Review of "PCG for the RKHS CP-ALS mode-k subproblem (missing data)"

Summary

The document describes how to solve the $nr \times nr$ linear system arising from the mode-k subproblem of a CP decomposition with RKHS-constrained modes and missing data, using preconditioned conjugate gradients (PCG). The main contributions are: (i) an implicit matrix-vector product in $O(n^2r + qr)$ via gather/scatter, (ii) a Kronecker-structured preconditioner motivated by uniform sampling, and (iii) a complexity analysis showing the method avoids $O(N)$ computation and the $O(n^3r^3)$ cost of a dense solve.

Overall assessment: The core mathematical content is **correct**. The matvec procedure, the preconditioner diagonalization, and the complexity claims all check out, and are confirmed by the accompanying numerical scripts. The document is concise and largely well-written. There are several presentation issues and a few substantive gaps that should be addressed before the document can be considered complete.

Correctness verification

- **Kronecker-vec identity** (Eq. 2): $(Z \otimes K)\,\mathrm{vec}(X) = \mathrm{vec}(KXZ^\top)$ is a standard identity and is used correctly. Dimensions check: $Z \in \mathbb{R}^{M\times r}$, $K \in \mathbb{R}^{n\times n}$, $X \in \mathbb{R}^{n\times r}$, result $\in \mathbb{R}^{nM}$. ✓
- **Gather/scatter matvec** (Steps 1–4): The procedure correctly implements $Ax = (Z\otimes K)^\top P(Z\otimes K)x + \lambda(I_r \otimes K)x$ via (1) compute $G = KX$, (2) gather $u_t = G_{i_t,:} \cdot z_t$, (3) scatter/accumulate $H = \widetilde{U}Z$, (4) output $\mathrm{vec}(KH + \lambda G)$. The key observation that $(Z \otimes K)^\top \mathrm{vec}(\widetilde{U}) = \mathrm{vec}(K\widetilde{U}Z)$ is correct by the adjoint Kronecker identity. Verified numerically to machine precision. ✓
- **Preconditioner diagonalization** (Section 3): The formula $A_0 = \alpha(Z^\top Z) \otimes K^2 + \lambda(I_r \otimes K)$ diagonalizes in the $(V \otimes U)$ basis with eigenvalues $\alpha\sigma_a\lambda_b^2 + \lambda\lambda_b$. Verified numerically. ✓
- **SPD argument** (Section 1): The claim that $A \succ 0$ when $K \succ 0$ and $\lambda > 0$ is correct. The first term is PSD (being of the form $C^\top DC$ with $D \succeq 0$) and the second is PD. ✓
- **RHS computation**: The sparse MTTKRP $B_{i_t,:} \mathrel{+}= t_t z_t$ correctly computes $B = TZ$ since T has zeros at unobserved positions. Verified numerically. ✓
- **Complexity claims**: All operation counts are correct: matvec $O(n^2r + qr)$, preconditioner $O(n^2r + nr^2)$, RHS $O(n^2r + qr)$. ✓

Issues and suggestions

1. **Variable name collision (G).** The symbol G is used for two different objects:
   - In Section 2, Step 1: $G \leftarrow KX$ (temporary in the matvec algorithm).
   - In Section 3: $G \equiv Z^\top Z$ (the Khatri–Rao Gram matrix).

This collision is confusing and should be resolved. For instance, rename the matvec temporary to Γ or $\widehat{G}$, or rename the Gram matrix to Φ (as is common in CP-ALS literature, where it is often called the "Gamma" matrix).

2. Notation t_t for observed values. In the RHS paragraph, "for each observed value t_t at (i_t, j_t)" reuses the letter t for both the index and the value, creating confusion with the tensor T and the loop index. A different symbol (e.g., v_t, τ_t, or y_t) would be clearer.

3. Scatter description (line after Eq. 5). The text says "Su is the sparse $n \times M$ matrix with nonzeros u_t at (i_t, j_t)." Strictly, $Su \in \mathbb{R}^N$ is a *vector*; it becomes an $n \times M$ matrix after reshaping via vec^{-1}. This distinction matters because the document is otherwise careful about the vec operator.

4. Missing convergence / iteration count analysis. The PCG iteration count m is left unspecified with the remark "typically $m \ll nr$ with a good preconditioner." For a complete complexity comparison with the $O(n^3 r^3)$ dense solve, a bound on m (or at least a condition number estimate for $A_0^{-1}A$) is needed. Without such a bound the claimed complexity advantage is informal.

Concretely: the preconditioner is motivated by the approximation $P \approx \alpha I$, but the quality of this approximation depends on the sampling pattern. For highly non-uniform or very sparse observation masks, A_0 may be a poor approximation of A, leading to large m. A brief discussion of when the preconditioner is effective (e.g., $m = O(1)$ under near-uniform sampling) or a reference to relevant literature on incomplete-data preconditioning would strengthen the contribution.

5. Preconditioner choice of α. The document sets $\alpha = q/N$ with the motivation $\mathbb{E}[P] = (q/N)I$ under uniform sampling. In practice, the observations may not be uniformly distributed. It would be useful to briefly discuss: (a) whether other choices of α are preferable (e.g., $\alpha = \|P(Z \otimes K)\|_F^2 / \|Z \otimes K\|_F^2$), and (b) sensitivity of PCG iteration count to the choice of α.

6. One-time setup costs. The eigendecompositions of K ($O(n^3)$) and G ($O(r^3)$) are described as one-time costs but are not included in the total complexity formula in Section 4. Since these are amortized over all PCG iterations (and potentially over outer ALS iterations), this is acceptable, but should be stated explicitly to avoid confusion.

7. Block-diagonal alternative (last sentence of Section 3). The "cheaper alternative" of replacing G by $\text{diag}(G)$ is mentioned in passing but not developed. Since this decouples the r columns and reduces the preconditioner application to r independent $n \times n$ solves, it could be a practical default. Either develop this or remove the mention to avoid raising unanswered questions.

8. Minor: Hadamard product notation. The Hadamard product is denoted $*$ (with $*$ in the displayed equation), which could be confused with convolution or the $*$-algebra operation. The more standard notations $\odot$ or $\circ$ would be preferable. (Note that $\odot$ is already used for the Khatri–Rao product in the problem statement, but the document could clarify the dual usage or adopt $\circ$.)

Questions for the authors

1. Can you provide a condition number bound for $A_0^{-1}A$ under reasonable assumptions on the sampling pattern? Even an asymptotic estimate as $q/N \to \alpha$ would be informative.

2. For the case K only PSD (not PD), you suggest adding a nugget εI. Have you considered instead working in the range of K directly (reducing to mr unknowns where $m = \text{rank}(K)$)? This is briefly mentioned but not developed.

3. How does the PCG iteration count scale empirically with n, r, and q/N? Even a small table or figure from the existing numerical experiments would significantly strengthen the practical claims.

Verdict

The mathematical content is **correct** and the algorithmic contribution is sound. The main weakness is the lack of convergence analysis (issue 4 above). I would characterize this as a **minor revision**: the issues identified are presentation and completeness concerns, not correctness problems.