INTERNATIONAL JOURNAL FOR NUMERICAL METHODS IN FLUIDS
Int. J. Numer. Meth. Fluids 2006; **52**:843–865
Published online 16 March 2006 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/fld.1198

Theoretical analysis for achieving high-order spatial accuracy in Lagrangian/Eulerian source terms

David P. Schmidt*,†

Room 219 Engineering Laboratory, Department of Mechanical and Industrial Engineering, University of Massachusetts, 160 Governors Dr., Amherst, MA 01003, U.S.A.

SUMMARY

In a fully coupled Lagrangian/Eulerian two-phase calculation, the source terms from computational particles must be agglomerated to nearby gas-phase nodes. Existing methods are capable of accomplishing this particle-to-gas coupling with second-order accuracy. However, higher-order methods would be useful for applications such as two-phase direct numerical simulation and large eddy simulation. A theoretical basis is provided for producing high spatial accuracy in particle-to-gas source terms with low computational cost. The present work derives fourth- and sixth-order accurate methods, and the procedure for even higher accuracy is discussed. The theory is also expanded to include two- and three-dimensional calculations. One- and two-dimensional tests are used to demonstrate the convergence of this method and to highlight problems with statistical noise. Finally, the potential for application in computational fluid dynamics codes is discussed. It is concluded that high-order kernels have practical benefits only under limited ranges of statistical and spatial resolution. Additionally, convergence demonstrations with full CFD codes will be extremely difficult due to the worsening of statistical errors with increasing mesh resolution. Copyright © 2006 John Wiley & Sons, Ltd.

KEY WORDS: two-phase and multiphase flows; numerical methods; stochastic particle methods; particle methods and lattice-gas methods

INTRODUCTION

Usually, large eddy simulations (LES) and direct numerical simulations (DNS) of single-phase flows employ numerical schemes with a high formal order of accuracy. Spectral methods or high-order finite difference schemes are often favoured for providing the best use of available computational resources. The numerical techniques for high-accuracy single-phase calculations are well established, but for two-phase Lagrangian/Eulerian calculations, such high-order

*Correspondence to: D. P. Schmidt, Room 219 Engineering Laboratory, Department of Mechanical and Industrial Engineering, University of Massachusetts, 160 Governors Dr., Amherst, MA 01003, U.S.A.
†E-mail: schmidt@ecs.umass.edu

Received 27 July 2005
Revised 3 January 2006
Accepted 8 January 2006

accuracy constructions have not been fully accomplished. Unfortunately, there is no established numerical framework that is known to give better than second-order convergence for coupled Eulerian gas/Lagrangian two-phase simulations. The current work is dedicated to resolving one of the difficulties of achieving high-order accuracy in two-phase finite difference simulations.

In order to analyse the accuracy of two-phase schemes, it is helpful to divide the simulation into composite parts. Each of these parts can be studied independently, and it is expected that the overall scheme will possess accuracy commensurate with its components. This approach was used successfully by Are *et al.* [1] to establish second-order accuracy in a Lagrangian/Eulerian spray code. For the current discussion, the parts of the two-phase scheme will be divided as follows:

(1) The gas-phase solver.
(2) integration of the ODEs for particle position and momentum.
(3) interpolation of gas-phase properties to particle locations.
(4) summation of dispersed-phase sources.
(5) spray sub-models.

Item 1 is considered to be well understood already. Item 2 is also not expected to be extremely difficult. Many computations use high-order Runge–Kutta schemes, for example, to move particles and integrate the acceleration equation. The third item, interpolation of gas-phase properties to the particle locations, is another part of the simulation that has already been treated with higher order accuracy. Interpolation is required because the particles in the calculation are not located exactly on cell vertices. Some scheme is required to take the gas-phase quantities known at surrounding nodes and estimate what the values of these quantities would be at the particle locations. Linear interpolation was shown to be adequate for second-order accuracy by Aggarwal *et al.* [2]. Higher-order methods are available for this 'gas-to-particle' interpolation. For example, Miller and Bellan [3] used fourth-order Lagrange polynomial interpolation for their DNS scheme.

It is the fourth item which presently causes difficulties and which the current work addresses. In this part of a two-phase code, the effect of the particle on the gas is calculated. Particles must exchange momentum, mass, and energy with the gas phase. The rate of exchange varies spatially and shows up as a source term in the gas-phase equations. This source term must be calculated at each node location, which is typically done by a weighted summation of contributions from particles in the vicinity. For example, Miller and Bellan [3] uses a variant of Aggarwal *et al.*'s cloud-in-cell weighting [2]. Are *et al.* [1] analysed particle-to-gas coupling with the goal of achieving second-order accuracy. Their results provided a method of deriving the accuracy of weighting schemes and were contrary to some of Aggarwal *et al.*'s observations. The method of Are, Hou, and Schmidt can be extended to fourth-, sixth-, or even eighth-order accuracy and can be used in one, two, or three dimensions. This extension is the goal of the present work: to use analytical methods to derive these high-order schemes.

One of the challenges to any formally high-order scheme is to succeed in the face of large statistical errors. The statistical difficulties of Lagrangian particle tracking have been discussed by Miller and Bellan [3], Xu and Pope [4], and Subramaniam [5]. Subramaniam's work provides a theoretical framework, while Xu and Pope combined numerical tests with statistical analysis of a coupled Lagrangian/Eulerian calculation for turbulent single-phase flow. An idea from Miller and Bellan was that source terms produced large amounts of numerical

noise, and thus they used smoothing to ameliorate the errors. Subramaniam and Haworth fitted multidimensional functions to local source terms in order to reduce statistical noise in multi-dimensional PDF-based turbulence simulations using a least-squares approach [6]. This approach, though for single-phase simulations, recognized the statistical noise inherent in mixed Monte-Carlo Lagrangian/Eulerian solutions. The least-squares fitting demonstrated better resistance to statistical noise than a direct kernel approach.

The contribution of the present work is to derive methods that could potentially achieve high formal order of accuracy. Much of the analysis below is done for a single dimension, though the results are also applicable to multiple dimensions. The analytical extension to three dimensions is shown to be extremely simple. One- and two-dimensional kernels are tested using a problem that admits a closed-form solution. The results verify the formal order of accuracy of each kernel and also show the conditions under which high-order kernels are useful. A least-squares method, known to be tolerant of statistical noise, is compared to the higher-order kernels. Finally, the conclusions discuss how the results of this paper may be used to construct a CFD code with high spatial order of accuracy.

SPATIAL ERROR IN PARTICLE-TO-GAS COUPLING

The gas-phase Navier–Stokes equations in two-phase calculations include a volumetric source term, f, that represents the effect of the dispersed phase on the velocity field for a given time step. For a finite-difference scheme, the discretized equations thus require an estimate of f at each node location x. Spatial error occurs, due to the fact that the particles are not located at the gas-phase node, but are instead scattered in space. Statistical error also occurs due to the finite number of particles. (The current study is only dedicated to understanding errors in finite-difference calculations. In a finite-volume approach a particle is either in or out of the volume, resulting in statistical error and no spatial error.) This problem has been studied in a general form by mathematicians such as Bartlett [7] and Parzen [8]. Parzen studied second-order accurate kernels and emphasized statistical error. Rosenblatt [9] and Parzen [8] limited themselves to non-negative kernels, which will be shown below, to limit their analyses to second-order accurate methods. Bartlett, using different methods than the current work, analysed both second- and fourth-order kernels. Bartlett stopped at fourth-order, saying it was doubtful that higher order accurate methods would be useful, since statistical error was so significant. The current paper goes up to sixth order and explains how to go further, and to generalize the results to two and three dimensions. The beginning of the present derivation repeats Are *et al.* [1], who stopped at second order.

The following analysis will examine spatial errors in the volumetric source term f_n at a gas-phase node calculated from a set of n nearby particles. Each particle, numbered by the index i, is located at x_{p}^i and contributes an amount s^i to the value of f_n. Note that s^i has dimensions of mass, momentum, or energy but f_n has units of the respective quantity per volume, which is typical of a finite-difference scheme, and that each computational particle may represent a number of physical particles.

The exact value of a particle's contribution depends on the size, the velocity, and number of physical particles represented by each computational particle. However, f varies in space, and so none of the values of $f(x_{\mathrm{p}}^i)$ exactly correspond to $f(x)$. A spatial error occurs due to the distance between each particle and the gas-phase node as well as the variation in f.

To begin the analysis, the value of f_n is calculated from a weighted sum of the contributions of the particles. The weighting factor, W, depends on the difference between the node location, x and the particle location, x_{p}^i. The total value estimated for f is denoted as f_n.

$$f_n = \sum_{i=1}^{n} W(x_{\mathrm{p}}^i - x) s(x_{\mathrm{p}}^i) \tag{1}$$

In real calculations, the summation over a finite number of particles gives rise to a statistical error. In order to isolate the spatial error, an infinite number of particles is used here. This assumption allows the analysis to proceed based on the expectation of the dispersed-phase source, but discards information about statistical error. Thus, it is assumed that the numerical estimate $\langle f \rangle$, based on an infinite number of particles, includes only spatial errors (temporal errors are beyond the scope of the current work). The problem of statistical noise will be visited later in the paper.

The summation of Equation (1) is then recast as an integral over space around the node. The limits of the integral would depend on the width of the smoothing kernel, but are generally one to two gas-phase cell widths. The superscript on the particle location is dropped from this point onward, since the individual particles are no longer being considered. The effect of the particle on the gas-phase source term is given in the following equation:

$$\langle f(x) \rangle = \int W(x_{\mathrm{p}} - x) f(x_{\mathrm{p}}) \, \mathrm{d}x_{\mathrm{p}} \tag{2}$$

Despite the fact that the source term is normally considered at discrete locations, the exact function, f, is smooth and may be estimated with a Taylor series expansion. The value of f can be estimated at each particle location a small distance $(x_{\mathrm{p}} - x)$ from the node, as given in the following equation

$$f(x_{\mathrm{p}}) = f(x) + \sum_{n=1}^{\infty} \frac{1}{n!} \left. \frac{\partial^n f}{\partial x^n} \right|_x (x_{\mathrm{p}} - x)^n \tag{3}$$

This expression uses the Taylor series to quantify the error that occurs from the use of particles in the vicinity of gas-phase node. Errors are due to the variation in the source terms and the finite distance between the node and the particles. Equation (3) can be substituted into the integrand in Equation (2) to relate the particle contributions to the estimate $\langle f(x) \rangle$.

$$\langle f(x) \rangle = \int W(x_{\mathrm{p}} - x) f(x) \, \mathrm{d}x_{\mathrm{p}} + \int W(x_{\mathrm{p}} - x) \sum_{n=1}^{\infty} \frac{1}{n!} \left. \frac{\partial^n f}{\partial x^n} \right|_x (x_{\mathrm{p}} - x)^n \, \mathrm{d}x_{\mathrm{p}} \tag{4}$$

The $f(x)$ term can be pulled out of the integral in the first term on the right-hand side. Further, the derivatives in the right-most term are evaluated at the gas-phase node location and are not a function of the variable of integration. Hence, the order of the summation and integration can be exchanged.

$$\langle f(x) \rangle = f(x) \int W(x_{\mathrm{p}} - x) \, \mathrm{d}x_{\mathrm{p}} + \sum_{n=1}^{\infty} \frac{1}{n!} \left. \frac{\partial^n f}{\partial x^n} \right|_x \int W(x_{\mathrm{p}} - x)(x_{\mathrm{p}} - x)^n \, \mathrm{d}x_{\mathrm{p}} \tag{5}$$

The left side of Equation (5) is the expected value of the source term used in the gas-phase Navier–Stokes equations. By comparing the exact answer $f(x)$, and the expected numerical value $\langle f(x) \rangle$ for Taylor polynomials of increasing degree, one can note the error.

For example, Equation (5) gives a constraint for the weighting function so that if f is a constant, the gas-to-particle coupling gives the right answer. Equation (5) would reduce to

$$1 = \int W(x_p - x)\,dx_p \tag{6}$$

Equation (6) gives a consistency constraint for kernels which is easy to satisfy. If one uses the simplest possible scheme, then W is a constant close to x and zero outside of the integration area. This method is referred to as the 'nearest node' weighting function and may be defined as

$$W(x_p - x) = \begin{cases} \dfrac{1}{\Delta x} & \text{if } |x_p - x| \leqslant \ell \\ 0 & \text{if } |x_p - x| > \ell \end{cases} \tag{7}$$

This definition presumes that the width of the non-zero portion of the kernel is 2ℓ. The kernel need not be the same width as a gas-phase cell.

The starting point of attaining higher-order schemes is Equation (5). Applying Equation (5) to increasingly higher-order polynomial forms of f can produce further constraints for achieving a higher order of accuracy. Consider next, a first-order variation in f, of the form

$$f(x) = A + Bx \tag{8}$$

Substituting this function on both sides of Equation (5) and using the consistency constraint, Equation (6), gives the following:

$$\int W(x_p - x)(x_p - x)\,dx_p = 0 \tag{9}$$

To satisfy this constraint, it is sufficient that the kernel be a symmetric function. If $W(x - x_p) = W(x_p - x)$ then Equation (9), the second constraint is satisfied. This gives second-order global accuracy (see Reference [1] for a demonstration of this assertion with a spray calculation) The use of a second-order polynomial of the form

$$f(x) = A + Bx + Cx^2 \tag{10}$$

gives rise to the third constraint

$$\int W(x_p - x)(x_p - x)^2\,dx_p = 0 \tag{11}$$

Since $(x_p - x)^2$ is always positive, the kernel must change signs over the domain to satisfy Equation (11). No all-positive kernel can satisfy this constraint; *a kernel must change signs in order to exceed second-order accuracy*.

The above process can be continued for higher-order accuracy. For generality, no specific kernel width is assumed in the above derivations. The truncation error is proportional to the kernel width raised to the order of accuracy of the scheme. Narrower kernels offer a higher spatial accuracy, but wider kernels have the advantage of a larger sample size. The kernel widths could be significantly larger than the gas-phase node spacing in order to reduce statistical error. Given the notoriously slow convergence of Monte-Carlo schemes, the larger

kernel width may be advantageous. This must be balanced, of course, with the need to resolve important physical scales in the calculation. Note that the property of strong conservation, that the amount of momentum, energy, or mass lost by the particle phase is exactly gained by the gas phase, cannot be preserved with higher order kernels or kernels that overlap each other from neighbouring cells.

For simplicity, the dummy variable in the integration is changed to y, where y is the distance between a node and the particle. The limits of integration will simply be assumed to be from $-\ell$ to ℓ, representing the 'support' of the kernel. For a kernel that is one-cell wide, (the same width as nearest-node) the value of ℓ is $\Delta x/2$. For the simplest overlapping kernels, such as the cloud-in-cell approach or the two-dimensional equivalent used by Aggarwal *et al.* [2], ℓ is Δx. Hence Equation (11) becomes slightly simpler, as shown in the following equation:

$$\int_{-\ell}^{\ell} W(y) y^2 \, \mathrm{d}y = 0 \tag{12}$$

This relationship can be satisfied by any number of kernels. The logical choice would be an even function of y that can change signs. Similar to the work of Nordin [10], one might chose an inverse square weighting. However, this approach has the unfortunate characteristic of being unbounded at $y = 0$. A good choice is an even-order polynomial, as shown in the following equation

$$W(y) = a + by^2 \tag{13}$$

Substituting this equation in Equation (12) and noting that the integral from $-\ell$ to ℓ can be converted to an integral from 0 to ℓ gives the following:

$$\int_{0}^{\ell} (ay^2 + by^4) \, \mathrm{d}y = 0 \tag{14}$$

Evaluation of the integral gives

$$a = -\tfrac{3}{5} b\ell^2 \tag{15}$$

Using Equation (6) and integrating from $-\ell$ to ℓ uniquely determines a and b.

$$\begin{aligned} a &= \frac{1}{\ell}\frac{9}{8} \\ b &= \frac{1}{\ell^3}\frac{-15}{8} \end{aligned} \tag{16}$$

while studying how to estimate densities of discrete quantities, Bartlett [7] also derived Equations (13) and (16). Until now, Bartlett's results have not previously been noted in the multiphase-flow literature, and more importantly, have not been extended to source terms in multiple dimensions, as is required for practical use in CFD. The mathematical analysis used here is intended to be more general.

A plot of the quadratic kernel, as defined by Equations (13) and (16), is given in Figure 1. Note that, as required by Equation (11), the kernel is negative over part of the interval. This

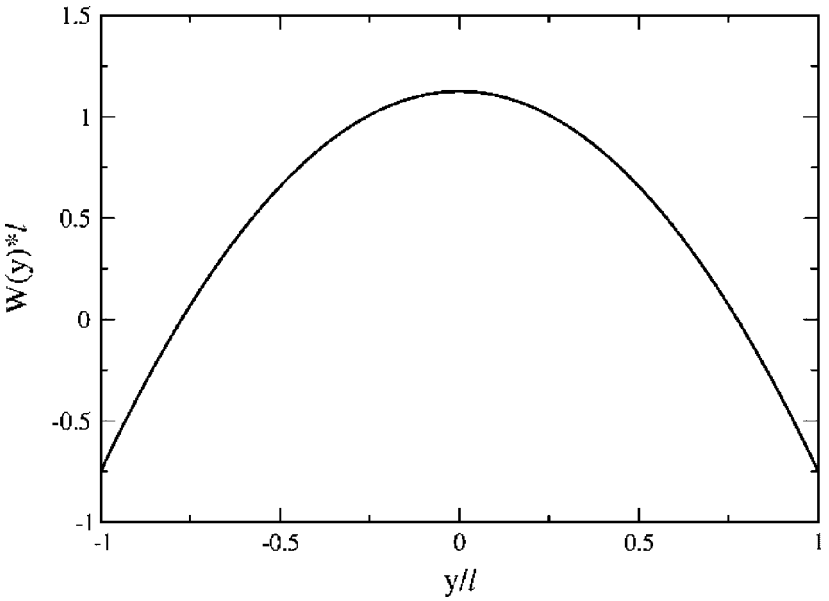

Figure 1. A fourth-order kernel given by Equations (13) and (16).

may seem odd that a source should change sign (e.g. that a droplet evaporating would cause the gas phase to lose vapour), but one should remember that high-order finite differences are often oscillatory in nature. For spatial error, it is not the behaviour any single particle that matters, but rather the agglomerate behaviour of many particles.

This kernel was designed to eliminate the second error term in the summation in Equation (5). Hence, the next leading error term appears to be a third-order term. Thus the above scheme produces at least third-order accuracy. However, a close examination of the next constraint is warranted. This would be

$$\int_{-\ell}^{\ell} W(y) y^3 \, \mathrm{d}y = 0 \tag{17}$$

This constraint is satisfied by any even kernel, and hence the scheme shown in Figure 1 is actually fourth-order accurate. In fact, the use of even kernels automatically eliminates all odd-order error terms. From examination of Equation (5), one can thus conclude that satisfying Equation (6) and

$$\int_{-\ell}^{\ell} W(y) y^m \, \mathrm{d}y = 0 \tag{18}$$

for all integers m from 2 to n provides an exact answer for a source distribution intensity that follows a polynomial of order n or less. Thus, the leading error term for such a kernel is at least order $n+1$. Any even kernel automatically satisfies Equation (18) for odd m. Hence, the leading error term is actually $n+2$ for even values of n. Thus, sixth-order accuracy is

achieved by satisfying Equations (6), (12), and (19).

$$\int_{-\ell}^{\ell} W(y) y^4 \, \mathrm{d}y = 0 \tag{19}$$

The second-order polynomial weighting used in Equation (13) cannot satisfy all three equations (6), (12), and (19). However, a higher-order polynomial could be satisfactory. Using an even kernel automatically satisfies Equations (9) and (17), as well as the corresponding fifth-order equation. Hence, a good choice for a sixth-order kernel is an even, fourth-order polynomial.

$$W(y) = a + by^2 + cy^4 \tag{20}$$

Substituting Equation (20) into Equations (6), (12), and (19) gives the three equations required to solve for the three constants in Equation (20). Again, it simplifies the integration if one uses symmetry to change the integration of the equations to the interval 0 to ℓ. The three resulting equations are written in matrix form in Equation (21).

$$\begin{bmatrix} \ell & \frac{\ell^3}{3} & \frac{\ell^5}{5} \\ \frac{\ell^3}{3} & \frac{\ell^5}{5} & \frac{\ell^7}{7} \\ \frac{\ell^5}{5} & \frac{\ell^7}{7} & \frac{\ell^9}{9} \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} \frac{1}{2} \\ 0 \\ 0 \end{bmatrix} \tag{21}$$

The solution for a, b, and c are given in the following equation:

$$\begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} \frac{225}{128\ell} \\ \frac{-1050}{128\ell^3} \\ \frac{945}{128\ell^5} \end{bmatrix} \tag{22}$$

A plot of this kernel is shown in Figure 2. This kernel will be sixth-order accurate. The procedure can be continued to produce any even order of accuracy. For example, an eighth-order kernel uses a sixth-order polynomial and requires solving for the four unknown coefficients using the additional constraint given in the following equation:

$$\int_{-\ell}^{\ell} W(y) y^6 \, \mathrm{d}y = 0 \tag{23}$$

STATISTICAL ERROR

Statistical errors have been removed from the previous discussion by taking the limit of an infinite number of particles. However, in practical computations, statistical noise levels can be larger than spatial error. The obvious statistical factor is that 'larger samples are better'.

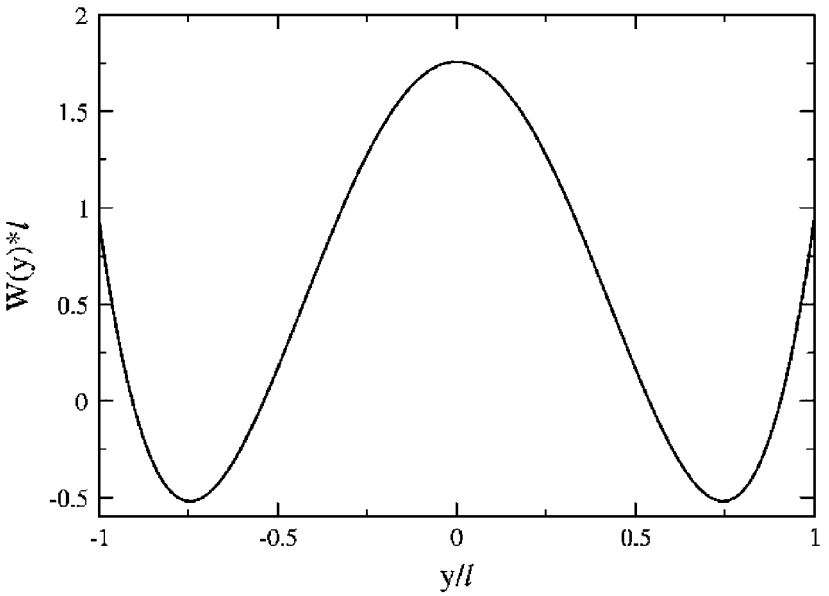

Figure 2. A sixth-order accurate kernel given by Equations (20) and (22).

One way that this can be accomplished is by widening the support of the kernel. The above derivations are all based on a kernel width of 2ℓ. The wider the extent of the kernel, the larger the sample size. However, there is a trade-off with spatial accuracy. With high-order accurate schemes, the spatial error will quickly decrease to a point where it is comparable to the statistical noise level. Applied mathematicians refer to this field of study as 'non-paramentric density estimation' and have studied the errors extensively. Normally the purpose of their analysis is to determine the optimal kernel width. For example, Rosenblatt [9] provides some discussion of statistical error for non-negative kernels. Bartlett [7] offers analysis of second- and fourth-order kernels, which requires the assumption of non-negativity to be dropped. Using a finite sample size, the mean square error can be estimated as a sum of statistical error, due to variance, and the spatial error, which represents a bias. (The word 'bias' is used in the turbulence PDF literature to refer to a source of error that is not considered in the present work, but is due to feedback from particle–particle effects. The dual meaning of the term can be confusing.) As explained by Rosenblatt, this expression is

$$E^2 = \sigma^2[f_n(x)] + |E[f_n(x)] - f(x)|^2 \tag{24}$$

where E^2 is the mean square error, σ^2 is the variance, $E[X]$ represents the expectation of the random variable X, and $f_n(x)$ is the estimated value based on a sample size n. The variance can be approximated by the following equation [7]:

$$\sigma^2[f_n(x)] \approx \frac{f(x)}{n} \int_{-\ell}^{\ell} W^2(y)\,\mathrm{d}y \tag{25}$$

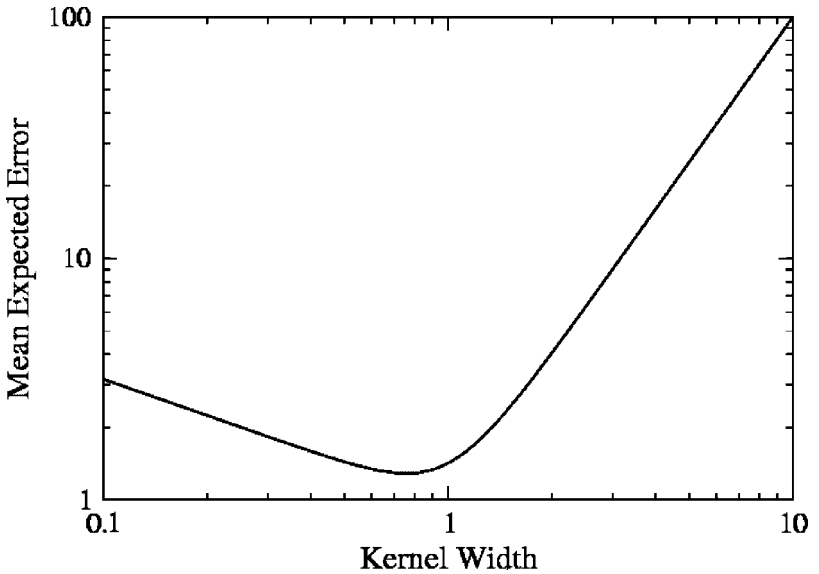

Figure 3. General trend of mean expected error versus resolution predicted by Equation (26). The units of both axes are arbitrary.

For the nearest-node kernel, the mean expected error in a single kernel width would be approximately [7]:

$$E \approx \sqrt{\frac{f}{n\ell} + \frac{1}{4}\frac{\ell^4}{144}\left(\frac{\mathrm{d}^2 f}{\mathrm{d}x^2}\right)^2} \tag{26}$$

The first term under the radical represents the statistical error and the second term is spatial error. As is typical of stochastic simulations, the error decreases with the number of particles in the sample to the negative one-half power. What is particularly unfortunate is that the mesh resolution appears in the denominator of the statistical error term, meaning that the convergence with mesh refinement is conditional on the behaviour of n. Performing refinement while keeping the number of particles per cell constant is not sufficient for demonstrating convergence unless the number of particles is so large, that the statistical error is suppressed. In general, the statistical error will start to eclipse the spatial error, as shown in Figure 3. On the right side of the plot, where the mesh would be coarse, the spatial error is dominant and on the left side of the figure, where the mesh would be fine, the statistical error dominates. In between, there is an optimal resolution (sometimes referred to as 'bandwidth') that minimizes total error [6].

To always be able to demonstrate convergence, the number of particles per cell must increase super-linearly. If one wishes to guarantee that the mean expected error converges at the same rate as spatial error, then the first term under the square root must have the same, or higher power, dependence on mesh resolution. This means that for the nearest-node approach, the number of particles per cell must be proportional to Δx to the negative fifth power. *Thus the total number of particles in a one-dimensional domain must increase with*

Table I. Comparison of the variances of several kernels.

Kernel	Support	Variance
Nearest node	$\left[-\frac{\Delta x}{2}, \frac{\Delta x}{2}\right]$	$\frac{f}{n\Delta x}$
$W(y)=\frac{1}{2\ell}$	$[-\ell,\ell]$	$\frac{1}{2}\frac{f}{n\ell}$
Fourth order	$[-\ell,\ell]$	$\frac{9}{8}\frac{f}{n\ell}$
Sixth order	$[-\ell,\ell]$	$\frac{225}{128}\frac{f}{n\ell}$

The second kernel is the nearest-node kernel, but spread out to the same width as the fourth- and sixth-order kernels.

Δx to the sixth power in order to guarantee numerical convergence. This result may explain why convergence of two-phase CFD results is so difficult to observe. Only with a very large number of particles can one observe the effect of the spatial error term. For example, Are *et al.* [1] used 300 000 particles in their two-dimensional convergence tests, which was apparently sufficiently large to minimize the effect of the first term in Equation (26).

For higher order accurate kernels, the trend is slightly worse. Following Bartlett, the mean expected error for a cell using the fourth-order accurate method is given by the following equation:

$$E \approx \sqrt{\frac{9}{8}\frac{f}{n\ell} + \left(\frac{9}{24}\frac{1}{110}\right)^2 \ell^8 \left(\frac{\mathrm{d}^4 f}{\mathrm{d}x^4}\right)^2} \tag{27}$$

Though the spatial error term is indeed fourth-order accurate, the statistical error term has a slightly higher coefficient. This trend continues with higher order kernels. The variance is summarized for four kernels in Table I. As mentioned in the previous section, the width of the kernels is somewhat arbitrary. The nearest node kernel, as inferred by its name, spans a half node spacing in each direction. The higher-order kernels are applied to twice that spacing. The second kernel in the table is constant, like the nearest-node, but spread out over the same width as the higher-order kernels. One can observe that widening a kernel reduces its variance, and the higher-order kernels tend to have increasing variance. Thus one would expect the higher-order kernels to be more susceptible to statistical noise. In the case of the fourth-order kernel, one would have to increase the number of particles per cell with resolution to the negative ninth-power, and the total number of particles in one-dimension to the tenth power in order to keep statistical noise in check.

Another method of estimating the source term strength is to use least-squares fitting. For example, Subramaniam and Haworth [6], successfully applied this approach to a three-dimensional CFD calculation. Least-squares fits are tolerant of noise and not computationally expensive. The procedure is to fit a polynomial to kernel estimates of the source term at several points around the node of interest. The number of points is arbitrary, so long as it exceeds the minimum to uniquely determine the coefficients of the fitted polynomial. Since

the above kernel analysis shows that a symmetric weighting has no odd-order error terms, it is reasonable to assume that the span of points used in the least-squares fit should also be symmetric around the node of interest. The order of the polynomial fit should be one less than the desired order of accuracy. For example, to achieve fourth-order accuracy, one should use a third-order polynomial.

Least-squares fits have the advantage that they reduce statistical error. However, one should expect lower spatial accuracy, since the fitting process uses data that are further from the node than the relatively compact kernels described above. In the tests below, least-squares fitting will be compared to the high-order kernels.

EXTENSION TO THREE DIMENSIONS

The above analysis must be extended to multiple dimensions. If one chooses to use a multidimensional kernel that is a product of the above one-dimensional kernels, the extension is simple. For example, the kernel would become:

$$W(x, y, z) = W_x(x)W_y(y)W_z(z) \tag{28}$$

In this form, the one-dimensional constraints remain unchanged. For example, the requirement for second-order accuracy, given in Equation (9) becomes trivially different, when applied to Equation (28). Since each component of the three-dimensional kernel is a function of only one variable, the other two components can be pulled outside the integral, leaving the x constraint unchanged. New cross-derivative error terms which appear in the multi-dimensional Taylor series expansion do not present any difficulties, since lower order constraints appear in these terms. The lower-order constraints will have already been satisfied with the higher-order kernels, nullifying the new terms. For example, if one is trying to achieve fourth-order accuracy, the third-order cross-derivative terms will give rise to first- and second-order constraints that have already been satisfied.

The full three-dimensional derivation is both laborious and unnecessary. For clarity, a two-dimensional derivation will be given here. By using separable multi-dimensional kernels, lower-order constraints are easily noted, which are known to integrate to zero. The three-dimensional case would be perfectly analogous.

As in Equation (2), the integral of the weighted sources over the two-dimensional cell would be:

$$\langle f(x, y)\rangle = \int\int W_x(x_\mathrm{p} - x) \cdot W_y(y_\mathrm{p} - y) f(x_\mathrm{p}, y_\mathrm{p})\, \mathrm{d}x_\mathrm{p}\, \mathrm{d}y_\mathrm{p} \tag{29}$$

The expansion of the source term in two dimensions in the neighbourhood of (x, y) is

$$\begin{aligned} f(x_\mathrm{p}, y_\mathrm{p}) \cong f(x, y) &+ \frac{\partial f}{\partial x}(x_\mathrm{p} - x) + \frac{\partial f}{\partial y}(y_\mathrm{p} - y) \\ &+ \frac{1}{2}\frac{\partial^2 f}{\partial x^2}(x_\mathrm{p} - x)^2 + \frac{1}{2}\frac{\partial^2 f}{\partial y^2}(y_\mathrm{p} - y)^2 + \frac{\partial^2 f}{\partial x \partial y}(x_\mathrm{p} - x)(y_\mathrm{p} - y) \end{aligned} \tag{30}$$

Inserting Equation (30) into Equation (29) and noting that the partial derivatives are all taken at a fixed (x, y) location, gives the following:

$$\begin{aligned}\langle f(x,y)\rangle = f(x,y)\int W_y(y_\mathrm{p}-y)\left(\int W_x(x_\mathrm{p}-x)\,\mathrm{d}x_\mathrm{p}\right)\mathrm{d}y_\mathrm{p}\\ +\frac{\partial f}{\partial x}\int W_y(y_\mathrm{p}-y)\left(\int W_x(x_\mathrm{p}-x)(x_\mathrm{p}-x)\,\mathrm{d}x_\mathrm{p}\right)\mathrm{d}y_\mathrm{p}\\ +\frac{\partial f}{\partial y}\int W_x(x_\mathrm{p}-x)\left(\int W_y(y_\mathrm{p}-y)(y_\mathrm{p}-y)\,\mathrm{d}y_\mathrm{p}\right)\mathrm{d}x_\mathrm{p}\\ +\frac{1}{2}\frac{\partial^2 f}{\partial x^2}\int W_y(y_\mathrm{p}-y)\left(\int W_x(x_\mathrm{p}-x)(x_\mathrm{p}-x)^2\,\mathrm{d}x_\mathrm{p}\right)\mathrm{d}y_\mathrm{p}\\ +\frac{1}{2}\frac{\partial^2 f}{\partial y^2}\int W_x(x_\mathrm{p}-x)\left(\int W_y(y_\mathrm{p}-y)(y_\mathrm{p}-y)^2\,\mathrm{d}y_\mathrm{p}\right)\mathrm{d}x_\mathrm{p}\\ +\frac{\partial^2 f}{\partial x\partial y}\int W_x(x_\mathrm{p}-x)(x_\mathrm{p}-x)\left(\int W_y(y_\mathrm{p}-y)(y_\mathrm{p}-y)\,\mathrm{d}y_\mathrm{p}\right)\mathrm{d}x_\mathrm{p}\\ +\ldots\end{aligned} \tag{31}$$

This equation can be reduced by noting that the integrals in the first term on the right side are unity, and for the quadratic kernel, all the inner integrals in subsequent terms are zero. The first non-zero integrand, after the leading term, is when a fourth power occurs in the integral. Hence, Equation (31) reduces to

$$\langle f(x,y)\rangle = f(x,y) + O[(x-x_\mathrm{p})^4] + O[(y-y_\mathrm{p})^4] \tag{32}$$

This quadratic kernel can be plotted for the two-dimensional case, and is shown in Figure 4. This kernel is simply the product of the two one-dimensional kernels. Quite simply, using a separable kernel works like independent one-dimensional integrals when working in two or three dimensions.

NUMERICAL TESTS

In order to test the theoretical and practical accuracy of the high-dimensional kernels, a simple scenario was contrived. This scenario is general enough to provide useful information about how these techniques perform, but is simple enough to admit an analytical solution which can be used for checking convergence rates. Consider the source term that appears in the continuum phase of a two-phase Lagrangian/Eulerian fluid dynamics calculation representing the momentum per unit volume exchanged between phases. There is variation in the source term due to spatially varying particle number density and relative velocity. This term is represented by f on the right side of the Navier–Stokes equations as a source term, as shown

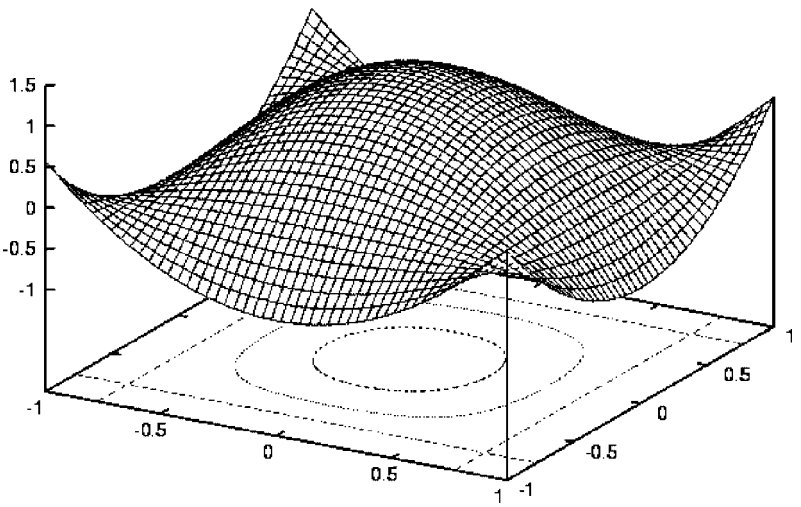

Figure 4. A quadratic, fourth-order kernel constructed from the product of two one-dimensional kernels. The two horizontal axes are $(x - x_p)$ and $(y - y_p)$ and the surface shows the kernel weighting at each point.

in the following equation for the velocity component u in the x direction:

$$\frac{Du}{Dt} = -\frac{1}{\rho}\frac{\partial p}{\partial x} + \frac{\partial \tau_{xx}}{\partial x} + \frac{\partial \tau_{xy}}{\partial y} + \frac{\partial \tau_{xz}}{\partial z} + f \tag{33}$$

The value of f is normally calculated in Lagrangian/Eulerian simulations by tallying the drag on nearby particles. In the current one-dimensional test case, f will be estimated for one instant in time for a two-phase flow. The Navier–Stokes equations will not be solved, nor will the particles be moved; this test case is only concerned with the inter-phase source term as a function of space.

The spatially varying particle number density used in this scenario has an interesting consequence: it means that the centre of particle mass on any interval does not coincide with the centre of the interval. The information about number density, which comes from the particles, is always spatially biased. In this particular test case, the number density varies exponentially, according to the following equation:

$$n(x) = n_0 \exp(-\alpha x) \tag{34}$$

Particles, are distributed according to the above distribution on the line from zero to one. The statistical weight per numerical particle is set so that the total number of physical particles is fixed at 10^9. The value chosen for α is $1/10$ and the value of n_0, the peak number density is set so that the integral of number density produces the correct number of particles on the interval. The relative velocity is assumed to vary exponentially, using the same form as the number density.

$$u(x) = u_0 \exp(-\alpha x) \tag{35}$$

Next, it is assumed that the drag force is proportional to the area of the particle and the relative velocity squared. This permits calculation of the exact the source term at any point x as given by the following equation and plotted in Figure 6. Note that the chosen value for α provides a rapid change in the source with position along the line.

$$f_{\text{exact}}(x) = Cr^2 u_0^2 n_0 \exp(-3\alpha x) \tag{36}$$

The value of r, representing the radius, was chosen to be a constant. The value of the constant, C, was set to unity. To quantify error in the numerical estimates, an L^2 norm was used, as given in the following equation. The summation is over all interior points; boundary points were excluded.

$$L^2 = \sqrt{\Delta x \sum_i (f_n(x) - f_{\text{exact}}(x))^2} \tag{37}$$

The tests evaluated the performance of the nearest-node, the quadratic (fourth-order), and quartic (sixth-order) kernels. A least-squares technique was also included for comparison. This least-squares technique used here is a simplified form of the approach used by Subramaniam and Haworth [6], since the current test is only in one dimension. The basic steps for the least-squares fit are:

1. Calculate all node values using a kernel estimate.
2. Use the node values from the previous step as 'knots' for fitting least-squares polynomials. The knots are the locations of the functional values used in the least-squares fits. A different set of coefficients is calculated for each node based on the nearest 11 kernel estimates from step 1. These 11 points were centred on the node of interest, except where boundaries interfered, in which case the 11 points were shifted to one side.
3. Evaluate the polynomial at the node location.

For the current work, a fourth-order accurate least-squares fit was attempted. In step 1, two different kernels were tested: a nearest-node kernel, as used in Reference [6], and a fourth-order quadratic kernel. A third-order polynomial fit was used in step 2, since this would be expected to give fourth-order accuracy. Figure 5 shows that the least-squares fit converges regardless of whether second- or fourth-order kernels are used to estimate the values at the knots. Both methods give similar errors, though the second-order knots are slightly more accurate. All subsequent use of least-squares fits in the current work have been performed using the nearest-node estimates at the knots.

The point of this test case is to predict a source term that varies in one dimension. Figure 6 shows the predicted source term compared to the exact value at low resolution. In this figure, an average of 100 particles per cell is used. The source appears as a peak near $x = 0$ and decays rapidly with increasing values of x. Figure 6 shows that when the source term is modestly resolved, the kernels all perform well and least-squares method overshoots as the function decreases. The performance of the kernels, like the high-order methods traditionally used for single-phase CFD calculations, permits the user to capture strong gradient with minimal resolution.

For convergence tests of the kernels, a large number of particles were used. In order to suppress statistical error, the number of particles was set to give a average of 10^9 particles per

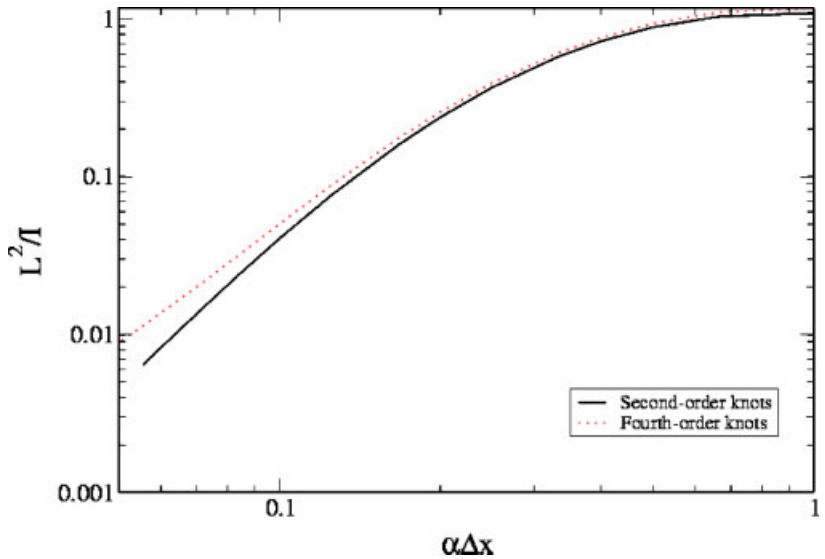

Figure 5. Convergence rates of the least-squares treatment of source terms using two different methods of calculating the values at the knots. The horizontal axis is the non-dimensional resolution and the vertical axis is non-dimension error.

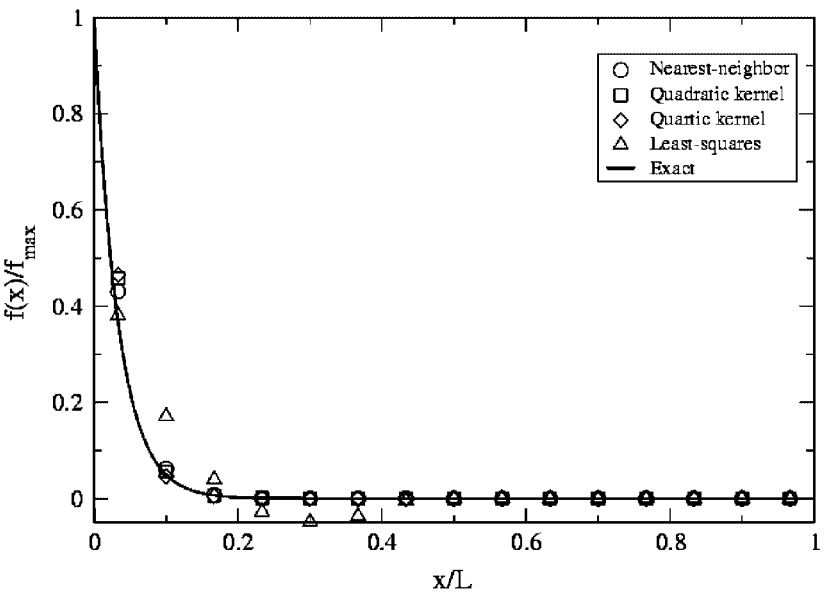

Figure 6. Estimates of the spatially varying source using several different kernels and the least-squares technique. Fifteen points are used and an average of 15 particles per point. The calculated source is normalized by the maximum value, which occurs at $x/L = 0$.

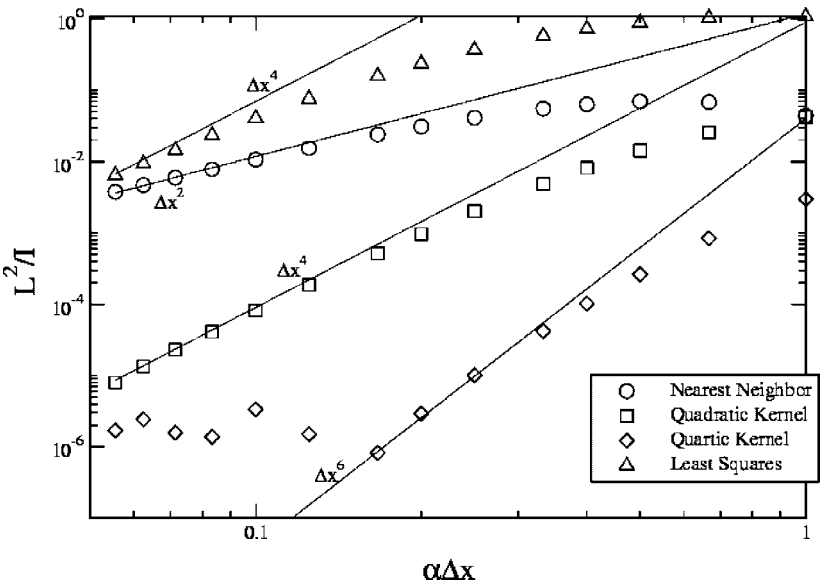

Figure 7. Convergence rates of various numerical schemes using a large number of particles. Numerical results (symbols) are compared to theoretical convergence rates (lines).

node. Hence, the number of samples exceeded the number of physical particles. As mentioned in the discussion of statistical error, the number of computational particles should increase super-linearly as the domain is refined in order to judge convergence. However, by using such a large number of particles, the statistical error was generally kept low enough to observe spatial convergence. The results of this test are shown in Figure 7. In this figure, the non-dimensionalized error is plotted versus the mesh resolution. The L^2 norm is divided by the integral of the exact answer, I, in order to produce a measure of relative error. The point of this figure is to show the asymptotic convergence and to demonstrate that the above kernel derivations are indeed correct.

One can see from the results that when the numbers of computational particles are large enough, good convergence can be observed from all the methods. The observed slopes tend to asymptotically approach the numerical data, with only a few exceptions. The highest order kernel, the quartic polynomial, indeed approaches sixth-order accuracy. However, it eventually reaches a level where the error is dominated by statistical noise and does not further converge. The figure also shows that the least-squares fit approaches the theoretical fourth-order slope, but suffers from a higher overall error level than the second-order nearest-neighbour kernel. As mentioned previously, this large error is expected because the least-squares method uses data from so far away from the node of interest. This figure confirms the theoretical expectations deduced from the analysis of various kernels.

The next step was to reduce the sample size and observe the behaviour of the kernels with only 100 particles per cell. The methods were otherwise the same as used for Figure 7. The results of the tests are shown in Figure 8. These results show that despite the problems with statistical noise, the high-order kernels are beneficial. Most CFD

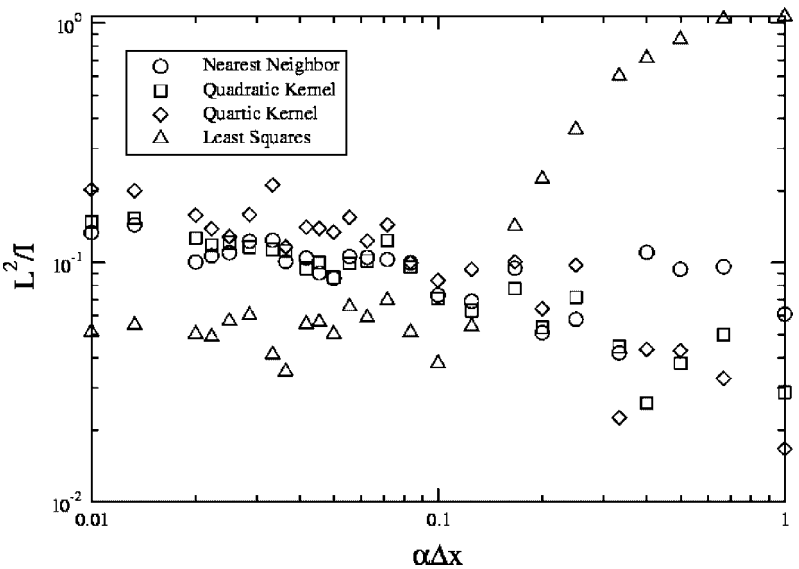

Figure 8. Convergence rates of various numerical schemes using an average of 100 particles per kernel width.

calculations operate with a mesh resolution that is just sufficient to capture the flowfield gradients, since the cost of obtaining a solution increases rapidly as the cell size decreases. Hence, most calculations would operate on the right side of Figure 8, where the gradients are barely resolved. Values of $\alpha\Delta x$ of unity would correspond to a mesh resolution comparable to the length scale of variation in the calculation. At $\alpha\Delta x$ of 0.01, the calculation would be exceptionally well-resolved. The results show that instead of such fine mesh resolution, one would do better to spend their resources on more computational particles (presuming that the numerical particles are a statistical representation of the physical particles).

Figure 9 shows similar results, but with the average number of particles per cell set to 1000. The shift from spatially dominated error to statistical error is now more evident, as the data from the nearest-neighbour and least-squares methods both pass through minima. Since this test case represents a single moment in time, Figure 9 also represents an ensemble of ten realizations of simulations with 100 particles per cell.

Clearly, statistical noise is dominating the mean expected error in the source term at extremely fine mesh resolutions. In the limit of fine mesh resolution, the error follows the left side of the curve shown in Figure 3, of the theoretical error. For the kernel schemes, error is slowly growing as the mesh is refined. Only the least-squares fit, with a relatively large spatial error and small statistical error, shows the complete transition from spatial to statistical error. One would expect that most two-phase flow simulations would use the minimum acceptable mesh resolution to capture spatial gradients, and thus the non-dimensional resolutions ($\alpha\Delta x$) would be approximately unity. At these more realistic resolutions, the high-order schemes do show benefits, even with modest numbers of particles.

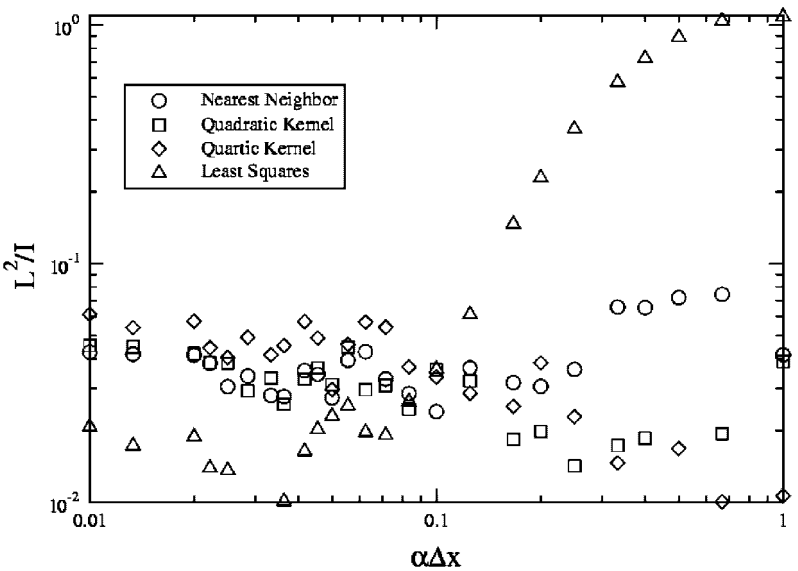

Figure 9. Convergence rates of various numerical schemes using an average of 1000 particles per kernel width.

TWO-DIMENSIONAL TESTS

A two-dimensional test was developed to verify the extension of high-dimensional kernels to multiple dimensions. Like the previous test, the goal is to estimate the inter-phase coupling source term for a single moment in time. Again, a test case that permits an analytical solution was used so that the numerical error could be calculated. In this test case, particles were distributed over a unit square with a spatially varying number density. The number density was proportional to $\sin(\pi x)$ and $\sin(\pi y)$. Each particle contributed a uniform amount to the source term calculation, but the spatial variation in number density created a source as a function of x and y, as given by the following equation and shown in Figure 10:

$$f_{\text{exact}}(x, y) = \frac{n_0 \pi^2}{4} \sin(\pi x) \sin(\pi y) \tag{38}$$

In Equation (38), the symbol n_0 represents the number density of particles per unit area. A value of 10^9 particles per unit area was used in the current calculations and in Figure 10. First, in order to confirm the theoretical convergence of the kernel, increasingly large numbers of particles were used and the error between the exact and the numerical solution was plotted in Figure 11. The error is defined as in the following equation:

$$\text{Error} \equiv \frac{\sqrt{\sum (f_n(x, y) - f_{\text{exact}}(x, y))^2 \Delta x \Delta y}}{A n_0 \pi^2 / 4} \tag{39}$$

The summation is done over all the interior points in the domain. The denominator normalizes the error, and the value of A is unity (similarly, the length scale L mentioned in the x-axis

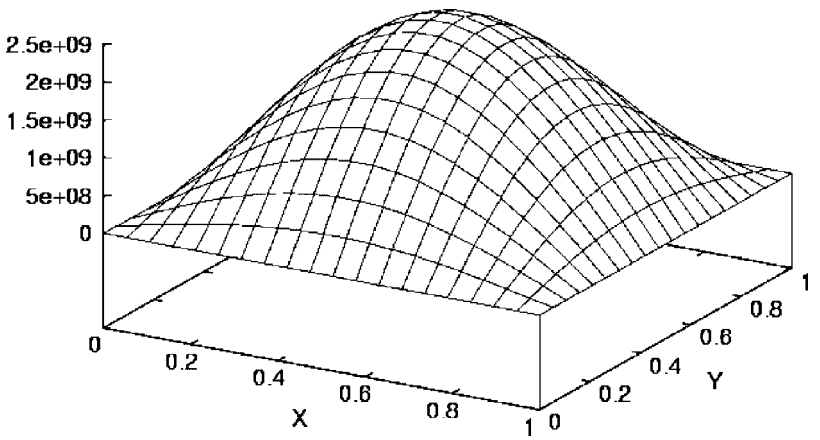

Figure 10. The exact solution of the spatially varying source term in the two-dimensional test case (see Equation (8)).

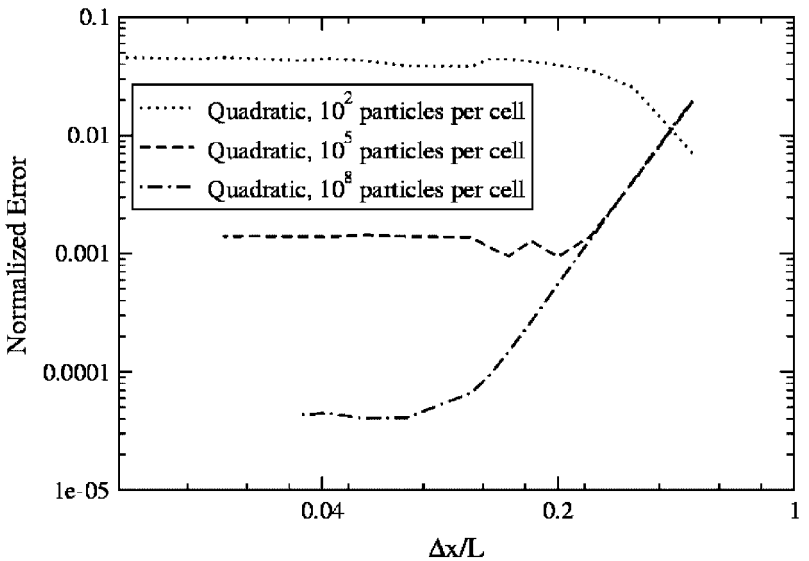

Figure 11. The normalized error, as defined in Equation (9), plotted versus mesh resolution for differing numbers of particles per kernel width.

label in Figure 11 is also unity). For very large numbers of particles (at least 10^5) and modest resolution ($\Delta x/L$ approximately 0.25), the quadratic kernel indeed provides fourth-order convergence in two-dimensions.

Figure 12 compares the quadratic kernel to other methods of calculating the source term in two dimensions. The nearest-node method accumulates all of a particle's contribution to the source at whatever continuum phase node is closest. The cloud-in-cell method recommended

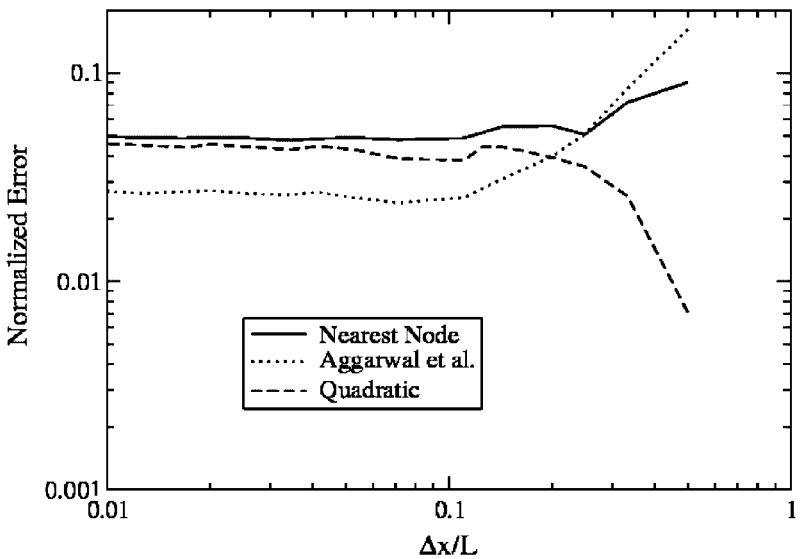

Figure 12. The normalized error, as defined in Equation (9), for various kernels with 100 particles per kernel width.

by Aggarwal *et al.* [2] is currently the standard method for most direct numerical simulation codes, and provides a weighting function developed from geometric arguments. This method was also tested and plotted in Figure 12. Finally, the quadratic kernel is shown for comparison. At resolutions of a quarter to half of the governing length scale, the quadratic kernel provides the most accurate estimate of the source term. For over-resolved cases, where the mesh resolution is very small compared to the length scales of the gradients, the statistical errors dominate and the method of Aggarwal *et al.* is the best.

As mentioned previously, the property of strong conservation is not retained for the higher order kernel. This means that there is little reason to limit the width of the kernel to the span of a cell. Under some conditions, the kernel width can be optimized so that it is fine enough to respect physical length scales and large enough to include numerous particles. With kernel width as a free parameter, one is free to look at Figure 8 and choose the method and size to minimize total error. This could lead to either a high-order kernel with low resolution or a least-squares method with high resolution (for a more rigorous comparison, one must let the sample size be a function of the kernel width). From the two-dimensional results shown in Figure 12, the lowest overall error comes from choosing a large kernel width and the quadratic kernel. If the simulation constraints permit the latitude of using these wider kernels, the accuracy of the inter-phase coupling can be improved. It remains to be seen, how methods for optimal kernel width could be constructed and how effective these methods will be.

CONCLUSIONS

Theoretically, it is possible to achieve high-order spatial accuracy in the calculation of gas-phase source terms in Lagrangian–Eulerian CFD codes. This high-order accuracy would bring

gas-to-particle coupling up to the formal accuracy of the other parts of the Lagrangian–Eulerian calculation. An analytical derivation has been provided for fourth- and sixth-order accuracy in one dimension, and it has been shown that two- or three-dimensional methods can easily be constructed by taking the product of one-dimensional kernels.

Numerical tests show that indeed, the predictions of theoretical accuracy are correct; sixth-order accuracy (and presumably even higher) can be achieved with little computational effort. Two-dimensional tests show that the extension to higher dimensions is simple.

Practically, there are still several obstacles to observing a large benefit in CFD simulations. Past work in non-parametric function analysis has shown that for modest sample sizes, statistical noise makes it difficult to observe convergence. The number of computational particles per cell must increase rapidly in order to keep statistical error from dominating spatial error, as the computational mesh is resolved. Even least-squares techniques, which tend to be resilient in the presence of statistical noise, provide only a modest benefit when the source term is well resolved.

In actual CFD calculations, high-order schemes could reduce spatial error only under certain conditions. In the current two-dimensional tests, the high-order kernel was superior for 100 particles per cell and at resolutions about one quarter of the variational length scale. For example, high-order methods could produce a significant benefit when sources change rapidly over a few mesh cells. Since one generally expects a modeller to use a resolution just sufficient for the problem at hand, many practical situations could show some benefit from using high-order methods. Using kernel widths larger than the gas-phase cells further allow the inter-phase coupling error to be minimized. This idea was shown to be beneficial with low-order methods for one-dimensional tests by Dreeben and Pope [11].

Further, the test results shown in this paper are for a single instant in time, not for an evolving transient calculation. In a full CFD simulation, cell source terms are calculated for a large number of cells over a large number of time steps. It is possible that the large amount of repetition in real CFD calculations may provide some moderation of the statistical errors. In the case of steady-state CFD calculations, averaging the source terms over multiple iterations would certainly provide statistical benefit. One also has the option of broadening the support of the kernels to include a larger sample size since the kernel width need not be equal to the cell size. Furthermore, the CPU cost of adding more particles to the calculation generally incurs a linear cost and is far easier than adding mesh resolution. Given the low cost of using high-order kernels, these options may prove attractive in the future if sufficient progress can be made on algorithms for optimizing the kernel width. Currently, for very fine meshes or low numbers of particles per cell, least-squares methods are likely to be more stable and more accurate.

NOMENCLATURE

a, b, c	coefficients which appear in the assumed polynomial form of a high-order kernel
A, B	coefficients which appear in the assumed polynomial variation of a source term
E	mean expected error; the total error, including both spatial and statistical contribution
f	the exact source contribution to the gas phase per unit volume

f_n	a numerical estimate of f based on n particles
$\langle f \rangle$	a numerical estimate of f based on an infinite number of particles, the expected value of f
ℓ	the bounds of the kernel support
L	the width of the domain used in the test cases
L^2	a norm of a vector, used for quantifying error
$n(x)$	number density of physical particles
s^i	a source contribution of mass, momentum, or energy from the ith particle
$u(x)$	relative velocity between particles and gas
W	a kernel, or weighting function, representing the relationship between location and contribution of source terms
x	spatial location
x_p	the spatial location of a particle
y	the distance between a node and a particle, e.g. $(x_\mathrm{p} - x)$

Greek letters

α	a coefficient representing how rapidly the test function varies in space
σ^2	the variance of a function or random variable

ACKNOWLEDGEMENTS

Sincere thanks are due to Shankar Subramaniam and Rahul Garg of Iowa State for sharing their knowledge of the relevant literature.

REFERENCES

1. Are S, Hou S, Schmidt DP. Second-order spatial accuracy in Lagrangian–Eulerian spray calculations. *Numerical Heat Transfer Part B*: *Fundamentals* 2005, in press.
2. Aggarwal SK, Fix GJ, Lee DN, Sirignano WA. Numerical optimization studies of axisymmetric unsteady sprays. *Journal of Computational Physics* 1983; **50**:101–115.
3. Miller RS, Bellan J. Direct numerical simulation of a confined three-dimensional gas mixing layer with one evaporating hydrocarbon-droplet laden stream. *Journal of Fluid Mechanics* 1999; **384**:293–338.
4. Xu J, Pope SB. Assessment of numerical accuracy of PDF/Monte Carlo methods for turbulent reacting flows. *Journal of Computational Physics* 1999; **152**:192–230.
5. Subramaniam S. Statistical representation of a spray as a point process. *Physics of Fluids* 2000; **12**(10).
6. Subramaniam S, Haworth DC. A probability density function method for turbulent mixing and combustion on three-dimensional unstructured deforming meshes. *Journal of Engine Research* 2000; **1**(2):171–190.
7. Bartlett MS, Statistical estimation of density functions. *Sankhyā*, *Series A* 1963; **25**:245–254.
8. Parzen, Emanuel. On estimation of a probability density function and mode. *The Annals of Mathematical Statistics* 1962; **33**(3):1065–1076.
9. Rosenblatt M. Curve estimates. *Annals of Mathematical Statistics* 1971; **42**(6):1815–1842.
10. Nordin N. Complex chemistry modeling of diesel spray combustion. *Ph.D. Thesis*, Chalmers University, 2001.
11. Dreeben TD, Pope SB. Nonparametric estimation of mean fields with application to particle methods for turbulent flows. *Technical Report FDA 92-13*, Cornell University, 1992.

