



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

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

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

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

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.

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

$$A_1 = BB^\top \in \mathbb{R}^{n_1 \times n_1}$$

$$A_2 = B^\top B \in \mathbb{R}^{n_2 \times n_2}$$

Spectral Theorem (extensions)

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

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

Singular Value Decomposition (SVD)

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

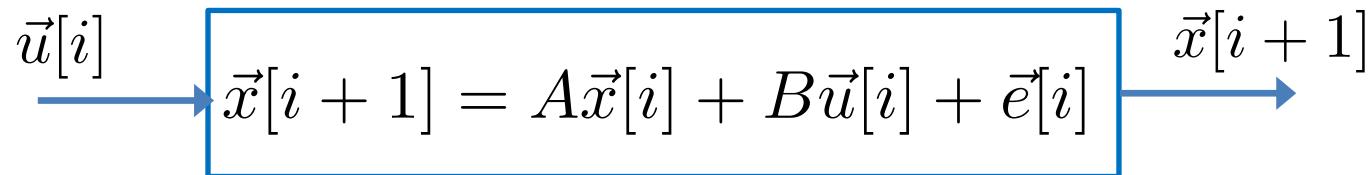
Triangularization for $A \in \mathbb{R}^{n \times n}$ with real **eigenvalues**: $U^{-1}AU = U^\top AU = \begin{bmatrix} t_{11} & t_{12} & \cdots & t_{1n} \\ 0 & t_{22} & \cdots & t_{2k} \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & t_{nn} \end{bmatrix}$

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

$m \geq n$: $y = A\vec{x}$? $m < n$: $\vec{y} = A\vec{x}$?

Over-determined: System Identification

Problem: consider the discrete linear system:



Given: observed inputs and outputs:

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

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

Objective: learn the system parameters: A, B

Least Squares: Some Extensions

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

unknown

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

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

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

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

3. Ridge regression

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

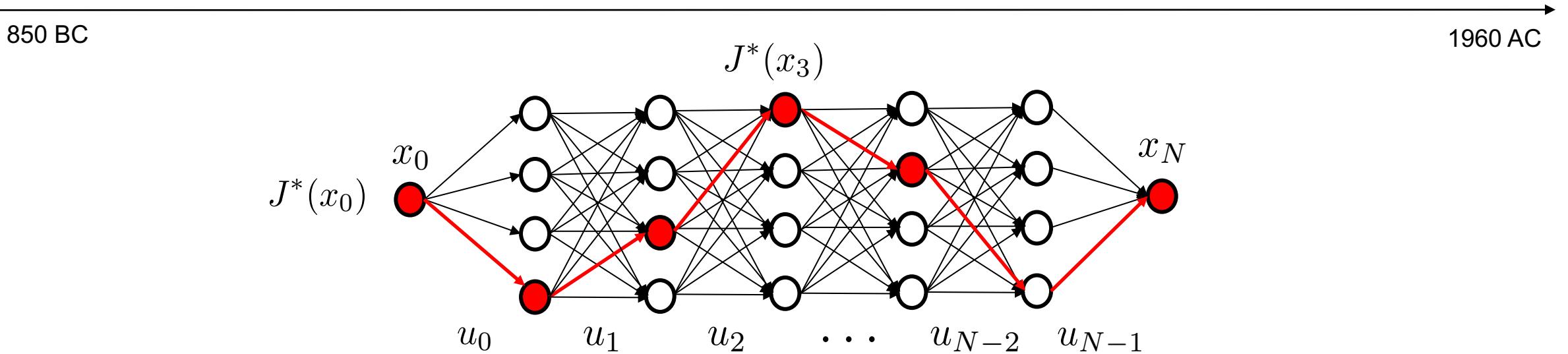
Under-determined: Minimum-Norm Control

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 $u[0], \dots, u[\ell]$ such that $\vec{x}[\ell] = \vec{x}_f$

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

Principle of (Path) Optimality

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



Principle of Optimality (Richard Bellman'54):
An optimal path has the property that any subsequent portion is optimal.

Minimum Norm Solution

Theorem: Let $A \in \mathbb{R}^{m \times n}$ have full row rank, i.e. $\text{rank}(A) \leq m$. 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 $\vec{x}_* = A^\top (AA^\top)^{-1} \vec{y}$.

Proof:

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

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

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

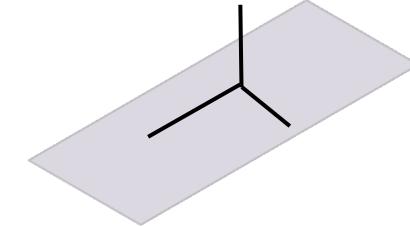
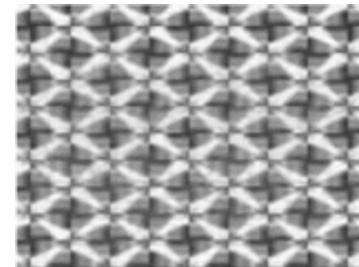
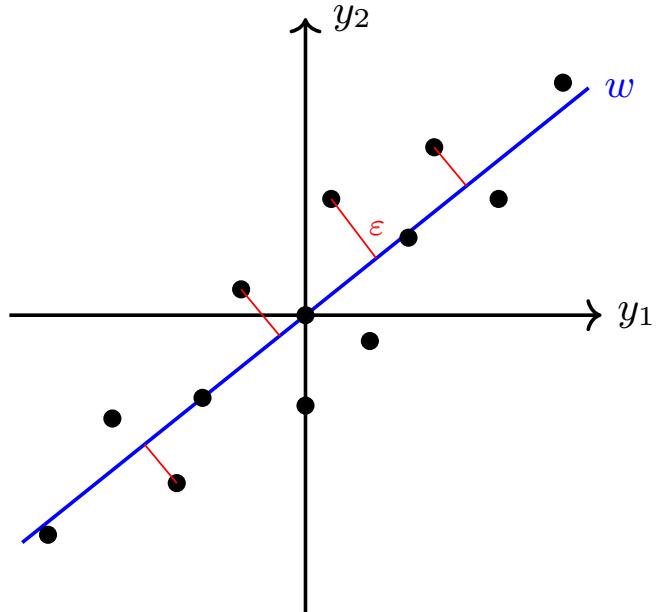
$A \in \mathbb{R}^{m \times n}$ not full column or row rank?

“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*

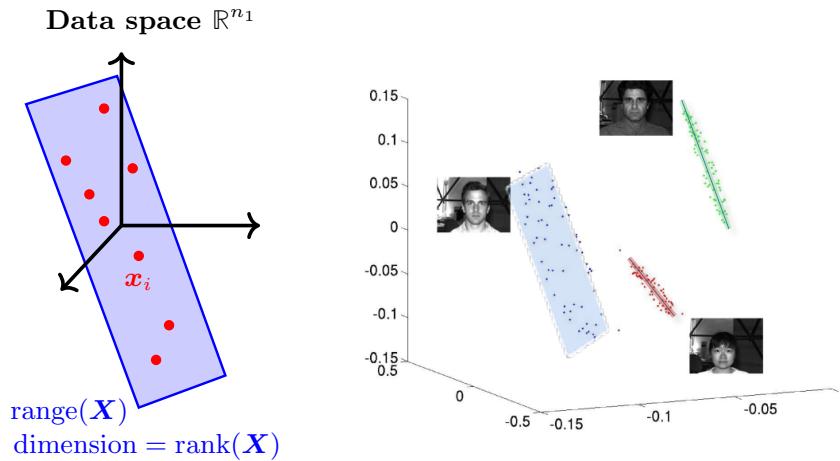
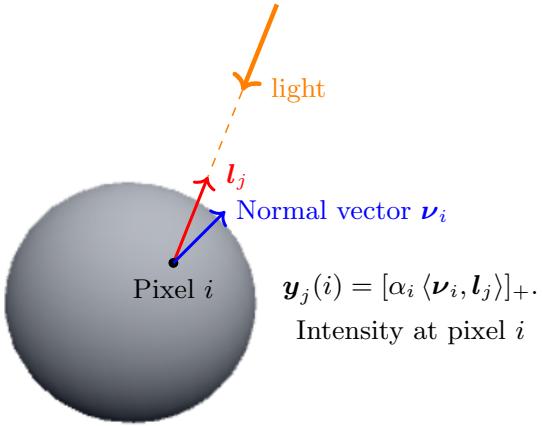
Identifying a Low-dim Linear Subspace

“Principal Component”



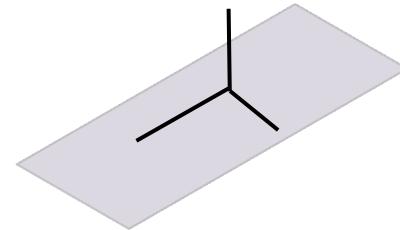
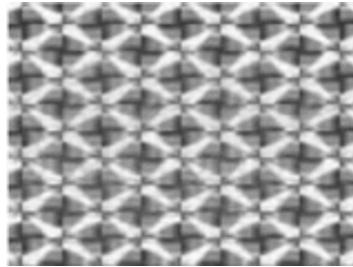
$$X = [\vec{x}_1, \vec{x}_2, \dots, \vec{x}_n] \in \mathbb{R}^{m \times n}$$

Identifying Low-dim Linear Subspaces



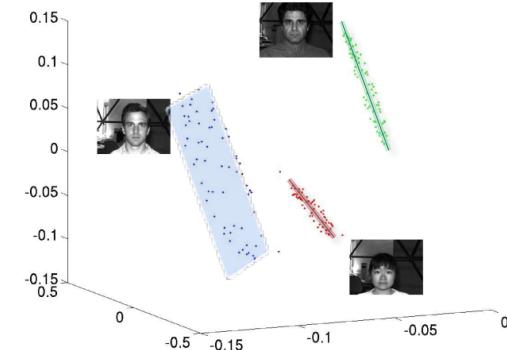
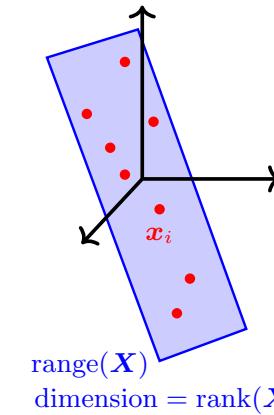
Recovering a Low-dim Linear Subspace

One low-dim subspace

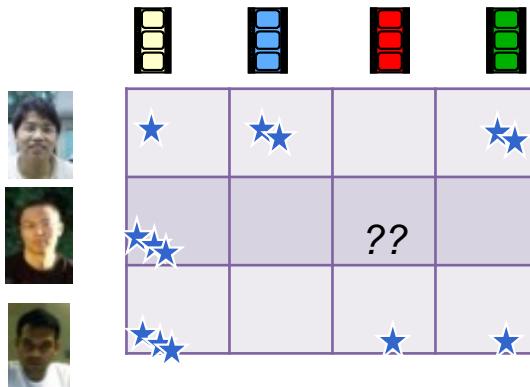


Multiple low-dim subspaces

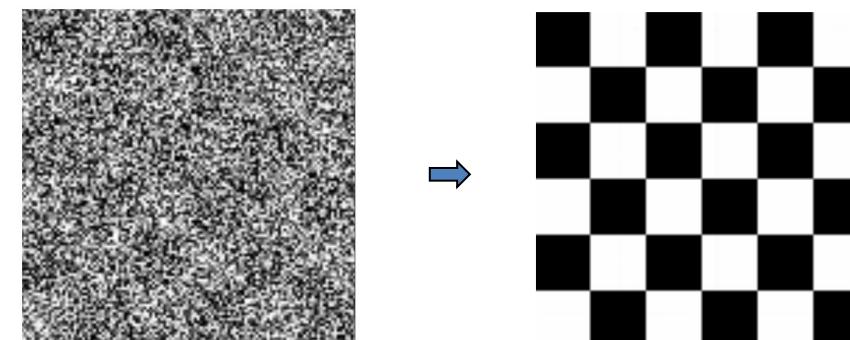
Data space \mathbb{R}^{n_1}



Incomplete low-rank matrix



Corrupted low-rank matrix



Identifying a Low-dim Linear Subspace

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

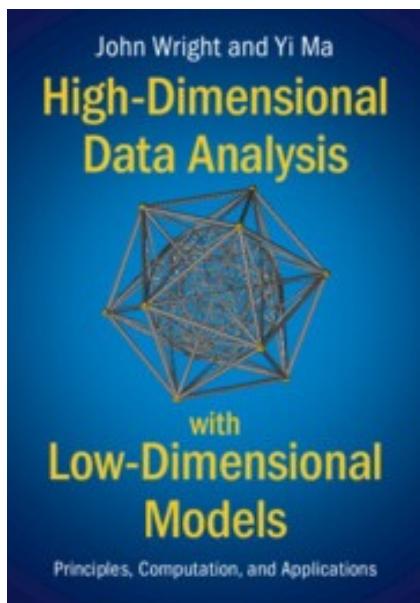
High-dimensional Data Analysis

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...)