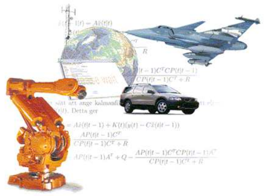
Identification of Linear and Nonlinear Dynamical Systems

Theme 2: Linear Models



Lennart Ljung

Division of Automatic Control
Linköping University
Sweden

Summary Theme 1: Curve Fitting

- We have used the simple case of curve fitting to illustrate basic issues, frameworks and techniques for linear and nonlinear system identification
- Choice of model parametrization, model size and parameter values.
- Parametric – Nonparametric methods
- Parameter values easy: Some version of least squares fit.
- Basic asymptotic properties: $\hat{\theta}_N \to \theta^*$, best possible approximation available in the parameterization (for the used x_t-sequence)
- $\sqrt{N}(\hat{\theta}_N - \theta^*) \sim N(0, P)$, $P = \lambda[E\psi(t)\psi^T(t)]^{-1}$ (Normal distribution)
- Choice of parametric model structure guided by bias-variance trade off (number of parameters)
- Choice of nonparametric method guided by bias-variance trade off (bandwidth of the kernel)

Today: Goal

Goal: Estimate a linear model in discrete or continuous time with or without an additive noise model.

$$y(t) = G(\sigma)u(t) + v(t)$$

σ is differentiation operator p or shift operator q.
The corresponding frequency response function (FRF) is $G(i\omega)$ or $G(e^{i\omega T})$.
Estimating a linear system is the same as estimating its FRF-curve.

Focus in This Theme

1. Frequency response function (FRF)
2. Data in time and frequency domain
3. The richness of parameterizations of linear models
4. Fitting parameterized linear models to data
   - Time domain data
   - Frequency domain data
5. Noise models
6. The asymptotic properties of the estimates
7. Subspace methods
8. Spectral Analysis

Background Material

- Linear differential (difference) equations
- Linear system, State-space model, Transfer function,
- Frequency response function, Bode plot
- Fourier transform, DFT, Signal spectrum
- Noise representation, Kalman Filter

The Frequency Response Function, FRF

A linear system is characterized by its transfer function $G(s)$ (the Laplace transform of its impulse response).
Evaluated on the imaginary axis, this gives the FRF $G(i\omega)$, which describes the response to sinusoidal inputs:

$$u(t) = A\cos(\omega t), \quad y(t) = A_1\cos(\omega t + \phi)$$
$$A_1 = |G(i\omega)|A, \quad \phi = \arg G(i\omega)$$

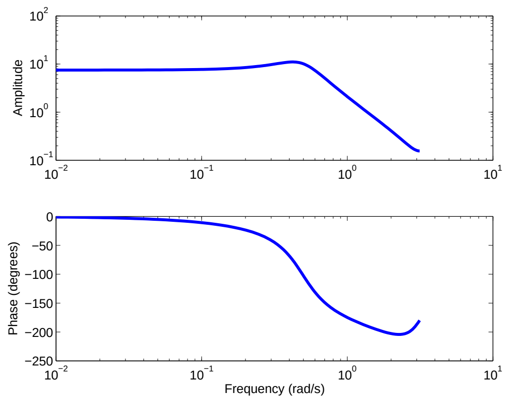
This could be a way of determining G (frequency analysis).
All Frequencies at the same time:

$$Y(i\omega) = G(i\omega)U(i\omega)$$

Y and U are the Fourier transforms of the output and input.

.

The Bode Plot



Focus in This Theme

1. Frequency response function (FRF)
2. Data in time and frequency domain
3. The richness of parameterizations of linear models
4. Fitting parameterized linear models to data
   - Time domain data
   - Frequency domain data
5. Noise models
6. The asymptotic properties of the estimates
7. Subspace methods
8. Spectral Analysis

Data from Dynamic Systems: Input-Output Data

- Discrete time
  - Time-domain: $\{u(1), y(1), u(2), y(2), \dots, u(N), y(N)\}$
  - Frequency-domain $\{U_N(e^{i\omega_1}), Y_N(e^{i\omega_1}), \dots, U_N(e^{i\omega_N}), Y_N(e^{i\omega_N})\}$
    DFT-grid: $\omega_k = 2\pi k/N$

$$U_N(z) = \frac{1}{\sqrt{N}} \sum_{k=1}^{N} u(k) z^{-k}$$

- Continuous time
  - Frequency-domain $\{U_N(i\omega_1), Y_N(i\omega_1), \dots, U_N(i\omega_N), Y_N(i\omega_N)\}$

$$U_N(s) = \frac{1}{\sqrt{N}} \int_0^N u(t) e^{-st} dt$$

(Band limited, periodic data)

Frequency Response Data (FRD)

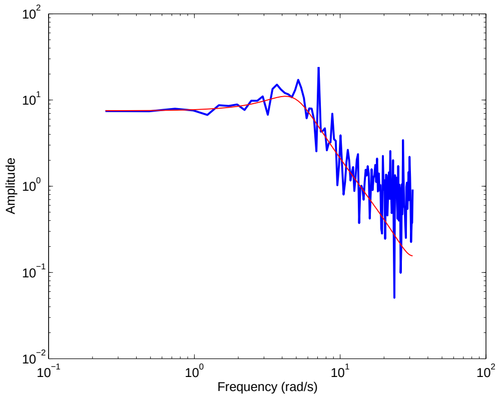
- From frequency analyzers or computed/estimated using FFT techniques
  - $\hat{\hat{G}}(i\omega_k)$ or $\hat{\hat{G}}(e^{i\omega_k}), k = 1, 2, \dots, N$
  - Possibly with uncertainty measures $W(i\omega_k)$
  - Simple estimate, ETFE:

$$\hat{\hat{G}}_N(i\omega) = \frac{Y_N(i\omega)}{U_N(i\omega)} \qquad \text{Variance: } W(i\omega) = \frac{\Phi_v(\omega)}{|U_N(i\omega)|^2}$$

where $\Phi_v(\omega)$ is the spectrum of the output disturbance

  - Other estimates (spectral analysis): smoothed versions of ETFE (more later)

ETFE: Empirical Transfer Function Estimate



Focus in This Theme

1. Frequency response function (FRF)
2. Data in time and frequency domain
3. The richness of parameterizations of linear models
4. Fitting parameterized linear models to data
   - Time domain data
   - Frequency domain data
5. Noise models
6. The asymptotic properties of the estimates
7. Subspace methods
8. Spectral Analysis

Linear Model Structures

Model Structure = Model Parameterization
First without noise model

$$y(t) = G(q, \theta) u(t), \qquad q \text{ shift operator: } qu(t) = u(t+1)$$

FIR: $y(t) = b_1 u(t-1) + b_2 u(t-2) + \dots + b_n u(t-n)$

$$y(t) = B(q) u(t), \quad \theta = [b_1, \dots, b_n]^T$$

$$B(q) = b_1 q^{-1} + \dots + b_n q^{-n}$$

OE: $y(t) = \frac{b_1 q^{-1} + \dots + b_n q^{-n}}{1 + f_1 q^{-1} + \dots + f_n q^{-n}} u(t)$

Linear difference equation

$$y(t) + f_1 y(t-1) + \dots + f_n y(t-n) = b_1 u(t-1) + \dots + b_n u(t-n)$$

Linear Model Structures, cont'd

State space: $x(t+1) = A(\theta) x(t) + B(\theta) u(t)$

$$y(t) = C(\theta) x(t) + D(\theta) u(t)$$

$$G(q, \theta) = C(\theta)[qI - F(\theta)]^{-1} B(\theta) + D(\theta)$$

The matrices can be arbitrarily parameterized by θ

Process model: Static gain, time constant, delay

$$T_p \dot{y}(t) + y(t) = K u(t - T_d)$$

$$G(s, \theta) = \frac{K}{1 + sT_p} e^{-T_d s}$$

Parameterizations of Linear Dynamic Models

A Linear Model Structure is (one way or another) a parameterization of the frequency function

$$G(e^{i\omega}, \theta) \quad \text{or} \quad G(i\omega, \theta)$$

Typical Cases: ($x = e^{i\omega}$ or $x = i\omega$)

- FIR: $G(x, \theta) = \theta_1 + \theta_2 x + \dots + \theta_n x^{n-1}$
- OE: $G(x, \theta) = \frac{\theta_1 + \theta_2 x + \dots + \theta_n x^{n-1}}{1 + \theta_{n+1} x + \dots + \theta_{n+m-1} x^{m-1}}$
- Laguerre and similar: $G(x, \theta) = \sum_{k=1}^{N} \theta_k L_k(x)$
- Via state space: $G(x, \theta) = C(\theta)(xI - A(\theta))^{-1} B(\theta) + D(\theta)$
  - Parameterization of the state space matrices A, B, C, D: Physical, free or canonical
- Local parameterization: $G(x, \theta)$ piecewise constant over a frequency grid (or smoothed versions thereof).
- Process Models: $G(i\omega, \theta) = \frac{K}{1 + i\omega T_p} e^{-i\omega T_d} \quad \theta = [K, T_p, T_d]$

Focus in This Theme

1. Frequency response function (FRF)
2. Data in time and frequency domain
3. The richness of parameterizations of linear models
4. Fitting parameterized linear models to data
   - Time domain data
   - Frequency domain data
5. Noise models
6. The asymptotic properties of the estimates
7. Subspace methods
8. Spectral Analysis

Fitting Models to Data – Parametric Curve Fitting

Bottom line curve-fitting: $y(k) = g(x_k) + v_k, E|v_k|^2 = \lambda_k$
$\min \sum |y(k) - g(x_k, \theta)|^2/\lambda_k$

Time domain: $y(t) = G(q, \theta)u(t) + v(t)\ Ev^2(t) = \lambda$

$$\min \sum |y(t) - G(q, \theta)u(t)|^2$$

Frequency domain:

$$Y_k = G(i\omega_k, \theta)U_k + V_k\ E|V_k|^2 = \lambda, \quad (Y_k = Y(i\omega_k))$$

$$\min \sum |Y_k - G(i\omega_k, \theta)U_k|^2$$

FRD: $\hat{\hat{G}}_k = G(i\omega_k, \theta) + \tilde{V}_k \qquad E|\tilde{V}_k|^2 = W_k$

$$\min \sum |\hat{\hat{G}}_k - G(i\omega_k, \theta)|^2/W_k$$

Connections between Parametric Approaches

- Input-Output Data, Discrete time, DFT frequency grid:

$$\sum |y(t) - G(q, \theta)u(t)|^2 \Longleftrightarrow \sum |Y_k - G(e^{i\omega_k}, \theta)U_k|^2$$

Parseval

- FRD being ETFE for the DFT frequency grid:

$$\hat{\hat{G}}_k = \frac{Y_k}{U_k}, \quad W_k = \frac{\Phi_v(\omega_k)}{|U_k|^2} \quad (\Phi_v(\omega) = \lambda)$$

$$\sum |\hat{\hat{G}}_k - G(e^{i\omega_k}, \theta)|^2/W_k = \sum |Y_k - G(e^{i\omega_k}, \theta)U_k|^2/\lambda$$

Pre-filtering: A Further Degree of Freedom

Pre-filter the data before the fit: $y_F(t) = L(q)y(t), \quad u_F(t) = L(q)u(t)$.
Gives

$$\sum |y_F(t) - G(q, \theta)u_F(t)|^2 = \sum |L(q)(y(t) - G(q, \theta)u(t))|^2$$
$$\sim \sum |L(e^{i\omega_k})|^2 |Y_k - G(e^{i\omega_k}, \theta)U_k|^2$$

"Relevance weighting" in curve fitting. The fit is focused to the frequency ranges where L is large. Think of L as a band-pass filter.

Focus in This Theme

1. Frequency response function (FRF)
2. Data in time and frequency domain
3. The richness of parameterizations of linear models
4. Fitting parameterized linear models to data
   - Time domain data
   - Frequency domain data
5. Noise models
6. The asymptotic properties of the estimates
7. Subspace methods
8. Spectral Analysis

Linear Models with Noise

Assume that the output is corrupted by additive noise with known or unknown properties.

$$y(t) = G(q, \theta) + v(t)$$

$v(t)$ has spectrum $\Phi_v(\omega, \theta) = \lambda |H(e^{i\omega}, \theta)|^2$

Same as

$$y(t) = G(q, \theta)u(t) + H(q, \theta)e(t)$$

where e is white noise with variance λ

$H(q, \theta)$ is monic, i.e. $H(0, \theta) = 1$.

Frequency Domain:

$$Y_k = G(e^{i\omega_k}, \theta)U_k + V_k, \quad E|V_k|^2 = \Phi(\omega_k, \theta), \quad (Y_k = Y(e^{i\omega_k}))$$

Some Typical Model Structures

Acronym	G	H	Equation
ARX	$\frac{B}{A}$	$\frac{1}{A}$	$A(q)y(t) = B(q)u(t) + e(t)$
OE	$\frac{B}{F}$	1	$y(t) = \frac{B(q)}{F(q)}u(t) + e(t)$
ARMAX	$\frac{B}{A}$	$\frac{C}{A}$	$A(q)y(t) = B(q)u(t) + C(q)e(t)$
BJ	$\frac{B}{F}$	$\frac{C}{D}$	$y(t) = \frac{B(q)}{F(q)}u(t) + \frac{C(q)}{D(q)}e(t)$

State-Space:

$$G(q, \theta) = C(\theta)(qI - F(\theta))^{-1}B(\theta), \quad H(q, \theta) = C(\theta)(qI - F(\theta))^{-1}K(\theta) + I$$
$$x(t+1) = F(\theta)x(t) + B(\theta)u(t) + K(\theta)e(t)$$
$$y(t) = C(\theta)x(t) + e(t)$$

- Common or different dynamics in input and noise channels.

One step ahead prediction for models with noise

Assume there is a delay in $G(q)$

$$y(t) = G(q)u(t) + H(q)e(t)$$
$$H^{-1}(q)y(t) = H^{-1}(q)G(q)u(t) + e(t)$$
$$y(t) = [1 - H^{-1}(q)]y(t) + H^{-1}(q)G(q)u(t) + e(t)$$

Since H is monic, $1 - H^{-1}(0) = 0$ so

$$[1 - H^{-1}(q)]y(t) = \tilde{h}_1 y(t-1) + \tilde{h}_2 y(t-2) + \ldots$$

so the RHS of the above expression is actually known at time $t-1$. Since $e(t)$ is unpredictable at time $t-1$ the predictor must be

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

Check the predictor formula

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

1. White additive noise: $H(q) = 1, \quad \hat{y}(t|t-1) = G(q)u(t)$
2. ARX model: (e white)

$$y(t) + a_1 y(t-1) + \ldots + a_n y(t-1) = b_1 u(t-1) + \ldots + b_n u(t-n) + e(t)$$

$$A(q)y(t) = B(q)u(t) + e(t), \quad G(q) = \frac{B(q)}{A(q)}, H(q) = \frac{1}{A(q)}$$
$$[1 - H^{-1}(q)] = -a_1 q^{-1} - \ldots - a_n q^{-n}$$
$$H^{-1}(q)G(q) = B(q)$$
$$\hat{y}(t|t-1) = -a_1 y(t-1) - \ldots - a_n y(t-n) + b_1 u(t-1) + \ldots + b_n u(t-n)$$

3. Kalman filter ...

Parametric Curve Fitting with Noise Models

Bottom line curve-fitting: $y(k) = g(x_k) + v_k, E|v_k|^2 = \lambda_k$

$\min \sum |y(k) - g(x_k, \theta)|^2 / \lambda_k$

Time domain:

$y(t) = \hat{y}(t|\theta) + e(t),\ Ee^2(t) = \lambda$

$\hat{y}(t|\theta) = \left[(1 - H^{-1}(q,\theta))y(t) + H^{-1}(q,\theta)G(q,\theta)u(t)\right]$

$\min \sum |y(t) - [1 - H^{-1}(q,\theta)]y(t) - H^{-1}(q,\theta)G(q,\theta)u(t)|^2$

$= \min \sum |[H^{-1}(q,\theta)(y(t) - G(q,\theta)u(t))]|^2$

Frequency domain:

$Y_k = G(e^{i\omega_k})U_k + V_k \quad E|V_k|^2 = \Phi(\omega_k) = \lambda |H(e^{i\omega_k},\theta)|^2$

$\min \sum |Y_k - G(e^{i\omega_k},\theta)U_k|^2 / \Phi(\omega_k)$

$= \min \sum |H^{-1}(e^{i\omega_k},\theta)[Y_k - G(e^{i\omega_k},\theta)U_k]|^2$

Focus in This Theme

1. Frequency response function (FRF)
2. Data in time and frequency domain
3. The richness of parameterizations of linear models
4. Fitting parameterized linear models to data
   - Time domain data
   - Frequency domain data
5. Noise models
6. The asymptotic properties of the estimates
7. Subspace methods
8. Spectral Analysis

Asymptotic Properties of Estimated Models

- Assume that data have been generated by $y(t) = G_0(q)u(t) + H_0(q)e(t)$, where e is white noise with variance λ.
- $$\hat{\theta}_N = \arg\min \sum |L(q)\varepsilon(t,\theta)|^2$$ $$\varepsilon(t,\theta) = H^{-1}(q,\theta)[y(t) - G(q,\theta)u(t)]$$
- Properties of the estimate $\hat{\theta}_N$
- A parameter-free assessment of model quality for linear systems and models
- How to affect the model quality?

Properties of the Estimate $\hat{\theta}_N$ as $N \to \infty$

- 1. $\hat{\theta}_N \to \theta^* = \arg\min \mathrm{E}[L(q)\varepsilon(t,\theta)]^2$

  "The estimate converges to the best possible approximation of the true system"

  "Best possible": ....

  If $\varepsilon(t,\hat{\theta}_N) \approx$ white noise then

  $\mathrm{Cov}\ \hat{\theta}_N \approx \frac{\lambda}{N}[\mathrm{E}\psi(t)\psi^T(t)]^{-1}$

  $\lambda = \mathrm{E}\varepsilon^2(t,\hat{\theta}_N)$

  $\psi(t) = \frac{d}{d\theta}\hat{y}(t|\theta)$ d-dimensional vector

  NB! This covariance can be estimated!

  Also, asymptotic normality of $\sqrt{N}(\hat{\theta}_N - \theta^*)$

Quality in The Frequency Domain: Bias

True system: $y(t) = G_0(q)u(t) + v(t)$

Model: $\hat{G}_N(e^{i\omega}) = G(e^{i\omega},\hat{\theta}_N), \quad \hat{H}_N(e^{i\omega}) = H(e^{i\omega},\hat{\theta}_N)$

Translate properties of $\hat{\theta}_N$ to $\hat{G}_N$ & $\hat{H}_N$

$$\hat{G}_N(e^{i\omega}) \to G^*(e^{i\omega}), \quad \hat{\theta}_N \to \theta^* \qquad \text{as } N \to \infty$$

$$\theta^* \approx \arg\min \int_{-\pi}^{\pi} |\ G_0(e^{i\omega}) - G(e^{i\omega},\theta)\ |^2 \cdot \frac{\Phi_u(\omega) \cdot |\ L(e^{i\omega})\ |^2}{|\ H(e^{i\omega},\theta^*)\ |^2} d\omega$$

$G^*(e^{i\omega})$ is closest to $G_0(e^{i\omega})$ in the norm $Q(\omega) = \dfrac{\Phi_u(\omega)|L(e^{i\omega})|^2}{|H(e^{i\omega},\theta^*)|^2}$

Φ_u : Input spectrum, $\quad L$: Pre-filter

A Paradox

A high order system is approximated in the ARX-structure

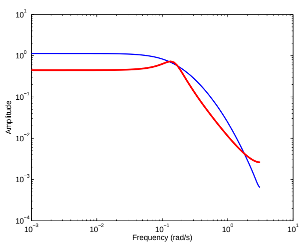
$$(1 + a_1q^{-1} + a_2q^{-2})y(t) = (b_1q^{-1} + b_2q^{-2})u(t) + e(t)$$

and in the OE-structure

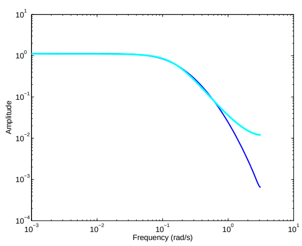
$$y(t) = \frac{b_1q^{-1} + b_2q^{-2}}{1 + f_1q^{-1} + f_2q^{-2}}u(t) + e(t)$$

Can you explain the difference?

True (thin) and Estimated (thick)

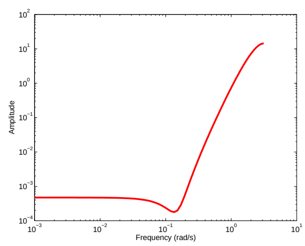


ARX

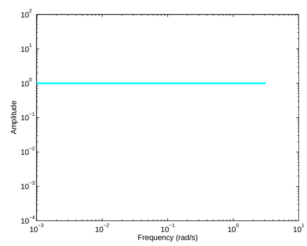


OE

The Weighting Functions



ARX

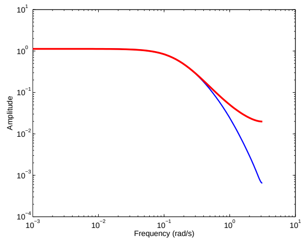


OE

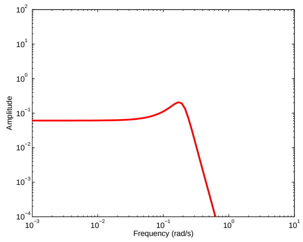
The Use of Pre-filtering

L is chosen as a low-pass filter

```
m=arx(data,[2 2 1],'focus',[0 0.2])
```



Estimated model



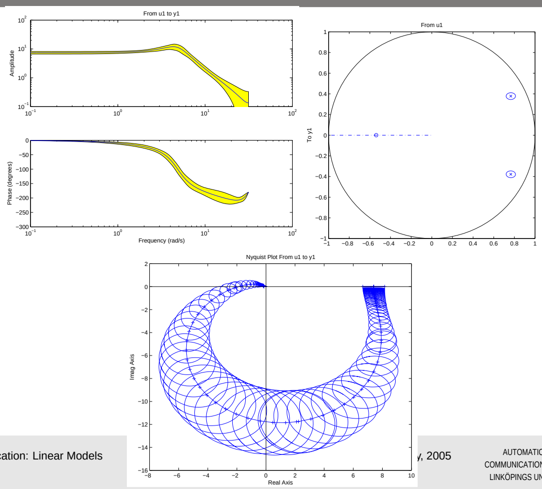
The weighting function

Model Quality in The Frequency Domain: Variance

The asymptotic distribution and variance of $\hat{\theta}_N$ can be translated to those of $G(e^{i\omega}, \hat{\theta}_N) = \hat{G}_N(e^{i\omega})$ by

Gauss' Approximation Formula: Cov $f(\hat{\theta}_N) \approx f'$Cov $\hat{\theta}_N (f')^T$

Example: Model Properties with Uncertainty



Model Quality in The Frequency Domain: Variance

The asymptotic distribution and variance of $\hat{\theta}_N$ can be translated to those of $G(e^{i\omega}, \hat{\theta}_N) = \hat{G}_N(e^{i\omega})$ by
Gauss' Approximation Formula: Cov $f(\hat{\theta}_N) \approx f'$Cov $\hat{\theta}_N (f')^T$
Approximate expressions for high order models:
Open Loop, Asymptotically in N and n

$$\text{Cov } \hat{G}_N(e^{i\omega}) \approx \frac{n}{N} \cdot \frac{\Phi_v(\omega)}{\Phi_u(\omega)}$$

n=model order, N=# of data

Φ_v: noise spectrum, Φ_u: input spectrum.

In General:

$$\text{Cov} \begin{bmatrix} \hat{G}_N(e^{i\omega}) \\ \hat{H}_N(e^{i\omega}) \end{bmatrix} \approx \frac{n}{N} \Phi_v(\omega) \begin{bmatrix} \Phi_u(\omega) & \Phi_{ue}(\omega) \\ \Phi_{ue}(-\omega) & \lambda \end{bmatrix}^{-1}$$

How To Affect the Model Quality

$$\text{Cov } \hat{\theta}_N \approx \frac{\lambda}{N} \cdot [\text{E}\psi(t)\psi^T(t)]^{-1}$$

$$\text{Cov } \hat{G}_N \approx \frac{n}{N} \cdot \frac{\Phi_v(\omega)}{\Phi_u(\omega)}$$

$$\hat{\theta}_N \to \theta^* = \arg\min \text{E}[L(q)\varepsilon(t,\theta)]^2)$$

$$\hat{G}_N \to G^* = \text{closest to } G_0 \text{ in the } \frac{\Phi_u(\omega) \mid L(e^{i\omega}) \mid^2}{\mid H^*(e^{i\omega}) \mid^2} \text{ norm}$$

Focus in This Theme

1. Frequency response function (FRF)
2. Data in time and frequency domain
3. The richness of parameterizations of linear models
4. Fitting parameterized linear models to data
   - Time domain data
   - Frequency domain data
5. Noise models
6. The asymptotic properties of the estimates
7. Subspace methods
8. Spectral Analysis

Subspace Methods

$$x(t+1) = Ax(t) + Bu(t) + w(t)$$
$$y(t) = Cx(t) + Du(t) + v(t)$$

Estimate the matrices A, B, C, D.
Suppose, for a second, that the states $x(t)$ were known. Then the above expression is a linear regression: Let

$$Y(t) = \begin{bmatrix} x(t+1) \\ y(t) \end{bmatrix}$$

$$\Phi(t) = \begin{bmatrix} x(t) \\ u(t) \end{bmatrix}$$

The Linear Regression

Then

$$Y(t) = \Theta\Phi + \nu(t)$$

with

$$\Theta = \begin{bmatrix} A & B \\ C & D \end{bmatrix}$$

$$\nu = \begin{bmatrix} w(t) \\ v(t) \end{bmatrix}$$

All matrices of interest, including the covariance matrix of ν could then be estimated using the Least Squares method. With the covariance matrix of ν, the optimal Kalman gain could then be computed.

.

How to Find the States and Conceptual Algorithm

Fact: All (interesting) states can be found as linear combinations of the k-step ahead predictors $\hat{y}(t+k|t)$, $k = 1, \ldots, n$ (the predicted value of $y(t+k)$ based on input-output data up to time t. No prediction of the effect of inputs after time t.)

So estimate these k-step ahead predictors using ARX-models, and determine from these the good linear combinations to form the states x.

Use these x to form the linear regression to estimate A, B, C, D.

.

Subspace Algorithms: Essential Features

- State space basis selected automatically
- Form sample covariances of y and u: One SVD and one QR-step
- No iterations
- Quality properties not fully understood

More Formal Calculations, 1/5

$$y(t+k) = \sum_{j=-\infty}^{t+k} h^u_{t+k-j} u(j) + h^e_{t+k-j} e(j)$$

$$\hat{y}(t+k|t) = \sum_{j=-\infty}^{t} h^u_{t+k-j} u(j) + h^e_{t+k-j} e(j)$$

Let

$$Y^x(t) = \begin{bmatrix} \hat{y}(t+1|t) \\ \vdots \\ \hat{y}(t+n|t) \end{bmatrix}$$

.

More Formal Calculations, 2/5

So, all (Kalman) states $x(t)$, in any state-space representation can be written as linear combinations of $Y^x(t)$:

$$x(t) = LY^x(t)$$

for some L. The (minimal) order of the state-space representation is the rank of $Y^x(t), t = 1, \ldots,$.
So, with $Y^x(t), t = 1, \ldots, N$ given, pick L, so that $x(t)$ becomes well conditioned. This includes the choice of dimension of x. Typically, apply SVD to

$$Y^N = \begin{bmatrix} Y^x(1) & Y^x(2) & Y^x(N) \end{bmatrix}$$

Once $x(t), t = 1, \ldots, N$ have been determined, proceed as above to find the state-space matrices.

.

More Formal Calculations, 3/5

How to estimate the predictors:

$$y(t+k) = \sum_{j=-\infty}^{t+k} h^u_{t+k-j} u(j) + h^e_{t+k-j} e(j) (*)$$

$$\hat{y}(t+k|t) = \sum_{j=-\infty}^{t} h^u_{t+k-j} u(j) + h^e_{t+k-j} e(j)$$

$e(t)$ and $y(t)$ have an invertible relationship.

$$y(t+k) = \sum_{j=-\infty}^{t+k-1} \tilde{h}^u_{t+k-j} u(j) + \tilde{h}^e_{t+k-j} y(j) + e(t+k) + \tilde{h}^u_0 u(t+k) (**)$$

so replace $e(j)$ in (*) by y and u from (**): $\ldots$

More Formal Calculations, 4/5

$$y(t+k) = \sum_{j=-\infty}^{t+k} \tilde{h}^u_{t+k-j} u(j) + \sum_{j=-\infty}^{t} \tilde{h}^e_{t+k-j} y(j) + \sum_{j=t+1}^{t+k} h^e_{t+k-j} e(j)$$

$$\hat{y}(t+k|t) = \sum_{j=-\infty}^{t} \tilde{h}^u_{t+k-j} u(j) + \tilde{h}^e_{t+k-j} y(j)$$

Now, truncate the first equation at $j = t - n_1$ rather than at $j = -\infty$, and estimate $\tilde{h}$ using the least squares method. Use these estimates in the second equation to estimate $\hat{y}$. The value n_1 corresponds to the "auxiliary order". All of this can be done numerically efficient by projections.

The Subspace Method Algorithm

The essence of the subspace methods is as follows

1. Select n and n_1 and estimate $Y^x(t), t = 1, \ldots, N$.
2. Determine a good choice of L in $x(t) = LY^x(t)$ (including dimension) using SVD or similar decomposition
3. Possibly determine n by visual inspection of the singular values in the above expression.
4. Estimate A, B, C and D by least squares in the state-space model, treating $x(t)$ as a measured sequence.
5. Use the covariance matrix of ν to compute the Kalman Filter gain K.

Nonparametric Curve Fitting

Smoothing bottom line:

$$\hat{g}_N(x) = \sum_{k=1}^{N} C(x, x_k) y(k), \quad \sum_{k=1}^{N} C(x, x_k) = 1 \ \forall x$$

Often $C(x, x_k) = \mathcal{N}\tilde{c}(x - x_k)/\lambda_k$ and $\tilde{c}(r) = 0$ for $|r| > \gamma$, $\gamma =$ the "bandwidth"

Curve fitting of the ETFE: $y(k) = \hat{\hat{G}}_k = \frac{Y_k}{U_k}, \quad \lambda_k = \frac{\Phi_v}{|U_k|^2}, \quad Y_k = Y(i\omega_k)$

$$\hat{G}(i\omega) = \frac{\sum \tilde{c}(\omega - \omega_k) Y_k U_k^*}{\sum \tilde{c}(\omega - \omega_k)|U_k|^2}$$

Bandwidth = Frequency Resolution – Could be frequency dependent $(C(\omega, \omega_k) = \tilde{c}(\omega - \omega_k, \omega))$

Spectral Analysis (with Frequency Dependent Resolution),

Conclusions Theme 2

- Rich possibilities to parameterize linear models (their FRFs)
- Curve fitting to parameterized FRFs
- Limit model is a weighted fit between the true FRF and the model one
- Variance of estimated FRF is approximatively proportional to noise-to-signal ratio frequency by frequency
- Subspace methods can be seen as a two step process: Estimate the states and find the state space matrices by linear regression
- Nonparametric curve fitting is the spectral analysis method