Lecture Notes for E211

System Identification

Fall 1999/00

Professor Jonathan P. How

SYSTEM IDENTIFICATION

- DEVELOPING AN APPROPRIATE MODEL OF A DYNAMIC SYSTEM USING OBSERVED DATA COMBINED WITH:
  - BASIC MECHANICS AND DYNAMICS
  - PRIOR KNOWLEDGE OF RELATIONSHIPS BETWEEN SIGNALS

$\Rightarrow$ INPUT / OUTPUT MODELS

- TENDS TO BE VERY EXPERIMENTAL / HEURISTIC
  - WE WILL TRY TO DEVELOP THE TOOLS REQUIRED TO PERFORM THE TASK

<u>BUT</u>

IT WILL TAKE MANY MORE TRIES (YEARS?) TO DEVELOP THE INTUITION NECESSARY TO GET <u>GOOD</u>, <u>LOW-ORDER</u> MODELS.

$\Rightarrow$ NEED TO WORK WITH REAL DATA AS MUCH AS POSSIBLE.

- WHY DO SYSTEM ID?
  - ALLOWS US TO DEVELOP MODELS FOR SYSTEMS WITH VERY COMPLEX DYNAMICS AND/OR SYSTEMS WITH UNKNOWN PHYSICAL PARAMETER VALUES.

⇒ REALLY SHOULD BE DONE IN PARALLEL WITH THE DEVELOPMENT OF AN ANALYTIC MODEL (WFF)

- PURPOSES OF IDENTIFICATION

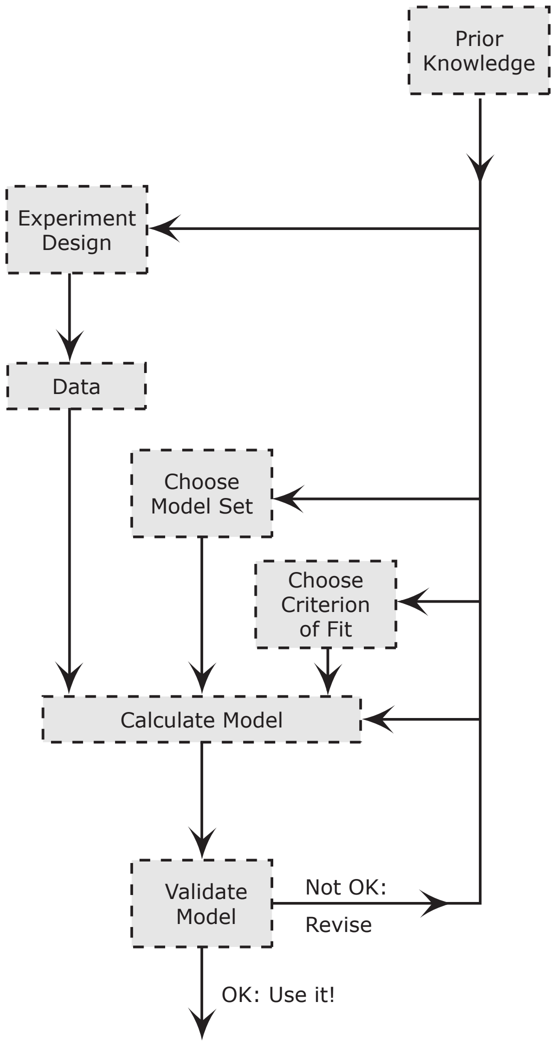
  ⇒ KEY POINT IS THAT THE MODEL ACCURACY REQUIREMENTS ARE A STRONG FUNCTION OF THE DESIRED APPLICATION.

  - CONTROL
  - ESTIMATION (OF STATES NOT AVAILABLE)
  - PREDICTION (OF RESPONSE TO DIFFERENT INPUTS)

- SYSTEM IDENTIFICATION PROCESS

- CLEARLY ITERATIVE, BUT AT MANY LEVELS

-



The System Identification Loop.

EXPERIMENT – NEED TO DESIGN EXPT. WELL TO GET GOOD DATA.

MODEL STRUCTURE – MANY CHOICES, PICK BASED ON OUR UNDERSTANDING OF SYSTEM DYNAMICS

FIT MODEL – OPTIMIZATION

EVALUATION – VALIDATE MODEL TO MAKE SURE THE FIT IS REASONABLE

EXPERIMENTS

- OPEN-LOOP OR CLOSED-LOOP ?
  - OFTEN NO CHOICE, BUT
  - CLP INTRODUCES MANY COMPLICATING FACTORS.
- WHAT IS THE INPUT SEQUENCE ?
  - FREQUENCY CONTENT ⇐ BIG IMPACT!
  - ACTUATOR LIMITS (SLEW RATES)
- SAMPLING RATE / DATA LENGTH (MEMORY)
- SISO / SIMO / MIMO SYSTEM
- DATA FILTERING
  - DRIFTS + BIASES
  - OUTLIERS
  - NOISE ATTENUATION

MODEL STRUCTURE

- NON-PARAMETRIC - TRANSFER FUNCTION PLOT
  - IMPULSE RESPONSE
- PARAMETRIC - CAPTURE DYNAMICS IN A SIMPLE STRUCTURE → $G(s) = \dfrac{s+\alpha}{s^2+\beta_1 s+\beta_2}$
- LINEAR / NONLINEAR
- MODEL SIZE (# POLES, # ZEROS)

FITTING THE MODEL

- BIG TRADE-OFF BETWEEN
  - ACCURACY
  - EASE OF SOLUTION
- DEGREE OF USER INPUT REQUIRED?
- DOES THE PROCESS ALWAYS WORK?

VALIDATION

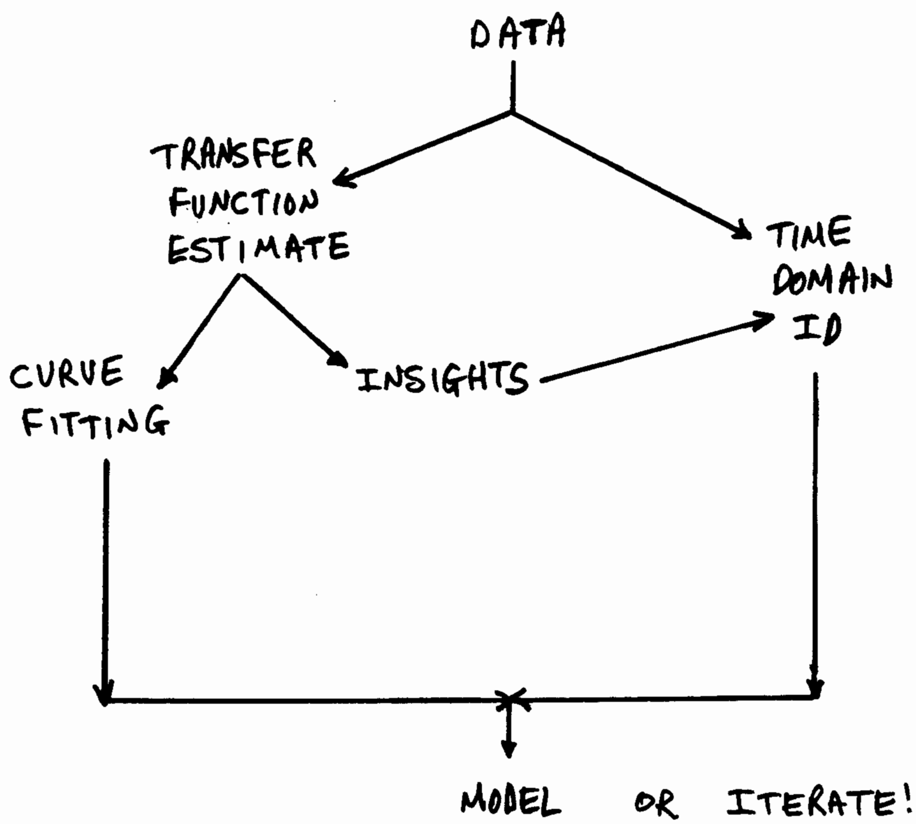
- PREDICTION AND SIMULATION
  - DIFFERENT DATA SETS
  - TIME + FREQ DOMAIN ANALYSIS OF THE ERROR
- STOCHASTIC ANALYSIS OF THE RESIDUAL ERROR
- DOES RESULT IMPLY THAT WE SHOULD MAKE CHANGES TO:
  - MODEL CHOICE (ORDER, TYPE, ...)
  - EXPT (INPUT SEQUENCE)
  - OBJECTIVE FTN FOR FIT
  - ALL?
- VALIDATE ON DIFFERENT DATA THAN THAT USED TO MAKE THE MODEL.

LECTURE #5

E211

- TRANSFER FUNCTIONS
  - ETFE + PROPERTIES
  - SMOOTHING
  - EXAMPLE

LL 6.3, 6.2, 6.4, 6.5, 6.6



E211

LECTURE # 2

- SOME ASPECTS OF DISCRETE LINEAR SYSTEM DYNAMICS
- IMPUSE RESPONSE MODELING / ESTIMATION
- DISCRETE STATE SPACE SYSTEMS
- SYSTEM REALIZATION THEORY

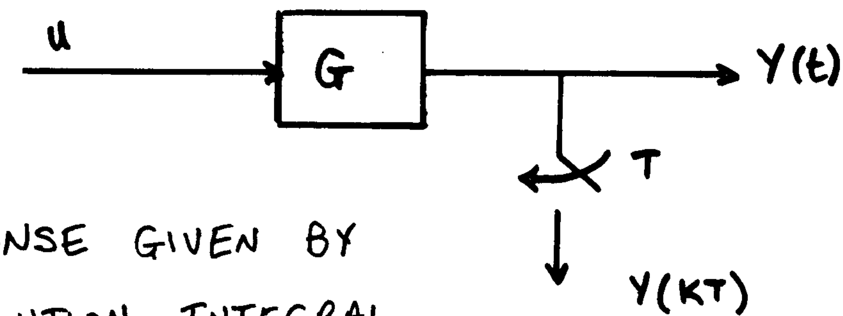
LINEAR SYSTEMS

- OUR DATA WILL TYPICALLY BE COLLECTED FROM REAL SYSTEMS

  $\Rightarrow$ DISCRETE DATA +

  DISCRETE MODELS

- CONSIDER THIS SCENARIO :



K INTEGER

- RESPONSE GIVEN BY CONVOLUTION INTEGRAL

$$y(t) = \int_{-\infty}^{\infty} g(\tau)\, u(t-\tau)\, d\tau$$

- TYPICALLY ASSUME THAT $g(\tau) = 0 \quad \forall\, \tau < 0$

  $\Rightarrow$ $g(\tau)$ DEFINES THE <u>CAUSAL</u> RELATIONSHIP BETWEEN THE INPUT $u(t)$ AND OUTPUT $y(t)$

  - CALLED THE <u>IMPULSE</u> <u>RESPONSE</u>

- COULD MODEL THE SYSTEM VERY WELL IF WE COULD FIND $g(t)$.

- ONE PROBLEM: MUST WORK IN DISCRETE TIME
  $\Rightarrow$ SLIGHTLY DIFFERENT IMPULSE RESPONSE

  - SAMPLE RESPONSE AT DISCRETE TIMES $t = KT$

$$y(KT) = \int_0^\infty g(\tau)\, u(KT-\tau)\, d\tau$$

  - TO SIMPLIFY THE ANALYSIS, ASSUME THAT THE INPUT $u(t)$ IS PIECE-WISE CONSTANT (OUTPUT OF $u(t)$ APPLIED TO A ZOH)

$$\Rightarrow \quad u(t) = u_K \quad \forall \quad KT \le t < (K+1)T$$

  - NOT ALWAYS VALID.

- NOW GET

$$y(KT) = \sum_{m=1}^{\infty} \int_{(m-1)T}^{mT} g(\tau)\, u(KT-\tau)\, d\tau$$

CONSTANT u_{K-m}

OR

$$y_K = \sum_{m=1}^{\infty} g_T(m)\, u_{K-m}$$

  - $g_T(m) = \displaystyle\int_{(m-1)T}^{mT} g(\tau)\, d\tau$ <u>IMPULSE RESPONSE OF A SAMPLED-DATA SYSTEM.</u>

$\Rightarrow$ CAN WE MEASURE THIS DIRECTLY?

TRANSFER FUNCTION FORMS

- FIRST WE INTRODUCE A LINEAR OPERATOR "q" THAT PERFORMS A "FORWARD SHIFT"

  I.E. $q\,u(t) = u(t+1)$

$\Rightarrow$ CAN REWRITE THE SYSTEM RESPONSE AS

$$y(t) = \sum_{K=0}^{\infty} g(K)\, u(t-K) \longrightarrow q^{-K} u(t)$$

$$= \left(\sum_{K=0}^{\infty} g(K)\, q^{-K} \right) u(t)$$

$$\triangleq G(q)\, u(t)$$

$\hookrightarrow$ SYSTEM TRANSFER FTN.

$$G(q) \triangleq \sum_{K=0}^{\infty} g(K)\, q^{-K}$$

- IN STATE SPACE

$$\begin{aligned} x_{K+1} &= A x_K + B u_K \\ y_K &= C x_K \end{aligned} \quad \Rightarrow \quad \begin{aligned} q\, x_K &= A x_K + B u_K \\ y_K &= C x_K \end{aligned}$$

$$\therefore \quad G(q) = \frac{y_K}{u_K} = C(qI - A)^{-1} B$$

E211

LECTURE 6

- MODELING AND MODEL FITTING
  - FREQUENCY FITS
  - TIME DOMAIN MODELS
  - BLACK BOX FORMS
  - HOW COMPUTE THE PARAMETERS
  - MODEL VALIDATION
  - DETAILED ANALYSIS LATER.

CURVE FITTING A MODEL

- SIMPLE ATTEMPT AT FIRST.

ASSUME A MODEL $G(q,\theta) = \dfrac{B(q)}{A(q)}$

$$B(q) = b_1 q^{-1} + b_2 q^{-2} + \dots + b_{n_b} q^{-n_b}$$

$$A(q) = 1 + a_1 q^{-1} + \dots + a_{n_a} q^{-n_a}$$

$$\theta = \left[a_1 \;\; a_2 \; \dots \; a_{n_a} \;\; b_1 \;\; b_2 \; \dots \; b_{n_b} \right]^T$$

$\Rightarrow$ PARAMETER VECTOR

- DEFINE ERROR CRITERION $V_N(\theta)$

$$V_N(\theta) = \sum_{\omega_i} \alpha_i \left| \hat{G}_N(e^{j\omega_i}) - G(e^{j\omega_i}, \theta) \right|^2$$

- PICK θ TO MIN V_N
- USUALLY USE $\alpha_i = |U_N(\omega)|^2$
- V_N NONLINEAR FUNCTION OF θ $\Rightarrow$ MUST USE GRADIENT SEARCH TECHNIQUES.

- GET A SIMPLER PROBLEM IF $\alpha_i = |U_N(\omega)|^2 \cdot |A(e^{i\omega})|^2$

$$\Rightarrow \hat{V}_N(\theta) = \sum_{\omega_i} \left| A(e^{i\omega}) Y_N(\omega) - B(e^{i\omega}) U_N(\omega) \right|^2$$

- QUADRATIC IN θ $\Rightarrow$ USE LEAST-SQUARES TECHNIQUES.
- GOOD PLACE TO START $\Rightarrow$ SEE HW4.

LOGARITHMIC LEAST SQUARES

- ONE OF THE PROBLEMS WITH THE LINEAR LEAST SQUARES FREQUENCY FIT IS THAT ZEROS OF THE SYSTEM ARE <u>FIT VERY POORLY</u>
  - ESPECIALLY TRUE FOR LIGHTLY DAMPED ZEROS

- REASON IS THAT THE INDEX ASSIGNS A VERY SMALL PENALTY TO ERRORS IN THE MATCH OF THE DATA/MODEL IN LOW-GAIN REGIONS
  $\Rightarrow$ NEAR THESE ZEROS THE ABSOLUTE DIFFERENCE IN THE FREQUENCY RESPONSES IS SMALL (RELATIVE ERRORS ARE LARGE)

- LINEAR LEAST SQUARES PUTS TOO MUCH EMPASIS ON FITTING THE POLES

- LOG. LEAST SQUARES BETTER FOR THE FITTING OF THE ZEROS SINCE IT WEIGHTS THE RATIO OF MODEL GAIN TO MEASUREMENT GAIN.

$$J_{LIN} = \sum_{\omega_i} \left| \hat{G}_N(e^{j\omega_i}) - G(e^{j\omega_i}, \theta) \right|^2$$

$$J_{LLS} = \sum_{\omega_i} \left| \mathrm{LOG}\left(G(e^{j\omega_i}, \theta) \right) - \mathrm{LOG}\left(\hat{G}(e^{j\omega_i}) \right) \right|^2$$

- J_{LLS} WORKS MUCH BETTER FOR A SYSTEM WITH LARGE DYNAMIC RANGE (SIDMAN, IEEE TAC, VOL 36 PAGE 1065, 1991)

- <u>EXAMPLE</u>: CONSIDER TWO POINTS IN THE TF

  - HAVE $\hat{G}(\omega_1) = 10$ MEASURED

    $\hat{G}(\omega_2) = 0.1$

  - OUR MODEL ESTIMATES THESE AS

    $G(\omega_1) = 9$

    $G(\omega_2) = 0.09$

$\Rightarrow$ CHECK CONTRIBUTION TO THE COST FUNCTIONS

$$|\hat{G} - G|^2 \rightarrow \quad @\,\omega_1 \quad (10-9)^2 = 1$$

$$@\,\omega_2 \quad (0.1-0.09)^2 = 0.0001 \ll 1$$

$$|\mathrm{LOG}(G) - \mathrm{LOG}(\hat{G})|^{2\ 2} \rightarrow \quad @\,\omega_1 \quad (2.3026-2.1972)^2 = 0.0111$$

$$(-2.3026+2.4079)^2 = 0.0111$$

$$\downarrow$$

$$\mathrm{LOG}(G/\hat{G})$$

- COMMERCIAL SOFTWARE AVAILABLE
  - WORK VERY WELL ON CLEAN DATA.

PARAMETRIC MODELS OF LINEAR SYSTEMS

- TO DEVELOP A MORE GENERAL RESULT, WE NEED TO SPECIFY WHAT SYSTEM MODELS WE WILL USE.

  - RECALL THAT WE ASSUMED THAT THE ACTUAL SYSTEM DYNAMICS ARE OF THE FORM:

$$y(t) = G_a(q)\, u(t) + H_a(q)\, e(t)$$

- $u(t)$ KNOWN (WE APPLIED IT)
- $e(t)$ IS A SEQUENCE OF IND. RANDOM VARIABLES
  - USUALLY ZERO MEAN
  - OFTEN ASSUME GAUSSIAN WHITE NOISE
  - IS $E[e^2]$ KNOWN?

- WRITE THE LINEAR TRANSFER FUNCTIONS AS A RATIO OF POLYNOMIALS IN q:

$$G_a(q) = \frac{B(q)}{A(q)} = q^{-\bar{n}_k} \frac{b_1 q^{-1} + \dots + b_{n_b} q^{-\bar{n}_b}}{1 + a_1 q^{-1} + \dots + a_{n_a} q^{-\bar{n}_a}}$$

n_k DELAYS

$$H_a(q) = C_a(q)/D_a(q)$$

- <u>ISSUES</u>: - DO ALL OF A_a, B_a, C_a, D_a EXIST?
  - VALUES FOR $\bar{n}_k, \bar{n}_a, \bar{n}_b, \bar{n}_c, \bar{n}_d$?

- **APPROACH** → MAKE SOME APPROXIMATIONS BY SELECTING FROM SEVERAL STANDARD MODEL FORMS.
  ⇒ "BLACKBOX"

⇒ SET OF MODELS $y = G(q,\theta)u + H(q,\theta)e$

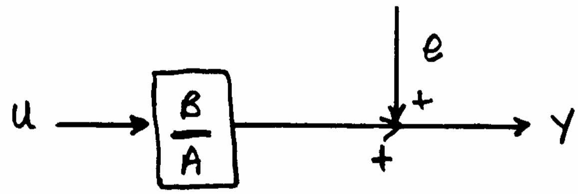
θ - PARAMETER VECTOR INCLUDING COEFFICIENTS OF TRANSFER FUNCTIONS (POSSIBLY $E[e^2] = \lambda$ AS WELL).

- **PROCEDURE**
  1. PICK A MODEL FORM
  2. SELECT $n_a, n_b, n_c, n_d, \ldots$

- **ISSUES**
  - GENERALITY OF THE NOISE MODEL (H)
  - COMPLEXITY OF THE PARAMETRIZATION
    ⇒ IMPACT ON THE OPTIMIZATION?
  - PHYSICAL INSIGHT? + HOW TO APPLY IT.

STANDARD MODEL FORMS

- **OUTPUT ERROR**



$G(q,\theta)$ $\quad H(q,\theta) = 1$

$$Y = \frac{B}{A} u + e$$

  - NOISE SOURCE IS THE DIFFERENCE (ERROR) BETWEEN ACTUAL / NOISE-FREE OUTPUT.
  - GOOD TO USE WHEN SYSTEM DOMINATED BY WHITE SENSOR NOISE
  - EXPECT PROBLEMS WHEN NOISE SPECTRUM IS SHAPED (COLORED NOISE / PROCESS NOISE).
    - WHY?

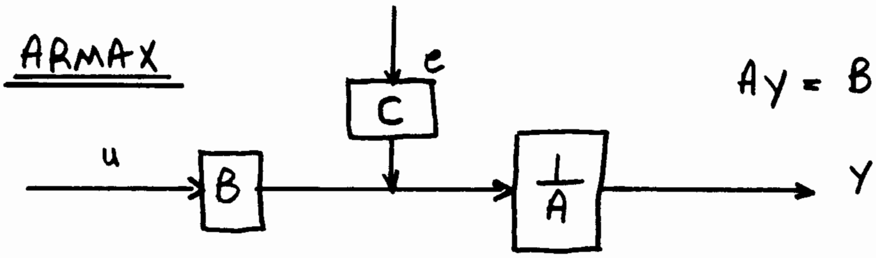
- **DIFFERENCE EQUATION**

$$w(t) + a_1 w(t-1) + \dots + a_{n_a} w(t-n_a)$$
$$= b_1 u(t-1) + \dots + b_{n_b} u(t-n_b)$$

$$y(t) = w(t) + e(t).$$

- SEE WHY THEY ARE CALLED "BLACK BOX"?

- **ARMAX**



$$Ay = Bu + Ce$$

$$G(q,\theta) = \frac{B}{A}$$

$$H(q,\theta) = \frac{C}{A}$$

- DISTURBANCE AND INPUT SUBJECT TO SAME POLES
- GOOD MODEL IF SHAPED OR PROCESS NOISE DOMINATES.
- PARAMETERIZED BY

$$\theta = [a_1 \ldots a_{n_a}\ b_1 \ldots b_{n_b}\ c_1 \ldots c_{n_c}]^T$$

$$A(q) = 1 + a_1 q^{-1} + \ldots + a_{n_a} q^{-n_a}$$

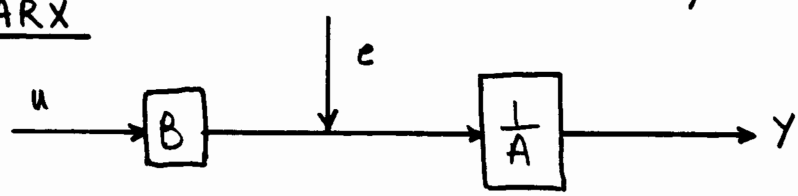
$$B(q) = b_1 q^{-1} + \ldots + b_{n_b} q^{-n_b}$$

$$C(q) = 1 + c_1 q^{-1} + \ldots + c_{n_c} q^{-n_c}$$

- **DIFFERENCE EQUATION**

$$y(t) + a_1 y(t-1) + \ldots + a_{n_a} y(t-n_a)$$
$$= b_1 u(t-1) + \ldots + b_{n_b} u(t-n_b)$$
$$+ e(t) + c_1 e(t-1) + \ldots + c_{n_c} e(t-n_c)$$

- ARX



$$Ay = Bu + e$$

$$G(q,\theta) = \frac{B}{A}$$

$$H(q,\theta) = \frac{1}{A}$$

- SIMPLIFIED DISTURBANCE MODEL
- NOT PARTICULARLY WELL MOTIVATED BY ANY PHYSICAL INTUITION.
- BENEFITS OF THIS MODEL ARE THAT IT LEADS TO A <u>VERY SIMPLE NUMERICAL FORMULATION.</u>

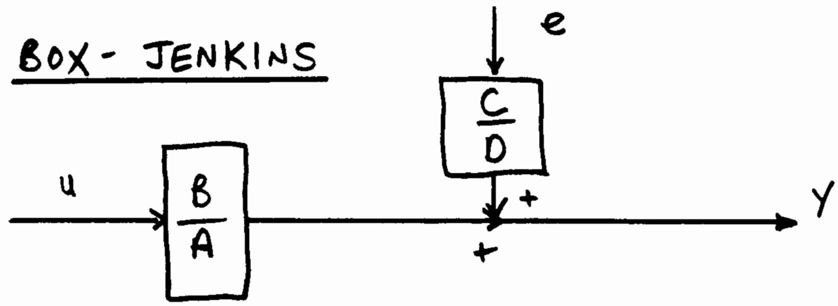
$$A(q) = 1 + a_1 q^{-1} + \dots + a_{n_a} q^{-n_a}$$

$$B(q) = b_1 q^{-1} + \dots + b_{n_b} q^{-n_b}$$

- DIFFERENCE EQUATION

$$y(t) + a_1 y(t-1) + \dots + a_{n_a} y(t-n_a)$$
$$= b_1 u(t-1) + \dots + b_{n_b} u(t-n_b) + e(t)$$

- BOX - JENKINS



$G(q,\theta) = \frac{B}{A}$

$H(q,\theta) = \frac{C}{D}$

  - VERY GENERAL FORM
  - INCLUDES ALL OTHERS AS SPECIAL CASES
  - NOW NEED TO PICK n_a, n_b, n_c, n_d

- NOTE THAT WE HAVE GIVEN THESE DIFF. EQUATIONS WITH NO EXTRA DELAYS ADDED.
  - $B(q)$ ASSUMES 1 DELAY
  - OFTEN NEED n_K DELAYS

$\Rightarrow$ $B(q) = q^{-n_K}\left(b_1 q^{-1} + \dots + b_{n_b} q^{-n_b}\right)$

$\Rightarrow$ DIFFERENCE EQUATION NOW OF THE FORM

$$\dots = b_1 u(t-n_K-1) + b_2 u(t-n_K-2) + \dots$$

POLYNOMIAL FORM

- MUCH OF THE TOOLBOX DEALS WITH TF GIVEN IN A POLYNOMIAL FORMAT
  - ORDER COEFFICIENTS IN ASCENDING POWER OF THE DELAY OPERATOR q^{-K}, $K = 0, 1, 2, \ldots$
  - KEY POINT: DELAYS ARE DENOTED BY LEADING ZEROS IN THE POLYNOMIAL

$$\frac{B(q)}{A(q)} = \frac{q^{-3}}{1 - 1.5q^{-1} + 0.7q^{-2}} \Rightarrow \begin{array}{l} B = [0\ 0\ 0\ 1] \\ A = [1\ \ -1.5\ \ 0.7] \end{array}$$

- IN DISCRETE TIME (Z FORM), WOULD NORMALLY WRITE THIS AS

$$\frac{z^{-3}}{1 - 1.5z^{-1} + 0.7z^{-2}} = \frac{1}{z^3 - 1.5z^2 + 0.7z} = \frac{NUM}{DEN}$$

$$\Rightarrow \begin{array}{l} NUM = 1 \\ DEN = [1\ \ -1.5\ \ 0.7\ \ 0] \end{array}$$

- THESE TWO REPRESENTATIONS ARE EQUIVALENT IF THE LENGTHS OF A AND B ARE EQUAL.

TH2POLY
POLY2TH

- MODELS SUMMARY
  - EACH MODEL VALID FOR DIFFERENT ASSUMPTIONS ON THE DYNAMICS / NOISE SPECIFICATIONS

  ⇒ OFTEN NOT CLEAR WHICH IS THE BEST TO USE !!

  ⇒ TRY SEVERAL AND SEE IF THE FIT IMPROVES. (HOW COMPARE?)

  - COEFFICIENT ORDER NOT OBVIOUS EITHER

  ⇒ USE FREQUENCY DOMAIN GRAPHS TO DEVELOP INSIGHTS?

  ⇒ TOOLS EXIST TO ENABLE YOU TO TRY A RANGE OF POSSIBLE VALUES.

  ⇒ TYPICALLY GET A BETTER FIT AS WE INCREASE THE ORDER OF A

  ⇒ AVOID "OVER FITTING" THE DATA.

  RESULT LOOKS GOOD ON THIS DATA SET, BUT MUCH POORER ON ANY OTHER.

FITTING PARAMETERIZED MODEL

- **BASIC APPROACH** - MINIMIZE THE PREDICTION ERRORS

  ⇒ USE APPROXIMATE MODEL TO PREDICT $y(t)$ GIVEN ALL DATA UPTO $t-1$

  ⇒ FORM PREDICTION ERROR

  $$\epsilon(t) = y(t) - \hat{y}(t|t-1)$$

  CLEARLY A FUNCTION OF θ

  ⇒ PICK PARAMETERS (θ) OF MODEL TO MINIMIZE $V_N(\theta)$.

  TYPICALLY $V_N(\theta) = \frac{1}{N}\sum_{t=1}^{N} \epsilon^2(t,\theta)$

  $$\hat{\theta}_N = \arg\min_{\theta} V_N(\theta).$$

- OTHER COST FUNCTIONS USED TOO

  $$V_N(\theta) = \frac{1}{N}\sum_{t=1}^{N} L(\epsilon(t,\theta))$$

  L - LOSS FUNCTION

  $L \geq 0$, SCALAR.

- GENERAL FORM OF THE PREDICTOR ON 6-12

$$\hat{y}(t|t-1) = (1-H^{-1})\,y(t) + H^{-1}G\,u(t)$$

- WHERE FROM? SEE APPENDIX. (A1)
  - EXPLICITLY ASSUMES THAT H AND H^{-1} ARE STABLE

- **NOTE :** $H = \dfrac{C(q)}{D(q)} = \dfrac{1 + c_1 q^{-1} + \ldots + c_{n_c} q^{-n_c}}{1 + d_1 q^{-1} + \ldots + d_{n_d} q^{-n_d}}$

  - ASSUME $n_c \geq n_d$ (ACTUALLY $n_c = n_d$ BELOW)

  $\Rightarrow \; 1 - H^{-1} = 1 - \left(\dfrac{C}{D}\right)^{-1} = \dfrac{C-D}{C}$

BUT $\dfrac{C-D}{C} = \dfrac{(c_1 - d_1)q^{-1} + \ldots + (c_{n_c} - d_{n_c})q^{-n_c}}{1 + c_1 q^{-1} + \ldots + c_{n_c} q^{-n_c}}$

$\therefore$ $(1 - H^{-1})\,y(t)$ ONLY CONTAINS OLD VALUES OF THE OUTPUT $[\,y(s),\; s \leq t-1\,]$

- SINCE $G \sim \dfrac{B}{A}$, $H^{-1}G\,u(t)$ ONLY INCLUDES OLD VALUES OF $u(t)$ $[\,u(s),\; s \leq t-1\,]$

- **SPECIAL CASES:**

(OE)

$$H = 1 \qquad G = \frac{B}{A} \qquad \Rightarrow \qquad 1 - H^{-1} = 0 \qquad H^{-1}G = G$$

$$\therefore \quad \hat{y} = \frac{B}{A} u(t)$$

- NOT A FUNCTION OF PAST OUTPUTS.

(ARX)

$$H = \frac{1}{A} \qquad G = \frac{B}{A} \qquad \Rightarrow \qquad 1 - H^{-1} = 1 - A \qquad H^{-1}G = B$$

$$\Rightarrow \quad \hat{y}(t|t-1) = (1-A)\, y(t) + B\, u(t)$$

$$= -a_1 y(t-1) - \dots - a_{n_a} y(t-n_a) + b_1 u(t-1) + \dots + b_{n_b} u(t-n_b)$$

- ARX USES OLD VALUES OF $y(t)$ AS WELL.

- OTHER CASES MORE COMPLICATED.

OPTIMIZATION

- CONSIDER QUADRATIC CASE FIRST.
  - LINEAR PREDICTION FORM OF ARX

WRITE

$$\hat{y}(t) = -a_1 y(t-1) - \dots - a_{n_a} y(t-n_a) + b_1 u(t-1) + \dots + b_{n_b} u(t-n_b)$$

(6-14)

$$= \theta^T \phi(t)$$

$$\theta = \begin{bmatrix} a_1 & a_2 & \dots & a_{n_a} & b_1 & \dots & b_{n_b} \end{bmatrix}^T$$

$$\phi(t) = \begin{bmatrix} -y(t-1) & \dots & -y(t-n_a) & u(t-1) & \dots & u(t-n_b) \end{bmatrix}^T$$

$\Rightarrow$ LINEAR REGRESSION $\phi(t)$ - REGRESSION VECTOR

- ANALYZE THIS CASE FIRST BECAUSE THE PREDICTION ERROR LINEAR IN θ

$$\epsilon(t,\theta) = y(t) - \hat{y}(t,\theta) = y(t) - \theta^T \phi(t)$$

$$\Rightarrow \quad V_N(\theta) = \frac{1}{N} \sum_{t=1}^{N} \epsilon^2(t,\theta)$$

QUADRATIC IN θ.

$\Rightarrow$ SIMPLEST NUMERICAL PROBLEM.

OPTIMIZATION WITH LINEAR REGRESSION

ARX

$$V_N = \frac{1}{N}\sum_{t=1}^{N}\left(y(t) - \theta^T\phi(t)\right)^2$$

$$= \frac{1}{N}\sum_{t=1}^{N}\left(y^2(t) - 2\,\theta^T\phi(t)\,y(t) + \theta^T\phi(t)\,\phi(t)^T\theta\right)$$

$$= \frac{1}{N}\sum_{t=1}^{N} y^2 - 2\theta^T f_N + \theta^T R_N\,\theta$$

- ASSUME THAT R_N INVERTIBLE

$$V_N = \frac{1}{N}\sum_{t=1}^{N} y^2(t) - f_N^T R_N^{-1} f_N + \underbrace{(\theta - R_N^{-1}f_N)^T R_N (\theta - R_N^{-1}f_N)}_{\text{POSITIVE SINCE } R_N \geq 0}$$

$\Rightarrow$ SMALLEST POSSIBLE $V_N(\theta)$ WHEN WE SELECT

$$\theta = \hat{\theta}_N = R_N^{-1} f_N$$

LET $X = \begin{bmatrix}\phi^T(1)\\ \phi^T(2)\\ \vdots\\ \phi^T(N)\end{bmatrix}$ $Y = \begin{bmatrix}y(1)\\ y(2)\\ \vdots\\ y(N)\end{bmatrix}$ $R_N = \frac{1}{N}X^T X$

$f_N = \frac{1}{N}X^T Y$

$\Rightarrow$ MORE COMPACT FORM $V_N = (Y - X\theta)^T(Y - X\theta)\frac{1}{N}$

$$\hat{\theta}_N = (X^T X)^{-1} X^T Y$$

LEAST-SQUARES ESTIMATE.

MORE GENERAL OPTIMIZATIONS

- TO FIND $\hat{\theta}$ TO $\min_{\theta} V_N(\theta)$

NEED TO SOLVE $\frac{\partial}{\partial \theta} V_N(\theta) = 0$

- PREVIOUS CASE

$$\frac{\partial}{\partial \theta} V_N = 2 R_N \theta - 2 f_N = 0$$

$$\Rightarrow \hat{\theta} = R_N^{-1} f_N \quad \therefore \text{CONSISTENT}$$

- MORE GENERAL CASE — MOST OTHER MODEL FORMS

  - $\epsilon(t)$ NOT LINEAR IN θ
  - V_N NOT QUADRATIC IN ϵ

  } V_N NOT QUADRATIC IN θ

$\Rightarrow$ OPTIMIZATION NOT AS SIMPLE.

NO CLOSED FORM SOLUTION AVAILABLE.

$\Rightarrow$ NONLINEAR OPTIMIZATION.

NONLINEAR OPTIMIZATION

- OBJECTIVE IS TO MINIMIZE THE FUNCTION $F(x)$, x IS A VECTOR
  ASSUME $F(x)$ SCALAR

  ARMAX
  OE
  B-J

$$x^* = \arg \min_x F(x)$$

- TYPICALLY END UP USING AN ITERATIVE ALGORITHM. GIVEN $\hat{x}_K$ AND A <u>SEARCH DIRECTION</u> p_K; FIND

$$\hat{x}_{K+1} = \hat{x}_K + \alpha_K p_K$$

  - HOW FIND p_K?
  - HOW FIND α_K? (LINE SEARCH)
  - HOW FIND $\hat{x}_0$? (INITIAL CONDITION)

- SEARCH DIRECTION.
  - TAYLOR EXPANSION OF $F(x)$ ABOUT CURRENT x_K

$$\Rightarrow \quad F_{K+1} = F(\hat{x}_K + \alpha p_K) \simeq F_K + \alpha_K g_K^T p_K$$

WHERE $g_K = \begin{bmatrix} \frac{\partial F}{\partial x_1} \\ \vdots \\ \frac{\partial F}{\partial x_n} \end{bmatrix}_{x_K}$

$$F(\hat{x}_{K+1}) \simeq F(\hat{x}_K) + \frac{\partial F}{\partial x} \underbrace{(\hat{x}_{K+1} - \hat{x}_K)}_{\alpha p_K}$$

STEPS

1) COLLECT DATA (2 SETS)
2) USE FREQ. RESP. ESTIMATE TO GAUGE THE CORRECT MODEL SIZE/COMPLEXITY
3) PICK A MODEL TYPE AND ORDER
4) OPTIMIZE TO GET THE BEST PARAMETERS
5) CHECK THE MODEL.

EXAMPLE

- ACTUAL MODEL CTS → DISCRETE USING ZOH

$$y = Gu + e$$

$$G = \frac{q^{-1}(0.1129) + 0.1038\,q^{-2}}{1 - 1.5622\,q^{-1} + 0.7788\,q^{-2}}$$

$E[e^2(t)] \sim \lambda = 0.1503$ BUT SCALES WITH Y. IN THE CODE

⇒ SYSTEM HAS "OE" STRUCTURE.

- GIVEN ESTIMATE OF $\hat{\theta}$, WE CAN USE THE RESIDUALS $\epsilon^2(t, \hat{\theta})$ TO ESTIMATE λ

$$\hat{\lambda} = \frac{1}{N}\sum_{t=1}^{N} \epsilon^2(t, \hat{\theta})$$

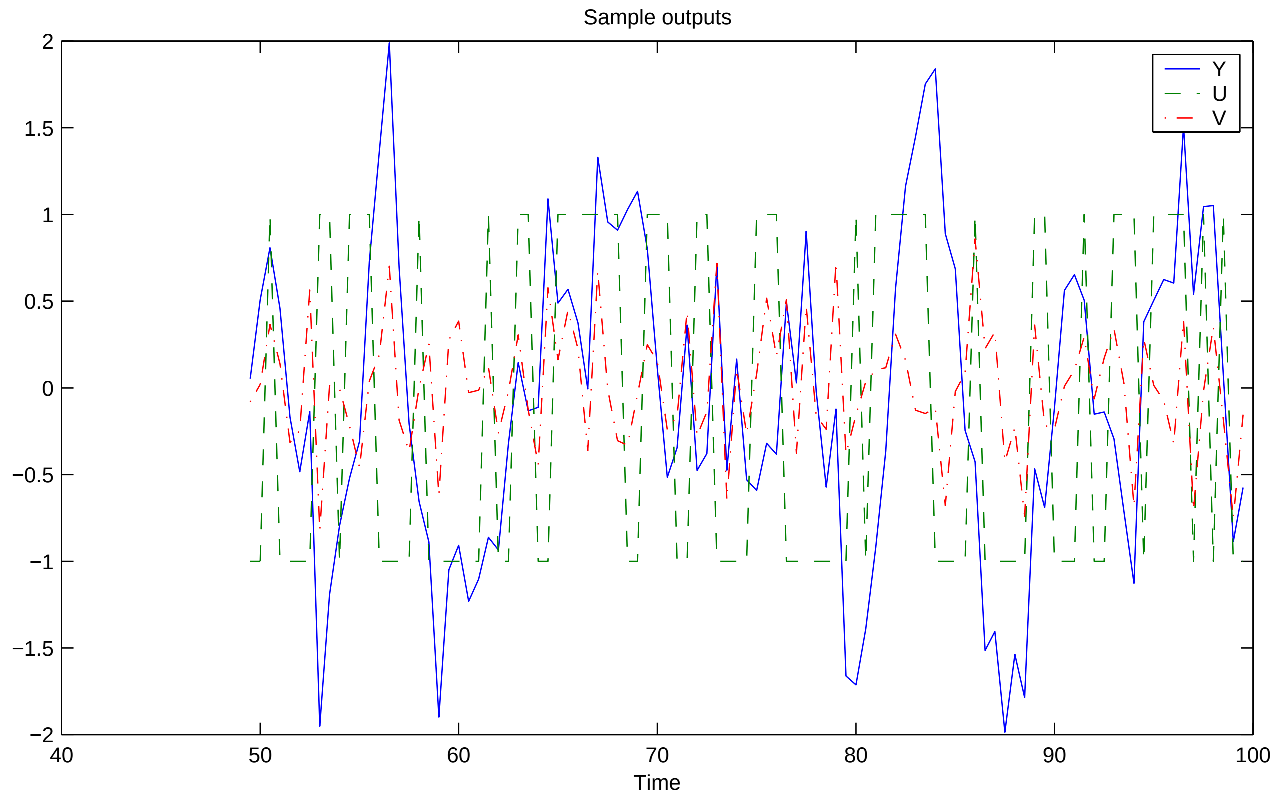
  - NATURAL SINCE $\epsilon(t, \hat{\theta}) = y(t) - \hat{y}(t, \hat{\theta})$ AND THIS DIFFERENCE IS A GOOD ESTIMATE OF $e(t)$.

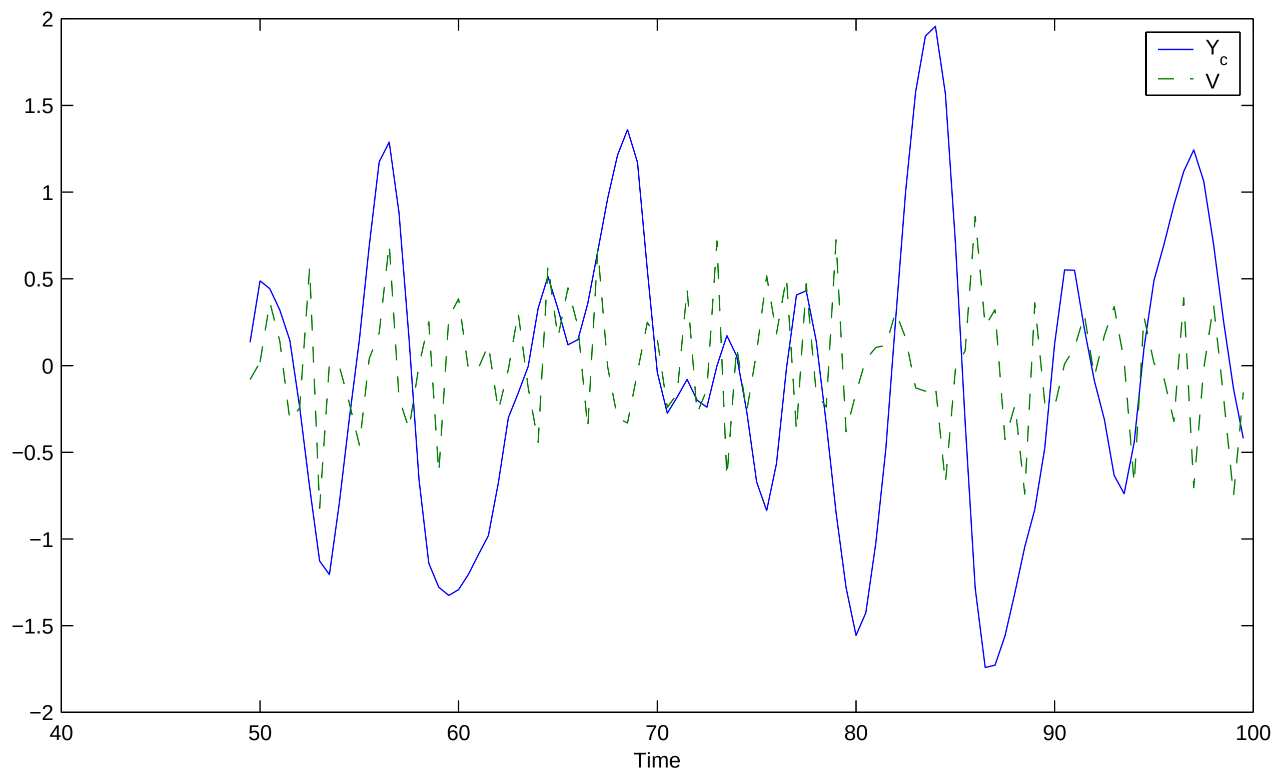
- RESULTS:

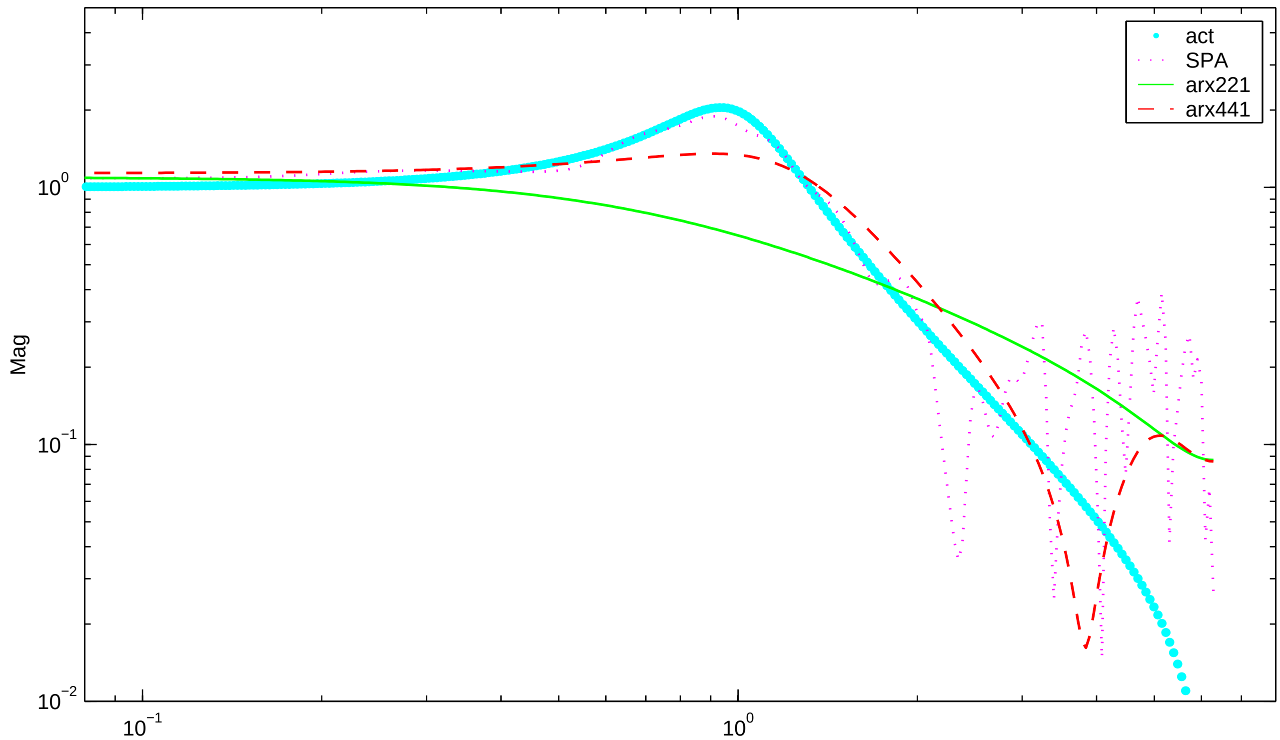
```
%res1 =
%ARX A  0     0.115      0.251
%ARX B  1    -0.616     -0.0497
%ARM A  0     0.117      0.105
%ARM B  1    -1.56       0.773
%BJ  A  0     0.118      0.103
%BJ  B  1    -1.56       0.777
%OE  A  0     0.117      0.104
%OE  B  1    -1.56       0.773
%Act A  0     0.113      0.104
%Act B  1    -1.56       0.779
%res2 =
%lam arx   0.3126
%lam arm   0.1679
%lam BJ    0.1672
%lam OE    0.1683
%lam act   0.1561
```

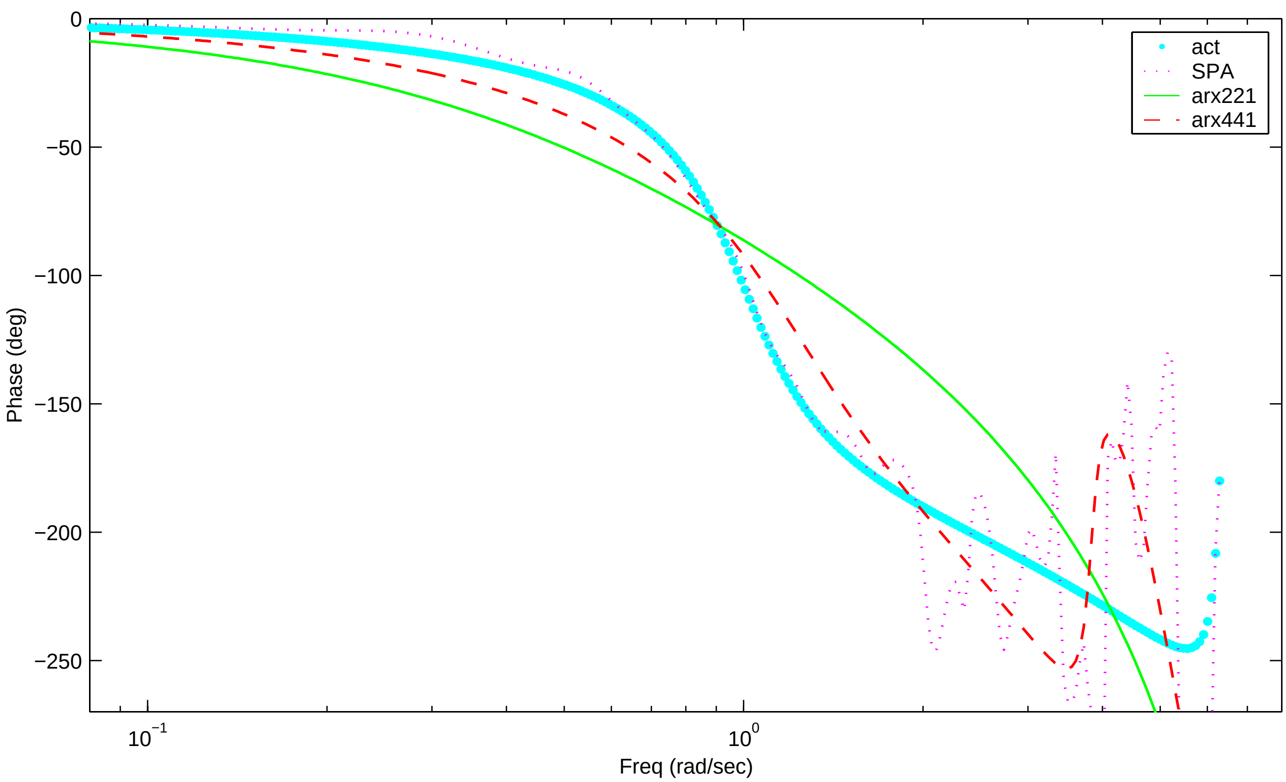
- 2^{ND} ORDER ARX NOT EVEN CLOSE
- OTHER 3 PROVIDE GOOD ESTIMATES OF $\hat{G}(q)$ AND $\hat{\lambda}$

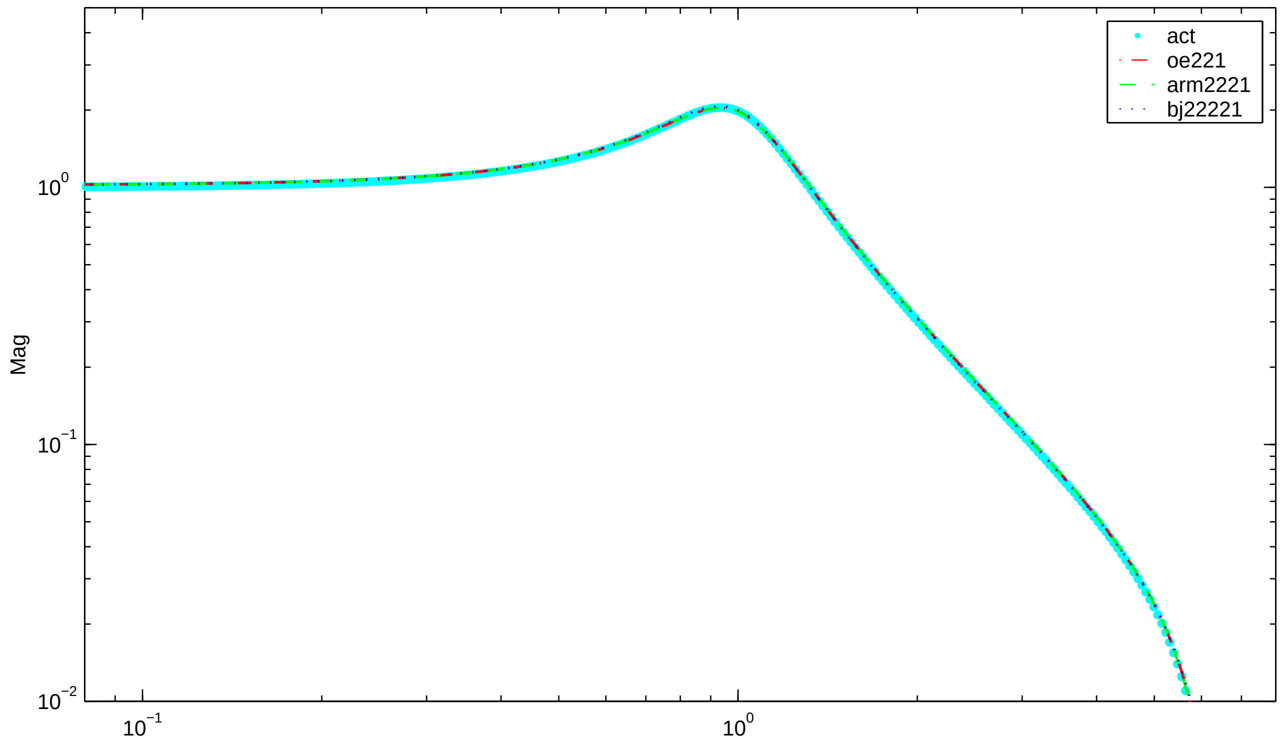
⇒ GOOD TF'S AND IMPULSE RESPONSE

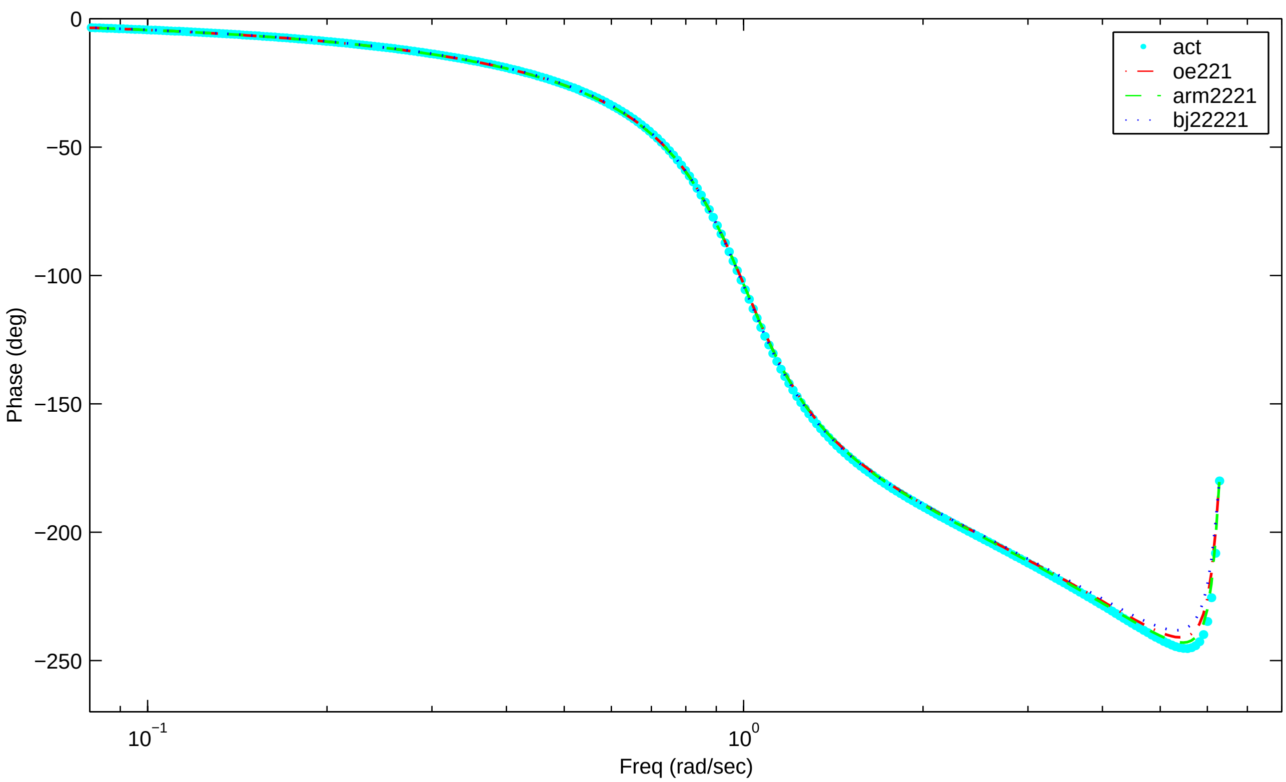

Sample outputs
Y
U
V
Time


Y_c
V
Time


act
SPA
arx221
arx441
Mag
10^0
10^{-1}
10^{-2}
10^{-1}
10^0


act
SPA
arx221
arx441
Phase (deg)
0
−50
−100
−150
−200
−250
10^{-1}
10^0
Freq (rad/sec)


act
oe221
arm2221
bj22221
Mag
10^{0}
10^{-1}
10^{-2}
10^{-1}
10^{0}


act
oe221
arm2221
bj22221
Phase (deg)
0
−50
−100
−150
−200
−250
10^{-1}
10^{0}
Freq (rad/sec)