SYSTEM IDENTIFICATION OF COMPLEX AND STRUCTURED SYSTEMS

Håkan Hjalmarsson

ACCESS Linnaeus Center, School of Electrical Engineering
KTH - Royal Insititute of Technology, Stockholm

European Control Conference
August 26, 2009

Acknowledgements

- Former PhD-students: Kristian Lindqvist, Henrik Jansson, Jonas Mårtensson, Märta Barenthin
- Collaborators: Xavier Bombois, László Gerencsér, Michel Gevers, Roland Hildebrand, Lennart Ljung, Brett Ninness, Cristian Rojas, Bo Wahlberg, James Welsh

What do we mean by complexity?

- Computer science: Computational complexity (Turing, Church, ...)
- Systems and control:
  - Computational complexity (see survey by Blondel and Tsitsiklis)
  - Feedback control under uncertainty (Zames; Egerstedt and Brockett; Delvenne and Blondel; Zhang and Guo)
- System identification:

???

Complexity in system identification

- Kolmogorov n-width (Zames)
- VC-dimension and PAC-learning (Vapnik-Chervonenkis, Vidyasagar)
- The minimum cost required to get within a given accuracy
- The minimum *experimental* cost
- Example: Time complexity of worst-case system identification (Poolla and Tikku)
- Here: *The minimum experimental cost required to achieve a certain performance in the application*

Experiment design for system identification

- Much work in 1970's:
  (Mehra; Goodwin and Payne; Ng, Goodwin and Söderström; Zarrop)
  - Scalar criteria, often not involving the application directly
  - All covariance matrices can be generated by sinusoidal inputs
- Renewed interest in mid 1980's:
  - Use of high order variance expressions (Gevers and Ljung)
  - Design tied to the application (e.g. minimum variance control)
- Revival in 2000's:
  - *Least-costly identification for robust control*
    (Bombois, Scorletti, Van den Hof, Gevers and Hildebrand)
  - Semi-definite programming (Cooley, Lee and Boyd; Lindqvist; Jansson)
  - Robust stability and robust performance criteria (Hildebrand and Gevers; Jansson)
  - Nonlinear systems (Mårtensson; Novara, Vincent and Poolla)
  - Robust input design (Mårtensson; Rojas, Welsh, Goodwin, Feuer)
  - Plant-friendly design (Rivera, Lee, Mittelmann and Braun)

Is there a problem?

The water-bed effect (Rojas, Welsh and Agüero):

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} \underbrace{N\Phi_u^{id}(e^{j\omega})}_{\text{Input energy density}} \ \text{Var}[\hat{G}(e^{j\omega})]\mathrm{d}\omega = \underbrace{n}_{\#\text{ parameters}} \ \underbrace{\lambda_e}_{\text{noise variance}}$$

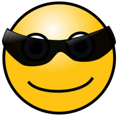
(output error models)

A fundamental limitation

YES, there is a problem!

An appetizer

Please say hello to:



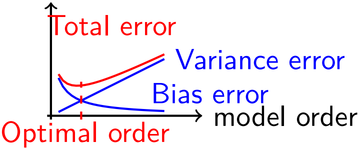
Izzy and Ozzy

Static gain estimate

Model order:	very low	true	very high
Constant input	good	good	good

Impulse response coefficient estimate

Model order:	very low	true	very high
White input	fair	good	good



Robustness against choice of model order. Why?

Outline

Cost of complexity

An alternative formulation

Output error models

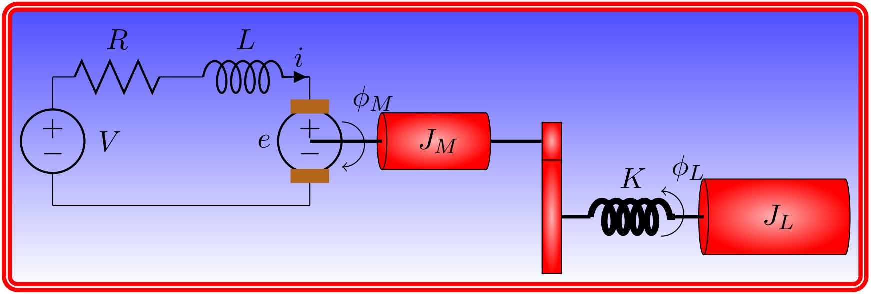
Some connections to the past

The impact of optimal experiments on the identification problem

Numerical computation of experiment designs

Implementation of experiment designs

An application example: MPC of a DC-motor

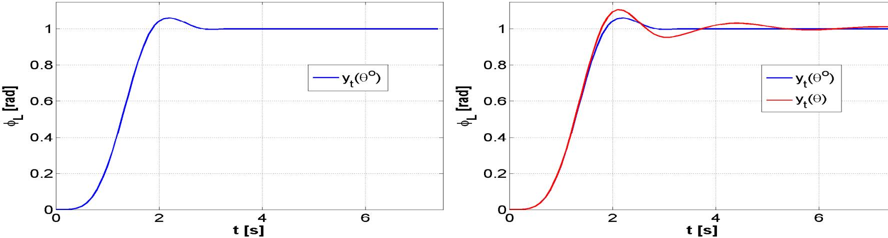


- Input: Voltage V
- Output: Angle ϕ_L
- Model parameters θ: Resistance R, Moment of inertia J_L, Elasticity K, ...
- True parameters: θ^o

An application example: MPC of a DC-motor

- Ideal response: $y_t(\theta^o)$ - true parameters used in MPC
- Actual response: $y_t(\theta)$ – parameter θ used in MPC

Constraint: Maximum input move 40

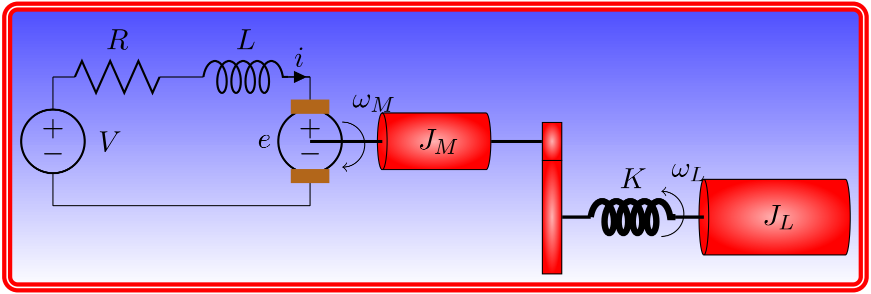


Performance degradation /Set of acceptable models

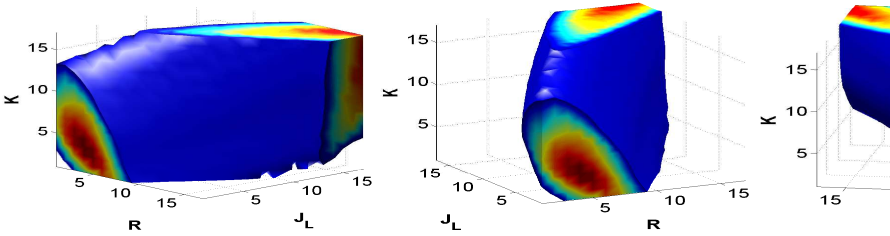
$$V_{app}(\theta) = \frac{1}{N}\sum_{t=1}^{N}(y_t(\theta^o) - y_t(\theta))^2$$

$$\mathcal{E}_{app} = \left\{\theta : V_{app}(\theta) \leq \frac{1}{\gamma}\right\} \qquad (\gamma = \text{accuracy})$$

An application example: MPC of a DC-motor



Set of acceptable models $\mathcal{E}_{app}$: (Maximum move size 3)



Summary of concepts

- Performance degradation for application: $V_{app}(\theta)$
- Set of acceptable models: $\mathcal{E}_{app} = \left\{\theta : V_{app}(\theta) \leq \frac{1}{\gamma}\right\}$
- Identification: Produce $\hat{\theta}_N \in \mathcal{E}_{app} \subset \mathbb{R}^n$ ($N =$ sample size)
- *Cost of complexity Q*
  =Minimum possible experimental cost required for $\hat{\theta}_N \in \mathcal{E}_{app}$
- Least-costly identification

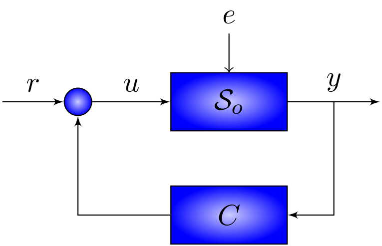
- *Quantification of Q*
- Here: Experimental cost $=$ input energy

Cost of complexity

$$Q := \min N\mathrm{E}[u_t^2]$$
$$s.t.\ \hat{\theta}_N \in \mathcal{E}_{app} \subset \mathbb{R}^n$$

An optimal experiment design problem

Identification recap



- Prediction error identification:
  - Prediction error: $\varepsilon_t(\theta)$
  - $\hat{\theta}_N = \arg\min \sum_{t=1}^{N} \varepsilon_t^2(\theta)$, $V_{id}(\theta) = \mathrm{E}[\varepsilon_t^2(\theta)] - \lambda_e \geq 0$
- Random noise (innovations (noise) variance λ_e)
- Stationary signals
- True system in the model set: $\mathcal{S}_o \Leftrightarrow \theta^o$ (to be relaxed later)
- High accuracy γ (implies large sample size N)
- $\sqrt{N}\left(\hat{\theta}_N - \theta^o\right) \sim \mathrm{As}\mathcal{N}\left(0, 2\lambda_e V_{id}''(\theta^o)^{-1}\right)$

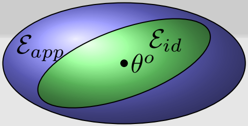
The cost of complexity

- Random noise $\Rightarrow \hat{\theta}_N$ random variable
- Cannot guarantee $\hat{\theta}_N \in \mathcal{E}_{app}$
- Relaxation: $\mathbf{Probability}(\hat{\theta}_N \in \mathcal{E}_{app}) = \alpha (= 99\%$ e.g.)
- In general difficult to compute
- Use standard asymptotic confidence ellipsoids:

$\mathbf{Probability}(\hat{\theta}_N \in \mathcal{E}_{id}) \approx \alpha$, where

Cost of complexity

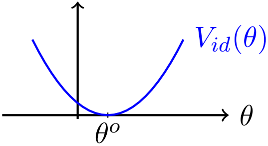
$$Q := \min N\mathrm{E}[u_t^2]$$
$$s.t.\ \mathcal{E}_{id} \subseteq \mathcal{E}_{app} \subset \mathbb{R}^n$$



An alternative expression for the confidence ellipsoid

$$\mathcal{E}_{id} = \left\{ \theta : \ \frac{N}{2} \left(\theta - \theta^o\right)^T V_{id}''(\theta^o) \left(\theta - \theta^o\right) \le \lambda_e \, n \right\}$$

Recall: $V_{id}(\theta) = \mathrm{E}[\varepsilon_t^2(\theta)] - \lambda_e$. High accuracy γ, i.e. $\mathcal{E}_{app}$ small



$$\Rightarrow V_{id}(\theta) \approx \frac{1}{2}(\theta - \theta^o)^T V_{id}''(\theta^o)(\theta - \theta^o)$$

$$\Rightarrow \qquad \mathcal{E}_{id} = \{\theta : \ N V_{id}(\theta) \le \lambda_e \, n\}$$

Confidence ellipsoid = Level set for identification criterion!

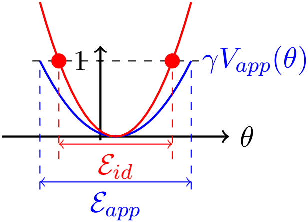
An alternative formulation of cost of complexity

Level sets:

$$\mathcal{E}_{app} = \left\{\theta : V_{app}(\theta) \leq \frac{1}{\gamma}\right\} = \{\theta : \gamma V_{app}(\theta) \leq 1\}$$

$$\mathcal{E}_{id} = \{\theta : \ N V_{id}(\theta) \leq \lambda_e \, n\} \ = \left\{\theta : \ \frac{N}{\lambda_e n} V_{id}(\theta) \leq 1\right\}$$

Recall: High accuracy $\gamma \Rightarrow \mathcal{E}_{app}$ small $\Rightarrow V_{id}$ and V_{app} quadratic:



$$\mathcal{E}_{id} \subseteq \mathcal{E}_{app} \quad \Leftrightarrow \quad \frac{N}{\lambda_e n} V_{id}(\theta) \geq \gamma V_{app}(\theta) \quad \forall \theta \in \mathcal{E}_{app}$$

$$\Leftrightarrow \quad N V_{id}(\theta) \geq \lambda_e \gamma n V_{app}(\theta) \quad \forall \theta \in \mathcal{E}_{app}$$

An alternative formulation of cost of complexity

Cost of complexity

$$Q := \min N\mathrm{E}[u_t^2]$$
$$s.t.\ NV_{id}(\theta) \geq \lambda_e \gamma n V_{app}(\theta),\ \forall \theta \in \mathcal{E}_{app}$$

Example (Estimation of an impulse response coefficient)

- Model: $y_t = \sum_{k=1}^{n} \theta_k u_{t-k} + e_t$
- Objective: Estimate θ_1
- $V_{app}(\theta) = (\theta_1^o - \theta_1)^2$
- Only one parameter matters!
- but our confidence ellipsoid includes all n parameters
- Use a confidence ellipsoid for that parameter only: $n \Rightarrow 1$

A generalized version

General case:

$n_{app} = \#$ non-singular directions of V_{app} ($=$ rank V''_{app})

Cost of complexity

$$Q := \min N\mathrm{E}[u_t^2]$$
$$s.t.\ NV_{id}(\theta) \geq \lambda_e \gamma n_{app} V_{app}(\theta),\ \forall\theta \in \mathcal{E}_{app}$$

Output error models

True system: $y_t = G_o(q)u_t + e_t$

Model: $y_t = G(q,\theta)u_t + e_t$

PE: $\varepsilon_t(\theta) = y_t - G(q,\theta)u_t = (G_o(q) - G(q,\theta))u_t + e_t$

$$\begin{aligned} V_{id}(\theta) &= \mathrm{E}[\varepsilon_t^2(\theta)] - \lambda_e \\ &= \frac{1}{2\pi}\int_{-\pi}^{\pi} \Phi_u^{id}(e^{j\omega})\, \left|G(e^{j\omega},\theta) - G_o(e^{j\omega})\right|^2 \,\mathrm{d}\omega \end{aligned}$$

Output error models

Cost of complexity

$$Q := \min N\mathrm{E}[u_t^2] = \frac{1}{2\pi}\int_{-\pi}^{\pi} N\Phi_u^{id}(e^{j\omega})\,\mathrm{d}\omega$$

$$s.t. \qquad \underbrace{NV_{id}(\theta)}_{\frac{1}{2\pi}\int_{-\pi}^{\pi} N\Phi_u^{id}(e^{j\omega})\,|G(e^{j\omega},\theta)-G_o(e^{j\omega})|^2\,\mathrm{d}\omega} \qquad \geq \lambda_e \gamma n_{app} V_{app}(\theta)$$

- Minimization with respect to energy density spectrum $N\Phi_u^{id}$
- Optimization tries to achieve

$$NV_{id}(\theta) = \lambda_e\,\gamma\,n_{app}V_{app}(\theta)$$

Identification cost matched to performance degradation

Output error models: Influence of $\lambda_e \gamma n_{app}$

$$Q := \frac{1}{2\pi}\int_{-\pi}^{\pi} N\Phi_u^{id}(e^{j\omega})\,\mathrm{d}\omega$$

$$s.t.\ \frac{1}{2\pi}\int_{-\pi}^{\pi} N\Phi_u^{id}(e^{j\omega})\,\left|G(e^{j\omega},\theta) - G_o(e^{j\omega})\right|^2\,\mathrm{d}\omega \geq \lambda_e \gamma n_{app} V_{app}(\theta)$$

- Cost & constraint linear in $N\Phi_u^{id}$

$\Rightarrow$ Problem scales with $\lambda_e \gamma n_{app}$

$\Rightarrow$ *Cost of complexity:* $Q = \lambda_e \gamma n_{app}\,\tilde{Q}$

Optimal energy density: $N\Phi_u^{id} = \lambda_e \gamma n \tilde{\Phi}_u$

where $\tilde{\Phi}_u$ the solution to

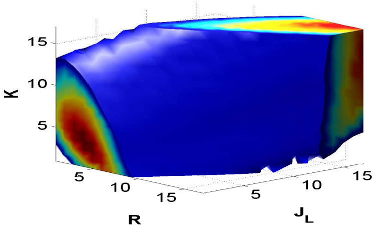
$$\tilde{Q} := \min \mathrm{E}[u^2(t)]$$
$$s.t.\ V_{id}(\theta) \geq V_{app}(\theta)$$

- Independent of sample size, noise variance, accuracy, # of parameters
- Normalized problem

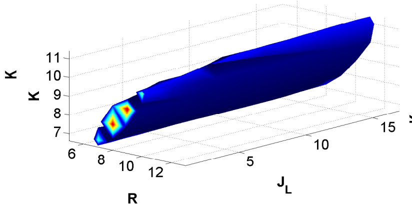
The normalized problem - Insights

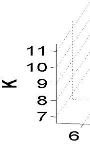
$$\tilde{Q} := \min \mathrm{E}[u^2(t)]$$
$$s.t.\ V_{id}(\theta) \geq V_{app}(\theta)$$

Maximum input move 3



Maximum input move 40

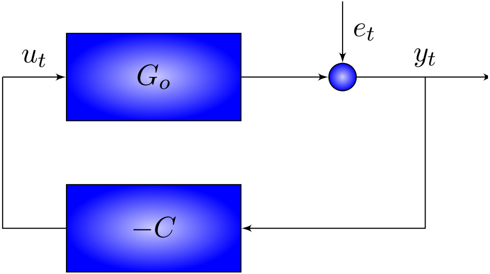


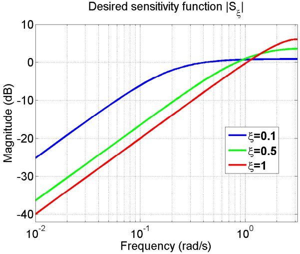


- Performance specifications determine the shape of $V_{app}(\theta)$
- Curvature of $V_{app}(\theta)$ increases when specs. are tightened

$\Rightarrow$ $\tilde{Q}$ reflects performance specifications in the application
We will use $0 \leq \xi \leq 1$ to indicate specs. i.e. $\tilde{Q}(\xi)$

Izzy and Ozzy goes to MRC (model reference control)



- Controller $C = C(G)$, G output error model
- Desired sensitivity function: S_ξ
- Achieved sensitivity function: $S(G) = \frac{1}{1+C(G)G_o}$
- Performance degradation: $V_{app}(G) := \left\| \frac{S(G)-S_\xi}{S_\xi} \right\|_2^2$

MRC: Cost of complexity

$$\tilde{Q} := \min \mathrm{E}[u^2(t)]$$
$$s.t.\ V_{id}(\theta) \geq V_{app}(\theta)$$

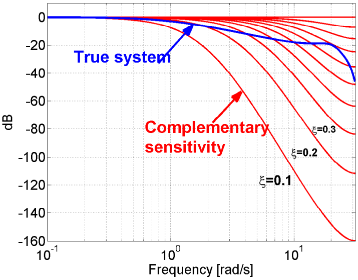
- Matching condition: $V_{id}(\theta) = V_{app}(\theta)$
- Output error:
  $V_{id}(\theta) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \tilde{\Phi}_u(e^{j\omega}) \left| G(e^{j\omega}, \theta) - G_o(e^{j\omega}) \right|^2 \mathrm{d}\omega$
- MRC:
  $V_{app}(G) := \left\| \frac{S(G) - S_\xi}{S_\xi} \right\|_2^2 \approx \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{\Phi_u^{desired}(e^{j\omega})}{\lambda_e} \left| G - G_o \right|^2 d\omega$
- Take $\tilde{\Phi}_u = \Phi_u^{desired}/\lambda_e \Rightarrow N\Phi_u^{id} = \gamma n \Phi_u^{desired}$
  Scaled version of desired operating conditions!

$\Rightarrow$ Upper bound: $Q \leq \gamma n \|\Phi_u^{desired}\|_1$

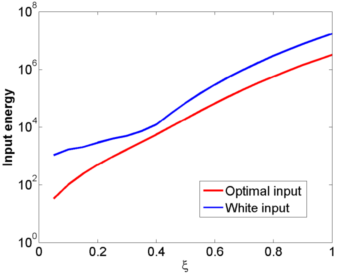
MRC: Cost of complexity

- $Q \leq \gamma n \|\Phi_u^{desired}\|_1 = \lambda_e \gamma n \left\| \frac{1-S_\xi}{G_o} \right\|_2^2$
- Allows user to make informed trade-offs:
- Performance specs. vs experimental cost
- Time vs excitation: $N\Phi_u^{id} = \gamma n \Phi_u^{desired}$:
- Minimum time stealth id.: $\mathrm{E}[u_t^2] \leq \lambda_u^{\max}$

$$\Rightarrow \quad \min N = \gamma n \frac{\mathrm{E}[u_{desired}^2]}{\lambda_u^{\max}}$$



Cost of complexity



Output error models: Cost of complexity - Summary

Cost of complexity

$$Q \approx \lambda_e \, \gamma \, n_{app} \, \tilde{Q}(\xi)$$

- λ_e: noise level
- γ: accuracy
- n_{app}: # of non-singular directions in the parameter space
- $\tilde{Q}(\xi)$: normalized cost

Implications:

- System complexity not important
- Performance specs of application determine cost
- Allows user to make informed trade-offs:
  specs. vs cost, time vs excitation

Some connections to the past

Identification experiment = desired closed loop operating conditions:

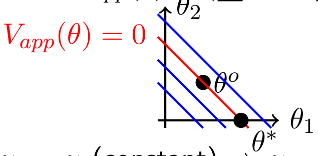
- Random errors
  - Minimum variance control (Gevers and Ljung 1986, Hjalmarsson, Gevers and De Bruyne 1996, Hildebrand and Solari 2007, Mårtensson, Rojas and Hjalmarsson 2009)
- Bias errors
  - Many contributions in the 1990s to identification for control, e.g.:
  - Control-relevant prefiltering (Rivera, Pollard, Garcia 1992)
  - Iterative identification and control (Schrama 1992, Zang, Bitmead and Gevers 1995)
  - Virtual feedback reference tuning (Campi, Lecchini, Savaresi 2002)
- Contributions here:
  - Results above different sides of the same coin (matching V_{id} and V_{app})
  - Matching not enough. Sufficient input energy required. ($N\Phi_u^{id} = \lambda_e \gamma n \Phi_u^{desired}$)

The Izzy & Ozzy problems revisited: Static gain estimation

Model order:	low	true	high
Constant input	good	good	good

$$y_t = \sum_{t=1}^{n} \theta_k u_{t-k} + e_t$$

Performance degradation: $V_{app}(\theta) = (\sum \theta_k - \sum \theta_k^o)^2$



- Optimal input: $u_t = u$ (constant) $\Rightarrow$ $y_t = \sum_k \theta_k^o u + e_t$
- Property of interest visible
- No other system property visible (due to min energy crit.)

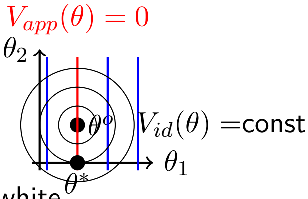
$\Rightarrow$ Perfect match $V_{id}(\theta) \propto V_{app}(\theta)$

- Same input optimal for high order system $\Rightarrow$ high order ok
- $V_{id}(\theta) \propto V_{app}(\theta) \Rightarrow$ Bias minimized!
- $V_{id}(\theta^*) = 0 \Rightarrow$ no unmodelled dynamics $\Rightarrow$ low order optimal

The Izzy & Ozzy problems revisited: Impulse response

Model order:	low	true	high
White input	fair	good	good

Performance degradation: $V_{app}(\theta) = (\theta_1 - \theta_1^o)^2$



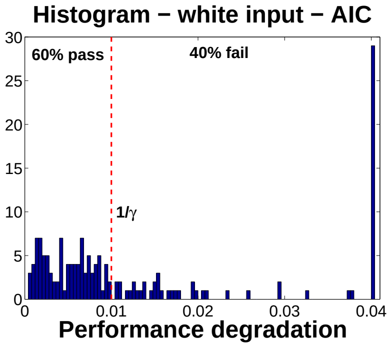
- Optimal input: white
- Property of interest visible
- All system properties visible (V_{id} cannot be shaped arbitrarily)

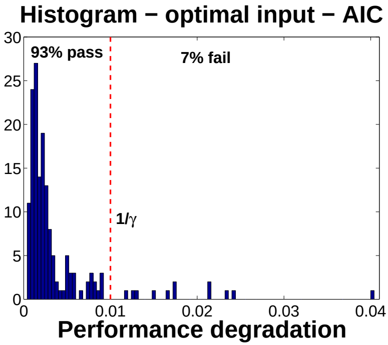
$\Rightarrow$ No match $V_{id}(\theta) \propto V_{app}(\theta)$

- Same input optimal for high order system $\Rightarrow$ high order ok
- V_{id} aligned to V_{app} $\Rightarrow$ Bias minimized!
- $V_{id}(\theta^*) > 0 \Rightarrow$ unmodelled dynamics $\Rightarrow$ low order not optimal
- NMP-zero estimation another example

MRC: Model selection using AIC

- AIC unbiased estimate of $\mathrm{E}[V_{id}(\hat{\theta}_N)]$
- Optimal experiment design: $V_{id} \propto V_{app}$
- Use AIC to estimate $\mathrm{E}[V_{app}(\hat{\theta}_N)]$
- Model order selection with the application in mind
- MRC example revisited:





Identification using optimal experiments: Summary

Optimal experiment

Aims at achieving

$$NV_{id}(\theta) = \lambda_e \, \gamma \, n_{app} V_{app}(\theta)$$

using minimum energy

- To achieve this requires *parsimonious excitation*:
  - i) System properties important to the application should be visible in the data
  - ii) System properties not important to the application should not be visible in the data, unless necessary for i).

    (The let sleeping dogs lie paradigm)
- As a result, the entire system may not have to be identified!
  - Choice of model structure less critical
  - Advice: Don't use too low order (c.f. impulse response). Use model reduction instead (c.f. the ASYM method by Zhu).

Numerical computation

Cost of complexity

$$Q := \min N\mathrm{E}[u^2(t)]$$
$$s.t.\ NV_{id}(\theta) \geq \lambda_e \gamma n_{app} V_{app}(\theta)$$

Recall: Output error models

$$NV_{id}(\theta) = \frac{1}{2\pi}\int_{-\pi}^{\pi} N\Phi_u^{id}(e^{j\omega}) \left|G(e^{j\omega},\theta) - G_o(e^{j\omega})\right|^2 \mathrm{d}\omega$$
$$N\mathrm{E}[u^2(t)] = \frac{1}{2\pi}\int_{-\pi}^{\pi} N\Phi_u^{id}(e^{j\omega})\,\mathrm{d}\omega$$

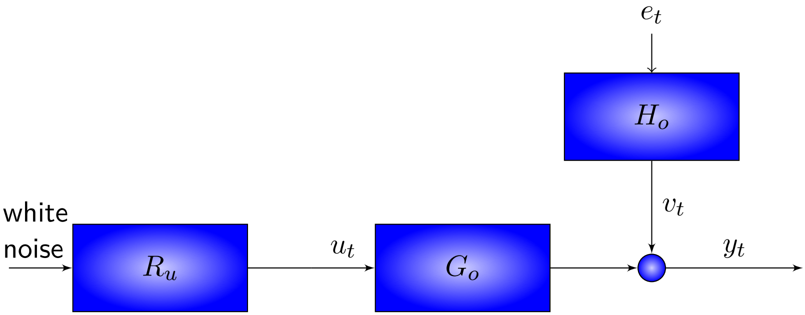
- Use finite dimensional parametrization of $N\Phi_u^{id}$

$\Rightarrow$ Semi-definite program!

- Generalizes to other model structures

Implementation

- Optimal experiment design $\Rightarrow$ Input spectrum Φ_u
- Spectral factorization: $\Phi_u^{id}(e^{j\omega}) = |R_u(e^{j\omega})|^2$



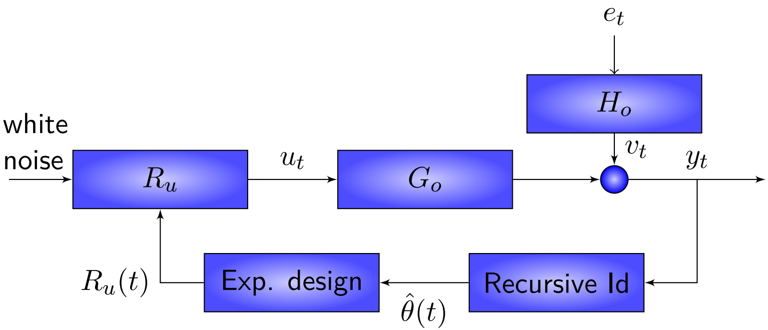
Implementation

Cost of complexity

$$Q := \min N \mathrm{E}[u^2(t)]$$
$$s.t.\ N V_{id}(\theta) \geq \lambda_e \gamma n_{app} V_{app}(\theta)$$

- Optimization problem depends on the unknown system!
- Major obstacle
- Solutions:
  - Robust experiment design (e.g. Rojas, Welsh, Goodwin, Feuer 2007)
  - Adaptive (sequential) experiment design

Adaptive input design



- An adaptive feedback system
- But measured signal not fed back directly
- Exp. design limits input power $\Rightarrow$ Stability when G_o stable

Key questions:

- Convergence?
- Accuracy?

Adaptive input design

Key questions:

- Convergence?
- Accuracy?

Theorem (Gerencsér's free lunch theorem for ARX-models)

- *True system in the model set*
- *System stable*

$\Rightarrow$ *Optimality when sample size grows*

Gerencsér

Adaptive input design

What happens when true system is not in the model set?

Example

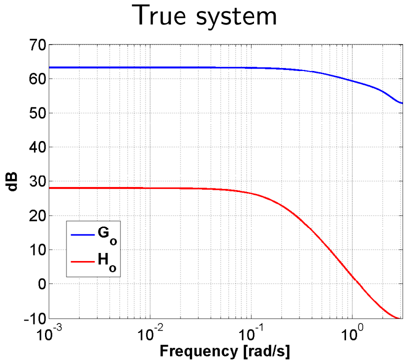
NMP-zero estimation

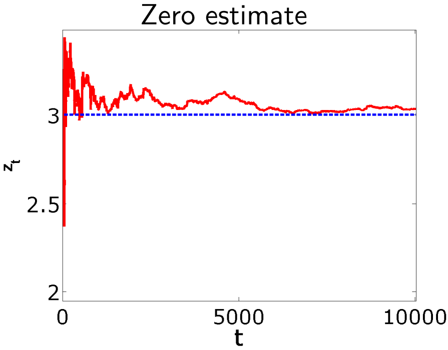
- Quantity of interest: z_o: $G_o(z_o) = 0$, $|z_o| > 1$
- Optimal input: $u_t = \frac{c}{z^{-1} - z_o} w_t$
- V_{id} and V_{app} not matched (c.f. impulse response problem)
- Still $y_t = \theta_1 u_t + \theta_2 u_{t-1} \Rightarrow$ consistent estimate

Example: Non-minimum phase zero estimation

$$\text{True system: } y_t = \frac{(q-3)(q-0.1)(q-0.2)(q+0.3)}{q^4(q-0.5)} u_t + \frac{q}{q-0.8} e_t^o$$

$$\text{Model: } y_t = \frac{\theta_1 q + \theta_2}{q^2} u_t + e_t$$





Example: Non-minimum phase zero estimation

Theorem (Rojas and Gerencsér)

$$\textit{True system: } y_t = G_o(q)u_t + H_o(q)e_t^o$$

with G_o and H_o stable and rational.

$$\hat{z}_t \longrightarrow \textit{ largest NMP-zero of system w.p.1}$$

What have we learnt?

- A framework for quantification of the experimental cost where the application is taken into account
- Allows the user to make trade-offs
- The optimal experiment matches the identification criterion to the performance degradation using parsimonious excitation
  (The let sleeping dogs lie paradigm)
- Simplifies the identification problem
- Adaptive input design practical implementation
- Focus on the application!

Future directions:

- Nonlinear systems
- Structured systems (e.g. decentralized and networked)
- Communication systems
- Adaptive control

A lot of exciting problems remain!!!