System identification in a historical perspective

Michel Gevers
UCL, Louvain la Neuve, Belgium

DISC Summer School on
Data-Driven Modelling for Control
16 June, 2014

Updated version of
M. Gevers, «A personal view of the development of system identification ...», IEEE Control Systems Magazine, 2006

Outline

- The origins of system identification
- 1965: the birth of identification in engineering
- The parametric Prediction Error framework
- Search for the model structure: key stumbling block
- Major challenge today: reduce the cost of identification
- Non-parametric approaches
- Reducing the cost of the experiment
- Application-oriented experiment design
- Present research areas
- Conclusions

The origins of system identification

- The early work in system identification was developed by the statistics and time series communities

- It has its roots in the work of Gauss (1809) and Fisher (1912) and the theory of stochastic processes

- See Deistler (2002) for an excellent survey of this work

Start the of the model-based control era

1960: Kalman's key papers - start of the state-space era
(Kalman:1960a,b)

$$\begin{cases} x_{t+1} &= Ax_t + Bu_t \\ y_t &= Cx_t \end{cases}$$

- Development of a model-based theory for prediction, filtering and control
- Kalman filter replaces Wiener filter
- Pole placement and LQG control
- Applications initially in areas where good models are available (aerospace, mechanical, electrical systems)

Growing pressure to apply these modern techniques to areas where models are not available from physics

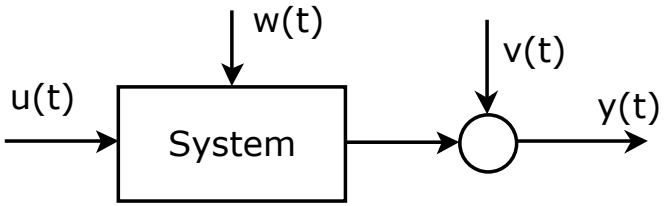
$\longrightarrow$ Need for system identification

Identification in a nutshell

What is system identification ?

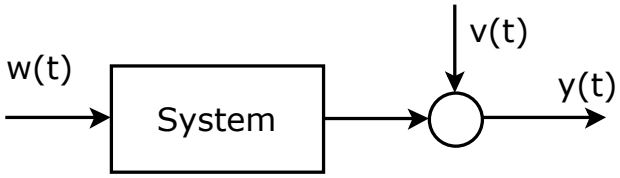
Identification is the task of constructing a dynamical model that can predict the outputs of a dynamical system:

Either driven by input and noise:



Predict y(t) from past u(), y() data

Or driven by noise only:



Predict y(t) from past y() data only

What kind of dynamical models ?

State-space models

with measured input:

$$\left\{ \begin{array}{lcl} x_{t+1} & = & Ax_t + Bu_t + Ke_t \\ y_t & = & Cx_t + e_t \end{array} \right.$$

without measured input:

$$\left\{ \begin{array}{lcl} x_{t+1} & = & Ax_t + Ke_t \\ y_t & = & Cx_t + e_t \end{array} \right.$$

Input/output models

with measured input:

$$\begin{aligned} & y_t + a_1 y_{t-1} + \ldots + a_n y_{t-n} \\ & \quad = b_1 u_{t-1} + \ldots + u_{t-m} \\ & \quad + e_t + c_1 e_{t-1} + \ldots + c_p e_{t-p} \end{aligned}$$

without measured input:

$$\begin{aligned} & y_t + a_1 y_{t-1} + \ldots + a_n y_{t-n} \\ & \quad = e_t + c_1 e_{t-1} + \ldots + c_p e_{t-p} \end{aligned}$$

Outline

- The origins of system identification
- 1965: the birth of identification in engineering
- The parametric Prediction Error framework
- Search for the model structure: key stumbling block
- Major challenge today: reduce the cost of identification
- Non-parametric approaches
- Reducing the cost of the experiment
- Application-oriented experiment design
- Present research areas
- Conclusions

Identification in the engineering world:

The two milestone papers of 1965

- Ho and Kalman (1965), `Effective construction of linear state-variable models from input-output data', *Regelungstechnik,* Vol. 12, pp. 545-548.

  Gave birth to realization theory ⟶ subspace identification

- Aström and Bohlin (1965), `Numerical identification of linear dynamic systems from normal operating records', *Proc. IFAC Symp on Self Adaptive Systems, Teddington, UK.*

  Gave birth to Prediction Error Identification (PEI)

Realization theory

How to go from an infinite description of a system

$$H(z) = \sum_{k=1}^{\infty} H_k z^{-k}$$

to a finite description :

$$H(z) = C(zI - A)^{-1}B$$

$\{H_1, H_2, H_3, \ldots\} \quad \Longrightarrow \{A, B, C\}$ with $A \in \mathbb{R}^{n\times n}, B \in \mathbb{R}^{n\times m}, C \in \mathbb{R}^{p\times n}$

Two aspects :
- find the McMillan degree of $H(z) \Longrightarrow$ **dim** A
- find the parameters of A, B, C.

$$y_t = \sum_{k=1}^{\infty} H_k u_{t-k} \quad \Longleftrightarrow \quad \left\{ \begin{array}{lcl} x_{t+1} & = & Ax_t + Bu_t \\ y_t & = & Cx_t \end{array} \right.$$

Key tool: the Hankel matrix

$$\mathcal{H} = \begin{bmatrix} H_1 & H_2 & H_3 & H_4 & \dots \\ H_2 & H_3 & H_4 & H_5 & \dots \\ H_3 & H_4 & H_5 & H_6 & \dots \\ \vdots & \vdots & \vdots & \vdots & \end{bmatrix} = \begin{bmatrix} C \\ CA \\ CA^2 \\ \vdots \end{bmatrix} \begin{bmatrix} B & AB & A^2B & \dots \end{bmatrix}$$

If the McMillan degree of $H(z)$ is n, then

1. rank $\mathcal{H} = n$
2. $\exists\ A, B, C$ such that $H_k = CA^{k-1}B$ with $A \in \mathbb{R}^{n \times n}, B \in \mathbb{R}^{n \times m}, C \in \mathbb{R}^{p \times n}$

Stochastic realization theory (1970-1975)

(Akaike, 1974)

Combines realization theory and innovations theory

Given the covariance sequence $\{R_0, R_1, R_2, \ldots\}$ of a zero-mean stochastic process $\{y_t\}$, where $R_k \triangleq E\{y_t y_{t-k}^T\}$, find a minimal Markovian representation for $\{y_t\}$:

$$\begin{cases} x_{t+1} &= Ax_t + K\varepsilon_t \\ y_t &= Cx_t + \varepsilon_t \end{cases} \tag{1}$$

where ε_t is a zero-mean white noise sequence, i.e. find the state-space matrices $\{A, K, C\}$ with $n = dim(A)$ minimal, such that the covariance of the output of (1) is exactly R_k.

Solution based on the Hankel matrix formed from the R_k.

Basis for subspace identification, developed much later

ARMAX, Likelihood function and the Prediction Error framework

Åström and Bohlin (1965)

- Input–output formulation : the ARMAX model structure

$$A(z^{-1})y_t = B(z^{-1})u_t + \lambda C(z^{-1})e_t, \qquad \{e_t\} : i.i.d.$$

- Prediction errors $\{\varepsilon_t\}$: $\quad C(z^{-1})\varepsilon_t = A(z^{-1})y_t - B(z^{-1})u_t$
- Likelihood function for a Gaussian p.d.f.:

$$L(\theta) = -\frac{1}{2\lambda^2}\sum_{t=1}^{N}\varepsilon_t^2(\theta) - N\log\lambda + \text{constant}$$

$$\max_\theta L(\theta) \Longleftrightarrow \min_\theta \frac{1}{2}\sum_{t=1}^{N}\varepsilon_t^2(\theta) = \min_\theta V_N(\theta)$$

$$\hat\theta = \arg\min_\theta V_N(\theta), \qquad \hat\lambda^2 = \frac{2}{N}V_N(\hat\theta)$$

- $\Rightarrow$ adopt $V(\theta) = \frac{1}{2}\sum_{t=1}^{N}\varepsilon_t^2(\theta)$ as a reasonable criterion even in the absence of a Gaussian probability function: Prediction Error framework; originally suggested by Gauss (1809): see Åström (1980).

State of the art around 1975

- Two fundamentally different approaches:
  - **State-space model: by Hankel matrix factorization.** Projection methods. Easy but not optimal. No need for parametrization.
  - **I/O model : by minimization of PE criterion**. Slower but optimal. Requires choice of model structure. Allows characterization of variance errors through Fisher information matrix.
- After 1975 the parametric (prediction error) approach took over the field.
- Main driving force behind the success of PE methods: Lennart Ljung, and software development (with increased computation speed)
- Hankel matrix approach re-emerged after 1990: subspace state-space methods.

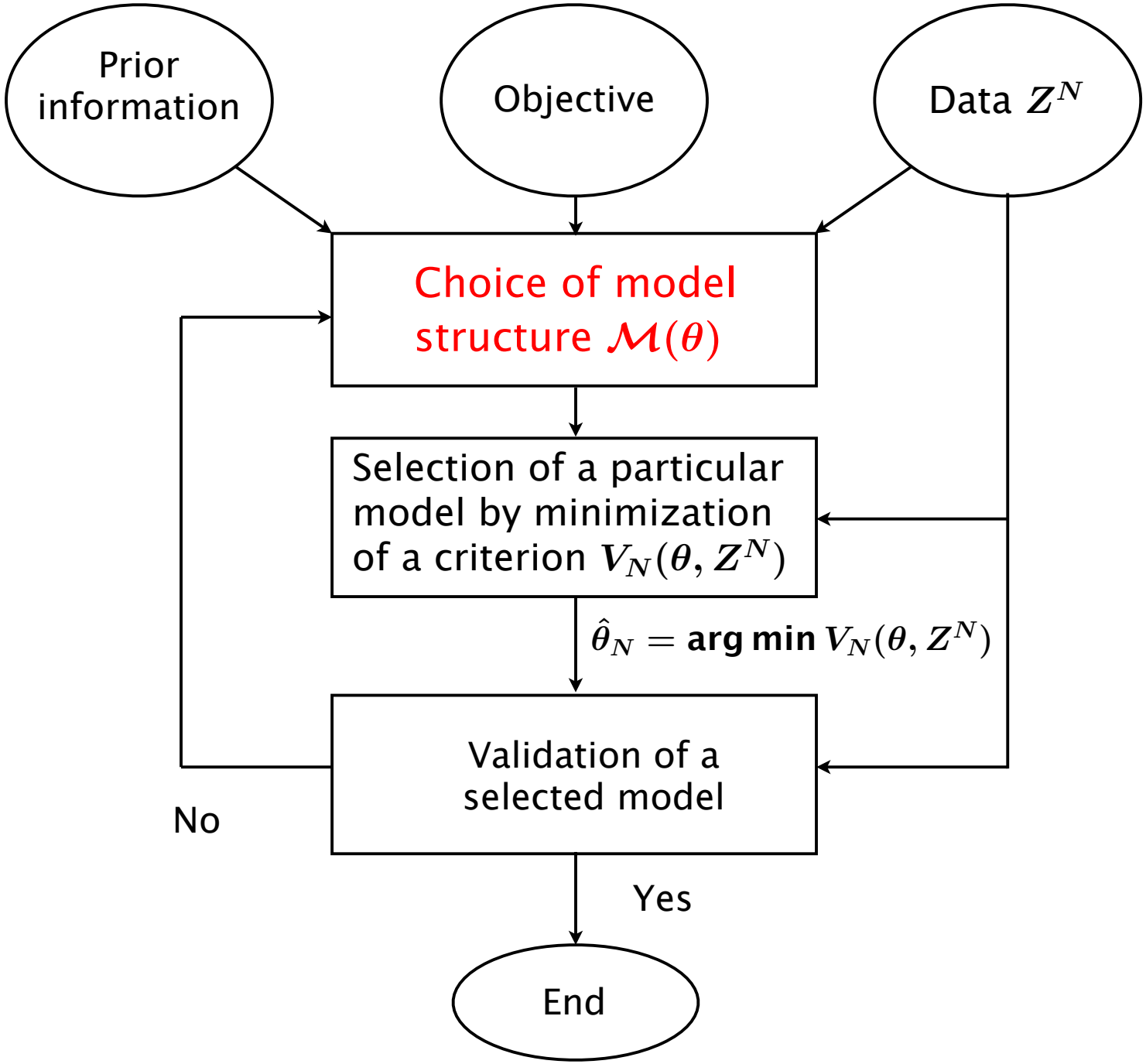
Outline

- The origins of system identification
- 1965: the birth of identification in engineering
- The parametric Prediction Error framework
- Search for the model structure: key stumbling block
- Major challenge today: reduce the cost of identification
- Non-parametric approaches
- Reducing the cost of the experiment
- Application-oriented experiment design
- Present research areas
- Conclusions

Prediction Error Identification in one slide

- True system $\mathcal{S}$: $y_t = G_0(z)u_t + \underbrace{H_0(z)e_t}_{v_t}$
- Model set $\mathcal{M}$: $y_t = G(z,\theta)u_t + H(z,\theta)\epsilon_t$
- Model $\Leftrightarrow$ predictor: $\hat{y}_{t|t-1}(\theta) = H^{-1}(z,\theta)G(z,\theta)u_t + (1 - H^{-1}(z,\theta))y_t$
- Prediction error: $\varepsilon_t(\theta) = y_t - \hat{y}_{t|t-1}(\theta)$
- Criterion $V_N(\theta, Z^N) = \frac{1}{N}\sum_{t=1}^{N} ||\varepsilon_t(\theta)||^2$
- The parameter estimate: $\hat{\theta}_N = \textbf{arg min}_{\theta\in\mathcal{D}_\theta} V_N(\theta, Z^N)$
- Transfer function estimates: $G(z,\hat{\theta}_N), \quad H(z,\hat{\theta}_N)$
- If $\exists\theta_0$: $\quad G(z,\theta_0) = G_0(z)$ and $H(z,\theta_0) = H_0(z)$, then:

$$\sqrt{N}(\hat{\theta}_N - \theta_0) \overset{N\to\infty}{\longrightarrow} N(0, P_\theta)$$


Prior information
Objective
Data Z^N
Choice of model structure $\mathcal{M}(\theta)$
Selection of a particular model by minimization of a criterion $V_N(\theta, Z^N)$
$\hat{\theta}_N = \arg\min V_N(\theta, Z^N)$
Validation of a selected model
No
Yes
End

Key new concept: identification as a design problem (1986)

(Gevers&Ljung,1986; Wahlberg&Ljung,1986)

Basic observation:

every model is at best an approximation of the exact system

$\longrightarrow$ it contains errors

Two aspects:

- Model structure and criterion should be tuned towards the application of the model
- Experiment should be tuned towards the application criterion

Prediction Error Identification is well suited to this design view

Example

When a system $y_t = G_0(z)u_t + v_t$ is modeled by a model structure $y_t = G(z,\theta)u_t + \epsilon_t$, the PE criterion minimizes

$$\hat{\theta}_N \rightarrow \textbf{arg min}\, V(\theta) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \left\{ \mid G(e^{j\omega},\theta) - G_0(e^{j\omega}) \mid^2 \Phi_u(\omega) + \Phi_v(\omega) \right\} d\omega$$

$\Longrightarrow$ If the application is to have a good model in a narrow bandwidth defined by a bandpass filter $W(e^{j\omega})$, then open loop identification with an input spectrum $\Phi_u(\omega) = |W(e^{j\omega})|^2 \Phi_v(\omega)$ will meet that objective.

1965 - 2000: what had been accomplished?

- Both approaches have been developed:
  - ⋆ Parametric Prediction Error approach
  - ⋆ Non-parametric state-space (subspace) approach
- A frequency domain approach has been developed, showing advantages of periodic excitation
- Theory is for the most part well understood
- Field of applications have been enormously widened
- Model Predictive Control has become the standard for most control applications
- Progress in nonlinear system identification

Bottlenecks and limitations

- Modeling and identification is still the most costly part of any advanced control design
- Model building accounts for 50% to 75% of total cost in an advanced control project
- Major bottleneck: search for the best model structure
- Prediction error criterion is non-convex: problem of local minima
- Identification of MIMO systems is still a difficult task
- Nonlinear identification is still in its infancy
- Identification of structured systems (e.g. distributed and networked control systems)

Outline

- The origins of system identification
- 1965: the birth of identification in engineering
- The parametric Prediction Error framework
- Search for the model structure: key stumbling block
- Major challenge today: reduce the cost of identification
- Non-parametric approaches
- Reducing the cost of the experiment
- Application-oriented experiment design
- Present research areas
- Conclusions

Key stumbling block in PE identification: choice of model structure

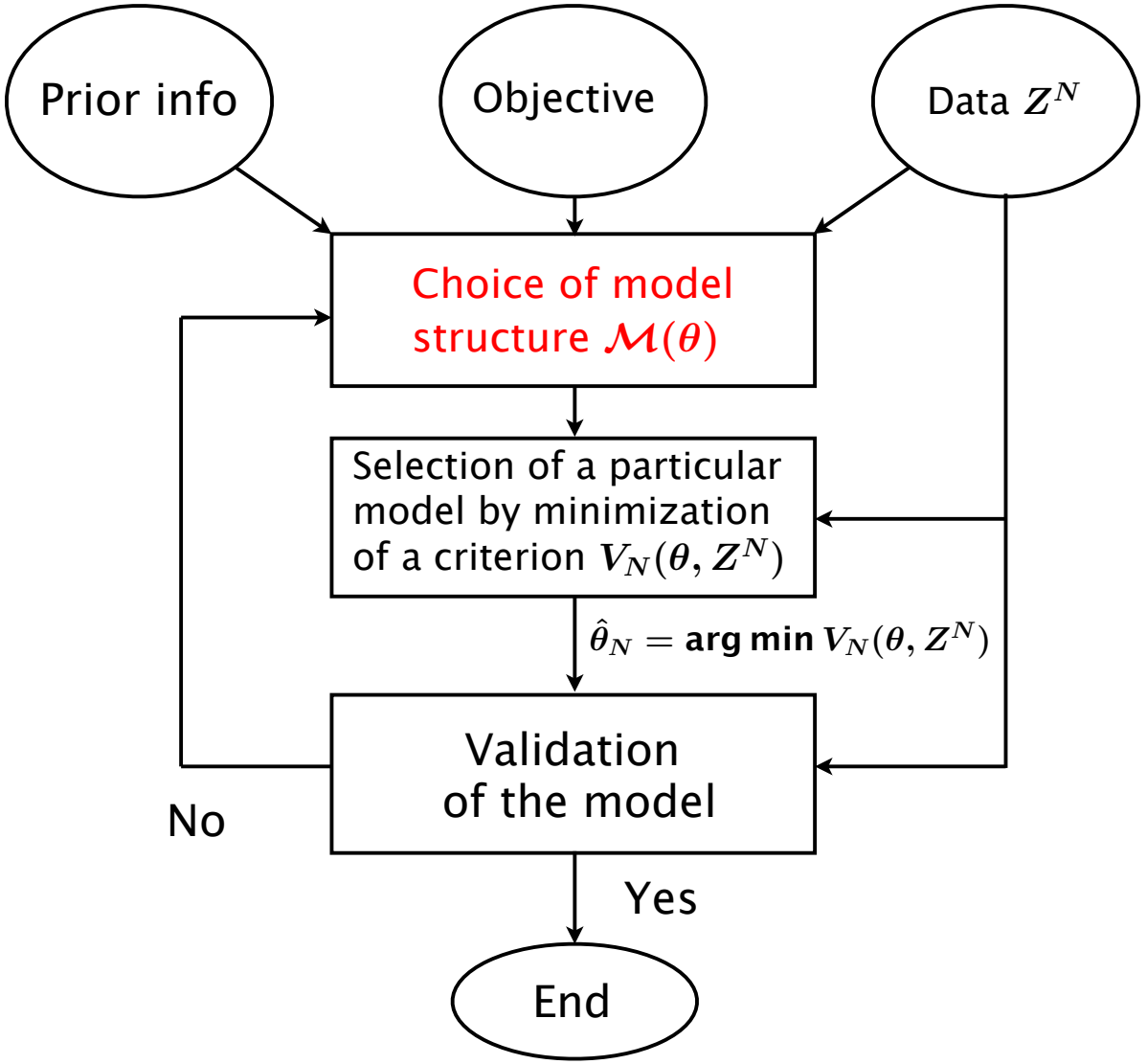


Illustration (courtesy Pintelon, Schoukens, Ljung)

7000 data are generated by a «true» BJ system:

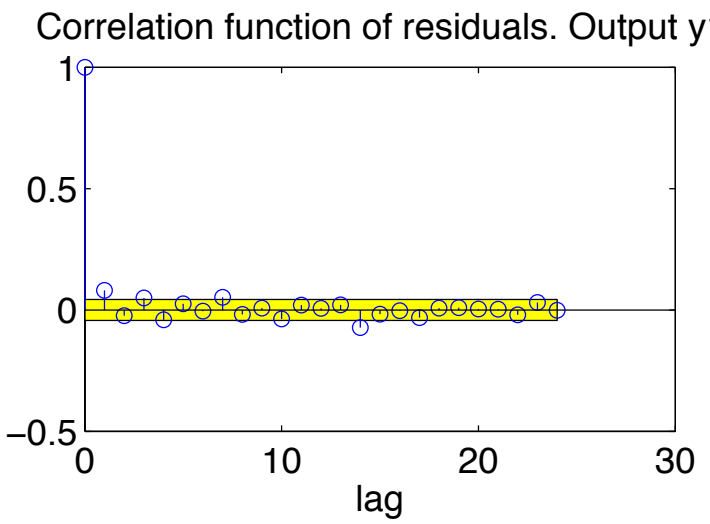
$$y_t = \frac{0.0947 + 0.2463z^{-1} + 0.0947z^{-2}}{1 - 0.5376z^{-1} + 0.7357z^{-2}}u_t + \frac{1 + 0.776z^{-1}}{1 - 0.1584z^{-1}}e_t$$

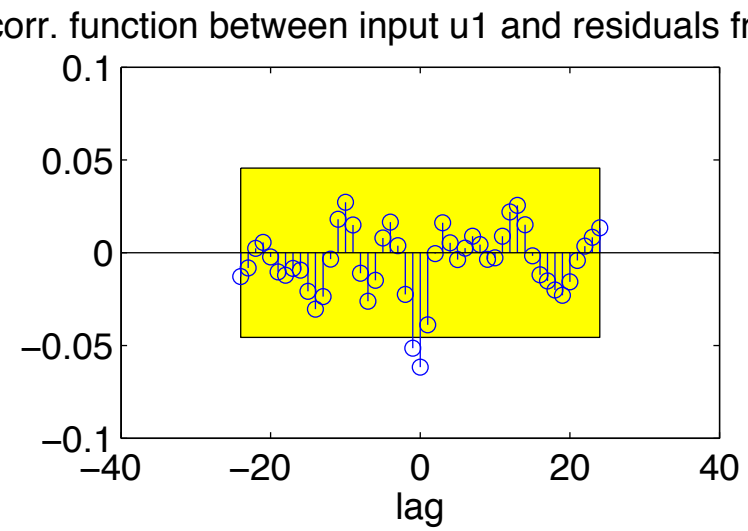
Apply the standard procedure of the Matlab SI toolbox:

- Split data set into two (estimation and validation data set)
- Estimate delay: d=0
- Initial guess with N4SID: yields state-space model m3 of order 3
- Compare with state-space models of order 1 and 2: m1 and m2
- Fits on validated data:
  - ➡ m1: 57.6 % fit
  - ➡ m2: 86.1 % fit
  - ➡ m3: 87.1 % fit

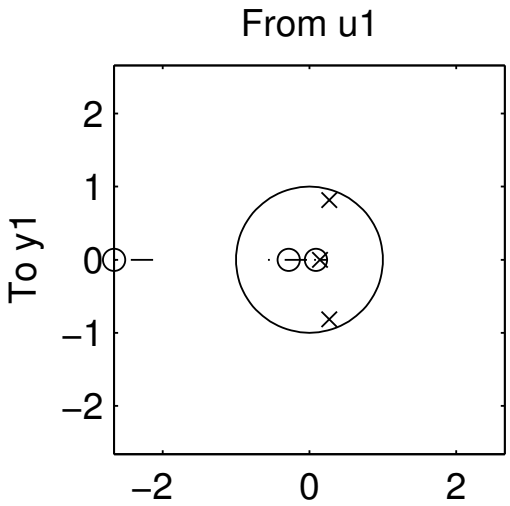
Fit = part of the output variation that is explained by the model

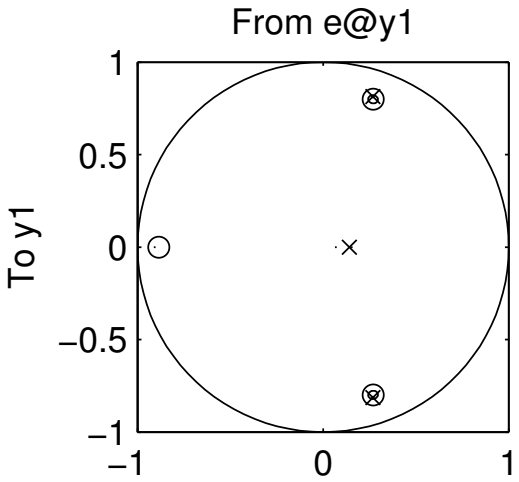
Analysis of model m3:





Poles and zeros of $\widehat{G}$ and $\widehat{H}$:



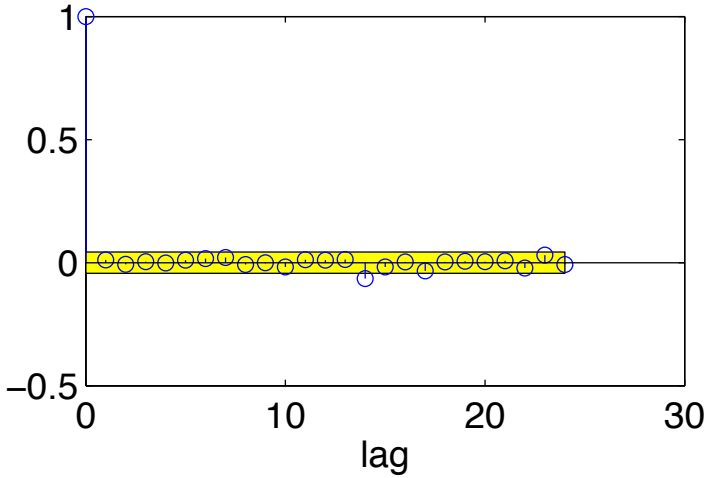


Conclusion of the analysis

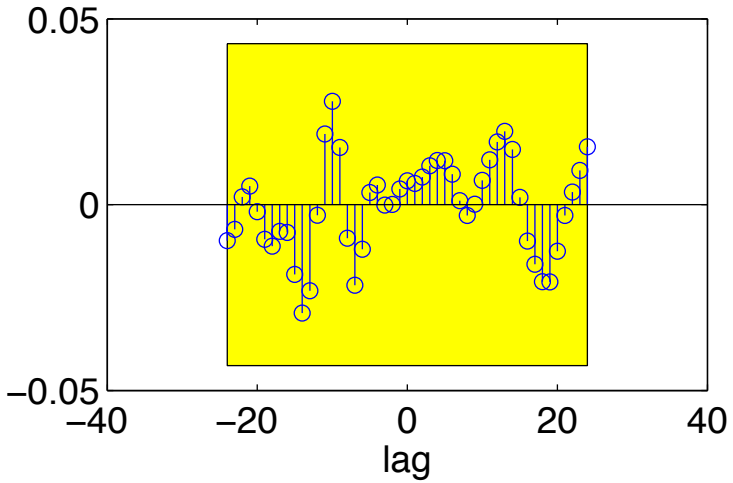
- Common poles and zeros in I/O and noise model
- Eliminate common poles and zeroes and try BJ model of order 2
- This yields an estimated model with fit 87.2%

$$y_t = \frac{0.0902 + 0.2493z^{-1} + 0.0912z^{-2}}{1 - 0.5377z^{-1} + 0.7357z^{-2}}u_t + \frac{1 + 0.9726z^{-1}}{1 - 0.1579z^{-1}}e_t$$

Correlation function of residuals. Output y1



Cross corr. function between input u1 and residuals from output y1



Outline

- The origins of system identification
- 1965: the birth of identification in engineering
- The parametric Prediction Error framework
- Search for the model structure: key stumbling block
- Major challenge today: reduce the cost of identification
- Non-parametric approaches
- Reducing the cost of the experiment
- Application-oriented experiment design
- Present research areas
- Conclusions

Major goal of present research: reducing the cost of the identification

What is the cost of modeling and identification ?

1. Man-hour costs

   → Reduce human intervention, particularly expert intervention. User-friendly identification.

2. Cost of the experiment

   → Reduce experiment time and performance degradation during data collection.

   Optimal experiment design.

3. Cost of estimating useless system properties

   → Cost of complexity.

   Application-oriented experiment design

Outline

- The origins of system identification
- 1965: the birth of identification in engineering
- The parametric Prediction Error framework
- Search for the model structure: key stumbling block
- Major challenge today: reduce the cost of identification
- Non-parametric approaches
- Reducing the cost of the experiment
- Application-oriented experiment design
- Present research areas
- Conclusions

Steps towards user-friendly identification

Main idea:

- first compute a nonparametric estimate of the system and noise distribution
- use these as a starting point for more efficient parametric PE identification

Two possible strategies:

- state-space subspace identification
  - ★ CVA (Larimore, 1990, 1996)
  - ★ N4SID (Van Overschee and De Moor, 1994)
  - ⋆ MOESP (Verhaegen, 1994)
- nonparametric frequency domain identification
  (VUB group: Pintelon, Schoukens, Rolain, Vandersteen, et al.)

Key step in the subspace methods

- An estimate of the Hankel matrix of impulse responses is obtained from I/O data.
- Recall the basic tool $y(t) = \sum_{k=1}^{\infty} H_k \, u(t-k)$

$$\mathcal{H} = \begin{bmatrix} H_1 & H_2 & H_3 & H_4 & \dots \\ H_2 & H_3 & H_4 & H_5 & \dots \\ H_3 & H_4 & H_5 & H_6 & \dots \\ \vdots & \vdots & \vdots & \vdots & \end{bmatrix} = \begin{bmatrix} C \\ CA \\ CA^2 \\ \vdots \end{bmatrix} \begin{bmatrix} B & AB & A^2B & \dots \end{bmatrix}$$

- Perform an SVD of the Hankel matrix
- Truncate by deleting all singular values below some threshold
- This fixes the order of the state space system, i.e. size of A
- A second LS step is needed to compute the remaining matrices

Estimation of the Frequency Response Function $\widehat{G}(e^{j\omega})$

Time versus frequency domain identification

System and noise model:

$$y(t) = G_0(z)u(t) + v(t) = G_0(z)u(t) + H_0(z)e(t)$$

Time domain PE criterion:

$$V_N(\theta) = \frac{1}{N}\sum_{t=1}^{N}\left(H^{-1}(z,\theta)[y(t) - G(z,\theta)u(t)]\right)^2$$

Frequency domain PE criterion:

$$V_N(\theta) = \frac{1}{N}\sum_{k=-\frac{N}{2}+1}^{\frac{N}{2}}\left(H^{-1}(k,\theta)[Y(k) - G(k,\theta)U(k)]\right)^2$$

where $G(k,\theta) = G(e^{j2\pi k/N})$, $H(k,\theta) = H(e^{j2\pi k/N})$ and

$X(k) = \frac{1}{\sqrt{N}}\sum_{t=0}^{N-1} x(t)e^{-j2\pi kt/N}$ = DFT of $x(t)$

Precise frequency domain formulation

For a finite record $t = 0, \ldots, N-1$:

$$y(t) = G_0(q)u(t) + t_G(t) + H_0(q)e(t) + t_H(t).$$

Frequency domain formulation:

$$\begin{aligned} Y(k) &= G_0(\Omega_k)U(k) + T_G(\Omega_k) + H_0(\Omega_k)E(k) + T_H(\Omega_k) \\ &= G_0(\Omega_k)U(k) + H_0(\Omega_k)E(k) + \underbrace{T(\Omega_k)}_{\text{leakage term}} \end{aligned}$$

Frequency domain criterion becomes:

$$V_N(\theta) = \frac{1}{N} \sum_{k=-\frac{N}{2}+1}^{\frac{N}{2}} \left(H_k^{-1}(\theta)[Y(k) - G_k(\theta)U(k) - T_k(\theta)] \right)^2$$

where $G_k(\theta) \triangleq G(\Omega_k, \theta)$, $H_k(\theta) \triangleq H(\Omega_k, \theta)$, $T_k(\theta) \triangleq T(\Omega_k)$

The Local Polynomial Method (LPM)

(VUB group: Schoukens, Pintelon, Barbé, Rolain, Vandersteen, 2009)

$$Y(k) = G_0(\Omega_k)U(k) + H_0(\Omega_k)E(k) + \underbrace{T(\Omega_k)}_{\text{leakage term}}$$

Observation: $G_0(\Omega_k)$ and $T(\Omega_k)$ are smooth rational functions.

$$\begin{aligned} G_0(\Omega_{k+r}) &\approx G_0(\Omega_k) + \sum_{s=1}^{R} g_s(k) r^s, \\ T_0(\Omega_{k+r}) &\approx T_0(\Omega_k) + \sum_{s=1}^{R} t_s(k) r^s, \quad \text{for } r = 0, \pm 1, \ldots, \pm n \end{aligned}$$

Define

$$\theta_k \triangleq [G_0(\Omega_k)\ g_1(k) \ldots\ g_R(k); T_0(\Omega_k)\ t_1(k) \ldots\ t_R(k)]^T$$

We can estimate $\theta_k, \quad k = 1, \ldots, N$ from local measurements

$$\{U(k-n), \ldots, U(k), \ldots, U(k+n);\ \ Y(k-n), \ldots, Y(k), \ldots, Y(k+n)\}$$

over a local frequency window of size $2n$ around Ω_k.

Benefits of the Local Polynomial Method (LPM)

- Excellent nonparametric estimate of $\widehat{G}(\Omega_k)$
- Plus nonparametric estimate of $\widehat{T}(\Omega_k)$, and of $\widehat{\Phi}_v(\Omega_k)$ from residuals
- Requires little or no expert user intervention
- Both are very useful starting points for parametric estimates
  - $\widehat{G}(\Omega)$ gives good idea of what $G_0(\Omega)$ looks like
  - $\widehat{\Phi}_v(\Omega_k)$ can be used to simplify the criterion for estimation of $G(\Omega, \theta)$:

$$V_N(\theta) = \frac{1}{N} \sum_{k=-\frac{N}{2}+1}^{\frac{N}{2}} \frac{\left[Y(k) - G_k(\theta)U(k) - \widehat{T}(\Omega_k)\right]^2}{\widehat{\Phi}_v(\Omega_k)}$$

- Risk of local minima with the latter criterion is greatly reduced (Schoukens et al; CDC-ECC2011)

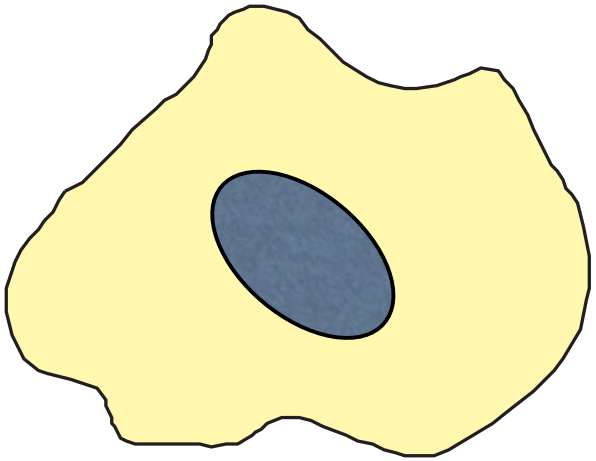
Outline

- The origins of system identification
- 1965: the birth of identification in engineering
- The parametric Prediction Error framework
- Search for the model structure: key stumbling block
- Major challenge today: reduce the cost of identification
- Non-parametric approaches
- Reducing the cost of the experiment
- Application-oriented experiment design
- Present research areas
- Conclusions

Least costly identification for robust control

Classical design:

Minimize a measure of the uncertainty subject to constraints

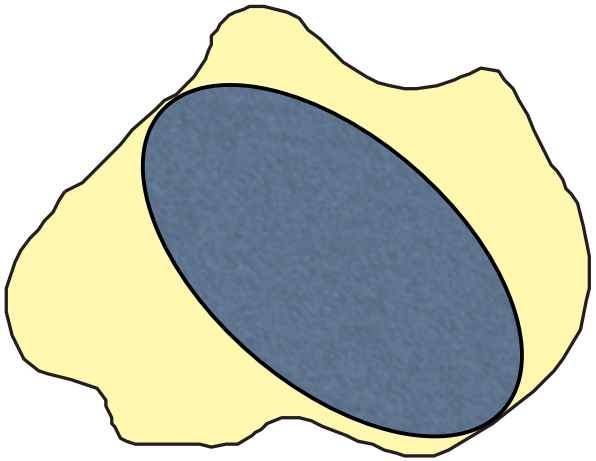


Dual idea:

Minimize the identification cost subject to a required quality

Cheapest cost:

Can be the length of the experiment, the applied signal energy, or the perturbation cost.



Bombois, Scorletti, Gevers, Van den Hof, Hildebrand, 2004–2006

The dual approach to experiment design

Illustration with open loop design

Classical approach:

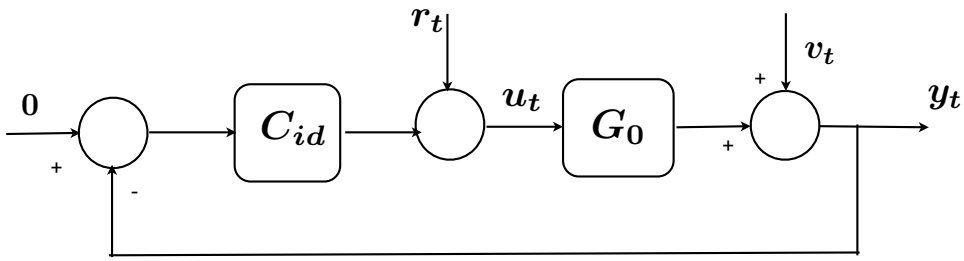
$$\left.\begin{array}{l} \min_{\Phi_u(\omega)} det(P_\theta) \quad \text{or} \\ \min_{\Phi_u(\omega)} \int_{-\pi}^{\pi} Var\{G(e^{j\omega}, \hat{\theta}_N)\} d\omega \end{array}\right\} \text{ subject to } \int_{-\pi}^{\pi} \Phi_u(\omega) d\omega \leq \alpha$$

Dual approach:

$$\min_{\Phi_u(\omega)} \int_{-\pi}^{\pi} \Phi_u(\omega) d\omega \text{ subject to } \left\{ \begin{array}{l} \int_{-\pi}^{\pi} Var\{G(e^{j\omega}, \hat{\theta}_N)\} \leq \gamma \quad \text{or} \\ \int_{-\pi}^{\pi} Var\{C(G(e^{j\omega}, \hat{\theta}_N))\} \leq \gamma \end{array} \right.$$

Application to least costly experiment design

Disturbance rejection problem



Normal operation:

$$\left.\begin{array}{l} y_t = G_0(z)u_t + v_t \\ u_t = -C(z)y_t \end{array}\right\} \Leftrightarrow y_t = \frac{1}{1+C(z)G_0(z)}v_t = S(z)v_t$$

During identification, an excitation signal r_t is added:

$$\begin{aligned} y_t &= S(z)v_t + \underbrace{G_0(z)S(z)r_t}_{y^r} \\ u_t &= -C(z)S(z)v_t + \underbrace{S(z)r_t}_{u^r} \end{aligned}$$

$\Longrightarrow$ Performance degradation cost due to y^r and u^r:

$$\begin{aligned} \mathcal{J}_r &= \frac{N}{2\pi}\int_{-\pi}^{\pi} (\alpha_y \Phi_{y_r} + \alpha_u \Phi_{u_r})\, d\omega \\ &= \frac{N}{2\pi}\int_{-\pi}^{\pi} \left(\alpha_y |G_0(e^{j\omega})S(e^{j\omega})|^2 + \alpha_u |S(e^{j\omega})|^2\right) \Phi_r(\omega)\, d\omega \end{aligned}$$

$$\min_{N,\Phi_r(\omega)} \mathcal{J}_r \text{ subject to } \left\| \frac{W(z)}{1+C(G(\theta))G(\theta)} \right\|_\infty \leq \gamma \quad \forall \theta \in \mathcal{D}(\hat{\theta}_N, P_\theta)$$

But is it all worth the effort ?

Illustration with Landau's flexible transmission system
(Bombois, Scorletti, 2012)

Experiment design objective

Perform closed-loop identification with fixed controller C_{id} while

- minimizing performance degradation J_r due to external $\Phi_r(\omega)$:

$$J_r = \beta_u Var\{u_r\} + \beta_y Var\{y_r\}$$

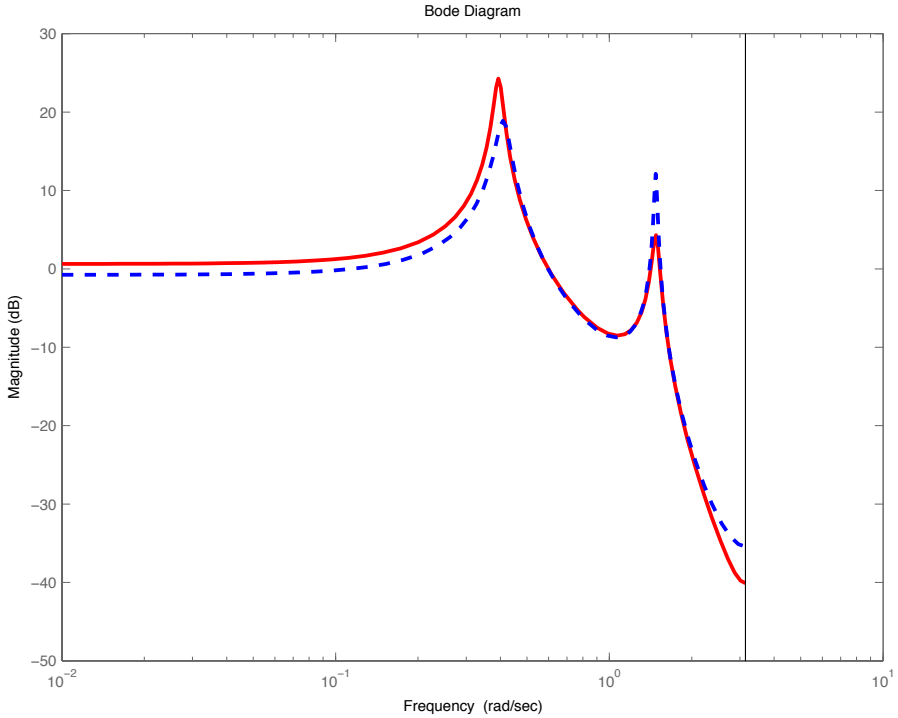
- maintaining accuracy required by robust control objective:

$$|G_0(e^{j\omega}) - G(e^{j\omega}, \hat{\theta}_N)| \leq r_{adm}(\omega) \quad \forall\omega \quad \text{w.p. } 95\%$$

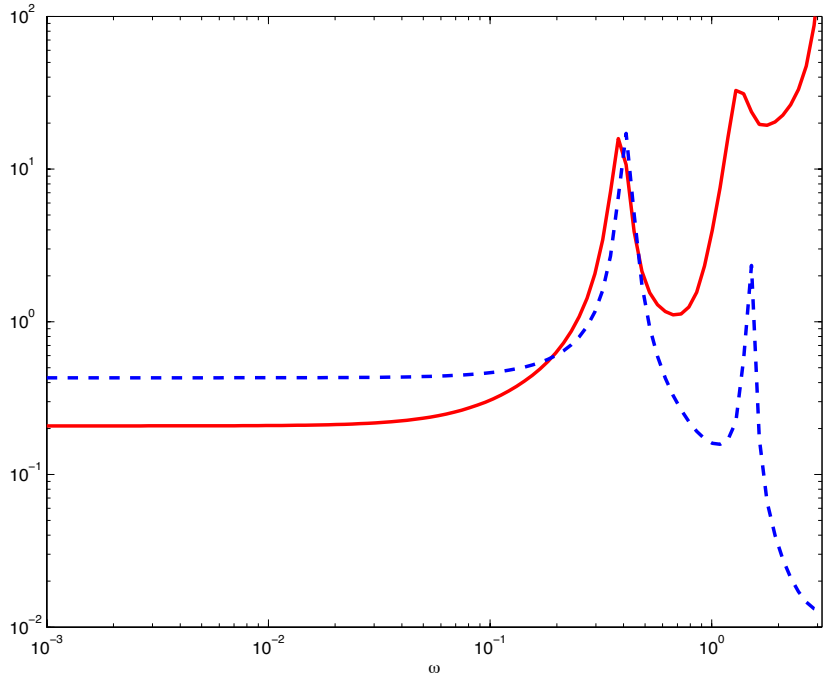
Purpose: use this model to design a better controller

Step 1: obtain a cheap initial model

Open-loop experiment with 100 data yields G_{init}



$|G_0(e^{j\omega})|$ and $|G_{init}(e^{j\omega})|$



$|G_0(e^{j\omega}) - G_{init}(e^{j\omega})|$ and $r_{adm}(\omega)$

- G_{init} does not satisfy quality constraint for robust control design
- but will be used to compute an optimal $\Phi_r(\omega)$

Step 2: computation of closed-loop optimal experiment

G_{init} is used to solve the closed-loop optimal design problem.

Two suboptimal solutions for $\Phi_r(\omega)$ are sought:

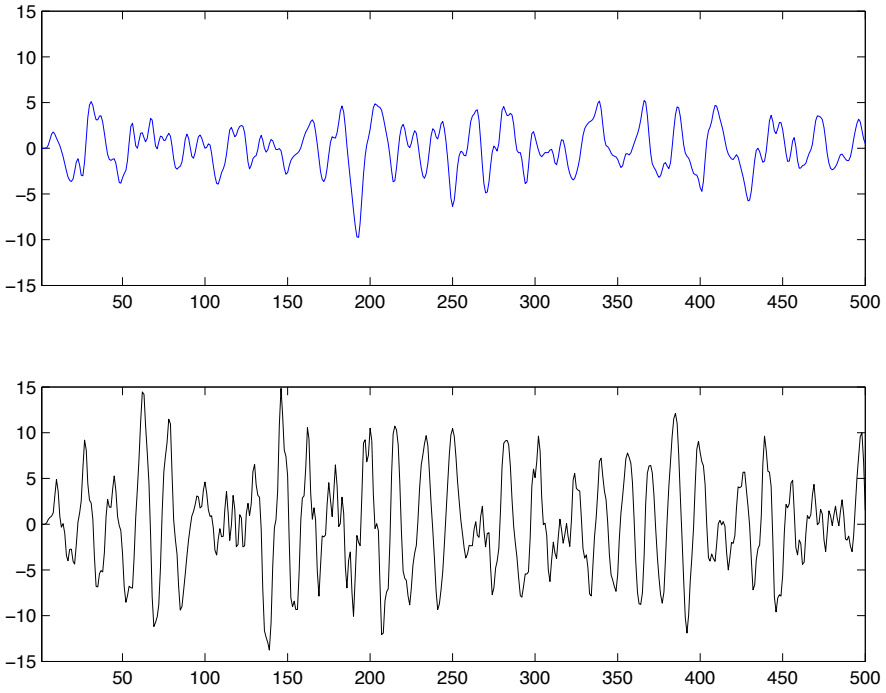
- white noise: $\Phi_r(\omega) =$ constant
- $\Phi_{r,10}(\omega) = \sum_{k=-10}^{10} c_k \, e^{j\omega k} \geq 0$

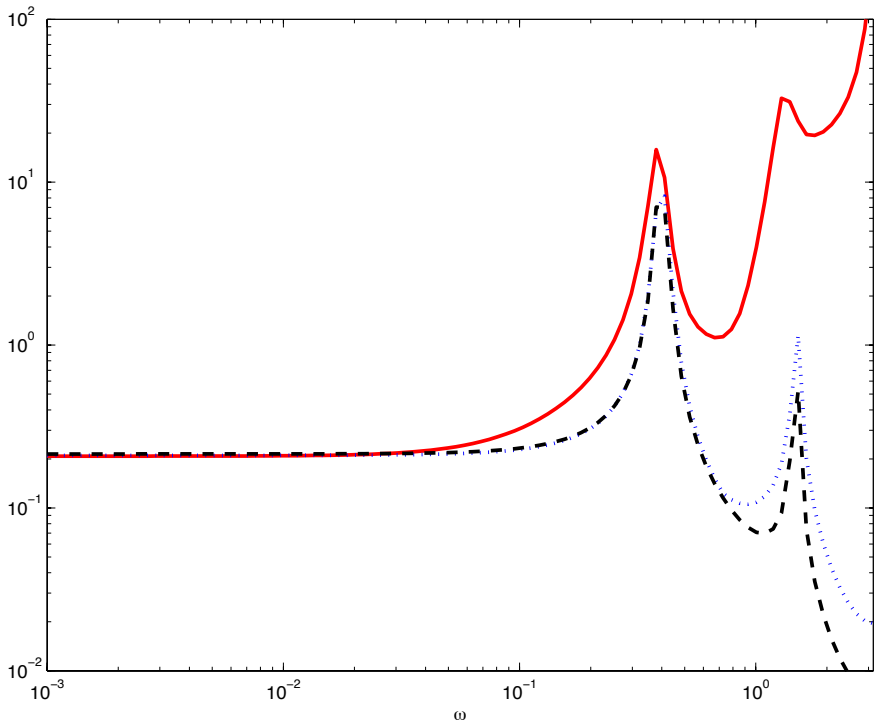
Closed-loop identification (500 data) with these 2 suboptimal spectra:

- with white noise: $\Rightarrow \quad J_r = 22$
- with optimal $\Phi_{r,10}(\omega)$: $\Rightarrow \quad J_r = 9.3$

We note that the performance degradation induced by the (sub) optimal excitation signal is more than two times smaller than with white noise excitation

Some comparisons:





Output perturbation y_r induced by

$\Phi_{r,10}$ and by Φ_r = white

$r_{adm(\omega)}$ compared with $|G_0(e^{j\omega}) - G(e^{j\omega}, \hat{\theta}_N)|$

obtained with $\Phi_{r,10}$ (dotted) and Φ_r = white (dashed)

Alternatively, the (sub)optimal excitation allows us to achieve the same required accuracy in a much shorter time

Outline

- The origins of system identification
- 1965: the birth of identification in engineering
- The parametric Prediction Error framework
- Search for the model structure: key stumbling block
- Major challenge today: reduce the cost of identification
- Non-parametric approaches
- Reducing the cost of the experiment
- Application-oriented experiment design
- Present research areas
- Conclusions

Basic concepts

(Hjalmarsson, ECC 09 plenary)

A. Waterbed effect - fundamental limitation (Rojas, Welsh, Agüero, 2009)

For open loop identification with OE model:

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} N\Phi_u(\omega) Var\{G(e^{j\omega},\hat{\theta}_N)\}d\omega = n_G\,\sigma_e^2$$

Discussion:

- Energy cost for open loop identification is $\int_{-\pi}^{\pi} N\Phi_u(\omega)d\omega$
- Requiring low variance everywhere is expensive
- Cost is proportional to number of parameters and to noise variance
- To get $Var\{G(e^{j\omega},\hat{\theta}_N)\} \leq \frac{1}{\gamma}\ \ \forall\omega$ requires $NE[u^2(t)] \geq \gamma n_G\sigma_e^2$.
- $\sup_\omega Var\{G(e^{j\omega},\hat{\theta}_N)\} \geq \frac{n_G\sigma_e^2}{NE[u^2(t)]}$
- $\Rightarrow$ Do not waste your energy at frequencies where you do not need it !!

B. Performance degradation and acceptable models

(Gevers & Ljung, 1986; Hjalmarsson, 2009)

- Let $\mathcal{J}$ be the quantity of interest for the application
- With true system: $\mathcal{J}(S_0)$; with a model: $\mathcal{J}(M(\theta))$. Assume $M(\theta_0) = S_0$.
- **Performance degradation measure**: differentiable function

  $V_{app}(\theta) = ||\mathcal{J}(M(\theta)) - \mathcal{J}(S_0)||$ such that $V_{app}(\theta_0) = 0$.
- **Set of acceptable models** for the application:

$$\mathcal{D}_{app} \triangleq \left\{\theta \ : V_{app}(\theta) \leq \tfrac{1}{\gamma}\right\} \quad (\gamma = \text{accuracy})$$

Example 1: step response application

With true θ_0 : $y_t(\theta_0), t = 1, \ldots, N$

With estimated $\hat{\theta}_N$: $y_t(\hat{\theta}_N), t = 1, \ldots, N$

$$\Rightarrow V_{app}(\theta) = \tfrac{1}{N} \sum_{t=1}^{N} [y_t(\theta_0) - y_t(\hat{\theta}_N)]^2$$

Example 2: steady-state gain for FIR(n) system

$$\Rightarrow V_{app}(\theta) = [\sum_{t=1}^{n} \theta_k^0 - \sum_{t=1}^{n} \hat{\theta}_k]^2$$

C. Identification uncertainty set

- Define identification criterion and assume system is in model set:
- $V_{id}(\theta) \triangleq \frac{1}{2}\{E[\epsilon_t(\theta)]^2 - \sigma_e^2\}$. Note that $V_{id}(\theta_0) = 0$.
- $V_{id}(\theta) \approx (\theta - \theta_0)^T I(\theta_0)(\theta - \theta_0)$ defines ellipsoid
- With probability α, $\hat{\theta}_N \in \mathcal{D}_{id}$ **(identification uncertainty set)**:

$$\mathcal{D}_{id} = \{\theta : NV_{id}(\theta) \leq \sigma_e^2 \chi_\alpha^2(n_\theta)\} \text{ where } n_\theta = \#\text{ parameters}$$

- $\chi_\alpha^2(n) \approx (\beta + \sqrt{n})^2 = O(n)$

D. Identification with minimal cost

In open loop: $\min_{N, \Phi_u(\omega)} \int_{-\pi}^{\pi} N\Phi_u(\omega)d\omega$ such that $\mathcal{D}_{id} \subseteq \mathcal{D}_{app}$

Interpretation of $\mathcal{D}_{id} \subseteq \mathcal{D}_{app}$

Let $V_{app}(\theta)$ be three times differentiable. Then:

$$\mathcal{D}_{app} \approx \left\{\theta \ : \ \gamma(\theta-\theta_0)^T V''_{app}(\theta_0)(\theta-\theta_0) \leq 1\right\}$$

while

$$\mathcal{D}_{id} \approx \left\{\theta : \frac{N}{\sigma_e^2 \chi_\alpha^2(n_\theta)}(\theta-\theta_0)^T I(\theta_0)(\theta-\theta_0) \leq 1\right\}$$

Therefore:

$$\mathcal{D}_{id} \subseteq \mathcal{D}_{app} \iff \underbrace{N.I(\theta_0)} \geq \gamma.\sigma_e^2.\chi_\alpha^2(n_\theta).V''_{app}(\theta_0)$$

identification effort:
data length x information matrix

Application-oriented experiment design

$$\mathcal{D}_{id} \subseteq \mathcal{D}_{app} \iff \underbrace{N.I(\theta_0)}_{\text{identification effort:}} \geq \gamma.\sigma_e^2.\chi_\alpha^2(n_\theta).V_{app}''(\theta_0)$$

identification effort:
data length x information matrix

If $V_{app}''(\theta_0)$ is singular, of rank $m < n_\theta$, then $I(\theta_0)$ can be singular and $\chi_\alpha^2(n_\theta)$ replaced by $\chi_\alpha^2(m) \Longrightarrow$ identification cost is reduced.

Example: estimation of steady state gain of FIR(n) system

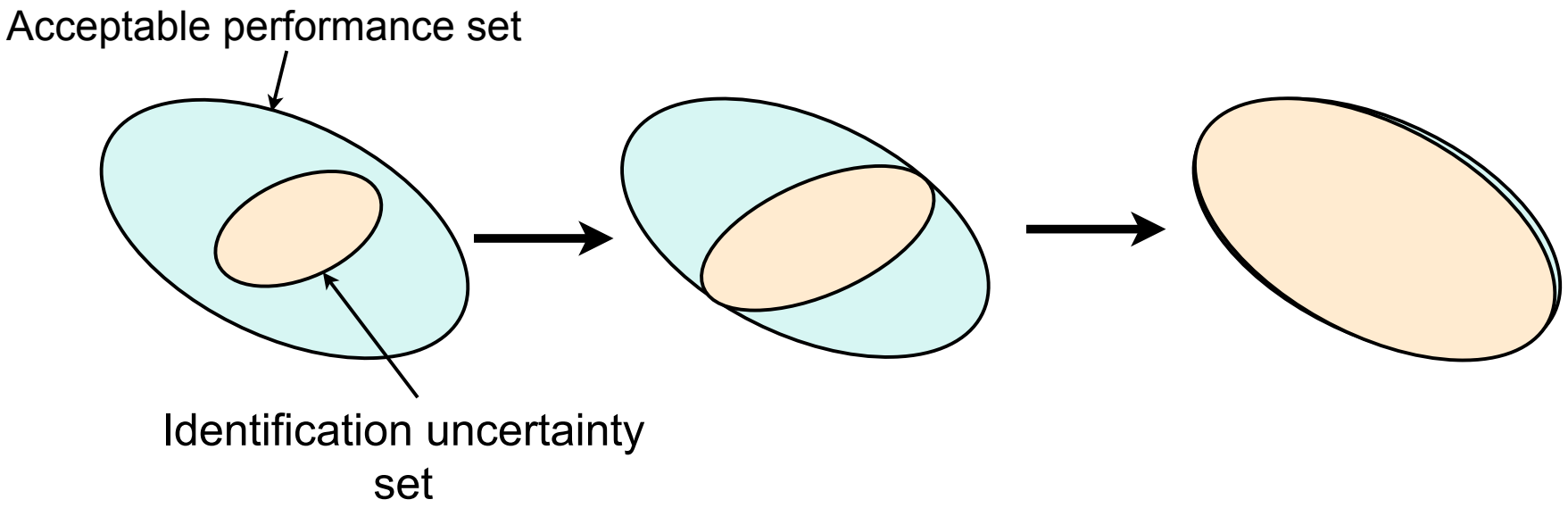
- $y_t = \sum_{k=0}^{n-1} \theta_k u_{t-k} + e_t$
- Compare white noise input u with variance σ_u^2 with constant input $u_t = \sigma_u$.
- With constant input, same precision is obtained with n times less input energy (ex: required data length is n times smaller)

See also Mårtensson and Hjålmarsson, 2011

Application-oriented experiment design

$$\mathcal{D}_{id} \subseteq \mathcal{D}_{app} \iff \underbrace{N.I(\theta_0)}_{} \geq \gamma.\sigma_e^2.\chi_\alpha^2(n_\theta).V''_{app}(\theta_0)$$

identification effort:
data length x information matrix



Optimal experiment: make $N.I(\theta_0) = \gamma.\sigma_e^2.\chi_\alpha^2(n_\theta).V''_{app}(\theta_0)$

The two ellipsoids coincide.
Usuallly not possible !

Outline

- The origins of system identification
- 1965: the birth of identification in engineering
- The parametric Prediction Error framework
- Search for the model structure: key stumbling block
- Major challenge today: reduce the cost of identification
- Non-parametric approaches
- Reducing the cost of the experiment
- Application-oriented experiment design
- Present research areas
- Conclusions

Active research areas

They can be broadly classified in two categories

1. Improving the quality of the estimate

- Regularization methods, Bayesian methods
- Minimizing the Mean Square Error
- Bias-variance trade-off
- Optimal model order selection

2. Experiment design questions

- Optimal experiment design
  - ✦ Open loop, closed loop, joint closed loop design
  - ✦ Linear systems and nonlinear systems
  - ✦ Parametrization of all input spectra
- Informative experiments and identifiability
  - ✦ For networks of dynamical systems
  - ✦ For classes of nonlinear models

1. Regularisation, Bayesian estimation and kernel methods

Connects «classical» system identification with machine learning and Bayesian estimation.

Simple introduction: consider a linear regression model

$$\boldsymbol{Y_N} = \boldsymbol{\Phi_N}\boldsymbol{\theta} + \boldsymbol{V_N}$$

Least squares estimate:

$$\hat{\theta} = \textbf{arg min}_{\theta} ||Y_N - \Phi_N\theta||^2 = (\Phi_N^T\Phi_N)^{-1}Y_N$$

$\Phi_N^T\Phi_N$ may be singular or ill-conditioned
$\Longrightarrow$ add **regularization term**:

$$\begin{aligned} \hat{\theta} &= \textbf{arg}\min_{\theta} ||Y_N - \Phi_N\theta||^2 + \gamma\theta^T P^{-1}\theta \\ &= P\Phi^T(\Phi P\Phi + \gamma I)^{-1}Y_N \end{aligned}$$

Regularisation, Bayesian estimation, kernel methods (cont'd)

Much present research is on the search for the best regularization term and on the interpretation: see the excellent survey paper
Pillonetto, Dinuzzo, Chen, De Nicolao, Ljung, 2014

• Kernel-based approach: function estimation in infinite-dimensional space with added regularization term. Introduced into system identification by
Pillonetto and De Nicolao, 2010

• Empirical Bayes approach: impulse response is modeled as a zero-mean Gaussian process; the parameters of the pdf are estimated by Max Likelihood: Rasmussen and Williams, 2006

• This idea was first proposed in the stochastic embedding approach:
Goodwin, Gevers, Ninness, 1992. Connection is established in Ljung, Goodwin, Aguero, 2014

A key element of the discussion is the model order selection problem which is closely linked to the bias-variance tradeoff.

2. Optimal experiment design questions

Key references for optimal experiment design: Goodwin & Payne, 1977; Zarrop,1979

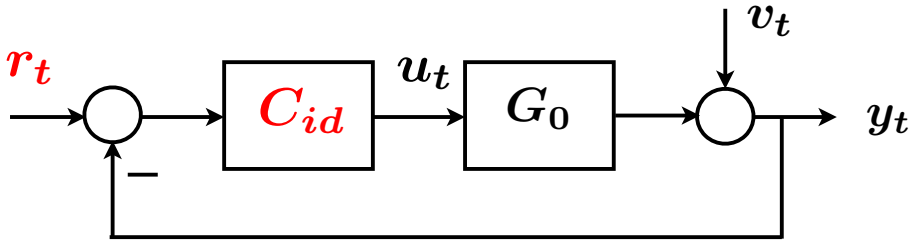
Joint optimal experiment design problem for closed-loop linear systems:

$$\min_{C_{id}(z),\ \Phi_r(\omega)} J, \quad \text{where } J \text{ can be}$$

- quality criterion. Ex: $J = \int_{-\pi}^{\pi} Var(C(\hat{G}_N(\omega)))d\omega$ or $Var(\hat{G}) < \gamma(\omega) \ \ \forall\omega$
- or cost criterion. Ex: $J = \int_{-\pi}^{\pi} \Phi_r(\omega)d\omega$

Always subject to constraints.

Examples: $\int_{-\pi}^{\pi} \Phi_y(\omega)d\omega \leq \alpha$, or $|\Phi_y(\omega)| \leq \alpha \ \ \forall\omega$



Optimal solution for integral criteria and integral constraints only:
Hildebrand, Gevers, Solari, 2014

Solution over a restricted set of spectra for integral and pointwise constraints:
Hjalmarsson & Jansson, 2008

Open loop design for classes of nonlinear systems

Optimal input design for nonlinear systems is very hard.

Optimal or suboptimal solutions are obtained by
- ✦ restricting the class of model structures
- ✦ restricting the class of inputs

Example:

- Model class: Nonlinear FIR

$$y_t(\theta) = G_{NL}(u_t, u_{t-1}, \ldots, u_{t-n+1}, \theta)$$

- Input signal: finite set of possible values

$$u_t \in \{u_1, \ldots, u_A\}$$

Larsson, Hjalmarsson and Rojas, 2010; De Cock, Gevers, Schoukens, 2014

Parametrization of all input spectra yielding the same covariance

Consider open loop identification and an OE model structure:

$$y_t = G(z,\theta)u_t + e_t \quad \text{with} \quad \dim(\theta) = n$$

Then the normalized asymptotic covariance matrix of $\hat{\theta}_N$ can be written as

$$P_\theta^{-1} = \frac{1}{2\pi}\int_{-\pi}^{\pi}\left(\frac{\partial G(e^{j\omega},\theta)}{\partial\theta}\right)\left(\frac{\partial G(e^{j\omega},\theta)}{\partial\theta}\right)^{*}\Phi_u(\omega)d\omega + M_e$$

Then the set of all achievable P_θ (for all possible $\Phi_u(\omega)$)

can be parametrized as a linear combination of n matrices:

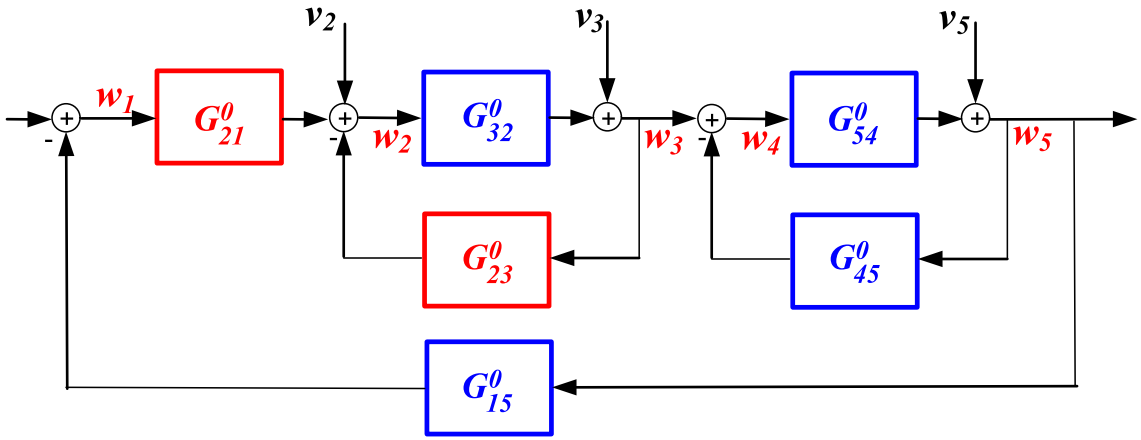
$$P_\theta^{-1} = m_1 R_1 + \ldots + m_n R_n + M_e$$

This gives a parametrization of all spectra yielding the same covariance P_θ.

Mahata, 2013

3. Informative experiments and identifiability

Identification of dynamic systems within networks



- w_j: measured
- v_j: unmeasured

All kinds of interesting questions, that can be summed up as:

- Identifiability and experiment design questions:
  Example: which signals do I need to identify G^0_{21}?
  Answer: w_1 and w_2 suffice

- Accuracy questions:
  Example: do I get a better estimate of G^0_{21} if I also measure w_3?
  Answer: yes, by simultaneously estimating G^0_{21} and G^0_{23}.

Van den Hof, Dankers, Heuberger, Bombois, 2013, 2014

Identifiability and informativity for nonlinear systems

- For the very broad model class

$$\dot{x} \quad = \quad f(x, u, \theta), \qquad y = h(x, u, \theta)$$

Sufficient conditions for identifiability of θ

based on application of the Ritt algorithm: Ljung and Glad, 1994

Powerful but huge computational requirements.

- For nonlinear model classes that are affine in θ

Recent results on identifiability and informative experiments:

Gevers, Bazanella, Coutinho, Dasgupta, 2014

Conclusions

- ✤ System identification in engineering started almost 50 years ago
- ✤ Enormous progress has been accomplished
- ✤ It is still the major task in any advanced control project
- ✤ Major challenge today: reduce the cost of identification
  - Make it more user friendly and data-driven
  - Optimize the experiment: less time, less energy
  - Tune the experiment towards the application: don’t waste !
- ✤ Many research challenges remain, new ones have appeared:
  - Structured systems
  - Nonlinear systems
  - Large distributed and network controlled systems

References

Akaike, H. (1974). Stochastic theory of minimal realization. IEEE Trans. Automatic Control **26**, 667--673.

Åström, K. J. (1980). Maximum likelihood and prediction error methods. Automatica **16**, 551--574.

Åström, K. J. and T. Bohlin (1965). Numerical identification of linear dynamic systems from normal operating records. In: Proc. IFAC Symposium on Self-Adaptive Systems. Teddington, UK. pp. 96--111.

Barenthin, Märta, Henrik Jansson and Håkan Hjalmarsson (2005). Applications of mixed H_2 and H_∞ input design in identification. In: 16th IFAC World Congress on Automatic Control, paper 03882. Prague. Paper Tu-A13-TO/1.

Bombois, X. and G. Scorletti (2012). Design of least costly identification experiments. Journal européen des systèmes automatisés (6-7), 587--610.

Bombois, X., G. Scorletti, M. Gevers, P.M.J. Van den Hof and R. Hildebrand (2006). Least costly identification experiment for control. Automatica **42**(10), 1651--1662.

Bombois, X., G. Scorletti, M. Gevers, R. Hildebrand and P.M.J. Van den Hof (2004). Cheapest open-loop identification for control. In: CD-ROM Proc. 33rd IEEE Conf on Decision and Control. The Bahamas. pp. 382--387.

Cock, A. De, M. Gevers and J. Schoukens (2014). D-optimal input design for FIR-type nonlinear systems: A dispersion-based approach. submitted to Automatica.

Dankers, A., P.M.J. Van den Hof, X. Bombois and P.S.C. Heuberger (2014). Identification of dynamic models in complex networks with prediction error methods - predictor input selection. Submitted to IEEE Trans. Automatic Control.

Deistler, M. (2002). System identification and time series analysis: Past, present, and future. In: `Stochastic Theory and Control', Festschrift for Tyrone Duncan. Springer. Kansas, USA. pp. 97--108.

Fisher, R.A. (1912). On an absolute criterion for fitting frequency curves. Mess. Math. **41**, 155.

Gauss, C.F. (1809). Teoria Motus Corporum Coelestium in Sectionibus Conicis Solem Ambientium. Reprinted translation: `Theory of Motion of the Heavenly Bodies Moving about the Sun in Conic Sections'. Dover. New York.

Gevers, M. (2005). Identification for control: from the early achievements to the revival of experiment design. European Journal of Control **11**, 1--18.

Gevers, M. (2006). A personal view of the development of system identification. IEEE Control Systems Magazine **26**(6), 93--105.

Gevers, M., A. Bazanella, D. Coutinho and S. Dasgupta (2014). Identifiability and excitation of linearly parametrized rational systems. submitted for publication to Automatica.

Gevers, M. and L. Ljung (1986). Optimal experiment designs with respect to the intended model application. Automatica **22**, 543--554.

Goodwin, G. C. and R. L. Payne (1977). Dynamic System Identification: Experiment Design and Data Analysis. Academic Press. New York.

Hildebrand, R., M. Gevers and G. Solari (2014). Closed-loop optimal experiment design: solution via the moment extension. submitted for publication to IEEE Trans. Auto. Control.

Hjalmarsson, H. (2009). System identification of complex and structured systems. European Journal of Control **15**(3-4), 275--310.

Ho, B. L. and R. E. Kalman (1965). Effective construction of linear state-variable models from input/output functions. In: Proc 3rd Allerton Conference; [also Regelungstechnik, Vol. 14, pp. 545-548, 1966]. USA. pp. 449--459.

Kalman, R.E. (1960a). Contributions to the theory of optimal control. Boletin de la Sociedad Matematica Mexicana **5**, 102--119.

Kalman, R.E. (1960b). A new approach to linear filtering and prediction problems. Transactions ASME, Series D, J. Basic Eng. **82**, 34--45.

Larimore, W. E. (1990). Canonical variate analysis in identification, filtering, and adaptive control. In: Proc. 29th IEEE Conf. on Decision and Control. Honolulu, Hawaii. pp. 596--604.

Larimore, W.E. (1996). Statistical optimality and canonical variate analysis system identification. Signal Processing **52**(2), 131--144.

Ljung, L. (1999). System Identification: Theory for the User, 2nd Edition. Prentice-Hall. Englewood Cliffs, NJ.

Ljung, L. and T. Glad (1994). On global identifiability for arbitrary model parametrizations. Automatica **30**(2), 265--276.

Ljung, L., G. Goodwin and J. Agüero (2014). Stochastic embedding revisited: A modern interpretation. In: submitted to IEEE Conference on Decision and Control.

Mahata, K. (2013). Variance error, interpolation and experiment design. Automatica **49**(5), 1117--1125.

Mårtensson, Jonas and Håkan Hjalmarsson (2011). How to make bias and variance errors insensitive to system and model complexity in identification. IEEE Transactions Automatic Control **56**(1), 100--112.

Pillonetto, G. and G. De Nicolao (2010). A new kernel-based approach for linear system identification. Automatica **46**(1), 81--93.

Pillonetto, Gianluigi, Francesco Dinuzzo, Tianshi Chen, Giuseppe De Nicolao and Lennart Ljung (2014). Kernel methods in system identification, machine learning and function estimation: A survey. Automatica **50**(3), 657--682.

Rasmussen, C. and C. Williams (2006). Gaussian processes for machine learning. MIT Press.

Rojas, C.R., J.S. Welsh and J.C. Agüero (2009). Fundamental limitations on the variance of estimated parametric models. IEEE Transactions Automatic Control **54**(5), 1077--1081.

Schoukens, J., G. Vandersteen, K. Barbé and R. Pintelon (2009). Nonparametric preprocessing in system identification: a powerful tool. European Journal of Control **15**(3-4), 260--274.

Schoukens, J., Y. Rolain and R. Pintelon (2006). Leakage reduction in frequency response function measurements. IEEE Trans. Instrumentation and Measurement **55**, 2286--2291.

Schoukens, J., Y. Rolain, G. Vandersteen and R. Pintelon (2011). User-friendly Box-Jenkins identification using nonparametric noise models. In: Proc. IEEE CDC and ECC 2011. Orlando, USA. pp. 2148--2153.

Van den Hof, P.M.J., A. Dankers, P.S.C. Heuberger and X. Bombois (2013). Identification of dynamic models in complex networks with prediction error methods- basic methods for consistent module estimates. Automatica **49**, 2994--3006.

Van Overschee, P. and B. De Moor (1994). N4SID: Subspace algorithms for the identification of combined deterministic-stochastic systems. Automatica **30**(1), 75--93.

Verhaegen, M. (1994). Identification of the deterministic part of MIMO state space models given in innovations form from input-output data. Automatica **30**(1), 61--74.

Wahlberg, B. and L. Ljung (1986). Design variables for bias distribution in transfer function estimation. IEEE Trans. Automatic Control **AC-31**, 134--144.

Zarrop, M. (1979). Optimal experiment design for dynamic system identification. Lecture Notes in Control and Information Sciences, Vol. 21, Springer Verlag. Berlin, New York.