Experiment Constrained Subset Selection based on Randomized QR with Column Pivoting

A Tensor $\star_{\mathbf{M}}$ Product Framework

Mitchell Scott

Department of Mathematics, Emory University

1st Annual SIAM New England Sectional Meeting

SIAM NE 2026

September 13, 2026

EMORY UNIVERSITY

Collaborators and Acknowledgments

Riley Murray, Sandia NL

Jamie Haddock, Harvey Mudd

Irene Simó Muñoz, Cornell University

Tanya Tafolla, UC, Merced

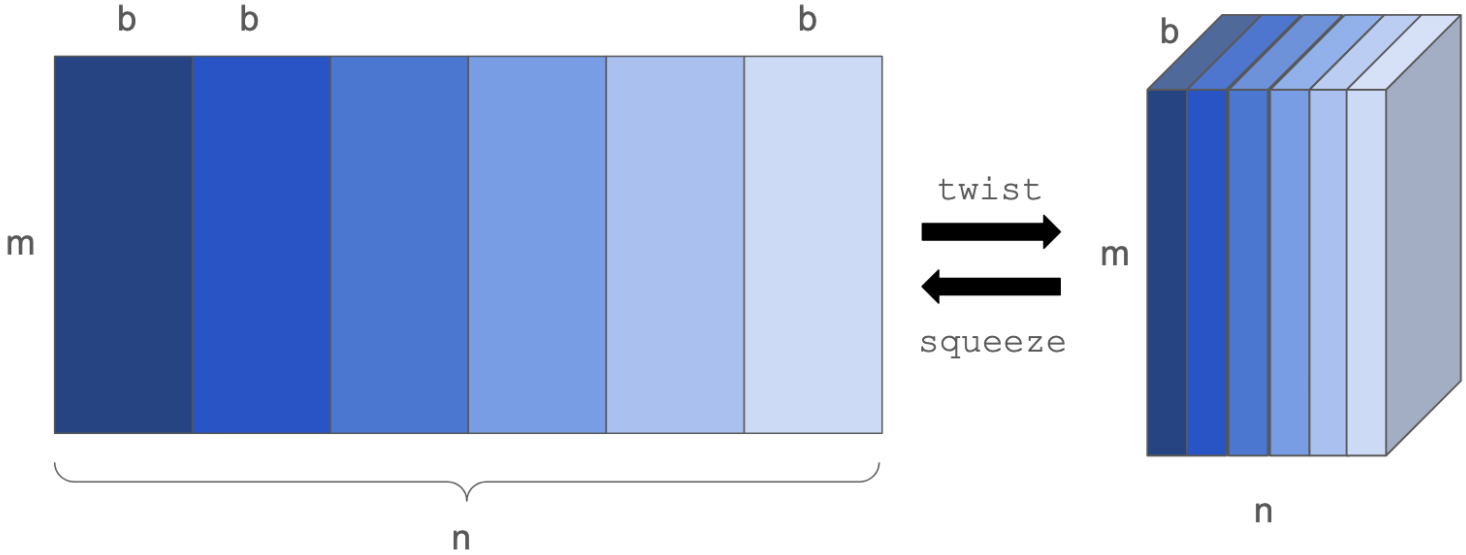
Part of this research was performed while the author was visiting the Institute for Pure and Applied Mathematics (IPAM), which is supported by the National Science Foundation (Grant No. DMS-1925919).

Motivation: Optimal Experiment Design

Goal: Find the best way to acquire more data, i.e. optimal experiment design (OED).

We are concerned with blocks because

- m is the number of model parameters,
- b is the number of time points per experiment, and
- $n = |\mathcal{C}|$ is the number of candidate experiments.



A candidate's b rows arrive together — you buy a whole experiment or none of it.

From Fisher information to block volume

Definition (Fisher Information Matrix)

For m parameters $\boldsymbol{\theta}$ and density $f(X;\boldsymbol{\theta})$,

$$[\mathbf{F}(c,\boldsymbol{\theta})]_{i,j} = \mathbb{E}\left[\left(\frac{\partial}{\partial\theta_i}\log f(X;\boldsymbol{\theta})\right)\left(\frac{\partial}{\partial\theta_j}\log f(X;\boldsymbol{\theta})\right)\,\middle|\,\boldsymbol{\theta}\right]$$

From Fisher information to block volume

Definition (Fisher Information Matrix)

For m parameters $\boldsymbol{\theta}$ and density $f(X;\boldsymbol{\theta})$,

$$[\mathbf{F}(c,\boldsymbol{\theta})]_{i,j} = \mathbb{E}\left[\left(\frac{\partial}{\partial\theta_i}\log f(X;\boldsymbol{\theta})\right)\left(\frac{\partial}{\partial\theta_j}\log f(X;\boldsymbol{\theta})\right)\,\middle|\,\boldsymbol{\theta}\right]$$

Gaussian noise + Gaussian prior $\Rightarrow$ after whitening,

$$\mathbf{F}(c,\boldsymbol{\theta}) \;=\; \sum_{i\in\mathcal{S}}\mathbf{J}_i^\top\mathbf{J}_i + \mathbf{I}_m \;=\; \mathbf{J}_{\mathcal{S}}^\top\mathbf{J}_{\mathcal{S}} + \mathbf{I}_m, \qquad \mathbf{J}_i = \left.\frac{\partial\mathbf{g}_i}{\partial\boldsymbol{\theta}}\right|_{\boldsymbol{\theta}_0} \in \mathbb{R}^{b\times m}$$

From Fisher information to block volume

Definition (Fisher Information Matrix)

For m parameters $\boldsymbol{\theta}$ and density $f(X;\boldsymbol{\theta})$,

$$[\mathbf{F}(c,\boldsymbol{\theta})]_{i,j} = \mathbb{E}\left[\left(\frac{\partial}{\partial\theta_i}\log f(X;\boldsymbol{\theta})\right)\left(\frac{\partial}{\partial\theta_j}\log f(X;\boldsymbol{\theta})\right)\,\middle|\,\boldsymbol{\theta}\right]$$

Gaussian noise + Gaussian prior $\Rightarrow$ after whitening,

$$\mathbf{F}(c,\boldsymbol{\theta}) \;=\; \sum_{i\in\mathcal{S}}\mathbf{J}_i^\top\mathbf{J}_i + \mathbf{I}_m \;=\; \mathbf{J}_\mathcal{S}^\top\mathbf{J}_\mathcal{S} + \mathbf{I}_m, \qquad \mathbf{J}_i = \left.\frac{\partial\mathbf{g}_i}{\partial\boldsymbol{\theta}}\right|_{\boldsymbol{\theta}_0} \in \mathbb{R}^{b\times m}$$

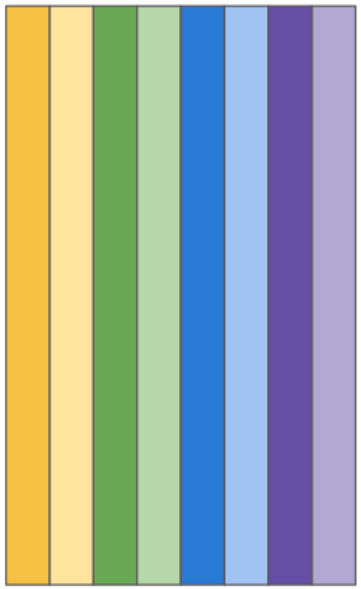
D-optimal design: $\max\limits_{|\mathcal{S}|=k}\ \phi\big(\mathbf{F}(c,\boldsymbol{\theta})\big) = \max\limits_{|\mathcal{S}|=k}\ \frac{1}{2}\log\det\mathbf{F}(c,\boldsymbol{\theta})$

EMORY UNIVERSITY

Existing Block QRCP is inadequate: a toy example

Let $\mathbf{A} = [\mathbf{A}_1\ \mathbf{A}_2\ \mathbf{A}_3\ \mathbf{A}_4]$, where $\mathbf{A}_i \in \mathbb{R}^{m\times 2}$, $\|\mathbf{A}_i(:,1)\| = 1$, $\|\mathbf{A}_i(:,2)\| = \varepsilon$, for all $i = 1, \ldots, 4$.

Goal: choose three blocks (six rows) such that we maximize $\mathrm{vol}([\mathbf{A}_{\pi(1)}\ \mathbf{A}_{\pi(2)}\ \mathbf{A}_{\pi(3)}])$

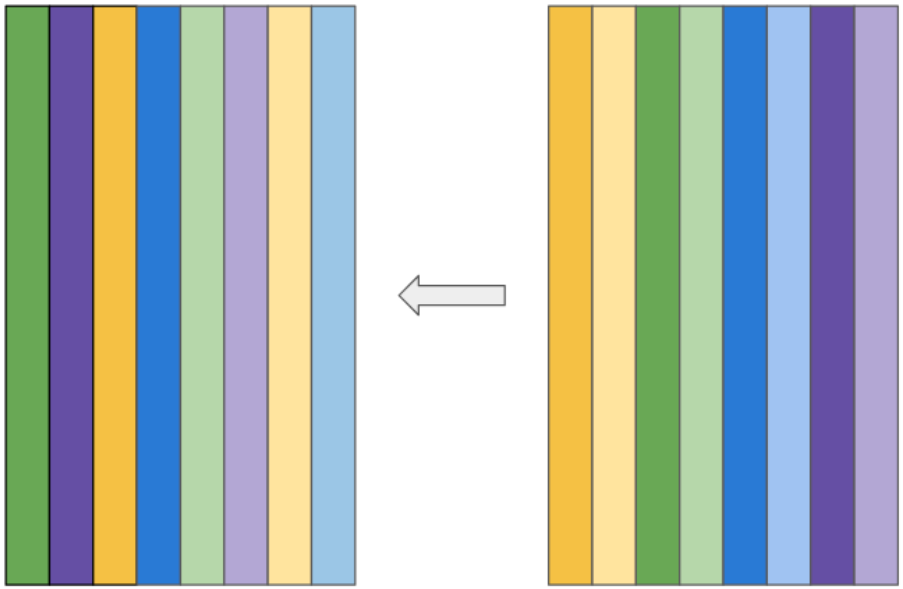


Block matrix

Existing Block QRCP is inadequate: a toy example

Let $\mathbf{A} = [\mathbf{A}_1\ \mathbf{A}_2\ \mathbf{A}_3\ \mathbf{A}_4]$, where $\mathbf{A}_i \in \mathbb{R}^{m\times 2}$, $\|\mathbf{A}_i(:,1)\| = 1$, $\|\mathbf{A}_i(:,2)\| = \varepsilon$, for all $i = 1, \ldots, 4$.

Goal: choose three blocks (six rows) such that we maximize $\mathrm{vol}([\mathbf{A}_{\pi(1)}\ \mathbf{A}_{\pi(2)}\ \mathbf{A}_{\pi(3)}])$

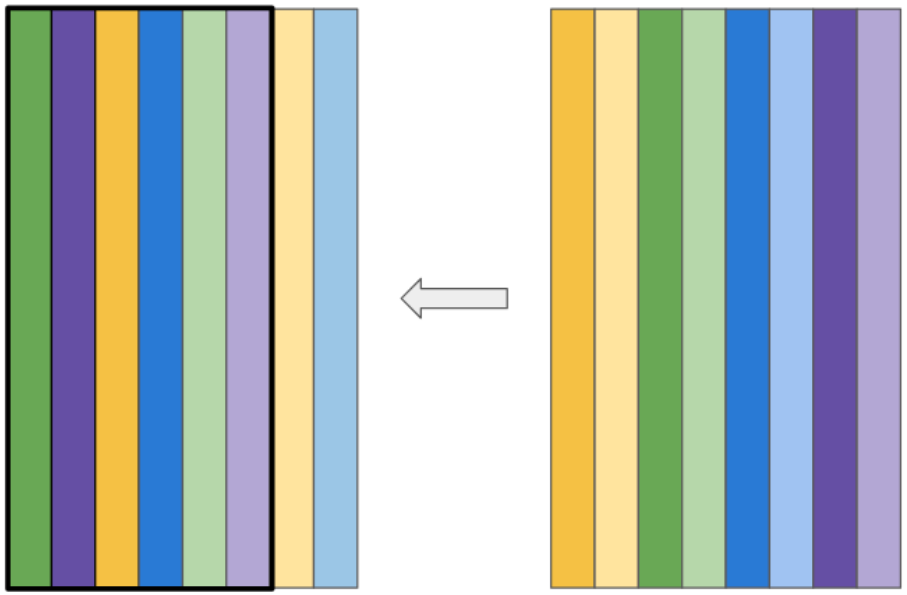


max vol Block matrix

Existing Block QRCP is inadequate: a toy example

Let $\mathbf{A} = [\mathbf{A}_1\ \mathbf{A}_2\ \mathbf{A}_3\ \mathbf{A}_4]$, where $\mathbf{A}_i \in \mathbb{R}^{m\times 2}$, $\|\mathbf{A}_i(:,1)\| = 1$, $\|\mathbf{A}_i(:,2)\| = \varepsilon$, for all $i = 1, \ldots, 4$.

Goal: choose three blocks (six rows) such that we maximize $\mathrm{vol}([\mathbf{A}_{\pi(1)}\ \mathbf{A}_{\pi(2)}\ \mathbf{A}_{\pi(3)}])$

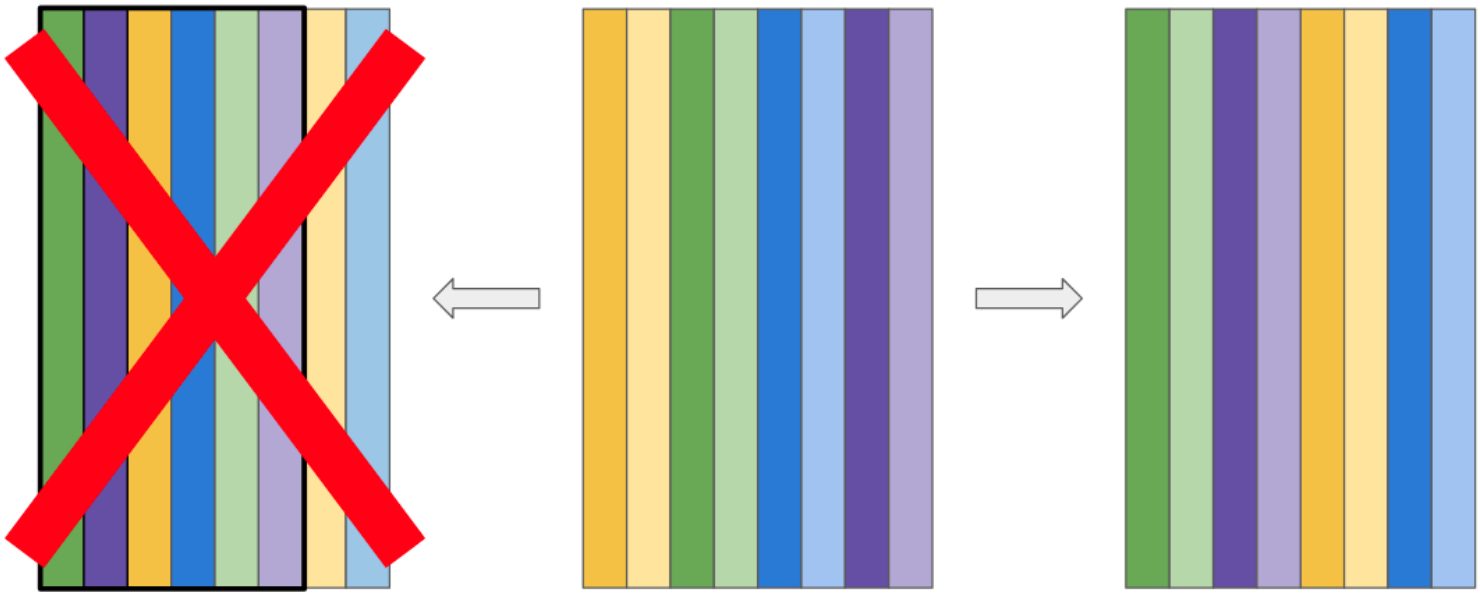


max vol Block matrix

Existing Block QRCP is inadequate: a toy example

Let $\mathbf{A} = [\mathbf{A}_1\ \mathbf{A}_2\ \mathbf{A}_3\ \mathbf{A}_4]$, where $\mathbf{A}_i \in \mathbb{R}^{m\times 2}$, $\|\mathbf{A}_i(:,1)\| = 1$, $\|\mathbf{A}_i(:,2)\| = \varepsilon$, for all $i = 1, \ldots, 4$.

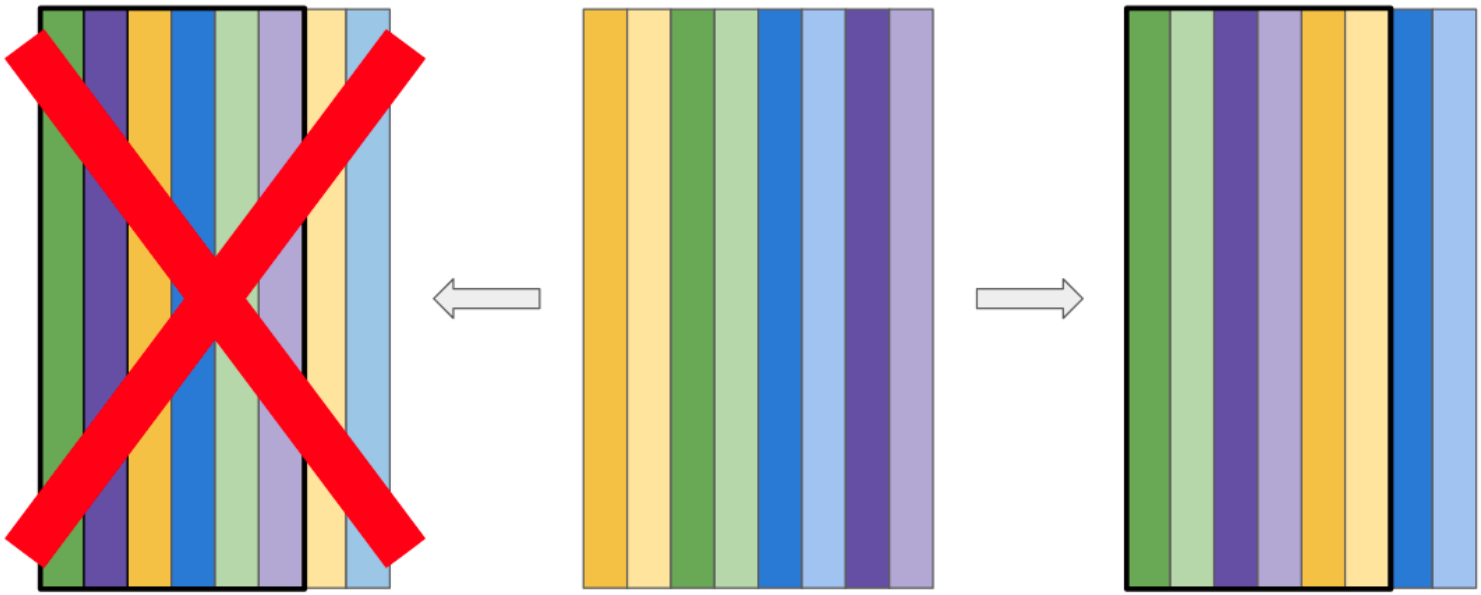
Goal: choose three blocks (six rows) such that we maximize $\mathrm{vol}([\mathbf{A}_{\pi(1)}\ \mathbf{A}_{\pi(2)}\ \mathbf{A}_{\pi(3)}])$



Existing Block QRCP is inadequate: a toy example

Let $\mathbf{A} = [\mathbf{A}_1\ \mathbf{A}_2\ \mathbf{A}_3\ \mathbf{A}_4]$, where $\mathbf{A}_i \in \mathbb{R}^{m\times 2}$, $\|\mathbf{A}_i(:,1)\| = 1$, $\|\mathbf{A}_i(:,2)\| = \varepsilon$, for all $i = 1, \ldots, 4$.

Goal: choose three blocks (six rows) such that we maximize $\mathrm{vol}([\mathbf{A}_{\pi(1)}\ \mathbf{A}_{\pi(2)}\ \mathbf{A}_{\pi(3)}])$



Greedy OED as block pivoted QR

Let $b \geq 2$ be the block size. Define the helper function `expand` such that

$$\texttt{expand([])} = \texttt{[]} \qquad \text{and} \qquad \texttt{expand}([\texttt{i}] + \mathcal{S}) = \texttt{i}b : (\texttt{i}+1)b - 1 \; + \texttt{expand}(\mathcal{S})$$

Greedy OED as block pivoted QR

Let $b \geq 2$ be the block size. Define the helper function `expand` such that

$$\texttt{expand([])} = \texttt{[]} \qquad \text{and} \qquad \texttt{expand}(\texttt{[i]} + \mathcal{S}) = \texttt{i}b : (\texttt{i}+1)b - 1 \; + \texttt{expand}(\mathcal{S})$$

Sylvester ⇒ D-optimality is volume maximization

With $\mathbf{A}_{\mathcal{S}} = \begin{bmatrix} \mathbf{J}_{\mathcal{C}}^{\top} \\ \mathbf{I}_{|\mathcal{C}|b} \end{bmatrix} \big(:, \texttt{expand}(\mathcal{S}) \big)$ and $[\sim, \mathbf{R}_{\mathcal{S}}] = \texttt{qr}(\mathbf{A}_{\mathcal{S}})$,

$$\phi\big(\mathbf{F}(c, \boldsymbol{\theta})\big) = \tfrac{1}{2} \log\det\left(\mathbf{J}_{\mathcal{S}}\mathbf{J}_{\mathcal{S}}^{\top} + \mathbf{I}_{|\mathcal{S}|b}\right) = \log\det \mathbf{R}_{\mathcal{S}} = \log \mathbb{V}\mathrm{ol}(\mathbf{A}_{\mathcal{S}})$$

EMORY UNIVERSITY

Greedy OED as block pivoted QR

Let $b \geq 2$ be the block size. Define the helper function `expand` such that

$$\texttt{expand([])} = \texttt{[]} \quad \text{and} \quad \texttt{expand}([\texttt{i}] + \mathcal{S}) = \texttt{i}b : (\texttt{i}+1)b - 1 \; + \texttt{expand}(\mathcal{S})$$

Sylvester ⇒ D-optimality is volume maximization

With $\mathbf{A}_\mathcal{S} = \begin{bmatrix} \mathbf{J}_\mathcal{C}^\top \\ \mathbf{I}_{|\mathcal{C}|b} \end{bmatrix} \left(:, \texttt{expand}(\mathcal{S}) \right)$ and $[\sim, \mathbf{R}_\mathcal{S}] = \texttt{qr}(\mathbf{A}_\mathcal{S})$,

$$\phi\big(\mathbf{F}(c, \boldsymbol{\theta})\big) = \tfrac{1}{2} \log\det\left(\mathbf{J}_\mathcal{S}\mathbf{J}_\mathcal{S}^\top + \mathbf{I}_{|\mathcal{S}|b}\right) = \log\det \mathbf{R}_\mathcal{S} = \log \mathbb{V}\mathrm{ol}(\mathbf{A}_\mathcal{S})$$

Idea: Define a suitable `block_pivoted_qr` function that we call as

$$[\sim, \mathbf{R}, \texttt{piv}] = \texttt{block\_pivoted\_qr}\left(b, \begin{bmatrix} \mathbf{J}_\mathcal{C}^\top \\ \mathbf{I}_{|\mathcal{C}|b} \end{bmatrix}\right),$$

and use $\mathcal{C}_{\text{ranked}} = \texttt{expand}^{-1}(\texttt{piv})$. Truncate $\mathcal{C}_{\text{ranked}}$ to get $\mathcal{S}$.

EMORY UNIVERSITY

The $\star_{\mathbf{M}}$ Product

Transform Domain
Let $\hat{\mathcal{T}} \in \mathbb{R}^{m\times n\times b}$ with i,j^{th} tube equal to the product of the i,j^{th} tube of tensor $\mathcal{T} \in \mathbb{R}^{m\times n\times b}$ with matrix $\mathbf{M} \in \mathbb{R}^{b\times b}$ by $\hat{\mathcal{T}} = \mathcal{T} \times_3 \mathbf{M}$,

$$(\mathcal{T} \times_3 \mathbf{M})(i,j,:) = \mathbf{M}\mathcal{T}(i,j,:).$$

Facewise-product
Here, the frontal facewise-product is defined in frontal slice k as

$$(\mathcal{T}_1 \Delta \mathcal{T}_2)(:,:,k) = \mathcal{T}_1(:,:,k)\mathcal{T}_2(:,:,k).$$

Definition ($\star_{\mathbf{M}}$ product[1])

The $\star_{\mathbf{M}}$ *product* for invertible $\mathbf{M}$ is defined as

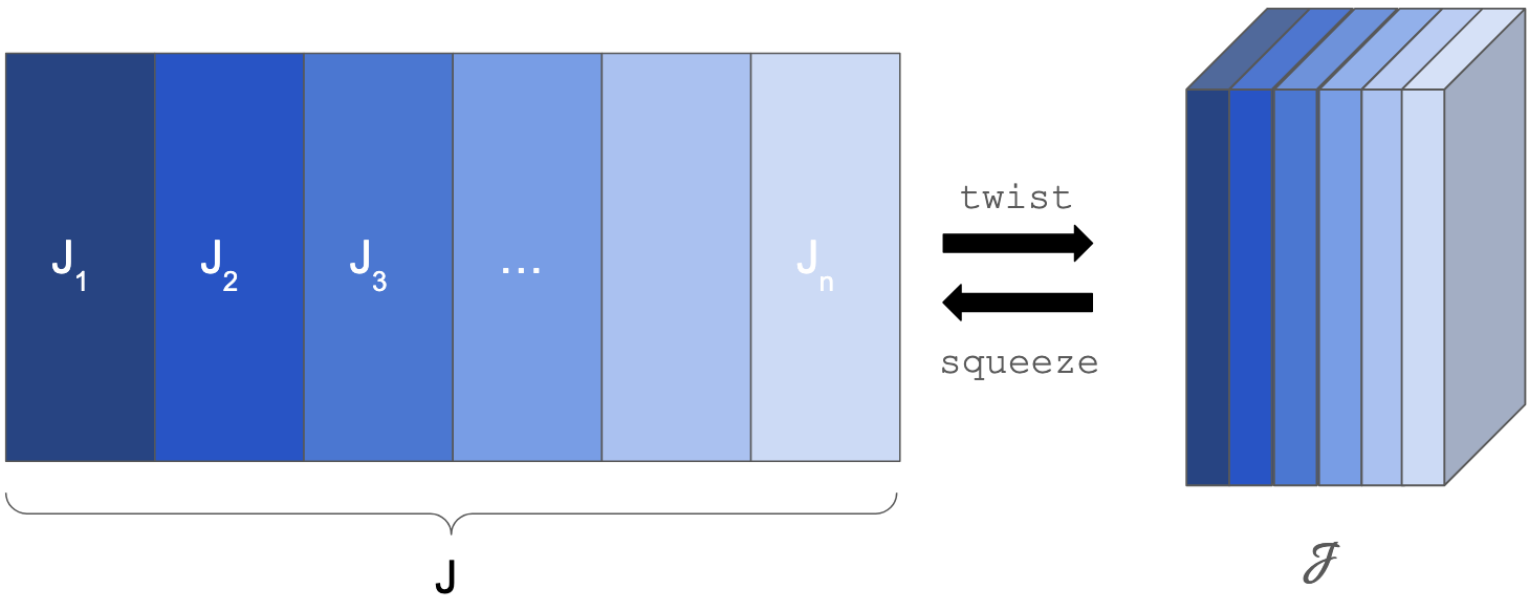
$$\mathcal{A} \star_{\mathbf{M}} \mathcal{B} := ((\mathcal{A} \times_3 \mathbf{M})\Delta(\mathcal{B} \times_3 \mathbf{M})) \times_3 \mathbf{M}^{-1}. \tag{1}$$

EMORY UNIVERSITY

[1] M. E. Kilmer and C. D. Martin. "Factorization strategies for third-order tensors". In: *Linear Algebra Appl.* 435.3 (2011).

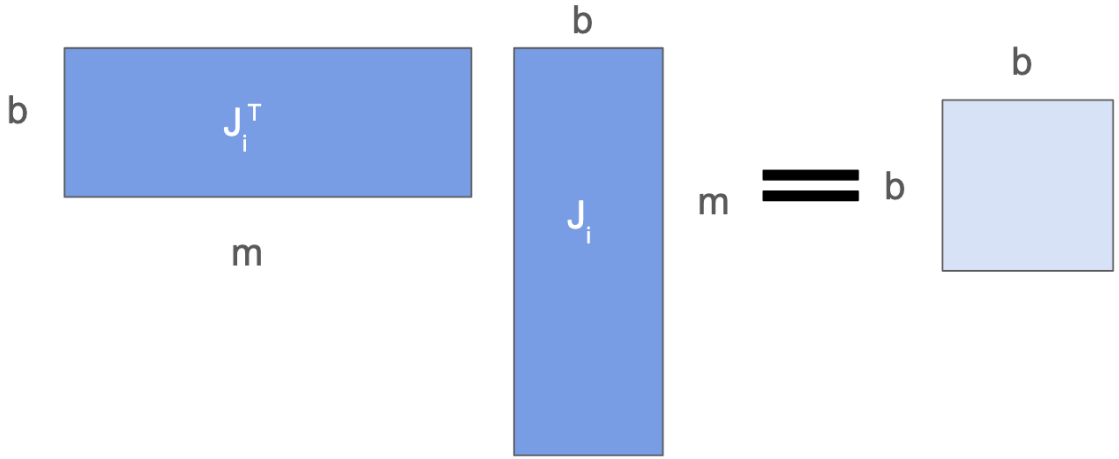
Start with your full Jacobian matrix **J**.

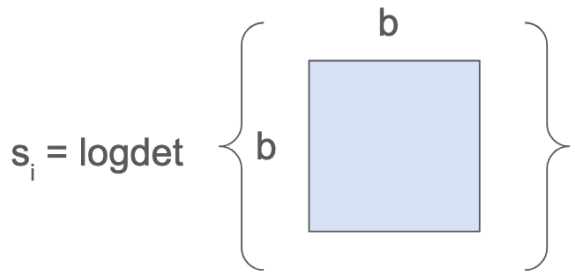
We can think of **J** as a block column matrix, or transform the blocks into lateral slices.



Determine the "most important" prior experiment

Initialize the permutation $\pi = (1, \cdots, n)$.
For each $1 \leq i \leq n$, compute $s_i = \log\det\left(\mathbf{J}_i^\top \mathbf{J}_i\right)$.



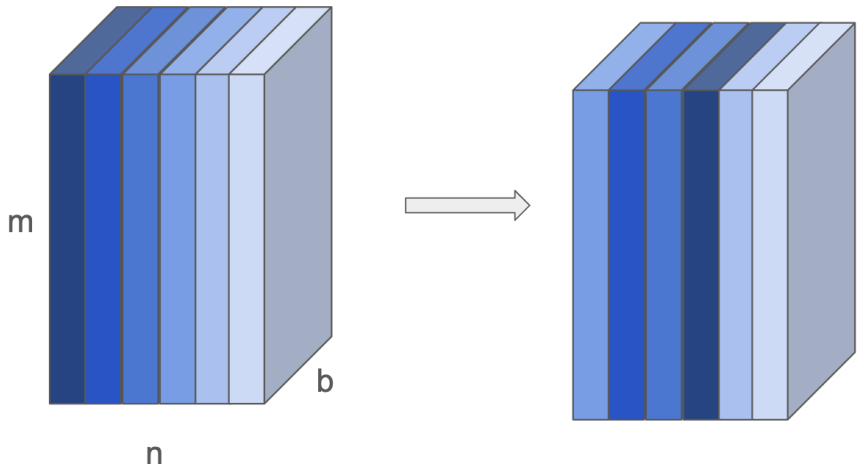


Then $k = \arg\max\{s_1, \cdots, s_n\}$.

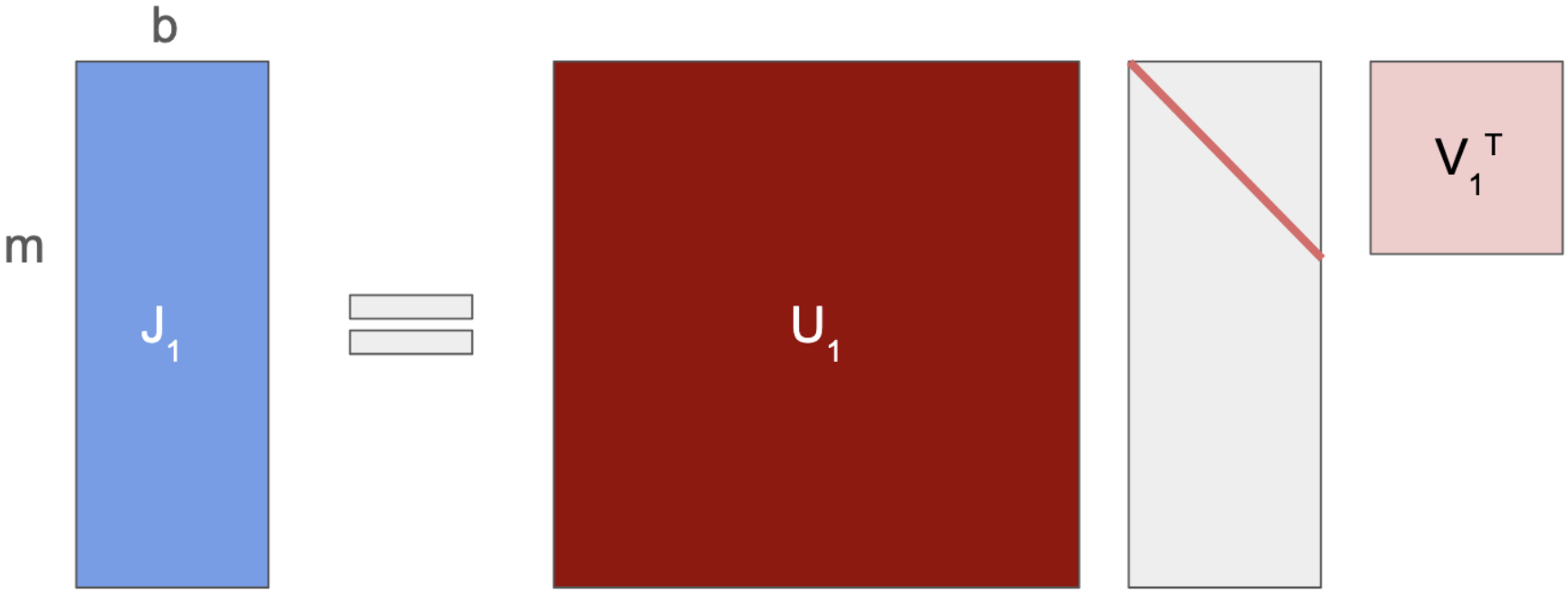
Permute the highest scored experiment to the front.

Since $k = \arg\max\{s_1, \cdots, s_n\}$, we let

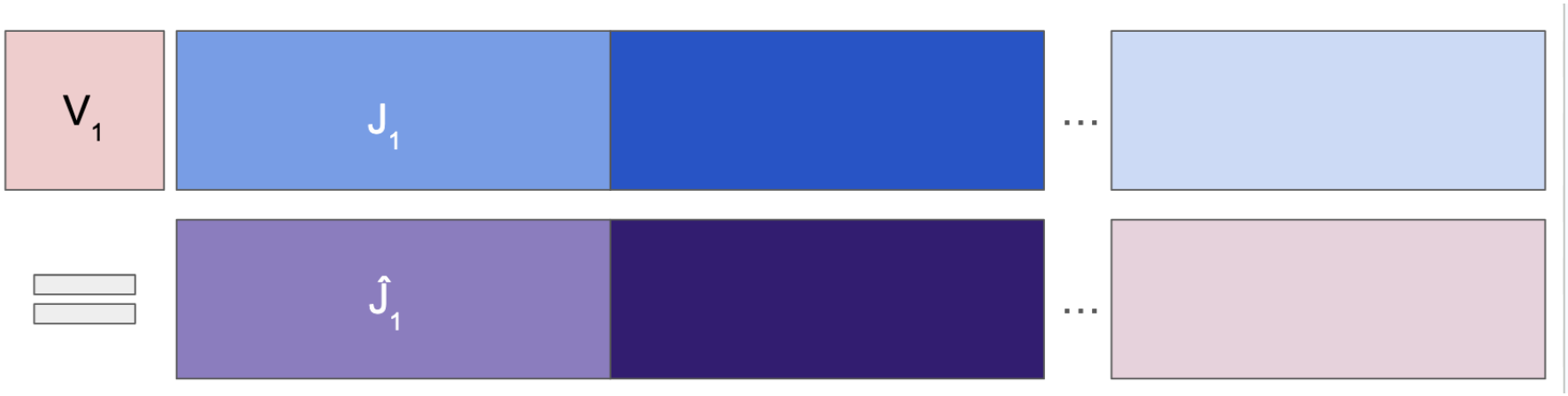
- $\mathcal{A}(:, 1, :) \leftrightarrow \mathcal{A}(:, k, :)$ and
- $\pi(1) \leftrightarrow \pi(k)$



Obtain the right singular vectors of the best experiment

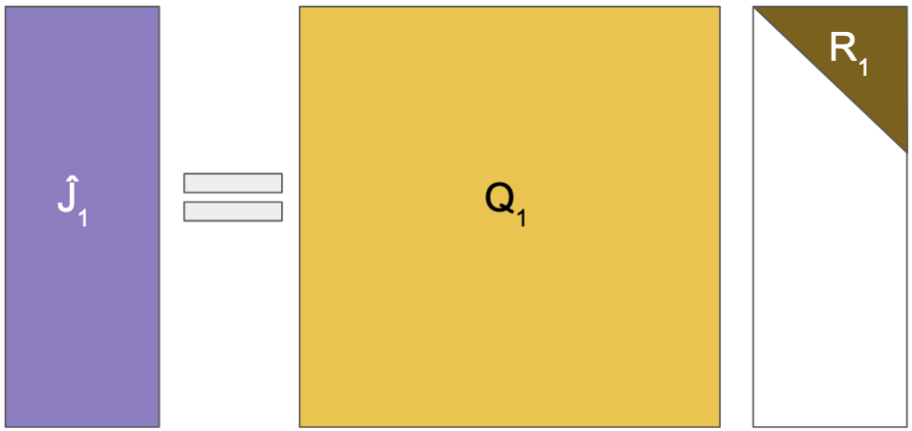


Update $\mathcal{J} \leftarrow \mathcal{J} \times_3 \mathbf{V}_1$

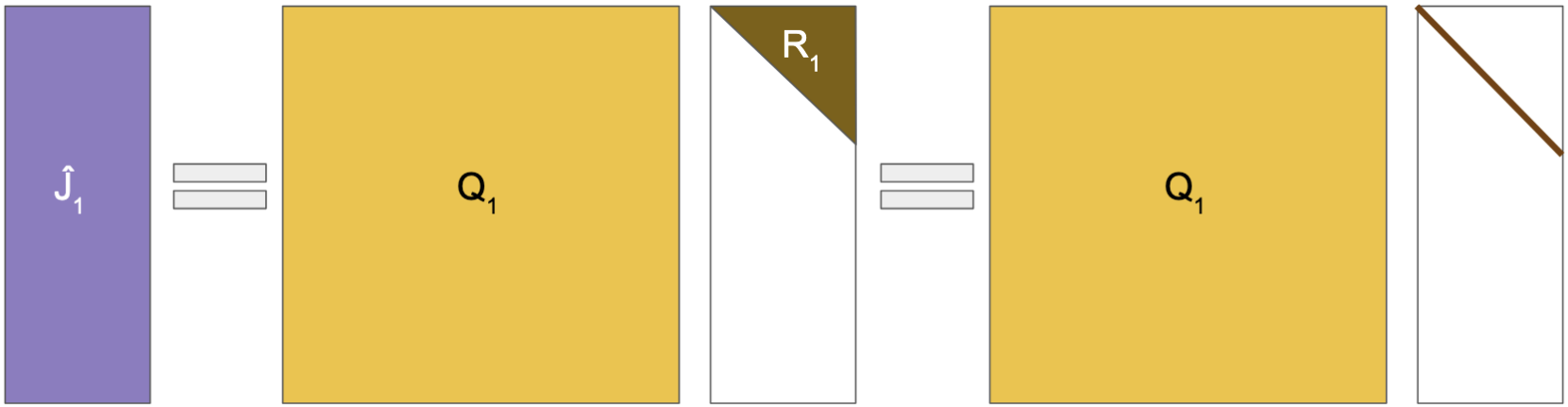


This is equivalent to lateral slice updates $\mathbf{J}_i \leftarrow \mathbf{J}_i \mathbf{V}_1^\top$

Compute the matrix QR on the transformed first lateral slice

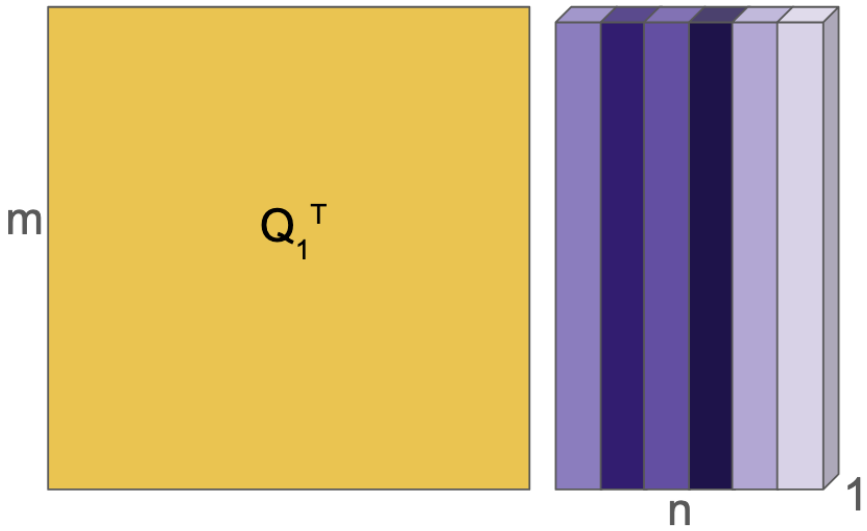


Compute the matrix QR on the transformed first lateral slice



Update the frontal slices

With access to $\mathbf{Q}_1^\top$, we perform $\mathcal{J}(:,:,\ell) \leftarrow \mathbf{Q}_1^\top \mathcal{J}(:,:,\ell)$ for all $\ell \in \{1,\ldots,b\}$



This is equivalent to $\mathbf{J}_i \leftarrow \mathbf{Q}_1^\top \mathbf{J}_i$

EMORY UNIVERSITY

The $\star_{\mathbf{M}}$ product Decomposition

Let $\mathcal{J}^{(1)}$ denote the value of $\mathcal{J}$ on *input* to the above procedure. By the end of the procedure we have a decomposition of the form

$$\mathcal{J}^{(1)} \star_{\mathbf{V}_1} \mathcal{P}^{(1)} = \mathcal{Q}^{(1)} \star_{\mathbf{V}_1} \mathcal{R}^{(1)}, \tag{2}$$

where $\mathcal{P}^{(1)}$, $\mathcal{Q}^{(1)}$, and $\mathcal{R}^{(1)}$ are defined implicitly; using π and $\mathbf{V}_1$ as they exist at step 3, and using $\mathbf{Q}_1$ from step 5, we have

$$[\mathcal{P}^{(1)} \times_3 \mathbf{V}_1](:,:,\ell) = \mathbf{I}_n(:,\pi) \tag{3}$$

$$[\mathcal{Q}^{(1)} \times_3 \mathbf{V}_1](:,:,\ell) = \mathbf{Q}_1 \tag{4}$$

$$[\mathcal{R}^{(1)} \times_3 \mathbf{V}_1](:,:,\ell) = \mathbf{Q}_1^\top [\mathcal{J}^{(1)} \times_3 \mathbf{V}_1](:,:,\ell)\mathbf{I}_n(:,\pi), \tag{5}$$

for each $\ell \in \{1,\ldots,b\}$.

We recurse on the $m \times (n-1) \times b$ tensor $\mathcal{J}^{(2)}$ getting a $\star_{\mathbf{V}_2}$ product.

How suboptimal is greedy? Exponentially

Theorem (Block submodularity, (Thm 4.1)[2])

Let $f(\mathcal{S}) = \log\det\left(\mathbf{X}_{\mathcal{S}}^{\top}\mathbf{X}_{\mathcal{S}}\right) = 2\log\mathbb{V}\mathrm{ol}(\mathbf{X}_{\mathcal{S}})$ *on the* n *blocks.*
For full column-rank $\mathbf{X}$, f *is submodular on* $2^{[n]}$; *if additionally* $\sigma_{\min}(\mathbf{X}) \geq 1$, *it is non-negative and monotone. This means*

$$0 \leq \log\mathbb{V}\mathrm{ol}(\mathbf{X}_{\mathcal{S}}^{*}) - \log\mathbb{V}\mathrm{ol}(\mathbf{X}_{\mathcal{S}}) \leq \left(1 - \frac{1}{k}\right)^{k} \log\mathbb{V}\mathrm{ol}(\mathbf{X}_{\mathcal{S}}^{*}) \lesssim 0.37\log\mathbb{V}\mathrm{ol}(\mathbf{X}_{\mathcal{S}}^{*})$$

EMORY UNIVERSITY

[2] Ilse CF Ipsen and Arvind K Saibaba. "Revisiting column subset selection through the lens of submodularity". In: *arXiv preprint arXiv:2607.13823* (2026).

How suboptimal is greedy? Exponentially

Theorem (Block submodularity, (Thm 4.1)[2])

Let $f(\mathcal{S}) = \log\det\left(\mathbf{X}_{\mathcal{S}}^{\top}\mathbf{X}_{\mathcal{S}}\right) = 2\log\mathbb{V}\mathrm{ol}(\mathbf{X}_{\mathcal{S}})$ *on the n blocks.*
For full column-rank $\mathbf{X}$, f *is submodular on* $2^{[n]}$; *if additionally* $\sigma_{\min}(\mathbf{X}) \geq 1$, *it is non-negative and monotone. This means*

$$0 \leq \log\mathbb{V}\mathrm{ol}(\mathbf{X}_{\mathcal{S}}^{*}) - \log\mathbb{V}\mathrm{ol}(\mathbf{X}_{\mathcal{S}}) \leq \left(1 - \frac{1}{k}\right)^{k} \log\mathbb{V}\mathrm{ol}(\mathbf{X}_{\mathcal{S}}^{*}) \lesssim 0.37\log\mathbb{V}\mathrm{ol}(\mathbf{X}_{\mathcal{S}}^{*})$$

Since $\mathcal{S}$ is an arbitrary subset selection, our experiment constrained subset works!

EMORY UNIVERSITY

[2] Ilse CF Ipsen and Arvind K Saibaba. "Revisiting column subset selection through the lens of submodularity". In: *arXiv preprint arXiv:2607.13823* (2026).

How suboptimal is greedy? Exponentially

Theorem (Block submodularity, (Thm 4.1)[2])

Let $f(\mathcal{S}) = \log\det\left(\mathbf{X}_{\mathcal{S}}^{\top}\mathbf{X}_{\mathcal{S}}\right) = 2\log\mathbb{V}\mathrm{ol}(\mathbf{X}_{\mathcal{S}})$ *on the n blocks.*
For full column-rank $\mathbf{X}$, f *is submodular on* $2^{[n]}$; *if additionally* $\sigma_{\min}(\mathbf{X}) \geq 1$, *it is non-negative and monotone. This means*

$$0 \leq \log\mathbb{V}\mathrm{ol}(\mathbf{X}_{\mathcal{S}}^{*}) - \log\mathbb{V}\mathrm{ol}(\mathbf{X}_{\mathcal{S}}) \leq \left(1 - \frac{1}{k}\right)^{k} \log\mathbb{V}\mathrm{ol}(\mathbf{X}_{\mathcal{S}}^{*}) \lesssim 0.37\log\mathbb{V}\mathrm{ol}(\mathbf{X}_{\mathcal{S}}^{*})$$

Since $\mathcal{S}$ is an arbitrary subset selection, our experiment constrained subset works!

Additionally, since $\mathbf{A} = \begin{bmatrix} \mathbf{J}_{\mathcal{C}}^{\top} \\ \mathbf{I}_{|\mathcal{C}|b} \end{bmatrix}$ is augmented by $\mathbf{I}$, then $\sigma_{\min} \geq 1$.

EMORY UNIVERSITY

[2] Ilse CF Ipsen and Arvind K Saibaba. "Revisiting column subset selection through the lens of submodularity". In: *arXiv preprint arXiv:2607.13823* (2026).

How suboptimal is greedy? Exponentially

Theorem (Block submodularity, (Thm 4.1)[2])

Let $f(\mathcal{S}) = \log\det\left(\mathbf{X}_{\mathcal{S}}^{\top}\mathbf{X}_{\mathcal{S}}\right) = 2\log \mathbb{V}\mathrm{ol}(\mathbf{X}_{\mathcal{S}})$ *on the n blocks.*
For full column-rank $\mathbf{X}$, f *is submodular on* $2^{[n]}$; *if additionally* $\sigma_{\min}(\mathbf{X}) \geq 1$, *it is non-negative and monotone. This means*

$$0 \leq \log \mathbb{V}\mathrm{ol}(\mathbf{X}_{\mathcal{S}}^{*}) - \log \mathbb{V}\mathrm{ol}(\mathbf{X}_{\mathcal{S}}) \leq \left(1 - \frac{1}{k}\right)^{k} \log \mathbb{V}\mathrm{ol}(\mathbf{X}_{\mathcal{S}}^{*}) \lesssim 0.37 \log \mathbb{V}\mathrm{ol}(\mathbf{X}_{\mathcal{S}}^{*})$$

Since $\mathcal{S}$ is an arbitrary subset selection, our experiment constrained subset works!

Additionally, since $\mathbf{A} = \begin{bmatrix} \mathbf{J}_{\mathcal{C}}^{\top} \\ \mathbf{I}_{|\mathcal{C}|b} \end{bmatrix}$ is augmented by $\mathbf{I}$, then $\sigma_{\min} \geq 1$.

Let $\mathcal{G}$ be greedy's k blocks and $\mathcal{S}^{\star}$ the optimum. Our framework has the bound:

$$\mathbb{V}\mathrm{ol}(\mathbf{A}_{\mathcal{S}^{*}})^{0.63} \lesssim \mathbb{V}\mathrm{ol}(\mathbf{A}_{\mathcal{S}^{*}})^{\left(1-\left(1-\frac{1}{k}\right)^{k}\right)} \leq \mathbb{V}\mathrm{ol}(\mathbf{A}_{\mathcal{G}}).$$

EMORY UNIVERSITY

[2] Ilse CF Ipsen and Arvind K Saibaba. "Revisiting column subset selection through the lens of submodularity". In: *arXiv preprint arXiv:2607.13823* (2026).

How suboptimal is greedy? Multiplicatively

Theorem (Multiplicative Suboptimal bound, (Thm 11)[3])

For $b = 1$,

$$\frac{1}{k!}\,\mathbb{V}\mathrm{ol}(\mathbf{X}_{\mathcal{S}}^{*}) \leq \mathbb{V}\mathrm{ol}(\mathbf{X}_{\mathcal{G}})$$

[3] Ali Çivril and Malik Magdon-Ismail. "On selecting a maximum volume sub-matrix of a matrix and related problems". en. In: *Theoretical Computer Science* 410.47-49 (2009). (Visited on 08/11/2025)

How suboptimal is greedy? Multiplicatively

Theorem (Multiplicative Suboptimal bound, (Thm 11)[3])

For $b = 1$,

$$\frac{1}{k!}\,\mathbb{V}\mathrm{ol}(\mathbf{X}_{\mathcal{S}}^{*}) \leq \mathbb{V}\mathrm{ol}(\mathbf{X}_{\mathcal{G}})$$

For $b \geq 2$, is there an $f(k,b) > 0$ such that

$$f(k,b)\,\mathbb{V}\mathrm{ol}(\mathbf{X}_{\mathcal{S}}^{*}) \leq \mathbb{V}\mathrm{ol}(\mathbf{X}_{\mathcal{G}})?$$

[3] Ali Çivril and Malik Magdon-Ismail. "On selecting a maximum volume sub-matrix of a matrix and related problems". en. In: *Theoretical Computer Science* 410.47-49 (2009). (Visited on 08/11/2025)

How suboptimal is greedy? Multiplicatively

Theorem (Multiplicative Suboptimal bound, (Thm 11)[3])

For $b = 1$,

$$\frac{1}{k!}\,\mathbb{V}\mathrm{ol}(\mathbf{X}_{\mathcal{S}}^{*}) \leq \mathbb{V}\mathrm{ol}(\mathbf{X}_{\mathcal{G}})$$

For $b \geq 2$, is there an $f(k,b) > 0$ such that

$$f(k,b)\,\mathbb{V}\mathrm{ol}(\mathbf{X}_{\mathcal{S}}^{*}) \leq \mathbb{V}\mathrm{ol}(\mathbf{X}_{\mathcal{G}})?$$

No! no $f(k,b) > 0$ can exist.

[3] Ali Çivril and Malik Magdon-Ismail. "On selecting a maximum volume sub-matrix of a matrix and related problems". en. In: *Theoretical Computer Science* 410.47-49 (2009). (Visited on 08/11/2025)

Why no $f(k, b) > 0$ exists: one 6×6 matrix

Fix $\varepsilon \in (0, 1)$, let $\delta > 0$, and take block size $b = 2$, rank $k = 2$, $n = 3$:

$$X = \left(\begin{array}{cc|cc|cc} 1 & 0 & \sqrt{1-\varepsilon^2} & 0 & 0 & 0 \\ 0 & 0 & \varepsilon & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & \sqrt{1-\varepsilon^2} & 0 \\ 0 & 0 & 0 & 0 & \varepsilon & 1 \\ \delta & 0 & 0 & 0 & 0 & 0 \\ 0 & \delta & 0 & 0 & 0 & 0 \end{array}\right)$$

$$\mathbb{V}\mathrm{ol}(X_{B_1}) = 1 + \delta^2 \;>\; \mathbb{V}\mathrm{ol}(X_{B_2}) = \mathbb{V}\mathrm{ol}(X_{B_3}) = \sqrt{1-\varepsilon^2}$$

Why no $f(k,b) > 0$ exists: one 6×6 matrix

Fix $\varepsilon \in (0,1)$, let $\delta > 0$, and take block size $b = 2$, rank $k = 2$, $n = 3$:

$$X = \left(\begin{array}{cc|cc|cc} 1 & 0 & \sqrt{1-\varepsilon^2} & 0 & 0 & 0 \\ 0 & 0 & \varepsilon & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & \sqrt{1-\varepsilon^2} & 0 \\ 0 & 0 & 0 & 0 & \varepsilon & 1 \\ \delta & 0 & 0 & 0 & 0 & 0 \\ 0 & \delta & 0 & 0 & 0 & 0 \end{array}\right)$$

$$\mathbb{V}\mathrm{ol}(X_{B_1}) = 1 + \delta^2 \;>\; \mathbb{V}\mathrm{ol}(X_{B_2}) = \mathbb{V}\mathrm{ol}(X_{B_3}) = \sqrt{1-\varepsilon^2}$$

Step 1: B_1 takes the first pivot and then poisons the first column of *both* B_2 and B_3:

$$\mathbb{V}\mathrm{ol}(X_{B_\mathcal{G}}) = (1+\delta^2)\frac{\delta\sqrt{1-\varepsilon^2}}{\sqrt{1+\delta^2}} = \delta\sqrt{1-\varepsilon^2}\sqrt{1+\delta^2}, \qquad \mathbb{V}\mathrm{ol}(X_{B_{S^\star}}) = 1-\varepsilon^2, \quad S^\star = \{2,3\}$$

$$\Longrightarrow \quad \frac{\mathbb{V}\mathrm{ol}(X_{B_\mathcal{G}})}{\mathbb{V}\mathrm{ol}(X_{B_{S^\star}})} = \frac{\delta\sqrt{1+\delta^2}}{\sqrt{1-\varepsilon^2}} \xrightarrow[\delta\to 0]{} 0.$$

Why no $f(k, b) > 0$ exists: one 6×6 matrix

Fix $\varepsilon \in (0, 1)$, let $\delta > 0$, and take block size $b = 2$, rank $k = 2$, $n = 3$:

$$X = \left(\begin{array}{cc|cc|cc} 1 & 0 & \sqrt{1-\varepsilon^2} & 0 & 0 & 0 \\ 0 & 0 & \varepsilon & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & \sqrt{1-\varepsilon^2} & 0 \\ 0 & 0 & 0 & 0 & \varepsilon & 1 \\ \delta & 0 & 0 & 0 & 0 & 0 \\ 0 & \delta & 0 & 0 & 0 & 0 \end{array}\right)$$

$$\mathbb{V}\mathrm{ol}(X_{B_1}) = 1 + \delta^2 \;>\; \mathbb{V}\mathrm{ol}(X_{B_2}) = \mathbb{V}\mathrm{ol}(X_{B_3}) = \sqrt{1-\varepsilon^2}$$

Step 1: B_1 takes the first pivot and then poisons the first column of *both* B_2 and B_3:

$$\mathbb{V}\mathrm{ol}(X_{B_G}) = (1+\delta^2)\frac{\delta\sqrt{1-\varepsilon^2}}{\sqrt{1+\delta^2}} = \delta\sqrt{1-\varepsilon^2}\sqrt{1+\delta^2}, \qquad \mathbb{V}\mathrm{ol}(X_{B_{S^\star}}) = 1-\varepsilon^2, \quad S^\star = \{2, 3\}$$

$$\Longrightarrow \quad \frac{\mathbb{V}\mathrm{ol}(X_{B_G})}{\mathbb{V}\mathrm{ol}(X_{B_{S^\star}})} = \frac{\delta\sqrt{1+\delta^2}}{\sqrt{1-\varepsilon^2}} \xrightarrow[\delta \to 0]{} 0.$$

On Gaussian blocks, greedy is exactly optimal in 60–90% of draws, and never below 0.72 in 1000 trials — the collapse is worst-case, not typical.

EMORY UNIVERSITY

Randomization can help!

The matrix $\mathbf{A} = \begin{bmatrix} \mathbf{J}_{\mathcal{C}}^{\top} \\ \mathbf{I}_{|\mathcal{C}|b} \end{bmatrix}$ is large!

Suppose we know $\mathbf{J}_{\mathcal{C}}$ has m columns, and we want to narrow down to k experiments.

For a tuning parameter $\gamma \in [1, 2]$, set $d = \lceil \gamma bk \rceil$.

Remark (Tuning parameter γ)

If $\gamma = 1$, then $d = bk$, which is the exact number of columns for k experiments. Just like Halko, Martinsson, and Tropp[4] recommend an oversampling parameter to be $bk + 5$ we use γ as a multiplicative oversampling parameter.

[4] Nathan Halko, Per-Gunnar Martinsson, and Joel A Tropp. "Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions". In: *SIAM review* 53.2 (2011)

Randomized Sketching is important.

Our sketching matrix becomes

$$\boldsymbol{\Omega} = \begin{bmatrix} \boldsymbol{\Omega}_1 & \boldsymbol{\Omega}_2 \end{bmatrix} \tag{6}$$

where $\boldsymbol{\Omega}_1 \in \mathbb{R}^{d \times m}$ and $\boldsymbol{\Omega}_2 \in \mathbb{R}^{d \times |\mathcal{C}|b}$ are i.i.d. mean zero unit variance Gaussian entries.

Remark (Why Gaussian Sketching?)

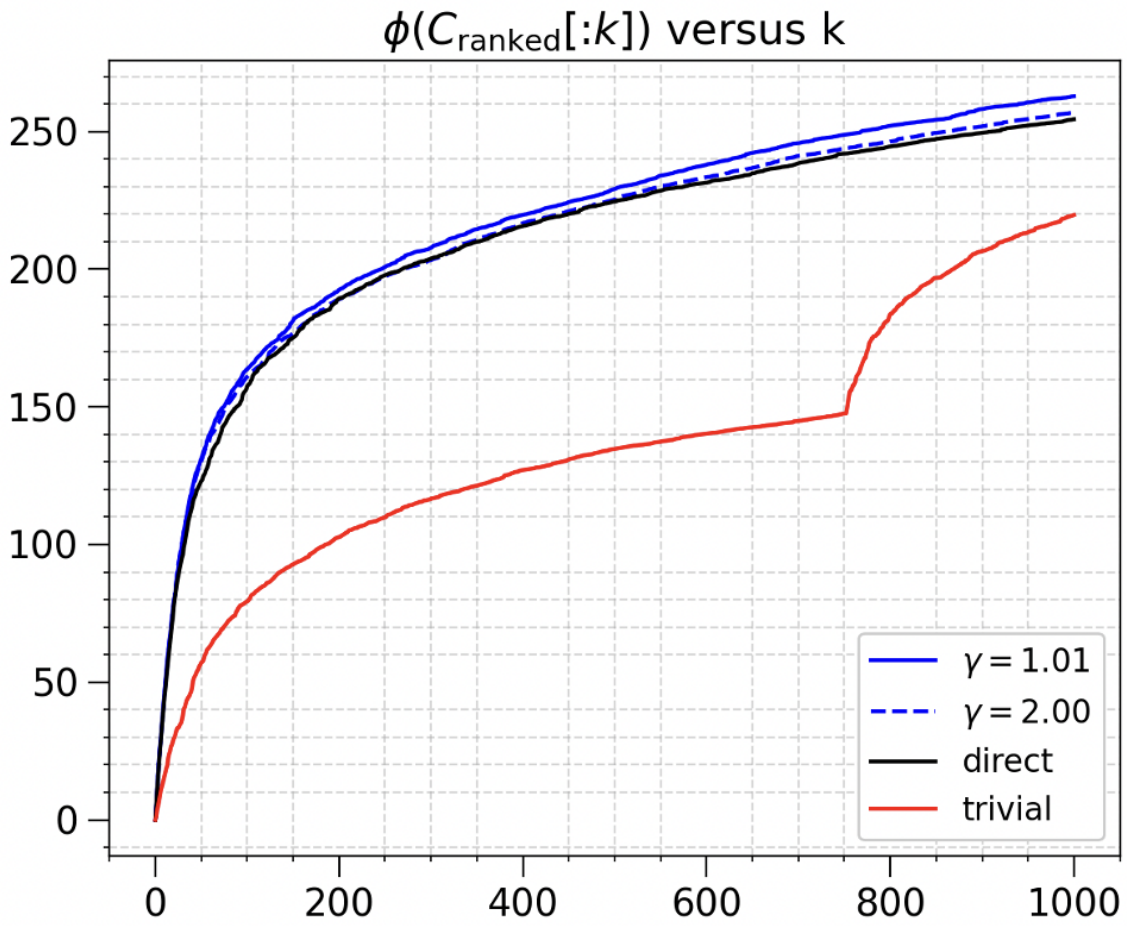
Because of noise in the quantum experiment matrix, the matrix is dense. Additionally, sketching is not the most expensive part of this procedure, so Gaussian works well.

Instead of running the algorithm on $\mathbf{A}$, instead we can make pivot decisions on $\boldsymbol{\Omega}\mathbf{A}$

$$\boldsymbol{\Omega}\mathbf{A} = \boldsymbol{\Omega}_1 \mathbf{J}_{\mathcal{C}}^{\top} + \boldsymbol{\Omega}_2$$

EMORY UNIVERSITY

Quantum Gate Set Tomography: 7× faster with no loss in objective



Use randLAPACK's randomized block QRCP algorithm `BQRRP`[5] to make 1000 block pivot decisions.

Here **S** represents the deterministically updated sketch.

Total time for block pivoted QR on sketch: **36 sec**

Total time for each input matrix:

- **SA** (direct): 48 minutes
- $\mathbf{S\Omega}_{2.00}\mathbf{A}$: 18 minutes
- $\mathbf{S\Omega}_{1.01}\mathbf{A}$: 7 minutes

EMORY UNIVERSITY

[5] Maksim Melnichenko et al. *Anatomy of High-Performance Column-Pivoted QR Decomposition.* en. 2025. (Visited on 08/11/2025)

Three things to take away

1. **Blocking is a constraint, not a convenience.** An experiment cannot be split, so the max-volume *column* subset is typically infeasible.
2. **Block-pivoted QR is a $\star_{\mathbf{V}_1}$ factorization with a *learned* transform.** $\mathbf{V}_1$ comes from the leading pivot's SVD, not a fixed DFT/DCT.
3. **The guarantee landscape splits at $b = 2$,** uniformly in k and b:

exponential $\mathbb{V}\mathrm{ol}(X_{B_{S^\star}})^{1-(1-1/k)^k}$	survives all b	✓
any $f(k,b) > 0$	$b \geq 2$	×

Next Steps

Open: what controls the total curvature κ for block Gram matrices?

Code and the counterexample family: `mitchell.scott@emory.edu`

EMORY UNIVERSITY

References

[1] M. E. Kilmer and C. D. Martin. "Factorization strategies for third-order tensors". In: *Linear Algebra Appl.* 435.3 (2011), pp. 641–658.

[2] Ilse CF Ipsen and Arvind K Saibaba. "Revisiting column subset selection through the lens of submodularity". In: *arXiv preprint arXiv:2607.13823* (2026).

[3] Ali Çivril and Malik Magdon-Ismail. "On selecting a maximum volume sub-matrix of a matrix and related problems". en. In: *Theoretical Computer Science* 410.47-49 (Nov. 2009), pp. 4801–4811. ISSN: 03043975. DOI: 10.1016/j.tcs.2009.06.018. URL: https://linkinghub.elsevier.com/retrieve/pii/S0304397509004101 (visited on 08/11/2025).

[4] Nathan Halko, Per-Gunnar Martinsson, and Joel A Tropp. "Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions". In: *SIAM review* 53.2 (2011), pp. 217–288.

References (cont.)

[5] Maksim Melnichenko et al. *Anatomy of High-Performance Column-Pivoted QR Decomposition*. en. arXiv:2507.00976 [cs]. July 2025. DOI: 10.48550/arXiv.2507.00976. URL: http://arxiv.org/abs/2507.00976 (visited on 08/11/2025).

[6] Anil Damle et al. *How to reveal the rank of a matrix?* arXiv:2405.04330 [math]. June 2024. DOI: 10.48550/arXiv.2405.04330. URL: http://arxiv.org/abs/2405.04330 (visited on 10/18/2025).

[7] Peter Businger and Gene H. Golub. "Linear least squares solutions by householder transformations". en. In: *Numerische Mathematik* 7.3 (June 1965), pp. 269–276. ISSN: 0945-3245. DOI: 10.1007/BF01436084. URL: https://doi.org/10.1007/BF01436084 (visited on 10/04/2025).

Connection between Volume, Singular Values, and Rank Revealers

Definition (Rank-revealer [6])

If there is a $\mu_{m,n,k} \in \text{poly}(m, n, k)$ such that for any $\mathbf{A} \in \mathbb{R}^{m \times n}$ and any $1 \leq k \leq \min(m, n)$, the *rank revealer* computes a rank $\leq k$ matrix $\mathbf{A}_k$ such that

$$\frac{1}{\mu_{m,n,k}} \sigma_j(\mathbf{A}) \leq \sigma_j(\mathbf{A}_k) \leq \mu_{m,n,k} \sigma_j(\mathbf{A}), \quad 1 \leq j \leq k \tag{7}$$

EMORY UNIVERSITY

[6] Anil Damle et al. *How to reveal the rank of a matrix?* 2024. (Visited on 10/18/2025)

Connection between Volume, Singular Values, and Rank Revealers

Definition (Rank-revealer [6])

If there is a $\mu_{m,n,k} \in \text{poly}(m, n, k)$ such that for any $\mathbf{A} \in \mathbb{R}^{m \times n}$ and any $1 \leq k \leq \min(m, n)$, the *rank revealer* computes a rank $\leq k$ matrix $\mathbf{A}_k$ such that

$$\frac{1}{\mu_{m,n,k}} \sigma_j(\mathbf{A}) \leq \sigma_j(\mathbf{A}_k) \leq \mu_{m,n,k} \sigma_j(\mathbf{A}), \quad 1 \leq j \leq k \tag{7}$$

and

$$\frac{1}{\mu_{m,n,k}} \sigma_{k+j}(\mathbf{A}) \leq \sigma_j(\mathbf{A} - \mathbf{A}_k) \leq \mu_{m,n,k} \sigma_{k+j}(\mathbf{A}), \quad 1 \leq j \leq \min(m, n) - k \tag{8}$$

EMORY UNIVERSITY

[6] Anil Damle et al. *How to reveal the rank of a matrix?* 2024. (Visited on 10/18/2025)

Connection between Volume, Singular Values, and Rank Revealers

Definition (Rank-revealer [6])

If there is a $\mu_{m,n,k} \in \text{poly}(m, n, k)$ such that for any $\mathbf{A} \in \mathbb{R}^{m \times n}$ and any $1 \leq k \leq \min(m, n)$, the *rank revealer* computes a rank $\leq k$ matrix $\mathbf{A}_k$ such that

$$\frac{1}{\mu_{m,n,k}} \sigma_j(\mathbf{A}) \leq \sigma_j(\mathbf{A}_k) \leq \mu_{m,n,k} \sigma_j(\mathbf{A}), \quad 1 \leq j \leq k \tag{7}$$

and

$$\frac{1}{\mu_{m,n,k}} \sigma_{k+j}(\mathbf{A}) \leq \sigma_j(\mathbf{A} - \mathbf{A}_k) \leq \mu_{m,n,k} \sigma_{k+j}(\mathbf{A}), \quad 1 \leq j \leq \min(m, n) - k \tag{8}$$

Truncated SVD:

- rank-revealer:
  $\mu = 1$ ☺

EMORY UNIVERSITY

[6] Anil Damle et al. *How to reveal the rank of a matrix?* 2024. (Visited on 10/18/2025)

Connection between Volume, Singular Values, and Rank Revealers

Definition (Rank-revealer [6])

If there is a $\mu_{m,n,k} \in \text{poly}(m, n, k)$ such that for any $\mathbf{A} \in \mathbb{R}^{m \times n}$ and any $1 \leq k \leq \min(m, n)$, the *rank revealer* computes a rank $\leq k$ matrix $\mathbf{A}_k$ such that

$$\frac{1}{\mu_{m,n,k}} \sigma_j(\mathbf{A}) \leq \sigma_j(\mathbf{A}_k) \leq \mu_{m,n,k} \sigma_j(\mathbf{A}), \quad 1 \leq j \leq k \tag{7}$$

and

$$\frac{1}{\mu_{m,n,k}} \sigma_{k+j}(\mathbf{A}) \leq \sigma_j(\mathbf{A} - \mathbf{A}_k) \leq \mu_{m,n,k} \sigma_{k+j}(\mathbf{A}), \quad 1 \leq j \leq \min(m, n) - k \tag{8}$$

Truncated SVD:

- rank-revealer: $\mu = 1$ ☺
- slow: $\mathcal{O}(mn \min(m, n))$ ☹

[6] Anil Damle et al. *How to reveal the rank of a matrix?* 2024. (Visited on 10/18/2025)

EMORY UNIVERSITY

Connection between Volume, Singular Values, and Rank Revealers

Definition (Rank-revealer [6])

If there is a $\mu_{m,n,k} \in \text{poly}(m, n, k)$ such that for any $\mathbf{A} \in \mathbb{R}^{m \times n}$ and any $1 \leq k \leq \min(m, n)$, the *rank revealer* computes a rank $\leq k$ matrix $\mathbf{A}_k$ such that

$$\frac{1}{\mu_{m,n,k}} \sigma_j(\mathbf{A}) \leq \sigma_j(\mathbf{A}_k) \leq \mu_{m,n,k} \sigma_j(\mathbf{A}), \quad 1 \leq j \leq k \tag{7}$$

and

$$\frac{1}{\mu_{m,n,k}} \sigma_{k+j}(\mathbf{A}) \leq \sigma_j(\mathbf{A} - \mathbf{A}_k) \leq \mu_{m,n,k} \sigma_{k+j}(\mathbf{A}), \quad 1 \leq j \leq \min(m, n) - k \tag{8}$$

Truncated SVD:

- rank-revealer: $\mu = 1$ ☺
- slow: $\mathcal{O}(mn \min(m, n))$ ☹

Column Pivoted QR [7]:

- practical: $\mathcal{O}(kmn)$ ☺

[6] Anil Damle et al. *How to reveal the rank of a matrix?* 2024. (Visited on 10/18/2025)

[7] Peter Businger and Gene H. Golub. "Linear least squares solutions by householder transformations". en. In: *Numerische Mathematik* 7.3 (1965). (Visited on 10/04/2025)

EMORY UNIVERSITY

Connection between Volume, Singular Values, and Rank Revealers

Definition (Rank-revealer [6])

If there is a $\mu_{m,n,k} \in \text{poly}(m, n, k)$ such that for any $\mathbf{A} \in \mathbb{R}^{m \times n}$ and any $1 \leq k \leq \min(m, n)$, the *rank revealer* computes a rank $\leq k$ matrix $\mathbf{A}_k$ such that

$$\frac{1}{\mu_{m,n,k}} \sigma_j(\mathbf{A}) \leq \sigma_j(\mathbf{A}_k) \leq \mu_{m,n,k} \sigma_j(\mathbf{A}), \quad 1 \leq j \leq k \tag{7}$$

and

$$\frac{1}{\mu_{m,n,k}} \sigma_{k+j}(\mathbf{A}) \leq \sigma_j(\mathbf{A} - \mathbf{A}_k) \leq \mu_{m,n,k} \sigma_{k+j}(\mathbf{A}), \quad 1 \leq j \leq \min(m, n) - k \tag{8}$$

Truncated SVD:

- rank-revealer: $\mu = 1$ ☺
- slow: $\mathcal{O}(mn \min(m, n))$ ☹

Column Pivoted QR [7]:

- practical: $\mathcal{O}(kmn)$ ☺
- not rank-revealer: $\mu = 2^k \sqrt{n - k}$ ☹

[6] Anil Damle et al. *How to reveal the rank of a matrix?* 2024. (Visited on 10/18/2025)

[7] Peter Businger and Gene H. Golub. "Linear least squares solutions by householder transformations". en. In: *Numerische Mathematik* 7.3 (1965). (Visited on 10/04/2025)

Connection between Volume, Singular Values, and Rank Revealers

Definition (Rank-revealer [6])

If there is a $\mu_{m,n,k} \in \text{poly}(m, n, k)$ such that for any $\mathbf{A} \in \mathbb{R}^{m \times n}$ and any $1 \leq k \leq \min(m, n)$, the *rank revealer* computes a rank $\leq k$ matrix $\mathbf{A}_k$ such that

$$\frac{1}{\mu_{m,n,k}} \sigma_j(\mathbf{A}) \leq \sigma_j(\mathbf{A}_k) \leq \mu_{m,n,k} \sigma_j(\mathbf{A}), \quad 1 \leq j \leq k \tag{7}$$

and

$$\frac{1}{\mu_{m,n,k}} \sigma_{k+j}(\mathbf{A}) \leq \sigma_j(\mathbf{A} - \mathbf{A}_k) \leq \mu_{m,n,k} \sigma_{k+j}(\mathbf{A}), \quad 1 \leq j \leq \min(m, n) - k \tag{8}$$

Truncated SVD:

- rank-revealer: $\mu = 1$ ☺
- slow: $\mathcal{O}(mn \min(m, n))$ ☹

Column Pivoted QR [7]:

- practical: $\mathcal{O}(kmn)$ ☺
- not rank-revealer: $\mu = 2^k \sqrt{n - k}$ ☹

ν-Near local max vol QR[1]:

- rank-revealer: $\mu = \sqrt{1 + 5\nu^2 kn}$ ☺

[6] Anil Damle et al. *How to reveal the rank of a matrix?* 2024. (Visited on 10/18/2025)

[7] Peter Businger and Gene H. Golub. "Linear least squares solutions by householder transformations". en. In: *Numerische Mathematik* 7.3 (1965). (Visited on 10/04/2025)

EMORY UNIVERSITY

Connection between Volume, Singular Values, and Rank Revealers

Definition (Rank-revealer [6])

If there is a $\mu_{m,n,k} \in \text{poly}(m, n, k)$ such that for any $\mathbf{A} \in \mathbb{R}^{m \times n}$ and any $1 \leq k \leq \min(m, n)$, the *rank revealer* computes a rank $\leq k$ matrix $\mathbf{A}_k$ such that

$$\frac{1}{\mu_{m,n,k}} \sigma_j(\mathbf{A}) \leq \sigma_j(\mathbf{A}_k) \leq \mu_{m,n,k} \sigma_j(\mathbf{A}), \quad 1 \leq j \leq k \tag{7}$$

and

$$\frac{1}{\mu_{m,n,k}} \sigma_{k+j}(\mathbf{A}) \leq \sigma_j(\mathbf{A} - \mathbf{A}_k) \leq \mu_{m,n,k} \sigma_{k+j}(\mathbf{A}), \quad 1 \leq j \leq \min(m, n) - k \tag{8}$$

Truncated SVD:

- rank-revealer: $\mu = 1$ ☺
- slow: $\mathcal{O}(mn \min(m, n))$ ☹

Column Pivoted QR [7]:

- practical: $\mathcal{O}(kmn)$ ☺
- not rank-revealer: $\mu = 2^k \sqrt{n - k}$ ☹

ν-Near local max vol QR[1]:

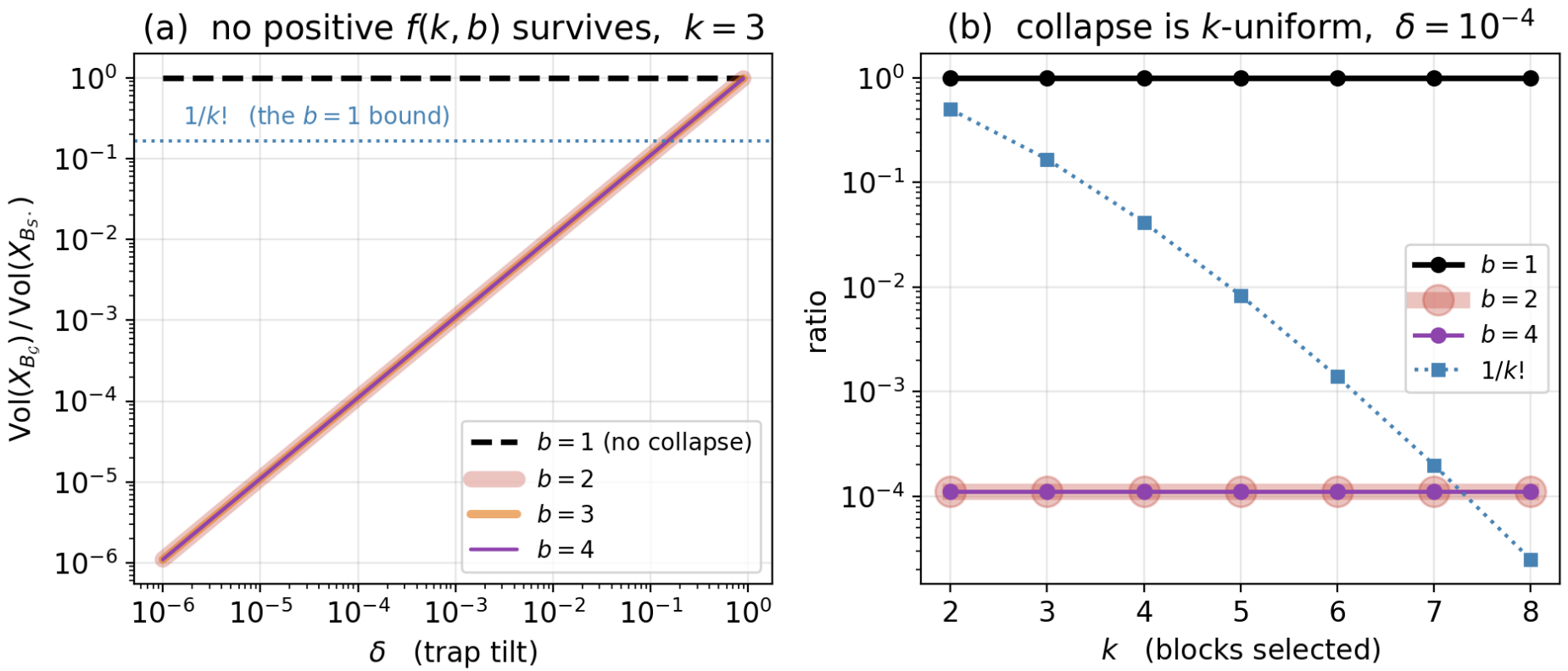
- rank-revealer: $\mu = \sqrt{1 + 5\nu^2 kn}$ ☺
- fast: $\mathcal{O}\left(\sqrt{kn}\right)$ ☺

[6] Anil Damle et al. *How to reveal the rank of a matrix?* 2024. (Visited on 10/18/2025)

[7] Peter Businger and Gene H. Golub. "Linear least squares solutions by householder transformations". en. In: *Numerische Mathematik* 7.3 (1965). (Visited on 10/04/2025)

EMORY UNIVERSITY

The collapse is uniform in k and in b

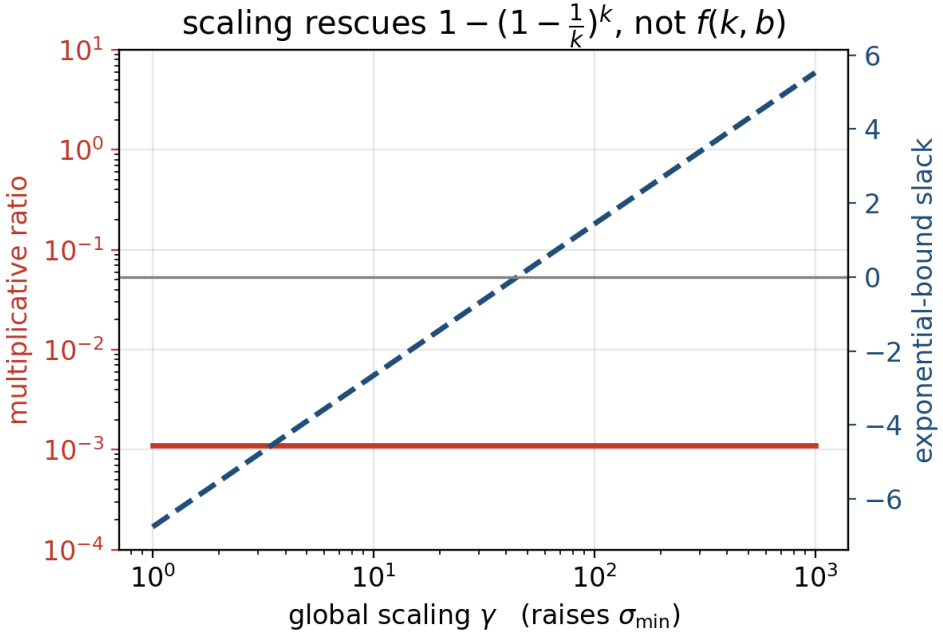


(a) ratio $= \Theta(\delta)$ for every $b \geq 2$, independent of b; the $b = 1$ curve never leaves 1.

(b) no dependence on k either — padding carries the 6×6 gadget to every $k \geq 2$, $b \geq 2$, so no $f(k, b)$ survives.

EMORY UNIVERSITY

Appendix: normalization cannot buy a multiplicative bound



Under $X \mapsto \gamma X$ both volumes pick up γ^{kb}, so

$$\frac{\mathbb{V}\mathrm{ol}(X_{B_{\mathcal{G}}})}{\mathbb{V}\mathrm{ol}(X_{B_{S^\star}})} \quad \text{is scale invariant.}$$

- Raising $\sigma_{\min}(X)$ to ≥ 1 *does* restore the exponential guarantee (slack crosses 0 near $\gamma \approx 45$).
- It leaves the multiplicative ratio untouched.
- The 6×6 has $\sigma_{\min} \approx 6.7 \times 10^{-3}$: the hypothesis of the corollary is doing real work.

EMORY UNIVERSITY

The collapse is worst case, not typical

				$\mathbb{V}\mathrm{ol}(X_{B_{\mathcal{G}}}) \,/\, \mathbb{V}\mathrm{ol}(X_{B_{S^\star}})$			
b	k	n	draws	$\mathbb{P}[=1]$	p_{05}	min	median | miss
i.i.d. Gaussian blocks							
2	3	8	300	87%	0.921	0.831	0.936
2	3	12	200	86%	0.938	0.858	0.946
2	3	20	120	78%	0.914	0.786	0.955
3	3	12	150	81%	0.873	0.739	0.915
4	3	12	150	78%	0.805	0.715	0.923
2	5	12	120	74%	0.830	0.732	0.894
3	5	12	100	59%	0.808	0.718	0.912
trap-and-twins							
any	any	$k{+}1$	—	0%	$\Theta(\delta)$	(1.1×10^{-4} at $\delta = 10^{-4}$)	

Greedy is *exactly* optimal on most random draws, and a miss costs only a few percent — never below 0.72 in 1,140 draws. The adversarial family drives the same quantity to 0.

So any useful guarantee must be instance-dependent: what parameter separates these two blocks of rows?

EMORY UNIVERSITY