4. Model Parametrization

- model classification
- general model structure
- time series models
- state-space models
- uniqueness properties

Model Classification

- SISO/MIMO models
- linear/nonlinear models
- parametric/nonparametric models
- time-invariant/time-varying models
- time domain/frequency domain models
- lumped/distributed parameter models
- deterministic/stochastic models

Transfer function and operator

impulse response of a time-invariant discrete-time linear model

$$y(t) = \sum_{k=0}^{\infty} g(k)u(t-k), \quad t = 0, 1, 2, \ldots$$

transfer function from u to y is the z-transform

$$G(z) = \sum_{k=0}^{\infty} g(k)z^{-k}$$

if define a **delay operator**: $Ly(t) = y(t-1)$, the **transfer operator** from u to y is

$$G(L) = \sum_{k=0}^{\infty} g(k)L^{k}$$

note: we replace the argument z^{-1} in the transfer function to L to obtain the transfer operator – we abuse the notations by simply using the same G's

General model structure

$$\mathcal{M}(\theta): \quad y(t) = G(L;\theta)u(t) + H(L;\theta)e(t)$$
$$\mathbf{E}e(t)e(s)^T = \Lambda(\theta)\delta(t,s)$$

- $y(t)$ is ny-dimensional output
- $u(t)$ is nu-dimensional input
- $e(t)$ is an i.i.d. random variable with zero mean (white noise)
- L is the delay (or lag) operator
- H, G, Λ are functions of the parameter vector θ
- this model is a genearal linear model in u and e

Feasible set of parameters

θ takes the values such that

- H^{-1} and $H^{-1}G$ are asymptotically stable
- $G(0;\theta) = 0$ and $H(0;\theta) = I$
- $\Lambda(\theta) \succeq 0$

General SISO model structure

$$A(L)y(t) = \frac{B(L)}{F(L)}u(t) + \frac{C(L)}{D(L)}e(t), \quad \mathbf{E}[e(t)e(t)^T] = \lambda^2$$

where

$$\begin{aligned}
A(q^{-1}) &= 1 + a_1 L + \cdots + a_p L^p \\
B(q^{-1}) &= b_1 L + b_2 L^2 + \cdots + b_n L^n \\
C(q^{-1}) &= 1 + c_1 L + \cdots + c_m L^m \\
D(q^{-1}) &= 1 + d_1 L + \cdots + d_s L^s \\
F(q^{-1}) &= 1 + f_1 L + \cdots + f_r L^r
\end{aligned}$$

note that $B(0) = 0$ (causal system)

Special cases

output error structure

$$y(t) = \frac{B(L)}{F(L)}u(t) + e(t)$$

in this case $H(L;\theta) = 1$

the output error is the difference between the measurable output $y(t)$ and the model output $B(L)/F(L)u(t)$

if $A(L) = 1$ in the general model structure

$$y(t) = \frac{B(L)}{F(L)}u(t) + \frac{C(L)}{D(L)}e(t)$$

- G and H have no common paramater
- possible to estimate G consistently even if choice of H is not appropriate

Model usages

- simulation: simulate the response where e can be randomly generated
- prediction: estimate $y(t)$ given the information up to time $t-1$

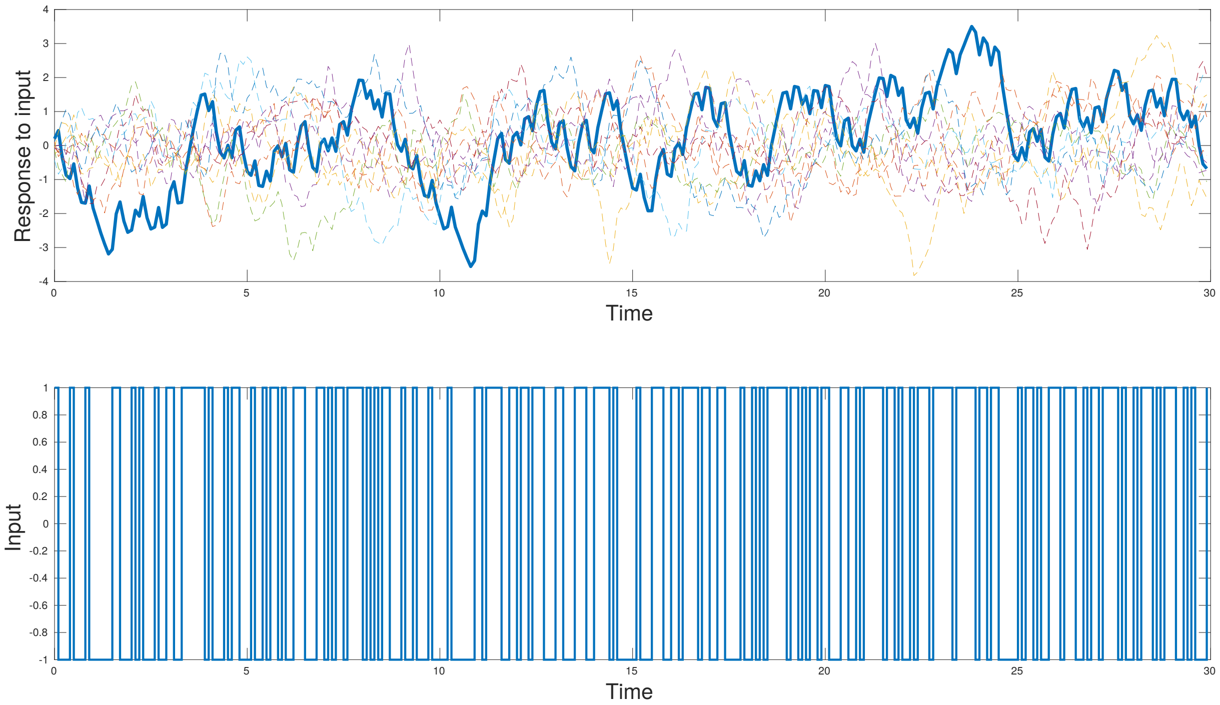
$$\hat{y}(t|t-1) = H^{-1}(L)G(L)u(t) + [1 - H^{-1}(L)]y(t)$$

- inference: use model parameters to explain some statistical properties

System simulation

example: u is binary signal; e is generated with variance 0.1 (10 realizations)

$$y(t) = \left(\frac{0.2 + 0.5L}{1 - 0.7L - 0.18L^2}\right) u(t) + \left(\frac{1 + 0.7L}{1 - 0.7L - 0.18L^2}\right) e(t)$$

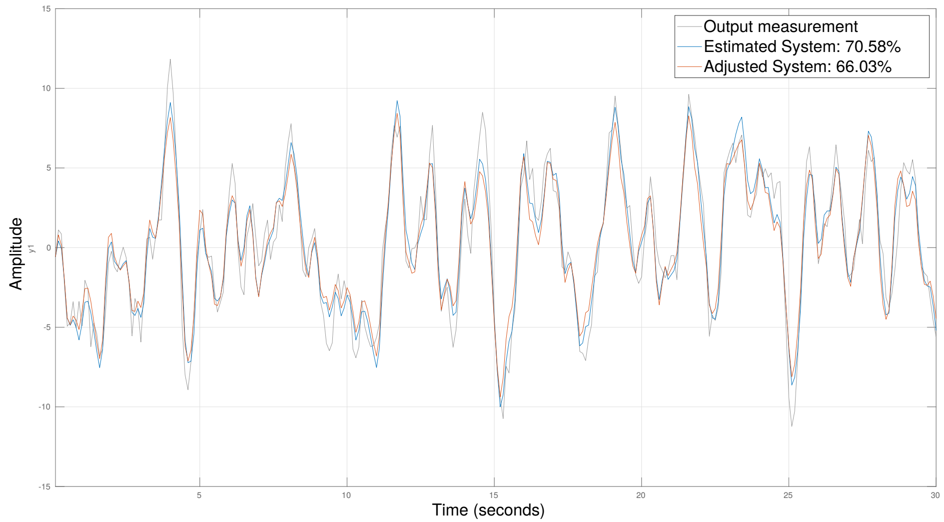


bold line is the response Gu and dashed line is He

One-step prediction

it can be derived that the one-step prediction of y (best in MMSE sense) is

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



- example of 1-step prediction of an estimated ARMAX model
- adjusted some model coefficients can lead to a large change in model dynamics

Model inference

some statistical properties can be drawn from model parameters

- zero entries in AR coefficients explain zero Granger causality

$$y(t) = A_1 y(t-1) + A_2 y(t-2) + \cdots + A_p y(t-p) + e_t$$

- zero entries in the inverse spectrum explain conditional independence

$$\text{example of AR spectrum:} \quad S(\omega) = A(\omega)^{-H} \Sigma A(\omega)^{-1}$$

$$A(\omega) = I - (A_1 e^{-i\omega} + A_2 e^{-2i\omega} + \cdots + A_p e^{-pi\omega})$$

- zero entries in MIMO transfer function suggest zero effect from u

$$H(s) = \begin{bmatrix} \frac{6}{(s^2+5s+6)} & \frac{(s+4)}{(s^2+5s+6)} & 0 \\ 0 & \frac{(s+2)}{(s^2+3s+2)} & 0 \end{bmatrix}$$

State-space models

a linear stochastic model:

$$\begin{aligned} x(t+1) &= A(\theta)x(t) + B(\theta)u(t) + w(t) \\ y(t) &= C(\theta)x(t) + D(\theta)u(t) + v(t) \end{aligned}$$

$w(t)$ is called *process* noise and $v(t)$ is *measurement* noise

- (A, B) defines the system controllability, while (A, C) explains the system observability
- both w and v are often assumed *white noise* sequences with zero means and

$$\mathbf{E} \begin{bmatrix} w(t) \\ v(t) \end{bmatrix} \begin{bmatrix} w(s) \\ v(s) \end{bmatrix}^T = \begin{bmatrix} Q & S \\ S^T & R \end{bmatrix} \delta(t, s)$$

- if A is stable, the processes $x(t)$ and $y(t)$ are wide-sense stationary in steady-state

steady-state sense: the covariance function of x is

$$\lim_{t\to\infty} C(t+k,t) = C(k) = \begin{cases} A^k\Sigma_x, & k \geq 0 \\ \Sigma_x(A^T)^{|k|}, & k \leq 0 \end{cases}$$

where $\Sigma_x = \lim_{t\to\infty} C(t,t)$ and can be obtained via the Lyapunov equation

$$\Sigma_x = A\Sigma_x A^T + BQB^T$$

- when A is stable, there exists a positive solution Σ_x (hence, a valid covariance)
- the decay rate of covariance sequence depends on the eigenvalues of A

Innovation form of state-space model

a standard state-space model can be transformed into the **innovation form**

$$\hat{x}(t+1) = A\hat{x}(t) + Bu(t) + Ke(t), \quad y(t) = C\hat{x}(t) + Du(t) + e(t)$$

- arises from the problem of estimating $x(t)$ conditioning on data up to $t-1$ in MMSE sense
- the estimated x is $\hat{x}$ and estimated output is $\hat{y}(t) = C\hat{x}(t) + Du(t)$
- $e(t) = y(t) - \hat{y}(t)$ is called an **innovation**, explains the residual error after prediction
- the best prediction of x can be represented in the state-space form, proved by Kalman – K is called the **Kalman gain**
- the term $Ke(t)$ is a compensation of approximation error in $y(t)$ to $\hat{x}(t+1)$
- the innovation form has only one error (noise) term in the model

Time series models

stationary models

- ARMAX: AutoRegressive Moving Average model with Exogenous inputs
- ARMA: AutoRegressive Moving Average model
- ARX: AutoRegressive model with Exogenous inputs
- AR: AutoRegressive model
- MA: Moving Average model

non-stationary models

- ARIMA: AutoRegressive Integrated Moving Average model
- ARCH, GARCH (not discussed here)

ARMAX models

an autoregressive moving average model with an exogenous input:

$$A(L)y(t) = B(L)u(t) + C(L)e(t)$$

where e is a white noise with covariance Σ and matrix polynomials are

$$\begin{aligned} A(L) &= I - (A_1 L + A_2 L^2 + \cdots + A_p L^p), \\ B(L) &= B_1 L + B_2 L^2 + \cdots + B_m L^m, \\ C(L) &= I + C_1 L + C_2 L^2 + \cdots + C_q L^q \end{aligned}$$

applying the backward shift (lag) operator explicitly

$$\begin{aligned} y(t) = A_1 y(t-1) + \cdots + A_p y(t-p) + B_1 u(t-1) + \cdots + B_m u(t-m) \\ e(t) + C_1 e(t-1) + \cdots + C_q e(t-q) \end{aligned}$$

the parameter vector is $\theta = (A_1, \ldots, A_p, B_1, \ldots, B_m, C_1, \ldots, C_q, \Sigma)$

Special cases of ARMAX models

model	equation
ARMA	$A(L)y(t) = C(L)e(t)$
AR	$A(L)y(t) = e(t)$
ARX	$A(L)y(t) = B(L)u(t) + e(t)$
MA	$y(t) = C(L)e(t)$
FIR	$y(t) = B(L)u(t) + e(t)$

special cases:

- autoregressive moving average: ARMA(p, q)

$$y(t) = A_1 y(t-1) + \cdots + A_p y(t-p) + e(t) + C_1 e(t-1) + \cdots + C_q e(t-q)$$

- autoregressive: AR(p)

$$y(t) = A_1 y(t-1) + \cdots + A_p y(t-p) + e(t)$$

- moving average: MA(q)

$$y(t) = e(t) + C_1 e(t-1) + \cdots + C_q e(t-q)$$

y consists of a finite sum of stationary white noise (e), so y is also stationary

- finite impulse response: FIR(m)

$$y(t) = B_1 u(t-1) + \cdots + B_m u(t-m) + e(t)$$

- autoregressive with exogenous input: ARX(p, m)

$$y(t) = A_1 y(t-1) + \cdots + A_p y(t-p) + B_1 u(t-1) + \cdots + B_m u(t-m) + e(t)$$

Equivalent representation of AR(1)

write the first-order AR model recursively

$$\begin{aligned}
y(t) &= Ay(t-1) + e(t) \\
&= A(Ay(t-2) + e(t-1)) + e(t) \\
&= A^2 y(t-2) + Ae(t-1) + e(t) \\
&= A^2(Ay(t-3) + e(t-2)) + Ae(t-1) + e(t) \\
&= A^3 y(t-3) + A^2 e(t-2) + Ae(t-1) + e(t) \\
&\vdots \\
&= \sum_{k=0}^{\infty} A^k e(t-k)
\end{aligned}$$

- by assuming that i) t can be extended to negative index and ii) $|\lambda(A)| < 1$
- y can be represented as *infinite moving average*

State-space form of AR models

define the state variable

$$x(t) = (y(t-1), y(t-2), \ldots, y(t-p))$$

the state-space form of AR model is

$$x(t+1) = \underbrace{\begin{bmatrix} A_1 & A_2 & \cdots & A_p \\ I & 0 & & 0 \\ \vdots & \ddots & & \vdots \\ 0 & 0 & I & 0 \end{bmatrix}}_{\mathcal{A}} x(t) + \begin{bmatrix} I \\ 0 \\ \vdots \\ 0 \end{bmatrix} e(t)$$

- the characteristic polynomial of the dynamic matrix is

$$\det(zI - \mathcal{A}) = \det(z^p - (A_1 z^{p-1} + A_2 z^{p-2} + \cdots + A_p))$$

- the AR process is wide-sense stationary if its dynamic matrix $\mathcal{A}$ is stable

Non-uniqueness of MA models

consider examples of two MA models

$$\begin{aligned} y(t) &= e(t) + (1/5)e(t-1), \quad e(t) \sim \mathcal{N}(0,25) \\ x(t) &= v(t) + 5v(t-1), \quad v(t) \sim \mathcal{N}(0,1) \end{aligned}$$

that their output spectrum cannot be distinguished

- note that MA and AR processes are the inverse to each other (by swapping the role of y and e)

$$y(t) = -(1/5)y(t-1) + e(t), \quad x(t) = -5x(t-1) + v(t)$$

- an MA model is called **invertible** if it corresponds to a *causal* infinite AR representation – e.g., process with coefficient $1/5$

Properties of ARMA models

important properties of ARMA model:

$$A(L)y(t) = C(L)e(t)$$

- the process is **stationary** if the roots of the determinant of

$$A(z) = I - (A_1 z + A_2 z^2 + \cdots + A_p z^p)$$

are outside the unit circle

- the process is said to be **causal** if it can be written as

$$y(t) = \sum_{k=0}^{\infty} \Psi(k) e(t-k), \quad \sum_{k=0}^{\infty} |\Psi(k)| \leq \infty$$

(the process cannot depend on the future input)

- the ARMA process is **causal** if and only if the roots of the determinant of $A(z)$ lie outside the unit circle

- the process is **invertible** if the roots of the determinant of

$$C(z) = I + C_1 z + \cdots + C_q z^q$$

lie outside the unit circle

Non-stationary models

examples of non-stationarity and the use of differencing

- random walk: $x(t) = x(t-1) + w(t)$ (covariance depends on t)

$$z(t) \triangleq x(t) - x(t-1) = w(t)$$

  $z(t)$ is white noise which is stationary

- linear static trend: $x(t) = a + bt + w(t)$

$$z(t) \triangleq x(t) - x(t-1) = b + w(t) - w(t-1)$$

  $z(t)$ is a MA process

can we recover the original model from the fitted differenced series ?

Integrated model

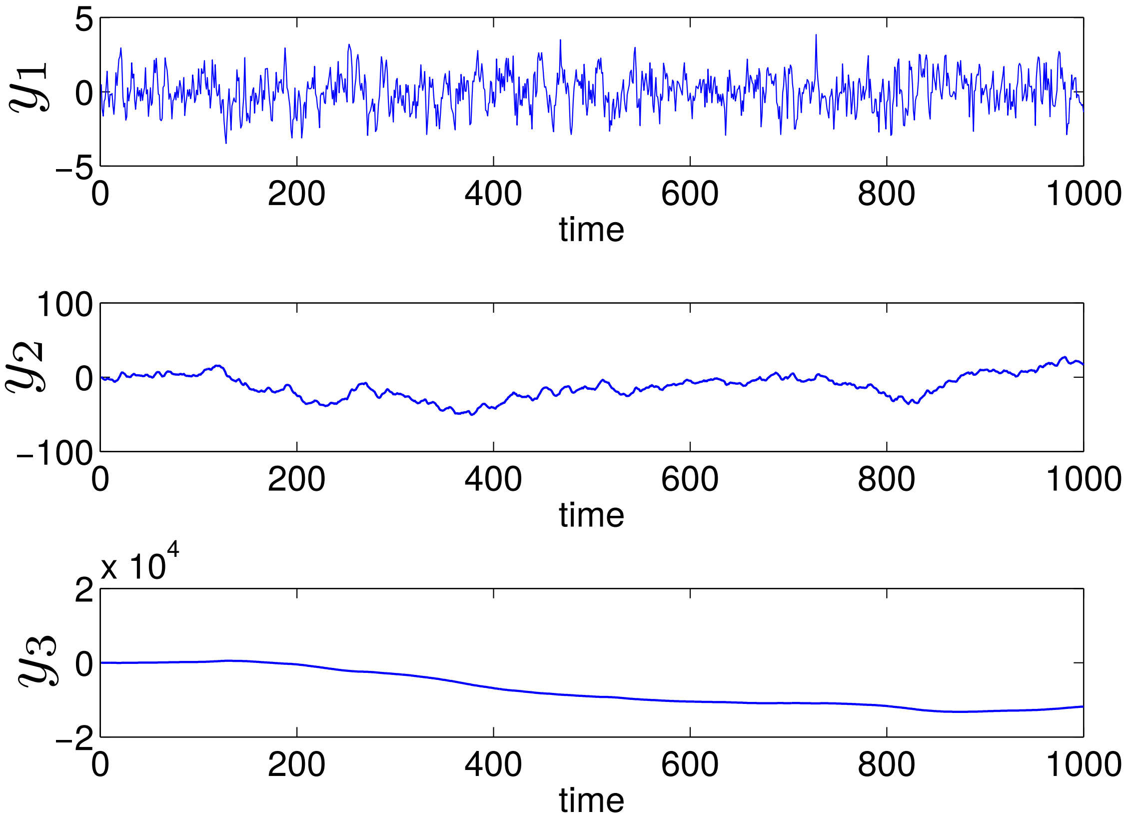
denote L a lag operator; a series $x(t)$ is **integrated** of order d if

$$(I - L)^d x(t)$$

is stationary (after d^{th} differencing)

- we use $I(d)$ to denote the integrated model of order d
- random walk is the first-order integrated model
- the lag of differencing is used to reduce a series with a trend
- for example, 12-lag of differencing removes additive seasonal effect

example: y_1 is a first-order AR process with coefficient 0.4 and is $I(0)$



- $y_2(t) = \sum_{k=0}^{t} y_1(k)$ (cumulative sum of y_1 is $I(1)$ – no exact reverting)
- $y_3(t) = \sum_{k=0}^{t} y_2(k)$ (cumulative sum of y_2 is $I(2)$ – momentum effect)

ARIMA models

$x(t)$ is an ARIMA(p, d, q) process if the dth differences of $x(t)$ is an ARMA(p,q)

$$A(L)(I - L)^d x(t) = C(L)e(t)$$

examples of scalar ARIMA models

- $x(t) = x(t-1) + e(t) + ce(t-1)$ can be arranged as

  $$(1 - L)x(t) = (1 + cL)e(t)$$

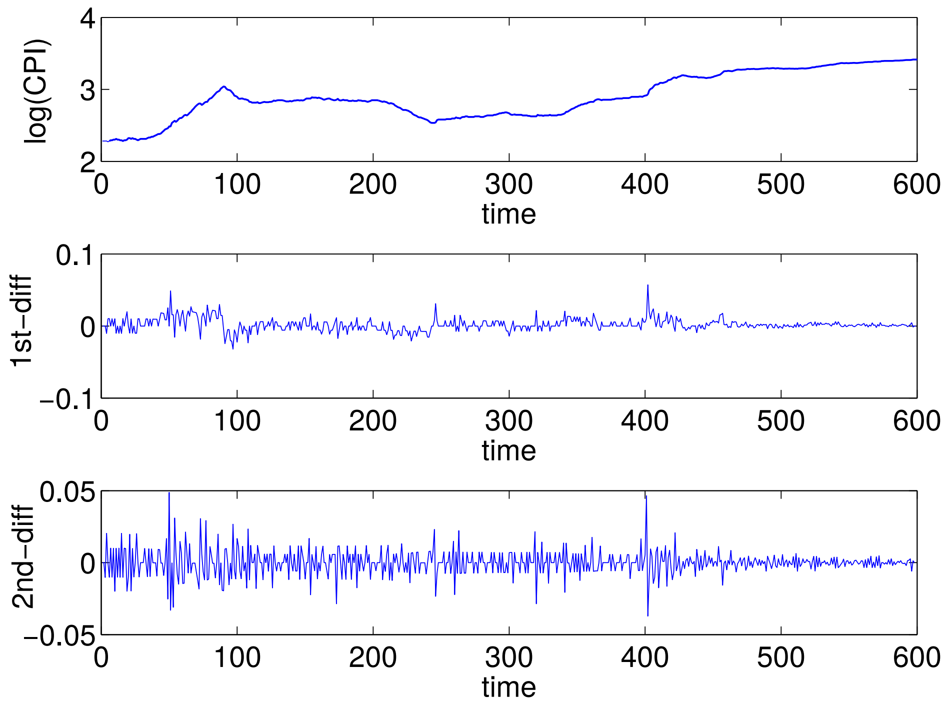
  which is ARIMA(0,1,1) or sometimes called *integrated moving average*

- $x(t) = ax(t-1) + x(t-1) - ax(t-2) + w(t)$ can be arranged as

  $$(1 - aL)(1 - L)x(t) = w(t)$$

  which is ARIMA(1,1,0)

example: $\log$ of CPI - consumer production index and its first, second differences



- $\log$ CPI shows the momentum type – characteristics of $I(2)$
- the first difference has no momentum but no mean-reverting
- the second difference seems to be mean-reverting and behaves like white noise

Uniqueness properties

question: can we describe a system *adequately* and *uniquely* ?

define $\mathcal{D}$ the set of θ for which

$(\hat{G}, \hat{H}, \hat{\Lambda})$ gives a *perfect description* of the true system

three possibilities of this set can occur:

- the set $\mathcal{D}$ is empty or underparametrization
- the set $\mathcal{D}$ contains one point
- the set $\mathcal{D}$ consists of several points or overparametrization

Non-uniqueness of general state-space models

consider the multivariable model

$$\begin{aligned} x(t+1) &= A(\theta)x(t) + B(\theta)u(t) + w(t) \\ y(t) &= C(\theta)x(t) + v(t) \end{aligned}$$

$w(t)$ and $v(t)$ are independent zero-mean white noise with covariance R_1, R_2

also consider a second model

$$\begin{aligned} z(t+1) &= \bar{A}(\theta)z(t) + \bar{B}(\theta)u(t) + \bar{w}(t) \\ y(t) &= \bar{C}(\theta)z(t) + v(t) \end{aligned}$$

where $\mathbf{E}[\bar{w}(t)\bar{w}(s)^T] = \bar{R}_1\delta(t,s)$ and

$$\bar{A} = QAQ^{-1}, \quad \bar{B} = QB, \quad \bar{C} = CQ^{-1}, \quad \bar{R}_1 = QR_1Q^T$$

for some nonsingular matrix Q $\bar{A}$ is simply a similarity transform of A

the two models are equivalent:

- they have the same transfer function from u to y

$$G(z) = \bar{C}(zI - \bar{A})^{-1}\bar{B} = CQ^{-1}(zI - QAQ^{-1})^{-1}QB = C(zI - A)^{-1}B$$

- the outputs y from the two models have the same second-order properties, *i.e.*, the spectral densities are the same

$$\begin{aligned}
S_y(\omega) &= \bar{C}(e^{\mathrm{i}\omega} - \bar{A})^{-1}\bar{R}_1(e^{\mathrm{i}\omega} - \bar{A})^{-*}\bar{C}^* + R_2 \\
&= CQ^{-1}(e^{\mathrm{i}\omega} - \bar{A})^{-1}QR_1Q^*(e^{\mathrm{i}\omega} - \bar{A})^{-*}Q^{-*}C^* + R_2 \\
&= C[Q^{-1}(e^{\mathrm{i}\omega} - \bar{A})Q]^{-1}R_1[Q^*(e^{\mathrm{i}\omega} - \bar{A})^*Q^{-*}]^{-1}C^* + R_2 \\
&= C(e^{\mathrm{i}\omega} - A)^{-1}R_1(e^{\mathrm{i}\omega} - A)^{-*}C^* + R_2
\end{aligned}$$

the model is not unique since Q can be chosen arbitrarily

Choosing a class of model structures

important factors:

- **flexibility:** the model structure should describe most of the different system dynamics expected in the application
- **parsimony:** the model should contain the smallest number of free parameters required to explain the data adequately
- **algorithm complexity:** the form of model structure can considerably influence the computational cost
- **properties of the criterion function:** for example, the asymptotic properties of prediction-error method depends crucially on the criterion function and the model structure

References

Chapter 6 in
T. Söderström and P. Stoica, *System Identification*, Prentice Hall, 1989

P.S.P. Cowpertwait and A.V. Metcalfe, *Introductory Time Series with R*, Springer, 2009

R.H. Shumway and D.S. Stoffer, *Time Series Analysis and Its Applications: with R Examples*, 3rd edition, Springer, 2011

Chapter 12-13 in
D. Ruppert and D.S. Matteson, *Statistics and Data Analysis for Financial Engineering*, 2nd edition, Springer, 2015