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

$$\theta = [\begin{array}{cccccccccccc} a_1 & \cdots & a_{na} & b_1 & \cdots & b_{nb} & c_1 & \cdots & c_{nc} & d_1 & \cdots & d_{nd} \end{array}]^T$$

- 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	$y_t = \frac{B(q)}{A(q)}u_t + w_t$

- 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

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

DETAILS: OPTIMAL ONE-STEP AHEAD PREDICTOR FOR LTI SYSTEMS

$$y_t = G(q)u_t + H(q)w_t = G(q)u_t + [H(q) - I]w_t + w_t$$

$G(q)u_t$ depends on u_{t-1}, ..., so it can be computed at time $t-1$

Notice that the elements of $H(q)-I$ have relative degree >0, so $[H(q)-I]w_t$ depends on past values of w_t. This term can be computed by noting that

$$w_t = H^{-1}(q)[y_t - G(q)u_t]$$

Note that w_t depends on $y_t, y_{t-1}, \ldots$ and $u_{t-1}, u_{t-2}, \ldots$, so it *cannot* be computed at time $t-1$. But $w_{t-1}, \ldots$ *can* be computed at time $t-1$

Therefore:

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

INNOVATIONS FORM

The system

$$y_t = G(q)u_t + H(q)w_t = [G(q) \quad H(q) - I]\begin{bmatrix} u_t \\ w_t \end{bmatrix} + w_t$$

where $H(q)$ is stable and minimum phase, can be written in *innovations form*

$$\begin{aligned} x_{t+1} &= Ax_t + Bu_t + Kw_t \\ y_t &= Cx_t + Du_t + w_t \end{aligned}$$

K : *Kalman gain*

ML ESTIMATION FOR THE PREDICTION ERROR FORMULATION

If $\{\varepsilon_t\}$ are i.i.d. 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: $\dfrac{1}{2N}\sum_{t=1}^{N}\varepsilon_t^2(\theta)$ $(SISO)$ or $\det\left[\dfrac{1}{N}\sum_{t=1}^{N}\varepsilon_t(\theta)\varepsilon_t^T(\theta)\right]$ $(MIMO)$

DETAILS: ML ESTIMATION OF LTI MODELS

To compute the MLE of θ given $U_N := \{u_N, \ldots, u_1\}$ and $Y_N := \{y_N, \ldots, y_1\}$, we need $P\{Y_N \mid U_N, \theta\}$. By Bayes' Theorem,

$$P\{Y_N \mid U_N, \theta\} = \prod_{t=1}^{N} P\{y_t \mid Y_{t-1}, U_t; \theta\}$$

Now, by the rule for transforming random variables:

$$P\{y_t \mid Y_{t-1}, U_t; \theta\} = P_\varepsilon\{\varepsilon_t \mid Y_{t-1}, U_t; \theta\}\Big|_{\varepsilon_t = y_t - f(Y_{t-1}, U_t, t; \theta)} \cdot \left|\det\left(\frac{\partial \varepsilon_t}{\partial y_t}\right)\right| = P_\varepsilon\{\varepsilon_t(\theta); \theta\}$$

where $\varepsilon_t(\theta) := y_t - f(Y_{t-1}, U_t, t; \theta)$, and $P_\varepsilon\{\,\cdot\,; \theta\}$ is the conditional density for ε_t given Y_{t-1} and U_t, and may depend on θ. Thus,

$$P\{Y_N \mid U_N, \theta\} = \prod_{t=1}^{N} P_\varepsilon\{\varepsilon_t(\theta); \theta\}$$

DETAILS: ML ESTIMATION OF LTI MODELS (CONT.)

When $\{\varepsilon_t\}$ are i.i.d. Gaussian with covariance Σ,

$$P\{Y_N \mid U_N, \theta\} = \prod_{t=1}^{N} \left([(2\pi)^m \det \Sigma]^{-1/2} \exp\left\{ -\frac{1}{2} \varepsilon_t^T(\theta) \Sigma^{-1} \varepsilon_t(\theta) \right\} \right)$$

$$= [(2\pi)^m \det \Sigma]^{-N/2} \exp\left\{ -\frac{1}{2} \sum_{t=1}^{N} \varepsilon_t^T(\theta) \Sigma^{-1} \varepsilon_t(\theta) \right\}$$

Hence the log likelihood function is

$$l(\theta) = \ln P\{Y_N \mid U_N, \theta\} = -\frac{Nm}{2} \ln 2\pi - \frac{N}{2} \ln \det \Sigma - \frac{1}{2} \sum_{t=1}^{N} \varepsilon_t^T(\theta) \Sigma^{-1} \varepsilon_t(\theta)$$

Therefore, the MLE is

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

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

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

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 $\{w_t\}$ ($\bar{E}\{\zeta_t w_{t-\tau}\}=0$ for all τ)

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

FREQUENCY DOMAIN IDENTIFICATION

The ML estimator of θ can be computed from frequency domain data:

$$U_k = \frac{1}{\sqrt{N}} \sum_{t=0}^{N-1} u_t e^{-i\omega_k t}, \quad Y_k = \frac{1}{\sqrt{N}} \sum_{t=0}^{N-1} y_t e^{-i\omega_k t}, \quad \omega_k = \frac{2\pi}{N} k, \quad k = 0, \ldots, N-1$$

Due to the invariance principle, the ML estimate should be exactly the same if we compute it in the time or frequency domain... however, for some time researchers claimed that these estimates differ!

Reason for discrepancy: treatment of initial (and final) conditions!

Time domain:

1. $x_0 = 0$
2. x_0 deterministic (parameter)
3. $x_0 \sim N(\mu_{x_0}, \Sigma_0)$

Frequency domain:

1. $\alpha := x_0 - x_N = 0$
2. α deterministic (parameter)
3. α hidden random variable

Frequency Domain Identification (cont.)

Time domain ML: (state-space, innovations form)

$$y_t = F(q)\delta_t x_0 + G(q)u_t + H(q)w_t = CA^t x_0 + \left[\sum_{l=0}^{t-1} CA^{t-l-1} Bu_l\right] + Du_t + \left[\sum_{l=0}^{t-1} CA^{t-l-1} Kw_l\right] + w_l$$

For $x_0 \sim N(\mu_{x_0}, \Sigma_0)$: $l(\theta, \mu_{x_0}, \Sigma_0) = -\frac{1}{2}[N\ln(2\pi) + \ln\det\Sigma_{Y_N} + (Y_N - \mu_{Y_N})^T \Sigma_{Y_N}^{-1}(Y_N - \mu_{Y_N})]$

where: $\mu_{Y_N} = \Lambda U_N + \Gamma\mu_{x_0}, \quad \Sigma_{Y_N} = \Gamma\Sigma_0\Gamma^T + \sigma_w^2 \Omega\Omega^T$

$$\Gamma = \begin{bmatrix} C \\ CA \\ \vdots \\ CA^{N-1} \end{bmatrix}, \Omega = \begin{bmatrix} I & & & 0 \\ CK & I & & \\ \vdots & & \ddots & \\ CA^{N-2}K & CA^{N-3}K & \cdots & I \end{bmatrix}, \Lambda = \begin{bmatrix} D & & & 0 \\ CB & D & & \\ \vdots & & \ddots & \\ CA^{N-2}B & CA^{N-3}B & \cdots & D \end{bmatrix}$$

The other 2 assumptions on x_0 follow easily from this expression (*exercise*!)

Frequency domain ML:

Difficulty: U_k, Y_k are complex for $k \neq 0, N/2$!

$$Z \sim N(0,P) \quad \text{means} \quad p(Z) = \begin{cases} \dfrac{\exp\{-(1/2) Z^T P^{-1} Z\}}{\sqrt{(2\pi)^N \det P}}, & Z \text{ real} \\ \dfrac{\exp\{-Z^T P^{-1} Z\}}{\pi^N \det P}, & Z \text{ complex} \end{cases}$$

Taking the DFT of the state space equations gives

$$e^{i\omega_k} X_k + \frac{e^{i\omega_k}}{\sqrt{N}} (x_N - x_0) = A X_k + B U_k + K W_k$$
$$Y_k = C X_k + D U_k + W_k$$

Frequency Domain Identification (cont.)

Equivalence:

Due to the invariance principle, and that the DFT is a linear transformation, it can be shown that the time- and frequency-domain ML are identical. In particular, if

$$\begin{bmatrix} \alpha \\ W_N \end{bmatrix} \sim N\left(\begin{bmatrix} \mu_\alpha \\ 0 \end{bmatrix}, \begin{bmatrix} \Sigma_\alpha & \Sigma_{\alpha W_N} \\ \Sigma_{\alpha W_N} & \sigma_w^2 I \end{bmatrix} \right)$$

then for different choices of μ_α, Σ_α and $\Sigma_{\alpha W_N}$ the ML estimator gives the same result as for the 3 time-domain and 2 remaining frequency domain assumptions

For details, see (Agüero, Yuz, Goodwin & Delgado, 2010).