Weighted Null Space Fitting
A link between the
Prediction Error Method and Subspace Identification

Håkan Hjalmarsson

Miguel Galrinho, Niklas Everitt, Mengyuan Fang, Cristian R. Rojas

Divison of Decision and Control Systems
School of Electrical Engineering and Computer Science
and
Centre for Advanced Bioproduction
KTH Royal Institute of Technology

2019-05-13

Outline

Problem Setting

True System:

$$y_t = \frac{L_\circ(q)}{F_\circ(q)} u_t + \frac{C_\circ(q)}{D_\circ(q)} e_t$$

Model:

$$y_t = \frac{L(q,\theta)}{F(q,\theta)} u_t + v_t$$

$$v_t = H(q,\theta) e_t$$

$$H(q,\theta) = \frac{C(q,\theta)}{D(q,\theta)}$$

Estimate θ!

Prediction Error Method

Minimize cost function:

$$V_N^{\mathsf{PEM}}(\theta) = \sum_{t=1}^{N} \left[\frac{1}{H(q,\theta)} \left(y_t - \frac{L(q,\theta)}{F(q,\theta)} u_t \right) \right]^2$$

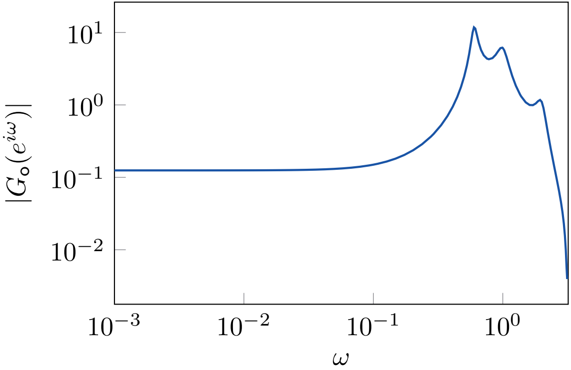
Non-convex!

Test example

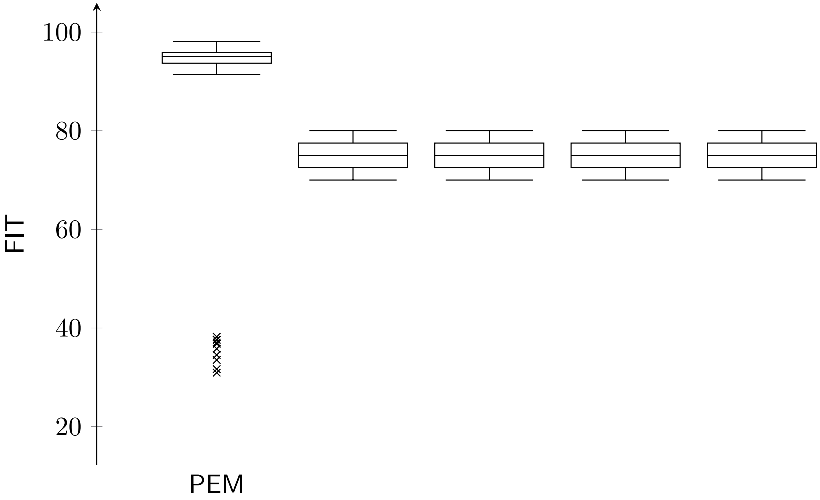
Box-Jenkins:

$$y_t = \frac{L_\circ(q)}{F_\circ(q)} u_t + \frac{1 + 0.8q^{-1}}{1 - 0.9q^{-1}} e_t$$

System (6^{th} order):



Simulation



Simulation

Alternatives and complements to PEM:

- Instrumental variable methods
- Subspace methods
- Iterative least-squares methods
- Multi-step high-order least-squares methods

Iterative Least Squares Methods

Output Error Models: $y_t = \frac{L(q)}{F(q)} u_t + e_t$

PEM: $V_N^{\text{PEM}}(\theta) = \sum_{t=1}^{N} (y_t - \frac{L(q)}{F(q)} u_t)^2$

Non-convex :(Tempting to try:

$$\sum_{t=1}^{N} (F(q) y_t - L(q) u_t)^2$$

Modified PEM is Least-Squares!
but biased since we (for open loop data) are minimizing

$$\sum_{t=1}^{N} \left(F(q) \left(\frac{L_o(q)}{F_o(q)} u_t + e_t \right) - L(q) u_t \right)^2 \approx \sum_{t=1}^{N} \left(\frac{F(q) L_o(q) - F_o(q) L(q)}{F_o(q)} u_t \right)^2 + \sum_{t=1}^{N} (F(q) e_t)^2$$

Steiglitz-McBride

Step 1: Estimate $F(q,\theta)y_t = L(q,\theta)u_t + e_t \quad \Longrightarrow \quad \hat{\theta}_N^1$

Step 2: Filter the data according to

$$y_t^f = \frac{1}{F(q,\hat{\theta}_N^1)} y_t, \qquad u_t^f = \frac{1}{F(q,\hat{\theta}_N^1)} u_t$$

Step 3: Estimate $F(q,\theta)y_t^f = L(q,\theta)u_t^f + e_t \quad \Longrightarrow \quad \hat{\theta}_N^2$

Iterate!

$$V_N(\theta) = \frac{1}{N}\sum_{t=1}^{N}\left[\frac{F(q,\theta)}{F(q,\hat{\theta}_N^k)} y_t - \frac{L(q,\theta)}{F(q,\hat{\theta}_N^k)} u_t\right]^2$$

$\theta_N^k \to \theta_o$ as $k \to \infty$ and $N \to \infty$

...but the noise must be white and $\hat{\theta}_N$ is not asymptotically efficient!

Multi-Step High Order Methods

- Prefiltering
- Residual estimation
- Optimal model reduction
- Weighted Null Space Fitting (WNSF)

High-Order ARX Models

$$y_t = G_\circ(q)u_t + H_\circ(q)e_t \iff A_\circ(q)y_t = B_\circ(q)u_t + e_t$$

$$A_\circ(q) = \frac{1}{H_\circ(q)} = 1 + a_1^\circ q^{-1} + a_2^\circ q^{-2} + \cdots$$

$$B_\circ(q) = \frac{G_\circ(q)}{H_\circ(q)} = b_1^\circ q^{-1} + b_2^\circ q^{-2} + \cdots$$

$$A(q,\eta^n)y_t = B(q,\eta^n)u_t + e_t$$

$$A(q,\eta^n) = 1 + a_1 q^{-1} + \cdots + a_n q^{-n} \qquad B(q,\eta^n) = b_1 q^{-1} + \cdots + b_n q^{-n}$$

$$\eta^n = \begin{bmatrix} a_1 & \cdots & a_n & b_1 & \cdots & b_n \end{bmatrix}^\top$$

Choose n “sufficiently large” for the truncation error to be “sufficiently small!”

Special case: High-Order FIR

$$y_t = B(q, \eta^n)u_t + e_t = G(q, \eta^n)u_t + e_t$$

Matrix form:

$$Y = T_{N\times n}(u)\eta + E, \quad T_{N\times n}(u) = \begin{bmatrix} u_1 & 0 & 0 & \dots & 0 \\ u_2 & u_1 & 0 & \dots & 0 \\ u_3 & u_2 & u_1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ u_N & u_{N-1} & u_{N-2} & \dots & u_{N-n+1} \end{bmatrix}$$

$$\text{Cov } \hat{\eta}^n = \sigma^2 \left(T_{N\times n}^T(u) T_{N\times n}(u) \right)^{-1}$$

Prefiltering

$$y_t = \frac{L_o(q)}{F_o(q)} u_t + H_o(q) e_t, \quad \hat{A}(q) \approx H_o^{-1} \Rightarrow$$

$$\hat{A}(q) y_t = \frac{L_o(q)}{F_o(q)} \hat{A}(q) u_t + \hat{A}(q) H_o(q) e_t \approx \frac{L_o(q)}{F_o(q)} \hat{A}(q) u_t + e_t$$

Now use SM on prefiltered data $\{\hat{A}(q) y_t, \hat{A}(q) u_t\}$

The Box-Jenkins Steiglitz McBride Method.

For open loop data, BJSM is

- consistent
- asymptotically efficient for Gaussian noise (even for OE models!)
- still need to iterate ($k \to \infty$)

Prefiltering

Notice BJSM uses both high-order model and data when estimating the model.

However, the high-order ARX model is (almost) a sufficient statistic, so from a statistical perspective we should be able to use this model only when estimating the model.

Residual Estimation

For a while we will for simplicity consider open loop data and the OE-case:

$$y_t = \frac{L_o(q)}{F_o(q)} u_t + e_t$$

and use a high-order FIR model

$$y_t = G(q) u_t + e_t, \; G(q) = \sum_{k=1}^{n} \eta_k q^{-k}$$

- High order predictor: $\hat{y}_t = \hat{G}(q) u_t$
- Form residuals: $\varepsilon_t = y_t - \hat{y}_t$
- Use residuals in estimation: $y_t = \frac{L(q)}{F(q)} u_t + \varepsilon_t$

Net result:

$$y_t = \frac{L(q)}{F(q)} u_t + y_t - \hat{y}_t \;\Rightarrow\; \hat{y}_t = \frac{L(q)}{F(q)} u_t$$

$V_N^{\mathsf{RE}}(\theta) = \sum_t \left(\hat{y}_t - \frac{L(q)}{F(q)} u_t \right)^2$

Simulated output used instead of the real output - only high order model used!

Optimal model reduction

Model reduction taking the statistical properties of the high order estimate into account.

- Use the (asymptotic) distribution of $\hat{\eta}$

$$V_N^{\text{E-ML}}(\theta) = \left(\eta^n(\theta) - \hat{\eta}_N^n\right)^\top \text{cov}[\hat{\eta}_N^n]^{-1} \left(\eta^n(\theta) - \hat{\eta}_N^n\right)$$

The Extended Invariance Principle (EXIP)

- Use the asymptotic distribution of $G(e^{i\omega}, \hat{\eta}_N^n)$

$$\sqrt{N}(G(e^{i\omega}, \hat{\eta}_N^n) - G_o(e^{i\omega}))) \sim AsN\left(0, \frac{\sigma^2}{\Phi_u(\omega)}\right)$$

$$V_N^{\text{A-ML}}(\theta) = \int_0^{2\pi} |G(e^{i\omega}, \hat{\eta}_N^n) - G(e^{i\omega}, \theta)|^2 \Phi_u(\omega) d\omega$$

Asymptotic ML

Weighted Null Space Fitting

$$\frac{L(q)}{F(q)} = G(q) \quad \Rightarrow \quad F(q)G(q) - L(q) = 0$$

$$\Rightarrow \quad F(q)\hat{G}(q) - L(q) = F(q)(G(q) + \Delta_G(q)) - L(q) = F(q)\Delta_G(q)$$

Find F and L s.t. $F(q)\hat{G}(q) - L(q)$ behaves in a statistical way as $F(q)\Delta_G(q)$

In equation form: $F(q) = 1 + f_1 q^{-1} + \ldots f_m q^{-m} = 1 + \tilde{F}(q) \Rightarrow$

$$F(q)G(q) - L(q) = (1 + \tilde{F}(q))G(q) - L(q) = G(q) - \begin{bmatrix} 1 & -G(q) \end{bmatrix} \begin{bmatrix} L(q) \\ \tilde{F}(q) \end{bmatrix}$$

$$= g_1 q^{-1} + \ldots g_n q^{-n} - \begin{bmatrix} 1 & -(g_1 q^{-1} + \ldots + g_n q^{-n}) \end{bmatrix} \begin{bmatrix} l_1 q^{-1} + l_m q^{-m} \\ f_1 q^{-1} + \ldots + f_m q^{-m} \end{bmatrix}$$

$$\Leftrightarrow \ \eta - Q(\eta)\theta$$

$$V_N^{\text{WNSF}}(\theta) = (\hat{\eta}_N^n - Q(\hat{\eta}_N^n)\theta)^\top \left(T_{n\times n}(F(q,\theta)) \text{cov}[\hat{\eta}_N^n] T_{n\times n}^T(F(q,\theta)) \right)^{-1} (\hat{\eta}_N^n - Q(\hat{\eta}_N^n)\theta)$$

$T_{n\times n}(F(q,\theta))$ $n \times n$ lower Toeplitz matrix of coefficients of $F(q,\theta)$

Asymptotic Equivalence

Residual estimation, optimal model order reduction and WNSF are equivalent for large sample sizes.

"Proof"

- Residual estimation: $V_N^{\text{RE}}(\theta) = \sum_t \left(\hat{y}_t - \frac{L(q)}{F(q)} u_t\right)^2$
- Asymptotic ML: $V_N^{\text{A-ML}}(\theta) = \int_0^{2\pi} |G(e^{i\omega}, \hat{\eta}_N^n) - G(e^{i\omega}, \theta)|^2 \Phi_u(\omega) d\omega \approx$ $\int_0^{2\pi} |G(e^{i\omega}, \hat{\eta}_N^n) - G(e^{i\omega}, \theta)|^2 |U_N(e^{i\omega})|^2 d\omega =$ (Parseval) $= \sum_t \left(\hat{y}_t - \frac{L(q)}{F(q)} u_t\right)^2$
- EXIP: $V_N^{\text{E-ML}}(\theta) = \left(\hat{\eta}^n(\theta) - \hat{\eta}_N^n\right)^\top \text{cov}[\hat{\eta}_N^n]^{-1} \left(\eta^n(\theta) - \hat{\eta}_N^n\right)$
  but $\sigma^2 \text{cov}[\hat{\eta}_N^n]^{-1} = T_{N\times n}(u)^T T_{N\times n}(u)$, and

$$T_{N\times n}(u)(\eta^n(\theta) - \hat{\eta}_N^n) = \begin{bmatrix} G(q,\theta)u_1 - G(q,\hat{\eta}_N^n)u_1 \\ \vdots \\ G(q,\theta)u_N - G(q,\hat{\eta}_N^n)u_N \end{bmatrix} = \begin{bmatrix} G(q,\theta)u_1 - \hat{y}_1 \\ \vdots \\ G(q,\theta)u_N - \hat{y}_N \end{bmatrix}$$

- WNSF:
  $V_N^{\text{WNSF}}(\theta) = (\hat{\eta}_N^n - Q(\hat{\eta}_N^n)\theta)^\top \left(T_{n\times n}(F(q,\theta)) \text{cov}[\hat{\eta}_N^n] T_{n\times n}^T(F(q,\theta))\right)^{-1} (\hat{\eta}_N^n - Q(\hat{\eta}_N^n)\theta)$
  $\hat{\eta}_N^n - Q(\hat{\eta}_N^n)\theta \Leftrightarrow$

$$\begin{aligned} F(q,\theta)G(q,\hat{\eta}_N^n) - L(q,\theta) &= F(q,\theta)G(q,\hat{\eta}_N^n) - L(q,\theta) - \underbrace{(F(q,\theta)G(q,\eta(\theta)) - L(q,\theta))}_{0} \\ &= F(q,\theta)(G(q,\hat{\eta}_N^n) - G(q,\eta(\theta))) \Leftrightarrow T_{n\times n}(F(q,\theta))(\hat{\eta}_N^n - \eta(\theta)) \end{aligned}$$

Towards Least-Squares

Residual Estimation: $\sum_t (\hat{y}_t - \frac{L(q)}{F(q)} u_t)^2$

Still as non-convex as PEM. Advantage??

Modified cost function: $\sum_t (F(q)\hat{y}_t - L(q)u_t)^2$

Least-Squares! but is it any good, c.f. modified PEM?

Let the order n of the FIR model G grow to infinity: $n(N) \to \infty$ as $N \to \infty$.
$\Rightarrow \hat{G}(q) \to G_o(q) \Rightarrow$ Error in $\hat{y}_t$ vanishes as $N \to \infty \Rightarrow \hat{y}_t = G_o(q)u_t$

Least-squares estimate consistent (unlike modified PEM)!

Consistent Least-Squares Estimation

- Residual Estimation: $\sum_t (F(q)\hat{y}_t - L(q)u_t)^2$
- Optimal model order reduction (Asymptotic ML): $\int_0^{2\pi} |F(e^{i\omega})G(e^{i\omega}, \hat{\eta}_N^n) - L(e^{i\omega})|^2 \Phi_u(\omega) d\omega$
- WNSF: $(\hat{\eta}_N^n - Q(\hat{\eta}_N^n)\theta)^\top \left(T_{n\times n}(1)\text{cov}[\hat{\eta}_N^n]T_{n\times n}^T(1)\right)^{-1} (\hat{\eta}_N^n - Q(\hat{\eta}_N^n)\theta)$

All consistent if $n(N) \to \infty$ at a suitable rate:

- Not too slow: $n(N)/(\log(N))^{1+\delta} \to \infty$ for some $\delta > 0$
- Not too fast: $n^{4+\delta}(N)/N \to 0$

Towards Asymptotically Efficient Least-Squares

Residual Estimation: $\sum_t (F(q)\hat{y}_t - L(q)u_t)^2$

$$\hat{y}_t = G(q, \hat{\eta}_N^n)u_t = G_o(q)u_t + \Delta_G(q)u_t$$

$$F(q)\hat{y}_t - L(q)u_t = (F(q)G_o(q) - L(q))u_t + F(q)\Delta_G(q)u_t$$

$F(q)\Delta_G(q)u_t$ random error term. Has to be white for asymptotic efficiency.

$\Delta_G(q) \Leftrightarrow \hat{\eta}_N^n - \eta_o$ which has covariance $\sigma^2 \left(T_{N\times n}^T(u) T_{N\times n}(u)\right)^{-1}$

$\Rightarrow \Delta_G(q)u_t$ is temporally white!

$\Rightarrow F(q)\Delta_G(q)u_t$ is NOT temporally white :(

Asymptotically Efficient Least-Squares

Idea: Two-steps

Residual Estimation:

1. Minimize $\sum_t (F(q)\hat{y}_t - L(q)u_t)^2 \quad \Rightarrow \hat{L},\ \hat{F}$ (consistent)
2. $\sum_t (F(q)\frac{1}{\hat{F}(q)}\hat{y}_t - L(q)\frac{1}{\hat{F}(q)}u_t)^2 \quad \Rightarrow \hat{\hat{L}},\ \hat{\hat{F}}$

Result: $\hat{\hat{L}},\ \hat{\hat{F}}$ asymptotically efficient if $n(N) \to \infty$ at a suitable rate.

- Optimal model order reduction (Asymptotic ML):

$$\int_0^{2\pi} |F(e^{i\omega})G(e^{i\omega}, \hat{\eta}_N^n) - L(e^{i\omega})|^2 \frac{\Phi_u(\omega)}{|\hat{F}(e^{i\omega})|^2} d\omega$$

- WNSF: $(\hat{\eta}_N^n - Q(\hat{\eta}_N^n)\theta)^\top \left(T_{n\times n}(\hat{F}(q))\text{cov}[\hat{\eta}_N^n]T_{n\times n}^T(\hat{F}(q))\right)^{-1}(\hat{\eta}_N^n - Q(\hat{\eta}_N^n)\theta)$

Summary: Three steps: i) High order LS, ii) OLS, iii) WLS

A Brief History of Iterative Least-Squares and High-Order Methods

- Durbin 1959: MA. High order AR-model. Clever way of using that model as weighting as well $\Rightarrow$ Only two steps!
- Durbin 1960: ARMA. High order AR-model. Alternate between MA-part (using previous result), and AR-part (easy).
- Santathan & Koerner 1963: Steiglitz-McBride in frequency domain.
- Steiglitz-Mcbride 1965: Steiglitz-McBride iterations.
- Mayne & Firoozan 1982: ARMA. Residual estimation. All three steps. Consistency & asymptotic efficiency but when first N and then n tends to infinity.
- Hannan & Rissanen 1982: ARMA. Residual estimation. Uses model in step 2 to form new residual estimate. Order estimation. Recursive. $n = n(N)$. Consistency and asymptotic efficiency.
- Hannan & Kavaleris 1983: As Mayne & Firoozan but consistency analyzed for $n = n(N)$.
- Mayne, Åström & Clark 1984: As Mayne & Firoozan but recursive.
- Hannan & Kavaleris 1984: As Hannan & Rissanen but multivariate & order recursive.

A Brief History of Iterative Least-Squares and High-Order Methods

- Zhu 1989: ASYM - Asymptotic ML in time-domain (=Residual estimation)
- Wahlberg 1989: Asymptotic ML.
- Zhu 2011: Box-Jenkins Steiglitz-Mcbride. Prefiltering method.
- Dufour & Jouini 2014: VARMA. Multi-step.
- Galrinho, Rojas and Hjalmarsson 2014: WNSF.
- Everitt, Galrinho and Hjalmarsson 2017: MORSM. Residual estimation.
- Fang, Galrinho & Hjalmarsson 2017: WNSF. Recursive.

Multi-step LS vs subspace identification

Subspace id:

1) Estimate Hankel matrix

$$\mathcal{H} = \begin{bmatrix} g_1 & g_2 & g_3 & \cdots \\ g_2 & g_3 & g_4 & \cdots \\ g_3 & g_4 & g_5 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix} = \mathcal{O}_e \mathcal{C}_e$$

2) Obtain estimate of extended observability matrix $\mathcal{O}_e$ using SVD

$$\begin{aligned} W_1 \hat{\mathcal{H}} W_2 &= U S V^T \\ \mathcal{O}_e &= W_1^{-1} \bar{U} \bar{S}^{1/2} \end{aligned}$$

where $\bar{U}$ and $\bar{S}$ truncated versions of U and S.

This means that the range space of $\mathcal{H}$ is estimated.

3) Estimate state-space matrices from $\mathcal{O}_e$.

Data only used in Step 1. W_1 can be used to affect the statistical accuracy. Not clear what the optimal weighting is.

Multi-step LS vs subspace identification

WNSF:

$$F(q)G(q) - L(q) = 0 \;\Rightarrow\; (1+f_1q^{-1}+\ldots+f_mq^{-m})(g_1q^{-1}+\ldots g_nq^{-n}) - (l_1q^{-1}+\ldots+l_mq^{-m}) = 0$$

Look at delays higher than m:

$$\begin{bmatrix} g_1 & g_2 & g_3 & \cdots & g_f \\ g_2 & g_3 & g_4 & \cdots & g_{f+1} \\ g_3 & g_4 & g_5 & \cdots & g_{f+2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ g_f & g_{f+1} & g_{f+2} & \cdots & g_{2f-1} \end{bmatrix} \begin{bmatrix} f_m \\ \vdots \\ f_1 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = 0$$

Multi-step LS vs subspace identification

but also

$$\begin{bmatrix} g_1 & g_2 & g_3 & \cdots & g_f \\ g_2 & g_3 & g_4 & \cdots & g_{f+1} \\ g_3 & g_4 & g_5 & \cdots & g_{f+2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ g_f & g_{f+1} & g_{f+2} & \cdots & g_{2f-1} \end{bmatrix} \begin{bmatrix} f_m & 0 & 0 & \ldots & 0 \\ f_{m-1} & f_m & 0 & \ldots & 0 \\ f_{m-2} & f_{m-1} & f_m & \ldots & 0 \\ \vdots & \vdots & \vdots & \ldots & \vdots \\ f_1 & f_2 & f_3 & \ldots & 0 \\ 1 & f_1 & f_2 & \ldots & f_m \\ 0 & 1 & f_1 & \ldots & f_{m-1} \\ 0 & 0 & 1 & \ldots & f_{m-2} \\ \vdots & \vdots & \vdots & \ldots & \vdots \\ 0 & 0 & 0 & \ldots & 1 \end{bmatrix} = 0$$

- $\mathcal{H}$ has rank $m \Rightarrow$ Nullspace of $\mathcal{H}$ has dim $f - m$
- Right hand factor has $f - m$ columns
- Right hand factor has full column rank
- Parametrization of null-space of $\mathcal{H}$!

Multi-step LS vs subspace identification

Estimate $f_1, \dots, f_m$ by solving

$$\begin{bmatrix} \hat{g}_1 & \hat{g}_2 & \hat{g}_3 & \dots & \hat{g}_f \\ \hat{g}_2 & \hat{g}_3 & \hat{g}_4 & \dots & \hat{g}_{f+1} \\ \hat{g}_3 & \hat{g}_4 & \hat{g}_5 & \dots & \hat{g}_{f+2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \hat{g}_f & \hat{g}_{f+1} & \hat{g}_{f+2} & \dots & \hat{g}_{2f-1} \end{bmatrix} \begin{bmatrix} f_m & 0 & 0 & \dots & 0 \\ f_{m-1} & f_m & 0 & \dots & 0 \\ f_{m-2} & f_{m-1} & f_m & \dots & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots \\ f_1 & f_2 & f_3 & \dots & 0 \\ 1 & f_1 & f_2 & \dots & f_m \\ 0 & 1 & f_1 & \dots & f_{m-1} \\ 0 & 0 & 1 & \dots & f_{m-2} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 0 & 0 & 0 & \dots & 1 \end{bmatrix} = 0$$

$\hat{g}$ noisy. $\Rightarrow$ Need to take statistics into account (c.f. subspace id)

Same problem as in subspace id? No!!!

In this case we can vectorize the system of equations $\Rightarrow$ WNSF!

Simpler to in a statistically efficient way estimate elements in the null-space than elements in the range space of a matrix

Multi-step LS vs subspace identification

Summary:

Method	Subspace id	Multi-step LS
Subspace	Range space	Null space
Weighting of	$\hat{\mathcal{H}}$	vec $\left\{\hat{\mathcal{H}}\right\}$
Estimation method	SVD+LS	LS
Can incorporate structural information	NO	YES
Consistency	YES	YES
Asymptotic efficiency	Special cases	YES

MIMO models

Matrix-Fraction Description (MFD) OE-MIMO:

$$y_t = F^{-1}(q)L(q)u_t + e_t$$

High order model: MIMO-FIR

$$y_t = G(q)u_t + e_t$$

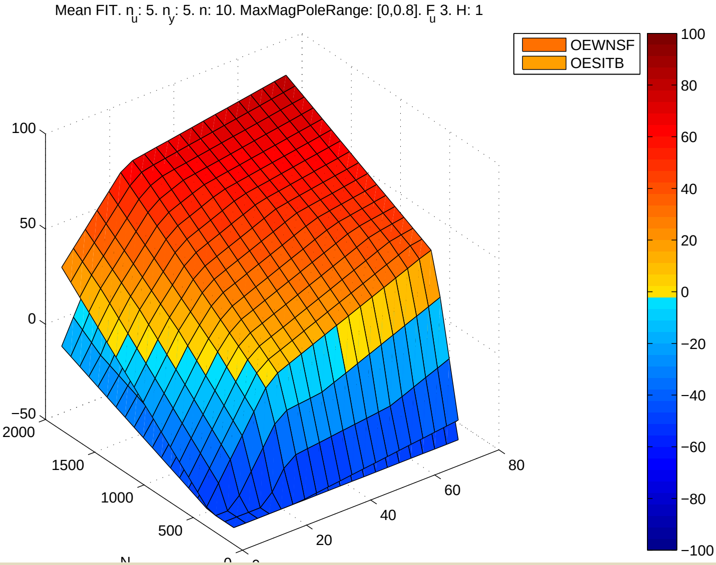
$$F^{-1}(q)L(q) = G(q) \;\Leftrightarrow F(q)G(q) - L(q) = 0$$

Same as in the SISO case!

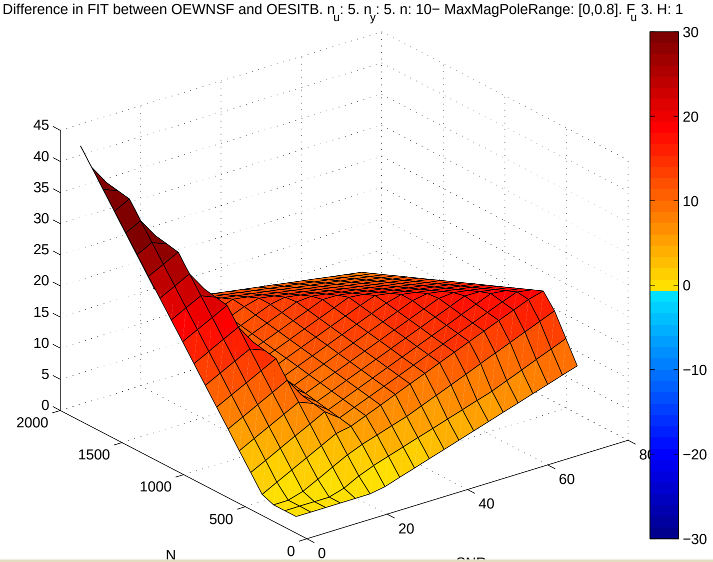
OE, ARMAX, BJ, MAX, ...

MFD, element-wise parameterizations

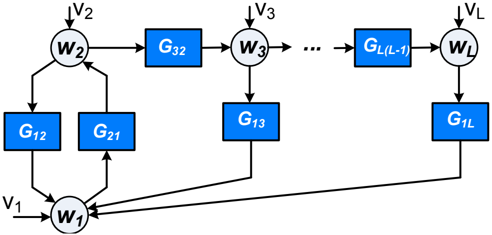
MIMO models: Output error with element wise parametrization



MIMO models: Output error with element wise parametrization



Dynamic Network Identification



$$w(t) = G(q)w(t) + R(q)r(t) + H(q)e(t)$$

Interconnection structure given by

$$G(q) = \begin{bmatrix} 0 & G_{12}(q) & G_{13}(q) \\ G_{21}(q) & 0 & G_{23}(q) \\ G_{31}(q) & G_{32}(q) & 0 \end{bmatrix}$$

Dynamic Network Identification

$$G(q) = \begin{bmatrix} 0 & G_{12}(q) & G_{13}(q) \\ G_{21}(q) & 0 & G_{23}(q) \\ G_{31}(q) & G_{32}(q) & 0 \end{bmatrix}$$

Suppose

$$G(q) = D^{-1}(q)N_G(q), \quad R(q) = D^{-1}(q)N_R(q), \quad H(q) = D^{-1}(q)N_H(q)$$

$$w = Gw + Rr + He \;\Leftrightarrow\; D(q)w(t) = N_G(q)w(t) + N_R(q)r(t) + N_H e(t)$$

which can be written

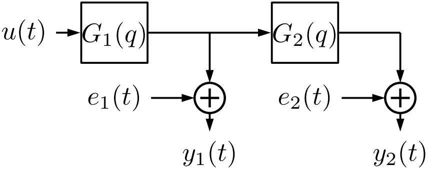
$$(D(q) - N_G(q))w(t) = N_R(q)r(t) + N_H e(t)$$

ARMA!

A range of structures can be accomodated for. For example

$D(q)$ diagonal: All transfer functions to one node have the same poles.

Cascade Networks



$$\begin{bmatrix} y_1(t) \\ y_2(t) \end{bmatrix} = \begin{bmatrix} \frac{L_1(q,\theta)}{F_1(q,\theta)} \\ \frac{L_2(q,\theta)}{F_2(q,\theta)} \frac{L_1(q,\theta)}{F_1(q,\theta)} \end{bmatrix} u(t) + e(t)$$

PEM: can be difficult...

Cascade Networks

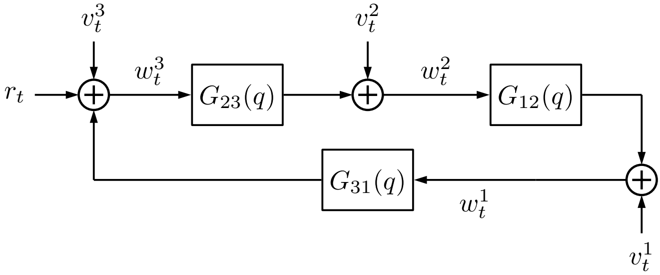
$$\begin{bmatrix} y_1(t) \\ y_2(t) \end{bmatrix} = \begin{bmatrix} \frac{L_1(q,\theta)}{F_1(q,\theta)} \\ \frac{L_2(q,\theta)}{F_2(q,\theta)} \frac{L_1(q,\theta)}{F_1(q,\theta)} \end{bmatrix} u(t) + e(t)$$

$$\approx \begin{bmatrix} \sum_{k=1}^{n} g_k^{(1)} q^{-k} \\ \sum_{k=1}^{n} g_k^{(21)} q^{-k} \end{bmatrix} u(t) + e(t) \qquad \Longrightarrow \qquad \hat{g}_k^{(1)}, \hat{g}_k^{(21)} \text{ (LS)}$$

$$\begin{cases} \frac{L_1(q,\theta)}{F_1(q,\theta)} = \bar{G}_1(q,g) \\ \frac{L_2(q,\theta)}{F_2(q,\theta)} \bar{G}_1(q,g) = \bar{G}_{21}(q,g) \end{cases} \Leftrightarrow \begin{cases} F_1(q,\theta)\bar{G}_1(q,g) - L_1(q,\theta) = 0 \\ F_2(q,\theta)\bar{G}_{21}(q,g) - L_2(q,\theta)\bar{G}_1(q,g) = 0 \end{cases}$$

WNSF can be applied with optimal asymptotic properties

Errors-in-Variables Problems in Dynamic Networks

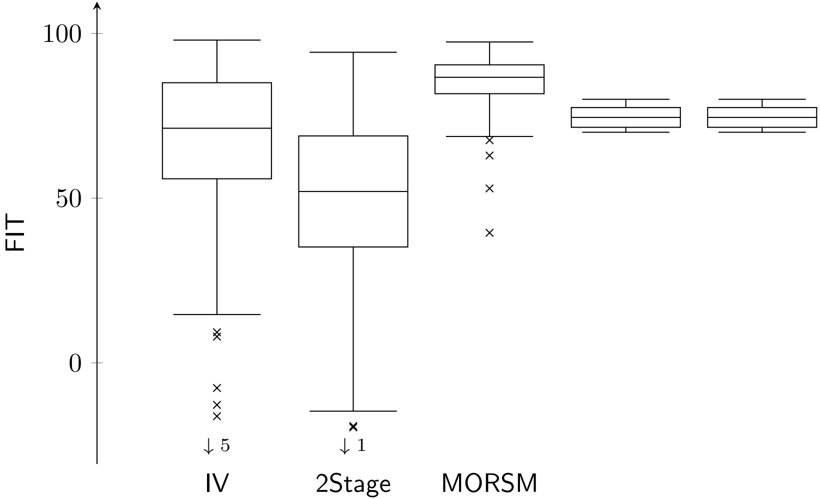


Measurements: $\tilde{w}_t = w_t + s_t$

Estimate G_{12}. Errors-in-variables problem!

- IV
- Two-stage methods
- WNSF, but requires more steps (no time for this, unfortunately)

Dynamic Networks: Simulation



Noise with High-order Dynamics

A low-order parametrization of the noise model...

- ...will not give asymptotic efficiency in open loop
- ...will not give consistency in closed loop

In Step 1, WNSF capture the noise with the non-parametric model

In Step 2, we may only compute a parametric model of the plant!

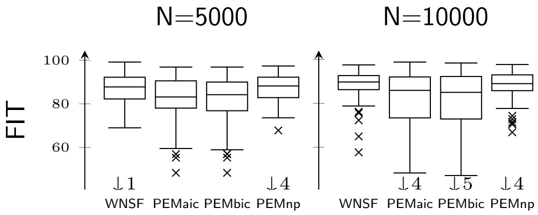
$$H(q,\theta) = \frac{1}{A(q,\eta)} \qquad G(q,\theta) = \frac{B(q,\eta)}{A(q,\eta)}$$

Noise with High-order Dynamics

True noise model given by very long FIR without a low-order parametrization.

Noise models:

- WNSF and PEMnp use a noise model $H(q,a) = 1/[1+\sum_{k=1}^{70} a_k q^{-k}]$.
- PEMaic and PEMbic use a noise model $H(q;c,d) = [1+\sum_{k=1}^{m} c_k q^{-k}]/[1+\sum_{k=1}^{m} d_k q^{-k}]$, $m = \{1, ..., 30\}$, with m decided with AIC/BIC.



Average computational times [s]

N	5000	10000
WNSF	0.907	1.29
$\text{PEM}_{\text{aic,bic}}$	26.8	38.1
PEM_{np}	133	236

Online Identification

Recursive PEM:

$$\hat{\theta}_t = \hat{\theta}_{t-1} + \frac{1}{t}\Lambda_t^{-1}\psi_t(\hat{\theta}_{t-1})\varepsilon_t(\hat{\theta}_{t-1})$$

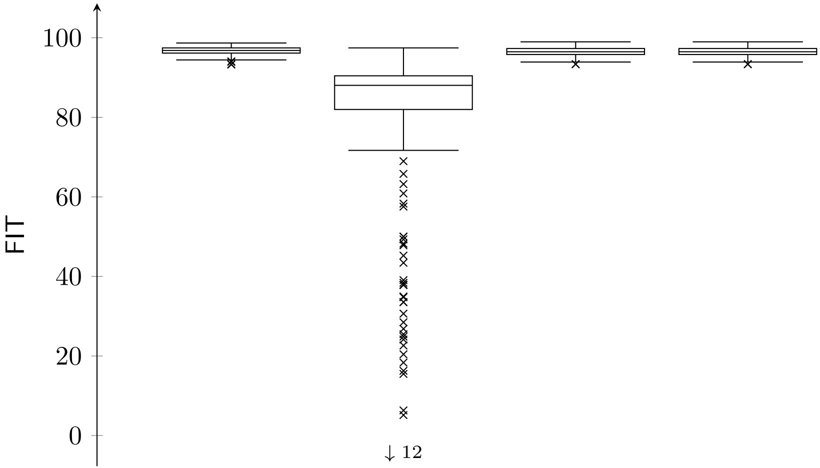
The gradient $\psi_t(\theta)$ and the prediction error $\varepsilon_t(\theta)$ cannot be computed with fixed-size memory $\Longrightarrow$ approximations!

Recursive WNSF:

- ARX model can be computed recursively
- Parametric estimate update identical to offline method

$$\hat{\theta}_t = \left[Q^\top(\hat{\eta}_t^n)W(\hat{\theta}_{t-1})Q(\hat{\eta}_t^n)\right]^{-1} Q^\top(\hat{\eta}_t^n)W(\hat{\theta}_{t-1})\hat{\eta}_t^n \quad (1)$$

Online Identification



Nonlinear Models

Rational in parameters models:

$$y(t) = \frac{f(\varphi(t))\theta}{1 + g(\varphi(t))\theta} + e(t)$$

where $\varphi(t)$ function of past inputs and outputs.

Multi-step LS procedure:

i) High order expansion: $y(t) = \sum_{k=0}^{\infty} \alpha_k \gamma_k(\varphi(t))$
for example Taylor expansion

ii) Truncation and LS-estimation: $\hat{\alpha} = \text{argmin}_\alpha \sum_{t=1}^{N} \left(y(t) - \sum_{k=1}^{n} \alpha_k \gamma_k(\varphi(t))\right)^2$

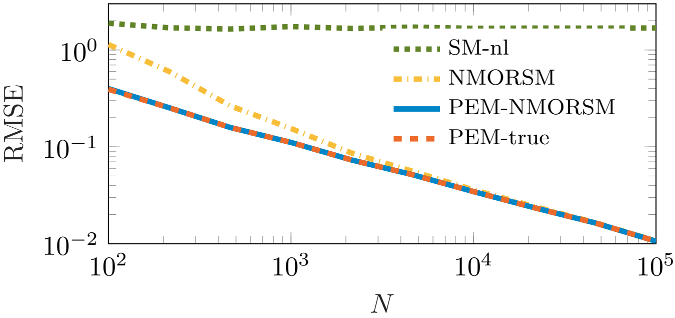
ii) Simulated model output: $\hat{y}(t) := \sum_{k=1}^{n} \hat{\alpha}_k \gamma_k(\varphi(t))$

iii) Residual estimation: $\hat{y}(t) \approx \frac{f(\varphi(t))\theta}{1+g(\varphi(t))\theta}$

iv) Multi-step LS: $\hat{\theta}_{k+1} = \text{argmin}_\theta \sum_{t=1}^{N} \left(\frac{\hat{y}(t)(1+g(\varphi(t))\theta) - f(\varphi(t))\theta}{1+g(\varphi(t))\hat{\theta}_k}\right)^2$

Nonlinear Models

$$y(t) = \frac{\theta_1 u(t-1)}{1+\theta_2 u^2(t)} + e(t) = \sum_{k=0}^{\infty} \alpha_k u(t-1) u^{2k}(t)$$



Conclusions

- PEM may require very accurate initial conditions to converge to the global minimum
  - Systems with several resonance peaks
  - Systems with widely spread eigenvalues
- Multi-step high order LS may be appropriate to handle these scenarios:
  - Less sensitive to the effect of the initial condition
  - Faster convergence
  - Asymptotically efficient
- Other scenarios where multi-step high-order LS may be useful:
  - MIMO
  - Online identification
  - Dynamic networks
  - Overparametrized models
  - Non-parametric noise spectra
  - Non-linear models