



Queensland University of Technology
Brisbane Australia

This is the author's version of a work that was submitted/accepted for publication in the following source:

[Belward, John A.](#), [Turner, Ian W.](#), & [Oqielat, Moa'Ath](#) (2009) Numerical investigations of linear least squares methods for derivative estimation. *ANZIAM Journal*, 50, C844-C857.

This file was downloaded from: <http://eprints.qut.edu.au/30160/>

Notice: *Changes introduced as a result of publishing processes such as copy-editing and formatting may not be reflected in this document. For a definitive version of this work, please refer to the published source:*

Numerical investigations of linear least squares methods for derivative estimation

John A. Belward* Ian W. Turner* Moa'ath N.Oqielat*

September 1, 2009

Dedicated to Ian Sloan on the occasion of his seventieth birthday

Abstract

The results of a numerical investigation into the errors for least squares estimates of function gradients are presented. The underlying algorithm is obtained by constructing a least squares problem using a truncated Taylor expansion. An error bound associated with this method contains in its numerator terms related to the Taylor series remainder, while its denominator contains the smallest singular value of the least squares matrix. Perhaps for this reason the error bounds are often found to be pessimistic by several orders of magnitude. The circumstance under which these poor estimates arise is elucidated and an empirical correction of the theoretical error bounds is conjectured and investigated numerically. This is followed by an indication of how the conjecture is supported by a rigorous argument.

Contents

1	Introduction	2
2	Construction of a leaf surface	3
3	Theoretical error bounds	4
4	Numerical results	7

*School of Mathematical Sciences, Queensland University of Technology, Brisbane, AUSTRALIA. <mailto:john.belward@qcscm.com>

1 Introduction

Results are presented for the numerical investigation of an algorithm to determine the gradient of a function from planar scattered data values. The particular method analysed was chosen to provide accurate gradient estimates combined with the opportunity for computational efficiency with large data sets. The current target of the work is in the construction of models of leaf surfaces [4, 5], which assists the study of local effects such as the movement of droplets across the surface.

The work also has the potential to contribute to simulation models for the growth of plants. This has been facilitated by the development of L-systems that are used to capture the architecture of a plant, or tree, as it grows. Since leaves are an essential part of a plant, the understanding of how they respond to physical inputs is crucial. Knowing where the surface is located is essential for light interception and deposition of water, liquids and perhaps powders. While a piecewise linear representation may be adequate in certain situations, the applications envisaged in this work call for a smoothly varying surface normal, thus continuity of the gradient will be built into the algorithms described in Section 2. A theoretical error bound is derived for the least squares gradient estimate in Section 3. This bound is evaluated in Section 4 and found to be several orders of magnitude pessimistic. An examination of the terms comprising the bound suggests a modification and although the evidence at this stage is purely experimental, a heuristic argument is developed that suggests the means by which a more accurate bound has been derived.

An important observation is the occurrence of the smallest non-zero singular value of the least squares matrix in the denominator of the error bound. The progressive accumulation of the approximants by computation of the singular value expansion is tabulated. The results quantify the contributions of the elements of the singular value decomposition and demonstrate the contribution of each term to the gradient approximation. As a result of this numerical investigation it is conjectured that rather than using the smallest singular value in the error bound it appears more appropriate to use the smallest singular value of the reduced form of the least squares matrix considered by Belward, Turner and Ilic [1]. The results of this conjecture appear quite promising and point to how this new result has been justified analytically.

2 Construction of a leaf surface

Both sonic and laser scanner devices are used to capture leaf surfaces and in both cases the raw data comprises three-dimensional coordinates of points. A recent presentation by Oqielat, Belward, Turner and Loch [4] describes the use of piecewise polynomial basis functions on a triangulation of the data points to generate smooth representations of the scanned surface. The image is constructed by a surface fit to the scattered data and realism may be conveyed by adding texture to the surface plots. Clough–Tocher elements are triangular elements that are comprised of three triangular micro-elements, which each have their common interior vertex at the incentre of the element. A piecewise cubic approximant may be constructed with a continuous gradient using this configuration provided that gradient values are available at the nodes, together with the normal derivative at the midpoints of the edges. Recently, Oqielat et al. [4] gave results for leaf surface fitting that utilise the Clough–Tocher representation to interpolate the scattered data to ensure this continuous gradient. The approximant is expressed as a linear combination of twelve basis functions; a cardinal basis for a standard triangle is given by Lancaster and Salkauskas [3]. Since for the leaf surface only data values are available at the scanned data points, gradient estimates are needed and a brief description of an accurate method to provide these estimates is given below.

Taylor’s Theorem for several variables states for $f : \mathbb{R}^2 \rightarrow \mathbb{R}$, in open convex set $D \subset \mathbb{R}^2$ that

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

where the remainder R_n has the integral form

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

In this work, in order to approximate the gradient at a point a , the Taylor series is truncated and the first three terms of the right hand side of (1) evaluated at a scattered set of points $v_i = a + h_i\nu_i, i = 1, \dots, m$. The term $f(a)$ is transferred to the left hand side and the equation divided by $h_i = \|v_i - a\|$ to obtain

$$\frac{f(a + h_i\nu_i) - f(a)}{h_i} = (\nu_i \cdot \nabla)f(a) + \frac{h_i}{2} \nu_i^T H_f(a) \nu_i,$$

where $H_f(a)$ is the Hessian of f evaluated at a . This equation is applied at each of a neighbouring set of m points $v_i, i = 1, \dots, m$ near a to obtain the

overdetermined linear system

$$A\gamma \approx q, \quad (2)$$

where vector q has as its elements the difference quotients for f in the direction of the unit vectors $\nu_i = (\nu_{x_i}, \nu_{y_i})^T$. The elements of γ are approximations to the five partial derivatives of f , namely

$$\gamma \approx \left(\frac{\partial f(a)}{\partial x}, \frac{\partial f(a)}{\partial y}, \frac{\partial^2 f(a)}{\partial x^2}, \frac{\partial^2 f(a)}{\partial x \partial y}, \frac{\partial^2 f(a)}{\partial y^2} \right)^T,$$

and the matrix $A \in \mathbb{R}^{m \times 5}$ takes the following form

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

The least squares solution of equation (2) for $\gamma = \arg \min_{x \in \mathbb{R}^5} \|Ax - q\|_2$ yields estimates of the gradient of f at a that are $O(h^2)$ accurate.

In earlier work by Belward, Turner and Ilić [1] it was observed that one could either take the direct approach and estimate $\nabla f(a) \approx E_1 A^\dagger q$ where $E_1 = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \end{pmatrix}$ and A^\dagger denotes the pseudoinverse or generalised inverse of A as given by Lawson and Hansen [2]; or perform an orthogonal reduction of columns 3–5 in $A = (A_1|A_2)$ as $Q^T A_2 = \begin{pmatrix} A_{12} \\ 0 \end{pmatrix}$ to obtain $Q^T A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & 0 \end{pmatrix}$, thus enabling the approximation $\nabla f(a) \approx \arg \min_{y \in \mathbb{R}^2} \|A_{21}y - \tilde{q}\|_2$ where \tilde{q}_2 represents the last $m - 3$ entries in $Q^T q$. By Belward et al. [1] these methods provide the same least squares solution and error.

3 Theoretical error bounds

The following lemma and proposition establish a bound for the gradient approximation computed using the algorithms described in the previous section.

Lemma 1. *Let $f : D \subset \mathbb{R}^2 \rightarrow \mathbb{R}$ in an open convex set D and $f \in C^2(D)$. Suppose that $H_f \in \text{Lip}_\lambda(D)$. Then for any $a + h\nu \in D$ with $\|\nu\|_2 = 1$,*

$$\left| \frac{h}{2} \nu^T H_f(a) \nu - \left\{ \frac{f(a + h\nu) - f(a)}{h} \right\} + \nu^T \nabla f(a) \right| \leq \frac{\lambda h^2}{6}. \quad (3)$$

Proof. Rearranging the multivariable Taylor series for f about the point a gives

$$\frac{f(a + \nu h) - f(a)}{h} - \nu^T \nabla f(a) = h \int_0^1 (1-t) \nu^T H_f(a + th\nu) \nu dt,$$

and therefore

$$\frac{h}{2} \nu^T H_f(a) \nu - \left\{ \frac{f(a + h\nu) - f(a)}{h} \right\} + \nu^T \nabla f(a) = h \int_0^1 (1-t) \nu^T \left\{ H_f(a) - H_f(a + th\nu) \right\} \nu dt.$$

Hence,

$$\left| \frac{h}{2} \nu^T H_f(a) \nu - \left\{ \frac{f(a + h\nu) - f(a)}{h} \right\} + \nu^T \nabla f(a) \right| \leq h \int_0^1 \left| (1-t) \nu^T \left\{ H_f(a) - H_f(a + th\nu) \right\} \nu \right| dt.$$

The result follows by invoking Cauchy–Schwarz inequality and Lipschitz continuity and noting that $\|\nu\| = 1$. \blacksquare

To estimate the Lipschitz constant λ in (3) we note that

$$\|H_f(a) - H_f(b)\|_2 \leq \sqrt{2} \|H_f(a) - H_f(b)\|_\infty.$$

By defining

$$\theta = \max_{x \in D} \left\{ \left| \frac{\partial^3 f}{\partial x^{3-i} \partial y^i} \right|, i = 0, 1, 2, 3 \right\}$$

it is shown using the mean value theorem that $\|H_f(a) - H_f(b)\|_\infty \leq 2\sqrt{2}\theta \|a - b\|_2$. Therefore, we conclude that $\|H_f(a) - H_f(b)\|_2 \leq 4\theta \|a - b\|$ and $\lambda = 4\theta$. In §4 these maxima were determined using the Maple software. Another strategy for estimating the Lipschitz constant is elaborated by Turner, Belward and Oqielat [6].

Proposition 1. *Suppose around point a we have m neighbouring points $v_i, i = 1, \dots, m$ with $a, v_1, \dots, v_m \in D$; D an open convex set $\subset \mathbb{R}^2$. Suppose further that $f \in C^2(D)$ with $H_f \in Lip_\lambda(D)$ and we approximate the gradient locally at a by $E_1\gamma$ via the least squares solution $\gamma = \arg \min_{x \in \mathbb{R}^5} \|Ax - q\|$, where*

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

and $h_i = \|v_i - a\|_2$ with $h_i v_i = v_i - a$; and $\nu_{i1}^T = (\nu_{x_i}, \nu_{y_i})$, $\nu_{i2}^T = (\frac{h_i}{2} \nu_{x_i}^2, h_i \nu_{x_i} \nu_{y_i}, \frac{h_i}{2} \nu_{y_i}^2)$.
Then,

$$\frac{\|\nabla f(a) - E_1 \gamma\|_2}{\|\nabla f(a)\|_2} \leq \frac{\lambda \sqrt{m} h_{max}^2}{6 \sigma_1 \|\nabla f(a)\|_2}, \quad (4)$$

where σ_1 is the smallest singular value of matrix A , which is assumed to have $\text{rank}(A) = 5$, and $h_{max} = \max_{1 \leq i \leq m} h_i$.

Proof. Let $E_2 \in \mathbb{R}^{3 \times 5}$ be the last 3 rows of the identity matrix I_5 and

$$U = \left(\frac{\partial f}{\partial x}(a), \frac{\partial f}{\partial y}(a), \frac{\partial^2 f}{\partial x^2}(a), \frac{\partial^2 f}{\partial x \partial y}(a), \frac{\partial^2 f}{\partial y^2}(a) \right)^T$$

be the exact values of the derivatives at a . Now $U - \gamma$ is partitioned as

$$U - \gamma = \begin{pmatrix} E_1(U - \gamma) \\ E_2(U - \gamma) \end{pmatrix}$$

with E_1 defined above and hence it follows that,

$$\|U - \gamma\|_2^2 = \|E_1(U - \gamma)\|_2^2 + \|E_2(U - \gamma)\|_2^2 \geq \|E_1(U - \gamma)\|_2^2 = \|\nabla f(a) - E_1 \gamma\|_2^2.$$

Next, with $\gamma = A^\dagger q$ and $A^\dagger A = I_5$, we obtain

$$\|U - \gamma\|_2^2 = \|U - A^\dagger q\|_2^2 = \|A^\dagger(AU - q)\|_2^2 \leq \|A^\dagger\|_2^2 \|AU - q\|_2^2 = \frac{1}{\sigma_1^2} \|AU - q\|_2^2.$$

Now using the result in lemma 1, the following upper bound is derived

$$\begin{aligned} \|AU - q\|_2^2 &= \sum_{i=1}^m \left| \frac{h}{2} \nu_i^T H_f(a) \nu_i - \left\{ \frac{f(a + h_i \nu_i) - f(a)}{h_i} \right\} + \nu_i^T \nabla f(a) \right|^2 \\ &\leq \sum_{i=1}^m \left(\frac{\lambda h_i^2}{6} \right)^2 \quad (\text{using lemma 1}) \\ &\leq \left(\frac{\lambda}{6} \right)^2 \sum_{i=1}^m \left(h_{max}^2 \right)^2 = \left(\frac{\lambda}{6} \right)^2 h_{max}^4 m. \end{aligned}$$

The result follows by taking square roots and noting that

$$\|\nabla f(a) - E_1 \gamma\|_2 \leq \|U - \gamma\|_2 \leq \frac{\lambda h_{max}^2}{6 \sigma_1} \sqrt{m}$$

and then dividing both sides by $\|\nabla f(a)\|_2$. ■

Although the bound (4) the approximation scheme has a truncation error of order two, the conclusion of the proposition does not guarantee second

order accuracy because of the presence of the smallest singular value σ_1 of the least squares matrix in the error bound. The numerical results presented here and in earlier work provide strong evidence that such a conclusion would be valid. Confining the discussion to the gradient approximation, the results of the next section show that the bound (4) is pessimistic and additional numerical investigation suggests that a better bound is possible by utilising the singular value decomposition of A_{21} .

4 Numerical results

The tests described here were run on the function for $(\sin r)/r$, for $r = (x^2 + y^2)^{\frac{1}{2}}$. One hundred points $(r_i \cos \theta_i, r_i \sin \theta_i)^T$ were generated such 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 30 points nearest the origin were chosen and used as displacements from the point (3, 4) to generate a scattered data set. The radial distances were scaled to produce the annuli with inner radii shown in the top row of Table 1. The results given in the table provide the numerical evidence for second order and first order convergence of, respectively, the gradient and the Hessian approximations.

Table 1: Norms of the errors in the gradient and Hessian with differing radii

radius	2.5e-01	2.5e-02	2.5e-03	2.5e-04	2.5e-05
gradient error	1.3315e-02	1.3363e-04	1.3363e-06	1.3363e-08	1.3537e-10
hessian error	2.2085e-02	2.5141e-03	2.5457e-04	2.5486e-05	2.7235e-06

Table 2: Error bound and modified error bound values for varying radii

radius	error bound	modified error bound
2.5e-01	0.888e+01	.983e+00
2.5e-02	0.888e+00	.983e-02
2.5e-03	0.888e-01	.983e-04
2.5e-04	0.888e-02	.983e-06
2.5e-05	0.888e-03	.983e-08

Table 2 contains two sets of results. Column two lists the error bounds computed from expression (4) whereas column three shows the values computed from a proposed amendment to expression (4). Column one contains

the corresponding radii. Examination of the components of expression (4) reveals that it is the smallest singular value of the least squares matrix that produces the poor error estimates exhibited in column two of Table 2. From the singular value analysis of the least squares matrix A given in Table 3 it is observed that the singular values fall into a group of two and a group of three identified by their magnitudes. In terms of the ordering of the columns of the least squares matrix these are associated with, respectively, the gradient and the Hessian. It therefore seems quite plausible that a closer bound would be obtained if the smallest singular value of the group of two were used in place of the smallest singular value of the whole system. This conjecture is apparent in the third column of Table 2, which provides a much tighter bound on the gradient estimates given in Table 1.

The argument for this amendment is strengthened by examining the elimination method described at the end of §2. After the orthogonal reduction of A_2 and the omission of the first three equations a new least squares matrix A_{21} is obtained that has just two non-zero columns and two non-zero singular values. Results displayed in Table 3 show that the singular values of the new system become increasingly close to those of the full least squares matrix as the radius of the test point set is reduced. Note that the first two columns of the full least squares matrix A are independent of the radii of the annuli. The columns of the elimination matrix A_{21} have a dependence on h through the orthogonal reduction, although this dependence diminishes as h is reduced.

A further calculation that also demonstrates the importance of the first two singular values is to use the singular value decomposition of the least squares matrix to progressively construct the least squares solution. As the

Table 3: Singular values (sv's) of the least squares and elimination matrices for various radii

radius	2.5e-01	2.5e-02	2.5e-03	2.5e-04	2.5e-05
sv's of	1.3486e-01	1.3476e-01	1.3476e-01	1.3476e-01	1.3476e-01
A	1.2321e-01	1.2318e-01	1.2318e-01	1.2318e-01	1.2318e-01
	1.9399e-02	1.9413e-03	1.9413e-04	1.9413e-05	1.9413e-06
	1.6844e-02	1.6848e-03	1.6848e-04	1.6848e-05	1.6848e-06
	1.3650e-02	1.3651e-03	1.3651e-04	1.3651e-05	1.3651e-06
sv's of	1.3476e-01	1.3476e-01	1.3476e-01	1.3476e-01	1.3476e-01
A_{21}	1.2318e-01	1.2318e-01	1.2318e-01	1.2318e-01	1.2318e-01

Table 4: Cumulative sums of the singular expansion solution for the derivatives for the point (3,4) for inner annular radii of 2.5e-1, 2.5e-3 and 2.5e-5

number of terms	f_x	f_y	radius 2.5e-1		
			f_{xx}	f_{xy}	f_{yy}
1	2.5279e-02	-9.1398e-03	-6.7030e-04	8.0706e-04	-2.5922e-04
2	5.6724e-02	7.7756e-02	6.4653e-05	1.1226e-03	1.4595e-03
3	5.8838e-02	7.6049e-02	5.9274e-02	-2.6472e-02	2.8863e-02
4	5.6445e-02	7.5021e-02	6.9913e-02	6.9343e-02	1.0248e-01
5	5.6462e-02	7.4987e-02	6.7879e-02	6.7568e-02	1.0508e-01
number of terms	f_x	f_y	radius 2.5e-3		
			f_{xx}	f_{xy}	f_{yy}
1	2.6256e-02	-9.4525e-03	-6.7977e-06	8.2338e-06	-2.6297e-06
2	5.7056e-02	7.6099e-02	2.5457e-07	1.1245e-05	1.3981e-05
3	5.7077e-02	7.6082e-02	5.9950e-02	-2.8056e-02	2.7745e-02
4	5.7053e-02	7.6072e-02	7.0680e-02	6.7452e-02	1.0122e-01
5	5.7054e-02	7.6071e-02	6.7528e-02	6.4704e-02	1.0525e-01
number of terms	f_x	f_y	radius 2.5e-5		
			f_{xx}	f_{xy}	f_{yy}
1	2.6263e-02	-9.4550e-03	-6.7994e-08	8.2359e-08	-2.6304e-08
2	5.72054e-02	7.6072e-02	2.5077e-09	1.1246e-07	1.3976e-07
3	5.7054e-02	7.6072e-02	5.9956e-02	-2.8070e-02	2.7733e-02
4	5.7054e-02	7.6072e-02	7.0685e-02	6.7430e-02	1.0120e-01
5	5.7054e-02	7.6072e-02	6.7521e-02	6.4671e-02	1.0525e-01

terms are added the progress towards the solution is observed. Three sets of results are shown in Table 4. The dominance of the first two terms is increasingly marked with the diminution of the radius; by the time the radius reaches $2.5e - 05$ only the first two terms are needed. Progress towards the solution for the Hessian is more steady; the results of Table 4 conform to $O(h)$ convergence as do the bounds in column 2 of Table 2.

5 Conclusion

Numerical investigations have been made of the theoretical bounds of the errors in gradient estimation from scattered data values using a least squares algorithm. These bounds have been compared with the errors that are observed when the algorithm is implemented.

The smallest singular value of the least squares matrix A has a large

detrimental influence on the realism of the bounds. On the basis of numerical experimentation, a modification is suggested in which certain singular values are deemed associated with the gradient. This modification is more apparent when a version of the algorithm is used wherein the Hessian terms are eliminated.

These results, which are speculative will be shown to be rigorous in further rigorous research [6].

Acknowledgements: J.A.Belward thanks the Queensland Consortium for Scientific Computation and Mathematics and Queensland University of Technology for financial support.

References

- [1] J.A. Belward, I.W. Turner, and M. Ilic. On derivative estimation and the solution of least squares problems. *Journal of Computational and Applied Mathematics*, 222:511–523, 2008.
- [2] C.L.Lawson and R.J.Hansen. *Solving Least Squares Problems*. Prentice-Hall, 1974.
- [3] P. Lancaster and K. Salkauskas. *Curve and Surface Fitting, An Introduction*. Academic Press, San Diego, 1986.
- [4] M.N. Oqielat, J.A. Belward, I.W. Turner, and B.I. Loch. A Hybrid Clough-Tocher Radial Basis Function Method for Modelling Leaf Surfaces. In *Oxley, L. and Kulasiri, D. (eds) MODSIM 2007 International Congress on Modelling and Simulation*. http://mssanz.org.au/MODSIM2007/papers/7_s50, pages 400–406, 2007.
- [5] M.N. Oqielat, I.W. Turner, and J.A. Belward. A hybrid clough-tocher method for surface fitting with application to leaf data. *Applied Mathematical Modeling*. <http://doi:10.1016/j.apm.2008.07.023>, 2008.
- [6] I.W. Turner, J.A. Belward, and M.N. Oqielat. Error bounds for least squares gradient estimates. *In preparation*, 2008.