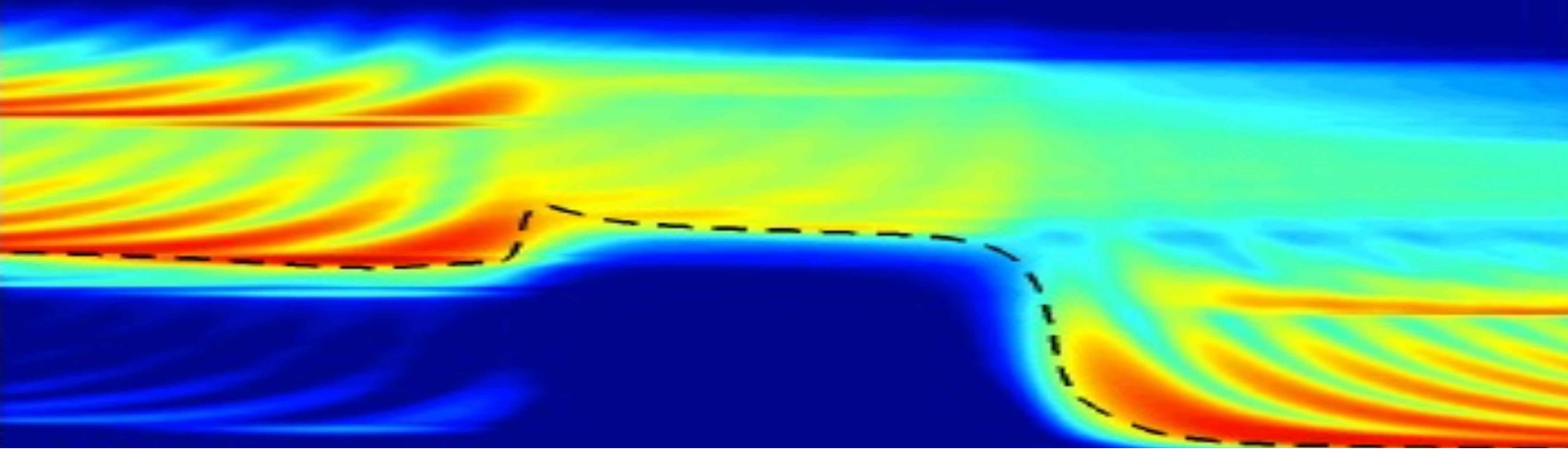


EECS 16B
Designing Information Devices and Systems II
Lecture 20

Prof. Yi Ma
Department of Electrical Engineering and Computer Sciences, UC Berkeley,
yima@eecs.berkeley.edu

Outline

- Spectral Theorem (finish)
- Singular Value Decomposition (Motivations)
- Least Squares and Minimum Norm Solution
- Identifying Low-dim Linear Subspace

Spectral Theorem

Diagonalization for $A \in \mathbb{R}^{n\times n}$ with n independent eigenvectors: $V^{-1}AV = \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \lambda_n \end{bmatrix}$

V.

$A = V \Lambda V^{-1}$ $[v_1, v_2, \ldots, v_n]$

Triangularization for $A \in \mathbb{R}^{n\times n}$ with real eigenvalues: $U^{-1}AU = U^{\top}AU = \begin{bmatrix} t_{11} & t_{12} & \cdots & t_{1n} \\ 0 & t_{22} & \cdots & t_{2k} \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & t_{nn} \end{bmatrix}$

U ortlogo.

proof: Gram-Schmidt + induction

For real symmetric matrices $A = A^{\top} \in \mathbb{R}^{n\times n}$:

$$V^{-1}AV = V^{\top}AV = \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \lambda_n \end{bmatrix}$$

$$M = \frac{M+M^{T}}{2} + \frac{M-M^{T}}{2}$$

sym. anti.

$n_1 \neq n_2$

$\mathbb{R}^{n_1} \ni B\underline{u} \neq \lambda \underline{u}$? $\in \mathbb{R}^{n_2}$

Spectral Theorem

Theorem: Let $A = A^\top \in \mathbb{R}^{n\times n}$ be a *real and symmetric* matrix. Then

1. All eigenvalues of A are real.
2. A is diagonalizable.
3. All eigenvectors are orthogonal to each other.

$A = A^T$

$A = U \Lambda U^T$ $T = T^T$

$AU = U\Lambda \Rightarrow A\vec{u}_i = \lambda_i \vec{u}_i$

An Important Example: for any $B \in \mathbb{R}^{n_1\times n_2}$, we have two associated symmetric matrices:

$$A_1 = BB^\top \in \mathbb{R}^{n_1\times n_1} \qquad A_2 = B^\top B \in \mathbb{R}^{n_2\times n_2}$$

$A_1 = V\Lambda_1 V^T \quad (VV^T = I)$ $A_2 = U\Lambda_2 U^T \quad (UU^T = I)$

$B \quad (V, U, \Lambda)$

Spectral Theorem (extensions)

$j b$

What if A is real and *anti-symmetric*: $A^\top = -A \in \mathbb{R}^{n\times n}$

$$A\vec{u} = \lambda \vec{u} \qquad \bar{\vec{u}}^T A^T = \bar{\lambda}\, \bar{\vec{u}}^T$$

$$\bar{\vec{u}}^T A^T \vec{u} = \bar{\lambda}\, \bar{\vec{u}}^T \vec{u}$$

$$A^T\vec{u} = -A\vec{u}$$

$$-\lambda \bar{\vec{u}}^T \vec{u} = \bar{\lambda}\, \bar{\vec{u}}^T \vec{u}$$

$$-\lambda = \bar{\lambda}$$

$$\overline{a+jb} = a - jb$$

$$\overline{\begin{bmatrix} a_1 + j b_1 \\ a_2 + j b_2 \end{bmatrix}}^T = [a_1 - j b_1, a_2 - j b_2]$$

An Important Example: $R(t)$ is a continuous rotation $R(t)^T R(t) = I$

$$R(t) \in \mathbb{R}^{3\times 3}$$

$$\frac{d\left(R(t)^T R(t)\right)}{dt} = 0$$

$$\left(\dot{R}(t)^T R(t)\right)^T = R(t)^T \dot{R}(t)$$

$$\dot{R}(t)^T R(t) + R(t)^T \dot{R}(t) = 0$$

$$\omega(t) + \omega(t)^T = 0$$

Singular Value Decomposition (SVD)

Diagonalization for $A \in \mathbb{R}^{n \times n}$ with n independent **eigenvectors**: $V^{-1}AV = \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \lambda_n \end{bmatrix}$

$y = Ax \qquad \underline{x = A^{-1} y}$

Triangularization for $A \in \mathbb{R}^{n \times n}$ with real **eigenvalues**: $U^{-1}AU = U^{\top}AU = \begin{bmatrix} t_{11} & t_{12} & \cdots & t_{1n} \\ 0 & t_{22} & \cdots & t_{2k} \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & t_{nn} \end{bmatrix}$

What about a non-square matrix: $A \in \mathbb{R}^{m \times n}$?

$\underline{m \geq n: \quad y = A\vec{x}\,?} \qquad \underline{m < n: \quad \vec{y} = A\vec{x}\,?}$

$Q^{T} y = Rx \qquad \boxed{A} = QR \qquad \boxed{A}$

Over-determined: System Identification

Problem: consider the discrete linear system:

$$\vec{u}[i] \longrightarrow \boxed{\vec{x}[i+1] = A\vec{x}[i] + B\vec{u}[i] + \vec{e}[i]} \longrightarrow \vec{x}[i+1]$$

Given: observed inputs and outputs:

$$\vec{u}[0], \vec{u}[1], \ldots, \vec{u}[l], \ldots$$
$$\vec{x}[0], \vec{x}[1], \ldots, \vec{x}[l], \ldots$$

Objective: learn the system parameters: A, B

n

$$\begin{bmatrix} \vec{x}(1)^T \\ \vec{x}(2)^T \\ \vdots \\ \vec{x}(l) \end{bmatrix} = \begin{bmatrix} x[0]^T \; u^T \\ \vdots \\ \\ \end{bmatrix} \begin{bmatrix} A^T \\ B^T \end{bmatrix} + \vec{e}$$

$$\Rightarrow \vec{S} = D\vec{P} + \vec{e}$$

n

$l \gg n.$

Least Squares: Some Extensions

$\vec{s} \in \mathbb{R}^l, \quad D \in \mathbb{R}^{l\times q}, \quad \vec{p} \in \mathbb{R}^q, \quad \vec{e} \in \mathbb{R}^l \qquad \vec{s} = D \underbrace{\vec{p}}_{\text{unknown}} + \vec{e}$

1. Over-determined $(l \geq q,\ \operatorname{rank}[D] = q)$

$$\vec{p}_\star = \arg\min_{\vec{p}} \|\vec{s} - D\vec{p}\|_2^2 \quad = (D^\top D)^{-1} D^\top \vec{s}$$

$\boxed{D}$

2. Under-determined $(l < q,\ \operatorname{rank}[D] = l)$

$\vec{s} = \boxed{\quad D \quad}\ \vec{p}$

$$\vec{p}_\star = \arg\min_{\vec{p}} \|\vec{p}\|_2^2 \ \text{ s.t. } \ \vec{s} = D\vec{p} \quad = D^\top (DD^\top)^{-1} \vec{s}.$$

$\vec{p} + \vec{v} \qquad \vec{v} \in \mathrm{Nu}(D)$

3. Ridge regression

$$\vec{p}_\star = \arg\min_{\vec{p}} \|\vec{s} - D\vec{p}\|_2^2 + \lambda\|\vec{p}\|_2^2 \quad = (D^\top D + \lambda I)^{-1} D^\top \vec{s}.$$

Under-determined: Minimum-Norm Control

Definition: a system $\vec{x}[i+1] = A\vec{x}[i] + Bu[i]$ is said to be ***controllable*** if given any target state $\vec{x}_f \in \mathbb{R}^n$ and initial state $\vec{x}[0]$, we can find a time $i = \ell$ and a sequence of control input $u[0], \ldots, u[\ell]$ such that $\vec{x}[\ell] = \vec{x}_f$

$$\vec{x}[\ell] = A^\ell \vec{x}[0] + \mathcal{C}_\ell \vec{u}[\ell] \qquad \mathcal{C}_\ell \doteq [A^{\ell-1}B \mid \cdots \mid AB \mid B] \in \mathbb{R}^{n\times\ell}$$

$$\underbrace{\vec{x}(\ell)}_{\vec{x}_f} = A^\ell \vec{x}[0] + \underbrace{[A^{\ell-1}B, \ldots, AB, B]}_{C_\ell} \underbrace{\begin{bmatrix} u[0] \\ u[1] \\ \vdots \\ u[\ell] \end{bmatrix}}_{\vec{u}[\ell]?}$$

$$\underbrace{[\vec{x}_f - A^\ell \vec{x}[0]]}_{\mathbb{R}^n} = \underbrace{C_\ell}_{\mathbb{R}^{n\times\ell}} \underbrace{\vec{u}[\ell]}_{\mathbb{R}^\ell}$$

$\ell \geq n \qquad \ell > n$

$\vec{y} = A\vec{x}$ $\leftarrow$ $\vec{x} = \vec{u}$ $\qquad \vec{u} + Nu(A)$ $\leftarrow$ infinite.

$\|\vec{u}\|_2^2 = u[0]^2 + u[1]^2 + \cdots + u[l]^2$

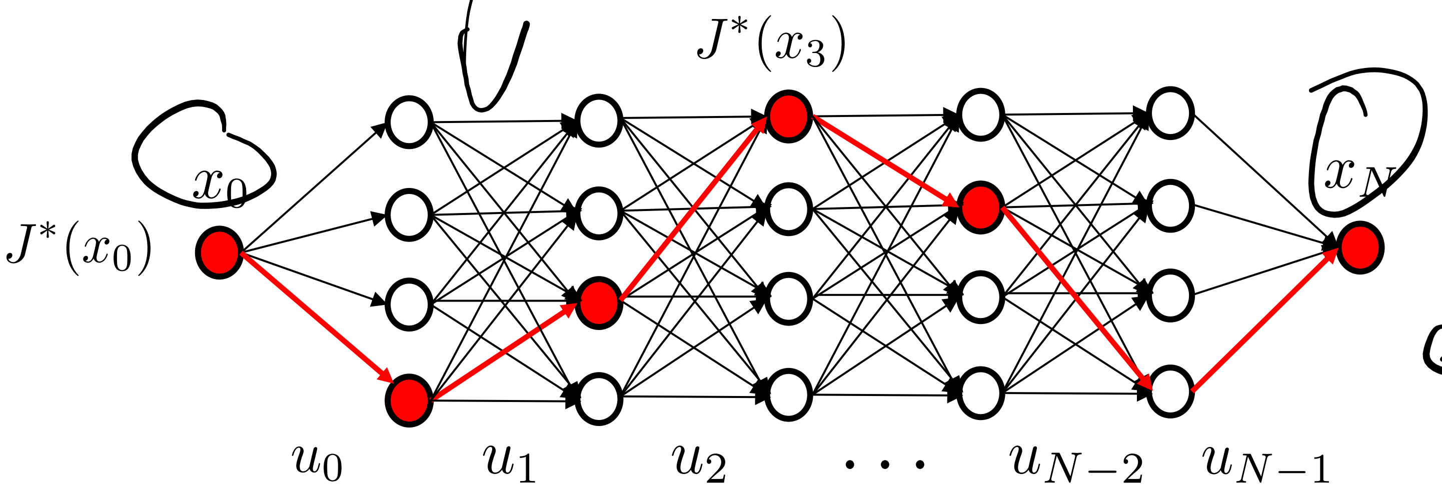
"energy"

minimum (l^2) norm

Principle of (Path) Optimality

Dido of Carthage…, Euler, Lagrange, Newton, Hamilton, Jacobi, Pontryagin, Bellman, Ford, Kalman

850 BC 1960 AC



x_f

Principle of Optimality (*Richard Bellman'54*): 1954.

An optimal path has the property that any subsequent portion is optimal.

RL.

Minimum Norm Solution

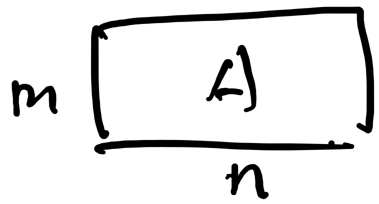
Theorem: Let $A \in \mathbb{R}^{m\times n}$ have full row rank, i.e. $\mathrm{rank}(A) = m$. Then for any $\vec{y} \in \mathbb{R}^m$ the following problem

$$\min \|\vec{x}\|_2^2 \quad \text{subject to} \quad \vec{y} = A\vec{x}$$

has a unique optimal solution $\vec{x}_\star = A^\top (AA^\top)^{-1}\vec{y}$.

Proof:

① $A\vec{x}_\star = \underbrace{AA^T(AA^T)^{-1}}\vec{y} = I \cdot \vec{y}$



$AA^T_{m\times m}$

② $A\vec{x}' = \vec{y}$ $\quad \|\vec{x}'\|_2^2 \geq \|\vec{x}_\star\|_2^2$

$$\|\vec{x}'\|_2^2 = \|\underbrace{\vec{x} + (\vec{x}' - \vec{x})}_{v}\|_2^2 = \underline{\|\vec{x}\|_2^2 + 2\langle \vec{x}, (\vec{x}' - \vec{x})\rangle + \|\vec{x}' - \vec{x}\|_2^2}$$

$v^T v$

$\vec{x}_\star$

Minimum Norm Solution

Theorem: Let $A \in \mathbb{R}^{m\times n}$ have full row rank, i.e. $\operatorname{rank}(A) \leq m$. Then ~~for any $\vec{y} \in \mathbb{R}^m$~~ the following problem

$$\min \|\vec{x}\|_2^2 \quad \text{subject to} \quad \vec{y} = A\vec{x}$$

has a unique optimal solution $\vec{x}_\star = A^\top(AA^\top)^{-1}\vec{y}$.

$$\langle Av, u\rangle = v^TA^Tu = \langle v, A^Tu\rangle$$

Proof:

$$\|\vec{x}'\|_2^2 = \|\vec{x}_* + (\vec{x}' - \vec{x}_*)\|_2^2$$

$$= \|\vec{x}_*\|_2^2 + 2\langle \vec{x}_*, \vec{x}' - \vec{x}_*\rangle + \|\vec{x}' - \vec{x}_*\|_2^2$$

$$= \|\vec{x}_*\|_2^2 + 2\langle A^T(AA^T)^{-1}\vec{y}, \vec{x}' - \vec{x}_*\rangle + \|\vec{x}' - \vec{x}_*\|_2^2$$

$$\geq\!\!= \|\vec{x}_*\|_2^2 + 2\langle \underbrace{(AA^T)^{-1}\vec{y}}_{v}, \underbrace{A(\vec{x}' - \vec{x}_*)}_{u}\rangle + \|\vec{x}' - \vec{x}_*\|_2^2$$

$$0 \qquad \geq 0$$

$$A\vec{x}_* = \vec{y} \qquad A\vec{x}' = \vec{y}$$

Least-Squares vs Minimum-Norm Solutions

Moore-Penrose pseudo inverse of $A \in \mathbb{R}^{m\times n}$: $\vec{y} = A\vec{x}, \quad \vec{x} = A^{\dagger}\vec{y}$. Moore-Penrose pseudo inverse

$A^{-1} y$

$$m \geq n \text{ and } \operatorname{rank}(A) = n : \quad A^{\dagger} = (AA^{\top})^{-1}A^{\top}$$

$\boxed{A}$

least squares.

$$m \leq n \text{ and } \operatorname{rank}(A) = m : \quad A^{\dagger} = A^{\top}(AA^{\top})^{-1}$$

$\boxed{A}$

$A^{\dagger}$

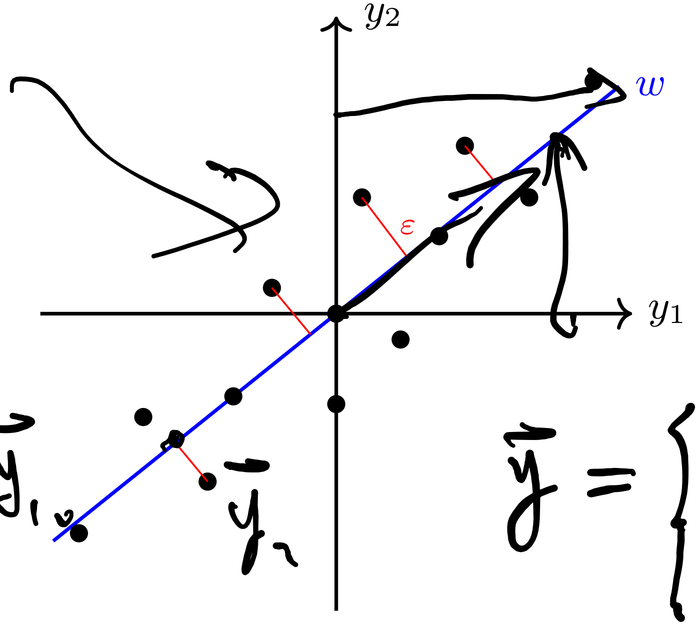
$A \in \mathbb{R}^{m\times n}$ not full column or row rank?

"It is quite probable that our mathematical insights and understandings are often used to achieve things that could in principle also be achieved computationally but where blind computation without much insight may turn out to be so inefficient that it is unworkable."

-- Roger Penrose, *Shadows of the Mind*

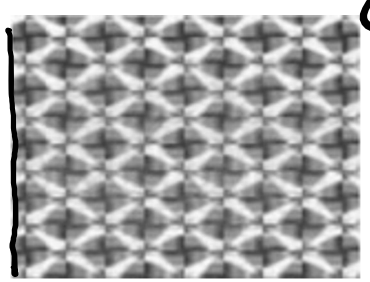
Identifying a Low-dim Linear Subspace

"Principal Component"



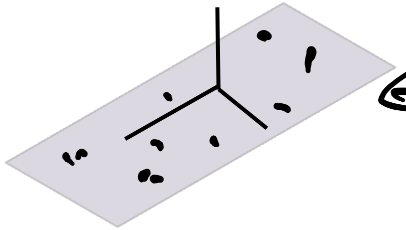
$$\vec{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \end{bmatrix}$$

100



200

$\mathbb{R}^{100}$



S^{10}

$$X = [\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_n] \in \mathbb{R}^{m\times n}$$

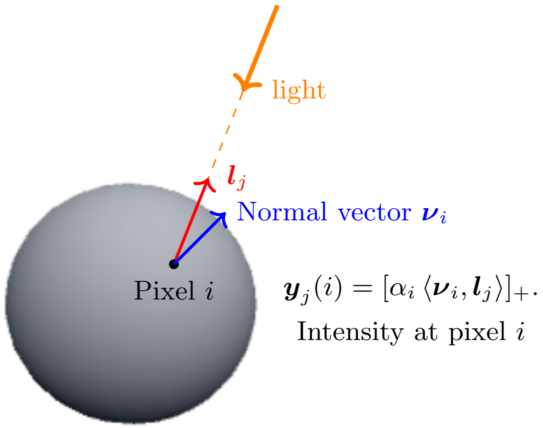
$$x_i = [w_1 \; w_2 \cdots w_{10}] \begin{bmatrix} \alpha_1^i \\ \vdots \\ \alpha_n^i \end{bmatrix}$$

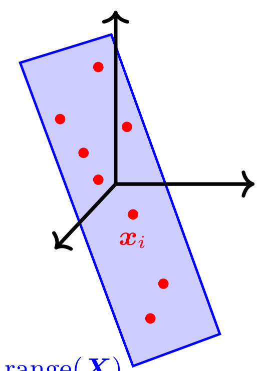
$$\{\vec{y}_1, \vec{y}_2, \ldots, \vec{y}_n] \in \mathbb{R}^{2\times n}$$

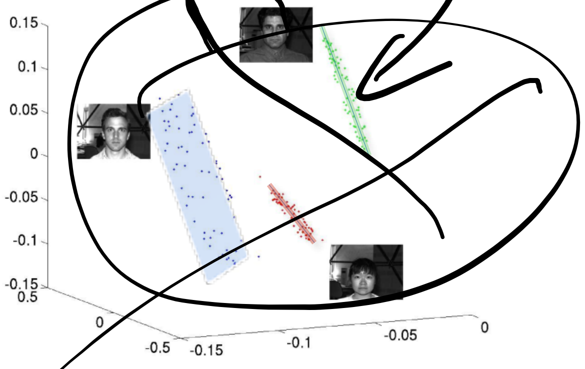
$$X = W\cancel{X}$$

$$[\alpha_1 \vec{w}, \alpha_2 \vec{w}, \ldots, \alpha_n \vec{w}] + e = \underline{\{w_1, \ldots, w_{10}]} [\vec{\alpha}_1, \vec{\alpha}_2, \ldots, \vec{\alpha}_n]$$

Identifying Low-dim Linear Subspaces







10^6

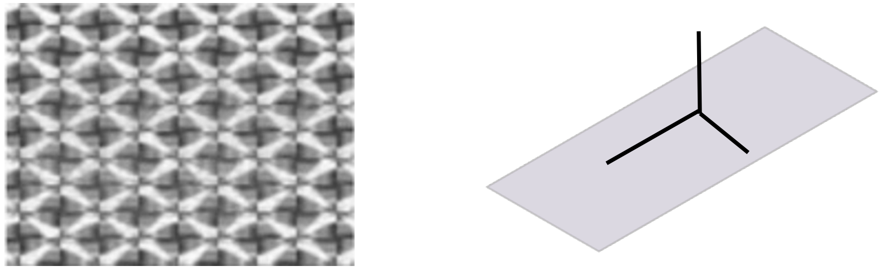
rank ≤ 3

≤ 9

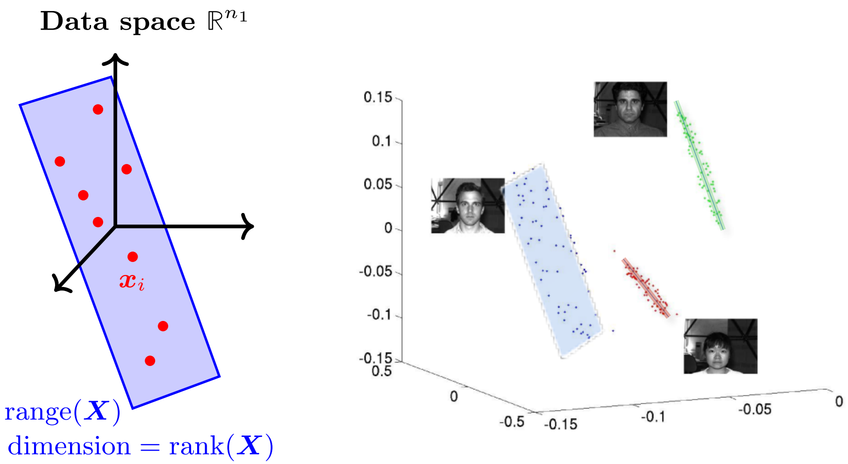
$10 - 200$

Recovering a Low-dim Linear Subspace

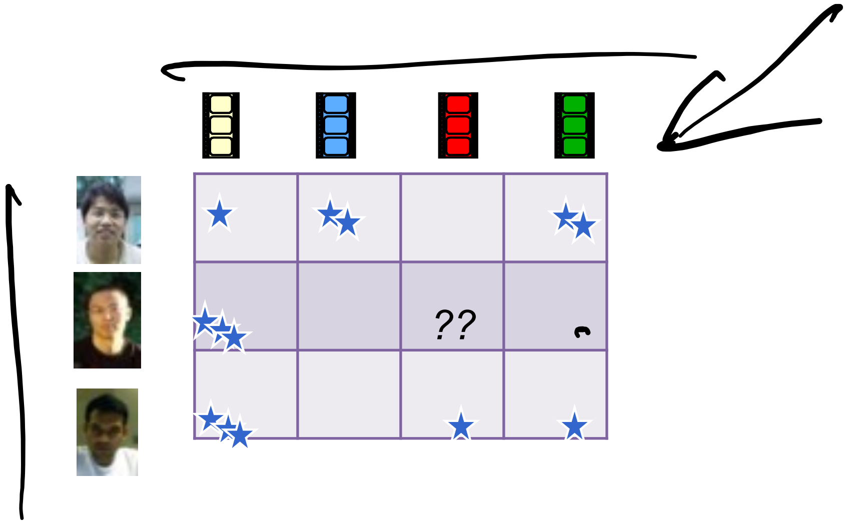
One low-dim subspace



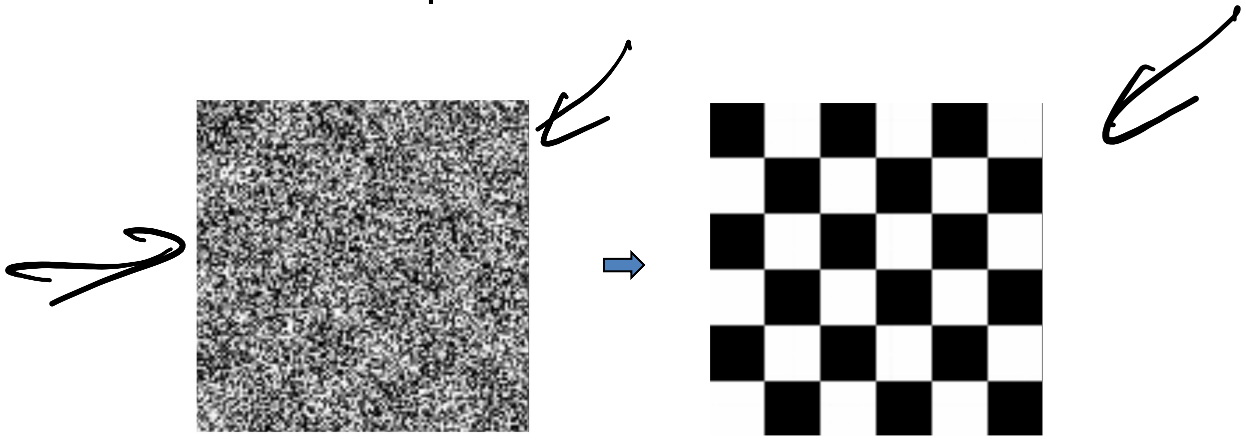
Multiple low-dim subspaces



Incomplete low-rank matrix



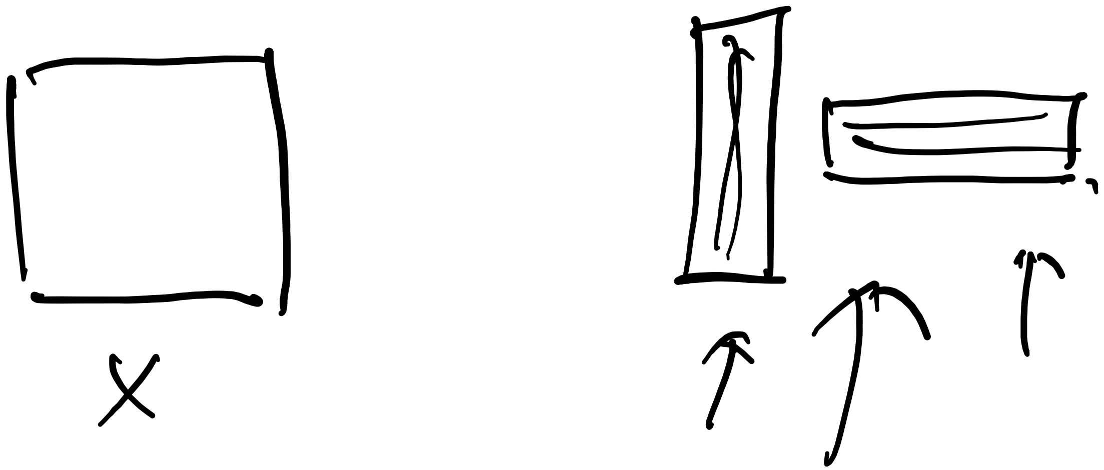
Corrupted low-rank matrix



Identifying a Low-dim Linear Subspace

Given $X = [\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_n] \in \mathbb{R}^{m \times n}$, find a low rank L : $\min_L \|X - \underline{L}\|_F^2, \;\; \text{s.t.} \;\; \text{rank}(L) \le r.$

$$X = L = U V^T$$



M_1, M_2

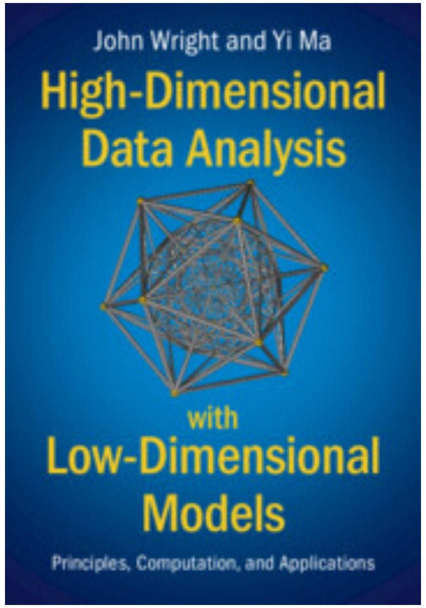
High-dimensional Data Analysis

Principal Component Analysis: Finding one linear subspace

Compressive Sensing: Finding multiple low-dim linear structures

- Solving under-determined systems of linear equations
- Low-rank matrix approximation or recovery

Deep Learning: Finding non-linear low-dimensional structures



EECS 208: [Computational Principles for High-Dimensional Data Analysis]

(from SVD/PCA, to Generalized PCA, Robust PCA, Nonlinear PCA, and to Deep Networks…)