

Taking the Guess Work Out of the Initial Guess: A Solution Interval Method for Least-Squares Parameter Estimation in Nonlinear Models

Guanglu Zhang¹

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

Douglas Allaire

Mem. ASME
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

Fitting a specified model to data is critical in many science and engineering fields. A major task in fitting a specified model to data is to estimate the value of each parameter in the model. Iterative local methods, such as the Gauss–Newton method and the Levenberg–Marquardt method, are often employed for parameter estimation in nonlinear models. However, practitioners must guess the initial value for each parameter to initialize these iterative local methods. A poor initial guess can contribute to non-convergence of these methods or lead these methods to converge to a wrong or inferior solution. In this paper, a solution interval method is introduced to find the optimal estimator for each parameter in a nonlinear model that minimizes the squared error of the fit. To initialize this method, it is not necessary for practitioners to guess the initial value of each parameter in a nonlinear model. The method includes three algorithms that require different levels of computational power to find the optimal parameter estimators. The method constructs a solution interval for each parameter in the model. These solution intervals significantly reduce the search space for optimal parameter estimators. The method also provides an empirical probability distribution for each parameter, which is valuable for parameter uncertainty assessment. The solution interval method is validated through two case studies in which the Michaelis–Menten model and Fick’s second law are fit to experimental data sets, respectively. These case studies show that the solution interval method can find optimal parameter estimators efficiently. A four-step procedure for implementing the solution interval method in practice is also outlined. [DOI: 10.1115/1.4048811]

Keywords: parameter estimation, least squares, curve fitting, initial guess, Michaelis–Menten model, Fickian diffusion, data-driven engineering, engineering informatics, inverse methods for engineering applications, model-based systems engineering, multidisciplinary optimization

1 Introduction

Fitting a specified model to data, also called curve fitting, is an inverse problem that has wide application in many science and engineering fields. For example, researchers often fit a fatigue model, such as the stress-life (S - N) model and the strain-life (ϵ - N) model, to the cycles to failure experimental data of different metals [1]. Fatigue curves, which play a significant role in machine component design [2], are then constructed accordingly [3,4]. Similarly, Fick’s laws govern the moisture diffusion in composite materials [5]. Practitioners usually fit Fick’s second law to experimental data in order to estimate mass diffusivities of different composite materials [6]. These estimated mass diffusivities are then used to design composite structures (e.g., wings of passenger aircraft and the bodies of racing cars) under different service environments [7].

A major task in fitting a specified model to data is to estimate the value of each parameter in the model. Several objective functions, such as least-squares error and maximum likelihood, are available for parameter estimation [8]. Practitioners need to find the optimal parameter values that minimize or maximize a chosen objective function. Linear regression has been applied successfully to estimate parameter values in linear models [9]. However, many

scientific and engineering models are nonlinear [10]. Iterative local methods, such as the Gauss–Newton method [9] and the Levenberg–Marquardt method [11], and Bayesian approaches, are usually employed for parameter estimation in nonlinear models [12–14]. Importantly, practitioners must guess the initial value or the prior distribution of each parameter to initialize these methods. A poor initial guess can contribute to non-convergence of these methods or lead these methods to converge to a wrong solution or a sub-optimal solution (e.g., local optimum) [15]. Moreover, practitioners may not know whether the result is the optimal solution or not even when a good initial guess is employed to initialize these methods.

In this paper, a broadly applicable method is introduced to estimate parameter values in a nonlinear model that minimizes the squared error of the fit when the number of unknown parameters in the model is less than the number of data points. To initialize this method, it is not necessary for practitioners to guess the initial value of each parameter in a nonlinear model. To our knowledge, an initial guess free method for least-squares parameter estimation that can be applied to fit any specified nonlinear models to data is a new contribution to curve fitting. This method establishes an empirical probability distribution for each parameter in a nonlinear model. This method also constructs a solution interval for each parameter in a nonlinear model. It is proved that the optimal estimator for each parameter in a nonlinear model that minimizes the squared error of the fit belongs to its corresponding solution interval

¹Corresponding author.

Manuscript received February 23, 2020; final manuscript received October 13, 2020; published online December 10, 2020. Assoc. Editor: John Michopoulos.

under certain conditions. The optimal least-squares estimator for each parameter could be found through exhaustive search with specified significant digits (e.g., four significant digits) in the solution interval. The procedure to compute an initial value for each parameter in a nonlinear model is provided when an iterative local method is preferred for parameter estimation. A stochastic algorithm is also developed to find the optimal parameter estimators efficiently.

This paper begins with a short review of currently available methods for parameter estimation in nonlinear models in Sec. 2. The new method to construct a solution interval and search for the optimal least-squares estimator for each parameter in a nonlinear model is presented in Sec. 3. To ease the implementation of this method in practice, four steps to implement this method are summarized in Sec. 4. In Sec. 5, the application of this method is illustrated by fitting the Michaelis–Menten model and Fick’s second law to experimental data sets, respectively. The paper concludes with a brief discussion of the contribution of the work presented here and future research directions.

2 Background and Related Work

The goal of parameter estimation in nonlinear models is to find the optimal estimator for each parameter in a specified nonlinear model that minimizes or maximizes a chosen objective function. Available objective functions include, but are not limited to, least-squares error, weighted least-squares error, maximum likelihood, and robust regression functions for M-estimators [8]. Due to space limitations, this section only summarizes several popular methods to find the optimal parameter estimators in nonlinear models that minimize the squared error of the fit. A brief review of Bayesian approaches for parameter estimation is also provided. A thorough review of available methods for parameter estimation in nonlinear models using different objective functions could be found in books, review papers, and a technical report written by Bates and Watts [16], Björck [13], Jennrich and Ralston [8], Motulsky and Ransnas [15] and Fraley [17], among others.

A simple approach for parameter estimation in nonlinear models is to transform a nonlinear model into a linear model. Linear regression is then performed to solve the estimator for each parameter in the transformed model. Approaches based on linear transformations are easy to implement, but such approaches usually cannot produce the optimal least-squares parameter estimators [9]. For example, Moore’s law is commonly used in the semiconductor industry to describe the performance change of central processing unit (CPU) and dynamic random-access memory (RAM) over time [18]. Moore’s law is an exponential function as

$$y = Ae^{Bt} \quad (1)$$

where y is the technology performance (e.g., CPU transistor count), t is the time, and A and B are constant parameters. A natural logarithm transformation is often made to Eq. (1) as

$$\ln(y) = \ln(A) + Bt \quad (2)$$

when practitioners fit Moore’s law to technology performance historical data. Linear regression is sometimes used to solve the values of parameters A and B for technological forecasting. However, the technology performance data, y , have been distorted by the natural logarithm transformation in Eq. (2). The linear least-squares estimators for parameters A and B found from solving Eq. (2) minimize the summation of squared residuals on $\ln(y)$ rather than on y ; thus, these linear least-squares estimators are not the optimal least-squares parameter estimators for the original problem represented by Eq. (1). A similar example could be found in Sec. 5.1 where the Michaelis–Menten model is fit to experimental data.

An iterative least-squares local method is commonly employed to estimate parameter values in nonlinear models since the linear transformation approaches usually cannot produce the optimal least-squares parameter estimators, and many nonlinear models

even cannot be transformed into a linear model. An iterative local approach begins with an initial guess for the value of each parameter in a nonlinear model and then alters the values of parameters iteratively to reduce the squared error of the fit. Among currently available local methods that have been developed to direct these iterations, the Gauss–Newton method and the Levenberg–Marquardt method are two popular methods in science and engineering fields [9,13].

The Gauss–Newton method is based on a sequence of linear approximation of least-squares error [13]. Using Taylor’s theorem, the residuals of a nonlinear fit are approximated as

$$\mathbf{r}(\boldsymbol{\theta}) \approx \mathbf{r}(\boldsymbol{\theta}^{(s)}) + \mathbf{J}(\boldsymbol{\theta}^{(s)})\Delta \quad (3)$$

where $\Delta = \boldsymbol{\theta} - \boldsymbol{\theta}^{(s)}$, $\mathbf{r} = [r_1, r_2, \dots, r_m]$ is a column vector of residuals, $\boldsymbol{\theta} = [\theta_1, \theta_2, \dots, \theta_n]$ is a column vector of parameters in a nonlinear model $y = f(x_1, x_2, \dots, x_q, \boldsymbol{\theta})$ that is fit to m data points $[x_{1(i)}, x_{2(i)}, \dots, x_{q(i)}, y_{(i)}]$, and \mathbf{J} is $m \times n$ Jacobian matrix of residuals as

$$(\mathbf{J})_{ij} = \frac{\partial r_i}{\partial \theta_j} \quad (4)$$

where $i \in \{1, 2, \dots, m\}$, $j \in \{1, 2, \dots, n\}$. The squared error of the fit, S , is the squared Euclidean norm of residual vector \mathbf{r} as

$$\begin{aligned} S(\boldsymbol{\theta}) &= \sum_{i=1}^m [y_{(i)} - f(x_{1(i)}, x_{2(i)}, \dots, x_{q(i)}, \boldsymbol{\theta})]^2 \\ &= \sum_{i=1}^m [r_i(\boldsymbol{\theta})]^2 = \|\mathbf{r}(\boldsymbol{\theta})\|_2^2 \end{aligned} \quad (5)$$

The iterative equation of the Gauss–Newton method is

$$\boldsymbol{\theta}^{(s+1)} = \boldsymbol{\theta}^{(s)} + \Delta^{(s)} \quad (6)$$

The iterative increment vector $\Delta^{(s)}$ is derived through minimizing approximate least-squares error defined by Eq. (3) as

$$\min \|\mathbf{r}(\boldsymbol{\theta}^{(s)}) + \mathbf{J}(\boldsymbol{\theta}^{(s)})\Delta^{(s)}\|_2^2 \quad (7)$$

Equation (7) could be converted to solve n simultaneous linear equations as

$$\mathbf{J}(\boldsymbol{\theta}^{(s)})^T \mathbf{J}(\boldsymbol{\theta}^{(s)})\Delta^{(s)} = -\mathbf{J}(\boldsymbol{\theta}^{(s)})^T \mathbf{r}(\boldsymbol{\theta}^{(s)}) \quad (8)$$

where \mathbf{J}^T is the transpose of the Jacobian matrix \mathbf{J} . The vector $\Delta^{(s)}$ may be solved, for example, via a decomposition of the matrix into an orthogonal matrix and a triangular matrix (QR decomposition) at each iteration step [13,19].

The Gauss–Newton method has the advantage that it is only necessary to solve linear equations represented by Eq. (8) in each iteration step. However, the Gauss–Newton method may converge slowly for large residual problems [13,20]. Moreover, the method fails to converge when the matrix $\mathbf{J}^T \mathbf{J}$ is singular [21]. To solve the convergence issue of the Gauss–Newton method, several modified Gauss–Newton methods, also known as Newton-type methods [13], have been developed. These modified Gauss–Newton methods [22] take second-order derivatives (i.e., Hessian matrices) into account when the residuals of a nonlinear fit are approximated compared to the first-order approximation in Eq. (3). It is usually expensive to compute Hessian matrices of the residuals in each iteration step, and researchers have therefore suggested different methods of approximating Hessian matrices [23,24].

Levenberg [25] and Marquardt [26] introduced another approach to solve the convergence issue of the Gauss–Newton method. The Levenberg–Marquardt method, also known as a trust-region method [27] or restricted step method [28], use

$$[\mathbf{J}(\boldsymbol{\theta}^{(s)})^T \mathbf{J}(\boldsymbol{\theta}^{(s)}) + \lambda^{(s)} \mathbf{D}^{(s)}]\Delta^{(s)} = -\mathbf{J}(\boldsymbol{\theta}^{(s)})^T \mathbf{r}(\boldsymbol{\theta}^{(s)}) \quad (9)$$

to compute the increment vector Δ in each iteration step, where λ is the Marquardt parameter that controls the trust-region size, and \mathbf{D} is

a diagonal scaling matrix. Both parameter λ and matrix D are updated in each iteration step based on the change of least-squares error [26]. A variety of algorithms [27] have been developed to compute parameter λ and matrix D in each iteration step and to approximate the residuals of a nonlinear fit after the pioneering work of Marquardt [26] in 1963. The Levenberg–Marquardt method also has been extended and applied to least-squares parameter estimation subject to bounds (e.g., $\theta_{lb} < \theta_j < \theta_{ub}$, where θ_{lb} and θ_{ub} are the lower bound and upper bound of parameter θ_j respectively, in a nonlinear model) [29].

Of note, least-squares parameter estimation in nonlinear models could be regarded as a special case of engineering optimization, where least-squares error is the optimization objective and parameters in a specified nonlinear model are the design variables. Parameter estimation methods, especially trust-region methods, have been generalized and commonly employed to solve unconstrained and constrained optimization problems [30,31]. Similarly, several guided random search techniques invented for engineering optimization [32], such as genetic algorithms, simulated annealing, and particle swarm optimization, also have been applied for parameter estimation in nonlinear models [33–36]. These guided random search techniques are iterative methods that usually begin with multiple initial guesses (i.e., initial population) for the value of each parameter in a nonlinear model. A process involving random variation in parameter values is then performed to improve an objective function in each iteration step [32].

Bayesian approaches are also popular for parameter estimation in nonlinear models. Parameters in a nonlinear model are considered random variables in Bayesian approaches. The distribution of parameter θ conditional on observed data \mathbf{B} , $P(\theta|\mathbf{B})$, also called the posterior distribution, is derived using Bayes' theorem as

$$P(\theta|\mathbf{B}) = \frac{P(\theta, \mathbf{B})}{P(\mathbf{B})} = \frac{P(\theta)P(\mathbf{B}|\theta)}{P(\mathbf{B})} = \frac{P(\theta)P(\mathbf{B}|\theta)}{\int P(\theta)P(\mathbf{B}|\theta)d\theta} \quad (10)$$

where $P(\theta)$ is a user-specified prior distribution of parameter θ , and $P(\mathbf{B}|\theta)$ is a data distribution [14]. The posterior expectation of parameter θ is

$$E(\theta|\mathbf{B}) = \int \theta P(\theta|\mathbf{B})d\theta = \frac{\int \theta P(\theta)P(\mathbf{B}|\theta)d\theta}{\int P(\theta)P(\mathbf{B}|\theta)d\theta} \quad (11)$$

In most applications involving nonlinear models, it is difficult to evaluate the integrations in Eq. (11) analytically. Laplace approximation [37] and Markov chain Monte Carlo methods [38] have been applied to obtain the posterior expectation for each parameter in a nonlinear model.

Iterative least-squares local methods (e.g., the Gauss–Newton method, the Levenberg–Marquardt method, and guided random search techniques) and Bayesian approaches have been applied to estimate parameter values in nonlinear models in many science and engineering fields [13,14,32,39]. A successful application of any method mentioned previously relies on a smart initial guess. Specifically, an initial value for each parameter in a nonlinear model must be specified when an iterative least-squares local method is employed. Bayesian approaches require a user-specified prior distribution for each parameter. A poor initial guess can contribute to non-convergence of these methods or lead these methods to converge to a wrong solution or a sub-optimal solution (e.g., local optimum) [15,40]. However, there is no standard procedure available to specify initial values for least-squares iterative local methods or prior distributions for Bayesian approaches [15,41,42]. In practice, an initial guess for parameter estimation in nonlinear models is made on a case-by-case basis and is often based on previous experience, expert advice, linear least-squares estimators (if linear transformation is feasible for a nonlinear model), graphical exploration, or a grid search (i.e., “brute force” approach) [15,40–42].

3 A Solution Interval Method to Estimate Parameter Values in Nonlinear Models

Based on the research gap discussed in Sec. 2, a solution interval method for least-squares parameter estimation in nonlinear models is introduced in this section. This method includes three algorithms that require different levels of computational power to find the optimal parameter estimators. Importantly, it is not necessary to guess the initial value for each parameter in a nonlinear model using the method.

As a generic problem of curve fitting, there are m data points $[x_{1(i)}, x_{2(i)}, \dots, x_{q(i)}, y(i)]$, where $i \in \{1, 2, \dots, m\}$. Practitioners want to fit a nonlinear model $y=f(\mathbf{x}, \theta)$ to the m data points and find the optimal values for parameters θ that minimize the squared error of the fit, where $\theta = [\theta_1, \theta_2, \dots, \theta_n]$ represents n constant parameters ($n < m$) in the nonlinear model, and $\mathbf{x} = [x_1, x_2, \dots, x_q]$. Here, $y=f(\mathbf{x}, \theta)$ is a specified nonlinear model, and model selection is beyond the scope of this paper. The least-squares error of the fit is

$$S(\hat{\theta}) = \sum_{i=1}^m [y(i) - f(x_{1(i)}, x_{2(i)}, \dots, x_{q(i)}, \hat{\theta})]^2 \quad (12)$$

where $\hat{\theta} = [\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_n]$ are the estimated values of parameters θ .

To find the optimal parameter estimators using the solution interval method, the first step is to derive *parameter solution equations*. n data points $[x_{1(i_k)}, x_{2(i_k)}, \dots, x_{q(i_k)}, y(i_k)]$ are chosen from the m data points $[x_{1(i)}, x_{2(i)}, \dots, x_{q(i)}, y(i)]$, where $k \in \{1, 2, \dots, n\}$, $i_k \in \{1, 2, \dots, m\}$, and $i \in \{1, 2, \dots, m\}$. There are h combinations for choosing n data points $[x_{1(i_k)}, x_{2(i_k)}, \dots, x_{q(i_k)}, y(i_k)]$ from the m data points $[x_{1(i)}, x_{2(i)}, \dots, x_{q(i)}, y(i)]$, where

$$h = C_m^n = \frac{m!}{(m-n)!n!} \quad (13)$$

n equations are derived when any chosen n data points are plugged into the nonlinear model $y=f(\mathbf{x}, \theta)$ respectively as

$$y(i_k) - f(x_{1(i_k)}, x_{2(i_k)}, \dots, x_{q(i_k)}, \theta) = 0 \quad (14)$$

The n equations represented by Eq. (14) are called *parameter solution equations*. The unknown variables in these n equations are $\theta = [\theta_1, \theta_2, \dots, \theta_n]$. If the n equations represented by Eq. (14) only have one solution for one combination as $\theta_{(s)} = [\theta_{1(s)}, \theta_{2(s)}, \dots, \theta_{n(s)}]$, where $s \in \{1, 2, \dots, h\}$, there are h solutions for each parameter θ_j in total, where $j \in \{1, 2, \dots, n\}$. Of note, although $y=f(\mathbf{x}, \theta)$ is a nonlinear model, the n equations represented by Eq. (14) may be linear equations or nonlinear equations that have analytical solutions. Numerical methods, such as the bisection method, Brent's method, and the Newton–Raphson method [19], could be employed to solve these n equations when analytical solutions are not available. If there are any bounds applied to parameter θ_j (e.g., $\theta_{lb} < \theta_j < \theta_{ub}$, where θ_{lb} and θ_{ub} are the lower bound and upper bound of parameter θ_j), it is only necessary to find the solution of Eq. (14) for θ_j that satisfies these bounds.

The h solutions for each parameter θ_j are sorted from the smallest solution θ_j^{\min} to the largest solution θ_j^{\max} , where $\theta_j^{\min} \leq \theta_{j(s)} \leq \theta_j^{\max}$, $j \in \{1, 2, \dots, n\}$, and $s \in \{1, 2, \dots, h\}$. The empirical probability distribution of each parameter θ_j could be built through setting the same probability, $1/h$, at each solution $\theta_{j(1)}, \theta_{j(2)}, \dots, \theta_{j(h)}$ if necessary. The length of the interval $[\theta_j^{\min}, \theta_j^{\max}]$ is defined as

$$L_j = \theta_j^{\max} - \theta_j^{\min} \quad (15)$$

The midpoint of the interval $[\theta_j^{\min}, \theta_j^{\max}]$ is

$$\theta_j^{\text{mid}} = \frac{\theta_j^{\min} + \theta_j^{\max}}{2} \quad (16)$$

It is proven in Appendix A that the optimal estimator for each parameter θ_j that minimizes the squared error S belongs to the interval $[\theta_j^{\text{mid}} - L_j, \theta_j^{\text{mid}} + L_j]$ when the monotonic conditions and the

symmetric condition defined by Eqs. (A8)–(A10) are satisfied. $[\theta_j^{\text{mid}} - L_j, \theta_j^{\text{mid}} + L_j]$ is called the *solution interval* of parameter θ_j . Future research may relax these conditions given by Eqs. (A8)–(A10) in Appendix A and define a solution interval for the more general case or derive the solution interval on a case-by-case basis.

Here, three algorithms that require different levels of computational power are provided to find the optimal estimator for each parameter in a nonlinear model using the solutions of *parameter solution equations*, $\theta_{j(s)}$, and the solution interval $[\theta_j^{\text{mid}} - L_j, \theta_j^{\text{mid}} + L_j]$ for each parameter θ_j . As the first algorithm, the optimal parameter estimators that minimize the squared error S could be found through an exhaustive search, also called grid search, with specified significant digits (e.g., four significant digits) in the *solution interval* of each parameter θ_j if abundant computational power is available. The second algorithm could be employed if exhaustive search is not feasible. For example, the solution intervals derived by Eqs. (15) and (16) may be extremely large because of outliers in the data set or numerical ill-conditioning in solving the parameter solution equations. The median of the h solutions of each parameter θ_j , θ_j^{med} could be found in that case. The median of parameter solutions, $\theta^{\text{med}} = [\theta_1^{\text{med}}, \theta_2^{\text{med}}, \dots, \theta_n^{\text{med}}]$, is supplied as the initial values to a currently available iterative least-squares local method (e.g., the Gauss–Newton method or the Levenberg–Marquardt method) to search the optimal estimator for each parameter.

It is necessary to solve the n equations represented by Eq. (14) h times using either algorithms mentioned above (i.e., exhaustive search in the solution interval or use the median of h parameter solutions as initial values of an iterative least-squares local method). In some cases, it may be expensive to solve these n equations h times. For example, a nonlinear model with 10 parameters is fit to 1000 data points (i.e., $m = 1000$, and $n = 10$). There are 2.63×10^{23} combinations to choose 10 data points from the 1000 data points (i.e., $h = 2.63 \times 10^{23}$). Enormous computational power is required to solve the 10 equations represented by Eq. (14) 2.63×10^{23} times if analytical solutions of these 10 equations are not available, and a numerical method has to be employed. The following stochastic algorithm, as the third algorithm, is an alternative to search for the optimal value for each parameter in a nonlinear model when there is not enough computational power available to solve these n equations h times.

A flowchart of the stochastic algorithm appears in Fig. 1. As defined earlier, there are h combinations to choose n data points from the m data points ($n < m$). Three combinations are chosen from the h combinations randomly without replacement. Each combination corresponds to one solution for each parameter θ_j through solving the n equations represented by Eq. (14), where $j \in \{1, 2, \dots, n\}$. A solution interval $[a_j^{(1)}, b_j^{(1)}] = [\theta_j^{\text{mid}} - L_j, \theta_j^{\text{mid}} + L_j]$ could be built for each parameter θ_j using these three solutions. The median of these three solutions is used as the initial value for each parameter θ_j of a specified iterative least-squares local method (e.g., the Gauss–Newton method or the Levenberg–Marquardt method) to search for the optimal parameter estimators. The search result for parameter θ_j is denoted by $\theta_j^{(1)}$. If search result $\theta_j^{(1)}$ for each parameter θ_j belongs to the corresponding solution interval $[a_j^{(1)}, b_j^{(1)}]$, $\theta^{(1)} = [\theta_1^{(1)}, \theta_2^{(1)}, \dots, \theta_n^{(1)}]$ is the final parameter estimators. If any search result $\theta_j^{(1)}$ does not belong to the corresponding solution interval $[a_j^{(1)}, b_j^{(1)}]$, another two combinations are then chosen from the $(h-3)$ combinations randomly without replacement. A solution interval $[a_j^{(2)}, b_j^{(2)}] = [\theta_j^{\text{mid}} - L_j, \theta_j^{\text{mid}} + L_j]$ is built using all five available solutions for each parameter θ_j , and the median of the five solutions is used as the initial value for each parameter θ_j of the specified iterative least-squares local method. If search result $\theta_j^{(2)}$ for each parameter θ_j belongs to the corresponding solution interval, $[a_j^{(2)}, b_j^{(2)}]$, $\theta^{(2)} = [\theta_1^{(2)}, \theta_2^{(2)}, \dots, \theta_n^{(2)}]$ is the final parameter estimators. If any search result $\theta_j^{(2)}$ does not belong to the corresponding solution interval $[a_j^{(2)}, b_j^{(2)}]$, another two combinations are then chosen from the remaining combinations randomly without replacement repeatedly until the new search

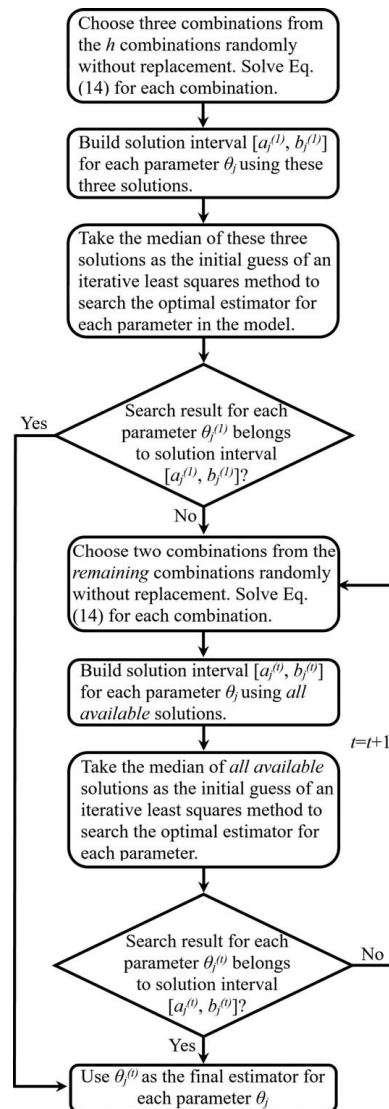


Fig. 1 A stochastic algorithm to estimate parameter values in a nonlinear model

result $\theta_j^{(t)}$ for each parameter θ_j belongs to the corresponding solution interval $[a_j^{(t)}, b_j^{(t)}]$, where $t < h/2$. The convergence rate of the stochastic algorithm depends on the specified nonlinear model and the iterative least-squares local method employed in each iteration step. A case study using the stochastic algorithm could be found in Sec. 5.2 where Fick’s second law is fit to experimental data and the Levenberg–Marquardt method is employed in each iteration step. The stochastic algorithm reaches 100% convergence within five iteration steps, and the parameter solution equation represented by Eq. (14) is solved less than $0.2 \cdot h$ times in the case study. The convergence rate of the stochastic algorithm could be investigated through more case studies in future research.

As stated earlier, the solution interval method provides practitioners three algorithms to choose from based on available computational power. The first algorithm is more computationally expensive than the other two algorithms. However, the first algorithm is complete since exhaustive search in the solution interval of each parameter is performed. It is proven in Appendix A that the optimal estimator for each parameter belongs to its corresponding solution interval when the monotonic conditions and the symmetric condition defined by Eqs. (A8)–(A10) are satisfied. The soundness of the second algorithm (i.e., using median of solutions as initial value) and the third algorithm (i.e., the stochastic algorithm shown in Fig. 1) depends on whether a sound iterative local method

is employed in these algorithms. Of note, the third algorithm is equivalent to the second algorithm when the stochastic algorithm does not converge before all h combinations are chosen. All h combinations have been chosen when the third algorithm terminates in the end, and h solutions are derived for each parameter. The median of the h parameter solutions is then taken as the initial value for the specified iterative local method, which is identical to the second algorithm.

4 Four Steps to Estimate Parameter Values in a Nonlinear Model

In Sec. 3, a generic least-squares method, called the solution interval method, is introduced to estimate parameter values in a nonlinear model. The method includes three algorithms that require different levels of computational power to find the optimal parameter estimators, and practitioners do not need to guess the initial value for each parameter in a nonlinear model using the method. In this section, the procedure to implement the

solution interval method is summarized in four steps. These four steps are illustrated in Fig. 2, and the three algorithms appear as the two blocks on the right side of step 2 and step 3 and the block of step 4 in Fig. 2. Practitioners can follow these guidelines to choose an appropriate algorithm and find the optimal estimator for each parameter in a nonlinear model that minimizes the squared error of the fit.

Step 1—Data collection and model selection. Practitioners first collect data (i.e., m data points) from scientific experiment, observational study, simulation, or survey. An appropriate mathematical model with n unknown parameters ($n < m$) is selected from several candidate models to fit the data. If the mathematical model is linear, linear regression is applied to solve the parameter values in the model.

Step 2—Parameter solution equations derivation. If the chosen mathematical model is nonlinear, the expressions of n equations represented by Eq. (14) are derived from the specified nonlinear model. These n equations are then solved analytically. If analytic solutions are not available, a numerical

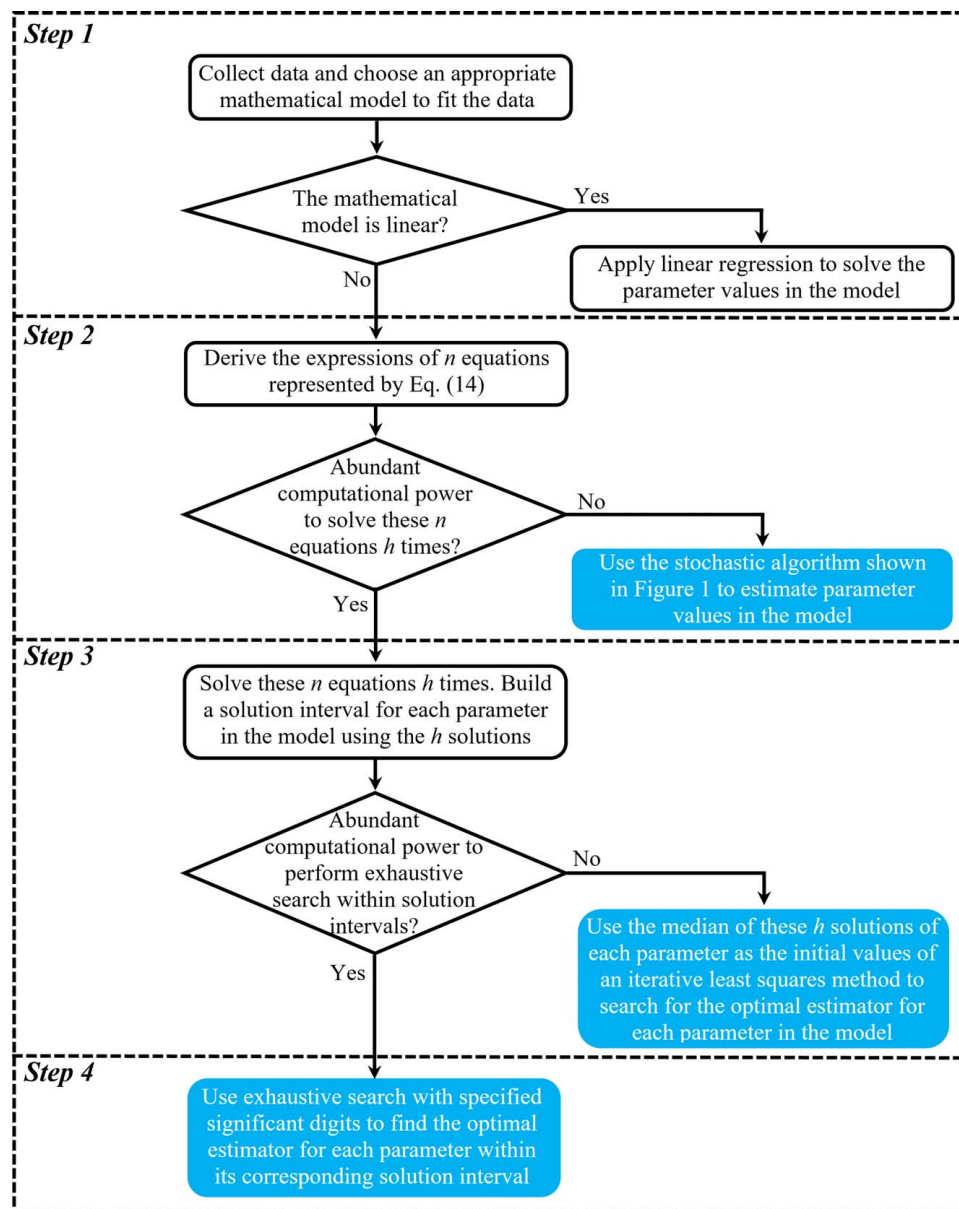


Fig. 2 Four steps to apply the solution interval method for parameter estimation in a nonlinear model

method (e.g., the bisection method, Brent's method, or the Newton–Raphson method) should be specified to solve these n equations. Practitioners evaluate whether there is abundant computational power to solve these n equations h times using the numerical method, where h is defined in Eq. (13). If limited computational power is available, practitioners should use the stochastic algorithm shown in Fig. 1 to estimate the value for each parameter in the nonlinear model.

Step 3—Solution interval construction. If there is abundant computational power to solve the n equations represented by Eq. (14) h times, n data points are chosen from the m data points exhaustively and plugged into these n equations, respectively. h solutions are derived for each parameter in the nonlinear model. These h solutions are sorted from the smallest solution to the largest solution, and a solution interval $[\theta_j^{\text{mid}} - L_j, \theta_j^{\text{mid}} + L_j]$ is constructed for each parameter θ_j using Eqs. (15)–(16). Practitioners evaluate whether there is abundant computational power to perform an exhaustive search within the solution interval of each parameter to find the optimal estimator. If limited computational power is available, practitioners should use the median of these h solutions of each parameter as the initial value of an iterative least-squares local method (e.g., the Gauss–Newton method or the Levenberg–Marquardt method) to search for the optimal estimator for each parameter.

Step 4—Optimal parameter estimator search. If abundant computational power is available for exhaustive search in the solution interval of each parameter, practitioners specify the significant digits (e.g., four significant digits) required for final parameter estimators and perform an exhaustive search, also called grid search, in solution intervals to find the optimal estimator for each parameter in the nonlinear model that minimizes the squared error of the fit.

5 Case Studies Using Michaelis–Menten Model and Fick's Second Law

To illustrate the four-step procedure outlined in Sec. 4, the Michaelis–Menten model and Fick's second law are fit to experimental data sets, respectively, as two case studies in this section. These two case studies show that the optimal least-squares parameter estimators belong to the solution intervals defined in Sec. 3 (i.e., $[\theta_j^{\text{mid}} - L_j, \theta_j^{\text{mid}} + L_j]$ for parameter θ_j), although it is difficult to prove analytically whether the Michaelis–Menten model and Fick's second law satisfy the conditions defined by Eqs. (A8)–(A10). These optimal least-squares parameter estimators have also been found more efficiently using the solution interval method compared with using currently available methods (e.g., the linear transformation method and iterative least-squares local methods) in these case studies. These results validate the solution interval method and associated assumptions.

5.1 Fitting Michaelis–Menten Model to Experiment Data. Michaelis–Menten model [43] is well known in biochemistry to describe the rate of enzymatic reactions [44]. Fitting the Michaelis–Menten model to experiment data has also been used as a benchmark problem to test different least-squares methods [9,16,45]. Here, the Michaelis–Menten model is fit to an experimental data set for the initial velocity of a reaction for an enzyme treated with puromycin [9]. There are 12 data points (i.e., $m = 12$) as shown in Table 1, and the Michaelis–Menten model

$$y = \frac{\theta_1 x}{x + \theta_2} \quad (17)$$

has two parameters θ_1 and θ_2 (i.e., $n = 2$), where x represents substrate concentration in the unit of parts per million (ppm), and y denotes the initial velocity of the reaction in the unit of counts/min².

The solution interval method presented in Secs. 3 and 4 is employed to find the optimal estimators for parameters θ_1 and θ_2

Table 1 Reaction velocity and substrate concentration for puromycin experiment [9]

No.	Substrate concentration, x (parts per million)	Initial 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

in the Michaelis–Menten model. There are h combinations to choose two data points $[x_{(i_1)}, y_{(i_1)}]$ and $[x_{(i_2)}, y_{(i_2)}]$ from the 12 data points, where

$$h = C_m^n = C_{12}^2 = \frac{12!}{(12-2)!2!} = 66 \quad (18)$$

$i_1 \in \{1, 2, \dots, 12\}$, $i_2 \in \{1, 2, \dots, 12\}$. Of note, as shown in Table 1, there are six combinations where the two chosen data points have different initial velocities of the reaction but the same substrate concentration (i.e., $y_{(i_1)} \neq y_{(i_2)}$, $x_{(i_1)} = x_{(i_2)}$). These six combinations are removed, and the remaining 60 combinations are used in the following steps.

Any two chosen data points $[x_{(i_1)}, y_{(i_1)}]$ and $[x_{(i_2)}, y_{(i_2)}]$ are plugged into Eq. (17) as

$$y_{(i_1)} = \frac{\theta_1 x_{(i_1)}}{x_{(i_1)} + \theta_2} \quad (19)$$

$$y_{(i_2)} = \frac{\theta_1 x_{(i_2)}}{x_{(i_2)} + \theta_2} \quad (20)$$

Equations (19) and (20) give two linear equations for parameters θ_1 and θ_2 as

$$x_{(i_1)}\theta_1 - y_{(i_1)}\theta_2 = x_{(i_1)}y_{(i_1)} \quad (21)$$

$$x_{(i_2)}\theta_1 - y_{(i_2)}\theta_2 = x_{(i_2)}y_{(i_2)} \quad (22)$$

Analytic solutions of parameters θ_1 and θ_2 are derived from Eqs. (21)–(22) as

$$\theta_1 = \frac{[x_{(i_1)} - x_{(i_2)}]y_{(i_1)}y_{(i_2)}}{x_{(i_1)}y_{(i_2)} - x_{(i_2)}y_{(i_1)}} \quad (23)$$

$$\theta_2 = \frac{[y_{(i_1)} - y_{(i_2)}]x_{(i_1)}x_{(i_2)}}{x_{(i_1)}y_{(i_2)} - x_{(i_2)}y_{(i_1)}} \quad (24)$$

The values of parameters θ_1 and θ_2 are solved for 60 combinations using Eqs. (23)–(24). The solution distributions of parameters θ_1 and θ_2 appear in Figs. 3 and 4, respectively, where $112.5 \leq \theta_1 \leq 295.8$, and $-0.005646 \leq \theta_2 \leq 0.1476$. The solution intervals for parameters θ_1 and θ_2 are derived based on Eqs. (15)–(16) as

$$\theta_1: [\theta_1^{\text{mid}} - L_1, \theta_1^{\text{mid}} + L_1] = [20.85, 387.5] \quad (25)$$

$$\theta_2: [\theta_2^{\text{mid}} - L_2, \theta_2^{\text{mid}} + L_2] = [-0.08227, 0.2242] \quad (26)$$

Exhaustive search with four significant digits in these solution intervals yields the optimal least-squares estimators for parameters θ_1 and θ_2 as $\theta_1 = 212.7$ and $\theta_2 = 0.06412$, which is the same result as previous studies produce [9,16]. The coefficient of determination, R^2 , for this result is 0.7206. The residual sum of squares, RSS, is 1195.

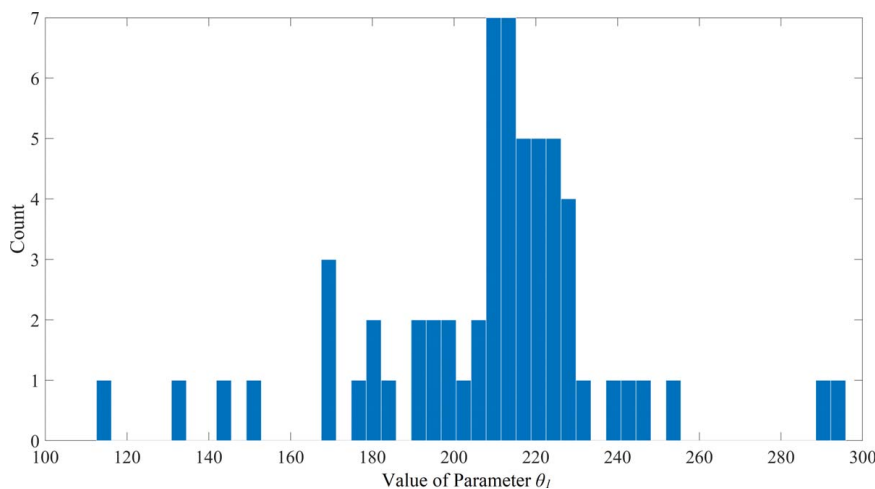


Fig. 3 Solution distribution of parameter θ_1 in the Michaelis–Menten model

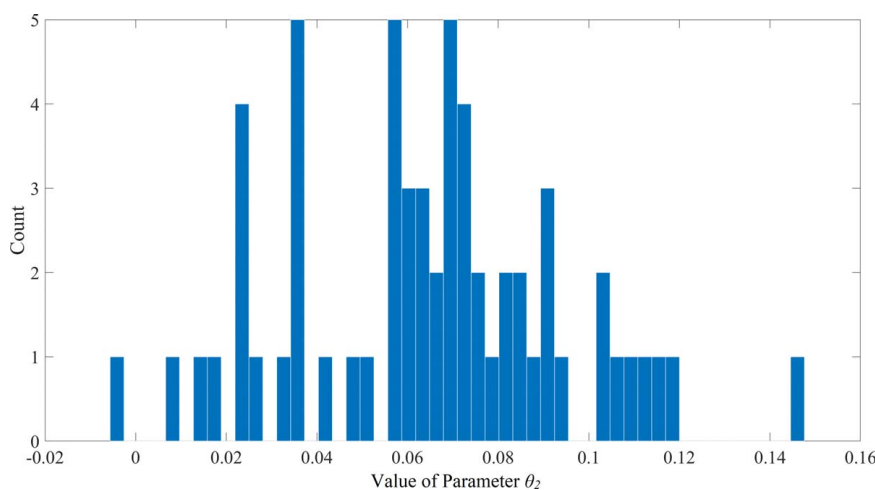


Fig. 4 Solution distribution of parameter θ_2 in the Michaelis–Menten model

Although exhaustive search in solution intervals is feasible in this case study, it is more efficient to find the optimal least-squares parameter estimators by using the median of the 60 solutions of each parameter. It is not necessary to employ the stochastic algorithm shown in Fig. 1 in this case study because Eqs. (19)–(20) have analytic solutions, and it is not expensive to solve Eqs. (19)–(20) 60 times. The median of the 60 solutions for parameter θ_1 , θ_1^{med} , is 213.7, and the median of the 60 solutions for parameter θ_2 , θ_2^{med} , is 0.06693. Of note, these medians are close to the optimal least-squares estimators for parameters θ_1 and θ_2 derived by exhaustive search. These two medians are supplied as the initial values of the Levenberg–Marquardt method, and the same optimal parameter estimators (i.e., $\theta_1 = 212.7$ and $\theta_2 = 0.06412$) are derived by the Levenberg–Marquardt method.

The curve fitting accuracy (i.e., R^2 and RSS) of the solution interval method could be compared with the linear transformation method. Equation (17) can be linearized as [9]

$$\frac{1}{y} = \frac{x + \theta_2}{\theta_1 x} = \frac{1}{\theta_1} + \frac{\theta_2}{\theta_1 x} = \beta_0 + \beta_1 u \quad (27)$$

The following linear model

$$y^* = \beta_0 + \beta_1 u \quad (28)$$

is fit to the experimental data set, where $y^* = 1/y$ and $u = 1/x$. The result of linear regression is $\beta_0 = 0.005107$ and $\beta_1 = 0.0002472$, so $\theta_1 = 195.8$ and $\theta_2 = 0.04841$. Figure 5 shows the fitted curves

with the original data using the solution interval method and the linear transformation method. The original data points have been distorted by the linear transformation shown in Eq. (27), so the linear model, Eq. (28), does not fit the original data well when the substrate has high concentration. The curve fitting accuracy of the linear transformation method is lower than that of the solution interval method. The coefficient of determination, R^2 , for the fit using the linear transformation method is 0.5513, and the residual sum of squares, RSS, is 1920.

The effectiveness of the solution interval method also could be compared with extant iterative least-squares local methods. Here, the Levenberg–Marquardt method is employed to fit the Michaelis–Menten model to the experimental data set, and 1000 random values are chosen as initial guesses for parameters θ_1 and θ_2 , respectively. The results show that only three out of the 1000 random initial guesses lead the Levenberg–Marquardt iterative algorithm to converge to the optimal parameter estimators (i.e., $\theta_1 = 212.7$ and $\theta_2 = 0.06412$). In other words, practitioners need to run the Levenberg–Marquardt iterative algorithm with a random initial guess 333 times on average to find the optimal parameter estimators. In practice, practitioners may run the Levenberg–Marquardt iterative algorithm with a random initial guess much more than 333 times to check whether the best available parameter estimators are the optimal estimators. The curve fitting results using the Levenberg–Marquardt method with six different initial guesses are shown in Table 2. These results in Table 2 indicate that the estimators of parameters θ_1 and θ_2 derived through the Levenberg–Marquardt

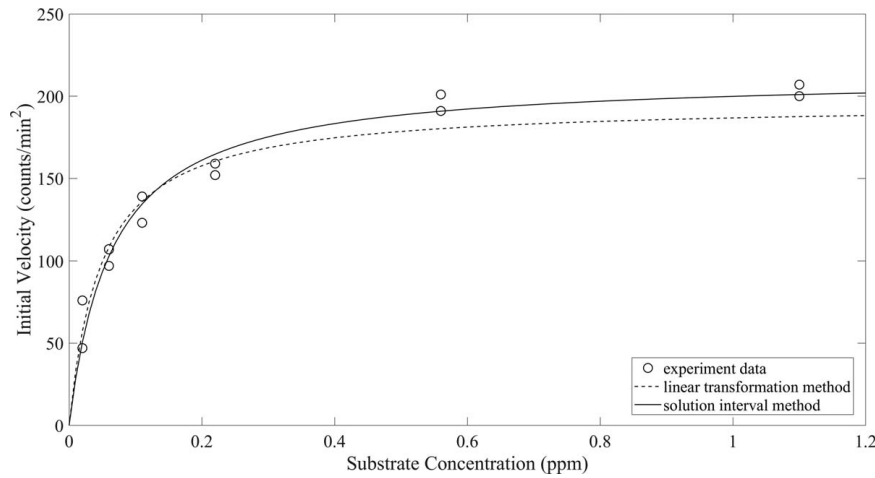


Fig. 5 Fitted curves using the solution interval method and the linear transformation method in the Michaelis–Menten model case study

Table 2 Curve fitting results using the Levenberg–Marquardt method in the Michaelis–Menten model case study

No.	Initial guess		Solution		RSS	R^2
	θ_{10}	θ_{20}	θ_1	θ_2		
1	1.0	1.0	-3.884×10^8	-1.583×10^6	80272	-17.75
2	1.0	0.1	25.98	-0.4894	181002	-41.30
3	10	0.1	68.16	-0.03782	130086	-29.40
4	10	1.0	68.15	-0.03782	130086	-29.40
5	100	0.1	212.7	0.06412	1195	0.7206
6	213.7	0.06693	212.7	0.06412	1195	0.7206

method are sensitive to the initial guess. The Levenberg–Marquardt iterative algorithm converges to the optimal parameter estimators (i.e., $\theta_1 = 212.7$ and $\theta_2 = 0.06412$) only when the initial guesses of the parameters are close to their optimal estimators. In contrast, the solution interval method is less computationally expensive than the Levenberg–Marquardt method with random initial guesses in this case study. Practitioners only need to solve the linear Eqs. (19)–(20) 60 times and take the median of the 60 solutions using the solution interval method. As stated earlier, the optimal parameter estimators can be found when the median of parameter solutions provided by the solution interval method is used as the initial value for each parameter.

5.2 Fitting Fick’s Second Law to Experiment Data. Fick’s laws govern the transportation of mass through diffusion [5,46]. One-dimensional moisture absorption of solid materials is governed by Fick’s second law as [46]

$$\frac{\partial c}{\partial t} = D \frac{\partial^2 c}{\partial x^2} \quad (29)$$

$$c = c_i \quad 0 < x < h; \quad t \leq 0 \quad (30)$$

$$c = c_m \quad x = 0; \quad x = h; \quad t > 0 \quad (31)$$

where a plate of thickness h exposes on two sides to the same moist environment. The plate is assumed to be infinite in the y and z directions so that the moisture content inside the plate varies only in the x -direction. The temperature in the moist environment and the temperature inside the plate are the same and do not change with time. In Eqs. (29)–(31), c is the moisture concentration; c_i is the initial moisture concentration; c_m is the moisture concentration at the

surface; and D is the mass diffusivity. The solution of Eqs. (29)–(31) is given by [46,47]

$$\frac{c - c_i}{c_m - c_i} = 1 - \frac{4}{\pi} \sum_{j=0}^{\infty} \frac{1}{(2j+1)} \sin \frac{(2j+1)\pi x}{h} \exp \left[-\frac{(2j+1)^2 \pi^2 D t}{h^2} \right] \quad (32)$$

The total weight of moisture in the plate is given by the integration of moisture concentration c over the plate thickness

$$m = \int_0^h c dx \quad (33)$$

The integration of Eq. (32) is [46,47]²

$$\frac{m - m_i}{m_m - m_i} = 1 - \frac{8}{\pi^2} \sum_{j=0}^{\infty} \frac{\exp \left[-(2j+1)^2 \pi^2 \left(\frac{D t}{h^2} \right) \right]}{(2j+1)^2} \quad (34)$$

where m_i is the initial weight of the moisture in the plate ($m_i = c_i h$), and m_m is the weight of moisture in the fully saturated plate ($m_m = c_m h$).

Here, Fick’s second law is fit to an experimental data set for percent mass change of an epoxy plate immersed in deionized water at room temperature [48] in order to estimate the mass diffusivity of the epoxy plate. A standard test method [6,49] has been followed to collect experimental data. A square epoxy plate is dried in an oven before the test. The oven-dry specimen mass is recorded as

²There is a typo in Ref. [47] for this equation. The summation should begin with $j = 0$ rather than $j = 1$.

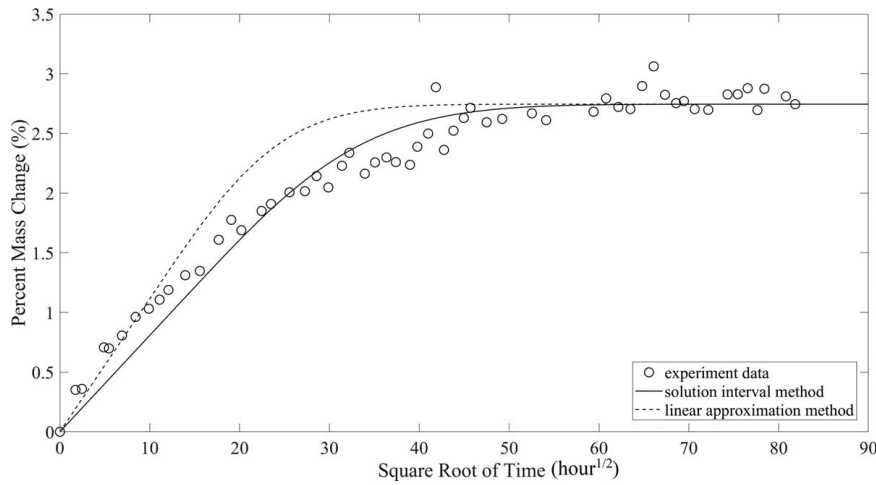


Fig. 6 Fitted curves using the solution interval method and the linear approximation method in Fick's second law case study

the baseline mass, W_b , for moisture absorption. The specimen is then placed in a conditioning chamber, which has previously reached a specified steady-state moist environment. At the beginning of the test ($t=0$), moisture concentration, c_i , and the total weight of moisture, m_i , are equal to zero because the specimen is dry. The specimen is removed from the conditioning chamber and weighted at the end of each specified weighting time interval. The weight of the specimen in each measurement is recorded as W_t . The percent specimen mass change ΔM is given by

$$\Delta M = \frac{W_t - W_b}{W_b} \times 100\% = \frac{m}{W_b} \times 100\% \quad (35)$$

The test stops when the specimen is fully saturated, and the specimen weight is W_f . Effective moisture equilibrium content, M_m , is

$$M_m = \frac{W_f - W_b}{W_b} \times 100\% = \frac{m_m}{W_b} \times 100\% \quad (36)$$

Equation (34) is converted to

$$\frac{\Delta M}{M_m} = \frac{m}{m_m} = 1 - \frac{8}{\pi^2} \sum_{j=0}^{\infty} \frac{\exp\left[-(2j+1)^2 \pi^2 \left(\frac{Dt}{h^2}\right)\right]}{(2j+1)^2} \quad (37)$$

Equation (37) is fit to the experimental data set of percent mass change, ΔM , of the epoxy plate. There are 57 data points (i.e., $m=57$), as shown in Fig. 6, and the last data point is employed as effective moisture equilibrium content, M_m [6].

The solution interval method presented in Secs. 3 and 4 is employed to find the optimal estimator for the parameter in Eq. (37). Mass diffusivity, D , is the only parameter in Eq. (37) (i.e., $n=1$). There are h combinations to choose one data point ($t, \Delta M$) from the 57 data points, where

$$h = C_m^n = C_{57}^1 = \frac{57!}{(57-1)!1!} = 57 \quad (38)$$

The chosen data point is plugged into Eq. (37) to solve the mass diffusivity, D . Analytic solution is not available in this case. Brent's method [50], which is a combination of the bisection method, the secant method, and the inverse quadratic interpolation, is employed to solve mass diffusivity, D , in Eq. (37) numerically. A wide range [0, 1] is set as the initial interval for Brent's method since it is assumed that prior knowledge about the range of the mass diffusivity of the epoxy plate is not available. The solution distribution of mass diffusivity, D (mm^2/s), solving by Brent's method appears in Fig. 7, where $2.411 \times 10^{-8} \leq D \leq 4.045 \times 10^{-7}$. Of note, only 43 solutions are obtained from solving Eq. (37) numerically for

the 57 combinations because the last data point is employed as M_m [6], and Eq. (37) has no solution for mass diffusivity, D , when percent mass change, ΔM , is greater than or equal to M_m . The solution interval for mass diffusivity is derived based on Eqs. (15)–(16) as

$$D: [D^{\text{mid}} - L, D^{\text{mid}} + L] = [-1.661 \times 10^{-7}, 5.947 \times 10^{-7}] \quad (39)$$

The left bound of the solution interval could be truncated at $D=0$ because mass diffusivity is always non-negative. Exhaustive search with four significant digits in this truncated solution interval (i.e., $0 \leq D \leq 5.947 \times 10^{-7}$) yields the optimal least-squares estimator for the mass diffusivity as $D=6.414 \times 10^{-8} \text{ mm}^2/\text{s}$, which is the same result as exhaustive search in the wide range [0, 1] (i.e., $0 \leq D \leq 1$), as shown in Fig. 8. The coefficient of determination, R^2 , for this result is 0.9555. The residual sum of squares, RSS, is 1.511×10^{-4} .

It is expensive to perform an exhaustive search in the solution interval in this case study because Eq. (37) includes the summation of infinite terms. The median of the 43 solutions, $D^{\text{med}}=5.662 \times 10^{-8} \text{ mm}^2/\text{s}$, could be supplied as the initial value of the Levenberg–Marquardt method, and the same optimal mass diffusivity estimator (i.e., $D=6.414 \times 10^{-8} \text{ mm}^2/\text{s}$) is derived by the Levenberg–Marquardt method. Like the case study using the Michaelis–Menten model in Sec. 5.1, the median is close to the optimal estimator for the mass diffusivity, D , derived by exhaustive search in this case study.

The stochastic algorithm shown in Fig. 1 could be employed to find the optimal estimator for the mass diffusivity in Eq. (37) more efficiently because the stochastic algorithm usually does not require solving Eq. (37) numerically 57 times. The Levenberg–Marquardt method is used as the iterative least-squares local method in the stochastic algorithm. In each iteration step, a new combination is chosen without replacement and plugged into Eq. (37) to solve D if Eq. (37) has no solution for the previous combination. To test the convergence of the stochastic algorithm, the stochastic algorithm shown in Fig. 1 has been performed 100 times, and the results show that the stochastic algorithm always converges to the optimal mass diffusivity estimator (i.e., $D=6.414 \times 10^{-8} \text{ mm}^2/\text{s}$). Specifically, the algorithm converges within one iteration step (i.e., using three solutions) 86 times, converges within two iteration steps (i.e., using five solutions) 98 times, and converges within five iteration steps (i.e., using 11 solutions) 100 times. The solution interval derived by five iteration steps using 11 solutions covers

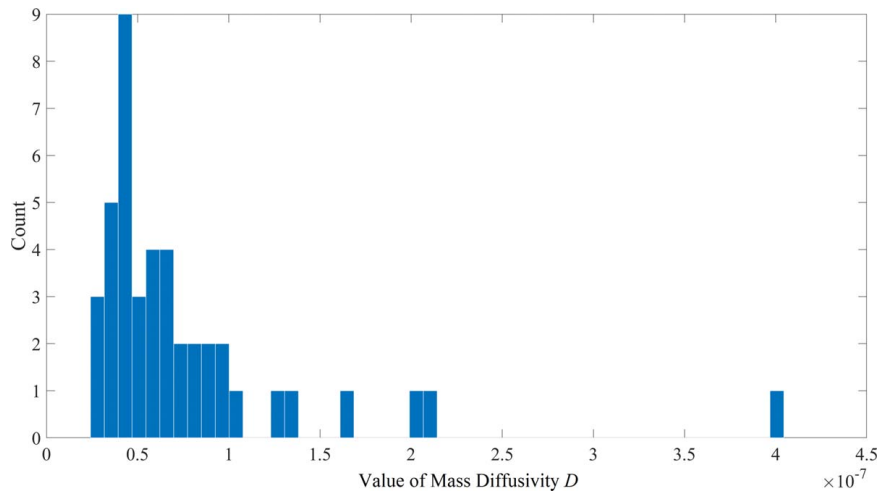


Fig. 7 Solution distribution of mass diffusivity D in Fick's second law

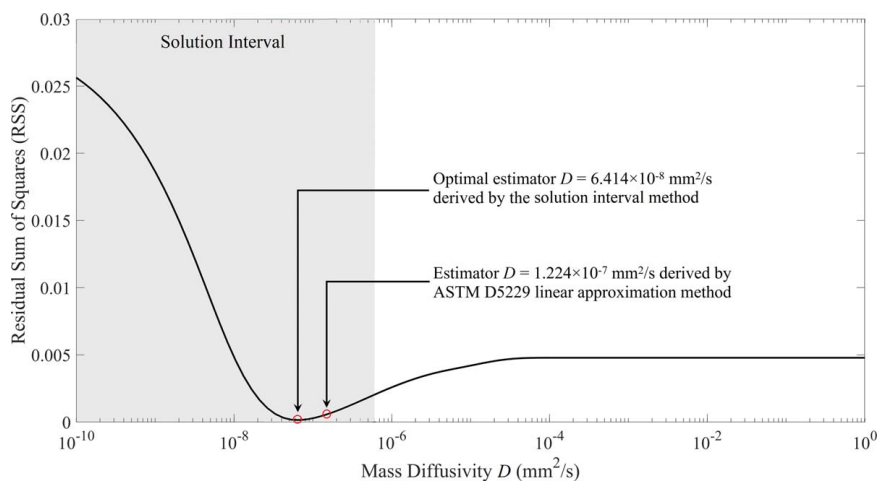


Fig. 8 Residual sum of squares of the fit with different mass diffusivity values in Fick's second law case study

51.36% of the full solution interval using all 43 solutions (i.e., $-1.661 \times 10^{-7} \leq D \leq 5.947 \times 10^{-7}$) on average in this case study.

The curve fitting accuracy (i.e., R^2 and RSS) of the solution interval method could be compared with the linear approximation method suggested by ASTM D5229 [6]. Equation (37) approximates a linear relationship between the percent specimen mass change, ΔM , and a square root of time, \sqrt{t} , for a short time after the specimen is exposed to the moist environment [47]. In ASTM D5229, the mass diffusivity of the specimen is calculated through [6,47]

$$D = \pi \left(\frac{h}{4M_m} \right)^2 \left(\frac{M_2 - M_1}{\sqrt{t_2} - \sqrt{t_1}} \right)^2 \quad (40)$$

where h is the specimen thickness, M_m is the effective moisture equilibrium content defined by Eq. (36), $(M_2 - M_1)/(\sqrt{t_2} - \sqrt{t_1})$ is the slope of moisture absorption plot in the initial linear portion of the $\sqrt{t} - \Delta M$ curve. In practice, it is often difficult to identify the endpoint of the initial linear portion of the $\sqrt{t} - \Delta M$ curve visually. Researchers who published the experimental data shown in Fig. 6 take the first nine data points as the linear portion of $\sqrt{t} - \Delta M$ curve, which is a reasonable visual pick from the curve. The last data point is employed as effective moisture equilibrium content, M_m [6]. The mass diffusivity derived through Eq. (40) is $D = 1.224 \times 10^{-7} \text{ mm}^2/\text{s}$, which is 1.91 times of the optimal mass diffusivity estimator $D = 6.414 \times 10^{-8} \text{ mm}^2/\text{s}$. Figure 6 shows the fitted curves using the solution interval method and the linear approximation method suggested by ASTM

D5229. The linear approximation method leads to a poor fit in the middle portion of the curve (i.e., $15 < \sqrt{t} < 45$) compared with the solution interval method because the linear approximation method only takes the first nine data points into account. In addition, the value of the mass diffusivity derived through Eq. (40) varies significantly if a different data point is chosen as the endpoint of the initial linear portion of the $\sqrt{t} - \Delta M$ curve. The mass diffusivity results given by Eq. (40) using seven different endpoints appear in Table 3. These results show that the mass diffusivity derived through Eq. (40) could vary more than 60% using different endpoints to make linear approximations for the initial portion of the $\sqrt{t} - \Delta M$ curve. The R^2 and RSS results in Table 3 also show that the curve fitting accuracy of the linear approximation method is lower than that of the solution interval method (i.e., $R^2 = 0.9555$ and $\text{RSS} = 1.511 \times 10^{-4}$).

The effectiveness of the solution interval method also could be compared with extant iterative least-squares local methods. As an example, the Levenberg–Marquardt method is employed to fit Eq. (37) to the experimental data set, and 1000 random values are chosen as initial guesses for mass diffusivity, D . The last data point is employed as effective moisture equilibrium content, M_m [6]. The results show that none of the 1000 random initial guesses lead the Levenberg–Marquardt iterative algorithm to converge to the optimal parameter estimators (i.e., $D = 6.414 \times 10^{-8} \text{ mm}^2/\text{s}$). The curve fitting results using the Levenberg–Marquardt method with nine different initial guesses are shown in Table 4. The Levenberg–Marquardt iterative algorithm converges to the

Table 3 Mass diffusivity results calculated through Eq. (40) using different endpoints

No.	Linear portion of the curve	Slope of the linear portion	Mass diffusivity (mm ² /s)	RSS	R ²
1	First 6 data points	0.1298	1.655×10^{-7}	6.596×10^{-4}	0.8056
2	First 7 data points	0.1237	1.503×10^{-7}	5.764×10^{-4}	0.8301
3	First 8 data points	0.1168	1.340×10^{-7}	4.837×10^{-4}	0.8575
4	First 9 data points	0.1116	1.224×10^{-7}	4.158×10^{-4}	0.8775
5	First 10 data points	0.1081	1.148×10^{-7}	3.717×10^{-4}	0.8905
6	First 11 data points	0.1044	1.071×10^{-7}	3.270×10^{-4}	0.9036
7	First 12 data points	0.1000	9.824×10^{-8}	2.773×10^{-4}	0.9183

optimal mass diffusivity estimator (i.e., $D = 6.414 \times 10^{-8}$ mm²/s) only when the initial guess is close to the optimal estimator. As shown in Fig. 8, the curve of RSS is flat when the value of the mass diffusivity is larger than 10^{-4} mm²/s, and the Levenberg–Marquardt iterative algorithm therefore cannot converge to the optimal mass diffusivity estimator if the initial guess is chosen in that flat region. In contrast, the solution interval method establishes a solution interval for the mass diffusivity, shown as the gray area in Fig. 8. The solution interval significantly reduces the search space for the optimal estimator of the mass diffusivity in this case study. As stated earlier, any of the three algorithms associated with the solution interval method (i.e., exhaustive search in solution interval, using median of solutions as the initial value, and the stochastic algorithm shown in Fig. 1) can find the optimal estimator of the mass diffusivity more effectively in this case study.

6 Conclusions and Future Work

A solution interval method is introduced for least-squares parameter estimation in nonlinear models. The novelty of this method is that practitioners do not need to guess the value for each parameter in a nonlinear model in order to initialize the method. The solution interval method includes three algorithms (i.e., blue blocks in Fig. 2) that require different levels of computational power to find the optimal parameter estimators. The method provides the empirical probability distribution of each parameter in a nonlinear model. A solution interval for each parameter in a nonlinear model is also constructed by the method. It is proved that the optimal estimator for each parameter in a nonlinear model that minimizes the squared error of the fit belongs to its corresponding solution interval under certain conditions. Four steps to implement the solution interval method in practice are outlined. The Michaelis–Menten model and Fick’s second law are fit to experimental data sets respectively as two case studies to illustrate these steps and validate the method.

Based on available computational power, practitioners can find the optimal least-squares parameter estimators in a specified nonlinear model efficiently using one of the three algorithms associated with the solution interval method. The solution intervals constructed by the method significantly reduce the search space for the optimal parameter estimators. The empirical probability distribution of each parameter is valuable for researchers to assess parameter uncertainty that pertains to curve fitting.

This work has limitations that provide opportunities for future research. As stated in Sec. 3, parameter solution equations represented by Eq. (14) are derived when the chosen data points are plugged into the nonlinear model. In this paper, it is assumed that these parameter solution equations only have one solution for each parameter. Future research may develop methodologies to find the optimal parameter estimators when these equations have more than one solution for each parameter. Moreover, the solution interval method presented in this paper provides an empirical probability distribution for each parameter in the nonlinear model. Future research could explore how to utilize these probability distributions. To assess parameter uncertainty, these probability distributions may be employed as prior distributions of a Bayesian

Table 4 Curve fitting results using the Levenberg–Marquardt method in Fick’s second law case study

No.	Initial guess D_0	Solution D	RSS	R ²
1	1.0	1.0	4.789×10^{-3}	-0.4114
2	0.1	0.1	4.789×10^{-3}	-0.4114
3	0.01	0.01	4.789×10^{-3}	-0.4114
4	1×10^{-3}	1×10^{-3}	4.789×10^{-3}	-0.4114
5	1×10^{-4}	Non-convergence	N/A	N/A
6	1×10^{-5}	Non-convergence	N/A	N/A
7	1×10^{-6}	Non-convergence	N/A	N/A
8	1×10^{-7}	6.414×10^{-8}	1.511×10^{-4}	0.9555
9	5.662×10^{-8}	6.414×10^{-8}	1.511×10^{-4}	0.9555

approach. These probability distributions may be used to identify potential outliers in a data set as well. For example, if a data point leads to its corresponding parameter solution equation(s) which have no solution or the solution locates at the margin of the probability distribution for each parameter, practitioners need to check whether that data point is an outlier. Practitioners may also allocate the computational resource in an exhaustive search (i.e., the first algorithm in Sec. 3) based on the empirical probability distribution of each parameter in the nonlinear model when the solution interval is extremely large. For example, a finer grid may be applied in the high-density regions of the probability distribution for each parameter. In addition, the solution interval of each parameter is defined as $[\theta_j^{\text{mid}} - L_j, \theta_j^{\text{mid}} + L_j]$ rather than $[\theta_j^{\text{min}}, \theta_j^{\text{max}}]$ in this paper because it is proven in Appendix A that the optimal estimator for each parameter θ_j that minimizes the squared error of the fit belongs to the interval $[\theta_j^{\text{mid}} - L_j, \theta_j^{\text{mid}} + L_j]$ when conditions given by Eqs. (A8)–(A10) are satisfied. However, two case studies in Sec. 5 show that the optimal parameter estimator for each parameter belongs to $[\theta_j^{\text{min}}, \theta_j^{\text{max}}]$. It may not be necessary to extend the solution interval from $[\theta_j^{\text{min}}, \theta_j^{\text{max}}]$ to $[\theta_j^{\text{mid}} - L_j, \theta_j^{\text{mid}} + L_j]$. Future research may relax the conditions given by Eqs. (A8)–(A10), which are stronger than convexity for the least-squares error $R(\hat{\theta})$, and define a solution interval for the more general case or develop a procedure to derive the solution interval on a case-by-case basis. Lastly, both the number of unknown parameters in the nonlinear models and the number of data points are small (i.e., $n < 3$ and $m < 60$) in the Michaelis–Menten model case study and Fick’s second law case study in Sec. 5. The solution interval method presented in this paper may be applied to high-dimensional (i.e., large n) and big data (i.e., large m) engineering problems in future research.

Acknowledgment

This material is based upon a work supported by the Defense Advanced Research Projects Agency through cooperative agreement No. N66001-17-1-4064. 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.

Appendix A: Mathematical Proof of the Solution Interval

DEFINITION. As discussed in Sec. 3, a generic nonlinear model $y=f(\mathbf{x}, \boldsymbol{\theta})$ is fit to m data points $[x_{1(i)}, x_{2(i)}, \dots, x_{q(i)}, y_{(i)}]$, where $i \in \{1, 2, \dots, m\}$, $\mathbf{x}=[x_1, x_2, \dots, x_q]$, and $\boldsymbol{\theta}=[\theta_1, \theta_2, \dots, \theta_n]$ represents n constant parameters ($n < m$) in the nonlinear model. The least-squares error of the fit is

$$S(\hat{\boldsymbol{\theta}}) = \sum_{i=1}^m [y_{(i)} - f(x_{1(i)}, x_{2(i)}, \dots, x_{q(i)}, \hat{\boldsymbol{\theta}})]^2 \quad (\text{A1})$$

where $\hat{\boldsymbol{\theta}} = [\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_n]$ is the estimated values of parameters $\boldsymbol{\theta}$.

There are h combinations for choosing n data points $[x_{1(i_k)}, x_{2(i_k)}, \dots, x_{q(i_k)}, y_{(i_k)}]$ from the m data points $[x_{1(i)}, x_{2(i)}, \dots, x_{q(i)}, y_{(i)}]$, where $k \in \{1, 2, \dots, h\}$, $i_k \in \{1, 2, \dots, m\}$, $i \in \{1, 2, \dots, m\}$, and

$$h = C_m^n = \frac{m!}{(m-n)!n!} \quad (\text{A2})$$

If any n chosen data points are plugged into the nonlinear model $y=f(\mathbf{x}, \boldsymbol{\theta})$, n equations are derived as

$$y_{(i_k)} - f(x_{1(i_k)}, x_{2(i_k)}, \dots, x_{q(i_k)}, \boldsymbol{\theta}) = 0 \quad (\text{A3})$$

If the n equations represented by Eq. (A3) only have one solution for one combination as $\boldsymbol{\theta}_{(s)} = [\theta_{1(s)}, \theta_{2(s)}, \dots, \theta_{n(s)}]$, where $s \in \{1, 2, \dots, h\}$, there are h solutions for each parameter θ_j in total, where $j \in \{1, 2, \dots, n\}$. For each parameter θ_j , these h solutions are sorted from the smallest solution θ_j^{\min} to the largest solution θ_j^{\max} , where $\theta_j^{\min} \leq \theta_{j(s)} \leq \theta_j^{\max}$. The solution interval of parameter θ_j is defined as $[\theta_j^{\min} - L_j, \theta_j^{\max} + L_j]$, where

$$L_j = \theta_j^{\max} - \theta_j^{\min} \quad (\text{A4})$$

$$\theta_j^{\text{mid}} = \frac{\theta_j^{\min} + \theta_j^{\max}}{2} \quad (\text{A5})$$

For any n data points $[x_{1(i_k)}, x_{2(i_k)}, \dots, x_{q(i_k)}, y_{(i_k)}]$ that are chosen from the m data points $[x_{1(i)}, x_{2(i)}, \dots, x_{q(i)}, y_{(i)}]$, the least-squares error of these n data points is

$$R(\hat{\boldsymbol{\theta}}) = \sum_{k=1}^n [y_{(i_k)} - f(x_{1(i_k)}, x_{2(i_k)}, \dots, x_{q(i_k)}, \hat{\boldsymbol{\theta}})]^2 \quad (\text{A6})$$

These n chosen data points $[x_{1(i_k)}, x_{2(i_k)}, \dots, x_{q(i_k)}, y_{(i_k)}]$ are plugged into Eq. (A3), and the solution of Eq. (A3) is $\boldsymbol{\theta}_{(s)}$. If $\boldsymbol{\theta}_{(s)}$ is taken as the parameter estimators $\hat{\boldsymbol{\theta}}$ for these n data points, the least-squares error is

$$\begin{aligned} R(\boldsymbol{\theta}_{(s)}) &= R(\theta_{1(s)}, \dots, \theta_{n(s)}) \\ &= \sum_{k=1}^n [y_{(i_k)} - f(x_{1(i_k)}, x_{2(i_k)}, \dots, x_{q(i_k)}, \boldsymbol{\theta}_{(s)})]^2 = 0 \end{aligned} \quad (\text{A7})$$

LEMMA. If the least-squares error $R(\hat{\boldsymbol{\theta}})$ is symmetric with respect to $\hat{\boldsymbol{\theta}} = \boldsymbol{\theta}_{(s)}$ as

$$R(\theta_{1(s)}, \dots, \theta_{j(s)} + d_j, \dots, \theta_{n(s)}) = R(\theta_{1(s)}, \dots, \theta_{j(s)} - d_j, \dots, \theta_{n(s)}) \quad (\text{A8})$$

and the least-squares error $R(\hat{\boldsymbol{\theta}})$ monotonically increase when $\hat{\boldsymbol{\theta}}$ departs from $\boldsymbol{\theta}^*$ as

$$\begin{aligned} R(\theta_{1(s)} + d_1, \dots, \theta_{j(s)} + d_j, \dots, \theta_{n(s)} + d_n) \\ < R(\theta_{1(s)} + d_1, \dots, \theta_{j(s)} + d_j + \varepsilon, \dots, \theta_{n(s)} + d_n) \end{aligned} \quad (\text{A9})$$

$$\begin{aligned} R(\theta_{1(s)} - d_1, \dots, \theta_{j(s)} - d_j, \dots, \theta_{n(s)} - d_n) \\ < R(\theta_{1(s)} - d_1, \dots, \theta_{j(s)} - d_j - \varepsilon, \dots, \theta_{n(s)} - d_n) \end{aligned} \quad (\text{A10})$$

where d_j is an arbitrary positive number, and ε is a small positive number, then the optimal parameter estimators that minimize the squared error S for all m data points belong to solution interval $[\theta_j^{\text{mid}} - L_j, \theta_j^{\text{mid}} + L_j]$.

Proof. If the midpoint of solution interval $[\theta_j^{\text{mid}} - L_j, \theta_j^{\text{mid}} + L_j]$, θ_j^{mid} , is taken as the parameter estimator, the least-squares error for all m data points is

$$S(\boldsymbol{\theta}^{\text{mid}}) = S(\theta_1^{\text{mid}}, \dots, \theta_n^{\text{mid}}) = \sum_{i=1}^m [y_{(i)} - f(x_{1(i)}, \dots, x_{q(i)}, \boldsymbol{\theta}^{\text{mid}})]^2 \quad (\text{A11})$$

n times the least-squares error $S(\boldsymbol{\theta}^{\text{mid}})$ as

$$\begin{aligned} nS(\boldsymbol{\theta}^{\text{mid}}) &= n \sum_{i=1}^m [y_{(i)} - f(x_{1(i)}, \dots, x_{q(i)}, \boldsymbol{\theta}^{\text{mid}})]^2 \\ &= [y_{(1)} - f(x_{1(1)}, \dots, x_{q(1)}, \boldsymbol{\theta}^{\text{mid}})]^2 + \dots \end{aligned} \quad (\text{A12})$$

Equation (A12) could be expanded as the summation of $n \times m$ terms in sequence. The summation of the first n terms is defined as $R_1(\boldsymbol{\theta}^{\text{mid}})$, the summation of the second n terms is defined as $R_2(\boldsymbol{\theta}^{\text{mid}})$, and so on. Equation (A12) is converted to

$$nS(\boldsymbol{\theta}^{\text{mid}}) = n \sum_{i=1}^m [y_{(i)} - f(x_{1(i)}, \dots, x_{q(i)}, \boldsymbol{\theta}^{\text{mid}})]^2 = \sum_{i=1}^m R_i(\boldsymbol{\theta}^{\text{mid}}) \quad (\text{A13})$$

If θ_j^{out} is chosen as the parameter estimator, and θ_j^{out} does not belong to the solution interval $[\theta_j^{\text{mid}} - L_j, \theta_j^{\text{mid}} + L_j]$, thus $\theta_j^{\text{out}} < \theta_j^{\text{mid}} - L_j$ or $\theta_j^{\text{out}} > \theta_j^{\text{mid}} + L_j$. The least-squares error for all m data points is

$$S(\boldsymbol{\theta}^{\text{out}}) = S(\theta_1^{\text{out}}, \dots, \theta_n^{\text{out}}) = \sum_{i=1}^m [y_{(i)} - f(x_{1(i)}, \dots, x_{q(i)}, \boldsymbol{\theta}^{\text{out}})]^2 \quad (\text{A14})$$

n times the least-squares error $S(\boldsymbol{\theta}^{\text{out}})$ as

$$\begin{aligned} nS(\boldsymbol{\theta}^{\text{out}}) &= n \sum_{i=1}^m [y_{(i)} - f(x_{1(i)}, \dots, x_{q(i)}, \boldsymbol{\theta}^{\text{out}})]^2 \\ &= [y_{(1)} - f(x_{1(1)}, \dots, x_{q(1)}, \boldsymbol{\theta}^{\text{out}})]^2 + \dots \end{aligned} \quad (\text{A15})$$

Equation (A15) also could be expanded as the summation of $n \times m$ terms in sequence. The summation of the first n terms is defined as $R_1(\boldsymbol{\theta}^{\text{out}})$, the summation of the second n terms is defined as $R_2(\boldsymbol{\theta}^{\text{out}})$, and so on. Equation (A15) is converted to

$$nS(\boldsymbol{\theta}^{\text{out}}) = n \sum_{i=1}^m [y_{(i)} - f(x_{1(i)}, \dots, x_{q(i)}, \boldsymbol{\theta}^{\text{out}})]^2 = \sum_{i=1}^m R_i(\boldsymbol{\theta}^{\text{out}}) \quad (\text{A16})$$

The same n data points are used to calculate $R_i(\boldsymbol{\theta}^{\text{mid}})$ in Eq. (A13) and $R_i(\boldsymbol{\theta}^{\text{out}})$ in Eq. (A16), where $i \in \{1, 2, \dots, m\}$. These n data points could be plugged into Eq. (A3), and the solution of Eq. (A3) is $\boldsymbol{\theta}_{(s)}$. The least-squares error for these n data points, $R_i(\boldsymbol{\theta}_{(s)})$, equals to zero as defined in Eq. (A7). For each parameter θ_j , the solution $\theta_{j(s)}$ belongs to the interval $[\theta_j^{\min}, \theta_j^{\max}]$. θ_j^{mid} is defined in Eq. (A5) as the midpoint of the interval $[\theta_j^{\min}, \theta_j^{\max}]$. It is also known that θ_j^{out} does not belong to the solution interval $[\theta_j^{\text{mid}} - L_j, \theta_j^{\text{mid}} + L_j]$. Thus,

$$|\theta_j^{\text{mid}} - \theta_{j(s)}| \leq \frac{L_j}{2} < |\theta_j^{\text{out}} - \theta_{j(s)}| \quad (\text{A17})$$

for any $j \in \{1, 2, \dots, n\}$. Equation (A17) shows that the distance between θ_j^{out} and $\theta_{j(s)}$ is always greater than the distance between θ_j^{mid} and $\theta_{j(s)}$.

The least-squares error $R_i(\theta^{\text{mid}})$ is always smaller than $R_i(\theta^{\text{out}})$ as

$$R_i(\theta^{\text{mid}}) < R_i(\theta^{\text{out}}) \quad (\text{A18})$$

based on Eq. (A17) for any $i \in \{1, 2, \dots, m\}$ and the monotonic conditions and the symmetric condition defined by Eqs. (A8)–(A10). Of note, the symmetric condition represented by Eq. (A8) is necessary to derive Eq. (A18) because θ_j^{mid} and θ_j^{out} may lay on the different side of $\theta_{j(s)}$ (e.g., $\theta_j^{\text{mid}} < \theta_{j(s)} < \theta_j^{\text{out}}$). Thus,

$$nS(\theta^{\text{mid}}) < nS(\theta^{\text{out}}) \quad (\text{A19})$$

based on Eqs. (A13) and (A16), where n is a positive integer. Thus,

$$S(\theta^{\text{mid}}) < S(\theta^{\text{out}}) \quad (\text{A20})$$

Equation (A20) shows that the least-squares error for all m data points is always smaller when the midpoint of solution interval, θ_j^{mid} , is chosen as the estimator for each parameter compared to when any number, θ_j^{out} , which does not belong to solution interval $[\theta_j^{\text{mid}} - L_j, \theta_j^{\text{mid}} + L_j]$, is chosen as the estimator for each parameter. Thus, the optimal parameter estimators that minimize the squared error S for all m data points belong to solution interval $[\theta_j^{\text{mid}} - L_j, \theta_j^{\text{mid}} + L_j]$ if conditions Eqs. (A8)–(A10) are satisfied. ■

References

- [1] Bannantine, J., Comer, J., and Handrock, J., 1990, *Fundamentals of Metal Fatigue Analysis*, Prentice Hall, Englewood Cliffs, NJ.
- [2] Juvinall, R. C., and Marshek, K. M., 2000, *Fundamentals of Machine Component Design*, John Wiley & Sons, Inc, New York, NY.
- [3] Wirsching, P. H., 1983, *Statistical Summaries of Fatigue Data for Design Purposes*, NASA Contractor Report 3697, Washington, DC, <https://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/19830021460.pdf>
- [4] ASTM, 2015, *E739-10. Standard Practice for Statistical Analysis of Linear or Linearized Stress-Life (S-N) and Strain-Life (ε-N) Fatigue Data*, ASTM International, West Conshohocken, PA, <https://www.astm.org/Standards/E739>
- [5] Crank, J., 1975, *The Mathematics of Diffusion*, Oxford University Press, Oxford.
- [6] 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/Standards/D5229.htm>
- [7] McKague, E., Reynolds, J., and Halkias, J., 1975, *Life Assurance of Composite Structures. Volume I. Moisture Effects*, AFML-TR-75-51, Fort Worth Division of General Dynamics, Wright-Patterson Air Force Base, Ohio., <https://apps.dtic.mil/dtic/tr/fulltext/u2/b006906.pdf>
- [8] Jennrich, R. I., and Ralston, M. L., 1979, "Fitting Nonlinear Models to Data," *Annu. Rev. Biophys. Bioeng.*, **8**(1), pp. 195–238.
- [9] Montgomery, D. C., Peck, E. A., and Vining, G. G., 2006, *Introduction to Linear Regression Analysis*, Wiley-Interscience, Hoboken, NJ.
- [10] Hauser, J. R., 2009, *Numerical Methods for Nonlinear Engineering Models*, Springer Science & Business Media, Berlin/Heidelberg, Germany.
- [11] Moré, J. J., 1978, "The Levenberg-Marquardt Algorithm: Implementation and Theory," *Numerical Analysis*, Springer, Berlin/Heidelberg, Germany, pp. 105–116, <https://link.springer.com/book/10.1007%2FBFb0067690#page=116>
- [12] Hansen, P. C., Pereyra, V., and Scherer, G., 2013, *Least Squares Data Fitting With Applications*, The Johns Hopkins University Press, Baltimore, ML.
- [13] Björck, Å., 1996, *Numerical Methods for Least Squares Problems*, Society for Industrial and Applied Mathematics, Philadelphia, PA.
- [14] 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.
- [15] 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.
- [16] Bates, D. M., and Watts, D. G., 1988, *Nonlinear Regression Analysis and Its Applications*, John Wiley & Sons, New York, <https://onlinelibrary.wiley.com/doi/book/10.1002/9780470316757>
- [17] Fraley, C., 1988, *Algorithms for Nonlinear Least-Squares Problems*, SOL 88-16, Systems Optimization Laboratory, Stanford, CA, <https://apps.dtic.mil/dtic/tr/fulltext/u2/a201848.pdf>
- [18] Schaller, R. R., 1997, "Moore's Law: Past, Present and Future," *IEEE Spectrum*, **34**(6), pp. 52–59.
- [19] Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P., 2007, *Numerical Recipes: The Art of Scientific Computing*, Cambridge University Press, Cambridge.
- [20] Ramsin, H., and Wedin, P.-Å., 1977, "A Comparison of Some Algorithms for the Nonlinear Least Squares Problem," *BIT Numer. Math.*, **17**(1), pp. 72–90.
- [21] Gill, P. E., and Murray, W., 1976, "Nonlinear Least Squares and Nonlinearly Constrained Optimization," *Numerical Analysis*, G. A. Watson, ed., Springer-Verlag Berlin, Germany, 134–147.
- [22] Dennis, J. E., Jr, and Schnabel, R. B., 1996, *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*, Society for Industrial and Applied Mathematics, Philadelphia, PA.
- [23] Gill, P. E., and Murray, W., 1978, "Algorithms for the Solution of the Nonlinear Least-Squares Problem," *SIAM J. Numer. Anal.*, **15**(5), pp. 977–992.
- [24] Dennis, J. E., Jr, Gay, D. M., and Welsch, R. E., 1981, "An Adaptive Nonlinear Least Square Algorithm," *ACM Trans. Math. Software*, **7**(3), pp. 348–368.
- [25] Levenberg, K., 1944, "A Method for the Solution of Certain Non-Linear Problems in Least Squares," *Quarterly Appl. Math.*, **2**(2), pp. 164–168.
- [26] Marquardt, D. W., 1963, "An Algorithm for Least-Squares Estimation of Nonlinear Parameters," *J. Soc. Ind. Appl. Math.*, **11**(2), pp. 431–441.
- [27] Moré, J. J., 1983, "Recent Developments in Algorithms and Software for Trust Region Methods," *Mathematical Programming The State of the Art*, Springer-Verlag, Berlin, Germany, 258–287, https://link.springer.com/chapter/10.1007/978-3-642-68874-4_11#citeas
- [28] Fletcher, R., 1987, *Practical Methods of Optimization*, John Wiley & Sons, Chichester.
- [29] 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.
- [30] Conn, A. R., Gould, N. I., and Toint, P. L., 2000, *Trust Region Methods*, Society for Industrial and Applied Mathematics and Mathematical Programming Society, Philadelphia, PA.
- [31] Yuan, Y.-x., 2015, "Recent Advances in Trust Region Algorithms," *Math. Program.*, **151**(1), pp. 249–281.
- [32] Sobieszcanski-Sobieski, J., Morris, A., and VanTooren, M., 2015, *Multidisciplinary Design Optimization Supported by Knowledge Based Engineering*, John Wiley & Sons, The Atrium, Southern Gate, Chichester.
- [33] Chang, W.-D., 2007, "Nonlinear System Identification and Control Using A Real-Coded Genetic Algorithm," *Appl. Math. Model.*, **31**(3), pp. 541–550.
- [34] Mohan, S., 1997, "Parameter Estimation of Nonlinear Muskingum Models Using Genetic Algorithm," *J. Hydraul. Eng.*, **123**(2), pp. 137–142.
- [35] Wong, K.-P., Meikle, S. R., Feng, D., and Fulham, M. J., 2002, "Estimation of Input Function and Kinetic Parameters Using Simulated Annealing: Application in a Flow Model," *IEEE Trans. Nucl. Sci.*, **49**(3), pp. 707–713.
- [36] Schwaab, M., Biscacia, E. C., Jr, Monteiro, J. L., and Pinto, J. C., 2008, "Nonlinear Parameter Estimation Through Particle Swarm Optimization," *Chem. Eng. Sci.*, **63**(6), pp. 1542–1552.
- [37] Tierney, L., and Kadane, J. B., 1986, "Accurate Approximations for Posterior Moments and Marginal Densities," *J. Am. Stat. Assoc.*, **81**(393), pp. 82–86.
- [38] Gilks, W. R., Richardson, S., and Spiegelhalter, D., 1996, *Markov Chain Monte Carlo in Practice*, Chapman and Hall/CRC, Boca Raton, FL.
- [39] Congdon, P., 2014, *Applied Bayesian Modelling*, John Wiley & Sons, The Atrium, Southern Gate, Chichester.
- [40] 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.
- [41] Archontoulis, S. V., and Miguez, F. E., 2015, "Nonlinear Regression Models and Applications in Agricultural Research," *Agron. J.*, **107**(2), pp. 786–798.
- [42] Garthwaite, P. H., Kadane, J. B., and O'Hagan, A., 2005, "Statistical Methods for Eliciting Probability Distributions," *J. Am. Stat. Assoc.*, **100**(470), pp. 680–701.
- [43] 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.
- [44] Voet, D., and Voet, J. G., 2010, *Biochemistry*, John Wiley & Sons, Hoboken, NJ.
- [45] Atkins, G., and Nimmo, I., 1975, "A Comparison of Seven Methods for Fitting the Michaelis-Menten Equation," *Biochem. J.*, **149**(3), pp. 775–777.
- [46] Jost, W., 1952, *Diffusion in Solids, Liquids, Gases*, Academic Press, New York.
- [47] Shen, C.-H., and Springer, G. S., 1976, "Moisture Absorption and Desorption of Composite Materials," *J. Compos. Mater.*, **10**(1), pp. 2–20.
- [48] Fan, Y., Gomez, A., Ferraro, S., Pinto, B., Muliana, A., and La Saponara, V., 2017, "The Effects of Temperatures and Volumetric Expansion on the Diffusion of Fluids Through Solid Polymers," *J. Appl. Polym. Sci.*, **134**(31), p. 45151.
- [49] ISO, 2008, *62 Plastics—Determination of Water Absorption*, ISO, Geneva, Switzerland, <https://www.iso.org/standard/41672.html>
- [50] Brent, R. P., 1973, *Algorithms for Minimization Without Derivatives*, Prentice-Hall, Englewood Cliffs, New Jersey.