CSE 392: Matrix and Tensor Algorithms for Data

Instructor: Shashanka Ubaru

University of Texas, Austin
Spring 2024

Lecture 18: Randomized CP - II

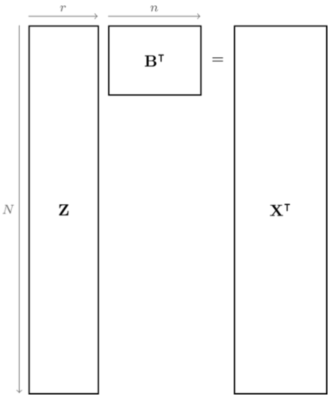
Outline

1. CP-ARLS-Mix

2. CP-ARLS-Lev

Kronecker FJLTs

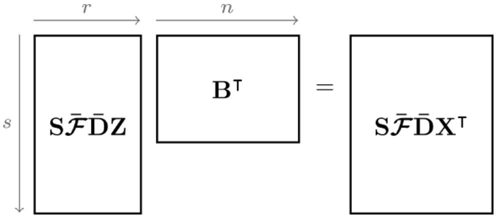
$$\min_{\boldsymbol{B}} \|\boldsymbol{Z}\boldsymbol{B}^\top - \boldsymbol{X}^\top\|_F^2$$



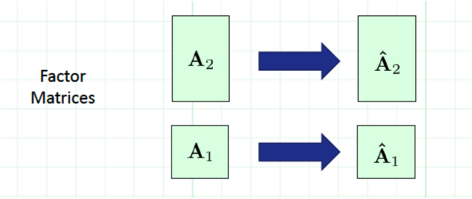
$$\boldsymbol{Z} = \boldsymbol{A}_d \odot \cdots \odot \boldsymbol{A}_{k+1} \odot \boldsymbol{A}_{k-1} \odot \cdots \odot \boldsymbol{A}_1$$

$$\min_{\boldsymbol{B}} \|\boldsymbol{S}\bar{\mathcal{F}}\bar{\boldsymbol{D}}\boldsymbol{Z}\boldsymbol{B}^\top - \boldsymbol{S}\bar{\mathcal{F}}\bar{\boldsymbol{D}}\boldsymbol{X}^\top\|_F^2$$

- $\boldsymbol{S}$ is $s \times N$ sampling matrix
- $\bar{\mathcal{F}} = \mathcal{F}_d \otimes \cdots \otimes \mathcal{F}_{k+1} \otimes \mathcal{F}_{k-1} \otimes \cdots \otimes \mathcal{F}_1$.
- $\bar{\boldsymbol{D}} = \boldsymbol{D}_d \otimes \cdots \otimes \boldsymbol{D}_{k+1} \otimes \boldsymbol{D}_{k-1} \otimes \cdots \otimes \boldsymbol{D}_1$.



Mixing KRP Efficiently Using Kronecker FJLT



$$\begin{aligned}
\boldsymbol{S}\bar{\mathcal{F}}\bar{\boldsymbol{D}}\boldsymbol{Z} &= \boldsymbol{S}(\mathcal{F}_2 \otimes \mathcal{F}_1)(\boldsymbol{D}_2 \otimes \boldsymbol{D}_1)(\boldsymbol{A}_2 \odot \boldsymbol{A}_1) \\
&= \boldsymbol{S}\left((\mathcal{F}_2\boldsymbol{D}_2) \otimes (\mathcal{F}_1\boldsymbol{D}_1)\right)(\boldsymbol{A}_2 \odot \boldsymbol{A}_1) \\
&= \boldsymbol{S}\left((\mathcal{F}_2\boldsymbol{D}_2\boldsymbol{A}_2) \odot (\mathcal{F}_1\boldsymbol{D}_1\boldsymbol{A}_1)\right) \\
&= \boldsymbol{S}(\hat{\boldsymbol{A}}_2 \odot \hat{\boldsymbol{A}}_1)
\end{aligned}$$

Pre-Mixing Tensor

Need to compute sketched right hand side...

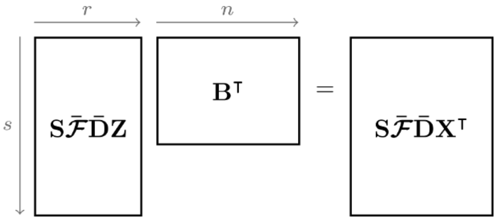
$$\boldsymbol{S}\bar{\mathcal{F}}\bar{\boldsymbol{D}}\boldsymbol{X}^\top = \boldsymbol{S}(\mathcal{F}_2 \otimes \mathcal{F}_1)(\boldsymbol{D}_2 \otimes \boldsymbol{D}_1)\boldsymbol{X}_{(3)}^\top$$

$$\min_{\boldsymbol{B}} \|\boldsymbol{S}\bar{\mathcal{F}}\bar{\boldsymbol{D}}\boldsymbol{Z}\boldsymbol{B}^\top - \boldsymbol{S}\bar{\mathcal{F}}\bar{\boldsymbol{D}}\boldsymbol{X}^\top\|_F^2$$

Pre-mixed tensor

$$\tilde{\mathcal{X}} = \mathcal{X} \times_1 \mathcal{F}_1\boldsymbol{D}_1 \times_2 \mathcal{F}_2\boldsymbol{D}_2 \times_3 \mathcal{F}_3\boldsymbol{D}_3$$

$$\tilde{\boldsymbol{X}}_{(3)}^\top = (\mathcal{F}_2\boldsymbol{D}_2 \otimes \mathcal{F}_1\boldsymbol{D}_1)\boldsymbol{X}_{(3)}^\top(\mathcal{F}_3\boldsymbol{D}_3)^\top$$



r
n
s
$\mathbf{S}\bar{\mathcal{F}}\bar{\mathbf{D}}\mathbf{Z}$
$\mathbf{B}^\intercal$
=
$\mathbf{S}\bar{\mathcal{F}}\bar{\mathbf{D}}\mathbf{X}^\intercal$

Sample before unmixing

$$\boldsymbol{S}\bar{\mathcal{F}}\bar{\boldsymbol{D}}\boldsymbol{X}^\top = (\boldsymbol{S}\tilde{\boldsymbol{X}}_{(3)}^\top)\boldsymbol{D}_3\mathcal{F}_3^*$$

CP-ARLS-Mix Algorithm

Inputs: Tensor $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d}$, desired rank $r \in \mathbb{N}$, number of samples $s \in \mathbb{N}$.

```
Initialize $\boldsymbol{A}_k \in \mathbb{R}^{n_k \times r}$ for all $k \in [d]$
Draw random diagonal $\boldsymbol{D}_k$ for all $k \in [d]$
Compute $\tilde{\boldsymbol{A}}_k = \mathcal{F}_k \boldsymbol{D}_k \boldsymbol{A}_k$for all $k \in [d]$
Compute $\tilde{\mathcal{X}} = \mathcal{X} \times_1 \mathcal{F}_1 \boldsymbol{D}_1 \times_2 \mathcal{F}_2 \boldsymbol{D}_2 \cdots \times_d \mathcal{F}_d \boldsymbol{D}_d$
$\Omega \leftarrow$sampled indices for function value estimation
repeat
    for $k = 1, \ldots, d$ do
        $\boldsymbol{S} \leftarrow$ random rows of $\boldsymbol{I}$ scaled by $1/\sqrt{s}$.
        $\hat{\boldsymbol{Z}} \leftarrow \mathrm{SKRP}(\boldsymbol{S}, \tilde{\boldsymbol{A}}_1, \ldots, \tilde{\boldsymbol{A}}_{k-1}, \tilde{\boldsymbol{A}}_{k+1}, \ldots, \tilde{\boldsymbol{A}}_d)$
        $\hat{\boldsymbol{X}} \leftarrow \mathcal{F}_k^* \boldsymbol{D}_k \left(\mathrm{STU}(\boldsymbol{S}, \tilde{\mathcal{X}}, k)\right)$
        $\boldsymbol{A}_k \leftarrow \arg\min_{\boldsymbol{B}} \|\hat{\boldsymbol{Z}} \boldsymbol{B}^\top - \hat{\boldsymbol{X}}^\top\|_F^2$
        $\tilde{\boldsymbol{A}}_k \leftarrow \mathcal{F}_k \boldsymbol{D}_k \boldsymbol{A}_k$
    end
until SFV$(\Omega, \mathcal{X}, \mathbf{A}_1, \boldsymbol{A}_2, \ldots, \boldsymbol{A}_d)$ ceases to decrease
```

Is the KFJLT adequate? YES

$$\min_{\boldsymbol{B}} \|\boldsymbol{S}\bar{\mathcal{F}}\bar{\boldsymbol{D}}\boldsymbol{Z}\boldsymbol{B}^\top - \boldsymbol{S}\bar{\mathcal{F}}\bar{\boldsymbol{D}}\boldsymbol{X}^\top\|_F^2$$

- $\boldsymbol{S}$ is $s \times N$ sampling matrix
- $\bar{\mathcal{F}} = \mathcal{F}_d \otimes \cdots \otimes \mathcal{F}_{k+1} \otimes \mathcal{F}_{k-1} \otimes \cdots \otimes \mathcal{F}_1$.
- $\bar{\boldsymbol{D}} = \boldsymbol{D}_d \otimes \cdots \otimes \boldsymbol{D}_{k+1} \otimes \boldsymbol{D}_{k-1} \otimes \cdots \otimes \boldsymbol{D}_1$.

- R. Jin, T. G. Kolda, and R. Ward. *Faster Johnson Lindenstrauss Transforms via Kronecker Products*, Information and Inference, 2020
- O. A. Malik, and S. Becker. *Guarantees for the Kronecker Fast Johnson Lindenstrauss Transform Using a Coherence and Sampling Argument*, Linear Algebra and its Applications, 2020

Recall: JL Lemma

JL Lemma

Let $\Phi \in \mathbb{R}^{m \times N}$ have independent entries $s_{ij} \sim \frac{1}{\sqrt{m}}\mathcal{N}(0,1)$. If $m = O\left(\frac{\log(p)}{\epsilon^2}\right)$, then for any set of n data points $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_p \in \mathbb{R}^N$, with high probability:

$$(1-\epsilon)\|\boldsymbol{x}_i - \boldsymbol{x}_j\|_2 \leq \|\Phi\boldsymbol{x}_i - \Phi\boldsymbol{x}_j\|_2 \leq (1+\epsilon)\|\boldsymbol{x}_i - \boldsymbol{x}_j\|_2$$

Recall: JL Lemma

JL Lemma

Let $\Phi \in \mathbb{R}^{m \times N}$ have independent entries $s_{ij} \sim \frac{1}{\sqrt{m}}\mathcal{N}(0,1)$. If $m = O\left(\frac{\log(p)}{\epsilon^2}\right)$, then for any set of n data points $\boldsymbol{x}_1, \dots, \boldsymbol{x}_p \in \mathbb{R}^N$, with high probability:

$$(1-\epsilon)\|\boldsymbol{x}_i - \boldsymbol{x}_j\|_2 \leq \|\Phi\boldsymbol{x}_i - \Phi\boldsymbol{x}_j\|_2 \leq (1+\epsilon)\|\boldsymbol{x}_i - \boldsymbol{x}_j\|_2$$

The Fast JL Lemma

Let $\Phi = \boldsymbol{SHD} \in \mathbb{R}^{m \times N}$ be a subsampled randomized Hadamard transform with $m = O\left(\frac{\log(N)\log(p)}{\epsilon^2}\right)$. Then for any set of p data points $\boldsymbol{x}_1, \dots, \boldsymbol{x}_p \in \mathbb{R}^N$, with high probability,

$$\|\Phi\boldsymbol{x}_i\|_2 = (1 \pm \epsilon)\|\boldsymbol{x}_i\|_2.$$

KFJLT

KFJL Result

Let $\Phi = \boldsymbol{S}(\mathcal{F}_d \boldsymbol{D}_d \otimes \cdots \otimes \mathcal{F}_1 \boldsymbol{D}_1) \in \mathbb{R}^{m \times N}$ be a KFJLT with $m = O\left(\frac{\log(N)\log^{2d-1}(p)}{\epsilon^2}\right)$. Then for any set of p data points $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_p \in \mathbb{R}^N$, with high probability,

$$\|\Phi \boldsymbol{x}_i\|_2 = (1 \pm \epsilon)\|\boldsymbol{x}_i\|_2.$$

- R. Jin, T. G. Kolda, and R. Ward. *Faster Johnson Lindenstrauss Transforms via Kronecker Products*, Information and Inference, 2020

KFJLT and LS regression

KFJLT-Sketch and solve

Given a matrix $\boldsymbol{A} \in \mathbb{R}^{N \times r}$ and a fixed vector $\boldsymbol{b} \in \mathbb{R}^{N}$, let $\boldsymbol{x}^* = \min_{\boldsymbol{x} \in \mathbb{R}^d} \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}\|_2$. Let $\Phi = \boldsymbol{S}(\mathcal{F}_d \boldsymbol{D}_d \otimes \cdots \otimes \mathcal{F}_1 \boldsymbol{D}_1) \in \mathbb{R}^{m \times N}$ be a KFJLT with

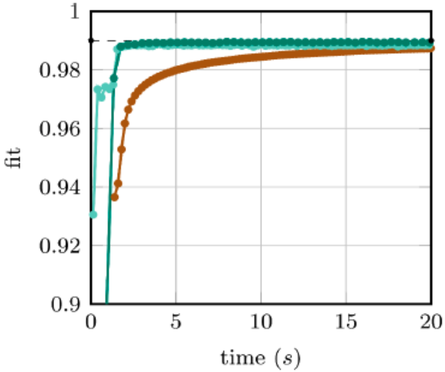
$$m = O\left(\frac{r^{2d} \log(N) \log^{2d-1}(r)}{\epsilon} \right),$$

and if $\tilde{\boldsymbol{x}} = \min_{\boldsymbol{x} \in \mathbb{R}^d} \|\Phi(\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b})\|_2$, then, with high probability,

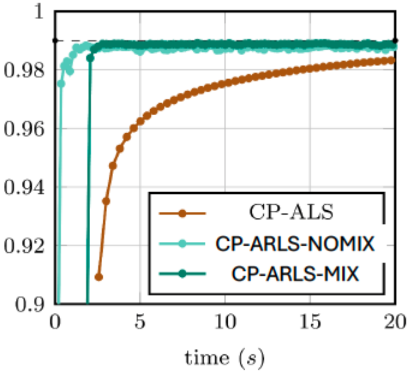
$$\|\boldsymbol{A}\tilde{\boldsymbol{x}} - \boldsymbol{b}\|_2 \leq (1 + \epsilon)\|\boldsymbol{A}\boldsymbol{x}^* - \boldsymbol{b}\|_2.$$

- R. Jin, T. G. Kolda, and R. Ward. *Faster Johnson Lindenstrauss Transforms via Kronecker Products*, Information and Inference, 2020

CP-ARLS faster than CP-ALS



300 x 300 x 300 Random Rank-5 Tensor
with 1% Noise



80 x 80 x 80 x 80 Random Rank-5 Tensor
with 1% Noise

Battaglino, Ballard, & Kolda 2017

Leverage scores and incoherence

Leverage scores

Given $\boldsymbol{A} \in \mathbb{R}^{N \times r}$, and an orthonormal basis $\boldsymbol{U}$ for $span(\boldsymbol{A})$, for $i \in [n]$, the ith *leverage score*

$$\ell_i(\boldsymbol{A}) = \sup_{\boldsymbol{x}} \frac{(\boldsymbol{A}_{i*}\boldsymbol{x})^2}{\|\boldsymbol{A}\boldsymbol{x}\|^2} = \|\boldsymbol{U}_{i*}\|^2.$$

Leverage scores and incoherence

Leverage scores

Given $\boldsymbol{A} \in \mathbb{R}^{N \times r}$, and an orthonormal basis $\boldsymbol{U}$ for *span*$(\boldsymbol{A})$, for $i \in [n]$, the ith *leverage score*

$$\ell_i(\boldsymbol{A}) = \sup_{\boldsymbol{x}} \frac{(\boldsymbol{A}_{i*}\boldsymbol{x})^2}{\|\boldsymbol{A}\boldsymbol{x}\|^2} = \|\boldsymbol{U}_{i*}\|^2.$$

Coherence

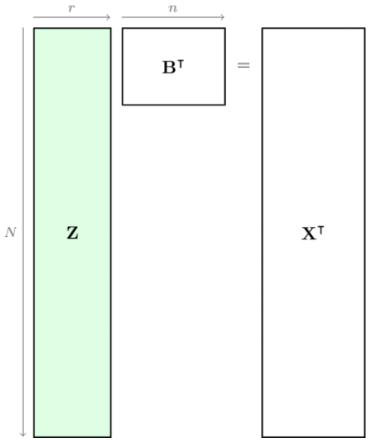
The coherence of $\boldsymbol{A} \in \mathbb{R}^{N \times r}$, denoted by $\mu(\boldsymbol{A})$ is the *maximum leverage score*, i.e.,

$$\mu(\boldsymbol{A}) = \max_{i \in [N]} \ell_i(\boldsymbol{A}).$$

We have $\frac{r}{N} \leq \mu(\boldsymbol{A}) \leq 1$.
We say $\boldsymbol{A}$ is **incoherent** if $\mu(\boldsymbol{A}) \approx \frac{r}{N}$.

Is the KRP incoherent?

$$\min_{\boldsymbol{B}} \|\boldsymbol{Z}\boldsymbol{B}^\top - \boldsymbol{X}^\top\|_F^2$$



Khatri-Rao Product:

$$\boldsymbol{Z} = \boldsymbol{A}_d \odot \cdots \odot \boldsymbol{A}_{k+1} \odot \boldsymbol{A}_{k-1} \odot \cdots \odot \boldsymbol{A}_1$$

- Lemma 1: $\mu(\boldsymbol{A} \otimes \boldsymbol{B}) = \mu(\boldsymbol{A})\mu(\boldsymbol{B})$
- Lemma 2: $\mu(\boldsymbol{A} \odot \boldsymbol{B}) \leq \mu(\boldsymbol{A})\mu(\boldsymbol{B})$

KRP is incoherent if the factor matrices are!

Recall: Leverage score sampling

Sampling for LS

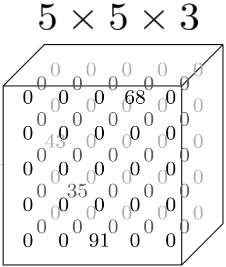
Given a matrix $\boldsymbol{A} \in \mathbb{R}^{N \times r}$ and a fixed vector $\boldsymbol{b} \in \mathbb{R}^N$, let $\boldsymbol{x}^* = \min_{\boldsymbol{x} \in \mathbb{R}^d} \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}\|_2$. Let $\boldsymbol{S} \in \mathbb{R}^{m \times N}$ be a sampling matrix with probabilities $p_i = \ell_i / r$, and $\boldsymbol{S}_{i*} = \boldsymbol{e}_j / \sqrt{m p_j}$ with $\Pr(j = i) = p_i$. If $m = O(r \log(r/\delta)/\epsilon)$ and $\tilde{\boldsymbol{x}} = \min_{\boldsymbol{x} \in \mathbb{R}^d} \|\boldsymbol{S}(\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b})\|_2$, then, with high probability,

$$\|\boldsymbol{A}\tilde{\boldsymbol{x}} - \boldsymbol{b}\|_2 \leq (1 + \epsilon)\|\boldsymbol{A}\boldsymbol{x}^* - \boldsymbol{b}\|_2.$$

We also saw a procedure for approximately estimating the leverage scores.

Sparse Tensors

We can store a sparse tensor in size proportion to its number of nonzeros nnz

$\mathcal{X} =$ 

$$N = \prod_{k=1}^{3} n_k = 75$$

$$q = 4 \text{ nonzeros}$$

$$\mathbf{C} = \begin{bmatrix} 1 & 4 & 1 \\ 3 & 1 & 3 \\ 4 & 2 & 2 \\ 5 & 3 & 1 \end{bmatrix} \in \mathbb{N}^{q \times d} \quad \text{and} \quad \mathbf{v} = \begin{bmatrix} 68 \\ 43 \\ 35 \\ 91 \end{bmatrix} \in \mathbb{R}^q,$$

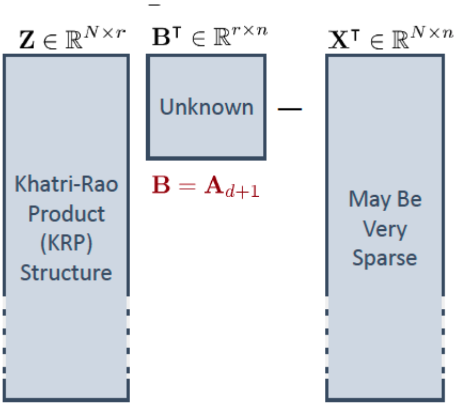
CP for sparse tensors

$N \gg r, n$

Linking back to mode-$(d+1)$ least squares subproblem

$$N = \prod_{k=1}^{d} n_k$$

$$n = n_{d+1}$$



$\mathbf{Z} \in \mathbb{R}^{N \times r}$ $\mathbf{B}^\intercal \in \mathbb{R}^{r \times n}$ $-$ $\mathbf{X}^\intercal \in \mathbb{R}^{N \times n}$

Khatri-Rao Product (KRP) Structure

Unknown

May Be Very Sparse

$\mathbf{B} = \mathbf{A}_{d+1}$

$\mathbf{Z} = \mathbf{A}_d \odot \cdots \odot \mathbf{A}_1$

$\mathbf{X} = \mathbf{X}_{(d+1)}$

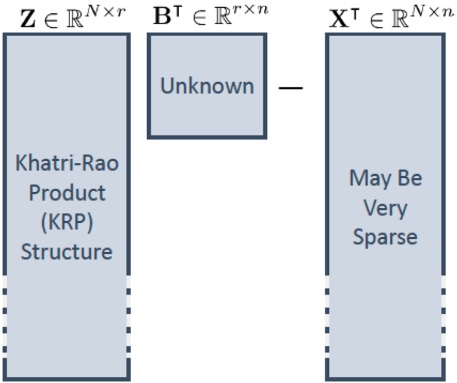
- KRP costs $O(Nr)$ to form
- System costs $O(Nnr^2)$ to solve
- KRP structure
  - Cost reduced to $O(Nnr)$
- KRP structure + data sparse
  - Cost reduced to $O(r\,\mathrm{nnz}(\mathbf{X}))$

For Reddit (mode 3)
N = 1.2B
$\mathrm{nnz}(\mathbf{X})$ = 4.7B

$$\min_{\boldsymbol{B}} \|\boldsymbol{Z}\boldsymbol{B}^\top - \boldsymbol{X}^\top\|_F^2$$

CP by sampling

$$\min_{\boldsymbol{B}} \|\boldsymbol{Z}\boldsymbol{B}^\top - \boldsymbol{X}^\top\|_F^2 \qquad \min_{\boldsymbol{B}} \|\Omega\boldsymbol{Z}\boldsymbol{B}^\top - \Omega\boldsymbol{X}^\top\|_F^2$$



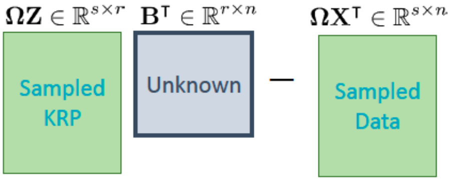
$\mathbf{Z} \in \mathbb{R}^{N\times r}$ $\mathbf{B}^\intercal \in \mathbb{R}^{r\times n}$ $\mathbf{X}^\intercal \in \mathbb{R}^{N\times n}$

Khatri-Rao Product (KRP) Structure

Unknown

$-$

May Be Very Sparse



$\mathbf{\Omega Z} \in \mathbb{R}^{s\times r}$ $\mathbf{B}^\intercal \in \mathbb{R}^{r\times n}$ $\mathbf{\Omega X}^\intercal \in \mathbb{R}^{s\times n}$

Sampled KRP

Unknown

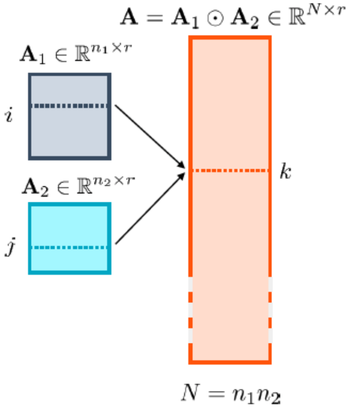
$-$

Sampled Data

Complexity reduced from $O(Nnr)$ to $O(snr^2)$

$N \gg r, n$

Bounding Leverage Scores


$\mathbf{A} = \mathbf{A}_1 \odot \mathbf{A}_2 \in \mathbb{R}^{N \times r}$
$\mathbf{A}_1 \in \mathbb{R}^{n_1 \times r}$
i
$\mathbf{A}_2 \in \mathbb{R}^{n_2 \times r}$
j
k
$N = n_1 n_2$

Upper Bound on Leverage Score

Lemma (Cheng et al., NIPS 2016; Battaglino et al., SIMAX 2018):

$$\ell_k(\mathbf{A}) \le \ell_i(\mathbf{A}_1)\ell_j(\mathbf{A}_2)$$

Too expensive to calculate $\mathcal{O}(Nr^2)$

Cheap to calculate leverage scores for each submatrix $\mathcal{O}((n_1 + n_2)r^2)$

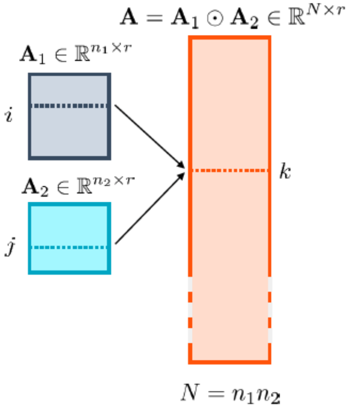
Probability of Sampling row k in $\mathbf{A}$:

$$p_k = \frac{\ell_i(\mathbf{A}_1)\ell_j(\mathbf{A}_2)}{r^2}$$

1-1 Correspondence between k and (i, j)

$$k \in [N] \quad \Leftrightarrow \quad (i, j) \in [n_1] \otimes [n_2]$$

Sampling Piecemeal


$\mathbf{A} = \mathbf{A}_1 \odot \mathbf{A}_2 \in \mathbb{R}^{N \times r}$
$\mathbf{A}_1 \in \mathbb{R}^{n_1 \times r}$
i
$\mathbf{A}_2 \in \mathbb{R}^{n_2 \times r}$
j
k
$N = n_1 n_2$

Upper Bound on Leverage Score

Lemma (Cheng et al., NIPS 2016; Battaglino et al., SIMAX 2018):

$$\ell_k(\mathbf{A}) \le \ell_i(\mathbf{A}_1)\ell_j(\mathbf{A}_2)$$

Too expensive to calculate $\mathcal{O}(Nr^2)$

Cheap to calculate leverage scores for each submatrix $\mathcal{O}((n_1 + n_2)r^2)$

1-1 Correspondence between k and (i, j)

$$k \in [N] \quad \Leftrightarrow \quad (i, j) \in [n_1] \otimes [n_2]$$

Probability of Sampling row k in $\mathbf{A}$:

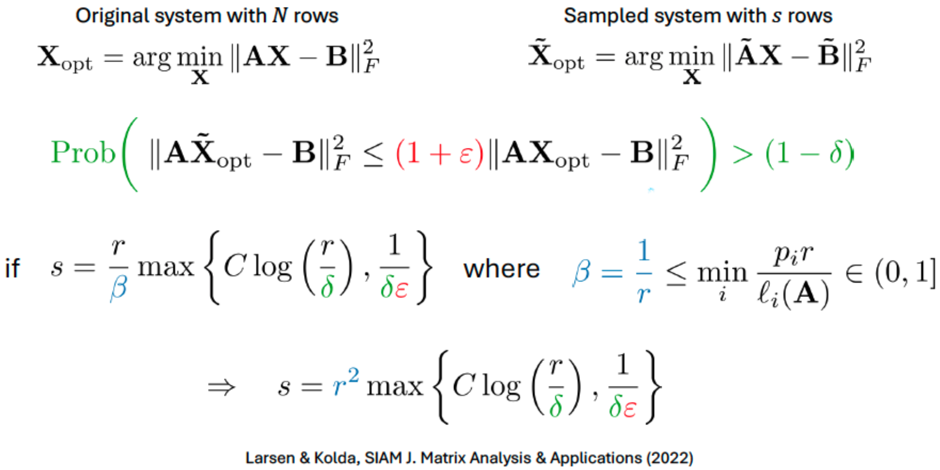
$$p_k = \frac{\ell_i(\mathbf{A}_1)\ell_j(\mathbf{A}_2)}{r^2}$$

Choose $i \sim p_i = \ell_i(\mathbf{A}_1)/r$

Choose $j \sim p_j = \ell_j(\mathbf{A}_2)/r$

$$k = i + (j - 1)n_1$$

Accuracy of Sketched ALS for 3-way Tensors



Original system with N rows

$$\mathbf{X}_{\text{opt}} = \arg\min_{\mathbf{X}} \|\mathbf{A}\mathbf{X} - \mathbf{B}\|_F^2$$

Sampled system with s rows

$$\tilde{\mathbf{X}}_{\text{opt}} = \arg\min_{\mathbf{X}} \|\tilde{\mathbf{A}}\mathbf{X} - \tilde{\mathbf{B}}\|_F^2$$

$$\text{Prob}\left(\|\mathbf{A}\tilde{\mathbf{X}}_{\text{opt}} - \mathbf{B}\|_F^2 \le (1+\varepsilon)\|\mathbf{A}\mathbf{X}_{\text{opt}} - \mathbf{B}\|_F^2 \right) > (1-\delta)$$

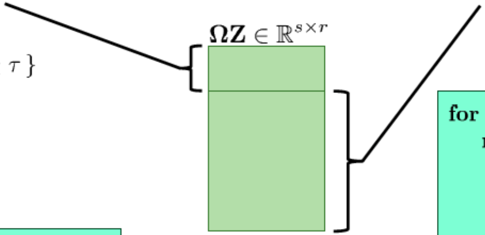
if $\quad s = \dfrac{r}{\beta} \max\left\{ C \log\left(\dfrac{r}{\delta}\right), \dfrac{1}{\delta\varepsilon} \right\}$ where $\beta = \dfrac{1}{r} \le \min_i \dfrac{p_i r}{\ell_i(\mathbf{A})} \in (0,1]$

$$\Rightarrow \quad s = r^2 \max\left\{ C \log\left(\frac{r}{\delta}\right), \frac{1}{\delta\varepsilon} \right\}$$

Larsen & Kolda, SIAM J. Matrix Analysis & Applications (2022)

Hybrid Deterministic and Randomly Sampled Rows

Deterministic Rows



$\mathbf{\Omega Z} \in \mathbb{R}^{s\times r}$

Random Rows

$$s_{\text{rnd}} = s - s_{\text{det}}$$

$$\mathcal{D}_\tau = \{\, i \in [N] \mid p_i \geq \tau \,\}$$

$$s_{\text{det}} = |\mathcal{D}_\tau|$$

$$p_{\text{det}} = \sum_{i\in\mathcal{D}_\tau} p_i$$

```
for $i \in \mathcal{D}_\tau$ do
    add row $\mathbf{A}_1(i_1,:) * \cdots * \mathbf{A}_d(i_d,:)$
end for
```

```
for $j = 1 \ldots, s_{\text{rnd}}$ do
    repeat
        for $k = 1 \ldots, d$ do
            $i_k \leftarrow \texttt{multi}(\ell(\mathbf{A}_k)/r)$
        end for
    until $i \notin \mathcal{D}_\tau$
    $\omega \leftarrow \sqrt{(1 - p_{\text{det}})/(s_{\text{rnd}}\, p_i)}$
    add row $\omega\,(\mathbf{A}_1(i_1,:) * \cdots * \mathbf{A}_d(i_d,:))$
end for
```

$$p_i \equiv \frac{1}{r^d} \prod_{k=1}^{d} \ell_{i_k}(\mathbf{A}_k)$$

1-1 Correspondence between *linear index and multi index:*

$$i \in [N] \Leftrightarrow (i_1, \ldots, i_d) \in [n_1] \otimes \cdots \otimes [n_d]$$

Find All High Probability Rows without Computing All Probabilities

- Recall

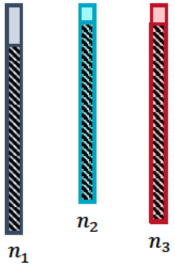
$$p_i \equiv \frac{1}{r^d} \prod_{k=1}^{d} \ell_{i_k}(\mathbf{A}_k)$$

- For given tolerance $\tau > 1/N$, define the set of deterministic rows to include

$$\mathcal{D}_\tau = \{\, i \in [N] \mid p_i \geq \tau \,\}$$

  - Compute *without* computing all p_i values
  - A few high leverage scores means all the others are necessarily low!
  - Use bounding procedure to eliminate most options
  - Compute products of at most a top few leverage scores in each mode

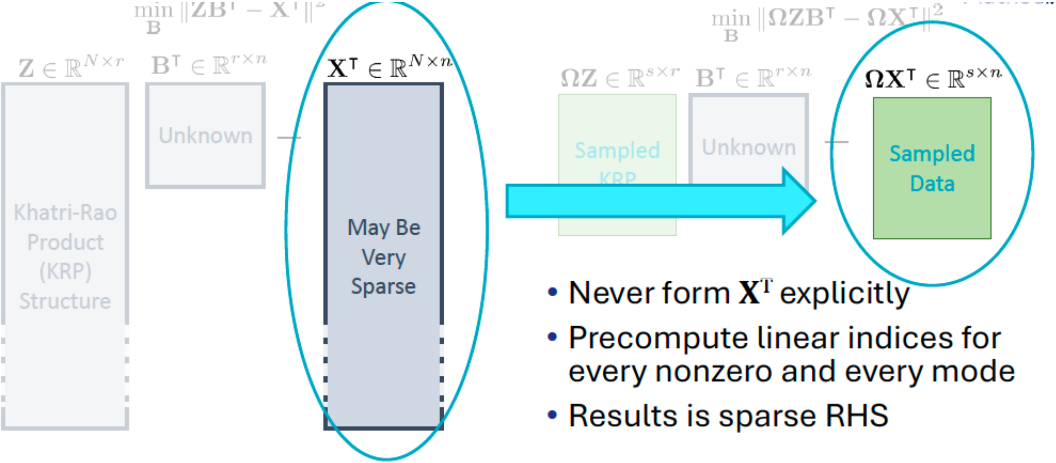
Sorted Leverages Scores (Descending)



1-1 Correspondence between *linear index and multi index:*

$i \in [N] \Leftrightarrow (i_1, \ldots, i_d) \in [n_1] \otimes \cdots \otimes [n_d]$

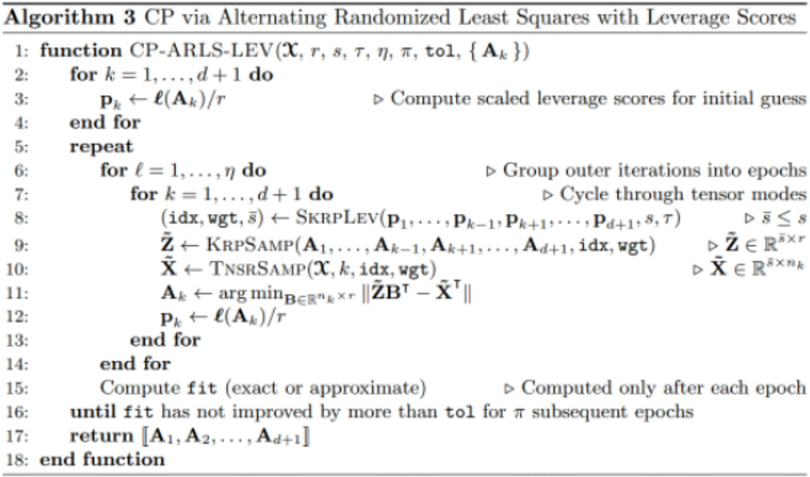
Efficiently Extract RHS from (Sparse) tensor


$\min_{\mathbf{B}} \|\mathbf{Z}\mathbf{B}^\intercal - \mathbf{X}^\intercal\|^2$
$\mathbf{Z} \in \mathbb{R}^{N \times r}$
$\mathbf{B}^\intercal \in \mathbb{R}^{r \times n}$
$\mathbf{X}^\intercal \in \mathbb{R}^{N \times n}$
Khatri-Rao Product (KRP) Structure
Unknown
−
May Be Very Sparse
$\min_{\mathbf{B}} \|\Omega\mathbf{Z}\mathbf{B}^\intercal - \Omega\mathbf{X}^\intercal\|^2$
$\Omega\mathbf{Z} \in \mathbb{R}^{s \times r}$
$\mathbf{B}^\intercal \in \mathbb{R}^{r \times n}$
$\Omega\mathbf{X}^\intercal \in \mathbb{R}^{s \times n}$
Sampled KRP
Unknown
−
Sampled Data

- Never form $\mathbf{X}^\intercal$ explicitly
- Precompute linear indices for every nonzero and every mode
- Results is sparse RHS

Similar in spirit to ideas for dense tensors in Battaglino et al., SIMAX 2018

CP-ARLS-Lev

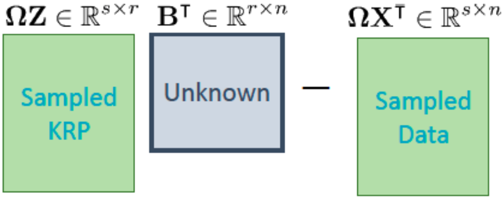


Algorithm 3 CP via Alternating Randomized Least Squares with Leverage Scores

```
function CP-ARLS-LEV($\mathcal{X}$, $r$, $s$, $\tau$, $\eta$, $\pi$, tol, $\{\mathbf{A}_k\}$)
    for $k = 1, \dots, d+1$ do
        $\mathbf{p}_k \leftarrow \boldsymbol{\ell}(\mathbf{A}_k)/r$        ▷ Compute scaled leverage scores for initial guess
    end for
    repeat
        for $\ell = 1, \dots, \eta$ do        ▷ Group outer iterations into epochs
            for $k = 1, \dots, d+1$ do        ▷ Cycle through tensor modes
                $(\texttt{idx}, \texttt{wgt}, \bar{s}) \leftarrow$ SKRPLEV$(\mathbf{p}_1, \dots, \mathbf{p}_{k-1}, \mathbf{p}_{k+1}, \dots, \mathbf{p}_{d+1}, s, \tau)$        ▷ $\bar{s} \leq s$
                $\tilde{\mathbf{Z}} \leftarrow$ KRPSAMP$(\mathbf{A}_1, \dots, \mathbf{A}_{k-1}, \mathbf{A}_{k+1}, \dots, \mathbf{A}_{d+1}, \texttt{idx}, \texttt{wgt})$        ▷ $\tilde{\mathbf{Z}} \in \mathbb{R}^{\bar{s} \times r}$
                $\tilde{\mathbf{X}} \leftarrow$ TNSRSAMP$(\mathcal{X}, k, \texttt{idx}, \texttt{wgt})$        ▷ $\tilde{\mathbf{X}} \in \mathbb{R}^{\bar{s} \times n_k}$
                $\mathbf{A}_k \leftarrow \arg\min_{\mathbf{B} \in \mathbb{R}^{n_k \times r}} \|\tilde{\mathbf{Z}}\mathbf{B}^\intercal - \tilde{\mathbf{X}}^\intercal\|$
                $\mathbf{p}_k \leftarrow \boldsymbol{\ell}(\mathbf{A}_k)/r$
            end for
        end for
        Compute fit (exact or approximate)        ▷ Computed only after each epoch
    until fit has not improved by more than tol for $\pi$ subsequent epochs
    return $[\mathbf{A}_1, \mathbf{A}_2, \dots, \mathbf{A}_{d+1}]$
end function
```

Hybrid leverage score sampling

Single Least Squares Problem with N = 46M rows, r = 10 columns, n = 183 right-hand sides



$\mathbf{\Omega Z} \in \mathbb{R}^{s\times r}$ $\mathbf{B}^\intercal \in \mathbb{R}^{r\times n}$ $\mathbf{\Omega X}^\intercal \in \mathbb{R}^{s\times n}$

Sampled KRP

Unknown

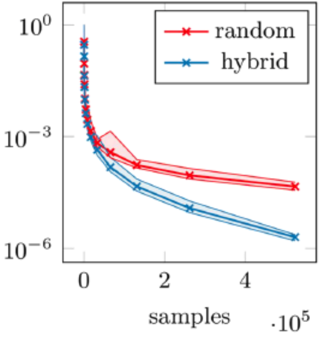
−

Sampled Data

$$\tilde{\mathbf{B}}_* \equiv \arg\min_{\mathbf{B}\in\mathbb{R}^r} \|\mathbf{\Omega Z B}^\intercal - \mathbf{\Omega X}^\intercal\|_2^2$$

$$\mathbf{B}_* \equiv \arg\min_{\mathbf{B}\in\mathbb{R}^r} \|\mathbf{Z B}^\intercal - \mathbf{X}^\intercal\|_2^2$$

$$\frac{\left| \|\mathbf{Z B}_*^\intercal - \mathbf{X}^\intercal\|_2 - \|\mathbf{Z}\tilde{\mathbf{B}}_*^\intercal - \mathbf{X}^\intercal\|_2 \right|}{\max\{1, \|\mathbf{Z B}_*^\intercal - \mathbf{X}^\intercal\|_2\}}$$



Matlab Demo