# Quantization for Low-Rank Matrix Recovery

Eric Lybrand, Rayan Saab

# Overview

# Low Rank Matrix Recovery

Suppose $\mathcal{X} = \{X \in \mathbb{R}^{n_1 \times n_2} : \text{rank}(X) \leq k \ll n_1, n_2\}$

[1] www-bcf.usc.edu/

# Low Rank Matrix Recovery

Suppose $\mathcal{X} = \{X \in \mathbb{R}^{n_1 \times n_2} : \text{rank}(X) \leq k \ll n_1, n_2\}$
Shows up in

[1] www-bcf.usc.edu/

# Low Rank Matrix Recovery

Suppose $\mathcal{X} = \{X \in \mathbb{R}^{n_1 \times n_2} : \text{rank}(X) \leq k \ll n_1, n_2\}$
Shows up in

Global Positioning, Sensor Localization

[1] www-bcf.usc.edu/

# Low Rank Matrix Recovery

Suppose $\mathcal{X} = \{X \in \mathbb{R}^{n_1 \times n_2} : \text{rank}(X) \leq k \ll n_1, n_2\}$
Shows up in

- Global Positioning, Sensor Localization
- Collaborative Filtering

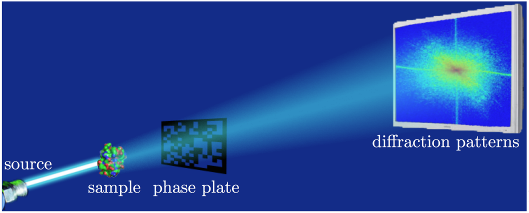
[1] www-bcf.usc.edu/

# Low Rank Matrix Recovery

Suppose $\mathcal{X} = \{X \in \mathbb{R}^{n_1 \times n_2} : \text{rank}(X) \leq k \ll n_1, n_2\}$
Shows up in

Global Positioning, Sensor Localization
Collaborative Filtering
Quantum State Tomography, X-ray Crystallography

$$y_i = |\langle a_i, x\rangle|^2 = \langle a_i a_i^*, xx^*\rangle =: \mathcal{M}(xx^*)$$



[1]

[1] www-bcf.usc.edu/

# Low Rank Matrix Recovery

Natural first guess:

# Low Rank Matrix Recovery

Natural first guess:

$$X^\sharp := \arg\min_Z \operatorname{rank}(Z) \quad \text{subject to } \mathcal{M}(Z) = y$$

# Low Rank Matrix Recovery

Natural first guess:

$$X^\sharp := \arg\min_Z \operatorname{rank}(Z) \quad \text{subject to } \mathcal{M}(Z) = y$$

**Problem:** Solving the above is NP-Hard

# Low Rank Matrix Recovery

Natural first guess:

$$X^\sharp := \arg\min_{Z} \operatorname{rank}(Z) \quad \text{subject to } \mathcal{M}(Z) = y$$

**Problem:** Solving the above is NP-Hard
Take convex relaxation (Maryam Fazel, '02)

# Low Rank Matrix Recovery

Natural first guess:

$$X^\sharp := \arg\min_Z \operatorname{rank}(Z) \quad \text{subject to } \mathcal{M}(Z) = y$$

**Problem:** Solving the above is NP-Hard
Take convex relaxation (Maryam Fazel, '02)

$$X^\sharp := \arg\min_Z \|Z\|_* \quad \text{subject to } \mathcal{M}(Z) = y,$$

$$\|Z\|_* = \sum_{j=1}^{r} \sigma_j(Z)$$

# Nuclear Norm Intuition

Low rank matrices have few singular values, i.e. vector of singular values is sparse

[2] dustingmixon.wordpress.com

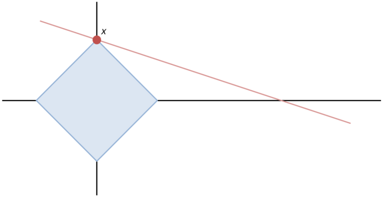
# Nuclear Norm Intuition

Low rank matrices have few singular values, i.e. vector of singular values is sparse

Use $\ell_1$ minimization

$$x^\sharp := \arg\min_Z \|x\|_1 \quad \text{subject to } \mathcal{M}(x) = y,$$

$$\|x\|_1 = \sum_{j=1}^{r} |x_j|$$




[2]

[2]dustingmixon.wordpress.com

# Random $\mathcal{M}$ Work Well

**Theorem (E. Candès, Y. Plan, '10)**

*Suppose $m \geq Ck \max\{n_1, n_2\}$, and let $\mathcal{M}(X) := \sum_{j=1}^{m} \langle A_i, X \rangle$ where $A_i$ are matrices with i.i.d. Gaussian entries. Then with high probability on the draw of $\mathcal{M}$ the following is true for all $X \in \mathbb{R}^{n_1 \times n_2}$ with $\mathrm{rank}(X) \leq k$:*
*$X$ is the unique minimizer of*

$$X^\sharp := \arg\min_Z \|Z\|_* \quad \textit{subject to } \mathcal{M}(Z) = \mathcal{M}(X)$$

# Random $\mathcal{M}$ Work Well

**Theorem (E. Candès, Y. Plan, '10)**

*Suppose $m \geq Ck \max\{n_1, n_2\}$, and let $\mathcal{M}(X) := \sum_{j=1}^{m} \langle A_i, X \rangle$ where $A_i$ are matrices with i.i.d. Gaussian entries. Then with high probability on the draw of $\mathcal{M}$ the following is true for all $X \in \mathbb{R}^{n_1 \times n_2}$ with rank$(X) \leq k$:*
*$X$ is the unique minimizer of*

$$X^\sharp := \arg\min_Z \|Z\|_* \text{ subject to } \mathcal{M}(Z) = \mathcal{M}(X)$$

More generally, linear maps which satisfy the matrix Restricted Isometry Property work well.

# Analog to Digital

Nuclear norm minimization necessitates the use of computers...must store measurements with finitely many bits.

# Analog to Digital

Nuclear norm minimization necessitates the use of computers...must store measurements with finitely many bits.

How should we represent the continuum with a finite set?

# Analog to Digital

Nuclear norm minimization necessitates the use of computers...must store measurements with finitely many bits.

How should we represent the continuum with a finite set?

Are the previous results robust to quantization error?

# MSQ

Suppose we have some finite set (alphabet) $\mathcal{A} \subset \mathbb{R}$ (e.g. $\mathcal{A} = \{\pm 1\}$).

# MSQ

Suppose we have some finite set (alphabet) $\mathcal{A} \subset \mathbb{R}$ (e.g. $\mathcal{A} = \{\pm 1\}$).

**First Idea:** "Round" each $y_i$ and proceed as usual (AKA "Memoryless Scalar Quantization" or MSQ)

# MSQ

Suppose we have some finite set (alphabet) $\mathcal{A} \subset \mathbb{R}$ (e.g. $\mathcal{A} = \{\pm 1\}$).

**First Idea:** "Round" each $y_i$ and proceed as usual (AKA "Memoryless Scalar Quantization" or MSQ)

In the simplest case, take

$$\mathcal{Q} : \mathbb{R} \to \{\pm 1\} \qquad \mathcal{D} : \{\pm 1\} \to \mathbb{R}$$

$$\mathcal{Q}(y) = \text{sign}(y) \qquad \mathcal{D}(q) = q$$

# MSQ

Suppose we have some finite set (alphabet) $\mathcal{A} \subset \mathbb{R}$ (e.g. $\mathcal{A} = \{\pm 1\}$).

**First Idea:** "Round" each $y_i$ and proceed as usual (AKA "Memoryless Scalar Quantization" or MSQ)

In the simplest case, take

$$\mathcal{Q} : \mathbb{R} \to \{\pm 1\} \qquad \mathcal{D} : \{\pm 1\} \to \mathbb{R}$$
$$\mathcal{Q}(y) = \operatorname{sign}(y) \qquad \mathcal{D}(q) = q$$

Control error in recovering $X$ by increasing size of $\mathcal{A}$ (resp. bits) so that $\mathcal{D} \circ \mathcal{Q}(y) \approx y$.

# MSQ

Suppose we have some finite set (alphabet) $\mathcal{A} \subset \mathbb{R}$ (e.g. $\mathcal{A} = \{\pm 1\}$).

**First Idea:** "Round" each $y_i$ and proceed as usual (AKA "Memoryless Scalar Quantization" or MSQ)

In the simplest case, take

$$\mathcal{Q} : \mathbb{R} \to \{\pm 1\} \qquad \mathcal{D} : \{\pm 1\} \to \mathbb{R}$$
$$\mathcal{Q}(y) = \mathrm{sign}(y) \qquad \mathcal{D}(q) = q$$

Control error in recovering $X$ by increasing size of $\mathcal{A}$ (resp. bits) so that $\mathcal{D} \circ \mathcal{Q}(y) \approx y$.

**Problem:** It could be expensive to increase the number of bits used

# Oversampling

If the number of bits is fixed, try taking more measurements

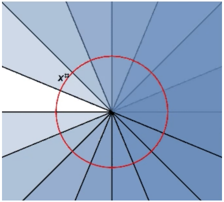
**Intuition:** Measurements $\text{sign}(\langle A_j, X\rangle)$ defines a half-space $X$ lies in.

Minimizing quantization error $\iff$ minimizing volumes

# The Shortcomings of MSQ

Volume of regions (i.e. reconstruction error) decay like $m^{-1}$

Vivek Goyal et al ('98): reconstruction error from MSQ quantized frame coefficients can't decay faster than $O(m^{-1})$.

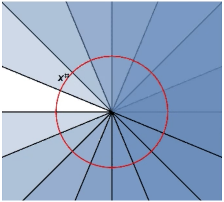


# The Shortcomings of MSQ

Volume of regions (i.e. reconstruction error) decay like $m^{-1}$

Vivek Goyal et al ('98): reconstruction error from MSQ quantized frame coefficients can't decay faster than $O(m^{-1})$.

Candés, Romberg, and Tao (2005): for sparse $x$ if $\|y - q\|_2 \leq \varepsilon$, then $\|x - \hat{x}\|_2 \leq \frac{c_1}{\sqrt{m}}\varepsilon$.



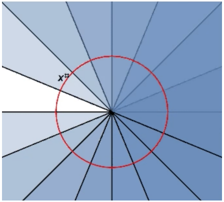
# The Shortcomings of MSQ

Volume of regions (i.e. reconstruction error) decay like $m^{-1}$

Vivek Goyal et al ('98): reconstruction error from MSQ quantized frame coefficients can't decay faster than $O(m^{-1})$.

Candés, Romberg, and Tao (2005): for sparse $x$ if $\|y - q\|_2 \leq \varepsilon$, then $\|x - \hat{x}\|_2 \leq \frac{c_1}{\sqrt{m}}\varepsilon$.

Alphabet resolution $\beta \implies \|y - q\|_2 \leq \sqrt{m}\beta \implies \|x - \hat{x}\|_2 \leq c_1\beta$ i.e. the error bound does not decrease with $m$.



# ΣΔ Quantization

Proposed by Inose & Yasuda, 1963 for quantizing bandlimited functions

Keeps track of $r$ previous quantization errors in a state variable $u$ to "shape" the quantized values

$$q_i = \mathcal{Q}(\rho_r(u_{i-1}, \dots, u_{i-r}, y_i, \dots, y_{i-r+1}))$$

$$D^r u = y - q, \quad (Du)_i = u_i - u_{i-1}$$

# ΣΔ Quantization

Proposed by Inose & Yasuda, 1963 for quantizing bandlimited functions

Keeps track of $r$ previous quantization errors in a state variable $u$ to "shape" the quantized values

$$q_i = \mathcal{Q}(\rho_r(u_{i-1}, \dots, u_{i-r}, y_i, \dots, y_{i-r+1}))$$
$$D^r u = y - q, \quad (Du)_i = u_i - u_{i-1}$$

For example, when $r = 1$,

$$q_i = \mathcal{Q}(y_i + u_{i-1})$$
$$u_i = u_{i-1} + y_i - q_i.$$

# ΣΔ Quantization

Proposed by Inose & Yasuda, 1963 for quantizing bandlimited functions

Keeps track of $r$ previous quantization errors in a state variable $u$ to "shape" the quantized values

$$q_i = \mathcal{Q}(\rho_r(u_{i-1}, \ldots, u_{i-r}, y_i, \ldots, y_{i-r+1}))$$
$$D^r u = y - q, \quad (Du)_i = u_i - u_{i-1}$$

For example, when $r = 1$,

$$q_i = \mathcal{Q}(y_i + u_{i-1})$$
$$u_i = u_{i-1} + y_i - q_i.$$

For example, could use equispaced grid where for some fixed $L > 0$ and resolution $\beta > 0$

$$\mathcal{A} := \{\pm(j - 1/2)\beta, \ j \in [L]\}.$$

# ΣΔ Quantization

Critically important that for a given $\mathcal{A}$, $\rho_r$ is chosen so that $\|u\|_\infty < \gamma(r)$ (Stability)

# ΣΔ Quantization

Critically important that for a given $\mathcal{A}$, $\rho_r$ is chosen so that $\|u\|_\infty < \gamma(r)$ (Stability)

- Daubechies & Devore (2003): first provably stable family for bandlimited functions

# $\Sigma\Delta$ Quantization

Critically important that for a given $\mathcal{A}$, $\rho_r$ is chosen so that $\|u\|_\infty < \gamma(r)$ (Stability)

> Daubechies & Devore (2003): first provably stable family for bandlimited functions

Trade off between bit complexity of alphabet and stability constant.

# A More General View of Noise Shaping

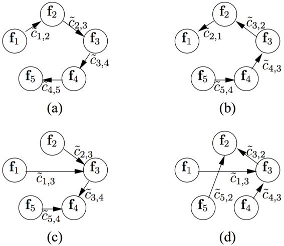
$\Sigma\Delta$ pushes quantization error of previous measurements forward "in time."

[3] P. T. Boufounos, "Quantization and erasures in frame representations."

# A More General View of Noise Shaping

$\Sigma\Delta$ pushes quantization error of previous measurements forward "in time."

More general noise shaping could involving pushing quantization error for $\ell^{th}$ coefficient to the $\ell_k^{th}$ coefficient to compensate (Boufounos, 2006).




[3]

[3] P. T. Boufounos, "Quantization and erasures in frame representations."

# The Perks of Noise Shaping: Sparse Vectors

**Theorem (R. Saab, R. Wang, O. Yilmaz, 2015)**

*Let $A \in \mathbb{R}^{m\times N}$ be a Gaussian matrix with $m \geq C_1 k \log(eN/k)$. Then with high probability the following is true for any $k$-sparse $x \in \mathbb{R}^N$: let $q = \mathcal{Q}_{\Sigma\Delta}^{(r)}(Ax)$. The solution*

$$\hat{x} := \arg\min_z \|z\|_1 \quad s.t. \ \|D^{-r}(Az - q)\|_2 \leq \gamma(r)\sqrt{m}$$

*satisfies*

$$\|\hat{x} - x\|_2 \leq C_2 \beta \left(\frac{m}{\ell}\right)^{-r+1/2}.$$

# The Perks of Noise Shaping: Sparse Vectors

**Theorem (R. Saab, R. Wang, O. Yilmaz, 2015)**

*Let $A \in \mathbb{R}^{m \times N}$ be a Gaussian matrix with $m \geq C_1 k \log(eN/k)$. Then with high probability the following is true for any $k$-sparse $x \in \mathbb{R}^N$: let $q = \mathcal{Q}^{(r)}_{\Sigma\Delta}(Ax)$. The solution*

$$\hat{x} := \arg\min_z \|z\|_1 \quad s.t. \quad \|D^{-r}(Az - q)\|_2 \leq \gamma(r)\sqrt{m}$$

*satisfies*

$$\|\hat{x} - x\|_2 \leq C_2 \beta \left(\frac{m}{\ell}\right)^{-r+1/2}.$$

Result is stable w.r.t noise.

# The Perks of Noise Shaping: Sparse Vectors

**Theorem (R. Saab, R. Wang, O. Yilmaz, 2015)**

*Let $A \in \mathbb{R}^{m\times N}$ be a Gaussian matrix with $m \geq C_1 k \log(eN/k)$. Then with high probability the following is true for any $k$-sparse $x \in \mathbb{R}^N$: let $q = \mathcal{Q}^{(r)}_{\Sigma\Delta}(Ax)$. The solution*

$$\hat{x} := \arg\min_z \|z\|_1 \quad s.t. \ \|D^{-r}(Az - q)\|_2 \leq \gamma(r)\sqrt{m}$$

*satisfies*

$$\|\hat{x} - x\|_2 \leq C_2 \beta \left(\frac{m}{\ell}\right)^{-r+1/2}.$$

Result is stable w.r.t noise.

Result is robust to sparsity assumption.

# Generalizing to Matrices

## Theorem (E.L. and R. Saab, 2018)

*Let $\mathcal{M} : \mathbb{R}^{n_1 \times n_2} \to \mathbb{R}^m$ be a sub-Gaussian linear map.*

# Generalizing to Matrices

## Theorem (E.L. and R. Saab, 2018)

*Let $\mathcal{M} : \mathbb{R}^{n_1 \times n_2} \to \mathbb{R}^m$ be a sub-Gaussian linear map. If $m \geq \ell \geq c_1 k \max\{n_1, n_2\}$ then w.h.p. on the draw of $\mathcal{M}$ the following holds uniformly for all $X \in \mathbb{R}^{n_1 \times n_2}$:*

# Generalizing to Matrices

## Theorem (E.L. and R. Saab, 2018)

*Let $\mathcal{M} : \mathbb{R}^{n_1 \times n_2} \to \mathbb{R}^m$ be a sub-Gaussian linear map. If $m \geq \ell \geq c_1 k \max\{n_1, n_2\}$ then w.h.p. on the draw of $\mathcal{M}$ the following holds uniformly for all $X \in \mathbb{R}^{n_1 \times n_2}$: let $q = \mathcal{Q}^{(r)}_{\Sigma\Delta}(\mathcal{M}(X) + \eta)$ with $\|\eta\|_\infty \leq \varepsilon$.*

# Generalizing to Matrices

## Theorem (E.L. and R. Saab, 2018)

*Let $\mathcal{M} : \mathbb{R}^{n_1 \times n_2} \to \mathbb{R}^m$ be a sub-Gaussian linear map. If $m \geq \ell \geq c_1 k \max\{n_1, n_2\}$ then w.h.p. on the draw of $\mathcal{M}$ the following holds uniformly for all $X \in \mathbb{R}^{n_1 \times n_2}$: let $q = \mathcal{Q}^{(r)}_{\Sigma\Delta}(\mathcal{M}(X) + \eta)$ with $\|\eta\|_\infty \leq \varepsilon$. Define*

$$(X^\sharp, \nu^\sharp) := \arg\min_{(Z,\nu)} \|Z\|_* \quad s.t. \quad \|D^{-r}(\mathcal{M}(Z) + \nu - q)\|_2 \leq \gamma(r)\sqrt{m}$$
$$\text{and} \quad \|\nu\|_2 \leq \varepsilon\sqrt{m}.$$

# Generalizing to Matrices

## Theorem (E.L. and R. Saab, 2018)

*Let $\mathcal{M} : \mathbb{R}^{n_1 \times n_2} \to \mathbb{R}^m$ be a sub-Gaussian linear map. If $m \geq \ell \geq c_1 k \max\{n_1, n_2\}$ then w.h.p. on the draw of $\mathcal{M}$ the following holds uniformly for all $X \in \mathbb{R}^{n_1 \times n_2}$: let $q = \mathcal{Q}^{(r)}_{\Sigma\Delta}(\mathcal{M}(X) + \eta)$ with $\|\eta\|_\infty \leq \varepsilon$. Define*

$$(X^\sharp, \nu^\sharp) := \arg\min_{(Z,\nu)} \|Z\|_* \quad s.t. \quad \|D^{-r}(\mathcal{M}(Z) + \nu - q)\|_2 \leq \gamma(r)\sqrt{m}$$
$$\text{and} \quad \|\nu\|_2 \leq \varepsilon\sqrt{m}.$$

*Then $X^\sharp$ satisfies*

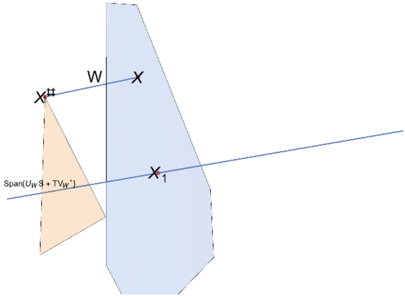
$$\|X^\sharp - X\|_F \lesssim_r \left(\frac{m}{\ell}\right)^{-r+1/2} \beta + \frac{\sigma_k(X)_*}{\sqrt{k}} + \sqrt{\frac{m}{\ell}}\epsilon.$$

# Proof Sketch

**Goal:** Control $\|X^\sharp - X\|_F := \|W\|_F$

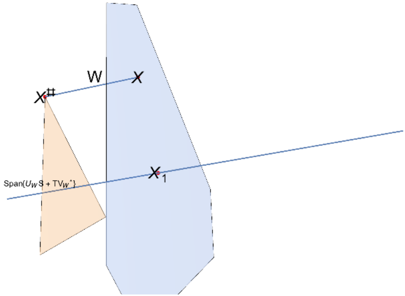
# Proof Sketch

**Goal:** Control $\|X^\sharp - X\|_F := \|W\|_F$
Non-commutativity makes things difficult. Try and reduce it to the vector setting.



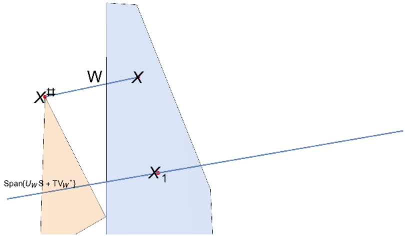
# Proof Sketch

**Goal:** Control $\|X^\sharp - X\|_F := \|W\|_F$
Non-commutativity makes things difficult. Try and reduce it to the vector setting.
**New Goal:** Formulate a corresponding vector optimization problem where error between minimizer and truth is $\|W\|_F$.



# Proof Sketch

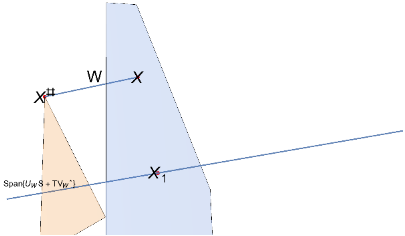
Let $W := U_W \Sigma_W V_W^*$, and set $X_1 := -U_W \Sigma_X V_W^*$



# Proof Sketch

Let $W := U_W \Sigma_W V_W^*$, and set $X_1 := -U_W \Sigma_X V_W^*$

Define $\mathcal{M}_{U_W, V_W}(x) := \mathcal{M}(U_W \operatorname{diag}(x) V_W^*)$, and
$y := D^{-r} \left( M_{U_W, V_W}(-\vec{\sigma}(X)) + e \right) + u$.

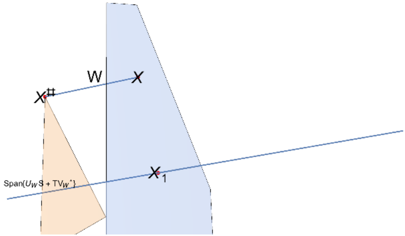


# Proof Sketch

Let $W := U_W \Sigma_W V_W^*$, and set $X_1 := -U_W \Sigma_X V_W^*$

Define $\mathcal{M}_{U_W, V_W}(x) := \mathcal{M}(U_W \operatorname{diag}(x) V_W^*)$, and
$y := D^{-r}\left(M_{U_W, V_W}(-\vec{\sigma}(X)) + e\right) + u$.

Show $\vec{\sigma}(W) - \vec{\sigma}(X)$ is feasible to the vector optimization problem with $A = M_{U_W, V_W}$ and $D^{-r} q = y$.





UC San Diego

# Proof Sketch

Use a lemma from Oymak et al (2011) which buys us

$$\|\vec{\sigma}(W) - \vec{\sigma}(X)\|_1 = \|X_1 + W\|_* \leq \|X_1\|_* = \|\vec{\sigma}(X)\|_1$$

# Proof Sketch

Use a lemma from Oymak et al (2011) which buys us

$$\|\vec{\sigma}(W) - \vec{\sigma}(X)\|_1 = \|X_1 + W\|_* \leq \|X_1\|_* = \|\vec{\sigma}(X)\|_1$$

All that's left is to show that $\frac{1}{\sqrt{\ell}} P_\ell V^* M_{U_W, V_W}$ satisfies the RIP for all unitary $U_W, V_W$, as then we can invoke the theorem for vector recovery.

# Proof Sketch

Use a lemma from Oymak et al (2011) which buys us

$$\|\vec{\sigma}(W) - \vec{\sigma}(X)\|_1 = \|X_1 + W\|_* \leq \|X_1\|_* = \|\vec{\sigma}(X)\|_1$$

All that's left is to show that $\frac{1}{\sqrt{\ell}} P_\ell V^* M_{U_W, V_W}$ satisfies the RIP for all unitary $U_W, V_W$, as then we can invoke the theorem for vector recovery.

We use the chaining technique as proposed by Talagrand.

# Root Exponential Accuracy

## Corollary (E.L. and R. Saab, 2018)

*Let* $q = \mathcal{Q}^{(r)}_{\Sigma\Delta}(\mathcal{M}(X))$ *denote quantized noiseless measurements and* $X \in \mathbb{R}^{n_1 \times n_2}$ *with* $\mathrm{rank}(X) = k$. *Then there exist constants* $c, c_1, C_1, C_2 > 0$ *so that when*

$$\lambda := \frac{m}{\lceil ck \max(n_1, n_2) \rceil}$$

$$r := \left\lfloor \frac{\lambda}{2C_1 e} \right\rfloor^{1/2}$$

$$q := Q^{r}_{\Sigma\Delta}(\mathcal{M}(X)).$$

*the minimizer* $X^\sharp$ *satisfies* $\|X^\sharp - X\|_F \lesssim \beta e^{-c_1\sqrt{\lambda}}$.

# Exponential Accuracy with Random Encoding

**Corollary (E.L. and R. Saab, 2018)**

*Let $B : \mathbb{R}^m \to \mathbb{R}^L$ be a Bernoulli random matrix whose entries are $\pm 1$. Whenever $m \gtrsim_r L \gtrsim_r k \max(n_1, n_2)$ the following is true w.h.p. on the draw of $\mathcal{M}$ and $B$: the solution of*

$$(\hat{X}, \hat{\nu}) := \arg\min_{(Z,\nu)} \|Z\|_* \quad s.t. \quad \|BD^{-r}(\mathcal{M}(Z) + \nu - q)\|_2 \leq 3m\gamma(r)$$

$$\text{and} \quad \|\nu\|_2 \leq \epsilon\sqrt{m}.$$

*satisfies*

$$\|\hat{X} - X\|_F \lesssim \left(\frac{m}{L}\right)^{-r/2+3/4} \beta + \frac{\sigma_k(X)_*}{\sqrt{k}} + \sqrt{\frac{m}{L}}\varepsilon.$$

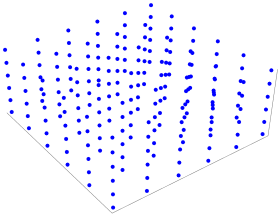
# Exponential Accuracy with Random Encoding

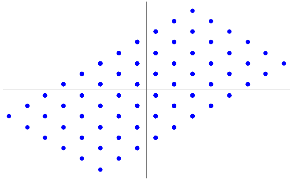
For noiseless measurements of rank $k$ matrices, this means reconstruction error decays exponentially w.r.t. rate (number of bits).

# Exponential Accuracy with Random Encoding

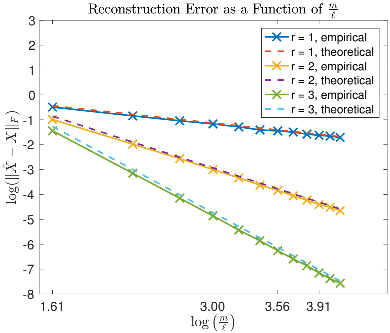
For noiseless measurements of rank $k$ matrices, this means reconstruction error decays exponentially w.r.t. rate (number of bits).

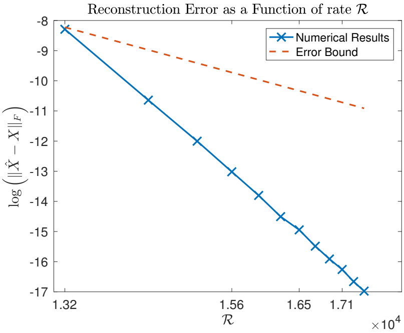
Random encoding "reduces complexity" of alphabet $\mathcal{A}$.





# Numerical Illustrations




Experimental DL: reconstruct rank 5, $20 \times 20$ Gaussian matrices from noiseless Gaussian measurements, averaged over 20 draws of true matrix.

# Future Directions

Taking sub-gaussian measurements is, in general, slow. Do the results hold for partial random circulant matrices, etc?

# Future Directions

Taking sub-gaussian measurements is, in general, slow. Do the results hold for partial random circulant matrices, etc?

How can we modify these results to apply in the matrix completion setting?

# Future Directions

Taking sub-gaussian measurements is, in general, slow. Do the results hold for partial random circulant matrices, etc?

How can we modify these results to apply in the matrix completion setting?

Experiments show the exponent for noiseless encoding bound

$$\|\hat{X} - X\|_F \lesssim \left(\frac{m}{L}\right)^{-r/2+3/4}$$

is sub-optimal. Can we prove that it holds with the bound $\left(\frac{m}{L}\right)^{-r+3/4}$?

*Fin*

## Theorem (R. Saab, R. Wang, O. Yilmaz, 2015)

*Let $A \in \mathbb{R}^{m\times N}$, $P_\ell : \mathbb{R}^m \to \mathbb{R}^m$ the projection onto the first $\ell$ coordinates, and $V^*$ as in the singular value decomposition of $D^{-r}$. Suppose that $\frac{1}{\sqrt{\ell}} P_\ell V^* A$ has the vector-RIP of order $k$ and constant $\delta_k < 1/9$. Then any feasible $\hat{x}$ of*

$$(\hat{x}, \hat{\nu}) := \arg\min_{(z,\nu)} \|z\|_1 \quad s.t. \;\; \|D^{-r}(Az + \nu - q)\|_2 \leq \gamma(r)\sqrt{m}$$
$$\text{and} \;\; \|\nu\|_2 \leq \varepsilon$$

*with $\|\hat{x}\|_1 \leq \|x\|_1$ and $q$ satisfying $Ax + e - D^r u = q$ with $\|u\|_\infty \leq \gamma(r) < \infty$ and $\|e\|_2 \leq \varepsilon$ satisfies*

$$\|\hat{x} - x\|_2 \leq C\left( \left(\frac{m}{\ell}\right)^{-r+1/2} \beta + \frac{\sigma_k(x)_1}{\sqrt{k}} + \sqrt{\frac{m}{\ell}}\epsilon \right),$$

*where $\sigma_k(x)_1 = \text{argmin} \ldots \|x - z\|_1$*

# A Stronger RIP

## Definition

A linear map $\mathcal{M} : \mathbb{R}^{n_1 \times n_2} \to \mathbb{R}^m$ satisfies the matrix RIP of order $k$ with constant $\delta_k$ if for any matrix $X$ with $\text{rank}(X) \leq k$

$$(1 - \delta_k)\|X\|_F^2 \leq \|\mathcal{M}(X)\|_2^2 \leq (1 + \delta_k)\|X\|_F^2$$

# A Stronger RIP

## Definition

A linear map $\mathcal{M} : \mathbb{R}^{n_1 \times n_2} \to \mathbb{R}^m$ satisfies the matrix RIP of order $k$ with constant $\delta_k$ if for any matrix $X$ with $\text{rank}(X) \leq k$

$$(1 - \delta_k)\|X\|_F^2 \leq \|\mathcal{M}(X)\|_2^2 \leq (1 + \delta_k)\|X\|_F^2$$

## Lemma (S. Oymak, K. Mohan, M. Fazel, B. Hassibi, 2011)

*If $\mathcal{M}$ satisfies the matrix RIP of order $k$ with constant $\delta_k$, then for any unitary matrices $U, V$ the linear map $\mathcal{M}_{U,V}$ satisfies the vector RIP of order $k$ with constant $\delta_k$.*

# Proof Sketch

So it suffices to show that the linear map $\frac{1}{\sqrt{\ell}}P_\ell V^*\mathcal{M}$ satisfies the matrix RIP.

# Proof Sketch

So it suffices to show that the linear map $\frac{1}{\sqrt{\ell}} P_\ell V^* \mathcal{M}$ satisfies the matrix RIP.

Consider the stochastic process

$$Z_X := \left| \frac{1}{\ell} \|P_\ell V^* \mathcal{M}(X)\|_F^2 - \|X\|_F^2 \right|$$

# Proof Sketch

So it suffices to show that the linear map $\frac{1}{\sqrt{\ell}} P_\ell V^* \mathcal{M}$ satisfies the matrix RIP.

Consider the stochastic process

$$Z_X := \left| \frac{1}{\ell} \| P_\ell V^* \mathcal{M}(X) \|_F^2 - \|X\|_F^2 \right|$$

**Goal:** Control

$$\mathbb{P}\left( \sup_X Z_X \geq t \right)$$

# The Generic Chaining

**Motivating Idea:** Suppose that $X$ were drawn from a finite set $T$. We could always union bound:

$$\mathbb{P}\left(\sup_{X\in T} Z_X \geq t\right) \leq \sum_{X\in T} \mathbb{P}\left(Z_X \geq t\right)$$

# The Generic Chaining

**Motivating Idea:** Suppose that $X$ were drawn from a finite set $T$. We could always union bound:

$$\mathbb{P}\left(\sup_{X \in T} Z_X \geq t\right) \leq \sum_{X \in T} \mathbb{P}\left(Z_X \geq t\right)$$

This upper bound will be too pessimistic if the $Z_X$ are correlated.

## The Generic Chaining

**Motivating Idea:** Suppose that $X$ were drawn from a finite set $T$. We could always union bound:

$$\mathbb{P}\left(\sup_{X\in T} Z_X \geq t\right) \leq \sum_{X\in T} \mathbb{P}\left(Z_X \geq t\right)$$

This upper bound will be too pessimistic if the $Z_X$ are correlated.

Michel Talagrand (1996) established a technique which cleverly "groups" correlated draws of $Z_X$ to make union bounding effective.

# The Generic Chaining

**Motivating Idea:** Suppose that $X$ were drawn from a finite set $T$. We could always union bound:

$$\mathbb{P}\left(\sup_{X\in T} Z_X \geq t\right) \leq \sum_{X\in T} \mathbb{P}\left(Z_X \geq t\right)$$

This upper bound will be too pessimistic if the $Z_X$ are correlated.

Michel Talagrand (1996) established a technique which cleverly “groups” correlated draws of $Z_X$ to make union bounding effective.

Built off of an increment property: it is assumed that there exists a metric $d$ so that

$$\mathbb{P}\left(|Z_X - Z_Y| \geq t\right) \leq 2\exp\left(\frac{-t^2}{d^2(X,Y)}\right).$$

# The Generic Chaining

Successively approximate $Z_X$ by

$$Z_X = Z_X - Z_{\pi_1(X)} + Z_{\pi_1(X)} = Z_X - \sum_j Z_{\pi_j(X)} - Z_{\pi_{j-1}(X)}$$

where $\pi_j$ projects $T$ onto some finite subset $T_j \subset T$.
Intuitively, elements in the fiber $\pi_j^{-1}(t)$ "are the same".

# The Generic Chaining

Successively approximate $Z_X$ by

$$Z_X = Z_X - Z_{\pi_1(X)} + Z_{\pi_1(X)} = Z_X - \sum_j Z_{\pi_j(X)} - Z_{\pi_{j-1}(X)}$$

where $\pi_j$ projects $T$ onto some finite subset $T_j \subset T$. Intuitively, elements in the fiber $\pi_j^{-1}(t)$ "are the same".

Use the increment property on each of the residuals $Z_{\pi_j(X)} - Z_{\pi_{j-1}(X)}$ and union bound over the fibers of $\pi_j$.

# Results Using Chaining

Unsurprisingly, the geometry induced on $T$ by the metric $d$ will govern the tail bound

# Results Using Chaining

Unsurprisingly, the geometry induced on $T$ by the metric $d$ will govern the tail bound

A result by Krahmer, Mendelson, and Rauhut (2013) using chaining allows us to bound the deviation of

$$\left| \frac{1}{\ell} \| P_\ell V^* \mathcal{M}(X) \|_F^2 - \|X\|_F^2 \right|$$

in terms of the "sizes" of the set

$$\{X \in \mathbb{R}^{n_1 \times n_2} : \|X\|_F = 1,\ \operatorname{rank}(X) \leq k\}.$$

# Results Using Chaining

Unsurprisingly, the geometry induced on $T$ by the metric $d$ will govern the tail bound

A result by Krahmer, Mendelson, and Rauhut (2013) using chaining allows us to bound the deviation of

$$\left| \frac{1}{\ell} \| P_\ell V^* \mathcal{M}(X) \|_F^2 - \|X\|_F^2 \right|$$

in terms of the "sizes" of the set

$$\{X \in \mathbb{R}^{n_1 \times n_2} : \|X\|_F = 1,\ \operatorname{rank}(X) \leq k\}.$$

The low dimensionality of the above set is what allows us to undersample and obtain the matrix RIP.