AIAA JOURNAL
Vol. 42, No. 5, May 2004

In-Phase Error Estimation of Experimental Data and Optimal First Derivatives

Jay Frankel,* Majid Keyhani,† and Kunihiko Taira‡
University of Tennessee, Knoxville, Tennessee 37996-2210

The development of in-phase, a posteriori error estimates and the best functional representation for the solution and its derivative are examined. Several new and fundamental findings pertinent to both the experimentalist and analyst are presented. Applications requiring a high degree of precision such as associated with the defense, aerospace, and heat treatment arenas may benefit from these new findings. The basic framework involves the classical least-squares method, which is reexamined with the intent of 1) approximating local measurement errors necessary for estimating uncertainty and 2) predicting optimal solutions in both the primitive function and its derivative. With accurate local-error estimates, the initial data can be corrected by adding these estimates to the initial data to substantially reduce the root-mean-square error, that is, uncertainty. Additionally, the clarification between best curve fit and best prediction or solution is elucidated to provide insight into the meaning of a desirable resolution. The best prediction permits the proper interpretation of the derivative, which is essential to many contemporary studies involving preprocessing of data. The framework offered presents an elegant but practical method for approximating local measurement errors and for developing a functional representation of the desired function that best represents the solution in the presence of errorless data.

Nomenclature

a_j = exact expansion coefficient; Eq. (5a)
a_j^N = expansion coefficient; Eq. (5b)
$\bar{B}_N^2$ = deterministic bias; Eq. (19a)
b_j^N = expansion coefficient; Eq. (9)
M = total number of data
N = number of terms in expansions
p = degree of polynomial
$R_N(t)$ = residual function; Eq. (6)
$\bar{R}_N(t)$ = residual function; Eq. (8)
S = total least-squares error; Eq. (13)
$\bar{S}$ = total least-squares error; Eq. (14)
T_i = data from measurement system
$T^N(t)$ = approximation of $T(t)$; Eq. (5b)
$T(t)$ = ideal (continuous) function
T_0 = initial condition
$\tilde{T}_{i,p}$ = corrected data using pth corrector
$\bar{T}^N(t)$ = approximate solution of $T(t)$; Eq. (9)
t = time
t_i = discrete time where data are collected
t_j = jth multiquadric center, $jt_{\max}/N$; Eq. (21)
$t_{\max}$ = cutoff time for collection of data
β = shape factor in multiquadric; Eq. (21)
γ = induced noise level factor; Eq. (16)
$\bar{\epsilon}_i$ = local error; Eq. (7b)
μ_N^2 = variance defined in Eq. (15)
μ_n^2 = variance defined in Eq. (4)
$\bar{\mu}_N^2$ = variance defined in Eq. (20)
σ = induced uncertainty, $\sigma = \sqrt{(\|\epsilon_i\|_2^2/M)}$
$\bar{\sigma}_{c,p}$ = corrected uncertainty, $\bar{\sigma}_{c,p} = \sqrt{[\|T - (T_i + \tilde{\epsilon}_{i,p})\|_2^2/M]}$
$\bar{\sigma}_N^2$ = variance defined in Eq. (19b)
$\bar{\sigma}_p$ = estimated uncertainty, $\bar{\sigma}_p = \sqrt{(\|\tilde{\epsilon}_{i,p}\|_2^2/M)}$
$\Omega_j(t)$ = trial function described in Eq. (5)

Received 24 December 2002; revision received 13 September 2003; accepted for publication 4 December 2003.

*Professor, Mechanical and Aerospace Engineering and Engineering Science Department; frankel@titan.engr.utk.edu.
†Professor, Mechanical and Aerospace Engineering and Engineering Science Department.
‡Undergraduate Student, Mechanical and Aerospace Engineering and Engineering Science Department.

Introduction

In the present context, noisy data are defined as data containing only a random error component. In error analysis, a single value for the global uncertainty is presented or assumed. In stark contrast, this study involves a local examination of the error and attaches an error estimation to each corresponding data value. If desired, a global uncertainty value can then be produced. Representing data by a function is advantageous in many situations. The best solution is developed in lieu of the best curve fit. This subtlety involves the development of a new methodology that in effect attempts to minimize$\|f - \bar{f}_N\|_2^2 = \|\epsilon_i - R_N\|_2^2$, where f is the ideal function truly being sought in the presence of errorless (continuous) data, $\bar{f}_N$ is the approximate function being developed using contaminated data, ϵ_i is the local error for the ith data point defined as $f(x_i) - f_i$, and R_N is the traditional local residual defined as $\bar{f}_N(x_i) - f_i$, where x_i is the ith collection point and f_i is the corresponding data value to form the data set denoted as $\{x_i, f_i\}_{i=1}^M$, where M represents the total number of data collected from the experiment. Here, $\|\cdot\|_2$ represents the discrete two norm.

Numerical differentiation of noisy data is known to be ill posed in the sense of Hadamard (see Ref. 1) because small perturbations in the function can lead to large variations in the derivative. This can quickly be understood by noting that the absolute error associated with forward difference approximation to $f'_\epsilon(x)$ in the presence of noisy data results in the error bound $f'_\epsilon < \mathcal{O}(h) + 2\delta/h$, where δ is the maximum absolute error, that is, $|f - f_i| < \delta$, and where h is the conventional step size. Therefore, for $\delta > 0$, as $h \to 0$, the resulting error bound blows up. Thus, $h \to 0$ cannot be performed in an arbitrary manner.

The least-squares method[2] involves selecting an n-parameter trial function normally given as $\bar{\Phi}_n(x)$ to a set of M data $\{x_i, \Phi_i\}_{i=1}^M$, where $n < M$ in such a manner that the n undetermined coefficients $\{a_j^n\}_{j=1}^n$ are obtained by minimizing the two-norm square of the residual, namely,

$$\min\left(S\left\{a_j^n\right\}_{j=1}^n\right) = \min\|R_n(x_i)\|_2^2 = \min\|\bar{\Phi}_n(x_i) - \Phi_i\|_2^2$$

$$= \min\sum_{i=1}^{M}[\bar{\Phi}_n(x_i) - \Phi_i]^2 \tag{1}$$

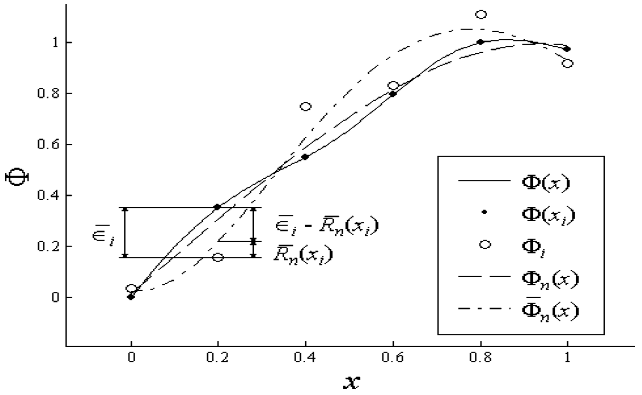

Fig. 1 Demonstration of functions.

Here, the residual is given as $R_n(x_i)$ and is defined as the deviation from the n-parameter approximate solution, $\bar{\Phi}_n(x)$ to the local data point Φ_i at $x = x_i$ for all $i = 1, 2, \ldots, M$. Hence, with regard to the true solution, the conventional least-squares methods as given in Eq. (1) merely provides a representative function that is intended to fit the data. The optimal curve fit is obtained by truncating the series at n terms (in the case of a linear least-squares method). This optimal condition is normally chosen with either the aid of the variance defined as

$$\sigma_n^2 = \frac{\|R_n(x_i)\|_2^2}{M-n}, \qquad n < M \tag{2}$$

where $M - n$ represents the degrees of freedom, or the determination index. Where the variance concept is used, the optimal curve fit is normally defined as the value of n that permits σ_n^2 to reach its first significant minimum.

The conventional least-squares method does not produce the best solution but merely produces the best representation or best curve fit to the data in the two-norm sense under the normally required statistical assumptions associated with the data.[2] Figure 1 shows several key ingredients necessary for the complete understanding of the mathematical formalism to be described. Exaggerated data and curves are used to assist in clarifying the concepts.

Figure 1 shows three curves and two sets of discrete data. The exact continuous solution is given by $\Phi(x)$ from which all exact and noisy data are simulated. Exact discrete data are denoted as $\Phi(x_i)$ and would lie on this curve at discrete locations $\{x_i\}_{i=1}^M$. Discrete data containing noise are denoted by open circles and given as Φ_i. The n-parameter approximate solution using the exact data set $\{x_i, \Phi(x_i)\}_{i=1}^M$ is given as $\Phi_n(x)$. Finally, the n-parameter approximate solution using the noisy data set $\{x_i, \Phi_i\}_{i=1}^M$ is given as $\bar{\Phi}_n(x)$.

From Fig. 1, it is evident that one would wish to drive the curve describing $\bar{\Phi}_n(x)$ to the exact solution curve $\Phi(x)$ by some manner. Here, minimizing the residual given as $R_n(x_i)$ could place $\bar{\Phi}_n(x)$ as indicated in Fig. 1. Substantial errors may be induced if derivative information is required by this approximation. It is well known that the minimum variance as defined in Eq. (2) will not necessarily produce the optimal prediction. Figure 1 indicates that the true goal should involve minimizing $\epsilon_i - R_n(x_i)$ in some norm sense. This difference can alternatively be expressed as $\Phi(x_i) - \bar{\Phi}_n(x_i)$ because the local error ϵ_i is given by $\epsilon_i = \Phi(x_i) - \Phi_i$ and the residual is given by $R_n(x_i) = \bar{\Phi}_n(x_i) - \Phi_i$; thus, we note $\epsilon_i - R_n = \Phi - \bar{\Phi}_n$. The goal should involve

$$\min\|\epsilon_i - R_n(x_i)\|_2^2 = \min \sum_{i=1}^{M} [\epsilon_i - R_n(x_i)]^2 \tag{3a}$$

or

$$\min\|\Phi(x_i) - \bar{\Phi}_n(x_i)\|_2^2 = \min \sum_{i=1}^{M} [\Phi(x_i) - \bar{\Phi}_n(x_i)]^2 \tag{3b}$$

With this, one would find the best solution or representation for $\Phi(x)$ denoted as $\bar{\Phi}_n(x)$. The optimal or best solution could be defined with the aid of variance defined by

$$\mu_n^2 = \frac{\|\epsilon_i - R_n(x_i)\|_2^2}{M-n} = \frac{\|\Phi(x_i) - \bar{\Phi}_n(x_i)\|_2^2}{M-n}, \qquad n < M \tag{4}$$

Thus, the best solution would be interpreted as the value of n that would produce the first significant minimum in μ_n^2. At first glance, this appears to be a hopeless goal because neither ϵ_i nor $\Phi(x)$ are known a priori. Equation (4) suggests that the discrete residual function should attempt to emulate the local errors at the optimal condition.

Mathematical Formulation

To begin the mathematical analysis, we consider the idealized situation involving perfect data denoted as the collection in the set $\{t_i, T(t_i)\}_{i=1}^M$, that is, data in the absence of any error. Here, time t is representative of the domain space, whereas $T(t)$ is representative of the range space. Several observations are available that when combined with the situation of noisy data denoted by $\{t_i, T_i\}_{i=1}^M$ permit the development of several fundamental mathematical relations to be derived and later exploited. Again, the goal is 1) to develop local error estimators and 2) to recover the best prediction for both $T(t)$ and $(\mathrm{d}T/\mathrm{d}t)(t)$ (Ref. 1). Here, $T(t)$ and $(\mathrm{d}T/\mathrm{d}t)(t)$ are interpreted as the theoretical exact functions in the presence of perfect, continuous data. Furthermore, let us denote the time frame of interest (for transient problems) as $t \epsilon [0, t_M]$ where $t_M = t_{\max}$. Let us further assume that leader data are available that can be used to establish $T(0) = T_0$, which leads to a constrained least-squares approach. This condition is not necessary to impose for the present algorithm.

To begin, consider the availability of exact continuous data and that we wish to approximate $T(t)$ by a finite series representation denoted as $T^N(t)$. For demonstration purposes, let[3]

$$T(t) = T_0 + \sum_{j=1}^{\infty} a_j \Omega_j(t), \qquad t \geq 0 \tag{5a}$$

or on truncating after N terms, we obtain

$$T(t) \approx T^N(t) = T_0 + \sum_{j=1}^{N} a_j^N \Omega_j(t), \qquad t \geq 0 \tag{5b}$$

where the set $\{\Omega_j(t)\}_{j=1}^N$ represents some predefined set of trial functions having a corresponding set of undetermined coefficients denoted by $\{a_j^N\}_{j=1}^N$. It is hoped that $a_j \approx a_j^N$ as N grows in the presence of perfect data. This is not the case when imperfect or noisy data are introduced.

The local residual $R_N(t_i)$ for this case is given by

$$R_N(t_i) = T^N(t_i) - T(t_i), \qquad i = 1, 2, \ldots, M \tag{6}$$

This residual represents the distance between the approximating function and the errorless discrete data. Here, M represents the total number of data in the set. In this case, the local error, that is, the distance between the data $T(t_i)$ and the exact solution $T(t_i)$, is zero because $\bar{\epsilon}_i = T(t_i) - T(t_i)$, $i = 1, 2, \ldots, M$.

In the noisy data case, the $T(t_i)$ can be decomposed into two parts as

$$T(t_i) = T_i + \bar{\epsilon}_i, \qquad i = 1, 2, \ldots, M \tag{7}$$

where T_i is the interpreted data from a measurement device and $\bar{\epsilon}_i$ is the local error of T_i at time t_i. (Note that an overbar over any function serves to indicate that noisy data are used.) It is tacitly assumed that no measurement error exists in time t. Unfortunately, in real problems, the exact value of $\bar{\epsilon}_i$ is unknown, although some bound can be determined by uncertainty analysis. The local residual is now expressed as

$$\bar{R}_N(t_i) = \bar{T}^N(t_i) - T_i, \qquad i = 1, 2, \ldots, M \tag{8}$$

where $\bar{R}_N(t_i)$ represents the discrete residual associated with the distance between the approximating function $\bar{T}^N(t_i)$ and the ith noisy data point denoted as T_i. Here, the numerical approximation for the unknown function $\bar{T}(t)$ is denoted as $\bar{T}^N(t)$ and is represented using the truncated series

$$\bar{T}^N(t) = T_0 + \sum_{j=1}^{N} b_j^N \Omega_j(t) \tag{9}$$

where the trial functions are given by $\{\Omega_j(t)\}_{j=1}^N$ having corresponding unknown expansion coefficients denoted by the set $\{b_j^N\}_{j=1}^N$. To reiterate, $T(t)$ represents the exact solution using perfect, continuous data; $T^N(t)$ represents the approximate solution in the presence of errorless data; and $\bar{T}^N(t)$ represents the approximate solution in the presence of inexact or noisy data. It should be clear that the missing ingredients to the actual reconstruction of either $T^N(t)$ from Eq. (5b) or $\bar{T}^N(t)$ from Eq. (9) are the numerical values for their respective expansion coefficients. Again, note that Eqs. (7) and (8) can be combined via subtraction to form

$$\bar{\epsilon}_i - \bar{R}_N(t_i) = [T(t_i) - T_i] - \left[\bar{T}^N(t_i) - T_i\right] = T(t_i) - \bar{T}^N(t_i)$$
$$i = 1, 2, \ldots, M \tag{10}$$

Qualitatively speaking and as previously noted, one should conceptualize four "plots" in the data domain, namely, the discrete data as interpreted by the measurement device leading to T_i; the exact errorless history, $T(t)$; the reconstructed functional representation of the data in the presence of perfect data, $T^N(t)$; and the reconstructed functional representation of the data, $\bar{T}^N(t)$ using inexact, discrete data. With these, the classical definitions for both bias and variance can be defined in the context of the mean-square error (Ref. 4, pp. 154–156).

It now appears germane to derive some relationships among several important functions to support the development of error estimation. Consider subtracting Eq. (6) from Eq. (8) to yield

$$R_N(t_i) - \bar{R}_N(t_i) = -\bar{\epsilon}_i + \left[T^N(t_i) - \bar{T}^N(t_i)\right]$$
$$N = 1, 2, \ldots, \qquad i = 1, 2, \ldots, M \tag{11}$$

because $\bar{\epsilon}_i = T(t_i) - T_i$. Equation (11) is a single equation containing three unknowns, namely, $\bar{\epsilon}_i$, $R_N(t_i)$, and $T^N(t_i)$. Clearly, a closure problem arises at this juncture. This fundamental equation and its proper interpretation contains the key ingredients for obtaining error estimates. For the moment, assume that $|T^N(t_i) - \bar{T}^N(t_i)| \ll |\bar{\epsilon}_i|$ for $N = 1$, thus, reducing Eq. (11), in an approximate sense, to

$$R_N(t_i) - \bar{R}_N(t_i) \approx -\bar{\epsilon}_i, \qquad i = 1, 2, \ldots, M \tag{12}$$

when $N = 1$. Observe that $T^N(t) - \bar{T}^N(t)$ is a smooth, continuous contributor in Eq. (11).

Equation (12) is a single equation containing two unknowns, namely, $\bar{\epsilon}_i$ and $R_N(t_i)$. Still a closure problem exists, although it now appears possible to form a reasonable estimate to $\{R_N(t_i)\}_{i=1}^M$ when $N = 1$ because the behavior of $\{\bar{R}_1(t_i)\}_{i=1}^M$ may closely represent the case of errorless data. That is, the high degree of fluctuation may not be evident in this special case. If so, then an estimate for $\bar{\epsilon}_i$, denoted by $\bar{\epsilon}_{i,p}$, may be developed using a least-squares representation. This counterintuitive concept can be, at this moment, qualitatively clarified by considering the effect of bias and variance. The bias involves the two norm of $T(t) - T^N(t)$, and the variance involves the two norm of $T^N(t) - \bar{T}^N(t)$. For small N, the bias is dominant over the variance. That is, the approximation is driven by numerical errors and not data errors. Equation (11) contains the difference between $T^N(t)$ and $\bar{T}^N(t)$, which represents a local building block in the evaluation of the variance. For $N = 1$, the contribution of this term is hopefully small. In the two-norm square sense, it is later illustrated that $\|T^N - \bar{T}^N\|_2^2 \ll \|\bar{\epsilon}_i\|_2^2$ when $N = 1$.

Least-Squares Formalism

The minimization of the residual follows classical doctrine, and thus, only a cursory discussion is offered for the cases involving 1) errorless data and 2) noisy data. In the presence of ideal data, let us write

$$S\left(\left\{a_j^N\right\}_{j=1}^N\right) = \sum_{i=1}^{M} [R_N(t_i)]^2 \tag{13a}$$

and on substituting Eq. (6) for $R_N(t_i)$ into Eq. (13a) where the expansion displayed in Eq. (5b) is used to represent $T^N(t)$, we arrive at

$$S\left(\left\{a_j^N\right\}_{j=1}^N\right) = \sum_{i=1}^{M}\left[T_0 - T(t_i) + \sum_{j=1}^{N} a_j^N \Omega_j(t_i)\right]^2 \tag{13b}$$

where $\Omega_j(u)$ is a trial function that satisfies $\Omega_j(0) = 0$, $j = 1, 2, \ldots, N$. The procedure for uniquely determining the expansion coefficients $\{a_j^N\}_{j=1}^N$ is well known.[3]

In the case of noisy data, let us write

$$\bar{S}\left(\left\{b_j^N\right\}_{j=1}^N\right) = \sum_{i=1}^{M} \bar{R}_N^2(t_i) \tag{14a}$$

and on substituting Eq. (8) for $\bar{R}_N(t_i)$ into Eq. (14a) where the expansion displayed in Eq. (9) is used to represent $\bar{T}^N(t)$, we arrive at

$$\bar{S}\left(\left\{b_j^N\right\}_{j=1}^N\right) = \sum_{i=1}^{M}\left[T_0 - T_i + \sum_{j=1}^{N} b_j^N \Omega_j(t_i)\right]^2 \tag{14b}$$

with $\Omega_j(u)$ defined earlier. The procedure for determining the expansion coefficients $\{b_j^N\}_{j=1}^N$ is well known.[3]

Error Analysis

This section relates several analytic concepts necessary for defining prediction optimality in the two-norm sense. In this presentation, we define the discrete two-norm square as

$$\|\Phi(t) - \Phi_N(t)\|_2^2 = \sum_{i=1}^{M} [\Phi(t_i) - \Phi_N(t_i)]^2$$

for arbitrary functions $\Phi(t)$ and $\Phi_N(t)$, where M is the number of collected data.

To begin, consider numerical accuracy in the presence of errorless data. The question that now arises involves determining a process whereby the optimal N can be determined. The procedure used to generate simulated data is now described. An analytic function for $T(t)$ subject to $T(0) = T_0$ is prescribed; then it is sampled at M equidistant times given by t_i, $i = 1, 2, \ldots, M$ to generate the data set $\{t_i, T(t_i)\}_{i=1}^M$. Note that the choice of equidistant sampling is merely a convenience. This produces the set of ideal data denoted by $T(t_i)$, $i = 1, 2, \ldots, M$.

The square error of the function in the approximation process containing errorless discrete data, $e_N^2(T)$, is given as $e_N^2(T) = \|T(t) - T^N(t)\|_2^2$, whereas the square error of its temporal derivative is given as $e_N^2(\dot{T}) = \|\dot{T} - \dot{T}^N\|_2^2$. Finally, the square error of the data itself is given by $e_i^2(T) = \|T(t_i) - T_i\|_2^2 = \|\bar{\epsilon}_i\|_2^2$, where for this case $e_i^2(T) = 0$ because $T_i = T(t_i)$, $i = 1, 2, \ldots, M$. The residual $R_N(t_i)$ is numerically obtained from Eq. (6) once the least-squares procedure is concluded and the expansion coefficients $\{a_j^N\}_{j=1}^N$ are known for each choice of N. Using the least-square procedure defined with the aid of Eq. (13b) and forming the normal equations, one can uniquely determine the necessary expansion coefficients $\{a_j^N\}_{j=1}^N$ for fixed value of N. Also, we note that

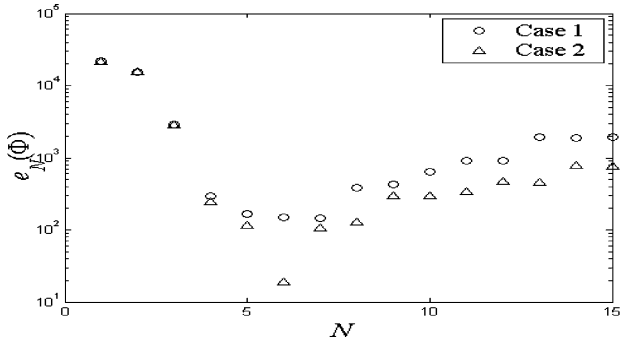

Fig. 2 Error trends.

$\|R_N\|_2^2 = \|T(t_i) - T^N(t_i)\|_2^2$. The conventional least-squares variance measure is, thus, defined as

$$\mu_N^2 = \frac{\sum_{i=1}^{M}[R_N(t_i)]^2}{M-N} = \frac{\|R_N\|_2^2}{M-N}, \qquad N < M \tag{15}$$

Next, consider the case where noisy data are generated with the aid of

$$T_i = T(t_i) - \gamma \text{random}(i)\|T(t)\|_\infty \tag{16}$$

where $\|T(t)\|_\infty = \max_{t\epsilon[0,t_{\max}]}|T(t)|$ and the term random(i) denotes the ith randomly drawn number from the closed interval $[-1, 1]$.

The approximate reconstruction of $T(t)$ denoted by $\bar{T}^N(t)$ contains two sources of errors, namely, data errors and truncation errors. Using the least-square procedure defined with the aid of Eq. (14) and forming the normal equations, one can uniquely determine the necessary expansion coefficients $\{b_j^N\}_{j=1}^N$ for fixed value of N. Again, the question arises, which value of N do we choose? The task at hand involves locating the point, or more precisely, the value of N in the series displayed in Eq. (9) for $\bar{T}^N(t)$ based on the two norm that can best replicate the exact errorless solution given as $T(t)$, that is, the best replicant of $T(t)$ is desired and not the data $\{T_i\}_{i=1}^M$. Again, this is not establishing the best curve fit as normally considered and defined with the variance shown in Eq. (15) [where $R_N(t_i)$ is replaced by $\bar{R}_N(t_i)$]. At that juncture, we stop adding terms to the series. This turning point is denoted as N_{opt}.

A graphical depiction of $\bar{e}_N^2(T) = \|T(t) - \bar{T}^N(t)\|_2^2$ against N, as shown in Fig. 2, would indicate that, as N increases, the error decreases rapidly up to the point defined as N_{opt} at which we subsequently see the undesirable effect of error amplification Detection of this occurrence is crucial. That is, the minimum error is observed at $N = N_{\text{opt}}$ before error amplification creeps into the prediction. Unfortunately, this quantity is not available for calculation because the exact solution $T(t)$ is unknown. Additionally, the value of N_{opt} depends on which of the two functions $\bar{T}^N(t)$ or $(\mathrm{d}\bar{T}^N/\mathrm{d}t)(t)$ is of key interest. Often, both collapse to a single value of N_{opt}, but it is not necessarily true for all cases. Two error patterns tend to exist. Figure 2 shows qualitatively these common patterns, indicating distinct characteristics that require different rules for identifying N_{opt}. Case 1 shows a flat region near N_{opt} whereas case 2 displays a sharp, single value for N_{opt}. Figure 2 suggests that the establishment of a single rule for finding N_{opt} will be an elusive task.

At the point of optimality, note $(\bar{e}^2_{N_{\text{opt}}} - \bar{e}^2_{N_{\text{opt}-1}}) < 0$, whereas $(\bar{e}^2_{N_{\text{opt}+1}} - \bar{e}^2_{N_{\text{opt}}}) > 0$. Additionally, one can mathematically establish

$$\bar{e}_N^2(T) - \bar{e}_{N-1}^2(T) = \|T(t) - \bar{T}^N(t)\|_2^2 - \|T(t) - \bar{T}^{N-1}(t)\|_2^2 \tag{17a}$$

Furthermore, with the aid of Minkowski's triangle inequality (see Ref. 3), one can show

$$\|T - \bar{T}^N\|_2 \le \|T - \bar{T}^{N-1}\|_2 + \|\bar{T}^{N-1} - \bar{T}^N\|_2 \tag{17b}$$

or

$$\bar{e}_N - \bar{e}_{N+1} \le \|\bar{T}^{N+1} - \bar{T}^N\|_2 \tag{17c}$$

The right-hand side is a computable quantity, which could indicate the progression of the errors under certain circumstances (such as indicated by the flat set of data near N_{opt} in Fig. 2).

The square error of the reconstructed function $\bar{T}_N(t)$ is defined as $\bar{e}_N^2(T)$ as given by

$$\bar{e}_N^2(T) = \|\bar{\epsilon}_i - \bar{R}_N(t_i)\|_2^2 \tag{18a}$$

where $\bar{\epsilon}_i - \bar{R}_N(t_i)$ is suggested earlier. The square error $\bar{e}_N^2(T)$ can be expressed in terms of the bias $\bar{B}_N^2$ and variance $\bar{\sigma}_N^2$ as

$$\|\bar{\epsilon}_i - \bar{R}_N(t_i)\|_2 = \|T(t) - \bar{T}^N(t)\|_2 \le \|T(t) - T^N(t)\|_2 + \|T^N(t) - \bar{T}^N(t)\|_2 \tag{18b}$$

where[5]

$$\bar{B}_N^2 = \|T(t) - T^N(t)\|_2^2 = \sum_{i=1}^{M}\left[T(t_i) - T^N(t_i)\right]^2 \tag{19a}$$

$$\bar{\sigma}_N^2 = \|T^N(t) - \bar{T}^N(t)\|_2^2 = \sum_{i=1}^{M}\left[T^N(t_i) - \bar{T}^N(t_i)\right]^2 \tag{19b}$$

Error amplification is the result of the conflict between a decreasing bias $\bar{B}_N^2$ and an increasing variance $\bar{\sigma}_N^2$ (Ref. 4, pp. 153–156) for increasing N.

The variance μ_N^2 as defined in Eq. (15) is an indicator for the case of errorless data. A similar measure can be expressed in the presence of noisy data, namely,

$$\bar{\mu}_N^2 = \frac{\|\bar{\epsilon}_i - \bar{R}_N(t_i)\|_2^2}{M-N} = \frac{\|T(t) - \bar{T}^N(t)\|_2^2}{M-N} \le \frac{(\|\bar{\epsilon}_i\|_2 + \|\bar{R}_N(t_i)\|_2)^2}{M-N}, \qquad N = 1, 2, \ldots \tag{20}$$

Equation (20) ensures the best approximation for $T(t)$, namely, the reconstructed function $\bar{T}^N(t)$. Unfortunately, both $\bar{\epsilon}_i$ and $T(t)$ are elusive in an exact sense at this stage. However, it appears possible to form estimates based on a recent observation for approximating $\bar{\epsilon}_i$ by exploiting the nature of bias and variance.[3] Note that the inequality shown in Eq. (20) will often lead to gross overestimations of the numerical value for the variance, though the trend may be qualitatively correct.

Results

The goal of the present investigation involves developing a quantifiable procedure for 1) approximating the local errors to form a global uncertainty value and to correct the local measured values and 2) identifying the optimal prediction for the best representation of the true solution and its derivative. Hardy multiquadric radial basis functions are presently used as the basis functions to form the trial functions denoted by $\Omega_j(t)$, $j = 1, 2, \ldots, N$ with $\Omega_j(0) = 0$, $j = 1, 2, \ldots, N$. The trial function are chosen as[3]

$$\Omega_j(t) = \sqrt{\beta_j + (t - t_j)^2} - \sqrt{\beta_j + t^2} - \sqrt{\beta_j + t_j^2} + \sqrt{\beta_j} \qquad j = 1, 2, \ldots, N \tag{21}$$

where t_j is the jth center and β_j is the jth shape factor corresponding to the jth center. Based on physical insight, we let $\beta_j = \beta$, $j = 1, 2, \ldots, N$, given by $\beta_j = \beta = (t_{\max}/N)^2$ and choose equidistant centers defined as $t_j = jt_{\max}/N$.

Figure 3 presents results using ideal data. Figure 3a shows the idealized continuous data function $T(t)$ to be sampled and later perturbed to generate the noisy data. For sake of discussion, let $t_{\max} = 2.5$, $T_0 = 0$, and $M = 50$ data points. This function has its origins in inverse heat conduction studies.[1,4] Discrete exact data $\{T(t_i)\}_{i=1}^M$ are extracted from the $T(t)$ curve at equidistant temporal locations governed by $t_i = i\Delta t$, $i = 1, 2, \ldots, M$, such that $\Delta t = t_{\max}/M$. Figure 3a shows the discrete data by solid dots. The

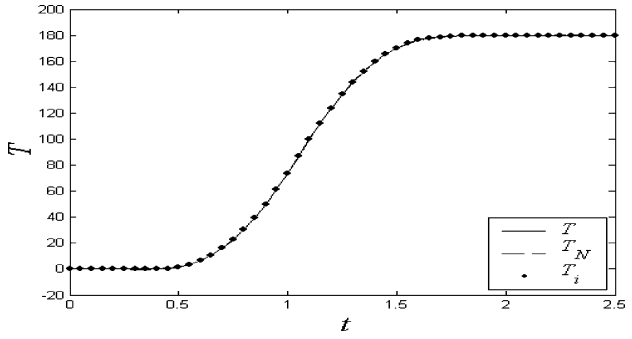

Fig. 3a Function $T(t)$.

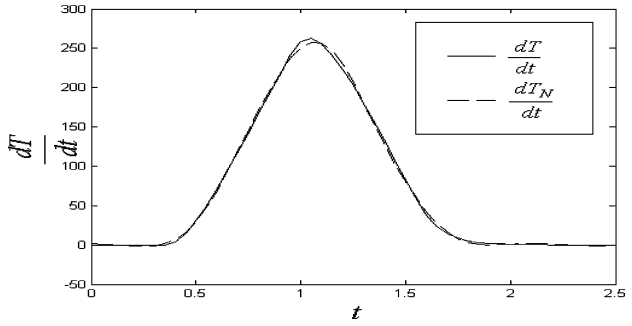

Fig. 3b Function derivative $\dot{T}(t)$.

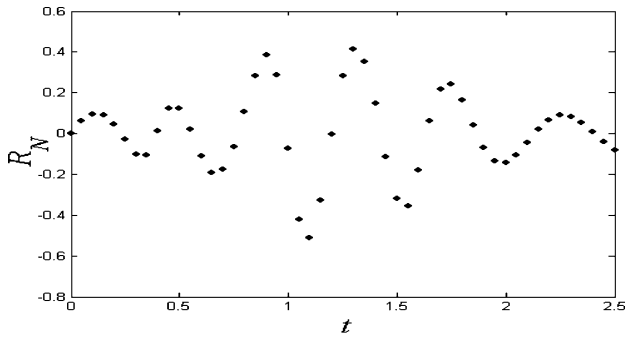

Fig. 3c Residuals $R_N(t_i)$.

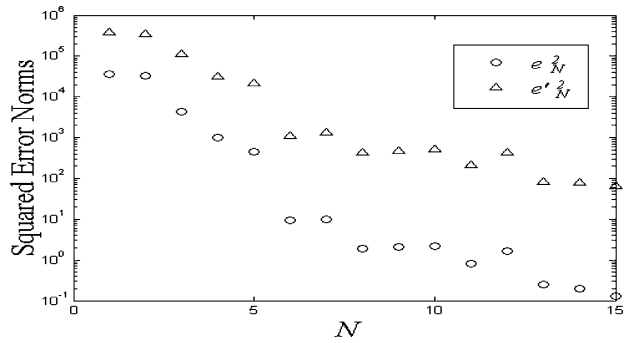

Fig. 4 Square error norms.

exact function $T(t)$ is the solid line, whereas the optimal prediction ($N=8$) $T^N(t)$ is the dashed curve. (Two function curves are indistinguishable.) Figure 3b shows the exact and predicted time derivatives of these functions. Figure 3c shows the smooth and oscillatory nature of the resulting residuals $R_N(t_i)$.

Figure 4 presents the errors $e_N^2(T)$ and $\dot{e}_N^2(\dot{T})$ for increasing values of N. The first significant minimum in the errors for both $e_N^2(T)$ and $\dot{e}_N^2(\dot{T})$ takes place at $N=8$. Above this value of N, no appreciable change in the prediction is noted. The goodness of fit is often qualitatively assessed through the variance and the determination index. The variance value μ_N^2 described by Eq. (15) also predicts that the first significant minimum takes place at $N=8$.

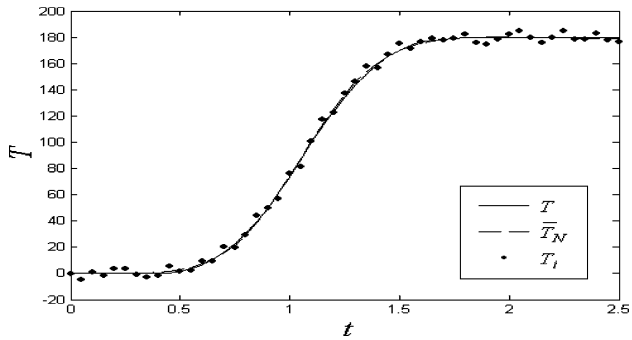

Fig. 5a Function, approximation, and data.

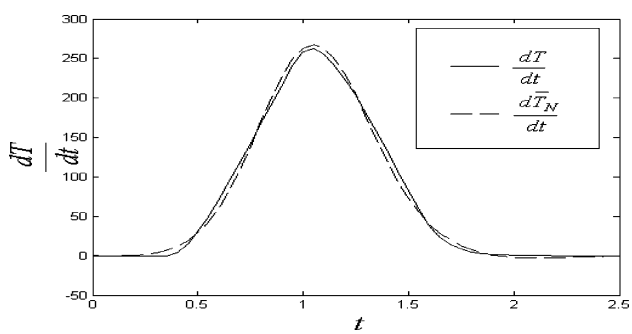

Fig. 5b Function derivative and approximation.

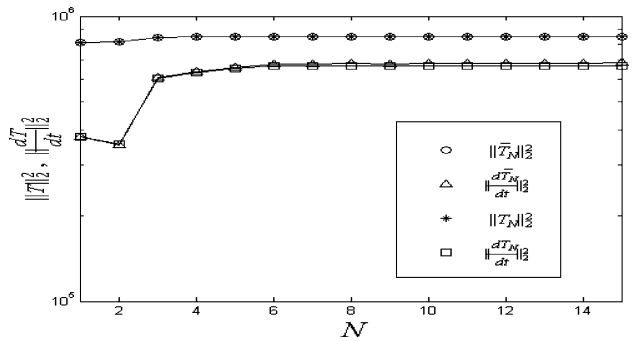

Fig. 5c Function square norms.

Next, consider the noisy data set $\{t_i, T_i\}_{i=1}^M$ as numerically simulated in accordance to Eq. (16). For the sake of discussion, let $\gamma=0.03$ and continue using $t_{\max}=2.5$, $T_0=0$, and $M=50$. Figure 5 presents results germane to this situation. In Fig. 5a, the simulated noisy data $\{T_i\}_{i=1}^M$ are the solid dots, the exact function $T(t)$ is the solid line and the reconstructed prediction $\bar{T}^N(t)$ is the dashed line. Here, it is found that the optimal prediction occurs when $N=6$, in accordance to $\bar{e}_N^2(T)$ and $\dot{\bar{e}}_N^2(\dot{T})$. Figure 5b shows the exact and predicted time derivatives when $N=6$. In both cases, highly accurate predictions are shown. Figure 5c shows the two-norm square of the functions $T^N(t)$, $\bar{T}^N(t)$, and their derivatives as N is increased.

Here, the connection between the points is presented only to assist the reader in establishing the asymptotic behavior as $N\to\infty$. As expected, the predictions involving ideal data tend toward an asymptotic value, whereas the predictions involving the noisy data do not. Low values of N can be eliminated with this understanding.

Figure 6 shows the error $\{\bar{\epsilon}_i\}_{i=1}^M$ and the resulting residuals $\{\bar{R}_N(t_i)\}_{i=1}^M$ at the optimal prediction, $N=6$. As expected, they should appear similar as noted by Eq. (10). Figure 7a shows the error trends for the noisy data set for increasing values of N. For this set, $N=6$ is clearly the optimal value. Figure 7b shows the conflict between the bias $\bar{B}_N^2$ described in Eq. (19a) and the variance $\bar{\sigma}_N^2$ given in Eq. (19b). For small N, it is evident that the bias dominates over the variance, whereas for large N the variance dominates over the bias. A balance between both takes place at optimal N.

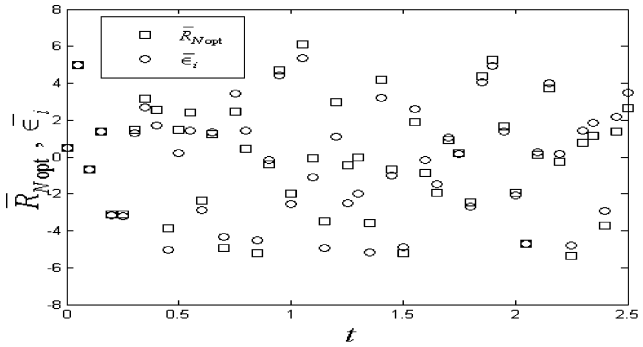

Fig. 6 Errors and optimal residuals.

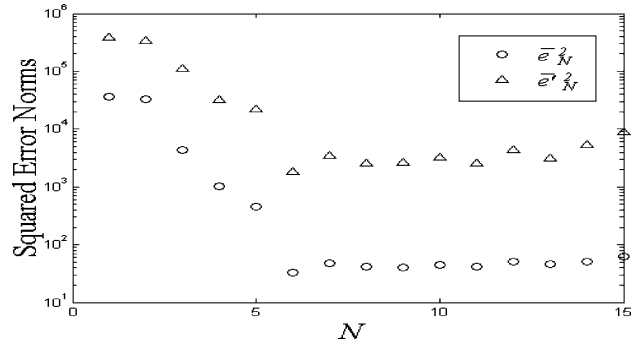

Fig. 7a Square error norms.

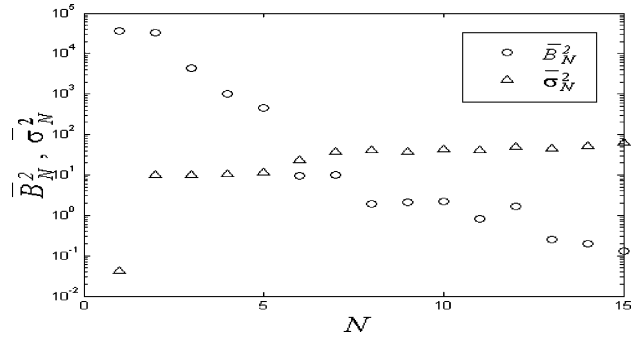

Fig. 7b Bias and variance.

This conflict is exploited in arriving at Eq. (12) and will shortly be qualified further.

Up to now, the reported results show that it is possible to predict the optimal or best solution when provided true values for the induced errors. In reality, estimating errors will be required to implement the preceding analyses. The process begins with the closure problem described in Eq. (11) with the assumption used to arrive at Eq. (12). To illustate the building blocks required to approximate the errors, we begin by investigating the worst-case approximation of the data. This counterintuitive ideas leads to the reduction displayed in Eq. (12). Figure 8 displays the predictions for the function $T(t)$ (Fig. 8a) and its derivative $\mathrm{d}T/\mathrm{d}t$ when $N = 1$ (Fig. 8b) using the noisy data displayed in Fig. 5a. The solid lines in Figs. 8a and 8b indicate the exact, continuous functions. The numerical results are indicated by the dashed lines and are clearly indicative of poor predictions.

The next ingredient in the recipe involves estimating the exact, discrete residuals $\{R_1(t_i)\}_{i=1}^M$ for $N = 1$ in the presence of noiseless data. To accomplish the task of closure, as necessitated by Eq. (12), we form a model based on approximating $R_1(t_i)$ by using $\{\bar{R}_1(t_i)\}_{i=1}^M$ as data for generating the best representation for $R_1(t)$, denoted by $\bar{R}_{1,p}(t)$, that is, $R_1(t_i) \approx \bar{R}_{1,p_{\mathrm{opt}}}(t_i)$, $t\epsilon[0, t_{\max}]$, where p_{opt} represents the optimal prediction. To justify the assumption indicated in Eq. (12) when $N = 1$, we begin by reporting the numerical value of the expansion coefficient based on perfect, discrete

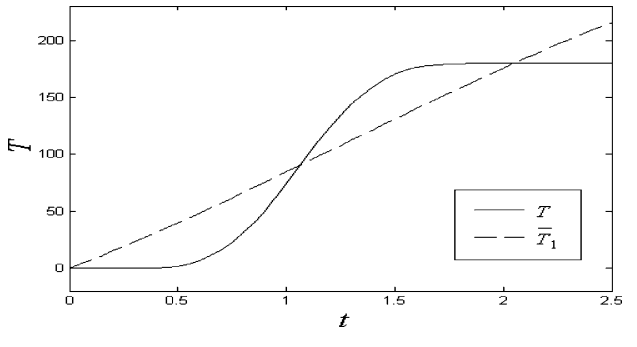

Fig. 8a Function and $N = 1$ approximation.

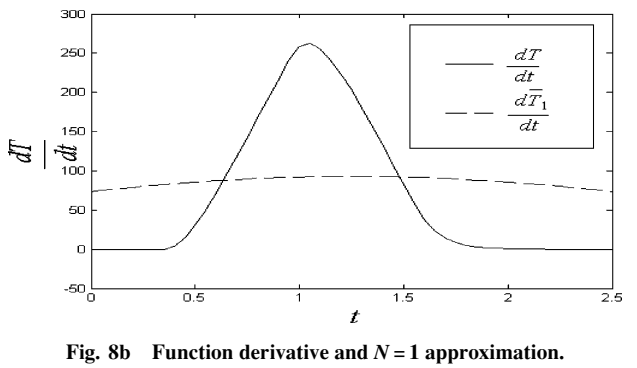

Fig. 8b Function derivative and $N = 1$ approximation.

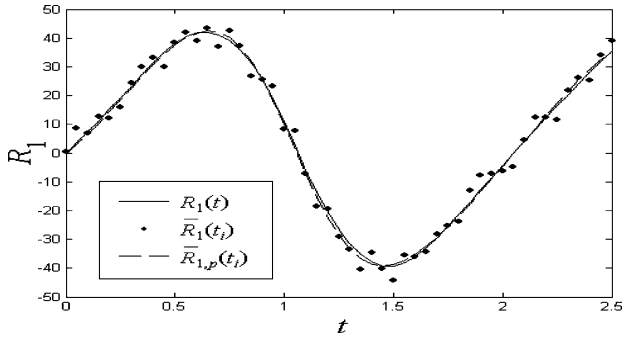

Fig. 9a Residuals and function residuals, $N = 1$.

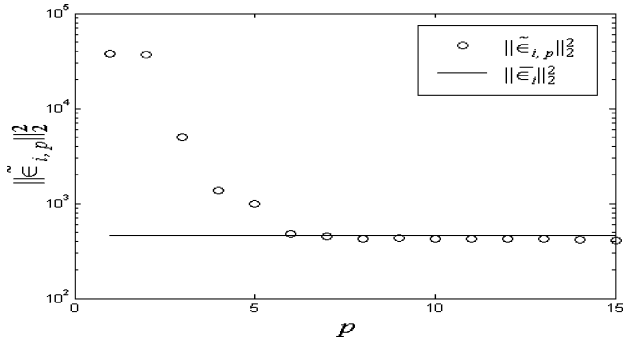

Fig. 9b Square error norm of approximate errors.

data as $a_1^N = -104.0942$. The corresponding expansion coefficient in the presence of the noisy data used in Fig. 5a is $b_1^N = -104.1179$. With these numerical values, the expansions for $T^N(t)$ and $\bar{T}^N(t)$ render graphically identical results. This indicates that for $N = 1$ the approximation is dominated by numerical errors and not data errors.

Figure 9a presents the residual $\{\bar{R}_1(t_i)\}_{i=1}^M$, denoted by solid dots, resulting from the data used in Fig. 5a. The solid line represents the corresponding exact residual function $R_1(t)$ in the presence of errorless discrete data. [Though these values are actually discrete,

poetic license is taken by connecting these points to clearly indicate the scattering effect about the errorless situation and motivates the modeling concept. This continuous curve $R_1(t)$ is reminiscent of $R_1^*(t)$ that would be obtained by a continuous least-squares solution using function data.] With these data randomly scattered about $R_1(t)$, it appears quite suggestive that a curve fit may be used nearly to replicate $R_1(t)$. As such, we begin by developing a crude, first-cut approach. Let us approximate $R_1(t)$ using a simple curve-fitting routine using the discrete $\bar{R}_1(t_i)$ data with the realization that an optimal curve fit does not necessary equate to $R_1(t)$ as shown earlier. Radial basis functions (Hardy multiquadrics) are again used as the trial functions in the least-squares algorithm. Recall, $\bar{R}_{1,p}(t)$ is the approximation being sought, where p is the number of terms in the representative expansion of the function using $\{\bar{R}_1(t_i)\}_{i=1}^M$ as the data. For the least-squares method, the local residual is now defined as $\bar{r}_{p,i} = \bar{R}_{1,p}(t_i) - \bar{R}_1(t_i)$, $i = 1, 2, \ldots, M$. Determining p_{opt} places us in a similar dilemma as before; however, at this stage, it is at a secondary approximation level. Once the p spectrum of results from, for example, $p = 1, 2, \ldots, N'$, are complete, expansion coefficients for the p-term representation of $\bar{R}_{1,p}(t)$ are known. Once the optimal p is chosen from the N' choices, then an approximation for the local errors $\bar{\epsilon}_i$, namely, $\tilde{\epsilon}_{i,p}$, are expressible with the aid of the modified form of Eq. (12), that is, $\bar{R}_{1,p}(t_i) - \bar{R}_1(t_i) = -\tilde{\epsilon}_{i,p}$, $i = 1, 2, \ldots, M$, where $\{\tilde{\epsilon}_{i,p}\}_{i=1}^M$ is obtained after the approximation for the residual function is provided, normally at $p = p_{\text{opt}}$. Figure 9b shows how p_{opt} is defined in the present context. The estimated two-norm square error $\{\bar{\epsilon}_{i,p}\}_{i=1}^M$ is plotted against increasing values of p in Fig. 9b. Figure 9b indicates that a leveling-off region occurs as p is increased past the minimum threshold value of $p = 5$. The solid, horizontal line represents $\|\bar{\epsilon}_i\|_2^2$ (true, induced errors). The classical (least-squares) variance analysis could be used in identifying the optimal curve fit for the data. Note that the choice of p greater than 6 would have little effect on correcting the initial data, as will be described later. Thus, if the main purpose of a study involves data correction, that is, $\tilde{T}_{i,p} = T_i + \tilde{\epsilon}_{i,p}$ and uncertainty estimation, then a great deal of flexibility exists in choosing p as long as it is in proximity of the elbow in Fig. 9b.

Figure 10a contrasts the induced errors $\{\bar{\epsilon}_i\}_{i=1}^M$ to the estimated errors $\{\tilde{\epsilon}_{i,p}\}_{i=1}^M$ when $p = 6$ using $N = 1$ results. Figure 10a shows that the local error estimates are in-phase with the actual errors. Note the remarkable similarity between the estimated and induced errors.

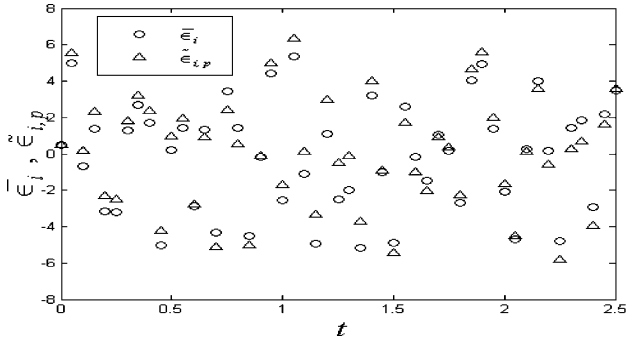

Fig. 10a Errors and approximate errors.

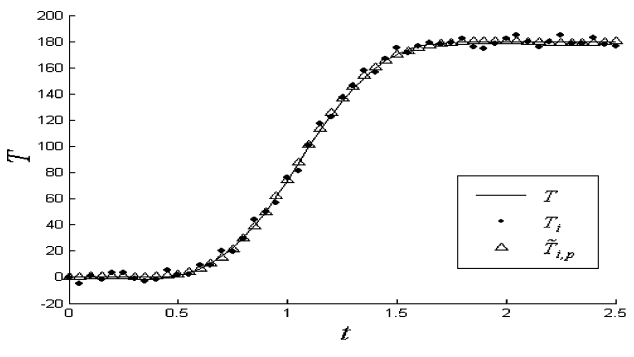

Fig. 10b Function, data, and corrected data.

The induced uncertainty is $\sigma = 3.0323$, whereas the estimated uncertainty $\bar{\sigma}_p$ is 3.0891. Figure 10b shows the effect of adding the estimated errors to the data, namely, $\tilde{T}_{i,p} = T_i + \tilde{\epsilon}_{i,p}$, $i = 1, 2, \ldots, M$. These adjusted data are denoted by the triangles, whereas the initial data are presented as solid dots. The exact function is given by the solid line. When the estimated errors are added to the initial data, the uncertainty of the corrected data is $\bar{\sigma}_{c,p} = 0.86948$. This represents a substantial reduction in the uncertainty.

To find N_{opt}, and due to space limitations, one concept proposed by the authors involves using the adjusted data to find an estimate of the derivative of the function, now denoted as $\tilde{\dot{T}}_{i,p}$ taken at $p = p_{\text{opt}}$. This derivative is developed with the aid of three- and five-point finite difference rules.[5] Additionally, although not described here, Tikhonov regularization (see Ref. 4) can also be applied. The authors have developed a cost functional with a penalty term containing the two norm of the second derivative of the function of interest. Results using this approach will be reported at a later time. Figure 11 shows the estimated derivatives as implemented using the corrected data for both rules. The three-point rule results are triangles, whereas the five-point results are circles. The exact function $\dot{T}(t)$ is the solid line.

Once these derivatives are formed, $\|\dot{\bar{T}}^N(t) - \dot{\tilde{T}}_{i,p}\|_2^2$ is studied with the purpose of estimating N_{opt}. The value for N_{opt} is chosen such that this norm reaches its first minimum. Figure 12a shows $\|\bar{T}^N(t) - \tilde{T}_{i,p}\|_2^2$ over increasing N for fixed $p = p_{\text{opt}} = 6$, whereas Fig. 12b shows $\|\bar{T}^N(t) - \tilde{T}_{i,p}\|_2^2$ for both finite difference

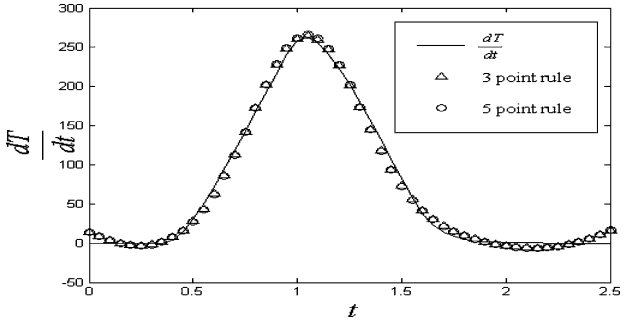

Fig. 11 Function derivative and finite differences.

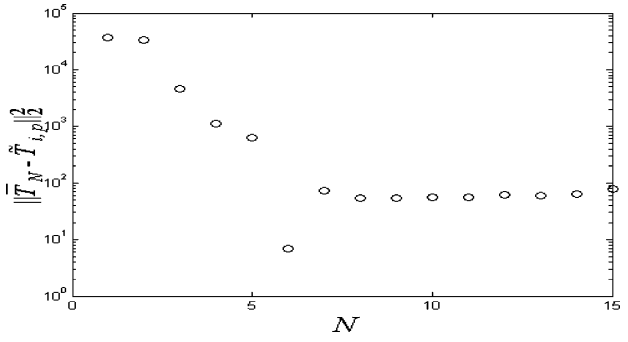

Fig. 12a Square error norm using corrected data.

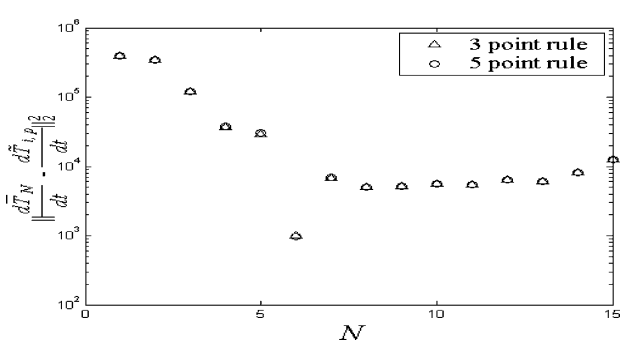

Fig. 12b Square error norm of corrected derivative.

implementations on the corrected data. Again, N_{opt} is indicated at the correct value of $N = 6$. Note that because of the sharp error nature indicated by Fig. 7a this rule works well. Simultaneously, Eq. (17c) is also reviewed. For the sharp error behavior, Eq. (17c) will tend to choose a larger value of N than indicated by the present rule. Thus, the minimum of the two N values is chosen.

Conclusions

This antecedent paper is predicated on the existence of a mathematical foundation for approximating 1) local errors and 2) the optimal truncation condition for the predicting the best solution based on a careful examination of the residuals. A suggestive framework is developed for exploiting the classical least-squares method and broadening its appeal through a basic understanding of the residuals. It is demonstrated that the global uncertainty and temporal error of noisy data can be estimated with a remarkable accuracy (unknown $\sigma = 3.0323$ and estimated $\bar{\sigma}_p = 3.0891$). Furthermore, the estimated temporal errors are in-phase, such that, by adding the estimated errors to the noisy data, the uncertainty of the corrected data is reduced to $\bar{\sigma}_{c,p} = 0.86948$.

References

[1]Frankel, J. I., and Keyhani, M., "Inverse Heat Conduction: The Need of $\partial T/\partial t$ Data for Design and Diagnostic Purposes," International Association of Science and Technology for Development, Modelling Identification, and Control, Feb. 1999.

[2]Mendenhall, W., and Scheaffer, R. L., *Mathematical Statistics with Applications*, Duxbury, Boston, 1973, pp. 375–430.

[3]Frankel, J. I., Keyhani, M., and Taira, K., "A Modified Least-Squares Method for Optimal Solutions in Inverse Problems," AIAA Paper 2002-0657, Jan. 2002.

[4]Beck, J. V., Blackwell, B., and St. Clair, C. R., Jr., *Inverse Heat Conduction*, Wiley, New York, 1985, pp. 153–156.

[5]Abramowitz, M., and Stegun, I. A. (eds.), *Handbook of Mathematical Functions*, Dover, New York, 1972, pp. 883, 914.

S. Saigal
Associate Editor