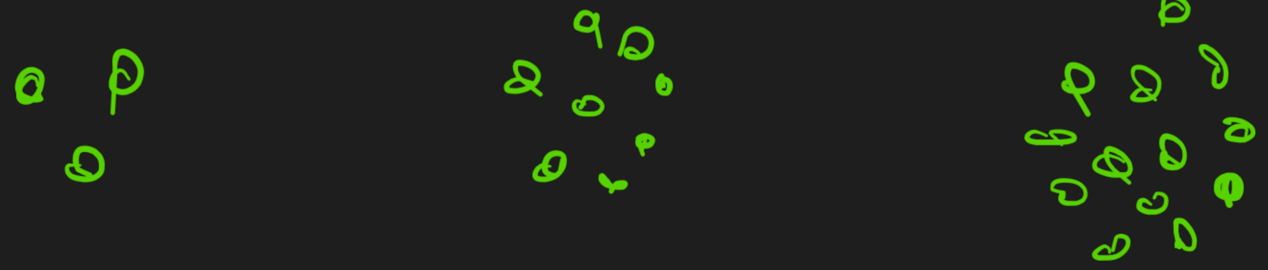
INTRODUCTION TO MACHINE LEARNING COMPSCI 4ML3

Lecture 8

Hassan Ashtiani

HIGH-DIMENSIONAL DATA VISUALIZATION

- We have n data points
  - Each d dimensional
- How can we visualize $\boxed{X_{n \times d}?}$
  - For now assume there is no Y value
- If $d = 1$ or $d = 2$ (or maybe $d = 3$)?

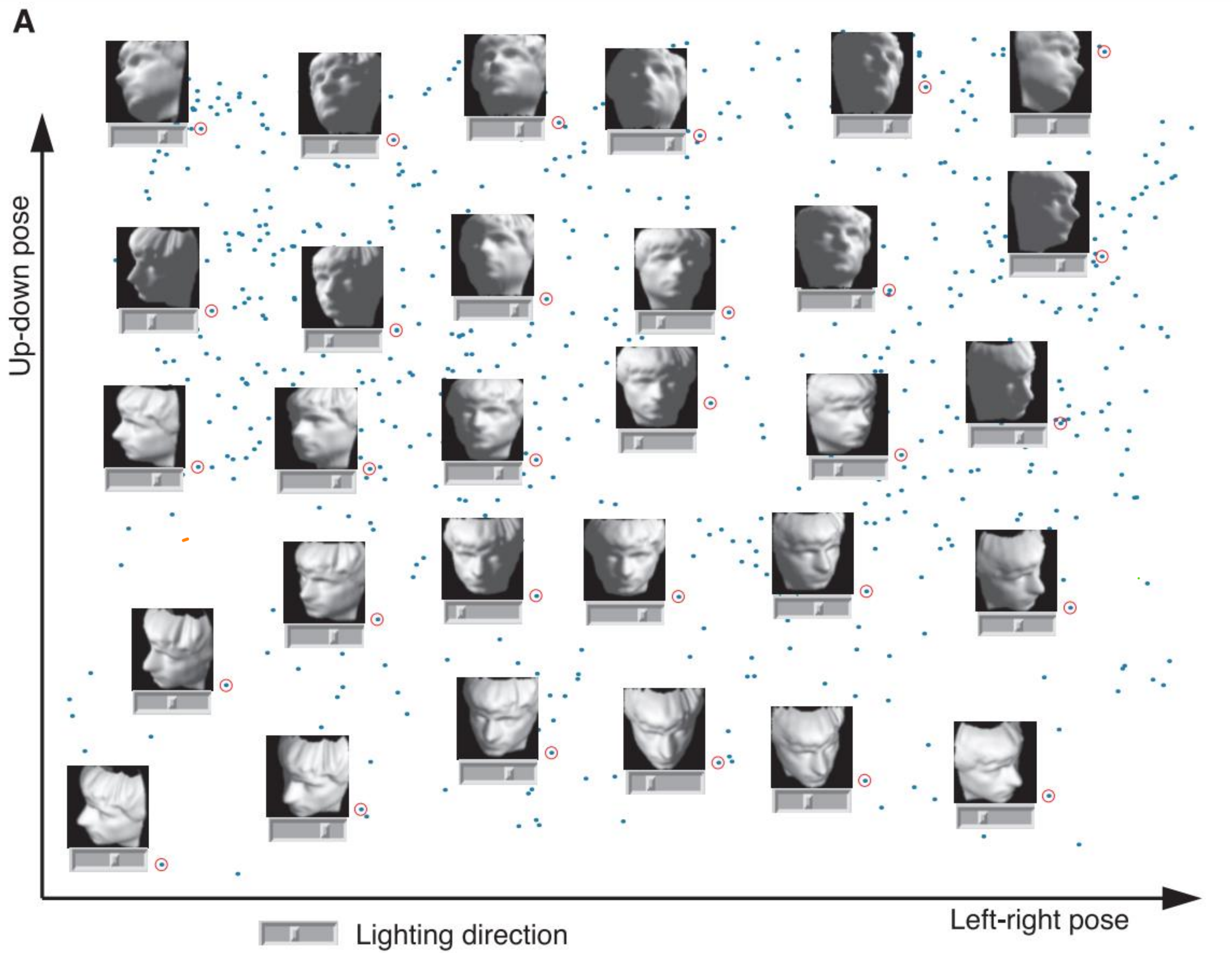


HIGH-DIMENSIONAL DATA

SAY YOU HAVE A DATA SET OF IMAGES

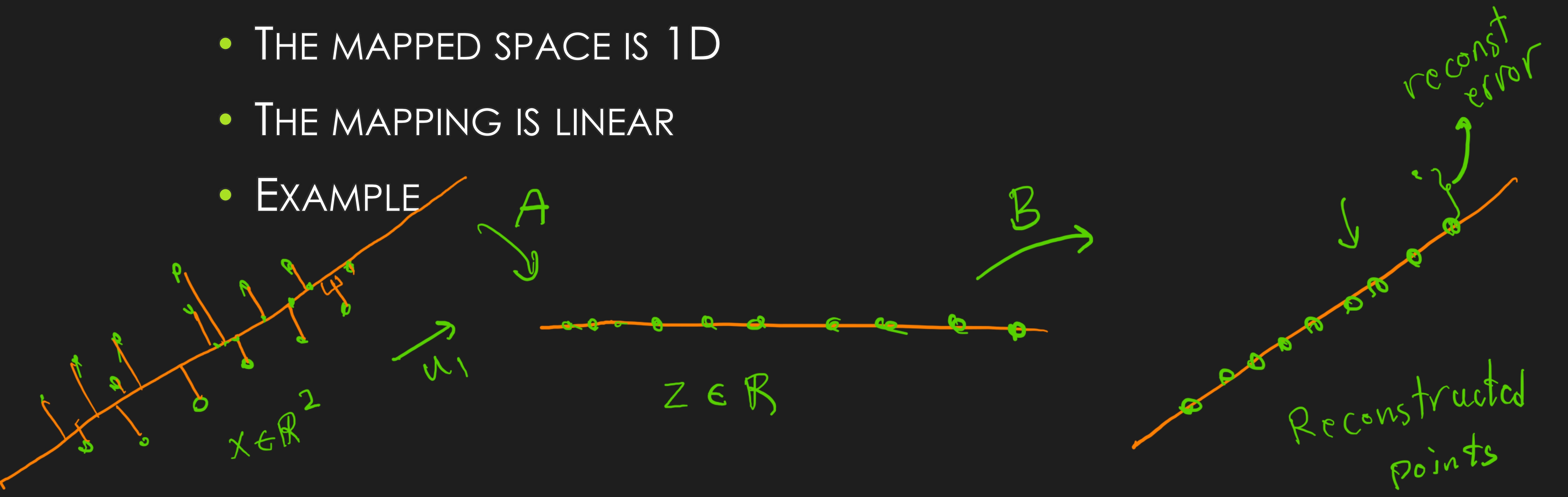
HOW TO VISUALIZE?

- MAP THE DATA SET TO A LOW DIMENSIONAL SPACE
  - OPPOSITE OF WHAT WE DID FOR NON-LINEAR CURVE-FITTING!
  - HOW TO MAP THE DATA?

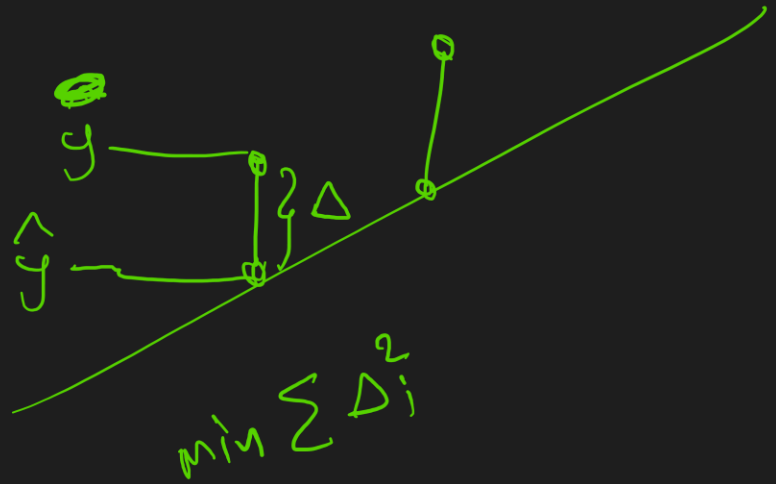

A
Up-down pose
Left-right pose
Lighting direction

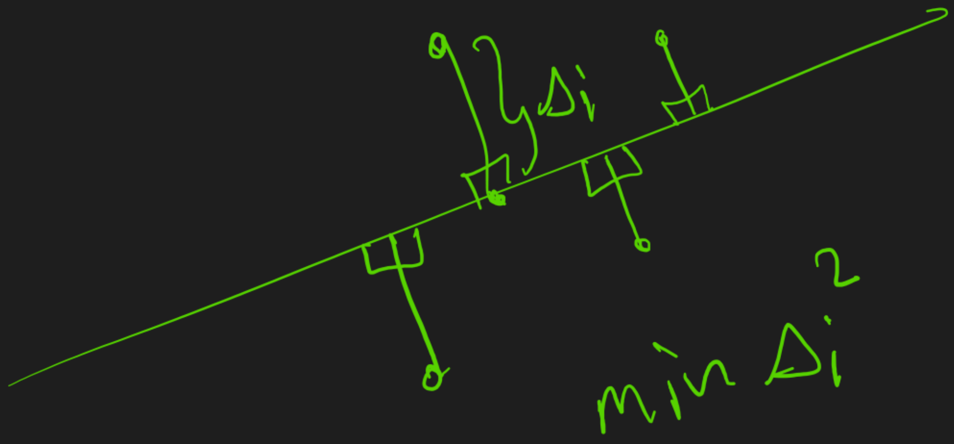
FINDING A GOOD MAPPING

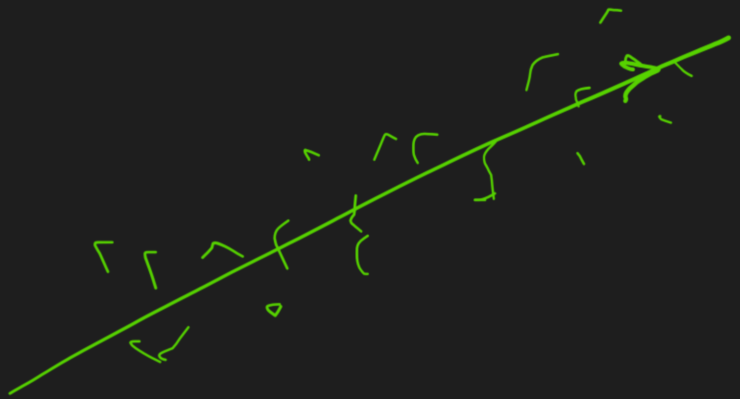
- Simple case:
  - The original space is 2D
  - The mapped space is 1D
  - The mapping is linear
  - Example



IN CONTRAST TO LS? (EXAMPLE)







PROBLEM FORMULATION

$Ax + \cancel{b}$

- Map $x \in R^d$ to $z \in R^q$ with $q < d$
- A $q \times d$ matrix can represent a linear mappting:

$$\boxed{z = Ax} , \quad z_j = \langle u_j , x \rangle$$

$q \times d$

  - Homogeneous mapping
  - We assume $AA^T = I$ (why?) $\quad \langle u_i , u_j \rangle = 0$
  - But $A^T A \neq I$! (why?) $\quad \mathrm{Rank}(A^T A) \leq q \quad \boxed{A^T A : d \times d}$
- Which mapping A is "good"? How to find it?

$$A = \begin{bmatrix} u_1^T \\ \vdots \\ u_q^T \end{bmatrix}$$

RECONSTRUCTING DATA

- A linear mapping to a lower dimensional space may inevitably lose some information
- Try to ``reconstruct the data''
  - Measure the reconstruction error
- Example

MINIMIZING THE RECONSTRUCTION ERROR

- $\min_{A,B} \sum_i \left\| x^i - BAx^i \right\|_2^2$

WHERE

- A IS $q \times d$
- B IS $d \times q$
- $AA^T = I_{q \times q}$

- EXAMPLE: WHAT IF $q = d$?

$A \cdot B = I$, error $= 0$

$\cancel{B = A^{-1}}$

MINIMIZING THE RECONSTRUCTION ERROR

- $\min_{A,B} \sum_i \|x^i - BAx^i\|_2^2$
  - S.T. $AA^T = I_{q \times q}$
- GIVEN A, WHAT IS THE BEST INVERSE MAPPING?
  - $B = A^T$

[show?]

SOLUTION TO THE MINIMIZATION PROBLEM

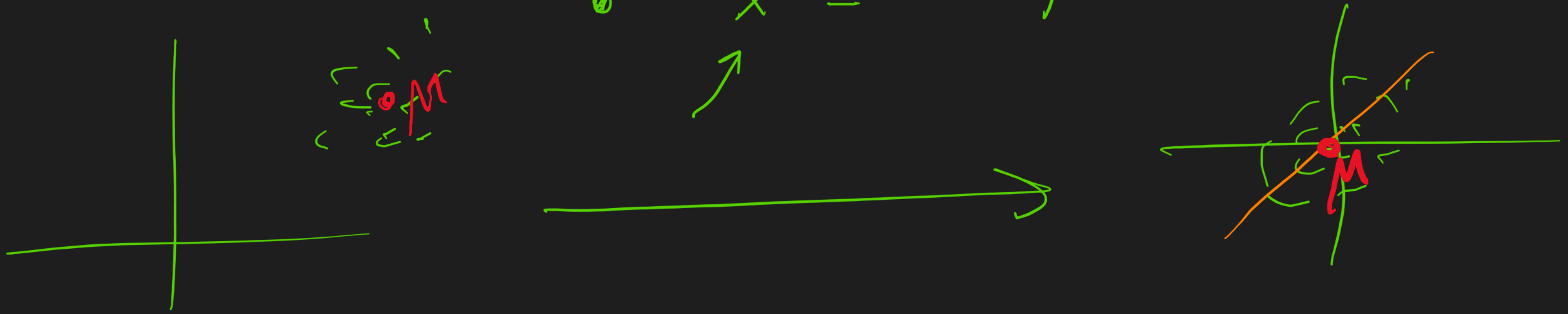
- $B = A^T$
- $\min_A \sum_i \left\| x^i - A^T A x^i \right\|^2$ SUBJECT TO $AA^T = I_{q \times q}$
- ASSUME THE DATA IS CENTERED: $\frac{1}{\mathrm{n}} \sum_i x^i = [0, \dots, 0]^T$
- WHAT TO DO IF DATA IS NOT CENTERED?

CENTERING THE DATA

$\mu = \frac{1}{n} \sum x^i$

new data set :

$\tilde{x}^i = x^i - \mu$



SOLUTION TO THE MINIMIZATION PROBLEM

- $B = A^T$
- $\min_A \sum_i \left\| x^i - A^T A x^i \right\|^2$ SUBJECT TO $AA^T = I_{q \times q}$
- ASSUME THE DATA IS CENTERED: $\frac{1}{\text{n}} \sum_i x^i = [0, \ldots, 0]^T$
- OPTIMAL MATRIX A CONSISTS OF THE **EIGEN-VECTORS** OF THE COVARIANCE MATRIX $X^T X$ (CORRESPONDING TO THE TOP q EIGENVALUES)

EIGENVALUE DECOMPOSITION

$X^T X$: symmetric real

- EIGENVECTORS AND EIGENVALUES:
  - $X^T X u = \lambda u$
  - $X^T X = U^T \Lambda U$
  - $\Lambda = diag(\lambda_1, \lambda_2, \dots, \lambda_d)$, $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_d$
  - $U^T U = U U^T = I$ → $\langle u_i, u_j \rangle = 0$ when $i \neq j$
  - THE COLUMNS OF U ARE THE EIGENVECTORS

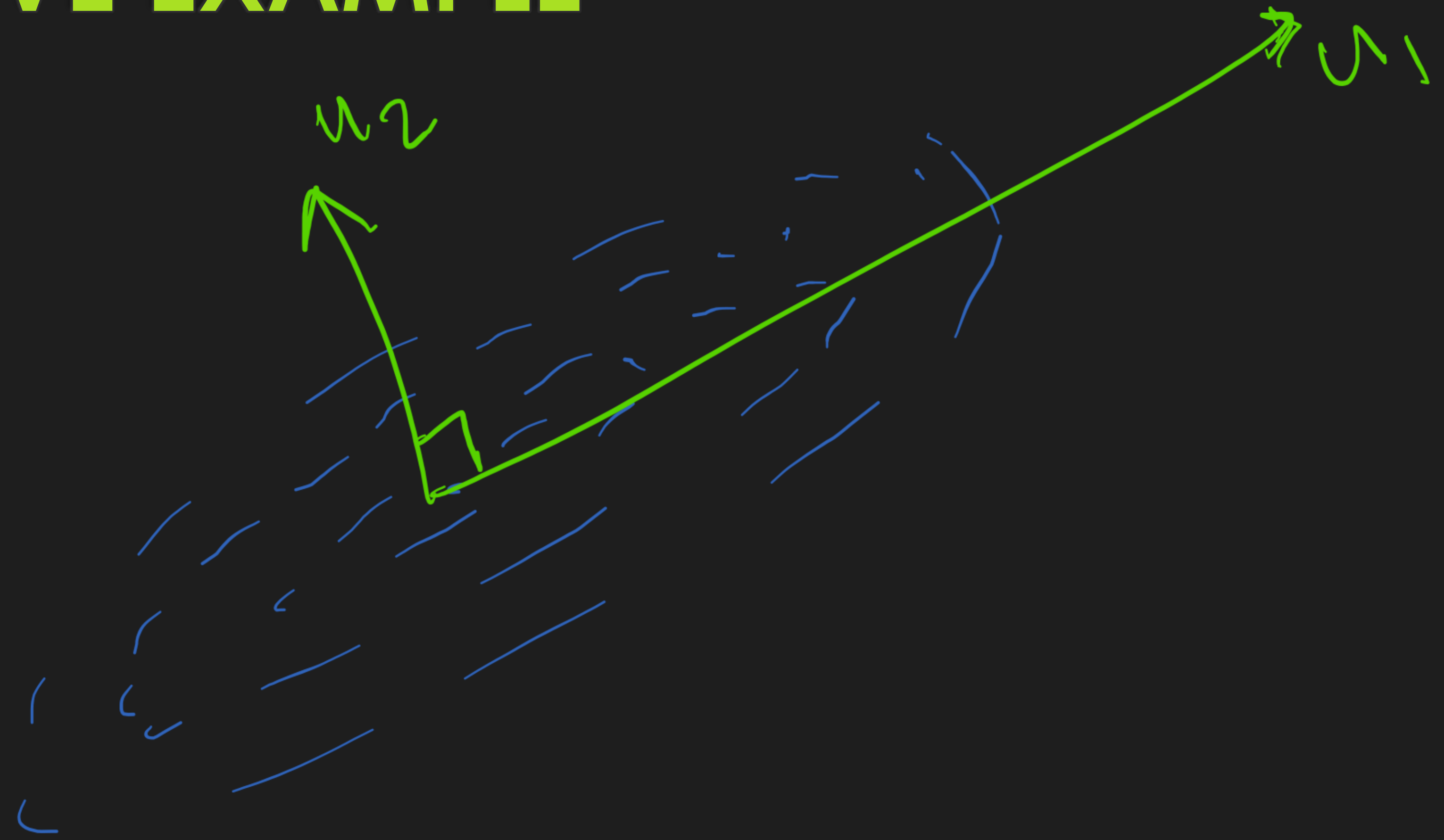
$U = \left[u_1 \middle| u_2 \middle| \cdots \middle| u_d \right]$, $\Lambda = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_d \end{bmatrix}$

PRINCIPAL COMPONENT ANALYSIS (PCA)

- INPUT: $x^1, \dots, x^n \in R^d$
- $\mu = \frac{1}{n} \sum x^i$
- $C = \sum (x^i - \mu)(x^i - \mu)^T$
- FIND EIGEN-VECTORS/VALUES OF C
  - $C = U^{\mathrm{T}} \Lambda \mathrm{U}$
- FIND EIGEN-VECTORS/VALUES OF C
- OPTIMAL MATRIX $A = \begin{bmatrix} u_1^T \\ u_2^T \\ \vdots \\ u_q^T \end{bmatrix}$
-
-

Top q eigen vectors

INTUITIVE EXAMPLE



ALTERNATIVE TO FIND EIGENVECTORS OF C

- $C = \sum(x^i - \mu)(x^i - \mu)^T$ IS $d \times d$
  - HARD TO DEAL WITH WHEN d IS SUPER LARGE
- WHEN $d \gg n$, WE CAN INSTEAD DO SINGULAR VALUE DECOMPOSITION (SVD) ON X
  - ASSUME THE DATA IS CENTERED
  - $X = U\Sigma V^T$
  - $C = V\Sigma U^T U\Sigma V^T = V\Lambda V^T$

$X^T X$

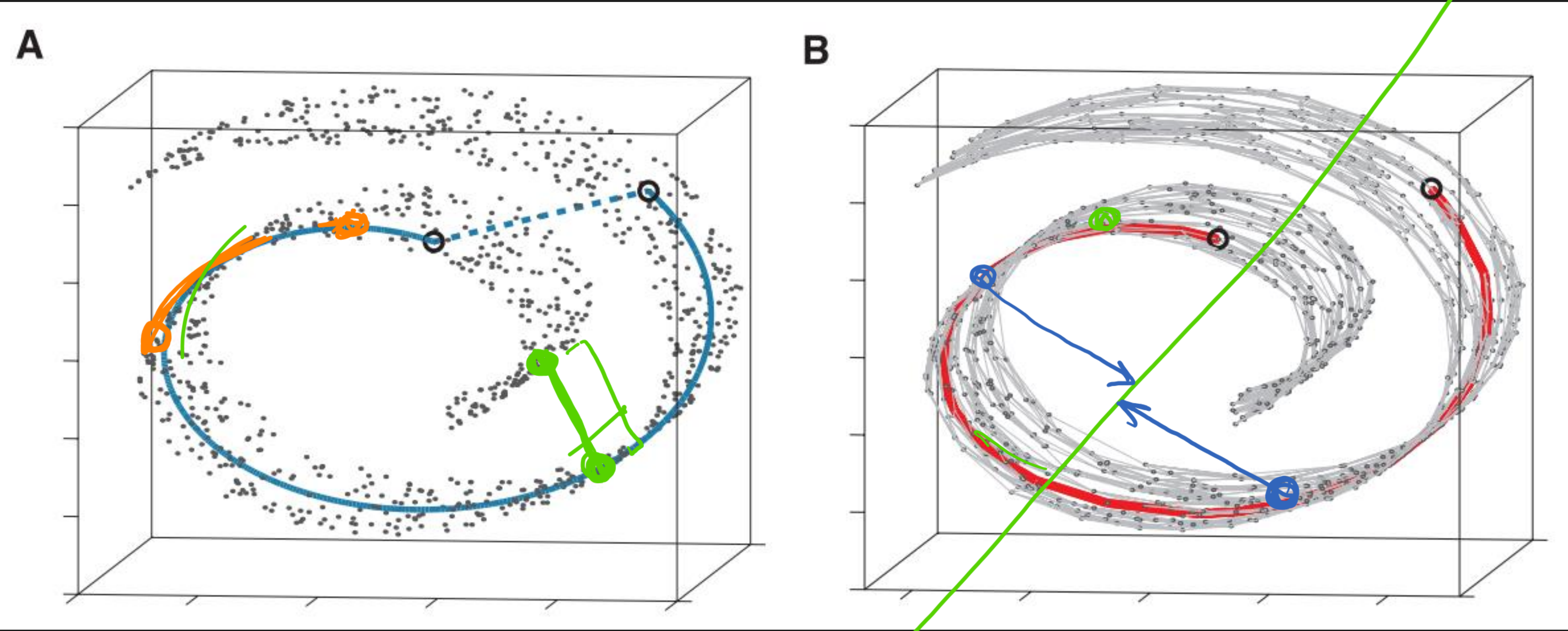
$n \times d$

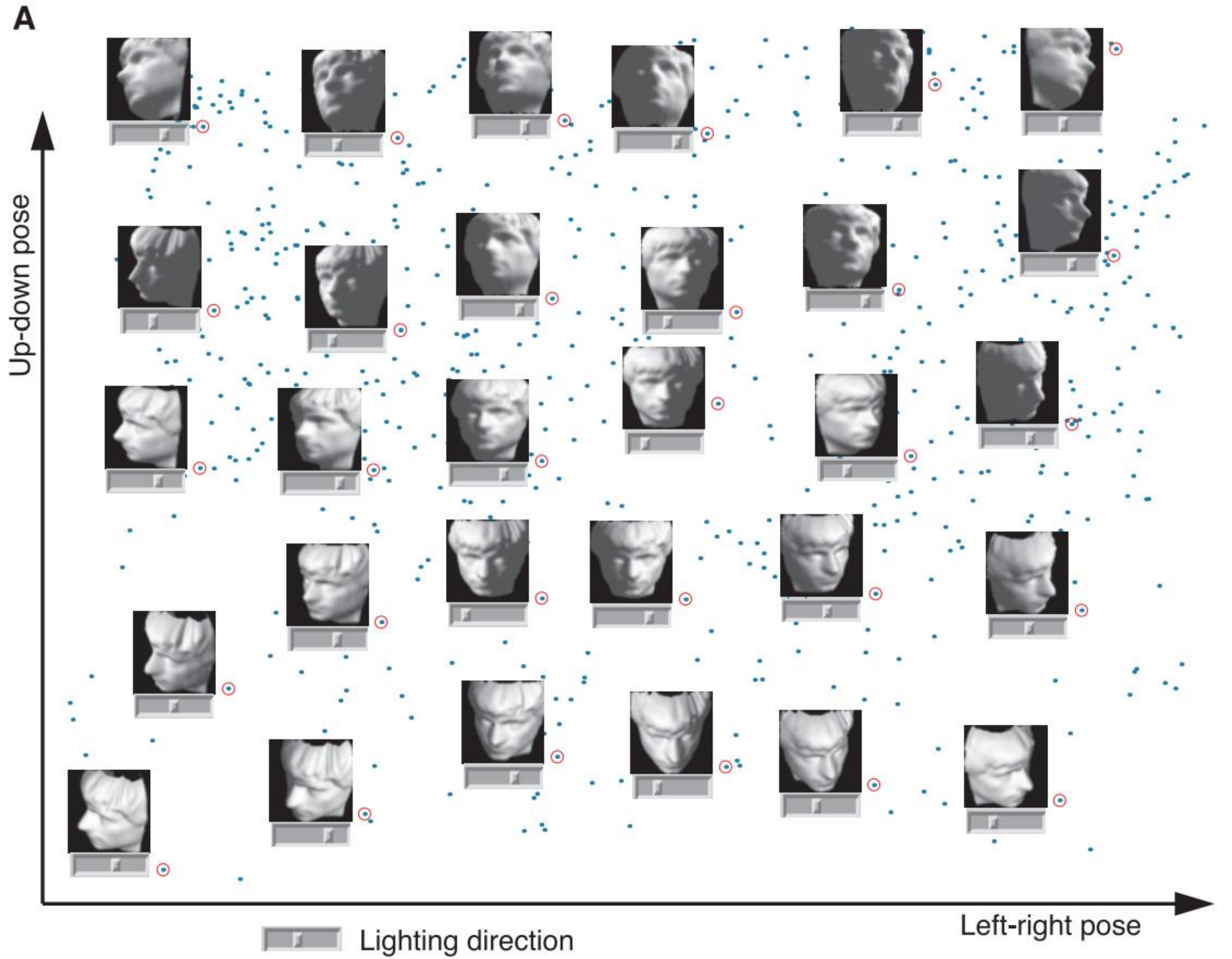
MAP AND THE INVERSE MAP

BENEFITS OF REDUCING DIMENSIONALITY

- Visualization
- Computational benefits
- Curse of dimensionality?

LINEAR VS NON-LINEAR DIMENSIONALITY REDUCTION




A
Up-down pose
Left-right pose
Lighting direction

IDEAS FOR NONLINEAR DIMENSIONALITY REDUCTION?