Data Scientist vs Statistician

EE16B
Designing Information Devices and Systems II

Lecture 14A

Low Dimensional Data Analysis with SVD

General Procedure for SVD

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

based on A^TA

1: Find eigenvalues of A^TA and order them from biggest to smallest

2: Find orthonormal vectors of A^TA

$$A^TA\vec{v}_i = \lambda_i \vec{v}_i$$

3: Set $\sigma_i = \sqrt{\lambda_i}$ $\quad \vec{u}_i = \frac{1}{\sigma_i} A\vec{v}_i$

based on AA^T

1: Find eigenvalues of AA^T and order them from biggest to smallest

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_r > 0 \cdots 0$$

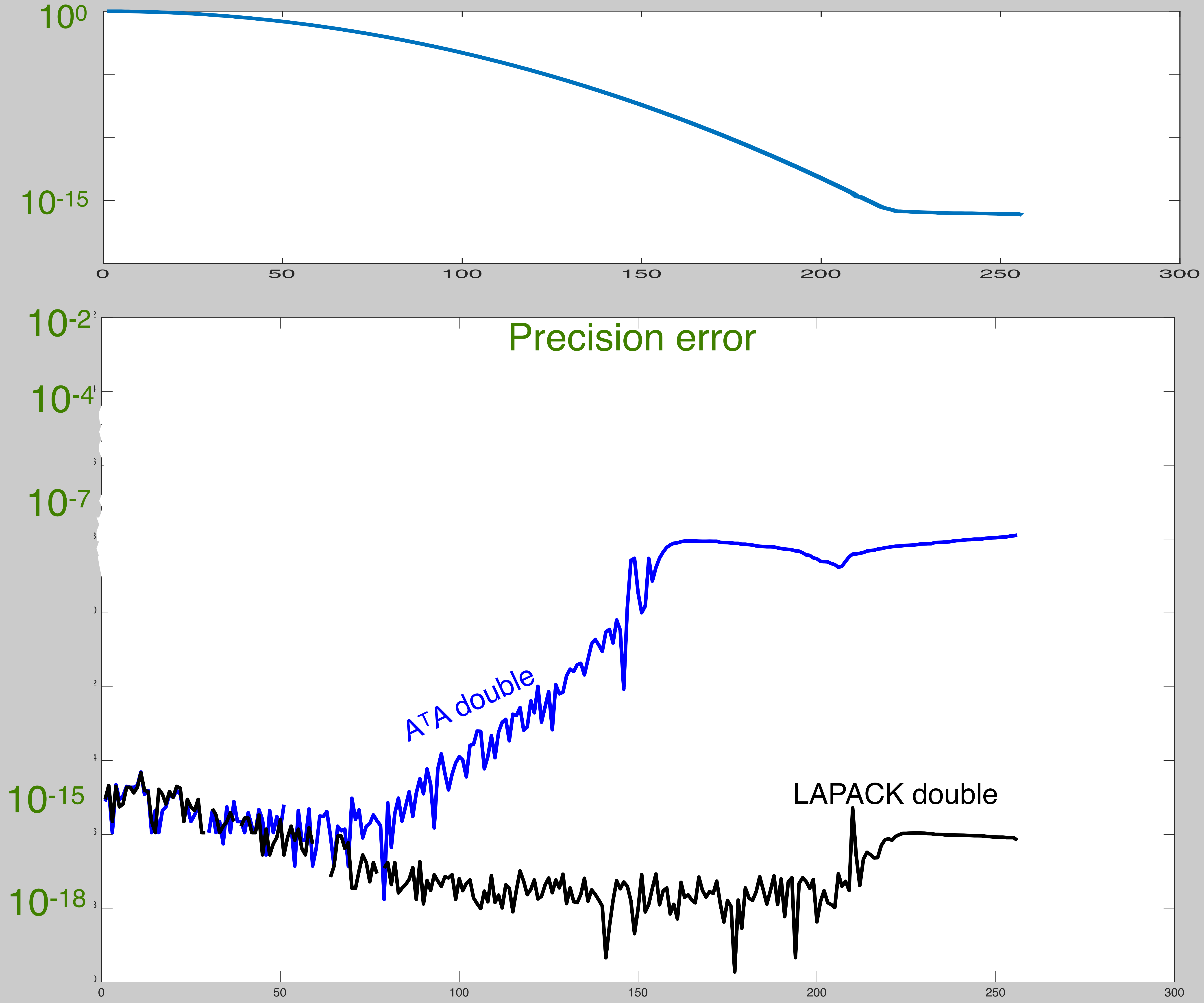
2: Find orthonormal vectors of AA^T

$$AA^T\vec{u}_i = \lambda_i \vec{u}_i$$

3: Set $\sigma_i = \sqrt{\lambda_i}$ $\quad \vec{v}_i = \frac{1}{\sigma_i} A^T\vec{u}_i$

Accuracy with Finite Precision

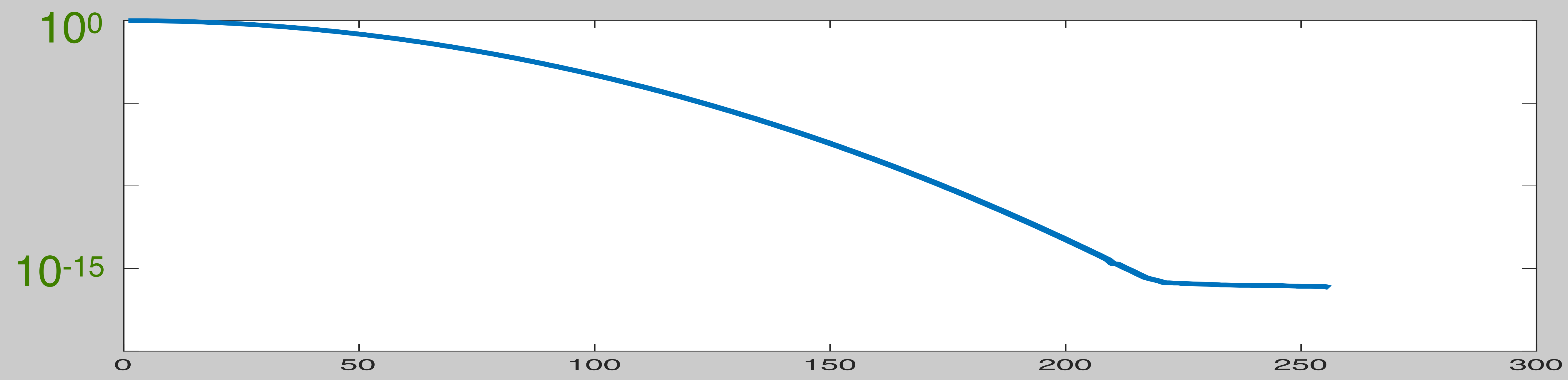
Consider matrix $A \in R^{512x256}$ with the following singular values:

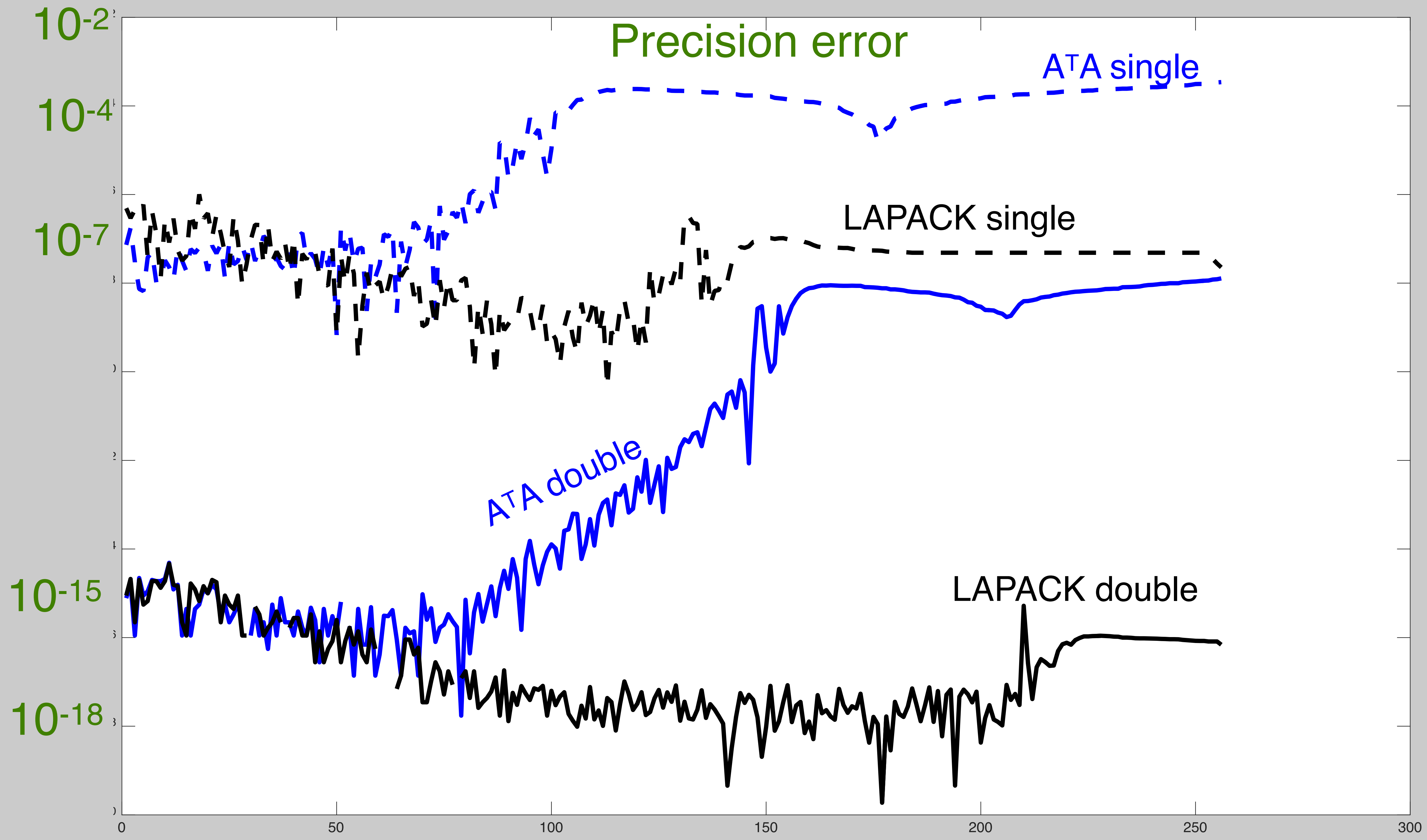


	sign	exponent	mantissa
format	bit	bits	bits
IEEE 32-bit	1	8	23
IEEE 64-bit	1	11	52

Accuracy with Finite Precision

Consider matrix $A \in R^{512x256}$ with the following singular values:

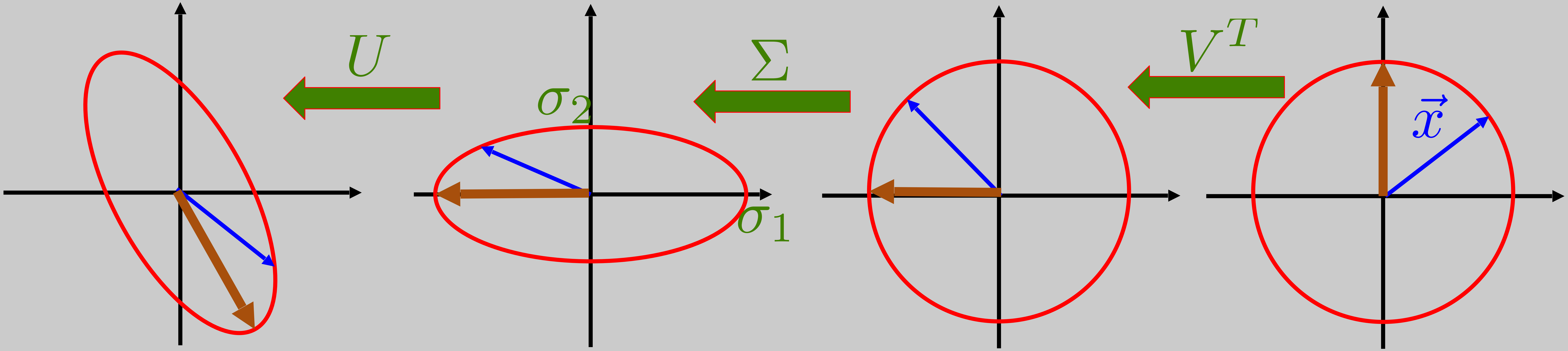




	sign	exponent	mantissa
format	bit	bits	bits
IEEE 32-bit	1	8	23
IEEE 64-bit	1	11	52

Geometric Interpretation

$$A = U \Sigma V^T \qquad A\vec{x}$$



$$||A\vec{x}|| \leq \sigma_1 ||\vec{x}||$$

Q: What vector would amplify the most?

A: Aligns with $\vec{v}_1$

$$U\Sigma V^T$$

Top-Secret (1984)

Rank 1 Matrix

Consider the following matrix:

$$\begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 2 & 2 & 2 & 2 & 2 & 2 \\ 3 & 3 & 3 & 3 & 3 & 3 \end{bmatrix}$$

Rank = 1

We can decompose a rank-1 matrix as an outer product:

$$\vec{u}\vec{v}^T \in \mathrm{R}^{m \times n}$$

$$\vec{u} \in \mathrm{R}^m$$
$$\vec{v} \in \mathrm{R}^n$$

$$\underset{m \times 1}{\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}} \underset{1 \times n}{\begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}}$$

SVD

SVD decomposes a rank r matrix $A \in \mathrm{R}^{m \times n}$
into a sum of r rank-1 matrices:

$$A = \sigma_1 \vec{u}_1 \vec{v}_1^T + \sigma_2 \vec{u}_2 \vec{v}_2^T + \cdots + \sigma_r \vec{u}_r \vec{v}_r^T$$

1) $\vec{u}_i^T \vec{u}_j = \begin{cases} 0 & i \neq j \\ 1 & i = j \end{cases} \Rightarrow ||\vec{u}_i|| = 1 \quad \vec{u}_i \perp \vec{u}_j$

2) $\vec{v}_i^T \vec{v}_j = \begin{cases} 0 & i \neq j \\ 1 & i = j \end{cases} \Rightarrow ||\vec{v}_i|| = 1 \quad \vec{v}_i \perp \vec{v}_j$

3) $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$

SVD

$$A = \sigma_1 \vec{u}_1 \vec{v}_1^T + \sigma_2 \vec{u}_2 \vec{v}_2^T + \cdots + \sigma_r \vec{u}_r \vec{v}_r^T$$

$m \times n$ $\quad m + n \quad m + n \quad m + n$

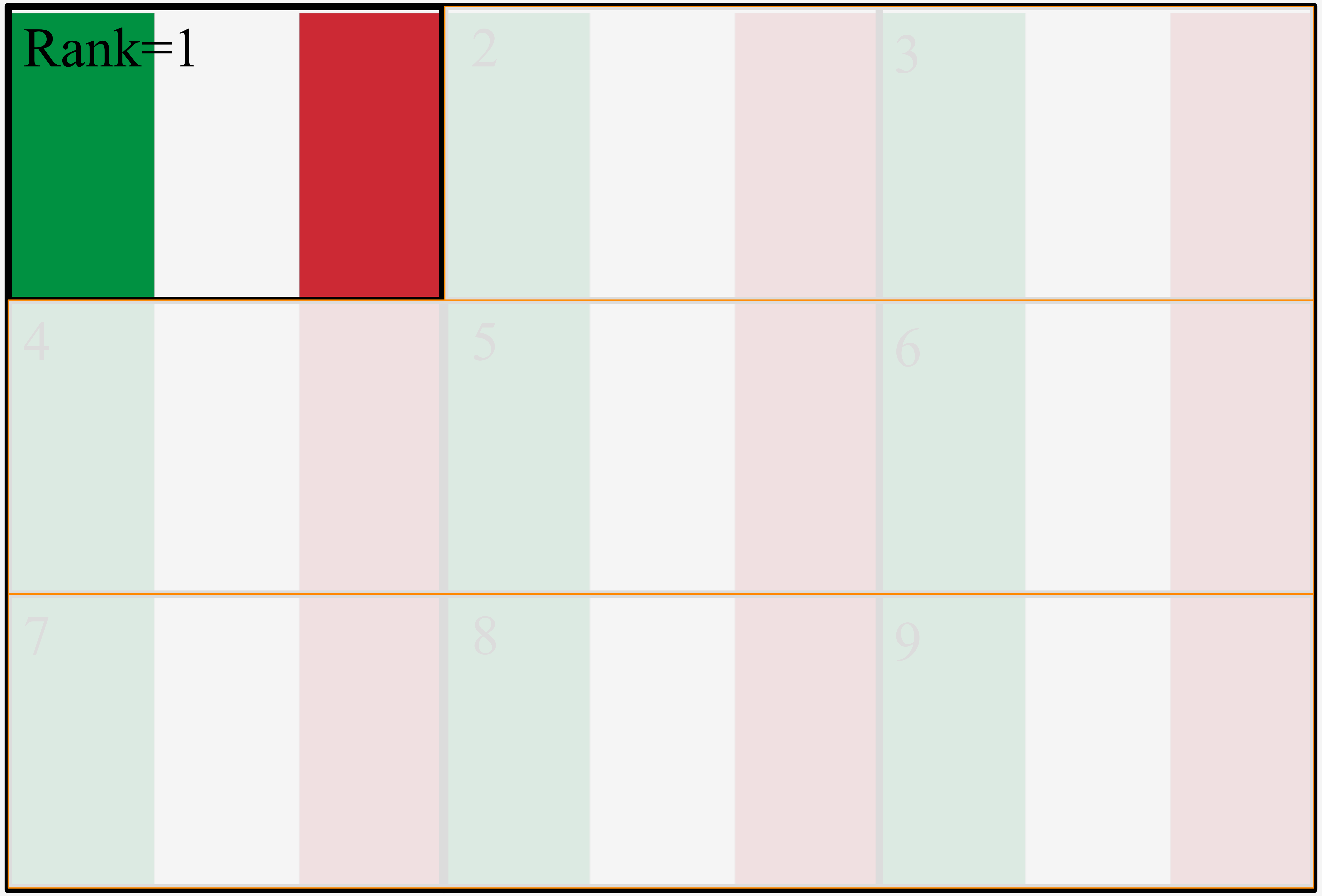
$r(m+n) \leq mn$ If m,n are large and r is small

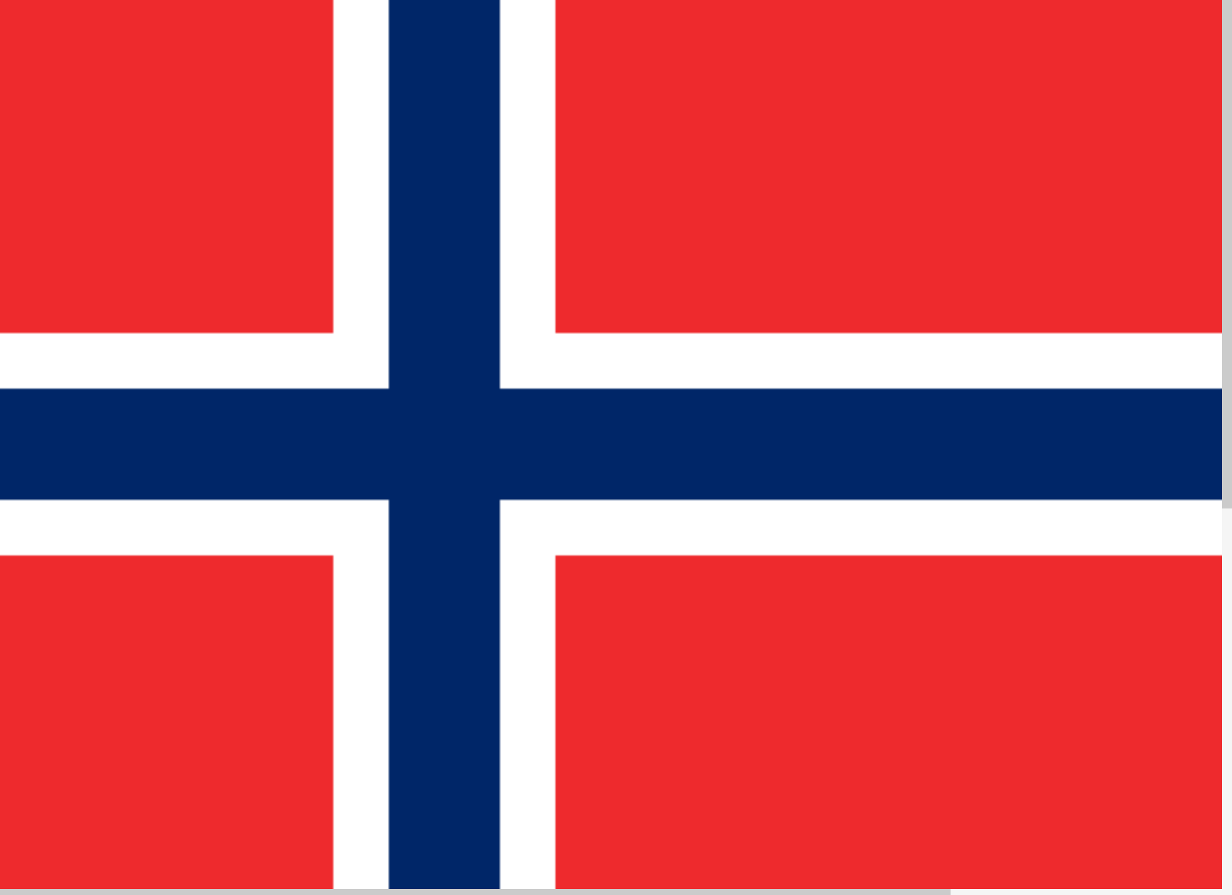
Typically, $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_{\hat{r}} >> \sigma_{\hat{r}+1} \geq \cdots \geq \sigma_r$

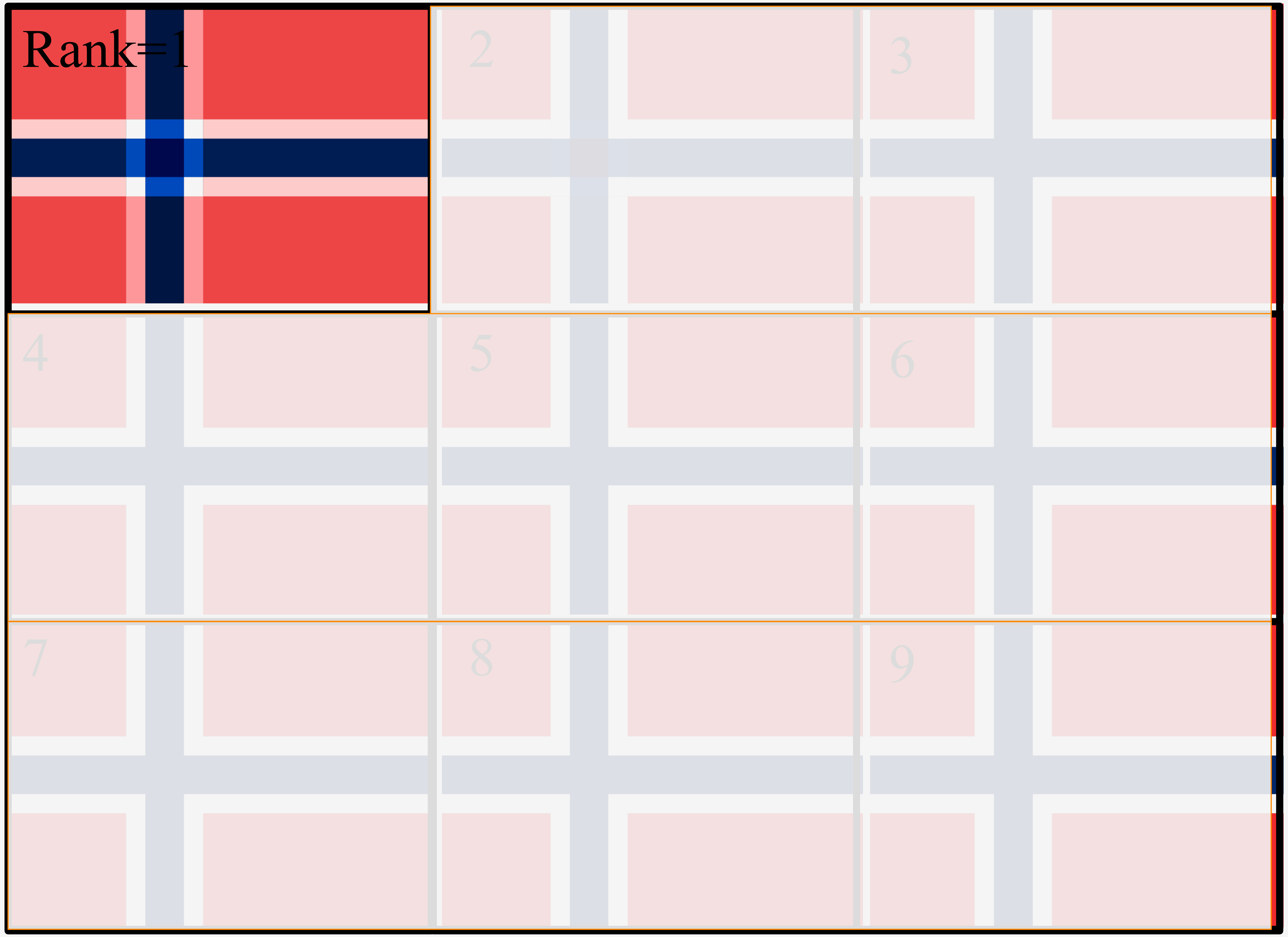
10 8 5 0.1 0.001

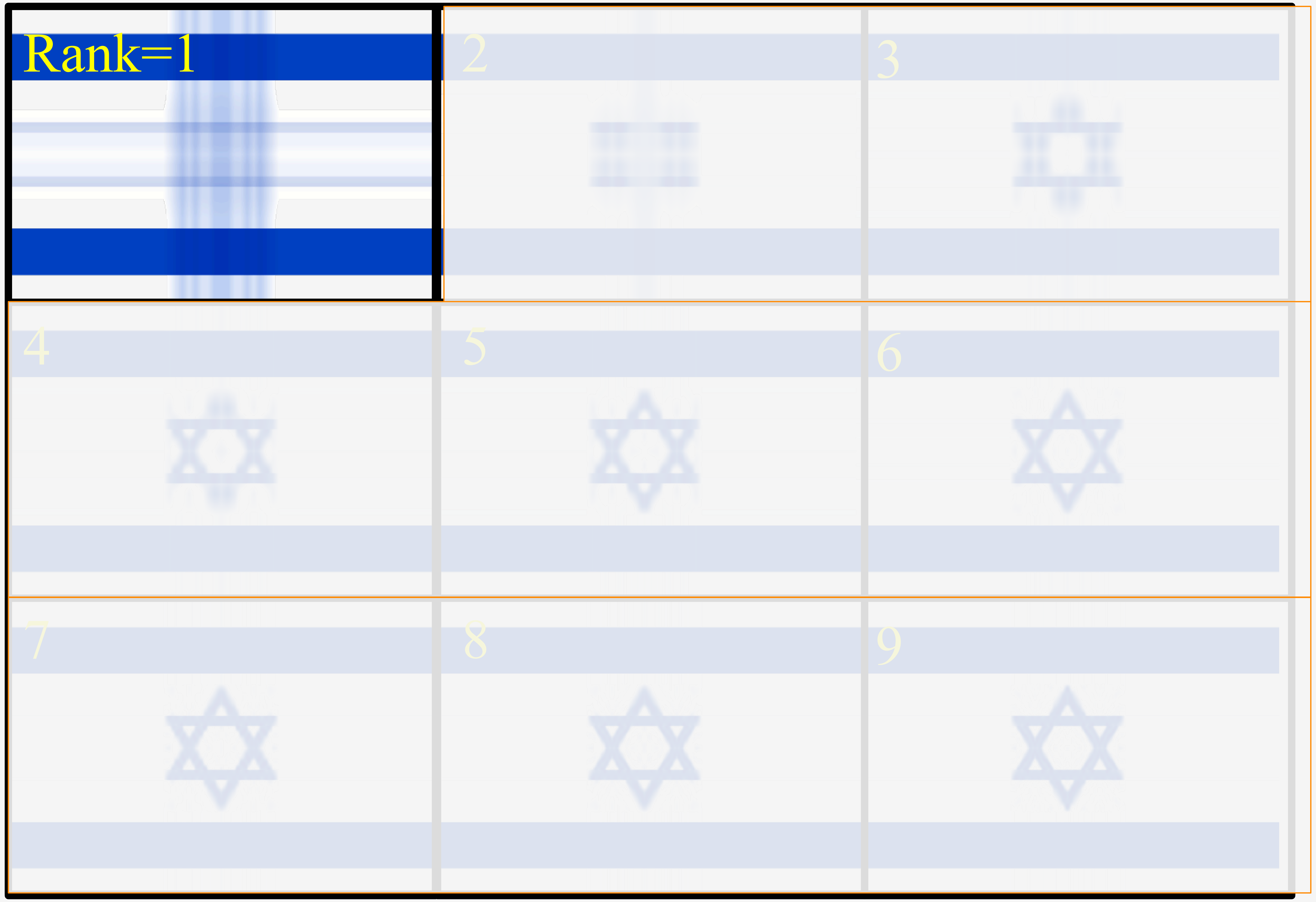
$$\begin{bmatrix} 1.02 & 0.99 & 0.98 & 1.03 & 1.01 & 1 \\ 2 & 1.98 & 2.01 & 2.03 & 1.99 & 1.97 \\ 3.01 & 2.98 & 3 & 2.99 & 3.03 & 3.02 \end{bmatrix}$$

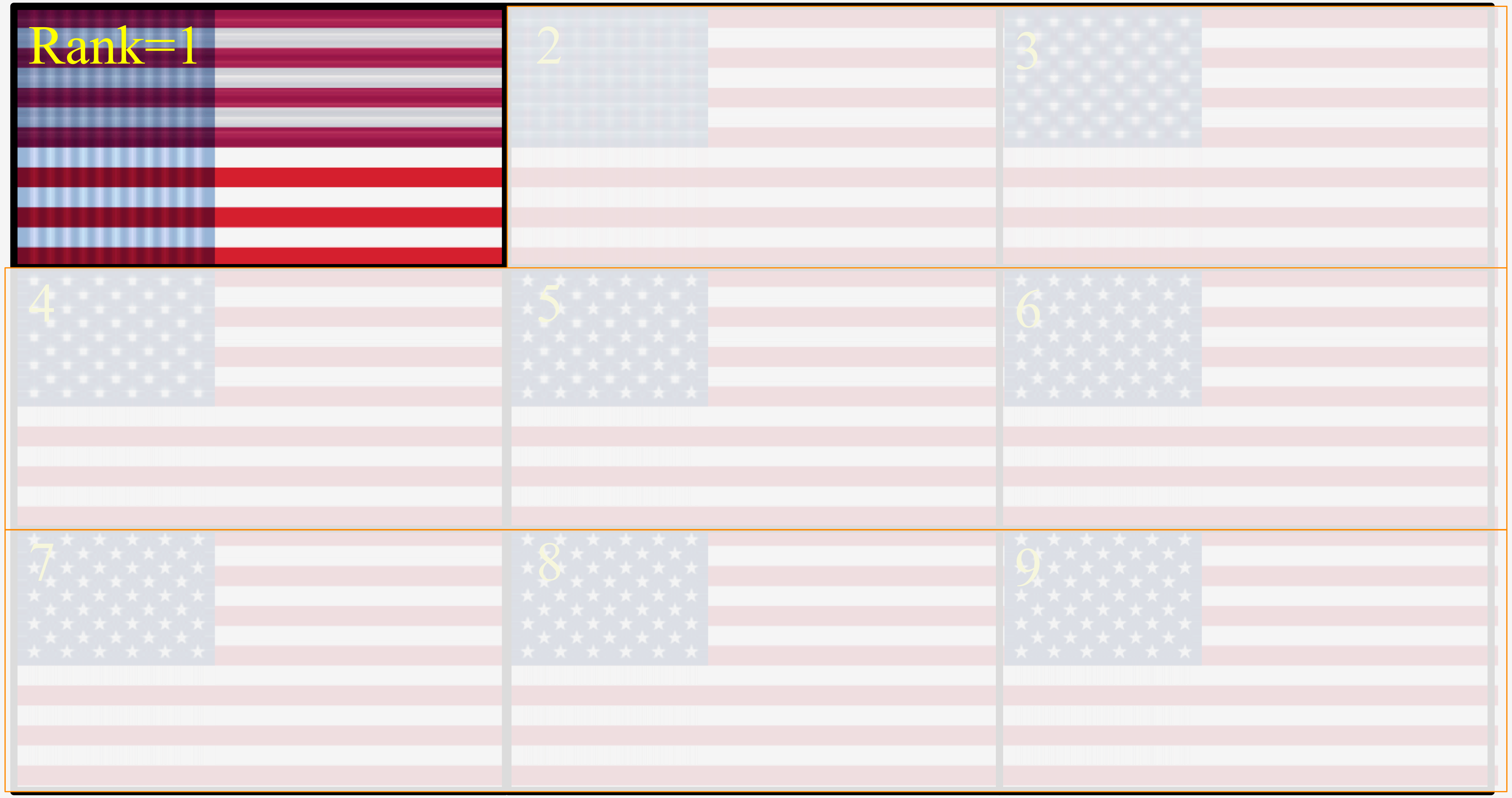
$$A \approx \sigma_1 \vec{u}_1 \vec{v}_1^T + \sigma_2 \vec{u}_1 \vec{v}_1^T + \cdots + \sigma_{\hat{r}} \vec{u}_{\hat{r}} \vec{v}_{\hat{r}}^T$$

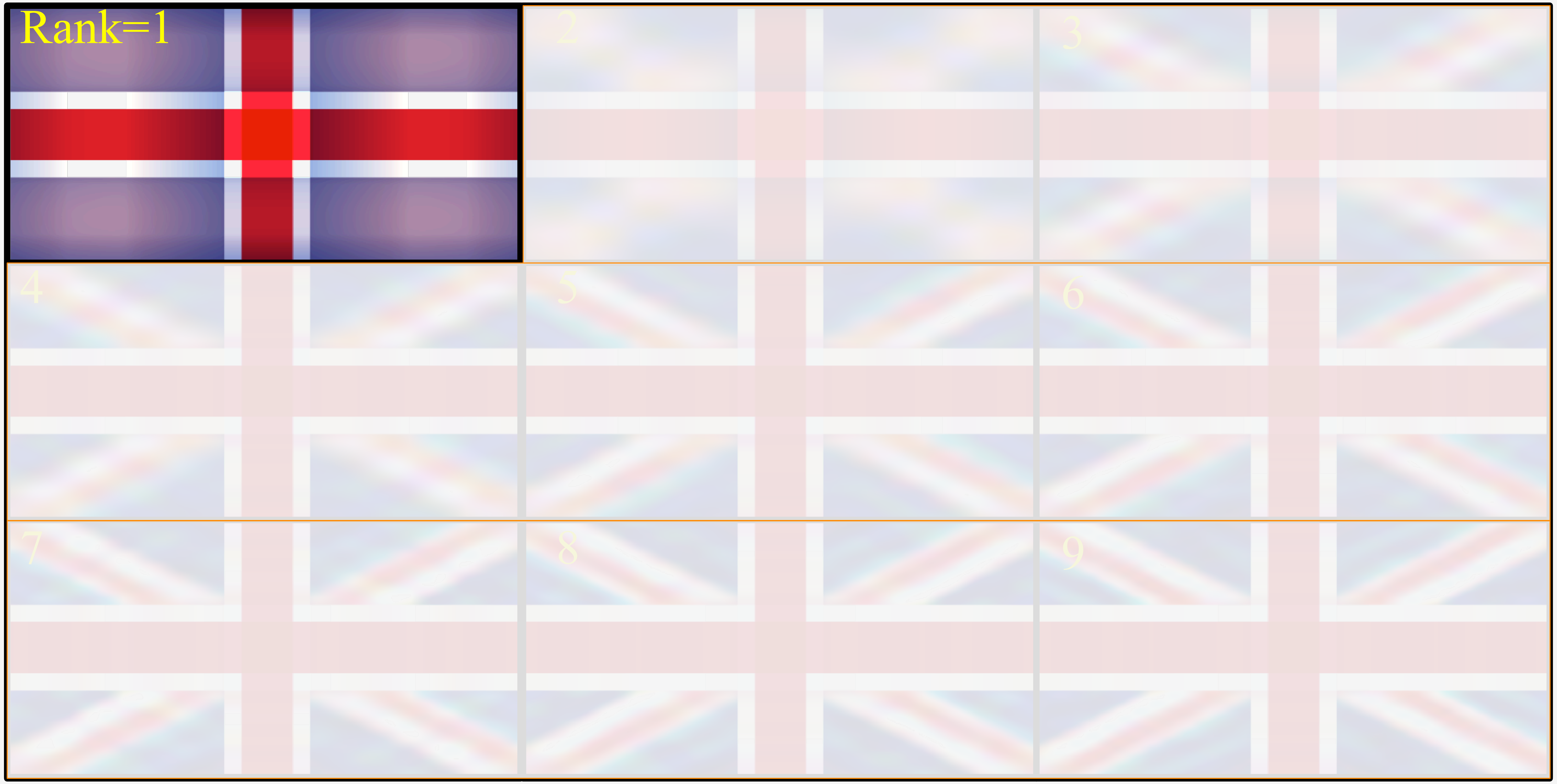

Rank=1
2
3
4
5
6
7
8
9




Rank=1
2
3
4
5
6
7
8
9


Rank=1
2
3
4
5
6
7
8
9

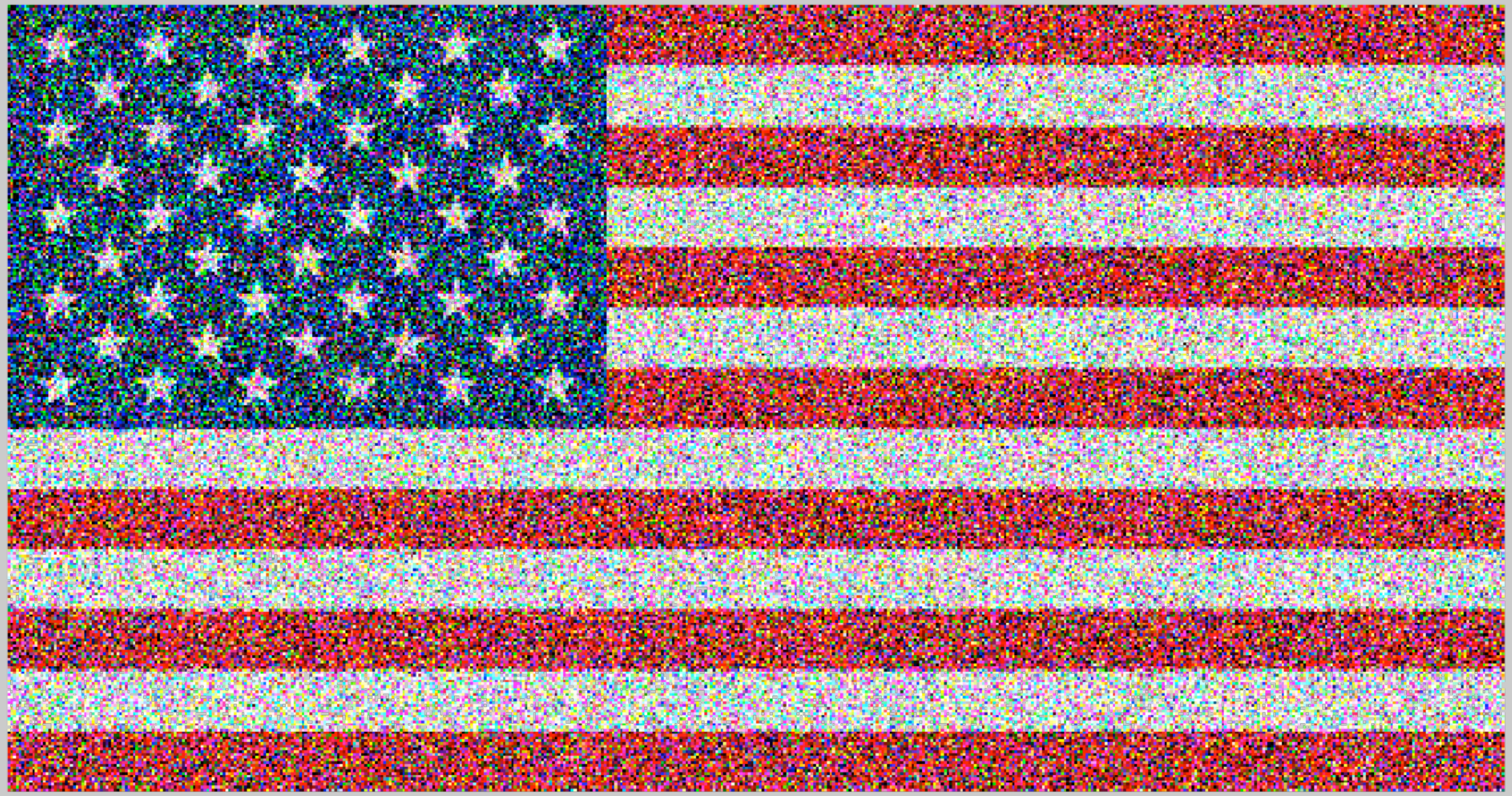

Rank=1
2
3
4
5
6
7
8
9

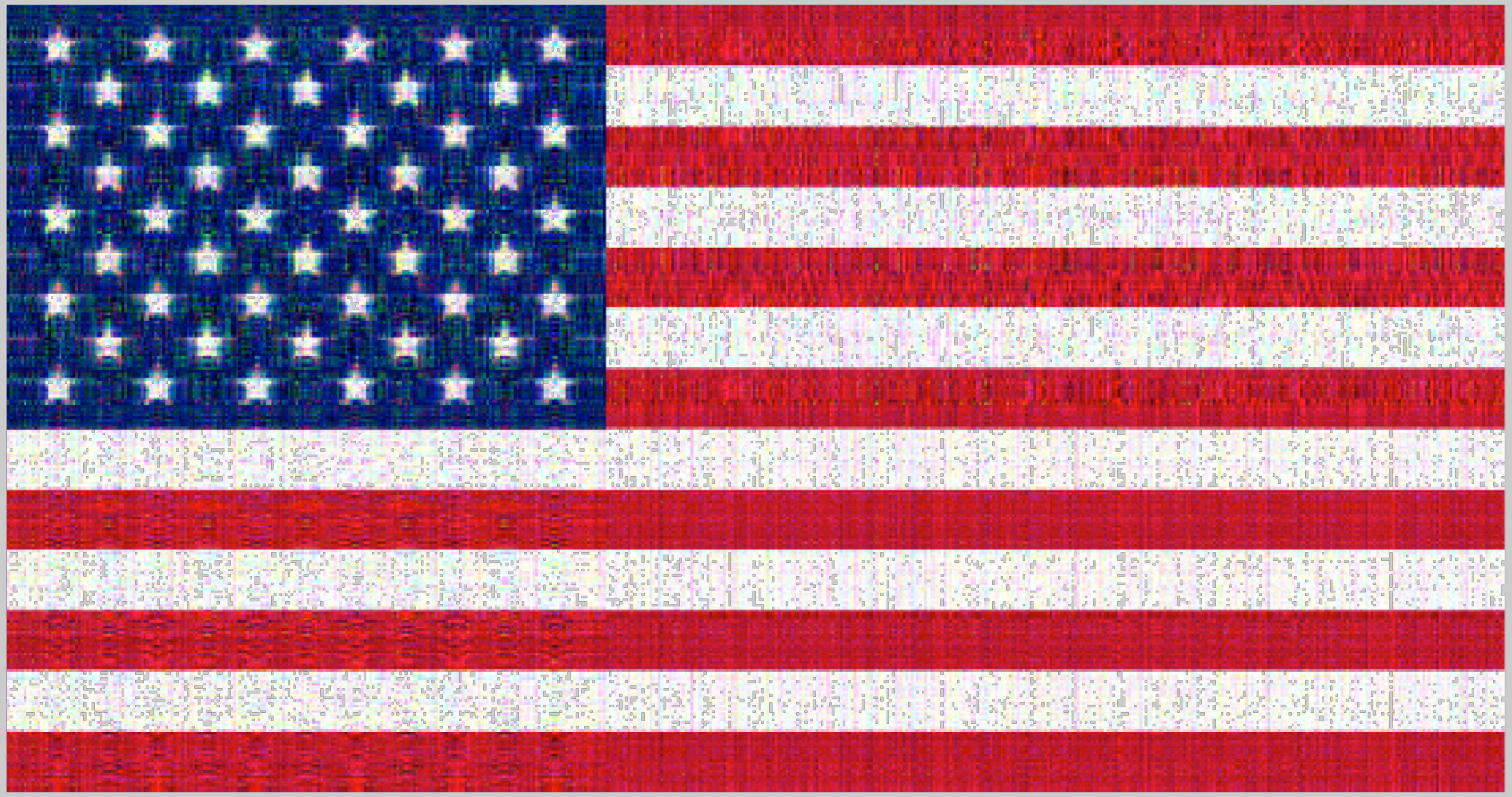

Rank=1
2
3
4
5
6
7
8
9

Noisy Flag

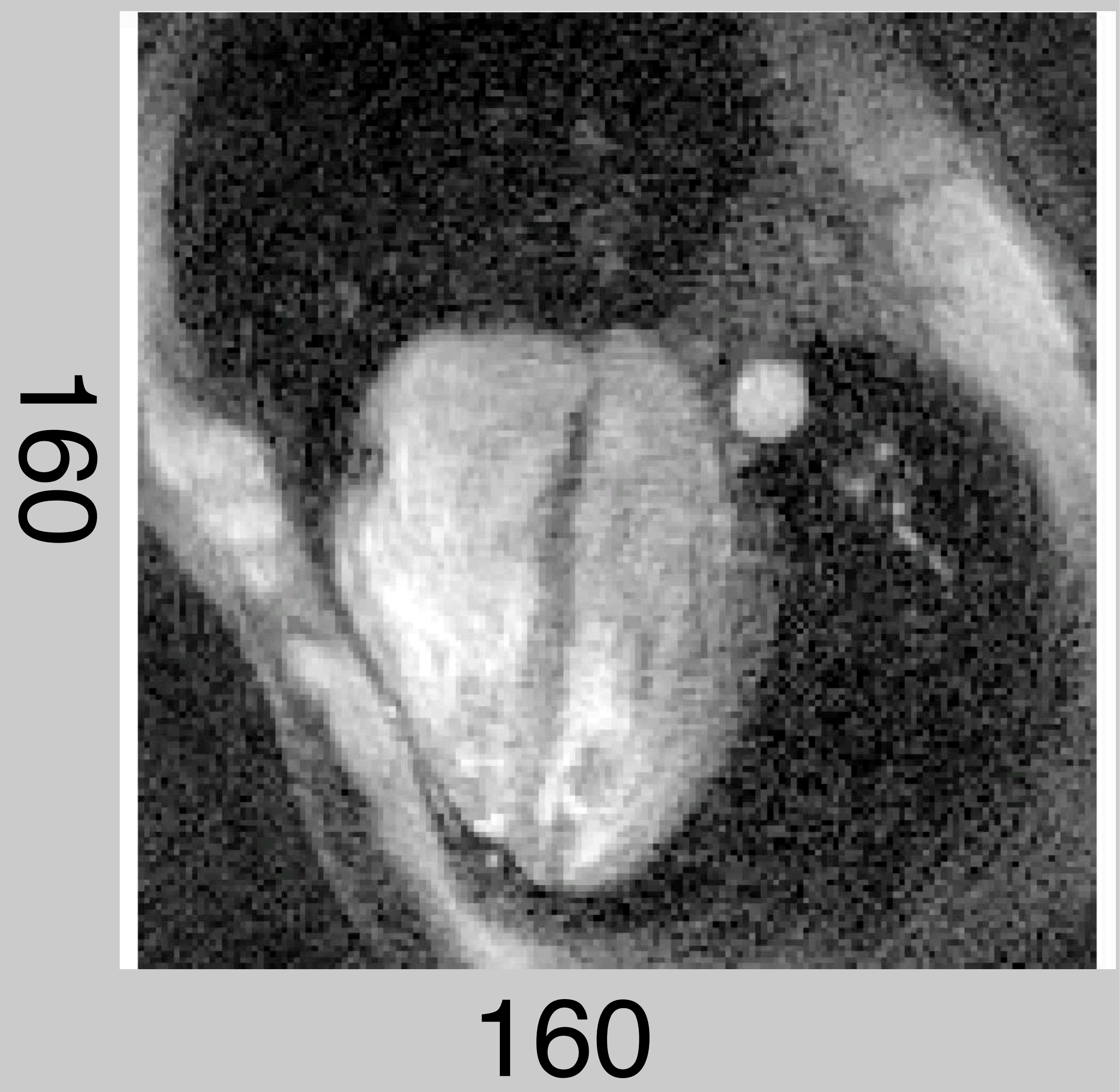
$$\begin{bmatrix} 1.02 & 0.99 & 0.98 & 1.03 & 1.01 & 1 \\ 2 & 1.98 & 2.01 & 2.03 & 1.99 & 1.97 \\ 3.01 & 2.98 & 3 & 2.99 & 3.03 & 3.02 \end{bmatrix}$$

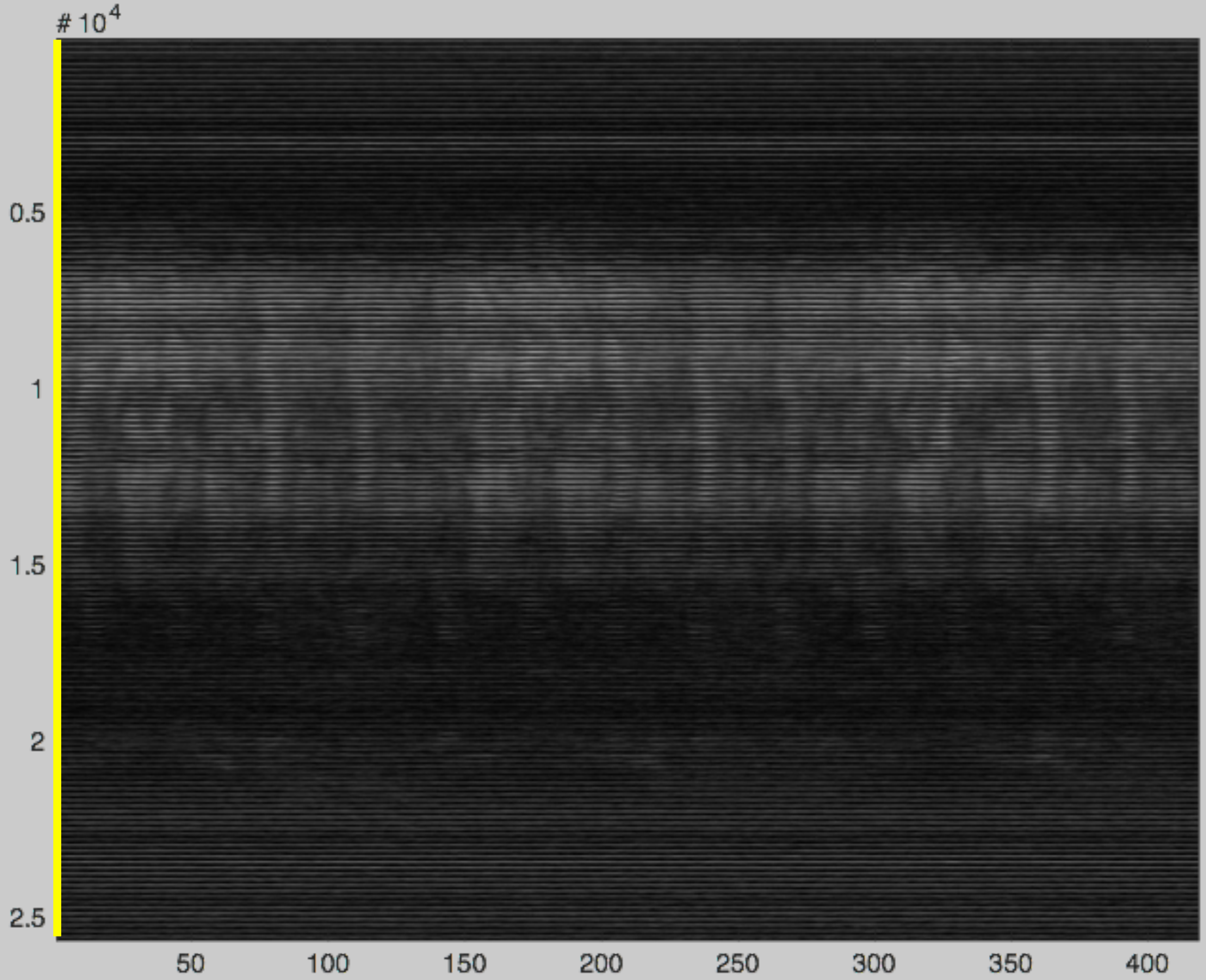
$$\sum_{i=0}^{5} \sigma_i \vec{u}_i \vec{v}_i^T$$





Video of a Heart



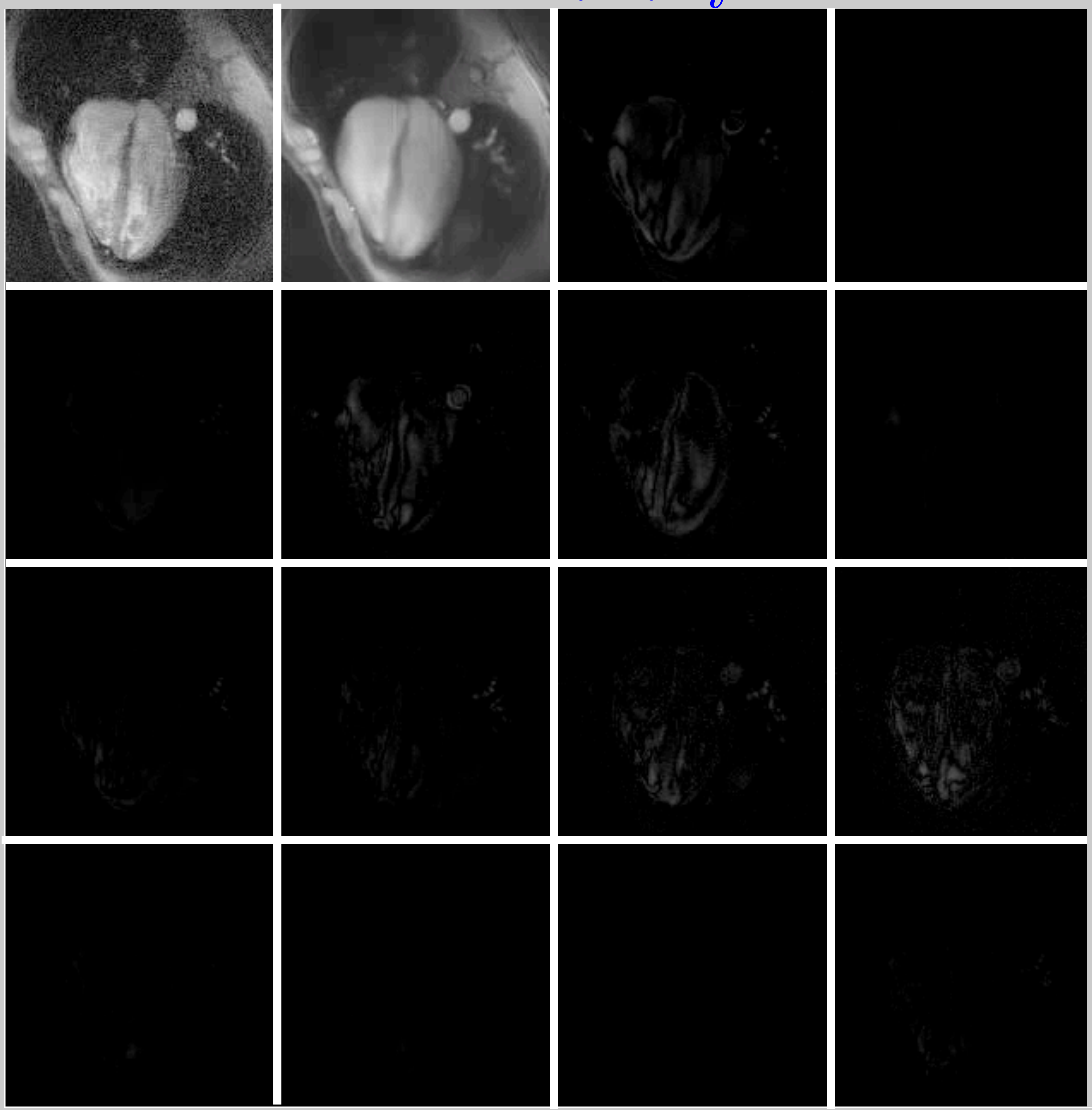


160^2

418

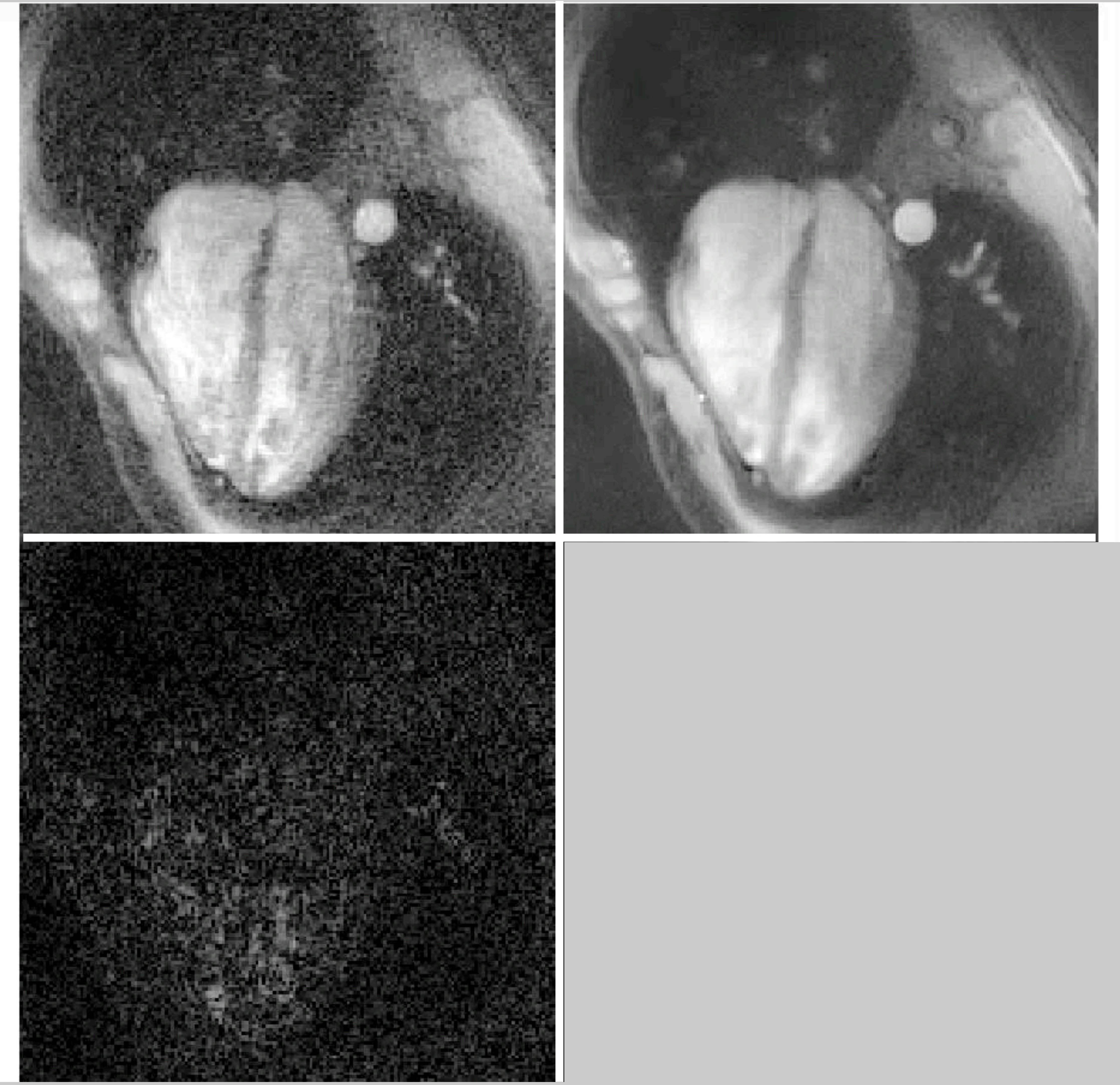
Heart Example

$$\sigma_i \vec{u}_i \vec{v}_i^T$$



$$\sum_{i=0}^{r} \sigma_i \vec{u}_i \vec{v}_i^T$$

Heart Example



$$\sum_{i=0}^{15} \sigma_i \vec{u}_i \vec{v}_i^T$$

$$\sum_{i=16}^{417} \sigma_i \vec{u}_i \vec{v}_i^T$$

Data Analysis with SVD

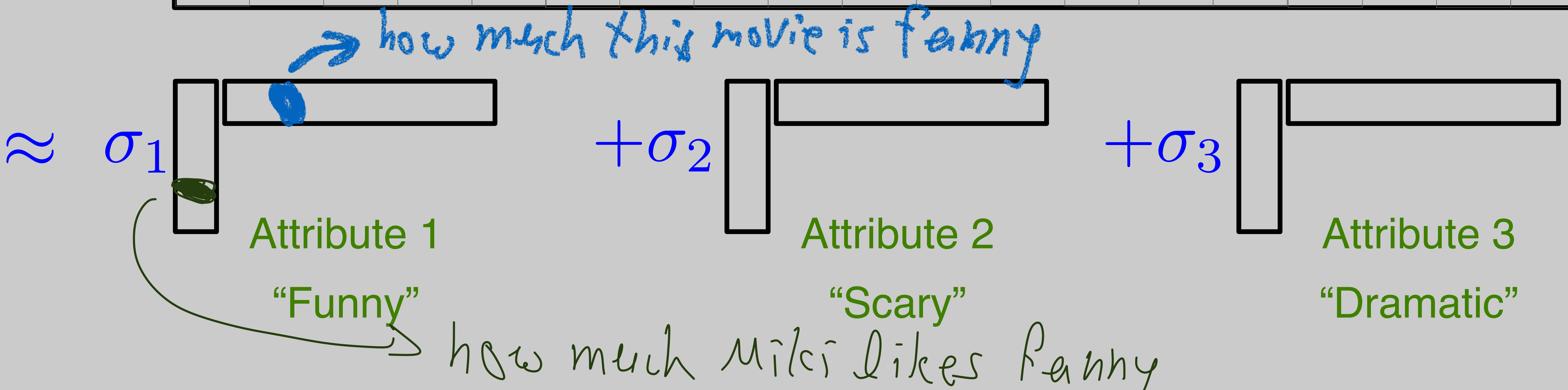
$$A \approx \sigma_1 \vec{u}_1 \vec{v}_1^T + \sigma_2 \vec{u}_1 \vec{v}_1^T + \cdots + \sigma_{\hat{r}} \vec{u}_{\hat{r}} \vec{v}_{\hat{r}}^T$$

effective

n movies ratings

m viewers

1	2	5	5	1	3	3	3	2	5	5	4	4	3	2	2	5	1	1
5	3	2	1	1	3	3	3	1	1	2	4	5	5	4	4	1	3	2
5	1	2	1	1	2	3	3	1	1	2	1	5	5	3	5	1	1	2
5	3	2	1	1	3	2	3	1	1	2	4	1	1	4	4	5	1	5
1	2	3	2	1	3	2	3	2	1	2	1	1	1	4	4	5	1	5
1	1	1	1	5	3	3	3	1	5	2	4	4	4	2	5	5	1	1

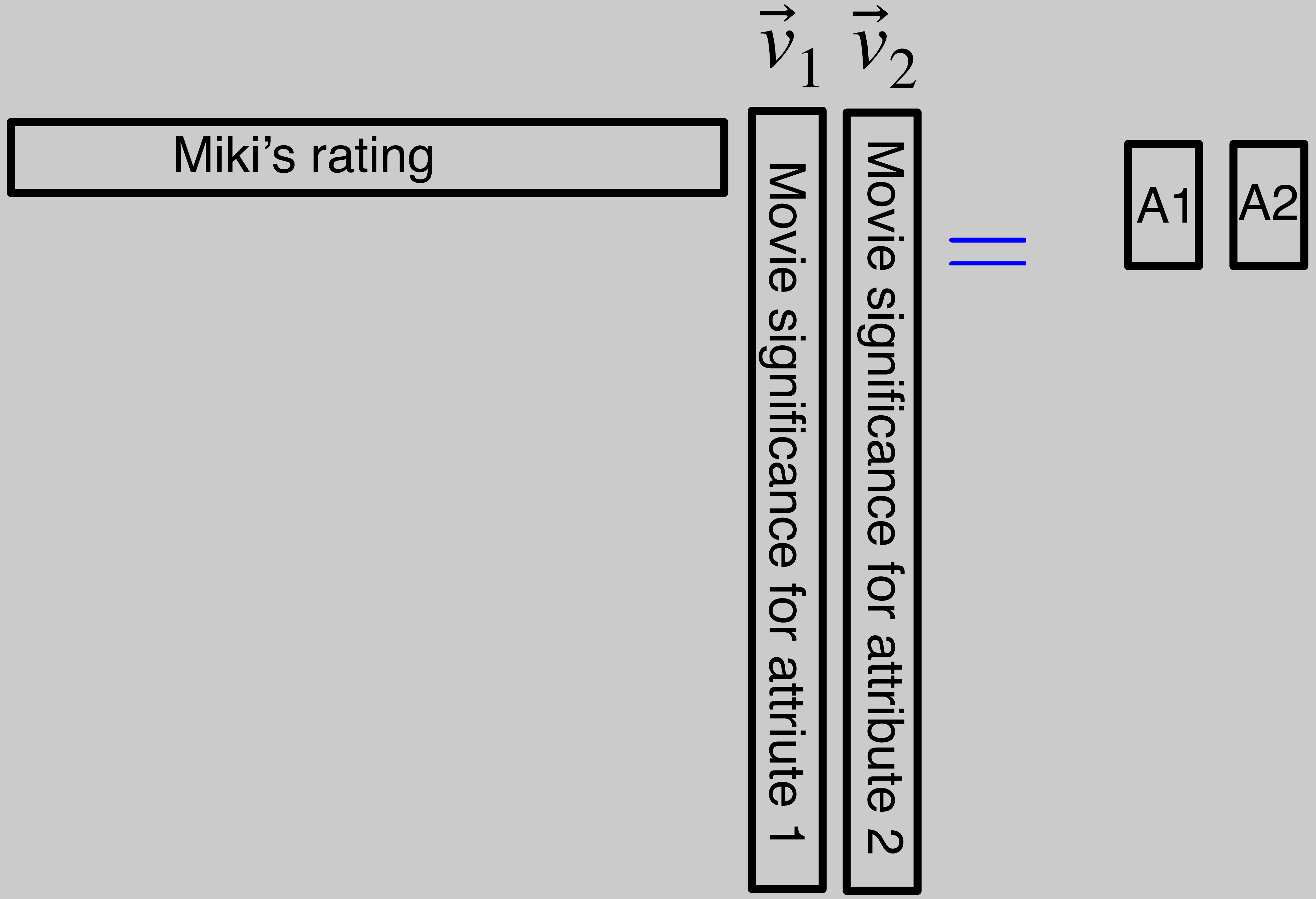


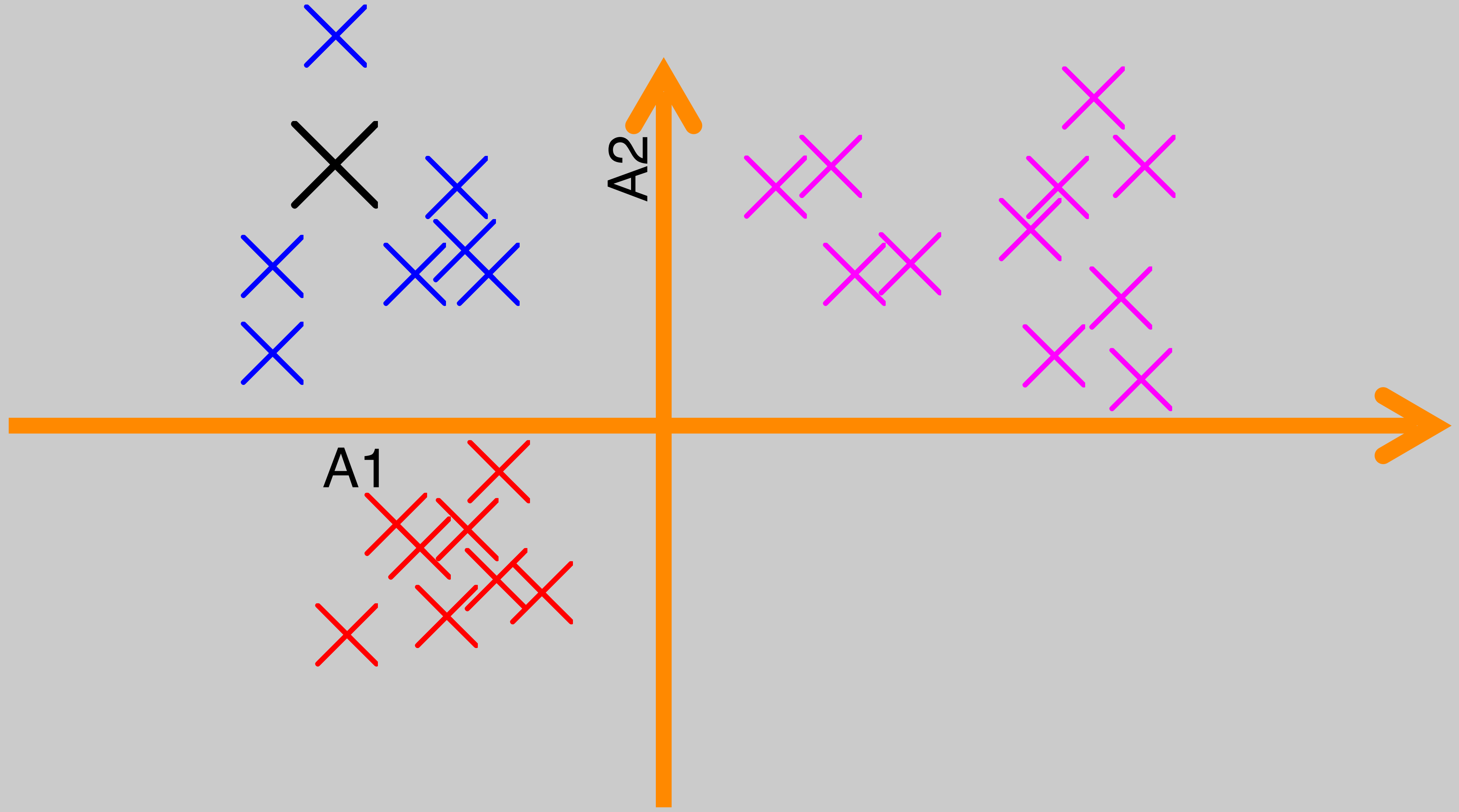
Classification with SVD

n movies

m viewers

1	2	5	5	1	3	3	3	2	5	5	4	4	3	2	2	5	1	1
5	3	2	1	1	3	3	3	1	1	2	4	5	5	4	4	1	3	2
5	1	2	1	1	2	3	3	1	1	2	1	5	5	3	5	1	1	2
5	3	2	1	1	3	2	3	1	1	2	4	1	1	4	4	5	1	5
1	2	3	2	1	3	2	3	2	1	2	1	1	1	4	4	5	1	5
1	1	1	1	5	3	3	3	1	5	2	4	4	4	2	5	5	1	1





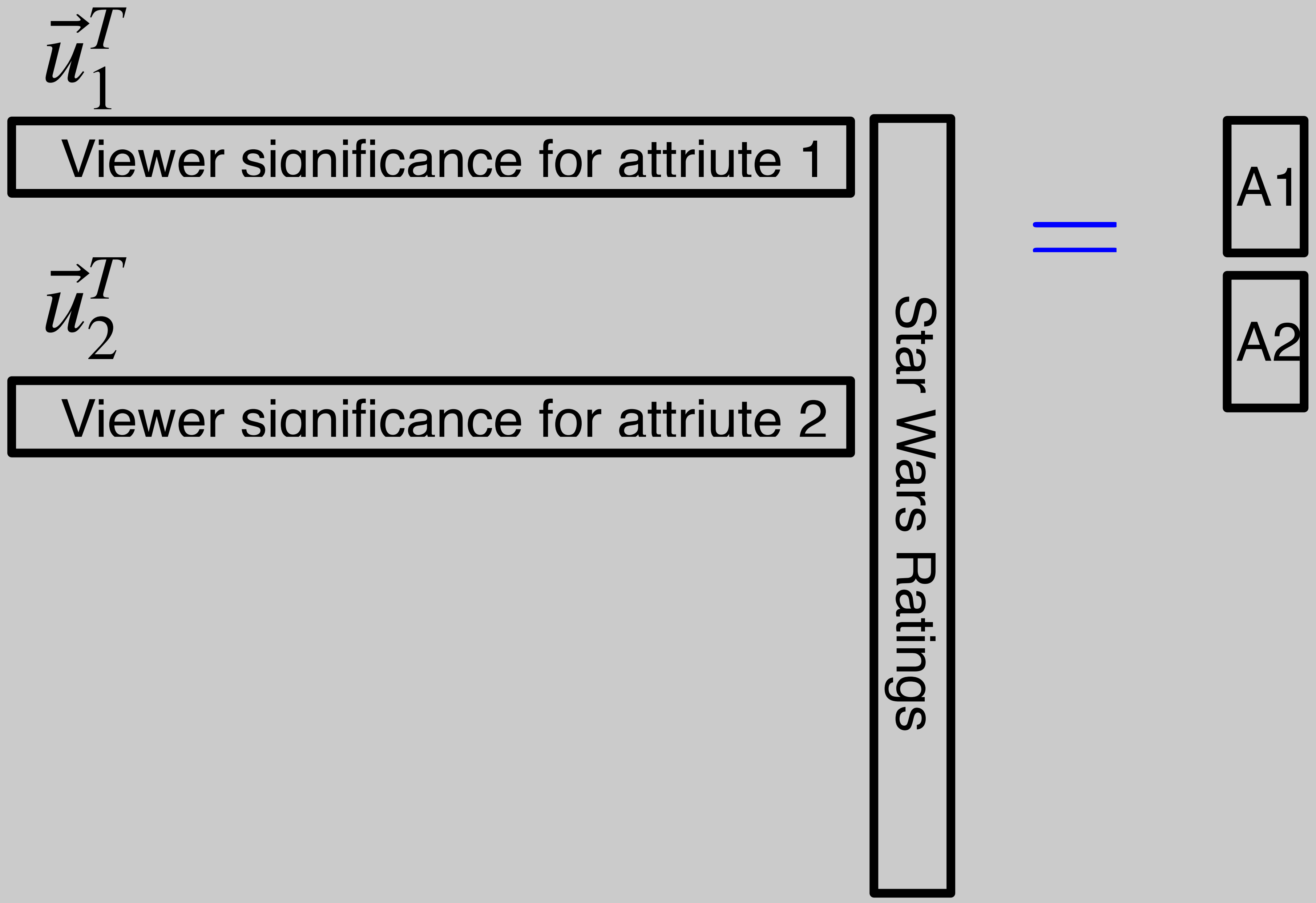
Miki belongs to class: like A2 don't like A1

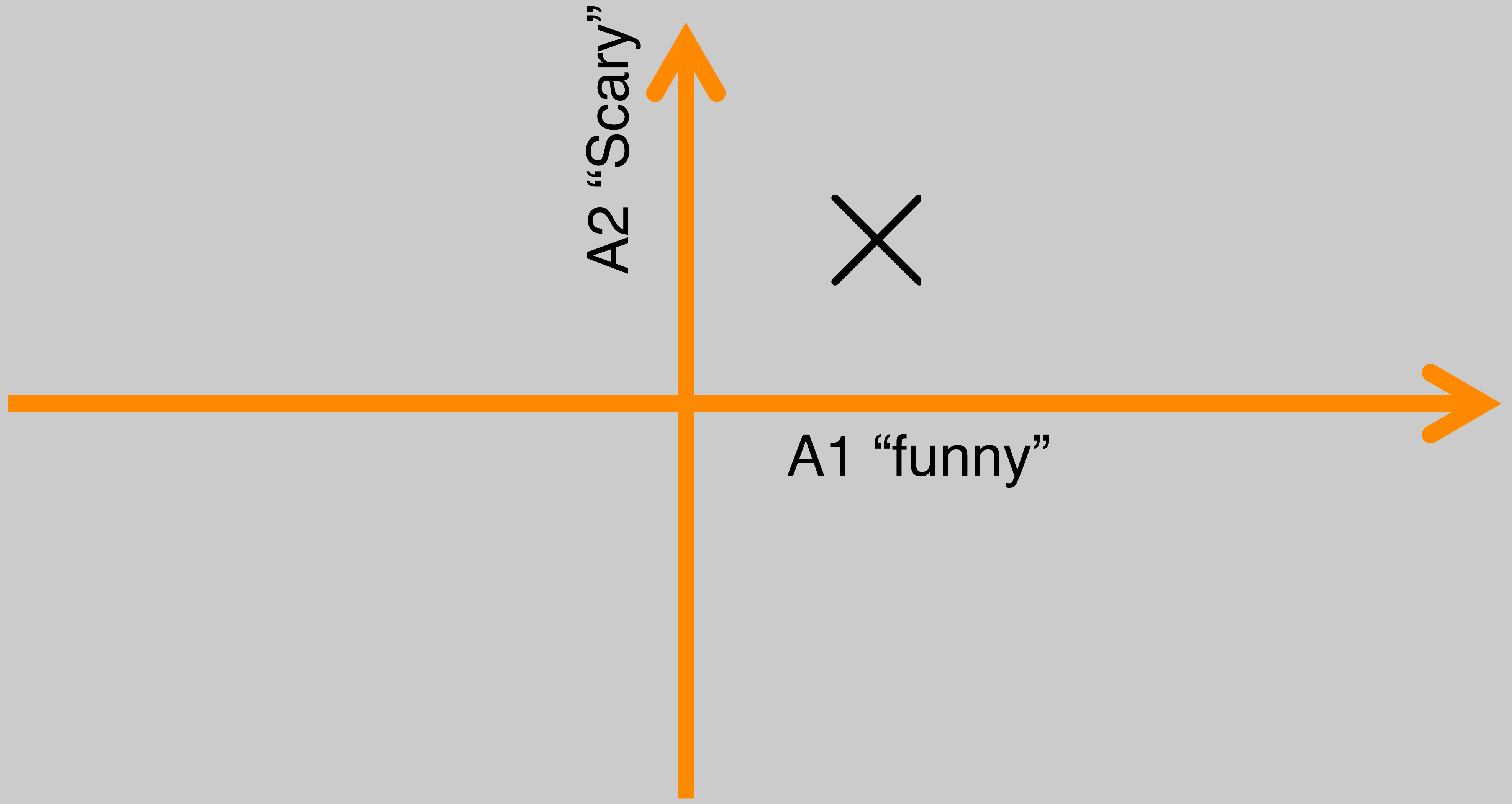
Classification with SVD

n movies

m viewers

1	2	5	5	1	3	3	3	2	5	5	4	4	3	2	2	5	1	1
5	3	2	1	1	3	3	3	1	1	2	4	5	5	4	4	1	3	2
5	1	2	1	1	2	3	3	1	1	2	1	5	5	3	5	1	1	2
5	3	2	1	1	3	2	3	1	1	2	4	1	1	4	4	5	1	5
1	2	3	2	1	3	2	3	2	1	2	1	1	1	4	4	5	1	5
1	1	1	1	5	3	3	3	1	5	2	4	4	4	2	5	5	1	1





Star Wars is somewhat funny and somewhat scary

Prediction with SVD

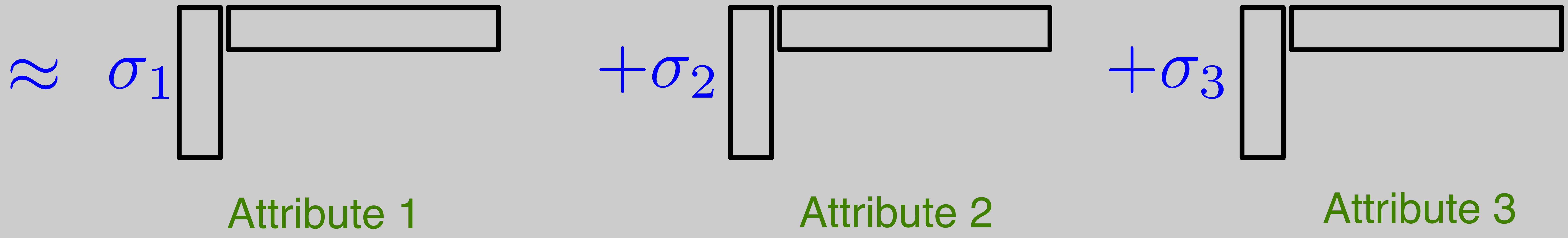
Can try to predict preferences of a new customer with few ratings

See homework!

n movies

m views

1	2	5	5	1	3	3	3	2	5	5	4	4	3	2	2	5	1	1
5	3	2	1	1	3	3	3	1	1	2	4	5	5	4	4	1	3	2
5	1	2	1	1	2	3	3	1	1	2	1	5	5	3	5	1	1	2
5	3	2	1	1	3	2	3	1	1	2	4	1	1	4	4	5	1	5
1	2	3	2	1	3	2	3	2	1	2	1	1	1	4	4	5	1	5
1	1	1	1	5	3	3	3	1	5	2	4	4	4	2	5	5	1	1
1	?	?	2	?	?	?	?	3	5	1	?	?	?	?	5	2	?	3



Low-rank Completion

What if my database is full of “holes”?

Should be still low-rank!

n movies

m views

1		5	5	1	3		3	2	5	5	4		3	2	2	5		1	
5	3		1	1	3	3		1	1	2		5		4	4	1	3	2	
	1	2	1	1		3	3		1	2	1	5	5	3	5		1	2	
5	3		1		3	2		1		2	4	1		4		5		5	
1		3	2	1	3	2	3	2	1		1		1	4		5	1	5	
1	1			5	3	3		5	2	4			2	5	5	1	1		

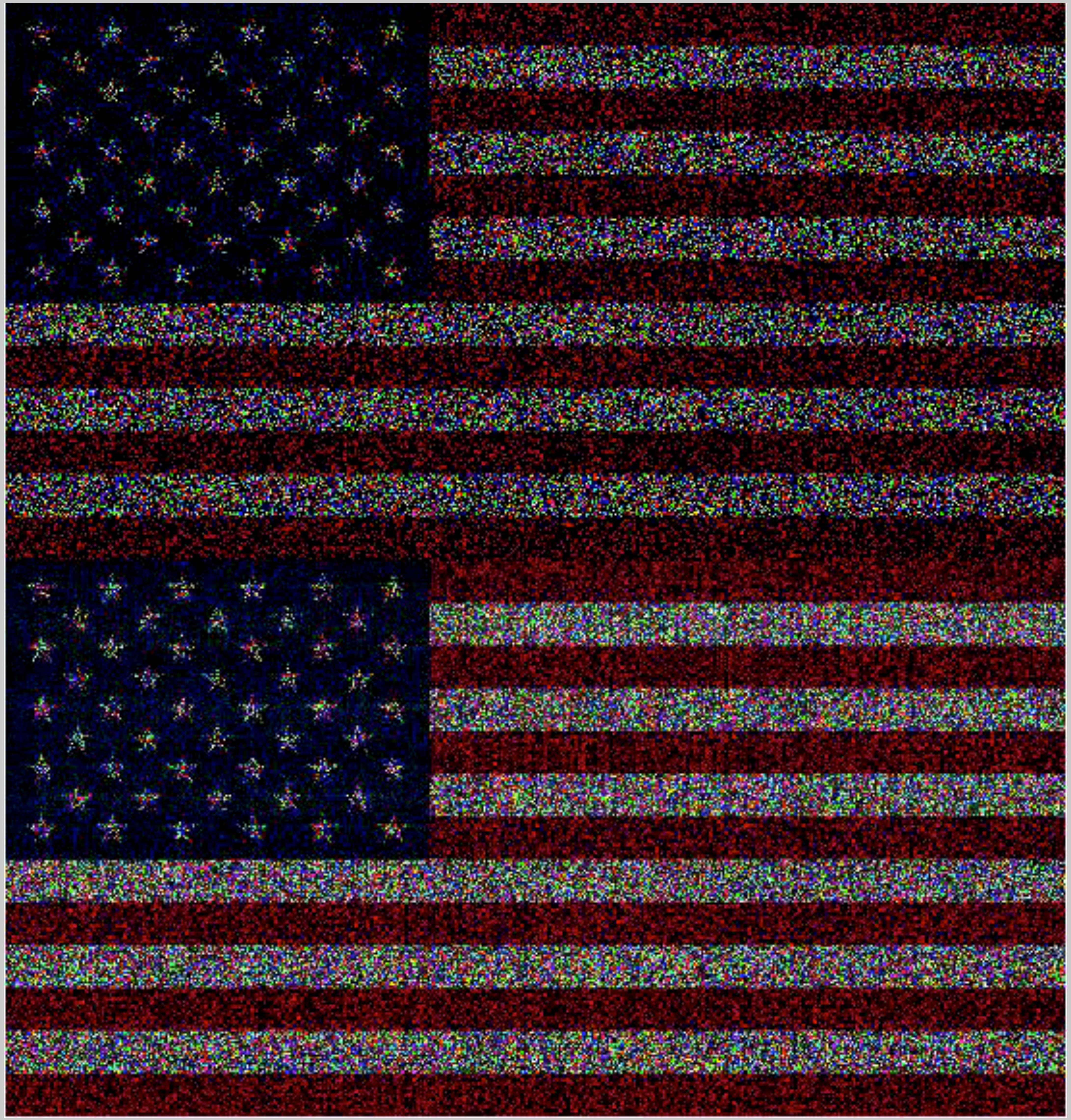
Q) Can we complete missing data?

A) Sometimes! Very recent mathematical and practical results show you can.
Keywords: Compressed Sensing, Low-rank completion, robust PCA

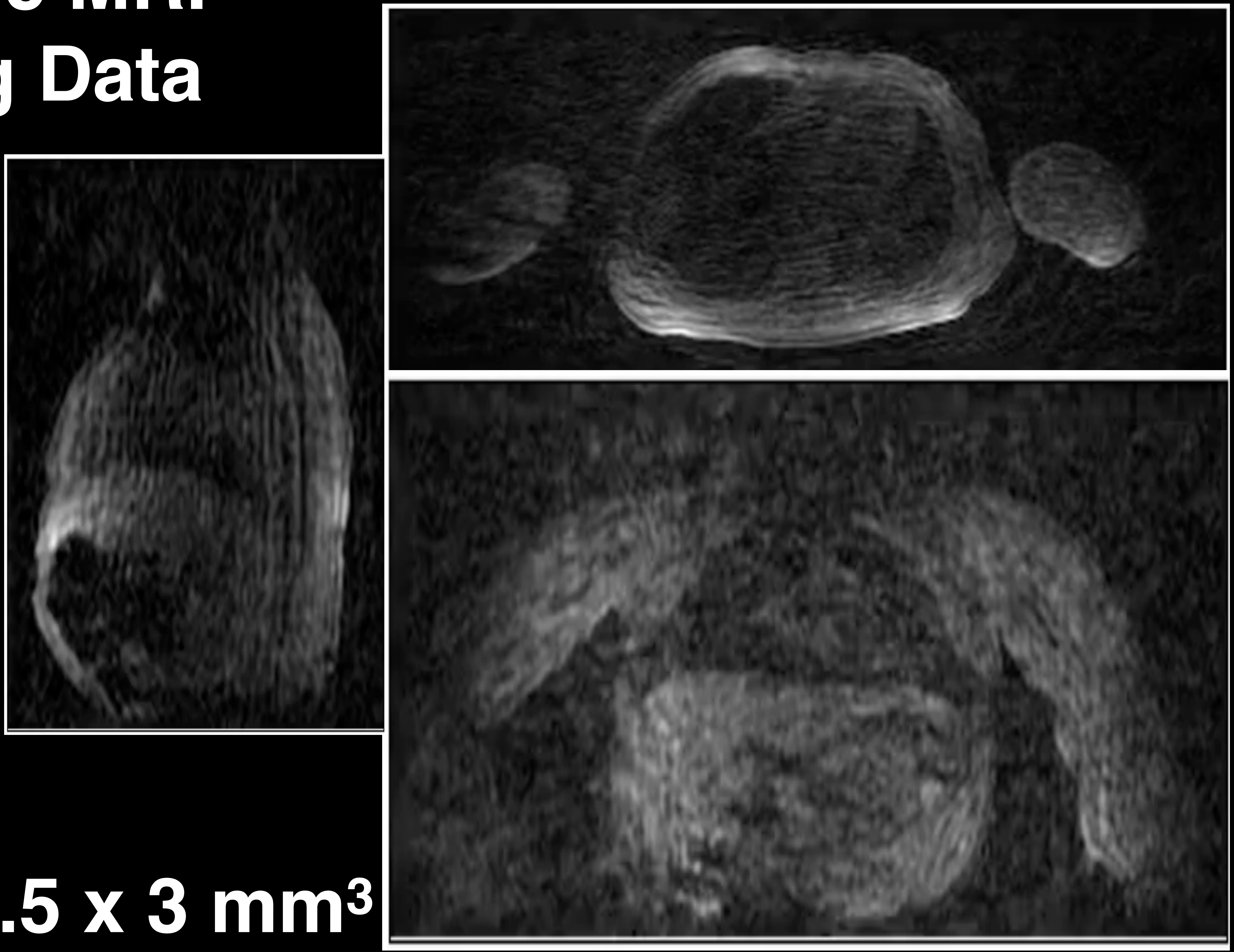
E. Candes and B. Recht, Foundation of Computational Mathematics, 2009;9:717

Low-Rank Recovery from 20% pixels

- Algorithm for low-rank completion:
  - flag_hat = flag
  - Compute [U,S,V] = svd(flag_hat)
  - $$\text{flag}_{\text{hat}} = \sum_{i=0}^{6} \sigma_i \vec{u}_i \vec{v}_i^T$$
  - update missing pixels in flag from flag_hat
  - repeat (250 times here)

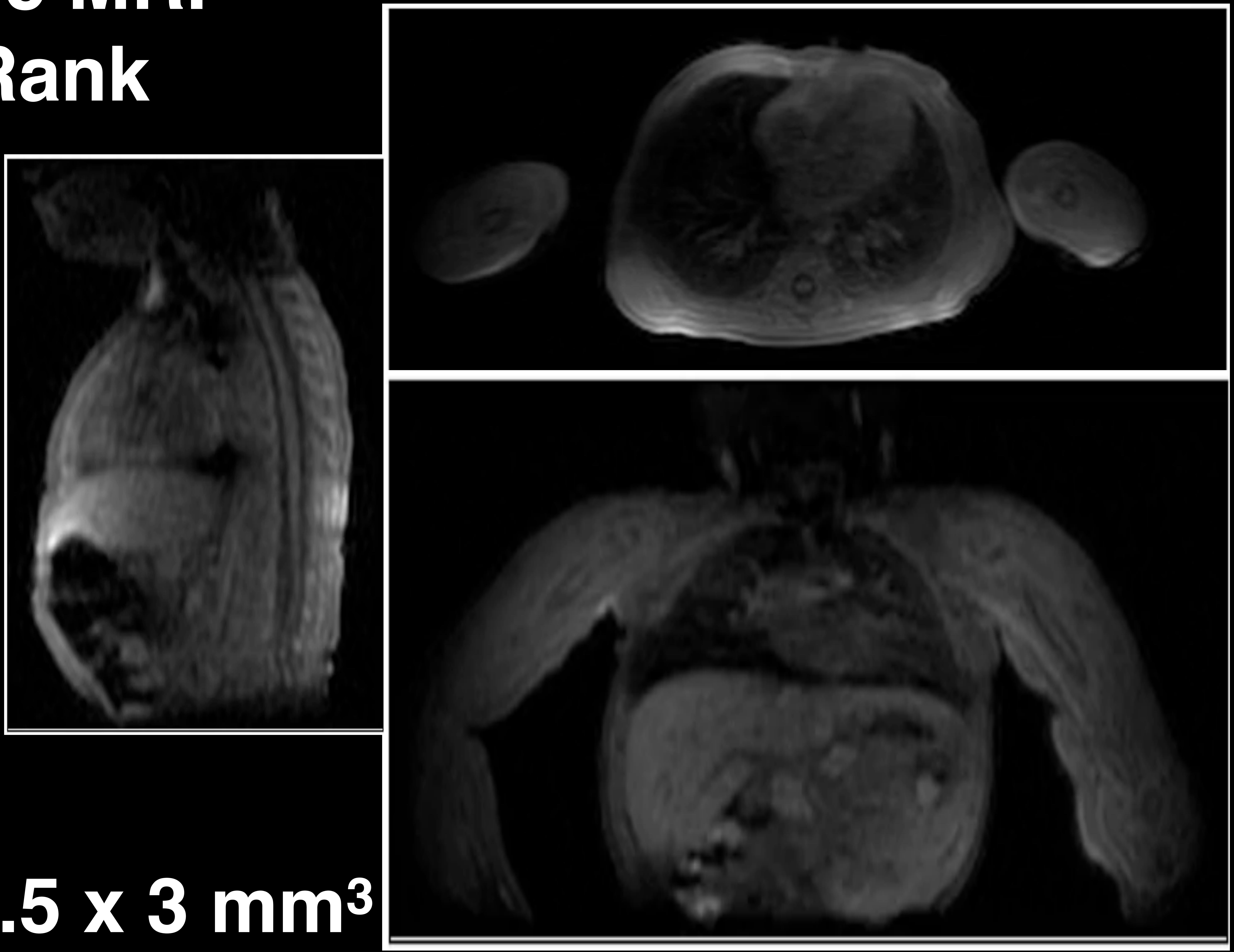


Dynamic MRI Missing Data



~1.5 x 1.5 x 3 mm^3

Dynamic MRI
Low-Rank

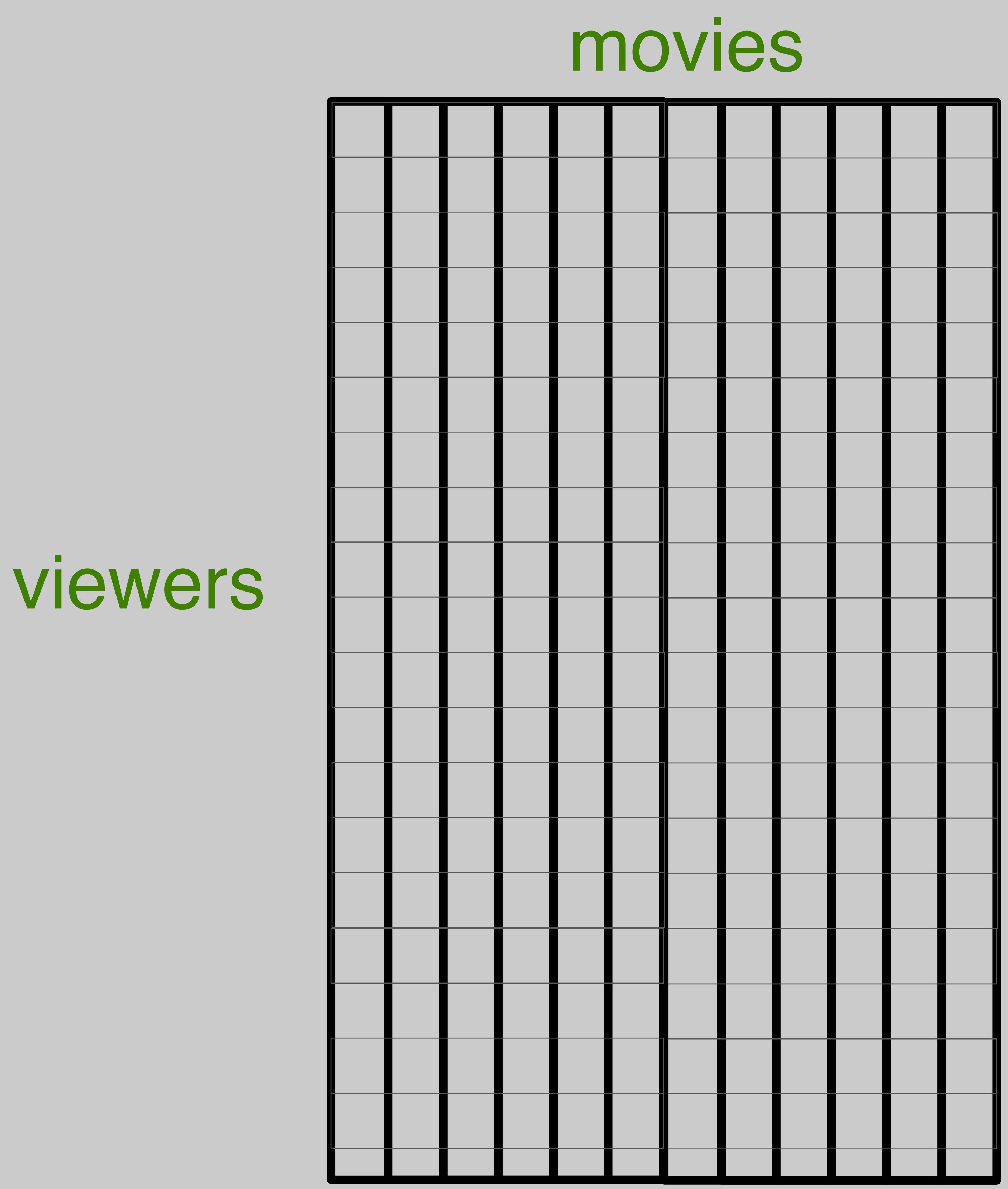


~1.5 x 1.5 x 3 mm^3

Principal Component Analysis

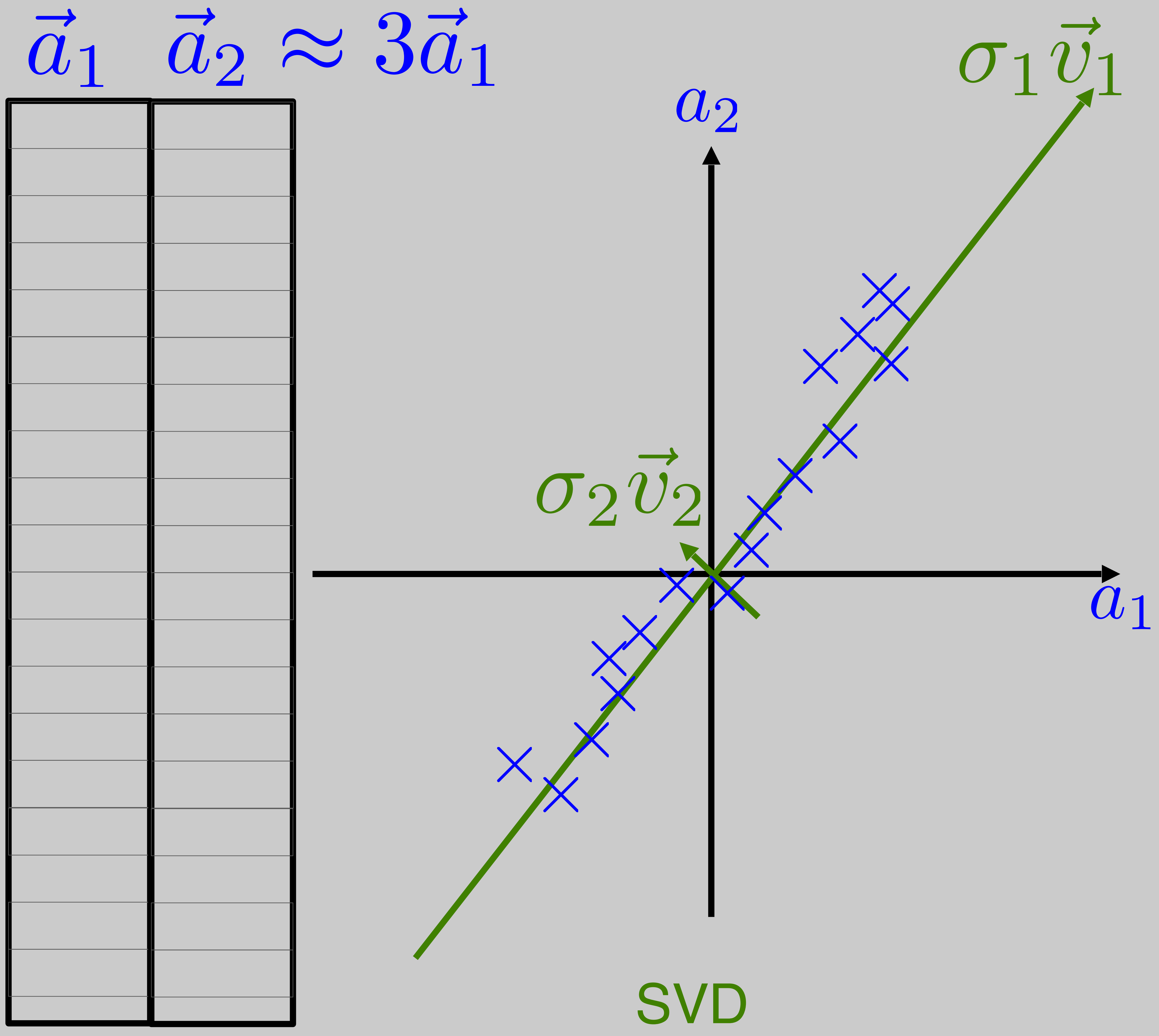
Application of the SVD to datasets to learn features

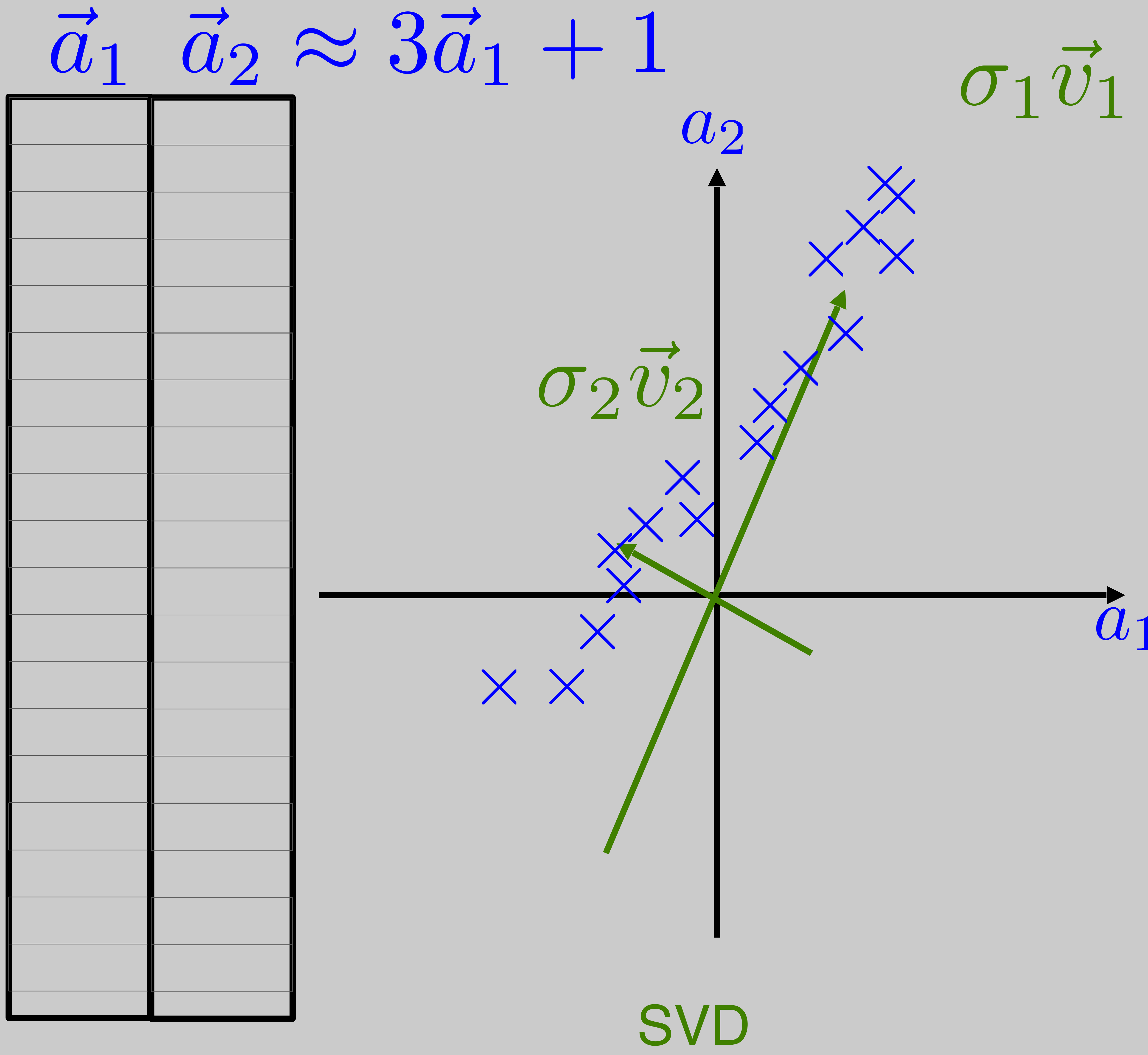
PCA is a tool in statistics and machine learning, which can be computed using SVD



Example -- PCA

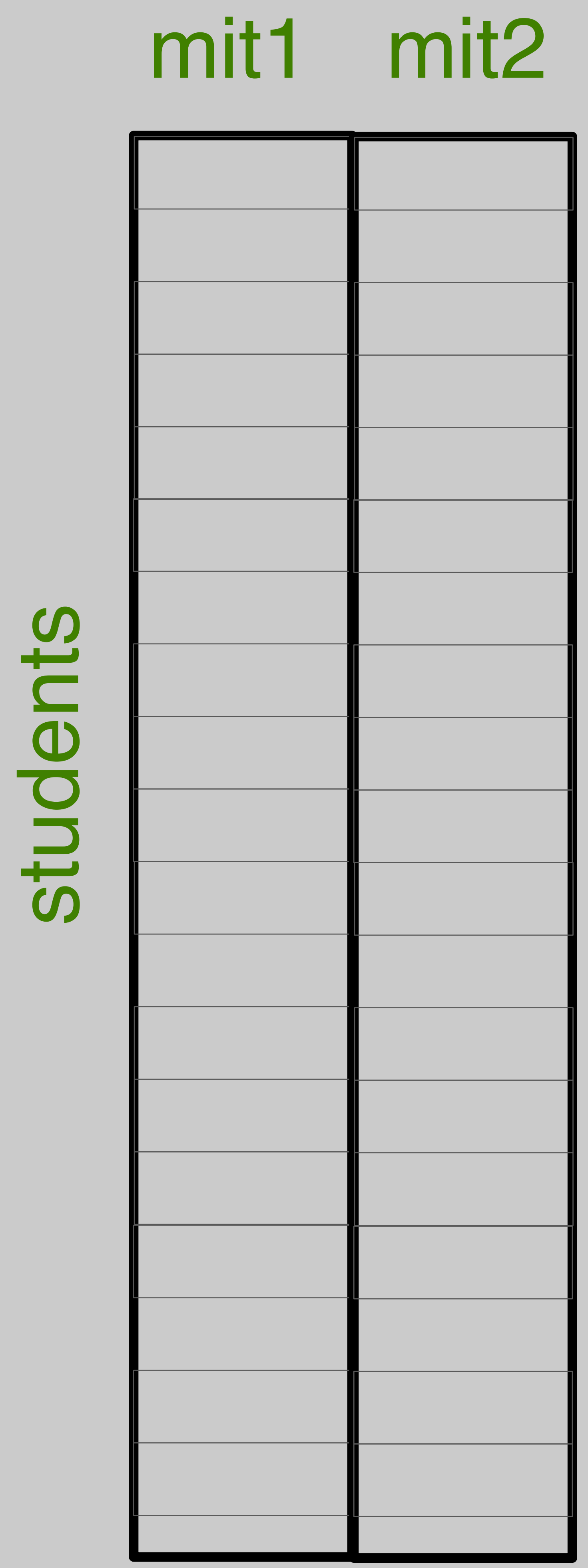
Consider data s.t.

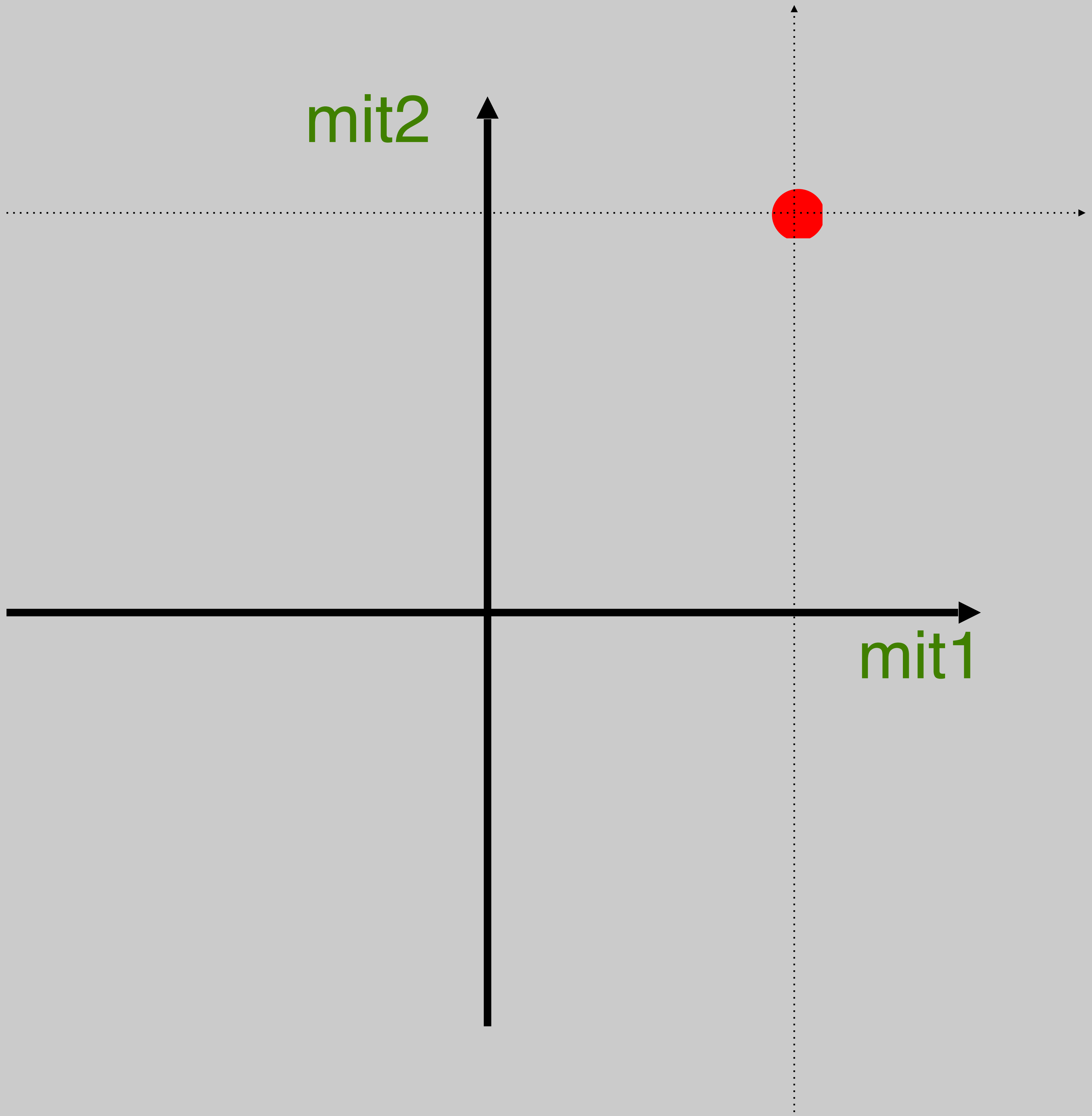




Example -- PCA

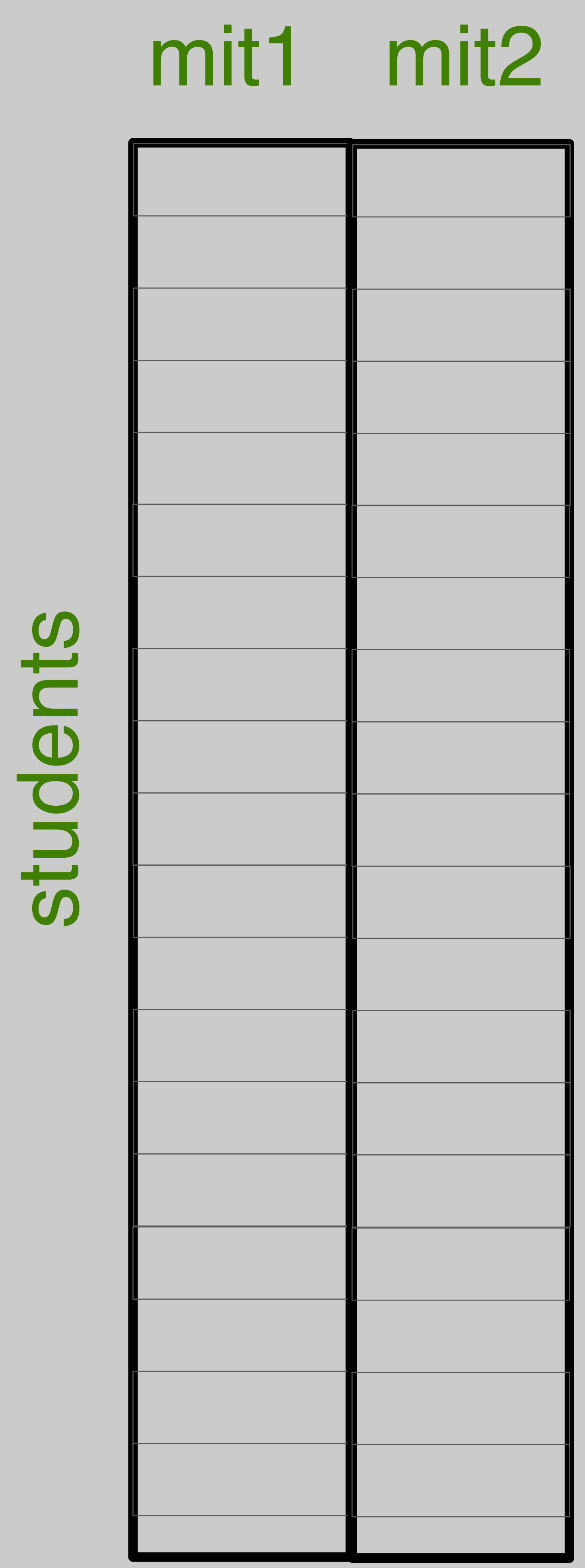
Consider miterm data

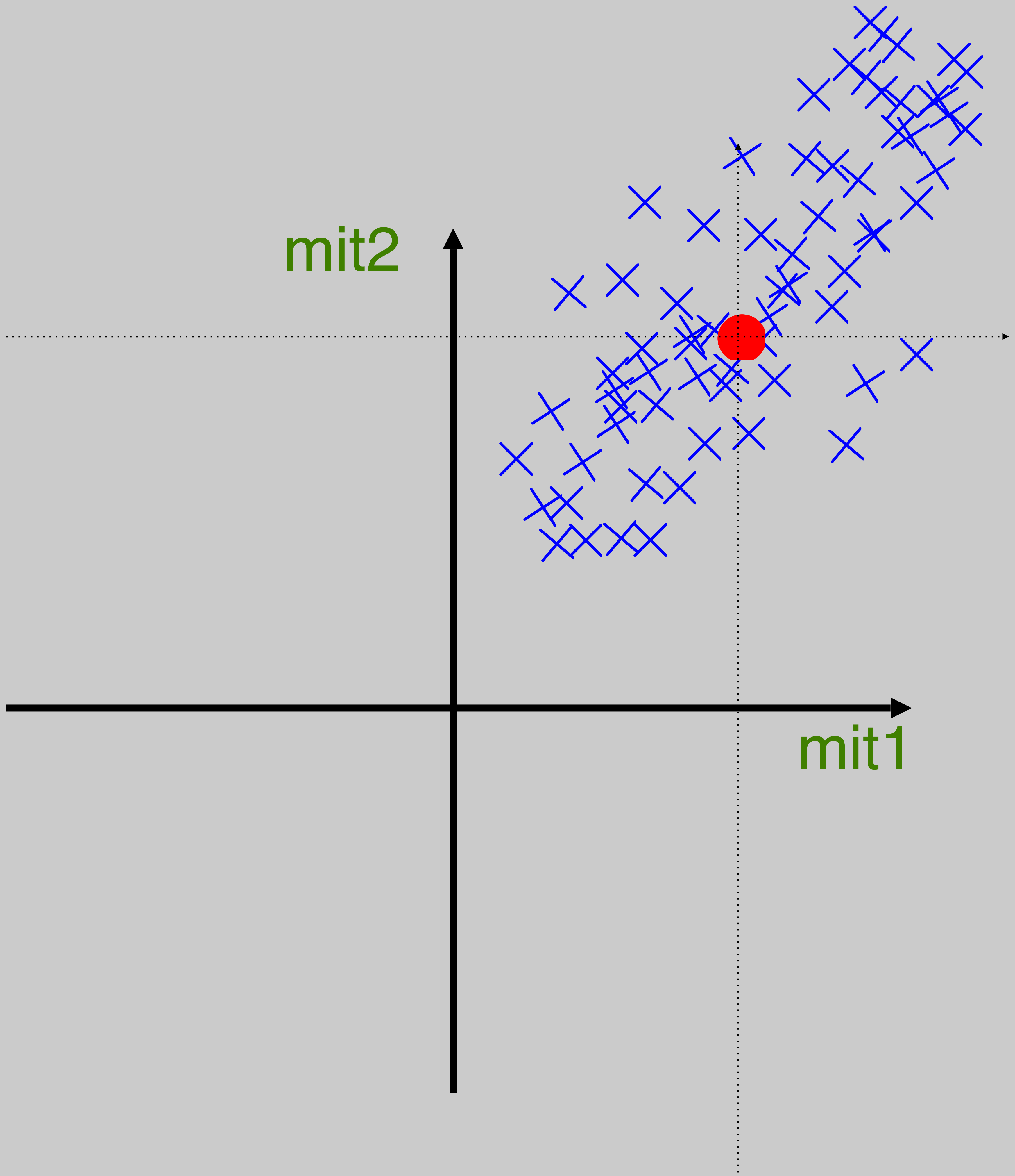




Example -- PCA

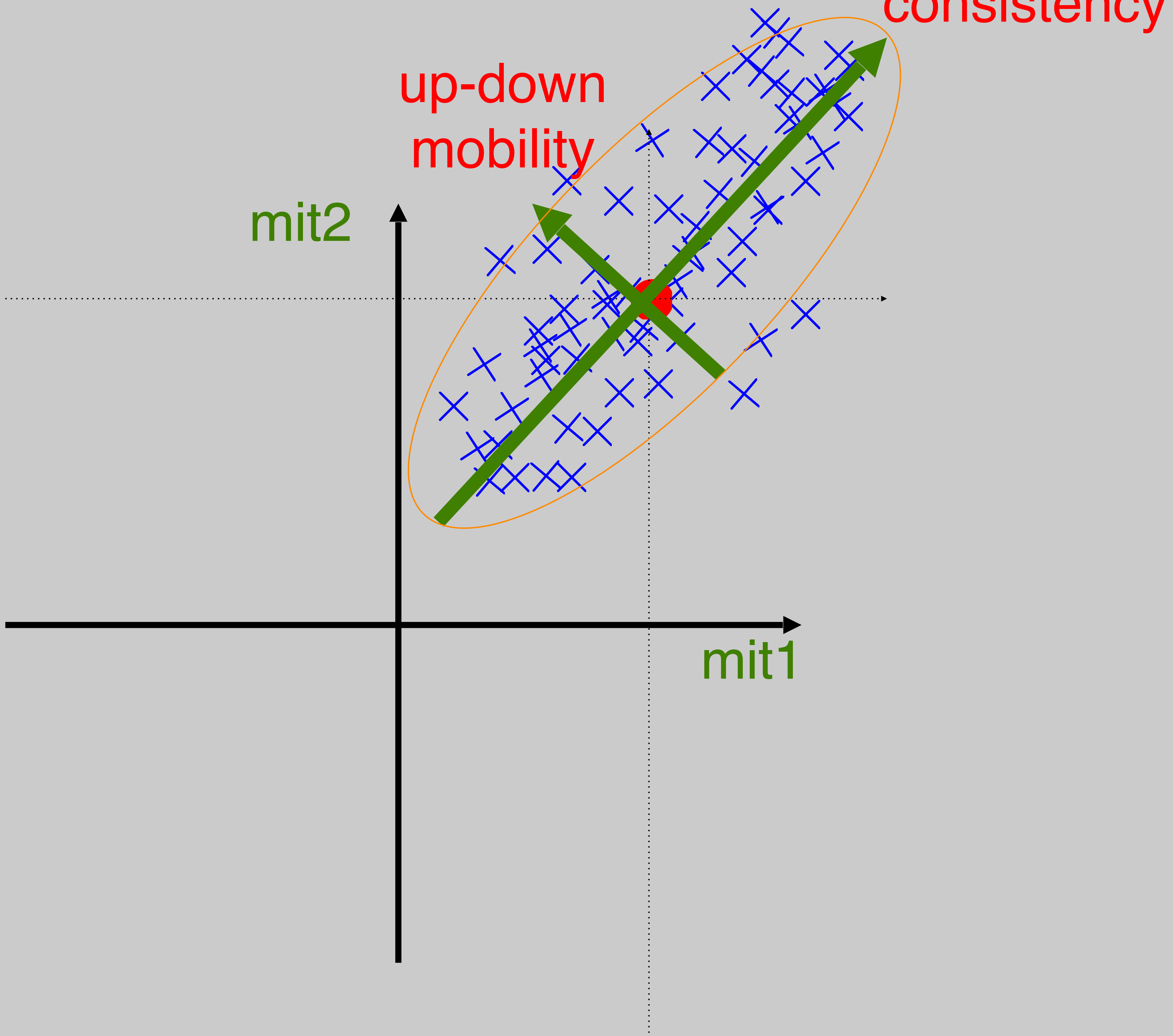
Consider miterm data





Example -- PCA

Consider miterm data

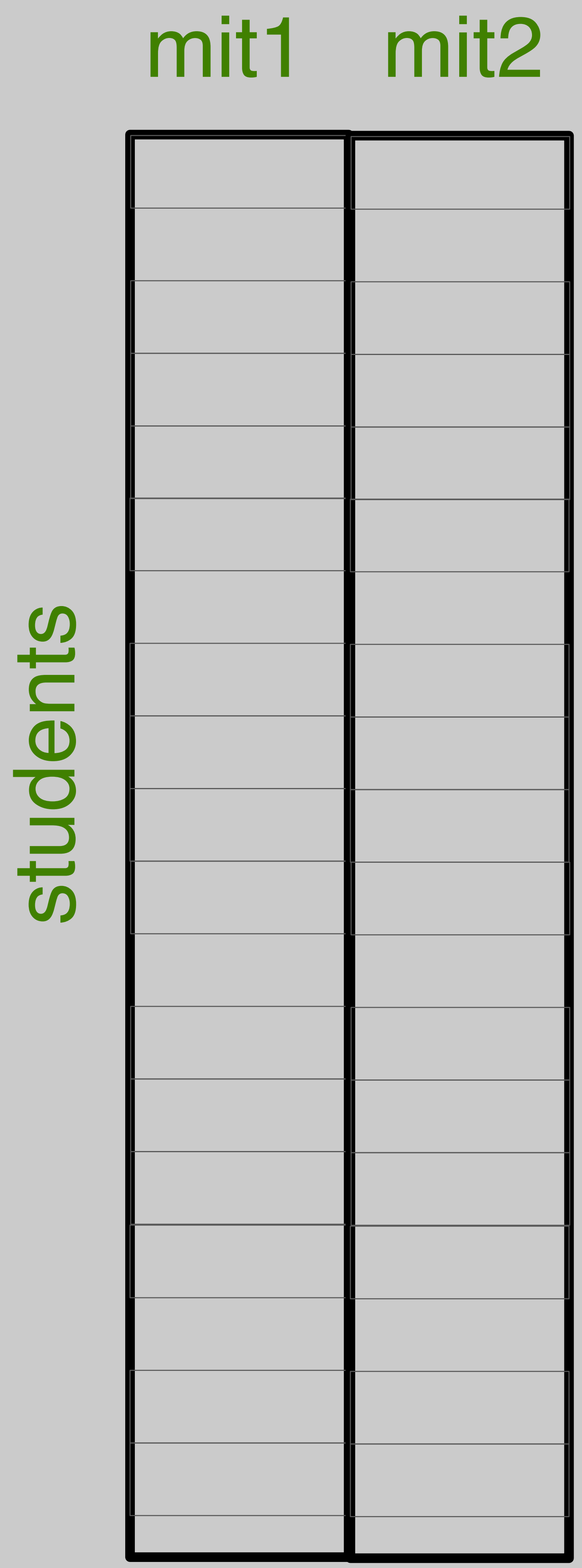


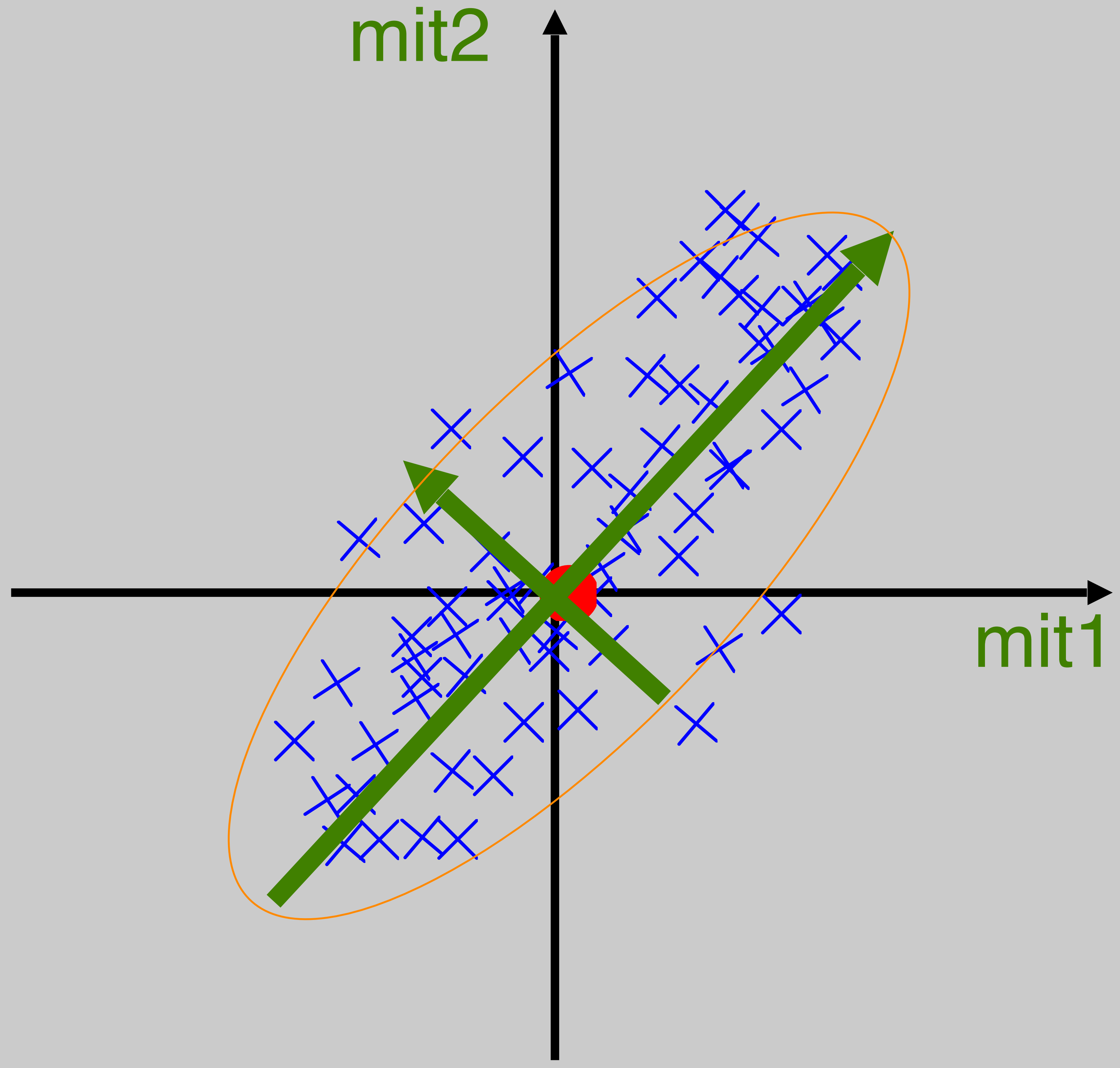
PCA Procedure

Remove averages from column of A

From A^TA, find $\sigma_i, \quad \vec{v_i}$

$\vec{v_i}$ are principal components!





$A^T A$ as sample covariance matrix

$$A = \vec{a} \qquad a_\mu = \frac{1}{N}\sum_{i=0}^{N-1} a_i \qquad \tilde{A} = \vec{a} - a_\mu \vec{1}$$

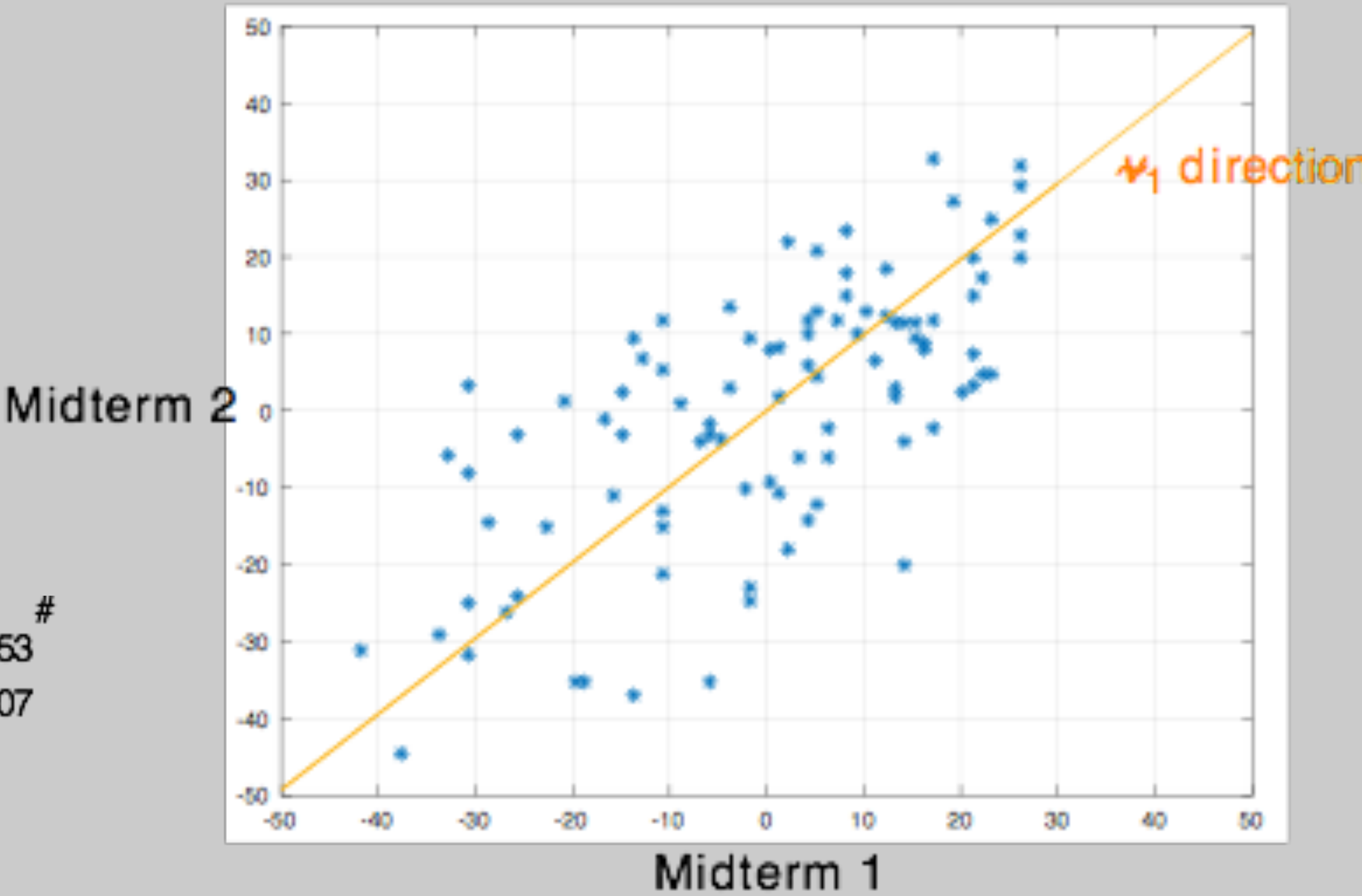
$$\tilde{A}^T \tilde{A} = (\vec{a} - a_\mu \vec{1})^T (\vec{a} - a_\mu \vec{1})$$

$$= \vec{a}^T \vec{a} - 2N a_\mu^2 + N a_\mu^2 \quad = \vec{a}^T \vec{a} - N a_\mu^2$$

$$\frac{1}{N}\tilde{A}^T \tilde{A} = \frac{1}{N}\vec{a}^T \vec{a} - a_\mu^2 = \frac{1}{N}\sum_{i=0}^{N-1} a_i^2 - a_\mu^2 \quad = a_\sigma^2$$

Sample
Variance!

Example midterm



$$\frac{1}{93} A^T A = \begin{bmatrix} 297.69 & 202.53 \\ 202.53 & 292.07 \end{bmatrix}$$