Welcome to

2.160

Identification, Estimation, and Learning

3-0-9 H-Level Graduate Credit

Prerequisite: 2.151 (2.14/2.140) or similar subject

Professor H. Harry Asada

Ford Professor of Mechanical Engineering

asada@mit.edu , Room 3-346, x3-6257

TA: Nicholas Selby, nselby@mit.edu

H. Harry Asada

- Regularly teaches
  - 2.12/2.120 Introduction to Robotics
- Has taught
  - 2.151 Advanced System Dynamics and Control
  - 2.165 Robotics
  - 2.004 Dynamics and Control
  - 2.03J (2.003) Dynamics
  - 2.14 Feedback Control
  - 2.671 Measurement and Instrumentation
  - 2.86 (2.008) Manufacturing
- Specializes in Robotics, Control, and Biological Engineering

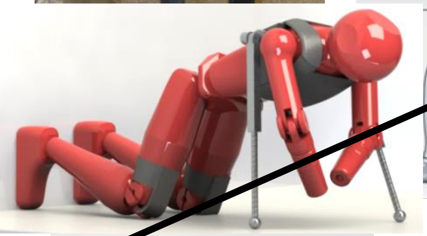
Quick introduction of the instructor:

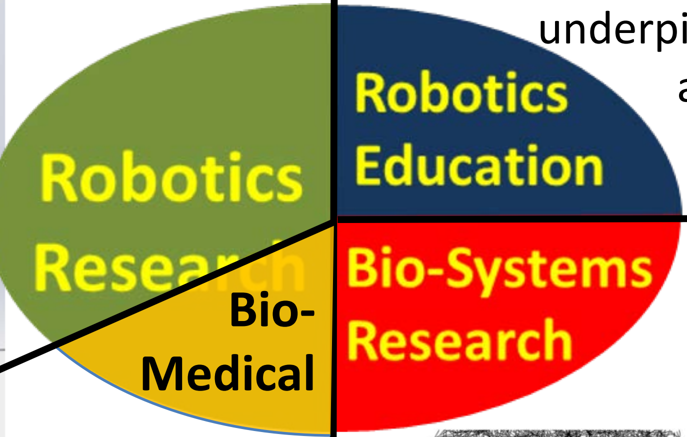
H. Harry Asada

Ford Professor of Engineering

Wearable Extra Limbs

Human Augmentation: People can possess extra arms, legs, and fingers for augmenting and compensating for the physical and cognitive abilities.





Robotics Research

Robotics Education

Bio-Systems Research

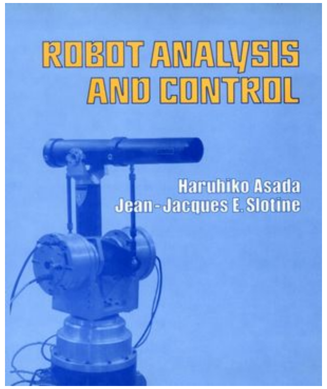
Bio-Medical

Wearable Health Monitoring

MIT Ring

Introduction to Robotics

Robotics provides students with clear, tangible, and graphical understanding of complex motion and underpinning math and physics.



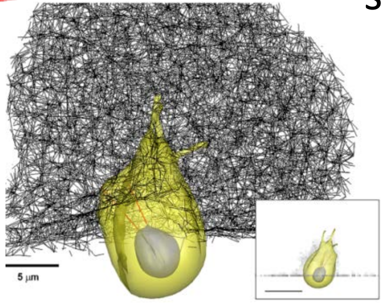
ROBOT ANALYSIS AND CONTROL

Haruhiko Asada
Jean-Jacques E. Slotine

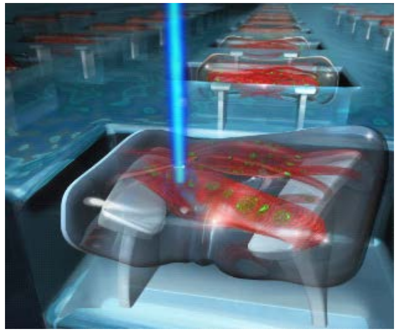
2.12:
Multi-disciplinary capstone course

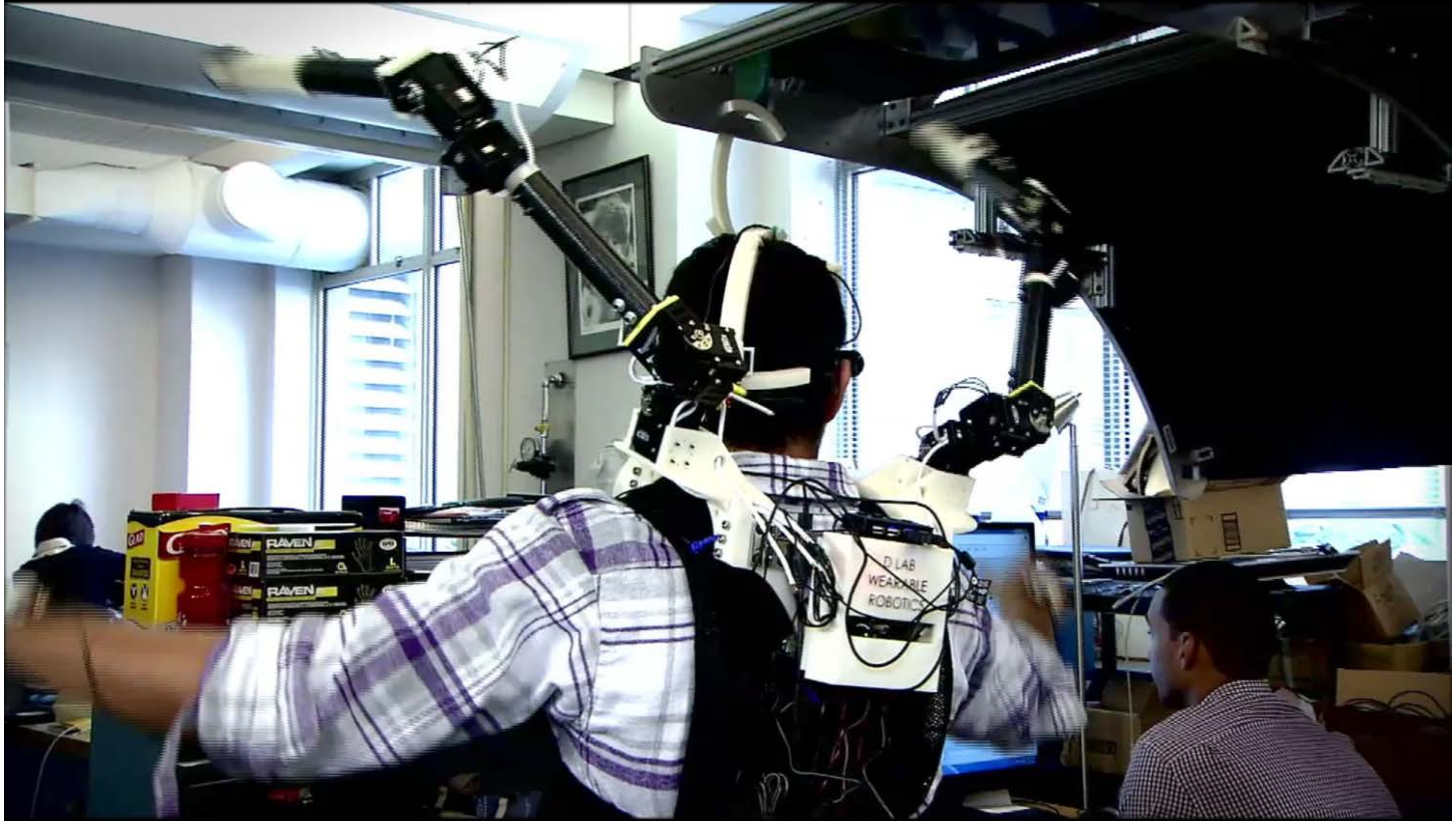
Integrated cellular systems

Optogenetic control of skeletal muscles



Modeling & Computation of 3D cell migration




D LAB
WEARABLE
ROBOTICS
RAVEN
RAVEN
Glad

Harry Asada
An inventor

Biologically-Inspired Actuators
US Patent 7,188,473

Ball-Wheel
Holonomic Wheelchair
US Patent 5,927,423

Fingernail
Sensors
US Patent 6,236,037
US Patent 6,388,247

RHOMBUS
Hybrid Bed/Wheelchair
US Patent 6,135,228

Direct-Drive Robot
US Patent 4,425,818

Surface Wave Actuators
US Patent 5,953,773

High Strain PZT
Modular Actuators

Broadcast Feedback

Wireless Networking
US Patent 6,553,535
US Patent 7,376,105

ABP Estimation
US Patent 6,413,223
US Patent 7,169,111

Ring Sensor
US Patent 6,402,690
US Patent 5,964,701
US Patent 6,699,199

Wearable
Goniometry

Health Chair
US Patent 6,947,781

Cable-Free Smart
Vest

Driver
Monitoring

Wearable Health
Monitoring

Teaching Style:
Theory for the User

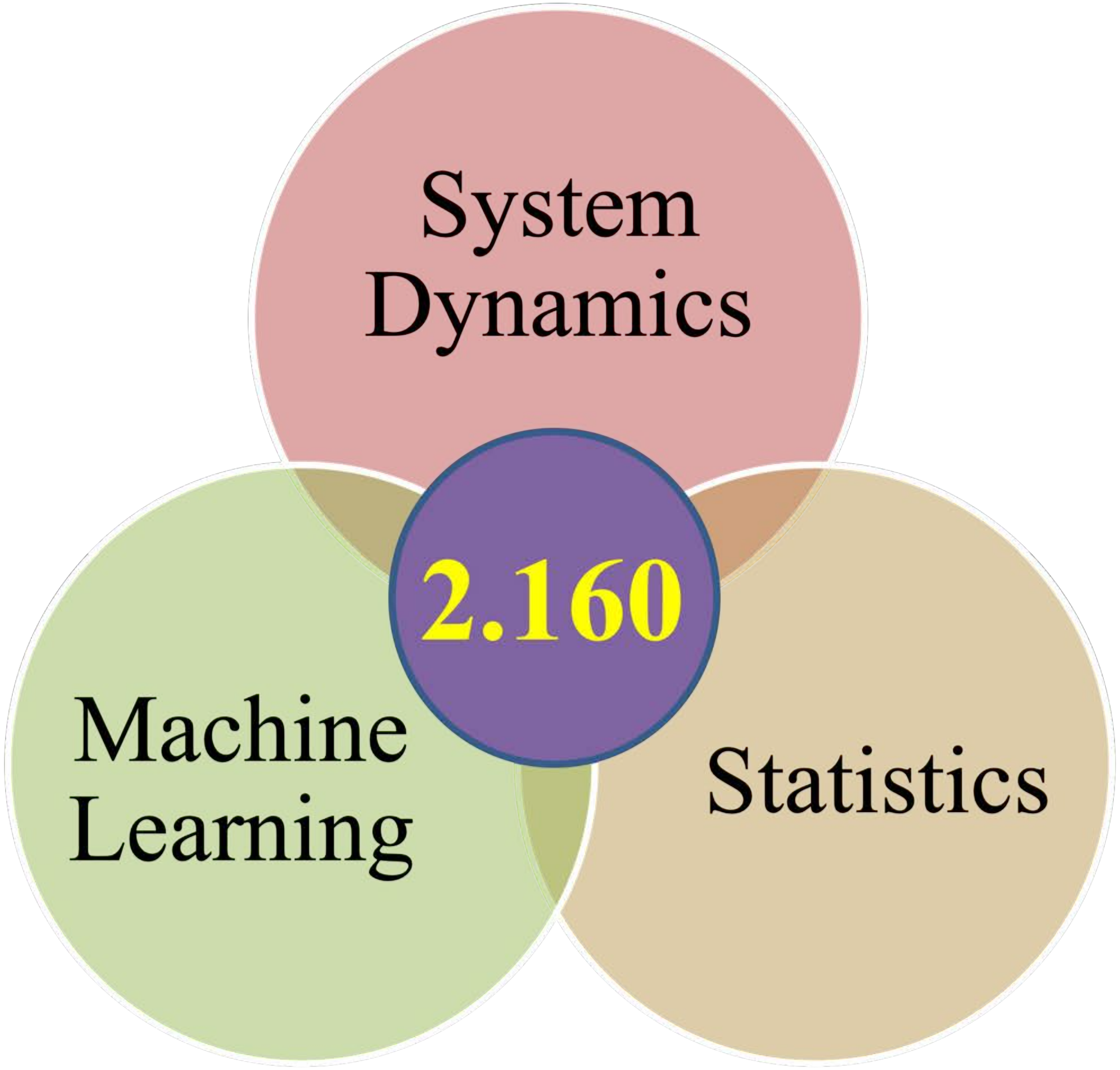
WHAT IS 2.160?

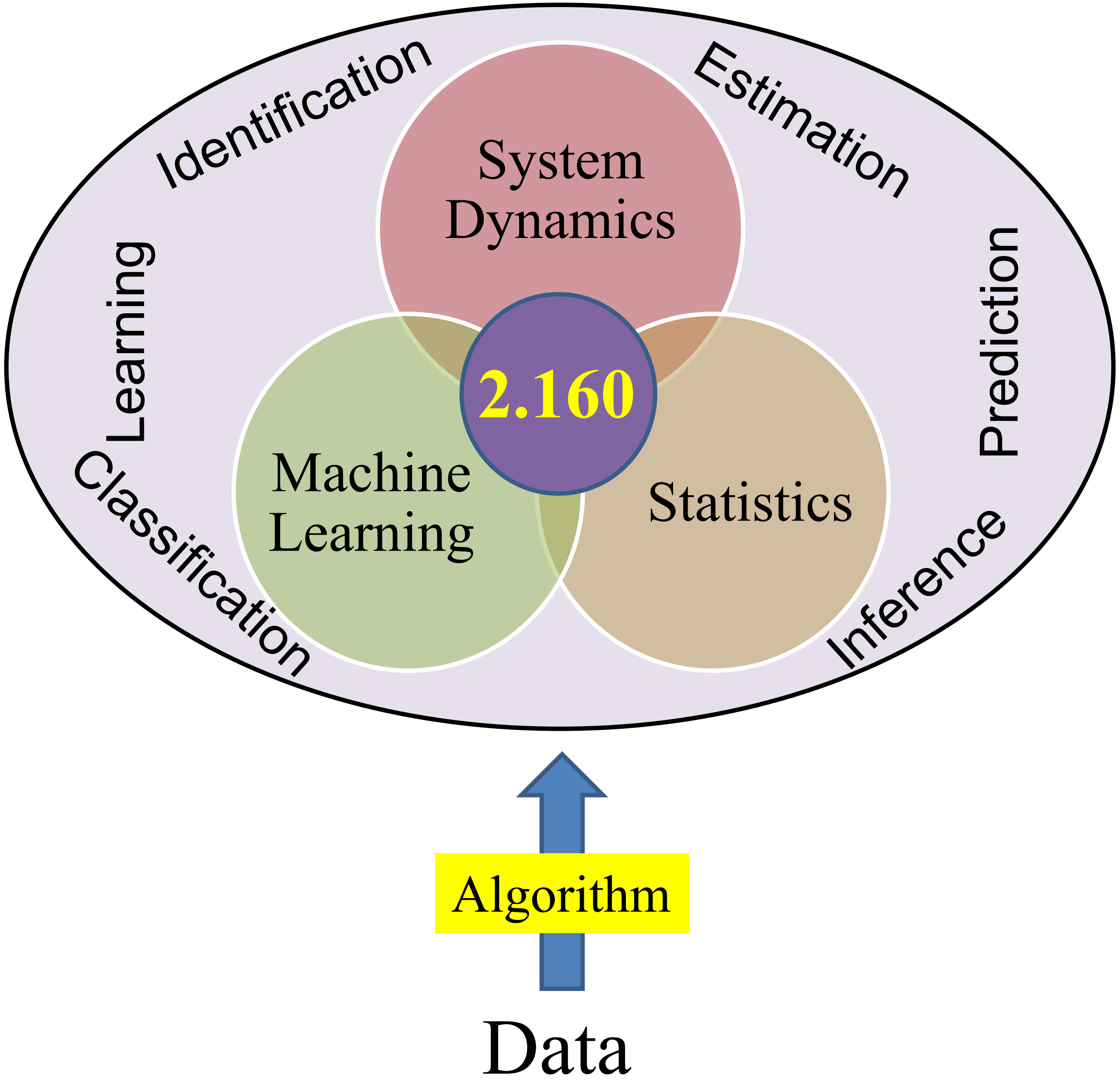
- Not a fancy subject, like robotics and design subjects.
- But, if you wish to learn something fundamental, establish a sold foundation, or apply an analytical and/or mathematical methodology to your thesis research, you will find 2.160 to be a useful subject.

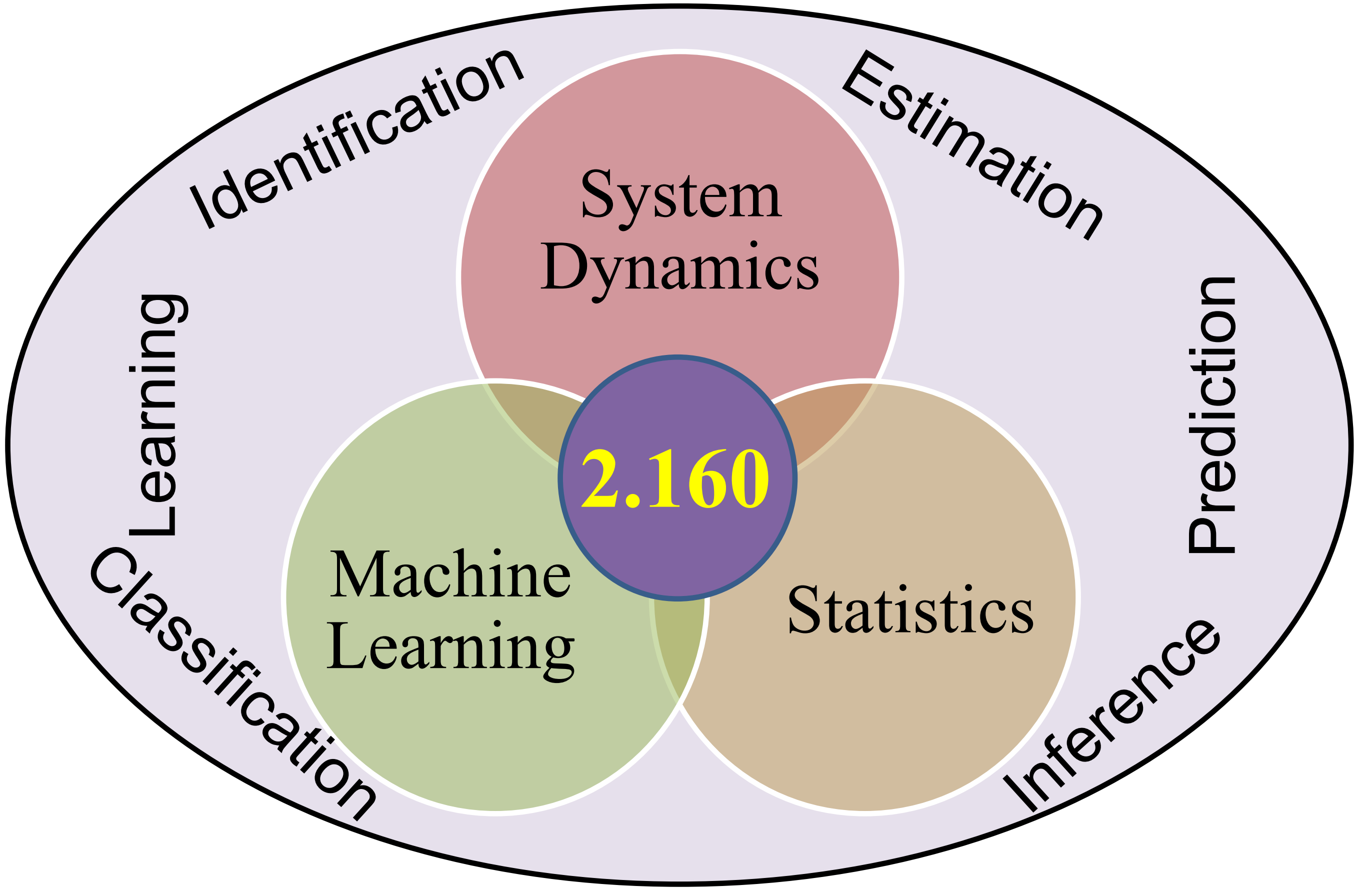
Goal of the Subject

2.160 seeks deeper understanding, clear insights, scientifically sound methodologies, and practically useful techniques for modeling, estimation, and learning.

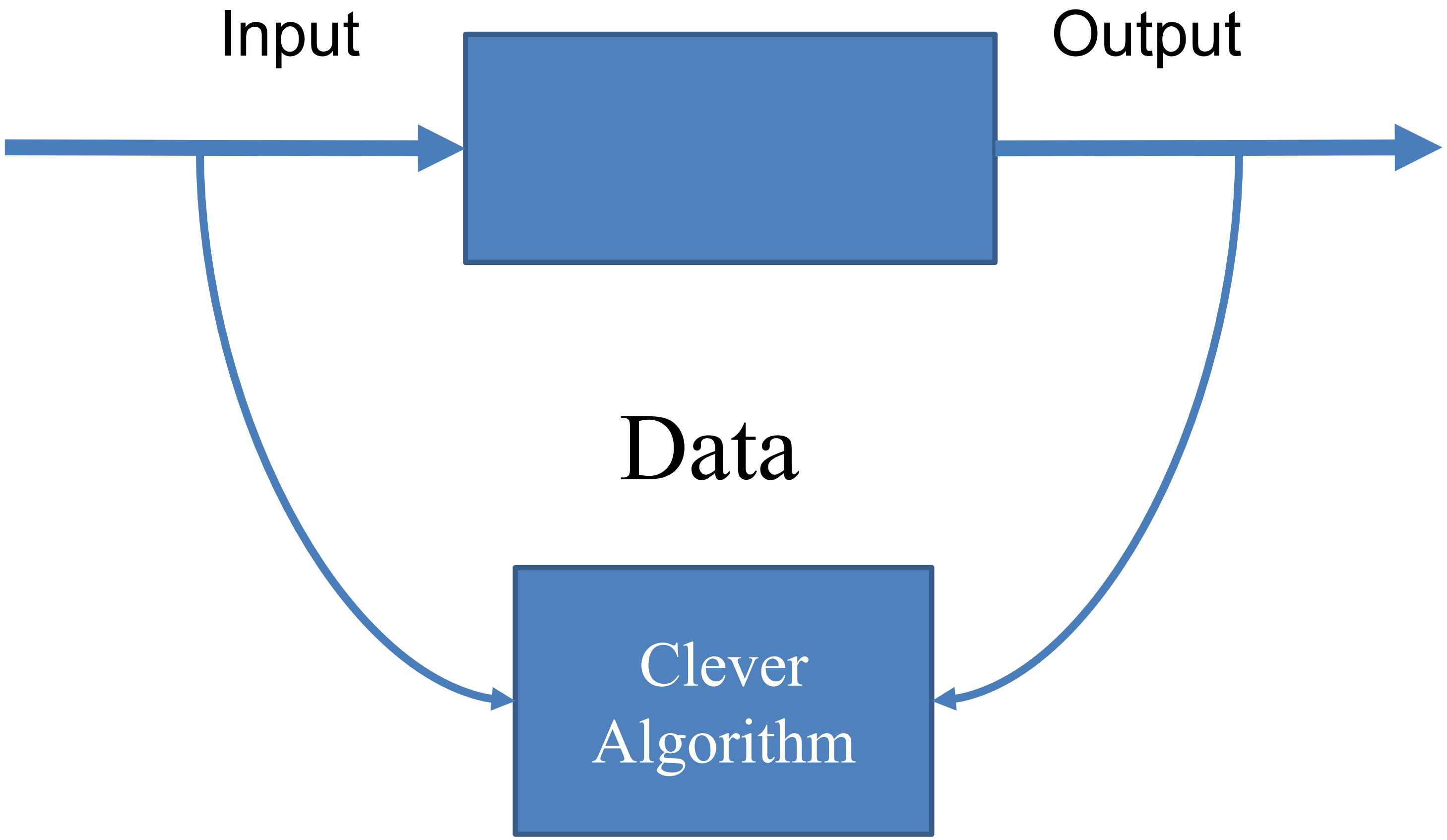
Cross-Disciplinary Study




Identification
Estimation
Learning
Prediction
Classification
Inference
System Dynamics
Machine Learning
Statistics
2.160
Algorithm
Data



2.160 addresses diverse topics in a cohesive manner. These include system identification, estimation, prediction, inference, classification, and learning. Although the objectives are different among these topics, the underpinning theories, techniques, and algorithms are common to them. These have been established at the cross-disciplinary area of system dynamics, machine learning, and statistics.

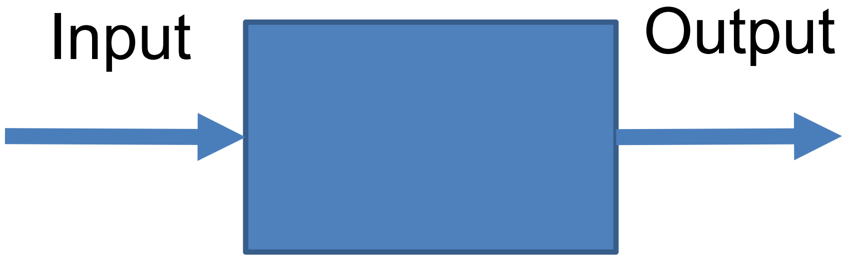


Input

Output

Data

Clever Algorithm

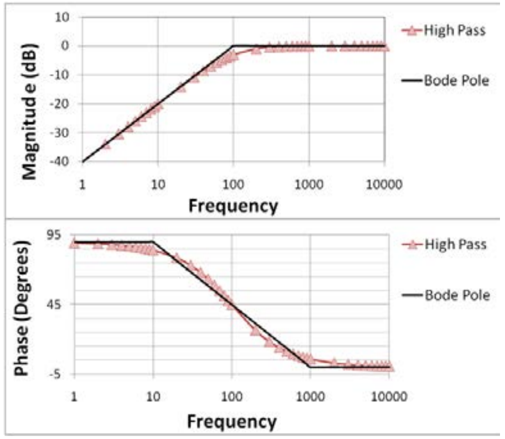


Input

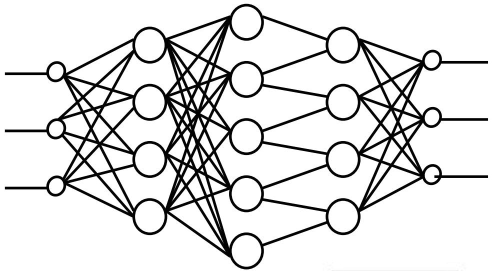
Output

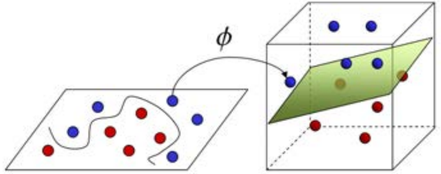
$$G(q)=\frac{b_1 q^{-1}+b_2 q^{-2}+\cdots+b_{n_b} q^{-n_b}}{1+a_1 q^{-1}+a_2 q^{-2}+\cdots+a_{n_a} q^{-n_a}}$$

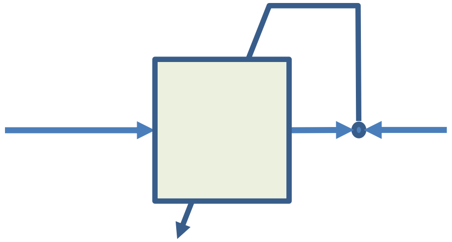
$$x(t+1)=Ax(t)+Bu(t)+\eta(t)$$
$$y(t)=Cx(t)+Du(t)+\upsilon(t)$$



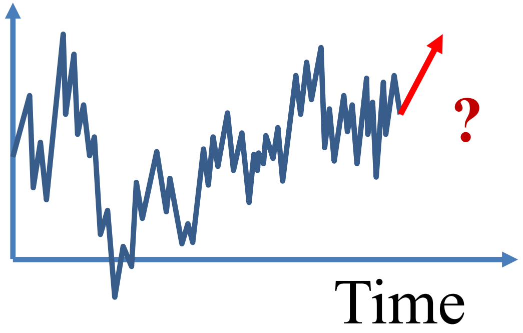
System Identification

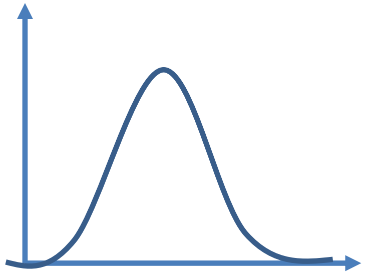




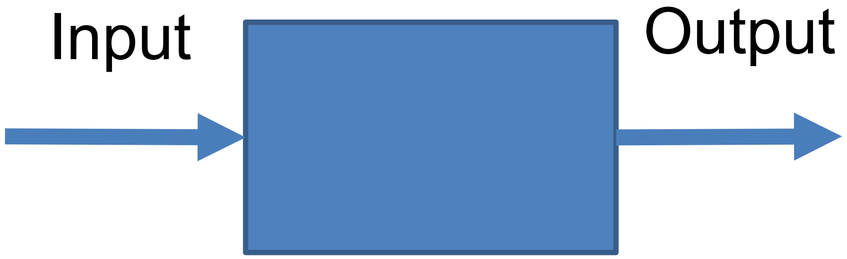


Learning



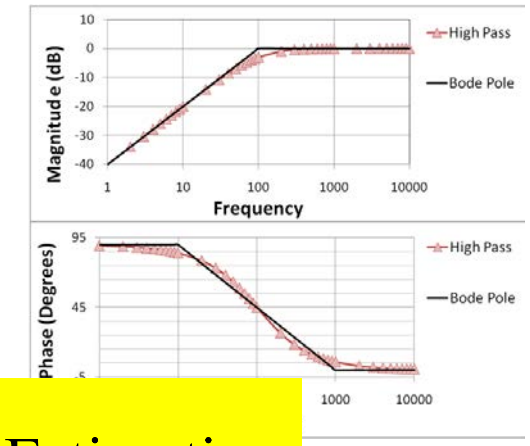


Prediction



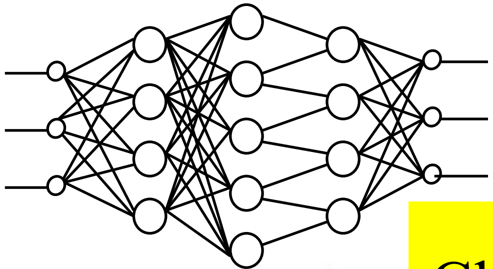
$$G(q) = \frac{b_1 q^{-1} + b_2 q^{-2} + \cdots + b_{n_b} q^{-n_b}}{1 + a_1 q^{-1} + a_2 q^{-2} + \cdots + a_{n_a} q^{-n_a}}$$

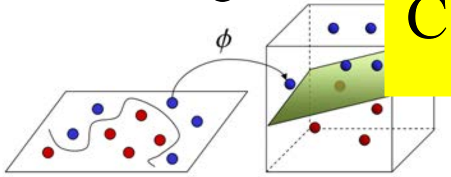
$$x(t+1) = Ax(t) + Bu(t) + \eta(t)$$
$$y(t) = Cx(t) + Du(t) + \upsilon(t)$$

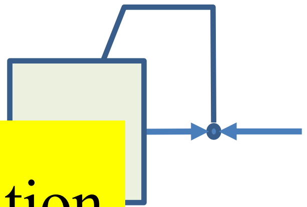


System Identification

Estimation

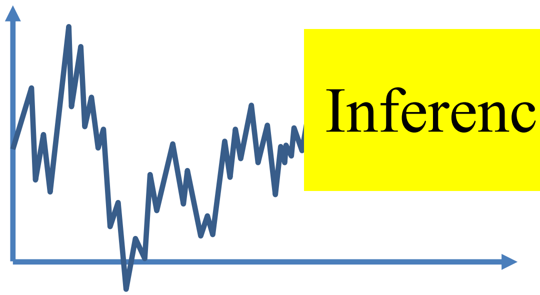




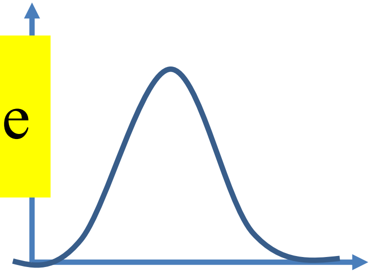


Learning

Classification



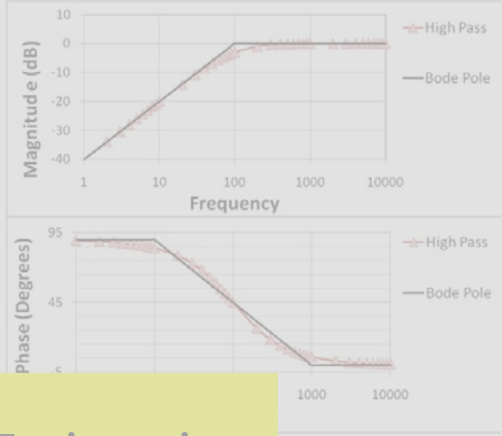
Inference



Prediction

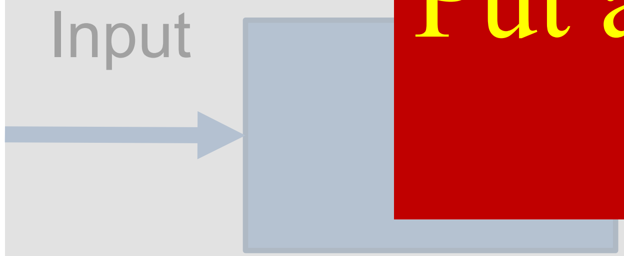
$$G(q)=\frac{b_1q^{-1}+b_2q^{-2}+\cdots+b_{n_b}q^{-n_b}}{1+a_1q^{-1}+a_2q^{-2}+\cdots+a_{n_a}q^{-n_a}}$$

$$x(t+1)=Ax(t)+Bu(t)+\eta(t)$$
$$y(t)=Cx(t)+Du(t)+\upsilon(t)$$



System Identification

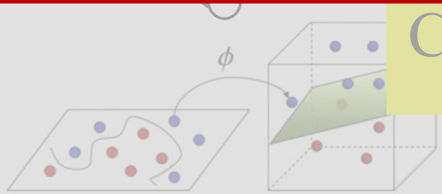
Estimation



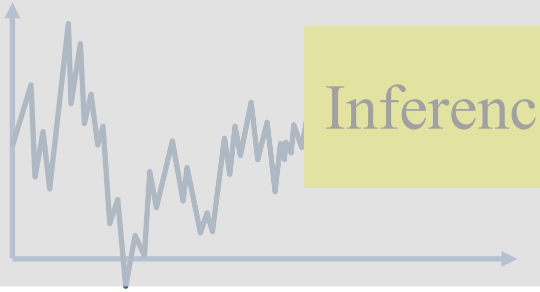
Put all in one subject, 25 lectures

2.160

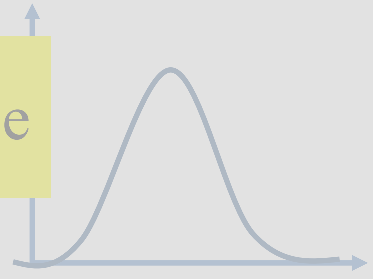
Learning



Classification



Inference



Prediction

General Course Information

Class

- Synchronous Lecture on Zoom
- Monday and Wednesday: 1:00 pm – 2:30 pm
- All lectures are recorded and posted for your review.
- Lecture slides will also be posted.

Lecture Notes

- Will be provided for (almost) every lecture
- Intuitive and helpful for understanding fundamental concepts
- Intensive and extensive: covering a lot of topics
- Examples
- Background materials and review
- Read them before going to reference books

No Formal Exams

- You will not have formal exams: no mid-term, no final exam;
- Instead, you will work on 4 Context-Oriented Mini Projects and 6 homework assignments.

Grading

Context-Oriented Projects* (4 projects)	50%
Homework Assignment* (6 assignments)	40 %
Participation**	10%
Total	100%

* Assignments will be given mostly on Monday, and will be due in the following week.

** Participation in study group meetings and active engagement with lectures

Study Group Meetings

- Student groups, each consisting of 6~8 students, will be formed.
- Each group will meet weekly on Friday for
  - Discussing Context-Oriented Mini Projects and Homework Assignments, and
  - Review and recitation of lecture materials
- TA, Nick Selby, and/or Professor Asada will participate in each study group meeting.
- Schedule will be discussed later.

Ethics

- Use of problem set solutions of previous terms is strictly prohibited.
- Students are encouraged to discuss problem assignments with one another. However, each student must submit his/her own solution to each problem set and mini project.

In Case of Difficulty,…

- Contact Professor Asada and the Student Support Service office.
- One assignment relief: You can skip one homework assignment without penalty.
- Late submission of one Mini Project without penalty
- Prior notification to Professor Asada is required.
- Your mental and physical health is more important than your work performance.

PS Grading

Eliminate the lowest score PS.

$$\underbrace{75,\ 85,\ 90,\ 67,\ 65,}_{\text{Counted toward the final grade.}}\text{and } \cancel{25}$$

Learning Management System

- We will use Canvas + Panopto
  - These tools are still relatively new to the MIT community.
  - We will also use Stellar in the beginning. All the lecture notes, handouts, lecture slides, assignments will be posted on Stellar.
  - We will move to Canvas as we get familiarized with the new system. This may occur before the first due date of assignment.
  - TA will make clear announcements what to do.

Perspective of 2.160

History, Theory, Key Concept, Application, Projects

2.160

Part 1: Regression

…5 lectures

Part 2: Kalman and Bayes Filters

…6 lectures

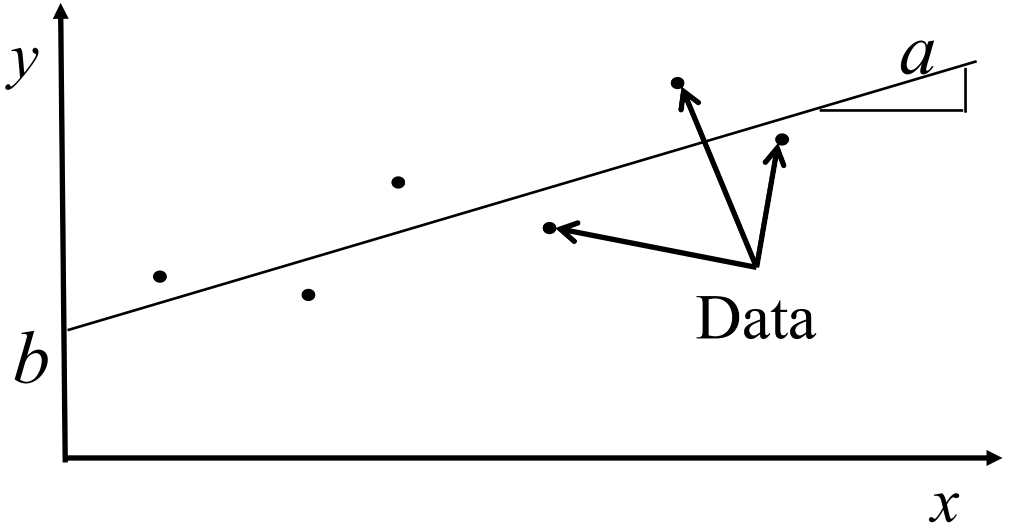
Part 3: System Identification of Linear Dynamical Systems

…6 lectures

Part 4: Machine Learning and Nonlinear System Modeling

…7 lectures

Part 1: Regression



Output

Input

$$y = ax + b$$

Parameters to estimate

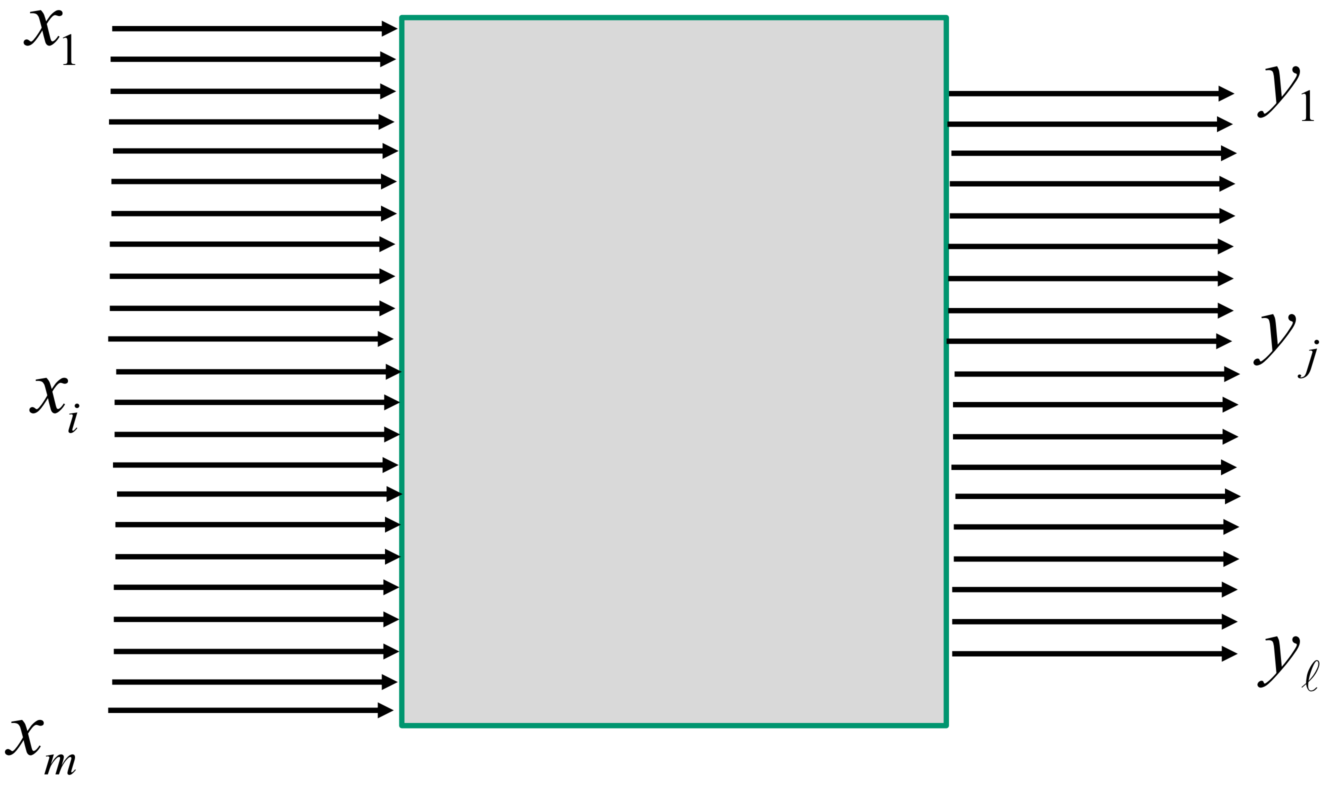
Find the parameter values that minimize the total squared error:

$$V_N(\theta) = \frac{1}{N}\sum_{t=1}^{N}(\hat{y}(t \mid \theta) - y(t))^2 \Rightarrow \min$$

Least Squares Estimation

Find input-output relationship.

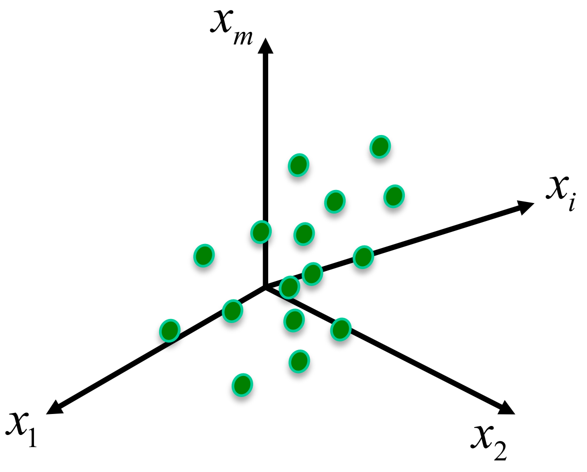
Challenge: High Dimensional Space

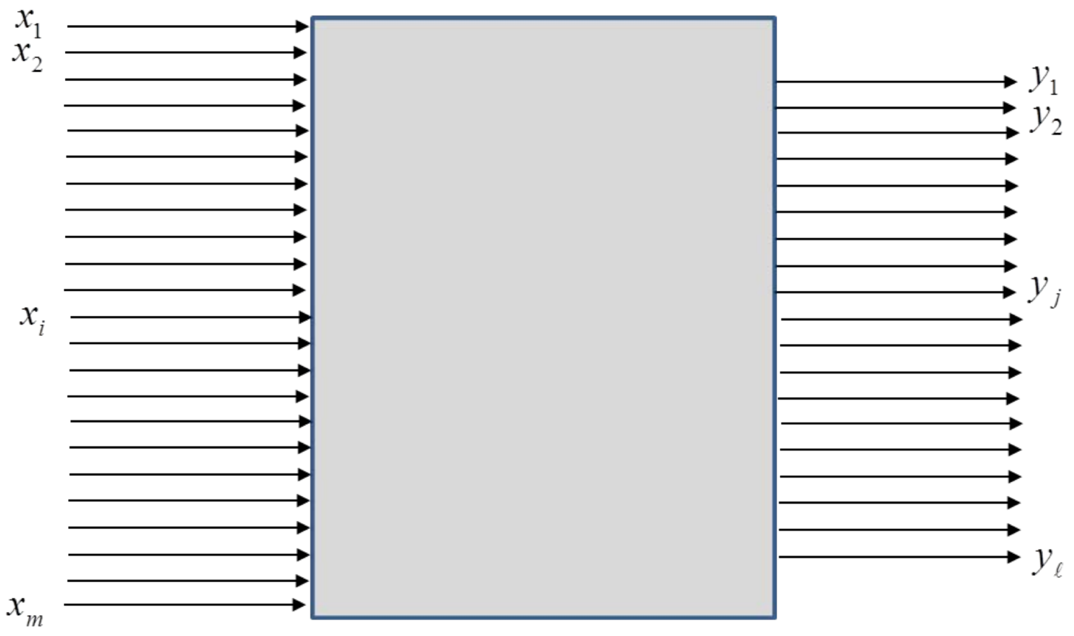


$$y_j = a_{j1}x_1 + a_{j2}x_2 + \cdots\cdots\cdots + a_{jm}x_m$$

Estimate parameters $a_1, \cdots\cdots, a_m$: an ill-posed question(?)

Input variables may be collinear;
The input data set may not contain samples in some directions in the input space

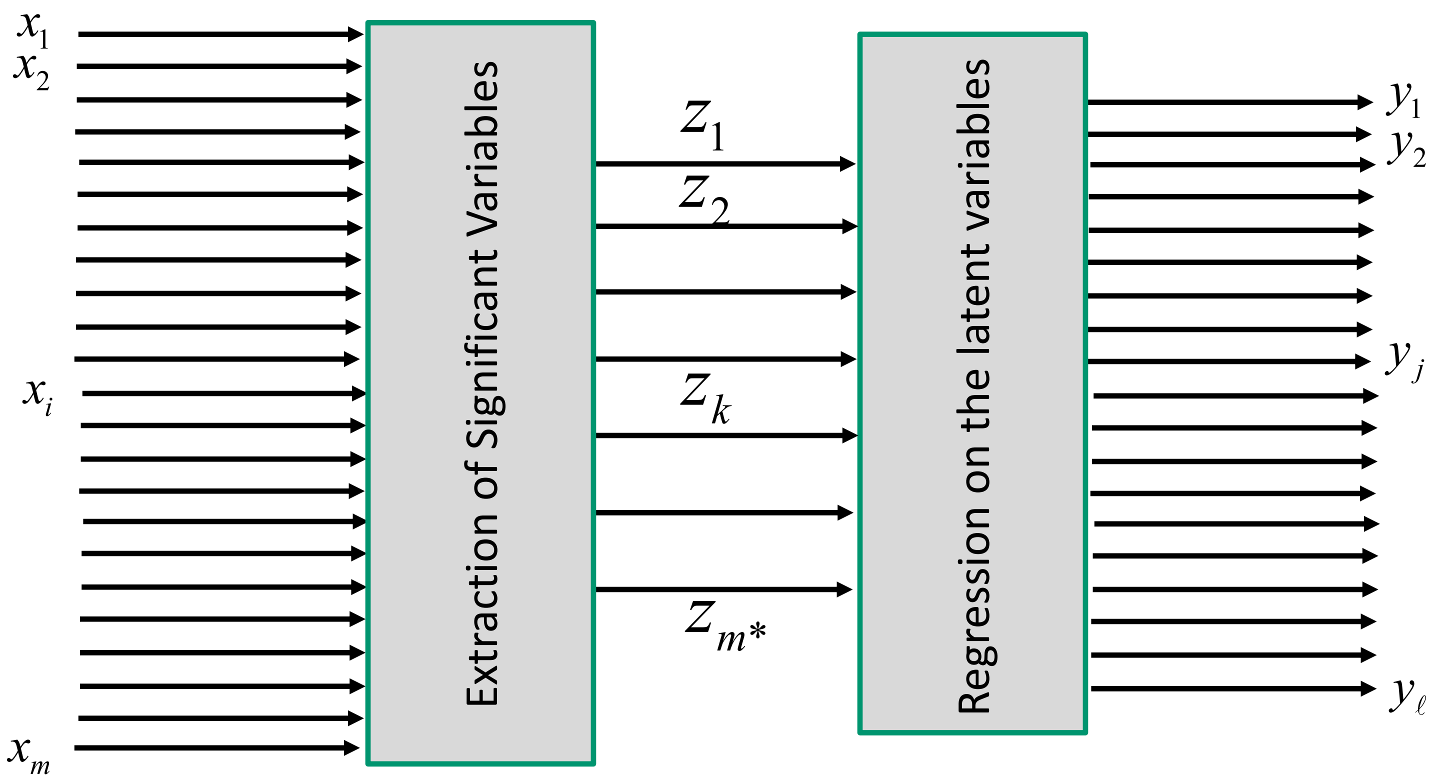




$$y_j = a_{j1}x_1 + a_{j2}x_2 + \cdots\cdots\cdots + a_{jm}x_m$$

Estimate parameters $a_{j1},\cdots\cdots,a_{jm}$: an ill-posed question(?)

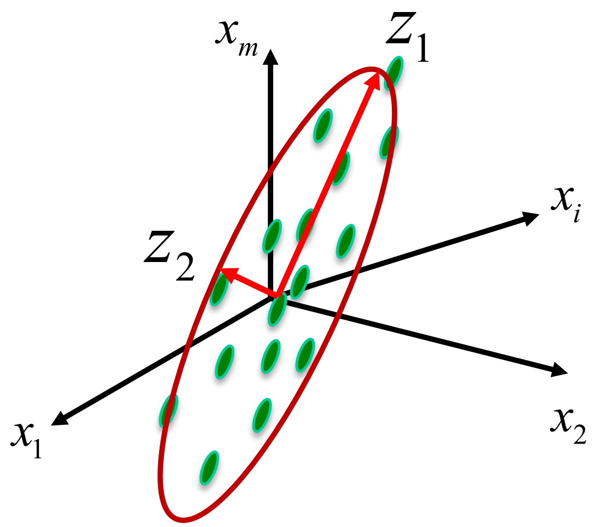
Extracting $\underbrace{\text{significant variables}}_{\textbf{\textit{Latent Variables}}}$ from the input data


x_1
x_2
x_i
x_m
Extraction of Significant Variables
z_1
z_2
z_k
z_{m*}
Regression on the latent variables
y_1
y_2
y_j
y_ℓ

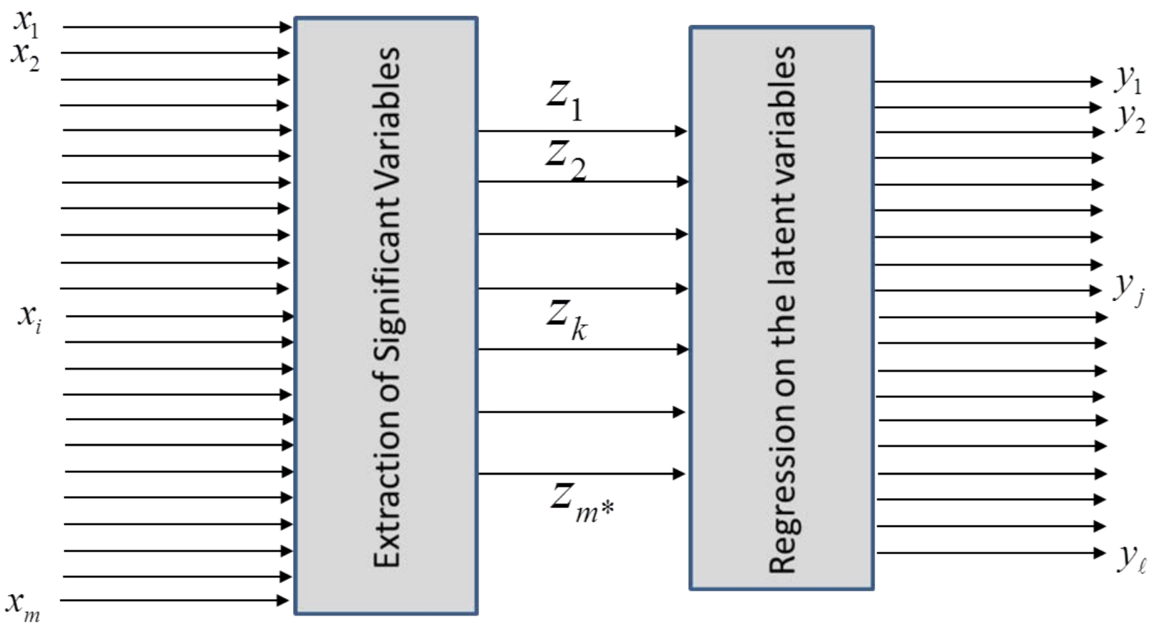
Regress output y on latent variables, $z_1, \cdots\cdots, z_{m*}$

$$y_j = b_{j1} z_1 + \cdots\cdots\cdots + b_{m*} z_{m*} \qquad m^* << m$$

Principal Component Regression: an example of latent variable method



Principal Components Analysis



Extraction of Significant Variables

Regression on the latent variables

Mean-centered Input Data Set $X=\left[x^1,\cdots,x^N\right]\in\Re^{m\times N}$ $x=\left(x_1,\cdots,x_m\right)^T$

Covariance

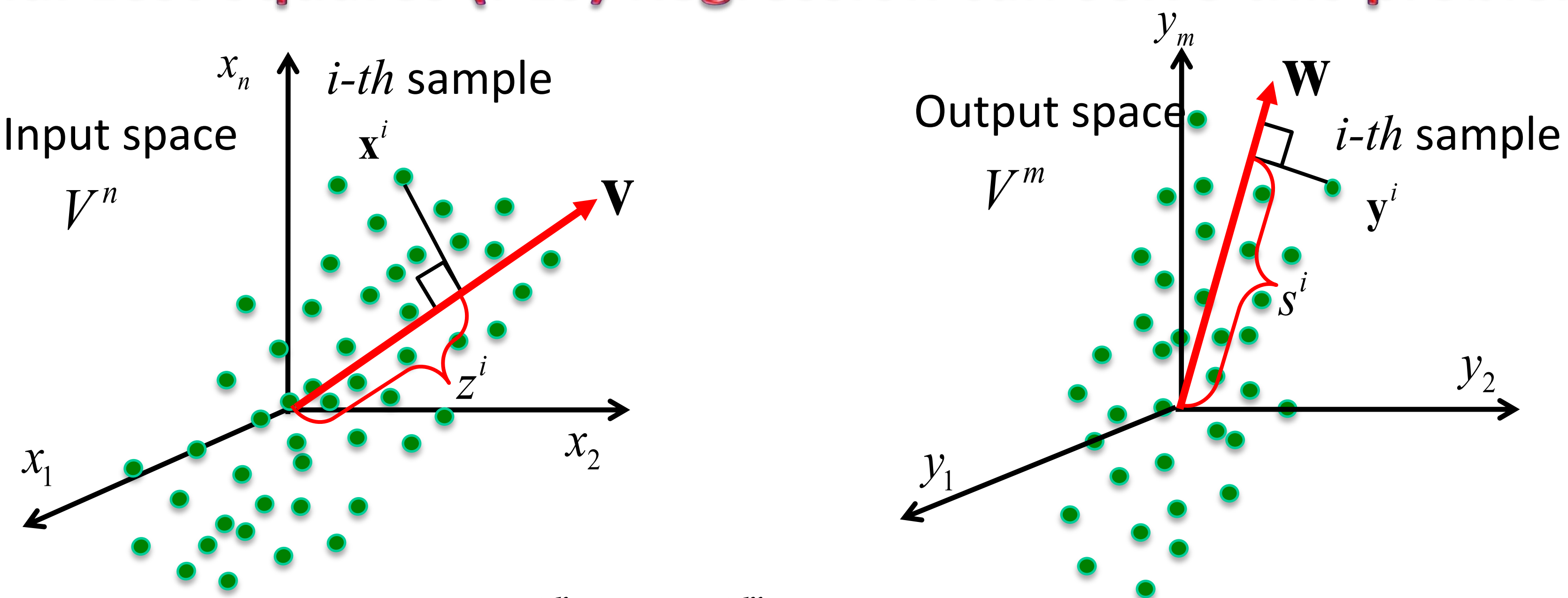
$$\lambda_1\ge\lambda_2\ge\cdots\ge\lambda_m\ge 0$$

$$XX^T=\left(v_1\ v_2\cdots v_m\right)\begin{pmatrix}\lambda_1&0&0\\0&\ddots&0\\0&0&\lambda_m\end{pmatrix}\begin{pmatrix}v_1^T\\\vdots\\v_m^T\end{pmatrix}=\lambda_1v_1v_1^T+\cdots+\lambda_{m*}v_{m*}v_{m*}^T\,\Big|\,+\cdots+\lambda_mv_mv_m^T$$

Truncate

A drawback of Principal Component Regression is that Components with a large eigenvalue in the input space are not necessarily significant for predicting the output; small eigenvalue components may be more correlated with the output.

Partial Lest Squares (PLS) Regression can solve this problem.



Find a pair of unit vectors, $\mathbf{v} \in \Re^n$, $\mathbf{w} \in \Re^m$, such that the correlation between $z = \mathbf{v}^T \mathbf{x}$ and $s = \mathbf{w}^T \mathbf{y}$ becomes maximum.

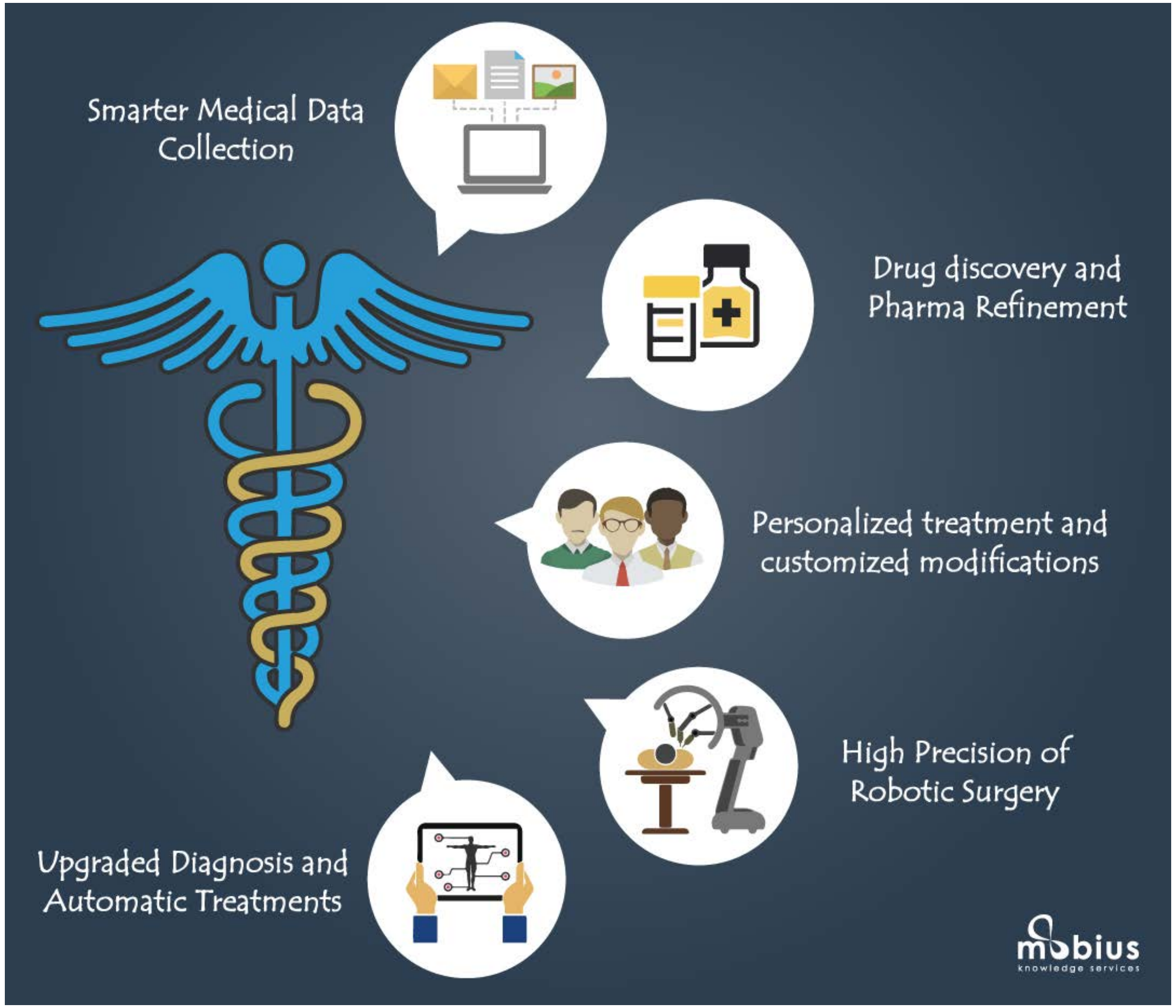
$$\left.\begin{matrix}\mathbf{v}\\ \mathbf{w}\end{matrix}\right\} = \underset{\substack{|\mathbf{v}|=1,\\ |\mathbf{w}|=1}}{\arg\max}\; E[z \cdot s] = \underset{\substack{|\mathbf{v}|=1,\\ |\mathbf{w}|=1}}{\arg\max}\; \mathbf{v}^T E[\mathbf{x}\mathbf{y}^T]\mathbf{w}$$

The optimal $\mathbf{v}$ and $\mathbf{w}$ are the left and right singular vectors associated with the largest singular value of Cross-Covariance

Singular-Value Decomposition

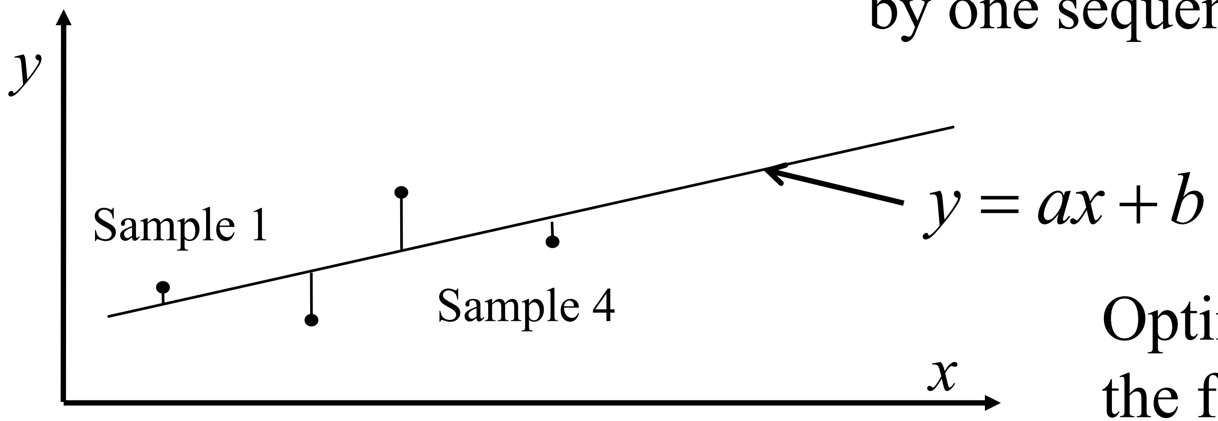
$$C_{XY} = E[\mathbf{x}\mathbf{y}^T] \rightarrow \text{SVD}$$

Where are these methods used?

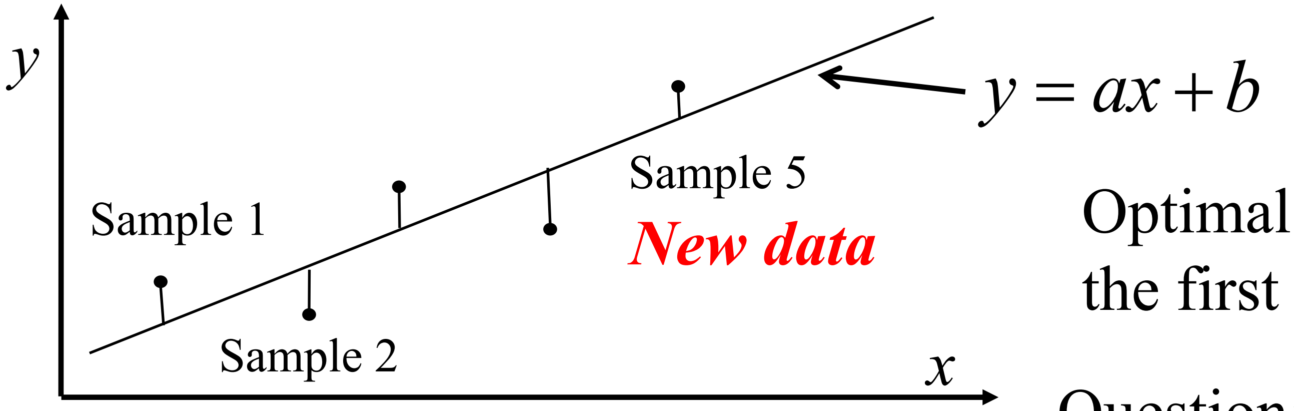

Smarter Medical Data Collection
Drug discovery and Pharma Refinement
Personalized treatment and customized modifications
High Precision of Robotic Surgery
Upgraded Diagnosis and Automatic Treatments
mobius knowledge services

Recursive Estimation

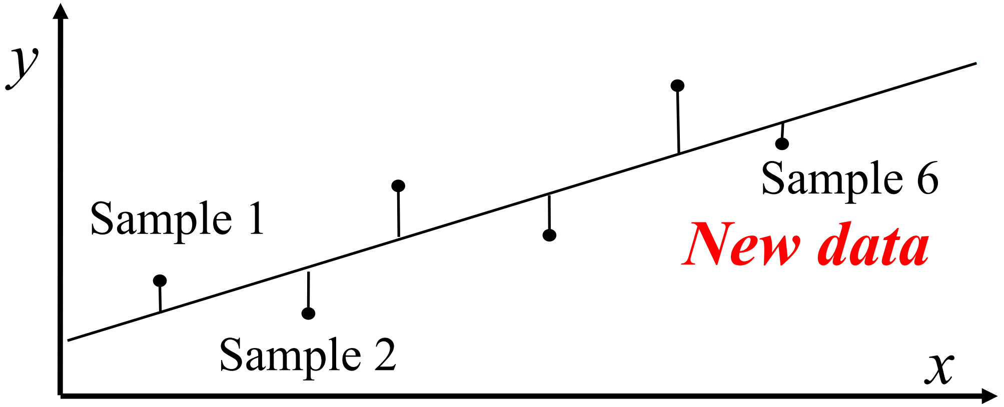
What if the data are obtained one by one sequentially?



Optimal a and b based on the first 4 samples



Optimal a and b based on the first 5 samples



Question: Do I need to compute a and b from scratch?

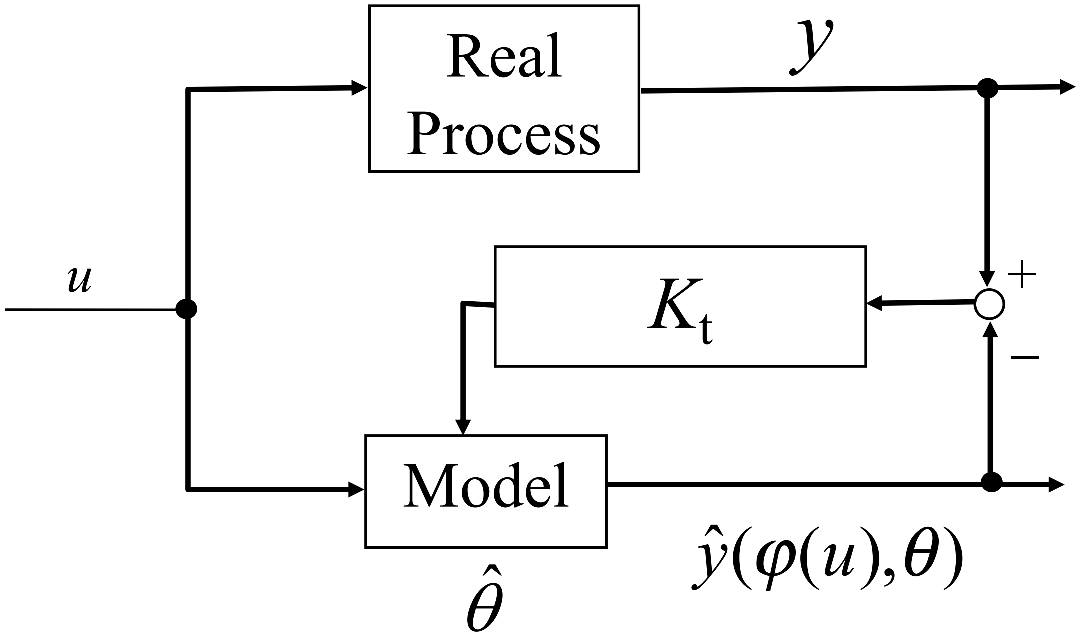
Answer: No, only small changes Δa and Δb must be made, given a new sample.

Gauss observed the movement of planets nightly and discovered the celebrated Recursive Least Squares Algorithm

Astronomy influenced Math & Science

Recursive Least Squares

- ❑ Recursively estimate parameters involved in the model. $\hat{\theta}$



$$\hat{\theta}(t) = \hat{\theta}(t-1) + \mathrm{K}_t \left(\underbrace{y(t) - \hat{y}(\varphi,\theta)}_{\text{Prediction Error}} \right)$$

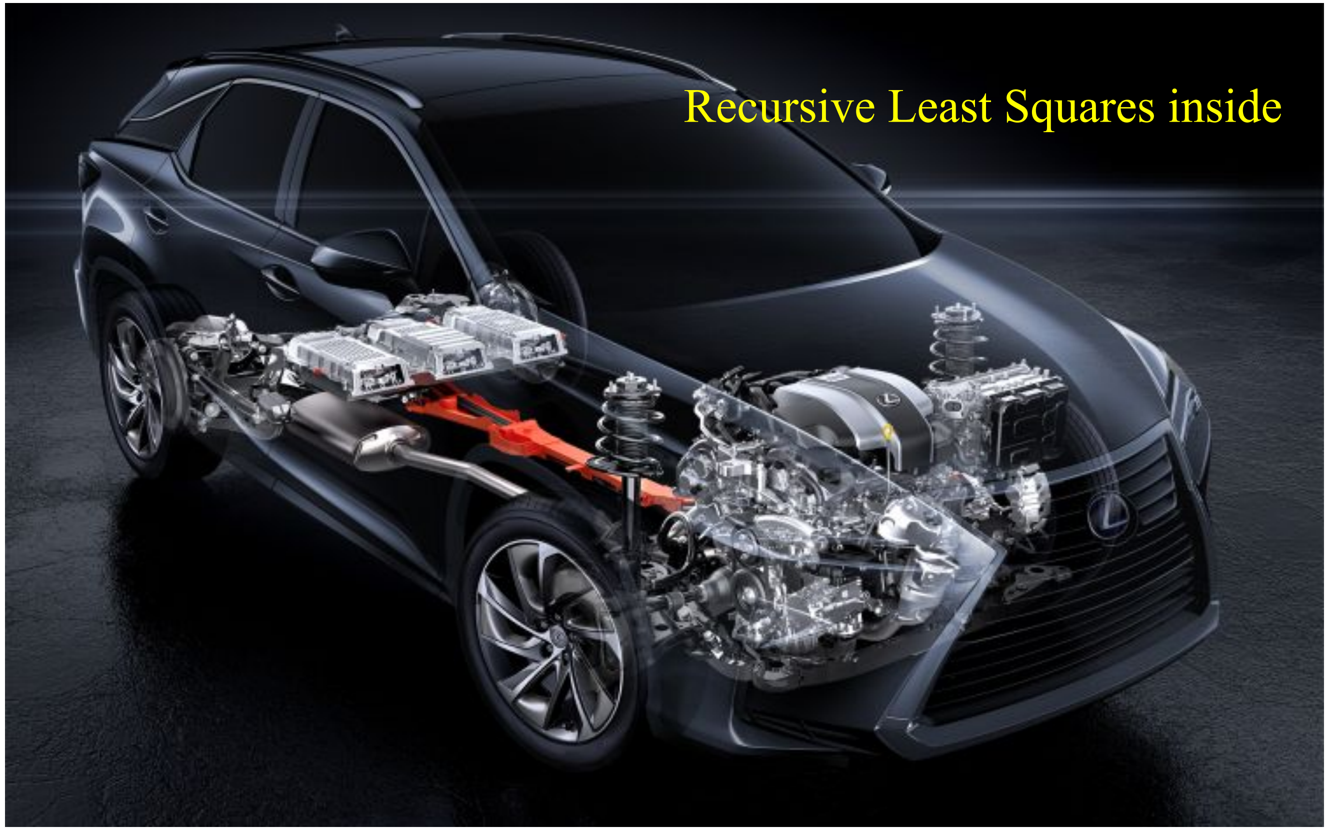
A type of gain for correcting the error

$$K_t = \frac{P_{t-1}\phi(t)}{1+\phi^T(t)P_{t-1}\phi(t)}$$

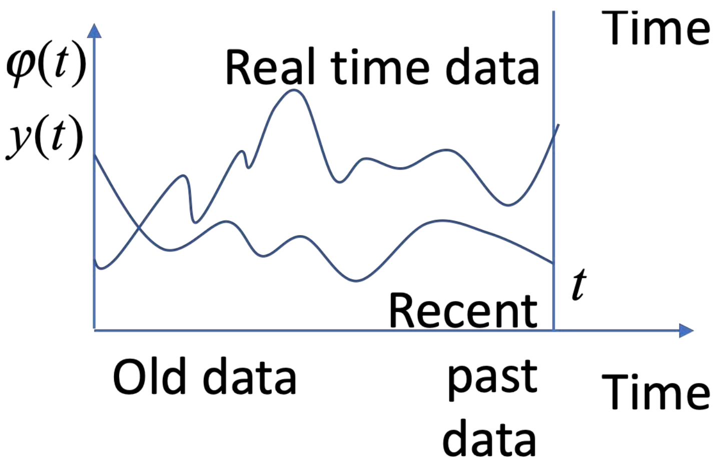
$$P_t = P_{t-1} - \frac{P_{t-1}\varphi(t)\varphi^T(t)P_{t-1}}{\left(1+\varphi^T(t)P_{t-1}\varphi(t)\right)}$$

This Recursive Least Squares Algorithm was originally developed by Gauss (1777 – 1855)

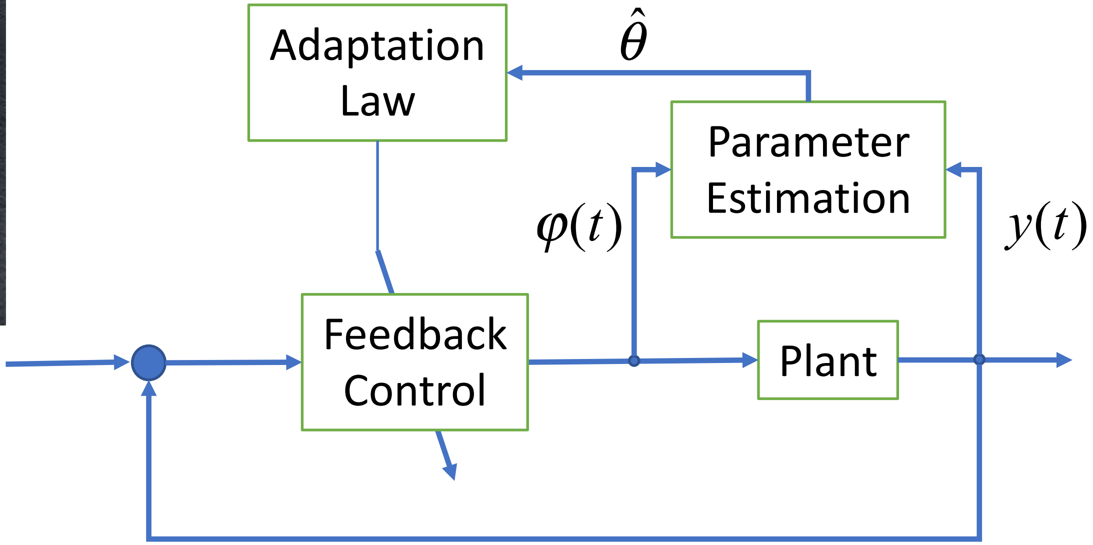
Smart traction + suspension control



Road conditions may vary.

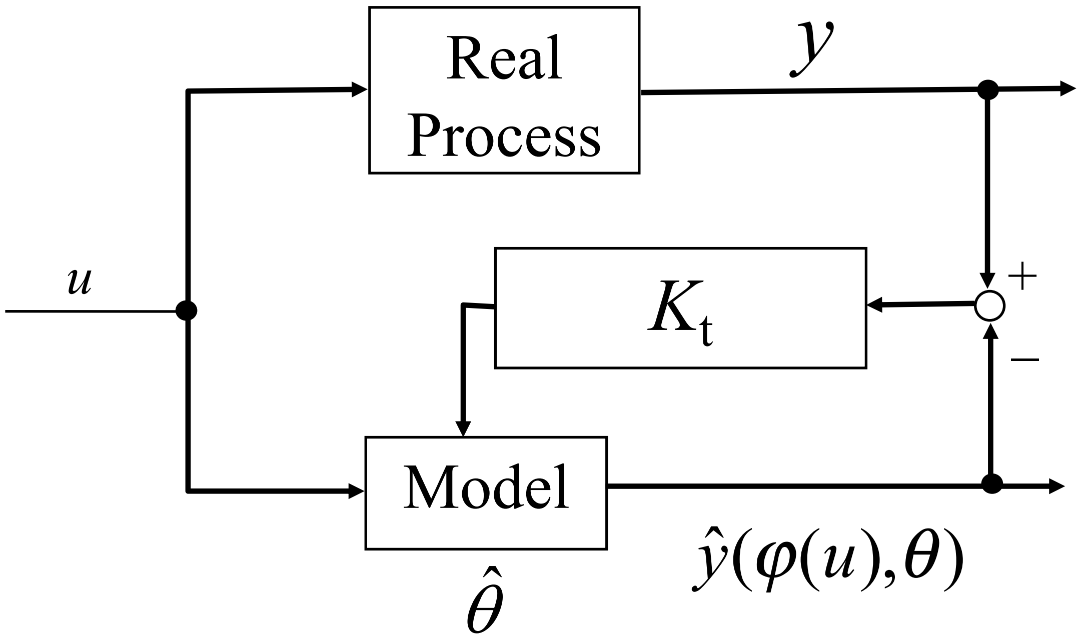


Indirect Adaptive Control



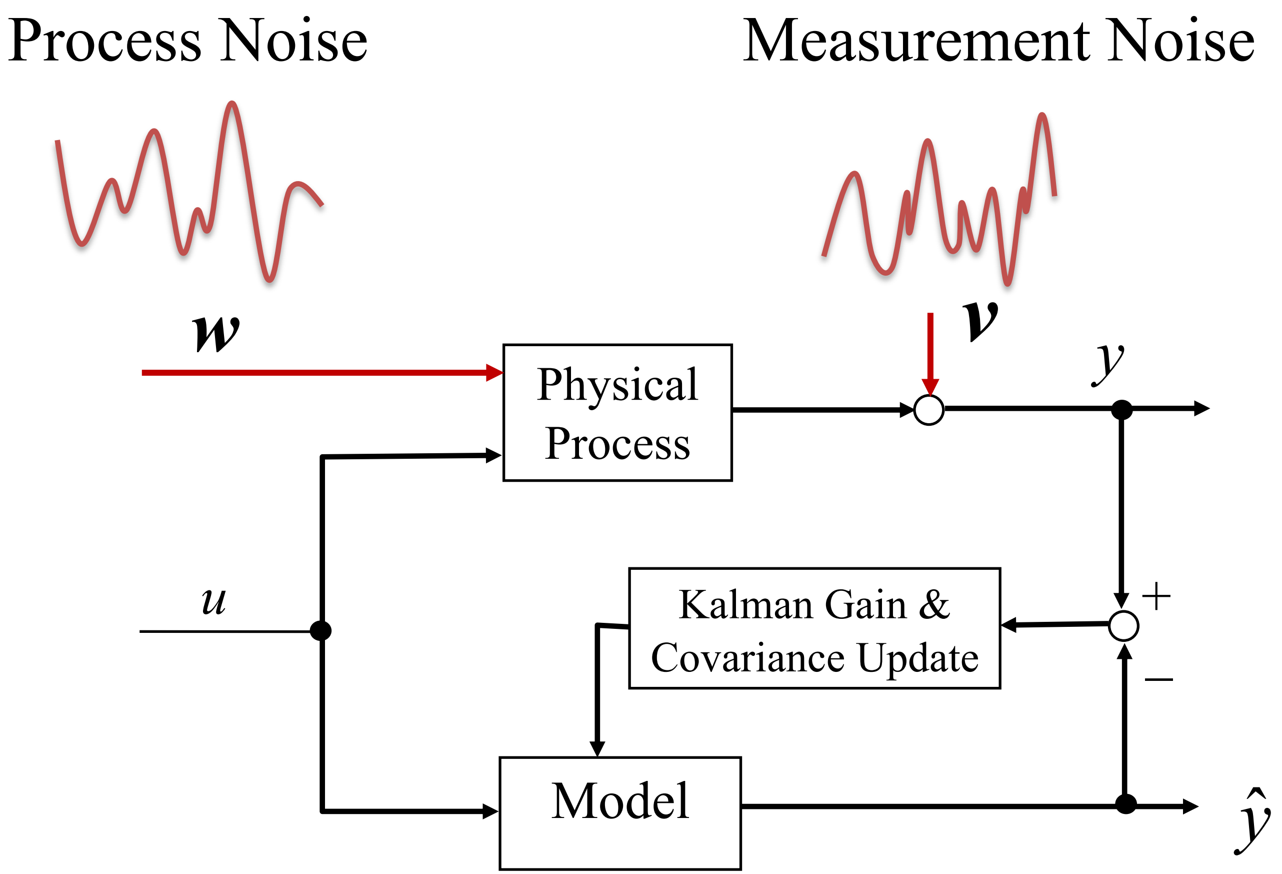
Recursive Least Squares

1821



Kalman Filter

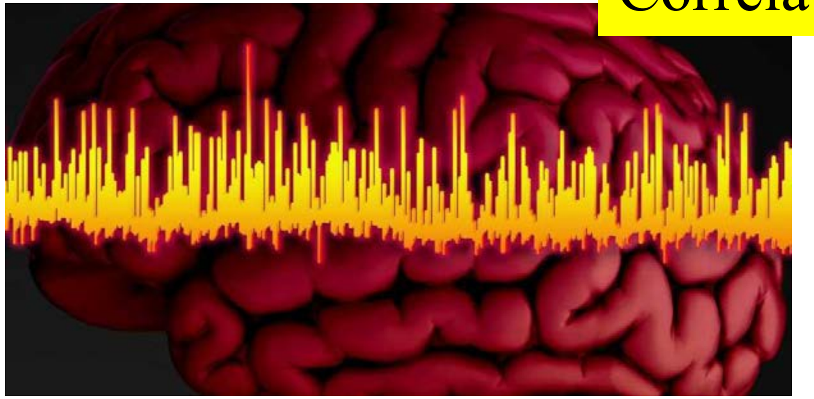
1960

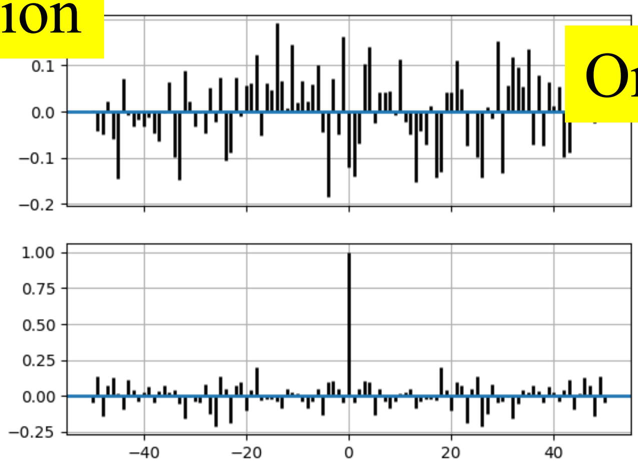


$$\hat{x}(t) = \hat{x}(t\,|\,t-1) + \mathrm{K}_t\left(y(t) - H(t)\hat{x}(t\,|\,t-1)\right)$$

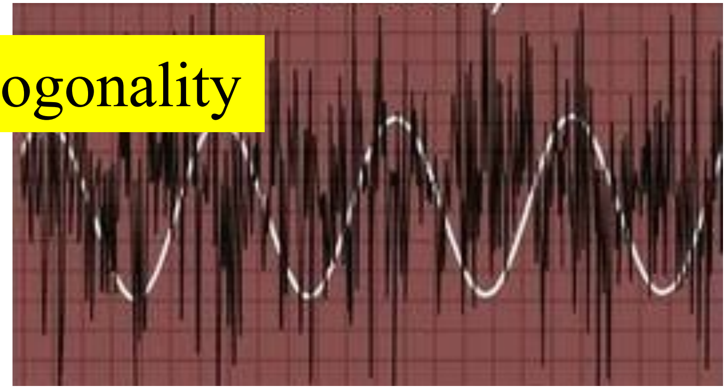
Random Processes

Correlation

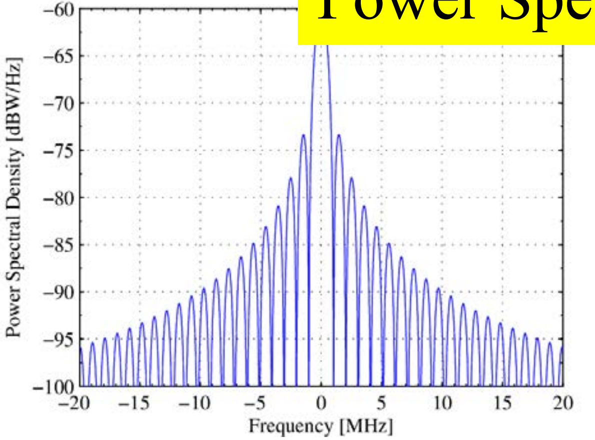




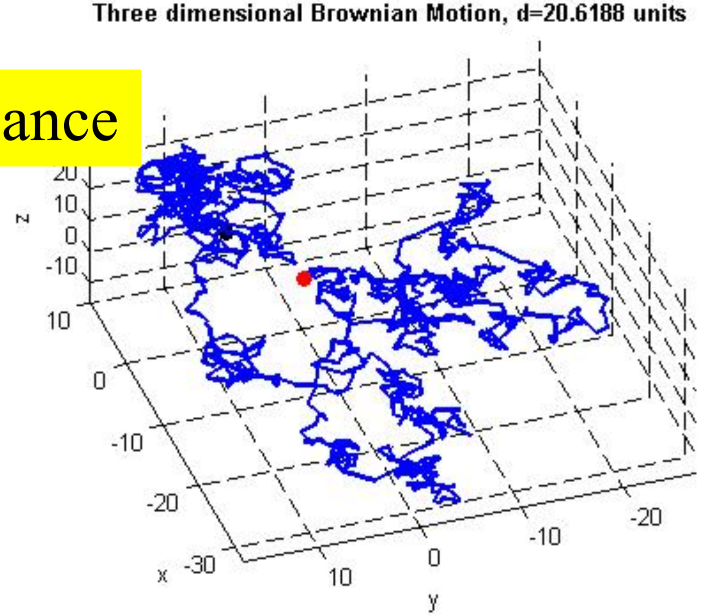
Orthogonality



Power Spectrum



Covariance



Random Variables and Random Processes

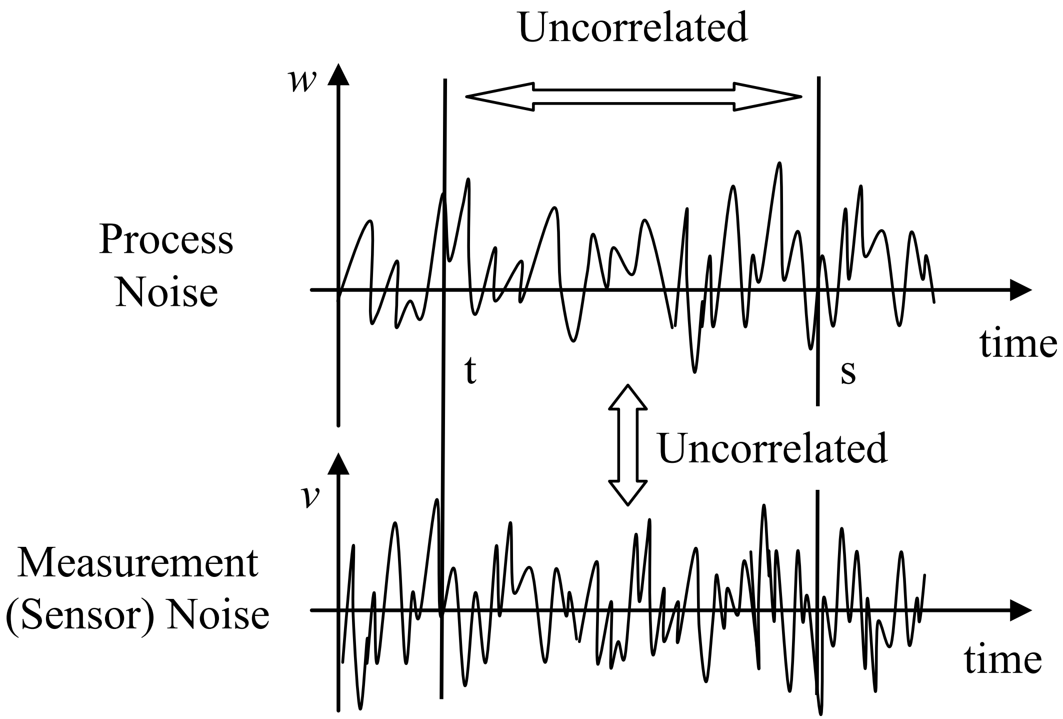


Figure 4-5 Noise characteristics

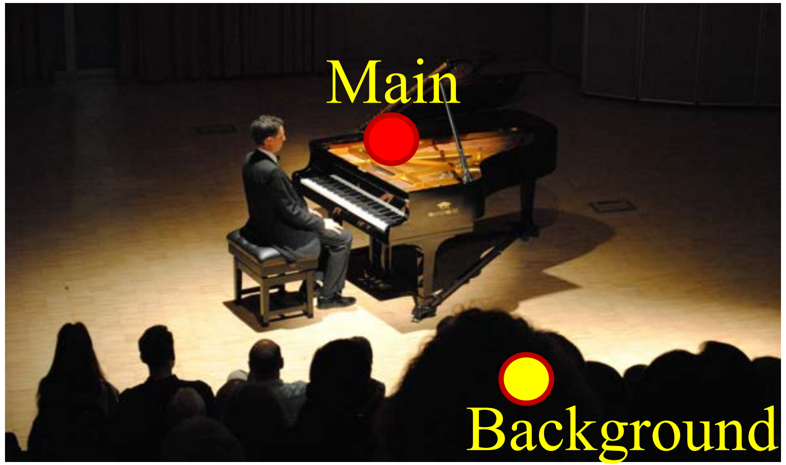
Correlation

$$E[XY] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy\, f_{XY}(x,y)\,dxdy$$

Covariance of X and $Y = E\,[(X\text{-}m_X)\,(Y\text{-}m_Y)]$

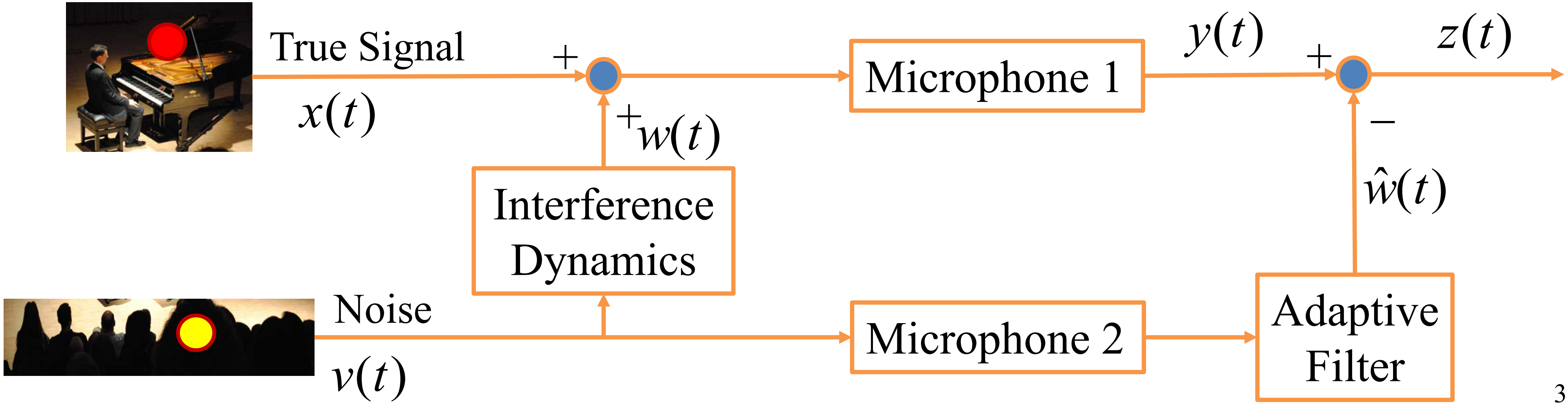
Math Policy:
Necessary math will be introduced/reviewed when needed / as needed.

Adaptive Noise Cancellation: An Application of Random Processes

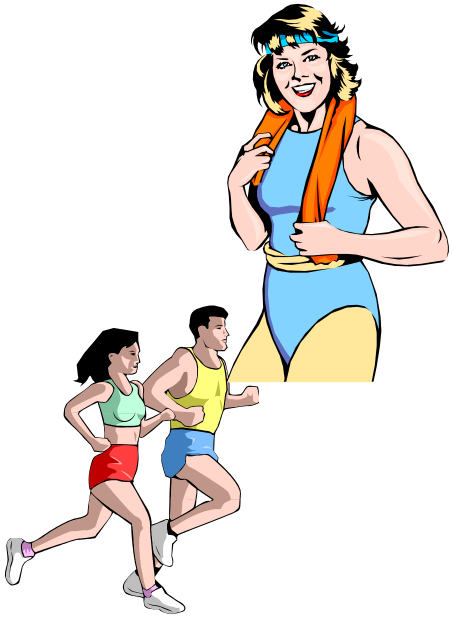


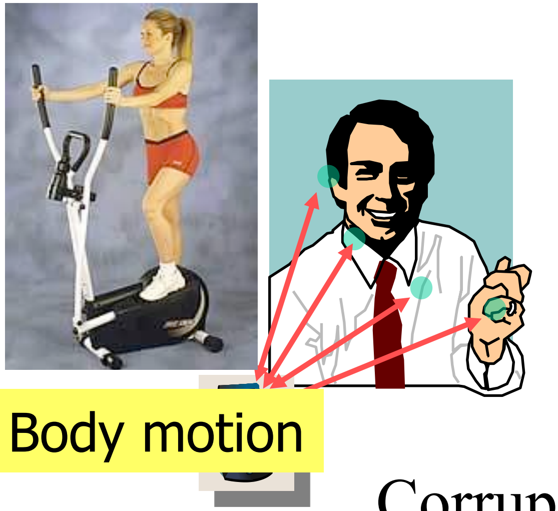
$$y(t) = x(t) + w(t)$$

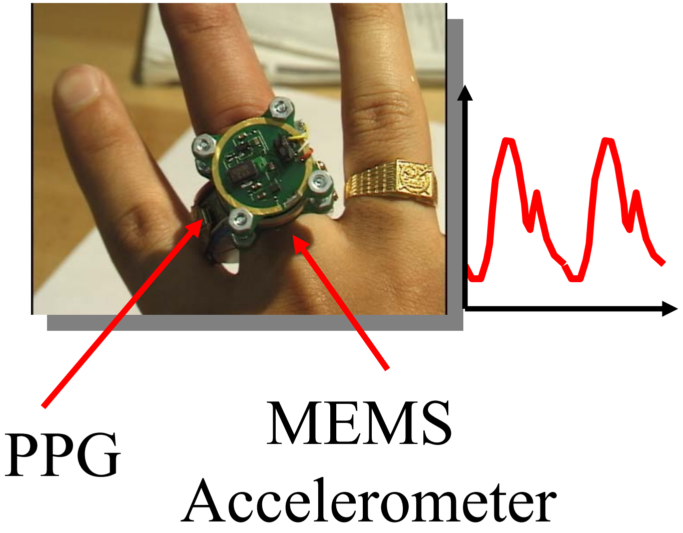
$$z(t) = y(t) - \hat{w}(t)$$



Context-Oriented Project #1: Active Noise Cancellation for Wearable Sensors







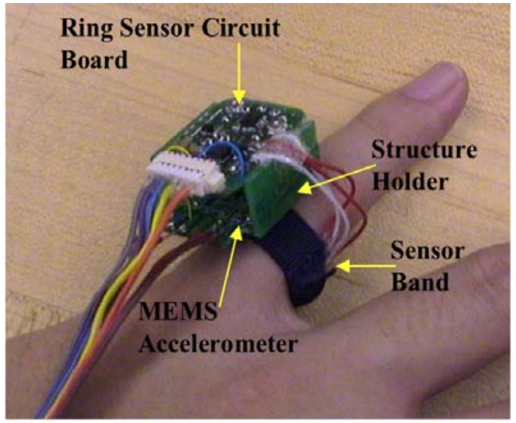
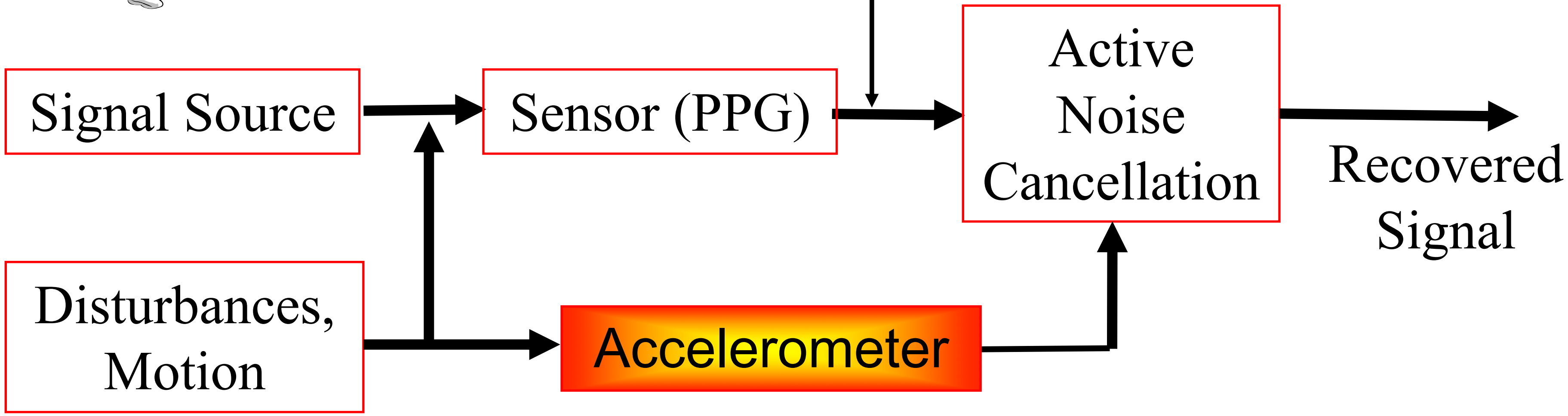


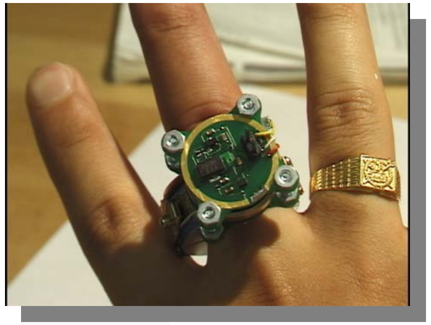
Figure 1 MEMS accelerometer collocated with PPG ring sensor

Asada, et al 2001



Apple iWatch

Active Noise Cancellation

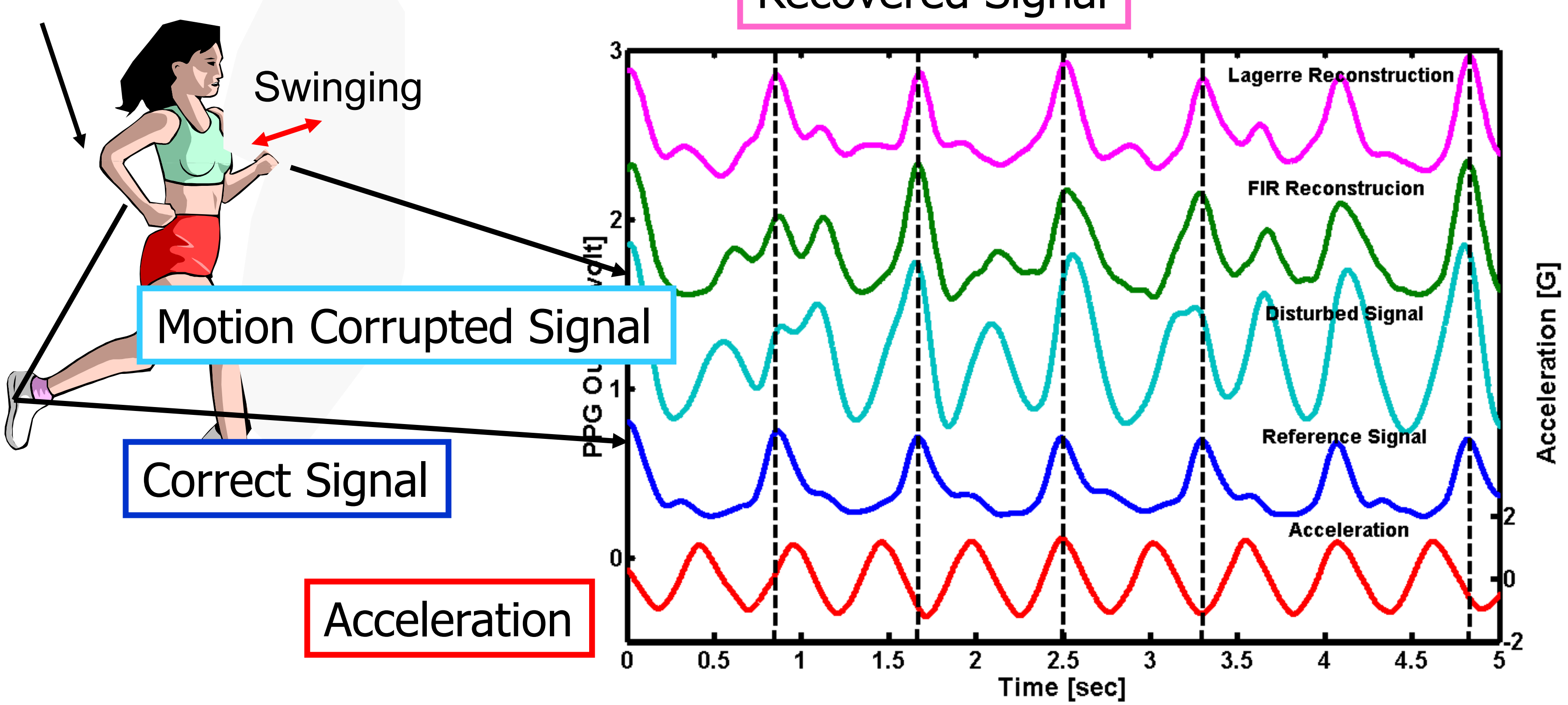


Orthogonality of two signals

$$E[XY] = 0$$

Stationary

Recovered Signal



Motion Corrupted Signal

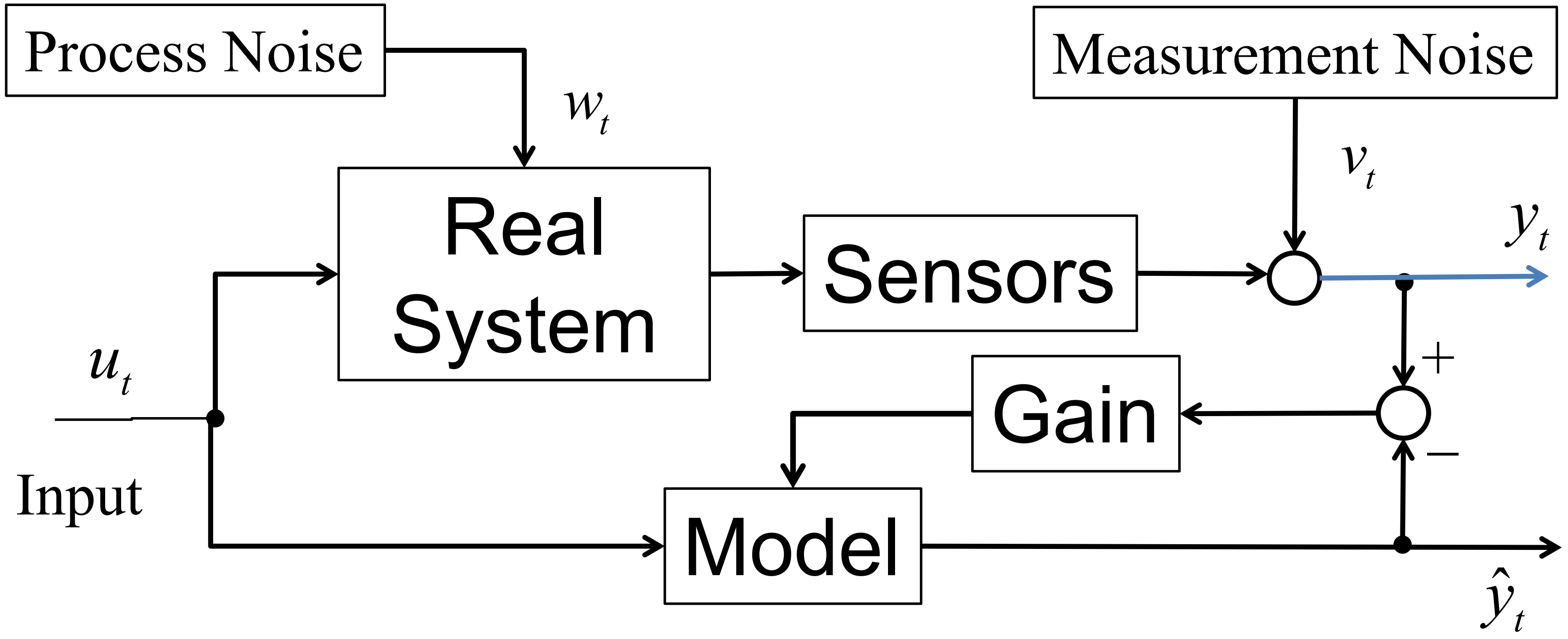
Correct Signal

Acceleration

Lecture 7

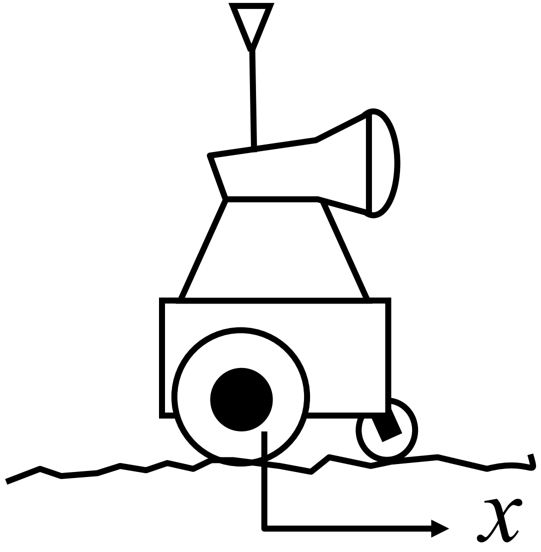
Part 2: Kalman and Bayes Filters

Kalman Filter Framework



$$[\text{Correction}] \propto \frac{[\text{Prediction Error}]}{[\text{Confidence of Previous Estimate}]\times[\text{Noise Variance}]}$$

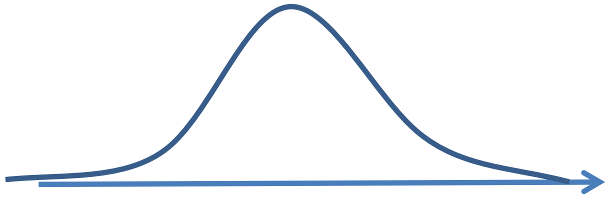
Knowledge about the Process



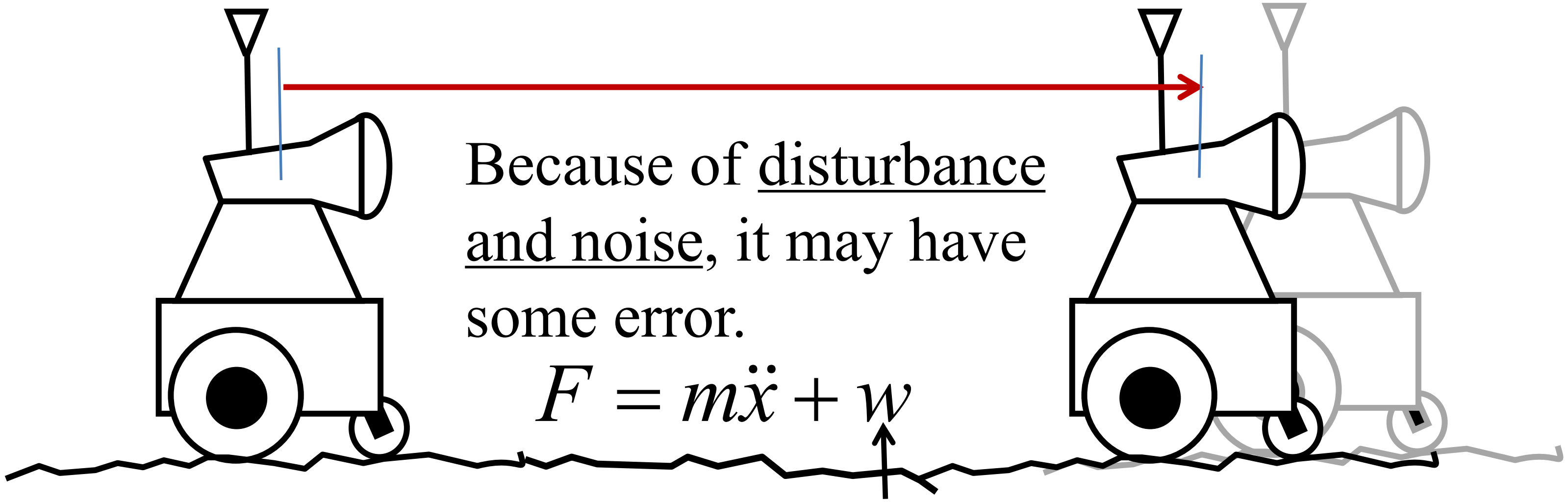
Vehicle Localization Problem:
Finding where it is now.

Model = What we know about the robot
and the environment.

Vehicle Dynamics: $F = m\ddot{x}$



We can predict the vehicle position by simulating the dynamic equation.

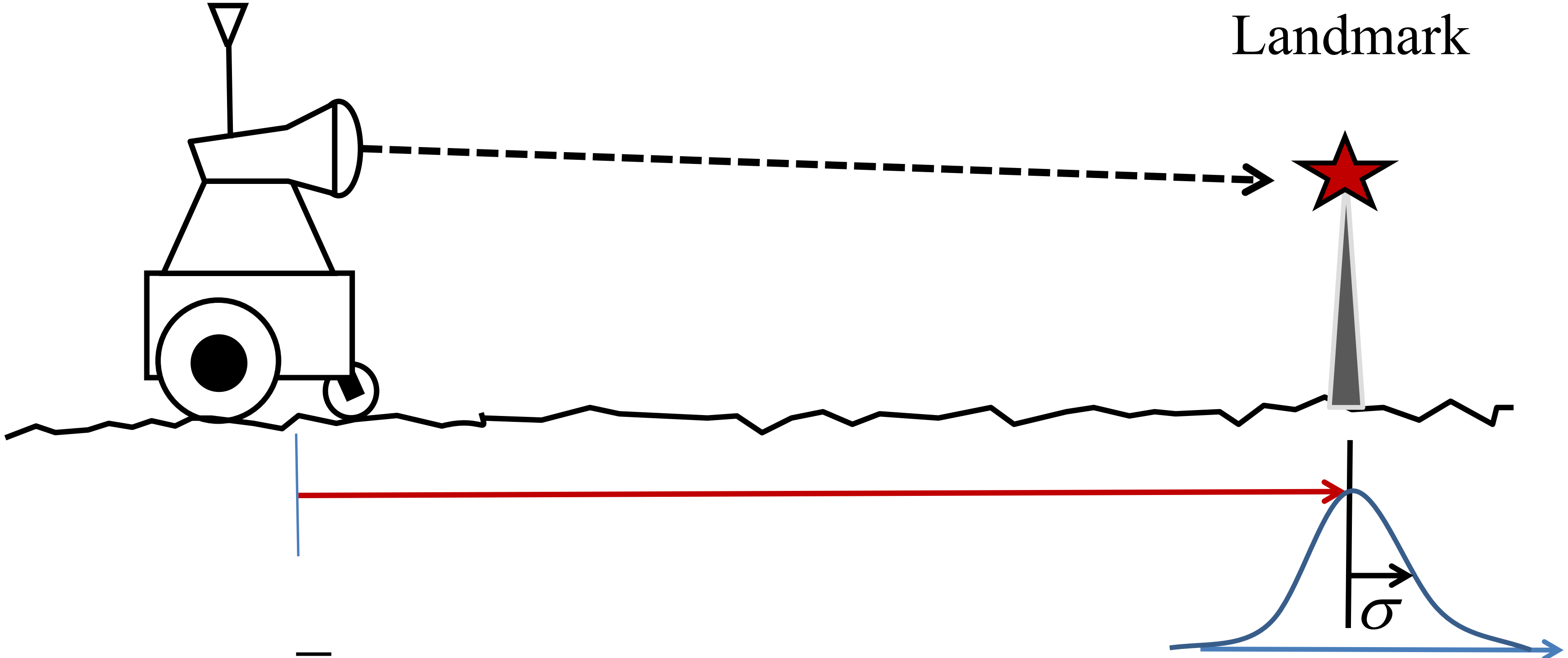


Because of disturbance and noise, it may have some error.

$$F = m\ddot{x} + \underset{\uparrow}{w}$$

Process Noise

Measurement Noise



$$y_t = \bar{y}_t + \underset{\uparrow}{v_t}$$

Measurement Noise

Mean = 0, Variance = σ^2

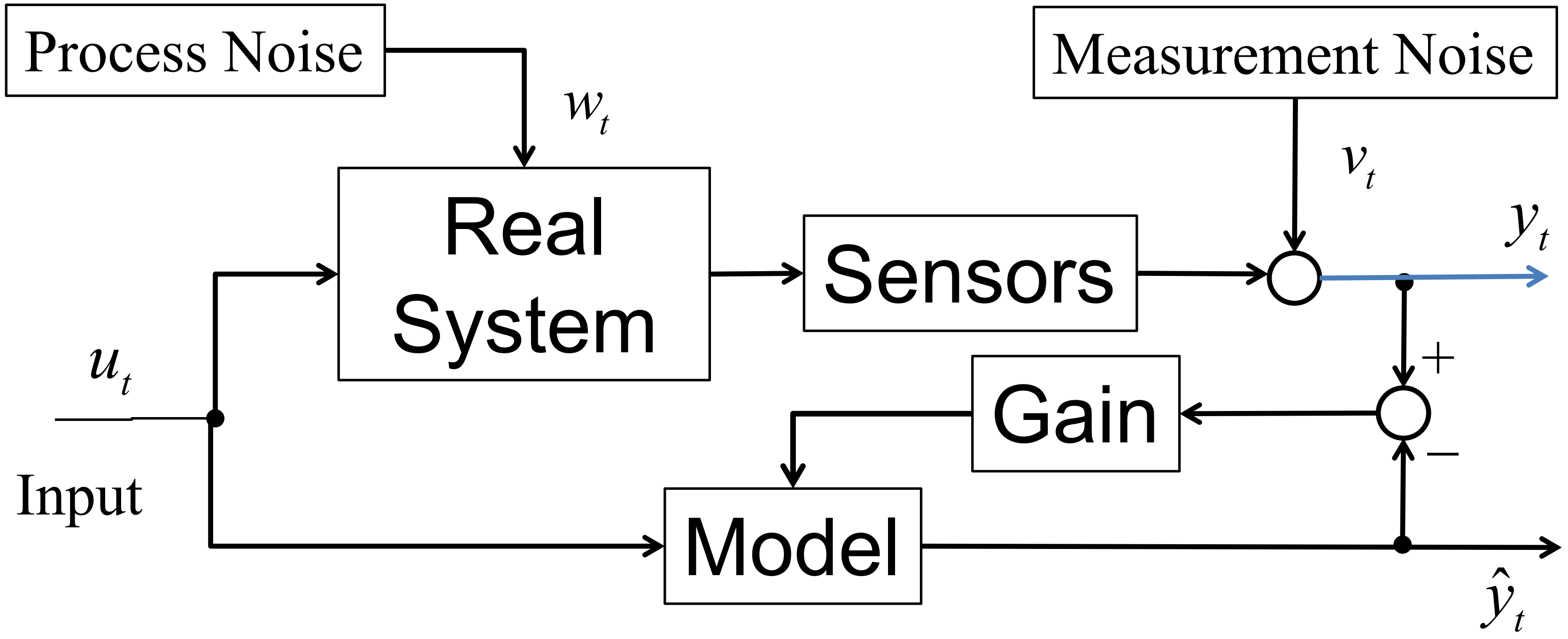
Experimentally determined

If two sensors are used,

$$R = \begin{pmatrix} \sigma_X{}^2 & 0 \\ 0 & \sigma_Y{}^2 \end{pmatrix}$$

Part 2: Kalman and Bayes Filters

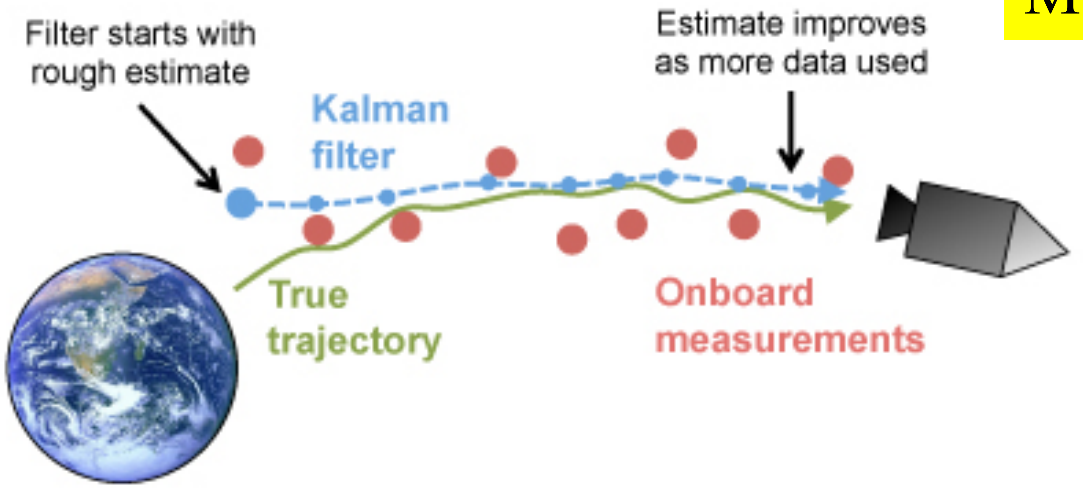
Kalman Filter Framework



$$[\text{Correction}] \propto \frac{[\text{Prediction Error}]}{[\text{Confidence of Previous Estimate}] \times [\text{Noise Variance}]}$$

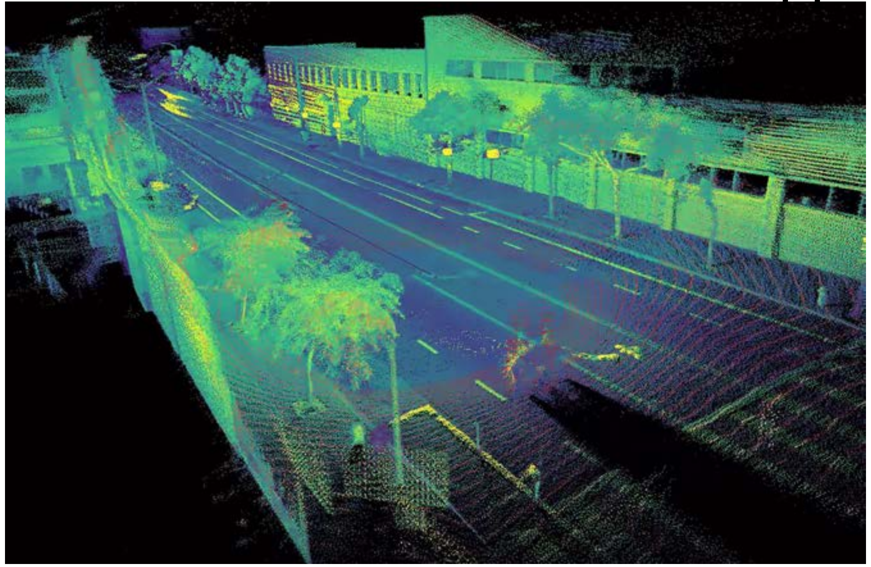
Kalman Filter applied to the Apollo Moon Mission

Moon Shot

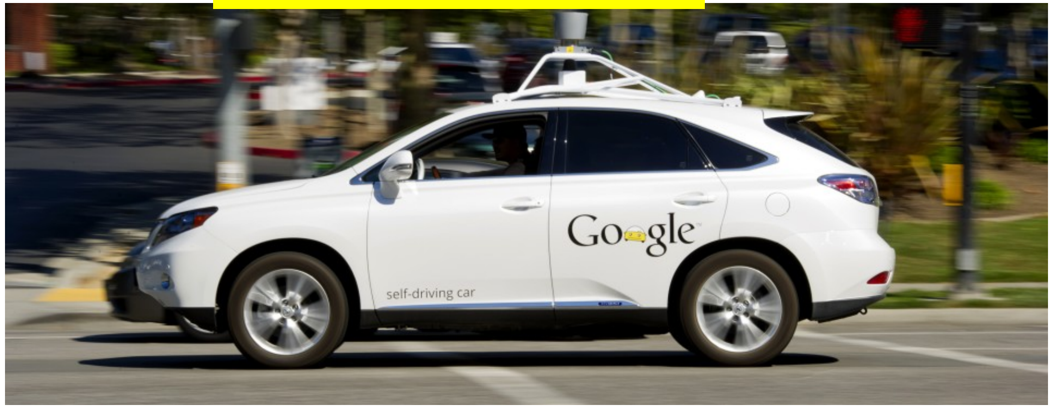

Filter starts with rough estimate
Kalman filter
Estimate improves as more data used
True trajectory
Onboard measurements

Rudolf E. Kalman

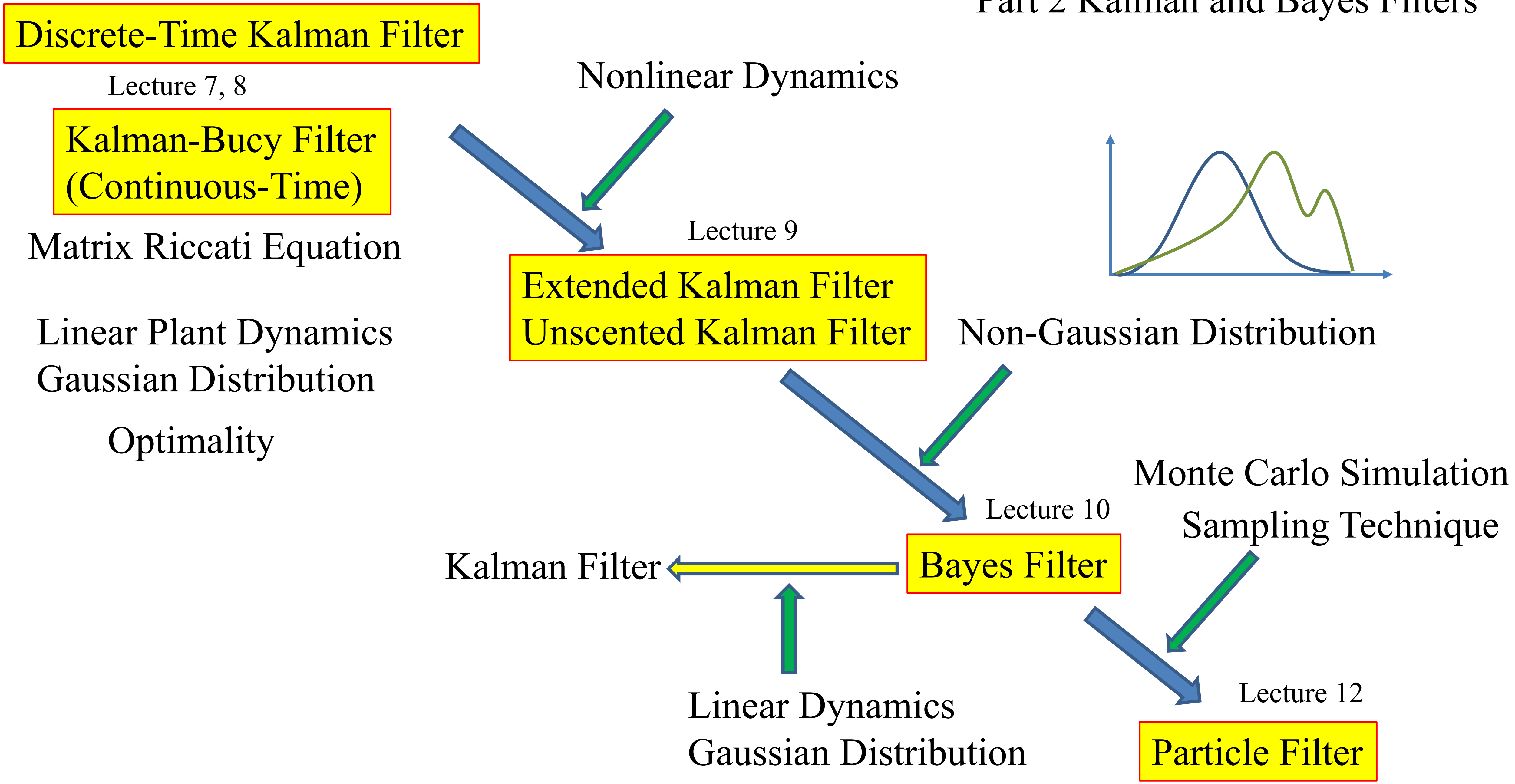
Simultaneous Localization And Mapping (SLAM)



Self-Driving Car



Part 2 Kalman and Bayes Filters



Discrete-Time Kalman Filter

Lecture 7, 8

Kalman-Bucy Filter (Continuous-Time)

Matrix Riccati Equation

Linear Plant Dynamics
Gaussian Distribution

Optimality

Nonlinear Dynamics

Lecture 9

Extended Kalman Filter
Unscented Kalman Filter

Non-Gaussian Distribution

Lecture 10

Bayes Filter

Kalman Filter

Linear Dynamics
Gaussian Distribution

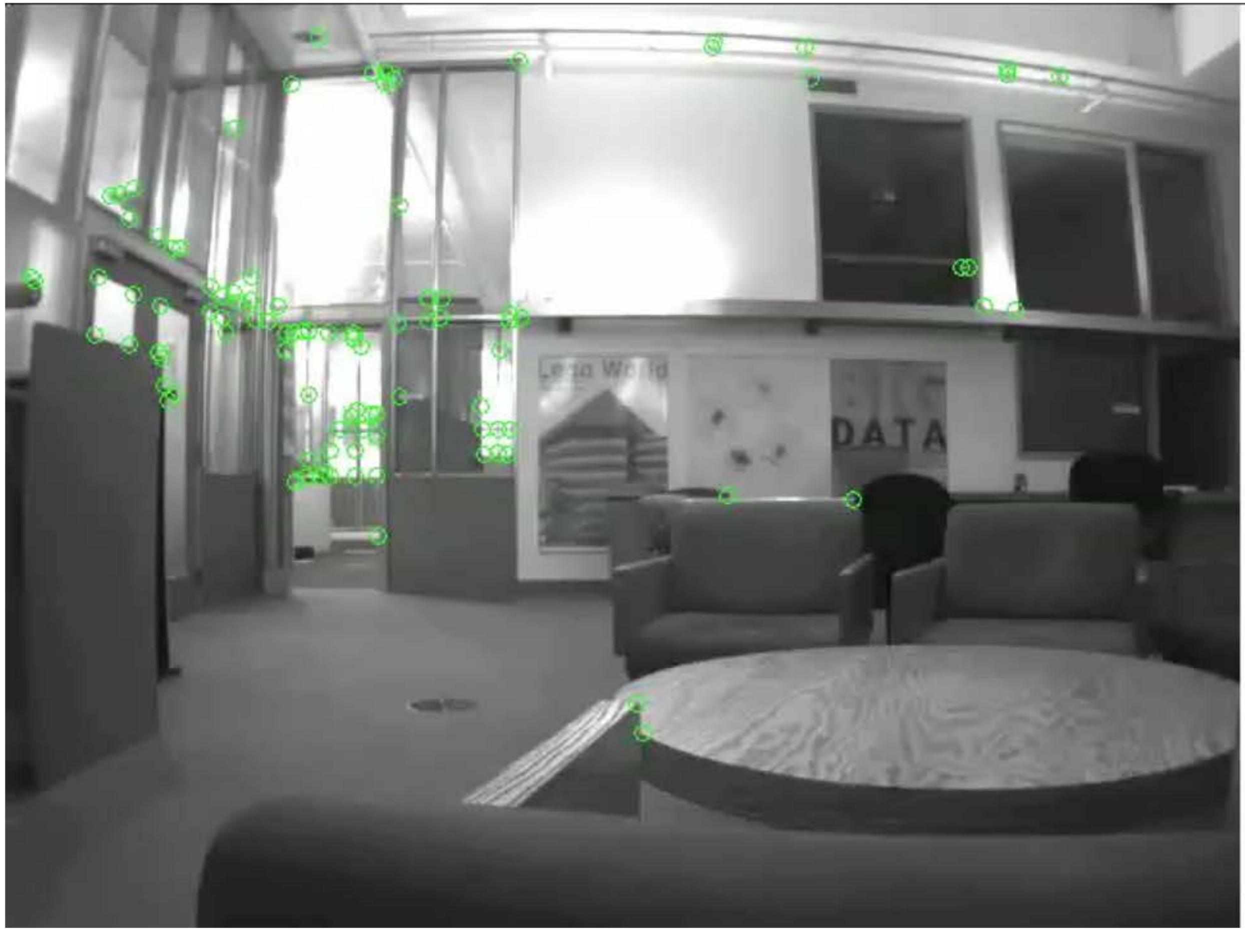
Monte Carlo Simulation
Sampling Technique

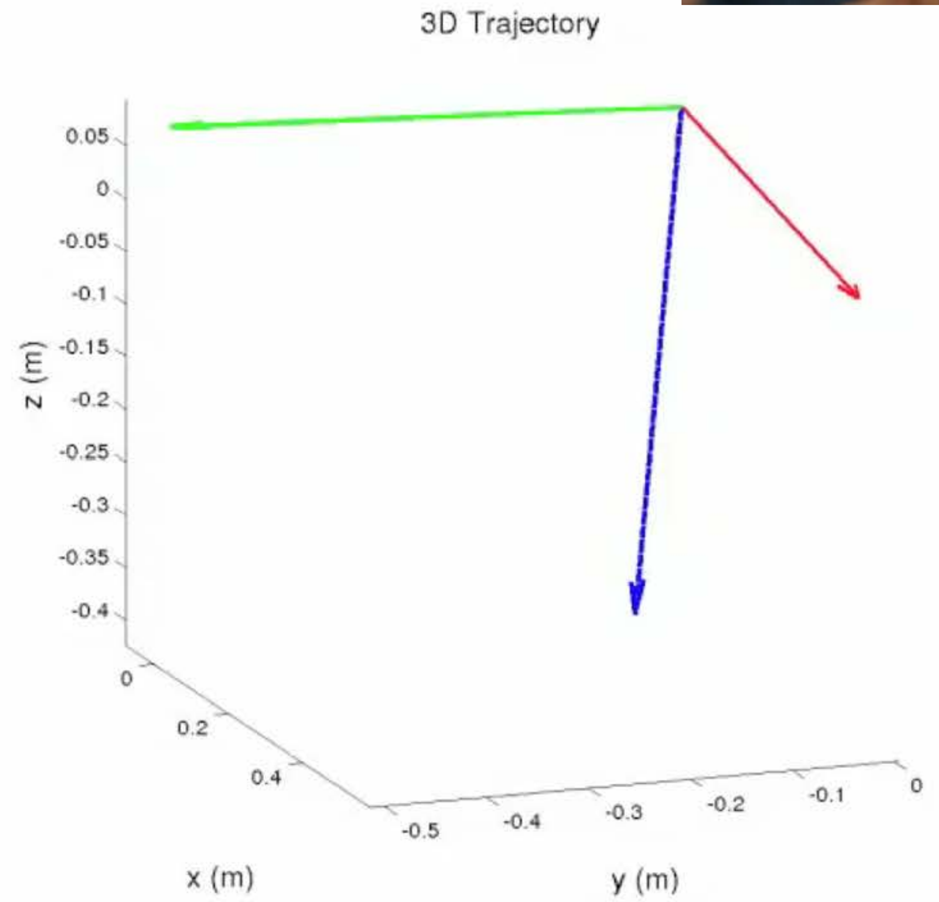
Lecture 12

Particle Filter

Lecture 11

Visual SLAM
Simultaneous Localization and Mapping





Professor John Leonard

Context-Oriented Project No.2

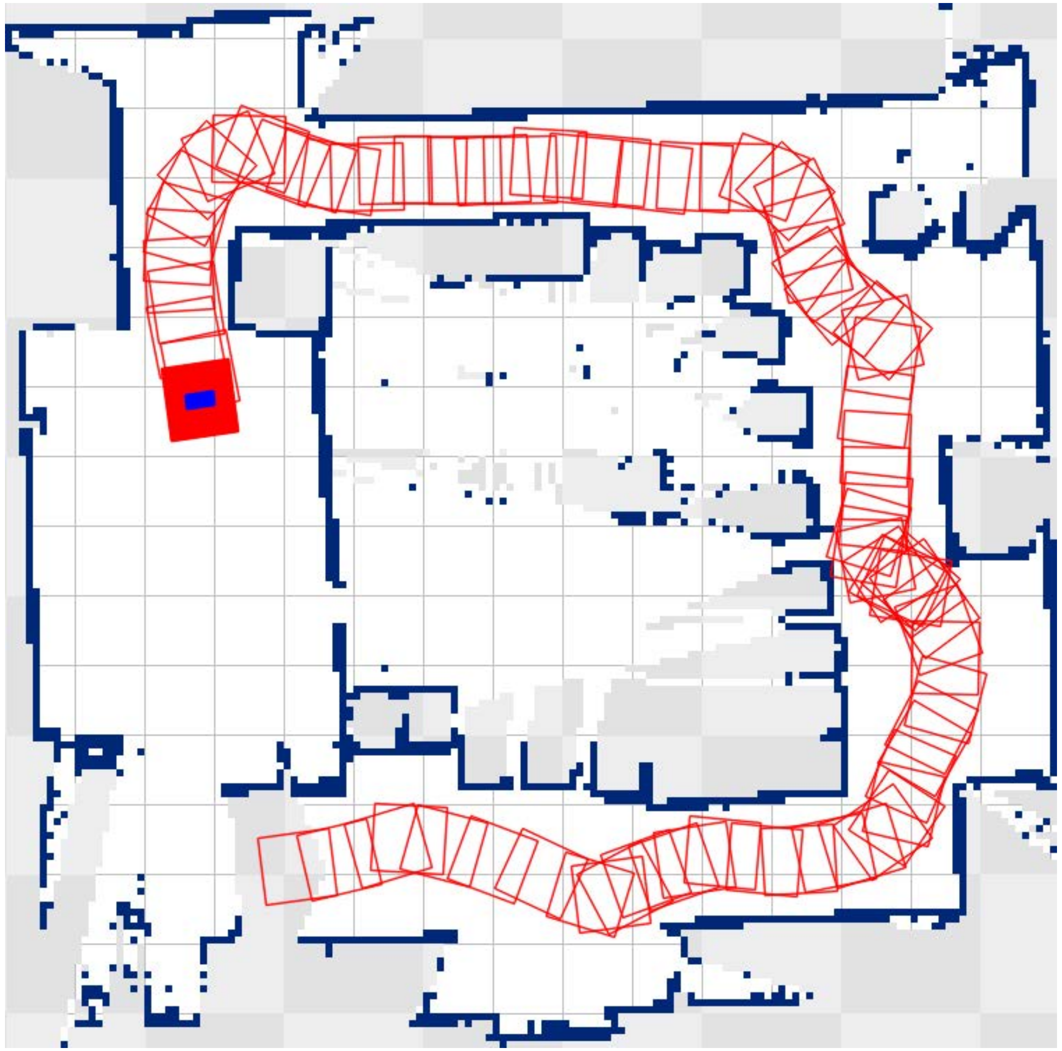
Robot Localization and Navigation Using advanced Kalman Filters:

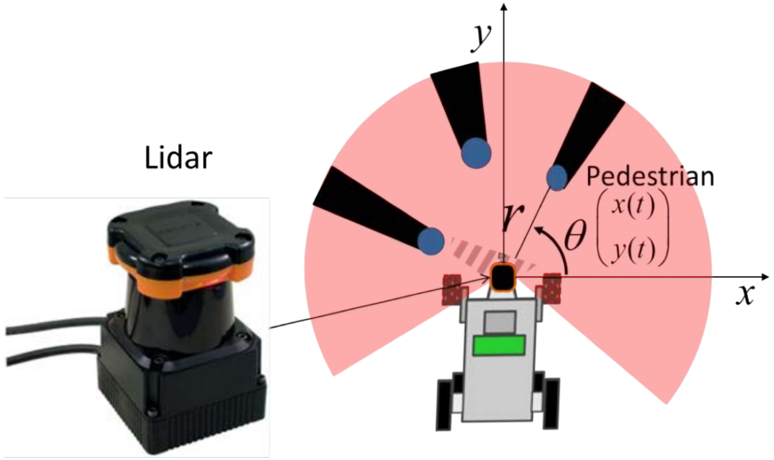
Extended KF

Unscented KF

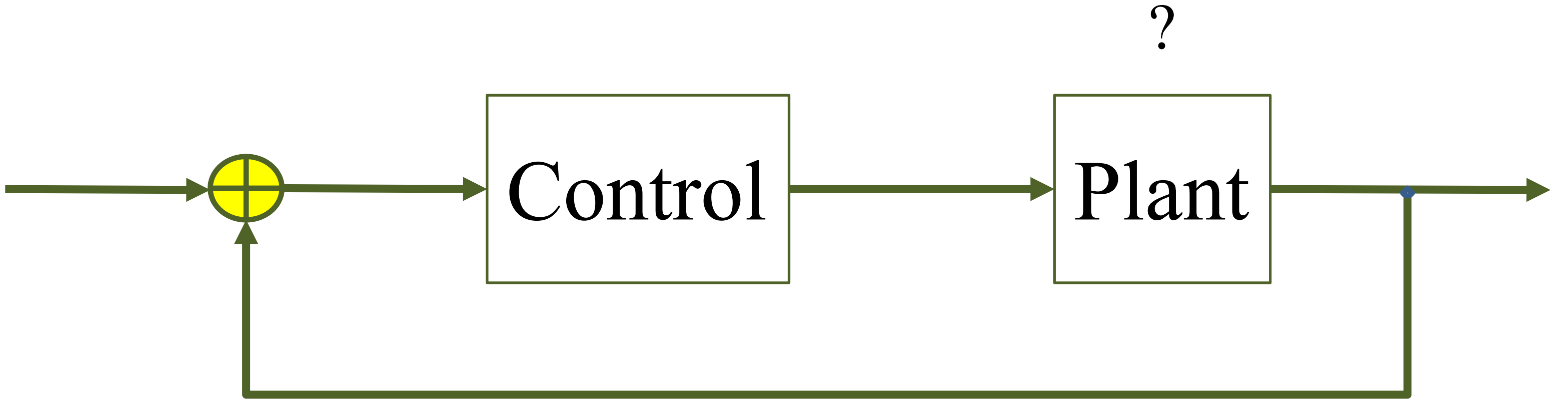
Bayes filter

Particle filter

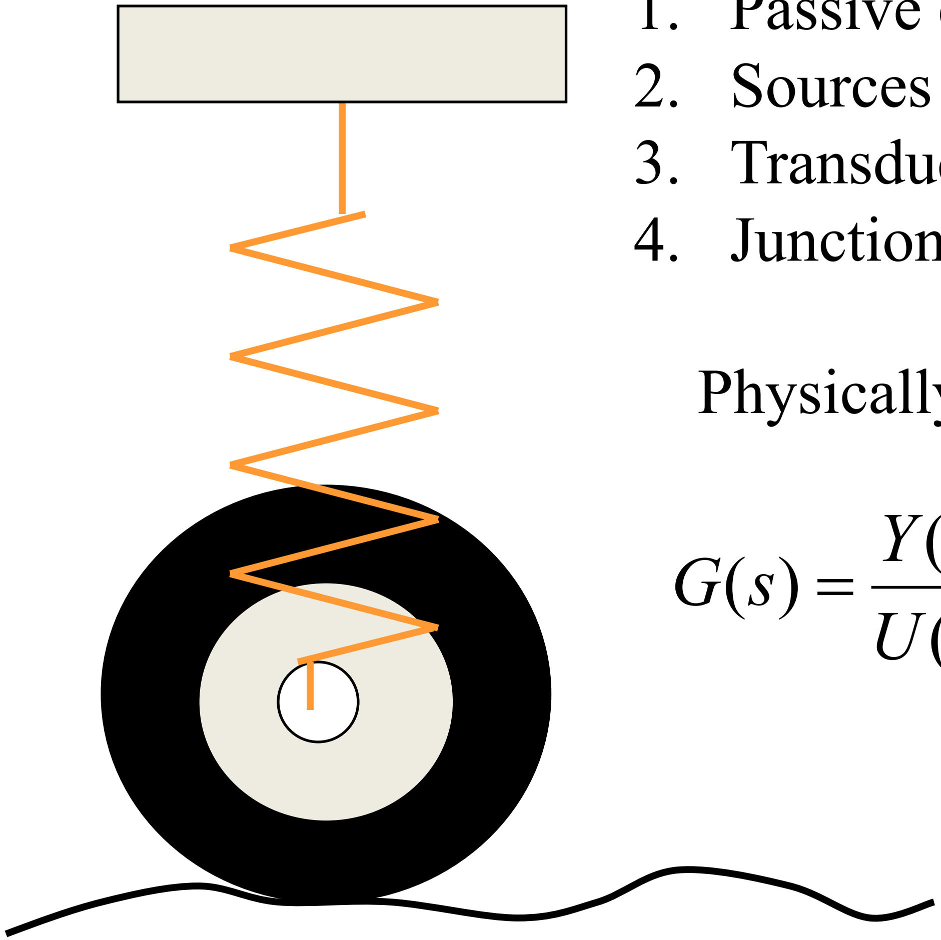




Part 3:
System Identification of Linear Dynamical Systems



Physical Modeling : 2.151/2.140, 2.032, 2.004, 2.003, etc.



1. Passive elements: mass, damper, spring
2. Sources
3. Transducers
4. Junction structure

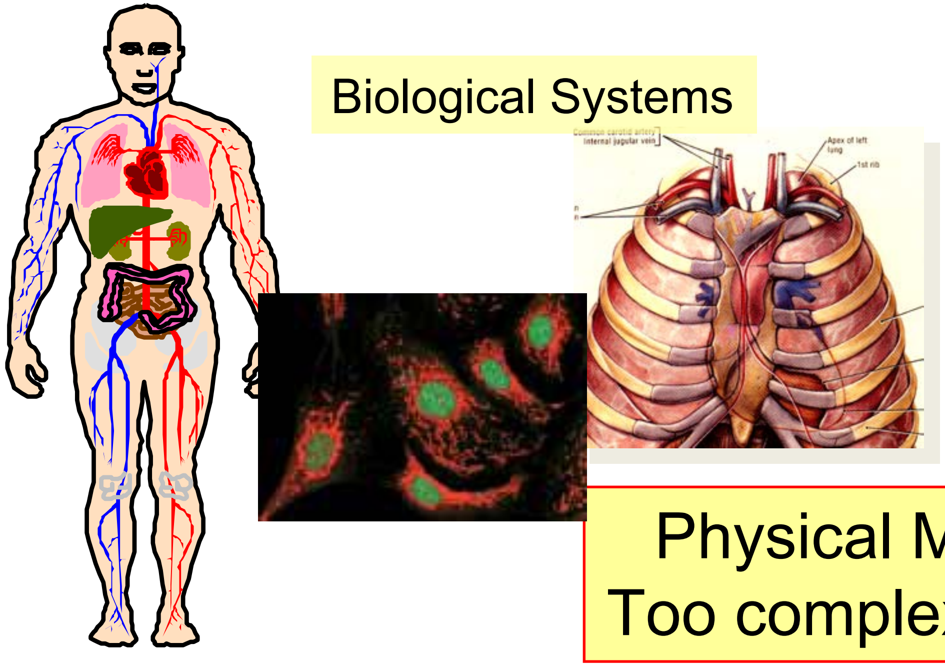
Physically meaningful parameters

$$G(s) = \frac{Y(s)}{U(s)} = \frac{b_0 s^m + b_1 s^{m-1} + \cdots + b_m}{s^n + a_1 s^{n-1} + \cdots + a_n}$$

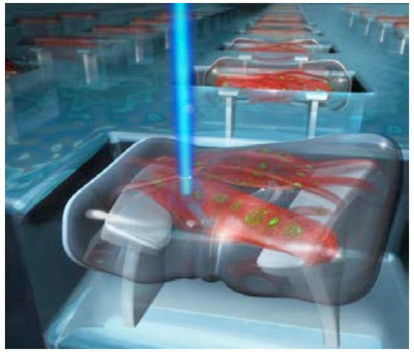
$$a_i = a_i(M, B, K)$$

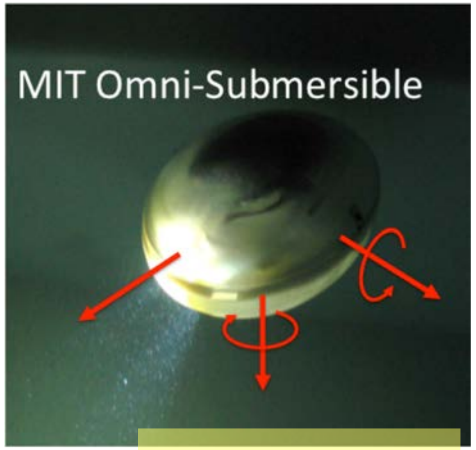
$$b_i = b_i(M, B, K)$$

Mathematical models of real-world systems are often too difficult to build based on first principles ***alone***.



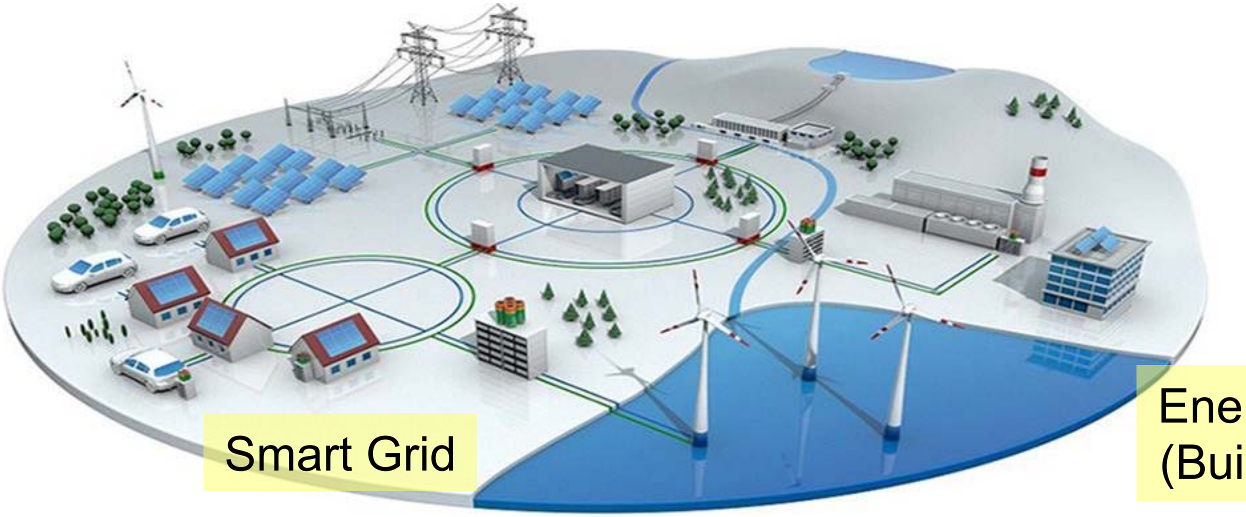
Biological Systems



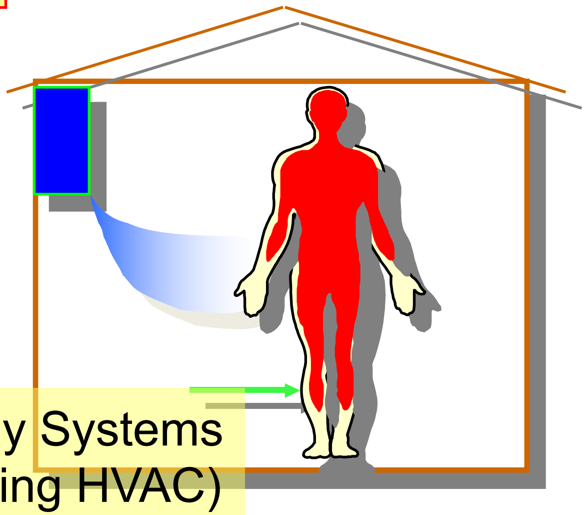


Fluid-thermal Systems

Physical Modeling: Too complex to model

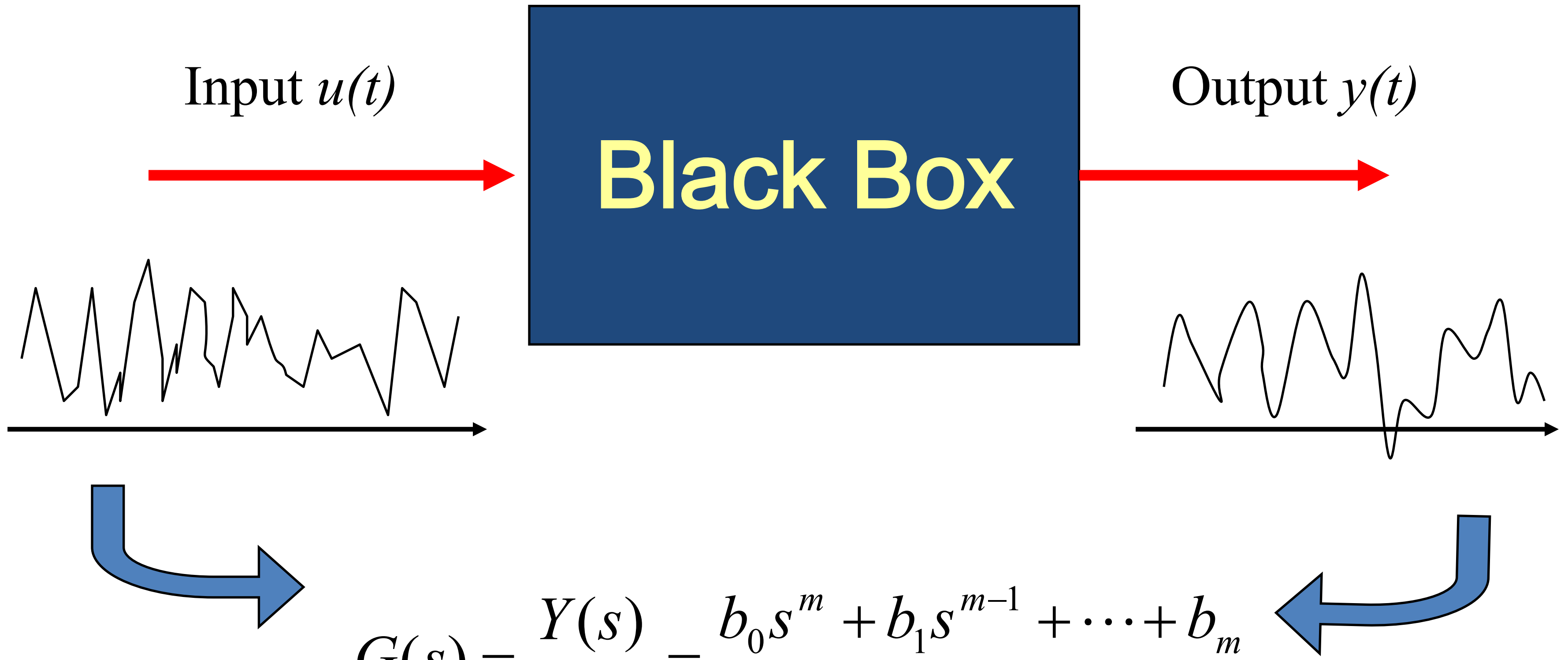


Smart Grid



Energy Systems (Building HVAC)

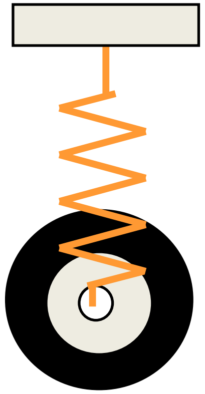
System Identification;
"Let the data speak about the system".



$$G(s) = \frac{Y(s)}{U(s)} = \frac{b_0 s^m + b_1 s^{m-1} + \cdots + b_m}{s^n + a_1 s^{n-1} + \cdots + a_n}$$

Find a model structure and determine parameter values that fit the data.

Physical modeling



Pros

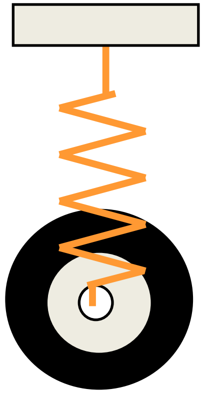
1. Physical insight and knowledge
2. Modeling a conceived system before hardware is built

Cons

1. Often leads to high system order with too many parameters
2. Input-output model has a complex parameter structure
3. Not convenient for parameter tuning
4. Complex system; too difficult to analyze

Comparison

Physical modeling



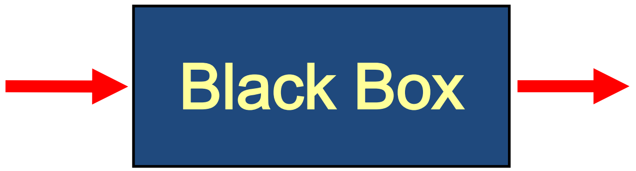
Pros

1. Physical insight and knowledge
2. Modeling a conceived system before hardware is built

Cons

1. Often leads to high system order with too many parameters
2. Input-output model has a complex parameter structure
3. Not convenient for parameter tuning
4. Complex system; too difficult to analyze

Data-Driven Approach



Pros

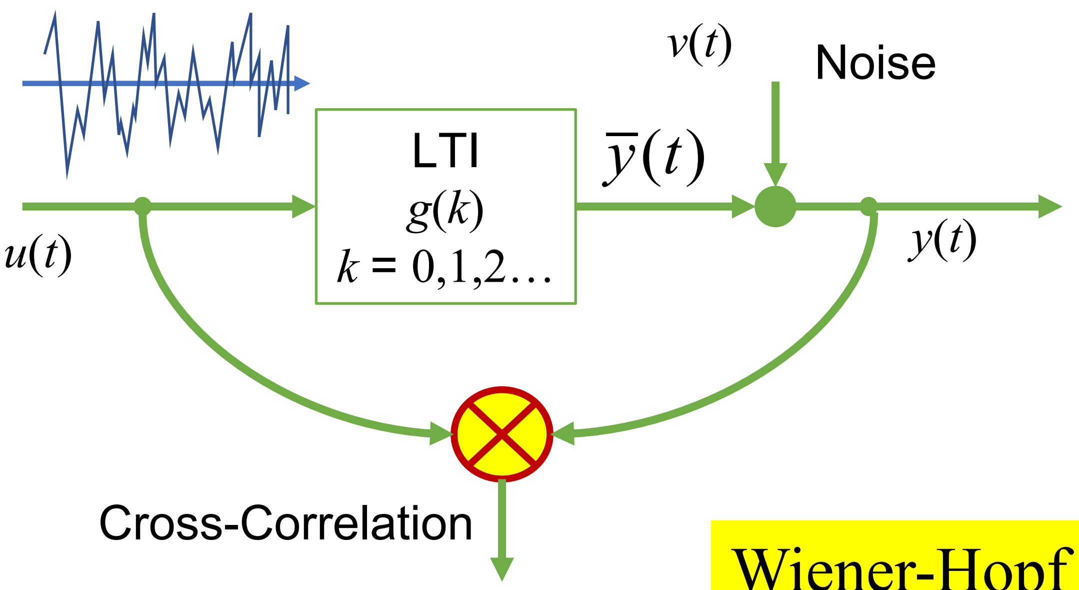
1. Close to the actual input-output behavior
2. Useful for complex systems; too difficult to build physical model
3. Quantify noise and uncertainty
4. Convenient structure for parameter tuning

Cons

1. No direct connection to physical parameters
2. No solid ground to support a model structure
3. Not available until an actual system has been built

Obtaining a Transfer Function from Input-Output Data

Time Domain



Wiener-Hopf Equation

$$R_{uy}(\tau) = E[u(t)y(t+\tau)]$$

$$= E[u(t)\overline{y}(t+\tau)] + E[u(t)v(t+\tau)]$$

$$= E[u(t)\overline{y}(t+\tau)]$$

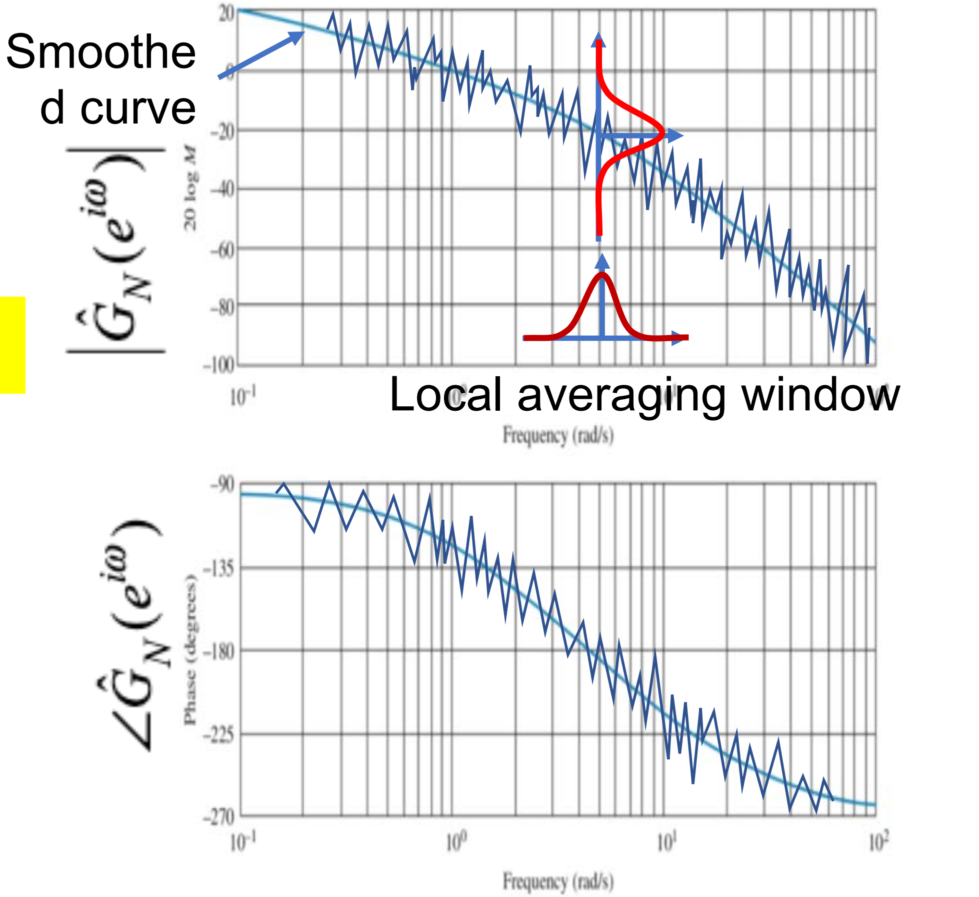
Impulse Response to White noise

$$\hat{g}(t) = \frac{1}{\lambda} R_{uy}(t), \quad t = 0,1,\cdots,N$$

Frequency Domain

$$G(e^{i\omega}) = \frac{\text{(Cross Spectrum)}}{\text{(Power Spectrum)}}$$

Bode Plot



Parametric System Identification

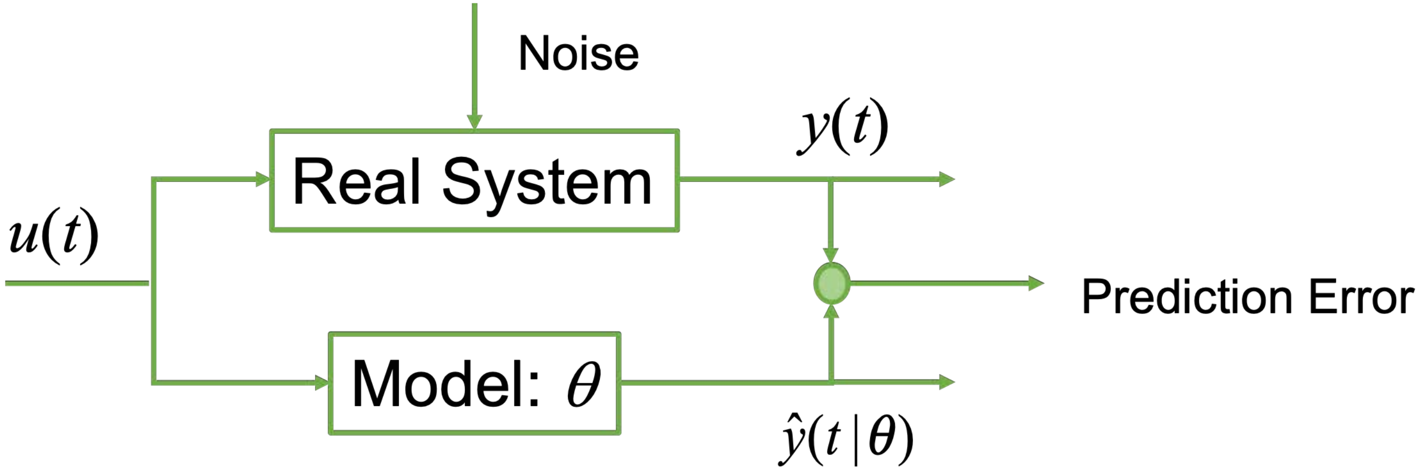
Parametric Model

$$G(q)=\frac{b_1q^{-1}+b_2q^{-2}+\cdots+b_{n_b}q^{-n_b}}{1+a_1q^{-1}+a_2q^{-2}+\cdots+a_{n_a}q^{-n_a}}$$

Parameters to identify

$$\theta=(a_1,\cdots,a_{n_a},b_1,\cdots,b_{n_b})^T$$

Prediction-Error Method



Noise

Real System

$u(t)$

$y(t)$

Prediction Error

Model: θ

$\hat{y}(t\,|\,\theta)$

$$\hat{\theta}_N=\arg\min_{\theta}\frac{1}{N}\sum(y(t)-\hat{y}(t\,|\,\theta))^2$$

How many data are required? What would be effective input signals?

Bode Plot

Non-Parametric

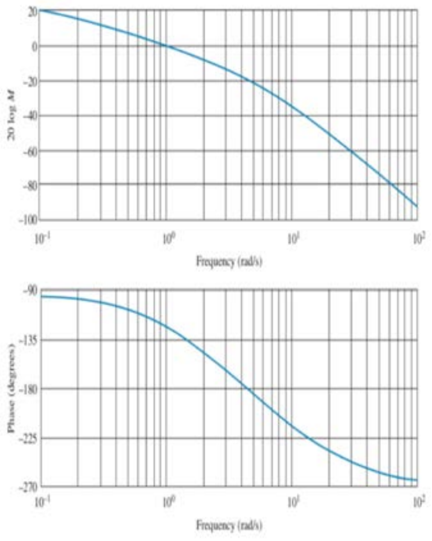
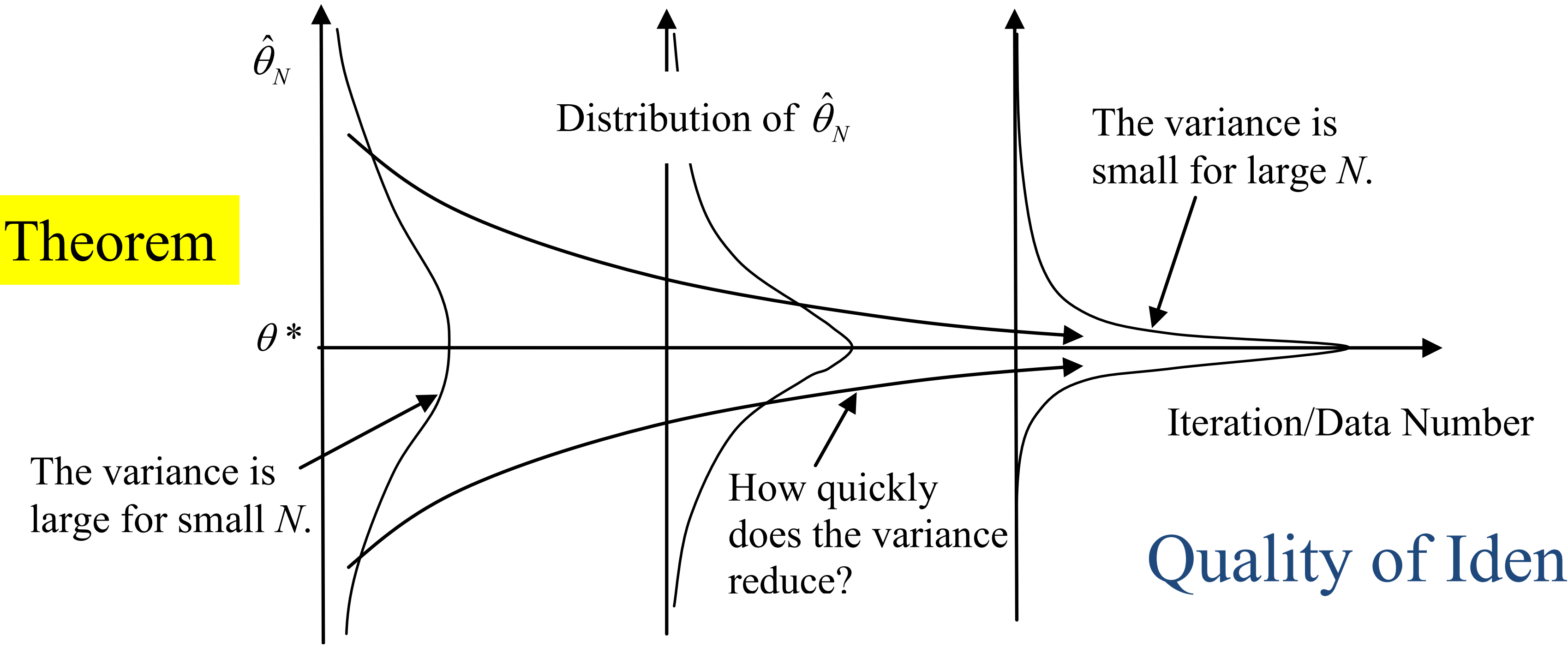


FIGURE P10.7

Lecture 15

Asymptotic Distribution of Parameter Estimates

Central Limit Theorem



$\hat{\theta}_N$

θ^*

Distribution of $\hat{\theta}_N$

The variance is small for large N.

The variance is large for small N.

How quickly does the variance reduce?

Iteration/Data Number

Quality of Identification

The main points to be obtained in this chapter

The variance analysis of this chapter will reveal

a) The estimate converges to θ^* at a rate proportional to $\frac{1}{\sqrt{N}}$

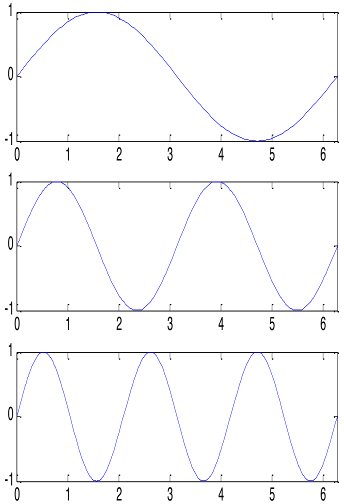
b) Distribution converges to a Gaussian distribution: $\boldsymbol{N}(0, Q)$.

c) Cov $\hat{\theta}_N$ depends on the parameters sensitivity of the predictor: $\frac{\partial \hat{y}}{\partial \theta}$

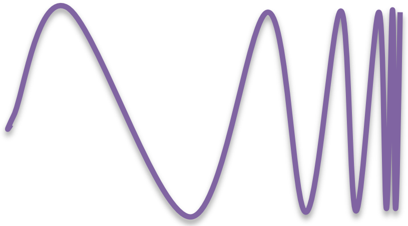
Identified model parameter $\hat{\theta}_N$ with cov $\hat{\theta}_N$: a “quality tag” confidence interval

Input Design

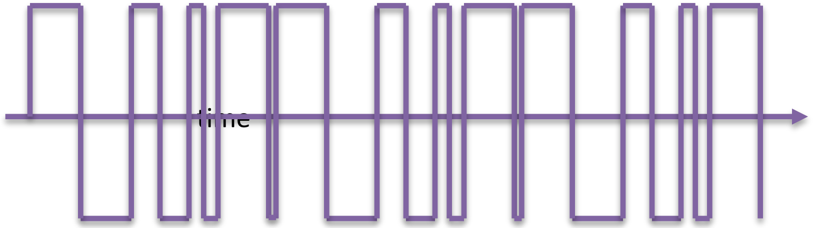
Sine

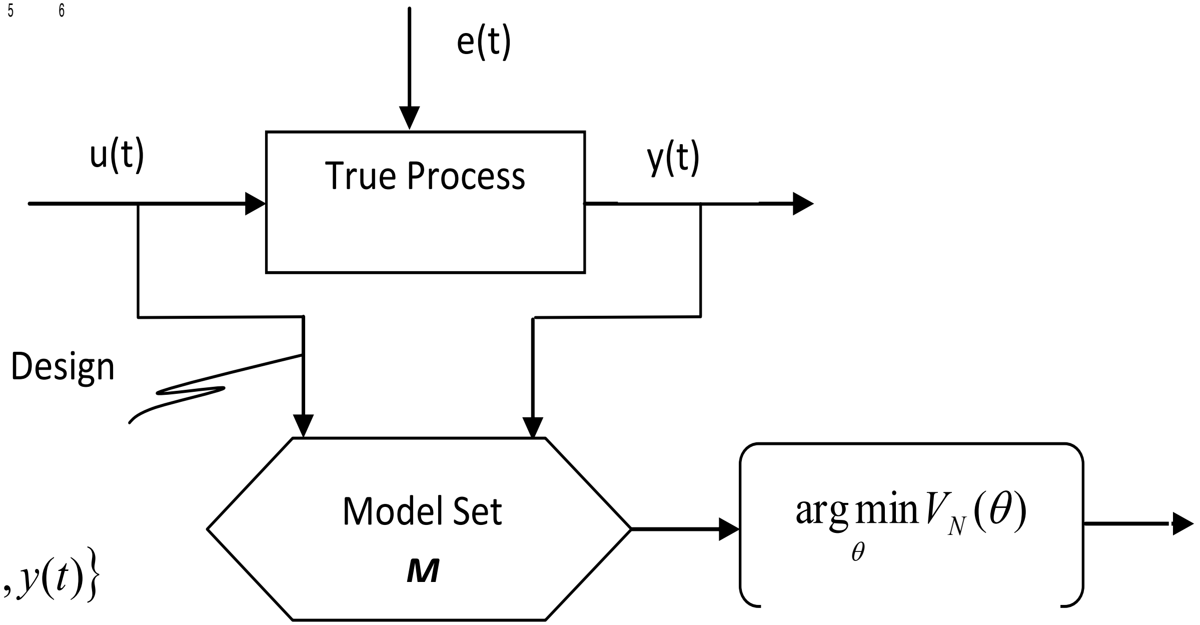


Chirp



Pseudo-Random Binary Signal (PRBS)





e(t)

u(t)

True Process

y(t)

Experiment Design

Data Set

$Z^N = \{u(t), y(t)\}$

Model Set

M

$\arg\min_{\theta} V_N(\theta)$

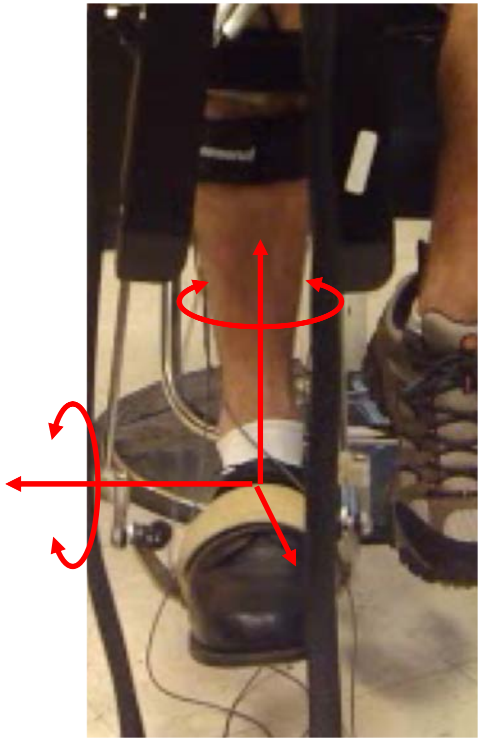
Consistency

$\hat{\theta}_N \sim \theta_0$

Identification of Ankle Impedance: Experiments

Hyunglae Lee

Professor Neville Hogan's Lab



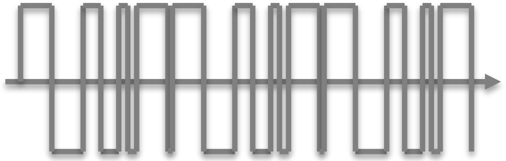
12 unimpaired young subjects

Measurements both in seated and standing postures

Two uncorrelated random perturbations (bandwidth of 100Hz) for 40 sec.

Muscle active conditions: Relaxed, TA active, SOL active, and Co-contraction

Target activation level: 10% MVC



Pseudo-Random Binary Signal

Seated

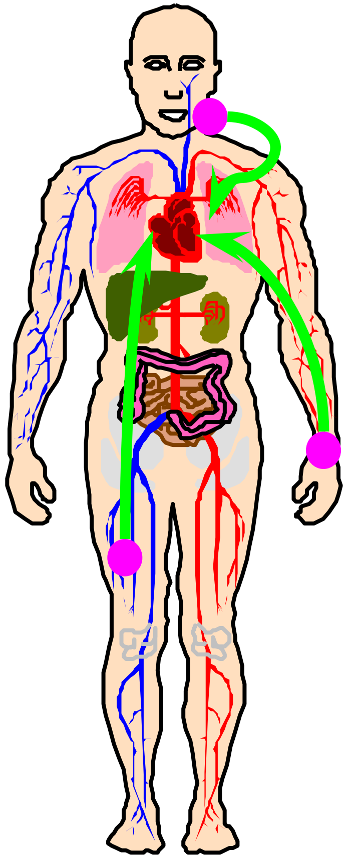
Standing

Context-Oriented Project No.3

Cardiovascular Monitoring

Identify the transfer function from cardiac output to peripheral pressure;

Based on the model estimate the cardiac output from the peripheral pressure measurement.



Noninvasive:
peripheral sensors

Arterial Tonometer

PPG Ring Sensor

Wearable

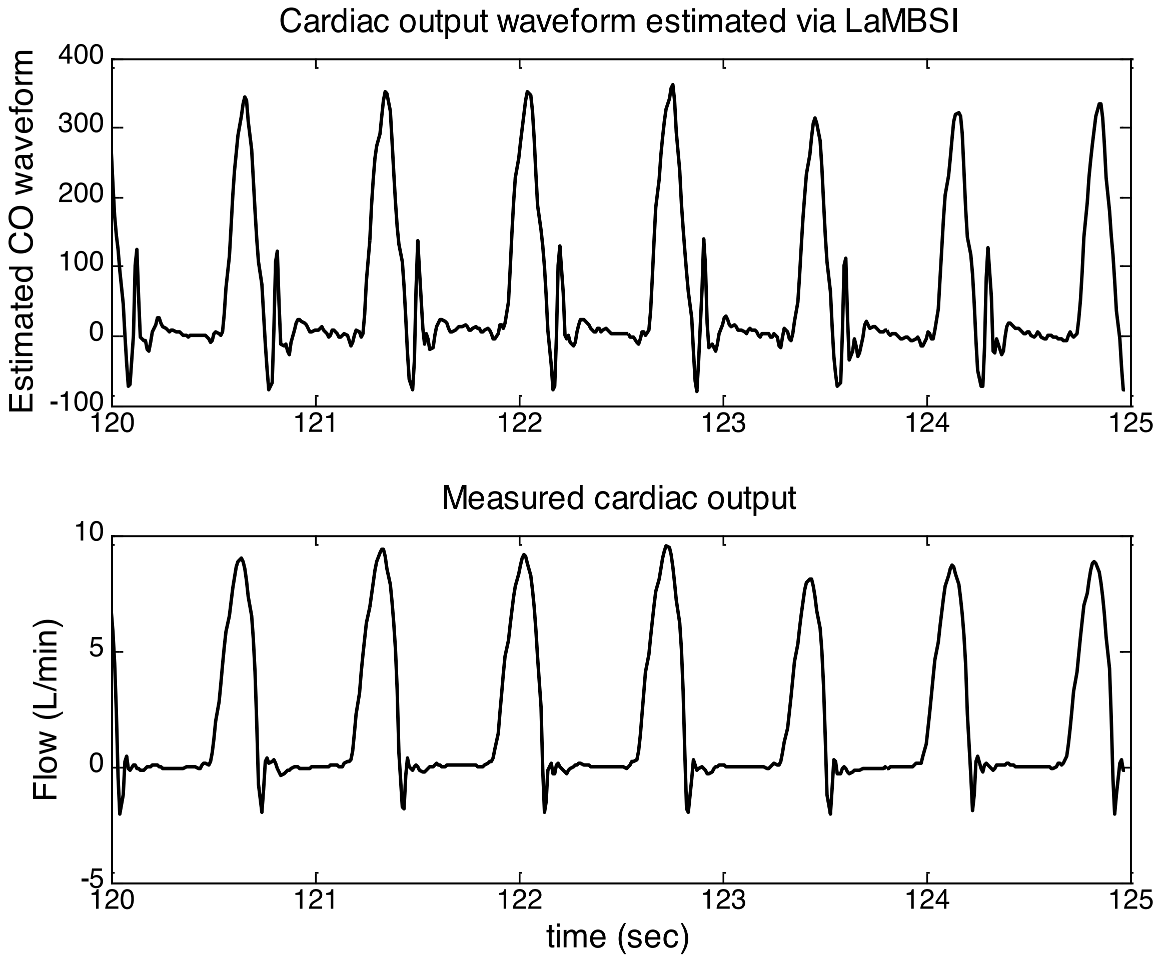
Deriving 'central' information
from 'peripheral' noninvasive measurements

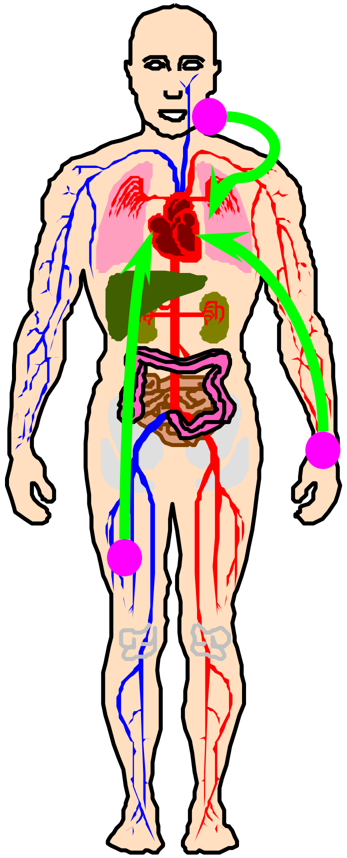
Multi-Channel Blind System Identification

Zhang and Asada, MIT

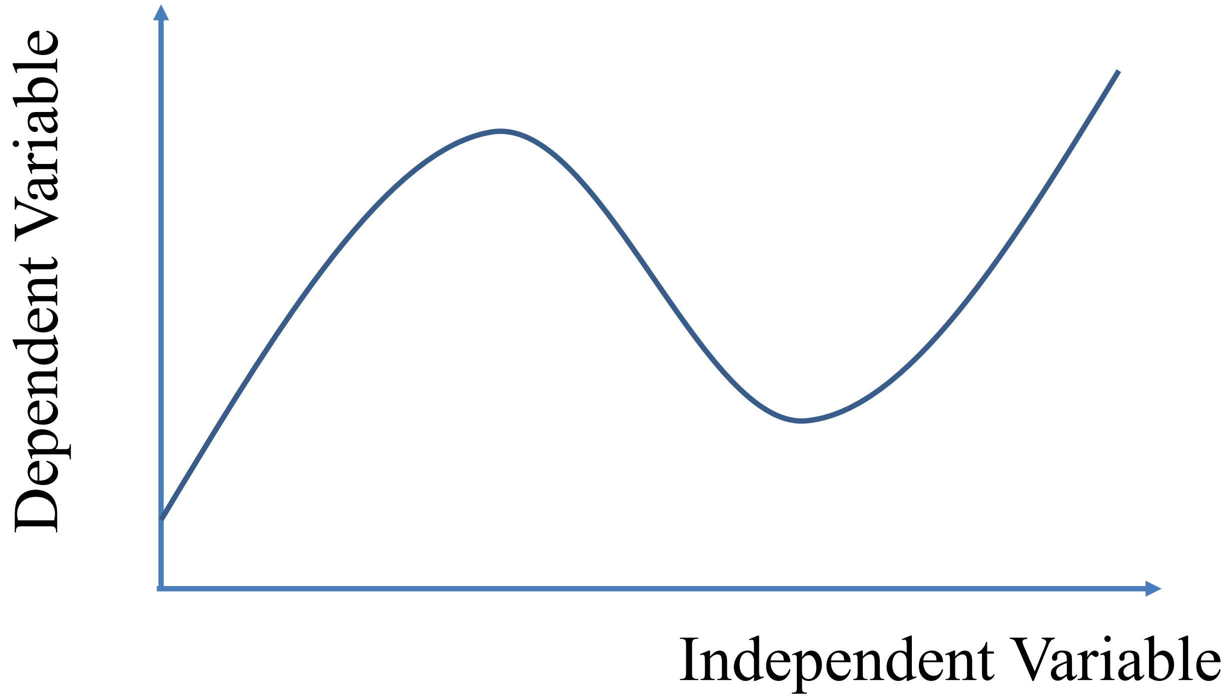
Context-Oriented Project No.3

Cardiac system identification and cardiac output waveform estimation using the Laguerre deconvolution algorithm

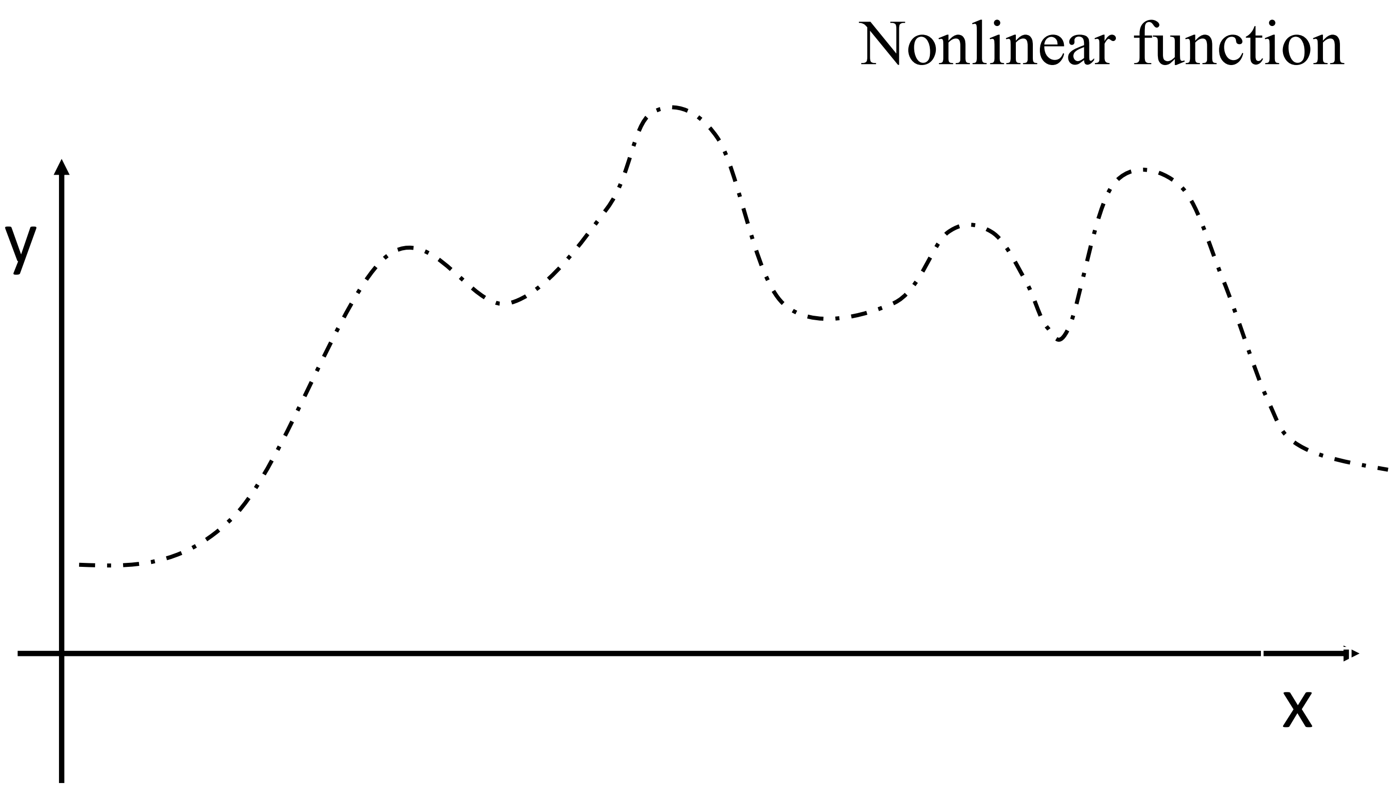




Part 4:
Machine Learning and Nonlinear System Modeling



Function Approximation and Learning

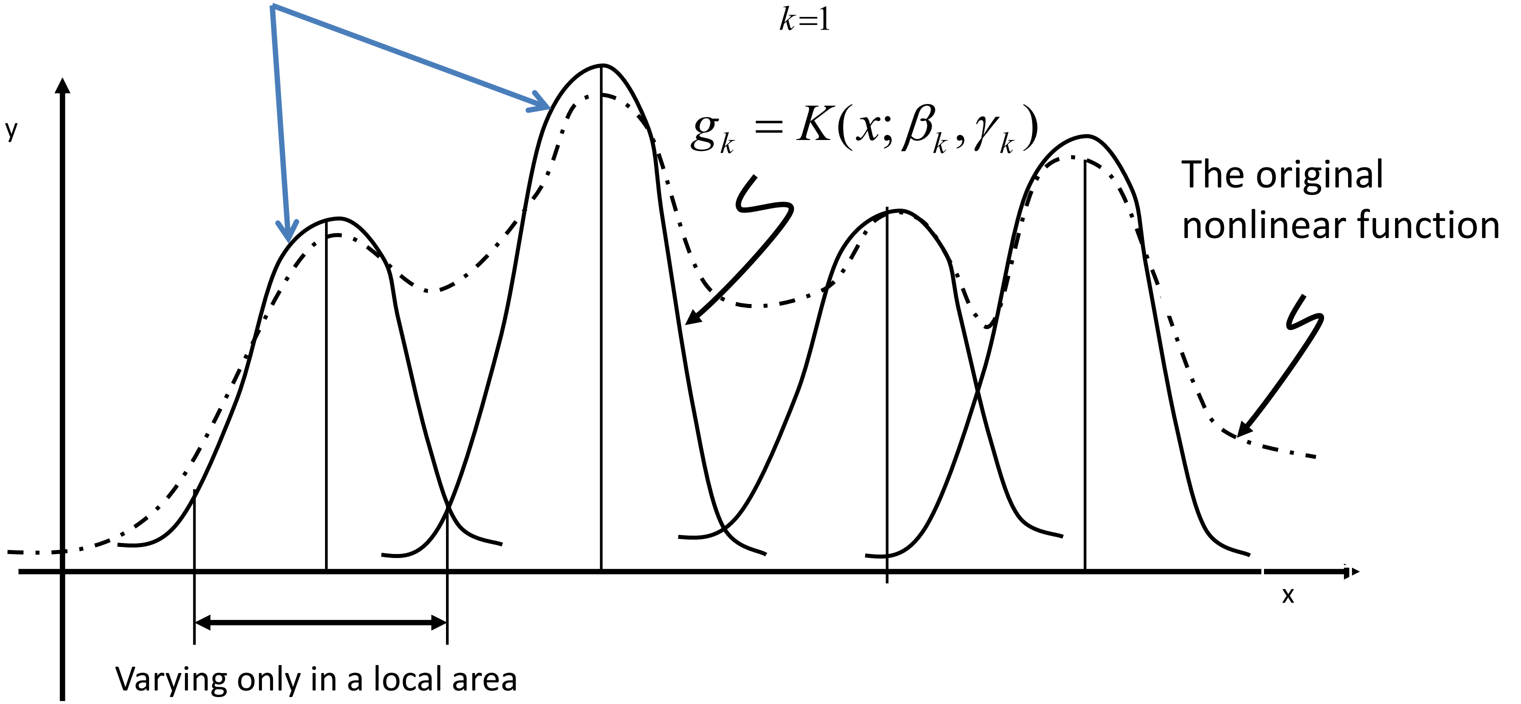


Function Approximation Theory

Expand the original nonlinear function to a series of basis functions:

Basis functions

$$g_0(x) \cong \sum_{k=1}^{m} \alpha_k g_k(x; \beta_k, \gamma_k)$$

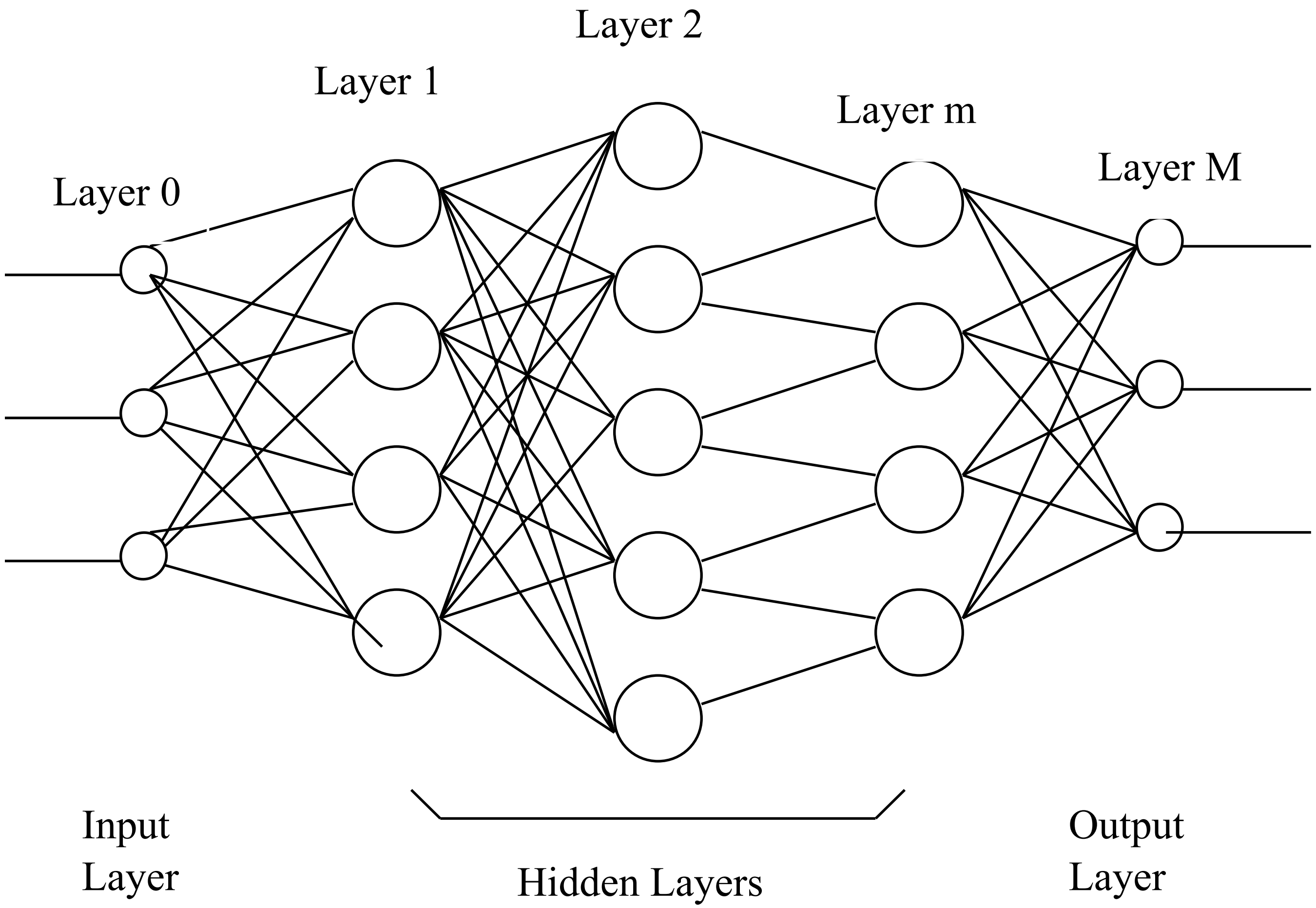

y
$g_k = K(x; \beta_k, \gamma_k)$
The original nonlinear function
x
Varying only in a local area

Basis functions:

- Radial basis functions
- Wavelets
- Neural nets

A broad class of nonlinear functions can be approximated to a series of basis functions to any accuracy. $\varepsilon - \delta$ proof

Multi-Layer Perception

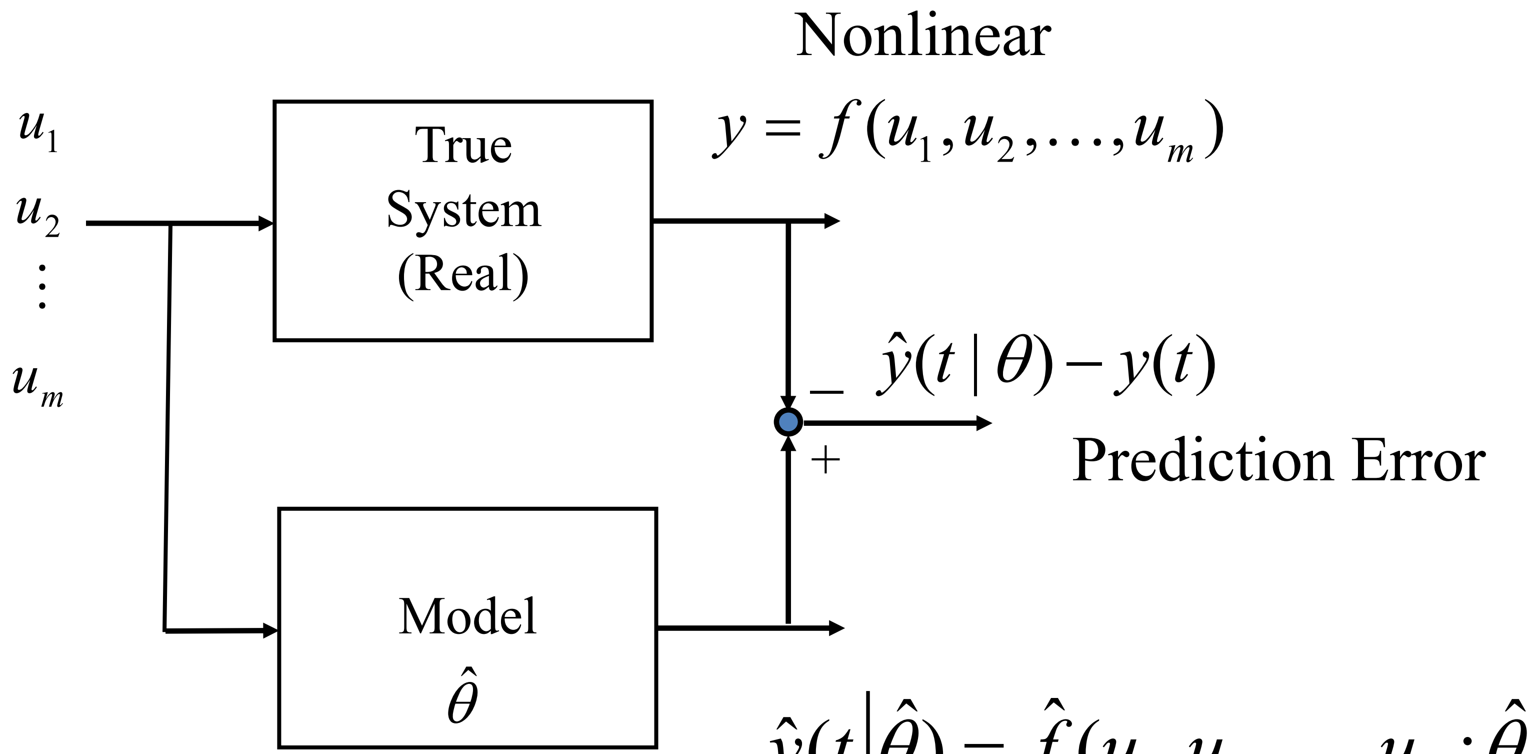


The Multi-Layer Perception is a universal approximation function that can approximate an arbitrary (measurable) function to any accuracy.

How do we train the multi-layer perceptron, given training data presented sequentially?

The Error Back Propagation Algorithm

Prediction Error Formalism extended to machine learning



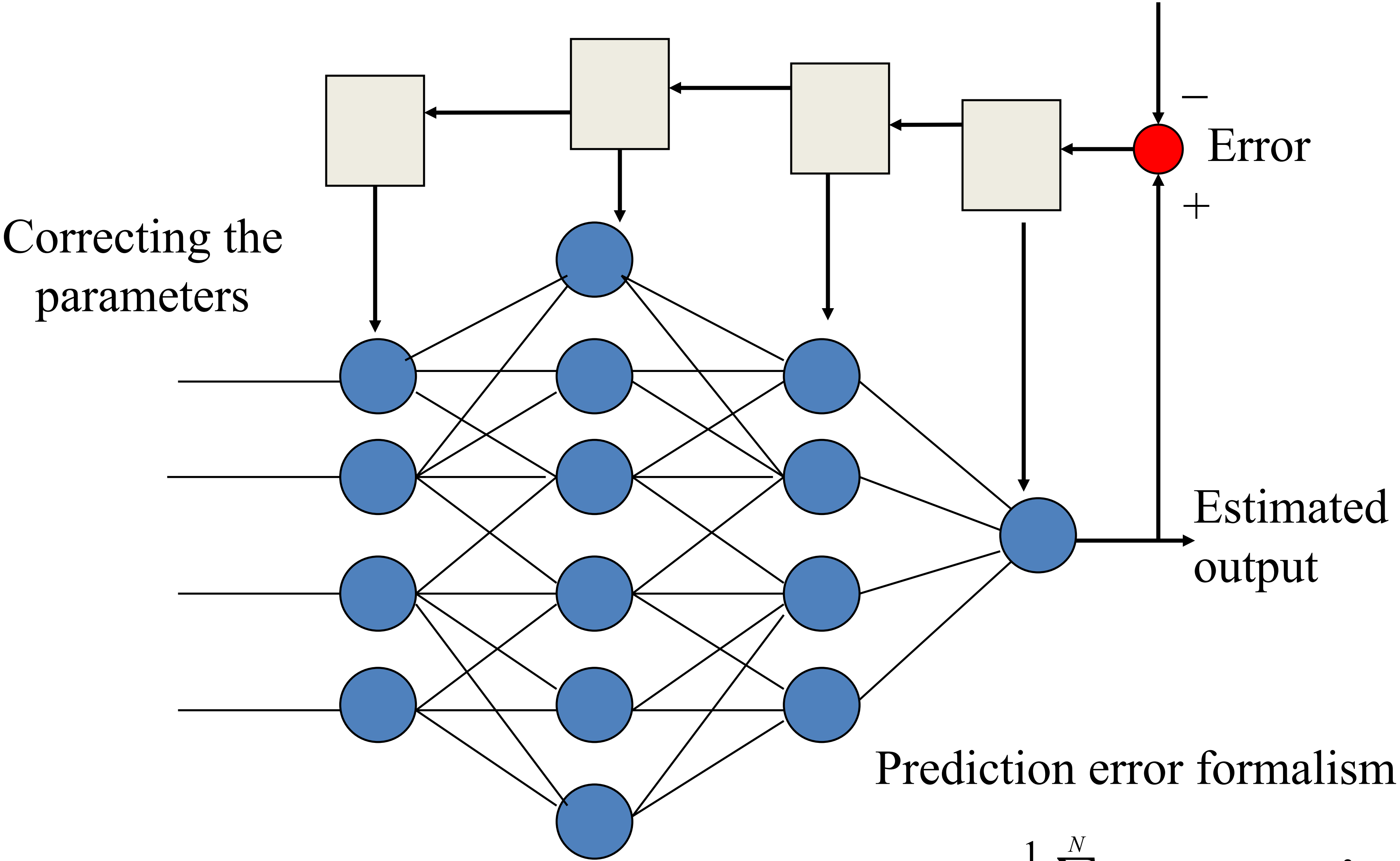
Neural Net

$$\hat{y}(t \big| \hat{\theta}) = \hat{f}(u_1, u_2, \ldots, u_m; \hat{\theta})$$

$$V_N(\theta) = \frac{1}{N} \sum_{t=1}^{N} (\hat{y}(t \mid \theta) - y(t))^2 \Rightarrow \min$$

Optimal Estimate: $\hat{\theta} = \arg\min_{\theta} V_N(\theta)$

Neural Network and Learning Algorithm



Prediction error formalism

$$V_N(\theta) = \frac{1}{N}\sum_{t=1}^{N}(\hat{y}(t\,|\,\theta) - y(t))^2$$

Error Backpropagation Algorithm

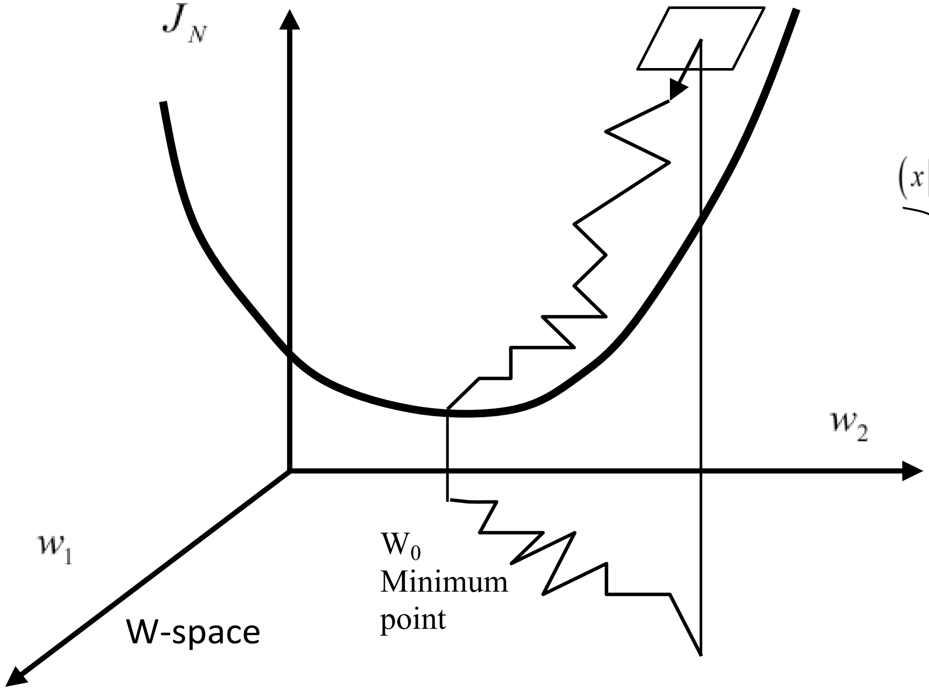
An alternative method is to execute updating the weight Δw_i *every time* the training data is presented.

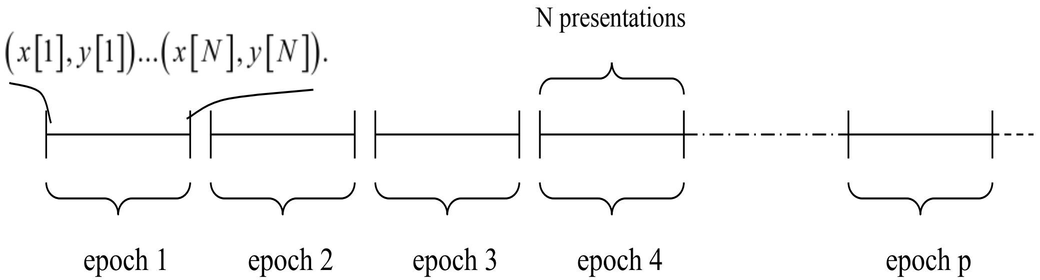
(8) $\Delta w_i[k] = \rho\delta[k]x_i[k]$ for the *k-th* presentation

(9) where $\delta(k) = y[k] - \sum w_l[k]x_l[k]$

Correct output for the training data presented at the k -th time

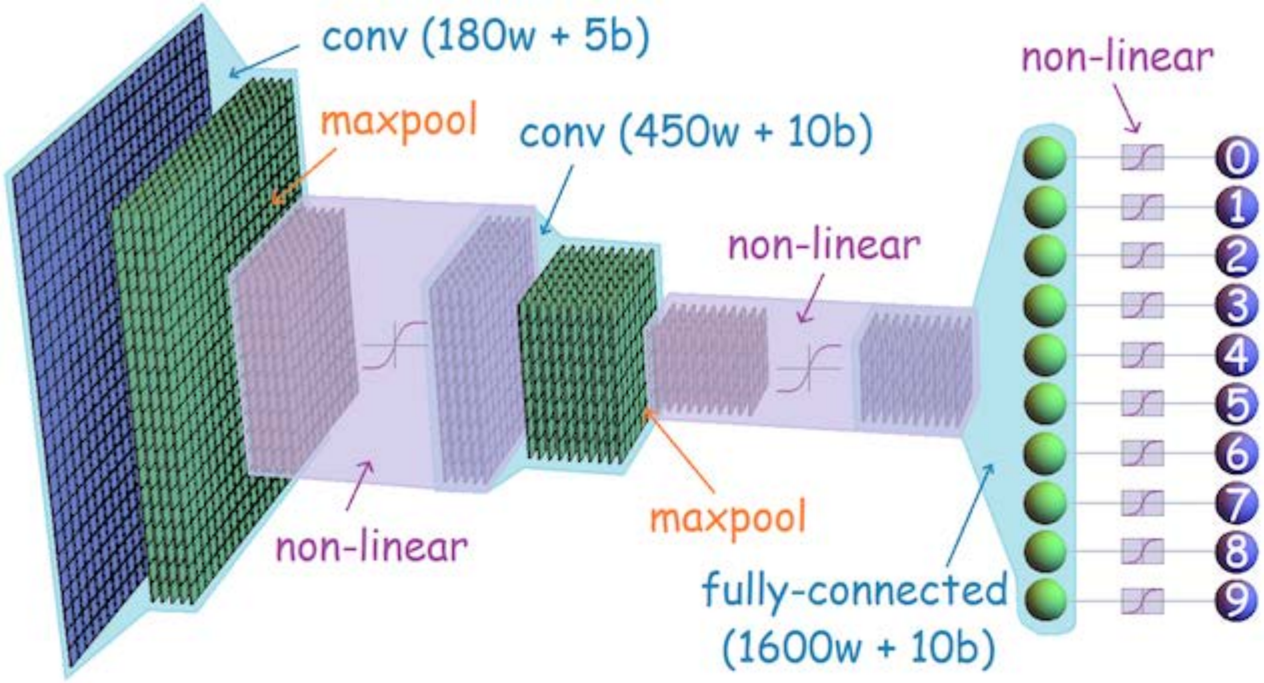
Predicted output based on the weights $w_i[k]$ for the training data presented at the k -th time





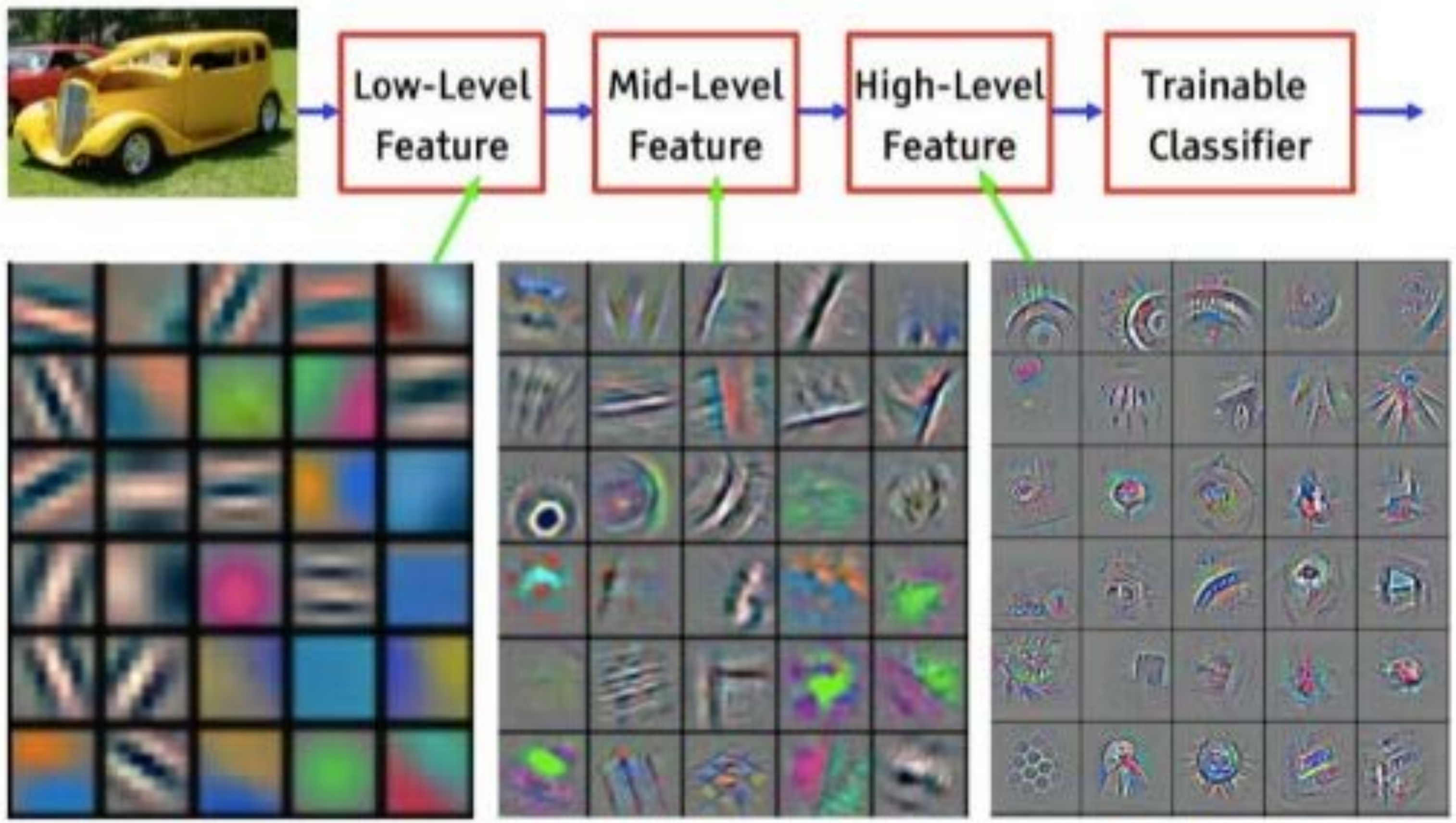
Deep Learning

- Many hidden layers : 5~20 layers
- Revised output functions
- Convolutional Neural Net (CNN)
- MaxPooling
- Big Data
- Computing power: GPU

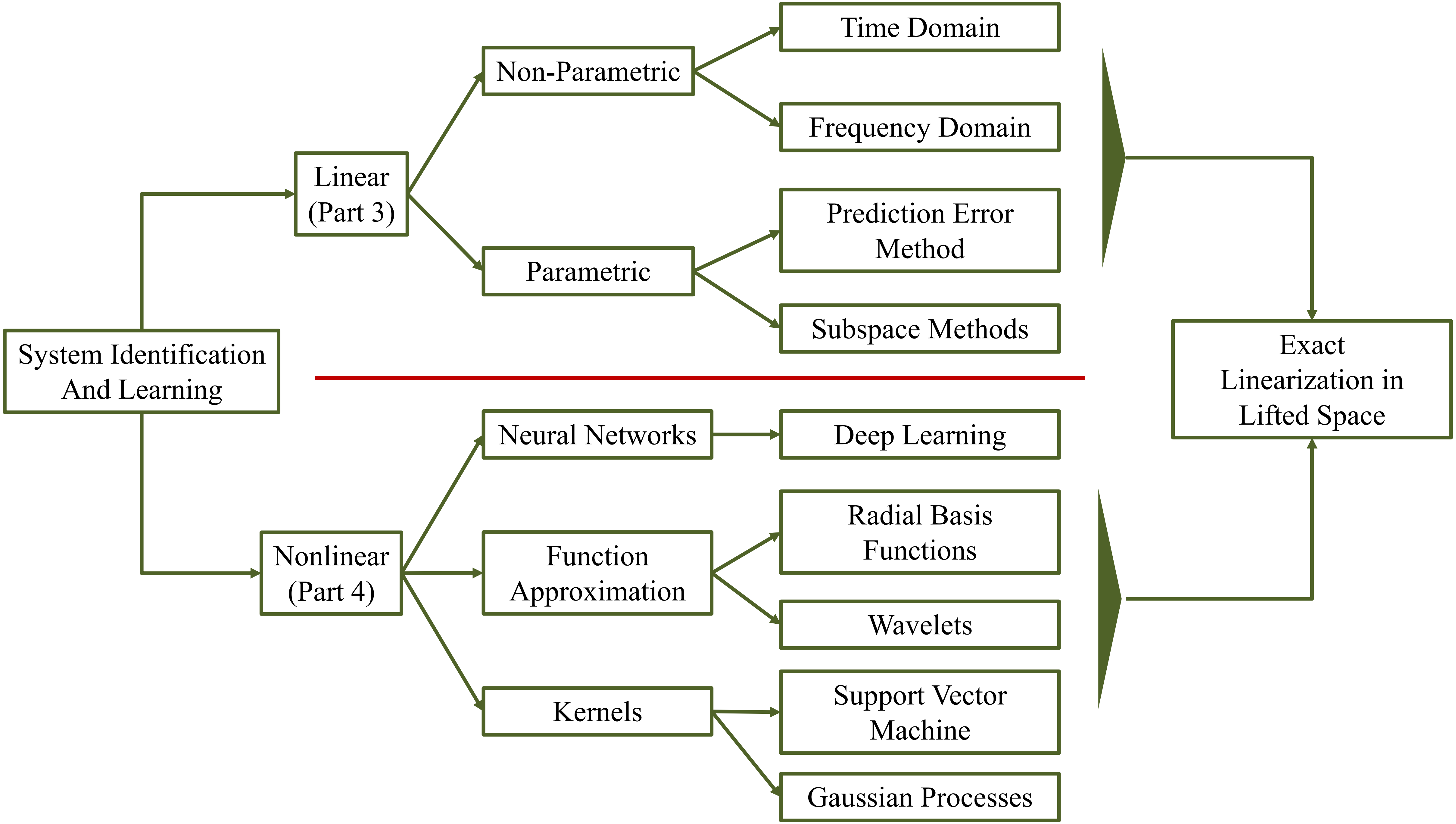


Correlation

~~Convolutional~~ Neural Network



Feature visualization of convolutional net trained on ImageNet from [Zeiler & Fergus 2013]


System Identification And Learning

Linear (Part 3)

Non-Parametric

Time Domain

Frequency Domain

Parametric

Prediction Error Method

Subspace Methods

Nonlinear (Part 4)

Neural Networks

Deep Learning

Function Approximation

Radial Basis Functions

Wavelets

Kernels

Support Vector Machine

Gaussian Processes

Exact Linearization in Lifted Space

Exact Linearization of Nonlinear Dynamical Systems

Unlike traditional linearization of nonlinear functions, Lifting Linearization is a novel method in which a nonlinear dynamical system is recast in a higher-dimensional space, so that the nonlinear system may behave linearly in the augmented, or lifted space.

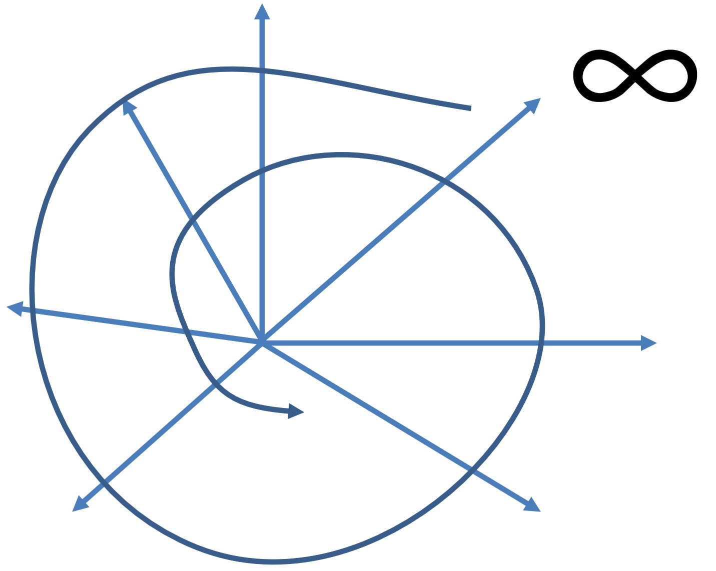
Koopman [1931] proved that a general nonlinear system can be represented as a linear system in an infinite dimensional space.

The Koopman Operator, however

- ❑ Only for autonomous systems with no input;
- ❑ Infinite dimensional space

$$\frac{dX}{dt} = F \cdot X$$

Linear



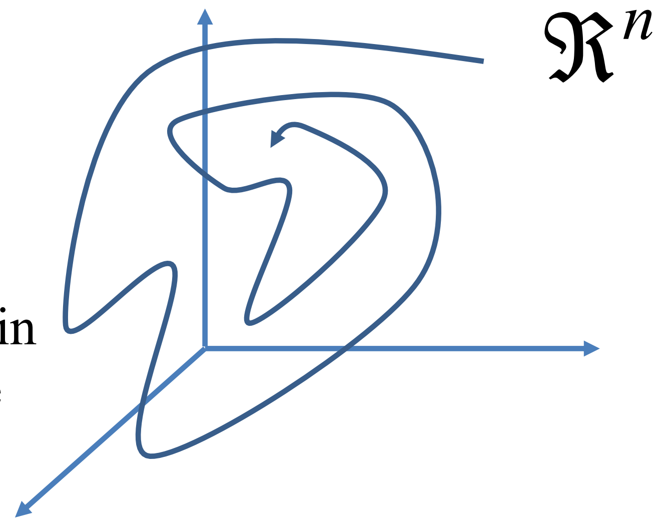
$$X = \begin{pmatrix} \varphi_1(x) \\ \varphi_2(x) \\ \varphi_3(x) \\ \vdots \end{pmatrix}$$

Observables

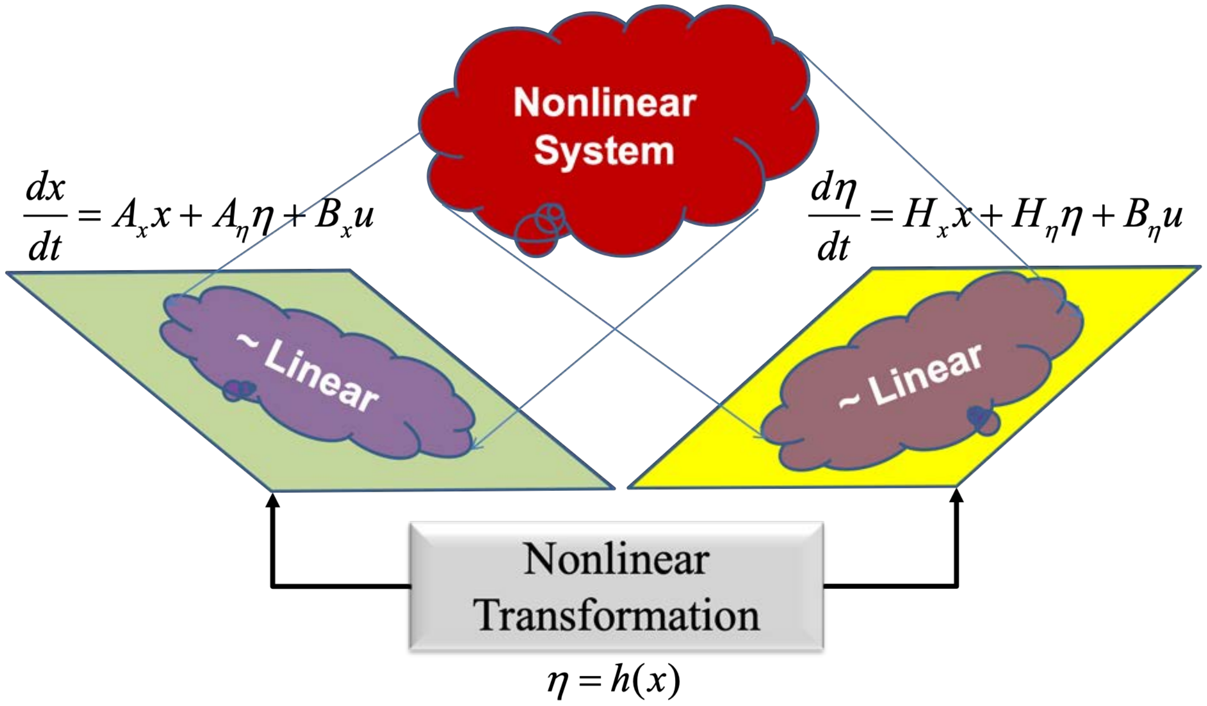
Lifting

$$\frac{dx}{dt} = f(x)$$

Complex nonlinearity in the original state space

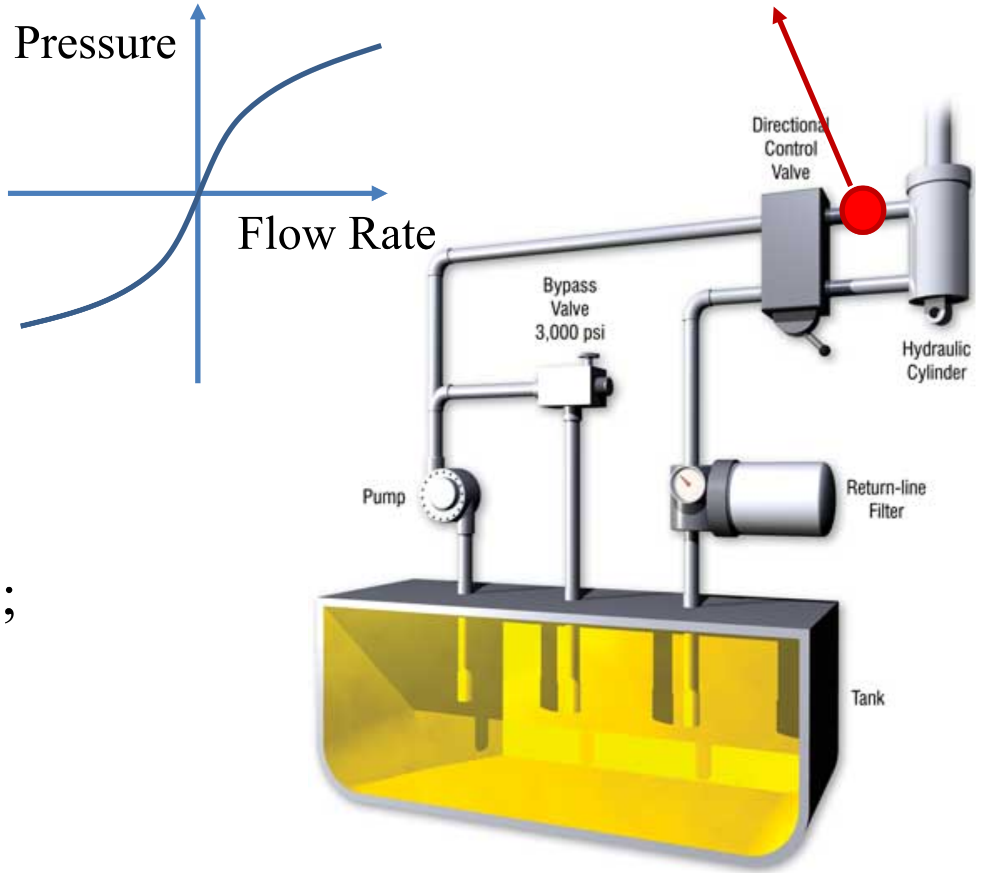


Dual-Faceted Linearization


Nonlinear System
$\frac{dx}{dt} = A_x x + A_\eta \eta + B_x u$
$\frac{d\eta}{dt} = H_x x + H_\eta \eta + B_\eta u$
~ Linear
~ Linear
Nonlinear Transformation
$\eta = h(x)$

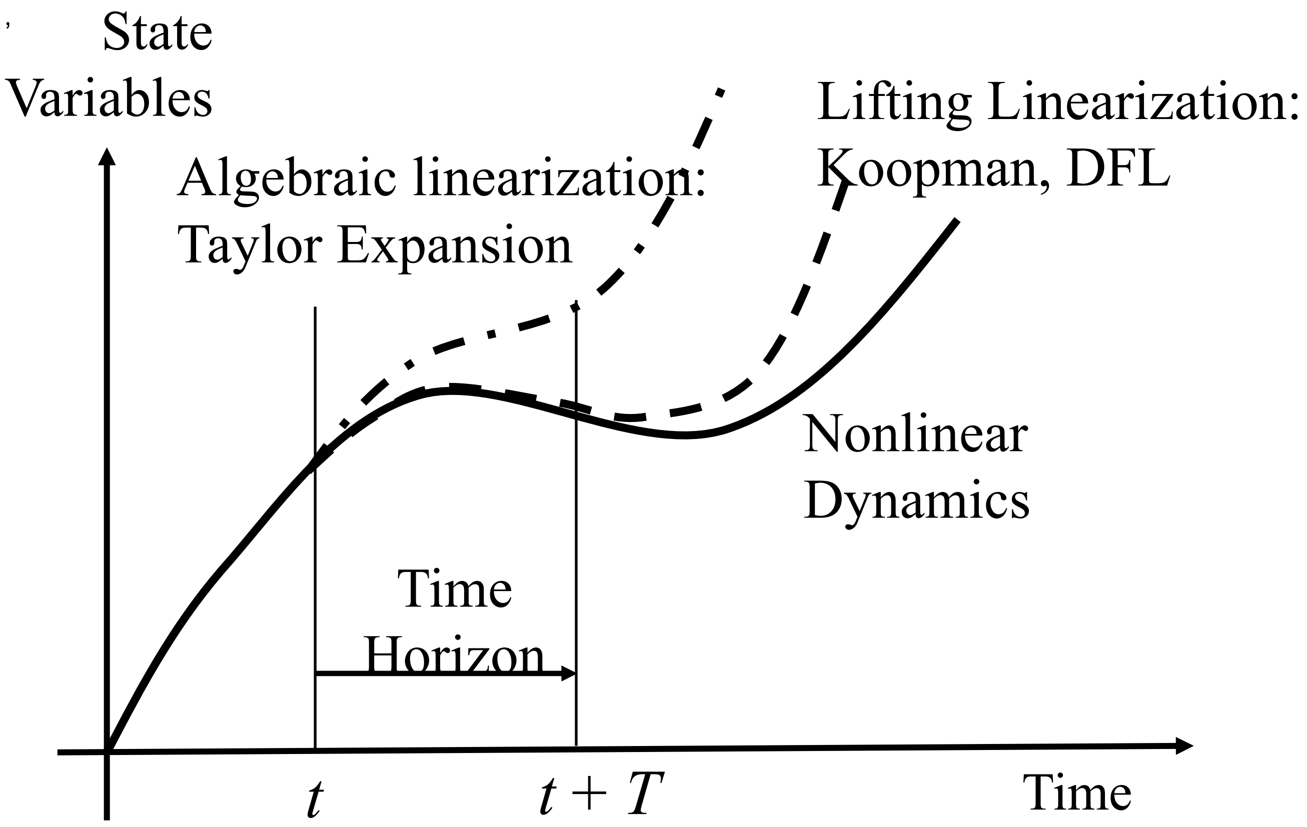
Hydraulic System: An Example

Measure both
Flow rate and pressure



- ❑ Applicable to controlled systems;
- ❑ Lifting to a finite dimensional (relatively low) space;
- ❑ Physically meaningful augmented state variables.

Application of DFL to Model Predictive Control (MPC) of Nonlinear Systems



State Variables

Algebraic linearization: Taylor Expansion

Lifting Linearization: Koopman, DFL

Nonlinear Dynamics

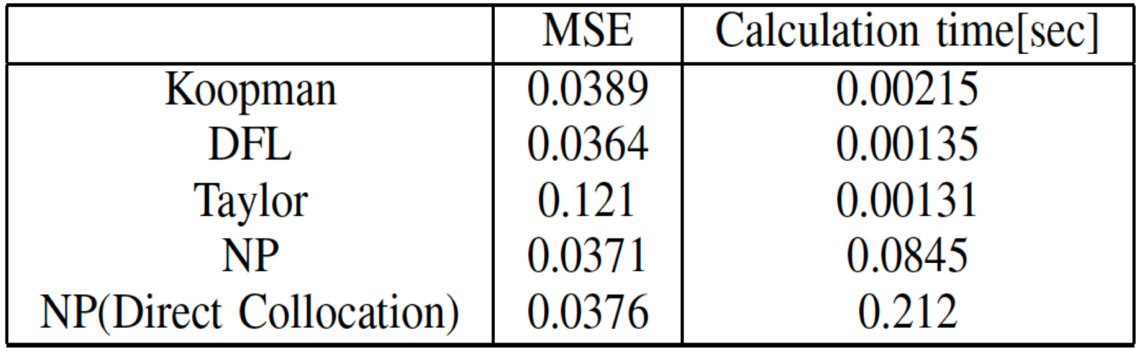
Time Horizon

t

$t + T$

Time

- ❑ The original system is nonlinear;
- ❑ Lifting the system using DFL for accurate linearization;
- ❑ Apply MPC to the linearized system;
- ❑ Convex optimization, fast computation



	MSE	Calculation time[sec]
Koopman	0.0389	0.00215
DFL	0.0364	0.00135
Taylor	0.121	0.00131
NP	0.0371	0.0845
NP(Direct Collocation)	0.0376	0.212

MPC

Minimize $P_1 = \int_t^{t+T} \ell(x(\tau),u(\tau))\,d\tau + \varphi(x(t+T))$

Subject to $\dot{x}(t) = f(x(t),u(t))$ $x_0(t) = x(t)$ $C(x(t),u(t)) \le 0$

Conclusion:

You will learn a lot in 2.160.

WHAT IS 2.160?

- Not a fancy subject, like robotics and design subjects.
- But, if you wish to learn something fundamental, establish a sold foundation, or apply an analytical and/or mathematical methodology to your thesis research, you will find 2.160 to be a useful subject.

Goal of the Subject

2.160 seeks deeper understanding, clear insights, scientifically sound methodologies, and practically useful techniques for modeling, estimation, and learning.

Syllabus and Schedule

1.	9/2-W	Introduction	*PS/Projects*	*Study Group*
	9/7-M	Labor Day Holiday, no class		
		Part 1. Regression		
2.	9/9-W	Least Squares Estimate	PS#1 Out	9/11 Kick-off
3.	9/14-M	Recursive Least Square (RLS) algorithms		
4.	9/16-W	Random processes, Active noise cancellation	Project #1 Out	9/18 PS#1
5.	9/21-M	Model Reduction – 1, Principal Component Regression	PS#1 Due,	
6.	9/23-W	Model Reduction – 2, Partial Least Squares Regression		9/25 Project#1
		Part 2. Kalman and Bayes Filters		
7.	9/28-M	Discrete Kalman Filter	Project#1 Due, PS#2 Out	
8.	9/30-W	Continuous Kalman Filter		10/2 PS#2
9.	10/5-M	Extended Kalman Filter and Unscented Kalman Filter	PS#2 Due, PS#3 Out	
10.	10/7-W	Bayes Filter		10/9 PS#3
11.	10/13-Tu	Simultaneous Localization and Mapping (SLAM)	PS#3 Due, Project #2 Out	
12.	10/14-W	Particle Filter		10/16 Project#2

			PS/Project	Study Group
		Part 3. Linear System Identification		
1.	10/19-M	Non-parametric linear system identification	Project #2 Due, PS #4 Out	
2.	10/21-W	Linear parametric systems identification		10/23 PS#4
3.	10/26-M	Asymptotic parameter distribution and Experiment design	PS#4 Due, Project#3 Out	
4.	10/28-W	Unbiased identification, Laguerre series expansion		10/30 Project#3
5.	11/2-M	Subspace methods – 1, Realization	Project#3 Due	
	11/4-W	Veterans Day Holiday, no class		
6.	11/9-M	Subspace methods - 2 , MOESP, N4SID	PS#5 Out	
		Part 4. Machine Learning and Nonlinear System Modeling		
7.	11/11-W	Universal function approximation; Radial basis function network		11/13 PS#5
8.	11/16-M	Multi-Layered Neural Network and Error Back Propagation	PS#5 Due, Project #4 Out	
9.	11/18-W	CNN, Recurrent network, and Deep neural network		11/20 Project#4
	11/23~25	Thanksgiving Holidays		
10.	11/30-M	Support Vector Machines and Kernel Trick	Project#4 Due, PS#6 Out	
11.	12/2-W	Gaussian processes for nonlinear system ID and prediction		12/4 PS#6
12.	12/7-M	Koopman operator theory for exact linearization of nonlinear dynamical systems, PS#6 Due		
13.	12/9-W	Dual-Faceted Linearization with application to nonlinear Model Predictive Control		