EFFICIENT SOLUTION OF THE RKHS-CONSTRAINED MODE SUBPROBLEM IN CP TENSOR COMPLETION VIA PRECONDITIONED CONJUGATE GRADIENTS

1. Setup

We adopt the notation and terminology of [5] for tensor operations. Let $\mathcal{T} \in \mathbb{R}^{n_1 \times \cdots \times n_d}$ be a partially observed d-way tensor, and suppose we seek a rank-r CP decomposition [5, Definition 3.1] via alternating optimization over the factor matrices $A_1, \ldots, A_d$. We consider the mode-k subproblem in which the factor matrices for all modes $i \neq k$ are fixed and mode k is constrained to a reproducing kernel Hilbert space (RKHS) in the sense of Aronszajn [1].

Write $n = n_k$, $M = \prod_{i \neq k} n_i$, and $N = nM$. Let $T \in \mathbb{R}^{n \times M}$ denote the mode-k unfolding of $\mathcal{T}$ [5, Section 2.4] with unobserved entries set to zero, and let $q \ll N$ denote the number of observed entries. Define the *selection matrix* $S \in \mathbb{R}^{N \times q}$ as the submatrix of I_N whose columns correspond to the observed entries, so that $S^T \operatorname{vec}(T)$ extracts the q known entries. Let

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

be the Khatri–Rao product [5, Section 2.6] of the fixed factor matrices, and let $B = TZ \in \mathbb{R}^{n \times r}$ be the matricized tensor times Khatri–Rao product (MTTKRP) [5, Section 3.3].

We parameterize the mode-k factor as $A_k = KW$, where $K \in \mathbb{R}^{n \times n}$ is the symmetric positive semi-definite kernel (Gram) matrix [1, Section 1] associated with the RKHS and $W \in \mathbb{R}^{n \times r}$ is unknown. The regularized least-squares subproblem yields the normal equations

$$\mathcal{A} \operatorname{vec}(W) = \mathbf{b}, \tag{1}$$

where

$$\mathcal{A} = (Z \otimes K)^T S\, S^T (Z \otimes K) + \lambda\, (I_r \otimes K), \qquad \mathbf{b} = (I_r \otimes K) \operatorname{vec}(B), \tag{2}$$

and $\lambda > 0$ is a regularization parameter. System (1) has dimension $nr \times nr$. A direct factorization costs $O(n^3 r^3)$ and requires explicit formation of $\mathcal{A}$; we show that preconditioned conjugate gradients (PCG) achieves per-iteration cost $O(qr + n^2 r + nr^2)$ with no computation of order N.

2. Properties of the system

Lemma 1. *The matrix $\mathcal{A}$ defined in (2) is symmetric positive semi-definite. If K is positive definite, then $\mathcal{A}$ is positive definite.*

Proof. Write $\mathcal{A} = C^T C + \lambda\, (I_r \otimes K)$ where $C = S^T (Z \otimes K)$. The first summand $C^T C$ is symmetric positive semi-definite, and $I_r \otimes K$ is positive semi-definite since K is. When $K \succ 0$, the second summand $\lambda(I_r \otimes K) \succ 0$, so $\mathcal{A} \succ 0$. □

Since $\mathcal{A}$ is symmetric positive definite (assuming $K \succ 0$), the conjugate gradient method [3] is applicable and converges in at most nr iterations.

3. Efficient matrix–vector products

The central tool is the following standard identity for Kronecker products and the vec operator.

Lemma 2 (cf. [4, Theorem 4.2.10])**.** *For matrices $A \in \mathbb{R}^{m \times n}$, $X \in \mathbb{R}^{n \times p}$, $B \in \mathbb{R}^{p \times s}$, one has*

$$(B^T \otimes A) \operatorname{vec}(X) = \operatorname{vec}(AXB).$$

Proposition 3. *Given $W \in \mathbb{R}^{n\times r}$, the product $\mathcal{A}\ \mathrm{vec}(W)$ can be evaluated in $O(qr+n^2r)$ operations without forming or storing any object of dimension N.*

Proof. We decompose the computation into three stages.

Stage 1: Forward map. By Lemma 2 with $A = K$, $X = W$, $B = Z^T$,

$$(Z \otimes K)\ \mathrm{vec}(W) = \mathrm{vec}(KWZ^T). \tag{3}$$

Forming the $n \times M$ matrix KWZ^T explicitly would cost $O(nMr)$, which is of order N. Instead, we first compute $P = KW \in \mathbb{R}^{n\times r}$ at cost $O(n^2r)$ and then evaluate only the q entries selected by S^T. Each observed entry is indexed by a pair $(i_s, j_s) \in \{1,\dots,n\} \times \{1,\dots,M\}$ for $s = 1,\dots,q$, and

$$y_s = (KWZ^T)_{i_s,j_s} = \sum_{\ell=1}^{r} P_{i_s,\ell}\, Z_{j_s,\ell}, \tag{4}$$

which requires $O(r)$ operations per entry and $O(qr)$ in total. Thus $y = S^T(Z \otimes K)\ \mathrm{vec}(W) \in \mathbb{R}^q$ is obtained in $O(n^2r + qr)$ operations without forming any N-dimensional quantity.

Stage 2: Adjoint map. We require $(Z \otimes K)^T S\, y = (Z^T \otimes K)\, Sy$. The vector $Sy \in \mathbb{R}^N$ has at most q nonzero entries; its mode-k matricization $V \in \mathbb{R}^{n\times M}$ satisfies $V_{i_s,j_s} = y_s$ for $s = 1,\dots,q$ and is zero elsewhere. By Lemma 2 with $A = K$, $X = V$, $B^T = Z^T$ (i.e., $B = Z$),

$$(Z^T \otimes K)\ \mathrm{vec}(V) = \mathrm{vec}(KVZ). \tag{5}$$

Since V has only q nonzero entries, the product $G = VZ \in \mathbb{R}^{n\times r}$ is computed by the scatter operation

$$G_{i,\ell} = \sum_{\substack{s=1,\dots,q \\ i_s=i}} y_s\, Z_{j_s,\ell}, \qquad i = 1,\dots,n, \quad \ell = 1,\dots,r,$$

at cost $O(qr)$. Subsequently, KG costs $O(n^2r)$. This stage therefore costs $O(qr + n^2r)$.

Stage 3: Regularization. By Lemma 2, $\lambda(I_r \otimes K)\ \mathrm{vec}(W) = \lambda\ \mathrm{vec}(KW) = \lambda\ \mathrm{vec}(P)$. The matrix $P = KW$ was already computed in Stage 1, so no additional work is required.

Assembly. Combining Stages 1–3,

$$\mathcal{A}\ \mathrm{vec}(W) = \mathrm{vec}(KG + \lambda P). \tag{6}$$

The total cost is $O(qr + n^2r)$, dominated by the sparse inner products (4) and the dense matrix multiplications with K. At no point is an N-dimensional vector or $n \times M$ dense matrix formed. □

Remark 4. *The right-hand side $\mathbf{b} = \mathrm{vec}(KB)$ requires the MTTKRP $B = TZ \in \mathbb{R}^{n\times r}$, which costs $O(qr)$ since the mode-k unfolding T has exactly q nonzero entries* [5, Section 3.3]*, followed by a multiplication with K at cost $O(n^2r)$.*

4. Preconditioner

The diagonal matrix $\Omega = SS^T \in \mathbb{R}^{N\times N}$ has $\Omega_{ii} = 1$ if entry i is observed and $\Omega_{ii} = 0$ otherwise. For the purpose of constructing a spectrally faithful preconditioner with exploitable Kronecker structure, we replace Ω by the scalar matrix $\frac{q}{N} I_N$. This substitution is justified by the following elementary observation.

Lemma 5. *If the index set of observed entries is drawn uniformly at random with inclusion probability $p = q/N$ per entry, then $\mathbb{E}[\Omega] = p\, I_N$.*

Proof. Each diagonal entry Ω_{ii} is an independent Bernoulli random variable with parameter p. □

Replacing Ω by its expectation in (2), we define

$$\mathcal{P} = \frac{q}{N}\,(Z \otimes K)^T(Z \otimes K) + \lambda\,(I_r \otimes K) = \frac{q}{N}\,(Z^T Z \otimes K^2) + \lambda\,(I_r \otimes K), \tag{7}$$

where the second equality uses the mixed-product property of Kronecker products [4, Theorem 4.2.10] together with the symmetry $K^T = K$.

Proposition 6. *After a one-time setup of cost $O(n^3 + r^3)$, linear systems of the form $\mathcal{P}\ \mathrm{vec}(X) = \mathrm{vec}(F)$ can be solved in $O(n^2 r + n r^2)$ operations.*

Proof. Compute the eigendecompositions

$$K = U\Sigma U^T, \qquad Z^T Z = V \Lambda V^T,$$

where $U \in \mathbb{R}^{n\times n}$ and $V \in \mathbb{R}^{r\times r}$ are orthogonal, $\Sigma = \mathrm{diag}(\sigma_1, \ldots, \sigma_n)$, and $\Lambda = \mathrm{diag}(\mu_1, \ldots, \mu_r)$. These decompositions cost $O(n^3)$ and $O(r^3)$, respectively [2, Section 8.1]. Substituting into (7),

$$\mathcal{P} = (V \otimes U)\left[\frac{q}{N}\,(\Lambda \otimes \Sigma^2) + \lambda\,(I_r \otimes \Sigma)\right](V^T \otimes U^T).$$

The bracketed matrix is diagonal with entries

$$d_{j,i} = \frac{q}{N}\,\mu_j\,\sigma_i^2 + \lambda\,\sigma_i, \qquad j = 1, \ldots, r, \quad i = 1, \ldots, n. \tag{8}$$

Each $d_{j,i} > 0$ provided $K \succ 0$ and $\lambda > 0$, since $\sigma_i > 0$ and $\mu_j \geq 0$.

Given $F \in \mathbb{R}^{n\times r}$, solve $\mathcal{P}\ \mathrm{vec}(X) = \mathrm{vec}(F)$ as follows.

1. Compute $\hat{F} = U^T F V \in \mathbb{R}^{n\times r}$. By Lemma 2, this effects the change of basis $(V^T \otimes U^T)\ \mathrm{vec}(F) = \mathrm{vec}(U^T F V)$. *Cost:* $O(n^2 r + n r^2)$.
2. Set $\hat{X}_{i,j} = \hat{F}_{i,j}/d_{j,i}$ for all i, j. *Cost:* $O(nr)$.
3. Compute $X = U\hat{X}V^T$. *Cost:* $O(n^2 r + n r^2)$.

The total solve cost is $O(n^2 r + n r^2)$. □

5. Preconditioned conjugate gradient algorithm

We apply the standard PCG iteration [3, 2] to system (1) with preconditioner $\mathcal{P}$. The full procedure is recorded in Algorithm 1 for completeness.

6. Complexity analysis

Proposition 7. *Suppose $K \succ 0$ and $\lambda > 0$. Algorithm 1 has the following operation counts:*

(i) ***One-time preprocessing:*** *$O(n^3 + r^3)$ for the eigendecompositions of K and $Z^T Z$, plus $O(qr + n^2 r)$ for the MTTKRP and right-hand side.*

(ii) ***Per iteration:*** *$O(qr + n^2 r + n r^2)$ for one matrix–vector product with $\mathcal{A}$ (Proposition 3) and one preconditioner solve with $\mathcal{P}$ (Proposition 6), together with $O(nr)$ for vector updates and inner products.*

(iii) ***Total after t iterations:*** *$O(n^3 + r^3 + t(qr + n^2 r + n r^2))$.*

Under the assumption $n, r \ll q$, the dominant per-iteration cost simplifies to $O(qr)$. No computation of order N is performed.

Proof. Claim (i) follows from Proposition 6 (eigendecompositions) and Remark 4 (right-hand side). Claim (ii) combines Propositions 3 and 6 with the observation that each CG iteration requires exactly one product with $\mathcal{A}$, one solve with $\mathcal{P}$, and a constant number of $\mathbb{R}^{nr}$-vector operations [2, Algorithm 11.5.1]. Claim (iii) follows by summation. Since $n^2 r \leq qr$ and $nr^2 \leq qr$ when $n, r \leq q^{1/2}$ (which is implied by $n, r \ll q$), the per-iteration cost is $O(qr)$.

To verify the absence of $O(N)$ work: all multiplications with K or K^2 involve matrices of size $n \times n$; products with Z or $Z^T Z$ are either scatter operations over q entries or involve the $r \times r$

Algorithm 1 PCG for the RKHS mode-k subproblem (1)

Require: Kernel matrix K, Khatri–Rao product Z, MTTKRP B, observation indices $\{(i_s, j_s)\}_{s=1}^{q}$, regularization $\lambda > 0$, tolerance $\varepsilon > 0$
Ensure: Approximate solution $W \in \mathbb{R}^{n \times r}$

Precompute: $U, \Sigma \leftarrow \mathrm{eig}(K)$; $V, \Lambda \leftarrow \mathrm{eig}(Z^T Z)$; diagonal (8)
$\mathbf{b} \leftarrow \mathrm{vec}(KB)$
$W_0 \leftarrow 0$; $\mathbf{r}_0 \leftarrow \mathbf{b}$
$\mathbf{z}_0 \leftarrow \mathcal{P}^{-1}\mathbf{r}_0$ ▷ Proposition 6
$\mathbf{p}_0 \leftarrow \mathbf{z}_0$
for $t = 0, 1, 2, \ldots$ **do**
  $\mathbf{v} \leftarrow \mathcal{A}\,\mathbf{p}_t$ ▷ Proposition 3
  $\alpha_t \leftarrow \mathbf{r}_t^T \mathbf{z}_t \,/\, \mathbf{p}_t^T \mathbf{v}$
  $\mathbf{w}_{t+1} \leftarrow \mathbf{w}_t + \alpha_t\,\mathbf{p}_t$
  $\mathbf{r}_{t+1} \leftarrow \mathbf{r}_t - \alpha_t\,\mathbf{v}$
  if $\|\mathbf{r}_{t+1}\| < \varepsilon\,\|\mathbf{b}\|$ **then return** W_{t+1} (reshaped from $\mathbf{w}_{t+1}$)
  end if
  $\mathbf{z}_{t+1} \leftarrow \mathcal{P}^{-1}\mathbf{r}_{t+1}$
  $\beta_t \leftarrow \mathbf{r}_{t+1}^T \mathbf{z}_{t+1} \,/\, \mathbf{r}_t^T \mathbf{z}_t$
  $\mathbf{p}_{t+1} \leftarrow \mathbf{z}_{t+1} + \beta_t\,\mathbf{p}_t$
end for

matrix $Z^T Z$; and the selection operator S is applied only through pointwise evaluation at the q observed indices (4) or through the sparse scatter of Stage 2 in the proof of Proposition 3. □

Remark 8. *The convergence rate of PCG is governed by the condition number* $\kappa(\mathcal{P}^{-1}\mathcal{A})$ [2, Theorem 11.3.3]*: after* t *iterations,*

$$\|\mathbf{w}_t - \mathbf{w}_*\|_{\mathcal{A}} \le 2\left(\frac{\sqrt{\kappa(\mathcal{P}^{-1}\mathcal{A})} - 1}{\sqrt{\kappa(\mathcal{P}^{-1}\mathcal{A})} + 1}\right)^t \|\mathbf{w}_0 - \mathbf{w}_*\|_{\mathcal{A}}.$$

Since the preconditioner $\mathcal{P}$ *is obtained from* $\mathcal{A}$ *by replacing* $\Omega = SS^T$ *with* $\frac{q}{N} I_N$ *(Lemma 5),* $\kappa(\mathcal{P}^{-1}\mathcal{A})$ *is controlled by the deviation of the sampling operator from its expectation. When the observation pattern is close to uniform,* $\mathcal{P}^{-1}\mathcal{A}$ *is well-conditioned and a small number of iterations suffices.*

Remark 9. *Forming the* $nr \times nr$ *system matrix* $\mathcal{A}$ *explicitly requires* $O(qnr + n^2r^2)$ *operations, and solving via Cholesky factorization costs* $O(n^3r^3)$ [2, Section 4.2]*. Algorithm 1 is therefore advantageous whenever the number of PCG iterations* t *satisfies* $t \ll n^2r^2/q$*, a condition that is easily met in practice.*

References

[1] N. Aronszajn, *Theory of reproducing kernels*, Trans. Amer. Math. Soc. **68** (1950), 337–404.
[2] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 4th ed., Johns Hopkins Univ. Press, Baltimore, 2013.
[3] M. R. Hestenes and E. Stiefel, *Methods of conjugate gradients for solving linear systems*, J. Research Nat. Bur. Standards **49** (1952), no. 6, 409–436.
[4] R. A. Horn and C. R. Johnson, *Topics in Matrix Analysis*, Cambridge Univ. Press, Cambridge, 1991.
[5] T. G. Kolda and B. W. Bader, *Tensor decompositions and applications*, SIAM Rev. **51** (2009), no. 3, 455–500.