Published in final edited form as:

Int Stat Rev. 2018 August ; 86(2): 169–188. doi:10.1111/insr.12233.

Optimal Adaptive Designs with Inverse Ordinary Differential Equations

Eugene Demidenko

Department of Biomedical Data Science, Dartmouth College, Hanover, NH 03755, USA;

Department of Mathematics, Dartmouth College, Hanover, NH 03755, USA

Summary

Many industrial and engineering applications are built on the basis of differential equations. In some cases, parameters of these equations are not known and are estimated from measurements leading to an inverse problem. Unlike many other papers, we suggest to construct new designs in the adaptive fashion ‘on the go’ using the A-optimality criterion. This approach is demonstrated on determination of optimal locations of measurements and temperature sensors in several engineering applications: (1) determination of the optimal location to measure the height of a hanging wire in order to estimate the sagging parameter with minimum variance (toy example), (2) adaptive determination of optimal locations of temperature sensors in a one-dimensional inverse heat transfer problem and (3) adaptive design in the framework of a one-dimensional diffusion problem when the solution is found numerically using the finite difference approach. In all these problems, statistical criteria for parameter identification and optimal design of experiments are applied. Statistical simulations confirm that estimates derived from the adaptive optimal design converge to the true parameter values with minimum sum of variances when the number of measurements increases. We deliberately chose technically uncomplicated industrial problems to transparently introduce principal ideas of statistical adaptive design.

Keywords

Diffusion equation; heat equation; non-linear least squares; optimal design of experiments; inverse problem

1 Introduction and Motivation

Many industrial and engineering designs are reduced to inverse differential equations or more specifically to boundary value problems. For example, several books, including Alifanov (1994), Beck & Arnold (1977) and Ozisik & Orlande (2000), discuss inverse problems for the heat transfer and diffusion–conduction models. Typically, non-linear least squares is used to identify parameters (2005). A few authors go beyond parameter identification and apply further statistical inference to parameter estimation including confidence intervals and hypothesis testing. Only a handful of papers discuss optimal design when the parameters of the differential equations are not known, Pronzato (2008) and

Glaholt *et al.* (2012); see Franceschini & Macchietto (2008) for a review. This discussion leaves a practically important question unanswered: where the next location of the measurement or sensor should be, or where should the additional measurements be taken to estimate unknown parameters of differential equations with maximum efficiency? We adopt the statistical theory of optimal experiment for non-linear parameter estimation to answer the location question.

Several excellent books on optimal design of experiments can be recommended, including Fedorov (1972), Pazman (1986), Silvey (1980), Pukelsheim (2006) and Atkinson & Donev (1992). In particular, the book by Ucinski (2000) discusses optimal engineering designs in detail. Several criteria for the optimal design have been developed. However, we advocate the A-optimality design criterion over a more popular D-optimality design criterion introduced and studied in seminal papers by Box & Lucas (1959) and Kiefer (1974). In both cases, the inverse of the Fisher information matrix, $\mathbf{C}$; is involved: in the A-optimality, the trace of $\mathbf{C}$ is minimised, and in the D-optimality, the determinant $|\mathbf{C}|$, sometimes called the generalised variance, is minimised. We argue that the A-optimality has more appeal from the application point of view:

- All optimal estimation results, such as the Gauss–Markov theorem are expressed in terms of the variance, not generalised variance. In a linear model, it is proved that the optimal estimate of a linear combination of parameters $\mathbf{a}'\boldsymbol{\beta}$ is $\mathbf{a}'\hat{\boldsymbol{\beta}}$ with the minimum variance $\mathbf{a}'\mathbf{C}\mathbf{a}$, not the generalised variance.
- Many goodness-of-fit measures, such as the coefficient of determination, are also expressed in terms of variances, not the generalised variance.
- When the model is used for prediction, the variance of the prediction to be minimised is the quadratic form with matrix $\mathbf{C}$.
- An often quoted invariance of the D-optimality criterion to rescaling and linear transformation in general, sometimes used to justify this criterion, holds for linear models and only occasionally for non-linear models. In particular, this property does hold for the models discussed in the present paper.
- For models with one parameter, the optimal designs using the A-optimality and D-optimality designs are the same.

Besides the choice of the optimality criterion, another problem to solve is how to break the circle: the optimal design depends on the true parameter, but the parameter value is unknown. The most popular approaches, as recently discussed by Duarte & Wong (2015) on the pages of this journal, are as follows: (1) obtain the values from previous studies and use them for optimal design, (2) use minimax or maximin approach by building design based on the worst-case scenario, (3) use Bayesian approach with a priori distribution on the parameter values. Obviously, all these approaches have shortcomings. For example, (1) and (3) require information that may be just not available. In contrast, we utilise the idea of the adaptive design when the design is re-evaluated ‘on the go’ based the current parameter value during the iterative process of parameter estimation. The adaptive design is popular in medical and pharmaceutical applications (Pong & Chow, 2011) but just briefly mentioned by

Atkinson & Donev (1992) and in a recent book by Fedorov & Leonov (2013) in the framework of the D-optimality criterion.

The goal of the present work is to promote the adaptive design with the A-optimality for non-linear problems with engineering and industrial applications. Most of the problems discussed in this paper pose the following question: where should the next measurement be made to make the total parameter variance as small as possible? The heat transfer or diffusion–conduction equation is chosen to illustrate the optimal design solution. Both situations are considered: when the ordinary differential equation (ODE) admits a closed-form solution and when the solution is found numerically via the finite difference approximation.

We use transparent and specific industry examples rather than general theory, and we believe that this is the shortest way to popularise statistical concepts of experiment design among applied statisticians and researchers working in diverse engineering and industry communities.

2 Toy Example: A Hanging Wire

In this section, we provide an illustrative (toy) example on how the statistical theory of optimal experiments can be used in the framework of inverse differential equations for parameter identification. We intentionally chose this simplistic inverse engineering problem to emphasise the statistical approach to the optimal design of experiments.

Two factors contribute to the degree of sagging: the weight and the elasticity of the wire or chain. There are two questions: Is it possible to identify these two parameters by measuring the distance from the wire to the ground? Where should these measurements be taken to obtain the maximum precision for these parameters?

It is well known that the shape of a flexible wire with fixed ends is governed by a non-linear differential equation of the second order,

$$y'' = \beta\sqrt{1 + y'^2}, \quad (1)$$

where $\beta = w/h$ is called the sagging parameter, h is the elasticity coefficient and w is the wire weight per unit length. Without loss of generality, we assume $-1 \leq x \leq 1$, so that the distance between the poles is 2 and the poles have height 1. See Figure 1 for an illustration. It is well known that the solution can be expressed in closed form as

$$y(x; \beta) = \frac{e^{-x\beta} + e^{x\beta}}{e^{-\beta} + e^{\beta}},$$

a symmetric function around $x = 0$.

Sometimes this solution is called a *catenary* (Latin for ‘chain’). The function that governs the shape of a hanging flexible wire was on the mind of physicists and mathematicians for

centuries. For example, Galileo claimed that the wire's curve is parabolic. Many famous mathematicians contributed to the correct formula as we know it today, including Leibniz, Huygens and Bernoulli (Wells, 1991). Note that we write this solution as a function of x given parameter β . Using elementary calculus or the symmetry argument, one can see that the maximum sag is in the middle, $x = 0$. We want to use the catenary solution to identify β by measuring the sag at some x_* . What should x_* be to estimate β as precisely as possible? Intuition says, 'In the middle'. We shall prove this mathematically. In order to answer the question, we need to specify the measure of accuracy of the beta-parameter estimation; for this reason, we employ the theory of non-linear regression (Seber & Wild, 1989) and optimal design of experiments (Atkinson & Donev, 1992). If $m(x_*)$ is the measured distance of the wire from the ground at point x_* , we can write

$$m(x_*) = y(x_*; \beta) + \varepsilon,$$

where ε is a random error with zero mean and variance σ^2 representing the measurement error. It is customary to use least squares to estimate parameters in non-linear regression; with a single measurement, we solve the equation $m(x_*) = y(x_*; \beta)$ for β given x_* . From the theory of non-linear regression, we know that the variance of the LS estimate $\hat{\beta}$ is approximated as $\text{var}(\hat{\beta}) = \sigma^2 / \|dy/d\beta\|^2$. Consequently, to make our measurement the most efficient, we have to find x_* that maximises the squared derivative of the catenary, $\|dy/d\beta\|^2$. Again, using elementary calculus or symmetry, one can show that this maximum is attained at $x = 0$. (The squared derivative takes maximum at $x = 0$ in Figure 1.) The maximum precision is gained when the measurement is taken in the middle of the wire.

Once the decision on the location of the height measurement is made ($x_* = 0$), we find an estimate of β as

$$\hat{\beta} = \ln \frac{1 + \sqrt{1 - m_0^2}}{m_0},$$

where m_0 is the height of the rope in the middle (maximum sag). We use the hat notation to indicate that $\hat{\beta}$ is an estimate of unknown β as is customary in statistics.

The next question arises: Where should the height of the wire be measured if several measurements can be made? It is easy to see that all measurements should be made in the centre, and then, m_0 is the average of these measurements.

In this example, the answer where to make measurements does not depend on the value of the unknown parameter. In most applications, this is not the case, and one has to adjust the location of the next measurement once the parameter has been estimated using the previous set of measurements. This leads us to the concept of adaptive design. To illustrate, we consider the same hanging wire problem but with different heights, say, H_1 and H_2 ($H_1 < H_2$). The general solution to the ODE (1) is

$$y(x; \beta) = \frac{1}{2\beta} \left(e^{\beta x + C_1 \beta} + e^{-\beta x - C_1 \beta} \right) + C_2,$$

where constants C_1 and C_2 are the solutions to the equations $y(-1) = H_1$ and $y(1) = H_2$. After some algebra, one can show that the solution to the hanging wire problem can be written as

$$y(x; \beta) = \frac{1}{2\beta} \left(\rho e^{\beta x} + \frac{1}{\rho} e^{-\beta x} \right) + \frac{-e^{4\beta} + 2\rho H_1 \beta e^{3\beta} + 1 - 2H_2 \beta \rho e^{\beta}}{2e^{\beta}(e^{2\beta} - 1)\beta \rho},$$

where $\rho > 0$ is the solution to the quadratic equation $(e^{2\beta} - 1)\rho^2 + 2\beta e^{\beta}(H_1 - H_2)\rho + 1 - e^{2\beta} = 0$. It is easy to see that this quadratic equation has a unique positive solution.

When the poles had the same heights, the optimal location of the measurement was in the middle regardless of β . Now we face a new problem: in order to find the optimal location, we need to know β . The idea of adaptive optimal design is to measure the height of the chain at some location, say, $x = 1/2$, obtain a preliminary estimate $\hat{\beta}_1$, and then to find the next location that minimises the squared derivative of y evaluated at $\beta = \hat{\beta}_1$, measure the height again, and so on. The reader can learn more about adaptive design in the next section.

3 The Inverse Heat Transfer Problem

In this section, we consider an example of an inverse boundary value problem with two unknown parameters, the heat conductivity coefficient and the ambient (surrounding) temperature. Much work has been carried out on the inverse heat transfer problem; see books by Beck & Arnold (1977) and Ozisik & Orlande (2000). Here, we consider a simple steady-state inverse problem. Several situations may be studied: when the conductivity coefficient is unknown and estimated from additional temperature measurements, when the temperature at the ends is to be estimated, when the conductivity and the ambient temperature are unknown and so on. We will consider the latter case—the study of other cases follows the same line.

The temperatures at the left and the right ends of a homogeneous rod of unit length are maintained at T_0 and T_1 , respectively. If T_a is the ambient temperature and α is the heat conductivity coefficient, the one-dimensional heat transfer differential equation in the steady state takes the form (Hoffman, 1992, p. 436)

$$T'' - \alpha^2 T = -\alpha^2 T_a, \quad (2)$$

where $T(x)$ is the temperature of the rod at x , $0 \leq x \leq 1$, $T(0) = T_0$ and $T(1) = T_1$. It is elementary to show that the solution can be expressed in closed form as

$$T(x; \alpha, T_a) = Ae^{\alpha x} + Be^{-\alpha x} + T_a, \quad (3)$$

where constants A and B are functions of the diffusion coefficient, α :

$$A = \frac{(T_1 - T_a) - (T_0 - T_a)e^{-\alpha}}{e^{\alpha} - e^{-\alpha}}, \quad B = \frac{(T_0 - T_a)e^{\alpha} - (T_1 - T_a)}{e^{\alpha} - e^{-\alpha}}.$$

Neither α nor T_a is known. They are to be estimated from at least two measurements at sensor locations $\{x_i, i = 1, 2, \dots, n\}$, where $n \geq 2$. Without loss of generality, we can assume that $x_i < x_{i+1}$. If y_i is the temperature measurement at location x_i and ε_i is the error, the statistical model takes the form of a non-linear regression,

$$y_i = Ae^{\alpha x_i} + Be^{-\alpha x_i} + T_a + \varepsilon_i, \quad (4)$$

with α and T_a estimated by minimising the non-linear least squares criterion:

$$\sum_{i=1}^n (y_i - Ae^{\alpha x_i} - Be^{-\alpha x_i} - T_a)^2.$$

The error term, ε_i , represents the temperature measurement error and/or the model mismatch. It is assumed that $\{\varepsilon_i\}$ are independently and identically distributed with zero mean and constant variance, σ^2 , and x_i contains no error. Instalment of temperature sensors is usually expensive; therefore, it is important to find their optimal locations to reduce the cost of installation—this problem is solved in the next section. In this paper, we assume that the deviation of the temperature measurements from the model (error terms) do not correlate and have constant variance (homoscedastic). These assumptions should be verified after model fitting although incorporation correlation and heteroscedasticity would require more assumptions and additional estimation of more parameters that may even worsen the estimates.

3.1 Where to Put Temperature Sensors?

In this subsection, we assume that the parameters of the model, or at least their estimates, are known. We show how to estimate these parameters along with determination of the sensor locations in the next subsection. To find the optimal sensor locations, we need to know the Jacobian of the heat transfer solution (3) as a matrix of partial derivatives of the temperature of the rod with respect to the unknown parameters. Let us assume here that values of α and T_a are known and we want to find the optimal locations of two sensors, x_1 and x_2 , where measurements y_1 and y_2 are to be made. In practical designs, temperature sensors are usually located at equal distance from each other. Because we have to have at least two locations to estimate the model, an obvious starting location is at $x_1 = 1/3$ and $x_2 = 2/3$ on the rod of unit

length. We call these the naive sensor locations. We show that this is a poor choice even when the temperatures at the ends are the same.

An important question is what criterion of optimality to use when finding the optimal location—we suggest using a statistical criterion that leads to the minimal total variance of estimated parameters computed using the Fisher information matrix. We refer the reader to Emery & Fadale (1996), Fadale *et al.* (1995), Emery *et al.* (2000) and Li *et al.* (2008) for a general discussion of the criteria in the context of the sensor location problem. In the case of two measurements, the Jacobian takes the form

$$\mathbf{J} = \begin{bmatrix} \left. \frac{\partial T}{\partial \alpha} \right|_{x=x_1} & \left. \frac{\partial T}{\partial T_a} \right|_{x=x_1} \\ \left. \frac{\partial T}{\partial \alpha} \right|_{x=x_2} & \left. \frac{\partial T}{\partial T_a} \right|_{x=x_2} \end{bmatrix},$$

where the partial derivative $\partial T / \partial \alpha$ and $\partial T / \partial T_a$ are straightforward to obtain from (3). From the theory of non-linear regression, it is known that the variance–covariance matrix of the unknown parameters α and T_a can be approximated by the inverse of the Fisher information matrix,

$$\text{cov} = \sigma^2(\mathbf{J}'\mathbf{J})^{-1}. \quad (5)$$

The variances for α and T_a are on the diagonal, $\text{var}(\hat{\alpha}) = \sigma^2(\mathbf{J}'\mathbf{J})_{11}^{-1}$ and $\text{var}(\hat{T}_a) = \sigma^2(\mathbf{J}'\mathbf{J})_{22}^{-1}$, where the sign ' on a vector or matrix (boldface), throughout the paper, means transposition. Finally, we find locations x_1 and x_2 where, according to the A-optimal design, the sum of these variances (total variance) is minimum:

$$\text{mintr}(\mathbf{J}'\mathbf{J})^{-1}, \quad (6)$$

as discussed in the literature on statistical optimal design theory. We illustrate the determination of the optimal sensor locations using two scenarios: the temperatures at the ends of the rod are the same ($T_0 = T_1$) and different ($T_0 \neq T_1$).

In Figure 2, we show the temperature distribution in the rod where the same temperature, 10 °C is fixed at both ends with $\alpha = 4 \text{ cm}^{-1}$ and $T_a = -5 \text{ °C}$. Obviously, the temperature is symmetric around $x = 0.5$. It is a matter of elementary calculus to show that the naive locations, $x_1 = 1/3$ and $x_2 = 2/3$, are poor choices because they lead to infinite variances. (The Jacobian is singular.) In fact, this is true for any α and T_a when $T_0 = T_1$ because of symmetry. The optimal sensor locations are $x_1 = 0.5$ and $x_2 = 0.88$ (or $x_1 = 0.88$ and $x_2 = 0.5$), which solve the optimisation problem (6).

In Figure 3, we show the temperature distribution when the temperatures at the ends are different, $T_0 = 0$ and $T_1 = 10$. In this case, $\text{tr}(\mathbf{J}'\mathbf{J})^{-1}$ reaches minimum when sensors are located at $x_1 = 0.34$ and $x_2 = 0.87$.

The contours of the sum of the variances as a function of the two locations are shown in Figure 4, again assuming that α and T_a are known and fixed. For the symmetric temperature distribution ($T_0 = T_1$, left panel), the optimal locations are $x_1 = 0.12$ and $x_2 = 0.5$, or equivalently $x_1 = 0.5$ and $x_2 = 0.88$. In the right panel, we show the asymmetric case ($T_0 \neq T_1$) with unique optimal sensor locations $x_1 = 0.34$ and $x_2 = 0.87$.

3.2 Adaptive Sensor Design

In the previous subsection, it was assumed that a priori values of α and T_a were available. In practice, we may not have them. Because temperature measurements are typically expensive, we want to make as few measurements as possible while achieving the desired accuracy of the ODE inversion. Thus, we pose the following problem. Let temperature measurements be taken at n sensor locations, $x_1, x_2, \dots, x_n$. Where should the next measurement be taken to minimise the total variance of the parameter estimates?

Let the Jacobian matrix based on n measurements with m unknown parameters be $\mathbf{J}$, a $n \times m$ matrix (in our case, $m = 2$). Then the variance–covariance matrix of parameters is expressed as (5). If a new measurement, y_{n+1} , is made at x_{n+1} , let $\mathbf{j} = \mathbf{f}'/\boldsymbol{\beta}$, a $m \times 1$ vector, where $\mathbf{f}$ is the solution to the ODE as a function of the parameter vector $\boldsymbol{\beta}$. The total variance is proportional to the trace of matrix, $\text{tr}(\mathbf{J}'\mathbf{J} + \mathbf{j}\mathbf{j}')^{-1}$. Thus, we want to minimise this criterion considering $\mathbf{J}$ fixed and $\mathbf{j} = \mathbf{j}(x_{n+1})$, a function of x_{n+1} . Letting $\mathbf{A} = \mathbf{J}'\mathbf{J}$ and using the formula (Abadir & Magnus, 2005)

$$(\mathbf{A} + \mathbf{j}\mathbf{j}')^{-1} = \mathbf{A}^{-1} - \frac{\mathbf{A}^{-1}\mathbf{j}\mathbf{j}'\mathbf{A}^{-1}}{1 + \mathbf{j}'\mathbf{A}^{-1}\mathbf{j}},$$

we obtain

$$\text{tr}(\mathbf{A} + \mathbf{j}\mathbf{j}')^{-1} = \text{tr}\mathbf{A}^{-1} - \frac{\mathbf{j}'\mathbf{A}^{-2}\mathbf{j}}{1 + \mathbf{j}'\mathbf{A}^{-1}\mathbf{j}}.$$

Finally, we arrive at an equivalent but simplified version of the total variance A-criterion

$$C(x_{n+1}) = \frac{\mathbf{j}'(x_{n+1})\mathbf{A}^{-2}\mathbf{j}(x_{n+1})}{1 + \mathbf{j}'(x_{n+1})\mathbf{A}^{-1}\mathbf{j}(x_{n+1})} \quad (7)$$

to be maximised as a function of x_{n+1} .

Finally, the adaptive algorithm for sensor location is formulated as follows.

1. Pick a few arbitrary, say, equidistant sensor locations (initial design).

2. Estimate parameters and compute matrix $\mathbf{A} = \mathbf{J}' \mathbf{J}$.
3. Maximise $C(x_{n+1})$ and add the maximiser to the previous set of sensor locations.

Iterate between steps 2 and 3 until the total variance does not decrease significantly. Particularly, iterations stop if the total variance increases.

It is instructive to find the solution to the adaptive design as the minimiser of criterion (7) when $\mathbf{j}$ is treated as a free vector with unit norm. In other words, we want to know what direction in R^m the next sensor location should be chosen to maximise C . Then criterion (7) is replaced with

$$C(\mathbf{j}) = \frac{\mathbf{j}' \mathbf{A}^{-2} \mathbf{j}}{1 + \mathbf{j}' \mathbf{A}^{-1} \mathbf{j}}.$$

Differentiating this function with respect to $\mathbf{j}$ one can show that its maximum is attained when $\mathbf{j}$ is the minimum eigenvector of matrix $\mathbf{A}$. Geometrically, the function $\mathbf{j}(x_{n+1})$ represents a curve in R^m . Then the next optimal location will have the direction of vector $\mathbf{j}(x_{n+1})$ close to the straight line defined by the minimum eigenvector of matrix $\mathbf{A}$.

We apply this algorithm to the aforementioned heat transfer problem (2) with results presented in Table 1. We start with three locations ($n = 3$) that yield $\hat{\alpha} = 4.050 \text{ cm}^{-1}$ and $\hat{T}_a = 0.715^\circ\text{C}$ with the sum of the variances 0.320. After each location is added, the total variance decreases with the final value 0.104. Note that the previously derived locations $x_1 = 0.34$ and $x_2 = 0.87$ assumed that α and T_a were known; see Figure 3. As we mentioned at the beginning of the section, other parameters can be unknown and reconstructed from additional temperature measurements, such as the ambient temperature of the rod (Beck *et al.*, 1985). Interestingly, the optimal locations at the right end of the rod are still close to 0.87 at the end of iterations, but the temperature measurement locations at the left spread around 0.23.

3.2.1 Adaptive design with fixed sensor locations—Sometimes measurements at new locations may be prohibitive because of either inaccessibility or cost. Can we still improve the precision of the parameters by making additional measurements at the existing sensor locations? Certainly, the answer is positive because taking more observations reduces the variance of the estimate. A less trivial question is, at what locations, among the existing ones, should we make the measurements?

In the general case, the choice of the site for additional measurements reduces to computation of criterion (7). Later, we find the closed-form solution for a special case with two locations, x_1 and x_2 .

Let the total number of measurements, n , be fixed. We want to know how many measurements should be taken at locations x_1 and x_2 , denoted n_1 and n_2 ; respectively ($n_1 + n_2 = n$), to obtain the non-linear least squares estimate as precise as possible. As follows, from formula (5), the search for the optimal design reduces to an optimisation problem:

$$\min_{n_1 + n_2 = n} \text{tr}(n_1 \mathbf{j}_1 \mathbf{j}_1' + n_2 \mathbf{j}_2 \mathbf{j}_2')^{-1}, \quad (8)$$

where $\mathbf{j}_1$ and $\mathbf{j}_2$ are the $m \times 1$ vector of derivatives at the two locations. After some algebra (see Appendix), one can show that the solution to this optimisation problem is

$$n_1 = \frac{\|\mathbf{j}_2\|n}{\|\mathbf{j}_1\| + \|\mathbf{j}_2\|}, \quad n_2 = \frac{\|\mathbf{j}_1\|n}{\|\mathbf{j}_1\| + \|\mathbf{j}_2\|}, \quad (9)$$

where $\|\cdot\|$ hereafter refers to the Euclidean norm. We emphasise the difference between this solution and the one-dimensional case where all measurements should be made at one location: it is simply impossible to do all measurements at one location in the two-dimensional parameter problem because then the matrix $\mathbf{J}'\mathbf{J}$ is singular and the sum of the variances is infinity.

Of course, in the case of non-linear identification, the true parameter is not known, so we do not know the derivatives even when the sensor locations are fixed and known. However, we can still apply this idea in the fashion of adaptive design assuming that the derivatives are fixed from the previous iteration.

We illustrate the adaptive design with fixed sensor locations using the example from the previous section. Now we want to estimate the heat conductivity coefficient and ambient temperature by repeated measurements assuming that the two sensors are located at $x_1 = 0.25$ and $x_2 = 0.5$ and new temperature sensors are prohibitive. The adaptive design algorithm is as follows: (1) given starting values for α and T_a , compute the 2×1 vectors $\mathbf{j}_1$ and $\mathbf{j}_2$ for x_1 and x_2 , respectively; (2) if $\|\mathbf{j}_1\| > \|\mathbf{j}_2\|$, take an additional temperature measurement at location x_1 ; otherwise, at x_2 ; (3) use non-linear regression to acquire new estimates for α and T_a and compute $\mathbf{j}_1$ and $\mathbf{j}_2$ with the updated parameters; (4) return to step (2) and iterate until the difference in parameters is small or the time limit is reached. In Table 2, we report the adaptive design iterations with increasing number of measurements at either location #1 or #2. (The true values of α and T_a are the same as in Table 1.) The iterations start from $\alpha = 5$ and $T_a = 0$; $\mathbf{j}_1$ and $\mathbf{j}_2$ are the 2×1 derivatives of (3) evaluated at $x_1 = 0.25$ and $x_2 = 0.5$, respectively (at this iteration $\mathbf{J}'\mathbf{J} = \mathbf{j}_1 \mathbf{j}_1' + \mathbf{j}_2 \mathbf{j}_2'$). Because $\|\mathbf{j}_2\| > \|\mathbf{j}_1\|$, the next temperature measurement should be taken at $x_2 = 0.5$ to minimise $\text{tr}(\mathbf{J}'\mathbf{J})^{-1}$. This yields the second iteration with $n_2 = 2$ and so forth. In this particular example, all subsequent measurements are taken at the second location—this location is pivotal in this example. Importantly, the quality of the estimation, $\text{tr}(\mathbf{J}'\mathbf{J})^{-1}$, improves from iteration to iteration—the adaptive design works.

3.3 D-design and Its Comparison with A-design

As was mentioned in Section 1, the D-design is very popular in the literature. Here, we derive the adaptive sensor design algorithm using the D-criterion that maximises $\det(\mathbf{J}'\mathbf{J})$ and compare it with the previously discussed A-design minimising $\text{tr}(\mathbf{J}'\mathbf{J})^{-1}$. Let $\mathbf{A}$ be the $m \times m$ matrix computed from the previous sensor locations as mentioned earlier. Using the identity

$$|\mathbf{A} + \mathbf{j}\mathbf{j}'| = |\mathbf{A}|(1 + \mathbf{j}'\mathbf{A}^{-1}\mathbf{j}),$$

we deduce that the best next sensor location, x_{n+1} , using the D-criterion, will maximise

$$C(x_{n+1}) = \mathbf{j}'(x_{n+1})\mathbf{A}^{-1}\mathbf{j}(x_{n+1}),$$

compared with (7) using the A-criterion. In terms of direction, $\mathbf{j}$ is the minimum eigenvector of matrix $\mathbf{A}$; the same as in the A-design—the two designs are equivalent in terms of the next direction. However, they are ultimately different in terms of the choice of x_{n+1} because $\mathbf{j}$ is not free but is a function of x_{n+1} .

Now we derive the answer for the adaptive design with fixed sensor locations using the D-criterion. Compared with (8), the optimal choice of n_1 and n_2 solves the following optimisation problem:

$$\max_{n_1 + n_2 = n} |n_1 \mathbf{j}_1 \mathbf{j}_1' + n_2 \mathbf{j}_2 \mathbf{j}_2'|.$$

After some algebra (see Appendix), we obtain that the optimal number of measurements at both sensor locations is equal, $n_1 = n_2 = n/2$. Interestingly, the A-criterion places the numbers relative to the Euclidean norm of the derivative (9), but the D-criterion suggests an equal number for each location. Clearly, the two designs are equivalent when $\|\mathbf{j}_1\| = \|\mathbf{j}_2\|$.

3.4 Assessing the Optimality of the Adaptive Design via Simulations

The goal of this section is to demonstrate by simulations that the optimal design algorithm described in Section 3.2 yields the beta-estimates that converge to the true value with the minimum sum of variances. The inverse heat transfer non-linear regression problem given by equation (3) will be used as an example. The asymptotically optimal designs are defined in the Appendix. Given the true values α and T_a , the minimum sum of asymptotic variance design for model (3) is defined by two optimal support points x_{*1} and x_{*2} such that $\text{tr}(w_{*1} \mathbf{j}_1 \mathbf{j}_1' + w_{*2} \mathbf{j}_2 \mathbf{j}_2')^{-1}$ reaches its minimum, where $\mathbf{j}_1$ and $\mathbf{j}_2$ are the 2×1 derivative vectors evaluated at x_{*1} and x_{*2} with w_{*1} and w_{*2} summing to one as optimal weights (see the Appendix). For model (3) with parameters specified in Table 1, we have $x_{*1} = 0.270$ and $x_{*2} = 0.846$ with optimal weights $w_{*1} = 0.574$ and $w_{*2} = 0.426$ yielding the minimum trace = 19.4. Figure 5 depicts simulation results for $n = 3, 4, \dots, 300$. Two typical patterns are shown: black depicts the optimal adaptive design, and green depicts the random point design

where the next x_j is chosen randomly on $[0, 1]$ to amplify the optimality of the adaptive design. Both designs produce the non-linear least squares estimates converging to the true values for α and T_α . However, the optimal adaptive design (A-criterion plot) produces estimates with the minimum sum of variances (19.4), but the A-criterion value for the random point design is considerably higher (around 35). The last plot depicts the kernel density distribution of the optimal design points at the 300th iteration. As expected, the optimal points cluster around asymptotically optimal support point values $x_{*1} = 0.270$ and $x_{*2} = 0.846$. Interestingly, the frequency repetition of the optimal design point matches the optimal weights $w_{*1} = 0.574$ and $w_{*2} = 0.426$ as the ratio of the peaks of the densities.

These simulations confirm that the non-linear least squares estimated derived in the adaptive design converge to the true values with the asymptotic covariance matrix that has minimum sum of variances.

4 Adaptive Designs with Numerically Solved ODEs

We have dealt with differential equations that admit a closed-form solution. Now we demonstrate how optimal adaptive designs can be applied to inverse differential equations when the solution is derived numerically. As in previous sections, we want to assess the location of the next measurements to identify the unknown parameters of the ODE with maximum precision.

4.1 One-parameter Inverse Diffusion Equation

In this section, we discuss the inversion of a one-dimensional diffusion/heat differential equation with unknown diffusion coefficient k on a rod of unit length with the Dirichlet-type boundary conditions (BCs) specified at 0 and 1. The inverse problem aims to estimate the diffusion coefficient by taking temperature measurements within the interval $(0, 1)$ assuming that the temperature measurements are not precise and contain random error.

Specifically, we write the second-order diffusion equation in a rod as

$$y''(x)k + f(x)y'(x) = g(x), \quad 0 \leq x \leq 1, \quad (10)$$

where the source function, $f(x)$, and the external heat function, $g(x)$, are known. The Dirichlet-type BCs are given as

$$y(0) = a, \quad y(1) = b,$$

where a and b are the constant temperatures at the ends of the rod. Because the closed-form solution of (10) is not available in general terms, we discretise the ODE using the central difference formulas

$$y''(x) \simeq \frac{1}{h^2}(y_{i+1} - 2y_i + y_{i-1}),$$

$$y(x) \simeq \frac{1}{2h}(y_{i+1} - y_{i-1}), \quad i = 2, \dots, n-1,$$

assuming that the interval $[0, 1]$ is split into n subintervals with step $h = 1/(n-1)$, so that $x_i = (i-1)/(n-1)$, and the BCs are replaced with $y_1 = a$ and $y_n = b$. Introducing an $n \times n$ matrix $\mathbf{A}$ such that $A_{11} = A_{nn} = 1$, $A_{i,i} = -2k/h^2$, $A_{i,i+1} = k/h^2 + 0.5k f_i/h$, $A_{i+1,i} = k/h^2 - 0.5k f_i/h$ for $i = 2, \dots, n-1$, and 0 elsewhere, we come to a system of linear equations $\mathbf{A}\mathbf{y} = \mathbf{g}$ with the solution as an approximation of the original ODE (10):

$$\mathbf{y} = \mathbf{A}^{-1}\mathbf{g}, \quad (11)$$

where the right-hand side is the $n \times 1$ vector $\mathbf{g} = (a, g_2, \dots, g_{n-1}, b)'$; see Holmes (2007) and Soetaert *et al.* (2012). The matrix $\mathbf{A}$ is not known and depends on the parameter k , so we write $\mathbf{A} = \mathbf{A}(k)$ to indicate this fact (vector $\mathbf{g}$ is k -independent). For some complex heat transfer problems, matrix $\mathbf{A}$ may be ill-posed. There exists a rich literature on ill-posed inverse problems, and it is not a goal of this paper to review the topic. We merely mention three books, Alifanov (1994), Aster *et al.* (2005), and Beck & Arnold (1977), where ill-posed problems emerge in the framework of parameter identification for the heat transfer problem.

In order to identify k , we need to have at least one measurement within the rod. However, because the temperature is subject to measurement error, with only one measurement, the estimate of the diffusion coefficient will be imprecise. Moreover, that would not allow assessing its standard error, and at least two measurements are needed. Let $J \geq 2$ of those measurements be available at locations $x_{j1}, x_{j2}, \dots, x_{jJ}$ with temperature readings $\tilde{y}_{j1}, \tilde{y}_{j2}, \dots, \tilde{y}_{jJ}$ that constitute a $J \times 1$ vector $\tilde{\mathbf{y}}$. Note that we can always choose the discretisation n large enough to make the locations of the measurements close to the x -grid. To express the values of $\mathbf{y}$ that correspond to the vector $\tilde{\mathbf{y}}$; we introduce a $J \times n$ index (or 'Boolean' in the language of Xiang *et al.*, 2003) matrix $\mathbf{L}$ with entries either 0 or 1 such that $\mathbf{L}\mathbf{y}$ returns the $J \times 1$ subvector of $\mathbf{y}$ where measurements were carried out on the x -grid. (Matrix $\mathbf{L}$ has full-rank $J < n$.) Then the unknown coefficient estimate, k , is obtained from the non-linear least squares criterion,

$$S(k) = \min_k \|\tilde{\mathbf{y}} - \mathbf{L}\mathbf{A}^{-1}(k)\mathbf{g}\|^2, \quad (12)$$

which can be rewritten as a non-linear regression criterion

$$S(k) = \|\tilde{\mathbf{y}} - \mathbf{f}(k)\|^2$$

letting $\mathbf{f}(k) = \mathbf{L}\mathbf{A}^{-1}(k)\mathbf{g}$. In order to minimise function S , but more importantly to derive the standard error of the estimate of k , we need the Jacobian of function $\mathbf{f}$. Several authors, including Schittkowski (2008), suggest finding the Jacobian by differentiating the original ODE with respect to the unknown parameters. An advantage of this approach is that it is independent of the method of the numerical solution of the ODE itself. However, because the differentiation leads to another ODE, solving it may be computationally demanding and may yield a derivative that is a poor approximation of the exact derivative of (12). We instead find the derivative of (12) with respect to k directly because the numerical solution is expressed as the inverse of a matrix as a function of k . Define the constant $n \times n$ matrix $\mathbf{D}$ as $\mathbf{D} = d\mathbf{A}/dk$. It is easy to see that $D_{i,i} = -2/h^2$, $D_{i,i+1} = 1/h^2 + 0.5f_i/h$, $D_{i+1,i} = 1/h^2 - 0.5f_i/h$ for $i = 2, \dots, n-1$ and 0 elsewhere. Then, using the chain rule and differentiating the inverse matrix (Abadir & Magnus, 2005; Demidenko, 2013), we finally obtain the Jacobian of $\mathbf{f}(k)$ as a $J \times 1$ vector:

$$\mathbf{J} = \frac{d\mathbf{f}}{dk} = -\mathbf{L}\mathbf{A}^{-1}\mathbf{D}\mathbf{A}^{-1}\mathbf{g}.$$

Therefore, the Gauss–Newton iterations for minimisation (12) can be written as

$$k_{s+1} = k_s + \frac{\mathbf{J}'_s(\tilde{\mathbf{y}} - \mathbf{L}\mathbf{A}^{-1}(k_s)\mathbf{g})}{\|\mathbf{J}_s\|^2}, \quad s = 0, 1, \dots,$$

where s is the iteration index and $\mathbf{J}_s = \mathbf{J}(k_s)$. This method of deriving the Jacobian and the Gauss–Newton iterations can easily be extended to the multi-parameter ODE.

As with the adaptive design, when the solution of the ODE was derived in closed form, we want to find the next location of the temperature measurement, $x = x_{n+1}$, that yields the maximum precision of the k coefficient. If σ is the standard error of the noise and

$$h = \left. \frac{df}{dk} \right|_{x=x_{n+1}},$$

we expect that the variance of $\hat{k}$, as an estimator of k based on $n+1$ measurements, is

$$\text{var}(\hat{k}_{n+1}) = \frac{\sigma^2}{\|\mathbf{J}_n\|^2 + h^2}, \quad (13)$$

where the subindex n indicates that the Jacobian is evaluated using n previous locations (Seber & Wild, 1989). Thus, we conclude that the next measurement should be made where the denominator is maximum or where the absolute value of the derivative with respect to k is maximum (Section 2). This adaptive choice of x continues until the desired precision of k is achieved and estimates converge.

4.2 Example: Heat Quality Control

Consider a real-life quality control industrial problem of a heating system described by ODE (10) with $f(x) = 5x$, $g(x) = -x$ and BCs $a = b = 1$. In the past, the diffusion coefficient was $k = 1$, but after an installation of a new rod, there was a concern that the diffusion coefficient has been changed. To test whether $k = 1$, temperature measurements are required. We solve the ODE using the finite difference approach with $N = 100$ points and compare it with the closed-form solution, which in this specific case is given by

$$y(x; k) = -\frac{1}{5}x + 1 + \frac{1}{5} \frac{2 \Phi(x\sqrt{5/k}) - 1}{2 \Phi(\sqrt{5/k}) - 1}, \quad (14)$$

where $\Phi(t)$ is the standard normal cumulative distribution function; see the left plot in Figure 6. The numerical and exact solutions are in good agreement ($k = 1$).

As was mentioned earlier, we have to measure the temperature inside the rod to reconstruct the diffusion coefficient and compare its value against $k = 1$. However, we want to reduce the number of measurements to a minimum, so they should be made at some optimal locations. In order to simulate real-life measurements, we convolute the temperature values depicted in Figure 6 with random noise (SD = 0.01). In the naive approach, we measure the temperature at n symmetric locations, say $n = 5$, $x = 0.2, 0.8, 0.35, 0.5, 0.65$. (We write the location coordinates not in ascending order to compare with the further adaptive design.) This symmetric design yields an estimate $\hat{k} = 1.013$ with SE = 0.160; see the last row of Table 3.

In the adaptive design, we start with a few locations, say $x_1 = 0.2$ and $x_2 = 0.8$ ($n = 2$), that gave $\hat{k} = 1.236$ with SE = 0.271. The next location, where the temperature should be measured is computed from formula (13), namely, where the expected variance (or equivalently SE) of k is minimum. The third optimal location ($n = 3$) is $x = 0.4385$; see the right panel of Figure 6. The new optimal location is added to the first two, new temperature measurements are taken and the process repeats. Two possibilities exist: (a) keep the previous temperature measurements after the new location is found (measure the temperature at the new location only) or (b) retake the measurements at all locations including the previous ones. (This strategy may be more appropriate if the previous measurements become obsolete.) We follow the first possibility. Although the estimated values may vary for small n , we comply with independence of $\{y_i\}$ unlike under the first rule. However, the estimates stabilise as the number of measurements increases. Importantly, SE $\hat{k}$ and $1/\|\mathbf{j}\|$ decrease from iteration to iteration: the precision of the k estimation improves with the adaptive design. In particular, SE with the adaptive design is twice smaller than with the symmetric design although the number of measurements is the same, $n = 5$.

4.3 Statistical Simulations for Design Comparison

In this section, we report the comparison of the adaptive and symmetric design (the x -values are the same as in the last row of Table 3) based on a large number of experiments using statistical simulations (the number of experiments = 10 000). Because the numerical solution

of the ODE (10) and its closed-form expression are very close, we work with the latter given by equation (14). The simulations were repeated for several SDs of the Gaussian noise, , to see how the designs behave with various degrees of measurement error; the true diffusion coefficient $k = 1$ and other parameters are set as mentioned earlier. Because we deal with one parameter, the adaptive design seeks the next x that maximises the squared derivative of $y(x; k)$ with respect to k evaluated at $k = \hat{k}_j$, the estimate of k from the previous set of x (recall that the A-optimality and D-optimality criteria lead to the same design), starting from the two-point design $x_1 = 0.2$ and $x_2 = 0.8$, as in Table 3. Fortunately, this maximisation problem admits a closed-form expression,

$$x_{\max} = \sqrt{k/5} \sqrt{1 - \frac{2\sqrt{5}\phi(\sqrt{5/k})}{\sqrt{k}(2\Phi(\sqrt{5/k}) - 1)}},$$

where $\phi = \Phi'$ is the density of the normal cumulative distribution function. We use the bias and the root mean square error to characterise the distribution of 10 000 estimates of k in each method. The results of the simulations are presented graphically in Figure 7. Obviously, the adaptive design is superior: it uses the same number of measurements as the symmetric one but has smaller bias and root mean square error. Analogous results were obtained under rule (a) when only one temperature measurement was taken at each new location.

5 Discussion and Summary Points

Some hardware configurations in industry and engineering based on non-linear models derived via ODE require estimation of unknown parameters leading to an inverse problem. Unlike the popular Bayesian approach, the adaptive design does not require a priori distribution. Instead, the optimal design is re-evaluated on the basis of the current value of the parameter during the iterative estimation process. Two situations may occur: the solution to the ODE exists in closed form or should be solved numerically. Additional measurements should be made to estimate unknown parameters. In this paper, we provide guidelines and illustrate them by examples on where to make these measurements, for example, where to put sensors, in order to estimate the parameters with maximum precision. We develop an adaptive optimal design that combines parameter identification with determination of the optimal sensor location. Typically, the D-optimality criterion is applied (minimise the generalised variance). We, however, advocate the A-optimality, by minimising the sum of the variances of the parameter estimates derived from the non-linear least squares. We argue that the A-optimality is more adequate from the statistical point of view because it may happen that the determinant of the $m \times m$ Fisher information matrix is almost zero (D-optimality is almost perfect), but $m - 1$ parameters of the model have large variance (the A-optimality is unacceptable).

We show that the A-optimality and D-optimality criteria for the adaptive sensor design are equivalent in terms of the direction but are ultimately different for the multi-parameter model. Our statistical simulations demonstrate the superiority of the adaptive design in a simple second-order diffusion equation inverse problem.

Traditionally, the optimality criteria use the Fisher information matrix and rely on the large sample normal distribution approximation of the parameter estimates. Only a handful of work, including Pazman (1986) and Pronzato & Pazman (2001) hint at the possibility of building designs for a small sample. Recent results on the exact distribution of the non-linear least squares and estimating equation approach (Demidenko, 2017) create novel opportunities in this direction.

In this paper, we deliberately use simplified models when the industry system is defined as a single ODE. Such exposition enables presenting the concepts of optimal designs with maximum clarity. The presented approach can be generalised to 2D or 3D differential equations, solved exactly or numerically by either finite difference or finite element methodology.

Appendix

Although the definitions of the asymptotic optimal design in non-linear regression can be found elsewhere but scattered through the literature, we succinctly formulate it here for the reader's convenience.

Optimal asymptotic properties of the adaptive design.

Define the non-linear regression model as $y = f(\mathbf{x}; \boldsymbol{\beta}) + \varepsilon$, where $\mathbf{x}$ is the $p \times 1$ design vector, $\boldsymbol{\beta}$ is the m -dimensional vector of unknown parameters, ε is the random error with zero mean and variance σ^2 . The goal of the optimal design is to choose design points $\mathbf{x}_i$ such that estimates of $\boldsymbol{\beta}$ based on measurements $y_i = f(\mathbf{x}_i; \boldsymbol{\beta}) + \varepsilon_i$ converge to the true value with minimum criterion function when $i = 1, 2, \dots, n \rightarrow \infty$. As before, it is assumed that error terms ε_i are independent and normally distributed with zero mean and constant variance σ^2 . This assumption implies that the maximum likelihood estimator is the non-linear least squares estimator, $\hat{\boldsymbol{\beta}}$. Suppose that for every m distinct design points $\{\mathbf{x}_k, k = 1, \dots, m\}$, matrix $\mathbf{J}'\mathbf{J} = \sum_{k=1}^m \mathbf{j}(\mathbf{x}_k; \boldsymbol{\beta})\mathbf{j}'(\mathbf{x}_k; \boldsymbol{\beta})$ is non-singular for every $\boldsymbol{\beta}$ where $\mathbf{j}(\mathbf{x}; \boldsymbol{\beta}) = f(\mathbf{x}; \boldsymbol{\beta}) / \boldsymbol{\beta}$ is the $m \times 1$ vector of derivatives with respect to parameters. For every known $\boldsymbol{\beta}$, define the optimal asymptotic design as m support design points $\mathbf{x}_{*1}, \mathbf{x}_{*2}, \dots, \mathbf{x}_{*m}$ and corresponding optimal positive weights $w_{*1}, w_{*2}, \dots, w_{*m}$ such that $\sum_{k=1}^m w_{*k} = 1$ and

$$\min_{\{\mathbf{x}_k\}, \{w_k\}} \operatorname{tr} \left(\sum_{k=1}^m w_k \mathbf{j}(\mathbf{x}_k; \boldsymbol{\beta}) \mathbf{j}'(\mathbf{x}_k; \boldsymbol{\beta}) \right)^{-1} = \operatorname{tr}(\mathbf{M}_*^{-1}), \quad (\text{A1})$$

where

$$\mathbf{M}_* = \sum_{k=1}^m w_{*k} \mathbf{j}(\mathbf{x}_{*k}; \boldsymbol{\beta}) \mathbf{j}'(\mathbf{x}_{*k}; \boldsymbol{\beta})$$

is the D-optimality asymptotic matrix. Let $\boldsymbol{\beta}$ be the true parameter vector and let for each sequence of the design points $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$ we have

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \mathbf{j}(\mathbf{x}_i; \boldsymbol{\beta}) \mathbf{j}'(\mathbf{x}_i; \boldsymbol{\beta}) = \mathbf{M}.$$

Then according to standard theory of non-linear regression

$$\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \simeq \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{M}^{-1}), \quad n \rightarrow \infty.$$

The asymptotic A-optimality design minimises the sum of asymptotic variances of the parameters $\text{tr}(\mathbf{M}_*^{-1}) \leq \text{tr}(\mathbf{M}^{-1})$, for any sequence of design points $\{\mathbf{x}_i, i = 1, 2, \dots, n\}$ when $n \rightarrow \infty$. We claim that the adaptive design is asymptotically optimal; that is, $\hat{\boldsymbol{\beta}}$ converges to the true parameter $\boldsymbol{\beta}$ with probability 1 with the asymptotic covariance matrix which trace reaches the lower bound (A1).

For the D-optimality design, the optimal support points and weights are found from the following optimisation problem:

$$\max_{\{\mathbf{x}_k\}, \{\mathbf{w}_k\}} \det \left(\sum_{k=1}^m w_k \mathbf{j}(\mathbf{x}_k; \boldsymbol{\beta}) \mathbf{j}'(\mathbf{x}_k; \boldsymbol{\beta}) \right). \quad (\text{A2})$$

The optimal weights given support points are found later.

Optimal weights.

Besides finding the optimal support points $\mathbf{x}_{*1}, \mathbf{x}_{*2}, \dots, \mathbf{x}_{*m}$, one has to determine their repetition frequency in the final design (A1), that is, the positive weights $w_{*1}, w_{*2}, \dots, w_{*m}$ summing to one. This problem has been solved by several authors for D-optimality design with regression linear in parameters, including Fedorov (1972) and Fedorov & Leonov (2013). Here we prove that, when the support points $\mathbf{x}_{*1}, \mathbf{x}_{*2}, \dots, \mathbf{x}_{*m}$ are known, for the D-optimality design (A2), the optimal weights are the same, $w_{*k} = 1/m$, and for A-optimality design, the weights are proportional to the squared norm of the gradient at the support points.

Indeed, combine m gradients at support points $\mathbf{j}_k$ as vector columns of an $m \times m$ matrix $\mathbf{J}$ such that $\det(\mathbf{J}) \neq 0$ and m weights into a diagonal $m \times m$ matrix $\mathbf{W}$. Then the D-optimality criterion can be written as maximisation of $\det(\mathbf{J}\mathbf{W}\mathbf{J}')$. Equivalently, we can maximise the log determinant,

$$\ln \det(\mathbf{J}\mathbf{W}\mathbf{J}') = \ln \det(\mathbf{W}\mathbf{J}'\mathbf{J}) = \ln \det(\mathbf{W}) + \ln \det(\mathbf{J}'\mathbf{J}).$$

Because $\ln \det(\mathbf{W}) = \sum_{k=1}^m \ln w_k$ and $\ln \det(\mathbf{J}'\mathbf{J})$ does not depend on weights, we reduced the problem of maximisation $\sum_{k=1}^m \ln w_k$ under constraint $\sum_{k=1}^m w_k = 1$. An elementary Lagrange multiplier technique yields that optimal weights of the D-optimality criterion are the same, $w_k = 1/m$.

Now we find optimal weights for the A-optimality criterion $\min \text{tr}(\mathbf{J}\mathbf{W}\mathbf{J}')^{-1}$. We have

$$\text{tr}(\mathbf{J}\mathbf{W}\mathbf{J}')^{-1} = \text{tr}(\mathbf{J}'^{-1}\mathbf{W}^{-1}\mathbf{J}^{-1}) = \text{tr}(\mathbf{W}^{-1}(\mathbf{J}'\mathbf{J})^{-1}) = \sum_{k=1}^m \frac{(\mathbf{J}'\mathbf{J})^k}{w_k},$$

where $(\mathbf{J}'\mathbf{J})^k$ denotes the k -th diagonal element of matrix $(\mathbf{J}'\mathbf{J})^{-1}$. Again, using elementary Lagrange multiplier technique, it is easy to show that the optimal weights are

$$w_k = \frac{\sqrt{(\mathbf{J}'\mathbf{J})^k}}{\sum_{l=1}^m \sqrt{(\mathbf{J}'\mathbf{J})^l}}.$$

In the case of a non-linear regression with two parameters, $m = 2$, we have $\mathbf{J}'\mathbf{J} = \mathbf{j}_1\mathbf{j}_1' + \mathbf{j}_2\mathbf{j}_2'$ and $w_{*1} = \|\mathbf{j}_2\| / (\|\mathbf{j}_1\| + \|\mathbf{j}_1\|)$ and $w_{*2} = \|\mathbf{j}_1\| / (\|\mathbf{j}_1\| + \|\mathbf{j}_1\|)$.

References

- Abadir KM & Magnus JR (2005). Matrix Algebra Cambridge: University Press.
- Alifanov OM (1994). Inverse Heat Transfer Problems New York: Springer-Verlag.
- Aster RC, Borchers B & Thurber CH (2005). Parameter Estimation and Inverse Problems Amsterdam: Elsevier Academic Press.
- Atkinson AC & Donev AN (1992). Optimum Experimental Designs, 2nd ed. Oxford: Clarendon Press.
- Beck JV & Arnold KJ (1977). Parameter Estimation in Engineering and Science New York: Wiley.
- Beck JV, Blackwell B & Clair CRS (1985). Inverse Heat Conduction New York: Wiley.
- Box GEP & Lucas HL (1959). Design of experiments in non-linear situations. Biometrika, 46, 77–90.
- Demidenko E (2013). Mixed Models: Theory and Applications with R, 2nd ed Wiley: Hoboken.
- Demidenko E (2017). Exact and approximate statistical inference for nonlinear regression and the estimating equation approach. Scand. J. Stat, 44, 636–665.
- Duarte BPM & Wong WK (2015). Finding Bayesian optimal designs for nonlinear models: a semidefinite programming-based approach. Int. Stat. Rev, 83, 239–262. [PubMed: 26512159]
- Emery AF & Fadale TD (1996). Design of experiments using uncertainty information. J Heat Transfer, 118, 532–538.
- Emery AF, Nenarokomov AV & Fadale TD (2000). Uncertainties in parameter estimation: the optimal experiment design. Int. J. Heat Mass Transfer, 43, 3331–3339.
- Fadale TD, Nenarokomov AV & Emery AF (1995). Two approaches to optimal sensor location. J. Heat Transfer, 117, 373–379.
- Fedorov VV (1972). Theory of Optimal Experiments New York: Academic Press.
- Fedorov VV & Leonov SL (2013). Optimal Design for Nonlinear Response Models Boca Raton: CRC Press.
- Franceschini G & Macchietto S (2008). Model-based design of experiments for parameter precision: state of the art. Chem. Eng. Sci, 63, 4846–4872.

- Glaholt SP, Chen CY, Demidenko E, Bugge DM, Folt CL & Shaw JR (2012). Adaptive iterative design (AID): a novel approach for evaluating the interactive effects of multiple stressors on aquatic organisms. *Sci. Total Environ*, 432, 57–64. [PubMed: 22717606]
- Hoffman JD (1992). *Numerical Methods for Engineers and Scientists*, 2nd ed. New York: Marcel Dekker.
- Holmes M (2007). *Introduction to Numerical Methods in Differential Equations* New York: Springer.
- Kiefer J (1974). General equivalence theory for optimum designs (approximate theory). *Ann. Stat*, 2, 849–879.
- Ozisik MN & Orlande HRN (2000). *Inverse Heat Transfer Fundamentals and Applications* New York: Taylor & Francis.
- Pazman A (1986). *Foundations of Optimum Experimental Design* Dordrecht: Kluwer.
- Pong A & Chow S-C. (2011). Overview of adaptive design methods in clinical trials. In *Handbook of Adaptive Designs in Pharmaceutical and Clinical Development*, Eds. Pong A & Chow S-C, pp. 1–19. Boca Raton: CRC Press.
- Pronzato L (2008). Optimal experimental design and some related control problems. *Automatica*, 44, 303–325.
- Pronzato L & Pazman A (2001). Using densities of estimators to compare pharmacokinetic experiments. *Comput. Biol. Med*, 31, 179–195. [PubMed: 11173056]
- Pukelsheim F (2006). *Optimal Design of Experiments* Philadelphia: SIAM.
- Schittkowski K (2008). Parameter identification and model verification in systems of partial differential equations applied to transdermal drug delivery. *Math. Comput. Simulat*, 79, 521–538.
- Seber GAF & Wild CJ (1989). *Nonlinear Regression* New York: Wiley.
- Silvey SD (1980). *Optimal Design* London: Chapman & Hall.
- Soetaert K, Cash J & Mazzia F (2012). *Solving Differential Equations in R* Heidelberg: Springer-Verlag.
- Ucinski D (2000). *Optimal Measurement Methods for Distributed Parameter System Identification* Boca Raton: CRC Press.
- Wells D (1991). *The Penguin Dictionary of Curious and Interesting Geometry* London: Penguin.
- Xiang ZH, Swoboda G & Cen ZZ (2003). On the optimal layout of displacement measurements for parameter identification process in geomechanics. *Int. J. Geomech*, 3, 205–216.
- Li Y, Xiang Z, Zhou M & Cen Z (2008). An integrated parameter identification method combined with sensor placement design. *Commun. Numer. Methods Eng*, 24, 1571–1585.

Figure 1.

The problem of a hanging wire. Where should measurements be made to estimate the sagging parameter; β , with the maximum precision?

Figure 2.

The inverse heat transfer problem in a homogeneous rod with $\alpha = 4\text{ cm}^{-1}$. The temperature at both ends is kept at 10°C with the ambient temperature $T_a = -5^\circ\text{C}$. The resulting temperature distribution is symmetric. The optimal sensor locations are $x_1 = 0.5$ and $x_2 = 0.88$.

Figure 3.

Asymmetric case: the inverse heat transfer problem when $T(0) \neq T(1)$. The optimal sensor locations are $x_1 = 0.34$ and $x_2 = 0.87$.

Figure 4.

Left panel ($T_0 = T_1$): the contours of the total variance as a function of two sensor locations.

One can see two solutions, but due to the symmetry of the temperature distribution in the rod, there is only one: the optimal sensor locations are $x_1 = 0.12$ and $x_2 = 0.5$, or $x_1 = 0.5$ and $x_2 = 0.88$.

Right panel ($T_0 = 0$ and $T_1 = 10$): the contours of the total variance in an asymmetric temperature distribution with optimal sensor location $x_1 = 0.34$ and $x_2 = 0.87$.

The values of α and T_α are known in both cases.

Figure 5. Simulation results confirming the optimality of the adaptive design for the inverse heat transfer non-linear regression problem.

Figure 6.

Left: exact and numerical solutions to ordinary differential equation (ODE) (10). Right: the third optimal location, $x = 0.4385$, maximises the squared Jacobian on $(0, 1)$ with $n = 3$. The triangles indicate the first two previous locations.

Figure 7. Comparison of the symmetric with adaptive design using statistical simulations (true $k = 1$, number of simulations = 10,000). The latter has a smaller bias and root mean square error for the entire range of the SD noise.

Table 1.

Iterations of optimal adaptive design for sensor locations.

n	$\hat{\alpha}(\text{cm}^{-1})$	$\hat{T}_d(^{\circ}\text{C})$	Var	x_1	x_2	x_3	x_4	x_5	x_6
3	4.378	0.270	2.084	0.25	0.50	0.75			
4	4.353	0.256	0.682	0.25	0.50	0.75	0.869		
5	4.516	0.452	0.365	0.25	0.50	0.75	0.869	0.263	
6	4.667	0.375	0.127	0.25	0.50	0.75	0.869	0.263	0.859

The data are simulated with true values $\alpha = 4 \text{ cm}^{-1}$ and $T_0 = 0^{\circ}\text{C}$, $T_1 = 10^{\circ}\text{C}$ and $T_d = 0^{\circ}\text{C}$, assuming a measurement error $\sigma = 0.25^{\circ}\text{C}$.

Iterations of the optimal adaptive design with fixed sensor locations at $x_1 = 0.25$ and $x_2 = 0.5$ (one new measurement per iteration).

Table 2.

iter	$\hat{\vartheta}(\text{cm})^{-1}$	$\hat{T}_d(^{\circ}\text{C})$	$\hat{\sigma}$	$r(\mathbf{J}'\mathbf{J})^{-1}$	$\ \mathbf{j}_1\ $	$\ \mathbf{j}_2\ $	n_1	n_2
1	5.000	0.000	0.830	1.481	0.709	0.929	1	1
2	3.924	-0.401	0.587	0.930	0.679	1.018	1	2
3	3.903	-0.414	0.525	0.841	0.680	1.022	1	3
4	3.527	-0.697	0.459	0.705	0.713	1.116	1	4
5	3.631	-0.607	0.414	0.706	0.700	1.085	1	5

Table 3.

Optimal temperature locations in the second-order diffusion ODE.

n	$\hat{k}$	$SE \hat{k}$	$1/\ \mathbf{j}\ $	x_1	x_2	x_3	x_4	x_5
Adaptive optimal design								
2	1.236	0.271	43.15	0.20	0.80			
3	1.067	0.133	21.75	0.20	0.80	0.4385		
4	0.974	0.100	15.05	0.20	0.80	0.4385	0.4206	
5	0.997	0.079	13.25	0.20	0.80	0.4385	0.4206	0.4088
Symmetric design								
5	1.013	0.160	14.84	0.20	0.80	0.35	0.50	0.65