RICE

A DEIM Induced CUR Factorization

D.C. Sorensen

Co-Author: M. Embree

Support: AFOSR and NSF

Salt Lake City • SIAM CSE 2015 • 14-18 March, 2015

CUR Factorization

Low Rank Approximation (Rank k)

$$\mathbf{A} \approx \mathbf{CUR}, \quad \mathbf{C} = \mathbf{A}(:, \mathbf{q}), \quad \mathbf{R} = \mathbf{A}(\mathbf{p}, :)$$

Purpose:

1) Bases **C** and **R** are *actual samples* of rows, and cols of **A**
Instead of *abstract* bases **V**, **W** from low rank SVD $\mathbf{A} \approx \mathbf{VSW}^T$

2) Data Mining: **C** and **R** are *important* samples (instances) of data

Applications:
Large Scale Scientific Data Analysis: Astronomy, Genetics, Term-Document, many others

Approaches to CUR

CUR approximations can be computed using various different strategies.

- Column pivoted QR factorizations [Stewart 1999]
  cf. Rank Revealing QR factorizations [Gu, Eisenstat 1996]
- Volume optimization [Goreinov, Tyrtyshnikov, Zamarashkin 1997], [Goreinov, Oseledets, Savostyanov, Tyrtyshnikov, Zamarashkin 2010], [Thurau, Kersting, Bauckhage 2012]
- Uniform sampling of columns e.g., [Chiu, Demanet 2012]
- Leverage scores (norms of rows of singular vector matrices) [Drineas, Mahoney, Muthukrishnan 20008], [Mahoney, Drineas 2009], [Boutsidis, Woodruff 2014]

Our DEIM-CUR, like leverage scores, requires (approximate) singular vectors.

CUR Approximation

How to choose p, q, U

Want $\|\mathbf{A} - \mathbf{CUR}\| = \textit{small}$, e.g. $\textit{small} = \mathcal{O}(\sigma_{k+1})$

Approximate Low Rank SVD

Using

$$\mathbf{A} \approx \mathbf{VSW}^T$$

Row indices $\mathbf{p}$ obtained from $\mathbf{V}$. Col indices $\mathbf{q}$ obtained from $\mathbf{W}$.

We use DEIM

The Discrete Empirical Interpolation Method (DEIM) overcomes a computational limitation of POD.
The method was first introduced in a finite element framework by [Barrault, Maday, Nguyen, Patera, 2004], and generalized/extended to discrete ODE's by [Chaturantabut & DCS, 2010].

Options for computing the SVD

Problem size dictates computation of SVD that feeds DEIM.

- For modest m or n,
  use the economy SVD: `[V,S,W] = svd(A,'econ')` .
- Krylov SVD routines compute the largest k singular vectors (`svds`). These algorithms access $\mathbf{A}$ and $\mathbf{A}^T$ through matrix-vector products. Need to access $\mathbf{A}$ often, but need minimal intermediate storage.
- Randomized range-finding techniques can find $\mathbf{V}$ with high probability [Halko, Martinsson, Tropp 2011]. These algorithms also access $\mathbf{A}$ and $\mathbf{A}^T$ through matrix-vector products. Like Krylov methods: access $\mathbf{A}$ often, need minimal intermediate storage.
- Incremental QR factorization approximates the SVD in one pass. Given the economy QR factorization $\mathbf{A} = \widehat{\mathbf{Q}}\widehat{\mathbf{R}}$ for $\widehat{\mathbf{Q}} \in \mathbb{R}^{m\times k}$, $\widehat{\mathbf{R}} \in \mathbb{R}^{k\times k}$, compute the SVD $\widehat{\mathbf{R}} = \widehat{\mathbf{V}}\Sigma\mathbf{W}^*$. Then $\mathbf{A} = (\widehat{\mathbf{Q}}\widehat{\mathbf{V}})\Sigma\mathbf{W}^*$ is an SVD of $\mathbf{A}$ cf. [Stewart 1999], [Baker, Gallivan, Van Dooren, 2011]. Intermediate storage depends on rank and sparsity of $\mathbf{A}$.

Incremental QR Factorization $\quad \mathbf{A} \approx \mathbf{VR}$

for $j = k+1 : n,$

$\quad \mathbf{a} = \mathbf{A}(:, j); \quad \mathbf{f} = (\mathbf{I} - \mathbf{V}\mathbf{V}^T)\mathbf{a}; \quad \rho = \|\mathbf{f}\|; \quad \mathbf{v} = \mathbf{f}/\rho;$

$\quad \mathbf{V} = [\mathbf{V},\ \mathbf{v}]; \quad \mathbf{R} \leftarrow \begin{bmatrix} \mathbf{R} & \mathbf{r} \\ \mathbf{0} & \rho \end{bmatrix}$

$\quad [\sigma, imin] = \min(\mathbf{rnorms});$

$\quad$ **if** $(\sigma > tol^2 \cdot \mathbf{FnormR}),$

$\quad\quad k = k+1;$

$\quad$ **else** $\quad$ % Deflate

$\quad\quad$ **if** $(imin < k+1),$

$\quad\quad\quad \mathbf{R}(imin, :) = \mathbf{R}(k+1, :);$

$\quad\quad\quad \mathbf{V}(:, imin) = \mathbf{V}(:, k+1);$

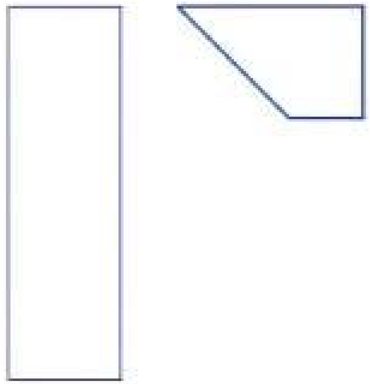
$\quad\quad\quad \mathbf{rnorms}(imin) = \mathbf{rnorms}(k+1);$

$\quad\quad$ **end**

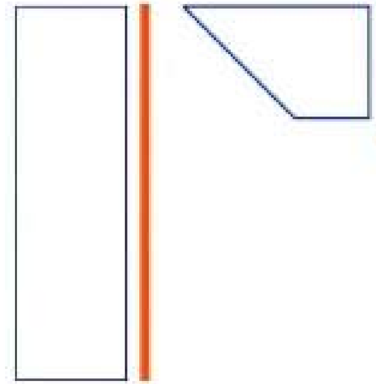
$\quad\quad \mathbf{V} = \mathbf{V}(:, 1:k); \quad \mathbf{R} = \mathbf{R}(1:k, :);$ $\quad$ Update **FnormR**

$\quad$ **end**

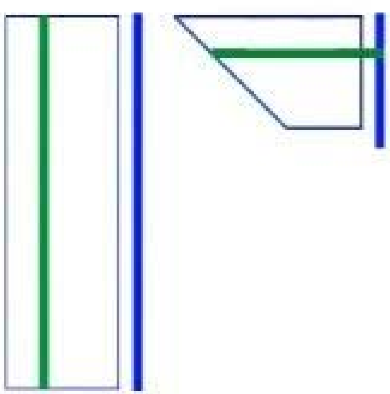
end



Partial **QR** factorization

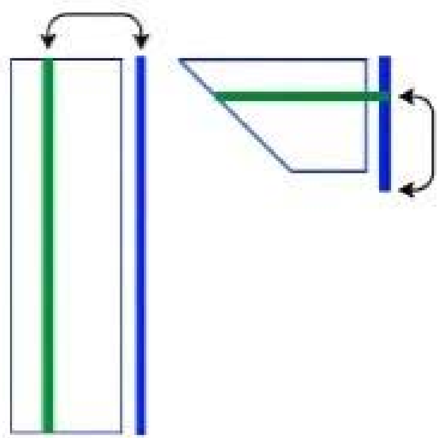


Extend with Gram–Schmidt

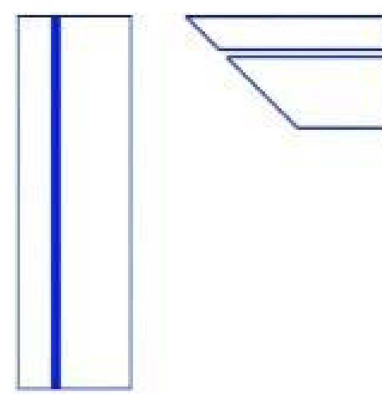


Find $\mathbf{q}_j$ with

$\|\mathbf{R}(j,:)\|^2 < \epsilon^2 \left(\|\mathbf{R}\|_F^2 - \|\mathbf{R}(j,:)\|^2\right)$



Replace $\mathbf{q}_j$, $\mathbf{R}(j,:)$



Truncate last col of **Q** and last row of **R**

Incremental One-Pass QR Factorization: Analysis

How does badly does this simple truncation strategy compromise the accuracy of the factorization?

Let $\mathbf{A}_k = \mathbf{A}(:,1:k)$ denote the first k columns of $\mathbf{A}$.

Theorem. Perform k steps of the incremental QR algorithm to get $\mathbf{A}_k \approx \mathbf{Q}_k\mathbf{R}_k$ using d_k deletions governed by the tolerance ϵ:

$$\mathbf{A}_k \in \mathbb{R}^{n\times k}, \qquad \mathbf{Q}_k \in \mathbb{R}^{n\times(k-d_k)}, \qquad \mathbf{R}_k \in \mathbb{R}^{(k-d_k)\times k}.$$

Then

$$\|\mathbf{A}_k - \mathbf{Q}_k\mathbf{R}_k\|_F \;\le\; \epsilon\, d_k\, \|\mathbf{R}_k\|_F.$$

Note that one can monitor this error bound as the method progresses.

CUR Factorization

Want

$$\mathbf{A} = \mathbf{CUR} + \mathbf{F}, \quad \|\mathbf{F}\| \leq \mu \|\mathbf{A}\|, \quad \mu \ \textit{small}$$

Given indices $\mathbf{p}, \mathbf{q}$, Construct $\mathbf{U}$ for good approximation

Motivation:

If $\mathbf{A} = \mathbf{CUR}$, exact Construct $\mathbf{Y} \in \mathbb{R}^{m \times k}$, $\mathbf{Z} \in \mathbb{R}^{n \times k}$:

$$\mathbf{Y}^T\mathbf{C} = \mathbf{RZ} = \mathbf{I}_k,$$

implies

$\mathbf{CY}^T$ *left projects* onto $Ran(\mathbf{C})$, $\mathbf{ZR}$ *right projects* onto $Col(\mathbf{R})$

Then

$$\mathbf{Y}^T\mathbf{AZ} = (\mathbf{Y}^T\mathbf{C})\mathbf{U}(\mathbf{RZ}) = \mathbf{U} \quad \Rightarrow \quad \mathbf{A} = \mathbf{CY}^T\mathbf{AZR}.$$

Choices for $\mathbf{Y}, \mathbf{Z}$

In General, given $\mathbf{p}, \mathbf{q}$, $\mathbf{Y}, \mathbf{Z}$, (with $\mathbf{Y}^T\mathbf{C} = \mathbf{R}\mathbf{Z} = \mathbf{I}_k$),

$$\mathbf{U} = \mathbf{Y}^T\mathbf{A}\mathbf{Z} \text{ and } \mathbf{F} \equiv \mathbf{A} - \mathbf{C}\mathbf{U}\mathbf{R}.$$

Reproduce: **Rows** $\mathbf{C} = \mathbf{A}(\mathbf{p}, :)$, **Cols** $\mathbf{R} = \mathbf{A}(:, \mathbf{q})$

Interpolatory "Inverses":

$$\mathbf{Y}^T = (\mathbf{P}^T\mathbf{C})^{-1}\mathbf{P}^T \text{ and } \mathbf{Z} = \mathbf{Q}(\mathbf{R}\mathbf{Q})^{-1}$$

$$\mathbf{P}^T\mathbf{a} = \mathbf{a}(\mathbf{p}) \text{ and } \mathbf{b}^T\mathbf{Q} = \mathbf{b}^T(\mathbf{q})$$

Interpolatory $\mathbf{Y}, \mathbf{Z}$

$$\mathbf{Y}^T = (\mathbf{P}^T\mathbf{C})^{-1}\mathbf{P}^T \text{ and } \mathbf{Z} = \mathbf{Q}(\mathbf{R}\mathbf{Q})^{-1}$$

Interpolatory Projectors: $\mathbf{C}\mathbf{Y}^T$ and $\mathbf{Z}\mathbf{R}$

Note

$$\mathbf{P}^T\mathbf{C} = \mathbf{C}(\mathbf{p},:) = \mathbf{A}(\mathbf{p},\mathbf{q}), \quad \mathbf{R}\mathbf{Q} = \mathbf{R}(:,\mathbf{q}) = \mathbf{A}(\mathbf{p},\mathbf{q}),$$

Implies $\mathbf{U} = \mathbf{A}(\mathbf{p},\mathbf{q})^{-1}$

This CUR approximation satisfies

$$\mathbf{A}(:,\mathbf{q}) = \mathbf{C}\mathbf{U}\mathbf{R}(:,\mathbf{q}) \text{ and } \mathbf{A}(\mathbf{p},:) = \mathbf{C}(\mathbf{p},:)\mathbf{U}\mathbf{R}$$

Reproduces Selected Rows, Cols

DEIM Point Selection

Input: $\mathbf{V}$ an $m \times k$ matrix (full rank)

Output: $\mathbf{p}$ integer vector with k non-repeated entries in $\{1 : m\}$

$$
\begin{aligned}
&\mathbf{v} = \mathbf{V}(:,1) \; ; \\
&[\nu, p_1] = \max(|\mathbf{v}|); \\
&\mathbf{p} = [p_1]; \\
&\textbf{for } j = 2, 3, \ldots, k, \\
&\quad \mathbf{v} = \mathbf{V}(:,j) \; ; \\
&\quad \mathbf{c} = \mathbf{V}(\mathbf{p}, 1 : j-1)^{-1}\mathbf{v}(p) \; ; \\
&\quad \mathbf{r} = \mathbf{v} - \mathbf{V}(:, 1 : j-1)\mathbf{c} \; ; \\
&\quad [\nu, p_j] = \max(|\mathbf{r}|); \\
&\quad \mathbf{p} = [\mathbf{p}; p_j]; \\
&\textbf{end}
\end{aligned}
$$

Key Lemmas

Lemma

$\mathrm{rank}(\mathbf{V}) = k \;\Rightarrow\; \mathbf{P}_j^T \mathbf{V}_j$ *nonsingular*, $1 \leq j \leq k$.

Lemma

If $\mathbf{V}^T\mathbf{V} = \mathbf{I}_k$ *then*

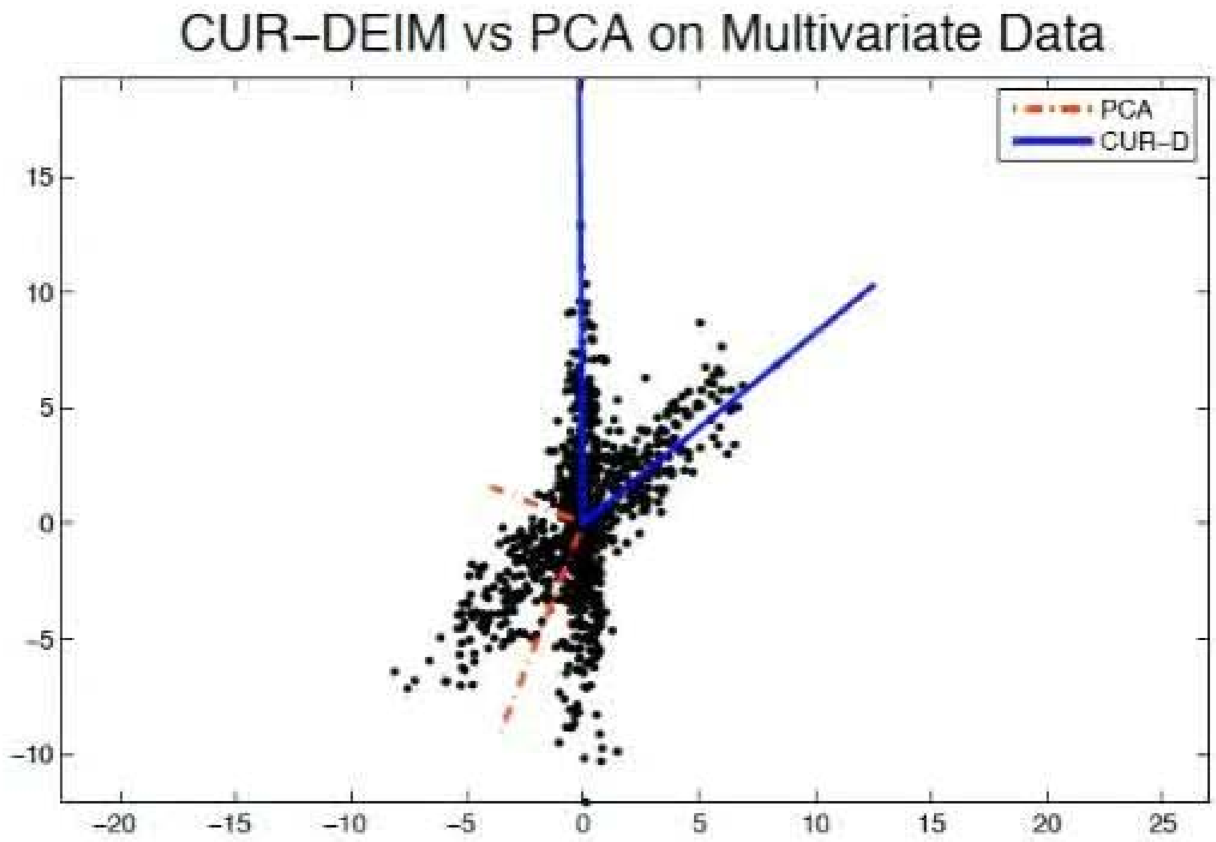
$$\|\mathbf{A} - \mathcal{P}\mathbf{A}\| \leq \|(\mathbf{P}^T\mathbf{V})^{-1}\| \|(\mathbf{I} - \mathbf{V}\mathbf{V}^T)\mathbf{A}\|.$$

If $\mathbf{V}$ *consists of the leading* k *left singular vectors of* $\mathbf{A}$ *then*

$$\|\mathbf{A} - \mathcal{P}\mathbf{A}\| = \|(\mathbf{I} - \mathcal{P})\mathbf{A}\| \leq \|(\mathbf{P}^T\mathbf{V})^{-1}\| \sigma_{k+1}.$$

where $\mathcal{P} = \mathbf{C}\mathbf{Y}^T = \mathbf{C}(\mathbf{P}^T\mathbf{C})^{-1}\mathbf{P}^T$

CUR-DEIM vs PCA on Multivariate Data



CUR based on DEIM versus Leverage Scores

Leverage Scores are a popular technique for computing the CUR factorization; [Mahoney & Drineas, 2009].

- Suppose we have $\mathbf{A} = \mathbf{V\Sigma W}^*$, $\mathbf{V} \in \mathbb{R}^{m\times r}$, $\mathbf{W} \in \mathbb{R}^{n\times r}$.
- To rank the importance of the *rows*, take the 2-norm of each *row* of $\mathbf{V}$:
$$\ell_{r,k} = \|\mathbf{V}(k,:)\|.$$
- To rank the importance of the *columns*, take the 2-norm of each *row* of $\mathbf{W}$:
$$\ell_{c,k} = \|\mathbf{W}(k,:)\|.$$
- Select rows and columns based on the highest leverage scores.
- Leverage scores can be highly influenced by latter columns of $\mathbf{V}$ and $\mathbf{W}$ that correspond to the *smaller* singular values.
- A perturbation theory has been developed by [Ipsen & Wentworth, 2014].

TechTC Data

Term Document Data:

4 datasets that cluster into 2 classes via PCA

Data set 1 $139 \times 15,170$

Source:

TechTC (Technion Repository of Text Categorization Datasets) from The Open Directory Project (ODP) (26)

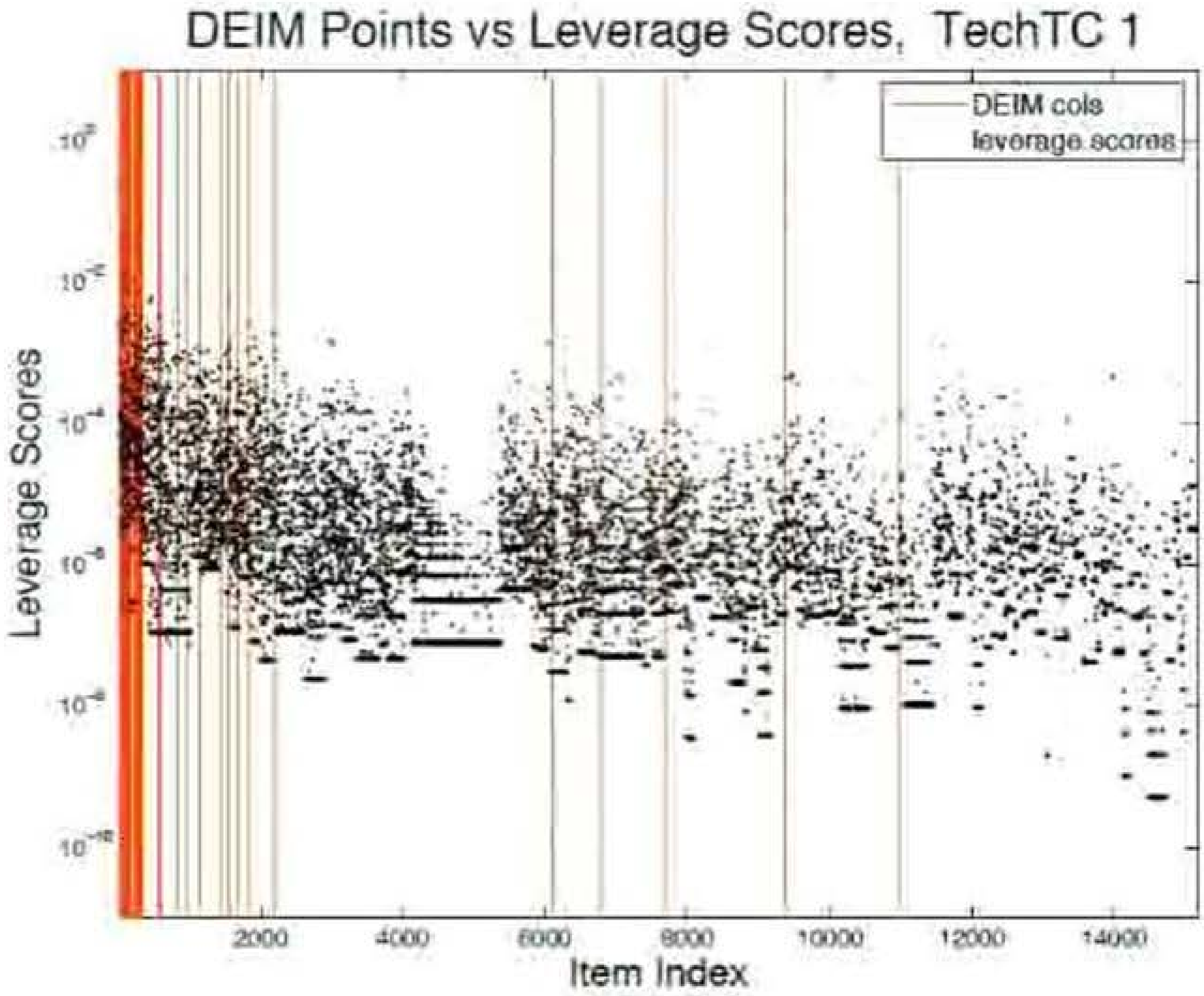
Goal:

Identify Important Terms

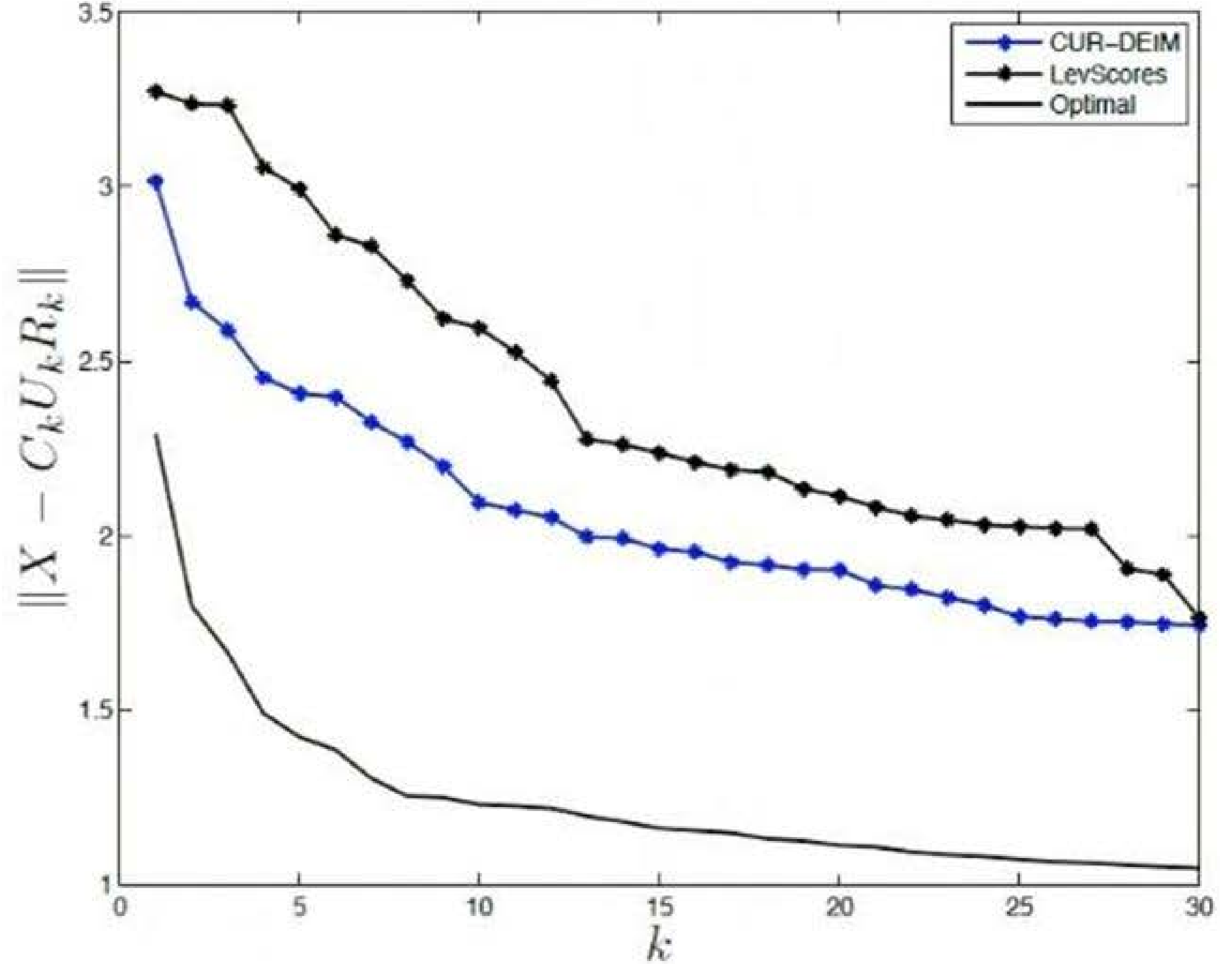
Evansville, Florida, South, Miami

M. Mahoney and P. Drineas
CUR matrix decompositions for improved data analysis
PNAS (2009)

Column Selection: DEIM vs Leverage Score



Comparison 2-Norm Error



DEIM Column and Term Selection

DEIM rank	column indx	term
1	10973	evansville
2	1	florida
3	1547	spacer
4	109	contact
5	209	service
6	50	miami
7	824	chapter
8	1841	health
9	171	information
10	234	events

Tumor Detection

Population:

107 patients, 58 have lung cancer.

Source:

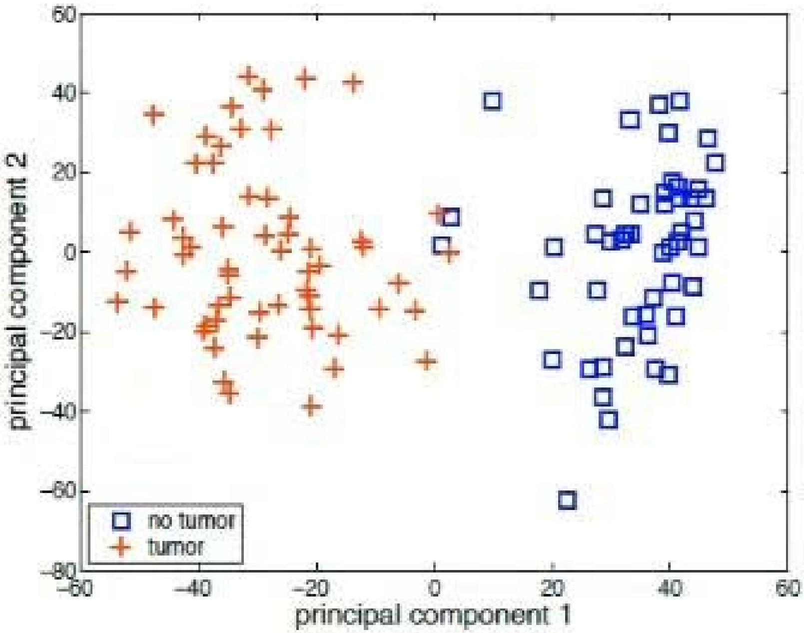
Data set GSE10072 from NIH 22,283 genetic probes

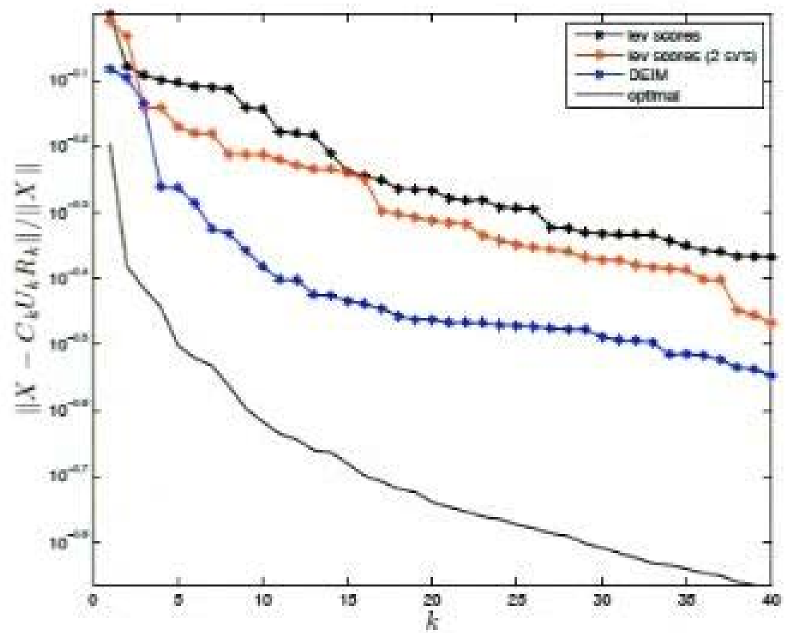
Goals:

- Identify most significant genes
- Discriminate between the healthy and unhealthy patients

A.Kundu, S. Nambirajan and P. Drineas,
Identifying Influential Entries in a Matrix,
arXiv:1310.3556v1 [cs.NA] 14 Oct. 2013

Tumor Detection





Leverage Score Selected Genes

row	probe ID	gene	#sick	#well	lev score
9565	210081_at	AGER	2	45	0.0023317
13766	214387_x_at	SFTPC	6	48	0.0021487
11135	211735_x_at	SFTPC	5	48	0.0021148
9361	209875_s_at	SPP1	50	2	0.0020957
5509	205982_x_at	SFTPC	5	48	0.0020882
9103	209613_s_at	ADH1B	2	47	0.0019479
14827	215454_x_at	SFTPC	0	46	0.0019448
9580	210096_at	CYP4B1	6	44	0.0018578
4239	204712_at	WIF1	5	43	0.0017589
3507	203980_at	FABP4	2	44	0.0016870

DEIM Selected Genes

row	probe ID	gene	#sick	#well
9565	210081_at	AGER	2	45
14270	214895_s_at	ADAM10	8	3
8650	209156_s_at	COL6A2	5	6
11057	211653_x_at	AKR1C2	18	1
14153	214777_at	IGKV4-1	27	3
18976	219612_s_at	FGG	17	17
3831	204304_s_at	PROM1	16	4
3351	203824_at	TSPAN8	17	4
4275	204748_at	PTGS2	18	14
1437	201909_at	RPS4Y1	21	34

Sparse Data Example

Sparse matrix constructed to have steady singular value decay, with a gap: $\mathbf{A} \in \mathbb{R}^{m \times n}$ for $m = 300,000$ and $n = 300$:
Sparse Vectors $\mathbf{x}_j, \quad \mathbf{y}_j$

Small Gap

$$\mathbf{A} = \sum_{j=1}^{10} \frac{2}{j} \mathbf{x}_j \mathbf{y}_j^T + \sum_{j=11}^{300} \frac{1}{j} \mathbf{x}_j \mathbf{y}_j^T .$$

Large Gap

$$\mathbf{A} = \sum_{j=1}^{10} \frac{1000}{j} \mathbf{x}_j \mathbf{y}_j^T + \sum_{j=11}^{300} \frac{1}{j} \mathbf{x}_j \mathbf{y}_j^T .$$

Conclusions

CUR-DEIM

- Good Approximation Properties

$$\|\mathbf{A} - \mathbf{CUR}\| \leq (\eta_p + \eta_q)\sigma_{k+1}$$

- Completely Deterministic with Incremental QR   One Pass
- Questionable Identification of Important Terms (Data items)
- D.C. Sorensen and M. Embree, A DEIM induced CUR factorization, Technical Report CAAM TR14-04, available on line at http://www.caam.rice.edu/tech reports.html,