Chemometrics and Intelligent Laboratory Systems 239 (2023) 104874

Contents lists available at ScienceDirect

Chemometrics and Intelligent Laboratory Systems

journal homepage: www.elsevier.com/locate/chemometrics

Optimum design for ill-conditioned models: K–optimality and stable parameterizations

Belmiro P.M. Duarte [a,b,*], Anthony C. Atkinson [c], Nuno M.C. Oliveira [b]

[a] Polytechnic Institute of Coimbra, ISEC, Department of Chemical & Biological Engineering, Rua Pedro Nunes, 3030–199 Coimbra, Portugal
[b] University of Coimbra, CIEPQPF, Department of Chemical Engineering, Rua Sílvio Lima — Pólo II, 3030–790 Coimbra, Portugal
[c] Department of Statistics, London School of Economics, London WC2A 2AE, United Kingdom

ARTICLE INFO

Keywords:
K-optimal design of experiments
Semidefinite programming
Nonlinear regression
Model reparameterization
Support points

ABSTRACT

Nonlinear regression is frequently used to fit nonlinear relations between response variables and regressors, for process data. The procedure involves the minimization of the square norm of the residuals with respect to the model parameters. Nonlinear least squares may lead to parametric collinearity, multiple optima and computational inefficiency. One of the strategies to handle collinearity is model reparameterization, i.e. the replacement of the original set of parameters by another with increased orthogonality properties. In this paper we propose a systematic strategy for model reparameterization based on the response surface generated from a carefully chosen set of points. This is illustrated with the support points of locally K-optimal experimental designs, to generate a set of analytical equations that allow the construction of a transformation to a set of parameters with better orthogonality properties. Recognizing the difficulties in the generalization of the technique to complex models, we propose a related alternative approach based on first-order Taylor approximation of the model. Our approach is tested both with linear and nonlinear models. The Variance Inflation Factor and the condition number as well as the orientation and eccentricity of the parametric confidence region are used for comparisons.

1. Motivation

Multiple linear regression is a widely used and, statistically, well understood method for modelling a response as a function of a set of predictor variables. Nonlinear least squares is much used in the pharmaceutical and other process industries to fit responses to the nonlinear models that frequently arise from chemical kinetics. Although the asymptotic properties of nonlinear least squares are also well understood, there remain problems, especially numerical, in the analysis of finite data sets. These include (i) the difficulty in achieving the minimization of the LS when the covariance matrix is ill-conditioned; and (ii) strong parameter collinearity found for some data and models. This last drawback is commonly related with model identifiability issues and may induce difficulties concerning model discrimination and the uncertainty (or even the accuracy) of the estimated parameters. Under these conditions, slight non-convergence of the minimizing algorithm can compound the difficulties in model building. Seber and Wild [1] provides a systematic summary of difficulties including plots of the narrow twisting valleys of some least squares surfaces. López et al. [2] provides a survey of more recent progress.

Least Squares (LS) minimizes the sum of square errors between observations and model predictions and implicitly relies on a set of assumptions: (i) the errors are identically and independently distributed (iid);(ii) the variance is constant; and (iii) the regressors are independent. Common techniques for model fitting are described in standard text books (see Seber and Wild [1] among others), and there is a broad range of computational tools supporting their use [3,4].

Several techniques have been proposed for improving the robustness of model fitting procedures in such situations. Among them are (i) variable transformation [5,6] for dependent or independent variables and the Box–Cox transformations of non-negative responses [7]; (ii) variable separation [8]; and (iii) model reparameterization [9–11]. Recently, Atkinson et al. [12] extended the Box–Cox transformation to responses that can be positive or negative; they can also be used for variable transformation.

Our paper focuses on model reparameterizations which replace the original set of parameters of the model by another set less correlated (leaving, whenever possible, all parameters mutually uncorrelated) via a transformation specific to the model. For models which are nonlinear

* Corresponding author at: Polytechnic Institute of Coimbra, ISEC, Department of Chemical & Biological Engineering, Rua Pedro Nunes, 3030–199 Coimbra, Portugal.
E-mail addresses: bduarte@isec.pt (B.P.M. Duarte), A.C.Atkinson@lse.ac.uk (A.C. Atkinson), nuno@eq.uc.pt (N.M.C. Oliveira).

https://doi.org/10.1016/j.chemolab.2023.104874
Received 1 March 2023; Received in revised form 20 May 2023; Accepted 23 May 2023
Available online 27 May 2023

in the parameters, these transformations use both *a priori* and *a posteriori* information, derived from the original model and the experimental design, including an estimate of the final values of the parameters, in the case of nonlinear models (i.e., assuming a local design).

Typically, *a priori* methods use (i) moments of the data distribution (see Schwaab and Pinto [13]); (ii) orthogonal polynomial analogies; and (iii) a family of analytical curves constructed through a set of judiciously chosen points in the domain of the regressors [14]. This important third strategy has never been totally exploited as it can be challenging to automate computationally; it requires a set of points chosen from the domain of the regressors (in Ross et al. [14] they are generally equally spaced) and a subsequent algebraic treatment of the response predictions at these points. This task also becomes somewhat specific for each atomic function that occurs in the model (e.g., exponential, logarithm, or polynomial functions). On the other hand, for some atomic functions, including the transcendental functions, this treatment can only be directly implemented approximately, leading to transformations for which the set of parameters is not completely orthogonal.

This paper also relies on the strategy of using a family of curves that pass through specific points for model reparameterization. The construction of these curves can be automated through the use of symbolic computational tools to generate the corresponding orthogonal transformations. A simpler alternative is further considered, based on the manipulation of a first-order Taylor approximation with respect to the parameters. In both cases, instead of choosing a set of discretionary domain points we reuse the support points of optimal designs (namely K-optimal designs), obtained via Semidefinite Programming (SDP), for a discretized design space. The K-optimal designs minimize the condition number of the Fisher Information Matrix (FIM); thus, the support points of these designs are obvious candidates for generating transformations that lead to a set of orthogonal or nearly orthogonal parameters, while simultaneously minimizing the condition of the FIM. A helpful property of K-optimal designs in our context is that they can reduce the Variance Inflation Factor and the error sensitivity for the least squares estimator for the original set of parameters [15]. The designs that we find are locally optimal, since they are determined assuming reference values for the parameters. However, as is illustrated in Section 5, the generated transformation is robust enough to cope with parameters different from those initially postulated for constructing the locally optimal design.

It should be stressed that the models with known untransformed parameters give identical relationships between response and regressors to the models with the transformation of these values. This relationship can be destroyed due to poor convergence of model fitting in the presence of high correlation. Our purpose is to reduce the collinearity of the parameter estimates together with the occurrence of multiple optima and to improve the computational efficiency. The parameter estimates from the two models will then give virtually identical response predictions.

1.1. Novelty statement and organization

This paper contains four elements of novelty: (i) a computationally automated approach for model reparameterization based on the use of families of curves that pass through a given set of points of the response surface; (ii) a Semidefinite Programming based tool to construct K-optimal designs, the support points of which are used to construct the analytic curves; (iii) a first-order Taylor based approximation for systematic handling of more complex models; and (iv) the application of the algorithm to several problems that are considered benchmark tests in model reparameterization.

The paper is organized as follows. Section 2 introduces the background and the notation used to formulate the model fitting problem as well as the fundamentals of Semidefinite Programming used to compute the support points and the nonlinear program (NLP) used to solve the nonlinear least squares problem. Section 3 introduces the various steps required by the tool and its test, specifically (i) the construction of support points to generate transformations; (ii) the construction of the transformations; (iii) the generation of data to test the parameterized model using other alphabetic optimal experimental designs; (iv) the fitting of the parameterized model via LS using previously generated data; and (v) the calculation of numerical (and graphical) figures of merit to compare the effects of reparameterization. Comparisons for different univariate response models are presented in Section 4. In Section 5 the robustness of the parameterized model is analysed with respect to (i) the "real" parameters used to generate the data; and (ii) the design space. Section 6 analyses the proposed tool in depth focusing on its ability to find orthogonal spaces of parameters and the effectiveness to handle ill-conditioned LS problems. Section 7 reviews the formulation and offers a summary of the results obtained.

2. Notation and background

In our notation bold face lowercase letters represent vectors, bold face capital letters stand for continuous domains, blackboard bold capital letters are used to denote discrete domains and capital letters are adopted for matrices. Finite sets containing $\imath$ elements are compactly represented by $[\![\imath]\!] \equiv \{1, \ldots, \imath\}$. The transpose operation of a matrix or vector is represented by "$^\intercal$". The cardinality of a vector is represented by $\mathrm{card}(\bullet)$ and the trace of a matrix by $\mathrm{tr}(\bullet)$.

Next, we introduce the fundamentals of our conceptualization. Then, the optimization tools used herein are briefly covered; in Section 2.1 the basics of SDP are introduced, and Section 2.2 offers the fundamentals about Nonlinear Programming used to solve the LS problems.

We consider a general univariate nonlinear model with the form

$$y = f(x, \boldsymbol{\theta}) + \epsilon, \tag{1}$$

where y is the response, x is a regression factor, $f(\bullet)$ a twice differentiable function, $\boldsymbol{\theta}$ the vector of parameters and $\epsilon \approx \mathcal{N}(0, \sigma)$ is the observational noise which has zero mean and standard deviation σ. Let the number of parameters of the model be n_θ, each of them being constrained to a compact domain, i.e. $\theta_i \in [\theta_i^L, \theta_i^U]$, where θ_i^L is the lower bound for parameter i, and θ_i^U the upper bound. Consequently, the Cartesian domain containing all the combinations of parameters is $\boldsymbol{\Theta} \equiv \bigotimes_{i=1}^{n_\theta} [\theta_i^L, \theta_i^U] \subset \mathbb{R}^{n_\theta}$. To distinguish between the generic vector $\boldsymbol{\theta}$ and a singleton vector of $\boldsymbol{\Theta}$, the latter is designated by $\mathbf{p}_0$. In turn, $x \in \mathbf{X} \subset \mathbb{R}$ where $\mathbf{X}$ is the closed domain containing all the possible values of the regressors.

For comparison, a standard univariate linear model

$$y' = X\boldsymbol{\theta}' + \epsilon \tag{2}$$

with design matrix X is also considered, since some of the features of the approach are also applicable and are easier to illustrate in this case. The model structure (2) can be used to describe a local approximation of the original model (1) around $(x_0, \mathbf{p}_0)$, using in this case the variable replacements $y' \to y - y(\mathbf{x}_0, \mathbf{p}_0)$, $\boldsymbol{\theta}' \to \boldsymbol{\theta} - \mathbf{p}_0$, with the sensitivity matrix:

$$X = \left.\frac{\partial f(\mathbf{x}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right|_{\mathbf{x}_0, \mathbf{p}_0}$$

As standard and in both cases, the ranges of the variables $\mathbf{x}$ and $\mathbf{y}$ are assumed to have been previously scaled from the original domain variables to comparable magnitudes (e.g., around unity) so that their relative effects can be easily compared, leading also to more scaled parameter values and matrix coefficients. Although this is an usually implicit step, this scaling is required for the success of the numerical implementation of the algorithm and for a meaningful interpretation of the some of the performance indicators considered in Section 3.6.

The corresponding global FIM of model (1) at a singleton point $\mathbf{p} \in \boldsymbol{\Theta}$ for optimal design ξ is

$$\mathcal{M}(\xi|\mathbf{x}, \boldsymbol{\theta}) = -\mathbb{E}\left[\frac{\partial}{\partial \boldsymbol{\theta}}\left(\frac{\partial \mathcal{L}(\xi|\mathbf{p})}{\partial \boldsymbol{\theta}^\intercal}\right)\right] = \sum_{j=1}^{n_x} w_j \; M(\mathbf{x}_j|\mathbf{p}) =$$

$$= \sum_{j=1}^{n_x} w_j \left. \frac{\partial f(\mathbf{x},\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^\intercal} \right|_{\mathbf{x}_j,\mathbf{p}} \left. \frac{\partial f(\mathbf{x},\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right|_{\mathbf{x}_j,\mathbf{p}}, \tag{3}$$

where $\mathbf{w}$ is the vector of weights of the support points in the design, n_x is the number of discrete (candidate) points, previously set by the user, $M(\mathbf{x}_j|\mathbf{p})$ is the elemental FIM at $\mathbf{x}_j$, and $\mathbb{E}[\bullet]$ stands for the expectation.

Now, we introduce two definitions that are to be used in the remaining parts of the paper.

Definition 1. Parametric transformation — a vector of functions, designated by $\mathcal{T}$, that maps the original vector of parameters $\boldsymbol{\theta}$ into another one, say $\boldsymbol{\vartheta} \in \mathbb{R}^{n_\theta}$, i.e. $\mathcal{T} : \boldsymbol{\theta} \mapsto \boldsymbol{\vartheta}$.

We note that in our context, Definition 1 might include linear and nonlinear functions, and the domain of the transformed vector $\boldsymbol{\vartheta}$ is closed but might be non-Cartesian.

Definition 2. Stable parameter vector — a set of parameters that, after transformation (see Definition 1), is, for a given design and response model, less intercorrelated than the estimates of the original model structure, as measured, for example, by the conditioning of the dispersion matrices, or graphically by the orientation and eccentricity of the likelihood contours [9].

Specifically, $\boldsymbol{\vartheta}$ is a stable parameter vector with respect to $\boldsymbol{\theta}$ if the parameters ϑ_i, $i \in [\![n_\theta]\!]$ are less intercorrelated than θ_i. In practice, they are closer to orthogonality, where the degree of orthogonality is strongly related to the condition number of the parametric covariance matrix, see Xu [16].

2.1. Semidefinite programming

In this Section, we introduce the fundamentals of Semidefinite Programming. This class of (convex) mathematical programming is employed to solve the optimal experimental design problems, given the discrete design domain $\mathbb{X}^{[\![n_x]\!]}$ populated with n_x experimental candidate points.

Let $\mathbb{S}_+^{n_\theta}$ be the space of $n_\theta \times n_\theta$ symmetric positive semidefinite matrices, and $\mathbb{S}^{n_\theta}$ the space of $n_\theta \times n_\theta$ symmetric matrices. A convex set $\mathbf{S} \in \mathbb{R}^{n_\theta}$ is semidefinite representable (SDr) if $\text{proj}_{\mathbf{S}^{\text{exp}}}(\boldsymbol{\zeta})$, $\forall \boldsymbol{\zeta} \in \mathbf{S}$, interpreted as the projection $\boldsymbol{\zeta}$ on to a higher dimensional set $\mathbf{S}^{\text{exp}}$, can be described by Linear Matrix Inequalities (LMIs).

In turn, a convex (or concave) function $\varphi : \mathbb{R}^{m_1} \mapsto \mathbb{R}$ is SDr if and only if the epigraph of φ, $\{(t,\boldsymbol{\zeta}) : \varphi(\boldsymbol{\zeta}) \le t\}$, or the hypograph, $\{(t,\boldsymbol{\zeta}) : \varphi(\boldsymbol{\zeta}) \ge t\}$, respectively, are SDr and can be casted by LMIs [17,18]. The optimal values, $\boldsymbol{\zeta}$, of SDr functions are then formulated as *semidefinite programs* of the form [18, §4.6.2]:

$$\max_{\boldsymbol{\zeta}} \quad \mathbf{d}^\intercal \boldsymbol{\zeta} \tag{4a}$$

$$\text{s.t.} \quad \sum_{i=1}^{m_1} \zeta_i \, M_{i,j} + M_{0,j} \preceq 0, \quad j \in [\![k]\!] \tag{4b}$$

$$M_0 \, \boldsymbol{\zeta} \preceq h \tag{4c}$$

$$M_{i,j} \in \mathbb{S}_+^k, \quad i \in \{0,\dots,m_1\}, \quad j \in [\![k]\!]. \tag{4d}$$

In our design context, $\mathbf{d}$ is a vector of known constants that depends on the design problem, and semidefinite positive matrices $M_{i,j}$, $i \in \{0,\dots,m_1\}, j \in [\![k]\!]$ contain local FIMs and other matrices produced by the reformulation of the functions $\varphi(\boldsymbol{\zeta})$ into LMIs. The decision variables in vector $\boldsymbol{\zeta}$ are the weights w_i, $i \in [\![n_x]\!]$, of the optimal design and other auxiliary variables required. The problem of calculating a design for a pre-specified set of candidate experiments $\mathbb{X}^{[\![n_x]\!]}$ of points $\mathbf{x}_i$, $\forall i \in [\![n_x]\!]$, is solved with the formulation (4) complemented by the linear constraints on $\mathbf{w}$: (i) $\mathbf{w} \ge 0$, and (ii) $\mathbf{1}_{n_x}^\intercal \mathbf{w} = 1$, where $\mathbf{1}_{n_x}^\intercal$ is a unitary column vector with n_x lines. The problem (4) is the classic Semidefinite Programming problem which includes LMIs representing conic constraints.

Ben-Tal and Nemirovski [17] provide a list of SDr functions useful for solving continuous optimal design problems with SDP formulations, see Boyd and Vandenberghe [18, Sec. 7.3]. Recently, Sagnol [19] showed that each criterion in the Kiefer class of optimality criteria is SDr for all rational values of $\delta \in (-\infty, 0[$ and general Semidefinite Programming formulations exist for them. Notice that A-optimality corresponds to $\delta = -1$, E-optimality to $\delta \to -\infty$ and D-optimality to $\delta \to 0$. Practically, the problem of finding optimal approximate plans of experiments for the most common convex (or concave) criteria can be formulated as a Semidefinite Programming problem falling into the general representation, see Vandenberghe and Boyd [20] and Duarte and Wong [21] among others.

2.2. Nonlinear programming

In this Section we introduce the fundamentals of Nonlinear Programming. NLP is used to solve the LS problem, and seeks to find the global optimum of a convex or nonconvex nonlinear function $\psi : \mathbf{X} \mapsto \mathbb{R}$ in a compact domain $\mathbf{X}$ with possibly nonlinear constraints. The general structure of the NLP problems is

$$\min_{\mathbf{x}\in\mathbf{X}} \quad \psi(\mathbf{x}) \tag{5a}$$

$$\text{s.t.} \quad \mathbf{g}(\mathbf{x}) \le \mathbf{0} \tag{5b}$$

$$\mathbf{h}(\mathbf{x}) = \mathbf{0} \tag{5c}$$

where (5b) represents a set of r_i inequalities, and (5c) represents a set of r_e equality constraints. The functions $\psi(\mathbf{x})$, $\mathbf{g}(\mathbf{x})$ and $\mathbf{h}(\mathbf{x})$ are assumed to be twice differentiable. In our context, the variable $\mathbf{x} \in \mathbf{X}$ are the parameters to be estimated, the objective function (5a) corresponds to the log-likelihood. If unconstrained LS are used the sets r_i and r_e are empty, but the decision variables can be bounded to $[\theta_i^L, \theta_i^U]$. When additional constraints are included in the problem, they populate the set of inequalities (5b) or the set of equalities (5c).

3. Algorithm for model reparameterization and testing

In this Section we introduce the framework proposed for model reparameterization and testing. First, the complete algorithm is discussed, then each of the steps is analysed in depth, see Section 3.1–Section 3.6.

Fig. 1 illustrates the basic sequence of steps of the proposed tool. We note that Step 2 is to discretize the design space and construct experimental candidate points which are used in Step 3 to find optimal designs of experiments for various alphabetic criteria. Then, the optimal designs are employed for two purposes: (i) the support points of the K-optimal design are used to construct a transformation to a stable vector of parameters (see Step 4); (ii) all the optimal designs found are used to generate data samples to test the parameterized model by fitting the stable set of parameters via LS (see Step 5). Step 6 is to fit the parameterized model using various data sets. Finally, Step 7 is to compute a set of numerical and graphical indicators to compare the performance of the reparameterization technique introduced.

3.1. Discretization of the design space

We consider the regression model (1). A uniformly spaced grid is used for discretization purposes, where Δx is the step. Consequently, the continuous design space $\mathbf{X}$ of the regressors is approximated by a (finite) discrete set of candidate points, $\mathbb{X}^{[\![n_x]\!]}$. Here, $n_x = 1 + (x^U - x^L)/\Delta x$ where x^U and x^L are the upper bound and lower bounds of $\mathbf{X}$, respectively. After discretizing the design space the *local* FIMs at each candidate point are constructed. When nonlinear models are considered the singleton vector of parameters used to construct the FIMs is $\mathbf{p}_0$.

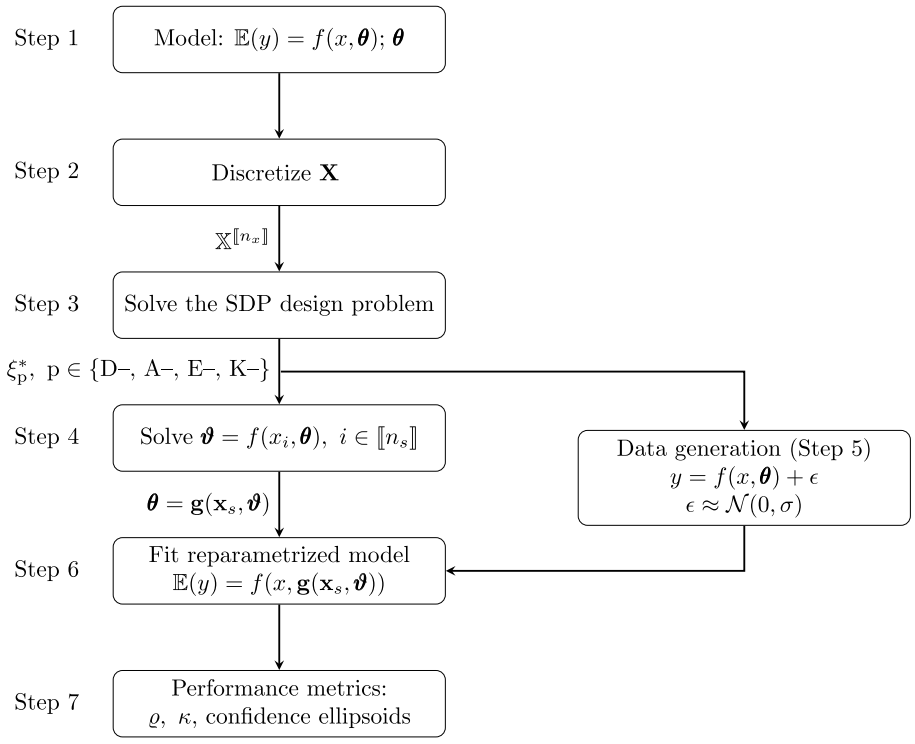


Fig. 1. Algorithm for model reparameterization and testing.

3.2. Finding alphabetic locally optimal designs of experiments via semidefinite programming

This Section presents the formulations for finding locally optimal alphabetic designs via SDP (Step 3 of the Algorithm). We notice that the Semidefinite Programming can be computationally challenging if the number of candidate experiments is large, although it ensures that the global optimum is found for a grid of discrete candidate points. The Semidefinite Programming formulations for all the criteria are of the general form (4). The formulations for D-, A-, E-optimality criteria are currently state of art and appear in Appendix A. The formulation for K-optimality is less familiar and therefore is analysed here in more detail.

The K-optimality criterion minimizes the condition number of the FIM by choice of the experimental design points. The condition number is defined as the ratio $\gamma[\mathcal{M}(\xi|\mathbf{x}, \theta)] = \lambda_{\max}[\mathcal{M}(\xi|\mathbf{x}, \theta)]/\lambda_{\min}[\mathcal{M}(\xi|\mathbf{x}, \theta)]$, where $\lambda_{\max}[\bullet]$ is the maximum eigenvalue and $\lambda_{\min}[\bullet]$ the minimum eigenvalue of the FIM. Then $\gamma[\bullet]$ can be used as a bound on the amplification factor between relative changes in the vector of experimental observations and their effects on the parameter estimates, in terms of norms of these quantities. Thus, by minimizing the condition number, we minimize the potential impact of the observational noise on the estimated parameters. The design problem can be represented by

$$\xi_K = \arg\min_{\xi \in \Xi} \frac{\lambda_{\max}[\mathcal{M}(\xi|\mathbf{x}, \mathbf{p})]}{\lambda_{\min}[\mathcal{M}(\xi|\mathbf{x}, \mathbf{p})]}, \tag{6}$$

where Ξ is the space of feasible K-optimal designs.

The formulation for K-optimal designs was firstly considered in Ye and Zhou [22] where the authors proved that the condition number of the Fisher Information Matrix, which is by nature positive semidefinite, is a smooth function for polynomial regression models and design intervals limited to $[-1, +1]$. Consequently, finding K-optimal designs can be represented by a Semidefinite Programming problem in the domain of the regressors. Herein, we also adopt a formulation similar to that of Ye and Zhou [22]. The SDP formulation for K-optimal designs is

$$\mathrm{Opt} \equiv \min_{s,t,\mathbf{z}} \quad s \tag{7a}$$

$$\text{s.t.} \quad s I_{n_\theta} - M(\xi) \succeq 0_{n_\theta} \tag{7b}$$

$$M(\xi) - I_{n_\theta} \succeq 0_{n_\theta} \tag{7c}$$

$$M(\xi) = \sum_{i=1}^{n_x} z_i \, \mathbb{M}(x_i, \mathbf{p}_0) \tag{7d}$$

$$\sum_{i=1}^{n_x} z_i = t \tag{7e}$$

$$z_i \geq 0, \quad i \in [\![n_x]\!], \; t > 0, \tag{7f}$$

where (7b) is the upper bound of the set of eigenvalues of the FIM, (7c) is to constrain the set of eigenvalues to positive values, (7d) constructs the global FIM from local FIMs at candidate points x_i, $i \in [\![n_x]\!]$ (designated by $\mathbb{M}(x_i, \mathbf{p}_0)$ where the notation $\mathbf{p}_0$ is to make explicit the dependence on the parameters when nonlinear models are considered), (7e) computes the sum of values $\mathbf{z}$, and Opt is the optimal objective value (designated as the *optimum*). After obtaining the solution, the weights of the optimal design are calculated by normalizing the vector $\mathbf{z}$:

$$w_i = z_i/t, \quad i \in [\![n_x]\!] \tag{8}$$

In our work, we solved the Semidefinite Programming problems using the `cvx` environment combined with the solver `Mosek` that uses an efficient Interior Point algorithm [23]. The relative and absolute tolerances used to solve the SDP problem were set to 1×10^{-5}.

3.3. Construction of parametric transformations that lead to stable parameters

This Section describes the techniques used to construct parametric transformations that lead to stable vectors of parameters (Step 4 of the Algorithm), see Definition 2. We consider the reparameterization technique based on a family of curves constructed such that they pass through a set of previously chosen points in the domain of the regressors, see Ross [24]. The sought parametric transformation maps the original set of parameters into an equivalent orthogonal (or nearly

orthogonal) set — the set of stable parameters. Here, instead of using equispaced points, we use the support points of the chosen optimal design, since they can provide the exact number of equations required by the method, and ensure that each transformed parameter is individually expressed in the model response, when the model is latter sampled at these design points. In this paper, the K-optimal design is used due to its distinguishing properties, although other optimal designs (such as the ones provided by the D-, A-, E-optimality criteria) could also be used, since they also yield similar orthogonality properties. Two distinct approaches are considered for the practical construction of these parametric transformations.

Approach 1. Let $\mathbf{x}_s$ be the vector of size n_s containing the support points of the K-optimal design obtained solving the corresponding problem (7) where $n_s \geq n_\theta$. The first approach used to find $\mathcal{T}$, here denoted as *Approach 1*, consists of formulating a set of algebraic equations where the vector of stable parameters represent the response surface at the support points, i.e.

$$\vartheta_i = f(x_i, \boldsymbol{\theta}), \quad i \in [\![n_\theta]\!] \tag{9}$$

where the components ϑ_i, $i \in [\![n_\theta]\!]$ appear individually in each equation. For the case where $n_s > n_\theta$, the question of which of the n_s support points should be selected to be used in (9) arises. This situation is not very common, even with nonlinear models, since optimal designs are typically saturated, i.e., they tend to produce $n_s = n_\theta$. But even if this does not happen, for the purpose of the use of the transformation (9) any subset of n_θ support points is feasible to be selected. The possible additional effects of having designs with $n_s > n_\theta$ is discussed in Section 6.

A rearrangement of (9) allows expressing $\boldsymbol{\theta}$ as a function of $\boldsymbol{\vartheta}$

$$\boldsymbol{\theta} = \mathbf{g}(\mathbf{x}_s, \boldsymbol{\vartheta}) \tag{10}$$

where $\mathbf{g}(\bullet)$ is a vector of functions forming the parametric transformation $\mathcal{T}$ (see Definition 1). This solution of (9) exists provided that the set of parameters $\boldsymbol{\theta}$ is *identifiable*, considering the support points x_i. In some cases, the analytic expression of (10) can be obtained through symbolic manipulation tools, such as Mathematica® [25]. In fact, if the functions $f(x, \boldsymbol{\theta})$ are expressed in closed form, and do not involve trigonometric or transcendental terms then (10) can also be exactly expressed in analytic closed form; otherwise, an implicit definition or the explicit form of some approximation need to be considered. For complex nonlinear models, these techniques can become rather specific and may not be entirely successful in its target goal. Therefore, an alternative approach which can be more generally implemented is considered next.

Approach 2. An alternative approach corresponds to the approximation of $f(x, \boldsymbol{\theta})$ by a Taylor expansion with respect to $\boldsymbol{\theta}$, here limited to first order terms. In this case the system (9) is replaced by a first-order approximation

$$\boldsymbol{\vartheta} \simeq \mathbf{f}_0 + \boldsymbol{J}(\mathbf{x}, \mathbf{p}_0)\ (\boldsymbol{\theta} - \mathbf{p}_0) \tag{11}$$

where the vector $\mathbf{f}_0 = (f(x_1, \mathbf{p}_0), \ldots, f(x_{n_s}, \mathbf{p}_0))^\intercal$ contains the model predictions at the support points using the original vector of parameters, $\mathbf{p}_0$. Further, $\boldsymbol{J}(\mathbf{x}, \boldsymbol{\theta})$ is the $n_s \times n_\theta$ Jacobian matrix

$$\boldsymbol{J}(\mathbf{x}, \mathbf{p}_0) = \begin{pmatrix} \nabla f(x_1, \mathbf{p}_0) \\ \vdots \\ \nabla f(x_{n_s}, \mathbf{p}_0) \end{pmatrix}$$

formed by $1 \times n_\theta$ vectors containing the derivatives of the model with respect to the parameters at support point i:

$$\nabla f(x_i, \boldsymbol{\theta}) = \left(\frac{\partial f(x_i, \boldsymbol{\theta})}{\partial \theta_j} \right)\Bigg|_{\boldsymbol{\theta}=\mathbf{p}_0}, \quad j \in [\![n_\theta]\!]$$

The solution for $\boldsymbol{\theta}$ is therefore

$$\boldsymbol{\theta} = \mathbf{p}_0 + [\boldsymbol{J}(\mathbf{x}, \mathbf{p}_0)]^{-1}\ (\boldsymbol{\vartheta} - \mathbf{f}_0)$$

where $[\boldsymbol{J}(\mathbf{x}, \mathbf{p}_0)]^{-1}$ is the inverse of the Jacobian matrix; consequently, $\mathbf{g}(\mathbf{x}_s, \boldsymbol{\vartheta}) = \mathbf{p}_0 + [\boldsymbol{J}(\mathbf{x}, \mathbf{p}_0)]^{-1}\ (\boldsymbol{\vartheta} - \mathbf{f}_0)$. Similarly to the previous condition for the existence of the solution (10), the Jacobian matrix in this last equation is invertible if the linearized model approximation is considered identifiable at the support points x_i, and parameter estimates $\mathbf{p}_0$.

3.4. Generation of data for testing the model reparameterization

Here, we present the approach used to construct data sets to test the model reparameterization (see Step 5 of the Algorithm) and compare the fit of the parameterized model with that of the original model.

Let the number of observations in each data set be n_o. The optimal experimental designs are used for sampling; we sample from the support points with the number of observations sampled from each support point obtained from the corresponding weight. We use the rounding procedure of Pukelsheim and Rieder [26] to determine the number of observations taken at each support point ensuring that the total sums to n_o. Each observation from model (1) is corrupted with independent normally distributed random errors with standard deviation σ from a normal random generator. The data sets from the designs obtained for each criteria will be different, depending as they do on the support points and on the weights.

3.5. Fitting the model

Now, we fit the parameterized model (see Step 6 of the Algorithm). Model fitting is a LS problem, where the sum of squared errors of the predictions is minimized. The NLP problem for this is

$$\min_{\boldsymbol{\vartheta}} \quad \sum_{i=1}^{n_o} \left(y_i - f[x_i, \mathbf{g}(\mathbf{x}_s, \boldsymbol{\vartheta})] \right)^2 \tag{12a}$$

$$\text{s.t.} \quad \boldsymbol{\vartheta} \in [\boldsymbol{\vartheta}^L, \boldsymbol{\vartheta}^U], \tag{12b}$$

where $\mathbf{g}(\mathbf{x}_s, \boldsymbol{\vartheta})$ is the transformation obtained in Step 4, $\boldsymbol{\vartheta}^L$, and $\boldsymbol{\vartheta}^U$ are the lower and upper bounds for parameters $\boldsymbol{\vartheta}$, which may be redundant, and (x_i, y_i), $i \in [\![n_o]\!]$ are the observations generated in Step 5. The Examples considered were solved with a straightforward Levenberg–Marquardt algorithm; the details are in Nielsen and Madsen [27]. Note that the same formulation is used for fitting the original model.

3.6. Construction of performance indicators

We now present the performance indicators used to compare the effect of model reparameterization on model fitting (see Step 7 of the Algorithm).

To compare the impact of the various designs on the least-squares problem we used two numerical indicators: (i) the maximum of the absolute cross-correlation among the parameter estimates (designated ϱ) and (ii) the condition number (κ) of the parameter covariance matrix at the solution. To compute ϱ, we first construct an approximation to the correlation matrix using the covariance matrix $C(\hat{\boldsymbol{\theta}})$, given by the linearized model from the LS algorithm at convergence where $\hat{\boldsymbol{\theta}}$ stands for the estimates of $\boldsymbol{\theta}$. Let $B(\hat{\boldsymbol{\theta}})$ be a diagonal (square) matrix of size n_θ containing the square roots of the diagonal elements of $C(\hat{\boldsymbol{\theta}})$, i.e. $B_{i,i} = \sqrt{C_{i,i}}$, $i \in [\![n_\theta]\!]$ and $B_{i,j} = 0$, $i, j \in [\![n_\theta]\!]$, $i \neq j$. The correlation matrix of the parameter estimates is then given by

$$R(\hat{\boldsymbol{\theta}}) = B^{-1}(\hat{\boldsymbol{\theta}})\ C(\hat{\boldsymbol{\theta}})\ B^{-1}(\hat{\boldsymbol{\theta}}).$$

and

$$\varrho = \max_{\substack{i,j \in [\![n_\theta]\!] \\ j \neq i}} |R_{i,j}|.$$

The Variance Inflation Factor (VIF) is a metric commonly used to measure the collinearity between a pair of parameters [28]. It is represented by $\upsilon = 1/(1 - R^2_{i,j})$, and thus:

$$\max \upsilon = \max_{\substack{i,j \in [\![n_\theta]\!] \\ j \neq i}} 1/(1 - R^2_{i,j}) = 1/(1 - \varrho^2)$$

Despite the use of the VIF for measuring collinearity which occurs when a parameter is a linear combination of others, Belsley et al. [29] and Rempel and Zhou [30] suggest the advantages of several condition indexes with the condition number, κ, being one of them. Here, $\kappa = \lambda_{\max}[C(\hat{\theta})]/\lambda_{\min}[C(\hat{\theta})]$, where $\lambda_{\max}[C(\hat{\theta})]$ is the maximum eigenvalue of the covariance matrix and $\lambda_{\min}[C(\hat{\theta})]$ the minimum. A lower condition number is an indication of reduced parametric collinearity. López et al. [2],Belsley et al. [29] provide guidelines for the orthogonality analysis, and suggest the arbitrary but convenient numbers of 0.90 as the cut-off value for the maximum absolute parametric correlation, and 20 as the threshold for the condition number. Despite their arbitrariness, they are helpful in assessing the performance of the reparameterization technique.

To complement the metrics ϱ and κ we also provide the graphical representation of the 95% confidence ellipsoids for selected pairs of parameters. The construction of these ellipsoids also uses the final linearized version of the Jacobian after convergence is achieved [1]. Considering the scaled variable ranges, confidence ellipsoids where the principal axes are not well aligned with the coordinate axes, associated with higher condition numbers of the parameter covariance matrix at the solution denote higher parameter collinearity (or equivalently, reduced orthogonality). Throughout the paper the eccentricity of these ellipsoids is denoted as the ratio of the lengths of the major semi-axis to the minor semi-axis of the confidence region for a pair of parameters. When the eccentricity is 1.0 the confidence region is circular and the parameters are uncorrelated. Contrarily, when the axes of the ellipses are not the coordinate axes, as the eccentricity tends to $+\infty$ the confidence region tends to a line, and the parameters become highly correlated.

The absolute and relative tolerances imposed on the SDP and NLP solvers were set equal to 1×10^{-5} and 1×10^{-6}, respectively. All computations in Section 4 used an Intel Core i7 machine running a 64 bits Windows 10 operating system with a 2.80 GHz processor.

4. Application examples

In this Section, we apply the formulations of Section 3 to find parameterizations for nonlinear regression models. The first Example considered is a linear (in the parameters) polynomial model while the remaining four are nonlinear. The first model, see Section 4.1, was considered by Ye and Zhou [22] to illustrate the construction of K-optimal designs with Semidefinite Programming. It is considered here for comparative purposes and serves to demonstrate that the formulation (7) is accurate for computing K-optimal designs. In Section 3.3 we introduced two approaches for the calculation of the transformations. The remaining Examples demonstrate the application of *Approach 1* (Examples 1 to 4, discussed in Section 4.1- Section 4.4) and of *Approach 2* (Examples 4 and 5, in Section 4.4- Section 4.5). All of the examples presented in the next Sections required less than 15 s of CPU time. The solution of the SDP problems required less than 3 s each and the LS determination less than 1 s. Finally, we note that the optimal designs obtained for some of the examples used for testing the algorithm have analytical solution or can be addressed with alternative numerical approaches. These tools can be used instead of the SDP-based formulations in Step 3 of the Algorithm.

From this point onwards we use the counter $i \in [\![n_\theta]\!]$ for ordering the parameters in $\boldsymbol{\theta}$ and $\boldsymbol{\vartheta}$.

4.1. Example 1 — Polynomial model of third degree

Our first example is the third degree polynomial

$$\mathbb{E}(y) = \theta_1 + \theta_2\, x + \theta_3\, x^2 + \theta_4\, x^3, \tag{13}$$

where $\boldsymbol{\theta} \equiv (\theta_0, \theta_1, \theta_2, \theta_3)^\intercal$ and the design domain is $\mathbf{X} = [-1, +1]$. For discretizing the domain we use a uniform grid with $\Delta x = 0.02$ (see Section 3.1); consequently, $n_x = 101$ and the set of candidate design points is $\mathbb{X}^{[\![101]\!]}$. The results are in Table 1, and are in agreement with those of Ye and Zhou [22]. To simplify the interpretation of the results, the first column of Table 1 lists the optimality criteria, the second column is for the corresponding optimal design obtained with the Semidefinite Programming formulations (in Section 3.2 and Appendix A) and the third column is for the optimum of the SDP problem (7). We note for this example the coincidence of the support points of all designs except those for E-optimality. However, their respective weights differ. As expected, for a problem in which $n_s = n_\theta$, D-optimality equidistributes the weights. The K-optimality criterion relatively overweights the middle support points, whereas the A- and E-optimality criteria lead to designs between these two extremes. The designs obtained for other polynomial orders (not shown here) are also very similar to those of Ye and Zhou [22].

To analyse the performance of D-, A-, E- and K-optimal designs in capturing the "real" model parameters we used the simulation-based approach described in Section 3.5. We set n_o to 200 observations, a value that is used in all the remaining examples. Then, we produced exact designs equivalent to the approximate designs in Table 1 by use of the rounding procedure of Pukelsheim and Rieder [26]; for instance for the D-optimality criterion the number of observations at each support point $x_i \in \{-1.0000, -0.4446, +0.4446, +1.0000\}$ is $\lfloor 0.2500 \times 200 \rceil = 50$. Similar calculations were made for the other criteria.

Next, we generated the observations using the model (13) with $\mathbf{p}_0 = (1.0, 2.0, -2.0, -4.0)^\intercal$ (this vector is interpreted as the vector of "real" model parameters), and corrupted the observations with zero mean normally distributed random noise with $\sigma = 0.1$. Then, we used LS (see Section 3.5) to fit the data generated from the design for each criterion to a model structurally equal to (13) and obtained the estimated parameters, represented as $\hat{\boldsymbol{\theta}}$. The estimated parameters produced from all the data sets should be close, but differ in the parametric confidence regions.

Finally, we constructed the parametric covariance matrix and subsequently determined the performance metrics (ϱ and κ) described in Section 3.6. Their values, obtained from fitting each data set, are listed in columns 4 and 5 of Table 1. The parametric covariance matrix is also used to compute the 95% confidence ellipsoids (see the above panel of Fig. 2). Several findings can be noted: (i) the optimal designs yield parameter estimates near the "real" values used for simulation; (ii) Fig. 2(a) shows no correlation between parameter estimates - $\hat{\theta}_3$ vs. $\hat{\theta}_2$ – a direct consequence of the elements $(2, 3)$ and $(3, 2)$ of the covariance matrix at convergence being nearly 0; (iii) Fig. 2(b) shows that a degree of collinearity exists between $\hat{\theta}_4$ and $\hat{\theta}_2$ (the ellipses exhibit some degree of eccentricity and their axes do not coincide with the coordinate axes); (iv) all optimality criteria produce similar confidence regions. We notice that the visual analysis of collinearity requires analysing all pairs of parameters; here, only two of them are plotted.

The correlation of some pairs of parameter estimates (one example being $\hat{\theta}_1$ and $\hat{\theta}_3$) is also revealed by the metrics ϱ and κ. The data set generated with the D-optimal design leads to a value of $\varrho = 0.95$ (above the cut-off). Likewise, all data sets fitted produce models with a parametric covariance matrix with κ above the threshold (20). The results also confirm the advantages of the K-optimality criterion in this context (i.e., to sample the system) relative to the other criteria since the condition number of the covariance matrix of the LS problem obtained from the K-optimal design is the lowest, as expected, since it minimizes the maximum parametric sensitivity to observational error. We note that the results obtained by sampling from all designs are above the threshold, which suggest that a model reparameterization would be beneficial; it will be considered next. From this point on we use $\hat{\vartheta}_i$ to designate the parameter estimates obtained fitting the reparameterized model.

Now, we parameterize model (13). The algebraic treatment of the Eqs. (9) leads to the following linear transformation:

$$\boldsymbol{\theta} = \begin{pmatrix} -0.1342\,\vartheta_1 + 0.6342\,\vartheta_2 + 0.6342\,\vartheta_3 - 0.1342\,\vartheta_4 \\ 0.1342\,\vartheta_1 - 1.3787\,\vartheta_2 + 1.3787\,\vartheta_3 - 0.1342\,\vartheta_4 \\ 0.6342\,\vartheta_1 - 0.6342\,\vartheta_2 - 0.6342\,\vartheta_3 + 0.6342\,\vartheta_4 \\ -0.6342\,\vartheta_1 + 1.3787\,\vartheta_2 - 1.3787\,\vartheta_3 + 0.6342\,\vartheta_4 \end{pmatrix} \tag{14}$$

Table 1
Polynomial model of third degree (13): optimal designs, $\mathbf{X} = [-1, +1]$ and $\Delta x = 0.02$.

Optimality criterion	Design				Optimum	$\varrho^{\dagger}$	$\kappa^{\dagger}$	$\varrho^{\ddagger}$	$\kappa^{\ddagger}$
D–	$\begin{pmatrix} -1.0000 & -0.4446 & 0.4446 & 1.0000 \\ 0.2500 & 0.2500 & 0.2500 & 0.2500 \end{pmatrix}$				0.2674	0.9485	47.3904	0.0339	1.0801
A–	$\begin{pmatrix} -1.0000 & -0.4600 & 0.4600 & 1.0000 \\ 0.1497 & 0.3503 & 0.3503 & 0.1497 \end{pmatrix}$				37.5254	0.8939	31.7578	0.0000	2.3333
E–	$\begin{pmatrix} -1.0000 & -0.5000 & 0.5000 & 1.0000 \\ 0.1267 & 0.3733 & 0.3733 & 0.1267 \end{pmatrix}$				0.0400	0.8781	30.2806	0.0858	3.1894
K–	$\begin{pmatrix} -1.0000 & -0.4600 & 0.4600 & 1.0000 \\ 0.0969 & 0.4031 & 0.4031 & 0.0969 \end{pmatrix}$				29.3651	0.8476	29.3569	0.0000	4.2083

† — based on original model.
‡ — based on parameterized model.

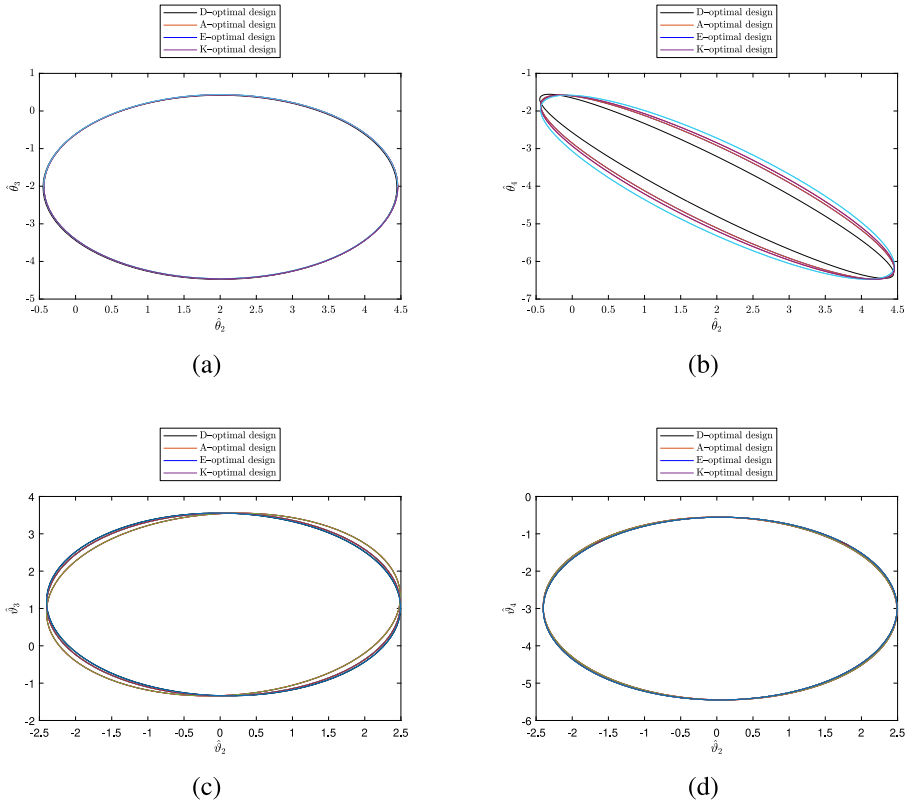


Fig. 2. Polynomial model of third degree: 95% confidence ellipses for parameters obtained with the optimal designs in Table 1 for: (a) $\hat{\vartheta}_3$ vs. $\hat{\vartheta}_2$; (b) $\hat{\vartheta}_4$ vs. $\hat{\vartheta}_2$; (c) $\hat{\vartheta}_3$ vs. $\hat{\vartheta}_2$ using transformation (14); (d) $\hat{\vartheta}_4$ vs. $\hat{\vartheta}_2$ using transformation (14).

Figs. 2(c) and 2(d) present the confidence regions for $\hat{\vartheta}_2$ vs. $\hat{\vartheta}_1$ and $\hat{\vartheta}_3$ vs. $\hat{\vartheta}_1$, respectively, after reparameterization. We note that the axis of the ellipses for the pair of parameters $\hat{\vartheta}_3$ and $\hat{\vartheta}_1$ (left column of the lower panel) are now aligned with the coordinate axes, thus showing the orthogonality of the new parameters. This is also shown by observing the values of ϱ and κ in Table 1 (see columns 6 and 7). The maximum of the absolute of cross-correlation is below 0.10 for all data sets, being exactly 0 for A- and K-optimality; this result was expected for K-optimality, since the K-optimal support points were used for model reparameterization and the algebraic manipulation of the analytic equations is exact. A similar result could be obtained for the remaining optimality criteria if linear transformations of the form (14) were considered, but based instead in the support points of the D- and E- designs. Now, the condition number is well below 20 for all criteria.

Finally, we note that the regression of Eq. (13) in $[-1, 1]$ leads to a coefficient or design matrix (of the over-determined system to be solved in the least squares sense) of Bernstein–Vandermonde kind [31]. The numerical condition of these matrices can grow exponentially with size [32]. Here, using optimal designs based on information criteria to choose a limited set of support points significantly helps to limit the condition of the resulting design matrices. The K-optimality criterion is an obvious choice in this particular example as it minimizes specifically the condition number of the covariance matrix.

4.2. Example 2 — Michaelis–Menten rate-based model

Here, we consider the Michaelis–Menten rate model

$$\mathbb{E}(y) = \theta_1 + \frac{\theta_2\, x}{\theta_3 + x}, \tag{15}$$

which has the form of a rational polynomial function and admits exact reparameterization solutions by applying purely algebraic manipulation techniques.

The vector of parameters is $\boldsymbol{\theta} \equiv (\theta_1, \theta_2, \theta_3)^{\intercal}$, and for testing purposes we consider the singleton $\mathbf{p}_0 = (1.0,\ 2.0,\ 5.0)^{\intercal}$. The design space is $\mathbf{X} = [0, 10]$, and we discretize it with a uniformly distributed grid with $\Delta x = 0.1$. The optimal experimental designs obtained via Semidefinite Programming are listed in Table 2. Again, we note that all criteria provide similar support points but different weights. The maximum of parameter intercorrelation is below 0.9 for all data sets (see column

Table 2
Michaelis–Menten model (15): optimal designs for $\mathbf{X} = [0, 10]$, $\Delta x = 0.1$ and $\mathbf{p}_0 = (1.0,\ 2.0,\ 5.0)^\intercal$.

Optimality criterion	Design			Optimum	$\varrho^\dagger$	$\kappa^\dagger$	$\varrho^\ddagger$	$\kappa^\ddagger$
D–	0.0000	2.5000	10.0000	0.0319	0.8547	2.1742E+3	0.0289	1.0816
	0.3334	0.3333	0.3333					
A–	0.0000	2.5000	10.0000	2.0827E+3	0.8060	1.9952E+3	0.0419	2.0345
	0.2475	0.4974	0.2551					
E–	0.0000	2.5000	10.0000	4.8495E−4	0.7594	2.1932E+3	0.0578	4.1066
	0.1642	0.6662	0.1696					
K–	0.0000	2.3000	10.0000	2.2297E+3	0.8212	2.0344E+3	0.0000	2.3056
	0.2854	0.4972	0.2175					

† — based on original model.
‡ — based on parameterized model.

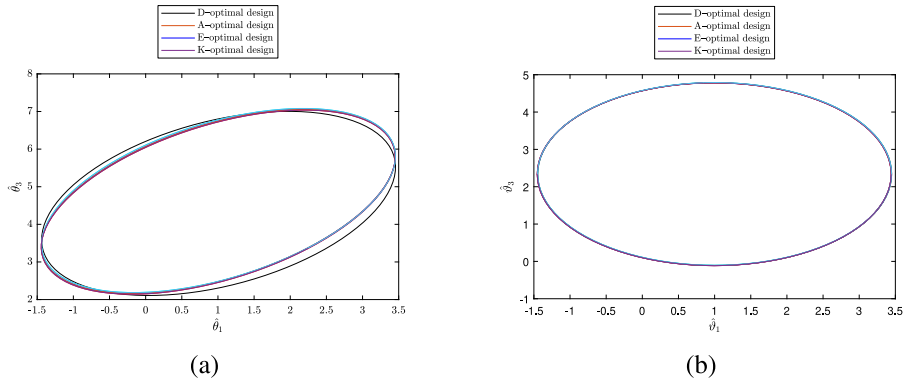


Fig. 3. Michaelis–Menten model (15): 95% confidence ellipses for parameters obtained with the optimal designs in Table 2 for $\hat{\theta}_3$ vs. $\hat{\theta}_1$: (a) considering the original parameters; (b) after reparameterization.

4 of Table 2). Contrarily, the condition number is larger than 1×10^3, denoting a ill-conditioned LS problem (see column 5).

Next, we used the support points of the K-optimal design in Table 2 and *Approach 1* in Section 3.3 to find the parametric transformations to a stable vector, namely:

$$\boldsymbol{\theta} = \left(\ \vartheta_1, \quad \frac{77.0\,(\vartheta_1\,\vartheta_2 + \vartheta_1\,\vartheta_3 - \vartheta_2\,\vartheta_3 - {\vartheta_1}^2)}{77.0\,\vartheta_1 - 100.0\,\vartheta_2 + 23.0\,\vartheta_3}, \quad \frac{230.0\,\vartheta_2 - 230.0\,\vartheta_3}{77.0\,\vartheta_1 - 100.0\,\vartheta_2 + 23.0\,\vartheta_3} \ \right)^\intercal \tag{16}$$

The performance metrics of the nonlinear LS fits for the parameterized model are in columns 6 and 7 of Table 2 and they show: (i) the maximum of parameter intercorrelations is close to 0; (ii) the condition number decreased from 1×10^3 to values below 5 for all the criteria (i.e., below the threshold); and (iii) the data set obtained by sampling with the K-optimal design gives a parametric intercorrelation of exactly 0. This is expected as the new parameters in the transformed model are fully orthogonal and no approximations were required in the algebraic treatment of the equations representing the response surface (see Eq. (9)).

Fig. 3(a) shows the ellipsoids for $\hat{\theta}_3$ vs. $\hat{\theta}_1$ obtained fitting the original model, and Fig. 3(b) is for the reparameterized model obtained using (16). Here, we note the value of one for $\hat{\vartheta}_1$ is also that of $\hat{\theta}_1$. Since it is the ordinate at the origin, this value, unlike the other parameters, was not affected by the reparameterization. Fig. 3(b) also shows that the reparameterization removes any evidence of collinearity whichever design criterion is used, with the major impact also holding for K–optimality for which $\varrho = 0.0$, as expected.

4.3. Example 3 — Exponential model

Here, we consider the exponential model

$$\mathbb{E}(y) = \theta_1 + \theta_2 \exp(-\theta_3\ x) \tag{17}$$

where $\boldsymbol{\theta} \equiv (\theta_1, \theta_2, \theta_3)^\intercal$ and the design domain is $\mathbf{X} = [0, 10]$. We follow Algorithm Fig. 1. For discretizing the design space we use a uniform grid with $\Delta x = 0.1$, i.e. $n_x = 101$. The locally optimal designs are found for $\mathbf{p}_0 \equiv (1.0,\ 1.0,\ 0.1)^\intercal$ and are in Table 3. We note that the K-optimal design is remarkably different from those obtained for the other optimality criteria.

To compare the designs for the ability to generate observations allowing capture of the "real" model parameters, we followed the simulation scheme presented in Section 3.4. The same vector $\mathbf{p}_0$ used to find optimal experimental designs was used to construct new data, with the observational noise used for its randomization following a normal distribution with zero mean and $\sigma = 0.10$. As in Section 4.1, the n_o points forming each data set were sampled using the respective approximate optimal designs previously found. Then, a nonlinear least squares procedure was used for model fitting to the each of the four data sets previously generated, see Section 3.5. The figures of merit of the models identified are in the fourth column, ϱ, and in the fifth column, κ, of Table 3. Both performance metrics show that the estimated model parameters are highly correlated, with the condition number of the covariance matrices at convergence being above 1×10^3 for all models. The 95% confidence ellipsoids are displayed in Fig. 4(a); they strengthen those findings. The confidence ellipsoids for parameters $\hat{\theta}_3$ vs $\hat{\theta}_1$, by their orientation and eccentricity, show that all the designs lead to strongly collinear parameter estimates.

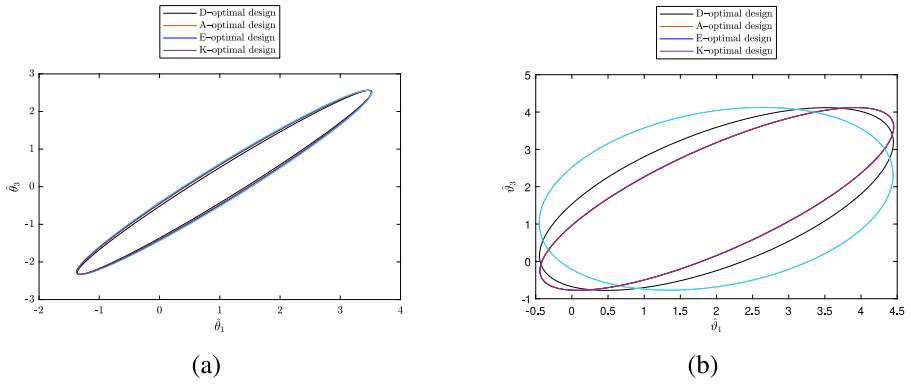
Now, we demonstrate the impact of the reparameterization of model (17). That is, we use the vector of support points of the K-optimal design to determine a transformation to a stable vector of parameters, see Section 3.3. Since this specific problem involves an exponential expression in one of the parameters, no closed form exact reparameterization was possible without simultaneous changes in the data variables. Instead, in this case, an approximate solution of the reparameterization Eqs. (10) was obtained using *Approach 1*, considering the method followed by Ross et al. [14]. Specifically, the system of algebraic equations was reformulated as a ratio and later approximated by an expression that depends only on θ_3:

$$\frac{\vartheta_3 - \vartheta_2}{\vartheta_2 - \vartheta_1} = \frac{\exp(-\theta_3\ x_3) - \exp(-\theta_3\ x_2)}{\exp(-\theta_3\ x_2) - \exp(-\theta_3\ x_1)} \approx \exp[-\theta_3\ (x_3 - x_2)] - 1$$

Table 3
Exponential model (17): optimal designs, $\mathbf{X} = [0, 10]$, $\Delta x = 0.1$, and $\mathbf{p}_0 = (1.0,\ 1.0,\ 0.1)^\intercal$.

Optimality criterion	Design			Optimum	$\varrho^\dagger$	$\kappa^\dagger$	$\varrho^\ddagger$	$\kappa^\ddagger$
D–	$\begin{pmatrix}0.0000\\0.3334\end{pmatrix}$	$\begin{matrix}4.1999\\0.3333\end{matrix}$	$\begin{matrix}10.0000\\0.3333\end{matrix}$	0.2051	0.9952	4.4778E+3	0.6192	9.9220
A–	$\begin{pmatrix}0.0000\\0.1567\end{pmatrix}$	$\begin{matrix}4.200\\0.4826\end{matrix}$	$\begin{matrix}10.0000\\0.3608\end{matrix}$	506.064	0.9896	4.8957E+3	0.7835	20.4597
E–	$\begin{pmatrix}0.0000\\0.1042\end{pmatrix}$	$\begin{matrix}4.2000\\0.6520\end{matrix}$	$\begin{matrix}10.0000\\0.2438\end{matrix}$	0.0019	0.9896	5.0800E+3	0.8065	20.7862
K–	$\begin{pmatrix}0.0000\\0.5458\end{pmatrix}$	$\begin{matrix}2.5000\\0.3399\end{matrix}$	$\begin{matrix}10.0000\\0.1143\end{matrix}$	3408.26	0.9996	2.8100E+3	0.4071	3.3525

† — based on original model.
‡ — based on parameterized model.



(a) (b)

Fig. 4. Exponential model (17): 95% confidence ellipses for parameters obtained with the optimal designs in Table 3 for $\hat{\theta}_3$ vs. $\hat{\theta}_1$: (a) considering the original parameters; (b) after reparameterization.

The approximation was used to relate θ_3 with $\boldsymbol{\vartheta}$ and the remaining parameters of the "original" model (i.e., θ_1 and θ_2) were determined by replacing it in the first two symbolic equalities, yielding:

$$\boldsymbol{\theta} = \left(-\frac{\left(\vartheta_2 - \vartheta_1 \left(\frac{\vartheta_2-\vartheta_3}{\vartheta_1-\vartheta_2}\right)^{1/3}\right)}{\left(\frac{\vartheta_2-\vartheta_3}{\vartheta_1-\vartheta_2}\right)^{1/3}-1.0},\quad -\frac{(\vartheta_1-\vartheta_2)}{\left(\frac{\vartheta_2-\vartheta_3}{\vartheta_1-\vartheta_2}\right)^{1/3}-1.0},\quad -\frac{2\,\ln\left(\frac{\vartheta_1-\vartheta_3}{\vartheta_1-\vartheta_2}\right)}{15} \right)^\intercal \tag{18}$$

The same data used to fit the "original" model was then used to fit the parameterized model where $\boldsymbol{\theta}$ is replaced by (18) and the nonlinear LS problem presented in Section 3.5 re-solved. The performance metrics of the parameterized model are in columns 6 and 7 of Table 3. It should be stressed that the unique parameterization (18) (specific to the K-optimal design support points) was used to generate all data in this Table. Because of this fact, the orthogonality behaviour of the remaining optimal designs needs to be considered as partially impaired, since their support points differ from the case considered. Repeating the same procedure, but considering this time the support points, e.g., of design D- would produce better orthogonal behaviour for the remaining criteria.

In the present case we notice that (i) all the parametric correlations are between −0.81 and 0.81, which indicates the reparameterization is efficient in treating the collinearity. This value is below the threshold, see Section 3.6; (ii) the condition numbers of the LS problems for the parameterized model are below 21 with those for data generated with D- and K-optimal designs below 20; (iii) as expected the lowest condition number is obtained from the data set generated with the K-optimal design, since the reparameterization used the curve passing exactly through the same support points as the design. The failure of the reparameterization (18) in providing a fully orthogonal new parameter set is explained by the approximations considered in the algebraic manipulation of the original Eqs. (9).

Instead of the approximated solution of Eqs. (10), the alternative application of Approach 2 would also be feasible; this is considered in the next example, also involving exponential terms. Finally, the confidence regions for parameters $\hat{\vartheta}_3$ vs $\hat{\vartheta}_1$ (obtained after reparameterization) are in Fig. 4(b). They show a visible decrease in eccentricity which denotes lower collinearity despite the orientation of the ellipses. Here, for the nonlinear model, the values of the parameters obtained for the reparameterized model are different from those found for the original form.

4.4. Example 4 — Logistic model

We consider now the logistic model:

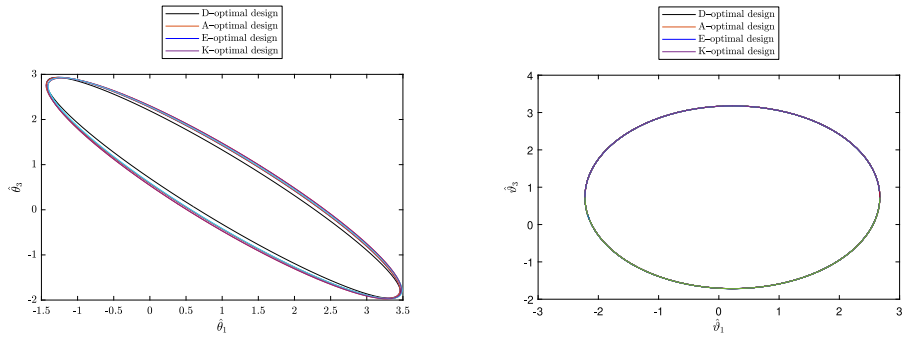
$$\mathbb{E}(y) = \frac{\theta_1}{1 + \exp[-\theta_2\ (x - \theta_3)]} \tag{19}$$

The vector of parameters is $\boldsymbol{\theta} \equiv (\theta_1, \theta_2, \theta_3)^\intercal$; for testing we consider the singleton $\mathbf{p}_0 = (1.0,\ 0.5,\ 0.5)^\intercal$. The design space is $\mathbf{X} = [-2.0, 2.5]$, and we discretize it with a uniformly distributed grid with $\Delta x = 0.045$, corresponding to 101 candidate points. The optimal designs are listed in Table 4. The maximum absolute intercorrelation metrics obtained considering the original parameters are above 0.98 for all data sets (see column 4). The condition number is also very large (see column 5 of the Table) which indicates collinear parameter estimates.

Similarly, we used the support points produced by the K-optimal design in Table 4 and *Approach* 1 in Section 3.3 to find a reparameterization for $\boldsymbol{\theta}$. Like in the previous example, during the algebraic manipulation some approximations were required. In this case, the system of algebraic equations was reformulated as a ratio and then approximated by an expression that depends only on θ_2:

$$\begin{aligned}\frac{1/\vartheta_3 - 1/\vartheta_2}{1/\vartheta_2 - 1/\vartheta_1} &= \frac{\exp[-\theta_2\ (x_3 - \theta_3)] - \exp[-\theta_2\ (x_2 - \theta_2)]}{\exp[-\theta_2\ (x_2 - \theta_3)] - \exp[-\theta_2\ (x_1 - \theta_3)]} \\ &\approx \exp[-\theta_2\ (x_3 - x_2)] - 1\end{aligned}$$

The approximation was used to relate θ_2 with $\boldsymbol{\vartheta}$ and the remaining parameters were determined by replacing this parameter in



(a) (b)

Fig. 5. Logistic model (19): 95% confidence ellipses for parameters obtained with the optimal designs in Table 4 for $\hat{\theta}_3$ vs. $\hat{\theta}_1$: (a) considering the original parameters; (b) after reparameterization.

Table 4
Logistic model (19): optimal designs, $\mathbf{X} = [-2.5, 2.0]$, $\Delta x = 0.045$, and $\mathbf{p}_0 = (1.0,\ 0.5,\ 0.5)^\intercal$.

Optimality criterion	Design			Optimum	$\varrho^\dagger$	$\kappa^\dagger$	$\varrho^\ddagger$	$\kappa^\ddagger$
D–	$\begin{pmatrix} -2.0000 & 0.6102 & 2.5000 \\ 0.3334 & 0.3332 & 0.3334 \end{pmatrix}$			0.0177	0.9909	1.6814E+3	0.1377	1.3851
A–	$\begin{pmatrix} -2.0000 & 0.6550 & 2.5000 \\ 0.2215 & 0.4918 & 0.2868 \end{pmatrix}$			4.7069E+3	0.9890	1.4175E+3	0.2211	2.5908
E–	$\begin{pmatrix} -2.0000 & 0.6550 & 2.5000 \\ 0.1477 & 0.6595 & 0.1928 \end{pmatrix}$			2.1322E−4	0.9919	1.5985E+3	0.3038	5.0802
K–	$\begin{pmatrix} -2.0000 & 0.2500 & 2.5000 \\ 0.4209 & 0.4317 & 0.1475 \end{pmatrix}$			1.1668E+3	0.9898	1.1338E+3	0.0000	2.9189

† — based on original model.
‡ — based on parameterized model.

the first and third symbolic equalities, leading to the approximate reparameterization:

$$\boldsymbol{\theta} = \begin{pmatrix} -\dfrac{\vartheta_2\,(\vartheta_1\,\vartheta_2 - 2.0\,\vartheta_1\,\vartheta_3 + \vartheta_2\,\vartheta_3)}{\vartheta_1\,\vartheta_3 - \vartheta_2^{\,2}} \\ -\dfrac{4\,\ln\left(\dfrac{\vartheta_1\,(\vartheta_2 - \vartheta_3)}{\vartheta_3\,(\vartheta_1 - \vartheta_2)}\right)}{9} \\ 2.25\,\ln\left(-\dfrac{\vartheta_3\,(\vartheta_1 - \vartheta_2)^2\left(\dfrac{\vartheta_1\,(\vartheta_2 - \vartheta_3)}{\vartheta_3\,(\vartheta_1 - \vartheta_2)}\right)^{8/9}}{\vartheta_1\,(\vartheta_1\,\vartheta_3 - \vartheta_2^{\,2})}\right) \\ -\dfrac{}{\ln\left(\dfrac{\vartheta_1\,(\vartheta_2 - \vartheta_3)}{\vartheta_3\,(\vartheta_1 - \vartheta_2)}\right)} \end{pmatrix} \tag{20}$$

The results for the parameterized model are in columns 6 and 7 of Table 4 and they show: (i) the parametric intercorrelation coefficients are all below or equal to 0.3; (ii) the condition number is below 5.1 for all criteria; and (iii) the data set generated with the K-optimal design yields non-correlated parameters ($\varrho = 0$); the parameters are orthogonal in the transformed model. Fig. 5(a) shows the ellipsoids for $\hat{\theta}_3$ vs. $\hat{\theta}_1$ obtained fitting the original model, while Fig. 5(b) is for the reparameterized model. As in the previous examples, Fig. 5(b) demonstrates that the K–optimal design based parameterization shows no evidence of parameter collinearity. A similar behaviour is expected for the remaining optimal designs, provided that the exact support points were considered in the reparameterization (20).

Now, we apply *Approach 2* in Section 3.3 to reparameterize the model (19) and compare the results obtained with *Approach 1*. The vector of stable parameters obtained with the first-order Taylor approximation is:

$$\boldsymbol{\theta} = \begin{pmatrix} 4.6595\,\vartheta_1 - 6.4779\,\vartheta_2 + 4.1024\,\vartheta_3 \\ -3.8017\,\vartheta_1 + 3.5006\,\vartheta_2 - 1.0867\,\vartheta_3 + 0.5000 \\ 19.4439\,\vartheta_1 - 34.1709\,\vartheta_2 + 15.9889\,\vartheta_3 + 0.5000 \end{pmatrix} \tag{21}$$

Table 5 compares the figures of merit of the LS problem obtained with (i) the original set of parameters; (ii) the vector (20); and (iii) the vector of parameters (21) obtained by approximating $f(x, \boldsymbol{\theta})$ to a first order polynomial via Taylor expansion. The results show that the maximum parameter intercorrelation and the condition number of the LS problems for this latest vector of parameters are similar to those of (20) (compare columns 3 and 5, and 4 and 6, respectively). The condition number obtained for all the data sets is below 6, well below the threshold. Now, the value of ϱ for data set sampled with the K-optimal design is not 0, unlike that obtained with the transformation (20). This indicates that, due to the first-order approximation, the vector (21) is not fully orthogonal; nevertheless, it still provides an acceptable set of stable parameters.

Table 5
Logistic model (19): optimal designs, $\mathbf{X} = [-2.5, 2.0]$, $\Delta x = 0.045$, and $\mathbf{p}_0 = (1.0,\ 0.5,\ 0.5)^\intercal$.

Optimality criterion	$\varrho^\dagger$	$\kappa^\dagger$	$\varrho^\ddagger$	$\kappa^\ddagger$	$\varrho^\star$	$\kappa^\star$
D–	0.9909	1.6814E+3	0.1377	1.3851	0.1807	1.7779
A–	0.9890	1.4175E+3	0.2211	2.5908	0.2180	2.5200
E–	0.9919	1.5985E+3	0.3038	5.0802	0.3112	5.1429
K–	0.9898	1.1338E+3	0.0000	2.9189	0.0853	4.2371

† — based on original model.
‡ — based on the reparameterization (20).
⋆ — based on the reparameterization (21).

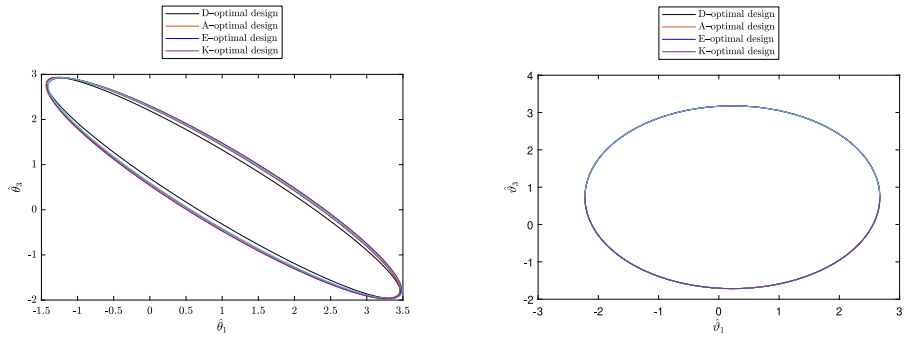
Fig. 6(a) shows the ellipsoids for $\hat{\vartheta}_3$ vs. $\hat{\vartheta}_1$ obtained with the set of stable parameters (20) while Fig. 6(b) is for the reparameterized model obtained with the Taylor expansion, i.e. vector (21). Fig. 6(b) demonstrates that the approximation has only a marginal effect on the confidence ellipses and that the eccentricity is hardly changed.

4.5. Example 5 — Four-parameter hill model

Finally, we consider the four-parameter Hill model [33]:

$$\mathbb{E}(y) = \theta_1 + \frac{(\theta_2 - \theta_1)\ x^{\theta_4}}{x^{\theta_4} + \theta_3}. \tag{22}$$

The vector of parameters is $\boldsymbol{\theta} \equiv (\theta_1, \theta_2, \theta_3, \theta_4)^\intercal$, and for testing we consider the singleton $\mathbf{p}_0 = (0.1,\ 0.9,\ 0.25,\ 1.5)^\intercal$. The domain is



(a) (b)

Fig. 6. Logistic model (19): 95% confidence ellipses for parameters obtained with the optimal designs in Table 4 for $\hat{\theta}_3$ vs. $\hat{\theta}_1$: (a) considering the stable parameters (20); (b) considering the stable parameters (21).

Table 6
Four-parameter Hill model (22): optimal designs, $\mathbf{X} = [1 \times 10^{-4}, 5.0]$, $\Delta x = 0.05$, and $\mathbf{p}_0 = (0.1,\ 0.9,\ 0.2,\ 1.5)^\intercal$.

Optimality criterion	Design				Optimum	$\varrho^\dagger$	$\kappa^\dagger$	$\varrho^\star$	$\kappa^\star$
D–	0.0001	0.1958	0.7253	5.0000	0.0828	0.8004	196.89	0.2887	2.1487
	0.2496	0.2510	0.2496	0.2498					
A–	0.0001	0.1500	0.8000	5.0000	2.7663E+2	0.7861	168.47	0.1236	1.9552
	0.1992	0.3168	0.2989	0.1851					
E–	0.0001	0.1500	0.8000	5.0000	3.9611E−3	0.7941	170.12	0.1266	2.1132
	0.1925	0.3253	0.3060	0.1763					
K–	0.0001	0.1500	0.9000	5.0000	1.5825E+2	0.7779	163.36	0.0134	1.3143
	0.2185	0.2683	0.2797	0.2335					

† — based on original model.
‡ — based on parameterized model after Taylor expansion.

$\mathbf{X} = [1.0 \times 10^{-4}, 5.0]$, and we discretize it with a uniformly distributed grid with $\Delta x = 0.05$. The optimal designs are listed in Table 6. The maximum intercorrelation coefficients obtained considering the original parameters are below 0.9 (see column 4) but the condition number is considered large, above 150 for all the criteria (see column 5), which indicates ill-conditioning in the original LS problem. The K-optimality criterion leads to the least badly scaled problem but κ is still above 160.

Now, we use the Taylor expansion approximation introduced in Section 3.3 to find a vector of stable parameters. The algebraic treatment of the Eqs. (9) leads to:

$$\theta = \begin{pmatrix} 1.0000\,\vartheta_1 \\ -0.0933\,\vartheta_1 + 0.1635\,\vartheta_2 - 0.2920\,\vartheta_3 + 1.2217\,\vartheta_4 \\ 0.1962\,\vartheta_1 + 0.3552\,\vartheta_2 - 2.3086\,\vartheta_3 + 1.7572\,\vartheta_4 + 0.25 \\ 3.0052\,\vartheta_1 - 4.9230\,\vartheta_2 + 4.6308\,\vartheta_3 - 2.7131\,\vartheta_4 + 1.5 \end{pmatrix} \tag{23}$$

The figures of merit of this new system of parameters are in columns 6–7 of Table 6. The technique is efficient in reducing both the maximum parametric correlations and the condition numbers which are now below 2.5. Fig. 7(a) shows the ellipsoids for $\hat{\theta}_4$ vs. $\hat{\theta}_2$ obtained fitting the model (22) with the original set of parameters. Fig. 7(b) shows the ellipses with the set of stable parameters (23). The axes of the confidence regions are now virtually aligned with the parameter axes, supporting the conclusions from the figures of merit.

5. Robustness of the reparameterization

The algorithm we describe for model reparameterization is "local" as it depends on the support points of the optimal designs which are found for a given singleton θ (called $\mathbf{p}_0$). In this Section we test the robustness at two levels. We consider the Michaelis–Menten model (15), and submit it to two simulation tests.

First, we analyse the robustness of the transformation found (see (16)) when there is a change in the values of the parameters of the model used to generate the data. Here, we increased all the parameters by 10%, so that the new vector of parameters is $\mathbf{p}_0 = (1.1,\ 2.2,\ 5.5)^\intercal$.

Table 7
Michaelis–Menten model (15): metrics of performance when the parameters and the design space are modified while the reparameterization (16) is maintained, $\Delta x = 0.1$.

Optimality	$\mathbf{p}_0 = (1.1,\ 2.2,\ 5.5)^\intercal$, $\mathbf{X} = [0, 10]$		$\mathbf{p}_0 = (1.0,\ 2.0,\ 5.0)^\intercal$, $\mathbf{X} = [0, 20]$	
criterion	$\varrho^\dagger$	$\kappa^\dagger$	$\varrho^\ddagger$	$\kappa^\ddagger$
D–	0.0461	1.1260	0.6159	5.1442
A–	0.0640	2.0662	0.4377	4.4406
E–	0.0641	2.0747	0.4333	4.4531
K–	0.0210	2.3040	0.4338	3.3854

Table 7 shows the metrics for such a perturbation in columns 2 and 3. We note that ϱ and κ increase from the reference values (see the values in columns 6–7 of Table 2) but they are still well below the thresholds considered. The confidence ellipsoids in Fig. 8(a) also show evidence of the trends observed.

Secondly, we maintain the original vector of parameters, $\mathbf{p}_0 = (1.0,\ 2.0,\ 5.0)^\intercal$, but introduce a larger design region, $\mathbf{X} = [0, 20]$, while keeping $\Delta x = 0.1$. We find the optimal design for this larger region and use it for sampling, but stay with the reparameterization optimized for the smaller region. The performance metrics (in columns 4 and 5 of Table 7) present a remarkable increase, particularly in the value of ϱ, although the values remain within the thresholds. The ellipses of Fig. 8(b) support these conclusions. The results show that the reparameterization, although based on a local linearization, is robust against modifications in the model parameters and changes in the design region.

Relatively to the choices of the discretization step for the domain and the exact number of observations sampled from each support point, these parameters retain their usual effect from traditional optimal designs, since the orthogonalization capabilities of the proposed method are not dependent on particular choices of these two parameters. With numbers of support points/replicas similar to the values in the examples considered, the approximation error in the location of the optimal designs will be small, but in each case the orthogonality of

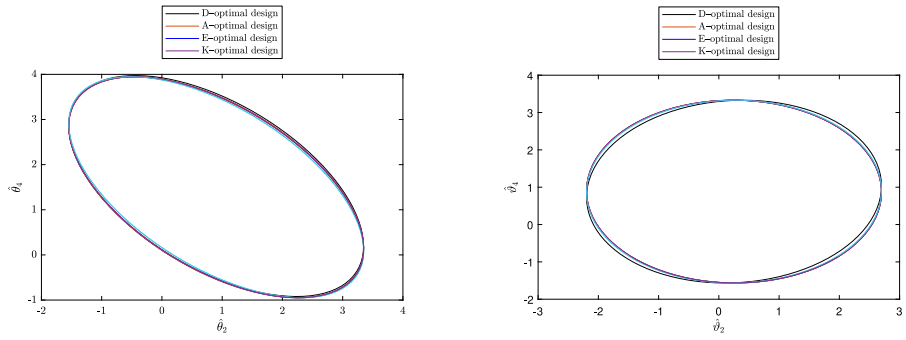


Fig. 7. Four-parameter Hill model (22): 95% confidence ellipses for parameters obtained with the optimal designs in Table 6 for $\hat{\theta}_4$ vs. $\hat{\theta}_2$: (a) considering the original parameters; (b) considering the stable parameters (23).

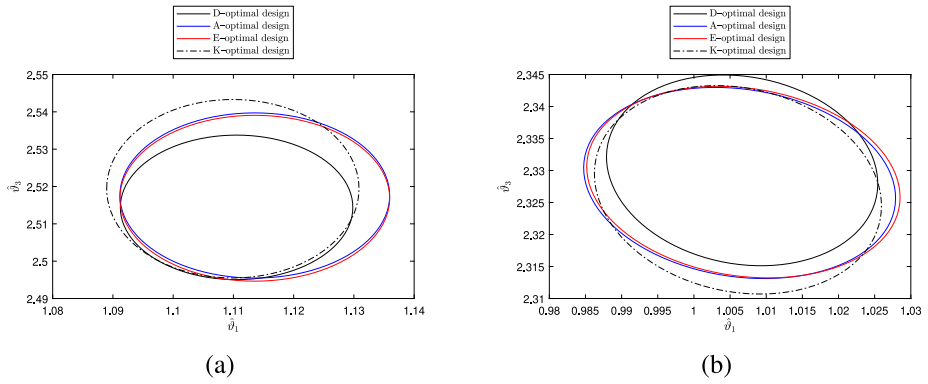


Fig. 8. Michaelis–Menten model (15): 95% confidence ellipses for parameters obtained with the optimal designs in Table 2 for $\hat{\vartheta}_3$ vs. $\hat{\vartheta}_1$: (a) considering $\mathbf{p}_0 = (1.1,\ 2.2,\ 5.5)^\intercal$ and $\mathbf{X} = [0, 10]$; (b) considering $\mathbf{p}_0 = (1.0,\ 2.0,\ 5.0)^\intercal$ and $\mathbf{X} = [0, 20]$.

the parameters estimated can be guaranteed through the exact solution of (9).

6. Results interpretation and overall effectiveness of the procedure

For the proposed methodology, given its practical ability to reformulate the original problems into transformed versions where essentially good orthogonality properties were observed with all of the examples considered, two significant questions arise, namely: (i) whether this property can be generalized for all classes of models, including cases where severe ill-conditioning or even where non-identifiable parameters might exist; (ii) the exact basis for the effectiveness of this procedure. The answer to both of these questions is interrelated, and can be elucidated from the structure of Eqs. (9).

We first consider the most common case where $n_s = n_\theta$, for saturated designs where the number of support points in the optimal design correspond to the number of parameters in the model. Since the transformed model seeks to produce responses (y) equal to the values of the model parameters ($\boldsymbol{\vartheta}$) at the chosen sampling points, its original form (1) can be replaced at these points by the equivalent form

$$y(x) = \begin{cases} \vartheta_1, & \text{if} \quad x = x_1 \\ \vartheta_2, & \text{if} \quad x = x_2 \\ \vdots & \\ \vartheta_{n_\theta}, & \text{if} \quad x = x_{n_\theta}. \end{cases} \tag{24}$$

This occurs if the transformation (10) is obtained without approximation (Approach 1), and also for the linearization considered in Approach 2, provided that the first order approximation involved is computed exactly. Eq. (24) corresponds to a linear model of the form (2) where the design matrix has a special block structure of the form $X = \begin{bmatrix} H_1 & H_2 & \cdots & H_{n_\theta} \end{bmatrix}^\intercal$ and each block $H_i \in R^{n_{o,i} \times n_\theta}$ has a single unitary column i and the remaining columns are filled by zeros, e.g.,

$$H_1 = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & \cdots & 0 \end{bmatrix}, \qquad H_2 = \begin{bmatrix} 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 1 & \cdots & 0 \end{bmatrix}, \qquad \text{etc.},$$

where $n_{o,i}$ is the number of observations considered at the support point i. This simplifies the analytic solution of the normal equations resulting from the least-squares formulation of the model (2). In this case the estimation of each parameter becomes decoupled, and its least squares estimation corresponds to the arithmetic mean of a subset of the samples, that is

$$\hat{\vartheta}_i = \frac{\sum_i^{n_{o,i}} g_i(x_i, \boldsymbol{\vartheta})}{n_{o,i}}, \quad i \in [\![n_\theta]\!].$$

Hence the application of the transformation (9) allows the replacement of the computation of the least squares solution on the original model (1) by the equivalent solution of the simultaneous set of Eqs. (10). For linear models of the form (2) the effort is comparable, since the solution of (10) also involves the solution of a linear set of equations, of the same size. For nonlinear models, (10) represents the solution of a set of nonlinear equations, also comparable in terms of numerical effort to the numerical solution of the nonlinear least squares problem. When the analytic solution of (10) is readily available, the application of this transformation has the advantage of allowing the estimation of the transformed parameters without collinearity; consequently, it takes a numerically more stable form. In either situation of Approaches 1 and 2, the computation of the model functions $f(x, \boldsymbol{\theta})$ depends on knowing the exact value of the parameters $\boldsymbol{\theta}$, which is

usually not available until the parameter estimation task is completed. This suggests the possible application of the transformation algorithm iteratively, where the present estimate of the values of the parameters is used to provide a local reparameterization to improve these estimates; each step of the iterative procedure includes the model reparameterization and updating the parameter estimates. Although conceptually feasible, the convergence properties of this approach were not studied in this paper.

In the cases where $n_s > n_\theta$, more support points are required in the optimal design than the number of parameters in the model. This situation is uncommon, since it is indicative of structural singularities in the global FIM matrix (3) constructed from the support points selected. As mentioned previously, it is possible here to select a subset of n_θ support points to write the Eqs. (24). Thus the predictive model equations at the selected support points would also have the linear form (24), but the model equations for the remaining support points would retain the original form (1) or (2). In this case, a full parameter orthogonalization would therefore not be reached, although a reduction in the overall parameter collinearity would still be achieved. Since this configuration occurs rarely in practice, a detailed study of this situation was not considered in this work. A possible extension to models resulting in singular FIMs may require working with the Moore–Penrose generalized inverse to generate the (optimal) set of support points to construct the response curve and inverting the equalities representing the stable transformation.

Now, the effect of the model reparameterization on the possible ill-conditioning of the system relative to the model parameters can be described. Since this effect is common to both linear and nonlinear model structures, it will be examined by considering the least squares solution of the predictive Eqs. (2)

$$\min_{\theta} \quad \|y - X\theta\|_2,$$

dropping the $'$ superscripts for simplicity of notation. Ill-condition or non-unique identifiability of the parameters θ can be diagnosed when the design matrix X has a high condition number or has linearly dependent columns, respectively. Contrarily, a sufficient condition for this least squares problem to be perfectly conditioned is that the matrix X is the identity matrix (or the identity matrix multiplied by some constant).

The proposed reformulation rewrites (9) as $\vartheta = X\theta$ and (10) produces $\theta = X^+\vartheta$ where the matrix $X^+ \in \mathbb{R}^{n_o \times n_\theta}$ corresponds to the pseudo inverse or Moore–Penrose inverse of the design matrix X. Thus, the model reparameterization (9) has the basic effect of producing a suitable a priori model inversion of the original model equations, so that the resulting estimation problem is always numerically well conditioned. A similar effect is also observed with nonlinear models of the form (1), this time involving the composition of the functions $f(\bullet)$ and $g(\bullet)$ as defined, when the equivalent transformed model can be written in the form (24). If the linear model is originally ill-conditioned with respect to the "original" parameters it is first transformed to a always well conditioned model in the new parameter space containing ϑ, where these parameters can be estimated without major difficulties, producing relatively small confidence regions with low eccentricity. However, in this case the inverse transformation matrix X^+ will also be ill-conditioned, and therefore larger and intercorrelated confidence regions for the original parameters θ will originate if these results are propagated back to the original parameter space. This can be easily seen for instance from a singular value decomposition $X = U\Sigma V^\intercal$, leading to $X^+ = V\Sigma^+U^\intercal$, where the singular value matrix Σ^+ is obtained from Σ by taking the reciprocal of each non-zero element on the diagonal of Σ, leaving the zeros in place, and then transposing the resulting matrix. If Σ is originally ill-conditioned, the transformation Σ^+ will also be ill-conditioned. However, the update of the uncertainty regions for the problem will only occur during the transformation (10), which is handled after the parameter estimation procedure has been completed. Thus, the parameter estimation for these systems can be performed in a controlled manner in the transformed parameter space, although the equivalent final confidence regions in the original parameter space will always reflect the effect of the uncertainties. Following a similar reasoning, it is also possible to conclude that if the original model has non uniquely identifiable parameters, this type of transformations will not be able to change this situation, in the sense that the reparameterized model will also have the same number of non uniquely identifiable parameters, since the number of non-zero elements in the matrices Σ and Σ^+ is equal.

7. Conclusions

We have considered the problem of robustifying LS fitting in nonlinear regression through the calculation and use of orthogonal (independent) parameters, which improve the computational efficiency and improve the local model identifiability. The strategy, based on model reparameterization, finds a transformation of the original set of parameters to a stable vector. Our approach constructs a response curve that passes through a given set of points of the regressor space. This curve allows the expression of the original set of parameters as a vector of functions of orthogonal (or nearly orthogonal) parameters. To automate the procedure we use the support points of the K-optimal design, obtained via Semidefinite Programming, to construct the analytic response surface. We propose two distinct approaches to handle the construction of stable transformations: (i) based on the algebraic manipulation of the model equations at the support points; and (ii) using a first-order Taylor approximation to the model. To compare the effect of reparameterization we use three metrics (i) the maximum parametric intercorrelation at LS convergence; (ii) the condition number of the covariance matrix; and (iii) the orientation and eccentricity of the parametric confidence region.

The results in Section 4 show the application to four models and illustrate the strong improvement in the orthogonality of the new parameters. Using the Michaelis–Menten model as an example, Section 5 briefly analyses the robustness of the reparameterization to modifications in the parameters and the design region. Both analyses show that the reparameterization is robust despite being based on a locally optimal design. Thus the reparameterization approach proposed herein can handle satisfactorily and systematically problems where the parameter estimates are highly collinear, typical of models that produce FIMs with a large (but finite) condition number.

We have considered the support points of the K-optimal design of experiments for constructing the surface response to find the transformation used for model reparameterization. The main reason for choosing the K-optimality criterion is that it allows minimizing an upper bound on the sensitivity of the estimated parameters in the original model to observational noise; that is, the support points in the covariance space are those that minimize the condition number of the covariance matrix in LS fitting. However, the orthogonality of the transformed parameters is guaranteed by definition and depends mainly on the invertibility of the set of equations (or their approximation) representing the surface response. Consequently, the support points maximizing other information optimality criteria or even equispaced points, as in Ross [24], can also be used. Practically, the points used only have to assure the invertibility of the system of algebraic equations, and other more accurate approaches can be used in this step, such as the numerical parameterization of the solution.

At the end of a discussion of ill-conditioning in nonlinear models, Seber and Wild [1, p. 118] write "good experimental design can reduce the problem, but it may not be able to eliminate it". In this paper we have shown that parameter transformation, coupled with appropriate experimental design, can effectively deal with the problem of estimation of ill-conditioned models through the suitable characterization of an equivalent parameter space for the original problem.

Code availability

The code used for obtaining the results in the paper was developed in `Matlab`® (version 2022). To run it requires the installation of the convex optimization environment `cvx` [34] (version 2.2) and the solver `Mosek` [35] (version 9). A sample is presented in Appendix B of the Supplementary Material.

CRediT authorship contribution statement

Belmiro P.M. Duarte: Conceptualization, Methodology, Software, Writing, Reviewing. **Anthony C. Atkinson:** Conceptualization, Methodology, Writing, Validation, Reviewing. **Nuno M.C. Oliveira:** Conceptualization, Methodology, Writing, Validation, Reviewing.

Declaration of competing interest

The authors whose names are listed immediately below certify that they have NO affiliations with or involvement in any organization or entity with any financial interest (such as honoraria; educational grants; participation in speakers' bureaus; membership, employment, consultancies, stock ownership, or other equity interest; and expert testimony or patent-licensing arrangements), or non-financial interest (such as personal or professional relationships, affiliations, knowledge or beliefs) in the subject matter or materials discussed in this manuscript.

Data availability

No data was used for the research described in the article.

Appendix A

A.1. Formulations to determine the optimal allocation via semidefinite programming

Here we list the SDP formulations for the D-, A- and E-optimality criteria. The first three were introduced in Vandenberghe and Boyd [20, 36] and Ben-Tal and Nemirovski [17]. We start with the formulation for D-optimal designs:

$$\begin{aligned}
\text{Opt} \equiv \max_{\xi,B} \quad & t && \text{(A.1a)}\\
\text{s.t.} \quad & \begin{pmatrix} M(\xi) & B^\intercal \\ B & \text{diag}(B) \end{pmatrix} \succeq 0_{n_\theta} && \text{(A.1b)}\\
& t \le \prod_{i=1}^{n_\theta} B_{i,i}^{1/n_\theta} && \text{(A.1c)}\\
& \sum_{i=1}^{k} w_i = 1 && \text{(A.1d)}\\
& 0 \le w_i \le 1, \quad i \in [\![k]\!]. && \text{(A.1e)}
\end{aligned}$$

The formulation for computing A-optimal designs is:

$$\begin{aligned}
\text{Opt} \equiv \min_{\xi,B} \quad & t && \text{(A.2a)}\\
\text{s.t.} \quad & \begin{pmatrix} M(\xi) & I_{n_\theta} \\ I_{n_\theta} & B \end{pmatrix} \succeq 0_{2\times n_\theta} && \text{(A.2b)}\\
& t \ge \sum_{i=1}^{n_\theta} B_{i,i} && \text{(A.2c)}\\
& \sum_{i=1}^{k} w_i = 1 && \text{(A.2d)}\\
& 0 \le w_i \le 1, \quad i \in [\![k]\!], && \text{(A.2e)}
\end{aligned}$$

Finally, for E-optimal designs, we have:

$$\begin{aligned}
\text{Opt} \equiv \max_{\xi,t} \quad & t && \text{(A.3a)}\\
\text{s.t.} \quad & M(\xi) - t\, I_{n_\theta} \succeq 0_{n_\theta} && \text{(A.3b)}\\
& \sum_{i=1}^{k} w_i = 1 && \text{(A.3c)}\\
& 0 \le w_i \le 1, \quad i \in [\![k]\!]. && \text{(A.3d)}
\end{aligned}$$

Appendix B. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.chemolab.2023.104874.

References

[1] G.A.F. Seber, C.J. Wild, Nonlinear Regression, John Wiley & Sons, New York, 2003.

[2] D.C. López, T. Barz, S. Korkel, G. Wozny, Nonlinear ill-posed problem analysis in model-based parameter estimation and experimental design, Comput. Chem. Eng. 77 (2015) 24–42.

[3] F. Baty, C. Ritz, S. Charles, M. Brutsche, J.-P. Flandrois, M.-L. Delignette-Muller, A toolbox for nonlinear regression in R: the package `nlstools`, J. Stat. Softw. 66 (2015) 1–21.

[4] S. Huet, A. Bouvier, M.-A. Poursat, E. Jolivet, A. Bouvier, Statistical Tools for Nonlinear Regression: A Practical Guide with S-PLUS and R Examples, Springer, 2004.

[5] A. Celminš, Least square model fitting with transformation of variable, J. Stat. Comput. Simul. 14 (1) (1981) 17–39.

[6] M.H. Kutner, C.J. Nachtsheim, J. Neter, W. Li, Applied Linear Statistical Models, fifth ed., Irwin Chicago, 1996.

[7] G.E. Box, D.R. Cox, An analysis of transformations, J. R. Stat. Soc. Ser. B Stat. Methodol. 26 (2) (1964) 211–243.

[8] G.H. Golub, V. Pereyra, The differentiation of pseudo-inverses and nonlinear least squares problems whose variables separate, SIAM J. Numer. Anal. 10 (2) (1973) 413–432.

[9] G.J.S. Ross, The efficient use of function minimization in non-linear maximum-likelihood estimation, J. R. Stat. Soc. Ser. C. Appl. Stat. 19 (3) (1970) 205–221.

[10] C.-L. Tsai, Power transformations and reparameterizations in nonlinear regression models, Technometrics 30 (4) (1988) 441–448.

[11] J.-S. Chen, R.I. Jennrich, Transformations for improving linearization confidence intervals in nonlinear regression, J. Amer. Statist. Assoc. 90 (432) (1995) 1271–1276.

[12] A.C. Atkinson, M. Riani, A. Corbellini, The Box-Cox transformation: review and extensions, Statist. Sci. 36 (2021) 239–255, http://dx.doi.org/10.1214/20-STS778.

[13] M. Schwaab, J.C. Pinto, Optimum reparameterization of power function models, Chem. Eng. Sci. 63 (18) (2008) 4631–4635.

[14] G.J. Ross, Prajneshu, C. Sarada, Reparameterization of nonlinear statistical models: a case study, J. Appl. Stat. 37 (12) (2010) 2015–2026.

[15] Z. Yue, X. Zhang, P. Driessche, J. Zhou, Constructing K-optimal designs for regression models, Statist. Papers (2022) 1–22.

[16] P. Xu, Experimental quality evaluation of lattice basis reduction methods for decorrelating low-dimensional integer least squares problems, EURASIP J. Adv. Signal Process. 2013 (1) (2013) 1–29.

[17] A. Ben-Tal, A.S. Nemirovski, Lectures on Modern Convex Optimization: Analysis, Algorithms, and Engineering Applications, Society for Industrial and Applied Mathematics, Philadelphia, 2001.

[18] S. Boyd, L. Vandenberghe, Convex Optimization, University Press, Cambridge, 2004.

[19] G. Sagnol, On the semidefinite representation of real functions applied to symmetric matrices, Linear Algebra Appl. 439 (10) (2013) 2829–2843.

[20] L. Vandenberghe, S. Boyd, Applications of semidefinite programming, Appl. Numer. Math. 29 (1999) 283–299.

[21] B.P.M. Duarte, W.K. Wong, Finding Bayesian optimal designs for nonlinear models: A semidefinite programming-based approach, Internat. Statist. Rev. 83 (2) (2015) 239–262.

[22] J.J. Ye, J. Zhou, Minimizing the condition number to construct design points for polynomial regression models, SIAM J. Optim. 23 (1) (2013) 666–686.

[23] Y. Ye, Interior Point Algorithms: Theory and Analysis, John Wiley & Sons, New York, 1997.

[24] G.J.S. Ross, Nonlinear Estimation, Springer Verlag, New York, NY, 1990, Springer Series in Statistics.

[25] Wolfram Research, Inc., Mathematica, version 12.0, 2022, URL https://www.wolfram.com/mathematica, Champaign, IL.

[26] F. Pukelsheim, S. Rieder, Efficient rounding of approximate designs, Biometrika 79 (1992) 763–770.
[27] H.B. Nielsen, K. Madsen, Introduction to Optimization and Data Fitting, Informatics and Mathematical Modelling, Technical University of Denmark, DTU, Richard Petersens Plads, Building 321, DK-2800 Kgs. Lyngby, 2010, p. 176.
[28] D.W. Marquardt, Generalized inverses, ridge regression, biased linear estimation, and nonlinear estimation, Technometrics 12 (3) (1970) 591–612.
[29] D.A. Belsley, E. Kuh, R.E. Welsch, Regression Diagnostics: Identifying Influential Data and Sources of Collinearity, John Wiley & Sons, 2005.
[30] M.F. Rempel, J. Zhou, On exact K-optimal designs minimizing the condition number, Comm. Statist. Theory Methods 43 (6) (2014) 1114–1131.
[31] A. Marco, J.-J. Martí, Polynomial least squares fitting in the Bernstein basis, Linear Algebra Appl. 433 (7) (2010) 1254–1264.
[32] P.D. Brubeck, Y. Nakatsukasa, L.N. Trefethen, Vandermonde with Arnoldi, SIAM Rev. 63 (2) (2021) 405–415.
[33] A.V. Hill, The possible effects of the aggregation of the molecules of hæmoglobin on its dissociation curves, J. Physiol. 40 (1910) i–vii.
[34] M. Grant, S. Boyd, Y. Ye, cvx Users Guide for cvx Version 1.22, CVX Research, Inc., 1104 Claire Ave., Austin, TX 78703-2502, 2012.
[35] E. Andersen, B. Jensen, J. Jensen, R. Sandvik, U. Worsøe, MOSEK Version 6, Technical Report TR–2009–3, MOSEK, 2009.
[36] L. Vandenberghe, S. Boyd, Semidefinite programming, SIAM Rev. 8 (1996) 49–95.