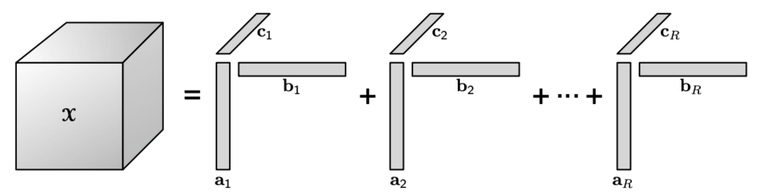
Error Analysis of Pairwise Perturbation for Tensor Decomposition

Linjian Ma and Edgar Solomonik

Department of Computer Science, University of Illinois at Urbana-Champaign

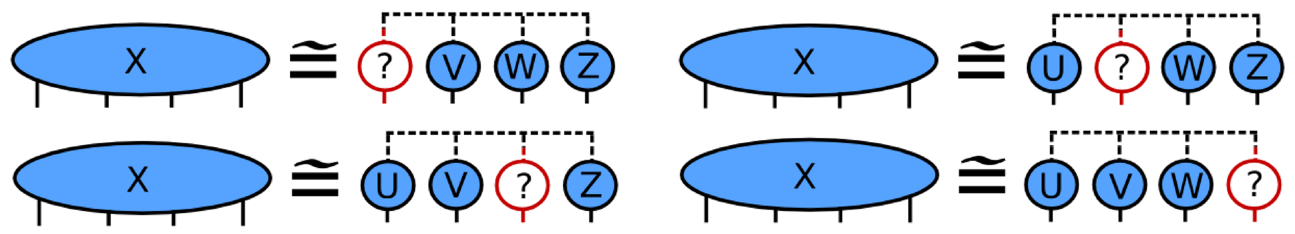
CP Decomposition with ALS



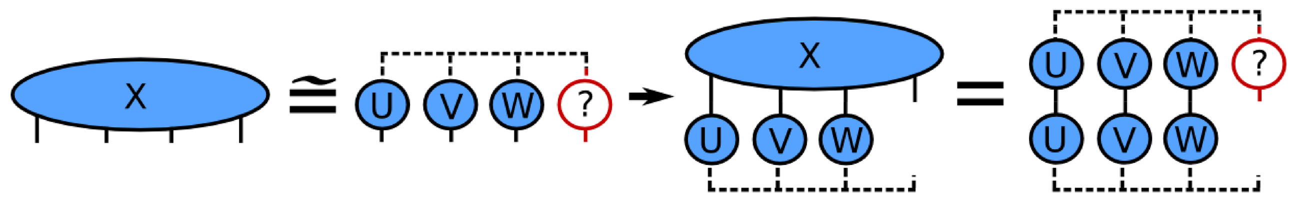
Consider rank R CP decomposition of an $s \times s \times s \times s$ tensor

$$x_{ijkl} \approx \sum_{r=1}^{R} u_{ir} v_{jr} w_{kr} z_{lr}$$

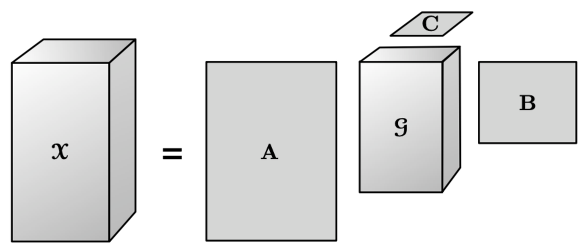
ALS updates factor matrices in an alternating manner



Each quadratic subproblem is typically solved via normal equations

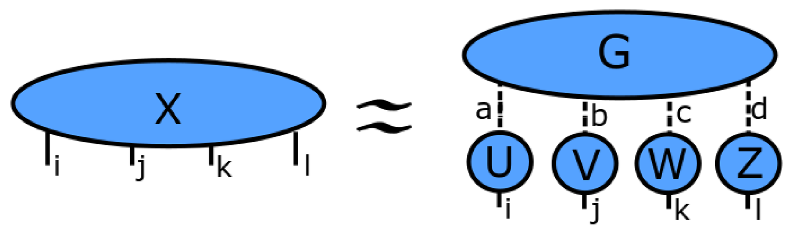


Tucker Decomposition with ALS

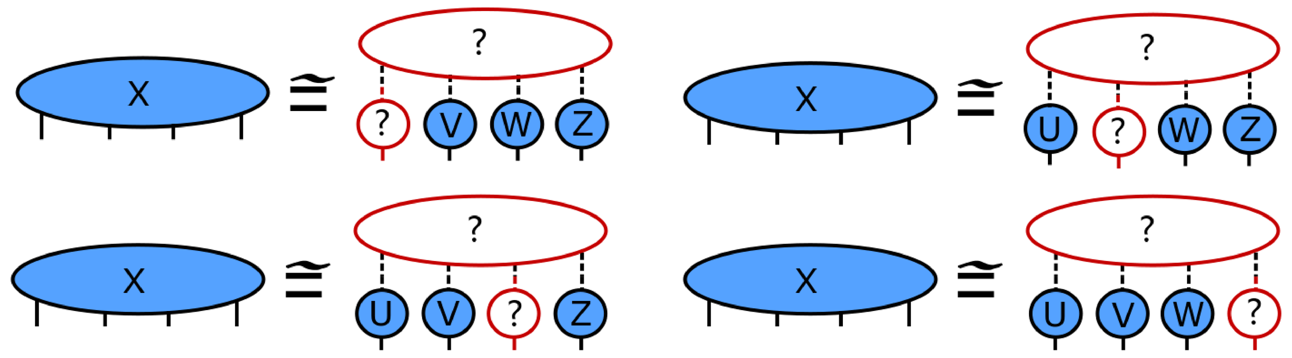


Consider rank R Tucker decomposition of an $s \times s \times s \times s$ tensor

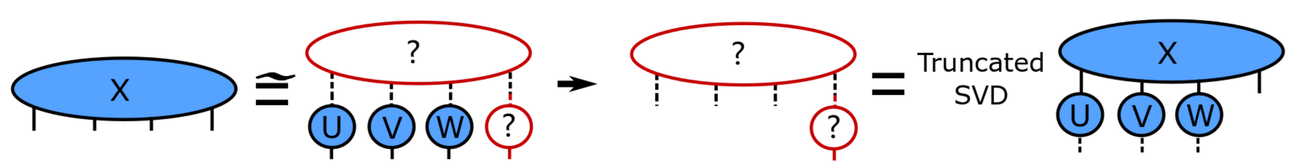
$$x_{ijkl} \approx \sum_{a,b,c,d} g_{abcd} u_{ia} v_{jb} w_{kc} z_{ld}$$



ALS updates factor matrices in an alternating manner



Tucker-ALS is usually solved with HOOI (Higher-Order Orthogonal Iteration)

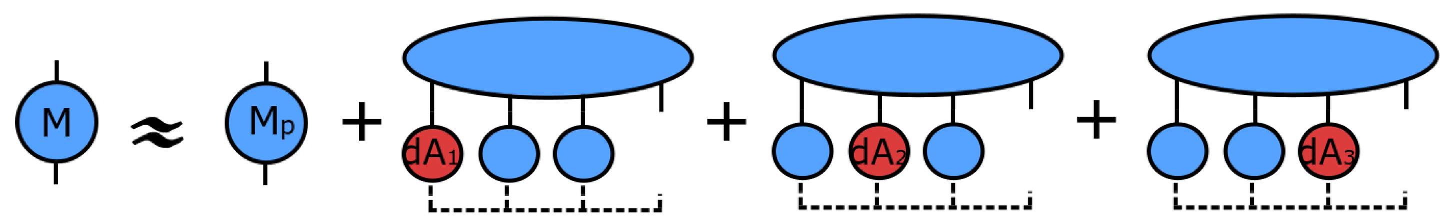


Pairwise Perturbation

For CP, Pairwise perturbation (PP) approximates MTTKRP result $M^{(n)} \approx \tilde{M}^{(n)}$ using pairwise perturbation operators $\mathcal{M}_p^{(i,n)}$

- Write $A^{(n)} = A_p^{(n)} + dA^{(n)} \rightarrow M^{(n)} = X_{(n)} \bigodot_{i=1,i\neq n}^{N} (A_p^{(i)} + dA^{(i)})$
- Element-wise,

$$M^{(n)}(y,k) = M_p^{(n)}(y,k) + \sum_{i=1,i\neq n}^{N} \sum_{x=1}^{s_i} \mathcal{M}_p^{(i,n)}(x,y,k) dA^{(i)}(x,k) + \sum_{i=1,i\neq n}^{N} \sum_{j=i+1,j\neq n}^{N} \sum_{x=1}^{s_i} \sum_{z=1}^{s_j} \mathcal{M}_p^{(i,j,n)}(x,z,y,k) dA^{(i)}(x,k) dA^{(j)}(z,k) + \cdots$$



	DT ALS	PP initialization step	PP approximate step
CP	$4s^N R$	$6s^N R$	$2N^2 s^2 R$
Tucker	$4s^N R$	$6s^N R$	$2N^2 s^2 R^{N-1}$

Error Analysis for Tucker

Consider order $N = 3$ tensor $\boldsymbol{\mathcal{X}}$, let $\boldsymbol{\mathcal{Y}}^{(3)}$ be the HOOI result needed to form the third factor matrix $A^{(3)}$

- Bound relative error of $\tilde{\boldsymbol{\mathcal{Y}}}^{(3)}$ computed by PP middle step
- The ith factor matrix changed by $dA^{(i)}$ since the first step of PP
- The spectral norm of the tensor corresponds to $||\boldsymbol{\mathcal{X}}||_2 = \sup\{||f_{\boldsymbol{\mathcal{X}}}||_2\}$, where $f_{\boldsymbol{\mathcal{X}}} \in \mathbb{R}^s \times \mathbb{R}^s \rightarrow \mathbb{R}^s$,

$$z = f_{\boldsymbol{\mathcal{X}}}(u,v) \Rightarrow z_k = \sum_{i,j} x_{ijk} u_i v_j$$

***Theorem* 0.1** (Error Bound with Bounded Residual). *If $||dA^{(l)}||_2 \leq \epsilon \ll 1$ for $l \in \{1,2,3\}$ and residual spectral norm $\leq \frac{1}{3}||\boldsymbol{\mathcal{X}}||_2$,*

$$\frac{||\tilde{\boldsymbol{\mathcal{Y}}}^{(3)} - \boldsymbol{\mathcal{Y}}^{(3)}||_2}{||\boldsymbol{\mathcal{Y}}^{(3)}||_2} = O(\epsilon^2).$$

***Theorem* 0.2** (Error Bound when Tucker starts with interlaced HOSVD). *If $||dA^{(l)}||_F \leq \epsilon \ll 1$ for $l \in \{1,2,3\}$ and*
1. interlaced HOSVD is used to initialize Tucker-ALS
2. the decomposition residual is no higher than that attained by HOSVD

$$\frac{||\tilde{\boldsymbol{\mathcal{Y}}}^{(n)} - \boldsymbol{\mathcal{Y}}^{(n)}||_F}{||\boldsymbol{\mathcal{Y}}^{(n)}||_F} = O\Big(\epsilon^2 \Big(\frac{s}{R}\Big)^{N/2}\Big).$$

The error bound is independent of the input tensor conditioning

Error Analysis for CP

Consider order $N = 3$ tensor $\boldsymbol{\mathcal{X}}$, let $M^{(3)}$ be the MTTKRP result needed to form the third factor matrix $A^{(3)}$

- Bound columnwise error of $\tilde{M}^{(3)}$ computed by PP middle step
- The ith factor matrix changed by $dA^{(i)}$ since the first step of PP
- Error bound based on conditioning bound of $f_{\boldsymbol{\mathcal{X}}}$

***Theorem* 0.3** (Columnwise Error Bound from Tensor Conditioning). *If $||da_k^{(l)}||_2/||a_k^{(l)}||_2 \leq \epsilon$ for $l \in \{1,2,3\}$,*

$$\frac{||\tilde{m}_k^{(3)} - m_k^{(3)}||_2}{||m_k^{(3)}||_2} \leq \frac{\max_{u,v\in\mathbb{S}^{s-1}} ||f_{\boldsymbol{\mathcal{X}}}(u,v)||_2}{\min_{y,z\in\mathbb{S}^{s-1}} ||f_{\boldsymbol{\mathcal{X}}}(y,z)||_2} O(\epsilon^2).$$

- If $\min_{u,v\in\mathbb{S}^{s-1}} ||f_{\boldsymbol{\mathcal{X}}}(u,v)||_2 = 0$, bound is trivial
- There exist $2 \times 2 \times 2$, $4 \times 4 \times 4$, and $8 \times 8 \times 8$ tensors for which $||f_{\boldsymbol{\mathcal{X}}}(u,v)||_2 = 1$ for all $u, v \in \mathbb{S}^{s-1}$
- However, for any $s \notin \{1,2,4,8\}$, any $s \times s \times s$ tensor $\boldsymbol{\mathcal{X}}$ has $\min_{u,v\in\mathbb{S}^{s-1}} ||f_{\boldsymbol{\mathcal{X}}}(u,v)||_2 = 0$

We can bound columnwise error of *approximate update* $\tilde{H}^{(1,3)}$ to $\tilde{M}^{(3)}$ computed by PP middle step due to change in $A^{(1)}$

- For simplicity, assume $N = 3$ and $R = 1$, so that

$$\boldsymbol{\mathcal{X}} \approx a^{(1)} \circ a^{(2)} \circ a^{(3)}$$

- Updating $a^{(n)}$ by $\delta a^{(n)}$ yields update $h^{(m,n)}$ to mth factor matrix
- $h^{(1,3)} = R^{(3)} a^{(2)}$ where $R^{(3)} \in \mathbb{R}^{s\times s}$ is $\boldsymbol{\mathcal{T}}$ contracted along the last mode with with $\delta a^{(3)}$

***Theorem* 0.4** (Columnwise Error Bound from Matricization Conditioning). *PP approximate step performs update $\tilde{h}^{(1,3)}$ with relative error*

$$\frac{||h^{(1,3)} - \tilde{h}^{(1,3)}||_2}{||h^{(1,3)}||_2} \leq \kappa(R^{(3)}) ||da^{(2)}||_2,$$

where $da^{(2)}$ is the change to $a^{(2)}$ since PP initialization

References

1. L. Ma and E. Solomonik. Accelerating alternating least squares for tensor decomposition by pairwise perturbation. arXiv preprint arXiv:1811.10573, 2018.
2. T. G. Kolda and B. W. Bader. Tensor decompositions and applications. SIAM review, 51(3):455–500, 2009.
3. https://github.com/cyclops-community/ctf