Statistics and Computing manuscript No.
(will be inserted by the editor)

Optimal experimental design for linear time invariant state–space models

Belmiro P.M. Duarte · Anthony C. Atkinson · Nuno M.C. Oliveira

Received: date / Accepted: date

Abstract The linear time invariant state–space model representation is common to systems from several areas ranging from engineering to biochemistry. We address the problem of systematic optimal experimental design for this class of model. We consider two distinct scenarios: (i) steady-state model representations and (ii) dynamic models described by discrete-time representations. We use our approach to construct locally D–optimal designs by incorporating the calculation of the determinant of the Fisher Information Matrix and the parametric sensitivity computation in a Nonlinear Programming formulation. A global optimization solver handles the resulting numerical problem. The Fisher Information Matrix at convergence is used to determine model identifiability. We apply the methodology proposed to find approximate and exact optimal experimental designs for static and dynamic experiments for models representing a biochemical reaction network where the experimental purpose is to estimate kinetic constants.

Keywords Optimal design of experiments · Linear time invariant systems · State-space models · Model identifiability · Biochemical reaction networks

Mathematics Subject Classification (2000) 62K05 · 90C47

Belmiro P.M. Duarte
Instituto Politécnico de Coimbra, Instituto Superior de Engenharia de Coimbra, Department of Chemical and Biological Engineering, Rua Pedro Nunes, Quinta da Nora, 3030−199 Coimbra, Portugal, and CIEPQPF, Department of Chemical Engineering, University of Coimbra, Rua Sílvio Lima – Pólo II, 3030-790 Coimbra, Portugal. Tel.: +351-239-790200, Fax: +351-239-790201, E-mail: bduarte@isec.pt

Anthony C. Atkinson
Department of Statistics, London School of Economics, London WC2A 2AE, United Kingdom. E-mail: A.C.Atkinson@lse.ac.uk

Nuno M.C. Oliveira
CIEPQPF, Department of Chemical Engineering, University of Coimbra, Rua Sílvio Lima – Pólo II, 3030-790 Coimbra, Portugal. E-mail: nuno@eq.uc.pt

1 Motivation

As a motivating example we consider a (bio-) chemical reaction network modeled by a Linear-Time Invariant (LTI) state–space model, which also provides the basis for our numerical exploration of the calculation and properties of optimum experimental designs. For a recent review of (bio-) chemical reaction networks the reader is referred to Loskot et al. (2019). The reaction kinetic model describing the convective-based transfer network of a given component will only have linear kinetics. Specifically, the inflow and outflow terms are represented using zero order kinetics with the intermediate transitions being of first order (Hangos et al., 2013). The n chemical species (metabolites) of a reaction network are denoted as $\chi_1, \cdots, \chi_n$, the concentrations of which are respectively $x_1, \cdots, x_n$, forming the vector $\mathbf{x}$ of state variables. A network can be represented by a stoichiometric matrix $T \in \mathbb{R}^{n_x \times n_r}$ where each entry $T_{i,j}$ represents the production or consumption of metabolite $i \in \{1, \cdots, n_x\}$ in reaction $j \in \{1, \cdots, n_r\}$. The $T_{i,j}$ are +1 if the metabolite is produced, −1 if it is consumed and 0 otherwise; n_r is the number of reactions in the network.

In modern biological research, it is very common to collect detailed information on time dependent chemical concentration data for large networks of biochemical reactions (Crampin et al., 2004). One technique uses the differential uptake of isotopes of carbon. Tracers containing an increased amount of Carbon-13 can be introduced into the network to aid the identification of molecules. Examples, with a discussion of experimental design, include Bouvin et al. (2015) and Wiechert et al. (2001). A first step in design is to ensure that proposed measurements lead to identification of potential models. The main purpose in the chemical networks with which this paper is concerned is then to identify the exact structure of the network of chemical reactions and to provide efficient estimates of the respective rate constants. We consider model identifiability and parameter estimation problems for both steady-state and dynamic models.

This paper addresses the D–optimal design of experiments for parameterizing LTI state–space models, handling both *static* and *dynamic* kinds of experiments. *Static experiments* are those where inputs are initially chosen and kept constant over time until the system reaches a steady state, and the underlying model is the steady-state LTI state–space model. Here, the experimenter will run the experiments several times, with a different constant input $\mathbf{u}$ at each trial, and the response $\mathbf{y}$ will be observed after the system has converged to a new steady-state. In *dynamic experiments* the inputs can vary during the experiment at a previously defined grid of time instants of the experimental horizon. These experiments are to be run a single time, and the inputs remain constant during the time slots forming the discretized grid, but are allowed to change at their bounds. The process is also sampled at a previously set grid of times, that may (or may not) coincide with the grid used to manipulate the inputs.

1.1 Models and related literature

Here, we introduce the formalism of LTI space-state models. In this paper we employ the nomenclature commonly used in systems theory. Specifically, $\mathbf{x} \in \mathbf{X} \subset \mathbb{R}^{n_x}$

denotes the vector of *state variables* that fully characterize the state of the system, $\mathbf{u} \in \mathbf{U} \subset \mathbb{R}^{n_u}$ is the vector of *control variables*, known without error, and used as *control factors* in the experiments, $\mathbf{y}$ is the vector of variables measured in the experiment, called *responses*, i.e. $\mathbf{y} \in \mathbf{Y} \subset \mathbb{R}^{n_y} \subseteq \mathbb{R}^{n_x}$ where $n_y (\leq n_x)$ is the number of response variables, and $\boldsymbol{\theta} \in \boldsymbol{\Theta} \subset \mathbb{R}^{n_\theta}$ the vector of parameters to be estimated. Here, $\mathbf{U}, \mathbf{X}, \mathbf{Y}$ and $\boldsymbol{\Theta}$ are compact domains of factors, states, responses and parameters, respectively; $\boldsymbol{\Theta} = \bigotimes_{j=1}^{n_\theta} [\theta_j^{\mathrm{LO}}, \theta_j^{\mathrm{UP}}]$ is a compact set in the domain of the parameters, θ_j represents a local value of parameter j and θ_j^{LO} and θ_j^{UP} are the lower and upper values admissible for θ_j. The LTI state–space model relates control variables, states and responses as follows

$$\frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t} = A(\boldsymbol{\theta})\, \mathbf{x} + B\, \mathbf{u}(t) \tag{1a}$$

$$\mathbf{y} = C\, \mathbf{x} + \boldsymbol{\epsilon}, \tag{1b}$$

where $A(\boldsymbol{\theta}) \in \mathbb{R}^{n_x \times n_x}$ is a time-invariant (i.e., independent of time) matrix, $B \in \mathbb{R}^{n_x \times n_u}$ is the matrix of constants relating the state variables to the inputs, and $C \in \mathbb{R}^{n_y \times n_x}$ is the matrix of coefficients relating the responses to the state variables. Equation (1a) is the state equation; $A(\boldsymbol{\theta})$ can be sparse with many entries 0. The time $t \in [0, H]$ is bounded by H, the horizon to be considered in the simulation (or experiment). B and C are also time invariant and known. Equation (1b) is the measurement equation. Let $\boldsymbol{\epsilon} \in \mathbb{R}^{n_y}$ be the vector of observational errors; each component ϵ_i is described by an i.i.d. Gaussian probability distribution $\mathcal{N}(0, \sigma_i)$ with mean 0 and standard error σ_i, $i \in \{1, \cdots, n_y\}$. The dynamic behavior of process states (and responses) can be adjusted by manipulating the control factors. Thus, the dynamic model includes the time dependence of $\mathbf{u}$, i.e. $\mathbf{u}(t)$. A steady-state version of the model (1) is

$$\mathbf{x} = -A(\boldsymbol{\theta})^{-1}\, B\, \mathbf{u} \tag{2a}$$

$$\mathbf{y} = C\, \mathbf{x} + \boldsymbol{\epsilon}. \tag{2b}$$

Here, none of the states, responses and control factors are dependent on time as they refer to a time instant where the accumulation term $(\mathrm{d}\mathbf{x}/\mathrm{d}t)$ is null and we omit the dependence of $\mathbf{u}$ on t. $A(\boldsymbol{\theta})$ is assumed invertible to avoid underdetermined parametrization.

More generally, state–space models are mathematical representations of the dynamics of general systems relating input, state and output variables. They are formalized as first order differential equations and allow a convenient algebrization and compactness. Fundamental theoretical results for establishing the properties of state–space models and constructing optimal input signals for system identification can be found in a vast range of references (Goodwin and Payne, 1977; Kalaba and Spingarn, 1982; Ljung, 1999; Titterington, 1980). The optimal design of experiments in the time domain and in the frequency domain has been considered for the estimation of correctly parameterized models. The problem in the time domain reduces to a nonlinear optimal control problem; the complexity was one of the reasons that motivated researchers to find input designs in the frequency domain. The problem in the frequency domain aims at finding a set of finitely parametrized inputs that

parametrizes all achievable information matrices (Goodwin and Payne, 1977; Mehra, 1974; Zarrop, 1979). One of the problems of frequency domain based methods is their inability to take time domain constraints into consideration. Here, we address the problem in the time domain. An early reference to ODoE for time discrete model identification is Goodwin and Payne (1977, Chap. 6) where an adjoint state approach was considered for the solution (Kalaba and Spingarn, 1973). Additional results can be found in Zarrop (1979, Chap. 2). The formalization of the problem as an optimal control problem appeared in Asprey and Macchietto (2000); Espie and Macchietto (1989); Körkel et al. (2004); Rudolph and Herrendörfer (1995) among others. The time domain is discretized and the decision variables and control actions parametrized in each interval so that the number of decision variables is finite (Bryson, 1999). This approach is known as *dynamic optimization*; the original problem is approximated by an algebraic representation and may be cast as a NLP solved with convenient algorithms. Herein, we use a dynamic optimization-based approach to handle the optimal ODoE problem for time discrete models.

The representation of (bio)- chemical reaction networks is an area of chemical and biological engineering where state–space models find extensive application (Anderson et al., 2011). Very often the parametrization and identification of reaction networks require experimental work, and the application of the fundamentals of optimal design of experiments (ODoE) may rationalize and reduce the amount of work needed. Here, we use the customary parametrization of (bio)- chemical reaction networks as the motivating example for proposing general formulations for optimal design of static and dynamic experiments for systems represented by Linear-Time Invariant state–space models.

The steady and time discrete LTI state–space models find direct application in NMR spectroscopy, network traffic flow, signal processing, control theory and reaction network modeling among others. The optimal design of experiments considered in this paper aims at determining conditions that provide measurements so that the parameters $\boldsymbol{\theta}$ in $A(\boldsymbol{\theta})$ are estimated with minimum confidence region. This, in turn, requires maximizing a measure of the Fisher Information Matrix (FIM). The choice of optimal sampling strategies for system identification was considered by Mehra (1974) and Ng and Goodwin (1976), among others. The construction of optimal input signals for biological systems identification was considered by Cobelli and Thomaseth (1985), and the identifiability of the state–space model was analyzed by Walter (2013).

The applications of dynamic experimentation aim at finding the optimal sequence of actions on input variables and/or time instants at which sampling is required so that the information obtained from experiments is maximized (Asprey and Macchietto, 2000; Espie and Macchietto, 1989). The problem is formulated as an optimal control problem (Pronzato, 2008; Zarrop, 1979), handled numerically with dynamic optimization techniques (Hoang et al., 2013; Körkel et al., 2004). Recent applications include systems with continuous measurement (Galvanin et al., 2011), online redesign of experiments considering the amount of information gathered previously and model inaccuracy (Galvanin et al., 2012), the design of robust experiments taking into account the uncertainty of the model and violation of the constraints (Telen et al., 2018) and an application to a real case study where local identifiability is simultaneously monitored and used to transform the problem into a well-conditioned equivalent form

(Barz et al., 2013). These applications focus on the optimal design of experiments for dynamic models, and address locally optimal designs for general nonlinear state–space models; but they disregard the static experimentation that can be considered for some processes. The LTI state–space model has specificities that broaden its application to a large range of systems, namely a structure that:

i. allows improving the numerical efficiency of the formulations for computing optimal experimental designs through taking advantage of the sparsity of $A(\boldsymbol{\theta})$; and
ii. generalizes to the use of both static and dynamic experiments for parameter estimation.

The ODoE for LTI models represented by state–space models has not been addressed consistently although there is a recognized interest in the optimization of the experimental work for estimating the parameters and checking whether or not the model is identifiable. The problem was addressed by Brown et al. (2008) for measurement selection in Chemical Reaction Network (CRN) characterization; Maidens and Arcak (2016) considered the problem of finding the optimal substrate injection to characterize metabolic networks using magnetic resonance imaging. Their proposed solution used a Semidefinite Programming formulation; another context where ODoE was applied to LTI systems is in inference about traffic flow networks (Sagnol, 2010; Singhal and Michailidis, 2010) and optimal design of Kalman filters (Sagnol and Harman, 2015b). The application of nonlinear state–space models to CRN inference was considered, among others, by Chis et al. (2016); Eisenberg and Hayashi (2014); Telen et al. (2014); Villaverde (2019). We observe that there is a lack of systematic methods for finding optimal experimental designs specifically for LTI state–space models which simultaneously take advantage of the topology of the model representation and of the accuracy and efficiency of the optimization algorithms currently available. Further, a strategy that can adapt to find optimal experimental designs for both steady-state and dynamic LTI state–space models is certainly a research topic worth pursuing. Finally, since the optimal design of experiments can be viewed as maximizing a (quantitative) measure of model identifiability, where the usual practice is to maximize some functional of the FIM (Walter and Pronzato, 1985, 1988), one can use the FIM at convergence of the ODoE problem as a local check of model identifiability.

1.2 The LTI state–space model representation of CRNs

Now we recall the CRN reaction rate estimation problem conceptualized in §1 and demonstrate that under certain conditions it can be represented by an LTI state–space model of form (1). We note the formulations developed are general and can be applied to models of different areas with the CRN parametrization being one of them.

Let $\mathbf{v} \in \mathbb{R}^{n_r}_{\geq 0}$ be the vector of fluxes (or reaction rates) expressed in units of quantity of matter consumed (or produced) per time. When the network only involves first order kinetics, $\mathbf{v} = L(\boldsymbol{\theta})\ \mathbf{x}$, and the model representing the species concentration network

becomes

$$\frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t} = S\,\mathbf{v} = S\,L(\boldsymbol{\theta})\,\mathbf{x} = A(\boldsymbol{\theta})\,\mathbf{x} + B\,\mathbf{u}(t), \tag{3}$$

where S is the constant stoichiometric matrix (Varma and Palsson, 1994) and $L(\boldsymbol{\theta})$ contains all the kinetic parameters, even those referring to the conceptual zero-order reactions converting inflow terms into intermediate metabolites. Matrix $L(\boldsymbol{\theta})$ can be decoupled into matrices $A(\boldsymbol{\theta})$, which includes all the kinetic rates used for representing first order kinetics, and B assumed to be known. In this context $\mathbf{u}$ is the vector of inflow terms and refers to reactant species entering the network (with zero-order kinetics). The rates of the conceptual zero-order reactions give the desired values of $\mathbf{u}$. That is, choosing u_i is choosing the conceptual rate of the zero-order (or saturated) reactions modeling inflows. When $u_i = 0$, the flux is deactivated, otherwise, when $u_i > 0$, the flux is activated. In the optimal design of experiments, the vector of support points $\mathbf{u}$ is chosen to maximize a given information criterion of the network parametrization. This optimal choice ensures the local identifiability of the model parameters (when this is possible) and the most precise estimation of the model parameters, in the sense of a scalar function of their asymptotic covariances being minimal. In turn, the vector of measurements corresponds to the set of states that are measured, and Equation (1b) is used for forecasting the responses.

As an example, consider the kinetic network formed by 6 state variables ($n_x = 6$) with 3 input variables ($n_u = 3$) and 10 parameters to be estimated ($n_\theta = 10$) represented by matrices (Frøysa et al., 2020)

$$A(\boldsymbol{\theta}) = \begin{pmatrix} -\theta_1-\theta_2-\theta_3 & 0 & 0 & 0 & 0 & 0 \\ \theta_1 & -\theta_4-\theta_5 & 0 & 0 & 0 & 0 \\ 0 & \theta_4 & -\theta_6-\theta_9 & 0 & 0 & 0 \\ \theta_2 & 0 & 0 & -\theta_7 & 0 & 0 \\ \theta_3 & \theta_5 & 0 & \theta_7 & -\theta_8 & 0 \\ 0 & 0 & \theta_6 & 0 & \theta_8 & -\theta_{10} \end{pmatrix}, \quad B = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

and graphically represented in Figure 1. Here, we consider that all process states are measured from the experiments, i.e., the experimental response variables correspond to the states affected by observational error. It means that the matrix C in (2b) is the identity matrix of size 6.

Briefly, the ODoE aims at finding the set of combinations of $\mathbf{u}$ that assure maximization of a given measure of the information content. The example introduced above of the identification of CRN is used for demonstration, since the characterization of chemical reaction networks is currently of appreciable interest, playing an important role in systems biology, (bio-) chemical engineering, and the emerging field of synthetic biology (Loskot et al., 2019; van der Schaft et al., 2016). The problem considered by Frøysa et al. (2020) offered a motivating example and we use their results for comparison. Their strategy to find the optimal design of experiments involves (i) generating a set of potential control action vectors using the vertex of the design space; and (ii) find the optimal combination. This approach does not consider explicitly the time in the dynamics (or the time discretization interval Δt at which the system is sampled and actions implemented), only finds the set of actions that lead to information maximization. Practically, it is comparable to our framework to

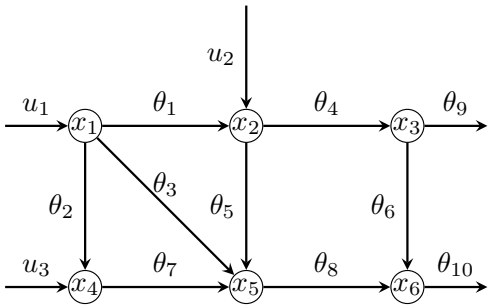


Figure 1 Example of kinetic metabolic network.

determine optimal static experiments. We extend the results in their paper and include the time explicitly in the model. Thus, we find the optimal vectors of actions (which can be similar) and the optimal sequence having into account the system dynamics; this sequence can not be interchanged without loosing information. Practically, our formulation for dynamic experimentation relies on the representation of the problem as an optimal control problem, and we use a dynamic optimization approach to handle the problem numerically which requires discretizing the horizon of the experiment. Next, the NLP problem is solved using a simultaneous approach exploiting the representation of dynamic profiles by parametric approximations.

1.3 Novelty and organization

This paper contains five elements of novelty:

i. the development of general Nonlinear Programming (NLP) formulations to find continuous and exact locally D–optimal experimental designs to estimate the parameters in matrix $A(\boldsymbol{\theta})$ in steady-state LTI state–space models employing static experiments;
ii. the development of general NLP formulations to find locally D–optimal experimental designs for time discrete LTI state–space models employing dynamic experimentation where the control actions are optimized so that they assure the maximization of the information gathered in the experiment and the dynamics of the system is explicitly considered. The resulting optimal control problem is cast as a dynamic optimization problem after discretizing the time domain and parametrizing the variables profiles at the grid; the problem is then solved with a simultaneous-based technique;
iii. to avoid the effect of inaccuracy in the estimation of variables on the amount of information, we use a dual time grid, where a tighten grid is used for updating the state and response variables and a coarser grid used to sample the process and actuate;
iv. the diagnostic of model identifiability using the results of the ODoE, specifically the FIM at convergence; and

v. demonstrating the application of the formulations proposed to the identification of biochemical reaction networks.

The paper is organized as follows. Section 2 introduces the background and the notation used to formulate the problem as well as the fundamentals of nonlinear programming and the time discrete LTI model. Section 3 presents the mathematical programming formulation for finding D–optimal designs for steady-state and dynamic state–space models. Details of the construction of the FIM are given, which in turn requires the calculation of the sensitivity coefficients. Section 4 applies the previous formulations to finding optimal designs. Comparisons involving uniform and non-uniform optimal designs for steady-state models are provided and compared to ODoE for dynamic models. After solving the ODoE problem, model identifiability is analyzed. Finally, Section 5 reviews the formulation and offers a summary of the results obtained.

2 Notation and background

This section establishes the nomenclature used in the representation of the models. In §2.1 we present the experimental design problems outlined above and in §2.2 we address the dynamic solution of the LTI model considering a time-discrete representation as well as the construction of the FIM. Then, §2.3 overviews the fundamentals of NLP.

2.1 Optimal experimental design

Bold face lowercase letters represent vectors, bold face capital letters continuous domains, blackboard bold capital letters discrete domains and capital letters are for matrices. Finite sets containing ι elements are compactly represented by $[\![\iota]\!] \equiv \{1, \cdots, \iota\}$. The transpose operation of a matrix/vector is represented by "$^\intercal$".

To introduce the theoretical background to formalize the ODoE problem we consider static experimentation. Accordingly, the steady-state model (2) is used for describing the process, and $\mathbf{u}$ is independent of time. We consider continuous designs with K support points at $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_K$ where each vector $\mathbf{u}_i,\ i \in [\![K]\!]$, corresponds to a combination of control factors (constants) used in each trial; the process is observed after reaching the steady-state. The weights representing the relative effort at these points are, respectively, $w_1, w_2, \ldots, w_K$ where K is chosen by the user so that $K \times n_y \geq n_\theta$.

Let N be the total number of experiments of the experimental plan. *Continuous designs* are used to represent experimental setups where $N \to +\infty$; consequently the weights vary continuously on $[0,\ 1]$. To implement continuous designs we take roughly $N \times w_k$ experiments at $\mathbf{u}_k$, $k \in [\![K]\!]$, subject to $N \times w_1 + \cdots + N \times w_K = N$, and each summand is an integer. For models with n_u control factors, we denote the kth support point by $\mathbf{u}_k^\intercal = (u_{k,1}, \ldots, u_{k,n_u})$ and represent the design ξ by K rows $(\mathbf{u}_k^\intercal, w_k),\ k \in [\![K]\!]$ with $\sum_{k=1}^{K} w_k = 1$. In what is to follow, we let $\Xi \equiv \mathbf{U}^K \times \Sigma$ be the space of feasible K-point designs over $\mathbf{U}$ where Σ is the $K-1$-simplex in the domain

of weights, i.e. $\Sigma = \{w_k : w_k \geq 0, \forall k \in [\![K]\!], \sum_{k=1}^{K} w_k = 1\}$. We notice that uniform designs have $w_i = 1/K$. In finding continuous optimal designs the weights are not restricted and these designs are usually non-uniform, that is some of the weights are larger than others.

Exact designs are experimental plans where the weights w_k are ratios n_k/N satisfying the conditions: (i) all n_k's are integer (or null); and (ii) sum to N. In practice, exact designs are obtained from continuous designs considering an experimental plan with N experiments and using a rounding procedure (Pukelsheim and Rieder, 1992) or Mixed Integer Nonlinear Programming formulations (Duarte et al., 2020) to allocate them to support points.

The log-likelihood function for the parameter estimation problem after the experimental data are available reduces to the least squares problem

$$\mathcal{L}(\mathbf{y}, \boldsymbol{\theta}) = \sum_{i=1}^{n_y} \sum_{j=1}^{n_e} (\eta_{i,j}^{\text{obs}} - \eta_{i,j})\, V^{-1}\, (\eta_{i,j}^{\text{obs}} - \eta_{i,j})^{\intercal}. \tag{4}$$

See, for example, (Fedorov and Leonov, 2014, Chap. 1). Here, V is the (constant) variance-covariance matrix, $\eta_{i,j}^{\text{obs}}$ refers to measurements of y_i from the jth experiment, $\eta_{i,j}$ stands for predictions constructed using model (1), and n_e is the number of experiments. Consequently, the corresponding global FIM at a singleton point $\mathbf{p} \in \boldsymbol{\Theta}$ for continuous optimal design ξ is

$$\begin{aligned}\mathcal{M}(\xi|\mathbf{u}, \boldsymbol{\theta}) &= -\mathbb{E}\left[\frac{\partial}{\partial \boldsymbol{\theta}}\left(\frac{\partial \mathcal{L}(\xi|\mathbf{p})}{\partial \boldsymbol{\theta}^{\intercal}}\right)\right] = \sum_{j=1}^{K} w_j\, M(\mathbf{u}_j|\mathbf{p}) = \\ &= \sum_{j=1}^{K} w_j\, F(\mathbf{u}_j|\mathbf{p})^{\intercal}\, V^{-1}\, F(\mathbf{u}_j|\mathbf{p}),\end{aligned} \tag{5}$$

where $\mathbf{w}$ is the vector of weights of the support points in the design (or experiments if a static experimental setup is adopted), K is the number of support points, previously set by the user, $M(\mathbf{u}_j|\mathbf{p})$ is the elemental FIM at $\mathbf{u}_j$. Further, we assume V is a $n_y \times n_y$ identity matrix as in Draper and Hunter (1966), i.e. the measurement error of each of the responses is independent of the others and their standard error is equal; $\mathbb{E}[\bullet]$ stands for expectation. Let $F(\mathbf{u}_j|\mathbf{p})$ be the sensitivity of the measurements with respect to the parameters at support point j, i.e., $F(\mathbf{u}_j|\mathbf{p}) = C\, \partial\mathbf{x}/\partial\boldsymbol{\theta}|_{\mathbf{u}_j,\mathbf{p}}$.

To derive the sensitivity matrix consider the steady-state LTI state–space model (2). Let $A(\boldsymbol{\theta}) = \sum_{i=1}^{n_\theta} \theta_i\, E_i$ where E_i is a $n_\theta \times n_\theta$ matrix populated with elements "+1", "0" and "−1" such that $E_i = \frac{\partial A(\boldsymbol{\theta})}{\partial \theta_i}$. Using (2a), chain-rule differentiation leads to

$$F(\mathbf{u}|\boldsymbol{\theta}) = \frac{\partial \mathbb{E}(\mathbf{y})}{\partial \boldsymbol{\theta}} = \oplus_{i=1}^{n_\theta}\, C\, A(\boldsymbol{\theta})^{-1}\, E_i\, A(\boldsymbol{\theta})^{-1}\, B\, \mathbf{u}. \tag{6}$$

where the symbol $\oplus$ is used to represent the concatenation of columns into a matrix, and $F(\mathbf{u}|\boldsymbol{\theta})$ is an $(n_y \times n_\theta)$-matrix, whose i^{th} column is $C\, A(\boldsymbol{\theta})^{-1}\, E_i\, A(\boldsymbol{\theta})^{-1}\, B\, \mathbf{u}$. Despite the linearity of $A(\boldsymbol{\theta})$ its inversion leads to nonlinear dependence of $\mathbf{y}$ on $\boldsymbol{\theta}$. Thus, we focus on locally optimal designs, as they do ensure optimality for a vector,

i.e. $\boldsymbol{\theta} \equiv \mathbf{p}$. Similar approaches for calculating sensitivities are discussed in Perry et al. (2006).

Herein, we focus on the class of design criteria proposed by Kiefer (1974) where each member in the class, indexed by a parameter δ, is positively homogeneous and defined on the set of symmetric $n_\theta \times n_\theta$ semi-positive definite matrices given by

$$\Phi_\delta[\mathcal{M}(\xi|\mathbf{u},\boldsymbol{\theta})] = \left[\frac{1}{n_\theta}\ \mathrm{tr}(\mathcal{M}(\xi|\mathbf{u},\boldsymbol{\theta})^\delta)\right]^{1/\delta}. \tag{7}$$

The maximization of Φ_δ for $\delta \neq 0$ is equivalent to minimizing $\mathrm{tr}(\mathcal{M}(\xi|\mathbf{u},\boldsymbol{\theta})^\delta)$ when $\delta < 0$. Practically, Φ_δ becomes $[\mathrm{tr}(\mathcal{M}(\xi|\mathbf{u},\boldsymbol{\theta})^{-1})]^{-1}$ for $\delta = -1$, which is A–optimality, and $[\det[\mathcal{M}(\xi|\mathbf{u},\boldsymbol{\theta})]]^{1/n_\theta}$ when $\delta \to 0$, which is D–optimality. These design criteria are suitable for estimating model parameters as they maximize the FIM in various ways. For the D–optimality criterion the volume of the confidence region of $\boldsymbol{\theta}$ is proportional to $\det[\mathcal{M}^{-1/2}(\xi|\mathbf{u},\boldsymbol{\theta})]$. Then, maximizing the determinant of the FIM leads to the smallest possible volume. Consequently, the ODoE problem can be cast as an optimization problem. For example, when $\mathbf{p}$ is fixed, the locally D–optimal design is defined by

$$\xi_D = \arg\max_{\xi\in\Xi} \log\left\{\det[\mathcal{M}(\xi|\mathbf{u},\mathbf{p})]\right\}, \tag{8}$$

where the criterion (8) is $+\infty$ for designs with singular FIM. Herein we limit our analysis to D–optimal designs since these are the most commonly used in practical applications. Without loss of generality, the formulations proposed in the following sections can easily be extended to other criteria of the Kiefer's class as well as V– and I–optimality when interest is in prediction rather than parameter estimation.

When the design criterion is convex (which is the case for the D–optimality criteria originally formulated), the global optimality of a design ξ in $\mathbf{U}$ can be verified using an equivalence theorem based on the consideration of the directional derivative of the objective function (Fedorov, 1972; Kiefer, 1974; Kiefer and Wolfowitz, 1960; Pukelsheim, 1993; Silvey, 1980; Whittle, 1973). For instance, if we let δ_u be the degenerate design putting weight one at the point $\mathbf{u} \in \mathbf{U}$, the equivalence theorem for D–optimality is as follows: ξ_D is D–optimal if and only if

$$\mathrm{tr}\left\{[\mathcal{M}(\xi_D|\mathbf{u},\boldsymbol{\theta})]^{-1}\ M(\boldsymbol{\delta}_u)\right\} - n_\theta \leq 0, \quad \forall \mathbf{u} \in \mathbf{U}. \tag{9}$$

Herein, for convenience we reformulate of the D–optimality criterion as a maximization problem where the objective function is concave (Whittle, 1973). To compare the information content obtained from two different designs, say ξ_D and ξ_D^{ref}, where the latter one is the reference, we use the D–optimality efficiency (Atkinson et al., 2007, Chap. 11)

$$\mathrm{Eff}_D = \left\{\frac{\det[\mathcal{M}(\xi_D|\mathbf{u},\boldsymbol{\theta})]}{\det[\mathcal{M}(\xi_D^{\mathrm{ref}}|\mathbf{u},\boldsymbol{\theta})]}\right\}^{1/n_\theta}. \tag{10}$$

In determining model identifiability we use the FIM at convergence and determine the eigenvalues. The smallest eigenvalue, $\lambda_{\min}[\mathcal{M}(\xi_D|\mathbf{u},\boldsymbol{\theta})]$, is subsequently compared with the tolerance employed to solve the NLP problem.

Mathematical programming algorithms can currently solve complex, high–dimensional optimization problems, especially when they are convex and a self-concordant barrier is available for the constraints. Among the mathematical programming methods two strategies are commonly followed: (i) the design domain is discretized, when the optimal design problem reduces to weight optimization where convex programming can be used; and (ii) the optimal design problem involves optimizing the weights and location of the support points simultaneously where the design domain is continuous. The latter approaches require nonlinear programming based methods. Examples of applications of convex programming based algorithms for finding continuous optimal designs are Linear Programming (LP) (Gaivoronski, 1986; Harman and Jurík, 2008), Second Order Conic Programming (SOCP) (Sagnol, 2011; Sagnol and Harman, 2015a), and Semidefinite Programming (SDP) (Duarte and Wong, 2015; Papp, 2012; Vandenberghe and Boyd, 1999). Examples of applications requiring nonlinear solvers include: Semi Infinite Programming (SIP) (Duarte and Wong, 2014; Duarte et al., 2015), Nonlinear Programming (NLP) (Chaloner and Larntz, 1989; Molchanov and Zuyev, 2002), and Global optimization (Boer and Hendrix, 2000; Duarte et al., 2016). Yang et al. (2013) and Pronzato and Zhigljavsky (2014) consider the joint problem of weight optimization and choice of support points in a compact (continuous) set and propose specific methods with guaranteed convergence to the optimum. Applications based on optimization procedures relying on metaheuristic algorithms are also reported in the literature. See, among others, Heredia-Langner et al. (2004) for Genetic Algorithms, Woods (2010) for Simulated Annealing, Chen et al. (2015) for Particle Swarm Optimization and Masoudi et al. (2019) for the Imperialist Competitive Algorithm.

The proposed approach for solving the design problem (8) relies on nonlinear programming algorithms as we determine the weights and the support points (control actions) simultaneously. Our formulation leads to an optimization problem of the NLP class; since the problem may have multiple local optima, a global optimizer is used. The equations representing the model and the sensitivity construction are embedded in the optimal design problem as additional constraints. The same holds for matrix algebra operations required for computing D–optimality criteria. This strategy allows us to find optimal designs that satisfy the model equations and guarantees that all the solutions in the convergence process are feasible. For a detailed analysis of the formulation that allows the automation of determinant computation the reader is referred to Duarte et al. (2020).

2.2 Time-discrete state–space models

In this section we consider the solution of the model (1). We solve the model at a grid with intervals Δt. The sampling (and actuation) interval, $\Delta\tau$, is an integer multiple of Δt. In general the instants at which the system is sampled and control actions are applied form a coarser grid, the points of which coincide with some of the instants at which the variables are recalculated, see Figure 2 where a discretization scheme with $\Delta\tau = 2\,\Delta t$ is exemplified.

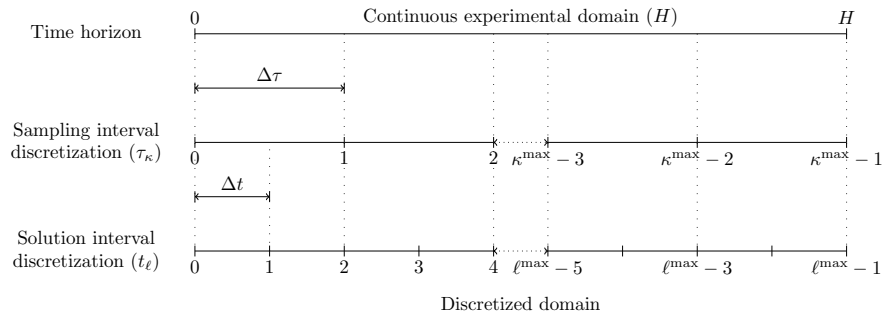


Figure 2 Discretization scheme for a sampling and actuation interval twice that of solution interval.

The solution considering a discrete time representation and successive step disturbances at control variables $\mathbf{u}$ is (Bay, 1999):

$$\mathbf{x}_{\ell+1} = \exp[A(\boldsymbol{\theta})\,\Delta t]\,\mathbf{x}_\ell + A(\boldsymbol{\theta})^{-1}\,\left\{\exp[A(\boldsymbol{\theta})\,\Delta t] - I_{n_x}\right\}\,B\,\mathbf{u}_\ell, \tag{11a}$$

$$\mathbf{y}_\ell = C\,\mathbf{x}_\ell + \boldsymbol{\epsilon}, \tag{11b}$$

$$\mathbf{x}_0 = \mathbf{x}_{\text{in}}, \tag{11c}$$

where $\ell \in \{0, \cdots, \ell^{\max} - 1\}$ is the counter of discrete time instants at which the system is to be updated, $\ell^{\max} - 1$ is its maximum number, which is previously given; the discretization time instants at intervals Δt are denoted by t_ℓ where $t_{\ell+1} = t_\ell + \Delta t$. Here, $\mathbf{x}_\ell$ is the vector of states and $\mathbf{y}_\ell$ the vector of measurement variables observed at t_ℓ, I_{n_x} is the identity matrix of size n_x and $\mathbf{x}_{\text{in}}$ the initial state of the system. The grid of sampling points $\kappa \in \{0, \cdots, \kappa^{\max} - 1\}$ is formed by $\tau_\kappa = \kappa \times \Delta\tau$ time instants at which the system is sampled and $\mathbf{u}_\kappa$, the vector of inputs, is optimally chosen (and implemented) to maximize the amount of information gathered from the complete experiment. In our conceptualization we distinguish between the discretization grid and the sampling (and actuation) grid. The former controls the accuracy of the predictions, and the latter is related to the amount of information gathered. The accuracy of estimated variables increases as $\Delta t \to 0$ and may have impact on the amount of information gathered. If the prediction error becomes large the discretization interval can be reduced without affecting the sampling grid. On the other hand, since the ODoE problem is reformulated as a dynamic optimization problem, the dual grid has the same purpose of using more nodes as when simultaneous approaches relying on orthogonal collocation are used (Hoang et al., 2013). It also corresponds to lower tolerances in error-based step adaptation techniques when sequential approaches are considered (Banga et al., 2002).

In this framework, the control actions are time dependent and can vary over the horizon of the experiment. After discretization we consider they are constant over discrete sampling intervals, changing only at their limits. Thus, $\mathbf{u}_\kappa$ designates the vector of actions to be implemented in sampling interval $(\kappa - 1)$. Within the dynamic experimental setup required by (11), we take the input variables to be constant

piecewise functions, i.e.

$$\mathbf{U} = \begin{cases} \mathbf{u}_0 & \text{if } t \in [\tau_0, \tau_1) \\ \cdots & \cdots \\ \mathbf{u}_\kappa & \text{if } t \in [\tau_\kappa, \tau_{\kappa+1}) \\ \cdots & \cdots \\ \mathbf{u}_{\kappa^{\max}-1} & \text{if } t \in [\tau_{\kappa^{\max}-1}, \tau_{\kappa^{\max}}), \end{cases}$$

with the steps occurring at the discrete time instants τ_κ. Here, we consider only single experiment setups and the optimal design is formed by κ tuples $\mathbf{u}_\kappa$, $\kappa \in \{0, \cdots, \kappa^{\max} - 1\}$,

$$\xi = \begin{pmatrix} \mathbf{u}_0^\top, & \cdots, & \mathbf{u}_\kappa^\top & \cdots, & \mathbf{u}_{\kappa^{\max}-1}^\top \\ [\tau_0,\ \tau_1), & \cdots, & [\tau_\kappa,\ \tau_{\kappa+1}), & \cdots, & [\tau_{\kappa^{\max}-1},\ \tau_{\kappa^{\max}}) \end{pmatrix}.$$

The optimal design aims at finding the optimal sequence of input levels that maximizes the information content of the complete experiment given a grid of sampling times.

The sensitivity matrix of the states with respect to $\boldsymbol{\theta}$ at time instant t_ℓ is denoted by $S_\ell^{\mathbf{x},\boldsymbol{\theta}} = \partial \mathbf{x}_\ell / \partial \boldsymbol{\theta}$ with $S_\ell^{\mathbf{x},\boldsymbol{\theta}} \in \mathbb{R}^{n_x \times n_\theta}$. Similarly, the sensitivity matrix of the measurement variables with respect to the parameters is $F(\mathbf{u}_\ell|\mathbf{p}) = C\ \partial \mathbf{x}/\partial \boldsymbol{\theta}|_{\mathbf{u}_\ell,\mathbf{p}}$. After algebraic manipulation they are:

$$\begin{aligned} S_{\ell+1}^{\mathbf{x},\boldsymbol{\theta}} =& \Delta t\ \exp[A(\boldsymbol{\theta})\ \Delta t]\ \frac{\partial A(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\ \mathbf{x}_\ell + \exp[A(\boldsymbol{\theta})\ \Delta t]\ S_\ell^{\mathbf{x},\boldsymbol{\theta}} + \\ &+ \Delta t\ A(\boldsymbol{\theta})^{-1}\ \exp[A(\boldsymbol{\theta})\ \Delta t]\ \frac{\partial A(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\ B\ \mathbf{u}_\ell - \\ &- A(\boldsymbol{\theta})^{-1}\ \frac{\partial A(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\ A(\boldsymbol{\theta})^{-1}\ \left\{\exp[A(\boldsymbol{\theta})\ \Delta t] - I_{n_x}\right\}\ B\ \mathbf{u}_\ell,\ \ell \geq 1, \end{aligned} \tag{12a}$$

$$F(\mathbf{u}_\ell|\boldsymbol{\theta}) = C\ S_\ell^{\mathbf{x},\boldsymbol{\theta}},\ \ell \in \{0, \cdots, \ell^{\max} - 1\} \tag{12b}$$

$$\mathbf{x}_0 = \mathbf{x}_{\text{in}}, \tag{12c}$$

$$S_0^{\mathbf{x},\boldsymbol{\theta}} = 0_{n_x \times n_\theta}. \tag{12d}$$

Here $S_0^{\mathbf{x},\boldsymbol{\theta}}$ is the matrix of sensitivities at t_0, and the exponential matrix $\exp[A(\boldsymbol{\theta})\ \Delta t]$ is computed via eigendecomposition, i.e. $\exp[A(\boldsymbol{\theta})\ \Delta t] = V(\boldsymbol{\theta})\ \exp[\Lambda(\boldsymbol{\theta})\ \Delta t]\ V(\boldsymbol{\theta})^{-1}$; $V(\boldsymbol{\theta})$ is the matrix of eigenvectors of $A(\boldsymbol{\theta})$, $V(\boldsymbol{\theta})^{-1}$ its inverse and $\Lambda(\boldsymbol{\theta})$ the corresponding diagonal matrix containing the eigenvalues. Similarly to Equation (11a), in (12a), $\mathbf{u}_\ell = \mathbf{u}_\kappa$ for all points of the discretization grid, t_ℓ, $\ell \in \{0, \cdots, \ell^{\max} - 1\}$, falling in the κth interval of the sampling grid. The sensitivities $F(\mathbf{u}_\ell|\boldsymbol{\theta})$ are updated at all discretization points, but only those obtained at sampling points $\kappa \in \{0, \cdots, \kappa^{\max} - 1\}$ are used for constructing the FIM. The term $\partial A(\boldsymbol{\theta})/\partial \boldsymbol{\theta}$ produces a three-dimensional tensor where each slice contains the derivatives $\partial A(\boldsymbol{\theta})/\partial \theta_i = E_i$, $i \in [\![n_\theta]\!]$.

2.3 Nonlinear Programming

In this section we introduce the fundamentals of NLP which are used to solve the design problem (8). Nonlinear Programming seeks to find the global optimum $\mathbf{x}$ of

a convex or nonconvex nonlinear function $f : \mathbf{X} \mapsto \mathbb{R}$ in a compact domain $\mathbf{X}$ with possibly nonlinear constraints. The general structure of NLP problems is:

$$\min_{\mathbf{x}\in\mathbf{X}} f(\mathbf{x}) \tag{13a}$$

$$\text{s.t. } \mathbf{g}(\mathbf{x}) \leq \mathbf{0} \tag{13b}$$

$$\mathbf{h}(\mathbf{x}) = \mathbf{0}, \tag{13c}$$

where (13b) represents a set of r_i inequalities, and (13c) represents a set of r_e equality constraints. The functions $f(\mathbf{x})$, $\mathbf{g}(\mathbf{x})$ and $\mathbf{h}(\mathbf{x})$ are twice differentiable. In our context, the variable $\mathbf{x} \in \mathbf{X}$ includes the location of the support points as well as the weights quantifying the relative effort required at each one. By construction $\mathbf{X}$ in (13a) is closed which is what we have for Ξ.

Nested and gradient projection methods are commonly used to solve NLP problems. Some examples are the General Reduced Gradient (GRG) (Drud, 1985, 1994) and the Trust-Region (Coleman and Li, 1994) algorithms. Other common methods are Sequential Quadratic Programming (SQP) (Gill et al., 2005) and the Interior-Point (IP) (Byrd et al., 1999). Ruszczyński (2006) provides an overview of NLP algorithms.

3 Finding D-optimal designs

In this section we describe the numerical procedure for finding D–optimal experimental designs for estimation of the parameters $\boldsymbol{\theta}$ involved in $A(\boldsymbol{\theta})$. First, in §3.1 we consider the steady-state model (2), and subsequently, in §3.2 the time-discrete state–space model (11) is addressed. Section 3.3 overviews the implementation details. For clarification, the first model is to be designated as the SS-LTI model, and the second as the TD-LTI model.

3.1 Locally D–optimal design for SS-LTI model

Here, we consider the locally D–optimal continuous design problem for the SS-LTI model for a given vector $\mathbf{p} \in \boldsymbol{\Theta}$. Let us recall that the design problem consists of finding the combination of inputs $\mathbf{u}$ and weights $\mathbf{w}$ maximizing a given criterion of the information extracted from a set of K experiments in the feasibility domain Ξ. The optimization problem for finding uniform (exact) D–optimal designs is similar to that used for continuous designs except for the weights which are set equal to $1/K$ and fixed.

Practically, this setup can be seen as a static experimental plan where the complete set of experiments is planned at one time, and the results at the end of the experiments characterizing the steady-states of the system serve to estimate the parameters. The experimental design is represented by

$$\xi_{SS-LTI} = \begin{pmatrix} \mathbf{u}_1^\top, \cdots, \mathbf{u}_K^\top \\ w_1, \cdots, w_K \end{pmatrix} \in \Xi, \tag{14}$$

where $\Xi = \mathbf{U}^K \times \Sigma$ and solves

$$\max_{\xi \in \Xi} \ \log\{\det[\mathcal{M}(\xi|\mathbf{u},\mathbf{p})]\} \tag{15a}$$

$$\text{s.t. Model (2)} \tag{15b}$$

$$\text{Sensitivity Equations (6)} \tag{15c}$$

$$\sum_{i=1}^{K} w_i = 1. \tag{15d}$$

Let each element of $\mathcal{M}(\xi|\mathbf{u},\mathbf{p})$ be designated as $m_{i,j},\ i,j \in [\![n_\theta]\!]$, and let $\mathcal{M}(\xi|\mathbf{u},\mathbf{p})$ aggregate the information of all the response variables at all the support points. Each element of the FIM is determined from (5) using the sensitivity matrices generated at the support points, see Fedorov (1971).

Maximizing $\log\{\det[\mathcal{M}(\xi|\mathbf{u},\mathbf{p})]\}$ is equivalent to maximizing $\det[\mathcal{M}(\xi|\mathbf{u},\mathbf{p})]$ which is a concave function of the FIM. The determinant of the FIM is calculated applying the Cholesky decomposition. Let $\log\{\det[\mathcal{M}(\xi|\mathbf{u},\mathbf{p})]\} = 2\sum_{i=1}^{n_\theta} \log(q_{i,i})$, where $q_{i,i},\ i \in [\![n_\theta]\!]$ are the diagonal element of the lower triangular matrix that results of the decomposition of the FIM denoted by $Q(\xi|\mathbf{u},\mathbf{p})$. Then, maximizing $\det[\mathcal{M}(\xi|\mathbf{u},\mathbf{p})]$ is equivalent to maximizing the sum of the logarithms of the diagonal elements of $Q(\xi|\mathbf{u},\mathbf{p})$. The NLP formulation used for determining locally D–optimal designs for the SS-LTI model is

$$\max_{\xi \in \Xi} \ 2 \sum_{i=1}^{n_\theta} \log(q_{i,i}) \tag{16a}$$

$$\text{s.t. } F_k = C\ A(\mathbf{p})^{-1} \left.\frac{\partial A(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right|_{\mathbf{p}} A(\mathbf{p})^{-1}\ B\ \mathbf{u}_k, \qquad k \in [\![K]\!] \tag{16b}$$

$$m_{i,j} = \sum_{k=1}^{K} w_k\ F_{k,i,j}^{\mathsf{T}}\ V^{-1}\ F_{k,i,j}, \qquad i,\ j \in [\![n_\theta]\!] \tag{16c}$$

$$m_{i,j} = \sum_{l=1}^{n_\theta} q_{i,l} q_{j,l}, \qquad i,\ j \in [\![n_\theta]\!],\ i \le j \tag{16d}$$

$$q_{i,i} \ge \zeta, \qquad i \in [\![n_\theta]\!] \tag{16e}$$

$$q_{i,j} = 0, \qquad i,\ j \in [\![n_\theta]\!],\ i < j, \tag{16f}$$

$$m_{i,i} \ge q_{i,j}^2 \qquad i,\ j \in [\![n_\theta]\!], \tag{16g}$$

$$\sum_{i=1}^{K} w_i = 1, \tag{16h}$$

where ζ is a small positive constant to ensure the semi-positiveness of the FIM. In all examples presented in the following sections we set $\zeta = 1 \times 10^{-5}$. Equation (16b) generates the matrices of sensitivity of the response variables with respect to $\boldsymbol{\theta}$ at the support points (which are saved in three-dimensional matrices F of size $K \times n_\theta \times n_\theta$). Equation (16c) is to construct the FIM, (16d) represents the Cholesky decomposition, (16e) is to guarantee the positiveness of the diagonal elements of $Q(\xi|\mathbf{u},\mathbf{p})$, and

(16f) ensures that $Q(\xi|\mathbf{u},\mathbf{p})$ is upper diagonal. Equation (16g) is a numerical stability condition imposed on the Cholesky factorization of positive semidefinite matrices (Golub and Van Loan, 2013, Theorem 4.2.8), and (16h) is to constrain the sum of weights.

It is noteworthy that the optimal design of static experiments (Problem 15) can be handled with convex techniques such as SOCP and SDP combined with adaptive strategies or cutting planes for refining the location of the support points. Typically, these techniques iterate between the solution of the problem for finding the optimal weights for a previously defined (or updated) grid of candidate points, and the improvement of the support points location using adaptive schemes as in Duarte et al. (2018) or cutting planes as in Pronzato and Pázman (2013). In both cases the convergence may require a large amount of CPU time, especially for models including several parameters. Instead, NLP can handle the optimal design problem at once and solve simultaneously for the weights and the locations of support points. The resulting optimization problem is NP-hard and may include several optima, thus requiring global optimizers that allow certifying the global optimality of the design. Since the optimal design problems are typically of small/medium scale, and the Jacobian and Hessian matrices can be generated analytically using automatic differentiation, NLP guarantees a good compromise between accuracy and numerical efficiency. Also, since the design of dynamic experiments (see §3.2) cannot be solved with convex techniques and we aim at formalizing a strategy that can easily adapt to static and dynamic experiments, NLP was chosen so it allows this generalization.

3.2 Locally D–optimal design for TD-LTI model

Here, we present the formulation for finding D–optimal designs for time-discrete state–space LTI model (11). The prescribed problem in the ODoE has the form of a *dynamic experiment* in which the input variables are changed along the horizon of the experiment so that the information obtained is maximized. The optimal design consists of the optimal set of combinations of $\mathbf{u}$ in each interval $[\kappa\,\Delta t,\ (\kappa+1)\,\Delta t)$, $\kappa \in \{0,\cdots,\kappa^{\max}-1\}$, i.e.,

$$\xi_{TD-LTI} = \begin{pmatrix} \mathbf{u}_0^{\mathsf{T}}, & \cdots, & \mathbf{u}_{\kappa^{\max}-1}^{\mathsf{T}} \\ [t_0,\ t_1), & \cdots, & [t_{\kappa^{\max}-1}, t_{\kappa^{\max}}) \end{pmatrix} \in \mathbf{U}^{\kappa^{\max}},$$

where $\mathbf{u}$ is formed by successive step jumps, $\kappa^{\max}$ is the number of sampling instants of the experiment, $t_0 = 0$ and the time horizon of the experiment is $H = \kappa^{\max} \times \Delta t$. The last action implemented occurs at $t = (\kappa^{\max}-1)\times\Delta t$ but the system is monitored during the horizon of the experiment, i.e., the responses are measured at $\kappa^{\max}$ time instants and the FIM matrix is so scaled. We note that the number of observations produced by an experiment is $\kappa^{\max} \times n_y$. The optimal design problem is as follows:

$$\max_{\xi\in\mathbf{U}^{\kappa^{\max}},\ \mathbf{x}\in\mathbf{X}} \ 2\sum_{i=1}^{n_\theta} \log(q_{i,i}) \tag{17a}$$

$$\text{s.t. } \mathbf{x}_{\ell+1} = \exp[A(\mathbf{p})\,\Delta t]\,\mathbf{x}_\ell + A(\mathbf{p})^{-1}\,\left\{\exp[A(\mathbf{p})\,\Delta t] - I_{n_x}\right\}\,B\,\mathbf{u}_\ell,$$

$$\ell \in [\![\ell^{\max} - 1]\!], \tag{17b}$$

$$\mathbf{y}_\ell = C\,\mathbf{x}_\ell, \quad \ell \in [\![\ell^{\max} - 1]\!] \tag{17c}$$

$$\mathbf{x}_0 = \mathbf{x}_{\text{in}} \tag{17d}$$

$$\begin{aligned} S^{\mathbf{x},\boldsymbol{\theta}}_{\ell+1} &= \Delta t\ \exp[A(\mathbf{p})\,\Delta t]\ \frac{\partial A(\mathbf{p})}{\partial \boldsymbol{\theta}}\,\mathbf{x}_\ell + \exp[A(\mathbf{p})\,\Delta t]\ S^{\mathbf{x},\boldsymbol{\theta}}_{\ell} + \\ &+ \Delta t\ A(\mathbf{p})^{-1}\ \exp[A(\mathbf{p})\,\Delta t]\ \frac{\partial A(\mathbf{p})}{\partial \boldsymbol{\theta}}\ B\,\mathbf{u}_\ell - \\ &- A(\mathbf{p})^{-1}\ \frac{\partial A(\mathbf{p})}{\partial \boldsymbol{\theta}}\ A(\mathbf{p})^{-1}\ \left\{\exp[A(\mathbf{p})\,\Delta t] - I_{n_x}\right\}\ B\,\mathbf{u}_\ell, \\ &\ell \in [\![\ell^{\max} - 1]\!] \end{aligned} \tag{17e}$$

$$F_\ell = C\,S^{\mathbf{x},\boldsymbol{\theta}}_{\ell}, \quad \ell \in \{0, \cdots, \ell^{\max} - 1\} \tag{17f}$$

$$S^{\mathbf{x},\boldsymbol{\theta}}_{0} = 0_{n_x \times n_\theta} \tag{17g}$$

$$\mathbf{u}_\ell = \mathbf{u}_\kappa, \quad \ell \in \{\ell :\ t_\ell \in [\tau_\kappa,\ \tau_{\kappa+1})\},\ \kappa \in \{0, \cdots, \kappa^{\max} - 1\} \tag{17h}$$

$$m_{i,j} = \sum_{\kappa=0}^{\kappa^{\max}-1} F^{\mathsf{T}}_{k,i,j}\,V^{-1}\,F_{k,i,j}, \quad i,\ j \in [\![n_\theta]\!] \tag{17i}$$

$$\text{Equations (16}d - 16g\text{)} \tag{17j}$$

Equation (17b) is for the prediction of state variables, (17c) for measurement prediction and (17d) is the initialization of the state variables. Equation (17e) is for the computation of the sensitivities of the state variables, (17f) the sensitivities of the measurements wrt $\boldsymbol{\theta}$, and (17g) is the initialization of the sensitivities. Finally, (17h) sets the control actions used for updating state variables and sensitivities at discretization time instants to the values prescribed for κ^{th} interval of the sampling grid, (17i) is to construct the FIM.

3.3 Implementation aspects

Here we detail the implementation aspects related to the numerical approach for solving the optimal design problem.

Formulations (16) and (17) are coded in `The General Algebraic Modeling System` environment, commonly known by the initials `GAMS` (GAMS Development Corporation, 2013). `GAMS` is a general modeling system that supports mathematical programming applications in several areas. Upon execution, the code describing the mathematical program is automatically compiled, symbolically transcribed into a set of numerical structures, and all information regarding the gradient and matrix Hessian is generated using the automatic differentiation tool and made available to the solver. We provide a sample of such a code in the Supplementary Material.

The convexity properties of ODoE problems are rather challenging. The calculations require matrix algebra operations embedded in the optimization problems which in turn produce problems with multinomial terms and variables of different scales, that may, in principle, lead to multiple local optima. We did not encounter such problems in our numerical examples.

To determine the optimal design (i.e., solve the ODoE problems) we used a multistart heuristic algorithm-based solver, `OQNLP`. The algorithm calls an NLP solver from multiple starting points, keeps all the feasible solutions found, and picks the best as the optimal solution of the problem (Ugray et al., 2005). The starting points are computed with a random sampling driver that uses independent normal random variables for initializing each decision variable. Contrarily to deterministic global optimization solvers, `OQNLP` does not certify that the final solution is a global optimum, but it has been successfully tested on a large set of global optimization problems. To build the initial sampling points the variables need to be bounded, which is what we have since the design space and the region of plausible values are all compact by assumption. The NLP solver called by `OQNLP` is `CONOPT`, which in turn uses the Generalized Reduced Gradient (GRG) algorithm (Drud, 1985).

The maximum number of starting points allowed is set to 5000 and the procedure terminates when 100 consecutive NLP solver calls result in an improvement less than 1×10^{-4}. The absolute and relative tolerances of the solver were set equal to 1×10^{-5} and 1×10^{-6}, respectively, with the absolute tolerance being equal to ζ which is the minimum value allowed for the diagonal entries in the FIM so that it is positive definite. All computations in §4 used an Intel Core i7 machine running a 64 bits Windows 10 operating system with a 2.80 GHz processor.

4 Locally optimal designs

This section presents the locally D–optimal designs calculated for steady-state and time-discrete LTI models employing the formulations derived in §3. In Section 4.1 we consider the steady-state LTI model, and in §4.2 the time-discrete LTI model is solved and optimal designs for dynamic experiments are obtained. For demonstration we consider the biochemical reaction network of Figure 1, and the corresponding state–space representations which involve 10 kinetic rates to be estimated.

4.1 Locally optimal designs for SS-LTI model

In this section we consider steady-state LTI models and analyze (i) the impact of $\mathbf{p}$ for which the design is to be obtained; (ii) the comparison of exact and continuous optimal designs for the same vectors of parameters; and (iii) the impact of constraining the individual elements in $\mathbf{u}_k$ to smaller allowable maximum levels. We recall that this setup is adopted for static experiments, where the optimal vectors of actions (support points) found give the conditions for running different trials. In each trial, the system is observed after achieving the (new) steady-state. We take the vectors $\mathbf{u}_k,\ k \in [\![K]\!]$, to be constrained to the unit simplex set, i.e. $\{\mathbf{u}_k \in \mathbb{R}^3,\ k \in [\![K]\!] :\ \mathbf{1}^\intercal\ \mathbf{u}_k = 1\}$, which represents the limitation of activating no more than one entering flux in each trial; $\mathbf{1}$ is the unitary vector of size 3. The exception occurs when the number of support points is restricted to 2; the optimal design may require fractionally activating more than one flux at once to assure the FIM is non-singular. This constraint is explicitly included in the design space, $\mathbf{U}$, which may include additional bounds.

The first numerical cases reported in Tables 1-4, were solved limiting the upper bound of control actions to 1, corresponding to fully activating a single flux. This constraint assures bounded optimization problems since the determinant of the FIM depends on combinations of square vectors $\mathbf{u}$, thus being naturally unbounded if $\mathbf{u} \to \infty$. Here, $\mathbf{U} = \{\mathbf{u}_k \in \mathbb{R}^3,\ k \in [\![K]\!] :\ \mathbf{1}^\intercal\ \mathbf{u}_k = 1,\ \mathbf{u}_k \leq 1\}$. The results in Tables 5-6 were obtained with an upper bound of 0.5 imposed to control actions which corresponds to fractionally activating the fluxes with upper limits of 50 %. Now, we have $\mathbf{U} = \{\mathbf{u}_k \in \mathbb{R}^3,\ k \in [\![K]\!] :\ \mathbf{1}^\intercal\ \mathbf{u}_k = 1,\ \mathbf{u}_k \leq 0.5\}$.

To study the impact of $\mathbf{p}$ on the optimal design and model identifiability, we consider two distinct scenarios where (i) the parameters $\theta_i,\ i \in [\![n_\theta]\!]$ are all equal to 1, i.e. $\mathbf{p}_1 = \mathbf{1}_{n_\theta}^\intercal$; and (ii) the parameters $\theta_i,\ i \in [\![n_\theta]\!]$ are chosen using a uniform random generator in interval [0.5, 1.5]; the main purpose of this numerical test is generalizing the application to every $\boldsymbol{\theta} \in \boldsymbol{\Theta}$. Specifically, the vector obtained for the later scenario is $\mathbf{p}_2 =$ [1.3147, 1.4057, 0.6269, 1.4133, 1.1323, 0.5975, 0.7784, 1.0468, 1.4575, 1.4648].

We recall that the formulations developed in §3 are for continuous designs. To determine exact optimal designs we solve the optimal design problem (16) after setting $w_k = 1/n_s,\ k \in [\![n_s]\!]$, i.e. the optimization problem aims at finding the support points maximizing the optimality criterion given that each one of them have equal weights $1/n_s$. Obviously, for some values of n_s, the exact optimal design may include replicates corresponding to support points with weights integer multiples of $1/n_s$. The corresponding continuous designs are determined relaxing the weights w_k and solving the problem (16) for both $\mathbf{u}_k,\ k \in [\![n_s]\!]$, and w_k. The continuous optimal designs will be at least as efficient as the equivalent exact optimal designs as they do not have to obey the constraint $n_k \in \mathbb{N}$. To evaluate the efficiency of the exact designs obtained for different values of n_s (number of support points) we calculate the D–efficiency with (10). Here, the continuous designs obtained for $n_s = 3$ in Tables 3 and 4 are used as reference for evaluating the efficiency of the exact designs in Tables 1 and 2 as well as that of the continuous designs obtained from $n_s = 2$, also displayed in 3 and 4.

To assess model identifiability we use the value of the minimum eigenvalue of the FIM at convergence and compare it with the tolerance imposed on the NLP solver, ζ. If $\lambda_{\min}[\mathcal{M}(\xi_D|\mathbf{u},\boldsymbol{\theta})] < \zeta$ the model is assumed unidentifiable ($Id = 0$) otherwise it is locally identifiable with the identifier $Id = 1$. The designs presented in Tables 1–4 were obtained for $\mathbf{u}_k \in [0,\ 1]^3,\ k \in [\![n_s]\!]$ where the minimum value (0) corresponds to inactive entering fluxes and (1) to active fluxes. For compactness we use $\{\det[\mathcal{M}]\}^{1/n_\theta}$ for representing $\{\det[\mathcal{M}(\xi|\mathbf{u},\boldsymbol{\theta})]\}^{1/n_\theta}$ and $\lambda_{\min}[\mathcal{M}]$ for $\lambda_{\min}[\mathcal{M}(\xi_D|\mathbf{u},\boldsymbol{\theta})]$. The Tables also report the CPU time required to solve the ODoE problem with `OQNLP` which is relatively high because of the need of ensuring that global optimality is attained.

Table 1 presents the exact optimal designs obtained for $\mathbf{p}_1$. To help in the interpretation of the results, each of columns 2 to 4 in the table is for one support point, with the respective weights listed below for $n_s \in \{3, \cdots, 6\}$. These extreme support points correspond to activation of just one flux. In contrast, the optimal design for $n_s = 2$ in the last line of Table 1 contains one support point which is not extreme; the second support point indicates activation of a linear combination of inputs u_2 and u_3.

The efficiency of the exact designs in Table 1, computed using the continuous design for $n_s = 3$ in Table 3 as reference, is close to 100 %, and very similar for $n_s \in \{3, \cdots, 6\}$. However, two patterns are discernible. The design for $n_s = 6$ twice replicates the design for $n_s = 3$, so the two designs have identical properties since the ξ are identical. The highest determinant is from $n_s = 5$, which is a close approximation to the optimal continuous design in Table 3. The determinant of the FIM for $n_s = 2$ is one order of magnitude below the values obtained for other values of n_s, with the efficiency being about 80 %. Considering the criterion $\lambda_{\min}[\mathcal{M}(\xi_D|\mathbf{u}, \boldsymbol{\theta})]$, we notice the model is identifiable for $n_s \in \{2, \cdots, 6\}$.

The exact optimal design obtained for $n_s = 2$ includes a support point where the levels of u_2 and u_3 are fractional; a linear combination of both allows exciting the complete network as u_2 and u_3 excite different parts (i.e., u_2 excites the fluxes $x_2 \rightarrow x_3$, $x_3 \rightarrow$ output, $x_2 \rightarrow x_5$, $x_5 \rightarrow x_6$ and $x_6 \rightarrow$ output while u_3 excites $x_4 \rightarrow x_5$, $x_5 \rightarrow x_6$ and $x_6 \rightarrow$ output). The activation of u_1 excites the full network but only provides 6 measurements, insufficient to identify the model.

Table 2 contains the exact optimal designs for a different parameter vector $\mathbf{p}_2$. Although the parameters are different, the designs are the same as those in Table 1. The model is again identifiable for $n_s \in \{2, \cdots, 6\}$, and the behavior of $\det[\mathcal{M}(\xi|\mathbf{u}, \boldsymbol{\theta})]$ and $\lambda_{\min}[\mathcal{M}(\xi_D|\mathbf{u}, \boldsymbol{\theta})]$ with the number of support points is similar to that observed for the previous paremeter vector. Although the model is identifiable for both of the vectors $\mathbf{p}$ tested, model identifiability is pointwise in the space of parameters, and no general conclusion can be extended to the global identifiability of the model. The efficiency of the exact designs computed using the continuous designs in Table 4 for reference is again very close to 100 %, except for $n_s = 2$.

The diagnosis of global identifiability of the state–space model describing the network in Figure 1 is out of the scope of this paper as it requires symbolic algebra-based approaches, see Saccomani et al. (1997). The lower eigenvalue of the FIM at convergence can only be used for checking local identifiability; there are many models that are locally identifiable despite being globally non-identifiable (Guillaume et al., 2019). However, to check the global identifiability of the space-state model representing the network we ran the analysis in `DAISY`, a software tool developed for testing the global identifiability of state–space models (Bellu et al., 2007). Practically, we tested the model (1) representing the network in Figure 1, and the result is that it is globally identifiable. Thus, we expect local identifiability holds for all parameter vectors used for finding optimal experimental designs.

Table 3 presents the continuous optimal designs obtained for $\mathbf{p}_1$. Only the designs obtained for $n_s \in \{2,\ 3\}$ are listed' since the numerical experiments for $K > 3$ produce designs identical to that for $n_s = 3$ after collapsing some support points. The D–optimal designs obtained for a single output ($n_y = 1$) and $n_s = n_\theta$ allocate equal weights $w_i = 1/n_\theta$ to each support point. We note the designs in Table 3 do not follow this rule because $n_y > 1$.

Practically, D–optimal designs based on two or more support points allow full identification of the model; designs with more than 3 support points collapse into the 3 support point design. As for the exact designs, the continuous optimal designs for $n_s = 2$ include a support point formed by a fractional combination of u_2 and u_3; the weights of the support are unequal but close to $1/2$. Table 4 shows the optimal designs

Table 1 Steady state model: exact optimal experimental designs for $\mathbf{p}_1$ ($\mathbf{U} = \{\mathbf{u}_k \in \mathbb{R}^3,\ k \in [\![K]\!] :\ \mathbf{1}^\top \mathbf{u}_k = 1,\ \mathbf{u}_k \leq 1\}$).

	Support points							
n_s	$\begin{pmatrix}1.0000\\0.0000\\0.0000\end{pmatrix}$	$\begin{pmatrix}0.0000\\1.0000\\0.0000\end{pmatrix}$	$\begin{pmatrix}0.0000\\0.0000\\1.0000\end{pmatrix}$	$\{\det[\mathcal{M}]\}^{1/n_\theta}$	$\lambda_{\min}[\mathcal{M}]$	Id	Eff_D (%)	CPU (s)
3	1/3	1/3	1/3	0.2176	4.9911×10^{-3}	1	99.10	28.31
4	2/4	1/4	1/4	0.2157	3.7783×10^{-3}	1	98.25	30.69
5	2/5	2/5	1/5	0.2192	5.0072×10^{-3}	1	99.80	35.15
6	2/6	2/6	2/6	0.2176	4.9911×10^{-3}	1	99.10	31.65
	Support points							
	$\begin{pmatrix}1.0000\\0.0000\\0.0000\end{pmatrix}$	$\begin{pmatrix}0.0000\\0.8383\\0.1617\end{pmatrix}$						
2	1/2	1/2		0.1746	2.1388×10^{-3}	1	79.50	24.13

Table 2 Steady state model: exact optimal experimental designs for $\mathbf{p}_2$ ($\mathbf{U} = \{\mathbf{u}_k \in \mathbb{R}^3,\ k \in [\![K]\!] :\ \mathbf{1}^\top \mathbf{u}_k = 1,\ \mathbf{u}_k \leq 1\}$).

	Support points							
n_s	$\begin{pmatrix}1.0000\\0.0000\\0.0000\end{pmatrix}$	$\begin{pmatrix}0.0000\\1.0000\\0.0000\end{pmatrix}$	$\begin{pmatrix}0.0000\\0.0000\\1.0000\end{pmatrix}$	$\{\det[\mathcal{M}]\}^{1/n_\theta}$	$\lambda_{\min}[\mathcal{M}]$	Id	Eff_D (%)	CPU (s)
3	1/3	1/3	1/3	0.1853	3.4247×10^{-3}	1	99.34	34.88
4	2/4	1/4	1/4	0.1837	2.5821×10^{-3}	1	98.50	40.57
5	2/5	2/5	1/5	0.1859	3.3626×10^{-3}	1	99.67	39.20
6	2/6	2/6	2/6	0.1853	3.4247×10^{-3}	1	99.34	33.85
	Support points							
	$\begin{pmatrix}1.0000\\0.0000\\0.0000\end{pmatrix}$	$\begin{pmatrix}0.0000\\0.8474\\0.1526\end{pmatrix}$						
2	1/2	1/2		0.1437	1.2233×10^{-3}	1	77.02	20.34

for $\mathbf{p}_2$. Although the designs are not identical to those of Table 3, we observe similar trends to those found for $\mathbf{p}_1$.

Now we consider a constrained design region with $\mathbf{U} = \{\mathbf{u} \in \mathbb{R}^3 :\ 0 \leq \mathbf{u} \leq 0.5\}$. The exact optimal designs obtained for $\mathbf{p}_1$ are in Table 5, and the corresponding continuous optimal designs in Table 6. The efficiency of the exact designs relative to the continuous design for $n_s = 3$ is close to 100 %. The comparison of the determinant values with those for the designs in Tables 1 and 2 reveals that the constraint on input variables decreases the efficiency by about 30 %. In Tables 1–2 the optimal designs take measurements when each input in turn equals one, with the other two zero. The effect of the constraint on the values of the u_k in the designs of Tables 5–6 is to have two inputs at the maximum for each design point and one at zero. The weights for the continuous optimal design in Table 6 are less equal than those for the continuous design in Table 3, with the effect in Table 5 that 3 trials are at support point 1 when $n_s = 6$.

Table 3 Steady state model: continuous optimal experimental designs for $\mathbf{p}_1$ ($\mathbf{U} = \{\mathbf{u}_k \in \mathbb{R}^3,\ k \in [\![K]\!] :\ \mathbf{1}^\top \mathbf{u}_k = 1,\ \mathbf{u}_k \leq 1\}$).

n_s	Support points $\begin{pmatrix}1.0000\\0.0000\\0.0000\end{pmatrix}$	$\begin{pmatrix}0.0000\\1.0000\\0.0000\end{pmatrix}$	$\begin{pmatrix}0.0000\\0.0000\\1.0000\end{pmatrix}$	$\{\det[\mathcal{M}]\}^{1/n_\theta}$	$\lambda_{\min}[\mathcal{M}]$	Id	Eff_D (%)	CPU (s)
3	0.3887	0.3757	0.2356	0.2196	4.9949×10^{-3}	1	100.00	78.42
	Support points $\begin{pmatrix}1.0000\\0.0000\\0.0000\end{pmatrix}$	$\begin{pmatrix}0.0000\\0.8381\\0.1619\end{pmatrix}$						
2	0.5143	0.4857		0.1746	2.1233×10^{-3}	1	79.51	44.18

Table 4 Steady state model: continuous optimal experimental designs for $\mathbf{p}_2$ ($\mathbf{U} = \{\mathbf{u}_k \in \mathbb{R}^3,\ k \in [\![K]\!] :\ \mathbf{1}^\top \mathbf{u}_k = 1,\ \mathbf{u}_k \leq 1\}$).

n_s	Support points $\begin{pmatrix}1.0000\\0.0000\\0.0000\end{pmatrix}$	$\begin{pmatrix}0.0000\\1.0000\\0.0000\end{pmatrix}$	$\begin{pmatrix}0.0000\\0.0000\\1.0000\end{pmatrix}$	$\{\det[\mathcal{M}]\}^{1/n_\theta}$	$\lambda_{\min}[\mathcal{M}]$	Id	Eff_D (%)	CPU (s)
3	0.3836	0.3649	0.2515	0.1865	3.3758×10^{-3}	1	100.00	93.21
	Support points $\begin{pmatrix}1.0000\\0.0000\\0.0000\end{pmatrix}$	$\begin{pmatrix}0.0000\\0.8471\\0.1529\end{pmatrix}$						
2	0.5188	0.4812		0.1437	1.2128×10^{-3}	1	77.05	40.21

Table 5 Steady state model: exact optimal experimental designs for $\mathbf{p}_1$ ($\mathbf{U} = \{\mathbf{u}_k \in \mathbb{R}^3,\ k \in [\![K]\!] :\ \mathbf{1}^\top \mathbf{u}_k = 1,\ \mathbf{u}_k \leq 0.5\}$).

n_s	Support points $\begin{pmatrix}0.5000\\0.5000\\0.0000\end{pmatrix}$	$\begin{pmatrix}0.5000\\0.0000\\0.5000\end{pmatrix}$	$\begin{pmatrix}0.0000\\0.5000\\0.5000\end{pmatrix}$	$\{\det[\mathcal{M}]\}^{1/n_\theta}$	$\lambda_{\min}[\mathcal{M}]$	Id	Eff_D (%)	CPU (s)
2	1/2	1/2	0	0.1516	9.2594×10^{-4}	1	98.23	21.35
3	1/3	1/3	1/3	0.1511	1.3115×10^{-3}	1	97.92	39.15
4	2/4	1/4	1/4	0.1522	1.3212×10^{-3}	1	98.69	42.38
5	2/5	2/5	1/5	0.1536	1.5323×10^{-3}	1	99.54	45.68
6	3/6	2/6	1/6	0.1541	1.4961×10^{-3}	1	99.87	37.06

The convergence of the global optimizer ensures the global optimality of all the designs obtained in subsequent sections. Nonetheless, the optimality of the static designs was checked graphically by plotting the dispersion function and validating the equivalence theorem. Here, for demonstration purposes we consider the continuous D–optimal design with three support points obtained for vector $\mathbf{p}_2$ (in Table 4), and compute the directional derivative (9). The display is shown in Figure 3; the dispersion function is bounded from above by zero and is maximized at the support points, so the design is indeed locally D-optimal. We note the design problem has three decision

Table 6 Steady state model: continuous optimal experimental designs for $\mathbf{p}_1$ ($\mathbf{U} = \{\mathbf{u}_k \in \mathbb{R}^3,\ k \in [\![K]\!] :\ \mathbf{1}^\top \mathbf{u}_k = 1,\ \mathbf{u}_k \leq 0.5\}$).

	Support points							
n_s	$\begin{pmatrix}0.5000\\0.5000\\0.0000\end{pmatrix}$	$\begin{pmatrix}0.5000\\0.0000\\0.5000\end{pmatrix}$	$\begin{pmatrix}0.0000\\0.5000\\0.5000\end{pmatrix}$	$\{\det[\mathcal{M}]\}^{1/n_\theta}$	$\lambda_{\min}[\mathcal{M}]$	Id	Eff_D (%)	CPU (s)
2	0.5502	0.4498	0.0000	0.1519	8.6006×10^{-4}	1	98.47	32.18
3	0.4864	0.3757	0.1379	0.1543	1.4393×10^{-3}	1	100.00	46.40

variables but because of the constraint $\sum_{k=1}^{3} u_k = 1$ it reduces to variables u_1 and u_2, and the directional derivate can be represented in a three-dimensional plot. Similar plots were constructed for all designs obtained for static experiments and all satisfy the optimality conditions.

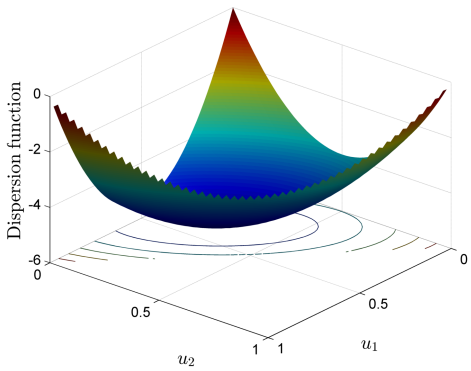


Figure 3 Directional derivative (9) of the continuous D–optimal design for $\mathbf{p}_2$ assuming 3 support points (see Table 4).

4.2 Locally optimal designs for TD-LTI model

We now consider optimal design for time-discrete LTI models and solve the optimization problem (17). This setup is adopted for dynamic experiments, where the optimal sequence of actions found serve to run a single experiment. The sequence of actuations prescribed as well as the process sampling are implemented at a previously defined grid of discrete time instants of the experimental horizon. We set $\kappa^{\max}$ to 6 and varied the interval $\Delta\tau$ between sampling (and control) instants. In all dynamic experiments considered in this section, the factor domain is $\mathbf{U} = \{\mathbf{u} \in \mathbb{R}^{\kappa^{\max}} :\ \sum_{k=1}^{3} u_k = 1,\ \mathbf{u} \geq 0\}$. The impact of $\mathbf{p}$ on local model identifiability is also assessed. For testing the formulation we used the parameter vectors considered in §4.1, i.e. $\mathbf{p}_1$ and $\mathbf{p}_2$. The optimal designs are presented as the set of values of the control factors $u_i,\ i \in [\![n_y]\!]$, at

each discrete time instant $\kappa \in \{0, \cdots, \kappa^{\max} - 1\}$. The state and measurement (algebraic) equations (11) and the respective sensitivity equations (12a-12b) are solved for $\ell \in \{0, \cdots, \ell^{\max} - 1\}$ at once. Initially, at time $t = 0$ (corresponding to discrete time t_0) the network is considered deactivated, consequently $\mathbf{x}_{\text{in}} = \mathbf{0}$.

The sampling interval $\Delta\tau$ should be lower than the Shannon period for the input signal (Franklin et al., 1990). That is, $\Delta\tau$ should be smaller than $1/2$ of the time constant associated with the slowest system dynamics. For vector $\mathbf{p}_1$, this is governed by the eigenvalue $\lambda = -1$. Thus, we consider $\Delta\tau = 1/3$ and subsequently reduce it to $1/6$ to analyze the impact of the sampling interval reduction on the optimal design. The time interval Δt at which the model predictions and sensitivities are updated is $\Delta t = 1/6$ in all the numerical experiments, which is the minimum value of $\Delta\tau$ tested. Consequently, in these experiments we set $\ell^{\max} = 12$ which allows comparing optimal designs independently of the grid at which the variables are recalculated. We thus disaggregate the influence on the amount of information gathered of the integration interval from that of the sampling interval and the horizon of the experiment.

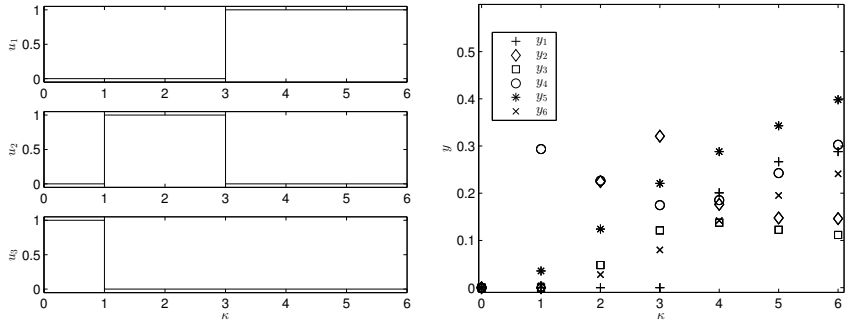
Table 7 presents the optimal designs for $\Delta\tau = 1/3$; we observe the inputs take extreme values corresponding to activation/deactivation of inflow fluxes. It is well known that D–optimal designs for linear models choose experimental points at the extremes of the design region. The same is found by Zarrop (1979) for D–optimal designs for linear control systems, resulting in persistent excitation in which at least one control is non-zero. Here, the optimal profiles of actions likewise require extreme variations of the input, going suddenly from lower to upper bounds or vice-versa. The two optimal designs are identical, having one input at its maximum and the others at the zero, the input at the maximum changing over the horizon of the experiment.

We found that the matrix of eigenvectors, $V(\boldsymbol{\theta})$, for the first vector of parameters ($\mathbf{p}_1$) is non-invertible, so $V^{-1}(\boldsymbol{\theta})$ cannot be computed. To overcome this issue the matrix exponential $\exp[A(\mathbf{p})\,\Delta t]$ is computed using the power series method with 30 terms, i.e. $\exp[A(\mathbf{p})\,\Delta t] = \sum_{i=0}^{29}[A(\mathbf{p})\,\Delta t]^i/i!$.

The optimum design for $\Delta\tau = 1/3$ and $\mathbf{p}_2$ is plotted in Figure 4 where 4(a) is for the sequence of control actions and 4(b) for measurement predictions. Figure 4(a) highlights the "bang-bang" form of actions when the goal is to maximize the information content gathered from dynamic experiments.

Table 7 Discrete-time dynamic model: D–optimal experimental designs ($\Delta\tau = 1/3,\ H = 2$).

	Values of $\mathbf{u}$					
	$\mathbf{p}_1$			$\mathbf{p}_2$		
κ	u_1	u_2	u_3	u_1	u_2	u_3
0	0.0000	0.0000	1.0000	0.0000	0.0000	1.0000
1	1.0000	0.0000	0.0000	0.0000	1.0000	0.0000
2	1.0000	0.0000	0.0000	0.0000	1.0000	0.0000
3	0.0000	1.0000	0.0000	1.0000	0.0000	0.0000
4	0.0000	1.0000	0.0000	1.0000	0.0000	0.0000
5	0.0000	1.0000	0.0000	1.0000	0.0000	0.0000
$\{\det[\mathcal{M}]\}^{1/n_\theta}$	0.2002			0.1779		
$\lambda_{\min}[\mathcal{M}]$	2.8527×10^{-4}			1.5119×10^{-4}		
CPU (s)	167.86			221.51		



(a) (b)

Figure 4 Optimal experimental design for TD-LTI model; $\Delta\tau = 1/3$ and $\mathbf{p}_2$: (a) control factors; (b) measurements.

Table 8 shows the optimal designs obtained for the same parameter vectors by reducing the horizon and considering a smaller sampling interval, $\Delta\tau = 1/6$ such that the number of observations obtained is equal to that of the previous setup. Here, $\Delta\tau = \Delta t$; consequently, $\ell^{\max} = 6$. Practically, we analyze the effect of reducing the sampling interval keeping the number of observations produced from the experimental plan fixed as well as the interval at which variables are updated. The model is still identifiable for both parameter vectors since the minimum eigenvalue of the FIM at convergence is larger than 1×10^{-5}. The efficiencies of the designs in Table 7 ($H = 2$ and $\Delta\tau = 1/3$) relative to those of Table 8 ($H = 1$ and $\Delta\tau = 1/6$) are 94.57 % and 97.97 % for vectors $\mathbf{p}_1$ and $\mathbf{p}_2$, respectively. We note that the experiments with smaller sampling interval are slightly more efficient when the number of sampling points is fixed, although the difference is small. This finding is due to appreciable information at the beginning of the experiment, as the responses change relatively rapidly, which is partially lost if $\Delta\tau$ is too large. A further advantage is the economic one that a shorter experiment is cheaper to run. Nevertheless, these findings are dependent on the model and sampling interval, and cannot be generalized.

Table 8 Discrete-time dynamic model: D–optimal experimental design ($\Delta\tau = 1/6,\ H = 1$).

	Values of $\mathbf{u}$					
	$\mathbf{p}_1$			$\mathbf{p}_2$		
κ	u_1	u_2	u_3	u_1	u_2	u_3
0	1.0000	0.0000	0.0000	1.0000	0.0000	0.0000
1	1.0000	0.0000	0.0000	1.0000	0.0000	0.0000
2	0.0000	0.0000	1.0000	0.0000	0.0000	1.0000
3	0.0000	1.0000	0.0000	0.0000	1.0000	0.0000
4	0.0000	1.0000	0.0000	0.0000	1.0000	0.0000
5	0.0000	1.0000	0.0000	0.0000	1.0000	0.0000
$\{\det[\mathcal{M}]\}^{1/n_\theta}$	0.2117			0.1816		
$\lambda_{\min}[\mathcal{M}]$	1.8816×10^{-4}			3.6568×10^{-5}		
CPU (s)	191.31			280.64		

To analyze the impact of varying $\Delta\tau$ but conserving the horizon we solved the problem for $\Delta\tau = \Delta t = 1/6$ and $H = 2$ with the two sets of parameter values of Table 7. Here, we analyze the effect of doubling the number of observations. Now $\kappa^{\max} = \ell^{\max} = 12$ and the optimal designs obtained are in Table 9. The comparison of the results of Table 9 with those of Table 7 shows an increase in information when the horizon is maintained and the sampling interval is reduced to one half. This effect is stronger for the parameter value in the right-hand half of the table.

Table 9 Discrete-time dynamic model: D–optimal experimental designs ($\Delta\tau = 1/6,\ H = 2$).

	Values of $\mathbf{u}$					
	$\mathbf{p}_1$			$\mathbf{p}_2$		
κ	u_1	u_2	u_3	u_1	u_2	u_3
0	0.0000	0.0000	1.0000	0.0000	0.0000	1.0000
1	0.0000	0.0000	1.0000	0.0000	0.0000	1.0000
2	1.0000	0.0000	0.0000	0.0000	0.0000	1.0000
3	1.0000	0.0000	0.0000	0.0000	1.0000	0.0000
4	1.0000	0.0000	0.0000	0.0000	1.0000	0.0000
5	1.0000	0.0000	0.0000	0.0000	1.0000	0.0000
6	0.0000	1.0000	0.0000	0.0000	1.0000	0.0000
7	0.0000	1.0000	0.0000	1.0000	0.0000	0.0000
8	0.0000	1.0000	0.0000	1.0000	0.0000	0.0000
9	0.0000	1.0000	0.0000	1.0000	0.0000	0.0000
10	0.0000	1.0000	0.0000	1.0000	0.0000	0.0000
11	0.0000	1.0000	0.0000	1.0000	0.0000	0.0000
$\{\det[\mathcal{M}]\}^{1/n_\theta}$	0.3040			0.2754		
$\lambda_{\min}[\mathcal{M}]$	6.7658×10^{-4}			4.2541×10^{-4}		
CPU (s)	558.28			740.26		

To compare the amount of information gathered from the experimental plans obtained for (i) $\Delta\tau = 1/3,\ H = 2$; (ii) $\Delta\tau = 1/6,\ H = 1$; and (iii) $\Delta\tau = 1/6,\ H = 2$, we compute the D–optimal efficiency of the two former designs using Eq. (10) with the reference design used for computation being that obtained for setup (iii). The efficiency of the reference design is omitted from Table 10 since it is 100 % by assumption. The D–optimal efficiencies are in Table 10, and is noteworthy that the number of observations of the plan increases the amount of information. The reference design involves 12 observations and the other two only 6. We note the D–efficiency of the latter two is around 65 %. The consequence is that it is more efficient to repeat independent observations with multiple sensors in setups (i) and (ii) than to use (iii).

Table 10 Discrete-time dynamic model: D–optimal efficiency obtained for parameter vectors $\mathbf{p}_1$ and $\mathbf{p}_2$ (expressed in percentage). The reference design is that obtained for $\Delta\tau = 1/6,\ H = 2$ (see Table 9).

Setup	$\mathbf{p}_1$	$\mathbf{p}_2$
$\Delta\tau = 1/3,\ H = 2$	65.86	64.62
$\Delta\tau = 1/6,\ H = 1$	69.64	65.96

Now, we compare the volume of the parametric confidence regions estimated from static experimental setup with $\mathbf{U} = \{\mathbf{u}_k \in \mathbb{R}^3,\ k \in [\![K]\!] :\ \mathbf{1}^\intercal\ \mathbf{u}_k = 1,\ \mathbf{u}_k \leq 1\}$ (see Tables 1-4) and dynamic setup with $\Delta\tau = 1/6,\ H = 1$ (see Table 8) for plans producing the same number of measurements using the metric $\{\det[\mathcal{M}(\xi|\mathbf{u},\boldsymbol{\theta})]\}^{1/n_\theta}$. The aim is to determine if any of the approaches have clear advantages. Specifically, we compared static plans obtained for $n_s = 6$ with the dynamic plan in Table 8. The dynamic experiment allows obtaining 97.29 % and 98.00 % of the information gathered from the exact static design for $\mathbf{p}_1$ and $\mathbf{p}_2$, respectively, and 96.40 % and 97.37 % of the information produced by equivalent continuous designs. Consequently, the confidence region obtained from static experimental design is slightly smaller but it requires 6 trials while the dynamic experimental plan requires only one setting of the control factors, although the system is, of course, sampled at 6 time instants.

We conclude with a simulation to illustrate the effect of the proposed approach in accurately estimating the parameters θ. We could, as in the preceding paragraph, compare the values of determinants of information matrices for two designs, but we instead provide simulated confidence intervals for the values of the individual parameters. We take $H = 2, \Delta\tau = 1/3$ with $\theta = \mathbf{p}_2$. The precision of the estimated $\hat{\theta}$ is computed by simulation for the profile of actions resulting from the optimal dynamic design and for an alternative (non-optimal) design. To construct the reference scenario we used the design obtained from solving the ODoE problem. Then, the response variables were simulated and corrupted with observational error normally distributed where each component ϵ_i is described by an i.i.d. Gaussian probability distribution $\mathcal{N}(0, \sigma_i)$. Here, we set $\sigma_i = 2 \times 10^{-3}$ for all the components.

Each simulated sample provides a total of $6{\times}n_y$ measurements used to estimate the model parameters with a multiresponse least squares (MLS) method. This problem is formulated as a NLP problem that minimizes $\sum_{i=1}^{n_y} \sum_{j=1}^{\kappa^{\max}} (\eta_{i,j}^{\text{obs}} - \eta_{i,j})\, V^{-1}\, (\eta_{i,j}^{\text{obs}} - \eta_{i,j})^\intercal$, and is solved for each simulated sample.

To construct non-optimal profiles of actions we randomly sample a 6-element vector from the set of integers $\{1,\ 2,\ 3\}$. Then, a 6×3 matrix of zeros is constructed and the elements with row indices equal to the vector of integers previously generated are set to 1. Next, this profile allows estimating the response variables, which are then corrupted with noise with the same characteristics applied to reference scenario, and used to estimate the model parameters with MLS. This procedure is also repeated 500 times. The complete set of parameter estimates allows computing the average and 95 % confidence intervals. The results are displayed in Table 11. Column 2 shows the true values used for generating the data, column 3 the estimates obtained with the profile of actions prescribed by the optimal design, and in the last column are the estimates obtained with the non-optimal sequence of actions. The estimates resulting from the optimal design are closer to the true values and, more importantly, for most of the parameters the confidence intervals are tighter. Consequently, the size of the parametric confidence regions for the individual parameters found from data obtained with the optimal profile of actions is smaller, which is the primary objective of ODoE.

The convergence of the global optimizer indicates the global optimality of all the designs obtained in this section. A full certification requires using a spatial branch and bound algorithm.

Table 11 Simulation-based analysis for $\Delta\tau = 1/3$, $H = 2$, $\mathbf{p}_2$ and 500 simulated samples. (in the representation $x.xxxx \pm y.yyyy$ the first number indicates the mean and the second the 95 % confidence limits)

Parameter	True value	Value obtained from reference simulated samples	Value obtained from non-optimal simulated samples
θ_1	1.3147	1.3165±0.0699	1.3610±0.0748
θ_2	1.4057	1.4058±0.0410	1.3963±0.1054
θ_3	0.6269	0.6231±0.0682	0.6116±0.1241
θ_4	1.4133	1.4141±0.0485	1.3918±0.1290
θ_5	1.1323	1.1331±0.0653	1.2744±0.1118
θ_6	0.5975	0.6045±0.1163	0.6285±0.1835
θ_7	0.7784	0.7777±0.0169	0.7897±0.0419
θ_8	1.0468	1.0456±0.0382	1.0658±0.0674
θ_9	1.4575	1.4557±0.0660	1.4382±0.0396
θ_{10}	1.4648	1.4667±0.0551	1.4535±0.0555

5 Conclusions

We have considered the optimal design of experiments for Linear Time Invariant State-Space models, and have proposed general formulations for finding locally optimal designs for steady-state and time-discrete representations. Here, the formulations were applied to the identification of biochemical reaction networks; the analysis of the Fisher Information Matrix at convergence provides a check on the local identifiability of the model through comparison of the minimum eigenvalue of the FIM with a previously set threshold value.

The static experimental designs for the SS-LTI model demonstrate only a very slight loss in D–efficiency from small exact designs compared with the optimal continuous designs. The actual number of experiments should be chosen to reflect the required level of accuracy in the estimates of the parameters, or derived confidence intervals. The dynamic optimal experimental designs for the TD-LTI model show that the sampling interval is a crucial factor in design efficiency; however, the optimal sampling interval depends on the specific system under analysis.

Our formulation addresses the D–optimality criterion and includes: (i) the generation of the sensitivity coefficients; and (ii) the computation of the determinant of the FIM. This approach can be applied to any criterion formulated as the maximization/minimization of a convex function of the FIM such as those of the Kiefer (1974) class. The first step requires solving/expanding the sensitivity equations derived from the chain rule of differentiation, and the second allows optimizing concave/convex functions of the FIM, such as the determinant. The resulting optimization problem for approximate and exact designs can have multiple local optima, so a global optimizer is required to ensure that a global optimum is attained.

Acknowledgements We wish to thank to the Associate Editor and two anonymous reviewers for their many comments and suggestions that helped to improve the quality of the paper.

References

Anderson, J., Chang, Y.C., Papachristodoulou, A.: Model decomposition and reduction tools for large-scale networks in systems biology. Automatica **47**(6), 1165–1174 (2011). DOI https://doi.org/10.1016/j.automatica.2011.03.010

Asprey, S., Macchietto, S.: Statistical tools for optimal dynamic model building. Computers & Chemical Engineering **24**(2), 1261 – 1267 (2000)

Atkinson, A.C., Donev, A.N., Tobias, R.D.: Optimum Experimental Designs, with SAS. Oxford University Press, Oxford (2007)

Banga, J.R., Versyck, K.J., Van Impe, J.F.: Computation of optimal identification experiments for nonlinear dynamic process models: a stochastic global optimization approach. Industrial Engineering Chemistry Research **41**, 2425–2430 (2002)

Barz, T., López Cárdenas, D.C., Arellano-Garcia, H., Wozny, G.: Experimental evaluation of an approach to online redesign of experiments for parameter determination. AIChE Journal **59**(6), 1981–1995 (2013)

Bay, J.: Fundamentals of Linear State Space Systems. WCB/McGraw-Hill (1999)

Bellu, G., Saccomani, M.P., Audoly, S., D'Angiò, L.: DAISY: A new software tool to test global identifiability of biological and physiological systems. Comput. Methods Programs Biomed. **88**(1), 52–61 (2007)

Boer, E., Hendrix, E.: Global optimization problems in optimal design of experiments in regression models. Journal Global Optimization **18**, 385–398 (2000)

Bouvin, J., Cajot, S., D'Huys, P.J., Ampofo-Asiama, J., Anné, J., Van Impe, J., Geeraerd, A., Bernaerts, K.: Multi-objective experimental design for 13C-based metabolic flux analysis. Mathematical Biosciences **268**, 22 – 30 (2015)

Brown, M., He, F., Yeung, L.F.: Robust measurement selection for biochemical pathway experimental design. Int J Bioinform Res Appl. **4**(4), 400–416 (2008)

Bryson, A.: Dynamic Optimization. Bibliyografya Ve Indeks. Addison Wesley Longman (1999)

Byrd, R.H., Hribar, M.E., Nocedal, J.: An interior point algorithm for large-scale nonlinear programming. SIAM J. on Optimization **9**(4), 877–900 (1999)

Chaloner, K., Larntz, K.: Optimal Bayesian design applied to logistic regression experiments. Journal of Statistical Planning and Inference **59**, 191–208 (1989)

Chen, R.B., Chang, S.P., Wang, W., Tung, H.C., Wong, W.K.: Minimax optimal designs via particle swarm optimization methods. Statistics and Computing **25**(5), 975–988 (2015)

Chis, O.T., Villaverde, A.F., Banga, J.R., Balsa-Canto, E.: On the relationship between sloppiness and identifiability. Mathematical Biosciences **282**, 147 – 161 (2016)

Cobelli, C., Thomaseth, K.: Optimal input design for identification of compartmental models. Theory and application to a model of glucose kinetics. Mathematical Biosciences **77**(1), 267 – 286 (1985)

Coleman, T.F., Li, Y.: On the convergence of reflective Newton methods for large-scale nonlinear minimization subject to bounds. Mathematical Programming **67**(2), 189–224 (1994)

Crampin, E.J., Schnell, S., McSharry, P.E.: Mathematical and computational techniques to deduce complex biochemical reaction mechanisms. Progress in Biophysics and Molecular Biology **86**(1), 77 – 112 (2004)

Draper, N.R., Hunter, W.G.: Design of experiments for parameter estimation in multiresponse situations. Biometrika **53**(3/4), 525–533 (1966)

Drud, A.: `CONOPT`: A GRG code for large sparse dynamic nonlinear optimization problems. Mathematical Programming **31**, 153–191 (1985)

Drud, A.: `CONOPT` - A large–scale GRG code. ORSA Journal on Computing **6**(2), 207–216 (1994)

Duarte, B.P.M., Granjo, J.F.O., Wong, W.K.: Optimal exact designs of experiments via Mixed Integer Nonlinear Programming. Statistics and Computing **30**, 93–112 (2020)

Duarte, B.P.M., Wong, W.K.: A semi-infinite programming based algorithm for finding minimax optimal designs for nonlinear models. Statistics and Computing **24**(6), 1063–1080 (2014)

Duarte, B.P.M., Wong, W.K.: Finding Bayesian optimal designs for nonlinear models: A semidefinite programming-based approach. International Statistical Review **83**(2), 239–262 (2015)

Duarte, B.P.M., Wong, W.K., Atkinson, A.C.: A semi-infinite programming based algorithm for determining $T-$optimum designs for model discrimination. Journal of Multivariate Analysis **135**, 11 – 24 (2015)

Duarte, B.P.M., Wong, W.K., Dette, H.: Adaptive grid semidefinite programming for finding optimal designs. Statistics and Computing **28**(2), 441–460 (2018)

Duarte, B.P.M., Wong, W.K., Oliveira, N.M.C.: Model-based optimal design of experiments - Semidefinite and nonlinear programming formulations. Chemometrics and Intelligent Laboratory Systems **151**, 153 – 163 (2016)

Eisenberg, M.C., Hayashi, M.A.: Determining identifiable parameter combinations using subset profiling. Mathematical Biosciences **256**, 116 – 126 (2014)

Espie, D., Macchietto, S.: The optimal design of dynamic experiments. AIChE Journal **35**(2), 223–229 (1989)

Fedorov, V.V.: The design of experiments in the multiresponse case. Theory of Probability and its Applications **16**, 323 – 332 (1971)

Fedorov, V.V.: Theory of Optimal Experiments. Academic Press (1972)

Fedorov, V.V., Leonov, S.L.: Optimal Design for Nonlinear Response Models. Chapman and Hall/CRC Press, Boca Raton (2014)

Franklin, G.F., Powell, J.D., Workman, M.L.: Digital Control of Dynamic Systems. 3rd edn. Addison-Wesley, Reading, MA (1990)

Frøysa, H.G., Skaug, H.J., Alendal, G.: Experimental design for parameter estimation in steady-state linear models of metabolic networks. Mathematical Biosciences **319**, 108,291 (2020)

Gaivoronski, A.: Linearization methods for optimization of functionals which depend on probability measures. In: Prékopa, A., Wets, R.J.B. (eds.) Stochastic Programming 84 Part II, *Mathematical Programming Studies*, vol. 28, pp. 157–181. Springer Berlin Heidelberg (1986)

Galvanin, F., Barolo, M., Pannocchia, G., Bezzo, F.: Online model-based redesign of experiments with erratic models: A disturbance estimation approach. Computers & Chemical Engineering **42**, 138 – 151 (2012). European Symposium of Computer Aided Process Engineering - 21

Galvanin, F., Boschiero, A., Barolo, M., Bezzo, F.: Model-based design of experiments in the presence of continuous measurement systems. Industrial & Engineering Chemistry Research **50**(4), 2167–2175 (2011)

GAMS Development Corporation: GAMS - A User's Guide, GAMS Release 24.2.1. GAMS Development Corporation, Washington, DC, USA (2013)

Gill, P.E., Murray, W., Saunders, M.A.: SNOPT: An SQP algorithm for large-scale constrained optimization. SIAM Rev. **47**(1), 99–131 (2005)

Golub, G., Van Loan, C.: Matrix Computations. Johns Hopkins Studies in the Mathematical Sciences. Johns Hopkins University Press (2013)

Goodwin, G.C., Payne, R.L.: Dynamic System Identification, Experiment Design and Data Analysis. Academic Press, New York (1977)

Guillaume, J.H., Jakeman, J.D., Marsili-Libelli, S., Asher, M., Brunner, P., Croke, B., Hill, M.C., Jakeman, A.J., Keesman, K.J., Razavi, S., Stigter, J.D.: Introductory overview of identifiability analysis: A guide to evaluating whether you have the right type of data for your modeling purpose. Environmental Modelling & Software **119**, 418 – 432 (2019)

Hangos, K.M., Szederkényi, G., Alonso, A.A.: Reaction kinetic form for lumped process system models. IFAC Proceedings Volumes **46**(14), 48 – 53 (2013). 1st IFAC Workshop on Thermodynamic Foundations of Mathematical Systems Theory

Harman, R., Jurík, T.: Computing $c-$optimal experimental designs using the Simplex method of linear programming. Comput. Stat. Data Anal. **53**(2), 247–254 (2008)

Heredia-Langner, A., Montgomery, D.C., Carlyle, W.M., Borror, C.M.: Model-robust optimal designs: A Genetic Algorithm approach. Journal of Quality Technology **36**, 263–279 (2004)

Hoang, M.D., Barz, T., Merchan, V.A., Biegler, L.T., Arellano-Garcia, H.: Simultaneous solution approach to model-based experimental design. AIChE Journal **59**(11), 4169–4183 (2013)

Kalaba, R., Spingarn, K.: Control, Identification, and Input Optimization. Springer US (1982)

Kalaba, R.E., Spingarn, K.: Optimal inputs and sensitivities for parameter estimation. Journal of Optimization Theory and Applications **11**, 56–67 (1973). DOI https://doi.org/10.1007/BF00934291

Kiefer, J.: General equivalence theory for optimum design (approximate theory). Annals of Statistics **2**, 849–879 (1974)

Kiefer, J., Wolfowitz, J.: The equivalence of two extremum problem. Canadian Journal of Mathematics **12**, 363–366 (1960)

Körkel, S., Kostina, E., Bock, H.G., Schlöder, J.P.: Numerical methods for optimal control problems in design of robust optimal experiments for nonlinear dynamic processes. Optimization Methods and Software **19**(3-4), 327–338 (2004)

Ljung, L.: System Identification: Theory for the User. 2nd. edn. Prentice-Hall, Inc., USA (1999)

Loskot, P., Atitey, K., Mihaylova, L.: Comprehensive review of models and methods for inferences in bio-chemical reaction networks. Frontiers in Genetics **10**, 549 (2019). DOI 10.3389/fgene.2019.00549

Maidens, J., Arcak, M.: Semidefinite relaxations in optimal experiment design with application to substrate injection for hyperpolarized MRI. In: 2016 American

Control Conference (ACC), pp. 2023–2028 (2016)
Masoudi, E., Holling, H., Duarte, B.P.M., Wong, W.K.: A metaheuristic adaptive cubature based algorithm to find Bayesian optimal designs for nonlinear models. Journal of Computational and Graphical Statistics **0**(0), 1–16 (2019)
Mehra, R.: Optimal input signals for parameter estimation in dynamic systems–Survey and new results. IEEE Transactions on Automatic Control **19**(6), 753–768 (1974)
Molchanov, I., Zuyev, S.: Steepest descent algorithm in a space of measures. Statistics and Computing **12**, 115–123 (2002)
Ng, T.S., Goodwin, G.C.: On optimal choice of sampling strategies for linear system identification. International Journal of Control **23**(4), 459–475 (1976)
Papp, D.: Optimal designs for rational function regression. Journal of the American Statistical Association **107**, 400–411 (2012)
Perry, M., Wynn, H., Bates, R.: Principal components analysis in sensitivity studies of dynamic systems. Probabilistic Engineering Mechanics **21**(4), 454 – 460 (2006)
Pronzato, L.: Optimal experimental design and some related control problems. Automatica **44**, 303–325 (2008)
Pronzato, L., Pázman, A.: Algorithms: A survey. In: Design of Experiments in Nonlinear Models: Asymptotic Normality, Optimality Criteria and Small-Sample Properties, pp. 277–333. Springer New York, New York, NY (2013). DOI 10.1007/978-1-4614-6363-4_9
Pronzato, L., Zhigljavsky, A.A.: Algorithmic construction of optimal designs on compact sets for concave and differentiable criteria. Journal of Statistical Planning and Inference **154**, 141 – 155 (2014)
Pukelsheim, F.: Optimal Design of Experiments. SIAM, Philadelphia (1993)
Pukelsheim, F., Rieder, S.: Efficient rounding of approximate designs. Biometrika **79**, 763–770 (1992)
Rudolph, P.E., Herrendörfer, G.: Optimal experimental design and accuracy of parameter estimation for nonlinear regression models used in long-term selection. Biometrical Journal **37**(2), 183–190 (1995). DOI https://doi.org/10.1002/bimj.4710370209
Ruszczyński, A.: Nonlinear Optimization. Princeton University Press, New Jersey, USA (2006)
Saccomani, M.P., Audoly, S., Bellu, G., D'Angiò, L., Cobelli, C.: Global identifiability of nonlinear model parameters. IFAC Proceedings Volumes **30**(11), 233 – 238 (1997). IFAC Symposium on System Identification (SYSID'97), Kitakyushu, Fukuoka, Japan, 8-11 July 1997
Sagnol, G.: Plans d'expériences optimaux et application à l'estimation des matrices de trafic dans les grands réseaux. Ph.D. thesis, L'École Nationale Supérieure des Mines de Paris (2010)
Sagnol, G.: Computing optimal designs of multiresponse experiments reduces to second-order cone programming. Journal of Statistical Planning and Inference **141**(5), 1684–1708 (2011)
Sagnol, G., Harman, R.: Computing exact $D-$optimal designs by mixed integer second order cone programming. Annals of Statistics **43**(5), 2198–2224 (2015a)
Sagnol, G., Harman, R.: Optimal designs for steady-state Kalman filters. In: Steland, A., Rafajłowicz, E., Szajowski, K. (eds.) Stochastic Models, Statistics and Their

Applications, pp. 149–157. Springer International Publishing, Cham (2015b)
van der Schaft, A.J., Rao, S., Jayawardhana, B.: A network dynamics approach to chemical reaction networks. International Journal of Control **89**(4), 731–745 (2016)
Silvey, S.D.: Optimal Design. Chapman & Hall, London (1980)
Singhal, H., Michailidis, G.: Optimal experiment design in a filtering context with application to sampled network data. Ann. Appl. Stat. **4**(1), 78–93 (2010)
Telen, D., Logist, F., Quirynen, R., Houska, B., Diehl, M., Van Impe, J.: Optimal experiment design for nonlinear dynamic (bio)chemical systems using sequential semidefinite programming. AIChE Journal **60**(5), 1728–1739 (2014)
Telen, D., Nimmegeers, P., Impe, J.V.: Uncertainty in optimal experiment design: comparing an online versus offline approaches. IFAC-PapersOnLine **51**(2), 771 – 776 (2018). 9th Vienna International Conference on Mathematical Modelling
Titterington, D.M.: Aspects of optimal design in dynamic systems. Technometrics **22**(3), 287–299 (1980)
Ugray, Z., Lasdon, L., Plummer, J., Glover, F., Kelly, J., Martí, R.: A multistart scatter search heuristic for smooth nlp and minlp problems. In: Metaheuristic Optimization via Memory and Evolution, pp. 25–51. Springer (2005)
Vandenberghe, L., Boyd, S.: Applications of semidefinite programming. Applied Numerical Mathematics **29**, 283–299 (1999)
Varma, A., Palsson, B.O.: Metabolic flux balancing: Basic concepts, scientific and practical use. Bio/Technology **12**(10), 994–998 (1994). DOI 10.1038/nbt1094-994
Villaverde, A.F.: Observability and structural identifiability of nonlinear biological systems. Complexity **Article ID 8497093**, 1 – 12 (2019)
Walter, E.: Identifiability of State Space Models: with applications to transformation systems. Lecture Notes in Biomathematics. Springer Berlin Heidelberg (2013)
Walter, E., Pronzato, L.: How to design experiments that are robust to parameter uncertainty. IFAC Proceedings Volumes **18**(5), 921 – 926 (1985)
Walter, E., Pronzato, L.: Qualitative and quantitative experiment design for nonlinear models. IFAC Proceedings Volumes **21**(1), 69 – 80 (1988)
Whittle, P.: Some general points in the theory of optimal experimental design. Journal of the Royal Statistical Society, Ser. B **35**, 123–130 (1973)
Wiechert, W., Möllney, M., Petersen, S., de Graaf, A.A.: A universal framework for 13C metabolic flux analysis. Metabolic Engineering **3**, 265 – 283 (2001)
Woods, D.C.: Robust designs for binary data: applications of simulated annealing. Journal of Statistical Computation and Simulation **80**(1), 29–41 (2010)
Yang, M., Biedermann, S., Tang, E.: On optimal designs for nonlinear models: A general and efficient algorithm. Journal of the American Statistical Association **108**(504), 1411–1420 (2013)
Zarrop, M.B.: Optimal Experimental Design for Dynamic System Identification: Lecture Notes in Control and Information Sciences 21. Springer-Verlag, New York (1979)