Low Rank Approximation
Lecture 2

Daniel Kressner
Chair for Numerical Algorithms and HPC
Institute of Mathematics, EPFL
daniel.kressner@epfl.ch

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Basic randomized algorithm for low-rank approx

Must read: Halko/Martinsson/Tropp'2010: Finding Structure with Randomness...

Randomized Algorithm:

1. Draw standard Gaussian random matrix $\Omega \in \mathbb{R}^{n \times k}$.
2. Perform block mat-vec $Y = A\Omega$.
3. Compute (economic) QR decomposition $Y = QR$.
4. Form $B = Q^T A$.

Exact recovery: If A has rank k, we recover $\widehat{A} = A$ with probability 1.

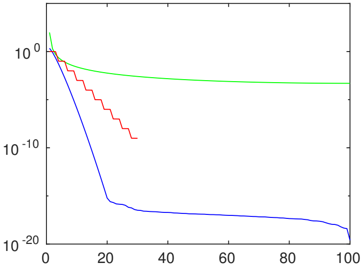
Three test matrices

(a) The 100×100 Hilbert matrix A defined by $A(i,j) = 1/(i+j-1)$.

(b) The matrix A defined by $A(i,j) = \exp(-\gamma|i-j|/n)$ with $\gamma = 0.1$.

(c) 30×30 diagonal matrix with diagonal entries

$$1, 0.99, 0.98, \frac{1}{10}, \frac{0.99}{10}, \frac{0.98}{10}, \frac{1}{100}, \frac{0.99}{100}, \frac{0.98}{100}, \dots$$



Singular values of test matrices

Randomized algorithm applied to test matrices

errors measured in spectral norm:

(a) Hilbert matrix, $k = 5$:

Exact	mean	std
0.0019	0.0092	0.0099

(b) Matrix with slower decay, $k = 25$:

Exact	mean	std
0.0034	0.012	0.002

(c) Matrix with staircase sv, $k = 7$:

Exact	mean	std
0.010	0.038	0.025

Randomized algorithm applied to test matrices

errors measured in Frobenius norm:

(a) Hilbert matrix, $k = 5$:

Exact	mean	std
0.0019	0.0093	0.0099

(b) Matrix with slower decay, $k = 25$:

Exact	mean	std
0.011	0.024	0.001

(c) Matrix with staircase sv, $k = 7$:

Exact	mean	std
0.014	0.041	0.024

Basic randomized algorithms for low-rank approx

Add oversampling. (usually small) integer p

Randomized Algorithm:

1. Draw standard Gaussian random matrix $\Omega \in \mathbb{R}^{n\times(k+p)}$.
2. Perform block mat-vec $Y = A\Omega$.
3. Compute (economic) QR decomposition $Y = QR$.
4. Form $B = Q^T A$.
5. Optional: Set $\mathcal{T}_k(A) \approx \widehat{A} := Q\mathcal{T}_k(B)$

Gold standard best rank-k approximation error:

- spectral norm: σ_{k+1}
- Frobenius norm: $\sqrt{\sigma_{k+1}^2 + \cdots + \sigma_n^2}$.

Randomized algorithm applied to test matrices

errors measured in spectral norm:

(a) Hilbert matrix, $k = 5$:

Exact	mean	std	
0.0019	0.0092	0.0099	$p = 0$
0.0019	0.0026	0.0019	$p = 1$
0.0019	0.0019	0.0001	$p = 2$

(b) Matrix with slower decay, $k = 25$:

Exact	mean	std	
0.0034	0.012	0.002	$p = 0$
0.0034	0.011	0.0017	$p = 1$
0.0034	0.010	0.0015	$p = 2$
0.0034	0.0064	0.0008	$p = 10$
0.0034	0.0037	0.0002	$p = 25$

(c) Matrix with staircase sv, $k = 7$:

Exact	mean	std	
0.010	0.038	0.025	$p = 0$
0.010	0.021	0.012	$p = 1$
0.010	0.012	0.005	$p = 2$

Analysis: basic setting

Goal: Say something sensible about $\|(I - QQ^T)A\|$. Expected value, tail bounds, ...

Let $X \sim \mathcal{N}(0, I_n)$ be m-dimensional Gaussian random vector. Then

- $V^T X \sim \mathcal{N}(0, V^T V)$
- if V is $m \times n$ matrix with orthonormal columns $\rightsquigarrow V^T X \sim \mathcal{N}(0, I_n)$

This allows us to assume w.l.o.g. in the analysis that $A = \Sigma$ diagonal (and square).

Analysis: $k = 1, p = 0$

Partition

$$\Omega = \begin{pmatrix} \omega_1 \\ \omega_2 \end{pmatrix}, \qquad A = \begin{pmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{pmatrix}.$$

Then

$$Y = A\Omega = \begin{pmatrix} \sigma_1\omega_1 \\ \Sigma_2\omega_2 \end{pmatrix} = \sigma_1\omega_1 \begin{pmatrix} 1 \\ \frac{1}{\sigma_1\omega_1}\Sigma_2\omega_2 \end{pmatrix}.$$

Problem: Need to control norm of $\frac{1}{\sigma_1\omega_1}\Sigma_2\omega_2$ but ω_1^{-1} is Cauchy distribution with undefined mean and variance.

Need to consider $p \geq 2$.

Analysis: $k = 1, p \geq 2$

Lemma

Let $\Pi_{\tilde{\mathcal{Y}}}$, $\Pi_{\mathcal{Y}}$ be orthogonal projectors onto subspaces $\tilde{\mathcal{Y}} \subset \mathcal{Y}$. Then

$$\|\Pi_{\tilde{\mathcal{Y}}} A\|_F \geq \|\Pi_{\mathcal{Y}} A\|_F, \qquad \|(I - \Pi_{\tilde{\mathcal{Y}}})A\|_F \leq \|(I - \Pi_{\mathcal{Y}})A\|_F.$$

for any matrix A.

Proof. EFY.

$$Y = A\Omega = \begin{pmatrix} \sigma_1 \Omega_1 \\ \Sigma_2 \Omega_2 \end{pmatrix}, \qquad \tilde{Y} = \begin{pmatrix} 1 \\ \sigma_1^{-1} \Sigma_2 \Omega_2 \Omega_1^{\dagger} \end{pmatrix},$$

where $\Omega_1^{\dagger}$ is pseudoinverse of Ω_1. In this case, Ω_1 is (nonzero) row vector and $\Omega_1^{\dagger} = \Omega_1^T / \|\Omega_1\|_2^2$. We have

$$\text{range}(\tilde{Y}) \subset \text{range}(Y).$$

Thus, with $f = \sigma_1^{-1} \Sigma_2 \Omega_2 \Omega_1^{\dagger}$, we get

$$\|(I - QQ^T)A\|_F \leq \|(I - \tilde{Q}\tilde{Q}^T)A\|_F, \qquad \tilde{Q} = \frac{1}{\sqrt{1 + \|f\|_2^2}} \begin{pmatrix} 1 \\ f, \end{pmatrix}$$

Analysis: $k = 1, p \geq 2$

$$\|(I - \tilde{Q}\tilde{Q}^T)A\|_F^2 = \|A\|_F^2 - \|\tilde{Q}^T A\|_F^2$$

$$\begin{aligned} \|\tilde{Q}^T A\|_F^2 &= \frac{1}{1 + \|f\|_2^2}(\sigma_1^2 + \|\Sigma_2 f\|_2^2) \geq \frac{\sigma_1^2}{1 + \|f\|_2^2} \\ &\geq \sigma_1^2(1 - \|f\|_2^2) = \sigma_1^2 - \|\Sigma_2 \Omega_2 \Omega_1^\dagger\|_2^2 \end{aligned}$$

In summary, we have:

Lemma

$\|(I - \tilde{Q}\tilde{Q}^T)A\|_F^2 \leq \|\Sigma_2\|_F^2 + \|\Sigma_2 \Omega_2 \Omega_1^\dagger\|_F^2$.

To analyze red term, we use

$$\mathbb{E}\|\Sigma_2 \Omega_2 \Omega_1^\dagger\|_F^2 = \mathbb{E}\big(\mathbb{E}\big(\|\Sigma_2 \Omega_2 \Omega_1^\dagger\|_F^2 \,|\, \Omega_1\big)\big) = \|\Sigma_2\|_F^2 \cdot \mathbb{E}\|\Omega_1^\dagger\|_F^2.$$

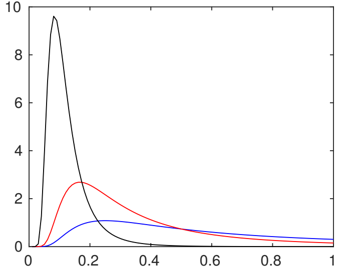
(EFY: Prove $\mathbb{E}\|A\Omega B\|_F^2 = \|A\|_F^2 \|B\|_F^2$ for Gaussian matrix Ω and constant matrices A, B.)

Analysis: $k = 1, p \geq 2$

$$\|\Omega_1^\dagger\|_F^2 = \frac{1}{\|\Omega_1\|_F^2} = \Big(\sum_{i=1}^{p} \Omega_{1,i}^2\Big)^{-1} \sim \mathrm{Inv} - \chi^2(p+1),$$

the inverse-chi-squared distribution with p degrees of freedom. Pdf given by

$$\frac{2^{-(p+1)/2}}{\Gamma((p+1)/2)} x^{-(p+1)/2-1} \exp(-1/(2x)).$$



pdf for $p = 1$, $p = 3$, $p = 9$

Analysis: $k = 1, p \geq 2$

Standard results imply

- $\mathbb{E}\|\Omega_1^\dagger\|_F^2 = (p-1)^{-1}$
- $\mathbb{P}\left[\|\Omega_1\|_F^2 \leq p+1-t\right] \leq \exp\left(-\frac{t^2}{4(p+1)}\right)$

Theorem

For $k = 1, p \geq 2$, we have

$$\mathbb{E}\|(I - QQ^T)A\|_F \leq \sqrt{1 + \frac{1}{p-1}}\|\Sigma_2\|_F.$$

Probability of deviating from this upper bound decays exponentially, by tail bound for χ^2.

Analysis: general k, $p \geq 2$

Lemma

$\|(I - QQ^T)A\|_F^2 \leq \|\Sigma_2\|_F^2 + \|\Sigma_2\Omega_2\Omega_1^\dagger\|_F^2$.

Proof.

$$Y = A\Omega = \begin{pmatrix} \Sigma_1\Omega_1 \\ \Sigma_2\Omega_2 \end{pmatrix}, \qquad \tilde{Y} = \begin{pmatrix} I_k \\ F \end{pmatrix}, \qquad F = \Sigma_2\Omega_2\Omega_1^\dagger\Sigma_1^{-1}.$$

ONB

$$\tilde{Q} = \begin{pmatrix} I_k \\ F \end{pmatrix} (I_k + F^T F)^{-1/2}.$$

$$\|(I - QQ^T)A\|_F^2 \leq \|(I - \tilde{Q}\tilde{Q}^T)A\|_F^2 = \|A\|_F^2 - \|\tilde{Q}^T A\|_F^2.$$

$$\begin{aligned} \|\tilde{Q}^T A\|_F^2 &= \left\|(I_k + F^T F)^{-1/2}\Sigma_1\right\|_F^2 + \left\|(I_k + F^T F)^{-1/2} F^T \Sigma_2\right\|_F^2 \\ &\geq \left\|(I_k + F^T F)^{-1/2}\Sigma_1\right\|_F^2 = \operatorname{trace}(\Sigma_1 (I_k + F^T F)^{-1}\Sigma_1) \\ &\geq \operatorname{trace}(\Sigma_1 (I_k - F^T F)\Sigma_1) = \|\Sigma_1\|_F^2 - \|\Sigma_2\Omega_2\Omega_1^\dagger\|_F^2 \end{aligned}$$

Analysis: general k, $p \geq 2$

Again use

$$\mathbb{E}\|\Sigma_2\Omega_2\Omega_1^\dagger\|_F^2 = \|\Sigma_2\|_F^2 \cdot \mathbb{E}\|\Omega_1^\dagger\|_F^2.$$

By standard results in multivariate statistics[1], we have

$$\mathbb{E}\|\Omega_1^\dagger\|_F^2 = \frac{k}{p-1}.$$

Theorem

For $p \geq 2$, we have

$$\mathbb{E}\|(I - QQ^T)A\|_F \leq \sqrt{1 + \frac{k}{p-1}}\|\Sigma_2\|_F,$$

$$\mathbb{E}\|(I - QQ^T)A\|_2 \leq \left(1 + \sqrt{\frac{k}{p-1}}\right)\|\Sigma_2\|_2 + \frac{e\sqrt{k+p}}{p}\|\Sigma_2\|_F,$$

For proof of spectral norm and tail bounds, see [Halko/Martinsson/Tropp'2010].

[1] R. J. Muirhead, Aspects of Multivariate Statistical Theory, Wiley, New York, NY, 1982.

Randomized subspace iteration

1. Draw standard Gaussian random matrix $\Omega \in \mathbb{R}^{n\times(k+p)}$.
2. Perform block mat-vec $Y_0 = A\Omega$.
3. Perform q steps of subspace iteration: $Y = (AA^T)^q Y_0$.
4. Compute (economic) QR decomposition $Y = QR$.
5. Form $B = Q^T A$.
6. Optional: Set $\mathcal{T}_k(A) \approx \widehat{A} := Q\mathcal{T}_k(B)$

Theorem

For $p \geq 2$, $\mathbb{E}\|(I - QQ^T)A\|_2$ is bounded by

$$\left[\left(1 + \frac{k}{p-1}\right)\sigma_{k+1}^{2q+1} + \frac{e\sqrt{k+p}}{p}\left(\sigma_{k+1}^{2(2q+1)} + \cdots\right)^{1/2}\right]^{1/(2q+1)}$$

$$\leq \quad \sigma_{k+1}\left[1 + \frac{k}{p-1} + \frac{e\sqrt{(k+p)(n-k)}}{p}\right]^{1/(2q+1)}$$

EFY. Implement this algorithm and repeat the experiments from Slide 7.

A posteriori error estimate and adaptive choice of k

Lemma

Let $\omega^{(i)}$, $i = 1, \ldots, r$ be n-dimensional random Gaussian vectors. Then for any $m \times n$ matrix C the inequality

$$\|C\|_2 \leq 10\sqrt{2/\pi} \max_{i=1,\ldots,r} \|C\omega^{(i)}\|_2.$$

holds with probability $1 - 10^{-r}$.

Proof. See [F. Woolfe, E. Liberty, V. Rokhlin, and M. Tygert, A fast randomized algorithm for the approximation of matrices, Appl. Comp. Harmon. Anal., 25 (2008), pp. 335–366].

Given ONB Q returned by randomized algorithm, use result for $C = (I - QQ^T)A$. Can be combined into adaptive algorithm for choosing k.

Assuming that one knows $\|A\|_F$ the Frobenius norm error can be estimated from

$$\|A\|_F^2 = \|QQ^T A\|_F^2 + \|(I - QQ^T)A\|_F^2.$$