# Fast and accurate randomized algorithms for low-rank tensor decompositions

Linjian Ma and Edgar Solomonik

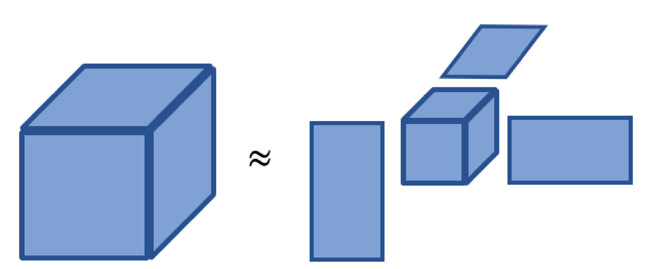
Department of Computer Science, University of Illinois at Urbana Champaign

## Abstract

Low-rank Tucker [3] tensor decomposition is a powerful tool in data analytics. However, the widely used alternating least squares (ALS) method is costly for large and sparse tensors. We propose a fast and accurate sketched ALS algorithm for Tucker decomposition, which solves a sequence of sketched rank-constrained linear least squares subproblems.

## Tucker decomposition

$$\boldsymbol{T} \approx \boldsymbol{X} \times_1 \boldsymbol{A} \times_2 \boldsymbol{B} \times_3 \boldsymbol{C}$$



$\boldsymbol{T} \in \mathbb{R}^{s\times s\times s}$, $\boldsymbol{X} \in \mathbb{R}^{R\times R\times R}$, $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C} \in \mathbb{R}^{s\times R}$ with orthonormal columns, $R < s$

## Existing optimization methods

1. Higher order orthogonal iteration (HOOI) [4]

$$\min_{(\boldsymbol{A},\boldsymbol{X})} \left\| (\boldsymbol{C}\otimes\boldsymbol{B})\boldsymbol{X}_{(1)}^T\boldsymbol{A}^T - \boldsymbol{T}_{(1)}^T \right\|_F^2$$

- Update sequence: $(\boldsymbol{A},\boldsymbol{X}), (\boldsymbol{B},\boldsymbol{X}), (\boldsymbol{C},\boldsymbol{X})$
- Fast convergence
- Costs $\Theta(s^3R)$ (dense case) or $\Omega(\text{nnz}(\boldsymbol{T})R)$ (sparse case)

2. Alternating unconstrained least squares (AULS)

$$\min_{\boldsymbol{A}} \left\| (\boldsymbol{C}\otimes\boldsymbol{B})\boldsymbol{X}_{(1)}^T\boldsymbol{A}^T - \boldsymbol{T}_{(1)}^T \right\|_F^2$$

$$\min_{\boldsymbol{X}} \left\| (\boldsymbol{C}\otimes\boldsymbol{B}\otimes\boldsymbol{A})\text{vec}(\boldsymbol{X}) - \text{vec}(\boldsymbol{T}) \right\|_F^2$$

- Update sequence: $(\boldsymbol{A}), (\boldsymbol{B}), (\boldsymbol{C}), (\boldsymbol{X})$
- Slow convergence
- Costs $\Theta(s^3R)$ (dense case) or $\Omega(\text{nnz}(\boldsymbol{T})R)$ (sparse case)

3. Sketched AULS with TensorSketch [1]

$$\min_{\boldsymbol{A}} \left\| \boldsymbol{S}(\boldsymbol{C}\otimes\boldsymbol{B})\boldsymbol{X}_{(1)}^T\boldsymbol{A}^T - \boldsymbol{S}\boldsymbol{T}_{(1)}^T \right\|_F^2$$

$$\min_{\boldsymbol{X}} \left\| \boldsymbol{S}(\boldsymbol{C}\otimes\boldsymbol{B}\otimes\boldsymbol{A})\text{vec}(\boldsymbol{X}) - \boldsymbol{S}\text{vec}(\boldsymbol{T}) \right\|_F^2$$

- $\boldsymbol{S}$: sketching matrix, TensorSketch [2] is used in the reference
- Advantage: overall cost with $t$ sweeps reduced from $\Omega(t\text{nnz}(\boldsymbol{T})R)$ to $O\left(\text{nnz}(\boldsymbol{T}) + t\left(sR^5 + R^7\right)\right)$
- Disadvantage: slow convergence since based on Tucker-AULS

## Our approach: sketched HOOI

Let $\boldsymbol{Q} = \boldsymbol{C}\otimes\boldsymbol{B}$, $\boldsymbol{Y} = \boldsymbol{T}_{(1)}^T$

HOOI: solve and truncate

$$\boldsymbol{P}_{\text{opt}} \leftarrow \underset{\boldsymbol{P}\in\mathbb{R}^{s\times R^2}}{\text{argmin}} \left\| \boldsymbol{Q}\boldsymbol{P}^T - \boldsymbol{Y} \right\|_F^2$$

$$\boldsymbol{A}\boldsymbol{X}_{(1)} \leftarrow \boldsymbol{P}_R$$

Sketched HOOI: sketch, solve and truncate

$$\hat{\boldsymbol{P}}_{\text{opt}} \leftarrow \underset{\boldsymbol{P}\in\mathbb{R}^{s\times R^2}}{\text{argmin}} \left\| \boldsymbol{S}\boldsymbol{Q}\boldsymbol{P}^T - \boldsymbol{S}\boldsymbol{Y} \right\|_F^2$$

$$\hat{\boldsymbol{A}}\hat{\boldsymbol{X}}_{(1)} \leftarrow \hat{\boldsymbol{P}}_R$$

- $\boldsymbol{P}_R, \hat{\boldsymbol{P}}_R$: best rank-$R$ approximation of $\boldsymbol{P}_{\text{opt}}, \hat{\boldsymbol{P}}_{\text{opt}}$
- $\boldsymbol{S} \in \mathbb{R}^{m\times s^2}$: sketching matrix, $m < s^2$ is the sketch size
- $\boldsymbol{Q}$ has **orthonormal columns**
- Sketched **rank-constrained** linear least squares problem

## $(1/2, \delta, \epsilon)$-accurate sketching matrix

$\boldsymbol{S}$ is a $(1/2, \delta, \epsilon)$-accurate sketching matrix for $\boldsymbol{Q}$ if with probability at least $1-\delta$,

- each singular value $\sigma$ of $\boldsymbol{S}\boldsymbol{Q}$ satisfies $1 - 1/2 \le \sigma^2 \le 1 + 1/2$,
- and for any fixed matrix $\boldsymbol{M}$ $\|\boldsymbol{Q}^T\boldsymbol{S}^T\boldsymbol{S}\boldsymbol{M} - \boldsymbol{Q}^T\boldsymbol{M}\|_F^2 \le \epsilon^2 \cdot \|\boldsymbol{M}\|_F^2$.

With $\boldsymbol{Q} = \boldsymbol{C}\otimes\boldsymbol{B} \in \mathbb{R}^{s^2\times R^2}$, sketching techniques below are $(1/2, \delta, \epsilon)$-accurate

- TensorSketch (a tensorized CountSketch) [2] with sketch size $O(R^2/\delta \cdot (R^2 + 1/\epsilon^2))$
- Leverage score sampling (Importance sampling based on the leverage score of $\boldsymbol{Q}$) with sketch size $O(R^2/(\epsilon^2\delta))$

## Sketched rank-constrained linear least squares

**New theoretical contribution**: when $\boldsymbol{S}$ is a $(1/2, \delta, \epsilon)$-accurate sketching matrix for $\boldsymbol{Q}$, then with probability at least $1-\delta$

$$\left\| \boldsymbol{Q}\hat{\boldsymbol{P}}_R - \boldsymbol{Y} \right\|_F^2 \le (1 + O(\epsilon)) \left\| \boldsymbol{Q}\boldsymbol{P}_R - \boldsymbol{Y} \right\|_F^2. \quad (1)$$

Comparison of sufficient conditions to guarantee Equation 1:

| | rank-constrained LS with $\boldsymbol{Q}$ having orthonormal columns | unconstrained LS |
|---|---|---|
| $\boldsymbol{S}$ | $(1/2, \delta, \epsilon)$-accurate | $(1/2, \delta, \sqrt{\epsilon})$-accurate |

- Tighter bound on $\epsilon$ for $\boldsymbol{S}$ is needed for rank-constrained LS to be $(1+O(\epsilon))$-accurate

Proof sketch: when $\boldsymbol{S}$ is a $(1/2, \delta, \epsilon)$-accurate sketching matrix

$$\left\| \boldsymbol{Q}\boldsymbol{P}_R - \boldsymbol{Y} \right\|_F^2 = \left\| \boldsymbol{Y}^\perp \right\|_F^2 + \underbrace{\left\| \boldsymbol{P}_R - \boldsymbol{P}_{\text{opt}} \right\|_F^2}_{\text{low rank truncation error}}$$

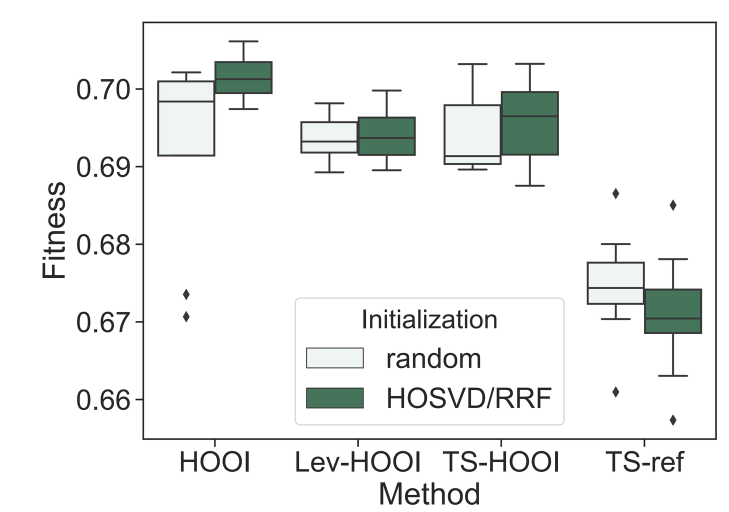
$$\left\| \boldsymbol{Q}\hat{\boldsymbol{P}}_R - \boldsymbol{Y} \right\|_F^2 = \left\| \boldsymbol{Y}^\perp \right\|_F^2 + \underbrace{\left\| \hat{\boldsymbol{P}}_{\text{opt}} - \boldsymbol{P}_{\text{opt}} \right\|_F^2}_{\text{sketched least squares error}} + \underbrace{\left\| \hat{\boldsymbol{P}}_R - \hat{\boldsymbol{P}}_{\text{opt}} \right\|_F^2 + 2\left\langle \hat{\boldsymbol{P}}_R - \hat{\boldsymbol{P}}_{\text{opt}}, \hat{\boldsymbol{P}}_{\text{opt}} - \boldsymbol{P}_{\text{opt}} \right\rangle_F}_{\text{sketched low rank truncation error}}$$

- $\left\| \hat{\boldsymbol{P}}_{\text{opt}} - \boldsymbol{P}_{\text{opt}} \right\|_F^2 = O(\epsilon^2) \left\| \boldsymbol{Y}^\perp \right\|_F^2$ [6]
- $\left\| \hat{\boldsymbol{P}}_R - \hat{\boldsymbol{P}}_{\text{opt}} \right\|_F^2 = \left\| \boldsymbol{P}_R - \boldsymbol{P}_{\text{opt}} \right\|_F^2 + O(\epsilon) \left\| \boldsymbol{Q}\boldsymbol{P}_R - \boldsymbol{Y} \right\|_F^2$ (Mirsky's inequality [5])
- $\left\langle \hat{\boldsymbol{P}}_R - \hat{\boldsymbol{P}}_{\text{opt}}, \hat{\boldsymbol{P}}_{\text{opt}} - \boldsymbol{P}_{\text{opt}} \right\rangle_F = O(\epsilon) \left\| \boldsymbol{Q}\boldsymbol{P}_R - \boldsymbol{Y} \right\|_F^2$ (Mirsky's inequality)

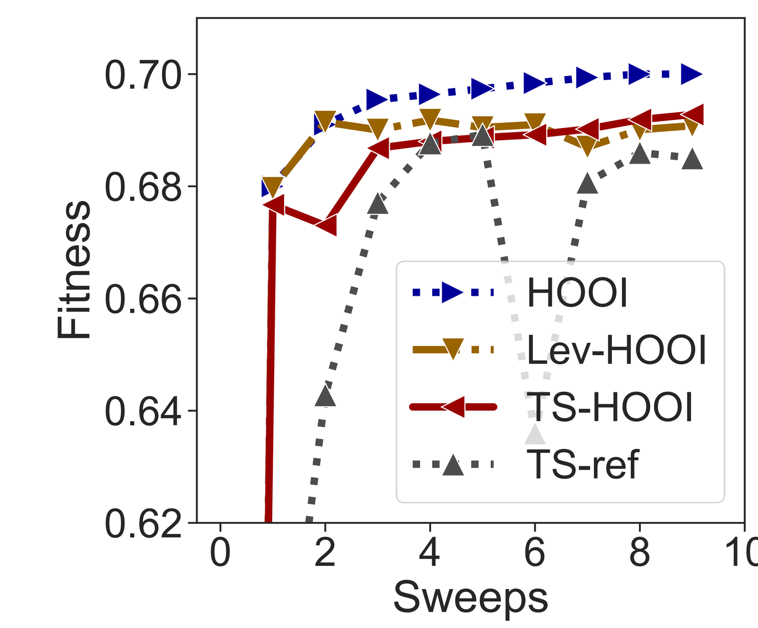
## Cost Comparison

| Algorithm for Tucker | LS subproblem cost | Sketch size ($m$) |
|---|---|---|
| HOOI | $\Omega(\text{nnz}(\boldsymbol{T})R)$ | / |
| AULS + TensorSketch | $\tilde{O}(msR + mR^3)$ | $O(R^2/\delta \cdot (R^2 + 1/\epsilon))$ |
| **HOOI + TensorSketch** | $O(msR + mR^4)$ | $O(R^2/\delta \cdot (R^2 + 1/\epsilon^2))$ |
| **HOOI + leverage score sampling** | $O(msR + mR^4)$ | $O(R^2/(\epsilon^2\delta))$ |

## Experiments: tensors with spiked signal




(a) 5 sweeps, sample size $16R^2$




(b) sample size $16R^2$

- $\boldsymbol{T} = \boldsymbol{T}_0 + \sum_{i=1}^5 \lambda_i \boldsymbol{a}_i \circ \boldsymbol{b}_i \circ \boldsymbol{c}_i$, each $\boldsymbol{a}_i, \boldsymbol{b}_i, \boldsymbol{c}_i$ has unit 2-norm, $\lambda_i = 3\frac{\|\boldsymbol{T}_0\|_F}{i^{1.5}}$
- Leading low-rank components obey the power-law distribution
- Tensor size $200 \times 200 \times 200$, $R = 5$
- TS-ref: sketched AULS with TensorSketch [1]

## Other contributions: more experiments, the initialization scheme, and CP decomposition

- Detailed comparison of TensorSketch and leverage score sampling
- An initialization scheme based on randomized range finder that improves the accuracy of leverage score sampling based sketching
- CP decomposition can be more efficiently calculated on top of sketched HOOI

## References

[1] O. A. Malik and S. Becker, *Low-rank Tucker decomposition of large tensors using TensorSketch*, NeurIPS'18.

[2] R. Pagh, *Compressed matrix multiplication*, TOCT 2013.

[3] L. R. Tucker, *Some mathematical notes on three-mode factor analysis*, Psychometrika 1966.

[4] C. A. Andersson and R. Bro, *Improving the speed of multi-way algorithms: Part I. Tucker3*, 1998.

[5] L. Mirsky, *Symmetric gauge functions and unitarily invariant norms*, 1960

[6] P. Drineas, M. W. Mahoney, S. Muthukrishnan, and T. Sarlos, *Faster least squares approximation*, 2011