Today: Nk^2 $\quad N\overset{k}{[\,]} \leftarrow Q$

proj-form LRA $\quad A \approx QQ^TA$

two story holes:

- Q from where?
- Computing Q^TA is a killer

"Interpolative Decomposition" $\to$ ID

Giving up optimality

What problem should we actually solve then?

$$\|A - QQ^T A\|_2 = \min_{\operatorname{rank}(X) \le k} \|A - X\|_2 = \sigma_{k+1}$$

$$\|A - QQ^T A\|_2 = \min_{\operatorname{rank}(X) \le k} \|A - X\|_2 = \sigma_{k+1} \cdot C$$

column basis allowed to larger,

$k+p$

Recap: The Power Method

How did the power method work again?

A diagonalizable w/ eigenvalues $\lambda_1 \ldots \lambda_n$ and eigvec $x_1 \ldots x_n$

$$|\lambda_1| > |\lambda_2| > \ldots |\lambda_n| \geq 0$$

$$y = \alpha_1 x_1 + \cdots \alpha_n x_n$$

$$\frac{Ay}{\lambda_1} = \alpha_1 x_1 + \cdots + \frac{\lambda_n}{\lambda_1} \alpha_n x_n$$

How do we construct the LRA basis?

Put randomness to work:

1. Draw a $n \times l$ Gaussian (iid) matrix Ω
2. $Y = A\Omega$
3. $Y = QR$
4. Carry on with Q

$$= \underbrace{\qquad\qquad} \ \Omega$$

Tweaking the Range Finder (I)

Can we accelerate convergence?

$$Y = (AA^T)^q A\Omega$$

$$A = U\Sigma V^T$$

$$(U\Sigma \cancel{V^T V}\Sigma \cancel{U^T) ^q U}\Sigma V^T$$

$$= U\Sigma^q V^T$$

Tweaking the Range Finder (II)

What is ~~one~~ possible issue with the power method?

- overflow / FP problems
→ QR after each application of A or A^T can help
- Ω not in the correct space
→ errors will eventually amplify the "bigger" left singular vectors

Even Faster Matvecs for Range Finding

$$A\Omega \to n^2\ell$$

Assumptions on Ω are pretty weak–can use more or less anything we want.
$\to$ Make it so that we can apply the matvec $A\Omega$ in $O(n \log \ell)$ time.
How? Pick Ω as a carefully-chosen subsampling of the Fourier transform.

$$zx$$

Errors in Random Approximations

If we use the randomized range finder, how close do we get to the optimal answer?

Theorem

For an $m \times n$ matrix A, a target rank $k \geqslant 2$ and an oversampling parameter $p \geqslant 2$ with $k + p \leqslant \min(m, n)$, with probability $1 - 6 \cdot p^{-p}$,

$$\left|A - QQ^T A\right|_2 \leqslant \left(1 + 11\sqrt{k+p}\sqrt{\min(m,n)}\right)\sigma_{k+1}.$$

(given a few more very mild assumptions on p)

[Halko/Tropp/Martinsson '10, 10.3]

Message: We can *probably* (!) get away with oversampling parameters as small as $p = 5$.

A-posteriori and Adaptivity

The result on the previous slide was *a-priori*. Once we're done, can we find out 'how well it turned out'?

estimate $\|\underbrace{A - QQ^T A}\|_2$

$E = (I - QQ^T)A$

We're interested in $\sigma_1(E)$

rand. vec ω with $\|\omega\|_2 = 1$

Use $\|E\|_2 \approx \dfrac{\|A\omega\|_2}{\|\omega\|_2}$

Adaptive Range Finding: Algorithm

- Compute a small-ish CRA
  - Check whether it's OK (by the estimation proc.)
  - Too big? Continue with more rand. vec.

Rank-revealing/pivoted QR

Sometimes the SVD is too *good* (aka expensive)–we may need less accuracy/weaker promises, for a significant decrease in cost.

$A \in \mathbb{R}^{m \times n}$

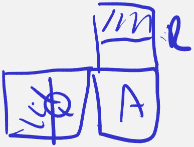
$$A\Pi = QR = Q\begin{pmatrix} R_{11} & R_{12} \\ & R_{22} \end{pmatrix}$$

where

$R_{11} \in \mathbb{R}^{k \times k}$

$\|R_{22}\|_2$ is "small"

$Q^T Q = I$



Using RRQR for LRA

G/vL ch.5

- $\sigma_{k+1} \leq \|R_{22}\|_2$ (it won't beat the SVD)
- To precision $\|R_{22}\|_2$, A has num. rank k.

Interpolative Decomposition (ID): Definition

Would be helpful to know *columns of A* that contribute 'the most' to the rank.
(orthogonal transformation like in QR 'muddies the waters')

For a rank-k matrix A

$$\underset{m\times n}{A} = \underset{m\times k}{A_{[:,J]}} \underset{k\times n}{P}$$

– P is well-conditioned (the magnitude of all entries ≤ 2)

ID: Computation

How do we construct this (from RRQR): (short/fat case)

$$A\Pi = Q\begin{pmatrix} R_{11} & R_{12} \end{pmatrix} \qquad B = QR_{11} = A_{[:,J]}$$

Q: What is P, in terms of the RRQR?

$$P = \begin{bmatrix} \mathrm{Id} & R_{11}^{-1}R_{12} \end{bmatrix}$$

$$BP = QR_{11}\begin{bmatrix} \mathrm{Id} & R_{11}^{-1}R_{12} \end{bmatrix}$$

$$= Q\begin{bmatrix} R_{11} & R_{12} \end{bmatrix} = A\Pi$$

ID Q vs ID A

What does row selection mean for the LRA?

$$A \approx Q Q^T A$$

$$Q = P Q_{[J,:]}$$

$$A_{[J,:]} = \cancel{P_{[J,:]}} Q_{[J,:]} Q^T A$$

$$P A_{[J,:]} = \underbrace{P \; Q_{[J,:]}}_{\approx Q} Q^T A$$

[Martinsson, Rokhlin, Tygert '06]

Demo: Interpolative Decomposition

$$A = U\Sigma V^T$$

$$A_{[J,:]} = U_{[J,:]}\,\Sigma V^T$$

not orth.

What does the ID buy us?

Name a property that the ID has over other factorizations.

All our randomized tools have two stages:

1. Find ONB of approximate range
2. Do actual work only on approximate range

Complexity?

What is the impact of the ID?

ID-based Complexity Reduction

How can we reduce factorization complexity with the ID?