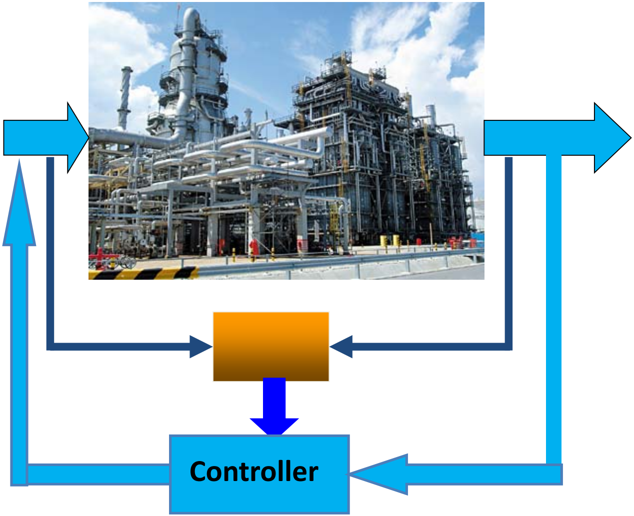
FUNDAMENTALS OF DATA DRIVEN CONTROL DESIGN

Håkan Hjalmarsson

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

IST, Universität Stuttgart
December 6, 2016

The problem



Ogunnaike 1996:

... obtaining the process model is the single most time consuming task in the application of model based control

I'm afraid this still describes state-of-the art....

My assistants

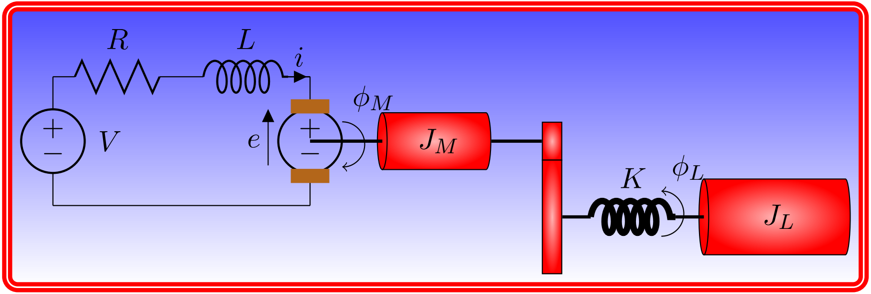
Please say hi to

Izzy

and

Ozzy

Preview: MPC of a DC-motor

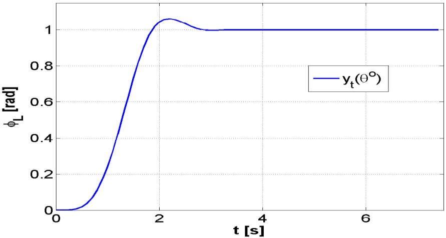


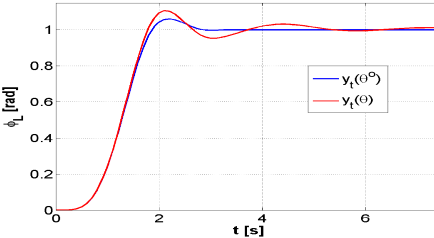
- Input: Voltage V
- Output: Angle ϕ_L
- Model parameters θ: Resistance R, Moment of inertia J_L, Elasticity K, ...
- True parameters: θ^o

Preview: 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



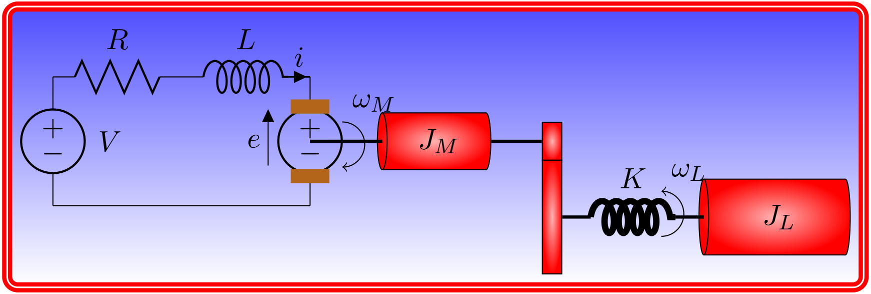


Performance degradation /Set of acceptable models

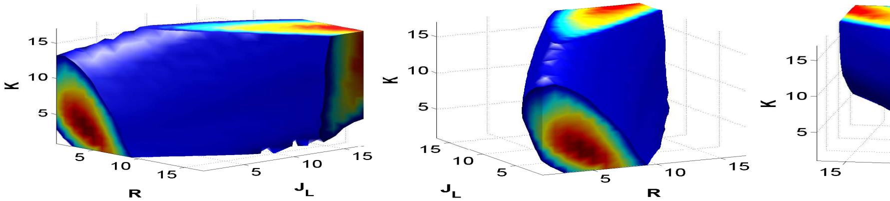
$$V_{app}(\theta) = \frac{1}{N}\sum_{t=1}^{N}(y_t(\theta^o) - y_t(\theta))^2$$

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

Preview: MPC of a DC-motor



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

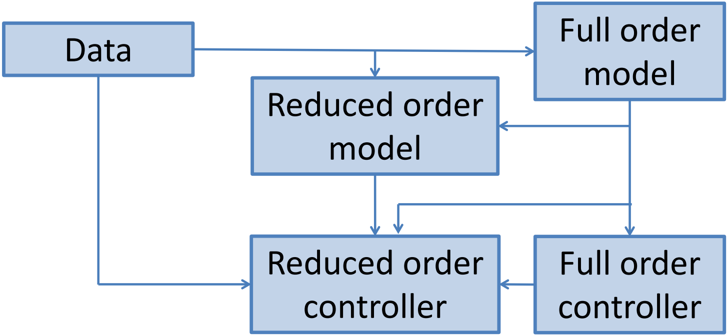


We want $\hat{\theta}_N \in \mathcal{E}$

The issues involved

- The shape of the set of acceptable controllers: Robust control
  - The forgiving nature of feedback
  - Fundamental limitations of feedback
- How to construct the map Data $\to$ Controller
- How to quantify errors (noise) in data
- How to generate the data

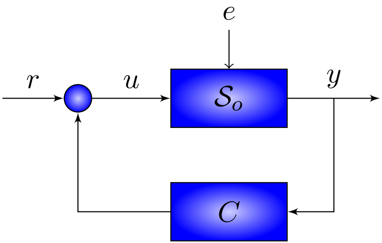
How to construct the map Data $\rightarrow$ Controller



Which path to take???

How to construct the map Data $\rightarrow$ Controller

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: $\mathcal{S}_o \Leftrightarrow \theta^o$
- High accuracy γ (implies large sample size N)
- $\sqrt{N}\left(\hat{\theta}_N - \theta^o\right) \sim \mathrm{As}\mathcal{N}\left(0, 2\lambda_e V_{id}''(\theta^o)^{-1}\right)$

How to construct the map Data $\rightarrow$ Controller

A lot of work in the past 30 years

Linear model identification:

$$y_t = G(q,\theta)u_t + H(q,\theta)e_t$$
$$\varepsilon_t(\theta) = H^{-1}(q,\theta)(y_t - G(q,\theta)u_t)$$
$$V_{id} = \mathrm{E}[\varepsilon_t^2(\theta)] - \lambda_e = \frac{1}{2\pi}\int_{-\pi}^{\pi} |G_o(e^{i\omega}) - G(e^{i\omega},\theta)|^2 \frac{\Phi_u(\omega)}{|H(e^{i\omega},\theta)|^2}$$

(defines estimate as $N \rightarrow \infty$)

Model reference control (MRC):

Reference model: $y_d = T(q)r$

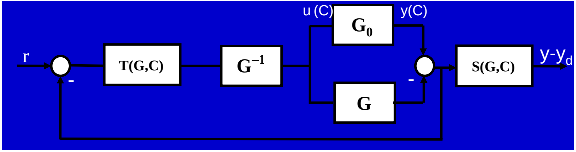
MRC: $C(G(\theta)) = \frac{1}{G(\theta)}\frac{T}{1-T}$

$$\min_{\theta} \mathrm{E}[(y(C(G(\theta))) - y_d)^2]$$

How to construct the map Data → Controller

$$\varepsilon_t(\theta) = H^{-1}(q,\theta)(y_t - G(q,\theta)u_t)$$

$$\min_\theta \mathrm{E}[(y(C(G(\theta))) - y_d)^2]$$



$$y(C(G(\theta))) - y_d = S(G(\theta), C(G(\theta)))(y(C(G(\theta))) - G(\theta)u(C(G(\theta))))$$

Tuning of the bias error:

- Prefiltering (Rivera et al)
- Model reference control (Karimi et al)
- VRFT (Campi et al) - Direct model reference control $G(\theta) = T/(1-T)/C(\theta)$
- Iterative closed loop techniques (Zang/Gevers/Bitmead, Schrama/Van den Hof)

How to construct the map Data $\rightarrow$ Controller

Let's take a statistical approach:

Suppose that

$$y_t = f_t(u_{1:t}, e_{1:t}, \theta),\ t = 1, 2, \ldots,\ e_t \text{ white noise}$$

Then the Cramér-Rao lower bound (CRLB)

$$\text{Cov}\, \hat{\theta} \geq I_F^{-1}$$

applies to every unbiased estimate $\hat{\theta}$ of θ.

Here I_F is the Fisher information matrix.

But also

$$\text{Cov}\, \widehat{C(\theta)} \geq C'(\theta) I_F^{-1} [C'(\theta)]^T$$

for any unbiased estimate $\widehat{C(\theta)}$ of the function $C(\theta)$.

How to construct the map Data $\to$ Controller

The ML (Maximum Likelihood) estimate, $\hat{\theta}_{ML}$ say,
typically achieves the CRLB, at least as the sample size grows.

But it also holds that:

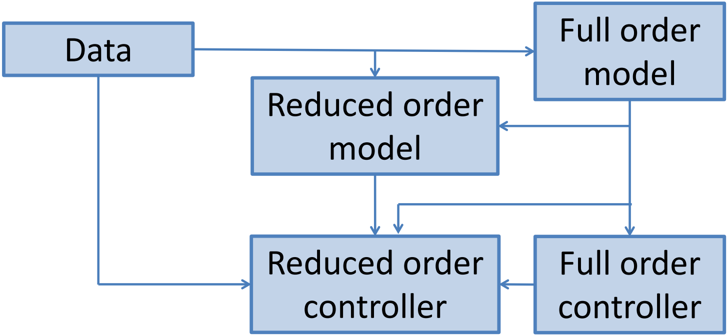
$C(\hat{\theta}_{ML})$ is the ML estimate of $C(\theta)$,
and thus also typically achieves the CRLB for $C(\theta)$!

How to construct the map Data $\rightarrow$ Controller

Translation into control terms:

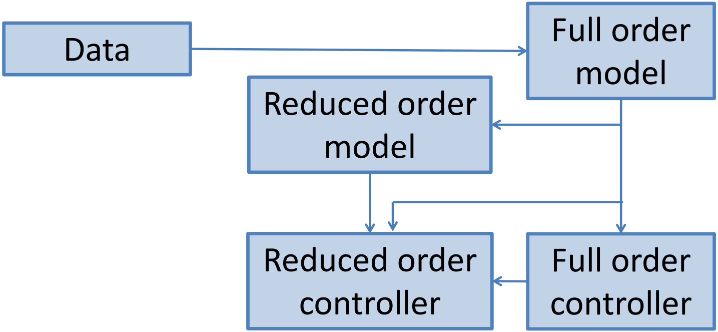
- θ represents the true underlying system
- $C(\theta)$ represents the desired controller, not necessarily full order but it is a function of the underlying system (as defined by some criterion)
- ML estimate of the desired controller, obtained by
  1) First computing the ML estimate $\hat{\theta}_{ML}$ of the entire system
  2) Then computing the certainty equivalence controller $C(\hat{\theta}_{ML})$
- This controller has optimal statistical performance, at least when the sample size is large
- This is *the whatever you do, I can do better* theorem

How to construct the map Data $\rightarrow$ Controller



Which path to take???

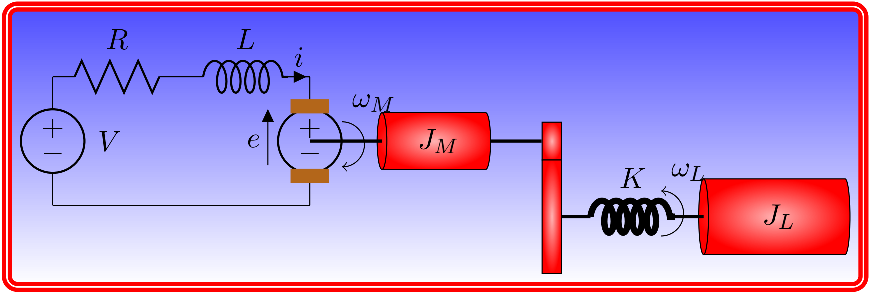
How to construct the map Data → Controller

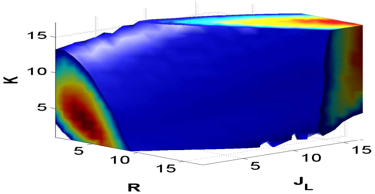


Many paths can be taken but start with as good model as possible

- Clean away as much noise first as possible
- Enables correct uncertainty characterization

How to quantify errors (noise) in data





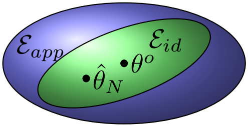
We want $\hat{\theta}_N \in \mathcal{E}_{app}$

How to quantify errors (noise) in data

- 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

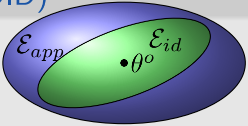
Desired situation:



How to generate the data

Application oriented experiment design (AOID)

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



- $Q =$ energy required in the identification experiment.
- Called the cost of complexity
- Least-costly identification (Bombois et al. Automatica 2006.)

An alternative formulation of AOID

Application oriented experiment design (AOID)

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

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

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

Illustration of AOID: Output error models

Applications oriented experiment design (AOID)

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

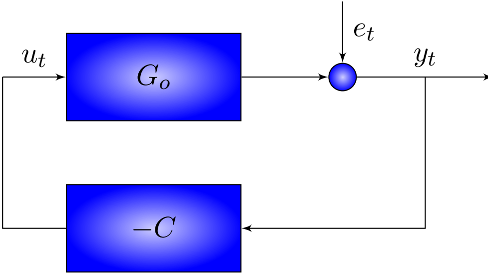
$$s.t. \quad \underbrace{NV_{id}(\theta)}_{\frac{1}{2\pi}\int_{-\pi}^{\pi} N\Phi_u^{id}(e^{j\omega})\,|G(e^{j\omega},\theta)-G_o(e^{j\omega})|^2\,\mathrm{d}\omega} \geq \lambda_e \gamma n V_{app}(\theta)$$

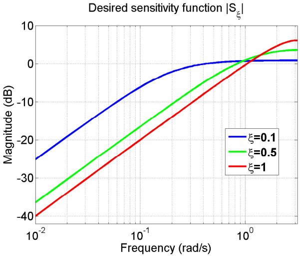
- Minimization with respect to energy density spectrum $N\Phi_u^{id}$
- Optimization tries to achieve

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

Identification cost matched to performance degradation

Izzy and Ozzy goes to MRC (model reference control)





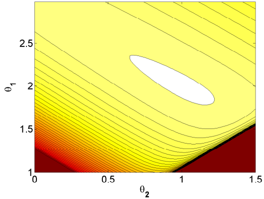
- Controller $C = C(G)$, G output error model
- Desired sensitivity function: S_ξ
- Achieved sensitivity function: $S(G) = \frac{1}{1+C(G)G_o}$
- Performance degradation: $V_{app}(G) := \left\| \frac{S(G)-S_\xi}{S_\xi} \right\|_2^2$

MRC: The impact of the performance specs.

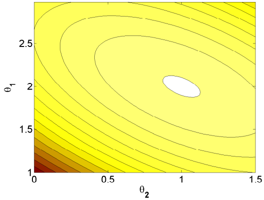
$$y_t = G_o(q)u_t + e_t = \theta_1 u_t + \theta_2 u_{t-1} + e_t$$

- Low bandwidth $\xi = 0.025$. Mainly static gain important:

V_{app}

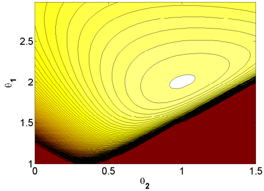


V_{id}

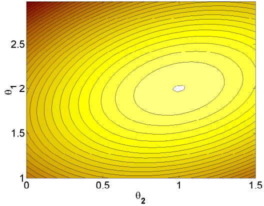


- High bandwidth $\xi = 1$. Entire frequence response important:

V_{app}



V_{id}



MRC: Cost of complexity

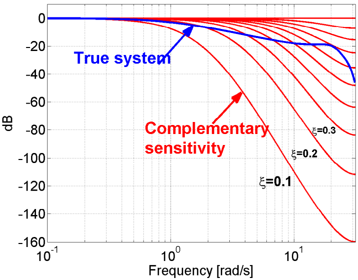
$$Q := \min N\mathrm{E}[u^2(t)]$$
$$s.t.\ NV_{id}(\theta) \geq \lambda_e \gamma n V_{app}(\theta)$$

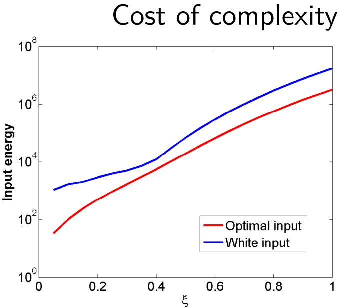
- Matching condition: $V_{id}(\theta) = V_{app}(\theta)$
- Output error:
  $NV_{id}(\theta) = \frac{1}{2\pi}\int_{-\pi}^{\pi} N\Phi_u^{id}(e^{j\omega})\left|G(e^{j\omega},\theta) - G_o(e^{j\omega})\right|^2 \mathrm{d}\omega$
- MRC:
  $V_{app}(G) := \left\|\frac{S(G)-S_\xi}{S_\xi}\right\|_2^2 \approx \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{\Phi_u^{desired}(e^{j\omega})}{\lambda_e}\left|G - G_o\right|^2 d\omega$
- Take $N\Phi_u^{id} = \gamma n \Phi_u^{desired}$
  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

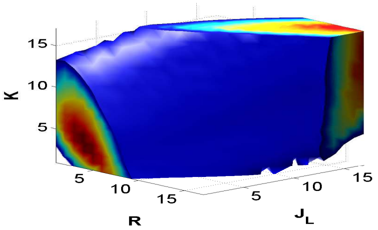




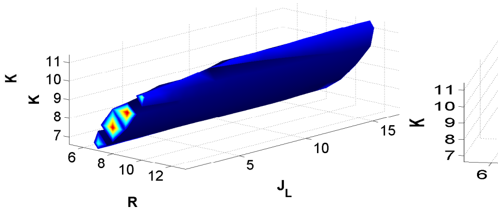
Further insights: The MPC example

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

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

Outline

How to construct the map Data $\rightarrow$ Controller

How to quantify errors (noise) in data

How to generate the data

Some connections to the past

Identification experiment = desired closed loop operating conditions:

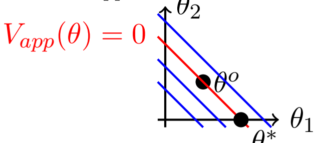
- Random errors
  - Minimum variance control (Gevers and Ljung 1986, Hjalmarsson, Gevers and De Bruyne 1996, Hildebrand and Solari 2007, Mårtensson, Rojas and Hjalmarsson 2009)
- Bias errors
  - Many contributions in the 1990s to identification for control, e.g.:
  - Control-relevant prefiltering (Rivera, 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 impact of AOID on the sysid problem

Static gain estimation

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

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

Applications oriented input design: Summary

AOID

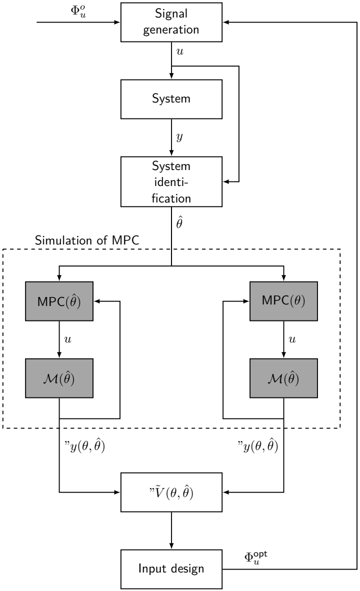
Aims at achieving

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

using minimum input energy

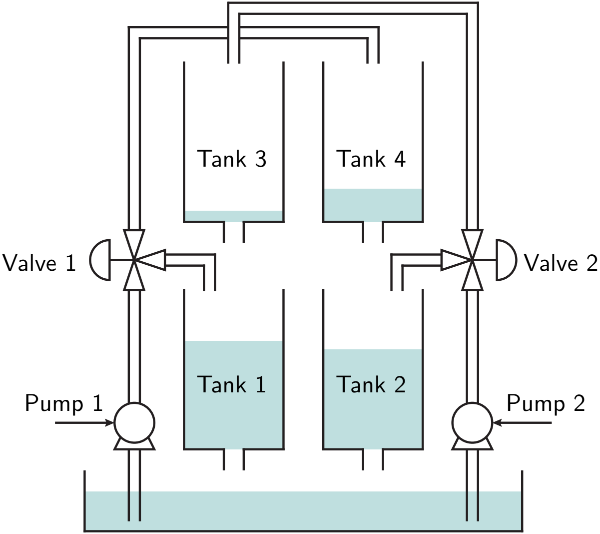
- To achieve this requires *parsimonious excitation*:
  i) System properties important to the application should be visible in the data
  ii) System properties not important to the application should not be visible in the data, unless necessary for i).
  (The let sleeping dogs lie paradigm)
- As a result, the entire system may not have to be identified!
  - The identification criterion measures V_{app} (modulo scaling).
  - 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).

AOID for MPC

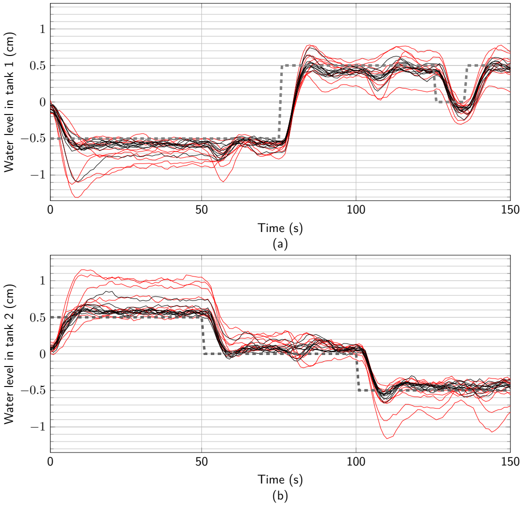


Φ_u^o
Signal generation
u
System
y
System identi- fication
$\hat{\theta}$
Simulation of MPC
MPC($\hat{\theta}$)
MPC(θ)
u
u
$\mathcal{M}(\hat{\theta})$
$\mathcal{M}(\hat{\theta})$
"$y(\theta, \hat{\theta})$
"$y(\theta, \hat{\theta})$
"$\tilde{V}(\theta, \hat{\theta})$
Input design
Φ_u^{opt}

AOID for MPC: Water tank labprocess



Response in 20 experiments:



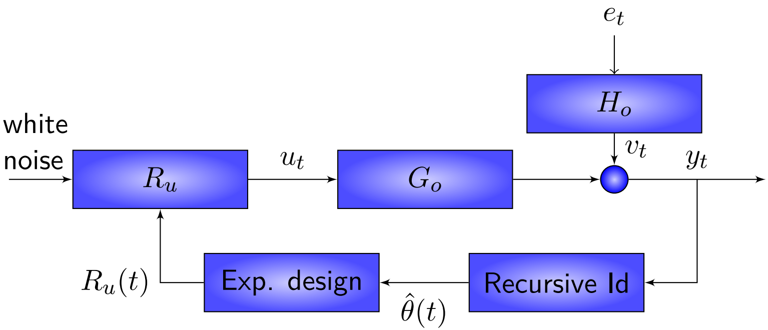
Implementation of AOID

Cost of complexity

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

- 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*

Adaptive input design

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

Example

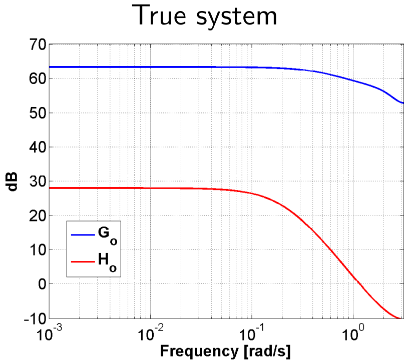
NMP-zero estimation

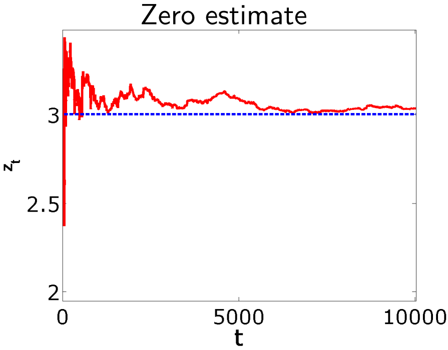
- Quantity of interest: z_o: $G_o(z_o) = 0$, $|z_o| > 1$
- Optimal input: $u_t = \frac{c}{z^{-1} - z_o} w_t$
- V_{id} and V_{app} not matched (c.f. impulse response problem)
- Still $y_t = \theta_1 u_t + \theta_2 u_{t-1} \Rightarrow$ consistent estimate

Example: Non-minimum phase zero estimation

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

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





Adaptive input design

MPC-X - Model Predictive Control with eXperiment design

- Add "$\mathcal{E}_{id} \subset \mathcal{E}_{app}$" as (final) constraint to standard MPC
- Related work: Persistence of exciation (Marafioti/Bitmead, Rathousky/Havlena et al)
- Tested on a depropanizer (SASOL, Sekunda, South Africa)

Izzy and Ozzy interrupts

- But hey, through the entire talk you have assumed that the true system is in the model set. This is never the case!

- Yeah, you have been pulling this spiel for ten years now. Surely you can do better, or?

- Yeah, what about this factor little n? Not so little in reality, no?

Recall that n is the number of estimated parameters, the complexity of the model.
The experimental cost is proportional to n....

Is there a problem?

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

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

(output error models)

A fundamental limitation of full order models

In fact

There's only one system!

Hmm, n seems to be pretty big ($= \infty$?).

How handle $\frac{1}{2\pi}\int_{-\pi}^{\pi} \text{Cov}\, G(e^{i\omega}, \hat{\theta}_N)\frac{\Phi_u(\omega)}{|H_o(e^{i\omega})|^2}d\omega = \infty$???

Handling complexity

The situation is clearly absurd.

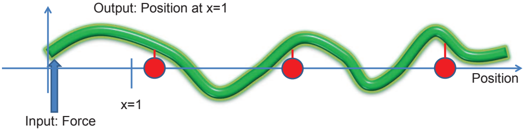
Consider an exponentially stable linear system that is not finite dimensional.

The preceeding theory cannot cope with this type of system, although this is the type of system that we model and control every day (I suspect).

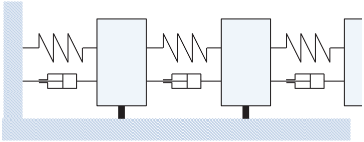
Handling complexity

- Split model in two parts:
  - “Core” with, say, n parameters
    (The part that has a large impact on the application - what we usually call the model)
  - “Dust”, the left-overs (that may have many parameters)
- Estimate “Core”+”Dust”, i.e. entire system (as before), using the data
  - Error in “Core” model of the same type as we have used (with some minor modifications)
  - The error $\int_{-\pi}^{\pi} |''\text{true dust}''|^2$ can be estimated from the “dust” model, with accuracy depending on the experimental conditions.
- Implications for AOID:
  - $\mathcal{E}_{app}$ should have two parts: “Core”$\subset \mathcal{E}_{app}$ & $\int_{-\pi}^{\pi} |''\text{true dust}''|^2 \leq \delta$.
  - The experiment design tries to:
    - Emphasize “Core”
    - Bound the “Dust” integral (below δ)
    - Minimize the correlation between the two
  - Performance requirements determine n!

Philosphical intermission: The Chained Snake

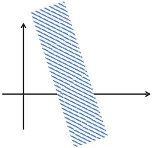


The Chained Snake – A simplified analog



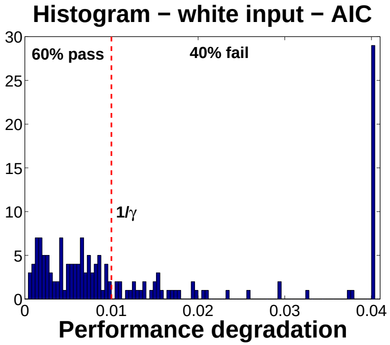
- The more excitation, the more complex behaviour
- Little excitation, essentially a second order system
- $n = 2$. A simple problem!
- The n we talk about pertains to the parts of the system that are excited

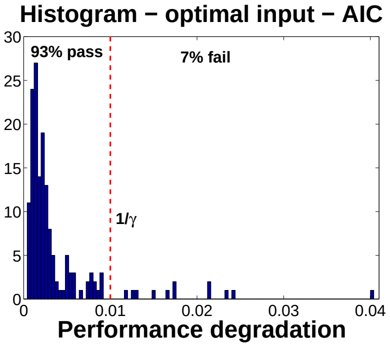
Confidence ellipsoids always look like



Another aspect of model complexity: Model selection

- 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:





Summary: What have we learnt?

- The *whatever you do, I can to better* theorem, or always first model as well as possible.
  Caveat: Result asymptotic in sample size. For finite simple size there may be better methods.
- A framework for applications oriented experiment design (AOID)
- Allows the user to make trade-offs between end-performance and experimental effort
- The *let sleeping dogs lie* paradigm
  - The optimal experiment matches the identification criterion to the performance degradation using parsimonious excitation
  - Simplifies the identification problem
- Adaptive input design practical implementation. e.g. MPC-X
- *Core and dust* modeling
  - Handle on system complexity
  - Model order determined by performance requirement and prior knowledge

Summary: What we haven't learnt?

- The world abound with data. But process data is always in closed loop. Very hard to handle in identification. (Non-stationary) disturbances causing inverse controller responses. Still biggest challenge.
- Development of regularization based algorithms has exploded.
- Interesting results using Bayesian non-parametric methods, e.g. Gaussian Processes (yet for me to understand how to understand what is meant by the posterior distribution). Also linked to previous item.
- New potent algorithms for estimating MIMO systems emerging (Weighted Null Space Fitting & Model Order Reduction Steiglitz-McBride)
- Impressive developments in nonlinear identification for systems with process noise