Stochastic Methods for Low-Rank Tensor Recovery

Deanna Needell*
R. Grotheer, Shuang Li, Anna Ma, Jing Qin

*Dept. of Mathematics
UCLA

Feb. 4, 2020

Motivation: Low-rank Matrix Approximation

A Common Problem

Given matrix $A \in \mathbb{R}^{n \times d}$, find a rank-$k$ matrix B where $||A - B||$ is small.

- Motivation in CS: sparse vectors can be recovered from incomplete measurements
- Usually uses nuclear norm minimization (convex relaxation)

$$\min_{\mathbf{X} \in \mathbb{R}^{n_1 \times n_2}} ||\mathbf{X}||_* \qquad \text{s. t.} \qquad \mathcal{A}(\mathbf{X}) = \mathbf{y}$$

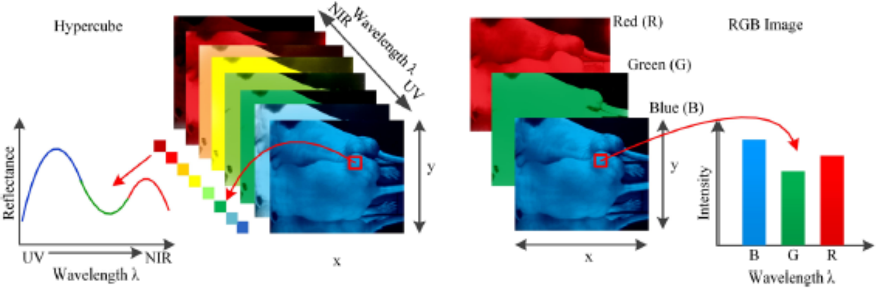
where $||\mathbf{X}||_* = \text{tr}\left((\mathbf{X}^*\mathbf{X})^{1/2}\right)$.

Problem

Question

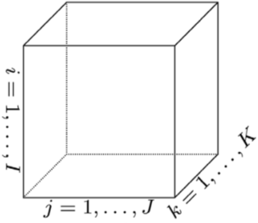
How do we deal with large dimensional data sets?

Example:



What is a tensor?

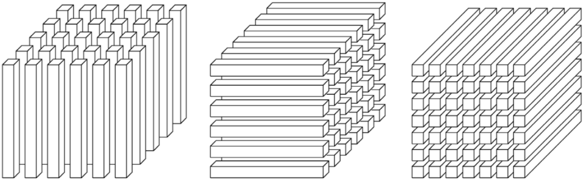
- High-dimensional extension of vectors and matrices:
  $\mathbf{X} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d}$.



- Examples: hyperspectral imaging, video processing, signal processing

Tensor Operations

- Matricize/Unfolding: $\mathbf{X}^{\{i\}} \in \mathbb{R}^{n_i \times (n_1 n_2 \cdots n_{i-1} n_{i+1} \cdots n_d)}$



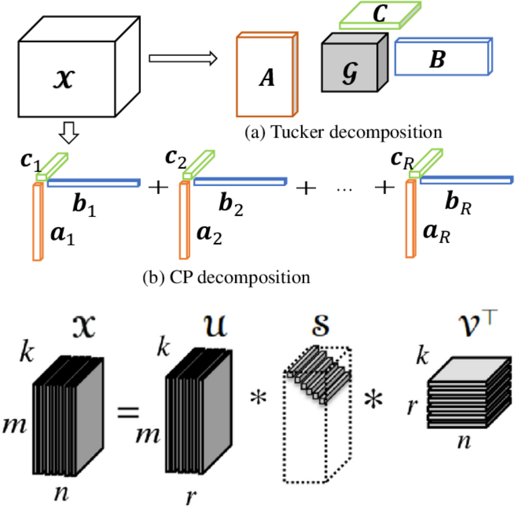
(a) Mode-1 (column) fibers: $\mathbf{x}_{:jk}$ (b) Mode-2 (row) fibers: $\mathbf{x}_{i:k}$ (c) Mode-3 (tube) fibers: $\mathbf{x}_{ij:}$

- Vectorize: $\text{vec}(\mathbf{X}) = \text{vec}(\mathbf{X}^{\{i\}})$.
- Multiplication: mode-i (matrix) product $(\mathbf{X} \times_i \mathbf{B})$ product of tensor and matrix along i-th mode
- Inner Product: $\langle \mathbf{X}_1, \mathbf{X}_2 \rangle := \text{vec}(\mathbf{X}_2)^\intercal \text{vec}(\mathbf{X}_1)$
- (induced) Frobenius norm: $||\mathbf{X}||_F := \sqrt{\langle \mathbf{X}, \mathbf{X} \rangle}$

Images from [KB09]

Tensor Decomposition

$$\mathbf{X} \approx \mathbf{G} \times_1 \mathbf{A} \times_2 \mathbf{B} \times_3 \mathbf{C}$$

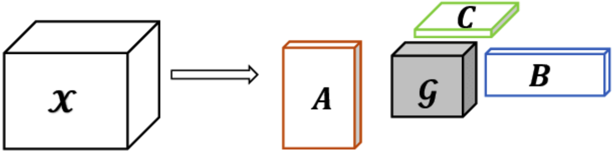


(c) Others (tubal)

Tucker Rank

One method: Higher-Order SVD (HOSVD)

$$\mathbf{X} = \mathbf{G} \times_1 \mathbf{U}^{(1)} \times_2 \cdots \times_d \mathbf{U}^{(d)}$$



Tucker rank:

$$\mathbf{r} = (r_1, \ldots, r_d), \qquad r_i = \text{rank}(\mathbf{X}^{\{i\}})$$

Tensor Recovery

Linear measurements $\mathbf{y} = \mathcal{A}(\mathbf{X}) \in \mathbb{R}^m$, $\mathcal{A} : \mathbb{R}^{n_1 \times n_2 \times \cdot \times n_d} \to \mathbb{R}^m$, linear sensing operator and

$$\mathbf{y}(i) = \mathcal{A}_i(\mathbf{X}) = \langle \mathbf{A}_i, \mathbf{X} \rangle, \qquad i = 1, \ldots, m$$

Goal:

$\min_{\mathbf{X}} F(\mathbf{X})$ subject to $\operatorname{rank}(\mathbf{X}) \leq r$ where

$$F(\mathbf{X}) = \frac{1}{2m} ||\mathbf{y} - \mathcal{A}(\mathbf{X})||_2^2 = \frac{1}{M} \sum_{i=1}^{M} \sum_{i=1}^{M} f_i(\mathbf{X})$$

$\mathcal{A}$ measurement operator.

Define functions f_i using tensor inner product:

$$f_i(\mathbf{X}_t) := \frac{1}{2b} \sum_{j=(i-1)b+1}^{ib} (y(j) - \langle \mathbf{A}_j, \mathbf{X} \rangle)^2$$

Extension to Tensors

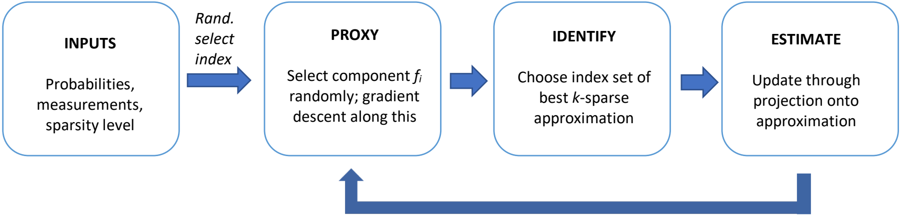
Extension of low-rank matrix recovery algorithms to tensors not obvious

- Computation of nuclear norm for tensors NP-hard
- Computation of rank of tensor NP-hard

Stochastic IHT Algorithm

Used for applications in Compressed Sensing and low-rank matrix recovery:

$$\min_{\mathbf{x}\in\mathbb{R}^n} \frac{1}{M}\sum_{i=1}^{M} f_i(\mathbf{x}), \qquad \text{subject to} \qquad ||x||_{0,\mathcal{D}} \leq k.$$



Tensor IHT

- First, extended IHT to tensors by using sum nuclear norm of matricizations [GRY11, MHWG14] $\rightarrow$ nonoptimal
- Rauhut et al [RSS17] extend to TIHT:

$$\begin{aligned} \tilde{\mathbf{X}}^t &= \mathbf{X}^t + \mu\mathcal{A}^*(\mathbf{y} - \mathcal{A}(\mathbf{X}^t)) \\ \mathbf{X}^{t+1} &= \mathcal{H}_r(\tilde{\mathbf{X}}^t) \end{aligned}$$

- $\mathcal{H}_r$ computes (quasi-best) rank-$\mathbf{r}$ approx using successive HOSVD.

Tensor RIP

Let $\mathcal{A} : \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d} \to \mathbb{R}^m$. For a fixed tensor Tucker decomposition and a corresponding Tucker rank r, we say $\mathcal{A}$ and $\mathcal{A}_{b_i}$ satisfy the TRIP if there exists a tensor restricted isometry constant δ_r such that

$$(1 - \delta_r)\|\mathbf{X}\|_F^2 \leq \|\mathcal{A}(\mathbf{X})\|_2^2 \leq (1 + \delta_r)\|\mathbf{X}\|_F^2$$

hold for all tensors $\mathbf{X} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d}$ of Tucker-rank at most r.

Tensor IHT

- First, extended IHT to tensors by using sum nuclear norm of matricizations [GRY11, MHWG14] $\to$ nonoptimal
- Rauhut et al [RSS17] extend to TIHT:

$$\begin{aligned} \tilde{\mathbf{X}}^t &= \mathbf{X}^t + \mu \mathcal{A}^*(\mathbf{y} - \mathcal{A}(\mathbf{X}^t)) \\ \mathbf{X}^{t+1} &= \mathcal{H}_r(\tilde{\mathbf{X}}^t) \end{aligned}$$

- $\mathcal{H}_r$ computes (quasi-best) rank-**r** approx using successive HOSVD.
- Provable convergence guarantees under TRIP and mild assumptions

This may ring some bells...

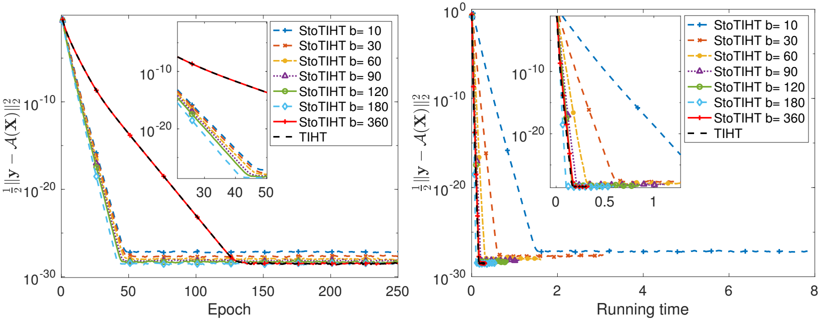
Stochastic TIHT

- Update estimate of $\mathbf{X}$ using the vectorized versions of $\mathbf{X}$ and sensing tensors $\mathbf{A}_i$ using only randomly selected i_t-th block of $\mathbf{A}$:

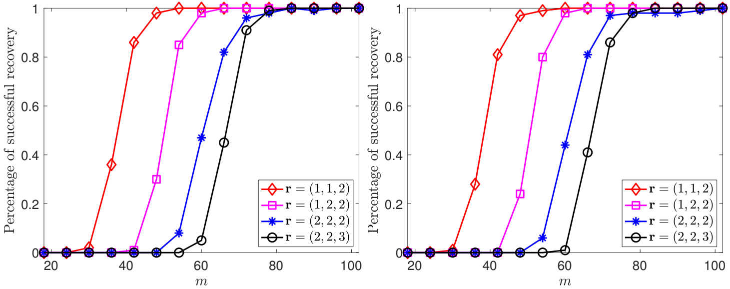
$$\tilde{\mathbf{X}}^t = \mathbf{X}^t - \frac{\mu}{Mp(i_t)} \nabla f_{i+t}(\mathbf{X}^t)$$

  where $f_{i_t} := \sum \frac{1}{2b} \sum_{j=(i_t-1)b+1}^{i_t b} (\mathbf{y}(j) - \langle \mathbf{A}_j, \mathbf{X} \rangle)^2$.
- Linear convergence (using that $\mathcal{A}, \mathcal{A}_{b_i}$ satisfy TRIP)

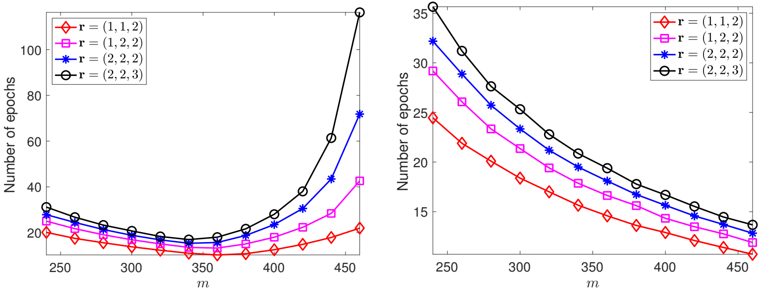
Results (5 × 5 × 6 tensor)



Results Cont

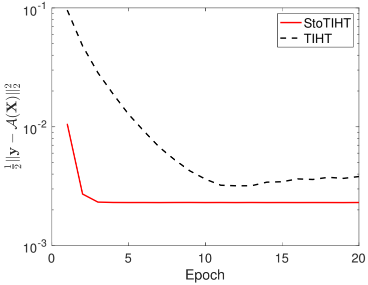


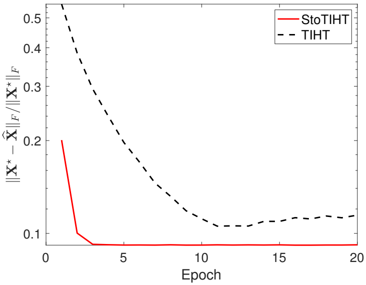
Results Cont



Left: TIHT, Right: StoTIHT

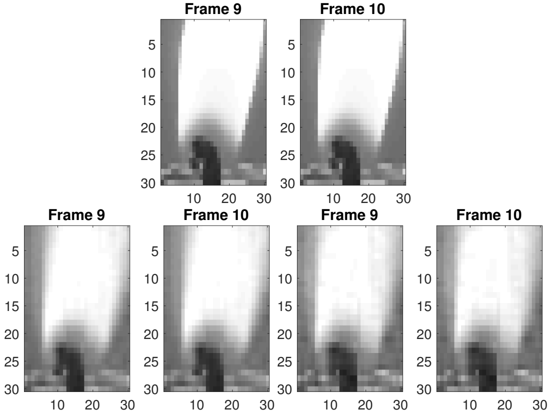
Results Cont





Left: Cost function, Right: Relative error

Results Cont.



The (a) true, (b) StoTIHT, and c) TIHT reconstructions.

Conclusions and Next Steps

- Conclusion: Stochastic version of the TIHT algorithm faster convergence, lower recovery error in large-scale setting
- Working on extending greedy stochastic algorithms for tubal rank
- Apply tensor algorithms to more real-world applications

References I

Silvia Gandy, Benjamin Recht, and Isao Yamada, *Tensor completion and low-n-rank tensor recovery via convex optimization*, Inverse Problems **27** (2011), no. 2, 025010.

Tai-Xiang Jiang, Ting-Zhu Huang, Xi-Le Zhao, and Liang-Jian Deng, *A novel nonconvex approach to recover the low-tubal-rank tensor data: when t-svd meets pssv*, arXiv preprint arXiv:1712.05870 (2017).

Tamara G Kolda and Brett W Bader, *Tensor decompositions and applications*, SIAM review **51** (2009), no. 3, 455–500.

Xiao-Yang Liu, Shuchin Aeron, Vaneet Aggarwal, and Xiaodong Wang, *Low-tubal-rank tensor completion using alternating minimization*, Modeling and Simulation for Defense Systems and Applications XI, vol. 9848, International Society for Optics and Photonics, 2016, p. 984809.

References II

Cun Mu, Bo Huang, John Wright, and Donald Goldfarb, *Square deal: Lower bounds and improved relaxations for tensor recovery*, International conference on machine learning, 2014, pp. 73–81.

Holger Rauhut, Reinhold Schneider, and Željka Stojanac, *Low rank tensor recovery via iterative hard thresholding*, Linear Algebra and its Applications **523** (2017), 220–262.