

Multiple facets of modern optimal experimental design

Dariusz Uciński

Institute of Control and Computation Engr., University of Zielona Góra, Poland

9th German-Polish Conference on Optimization Methods and
Applications, Białystok, 21 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.

Measurement accuracy: intro to design

The weights of objects A , B , C and D 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.



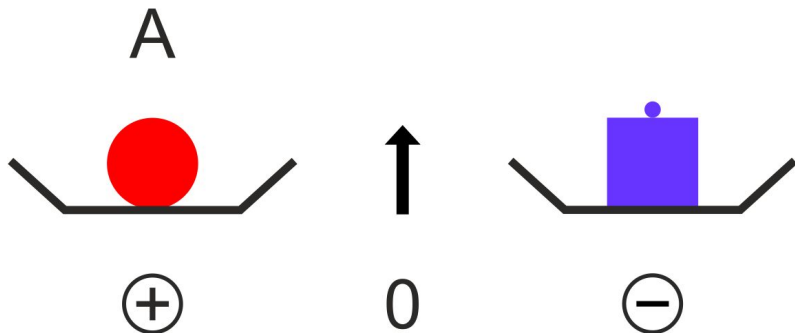
Weighing four objects

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 , m_B , m_C and m_D the true weights.

Experimental budget: We may take only four measurements.

Weighing four objects: Strategy 1

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



Weighing four objects: Strategy 1

Obviously, we get

$$m_A = y_1 \pm \sigma$$

$$m_B = y_2 \pm \sigma$$

$$m_C = y_3 \pm \sigma$$

$$m_D = y_4 \pm \sigma$$

For example, if y_1 , y_2 , y_3 and y_4 are 10, 25, 15 and 30, respectively, and σ is 0.1 gramme, we have

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

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

$$m_C = 15 \pm 0.1 \text{ grammes}$$

$$m_D = 30 \pm 0.1 \text{ grammes}$$

Weighing four objects: Strategy 1

We can express these results using the matrix notation

$$y = F\theta + \epsilon$$

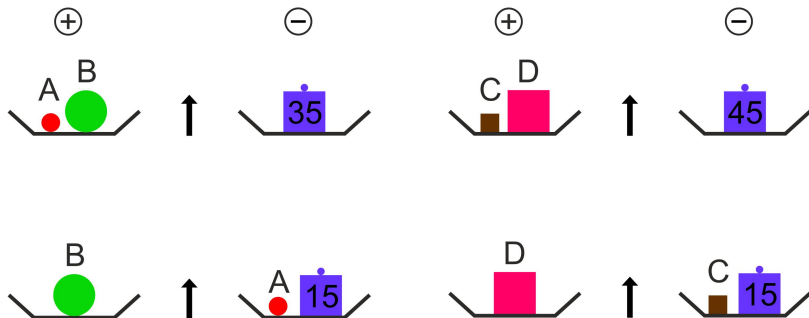
where

$$y = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix}, \quad F = \begin{bmatrix} +1 & 0 & 0 & 0 \\ 0 & +1 & 0 & 0 \\ 0 & 0 & +1 & 0 \\ 0 & 0 & 0 & +1 \end{bmatrix}$$
$$\theta = \begin{bmatrix} m_A \\ m_B \\ m_C \\ m_D \end{bmatrix}, \quad \epsilon = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \epsilon_4 \end{bmatrix}$$

Observe that $\det(F^T F) = 1$.

Weighing four objects: Strategy 2

But consider another strategy:



Weighing four objects: Strategy 2

Neglecting measurement errors, we get

$$\left\{ \begin{array}{l} +m_A + m_B = y_1 \\ -m_A + m_B = y_2 \\ +m_C + m_D = y_3 \\ -m_C + m_D = y_4 \end{array} \right. \Rightarrow \left\{ \begin{array}{l} m_A = \frac{1}{2}(y_1 - y_2) \\ m_B = \frac{1}{2}(y_1 + y_2) \\ m_C = \frac{1}{2}(y_3 - y_4) \\ m_D = \frac{1}{2}(y_3 + y_4) \end{array} \right.$$

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

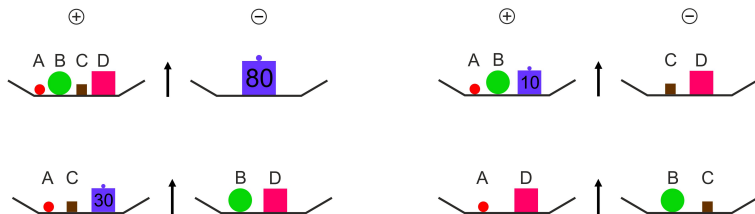
We have

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

Additionally, observe that $\det(F^T F) = 16$.

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

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

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

Weighing four objects: Strategy 3

We have

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

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

Better precision coincides with higher values of $\det(F^T F)$.

The matrix $M = F^T F$ is called the **Fisher information matrix** and experimental conditions maximizing its determinant are called **D-optimal**.

Why is $F^T F$ so important?

Assume that the observations are modelled by

$$y = F\theta + \varepsilon$$

where ε is measurement noise satisfying

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

and σ^2 is a constant noise variance.

Why is $F^T F$ so important?

Assume that the observations are modelled by

$$y = F\theta + \varepsilon$$

where ε is measurement noise satisfying

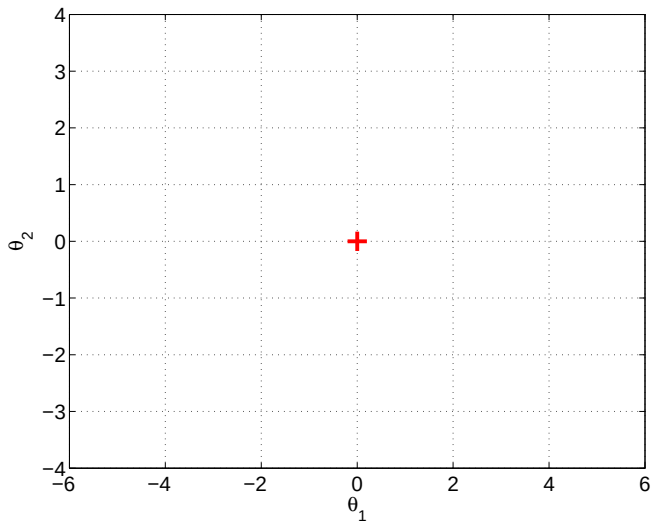
$$E[\varepsilon] = \mathbf{0}, \quad \text{cov}(\varepsilon) = E[\varepsilon\varepsilon^T] = \sigma^2 I$$

and σ^2 is a constant noise variance.

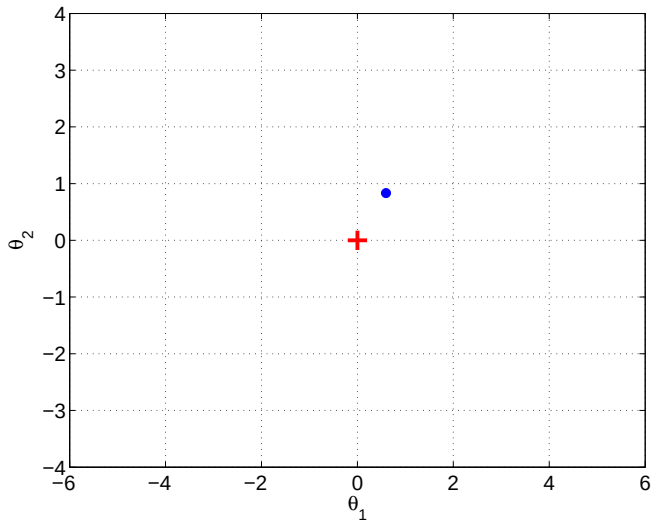
The parameter estimate $\hat{\theta} = (F^T F)^{-1} F^T y$ is then a random vector and

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

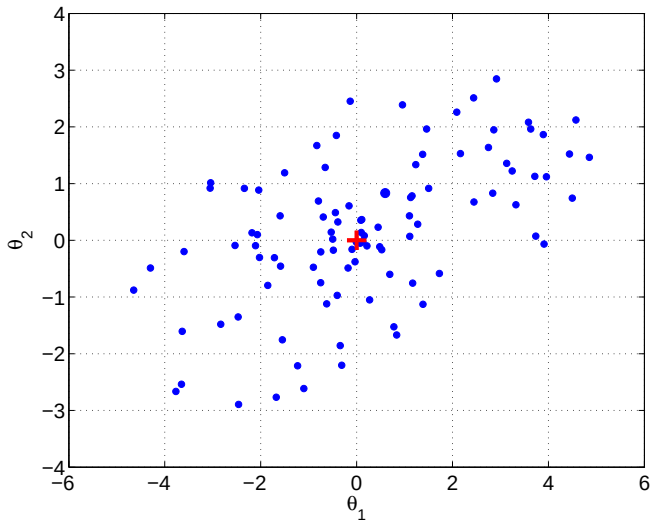
Uncertainty ellipsoid



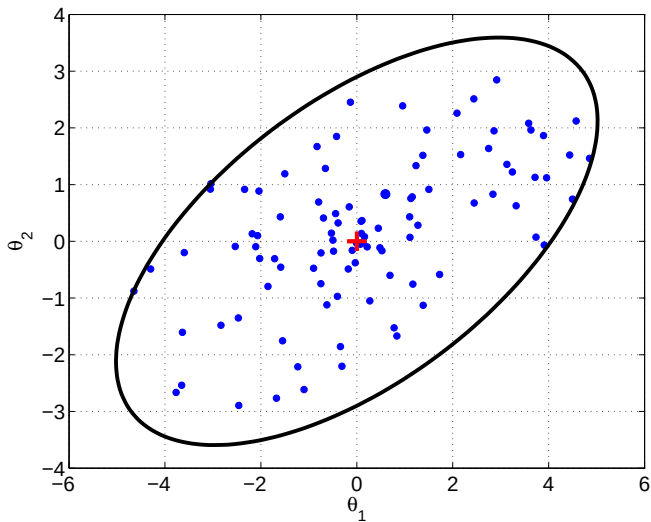
Uncertainty ellipsoid



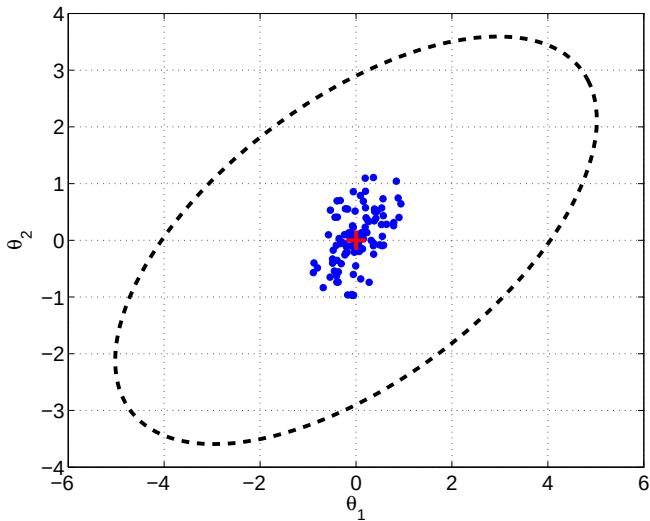
Uncertainty ellipsoid



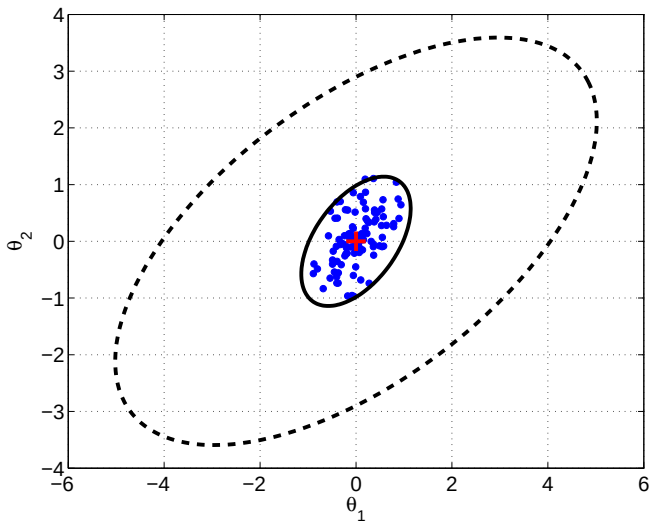
Uncertainty ellipsoid



Uncertainty ellipsoid



Uncertainty ellipsoid



Alphabetic optimality criteria

Setting $M = F^T F$, we define

- **D-optimality** criterion:

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

Alphabetic optimality criteria

Setting $M = F^T F$, we define

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

Alphabetic optimality criteria

Setting $M = F^T F$, we define

- **D-optimality** criterion:
 $\Psi(M) = \log \det(M^{-1}) = -\log \det(M)$ (the volume of the uncertainty ellipsoid is minimized)
- **A-optimality** criterion: $\Psi(M) = \text{trace}(M^{-1})$ (the mean axis length of the uncertainty ellipsoid is minimized)
- **E-optimality** criterion: $\Psi(M) = \lambda_{\max}(M^{-1})$ (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.

2^m factorial design

Consider the model

$$\begin{aligned} z = & \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 \\ & + \beta_4 x_1 x_2 + \beta_5 x_1 x_3 + \beta_6 x_2 x_3 \\ & + \beta_7 x_1 x_2 x_3 + \varepsilon \end{aligned}$$

where

- z – observed output
- $\mathbf{x} = [x_1, x_2, x_3]^T$ – input vector consisting of $m = 3$ factors
- $\boldsymbol{\theta} = [\beta_0, \dots, \beta_7]^T$ – vector of unknown parameters
- ε – random error (zero mean, independent and identically distributed)

Here $x_1 x_2$, $x_1 x_3$, $x_2 x_3$ and $x_1 x_2 x_3$ are two- and three-way interactions among the input factors.

Two-level factors

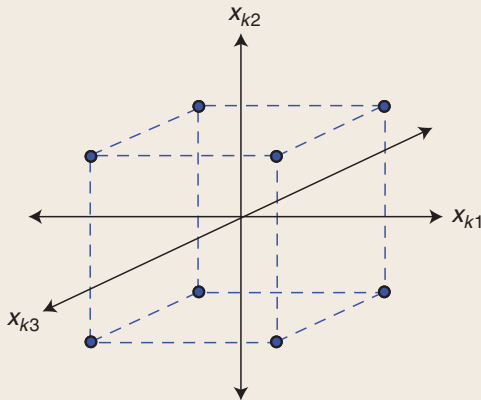
Factors may be quantitative, e.g.,

- 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

One replicate yields 2^m output values



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

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

Specifically,

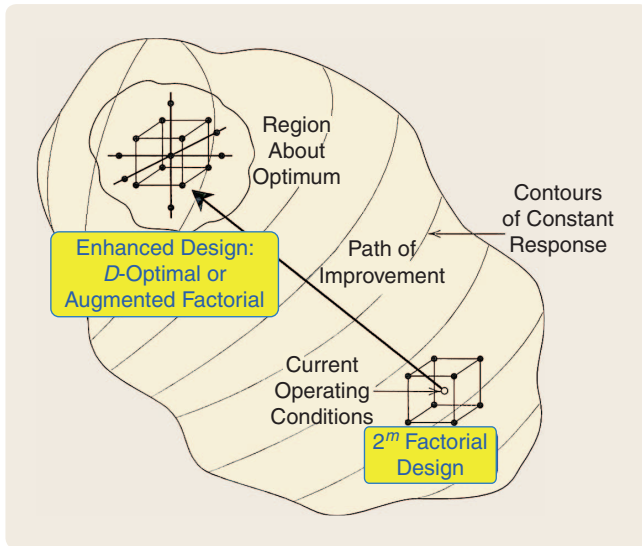
$$\hat{\beta}_1 = \frac{(z_2 + z_4 + z_6 + z_8) - (z_1 + z_3 + z_5 + z_7)}{8}$$

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

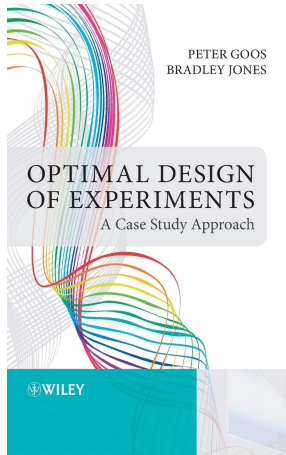
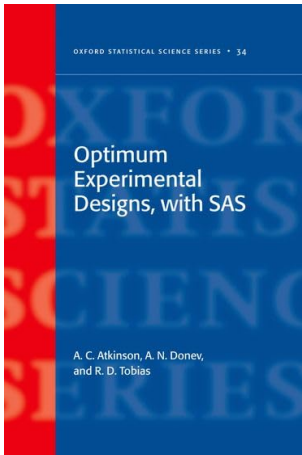
$$\hat{\beta}_3 = \frac{(z_5 + z_6 + z_7 + z_8) - (z_1 + z_2 + z_3 + z_4)}{8}$$

Factorial designs in this case are also D-optimal!

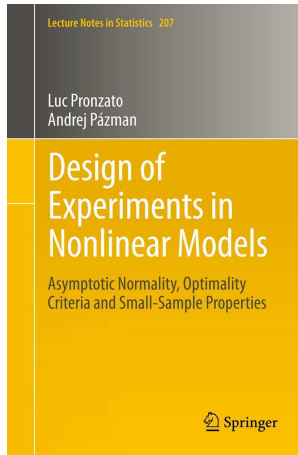
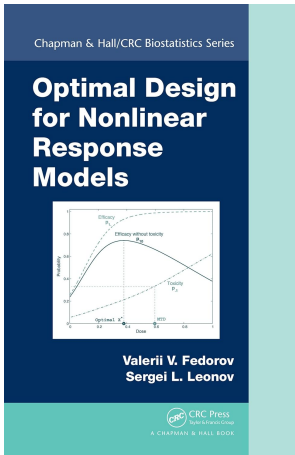
Response surface methodology



For more information



For more information



Optimum input design

Consider the linear dynamic system

$$\begin{aligned}\dot{x}(t) &= -a x(t) + u(t), \quad t \in [0, t_f] \\ x(0) &= c\end{aligned}$$

It is desired to estimate a based on noisy observations

$$z(t) = x(t) + \varepsilon(t), \quad t \in [0, t_f]$$

where $\varepsilon(t)$ is zero-mean white noise, $E[\varepsilon(t)\varepsilon(\tau)] = \sigma^2\delta(t - \tau)$.

Optimum input design

The information matrix (here scalar) is

$$M = \frac{1}{\sigma^2} \int_0^{t_f} x_a^2(t) dt$$

where $x_a(t) = \frac{\partial x(t; a)}{\partial a}$ is called the **sensitivity coefficient**.

The equation defining the sensitivity coefficient is obtained by differentiating the system equation with respect to a :

$$\begin{aligned}\dot{x}_a(t) &= -x(t) - a x_a(t) \\ x_a(0) &= 0\end{aligned}$$

Optimal control formulation

Find control $u(t)$ to maximize

$$\int_0^{t_f} x_a^2(t) dt$$

subject to

$$\begin{aligned}\dot{x}(t) &= -a x(t) + u(t), & x(0) &= c \\ \dot{x}_a(t) &= -x(t) - a x_a(t), & x_a(0) &= 0\end{aligned}$$

and

$$\int_0^{t_f} u^2(t) dt \leq E$$

where E is the maximal allowed input energy.

Optimal control formulation

Find control $u(t)$ to maximize

$$\int_0^{t_f} x_a^2(t) dt$$

subject to

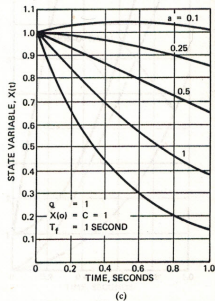
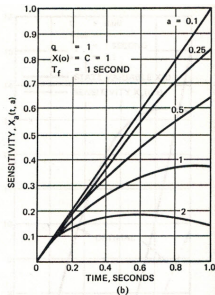
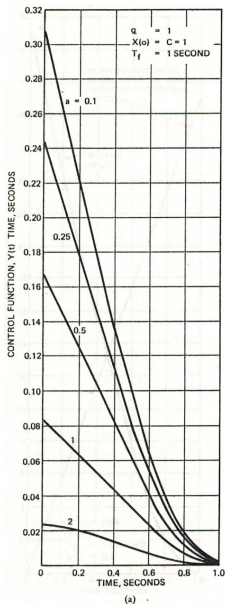
$$\begin{aligned}\dot{x}(t) &= -a x(t) + u(t), & x(0) &= c \\ \dot{x}_a(t) &= -x(t) - a x_a(t), & x_a(0) &= 0\end{aligned}$$

and

$$\int_0^{t_f} u^2(t) dt = E$$

where E is the maximal allowed input energy.

Numerical results



MATHEMATICAL CONCEPTS AND METHODS
IN SCIENCE AND ENGINEERING
Series Editor: Angelo Miele
Volume 25

Control, Identification, and Input Optimization

Robert Kalaba and Karl Spingarn

Plenum Press, New York and London

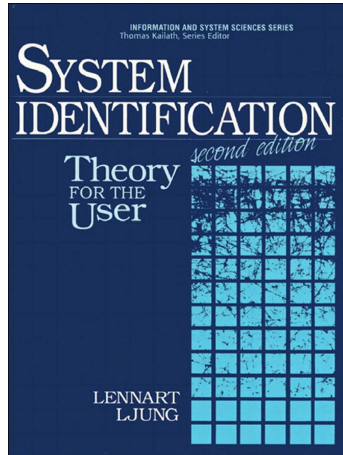
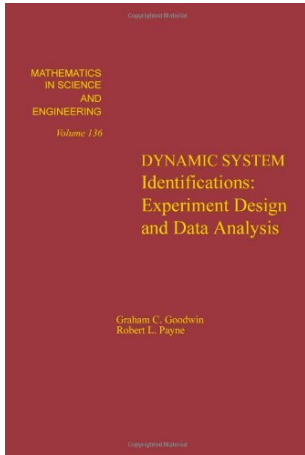
Dependence on the estimated parameter

The optimal input depends on the 'true' value of a since the state depends nonlinearly on a . Indeed,

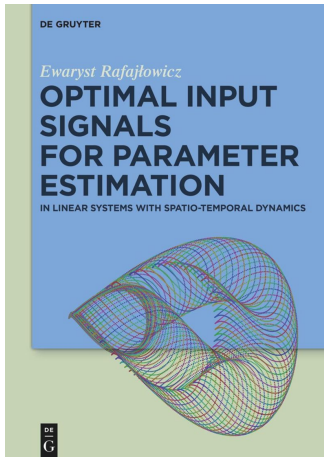
$$x(t) = \exp(-a t)c + \int_0^t \exp(-a(t - \tau))u(\tau) d\tau$$

This is typical when the output depends nonlinearly on the unknown parameters.

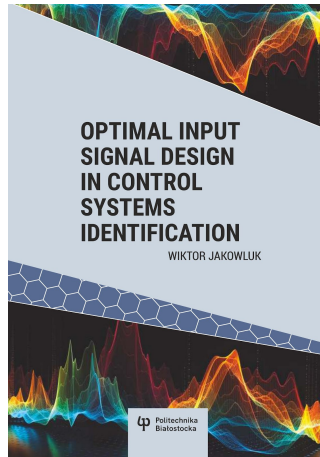
For more information



Recent monographs

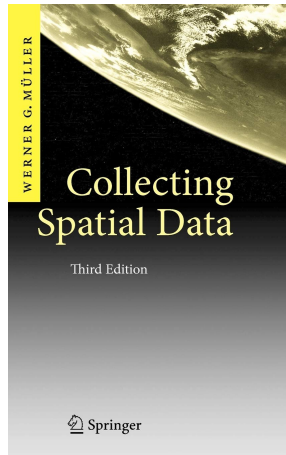
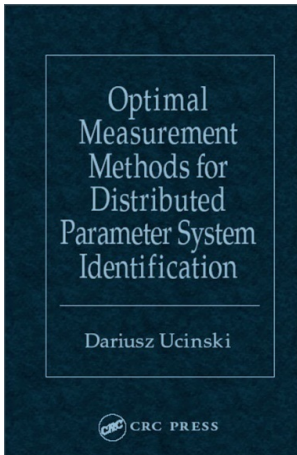


De Gruyter, 2022

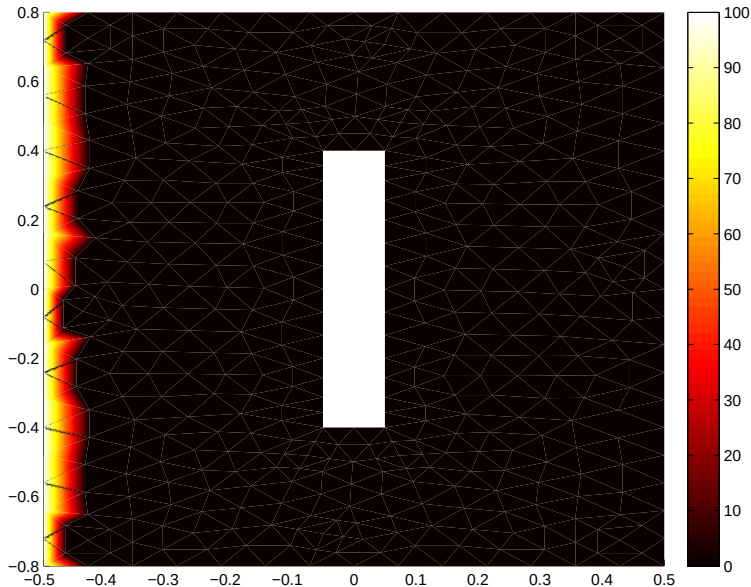


BUT, 2024

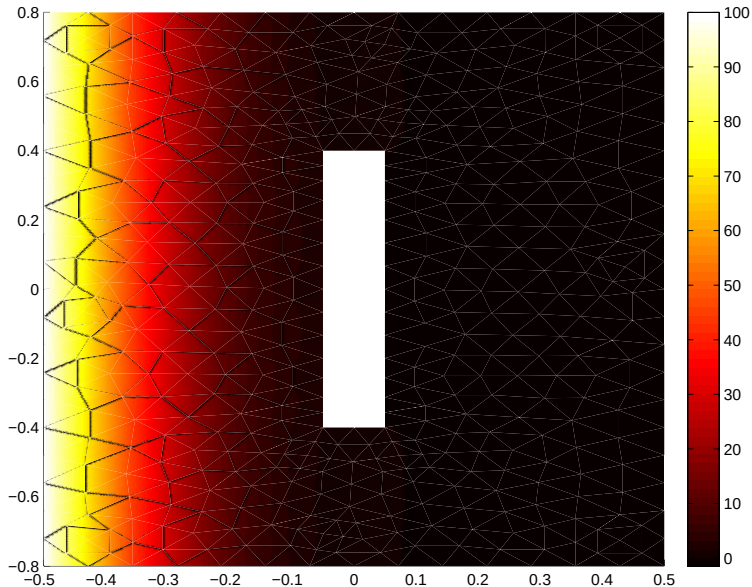
And now optimal sensor location



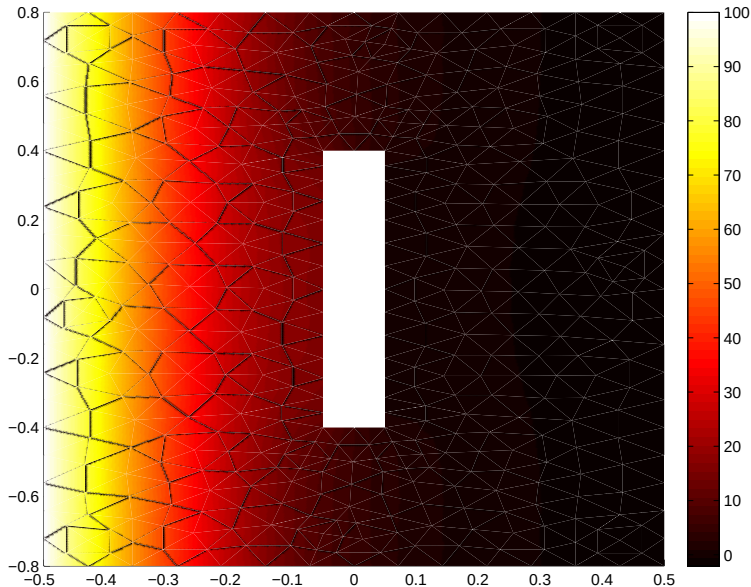
Heating a metal block with a crack



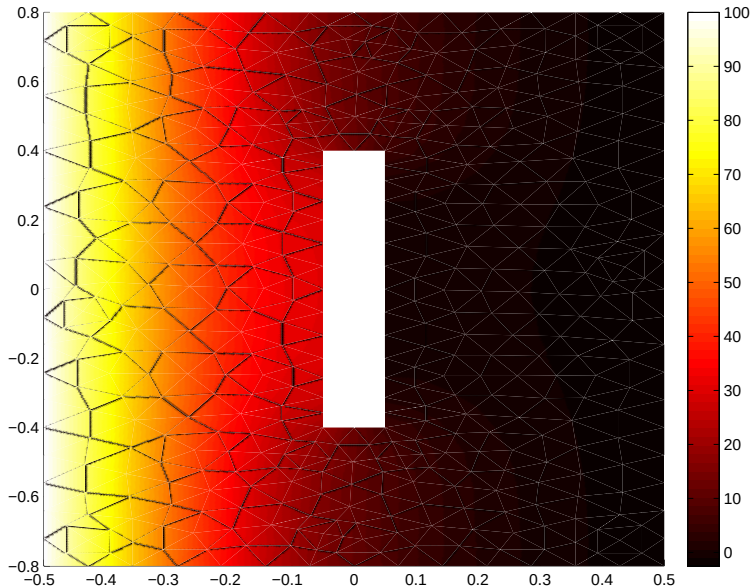
Heating a metal block with a crack



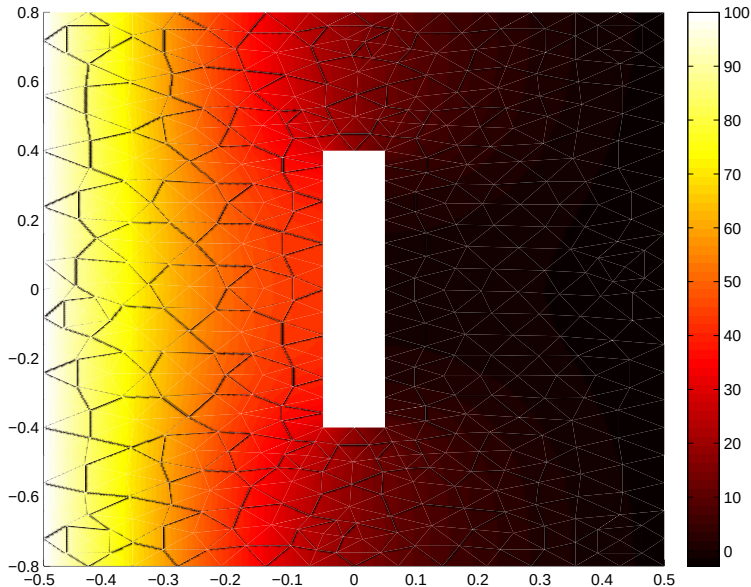
Heating a metal block with a crack



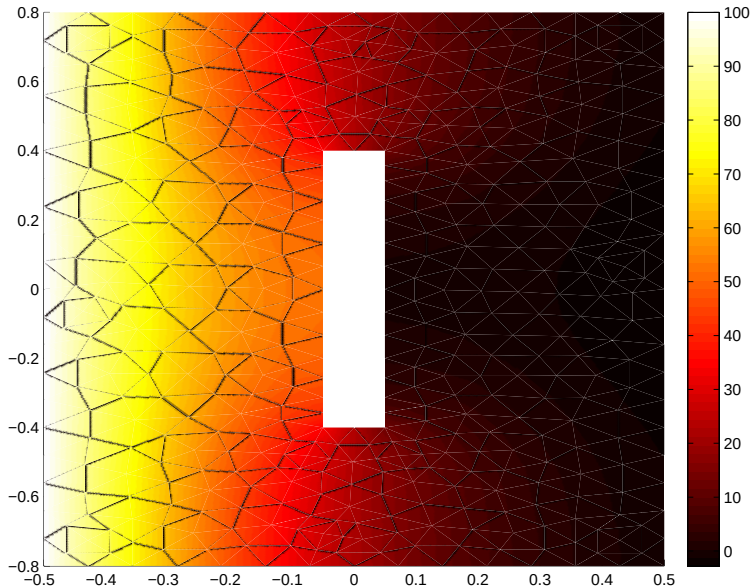
Heating a metal block with a crack



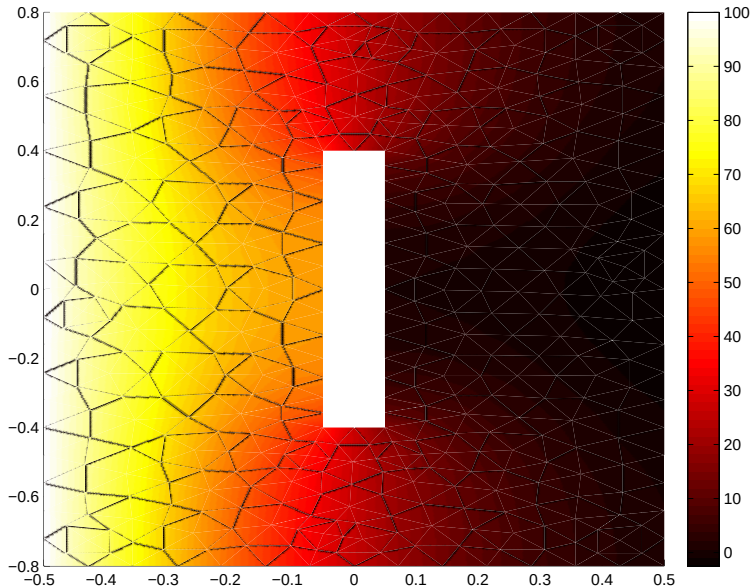
Heating a metal block with a crack



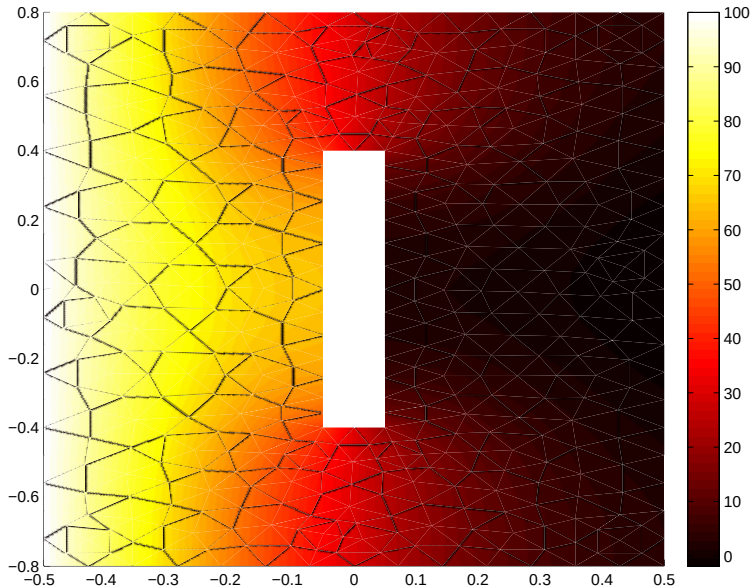
Heating a metal block with a crack



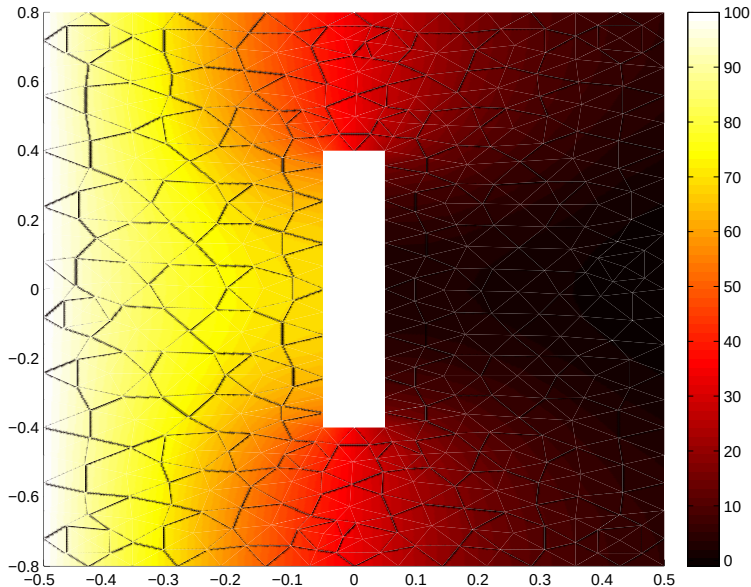
Heating a metal block with a crack



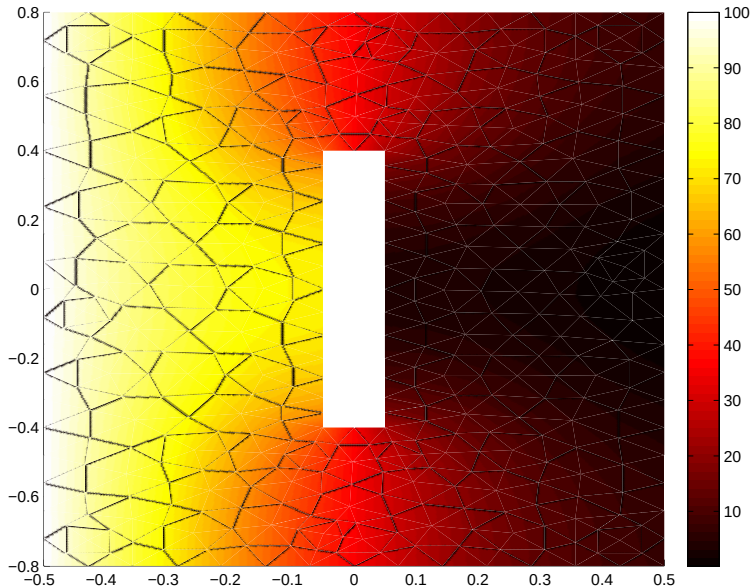
Heating a metal block with a crack



Heating a metal block with a crack



Heating a metal block with a crack



Example ctd': Mathematical model

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

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

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

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

and the **boundary conditions**

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

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

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

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.

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 ?

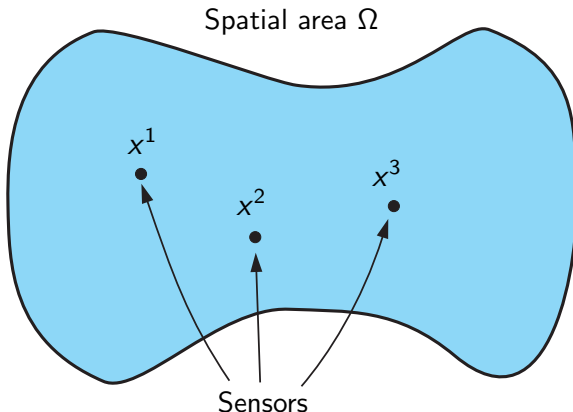
Such situations (PDEs and related sensor location problems) are common in engineering practice.

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 . . .

Strategies of taking measurements (1)

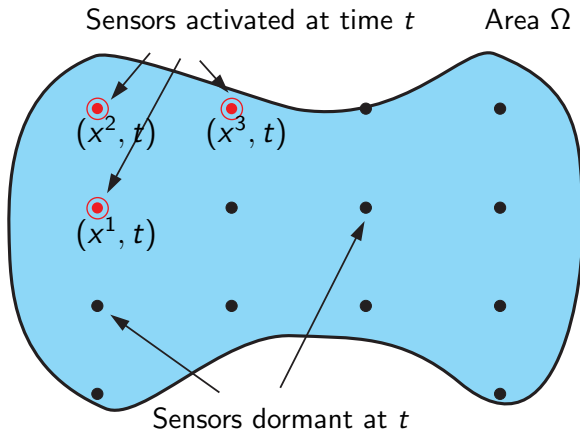
Stationary sensors



How to determine the best sensor configuration?

Strategies of taking measurements (2)

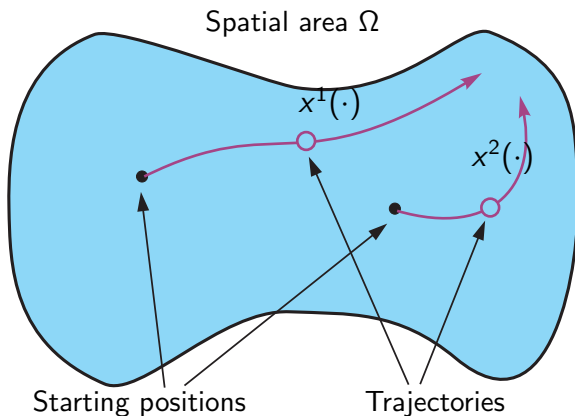
Scanning



How to select optimal locations at given time instants?

Strategies of taking measurements (3)

Mobile sensors



How to design optimal trajectories?

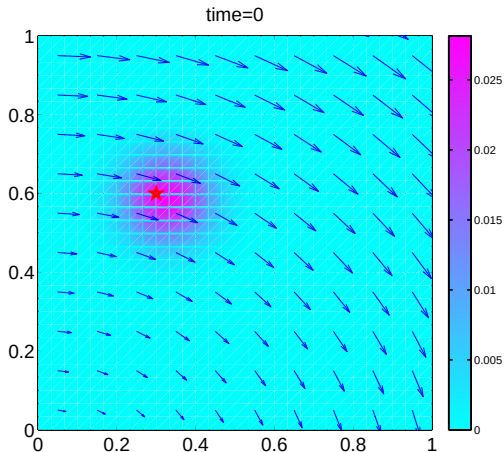
Approaches to optimal sensor location

- 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; Quereshi *et al.*, 1980; Rafajłowicz, since 1978; Kammer, 1990; 1992; Sun, 1994; Point *et al.*, 1996; Vande Wouwer *et al.*, 1999; Uciński, since 1990; Patan, since 2001).
- The dawn of 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).

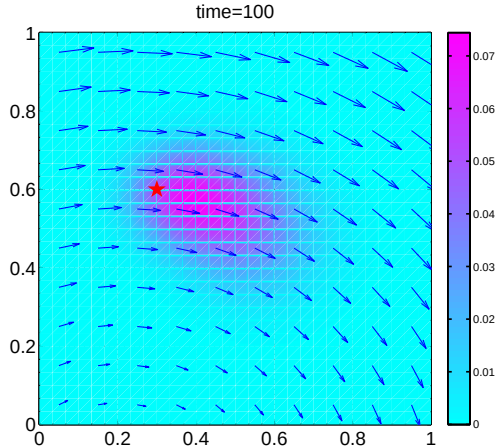
Approaches to optimal sensor location

- 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; Quereshi *et al.*, 1980; Rafajłowicz, 1978–1995; Kammer, 1990; 1992; Sun, 1994; Point *et al.*, 1996; Vande Wouwer *et al.*, 1999; Uciński, since 1990; Patan, since 2001);
- The dawn of 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).

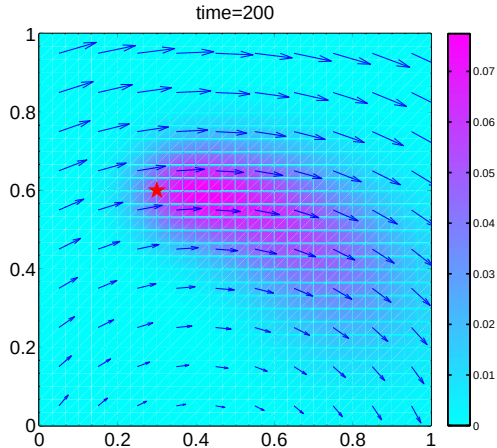
Example: Air pollution monitoring



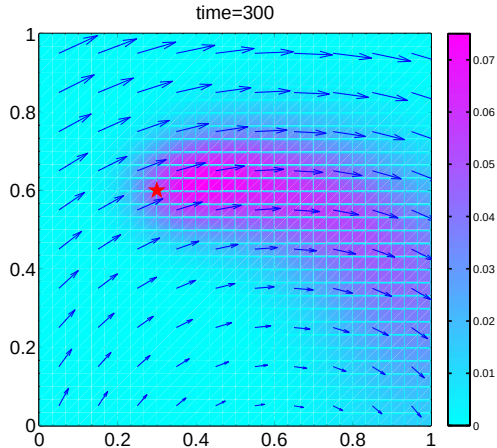
Example: Air pollution monitoring



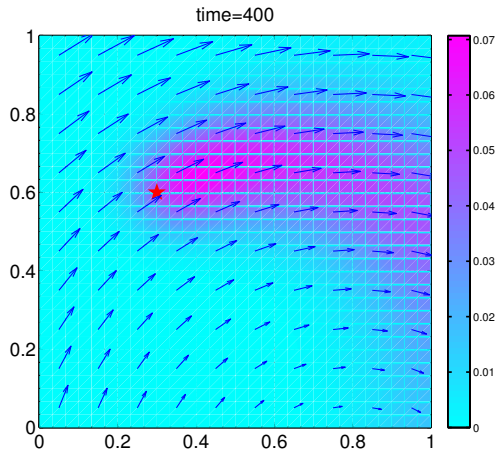
Example: Air pollution monitoring



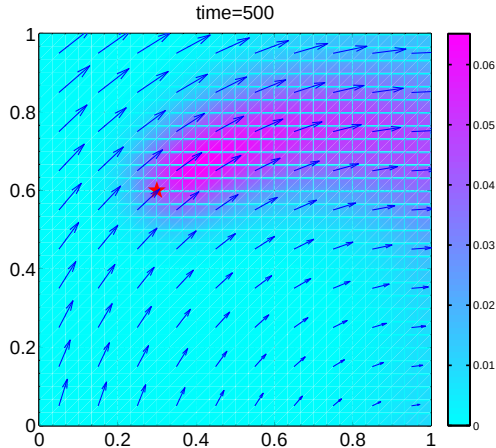
Example: Air pollution monitoring



Example: Air pollution monitoring



Example: Air pollution monitoring



Example: Air pollution

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

Advection-diffusion equation

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

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
- $\boldsymbol{\nu}(\mathbf{x}) = (\nu_1(\mathbf{x}), \nu_2(\mathbf{x}))$ – wind velocity field

Observation

For pointwise stationary sensors we have

$$z(t) = y_m(t) + \varepsilon_m(t), \quad t \in T$$

where

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

$\mathbf{x}^j \in \bar{\Omega}$, $j = 1, \dots, N$ – 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.

Towards a criterion

Spatially independent measurements

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

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

Fisher Information Matrix (FIM)

$$M = \frac{1}{\sigma^2} \sum_{j=1}^N \int_0^{t_f} \mathbf{g}(\mathbf{x}^j, t) \mathbf{g}^\top(\mathbf{x}^j, t) dt$$

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

Well, there are some drawbacks...

Unfortunately, the remainder does not reduce to numerically minimizing an alphabetical optimality criterion.

- 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

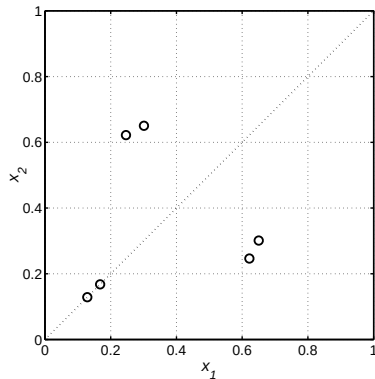
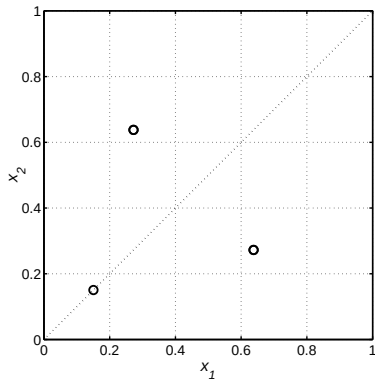
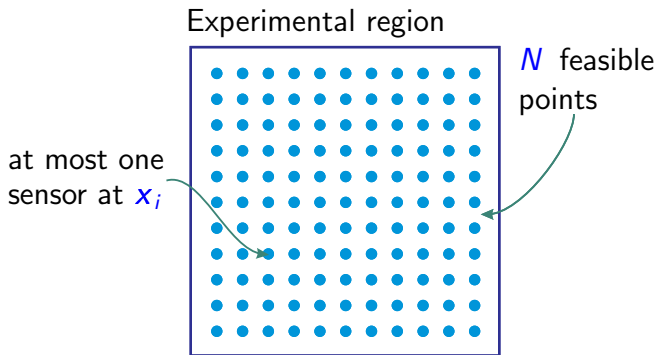


Figure: Locating six sensors through the direct use of non-linear programming techniques: (a) independent measurements, (b) correlated measurements.

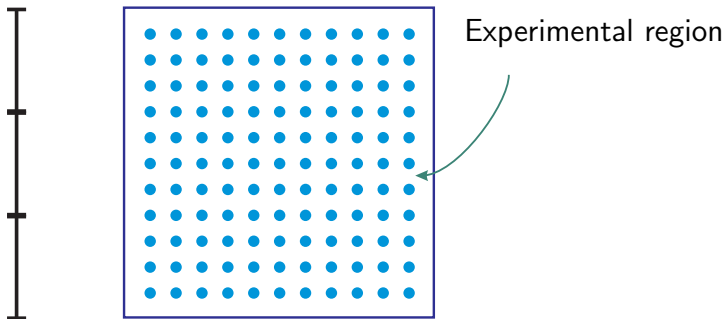
Sensor selection

Given n sensors and a finite set $\{x_1, \dots, x_N\}$ (here $N > n$) of candidate points for their location, consider an 'optimal' selection of these observation points.



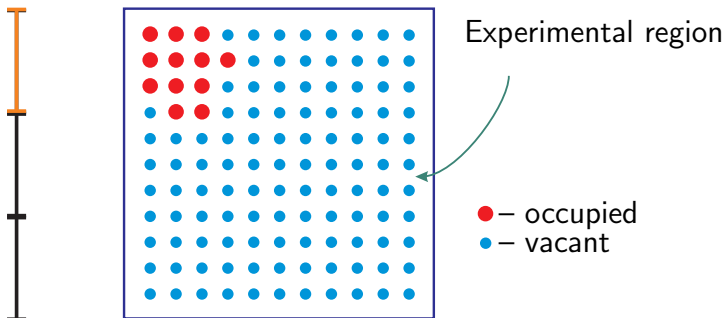
Observations

Combinatorial problem: Which n -element subset to select?



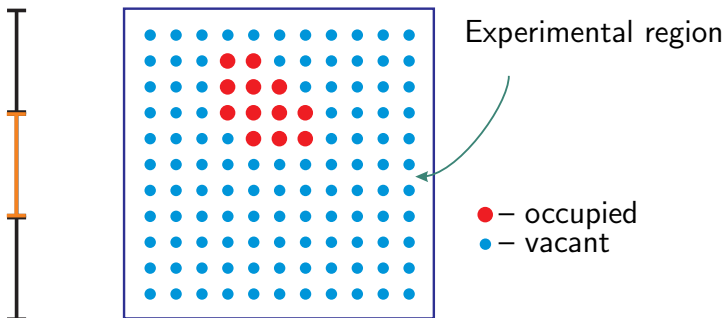
Observations

Combinatorial problem: Which n -element subset to select?



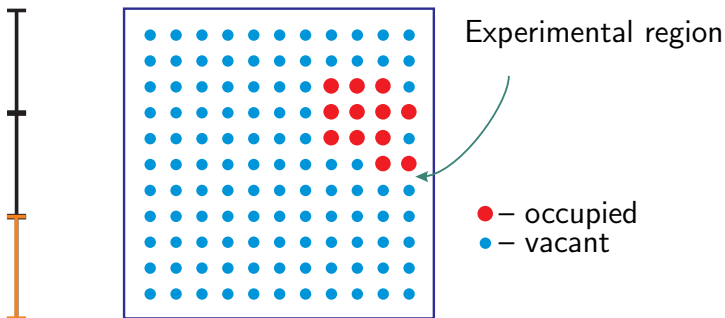
Observations

Combinatorial problem: Which n -element subset to select?



Observations

Combinatorial problem: Which n -element subset to select?



Binary decision variables

Given candidate points $\mathbf{x}_1, \dots, \mathbf{x}_N$, introduce binary flags (indicator variables) $\mathbf{v} = (v_1, \dots, v_N)$ satisfying

$$v_i = \begin{cases} 1 & \text{if a sensor is located at } \mathbf{x}_i \\ 0 & \text{otherwise} \end{cases}$$

Thus looking for a best n -element subset of N design points amounts to choosing the best element of the feasible set of binary weights

$$\mathcal{V}_{\text{bin}} = \{ \mathbf{v} \in \{0, 1\}^N : \mathbf{v}^T \mathbf{1} = n \}$$

Notation: $\mathbf{1}$ – vector of ones

Fisher Information Matrix (FIM)

Up to a constant multiplier, we have

$$M(\mathbf{v}) = M_0 + \sum_{i=1}^N v_i M_i, \quad M_i = \int_0^{t_f} \mathbf{g}(x_i, t) \mathbf{g}^T(x_i, t) dt$$

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

(Relaxed) problem of best sensor selection

Find $v_1, \dots, v_N$ so as to maximize the Φ -optimum design criterion

$$\mathcal{P}_\Phi(\mathbf{v}) = \Phi\left(M_0 + \sum_{i=1}^N v_i M_i\right)$$

subject to

$$\sum_{i=1}^N v_i = n$$
$$v_i \in \{0, 1\}, \quad i = 1, \dots, N$$

Φ is assumed to be concave, e.g., for D-optimality

$$\Phi_D(M) = \log \det(M) \quad \text{or} \quad \Phi_D(M) = \det^{1/m}(M)$$

and its maximization yields the minimal volume of the confidence ellipsoid.

(Relaxed) problem of best sensor selection

Find $v_1, \dots, v_N$ so as to maximize the Φ -optimum design criterion

$$\mathcal{P}_\Phi(\mathbf{v}) = \Phi\left(M_0 + \sum_{i=1}^N v_i M_i\right)$$

subject to

$$\sum_{i=1}^N v_i = n$$

$$v_i \in [0, 1], \quad i = 1, \dots, N$$

Φ is assumed to be concave, e.g., for D-optimality

$$\Phi_D(M) = \log \det(M) \quad \text{or} \quad \Phi_D(M) = \det^{1/m}(M)$$

and its maximization yields the minimal volume of the confidence ellipsoid.

(Relaxed) problem of best sensor selection

Find $v_1, \dots, v_N$ so as to maximize the Φ -optimum design criterion

$$\mathcal{P}_\Phi(\mathbf{v}) = \Phi\left(M_0 + \sum_{i=1}^N v_i M_i\right)$$

subject to

$$\sum_{i=1}^N v_i = n$$

$$v_i \in [0, 1], \quad i = 1, \dots, N$$

Φ is assumed to be concave, e.g., for E-optimality

$$\Phi_E(M) = \lambda_{\min}(M)$$

and its maximization yields the minimal length of the largest axis of the confidence ellipsoid.

(Relaxed) problem of best sensor selection

Find $v_1, \dots, v_N$ so as to maximize the Φ -optimum design criterion

$$\mathcal{P}_\Phi(\mathbf{v}) = \Phi\left(M_0 + \sum_{i=1}^N v_i M_i\right)$$

subject to

$$\sum_{i=1}^N v_i = n$$

$$v_i \in [0, 1], \quad i = 1, \dots, N$$

Φ is assumed to be concave, e.g., for A-optimality

$$\Phi_A(M) = (\text{trace}(M^{-1}))^{-1}$$

and its maximization yields the minimal sum of the squared lengths of the axes of the confidence ellipsoid.

Relaxed solutions $\rightarrow$ sensor configurations?

There is no guarantee that the number of positive components of $\mathbf{v}$ will not exceed the number of sensors. Therefore, **sensor locations are drawn without replacement from the set of points with positive weights in the optimal relaxed solution**, with the probabilities proportional to the values of these weights. (Repeat a pre-defined number of times and select best.)

Which optimality criterion to choose?

Different optimality criteria focus on different aspects of the FIM. As a result, they may lead to radically different sensor locations and an optimal solution for one criterion may prove poor for another criterion of interest.

Therefore, practitioners can be interested in designs representing the best compromise between a sufficiently broad class of optimum design criteria.

Orthogonally invariant information criteria

A criterion Φ is called **orthogonally invariant** if

$$\Phi(\mathbf{U} \mathbf{M} \mathbf{U}^T) = \Phi(\mathbf{M})$$

for any FIM and any orthogonal (i.e., satisfying $\mathbf{U}^T \mathbf{U} = \mathbf{I}_m$) matrix $\mathbf{U} \in \mathbb{R}^{m \times m}$. Equivalently, $\Phi(\mathbf{M})$ only depends on the eigenvalues of $\mathbf{M}$.

The class $\mathcal{O}$ of orthogonally invariant design criteria is broad enough to include all ‘alphabetical’ criteria widely used in practice (D-, A-, E-optimal, and many others).

Maximin-efficient sensor location

The Φ -suboptimality of a sensor location $\mathbf{v}$ is quantified by its efficiency:

$$\text{eff}(\mathbf{v} \mid \Phi) = \frac{\Phi(M(\mathbf{v}))}{\Phi(M(\mathbf{v}_{\Phi}^*))}$$

where $\mathbf{v}_{\Phi}^*$ is an Φ -optimum sensor location. Note that $0 \leq \text{eff}(\mathbf{v} \mid \Phi) \leq 1 = \text{eff}(\mathbf{v}_{\Phi}^* \mid \Phi)$.

Maximin-efficient sensor location

The Φ -suboptimality of a sensor location $\mathbf{v}$ is quantified by its **efficiency**:

$$\text{eff}(\mathbf{v} \mid \Phi) = \frac{\Phi(M(\mathbf{v}))}{\Phi(M(\mathbf{v}_\Phi^*))}$$

where $\mathbf{v}_\Phi^*$ is an Φ -optimum sensor location. Note that $0 \leq \text{eff}(\mathbf{v} \mid \Phi) \leq 1 = \text{eff}(\mathbf{v}_\Phi^* \mid \Phi)$.

Problem

Find $\mathbf{v}^* \in \mathcal{V}$ to maximize the $\mathbb{O}$ -maximin-efficiency design criterion

$$P(\mathbf{v}) = \inf_{\Phi \in \mathbb{O}} \text{eff}(\mathbf{v} \mid \Phi)$$

Thus, we are to maximize the worst efficiency within the class of orthogonally invariant criteria $\mathbb{O}$.

Important simplification

Define the E_k -optimality criterion

$$\Phi_{E_k}(M) = \sum_{\ell=1}^k \lambda_{\ell}(M)$$

where $\lambda_1(M) \leq \dots \leq \lambda_m(M)$ denote the eigenvalues of M in increasing order.

Important simplification

Define the E_k -optimality criterion

$$\Phi_{E_k}(M) = \sum_{\ell=1}^k \lambda_{\ell}(M)$$

where $\lambda_1(M) \leq \dots \leq \lambda_m(M)$ denote the eigenvalues of M in increasing order.

We have (Harman, 2004)

$$\inf_{\Phi \in \mathbb{O}} \text{eff}(\mathbf{v} \mid \Phi) = \min_{k=1, \dots, m} \text{eff}(\mathbf{v} \mid \Phi_{E_k})$$

Hence the minimal efficiency w.r.t. the uncountable set of all orthogonally invariant criteria is simply the minimal efficiency w.r.t. the set of criteria Φ_{E_k} which has only m elements.

Reduction to semi-infinite programming (SIP)

Using the Ky Fan Theorem, for $M \succ 0$ we get

$$\Phi_{E_k}(M) = \min_{Q \in \mathcal{Q}_k} \text{trace}(Q^T M Q)$$

where

$$\mathcal{Q}_k = \{Q \in \mathbb{R}^{m \times k} : Q^T Q = I_k\}$$

Reduction to semi-infinite programming (SIP)

Using the Ky Fan Theorem, for $M \succ 0$ we get

$$\Phi_{E_k}(M) = \min_{Q \in \mathcal{Q}_k} \text{trace}(Q^T M Q)$$

where

$$\mathcal{Q}_k = \{Q \in \mathbb{R}^{m \times k} : Q^T Q = I_k\}$$

Problem

Find $v \in \mathcal{V}$ to maximize

$$P(v) = \min_{k=1, \dots, m} \left(\frac{1}{c_k} \min_{Q \in \mathcal{Q}_k} \text{trace}(Q^T M(v) Q) \right)$$

where $c_k = \max_{v \in \mathcal{V}} \Phi_{E_k}(M(v))$.

Reduction to SIP (cont'd)

Problem

Find $(\mathbf{v}, \alpha) \in \mathcal{V} \times \mathbb{R}$ so as to maximize α subject to the constraints

$$\frac{1}{c_k} \text{trace}(\mathbf{Q}_k^T \mathbf{M}(\mathbf{v}) \mathbf{Q}_k) \geq \alpha, \quad \forall \mathbf{Q}_k \in \mathcal{Q}_k$$

for $k = 1, \dots, m$.

Reduction to SIP (cont'd)

Problem

Find $(\mathbf{v}, \alpha) \in \mathcal{V} \times \mathbb{R}$ so as to maximize α subject to the constraints

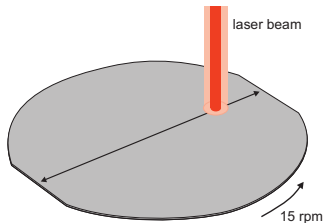
$$\frac{1}{c_k} \text{trace}(\mathbf{Q}_k^T \mathbf{M}(\mathbf{v}) \mathbf{Q}_k) \geq \alpha, \quad \forall \mathbf{Q}_k \in \mathcal{Q}_k$$

for $k = 1, \dots, m$.

This problem is solved using the method of outer approximations. It alternates between computing the eigenvalues and their associated orthonormal eigenvectors, and solving a linear-programming problems with respect to the design weights $\mathbf{v}$.

Simulation example

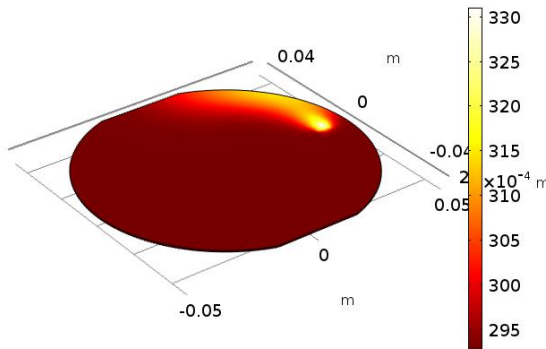
Consider the wafer staging process in semiconductor manufacturing. A circular silicon wafer is heated up by a laser that moves radially in and out over time while the wafer itself rotates on its stage. The wafer is 1 mm thick and has a diameter of 10 cm. It is trimmed by 2.5 mm on the left and right sides for mounting in stage.



The laser operates on its upper surface for a time of $t_f = 30$ seconds. The silicon disk rotates counterclockwise with a constant rotational speed of 15 rpm.

Simulation example

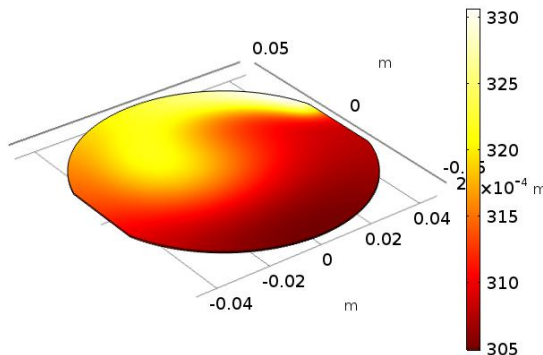
The thermal process in the wafer is governed by a convection-diffusion PDE. Here is the temperature evolution:



$t = 1$ s

Simulation example

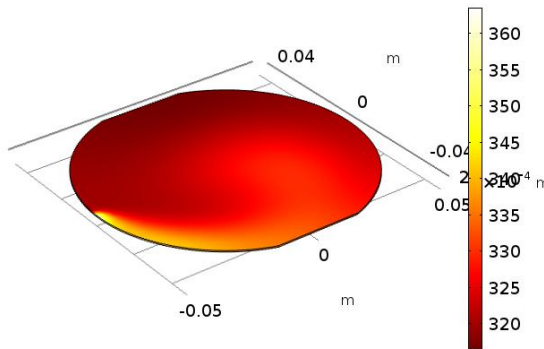
The thermal process in the wafer is governed by a convection-diffusion PDE. Here is the temperature evolution:



$t = 10$ s

Simulation example

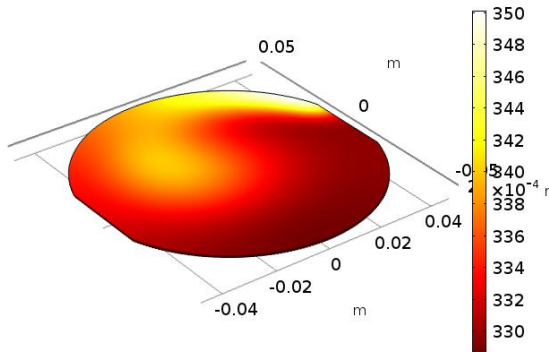
The thermal process in the wafer is governed by a convection-diffusion PDE. Here is the temperature evolution:



$t = 15$ s

Simulation example

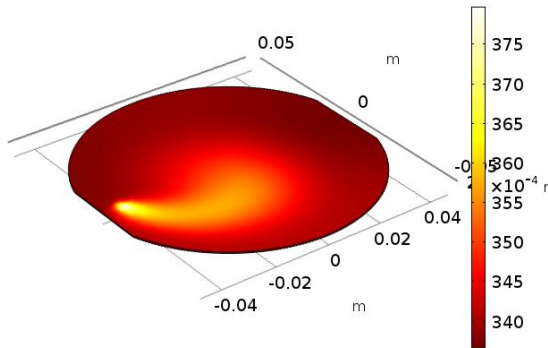
The thermal process in the wafer is governed by a convection-diffusion PDE. Here is the temperature evolution:



$$t = 20 \text{ s}$$

Simulation example

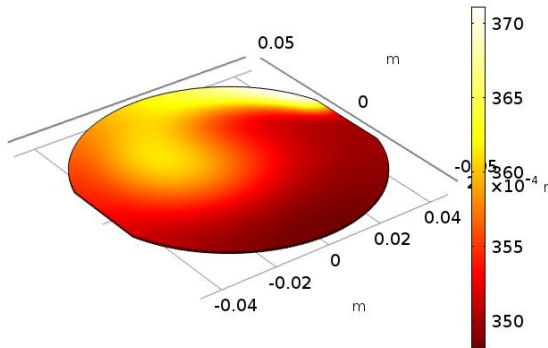
The thermal process in the wafer is governed by a convection-diffusion PDE. Here is the temperature evolution:



$t = 24$ s

Simulation example

The thermal process in the wafer is governed by a convection-diffusion PDE. Here is the temperature evolution:



$t = 30$ s

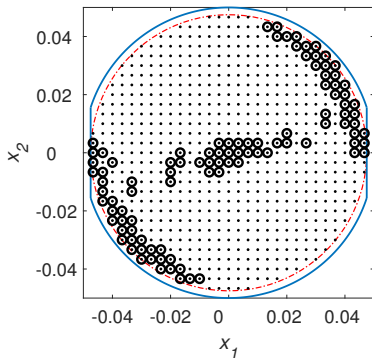
Sensor selection problem

The temperature can be measured at the lower surface using infrared detectors placed at sites picked up from a total of 641 candidate sites. Once placed, each detector take measurements at time instants $t_k = 0.5k$, $k = 1, \dots, 60$.

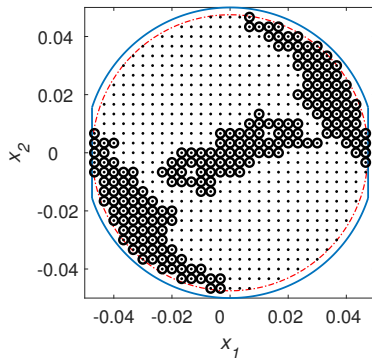
The objective is to determine a maximin-efficient subset of n gaged sites for estimaton of four scalar parameters:

- thermal conductivity (PDE)
- surface emissivity (BC)
- conservative flux source term (PDE)
- volumetric thermal capacity (PDE)

Maximix-efficient locations

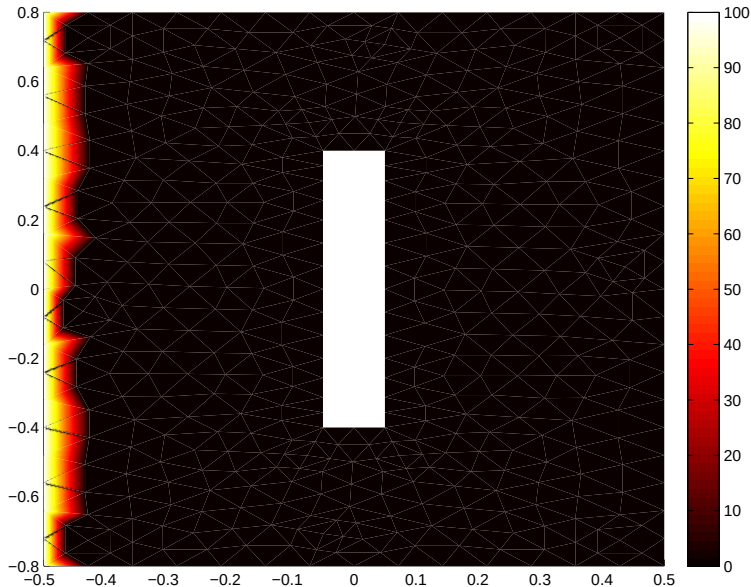


(a) $n = 100$, maximin eff. 0.8802

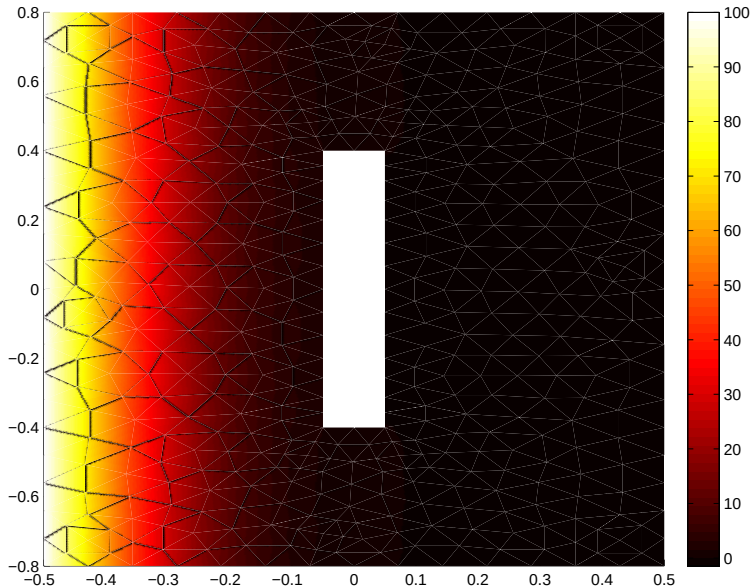


(b) $n = 200$, maximin eff. 0.8853

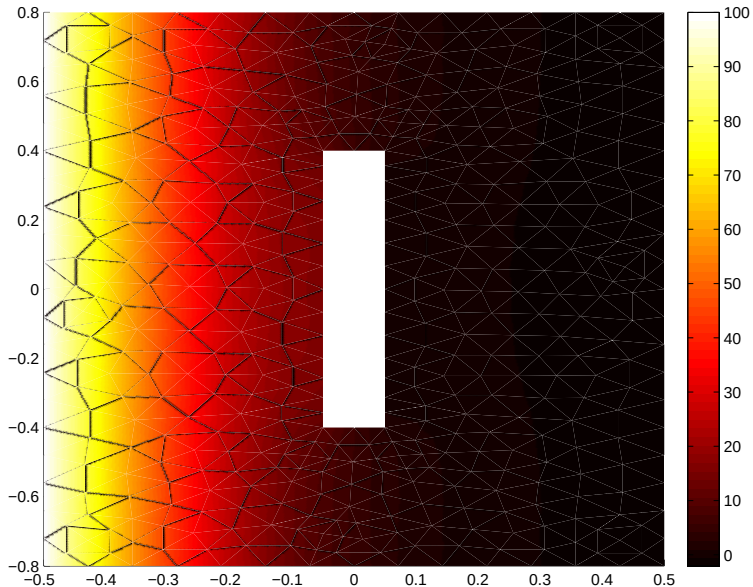
Heating a metal block with a crack – ctd'



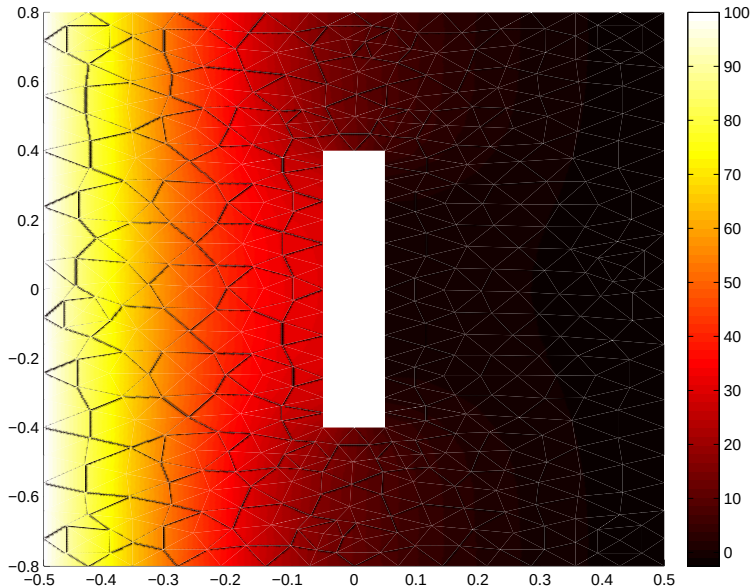
Heating a metal block with a crack – ctd'



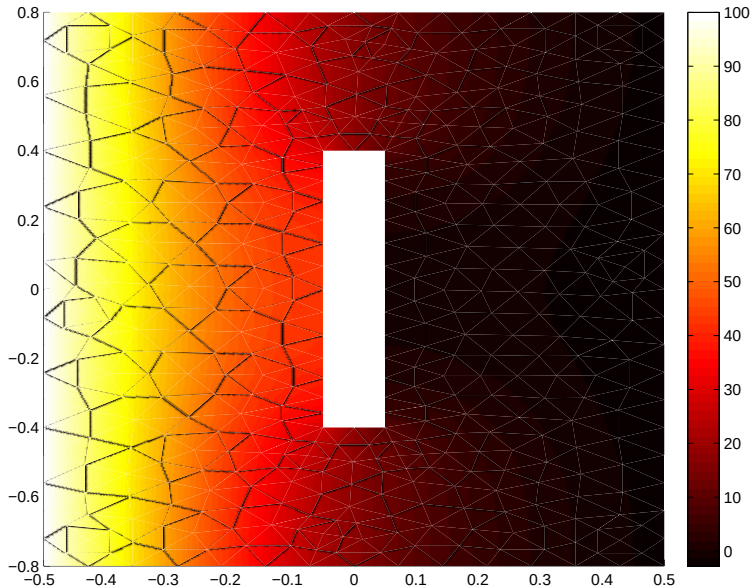
Heating a metal block with a crack – ctd'



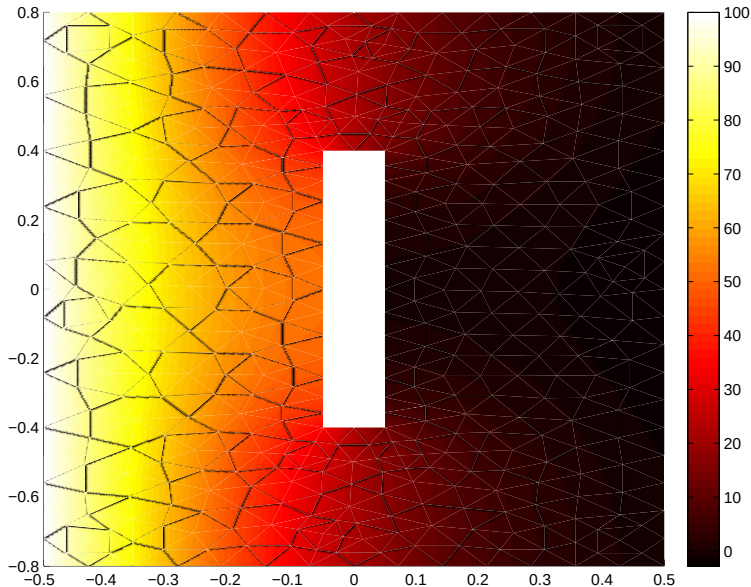
Heating a metal block with a crack – ctd'



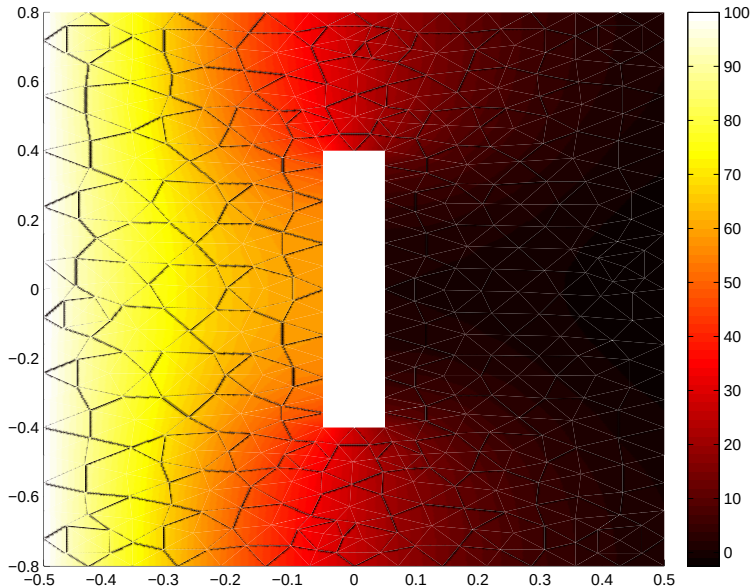
Heating a metal block with a crack – ctd'



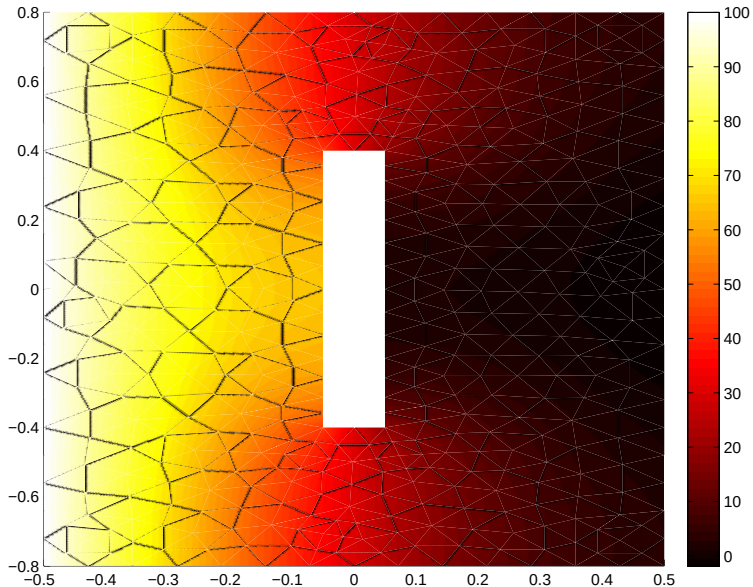
Heating a metal block with a crack – ctd'



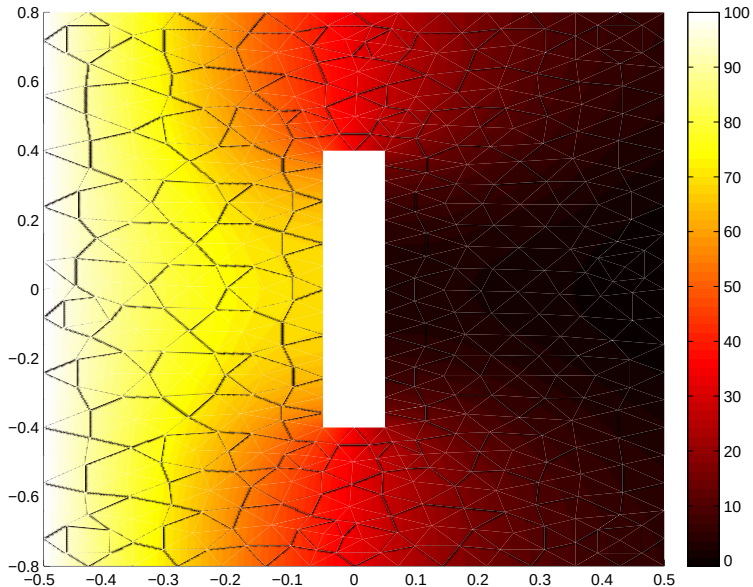
Heating a metal block with a crack – ctd'



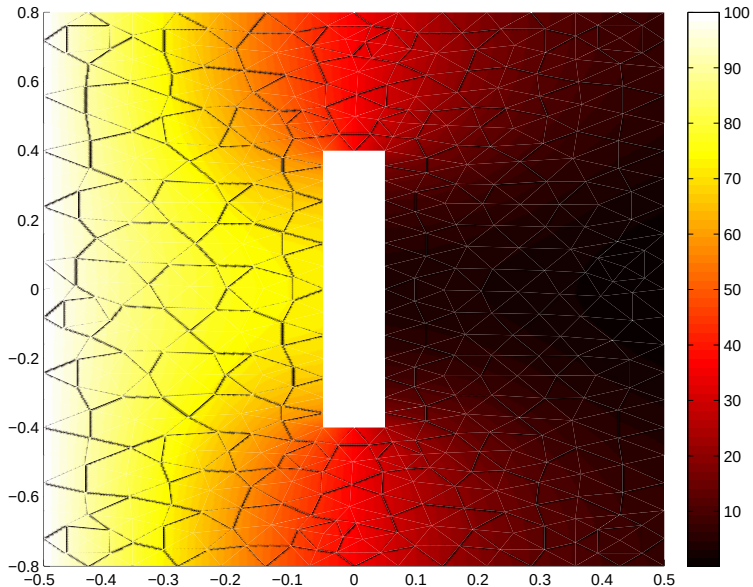
Heating a metal block with a crack – ctd'



Heating a metal block with a crack – ctd'



Heating a metal block with a crack – ctd'



Example ctd': Mathematical model

Let $T(\chi, t)$ be the temperature at spatial point χ and time t . Its evolution is governed by the **partial differential equation (PDE)**

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

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

$$T = T_0 \quad \text{in } \Omega \text{ at } t = 0$$

and the **boundary conditions**

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

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

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

Example ctd': Calibration and measurement design

If thermal conductivity α and initial condition T_0 are unknown, which is rather typical, we could observe the temperature evolution at a fixed point and tune α and T_0 so that our model fits the data as best as it can.

Example ctd': Calibration and measurement design

If thermal conductivity α and initial condition T_0 are unknown, which is rather typical, we could observe the temperature evolution at a fixed point and tune α and T_0 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 α and T_0 ?

Example ctd': Calibration and measurement design

If thermal conductivity α and initial condition T_0 are unknown, which is rather typical, we could observe the temperature evolution at a fixed point and tune α and T_0 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 α and T_0 ?

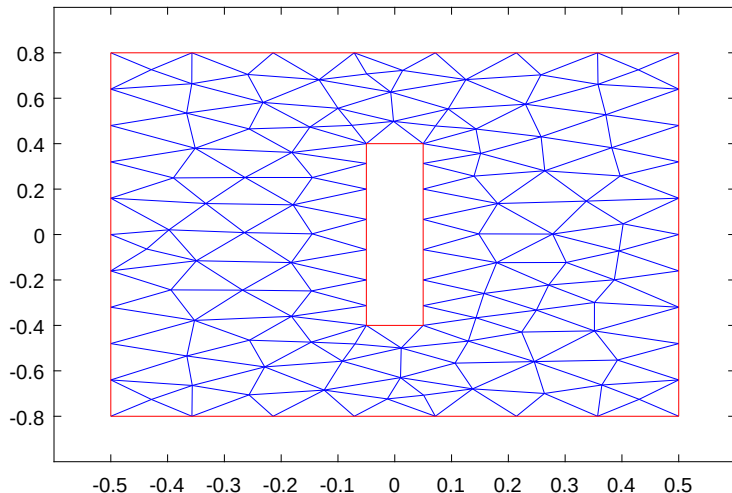
Crucial difficulty: In the previous example, the component T_0 of $\theta = (\alpha, T_0)$ is an element of an infinite dimensional functional space (here $L^2(\Omega)$)

Semi-discretization (method of lines)

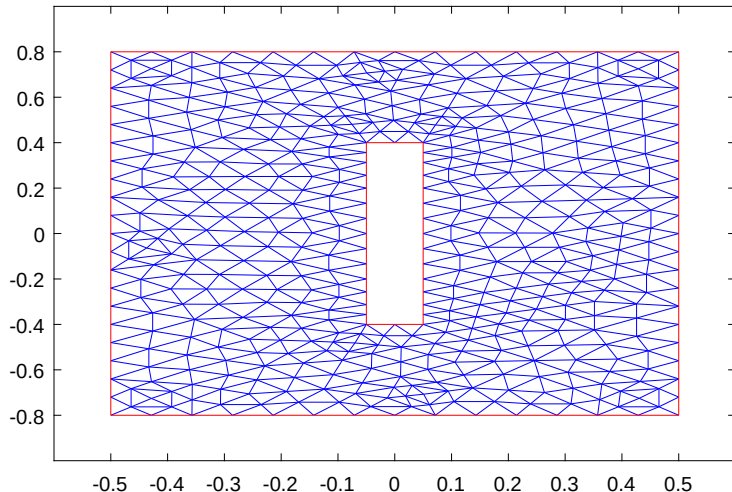
The **finite-element method (FEM)** is a technique for solving PDEs by first discretizing these equations in their space dimensions. The discretization is carried out locally over small regions of simple but arbitrary shape (the **finite elements**). Each vertex is called a **node**.

The solution should be simple on each element. Piecewise linear functions are a good choice.

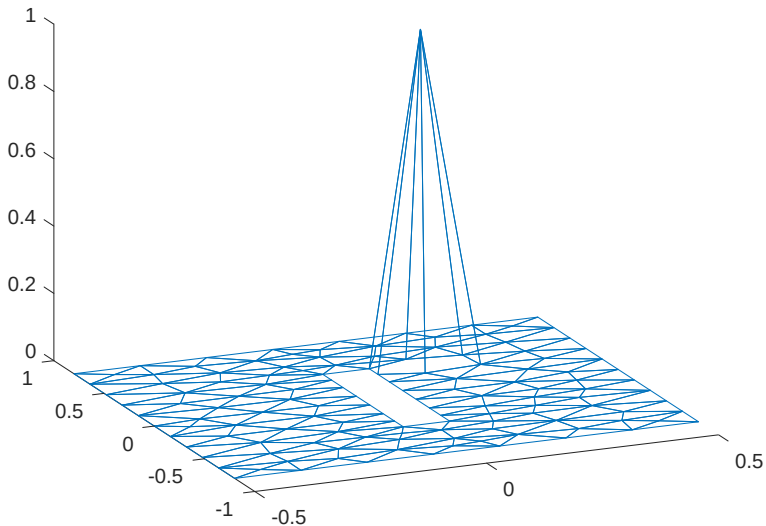
Triangulation of the spatial domain



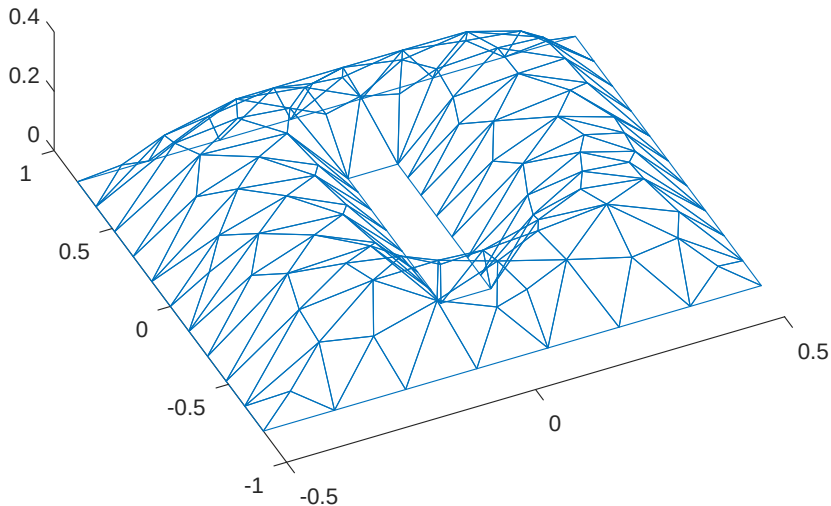
Triangulation of the spatial domain



Basis functions



Piecewise linear function



Resulting approximations

Triangular finite elements and piecewise linear continuous Lagrange basis functions $\{\varphi_j\}_{j=1}^n$ are used here. The basis functions correspond to nodal points $\{\chi_j\}_{j=1}^n$ such that

$$\varphi_j(\chi_i) = \delta_{ij} \quad \text{for } i, j \in \{1, \dots, n\}$$

We have approximations

$$T(\chi, t) \approx \sum_{j=1}^n x_j(t) \varphi_j(\chi), \quad T_0(\chi) \approx \sum_{j=1}^n x_{0,j} \varphi_j(\chi)$$

Bayesian setting

Assume that we are given a Gaussian prior distribution with mean θ_{prior} and covariance matrix Γ_{prior} . Using Bayes' formula, the maximum *a posteriori* (MAP) estimate θ_{MAP} of the posterior measure can be determined.

Applying a linearized Bayesian formulation, we can get a Gaussian approximation to the posterior distribution, for which the posterior covariance matrix $\Gamma_{\text{post}} \approx M^{-1}$ (the expression for M is quite similar to what has been presented before).

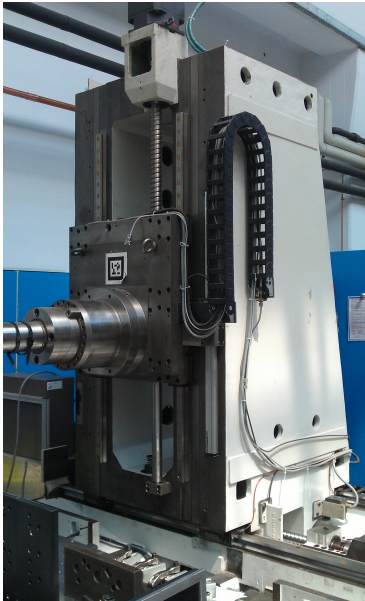
Digression: Randomized trace estimator

Randomized Gaussian trace estimator:

$$\text{trace}(\mathbf{\Gamma}_{\text{post}}) = \frac{1}{\ell} \sum_{i=1}^{\ell} z_i \underbrace{M^{-1} z_i}_{=q_i}$$

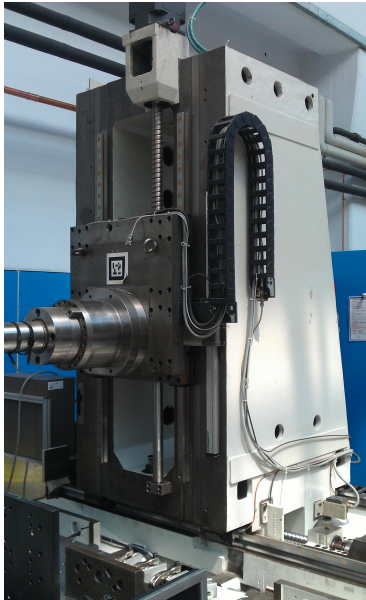
where the z_i are ℓ independent random vectors whose entries are i.i.d. standard normal variables. (q_i is evaluated by solving $Mq_i = z_i$.)

Example: Deformation of machine tools



The prototype machine column Auerbach ACW 360 **deforms** during its operation **due to waste heat from two external electrical drives** (the one at the top moves the sledge holding the main spindle and the one at the bottom moves the entire machine column).

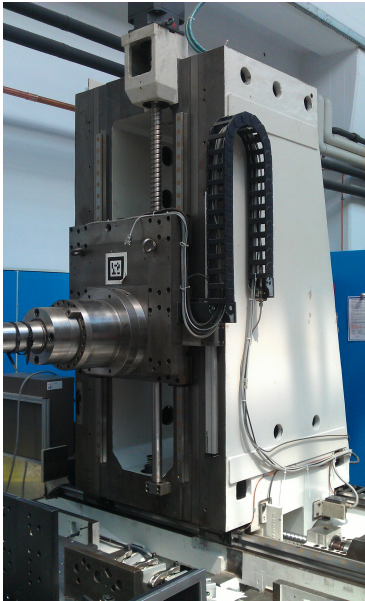
Example: Deformation of machine tools



The inhomogeneous transient temperature field displaces the **tool centre point (TCP)** and thus reduces production accuracy and product quality.

Key idea: Predict the machine tool deformation based on measurements from temperature sensors and then correct it on line.

Example: Deformation of machine tools



Key problem: Temperature sensors may be placed only on the surface of the machine column. But which locations are best?

(joint work with Roland Herzog from Heidelberg University)

Thermo-mechanical system

Heat equation

$$\begin{aligned}\rho c_p \dot{T} - \operatorname{div}(\lambda \nabla T) &= 0 && \text{in } \Omega \times (0, t_f) \\ \lambda \frac{\partial T}{\partial n} + \alpha(\chi) (T - T_{\text{ref}}) &= r(\chi, t) && \text{on } \partial\Omega \times (0, t_f) \\ T(\chi, 0) &= T_0(\chi) && \text{in } \Omega\end{aligned}$$

T	temperature
r	thermal surface load
ρ	density
c_p	specific heat at constant pressure
λ	thermal conductivity
α	coefficient of heat transfer
T_{ref}	ambient temperature
T_0	initial temperature

Thermo-mechanical system

Linear elasticity (BCs omitted)

$$-\operatorname{div} \sigma(\varepsilon(\mathbf{u}), \mathbf{T}(t_f)) = \mathbf{0} \quad \text{in } \Omega$$

$$\sigma(\varepsilon(\mathbf{u}), \mathbf{T}(t_f)) = \sigma^{\text{el}}(\varepsilon(\mathbf{u})) + \sigma^{\text{th}}(\mathbf{T}(t_f))$$

$$\sigma^{\text{el}}(\varepsilon(\mathbf{u})) = \frac{E}{1+\nu} \varepsilon(\mathbf{u}) + \frac{E\nu}{(1+\nu)(1-2\nu)} \operatorname{trace}(\varepsilon(\mathbf{u})) \operatorname{id}$$

$$\sigma^{\text{th}}(\mathbf{T}(t_f)) = -\frac{E}{1-2\nu} \beta (\mathbf{T}(t_f) - T_{\text{ref}}) \operatorname{id}_3$$

$$\varepsilon(\mathbf{u}) = \frac{1}{2}(\nabla \mathbf{u} + \nabla \mathbf{u}^\top)$$

$\mathbf{u}$ displacement

σ stress

ε strain

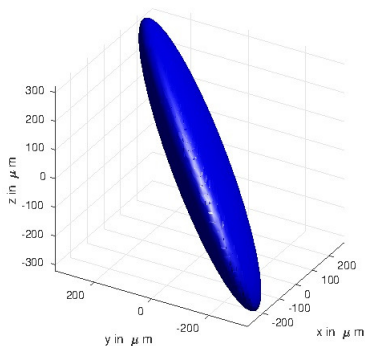
ν, E Poisson's ratio and modulus of elasticity, resp.

β thermal volumetric expansion coefficient

Results: Optimal positions of 10 sensors

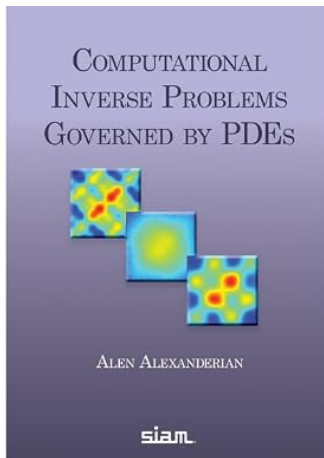


Covariance matrix of the QOI estimator



More details in R. Herzog, I. Riedel, D. Uciński: Optimal sensor placement for joint parameter and state estimation problems in large-scale dynamical systems with applications to thermo-mechanics, *Optimization and Engineering* (2018) 19:591–627

Recent monograph



SIAM, 2025

Persisting open problems

- Design for correlated errors
- Nonparametric models
- Active learning
- Development of computationally efficient algorithms for the Bayesian framework
- Joint design of sensor location and inputs
- Design for networks
- Design for digital twins

- D. Uciński and M. Patan, "Maximin Efficient Sensor Location for Parameter Estimation of Spatiotemporal Systems," *60th IEEE Conference on Decision and Control (CDC)*, Austin, TX, USA, 2021, pp. 2601–2606.
- R. Harman, "Minimal efficiency of designs under the class of orthogonally invariant information criteria," *Metrika*, vol. 60, no. 2, pp. 137–153, 2004.
- D. Uciński, D., "Sensor selection with nonsmooth design criteria based on semi-infinite programming." *IFAC-PapersOnLine*, vol. 53, no. 2, pp. 7539–7544, 2020.