Finding Structure with Randomness

Joel A. Tropp

Computing & Mathematical Sciences
California Institute of Technology
jtropp@acm.caltech.edu

Joint with P.-G. Martinsson and N. Halko
Applied Mathematics, Univ. Colorado at Boulder

Research supported in part by NSF, DARPA, and ONR

Top 10 Scientific Algorithms

list (here, the list is in chronological order; however, the articles appear in no particular order):

- Metropolis Algorithm for Monte Carlo
- Simplex Method for Linear Programming
- Krylov Subspace Iteration Methods
- The Decompositional Approach to Matrix Computations
- The Fortran Optimizing Compiler
- QR Algorithm for Computing Eigenvalues
- Quicksort Algorithm for Sorting
- Fast Fourier Transform
- Integer Relation Detection
- Fast Multipole Method

With each of these algorithms or approaches, there is a person or group receiving credit for inventing or

Source: Dongarra and Sullivan, *Comput. Sci. Eng.*, 2000.

The Decompositional Approach

"The underlying principle of the decompositional approach to matrix computation is that it is not the business of the matrix algorithmicists to solve particular problems but to construct computational platforms from which a variety of problems can be solved."

- A decomposition solves not one but many problems
- Often expensive to compute but can be reused
- Shows that apparently different algorithms produce the same object
- Facilitates rounding-error analysis
- Can be updated efficiently to reflect new information
- Has led to highly effective black-box software

Source: Stewart 2000.

Low-Rank Matrix Approximation

$$\underset{m \times n}{\boldsymbol{A}} \approx \underset{m \times k}{\boldsymbol{B}} \; \underset{k \times n.}{\boldsymbol{C},}$$

Benefits:

- Exposes structure of the matrix
- Allows efficient storage
- Facilitates multiplication with vectors or other matrices

Applications:

- Principal component analysis
- Low-dimensional embedding of data
- Approximating continuum operators with exponentially decaying spectra
- Model reduction for PDEs with rapidly oscillating coefficients

Model Problem

Given:

- An $m \times n$ matrix $\boldsymbol{A}$ with $m \geq n$
- Target rank k
- Oversampling parameter p

Construct an $n \times (k+p)$ matrix $\boldsymbol{Q}$ with orthonormal columns s.t.

$$\|\boldsymbol{A} - \boldsymbol{Q}\boldsymbol{Q}^* \boldsymbol{A}\| \approx \min_{\operatorname{rank}(\boldsymbol{B}) \leq k} \|\boldsymbol{A} - \boldsymbol{B}\|,$$

- $\boldsymbol{Q}\boldsymbol{Q}^*$ is the orthogonal projector onto the range of $\boldsymbol{Q}$
- The basis $\boldsymbol{Q}$ can be used to construct matrix decompositions

From Basis to Decomposition

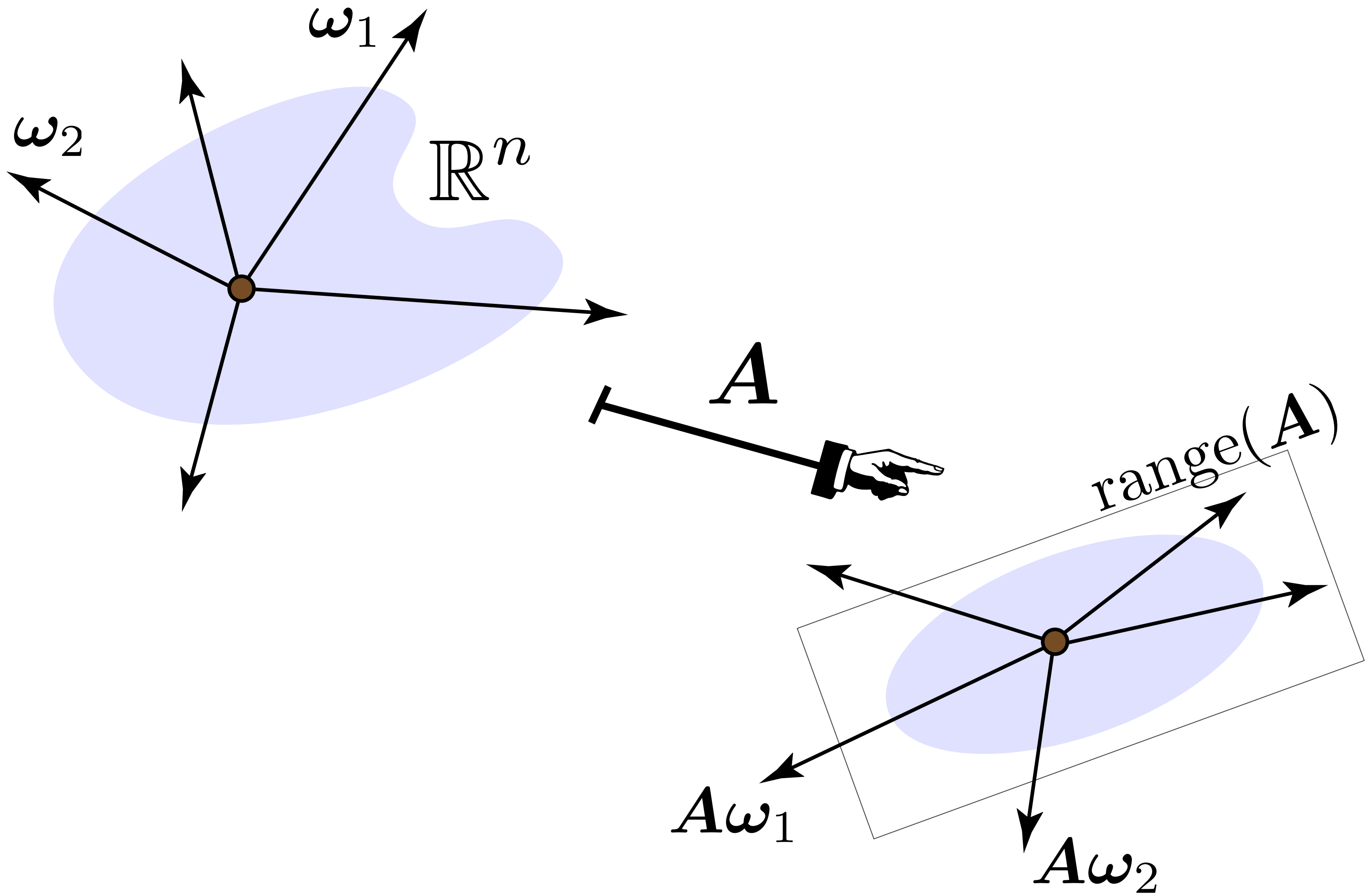
Problem: Given the basis $\boldsymbol{Q}$, where do we get a factorization?

Example: Singular value decomposition

Assume $\boldsymbol{A}$ is $m \times n$ and $\boldsymbol{Q}$ is $m \times k$ where $\boldsymbol{A} \approx \boldsymbol{Q}\boldsymbol{Q}^*\boldsymbol{A}$.

1. Form $k \times n$ matrix $\boldsymbol{B} = \boldsymbol{Q}^*\boldsymbol{A}$
2. Factor $\boldsymbol{B} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^*$
3. Conclude $\boldsymbol{A} \approx (\boldsymbol{Q}\boldsymbol{U})\boldsymbol{\Sigma}\boldsymbol{V}^*$

Random Sampling: Intuition



Proto-Algorithm for Model Problem

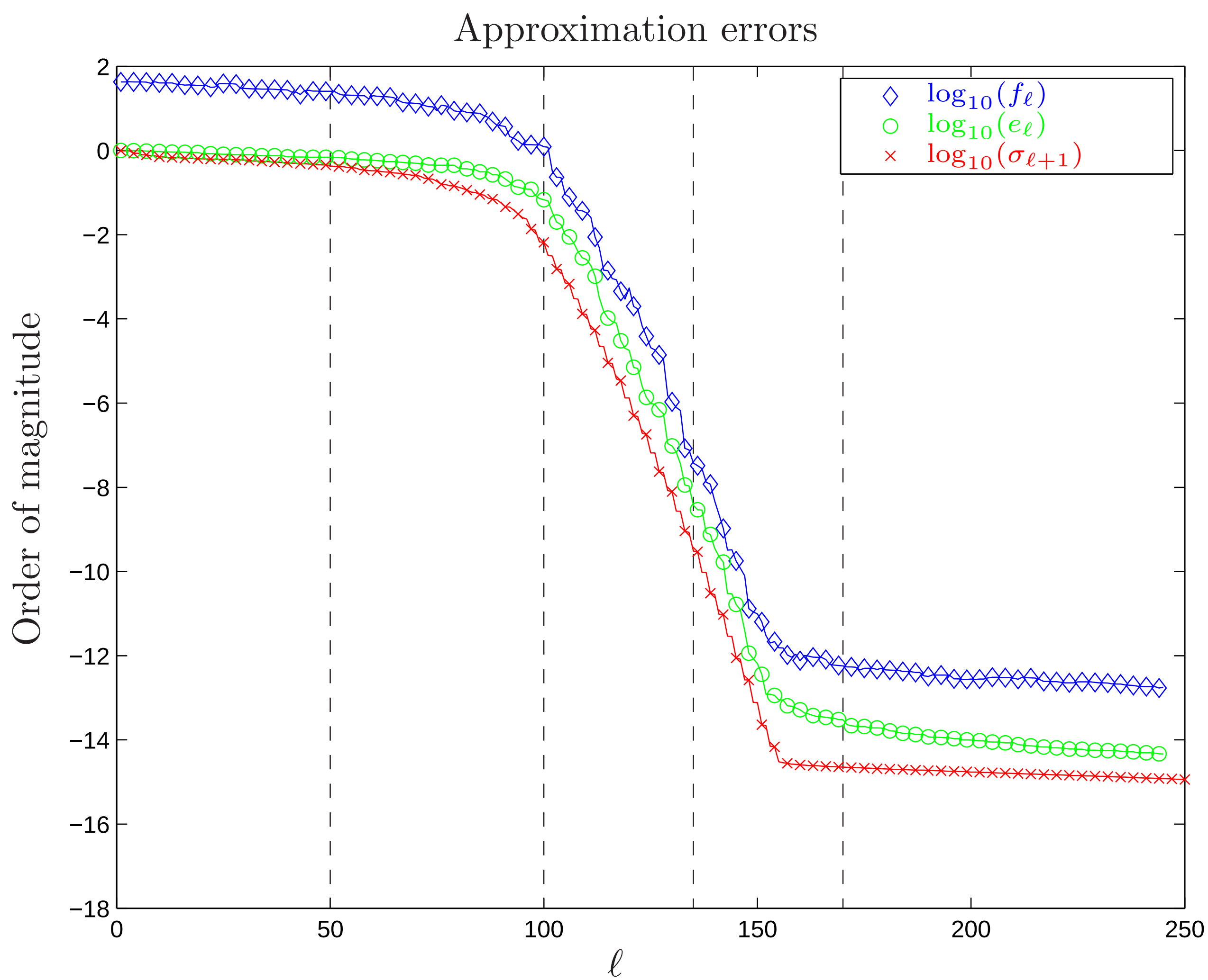
❧ Converting this intuition into a computational procedure...

Input: An $m \times n$ matrix $\boldsymbol{A}$, a target rank k, an oversampling parameter p

Output: An $m \times (k+p)$ matrix $\boldsymbol{Q}$ with orthonormal columns

1. Draw an $n \times (k+p)$ random matrix $\boldsymbol{\Omega}$.
2. Form the matrix product $\boldsymbol{Y} = \boldsymbol{A}\boldsymbol{\Omega}$.
3. Construct an orthonormal basis $\boldsymbol{Q}$ for the range of $\boldsymbol{Y}$.

Approximating a Helmholtz Integral Operator



(Simplified) Error Bound for Proto-Algorithm

Theorem 1. [HMT 2011] **Assume**

- *the matrix* $\boldsymbol{A}$ *is* $m \times n$ *with* $m \geq n$;
- *the optimal error* $\sigma_{k+1} = \min_{\operatorname{rank}(\boldsymbol{B}) \leq k} \|\boldsymbol{A} - \boldsymbol{B}\|$;
- *the test matrix* $\boldsymbol{\Omega}$ *is* $n \times (k+p)$ *standard Gaussian.*

Then *the basis* $\boldsymbol{Q}$ *computed by the proto-algorithm satisfies*

$$\mathbb{E}\,\|\boldsymbol{A} - \boldsymbol{Q}\boldsymbol{Q}^*\boldsymbol{A}\| \leq \left[1 + \frac{4\sqrt{k+p}}{p-1} \cdot \sqrt{n}\right] \sigma_{k+1}.$$

The probability of a substantially larger error is negligible.

Proto-Algorithm + Power Scheme

Problem: The singular values of the data matrix often decay slowly

Remedy: Apply the proto-algorithm to $(\boldsymbol{A}\boldsymbol{A}^*)^q\boldsymbol{A}$ for small q

---

Input: An $m \times n$ matrix $\boldsymbol{A}$, a target rank k, an oversampling parameter p

Output: An $m \times (k+p)$ matrix $\boldsymbol{Q}$ with orthonormal columns

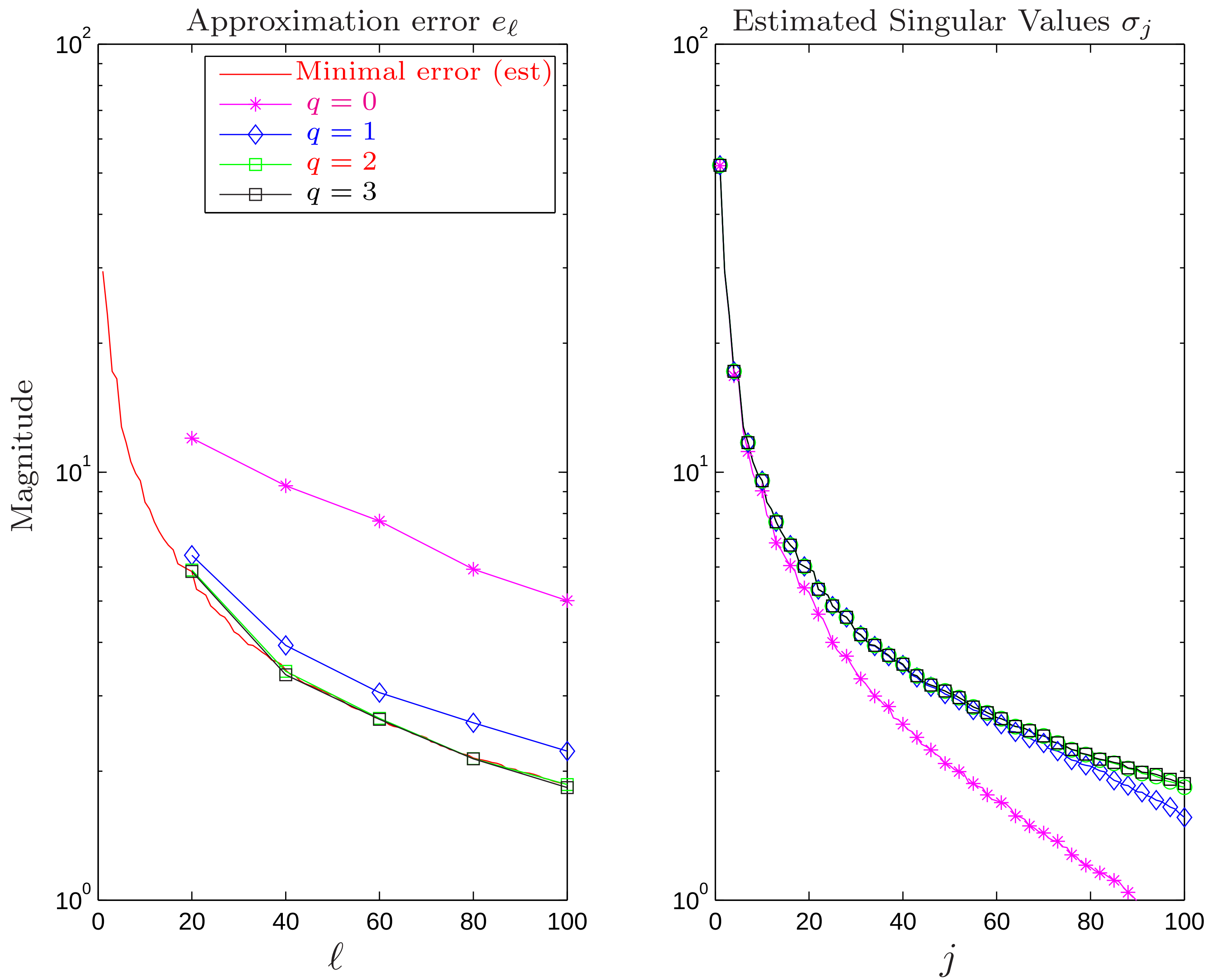
1. Draw an $n \times (k+p)$ random matrix $\boldsymbol{\Omega}$.
2. Form the matrix product $\boldsymbol{Y}_0 = \boldsymbol{A}\boldsymbol{\Omega}$.
3. Sequentially form $\boldsymbol{Y}_k = (\boldsymbol{A}\boldsymbol{A}^*)\boldsymbol{Y}_k$ for $k = 1, 2, \dots, q$.
4. Construct an orthonormal basis $\boldsymbol{Q}$ for the range of $[\,\boldsymbol{Y}_0 \mid \boldsymbol{Y}_1 \mid \dots \mid \boldsymbol{Y}_q\,]$.

Open Question: Can we improve using Lanczos?

Eigenfaces

- Database consists of $7,254$ photographs with $98,304$ pixels each
- Form $98,304 \times 7,254$ data matrix $\widetilde{\boldsymbol{A}}$
- **Total storage:** 5.4 Gigabytes (uncompressed)
- Center each column and scale to unit norm to obtain $\boldsymbol{A}$
- The dominant left singular vectors are called **eigenfaces**
- Attempt to compute first 100 eigenfaces using power scheme

Image: Scholarpedia article "Eigenfaces," 12 October 2009


Approximation error e_ℓ
Minimal error (est)
$q = 0$
$q = 1$
$q = 2$
$q = 3$
Magnitude
10^2
10^1
10^0
0
20
40
60
80
100
ℓ
Estimated Singular Values σ_j
j

(Simplified) Error Bound for Power Scheme

Theorem 2. [HMT 2011] **Assume**

- *the matrix* $\boldsymbol{A}$ *is* $m \times n$ *with* $m \geq n$*;*
- *the optimal error* $\sigma_{k+1} = \min_{\text{rank}(\boldsymbol{B}) \leq k} \|\boldsymbol{A} - \boldsymbol{B}\|$*;*
- *the test matrix* $\boldsymbol{\Omega}$ *is* $n \times (k+p)$ *standard Gaussian.*

Then *the basis* $\boldsymbol{Q}$ *computed by the power scheme satisfies*

$$\mathbb{E}\,\|\boldsymbol{A} - \boldsymbol{Q}\boldsymbol{Q}^*\boldsymbol{A}\| \leq \left[1 + \frac{4\sqrt{k+p}}{p-1} \cdot \sqrt{n}\right]^{1/(2q+1)} \sigma_{k+1}.$$

The probability of a substantially larger error is negligible.

- The power scheme drives the extra factor to one exponentially fast!
- Qualitative improvement for error bound (various authors, 2015)

To learn more...

E-mail: `jtropp@cms.caltech.edu`

Web: `http://users.cms.caltech.edu/~jtropp`

Papers:

- HMT, "Finding Structure with Randomness: Probabilistic Algorithms for Computing Approximate Matrix Decompositions," *SIREV* 2011.
- T, "Improved Analysis of the Subsampled Randomized Hadamard Transform," *AADA* 2011.