero, so $\|L_K^\top X L_R\|_F^2 > 0$. □

7.3 Inverse and application formula

By Lemma 3.3 (applicable since R and K are both invertible under Assumptions 2.2–2.3):

$$P^{-1} = (R \otimes K)^{-1} = R^{-1} \otimes K^{-1}. \tag{33}$$

Proposition 7.3 (Preconditioner application)**.** *For any $V \in \mathbb{R}^{n \times r}$,*

$$P^{-1} \operatorname{vec}(V) = \operatorname{vec}\big(K^{-1} V\, R^{-1}\big). \tag{34}$$

Proof. Apply Lemma 3.1 with $A = K^{-1} \in \mathbb{R}^{n \times n}$, $B = R^{-1} \in \mathbb{R}^{r \times r}$, $X = V \in \mathbb{R}^{n \times r}$:

$$(R^{-1} \otimes K^{-1})\ \operatorname{vec}(V) = \operatorname{vec}\big(K^{-1} V\, (R^{-1})^\top\big) = \operatorname{vec}\big(K^{-1} V\, R^{-1}\big),$$

where the last equality uses $R^{-1} = (R^\top)^{-1} = (R^{-1})^\top$ (since R is symmetric). □

Practical application procedure. Given $v = \operatorname{vec}(V)$ with $V \in \mathbb{R}^{n \times r}$, compute $P^{-1} v$ as follows (without forming any inverse matrix explicitly):

(i) *Left solve.* Solve $KX = V$ for $X \in \mathbb{R}^{n \times r}$ (r right-hand sides) using a precomputed Cholesky factorization $K = L_K L_K^\top$: first solve $L_K W_1 = V$ by forward substitution, then $L_K^\top X = W_1$ by back substitution.

(ii) *Right solve.* Solve $Y R^\top = X$ for $Y \in \mathbb{R}^{n \times r}$ (equivalently, $R Y^\top = X^\top$) using a precomputed Cholesky factorization $R = L_R L_R^\top$: first solve $L_R W_2 = X^\top$ by forward substitution, then $L_R^\top Y^\top = W_2$ by back substitution. Since R is symmetric, $R^\top = R$, and $Y = X R^{-1}$.

(iii) Return $\operatorname{vec}(Y) = P^{-1} v$.

7.4 Computing $Z^\top Z$ without forming Z

Recall that $Z = A_d \odot \cdots \odot A_{k+1} \odot A_{k-1} \odot \cdots \odot A_1$ (eq. (4)). By Lemma 3.4, applied with $C_\alpha = A_{i_\alpha}$ for $i_\alpha \in \{1, \dots, d\} \setminus \{k\}$ (listed in the order used in (4)):

$$Z^\top Z = (A_d^\top A_d) * \cdots * (A_{k+1}^\top A_{k+1}) * (A_{k-1}^\top A_{k-1}) * \cdots * (A_1^\top A_1), \tag{35}$$

where each $A_i^\top A_i \in \mathbb{R}^{r \times r}$ and $*$ is the Hadamard product.

Algorithm.

(i) For each $i \in \{1, \dots, d\} \setminus \{k\}$, form the Gram matrix $G_i := A_i^\top A_i \in \mathbb{R}^{r \times r}$, at cost $O(n_i r^2)$.

(ii) Compute the Hadamard product of the $d-1$ matrices G_i to obtain $Z^\top Z$, at cost $O((d-2)r^2)$.

(iii) Form $R = Z^\top Z + \lambda I_r$.

7.5 Setup and per-application complexity

Proposition 7.4 (Preconditioner costs)**.** *Under the standing assumptions, the preconditioner $P = R \otimes K$ has the following costs (with d kept explicit).*

(a) ***Setup*** *(one-time, before iterating):*

  (i) *Cholesky factorization of $K \in \mathbb{R}^{n \times n}$: $O(n^3)$.*

  (ii) *Forming R: Gram matrices G_i cost $O\big(\sum_{i \neq k} n_i\, r^2\big)$; Hadamard products cost $O((d-2)\, r^2)$; total $O\big((\sum_{i \neq k} n_i)\, r^2 + (d-2)\, r^2\big)$.*

  (iii) *Cholesky factorization of $R \in \mathbb{R}^{r \times r}$: $O(r^3)$.*

(b) ***Per application*** *of P^{-1} (one solve per PCG iteration):*

  (i) *Left solve $KX = V$ (r right-hand sides via Cholesky): $O(n^2 r)$.*

  (ii) *Right solve $RY^\top = X^\top$ (n right-hand sides via Cholesky): $O(nr^2)$.*

  Total per application: $O(n^2 r + nr^2)$.

Proof. All cost claims follow from standard dense linear algebra: Cholesky factorization of an $m \times m$ SPD matrix costs $O(m^3)$; a triangular solve with p right-hand sides of length m costs $O(m^2 p)$; each Gram matrix $A_i^\top A_i$ is a product of an $r \times n_i$ matrix by an $n_i \times r$ matrix, costing $O(n_i r^2)$; each Hadamard product of two $r \times r$ matrices costs $O(r^2)$. □

7.6 Verification of the "no $O(N)$" property

The preconditioner setup and application never require forming or accessing any object of size M, N, or larger:

(i) $Z \in \mathbb{R}^{M \times r}$ is never formed; only the small Gram matrices $A_i^\top A_i \in \mathbb{R}^{r \times r}$ are computed, one per mode $i \neq k$.

(ii) $Z \otimes K \in \mathbb{R}^{N \times nr}$ is never formed.

(iii) The only dense matrices stored are $K \in \mathbb{R}^{n \times n}$ (and its Cholesky factor) and $R \in \mathbb{R}^{r \times r}$ (and its Cholesky factor).

All operations involve matrices of size at most $n \times n$, $n \times r$, or $r \times r$.

8 Preconditioned Conjugate Gradient Solver

8.1 Problem in vector form

Let $w := \mathrm{vec}(W) \in \mathbb{R}^{nr}$. The system to be solved is

$$Hw = b,$$

where $H \in \mathbb{R}^{nr\times nr}$ is SPD (Theorem 4.4) and $b \in \mathbb{R}^{nr}$ is defined in (8). The preconditioner $P \in \mathbb{R}^{nr\times nr}$ is SPD (Proposition 7.2). Neither H nor P is formed explicitly; they are accessed only through the routines developed in Sections 5 and 7, respectively.

8.2 Algorithm

We use the standard left-preconditioned conjugate gradient method. Given an initial guess $w_0 \in \mathbb{R}^{nr}$ (typically $w_0 = 0$) and a relative residual tolerance $\mathrm{tol} > 0$, the iteration proceeds as follows.

Trivial case. If $b = 0$, then $w_* = 0$ is the unique solution (since H is SPD), so we return $w = 0$ without iterating.

Algorithm: Preconditioned Conjugate Gradient (PCG)

Input: $w_0 \in \mathbb{R}^{nr}$, $\mathrm{tol} > 0$. Require $b \neq 0$ (see trivial case above).

1. $r_0 \leftarrow b - Hw_0$ *(matvec, Section 5)*
2. $z_0 \leftarrow P^{-1} r_0$ *(preconditioner apply, Section 7)*
3. $p_0 \leftarrow z_0$
4. **for** $t = 0, 1, 2, \ldots$
5. $\quad q_t \leftarrow Hp_t$ *(matvec, Section 5)*
6. $\quad \alpha_t \leftarrow (r_t^\top z_t) \,/\, (p_t^\top q_t)$
7. $\quad w_{t+1} \leftarrow w_t + \alpha_t\, p_t$
8. $\quad r_{t+1} \leftarrow r_t - \alpha_t\, q_t$
9. $\quad$ **if** $\|r_{t+1}\|_2 \,/\, \|b\|_2 \leq \mathrm{tol}$ **then stop**
10. $\quad z_{t+1} \leftarrow P^{-1} r_{t+1}$ *(preconditioner apply, Section 7)*
11. $\quad \beta_t \leftarrow (r_{t+1}^\top z_{t+1}) \,/\, (r_t^\top z_t)$
12. $\quad p_{t+1} \leftarrow z_{t+1} + \beta_t\, p_t$

Output: approximate solution w_{t+1}.

All vectors r_t, z_t, p_t, q_t, w_t lie in $\mathbb{R}^{nr}$. Lines 1 and 5 each require one evaluation of the matrix–vector product $v \mapsto Hv$ via the procedure of Sections 5.3–5.5 (Proposition 5.4). Lines 2 and 10 each require one application of P^{-1} via the Cholesky-based procedure of Section 7.3 (Proposition 7.3). The remaining operations (lines 6–8, 11–12) are inner products and AXPY operations on vectors of length nr.

8.3 Convergence

Define the *symmetrically preconditioned* operator

$$\widetilde{A} := P^{-1/2}\, H\, P^{-1/2} \;\in\; \mathbb{R}^{nr\times nr}, \tag{36}$$

where $P^{-1/2}$ is the symmetric positive definite square root of P^{-1} (which exists since P is SPD). Since H and P are both SPD, $\widetilde{A}$ is SPD and hence has positive eigenvalues $0 < \lambda_{\min}(\widetilde{A}) \le \cdots \le \lambda_{\max}(\widetilde{A})$. Define the *preconditioned condition number*

$$\kappa := \kappa(\widetilde{A}) = \frac{\lambda_{\max}(\widetilde{A})}{\lambda_{\min}(\widetilde{A})}. \tag{37}$$

The standard convergence theory for PCG (see, e.g., Golub and Van Loan, *Matrix Computations*, 4th ed., §11.5; Saad, *Iterative Methods for Sparse Linear Systems*, 2nd ed., §9.2) yields the following two results.

Proposition 8.1 (PCG convergence bound)**.** *Let $w_* := H^{-1}b$ denote the exact solution. The PCG iterates $\{w_t\}$ satisfy, for every $t \ge 0$,*

$$\|w_t - w_*\|_H \;\le\; 2\left(\frac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1}\right)^t \|w_0 - w_*\|_H, \tag{38}$$

where $\|v\|_H := \sqrt{v^\top H\, v}$ is the H-energy norm and $\kappa = \kappa(\widetilde{A})$ is as in (37).

Proof. This is the standard CG/PCG error bound expressed in the H-energy norm. The PCG iteration applied to $Hw = b$ with preconditioner P is equivalent to applying unpreconditioned CG to the transformed system $\widetilde{A}\,\tilde{w} = P^{-1/2}b$ with $\tilde{w} = P^{1/2}w$. The standard CG bound (Chebyshev polynomial argument) gives (38) with $\kappa = \kappa(\widetilde{A})$. □

Proposition 8.2 (Finite termination in exact arithmetic)**.** *In exact arithmetic, the PCG recurrence (Section 8.2, without early stopping) yields the exact solution in at most nr iterations: there exists $t \le nr$ such that $w_t = w_*$. In particular, $w_{nr} = w_*$.*

Proof. The search directions $\{p_0, \ldots, p_t\}$ are conjugate with respect to H (i.e., $p_i^\top H p_j = 0$ for $i \ne j$). Since $\mathbb{R}^{nr}$ has dimension nr, there can be at most nr H-conjugate directions. After nr steps, the Krylov subspace spans all of $\mathbb{R}^{nr}$ and the solution is exact. □

8.4 Per-iteration cost

Proposition 8.3 (Per-iteration cost)**.** *Each iteration of the PCG algorithm (Section 8.2) requires*

$$O\big(q(d-1)r + n^2 r + nr^2\big) \tag{39}$$

arithmetic operations, with d kept explicit.

Proof. The dominant costs per iteration are:

(i) One matrix–vector product $q_t = Hp_t$: $O(n^2 r + q(d-1)r)$ (Proposition 5.5).

(ii) One preconditioner application $z_{t+1} = P^{-1} r_{t+1}$: $O(n^2 r + nr^2)$ (Proposition 7.4(b)).

(iii) Two inner products ($r_t^\top z_t$ and $p_t^\top q_t$), two AXPY updates (w_{t+1}, r_{t+1}), one AXPY for p_{t+1}, and one norm computation $\|r_{t+1}\|_2$: each costs $O(nr)$.

Since $nr \le n^2 r$ for $n \ge 1$ and $nr \le nr^2$ for $r \ge 1$, the $O(nr)$ terms are dominated. Summing (i)–(ii) gives the stated bound. □

Remark 8.4 (Initialization cost)*.* The initial residual $r_0 = b - Hw_0$ (line 1) requires one matvec at cost $O(n^2 r + q(d-1)r)$, and the initial preconditioner solve $z_0 = P^{-1} r_0$ (line 2) costs $O(n^2 r + nr^2)$. These are each counted as part of the first iteration; the total cost over L iterations includes $L+1$ matvecs (counting line 1) and $L+1$ preconditioner solves (counting line 2). The asymptotic per-iteration cost (39) is unaffected.

9 Overall Complexity and Main Result

9.1 One-time setup costs

Before the PCG iteration begins, the following precomputations are performed.

Right-hand side construction (Section 6, Proposition 6.2).

(i) Sparse MTTKRP $B = TZ$ via scatter-add (29): $O(q(d-1)r)$.

(ii) Dense multiply $b = \mathrm{vec}(KB)$: $O(n^2 r)$.

Total RHS cost: $O(q(d-1)r + n^2 r)$.

Preconditioner setup (Section 7, Proposition 7.4(a)).

(i) Cholesky factorization of $K \in \mathbb{R}^{n\times n}$: $O(n^3)$.

(ii) Form $R = Z^\top Z + \lambda I_r$ via Gram matrices and Hadamard products (Section 7.4): $O\big((\sum_{i\neq k} n_i)\, r^2 + (d-2)\, r^2\big)$.

(iii) Cholesky factorization of $R \in \mathbb{R}^{r\times r}$: $O(r^3)$.

Total setup cost: $O\big(n^3 + r^3 + (\sum_{i\neq k} n_i)\, r^2 + (d-2)\, r^2\big)$.

9.2 Total runtime

Let L denote the number of PCG iterations performed (determined by the convergence criterion in Section 8.2).

Proposition 9.1 (Total arithmetic cost)**.** *The total cost of solving $Hw = b$ by L iterations of PCG, including all setup, is*

$$O\Big(\underbrace{n^3 + r^3 + \Big(\sum_{i\neq k} n_i\Big) r^2 + (d-2)\, r^2}_{\textit{preconditioner setup}} + \underbrace{q(d-1)r + n^2 r}_{\textit{RHS build}} + L\, \underbrace{\big(q(d-1)r + n^2 r + nr^2\big)}_{\textit{per iteration}}\Big). \tag{40}$$

All parameters n, r, q, d, and $\{n_i\}_{i\neq k}$ are kept explicit.

Proof. Sum the one-time costs (Section 9.1) and L times the per-iteration cost (Proposition 8.3). As noted in Remark 8.4, the actual count is $L+1$ matvecs and $L+1$ preconditioner solves (due to the initialization on lines 1–2), but $L+1$ and L differ by a constant factor absorbed into the big-O. □

Remark 9.2 (Iteration count). By Proposition 8.1, the number of iterations L required to reduce the H-energy error by a factor of ε satisfies $L = O(\sqrt{\kappa}\, \log(1/\varepsilon))$, where $\kappa = \kappa(\widetilde{A})$ is the preconditioned condition number defined in (37). We make no claim about the magnitude of κ for the specific preconditioner (31); bounding κ would require additional spectral analysis that we do not pursue here.

9.3 Memory requirements

The algorithm stores the following objects:

(i) Kernel matrix and its Cholesky factor: $O(n^2)$.

(ii) Factor matrices $A_i \in \mathbb{R}^{n_i \times r}$ for $i \neq k$: $O\big((\sum_{i\neq k} n_i)\, r\big)$.

(iii) Matrix $R \in \mathbb{R}^{r\times r}$ and its Cholesky factor: $O(r^2)$.

(iv) Observation list $\{(i_1(\ell), \dots, i_d(\ell))\}_{\ell=1}^{q}$: $O(qd)$.

(v) PCG work vectors (w_t, r_t, z_t, p_t, q_t, plus intermediate matrices $U, G, Y \in \mathbb{R}^{n\times r}$ from the matvec/preconditioner routines): $O(nr)$.

Optionally, caching all Khatri–Rao rows z_ℓ across the matvec computation (Remark 5.6) adds $O(qr)$ storage.

The total memory requirement is

$$O\Big(n^2 + \Big(\sum_{i\neq k} n_i\Big) r + r^2 + qd + nr\Big), \tag{41}$$

with an optional $O(qr)$ term for caching.

9.4 Absence of $O(N)$ operations

No step of the algorithm—setup, matvec, preconditioner application, or PCG bookkeeping—forms or accesses any dense object whose size depends on $M = N/n$ or $N = \prod_i n_i$:

(i) The Khatri–Rao product $Z \in \mathbb{R}^{M\times r}$ is never formed as a dense matrix; individual rows are computed on the fly in $O((d-1)r)$ (Section 5).

(ii) The Kronecker product $Z \otimes K \in \mathbb{R}^{N\times nr}$ is never formed.

(iii) The Gram matrix $Z^\top Z \in \mathbb{R}^{r\times r}$ is assembled from small factor Gram matrices $A_i^\top A_i$ via Lemma 3.4 (Section 7.4).

(iv) No dense vector of length N or M is ever allocated.

This was verified individually for the matvec (Section 5.1), the RHS construction (Section 6), and the preconditioner (Section 7.6).

9.5 Main theorem

Theorem 9.3 (PCG solver for the kernel-parameterized mode-k update)**.** *Under Assumptions 2.2–2.3 ($K \succ 0$, $\lambda > 0$), the following hold.*

(i) ***Well-posedness.*** *The system matrix H (7) is symmetric positive definite (Theorem 4.4), so the system $Hw = b$ has a unique solution $w_* = H^{-1}b$.*

(ii) ***Valid preconditioner.*** *The Kronecker preconditioner $P = R \otimes K$ (31) is symmetric positive definite (Proposition 7.2).*

(iii) ***Correctness.*** *The PCG iteration (Section 8.2), using the matrix–vector product of Proposition 5.4 and the preconditioner application of Proposition 7.3, produces iterates* $\{w_t\}$ *satisfying the convergence bound (38). In exact arithmetic, there exists an index* $t \leq nr$ *such that* $w_t = w_*$*; in particular* $w_{nr} = w_*$ *(Proposition 8.2).*

(iv) ***Complexity.*** *The total arithmetic cost for L PCG iterations (including all setup) is given by (40) (Proposition 9.1).*

(v) ***No*** $O(N)$ ***property.*** *Every arithmetic operation and every stored object has size depending only on* n*,* r*,* q*,* d*, and* $\{n_i\}_{i\neq k}$*, never on* M *or* N *(Section 9.4).*

Proof. Each claim has been proved in the referenced result: (i) Theorem 4.4; (ii) Proposition 7.2; (iii) Propositions 5.4, 7.3, 8.1, and 8.2; (iv) Proposition 9.1; (v) Section 9.4, which collects the individual verifications from Sections 5, 6, and 7. □