

THE ACCURACY OF LEAST-SQUARES APPROXIMATION ON HIGHLY STRETCHED MESHES

N. B. PETROVSKAYA

*School of Mathematics, University of Birmingham
Edgbaston, Birmingham, B15 2TT, UK
n.b.petrovskaya@bham.ac.uk*

Received 1 November 2007

Accepted 2 July 2008

The least-squares (LS) method is often used in computational aerodynamics to reconstruct a given function at certain points of a computational grid. In this paper we discuss the accuracy of the LS approximation on highly stretched meshes that are inherent in computational aerodynamics. A new definition of a distant point in a LS reconstruction stencil will be given in order to explain the poor performance of the method in a boundary layer region. Namely, based on the concept of outliers widely used in the statistics, we demonstrate that the definition of a distant point in a LS reconstruction stencil should take into account the solution properties and it cannot rely upon the geometric shape of the stencil only. Our approach is illustrated with numerical examples.

Keywords: Least-squares method; solution reconstruction; outliers; stretched mesh.

1. Introduction

The least-squares (LS) method is one of the best-known approaches to solving a problem of finding the best polynomial approximation to the input samples. It was originally developed for statistical regression, but nowadays a general concept of LS approximation is widely used for various applications beyond the statistics. The numerous examples include meshfree methods [Atluri and Shen (2005); Liu and Gu (2005)], computational mechanics [Belytschko *et al.* (1996)], computer-aided geometric design and computer graphics [Bajaj *et al.* (1993); Pratt (1987)], image processing [Unser *et al.* (1995)], etc.

In recent decades the LS approach has become popular in computational aerodynamics, where it is often used to reconstruct a given function at certain points of a computational grid [Barth (1991); Haselbacher (2006); Mavriplis (2003); Ollivier-Gooch and Van Altena (2002)]. One basic application of the LS method is to reconstruct a solution gradient in projection–evolution discretization schemes which are heavily exploited in modern industrial aerodynamic codes. A second-order accurate discretization requires the solution gradient reconstruction to maintain the accuracy of the scheme [Anderson (1994); Barth and Jespersen (1989)]. The solution

gradients should be reconstructed based upon a solution field at mesh nodes, and the LS method is an attractive approach for this task, as it only requires a local reconstruction stencil to compute the gradient. The need to reconstruct the solution locally may also arise from the requirements of the solution postprocessing for real-life aerodynamic applications. Furthermore, a LS reconstruction of a given function at grid nodes or edge midpoints may be required for the purpose of grid adaptation [Borouchaki *et al.* (1996); Buscaglia and Dari (1997)].

For all of the above-mentioned computational aerodynamics problems, accuracy is the main requirement for a function reconstruction procedure. Usually the LS method is considered a reliable means of approximation that can reproduce a given smooth function with the desired accuracy [Atluri and Shen (2005); Sonar (2005)]. However, all results about the accuracy of the LS method are related to a regular geometry of a computational grid. Meanwhile, most of the present aerodynamic solvers use unstructured grids where the results of LS approximation depend strongly on the grid geometry. The grid generation in aerodynamic problems often results in computational regions where grid cells are almost degenerate. In particular, grid cells with a very high cell aspect ratio inevitably appear as a result of grid generation in boundary layer regions. Those irregular grid geometries present a serious difficulty when a solution function or solution gradient should be reconstructed by the LS method.

While the accuracy tests made on quasiuniform grids demonstrate very good results (see e.g. [Ollivier-Gooch and Van Altena (2002)]), a LS reconstruction behaves differently on stretched meshes [Mavriplis (2003); Petrovskaya (2004); Petrovskaya and Volkov (2007)]. From a LS approximation viewpoint, the function reconstruction over highly stretched grids means the presence of remote points in a reconstruction stencil. A conventional LS approach allows one to choose an approximation stencil based on some *a priori* considerations, so that the remote points can be eliminated from the reconstruction procedure. This is impossible, however, in computational aerodynamics problems, as the geometry of stretched grid cells cannot be changed in computations. As a result, the LS method may degrade to unacceptable accuracy on stretched meshes [Mavriplis (2003); Petrovskaya (2007)]. This problem becomes especially important when a higher order reconstruction is considered, as a higher order LS method requires an expanded reconstruction stencil that often captures many distant points.

In this paper we address the issue of the accuracy of the LS method in the presence of distant points in a reconstruction stencil. For this purpose we consider a general concept of outliers in a data set that is widely used in the statistics [Neter *et al.* (1985)]. Namely, we employ the concept of outliers to explain poor performance of a LS approximation over distorted geometries. We discuss two types of outliers that are well known in statistics problems (see e.g. [Kenney and Keeping (1962); Neter *et al.* (1985)]) to demonstrate that both of them may affect the accuracy of a LS reconstruction. As a result of our study we conclude that a definition of a distant point in a LS reconstruction stencil cannot rely upon the geometric shape

of the stencil only, as such a definition should also take into account the information about a solution function. Our approach is illustrated with numerical examples.

2. The Problem Statement

Consider a data set $\mathbf{U} = (U_1, U_2, \dots, U_N)$, where the data U_i represent a continuous function $U(x, y)$ at points $P_i = (x_i, y_i)$, $i = 1, \dots, N$. We have to fit the data \mathbf{U} to the function

$$u_{LS}(x, y) = \sum_{k=0}^M u_k \phi_k(x, y), \quad M < N, \tag{1}$$

where (u_0, u_1, \dots, u_M) are fitting parameters, and $\phi_k(x, y)$, $k = 0, \dots, M$ are polynomial basis functions. A linear regression that we use to compute the vector $\mathbf{u} = (u_0, u_1, \dots, u_M)$ is based on the following definition of the merit function F^2 (see e.g. [Press *et al.* (1996)]):

$$F^2 = \sum_{i=1}^N \left[\frac{U_i - \sum_{k=0}^M u_k \phi_k(P_i)}{\sigma_i} \right]^2, \tag{2}$$

where parameter σ_i is the weight of the i th data point. In aerodynamic applications a widespread approach is to define the weights $1/\sigma_i$ as

$$\frac{1}{\sigma_i} = r_{0i}^{-p}, \tag{3}$$

where $P_0 = (x_0, y_0)$ is the point where the solution is reconstructed, r_{0i} is the distance between P_0 and P_i , and $p = 0, 1, 2, \dots$ is an integer polynomial degree. The unweighted reconstruction corresponds to $p = 0$, while $p > 0$ provides inverse distance weighting used to mitigate the impact of remote stencil points on the results of LS approximation.

The LS approach considers the vector \mathbf{u} as the best fit to a given data set, if \mathbf{u} minimizes the function (2). Thus the parameters u_k can be found from the $M + 1$ conditions

$$\frac{\partial F^2}{\partial u_k} = 0, \quad k = 0, \dots, M,$$

which are called the normal equations of the LS problem. Taking into account the definition (2), we obtain the normal equations in the form

$$\sum_{i=1}^N \frac{1}{\sigma_i^2} \left[U_i - \sum_{j=0}^M u_j \phi_j(P_i) \right] \phi_k(P_i) = 0, \quad k = 0, \dots, M.$$

We introduce the weighted data \mathbf{b} and the design matrix \mathbf{A} as follows:

$$b_i = \frac{U_i}{\sigma_i}, \quad A_{ij} = \frac{\phi_j(P_i)}{\sigma_i}, \quad i = 1, \dots, N, \quad j = 0, \dots, M.$$

The normal equations can then be written as

$$\mathbf{D}\mathbf{u} = \mathbf{f}, \quad (4)$$

where the matrix $\mathbf{D} = \mathbf{A}^T \mathbf{A}$, and the right-hand side $\mathbf{f} = \mathbf{A}^T \mathbf{b}$. They are to be solved for the vector of parameters $\mathbf{u} = (u_0, \dots, u_M)$:

$$\mathbf{u} = \mathbf{D}^{-1} \mathbf{f}. \quad (5)$$

It can be seen from the definition of the LS method that the fitting parameters \mathbf{u} in (1) depend on the geometry $\{P\}$. A standard LS approach allows one to determine the points P_i , where the observations U_i are made, based on some *a priori* considerations. Meanwhile, a LS reconstruction procedure used in aerodynamic codes has to deal with prescribed geometries that cannot be altered. In other words, the data \mathbf{U} should be taken from a solution function that is defined only at nodes of a computational grid. That restriction presents a serious computational drawback of a LS solution reconstruction. Namely, it is well known that, while the LS method provides accurate solution approximation on uniform meshes, the LS reconstruction fails on stretched meshes with irregular geometry of grid cells. We discuss below a NACA 0012 airfoil test case [Johnson (2007)] that illustrates the behavior of a LS reconstruction on stretched meshes.

In the NACA 00012 test case a numerical solution to the Navier–Stokes equations is reconstructed at edge midpoints on a coarse grid by a quadratic polynomial:

$$u_{\text{LS}}(x, y) = u_0 + u_1(x - x_0) + u_2(y - y_0) + u_3(x - x_0)^2 + u_4(x - x_0)(y - y_0) + u_5(y - y_0)^2. \quad (6)$$

An example of a reconstruction stencil for the edge midpoint P_0 is shown in Fig. 1. The two nodes n_1 and n_2 that determine the edge where the reconstruction should be made are identified, and the reconstruction stencil consists of all nodes that belong to edges incident to the node n_1 or n_2 . Since a reconstruction stencil may include very distant points, a weighting procedure (3) is used in the problem.

The grid where the reconstruction has been made is then uniformly refined, so that the edge midpoints on the original grid become grid nodes on a new grid. Those

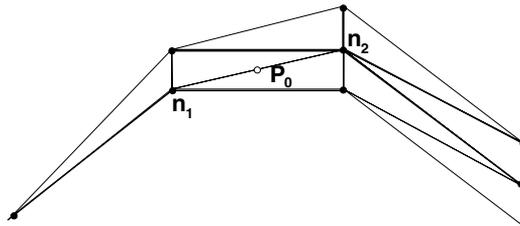


Fig. 1. An example of a LS reconstruction stencil. Stencil members for the solution reconstruction at point P_0 are shown as closed circles.

grid nodes are marked and stored, as they are further used to compute the error of the LS reconstruction. Since the exact solution is not available in the problem, a solution field is computed over a refined grid by a numerical method that employs higher order polynomial functions to approximate the solution (SUPG, [Hughes and Brooks (1979)]). Finally, an accurate “true” solution is compared with the LS reconstruction made over a set of marked grid nodes on a fine grid to validate the accuracy of the LS approximation.

A reconstructed solution field around the airfoil is shown in Fig. 2(a). While the overall accuracy of the LS approximation is good, a quadratic LS reconstruction

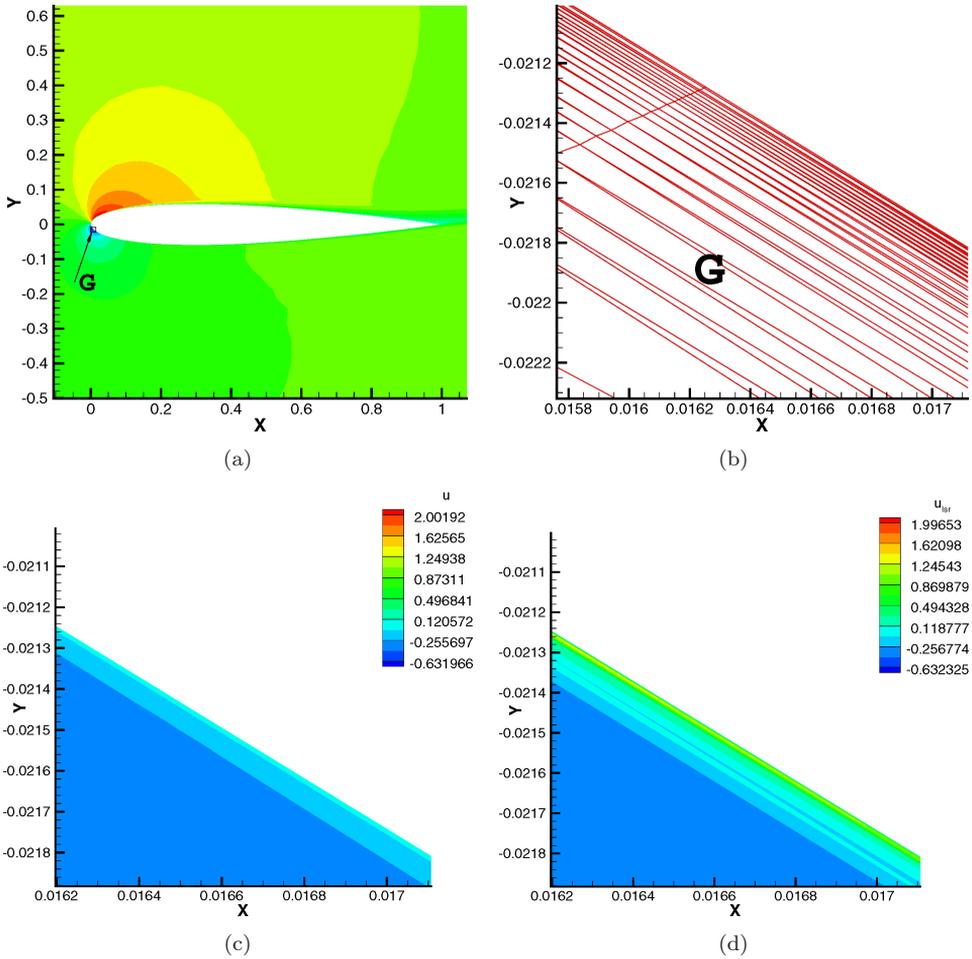


Fig. 2. (a) The solution field $u(x,y)$ (the x component of the velocity field) around the airfoil in the NACA 0012 test case. (b) A fragment of a computational grid near the wall (a close-up of domain G in Fig. 2a). (c) A “true” SUPG solution in domain G . (d) A quadratic LS reconstruction in domain G .

does not perform well in grid subdomains with stretched grid cells are present. A grid fragment near the wall [domain \mathbf{G} in Fig. 2(b)] has been selected to demonstrate the results of a quadratic reconstruction on cells with a high cell aspect ratio. The grid geometry in the boundary layer is shown in Fig. 2(b), while a “true” SUPG solution is presented in Fig. 2(c). A reconstructed solution is shown in Fig. 2(d), where it can be seen that the LS reconstruction results in a nonmonotone and inaccurate solution near the wall.

A reconstruction error computed in the boundary layer region confirms the above result. The error function $e(x, y)$ is

$$e(x, y) = |u_{\text{SUPG}}(x, y) - u_{\text{LS}}(x, y)|, \quad (7)$$

where $u_{\text{SUPG}}(x, y)$ and $u_{\text{LS}}(x, y)$ are a “true” solution and a quadratically reconstructed solution, respectively. The maximum error near the wall is defined as

$$e_{\max} = \max_{(x, y) \in D} e(x, y),$$

where $D = \{(x, y) : \sqrt{x^2 + y^2} < r\}$. We choose the value of r as 2.0, while the entire computational domain has a radius $R = 1000.0$. The maximum error for the quadratic LS reconstruction is

$$e_{\max} \equiv e(x_0, y_0) = 1.38595,$$

where $u_{\text{SUPG}}(x_0, y_0) = -0.0389462$, $u_{\text{LS}}(x_0, y_0) = 1.34701$, and the point (x_0, y_0) is located near the wall, $x_0 = 0.0162473$, $y_0 = -0.0212828$.

Similarly, a logarithmic error $e^l(x, y)$ is given by

$$e^l(x, y) = \log_{10} \frac{|u_{\text{SUPG}}(x, y)|}{|u_{\text{LS}}(x, y)|},$$

and the maximum logarithmic error near the wall is defined as

$$e_{\max}^l = \max_{(x, y) \in D} |e^l(x, y)|.$$

The maximum logarithmic error for the LS reconstruction is

$$|e^l(x_1, y_1)| = 2.49593, \quad x_1 = 0.274632, \quad y_1 = -0.059519,$$

where $u_{\text{SUPG}}(x_1, y_1) = 0.262319$, $u_{\text{LS}}(x_1, y_1) = 0.00029709$. Hence the LS reconstruction near the wall may lose several orders of magnitude, degrading to unacceptable accuracy on stretched meshes. Let us also note that the weighting (3) does not improve the results of the LS method, as it has been discussed in [Petrovskaya (2007)].

It has been demonstrated by Mavriplis [2003] that the poor accuracy of the LS approximation may cause serious problems resulting in an oscillating, and inaccurate solution. Thus the issue of an accurate LS reconstruction on stretched meshes

requires more careful study. We discuss below the impact of remote points on the results of a LS reconstruction.

3. The Impact of Grid Geometry on the Accuracy of a LS Reconstruction: r-outliers

In the statistics, a LS problem statement usually requires one to fit a range of observations to the properties of a given function. Suppose that we get data \mathbf{U} by mistake at some points which lie beyond an actual solution range of interest. Such observations are called outliers [Moore and McCabe (1999); Neter *et al.* (1985); Press *et al.* (1996)], and it is well known in the statistics that being included in the data set the outliers may seriously worsen the results of a LS procedure [Agee and Turner (1982); Neter *et al.* (1985)].

The detection of outliers in the statistics is based on the assumption that, for a given function $U(\mathbf{r}), \mathbf{r} = (x, y)$, an observation may be extreme with respect to its \mathbf{r} value or its U value, or both [Neter *et al.* (1985)]. In this section we demonstrate how the data that are outliers with respect to the \mathbf{r} value may affect the accuracy of a LS reconstruction. Such outliers are called \mathbf{r} -outliers further in the text.

Consider a set of points given by $P_1 = (-\Delta x, 0), P_2 = (0, \Delta y), P_3 = (\Delta x + \alpha, \alpha)$ and $P_4 = (0, -\Delta y)$, whose the points P_i can be thought of as nodes of an unstructured computational grid that has only two grid cells [see Fig. 3(a)]. The parameter $\alpha \in [0, \infty)$ allows one to control a geometry of the domain. The value $\alpha = 0$ corresponds to the regular geometry of a rectangle, while increasing α makes the grid cell (P_2, P_3, P_4) more distorted.

Let us use the LS method to reconstruct a given data set $U(x, y)$ at the origin $P_0 = (0, 0)$, assuming weights $\sigma_i = 1, \forall i = 1, \dots, 4$. Consider the approximation (1) with linear basis functions,

$$u_{LS}(x, y) = u_0 + u_1(x - x_0) + u_2(y - y_0). \tag{8}$$

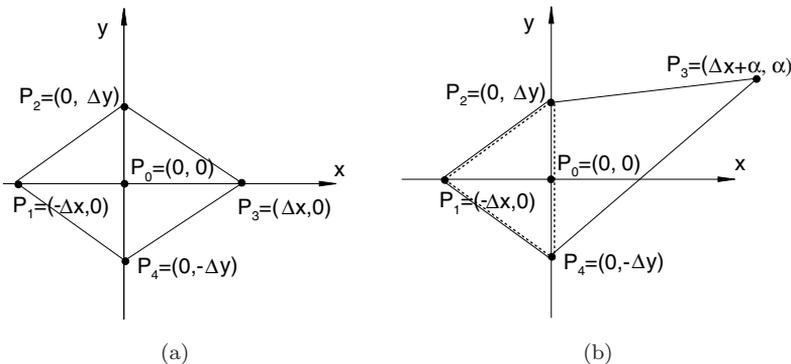


Fig. 3. A stencil for the LS reconstruction over a two-cell computational grid. The grid distortion is controlled by the parameter α . (a) A regular geometry, $\alpha = 0.0$. (b) A distorted geometry, $\alpha > 0$. A stencil for a three-point LS reconstruction is shown as dashed lines.

Let the parameter $\alpha = 0.0$, so that point P_3 is located at $(\Delta x, 0)$. For the regular geometry of Fig. 3(a) the matrix \mathbf{D}^{-1} is a diagonal matrix,

$$\mathbf{D}^{-1} = \begin{bmatrix} \frac{1}{4} & 0 & 0 \\ 0 & \frac{1}{2\Delta x^2} & 0 \\ 0 & 0 & \frac{1}{2\Delta y^2} \end{bmatrix},$$

and the reconstruction (8) is given by

$$u_{\text{LS}}(x, y) = \frac{U_1 + U_2 + U_3 + U_4}{4} + \frac{(U_3 - U_1)}{2\Delta x}x + \frac{(U_2 - U_4)}{2\Delta y}y,$$

where U_i , $i = 1, \dots, 4$, are the known solution values at points P_i . Hence, the LS procedure provides second-order accurate approximation to both the function $U(x, y)$ and the gradient $(\partial U(x, y)/\partial x, \partial U(x, y)/\partial y)$ at point P_0 .

Let us distort the configuration of Fig. 3(a) by increasing the value α [see Fig. 3(b)]. Now the matrix \mathbf{D}^{-1} is not diagonal any longer, and the solution of (5) results in a set u_k , where each fitting parameter u_k can be presented as

$$u_k = \sum_{l=1}^4 C_{kl}(\{P\})U_l, \quad k = 0, 1, 2.$$

For instance, the LS approximate solution at point P_0 is $u(P_0) \equiv u_0 = C_{01}U_1 + C_{02}U_2 + C_{03}U_3 + C_{04}U_4$. The coefficients C_{0l} , $l = 1, \dots, 4$, are defined as follows:

$$C_{01} = \frac{1}{d}(2\alpha^2\Delta^2y + 6\alpha\Delta x\Delta^2y + 4\Delta^2x\Delta^2y),$$

$$C_{02} = \frac{1}{d}(\alpha^2(\Delta^2x + 2\Delta^2y - \Delta x\Delta y) + 2\alpha(2\Delta x\Delta^2y - \Delta^2x\Delta y) + 4\Delta^2x\Delta^2y),$$

$$C_{03} = \frac{1}{d}(2\alpha\Delta x\Delta^2y + 4\Delta^2x\Delta^2y),$$

$$C_{04} = \frac{1}{d}(\alpha^2(\Delta^2x + 2\Delta^2y + \Delta x\Delta y) + 2\alpha(2\Delta x\Delta^2y + \Delta^2x\Delta y) + 4\Delta^2x\Delta^2y),$$

where $d = 16\Delta^2x\Delta^2y + 16\alpha\Delta x\Delta^2y + 6\alpha^2\Delta^2y + 2\alpha^2\Delta^2x$. It can be seen from the equations above that the LS method does not keep the order of approximation, as the solution reconstruction is not second-order accurate for arbitrary $\alpha > 0$.

Another important result is that for $\alpha \rightarrow \infty$ we still have an impact of remote point P_3 on the results of the LS reconstruction. Namely, in the case of an infinitely distant point, we have the following approximation of the data U at point P_0 :

$$u_{\text{LS}}(P_0) \equiv u_0 \rightarrow \frac{1}{4}U_1 + \frac{1}{4}U_2 + \frac{1}{2}U_4, \quad \text{as } \alpha \rightarrow \infty.$$

Meanwhile, if we removed point P_3 from the reconstruction stencil and considered the LS approximation based on a three-point stencil capturing points P_1 , P_2 and

P_4 [see Fig. 3(b)], we would obtain

$$u_{\text{LS}}(P_0) \equiv u_0 = \frac{1}{2}(U_2 + U_4),$$

where the linear interpolation between points P_2 and P_4 keeps the order of approximation.

A straightforward conclusion arising from the above consideration and supported by many authors [Barth (1991); Mavriplis (2003); Ollivier-Gooch and Van Altena (2002); Petrovskaya (2004)] is that all distant points should be removed from a reconstruction stencil. From a LS method viewpoint, an infinitely remote point in a geometric configuration is an extreme example of an \mathbf{r} -outlier. Obviously, the \mathbf{r} -outliers can be efficiently eliminated by means of an inverse distance weighting procedure (3) applied to the reconstruction stencil. However, in many real-life computations, such as the NACA 0012 test case considered in a previous section, a weighted LS method still results in an inaccurate reconstruction. Thus a more coherent definition of a remote point in a LS stencil should be provided prior to a decision to eliminate a given point from the geometric configuration. For this purpose we need to take into account data which are outlying with respect to its U value (U -outliers). That will be done in the next section.

4. The Impact of Data on the Accuracy of a LS Reconstruction: U -Outliers

The most important difference between \mathbf{r} -outliers and U -outliers is that the definition of U -outliers depends on a given data set under consideration. An outlier with respect to the U value is an observation that is numerically, but not geometrically, distant from the rest of the data [Neter *et al.* (1985)]. In other words, a point $P_i = (x_i, y_i)$, where $U(P_i)$ should be treated as an outlier for approximation of a given function $U(x, y)$, can belong to a correct range of observations if another data set is considered. Thus we need to introduce a new definition of a remote point P_r that will be based on the solution properties in order to take into account U -outliers in a given data set.

Let us evaluate the intrinsic lengths l_x and l_y from the requirement

$$\frac{U(l_x, 0) - U(0, 0)}{l_x} \sim 1, \quad \frac{U(0, l_y) - U(0, 0)}{l_y} \sim 1. \quad (9)$$

The point $P_i = (x_i, y_i)$ is then considered as a remote point in the reconstruction stencil, if we have at least one of the evaluates

$$x_i \gg l_x \quad \text{or} \quad y_i \gg l_y. \quad (10)$$

The above definition implies that only those points that lie beyond the actual intrinsic domain defined by (10) should be removed from the stencil. Apparently, if the distance $|\mathbf{r}_{0i}| \rightarrow \infty$, then the above conditions are true, and the point P_i is an outlier (see the previous section). However, the definition (10) may not agree with a simple geometric evaluation of P_i . The distance $|\mathbf{r}_{0i}|$ between the origin P_0 and

a given point P_i can be $|\mathbf{r}_{0i}| \gg d_0$, where d_0 is a length scale in the problem. (For instance, d_0 is the edge length, if we consider a grid edge which the origin P_0 belongs to.) Nevertheless, the point P_i may still belong to the domain (10) and should not be eliminated from the stencil. Another case is that a small distance $|\mathbf{r}_{0i}|$ may be beyond the solution domain of interest (10) if the lengths l_x and l_y are small, so that the point P_i should be eliminated from a stencil, though it can be located close to the origin.

The following test case illustrates the problem of U -outliers. Consider unit square G with vertices $P_1 = (0, 0)$, $P_2 = (0, 1)$, $P_3 = (1, 0)$, $P_4 = (1, 1)$. We also use the edge midpoints $P_5 = (0.5, 0)$, $P_6 = (1, 0.5)$, $P_7 = (0.5, 1)$, $P_8 = (0, 0.5)$ to reconstruct the solution at point P_0 (see domain I in Fig. 4). In our computations the origin $P_0 = (x_0, y_0)$ is chosen as $x_0 = 0.42, y_0 = 0.17$ to make a reconstruction stencil asymmetric.

We now consider a simple quadratic function,

$$U(x, y) = (y - ax)^2, \tag{11}$$

where $a \in [1, 200]$. Let us use the linear LS approximation (8) to reconstruct the solution at point P_0 . We are interested in both the solution and gradient errors computed at point P_0 . The solution error is defined as

$$e(x_0, y_0) = \frac{|U(x_0, y_0) - u_{\text{LS}}(x_0, y_0)|}{|U(x_0, y_0)|}, \tag{12}$$

where $u_{\text{LS}}(x_0, y_0)$ is a linear LS reconstruction at point P_0 . The gradient error is given by

$$e_{\nabla}(P_0) = \frac{\|\nabla U(x, y) - \nabla u_{\text{LS}}(x, y)\|_{|P_0}}{\|\nabla U(x, y)\|_{|P_0}}, \tag{13}$$

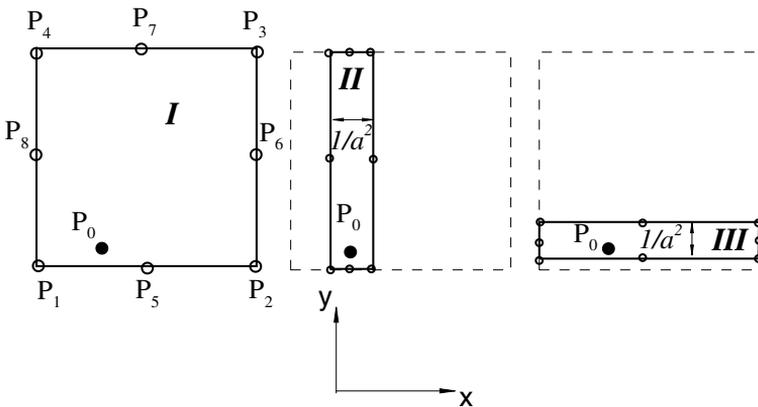


Fig. 4. The advection test case geometry. A stencil for the LS reconstruction consists of points P_1 through P_8 . Originally, the points are located in domain I. A domain contraction along the x axis (domain II) leads to a more accurate LS reconstruction, while a domain contraction along the y axis (domain III) does not reduce the reconstruction error.

Table 1. The linear LS approximation of a quadratic function (11). The solution error is e^I , e^{II} and e^{III} in domains I, II and III, respectively. The gradient error is e_{∇}^I , e_{∇}^{II} and e_{∇}^{III} in domains I, II and III, respectively.

a	e^I	e_{∇}^I	e^{II}	e_{∇}^{II}	e^{III}	e_{∇}^{III}
5.0	1.30744	0.362694	0.025881	0.162694	1.21834	0.200415
10.0	1.15244	0.116625	0.00527625	0.0799007	1.11541	0.197692
50.0	1.04974	0.176188	1.83753e-04	0.0157657	1.04347	0.192024
100.0	1.03807	0.183361	4.52328e-05	0.00786995	1.03501	0.19125
200.0	1.03233	0.186926	1.12229e-05	0.00393177	1.03081	0.190862

where ∇ is a formal notation for the gradient vector, $\nabla = (\partial/\partial x, \partial/\partial y)$, and the norm $\|\nabla g(x, y)\|$ is defined as

$$\sqrt{\left(\frac{\partial g(x, y)}{\partial x}\right)_{|P_0}^2 + \left(\frac{\partial g(x, y)}{\partial y}\right)_{|P_0}^2}.$$

Since the solution $U(x, y)$ depends linearly on the parameter a , both errors are scaled by the value of the exact solution at point P_0 or the solution gradient at P_0 , respectively. The results for a linear LS reconstruction are shown in Table 1, where the results of the LS approximation are compared for various values of a . The solution error is denoted as e^I for a reconstruction stencil defined by the original geometry I. The gradient error is denoted as e_{∇}^I .

We now evaluate the size of the intrinsic domain (10) in the problem. For the function (11) we have

$$\frac{U(l_x, 0) - U(0, 0)}{l_x} = a^2 l_x, \quad \frac{U(0, l_y) - U(0, 0)}{l_y} = l_y,$$

so that the estimate (9) results in $l_x \sim 1/a^2$, $l_y \sim 1$. In other words, the ratio $l_y/l_x \sim a^2$ and the domain (10) should be $l_x \ll l_y$ as $a \gg 1$. Thus all points belonging to original domain I in Fig. 4 can be considered as outliers and the original domain should be transformed to eliminate the outliers from a reconstruction stencil.

Let us transform the domain I to meet the conditions (10). The new coordinates of the points in the transformed domain II are defined as follows (see Fig. 4)

$$\begin{aligned} P_1 &= \left(x_0 - \frac{x_0}{a^2}, 0\right), & P_2 &= \left(x_0 - \frac{x_0}{a^2}, 1\right), \\ P_3 &= \left(x_0 + \frac{1-x_0}{a^2}, 0\right), & P_4 &= \left(x_0 + \frac{1-x_0}{a^2}, 1\right), \\ P_5 &= \left(x_0 - \frac{2x_0-1}{2a^2}, 0\right), & P_6 &= \left(x_0 + \frac{1-x_0}{a^2}, \frac{1}{2}\right), \\ P_7 &= \left(x_0 - \frac{2x_0-1}{2a^2}, 1\right), & P_8 &= \left(x_0 - \frac{x_0}{a^2}, \frac{1}{2}\right). \end{aligned} \tag{14}$$

The results of a LS reconstruction in the transformed domain II are shown in Table 1. It can be seen from the table that the transformation (14) results in a smaller solution error and gradient error.

Let us note that the diameter of the transformed domain II is smaller than that of the original domain I. Hence, we can expect better results of the LS approximation in domain II, as the accuracy of the approximation depends generally on the domain size. However, one important result about the evaluation of an intrinsic domain is that the transformation (14) is not equivalent to a simple reduction in the domain size. Let us apply a linear mapping similar to (14) to the y -coordinate of the original domain. We now require that

$$\begin{aligned}
 P_1 &= \left(0, y_0 - \frac{y_0}{a^2}\right), & P_2 &= \left(1, y_0 - \frac{y_0}{a^2}\right), \\
 P_3 &= \left(1, y_0 + \frac{1 - y_0}{a^2}\right), & P_4 &= \left(0, y_0 + \frac{1 - y_0}{a^2}\right), \\
 P_5 &= \left(\frac{1}{2}, y_0 - \frac{y_0}{a^2}\right), & P_6 &= \left(1, y_0 - \frac{2y_0 - 1}{2a^2}\right), \\
 P_7 &= \left(\frac{1}{2}, y_0 + \frac{1 - y_0}{a^2}\right), & P_8 &= \left(0, y_0 - \frac{2y_0 - 1}{2a^2}\right).
 \end{aligned} \tag{15}$$

The transformed domain III is shown in Fig. 4, and the results of a linear LS reconstruction in the new domain are displayed in Table 1. The diameter of domain III is the same as that of domain II. Nevertheless, it can be seen from the table that the mapping (15) does not reduce the LS reconstruction error. The solution error and the gradient error remain the same as the error computed for the LS reconstruction in the original domain I, because contraction along the y -axis still remains grid nodes beyond an intrinsic domain.

The above example demonstrates that the detection of the remote points in the reconstruction stencil should not rely upon the geometric shape of the stencil only, as numerically distant points can appear in a reconstruction stencil. In the latter case a weighting procedure used to eliminate remote points from the stencil will not be efficient, as the choice of weight coefficients in (3) is not related to the properties of a solution function. Hence, if all stencil points lie beyond the intrinsic domain, the weighting procedure will not result in a more accurate solution, as a uniform weighting of all stencil points will leave outliers in the reconstruction stencil.

5. Concluding Remarks and Future Work

In this work we have considered the problem of function approximation by the LS method on unstructured meshes with arbitrary grid cell geometry. The problem statement comes from aerodynamic applications where computational grid generation may result in grid cells with a very high cell aspect ratio. Those almost degenerate grid geometries present a serious difficulty when a solution function should be reconstructed by the LS method. It has been shown in the paper that the accuracy of a LS reconstruction can degrade to an unacceptable limit on irregular grids.

A concept of outliers that we have applied to a LS problem allows one to explain the poor performance of the LS method on stretched meshes. We have discussed the two types of outliers that may arise in reconstruction stencils on stretched meshes. Outliers of the first type are related to geometrically distant points. Those outliers can be easily removed from the stencil by a standard inverse distance weighting procedure. Outliers of the second type are related to the solution data and represent numerically distant points. Those data often appear in a boundary layer solution that has a strong solution gradient and requires a stretched mesh for its proper resolution. They are difficult to detect and eliminate from a stencil, as the location of such outliers is not directly related to the distance to a given stencil point.

Finally, let us note that the problem of outliers in a LS reconstruction stencil should be coupled with a grid generation problem in future work. Grid generation and solution grid adaptation is essentially based on solution information such as solution error estimates, and that information may help one to recognize U outliers in the reconstruction stencil and to choose stencil points in an optimal way. On the other hand, the well-developed methods of modern statistical analysis (see e.g. [Neter *et al.* (1985)]) can be applied to the problem to eliminate outliers from a reconstruction stencil. That may be considered as another topic of future research.

Acknowledgments

This research was supported by The Boeing Company under contract 66-ZB-B001-10A-533. The author is grateful to Dr. F. T. Johnson for stimulating discussions and for providing data for the NACA 0012 test case.

References

- Agee, W. S. and Turner, R. H. [1982] Robust range measurement preprocessing. Technical Report, National Range Operations Directorate, White Sands Missile Range NM.
- Aftosmis, M. J., Gaitonde, D. and Tavares, T. S. [1995] Behaviour of linear reconstruction techniques on unstructured meshes, *AIAA J* **33**(1), 2038–2049.
- Anderson, W. K. [1994] A grid generation and flow solution method for the Euler equations on unstructured grids, *J. Comput. Phys.* **110**(1), 23–38.
- Atluri, S. N. and Shen, S. [2005] The basis of meshless domain discretization: The meshless local Petrov–Galerkin (MLPG) method, *Adv. Comput. Math.* **23**, 73–93.
- Bajaj, C., Ihm, I. and Warren, J. [1993] Higher-order interpolation and least-squares approximation using implicit algebraic surfaces, *ACM Trans. Graph.* **12**(4), 327–347.
- Barth, T. J. [1991] A three-dimensional upwind Euler solver for unstructured meshes, *AIAA 91-1548*.
- Barth, T. J. and Jespersen, D. C. [1989] The design and application of upwind schemes on unstructured meshes, *AIAA 89-0366*.
- Belytschko, T., Krongauz, Y., Organ, D., Fleming, M. and Krysl, P. [1996] Meshless methods: An overview and recent developments, *Comput. Meth. Appl. Mech. Eng.* **139**, 3–47.
- Borouchaki, H., Castro-Diaz, M. J., George, P. L., Hecht, F. and Mohammadi, B. [1996] Anisotropic adaptive mesh generation in two dimensions for CFD, in *5th Int. Conf. Numerical Grid Generation in Computational Field Simulations*, **3**, 197–206.

- Buscaglia, G. S. and Dari, E. [1997] Anisotropic mesh optimization and its application in adaptivity, *Int. J. Numer. Meth. Eng.* **40**, 4119–4136.
- Haselbacher, A. [2006] On constrained reconstruction operators, *AIAA 2006-1274*.
- Hughes, T. J. R. and Brooks, A. [1979] A multidimensional upwind scheme with no cross-wind diffusion, in *Finite Element Methods for Convection Dominated Flows* (ASME, New York).
- Johnson, F. T. [2007] Boeing test cases for a higher order least-squares reconstruction. Personal Communication.
- Kenney, J. F. and Keeping, E. S. [1962] *Linear Regression and Correlation in Mathematics of Statistics*, Pt.1, 3rd edn. (Princeton, New Jersey, Van Nostrand), pp. 252–285.
- Liu, G. R. and Gu, Y. T. [2005] *An Introduction to Meshfree Methods and Their Programming* (Kluwer).
- Mavriplis, D. J. [2003] Revisiting the least-squares procedure for gradient reconstruction on unstructured meshes, *AIAA 2003-3986*.
- Moore, D. S. and McCabe, G. P. [1999] *Introduction to the Practice of Statistics*, 3rd edn. (W. H. Freeman, New York).
- Neter, J., Wasserman, W. and Kutner, M. H. [1985] *Applied Linear Statistical Models* (Richard D. Irwin, Illinois).
- Ollivier-Gooch, C. and Van Altena, M. [2002] A high-order-accurate unstructured mesh finite-volume scheme for the advection–diffusion equation, *J. Comp. Phys.* **181**, 729–752.
- Petrovskaya, N. B. [2004] The choice of weight coefficients for the least-squares gradient approximation, *Matematicheskoe Modelirovanie* **16**(5), 83–93.
- Petrovskaya, N. B. [2007] Some issues of implementation of higher order schemes in computational aerodynamics, *Boeing Technical Report 66-ZB-B001-10A-533* (The Boeing Company, Seattle, USA).
- Petrovskaya, N. B. and Wolkov, A. V. [2007] The issues of solution approximation in higher order schemes on distorted grids, *Int. J. Comput. Meth.* **4**(2), 367–382.
- Pratt, V. [1987] Direct least squares fitting of algebraic surfaces, *Comput. Graph.* **21**(3), 145–152.
- Press W. H. *et al.* [1996] *Numerical Recipes in Fortran 77: The Art of Scientific Computing* (Cambridge University Press).
- Sonar, T. [2005] Difference operators from interpolating moving least squares and their deviation from optimality, *M2AN* **39**(5), 883–908.
- Unser, M., Aldroubi, A. and Eden, M. [1995] Enlargement or reduction of digital images with minimum loss of information, *IEEE Trans. Image Process.* **4**, 247–258.