

Optimal experimental design for spatiotemporal system identification (Part 1)

Dariusz Uciński

Institute of Control and Computation Engineering
University of Zielona Góra

minicourse at the 9th German-Polish Conference on
Optimization: Methods and Applications
Białystok, 19 September 2026

Optimal experimental design

Optimum experimental design (OED) is a statistical methodology focused on **capturing maximum information** to build reliable predictive models across science, engineering, and industry.

In parameter estimation, OED guides the configuration of controllable input variables, the strategic placement of sensors, and the efficient allocation of experimental resources. Without such systematic planning, haphazard data collection risks substantial inefficiencies—either by probing the wrong operational domains or failing to resolve critical causal effects.

Spatiotemporal dynamics

Distributed parameter system—dynamic system whose state depends on both time and space; its model (a PDE) is known up to a vector of unknown parameters θ .

Observations—using sensors in order to estimate θ .

Structure of the minicourse

- **Part 1:** introduction to OED; sensor location problem for parameter estimation of PDEs; adaptation of OED theory to the sensor location setting
- **Part 2:** sensor selection for large-scale monitoring networks; overcoming the curse of dimensionality; design for scanning observations
- **Part 3:** design of mobile sensor trajectories

Schedule of Part 1

- 1 Introduction to optimum design
 - Linear regression
 - Optimum design

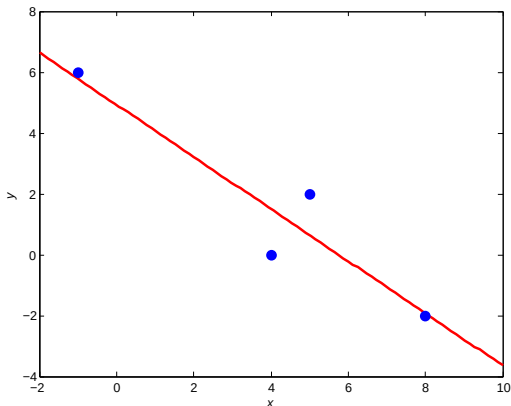
- 2 OED and optimal sensor location
 - Problem and its specificities
 - Adaptation of optimum design methodology

Linear regression

Example: For the following observations

x_i	8	4	5	-1
y_i	-2	0	2	6

we wish to find an equation describing best the linear trend:



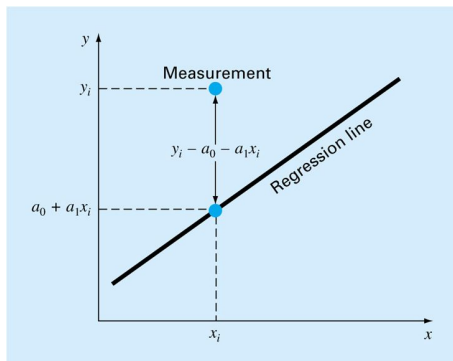
Linear regression

We approximate the response y by the linear model

$$\hat{y} = a_1x + a_0$$

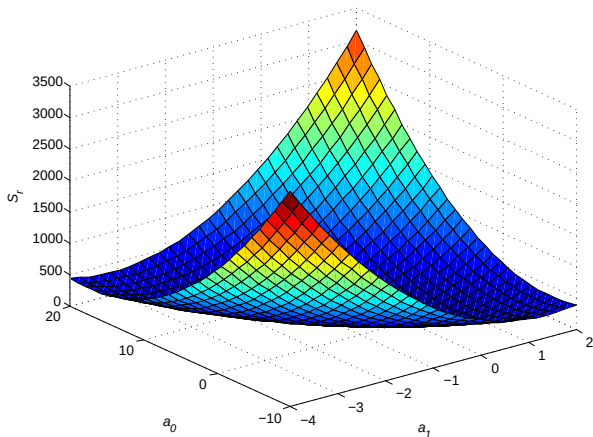
so as to minimize the individual prediction errors (residuals)

$$e_i = y_i - \hat{y}_i = y_i - a_0 - a_1x_i$$



Least-squares criterion

$$S_r = \sum_{i=1}^n e_i^2 = \sum_{i=1}^n (y_i - a_0 - a_1 x_i)^2 \rightarrow \min$$



Least-squares criterion

The optimality conditions

$$\frac{\partial S_r}{\partial a_0} = -2 \sum_{i=1}^n (y_i - a_0 - a_1 x_i) = 0$$

$$\frac{\partial S_r}{\partial a_1} = -2 \sum_{i=1}^n [(y_i - a_0 - a_1 x_i) x_i] = 0$$

yield the system of linear equations

$$\begin{cases} 0 = \sum y_i - \sum a_0 - \sum a_1 x_i \\ 0 = \sum y_i x_i - \sum a_0 x_i - \sum a_1 x_i^2 \end{cases}$$

Normal equations

After rearranging, we get the so-called *normal equations*:

$$\begin{cases} na_0 + \left(\sum x_i\right) a_1 = \sum y_i \\ \left(\sum x_i\right) a_0 + \left(\sum x_i^2\right) a_1 = \sum x_i y_i \end{cases}$$

Normal equations

After rearranging, we get the so-called *normal equations*:

$$\begin{cases} na_0 + \left(\sum x_i\right) a_1 = \sum y_i \\ \left(\sum x_i\right) a_0 + \left(\sum x_i^2\right) a_1 = \sum x_i y_i \end{cases}$$

or, in matrix form,

$$\mathbf{F}^T \mathbf{F} \boldsymbol{\theta} = \mathbf{F}^T \mathbf{y}$$

where

$$\mathbf{F} = \begin{bmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}, \quad \boldsymbol{\theta} = \begin{bmatrix} a_0 \\ a_1 \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$$

Linear regression

Its solution is $\hat{\theta} = (F^T F)^{-1} F^T y$. Specifically

$$\hat{a}_1 = \frac{n \sum x_i y_i - \sum x_i \sum y_i}{n \sum x_i^2 - (\sum x_i)^2}$$

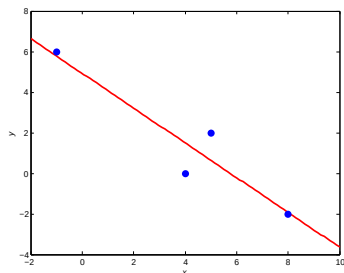
$$\hat{a}_0 = \bar{y} - a_1 \bar{x}$$

Example (ctd'):

$$a_1 = \frac{4 \cdot (-12) - 16 \cdot 6}{4 \cdot 106 - 16^2} = -0.857$$

$$a_0 = 1.5 - (-0.857)(4) = 4.929$$

$$\hat{y} = -0.857x + 4.929$$



Polynomial regression

Given data, fit a parabola:

$$\hat{y} = a_0 + a_1x + a_2x^2$$

Polynomial regression

Given data, fit a parabola:

$$\hat{y} = a_0 + a_1x + a_2x^2$$

Sum of squared residuals:

$$S_r(\boldsymbol{\theta}) = \sum_{i=1}^n (y_i - a_0 - a_1x_i - a_2x_i^2)^2 = \|\mathbf{y} - \mathbf{F}\boldsymbol{\theta}\|^2$$

where

$$\mathbf{F} = \begin{bmatrix} 1 & x_1 & x_1^2 \\ \vdots & \vdots & \vdots \\ 1 & x_n & x_n^2 \end{bmatrix}, \quad \boldsymbol{\theta} = \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$$

Polynomial regression

Given data, fit a parabola:

$$\hat{y} = a_0 + a_1x + a_2x^2$$

Sum of squared residuals:

$$S_r(\boldsymbol{\theta}) = \sum_{i=1}^n (y_i - a_0 - a_1x_i - a_2x_i^2)^2 = \|\mathbf{y} - \mathbf{F}\boldsymbol{\theta}\|^2$$

where

$$\mathbf{F} = \begin{bmatrix} 1 & x_1 & x_1^2 \\ \vdots & \vdots & \vdots \\ 1 & x_n & x_n^2 \end{bmatrix}, \quad \boldsymbol{\theta} = \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$$

Optimality conditions are the normal equations again:

$$(\mathbf{F}^T \mathbf{F})\boldsymbol{\theta} = \mathbf{F}^T \mathbf{y}$$

Polynomial regression

Explicitly, the normal equations are

$$\begin{cases} (n)a_0 + \left(\sum x_i\right)a_1 + \left(\sum x_i^2\right)a_2 = \sum y_i \\ \left(\sum x_i\right)a_0 + \left(\sum x_i^2\right)a_1 + \left(\sum x_i^3\right)a_2 = \sum x_i y_i \\ \left(\sum x_i^2\right)a_0 + \left(\sum x_i^3\right)a_1 + \left(\sum x_i^4\right)a_2 = \sum x_i^2 y_i \end{cases}$$

It is easy to generalize the approach to

$$\hat{y} = a_0 f_0(x) + a_1 f_1(x) + a_2 f_2(x) + \cdots + a_m f_m(x)$$

where $f_0, f_1, \dots, f_m$ are linearly independent functions.

Question

How much does measurement noise influence the accuracy of the estimates? Is it possible to reduce this influence?

Measurement accuracy: intro to optimal design

The weights of objects A and B are to be measured using a **pan balance** and a set of standard weights. Each weighing measures the weight difference between objects placed in the left pan vs. any objects placed in the right pan by adding calibrated weights to the lighter pan until the balance is in equilibrium.

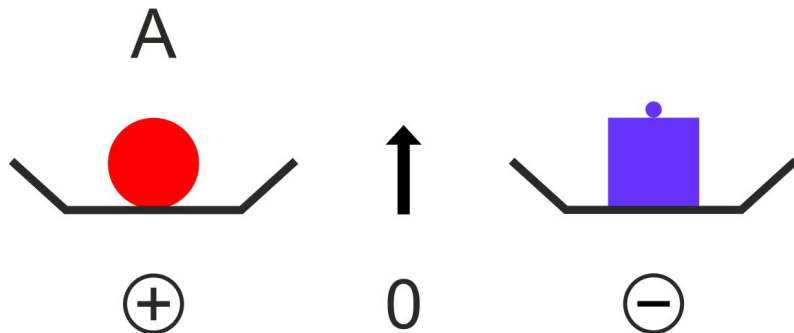


Intro to optimal design: Weighing two objects

Each measurement has a random error. The average error is zero; the standard deviations of the probability distribution of the errors is the same number σ on different weighings; and errors on different weighings are independent. Denote by m_A and m_B the true weights.

Weighing two objects

First, we put object A in one pan, with the other pan empty.



Weighing two objects

Let y_1 be its measured weight. As the error in this result is σ , we have

$$m_A = y_1 \pm \sigma$$

For object B , we perform the second measurement and get

$$m_B = y_2 \pm \sigma$$

For example, if y_1 and y_2 are 10 and 25 grammes, respectively, and σ is 0.1 gramme, we have

$$m_A = 10 \pm 0.1 \text{ grammes}$$

$$m_B = 25 \pm 0.1 \text{ grammes}$$

Weighing two objects

We can express these results using the matrix notation

$$\mathbf{y} = \mathbf{F}\boldsymbol{\theta} + \boldsymbol{\epsilon}$$

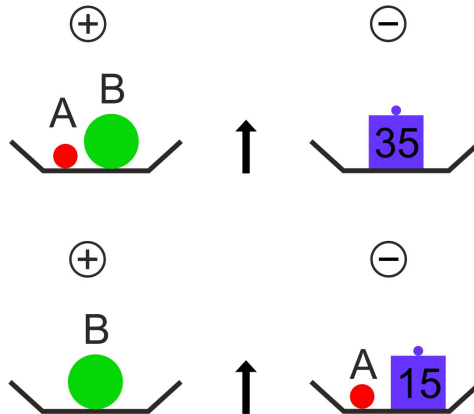
where

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}, \quad \mathbf{F} = \begin{bmatrix} +1 & 0 \\ 0 & +1 \end{bmatrix}, \quad \boldsymbol{\theta} = \begin{bmatrix} m_A \\ m_B \end{bmatrix}, \quad \boldsymbol{\epsilon} = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \end{bmatrix}$$

Note that we have $\det(\mathbf{F}^\top \mathbf{F}) = 1$.

Weighing two objects

But consider another strategy with the same experimental effort (i.e., two measurements):



Weighing two objects

Neglecting measurement errors, we get

$$\begin{cases} +m_A + m_B = y_1 \\ -m_A + m_B = y_2 \end{cases} \Rightarrow \begin{cases} m_A = \frac{1}{2}(y_1 - y_2) \\ m_B = \frac{1}{2}(y_1 + y_2) \end{cases}$$

Weighing two objects

Neglecting measurement errors, we get

$$\begin{cases} +m_A + m_B = y_1 \\ -m_A + m_B = y_2 \end{cases} \Rightarrow \begin{cases} m_A = \frac{1}{2}(y_1 - y_2) \\ m_B = \frac{1}{2}(y_1 + y_2) \end{cases}$$

What is the error for these estimates?

$$\begin{aligned} \text{var}(m_A) &= \frac{1}{4} [\text{var}(y_1) + \text{var}(y_2)] \\ &= \frac{1}{4} [\sigma^2 + \sigma^2] = \frac{1}{2}\sigma^2 \end{aligned}$$

It was reduced from σ to $\sigma/\sqrt{2}$:

$$m_A = 10 \pm 0.07 \text{ grammes}$$

$$m_B = 25 \pm 0.07 \text{ grammes}$$

Weighing two objects

Additionally, observe that now

$$\mathbf{F} = \begin{bmatrix} +1 & +1 \\ -1 & +1 \end{bmatrix}$$

and $\det(\mathbf{F}^\top \mathbf{F}) = 4$.

Weighing four objects: Strategy 1

Consider now weighing four objects A , B , C and D . First, we weigh each of them individually and get

$$m_A = y_1 \pm \sigma$$

$$m_B = y_2 \pm \sigma$$

$$m_C = y_3 \pm \sigma$$

$$m_D = y_4 \pm \sigma$$

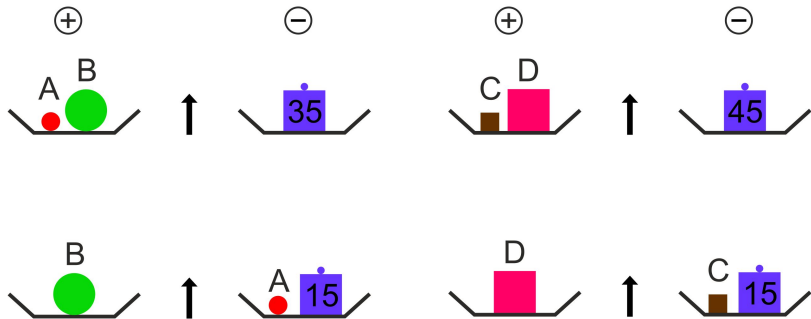
Suppose that $y_1 = 10$, $y_2 = 25$, $y_3 = 15$ and $y_4 = 30$.
Observe that

$$\mathbf{F} = \begin{bmatrix} +1 & 0 & 0 & 0 \\ 0 & +1 & 0 & 0 \\ 0 & 0 & +1 & 0 \\ 0 & 0 & 0 & +1 \end{bmatrix}$$

and $\det(\mathbf{F}^\top \mathbf{F}) = 1$.

Weighing four objects: Strategy 2

But consider another strategy:



Weighing four objects: Strategy 2

Neglecting measurement errors, we get

$$\begin{cases} +m_A + m_B = y_1 \\ -m_A + m_B = y_2 \\ +m_C + m_D = y_3 \\ -m_C + m_D = y_4 \end{cases} \Rightarrow \begin{cases} m_A = \frac{1}{2}(y_1 - y_2) \\ m_B = \frac{1}{2}(y_1 + y_2) \\ m_C = \frac{1}{2}(y_3 - y_4) \\ m_D = \frac{1}{2}(y_3 + y_4) \end{cases}$$

Weighing four objects: Strategy 2

What is the error for these estimates?

$$\begin{aligned}\text{var}(m_A) &= \frac{1}{4} [\text{var}(y_1) + \text{var}(y_2)] \\ &= \frac{1}{4} [\sigma^2 + \sigma^2] = \frac{1}{2}\sigma^2\end{aligned}$$

It was reduced from σ to $\sigma/\sqrt{2}$:

$$m_A = 10 \pm 0.07 \text{ grammes}$$

$$m_B = 25 \pm 0.07 \text{ grammes}$$

$$m_C = 15 \pm 0.07 \text{ grammes}$$

$$m_D = 30 \pm 0.07 \text{ grammes}$$

Weighing four objects: Strategy 2

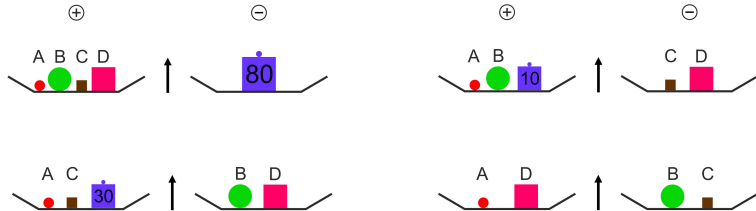
We have

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix} = \underbrace{\begin{bmatrix} +1 & +1 & 0 & 0 \\ -1 & +1 & 0 & 0 \\ 0 & 0 & +1 & +1 \\ 0 & 0 & -1 & +1 \end{bmatrix}}_{\mathbf{F}} \begin{bmatrix} m_A \\ m_B \\ m_C \\ m_D \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \epsilon_4 \end{bmatrix}$$

Additionally, observe that $\det(\mathbf{F}^\top \mathbf{F}) = 16$.

Weighing four objects: Strategy 3

But consider yet another strategy:



Weighing four objects: Strategy 3

Neglecting measurement errors, we get

$$+m_A + m_B + m_C + m_D = y_1$$

$$-m_A + m_B - m_C + m_D = y_2$$

$$-m_A - m_B + m_C + m_D = y_3$$

$$-m_A + m_B + m_C - m_D = y_4$$

Consequently,

$$m_A = \frac{1}{4} (y_1 - y_2 - y_3 - y_4)$$

$$m_B = \frac{1}{4} (y_1 + y_2 - y_3 + y_4)$$

$$m_C = \frac{1}{4} (y_1 - y_2 + y_3 + y_4)$$

$$m_D = \frac{1}{4} (y_1 + y_2 + y_3 - y_4)$$

Weighing four objects: Strategy 3

What is the error for these estimates?

$$\begin{aligned}\text{var}(m_A) &= \frac{1}{16} [\text{var}(y_1) + \text{var}(y_2) + \text{var}(y_3) + \text{var}(y_4)] \\ &= \frac{1}{16} [\sigma^2 + \sigma^2 + \sigma^2 + \sigma^2] = \frac{1}{4}\sigma^2\end{aligned}$$

It was further reduced from $\sigma/\sqrt{2}$ to $\sigma/2$:

$$m_A = 10 \pm 0.05 \text{ grammes}$$

$$m_B = 25 \pm 0.05 \text{ grammes}$$

$$m_C = 15 \pm 0.05 \text{ grammes}$$

$$m_D = 30 \pm 0.05 \text{ grammes}$$

It can be shown that this cannot be improved.

Weighing four objects: Strategy 3

We have

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix} = \underbrace{\begin{bmatrix} +1 & +1 & +1 & +1 \\ -1 & +1 & -1 & +1 \\ -1 & -1 & +1 & +1 \\ -1 & +1 & +1 & -1 \end{bmatrix}}_{\mathbf{F}} \begin{bmatrix} m_A \\ m_B \\ m_C \\ m_D \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \epsilon_4 \end{bmatrix}$$

Additionally, observe that $\det(\mathbf{F}^\top \mathbf{F}) = 256$.

Weighing four objects: Strategy 3

We have

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix} = \underbrace{\begin{bmatrix} +1 & +1 & +1 & +1 \\ -1 & +1 & -1 & +1 \\ -1 & -1 & +1 & +1 \\ -1 & +1 & +1 & -1 \end{bmatrix}}_{\mathbf{F}} \begin{bmatrix} m_A \\ m_B \\ m_C \\ m_D \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \epsilon_4 \end{bmatrix}$$

Additionally, observe that $\det(\mathbf{F}^T \mathbf{F}) = 256$.

Better precision coincides with higher values of $\det(\mathbf{F}^T \mathbf{F})$. The matrix $\mathbf{M} = \mathbf{F}^T \mathbf{F}$ is called the **Fisher information matrix** and experimental conditions maximizing its determinant are called **D-optimal**.

D-optimum designs for simple linear regression

Assume that the response is the form

$$y = a_0 + a_1x$$

We would like to make two measurements to estimate a_0 and a_1 , but we can select the corresponding settings of $x = x_1$ and $x = x_2$ before the experiment. If they are to belong to the interval $[-1, 1]$, how can we choose them best?

D-optimum designs for simple linear regression

Assume that the response is the form

$$y = a_0 + a_1 x$$

We would like to make two measurements to estimate a_0 and a_1 , but we can select the corresponding settings of $x = x_1$ and $x = x_2$ before the experiment. If they are to belong to the interval $[-1, 1]$, how can we choose them best?

Answer: We are going to get observations

$$y_1 \approx a_0 + a_1 x_1$$

$$y_2 \approx a_0 + a_1 x_2$$

D-optimum designs for simple linear regression

In matrix form, this is

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \underbrace{\begin{bmatrix} 1 & x_1 \\ 1 & x_2 \end{bmatrix}}_F \begin{bmatrix} a_0 \\ a_1 \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \end{bmatrix}$$

D-optimum designs for simple linear regression

In matrix form, this is

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \underbrace{\begin{bmatrix} 1 & x_1 \\ 1 & x_2 \end{bmatrix}}_{\mathbf{F}} \begin{bmatrix} a_0 \\ a_1 \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \end{bmatrix}$$

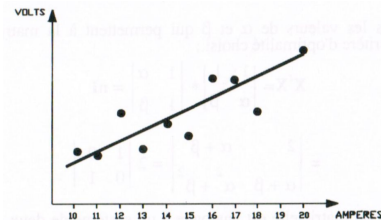
Adopting the D-optimality criterion as was for the weighing example, we get

$$\begin{aligned} \det(\mathbf{F}^T \mathbf{F}) &= \det \left(\begin{bmatrix} 1 & 1 \\ x_1 & x_2 \end{bmatrix} \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \end{bmatrix} \right) \\ &= \det \left(\begin{bmatrix} 2 & x_1 + x_2 \\ x_1 + x_2 & x_1^2 + x_2^2 \end{bmatrix} \right) = (x_1 - x_2)^2 \end{aligned}$$

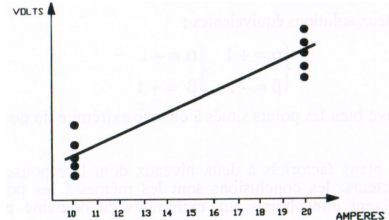
It is maximized for $x_1 = -1$ and $x_2 = +1$ (or $x_1 = +1$ and $x_2 = -1$), i.e., when the measurements are at the opposite ends of the design interval.

D-optimum design to measure resistance

We wish to estimate resistance R based on 10 measurements of the current (it may vary between 10 A and 20 A) and the corresponding voltage



arbitrary design



D-optimum design

Why is the information matrix so important?

Assume that the observations are modelled by

$$\mathbf{y} = \mathbf{F}\boldsymbol{\theta} + \boldsymbol{\varepsilon}$$

where $\boldsymbol{\varepsilon}$ is measurement noise satisfying

$$\mathbb{E}[\boldsymbol{\varepsilon}] = \mathbf{0}, \quad \text{cov}(\boldsymbol{\varepsilon}) = \mathbb{E}[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^T] = \sigma^2 \mathbf{I}$$

and σ^2 is a constant noise variance.

Why is the information matrix so important?

Assume that the observations are modelled by

$$\mathbf{y} = \mathbf{F}\boldsymbol{\theta} + \boldsymbol{\varepsilon}$$

where $\boldsymbol{\varepsilon}$ is measurement noise satisfying

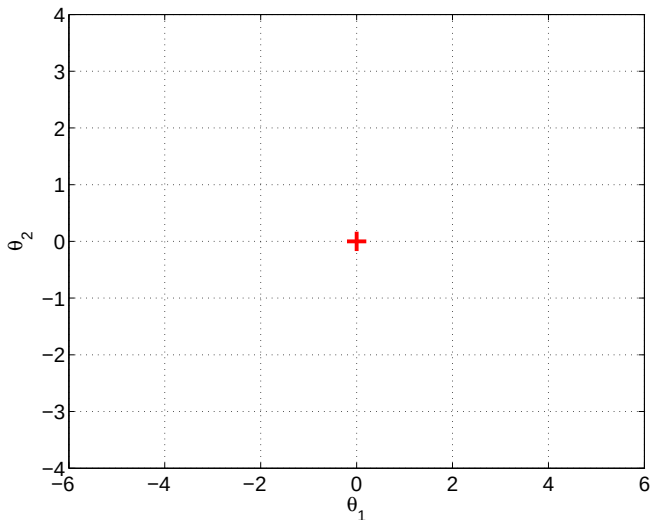
$$\mathbb{E}[\boldsymbol{\varepsilon}] = \mathbf{0}, \quad \text{cov}(\boldsymbol{\varepsilon}) = \mathbb{E}[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^T] = \sigma^2 \mathbf{I}$$

and σ^2 is a constant noise variance.

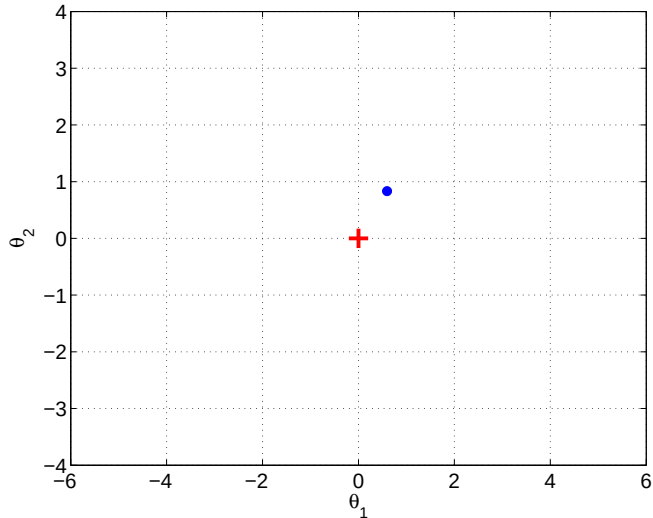
The parameter estimate $\hat{\boldsymbol{\theta}} = (\mathbf{F}^T \mathbf{F})^{-1} \mathbf{F}^T \mathbf{y}$ is then a random vector and

$$\begin{aligned} \mathbb{E}[\hat{\boldsymbol{\theta}}] &= \boldsymbol{\theta} && \text{(unbiasedness)} \\ \text{cov}(\hat{\boldsymbol{\theta}}) &= \sigma^2 (\mathbf{F}^T \mathbf{F})^{-1} && \text{(dispersion around } \boldsymbol{\theta} \text{)} \end{aligned}$$

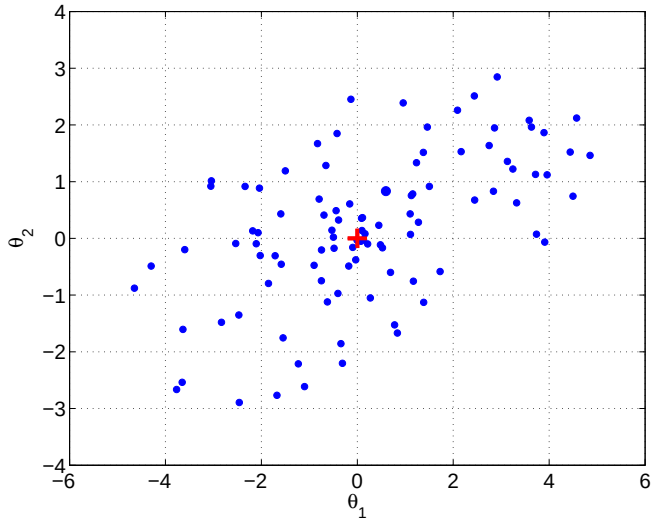
Uncertainty ellipsoid



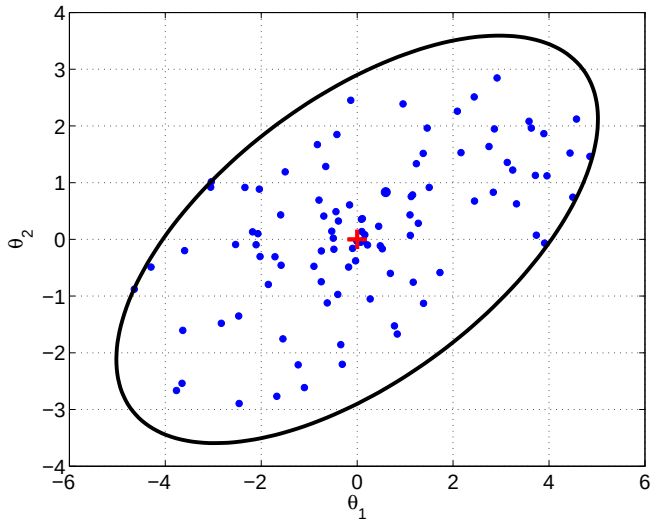
Uncertainty ellipsoid



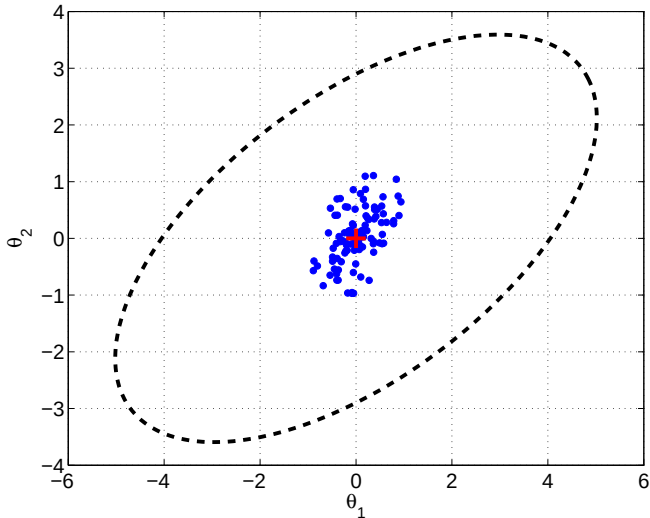
Uncertainty ellipsoid



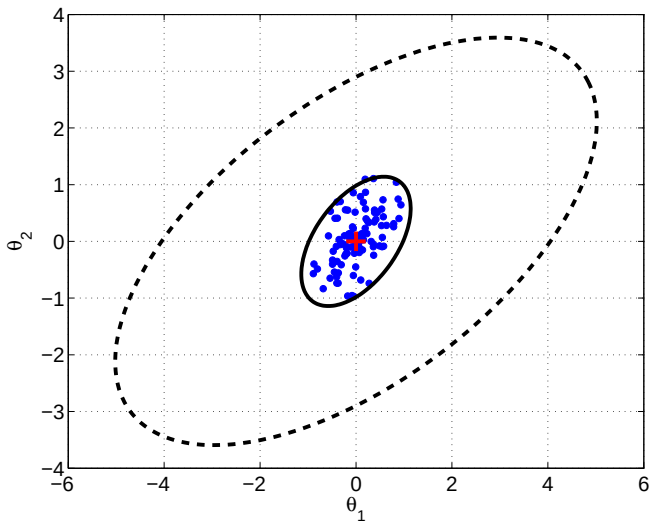
Uncertainty ellipsoid



Uncertainty ellipsoid



Uncertainty ellipsoid



Optimality criteria

Setting $\mathbf{M} = \mathbf{F}^\top \mathbf{F}$, we define

- **D-optimality** criterion:

$\Psi(\mathbf{M}) = \log \det(\mathbf{M}^{-1}) = -\log \det(\mathbf{M})$ (the volume of the uncertainty ellipsoid is minimized)

Optimality criteria

Setting $\mathbf{M} = \mathbf{F}^\top \mathbf{F}$, we define

- **D-optimality** criterion:

$\Psi(\mathbf{M}) = \log \det(\mathbf{M}^{-1}) = -\log \det(\mathbf{M})$ (the volume of the uncertainty ellipsoid is minimized)

- **A-optimality** criterion: $\Psi(\mathbf{M}) = \text{trace}(\mathbf{M}^{-1})$ (the mean axis length of the uncertainty ellipsoid is minimized)

Optimality criteria

Setting $\mathbf{M} = \mathbf{F}^T \mathbf{F}$, we define

- **D-optimality** criterion:

$\Psi(\mathbf{M}) = \log \det(\mathbf{M}^{-1}) = -\log \det(\mathbf{M})$ (the volume of the uncertainty ellipsoid is minimized)

- **A-optimality** criterion: $\Psi(\mathbf{M}) = \text{trace}(\mathbf{M}^{-1})$ (the mean axis length of the uncertainty ellipsoid is minimized)

- **E-optimality** criterion: $\Psi(\mathbf{M}) = \lambda_{\max}(\mathbf{M}^{-1})$ (the length of the largest axis of the uncertainty ellipsoid is minimized)

Notation: $\lambda_{\max}(\cdot)$ is the maximum eigenvalue of its matrix argument.

Most often, they yield different solutions!

Something about history

- 1940–1950 : development of factorial experiments and their applications in industry (mainly chemical engineering)
- 1950s: beginning of optimal experiment design (Kiefer-Wolfowitz theorem)
- 1969–1970 : first algorithms for numerical construction of optimum designs (V. Fedorov and H. Wynn)
- 1970 up to now : optimal input signals for parameter estimation in systems described by ODEs (R. Mehra)
- 1980s up to now – optimal sensor placement (1981) and input signal design (1983) for PDEs

All the above research directions are still far from being complete.

Workshops



Workshop on *Experiments for Processes With Time or Space Dynamics*,
Isaac Newton Institute for Mathematical Sciences, Cambridge, UK,
18–22 June 2011, organizers:
Dariusz Uciński and Andrew Curtis (University of Edinburgh)
www.newton.ac.uk/programmes/DAE/daew01.html

Workshops



Conference mODa 10



www.moda10.uz.zgora.pl

2^m factorial design

Consider the k -th run among multiple runs of an experiment

$$\begin{aligned} z_k = & \beta_0 + \beta_1 x_{k1} + \beta_2 x_{k2} + \beta_3 x_{k3} \\ & + \beta_4 x_{k1}x_{k2} + \beta_5 x_{k1}x_{k3} + \beta_6 x_{k2}x_{k3} \\ & + \beta_7 x_{k1}x_{k2}x_{k3} + \varepsilon_k \end{aligned}$$

where

- z_k – observed output
- $\mathbf{x}_k = [x_{k1}, x_{k2}, x_{k3}]^T$ – input vector consisting of $m = 3$ **factors**
- $\boldsymbol{\theta} = [\beta_0, \dots, \beta_7]^T$ – vector of unknown parameters
- ε_k – random error (zero mean, independent and identically distributed)

Here $x_{k1}x_{k2}$, $x_{k1}x_{k3}$, $x_{k2}x_{k3}$ and $x_{k1}x_{k2}x_{k3}$ are two- and three-way **interactions** among the input factors.

Factors

Factors may be quantitative, e.g.,

- pressure
- temperature
- time in a physical system
- money supply and interest rates in a macroeconomic system

or qualitative, e.g.,

- operation mode in a system
- the choice of open or closed-loop control
- the absence or presence of a human operator

Two-level factors and replications

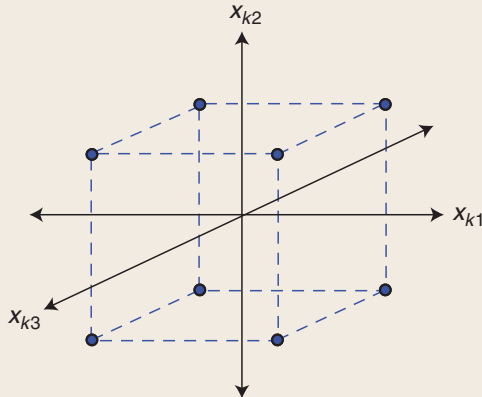
Let “ $-$ ” represent a low value of the given input and “ $+$ ” a high value. The unnormalized numerical values corresponding to a $+/-$ symbol different inputs may differ significantly according to the physical interpretations and units. It is often convenient, however, to normalize the numerical values to ± 1 .

Two-level factors and replications

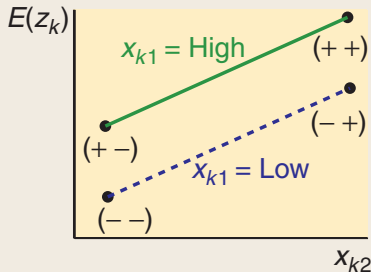
Let “ $-$ ” represent a low value of the given input and “ $+$ ” a high value. The unnormalized numerical values corresponding to a $+/-$ symbol different inputs may differ significantly according to the physical interpretations and units. It is often convenient, however, to normalize the numerical values to ± 1 .

Implementations of factorial design rely on mutually independent replications, where each replicate represents a set of 2^m runs. Multiple replicates decrease the variance of the outputs at different input levels, which increases the ability to detect differences in the influence of input factors.

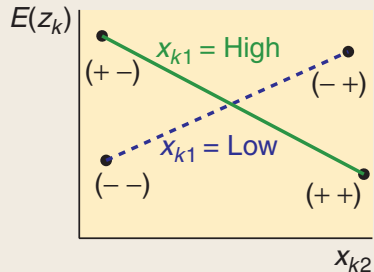
One replicate yields 2^m output values



Are interactions meaningful?



(a)



(b)

Orthogonality of the input matrix

For a single replicate we have

$$F = \begin{array}{c|cccc|cccc} & 1 & x_1 & x_2 & x_3 & x_1 x_2 & x_1 x_3 & x_2 x_3 & x_1 x_2 x_3 \\ \hline 1 & 1 & -1 & -1 & -1 & 1 & 1 & 1 & -1 \\ 2 & 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 \\ 3 & 1 & -1 & 1 & -1 & -1 & 1 & -1 & 1 \\ 4 & 1 & 1 & 1 & -1 & 1 & -1 & -1 & -1 \\ 5 & 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 \\ 6 & 1 & 1 & -1 & 1 & -1 & 1 & -1 & -1 \\ 7 & 1 & -1 & 1 & 1 & -1 & -1 & 1 & -1 \\ 8 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{array}$$

and $F^T F = 8I_8$ (eight times the identity matrix). In general, $F^T F = N2^m I_{2^m}$, where N is the number of replicates.

Main effects and interaction effects

Then from the normal equations we get

$$\hat{\theta} = (\mathbf{F}^T \mathbf{F})^{-1} \mathbf{F}^T \mathbf{z} = \frac{1}{8} \mathbf{F}^T \mathbf{z}$$

Specifically,

$$\hat{\beta}_1 = \frac{(z_2 + z_4 + z_6 + z_8) - (z_1 + z_3 + z_5 + z_7)}{8}$$

$$\hat{\beta}_2 = \frac{(z_3 + z_4 + z_7 + z_8) - (z_1 + z_2 + z_5 + z_6)}{8}$$

$$\hat{\beta}_3 = \frac{(z_5 + z_6 + z_7 + z_8) - (z_1 + z_2 + z_3 + z_4)}{8}$$

Main effects and interaction effects

Then from the normal equations we get

$$\hat{\theta} = (\mathbf{F}^T \mathbf{F})^{-1} \mathbf{F}^T \mathbf{z} = \frac{1}{8} \mathbf{F}^T \mathbf{z}$$

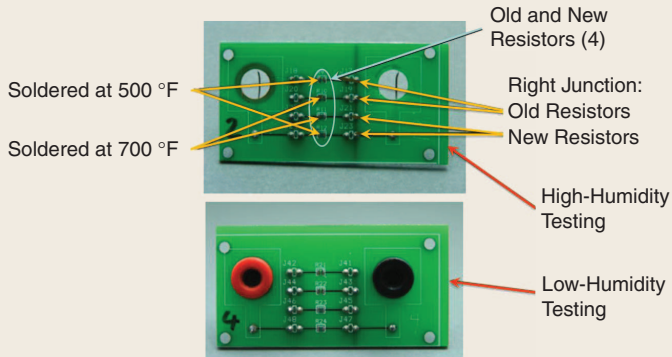
Specifically,

$$\begin{aligned}\hat{\beta}_1 &= \frac{(z_2 + z_4 + z_6 + z_8) - (z_1 + z_3 + z_5 + z_7)}{8} \\ \hat{\beta}_2 &= \frac{(z_3 + z_4 + z_7 + z_8) - (z_1 + z_2 + z_5 + z_6)}{8} \\ \hat{\beta}_3 &= \frac{(z_5 + z_6 + z_7 + z_8) - (z_1 + z_2 + z_3 + z_4)}{8}\end{aligned}$$

Factorial designs in this case are also D-optimal!

Factorial design for a circuit board

Shortly before the launch of the MESSENGER spacecraft from Cape Canaveral, Florida on August 3, 2004, an unexpected increase in the resistance of the resistors used in circuit boards on the spacecraft was observed. This threatened to delay the launch unless the cause was found.



Factorial design for a circuit board (ct'd)

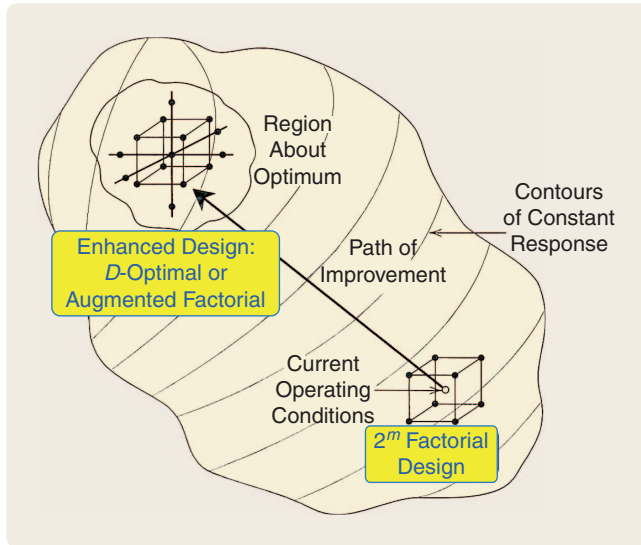
Solder Temp. (s)	Resistor Batch (r)	Humidity Level (h)	Initial Resist. (Ohms)	Final Resist. (Ohms)	z
—	—	—	90.87	90.87	0
+	—	—	90.89	90.88	−0.01
—	+	—	243.05	243.04	0
+	+	—	242.98	242.96	−0.01
—	—	+	90.68	200.00	120.6
+	—	+	90.71	198.00	118.3
—	+	+	241.64	244.43	1.155
+	+	+	241.63	248.90	3.009

For details, see James Spall, Factorial design for efficient experimentation, *IEEE Contr. Syst. Mag.*, October 2010, pp.38–53.

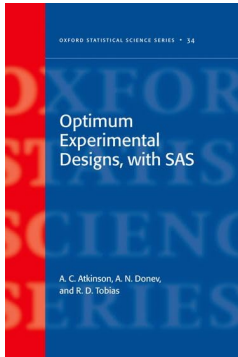
Factorial design for a circuit board (ct'd)

Parameter	Input Factor	Parameter	
		Estimate	<i>p</i> -value
β_0	Constant	30.38	0.0109
β_1	<i>s</i>	−0.058	0.929
β_2	<i>r</i>	−29.34	0.0113
β_3	<i>h</i>	30.39	0.0109
β_4	<i>sr</i>	0.52	0.5
β_5	<i>sh</i>	−0.053	0.935
β_6	<i>rh</i>	−29.34	0.0113

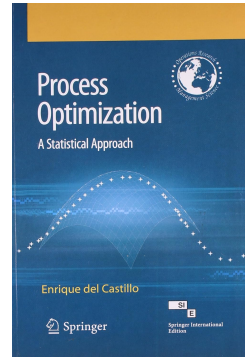
Response surface methodology



For more information



DESIGN AND ANALYSIS OF
EXPERIMENTS
TENTH EDITION
DOUGLAS C. MONTGOMERY



Connotations with machine learning?

Common hyperparameter tuning techniques:

- 1 Exhaustive search
- 2 Bayesian optimization
- 3 Heuristic search
- 4 Multi-fidelity optimization

Connotations with machine learning?

Common hyperparameter tuning techniques:

- 1 Exhaustive search
- 2 Bayesian optimization
- 3 Heuristic search
- 4 Multi-fidelity optimization

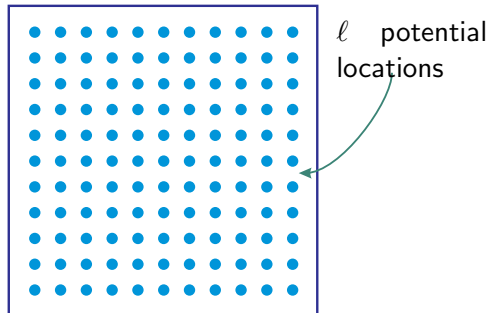
They can be augmented with RSM, see, e.g.,

- Lujan-Moreno, Gustavo A., Phillip R. Howard, Omar G. Rojas, and Douglas C. Montgomery. "Design of experiments and response surface methodology to tune machine learning hyperparameters, with a random forest case-study." *Expert Systems with Applications* 109 (2018): 195–205.
- Bozkurt Keser, Sinem, and Yeliz Buruk Sahin. "Response surface methodology to tune artificial neural network hyper-parameters." *Expert Systems* 38, no. 8 (2021): e12792.

And now the problem of sensor selection

Given m sensors and a finite set $\{\chi^1, \dots, \chi^\ell\} \subset \Omega \cup \partial\Omega$ of points at which they can be placed, consider their fixed configuration.

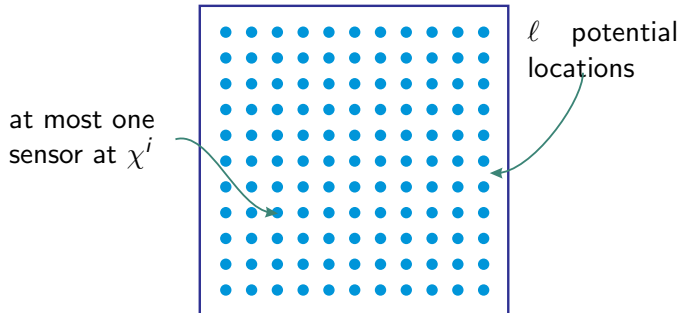
Spatial domain $\Omega \cup \partial\Omega$



And now the problem of sensor selection

Given m sensors and a finite set $\{\chi^1, \dots, \chi^\ell\} \subset \Omega \cup \partial\Omega$ of points at which they can be placed, consider their fixed configuration.

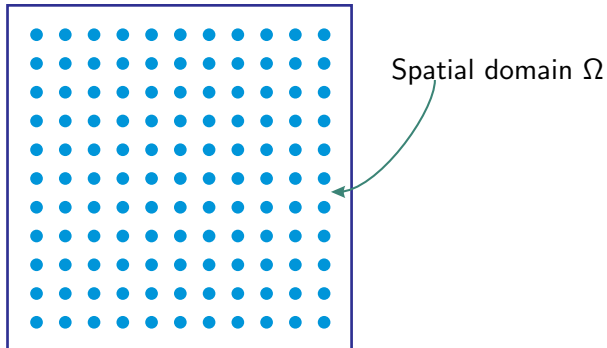
Spatial domain $\Omega \cup \partial\Omega$



Observations

Equivalently, we can treat $\{\chi^1, \dots, \chi^\ell\} \subset \Omega \cup \partial\Omega$ as gauged sites at which ℓ sensors reside and only m from among them are to be activated.

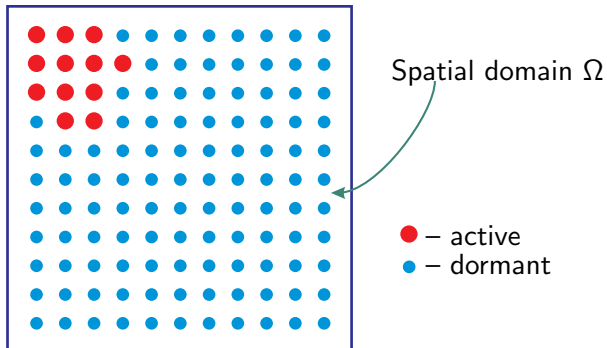
Combinatorial problem: Which m -element subset to select?



Observations

Equivalently, we can treat $\{\chi^1, \dots, \chi^\ell\} \subset \Omega \cup \partial\Omega$ as gauged sites at which ℓ sensors reside and only m from among them are to be activated.

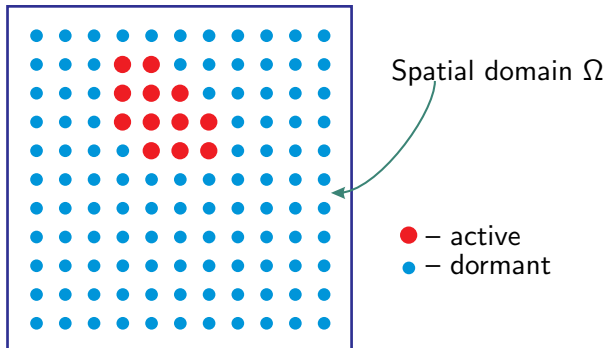
Combinatorial problem: Which m -element subset to select?



Observations

Equivalently, we can treat $\{\chi^1, \dots, \chi^\ell\} \subset \Omega \cup \partial\Omega$ as gauged sites at which ℓ sensors reside and only m from among them are to be activated.

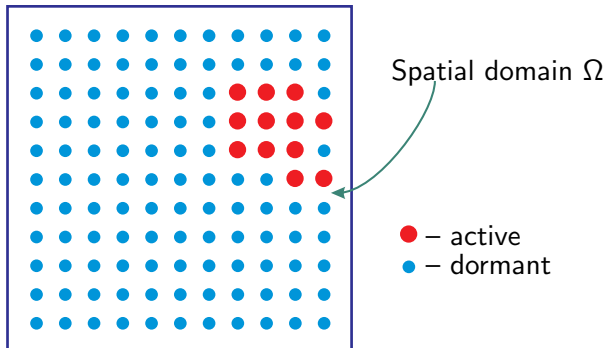
Combinatorial problem: Which m -element subset to select?



Observations

Equivalently, we can treat $\{\chi^1, \dots, \chi^\ell\} \subset \Omega \cup \partial\Omega$ as gauged sites at which ℓ sensors reside and only m from among them are to be activated.

Combinatorial problem: Which m -element subset to select?

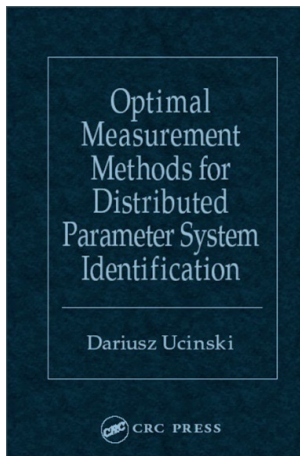


Connotations with machine learning?

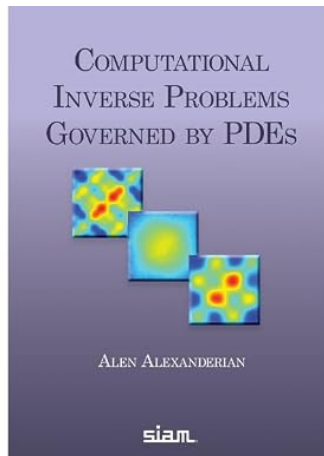
A critical step in big data analysis is data reduction. An approach is to work with a carefully selected small representative sample (called **subdata**) of size n from the big data of size N (called **full data**). The sample size n should be chosen so that appropriate statistical tools and methods can be applied to the subdata with sufficiently reduced computational complexity. Methods to identify such subdata are called **subdata selection methods**.

See, e.g., Singh, Rakhi, and John Stufken. "Subdata selection with a large number of variables." *The New England Journal of Statistics in Data Science* 1, no. 3 (2023).

Complements to this minicourse

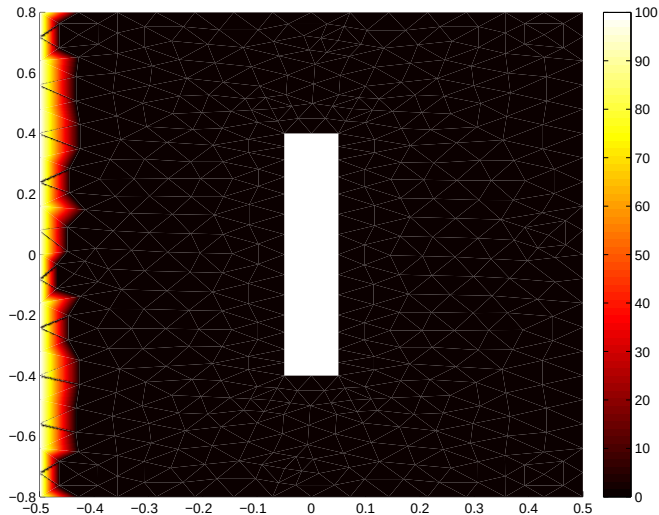


CRC Press, 2005

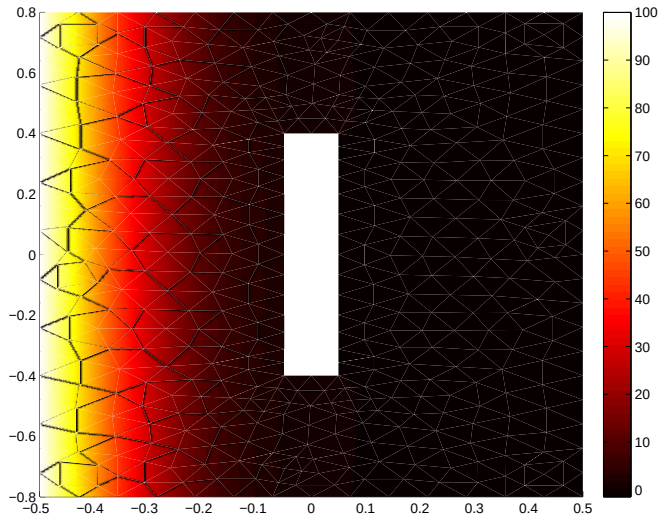


SIAM, 2025

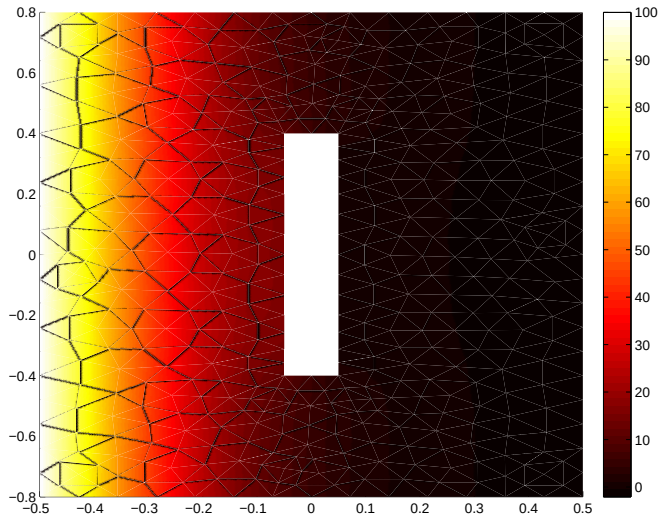
Example: Heating a metal block with a crack



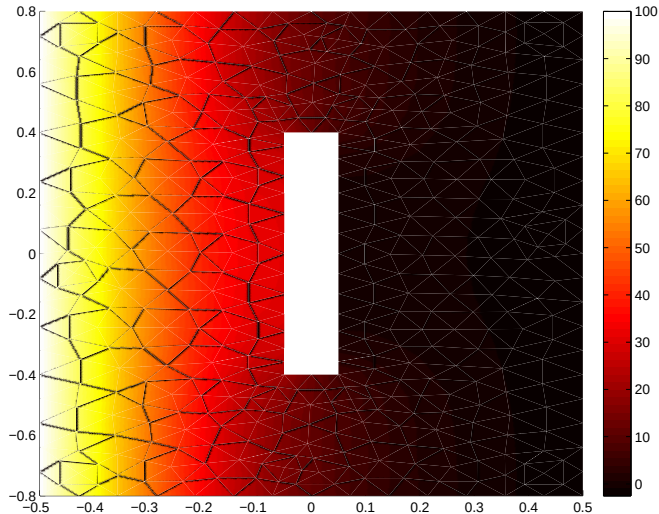
Example: Heating a metal block with a crack



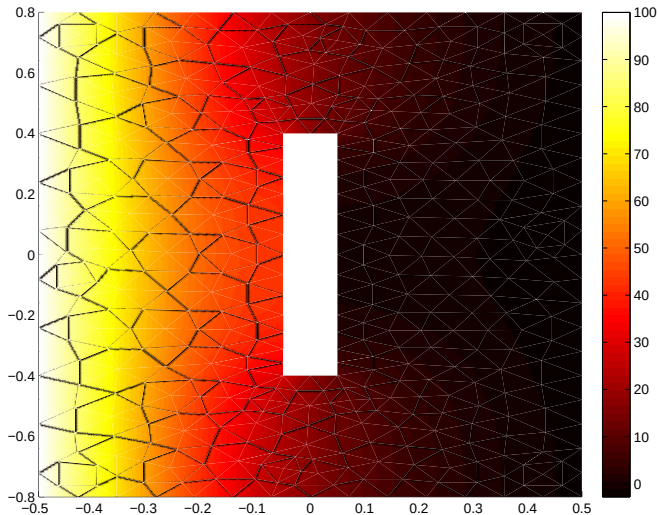
Example: Heating a metal block with a crack



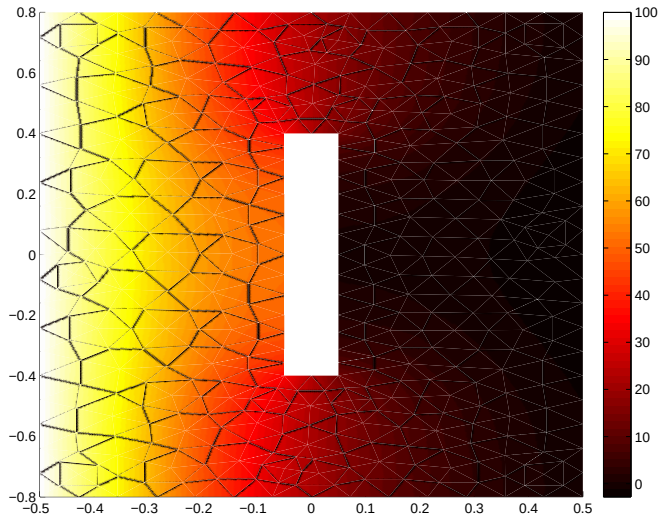
Example: Heating a metal block with a crack



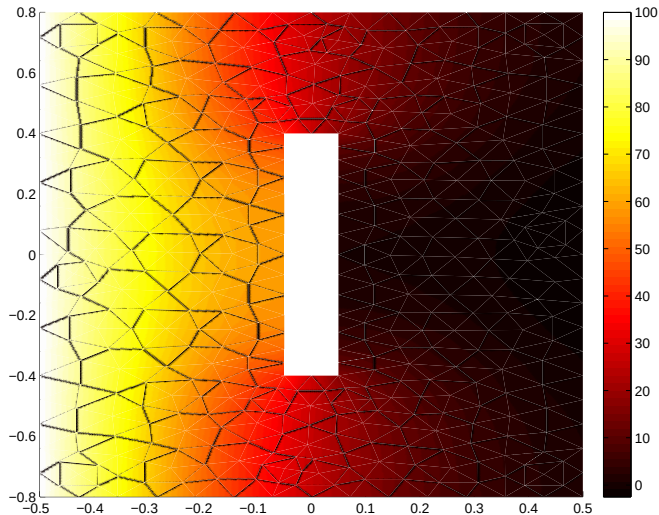
Example: Heating a metal block with a crack



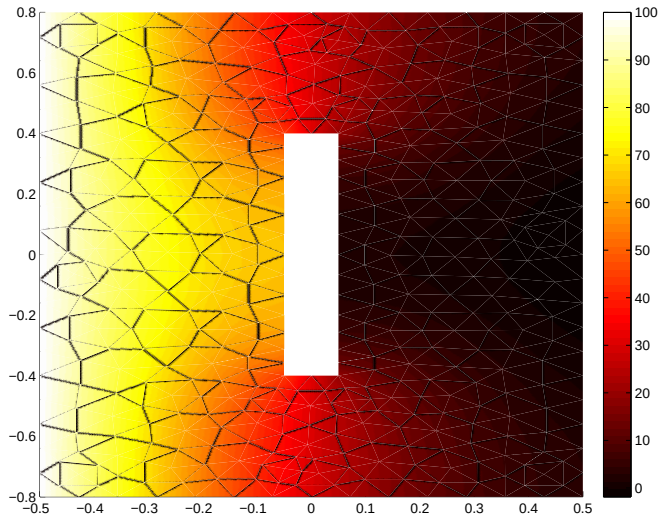
Example: Heating a metal block with a crack



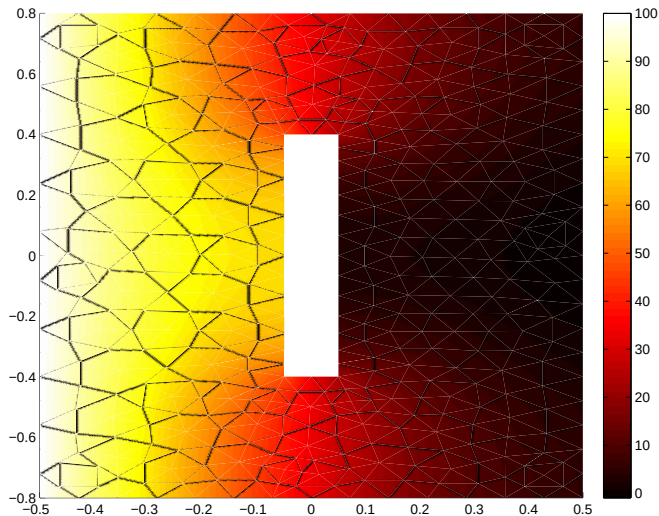
Example: Heating a metal block with a crack



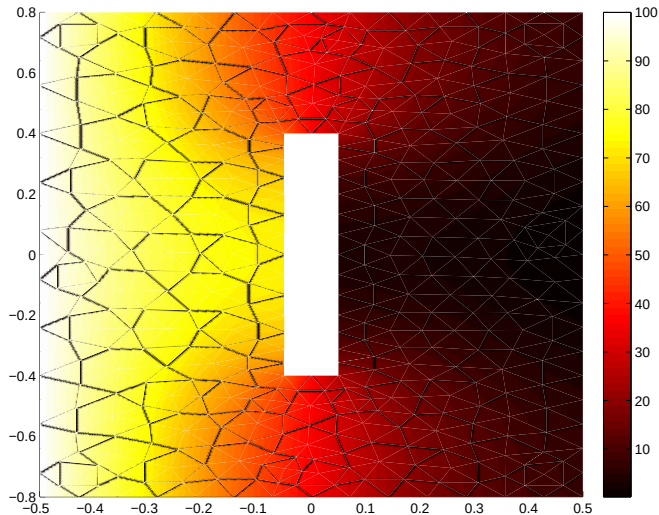
Example: Heating a metal block with a crack



Example: Heating a metal block with a crack



Example: Heating a metal block with a crack



Example ctd': Mathematical model

Let $y(\mathbf{x}, t)$ be the temperature at spatial point $\mathbf{x}$ and time t . Its evolution is governed by the **partial differential equation (PDE)**

$$\frac{\partial y}{\partial t} = \theta \Delta y \quad \text{for } \mathbf{x} \in \Omega \text{ and } t \in [0, t_f]$$

θ being a parameter, subject to the **initial condition**

$$y = 0 \quad \text{in } \Omega \text{ at } t = 0$$

and the **boundary conditions**

$$y = 100 \quad \text{on the left side of } \Omega \quad (\text{Dirichlet condition})$$

$$\frac{\partial y}{\partial \mathbf{n}} = -10 \quad \text{on the right side of } \Omega \quad (\text{Neumann condition})$$

$$\frac{\partial y}{\partial \mathbf{n}} = 0 \quad \text{on all other boundaries} \quad (\text{Neumann condition})$$

Example ctd': Calibration and measurement design

If thermal conductivity θ is unknown, which is rather typical, we could observe the temperature evolution at a fixed point and tune θ so that our model fits the data as best as it can.

Example ctd': Calibration and measurement design

If thermal conductivity θ is unknown, which is rather typical, we could observe the temperature evolution at a fixed point and tune θ so that our model fits the data as best as it can.

Problem

Where to place the pyrometer so as to get the most valuable information about θ ?

Such situations (PDEs and related sensor location problems) are common in engineering practice.

Other motivating examples

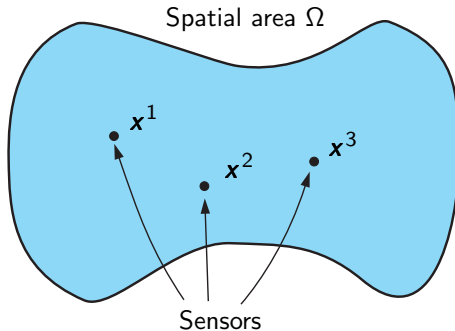
- environment observation and forecasting systems
 - air pollution monitoring,
 - groundwater resources management and river monitoring,
 - military applications: surveillance and inspection in hazardous environments,
- fault detection and isolation,
- emerging smart materials,
- intelligent building monitoring,
- habitat monitoring,
- intelligent traffic systems,
- computer-assisted tomography,
- recovery of valuable minerals and hydrocarbon from underground permeable reservoirs,
- and many more . . .

Spatiotemporal dynamics

Distributed parameter system—dynamic system whose state depends on both time and space; its model (a PDE) is known up to a vector of unknown parameters θ .

Observations—using sensors in order to estimate θ .

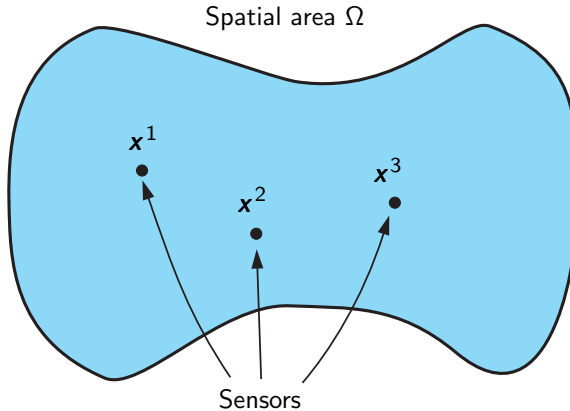
Subject of this minicourse



Problem: How to choose optimal sensor locations?

Strategies of taking measurements (Option 1)

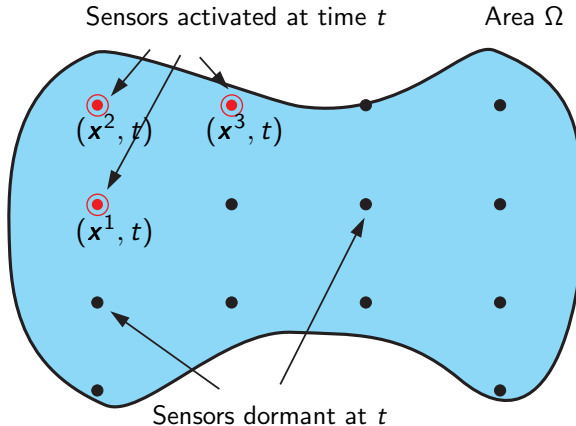
Stationary sensors



How to determine the best sensor configuration?

Strategies of taking measurements (Option 2)

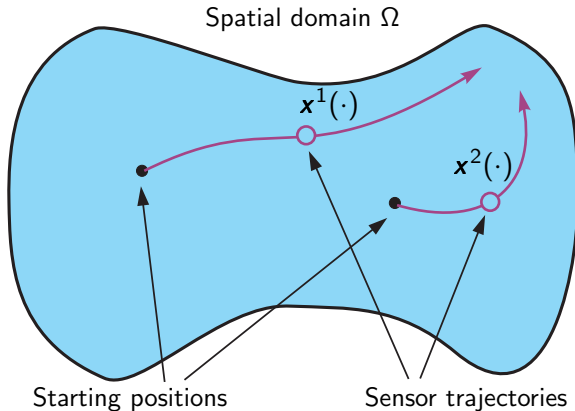
Scanning



How to select optimal locations at given time instants?

Strategies of taking measurements (Option 3)

Mobile sensors



How to design optimal trajectories?

Existing approaches to optimal sensor location

- Conversion to a non-linear problem of **state estimation** (Malebranche, 1988; Korbicz *et al.*, 1988);
- Employing **random fields analysis** (Kazimierczyk, 1989; Sun, 1994);
- Formulation in the spirit of **optimum experimental design** (Uspenskii and Fedorov, 1975; Quereschi *et al.*, 1980; Rafajłowicz, 1978–1995; Kammer, 1990; 1992; Sun, 1994; Point *et al.*, 1996; Vande Wouwer *et al.*, 1999; Uciński, 1991–present; Patan, 2001–present).
- Large-scale optimum **design for Bayesian inverse problems** (A. Alexanderian, N. Petra, G. Stadler, O. Ghattas, A. Attia, since 2014, I. Gedjadze and V. Shutyaev, since 2008).

Existing approaches to optimal sensor location

- Conversion to a non-linear problem of **state estimation** (Malebranche, 1988; Korbicz *et al.*, 1988);
- Employing **random fields analysis** (Kazimierczyk, 1989; Sun, 1994);
- Formulation in the spirit of optimum experimental design (Uspenskii and Fedorov, 1975; Quereschi *et al.*, 1980; Rafajłowicz, 1978–1995; Kammer, 1990; 1992; Sun, 1994; Point *et al.*, 1996; Vande Wouwer *et al.*, 1999; Uciński, 1991–present; Patan, 2001–present).
- Large-scale optimum **design for Bayesian inverse problems** (A. Alexanderian, N. Petra, G. Stadler, O. Ghattas, A. Attia, since 2014, I. Gedjadze and V. Shutyaev, since 2008).

System description

Consider a spatial domain $\Omega \subset \mathbb{R}^2$ and the PDE

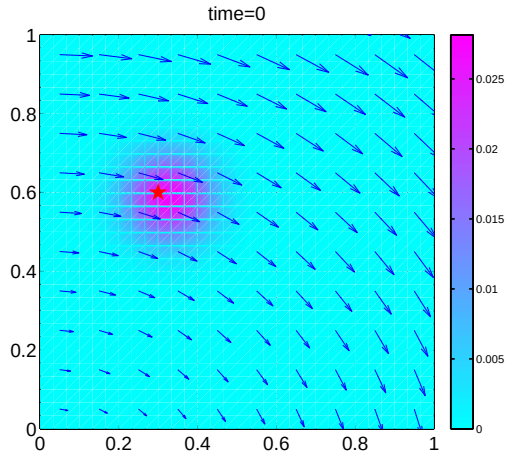
$$\frac{\partial y}{\partial t} = \mathcal{F} \left(\mathbf{x}, t, y, \boldsymbol{\theta} \right), \quad \mathbf{x} \in \Omega, \quad t \in T$$

subject to the appropriate initial and boundary conditions.

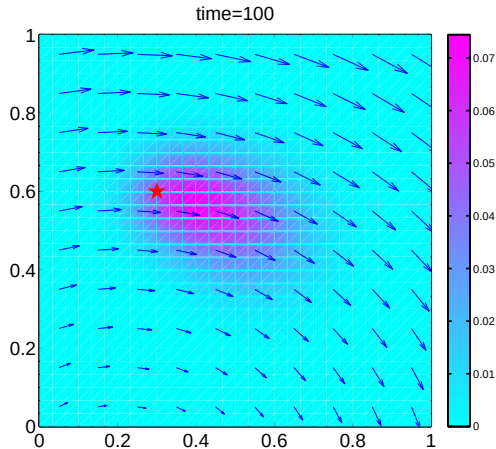
Notation:

- ▶ $\mathbf{x}$ – spatial variable, t – time, $T = [0, t_f]$;
- ▶ $y = y(\mathbf{x}, t)$ – state vector;
- ▶ $\mathcal{F}$ – given operator including spatial derivatives;
- ▶ $\boldsymbol{\theta} \in \mathbb{R}^m$ – vector of *unknown* (constant) parameters.

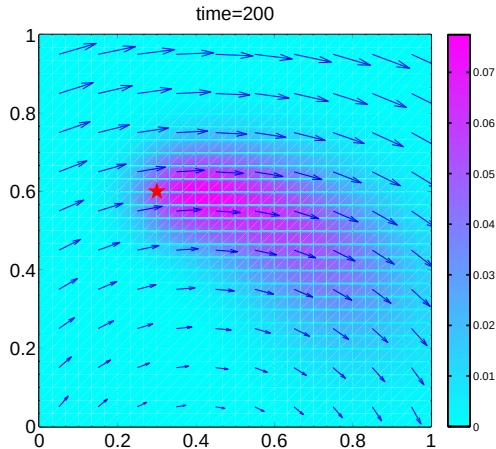
Example: Air pollution



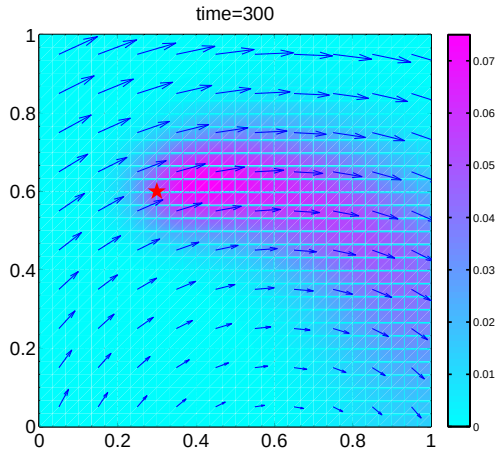
Example: Air pollution



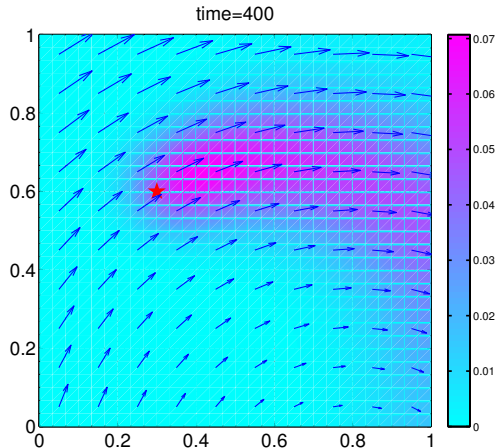
Example: Air pollution



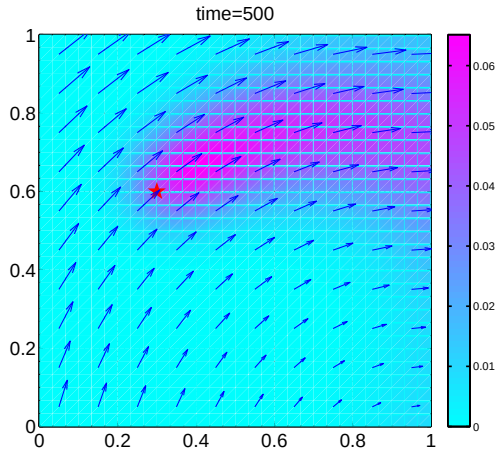
Example: Air pollution



Example: Air pollution



Example: Air pollution



Example: Air pollution

Consider a domain $\Omega \subset \mathbb{R}^2$ with sufficiently smooth boundary Γ .

Advection-diffusion equation

$$\frac{\partial}{\partial t} y(\mathbf{x}, t) = \nabla \cdot [\kappa(\mathbf{x}, \theta) \nabla y(\mathbf{x}, t)] + \nu(\mathbf{x}) \cdot \nabla y(\mathbf{x}, t) + f(\mathbf{x}, t, \theta)$$
$$y(\mathbf{x}, 0) = y_0(\mathbf{x}), \quad y(\mathbf{x}, t) \Big|_{\Gamma} = 0$$

Notation:

- $y(\mathbf{x}, t)$ – contaminant concentrat. at point $\mathbf{x}$ and time t
- $\kappa(\mathbf{x}, \theta)$ – diffusivity coefficient, parameterized by θ
- $f(\mathbf{x}, t, \theta)$ – source term
- $\nu(\mathbf{x})$ – wind velocity field

Observation

For pointwise stationary sensors we have

$$\mathbf{z}(t) = \mathbf{y}_m(t) + \boldsymbol{\varepsilon}_m(t), \quad t \in T$$

where

$$\mathbf{y}_m(t) = \text{col}[y(\mathbf{x}^1, t), \dots, y(\mathbf{x}^N, t)]$$

$$\boldsymbol{\varepsilon}_m(t) = \text{col}[\varepsilon(\mathbf{x}^1, t), \dots, \varepsilon(\mathbf{x}^N, t)]$$

$\mathbf{x}^j \in \bar{\Omega}$, $j = 1, \dots, N$ – sensor locations, $\varepsilon(\mathbf{x}, t)$ – Gaussian measurement noise such that

$$\mathbb{E}\{\varepsilon(\mathbf{x}, t)\} = 0, \quad \mathbb{E}\{\varepsilon(\mathbf{x}, t)\varepsilon(\mathbf{x}', t')\} = q(\mathbf{x}, \mathbf{x}', t)\delta(t - t')$$

δ – Diraca delta function.

LS criterion

The least-squares estimates of θ are the ones which minimize

$$\mathcal{J}(\theta) = \frac{1}{2} \int_T \|z(t) - \hat{y}_m(t; \theta)\|_{Q^{-1}(t)}^2 dt$$

where $Q(t) = [q(x^i, x^j, t)]_{i,j=1}^N \in \mathbb{R}^{N \times N}$ is positive-definite,

$$\|a\|_{Q^{-1}(t)}^2 = a^T Q^{-1}(t) a, \quad \forall a \in \mathbb{R}^N$$

$$\hat{y}_m(t; \theta) = \text{col}[\hat{y}(x^1, t; \theta), \dots, \hat{y}(x^N, t; \theta)]$$

and $\hat{y}(\cdot, \cdot; \theta)$ stands for the solution to the state equation corresponding to a given value of θ .

Optimal measurement problem

Formulation

Determine $\mathbf{x}^j$, $j = 1, \dots, N$ so as to maximize the identification accuracy in a sense.

Problem

How to quantify the identification accuracy?

Optimal measurement problem

Formulation

Determine $\mathbf{x}^j$, $j = 1, \dots, N$ so as to maximize the identification accuracy in a sense.

Problem

How to quantify the identification accuracy?

Answer: Make use of the Cramér-Rao inequality :

$$\text{cov}(\hat{\boldsymbol{\theta}}) = \text{E}\left\{(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^T\right\} \succeq \mathbf{M}^{-1}$$

We have $\text{cov}(\hat{\boldsymbol{\theta}}) = \mathbf{M}^{-1}$ provided that an estimator is efficient!
But what is $\mathbf{M}$?

Towards a criterion

Spatially independent measurements

$$E\{\varepsilon(\mathbf{x}^i, t)\varepsilon(\mathbf{x}^j, t')\} = \sigma^2 \delta_{ij} \delta(t - t')$$

where δ_{ij} is the Kronecker delta and $\sigma > 0$ signifies the standard deviation of the measurement noise.

Average Fisher Information Matrix (FIM)

$$\bar{\mathbf{M}} = \frac{1}{Nt_f} \sum_{j=1}^N \int_0^{t_f} \mathbf{g}(\mathbf{x}^j, t) \mathbf{g}^T(\mathbf{x}^j, t) dt$$

where $\mathbf{g}(\mathbf{x}^j, t) = \left(\frac{\partial y(\mathbf{x}^j, t; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right)_{\boldsymbol{\theta}=\boldsymbol{\theta}^0}^T$ is the vector of *sensitivity coefficients*, $\boldsymbol{\theta}^0$ stands for a preliminary estimate of $\boldsymbol{\theta}$.

$\bar{\mathbf{M}} = \mathbf{M}$ up to a constant multiplier.

Calculation of the sensitivity vector g

Options

- ① Finite-difference method
- ② Sensitivity-equation method (direct approach)
- ③ Variational method (indirect approach)

Calculation of the sensitivity vector g

Example (direct approach)

Consider

$$\frac{\partial y}{\partial t} = \nabla \cdot (\kappa \nabla y) \quad \text{in } \Omega \times T$$

where $\Omega \subset \mathbb{R}^2$ is the spatial domain with boundary Γ , $T = (0, t_f)$,

$$\kappa(\mathbf{x}) = a + b\psi(\mathbf{x}), \quad \psi(\mathbf{x}) = x_1 + x_2$$

subject to

$$y(\mathbf{x}, 0) = 5 \quad \text{in } \Omega$$

$$y(\mathbf{x}, t) = 5(1 - t) \quad \text{on } \Gamma \times T$$

Calculation of the sensitivity vector g

Example (ctd')

The sensitivities to a and b satisfy

$$\begin{aligned}\frac{\partial y_a}{\partial t} &= \nabla \cdot \nabla y + \nabla \cdot (\kappa \nabla y_a) \\ \frac{\partial y_b}{\partial t} &= \nabla \cdot (\psi \nabla y) + \nabla \cdot (\kappa \nabla y_b)\end{aligned}$$

supplemented with homogeneous initial and Dirichlet boundary conditions. We have to solve them simultaneously with the state equation and to store the solution in memory for its further use in a sensor location algorithm.

Ultimate formulation

Reformulated sensor location problem

Choose $\mathbf{x}^j, j = 1, \dots, N$ corresponding to an extreme of some scalar measure Ψ of the FIM, e.g. minimize

$$\Psi_{\gamma}(\mathbf{M}) = \begin{cases} \left[\frac{1}{m} \text{trace}(\mathbf{Q}\mathbf{M}^{-1}\mathbf{Q}^T)^{\gamma} \right]^{1/\gamma} & \text{if } \det(\mathbf{M}) \neq 0 \\ \infty & \text{if } \det(\mathbf{M}) = 0 \end{cases}$$

Most popular choices:

- $\gamma = 1$ and $\mathbf{Q} = \mathbf{I}_m \Rightarrow$ A-optimal design

$$J_A(\mathbf{x}^1, \dots, \mathbf{x}^N) = \text{trace } \mathbf{M}^{-1}(\mathbf{x}^1, \dots, \mathbf{x}^N) \longrightarrow \min$$

- $\gamma \rightarrow \infty$ and $\mathbf{Q} = \mathbf{I}_m \Rightarrow$ E-optimal design

$$J_E(\mathbf{x}^1, \dots, \mathbf{x}^N) = \lambda_{\max}(\mathbf{M}^{-1}(\mathbf{x}^1, \dots, \mathbf{x}^N)) \longrightarrow \min$$

- $\gamma \rightarrow 0$ and $\mathbf{Q} = \mathbf{I}_m \Rightarrow$ D-optimal design

$$J_D(\mathbf{x}^1, \dots, \mathbf{x}^N) = \det \mathbf{M}(\mathbf{x}^1, \dots, \mathbf{x}^N) \longrightarrow \max$$

- Sensitivity criterion (may yield very poor solutions!)

$$J_S(\mathbf{x}^1, \dots, \mathbf{x}^N) = \text{trace } \mathbf{M}(\mathbf{x}^1, \dots, \mathbf{x}^N) \longrightarrow \max$$

Well, there are some drawbacks. . .

- The problem dimensionality may be high and many local minima interfere with the optimization process.
- When $t_f < \infty$, the estimator is neither unbiased nor characterized by a normal distribution.
- Sensors sometimes tend to cluster.
- The optimal solution usually depends on the parameters to be identified.

Example

Consider the heat equation

$$\frac{\partial y(\mathbf{x}, t)}{\partial t} = \frac{\partial}{\partial x_1} \left(\kappa(\mathbf{x}) \frac{\partial y(\mathbf{x}, t)}{\partial x_1} \right) + \frac{\partial}{\partial x_2} \left(\kappa(\mathbf{x}) \frac{\partial y(\mathbf{x}, t)}{\partial x_2} \right) \quad \text{in } \Omega \times T$$

supplemented with

$$\begin{aligned} y(\mathbf{x}, 0) &= 5 && \text{in } \Omega \\ y(\mathbf{x}, t) &= 5(1 - t) && \text{on } \partial\Omega \times T \end{aligned}$$

where $\Omega = [0, 1]^2$, $T = (0, 1)$,

$$\begin{aligned} \kappa(\mathbf{x}) &= \theta_1 + \theta_2 x_1 + \theta_3 x_2 \\ \theta_1 &= 0.1, \quad \theta_2 = \theta_3 = 0.3 \end{aligned}$$

Results

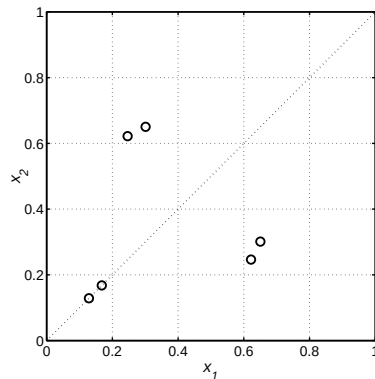
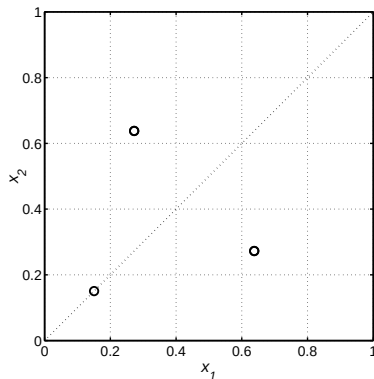


Figure: Locating six sensors through the direct use of non-linear programming techniques: (a) independent measurements, (b) correlated measurements.

Sensor clusterization

Example

Consider

$$\frac{\partial y(x, t)}{\partial t} = \theta_1 \frac{\partial^2 y(x, t)}{\partial x^2}, \quad x \in (0, \pi), \quad t \in (0, 1)$$

supplemented by

$$\begin{cases} y(0, t) = y(\pi, t) = 0, & t \in (0, 1) \\ y(x, 0) = \theta_2 \sin(x), & x \in (0, \pi) \end{cases}$$

Find D-optimal sensor locations for estimation of θ_1 and θ_2 .

Sensor clusterization

Example (ctd')

Solution The closed-form solution is

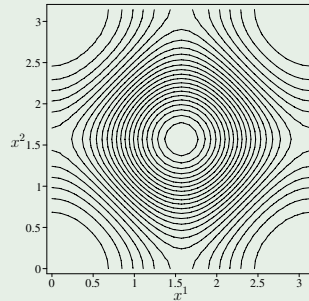
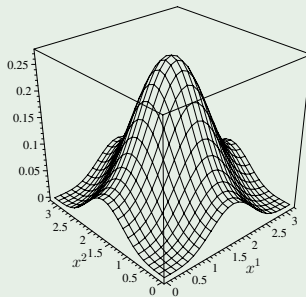
$$y(x, t) = \theta_2 \exp(-\theta_1 t) \sin(x)$$

Assuming $\sigma = 1$, on some easy calculations, we get

$$\begin{aligned} \det(M(x^1, x^2)) &= \underbrace{\frac{\theta_2^2}{16\theta_1^4} (-4\theta_1^2 \exp(-2\theta_1) - 2 \exp(-2\theta_1) + \exp(-4\theta_1) + 1)}_{\text{constant term}} \\ &\cdot \underbrace{(2 - \cos^2(x^1) - \cos^2(x^2))}_{\text{term dependent on } x^1 \text{ and } x^2}^2 \implies x^{1*} = x^{2*} = \frac{\pi}{2} \end{aligned}$$

Sensor clusterization

Example (ctd')



Dependence of designs on estimated parameters

Example

Consider the heat process

$$\frac{\partial y(x, t)}{\partial t} = \theta \frac{\partial^2 y(x, t)}{\partial x^2}, \quad x \in (0, \pi), \quad t \in (0, t_f)$$

complemented with

$$\begin{cases} y(0, t) = y(\pi, t) = 0, & t \in (0, t_f) \\ y(x, 0) = \sin(x) + \frac{1}{2} \sin(2x), & x \in (0, \pi) \end{cases}$$

Its solution can be easily found in explicit form as

$$y(x, t) = \exp(-\theta t) \sin(x) + \frac{1}{2} \exp(-4\theta t) \sin(2x)$$

Dependence of designs on estimated parameters

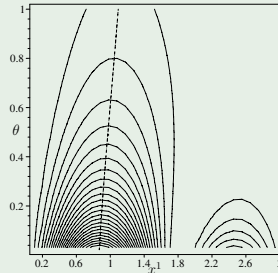
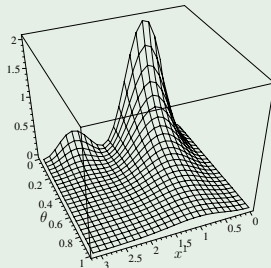
Example

The FIM (here scalar) is

$$\begin{aligned} M(x^1) &= \int_0^{t_f} \left(\frac{\partial y(x^1, t; \theta)}{\partial \theta} \right)^2 dt \\ &= \frac{1}{4\theta^3} \left\{ -\sin^2(x^1) (2\theta^2 t_f^2 + 2t_f\theta + 1) \exp(-2t_f\theta) \right. \\ &\quad - \frac{32}{125} \cos(x^1) \sin^2(x^1) (10t_f\theta + 2 + 25\theta^2 t_f^2) \exp(-5t_f\theta) \\ &\quad - \frac{1}{4} \cos^2(x^1) \sin^2(x^1) (1 + 32\theta^2 t_f^2 + 8t_f\theta) \exp(-8t_f\theta) \\ &\quad \left. + \frac{1}{500} \sin^2(x^1) (500 + 256 \cos(x^1) + 125 \cos^2(x^1)) \right\} \end{aligned}$$

Dependence of designs on estimated parameters

Example (ctd')



1st option: Adaptive design

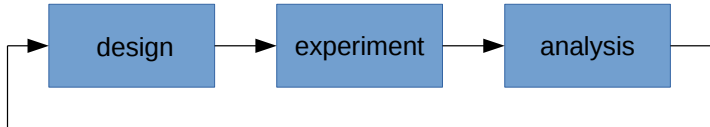
Since θ_{true} is unknown, a common approach is to use some prior estimate θ^0 of θ_{true} instead. But θ^0 may be far from θ_{true} and properties of locally optimal designs can be very sensitive to changes in θ .

Remedy

A first way out to obtain a robust design procedure is to provide adaptivity through application of a sequential design.

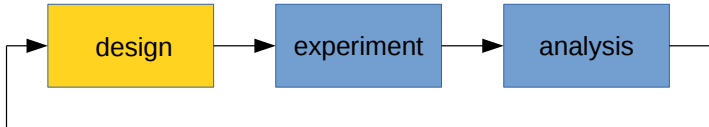
General scheme of adaptive design

Experimentation and estimation steps are alternated. At each stage, a locally optimal sensor location is determined based on the current parameter estimates (nominal parameter values can be assumed as initial guesses for the first stage), measurements are taken at the newly calculated sensor positions, and the data obtained are then analysed and used to update the parameter estimates.



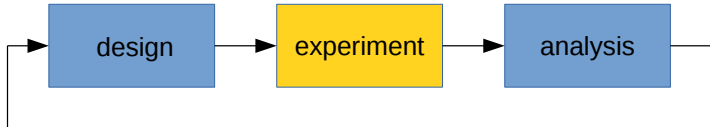
General scheme of adaptive design

Experimentation and estimation steps are alternated. At each stage, a locally optimal sensor location is determined based on the current parameter estimates (nominal parameter values can be assumed as initial guesses for the first stage), measurements are taken at the newly calculated sensor positions, and the data obtained are then analysed and used to update the parameter estimates.



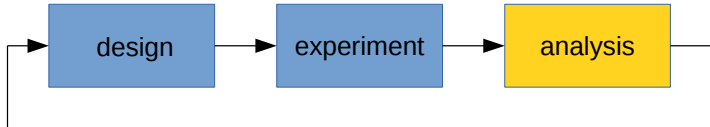
General scheme of adaptive design

Experimentation and estimation steps are alternated. At each stage, a locally optimal sensor location is determined based on the current parameter estimates (nominal parameter values can be assumed as initial guesses for the first stage), measurements are taken at the newly calculated sensor positions, and the data obtained are then analysed and used to update the parameter estimates.



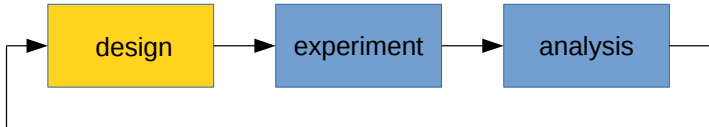
General scheme of adaptive design

Experimentation and estimation steps are alternated. At each stage, a locally optimal sensor location is determined based on the current parameter estimates (nominal parameter values can be assumed as initial guesses for the first stage), measurements are taken at the newly calculated sensor positions, and the data obtained are then analysed and used to update the parameter estimates.



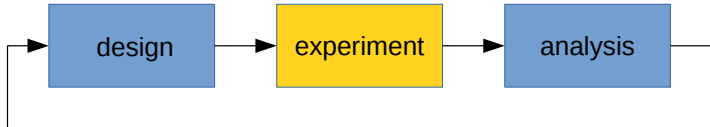
General scheme of adaptive design

Experimentation and estimation steps are alternated. At each stage, a locally optimal sensor location is determined based on the current parameter estimates (nominal parameter values can be assumed as initial guesses for the first stage), measurements are taken at the newly calculated sensor positions, and the data obtained are then analysed and used to update the parameter estimates.



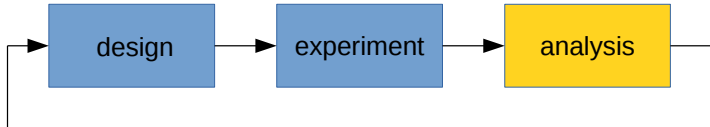
General scheme of adaptive design

Experimentation and estimation steps are alternated. At each stage, a locally optimal sensor location is determined based on the current parameter estimates (nominal parameter values can be assumed as initial guesses for the first stage), measurements are taken at the newly calculated sensor positions, and the data obtained are then analysed and used to update the parameter estimates.



General scheme of adaptive design

Experimentation and estimation steps are alternated. At each stage, a locally optimal sensor location is determined based on the current parameter estimates (nominal parameter values can be assumed as initial guesses for the first stage), measurements are taken at the newly calculated sensor positions, and the data obtained are then analysed and used to update the parameter estimates.



Locally optimal design for stationary sensors

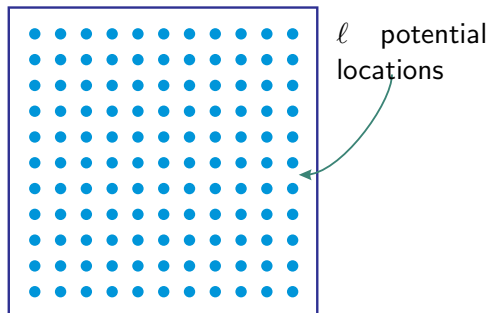
The approach adopted here

Adaptation of the notion of continuous designs (the approach set forth by Rafajłowicz in the 1980s) for a wide class of optimality criteria, as is common in modern optimum experimental design.

Observations

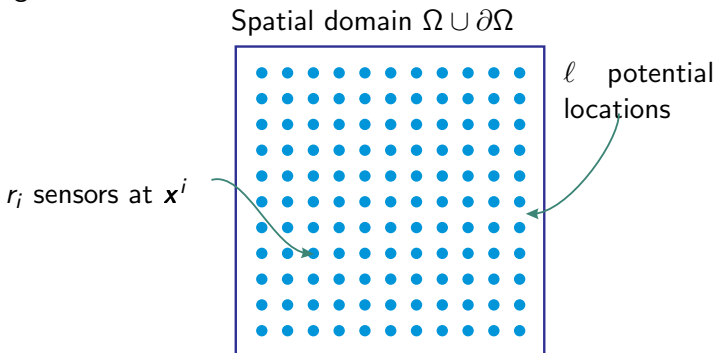
Given N sensors and a finite set $X = \{\mathbf{x}^1, \dots, \mathbf{x}^\ell\} \subset \Omega \cup \partial\Omega$ of points at which they can be placed, consider their fixed configuration.

Spatial domain $\Omega \cup \partial\Omega$



Observations

Given N sensors and a finite set $X = \{\mathbf{x}^1, \dots, \mathbf{x}^\ell\} \subset \Omega \cup \partial\Omega$ of points at which they can be placed, consider their fixed configuration.



Output equation

For pointwise stationary sensors we have

$$z_{ij}(t) = y(\mathbf{x}^i, t; \boldsymbol{\theta}) + \varepsilon_{ij}(t), \quad t \in T$$

for $i = 1, \dots, \ell$ and $j = 1, \dots, r_i$, where

$$\sum_{i=1}^{\ell} r_i = N$$

$\varepsilon_{ij}(\cdot)$ – white Gaussian measurement noise.

Output equation

For pointwise stationary sensors we have

$$z_{ij}(t) = y(\mathbf{x}^i, t; \boldsymbol{\theta}) + \varepsilon_{ij}(t), \quad t \in T$$

for $i = 1, \dots, \ell$ and $j = 1, \dots, r_i$, where

$$\sum_{i=1}^{\ell} r_i = N$$

$\varepsilon_{ij}(\cdot)$ – white Gaussian measurement noise.

Average Fisher Information Matrix (FIM)

Up to a constant multiplier, we have

$$\mathbf{M} = \frac{1}{Nt_f} \sum_{i=1}^{\ell} r_i \int_0^{t_f} \mathbf{g}(\mathbf{x}^i, t) \mathbf{g}^T(\mathbf{x}^i, t) dt$$

Problem of weight optimization

Minimize some **measure Ψ of the FIM**, e.g.,

$$\Psi(M) = \ln \det(M^{-1}) \quad \text{or} \quad \Psi(M) = \text{trace}(M^{-1})$$

by finding a solution (called **design**) of the form

$$\xi = \left\{ \begin{matrix} \mathbf{x}^1, & \dots, & \mathbf{x}^\ell \\ p_1, & \dots, & p_\ell \end{matrix} \right\}$$

where $p_i = r_i/N$, i.e., the weight p_i defining the proportion of sensors placed at $\mathbf{x}^i$, $i = 1, \dots, N$.

Problem of weight optimization

Minimize some measure Ψ of the FIM, e.g.,

$$\Psi(M) = \ln \det(M^{-1}) \quad \text{or} \quad \Psi(M) = \text{trace}(M^{-1})$$

by finding a solution (called design) of the form

$$\xi = \left\{ \begin{matrix} \mathbf{x}^1, & \dots, & \mathbf{x}^\ell \\ p_1, & \dots, & p_\ell \end{matrix} \right\}$$

where $p_i = r_i/N$, i.e., the weight p_i defining the proportion of sensors placed at $\mathbf{x}^i$, $i = 1, \dots, N$.

Problem relaxation

The problem still has a discrete nature, as the sought weights p_i are integer multiples of $1/N$ which sum up to unity. Consequently, calculus techniques cannot be exploited in the solution.

Problem relaxation

The problem still has a discrete nature, as the sought weights p_i are integer multiples of $1/N$ which sum up to unity. Consequently, calculus techniques cannot be exploited in the solution.

Thus we extend the definition of the design: when N is large, the feasible p_i 's can be considered as **any real numbers in the interval $[0, 1]$** which sum up to unity, and not necessarily integer multiples of $1/N$, i.e., elements of the **canonical simplex**.

Ultimate formulation

Find $\xi^* = \left\{ \begin{matrix} \mathbf{x}^1, \dots, \mathbf{x}^\ell \\ p_1^*, \dots, p_\ell^* \end{matrix} \right\}$ such that

$$\mathbf{p}^* = \arg \min_{\mathbf{p} \in \mathbb{S}} \Psi[M(\xi)]$$

where $\mathbb{S}$ is the canonical simplex, subject to

$$M(\xi) = \sum_{i=1}^{\ell} p_i \Upsilon(\mathbf{x}^i)$$

Here $\Upsilon(\mathbf{x}) = \frac{1}{t_f} \int_0^{t_f} \mathbf{g}(\mathbf{x}, t) \mathbf{g}^T(\mathbf{x}, t) dt$ is the fixed contribution of the point $\mathbf{x}$ to the FIM.

Ultimate formulation

Find $\xi^* = \left\{ \begin{matrix} \mathbf{x}^1, \dots, \mathbf{x}^\ell \\ p_1^*, \dots, p_\ell^* \end{matrix} \right\}$ such that

$$\mathbf{p}^* = \arg \min_{\mathbf{p} \in \mathbb{S}} \Psi[M(\xi)]$$

where $\mathbb{S}$ is the canonical simplex, subject to

$$M(\xi) = \sum_{i=1}^{\ell} p_i \mathbf{r}(\mathbf{x}^i)$$

Here $\mathbf{r}(\mathbf{x}) = \frac{1}{t_f} \int_0^{t_f} \mathbf{g}(\mathbf{x}, t) \mathbf{g}^T(\mathbf{x}, t) dt$ is the fixed contribution of the point $\mathbf{x}$ to the FIM.

Ultimate formulation

Find $\xi^* = \left\{ \begin{matrix} \mathbf{x}^1, \dots, \mathbf{x}^\ell \\ p_1^*, \dots, p_\ell^* \end{matrix} \right\}$ such that

$$\mathbf{p}^* = \arg \min_{\mathbf{p} \in \mathbb{S}} \Psi[M(\xi)]$$

where $\mathbb{S}$ is the canonical simplex, subject to

$$M(\xi) = \sum_{i=1}^{\ell} p_i \Upsilon(\mathbf{x}^i)$$

Here $\Upsilon(\mathbf{x}) = \frac{1}{t_f} \int_0^{t_f} \mathbf{g}(\mathbf{x}, t) \mathbf{g}^T(\mathbf{x}, t) dt$ is the fixed contribution of the point $\mathbf{x}$ to the FIM.

Ultimate formulation

Find $\xi^* = \left\{ \begin{matrix} \mathbf{x}^1, \dots, \mathbf{x}^\ell \\ p_1^*, \dots, p_\ell^* \end{matrix} \right\}$ such that

$$\mathbf{p}^* = \arg \min_{\mathbf{p} \in \mathbb{S}} \Psi[M(\xi)]$$

where $\mathbb{S}$ is the canonical simplex, subject to

$$M(\xi) = \sum_{i=1}^{\ell} p_i \mathbf{r}(\mathbf{x}^i)$$

Here $\mathbf{r}(\mathbf{x}) = \frac{1}{t_f} \int_0^{t_f} \mathbf{g}(\mathbf{x}, t) \mathbf{g}^T(\mathbf{x}, t) dt$ is the fixed contribution of the point $\mathbf{x}$ to the FIM.

Differentiability of the design criterion

Define

$$\dot{\Psi}(\xi) = \left. \frac{\partial \Psi(\mathbf{M})}{\partial \mathbf{M}} \right|_{\mathbf{M}=\mathbf{M}(\xi)} = \left[\left. \frac{\partial \Psi(\mathbf{M})}{\partial M_{ij}} \right] \right|_{\mathbf{M}=\mathbf{M}(\xi)}$$

Note that we have

$$\begin{aligned} \frac{\partial}{\partial \mathbf{M}} \log \det(\mathbf{M}^{-1}) &= -(\mathbf{M}^{-1})^T \\ \frac{\partial}{\partial \mathbf{M}} \text{trace}(\mathbf{M}^{-1}) &= -(\mathbf{M}^{-2})^T \\ \frac{\partial}{\partial \mathbf{M}} (-\text{trace}(\mathbf{M})) &= -\mathbf{I} \end{aligned}$$

Differentiability of the design criterion

We introduce crucial functions

$$c(\xi) = -\text{trace} \left[\dot{\Psi}(\xi) \mathbf{M}(\xi) \right]$$

$$\phi(\mathbf{x}, \xi) = -\text{trace} \left[\dot{\Psi}(\xi) \Upsilon(\mathbf{x}) \right]$$

$\Psi[\mathbf{M}(\xi)]$	$\phi(\mathbf{x}, \xi)$	$c(\xi)$
$\ln \det(\mathbf{M}^{-1}(\xi))$	$\frac{1}{t_f} \int_0^{t_f} \mathbf{g}^\top(\mathbf{x}, t) \mathbf{M}^{-1}(\xi) \mathbf{g}(\mathbf{x}, t) dt$	m
$\text{trace}(\mathbf{M}^{-1}(\xi))$	$\frac{1}{t_f} \int_0^{t_f} \mathbf{g}^\top(\mathbf{x}, t) \mathbf{M}^{-2}(\xi) \mathbf{g}(\mathbf{x}, t) dt$	$\text{trace}(\mathbf{M}^{-1}(\xi))$
$-\text{trace}(\mathbf{M}(\xi))$	$\frac{1}{t_f} \int_0^{t_f} \mathbf{g}^\top(\mathbf{x}, t) \mathbf{g}(\mathbf{x}, t) dt$	$\text{trace}(\mathbf{M}(\xi))$

Characterizations of solutions

Theorem (Existence of solution)

An optimal design exists comprising no more than $m(m+1)/2$ support points. Moreover, the set of optimal designs is convex.

Characterizations of solutions

Theorem (Existence of solution)

An optimal design exists comprising no more than $m(m+1)/2$ support points. Moreover, the set of optimal designs is convex.

Theorem (Equivalence theorem)

The following characterizations of an optimal design ξ^* are equivalent:

- ① the design ξ^* minimizes $\Psi[M(\xi)]$,
- ② the design ξ^* minimizes $\max_{\mathbf{x} \in X} \phi(\mathbf{x}, \xi) - c(\xi)$,
- ③ $\max_{\mathbf{x} \in X} \phi(\mathbf{x}, \xi^*) = c(\xi^*)$.

All the designs satisfying 1–3 and their convex combinations have the same $M(\xi^*)$.

Convex combination of designs

For any designs

$$\xi_1 = \left\{ \begin{matrix} \mathbf{x}^1, & \dots, & \mathbf{x}^\ell \\ p_1^{(1)}, & \dots, & p_\ell^{(1)} \end{matrix} \right\}, \quad \xi_2 = \left\{ \begin{matrix} \mathbf{x}^1, & \dots, & \mathbf{x}^\ell \\ p_1^{(2)}, & \dots, & p_\ell^{(2)} \end{matrix} \right\}$$

we define

$$\begin{aligned} \xi &= (1 - \lambda)\xi_1 + \lambda\xi_2 \\ &= \left\{ \begin{matrix} \mathbf{x}^1, & \dots, & \mathbf{x}^\ell \\ (1 - \lambda)p_1^{(1)} + \lambda p_1^{(2)}, & \dots, & (1 - \lambda)p_\ell^{(1)} + \lambda p_\ell^{(2)} \end{matrix} \right\} \end{aligned}$$

Numerical approach: A-optimality

For $\Psi[M(\xi)] = \text{trace}(M^{-1}(\xi))$, we have to find p and the additional variable $q \in \mathbb{R}^m$ to minimize

$$J(p, q) = \mathbf{1}^T q$$

subject to $\mathbf{1}^T p = 1$, $p \succeq \mathbf{0}$ and the Linear Matrix Inequalities (LMIs)

$$\left[\begin{array}{c|c} M(\xi) & e_j \\ \hline e_j^T & q_j \end{array} \right] \succeq 0, \quad j = 1, \dots, m.$$

where e_j is j -th coordinate vector.

Numerical approach: E-optimality

For $\Psi[M(\xi)] = \lambda_{\max}[M^{-1}(\xi)]$ we have to find p and $q \in \mathbb{R}$ to minimize

$$\tilde{J}(p, q) = -q$$

subject to

$$\mathbf{1}^T p = 1, \quad p \succeq \mathbf{0}$$

and the LMI $M(\xi) \succeq qI$.

Numerical approach: E-optimality

For $\Psi[M(\xi)] = \lambda_{\max}[M^{-1}(\xi)]$ we have to find p and $q \in \mathbb{R}$ to minimize

$$\tilde{J}(p, q) = -q$$

subject to

$$\mathbf{1}^T p = 1, \quad p \succeq \mathbf{0}$$

and the LMI $M(\xi) \succeq qI$.

Note that $\lambda_{\max}[M^{-1}(\xi)]$ is **non-differentiable**!

Numerical approach: Torsney's algorithm for D-optimality

For $\Psi[M(\xi)] = \ln \det[M^{-1}(\xi)]$ apply

- ① Guess $p_i^{(0)}$, $i = 1, \dots, \ell$. Set $k = 0$.
- ② If $\phi(\mathbf{x}^i, \xi^{(k)}) = \text{trace}(\mathbf{T}(\mathbf{x}^i)M^{-1}(\xi^{(k)})) \leq m$, $i = 1, \dots, \ell$, then STOP.
- ③ Evaluate

$$p_i^{(k+1)} = p_i^{(k)} \frac{\phi(\mathbf{x}^i, \xi^{(k)})}{m}, \quad i = 1, \dots, \ell.$$

Form $\xi^{(k+1)}$, increment k and go to Step 2.

Numerical approach: Torsney's algorithm for D-optimality

For $\Psi[M(\xi)] = \ln \det[M^{-1}(\xi)]$ apply

- 1 Guess $p_i^{(0)}$, $i = 1, \dots, \ell$. Set $k = 0$.
- 2 If $\phi(\mathbf{x}^i, \xi^{(k)}) = \text{trace}(\mathbf{\Upsilon}(\mathbf{x}^i)M^{-1}(\xi^{(k)})) \leq m$, $i = 1, \dots, \ell$, then STOP.
- 3 Evaluate

$$p_i^{(k+1)} = p_i^{(k)} \frac{\phi(\mathbf{x}^i, \xi^{(k)})}{m}, \quad i = 1, \dots, \ell.$$

Form $\xi^{(k+1)}$, increment k and go to Step 2.

Numerical approach: Torsney's algorithm for D-optimality

For $\Psi[M(\xi)] = \ln \det[M^{-1}(\xi)]$ apply

- 1 Guess $p_i^{(0)}$, $i = 1, \dots, \ell$. Set $k = 0$.
- 2 If $\phi(\mathbf{x}^i, \xi^{(k)}) = \text{trace}(\mathbf{T}(\mathbf{x}^i)M^{-1}(\xi^{(k)})) \leq m$, $i = 1, \dots, \ell$, then STOP.
- 3 Evaluate

$$p_i^{(k+1)} = p_i^{(k)} \frac{\phi(\mathbf{x}^i, \xi^{(k)})}{m}, \quad i = 1, \dots, \ell.$$

Form $\xi^{(k+1)}$, increment k and go to Step 2.

Properties of this multiplicative algorithm

The idea is reminiscent of the EM algorithm used for maximum likelihood estimation.

Theorem

Assume that $\{\xi^{(k)}\}$ is a sequence of iterates constructed by the outlined algorithm. Then the sequence $\{\det(\mathbf{M}(\xi^{(k)}))\}$ is nondecreasing and

$$\lim_{k \rightarrow \infty} \det(\mathbf{M}(\xi^{(k)})) = \max_{\xi} \det(\mathbf{M}(\xi))$$

Numerical example

Consider the nonlinear system

$$\frac{\partial y}{\partial t} = a \frac{\partial^2 y}{\partial x^2} - b y^3, \quad x \in (0, 1), \quad t \in (0, t_f)$$

subject to the conditions

$$\begin{aligned} y(x, 0) &= c \psi(x), & x &\in (0, 1) \\ y(0, t) &= y(1, t) = 0, & t &\in (0, t_f) \end{aligned}$$

where $t_f = 0.8$, $\psi(x) = \sin(\pi x) + \sin(3\pi x)$.

Adopted method

Determine a D-optimum design of the form

$$\xi^* = \left\{ \begin{matrix} x^1, & \dots, & x^\ell \\ p_1^*, & \dots, & p_\ell^* \end{matrix} \right\}$$

for recovering $\theta = (a, b, c)$, where the support points are

$$x^i = 0.05(i - 1), \quad i = 1, \dots, \ell = 21$$

Apply the multiplicative algorithm just mentioned. (MATLAB's routine `pdepe` comes here in handy.)

Results

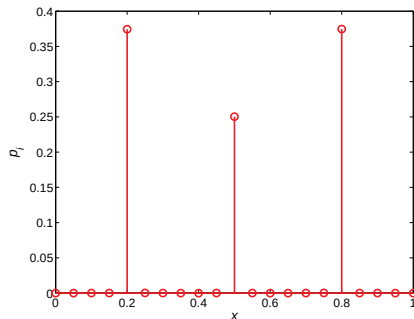


Figure: Optimal weights

Results

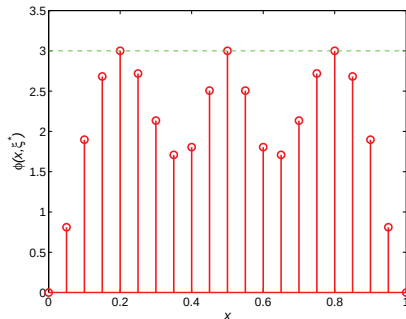


Figure: Derivative function $\phi(x, \xi^*)$

Results

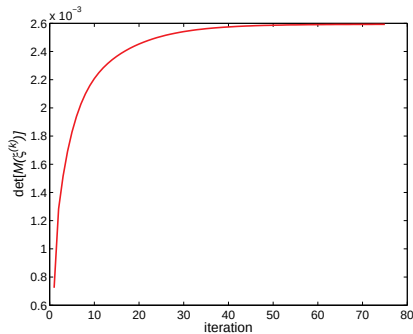
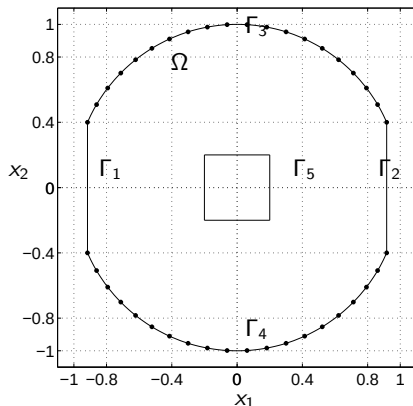
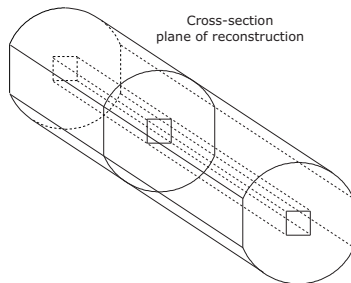


Figure: Monotonic increase in $\det[M(\xi^{(k)})]$

Simulation example: material flaw detection

Consider a metallic core with a rectangular crack inside.



The left and right boundaries Γ_1 and Γ_2 are subjected to constant voltages forcing the flow of a low density steady current.

Computer-assisted impedance tomography

Electrical voltage distribution y is described by the Laplace equation

$$-\operatorname{div}(\gamma(\mathbf{x})\nabla y(\mathbf{x})) = 0, \quad \mathbf{x} \in \Omega,$$

with boundary conditions

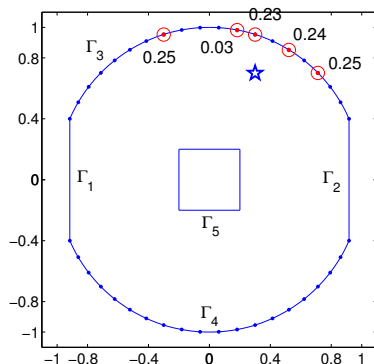
$$\begin{cases} y(\mathbf{x}) = 10 & \mathbf{x} \in \Gamma_1, \\ y(\mathbf{x}) = -10 & \mathbf{x} \in \Gamma_2, \\ \frac{\partial y(\mathbf{x})}{\partial \mathbf{n}} = 0 & \mathbf{x} \in \Gamma_i, \quad i = 3, 4, 5. \end{cases}$$

Conductivity coefficient

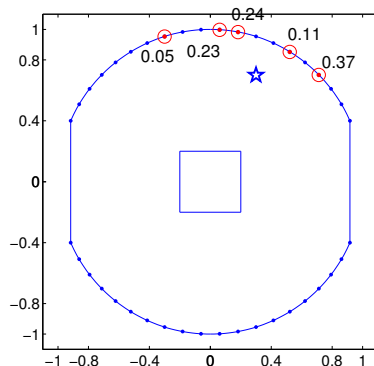
$$\gamma(\mathbf{x}) = 0.1 + 0.2 \exp \left(-\theta_1(x_1 - \theta_2)^2 - \theta_3(x_2 - \theta_4)^2 \right)$$

$$\theta^0 = (30.0, 0.3, 80.0, 0.7)$$

Results



D-optimal



A-optimal

Comments

The approach is useful as

Comments

The approach is useful as

- It indicates spatial regions which provide the most important information about the parameters, while quantifying their influence.

Comments

The approach is useful as

- It indicates spatial regions which provide the most important information about the parameters, while quantifying their influence.
- Instead of several sensors placed at a point, we can use a more precise measurement device at that point.

Comments

The approach is useful as

- It indicates spatial regions which provide the most important information about the parameters, while quantifying their influence.
- Instead of several sensors placed at a point, we can use a more precise measurement device at that point.
- Design weights can be interpreted as measurement frequencies.

Comments

The approach is useful as

- It indicates spatial regions which provide the most important information about the parameters, while quantifying their influence.
- Instead of several sensors placed at a point, we can use a more precise measurement device at that point.
- Design weights can be interpreted as measurement frequencies.
- Extensions towards simultaneously optimizing spatial points are possible.

Comments

The approach is useful as

- It indicates spatial regions which provide the most important information about the parameters, while quantifying their influence.
- Instead of several sensors placed at a point, we can use a more precise measurement device at that point.
- Design weights can be interpreted as measurement frequencies.
- Extensions towards simultaneously optimizing spatial points are possible.
- With a little turning up, the results can be employed when attacking the setting with purely binary weights ($r_i = 0$ or $r_i = 1$), see material on simplicial decomposition in Part 2.