On Optimal Input Design in System Identification for Control

Bo Wahlberg & Håkan Hjalmarsson

ACCESS and Automatic Control Lab
KTH Royal Institute of Technology, Stockholm, Sweden

*Joint work with Mariette Annergren, Christian Larsson and Cristian Rojas

Motivation

Advanced control methods are based on
Reliable Models

Experiment $\Rightarrow$ System identification $\Rightarrow$ Model

Question: How to design experiments to obtain models that are good enough for the intended application?

- Input excitation
- Signal to noise ratio
- Experimental time

How to Specify Performance?

Parametric Model: θ, **True System:** θ^o

Cost Function: $V_{appl}(\theta)$ such that

$V_{appl}(\theta^o) = 0$, $V'_{appl}(\theta^o) = 0$

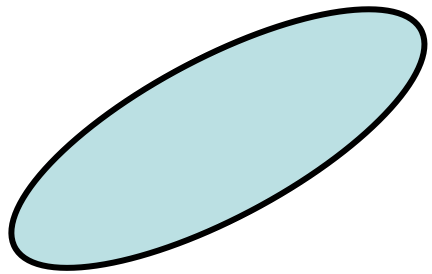
(true system optimal)

Second Order Approximation:

$$V_{appl}(\theta) \approx 0.5[\theta - \theta^o]^T V''_{appl}(\theta^o)[\theta - \theta^o]$$

Specification: $V_{appl}(\theta) \leq \frac{1}{\gamma}$

Set of admissible/good models $\Rightarrow$



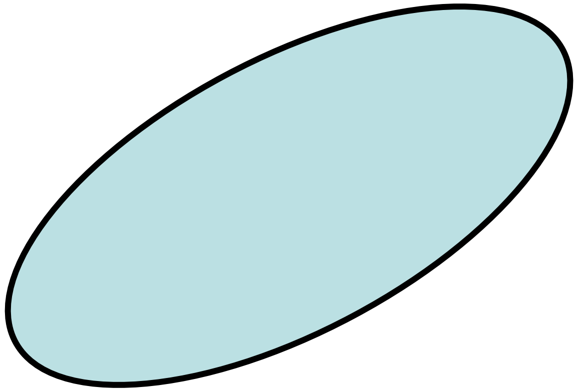
Typical Example

Closed Loop Step Response with Controller $C(\theta)$:

$$V_{appl}(\theta) = \frac{1}{N} \sum_{t=1}^{N} [y(t, C(\theta)) - y(t, C(\theta^o))]^2$$

The set $V_{appl}(\theta) \leq \frac{1}{\gamma}$ can be quite complex. Related to robust control.

Tight specifications $\Rightarrow$
Ellipsoidal approximation



Finite Impulse Response Model

Input/Excitation Signal: $u(t)$

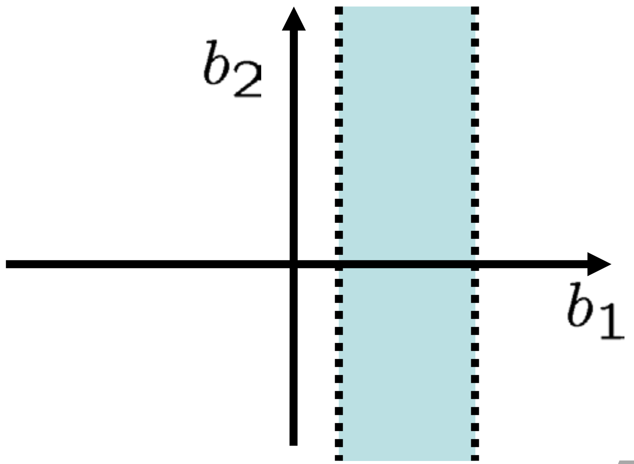
Measured Output Signal: $y(t)$

FIR Model:

$$y(t) = b_1 u(t-1) + b_2 u(t-2), \qquad \theta = (b_1 \; b_2)^T$$

Specification: $V_{appl}(\theta) = [b_1 - b_1^o]^2 \leq \frac{1}{\gamma}$

(for pedagogical use)



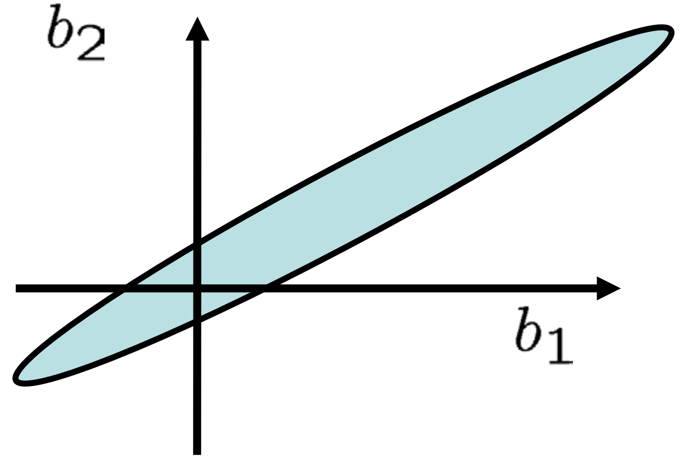
Control FIR Example

System:

$$y(t) = b_1^o u(t-1) + b_2^o u(t-2) + d(t)$$

Control objective: Reject a constant disturbance d

Feed-forward: $u(t) = \dfrac{-d(t)}{b_1 + b_2}$



Estimate the static gain $b_1^o + b_2^o$ *well!*

P-control FIR Example

Minimum phase system:

$$y(t) = b_1^o u(t-1) + b_2^o u(t-2) + d(t), \qquad b_2^o / b_1^o < 1$$

P-control $\quad u(t) = -K y(t) \qquad$ (high gain)

Closed loop stability condition: $\quad K < \dfrac{1}{b_1^o - b_2^o}$

$$K = \frac{\beta^2}{\beta b_1 - b_2}, \qquad 0 < \beta < 1, \qquad \Rightarrow$$

Nominal closed loop poles: $\quad -\beta, \quad \dfrac{-\beta b_2}{\beta b_1 - b_2}$

We need a good estimate of $\beta b_1^o - b_2^o$*!*

SI / Parameter Estimation in Two Slides

SI Cost Function: $V_{SI}(\theta, Data(N))$

PEM/ML Estimate: $\hat{\theta} = \arg\min V_{SI}(\theta, Data(N))$

For large data records ($N \rightarrow \infty$): $V_{SI}(\theta^o) = \lambda$, $V'_{SI}(\theta^o) = 0$

Second order approximation:

$$V_{SI}(\theta, Data(N)) \approx \lambda + 0.5[\theta - \theta^o]^T V''_{SI}(\theta^o)[\theta - \theta^o]$$

Average Fisher's Information Matrix,

Cramér-Rao Lower Bound

Key Result

Asymptotically in data size N:

$$\hat{\theta} \in \left\{[\theta - \theta^o]^T V''_{SI}(\theta^o)[\theta - \theta^o] \leq 2/\kappa\right\}, \; w.p.\alpha$$

where $\kappa = N/(\lambda\, \chi_2(\alpha, \mathrm{dim}[\theta]))$

Compare: $\mathrm{AsCov}\{\hat{\theta}\} = \frac{\lambda}{N}[0.5 V''_{SI}(\theta^o)]^{-1}$

See Ljung: System Identification (1999)

Example: FIR Estimation

Model:

$$y(t) = b_1 u(t-1) + b_2 u(t-2), \qquad \theta = (b_1 \; b_2)^T$$

True System:

$$y(t) = b_1^o u(t-1) + b_2^o u(t-2) + e(t),$$

where $\{e(t)\}$ is white noise with variance λ

Least squares linear regression:

$$V_{SI}(\theta, N) = \frac{1}{N} \sum_{t=1}^{N} [y(t) - (b_0 1(t-1) + b_2 u(t-2))]^2 \rightarrow$$

$$V_{SI}(\theta) = \mathsf{E}\{[y(t) - (b_1 u(t-1) + b_2 u(t-2))]^2\}, N \rightarrow \infty$$

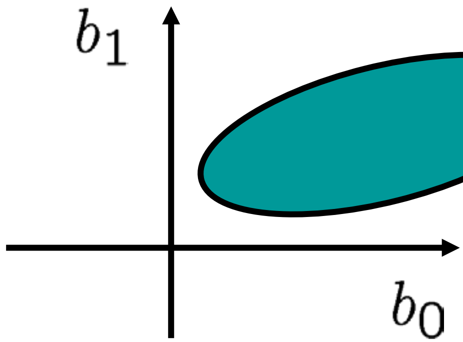
$$V''_{SI}(\theta) = 2 \begin{bmatrix} \mathsf{E}\{u^2(t)\} & \mathsf{E}\{u(t)u(t-1)\} \\ \mathsf{E}\{u(t)u(t-1)\} & \mathsf{E}\{u^2(t)\} \end{bmatrix}$$

Toeplitz Matrix

Observation: The **covariance function** r_τ, $\tau = 0, 1$, of the input signal determines the shape of the **ellipsoid:**

$$\widehat{\theta} \in \left\{ [\theta - \theta^o]^T V''_{SI}(\theta^o)[\theta - \theta^o] \leq 2/\kappa \right\}, \quad w.p.\alpha$$

$$\kappa = N/(\lambda\, \chi_2(\alpha, 2))$$



Idea

Optimize the experiment conditions
– here the excitation signal $\{u(t)\}$ –
to obtain the "*best*" possible model estimate!

- Signal to Noise Ratio
- Number of data N
- Excitation Characteristics

Classical Excitation Design Methods

Minimize the *"size"* of the covariance matrix of the parameter estimates:

$$\text{AsCov}(\hat{\theta}) = \frac{\lambda}{N}[0.5V''_{SI}]^{-1}$$

a function of the input signal $u(t)$

Quality measures: *trace, maximum eigenvalue, determinant*, ..

V''_{SI} is often an affine function in the covariance function/power spectral density of $u(t)$, while its inverse is more complicated.

Instead: Merge SI with Application!

If the **SI Set**

$$[\theta - \theta^o]^T V''_{SI}(\theta^o)[\theta - \theta^o] \le 2/\kappa$$

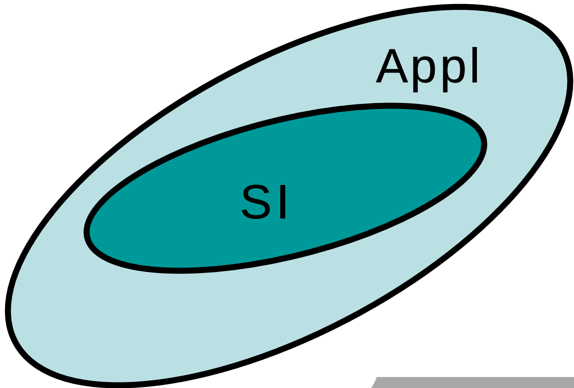
is **inside** the **Application Set**

$$[\theta - \theta^o]^T V''_{appl}(\theta^o)[\theta - \theta^o] \le 2/\gamma$$

then $\hat{\theta}$ **satisfies** the application **specifications** w.p. α.

True if $\kappa V''_{SI}(\theta^o) \ge \gamma V''_{appl}(\theta^o)$

(Matrix Inequality)



Modern Approach

$$\underset{u}{\text{minimize}}\ \mathsf{E}\{y^2 + \mu u^2\}$$

$$\text{subject to } \kappa V''_{SI}(\theta^o) \geq \gamma V''_{appl}(\theta^o).$$

Minimize the output and input power in the SI experiment subject to the application constraint.

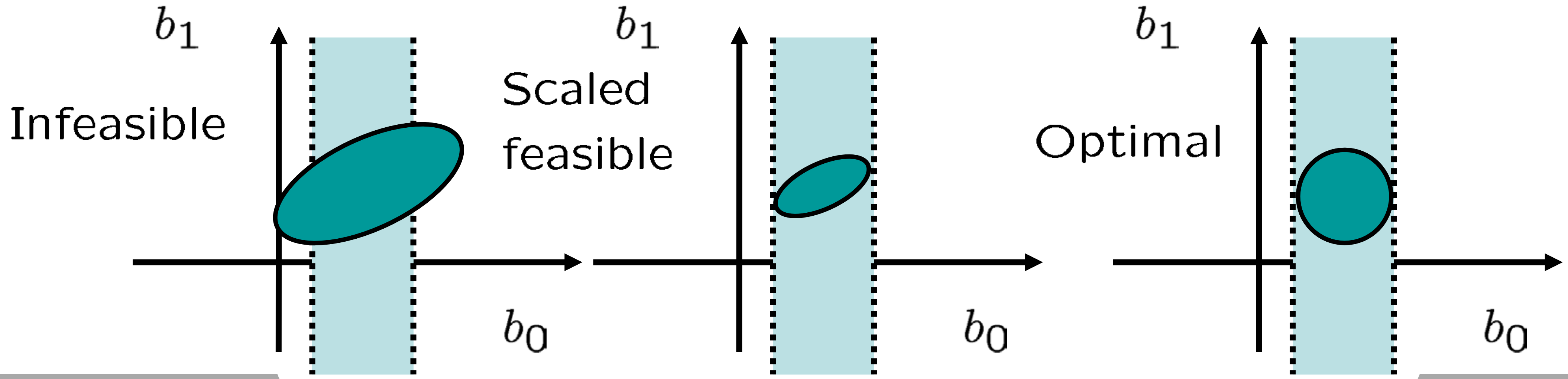
Convex optimization problem in the covariance function/power spectrum of the input signal!

FIR Example, cont.

$$\underset{u}{\text{Minimize}}\ \mathsf{E}\{u^2(t)\}$$

$$\text{subject to}\ \kappa \begin{bmatrix} \mathsf{E}\{u^2(t)\} & \mathsf{E}\{u(t)u(t-1)\} \\ \mathsf{E}\{u(t)u(t-1)\} & \mathsf{E}\{u^2(t)\} \end{bmatrix} \geq \gamma \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$$

$$\mathsf{E}\{u^2(t)\} > 0,\ \mathsf{E}\{u^2(t)\} \geq \mathsf{E}\{u(t)u(t-1)\}$$



FIR Example, cont

Minimize r_0

subject to

$$r_0 \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + r_1 \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \geq \gamma/\kappa \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$$

$$r_0 > 0,\ r_0 > r_1$$

Linear Matrix Inequality,

Semidefinite Program (SDP)!

Solution:

$$r_0 = \mathrm{E}\{u^2(t)\} = \gamma/\kappa$$

$$r_1 = \mathrm{E}\{u(t)u(t-1)\} = 0$$

FIR Example

$$0.5V''_{appl} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} 1 \begin{bmatrix} 1 & 0 \end{bmatrix}$$

Rank 1: It is enough to study b_0, $(\kappa = N/(\lambda\, \chi_2(\alpha, 1)))$

ONE degree of freedom estimation problem instead of two. (Idea: Low rank approx. V''_{appl})

It is the complexity of the application that is important rather than the number of parameters to be estimated in the model!

Frequency Response Estimation

$$G(e^{i\omega}) = b_1 e^{-i\omega} + b_1 e^{-2i\omega} = \Gamma^* \theta, \quad \theta = \begin{bmatrix} b_0 \\ b_1 \end{bmatrix}, \; \Gamma = \begin{bmatrix} e^{i\omega} \\ e^{2i\omega} \end{bmatrix}$$

$G(1) = b_1 + b_2$, $(\omega = 0)$, static gain control example

$G(-1) = b_2 - b_1$, $(\omega = \pi)$, P-control example

$$\text{AsVar}(\hat{G}(e^{i\omega})) \sim \Gamma^* \begin{bmatrix} r_0 & r_1 \\ r_1 & r_0 \end{bmatrix}^{-1} \Gamma$$

Old Optimal input design problem:

$$\min_{r_0, r_1} \text{AsVar}(\hat{G}(e^{i\omega})) \quad \text{s.t.} \quad |r_1| \leq r_0 \leq 1$$

Complicated problem due to the inverse matrix!

Alternative Formulation

New equivalent optimal input design problem:

$$\min_{r_0, r_1} r_0 \quad \text{s.t.} \begin{bmatrix} r_0 & r_1 \\ r_1 & r_0 \end{bmatrix} \geq \Gamma\Gamma^*$$

Simple LMI problem, with same solution as the previous old one (up to a scaling) !

Proof using Schur complements of

$$\begin{bmatrix} R & \Gamma \\ \Gamma^* & 1 \end{bmatrix} \geq 0, \quad R = \begin{bmatrix} r_0 & r_1 \\ r_1 & r_0 \end{bmatrix}$$

Analytic Solution for n=2

All solutions have the same frequency response variance at ω, but different input powers:

$$\text{Optimal solution}: \; r_0^\star = 1 + |\sin(\omega)|$$

$$\text{Sinusiodal solution}: \; r_0 = 2, \quad \omega \neq 0, \pi$$

$$\text{Sinusiodal solution}: \; r_0 = 1, \quad \omega = 0, \pi$$

$$\text{White noise solution}: \; r_0 = 2$$

The optimal solution is up to a factor of *two* more power efficient than white noise or a sinusiodal input!

The $n = 2$ case gives necessary conditions for $n \geq 2$.

The signal powers $r_0 = 1$, $(\omega = 0, \pi)$ and $r_0 = 2$ (otherwise) for a sinusoidal signal are sufficient for all model orders n (we can only estimate two parameters)

Bound for optimal solution for all n:

$$1 + |\sin(\omega)| \leq r_0^\star \leq 2, \quad \omega \neq 0, \pi$$

A sinusiodal signal is *optimal* for $\omega_k = k\pi/n$, $k = 0, \ldots n$, where n the model order
Compare $n = 2$. Also check $\omega = \pi/2$.

Application Set Approach

$$V''_{appl} = \mathrm{Re}\{\Gamma\Gamma^*\}, \quad \text{(real valued matrix)}$$

$$\min_{r_0, r_1} r_0 \quad \text{s.t.} \begin{bmatrix} r_0 & r_1 \\ r_1 & r_0 \end{bmatrix} \geq \begin{bmatrix} 1 & \cos(\omega) \\ \cos(\omega) & 1 \end{bmatrix}$$

Optimal solution: $u(t) = \sqrt{2}\cos(\omega t)$

Not the same as the variance constraint!

C.f. Chebyshev's inequality

$$P(|y| > \epsilon) \leq \frac{1}{\epsilon^2}\mathrm{E}\{y^2\}$$

What about OE models (CDC 2010)?

Stoica & Söderström (1982) : Useful par.

$$G(q) = \frac{bq^{-1}}{1+fq^{-1}} \Rightarrow$$

$$\Psi(q) = \begin{bmatrix} (1+fq^{-1})q^{-1} \\ bq^{-1}(-q^{-1}) \end{bmatrix} \frac{1}{(1+fq^{-1})^2}$$

$$= \begin{bmatrix} 1 & f \\ 0 & -b \end{bmatrix} \begin{bmatrix} q^{-1} \\ q^{-2} \end{bmatrix} \frac{1}{(1+fq^{-1})^2}$$

Sylvester matrix $\mathcal{S}\times$ FIR delay vector $\times$ filter

Define $\quad \bar{u}(t) = \frac{1}{(1+fq^{-1})^2} u(t)$

Information matrix: $\quad \mathcal{S} \begin{bmatrix} \bar{r}_0 & \bar{r}_1 \\ \bar{r}_1 & \bar{r}_0 \end{bmatrix} \mathcal{S}^T$

Optimal input design problem (SDP)

$$u(t) = (1 + fq^{-1})^2 \bar{u}(t)$$

$$\min r_0 = \begin{bmatrix} 1 & 2f & f^2 \end{bmatrix}^T \begin{bmatrix} \bar{r}_0 & \bar{r}_1 & \bar{r}_2 \\ \bar{r}_1 & \bar{r}_0 & \bar{r}_1 \\ \bar{r}_2 & \bar{r}_1 & \bar{r}_0 \end{bmatrix} \begin{bmatrix} 1 & 2f & f^2 \end{bmatrix}$$

w.r.t $\bar{r}_0, \bar{r}_1, \bar{r}_2$ s.t.

$$\begin{bmatrix} \bar{r}_0 & \bar{r}_1 \\ \bar{r}_1 & \bar{r}_0 \end{bmatrix} \geq \mathcal{S}^{-1} V''_{appl} \mathcal{S}^{-T}$$

$$\begin{bmatrix} \bar{r}_0 & \bar{r}_1 & \bar{r}_2 \\ \bar{r}_1 & \bar{r}_0 & \bar{r}_1 \\ \bar{r}_2 & \bar{r}_1 & \bar{r}_0 \end{bmatrix} \geq 0$$

Time Domain Data

- Asymptotic properties only depends on the input spectrum/ second order statistics

- Inputs are often amplitude constrained.

Many possible time realization,
e.g. filtered white noise

Example: Markov Chain Input

States: $u(t) \in \{C, \; -C\}$

Transition Probability Matrix:

$$\Pi = \begin{bmatrix} p & 1-p \\ 1-p & p \end{bmatrix}$$

Power Spectrum (AR):

$$\Phi(\omega) = \frac{C^2(1-\alpha^2)}{(e^{i\omega}-\alpha)(e^{-i\omega}-\alpha)}, \qquad \alpha = 2p-1$$

Explore Adaptivity in Input Design!

The optimal solution depends on the **true** system.

Here can concepts and results from iterative and adaptive control be very useful.

Robust min/max solution
Rojas et.al. *Automatica* 2007

H_∞-norm estimation,
Wahlberg et.al. *Automatica* 2010

Conclusions

Input design for SI is an established area with many important contributions: Goodwin, Payne, Ljung, Gevers, Hildebrand, Bombois,..

Many new results based on convex optimization methods.

Consider the APPLICATION of the model!

SI for MPC (ongoing project)