SIAM J. SCI. COMPUT.
Vol. 32, No. 4, pp. 2146–2166

ERROR BOUNDS FOR LEAST SQUARES GRADIENT ESTIMATES*

IAN W. TURNER†, JOHN A. BELWARD‡, AND MOA'ATH N. OQIELAT‡

Abstract. Least squares gradient estimates find application in many fields of computational science, particularly for the purposes of surface fitting and gradient reconstruction in computational fluid dynamics and data visualization. In this paper we derive error bounds for classical and weighted least squares gradient estimates. The bounds reflect how the number of points used in the least squares stencil and the smallest singular value of the least squares matrix impact the accuracy of these estimates. We show how an extrapolation method based on Householder transformations provides substantially tighter bounds. Numerical case studies are presented to elucidate our theory for a data set taken from Franke [*Math. Comp.*, 38 (1982), pp. 181–200].

Key words. surface fitting, truncated Taylor series, weighted least squares, extrapolation method

AMS subject classifications. 15A18, 65D10, 65D25, 65F20, 65G99

DOI. 10.1137/080744906

1. Introduction.

The accurate estimation of the gradient of a function $f : D \subset R^2 \to R$, for some domain of interest D, from a set of scattered function values arises in many important applications in applied and computational mathematics. One application of particular interest to the authors is the measurement of leaves of plants to capture their image for the modeling of droplet movement and absorption on the leaf surface. In this case the leaf surface representation requires a smooth fit to a set of scattered data, and our preferred method to obtain this surface is to represent the function by a set of cubic elements defined on a union of triangular domains using a Delaunay triangulation of the data points [13]. The surface is then represented using Clough–Tocher basis functions (see, for example, [12]), which enables a piecewise cubic surface with continuous gradient to be obtained if the function values and the gradients are known at the original data points and the gradient is also known at the midpoints of the edges of the triangulation. Given that only the data values are known at the scattered points, the implementation of our surface fitting strategy requires estimation of the gradient at the desired points; namely, given the values z_i at the points $(x_i, y_i)^T$, estimate the values of the gradient $\nabla f(x_i, y_i)$.

Two other important application areas where least squares gradient estimation is used are in computational fluid dynamics [2, 11, 15] and data visualization [16]. In the former, gradient reconstruction plays an important role in building accurate flux approximations at discrete cell faces, while in the latter the gradients are used for real-time volume rendering by providing a surface normal approximation that can be used for lighting, shading, and assigning opacity.

It is well known that multivariable Taylor expansions relate function values to derivatives, and these generate linear relations amongst the derivatives and function

*Received by the editors December 23, 2008; accepted for publication (in revised form) May 6, 2010; published electronically July 29, 2010.
http://www.siam.org/journals/sisc/32-4/74490.html
†Mathematical Sciences, Queensland University of Technology (QUT), GPO Box 2434, Brisbane, Queensland 4001, Australia (i.turner@qut.edu.au).
‡Queensland University of Technology (j.belward@qut.edu.au, m.oqielat@student.qut.edu.au).
The second and third authors wish to thank the Queensland University of Technology for research funding.

values. These expansions provide an excellent mechanism for derivative estimation, and when the data points are subject to error it seems only natural to form overdetermined systems of equations and then obtain gradient estimates by minimizing residuals via a least squares approach. We exploited this approach in [3], where we considered two different least squares strategies for approximating the local gradient estimates and analyzed the least squares errors associated with each method. An important question that arose from that work is "What is the spatial error associated with this type of estimation strategy?" Intuitively it seems quite plausible that this error will be $O(h^n_{\max})$, where n is the number of terms taken in the Taylor expansion and $h_{\max}$ is the maximum distance from the point of interest, say, a, and any of the cloud of neighboring points used for estimating the gradient. This assertion certainly seems to be well accepted in the literature, although the leading constant does not appear to have been explicitly identified. Many authors assert the order of piecewise polynomial approximations in $\mathbb{R}^2$; [1] is a typical example. Earlier, Stead [17] investigated gradient estimation by taking derivatives of interpolants based on Shepard's method and using hyperbolic multiquadrics. Goodman and Said [9] investigated C^1 triangular interpolants and later in [10] estimated the gradient as a convex combination of derivatives of triangular planes. Wei, Hon, and Wang [18] and Bozzini and Rossini [5] give estimates for the construction of numerical derivatives from noisy data, while Zuppa [19] has derived error bounds for derivative approximation based on algorithms using Gaussian elimination for a modified local Shepard's approximation. In the current work orthogonal transformations are used for the error analysis, thereby enabling the error bounds to be given in terms of the singular values of the least squares matrix used in our algorithm.

In recent work [4], we derived an error bound where this leading constant was given for a quadratic least squares gradient approximation, and it was conjectured, as a result of numerical experimentation and observation, that this bound could be further improved. This current work generalizes the results presented in [4] to an order n least squares gradient approximation and a proof of the conjecture is given. The main contribution is not only to investigate the order of the method but also to show that the smallest singular value of the least squares gradient coefficient matrix plays an important role; if this matrix is ill-conditioned, the greater the impact on the overall error. The importance of singular values in least squares approximation has been given prominence by Davydov [7], whose work uses powerful theoretical results to develop algorithms employing error estimates and error control. Davydov's methods are able to cater to contoured, tracked, and scattered data [6].

Another well-documented approach to local gradient estimation is to use a weighted least squares method, whereby the system is row-scaled in the sense that more importance is given to points closer to the point of interest a [2, 11]. Interestingly, our theory shows that this approach may not improve the situation at all, because the smallest singular value of the scaled matrix is usually smaller than that of the original matrix.

The paper is structured as follows. In section 2 a brief overview of least squares gradient estimation is given, and error bounds are derived for both the classical and weighted least squares estimates in section 3. We present two analyses of the order of convergence of these methods. The second in section 4 exploits a Householder reduction of certain columns of the least squares matrix to reveal that error estimates of subsets of the derivatives can be made using the singular values associated with those columns. An important corollary is that the gradient estimates are genuinely $\mathrm{O}(h^n)$ since these singular values are independent of h. In section 5.1 numerical results

are given exhibiting the predicted asymptotic behavior of the derivative estimates, while in section 5.2 we turn to a more practical situation and present results on point sets taken from Franke's celebrated paper [8]. Finally, in section 6 the main conclusions of the research are summarized.

2. Least squares gradient estimation. Although many representations of surfaces are possible, here we assume the representation is a function $f : D \subset R^2 \to R$, $z = f(x, y)$. Hence, a reference plane is assumed to exist with a unique ordinate at each data point in the xy-plane. Unless otherwise stated in the paper, $\|\cdot\|$ is assumed to be the Euclidean norm.

The gradient estimation strategy is now outlined. Suppose that point $a = (a_x, a_y)^T \in D$ is surrounded by m scattered data points $v_i = a + h_i\nu_i = (x_i, y_i)^T$, $i = 1, \ldots, m$, with $h_i = \|v_i - a\|$ and we require an estimate of the gradient $\nabla f(a)$. Assuming that f is of class C^{n+1} on an open set containing the line segment from a to $a + h\nu$, then Taylor's theorem for several variables states that

$$f(a + h\nu) = f(a) + h\frac{(\nu \cdot \nabla)f(a)}{1!} + \cdots + h^n \frac{(\nu \cdot \nabla)^n f(a)}{n!} + R_n, \tag{2.1}$$

where the remainder R_n has the integral form

$$R_n = \frac{h^{n+1}}{n!} \int_0^1 (1 - t)^n (\nu \cdot \nabla)^{n+1} f(a + th\nu)dt.$$

We now consider, as an example (see [4]), the use of relation (2.1) to write an overdetermined system of equations for the case $n = 2$, where we have the overdetermined system $A\gamma \approx q$ with $\gamma = (\frac{\partial f}{\partial x}(a), \frac{\partial f}{\partial y}(a), \frac{\partial^2 f}{\partial x^2}(a), \frac{\partial^2 f}{\partial x \partial y}(a), \frac{\partial^2 f}{\partial y^2}(a))^T$ and

$$A = \begin{pmatrix} \nu_{x_1} & \nu_{y_1} & \frac{1}{2}h_1\nu_{x_1}^2 & h_1\nu_{x_1}\nu_{y_1} & \frac{1}{2}h_1\nu_{y_1}^2 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \nu_{x_m} & \nu_{y_m} & \frac{1}{2}h_m\nu_{x_m}^2 & h_m\nu_{x_m}\nu_{y_m} & \frac{1}{2}h_m\nu_{y_m}^2 \end{pmatrix} \in R^{m\times 5}.$$

The entries in the matrix are defined as $h_i\nu_{x_i} = x_i - a_x, h_i\nu_{y_i} = y_i - a_y$, and the right-hand side vector $q \in R^{m\times 1}$ has as its ith component $q_i = \frac{f(a+h_i\nu_i)-f(a)}{h_i}$. The solution of the least squares problem is given by $\hat{\gamma} = \arg\min_{\gamma\in R^5}\|A\gamma - q\|$, which then enables the gradient estimate to be extracted from the first two components of $\hat{\gamma}$ as $\nabla f(a) \approx E_1\hat{\gamma} = E_1A^\dagger q$, where $E_1 \in \mathbb{R}^{2\times 5}$ is defined by the first two rows of the identity matrix I_5 and $A^\dagger$ is the pseudoinverse, or generalized inverse, of A (see, for example, [14]).

Note also that each of the estimates of the directional derivative may be weighted without loss of accuracy. This follows since the effect of a weight factor w_i is to introduce a diagonal matrix $W = \mathrm{diag}(w_1, w_2, \ldots, w_s)$, where typically one would use inverse distance, or inverse distance squared weights $w_i = \|a - v_i\|^{-d}, d = 1, 2$, to give more significance to points closer to a. In this case the overdetermined system becomes $WA\gamma \approx Wq$, and the least squares problem gives $\tilde{\gamma} = \arg\min_{\gamma\in R^5}\|WA\gamma - Wq\|$, which then enables the gradient estimate to be extracted as the first two components of $\tilde{\gamma}$ as $\nabla f(a) \approx E_1\tilde{\gamma} = E_1(WA)^\dagger Wq$, where $(WA)^\dagger$ is the pseudoinverse of WA.

Error bounds for both the classical $\|\nabla f(a) - E_1\hat{\gamma}\|$ and weighted $\|\nabla f(a) - E_1\tilde{\gamma}\|$ least squares gradient estimates for a general nth order approximation are derived in the next section. In both cases it is shown that the bounds take the form $\mathcal{C}\frac{h_{\max}^n \kappa_2(W)}{\sigma_1}$,

where $\mathcal{C}$ is an appropriately defined constant, $\kappa_2(W)$ (which is 1 for the classical case) is the condition number of the weight matrix W, σ_1 is the smallest singular value of the least squares matrix A, $h_{\max} = \max_{1\le i\le m} \|v_i - a\|$, and n is the degree of the truncated Taylor series given in (2.1).

3. Error bounds. In this section we consider a more general setting than that discussed in section 2 and derive error bounds for both the classical and weighted least squares gradient estimation methods for an $n+1$ term Taylor expansion.

3.1. Classical least squares gradient estimates. The error bound derived in Proposition 3.2 below requires the consideration of the following lemma, which effectively bounds the error in the Taylor series truncated at term n.

LEMMA 3.1. *Let $f : D \subset R^2 \to R$ be of class C^n, $a \in D$, and $\mathcal{B}(a) \subset D$ some neighborhood of a. Suppose that for $i = 0, 1, \ldots, n$ we have $\frac{\partial^n f}{\partial x^{n-i}\partial y^i} \in Lip_{\vartheta_i}(\mathcal{B}(a))$ with $\vartheta_{\max} = \max_{0\le i\le n} \vartheta_i$. Then for any $a + h\nu \in \mathcal{B}(a)$ with $\|\nu\| = 1$*

$$\left| \sum_{k=1}^{n} \frac{h^{k-1}}{k!} (\nu \cdot \nabla)^k f(a) - \left\{ \frac{f(a+h\nu) - f(a)}{h} \right\} \right| \le \frac{h^n}{(n+1)!} \vartheta_{\max} \|\nu\|_1^n. \tag{3.1}$$

Proof. Rearranging the multivariable Taylor series (2.1) for f about the point a, we obtain

$$\frac{f(a+\nu h) - f(a)}{h} - \sum_{k=1}^{n-1} \frac{h^{k-1}}{k!} (\nu\cdot\nabla)^k f(a) = \frac{h^{n-1}}{(n-1)!} \int_0^1 (1-t)^{n-1} (\nu\cdot\nabla)^n f(a+th\nu) dt,$$

and therefore

$$\begin{aligned}
&\frac{h^{n-1}}{n!} (\nu\cdot\nabla)^n f(a) - \left\{ \frac{f(a+h\nu) - f(a)}{h} \right\} + \sum_{k=1}^{n-1} \frac{h^{k-1}}{k!} (\nu\cdot\nabla)^k f(a) \\
&= \frac{h^{n-1}}{(n-1)!} \int_0^1 (1-t)^{n-1} \Big\{ (\nu\cdot\nabla)^n f(a) - (\nu\cdot\nabla)^n f(a+th\nu) \Big\} dt \\
&= \frac{h^{n-1}}{(n-1)!} \sum_{i=0}^{n} \binom{n}{i} \nu_x^{n-i} \nu_y^i \int_0^1 (1-t)^{n-1} \left\{ \frac{\partial^n f(a)}{\partial x^{n-i}\partial y^i} - \frac{\partial^n f(a+th\nu)}{\partial x^{n-i}\partial y^i} \right\} dt.
\end{aligned}$$

Hence,

$$\begin{aligned}
&\left| \sum_{k=1}^{n} \frac{h^{k-1}}{k!} (\nu\cdot\nabla)^k f(a) - \left\{ \frac{f(a+h\nu) - f(a)}{h} \right\} \right| \\
&= \left| \frac{h^{n-1}}{(n-1)!} \sum_{i=0}^{n} \binom{n}{i} \nu_x^{n-i} \nu_y^i \int_0^1 (1-t)^{n-1} \left\{ \frac{\partial^n f(a)}{\partial x^{n-i}\partial y^i} - \frac{\partial^n f(a+th\nu)}{\partial x^{n-i}\partial y^i} \right\} dt \right| \\
&\le \frac{h^{n-1}}{(n-1)!} \sum_{i=0}^{n} \binom{n}{i} |\nu_x|^{n-i} |\nu_y|^i \int_0^1 |1-t|^{n-1} \left| \frac{\partial^n f(a)}{\partial x^{n-i}\partial y^i} - \frac{\partial^n f(a+th\nu)}{\partial x^{n-i}\partial y^i} \right| dt.
\end{aligned} \tag{3.2}$$

We now invoke the Lipschitz continuity of the mixed partial derivatives to obtain a further inequality on (3.2) as

$$\begin{aligned}
&\leq \frac{h^{n-1}}{(n-1)!}\sum_{i=0}^{n}\left(\begin{array}{c} n \\ i \end{array}\right)|\nu_x|^{n-i}|\nu_y|^i\vartheta_i\int_0^1|1-t|^{n-1}\|a-a-th\nu\|dt \\
&\leq \frac{h^n}{(n+1)!}\vartheta_{\max}\sum_{i=0}^{n}\left(\begin{array}{c} n \\ i \end{array}\right)|\nu_x|^{n-i}|\nu_y|^i \\
&= \frac{h^n}{(n+1)!}\vartheta_{\max}\Big\{|\nu_x|+|\nu_y|\Big\}^n = \frac{h^n}{(n+1)!}\vartheta_{\max}\|\nu\|_1^n. \qquad \square
\end{aligned}$$

We are now in a position to prove the main result for this section, which provides a bound on the error in the classical least squares gradient estimate.

PROPOSITION 3.2. *Let $f : D \subset R^2 \to R$ be of class C^n, $a \in D$, and $\mathcal{B}(a) \subset D$ be some neighborhood of a. Suppose around point a we have m neighboring points $v_k, k = 1,\dots,m$, with $a, v_1,\dots,v_m \in \mathcal{B}(a) \subset D$. Suppose further that $\frac{\partial^n f}{\partial x^{n-i}\partial y^i} \in Lip_{\vartheta_i}(\mathcal{B}(a)), i = 0,1,\dots,n$, and we approximate the gradient locally at a by $E_1\hat{\gamma}$ via the least squares solution $\hat{\gamma} = \arg\min_{\gamma\in R^p}\|A\gamma - q\|$, where*

$$A = \begin{pmatrix} \nu_{11}^T & \nu_{12}^T \\ \nu_{21}^T & \nu_{22}^T \\ \vdots & \vdots \\ \nu_{m1}^T & \nu_{m2}^T \end{pmatrix} \in \mathbb{R}^{m\times p}, \quad q = \begin{pmatrix} \frac{f(a+h_1\nu_1)-f(a)}{h_1} \\ \frac{f(a+h_2\nu_2)-f(a)}{h_2} \\ \vdots \\ \frac{f(a+h_m\nu_m)-f(a)}{h_m} \end{pmatrix} \in \mathbb{R}^{m\times 1},$$

$E_1 = \left(\begin{smallmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \end{smallmatrix}\right) \in R^{2\times p}$, $h_k = \|v_k - a\|$ with $h_k\nu_k = v_k - a$; $\nu_{k1}^T = (\nu_{x_k}, \nu_{y_k})$, $\nu_{k2}^T = (\frac{h_k}{2}\nu_{x_k}^2, h_k\nu_{x_k}\nu_{y_k}, \frac{h_k}{2}\nu_{y_k}^2, \frac{h_k^2}{6}\nu_{x_k}^3, \frac{h_k^2}{2}\nu_{x_k}^2\nu_{y_k}, \dots, \frac{h_k^{n-1}}{n!}\nu_{y_k}^n)$, and $p = \frac{(n+1)(n+2)}{2} - 1$. Then a bound on the error in the least squares gradient estimate is given by

$$\|\nabla f(a) - E_1\hat{\gamma}\| \leq \frac{\vartheta_{\max}h_{\max}^n}{\sigma_1(n+1)!}\sqrt{\sum_{i=1}^{m}\|\nu_i\|_1^{2n}}, \tag{3.3}$$

where σ_1 is the smallest singular value of A, which is assumed to have $\operatorname{rank}(A) = p$, *$\vartheta_{\max}$ is as defined in Lemma* 3.1, *and $h_{\max} = \max_{1\leq k\leq m} h_k$.*

Proof. Let $E_2 \in R^{(p-2)\times p}$ be the last $p-2$ rows of the identity matrix I_p and $U = (\frac{\partial f}{\partial x}(a), \frac{\partial f}{\partial y}(a), \frac{\partial^2 f}{\partial x^2}(a), \frac{\partial^2 f}{\partial x\partial y}(a), \frac{\partial^2 f}{\partial y^2}(a), \frac{\partial^3 f}{\partial x^3}(a), \frac{\partial^3 f}{\partial x^2\partial y}(a), \dots \frac{\partial^n f}{\partial y^n}(a))^T \in \mathbb{R}^{p\times 1}$. Now $U - \hat{\gamma}$ can be partitioned as $\left(\begin{smallmatrix} E_1(U-\hat{\gamma}) \\ E_2(U-\hat{\gamma}) \end{smallmatrix}\right)$, and hence

$$\|U-\hat{\gamma}\|^2 = \|E_1(U-\hat{\gamma})\|^2 + \|E_2(U-\hat{\gamma})\|^2 \geq \|E_1(U-\hat{\gamma})\|^2 = \|\nabla f(a) - E_1\hat{\gamma}\|^2.$$

Next, with $\hat{\gamma} = A^\dagger q$, we have

$$\|U-\hat{\gamma}\|^2 = \|U - A^\dagger q\|^2 = \|A^\dagger(AU-q)\|^2 \leq \|A^\dagger\|^2\|AU-q\|^2 = \frac{1}{\sigma_1^2}\|AU-q\|^2.$$

Now using the result in Lemma 3.1, the following upper bound can be derived:

$$\begin{aligned}
\|AU - q\|^2 &= \sum_{i=1}^{m} \left| \sum_{k=1}^{n} \frac{h_i^{k-1}}{k!} (\nu_i \cdot \nabla)^k f(a) - \left\{ \frac{f(a + h_i\nu_i) - f(a)}{h_i} \right\} \right|^2 \\
&\le \sum_{i=1}^{m} \left(\frac{h_i^n}{(n+1)!} \vartheta_{\max} \|\nu_i\|_1^n \right)^2 \qquad \text{(using Lemma 3.1)} \\
&= \left(\frac{\vartheta_{\max}}{(n+1)!} \right)^2 \sum_{i=1}^{m} \left(h_i^n \|\nu_i\|_1^n \right)^2 \qquad (3.4) \\
&\le \left(\frac{\vartheta_{\max}}{(n+1)!} \right)^2 \sum_{i=1}^{m} \left(h_{\max}^n \|\nu_i\|_1^n \right)^2 \\
&= \left(\frac{\vartheta_{\max}}{(n+1)!} \right)^2 \left(h_{\max}^n \right)^2 \sum_{i=1}^{m} \|\nu_i\|_1^{2n}.
\end{aligned}$$

The result follows from

$$\|\nabla f(a) - E_1\hat{\gamma}\| \le \|U - \hat{\gamma}\| \le \frac{\vartheta_{\max}}{(n+1)!} \frac{h_{\max}^n}{\sigma_1} \sqrt{\sum_{i=1}^{m} \|\nu_i\|_1^{2n}}. \qquad \square$$

3.2. Weighted least squares gradient estimates. The bound presented in the previous section can be naturally extended to the weighted least squares case via the following proposition.

PROPOSITION 3.3. *Let $f : D \subset R^2 \to R$ be of class C^n, $a \in D$, and $\mathcal{B}(a) \subset D$ be some neighborhood of a. Suppose around point a we have m neighboring points $v_k, k = 1, 2, \ldots, m$, with $a, v_1, \ldots, v_m \in \mathcal{B}(a)$. Suppose further that $\frac{\partial^n f}{\partial x^{n-i} \partial y^i} \in Lip_{\vartheta_i}(\mathcal{B}(a))$, $i = 0, 1, 2, \ldots, n$, and we approximate the gradient locally at a by $E_1\tilde{\gamma}$ via the weighted least squares solution $\tilde{\gamma} = \arg\min_{\gamma \in R^p} \|\bar{A}\gamma - Wq\|$, where $\bar{A} = WA$ with $W = \mathrm{diag}(w_1, w_2, \ldots, w_m) \in R^{m \times m}$ being the weight matrix with positive weights w_i. Then, with the same notation as given in Proposition 3.2 and $w_{\max} = \max_{1 \le k \le m} w_k$, we have that a bound on the error in the weighted least squares gradient estimate is given by*

$$\|\nabla f(a) - E_1\tilde{\gamma}\| \le \frac{\vartheta_{\max} h_{\max}^n w_{\max}}{\bar{\sigma}_1 (n+1)!} \sqrt{\sum_{i=1}^{m} \|\nu\|_1^{2n}}, \qquad (3.5)$$

where $\bar{\sigma}_1$ is the smallest nonzero singular value of $\bar{A}$, which is assumed to have $\mathrm{rank}(\bar{A}) = p$.

Proof. The proof proceeds along the same lines of reasoning as that of Proposition 3.2, now with

$$\|U - \tilde{\gamma}\|^2 \ge \|\nabla f(a) - E_1 \bar{A}^\dagger W q\|^2$$

and

$$\begin{aligned}
\|U - \tilde{\gamma}\|^2 &= \|U - \bar{A}^\dagger W q\|^2 \\
&= \|\bar{A}^\dagger (\bar{A}U - Wq)\|^2 \\
&\le \|\bar{A}^\dagger\|^2 \|\bar{A}U - Wq\|^2 \\
&\le \|\bar{A}^\dagger\|^2 \|W\|^2 \|AU - q\|^2 \\
&= \frac{1}{\bar{\sigma}_1^2} \|W\|^2 \|AU - q\|^2.
\end{aligned}$$

Using inequality (3.4) and noting that $\|W\| = \max_{1\le k\le m} w_k = w_{\max}$ gives

$$\|\nabla f(a) - E_1\tilde{\gamma}\| \le \|U - \tilde{\gamma}\| \le \frac{\vartheta_{\max}}{(n+1)!}\frac{h_{\max}^n w_{\max}}{\bar{\sigma}_1}\sqrt{\sum_{i=1}^{m}\|\nu_i\|_1^{2n}}. \qquad \square$$

Remark 1. Noting that $\sigma_1 = \inf_{\|x\|=1}\|Ax\| \le \|W^{-1}\|\inf_{\|x\|=1}\|\bar{A}x\| = \|W^{-1}\|\bar{\sigma}_1$ the bound (3.5) can be also written in terms of the condition number of the weight matrix as follows:

$$\|\nabla f(a) - E_1\tilde{\gamma}\| \le \frac{\vartheta_{\max}h_{\max}^n\kappa_2(W)}{\sigma_1(n+1)!}\sqrt{\sum_{i=1}^{m}\|\nu\|_1^{2n}}. \tag{3.6}$$

Remark 2. It appears that rank deficiency is unlikely to occur, and we have not encountered any situations of rank deficiency throughout our numerical experimentation. We note further that there appears to be a lack of any criterion available in the literature that would ensure that the least squares matrix A has full rank. Further investigation of a suitable criterion will be a topic of future research to be undertaken by the authors.

4. Tighter error bounds. It will be apparent from the numerical experimentation performed in section 5 that the error bounds given in section 3 are somewhat pessimistic. To derive tighter bounds we return to the earlier work by Belward, Turner, and Ilić [3], where it was observed that the gradient $\nabla f(a)$ could be estimated using an orthogonal reduction of the matrix A_2 in the partitioned matrix $A = (A_1|A_2)$, $A_1 \in R^{m\times 2}$ and $A_2 \in R^{m\times(p-2)}$, as follows:

1. $Q^T A_2 = \begin{pmatrix} A_{12} \\ 0 \end{pmatrix}$ with $A_{12} \in R^{(p-2)\times(p-2)}$ upper triangular.
2. Partitioning $Q^T = \begin{pmatrix} Q_1^T \\ Q_2^T \end{pmatrix}$ yields $Q^T A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & 0 \end{pmatrix}$, with $A_{11} = Q_1^T A_1, A_{12} = Q_1^T A_2$, and $A_{21} = Q_2^T A_1$.
3. Then $\nabla f(a) \approx \arg\min_{g\in R^2} \left\|A_{21}g - Q_2^T q\right\|$.

It was proven in [3] that this strategy produces the same least squares error and gradient estimate as the direct approach described in section 3. Interestingly though, it is the orthogonal reduction approach that enables tighter bounds than those given in (3.3) and (3.5) to be derived. We begin with the following proposition.

PROPOSITION 4.1. *Let $A \in R^{m\times p}$ as defined above with* $\operatorname{rank}(A) = p$ *be column-partitioned as $A = (A_1|A_2)$ with $Q^T A_2 = \begin{pmatrix} A_{12} \\ 0 \end{pmatrix}$ being the orthogonal reduction of A_2, so that $Q^T A = \begin{pmatrix} Q_1^T \\ Q_2^T \end{pmatrix}(A_1|A_2) = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & 0 \end{pmatrix}$. Then the following hold.*

(i) *A_{21} has full column rank with* $\operatorname{rank}(A_{21}) = \operatorname{rank}(A_1)$.

(ii) *$\sigma_1 \le \hat{\sigma}_1$, where σ_1 is the smallest singular value of A and $\hat{\sigma}_1$ is the smallest singular value of A_{21}.*

Proof. (i) $\operatorname{rank}(A_{21}) = \operatorname{rank}(Q_2^T A_1) = \operatorname{rank}(A_1) - \dim \mathcal{N}(Q_2^T) \cap \mathcal{R}(A_1)$ (see, for example, [14]). Now $Q_2^T A_2 = 0 \Rightarrow \mathcal{R}(A_2) \subseteq \mathcal{N}(Q_2^T)$. Let $z \in \mathcal{N}(Q_2^T)$ and observe that $Q^T(A_2|z) = \begin{pmatrix} A_{12} & Q_1^T z \\ 0 & 0 \end{pmatrix}$ so that $\operatorname{rank}(A_2|z) = \operatorname{rank}(A_2) \Rightarrow z \in \mathcal{R}(A_2)$, i.e., $\mathcal{N}(Q_2^T) \subseteq \mathcal{R}(A_2)$, and hence it follows that $\mathcal{N}(Q_2^T) = \mathcal{R}(A_2)$. Thus, $\mathcal{N}(Q_2^T) \cap \mathcal{R}(A_1) = \emptyset$, since $\operatorname{rank}(A) = p$, and therefore $\operatorname{rank}(A_{21}) = \operatorname{rank}(A_1) \Rightarrow A_{21}$ has full column rank because any subset of a linearly independent set of column vectors must be also linearly independent.

(ii) $\sigma_1^2 = \inf_{\|x\|\neq 0} \frac{\|Ax\|^2}{\|x\|^2} = \inf_{\|x\|\neq 0} \frac{\|Q^T Ax\|^2}{\|x\|^2} = \inf_{\|x\|\neq 0} \frac{\|\left(\begin{smallmatrix} A_{11} & A_{12} \\ A_{21} & 0 \end{smallmatrix}\right)\left(\begin{smallmatrix} x_1 \\ x_2 \end{smallmatrix}\right)\|^2}{\|x_1\|^2+\|x_2\|^2}$. In particular, choose $x_1 = v_1$, $x_2 = -A_{12}^{-1}A_{11}v_1$, where $(\hat{\sigma}_1, u_1, v_1)$ is the singular triplet of A_{21} corresponding to the smallest singular value $\hat{\sigma}_1$. Hence, $\sigma_1 \leq \frac{\hat{\sigma}_1}{\sqrt{1+\|x_2\|^2}} \leq \hat{\sigma_1}$. ∎

PROPOSITION 4.2. *Under the hypotheses of Propositions* 3.2 *and* 4.1 *we have that*

$$\begin{aligned} \|\nabla f(a) - E_1 A^\dagger q\| &\leq \frac{\vartheta_{\max} h_{\max}^n}{\hat{\sigma}_1 (n+1)!} \sqrt{\sum_{i=1}^m \|\nu_i\|_1^{2n}} \\ &\leq \frac{\vartheta_{\max} h_{\max}^n}{\sigma_1 (n+1)!} \sqrt{\sum_{i=1}^m \|\nu_i\|_1^{2n}}. \end{aligned} \tag{4.1}$$

Proof.

$$\begin{aligned} \|\nabla f(a) - E_1 A^\dagger q\| &= \|\nabla f(a) - A_{21}^\dagger Q_2^T q\| \qquad \text{see [3] for proof} \\ &= \|A_{21}^\dagger (A_{21}\nabla f(a) - Q_2^T q)\| \quad \text{using Proposition 4.1(i)} \\ &\leq \frac{1}{\hat{\sigma}_1} \|Q_2^T (A_1 \nabla f(a) - q)\| \\ &= \frac{1}{\hat{\sigma}_1} \|Q_2^T \{(A_1|A_2)U - q\}\| \quad \text{using } Q_2^T A_2 = 0 \\ &\leq \frac{1}{\hat{\sigma}_1} \|AU - q\| \qquad \text{using the result } \|Q_2^T u\| \leq \|u\| \\ &\leq \frac{\vartheta_{\max} h_{\max}^n}{\hat{\sigma}_1 (n+1)!} \sqrt{\sum_{i=1}^m \|\nu_i\|_1^{2n}}. \end{aligned}$$

The second inequality holds from Proposition 4.1(ii) since $\sigma_1 \leq \hat{\sigma}_1 \Rightarrow \frac{1}{\hat{\sigma}_1} \leq \frac{1}{\sigma_1}$. ∎

All of the above results also hold for the weighted least squares problem given in section 3.2 where the orthogonal decomposition is now performed on WA_2 so that $\widetilde{Q}^T W A_2 = \left(\begin{smallmatrix} \widetilde{A}_{12} \\ 0 \end{smallmatrix}\right)$, where $\widetilde{A}_{12}$ is upper triangular. In this case $\widetilde{Q}^T (WA_1|WA_2) = \left(\begin{smallmatrix} \widetilde{A}_{11} & \widetilde{A}_{12} \\ \widetilde{A}_{21} & 0 \end{smallmatrix}\right)$, and we have the following proposition.

PROPOSITION 4.3. *Under the hypotheses of Proposition* 3.3 *and the extension of Proposition* 4.1 *to the weighted least squares case for* (i) $\widetilde{A}_{21}$ *has full column rank and* (ii) $\bar{\sigma}_1 \leq \tilde{\sigma}_1$, *where* $\tilde{\sigma}_1$ *is the smallest singular value of* $\widetilde{A}_{21}$, *we have that*

$$\begin{aligned} \|\nabla f(a) - E_1 \bar{A}^\dagger W q\| &\leq \frac{\vartheta_{\max} h_{\max}^n w_{\max}}{\widetilde{\sigma}_1 (n+1)!} \sqrt{\sum_{i=1}^m \|\nu_i\|_1^{2n}} \\ &\leq \frac{\vartheta_{\max} h_{\max}^n w_{\max}}{\bar{\sigma}_1 (n+1)!} \sqrt{\sum_{i=1}^m \|\nu_i\|_1^{2n}}. \end{aligned} \tag{4.2}$$

Proof.

$$\begin{aligned}
\|\nabla f(a) - E_1 \widetilde{A}^{\dagger} W q\| &= \|\nabla f(a) - \widetilde{A}_{21}^{\dagger} \widetilde{Q}_2^T W q\| \\
&= \|\widetilde{A}_{21}^{\dagger}(\widetilde{A}_{21}\nabla f(a) - \widetilde{Q}_2^T W q)\| \\
&\leq \frac{1}{\widetilde{\sigma}_1}\|\widetilde{Q}_2^T(WA_1\nabla f(a) - Wq)\| \\
&= \frac{1}{\widetilde{\sigma}_1}\|\widetilde{Q}_2^T\{(WA_1|WA_2)U - Wq\}\| \\
&\leq \frac{1}{\widetilde{\sigma}_1}\|W(A\nabla f(a) - q)\| \\
&\leq \frac{\|W\|}{\widetilde{\sigma}_1}\|A\nabla f(a) - q\| \\
&\leq \frac{\vartheta_{\max} h_{\max}^n w_{\max}}{\widetilde{\sigma}_1 (n+1)!}\sqrt{\sum_{i=1}^{m}\|\nu_i\|_1^{2n}}.
\end{aligned}$$

The second inequality holds since $\frac{1}{\widetilde{\sigma}_1} \leq \frac{1}{\bar{\sigma}_1}$. ∎

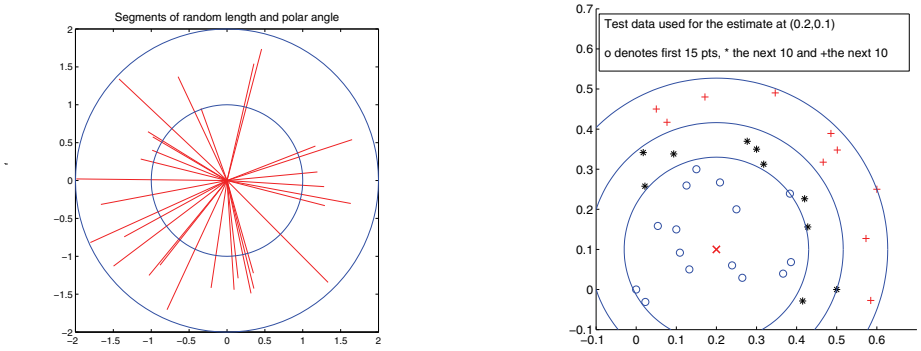
Remark 3. Proposition 3.2 appears at first to imply that the method has nth power accuracy in h; however, since the singular values of the least squares matrix A depend on h, this conclusion is false. The numerical results presented in [3] do exhibit a quadratic rate of convergence, and this is implied by Proposition 4.2. This follows because the singular values of the reduced matrix A_{21} are independent of h, being an orthogonal transformation of the first two columns of the least squares matrix A, which themselves are independent of h.

5. Numerical experiments. Two sets of numerical experiments are described in this section with the purpose of examining the veracity and utility of the error bounds derived in sections 3–4. All results appearing below exhibit relative errors in which the errors defined in the earlier sections are factored by $\|\nabla f(a)\|$. First, a set of results is described that confirms the error estimates and asymptotic behavior of the error as the test points move toward the data point in question. Then a contrasting situation is described with scattered data points taken from tests conducted by Franke [8]. The following three functions were chosen from Franke's set:

$$\begin{aligned}
F_1(x,y) &= \frac{1.25 + \cos(5.4y)}{6(1 + (3x-1)^2)}; \\
F_2(x,y) &= \frac{e^{-\frac{81}{16}((x-0.5)^2 + (y-0.5)^2)}}{3}; \\
F_3(x,y) &= \frac{\sqrt{64 - 81((x-0.5)^2 + (y-0.5)^2)}}{9} - 0.5.
\end{aligned}$$

5.1. Asymptotic results. A set of test points was generated by a pseudo-random selection of radial distances and polar angles. The polar angles were fixed and the radial distances were varied by scaling down in factors of 10.

Thirty points $(r_i \cos\theta_i, r_i \sin\theta_i)^T$ were generated so that $\{r_i\}$ and $\{\theta_i\}$ were distributed uniformly with $1 \leq r_i \leq 2$ and $0 \leq \theta_i \leq 2\pi$. From these points the 14 points nearest the origin were chosen and used as displacements from the point $a = (.2, .1)^T$. The inner radius of the annulus was scaled from 0.25 down to 2.5×10^{-6}. By this construction the angular distribution of points was unchanged as their distances from



(a) (b)

FIG. 5.1. (a) *Randomly generated direction vectors with magnitudes in the interval* $[1, 2]$. (b) *Sets of the* 15*,* 25*, and* 35 *data points closest to* $(0.2, 0.1)$.

the test point was varied and the data points were not permitted to lie favorably close to the test point. Figure 5.1(a) shows the segments used in the current tests. Earlier tests reported in [4] show little variation in the accuracy obtained as the number of test points is varied from their base level.

In addition to varying the distance between the data points and the evaluation point, the weighting of the data points by inverse powers of their distances was also investigated with exponents from -1 to -4.

The Lipschitz constant $\vartheta_{\max}$ in the error bounds can be estimated by an application of the mean value theorem to the mixed partial derivatives in the Taylor series remainder as

$$\vartheta_{\max} = \sqrt{2} \max_{\xi \in D} \left(\left| \frac{\partial^{n+1} f(\xi)}{\partial x^{n-i+1} \partial y^i} \right|, i = 0, \ldots, n \right).$$

The appearance of the Lipschitz constants in the error estimates presents a dilemma regarding the objectivity of the tests. A global estimate can lead to pessimistic error bounds, while an estimate local to a data point will cause fluctuation in the bounds. Here we have used global estimates based on derivative values determined using the software package Maple. Some of the estimates that use higher order derivatives cause the results to be undesirably large. Calculations for individual points can easily be made if individual local estimates are sought.

The results of our numerical experimentation are shown in Tables 5.1–5.3 for varying radii and in Tables 5.4–5.6 for the various weightings at the radius 2.5×10^{-3} for each of the test functions F_1, F_2, and F_3. Tables 5.1–5.3 exhibit the relative error in the gradient estimates at the test point a compared with $Bound_1$ from Proposition 3.2 and $Bound_2$ from Proposition 4.2 for the classical least squares estimates, while Tables 5.4–5.6 exhibit the relative error compared to $Bound_1$ from Proposition 3.3 and $Bound_2$ from Proposition 4.3 for the weighted case. The results are also plotted on logarithmic scales in Figure 5.2 for the quadratic estimates and Figure 5.3 for the cubic gradient estimates. The bounds given in Propositions 4.2 and 4.3 are clearly better than those in Propositions 3.2 and 3.3 for all cases considered.

TABLE 5.1

Asymptotic behavior of the relative error and the error bounds for the function F_1 using second and third order least squares gradient estimates.

	Second order			Third order		
Radius	Rel. error	$Bound_1$	$Bound_2$	Rel. error	$Bound_1$	$Bound_2$
2.5e-1	4.8981e-01	4.9819e+01	2.6869e+00	1.4095e-01	1.6826e+03	1.8366e+00
2.5e-2	7.4944e-03	4.9494e+00	2.6869e-02	2.1657e-04	1.6728e+02	1.8366e-03
2.5e-3	7.5058e-05	4.9491e-01	2.6869e-04	2.2999e-07	1.6727e+01	1.8366e-06
2.5e-4	7.5038e-07	4.9491e-02	2.6869e-06	2.3235e-10	1.6727e+00	1.8366e-09
2.5e-5	7.5037e-09	4.9491e-03	2.6869e-08	6.8197e-12	1.6727e-01	1.8366e-12

TABLE 5.2

Asymptotic behavior of the relative error and the error bounds for the function F_2 using second and third order least squares gradient estimates.

	Second order			Third order		
Radius	Rel. error	$Bound_1$	$Bound_2$	Rel. error	$Bound_1$	$Bound_2$
2.5e-1	2.3760e-01	2.4069e+01	1.2981e+00	4.1537e-02	5.3435e+02	5.8325e-01
2.5e-2	2.6968e-03	2.3912e+00	1.2981e-02	4.5445e-05	5.3125e+01	5.8325e-04
2.5e-3	2.6960e-05	2.3910e-01	1.2981e-04	4.5946e-08	5.3120e+00	5.8325e-07
2.5e-4	2.6956e-07	2.3910e-02	1.2981e-06	4.5712e-11	5.3120e-01	5.8325e-10
2.5e-5	2.6977e-09	2.3910e-03	1.2981e-08	2.5271e-12	5.3120e-02	5.8325e-13

TABLE 5.3

Asymptotic behavior of the relative error and the error bounds for the function F_3 using second and third order least squares gradient estimates.

	Second order			Third order		
Radius	Rel. error	$Bound_1$	$Bound_2$	Rel. error	$Bound_1$	$Bound_2$
2.5e-1	7.1965e-02	2.0562e+01	1.1090e+00	8.5304e-02	1.1333e+03	1.2370e+00
2.5e-2	7.2914e-04	2.0428e+00	1.1090e-02	2.5560e-05	1.1266e+02	1.2370e-03
2.5e-3	7.3165e-06	2.0426e-01	1.1090e-04	2.3839e-08	1.1266e+01	1.2370e-06
2.5e-4	7.3197e-08	2.0426e-02	1.1090e-06	2.5055e-11	1.1266e+00	1.2370e-09
2.5e-5	7.3197e-10	2.0426e-03	1.1090e-08	9.8809e-12	1.1266e-01	1.2370e-12

Tables 5.1–5.3 and the corresponding plots show that as the data points move closer to the evaluation point a, the superiority of the gradient estimates becomes more marked. We conjecture that this is due to the diminution of the columns of the full matrix from column 3 and beyond as h is reduced. Note also that the slight inconsistency of the bounds in Tables 5.1–5.3 for the cubic gradient estimates at radius 2.5×10^{-5} is thought to be due to roundoff error.

Two further observations that can be gleaned from the results for the three test functions are that cubic gradient estimates are more accurate than the quadratic estimates and that weighting appears to offer very little improvement in accuracy. Moreover, as a result of the scaling by the condition number of the weight matrix, the bounds are less spectacular for the weighted least squares gradient estimates. In fact, one might speculate that the insight gained from the derived error bounds may well indicate that weighted least squares, in this context of gradient estimation, is not beneficial as a direct consequence of this scaling.

5.2. Scattered data results. To assess the accuracy of the error bounds for a scattered data set that is more likely to arise in practice, we chose the 133 data points from Franke [8]. The point of interest was chosen as $a = (.2, .1)^T$ and subsets of $10, 15, \ldots, 35$ points, those being the closest to a, were used. Figure 5.1(b) shows

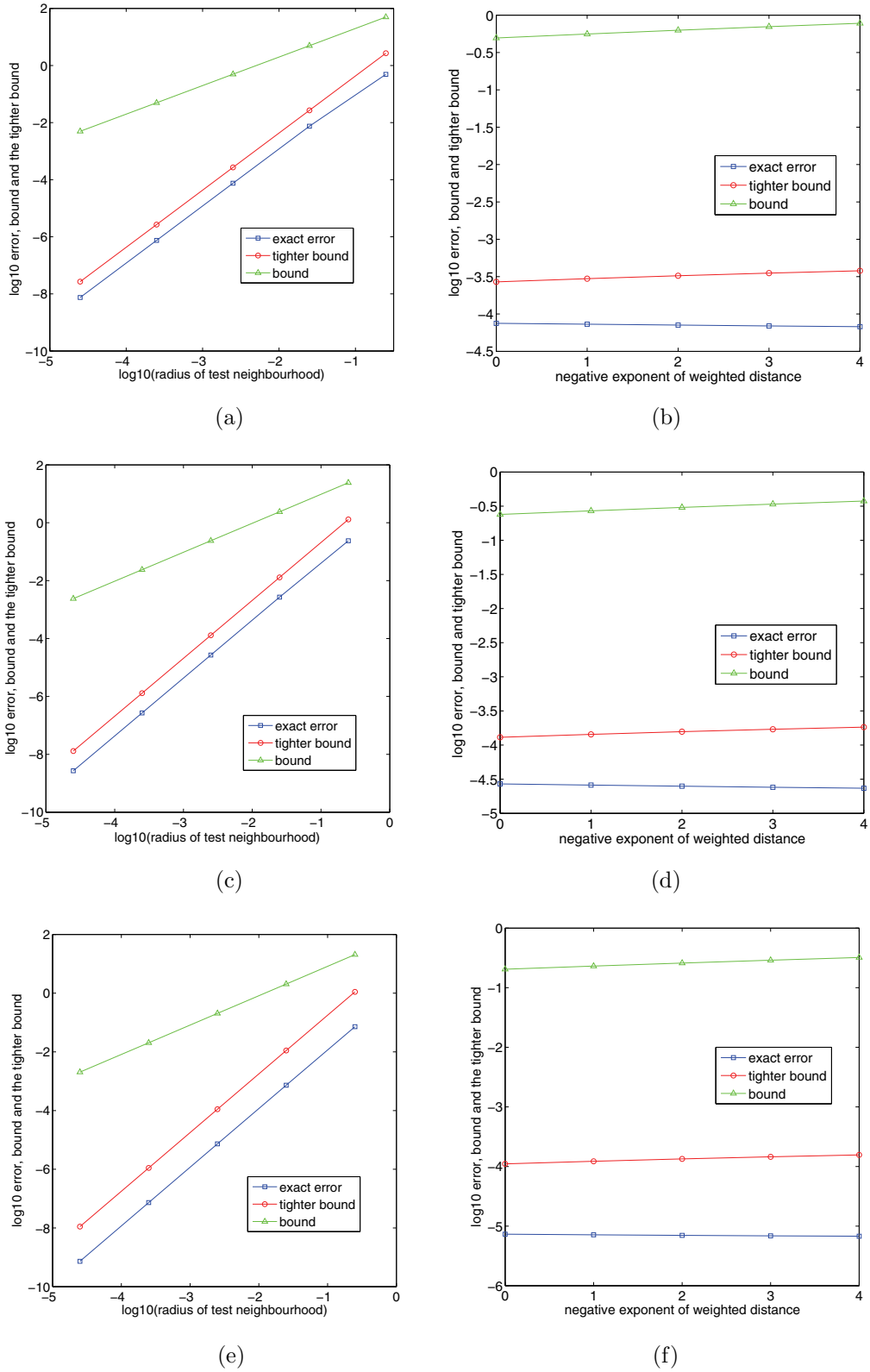


FIG. 5.2. *Second order least squares errors and error bounds;* (a) *and* (b) *for function* F_1, (c) *and* (d) *for function* F_2, (e) *and* (f) *for function* F_3. *Varying radius on the left; varying weightings at radius* 2.5e-03 *on the right.*

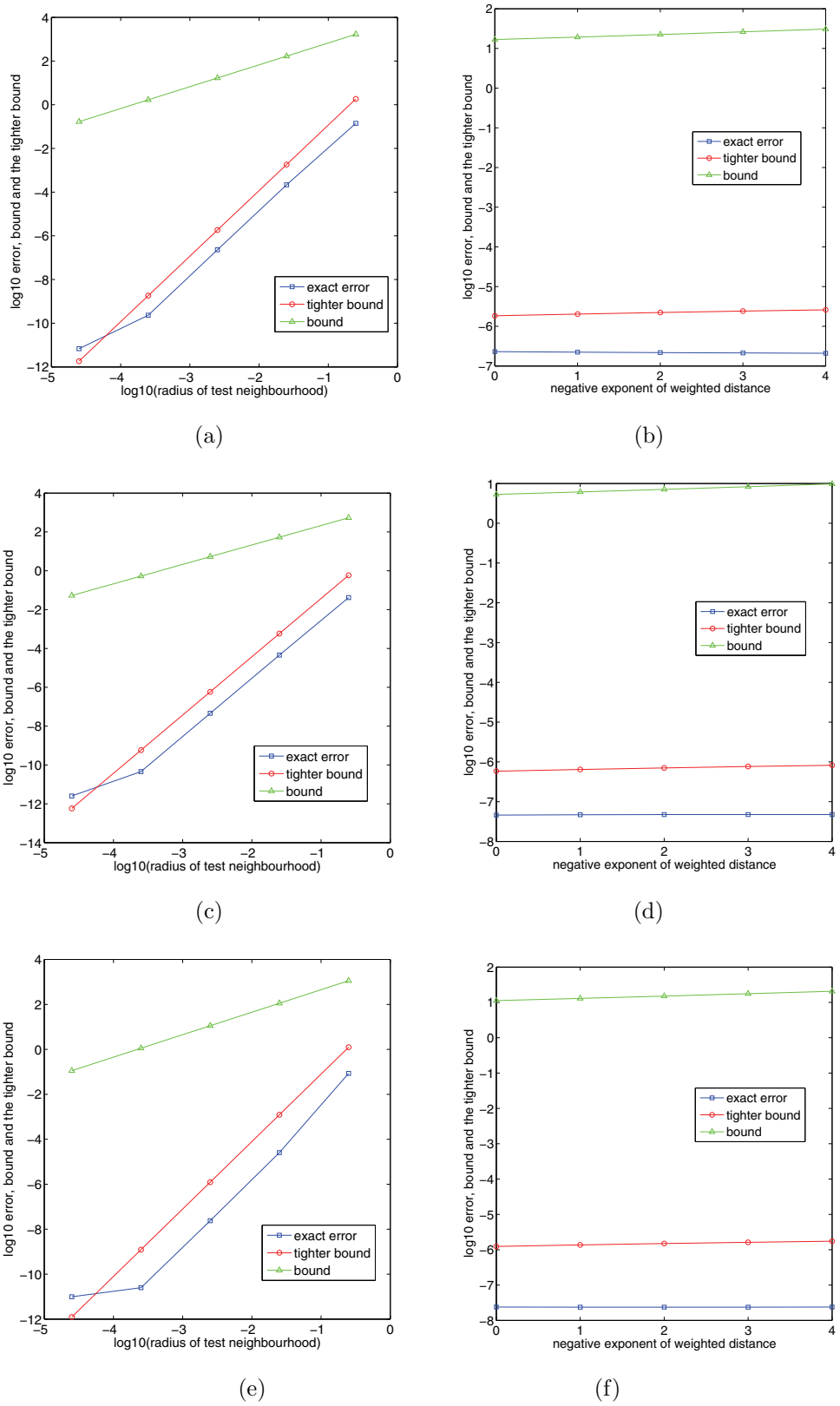


FIG. 5.3. *Third order least squares errors and error bounds;* (a) *and* (b) *for* F_1, (c) *and* (d) *for* F_2, (e) *and* (f) *for* F_3. *Varying radius on the left; varying weightings at radius* 2.5e-03 *on the right.*

TABLE 5.4
Weighted asymptotic behavior at radius 2.5e-3 *of the relative error and the error bounds for the function F_1 using weighted second and third order least squares gradient estimates.*

Weight	Second order			Third order		
exponent	Rel. error	$Bound_1$	$Bound_2$	Rel. error	$Bound_1$	$Bound_2$
0	7.5058e-05	4.9491e-01	2.6869e-04	2.2999e-07	1.6727e+01	1.8366e-06
-1	7.3037e-05	5.5877e-01	2.9682e-04	2.2310e-07	1.9309e+01	2.0289e-06
-2	7.1068e-05	6.2744e-01	3.2504e-04	2.1729e-07	2.2420e+01	2.2218e-06
-3	6.9239e-05	7.0060e-01	3.5272e-04	2.1260e-07	2.6169e+01	2.4110e-06
-4	6.7616e-05	7.7780e-01	3.7894e-04	2.0907e-07	3.0684e+01	2.5903e-06

TABLE 5.5
Weighted asymptotic behavior at radius 2.5e-3 *of the relative error and the error bounds for the function F_2 using weighted second and third order least squares gradient estimates.*

Weight	Second order			Third order		
exponent	Rel. error	$Bound_1$	$Bound_2$	Rel. error	$Bound_1$	$Bound_2$
0	2.6960e-05	2.3910e-01	1.2981e-04	4.5946e-08	5.3120e+00	5.8325e-07
-1	2.5936e-05	2.6996e-01	1.4340e-04	4.6820e-08	6.1320e+00	6.4435e-07
-2	2.4961e-05	3.0314e-01	1.5704e-04	4.7339e-08	7.1200e+00	7.0560e-07
-3	2.4073e-05	3.3848e-01	1.7041e-04	4.7496e-08	8.3104e+00	7.6565e-07
-4	2.3299e-05	3.7577e-01	1.8308e-04	4.7335e-08	9.7445e+00	8.2260e-07

TABLE 5.6
Weighted asymptotic behavior at radius 2.5e-3 *of the relative error and the error bounds for the function F_3 using weighted second and third order least squares gradient estimates.*

Weight	Second order			Third order		
exponent	Rel. error	$Bound_1$	$Bound_2$	Rel. error	$Bound_1$	$Bound_2$
0	7.3165e-06	2.0426e-01	1.1090e-04	2.3839e-08	1.1266e+01	1.2370e-06
-1	7.1454e-06	2.3062e-01	1.2250e-04	2.3663e-08	1.3005e+01	1.3665e-06
-2	6.9909e-06	2.5896e-01	1.3416e-04	2.3631e-08	1.5100e+01	1.4964e-06
-3	6.8628e-06	2.8916e-01	1.4558e-04	2.3715e-08	1.7625e+01	1.6238e-06
-4	6.7686e-06	3.2101e-01	1.5640e-04	2.3896e-08	2.0666e+01	1.7446e-06

the point $(.2, .1)$ marked as $\times$ and the closest points in sets of 15, 25, and 35 bounded by circular arcs centered on $(.2, .1)$. The Lipschitz constants required for the error bounds were determined using the Maple software across the unit square. We remark, as observed earlier in section 5.1, that the constant can become quite large for the third order case, which, as we will see, impacts the performance of the error bounds.

The results of the numerical experimentation are summarized in Tables 5.7–5.12 and the data is plotted on logarithm scales in Figures 5.4–5.6 as $h_{\max}$, or, equivalently, the number of points in the least squares stencil increases. We have compared, as this number of points is increased from 10 to 30 in steps of 5, the relative error for the first, second, and third order classical least squares gradient estimates in Tables 5.7, 5.9, and 5.11 with $Bound_1$ from Proposition 3.2 and $Bound_2$ from Proposition 4.2. The relative error is compared to $Bound_1$ from Proposition 3.3 and $Bound_2$ from Proposition 4.3 for the corresponding weighted cases in Tables 5.8, 5.10, and 5.12. Note that $Bound_2$ is not recorded for the first order estimate because the matrix A in this case has only two columns and the extrapolation method cannot be employed. Initial conclusions drawn from these results are that for function F_1 the quadratic estimate offers around the same accuracy as the linear estimate, but there is improvement for the cubic

estimates; however, for functions F_2 and F_3 there is a steady improvement in accuracy as the order of the truncated Taylor series moves from $n = 1$ through $n = 3$.

It is again evident from the results that $Bound_1$ increases as the order of the estimate increases, which is primarily due to the increase in the size of the Lipschitz constant across the domain for the mixed partial derivatives of the functions considered here. Furthermore, for the quadratic and cubic cases, $Bound_2$ is again superior to $Bound_1$, which is consistent with the findings presented in section 5.1. It is clear that the Franke points are not sufficiently dense for useful estimates to be obtained from $Bound_1$ and $Bound_2$. The threshold at which useful estimates may be obtained is with $h_{\max} \approx .025$, on the basis of the asymptotic results. This suggests a density of one or two orders of magnitude more dense compared with that of Franke's set.

There is only slight improvement in the gradient estimations when the weighted least squares method is employed, and the error bounds are higher than the bounds observed for the classical least squares estimates as a result of the scaling by the condition number of the weight matrix. Overall present evidence suggests that using the classical least squares method with $n = 3$ (cubic accuracy) with around $m = 15$ points in the least squares stencil provides the best gradient estimates for the functions under consideration, but the generalization of this observation requires considerably more experimentation. Interestingly, it can be observed from the plots that using more points than this for the cubic case can lead to an increase in the relative error for functions F_1 and F_3. A plausible explanation for this might be due to the local behavior of these functions away from the test point.

TABLE 5.7

Scattered data comparisons of the relative error and error bounds for the function F_1 using first, second, and third order least squares gradient estimates.

No. of	First order		Second order			Third order		
points	Rel. error	$Bound_1$	Rel. error	$Bound_1$	$Bound_2$	Rel. error	$Bound_1$	$Bound_2$
10	0.16	0.31e1	0.18	0.045e3	0.029e2	0.008	0.285e4	0.037e2
15	0.35	0.35e1	0.21	0.056e3	0.033e2	0.018	0.149e4	0.038e2
20	0.43	0.44e1	0.22	0.085e3	0.068e2	0.025	0.199e4	0.092e2
25	0.60	0.51e1	0.22	0.102e3	0.103e2	0.155	0.227e4	0.117e2
25	0.69	0.55e1	0.23	0.110e3	0.131e2	0.186	0.238e4	0.150e2
30	0.75	0.57e1	0.23	0.105e3	0.149e2	0.201	0.237e4	0.182e2

TABLE 5.8

Scattered data comparisons of the relative error and error bounds for the function F_1 using weighted first, second, and third order least squares gradient estimates.

No. of	First order		Second order			Third order		
points	Rel. error	$Bound_1$	Rel. error	$Bound_1$	$Bound_2$	Rel. error	$Bound_1$	$Bound_2$
10	0.122	0.065e2	0.143	0.111e3	0.062e2	0.008	0.0449e5	0.05e2
15	0.117	0.085e2	0.141	0.155e3	0.071e2	0.009	0.0460e5	0.07e2
20	0.138	0.127e2	0.144	0.273e3	0.145e2	0.002	0.0677e5	0.19e2
25	0.187	0.162e2	0.144	0.371e3	0.219e2	0.069	0.0901e5	0.30e2
30	0.220	0.190e2	0.143	0.438e3	0.284e2	0.096	0.1015e5	0.39e2
35	0.245	0.210e2	0.142	0.452e3	0.332e2	0.109	0.0999e5	0.47e2

TABLE 5.9

Scattered data comparisons of the relative error and error bounds for the function F_2 using first, second, and third order least squares gradient estimates.

No. of	First order		Second order			Third order		
points	Rel. error	$Bound_1$	Rel. error	$Bound_1$	$Bound_2$	Rel. error	$Bound_1$	$Bound_2$
10	0.14	0.24e1	0.05	0.22e2	0.14e1	0.011	0.148e4	0.019e2
15	0.04	0.27e1	0.06	0.28e2	0.16e1	0.013	0.078e4	0.020e2
20	0.04	0.35e1	0.06	0.42e2	0.33e1	0.011	0.104e4	0.048e2
25	0.04	0.40e1	0.06	0.50e2	0.50e1	0.031	0.118e4	0.061e2
30	0.06	0.43e1	0.06	0.54e2	0.64e1	0.033	0.124e4	0.078e2
35	0.08	0.45e1	0.05	0.51e2	0.73e1	0.039	0.124e4	0.095e2

TABLE 5.10

Scattered data comparisons of the relative error and error bounds for the function F_2 using weighted first, second, and third order least squares gradient estimates.

No. of	First order		Second order			Third order		
points	Rel. error	$Bound_1$	Rel. error	$Bound_1$	$Bound_2$	Rel. error	$Bound_1$	$Bound_2$
10	0.097	0.051e2	0.043	0.055e3	0.030e2	0.010	0.2335e4	0.03e2
15	0.068	0.066e2	0.048	0.076e3	0.035e2	0.015	0.2391e4	0.04e2
20	0.058	0.099e2	0.050	0.134e3	0.071e2	0.015	0.3521e4	0.10e2
25	0.059	0.127e2	0.060	0.182e3	0.107e2	0.009	0.4687e4	0.16e2
30	0.061	0.148e2	0.067	0.214e3	0.139e2	0.008	0.5283e4	0.21e2
35	0.067	0.164e2	0.069	0.221e3	0.163e2	0.011	0.5199e4	0.25e2

TABLE 5.11

Scattered data comparisons of the relative error and error bounds for the function F_3 using first, second, and third order least squares gradient estimates.

No. of	First order		Second order			Third order		
points	Rel. error	$Bound_1$	Rel. error	$Bound_1$	$Bound_2$	Rel. error	$Bound_1$	$Bound_2$
10	0.08	0.32e1	0.02	0.058e3	0.038e2	0.009	0.9638e4	0.124e2
15	0.09	0.35e1	0.04	0.073e3	0.043e2	0.009	0.5042e4	0.129e2
20	0.20	0.45e1	0.04	0.111e3	0.089e2	0.005	0.6760e4	0.313e2
25	0.27	0.52e1	0.05	0.133e3	0.134e2	0.014	0.7682e4	0.395e2
30	0.33	0.56e1	0.04	0.144e3	0.171e2	0.018	0.8067e4	0.508e2
35	0.38	0.58e1	0.04	0.137e3	0.195e2	0.019	0.8037e4	0.617e2

TABLE 5.12

Scattered data comparisons of the relative error and the error bounds for the function F_3 using weighted first, second, and third order least squares gradient estimates.

No. of	First order		Second order			Third order		
points	Rel. error	$Bound_1$	Rel. error	$Bound_1$	$Bound_2$	Rel. error	$Bound_1$	$Bound_2$
10	0.116	0.066e2	0.009	0.145e3	0.081e2	0.009	0.1518e5	0.018e3
15	0.053	0.086e2	0.025	0.202e3	0.092e2	0.011	0.1555e5	0.025e3
20	0.005	0.128e2	0.027	0.356e3	0.189e2	0.009	0.2289e5	0.066e3
25	0.025	0.164e2	0.029	0.484e3	0.286e2	0.001	0.3048e5	0.102e3
30	0.050	0.191e2	0.031	0.570e3	0.370e2	0.003	0.3435e5	0.134e3
35	0.071	0.212e2	0.031	0.588e3	0.433e2	0.004	0.3381e5	0.160e3

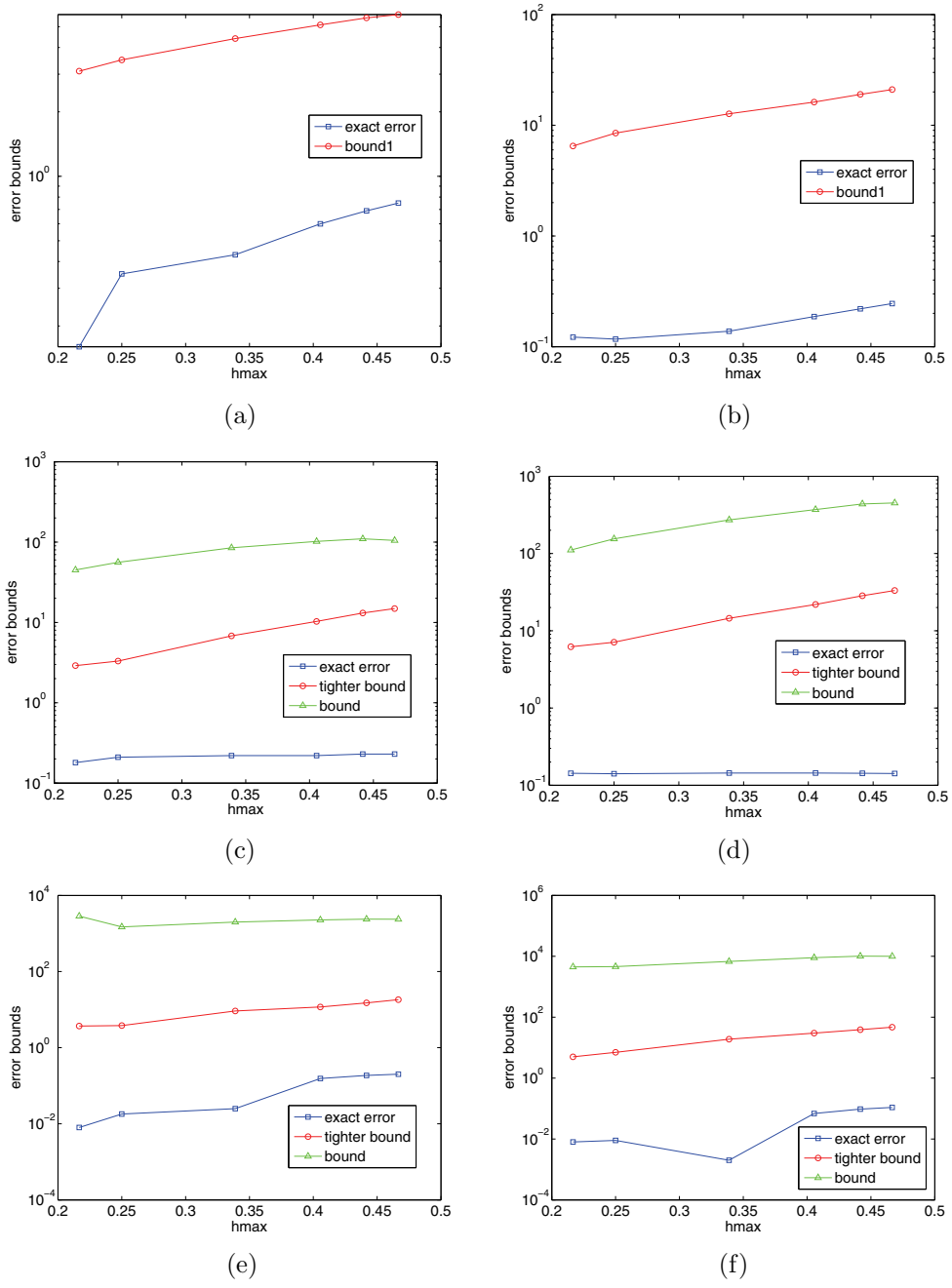


FIG. 5.4. *Relative error (line style ∘) and error bounds (line style △ and □) for function F_1 using* (a) *first order,* (c) *second order, and* (e) *third order least squares estimates. The corresponding weighted least squares estimates and their bounds are depicted in* (b), (d), *and* (f).

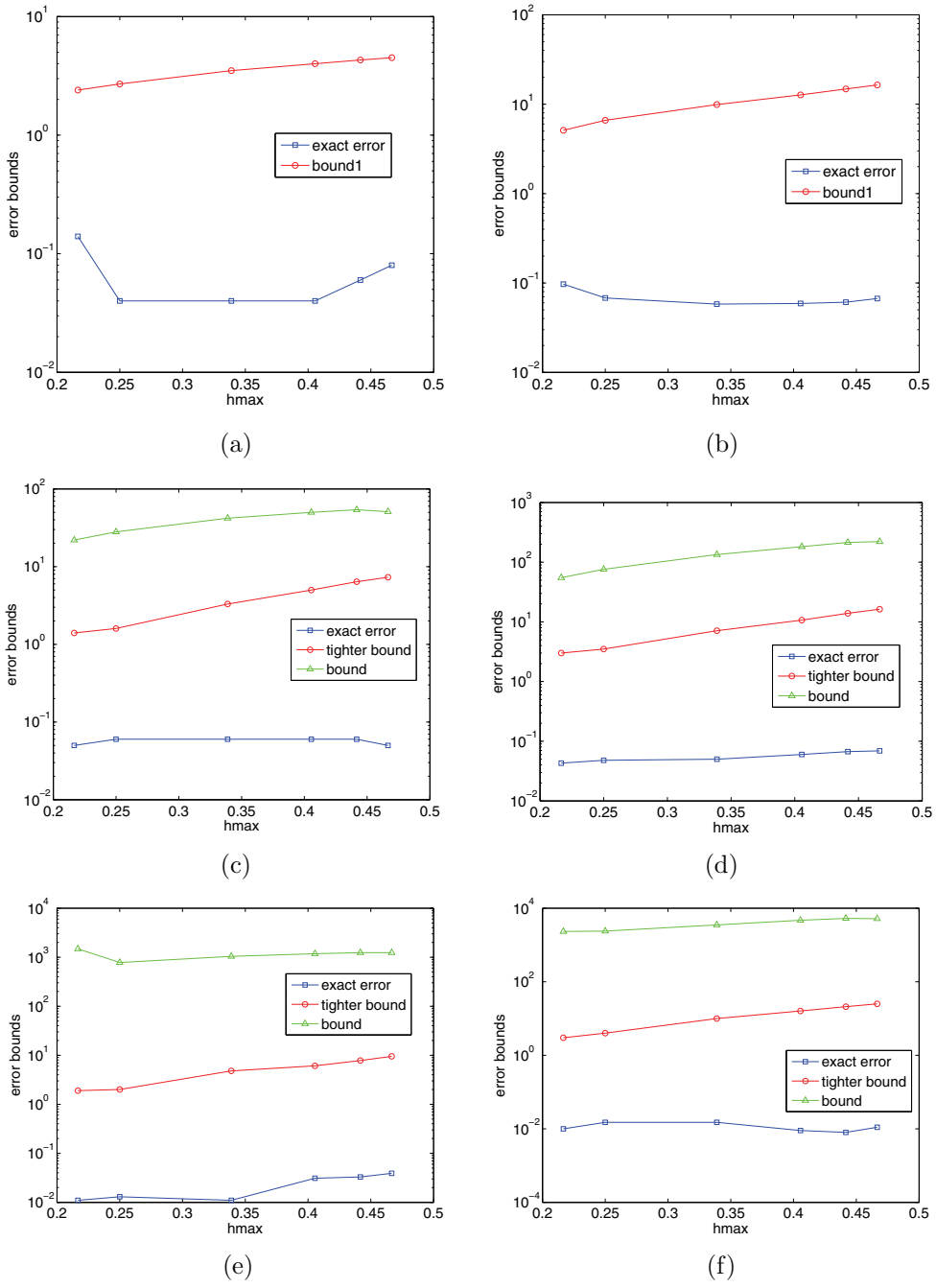


FIG. 5.5. *Relative error (line style ∘) and error bounds (line style △ and □) for function F_2 using* (a) *first order,* (c) *second order, and* (e) *third order least squares estimates. The corresponding weighted least squares estimates and their bounds are depicted in* (b), (d), *and* (f).

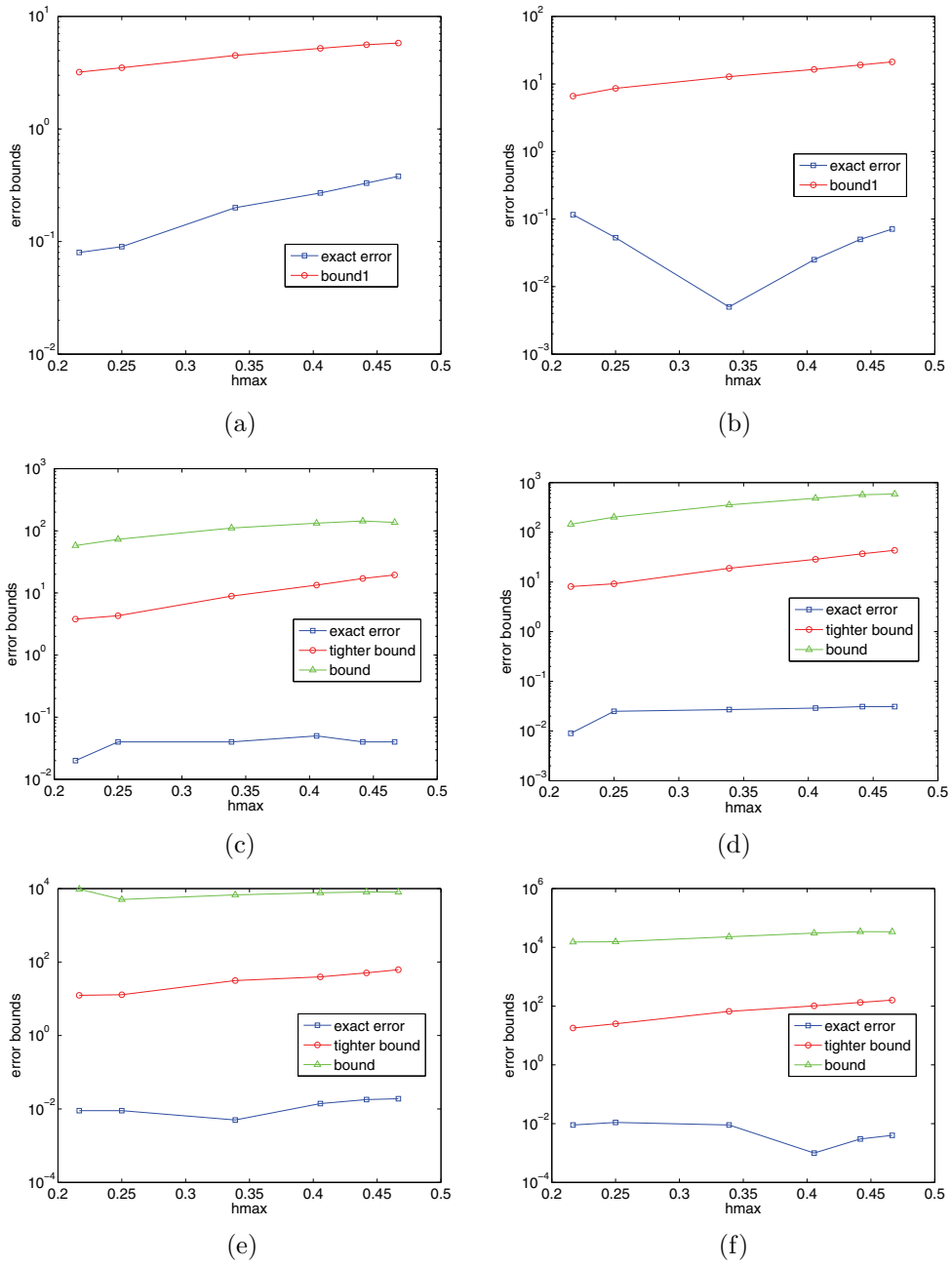


FIG. 5.6. *Relative error (line style ∘) and error bounds (line style △ and □) for function F_3 using* (a) *first order,* (c) *second order, and* (e) *third order least squares estimates. The corresponding weighted least squares estimates and their bounds are depicted in* (b), (d), *and* (f).

6. Conclusion. In this paper we have derived error bounds for the commonly used least squares gradient estimation strategies that are based on truncated Taylor series. We have used results from our previous research to improve and tighten these bounds. An important component of these bounds is the ratio of $h_{\max}^n$ (the maximum distance from the point of interest to any neighboring point in the least squares stencil raised to the order of the method) to the smallest singular value σ_1 of the least squares matrix A. The bounds have been tested to assess the error estimates and asymptotic behavior of the error as the test points move toward a chosen data point. Then the theory was analyzed for a practical scattered data set taken from the literature. The numerical experimentation highlights that the tighter bounds given in Propositions 4.2 and 4.3 are useful in gauging the accuracy of the least squares gradient estimates in that they capture the main trends in the relative error behavior. They also indicate that using a weighted least squares approach offers little improvement in accuracy over the classical least squares strategy.

Acknowledgment. The authors wish to thank the referees for several helpful criticisms that enabled us to improve the overall presentation significantly.

REFERENCES

[1] P. Alfeld and L. Schumaker, *Smooth macro-elements based on Powell–Sabin triangle splits*, Adv. Comput. Math., 16 (2002), pp. 29–46.
[2] T. J. Barth, *Aspects of unstructured grids and finite-volume solvers for the Euler and Navier–Stokes equations*, in Lecture Notes Presented at the VKI Lecture Series 1994–05, Feb. 1994, von Karman Institute, Rhode-St-Genese, Belgium, 1994.
[3] J. Belward, I. Turner, and M. Ilić, *On derivative estimation and the solution of least squares problems*, J. Comput. Appl. Math., 222 (2008), pp. 511–523.
[4] J. A. Belward, I. W. Turner, and M. N. Oqielat, *Numerical investigations of linear least squares methods for derivative estimation*, ANZIAM J., 50 (2008), pp. C844–C857.
[5] M. Bozzini and M. Rossini, *Numerical differentiation of 2D functions from noisy data*, Comput. Math. Appl., 45 (2003), pp. 309–327.
[6] O. Davydov and F. Zeilfelder, *Scattered data fitting by direct extension of local polynomials to bivariate splines*, Adv. Comput. Math., 21 (2004), pp. 223–271.
[7] O. Davydov, *On the approximation power of local least squares polynomials*, in Proceedings of Algorithms for Approximation IV, J. Levesley, I. J. Anderson, and J. C. Mason, eds., University of Huddersfield, Huddersfield, UK, 2002, pp. 346–353.
[8] R. Franke, *Scattered data interpolation: Tests of some methods*, Math. Comput., 38 (1982), pp. 181–200.
[9] T. N. T. Goodman and H. B. Said, *A triangle-based C^1 interpolatiion method*, Rocky Mountain J. Math., 14 (1984), pp. 223–237.
[10] T. N. T. Goodman, H. B. Said, and L. H. T. Chang, *Local derivative estimation for scattered data interpolation*, Appl. Math. Comput., 68 (1995), pp. 41–50.
[11] P. A. Jayantha and I. W. Turner, *A comparison of gradient approximation methods for use in the finite volume computational models for two dimensional diffusion equations*, Numer. Heat Transfer Part B: Fundamentals, 40 (2001), pp. 367–390.
[12] P. Lancaster and K. Salkauskas, *CURVE: Curve and Surface Fitting, An Introduction*, Academic Press, San Diego, 1986.
[13] M. Oqielat, I. Turner, and J. Belward, *A hybrid Clough-Tocher method for surface fitting with application to leaf data*, Appl. Math. Model., 33 (2009) pp. 2582–2595.
[14] C. D. Meyer, *Matrix Analysis and Applied Linear Algebra*, SIAM, Philadelphia, 2000.
[15] C. F. Ollivier-Gooch, *A new class of ENO schemes based on unlimited data-dependent least-squares reconstruction*, in Proceedings of the AIAA 34th Aerospace Sciences Meeting and Exhibit, Reno, NV, 1996, AIAA-96-0887.
[16] H. Pfister, F. Wessels, and A. Kaufman, *Sheared interpolation and gradient estimation for real-time volume rendering*, in Proceedings of the Eurographics Hardware Workshop, Oslo, Norway, 1994, pp. 1–10.

[17] S. E. Stead, *Estimation of gradients from scattered data*, Rocky Mountain J. Math., 14 (1984), pp. 265–279.
[18] T. Wei, Y. C. Hon, and Y. B. Wang, *Reconstruction of numerical derivatives from scattered noisy data*, Inverse Problems, 21 (2005), pp. 657–672.
[19] C. Zuppa, *Error estimates for modified local Shepard's interpolation formula*, Appl. Numer. Math., 49 (2004), pp. 245–259.