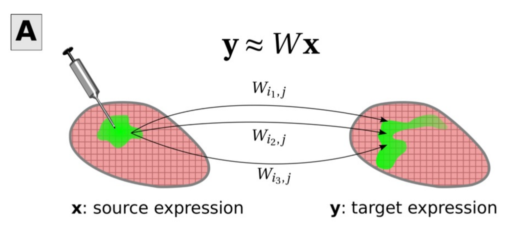
Sketched Column-based Matrix Approximation with Side Information

Jeongmin Chae[1], Praneeth Narayanamurthy[1], Selin Bac, Shaama Mallikarjun Sharada and Urbashi Mitra[1]
[1]Ming Hsieh Department of Electrical and Computer Engineering, USC, Los Angeles, CA
Department of Chemical Engineering, USC, Los Angeles, CA

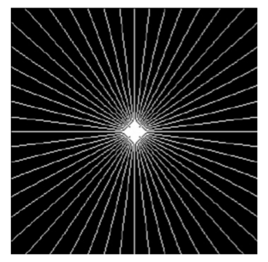
USC Viterbi School of Engineering

1 Applications

Random sampling is not always possible:

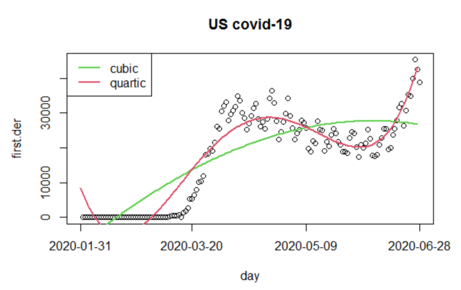


neural connectivity



undersampled radial MRI

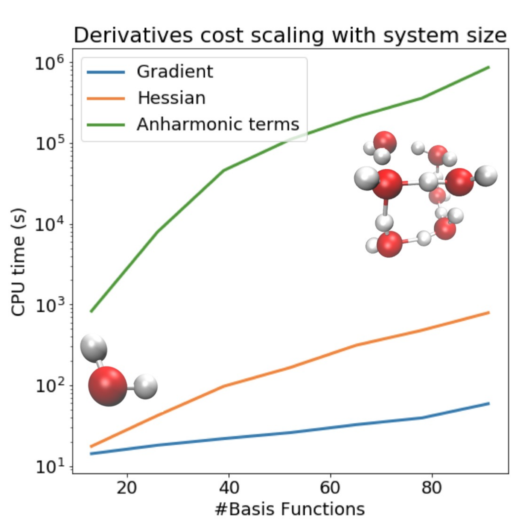
finance

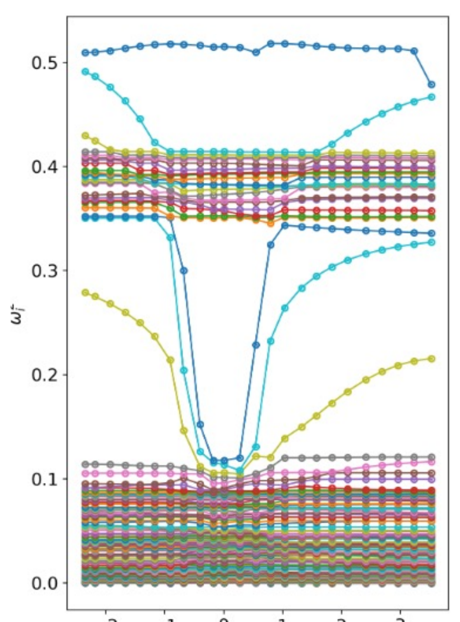


social science

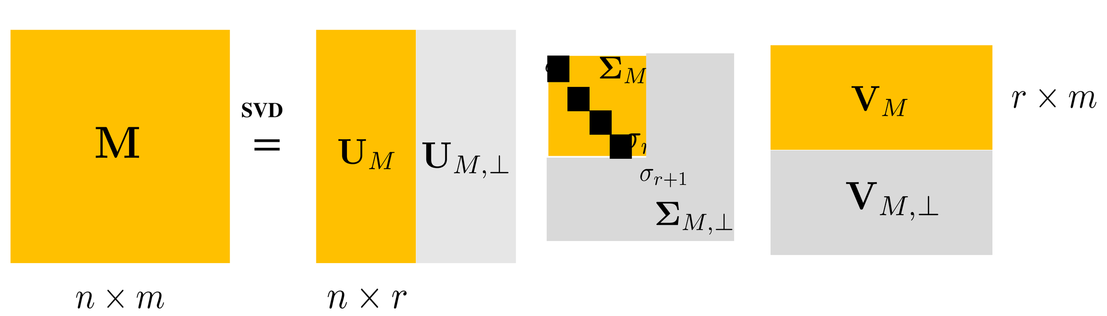
2 Motivation – Quantum Chemistry

Expensive eigenvalue computation along the chemical reaction time [1]





Our prior art QPMA
- (Expensive) SVD computaion [1]



Signal model

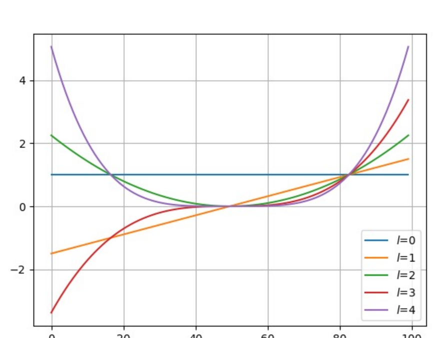
$$\mathbf{M} = \mathbf{QS} + \mathbf{E}$$

True matrix of rank k

unknown polynomial coefficient matrix

known polynomial information matrix

unknown pertubation matrix

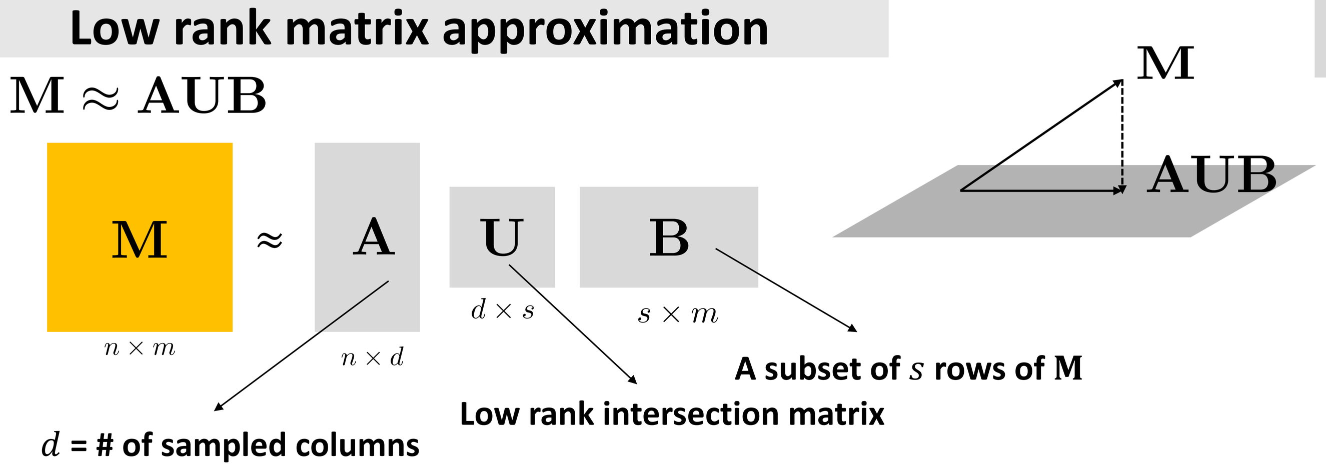


$$\mathbf{S} = \begin{bmatrix} 1 & 1 & \dots & 1 \\ s_1 & s_2 & \dots & s_m \\ s_1^2 & s_2^2 & \dots & s_m^2 \\ \vdots & \vdots & \vdots & \vdots \\ s_1^l & s_2^l & \dots & s_m^l \end{bmatrix}$$

3 Problem Solution – Column sampling and sketching

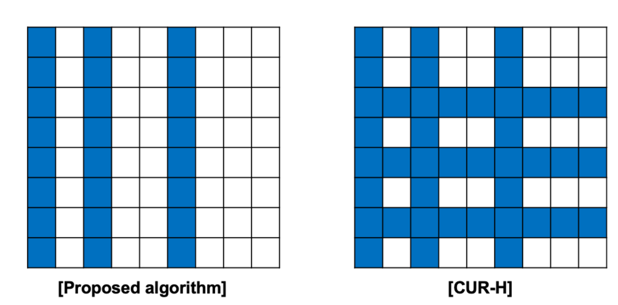
Low rank matrix approximation

$\mathbf{M} \approx \mathbf{AUB}$



Column-sampling formulation for A

- Observation: A matrix of sampled columns $\mathbf{M}\boldsymbol{\psi}$
- Column sampling operator $\boldsymbol{\psi} \in \{0,1\}^{m \times d}$



Sketching side information for B

Ω ($s \times n$) $\hat{Q}S$ ($n \times m$) → B ($s \times m$)

Conventional random sketching matrices
: produces a linear combination of rows, no structure

- CountSketch [4] $\begin{bmatrix} 0&0&0&1&-1&0&0\\ 0&0&1&0&0&1&0\\ 0&-1&0&0&0&0&0\\ 1&0&0&0&0&0&-1 \end{bmatrix}$
- Randomized Hadamard Transform $\frac{1}{2}\begin{bmatrix} 1&1&1&1\\ 1&-1&1&-1\\ 1&1&-1&-1\\ 1&-1&-1&1 \end{bmatrix}$

Proposed sketching matrices $\boldsymbol{\Omega}$

- Randomized row sampling matrix
- Maintains polynomial structure

$\begin{bmatrix} 0&0&0&0&1&0\\ 0&0&1&0&0&0\\ 0&1&0&0&0&0\\ 1&0&0&0&0&0 \end{bmatrix}$

Comparison of sampling complexity

Sketching matrix	Sampling complexity
CountSketch	$s = O(d^2/\varepsilon^2)$
Randomized Hadamard Trans.	$s = O(d/\varepsilon^2)$
Proposed method	$s = O(\mu(A) \cdot d\log d/\varepsilon^2)$

4 Key results

Theorem 1. If $s \geq O(\frac{\mu(\mathbf{A})}{(1-\epsilon)^2} d \log \frac{d}{\delta})$ and assuming $\text{rank}(\hat{\mathbf{M}}) \leq d \leq s \leq k \leq \min(m, n)$

$$\|\mathbf{M} - \mathbf{A}\hat{\mathbf{U}}\hat{\mathbf{B}}\|_2^2 \leq O(\tfrac{n}{\epsilon s}\|\mathbf{M} - \mathbf{AUB}\|_2^2) + O(\sigma_1^2(\mathbf{A})\sigma_1^2(\mathbf{B})\|(\boldsymbol{\Omega}\mathbf{A}^\dagger)\|_2^2\|\mathbf{V}_B^\top - \mathbf{V}_{\hat{B}}^\top\|_2^2)$$

Approximation error between **M** and **AUB**

$\mathbf{U} = \text{argmin}_{\mathbf{X}}\|\mathbf{M\Psi} - \mathbf{AX}(\boldsymbol{\Omega}\mathbf{M\Psi})\|_F^2$

Inaccurate side information estimation error

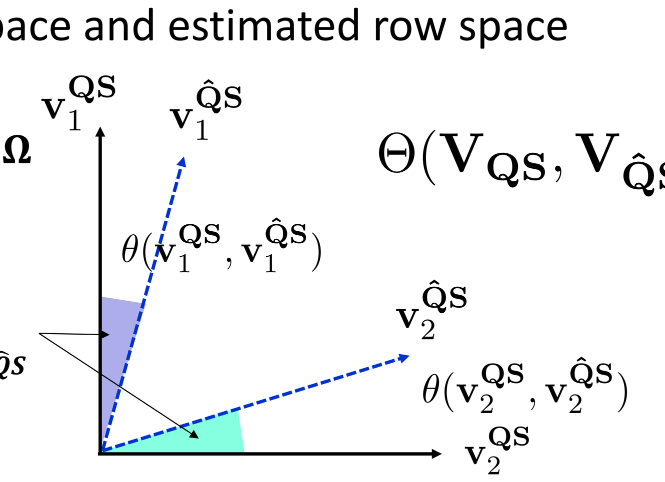
Parts of the error bound

- MSE 1 : Low rank approximation error with true columns & rows
- MSE 2 : Estimation error between sketched true row space and estimated row space

Loss incurred by our column sampling strategy & sketching matrix $\boldsymbol{\Omega}$

Residual between two row spaces

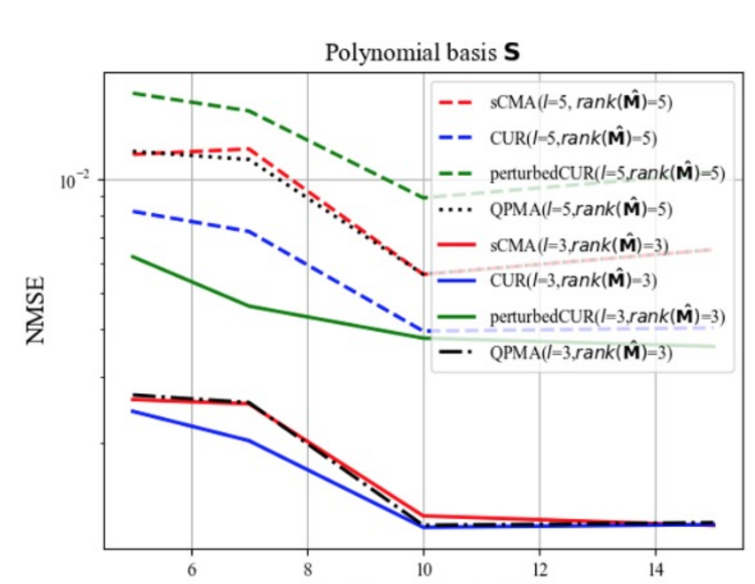
$$\Theta(\mathbf{V_{QS}}, \mathbf{V_{\hat{Q}S}}) \leq O(\sin^{-1}(\frac{\|\mathbf{E}\|_F^2\|(\mathbf{S\Psi})^\dagger\mathbf{S}\|_F}{\delta_2}))$$

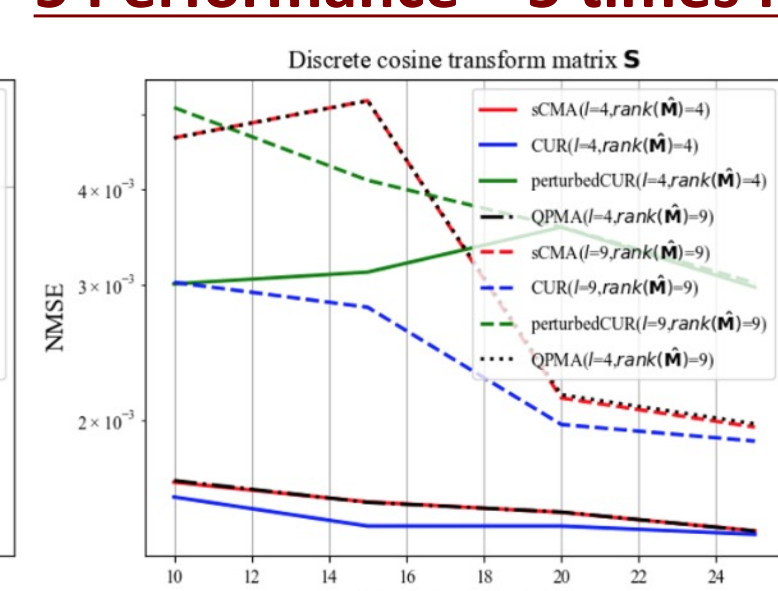


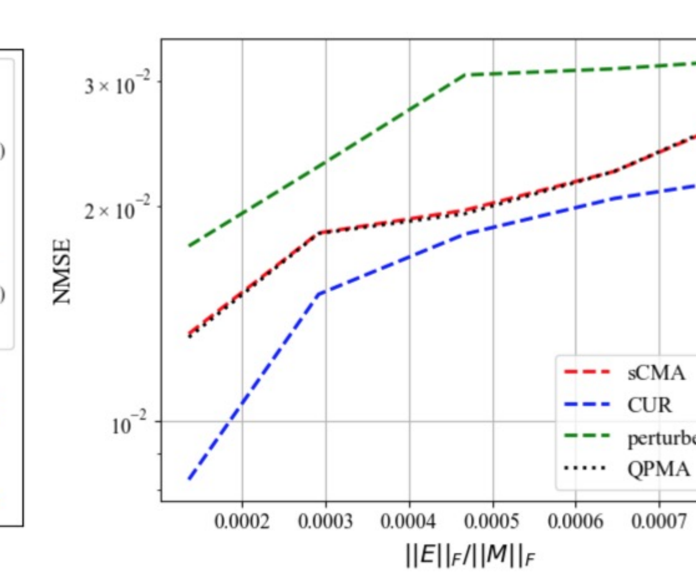
Canonical angles between V_{QS} and $V_{\hat{Q}S}$

Effective eigengap between **QS** and $\hat{\mathbf{Q}}\mathbf{S}$

5 Performance – 5 times faster







	Column	Full matrix	Row		Column	Full matrix	Row
CUR+ [2]	True d columns	M	True s rows	QPMA [1]	True d columns	$\hat{Q}S$	$V_{\hat{Q}S}^T$
Perturbed CUR [Hamm,2021]	True d columns	$\tilde{M}$	Perturbed s rows	sCMA	True d columns	$\hat{Q}S$	s rows of $\hat{Q}S$

6 sCMA Algorithm

- Input : $\mathbf{M\psi}$, side information $\mathbf{S}$, random row sketching $\boldsymbol{\Omega}$
- Parameters : degree of polynomial l, s and d

Unknown coefficient estimation

- Solve least-square problem to obtain $\hat{\mathbf{Q}} := \mathbf{A}(\mathbf{S\Psi})^\dagger$
- Obtain $\hat{\mathbf{Q}}\mathbf{S}$

Row sketch estimation

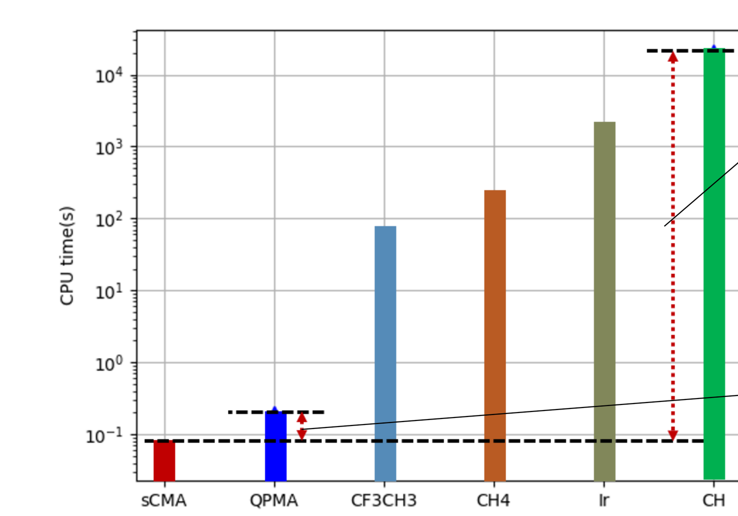
- Obtain $\hat{\mathbf{B}} := \boldsymbol{\Omega}\hat{\mathbf{Q}}\mathbf{S}$

Low rank matrix approximation

- Solve $\hat{\mathbf{U}} = \text{argmin}_{\mathbf{X}}\|\mathbf{M\Psi} - \mathbf{AX}\hat{\mathbf{B}}\mathbf{\Psi}\|_F^2$
- Obtain $\hat{\mathbf{U}} = \mathbf{A}^\dagger\mathbf{A}(\boldsymbol{\Omega}\hat{\mathbf{Q}}\mathbf{S\Psi})^\dagger$
- Construct the low rank approximation

$$\hat{\mathbf{M}} = \mathbf{A}\hat{\mathbf{U}}\hat{\mathbf{B}} = \mathbf{A}(\boldsymbol{\Omega}\hat{\mathbf{Q}}\mathbf{S\Psi})^\dagger\hat{\mathbf{B}}$$

Time complexity



- **sCMA**: max (O(d^2s), O(d^3)) = **O(d^2s)** (s ≥ O(d log d))
- **QPMA : O(ndr) + O(nmr)**

No dependence on n

7 Comparison to prior art [1,3]

M_d: The best rank- d approximation

- **Theorem 1** $\|\mathbf{M} - \hat{\mathbf{M}}\|_2 \leq O(\sqrt{\frac{n}{s}}\,\|\mathbf{M} - \mathbf{M}_d\|_2) + \epsilon\|\tilde{\mathbf{M}}\|_2$
- In [1,3], scaling of $(1 + \epsilon_1)$

if perfect row space information

8 Conclusions

- sCMA preserves the polynomial structure of the data
- $\boldsymbol{\Omega}$ randomly samples rows and its sparsity reduces the computational time 5-6 times compared to QPMA
- Informative columns & rows sampling sketching is motivated

- [1] Chae, Narayanamurthy & Mitra "Matrix approximation with side information: When column sampling is enough." *IEEE TSP, accepted*
- [2] Xu, Jin & Zhou, ICML' 15
- [3] Drineas, Mahoney & Muthukrishnan,,SIAM Journal on Matrix Analysis and Applications,", 08
- [4] Charikar et al., TCS, 04

Acknowledgments: NSF CCF-1817200, ARO W911NF1910269, DOE DE-SC0021417, Swedish Research Council 2018-04359, NSF CCF-2008927, NSF CCF-2200221, ONR 503400-78050, ONR N00014-15-1-2550 and USC+Amazon Trust AI center.