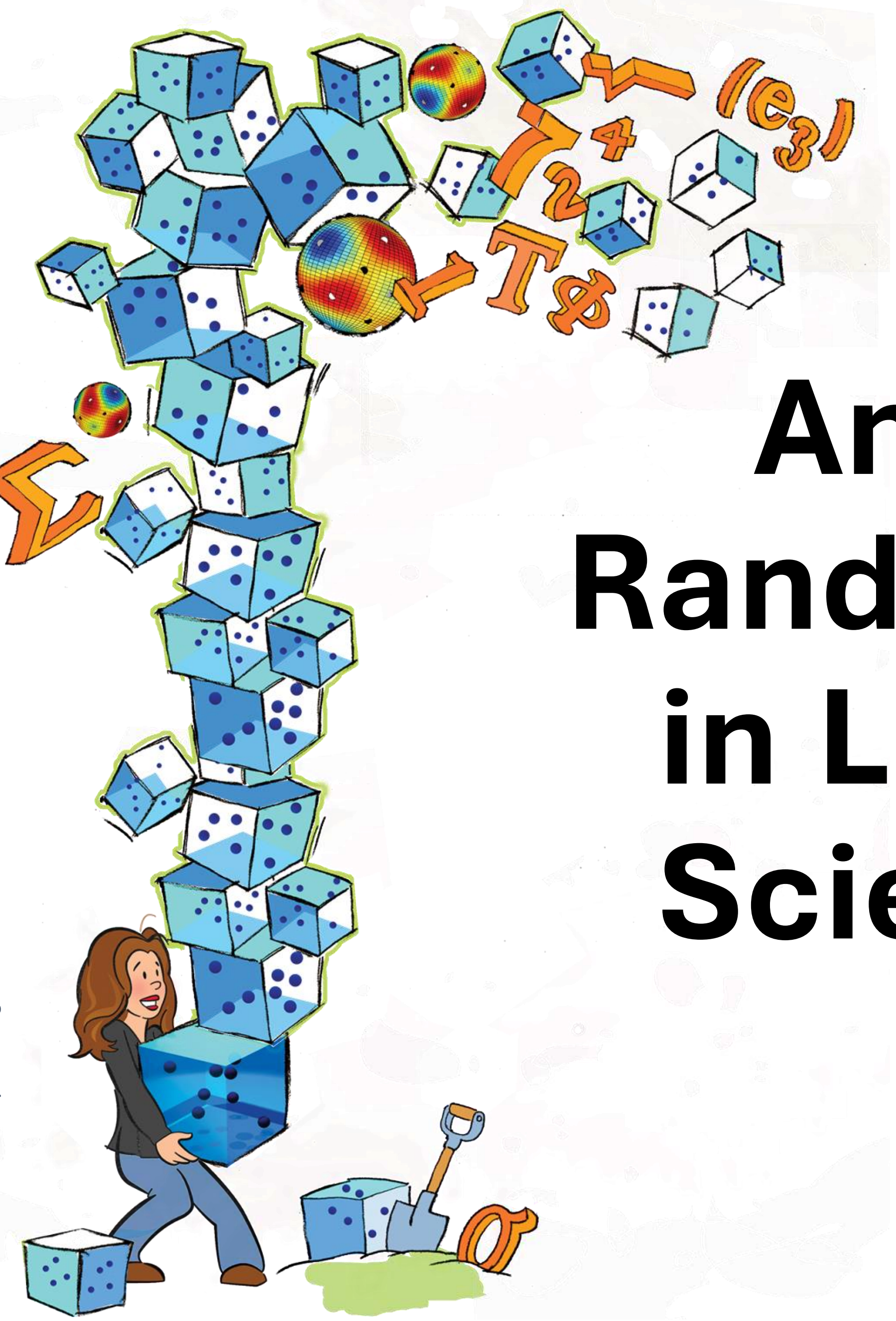


Illustration by Chris Brigman

MathSci.ai

An Introduction to Randomized Algorithms in Linear Algebra and Scientific Computing

Tammy Kolda & Brett Larsen

May 13, 2024

Resources and Schedule

MathSci.ai

- **Data for Labs:** tgkolda/randalgslabs (github.com)
- **Schedule**
  - 11:00 – Part I
    - Random Gaussian matrices
    - Randomized range finder
    - Randomized SVD
  - 11:35 – Labs for Part I
  - 11:50 – Part II
  - 12:25 – Labs for Part II
  - 12:40 – Adjourn

Part I

Randomized Range Finder & Randomized SVD

Notation

- $[n] = \{ 1, 2, \ldots, n \}$
- Expected value: $\mathbb{E}[X]$ or $\mathbb{E}X$
- Variance: $\mathrm{Var}[X] \equiv \mathbb{E}[(X - \mathbb{E}X)^2]$
- Probability: $\mathbb{P}\{ X > 5 \}$
- Independent and identically distributed: i.i.d.

Asymptotic versus Non-Asymptotic

Theorem (Asymptotic) Let $X_1, X_2, \ldots$ be a sequence of i.i.d. random variables with finite expectation. Then

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} X_i = \mathbb{E}[X]$$

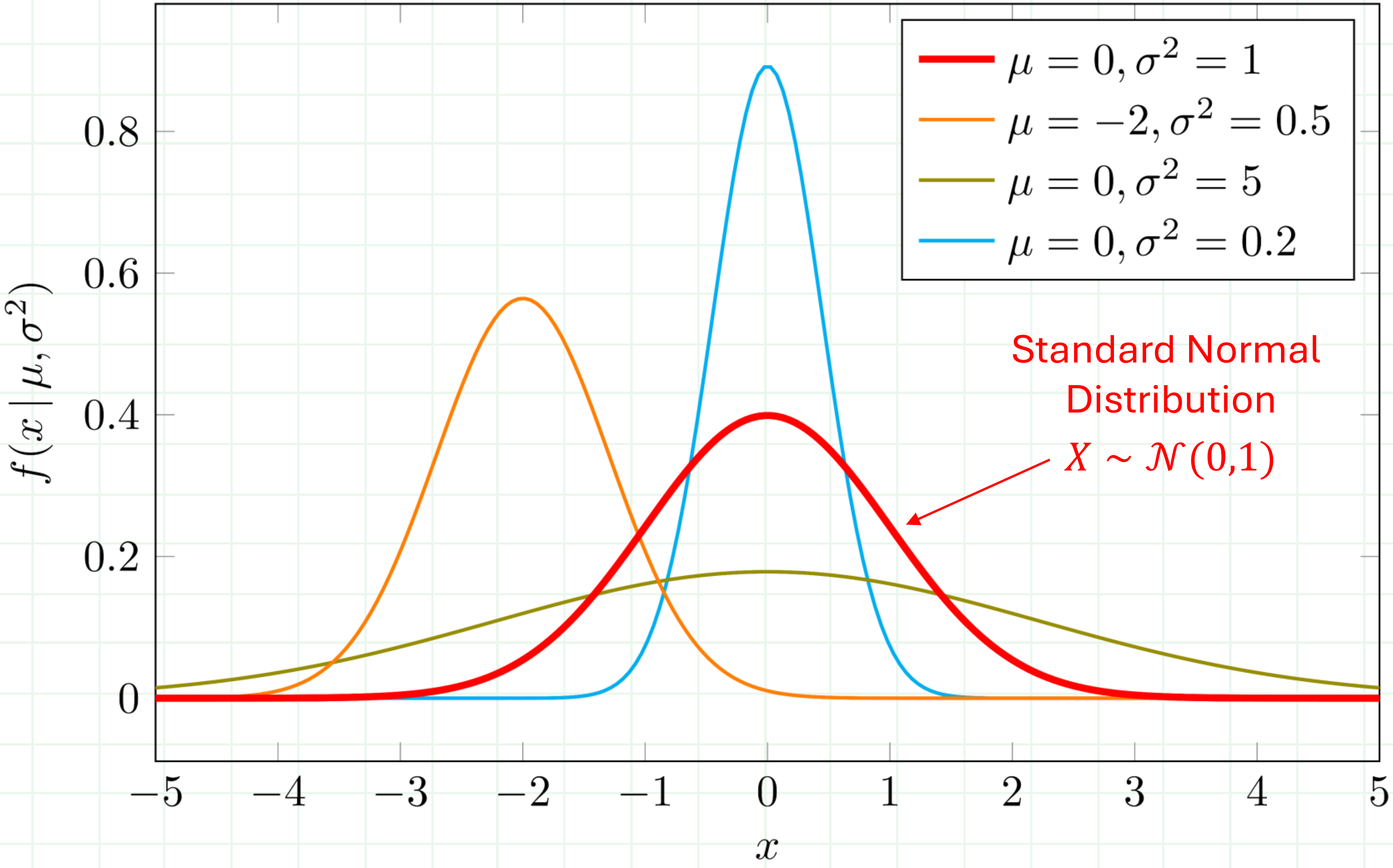
Theorem (Non-Asymptotic) Let $X_1, X_2, \ldots$ be a sequence of i.i.d. random variables with finite expectation. Then

$$\mathbb{P}\left\{\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \mathbb{E}[X]\right| > \epsilon\right\} = \frac{\mathrm{Var}[X]}{n\epsilon^2}$$

Gaussian Random Matrices

- Martinsson, P.-G.(2020). **Randomized methods in linear algebra and their applications in data science**.
  URL: http://users.oden.utexas.edu/~pgm/2020_kth_course/
- Vershynin, Roman (2018). **High-Dimensional Probability.** Cambridge University Press

Gaussian Distribution (the Bell Curve)



$$X \sim \mathcal{N}(0, \sigma^2)$$
$$X = \sigma Y$$
where $Y \sim \mathcal{N}(0, 1)$

Gaussian Random Variables

Fact If X and Y are independent random variables, then

$$\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$$

Proposition Let X be a real random variable such that $\mathbb{E}[X^2] < +\infty$. Then

$$\mathrm{Var}[X] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$$

$X, Y \sim \mathcal{N}(0,1)$ i.i.d.

What is $\mathbb{E}[XY]$?

$$\begin{aligned}\mathbb{E}[XY] &= \mathbb{E}[X]\mathbb{E}[Y] \\ &= (0)(0) \\ &= 0\end{aligned}$$

$X \sim \mathcal{N}(0,1)$

What is $\mathbb{E}[X^2]$?

$$\begin{aligned}\mathbb{E}[X^2] &= \mathrm{Var}[X] + (\mathbb{E}[X])^2 \\ &= 1 + 0^2 \\ &= 1\end{aligned}$$

Uses for Random Gaussian Matrices

Gaussian Random Matrix $\quad \boldsymbol{\Omega} \sim \mathcal{N}(0, \frac{1}{m})^{m \times n} \qquad \omega_{ij} \sim \frac{1}{\sqrt{m}}\mathcal{N}(0,1)$ i.i.d.

- Multiplication by $\boldsymbol{\Omega}$ of vector $\mathbf{x}$ "preserves" $\|\mathbf{x}\|_2$

  Expectation: $\mathbb{E}\|\boldsymbol{\Omega}\mathbf{x}\|_2 = \|\mathbf{x}\|_2$

  Probability: $\mathbb{P}\{\big|\|\boldsymbol{\Omega}\mathbf{x}\|_2 - \|\mathbf{x}\|_2\big| > \epsilon\} \leq \delta$

- Multiplication by $\boldsymbol{\Omega}$ of a set of vectors $\{\, \mathbf{x}_1, \ldots, \mathbf{x}_n \,\}$ "preserves" distances $\|\mathbf{x}_i - \mathbf{x}_j\|_2$ for all i, j
- Multiplication by $\boldsymbol{\Omega}$ of matrix $\mathbf{A}$ "preserves" $\|\mathbf{A}\|_2$ and $\|\mathbf{A}\|_F$
- Right multiplication by $\boldsymbol{\Omega}$ of matrix $\mathbf{A}$ "preserves" range$(\mathbf{A})$
- Left multiplication by $\boldsymbol{\Omega}$ of matrix $\mathbf{A}$ "preserves" range$(\mathbf{A}^\intercal)$

Vector Norm – One Sample

Adapted from Martinsson, 2020

MathSci.ai

Let $\mathbf{v} \in \mathbb{R}^n$ and $\boldsymbol{\omega} \sim \mathcal{N}(0,1)^n$

What is $\mathbb{E}[(\boldsymbol{\omega}^\intercal \mathbf{v})^2]$?

Fact $\mathbb{E}\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} \mathbb{E}[X_i]$

$$\mathbb{E}\left[(\boldsymbol{\omega}^\intercal \mathbf{v})^2\right] = \mathbb{E}\left[\left(\sum_{i=1}^{n} \omega_i v_i\right)^2\right] = \mathbb{E}\left[\sum_{i=1}^{n}\sum_{j=1}^{n} \omega_i \omega_j v_i v_j\right]$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{n} \mathbb{E}[\omega_i \omega_j] v_i v_j$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{n} \delta_{ij} v_i v_j = \sum_{i=1}^{n} v_i^2 = \|\mathbf{v}\|_2^2$$

Vector Norm – m Samples

Let $\mathbf{v} \in \mathbb{R}^n$ and $\boldsymbol{\Omega} \sim \mathcal{N}(0, \frac{1}{m})^{m \times n}$

n = vector length
m = number of dot products averaged

What is $\mathbb{E}[\|\boldsymbol{\Omega}\mathbf{v}\|_2^2]$?

$$\boldsymbol{\Omega} = \frac{1}{\sqrt{m}} \begin{bmatrix} \boldsymbol{\omega}_1^\top \\ \boldsymbol{\omega}_2^\top \\ \vdots \\ \boldsymbol{\omega}_m^\top \end{bmatrix}$$

$$\boldsymbol{\omega}_i \sim \mathcal{N}(0,1)^n$$

$$\begin{aligned} \mathbb{E}\left[\|\boldsymbol{\Omega}\mathbf{v}\|_2^2\right] &= \mathbb{E}\left[\sum_{i=1}^m \frac{1}{m}(\boldsymbol{\omega}_i^\top \mathbf{v})^2\right] \\ &= \frac{1}{m}\sum_{i=1}^m \mathbb{E}\left[(\boldsymbol{\omega}_i^\top \mathbf{v})^2\right] \\ &= \frac{1}{m}\sum_{i=1}^m \|\mathbf{v}\|^2 = \|\mathbf{v}\|^2 \end{aligned}$$

Gordon’s Theorem

Vershynin (2018)

MathSci.ai

Theorem (Gordon’s theorem for Gaussian matrices) Let $\mathbf{\Omega} \sim \mathcal{N}(0, \frac{1}{m})^{m\times n}$ be an $m\times n$ matrix with $m \geq n$ whose entries are independent normal random variables. Then

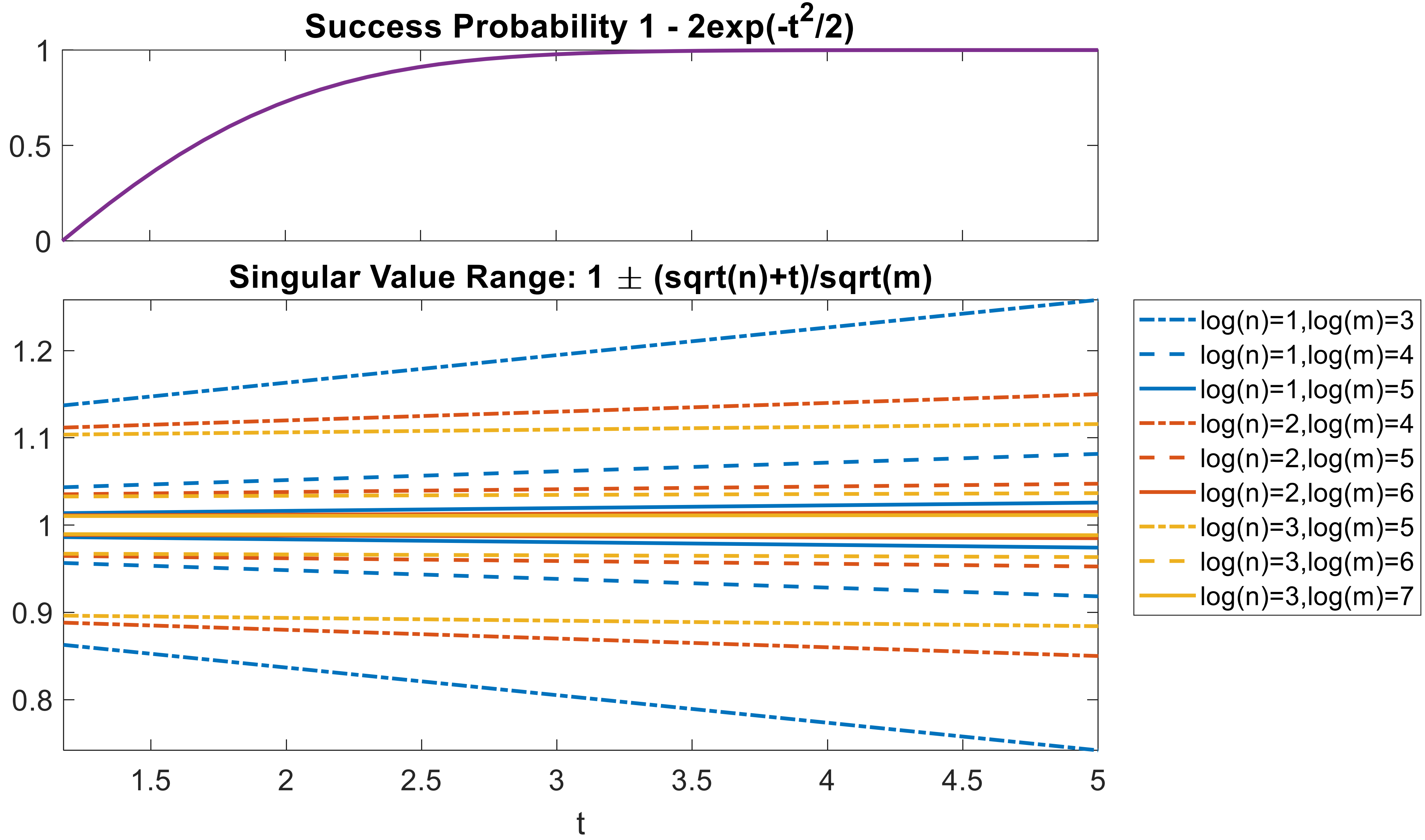
$$1 - \frac{\sqrt{n}}{\sqrt{m}} \leq \mathbb{E}[\sigma_{\min}(\mathbf{\Omega})] \leq \mathbb{E}[\sigma_{\max}(\mathbf{\Omega})] \leq 1 + \frac{\sqrt{n}}{\sqrt{m}}$$

Theorem (Gaussian Matrix Concentration) Let $\mathbf{\Omega} \sim \mathcal{N}(0, \frac{1}{m})^{m\times n}$ with $m \geq n$ whose entries are independent. Then for every $t \geq 0$ we have

$$1 - \frac{\sqrt{n} + t}{\sqrt{m}} \leq \sigma_{\min}(\mathbf{\Omega}) \leq \sigma_{\max}(\mathbf{\Omega}) \leq 1 + \frac{\sqrt{n} + t}{\sqrt{m}}$$

with probability at least $1 - 2\exp(-t^2/2)$

Spectral Norm Error $1 - \frac{\sqrt{n}+t}{\sqrt{m}} \le \sigma_{\min}(\mathbf{\Omega}) \le \sigma_{\max}(\mathbf{\Omega}) \le 1 + \frac{\sqrt{n}+t}{\sqrt{m}}$



Vector Norm Concentration

Theorem (Gaussian Matrix Concentration) Let $\mathbf{\Omega} \sim \mathcal{N}(0, \frac{1}{m})^{m \times n}$ with $m \geq n$ whose entries are independent. Then for every $t \geq 0$ we have

$$1 - \frac{\sqrt{n} + t}{\sqrt{m}} \leq \sigma_{\min}(\mathbf{\Omega}) \leq \sigma_{\max}(\mathbf{\Omega}) \leq 1 + \frac{\sqrt{n} + t}{\sqrt{m}}.$$

with probability at least $1 - 2\exp(-t^2/2)$.

n = vector length
m = number of dot products averaged

Let $\mathbf{v} \in \mathbb{R}^n$ and $\mathbf{\Omega} \sim \mathcal{N}(0, \frac{1}{m})$

$$\sigma_{\min}(\mathbf{\Omega})\|\mathbf{v}\|_2 \leq \|\mathbf{\Omega v}\|_2 \leq \sigma_{\max}(\mathbf{\Omega})\|\mathbf{v}\|_2$$

For $\epsilon = \frac{\sqrt{n}+t}{\sqrt{m}}$ and $\delta = 2\exp(-t^2/2)$, we have

$$\mathbb{P}\left\{(1-\epsilon)\|\mathbf{v}\|_2 \leq \|\mathbf{\Omega v}\|_2 \leq (1+\epsilon)\|\mathbf{v}\|_2\right\} > 1 - \delta$$

Randomized Range Finder

- Halko, N., P.-G. Martinsson, and J. A. Tropp (2011). **Finding Structure with Randomness: Probabilistic Algorithms for Constructing Approximate Matrix Decompositions**, *SIAM Review,* doi:10.1137/090771806
- Erichson, N. B., S. Voronin, S. L. Brunton, and J. N. Kutz (2016). **Randomized Matrix Decompositions using R**. doi: 10.18637/jss.v089.i11.
- Martinsson, P.-G.(2020). **Randomized methods in linear algebra and their applications in data science**.
  URL: http://users.oden.utexas.edu/~pgm/2020_kth_course/

Randomized Range Finder (RRF) Goal

MathSci.ai

Given:

$$\mathbf{A} \in \mathbb{R}^{m \times n} \quad \text{with} \quad \text{“rank”}(\mathbf{A}) < k < n \leq m$$

Find:

$$\text{orthonormal} \quad \mathbf{Q} \in \mathbb{R}^{m \times k} \quad \text{such that} \quad \text{range}(\mathbf{Q}) \approx \text{range}(\mathbf{A})$$

Applications:

- Singular Value Decomposition
- Principal Component Analysis
- Low-Rank Matrix Factorization
- Reduced-order Modeling, a.k.a. Model Reduction
- Manifold Learning

Orthogonal Projectors

Orthonormal Matrix: $\mathbf{Q}$

$$\mathbf{Q} \in \mathbb{R}^{m \times n} \text{ with } m > n \text{ and } \mathbf{Q}^\intercal \mathbf{Q} = \mathbf{I}_n$$

Orthogonal Projector: $\mathbf{Q}\mathbf{Q}^\mathrm{T}$

$$\mathbf{Q}\mathbf{Q}^\intercal \mathbf{x} \in \mathrm{range}(\mathbf{Q}) \quad \text{and} \quad \|\mathbf{Q}^\intercal \mathbf{x}\|_2 = \|\mathbf{Q}\mathbf{Q}^\intercal \mathbf{x}\|_2 \leq \|\mathbf{x}\|_2$$

$$\mathbf{x} \in \mathrm{range}(\mathbf{Q}) \Rightarrow \mathbf{Q}\mathbf{Q}^\intercal \mathbf{x} = \mathbf{x}$$

$$\mathbf{x} \in \mathrm{nullspace}(\mathbf{Q}) \Rightarrow \mathbf{Q}\mathbf{Q}^\intercal \mathbf{x} = \mathbf{0}$$

Norm Decomposition

$$\|\mathbf{x}\|_2^2 = \|\mathbf{Q}\mathbf{Q}^\intercal \mathbf{x}\|_2^2 + \|(\mathbf{I}_m - \mathbf{Q}\mathbf{Q}^\intercal)\mathbf{x}\|_2^2$$

Orthogonal Projectors and RRF

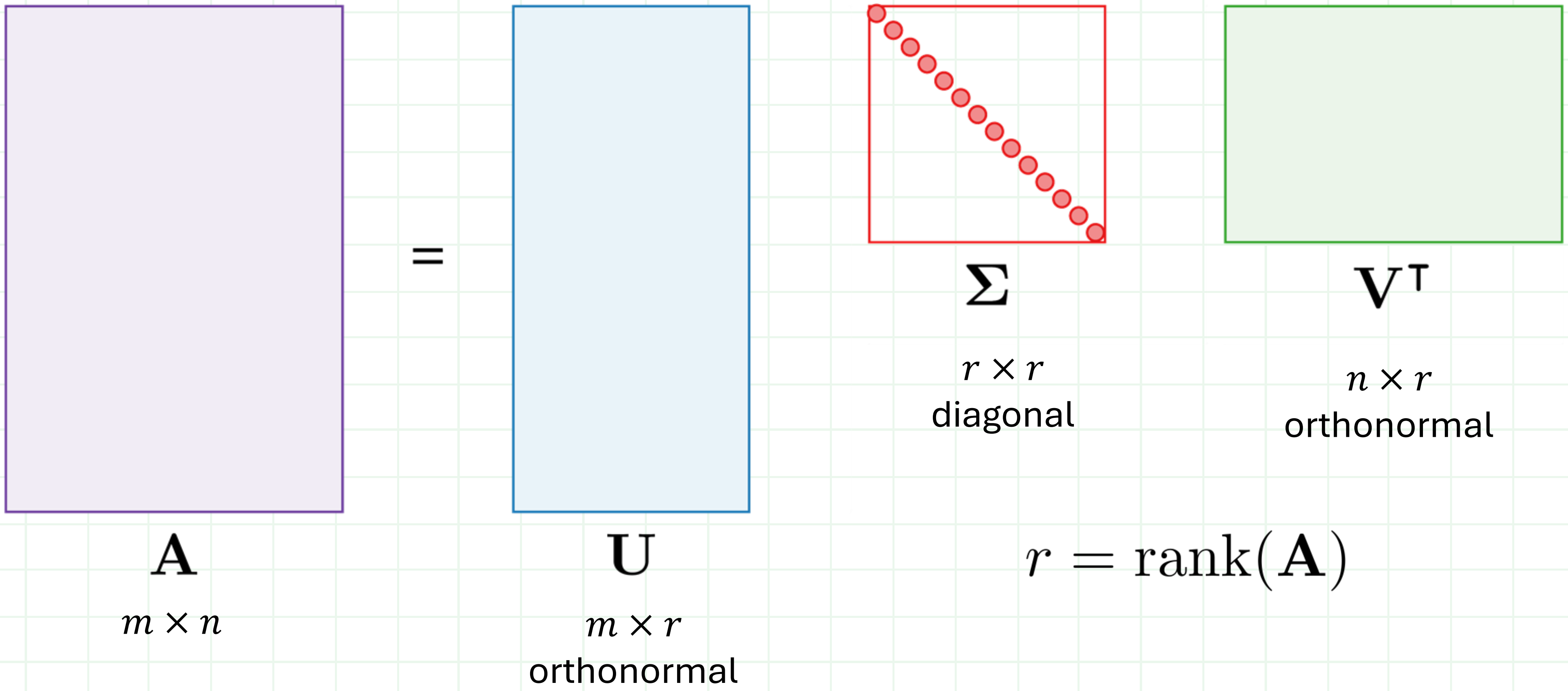
MathSci.ai

Given: $\mathbf{A} \in \mathbb{R}^{m \times n}$ with $m > n$

Find: $\mathbf{Q} \in \mathbb{R}^{m \times k}$ such that $\text{range}(\mathbf{Q}) \approx \text{range}(\mathbf{A})$

$$\|\mathbf{A}\|_{\mathrm{F}}^2 = \|\underbrace{\mathbf{Q}\mathbf{Q}^\intercal\mathbf{A}}_{\approx \mathbf{A}}\|_{\mathrm{F}}^2 + \underbrace{\|(\mathbf{I} - \mathbf{Q}\mathbf{Q}^\intercal)\mathbf{A}\|_{\mathrm{F}}^2}_{\text{(Squared) Subspace Error}}$$

(Compact) Singular Value Decomposition

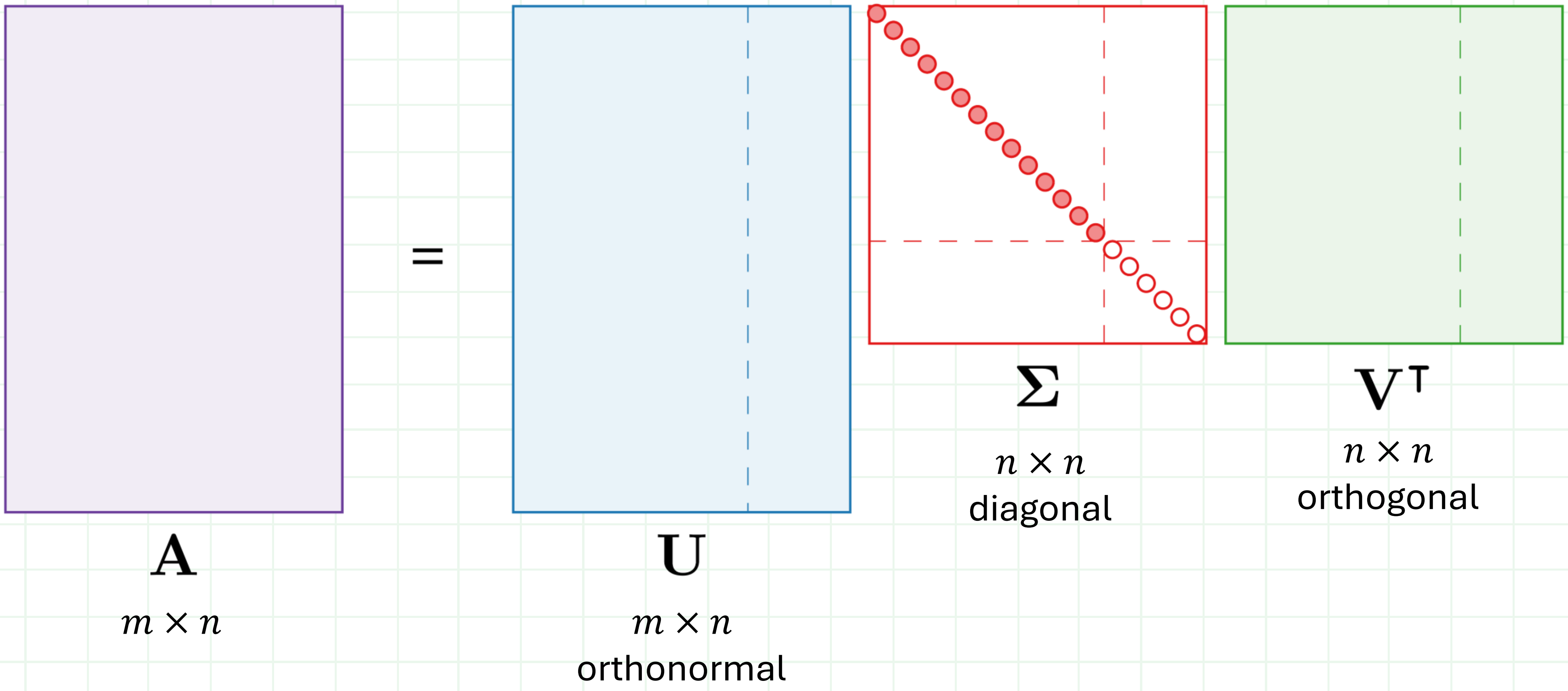


$$r = \operatorname{rank}(\mathbf{A})$$

Connections to Norms:

$$\|\mathbf{A}\|_{\mathrm{F}} = \sqrt{\sum_{i=1}^{r} \sigma_i^2(\mathbf{A})} \qquad \|\mathbf{A}\|_2 = \sigma_1(\mathbf{A})$$

(Economy) Singular Value Decomposition



Economy SVD

```
[U,S,V] = svd(A,"econ");
```

Eckart-Young Theorem, 1936

Let $\mathbf{A} \in \mathbb{R}^{m \times n}$ with $r = \operatorname{rank}(\mathbf{A})$

Denote its compact SVD by $\mathbf{A} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^\intercal = \sum_{i=1}^{r} \sigma_i \mathbf{u}_i \mathbf{v}_i^\intercal$

For $k < r$, the best rank-k approximation to $\mathbf{A}$ is the rank-k truncated SVD

$$\mathbf{A}_k = \mathbf{U}_k \boldsymbol{\Sigma}_k \mathbf{V}_k^\intercal = \mathbf{U}_k \mathbf{U}_k^\intercal \mathbf{A} = \sum_{j=1}^{k} \sigma_j \mathbf{u}_j \mathbf{v}_j^\intercal$$

In this case,

$$\left\|\mathbf{A} - \mathbf{A}_k\right\|_\mathrm{F} = \left(\sum_{j>k} \sigma_j^2\right)^{\frac{1}{2}} \quad \text{and} \quad \left\|\mathbf{A} - \mathbf{A}_k\right\|_2 = \sigma_{k+1}$$

Optimal Subspace Error

Optimal Rank-k Approximation: $\mathbf{A}_k = \mathbf{U}_k \mathbf{\Sigma}_k \mathbf{V}_k^\intercal$

Optimal k-Dimension Subspace: $\mathbf{Q}^* = \mathbf{U}_k$

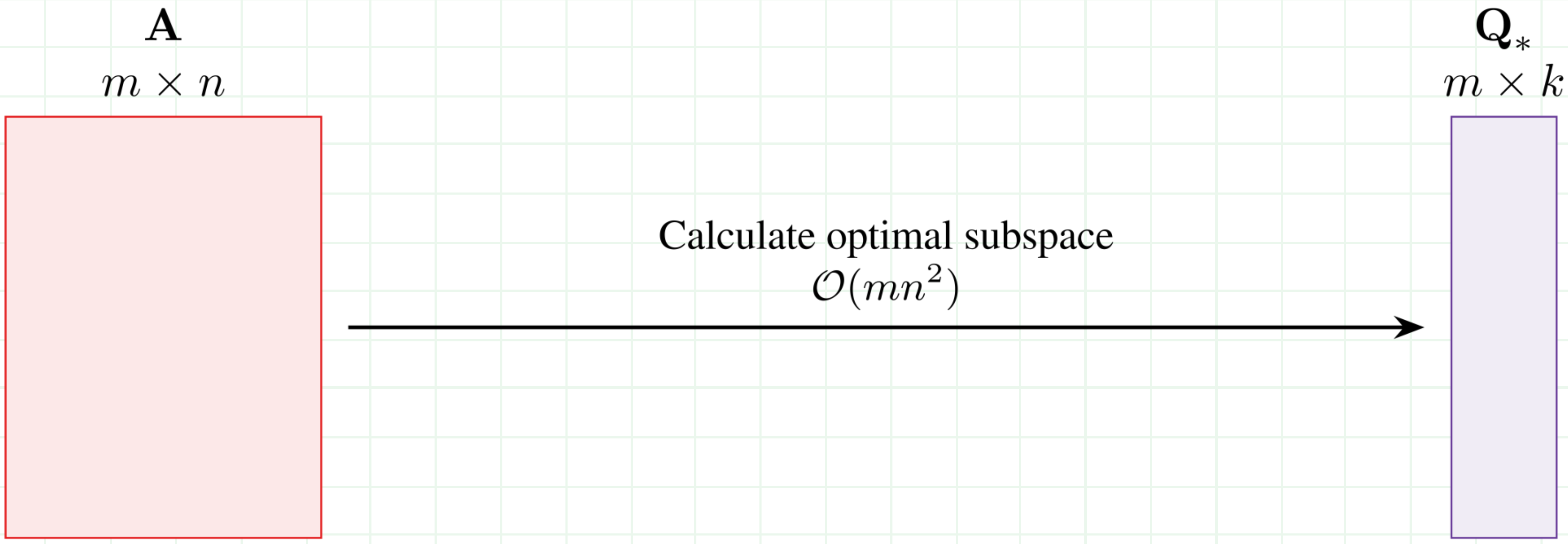
Frobenious Norm

$$\begin{aligned}\min_{\mathbf{Q}\in\mathbb{O}^{m\times k}} & \|(\mathbf{I}-\mathbf{Q}\mathbf{Q}^\intercal)\mathbf{A}\|_\mathrm{F} \\ &= \|(\mathbf{I}-\mathbf{U}_k\mathbf{U}_k^\intercal)\mathbf{A}\|_\mathrm{F} \\ &= \|\mathbf{A}-\mathbf{U}_k\mathbf{U}_k^\intercal\mathbf{A}\|_\mathrm{F} \\ &= \|\mathbf{A}-\mathbf{A}_k\|_\mathrm{F} \\ &= \left(\sum_{i=k+1}^{r}\sigma_i^2\right)^{\frac{1}{2}}\end{aligned}$$

2-Norm

$$\begin{aligned}\min_{\mathbf{Q}\in\mathbb{O}^{m\times k}} & \|(\mathbf{I}-\mathbf{Q}\mathbf{Q}^\intercal)\mathbf{A}\|_2 \\ &= \|(\mathbf{I}-\mathbf{U}_k\mathbf{U}_k^\intercal)\mathbf{A}\|_2 \\ &= \|\mathbf{A}-\mathbf{U}_k\mathbf{U}_k^\intercal\mathbf{A}\|_2 \\ &= \|\mathbf{A}-\mathbf{A}_k\|_2 \\ &= \sigma_{k+1}\end{aligned}$$

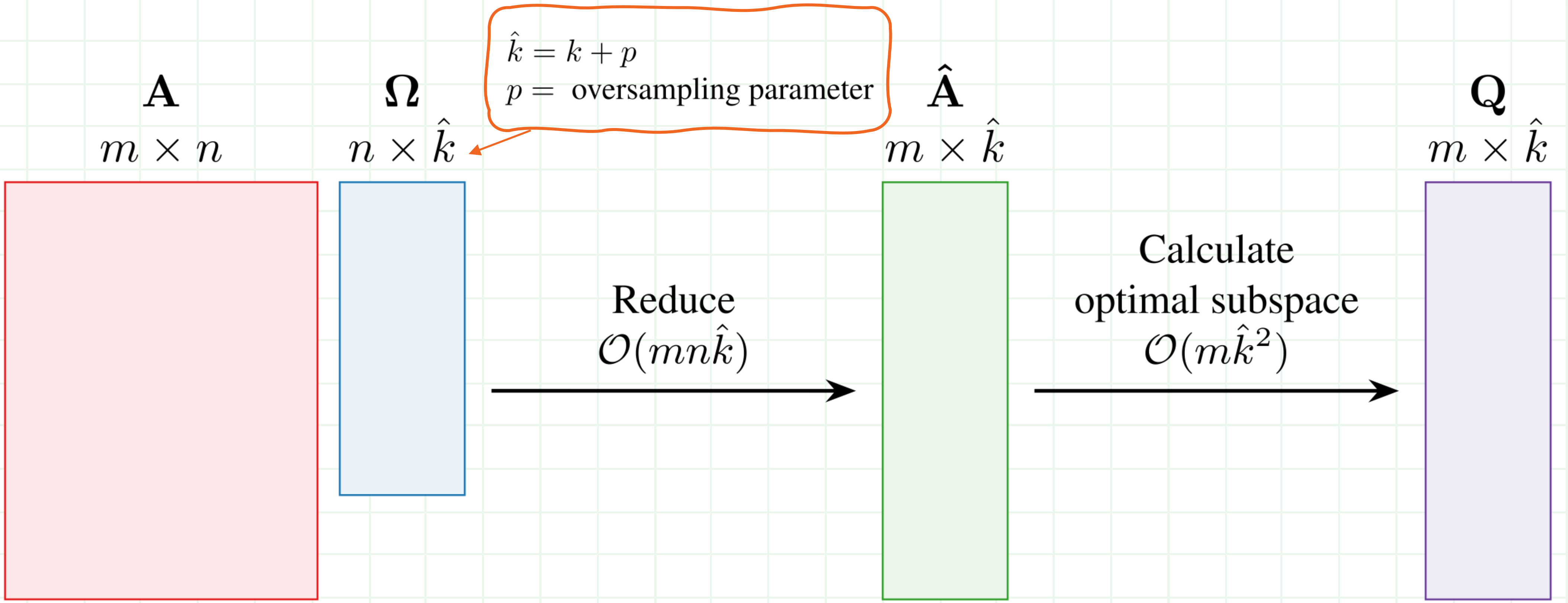
Calculating Optimal Subspace



MATLAB Code

```
[U,~,~] = svd(A,"econ");
Qstar = U(:,1:k);
```

Randomized Range Finder



RRF MATLAB Code

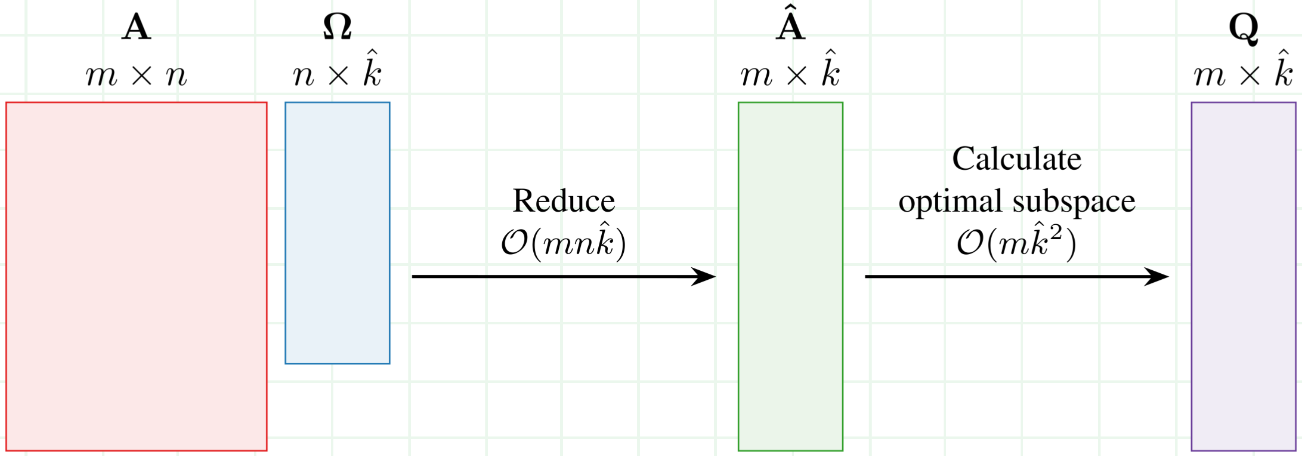
```
function Q = rrf(A,k,p)
Omega = rand(size(A,2),k+p);
Ahat = A * Omega;
[Q,~,~] = svd(Ahat,"econ");
end
```

Alternative to SVD

```
[Q,~,~] = qr(Ahat,"econ");
```

Average RRF Error Bounds

Halko, Martinsson, Tropp, 2011



$\hat{k} = k + p$

$p =$ oversampling parameter

Accuracy depends on decay of singular values!

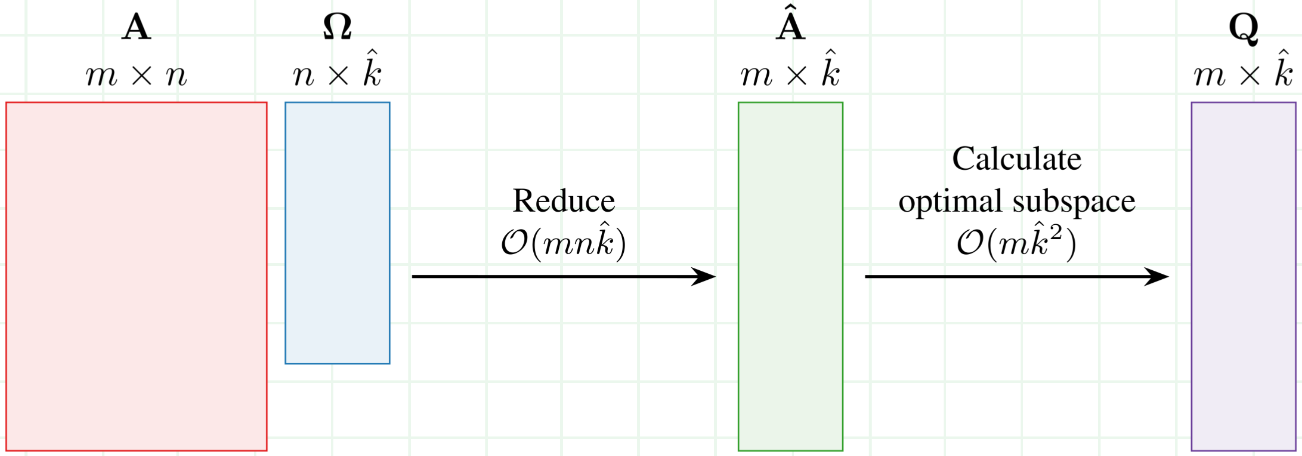
$$\mathbb{E}\left[\|(\mathbf{I} - \mathbf{Q}\mathbf{Q}^\intercal)\mathbf{A}\|_\mathrm{F}\right] \leq \left(1 + \frac{k}{p-1}\right)^{\frac{1}{2}} \underbrace{\left(\sum_{j>k} \sigma_j^2\right)^{\frac{1}{2}}}_{\text{optimal error}}$$

$$\mathbb{E}\left[\|(\mathbf{I} - \mathbf{Q}\mathbf{Q}^\intercal)\mathbf{A}\|_2\right] \leq \left(1 + \sqrt{\frac{k}{p-1}}\right)\sigma_{k+1} + \frac{e\sqrt{k+p}}{p}\left(\sum_{j=k+1}^{r} \sigma_j^2\right)^{\frac{1}{2}}$$

$$\leq \left(1 + \sqrt{\frac{k}{p-1}} + \frac{e\sqrt{k+p}}{p}\sqrt{\min\{m,n\} - k}\right) \underbrace{\sigma_{k+1}}_{\text{optimal error}}$$

Probabilistic RRF Error Bounds

Halko, Martinsson, Tropp, 2011



$\hat{k} = k + p$

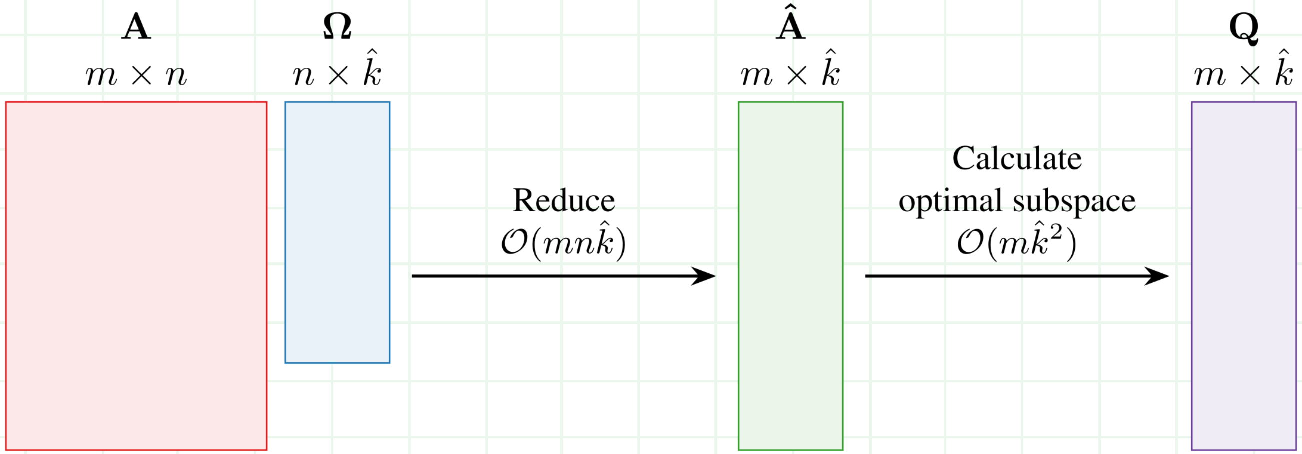
$p =$ oversampling parameter

$$\mathbb{E}\left[\|(\mathbf{I} - \mathbf{Q}\mathbf{Q}^\intercal)\mathbf{A}\|_\mathrm{F}\right] \leq \left(1 + \frac{k}{p-1}\right)^{\frac{1}{2}} \underbrace{\left(\sum_{j>k} \sigma_j^2\right)^{\frac{1}{2}}}_{\text{optimal error}}$$

For $p \geq 4$ and any $u, t \geq 0$:

$$\mathbb{P}\left\{\|(\mathbf{I} - \mathbf{Q}\mathbf{Q}^\intercal)\mathbf{A}\|_\mathrm{F} \leq \left(1 + t\sqrt{\frac{3k}{p-1}}\right)\left(\sum_{j>k} \sigma_j^2\right)^{\frac{1}{2}} + ut\frac{e\sqrt{k+p}}{p+1}\sigma_{k+1}\right\} \geq 1 - \left(2t^{-p} + e^{-\frac{u^2}{2}}\right)$$

RRF Error Bounds (Halko, Martinsson, Tropp, 2011)



$\hat{k} = k + p$

$p =$ oversampling parameter

Accuracy depends on decay of singular values!

$$\mathbb{E}\left[\|(\mathbf{I} - \mathbf{Q}\mathbf{Q}^\mathsf{T})\mathbf{A}\|_\mathrm{F}\right] \leq \left(1 + \frac{k}{p-1}\right)^{\frac{1}{2}} \underbrace{\left(\sum_{j>k} \sigma_j(\mathbf{A})\right)^{\frac{1}{2}}}_{\text{optimal error}}$$

To improve accuracy:
1) Increase p
2) Power iterations

$$\mathbb{E}\left[\|(\mathbf{I} - \mathbf{Q}\mathbf{Q}^\mathsf{T})\mathbf{A}\|_2\right] \leq \left(1 + \sqrt{\frac{k}{p-1}}\right)\sigma_{k+1} + \frac{e\sqrt{k+p}}{p}\left(\sum_{j>k} \sigma_j^2\right)^{\frac{1}{2}}$$

$$\leq \left(1 + \sqrt{\frac{k}{p-1}} + \frac{e\sqrt{k+p}}{p}\sqrt{\min\{m,n\} - k}\right) \underbrace{\sigma_{k+1}(\mathbf{A})}_{\text{optimal error}}$$

RRF with Power Iterations (Improved Error Bnd)

MathSci.ai

$$\mathrm{range}(\mathbf{A}) = \mathrm{range}\big((\mathbf{A}\mathbf{A}^\intercal)^q\mathbf{A}\big) \qquad (\mathbf{A}\mathbf{A}^\intercal)^q\mathbf{A} = \mathbf{U}\mathbf{\Sigma}^{2q+1}\mathbf{V}^\intercal$$

RRF MATLAB Code

```
function Q = rrf(A,k,p,q)
Omega = rand(size(A,2),k+p);
Ahat = A * Omega;
[Q,~,~] = svd(Ahat,"econ");
for i = 1:q
    [Q,~,~] = svd(A'*Q,"econ");
    [Q,~,~] = svd(A*Q,"econ");
end
end
```

$$\begin{aligned}&\mathrm{range}(\mathbf{A}\mathbf{A}^\intercal\mathbf{A}\mathbf{\Omega})\\ &\quad = \mathrm{range}(\mathbf{A}\mathbf{A}^\intercal\mathbf{Q}) \text{ with } \mathbf{Q} \leftarrow \mathrm{range}(\mathbf{A}\mathbf{\Omega})\\ &\quad = \mathrm{range}(\mathbf{A}\mathbf{Q}) \text{ with } \mathbf{Q} \leftarrow \mathrm{range}(\mathbf{A}^\intercal\mathbf{Q})\\ &\quad = \mathrm{range}(\mathbf{Q}) \text{ with } \mathbf{Q} \leftarrow \mathrm{range}(\mathbf{A}\mathbf{Q})\end{aligned}$$

$$\mathbb{E}\Big[\|(\mathbf{I}-\mathbf{Q}\mathbf{Q}^\intercal)\mathbf{A}\|_2\Big] \leq \left(1+\sqrt{\frac{k}{p-1}}+\frac{e\sqrt{k+p}}{p}\sqrt{\min\{m,n\}-k}\right)^{1/(2q+1)} \underbrace{\sigma_{k+1}(\mathbf{A})}_{\text{optimal error}}$$

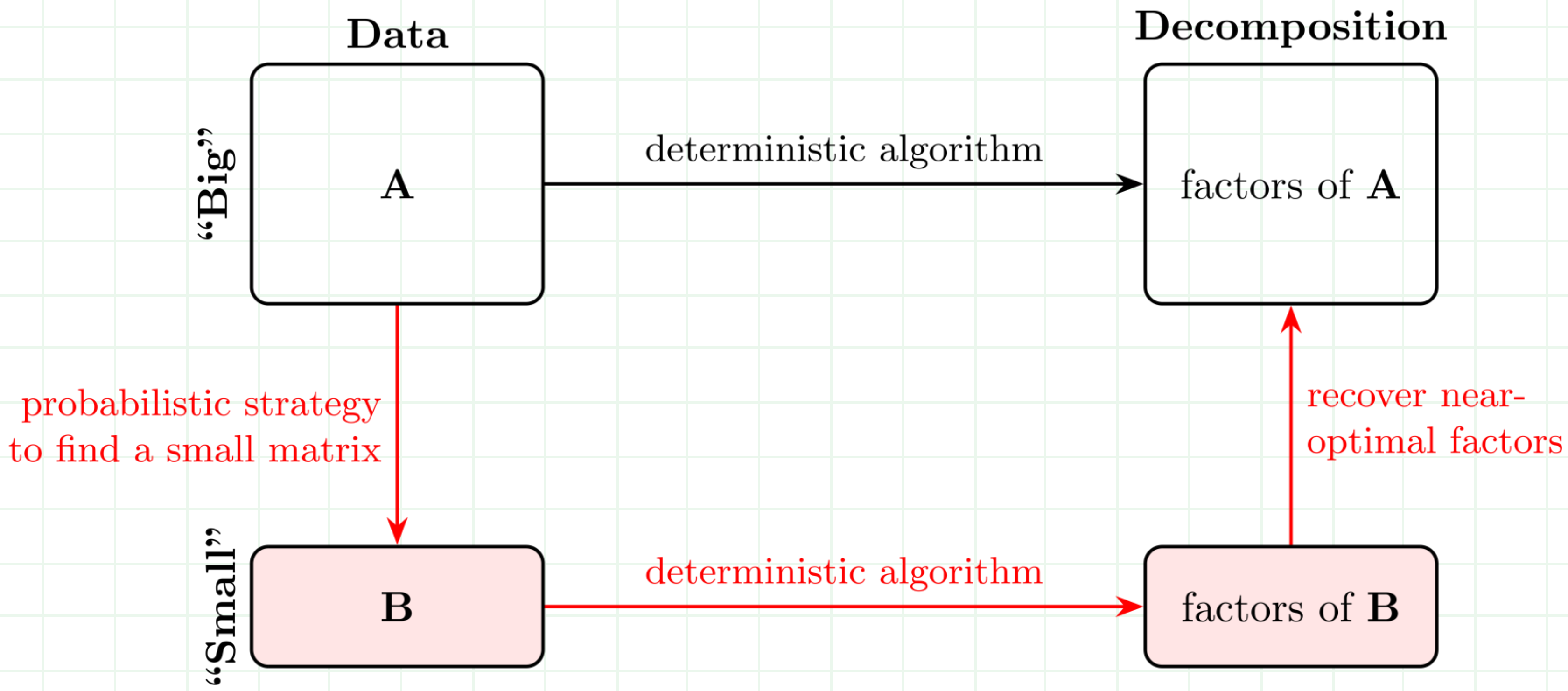
Halko, Martinsson, Tropp, 2011

Randomized SVD

- Halko, N., P.-G. Martinsson, and J. A. Tropp (2011). **Finding Structure with Randomness: Probabilistic Algorithms for Constructing Approximate Matrix Decompositions**, *SIAM Review,* doi:10.1137/090771806
- Brunton, S. L. and J. N. Kutz (2019). **Data Driven Science & Engineering: Machine Learning, Dynamical Systems, and Control.** Cambridge University Press
- Martinsson, P.-G.(2020). **Randomized methods in linear algebra and their applications in data science**. URL: http://users.oden.utexas.edu/~pgm/2020_kth_course/

Application of RRF: Randomized SVD (RSVD)

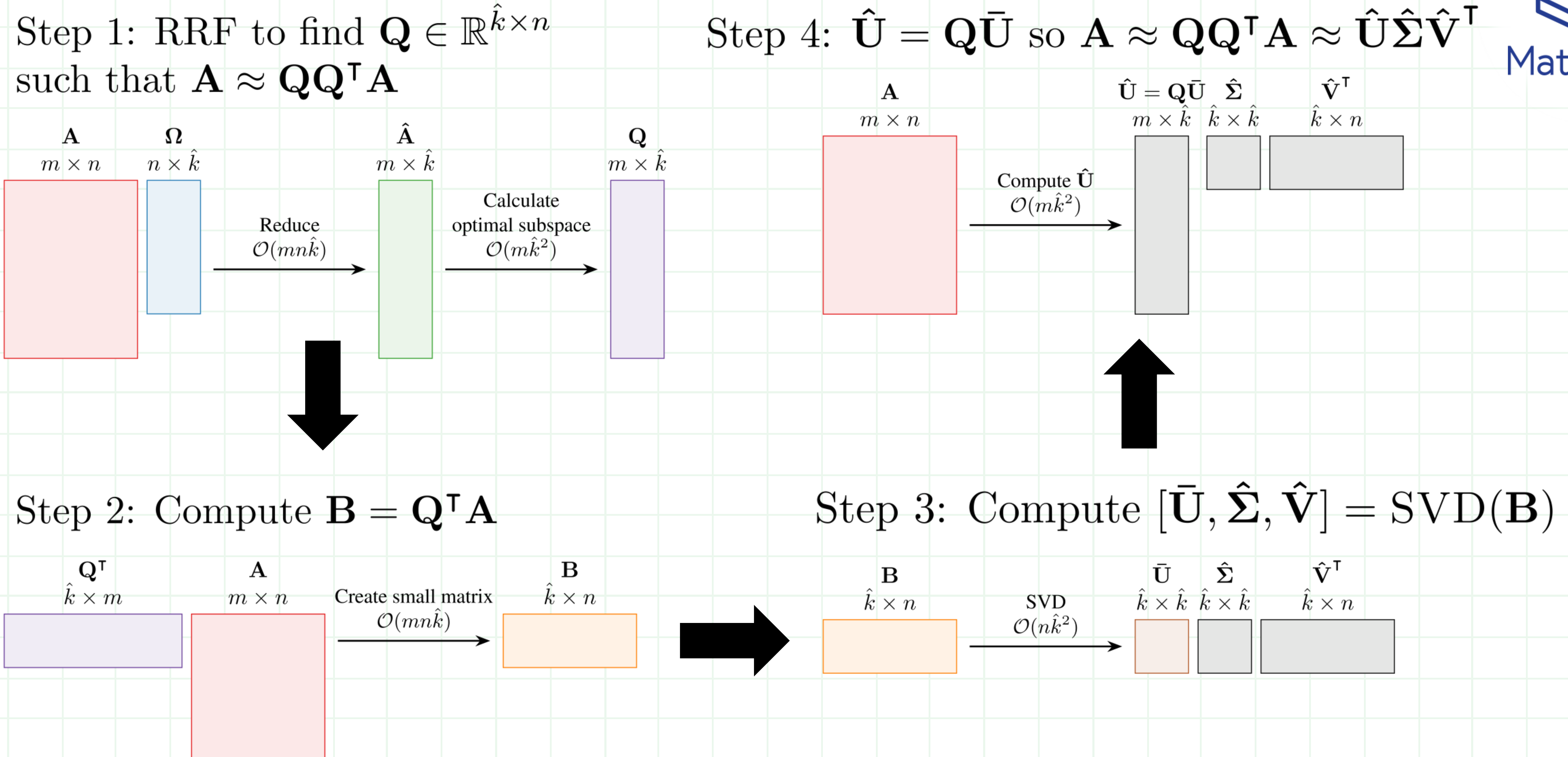
These methods don't work particularly well if the low-rank assumption is not satisfied!



Adopted from Erichson et al. (2016)

RSVD in Pictures

SVD complexity: $\mathcal{O}(mn^2)$



RSVD complexity: $\mathcal{O}(mn\hat{k})$

End of Part 1