

Reducing the Search Space for Global Minimum: A Focused Regions Identification Method for Least Squares Parameter Estimation in Nonlinear Models

Guanglu Zhang

Mem. ASME
Department of Mechanical Engineering,
Carnegie Mellon University,
Pittsburgh, PA 15213
e-mail: glzhang@cmu.edu

Douglas Allaire

Mem. ASME
J. Mike Walker '66 Department of Mechanical
Engineering,
Texas A&M University,
College Station, TX 77843
e-mail: dallaire@tamu.edu

Jonathan Cagan

Fellow ASME
Department of Mechanical Engineering,
Carnegie Mellon University,
Pittsburgh, PA 15213
e-mail: cagan@cmu.edu

Important for many science and engineering fields, meaningful nonlinear models result from fitting such models to data by estimating the value of each parameter in the model. Since parameters in nonlinear models often characterize a substance or a system (e.g., mass diffusivity), it is critical to find the optimal parameter estimators that minimize or maximize a chosen objective function. In practice, iterative local methods (e.g., Levenberg–Marquardt method) and heuristic methods (e.g., genetic algorithms) are commonly employed for least squares parameter estimation in nonlinear models. However, practitioners are not able to know whether the parameter estimators derived through these methods are the optimal parameter estimators that correspond to the global minimum of the squared error of the fit. In this paper, a focused regions identification method is introduced for least squares parameter estimation in nonlinear models. Using expected fitting accuracy and derivatives of the squared error of the fit, this method rules out the regions in parameter space where the optimal parameter estimators cannot exist. Practitioners are guaranteed to find the optimal parameter estimators through an exhaustive search in the remaining regions (i.e., focused regions). The focused regions identification method is validated through two case studies in which a model based on Newton's law of cooling and the Michaelis–Menten model are fitted to two experimental data sets, respectively. These case studies show that the focused regions identification method can find the optimal parameter estimators and the corresponding global minimum effectively and efficiently. [DOI: 10.1115/1.4054440]

Keywords: parameter estimation, least squares, curve fitting, global optimum, local optimum, outlier detection, nonlinear regression, inverse problem, Newton's law of cooling, Michaelis–Menten model, Lineweaver–Burk plot, big data and analytics, computational foundations for engineering optimization, data-driven engineering, inverse methods for engineering applications, machine learning for engineering applications

1 Introduction

Fitting nonlinear models to data, also called nonlinear curve fitting, is an important inverse problem for many science and engineering fields [1–3]. A major goal of nonlinear curve fitting is to estimate the value of each parameter in a specified nonlinear model. For example, Fick's second law is fitted to data collected from standard moisture diffusion tests to estimate the mass diffusivity of various composite materials [4,5]. Michaelis–Menten model is fitted to data measured in enzymatic reactions to estimate the maximum reaction rate and the Michaelis constant of these reactions [6,7].

In nonlinear curve fitting problems, parameters in the nonlinear model usually characterize a substance or a system and have clear interpretations, such as mass diffusivity between two species, viscosity of a fluid, and maximum velocity of a chemical reaction [8]. It is therefore critical to find the optimal parameter estimators that minimize or maximize a chosen objective function, such as the squared error of the fit. In practice, iterative local methods (e.g., Levenberg–Marquardt method [9,10] and Nelder–Mead method [11]) and heuristic methods (e.g., genetic algorithms [12] and simulated annealing [13]) are commonly employed for

parameter estimation in nonlinear models to minimize the squared error of the fit. However, practitioners are not able to know whether the parameter estimators derived through these extant methods are the optimal parameter estimators that correspond to the global minimum of the squared error of the fit. The iteration process of these methods also does not shed light on where the global minimum may be located. Recent research introduces a solution interval method for least squares parameter estimation in nonlinear models [14]. Practitioners do not need to guess the initial value of each parameter in a nonlinear model to initialize the method, but the solution interval method is still not guaranteed to find the optimal parameter estimators that correspond to the global minimum of the squared error of the fit.

In this paper, a generally applicable method is introduced to reduce the search space for the optimal least squares parameter estimators in nonlinear curve fitting problems. The method applies to nonlinear models that have a closed-form expression. Using expected fitting accuracy and derivatives of the squared error of the fit, this method rules out the suboptimal regions in parameter space where the optimal parameter estimators cannot exist. Notably, the ruled-out regions may include one or more local minima in which some iterative local methods can be trapped. The remaining regions in the parameter space are defined as the focused regions. If adequate computational power is available, practitioners are guaranteed to find the optimal parameter estimators that correspond to the *global minimum* of the squared error of the fit through an exhaustive search in the

Contributed by the Computers and Information Division of ASME for publication in the JOURNAL OF COMPUTING AND INFORMATION SCIENCE IN ENGINEERING. Manuscript received October 3, 2021; final manuscript received April 14, 2022; published online June 3, 2022. Assoc. Editor: Kazuhiro Saitou.

focused regions. In addition, a new approach to detect potential outliers using inequalities derived by expected fitting accuracy and derivatives of the squared error of the fit is also proposed for nonlinear curve fitting problems.

This paper begins with a brief review of extant least squares methods that are commonly employed to solve nonlinear curve fitting problems in Sec. 2. Section 3 presents the new method to identify the focused regions and search for the optimal least squares parameter estimators in nonlinear models. Four steps to implement the focused regions identification method in practice are provided in Sec. 4. The applications of the method are demonstrated through case studies of the Rumford cooling experiment and the puromycin experiment in Sec. 5. Section 6 presents an approach to detect potential outliers in data sets of nonlinear curve fitting problems using inequalities derived in Sec. 3. The paper concludes with a discussion of the contribution of the work presented here and future research directions.

2 Background and Related Work

Nonlinear models have been applied in many science and engineering fields. A review paper written by Archontoulis and Miguez provides 72 commonly employed nonlinear models that have a closed-form expression, such as exponential decay function, logistic function, Gompertz function, Gaussian function, and Michaelis–Menten model [3]. A major goal of fitting a specified nonlinear model to data, also called nonlinear curve fitting, is to estimate the value of each parameter in the nonlinear model. These parameter estimators should minimize or maximize a chosen objective function, such as squared error of the fit and maximum likelihood [15]. In nonlinear curve fitting problems, the nonlinear model usually has less than ten parameters. Each parameter in the nonlinear model has a clear interpretation (e.g., mass diffusivity and viscosity). Sometimes, the values of one or more parameters in the nonlinear model are bounded. For example, the viscosity of a fluid, μ , needs to be a positive constant ($\mu > 0$). However, inequality or equality constraints that establish relationships between several parameters in the nonlinear model are usually not applied to nonlinear curve fitting problems. In this section, several methods that are frequently used to solve nonlinear curve fitting problems are briefly reviewed. These methods estimate parameter values in nonlinear models to minimize the *squared error of the fit*. Detailed reviews of extant nonlinear parameter estimation methods that cover various objective functions can be found in books written by Bates and Watts [16], Björck [17], Gelman et al. [18], and Tarantola [8], among others.

Many nonlinear curve fitting problems are solved through linear transformation in practice. Specifically, the nonlinear model is transformed into a linear model, the value of each parameter in the model is then derived by linear regression. However, the parameter estimators derived through linear transformation are usually not the optimal estimators that minimize the squared error of the fit for the original problem. For example, Michaelis–Menten model, also called Michaelis–Menten equation, is a well-known nonlinear model in biochemistry [6]. The nonlinear model has the mathematical form

$$y = \frac{\theta_1 x}{x + \theta_2} \quad (1)$$

where x and y represent substrate concentration and reaction velocity in a biochemical reaction, respectively; parameter θ_1 denotes the maximum reaction velocity, and parameter θ_2 is defined as the Michaelis constant. Parameters θ_1 and θ_2 are, by definition, positive constants. To fit the Michaelis–Menten model to experimental data and estimate the values of parameters θ_1 and θ_2 in the model, Lineweaver and Burk transform the Michaelis–Menten model into a linear model as [19]

$$\frac{1}{y} = \frac{x + \theta_2}{\theta_1 x} = \frac{1}{\theta_1} + \frac{\theta_2}{\theta_1 x} \quad (2)$$

Equation (2) is often called the Lineweaver–Burk equation, and the graphical representation of Eq. (2) ($1/x$ versus $1/y$ graph) is known as the Lineweaver–Burk plot or the double reciprocal plot in biochemistry. The values of $1/\theta_1$ and θ_2/θ_1 in Eq. (2) are derived by regressing $1/y$ against $1/x$ using all data points collected from the biochemical experiment, and then the values of parameters θ_1 and θ_2 can be calculated accordingly. However, parameter estimators derived through the Lineweaver–Burk equation minimize the squared error of the fit for the reciprocal of the reaction velocity, $1/y$, rather than for the reaction velocity, y . In other words, the experimental data points with low reaction velocity (a low value of y), which often have greater percentage error, have a higher impact on the linear regression results compared to the data points with high reaction velocity (a high value of y). The parameter estimators derived through the Lineweaver–Burk equation are therefore not the optimal least squares estimators for the original problem of fitting the Michaelis–Menten model represented by Eq. (1) to experimental data. An example using the Lineweaver–Burk equation to estimate parameter values in the Michaelis–Menten model is provided in Sec. 5.2.

Since optimal least squares parameter estimators usually cannot be derived through linear transformation and many nonlinear models even cannot be transformed into a linear model, iterative local methods, such as the Gauss–Newton method [20], the Levenberg–Marquardt method [9,10], the Nelder–Mead method [11], and the trust region reflective method [21,22], are commonly employed to estimate parameter values in nonlinear models to minimize the squared error of the fit. Using these methods, the estimator for each parameter begins with an initial guess, and then the estimator is updated in an iterative manner. The iterative equation used in these methods is

$$\theta^{(s+1)} = \theta^{(s)} + \Delta^{(s)} \quad (3)$$

where $\theta = [\theta_1, \theta_2, \dots, \theta_n]$ is a column vector of n parameters in a specified nonlinear model $y = f(x_1, x_2, \dots, x_q, \theta)$ that is fit to m data points $[x_{1(i)}, x_{2(i)}, \dots, x_{q(i)}, y_{(i)}]$, and $i \in \{1, 2, \dots, m\}$; $\theta^{(s+1)}$ and $\theta^{(s)}$ represent parameter values at iteration step $s+1$ and s , respectively; $\Delta^{(s)}$ represents the increment vector at the iteration step s , which is specified differently in these iterative local methods. For example, when the Levenberg–Marquardt method is employed, the increment vector Δ in the iteration step s is computed through [10,23]

$$[\mathbf{J}(\theta^{(s)})^T \mathbf{J}(\theta^{(s)}) + \lambda^{(s)} \mathbf{D}^{(s)}] \Delta^{(s)} = -\mathbf{J}(\theta^{(s)})^T \mathbf{r}(\theta^{(s)}) \quad (4)$$

In Eq. (4), λ is a non-negative damping factor, \mathbf{D} is a diagonal scaling matrix, $\mathbf{r} = [r_1, r_2, \dots, r_m]$ is a column vector of m residuals [14], and \mathbf{J} is a $m \times n$ Jacobian matrix of residuals defined by

$$(\mathbf{J})_{ij} = \frac{\partial r_i}{\partial \theta_j} \quad (5)$$

where $i \in \{1, 2, \dots, m\}$, $j \in \{1, 2, \dots, n\}$.

Besides these iterative local methods, several heuristic methods, also call guided random search techniques [24], such as genetic algorithms [12], simulated annealing [13], particle swarm optimization [25], and tabu search [26,27], also have been applied for parameter estimation in nonlinear models to minimize the squared error of the fit. These heuristic methods are usually initialized by multiple initial guesses, known as initial population, as the initial estimators for each parameter in the nonlinear model. These parameter estimators are then updated in an iterative process involving random variation [24].

Importantly, in nonlinear curve fitting problems, it is critical to find the optimal estimator for each parameter that minimizes the squared error of the fit (i.e., global minimum) since each parameter in the nonlinear model characterizes a substance or a system (e.g., mass diffusivity between two species and maximum velocity of a chemical reaction). However, it is not guaranteed to find the optimal parameter estimators using these iterative local methods

and heuristic methods since these methods are based on sampling over the parameter space. Specifically, using these methods, the iterative process is usually stopped when the result reaches a pre-specified step tolerance, function tolerance, or maximum number of iterations. If the initial guess(es) and the stopping criteria are properly specified, these algorithms may converge to a local minimum where the first-order derivatives of the squared error of the fit equal (or approximate) zero [14,24]. However, since multiple or even an infinite number of local minima may exist for nonlinear curve fitting problems, practitioners are not able to know whether the local minimum derived from an iterative local method or a heuristic method is the global minimum that corresponds to the optimal parameter estimators, and thus it is difficult to determine when to stop the search process using these methods in practice. In other words, since the goal is to find the optimal parameter estimators that minimize the squared error of the fit (i.e., global minimum), when practitioners find a local minimum using these extant methods, they do not know whether they should stop the search process or keep searching the parameter space until they exhaust all available computational power (e.g., through using other initial guesses or increasing the population size). In addition, the iteration process of these extant methods also does not inform practitioners about where the global minimum may be located. Notably, in practice, it is also usually not feasible to perform an exhaustive search in the parameter space of nonlinear curve fitting problems since the parameter space is infinite in many cases. Recent research introduces a solution interval method for least squares parameter estimation in nonlinear models [14]. To initialize the solution interval method, practitioners do not need to guess the initial value of each parameter in a nonlinear model. It is also proved that the method can find the optimal estimator for each parameter that minimizes the squared error of the fit when the nonlinear curve fitting problem satisfies the specified monotonic and symmetric conditions [14]. However, many nonlinear curve fitting problems do not fulfill these conditions, such as the curve fitting problem defined in Sec. 5.2; it is therefore not guaranteed to find the optimal parameter estimators and the corresponding global minimum for nonlinear curve fitting problems using the solution interval method.

3 Focused Regions Identification for Least Squares Parameter Estimation in Nonlinear Models

Based on the research gap discussed in Sec. 2, a generally applicable method is introduced in this section to identify the focused regions for least squares parameter estimation in nonlinear models. Using expected fitting accuracy and derivatives of the squared error of the fit, this method rules out the suboptimal regions, where the optimal parameter estimators that correspond to the global minimum of the squared error of the fit are impossible to exist, from parameter space. The remaining regions in the parameter space are defined as the focused regions to search for the optimal parameter estimators. If adequate computational power is available, practitioners are guaranteed to find the optimal parameter estimators that correspond to the *global minimum* of the squared error of the fit through an exhaustive search in the focused regions. Notably, the method introduced in this section applies to nonlinear models that have a closed-form expression. The method becomes less efficient as the number of parameters in the nonlinear model increases, and ten parameters are found as a practical limit for the application of the method due to current typical computational resources. As stated in Sec. 2, many nonlinear models, such as exponential decay function, logistic function, Gaussian function, and Michaelis–Menten model, have less than ten parameters.

For a generic nonlinear curve fitting problem, there are m data points $[x_{1(i)}, x_{2(i)}, \dots, x_{q(i)}, y_{(i)}]$, where $i \in \{1, 2, \dots, m\}$. The objective is to fit a specified nonlinear model $y = f(\mathbf{x}, \boldsymbol{\theta})$ to the m data points by finding the optimal values for parameters $\boldsymbol{\theta}$ that minimize the squared error of the fit, where $\boldsymbol{\theta} = [\theta_1, \theta_2, \dots, \theta_n]$ represents n constant parameters in the nonlinear model, and $\mathbf{x} = [x_1, x_2, \dots, x_q]$.

Here, $y = f(\mathbf{x}, \boldsymbol{\theta})$ represents the closed-form expression of the nonlinear model using a finite number of standard operations that do not include limit, differentiation, and integration. Fitting a nonlinear model that does not have a closed-form expression (e.g., a differential equation) to data is beyond the scope of this paper, and model selection is also not discussed in this paper. The squared error of the fit for the problem, also called the residual sum of squares (RSS), is

$$S(\boldsymbol{\theta}) = \sum_{i=1}^m [y_{(i)} - f(x_{1(i)}, x_{2(i)}, \dots, x_{q(i)}, \boldsymbol{\theta})]^2 = \sum_{i=1}^m [y_{(i)} - f(\mathbf{x}_{(i)}, \boldsymbol{\theta})]^2 \quad (6)$$

where $\boldsymbol{\theta} = [\theta_1, \theta_2, \dots, \theta_n]$ are the estimated values of parameters $\boldsymbol{\theta}$. Sometimes there are bounds applied to parameter θ_j (e.g., $\theta_j^{(lb)} < \theta_j < \theta_j^{(ub)}$, where $\theta_j^{(lb)}$ and $\theta_j^{(ub)}$ are the lower bound and upper bound of parameter θ_j , respectively), and the n -dimensional parameter space is defined by the allowed value range of each parameter θ_j , where $j \in \{1, 2, \dots, n\}$.

The regions that cannot satisfy expected fitting accuracy are ruled out from parameter space first. When an expected fitting accuracy is specified, an upper limit is applied to the squared error of the fit or the squared error of each data point. The regions in which the squared error exceeds the upper limit are ruled out from parameter space. Notably, the ruled-out regions may include several or even an infinite number of local minima in which some iterative local methods, such as the Gauss–Newton method and the Levenberg–Marquardt method, can be trapped. The ruled-out regions identification through expected fitting accuracy is demonstrated in Fig. 1(a), where $n = 1$.

In curve fitting problems, the fitting accuracy is usually measured by the coefficient of determination, denoted R^2 , or mean squared error, denoted MSE. The coefficient of determination is defined as

$$R^2 = 1 - \frac{\sum_{i=1}^m [y_{(i)} - f(\mathbf{x}_{(i)}, \boldsymbol{\theta})]^2}{\sum_{i=1}^m [y_{(i)} - \bar{y}]^2} \quad (7)$$

where \bar{y} is the mean of the observed data $y_{(i)}$ as

$$\bar{y} = \frac{1}{m} \sum_{i=1}^m y_{(i)} \quad (8)$$

If the coefficient of determination is expected to be greater than or equal to a specified value, denoted $R_{(e)}^2$, the regions defined by Eq. (9) should be ruled out from the space of parameters $\boldsymbol{\theta}$:

$$\left\{ \boldsymbol{\theta} : R^2 = 1 - \frac{\sum_{i=1}^m [y_{(i)} - f(\mathbf{x}_{(i)}, \boldsymbol{\theta})]^2}{\sum_{i=1}^m [y_{(i)} - \bar{y}]^2} < R_{(e)}^2 \right\} \quad (9)$$

Equation (9) can be rewritten as

$$\left\{ \boldsymbol{\theta} : \sum_{i=1}^m [y_{(i)} - f(\mathbf{x}_{(i)}, \boldsymbol{\theta})]^2 > [1 - R_{(e)}^2] \sum_{i=1}^m [y_{(i)} - \bar{y}]^2 \right\} \quad (10)$$

Equation (10) includes the summation of m terms on each side of the inequality. In cases when the ranges of parameters $\boldsymbol{\theta}$ cannot be easily derived from Eq. (10), a more conservative criterion, based on the condition that each of the m terms on the left side of Eq. (10) is larger than its corresponding term on the right side of Eq. (10) can be used. In this case, the regions that should be ruled out from the parameter space can be defined by

$$\left\{ \boldsymbol{\theta} : |y_{(i)} - f(\mathbf{x}_{(i)}, \boldsymbol{\theta})| > \sqrt{1 - R_{(e)}^2} |y_{(i)} - \bar{y}| \text{ for all } i \in \{1, 2, \dots, m\} \right\} \quad (11)$$

Equation (11) represents the intersection of m inequalities related to parameters $\boldsymbol{\theta}$ after m data points are plugged into the equation. The regions defined by Eq. (11) are the subregions of the regions defined by Eq. (10). In other words, smaller suboptimal regions

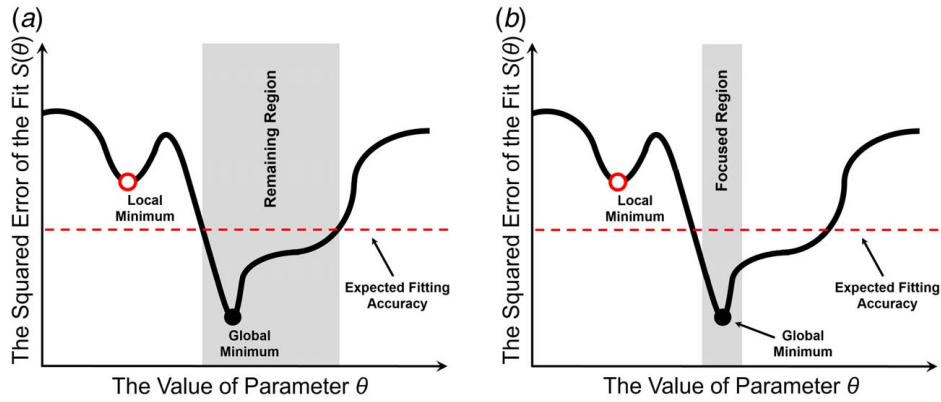


Fig. 1 A representative focused regions identification process: (a) ruled-out regions identification through expected fitting accuracy and (b) focused regions identification using derivatives of the squared error of the fit

are ruled out if Eq. (11) is employed instead of Eq. (10). Importantly, as demonstrated in the two case studies in Sec. 5, when Eq. (10) or Eq. (11) is employed to rule out suboptimal regions from the parameter space, sampling over the parameter space is not necessary given the closed-form nature of Eqs. (10) and (11). It is conservative to assign $R_{(e)}^2 = 0$ in Eqs. (10) and (11) since a successful curve fitting result must have its coefficient of determination, R^2 , greater than or equal to zero. In practice, a larger expected coefficient of determination, $R_{(e)}^2$, could be employed (e.g., 0.5 or 0.7) at the discretion of the practitioner. Cases when the whole parameter space is ruled out by Eq. (10) or Eq. (11) indicate that the curve fitting result cannot reach the expected coefficient of determination, and practitioners need to check data reliability and the fidelity of the nonlinear model employed to fit the data.

Similarly, the mean squared error of the fit is defined as

$$\text{MSE} = \frac{1}{m} \sum_{i=1}^m [y_{(i)} - f(\mathbf{x}_{(i)}, \boldsymbol{\theta})]^2 \quad (12)$$

If the mean squared error is expected to be less than or equal to a specified value, denoted $\text{MSE}_{(e)}$, the regions defined by Eq. (13) should be ruled out from the space of parameters $\boldsymbol{\theta}$:

$$\left\{ \boldsymbol{\theta} : \text{MSE} = \frac{1}{m} \sum_{i=1}^m [y_{(i)} - f(\mathbf{x}_{(i)}, \boldsymbol{\theta})]^2 > \text{MSE}_{(e)} \right\} \quad (13)$$

Equation (13) can be rewritten as

$$\left\{ \boldsymbol{\theta} : \sum_{i=1}^m [y_{(i)} - f(\mathbf{x}_{(i)}, \boldsymbol{\theta})]^2 > m \cdot \text{MSE}_{(e)} \right\} \quad (14)$$

The left side of the inequality represented by Eq. (14) includes the summation of m terms. In cases when the ranges of parameters $\boldsymbol{\theta}$ cannot be easily derived from Eq. (14), a more conservative set of regions that should be ruled out from the parameter space can be defined by

$$\left\{ \boldsymbol{\theta} : |y_{(i)} - f(\mathbf{x}_{(i)}, \boldsymbol{\theta})| > \sqrt{\text{MSE}_{(e)}} \text{ for all } i \in \{1, 2, \dots, m\} \right\} \quad (15)$$

Equation (15) is derived based on the condition that each term on the left side of Eq. (14) is greater than the expected $\text{MSE}_{(e)}$. Equation (15) represents the intersection of m inequalities related to parameters $\boldsymbol{\theta}$ after m data points are plugged into the equation. The regions defined by Eq. (15) are the subregions of the regions defined by Eq. (14). In practice, when the expected fitting accuracy

is specified as that the squared error for each data point is expected to be less than or equal to a specified value, denoted as $\text{MSE}_{(e)}$ in Eqs. (14) and (15), the ruled-out regions can be defined by the union of the m inequalities related to parameters $\boldsymbol{\theta}$ after m data points are plugged into Eq. (15). Cases when the whole parameter space is ruled out by Eq. (14) or Eq. (15) indicate that the curve fitting result cannot satisfy the expected mean squared error or the expected squared error for each data point.

The remaining regions in the parameter space derived through the expected fitting accuracy are then further reduced using derivatives of the squared error of the fit. The regions that are finally left over in the parameter space are defined as the focused regions, as demonstrated in Fig. 1(b). Specifically, if $\boldsymbol{\theta}^* = [\theta_1^*, \theta_2^*, \dots, \theta_n^*]$ are the optimal values of parameters $\boldsymbol{\theta}$, $S(\boldsymbol{\theta}^*)$ is the global minimum of the squared error of the fit. The squared error of the fit $S(\boldsymbol{\theta}^* + \mathbf{h})$ is then written as the Taylor series at $\boldsymbol{\theta}^*$ as

$$S(\boldsymbol{\theta}^* + \mathbf{h}) = S(\boldsymbol{\theta}^*) + \sum_{k=1}^{\infty} \frac{(\mathbf{h} \cdot \nabla)^k S(\boldsymbol{\theta}^*)}{k!} \quad (16)$$

where $\mathbf{h} = [h_1, h_2, \dots, h_n]$ are arbitrary small positive or negative values, and

$$(\mathbf{h} \cdot \nabla)^k = \sum_{|\alpha|=k} \frac{k!}{\alpha!} \mathbf{h}^\alpha \partial^\alpha \quad (17)$$

In Eq. (17), α is a multi-index defined by [28,29]

$$\alpha = (\alpha_1, \alpha_2, \dots, \alpha_j, \dots, \alpha_n) \quad (\alpha_j \in \{0, 1, 2, \dots\}) \quad (18)$$

$$|\alpha| = \alpha_1 + \alpha_2 + \dots + \alpha_n \quad (19)$$

$$\alpha! = \alpha_1! \alpha_2! \dots \alpha_n! \quad (20)$$

$$\mathbf{h}^\alpha = h_1^{\alpha_1} h_2^{\alpha_2} \dots h_n^{\alpha_n} \quad (21)$$

$$\partial^\alpha S(\boldsymbol{\theta}^*) = \frac{\partial^{|\alpha|} S(\boldsymbol{\theta}^*)}{\partial \theta_1^{\alpha_1} \partial \theta_2^{\alpha_2} \dots \partial \theta_n^{\alpha_n}} \quad (22)$$

Since $S(\boldsymbol{\theta}^*)$ is the global minimum of the squared error of the fit, $S(\boldsymbol{\theta}^* + \mathbf{h})$ must be greater than $S(\boldsymbol{\theta}^*)$ for any small positive or negative values of \mathbf{h} , and thus

$$\sum_{|\alpha|=k} \frac{k!}{\alpha!} \mathbf{h}^\alpha \partial^\alpha S(\boldsymbol{\theta}^*) > 0 \quad (23)$$

for any $k \in \{1, 2, \dots\}$ based on Eq. (16), where k represents the order of derivatives for the squared error of the fit. The regions in which Eq. (23) is not satisfied should be ruled out from parameter space. In practice, to further reduce the remaining regions in the parameter space derived through expected fitting accuracy, the order of derivatives for the squared error of the fit in Eq. (23) could be iteratively increased (i.e., assign $k=1$ in Eq. (23) first, and then $k=2, 3, \dots$) until the new inequalities including the k th-order derivatives of $S(\theta)$ derived from Eq. (23) cannot rule out any new area from the remaining regions for parameters θ . The iteration also could be halted when the remaining regions are small enough for an exhaustive search. For example, when $n=1$, Eq. (16) is simplified as

$$S(\theta_1^* + h_1) = S(\theta_1^*) + h_1 \left(\frac{dS}{d\theta_1} \right)_{\theta_1=\theta_1^*} + \frac{h_1^2}{2} \left(\frac{d^2S}{d\theta_1^2} \right)_{\theta_1=\theta_1^*} + \frac{h_1^3}{6} \left(\frac{d^3S}{d\theta_1^3} \right)_{\theta_1=\theta_1^*} + \dots \quad (24)$$

Given h_1 could be any small positive or negative value and each derivative term in Eq. (24) should have a positive value, the regions defined by Eqs. (25)–(28) must be ruled out from parameter space. For instance, as shown in Eq. (25), when $k=1$, since h_1 could have any small positive or negative value, the first-order derivative of the squared error of the fit must be zero at the global optimum θ_1^* , the regions where the first-order derivative is greater or less than zero should be ruled out from the parameter space. In practice, the regions defined by Eqs. (25)–(28) can be ruled out from parameter space one by one until the new inequalities including the k th-order derivative of $S(\theta_1)$ cannot rule out any new area from the remaining regions for parameter θ_1 or the remaining regions derived through the k th-order derivative of $S(\theta_1)$ are small enough for practitioners to perform an exhaustive search using the available computational power

$$\left\{ \theta_1 : \frac{dS}{d\theta_1} > 0 \text{ or } \frac{dS}{d\theta_1} < 0 \right\} \text{ when } k=1 \quad (25)$$

$$\left\{ \theta_1 : \frac{d^2S}{d\theta_1^2} \leq 0 \right\} \text{ when } k=2 \quad (26)$$

.....

$$\left\{ \theta_1 : \frac{d^k S}{d\theta_1^k} > 0 \text{ or } \frac{d^k S}{d\theta_1^k} < 0 \right\} \text{ when } k \text{ is an odd integer} \quad (27)$$

$$\left\{ \theta_1 : \frac{d^k S}{d\theta_1^k} \leq 0 \right\} \text{ when } k \text{ is an even integer} \quad (28)$$

.....

where k represents the order of derivative for the squared error of the fit. Notably, the squared error of the fit is the summation of m terms. In cases when the ranges of parameters θ cannot be easily derived from an inequality related to the derivative of the squared error of the fit, the subregions of the regions defined by the inequality could be ruled out instead. For example, the first inequality in Eq. (25) is expressed as

$$\left\{ \theta_1 : \frac{dS}{d\theta_1} = \sum_{i=1}^m -2 \frac{df(\mathbf{x}_{(i)}, \theta_1)}{d\theta_1} [y_{(i)} - f(\mathbf{x}_{(i)}, \theta_1)] > 0 \right\} \quad (29)$$

If the range of parameter θ_1 cannot be easily derived from Eq. (29), the regions that should be ruled out from the parameter

space can be defined conservatively by

$$\left\{ \theta_1 : \frac{df(\mathbf{x}_{(i)}, \theta_1)}{d\theta_1} [y_{(i)} - f(\mathbf{x}_{(i)}, \theta_1)] < 0 \text{ for all } i \in \{1, 2, \dots, m\} \right\} \quad (30)$$

which is the intersection of the regions defined by m inequalities represented by the equation. Equation (30) is derived based on the condition that each of the m terms on the left side of Eq. (29) has a positive value, and the regions defined by Eq. (30) are therefore the subregions of the regions defined by Eq. (29).

Ultimately, the optimal values for parameters θ that minimize the squared error of the fit could be searched in the focused regions identified through the expected fitting accuracy and the derivatives of the squared error of the fit. As shown in the two case studies in Sec. 5, the search space for the optimal parameter estimators that correspond to the global minimum of the squared error of the fit is significantly reduced using the expected fitting accuracy and the derivatives of the squared error of the fit. The optimal parameter estimators could be found through an exhaustive search, also called grid search, with specified significant digits (e.g., four significant digits) in the focused regions, if adequate computational power is available. If exhaustive search is not feasible, an iterative method that can solve nonlinear least squares problems subject to bounds, such as the trust region reflective method [21,22], could be employed to find the parameter estimators. The initial values of parameters to initialize the chosen iterative method could be specified based on previous experience, expert advice, linear least squares estimators (if linear transformation is feasible for the specified nonlinear model $y=f(\mathbf{x}, \theta)$) [3,30], graphical exploration [31], or the roots of parameter solution equations [14]. Notably, the parameter estimators derived through an iterative least squares method may not be the optimal parameter estimators since one or more local minima may exist in the focused regions.

4 Four Steps to Estimate Parameter Values in Nonlinear Models

In Sec. 3, a generic method is introduced to identify the focused regions in the parameter space for least squares parameter estimation in nonlinear models. The method rules out the suboptimal regions from the parameter space, and the remaining regions in the parameter space are defined as the focused regions to search for the optimal parameter estimators that correspond to the global minimum of the squared error of the fit. In this section, the procedure to implement the focused regions identification method is summarized in four steps. These four steps are illustrated in Fig. 2. Practitioners can follow these steps to identify the focused regions of their nonlinear curve fitting problems and search for the optimal parameter estimators in the focused regions.

Step 1—Data collection and model selection: Research data (m data points) are first collected from scientific experiments or computer simulations. An appropriate mathematical model with a closed-form expression including n parameters is selected to fit the research data. Practitioners need to check data reliability (e.g., potential outliers) and the fidelity of the mathematical model before fitting the model to data. If the selected mathematical model is linear, linear regression is employed to solve the values of these n parameters in the model to minimize the squared error of the fit.

Step 2—The ruled-out regions delineation using the expected fitting accuracy: If the selected mathematical model is nonlinear, the regions that cannot satisfy the expected fitting accuracy are ruled out from the parameter space. The n -dimensional parameter space is defined by the allowed value range of each parameter θ_j in the nonlinear model (e.g., $\theta_j^{(lb)} < \theta_j < \theta_j^{(ub)}$, where $\theta_j^{(lb)}$ and $\theta_j^{(ub)}$ are the lower bound and upper bound of parameter θ_j , respectively, where $j \in \{1, 2, \dots, n\}$). When the expected fitting accuracy is defined by the coefficient of determination (e.g., $R^2 \geq 0$), the

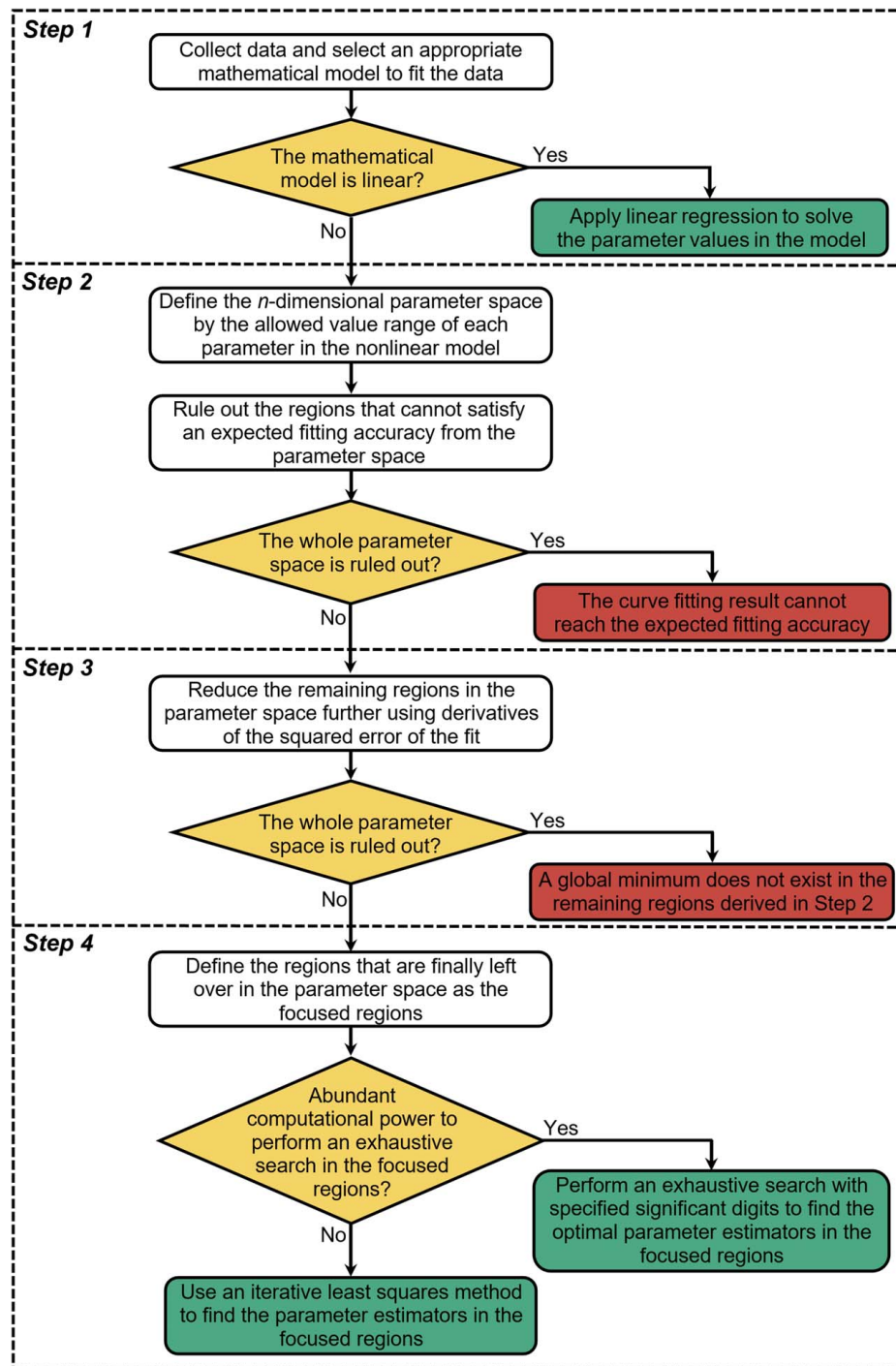


Fig. 2 Four steps to apply the focused regions identification method

regions defined by Eq. (10) are ruled out from the parameter space. In cases when the ranges of parameters θ cannot be easily derived from Eq. (10), the regions defined by Eq. (11) are ruled out instead. When the expected fitting accuracy is defined by the mean squared error (e.g., $MSE \leq 10^{-2}$) or squared error for each data point, Eq. (14) or Eq. (15) are employed in the similar manner. If the whole parameter space is ruled out in Step 2, it indicates that the fitting result cannot reach the expected fitting accuracy, and practitioners need to double check data reliability (e.g., potential outliers) and the fidelity of the nonlinear model employed to fit the data.

Step 3—The ruled-out regions delineation using the derivatives of the squared error of the fit: The remaining regions in the

parameter space are further reduced using the derivatives of the squared error of the fit if the whole parameter space is not ruled out in Step 2. For the first-order derivatives of the squared error of the fit ($k=1$), Eq. (23) leads to $2n$ inequalities. The regions defined by these $2n$ inequalities are ruled out from the parameter space first. In cases when the ranges of parameters θ cannot be easily derived from an inequality, the subregions of the regions defined by the inequality could be ruled out instead, as exemplified by Eq. (30). The order of the derivatives of the squared error of the fit is then iteratively increased ($k=2, 3, 4, \dots$) until the new inequalities derived from Eq. (23) for the k th-order derivatives of the squared error of the fit cannot rule any new area out from the remaining parameter space. Practitioners also can stop the iteration

when the remaining parameter space is small enough for an exhaustive search using the available computational power. If the whole parameter space is ruled out in Step 3, it indicates that a global minimum does not exist in the remaining regions derived in Step 2, and practitioners need to consider whether the expected fitting accuracy is overestimated in Step 2.

Step 4—The optimal parameter estimators search in the focused regions: If the whole parameter space is not ruled out in Step 3, the regions that are finally left over in the parameter space are defined as the focused regions. An exhaustive search with specified significant digits (e.g., four significant digits) is employed to find the optimal parameter estimators that correspond to the global minimum of the squared error of the fit in the focused regions if adequate computational power is available for the task. If the exhaustive search is not feasible, an iterative least squares method, such as the trust region reflective method [21,22], could be employed to find the parameter estimators in the focused regions. However, an iterative least squares method is not guaranteed to find the optimal parameter estimators (global minimum) in the focused regions.

5 Case Studies of the Rumford Cooling Experiment and the Puromycin Experiment

To demonstrate the four-step procedure outlined in Sec. 4, a model based on Newton's law of cooling and the Michaelis–Menten model are fitted to two data sets collected from the Rumford cooling experiment and the puromycin experiment, respectively, as two case studies. The results of these two case studies validate the effectiveness of the focused regions identification method. Notably, the original parameter space is infinite in both case studies. Using the method, the search space for the optimal least squares parameter estimators is significantly reduced from an infinite space to one or several finite focused regions. In addition, it is proved that the optimal parameter estimators corresponding to the global minimum of the squared error of the fit are located in the focused regions identified through the method in both case studies.

5.1 The Rumford Cooling Experiment. The Rumford cooling experiment is an influential experiment in the history of cooling law [32]. In the experiment, a blunt steel borer is shoved against the bottom of the bore of a cylinder by a strong screw. The boring is performed for 30 min, and the temperature of the cylinder is recorded after the boring is stopped. A book written by Roller [33] includes more details of the Rumford cooling experiment, such as Rumford's diagrams of his experimental apparatus. The cylinder cooling data set appears in Table 1. There are 13 data points in the experimental data set [16,33]. A model based on Newton's law of cooling [16], as shown in Eq. (31), is fitted to the experimental data set by finding the optimal least squares

Table 1 Temperature versus time data set for the Rumford cooling experiment in Sec. 5.1

No.	Time x (min)	Temperature y (°F)
1	4	126
2	5	125
3	7	123
4	12	120
5	14	119
6	16	118
7	20	116
8	24	115
9	28	114
10	31	113
11	34	112
12	37.5	111
13	41	110

estimator for parameter θ in the equation:

$$y = 60 + 70e^{-\theta x} \quad (31)$$

where x denotes time, y represents cylinder temperature, and parameter θ is a constant cooling coefficient in the model.

To find the optimal least squares estimator for the cooling coefficient θ in the model, the focused regions identification method presented in Secs. 3 and 4 is employed. The squared error of the fit for this problem, also called the RSS, is

$$S(\theta) = \sum_{i=1}^{13} [y_{(i)} - 60 - 70e^{-\theta x_{(i)}}]^2 \quad (32)$$

where $x_{(i)}$ and $y_{(i)}$ represent the 13 data points in Table 1. The regions for the value of parameter θ that cannot satisfy the expected fitting accuracy are ruled out from the parameter space $\theta \in (-\infty, \infty)$ first. Here, the expected fitting accuracy is specified as that the coefficient of determination, R^2 , is greater than or equal to zero ($R_e^2 = 0$) for demonstration purposes. Equation (11) is reduced to

$$\{\theta : |y_{(i)} - 60 - 70e^{-\theta x_{(i)}}| > |y_{(i)} - \bar{y}| \text{ for all } i \in \{1, 2, \dots, 13\}\} \quad (33)$$

Plugging the 13 data points into Eq. (33), the two regions that should be ruled out from parameter space are given by

$$\{\theta : \theta < -0.01699 \text{ or } \theta > 0.05102\} \quad (34)$$

Notably, the regions shown in Eq. (34) are the intersection of the regions defined by the 13 inequalities represented by Eq. (33). The remaining region $\theta \in [-0.01699, 0.05102]$ is shown as the gray area in Fig. 3(a).

The remaining region is further reduced using the derivatives of the squared error of the fit defined by Eq. (32). The first-order derivative of the squared error of the fit is

$$\frac{dS}{d\theta} = \sum_{i=1}^{13} 140x_{(i)}e^{-\theta x_{(i)}}(y_{(i)} - 60 - 70e^{-\theta x_{(i)}}) \quad (35)$$

Since $x_{(i)} > 0$ and $e^{-\theta x_{(i)}} > 0$, Eq. (35) leads to

$$\{\theta : y_{(i)} - 60 - 70e^{-\theta x_{(i)}} > 0 \text{ for all } i \in \{1, 2, \dots, 13\}\} \quad (36)$$

or

$$\{\theta : y_{(i)} - 60 - 70e^{-\theta x_{(i)}} < 0 \text{ for all } i \in \{1, 2, \dots, 13\}\} \quad (37)$$

The ruled-out regions are derived by plugging the 13 data points into Eqs. (36) and (37) as

$$\{\theta : \theta < 0.008207 \text{ or } \theta > 0.01505\} \quad (38)$$

The remaining region to search for the optimal parameter estimator is therefore reduced to $\theta \in [0.008207, 0.01505]$ using the first-order derivative of the squared error of the fit. The second-order derivative of the squared error of the fit is

$$\frac{d^2S}{d\theta^2} = \sum_{i=1}^{13} -140x_{(i)}^2e^{-\theta x_{(i)}}(y_{(i)} - 60 - 140e^{-\theta x_{(i)}}) \quad (39)$$

Since $x_{(i)}^2 > 0$ and $e^{-\theta x_{(i)}} > 0$, Eq. (39) leads to

$$\{\theta : y_{(i)} - 60 - 140e^{-\theta x_{(i)}} \geq 0 \text{ for all } i \in \{1, 2, \dots, 13\}\} \quad (40)$$

Plugging the 13 data points into Eq. (40), the ruled-out region derived by the second-order derivative of the squared error of the fit is

$$\{\theta : \theta \geq 0.1880\} \quad (41)$$

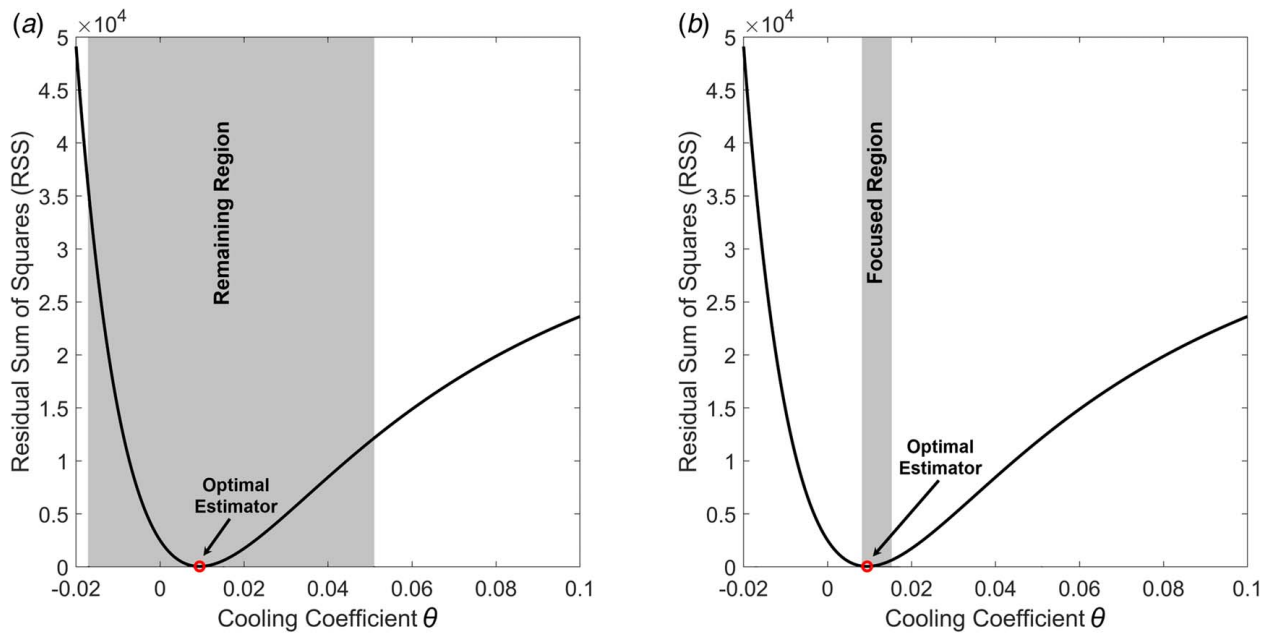


Fig. 3 Focused region identification process for the case study of the Rumford cooling experiment in Sec. 5.1: (a) ruled-out regions identification through expected fitting accuracy and (b) focused regions identification using derivatives of the squared error of the fit

Since no new area can be ruled out from the remaining region by Eq. (40), $\theta \in [0.008207, 0.01505]$ is defined as the focused region to search for the optimal parameter estimator that corresponds to the global minimum of the squared error of the fit. The focused region is shown as the gray area in Fig. 3(b). Figure 3(b) indicates that the focused regions identification method presented in this paper significantly reduces the search space for the optimal least squares parameter estimator θ^* in this case study from an infinite space $\theta \in (-\infty, \infty)$ to a small region with its length of 0.006843.

Exhaustive search with four significant digits in the focused region $\theta \in [0.008207, 0.01505]$ yields the optimal least squares parameter estimator for the cooling coefficient in the model as $\theta^* = 0.009415$, which is shown as the dot in Fig. 3(b). The coefficient of determination, R^2 , for this result is 0.8682. The RSS is 44.16, and it is the global minimum of the nonlinear least squares parameter estimation problem. Notably, this result is the same with the parameter estimator derived in previous studies [16]. However, previous studies using graphical exploration and iterative methods to search for the least squares parameter estimator cannot guarantee that the result is the optimal estimator that corresponds to the global minimum of the squared error of the fit, while the focused regions identification method does.

The effectiveness of the focused regions identification method could be compared with extant iterative least squares methods. The parameter estimators derived through the Levenberg–Marquardt method using eight different initial guesses are shown in Table 2. Notably, Fig. 3 only shows a small portion of the parameter space $\theta \in (-\infty, \infty)$, and several local minima exist in the parameter space. The results in Table 2 show that the parameter estimator derived through the Levenberg–Marquardt method is sensitive to the initial guess. The Levenberg–Marquardt iterative algorithm converges to the optimal parameter estimator $\theta^* = 0.009415$ only when the initial guess employed to initialize the iterative algorithm is close to the optimal estimator. Importantly, the Levenberg–Marquardt iterative algorithm stops when it finds a local minimum in this case study. Practitioners are not able to know whether the parameter estimator derived through the Levenberg–Marquardt method is the optimal estimator that corresponds to the global minimum of the squared error of the fit, so it is difficult for practitioners to determine when the search process should be halted. It is also not feasible to cover the infinite parameter space $\theta \in (-\infty, \infty)$ with an unlimited number of initial guesses in this case study. In contrast, using the focused regions identification method in this case study, the regions in which the optimal parameter estimator cannot be located

Table 2 The parameter estimators derived through the Levenberg–Marquardt method for the Rumford cooling experiment case study in Sec. 5.1

No.	Initial guess θ_0	Solution θ	RSS	R^2
1	-100	Nonconvergence	N/A	N/A
2	-10	-10	Infinite	-Infinite
3	-1	-0.7073	7.588×10^{28}	-2.266×10^{26}
4	0.01	0.009415	44.16	0.8682
5	1	0.009415	44.16	0.8682
6	5	4.996	4.269×10^4	-126.5
7	10	10	4.269×10^4	-126.5
8	100	100	4.269×10^4	-126.5

are ruled out from the parameter space, and the remaining region is defined as the focused region. Practitioners are therefore guaranteed to find the optimal parameter estimator by performing an exhaustive search in the focused region, and practitioners can stop the search process once the optimal parameter estimator is derived.

The effectiveness of the focused regions identification method also could be compared with extant heuristic methods. Genetic algorithm is employed to search for the optimal least squares parameter estimator in this case study. The population size of the genetic algorithm is specified as 1000 for demonstration purposes. Since random variation is involved in the iterative search process of the algorithm, the genetic algorithm initialized by 1000 random seeds is run 100 times and the parameter estimator that has the minimum RSS is chosen as the result. The parameter estimator derived by the genetic algorithm is $\theta=0.009771$, which has 3.781% deviation from the optimal parameter estimator. The coefficient of determination, R^2 , is 0.8605, and the RSS is 46.72. This result indicates that the fitting accuracy of the genetic algorithm with the population size of 1000 is lower than that of the focused regions identification method in this case study. Although the genetic algorithm with a larger population size may be able to find the optimal parameter estimator in this case study, practitioners never know whether they should run the algorithm again because they do not know the result is the global minimum or not. In contrast, using the focused regions identification method, practitioners are guaranteed that the optimal parameter estimator is located in the focused regions, and they can stop the search process once the optimal parameter estimator is derived.

5.2 The Puromycin Experiment. In the puromycin experiment, the substrate concentration and reaction velocity are measured once an enzyme is treated with puromycin. The 12 data points collected from the puromycin experiment appear in Table 3 [16]. The Michaelis–Menten model [7], also called the Michaelis–Menten equation, is fitted to the experimental data set by finding the optimal least squares estimators for the two parameters in the model. The Michaelis–Menten model is one of the best-known models of enzyme kinetics [6]. Fitting the Michaelis–Menten model to the data set collected from the puromycin experiment has been used as a benchmark problem in nonlinear regression analysis [16,34]. The mathematical form of the Michaelis–Menten model is

$$y = \frac{\theta_1 x}{x + \theta_2} \quad (42)$$

where x represents substrate concentration, y denotes reaction velocity, parameter θ_1 represents the maximum reaction velocity achieved in a biochemical experiment, and parameter θ_2 is the Michaelis constant. Both parameters θ_1 and θ_2 in Eq. (42) are

Table 3 Reaction velocity versus substrate concentration data set for the puromycin experiment in Sec. 5.2

No.	Substrate concentration x (parts per million)	Reaction velocity y (counts/min ²)
1	0.02	47
2	0.02	76
3	0.06	97
4	0.06	107
5	0.11	123
6	0.11	139
7	0.22	152
8	0.22	159
9	0.56	191
10	0.56	201
11	1.10	200
12	1.10	207

positive constants. These two parameters are often represented by V_{\max} and K_M , respectively, in biochemistry.

The focused regions identification method presented in Secs. 3 and 4 is employed to find the optimal least squares estimators for parameters θ_1 and θ_2 in Eq. (42). The squared error of the fit for the problem, also called the RSS, is expressed as

$$S(\theta) = \sum_{i=1}^{12} \left[y_{(i)} - \frac{\theta_1 x_{(i)}}{x_{(i)} + \theta_2} \right]^2 \quad (43)$$

where $\theta=[\theta_1, \theta_2]$, $x_{(i)}$ and $y_{(i)}$ represent the 12 data points in Table 3. The regions in the space of parameters θ_1 and θ_2 that cannot satisfy the expected fitting accuracy are ruled out from the parameter space first. In this case study, the expected fitting accuracy is specified as that the squared error for each data point is smaller than or equal to 1600 ($MSE_{(e)} = 1600$) for demonstration purposes. In other words, the absolute error of the predicted reaction velocity y at $x_{(i)}$ should be smaller than or equal to 40 for each data point in Table 3. Equation (15) is expressed as

$$\left\{ \theta : \left| y_{(i)} - \frac{\theta_1 x_{(i)}}{x_{(i)} + \theta_2} \right| > 40 \text{ for each } i \in \{1, 2, \dots, 12\} \right\} \quad (44)$$

Plugging the 12 data points into Eq. (44), the 16 regions that should be ruled out from the parameter space $\theta_1 \in (0, \infty)$ and $\theta_2 \in (0, \infty)$ are provided in Appendix A. Notably, the 16 regions provided in Appendix A are the union of the regions defined by the 12 inequalities represented by Eq. (44) since the absolute error for each data point cannot exceed 40. The remaining regions in the space of parameters θ_1 and θ_2 are shown in Fig. 4.

The remaining regions are further reduced using the derivatives of the squared error of the fit defined by Eq. (43). The first-order partial derivative for parameter θ_1 is

$$\frac{\partial S}{\partial \theta_1} = \sum_{i=1}^{12} \frac{2x_{(i)}}{(\theta_2 + x_{(i)})^2} (x_{(i)}\theta_1 - y_{(i)}\theta_2 - x_{(i)}y_{(i)}) \quad (45)$$

Since $x_{(i)} > 0$ and $(\theta_2 + x_{(i)})^2 > 0$, the ruled-out regions are defined by

$$\left\{ \theta : x_{(i)}\theta_1 - y_{(i)}\theta_2 - x_{(i)}y_{(i)} > 0 \text{ for all } i \in \{1, 2, \dots, 12\} \right\} \quad (46)$$

or

$$\left\{ \theta : x_{(i)}\theta_1 - y_{(i)}\theta_2 - x_{(i)}y_{(i)} < 0 \text{ for all } i \in \{1, 2, \dots, 12\} \right\} \quad (47)$$

The 12 data points in Table 3 are plugged into Eqs. (46) and (47), and the 12 regions that should be ruled out from the parameter space are provided in Appendix B. Similarly, the first-order partial derivative for parameter θ_2 is

$$\frac{\partial S}{\partial \theta_2} = \sum_{i=1}^{12} -\frac{2\theta_1 x_{(i)}}{(\theta_2 + x_{(i)})^3} (x_{(i)}\theta_1 - y_{(i)}\theta_2 - x_{(i)}y_{(i)}) \quad (48)$$

Since $x_{(i)} > 0$, $\theta_1 > 0$, and $\theta_2 > 0$, the ruled-out regions are also defined by Eqs. (46) and (47), and no new region can be ruled out from the space of parameters θ_1 and θ_2 using the first-order partial derivative for parameter θ_2 .

The overall ruled-out regions derived by the expected fitting accuracy and the first-order derivatives are provided in Appendix C, which are the union of the 16 regions defined in Appendix A and the 12 regions defined in Appendix B. The regions that are finally left over in the space of parameters θ_1 and θ_2 are shown in Fig. 5. Since the search space for the optimal least squares parameter estimators is significantly reduced and it is feasible to perform an exhaustive search in the leftover regions shown in Fig. 5, these leftover regions are defined as the focused regions in this case study. It is also proved that no new region can be ruled out using the second-order derivatives of

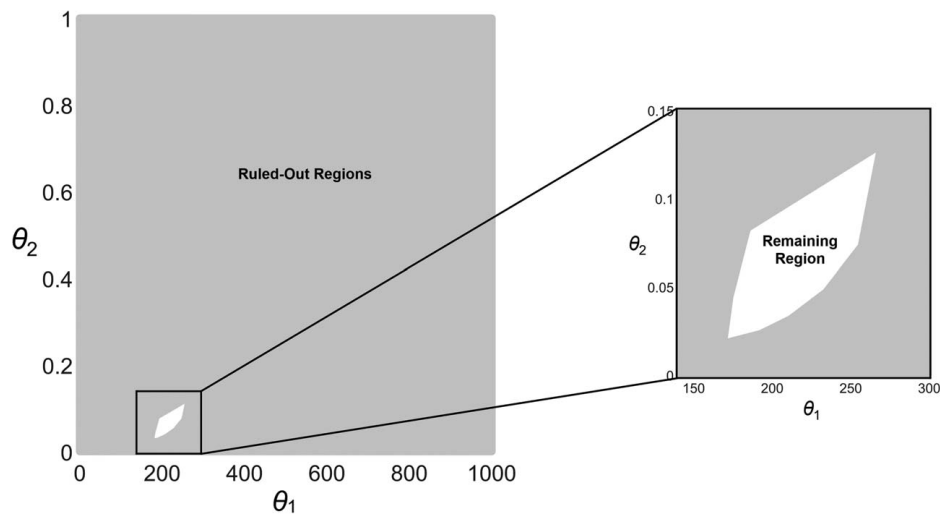


Fig. 4 Ruled-out regions derived by expected fitting accuracy in the case study of the puromycin experiment in Sec. 5.2

the squared error of the fit. Detailed mathematical derivations for the second-order derivatives are not included to conserve space.

The optimal least squares estimators for parameters θ_1 and θ_2 are derived through an exhaustive search with four significant digits in the focused regions as $\theta_1^* = 212.7$ and $\theta_2^* = 0.06412$. The optimal estimators are shown as the dot in Fig. 5. The coefficient of determination, R^2 , for this result is 0.9613. The RSS is 1195. This is the same result as previous studies yield [14,16,34]. However, to our knowledge, it is the first time that the result is proved to be the optimal parameter estimators that correspond to the global minimum of the squared error of the fit for this benchmark problem. In contrast, previous studies using other approaches cannot guarantee that the final parameter estimators derived through these approaches correspond to the global minimum of the problem. Specifically, using the focused regions identification method, for all parameter estimators that are located in the ruled-out regions defined in Appendix C, their corresponding squared error of the fit, RSS, is larger than 1600 or the first-order derivatives of the squared error of the fit are not zero, and therefore the parameter estimators $\theta_1^* = 212.7$ and $\theta_2^* = 0.06412$ derived through the exhaustive search in the focused regions are the optimal parameter

estimators that correspond to the global minimum of the squared error of the fit in this case study.

The fitting accuracy of the focused regions identification method is compared with the popular linear transformation method introduced by Lineweaver and Burk [19]. The Michaelis–Menten model is transformed into the Lineweaver–Burk equation, as shown in Eq. (2) in Sec. 2, and the values for parameters θ_1 and θ_2 are estimated through linear regression. The parameter estimators derived by the linear transformation method are $\theta_1 = 195.8$ and $\theta_2 = 0.04841$. The coefficient of determination, R^2 , is 0.9378, and the RSS is 1920. These results show that the fitting accuracy of the linear transformation method is lower than that of the focused regions identification method in this case study. Importantly, the parameter estimators derived by the linear transformation method have noticeable deviations from the optimal parameter estimators (7.945% for the maximum reaction velocity, θ_1 , and 24.50% for the Michaelis constant, θ_2). Such deviations can lead to a significant error when these estimated parameter values are employed to design other biochemical reactions.

The effectiveness of the focused regions identification method could be compared with extant iterative methods. The parameter estimators derived through the Levenberg–Marquardt method

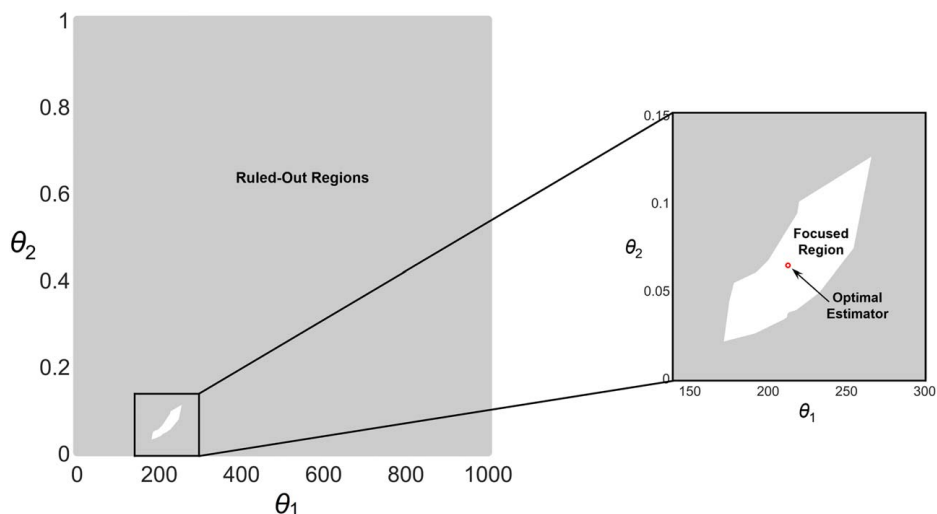


Fig. 5 Focused regions identified in the case study of the puromycin experiment in Sec. 5.2

using six different initial guesses are found in our previous study [14]. Similar to the results in Sec. 5.1, the parameter estimators derived by the Levenberg–Marquardt method are sensitive to the initial guesses. The Levenberg–Marquardt iterative algorithm converges to the optimal parameter estimators only when the initial guesses for parameter values are close to the optimal estimators. Although practitioners can find reasonable initial guesses using the solution interval method presented in our previous study [14], the solution interval method does not promise that it can output the optimal parameter estimators that minimize the squared error of the fit (global minimum) in this case study, and practitioners are therefore not able to know when to stop the search process using an iterative local method with multiple initial guesses. In contrast, practitioners are guaranteed that the parameter estimators derived by the focused regions identification method in this paper are the optimal estimators that correspond to the global minimum of the squared error of the fit in this case study, and practitioners can stop the search process once the optimal parameter estimators are derived.

The effectiveness of the focused regions identification method also could be compared with extant heuristic methods. The genetic algorithm with a population size of 1000 is run 100 times, and the parameter estimators that have the minimum RSS are selected from the 100 results. The parameter estimators derived by the genetic algorithm are $\theta_1 = 184.0$ and $\theta_2 = 0.03514$. The coefficient of determination, R^2 , is 0.8825, and the RSS is 3626. Notably, the parameter estimators derived by the genetic algorithm with the population size of 1000 have noticeable deviations from the optimal parameter estimators (13.49% for the maximum reaction velocity, θ_1 , and 45.20% for the Michaelis constant, θ_2), which can lead to significant error in biochemical reaction design. As stated in Sec. 2, the genetic algorithm with a larger population size still cannot guarantee that the results are the optimal parameter estimators that minimize the squared error of the fit, while the focused regions identification method does in this case study.

6 Outlier Detection for Nonlinear Curve Fitting Problems

The inequalities derived through the expected fitting accuracy and the derivatives of the squared error of the fit in Sec. 3, such as Eqs. (11), (15), and (23), also could be employed to detect potential outliers in nonlinear curve fitting problems. The principle of deleting one data point at a time is commonly used to detect potential outliers in linear curve fitting problems [35]. Specifically, θ represents the estimated values of parameters in a linear model computed from all the m data points, and $\theta(i)$ represents the estimated values of parameters θ in the linear model computed from $m-1$ data points (without the i th data point). The difference between θ and $\theta(i)$ indicates the impact of the i th data point on the estimated values of parameters θ in the linear model. However, parameter estimation in nonlinear models is usually more computationally expensive compared to linear models, and therefore it may not be feasible to compare θ and $\theta(i)$ for each of these m data points in practice when fitting a nonlinear model to data.

As an alternative, the inequalities derived through the expected fitting accuracy and the derivatives of the squared error of the fit in Sec. 3 could be employed to detect potential outliers in nonlinear curve fitting problems. For example, using Eq. (11), the regions that should be ruled out from the parameter space computed from all the m data points, denoted C^n , could be compared with the ruled-out regions computed from $m-1$ data points (without the i th data point), denoted $C^n(i)$, where $i \in \{1, 2, \dots, m\}$. When there is a significant difference between C^n and $C^n(i)$, practitioners are recommended to check whether the i th data point is an outlier.

The approach to detect potential outliers in nonlinear curve fitting problems is demonstrated and validated using the case study of the Rumford cooling experiment in Sec. 5.1. An outlier (time $x = 2$ min,

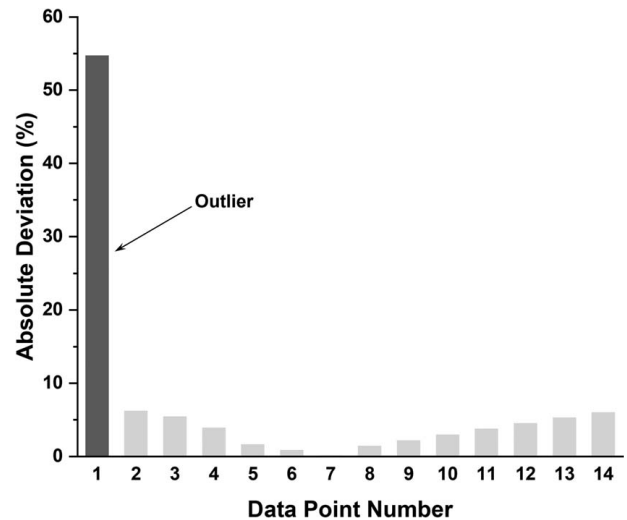


Fig. 6 Absolute percentage deviation from the benchmark length for each data point in Sec. 6

temperature $y = 128$ °F) is inserted into the original data set shown in Table 1 as the first data point in the modified data set for demonstration purposes. Equation (33) is employed to detect the potential outliers for the problem. Plugging the 14 data points (the inserted outlier and the 13 original data points in Table 1) into Eq. (33), the two regions defined by Eq. (49) should be ruled out from the parameter space $\theta \in (-\infty, \infty)$:

$$\{\theta : \theta < -0.05502 \text{ or } \theta > 0.09526\} \quad (49)$$

The remaining region is $\theta \in [-0.05502, 0.09526]$ with a length of 0.1503. The length derived by Eq. (33) using all 14 data points is set as a benchmark. The impact of the i th data point on the length of the remaining region derived by Eq. (33) is then evaluated through deleting the i th data point from the modified data set. Figure 6 shows the absolute deviation from the benchmark length for deleting each data point in the modified data set. The first data point (i.e., the inserted outlier) is identified as a potential outlier in Fig. 6 since its deviation is significantly greater than any other data point. Such a result validates the effectiveness of the potential outliers detection approach presented in this section, especially given the fact that it is difficult to identify the outlier by visual inspection in the case study of the Rumford cooling experiment since the outlier follows the temperature decreasing trend well.

7 Conclusions and Future Work

A focused regions identification method is introduced for least squares parameter estimation in nonlinear models. The novelty of this method is that the method uses expected fitting accuracy (e.g., the coefficient of determination, R^2 , and MSE) and derivatives of the squared error of the fit to rule out the suboptimal regions from the parameter space where the optimal parameter estimators cannot exist. A new approach to detect potential outliers based on the principle of deleting one data point at a time is also proposed for nonlinear curve fitting problems. The new approach employs the inequalities derived through expected fitting accuracy and derivatives of the squared error of the fit to measure the deviation that results from deleting each data point. The application of the focused regions identification method is demonstrated through two case studies in which a model based on Newton's law of cooling and the Michaelis–Menten model are fitted to two data sets collected from the Rumford cooling experiment and the purymycin experiment, respectively. The application of the potential outliers detection approach is demonstrated through the case study of the Rumford cooling experiment.

Using the focused regions identification method, practitioners can significantly reduce the search space for the optimal parameter estimators that correspond to the global minimum of the squared error of the fit for nonlinear curve fitting problems. The ruled-out regions may include one or more local minima in which some iterative local methods, such as the Gauss–Newton method and the Levenberg–Marquardt method, can be trapped. Practitioners can then find the guaranteed optimal parameter estimators and their corresponding global minimum for the nonlinear curve fitting problem through an exhaustive search in the remaining regions in the parameter space, defined as the focused regions, if adequate computational power is available for the exhaustive search. Potential outliers in the data sets of nonlinear curve fitting problems also can be detected efficiently using the new approach proposed in this paper.

This work has limitations that offer opportunities for future research. As stated in Sec. 3, the focused regions identification method presented in this paper only applies to nonlinear models that have a closed-form expression, and the method becomes inefficient when the nonlinear model has more than ten parameters. Future research may extend the method to nonlinear models without a closed-form expression, such as a differential equation, or nonlinear models that have a greater number of parameters. In addition, in both case studies in this paper, the alternative inequalities in which a summation is not included, such as Eqs. (11), (15), and (30), are employed to rule out suboptimal regions from the parameter space. Notably, the suboptimal regions ruled out by these alternative inequalities are the subregions of the regions defined by the original inequalities derived through expected fitting accuracy and derivatives of the squared error of the fit, such as, e.g., Eqs. (10), (14), and (29). These alternative inequalities may be modified in future research to rule out larger suboptimal regions from the parameter space, and then the resulting focused regions become smaller. It will be less computationally expensive to perform an exhaustive search in the reduced focused regions to find the optimal parameter estimators for nonlinear curve fitting problems.

Acknowledgment

This material is partially supported by the Air Force Office of Scientific Research (Grant No. FA9550-18-0088). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the sponsor.

Conflict of Interest

There are no conflicts of interest.

Data Availability Statement

The authors attest that all data for this study are included in the paper.

Appendix A: The Regions That can not Satisfy the Expected Squared Error for Each Data Point in Sec. 5.2

The following 16 regions are ruled out from the parameter space by Eq. (44) using the expected squared error for each data point in Sec. 5.2. The remaining regions are shown in Fig. 4.

$$0 < \theta_1 < 169.9 \text{ and } \theta_2 > 0 \quad (\text{A1})$$

$$169.9 \leq \theta_1 \leq 173.7 \text{ and } 0 < \theta_2 < 2.299 \times 10^{-4}(-87 + \theta_1) \quad (\text{A2})$$

$$169.9 \leq \theta_1 \leq 173.7 \text{ and } \theta_2 > 5.988 \times 10^{-4}(-1837 + 11\theta_1) \quad (\text{A3})$$

$$173.7 < \theta_1 \leq 184.8 \text{ and } 0 < \theta_2 < 2.299 \times 10^{-4}(-87 + \theta_1) \quad (\text{A4})$$

$$173.7 < \theta_1 \leq 184.8 \text{ and } \theta_2 > 1.739 \times 10^{-3}(-322 + 2\theta_1) \quad (\text{A5})$$

$$184.8 < \theta_1 \leq 192.2 \text{ and } 0 < \theta_2 < 2.299 \times 10^{-4}(-87 + \theta_1) \quad (\text{A6})$$

$$184.8 < \theta_1 \leq 192.2 \text{ and } \theta_2 > 5.556 \times 10^{-4}(-36 + \theta_1) \quad (\text{A7})$$

$$192.2 < \theta_1 \leq 211.1 \text{ and } 0 < \theta_2 < 1.460 \times 10^{-4}(-411 + 3\theta_1) \quad (\text{A8})$$

$$192.2 < \theta_1 \leq 211.1 \text{ and } \theta_2 > 5.556 \times 10^{-4}(-36 + \theta_1) \quad (\text{A9})$$

$$211.1 < \theta_1 \leq 233.6 \text{ and } 0 < \theta_2 < 6.135 \times 10^{-5}(-1793 + 11\theta_1) \quad (\text{A10})$$

$$211.1 < \theta_1 \leq 233.6 \text{ and } \theta_2 > 5.556 \times 10^{-4}(-36 + \theta_1) \quad (\text{A11})$$

$$233.6 < \theta_1 \leq 256.0 \text{ and } 0 < \theta_2 < 1.042 \times 10^{-4}(-2112 + 11\theta_1) \quad (\text{A12})$$

$$233.6 < \theta_1 \leq 256.0 \text{ and } \theta_2 > 5.556 \times 10^{-4}(-36 + \theta_1) \quad (\text{A13})$$

$$256.0 < \theta_1 \leq 268.1 \text{ and } 0 < \theta_2 < 4.167 \times 10^{-4}(-2640 + 11\theta_1) \quad (\text{A14})$$

$$256.0 < \theta_1 \leq 268.1 \text{ and } \theta_2 > 5.556 \times 10^{-4}(-36 + \theta_1) \quad (\text{A15})$$

$$\theta_1 > 268.1 \text{ and } \theta_2 > 0 \quad (\text{A16})$$

Appendix B: The Regions That are Ruled out Using the First-Order Derivatives in Sec. 5.2

The following 12 regions are ruled out from the parameter space by Eqs. (46) and (47) using the first-order derivatives in Sec. 5.2.

$$0 < \theta_1 \leq 47.00 \text{ and } \theta_2 > 0 \quad (\text{B1})$$

$$47.00 < \theta_1 \leq 192.0 \text{ and } \theta_2 > 4.255 \times 10^{-4}(-47 + \theta_1) \quad (\text{B2})$$

$$192.0 < \theta_1 \leq 198.9 \text{ and } \theta_2 > 8.130 \times 10^{-5}(-1353 + 11\theta_1) \quad (\text{B3})$$

$$198.9 < \theta_1 \leq 207.0 \text{ and } \theta_2 > 1.316 \times 10^{-4}(-1672 + 11\theta_1) \quad (\text{B4})$$

$$207.0 < \theta_1 \leq 213.6 \text{ and } 0 < \theta_2 < 4.831 \times 10^{-4}(-2277 + 11\theta_1) \quad (\text{B5})$$

$$207.0 < \theta_1 \leq 213.6 \text{ and } \theta_2 > 1.316 \times 10^{-4}(-1672 + 11\theta_1) \quad (\text{B6})$$

$$213.6 < \theta_1 \leq 214.0 \text{ and } 0 < \theta_2 < 1.990 \times 10^{-4}(-2814 + 14\theta_1) \quad (\text{B7})$$

$$213.6 < \theta_1 \leq 214.0 \text{ and } \theta_2 > 1.316 \times 10^{-4}(-1672 + 11\theta_1) \quad (\text{B8})$$

$$214.0 < \theta_1 \leq 217.1 \text{ and } 0 < \theta_2 < 2.632 \times 10^{-4}(-76 + \theta_1) \quad (\text{B9})$$

$$214.0 < \theta_1 \leq 217.1 \text{ and } \theta_2 > 1.316 \times 10^{-4}(-1672 + 11\theta_1) \quad (\text{B10})$$

$$\theta_1 > 217.1 \text{ and } 0 < \theta_2 < 2.632 \times 10^{-4}(-76 + \theta_1) \quad (\text{B11})$$

$$\theta_1 > 217.1 \text{ and } \theta_2 > 2.842 \times 10^{-3}(-387.0 + 1.935\theta_1) \quad (\text{B12})$$

Appendix C: Focused Regions Identified in Sec. 5.2

As a union of the regions defined by Eqs. (A1)–(A16) and the regions defined by Eqs. (B1)–(B12), the following 32 regions are ruled out from the parameter space. The remaining regions in the parameter space are defined as the focused regions for the nonlinear curve fitting problem in Sec. 5.2. The focused regions are shown in Fig. 5.

$$0 < \theta_1 < 169.9 \text{ and } \theta_2 > 0 \quad (\text{C1})$$

$$169.9 \leq \theta_1 \leq 173.7 \text{ and } 0 < \theta_2 < 2.299 \times 10^{-4}(-87 + \theta_1) \quad (\text{C2})$$

$$169.9 \leq \theta_1 \leq 173.7 \text{ and } \theta_2 > 5.988 \times 10^{-4}(-1837 + 11\theta_1) \quad (\text{C3})$$

$$173.7 < \theta_1 \leq 176.9 \text{ and } 0 < \theta_2 < 2.299 \times 10^{-4}(-87 + \theta_1) \quad (\text{C4})$$

$$173.7 < \theta_1 \leq 176.9 \text{ and } \theta_2 > 1.739 \times 10^{-3}(-322 + 2\theta_1) \quad (\text{C5})$$

$$176.9 < \theta_1 \leq 192.0 \text{ and } 0 < \theta_2 < 2.299 \times 10^{-4}(-87 + \theta_1) \quad (\text{C6})$$

$$176.9 < \theta_1 \leq 192.0 \text{ and } \theta_2 > 4.255 \times 10^{-4}(-47 + \theta_1) \quad (\text{C7})$$

$$192.0 < \theta_1 \leq 192.2 \text{ and } 0 < \theta_2 < 2.299 \times 10^{-4}(-87 + \theta_1) \quad (\text{C8})$$

$$192.0 < \theta_1 \leq 192.2 \text{ and } \theta_2 > 8.130 \times 10^{-5}(-1353 + 11\theta_1) \quad (\text{C9})$$

$$192.2 < \theta_1 \leq 198.9 \text{ and } 0 < \theta_2 < 1.460 \times 10^{-4}(-411 + 3\theta_1) \quad (\text{C10})$$

$$192.2 < \theta_1 \leq 198.9 \text{ and } \theta_2 > 8.130 \times 10^{-5}(-1353 + 11\theta_1) \quad (\text{C11})$$

$$198.9 < \theta_1 \leq 211.1 \text{ and } 0 < \theta_2 < 1.460 \times 10^{-4}(-411 + 3\theta_1) \quad (\text{C12})$$

$$198.9 < \theta_1 \leq 211.1 \text{ and } \theta_2 > 1.316 \times 10^{-4}(-1672 + 11\theta_1) \quad (\text{C13})$$

$$211.1 < \theta_1 \leq 213.4 \text{ and } 0 < \theta_2 < 6.135 \times 10^{-5}(-1793 + 11\theta_1) \quad (\text{C14})$$

$$211.1 < \theta_1 \leq 213.4 \text{ and } \theta_2 > 1.316 \times 10^{-4}(-1672 + 11\theta_1) \quad (\text{C15})$$

$$213.4 < \theta_1 \leq 213.6 \text{ and } 0 < \theta_2 < 4.831 \times 10^{-4}(-2277 + 11\theta_1) \quad (\text{C16})$$

$$213.4 < \theta_1 \leq 213.6 \text{ and } \theta_2 > 1.316 \times 10^{-4}(-1672 + 11\theta_1) \quad (\text{C17})$$

$$213.6 < \theta_1 \leq 214.0 \text{ and } 0 < \theta_2 < 1.990 \times 10^{-4}(-2814 + 14\theta_1) \quad (\text{C18})$$

$$213.6 < \theta_1 \leq 214.0 \text{ and } \theta_2 > 1.316 \times 10^{-4}(-1672 + 11\theta_1) \quad (\text{C19})$$

$$214.0 < \theta_1 \leq 217.1 \text{ and } 0 < \theta_2 < 2.632 \times 10^{-4}(-76 + \theta_1) \quad (\text{C20})$$

$$214.0 < \theta_1 \leq 217.1 \text{ and } \theta_2 > 1.316 \times 10^{-4}(-1672 + 11\theta_1) \quad (\text{C21})$$

$$217.1 < \theta_1 \leq 218.4 \text{ and } 0 < \theta_2 < 2.632 \times 10^{-4}(-76 + \theta_1) \quad (\text{C22})$$

$$217.1 < \theta_1 \leq 218.4 \text{ and } \theta_2 > 2.842 \times 10^{-3}(-387.0 + 1.935\theta_1) \quad (\text{C23})$$

$$218.4 < \theta_1 \leq 218.6 \text{ and } 0 < \theta_2 < 2.632 \times 10^{-4}(-76 + \theta_1) \quad (\text{C24})$$

$$218.4 < \theta_1 \leq 218.6 \text{ and } \theta_2 > 5.556 \times 10^{-4}(-36 + \theta_1) \quad (\text{C25})$$

$$218.6 < \theta_1 \leq 233.6 \text{ and } 0 < \theta_2 < 6.135 \times 10^{-5}(-1793 + 11\theta_1) \quad (\text{C26})$$

$$218.6 < \theta_1 \leq 233.6 \text{ and } \theta_2 > 5.556 \times 10^{-4}(-36 + \theta_1) \quad (\text{C27})$$

$$233.6 < \theta_1 \leq 256.0 \text{ and } 0 < \theta_2 < 1.042 \times 10^{-4}(-2112 + 11\theta_1) \quad (\text{C28})$$

$$233.6 < \theta_1 \leq 256.0 \text{ and } \theta_2 > 5.556 \times 10^{-4}(-36 + \theta_1) \quad (\text{C29})$$

$$256.0 < \theta_1 \leq 268.1 \text{ and } 0 < \theta_2 < 4.167 \times 10^{-4}(-2640 + 11\theta_1) \quad (\text{C30})$$

$$256.0 < \theta_1 \leq 268.1 \text{ and } \theta_2 > 5.556 \times 10^{-4}(-36 + \theta_1) \quad (\text{C31})$$

$$\theta_1 > 268.1 \text{ and } \theta_2 > 0 \quad (\text{C32})$$

References

- [1] Rhinehart, R. R., 2016, *Nonlinear Regression Modeling for Engineering Applications: Modeling, Model Validation, and Enabling Design of Experiments*, John Wiley & Sons, Chichester, UK.
- [2] Hauser, J. R., 2009, *Numerical Methods for Nonlinear Engineering Models*, Springer Science & Business Media, Berlin/Heidelberg, Germany.
- [3] Archontoulis, S. V., and Miguez, F. E., 2015, "Nonlinear Regression Models and Applications in Agricultural Research," *Agron. J.*, **107**(2), pp. 786–798.
- [4] ASTM, 2014, *D5229/D5229M-14 Standard Test Method for Moisture Absorption Properties and Equilibrium Conditioning of Polymer Matrix Composite Materials*, ASTM International, West Conshohocken, PA, https://www.astm.org/d5229_d5229m-12.html
- [5] Crank, J., 1975, *The Mathematics of Diffusion*, Oxford University Press, Oxford, UK.
- [6] Voet, D., and Voet, J. G., 2010, *Biochemistry*, John Wiley & Sons, Hoboken, NJ.
- [7] Johnson, K. A., and Goody, R. S., 2011, "The Original Michaelis Constant: Translation of the 1913 Michaelis–Menten Paper," *Biochemistry*, **50**(39), pp. 8264–8269.
- [8] Tarantola, A., 2005, *Inverse Problem Theory and Methods for Model Parameter Estimation*, Society for Industrial and Applied Mathematics and Mathematical Programming Society, Philadelphia, PA.
- [9] Levenberg, K., 1944, "A Method for the Solution of Certain Non-Linear Problems in Least Squares," *Quart. Appl. Math.*, **2**(2), pp. 164–168.
- [10] Marquardt, D. W., 1963, "An Algorithm for Least-Squares Estimation of Nonlinear Parameters," *J. Soc. Ind. Appl. Math.*, **11**(2), pp. 431–441.
- [11] Nelder, J. A., and Mead, R., 1965, "A Simplex Method for Function Minimization," *Comput. J.*, **7**(4), pp. 308–313.
- [12] Mitchell, M., 1998, *An Introduction to Genetic Algorithms*, MIT Press, Cambridge, MA.
- [13] Kirkpatrick, S., Gelatt, C. D., and Vecchi, M. P., 1983, "Optimization by Simulated Annealing," *Science*, **220**(4598), pp. 671–680.
- [14] Zhang, G., Allaire, D., and Cagan, J., 2021, "Taking the Guess Work Out of the Initial Guess: A Solution Interval Method for Least Squares Parameter Estimation in Nonlinear Models," *ASME J. Comput. Inf. Sci. Eng.*, **21**(2), p. 021011.

- [15] Jennrich, R. I., and Ralston, M. L., 1979, "Fitting Nonlinear Models to Data," *Ann. Rev. Biophys. Bioeng.*, **8**(1), pp. 195–238.
- [16] Bates, D. M., and Watts, D. G., 1988, *Nonlinear Regression Analysis and Its Applications*, John Wiley & Sons, New York.
- [17] Björck, Å., 1996, *Numerical Methods for Least Squares Problems*, Society for Industrial and Applied Mathematics, Philadelphia, PA.
- [18] Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., and Rubin, D. B., 2013, *Bayesian Data Analysis*, CRC Press, Boca Raton, FL.
- [19] Lineweaver, H., and Burk, D., 1934, "The Determination of Enzyme Dissociation Constants," *J. Am. Chem. Soc.*, **56**(3), pp. 658–666.
- [20] Gauss, C. F., 1857, *Theory of the Motion of the Heavenly Bodies Moving About the Sun in Conic Sections: A Translation of Gauss's "Theoria Motus."* With an Appendix, Little, Brown and Company, Boston, MA. <https://archive.org/details/motionofheavenly00gausrich/page/n15/mode/2up>
- [21] Coleman, T. F., and Li, Y., 1994, "On the Convergence of Interior-Reflective Newton Methods for Nonlinear Minimization Subject to Bounds," *Math. Program.*, **67**(1–3), pp. 189–224.
- [22] Coleman, T. F., and Li, Y., 1996, "An Interior Trust Region Approach for Nonlinear Minimization Subject to Bounds," *SIAM J. Optim.*, **6**(2), pp. 418–445.
- [23] Moré, J. J., 1983, "Recent Developments in Algorithms and Software for Trust Region Methods," *Mathematical Programming: The State of the Art*, A. Bachem, B. Korte, and M. Grötschel, eds., Springer-Verlag, Berlin, Germany, pp. 258–287.
- [24] Sobieszczanski-Sobieski, J., Morris, A., and VanTooren, M., 2015, *Multidisciplinary Design Optimization Supported by Knowledge Based Engineering*, John Wiley & Sons, Chichester, UK.
- [25] Kennedy, J., and Eberhart, R., 1995, "Particle Swarm Optimization," Proceedings of the International Conference on Neural Networks, Perth, WA, Australia, Nov. 27–Dec. 1, pp. 1942–.
- [26] Glover, F., 1989, "Tabu Search—Part I," *ORSA J. Comput.*, **1**(3), pp. 190–206.
- [27] Glover, F., 1990, "Tabu Search—Part II," *ORSA J. Comput.*, **2**(1), pp. 4–32.
- [28] Folland, G., 2005, "Higher-Order Derivatives and Taylor's Formula in Several Variables," Preprint, pp. 1–4. <http://sites.math.washington.edu/~folland/Math425/taylor2.pdf>
- [29] Reed, M., and Simon, B., 1980, *Methods of Modern Mathematical Physics I: Functional Analysis*, Academic Press, Inc., San Diego, CA.
- [30] Motulsky, H. J., and Ransnas, L. A., 1987, "Fitting Curves to Data Using Nonlinear Regression: A Practical and Nonmathematical Review," *FASEB J.*, **1**(5), pp. 365–374.
- [31] Motulsky, H., and Christopoulos, A., 2004, *Fitting Models to Biological Data Using Linear and Nonlinear Regression: A Practical Guide to Curve Fitting*, Oxford University Press, Oxford, UK.
- [32] Besson, U., 2012, "The History of the Cooling Law: When the Search for Simplicity Can Be an Obstacle," *Sci. Educ.*, **21**(8), pp. 1085–1110.
- [33] Roller, D., 1950, *CASE 3 The Early Development of the Concepts of Temperature and Heat—The Rise and Decline of the Caloric Theory*, Harvard University Press, Cambridge, MA, <https://www.degruyter.com/document/doi/10.4159/harvard.9780674599161.c3/html>
- [34] Montgomery, D. C., Peck, E. A., and Vining, G. G., 2006, *Introduction to Linear Regression Analysis*, Wiley-Interscience, Hoboken, NJ.
- [35] Rousseeuw, P. J., and Leroy, A. M., 1987, *Robust Regression and Outlier Detection*, John Wiley & Sons, New York.