Time-Varying and Nonlinear Models

"Happiness is finding something that is linear!"

"If it doesn't fit: Use a bigger hammer!"

Rough Outline

today's lecture

1. Time-varying linear systems
   - non-uniform sampling intervals
   - linearization around a trajectory
2. Linear Models + nonlinear blocks
   - input nonlinearities (Hammerstein models)
   - output nonlinearities (Wiener models)
   - nonlinear regressors (GMDH)
     (GMDH = Group Method of Data Handling)
   - physically-based models
3. Parametrized Model Families
   - basis functions and function expansions
   - Volterra, Fourier, Walsh, Wavelet
   - orthogonality
4. Network Models
   - neural networks
5. Fuzzy Sets and Rule-Based Models
   - parameterization
6. Summary

Time-Varying Models

Impulse response depends upon time at which impulse was applied.

① 

2 parameters : $\tilde{g}(t,s)$

t - response at time t

s - to an impulse applied at time s

$$y(t) = \sum_{s=-\infty}^{t-1} \tilde{g}(t,s)\,u(s) + v(t)$$

Alternate notation : $g_t(k)$

$$y(t) = \sum_{k=1}^{\infty} g_t(k)\,u(t-k) + v(t)$$

$$g_t(k) = \tilde{g}(t, t-k)$$

In a state-space framework,

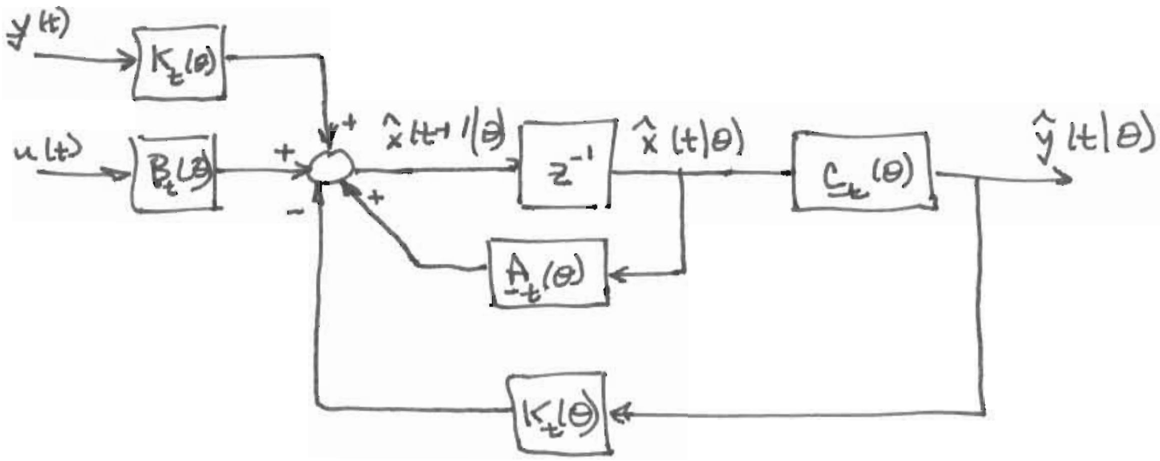
$$\underline{x}(t+1,\theta) = \underline{A}_t(\theta)\,\underline{x}(t,\theta) + \underline{B}_t(\theta)\,\underline{u}(t) + \underline{K}_t(\theta)\,e(t)$$

$$y(t) = \underline{C}_t(\theta)\,\underline{x}(t,\theta) + \underline{e}(t) \qquad (5.4)$$

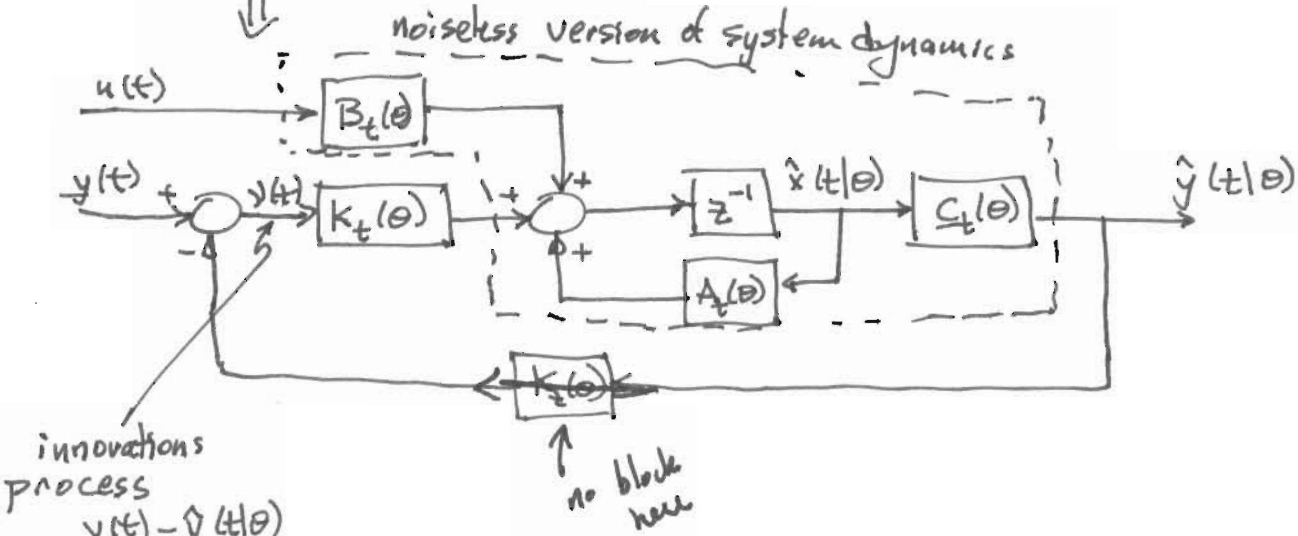
Given the parameter θ, the predictor is

$$\hat{\underline{x}}(t+1|\theta) = \left[\underline{A}_t(\theta) - \underline{K}_t(\theta)\,\underline{C}_t(\theta)\right]\hat{\underline{x}}(t,\theta) + \underline{B}_t(\theta)\,\underline{u}(t) + \underline{K}_t(\theta)\,\underline{y}(t) \qquad (5.5)$$

$$\hat{\underline{y}}(t|\theta) = \underline{C}_t(\theta)\,\hat{\underline{x}}(t,\theta)$$



$\Updownarrow$



Estimation problem: Given $\underline{\theta}$, find $\underline{K}_t(\theta)$ that

(a) minimizes statistic of error in state estimate, $\hat{\underline{x}}(t|\theta) - \underline{x}(t)$.

or

(b) makes the innovations process, the error between the actual (observed) and predicted output white and small.

Identification problem: Find both $\underline{\theta}$ and $\underline{k}_t(\theta)$ to satisfy (b).

Recursive ID problem: Recursively update $\underline{\theta}(t)$ and $\underline{K}_t(\theta)$ given $\underline{\theta}(t-1)$ and $y(t)$.

Note that from (5.5),

$$\hat{\underline{x}}(t+1|\theta) = [\underline{A}_t(\theta) - \underline{K}_t(\theta)\underline{C}_t(\theta)]\hat{\underline{x}}(t|\theta) + \underline{B}_t(\theta)\underline{u}(t) + \underline{K}_t(\theta)\underline{y}(t)$$

$$= [\underline{A}_t(\theta) - \underline{K}_t(\theta)\underline{C}_t(\theta)][\underline{A}_{t-1}(\theta) - \underline{K}_{t-1}(\theta)\underline{C}_{t-1}(\theta)]\hat{\underline{x}}(t-1|\theta)$$

$$+ [\underline{A}_t(\theta) - \underline{K}_t(\theta)\underline{C}_t(\theta)][\underline{B}_{t-1}(\theta)\underline{u}(t-1) + \underline{K}_{t-1}(\theta)\underline{y}(t-1)]$$

$$+ \underline{B}_t(\theta)\underline{u}(t) + \underline{K}_t(\theta)\underline{y}(t)$$

; this process can be repeated to obtain:

$$\hat{y}(t|\theta) = \sum_{k=1}^{\infty} w_t^u(k,\theta)\, u(t-k) + \sum_{k=1}^{\infty} w_t^y(k,\theta)\, y(t-k)$$

where

$$\underline{w}_t^u(k,\theta) = \underline{C}_t(\theta) \prod_{j=t-k}^{t-1} [\underline{A}_j(\theta) - \underline{K}_j(\theta)\underline{C}_j(\theta)]\, \underline{B}_{t-k}(\theta)$$

$$\underline{w}_t^y(k,\theta) = \underline{C}_t(\theta) \prod_{j=t-k}^{t-1} [A_j(\theta) - \underline{K}_j(\theta)\underline{C}_j(\theta)]\, \underline{K}_{t-k}(\theta)$$

which provides the discrete-time impulse response matrix, or weighting pattern, for ~~each~~ input to the output and the dependence upon past outputs.

Why do we end up with time-varying dynamics?

2 primary reasons:

(1) linearization, and

(2) non-uniform sampling of a continuous time process.

<u>Linearization</u>:

Assume nonlinear model

$$\underline{x}(t+1) = \underline{f}(\underline{x}(t),\underline{u}(t)) + \Gamma(\underline{x}(t),\underline{u}(t))\underline{w}(t)$$

$$\underline{y}(t) = \underline{h}(\underline{x}(t)) + \underline{m}(\underline{x}(t),\underline{u}(t))\,\underline{v}(t)$$

and a known trajectory in $\underline{X}\times\underline{U}$ $\quad \{(\underline{x}^*(t),\underline{u}^*(t))\}_{t=0}^{\infty}$,
with no noise ($\underline{w}\equiv\underline{0}$, $\underline{v}\equiv\underline{0}$).

Then

$$\underline{x}(t+1) = \underline{f}(\underline{x}^*(t),\underline{u}^*(t)) + \frac{\partial \underline{f}}{\partial \underline{x}}(x^*,u^*)\left[x(t)-x^*(t)\right] + \frac{\partial \underline{f}}{\partial \underline{u}}(x^*,u^*)\left[u(t)-u^*(t)\right] + \tilde{w}(t) + \text{h.o.t.}$$

where $\quad E[\underline{\tilde{w}}(t)] = \underline{\Gamma}(\underline{x}^*(t),\underline{u}^*(t))\,E[\underline{w}(t)]$

and $\quad E[\underline{\tilde{w}}(t)\,\underline{\tilde{w}}^T(s)] = \underline{\Gamma}(\underline{x}^*(t),\underline{u}^*(t))\,E[\underline{w}(t)\,\underline{w}^T(s)] \times \Gamma(\underline{x}^*(s),\underline{u}^*(s))\,\cancel{\delta(t-s)}$

by Taylor's theorem.

Neglecting the 2nd and higher order partials, which provides a valid approximation for sufficiently small $x(t) - x^*(t)$ and $u(t) - u^*(t)$, and defining

$$\Delta \underline{x}(t) = \underline{x}(t) - \underline{x}^*(t)$$

$$\Delta \underline{u}(t) = \underline{u}(t) - \underline{u}^*(t)$$

then

$$\Delta \underline{x}(t+1) = \underline{F}(t)\,\Delta \underline{x}(t) + \underline{G}(t)\,\Delta \underline{u}(t) + \underline{\bar{w}}(t)$$

where

$$\underline{F}(t) \triangleq \frac{\partial f}{\partial x}(x^*(t), u^*(t))$$

$$\underline{G}(t) \triangleq \frac{\partial f}{\partial u}(x^*(t), u^*(t))$$

A similar equation can be derived for Δy.

For the case where sampling intervals are not uniform, assume an underlying continuous time model

$$\underline{\dot{x}}(t) = \underline{\tilde{f}}(\underline{x}(t), \underline{y}(t), t) + w(t)$$

$$\underline{y}(t) = \tilde{g}(\underline{x}(t), \underline{y}(t), t) + v(t)$$

and sampling times $\{t_k\}_{k=0}^{\infty}$; $t_{k+1} > t_k \;\forall k$.

Integration from t_k to t_{k+1} $\forall k$ converts this model to the previous case.

Nonlinear Regression

Remember our basic linear model

$$y(t) + a_1 y(t-1) + \cdots + a_k y(t-k)$$
$$= b_0 u(t) + b_1 u(t-1) + \cdots + b_k u(t-k).$$

The identification problem, given $\{(y(t), u(t))\}_{t=0}^{N}$, reduces to finding a parameter vector

$$\underline{\theta} = \begin{bmatrix} a_1 \\ \vdots \\ a_k \\ b_0 \\ \vdots \\ b_k \end{bmatrix}$$

That minimizes

$$\sum_{t=K}^{N} \| y(t) - \underline{\varphi}^T(t)\, \underline{\theta} \|^2$$

where

$$\underline{\varphi}(t) = \begin{bmatrix} y(t-1) \\ \vdots \\ y(t-k) \\ u(t) \\ \vdots \\ u(t-k) \end{bmatrix}$$

which is a standard least squares problem.

Nothing in this formulation restricts the elements of $\underline{\varphi}(t)$ to be linearly related to the inputs and measurements.

Let

$$Z_t = [y(t-1) \cdots y(t-k)\ u(t) \cdots u(t-k)]^T$$

and redefine

$$\underline{\varphi}(t) = \begin{bmatrix} \varphi_1(t) \\ \vdots \\ \varphi_n(t) \end{bmatrix}$$

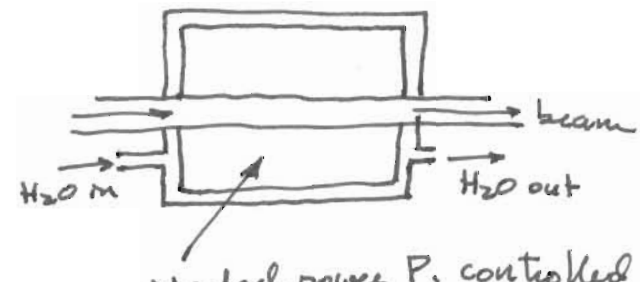
where each $\varphi_i(t)$ is a function of Z_t:

$$\varphi_i(t) \triangleq \psi_i(Z_t).$$

As before, choose $\underline{\theta}$ to minimize

$$\sum_{t=k}^{N} \| y(t) - \underline{\varphi}^T(t)\,\underline{\theta} \|^2$$

Example: Suppose a cavity resonator is used to produce microwave power at a specified frequency ω_0. The power is transferred to a beam of charged ions, except a small portion heats a ~~a~~ cooling jacket of water



injected power P, controlled by duty cycle of pulsed current.

Assume the resonant frequency ω of the cavity is linearly related to the $(\text{temperature})^{-1}$ of the cooling water. Physics tells us that ~~T_{in}~~ the flow rates in and out are roughly equal, and

~~$\dot{T} = (1-\alpha)T + \alpha T_{in} + \beta P$~~

$$\dot{T} = (\cancel{1} - \alpha f)T + \alpha f T_{in} + \beta P \qquad (*)$$

where T is the temperature of the cooling water jacket, T_{in} is the inlet water temperature, P is the input electrical power, and f is the water flow rate.

However, physics doesn't provide the entire story here, because we are neglecting the effects of non-uniform mixing, environmental conditions (air temperature), and nonlinearities of absorbed power as a function of frequency and duty cycle.

Assume we have installed 5 thermocouples on this unit, and can measure flow rate f, duty cycle, frequency of resonance ω, and inlet water temperature T_{in}. Our measurements are therefore

$$z_t = \begin{bmatrix} \omega \\ f \\ d \\ T_{in} \\ T_{amb} \\ T_1 \\ \vdots \\ T_5 \end{bmatrix}$$

The flow rate, duty cycle, inlet temperature, and ambient temperature are exogenous variables. We can consider $T_1 \cdots T_5$ to be state variables, and ω to be an output that depends upon $T_1 \cdots T_5$.

Our ultimate objective is to be able to ~~control~~ manipulate f as a function of the other measurements in order to minimize expected deviations of resonant frequency from a desired value ω_0. To accomplish this, we first need a model of how $T_1 \cdots T_5$ and ω respond to f, d, T_{in}, and T_{amb}. Given eq. (*), this is not likely to be linear. Can we outline a procedure to construct this model and identify its parameters?

We can break this problem down into ~~three~~ four steps:

(1) Postulate model structure

(2) Design of experiments

(3) Data collection

(4) Identify parameters.

(a) What do we know about model structure (from "first principles")?

1. $\dot{T}$'s should depend linearly on T's and fT's (including T_{in} and T_{amb}), and on duty cycle d
2. ω^{-1} should depend linearly on T's (with possible minor contributions by T_{in} and T_{amb}).

Therefore, an initial postulate for the model's structure is

$$T_i(t+\Delta t) = \cancel{T_i(t)+} \sum_{j=1}^{5} \sum_{k=0}^{n} \left[a_{ijk} T_j(t-k\Delta t) + b_{ijk} f T_j(t-k\Delta t) \right] + \sum_{k=0}^{n} \left[c_{ik} T_{in}(t-k\Delta t) + d_{ik} T_{amb}(t-k\Delta t) \right]$$

$$\omega(t) = \sum_{j=1}^{5} e_j T_j(t) + \cancel{g}\, h T_{in}(t) + l T_{amb}(t)$$

Note that there also might be an effect on the state from ω, so we may need to include that later.

(b) Experimental design, and (c) Data collection

We need to gather operational data for different environmental conditions (T_{in}, T_{ams}) — for example, summer and winter, or night and day — and for different flow rates and duty cycles — covering the range of expected operations.

The sampling rate, $1/\Delta t$, also needs to be chosen. This parameter depends upon the effects we wish to capture. For example, an extremely rapid sampling rate may end up ~~modeli~~ modeling turbulence in the cooling water (and miss the bulk thermal effects).

Remember that when a process is sampled a low-pass filter is needed to avoid aliasing.

(d) Identification

We can set up the identification problem as a least squares problem. The parameter vector $\underline{\theta}$ is composed of the a_{ijk}, b_{ijk}, c_{ik}, d_{ik}, e_j, h, and l. Choose $\underline{\Theta}^*$ to minimize

$$\sum_t \| Z_t - \hat{Z}_{t|\Theta} \|^2$$

Note that we have assumed a parameter n — the order of the model's various subsystems. $n=0$ is probably <u>not</u> sufficient, because it would approximate an exponential process over $[t, t+\Delta t]$ by a linear one. For this problem, since this is an IIR (ARMA) ~~(ARX)~~ model, $n=5$ is probably sufficient — meaning that if a 5^{th} order model does not provide good agreement with experiment, we probably need to re-visit step (a), or obtain additional measurement points.

This identification exercise will require a lot of collected data — and, therefore, a lot of computation. For this reason, a recursive identification procedure might be of value.

A final question: How should this type of model be validated?