SYSTEM ESTIMATION METHODS II: STRUCTURED ESTIMATION

Mathematical Models can be derived from

- Physical Modeling (Analytic approach)
- Identification (Experimental approach)

CLASSIFICATION OF MODELS

- SISO v/s MIMO
- Linear v/s Nonlinear
- Parametric v/s Nonparametric
- Time invariant v/s Time variant
- Time domain v/s Frequency domain
- Discrete time v/s Continuous time
- Deterministic v/s Stochastic

GENERAL LTI-SISO MODEL STRUCTURE (BOX-JENKINS)

$$y_t = G(q;\theta)u_t + H(q;\theta)w_t$$

$$G(q;\theta) = \frac{B(q)}{A(q)} = \frac{b_1 q^{-nk} + \cdots + b_{nb} q^{-nk-nb+1}}{1 + a_1 q^{-1} + \cdots + a_{na} q^{-na}}$$

$$H(q;\theta) = \frac{C(q)}{D(q)} = \frac{1 + c_1 q^{-1} + \cdots + c_{nc} q^{-nc}}{1 + d_1 q^{-1} + \cdots + a_{nd} q^{-nd}}$$

where $\{w_t\}$ is white noise of variance λ^2, $\{u_t\}$ is the input, and

$$\theta = [a_1 \quad \cdots \quad a_{na} \quad b_1 \quad \cdots \quad b_{nb} \quad c_1 \quad \cdots \quad c_{nc} \quad d_1 \quad \cdots \quad d_{nd}]^T$$

- Time delay $n_k \geq 1 \quad \Rightarrow \quad G(\infty;\theta) = 0 \qquad$ (also, $H(\infty;\theta) = 1$)
- $H^{-1}(q;\theta)$ and $H^{-1}(q;\theta)G(q;\theta)$ are asymptotically stable

Often, $H(q;\theta)$ is also required to be asymptotically stable

Time Series Models (without controllable input)

AR	$A(q)y_t = w_t$
MA	$y_t = C(q)w_t$
ARMA	$A(q)y_t = C(q)w_t$

Basic Models for Control

ARX	$A(q)y_t = B(q)u_t + w_t$
ARMAX	$A(q)y_t = B(q)u_t + C(q)w_t$
FIR	$y_t = B(q)u_t + w_t$
OE	$y_t = \dfrac{B(q)}{A(q)}u_t + w_t$

- Selection of model structure depends on prior knowledge and estimation method
- Preferable: model structure contains true system (*no undermodeling*)
- Preferable: different models predict different results (*identifiability*)

STATE SPACE MODELS

$$x_{t+1} = A(\theta)x_t + B(\theta)u_t + w_t$$
$$y_t = C(\theta)x_t + D(\theta)u_t + v_t$$

Advantages:

- Physical Insight (physical parameters)
- Natural extension to MIMO systems

Problem: Over-parameterization

$$\bar{x}_{t+1} = TA(\theta)T^{-1}\bar{x}_t + TB(\theta)u_t + Tw_t$$
$$y_t = C(\theta)T^{-1}\bar{x}_t + D(\theta)u_t + v_t$$

gives the same input-output relation!

PREDICTION ERROR FORMULATION

$$y_t = f(Y_{t-1}, U_t, t; \theta) + \varepsilon_t$$

where $Y_{t-1} := \{y_{t-1}, \ldots, y_1\}$, $U_t := \{u_t, \ldots, u_1\}$ and $\{\varepsilon_t\}$ is an *innovations sequence*, i.e.

$$E\{\varepsilon_t \mid Y_{t-1}, U_t\} = 0$$

Important Case: LTI models with wide-sense stationary disturbances

$$\begin{aligned} y_t &= G(q;\theta)u_t + H(q;\theta)w_t \\ &= \underbrace{[I - H^{-1}(q;\theta)]y_t + H^{-1}(q;\theta)G(q;\theta)u_t}_{\hat{y}_{t|t-1}(\theta) := E\{y_t | Y_{t-1}, U_{t-1}\}} + \underbrace{w_t}_{\varepsilon_t} \end{aligned}$$

ML ESTIMATION FOR THE PREDICTION ERROR FORMULATION

If $\{\varepsilon_t\}$ are IID Gaussian with (known) covariance Σ, the ML estimator is

$$\boxed{\hat{\theta}_{ML} = \arg\max_{\theta\in\Theta} l(\theta) = \arg\min_{\theta\in\Theta} \sum_{t=1}^{N} \varepsilon_t^T(\theta)\Sigma^{-1}\varepsilon_t(\theta)}$$

Interpretation:

$\hat{\theta}_{ML}$ *tries to make the prediction errors "small"*

$\Rightarrow$ *Prediction Error Methods* (PEM):

$$\hat{\theta}_{PEM} = \arg\min_{\theta\in\Theta} f(\{\varepsilon_t(\theta)\};\theta)$$

Typical cost functions: $\frac{1}{2N}\sum_{t=1}^{N}\varepsilon_t^2(\theta)$ $(SISO)$ or $\det\left[\frac{1}{N}\sum_{t=1}^{N}\varepsilon_t(\theta)\varepsilon_t^T(\theta)\right]$ $(MIMO)$

Linear Regression

For some model structures (e.g. ARX, FIR), the predictor is linear in θ:

$$\hat{y}_{t|t-1} = \varphi_t^T \theta$$

Then, the PEM cost function is a *least-squares (LS) criterion*:

$$V_N(\theta) := \frac{1}{2N} \sum_{t=1}^{N} [y_t - \varphi_t^T \theta]^2$$

The minimizer of $V_N(\theta)$ has an explicit expression:

$$\hat{\theta}_{LS} := \arg\min_{\theta} V_N(\theta) = \left[\frac{1}{N} \sum_{t=1}^{N} \varphi_t \varphi_t^T \right]^{-1} \frac{1}{N} \sum_{t=1}^{N} \varphi_t y_t$$

Linear Regression (cont.)

Properties:

If $y_t = \varphi_t^T \theta_0 + w_t$, and the signals are quasi-stationary, then

$$\hat{\theta}_{LS} - \theta_0 = \left[\frac{1}{N} \sum_{t=1}^{N} \varphi_t \varphi_t^T \right]^{-1} \frac{1}{N} \sum_{t=1}^{N} \varphi_t w_t \xrightarrow[N\to\infty]{a.s.} [R^*]^{-1} f^*$$

where

$$R^* = \bar{E}\{\varphi_t \varphi_t^T\}, \qquad f^* = \bar{E}\{\varphi_t w_t\}$$

Therefore, LS is consistent if:

- R^* is nonsingular (*persistence of excitation* and *identifiability*)
- $f^* = 0$, i.e., $\{w_t\}$ is uncorrelated with $\{\varphi_t\}$ (e.g. if $\{w_t\}$ is white noise, or for FIR)

LINEAR REGRESSION (CONT.)

Example: ARX

$$A(q)y_t = B(q)u_t + w_t \qquad \Rightarrow \qquad \hat{y}_{t|t-1} = [1 - A(q)]y_t + B(q)u_t$$

Then, $\qquad \varphi_t = [-y_{t-1} \quad \cdots \quad -y_{t-na} \quad u_{t-1} \quad \cdots \quad u_{t-nb}]^T$

and $\qquad \theta = [a_1 \quad \cdots \quad a_{na} \quad b_1 \quad \cdots \quad b_{nb}]^T$

What happens if the noise is colored, e.g. $A(q)y_t = B(q)u_t + \dfrac{1}{D(q)} w_t$?

One possibility: High-order ARX model

$$A(q)D(q)y_t = B(q)D(q)u_t + w_t$$

THE CORRELATION APPROACH (INSTRUMENTAL VARIABLES)

In LS: $$\frac{1}{N}\sum_{t=1}^{N}\varphi_t[y_t - \varphi_t^T\hat{\theta}_{LS}] = 0$$

There is a problem when $\{\varphi_t\}$ is correlated with $\{w_t\}$

Idea: Instrumental Variables

$$\frac{1}{N}\sum_{t=1}^{N}\zeta_t[y_t - \varphi_t^T\hat{\theta}_{IV}] = 0 \qquad \Rightarrow \qquad \hat{\theta}_{IV} = \left[\frac{1}{N}\sum_{t=1}^{N}\zeta_t\varphi_t^T\right]^{-1}\frac{1}{N}\sum_{t=1}^{N}\zeta_t y_t$$

where for consistency:

- $\{\zeta_t\}$ highly correlated with $\{u_t\}$ ($\bar{E}\{\zeta_t\varphi_t^T\}$ nonsingular)
- $\{\zeta_t\}$ uncorrelated with $\{w_t\}$ ($\bar{E}\{\zeta_t w_t\} = 0$)

Good choices: *past inputs*, *noiseless outputs* (from LS estimation)