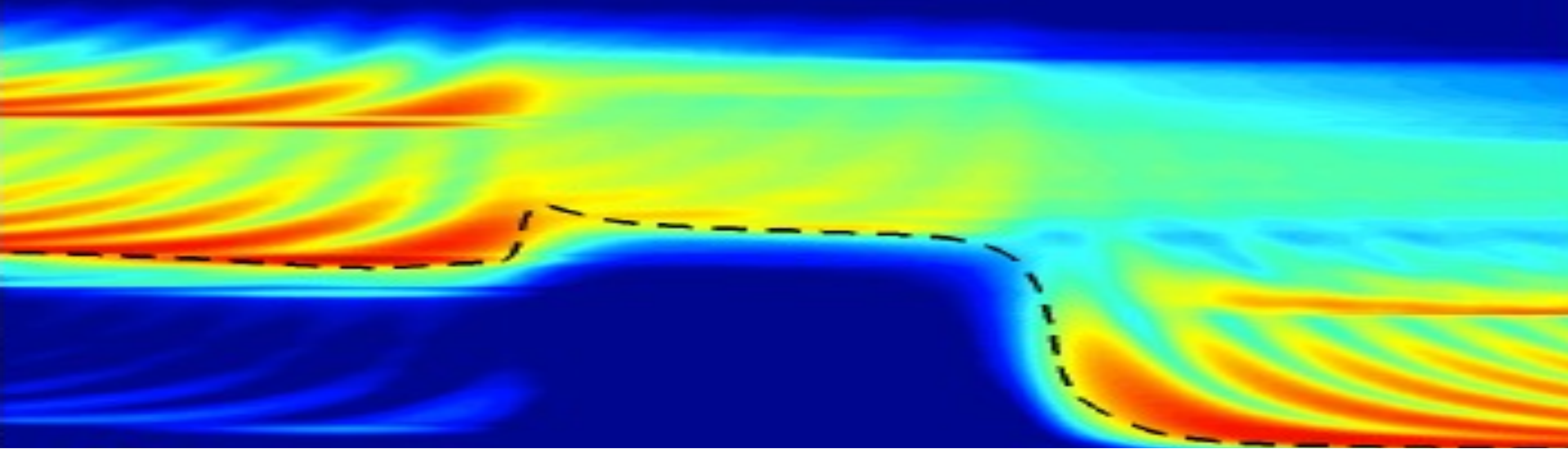


# EECS 16B
# Designing Information Devices and Systems II
# Lecture 25

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

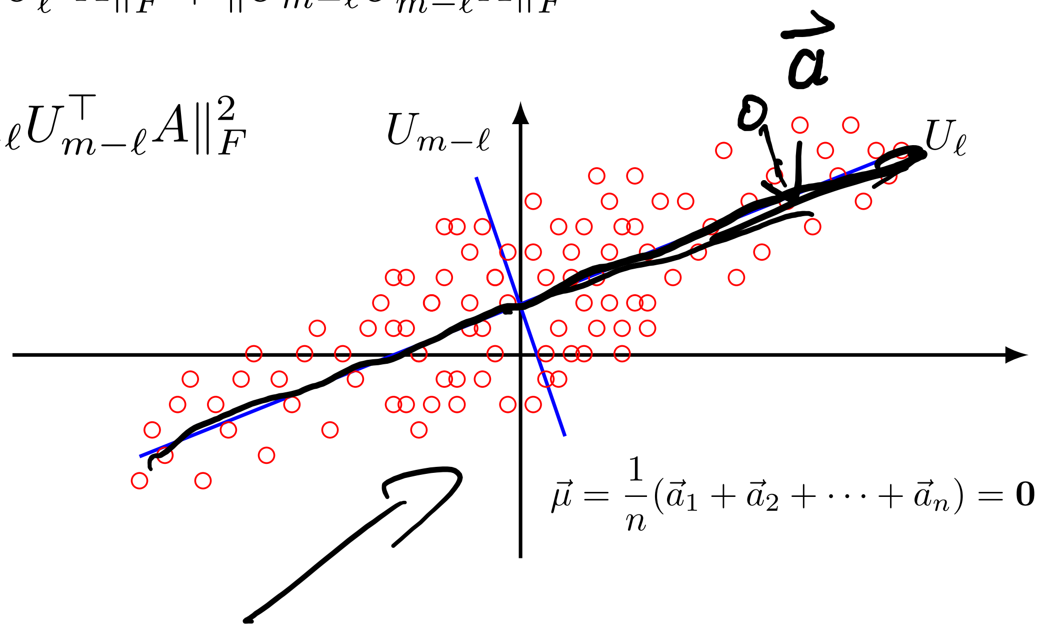
# Outline

- Principal Component Analysis (Statistics)
  - Least Squares versus Principal Components
- Linearization of Nonlinear Systems

# Principal Component Analysis (Statistics)

$U = [U_\ell, U_{m-\ell}] \in \mathbb{R}^{m\times m}$ orthogonal $\quad \|A\|_F^2 = \|UU^\top A\|_F^2 = \|U_\ell U_\ell^\top A\|_F^2 + \|U_{m-\ell}U_{m-\ell}^\top A\|_F^2$

$$\underline{\max_{U_\ell} \|U_\ell U_\ell^\top A\|_F^2} \Leftrightarrow \underline{\min_{U_\ell} \|A - U_\ell U_\ell^\top A\|_F^2} \Leftrightarrow \min_{U_{m-\ell}} \|U_{m-\ell}U_{m-\ell}^\top A\|_F^2$$




$A = \{\vec{a}_1, \vec{a}_2, \ldots, \vec{a}_n]$ $\quad \sum \vec{a}_i = 0$

$UU^\top A \leftarrow$ preserve most info.

$\|\vec{a}_i - U_\ell U_\ell^\top \vec{a}_i\|$ minimized.

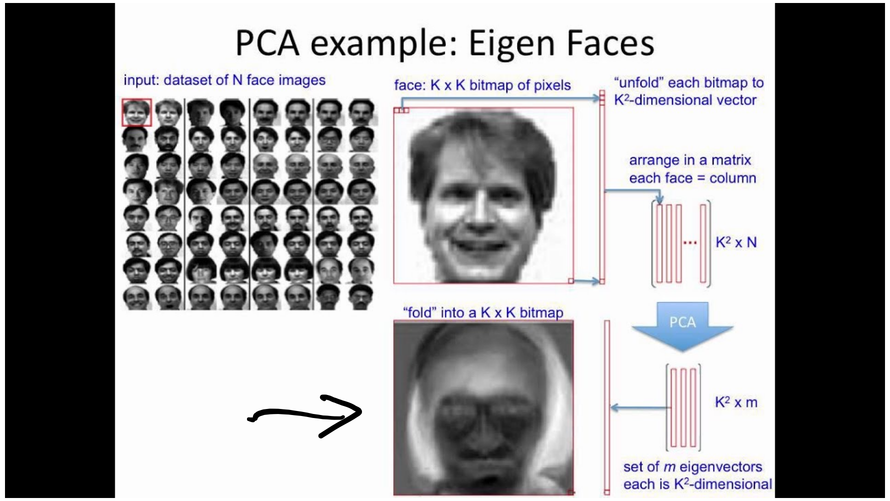
$U_\ell$ ? $\|U_\ell U_\ell^\top \vec{a} - \vec{a}\|$ small $\quad \ell << m$

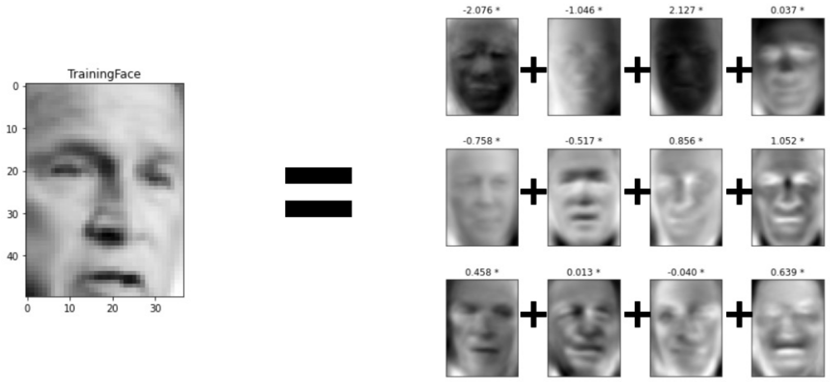
$U_\ell \vec{w} = w_1 \vec{u}_1 + \cdots + w_{12} \vec{u}_{12} \quad \vec{w} \in \mathbb{R}^\ell \quad \ell = 12$

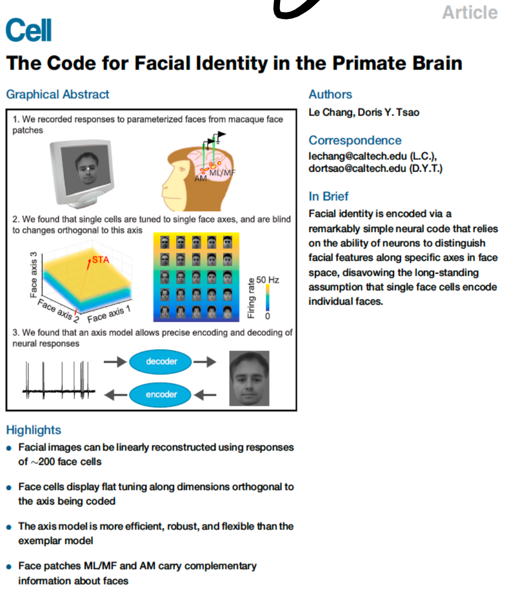
# Applications of PCA

l=12

- Eigenfaces [Turk & Pentland 1991]:





Article

Cell

The Code for Facial Identity in the Primate Brain

Graphical Abstract




Authors

Le Chang, Doris Y. Tsao

Correspondence

lechang@caltech.edu (L.C.),
dortsao@caltech.edu (D.Y.T.)

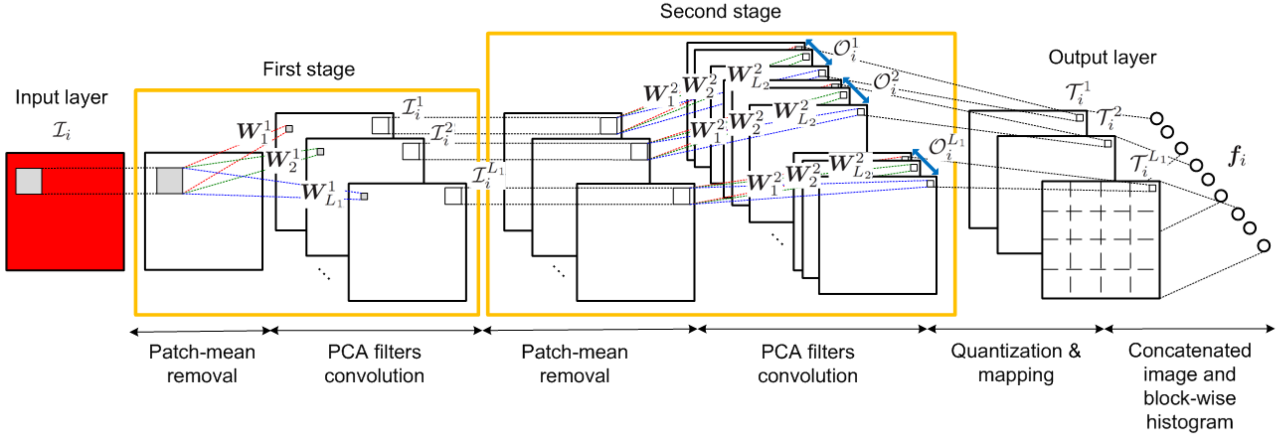

In Brief

Facial identity is encoded via a remarkably simple neural code that relies on the ability of neurons to distinguish facial features along specific axes in face space, disavowing the long-standing assumption that single face cells encode individual faces.

Highlights

- Facial images can be linearly reconstructed using responses of ~200 face cells
- Face cells display flat tuning along dimensions orthogonal to the axis being coded
- The axis model is more efficient, robust, and flexible than the exemplar model
- Face patches ML/MF and AM carry complementary information about faces

- PCANet [Chan & Ma et. al. 2015]:




Recognition rates (%) on FERET dataset.

| Probe sets | Fb | Fc | Dup-I | Dup-II | Avg. |
|---|---|---|---|---|---|
| LBP [18] | 93.00 | 51.00 | 61.00 | 50.00 | 63.75 |
| DMMA [25] | 98.10 | 98.50 | 81.60 | 83.20 | 89.60 |
| P-LBP [21] | 98.00 | 98.00 | 90.00 | 85.00 | 92.75 |
| POEM [26] | 99.60 | 99.50 | 88.80 | 85.00 | 93.20 |
| G-LQP [27] | 99.90 | 100 | 93.20 | 91.00 | 96.03 |
| LGBP-LGXP [28] | 99.00 | 99.00 | 94.00 | 93.00 | 96.25 |
| sPOEM+POD [29] | 99.70 | 100 | 94.90 | 94.00 | 97.15 |
| GOM [30] | 99.90 | 100 | 95.70 | 93.10 | 97.18 |
| PCANet-1 (Trn. CD) | 99.33 | 99.48 | 88.92 | 84.19 | 92.98 |
| PCANet-2 (Trn. CD) | 99.67 | 99.48 | 95.84 | 94.02 | 97.25 |
| PCANet-1 | 99.50 | 98.97 | 89.89 | 86.75 | 93.78 |
| PCANet-2 | 99.58 | 100 | 95.43 | 94.02 | 97.26 |

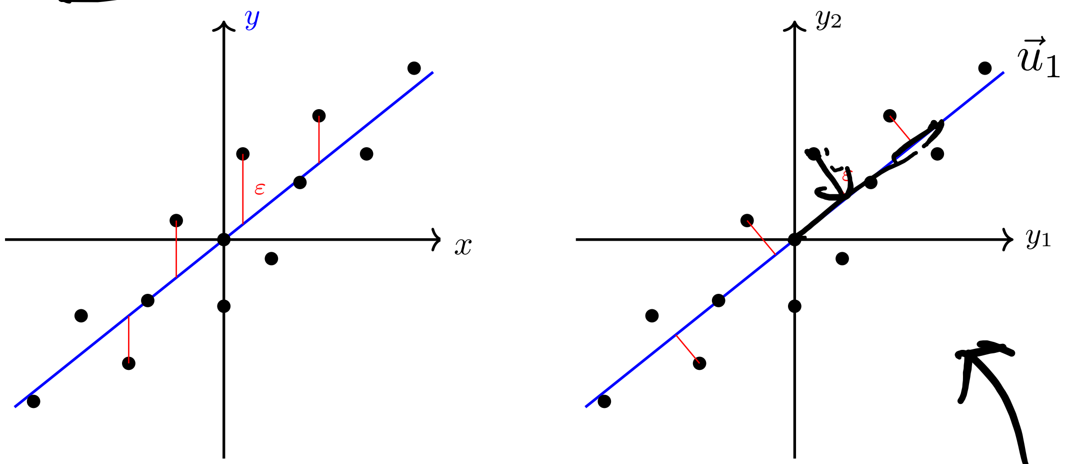
# Least Squares (Regression) versus PCA

$$A = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \\ y_1 & y_2 & \cdots & y_n \end{bmatrix} \leftarrow \begin{matrix} \vec{x}^T \\ \vec{y}^T \end{matrix}$$

$y = \alpha x + n$? ← regression

$$\|\vec{y} - \alpha \vec{x}\|_2^2$$

**Prediction** versus **Correlation**




**Supervised** versus **Unsupervised**

$$A = \begin{bmatrix} y_{11} & y_{12} & \cdots & y_{1n} \\ y_{21} & y_{22} & \cdots & y_{2n} \\ \vdots & \vdots & & \vdots \\ y_{m1} & y_{m2} & \cdots & y_{mn} \end{bmatrix} \longrightarrow \|\vec{u}_1^T A\|_2^2 \text{ maximized}$$

# Least Squares (Regression) versus PCA

**Example:** $A = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 1 & 1 \end{bmatrix} = \begin{bmatrix} \vec{x}^T \\ \vec{y}^T \end{bmatrix}$

$$\min \| \vec{y} - \alpha \vec{x} \|_2^2 \rightarrow \alpha = \frac{\vec{x}^T \vec{y}}{(\vec{x}^T \vec{x})}$$

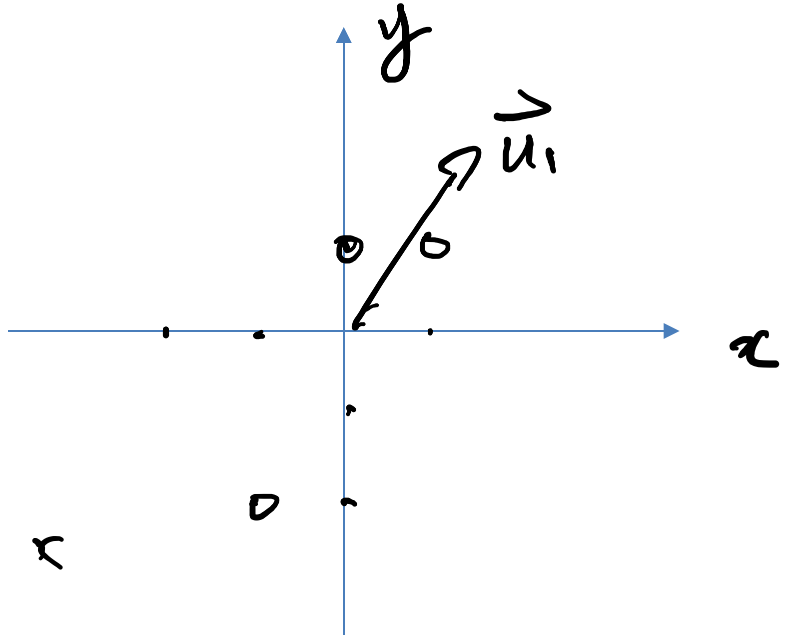
$$= \frac{3}{2} = 1.5$$

$$y = \frac{3}{2} x$$

$$AA^T = \begin{bmatrix} 2 & 3 \\ 3 & 6 \end{bmatrix} \leftarrow \vec{u}_1 = \begin{array}{c|c} \begin{bmatrix} 0.47 \\ 0.88 \end{bmatrix} & \begin{array}{c} x \\ y \end{array} \end{array}$$

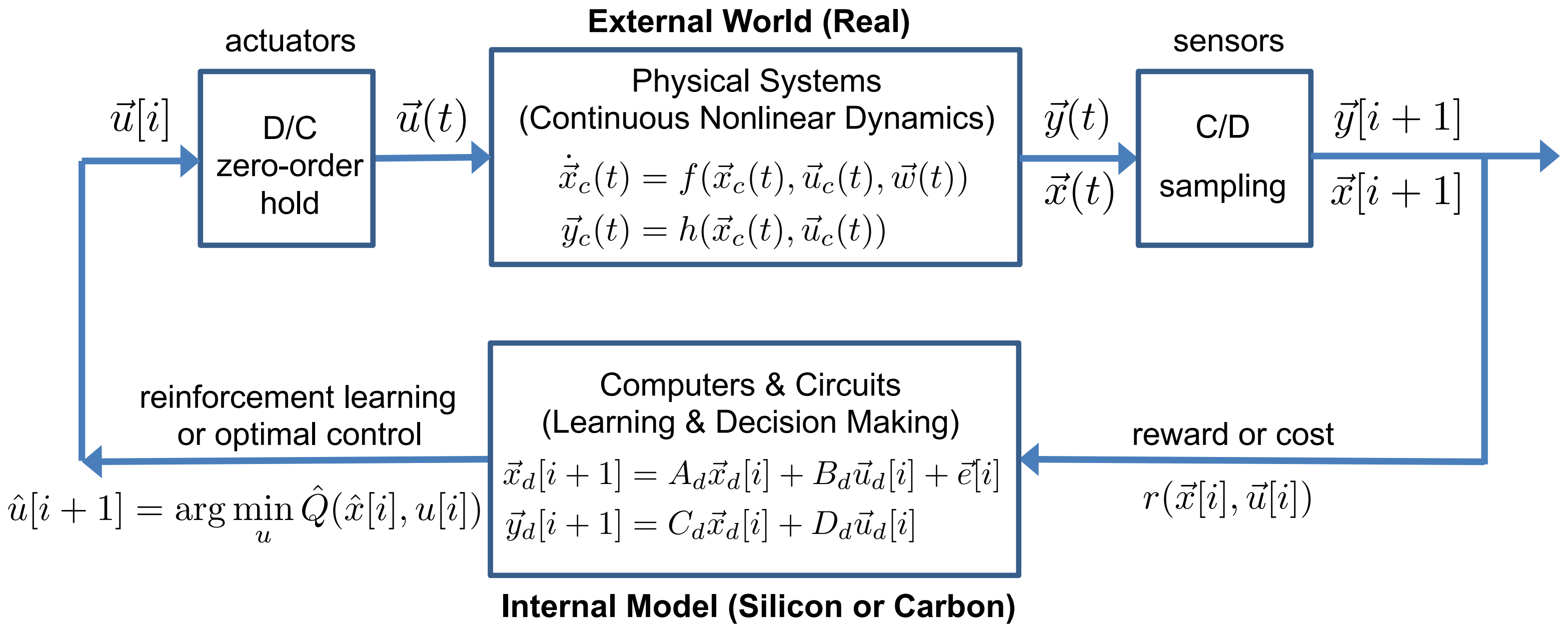
$$\text{slope} = \frac{0.88}{0.47} \approx 1.87.$$

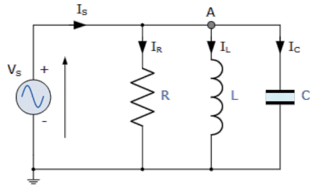
**Prediction versus Correlation**

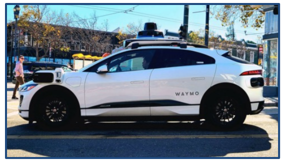


# System Modeling & Control

All autonomous intelligent (AI) systems rely on closed-loop learning and control:



mathematical modeling
from first principles

$$\dot{\vec{x}}_c(t) = f(\vec{x}_c(t), \vec{u}_c(t), \vec{w}(t))$$
$$\vec{y}_c(t) = h(\vec{x}_c(t), \vec{u}_c(t))$$

approximation
& linearization

$$\dot{\vec{x}}(t) = A\vec{x}(t) + B\vec{u}(t) + \vec{n}(t)$$
$$\vec{y}(t) = C\vec{x}(t) + D\vec{u}(t)$$

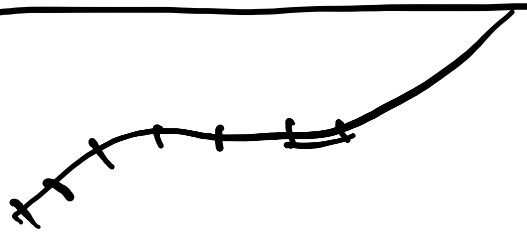
discretization
& digitization

$$\vec{x}_d[i+1] = A_d\vec{x}_d[i] + B_d\vec{u}_d[i] + \vec{e}[i]$$
$$\vec{y}_d[i+1] = C_d\vec{x}_d[i] + D_d\vec{u}_d[i]$$

# Linear versus Nonlinear Systems

**Objectives:** **Identification (learning), Analysis (stability), Control (closed-loop feedback)**

| | Continuous Time | Discrete Time |
|---|---|---|
| Linear Control Systems | $\frac{d\vec{x}(t)}{dt} = A\vec{x}(t) + B\vec{u}(t)$ | $\vec{x}[i+1] = A\vec{x}[i] + B\vec{u}[i]$ |
| Nonlinear Control Systems | $\frac{d\vec{x}(t)}{dt} = \vec{f}(\vec{x}(t), \vec{u}(t))$ | $\vec{x}[i+1] = \vec{f}(\vec{x}[i], \vec{u}[i])$ |
| Autonomous Systems | $\frac{d\vec{x}(t)}{dt} = \vec{f}(\vec{x}(t))$ | $\vec{x}[i+1] = \vec{f}(\vec{x}[i])$ |

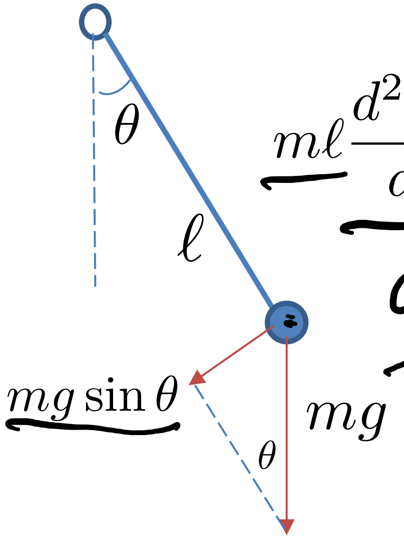


$$\vec{u}(t) = g(\vec{x}(t))$$

$$\frac{d\vec{x}(t)}{dt} = \vec{f}(\vec{x}(t), g(\vec{x}(t)))$$

$F = ma$

# Nonlinear Systems: Examples




$$\underline{m\ell}\underbrace{\frac{d^2\theta(t)}{dt^2}}_{a} = \underbrace{-k\ell\frac{d\theta(t)}{dt}}_{kv} - mg\sin\theta(t)$$

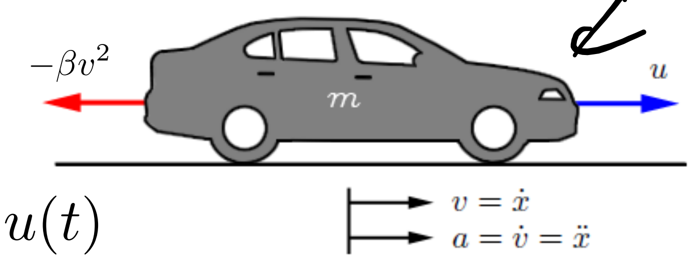
$x_1(t) = \theta(t)$

$x_2(t) = \dfrac{d\theta(t)}{dt}$

$$\begin{bmatrix} \dfrac{dx_1(t)}{dt} \\ \dfrac{dx_2(t)}{dt} \end{bmatrix} = \begin{bmatrix} x_2(t) \\ -\dfrac{g}{\ell}\sin x_1(t) - \dfrac{k}{m}x_2(t) \end{bmatrix}$$

$$\frac{d\vec{x}(t)}{dt} = \vec{f}(x_1(t), x_2(t)) \in \mathbb{R}^2$$

$ma = F$




$$m\frac{v(t)}{dt} = \underbrace{-\beta v(t)^2}_{drag} + u(t)$$

$x(t) = v(t)$

$$\frac{dx(t)}{dt} = -\frac{\beta}{m}x(t)^2 + \frac{1}{m}u(t)$$

$$= f(x(t), u(t)) \in \mathbb{R}^1$$

# Nonlinear Autonomous Systems: Equilibrium Points

$$\frac{d\vec{x}(t)}{dt} = \vec{f}(\vec{x}(t)) \quad \in \mathbb{R}^n$$

if $\vec{f}(\vec{x}^*) = \theta \in \mathbb{R}^n$, we call $\vec{x}^*$ equilibrium pts.

$\vec{x}(t) = \vec{x}^*$ $\quad \frac{d\vec{x}(t)}{dt} = \vec{f}(\vec{x}^*) = 0$

$$\vec{x}[i+1] = \vec{f}(\vec{x}[i]) \quad \in \mathbb{R}^n$$

$\vec{x}[i+1] = \vec{f}[\vec{x}[i]] = \vec{x}[i]$

an equilibrium pt. $\vec{x}^*$ is such that

$$\vec{f}(\vec{x}^*) = \vec{x}^*$$

fixed pt.

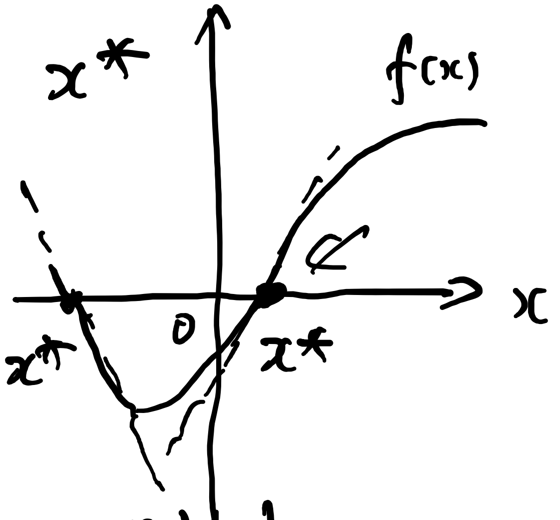
# Nonlinear Autonomous Systems: Linearization

**Scalar case:** $\dfrac{dx(t)}{dt} = f(x(t))$

at an equilibrium pt. $x^*$

$f(x^*) = 0$

around one $x^*$




① $f(x) = \underbrace{f(x^*)}_{0} + \boxed{f'(x^*)}\underbrace{(x - x^*)}_{\delta(t)} + \text{h.o.t.}$

$\rightarrow \delta(t) = x(t) - x^*$

$\boxed{\dfrac{d\delta(t)}{dt}} = \dfrac{d(x(t) - x^*)}{dt} = \dfrac{dx(t)}{dt} = \boxed{\overset{\lambda}{f'(x^*)}\,\delta(t)}$

**Vector case:** $\dfrac{d\vec{x}(t)}{dt} = \vec{f}(\vec{x}(t)) \in \mathbb{R}^n$

around some equilibrium point $\vec{x}^*$: $\vec{f}(\vec{x}^*) = 0$

$\vec{\delta}(t) = \vec{x}(t) - \vec{x}^*$

$\dfrac{d\vec{\delta}(t)}{dt} = \vec{f}(\vec{x}(t))$

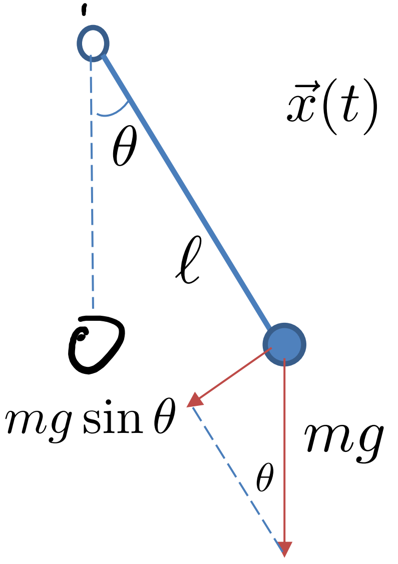
$= \boxed{\nabla_{\vec{x}} \vec{f}\,\Big|_{\vec{x}^*}}\,\vec{\delta}(t)$

$\dfrac{\partial \vec{f}}{\partial x}\Big|_{\vec{x}^*} = J_{\vec{x}} \vec{f}(\vec{x}^*)$

$$\vec{f} : \vec{x} \in \mathbb{R}^n \to \mathbb{R}^n \quad \begin{bmatrix} f_1(x_1, \ldots, x_n) \\ \vdots \\ f_n(x_1, \ldots, x_n) \end{bmatrix} = \vec{f}(\vec{x})$$

$$\left.\frac{\partial \vec{f}}{\partial \vec{x}}\right|_{\vec{x}^*} = \left.\begin{bmatrix} \frac{\partial f_1}{\partial x_1}, & \ldots, & \frac{\partial f_1}{\partial x_i} \\ \vdots & & \\ \frac{\partial f_n}{\partial x_i}, & \ldots, & \frac{\partial f_n}{\partial x_n} \end{bmatrix}\right|_{\vec{x}^*} = \nabla_{\vec{x}} f(\vec{x}^*) = J_{\vec{x}} f(\vec{x}^*)$$

# Nonlinear Autonomous Systems: Example




$$\vec{x}(t) = \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} = \begin{bmatrix} \theta(t) \\ \frac{d\theta(t)}{dt} \end{bmatrix} \in \mathbb{R}^2 \quad \begin{cases} \dfrac{dx_1(t)}{dt} = x_2(t) & = f_1(x_1(t), x_2(t)) \\ \dfrac{dx_2(t)}{dt} = \underline{-\dfrac{g}{\ell}\sin(x_1(t))} - \dfrac{k}{m}x_2(t) & = \underline{f_2(x_1(t), x_2(t))} \end{cases}$$

① equilibrium pt.

$\vec{f}(\vec{x}^*) = 0$

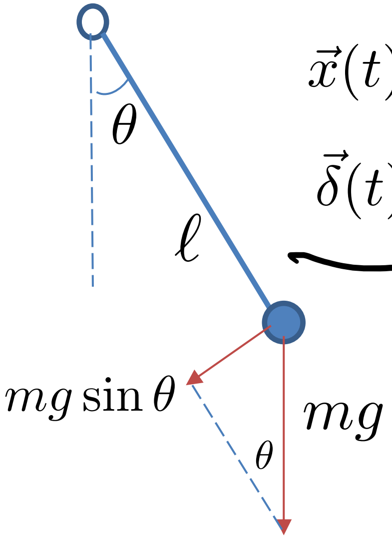
$f_1(x_1, x_2) = 0 \Rightarrow x_2 = 0$

$f_2(x_1, x_2) = -\frac{g}{\ell}\sin x_1 = 0 \Rightarrow x_1 = 0, \pi$

$\vec{x}^*$ $\underline{(0, 0)}$, $\underline{(\pi, 0)}$

points of interest

# Nonlinear Autonomous Systems: Example




$$\vec{x}(t) = \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} = \begin{bmatrix} \theta(t) \\ \frac{d\theta(t)}{dt} \end{bmatrix}$$

$$\vec{\delta}(t) = \vec{x}(t) - \vec{x}^*$$

$$\frac{d\vec{\delta}(t)}{dt} = \frac{d\vec{x}(t)}{dt} = \vec{f}(\vec{x}(t)) \approx \vec{f}(\vec{x}^*) + J_{\vec{x}}\vec{f}(\vec{x})\Big|_{\vec{x}^*}\vec{\delta}(t)$$

$$J_{\vec{x}}\vec{f}(\vec{x})\,|_{(x_1^*,x_2^*)} = \begin{bmatrix} \frac{\partial f_1}{\partial x_1}(x_1^*, x_2^*) & \frac{\partial f_1}{\partial x_2}(x_1^*, x_2^*) \\ \frac{\partial f_2}{\partial x_1}(x_1^*, x_2^*) & \frac{\partial f_2}{\partial x_2}(x_1^*, x_2^*) \end{bmatrix} \in \mathbb{R}^{2\times 2}$$

$$\frac{d\vec{\delta}(t)}{dt} = \boxed{J_{\vec{x}}\vec{f}(\vec{x})\Big|_{\vec{x}^*}}\,\vec{\delta}(t)$$

$$\frac{d\vec{\delta}(t)}{dt} = A\,\vec{\delta}(t)$$

① $\vec{x}^* = (0, 0)$, $J_{\vec{x}}\vec{f}\,\Big|_{\vec{x}^*} = \begin{bmatrix} 0 & 1 \\ -\frac{g}{\ell}\cdot 1 & -\frac{k}{m} \end{bmatrix}$ $A_1$

② $\vec{x}^* = (\pi, 0)$, $J_{\vec{x}}\vec{f}\,\Big|_{\vec{x}^*} = \begin{bmatrix} 0 & 1 \\ +\frac{g}{\ell} & \frac{-k}{m} \end{bmatrix}$ $A_2$

$A_1$ ? $\quad tr(A_1) < 0 \quad det(A_1) = \frac{g}{l} > 0$

$A_2$ ? $\quad tr(A_2) < 0, \quad det(A_2) = \frac{-g}{l} < 0$

$$(\lambda - \lambda_1)(\lambda - \lambda_2)$$
$$= \lambda^2 - \underline{(\lambda_1 + \lambda_2)}\lambda + \underline{\lambda_1 \lambda_2}$$

$$\lambda I - A = \underline{\begin{bmatrix} \lambda - a_{11} & -a_{12} \\ -a_{21} & \lambda - a_{22} \end{bmatrix}}$$

$$\det(\lambda I - A) =$$
$$= \lambda^2 - \underbrace{\boxed{(a_{11} + a_{22})}}_{\substack{tr(A) \\ < 0}} \lambda + \underbrace{(a_{11}a_{22} - a_{12}a_{21})}_{\substack{det(A) \\ > 0}}$$

# Nonlinear Control Systems: Operating Points

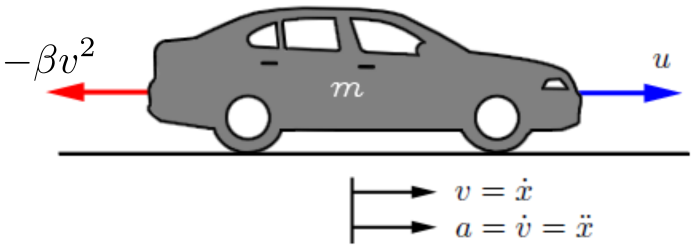
$$\frac{d\vec{x}(t)}{dt} = \vec{f}(\vec{x}(t), \vec{u}(t)) \in \mathbb{R}^n$$

$$\vec{x}[i+1] = \vec{f}(\vec{x}[i], \vec{u}[i]) \in \mathbb{R}^n$$

# Nonlinear Control Systems: Linearization

**Scalar case:** $$\frac{dx(t)}{dt} = f(x(t), u(t))$$

# Nonlinear Control Systems: Example




$$\frac{dx(t)}{dt} = -\frac{\beta}{m}x(t)^2 + \frac{1}{m}u(t) = f(x(t), u(t))$$

$$m\frac{v(t)}{dt} = -\beta v(t)^2 + u(t)$$