Time-domain design of experiments for linear and nonlinear systems

Marco Forgione

TU Delft — Delft University of Technology

Delft Center for Systems and Control

Promotor:
Paul van den Hof

Daily supervisor:
Adrie Huesman

30th Benelux Meeting on Systems and Control

Introduction

Scope of the work: fast data-driven development of mathematical models. Particular focus on process models

- medium- or large-scale models
- may be intrinsically non-linear (e.g. batch)
- severe structural and parametric uncertainties
- slow dynamics (time for advanced control available)

Example: batch cooling crystallization. Important separation and purification process in chemical and pharmaceutical sector

Introduction

Model building procedure

A rigorous procedure to develop a model from both data and knowledge. According to

G.Franceschini, S. Macchietto
Model-based design of experiments for parameter precision: State of the art
Chemical Engineering Science, Elvisier 2008

Figure: Model development procedure

Only the first step requires physical insight, the following steps constitue the central part of the procedure.

Focus on the last step: Experiment design for Parameter Precision

Introduction

Model building procedure

A rigorous procedure to develop a model from both data and knowledge. According to

G.Franceschini, S. Macchietto
Model-based design of experiments for parameter precision: State of the art
Chemical Engineering Science, Elvisier 2008

Figure: Model development procedure

Only the first step requires physical insight, the following steps constitue the central part of the procedure.
Focus on the last step: Experiment design for Parameter Precision

Design of Experiments for Parameter Precision

Overview

We have already determined a parametric model structure:

$$x_{k+1} = g(x_t, \dots, x_{t-n_a}, u_t, \dots, u_{t-n_b}; \theta^0)$$
$$y_t = x_t + e_t,\ \ e_t \text{ iid} \sim \mathcal{N}(0, \sigma_e^2)$$

- fixed and perfect process model structure (already determined)
- system in discrete time, finite-dimensional form
- fixed noise model structure (nonlinear OE)

Already quite strong hypotheses. . .

Problem: estimation of parameters θ of the model from experimental data with maximum precision.

Design of Experiments for Parameter Precision

Overview

We have already determined a parametric model structure:

$$x_{k+1} = g(x_t, \ldots, x_{t-n_a}, u_t, \ldots, u_{t-n_b}; \theta^0)$$
$$y_t = x_t + e_t,\ e_t \text{ iid} \sim \mathcal{N}(0, \sigma_e^2)$$

- fixed and perfect process model structure (already determined)
- system in discrete time, finite-dimensional form
- fixed noise model structure (nonlinear OE)

Already quite strong hypotheses. . .

Problem: estimation of parameters θ of the model from experimental data with maximum precision.

Parameter Estimation

Well-known results

Collecting an observation sequence of N samples:

$$Y = \Phi(U, \theta^0) + E_N, \qquad E_N \sim \mathcal{N}(0, \sigma^2 I_N)$$

We can define the Fisher Matrix

$$F(U, \theta^0) = \left.\frac{S^T S}{\sigma^2}\right|_{\theta=\theta^0}, \text{ with} \qquad S = \frac{\partial \Phi}{\partial \theta}$$

Any unbiased estimator $\hat{\theta}$ of θ^0 satisfies

$$\text{Var}[\hat{\theta}] \geq F^{-1}$$

Furthermore, equality is reached (at least asymptotically) by the ML estimator:

$$\theta^{ML} = \theta^{LS} = \arg\min_{\theta} V(\theta), \qquad V(\theta) = \|Y - G(U, \theta)\|_2$$

Parameter Estimation

Well-known results

Collecting an observation sequence of N samples:

$$Y = \Phi(U, \theta^0) + E_N, \qquad E_N \sim \mathcal{N}(0, \sigma^2 I_N)$$

We can define the Fisher Matrix

$$F(U, \theta^0) = \left. \frac{S^T S}{\sigma^2} \right|_{\theta=\theta^0}, \text{ with} \qquad S = \frac{\partial \Phi}{\partial \theta}$$

Any unbiased estimator $\hat{\theta}$ of θ^0 satisfies

$$\mathrm{Var}[\hat{\theta}] \geq F^{-1}$$

Furthermore, equality is reached (at least asymptotically) by the ML estimator:

$$\theta^{ML} = \theta^{LS} = \arg\min_{\theta} V(\theta), \qquad V(\theta) = \|Y - G(U, \theta)\|_2$$

Parameter Estimation

Fisher Matrix

The Fisher Matrix F is thus a measure of the amount of information contained in the data:

$$\mathrm{Var}[\hat{\theta}] \approx F^{-1}$$

Thus, if F is "large", F^{-1} is "small", and the estimated parameter has a small variance. For this reason, F is also called Information Matrix.

A well-designed experiment for parameter precision should lead to a big Fisher Matrix for the assumed model.

Parameter Estimation

Fisher Matrix

The Fisher Matrix F is thus a measure of the amount of information contained in the data:

$$\mathrm{Var}[\hat{\theta}] \approx F^{-1}$$

Thus, if F is "large", F^{-1} is "small", and the estimated parameter has a small variance. For this reason, F is also called Information Matrix.

A well-designed experiment for parameter precision should lead to a big Fisher Matrix for the assumed model.

Design of Experiments for Parameter Precision
Overview

The Optimal Experimental Design Problem can be cast into a Dynamic Optimization Problem (DOP):

$$U^0 = \arg\max_{U} f(F(U, \theta^0))$$

Accordintg to literature, different choices for $f(\cdot)$:

- $f = \det \Rightarrow$ D-optimality
- $f = \min eig \Rightarrow$ E-optimality
- $f = trace \Rightarrow$ A-optimality
- ...

Fisher Matrix Optimization

Static model

Consider the linear static model

$$y_i = \theta_0 + \theta_1 u_i + e_i, \qquad u \in [-1; 1]$$

$$Y = \overbrace{\begin{bmatrix} 1 & u_1 \\ 1 & u_2 \\ \vdots & \vdots \\ 1 & u_N \end{bmatrix}}^{\Phi(U)} \cdot \overbrace{\begin{bmatrix} \theta_0 \\ \theta_1 \end{bmatrix}}^{\theta} + E_N \Rightarrow F = \begin{bmatrix} N & \sum u_i \\ \sum u_i & \sum u_i^2 \end{bmatrix}, \det F = N \sum (u_i - \overline{u})^2$$

For N even, maximum for
$N/2$ u_i in -1,
$N/2$ u_i in $+1$

Note: awful design for model discrimination!

Fisher Matrix Optimization

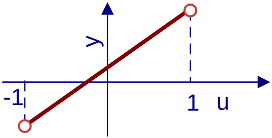
Static model

Consider the linear static model

$$y_i = \theta_0 + \theta_1 u_i + e_i, \qquad u \in [-1; 1]$$

$$Y = \overbrace{\begin{bmatrix} 1 & u_1 \\ 1 & u_2 \\ \vdots & \vdots \\ 1 & u_N \end{bmatrix}}^{\Phi(U)} \cdot \overbrace{\begin{bmatrix} \theta_0 \\ \theta_1 \end{bmatrix}}^{\theta} + E_N \Rightarrow F = \begin{bmatrix} N & \sum u_i \\ \sum u_i & \sum u_i^2 \end{bmatrix}, \det F = N \sum (u_i - \overline{u})^2$$

For N even, maximum for
$N/2$ u_i in -1,
$N/2$ u_i in $+1$



Note: awful design for model discrimination!

Fisher Matrix Optimization

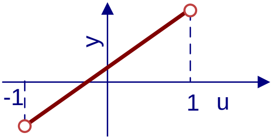
Static model

Consider the linear static model

$$y_i = \theta_0 + \theta_1 u_i + e_i, \qquad u \in [-1;1]$$

$$Y = \overbrace{\begin{bmatrix} 1 & u_1 \\ 1 & u_2 \\ \vdots & \vdots \\ 1 & u_N \end{bmatrix}}^{\Phi(U)} \cdot \overbrace{\begin{bmatrix} \theta_0 \\ \theta_1 \end{bmatrix}}^{\theta} + E_N \Rightarrow F = \begin{bmatrix} N & \sum u_i \\ \sum u_i & \sum u_i^2 \end{bmatrix}, \det F = N \sum (u_i - \overline{u})^2$$

For N even, maximum for
$N/2$ u_i in -1,
$N/2$ u_i in $+1$



Note: awful design for model discrimination!

Fisher Matrix Optimization

A simple dynamic model

A linear discrete-time FIR, N-length observation.

$$\begin{aligned} x_{t+1} &= \theta_0 u_t + \theta_1 u_{t-1} \\ y_t &= x_t + e_t \end{aligned} \Rightarrow Y = \begin{bmatrix} y_2 \\ y_3 \\ \vdots \\ y_{N+1} \end{bmatrix} = \overbrace{\begin{bmatrix} u_1 & u_0 \\ u_2 & u_1 \\ \vdots & \vdots \\ u_N & u_{N-1} \end{bmatrix}}^{\Phi(U)} \cdot \overbrace{\begin{bmatrix} \theta_0 \\ \theta_1 \end{bmatrix}}^{\theta} + E_N$$

Via straightforward computations:

$$F = \begin{bmatrix} \sum_{i=1}^{N} u_i^2 & \sum_{i=1}^{N} u_i u_{i-1} \\ \sum_{i=1}^{N} u_i u_{i-1} & \sum_{i=0}^{N-1} u_i^2 \end{bmatrix}, \det F = \sum_{i=1}^{N} u_i^2 \cdot \sum_{i=0}^{N-1} u_i^2 - \left(\sum_{i=1}^{N} u_i u_{i-1} \right)^2$$

Fisher Matrix Optimization

A simple dynamic model cont'd

$$U^o = \arg\max_U \det F = \sum_{i=1}^{N} u_i^2 \cdot \sum_{i=0}^{N-1} u_i^2 - \left(\sum_{i=1}^{N} u_i u_{i-1} \right)^2$$

Problem unbounded. We set the constraint $|u_i| \leq u_{max}$.

The maximum (for N even) is reached when

$$|u_i| = u_{max}, \quad \sum_{i=1}^{N} u_i u_{i-1} = 0$$

a Solution is

$$u = (-1)^{\lfloor \frac{n}{2} \rfloor} u_{max}$$

Note: The problem has multiple solutions.

Fisher Matrix Optimization

A simple dynamic model cont'd

$$U^o = \arg\max_U \det F = \sum_{i=1}^{N} u_i^2 \cdot \sum_{i=0}^{N-1} u_i^2 - \left(\sum_{i=1}^{N} u_i u_{i-1} \right)^2$$

Problem unbounded. We set the constraint $|u_i| \leq u_{max}$.
The maximum (for N even) is reached when

$$|u_i| = u_{max}, \quad \sum_{i=1}^{N} u_i u_{i-1} = 0$$

a Solution is

$$u = (-1)^{\lfloor \frac{n}{2} \rfloor} u_{max}$$

Note: The problem has multiple solutions.

Fisher Matrix Optimization

A simple dynamic model cont'd

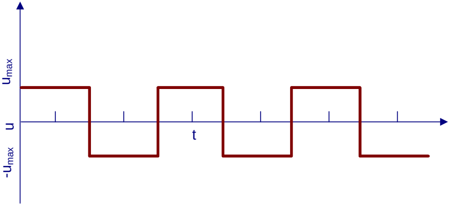
$$U^o = \arg\max_U \det F = \sum_{i=1}^{N} u_i^2 \cdot \sum_{i=0}^{N-1} u_i^2 - \left(\sum_{i=1}^{N} u_i u_{i-1}\right)^2$$

Problem unbounded. We set the constraint $|u_i| \leq u_{max}$.
The maximum (for N even) is reached when

$$|u_i| = u_{max}, \quad \sum_{i=1}^{N} u_i u_{i-1} = 0$$

a Solution is

$$u = (-1)^{\lfloor \frac{n}{2} \rfloor} u_{max}$$



Note: The problem has multiple solutions.

Fisher Matrix Optimization

A simple dynamic model cont'd

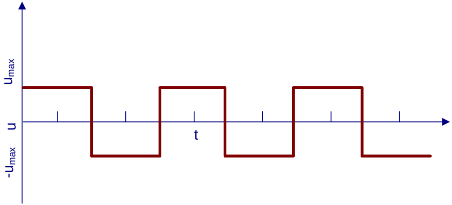
$$U^o = \arg\max_U \det F = \sum_{i=1}^{N} u_i^2 \cdot \sum_{i=0}^{N-1} u_i^2 - \left(\sum_{i=1}^{N} u_i u_{i-1}\right)^2$$

Problem unbounded. We set the constraint $|u_i| \leq u_{max}$.
The maximum (for N even) is reached when

$$|u_i| = u_{max}, \quad \sum_{i=1}^{N} u_i u_{i-1} = 0$$

a Solution is

$$u = (-1)^{\lfloor \frac{n}{2} \rfloor} u_{max}$$

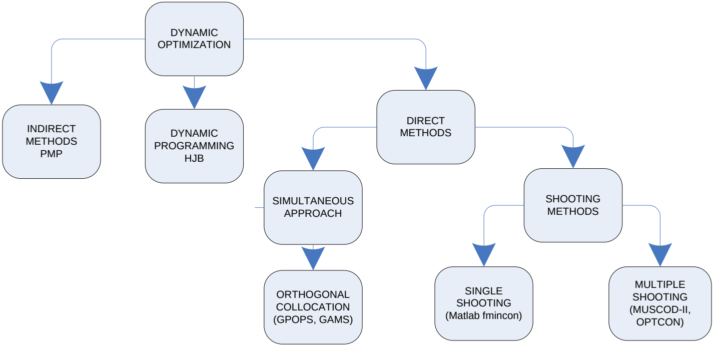


Note: The problem has multiple solutions.

Fisher Matrix Optimization

Numerical solutions

In general, the problem does not admit closed form solution.
Numerical techniques for the solution of the DOP:

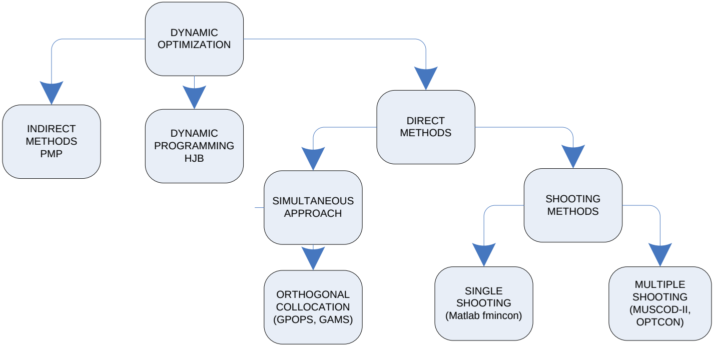


Direct methods are the most popular nowadays.

Fisher Matrix Optimization

Numerical solutions

In general, the problem does not admit closed form solution.
Numerical techniques for the solution of the DOP:



Direct methods are the most popular nowadays.

Fisher Matrix Optimization

First-order LTI example

$$x_{t+1} = g(x_t, u_t) = \theta_0 x_t + \theta_1 u_t$$
$$y_t = x_t + e_t$$

In this case, the lifted system is no more linear in θ:

$$Y = \Phi(U, \theta) + E_N$$

$$U^o = \arg\max_U \det F, \qquad F = \frac{\partial \Phi(U,\theta)}{\partial \theta}^T \frac{\partial \Phi(U,\theta)}{\partial \theta}, \qquad |u_i| \leq u_{max}$$

The objective function contains first-order derivatives $\frac{\partial \Phi}{\partial \theta}$:

- Numerical differentiation
- Sensitivity equations: $s_{t+1} = \frac{\partial x_{t+1}}{\partial \theta} = \frac{\partial g(x_t,u_t)}{\partial x_t} s_t + \frac{\partial g(x_t,u_t)}{\partial u_t}$

The second approach is prefered (notice that a gradient-based optimization requires further differentiations in u_i ...)

Fisher Matrix Optimization

First-order LTI example

$$x_{t+1} = g(x_t, u_t) = \theta_0 x_t + \theta_1 u_t$$
$$y_t = x_t + e_t$$

In this case, the lifted system is no more linear in θ:

$$Y = \Phi(U, \theta) + E_N$$

$$U^o = \arg\max_U \det F, \qquad F = \frac{\partial \Phi(U,\theta)}{\partial \theta}^T \frac{\partial \Phi(U,\theta)}{\partial \theta}, \qquad |u_i| \leq u_{max}$$

The objective function contains first-order derivatives $\frac{\partial \Phi}{\partial \theta}$:

- Numerical differentiation
- Sensitivity equations: $s_{t+1} = \frac{\partial x_{t+1}}{\partial \theta} = \frac{\partial g(x_t,u_t)}{\partial x_t} s_t + \frac{\partial g(x_t,u_t)}{\partial u_t}$

The second approach is prefered (notice that a gradient-based optimization requires further differentiations in u_i ...)

Fisher Matrix Optimization

First-order LTI example

$$x_{t+1} = g(x_t, u_t) = \theta_0 x_t + \theta_1 u_t$$
$$y_t = x_t + e_t$$

In this case, the lifted system is no more linear in θ:

$$Y = \Phi(U, \theta) + E_N$$

$$U^o = \arg\max_U \det F, \qquad F = \frac{\partial \Phi(U,\theta)}{\partial \theta}^T \frac{\partial \Phi(U,\theta)}{\partial \theta}, \qquad |u_i| \leq u_{max}$$

The objective function contains first-order derivatives $\frac{\partial \Phi}{\partial \theta}$:

- Numerical differentiation
- Sensitivity equations: $s_{t+1} = \frac{\partial x_{t+1}}{\partial \theta} = \frac{\partial g(x_t,u_t)}{\partial x_t} s_t + \frac{\partial g(x_t,u_t)}{\partial u_t}$

The second approach is prefered (notice that a gradient-based optimization requires further differentiations in u...)

Fisher Matrix Optimization

First-order LTI example cont'd

Numerical example

$$x_{t+1} = 0.8x_t + 0.2u_t$$

Implementation in Matlab `fmincon` using a shooting algorithm:

1. How to parametrize the optimal input?
2. How to initialize the estimate?

- Using a general parametrization e.g piecewise linear u, many local optima depending upon initialization.
  Strong parametrization of the input: $u_i = u_{max} \cos(\omega_c i)$
- For large N, we find $\omega_c \approx 0.13$. Same convergence starting from different point. Sensible, the bandwidth of the system is ≈ 0.2.
- Possible extension to multisine excitation.
- Using the sinusoidal as starting point with piecewise linear parametrization, we get a square wave of the same frequency.

Fisher Matrix Optimization

First-order LTI example cont'd

Numerical example

$$x_{t+1} = 0.8x_t + 0.2u_t$$

Implementation in Matlab `fmincon` using a shooting algorithm:

1. How to parametrize the optimal input?
2. How to initialize the estimate?

- Using a general parametrization e.g piecewise linear u, many local optima depending upon initialization.
  Strong parametrization of the input: $u_i = u_{max}\cos(\omega_c i)$
- For large N, we find $\omega_c \approx 0.13$. Same convergence starting from different point. Sensible, the bandwidth of the system is ≈ 0.2.
- Possible extension to multisine excitation.
- Using the sinusoidal as starting point with piecewise linear parametrization, we get a square wave of the same frequency.

Fisher Matrix Optimization

First-order LTI example cont'd

Numerical example

$$x_{t+1} = 0.8x_t + 0.2u_t$$

Implementation in Matlab `fmincon` using a shooting algorithm:

1. How to parametrize the optimal input?
2. How to initialize the estimate?

- Using a general parametrization e.g piecewise linear u, many local optima depending upon initialization.
  Strong parametrization of the input: $u_i = u_{max}\cos(\omega_c i)$
- For large N, we find $\omega_c \approx 0.13$. Same convergence starting from different point. Sensible, the bandwidth of the system is ≈ 0.2.
- Possible extension to multisine excitation.
- Using the sinusoidal as starting point with piecewise linear parametrization, we get a square wave of the same frequency.

Fisher Matrix Optimization

First-order LTI example cont'd

Numerical example

$$x_{t+1} = 0.8x_t + 0.2u_t$$

Implementation in Matlab `fmincon` using a shooting algorithm:

1. How to parametrize the optimal input?
2. How to initialize the estimate?

- Using a general parametrization e.g piecewise linear u, many local optima depending upon initialization.
  Strong parametrization of the input: $u_i = u_{max}\cos(\omega_c i)$
- For large N, we find $\omega_c \approx 0.13$. Same convergence starting from different point. Sensible, the bandwidth of the system is ≈ 0.2.
- Possible extension to multisine excitation.
- Using the sinusoidal as starting point with piecewise linear parametrization, we get a square wave of the same frequency.

Fisher Matrix Optimization

Process model example: batch crystallization

nonlinear dynamics

$$\frac{dm_0}{dt} = B$$
$$\frac{dm_j}{dt} = jGm_{j-1} + Br_0^j, \; j = 1, 2, 3$$
$$\frac{dC}{dt} = -3\rho_c k_v - \rho_c k_v B r_0^3$$

with

$$G = k_g S^g$$
$$B = k_b S^b m_3$$
$$S = C - C^*(T)$$
$$\theta = [g, \log k_g, b, \log k_b]^T$$
$$y_t = [I(m_2), C]^T$$

Temperature T is constrained to initial and final value, cooling rate is also limited. D-optimal design:

Approx 1 order of magnitude more accurate w.r.t. linear cooling.

Fisher Matrix Optimization

Process model example: batch crystallization

nonlinear dynamics

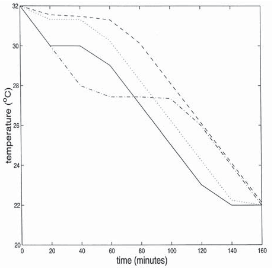
$$\frac{dm_0}{dt} = B$$
$$\frac{dm_j}{dt} = jGm_{j-1} + Br_0^j, \; j = 1, 2, 3$$
$$\frac{dC}{dt} = -3\rho_c k_v - \rho_c k_v B r_0^3$$

with

$$G = k_g S^g$$
$$B = k_b S^b m_3$$
$$S = C - C^*(T)$$
$$\theta = [g, \log k_g, b, \log k_b]^T$$
$$y_t = [I(m_2), C]^T$$

Temperature T is constrained to initial and final value, cooling rate is also limited. D-optimal design:

Approx 1 order of magnitude more accurate w.r.t. linear cooling.



Fisher Matrix Optimization

Considerations

The problem formulation is rather straightforward, but optimization problem is in general hampered by the large number of local optima.
Additional constraints might make the problem easier (many local optima are a-priori excluded).
Many examples from the process field are *per se* strongly constrained.

When the system is not linear in the parameters, the optimal design depends on the parameters. But this is what we want to estimate!

- Iterative procedure of design and identification
- Robust optimal experiment design, e.g. max-min

$$U^o = \arg\max_{U} \min_{\theta \in \Theta} f(F)$$

(additional numerical burden)

Fisher Matrix Optimization

Considerations

The problem formulation is rather straightforward, but optimization problem is in general hampered by the large number of local optima. Additional constraints might make the problem easier (many local optima are a-priori excluded).
Many examples from the process field are *per se* strongly constrained.

When the system is not linear in the parameters, the optimal design depends on the parameters. But this is what we want to estimate!

- Iterative procedure of design and identification
- Robust optimal experiment design, e.g. max-min

$$U^o = \arg\max_{U} \min_{\theta \in \Theta} f(F)$$

(additional numerical burden)

Fisher Matrix Optimization

Considerations cont'd

In general, models are used for some task (simulation, prediction, optimization, ...); the focus is not on parameter precision.
If we could define the performance objective as quadratic function:

$$\eta = (\hat{\theta} - \theta^0)^T A (\hat{\theta} - \theta^0)$$

then:

$$E[\eta] = \mathrm{tr}\, A\, \mathrm{Var}[\hat{\theta}] \qquad \text{(for } \theta^0 \text{ gaussian)}$$

One might use $E[\eta]$ as objective function for optimization.
A similar approach is the so-called D_A-optimal criterion:

$$U^o_{D_A} = \arg\min_U \det AF^{-1}A^T$$

For a non-quadratic performance objective, a Taylor expansion is straightforward.
However, one may argue about the accuracy of the approximation

Fisher Matrix Optimization

Considerations cont'd

In general, models are used for some task (simulation, prediction, optimization, . . .); the focus is not on parameter precision.
If we could define the performance objective as quadratic function:

$$\eta = (\hat{\theta} - \theta^0)^T A(\hat{\theta} - \theta^0)$$

then:

$$E[\eta] = \operatorname{tr} A \operatorname{Var}[\hat{\theta}] \qquad \text{(for } \theta^0 \text{ gaussian)}$$

One might use $E[\eta]$ as objective function for optimization.
A similar approach is the so-called D_A-optimal criterion:

$$U^o_{D_A} = \arg\min_U \det AF^{-1}A^T$$

For a non-quadratic performance objective, a Taylor expansion is straightforward.
However, one may argue about the accuracy of the approximation

Conclusions

In principle, optimization of the Fisher Matrix gives the most informative experiment, given an experimental framework. Open problems

Theoretical

- Strong assumptions on the model and the noise. Exact structure is known. Under modelling?
- Focus on parameter precision: difficult translation to model performance

Practical

- Non-convex, nonlinear optimization. Problem of local optima
- Lack of software off-the-shelf. Time waste, scarse reproducibility of the results

In practice, model-based experiment design techniques are not (yet) broadly applied.
The problems requires more attention from the control community.
Only discrete-time LTI models are well covered.

Conclusions

In principle, optimization of the Fisher Matrix gives the most informative experiment, given an experimental framework. Open problems

Theoretical

- Strong assumptions on the model and the noise. Exact structure is known. Under modelling?
- Focus on parameter precision: difficult translation to model performance

Practical

- Non-convex, nonlinear optimization. Problem of local optima
- Lack of software off-the-shelf. Time waste, scarse reproducibility of the results

In practice, model-based experiment design techniques are not (yet) broadly applied.
The problems requires more attention from the control community.
Only discrete-time LTI models are well covered.

Thank you for your attention