

# The Singular Value Decomposition (SVD)

The matrix factorization behind PCA

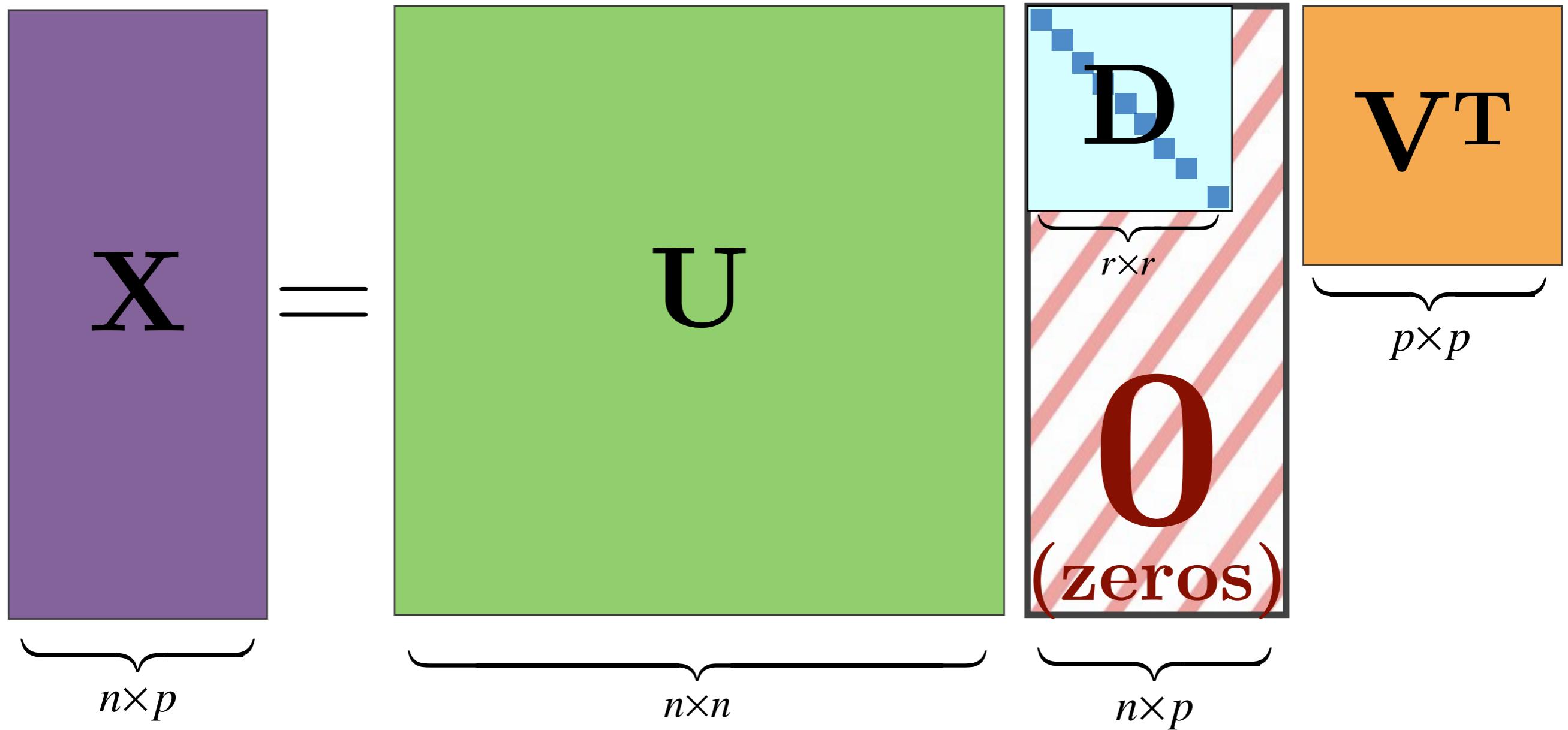
# Singular Value Decomposition (SVD)

For any  $n \times p$  matrix  $\mathbf{X}$  with rank= $r$

There exists orthogonal matrices  $\mathbf{U}_{n \times n}$  and  $\mathbf{V}_{p \times p}$   
and a diagonal matrix  $\mathbf{D}_{r \times r} = \text{diag}(\sigma_1, \sigma_2, \dots, \sigma_r)$  such that:

$$\mathbf{X} = \underbrace{\mathbf{U} \begin{pmatrix} \mathbf{D} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}}_{n \times p} \mathbf{V}^T \quad \text{with} \quad \sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r \geq 0$$

# SVD, Illustrated



# SVD, Illustrated

$$\mathbf{X} = \mathbf{U}_r \mathbf{D}_r \mathbf{V}_r^T$$

The diagram illustrates the Singular Value Decomposition (SVD) of a matrix  $\mathbf{X}$ . The matrix  $\mathbf{X}$  is shown as a purple rectangle labeled  $\mathbf{X}$  with dimensions  $n \times p$  indicated by a brace at the bottom. To its right is an equals sign. Following the equals sign is a green rectangle labeled  $\mathbf{U}_r$  with dimensions  $n \times r$  indicated by a brace at the bottom. To the right of  $\mathbf{U}_r$  is a red and white diagonal striped rectangle. To the right of this striped rectangle is a light blue square labeled  $\mathbf{D}_r$ , which contains several blue squares forming a checkerboard pattern. A brace below  $\mathbf{D}_r$  indicates its dimensions are  $r \times r$ . To the right of  $\mathbf{D}_r$  is another red and white diagonal striped rectangle. To the right of this striped rectangle is an orange rectangle labeled  $\mathbf{V}_r^T$ , with a brace at the bottom indicating its dimensions are  $r \times p$ . Below the orange rectangle is a large red zero symbol with the text "(zeros)" written underneath it, enclosed in a red rounded rectangle.

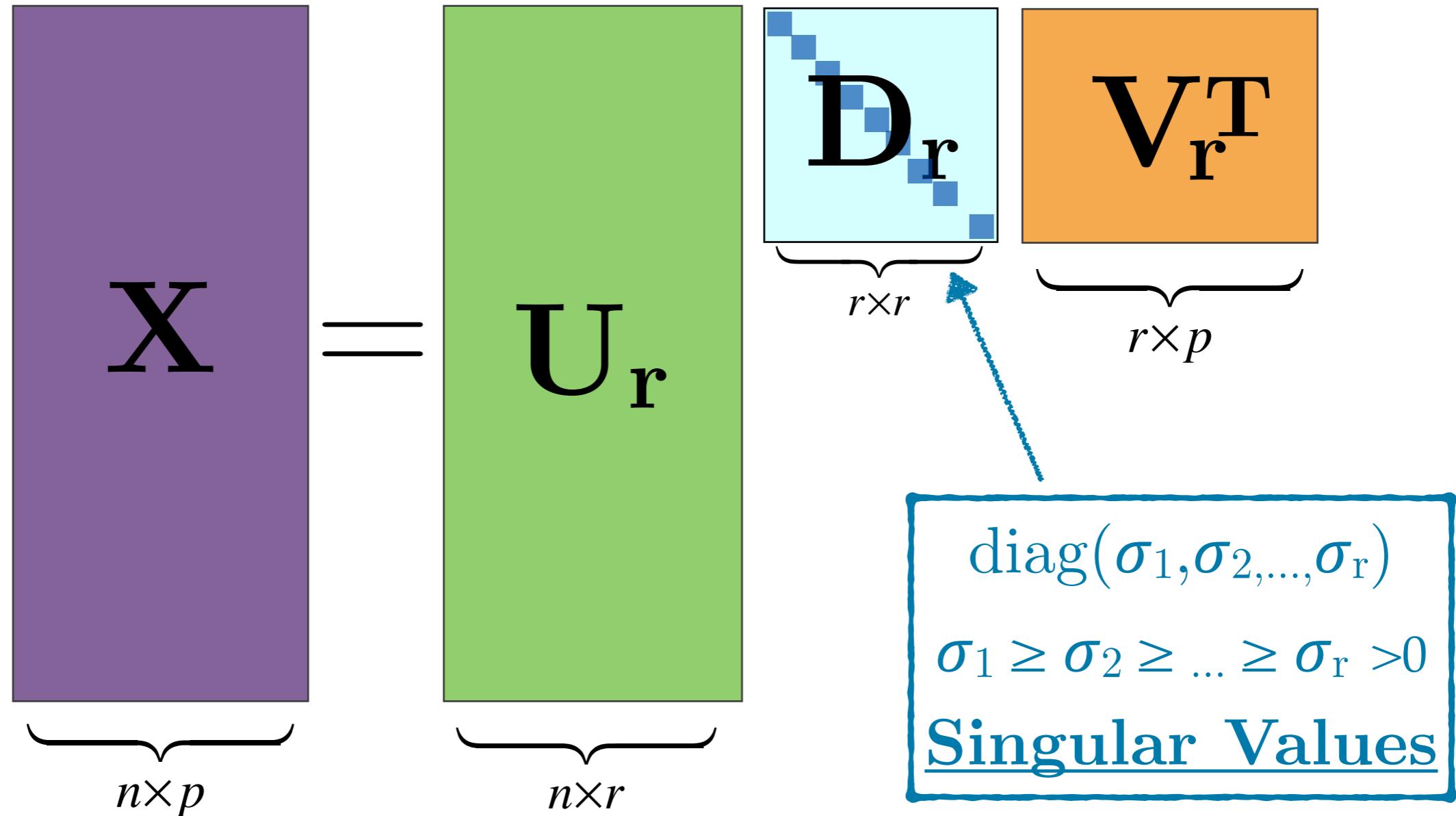
# Skinny SVD

$$\mathbf{X} = \mathbf{U}_r \mathbf{D}_r \mathbf{V}_r^T$$

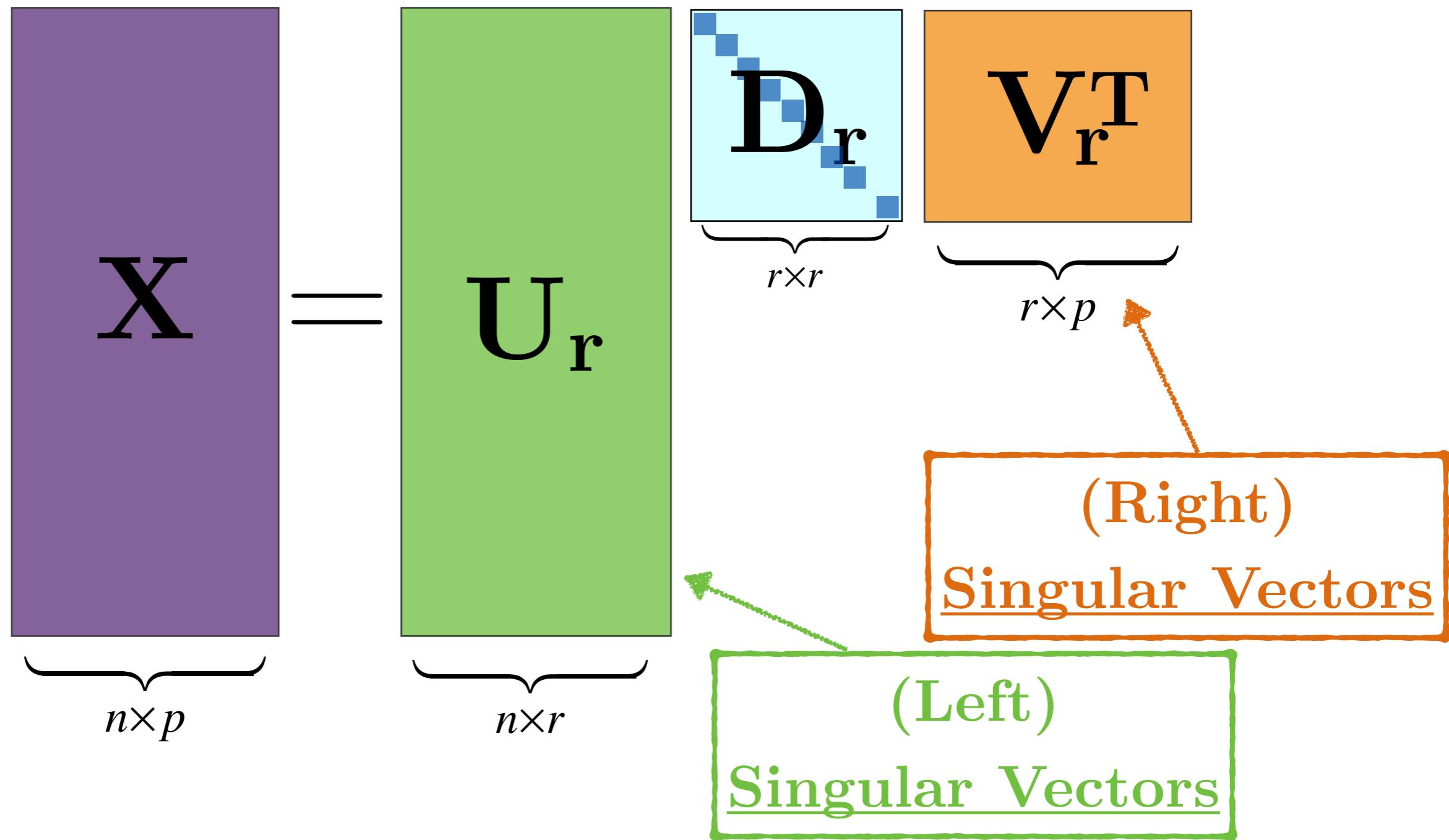
The diagram illustrates the Skinny SVD decomposition of matrix  $\mathbf{X}$ . Matrix  $\mathbf{X}$  is shown as a purple rectangle labeled  $n \times p$ . It is equated to the product of three matrices:  $\mathbf{U}_r$  (green rectangle,  $n \times r$ ),  $\mathbf{D}_r$  (blue square,  $r \times r$ ), and  $\mathbf{V}_r^T$  (orange rectangle,  $r \times p$ ). The  $\mathbf{D}_r$  matrix is depicted with blue squares in the diagonal and zeros elsewhere.

Typically  $r=p$   
because our matrix  
should be full rank!  
No Perfect Multicollinearity

# Skinny SVD



# Skinny SVD



# SVD Fun Facts

- Right singular vectors (rows of  $\mathbf{V}^T$ ) are the (orthonormal) eigenvectors of  $\mathbf{X}^T \mathbf{X}$
- Left singular vectors (columns of  $\mathbf{U}$ ) are the (orthonormal) eigenvectors of  $\mathbf{X} \mathbf{X}^T$
- Singular values are the square roots of the eigenvalues.  
 $(\mathbf{X} \mathbf{X}^T \text{ and } \mathbf{X}^T \mathbf{X} \text{ have the same eigenvalues.})$

# SVD Fun Facts

- ▶ Right singular vectors (rows of  $\mathbf{V}^T$ ) are the (orthonormal) eigenvectors of  $\mathbf{X}^T\mathbf{X}$   
If  $\mathbf{X}$  contains centered/standardized data then  $\mathbf{X}^T\mathbf{X}$  is the covariance/correlation matrix and the singular vectors are principal components! It's PCA!
- ▶ Left singular vectors (columns of  $\mathbf{U}$ ) are the (orthonormal) eigenvectors of  $\mathbf{X}\mathbf{X}^T$
- ▶ Singular values are the square roots of the eigenvalues.  
( $\mathbf{X}\mathbf{X}^T$  and  $\mathbf{X}^T\mathbf{X}$  have the same eigenvalues.)

# PCA from SVD

$$\mathbf{X} = \mathbf{U}_r \mathbf{D}_r \mathbf{V}_r^T$$

The diagram illustrates the Singular Value Decomposition (SVD) of a matrix  $\mathbf{X}$ . The matrix  $\mathbf{X}$  is shown as a purple rectangle labeled  $\mathbf{X}$  with dimensions  $n \times p$  indicated by a brace at the bottom. To its right is an equals sign. Next is a green rectangle labeled  $\mathbf{U}_r$  with dimensions  $n \times r$  indicated by a brace at the bottom. To its right is a blue square labeled  $\mathbf{D}_r$ , which is described as an  $r \times r$  diagonal matrix with blue squares on the diagonal. To the right of  $\mathbf{D}_r$  is an orange rectangle labeled  $\mathbf{V}_r^T$  with dimensions  $r \times p$  indicated by a brace at the bottom.

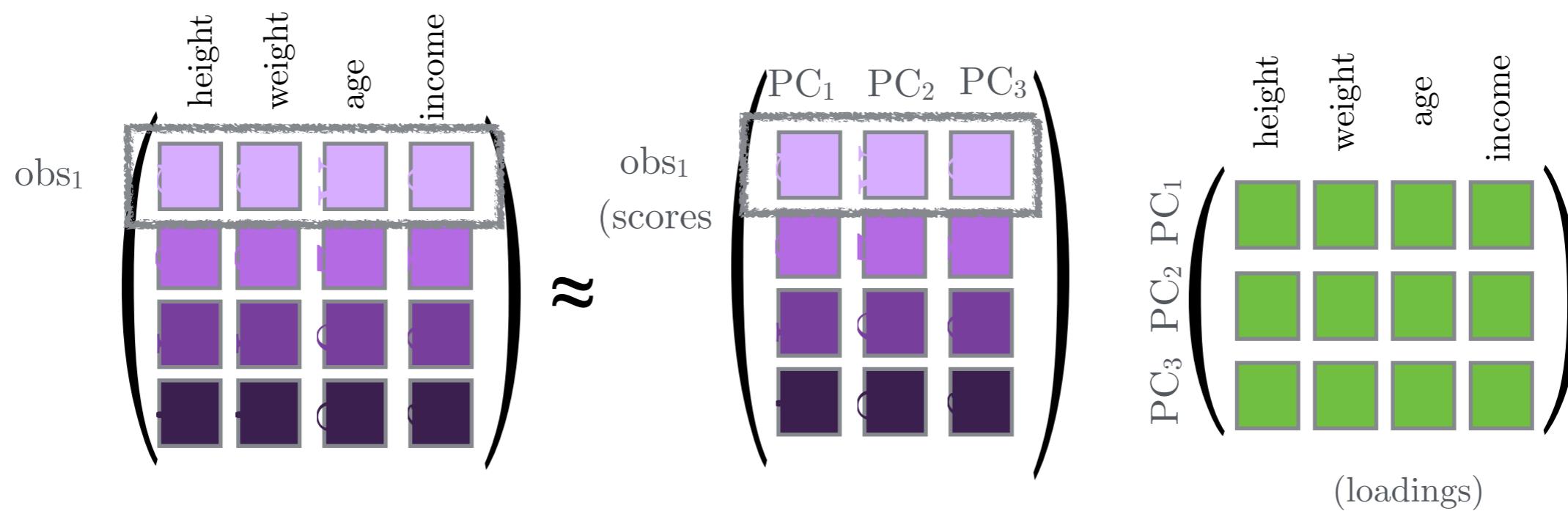
# PCA from SVD

$$\mathbf{X} = \underbrace{\begin{pmatrix} \mathbf{U}_r \\ \mathbf{D}_r \\ \mathbf{V}_r^T \end{pmatrix}}_{\text{Scores/Coordinates}} \quad \underbrace{\begin{pmatrix} \mathbf{U}_r & \mathbf{D}_r & \mathbf{V}_r^T \end{pmatrix}}_{\text{Factors/PCs/Components/Basis Vectors}}$$

# Slide Flashback

## (Factor Analysis Lecture)

$$\mathbf{X} = \mathbf{S}\mathbf{V}\mathbf{T}$$



# Slide Flashback

## (PCA Lecture)

### Coordinates in the New Basis

Original Data  
(Centered/Standardized)

$$\begin{pmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \mathbf{x}_3 & \mathbf{x}_4 & \mathbf{x}_5 \end{pmatrix}$$

Loadings

$$\begin{pmatrix} & & \\ & & \\ & & \\ & & \\ & & \end{pmatrix}$$

$\mathbf{Prin1} \quad \mathbf{Prin2} \quad \mathbf{Prin3}$

$$= \begin{pmatrix} & & \\ & & \\ & & \end{pmatrix}$$

$\boxed{\mathbf{XV} = \mathbf{S}}$

# Slide Flashback

## (Orthogonality Lecture)

### Orthogonal Matrix

- An orthogonal matrix is easy to maneuver inside matrix equations, since  $\mathbf{V}^{-1} = \mathbf{V}^T$
- For example if  $\mathbf{U}$  and  $\mathbf{V}$  are orthogonal, the following equations are equivalent:

$$\mathbf{X}\mathbf{V} = \mathbf{UD}$$

$$\mathbf{X} = \mathbf{UDV}^T$$

$$\mathbf{U}^T\mathbf{X} = \mathbf{DV}^T$$

$$\mathbf{U}^T\mathbf{X}\mathbf{V} = \mathbf{D}$$

# What's the point

- ▶ PCA IS the SVD on centered or standardized data.
- ▶ Sometimes, practitioners opt for the regular uncentered SVD rather than PCA.
  - ▶ True especially in genomics/text/image analysis

# Dimension Reduction



# Noise Reduction

# Resolving a Matrix into Components

Let  $\mathbf{U}_r = [\mathbf{U}_1 | \mathbf{U}_2 | \dots | \mathbf{U}_r]$  and  $\mathbf{V}_r^T =$   
(the left and right singular vectors)

$$\begin{bmatrix} \mathbf{V}_1^T \\ \mathbf{V}_2^T \\ \vdots \\ \mathbf{V}_r^T \end{bmatrix}$$

Then,

$$\begin{aligned}\mathbf{X} &= \mathbf{U}_r \mathbf{D}_r \mathbf{V}_r^T = \sum_{i=1}^r \sigma_i \mathbf{U}_i \mathbf{V}_i^T \\ &= \sigma_1 \mathbf{U}_1 \mathbf{V}_1^T + \sigma_2 \mathbf{U}_2 \mathbf{V}_2^T + \sigma_3 \mathbf{U}_3 \mathbf{V}_3^T + \dots + \sigma_r \mathbf{U}_r \mathbf{V}_r^T\end{aligned}$$

It's 'just' matrix multiplication - sum is visualized next slide

# Resolving a Matrix into Components

$$\mathbf{X} = \sum_{i=1}^r \sigma_i \mathbf{U}_i \mathbf{V}_i^T$$

The diagram illustrates the Singular Value Decomposition (SVD) of a matrix  $\mathbf{X}$ .  
- The matrix  $\mathbf{X}$  is shown as a purple rectangle labeled  $\mathbf{X}$  with dimensions  $n \times p$  indicated by a brace at the bottom.  
- The decomposition is represented by the equation  $\mathbf{X} = \sum_{i=1}^r \sigma_i \mathbf{U}_i \mathbf{V}_i^T$ .  
- The term  $\sum_{i=1}^r$  is shown with a large black sigma symbol, where  $r$  is the rank of the matrix.  
- The singular values  $\sigma_i$  are represented by blue squares.  
- The matrix  $\mathbf{U}_i$  is shown as a green vertical rectangle labeled  $\mathbf{U}_i$  with dimensions  $n \times 1$  indicated by a brace at the bottom.  
- The matrix  $\mathbf{V}_i^T$  is shown as an orange horizontal rectangle labeled  $\mathbf{V}_i^T$  with dimensions  $1 \times p$  indicated by a brace at the bottom.

(sum of rank 1 matrices)

# Signal-to-Noise Ratio

and

## Noise Reduction

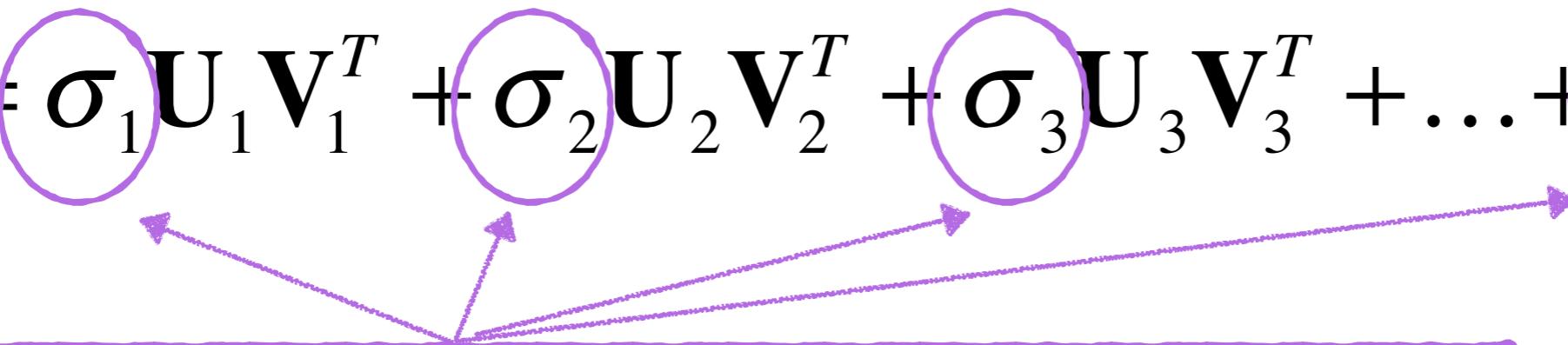
$$\mathbf{X} = \sigma_1 \mathbf{U}_1 \mathbf{V}_1^T + \sigma_2 \mathbf{U}_2 \mathbf{V}_2^T + \sigma_3 \mathbf{U}_3 \mathbf{V}_3^T + \dots + \sigma_r \mathbf{U}_r \mathbf{V}_r^T$$

Think of these as  
“unit basis directions”  
for the matrix  $\mathbf{X}$ .

# Signal-to-Noise Ratio

and

# Noise Reduction

$$\mathbf{X} = \sigma_1 \mathbf{U}_1 \mathbf{V}_1^T + \sigma_2 \mathbf{U}_2 \mathbf{V}_2^T + \sigma_3 \mathbf{U}_3 \mathbf{V}_3^T + \dots + \sigma_r \mathbf{U}_r \mathbf{V}_r^T$$


Think of these as coordinates that say how much “signal” or information of the matrix  $\mathbf{X}$  is directed along each basis direction.

The components are ordered by the magnitude of the signal.

$$\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r$$

# Signal-to-Noise Ratio

and

## Noise Reduction

$$\mathbf{X} = \sigma_1 \mathbf{U}_1 \mathbf{V}_1^T + \sigma_2 \mathbf{U}_2 \mathbf{V}_2^T + \sigma_3 \mathbf{U}_3 \mathbf{V}_3^T + \dots + \sigma_r \mathbf{U}_r \mathbf{V}_r^T$$

Anytime we have signal, we inevitably have some noise.

Our data is typically an imperfect depiction of reality.

# Signal-to-Noise Ratio

and

## Noise Reduction

$$\mathbf{X} = \sigma_1 \mathbf{U}_1 \mathbf{V}_1^T + \sigma_2 \mathbf{U}_2 \mathbf{V}_2^T + \sigma_3 \mathbf{U}_3 \mathbf{V}_3^T + \dots + \sigma_r \mathbf{U}_r \mathbf{V}_r^T$$

If we assume there is no pattern to the noise -  
That it is uniformly distributed “in every direction”

Then amount of noise in each of the terms in this sum is the same!

# Signal-to-Noise Ratio

and

## Noise Reduction

$$\mathbf{X} = \sigma_1 \mathbf{U}_1 \mathbf{V}_1^T + \sigma_2 \mathbf{U}_2 \mathbf{V}_2^T + \sigma_3 \mathbf{U}_3 \mathbf{V}_3^T + \dots + \sigma_r \mathbf{U}_r \mathbf{V}_r^T$$

The amount of signal in each of the terms in this sum is decreasing →

The amount of noise in each of the terms in this sum is the same.



The signal-to-noise ratio is higher in first terms.  
Last terms could be mostly noise

# Signal-to-Noise Ratio

and

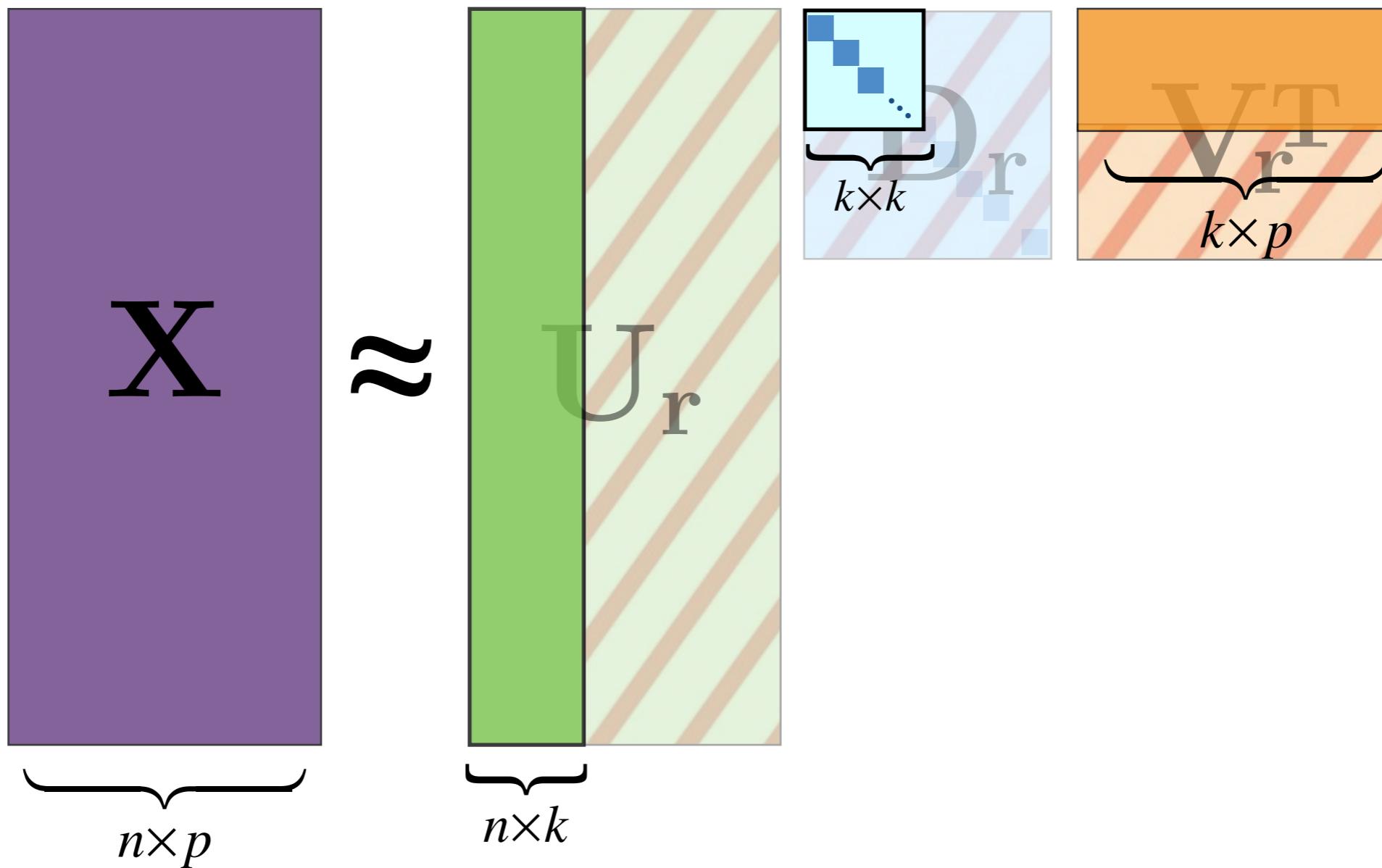
# Noise Reduction

$$\mathbf{X} = \sigma_1 \mathbf{U}_1 \mathbf{V}_1^T + \sigma_2 \mathbf{U}_2 \mathbf{V}_2^T + \sigma_3 \mathbf{U}_3 \mathbf{V}_3^T + \dots + \sigma_r \mathbf{U}_r \mathbf{V}_r^T$$


If the last terms have more noise, then we won't lose much information by omitting them, AND we may actually lose a good bit of noise. That's a perk of dimension reduction.

# Truncated SVD

$$\mathbf{X} = \sigma_1 \mathbf{U}_1 \mathbf{V}_1^T + \sigma_2 \mathbf{U}_2 \mathbf{V}_2^T + \sigma_3 \mathbf{U}_3 \mathbf{V}_3^T + \dots + \sigma_k \mathbf{U}_k \mathbf{V}_k^T$$



# Truncated SVD

$$\mathbf{X} = \sigma_1 \mathbf{U}_1 \mathbf{V}_1^T + \sigma_2 \mathbf{U}_2 \mathbf{V}_2^T + \sigma_3 \mathbf{U}_3 \mathbf{V}_3^T + \dots + \sigma_k \mathbf{U}_k \mathbf{V}_k^T$$

$k \leq r$

The diagram illustrates the truncated Singular Value Decomposition (SVD) of a matrix  $\mathbf{X}$ . The matrix  $\mathbf{X}$  is represented as a purple rectangle labeled  $n \times p$ . It is approximated by the equation  $\mathbf{X} \approx \sum_{k=1}^{r-k} \sigma_k \mathbf{U}_k \mathbf{V}_k^T$ , where  $r$  is the rank of  $\mathbf{X}$ .

The approximation consists of three main components:

- A green vertical rectangle representing the low-rank approximation, labeled  $n \times k$ .
- A blue square representing the singular values, labeled  $k \times k$ . It contains a diagonal of blue squares.
- An orange horizontal rectangle representing the error term, labeled  $k \times p$ .

Red circles highlight the  $k$ -th singular value and its corresponding columns in  $\mathbf{U}$  and  $\mathbf{V}$ .

# The SVD of Dr. Rappa

# Our Fearless Leader



# Let's start in B&W



Imagine this is  
your data matrix.

# Let's start in B&W



Each pixel represents  
a number between  
0 and 1.  
0=black 1=white

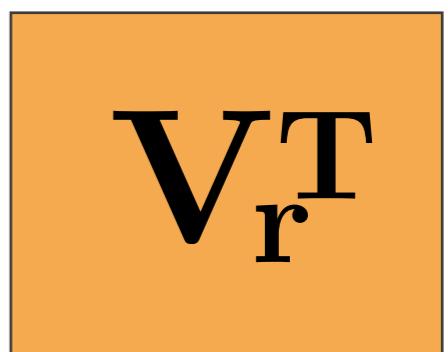
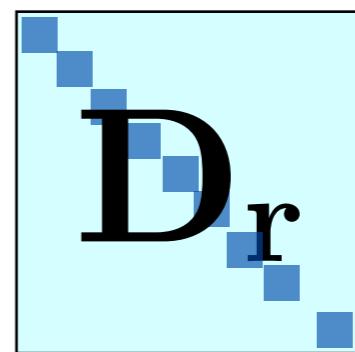
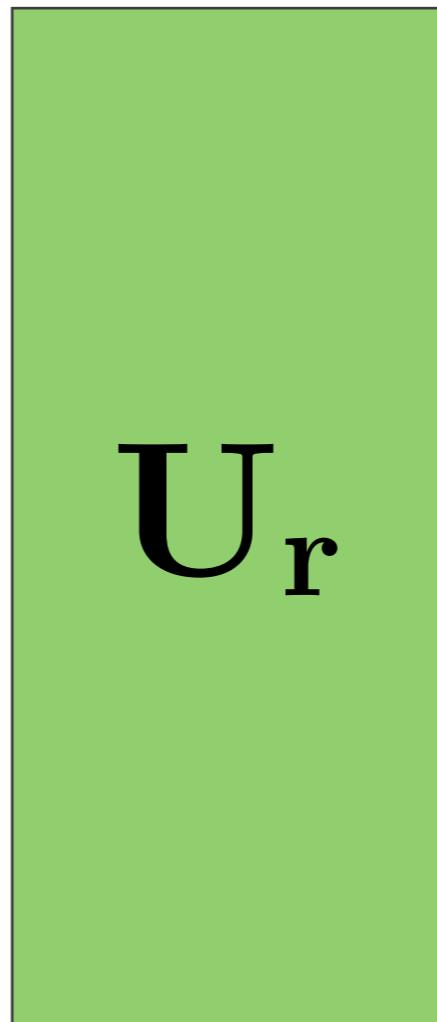
The matrix is  
160 x 250, and is  
called **rappa.grey**

# Take the SVD of the matrix

```
rappaSvd=svd(rappa.grey)  
U=rappaSvd$u  
d=rappaSvd$d  
Vt=t(rappaSvd$v)
```



=

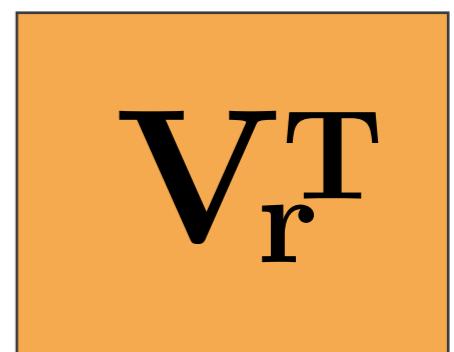
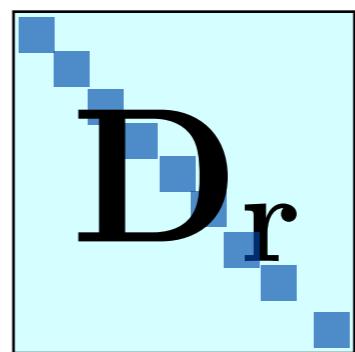
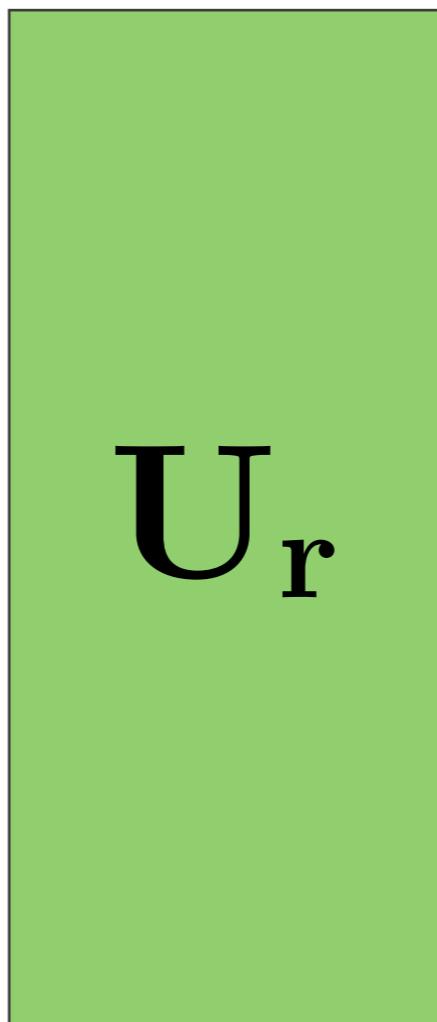


# Take the SVD of the matrix

```
rappasvd=svd(rappa.grey)  
U=rappasvd$u  
d=rappasvd$d  
Vt=t(rappasvd$v)
```



=



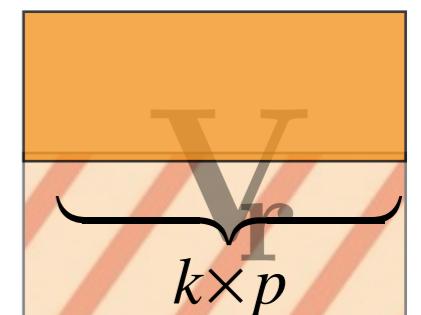
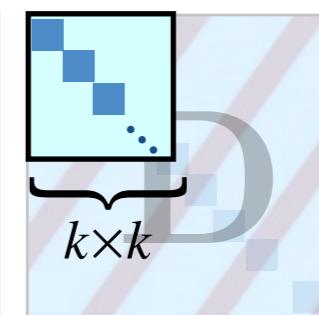
rappa.grey	num [1:250, 1]
rappasvd	Large list (3)
d:	num [1:160] 114.8 18.7 14.1 1
u:	num [1:250, 1:160] -0.107 -0.
v:	num [1:160, 1:160] -0.135 -0.
U	num [1:250, 1]
Vt	num [1:160, 1]
Values	
d	num [1:160] 1

# rank k approximations

```
RappaRank_k = U[,1:k] %*% diag(d[1:k]) %*% Vt[1:k,]  
image(RappaRank_k,  
      col=grey((0:1000)/1000),  
      main=paste(k,"dimensions"),  
      xaxt = 'n',  
      yaxt = 'n')
```



$\approx$

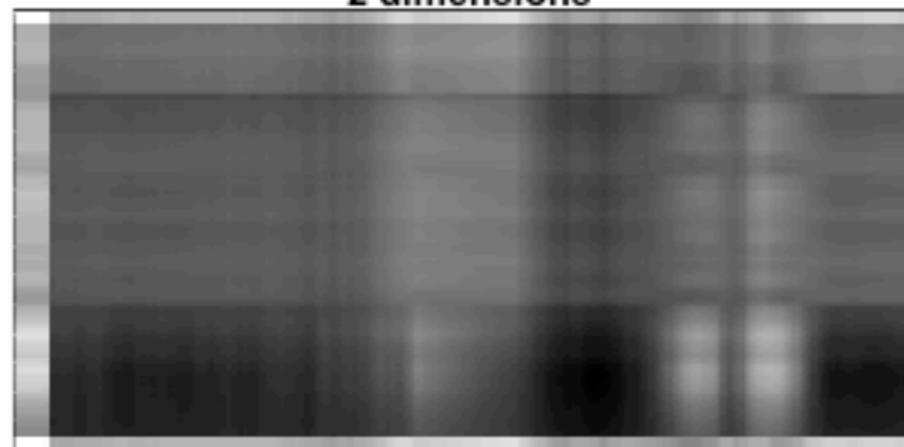


# rank k approximations

1 dimension



2 dimensions



3 dimensions



4 dimensions



5 dimensions



6 dimensions



7 dimensions



8 dimensions



9 dimensions



# rank k approximations

10 dimensions



20 dimensions



30 dimensions



40 dimensions



50 dimensions



60 dimensions



70 dimensions



80 dimensions

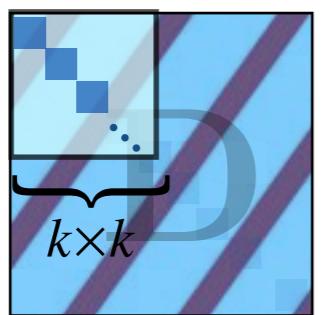
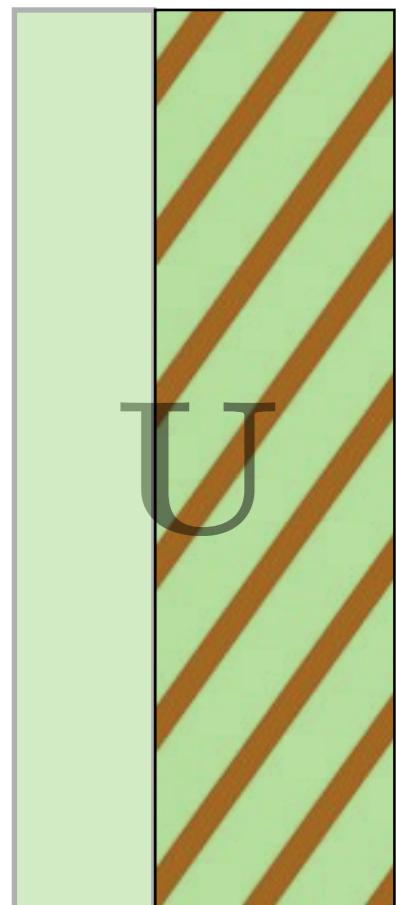


90 dimensions



# What about the dropped components?

```
RappaRank_n = U[,n:160] %*% diag(d[n:160]) %*% Vt[n:160,]  
image(RappaRank_n,  
      col=grey((0:1000)/1000),  
      main=paste("last", (160-n),"dimensions"),  
      xaxt = 'n',  
      yaxt = 'n')
```



$U$

$$X = \sigma_1 U_1 V_1^T + \sigma_2 U_2 V_2^T + \sigma_3 U_3 V_3^T + \dots + \sigma_r U_r V_r^T$$

what did we lose?

# What about the dropped components?

1 dimension



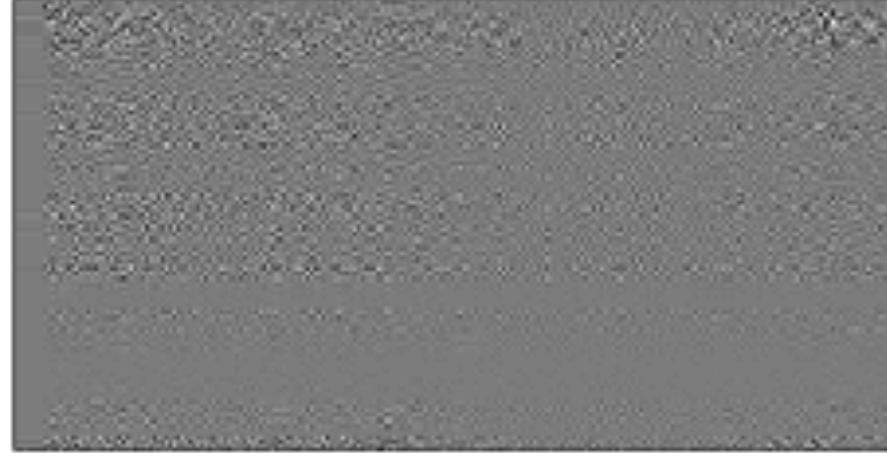
last 10 dimensions



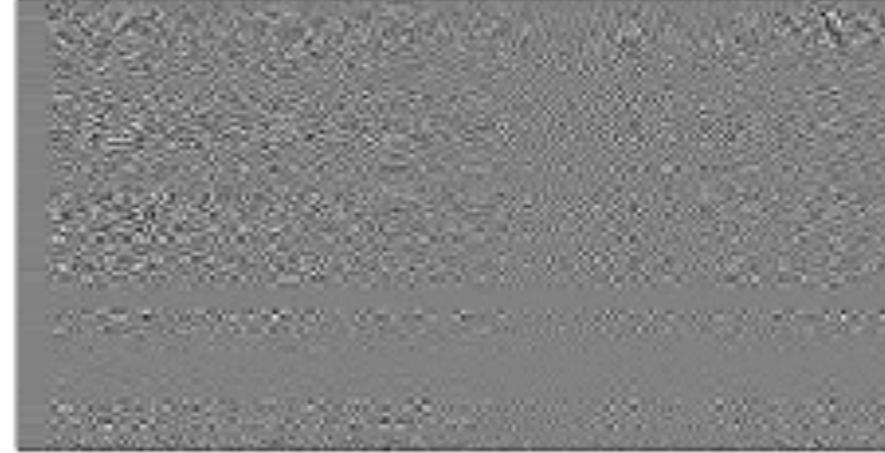
last 20 dimensions



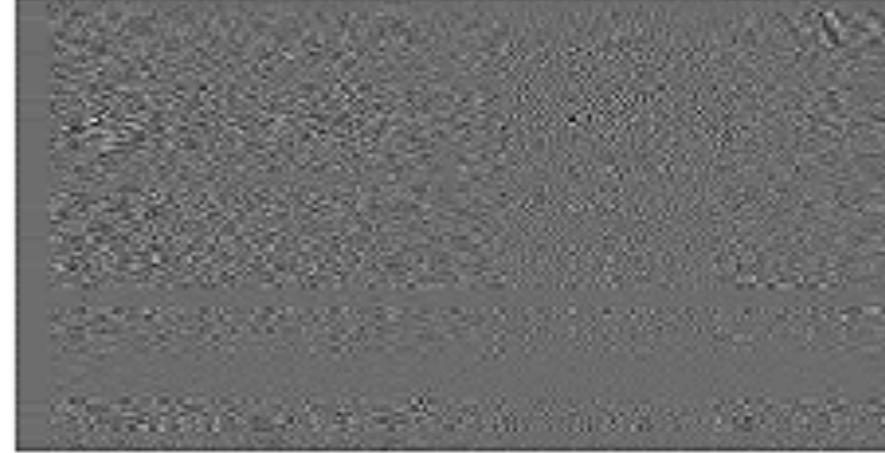
last 30 dimensions



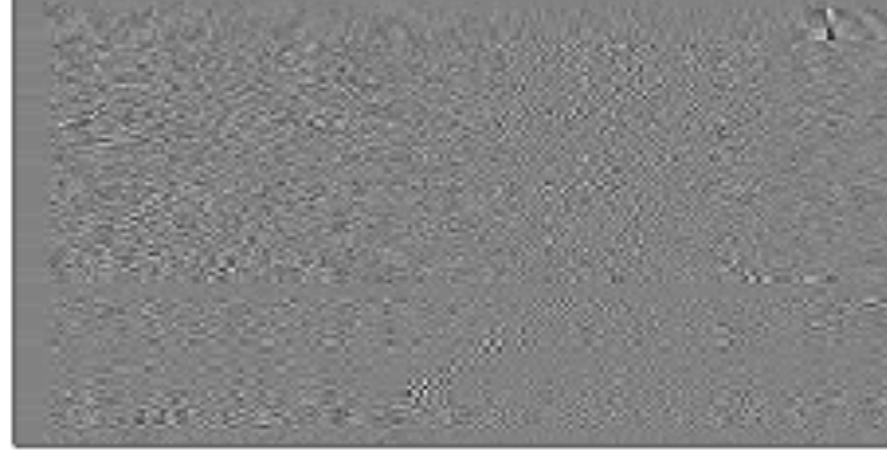
last 40 dimensions



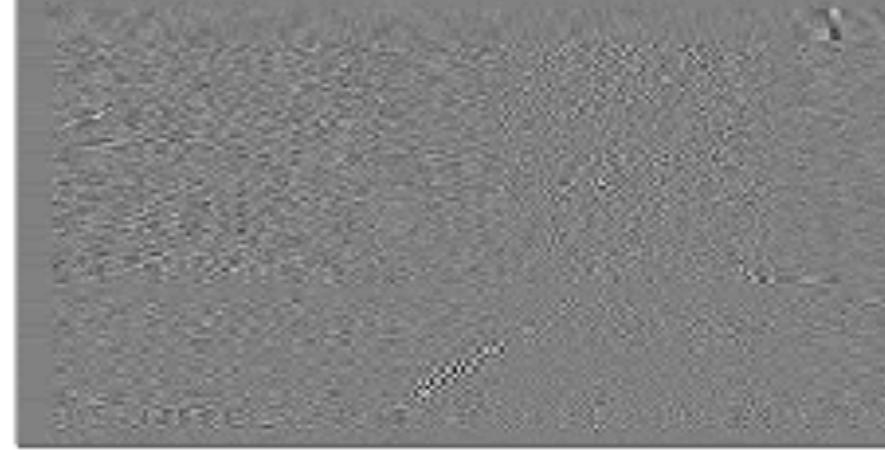
last 50 dimensions



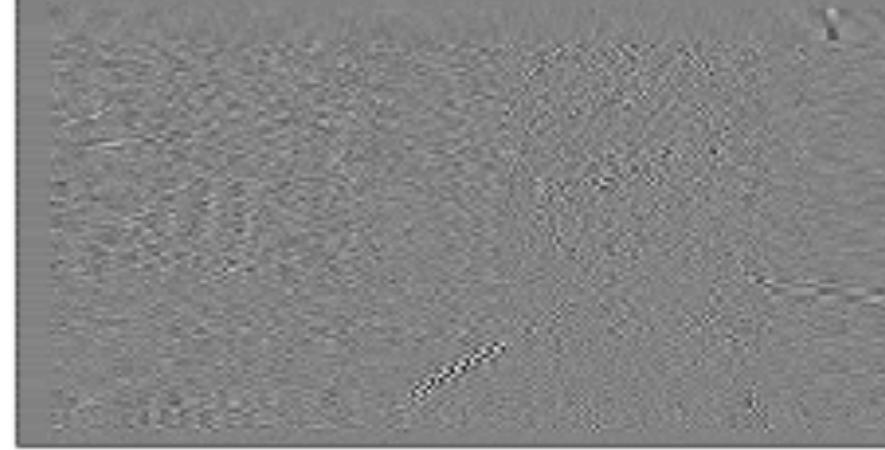
last 60 dimensions



last 70 dimensions



last 80 dimensions

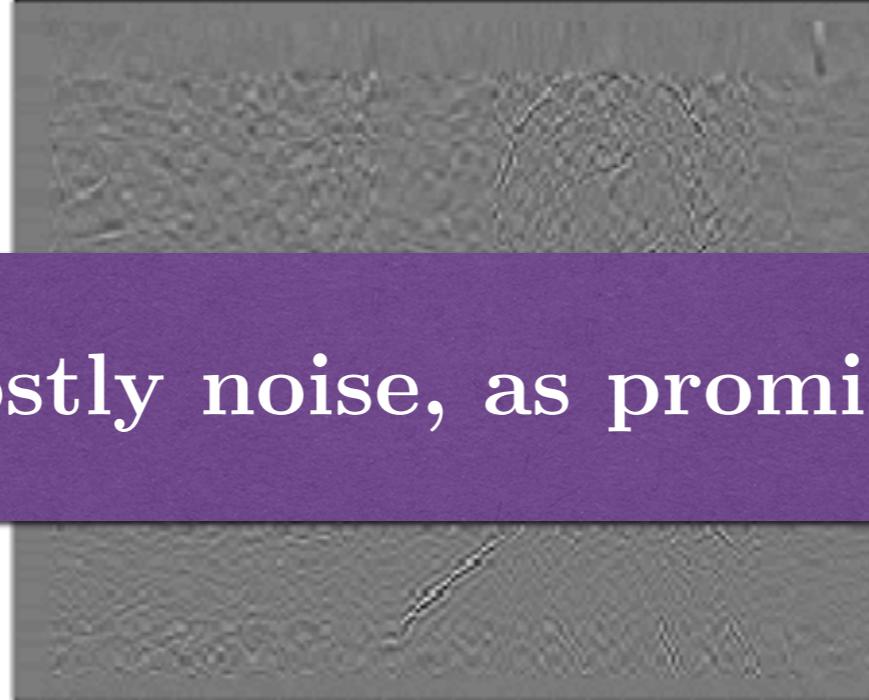


# What about the dropped components?

last 90 dimensions



last 100 dimensions

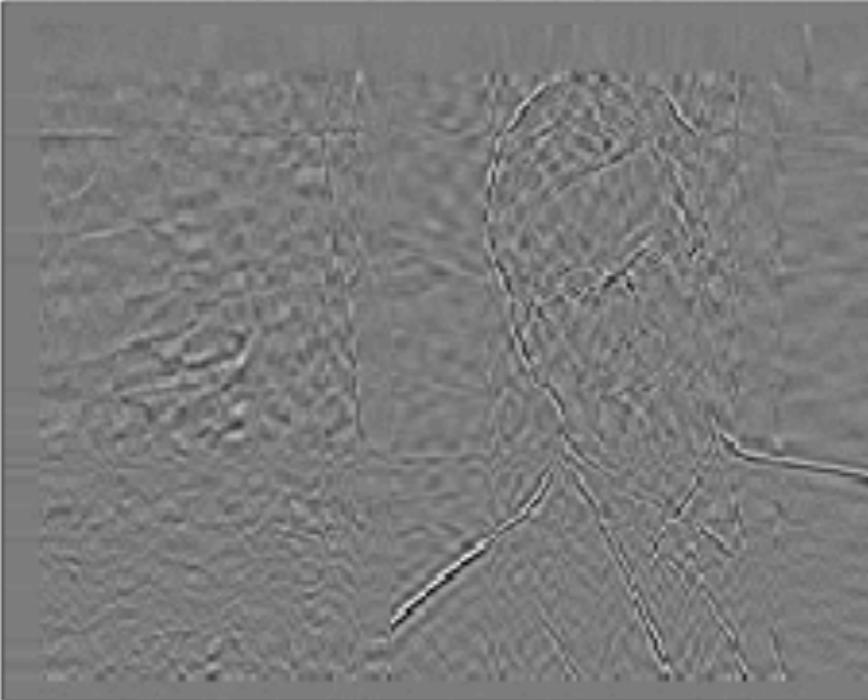


last 110 dimensions



Mostly noise, as promised!

last 120 dimensions



last 130 dimensions



last 140 dimensions



# What's the point?

- ▶ Orthogonal projections built on the theory of maximal variance don't tend to lie in the story they tell in the first few dimensions.
- ▶ Additional components can certainly help *resolve* the story - adding detail and clarity - but the theme remains the same.
- ▶ When you're reducing dimensionality of datasets, use the visual of 9-dimensional Dr. Rappa as an analogy to what you're seeing in the projection.