Robust Recovery of Low-rank Tensors from Noisy Sketches

Anna Ma
Assistant Professor
Department of Mathematics
University of California, Irvine

AWM Research Symposium
Special Session on Tensor Methods for Data Modeling

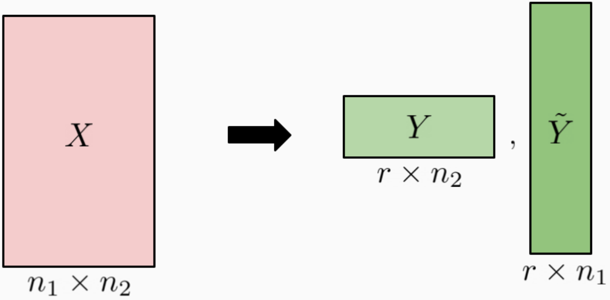
Data, data everywhere

- Data as matrices
- Data is typically too large to store, transport, or process locally.

Problem

Given a large-scale low-rank matrix or tensor, how can one (a) store and (b) retrieve the original matrix?

Matrix Sketching



Idea

Instead of storing large $n_1 \times n_2$ data X, keep lower dimensional sketches of the original data set.

Idea

Instead of storing large $n_1 \times n_2$ data X, keep lower dimensional sketches of the original data set and recover X via[1]

$$\hat{X} = \tilde{Y}^*(S\tilde{Y}^*)^\dagger Y.$$

[1] Fazel, Candes, Recht, and Parrilo, "Compressed sensing and robust recovery of low rank

Noisy Sketches

- Corruptions in data storage, transfer, or sensing
- Noise by design: Differential Privacy[2]

[2] Upadhyay "The Price of Privacy for Low-rank Factorization" (2018).

Problem: Matrix Setting

- Low rank matrix: $X_0 \in \mathbb{C}^{n_1 \times n_2}$: a matrix of rank r_0
- Sketching matrices: $S \in \mathbb{C}^{r \times n_1}, \tilde{S} \in \mathbb{C}^{r \times n_2}$ be two independent complex Gaussian random matrices with $r \geq r_0$
- Noise: $Z \in \mathbb{C}^{r \times n_2}, \tilde{Z} \in \mathbb{C}^{r \times n_1}$, and $\tilde{Z}$ is independent of S.
- Sketches[3]:

$$
\begin{aligned}
Y &= SX_0 + Z \\
\tilde{Y} &= \tilde{S}X_0^* + \tilde{Z}
\end{aligned}
$$

- Recovery:

$$
X = \tilde{Y}^*(S\tilde{Y}^*)^\dagger Y.
$$

[3] Fazel, Candes, Recht, and Parrilo. "*Compressed sensing and robust recovery of low rank matrices*" (2008).

Theorem (M., Stöger, Zhu, 2023)

Let $S \in \mathbb{C}^{r \times n_1}$, $\tilde{S} \in \mathbb{C}^{r \times n_2}$ be two independent standard complex Gaussian matrices. Let $Z \in \mathbb{C}^{r \times n_1}$ be any matrix, and $\tilde{Z} \in \mathbb{C}^{r \times n_2}$ be a matrix such that $(\tilde{S}, \tilde{Z})$ is independent of S, and $\tilde{Y} = \tilde{S} X_0^* + \tilde{Z}$ be almost surely of rank r with $r_0 < r < n_1$. For any $\delta_1, \delta_2, \epsilon > 0$ such that $1 > \delta_2 > \exp\left(-\left(\sqrt{r} - \sqrt{r_0}\right)^2\right)$ and $\epsilon < 1$, with probability at least $1 - \delta_1 - \delta_2 - \epsilon$, the output X satisfies

$$\|X - X_0\|_F \leq \frac{\sqrt{r(n_1 - r)}\|\tilde{Z}\|_F}{\sqrt{\delta_1}(\sqrt{r} - \sqrt{r_0} - \sqrt{\log(1/\delta_2)})} + \frac{\sqrt{r}\|Z\|_F}{\sqrt{\log(1/(1-\epsilon))}}.$$

Matrix Experiments

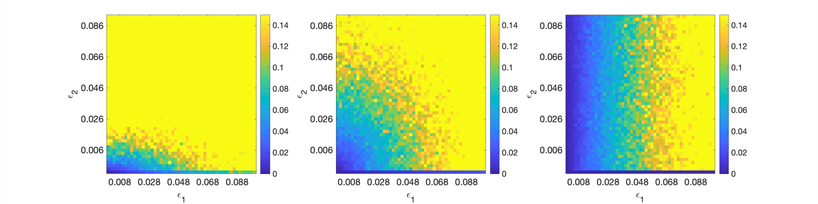


Figure 1: $\|Z\|_F = \varepsilon_1$ and $\|\tilde{Z}\|_F = \varepsilon_2$ for $X \in \mathbb{R}^{100\times 100}$, $r = 10$, and (left) $r = r_0 + 1$ (center) $r = 2r_0$ (right) $r = n - 1$.

Application to Low-rank Matrix Approximation

- Approximately Low rank matrix: $X_0 \in \mathbb{C}^{n_1 \times n_2}$: a matrix of rank r_0
- $X_0 = X_1 + E$ where X_1 is the best rank r_1 approximation of X_0
- Sketches:

$$\begin{aligned} Y &= SX_0 + Z = SX_1 + (SE + Z), \\ \tilde{Y} &= \tilde{S}X_0^* + \tilde{Z} = \tilde{S}X_1^* + (\tilde{S}E^* + \tilde{Z}). \end{aligned}$$

- Recovery:

$$X = \tilde{Y}^*(S\tilde{Y}^*)^\dagger Y.$$

Application: Low-rank Matrix Approximation

Corollary: Low-rank matrix approximation (M., Stöger, Zhu, 2023)

Let $r_1 < r < n_1$. With probability at least $1 - \delta_1 - 3\delta_2 - \epsilon$ and when $\tilde{Y}$ is of rank r, the output X satisfies

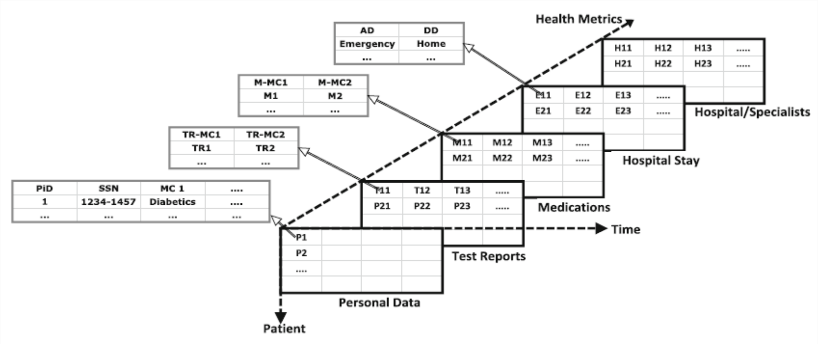
$$\|X - X_0\|_{2\to 2} \leq \sigma_{r_1+1}(X_0)\,\kappa + \sigma,$$

where

$$\kappa = \left(\frac{\sqrt{r(n_1 - r)}(\sqrt{r} + \sqrt{n_2} + \sqrt{\log(1/\delta_2)})}{\sqrt{\delta_1}(\sqrt{r} - \sqrt{r_1} - \sqrt{\log(1/\delta_2)})} + \frac{\sqrt{r}\left(\sqrt{r} + \sqrt{n_1} + \sqrt{\log(1/\delta_2)}\right)}{\sqrt{\log(1/(1-\epsilon))}} + 1 \right)$$

$$\sigma = \frac{\sqrt{r(n_1 - r)}\|\tilde{Z}\|_{2\to 2}}{\sqrt{\delta_1}(\sqrt{r} - \sqrt{r_1} - \sqrt{\log(1/\delta_2)})} + \frac{\sqrt{r}\|Z\|_{2\to 2}}{\sqrt{\log(1/(1-\epsilon))}}.$$

And Tensors too!



- Data as tensors
- Data can (still) be too large to store, transport, or process locally.

Problem: Tensor Setting

- Low-tubal-rank Tensor: $\mathcal{X}_0 \in \mathbb{C}^{n_1 \times n_2 \times n_3}$ with rank r_0
- Sketching tensors: Let $S \in \mathbb{C}^{r \times n_1}, \tilde{S} \in \mathbb{C}^{r \times n_2}$ be two independent complex standard Gaussian random matrices with $r_0 < r < n_1$.
  $\mathcal{S}_1 = S$, $\tilde{\mathcal{S}}_1 = \tilde{S}$ and $\mathcal{S}_k = \tilde{\mathcal{S}}_k = \mathbf{0}$ for all $k \in \{2, \ldots, n_3\}$
- Noise: $\mathcal{Z} \in \mathbb{C}^{r \times n_2 \times n_3}$, $\tilde{\mathcal{Z}} \in \mathbb{C}^{r \times n_1 \times n_3}$
- Sketches:

$$\mathcal{Y} = \mathcal{S} * \mathcal{X}_0 + \mathcal{Z},$$
$$\tilde{\mathcal{Y}} = \tilde{\mathcal{S}} * \mathcal{X}_0^* + \tilde{\mathcal{Z}},$$

- Previous related works:
  - Matricization: Related work considered the recovery of low-tubal-rank tensors through general linear Gaussian measurements of the form $y = A\mathrm{vec}(\mathcal{X})$ - e.g.,
    Lu, Feng, Lin, and Yan. *"Exact low tubal rank tensor recovery from Gaussian measurements"* (2018).
  - Noiseless: Qi and Yu.*"T-singular values and t-sketching for third order tensors"* (2021).

Tensor Set up: t-product

- The t-product was originally introduced in 2011 by Kilmer and Martin for order three tensors, motivated to be a natural extension of matrix multiplication
- Can be efficiently implemented using Fast Fourier Transforms
- Applied to problems in: Image deblurring, face recognition, video compression, and much more
- Has been generalized to higher order tensors and well as different transforms
- Advantage of using t-product: Linear-algebraic like framework (intuitive extensions of notions such as transpose, identity, etc.)
- Trade-off: Orientation dependence

Tensor Set up: Notation

Definition (Operations on tensors)

Let $\mathcal{A} \in \mathbb{C}^{n_1 \times n_2 \times n_3}$. The unfold of a tensor is defined to be the frontal slice stacking of that tensor. In other words,

$$\text{unfold}(\mathcal{A}) = \begin{pmatrix} A_1 \\ A_2 \\ \vdots \\ A_{n_3} \end{pmatrix} \in \mathbb{C}^{n_1 n_3 \times n_2},$$

where $A_i = \mathcal{A}_{:,:,i}$ denotes the i^{th} frontal slice of $\mathcal{A}$. We define the inverse of the unfold$(\cdot)$ as fold$(\cdot)$ so that fold(unfold($\mathcal{A}$)) $= \mathcal{A}$. The block circulant matrix of $\mathcal{A}$ is:

$$\text{bcirc}(\mathcal{A}) = \begin{pmatrix} A_1 & A_{n_3} & A_{n_3-1} & \dots & A_2 \\ A_2 & A_1 & A_{n_3} & \dots & A_3 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ A_{n_3} & A_{n_3-1} & A_{n_3-2} & \dots & A_1 \end{pmatrix} \in \mathbb{C}^{n_1 n_3 \times n_2 n_3}.$$

Tensor Set Up: Notation

Definition (Tensor t-product)

Let $\mathcal{A} \in \mathbb{C}^{n_1 \times \ell \times n_3}$ and $\mathcal{B} \in \mathbb{C}^{\ell \times n_2 \times n_3}$ then the t-product between $\mathcal{A}$ and $\mathcal{B}$, denoted $\mathcal{A} * \mathcal{B}$, is a tensor of size $n_1 \times n_2 \times n_3$ as is computed as:

$$\mathcal{A} * \mathcal{B} = \mathrm{fold}(\mathrm{bcirc}(\mathcal{A})\mathrm{unfold}(\mathcal{B})).$$

Definition (Mode-3 fast Fourier transformation (FFT))

The mode-3 FFT of a tensor $\mathcal{A}$, denoted $\widehat{\mathcal{A}}$, is obtained by applying the discrete Fourier Transform matrix, $F \in \mathbb{C}^{n_3 \times n_3}$, to each $\mathcal{A}_{i,j,:}$ of $\mathcal{A}$:

$$\widehat{\mathcal{A}}_{i,j,:} = F\mathcal{A}_{i,j,:}.$$

Here, F is a unitary matrix, $\mathcal{A}_{i,j,:}$ is an n_3-dimensional vector, and the product is the usual matrix-vector product.

Tensor Set Up: Low tubal rank

Definition (t-SVD)

The Tensor Singular Value Decomposition (t-SVD) of a tensor $\mathcal{M} \in \mathbb{C}^{n_1 \times n_2 \times n_3}$ is given by

$$\mathcal{M} = \mathcal{U} * \mathcal{S} * \mathcal{V}^*,$$

where $\mathcal{U} \in \mathbb{C}^{n_1 \times n_1 \times n_3}$ and $\mathcal{V} \in \mathbb{C}^{n_2 \times n_2 \times n_3}$ are unitary tensors and $\mathcal{S} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ is a tubal tensor (a tensor in which each frontal slice is diagonal), and $*$ denotes the t-product.

Definition (Tubal rank)

The tubal rank of a tensor $\mathcal{M} = \mathcal{U} * \mathcal{S} * \mathcal{V}^*$ is the number of non-zero singular tubes of $\mathcal{S}$.

After performing a mode-3 Fourier transformation (3) on the tensors, the measurements $\widehat{\mathcal{Y}}$ and $\widehat{\tilde{\mathcal{Y}}}$ can be decomposed into n_3 low-rank matrix double sketches:

$$\widehat{\mathcal{Y}}_i = \widehat{\mathcal{S}}_i \widehat{\mathcal{X}_0}_i + \widehat{\mathcal{Z}}, \;\; \widehat{\tilde{\mathcal{Y}}}_i = \widehat{\tilde{\mathcal{S}}}_i * \widehat{\mathcal{X}_0^*}_i + \widehat{\tilde{\mathcal{Z}}}_i, \quad i \in [n_3].$$

The double sketch algorithm outputs

$$\widehat{\mathcal{X}}_i = \widehat{\tilde{\mathcal{Y}}}_i^* (\widehat{\mathcal{S}}_i \widehat{\tilde{\mathcal{Y}}}_i^*)^\dagger \widehat{\mathcal{Y}}_i.$$

Main Result: Tensor Recovery

Corollary: Robust Recovery of Low-Tubal-Rank Tensors (M., Stöger, Zhu, 2023)

If $r_0 < r < n_1$ and for all $k \in [n_3]$, $\hat{\tilde{\mathcal{Y}}}_k$ is of rank r, then for any $\delta_1, \delta_2, \epsilon > 0$ such that $1 > \delta_2 > \exp(-(\sqrt{r} - \sqrt{r_0})^2)$ and $\epsilon < 1$, with probability at least $1 - (\delta_1 + \delta_2 + \epsilon)n_3$,

$$\|\mathcal{X} - \mathcal{X}_0\|_F^2 \leq \frac{2r(n_1 - r)\|\tilde{\mathcal{Z}}\|_F^2}{\delta_1(\sqrt{r} - \sqrt{r_0} - \sqrt{\log(1/\delta_2)})^2} + \frac{2r\|\mathcal{Z}\|_F^2}{\log(1/(1-\epsilon))}.$$

Tensor Experiments

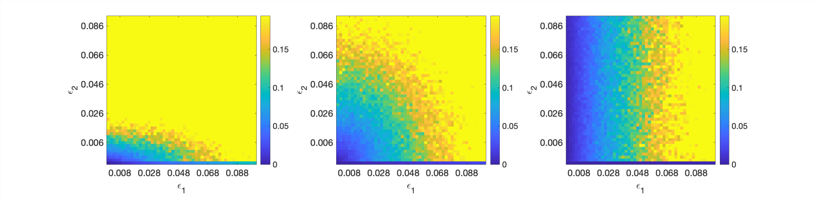


Figure 2: $\|\mathcal{Z}\|_F = \varepsilon_1$ and $\|\tilde{\mathcal{Z}}\|_F = \varepsilon_2$ when $\mathcal{X}_0 \in \mathbb{R}^{100\times100\times10}$, $r_0 = 10$, and (left) $r = r + 1$ (center) $r = 2r_0$ (right) $r = n - 1$.

Tensor Experiments

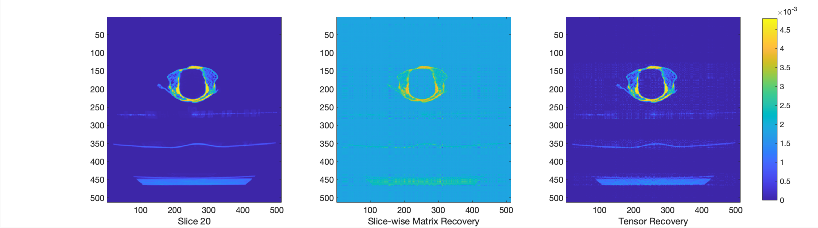


Figure 3: $\mathcal{X} \in \mathbb{R}^{512\times512\times47}$CT scan slices of the C1 vertebrae (left) ground truth 20th slice, (center) naive matrix sketching recovery (right) tensor sketching recovery

Approach	Memory	Error
Tensor Sketching	$rn_3(n_1+n_2)+rn_1$	0.3825
Naive Matrix Sketching I	$rn_3(n_1+n_2)+rn_1n_3$	0.8541
Naive Matrix Sketching II	$rn_3(n_1+n_2)+rn_1$	1.6181

Overview

- Proved first theoretical guarantees characterizing error for noisy double sketches
- Applied our results to obtain guarantees for low-rank matrix approximation using noisy double sketches
- Applied our results to low-tubal-rank tensor recovery using noisy double sketches

Thank you!

"Robust recovery of low-rank matrices and low-tubal-rank tensors from noisy sketches"

Work in collaboration with Dominik Stöger and Yizhe Zhu

arxiv:2206.00803

Github code available!