Journal of Process Control 70 (2018) 80–95

Contents lists available at ScienceDirect

Journal of Process Control

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

Set-membership nonlinear regression approach to parameter estimation

Nikola D. Perić[a], Radoslav Paulen[c], Mario E. Villanueva[b], Benoît Chachuat[a,*]

[a] Centre for Process Systems Engineering, Department of Chemical Engineering, Imperial College London, UK
[b] School of Information Science and Technology, ShanghaiTech University, Shanghai, China
[c] Faculty of Chemical and Food Technology, Slovak University of Technology in Bratislava, Slovakia

ARTICLE INFO

Article history:
Received 5 September 2017
Received in revised form 2 April 2018
Accepted 9 April 2018
Available online 25 September 2018

Keywords:
Parameter estimation
Nonlinear regression
Set-membership estimation
Statistical inference
Semi-infinite programming
Complete-search methods

ABSTRACT

This paper introduces *set-membership nonlinear regression* (SMR), a new approach to nonlinear regression under uncertainty. The problem is to determine the subregion in parameter space enclosing all (global) solutions to a nonlinear regression problem in the presence of bounded uncertainty on the observed variables. Our focus is on nonlinear algebraic models. We investigate the connections of SMR with (i) the classical statistical inference methods, and (ii) the usual set-membership estimation approach where the model predictions are constrained within bounded measurement errors. We also develop a computational framework to describe tight enclosures of the SMR regions using semi-infinite programming and complete-search methods, in the form of likelihood contour and polyhedral enclosures. The case study of a parameter estimation problem in microbial growth is presented to illustrate various theoretical and computational aspects of the SMR approach.

1. Introduction

Mathematical models capable of accurate prediction of physical phenomena have proved to be invaluable tools for engineers and scientists. In the area of process systems engineering, they routinely support the design, control and optimization of production processes, as a means of improving their economical profitability and reducing their environmental footprint. A majority of these models are nonlinear and contain adjustable parameters that need estimating from available experimental data, or else from other, more fundamental, mathematical descriptions. In this context, parameter estimation turns out to be a key step in the verification, and subsequent use, of the mathematical models.

Most commonly, parameter estimation in nonlinear models is cast as a nonlinear regression exercise, where selected parameter values are adjusted so that the model predictions match the available observations as close as possible, for instance in the least-squares or maximum-likelihood sense [1–4]. In order to avoid for the resulting parameter estimates to be biased, one can account for measurement errors in all of the variables, both independent and dependent variable observations, by following the so-called errors-in-variables approach [5,6]. This problem has been widely studied from a computational standpoint over the past decades, including the development of rigorous global optimization approaches for overcoming convergence to local optima [7,8].

Of course, there is more to model identification than just determining values for the unknown parameters. Systematic procedures have been devised to support the development and statistical verification of process models, which include testing structural identifiability, designing experiments for improved parameter precision, and inferring parameter confidence [9–12]. The focus in this paper is on the latter aspect, namely characterizing subregions in parameter space wherein the parameter values can be expected to lie. Other applications of such parameter confidence regions are in design under uncertainty [13,14], robust model predictive control [15–17], robust monitoring [18,19], and robust optimal design of experiments [20–22], to name but a few. For the scope of this paper, the emphasis is on models described by algebraic equations, but these ideas can be extended to dynamic or distributed models described by differential equations too.

Accounting for model mismatch and uncertain observations within the regression problem has spawned several schools of thought. Statistical approaches can be broadly classified as *frequentist* or *Bayesian*. The former seek to determine confidence regions around the regressed parameter values, typically a maximum-likelihood estimate, considered as the 'true' parameter values [1,2,4]. By construction, a $100(1-\alpha)\%$ frequentist confidence region comprises $100(1-\alpha)\%$ of the parameter values that would

* Corresponding author.
E-mail address: b.chachuat@imperial.ac.uk (B. Chachuat).

https://doi.org/10.1016/j.jprocont.2018.04.002

be obtained upon repetition of the parameter estimation using (hypothetical) new observations, considered as random variables. Approximate confidence regions, for instance based on the Wald test or the likelihood-ratio (LR) test, are known to converge to the exact confidence region in the limit of an infinite number of observations under certain conditions. Process modeling environments such as gPROMS and Aspen Custom Modeler have been relying on linear approximation and the Wald test to determine ellipsoidal confidence regions, a computationally efficient procedure for problems having several dozen unknown parameters, but one which may produce inaccurate results with large measurement errors and model mismatch or few measurement points. Confidence regions based on the LR test have been shown to yield superior approximations, but are computationally more involved since the corresponding parameter regions are complex sets in general (e.g., nonconvex, not simply connected) [23,24].

In practice, the term $100(1-\alpha)\%$ confidence region is often misused to refer to the range of parameter values that include $100(1-\alpha)\%$ of their probability distribution [25]. This description corresponds to so-called $100\,(1-\alpha)\%$ credible regions instead, which are defined in the Bayesian inference approach [26]. Bayesian estimation uses the available observations to construct a probability distribution of the parameters, called posterior distribution, based on a likelihood function and a prior probability distribution of the same parameters. In essence, this approach thus considers the unknown parameter values as random variables. Sampling-based techniques such as Markov-Chain Monte-Carlo (MCMC) [27,28] provide a means of constructing (approximate) credible regions, although the computational effort can become prohibitive for problems having upwards of 10 parameters [29]. A most probable estimate can be determined from the posterior distribution, which also corresponds to a maximum-likelihood estimate for a flat prior. Albeit classical frequentist and Bayesian inference regions can be reconciled in special cases, no equivalence can be drawn in general since Bayesian inference incorporates problem specific contextual information from the prior distribution, whereas frequentist inference is solely based on the data; see, e.g., [30, Chapter 5]. The debate on whether to use frequentist or Bayesian statistical inference continues to this day [25,31], but its intricacies are beyond the scope of this paper.

Regardless of whether a mathematical model's structure is correct or not, a frequentist confidence region will normally converge to the maximum- likelihood estimate as the number of observations increases. Likewise, a Bayesian posterior will normally converge to a point mass that corresponds to a most probable estimate, i.e., a point that maximizes the probability of the data given the (possibly wrong) model. An interesting alternative to these statistical approaches is *set-membership* estimation (SME). The traditional SME setting, also called guaranteed parameter estimation (GPE), seeks to determine the set of all possible parameter values for which a model's predictions are consistent with a set of observations subject to bounded errors [32–34]. The fact that this approach does not require a statistical description of the observation errors, solely bounds, is not only less demanding, but also more realistic in many practical applications, including biological systems where the measurements are often scarce and subject to large errors [21]. Beside parameter estimation, the distinctive yes-or-no answer provided by set-membership techniques can also be used for model inconsistency detection [35,36]. One caveat here is that the set of feasible parameter values may be empty in the presence of measurement outliers or due to an inadequate description of the measurement noise, thus calling for remedial strategies [37,38]. Another key challenge in nonlinear set-membership estimation is describing the feasible parameter set accurately, while remaining computationally tractable. This challenge is in fact similar to the one faced by aforementioned statistical inference methods for describing parameter confidence sets, and it may explain why set-membership estimation has not reached a wider diffusion to this day. Existing computational strategies are limited to problem with downwards of a dozen parameters. They range from approximation using sampling-based methods, including stochastic search [39], support vector machines (SVM) [40] and MCMC [41]; to rigorous complete-search methods based on interval analysis and other set arithmetics [42–44]; and to semidefinite relaxation techniques for semi-algebraic problems [45,46].

This paper introduces *set-membership regression* (SMR), a new approach to nonlinear regression. The SMR problem seeks to determine the subregion in parameter space enclosing all (global) solutions to a nonlinear regression problem in the presence of bounded uncertainty on the observed variables. By contrast with the traditional SME setting seeking for parameter values to satisfy certain feasibility constraints, the SMR approach method seeks for parameter values to satisfy an optimality condition. To the best knowledge of the authors, this problem has not been investigated in the general nonlinear setting so far. Milanese [47] studied optimality and convergence properties of least-squares estimates in the presence of unknown bounded disturbance, but their theoretical work is limited to linear problems. This paper sets out to investigate the connections of SMR with both statistical inference and set-membership estimation approaches for nonlinear algebraic models. Another principal contribution is a computational framework to describe tight enclosures of the SMR regions using complete-search methods.

The rest of the paper is organized as follows. Section 2 starts by reviewing classical results from both areas of statistical and set-membership estimation. Section 3 introduces the SMR approach and analyzes its properties, after which numerical solution strategies are developed in Section 4. A simple case study is used throughout Sections 2–4 to illustrate the main concepts and results. Section 5 presents a more challenging estimation problem in microbial growth to demonstrate the SMR approach. Finally, Section 6 concludes the paper and discusses future research opportunities.

2. Background

Our focus throughout this paper is on explicit models in the form

$$\mathbf{y} = \mathbf{g}(\mathbf{p}, \mathbf{u}),$$

where $\mathbf{p} \in \mathbb{R}^{n_p}$ is the vector of unknown parameters; and $(\mathbf{u}, \mathbf{y}) \in \mathbb{R}^{n_u} \times \mathbb{R}^{n_y}$ is the vector of observed variables, denoted collectively by $\mathbf{x}{:=}(\mathbf{u}, \mathbf{y}) \in \mathbb{R}^{n_x}$ for convenience. Notice that $\mathbf{u}$ and $\mathbf{y}$ often correspond to (either controlled or uncontrolled) input and output variables, respectively, in a practical setup. It is also worth pointing out that many of the concepts and methods presented herein can be applied to models described by implicit equation systems, such as $\mathbf{f}(\mathbf{p}, \mathbf{x}) = \mathbf{0}$, and models comprised of differential equations too.

Suppose that n_m observations $\mathbf{x}_k^{\mathrm{m}}{:=}(\mathbf{u}_k^{\mathrm{m}}, \mathbf{y}_k^{\mathrm{m}})$ of the input–output variables are available, and assume that all of these observation errors are independent and described by the probability density functions $p(\cdot \mid \boldsymbol{\psi})$ parameterized by $\boldsymbol{\psi}$. In the error-in-variables approach [6], the reconciled values $\mathbf{u}_1, \ldots, \mathbf{u}_{n_m}$ for the observations are estimated alongside the unknown model parameters $\mathbf{p}$. The joint probability of the prediction-observation mismatch in all data points for the parameter values $\boldsymbol{\theta}{:=}(\mathbf{p}, \mathbf{u}_1, \ldots, \mathbf{u}_{n_m}) \in \mathbb{R}^{n_\theta}$ is described by the following likelihood function:

$$\mathcal{L}(\boldsymbol{\theta} \mid \mathbf{x}^{\mathrm{m}}){:=}\prod_{k=1}^{n_m} p\left(\boldsymbol{\delta}\mathbf{u}_k \middle| \boldsymbol{\psi}_{\mathbf{u}_k}\right) \prod_{k=1}^{n_m} p\left(\boldsymbol{\delta}\mathbf{y}_k \middle| \boldsymbol{\psi}_{\mathbf{y}_k}\right), \tag{1}$$

with $\boldsymbol{\delta}\mathbf{u}_k{:=}\mathbf{u}_k - \mathbf{u}_k^{\mathrm{m}}$ and $\boldsymbol{\delta}\mathbf{y}_k{:=}\mathbf{g}(\mathbf{p}, \mathbf{u}_k) - \mathbf{y}_k^{\mathrm{m}}$. The error-in-equation approach instead, considers the input measurements $\mathbf{u}_k^{\mathrm{m}}$ to be

error-free; that is, the parameter vector $\boldsymbol{\theta}$ reduces to $\mathbf{p}$, and the likelihood function simplifies to

$$\mathcal{L}(\boldsymbol{\theta} \mid \mathbf{x}^{\mathrm{m}}) := \prod_{k=1}^{n_m} p\left(\boldsymbol{\delta}\mathbf{y}_k \middle| \boldsymbol{\psi}_{\mathbf{y}_k}\right). \tag{2}$$

Nonlinear regression in the maximum-likelihood sense seeks to determine values for $\boldsymbol{\theta}$ in order to maximize $\mathcal{L}$ or, equivalently, maximize $\log \mathcal{L}$. In the error-in-variables approach, this estimation entails the solution of an optimization problem in the form of

$$\hat{\boldsymbol{\theta}} \in \underset{\mathbf{p},\mathbf{u}_1,\ldots,\mathbf{u}_{n_m}}{\operatorname{argmax}} \sum_{k=1}^{n_m} \log p(\boldsymbol{\delta}\mathbf{u}_k|\boldsymbol{\psi}_{\mathbf{u}_k}) + \log p(\boldsymbol{\delta}\mathbf{y}_k|\boldsymbol{\psi}_{\mathbf{y}_k}). \tag{3}$$

If the parameters $\boldsymbol{\psi}$ describing the error distribution are also unknown, one may either approximate their values using an ad hoc estimator, or consider them as additional variables in the problem (3) [1].

In the special case of Gaussian-distributed errors, $p(\delta_{k,i} \mid v_{k,i}) = \frac{1}{\sqrt{2\pi v_{k,i}}} \exp\left(-\frac{\delta_{k,i}^2}{2v_{k,i}}\right)$ with zero mean and variance $v_{k,i}$, the maximum-likelihood problem (3) is equivalent to the following weighted least-squares problem

$$\hat{\boldsymbol{\theta}} \in \underset{\mathbf{p},\mathbf{u}_1,\ldots,\mathbf{u}_{n_m}}{\arg\min} \sum_{k=1}^{n_m} \left(\sum_{i=1}^{n_u} \frac{\left(\delta u_{k,i}\right)^2}{v_{u_k,i}} + \sum_{i=1}^{n_y} \frac{\left(\delta y_{k,i}\right)^2}{v_{y_k,i}} \right). \tag{4}$$

While least-squares (ℓ_2) regression is optimal amongst minimum-variance mean-unbiased estimators for normally distributed observation errors, outliers can greatly distort the least-squares estimates. As an alternative, least-absolute-values (ℓ_1) fitting may be preferable in the presence of outliers or if little is known about the distribution of the errors [48,49]. The ℓ_1 regression problem reads

$$\hat{\boldsymbol{\theta}} \in \underset{\mathbf{p},\mathbf{u}_1,\ldots,\mathbf{u}_{n_m}}{\arg\min} \sum_{k=1}^{n_m} \left(\sum_{i=1}^{n_u} \frac{\left|\delta u_{k,i}\right|}{v_{u_k,i}} + \sum_{i=1}^{n_y} \frac{\left|\delta y_{k,i}\right|}{v_{y_k,i}} \right), \tag{5}$$

where standard tricks can be used to reformulate or approximate the nonsmooth absolute value term in the objective function. The solutions to the ℓ_1 regression problem (5) can also be viewed as maximum-likelihood estimates if the observation errors follow the Laplacian distribution $p(\delta_{k,i} \mid v_{k,i}) = \frac{1}{\sqrt{2v_{k,i}}} \exp\left(-|\delta_{k,i}|\sqrt{\frac{2}{v_{k,i}}}\right)$ with zero mean and variance $v_{k,i}$. An ℓ_∞ regression problem can be constructed in a similar way [49].

2.1. Statistical inference

Classical frequentist confidence inference proceeds in two steps: (i) solve a regression problem, e.g., to determine a most-likely parameter estimate as described above; and (ii) construct confidence regions around this estimate.

Under the assumption that $\hat{\boldsymbol{\theta}}$ matches the (unique) 'true' value of the model parameters, both the *likelihood subset ratio statistic* $-2\log[\mathcal{L}(\boldsymbol{\theta} \mid \mathbf{x}^{\mathrm{m}})/\mathcal{L}(\hat{\boldsymbol{\theta}} \mid \mathbf{x}^{\mathrm{m}})]$, and the *Wald subset statistic*[1] $(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}})^T \mathbf{V}_{\hat{\boldsymbol{\theta}}}^{-1} (\boldsymbol{\theta}-\hat{\boldsymbol{\theta}})$, follow a chi-squared distribution with n_θ degrees of freedom with an increasing sample size $n_m \to \infty$ [4].

These asymptotic confidence results can be used to obtain (approximate) $100(1-\alpha)\%$ confidence regions, with the usual frequentist interpretation that the probability for a random confidence region to cover the true value of $\boldsymbol{\theta}$ is, in large samples, equal to $1-\alpha$ [24]:

- $100(1-\alpha)\%$ likelihood-based confidence region:

$$\Theta_{\mathrm{L}} := \left\{ \theta \in \Theta_0 \;\middle|\; -2\log\left(\frac{\mathcal{L}(\theta \mid \mathbf{x}^m)}{\mathcal{L}(\hat{\theta} \mid \mathbf{x}^m)}\right) \le \chi^2_{n_\theta}(1-\alpha) \right\} \tag{7}$$

- $100(1-\alpha)\%$ normal-theory (Wald) confidence region:

$$\Theta_{\mathrm{W}} := \left\{ \theta \in \Theta_0 \;\middle|\; (\theta-\hat{\theta})^T \mathbf{V}_{\hat{\theta}}^{-1} (\theta-\hat{\theta}) \le \chi^2_{n_\theta}(1-\alpha) \right\} \tag{8}$$

where $\Theta_0 \subseteq \mathbb{R}^{n_\theta}$ denotes the allowable (prior) parameter set; and $\chi^2_{n_\theta}(1-\alpha)$ is the $1-\alpha$ quantile of the chi-squared distribution with n_θ degrees of freedom. At this point, we note that confidence intervals can be inferred from any confidence region by bounding the range of values for each parameter θ_i. In the case of the Wald approximation, explicit confidence bounds are obtained as

$$\theta_i \in \left\{ \hat{\theta}_i \pm \sqrt{[\mathbf{V}_{\hat{\boldsymbol{\theta}}}]_{i,i} \chi^2_{n_\theta}(1-\alpha)} \right\}.$$

A classical result in statistical inference is that the confidence regions (7) and (8) are asymptotically equivalent [51,52], with a convergence rate $\propto n_m^{-1}$. However, unlike the likelihood-based confidence regions, the Wald confidence regions are not invariant to a model reparameterization because of the (approximate) covariance term $\mathbf{V}_{\hat{\boldsymbol{\theta}}}$. Conversely, computing a Wald confidence region is straightforward, whereas describing a likelihood-based confidence region for a nonlinear model is generally a hard task since this region may not be convex or not even simply connected.

Unlike the frequentist view, Bayesian estimation treats the parameters as random variables, whose (posterior) probability distribution, $p(\boldsymbol{\theta} \mid \mathbf{x}^{\mathrm{m}})$ can be inferred from Bayes' theorem,

$$p(\boldsymbol{\theta} \mid \mathbf{x}^{\mathrm{m}}) \propto \mathcal{L}(\boldsymbol{\theta} \mid \mathbf{x}^{\mathrm{m}}) \;\; p(\boldsymbol{\theta}), \tag{9}$$

where $p(\boldsymbol{\theta})$ is the so-called prior density of the parameters. Any subset $\Theta_{\mathrm{B}} \subseteq \mathbb{R}^{n_\theta}$ such that

$$\int_{\Theta_{\mathrm{B}}} p(\boldsymbol{\theta}|\mathbf{x}^{\mathrm{m}}) = 1-\alpha \tag{10}$$

is called a $100(1-\alpha)\%$ credible set. One particular kind of credible sets is the *highest posterior density* (HPD) set, given by

$$\Theta_{\mathrm{B}} := \{\boldsymbol{\theta} \mid p(\boldsymbol{\theta}|\mathbf{x}^{\mathrm{m}}) \ge \pi_\alpha\}, \tag{11}$$

where π_α is the largest value for which (10) holds. When a sampling approach is applied to estimate the posterior, for instance a MCMC sampler, the value of π_α can be estimated from a procedure that examines all available samples of $p(\boldsymbol{\theta} \mid \mathbf{x}^{\mathrm{m}})$ [28]. It is also worth mentioning that complete-search approaches to enclosing credible sets have been proposed as well [53,54].

The connections between Bayesian and non-Bayesian statistical inference have been studied since the 1960s, for instance with regards to matching credible and confidence intervals [55,56]; or, more recently, in order to reconcile Bayesian and frequentist higher-order asymptotic expansions for predictive probability densities [57]. In linear regression problems with normally distributed measurement errors, the Bayesian posterior takes the form of a multivariate Gaussian centered at the maximum-likelihood estimate and with covariance matrix $\mathbf{V}_{\hat{\boldsymbol{\theta}}}$ for non-informative priors, so the HPD credible regions match their frequentist counterparts. More generally, such matching can be made in cases where the Bayesian prior is invariant to model reparameterization, which is the case for Jeffreys or reference priors [58].

[1] The covariance matrix $\mathbf{V}_{\hat{\boldsymbol{\theta}}} \in \mathbb{S}_+^{n_\theta \times n_\theta}$ for the parameters at $\hat{\boldsymbol{\theta}}$ can be approximated in various ways [50], which are asymptotically equivalent; for instance [1, § 7-5],

$$\mathbf{V}_{\hat{\boldsymbol{\theta}}} := \hat{\mathcal{H}}^{-1} \frac{\partial^2 \log \mathcal{L}(\hat{\boldsymbol{\theta}}|\mathbf{x}^{\mathrm{m}})}{\partial \hat{\mathbf{x}} \partial \hat{\mathbf{x}}} \mathbf{V}_{\mathbf{e}} \frac{\partial^2 \log \mathcal{L}(\hat{\boldsymbol{\theta}}|\mathbf{x}^{\mathrm{m}})}{\partial \hat{\mathbf{x}} \partial \boldsymbol{\theta}} \hat{\mathcal{H}}^{-1}, \tag{6}$$

where $\mathbf{V}_{\mathbf{e}} \in \mathbb{S}_+^{n_x n_m \times n_x n_m}$ stands for the covariance matrix of the observation noise, and $\hat{\mathcal{H}} := \frac{\partial^2 \log \mathcal{L}(\hat{\boldsymbol{\theta}}|\mathbf{x}^{\mathrm{m}})}{\partial \boldsymbol{\theta}^2}$ is the Hessian matrix at $\hat{\boldsymbol{\theta}}$.

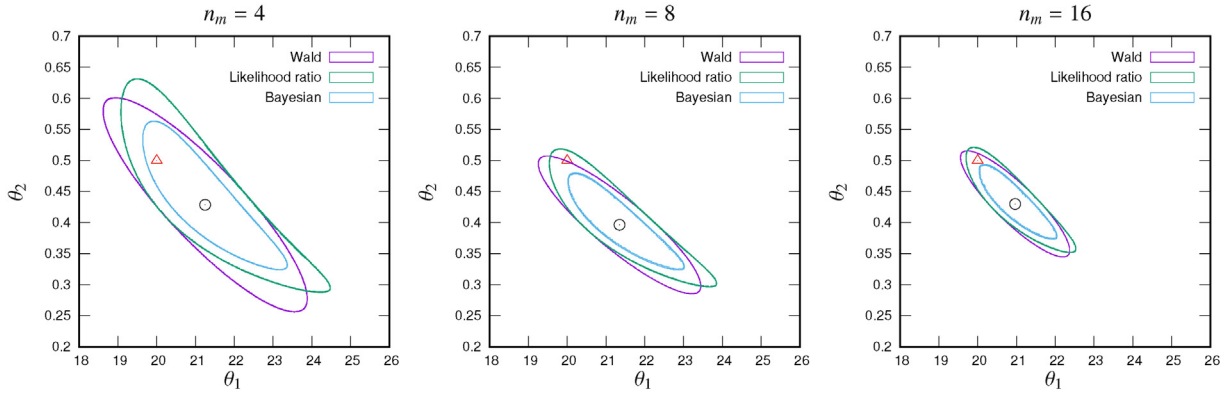

Fig. 1. 90% Wald and log-likelihood confidence regions and 90% HPD credible region for the BOD example with 4 (left), 8 (center), and 16 (right) measurement points. The circles and triangles represent the maximum-likelihood estimates and the real parameter values, respectively.

For simplicity, our focus in this paper is limited to uniform prior distributions with compact supports. Although such priors fail to be invariant under reparameterization, the resulting HPD sets correspond to contour levels of the likelihood function, similar to the likelihood-based confidence regions.

2.2. Set-membership estimation

The usual GPE problem in set-membership estimation seeks to determine a parameter subregion such that the predicted input–output observations are consistent with their matching measurements within given error bounds [32,33],

$$\Theta_G := \left\{ \boldsymbol{\theta} \in \Theta_0 \,\middle|\, \left(\boldsymbol{\delta u}_k, \boldsymbol{\delta y}_k\right)_{1\le k\le n_m} \in \mathbf{E} \right\}. \tag{12}$$

Here the error set $\mathbf{E} \subset \mathbb{R}^{n_x n_m}$ may be any compact set and does not need a statistical description of the uncertainty. In the usual scenario where independent error bounds $\pm\mathbf{e}_{\mathbf{u}_1}, \pm\mathbf{e}_{\mathbf{y}_1}, \ldots, \pm\mathbf{e}_{\mathbf{u}_{n_m}}, \pm\mathbf{e}_{\mathbf{y}_{n_m}}$ are given for each of the measurements, the set-membership estimation problem reads

$$\Theta_G := \left\{ \boldsymbol{\theta} \in \Theta_0 \,\middle|\, \begin{array}{l} \forall k = 1, \ldots, n_m, \\ -\mathbf{e}_{\mathbf{u}_k} \le \boldsymbol{\delta u}_k \le \mathbf{e}_{\mathbf{u}_k}, \quad -\mathbf{e}_{\mathbf{y}_k} \le \boldsymbol{\delta y}_k \le \mathbf{e}_{\mathbf{y}_k} \end{array} \right\}.$$

If statistical information about the observation error is nonetheless available, for instance a uniform or q-Gaussian probability distribution with compact support, one may take **E** directly as this support set. Even when the distribution support is not compact, one could decide to exclude those scenarios having a probability lower than a given threshold and use the corresponding HPD credible region as the error set **E**; see, e.g., [59].

It is not difficult to imagine a situation whereby no parameter value in Θ_0 can be found such that the model predictions are consistent with the observations for a given error set **E**, i.e., the guaranteed parameter region (12) is empty. This may happen in the presence of measurement outliers, or could be caused by a large model mismatch. The former situation is common with experimental data, e.g., due to a failing or drifting sensor. Methods have been developed for robustifying set-membership estimation against outliers [37,38], alongside classical approaches to detecting outliers [60]. Moreover, one can take advantage of the latter situation, for instance to invalidate candidate models that would present a systematic offset with a certain set of observations [35,36], typically after checking for outliers [38]. Another appeal of set-membership estimation lies in its ability to detect a lack of identifiability in parametric models, that is, when model responses corresponding to distinct parameter values are indistinguishable [9].

The vast majority of computational studies in set- membership estimation uses exhaustive-search techniques based on interval analysis or other set arithmetics to describe the parameter regions (12) [42–44,61] . A current bottleneck of these approaches is their applicability to problems having no more than 5–10 parameters. However, if one is ready to abandon guarantees, sampling-based techniques such as SVM or MCMC can be used to approximate the parameter regions, and these remain applicable for black-box models too [40,41].

Illustrative example. We use a simple estimation problem adapted from [3] to illustrate the main approaches described in this background section, and we use the same problem to illustrate the main properties of the SMR framework developed later on in Sections 3 and 4. The model describes the dynamic evolution of biological oxygen demand (BOD), c in a wastewater sample,

$$c = \theta_1(1 - \mathrm{e}^{-\theta_2 t}), \tag{13}$$

with parameters $(\theta_1, \theta_2) \in [0, 50] \times [0, 2]$, and time $t \ge 0$. For this problem, data points $(t_k^{\mathrm{m}}, c_k^{\mathrm{m}})$ have been generated by simulating the model (13) for the parameter values $\theta_1 = 20$ and $\theta_2 = 0.5$, and corrupting these values with a Gaussian white noise with variance $\sigma_c^2 = 1$. These data are reported in Appendix B for the sake of reproducibility.

Both 90% confidence regions and 90% HPD credible regions are compared in Fig. 1, in the case of an ℓ_2-regression problem. Various sets of measurements are considered, namely $n_m = 4$ measurement points (every other day), 8 measurement points (every day), and 16 measurement points (twice a day). The asymptotic convergence of the Wald and likelihood-based confidence regions with an increasing number of measurements is clearly visible. The HPD credible sets shown on these plots are generated from a flat prior, and are consistently smaller than their confidence counterparts; HPD credible sets constructed from a non-informative Jeffreys prior (not shown on the plots) would be identical to the likelihood-based confidence regions.

A comparison between guaranteed parameter regions for the same three sets of measurements, but corresponding to different measurement error sets in (12), is shown in Fig. 2. The first measurement error set corresponds to the usual assumption of independent error bounds on each measurement,

$$\mathbf{E}_1 := \left\{ \mathbf{e}_c \in \mathbb{R}^{n_m} \mid \ \forall k = 1 \ldots n_m, \quad e_{c,k}^2 \le \chi_1^2(0.9)\sigma_c^2 \right\}, \tag{14}$$

here for 90% confidence bounds, so that $\chi_1^2(0.9)\sigma_c^2 \approx 2.706$. Notice how the corresponding guaranteed parameter sets shrink when more measurements are added, as it becomes more challenging for the model predictions to match a larger measurement set in the presence of measurement noise. Such guaranteed parameter

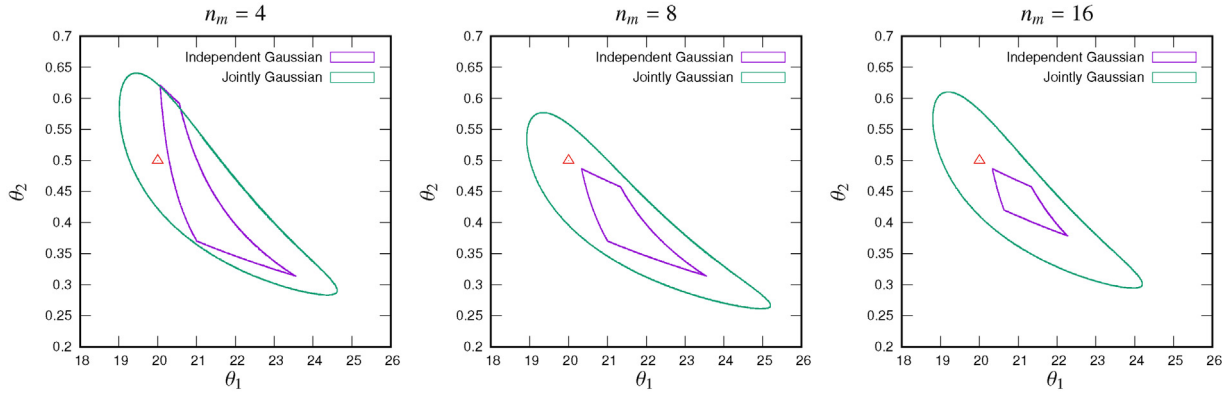

Fig. 2. Guaranteed parameter regions for the BOD example with 4 (left), 8 (center), and 16 (right) measurement points. The regions are for two measurement error sets corresponding to 90% HPD regions on either independent or jointly Gaussian distributions. The triangles represent the real parameter values.

regions could even be empty, which happens for instance with $e_{c_k}^2 \leq 1$ in (14), corresponding to 68% (1-sigma) confidence bounds. Also notice that the real parameter value (20, 0.5) lies outside the guaranteed regions due to the large measurement noise.

The other measurement error set in Fig. 2 is chosen as the HPD set of a joint Gaussian distribution,

$$\mathbf{E}_2 := \left\{ \mathbf{e}_c \in \mathbb{R}^{n_m} \;\middle|\; \mathbf{e}_c^T \mathbf{e}_c \leq \chi^2_{n_m}(0.9)\sigma_c^2 \right\}, \tag{15}$$

again for a 90% confidence limit. Guaranteed parameter sets so constructed do not shrink significantly as more measurements are added into the estimation problem, and they are thus more resilient to measurement noise than their counterpart sets constructed with independent error bounds on each measurement. This higher resilience is essentially due to an enlarged, and hence more flexible, measurement error set $\mathbf{E}_2$ compared to $\mathbf{E}_1$.

3. Set-membership nonlinear regression

The developed *set-membership regression* (SMR) approach seeks to describe the subregion Θ_R in parameter space enclosing *all* (global) solutions to a nonlinear regression problem under *all* possible measurement uncertainty scenarios. Given a bounded uncertainty set $\mathbf{E} \subset \mathbb{R}^{n_x n_m}$ on the observation errors, the SMR region Θ_R is mathematically defined as

$$\Theta_R := \{ \boldsymbol{\theta} \in \Theta_0 \,|\, \boldsymbol{\theta} \in \underset{\boldsymbol{\varpi}}{arg\,\max} \log \mathcal{L}(\boldsymbol{\varpi} \mid \mathbf{x}^m + \mathbf{e}) \;\; \exists \mathbf{e} \in \mathbf{E} : \}. \tag{16}$$

In the context of the ℓ_2-regression problem (4), SMR specializes to

$$\Theta_R^{\ell_2} := \left\{ \boldsymbol{\theta} \in \Theta_0 \;\middle|\; \begin{array}{l} \exists \left(\mathbf{e}_{\mathbf{u}_1}, \mathbf{e}_{\mathbf{y}_1}, \ldots, \mathbf{e}_{\mathbf{u}_{n_m}}, \mathbf{e}_{\mathbf{y}_{n_m}} \right) \in \mathbf{E} : \\ \boldsymbol{\theta} \in \underset{\mathbf{p}, \mathbf{u}_1, \ldots, \mathbf{u}_{n_m}}{\arg\min} \sum_{k=1}^{n_m} \left(\sum_{i=1}^{n_u} \frac{[\delta u_{k,i} - e_{u_k,i}]^2}{v_{u_k,i}} \right. \\ \left. + \sum_{i=1}^{n_y} \frac{[\delta y_{k,i} - e_{y_k,i}]^2}{v_{y_k,i}} \right) \end{array} \right\}, \tag{17}$$

and in the context of the ℓ_1- regression problem (5), to

$$\Theta_R^{\ell_1} := \left\{ \boldsymbol{\theta} \in \Theta_0 \;\middle|\; \begin{array}{l} \exists (\mathbf{e}_{\mathbf{u}_1}, \mathbf{e}_{\mathbf{y}_1}, \ldots, \mathbf{e}_{\mathbf{u}_{n_m}}, \mathbf{e}_{\mathbf{y}_{n_m}}) \in \mathbf{E} : \\ \boldsymbol{\theta} \in \underset{\mathbf{p}, \mathbf{u}_1, \ldots, \mathbf{u}_{n_m}}{\arg\min} \sum_{k=1}^{n_m} \left(\sum_{i=1}^{n_u} \frac{|\delta u_{k,i} - e_{u_k,i}|}{v_{u_k,i}} \right. \\ \left. + \sum_{i=1}^{n_y} \frac{|\delta y_{k,i} - e_{y_k,i}|}{v_{y_k,i}} \right) \end{array} \right\}. \tag{18}$$

Notice that the constraint feasibility condition in the traditional SME formulation (12) is replaced with an optimality condition in the SMR problem (16) , making the parameter regions in SMR expectedly more difficult to characterize. Numerical solution strategies for describing enclosures of an SMR region are presented later on in Section 4. The remainder of this section investigates connections between SMR and the well-established set-membership and statistical inference approaches, respectively in Sections 3.1 and 3.2.

3.1. Set-membership interpretation

By contrast with the usual approach to set-membership estimation (Section 2.2), SMR comes with a guarantee that the set Θ_R is always non-empty, no matter how large the model mismatch or the observation errors might be, since the regression problems in (16) are all feasible by construction. Therefore, the SMR formulation is inherently resilient to the presence of outlying observations, and it does not need for such outliers to be detected or removed from the observation set before computing the parameter regions [38]. In other words, the outlying observations can be dealt with directly into the SMR problem (16) via an appropriate likelihood function.

The following inclusion result holds between SMR and GPE under mild assumptions:

Theorem 1. *Suppose that the probability density functions $p(\cdot \mid \boldsymbol{\psi})$ participating in the likelihood function (1) are all maximal at 0. Then, for a given error set $\mathbf{E}$, the SMR region (16) contains the GPE region (12), $\Theta_G \subseteq \Theta_R$.*

Proof. Let $\bar{\boldsymbol{\theta}} \in \Theta_G$, so that $(\bar{\mathbf{u}}_1 - \mathbf{u}_1^m, \mathbf{g}(\mathbf{p}, \bar{\mathbf{u}}_1) - \mathbf{y}_1^m, \ldots, \bar{\mathbf{u}}_{n_m} - \mathbf{u}_{n_m}^m, \mathbf{g}(\mathbf{p}, \bar{\mathbf{u}}_{n_m}) - \mathbf{y}_{n_m}^m) \in \mathbf{E}$. It follows that $(\bar{\mathbf{u}}_1 - \mathbf{u}_1^m - \mathbf{e}_{\mathbf{u}_1}, \mathbf{g}(\mathbf{p}, \bar{\mathbf{u}}_1) - \mathbf{y}_1^m - \mathbf{e}_{\mathbf{y}_1}, \ldots, \bar{\mathbf{u}}_{n_m} - \mathbf{u}_{n_m}^m - \mathbf{e}_{\mathbf{u}_{n_m}}, \mathbf{g}(\mathbf{p}, \bar{\mathbf{u}}_{n_m}) - \mathbf{y}_{n_m}^m - \mathbf{e}_{\mathbf{u}_{n_m}}) = 0$ for some $\mathbf{e} := (\mathbf{e}_{\mathbf{u}_1}, \mathbf{e}_{\mathbf{y}_1}, \ldots, \mathbf{e}_{\mathbf{u}_{n_m}}, \mathbf{e}_{\mathbf{y}_{n_m}}) \in \mathbf{E}$. Since the probability density functions $p(\cdot \mid \boldsymbol{\psi})$ in $\mathcal{L}$ are all maximal at 0 by assumption, the log-likelihood function $\log \mathcal{L}(\cdot \mid \mathbf{x}^m + \mathbf{e})$ is (globally) maximal at $\bar{\boldsymbol{\theta}}$, and therefore $\bar{\boldsymbol{\theta}} \in \Theta_R$. □

Remark 1. The assumption on the likelihood function $\mathcal{L}$ in Theorem 1 is not very restrictive in practice. For instance, it is satisfied by both ℓ_2- and ℓ_1-regression problems in (17) and (18), so we have $\Theta_G \subseteq \Theta_R^{\ell_2}$ and $\Theta_G \subseteq \Theta_R^{\ell_1}$. It is also satisfied when the probability density functions are uniform on a compact support, as is the case with ℓ_∞-regression problems [49].

Illustrative example (continued). A comparison between GPE and SMR regions for both ℓ_1- and ℓ_2- regression is presented in Fig. 3, in the case of 8 measurements. The same measurement error sets $\mathbf{E}_1$ and $\mathbf{E}_2$ as introduced earlier in (14) and (15) are used in this comparison. For simplicity, we have applied a simple sampling procedure

to inner-approximate the SMR regions: 20,000 error vectors $\mathbf{e}_c^{(i)}$ are generated within the multi- dimensional error sets $\mathbf{E}_1$ and $\mathbf{E}_2$, here using Sobol quasi-random sampling; then, the following nonlinear regression problem is solved to global optimality to obtain a corresponding point $\boldsymbol{\theta}^{(i)} \in \Theta_R$,

$$\min_{\theta_1,\theta_2} \sum_{k=1}^{n_m} \frac{[c_k^m + e_{c,k}^{(i)} - \theta_1(1 - e^{-\theta_2 t_k^m})]^2}{\sigma_c^2}.$$

We start by noting that the inclusion result in Theorem 1 is indeed satisfied for both measurement error sets and both regression types. Moreover, the SMR regions obtained for either measurement error sets are comparable in size. In the case of independent error bounds on the measurements (set $\mathbf{E}_1$, left plot), the SMR regions do not shrink much when more measurements are added, which is unlike the corresponding GPE regions; compare Fig. 2. This also illustrates the higher resilience of SMR to noisy or outlying measurements than GPE. For both measurement error sets, the SMR-ℓ_2 regions are consistently smaller than their SMR-ℓ_1 counterparts. Interestingly, this observation is consistent with the classical Gauss-Markov theorem stating that the least-squares estimator provides the estimator with lowest variance in linear regression.

3.2. Statistical interpretation

Whenever statistical information is available for the observation errors, for instance in the form of a joint probability distribution, one may choose the error set $\mathbf{E}$ as the corresponding HPD region for a given credibility level $1-\alpha$. In the case of independent and Gaussian-distributed observation errors, such as those leading to the ℓ_2- regression problem (4), the $100(1-\alpha)$% HPD region is given by

$$\mathbf{E} := \left\{ \mathbf{e} \in \mathbb{R}^{n_x n_m} \;\middle|\; \mathbf{e}^T \mathbf{V}_\mathbf{e}^{-1} \mathbf{e} = \left\| \mathbf{V}_\mathbf{e}^{-1/2} \mathbf{e} \right\|_2^2 \leq \chi^2_{n_x n_m}(1-\alpha) \right\}, \tag{19}$$

with the diagonal error covariance matrix $\mathbf{V}_\mathbf{e} := \mathrm{diag}(\mathbf{v}_{\mathbf{u}_1}, \mathbf{v}_{\mathbf{y}_1}, \ldots, \mathbf{v}_{\mathbf{u}_{n_m}}, \mathbf{v}_{\mathbf{y}_{n_m}})$. Likewise, for Laplacian distributed errors as in the ℓ_1-regression problem (5), the $100(1-\alpha)$% HPD region comes in the form

$$\mathbf{E} := \left\{ \mathbf{e} \in \mathbb{R}^{n_x n_m} \;\middle|\; \left\| \mathbf{V}_\mathbf{e}^{-1/2} \mathbf{e} \right\|_1^2 \leq \Gamma_{n_x n_m}(1-\alpha) \right\}, \tag{20}$$

where $\Gamma_{n_x n_m}(1-\alpha)$ is the counterpart of the chi-squared value for a joint Laplacian distribution.

Notice that with the error sets in (19) and (20), the SMR regions Θ_R may not converge to a singleton (or a finite set) as more observations are added into the regression problem, since the HPD limits $\chi^2_{n_x n_m}(1-\alpha)$ and $\Gamma_{n_x n_m}(1-\alpha)$ are themselves increasing with n_m for a given confidence level $1-\alpha$. The SMR regions derived from such error sets are thus unrelated to their confidence and credible region counterparts in classical statistical inference (Section 2.1), which are both shrinking to a singleton as $n_m \to \infty$ (under certain regularity conditions). But while one would indeed expect convergence to some 'true' parameter value when a model's structure is correct, such an idea of 'true' parameter values becomes meaningless in the presence of structural model mismatch. By contrast, SMR does not make any assumption about the correctness of a model's structure, and a $100(1-\alpha)$% SMR region is comprised of those parameter values which are equally credible under the observation error set $\mathbf{E}$, in the sense of the regression problem at hand: a clear and unambiguous statistical interpretation.

To sum up, convergence of an SMR region Θ_R to a singleton is dependent on the choice of the measurement error set $\mathbf{E}$, but is unrelated to whether or not the model's structure is correct. A follow-up question then is identifying scenarios under which SMR regions would be asymptotically equivalent to classical confidence regions. The following result establishes one simple connection with the Wald confidence regions (8) under certain regularity conditions.

Theorem 2. *Let the error set in the SMR problem (16) be given by*

$$\mathbf{E} := \left\{ \mathbf{e} \in \mathbb{R}^{n_x n_m} \;\middle|\; \mathbf{e}^T \mathbf{V}_\mathbf{e}^{-1} \mathbf{e} \leq \chi^2_{n_\theta}(1-\alpha) \right\}, \tag{21}$$

for some confidence level $1-\alpha$, and covariance matrix $\mathbf{V}_\mathbf{e} \in \mathbb{S}_+^{n_x n_m \times n_x n_m}$. Assume that the likelihood function in (16) is twice continuously differentiable and the regression problems for $\mathbf{e} \in \mathbf{E}$ all have a unique, strict global optimum. Then, the SMR region Θ_R is asymptotically equivalent to the $100(1-\alpha)$% Wald confidence region Θ_W in (6) and (8),

$$d_H(\Theta_R, \Theta_W) \in \mathcal{O}(\mathrm{diam}(\mathbf{E})^2),$$

where d_H is the Hausdorff metric.

Proof. Let $\mathbf{e} \in \mathbf{E}$, and denote by $\bar{\boldsymbol{\theta}}(\mathbf{e}) \in \Theta_R$ the corresponding solution to the regression problem $\max_{\boldsymbol{\theta}} \log \mathcal{L}(\boldsymbol{\theta} \mid \mathbf{x}^m + \mathbf{e})$, so that $\frac{\partial \log \mathcal{L}}{\partial \boldsymbol{\theta}}(\bar{\boldsymbol{\theta}}(\mathbf{e}) \mid \mathbf{x}^m + \mathbf{e}) = \mathbf{0}$. Since we also have $\frac{\partial \log \mathcal{L}}{\partial \boldsymbol{\theta}}(\hat{\boldsymbol{\theta}} \mid \mathbf{x}^m) = \mathbf{0}$ at the maximum-likelihood estimate $\hat{\boldsymbol{\theta}}$, it follows by Taylor's theorem and the regularity assumptions that

$$\bar{\boldsymbol{\theta}}(\mathbf{e}) \in \hat{\boldsymbol{\theta}} - \hat{\mathcal{H}}^{-1} \frac{\partial^2 \log \mathcal{L}(\hat{\boldsymbol{\theta}} \mid \mathbf{x}^m)}{\partial \boldsymbol{\theta} \partial \mathbf{x}} \mathbf{e} + \mathcal{O}\left(\|\mathbf{e}\|^2 \right), \tag{22}$$

with $\hat{\mathcal{H}} := \frac{\partial^2 \log \mathcal{L}(\hat{\boldsymbol{\theta}} \mid \mathbf{x}^m)}{\partial \boldsymbol{\theta}^2}$. Now, let $\boldsymbol{\theta}$ be any point in Θ_R. From (22), we have

$$\boldsymbol{\theta}' = \hat{\boldsymbol{\theta}} - \hat{\mathcal{H}}^{-1} \frac{\partial^2 \log \mathcal{L}(\hat{\boldsymbol{\theta}} \mid \mathbf{x}^m)}{\partial \boldsymbol{\theta} \partial \mathbf{x}} \mathbf{e}, \tag{23}$$

for some $\boldsymbol{\theta}' \in \Theta_0$ with $\left\| \boldsymbol{\theta} - \boldsymbol{\theta}' \right\| \in \mathcal{O}(\mathrm{diam}(\mathbf{E})^2)$. The image of the error set (21) under the affine transformation (23) is an ellipsoid with center $\hat{\boldsymbol{\theta}}$ and shape matrix $\mathbf{V}_{\hat{\boldsymbol{\theta}}}$ as in (6), so that $\boldsymbol{\theta}' \in \Theta_W$. Conversely, let $\boldsymbol{\theta}'$ be any point in Θ_W, and let $\mathbf{e}$ be any point in $\mathbf{E}$ satisfying (23). Clearly, the point $\bar{\boldsymbol{\theta}}(\mathbf{e}) \in \Theta_R$ is such that $\left\| \bar{\boldsymbol{\theta}}(\mathbf{e}) - \boldsymbol{\theta}' \right\| \in \mathcal{O}(\mathrm{diam}(\mathbf{E})^2)$ by (22). □

Remark 2. In the special case of a linear regression, the equivalence between the SMR and Wald confidence regions in Theorem 2 turns out to be exact, not merely asymptotic. For an ℓ_2-regression and the model $\mathbf{y} = \mathbf{F}\boldsymbol{\theta}$, we have

$$\Theta_R^{\ell_2} = \left\{ \boldsymbol{\theta} \in \Theta_0 \;\middle|\; \left(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}} \right)^T \mathbf{F}^T \mathbf{V}_\mathbf{e}^{-1} \mathbf{F} \left(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}} \right) \leq \chi^2_{n_\theta}(1-\alpha) \right\},$$

which matches the likelihood-ratio confidence region Θ_L (2), as well as the Bayesian's HPD credible region Θ_B (11) for a uniform/non-informative prior. Both the frequentist and Bayesian inference regions are thus implied by the SMR framework in linear regression problems.

Remark 3. The key difference between the error set (21) in Theorem 2 and the $100(1-\alpha)$%-HPD region (19), is that the HPD limit in the former, namely $\chi^2_{n_\theta}(1-\alpha)$, is independent of the number of observations. This is also the reason why the error set (21) shrinks to the origin, and therefore Θ_R converges to the singleton set $\{\hat{\boldsymbol{\theta}}\}$ as $n_m \to \infty$ (under the assumptions of Theorem 2). Conversely, a $100(1-\alpha)$% confidence region may be regarded as the asymptotic equivalent to an SMR region with the confidence level $100(1-\beta)$% on the jointly Gaussian-distributed observation errors in (19) such that $\chi^2_{n_x n_m}(1-\beta) = \chi^2_{n_\theta}(1-\alpha)$. For instance, a 90%- confidence region in a two-parameter regression problem is asymptotically equivalent to an SMR region with 67%, 20% and 0.26% joint confidence for 4, 8 and 16 observations, respectively.

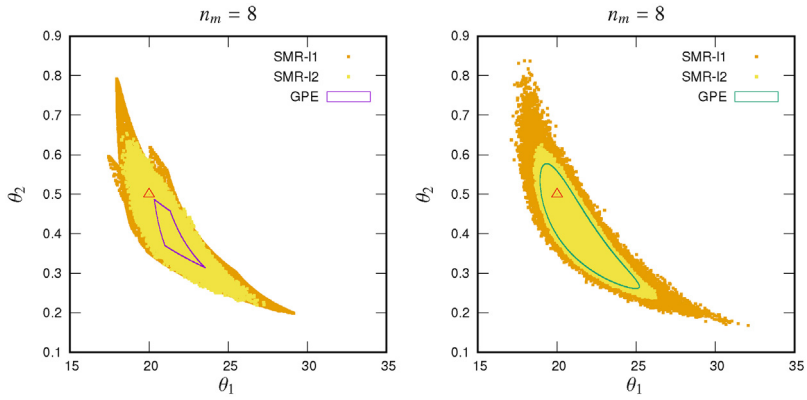

Fig. 3. Guaranteed parameter regions compared with sampled SMR-ℓ_1 and SMR-ℓ_2 regions for the BOD example with 8 measurement points. The left and right plots are for measurement error sets corresponding to 90% HPD regions on independent and joint Gaussian distributions, respectively. The triangles represent the real parameter values.

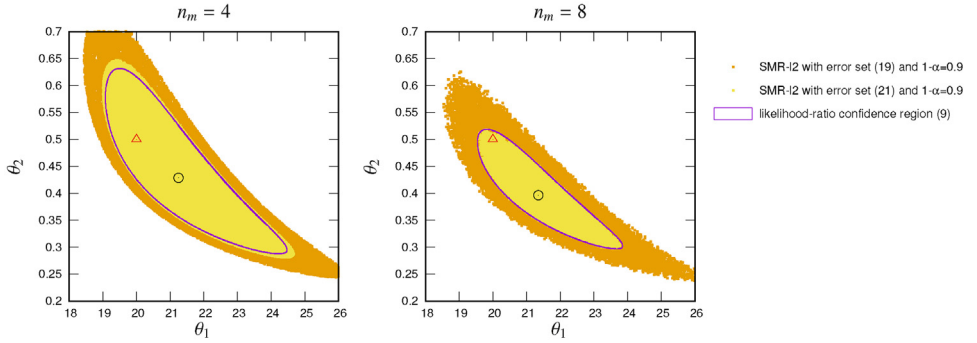

Fig. 4. 90% likelihood- ratio confidence regions compared with sampled SMR-ℓ_2 regions for the BOD example with 4 (left plot) and 8 (right plot) measurement points. The SMR regions are for two measurement error sets corresponding to the HPD regions (19) and (21) with $1-\alpha = 0.9$. The circles and triangles represent the maximum-likelihood estimates in ℓ_2- regression and the real parameter values, respectively.

Illustrative example (continued). A comparison between 90% likelihood-ratio confidence regions and two SMR-ℓ_2 regions corresponding to different measurement error sets is shown in Fig. 4, in the case of 4 and 8 measurement points. The SMR regions are inner-approximated using the same sampling strategy as previously.

The first error sets correspond to 90% HPD regions in (19) for the jointly Gaussian-distributed measurement errors—or, equivalently, the set $\mathbf{E}_2$ in (15). These SMR regions are found to be significantly larger than their 90% likelihood-ratio (or Wald) confidence counterparts. Also recall that, by Theorem 1, these SMR regions always enclose the GPE regions shown in Fig. 3 for the same error sets $\mathbf{E}_2$.

The second error sets are constructed per (21), in order to illustrate the asymptotic equivalence with classical confidence regions as established through Theorem 2; they correspond to 67% and 20% HPD regions for jointly Gaussian-distributed measurement errors with 4 and 8 measurements, respectively, as discussed in Remark 3. Such asymptotic convergence with an increasing number of measurements is clearly visible in Fig. 4, where the small discrepancy observed on the left plot for $n_m = 4$ cannot be seen anymore on the right plot for $n_m = 8$. The SMR framework is thus capable of providing equivalent confidence information as in classical statistical inference, with the attendant advantage of being able to switch between alternative error set descriptions or likelihood functions seamlessly.

4. Numerical solution and approximation

Describing the SMR region Θ_R as defined in (16) is a difficult task in general. A simple approach to enclosing Θ_R by a set of algebraic constraints, which would then allow the application of the same set-inversion techniques as for GPE (Section 2.2; Appendix A), entails a substitution of the regression problems by their optimality conditions. Since every element $\boldsymbol{\theta}$ in (the interior of) Θ_R should satisfy the first- and second-order optimality conditions

$$\frac{\partial \log \mathcal{L}(\boldsymbol{\theta} \mid \mathbf{x}^m + \mathbf{e})}{\partial \boldsymbol{\theta}} = \mathbf{0} \quad \text{and} \quad \frac{\partial^2 \log \mathcal{L}(\boldsymbol{\theta} \mid \mathbf{x}^m + \mathbf{e})}{\partial \boldsymbol{\theta}^2} \preceq 0 \tag{24}$$

for some observation error $\mathbf{e} \in \mathbf{E}$, we have

$$\left\{ \boldsymbol{\theta} \in \Theta_0 \;\middle|\; \begin{array}{l} \exists \mathbf{e} \in \mathbf{E}: \\ \dfrac{\partial \log \mathcal{L}(\boldsymbol{\theta} \mid \mathbf{x}^m + \mathbf{e})}{\partial \boldsymbol{\theta}} = \mathbf{0}, \quad \dfrac{\partial^2 \log \mathcal{L}(\boldsymbol{\theta} \mid \mathbf{x}^m + \mathbf{e})}{\partial \boldsymbol{\theta}^2} \preceq \mathbf{0} \end{array} \right\} \supseteq \Theta_R.$$

However, since the optimality conditions (24) hold for both local and global maxima of the likelihood function, as well as saddle points, this inclusion could end up being very conservative for nonlinear regression problems in general. Another important caveat with this approach is the computational penalty of applying a set-inversion algorithm in the $(n_\theta + n_x n_m)$-dimensional domain $\Theta_0 \times \mathbf{E}$, not merely in the original n_θ-dimensional domain Θ_0. The following subsections set out to develop more tractable, yet still conservative, bounding strategies to alleviate the computational burden of SMR, both in the form of confidence-like regions (Section 4.1) and polyhedral regions (Section 4.2).

4.1. Likelihood-contour enclosure

We consider the problem of enclosing the SMR region Θ_R within a confidence-like region of the form

$$\overline{\Theta}_R(\lambda) := \{\boldsymbol{\theta} \in \Theta_0 \mid \log \mathcal{L}(\boldsymbol{\theta} \mid \mathbf{x}^m) \geq \lambda\}, \tag{25}$$

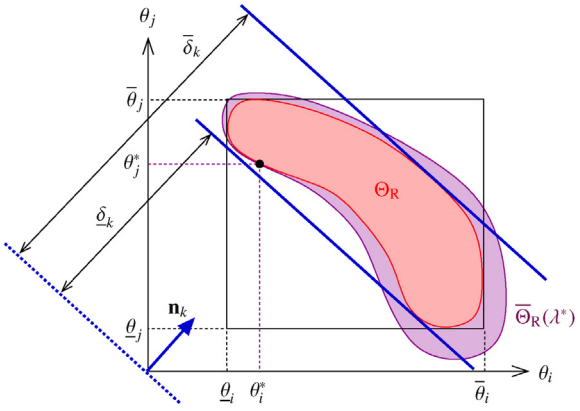

Fig. 5. Illustration of enclosure strategies for an SMR region Θ_R [red-shaded area], either in the form of a likely-contour enclosure $\overline{\Theta}_R(\lambda^*)$ [purple-shaded area], or by the box enclosure (32) [thin solid black lines] along with pairs of non- axis aligned cuts in the form (34) [thick solid blue lines]. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

for some constant $\lambda \geq 0$. Notice that the computational complexity of describing, or closely approximating, the relaxed region $\overline{\Theta}_R(\lambda)$ is then comparable to describing either a likelihood-based confidence region (7) or a GPE region (12), for instance by applying a set-inversion algorithm in the original n_θ- dimensional domain Θ_0.

The following theorem provides a systematic means of computing a value λ^* such that $\overline{\Theta}_R(\lambda^*)$ is a tight enclosure of Θ_R, upon specializing $\varphi(\boldsymbol{\theta}) := \log \mathcal{L}(\boldsymbol{\theta} \mid \mathbf{x}^m)$. This situation is depicted in Fig. 5.

Theorem 3. *Given any continuous function $\varphi : \mathbb{R}^{n_\theta} \to \mathbb{R}$, a valid enclosure $\{\boldsymbol{\theta} \in \Theta_0 \mid \varphi(\boldsymbol{\theta}) \geq \lambda\} \supseteq \Theta_R$ is obtained with $\lambda \geq \lambda^*$ and*

$$\begin{aligned} \lambda^* := & \min_{\boldsymbol{\theta} \in \Theta_0, \mathbf{e} \in \mathbf{E}} \ \varphi(\boldsymbol{\theta}) \\ & \text{s.t.} \quad \boldsymbol{\theta} \in \arg\max_{\boldsymbol{\varpi} \in \Theta_0} \log \mathcal{L}(\boldsymbol{\varpi} \mid \mathbf{x}^m + \mathbf{e}) \\ := & \min_{\boldsymbol{\theta} \in \Theta_0, \mathbf{e} \in \mathbf{E}} \ \varphi(\boldsymbol{\theta}) \\ & \text{s.t.} \, \forall \boldsymbol{\varpi} \in \Theta_0, \log \mathcal{L}(\boldsymbol{\varpi} \mid \mathbf{x}^m + \mathbf{e}) \leq \log \mathcal{L}(\boldsymbol{\theta} \mid \mathbf{x}^m + \mathbf{e}). \end{aligned} \tag{26}$$

Moreover, the enclosure with λ^* is tight in the sense that the two sets share one or more boundary points.

Proof. Let $\overline{\boldsymbol{\theta}} \in \Theta_R$. From (16), there exists $\overline{\mathbf{e}} \in \mathbf{E}$ such that

$$\forall \boldsymbol{\varpi} \in \Theta_0, \quad \log \mathcal{L}(\boldsymbol{\varpi} \mid \mathbf{x}^m + \overline{\mathbf{e}}) \leq \log \mathcal{L}(\overline{\boldsymbol{\theta}} \mid \mathbf{x}^m + \overline{\mathbf{e}}).$$

Therefore, $(\overline{\boldsymbol{\theta}}, \overline{\mathbf{e}})$ satisfies the semi-infinite constraint in (26), and $\varphi(\overline{\boldsymbol{\theta}}) \geq \lambda^*$ follows immediately by optimality. Conversely, any optimal pair $(\boldsymbol{\theta}^*, \mathbf{e}^*)$ corresponding to the optimal value λ^* of (26) is such that $\boldsymbol{\theta}^* \in \operatorname{argmax}_{\boldsymbol{\varpi} \in \Theta_0} \log \mathcal{L}(\boldsymbol{\varpi} \mid \mathbf{x}^m + \mathbf{e}^*)$, and so $\boldsymbol{\theta}^* \in \Theta_R$. Since $\boldsymbol{\theta}^* \in \Theta_0$ and $\varphi(\boldsymbol{\theta}^*) = \lambda^*$, we have that $\boldsymbol{\theta}^*$ is also a boundary point of $\overline{\Theta}_R(\lambda^*)$, and hence a boundary point of Θ_R too. □

Specializing the function φ in Theorem 3 to the log-likelihood function in (25) gives

$$\begin{aligned} \lambda^* = & \min_{\boldsymbol{\theta} \in \Theta_0, \mathbf{e} \in \mathbf{E}} \ \log \mathcal{L}(\boldsymbol{\theta} \mid \mathbf{x}^m) \\ & \text{s.t.} \quad \forall \boldsymbol{\varpi} \in \Theta_0, \quad \log \mathcal{L}(\boldsymbol{\varpi} \mid \mathbf{x}^m + \mathbf{e}) \leq \log \mathcal{L}(\boldsymbol{\theta} \mid \mathbf{x}^m + \mathbf{e}). \end{aligned} \tag{27}$$

Solving this SIP problem is hard in general, since both the semi-infinite constraint and the objective function are generally nonconvex for a nonlinear regression problem. Existing solution approaches to SIP rely on either one of two key ideas [62,63]. In *local reduction methods*, a semi-infinite constraint is represented locally by a finite number of instances of the constraint, upon invoking the implicit function theorem. Alternatively, *discretization (and exchange) methods* involve replacing the uncertain parameter set with a finite discretization so as to create a relaxation of the SIP, and then iteratively refining this discretization until convergence. The focus in the remainder of this paper is on the second type of methods, for which global optimality certificates can be provided upon solving the nonlinear programming (NLP) subproblems to global optimality using complete search methods [64–66].

More specifically, we apply the cutting-plane SIP algorithm by Blankenship and Falk [67] in order to construct a sequence of decreasing upper bounds λ^k on the upper bound λ^* given by (27); that is, we construct an inclusion sequence $\overline{\Theta}_R(\lambda^k) \supseteq \overline{\Theta}_R(\lambda^*) \supseteq \Theta_R$. Within the SMR framework, this algorithm entails an iteration between:

(i) the finite-dimensional nonlinear programming (NLP) subproblems

$$\begin{aligned} (\boldsymbol{\theta}^k, \mathbf{e}^k) \in & \arg\min_{\boldsymbol{\theta} \in \Theta_0, \mathbf{e} \in \mathbf{E}} \ \log \mathcal{L}(\boldsymbol{\theta} \mid \mathbf{x}^m) \\ & \text{s.t.} \quad \forall \boldsymbol{\varpi} \in \Theta_0^k, \quad \log \mathcal{L}(\boldsymbol{\varpi} \mid \mathbf{x}^m + \mathbf{e}) \leq \log \mathcal{L}(\boldsymbol{\theta} \mid \mathbf{x}^m + \mathbf{e}), \end{aligned} \tag{28}$$

where $\Theta_0^k := \{\boldsymbol{\varpi}^0, \ldots, \boldsymbol{\varpi}^k\}$ is a finite subset of Θ_0; and

(ii) the feasibility subproblems

$$\boldsymbol{\varpi}^{k+1} \in \arg\max_{\boldsymbol{\varpi} \in \Theta_0} \ \log \mathcal{L}(\boldsymbol{\varpi} \mid \mathbf{x}^m + \mathbf{e}^k). \tag{29}$$

The subset Θ_0^k at iteration $k = 1$ may be initialized as the empty set, or better, as a singleton set with the maximum-likelihood estimate $\hat{\boldsymbol{\theta}}$ (see Section 2).

Under the assumptions that the likelihood function $\mathcal{L}$ is jointly continuous in $(\boldsymbol{\theta}, \mathbf{e})$ and that the parameter set Θ_0 and the error set $\mathbf{E}$ are both compact, any point of accumulation $\boldsymbol{\theta}^*$ of the sequence $\{\boldsymbol{\theta}^k\}$ will correspond to the best possible lower bound λ^* in (27) [67,Theorem 2.1]. In practice, the iterations may be interrupted when the following termination criterion is satisfied for a certain tolerance $\epsilon > 0$,

$$\log \mathcal{L}(\boldsymbol{\varpi}^{k+1} \mid \mathbf{x}^m + \mathbf{e}^k) \leq \log \mathcal{L}(\boldsymbol{\theta}^k \mid \mathbf{x}^m + \mathbf{e}^k) + \epsilon. \tag{30}$$

Naturally, such a convergence property of the cutting-plane algorithm hinges on the ability to solve all of the nonconvex subproblems (28) and (29) to global optimality. Otherwise, the resulting threshold values λ^* could be underestimated, leading to likelihood contours that exclude parts of the corresponding SMR regions. The practical applicability of this approach may thus be hindered by its computational complexity.

One way to expedite convergence of the cutting-plane algorithm is via the addition of redundant constraints, namely constraints that do not alter the optimal solution set of the SIP (27) yet tighten the relaxations in (28); see, e.g., [68,69] for more details about KKT-based tightening in SIP. Provided that the likelihood function is sufficiently smooth, one can add the first- and second-order optimality cuts (24) as redundant constraints in the subproblem (28), so that[2]

$$\begin{aligned} (\boldsymbol{\theta}^k, \mathbf{e}^k) \in & \arg\min_{\boldsymbol{\theta} \in \Theta_0, \mathbf{e} \in \mathbf{E}} \ \log \mathcal{L}(\boldsymbol{\theta} \mid \mathbf{x}^m) \\ & \text{s.t.} \ \forall \boldsymbol{\varpi} \in \Theta_0^k, \ \log \mathcal{L}(\boldsymbol{\varpi} \mid \mathbf{x}^m + \mathbf{e}) \leq \log \mathcal{L}(\boldsymbol{\theta} \mid \mathbf{x}^m + \mathbf{e}) \\ & \frac{\partial \log \mathcal{L}(\boldsymbol{\theta} \mid \mathbf{x}^m + \mathbf{e})}{\partial \boldsymbol{\theta}} = \mathbf{0}, \quad \frac{\partial^2 \log \mathcal{L}(\boldsymbol{\theta} \mid \mathbf{x}^m + \mathbf{e})}{\partial \boldsymbol{\theta}^2} \preceq \mathbf{0}. \end{aligned} \tag{31}$$

[2] Given that most NLP solvers do not currently support constraints in the form of linear matrix inequalities (LMI), one can always substitute the LMI constraint $\frac{\partial^2 \log \mathcal{L}(\boldsymbol{\theta} \mid \mathbf{x}^m + \mathbf{e})}{\partial \boldsymbol{\theta}^2} \preceq \mathbf{0}$ in (31) by standard inequality constraints on the principal minors of $\frac{\partial^2 \log \mathcal{L}(\boldsymbol{\theta} \mid \mathbf{x}^m + \mathbf{e})}{\partial \boldsymbol{\theta}^2}$ [70].

In the case that none of the regression problems $\max_{\boldsymbol{\varpi}\in\Theta_0} \log \mathcal{L}(\boldsymbol{\varpi} \mid \mathbf{x}^{\mathrm{m}} + \mathbf{e})$ have local (suboptimal) solutions for any $\mathbf{e} \in \mathbf{E}$, enforcing the semi-infinite constraint in (27) is of course equivalent to satisfying the optimality conditions (24), and so the cutting-plane algorithm will trivially terminate after a single iteration. Otherwise, the intermediate solution points $\boldsymbol{\theta}^k$ to the NLP subproblems (31) might correspond to local optima of the regression problems for $\mathbf{e}^k \in \mathbf{E}$, and the algorithm thus keeps iterating by adding cutting planes until all of these local optima have been excluded. At this point, satisfying both the discretized semi-infinite and optimality constraints in (31) becomes equivalent to enforcing the original semi-infinite constraint in (27), and the algorithm will then terminate exactly – optimality gap $\epsilon = 0$ in (30) – at the next iteration. This behavior will be illustrated for the case study problem in Section 5.1.

4.2. Polyhedral enclosure

Applying a set-inversion approach to describe (an enclosure of) the SMR region Θ_{R} can prove computationally expensive, if at all tractable, especially for the estimation problems encountered in real-life situations. A computationally less demanding task entails the computation of a simple (axis-aligned) box enclosure for an SMR region; for instance, by solving a pair of optimization problems for each parameter θ_i, $i = 1 \ldots n_\theta$, as

$$\underline{\theta}_i/\overline{\theta}_i := \min_{\boldsymbol{\theta}\in\Theta_0, \mathbf{e}\in\mathbf{E}} / \max_{\boldsymbol{\theta}\in\Theta_0, \mathbf{e}\in\mathbf{E}} \ \theta_i \tag{32}$$
$$\text{s.t.} \quad \forall \boldsymbol{\varpi} \in \Theta_0, \quad \log \mathcal{L}(\boldsymbol{\varpi} \mid \mathbf{x}^{\mathrm{m}} + \mathbf{e}) \le \log \mathcal{L}(\boldsymbol{\theta} \mid \mathbf{x}^{\mathrm{m}} + \mathbf{e}).$$

Clearly, these bounds may be computed by applying a similar cutting-plane algorithm as in Section 4.1 above, whereby the discretization subproblem (28) is now replaced with

$$(\boldsymbol{\theta}^k, \mathbf{e}^k) \in \arg\min_{\boldsymbol{\theta}\in\Theta_0, \mathbf{e}\in\mathbf{E}} / \operatorname{argmax}_{\boldsymbol{\theta}\in\Theta_0, \mathbf{e}\in\mathbf{E}} \ \theta_i$$
$$\text{s.t.} \quad \forall \boldsymbol{\varpi} \in \Theta_0^k, \quad \log \mathcal{L}(\boldsymbol{\varpi} \mid \mathbf{x}^{\mathrm{m}} + \mathbf{e}) \le \log \mathcal{L}(\boldsymbol{\theta} \mid \mathbf{x}^{\mathrm{m}} + \mathbf{e}),$$

and possibly supplemented with the redundant optimality cuts (24) as in (31).

As an alternative to the direct solution of the SIP problems in (32), one can also use the likelihood-contour enclosure $\overline{\Theta}_{\mathrm{R}}(\lambda^*)$ in (25) with the lower bound λ^* from (27) in order to construct an NLP relaxation of the SIP problem. A conservative box enclosure can be computed in this way by solving the auxiliary (potentially nonconvex) NLP problems

$$\underline{\theta}_i/\overline{\theta}_i := \min_{\boldsymbol{\theta}\in\Theta_0} / \max_{\boldsymbol{\theta}\in\Theta_0} \ \theta_i \tag{33}$$
$$\text{s.t.} \quad \log \mathcal{L}(\boldsymbol{\theta} \mid \mathbf{x}^{\mathrm{m}}) \ge \lambda^*, \quad i = 1 \ldots n_\theta.$$

Of course, the presence of several disconnected subsets in an SMR region cannot be detected by a simple box enclosure, and information about correlations between the parameters θ_i in the actual SMR region is also lost. Part of this information could nonetheless be recovered by constructing a polyhedral enclosure of the SMR region, e.g., expressed in the form

$$\left\{ \boldsymbol{\theta} \in [\underline{\boldsymbol{\theta}}, \overline{\boldsymbol{\theta}}] \;\middle|\; \underline{\delta}_k \le \mathbf{n}_k^T \boldsymbol{\theta} \le \overline{\delta}_k, \quad k = 1 \ldots m \right\}, \tag{34}$$

for a set of vectors $\mathbf{n}_1 \ldots \mathbf{n}_m \in \mathbb{R}^{n_\theta}$ and scalars $\underline{\delta}_1 \ldots \underline{\delta}_m, \overline{\delta}_1 \ldots \overline{\delta}_m \in \mathbb{R}$.

Specializing the function $\varphi(\boldsymbol{\theta}) := \mathbf{n}_k^T \boldsymbol{\theta}$ in Theorem 3 provides a means of constructing such non-axis-aligned polyhedral cuts. Herein, the directions $\mathbf{n}_k$ are chosen in such a way that the cuts correspond to a (face or interior) diagonal of the box enclosure $[\underline{\boldsymbol{\theta}}, \overline{\boldsymbol{\theta}}]$,

$$\mathbf{n}_k := \sum_{i=1}^{n_\theta} \frac{\nu_i}{|\boldsymbol{\nu}|} \frac{1}{\overline{\theta}_i - \underline{\theta}_i} \tag{35}$$

with $\boldsymbol{\nu} \in \{-1, 0, 1\}^{n_\theta}$ and $|\boldsymbol{\nu}| = \sum_{i=1}^{n_\theta} \nu_i \ge 2$. Further, the limits $\underline{\delta}_k$, $\overline{\delta}_k$ in (34) such that the polyhedral cuts are tight can be computed via the solution of the auxiliary SIP problems

$$\underline{\delta}_k/\overline{\delta}_k := \min_{\boldsymbol{\theta}\in[\underline{\boldsymbol{\theta}}, \overline{\boldsymbol{\theta}}], \mathbf{e}\in\mathbf{E}} / \max_{\boldsymbol{\theta}\in[\underline{\boldsymbol{\theta}}, \overline{\boldsymbol{\theta}}], \mathbf{e}\in\mathbf{E}} \ \mathbf{n}_k^T \boldsymbol{\theta} \tag{36}$$
$$\text{s.t.} \quad \forall \boldsymbol{\varpi} \in \Theta_0, \quad \log \mathcal{L}(\boldsymbol{\varpi} \mid \mathbf{x}^{\mathrm{m}} + \mathbf{e}) \le \log \mathcal{L}(\boldsymbol{\theta} \mid \mathbf{x}^{\mathrm{m}} + \mathbf{e}),$$

possibly supplemented with the redundant optimality cuts (24) once again. Similar to the box enclosure (33) earlier, conservative, yet computationally less demanding, polyhedral cuts could be derived from the likelihood-contour enclosure $\overline{\Theta}_{\mathrm{R}}(\lambda^*)$ by solving the auxiliary NLP problems

$$\underline{\delta}_k/\overline{\delta}_k := \min_{\boldsymbol{\theta}\in[\underline{\boldsymbol{\theta}}, \overline{\boldsymbol{\theta}}]} / \max_{\boldsymbol{\theta}\in[\underline{\boldsymbol{\theta}}, \overline{\boldsymbol{\theta}}]} \ \mathbf{n}_k^T \boldsymbol{\theta} \tag{37}$$
$$\text{s.t.} \quad \log \mathcal{L}(\boldsymbol{\theta} \mid \mathbf{x}^{\mathrm{m}}) \ge \lambda^*.$$

Notice that the spans $(\overline{\delta}_k - \underline{\delta}_k)$ are bounded in $[0, 1]$ by construction.

The case of a 2-dimensional face diagonal, where $\nu_i = \nu_j = 1$ are the only nonzero elements in (35), is shown in Fig. 5 for illustration. Enumerating all such pairs of parameters $(\theta_i, \theta_j)_{1\le i<j\le n_\theta}$ calls for the solution of $2n_\theta(n_\theta - 1)$ auxiliary optimization problems. More generally with $|\boldsymbol{\nu}| \ge 2$ nonzero elements in the vector $\boldsymbol{\nu}$, the number of optimization problems is equal to $2^{|\boldsymbol{\nu}|} \binom{n_\theta}{|\boldsymbol{\nu}|}$. To manage this high combinatorial complexity when the number of parameters n_θ is high, it is of course possible to include only those cuts involving combinations of $|\boldsymbol{\nu}| = 2$ or 3 parameters in the polyhedral enclosure at the price of a more conservative polyhedral enclosure.

A simple way of detecting correlations among any parameter pair $(\theta_i, \theta_j)_{1\le i<j\le n_\theta}$ is by calculating the shortest-to-longest ratio between the spans $(\overline{\delta}_k - \underline{\delta}_k)$ obtained with $\nu_i = \nu_j = 1$ on the one hand, and $\nu_i = -\nu_j = 1$ on the other hand. A ratio close to 0 indicates an elongated set projection onto (θ_i, θ_j) in one of the diagonal directions, and therefore a large correlation between θ_i and θ_j; whereas, a ratio close to 1 indicates a more spherical set projection onto (θ_i, θ_j). This approach is the counterpart to the shortest-to-longest axis ratio in an ellipsoidal (Wald) confidence region, which is also the basis for the so-called modified E-optimality criterion in experimental design [11]. More generally, shortest-to-longest-span ratios could be computed with $|\boldsymbol{\nu}| > 2$ in order to unravel correlations among more than 2 parameters likewise. Other classical criteria, such as the A-optimality and D-optimality criteria, also have counterparts in the SMR framework, given by the sum of all the parameter ranges $\overline{\theta}_i - \underline{\theta}_i$ for $i = 1 \ldots n_\theta$ and the volume of the polytope (34), respectively.

To conclude this subsection, it is worth mentioning that the construction of such polyhedral enclosures is also relevant to the approximation of classical inference regions, for instance the likelihood-ratio confidence regions (7).

Illustrative example (continued). Various enclosures of SMR-ℓ_2 regions are presented in Fig. 6 for the BOD case study, here with either 4 or 8 measurement points. The measurement error set $\mathbf{E}$ is constructed based on (21) at the confidence level $1 - \alpha = 0.9$. The threshold values λ^* in the likelihood-contour enclosures $\overline{\Theta}_{\mathrm{R}}(\lambda^*)$ (25) are computed using the cutting-plane SIP algorithm described in Section 4.1, with first-order optimality cuts as in the discretized subproblem (31) . When the subsets Θ_0^k are initialized with the corresponding maximum- likelihood estimate $\hat{\boldsymbol{\theta}}$, the cutting-plane

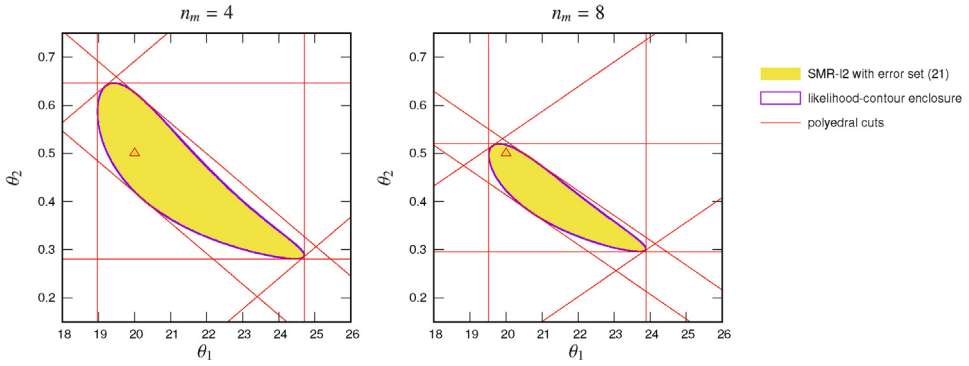

Fig. 6. Comparison of outer-approximation strategies to enclose the SMR-ℓ_2 regions for the BOD example with 4 (left) and 8 (right) measurement points: enclosures based on likelihood- contour cuts (25), and polyhedral cuts (34). The error set $\mathbf{E}$ is constructed based on (21) at the confidence level $1-\alpha = 0.9$. The triangles represent the real parameter values.

Table 1
Comparison between the thresholds λ^* and $\log \mathcal{L}(\hat{\boldsymbol{\theta}} \mid \mathbf{x}^m) - \frac{1}{2}\chi^2_2(0.9)$ corresponding to the SMR-ℓ_2 region (16) and the likelihood-based confidence region (7), respectively, for the BOD example with 4, 8 and 16 measurement points.

n_m	λ^*	$\log \mathcal{L}(\hat{\boldsymbol{\theta}} \mid \mathbf{x}^m) - \frac{1}{2}\chi^2_2(0.9)$
4	−7.66	−7.40
8	−11.90	−11.82
16	−21.20	−21.13

algorithm finds the exact solutions λ^* during the first iteration, irrespective of the number of measurement points. Even for such a simple estimation problem though, the solution of the discretized subproblem (31) to global optimality using GAMS-BARON proves computationally challenging as the number of measurement increases, here taking 404 CPU-sec for 4 measurement points, 3810 CPU-sec for 8 measurement points, and failing to close the gap within 7200 CPU-sec for 16 measurement points.[3] The GAMS code is provided as part of the Supplementary Information (see Appendix C) for the sake of reproducibility.

The likelihood-contour enclosures $\overline{\Theta}_R(\lambda^*)$ are found to provide a very close approximation of the SMR-ℓ_2 regions in Fig. 6—these enclosures are computed using the set- inversion algorithm described in Appendix A. This is expected given the fast convergence between the SMR and likelihood-based confidence regions already observed in Fig. 4, and confirmed by the comparison in Table 1 between the thresholds defining these two confidence regions.

For simplicity, the polyhedral cuts in Fig. 6 are constructed from the likelihood-contour enclosures $\overline{\Theta}_R(\lambda^*)$ rather than the actual SMR- ℓ_2 regions Θ_R here. The numerical solution of the auxiliary NLP subproblems (33) and (37) to global optimality using GAMS-BARON is fast in comparison with the SIP problems, taking <1 CPU-sec.

Finally, the shortest-to-longest-span ratios in the polyhedral enclosures of the SMR-ℓ_2 regions for 4, 8 and 16 measurement points are $\frac{0.284}{0.965} \approx 0.294$, $\frac{0.249}{0.970} \approx 0.257$ and $\frac{0.256}{0.967} \approx 0.265$, respectively. These small ratios (compared to 1) indicate that the SMR regions are 3- to 4-times flatter in one direction compared to the other direction, which unravels the presence of a strong correlation between θ_1 and θ_2 in (13), which is in agreement with the visual impression on Fig. 6.

[3] The reported CPU times are for an AMD Athlon 64 CPU at 2.2 GHz, running Red Hat 4.4.7-18, GAMS 25.0.2, and BARON 17.10.16 with default options, a relative convergence tolerance of 10^{-3} and time limit of 7200 CPU-sec.

5. Case study in temperature-dependent microbial growth

We now apply the SMR framework to a more challenging estimation problem in microbial growth, emphasizing their properties and drawing comparisons with other set-membership and statistical inference methods. Two models describing the effect of culture temperature, T on the growth rate, μ of a microbial population, each one comprising four parameters, are:

(i) The *Ratkowsky model*[71]:

$$\mu(T) = [b\,(T - T_{\min})(1 - \mathrm{e}^{c(T-T_{\max})})]^2,$$

where $T_{\min}$ and $T_{\max}$ (K) represent the minimal and maximal temperatures, respectively; while b ($\mathrm{K}^{-1}\,\mathrm{h}^{0.5}$) and c (K^{-1}) are extra parameters adding flexibility to the shape of the growth model.

The *cardinal temperature model*[72]:

$$\mu(T) = \mu_{\mathrm{opt}}\left[1 - \frac{(T - T_{\mathrm{opt}})^2}{(T - T_{\mathrm{opt}})^2 + T(T_{\max} + T_{\min} - T) - T_{\max}T_{\min}}\right],$$

where $T_{\min}$ and $T_{\max}$ (K) also represent the minimal and maximal temperatures, respectively; T_{opt} (K) corresponds to the optimal growth temperature; and μ_{opt} (h^{-1}) is the maximal growth rate attained at T_{opt}.

Experimental data used in the regression are from [71] for the bacterium *E. coli*. This data set comprises 15 measurement pairs (T_k, μ_k) within the temperature range 294–320 (K), and it is reproduced in Appendix B for completeness. The standard deviation of the growth rate measurements is taken as $\sigma_\mu = 0.1\,(\mathrm{h}^{-1})$ throughout. Results of a maximum-likelihood estimation with constant-variance and Gaussian-distributed errors – or, equivalently, a standard least-squares regression – are presented in Fig. 7. Both model predictions are found to be in good agreement with the experimental data, yet with a higher likelihood for the cardinal temperature model. Note also that errors are only taken into account for the growth rate measurements (outputs) herein, i.e. the temperature measurements (inputs) are considered to be exact.

5.1. Computational procedure and performance

For both candidate models we use the cutting-plane SIP algorithm of Section 4.1 to compute the threshold values λ^* (27), and we describe tight likelihood-contour enclosures $\overline{\Theta}_R(\lambda^*)$ (25) of the SMR regions using a set-inversion algorithm (see Appendix A) in turn. We apply a similar cutting-plane SIP algorithm to determine the box and polyhedral enclosures based on (32) and (36) with $|\boldsymbol{\nu}|$

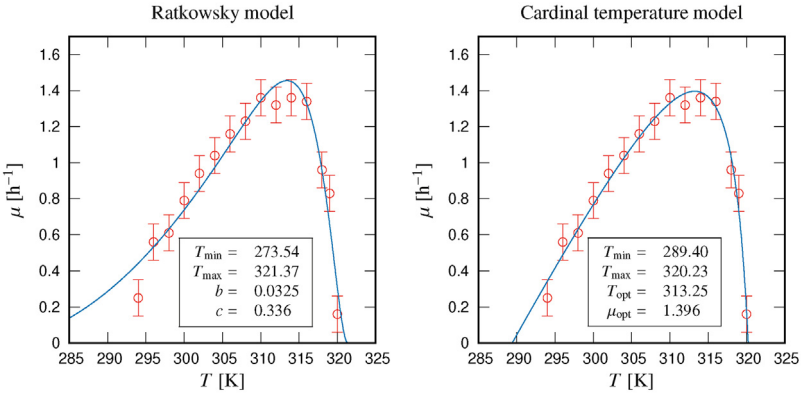

Fig. 7. Maximum- likelihood estimation results for the Ratkowsky (left) and cardinal temperature (right) models. The boxes on each plot report the maximum-likelihood parameter estimates.

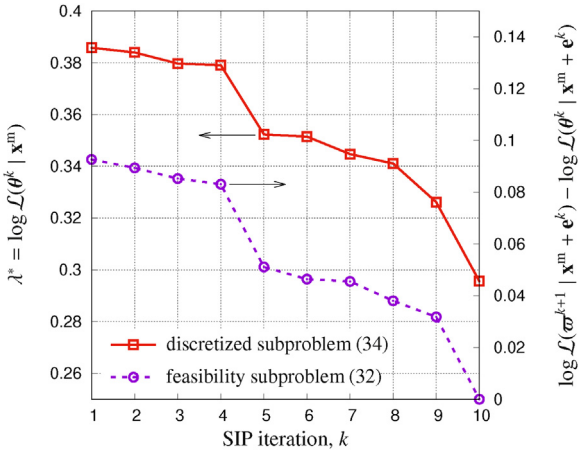

Fig. 8. Iterations of the cutting-plane SIP algorithm for computing the solution value λ^* of (27) for the Ratkowsky model. Left y- axis: solution value λ^* of subproblem (31) at iteration k. Right y-axis: optimal feasibility gap of subproblem (29) at iteration k.

= 2, as described in Section 4.2. First-order optimality cuts are added in the discretized NLP subproblems for all of the SIP problems, as in (31), in order to expedite the convergence of the cutting-plane algorithm, and the sets Θ_0^k are initialized with the maximum-likelihood estimates $\hat{\boldsymbol{\theta}}$ at iteration $k = 1$. All of the NLP subproblems in the SIP algorithm are solved with the global solver GAMS- BARON—these GAMS codes are provided as part of the Supplementary Information (see Appendix C) for reproducibility.[3] Lastly, the set-inversion computations are carried using our in-house library CRONOS [44], which is available from https://github.com/omega- icl/cronos.

In the case of the cardinal temperature model, a single iteration is needed to solve all of the SIP problems exactly – optimality gap $\epsilon = 0$ in (30). This behavior indicates that none of the regression problems for this model exhibit local, suboptimal solutions for the measurement error sets of interest in Sections 5.2 and 5.3 below. In the case of the Ratkowsky, the SIP problems are also solved exactly, but the cutting-plane algorithm terminates after several iterations due to the presence of local optima; for instance, computing the solution value λ^* of (27) for the SMR problem in Section 5.2 takes 10 iterations to terminate, as shown in Fig. 8.

Even though the cutting-plane SIP algorithms terminate exactly (after a single or several iterations), certifying global optimality for most discretized NLP subproblems is currently intractable with the state-of-the-art global solvers BARON [73] and ANTIGONE [66]. As already discussed in Section 4.1, this lack of guarantees could result in the likelihood contours or polyhedral cuts excluding parts of the actual SMR regions. The odds of missing a global optimum in a discretized NLP subproblem is nonetheless mitigated by letting BARON or ANTIGONE run up to a time limit of 7200 CPU-sec here.

5.2. SMR with jointly Gaussian-distributed errors

We consider SMR-ℓ_2 regions, where the error set $\mathbf{E}$ corresponds to the HPD region of a joint Gaussian distribution, as in (21). In order to draw on the asymptotic equivalence with a 95% confidence region in classical frequentist inference (Theorem 2), we select a 15% HPD region for the joint Gaussian distribution of the measurement errors here (see Remark 3). The likelihood-contour and polyhedral enclosures of these SMR regions are compared in Figs. 9 and 10 for the Ratkowsky and cardinal temperature models, respectively. The results from a random sampling are also shown on these plots, which lie inside the actual SMR regions.

Since the polyhedral cuts are tight by construction (Theorem 3), the seemingly large discrepancy between these cuts and the sampled SMR regions in Figs. 9 and 10 is mainly attributable to the sampling not being sufficiently exhaustive. Moreover, the comparisons between the polyhedral and likelihood-contour enclosures on these figures show that the conservatism introduced by the second remains small for both models in the present case of jointly Gaussian-distributed measurement errors.

Reported above each plot in Figs. 9 and 10 are the shortest-to-longest-span ratios in the polyhedral enclosure for the various parameter pairs (see Section 4.2). With the Ratkowsky model, all of these ratios happen to be smaller than 0.4, and even lower than 0.25 for the parameter pair ($T_{\min}$, b), thereby suggesting strong correlations in the parameter set ($T_{\min}$, $T_{\max}$, b, c). With the cardinal temperature model by contrast, most of the ratios are close to or above 0.5, suggesting much weaker correlations amongst the parameters ($T_{\min}$, $T_{\max}$, T_{opt}, μ_{opt}) thereof. Moreover, the SMR intervals for the parameters $T_{\min}$ and $T_{\max}$ – which participate and share the same interpretation in both models – are much larger for the Ratkowsky model than they are for the cardinal temperature model. On the basis of these results, a modeler would normally retain the cardinal temperature model over the Ratkowsky model.

Although the Wald confidence ellipsoids in Figs. 9 and 10 differ significantly from the SMR region enclosures, similar conclusions can nonetheless be drawn with regards to parameter precision and correlation for both the Ratkowsky and cardinal temperature models based on the main axes of the projected Wald ellipsoids. One can also compare the threshold of a likelihood-contour enclosure with its likelihood-based confidence region counterpart (7): for the Ratkowsky model, we find $\lambda^* \approx 5.9$ and $\log \mathcal{L}(\hat{\boldsymbol{\theta}} \mid \mathbf{x}^{\text{m}}) - \frac{1}{2}\chi^2_{n_\theta}(0.95) \approx 9.5$; whereas for the cardinal temperature model, we have $\lambda^* \approx 12.7$ and $\log \mathcal{L}(\hat{\boldsymbol{\theta}} \mid \mathbf{x}^{\text{m}}) - \frac{1}{2}\chi^2_{n_\theta}(0.95) \approx 14.0$. These values

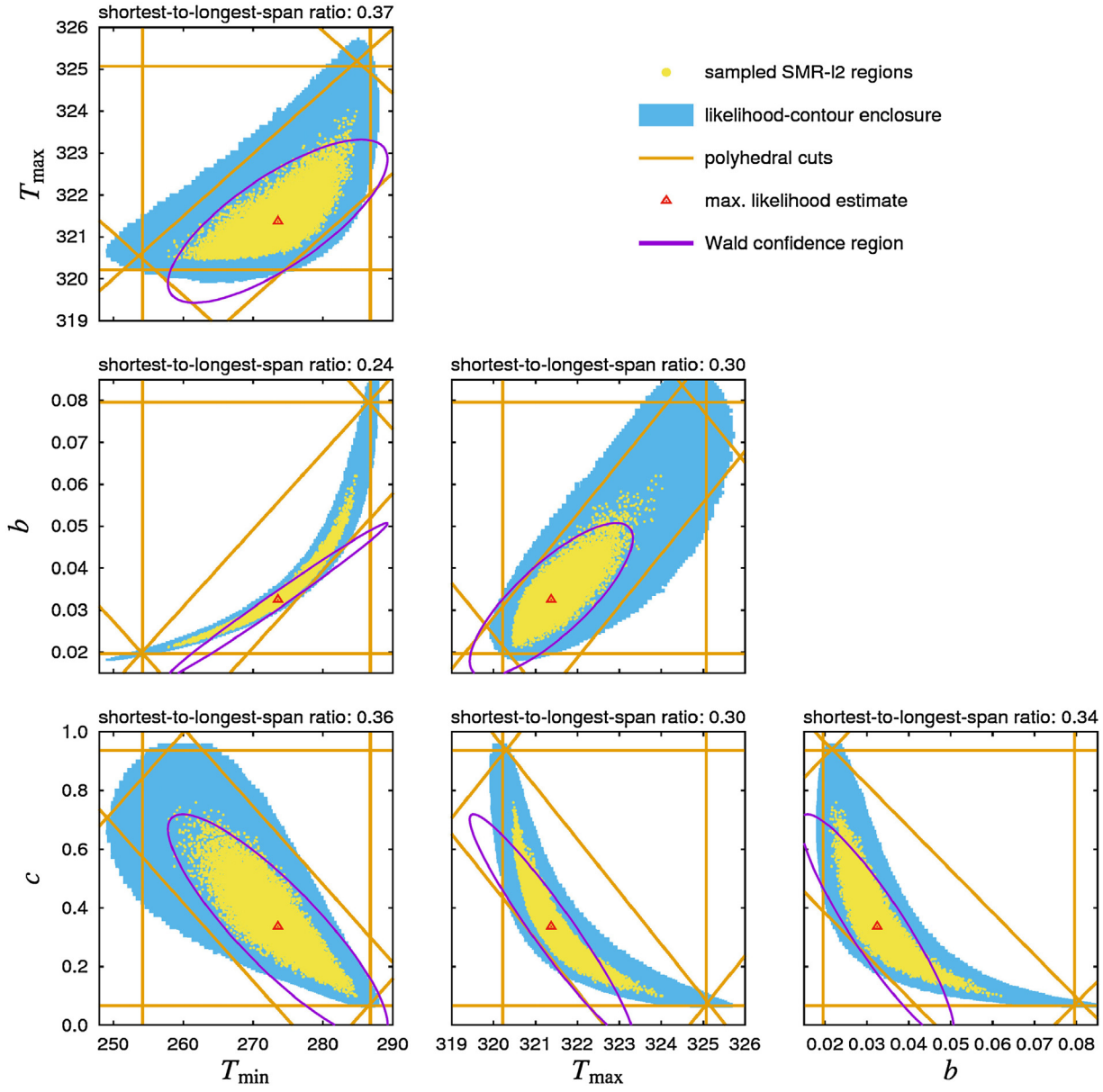

Fig. 9. Matrix plot comparing various parameter regions for the Ratkowsky model: sampled SMR regions, likelihood-contour and polyhedral enclosures of SMR regions, and Wald confidence region. The SMR region is based on ℓ_2-regression, with the error set being the 15% HPD region for the joint Gaussian distribution of the measurement errors.

being quite close to each other for both models provides yet another illustration of the asymptotic equivalence between classical statistical inference approaches and SMR for such choices of the error set (viz. Section 3.2). Also notice the higher likelihood threshold of the cardinal temperature model compared with the Ratkowsky model, which provides yet another indication of a much more confident estimation.

5.3. SMR with independently-distributed errors

We consider alternative SMR-ℓ_2 regions, where the error set $\mathbf{E}$ now comprises independent, 1-sigma error bounds on the measurements,

$$\mathbf{E} := \left\{ \mathbf{e}_\mu \in \mathbb{R}^{15} \;\middle|\; \forall k = 1 \ldots 15, \quad \left| e_{\mu,k} \right| \leq \sigma_\mu \right\}. \tag{38}$$

Similar to the jointly Gaussian-distributed case in Section 5.2 above, we compare various approximations of such an SMR region for the cardinal temperature model in Fig. 11; namely, the tight likelihood-contour and polyhedral enclosures, and an inner-approximation using a random sampling. Despite the error set $\mathbf{E}$ in (38) now being significantly different from a Gaussian HPD region, the polyhedral enclosures turn out to be comparable in shape and size to those in Fig. 10; and the shortest-to-longest-span ratios for the various parameter pairs are similar too. The likelihood-contour enclosure in Fig. 11 describes a rather close approximation of the SMR region too, albeit proving to be more conservative than for the jointly Gaussian- distributed case in Fig. 10. A similar behavior is obtained for the Ratkowsky model (results not shown).

In addition to SMR region approximations, Fig. 11 displays the guaranteed parameter region as given by (12), for the same error set (38). One can check that the inclusion property established in Theorem 1 holds. The guaranteed parameter region turns out to be much smaller than the SMR region here due to both the model mismatch and underestimating the measurement noise. For the Ratkowsky model, the guaranteed parameter region even happens to be empty for these data and error sets. Therefore, unlike SMR regions, guaranteed parameter regions do not provide a reliable means of detecting parameter correlations in the present case.

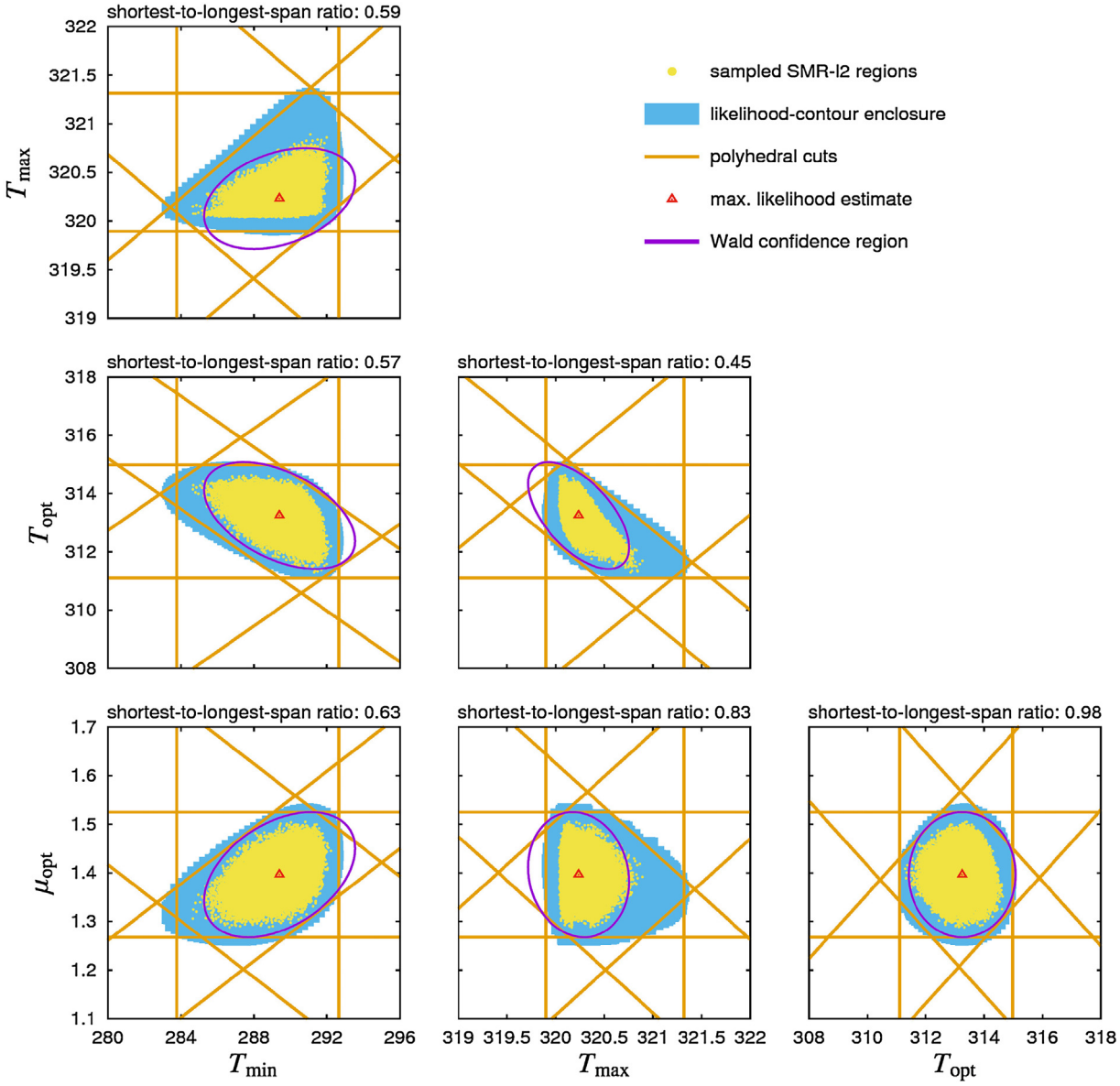

Fig. 10. Matrix plot comparing various parameter regions for the cardinal temperature model: sampled SMR regions, likelihood-contour and polyhedral enclosures of SMR regions, and Wald confidence region. The SMR region is based on ℓ_2-regression, with the error set being the 15% HPD region for the joint Gaussian distribution of the measurement errors.

6. Conclusions and future research directions

This paper has introduced set-membership regression (SMR), a new approach to parameter estimation which seeks to determine the subregion in parameter space enclosing all (global) solutions to a nonlinear regression problem subject to uncertain observations. An SMR region is thus understood as comprising those parameter values that are equally credible under the selected observation error set, in the sense of that regression problem. In particular, this interpretation is not conditional upon the model's structure being correct. Another distinctive feature of SMR is its ability to consider likelihood functions and error sets other than those corresponding to jointly Gaussian-distributed errors, including least-absolute-error (ℓ_1) regression, and independent error distributions or simple error bounds when the underlying statistics is unknown.

In a bounded-error context, SMR provides a means of robustifying existing guaranteed parameter estimation methods. By drawing on the principles of maximum likelihood estimation, an SMR region encloses the corresponding guaranteed parameter set, and unlike the latter, it may not become empty in the presence of large model mismatch or measurement errors and outliers. From a statistical inference viewpoint, SMR has been shown to be asymptotically equivalent to the Wald confidence regions for specific choices of the measurement error set. It will be important to keep developing the underlying SMR theory as part of future work, so as to better grasp the links with both frequentist and Bayesian statistical inference analysis.

Another important contribution of this paper is a computational framework for describing tight enclosures of the SMR regions, in the form of likelihood-contour and polyhedral enclosures. These enclosures can be described via the solution of auxiliary optimization problems, which are typically nonconvex and embed semi-infinite constraints. While tractable in principle using global optimization techniques based on complete search, our experience with such optimization problems is that they challenge state-of-the-art global optimization solvers such as BARON or ANTIGONE, even for small-scale estimation problems as exemplified with the

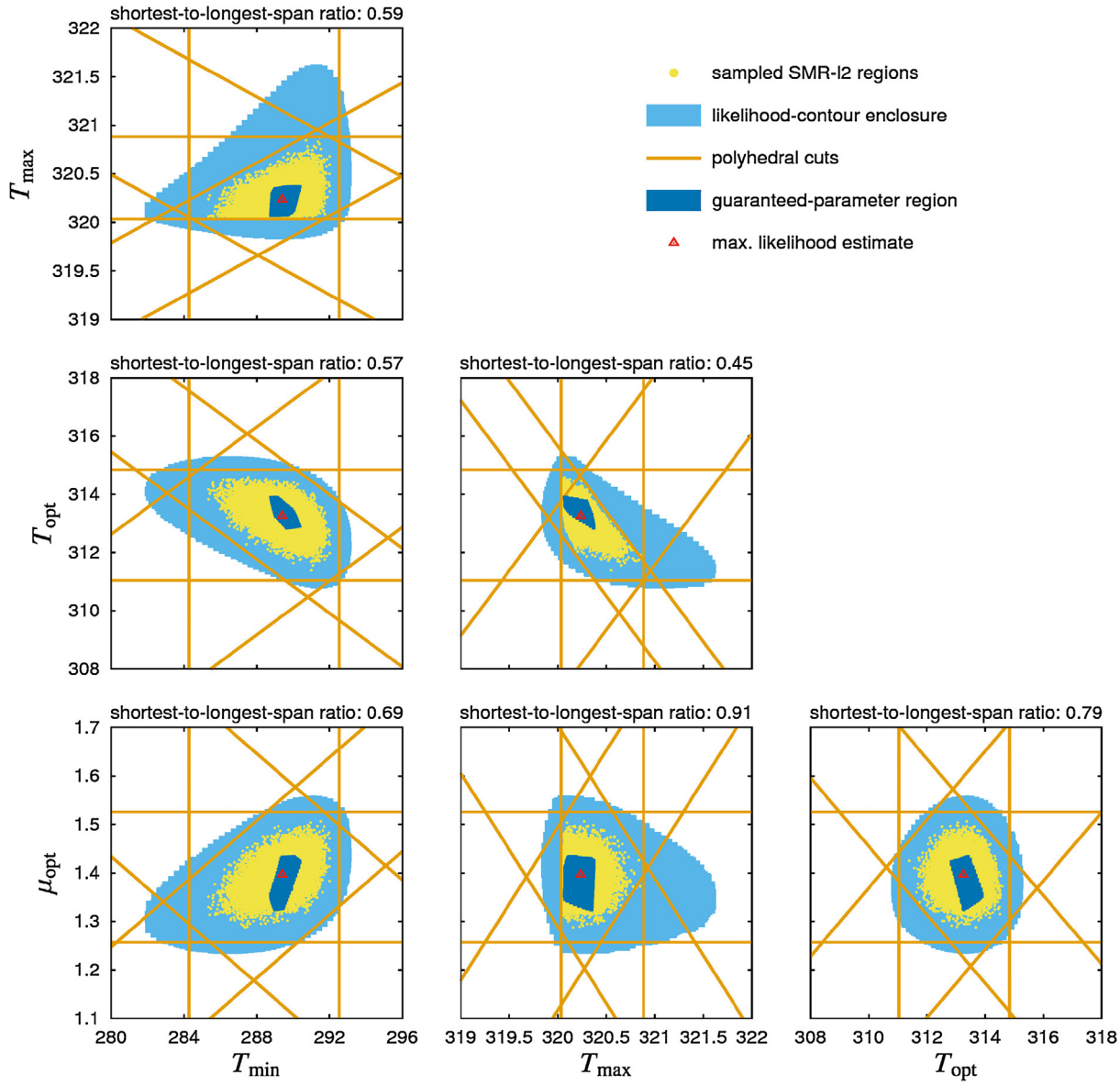

Fig. 11. Matrix plot comparing various parameter regions for the cardinal temperature model: sampled SMR regions, likelihood-contour and polyhedral enclosures of SMR regions, and Wald confidence region. The SMR region is based on ℓ_2-regression, with the error set corresponding to independent 1-sigma error bounds on the measured values.

BOD and microbial growth case studies. The tackling of larger-scale problems, including error-in-variables formulations, is a clear call for improved global search techniques; e.g., by exploiting problem structures or creating redundancy to strengthen the relaxations, or by combining with effective heuristics to increase the likelihood of finding a solution early on during the search [65].

One straightforward extension of the SMR methodology includes parameter estimation problems with other sources of uncertainty than just measurement errors. In principle, any set of nuisance parameters could be accounted for in the regression framework based on a description of the corresponding uncertainty set, similar to the measurement error set **E** in (16).

Lastly, it is worth reiterating that the SMR framework can be extended to parameter estimation in dynamic systems too. The main bottleneck in doing so is of computational rather than conceptual nature, since limited work has been published to date on SIP with differential equations embedded [74]. For instance, applying the cutting-plane SIP algorithm of Section 4 to the dynamic case should rely on efficient complete-search methods for global optimization and constraint satisfaction in dynamic optimization problems [44,75,76].

Data statement

No new data was collected in the course of this research.

Acknowledgements

NDP is grateful to EPSRC and the Department of Chemical Engineering at Imperial College London for Doctoral Training Award (DTA). RP gratefully acknowledges the contribution of the Slovak Research and Development Agency under the project APVV 15-0007. The authors would like to thank the anonymous reviewers for their thoughtful comments that led to substantial improvement of the article.

Appendix A. Set-inversion techniques

The problem in set inversion is describing, or approximating as closely as possible, a set Θ given in implicit form as

$$\Theta := \left\{ \boldsymbol{\theta} \in \Theta_0 \middle| \boldsymbol{\varphi}(\boldsymbol{\theta}) \in \Gamma \right\},$$

where $\Theta_0 \subset \mathbb{R}^{n_\theta}$ is the domain set; $\Gamma \subset \mathbb{R}^{n_\varphi}$, the target set; and $\boldsymbol{\varphi} : \mathbb{R}^{n_\theta} \to \mathbb{R}^{n_\varphi}$ is a continuous function. In other words, Θ is the pre-image of Γ under $\boldsymbol{\varphi}$ in Θ_0. The description of log-likelihood confidence regions (7) and guaranteed parameter regions (12), as well as SMR region enclosures (25), can all be cast as set- inversion problems. Existing computational approaches to set inversion range from sampling-based methods, including stochastic search [39] , support vector machines (SVM) [40] and MCMC [41], to rigorous complete-search methods based on interval analysis and other set arithmetics [42–44] or semidefinite relaxation techniques [45,46].

The focus herein is on branch-and-prune, a complete- search approach entailing the construction of partitions $\mathbb{P}_{\text{in}}$ and $\mathbb{P}_{\text{bnd}}$ such that

$$\bigcup_{P \in \mathbb{P}_{\text{in}}} P \subseteq \Theta \subseteq \bigcup_{P \in \mathbb{P}_{\text{in}} \cup \mathbb{P}_{\text{bnd}}} P,$$

with $\mathbb{P}_{\text{bnd}}$ sufficiently small (in the sense of a certain metric). A prototypical algorithm is the following [42,61]:

Algorithm 1. Basic branch-and-prune algorithm for set inversion.

Input: Termination tolerances $\epsilon_{\text{box}} \geq 0$
Initialization: Set partitions $\mathbb{P}_{\text{bnd}} = \{\Theta_0\}$ and $\mathbb{P}_{\text{in}} = \emptyset$; Set iteration counter $k = 0$
Main Loop:
1. Select a parameter box P in the partition $\mathbb{P}_{\text{bnd}}$ and remove it from $\mathbb{P}_{\text{bnd}}$
2. Compute an enclosure $\overline{\boldsymbol{\varphi}}(P) \supseteq \{\boldsymbol{\varphi}(\boldsymbol{\theta}) \,|\, \boldsymbol{\theta} \in P\}$
3. Exclusion Tests:
   (a) **If** $\overline{\boldsymbol{\varphi}}(P) \subset \Gamma$, insert P into $\mathbb{P}_{\text{in}}$
   (b) **Else if** $\overline{\boldsymbol{\varphi}}(P) \cap \Gamma = \emptyset$, fathom P
   (c) **Else** bisect P and insert subsets back into $\mathbb{P}_{\text{bnd}}$
4. **If** $\text{width}(P) \leq \epsilon_{\text{box}}$ for all $P \in \mathbb{P}_{\text{bnd}}$, **stop**
5. Increment counter k+=1; **Return** to step 1
Output: Partitions $\mathbb{P}_{\text{in}}$ and $\mathbb{P}_{\text{bnd}}$; Iteration count k

The basic requirements for finite convergence of Algorithm 1 are that: (i) the branching procedure is exhaustive; and (ii) the bounding is rigorous and convergent. Step 2 requires an enclosure of the reachable set of $\boldsymbol{\varphi}$ for the current parameter box P, which is the most critical step in Algorithm 1. Various bounding approaches are detailed in [44] (and references therein), with a focus on so-called *factorable* functions; namely, functions that can be represented by a finite number of binary sums, binary products and outer compositions with a univariate function. When the computed bounds are parameter-dependent, as is the case with McCormick relaxations or polynomial models, domain reduction techniques can be used within Step 2 in order to expedite convergence, e.g., via the solution of auxiliary optimization problems. In the case of polynomial models, these bounding subproblems may be nonconvex and it is therefore necessary to construct convex/polyhedral relaxations, for instance in the form of linear programs (LPs). This approach is the same as domain reduction in the context of branch- and-bound search for global optimization [64,65].

Appendix B. Data for the numerical case studies

The measurement data for the BOD example introduced at the end of Section 2 are reported in Table B.1 . Those for the microbial growth problem in Section 5 are reported in Table B.2 .

Table B.1
BOD concentrations at various time instants.

Times, t (day)	0.5	1.0	1.5	2.0	2.5	3.0	3.5	4.0
BOD, c (mg L^{-1})	3.696	6.197	11.289	12.626	13.589	14.702	17.105	17.425
Time, t (day)	4.5	5.0	5.5	6.0	6.5	7.0	7.5	8.0
BOD, c (mg L^{-1})	17.810	18.212	18.625	18.329	19.138	19.522	20.768	21.562

Table B.2
Specific growth rates of *E. coli* at various temperatures.

Temperature, T (K)	294	296	298	300	302	304	306	308
Spec. growth rate, μ (h^{-1})	0.25	0.56	0.61	0.79	0.94	1.04	1.16	1.23
Temperature, T (K)	310	312	314	316	318	319	320	
Spec. growth rate, μ (h^{-1})	1.36	1.32	1.36	1.34	0.96	0.83	0.16	

Appendix C. Supplementary data

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.jprocont.2018.04.002.

References

[1] Y. Bard, Nonlinear Parameter Estimation, Academic Press, New York, 1974.
[2] R.A. Gallant, Nonlinear Statistical Models, John Wiley & Sons, New York, 1987.
[3] D. Bates, D. Watts, Nonlinear Regression Analysis and Its Applications, John Wiley & Sons, 1988.
[4] G.A.F. Seber, C.J. Wild, Nonlinear Regression, John Wiley & Sons, New York, 1989.
[5] W.E. Deming, Statistical Adjustment of Data, Wiley, New York, 1943.
[6] I.-W. Kim, M.J. Liebman, T.F. Edgar, Robust error-in-variables estimation using nonlinear programming techniques, AIChE J. 36 (1990) 985–993.
[7] W.R. Esposito, C.A. Floudas, Global optimization in parameter estimation of nonlinear algebraic models via the error-in-variables approach, Ind. Eng. Chem. Res. 37 (1998) 1841–1858.
[8] C.-Y. Gau, M.A. Stadtherr, Deterministic global optimization for error-in-variables parameter estimation, AIChE J. 48 (2002) 1192–1197.
[9] E. Walter (Ed.), Identifiability of Parametric Models, Pergamon Press, Oxford, 1987.
[10] S.P. Asprey, S. Macchietto, Statistical tools for optimal dynamic model building, Comput. Chem. Eng. 24 (2000) 1261–1267.
[11] K.A. McLean, K.B. McAuley, Mathematical modelling of chemical processes – obtaining the best model predictions and parameter estimates using identifiability and estimability procedures, Can. J. Chem. Eng. 90 (2) (2012) 351–366.
[12] D. Bonvin, C. Georgakis, C.C. Pantelides, M. Barolo, M.A. Grover, D. Rodrigues, R. Schneider, D. Dochain, Linking models and experiments, Ind. Eng. Chem. Res. 24 (2016) 6891–6903.
[13] W. Rooney, L.T. Biegler, Incorporating joint confidence regions into design under uncertainty, Comput. Chem. Eng. 23 (1999) 1563–1575.
[14] W. Rooney, L. Biegler, Design for model parameter uncertainty using nonlinear confidence regions, AIChE J. 47 (2001) 1794–1804.
[15] W. Langson, S. Raković, I. Chryssochoos, D.Q. Mayne, Robust model predictive control using tubes, Automatica 40 (1) (2004) 125–133.
[16] V. Sakizlis, N. Kakalis, V. Dua, J.D. Perkins, E.N. Pistikopoulos, Design of robust model-based controllers via parametric programming, Automatica 40 (2) (2004) 189–201.
[17] M.E. Villanueva, R. Quirynen, M. Diehl, B. Chachuat, B. Houska, Robust MPC via min–max differential inequalities, Automatica 77 (2017) 311–321.
[18] J.-L. Gouzé, A. Rapaport, Z. Hadj-Sadok, Interval observers for uncertain biological systems, Ecol. Model. 133 (2000) 45–56.
[19] B. Chachuat, O. Bernard, Probabilistic observers for a class of uncertain biological processes, Int. J. Robust Nonlinear Control 16 (3) (2006) 157–171.
[20] C.R. Rojas, J.S. Welsh, G.C. Goodwin, A. Feuer, Robust optimal experiment design for system identification, Automatica 43 (6) (2007) 993–1008.
[21] S.W. Marvel, C.M. Williams, Set membership experimental design for biological systems, BMC Syst. Biol. 6 (1) (2012) 21.
[22] A.R. Gottu-Mukkula, R. Paulen, Model-based design of optimal experiments for nonlinear systems in the context of guaranteed parameter estimation, Comput. Chem. Eng. 99 (2017) 198–213.
[23] R.D. Cook, S. Weisberg, Confidence curves in nonlinear regression, J. Am. Stat. Assoc. 85 (1990) 544–551.
[24] W.Q. Meeker, L.A. Escobar, Teaching about approximate confidence regions based on maximum likelihood estimation, Am. Stat. 49 (1) (1995) 48–53.
[25] M.J. Bayarri, J.O. Berger, The interplay of Bayesian and frequentist analysis, Stat. Sci. 19 (1) (2004) 58–80.

[26] A. Gelman, J. Carlin, H. Stern, D. Rubin, Bayesian Data Analysis, 2nd ed., Chapman & Hall/CRC, 2004.
[27] A. Smith, G. Roberts, Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods, J. R. Stat. Soc. 55 (1993) 3–23.
[28] W. Gilks, S. Richardson, D. Spiegelhalter, Markov Chain Monte Carlo in Practice, 1st ed., Chapman & Hall/CRC, 1996.
[29] E. Laloy, B. Rogiers, J. Vrugt, D. Mallants, D. Jacques, Efficient posterior exploration of a high-dimensional groundwater model from two-stage Markov chain Monte Carlo simulation and polynomial chaos expansion, Water Resour. Res. 49 (2013) 2664–2682.
[30] C.P. Robert, The Bayesian Choice: From Decision-Theoretic Foundations to Computational Implementation, 2nd ed., Springer, 2001.
[31] J. Berger, The case for objective Bayesian analysis, Bayesian Anal. 1 (3) (2006) 385–402.
[32] K. Fedra, G. Van Straten, M.B. Beck, Uncertainty and arbitrariness in ecosystems modelling: a lake modelling example, Ecol. Model. 13 (1–2) (1981) 87–110.
[33] Special Issue on Parameter Identification with Error Bounds, Mathematics & Computers in Simulation 32 (1990) 447-607.
[34] M. Milanese, J.P. Norton, H. Piet- lahanier, E. Walter, Bounding Approaches to System Identication, Plenum Press, New York, 1996.
[35] J. Anderson, A. Papachristodoulou, On validation and invalidation of biological models, BMC Bioinform. 10 (2009) 132.
[36] P. Rumschinski, S. Borchers, S. Bosio, R. Weismantel, R. Findeisen, Set-based dynamical parameter estimation and model invalidation for biochemical reaction networks, BMC Syst. Biol. 4 (2010) 69.
[37] L. Jaulin, M. Kieffer, O. Didrit, E. Walter, Applied Interval Analysis, Springer-Verlag, London, 2001.
[38] S. Streif, M. Karl, R. Findeisen, Outlier analysis in set-based estimation for nonlinear systems using convex relaxations, Proceedings of the 2013 European Control Conference (2013) 2921–2926.
[39] T. Goerke, E. Engell, Application of evolutionary algorithms in guaranteed parameter estimation, 2016 IEEE Congress on Evolutionary Computation (CEC) (2016) 500–505.
[40] K.J. Keesman, R. Stappers, Nonlinear set-membership estimation: a support vector machine approach, J. Inverse Ill-Posed Probl. 12 (1) (2004) 27–41.
[41] E.W. Bai, H. Ishii, R. Tempo, A Markov chain Monte Carlo approach to nonlinear parameter system identification, IEEE Trans. Autom. Control 60 (9) (2015) 2542–2546.
[42] L. Jaulin, E. Walter, Set inversion via interval analysis for nonlinear bounded-error estimation, Automatica 29 (1993) 1053–1064.
[43] L. Jaulin, E. Walter, Guaranteed nonlinear parameter estimation from bounded-error data via interval analysis, Math. Comput. Simul. 35 (1993) 123–137.
[44] B. Chachuat, B. Houska, R. Paulen, N.D. Perić, J. Rajyaguru, M.E. Villanueva, Set-theoretic approaches in analysis, estimation and control of nonlinear systems, IFAC-PapersOnLine 48 (8) (2015) 981–995.
[45] V. Cerone, D. Piga, D. Regruto, Set-membership error-in-variables identification through convex relaxation techniques, IEEE Trans. Autom. Control 57 (2) (2012) 517–522.
[46] V. Magron, D. Henrion, J.B. Lasserre, Semidefinite approximations of projections and polynomial images of semialgebraic sets, SIAM J. Optim. 25 (4) (2015) 2143–2164.
[47] M. Milanese, Properties of least-squares estimates in set membership identification, Automatica 31 (2) (1995) 327–332.
[48] B.T. Poljak, J.Z. Tsypkin, Robust identification, Automatica 16 (1980) 53–63.
[49] A. van den Bos, Nonlinear least-absolute-values and minimax model fitting, Automatica 24 (6) (1988) 803–808.
[50] S.W. Marvel, C.M. Williams, Computational experience with confidence regions and confidence intervals for nonlinear least squares, Technometrics 29 (1) (1987) 67–82.
[51] D. Cox, D. Hinkley, Theoretical Statistics, 1st ed., Chapman and Hall, 1974.
[52] R.F. Engle, Wald, likelihood ratio, and Lagrange multiplier tests in econometrics, in: Z. Griliches, M.D. Intriligator (Eds.), Handbook of Econometrics, vol. 2, North Holland, 1984, pp. 775–826, Chapter 13.
[53] L. Jaulin, E. Walter, Guaranteed nonlinear parameter estimation via interval computations, Interval Comput. 3 (1993) 61–75.
[54] L. Jaulin, Computing minimal-volume credible sets using interval analysis; application to Bayesian estimation, IEEE Trans. Signal Process. 54 (2006) 3632–3636.
[55] B.L. Welch, H.W. Peers, On formulae for confidence points based on integrals of weighted likelihoods, J. R. Stat. Soc. B 25 (1991) 318–329.
[56] T.A. Severini, On the relationship between Bayesian and non-Bayesian interval estimates, J. R. Stat. Soc. B 53 (3) (1991) 611–618.
[57] L. Ventura, W. Racugno, A note on the relationships between Bayesian and non-Bayesian predictive inference, Atti della XLV Riunione Scientifica della SIS, Padova, 16–18 June 2010 (2010) 1–8.
[58] I. Smith, A. Ferrari, Equivalence between the posterior distribution of the likelihood ratio and a p-value in an invariant frame, Bayesian Anal. 9 (4) (2014) 939–962.
[59] L. Jaulin, Probabilistic set-membership approach for robust regression, J. Stat. Theory Pract. 4 (2010) 155–167.
[60] P.J. Rousseeuw, A.M. Leroy, Robust Regression and Outlier Detection, John Wiley & Sons, 1987.
[61] R.E. Moore, Parameter sets for bounded-error data, Math. Comput. Simul. 34 (2) (1992) 113–119.
[62] R. Hettich, K.O. Kortanek, Semi-infinite programming: theory, methods and applications, SIAM Rev. 35 (3) (1993) 380–429.
[63] M. Lopez, G. Still, Semi-infinite programming, Eur. J. Oper. Res. 180 (2007) 491–518.
[64] M. Tawarmalani, N. Sahinidis, Convexification and Global Optimization in Continuous and Mixed-Integer Nonlinear Programming: Theory, Algorithms, Software, and Applications, Kluwer Academic Publishers, 2002.
[65] A. Neumaier, Complete search in continuous global optimization and constraint satisfaction, Acta Numer. 13 (2004) 271–369.
[66] R. Misener, C.A. Floudas, ANTIGONE: algorithms for continuous/integer global optimization of nonlinear equations, J. Glob. Optim. 59 (2014) 503–526.
[67] J.W. Blankenship, J.E. Falk, Infinitely constrained optimization problems, J. Optim. Theory Appl. 19 (2) (1976) 261–281.
[68] C.A. Floudas, O. Stein, The adaptive convexification algorithm: a feasible point method for semi-infinite programming, SIAM J. Optim. 18 (4) (2007) 1187–1208.
[69] A. Mitsos, P. Lemonidis, C.K. Lee, P.I. Barton, Relaxation-based bounds for semi-infinite programs, SIAM J. Optim. 19 (1) (2008) 77–113.
[70] S. Boyd, L. El Ghaoui, E. Feron, V. Balakrishnan, Linear Matrix Inequalities in System and Control Theory Studies in Applied Mathematics, vol. 15, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, 1994.
[71] D. Ratkowsky, R. Lowry, T. McMeekin, A. Stokes, R. Chandler, A model for bacterial culture growth rate throughout the entire biokinetic temperature range, J. Bacteriol. 154 (1983) 1222–1226.
[72] J. Lobry, L. Rosso, J. Flandrois, A fortran subroutine for the determination of parameter confidence limits in non-linear models, Binary 3 (1991) 86–93.
[73] M. Tawarmalani, N.V. Sahinidis, A polyhedral branch-and-cut approach to global optimization, Math. Progr. 103 (2005) 225–249.
[74] A. Mitsos, B. Chachuat, P.I. Barton, Towards global bilevel dynamic optimization, J. Glob. Optim. 45 (1) (2009) 63–93.
[75] B. Chachuat, A. Singer, P. Barton, Global methods for dynamic optimization and mixed-integer dynamic optimization, Ind. Eng. Chem. Res. 45 (2006) 8373–8392.
[76] Y. Lin, M.A. Stadtherr, Deterministic global optimization of nonlinear dynamic systems, AIChE J. 53 (2007) 866–875.