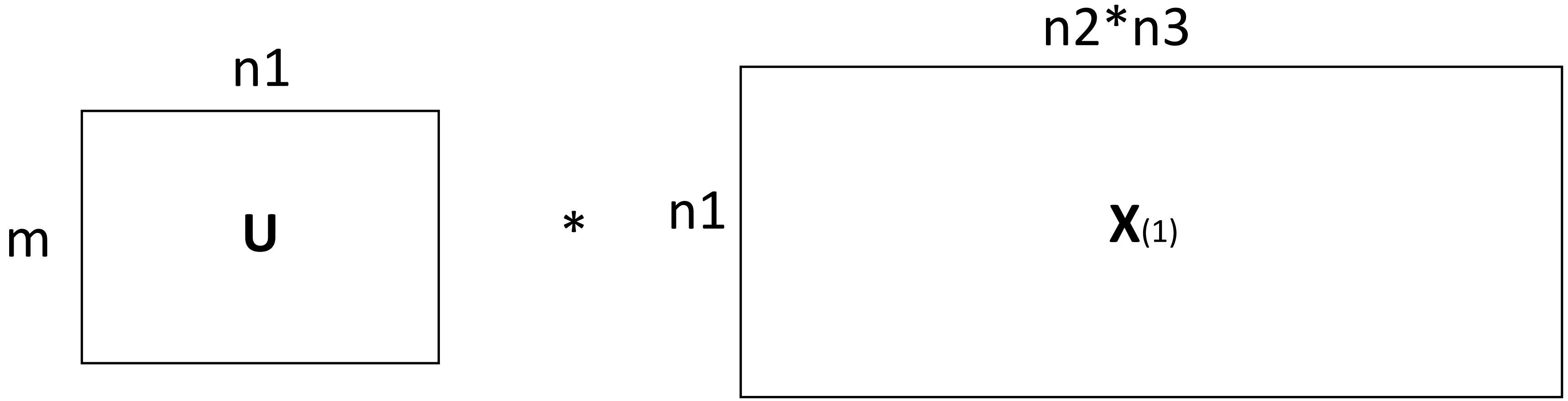
A randomized algorithm for a tensor-based generalization of the singular value decomposition

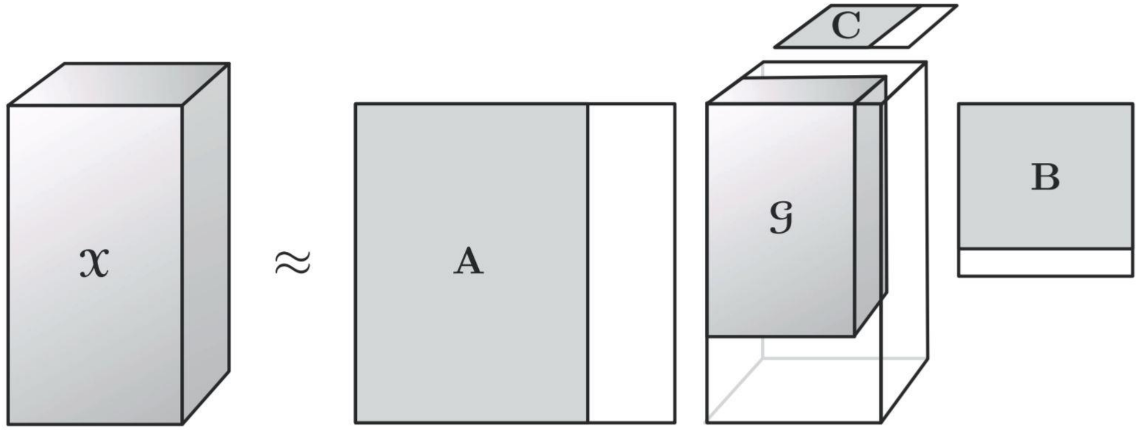
Review of tensor algebra

- n-mode (matrix) product of a tensor

$$\mathcal{Y} = \mathcal{X} \times_n \mathbf{U} \quad \Leftrightarrow \quad \mathbf{Y}_{(n)} = \mathbf{U}\mathbf{X}_{(n)}.$$



Tucker form



$$\mathcal{X} = \mathcal{G} \times_1 \mathbf{A}^{(1)} \times_2 \mathbf{A}^{(2)} \cdots \times_N \mathbf{A}^{(N)} = [\![\mathcal{G} \,; \mathbf{A}^{(1)}, \mathbf{A}^{(2)}, \ldots, \mathbf{A}^{(N)}]\!]$$

procedure HOSVD($\mathcal{X}$,R_1,$R_2,\ldots,R_N$)
  for $n = 1,\ldots,N$ **do**
    $\mathbf{A}^{(n)} \leftarrow R_n$ leading left singular vectors of $\mathbf{X}_{(n)}$
  end for
  $\mathcal{G} \leftarrow \mathcal{X} \times_1 \mathbf{A}^{(1)\mathsf{T}} \times_2 \mathbf{A}^{(2)\mathsf{T}} \cdots \times_N \mathbf{A}^{(N)\mathsf{T}}$
  return $\mathcal{G}, \mathbf{A}^{(1)}, \mathbf{A}^{(2)}, \ldots, \mathbf{A}^{(N)}$
end procedure

TensorSVD Algorithm

TENSORSVD Algorithm

Data : tensor $\mathcal{A} \in \mathbb{R}^{n_1 \times \cdots \times n_d}$, $k_i, 1 \leq k_i \leq n_i, i = 1, \ldots, d$.

Result : matrices $U_{[i],k_i} \in \mathbb{R}^{n_i \times k_i}$ for all $i = 1, \ldots, d$ such that

$$\mathcal{A} \approx \mathcal{A} \times_1 U_{[1],k_1} U_{[1],k_1}^T \times_2 U_{[2],k_2} U_{[2],k_2}^T \times_3 \cdots \times_d U_{[d],k_d} U_{[d],k_d}^T.$$

for $i = 1, \ldots, d$ **do**

  Compute the top k_i left singular vectors of $A_{[i]}$ and denote them by $U_{[i],k_i}$;

end

Error bound:

$$\mathscr{E} = \mathscr{A} - \mathscr{A} \times_1 U_{[1],k_1} U_{[1],k_1}^{\mathrm{T}} \times_2 \cdots \times_d U_{[d],k_d} U_{[d],k_d}^{\mathrm{T}} \leqslant \sum_{i=1}^{d} \| A_{[i]} - (A_{[i]})_{k_i} \|_{\mathrm{F}}.$$

Approximate tensor SVD

ApproxTensorSVD Algorithm

Data : tensor $\mathcal{A} \in \mathbb{R}^{n_1 \times \cdots \times n_d}$, $c_i, 1 \leq c_i \leq \prod_{j=1, j \neq i}^{d} n_j, i = 1, \ldots, d$.

Result : matrices $C_{[i]} \in \mathbb{R}^{n_i \times c_i}$ for all $i = 1, \ldots, d$ such that

$$\mathcal{A} \approx \mathcal{A} \times_1 C_{[1]} C_{[1]}^{+} \times_2 C_{[2]} C_{[2]}^{+} \times_3 \cdots \times_d C_{[d]} C_{[d]}^{+}.$$

for $i = 1, \ldots, d$ **do**

Form $C_{[i]}$ by calling

- the SelectColumnsSinglePass algorithm with input $A_{[i]}$ and c_i, **or**
- the SelectColumnsMultiPass algorithm with input $A_{[i]}$, c_i, and t ;

end

Randomized column selection algorithms

- If A is well approximated by a low-rank matrix, we would like

$$A \approx P_{\mathrm{span}(C)} A.$$

- Two randomized column selection algorithms

1. SelectColumnsSinglePass algorithm

"Fast Monte Carlo algorithms for matrices I: Approximating matrix multiplication." *SIAM Journal on Computing* (2006).

2. SelectColumnsMultiPass algorithm

"Matrix approximation and projective clustering via iterative sampling." (2005).

SELECTCOLUMNSSINGLEPASS Algorithm

Data : $A \in \mathbb{R}^{m\times n}$, $c \in \mathbb{Z}^+$ s.t. $1 \leq c \leq n$.

Result : $C \in \mathbb{R}^{m\times c}$, s.t. $CC^+A \approx A$.

Compute (for some positive $\beta \leq 1$) probabilities $\{p_i\}_{i=1}^n$ s.t.

$$p_i \geq \beta \left|A^{(i)}\right|^2 / \|A\|_F^2 \,,$$

where $A^{(i)}$ is the i-th column of A as a column vector.

$S = \{\}$;

for $t = 1, \ldots, c$ **do**

  Pick $i_t \in \{1, \ldots, n\}$ with $\mathbf{Pr}[i_t = \alpha] = p_\alpha$;

  $S = S \cup \{i_t\}$;

end

$C = A_S$;

Theorem 2. *Suppose* $A \in \mathbb{R}^{m\times n}$, *and let* C *be the* $m \times c$ *matrix constructed by sampling* c *columns of* A *with the* SELECTCOLUMNSSINGLEPASS *algorithm. If* $\eta = 1 + \sqrt{(8/\beta)\log(1/\delta)}$ *for any* $0 < \delta < 1$, *then, with probability at least* $1 - \delta$,

$$\|A - CC^+A\|_{\mathrm{F}}^2 \leqslant \|A - A_k\|_{\mathrm{F}}^2 + \epsilon\|A\|_{\mathrm{F}}^2,$$

if $c \geqslant 4\eta^2 k/(\beta\epsilon^2)$.

SELECTCOLUMNSMULTIPASS Algorithm

```
Data   : A ∈ ℝ^{m×n}, c ∈ ℤ^+ s.t. 1 ≤ c ≤ n, t ∈ ℤ^+.
Result : C ∈ ℝ^{m×tc}, s.t. CC^+A ≈ A.
S = {};
for ℓ = 1,...,t do
    if ℓ == 1 then
        E_1 = A;
    else
        E_ℓ = A − A_S A_S^+ A;
    end
    Compute (for some positive β ≤ 1) probabilities {p_i}_{i=1}^n s.t.
        p_i ≥ β |E_ℓ^{(i)}|^2 / ||E_ℓ||_F^2 ,
    where E_ℓ^{(i)} is the i-th column of E_ℓ as a column vector.
    for t = 1,...,c do
        Pick i_t ∈ {1,...,n} with Pr[i_t = α] = p_α;
        S = S ∪ {i_t};
    end
end
C = A_S;
```

Error bound

Theorem 3. *Suppose $A \in \mathbb{R}^{m \times n}$ and let C be the $m \times tc$ matrix constructed by sampling c columns of A in each of t rounds with the* SelectColumnsMultiPass *algorithm. If $\eta = 1 + \sqrt{(8/\beta)\log(1/\delta)}$ for any $0 < \delta < 1$, then, with probability at least $1 - t\delta$,*

$$\|A - CC^{+}A\|_{\mathrm{F}}^{2} \leqslant \frac{1}{1-\epsilon}\|A - A_k\|_{\mathrm{F}}^{2} + \epsilon^{t}\|A\|_{\mathrm{F}}^{2},$$

if $c \geqslant 4\eta^2 k/(\beta\epsilon^2)$ columns are picked in each of the t rounds.

- Proof

Let $C^1 = A_{S_1}$ we have with probability at least $1 - \delta$

$$\|A - C^1(C^1)^{+}A\|_{\mathrm{F}}^{2} \leqslant \frac{1}{1-\epsilon}\|A - A_k\|_{\mathrm{F}}^{2} + \epsilon\|A\|_{\mathrm{F}}^{2}$$

Let $(S_1, \ldots, S_{t-1})$ denote the set of columns picked in the first t-1 rounds.

Let $C^{t-1} = A_{(S_1,\ldots,S_{t-1})}$

Proof continue

Assume that by choosing $c \geqslant 4\eta^2 k/(\beta\epsilon^2)$ from the first t-1 rounds we have

$$\|A - C^{t-1}(C^{t-1})^+ A\|_{\mathrm{F}}^2 \leqslant \frac{1}{1-\epsilon}\|A - A_k\|_{\mathrm{F}}^2 + \epsilon^{t-1}\|A\|_{\mathrm{F}}^2$$

Holds with probability as least $1-(t-1)\delta$.

Define $E_t = A - C^{t-1}(C^{t-1})^+ A$ and let Z be a matrix that are included in the sample E_t . Then with probability at least $1-\delta$

$$\|E_t - ZZ^+ E_t\|_{\mathrm{F}}^2 \leqslant \|E_t - (E_t)_k\|_{\mathrm{F}}^2 + \epsilon\|E_t\|_{\mathrm{F}}^2$$

$$\Longrightarrow \|E_t - ZZ^+ E_t\|_{\mathrm{F}}^2 \leqslant \|E_t - (E_t)_k\|_{\mathrm{F}}^2 + \frac{\epsilon}{1-\epsilon}\|A - A_k\|_{\mathrm{F}}^2 + \epsilon^t\|A\|_{\mathrm{F}}^2$$

Holds with probability at least $1-t\delta$.

Proof continue

Since $E_t - ZZ^+E_t = A - C^t(C^t)^+A$ and

$$\begin{aligned}\|E_t - (E_t)_k\|_F^2 &= \|(I - C^{t-1}(C^{t-1})^+)A - ((I - C^{t-1}(C^{t-1})^+)A)_k\|_F^2 \\ &\leqslant \|(I - C^{t-1}(C^{t-1})^+)A - (I - C^{t-1}(C^{t-1})^+)A_k\|_F^2 \\ &\leqslant \|(I - C^{t-1}(C^{t-1})^+)(A - A_k)\|_F^2 \\ &\leqslant \|A - A_k\|_F^2.\end{aligned}$$

Thus, we probability at least $1 - t\delta$

$$\|A - CC^+A\|_F^2 \leqslant \frac{1}{1-\epsilon}\|A - A_k\|_F^2 + \epsilon^t\|A\|_F^2$$

Error bound of ApproxTensorSVD

Theorem 1. *Let* $\mathscr{A} \in \mathbb{R}^{n_1 \times \ldots \times n_d}$ *be a* d*-mode tensor, let* $\beta \in (0, 1]$*, and let* $\eta = 1 + \sqrt{(8/\beta)\log(1/\delta)}$*, for any* $0 < \delta < 1$*. Let matrices* $C_{[i]}, i \in \{1, \ldots, d\}$*, be computed by the* ApproxTensorSVD *algorithm (Algorithm 1).*

- If the columns are chosen with the SelectColumnsSinglePass algorithm, then with probability at least $1 - d\delta$

$$\|\mathscr{A} - \mathscr{A} \times_1 C_{[1]}C_{[1]}^{+} \times_2 \ldots \times_d C_{[d]}C_{[d]}^{+}\|_F \leqslant \sum_{i=1}^{d} \|A_{[i]} - (A_{[i]})_{k_i}\|_{\mathrm{F}} + d\epsilon\|\mathscr{A}\|_{\mathrm{F}},$$

if $c_i \geqslant 4\eta^2 k_i/(\beta\epsilon^2)$ *for all* $i = 1, \ldots, d$.

- If the columns are chosen with the SelectColumnsMultiPass algorithm, then with probability at least $1 - td\delta$

$$\|\mathscr{A} - \mathscr{A} \times_1 C_{[1]}C_{[1]}^{+} \times_2 \cdots \times_d C_{[d]}C_{[d]}^{+}\|_{\mathrm{F}} \leqslant \frac{1}{1-\epsilon} \sum_{i=1}^{d} \|A_{[i]} - (A_{[i]})_{k_i}\|_{\mathrm{F}} + d\epsilon^t \|\mathscr{A}\|_{\mathrm{F}}$$

if $c_i \geqslant 4\eta^2 k_i/(\beta\epsilon^2)$ *for every one of the* t *rounds and for all* $i = 1, \ldots, d$.

Proof

- Define $\widetilde{\mathscr{A}} = \mathscr{A} \times_1 C_{[1]}C_{[1]}^{+} \times_2 \cdots \times_d C_{[d]}C_{[d]}^{+}$

  Let $\widetilde{\mathscr{E}}_d = \widetilde{\mathscr{E}} = \mathscr{A} - \widetilde{\mathscr{A}}$ and $\widetilde{\Pi}_i = C_{[i]}C_{[i]}^{+}$ we have

$$
\begin{aligned}
\|\widetilde{\mathscr{E}}_d\|_{\mathrm{F}} &= \|\mathscr{A} - \mathscr{A} \times_d \widetilde{\Pi}_d + \mathscr{A} \times_d \widetilde{\Pi}_d - \mathscr{A} \times_1 \widetilde{\Pi}_1 \times_2 \cdots \times_d \widetilde{\Pi}_d\|_{\mathrm{F}} \\
&\leqslant \|\mathscr{A} - \mathscr{A} \times_d \widetilde{\Pi}_d\|_{\mathrm{F}} + \|(\mathscr{A} - \mathscr{A} \times_1 \widetilde{\Pi}_1 \times_2 \cdots \times_{d-1} \widetilde{\Pi}_{d-1}) \times_d \widetilde{\Pi}_d\|_{\mathrm{F}} \\
&\leqslant \|\mathscr{A} - \mathscr{A} \times_d \widetilde{\Pi}_d\|_{\mathrm{F}} + \|\mathscr{A} - \mathscr{A} \times_1 \widetilde{\Pi}_1 \times_2 \cdots \times_{d-1} \widetilde{\Pi}_{d-1}\|_{\mathrm{F}},
\end{aligned}
$$

Let $\widetilde{\mathscr{E}}_{d-1} = \mathscr{A} - \mathscr{A} \times_1 \widetilde{\Pi}_1 \times_2 \cdots \times_{d-1} \widetilde{\Pi}_{d-1}$, in the same manner

$$
\|\widetilde{\mathscr{E}}_{d-1}\|_{\mathrm{F}} \leqslant \|\mathscr{A} - \mathscr{A} \times_{d-1} \widetilde{\Pi}_{d-1}\|_{\mathrm{F}} + \|\mathscr{A} - \mathscr{A} \times_1 \widetilde{\Pi}_1 \times_2 \cdots \times_{d-2} \widetilde{\Pi}_{d-2}\|_{\mathrm{F}}
$$

Finally, we have

$$
\|\widetilde{\mathscr{E}}\|_{\mathrm{F}} \leqslant \sum_{i=1}^{d} \|\mathscr{A} - \mathscr{A} \times_i \widetilde{\Pi}_i\|_{\mathrm{F}}
$$

Restricting the approximation to matrices

$$\widetilde{A} = CC^{+}AR^{+}R = CUR$$

- If the columns and rows are chosen with the SelectColumnsSinglePass algorithm then, with probability at least $1 - 2\delta$

$$\|A - CC^{+}AR^{+}R\|_{\mathrm{F}} \leqslant 2\|A - A_k\|_{\mathrm{F}} + 2\epsilon\|A\|_{\mathrm{F}} \quad \textit{if } c, r \geqslant 4\eta^2 k/(\beta\epsilon^2).$$

- If the columns and rows are chosen with the SelectColumnsMultiPass algorithm then, with probability at least $1 - 2t\delta$

$$\|A - CC^{+}AR^{+}R\|_{\mathrm{F}} \leqslant \frac{2}{1-\epsilon}\|A - A_k\|_{\mathrm{F}} + 2\epsilon^t\|A\|_{\mathrm{F}}$$

$$\textit{if } c, r \geqslant 4\eta^2 k/(\beta\epsilon^2) \textit{ in each of the } t \textit{ passes.}$$