47th AIAA Aerospace Sciences Meeting Including The New Horizons Forum and Aerospace Exposition AIAA 2009-403
5 - 8 January 2009, Orlando, Florida

A Reconstruction Approach to High-Order Schemes Including Discontinuous Galerkin for Diffusion

H. T. Huynh
NASA Glenn Research Center, MS 5-11, Cleveland, OH 44135, USA; E-mail: huynh@grc.nasa.gov

Abstract. We introduce a new approach to high-order accuracy for the numerical solution of diffusion problems by solving the equations in differential form using a reconstruction technique. The approach has the advantages of simplicity and economy. It results in several new high-order methods including a simplified version of discontinuous Galerkin (DG). It also leads to new definitions of common value and common gradient—quantities at each interface shared by the two adjacent cells. In addition, the new approach clarifies the relations among the various choices of new and existing common quantities. Fourier stability and accuracy analyses are carried out for the resulting schemes. Extensions to the case of quadrilateral meshes are obtained via tensor products. For the two-point boundary value problem (steady state), it is shown that these schemes, which include most popular DG methods, yield exact common interface quantities as well as exact cell average solutions for nearly all cases.

Keywords. Discontinuous Galerkin methods, high-order methods, reconstruction, diffusion equation.

1. Introduction

In the field of Computational Fluid Dynamics, low-order methods are less accurate, but generally are robust and reliable; therefore, they are routinely employed in practical calculations. High-order (third and above) methods can provide accurate solutions; however, they are more complicated and less robust. The need to improve and simplify high-order methods as well as to develop new ones with more favorable properties has attracted the interest of many researchers. The current work is a step in this direction.

For simplicity, we discuss the case of one spatial dimension, which contains all of the key new ideas. For this case, the solution u is approximated in each cell (interval) by the data at K points called solution points ($K \geq 1$). These K pieces of data determine a polynomial of degree $K-1$, herein called a solution polynomial. Such polynomials collectively form a function, which is generally discontinuous across cell interfaces. One advantage of allowing such discontinuities is that the resulting method is local. Another advantage is that the method can better resolve discontinuities such as shocks in the exact solution. Calculating the first and second derivatives of u by those of the solution polynomials leads to incorrect results, since these quantities involve no interaction of the data between cells. For interaction to take place, at each interface, we need to define a value u and a derivative (gradient) value u_x common to the two adjacent cells; the first and second derivative estimates making use of these common quantities (value or derivative) would involve interaction. These common quantities can be employed in two different approaches.

The first approach, introduced by Reed and Hill (1973) for neutron transport equations, is the discontinuous Galerkin or DG approach. It was developed for fluid dynamics equations by Cockburn and Shu (1989), Bassi and Rebay (1997a, b), and numerous other researchers; see, e.g., the review paper by Cockburn, Karniadakis, and Shu (2000). Aside from the discontinuous representation of the solution, another key ingredient of the DG methods is the integral formulation using test functions. The DG approach was applied to solve diffusion-type problems by several authors, e.g., Arnold (1982), Bassi and Rebay (1997a), Cockburn and Shu (1998), Oden et al. (1998), Arnold et al. (2002). The common derivative at each interface, which leads to a scheme with a rather wide stencil called BR1, was presented in (Bassi and Rebay 1997a). It was modified so that the resulting scheme has a compact stencil including only immediate neighbors (Bassi and Rebay 2000). This modification, often called BR2, was analyzed in (Brezzi et al. 2000). In addition to the advantage of having a compact stencil, which simplifies the coding of boundary conditions, another advantage of the BR2 scheme is that for the steady state case, the coefficient matrix for the solution is invertible. The technique of compact stencil was applied by Peraire and Persson (2008) to the case of one-sided common quantities; the resulting method, which is a modification of the local DG or LDG scheme of Cockburn and Shu (1998), is named compact DG or CDG. Finally, Van Leer and Nomura (2006) obtained common quantities by a recovery principle: a polynomial of

degree $2K-1$ approximating the solution polynomials on a domain formed by the two cells adjacent to the interface is constructed via least squares.

The second approach, introduced in (Kopriva and Kolias 1996) and (Kopriva 1998) for a quadrilateral mesh, solves the differential form of the equation as opposed to the integral form. Here, the solutions are evaluated at the K Chebyshev-Gauss points, and the fluxes, at the $K+1$ Chebyshev-Lobatto points. For diffusion problems, the common quantities are defined as in the BR1 scheme. This approach on a triangular mesh results in a scheme called spectral difference (Liu et al. 2006). These solution-point and flux-point (or staggered-mesh) methods have the advantage of conceptual simplicity and the disadvantages of using two sets of points (or grids) as well as being mildly unstable (Huynh 2007, Van den Abeele et al. 2008).

In this paper, we present a new approach to high-order accuracy for diffusion problems using the differential form. The approach has the advantage of simplicity and economy. It employs one set of points, namely, the solution points, and fluxes are evaluated only at the cell boundaries. It was introduced to solve advection-type problems in (Huynh 2007), where the objective was the evaluation of the first derivative. Here, the approach is applied to solve diffusion problems, and the objective is the evaluation of the second derivative. To this end, first, the common value at each interface can be defined as the average of the two values just to its left and right as in the BR1 and BR2 schemes, or a one-sided value, say, the one to its left as in the LDG and CDG scheme (other criteria will be introduced later). Next, we reconstruct the solution by a function which is continuous across the interfaces and, in each cell, is a polynomial of one degree higher than that of the solution polynomial so that its derivative is of the same degree. To assure continuity across the interfaces, this reconstruction is required to take on the common interface value. In each cell, the reconstruction is obtained by adding two correction functions to the solution polynomial, one to match the common value at the left interface, and the other to match the common value at the right. Due to symmetry, we only need to focus on the correction for the left interface. By rescaling, the problem reduces to defining a correction function g on the interval $[-1,1]$ such that $g(-1)=1$, $g(1)=0$, and g is a polynomial of degree K approximating the zero function in some sense. The condition $g(-1)=1$ deals with the jump at the left interface whereas the condition $g(1)=0$ leaves the right interface value unchanged. The crucial condition of approximating the zero function can be met by the least-squares principle: g is required to be orthogonal to all polynomials of degree $K-2$ or less. Such a correction function turns out to be the Radau polynomial, and the resulting method is identical to DG. The current version of the scheme, however, is simpler than the standard version using integral formulation: at each solution point, it involves only a straightforward derivative formula and a simple correction term regardless of the choice of solution points. In addition to the Radau polynomial, two other choices for g that result in new methods will be discussed.

Next, the reconstruction for u is continuous across the interfaces, but its derivative can be discontinuous. For the evaluation of second derivative, we can apply the above correction procedure to the derivative of the reconstruction. Before this application, we need a common derivative at each interface, for which the simplest choice is the average derivative employed in the BR2 method. This common derivative was originally cast via the integral formulation; here, it is cast in a more intuitive manner via the differential formulation. The concept of reconstruction also clarifies and simplifies its motivation and calculation. Moreover, the current approach leads to several new criteria for the definition of common quantities. In particular, a scheme that shares the high accuracy property of Van Leer's recovery method without the disadvantage of interpolating across cells is derived. The comparison and trade-offs among the various choices of new and existing common values as well as choices of correction functions are carried out by Fourier analysis. Extension to the case of quadrilateral meshes is obtained via tensor products, and Fourier analyses in two dimensions are discussed. For the case of the two-point boundary value problem (steady state), it is shown that these reconstruction schemes, which include BR2, LDG, CDG, and recovery (RDG), yield exact common interface quantities as well as exact cell average solutions for almost all K. Finally, as a side benefit of the current approach, the (continuous) reconstruction polynomials are shown, via numerical examples, to be more accurate than the (discontinuous) solution polynomials.

In the hope of attracting the interest of researchers who are not familiar with these types of high-order schemes, this paper is written in a self-contained manner. It is organized as follows. The diffusion equation, solution polynomials, and various notations are introduced in Section 2. The evaluation of the first derivative via reconstruction is presented in §3, and that for the second derivative, in §4. Fourier analysis and the proof that the eigenvalues are independent of the solution points chosen can be found in §5. Stability and accuracy of various schemes are presented in §6. Extensions to the case of two spatial

dimensions are discussed in §7. Section 8 deals with numerical examples. Conclusions and discussions can be found in §9. Finally, the appendix concerns the boundary value problem (steady state).

2. Diffusion Equation

On the domain $[a,b]$, consider the diffusion equation

$$u_t = u_{xx} \tag{2.1}$$

with initial condition

$$u(x,0) = u_0(x)\,. \tag{2.2}$$

Let the domain $[a,b]$ be divided into possibly nonuniform cells or elements E_j of cell widths h_j and cell centers x_j, $j = 1,...,J$. For the moment, the solution is assumed to be periodic; therefore, boundary conditions are trivial: the data in the ghost cells $j = 0$ and $j = J+1$ are respectively identical to those in cells J and 1.

On each cell E_j, let the solution be approximated by the *K* pieces of data $u_{j,k}$, $k = 1, ..., K$ at locations $x_{j,k}$, which are called solution points. The *K* solution points are typically the Gauss or Lobatto points. For (at least) linear problems, equidistant points can also be employed. In fact, it will be shown that the Fourier stability and accuracy results are independent of the type of points chosen. For simplicity, we assume that the same type and same number of points are employed for all cells.

At each solution point, the solution $u_{j,k}(t)$ depends on t; $u_{j,k}(t^n)$, however, is abbreviated to $u_{j,k}$. At time level *n*, suppose the data $u_{j,k}$ are known for all *j* and *k*. We wish to calculate $du_{j,k}(t)/dt$ at time $t = t^n$, which is abbreviated to $du_{j,k}/dt$. That is, we wish to evaluate u_{xx} at the solution points $x_{j,k}$ in terms of the data. Then, we march in time by, say, a Runge-Kutta method.

Instead of dealing with the cell E_j or the global description, it is often more convenient to deal with the interval $I = [-1,1]$ or the local description; see, e.g., (Hughes 2000). With ξ varying on I and x on E_j, the linear function mapping I onto E_j and its inverse are

$$x(\xi) = x_j + \xi\, h_j/2 \quad \text{and} \quad \xi(x) = 2(x - x_j)/h_j\,. \tag{2.3a, b}$$

Let the solution points on I be denoted by ξ_k, $k = 1,...,K$. The solution points on E_j are, by (2.3a),

$$x_{j,k} = x(\xi_k) = x_j + \xi_k\, h_j/2\,. \tag{2.4}$$

A function $r_j(x)$ on E_j results in a function on I denoted by, for simplicity of notation, $r_j(\xi)$:

$$r_j(\xi) = r_j(x(\xi)) = r_j(x)\,.$$

The derivatives in the global and local descriptions are related by the chain rule,

$$\frac{dr_j(x)}{dx} = \frac{2}{h_j}\frac{dr_j(\xi)}{d\xi}\,. \tag{2.5}$$

On E_j, for each *k*, let $\phi_{j,k}$ be the Lagrange polynomial of degree $K-1$ that takes on the value 1 at $x_{j,k}$ and 0 at all other solution points. In the global and local descriptions, respectively,

$$\phi_{j,k}(x) = \prod_{l=1,\, l\neq k}^{K} \frac{x - x_{j,\,l}}{x_{j,\,k} - x_{j,\,l}} \quad \text{and} \quad \phi_k(\xi) = \prod_{l=1,\, l\neq k}^{K} \frac{\xi - \xi_l}{\xi_k - \xi_l}\,. \tag{2.6a, b}$$

These $\phi_{j,k}$ are called basis (or cardinal or shape) functions. The solution polynomial, i.e., the polynomial of degree $K-1$ interpolating $u_{j,k}$, $k = 1,...,K$, can be written as,

$$u_j(x) = \sum_{k=1}^{K} u_{j,k}\,\phi_{j,k}(x) \quad \text{and} \quad u_j(\xi) = \sum_{k=1}^{K} u_{j,k}\,\phi_k(\xi)\,. \tag{2.7a, b}$$

At each interface $x_{j+1/2}$, let $u^-_{j+1/2}$ and $u^+_{j+1/2}$ denote the values just to its left and right, respectively,

$$u^-_{j+1/2} = u_j(x_{j+1/2}) \quad \text{and} \quad u^+_{j+1/2} = u_{j+1}(x_{j+1/2})\,. \tag{2.8a,b}$$

In general, $u^-_{j+1/2} \neq u^+_{j+1/2}$. Using the local coordinate,

$$u^-_{j+1/2} = u_j(1) \qquad \text{and} \qquad u^+_{j+1/2} = u_{j+1}(-1)\,. \tag{2.9a,b}$$

3. Evaluation of the First Derivative

The first task is to estimate u_x at the solution points $x_{j,k}$. To this end, the polynomials u_j on E_j collectively (as j varies) form a function denoted by $\{u_j\}$, which is generally discontinuous across the interfaces. If we ignore these discontinuities and estimate u_x by $(u_j)_x$, we obtain erroneous results. Such a derivative estimate involves no interaction of the data between adjacent cells, and the solution never feels the effect of the boundary conditions. Therefore, to estimate u_x, we first reconstruct u by a piecewise polynomial function $\{u^C_j\}$, which is continuous at the cell interfaces and, on E_j, approximates u_j in some sense (the superscript 'C' stands for 'continuous' or 'corrected'). To maintain accuracy, the polynomial u^C_j is required to be of degree K so that $(u^C_j)_x$ is of the same degree as u_j. At the solution points, $(u^C_j)_x$ provides a derivative approximation that takes into account the data interaction.

Common interface values. In order for $\{u^C_j\}$ to be continuous at the interfaces, the reconstruction polynomials u^C_j and u^C_{j+1} must take on the same value at $x_{j+1/2}$. Such a value is often called 'numerical flux' (see the review paper by Cockburn, Karniadakis, and Shu (2000) or Bassi and Rebay (2000)); here, since the property of commonality relative to the two adjacent cells is essential, and we also need a similar quantity for the derivative, we use the term 'common'. Thus, at each interface, we need to define a common value and a common derivative. For an advection problem, between the left and right values $u^-_{j+1/2}$ and $u^+_{j+1/2}$, the common value is typically the upwind (flux) value. For diffusion problems, we use the following weighted average: with $0 \le \kappa \le 1$,

$$u^{\text{com}}_{j+1/2} = \kappa u^-_{j+1/2} + (1-\kappa) u^+_{j+1/2}\,. \tag{3.1}$$

If $\kappa = 1/2$, we obtain the average (centered formula) employed by Bassi and Rebay (1997a, 2000)

$$u^{\text{com}}_{j+1/2} = (1/2)(u^-_{j+1/2} + u^+_{j+1/2})\,. \tag{3.2}$$

If $\kappa = 1$ (or $\kappa = 0$), we have a one-sided formula employed by Cockburn and Shu (1998),

$$u^{\text{com}}_{j+1/2} = u^-_{j+1/2}\,. \tag{3.3}$$

Next, we require $u^C_j(x)$ to take on the values $u^{\text{com}}_{j-1/2}$ at $x_{j-1/2}$ and $u^{\text{com}}_{j+1/2}$ at $x_{j+1/2}$, to be of degree K, and to approximate $u_j(x)$ in some sense (see Fig. 3.1(a)). Instead of defining $u^C_j(x)$, we define $u^C_j(x) - u_j(x)$, which approximates the zero function. Switching to the local coordinate ξ on $I = [-1,1]$, at the left and right interfaces respectively, $u^C_j(\xi) - u_j(\xi)$ is required to take on the following known values:

$$u^C_j(-1) - u_j(-1) = u^{\text{com}}_{j-1/2} - u_j(-1) \qquad \text{and} \qquad u^C_j(1) - u_j(1) = u^{\text{com}}_{j+1/2} - u_j(1)\,. \tag{3.4a,b}$$

We now separate the prescription of the left interface value from that of the right. This separation is essential for the new approach to work. On the interval $I = [-1,1]$, let g^{LB} be the *correction function* for the left boundary defined by the following conditions:

$$g^{\text{LB}}(-1) = 1, \quad g^{\text{LB}}(1) = 0\,, \tag{3.5}$$

and g^{LB} is of degree K and approximates the zero function in some sense. The condition $g^{\text{LB}}(-1) = 1$ deals with the jump (3.4a), i.e., $u^{\text{com}}_{j-1/2} - u_j(-1)$, at the left interface, while the condition $g^{\text{LB}}(1) = 0$ leaves the right interface value $u_j(1)$ unchanged. Let g^{RB} be the correction function for the right boundary defined as the reflection of g^{LB}:

$$g^{\mathrm{RB}}(1)=1,\quad g^{\mathrm{RB}}(-1)=0\,, \tag{3.6}$$

and g^{RB} is of degree K and approximates the zero function. Employing g^{LB}, the polynomial

$$u_j^{\mathrm{LB}}(\xi) \;=\; u_j(\xi)+[u_{j-1/2}^{\mathrm{com}}-u_j(-1)]g^{\mathrm{LB}}(\xi) \tag{3.7}$$

provides a correction for $u_j(\xi)$ at the left interface by changing its value from $u_j(-1)$ to $u_{j-1/2}^{\mathrm{com}}$, while leaving the right interface value $u_j(1)$ unchanged. For later use, the polynomial

$$u_j^{\mathrm{RB}}(\xi) \;=\; u_j(\xi)+[u_{j+1/2}^{\mathrm{com}}-u_j(1)]g^{\mathrm{RB}}(\xi) \tag{3.8}$$

provides a correction for $u_j(\xi)$ at the right interface by changing its value from $u_j(1)$ to $u_{j+1/2}^{\mathrm{com}}$, while leaving the left interface value $u_j(-1)$ unchanged. See Fig. 3.1(b). Next, the polynomial

$$u_j^{C}(\xi) \;=\; u_j(\xi)+[u_{j-1/2}^{\mathrm{com}}-u_j(-1)]g^{\mathrm{LB}}(\xi)+\;[u_{j+1/2}^{\mathrm{com}}-u_j(1)]g^{\mathrm{RB}}(\xi) \tag{3.9}$$

provides the corrections to both interfaces: using (3.5) and (3.6), one can easily verify that $u_j^{C}(-1)=u_{j-1/2}^{\mathrm{com}}$ and $u_j^{C}(1)=u_{j+1/2}^{\mathrm{com}}$. See Fig. 3.1(a). Thus, the above $u_j^{C}(\xi)$ is of degree K, takes on the common values at the two interfaces, and approximates $u_j(\xi)$ in the same sense that g^{LB} and g^{RB} approximate the zero function. Finally, the derivative of $u_j^{C}(\xi)$ is

$$(u_j^{C})_{\xi}(\xi) \;=\; (u_j)_{\xi}(\xi)+[u_{j-1/2}^{\mathrm{com}}-u_j(-1)](g^{\mathrm{LB}})'(\xi)+\;[u_{j+1/2}^{\mathrm{com}}-u_j(1)](g^{\mathrm{RB}})'(\xi)\,. \tag{3.10a}$$

At the solution point ξ_k, the *corrected* derivative is given by

$$(u_x)_{j,k}=(2/h_j)(u_j^{C})_{\xi}(\xi_k)\,. \tag{3.10b}$$

Once the correction functions g^{LB} and g^{RB} are chosen on $[-1,1]$, the above calculations can be carried out in an economical and straightforward manner.

Note that the reconstruction polynomial $u_j^{C}(\xi)$ clarifies the ideas; in practice, however, we only need the values of its derivative at the solution points.

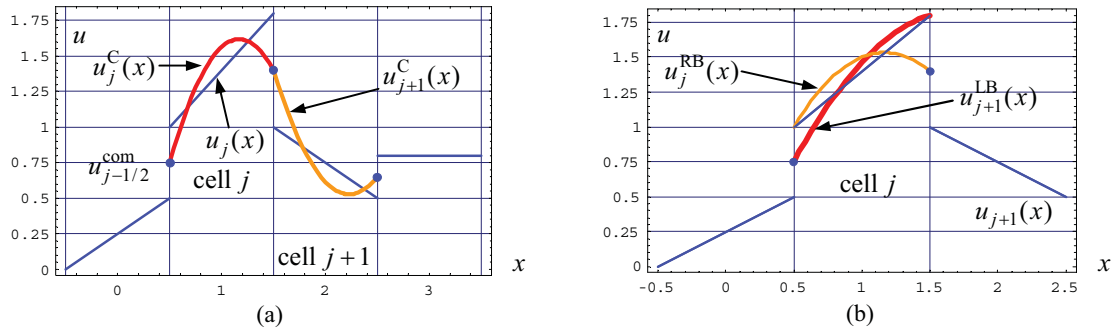


Figure 3.1. (a) Centered common values (dots) and reconstruction functions u_j^{C} and u_{j+1}^{C} for the case of piecewise linear solution polynomials; here, $\{u_j^{C}\}$ is continuous at all interfaces $x_{j+1/2}$. (b) Functions u_j^{LB} (corrected for the left boundary) and u_j^{RB} (corrected for the right boundary).

Correction functions. In the rest of this section, we define three correction functions. To this end, we need some reviews and preparations. Let the L^2 inner product of any two polynomials v and w on E_j be

$$(v,w)=\int_{x_{j-1/2}}^{x_{j+1/2}} v(\xi)w(\xi)d\xi\,.$$

For any integer $m \ge 0$, let $\boldsymbol{P}^m$ be the space of polynomials of degree m or less. Then $\boldsymbol{P}^m$ is a vector space of dimension $m+1$. When the domain, say, E_j, of the polynomials is important, we use the notation $\boldsymbol{P}^m(E_j)$. On $I=[-1,1]$, a polynomial v is orthogonal to $\boldsymbol{P}^m = \boldsymbol{P}^m(I)$ if, for each k, $0 \le k \le m$,

$$(v,\xi^k) = \int_{-1}^{1} v(\xi)\,\xi^k\,d\xi = 0\,.$$

Clearly, the criterion of being orthogonal to $\boldsymbol{P}^m$ provides $m+1$ conditions (or equations).

Again, on $I=[-1,1]$, for $k=0,1,2,\ldots$, let the Legendre polynomial P_k be the unique polynomial of degree k that is orthogonal to $\boldsymbol{P}^{k-1}$ and satisfies $P_k(1)=1$. Thus, if $k>m$, then P_k is orthogonal to $\boldsymbol{P}^m$. The Legendre polynomials are given by the following recurrence formula (see, e.g., Hildebrand 1987):

$$P_0 = 1,\ P_1 = \xi,$$

and, for $k \ge 2$,

$$P_k(\xi) = \frac{2k-1}{k}\,\xi\,P_{k-1}(\xi) - \frac{k-1}{k}\,P_{k-2}(\xi)\,. \tag{3.11}$$

The first few Legendre polynomials are plotted in Fig. 3.1(a). Properties of the Legendre polynomials that will be employed are listed below. First, P_k is an even function (involving only even powers of ξ) for even k, and an odd function for odd k. For all k, the values at the boundaries (or end points) are

$$P_k(-1) = (-1)^k\,, \qquad P_k(1) = 1\,. \tag{3.12a,b}$$

The derivative values at the end points are

$$P_k'(-1) = (-1)^{k-1}k(k+1)/2\,, \quad P_k'(1) = k(k+1)/2\,. \tag{3.13a,b}$$

In addition,

$$(P_k,P_k) = 2/(2k+1)\,, \text{ and, for } k \ne l\,,\ (P_k,P_l) = 0\,. \tag{3.14a,b}$$

Finally, the zeros of P_k are the k Gauss points on $[-1,1]$.

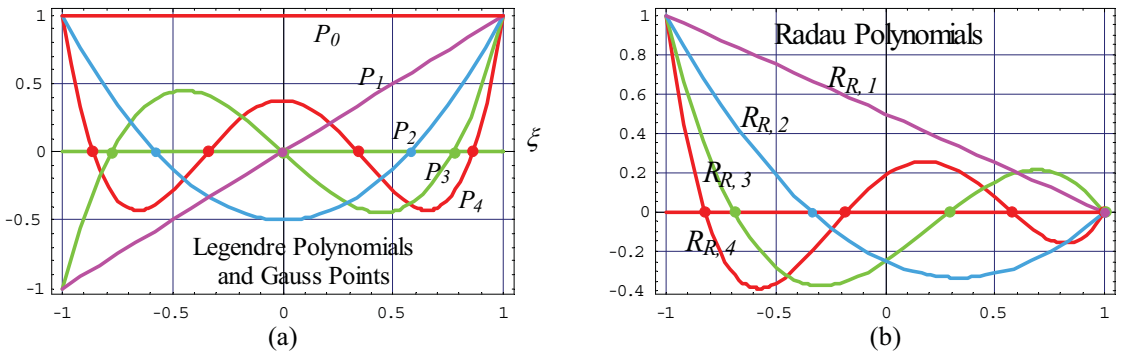


Figure 3.1. (a) Legendre polynomials and (b) right Radau polynomials.

The right Radau polynomial of degree k ($k \ge 1$) is defined by

$$R_{R,k} = \frac{(-1)^k}{2}(P_k - P_{k-1})\,. \tag{3.15}$$

The first few Radau polynomials are plotted in Fig. 3.1(b). Here, the letter R stands for 'Radau' and the subscript R for 'right'; the factor $(-1)^k$ is nonstandard and is needed so that (3.16a) below holds. Equation (3.15) implies that $R_{R,k}$ is orthogonal to $\boldsymbol{P}^{k-2}$. In addition,

$$R_{R,k}(-1)=1 \quad \text{and} \quad R_{R,k}(1)=0\,. \tag{3.16a,b}$$

Note that $R_{R,k}$, which is of degree k, is determined by the above two conditions and the $k-1$ conditions that it is orthogonal to $\boldsymbol{P}^{k-2}$. This definition of the Radau polynomial shows that it is a natural choice for the correction function. At the two boundaries, using (3.13), we have

$$R_{R,k}'(-1) = -k^2/2\,, \quad \text{and} \quad R_{R,k}'(1) = (-1)^{k-1}k/2\,. \tag{3.17a,b}$$

The zeros of the Radau polynomial $R_{R,k}$ are the k right Radau points.

The Lobatto polynomial of degree k ($k \geq 1$) is defined by

$$\mathrm{Lo}_k = P_k - P_{k-2}. \tag{3.18}$$

Using (3.15), the Lobatto polynomial can be expressed in terms of Radau polynomials:

$$\mathrm{Lo}_k = 2(-1)^k (R_{R,k} - R_{R,k-1}). \tag{3.19}$$

The zeros of the Lobatto polynomial of degree k are the k Lobatto points; they include the two boundary points ± 1. As shown in Fig. 3.1(b), they are also the ξ-coordinates of the intersections of the graphs of $R_{R,k}$ and $R_{R,k-1}$.

Returning to the correction functions, we always choose g^{RB} by reflection: $g^{\mathrm{RB}}(\xi) = g^{\mathrm{LB}}(-\xi)$. As a result, we need to focus only on g^{LB}. For simplicity of notation, set

$$g = g^{\mathrm{LB}}.$$

Since g is of degree K, it is determined by $K+1$ conditions. Two conditions, namely, $g(-1) = 1$ and $g(1) = 0$, are known; therefore, $K-1$ additional conditions remain. We discuss below three choices for g.

If the $K-1$ additional conditions are the requirement that g is orthogonal to $\boldsymbol{P}^{K-2}$, then g is the right Radau polynomial of degree K defined in (3.15). The resulting method yields a solution identical to that of DG (Huynh 2007). Therefore, we use the notation g_{DG}:

$$g_{\mathrm{DG}} = R_{R,K}. \tag{3.20}$$

The reason the DG scheme can be cast in the above reconstruction form is sketched below. In the current formulation, since the projections of g^{LB} and g^{RB} onto $\boldsymbol{P}^{K-2}(I)$ are zero, the projections of the polynomials $u_j^C(x)$ (degree K) and $u_j(x)$ (degree $K-1$) onto $\boldsymbol{P}^{K-2}(E_j)$ are identical. In the standard formulation, if ϕ is a test function of degree $K-1$, then ϕ_x is of degree $K-2$, and the expression $\int u_j \phi_x \, dx$ contains the projection of u_j onto $\boldsymbol{P}^{K-2}(E_j)$.

Loosely put, the current formulation is a finite-difference DG formulation (versus finite-element). It involves no quadratures and has the advantage of simplicity and economy. In addition, regardless of the choice of solution points, no matrix inversion is needed. For example, if we choose the Lobatto points as solution points, then, since the corresponding basis functions are not orthogonal, the standard DG formulation requires a matrix inversion, whereas the current formulation does not. In other words, the mass matrix inversion is built in.

For the next two choices, in addition to $g(-1) = 1$ and $g(1) = 0$, we require g to be orthogonal to $\boldsymbol{P}^{K-3}$ (yielding $K-2$ conditions) together with one additional condition. It was observed in (Huynh 2007) using Fourier analysis that the requirement of g being orthogonal to $\boldsymbol{P}^{K-3}$ gives rise to stable schemes for convection. (The converse is not true, however: there are correction functions not orthogonal to any $\boldsymbol{P}^k$ that result in stable schemes).

The second choice for g requires that it vanishes at the $K-1$ Gauss points. (These points are the zeros of P_{K-1} and are completely different from the K Gauss points, which typically are the solution points). For obvious reason, this choice is denoted by g_{Ga}. It will be shown later that

$$g_{\mathrm{Ga}} = \frac{K}{2K-1} R_{R,K} + \frac{K-1}{2K-1} R_{R,K-1}. \tag{3.21}$$

Note that this g results in a scheme very similar to the staggered-grid scheme of Kopriva and Kolias (1996) as well as the spectral-difference method of Liu et al. (2006). The key difference is that the current scheme employs only one grid, and it is stable, whereas the staggered-grid and spectral-difference methods employ two grids, and they are mildly unstable (Huynh 2007).

The third choice for g, requires that g' vanishes at $K-1$ of the K Lobatto points (Legendre-Lobatto, not Chebyshev-Lobatto). It can be written as (compared to g_{Ga}, the weights are switched),

$$g_{\text{Lump, Lo}} = \frac{K-1}{2K-1} R_{R,K} + \frac{K}{2K-1} R_{R,K-1}. \tag{3.22}$$

For this correction function, if *the solution points* ξ_k *are chosen to be the Lobatto points*, then the derivative $(u_j^{\text{LB}})_\xi$ at all solution points are the same as the uncorrected value $(u_j)_\xi$ except at the left boundary. In other words, this g lumps the correction to the left boundary—thus, the notation $g_{\text{Lump, Lo}}$.

We will also need the following derivative values at the left boundary: by (3.17a), (3.21), and (3.22),

$$g_{\text{DG}}'(-1) = -K^2/2, \quad g_{\text{Ga}}'(-1) = -[K(K-1)+1]/2, \text{ and } \quad g_{\text{Lump, Lo}}'(-1) = -K(K-1)/2 \tag{3.23}$$

The plots of the three correction functions for the cases of $K=2$ and $K=4$ are shown in Fig. 3.2. Loosely put, among the three correction functions, g_{DG} is the steepest, and $g_{\text{Lump, Lo}}$, least steep. Note that for each K, the ξ-coordinate of the intersections of the three curves are the K Lobatto points, and the derivative of $g_{\text{Lump, Lo}}$ vanishes at these points except at the left boundary $\xi=-1$. Also note that as K gets larger, g_{Ga} gets closer to $g_{\text{Lump, Lo}}$.

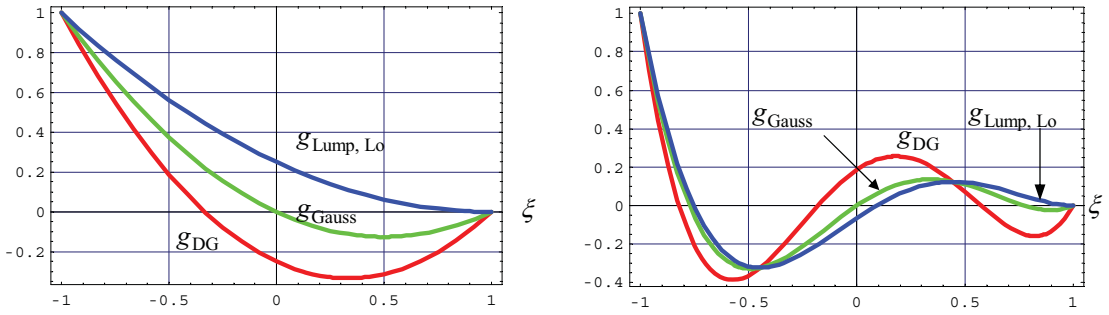


Figure 3.2. Correction functions for (a) $K=2$ *and (b)* $K=4$.

Concerning practical calculations, one can derive the equations for these correction functions (for arbitrary K) by a software package such as Mathematica or Maple.

Readers who are not interested in the proofs can skip the rest of this section with no loss in continuity. We now discuss the derivation of the last two correction functions. Since both $R_{R,K}$ and $R_{R,K-1}$ are orthogonal to $\boldsymbol{P}^{K-3}$, these two correction functions can be written as, with $0<\alpha<1$,

$$g = \alpha R_{R,K} + (1-\alpha) R_{R,K-1}. \tag{3.24}$$

Due to (3.16), the above obviously satisfies the conditions $g(-1)=1$ and $g(1)=0$.

Concerning g_{Ga} defined by (3.21), we need to prove that it vanishes at the $K-1$ Gauss points. To this end, by expanding (3.21) in terms of the Legendre polynomials using (3.15),

$$g_{\text{Ga}} = (-1)^K [K P_K - (2K-1) P_{K-1} + (K-1) P_{K-2}] \big/ [2(2K-1)].$$

By (3.11), or the recurrence formula for P_K, we have $(2K-1)\xi P_{K-1} = K P_K + (K-1) P_{K-2}$. Thus,

$$(2K-1)(\xi-1) P_{K-1} = K P_K - (2K-1) P_{K-1} + (K-1) P_{K-2}.$$

The above two equations imply that $g_{\text{Ga}} = [(-1)^K/2](\xi-1) P_{K-1}$; therefore, g_{Ga} and $(\xi-1)P_{K-1}$ have the same zeros. This completes the proof for g_{Ga}.

As for $g_{\text{Lump, Lo}}$, it is defined by (3.24) where α is chosen so that the point $\xi=1$ is a zero of multiplicity two, i.e., $g'(1)=0$. Using (3.17b), one can verify that $g'(1)=0$ if $\alpha=(K-1)/(2K-1)$. This α results in (3.22). It turns out that $(g_{\text{Lump, Lo}})'$ vanishes not only at $\xi=1$, but also at $K-1$ of the K Lobatto points. (The proof of this fact involves some algebra and can be found in (Huynh 2007).)

4. Evaluation of the Second Derivative

Unless otherwise stated, all discussions concerning stability and accuracy are based on Fourier (Von Neumann) analysis. Recall that for the diffusion equation, if an explicit time-stepping method is employed (say, Runge-Kutta), the time step size has a limit, above which the solution becomes unstable. This stability limit is inversely proportional to the largest magnitude of the eigenvalues of the discretization. It turns out that the eigenvalues are real and negative. Therefore, to compare the stability limits of schemes, we only need to compare the minimum eigenvalues: the scheme with a larger minimum eigenvalue (i.e., the minimum eigenvalue has a smaller magnitude) has the advantage of possessing a larger stability limit.

Next, a scheme using K solution points is super-convergent (or super-accurate) if its order of accuracy is higher than the expected order of K. All schemes discussed here are super-convergent for $K \geq 3$: their orders of accuracy are at least $2(K-1)$. In general, super-convergence does not carry over to nonlinear equations: due to nonlinear errors, the order of accuracy is typically only K. For comparison, we discuss the main trade-offs in accuracy and stability here; additional details will be provided in §6. In general, a less steep correction function for the solution points results in a scheme with the advantage of a larger stability limit and the disadvantage of lower accuracy.

Returning to our current task, recall that $\{u_j^C\}$ is continuous across the interfaces; as a result, $(u_j^C)_x$ provides an approximation to u_x that takes into account the data interaction. For each cell E_j, at the solution points $x_{j,k}$, set

$$v_{j,k} = (u_j^C)_x(x_{j,k}) . \tag{4.1}$$

These K (derivative) values define a polynomial of degree $K-1$ denoted by $v_j(x)$ that is identical to $(u_j^C)_x(x)$. Note that the function $\{v_j\} = \{v_j(x)\} = \{(u_j^C)_x\}$ can be and usually is discontinuous across the interfaces. Therefore, to calculate u_{xx}, we can apply the reconstruction procedure once again to the discontinuous function $\{v_j\} = \{(u_j^C)_x\}$. To this end, we must define a common derivative (or common gradient) at the cell interfaces.

Common derivatives. At each interface, in formula (3.1) for the common value, with $0 \leq \kappa \leq 1$, the weight for $u_{j+1/2}^-$ is κ and that for $u_{j+1/2}^+$, $1-\kappa$. To define the common derivative, we switch the two weights. Loosely put, this switch serves the purpose of canceling out the one-sidedness so that the scheme is consistent with the centered or unbiased nature of the diffusion process. Since the corrected derivative $(u_j^C)_x$ is readily available, an obvious choice is

$$(u_x)_{j+1/2}^{\text{com}} = (1-\kappa)(u_j^C)_x(x_{j+1/2}) + \kappa(u_{j+1}^C)_x(x_{j+1/2}) . \tag{4.2}$$

The function u_j^C involves the data in the cells $j-1$, j, and $j+1$. Consequently, the above formula has a four-cell stencil involving the data from E_{j-1} to E_{j+2} as can be seen in Fig. 4.1(a) for the centered case ($\kappa = 1/2$). Since the calculation of u_{xx} in the cell E_j employs $(u_x)_{j-1/2}^{\text{com}}$ and $(u_x)_{j+1/2}^{\text{com}}$, the corresponding scheme has a five-cell stencil.

We now define a common derivative at each interface $x_{j+1/2}$ in a manner that it involves only the data in the two adjacent cells, namely, E_j and E_{j+1}. A scheme with such a compact stencil is desirable since it is easy to code, the boundary conditions involved are simple, and the resulting implicit version has a sparser as well as invertible matrix. With $u_{j+1}^{\text{LB}}(x)$ by (3.7) and $u_j^{\text{RB}}(x)$ by (3.8), set

$$(u_x)_{j+1/2}^{\text{com}} = (1-\kappa)(u_j^{\text{RB}})_x(x_{j+1/2}) + \kappa(u_{j+1}^{\text{LB}})_x(x_{j+1/2}) . \tag{4.3}$$

The functions u_j^{RB} and u_{j+1}^{LB} for the centered case are shown in Fig. 4.1(b) and, for the one-sided case with $\kappa = 1$, Fig. 4.2(a). The above can be expressed as:

$$(u_x)^{\text{com}}_{j+1/2} = (1-\kappa)\,\frac{2}{h_j}\left\{(u_j)_\xi(1) + [u^{\text{com}}_{j+1/2} - u_j(1)]\,[(g^{\text{RB}})'(1)]\right\} + \kappa\,\frac{2}{h_{j+1}}\left\{(u_{j+1})_\xi(-1) + [u^{\text{com}}_{j+1/2} - u_{j+1}(-1)]\,[(g^{\text{LB}})'(-1)]\right\}. \tag{4.4}$$

Note the dependence only on $u^{\text{com}}_{j+1/2}$ and the data on E_j and E_{j+1}.

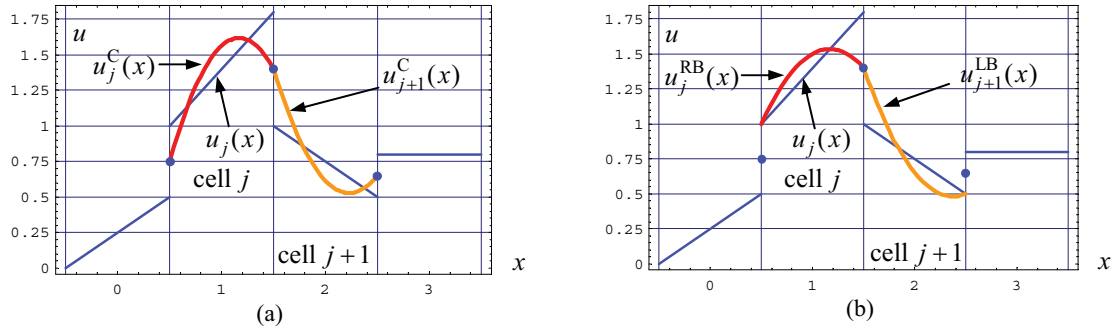


Figure 4.1. Centered common value and derivative. (a) $(u_x)^{\text{com}}_{j+1/2} = (1/2)\,[(u_j^C)_x(x_{j+1/2}) + (u_{j+1}^C)_x(x_{j+1/2})]$ *has a four-cell stencil. (b)* $(u_x)^{\text{com}}_{j+1/2} = (1/2)\,[(u_j^{\text{RB}})_x(x_{j+1/2}) + (u_{j+1}^{\text{LB}})_x(x_{j+1/2})]$ *has a two-cell stencil. In both figures, the solution polynomials are linear, and the correction function is* g_{DG}.

As for the boundary corrections $(g^{\text{LB}})'(-1)$ and $(g^{\text{RB}})'(1)$, they can be expressed in a formula. Recall that due to symmetry, $(g^{\text{RB}})'(1) = -(g^{\text{LB}})'(-1)$. The three correction functions in the order of increasing steepness are $g_{\text{Lump, Lo}}$, g_{Gauss}, and g_{DG}. The corresponding $(g^{\text{RB}})'(1)$ are, respectively, by (3.23),

$$K(K-1)/2, \quad [K(K-1)+1]/2, \quad \text{and} \quad K^2/2.$$

Note that for steady state problems, if we use $g_{\text{Lump, Lo}}$ for the definition of the common interface derivative, the resulting matrix for the solution of $u_{j,k}$ becomes singular. Thus, $g_{\text{Lump, Lo}}$ is to be avoided for steady state case. (It can still be used for the correction at solution points, however.)

Also note that at the common interface $x_{j+1/2}$, the common derivative evaluated by (4.3) or (4.4) involve corrections only at that interface. On the other hand, at the solution points of E_j, the derivative evaluated by (4.1) using u_j^{C} involves corrections at both interfaces of E_j.

With the common derivative defined by (4.4), set

$$v^{\text{com}}_{j+1/2} = (u_x)^{\text{com}}_{j+1/2}.$$

Finally, by applying the reconstruction procedure to the data $v_{j,k}$ of (4.1) and the above common quantities, at each solution point, the *corrected* second derivative is given by

$$(u_{xx})_{j,k} = (v_j^C)_x(x_{j,k}). \tag{4.5}$$

A few remarks are in order concerning the trade-offs between (4.2), which has a wide stencil, and (4.3), which has a compact stencil.

If the centered common value is employed, then as shown in Fig. 4.1, expressions (4.2) and (4.3) yield different schemes. These two schemes are comparable in terms of accuracy and stability; the scheme using (4.2) is slightly more accurate but its stability limit is smaller. However, the disadvantages of (4.2) are numerous: its stencil is large and thus, the coding of boundary conditions is more complex; it has a zero spurious eigenvalue with an undamped eigenfunction; and the steady state case results in a singular matrix. Therefore, unless otherwise stated, we employ only the compact stencil formula, namely, (4.3).

If g_{DG} is employed as correction functions at both the solution points and interfaces, then, with the centered common values, (4.2) results in a method identical to BR1, and (4.3), BR2 (Bassi and Rebay

1997a, 2000). These two BR schemes were originally formulated in the finite-element framework whereas the current versions are formulated in the finite-difference one.

In the case of common quantities via one-sided choice (e.g., the value to the left and the derivative to the right), for the one-dimensional case discussed here, the schemes corresponding to (4.2) and (4.3) are identical. For the two-dimensional case, however, they can be different. If, in addition, g_{DG} is employed as correction functions for both the interfaces and the solution points, then (4.2) results in a method identical to the local discontinuous Galerkin or LDG (Cockburn and Shu 1998), and (4.3), the compact discontinuous Galerkin or CDG (Peraire and Persson 2008). Again, these DG schemes were originally formulated in the finite-element framework whereas the current versions are formulated in the finite-difference one.

Comparing the centered versus one-sided common values with g_{DG} as correction function (i.e., BR2 versus CDG), the former is of order $2(K-1)$; the latter, order $2K$; the former, however, has the advantage that its stability limit is more than two times larger than the latter. Since super-convergence (accuracy higher than K) does not hold for the general case of nonlinear equations, the scheme using centered common values appear to have an edge. (For additional comparisons, see the numerical examples in §8.)

Criteria with high accuracy for common derivatives. Continuing with the derivation of the common derivative, new criteria with the advantage of high accuracy can be obtained as follows.

We reconstruct the solution on the domain $[x_{j-1/2}, x_{j+3/2}] = E_j \cup E_{j+1}$ in a manner that (a) it takes on the common value which is the weighted average (3.1) at $x_{j+1/2}$, and (b) it approximates the data on E_j and E_{j+1} without the requirement of leaving the interface values at $x_{j-1/2}$ and $x_{j+3/2}$, respectively, unchanged. In other words, for the cell E_{j+1}, to correct for its left boundary $x_{j+1/2}$, with $g = g^{\mathrm{LB}}$, we require that $g(-1) = 1$, while $g(1)$ is allowed to be arbitrary and, critically, we require g to approximate the zero function to a degree as high as possible. A similar statement holds for the cell E_j and its right boundary $x_{j+1/2}$. Thus, to calculate the common derivative (but *not* the derivative at the solution points) we can employ the Legendre polynomial of degree K, which is orthogonal to $\boldsymbol{P}^{K-1}$, as the correction function:

$$g_{\mathrm{Le}} = g^{\mathrm{LB}} = (-1)^K P_K \quad \text{and} \quad g^{\mathrm{RB}} = P_K . \tag{4.6}$$

By (3.13b), for this correction function,

$$(g^{\mathrm{RB}})'(1) \;=\; -(g^{\mathrm{LB}})'(-1) \;=\; K(K+1)/2 .$$

The common derivative is again given by (4.4).

For the case of centered common values using g_{DG} as correction function at the solution points, the method using (4.6) has the following order of accuracy: for odd K, the order is $2K$, and for even K, $2(K-1)$; e.g., both the schemes with $K=3$ and $K=4$ are of order 6.

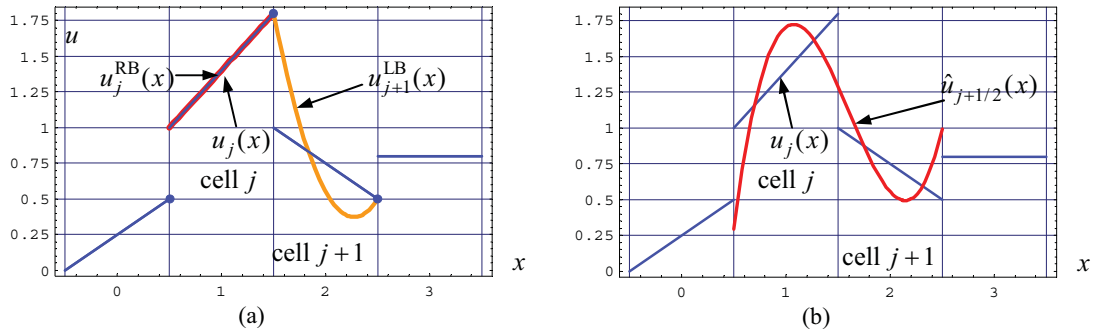


Figure 4.2. (a) The common derivative $(u_x)^{\mathrm{com}}_{j+1/2} = (u^{\mathrm{LB}}_{j+1})_x(x_{j+1/2})$, *which corresponds to (4.3) with* $\kappa = 1$, *has a two-cell stencil. (b) Using Van Leer and Nomura's recovery, the common value* $\hat{u}(x_{j+1/2})$ *and common derivative* $\hat{u}_x(x_{j+1/2})$ *have a two-cell stencil.*

We present now a new criterion for the common quantities (both value and derivative) that is centered (or unbiased) and, with g_{DG} as correction function at the solution points, yields a scheme of order $2K$ for all K. At each interface $x_{j+1/2}$, let $u^{\mathrm{com}}_{j+1/2}$ be an unknown. Using the Legendre polynomial (4.6) as correction function for the interface (a different g results in a lower order of accuracy), the corresponding corrected derivatives $(u_j^{\mathrm{RB}})_x(x_{j+1/2})$ and $(u_{j+1}^{\mathrm{LB}})_x(x_{j+1/2})$ can easily be obtained as linear functions of $u^{\mathrm{com}}_{j+1/2}$. The unknown $u^{\mathrm{com}}_{j+1/2}$ is then determined by requiring that these two derivatives are equal:

$$(u_j^{\mathrm{RB}})_x(x_{j+1/2}) = (u_{j+1}^{\mathrm{LB}})_x(x_{j+1/2}). \tag{4.7}$$

In other words, on $(x_{j-1/2}, x_{j+3/2})$, both the reconstruction and its derivative are required to be continuous across $x_{j+1/2}$. (Note that the continuity requirement simultaneously for all interfaces involving a tridiagonal solver was briefly discussed by Kopriva (1998) for his staggered-mesh methods. The current method is considerably simpler, however.)

For our final criterion, we define both $u^{\mathrm{com}}_{j+1/2}$ and $(u_x)^{\mathrm{com}}_{j+1/2}$ via the recovery approach of Van Leer and Nomura (2006). Let the recovery function $\hat{u}$ be a polynomial of degree $2K-1$ on the domain $[x_{j-1/2}, x_{j+3/2}] = E_j \cup E_{j+1}$ determined by the $2K$ conditions that: $\hat{u}$ and u_j have the same projection on $\boldsymbol{P}^{K-1}(E_j)$, and $\hat{u}$ and u_{j+1} have the same projection on $\boldsymbol{P}^{K-1}(E_{j+1})$. In other words, for $0 \le i \le K-1$,

$$\int_{E_j} \hat{u}(x)\,(x-x_j)^i\,dx = \int_{E_j} u_j(x)(x-x_j)^i\,dx, \tag{4.8a}$$

and

$$\int_{E_{j+1}} \hat{u}(x)\,(x-x_{j+1})^i\,dx = \int_{E_{j+1}} u_{j+1}(x)(x-x_{j+1})^i\,dx. \tag{4.8b}$$

The common quantities at the interface are then defined as $\hat{u}(x_{j+1/2})$ and $(\hat{u}_x)(x_{j+1/2})$. See Fig. 4.2(b).

Compared to the averages (3.1) and (4.3), the common quantities via recovery above has the advantage of higher accuracy and the disadvantage of being costly. In fact, based on Fourier analysis for K up to 10 (with $K > 10$, the errors become exceedingly small), the gain in accuracy is considerable: using g_{DG} as correction function at the solution points, for odd K, the order of the recovery scheme is $3K-1$, and for even K, $3K-2$. A peculiar property of these methods is the following: for $K \ge 4$, two of the spurious eigenvalues become complex for a certain range of wave numbers, which means that the corresponding components are propagated as well as damped. These complex eigenvalues, however, pose no problem since the corresponding components are damped very quickly. Also note that we can consider the averages (3.1) and (4.3) as lower order approximations to $\hat{u}(x_{j+1/2})$ and $(\hat{u}_x)(x_{j+1/2})$, respectively.

Algorithm. The algorithm for the diffusion equation is summarized below. Let the data $u_{j,k}$ at the solution points $x_{j,k}$ be given for all cells.

- At each interface $x_{j+1/2}$, calculate $u^-_{j+1/2}$ and $u^+_{j+1/2}$ by (2.8a,b) and the common value $u^{\mathrm{com}}_{j+1/2}$ by (3.1). Next, obtain the common derivative $(u_x)^{\mathrm{com}}_{j+1/2}$ by (4.4). Alternatively, $u^{\mathrm{com}}_{j+1/2}$ and $(u_x)^{\mathrm{com}}_{j+1/2}$ can be estimated via the continuity requirement (4.7) or via the recovery $\hat{u}$ defined by (4.8).
- For each cell, at the solution points, calculate $v_j(x_{j,k}) = (u_j^C)_x(x_{j,k})$ via (3.10). With $v^{\mathrm{com}}_{j+1/2} = (u_x)^{\mathrm{com}}_{j+1/2}$ at the interfaces, the values of u_{xx} at the solution points are given by $(v_j^C)_x(x_{j,k})$.

Boundary conditions. We end this section with a discussion of boundary conditions. Let the left boundary of the domain be $x_{l+1/2}$ for some l. If we have a Dirichlet condition u_D at this boundary, then we define the common value by

$$u^{\mathrm{com}}_{l+1/2} = u^{-}_{l+1/2} = u_D \,. \tag{4.9a}$$

This definition corresponds to $\kappa = 1$ in (3.1). Next, using (4.3) for the common derivative, set

$$(u_x)^{\mathrm{com}}_{l+1/2} \;=\; (u^{\mathrm{LB}}_{l+1})_x(x_{l+1/2}) \,. \tag{4.9b}$$

On the other hand, if we have a Neumann boundary condition $u_x = (u_x)_N$ at the left boundary $x_{l+1/2}$ for some real number $(u_x)_N$, then we define

$$(u_x)^{\mathrm{com}}_{l+1/2} = (u_x)^{-}_{l+1/2} = (u_x)_N \,. \tag{4.10a}$$

The value $u^{\mathrm{com}}_{l+1/2}$ is determined by the condition that $(u^{\mathrm{LB}}_{l+1})_x(x_{l+1/2}) = u_N$, i.e.,

$$u^{\mathrm{com}}_{l+1/2} \;=\; u_{l+1}(-1) \;+\; [(h_{l+1}/2)u_N - (u_{l+1})_\xi(-1)] \Big/ (g^{\mathrm{LB}})'(-1) \,. \tag{4.10b}$$

The boundary conditions on the right end of the domain follow by reflection.

As will be shown later, for the case where the exact solution is a parabola, and two Gauss points (linear element) are employed in each cell, the standard DG schemes do not recover the exact solution. Together with the above boundary conditions, the two schemes that recover the exact solution are: the scheme using the Legendre polynomial with the continuity requirement (4.7) and the recovery scheme (4.8).

5. Fourier (Von Neumann) Analysis

Fourier analysis provides information on both stability and accuracy. On (–∞, ∞), consider the equation (2.1), i.e., $u_t = u_{xx}$, with the periodic initial condition $u_{\mathrm{init}}(x) = e^{iwx}$, where the wave number w lies between $-\pi$ and π. Low frequency data corresponds to w of small magnitude; high frequency, to w near $\pm\pi$. The exact solution is $u_{\mathrm{exact}}(x,t) = e^{-w^2 t} e^{iwx}$. At $(x,t) = (0,0)$,

$$(u_{\mathrm{exact}})_t(0,0) = -w^2 \,. \tag{5.1}$$

The cells are $E_j = [j - 1/2, j + 1/2]$. The data are $u_{j,k} = \exp[\,i\,w(j + \xi_k/2)]$. However, it is not the data but their following property, which plays a key role in the calculation of eigenvalues,

$$u_{j-1,k} = e^{-iw} u_{j,k} \,. \tag{5.2}$$

To calculate the eigenvalues, the K solution points are grouped together as a vector: set

$$\boldsymbol{u}_j = \begin{pmatrix} u_{j,1} \\ \vdots \\ u_{j,K} \end{pmatrix} . \tag{5.3}$$

For the case of three-cell stencil, the solution can be expressed as

$$\frac{d\boldsymbol{u}_j}{dt} = \sum_{l=-1}^{1} \boldsymbol{C}_l \, \boldsymbol{u}_{j+l} \,. \tag{5.4}$$

Using (5.2), we replace $\boldsymbol{u}_{j+l}$ by $e^{ilw}\boldsymbol{u}_j$,

$$\frac{d\boldsymbol{u}_j}{dt} = \Big(\sum_{l=-1}^{1} e^{ilw}\boldsymbol{C}_l\Big)\; \boldsymbol{u}_j \tag{5.5}$$

where $\boldsymbol{C}_l$ are $K \cdot K$ matrixes. Thus, the spatial discretization (semidiscretization) results in

$$\frac{d\boldsymbol{u}_j}{dt} = \boldsymbol{S}\,\boldsymbol{u}_j \quad \text{where} \quad \boldsymbol{S} = \sum_{l=-1}^{1} e^{ilw}\boldsymbol{C}_l \,. \tag{5.6, 5.7}$$

Here, ***S***, for 'space' or 'semidiscrete', is a $K \cdot K$ matrix with *K* eigenvalues. The one approximating $-w^2$ (the exact eigenvalue for $\partial^2/\partial x^2$) is called the *principal eigenvalue* and is denoted by $S(w)$:

$$S(w) \approx -w^2 \,. \tag{5.8}$$

All others are spurious eigenvalues. For stability, all eigenvalues must have nonpositive real parts.

If an explicit time-stepping method such as the standard four-stage Runge-Kutta is employed, then the stability limit of a scheme is proportional to the inverse of the largest magnitude of all eigenvalues.

Accuracy. Concerning accuracy, note first that for a uniform mesh of cell width h, S approximates $h^2\partial^2/\partial x^2$. A scheme is accurate to order m if S approximates $h^2\partial^2/\partial x^2$ to $O(h^{m+2})$, i.e., for small w,

$$S(w) = -w^2 + O(w^{m+2}) . \tag{5.9}$$

In practice, it is difficult if not impossible to derive an expression for the principal eigenvalue $S(w)$ when the number of solution points is greater than 2 or 3. Therefore, we obtain the order of accuracy by the following approximation. Let w be fixed, say, $w = \pi/8$. We calculate the error

$$E(w) = S(w) + w^2 . \tag{5.10}$$

By halving w (i.e., doubling the number of cells), the error corresponding to $w/2$ (say, $w/2 = \pi/16$) is

$$E(w/2) = S(w/2) + (w/2)^2 .$$

Since $O((w/2)^{m+2}) = (1/2)^{m+2} O(w^{m+2})$, for a scheme to be m-th order accurate,

$$E(w/2) \approx (1/2)^{m+2} E(w) . \tag{5.11}$$

That is, $E(w)/E(w/2) \approx 2^{m+2}$. Using the logarithm function, the order of accuracy is given by the integer

$$m \approx \left[\mathrm{Log}\left(\frac{E(w)}{E(w/2)} \right) \Big/ \mathrm{Log}(2) \right] - 2 . \tag{5.12}$$

Choice of solution points. Next, we discuss the choice of solution points. The two common choices are the Gauss and Lobatto points. While these choices may be advantageous for the nonlinear case, from the consideration of linear stability and accuracy, all choices are equivalent. In fact, we claim that the eigenvalues of $\boldsymbol{S}$ are independent of the solution points chosen.

To prove this claim, consider the local coordinate ξ on $I = [-1,1]$ with solution points ξ_k and data u_k, $k = 1,...,K$. Denote the vector $\{u_k\}_{k=1}^K$ by $\boldsymbol{u}$. Recall that the solution polynomial is $p = \sum_{k=1}^K u_k \phi_k(\xi)$. Next, let $\widetilde{\xi}_l$, $l = 1,...,K$ be another set of solution points (we assume no two points are the same), and let $\widetilde{u}_l$ be the value $p(\widetilde{\xi}_l)$,

$$\widetilde{u}_l = \sum_{k=1}^K u_k \phi_k(\widetilde{\xi}_l) . \tag{5.13}$$

Set $\widetilde{\boldsymbol{u}} = \{\widetilde{u}_l\}_{l=1}^K$. In addition, for $1 \le k, l \le K$, set

$$m_{lk} = \phi_k(\widetilde{\xi}_l) . \tag{5.14}$$

Denote the $K \cdot K$ matrix $\{m_{lk}\}$ by $\boldsymbol{M}$. Then, it follows from the above two expressions that

$$\widetilde{\boldsymbol{u}} = \boldsymbol{M}\boldsymbol{u} . \tag{5.15}$$

Since we can obtain $\boldsymbol{u}$ from $\widetilde{\boldsymbol{u}}$, $\boldsymbol{M}$ is invertible. Next, equation (5.7a) takes the form,

$$d\boldsymbol{u}/dt = \boldsymbol{S}\,\boldsymbol{u} . \tag{5.16}$$

For the solution points $\widetilde{\xi}_l$, the corresponding equation is

$$d\widetilde{\boldsymbol{u}}/dt = \widetilde{\boldsymbol{S}}\,\widetilde{\boldsymbol{u}} . \tag{5.17}$$

We wish to show that $\widetilde{\boldsymbol{S}}$ and $\boldsymbol{S}$ are similar, and thus, have the same eigenvalues. To this end, multiplying (5.16) on the left by $\boldsymbol{M}$, we have

$$\boldsymbol{M}\, d\boldsymbol{u}/dt = \boldsymbol{M}\boldsymbol{S}\,\boldsymbol{u} .$$

Since $\boldsymbol{M}$ is independent of t and is invertible

$$d(\boldsymbol{M}\,\boldsymbol{u})/dt = \boldsymbol{M}\boldsymbol{S}\boldsymbol{M}^{-1}\boldsymbol{M}\,\boldsymbol{u} = (\boldsymbol{M}\boldsymbol{S}\boldsymbol{M}^{-1})(\boldsymbol{M}\,\boldsymbol{u}) . \tag{5.18}$$

Therefore, by (5.18),

$$d\widetilde{\boldsymbol{u}}/dt = (\boldsymbol{M}\boldsymbol{S}\boldsymbol{M}^{-1})\,\widetilde{\boldsymbol{u}} . \tag{5.19}$$

Comparing (5.17) with the above and, since $\widetilde{\boldsymbol{u}}$ is arbitrary, we have

$$\widetilde{\boldsymbol{S}} = \boldsymbol{M}\boldsymbol{S}\boldsymbol{M}^{-1} . \tag{5.20}$$

This completes the proof.

6. Stability and Accuracy of Schemes for Diffusion

In this section, we employ the Fourier analysis of the previous section; therefore, the mesh is uniform, and the cell width is 1. With K solution points, all schemes discussed have order of accuracy of at least $2(K-1)$. As a consequence, for $K \geq 2$, the order of accuracy is at least 2. For $K = 1$, however, the order of accuracy can be 0, which implies that the corresponding scheme is inconsistent. Again, concerning the super-convergence or super-accuracy property (order higher than K), for the nonlinear case such as the Navier-Stokes equations, due to nonlinear errors, it is expected that the schemes discussed are accurate to order no higher than K. Unless otherwise stated, all schemes have real negative eigenvalues (≤ 0).

A scheme is determined by the following choices. First, for the common quantities at each interface, the four choices are: (a) centered, (b) one-sided, (c) continuous (condition (4.7) or the derivative values to the left and right of the interface being equal), and (d) recovery, namely (4.8). Next, for the cases (a), (b), and (c), the derivative at the interface requires a correction function; in the order of decreasing steepness, the choices are: g_{Le}, $g_{\text{DG}} = g_{\text{Ra}}$, g_{Ga}, and $g_{\text{Lump, Lo}}$. To denote the scheme using a certain choice at the interface, we use the following self-explanatory notation: $\text{I-centered-}g_{\text{Ga}}$, or $\text{I-continuous-}g_{\text{DG}}$, or I-recovery. Finally, for the derivative values at the solution points, the correction functions in the order of decreasing steepness are: g_{DG}, g_{Ga}, and $g_{\text{Lump, Lo}}$. The corresponding schemes are denoted by, $\text{SP-}g_{\text{DG}}$, etc... We refer to a scheme by the choice at the interface and that at the solution points, e.g.,

$$\text{I-centered-}g_{\text{DG}}\,/\,\text{SP-}g_{\text{Ga}}\,.$$

That is, the common quantities are centered, and the derivatives just to the left and right of the interface are calculated via g_{DG}; for the solution points, the correction function is g_{Ga}.

6.1. The case $K = 1$. Here,

$$g = g_{\text{DG}} = g_{\text{Ga}} = g_{\text{Lump, Lo}} = (1-\xi)/2\,.$$

Therefore, there is only one correction function for the solution point. At the interface, however, there are two correction functions: the above g and the Legendre polynomial $g_{\text{Le}} = -\xi$.

The three schemes $\text{I-centered-}g_{\text{Le}}$, I-recovery, and $\text{I-one-sided-}g_{\text{DG}}$ yield a result identical to the standard second-order accurate method

$$(u_{xx})_j = u_{j+1} + u_{j-1} - 2u_j\,.$$

Next, for the scheme $\text{I-centered-}g_{\text{DG}}\,/\,\text{SP-}g_{\text{DG}}$, i.e., BR2 scheme, we have $u^{\text{com}}_{j+1/2} = (u_j + u_{j+1})/2$. Therefore,

$$(u_x)^{+}_{j+1/2} = u_{j+1} - u^{\text{com}}_{j+1/2} = (u_{j+1} - u_j)/2 \quad \text{and} \quad (u_x)^{-}_{j+1/2} = u^{\text{com}}_{j+1/2} - u_j = (u_{j+1} - u_j)/2\,.$$

Thus, $(u_x)^{\text{com}}_{j+1/2} = (u_{j+1} - u_j)/2$, and

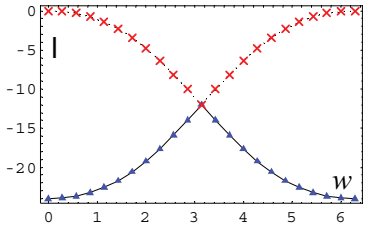
$$(u_{xx})_j = (u_{j+1} + u_{j-1} - 2u_j)/2\,.$$

This quantity approximates $u_{xx}/2$. As a result, this scheme has an error of $O(1)$ and is inconsistent.

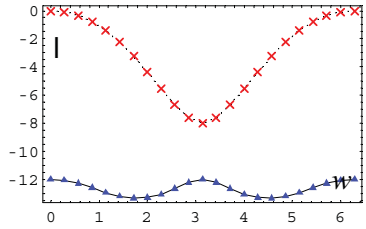
6.2. The case $K = 2$. All correction functions are different from each other (see Fig. 3.2(a)). The plots of eigenvalues as functions of wave numbers for six representative schemes are shown in Fig. 6.1.

The minimum eigenvalues, orders of accuracy, and errors of 18 schemes (with $K = 2$) are shown in Table 6.1. The last column shows the value $g'(-1)$ where g is the correction function for the left interface. Note that all schemes are of order 2 or higher. The three schemes 11, 14, and 15 are fourth-order accurate. Schemes 4, 7, and 10, which employ $g_{\text{Lump, Lo}}$ at the interfaces, have a zero spurious eigenvalue at $w = 0$; thus, the corresponding eigenvector is not damped. (For the steady state case, such a scheme results in a singular matrix for the solution and is to be avoided.) Scheme 10 is a limiting case: the spurious and principal eigenvalues become identical for all w. Scheme 15 (which is identical to CDG and LDG) has the drawback that the stability limit is rather small (the magnitude of the minimum eigenvalue is second largest on the list). Scheme 0 or $\text{I-centered-}3g_{\text{DG}}/\text{SP-}g_{\text{DG}}$ has the smallest stability limit of all with no improvement in accuracy. In general, a steeper correction for the derivative at interface results in a scheme with a smaller stability limit. From here on, therefore, we will not use a correction function steeper than

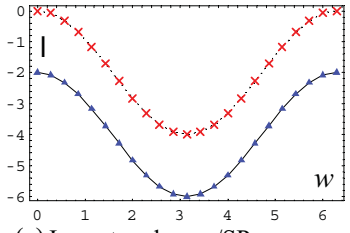
g_{Le} at the interface. As a limiting case, scheme I - continuous - $g_{\mathrm{Lump,\,Lo}}$ / SP - $g_{\mathrm{Lump,\,Lo}}$ (not listed) is inconsistent: it has an eigenvalue that is identically 0 for all w.



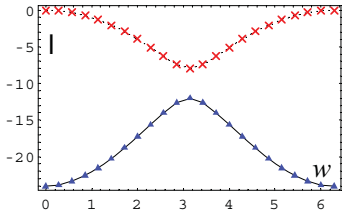
(a) I - centered - g_{Le}/SP - g_{DG}



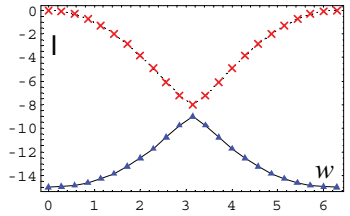
(b) I - centered - g_{DG}/SP - g_{DG}



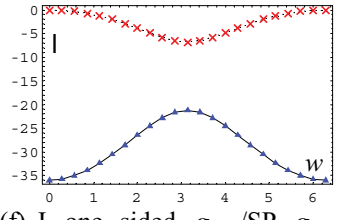
(c) I - centered - g_{Ga}/SP - $g_{\mathrm{Lump,\,Lo}}$



(d) I - continuous - g_{Le}/SP - g_{DG}



(e) I - recovery/SP - g_{DG}

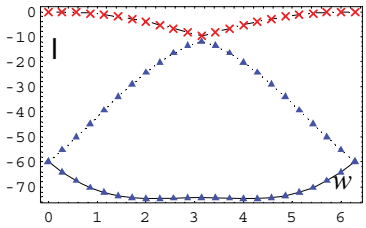


(f) I - one - sided - g_{DG}/SP - g_{DG}

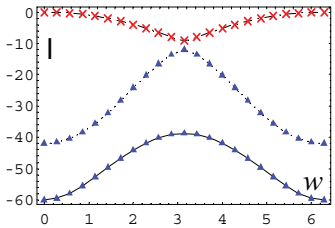
Figure 6.1. *Eigenvalues as functions of wave numbers for six representative schemes with* $K = 2$.

Scheme for $K = 2$	Min. Eigval.	Ord. Acc.	Error	For I - g, $g'(-1)$
0. I - centered - $3g_{\mathrm{DG}}$/SP - g_{DG}	–60.	2	$-w^4/12 + O(w^6)$	–6
1. I - centered - g_{Le}/SP - g_{DG}	–24.	2	$-w^4/12 + O(w^6)$	–3
2. I - centered - g_{DG}/SP - g_{DG}	–13.	2	$-w^4/12 + O(w^6)$	–2
3. I - centered - g_{Ga}/SP - g_{DG}	–12.	2	$-w^4/12 + O(w^6)$	–1.5
4. I - centered - $g_{\mathrm{Lump,\,Lo}}$/SP - g_{DG} *	–12.	2	$w^4/12 + O(w^6)$	–1
5. I - centered - g_{DG}/SP - g_{Ga}	–9.7	2	$-w^4/24 + O(w^6)$	–2
6. I - centered - g_{Ga}/SP - g_{Ga}	–8.	2	$-w^4/24 + O(w^6)$	–1.5
7. I - centered - $g_{\mathrm{Lump,\,Lo}}$/SP - g_{Ga} *	–8.	2	$w^4/12 + O(w^6)$	–1
8. I - centered - g_{DG}/SP - $g_{\mathrm{Lump,\,Lo}}$	–8.	2	$w^4/12 + O(w^6)$	–2
9. I - centered - g_{Ga}/SP - $g_{\mathrm{Lump,\,Lo}}$	–6.	2	$w^4/12 + O(w^6)$	–1.5
10. I - centered - $g_{\mathrm{Lump,\,Lo}}$/SP - $g_{\mathrm{Lump,\,Lo}}$ *	–4.	2	$w^4/12 + O(w^6)$	–1
11. I - continuous - g_{Le}/SP - g_{DG}	–24.	4	$w^6/1440 + O(w^8)$	–3
12. I - continuous - g_{DG}/SP - g_{DG}	–12.	2	$w^4/24 + O(w^6)$	–2
13. I - continuous - g_{Le}/SP - g_{Ga}	–16.	2	$w^4/24 + O(w^6)$	–3
14. I - recovery/SP - g_{DG}	–15.	4	$w^6/360 + O(w^8)$	*
15. I - one - sided - g_{DG}/SP - g_{DG}	–36.	4	$w^6/540 + O(w^8)$	–2
16. I - one - sided - $g_{\mathrm{Lump,\,Lo}}$/SP - $g_{\mathrm{Lump,\,Lo}}$	–10.5	2	$w^4/3 + O(w^6)$	–1
17. I - centered - g_{DG} (4 - cell)/SP - g_{DG}	–16.	2	$-w^4/24 + O(w^6)$	–2

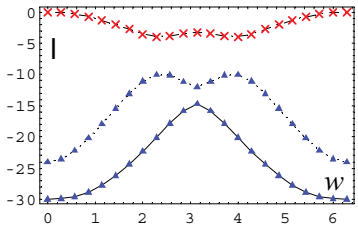
Table 6.1. *Minimum eigenvalues, orders of accuracy, and errors for schemes with* $K = 2$.



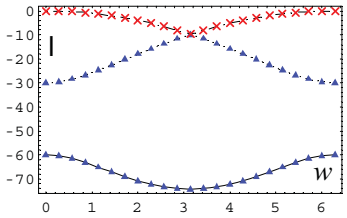
(a) I - centered - g_{Le}/SP - g_{DG}



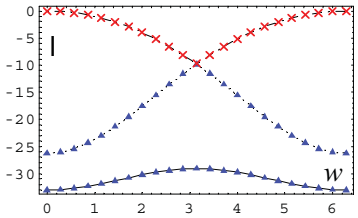
(b) I - centered - g_{DG}/SP - g_{DG}



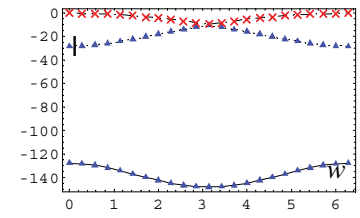
(c) I - centered - g_{Ga}/SP - $g_{\mathrm{Lump,\,Lo}}$



(d) I - continuous - g_{Le}/SP - g_{DG}



(e) I - recovery/SP - g_{DG}

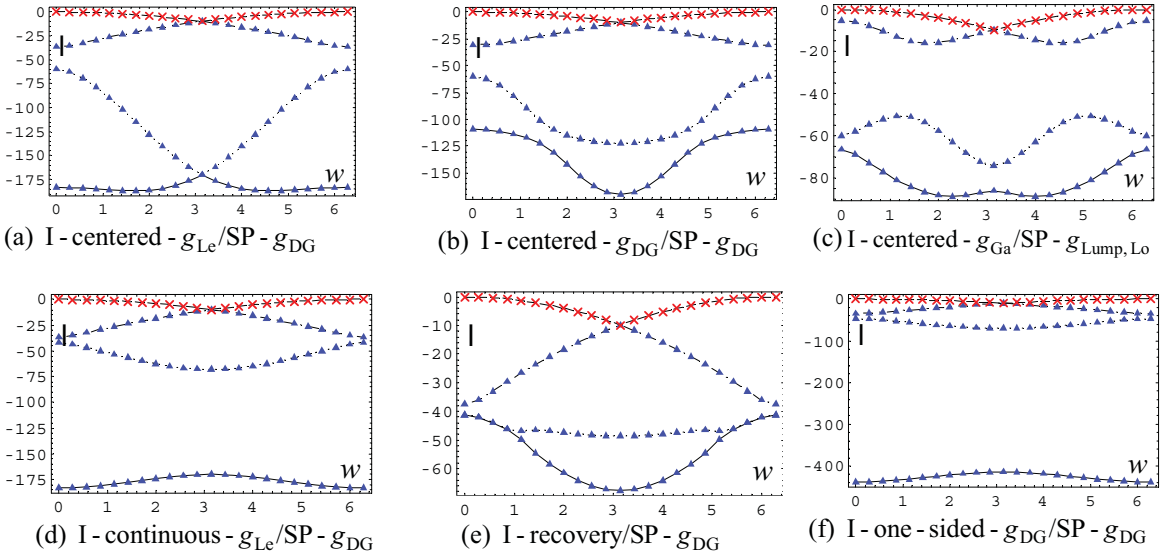


(f) I - one - sided - g_{DG}/SP - g_{DG}

Figure 6.2. *Eigenvalues as functions of wave numbers for six representative schemes with* $K = 3$.

Scheme for $K = 3$	Min. Eigval.	Ord. Acc.	Coarse mesh error, $w = \pi/8$	Fine mesh error, $w = \pi/16$	For I - g, $g'(-1)$
1. I - centered - g_{Le}/SP - g_{DG}	–75.	6	$-5.06\cdot 10^{-8}$	$-1.97\cdot 10^{-10}$	–6
2. I - centered - g_{DG}/SP - g_{DG}	–60.	4	$2.15\cdot 10^{-6}$	$3.40\cdot 10^{-8}$	–4.5
3. I - centered - g_{Ga}/SP - g_{DG}	–60.	4	$5.14\cdot 10^{-6}$	$7.97\cdot 10^{-8}$	–3.5
4. I - centered - $g_{\mathrm{Lump,\,Lo}}$/SP - g_{DG} *	–60.	4	$7.8\cdot 10^{-6}$	$1.2\cdot 10^{-7}$	–3
5. I - centered - g_{DG}/SP - g_{Ga}	–42.	4	$5.6\cdot 10^{-6}$	$8.72\cdot 10^{-8}$	–4.5
6. I - centered - g_{Ga}/SP - g_{Ga}	–36.	4	$8.62\cdot 10^{-6}$	$1.33\cdot 10^{-7}$	–3.5
7. I - centered - $g_{\mathrm{Lump,\,Lo}}$/SP - g_{Ga} *	–36.	4	$1.13\cdot 10^{-5}$	$1.73\cdot 10^{-7}$	–3
8. I - centered - g_{DG}/SP - $g_{\mathrm{Lump,\,Lo}}$	–42.	4	$9.93\cdot 10^{-6}$	$1.54\cdot 10^{-7}$	–4.5
9. I - centered - g_{Ga}/SP - $g_{\mathrm{Lump,\,Lo}}$	–30.	4	$1.3\cdot 10^{-5}$	$2.\cdot 10^{-7}$	–3.5
10. I - centered - $g_{\mathrm{Lump,\,Lo}}$/SP - $g_{\mathrm{Lump,\,Lo}}$ *	–26.	4	$1.57\cdot 10^{-5}$	$2.4\cdot 10^{-7}$	–3
11. I - continuous - g_{Le}/SP - g_{DG}	–74.	6	$1.87\cdot 10^{-9}$	$7.31\cdot 10^{-12}$	–6
12. I - continuous - g_{DG}/SP - g_{DG}	–42.	4	$2.18\cdot 10^{-6}$	$3.41\cdot 10^{-8}$	–4.5
13. I - continuous - g_{Le}/SP - g_{Ga}	–60.	4	$3.4\cdot 10^{-6}$	$5.31\cdot 10^{-8}$	–6
14. I - recovery/SP - g_{DG}	–33.	8	$4.75\cdot 10^{-11}$	$4.64\cdot 10^{-14}$	*
15. I - one - sided - g_{DG}/SP - g_{DG}	–148.	6	$4.5\cdot 10^{-9}$	$1.75\cdot 10^{-11}$	–4.5
16. I - one - sided - $g_{\mathrm{Lump,\,Lo}}$/SP - $g_{\mathrm{Lump,\,Lo}}$	–76.5	4	$1.55\cdot 10^{-5}$	$2.39\cdot 10^{-7}$	–3
17. I - centered - g_{DG} (4 - cell)/SP - g_{DG}	–65.	4	$6.84\cdot 10^{-8}$	$2.64\cdot 10^{-10}$	–4.5

Table 6.2. *Minimum eigenvalues, orders of accuracy, and errors for schemes with* $K = 3$.



(a) I - centered - g_{Le}/SP - g_{DG} (b) I - centered - g_{DG}/SP - g_{DG} (c) I - centered - g_{Ga}/SP - $g_{\mathrm{Lump,\,Lo}}$

(d) I - continuous - g_{Le}/SP - g_{DG} (e) I - recovery/SP - g_{DG} (f) I - one - sided - g_{DG}/SP - g_{DG}

Figure 6.3. *Eigenvalues as functions of wave numbers for six representative schemes with* $K = 4$.

Scheme for $K = 4$	Min. Eigval.	Ord. Acc.	Coarse mesh error, $w = \pi/8$	Fine mesh error, $w = \pi/16$	$g'(-1)$ (I - g)
1. I - centered - g_{Le}/SP - g_{DG}	–187.	6	$-5.31\cdot 10^{-9}$	$-2.16\cdot 10^{-11}$	–10
2. I - centered - g_{DG}/SP - g_{DG}	–170.	6	$-5.06\cdot 10^{-9}$	$-2.14\cdot 10^{-11}$	–8
3. I - centered - g_{Ga}/SP - g_{DG}	–170.	6	$-3.78\cdot 10^{-9}$	$-1.98\cdot 10^{-11}$	–6.5
4. I - centered - $g_{\mathrm{Lump,\,Lo}}$/SP - g_{DG} *	–170.	6	$4.5\cdot 10^{-9}$	$1.75\cdot 10^{-11}$	–6
5. I - centered - g_{DG}/SP - g_{Ga}	–122.	6	$-9.93\cdot 10^{-10}$	$-5.05\cdot 10^{-12}$	–8
6. I - centered - g_{Ga}/SP - g_{Ga}	–98.	6	below	$-3.89\cdot 10^{-12}$	–6.5
7. I - centered - $g_{\mathrm{Lump,\,Lo}}$/SP - g_{Ga} *	–98.	6	$6.64\cdot 10^{-9}$	$2.59\cdot 10^{-11}$	–6
8. I - centered - g_{DG}/SP - $g_{\mathrm{Lump,\,Lo}}$	–122.	6	$2.19\cdot 10^{-9}$	$7.63\cdot 10^{-12}$	–8
9. I - centered - g_{Ga}/SP - $g_{\mathrm{Lump,\,Lo}}$	–89.	6	$2.98\cdot 10^{-9}$	$8.55\cdot 10^{-12}$	–6.5
10. I - centered - $g_{\mathrm{Lump,\,Lo}}$/SP - $g_{\mathrm{Lump,\,Lo}}$ *	–83.	6	$8.43\cdot 10^{-9}$	$3.29\cdot 10^{-11}$	–6
11. I - continuous - g_{Le}/SP - g_{DG}	–183.	8	$8.57\cdot 10^{-13}$	$8.38\cdot 10^{-16}$	–10
12. I - continuous - g_{DG}/SP - g_{DG}	–122.	6	$2.24\cdot 10^{-9}$	$8.76\cdot 10^{-12}$	–8
13. I - continuous - g_{Le}/SP - g_{Ga}	–170.	6	$4.21\cdot 10^{-9}$	$1.64\cdot 10^{-11}$	–10
14. I - recovery/SP - g_{DG}	–68.	10	$-1.77\cdot 10^{-14}$	$-4.37\cdot 10^{-18}$	*
15. I - one - sided - g_{DG}/SP - g_{DG}	–439.	8	$1.96\cdot 10^{-12}$	$1.92\cdot 10^{-15}$	–8
16. I - one - sided - $g_{\mathrm{Lump,\,Lo}}$/SP - $g_{\mathrm{Lump,\,Lo}}$	–272.	6	$1.50\cdot 10^{-8}$	$5.85\cdot 10^{-11}$	–6
17. I - centered - g_{DG} (4 - cell)/SP - g_{DG}	–176.	6	$-1.39\cdot 10^{-9}$	$-5.46\cdot 10^{-12}$	–8

Table 6.3. *Minimum eigenvalues, orders of accuracy, and errors for schemes with* $K = 4$.

6.3. The case $K = 3$. The plots of eigenvalues as functions of wave numbers for six representative schemes are shown in Fig. 6.2.

The minimum eigenvalues, orders of accuracy, and errors of 17 schemes for $K = 3$ are shown in Table 6.2. Note that all schemes are of order at least 4 or $2(K-1)$. The three schemes 1, 11, and 15 are of order

6, and scheme 14 (recovery), order 8. Schemes 4, 7, and 10, which employ $g_{\text{Lump, Lo}}$ at interfaces, have a zero spurious eigenvalue at $w = \pi$ (not at $w = 0$ as in the case of $K = 2$); the corresponding eigenvector is not damped; and the matrix for steady state equation is singular. Schemes $\text{I - continuous - } g_{\text{Lump, Lo}}$ (not listed), regardless of the choice of $\text{SP - } g$, are inconsistent: they have an eigenvalue that is identically 0 for all w. The recovery scheme ($\text{I - recovery/SP - } g_{\text{DG}}$ or scheme 14) is more complicated than the others but, among the schemes of Table 6.2, it has the advantages of highest accuracy and third largest CFL limit (the amplitude of the minimum eigenvalue is 33).

6.4. The case $K = 4$. The plots of eigenvalues as functions of wave numbers for six representative schemes are shown in Fig. 6.3.

The minimum eigenvalues, orders of accuracy, and errors of 17 schemes for $K = 4$ are shown in Table 6.3. Again, the last column shows the value $g'(-1)$. Note that all schemes are of order at least 6 or $2(K-1)$. Schemes 11 and 15 are of order 8, and the scheme 14 (recovery), order 10. Schemes 4, 7, and 10, which employ $g_{\text{Lump, Lo}}$ at interfaces, have a zero spurious eigenvalue at $w = 0$; the corresponding eigenvector is not damped; and the matrix for steady state equation is singular. For scheme 6, or $\text{I - centered - } g_{\text{Ga}}/\text{SP - } g_{\text{Ga}}$, the error crosses zero near $w = \pi/8$. Therefore, to calculate the order of accuracy, we set $w = \pi/16$ for the coarse mesh and $w = \pi/32$ for the fine mesh. The errors are respectively $-3.89 \cdot 10^{-12}$ and $-1.98 \cdot 10^{-14}$, and the scheme is of order 6. Schemes $\text{I - continuous - } g_{\text{Lump, Lo}}$ (not listed), regardless of the choice of g for the solution points, are inconsistent: they have an eigenvalue that is identically 0 for all w. Note that for a certain range of w, two spurious eigenvalues of the recovery scheme become complex—a rather peculiar property. Since the real parts of these complex values are roughly –45, the corresponding eigenvectors are damped quickly. The property of possessing complex eigenvalues is unique for the recovery scheme and it holds true for all $K \geq 4$ (this author tested for K up to 10).

6.5. Arbitrary K. For a majority of schemes, the author tested with values K of up to 10. The following conclusions can be drawn.

All schemes are of order at least $2(K-1)$. In the order of increasing accuracy, scheme 1 or $\text{I - centered - } g_{\text{Le}}/\text{SP - } g_{\text{DG}}$ is of order $2K$ for odd K and $2(K-1)$ for even K. Schemes 11 and 15, i.e., $\text{I - continuous - } g_{\text{Le}}/\text{SP - } g_{\text{DG}}$ and $\text{I - one - sided - } g_{\text{DG}}/\text{SP - } g_{\text{DG}}$, respectively, are of order $2K$ for all K. Scheme 14 or $\text{I - recovery/SP - } g_{\text{DG}}$ is of order $3K-1$ for odd K and $3K-2$ for even K.

In general, $g_{\text{Lump, Lo}}$ is a limiting case in the sense that (a) a less steep correction function may lead to instability, and (b) for the steady state case, the scheme $\text{I - } g_{\text{Lump, Lo}}$ results in a singular solution matrix.

Based on the above results using Fourier analyses together with the fact that the super-convergence (super-accuracy) property does not hold for the general case of nonlinear systems, the following recommendations can be made. Scheme 2 (or $\text{I - centered - } g_{\text{DG}}/\text{SP - } g_{\text{DG}}$ or BR2) and 15 (or $\text{I - one - sided - } g_{\text{DG}}/\text{SP - } g_{\text{DG}}$ or CDG) appear to be good choices: they are simple to code and, concerning the stability limit, the BR2 method has an edge. Scheme 9 or $\text{I - centered - } g_{\text{Ga}}/\text{SP - } g_{\text{Lump, Lo}}$ is simple to code and, while not quite as accurate as the above two, has a rather large stability limit. If accuracy is a key concern, then at the cost of some additional coding, scheme 11 or $\text{I - continuous - } g_{\text{Le}}/\text{SP - } g_{\text{DG}}$ is of order $2K$. Among all methods, scheme 14 or $\text{I - recovery/SP - } g_{\text{DG}}$ is the most accurate and has the largest stability limit for $K \geq 4$, but it is also the most complicated since the recovery is across two cells (not one at a time as the other methods).

7. Two Dimensional Extensions

For the two-dimensional (2D) extension to a quadrilateral mesh of the current reconstruction approach, the solution procedure reduces to a series of one-dimensional (1D) operations in a manner similar to the extension of the staggered-grid scheme in (Kopriva 1998). The key difference is that the current extension is simpler since it involves only one grid instead of three.

With $u = u(x, y, t)$, consider the 2D diffusion equation

$$u_t = u_{xx} + u_{yy} \tag{7.1}$$

with initial condition

$$u(x, y, 0) = u_{\text{init}}(x, y) . \tag{7.2}$$

For the case of the Navier-Stokes equations, we have

$$\boldsymbol{u}_t + \nabla . \boldsymbol{F}(\boldsymbol{u}, \nabla \boldsymbol{u}) = 0$$

where $\boldsymbol{u}$ is the (column) vector of conservative variables and $\boldsymbol{F} = \boldsymbol{F}(\boldsymbol{u}, \nabla \boldsymbol{u})$ is the flux vector. Next, $\boldsymbol{F} = \boldsymbol{F}_a(\boldsymbol{u}) + \boldsymbol{F}_v(\boldsymbol{u}, \nabla \boldsymbol{u})$ where $\boldsymbol{F}_a$ and $\boldsymbol{F}_v$ are the advective and viscous flux vectors respectively. In addition, $\boldsymbol{F} = (\boldsymbol{f}, \boldsymbol{g})$ where $\boldsymbol{f}$ and $\boldsymbol{g}$ are the two Cartesian components; similarly, $\boldsymbol{F}_a = (\boldsymbol{f}_a, \boldsymbol{g}_a)$ and $\boldsymbol{F}_v = (\boldsymbol{f}_v, \boldsymbol{g}_v)$. The above equation can be written as

$$\boldsymbol{u}_t + \boldsymbol{f}_x + \boldsymbol{g}_y = 0 . \tag{7.3}$$

Clearly, (7.1) is a special case of (7.3): with $\boldsymbol{u} = u$ and $\boldsymbol{F}(\boldsymbol{u}, \nabla \boldsymbol{u}) = \nabla u$, i.e., $\boldsymbol{f} = u_x$ and $\boldsymbol{g} = u_y$, equation (7.3) takes the form

$$u_t = \nabla .(\nabla u) . \tag{7.4}$$

Let the domain of calculation be divided into an unstructured (or structured) mesh of quadrilaterals E_j where $j = 1, ..., J$. Dropping the subscript j, let the mapping from the biunit square $I \cdot I = I^2 = [-1,1]^2$ onto E be denoted by

$$\boldsymbol{x}(\xi, \eta) = (x(\xi, \eta), y(\xi, \eta)) . \tag{7.5}$$

This mapping can be isoparametric and the cell E can have curved edges. For simplicity, however, we assume that the mapping is bilinear: denoting the four corners of E by (counterclockwise) $\boldsymbol{x}_1 = \boldsymbol{x}(-1,-1)$, $\boldsymbol{x}_2 = \boldsymbol{x}(1,-1)$, $\boldsymbol{x}_3 = \boldsymbol{x}(1,1)$, and $\boldsymbol{x}_4 = \boldsymbol{x}(-1,1)$, then

$$\boldsymbol{x}(\xi, \eta) = (1/4) \{ (1-\xi)(1-\eta)\boldsymbol{x}_1 + (1+\xi)(1-\eta)\boldsymbol{x}_2 + (1+\xi)(1+\eta)\boldsymbol{x}_3 + (1-\xi)(1+\eta)\boldsymbol{x}_4 \} . \tag{7.6}$$

See Fig. 7.1. Under this mapping, (7.3) is transformed to

$$\widetilde{\boldsymbol{u}}_t + \frac{\partial \widetilde{\boldsymbol{f}}}{\partial \xi} + \frac{\partial \widetilde{\boldsymbol{g}}}{\partial \eta} = 0 \tag{7.7}$$

where

$$J = x_\xi y_\eta - x_\eta y_\xi , \quad \widetilde{\boldsymbol{u}} = J\boldsymbol{u} , \quad \widetilde{\boldsymbol{f}} = y_\eta \boldsymbol{f} - x_\eta \boldsymbol{g} , \quad \text{and} \quad \widetilde{\boldsymbol{g}} = -y_\xi \boldsymbol{f} + x_\xi \boldsymbol{g} . \tag{7.8}$$

Here, J represents the volume element. As for $\widetilde{\boldsymbol{f}}$ (and similarly for $\widetilde{\boldsymbol{g}}$), along a vertical edge, say, $\boldsymbol{x}_1 \boldsymbol{x}_4$, the covariant base vector is (x_η, y_η). The vector $(y_\eta, -x_\eta)$ represents the product of the unit normal with the edge length of $\boldsymbol{x}_1 \boldsymbol{x}_4$; therefore, $\widetilde{\boldsymbol{f}} = y_\eta \boldsymbol{f} - x_\eta \boldsymbol{g}$ represents the normal flux across this edge (with the edge length accounted for). That is, the above differential formulation can be considered to be equivalent to the integral formulation.

Similar to the discussion before (7.3), the flux vector $\widetilde{\boldsymbol{f}}$ consists of two parts: $\widetilde{\boldsymbol{f}}_a$ and $\widetilde{\boldsymbol{f}}_v$. The advective part $\widetilde{\boldsymbol{f}}_a$ is dealt with as in (Huynh 2007). As for the diffusive part, once we obtain the solution for the diffusion equation (7.4), with appropriate and trivial modifications, the solution for (7.3) follows. Therefore, we focus only on the scalar equation (7.4) with

$$\boldsymbol{f} = u_x , \quad \boldsymbol{g} = u_y , \tag{7.9}$$

and

$$\widetilde{\boldsymbol{f}} = y_\eta \boldsymbol{f} - x_\eta \boldsymbol{g} = y_\eta u_x - x_\eta u_y , \quad \text{and} \quad \widetilde{\boldsymbol{g}} = -y_\xi \boldsymbol{f} + x_\xi \boldsymbol{g} = -y_\xi u_x + x_\xi u_y . \tag{7.10}$$

Let the solution points on $I = [-1,1]$ be ξ_k, $k = 1, ..., K$. For simplicity (and not necessity), the solution points in the η direction is assumed to be the same as those in the ξ direction: $\eta_l = \xi_l$, $l = 1, ..., K$. The $K \cdot K$ solution points on the biunit square are (ξ_k, η_l), $k, l = 1, ..., K$. The solution points on E are

$$\boldsymbol{x}_{k,l} = (x_{k,l}, y_{k,l}) = \boldsymbol{x}(\xi_k, \eta_l) . \tag{7.11}$$

In addition, let the solution points along the edges, which are the small squares in Fig. 7.1, be called flux points. For the methods discussed here, *flux points always lie on cell edges*. On each cell E, let the

solution u be approximated by $K \cdot K$ pieces of data $u_{k,l}$ at the solution points $\boldsymbol{x}_{k,l}$, $k, l = 1,...,K$. Again, boundary conditions are assumed to be periodic and are omitted.

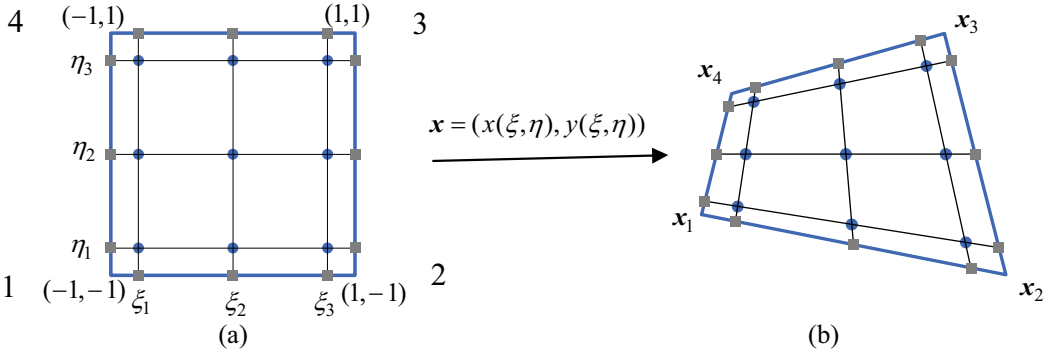


Figure 7.1. (a) Solution points (circular dots), which are Gauss points here, and flux points (squares along the edges) for $K = 3$ on the biunit square I^2. At each flux point, we need the common value and common gradient. (b) The corresponding quantities for cell E in the physical domain.

At time level n, suppose the data $u_{k,l}$ are known for all cells and all *k,* and *l*. For each cell, we wish to calculate $du_{k,l}(t)/dt$ at time $t = t^n$ (denoted by $du_{k,l}/dt$) in terms of the data. That is, we wish to calculate u_{xx} and u_{yy} at the solution points. Then, we march in time by, say, a Runge-Kutta method.

The first task is to approximate the solution by a polynomial. To this end, let ϕ_k denote the 1D basis function on I given in (2.6). Let the 2D basis functions on I^2 be defined via tensor product: for each (k,l), $k, l = 1,...,K$,

$$\phi_{k,l} = \phi_{k,l}(\xi,\eta) = \phi_k(\xi)\phi_l(\eta). \tag{7.12}$$

Thus, $\phi_{k,l}$ takes on the value 1 at (ξ_k, η_l) and 0 at all other $K^2 - 1$ solution points as shown in Fig. 7.2. Using the local coordinates (ξ,η) for the cell E, the polynomial interpolating $u_{k,l}$ at (ξ_k, η_l) is given by

$$u_E(\xi,\eta) = \sum_{k,l=1}^{K} u_{k,l}\,\phi_{k,l}(\xi,\eta) = \sum_{k,l=1}^{K} u_{k,l}\,\phi_k(\xi)\phi_l(\eta). \tag{7.13}$$

Note that for each fixed η, $u_E(\xi,\eta)$ is a polynomial of degree $K-1$ in ξ, and vice versa. That is, u_E is in $\boldsymbol{P}_{K-1,K-1} = \boldsymbol{P}_{K-1} \otimes \boldsymbol{P}_{K-1}$; e.g., for $K = 2$, u_E is bilinear; for $K = 4$, bicubic.

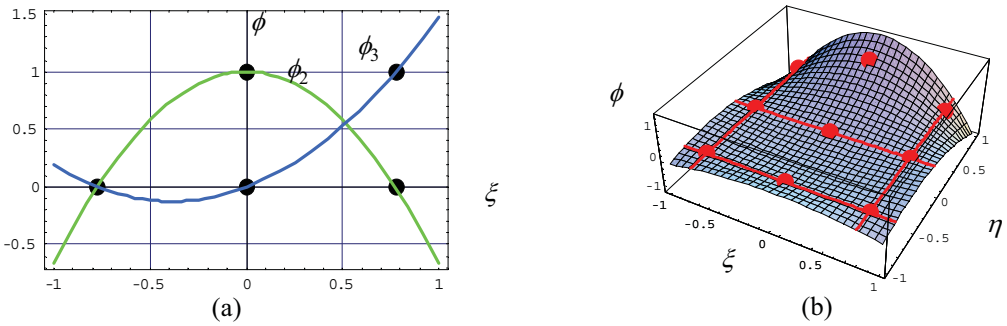


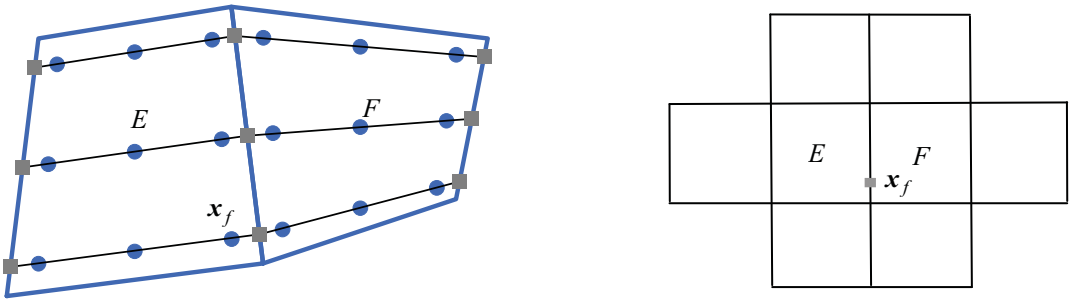
Figure 7.2. Basis functions for $K = 3$ with Gauss points as solution points. (a) The functions ϕ_2 and ϕ_3 for the 1D case. (b) The function $\phi_{2,3} = \phi_2(\xi)\phi_3(\eta)$ for the 2D case.

For each fixed η_l (Fig. 7.1(a)), the interpolation along the ξ-direction reduces to a 1D interpolation of the data $u_{k,l}$ where k varies. (A similar statement holds for each fixed ξ_k.) On E, we can evaluate

$(u_E)_\xi$ and $(u_E)_\eta$ (or the uncorrected gradient) at the solution points by the 1D procedure. We can also calculate the values of u_E at the flux points on the cell edge of E in a 1D manner.

Common interface values and corrected $\tilde{f}$ and $\tilde{g}$ at the solution points. On an arbitrary edge, say, in the local description, a vertical edge (similarly for a horizontal edge), let E and F denote the two adjacent cells as shown in Fig. 7.3(a). To define the common value and common derivative, let $\boldsymbol{x}_f$ be a flux point on this edge, e.g., $\boldsymbol{x}_f = \boldsymbol{x}_E(1, \eta_l) = \boldsymbol{x}_F(-1, \eta_l)$. At this point, with $u_E(\boldsymbol{x}_f) = u_E(1, \eta_l)$ and $u_F(\boldsymbol{x}_f) = u_F(-1, \eta_l)$, let the centered common value be defined by

$$u^{\text{com}} = u^{\text{com}}(\boldsymbol{x}_f) = (1/2)[u_E(\boldsymbol{x}_f) + u_F(\boldsymbol{x}_f)]. \tag{7.14}$$



(a) (b)

Figure 7.3. (a) Employing $u_E^{\text{RB}}(\xi)$ (corrected for the right boundary) and $u_F^{\text{LB}}(\xi)$ (corrected for the left boundary) to calculate the common gradient at $\boldsymbol{x}_f$ results in a formula of two-cell stencil. (b) Employing u_E^C and u_F^C to calculate the common gradient results in a formula of eight-cell stencil.

The one-sided common value can be defined in a manner similar to (3.1) provided that a selection of left and right sides is assigned to the interface. Next, on a general cell, say, E, with the common values at all flux points defined as above, we can obtain the corrected ξ-derivative as in the 1D case of (3.10): for each fixed η_l,

$$\begin{aligned}(u_E^C)_\xi(\xi, \eta_l) \;=\; & (u_E)_\xi(\xi, \eta_l) + [u^{\text{com}}(-1, \eta_l) - u_E(-1, \eta_l)](g^{\text{LB}})'(\xi) \\ & + [u^{\text{com}}(1, \eta_l) - u_E(1, \eta_l)](g^{\text{RB}})'(\xi)\end{aligned}. \tag{7.15}$$

The corrected η-derivative is calculated in the same manner. At the solution points, the values of the corrected ξ- and η-derivatives, namely $u_\xi = (u_E^C)_\xi(\xi_k, \eta_l)$ and $u_\eta = (u_E^C)_\eta(\xi_k, \eta_l)$ can easily be evaluated. The quantity ∇u follows by the chain rule:

$$u_x = u_\xi \xi_x + u_\eta \eta_x \quad \text{and} \quad u_y = u_\xi \xi_y + u_\eta \eta_y. \tag{7.16}$$

Then, the corrected $\tilde{f} = y_\eta u_x - x_\eta u_y$ and $\tilde{g} = -y_\xi u_x + x_\xi u_y$ can be obtained at all solution points of E.

Common gradients. We now define the common gradient or, more precisely, in the local description, $\tilde{f}^{\text{com}}$ on the vertical edges and $\tilde{g}^{\text{com}}$ on the horizontal ones. Again, on an arbitrary edge, let E and F denote the two adjacent cells as shown in Fig. 7.3(a). At a flux point $\boldsymbol{x}_f$ on this edge, similar to the 1D case, if we employ u_E^C and u_F^C to calculate the common gradient, the resulting formula generally has a large stencil involving eight cells as shown in Fig. 7.3(b).

A formula for $\tilde{f}^{\text{com}}$ that has a compact stencil involving only the two adjacent cells E and F can be obtained as follows. Without loss of generality, assume again the edge is vertical, i.e., for some fixed η_l, $\boldsymbol{x}_f = \boldsymbol{x}_E(1, \eta_l) = \boldsymbol{x}_F(-1, \eta_l)$ (see Fig. 7.3(a)). We can define $u_E^{\text{RB}}(\xi)$ (corrected for the right boundary of

E) and $u_F^{\mathrm{LB}}(\xi)$ (corrected for the left boundary of F) in a manner similar to (3.7) and (3.8). For the cell E, the corrected ξ-derivative at $\boldsymbol{x}_f$ is

$$u_{\xi,E} = (u_E)_\xi^{\mathrm{RB}}(1) \;=\; (u_E)_\xi(1) + [u^{\mathrm{com}} - u_E(1)]\,[(g^{\mathrm{RB}})'(1)] \tag{7.17a}$$

and, for the cell F,

$$u_{\xi,F} = (u_F)_\xi^{\mathrm{LB}}(-1) \;=\; (u_F)_\xi(-1) + [u^{\mathrm{com}} - u_F(-1)]\,[(g^{\mathrm{LB}})'(-1)]\,. \tag{7.17b}$$

We now discuss the evaluation of u_η at $\boldsymbol{x}_f$. Since the common values u^{com} at the flux points along the common edge are readily available, they can be employed to evaluate u_η. Alternatively, we can calculate u_η with no correction by first obtaining the values of u_E at the flux points along the common edge (and then u_F); however, this procedure is slightly costlier. Note that for the case of a uniform Cartesian mesh (used in Fourier analysis), u_η at $\boldsymbol{x}_f$ is not needed.

For cell E, with $u_\xi = (u_E)_\xi^{\mathrm{RB}}(1)$ via (7.17a) and u_η discussed above, we can evaluate $(\nabla u)_E$ at $\boldsymbol{x}_f$ by the chain rule. The normal flux along the common edge namely $(\widetilde{\boldsymbol{f}})_E$ can then be estimated by the formula $\widetilde{\boldsymbol{f}} = y_\eta u_x - x_\eta u_y$. Similarly, in the case of cell F, with $u_\xi = (u_F)_\xi^{\mathrm{LB}}(-1)$ via (7.17b) and u_η discussed above, we can evaluate $(\nabla u)_F$ at $\boldsymbol{x}_f$ by the chain rule, and then evaluate $(\widetilde{\boldsymbol{f}})_F$. To define $\widetilde{\boldsymbol{f}}^{\mathrm{com}}$, we could employ the weighted average (4.3) provided a selection of left and right sides is assigned to the interface. For simplicity, the centered formula is shown here:

$$\widetilde{\boldsymbol{f}}^{\mathrm{com}} = (1/2)[(\widetilde{\boldsymbol{f}})_E + (\widetilde{\boldsymbol{f}})_F]\,. \tag{7.18}$$

The above 2D extension can easily be applied to the schemes of type I - continuous. More precisely, at a flux point $\boldsymbol{x}_f$, the quantity u^{com} is considered to be an unknown, and it is calculated by requiring that $(\widetilde{\boldsymbol{f}})_E = (\widetilde{\boldsymbol{f}})_F$. Here, assuming that the common edge is of vertical type for cell E, the calculation of u_η with no correction by first obtaining the values of u_E at the flux points simplifies the scheme (see the discussion after (7.17)).

The 2D extension for scheme I - recovery or (4.8) to the case of a uniform mesh of squares is simple; however, that for a general quadrilateral mesh is considerably more involved.

Algorithm. The algorithm for the 2D diffusion equation is summarized below. Let the data $u_{k,l}$ at the solution points $\boldsymbol{x}_{k,l}$ be given for all cells.

- On each edge, compute u^{com} and $\widetilde{\boldsymbol{f}}^{\mathrm{com}}$ at the flux points as follows. Assume that edge e is of vertical type (horizontal type is similar). Let E and F be the two cells sharing this edge. First, obtain the values of u_E and u_F at the flux points on e. Next, at each flux point $\boldsymbol{x}_f$ of e, evaluate and store u^{com} by the centered formula (7.14) or the one-sided formula (3.1). Using this u^{com}, calculate the corrected (at one boundary only) ξ-derivative $u_{\xi,E}(\boldsymbol{x}_f)$ and $u_{\xi,F}(\boldsymbol{x}_f)$ by (7.17a) and (7.17b). Compute $u_\eta(\boldsymbol{x}_f)$ using the common values u^{com} at the flux points along e. Using $u_{\xi,E}(\boldsymbol{x}_f)$ and $u_\eta(\boldsymbol{x}_f)$, evaluate $(\nabla u)_E$ at $\boldsymbol{x}_f$ by the chain rule. Again at $\boldsymbol{x}_f$, estimate $(\widetilde{\boldsymbol{f}})_E$ via $\widetilde{\boldsymbol{f}} = y_\eta u_x - x_\eta u_y$. Evaluate $(\nabla u)_F$ and $(\widetilde{\boldsymbol{f}})_F$ in a similar manner. Calculate and store $\widetilde{\boldsymbol{f}}^{\mathrm{com}}$ by (7.18). Alternatively, u^{com} and $\widetilde{\boldsymbol{f}}^{\mathrm{com}}$ can be obtained via the continuity requirement (4.7). Store the values u^{com} and $\widetilde{\boldsymbol{f}}^{\mathrm{com}}$ at the flux points of all edges.
- For each cell, calculate u_t at the solution points as follows. Using u^{com} on the four edges of the cell, compute $(u_E^C)_\xi$ and $(u_E^C)_\eta$ at the solution points via (7.15). Again at the solution points, calculate ∇u by the chain rule, and then $\widetilde{\boldsymbol{f}} = y_\eta u_x - x_\eta u_y$ and, similarly, $\widetilde{\boldsymbol{g}} = -y_\xi u_x + x_\xi u_y$.

Next, using $\widetilde{\boldsymbol{f}}^{\,\mathrm{com}}$ at the flux points on the two edges of vertical type, obtain the corrected derivative $(\widetilde{\boldsymbol{f}})_{\xi}$ at the solution points. Similarly, using $\widetilde{\boldsymbol{g}}^{\,\mathrm{com}}$ at the flux points on the two edges of horizontal type, obtain the corrected derivative $(\widetilde{\boldsymbol{g}})_{\eta}$. Finally, calculate $\widetilde{\boldsymbol{u}}_t$ at the solution points via (7.7). This completes the algorithm.

Fourier analysis in 2D. We now carry out the Fourier (Von Neumann) accuracy and stability analysis for the 2D case. On the domain $(-\infty,\infty)\cdot(-\infty,\infty)$, consider the diffusion equation

$$u_t = u_{xx} + u_{yy}. \tag{7.19}$$

The cells are the squares $E_{i,j} = [i-1/2, i+1/2]\cdot[j-1/2, j+1/2]$ centered at (i,j). Denote by $\boldsymbol{u}_{i,j}$ the column vector of K^2 components

$$\boldsymbol{u}_{i,j} = (u_{i,j,1,1}, u_{i,j,1,2}, ..., u_{i,j,1,K}, u_{i,j,2,1}, ..., u_{i,j,2,K}, ..., u_{i,j,K,1}, ..., u_{i,j,K,K}). \tag{7.20}$$

Note that for the cell (i,j), at a fixed (k_0,l_0), the evaluation of u_{xx} is exactly the same as the 1D case using the data (k,l_0), where k varies in the cell (i,j), $(i-1,j)$, and $(i+1,j)$. A similar statement holds for u_{yy}. Thus, the 2D solution follows immediately from the 1D solution, and the result can be written as

$$\frac{d\boldsymbol{u}_{i,j}}{dt} = \boldsymbol{C}_{0,0}\,\boldsymbol{u}_{i,j} + \boldsymbol{C}_{-1,0}\,\boldsymbol{u}_{i-1,j} + \boldsymbol{C}_{1,0}\,\boldsymbol{u}_{i+1,j} + \boldsymbol{C}_{0,-1}\,\boldsymbol{u}_{i,j-1} + \boldsymbol{C}_{0,1}\,\boldsymbol{u}_{i,j+1}. \tag{7.21}$$

Here, $\boldsymbol{C}_{0,0}$, $\boldsymbol{C}_{-1,0}$, $\boldsymbol{C}_{1,0}$, $\boldsymbol{C}_{0,-1}$, and $\boldsymbol{C}_{0,1}$ are $K^2\cdot K^2$ matrices. Let the wave number in the x-direction be denoted by w_x, and that in the y-direction, w_y (involving no partial derivative here). Let the imaginary number be denoted by $\boldsymbol{I}$ instead of the standard notation i to avoid confusion with the index i. Replacing $\boldsymbol{u}_{i-1,j}$, $\boldsymbol{u}_{i+1,j}$, $\boldsymbol{u}_{i,j-1}$, and $\boldsymbol{u}_{i,j+1}$ respectively by $e^{-\boldsymbol{I}w_x}\boldsymbol{u}_{i,j}$, $e^{\boldsymbol{I}w_x}\boldsymbol{u}_{i,j}$, $e^{-\boldsymbol{I}w_y}\boldsymbol{u}_{i,j}$, and $e^{\boldsymbol{I}w_y}\boldsymbol{u}_{i,j}$, we obtain

$$\frac{d\boldsymbol{u}_{i,j}}{dt} = \left(\boldsymbol{C}_{0,0} + e^{-\boldsymbol{I}w_x}\boldsymbol{C}_{-1,0} + e^{\boldsymbol{I}w_x}\boldsymbol{C}_{1,0} + e^{-\boldsymbol{I}w_y}\boldsymbol{C}_{0,-1} + e^{\boldsymbol{I}w_y}\boldsymbol{C}_{0,1}\right)\boldsymbol{u}_{i,j}. \tag{7.22}$$

Set

$$\boldsymbol{S} = \boldsymbol{C}_{0,0} + e^{-\boldsymbol{I}w_x}\boldsymbol{C}_{-1,0} + e^{\boldsymbol{I}w_x}\boldsymbol{C}_{1,0} + e^{-\boldsymbol{I}w_y}\boldsymbol{C}_{0,-1} + e^{\boldsymbol{I}w_y}\boldsymbol{C}_{0,1}. \tag{7.23}$$

Then $\boldsymbol{S}$ is a $K^2\cdot K^2$ matrix. The semidiscretization results in

$$\frac{d\boldsymbol{u}_{i,j}}{dt} = \boldsymbol{S}\,\boldsymbol{u}_{i,j} \tag{7.24}$$

For all the schemes discussed here, $\boldsymbol{S}$ has K^2 eigenvalues. The eigenvalue that approximates the exact eigenvalue—there is one and only one such value—is called the *principal eigenvalue* and denoted by $S(w_x,w_y)$:

$$S(w_x,w_y) \approx -w_x^{\,2} - w_y^{\,2}. \tag{7.25}$$

All others are *spurious* eigenvalues. All eigenvalues must lie in the left half of the complex plane for the semidiscretization to be stable.

It turns out that for all methods discussed, via Fourier analysis, the 2D and corresponding 1D schemes have identical order of accuracy. Concerning stability limits, the 2D method has a limit of 1/2 that of the 1D version for nearly all schemes and, for a few exceptions, approximately 1/2. In other words, the largest magnitude of all eigenvalues for a 2D scheme is essentially twice that of the corresponding 1D method. This reduction by a factor of 2 in the stability limit for the 2D diffusion equation here is consistent with the reduction by a factor of $\sqrt{2}$ for the 2D advection equation in (Huynh2007): each derivative corresponds to a factor of $\sqrt{2}$.

The minimum eigenvalues, orders of accuracy, and errors of 16 schemes for the 2D case with $K=4$ are shown in Table 7.1. In the calculations of order of accuracy, the coarse mesh corresponds to $w_x=\pi/8$ and $w_y=\pi/10$, and the fine mesh, $w_x=\pi/16$ and $w_y=\pi/20$. Note that all schemes are of order at least 6 or $2(K-1)$, and the orders of accuracy are identical to those of the corresponding 1D schemes in Table

6.3. Again, schemes 11 and 15 are of order 8, and scheme 14 (recovery), order 10. For scheme 6, or I-centered-g_{Ga}/SP-g_{Ga}, the error crosses zero near $w_x = \pi/8$ and $w_y = \pi/10$. Therefore, to calculate the order of accuracy, we set $w_x = \pi/16$ and $w_y = \pi/20$ for the coarse mesh and $w_x = \pi/32$ and $w_y = \pi/40$ for the fine mesh. The errors are respectively $-4.64 \cdot 10^{-12}$ and $-2.32 \cdot 10^{-14}$, and the scheme is of order 6. The recovery scheme (I-recovery/SP-g_{DG} or 14) is more complicated than the others, but it has the advantages of highest accuracy and largest stability limit.

Schemes in 2D $K = 4$	Min. Eigval.	Ord. Acc.	Coarse mesh error, $w_x = \pi/8$ $w_y = \pi/10$	Fine mesh error, $w_x = \pi/16$ $w_y = \pi/20$
1. I-centered-g_{Le}/SP-g_{DG}	–374.	6	$-6.22 \cdot 10^{-9}$	$-2.53 \cdot 10^{-11}$
2. I-centered-g_{DG}/SP-g_{DG}	–340.	6	$-5.94 \cdot 10^{-9}$	$-2.50 \cdot 10^{-11}$
3. I-centered-g_{Ga}/SP-g_{DG}	–340.	6	$-4.51 \cdot 10^{-9}$	$-2.33 \cdot 10^{-11}$
4. I-centered-$g_{Lump,\,Lo}$/SP-g_{DG} *	–340.	6	$5.25 \cdot 10^{-9}$	$2.05 \cdot 10^{-11}$
5. I-centered-g_{DG}/SP-g_{Ga}	–244.	6	$-1.18 \cdot 10^{-9}$	$-5.93 \cdot 10^{-12}$
6. I-centered-g_{Ga}/SP-g_{Ga}	–196.	6	see explaination	$-4.64 \cdot 10^{-12}$
7. I-centered-$g_{Lump,\,Lo}$/SP-g_{Ga} *	–196.	6	$7.76 \cdot 10^{-9}$	$3.03 \cdot 10^{-11}$
8. I-centered-g_{DG}/SP-$g_{Lump,\,Lo}$	–244.	6	$2.54 \cdot 10^{-9}$	$8.89 \cdot 10^{-12}$
9. I-centered-g_{Ga}/SP-$g_{Lump,\,Lo}$	–177.	6	$3.42 \cdot 10^{-9}$	$9.91 \cdot 10^{-12}$
10. I-centered-$g_{Lump,\,Lo}$/SP-$g_{Lump,\,Lo}$ *	–165.	6	$9.85 \cdot 10^{-9}$	$3.84 \cdot 10^{-11}$
11. I-continuous-g_{Le}/SP-g_{DG}	–367.	8	$9.50 \cdot 10^{-13}$	$9.28 \cdot 10^{-16}$
12. I-continuous-g_{DG}/SP-g_{DG}	–244.	6	$2.62 \cdot 10^{-9}$	$1.02 \cdot 10^{-11}$
13. I-continuous-g_{Le}/SP-g_{Ga}	–340.	6	$4.91 \cdot 10^{-9}$	$1.92 \cdot 10^{-11}$
14. I-recovery/SP-g_{DG}	–135.	10	$-1.89 \cdot 10^{-14}$	$-4.67 \cdot 10^{-18}$
15. I-one-sided-g_{DG}/SP-g_{DG}	–878.	8	$2.17 \cdot 10^{-12}$	$2.12 \cdot 10^{-15}$
16. I-one-sided-$g_{Lump,\,Lo}$/SP-$g_{Lump,\,Lo}$	–544.	6	$1.75 \cdot 10^{-8}$	$6.83 \cdot 10^{-11}$

Table 7.1. *Minimum eigenvalues, orders of accuracy, and errors for schemes in 2D with* $K = 4$.

8. Numerical Results

As preliminary tests, we apply the schemes discussed above to solve the steady diffusion equations (or two-point boundary value problem). Here, the problem is solved not by time-marching, but by inversion of the implied coefficient matrix of $u_{j,k}$, $j = 1,...,J$ and $k = 1,...,K$. Unless otherwise stated, the solution points are the Gauss points and, in the graphs, the crosses represent the exact solutions, while the dots, the numerical ones.

The first test is the Poisson equation

$$u_{xx} = 2 \tag{8.1}$$

on the domain $[-1,1]$ with boundary conditions

$$u(-1) = u(1) = 1. \tag{8.2}$$

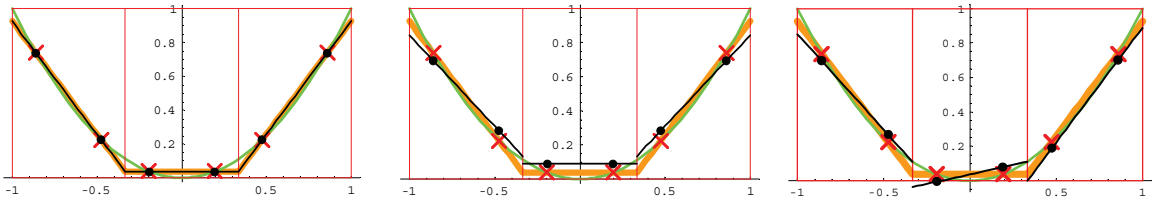
The exact solution is $u_{\text{exact}}(x) = x^2$. We employ piecewise linear methods to solve this problem. On each cell E_j of center x_j and length h_j, the two (Gauss) solution points are denoted by $x_{j,1}$ and $x_{j,2}$. The projection of x^2 onto the space of linear functions on E_j is

$$l_j(x) = x_j^{\,2} + h_j^{\,2}/12 + 2x_j(x - x_j).$$

That is, l_j provides the best linear approximation to x^2 on E_j in the least-squares sense. Note that l_j is identical to the line defined by the values of x^2 at the two Gauss points as shown in Fig. 8.1. (In general, if p is of degree K, then the projection of p onto $P^{K-1}(E_j)$ is determined by the values of p at the K Gauss points.) It appears desirable that the linear solution recovers l_j: at the Gauss points,

$$u_{j,1} = (x_{j,1})^2 \quad \text{and} \quad u_{j,2} = (x_{j,2})^2.$$

However, as shown in Fig. 8.1, only two schemes yield this solution: I - continuous - g_{Le}/SP - g_{DG} and I - recovery/SP - g_{DG}. All other schemes, e.g., BR2 and CDG do not. Concerning Fig. 8.1(c) for the CDG scheme, note that for each cell except the last one, the value of the solution polynomial at the right interface—thus, the common value—is exact. That is, at $x = -1$, i.e., the left boundary, $u^{\text{com}}_{1/2} = u^{-}_{1/2} = 1$; at each interior interface, $u^{\text{com}}_{j+1/2} = u^{-}_{j+1/2} = u_j(1)$. At the right boundary, however, the choice is switched: $u^{\text{com}}_{3+1/2} = u^{+}_{3+1/2} = 1$. This switch causes a loss in accuracy: loosely put, the cancellation of errors no longer holds. Also note that for all solutions of Fig. 8.1, the common interface values $u^{\text{com}}_{j+1/2}$ are exact. Concerning the cell average values, however, the BR2 and CDG schemes do not recover the exact solutions.



(a) I - continuous - g_{Le}/SP - g_{DG} and I - recovery/SP - g_{DG}

(b) Scheme BR2, i.e., I - centered - g_{DG}/SP - g_{DG}

(c) Scheme CDG or I - one sided - g_{DG}/SP - g_{DG}

Figure 8.1. Solutions for $u_{xx} = 2$ on $[-1,1]$ with boundary conditions $u(-1) = u(1) = 1$ using three cells. Only the schemes I - continuous - g_{Le}/SP - g_{DG} *and* I - recovery/SP - g_{DG} *recover the projection of the exact solution $u_{\text{exact}}(x) = x^2$ onto the space of piecewise linear functions (the lines through the crosses).*

The second test is the Poisson equation employed in (Van Leer and Nomura 2005)

$$u_{xx} = s(x) = -4\pi^2 \sin(2\pi x) \tag{8.3}$$

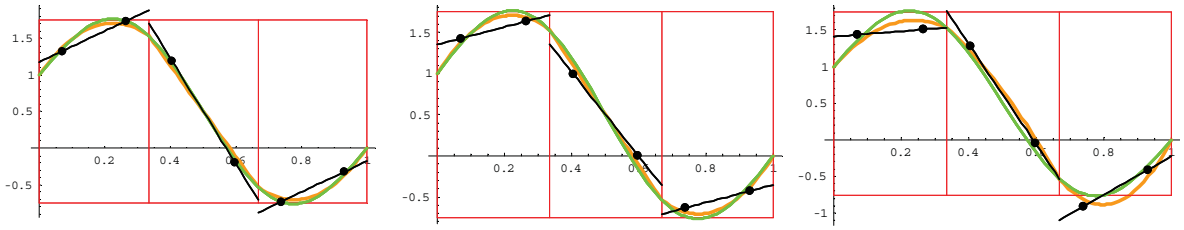
on the domain $[0,1]$ with boundary conditions

$$u_x(0) = 2\pi - 1 \quad \text{and} \quad u(1) = 0. \tag{8.4}$$

The exact solution is $u_{\text{exact}}(x) = \sin(2\pi x) + 1 - x$. The projection of $s(x)$ onto $\boldsymbol{P}^{K-1}(E_j)$ is denoted by $s_j(x)$. At the solution points, the source terms are set as, for $k = 1,...,K$, $s_{j,k} = s_j(x_{j,k})$.

Figure 8.2 shows the solutions with $K = 2$ using three cells: in the order of decreasing accuracy, (a) I - continuous - g_{Le}/SP - g_{DG}, (b) I - centered - g_{DG}/SP - g_{DG} or BR2, and (c) I - one - sided - g_{DG}/SP - g_{DG} or CDG. The solution by I - recovery/SP - g_{DG} is nearly identical to that by I - continuous - g_{Le}/SP - g_{DG} and is omitted. The smooth green curves are the exact solutions and the dots, the approximate solutions.

The brown curves are the reconstruction polynomials $\{u_j^C\}$; for all solutions, $\{u_j^C\}$ are more accurate than the piecewise linear solutions. Concerning Fig. 8.2(c) or the CDG solution, at each interior interface, $u_{j+1/2}^{\text{com}} = u_{j+1/2}^{-} = u_j(1)$, but at the right boundary, the choice is switched: $u_{3+1/2}^{\text{com}} = u_{3+1/2}^{+} = 0$. Note that this solution is asymmetric. Also note that $u_{j+1/2}^{\text{com}}$ for all solutions are again exact.



(a) I - continuous - g_{Le}/SP - g_{DG} (b) Scheme BR2 (c) Scheme CDG

Figure 8.2. Solutions for $u_{xx} = -4\pi^2 \sin(2\pi x)$ *on* $[0,1]$ *with boundary conditions* $u_x(-1) = 2\pi - 1$ and $u(1) = 0$ *using three cells.*

Scheme ($K = 4$)	Order of accuracy	Max. Error; 4 cells	Error ratio	Max. Error; 8 cells	Error ratio	Max. Error; 16 cells
I - recovery/SP - g_{DG} *		$.215 \cdot 10^{-5}$		$.712 \cdot 10^{-8}$		$.280 \cdot 10^{-10}$
	8 (or $2K$)		302.		236.	
I - recovery/SP - g_{DG}		$.769 \cdot 10^{-4}$		$.302 \cdot 10^{-5}$		$.100 \cdot 10^{-6}$
	5 (or $K+1$)		26.		30.	
I - continuous - g_{Le}/SP - g_{DG}		$.433 \cdot 10^{-4}$		$.170 \cdot 10^{-5}$		$.562 \cdot 10^{-7}$
	5 (or $K+1$)		25.		30.	
I - centered - g_{DG}/SP - g_{DG}		$.115 \cdot 10^{-2}$		$.106 \cdot 10^{-3}$		$.805 \cdot 10^{-5}$
	4 (or K)		11.		13.	
I - one sided - g_{DG}/SP - g_{DG}		$.115 \cdot 10^{-2}$		$.877 \cdot 10^{-4}$		$.574 \cdot 10^{-5}$
	4 (or K)		13.		15.	
I - centered - g_{Gauss}/SP - $g_{\text{Lump, Lo}}$		$.179 \cdot 10^{-2}$		$.151 \cdot 10^{-3}$		$.912 \cdot 10^{-5}$
	4 (or K)		12.		16.	

Table 8.1. Maximum errors and corresponding orders of accuracy. For the line I - recovery/SP - g_{DG} **, the exact values at the solution points in the error calculations are the values of the projection of the exact solution onto* $\boldsymbol{P}^{K-1}(E_j)$*; for all other cases, the exact values (with no projection) are employed.*

Table 8.1 shows the maximum errors at the solution points and orders of accuracy of various schemes with $K = 4$ for (boundary value) problem (8.3)-(8.4). As for the recovery scheme, we consider two types of errors: in the first type, the exact values at the solution points are the values of the projection of the exact solution onto $\boldsymbol{P}^{K-1}(E_j)$; in the second, the standard exact values (with no projection) are employed. Note that with the first type of errors, the recovery scheme is of order $2K$; with the second, order $K+1$. For all other schemes, the two types of errors are comparable; therefore, only errors of the second type, which are more straightforward, are listed. These results on order of accuracy were tested for K from 2 to 7. With $K = 2$, Fourier analyses showed that most schemes are of order $2(K-1) = 0$; for the steady state case considered here, however, these schemes are of order 2. Although not listed, all interface quantities as well as cell average values are exact (up to machine errors).

The third test is a 2D problem: on the domain $[0,1]\cdot[0,1]$, with $a=.2$, consider the Poisson equation

$$u_{xx}+u_{yy}=s(x,y)\ =\ -\frac{1}{a^2}\sin(3\pi x)\sin(3\pi y)\,. \tag{8.7}$$

The Dirichlet boundary conditions are

$$u(0,y)=u(1,y)=u(x,0)=u(x,1)=0\,. \tag{8.8}$$

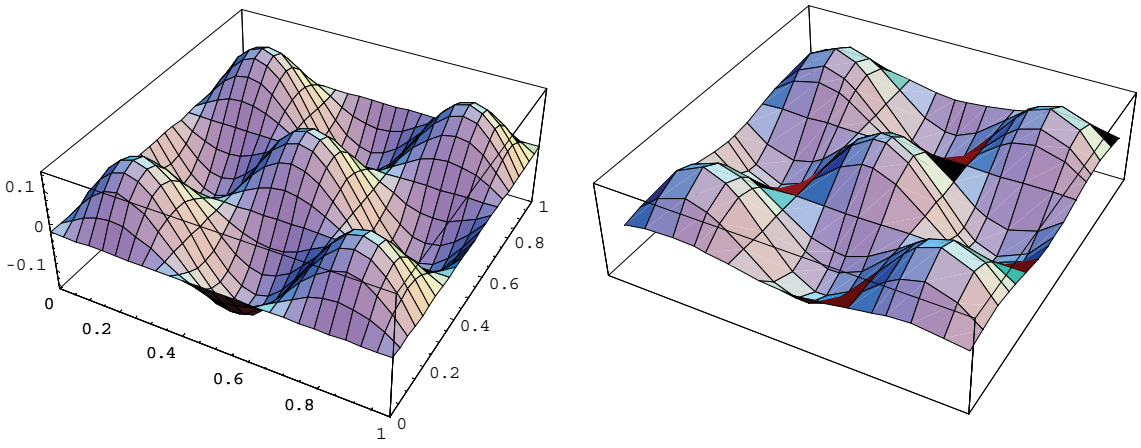
The exact solution is

$$u_{\text{exact}}(x,y)\ =\ -\frac{1}{2(3\pi a)^2}\sin(3\pi x)\sin(3\pi y)\,. \tag{8.9}$$

The domain $[0,1]\cdot[0,1]$ is divided into $J\cdot J$ uniform squares. On each cell (square), we employ $K\cdot K$ solution points, which are the Gauss points. For simplicity, at the solution points, the source term is evaluated with no projection: for $i,j=1,...,J$ and $k,l=1,...,K$,

$$s_{i,j,k,l}=s(x_{i,j,k,l})\,. \tag{8.10}$$

Figure 8.3 shows (a) the exact solution and (b) the solution with $4\cdot 4$ cells and $K=4$ using I - centered - g_{DG}/SP - g_{DG} or BR2 scheme. The solutions by other schemes are very similar.



(a) Exact solution (b) Solution by BR2 scheme

Figure 8.3. Solutions for problem (8.7), (8.8) on $[0,1]\cdot[0,1]$*: (a) exact and (b) with* $4\cdot 4$ *cells and* $K=4$ *using* I - centered - g_{DG}/SP - g_{DG}.

Scheme ($K=4$)	Order of accuracy	Max. Error; 2 cells	Error ratio	Max. Error; 4 cells	Error ratio	Max. Error; 8 cells
I - recovery/SP - g_{DG}		$.358\cdot 10^{-2}$		$.132\cdot 10^{-3}$		$.378\cdot 10^{-5}$
	5		27.		35.	
I - continuous - g_{Le}/SP - g_{DG}		$.237\cdot 10^{-2}$		$.720\cdot 10^{-4}$		$.214\cdot 10^{-5}$
	5		33.		34.	
I - centered - g_{DG}/SP - g_{DG}		$.117\cdot 10^{-1}$		$.851\cdot 10^{-3}$		$.116\cdot 10^{-3}$
	≈ 4		14		7	
I - one sided - g_{DG}/SP - g_{DG}		$.117\cdot 10^{-1}$		$.180\cdot 10^{-2}$		$.127\cdot 10^{-3}$
	≈ 4		7.		14.	
I - centered - g_{Gauss}/SP - $g_{\text{Lump, Lo}}$		$.450\cdot 10^{-1}$		$.147\cdot 10^{-2}$		$.134\cdot 10^{-3}$
	≈ 4		30		11	

Table 8.1. Maximum errors of values at solution points and corresponding orders of accuracy.

Table 8.2 shows the maximum errors at the solution points and orders of accuracy of various schemes with $K = 4$ for the 2D problem (8.7)-(8.8). Note that the orders of accuracy are essentially the same as those of Table 8.1.

9. Conclusions and discussion

In summary, a new approach to high-order accuracy for diffusion problems with the advantages of simplicity and economy was introduced. The approach evaluates the derivative for the discontinuous solution polynomial by first obtaining the derivatives at the solution points in a straightforward manner. These derivatives are then corrected, in each cell, by quantities proportional to the jumps at the two cell interfaces. The key idea is to solve the equations in differential form using a reconstruction technique where the discontinuous solution polynomials are approximated by piecewise polynomials that are continuous across the interfaces and of one degree higher than that of the solution polynomials. The approach resulted in several new methods including a simplified version of the DG scheme. It also led to new choices of common value and common gradient. Fourier stability and accuracy analyses were carried out. For the two-point boundary value problem, it was shown that nearly all of these reconstruction schemes (which include a majority of popular DG methods) yield exact common interface quantities as well as exact cell average solutions.

The approach can be extended to the case of the Navier-Stokes equations on a 2D quadrilateral mesh via tensor products. The extension to the case of a triangular mesh appears to be more complex and remains to be explored since the concept of tensor product is no longer available.

References

D.N. Arnold, "An interior penalty finite element method with discontinuous elements", SIAM J. Numer. Anal., v.19, pp. 742-760, 1982.

D. N. Arnold, F. Brezzi, B. Cockburn, and L. D. Marini, "Unified analysis of discontinuous Galerkin methods for elliptic problems", SIAM J. Numer. Anal., 39 (2002), pp. 1749–1779

F. Bassi and S. Rebay, "A high-order accurate discontinuous finite element method for the numerical solution of the compressible Navier-Stokes equations", J. Comput. Phys., 131 (1997), pp. 267-279.

F. Bassi and S. Rebay, "High-order accurate discontinuous finite element solution of the 2D Euler equations", J. Comput. Phys., 138 (1997), pp. 251-285.

F. Bassi and S. Rebay, "A high order discontinuous Galerkin method for compressible turbulent flows", in "Discontinuous Galerkin methods: Theory, Computation, and Application", B. Cockburn, G. Karniadakis, and C.-W. Shu, editors, Lecture Notes in Computational Science and Engineering, Springer (2000), pp. 77-88.

F. Brezzi, M. Manzini, D. Marini, P. Pietra, A. Russo, Discontinuous Galerkin approximations for elliptic problems, Numer. Methods Partial Differential Eq., 16, pp. 365-278, 2000.

B. Cockburn, G. Karniadakis, and C.-W. Shu, "The development of discontinuous Galerkin methods", in Discontinuous Galerkin methods: Theory, Computation, and Application, B. Cockburn, G. Karniadakis, and C.-W. Shu, editors, Lecture Notes in Computational Science and Engineering, Springer (2000), pp. 3-50

B. Cockburn and C.-W. Shu, "TVB Runge-Kutta local projection P^1-discontinuous Galerkin finite element method for conservation laws II: General framework", Math. Comp. , 52 (1989), pp. 411-453.

B. Cockburn and C.-W. Shu, "The local discontinuous Galerkin methods for time-dependent convection diffusion systems", SIAM J. Numer. Anal., 35 (1998), pp. 2440-2463

F. B. Hildebrand, Introduction to Numerical Analysis: Second Edition (Dover Books on Advanced Mathematics), 1987.

T.J.R. Hughes, "The finite element method", Dover (2000).

H. T. Huynh, "A Flux Reconstruction Approach to High-Order Schemes Including Discontinuous Galerkin Methods", 18th AIAA-CFD Conference, AIAA paper 2007-4079.

D.A. Kopriva, "A staggered-grid multidomain spectral method for the compressible Navier-Stokes equations", J. Comp. Phys., 143 (1998), pp. 125-158.

D.A. Kopriva and J.H. Kolias, " A conservative staggered-grid Chebyshev multidomain method for compressible flows", J. Comp. Phys., 125 (1996), pp. 244-261.

P. LeSaint and P.A. Raviart, "On the finite element method for solving the neutron transport equation", in C. de Boor, editor, "Mathematical aspects of finite elements in partial differential equations", Academic Press (1974), pp. 89-145.

Y. Liu, M. Vinokur, and Z.J. Wang, "Discontinuous Spectral Difference Method for Conservation Laws on Unstructured Grids," J. Comp. Phys., 216 (2006), pp. 780-801.

J. T. Oden, Ivo Babuska, and C. E. Baumann, "A discontinuous hp finite element method for diffusion problems", J. Comp. Phys., 146 (1998), pp. 491-519.

J. Peraire AND P.-O. Persson, "The compact discontinuous Galerkin (CDG) method for elliptic problems", SIAM J. Sci. Comput., Vol. 30, No. 4 (2008) pp. 1806-1824.

W.H. Reed and T.R. Hill, Triangular mesh methods for the neutron transport equation, Tech. Report LA-UR-73-479, Los Alamos Scientific Laboratory, 1973.

K. van den Abeele, C. Lacor and Z.J. Wang, "On the stability and accuracy of the spectral volume method", J. of Sci. Comput., in press.

B. van Leer, M. Lo, and M. van Raalte "A discontinuous Galerkin method for diffusion based on recovery", 18th AIAA-CFD Conference, 2007, AIAA paper 2007-4083.

B. van Leer and S. Nomura, "Discontinuous Galerkin for diffusion", 17th AIAA-CFD Conference, 2005, AIAA paper 2005-5108.

Appendix. Boundary Value problem (or Steady State Case)

In this appendix, we show that for the 1D boundary value problem, assuming that the solution exists, then the common interface values and derivatives and the cell average values of the solution are exact for essentially all schemes discussed.

More precisely, on the domain $[a,b]$, consider the Poisson equation

$$u_{xx} = s \tag{A.1}$$

with the Neumann boundary condition on the left and Dirichlet boundary condition on the right,

$$u_x(a) = \hat{\alpha} \quad \text{and} \quad u(b) = \beta \tag{A.2}$$

where $\hat{\alpha}$ and β are given constants and, for simplicity, $s = s(x)$ is assumed to be piecewise continuous. Let u_{exact} be the exact solution and $v_{\text{exact}} = (u_{\text{exact}})_x$. Then, for y in $[a,b]$,

$$v_{\text{exact}}(y) = \hat{\alpha} + \int_a^y s(z)\,dz \tag{A.3}$$

and, by integrating from right to left,

$$u_{\text{exact}}(x) = \beta + \int_b^x v_{\text{exact}}(y)\,dy\,. \tag{A.4}$$

Next, consider the following three sets of boundary conditions:

$$u(a) = \alpha\,, \quad u(b) = \beta\,, \tag{A.5a}$$

$$u(a) = \alpha\,, \quad u_x(a) = \hat{\alpha}\,, \text{ and} \tag{A.5b}$$

$$u(b) = \beta\,, \quad u_x(b) = \hat{\beta}\,. \tag{A.5c}$$

We now show that each of them can be recast in the form (A.2). For the case (A.5a), by (A.4) and (A.3),

$$u_{\text{exact}}(x) = \beta + \hat{\alpha}(x-b) + \int_b^x \int_a^y s(z)\,dz\,dy\,.$$

Therefore,

$$u_{\text{exact}}(a) = \beta - \hat{\alpha}(b-a) - \int_a^b \int_a^y s(z)\,dz\,dy\,.$$

Or

$$\alpha = \beta - \hat{\alpha}(b-a) - \int_a^b \int_a^y s(z)\,dz\,dy\,.$$

If α and β are given, we can evaluate $\hat{\alpha}$ in terms of α and β by the above equation. Therefore, (A.5a) can be cast in the form (A.2). The other two cases are similar and are omitted.

On domain $I = [-1,1]$. First, we consider the case of only one cell on the domain $[a,b] = [-1,1]$. Assume that $K \geq 1$. Denote by $\tilde{s}$ the projection of s onto $\boldsymbol{P}^{K-1}(I)$, the space of polynomials of degree $K-1$ or less on I. Using P_k, the Legendre polynomial of degree k, set

$$\tilde{s} = \text{Pr}_{K-1}(s) = \sum_{k=0}^{K-1} \gamma_k P_k$$

where, for $k = 0,...,K-1$, γ_k is defined by

$$\gamma_k(P_k, P_k) = (s, P_k)\,. \tag{A.6}$$

The exact solutions for the problem (A.1) and (A.2) on the domain $[-1,1]$ are

$$v_{\text{exact}}(y) = \hat{\alpha} + \int_{-1}^{y} s(\xi)\,d\xi \tag{A.7}$$

and

$$u_{\text{exact}}(x) = \beta + \int_{1}^{x} v_{\text{exact}}(y)\,dy\,. \tag{A.8}$$

The exact solutions for the same problem with s replaced by $\tilde{s}$ are

$$\tilde{v}(y) = \hat{\alpha} + \int_{-1}^{y} \tilde{s}(\xi)\,d\xi \tag{A.9}$$

and

$$\tilde{u}(x) = \beta + \int_{1}^{x} \tilde{v}(y)\,dy\,. \tag{A.10}$$

Here, since $\tilde{s}$ is of degree $K-1$, $\tilde{v}$ is of degree K, and $\tilde{u}$, degree $K+1$.

We claim that

$$\int_{-1}^{1} \tilde{s}(\xi)\,d\xi = \int_{-1}^{1} s(\xi)\,d\xi\,;\ \text{equivalently,}\ \tilde{v}(1) = v_{\text{exact}}(1)\,. \tag{A.11a,b}$$

In addition, if $K \geq 2$, then

$$\int_{-1}^{1} \tilde{v}(\xi)\,d\xi = \int_{-1}^{1} v_{\text{exact}}(\xi)\,d\xi\,;\ \text{equivalently,}\ \tilde{u}(-1) = u_{\text{exact}}(-1)\,. \tag{A.12a,b}$$

And, if $K \geq 3$, then

$$\int_{-1}^{1} \tilde{u}(\xi)\,d\xi = \int_{-1}^{1} u_{\text{exact}}(\xi)\,d\xi\,. \tag{A.13}$$

To prove this claim, note first that the projection of $\tilde{s}$ and s onto $\boldsymbol{P}^{K-1}(I)$ are equal; that is, for any polynomial p of degree $K-1$ or less,

$$(\tilde{s}, p) = (s, p)\,. \tag{A.14}$$

If $K \geq 1$, then $\boldsymbol{P}^{K-1}(I)$ contains $p = 1$. The above for $p = 1$ yields

$$\int_{-1}^{1} \tilde{s}(\xi)\,d\xi = \int_{-1}^{1} s(\xi)\,d\xi\,, \tag{A.15}$$

which is (A.11a). By (A.7) and (A.9), equation (A.11a) is equivalent to equation (A.11b).

Next, if $K \geq 2$, then $\boldsymbol{P}^{K-1}(I)$ contains $p(\xi) = \xi$. Equation (A.14) for this p yields

$$\int_{-1}^{1} \tilde{s}(\xi)\xi\,d\xi = \int_{-1}^{1} s(\xi)\xi\,d\xi\,. \tag{A.16}$$

Note that by (A.9), $\tilde{v}\,'(\xi) = \tilde{s}(\xi)$. Applying integration by parts to the left hand side above, we obtain

$$\int_{-1}^{1} \tilde{s}\,\xi\,d\xi = \int_{-1}^{1} \tilde{v}\,'\xi\,d\xi = \tilde{v}(1) + \tilde{v}(-1) - \int_{-1}^{1} \tilde{v}\,d\xi\,.$$

As for the right hand side of (A.16),

$$\int_{-1}^{1} s\,\xi\,d\xi = \int_{-1}^{1} v_{\text{exact}}\,'\xi\,d\xi = v_{\text{exact}}(1) + v_{\text{exact}}(-1) - \int_{-1}^{1} v_{\text{exact}}\,d\xi\,.$$

Since $\tilde{v}(-1) = v_{\text{exact}}(-1) = \hat{\alpha}$ and, by (A.11), $\tilde{v}(1) = v_{\text{exact}}(1)$, the above three equations imply

$$\int_{-1}^{1} \tilde{v}\,d\xi = \int_{-1}^{1} v_{\text{exact}}\,d\xi\,, \tag{A.17}$$

which is (A.12a). By (A.8) and (A.10), equation (A.12a) is equivalent to equation (A.12b).

Finally, if $K \geq 3$, then $\boldsymbol{P}^{K-1}(I)$ contains $p = \xi^2$. Equation (A.14) for this p yields

$$\int_{-1}^{1} \widetilde{s}(\xi)\xi^2 \, d\xi = \int_{-1}^{1} s(\xi)\xi^2 \, d\xi \,. \tag{A.18}$$

Again, by (A.9), $\widetilde{v}\,'(\xi) = \widetilde{s}(\xi)$. Applying integration by parts to the left hand side above, we have

$$\int_{-1}^{1} \widetilde{s}\,\xi^2 \, d\xi = \int_{-1}^{1} \widetilde{v}\,'\xi^2 \, d\xi = \widetilde{v}(1) - \widetilde{v}(-1) - 2\int_{-1}^{1} \widetilde{v}\,\xi \, d\xi \,. \tag{A.19}$$

In a similar manner, by (A.10), $\widetilde{u}\,' = \widetilde{v}$. Applying integration by parts to the last integral above, we obtain

$$\int_{-1}^{1} \widetilde{v}\,\xi \, d\xi = \int_{-1}^{1} \widetilde{u}\,'\xi \, d\xi = \widetilde{u}(1) + \widetilde{u}(-1) - \int_{-1}^{1} \widetilde{u} \, d\xi \,. \tag{A.20}$$

Substituting the above into (A.19), i.e., after integrating by parts twice, the result is

$$\int_{-1}^{1} \widetilde{s}\,\xi^2 \, d\xi = \widetilde{v}(1) - \widetilde{v}(-1) - 2\widetilde{u}(1) - 2\widetilde{u}(-1) + 2\int_{-1}^{1} \widetilde{u} \, d\xi \,.$$

A similar argument for the right hand side of (A.18) leads to

$$\int_{-1}^{1} s\,\xi^2 \, d\xi = v_{\text{exact}}(1) - v_{\text{exact}}(-1) - 2u_{\text{exact}}(1) - 2u_{\text{exact}}(-1) + 2\int_{-1}^{1} u_{\text{exact}} \, d\xi \,.$$

Recall that that $\widetilde{v}(-1) = v_{\text{exact}}(-1) = \hat{\alpha}$ and, by (A.11), $\widetilde{v}(1) = v_{\text{exact}}(1)$; in addition, $\widetilde{u}(1) = u_{\text{exact}}(1) = \beta$ and, by (A.12), $\widetilde{u}(-1) = u_{\text{exact}}(-1)$. Therefore, the above two equations and (A.18) imply

$$\int_{-1}^{1} \widetilde{u}(\xi) d\xi = \int_{-1}^{1} u_{\text{exact}}(\xi) d\xi \,.$$

This completes the proof.

Note that an argument along the same line as the above for the recovery scheme was discussed in (Van Leer et al. 2007).

Exactness of common interface quantities and cell average values regardless of the scheme. Let the domain $[a,b]$ be partitioned into (possibly irregular) cells E_j of center x_j and length h_j, $j = 1,...,J$. Assume that $K \geq 1$. Let $x_{j,k}$, $k = 1,...,K$, be any set of solution points. On each cell E_j, denote by $s_j = s_j(x)$ the polynomial of degree $K-1$ that is the projection of s onto $\boldsymbol{P}^{K-1}(E_j)$:

$$s_j = \text{Pr}_{K-1}(s) \,. \tag{A.21}$$

For the s values at the solution points, we employ, with $k = 1,...,K$,

$$s_{j,k} = s_j(x_{j,k}) \,. \tag{A.22}$$

The problem takes the form: find $u_{j,k}$ such that for all j and k

$$(u_{xx})_{j,k} = s_{j,k}$$

with the following boundary conditions:

$$(u_x)_{1/2} = (u_x)^{\text{com}}_{1/2} = \hat{\alpha} \quad \text{and} \quad u_{J+1/2} = u^{\text{com}}_{J+1/2} = \beta \,.$$

At each interface $j+1/2$, let $u^{\text{com}}_{j+1/2}$ be the common value and $(u_x)^{\text{com}}_{j+1/2}$ the common derivative. They can be calculated from $u_{j,k}$ and $u_{j+1,k}$, $k = 1,...,K$.

The following claim is in order. Assuming that the solution exists, then, regardless of the method, (a) the common derivatives are exact; (b) if $K \geq 2$ and the correction function g for the solution points has a zero average on $[-1,1]$, then the common interface values are exact. In addition, (c) if $K \geq 3$ and the correction function g for the solution points is orthogonal to $\boldsymbol{P}^1$, then the cell average values of the solution are also exact.

Concerning statement (b) in the above claim, since g is orthogonal to $\boldsymbol{P}^{K-3}$, g has a zero average if $K \geq 3$. Moreover, if $g = g_{\mathrm{DG}}$, then, since g_{DG} is orthogonal to $\boldsymbol{P}^{K-2}$, it has a zero average if $K \geq 2$. Concerning statement (c), since g is orthogonal to $\boldsymbol{P}^{K-3}$, g is orthogonal to $\boldsymbol{P}^{1}$ if $K \geq 4$. Moreover, if $g = g_{\mathrm{DG}}$, then, since g_{DG} is orthogonal to $\boldsymbol{P}^{K-2}$, g is orthogonal to $\boldsymbol{P}^{1}$ if $K \geq 3$.

To prove the claim, we employ the reconstruction functions. Recall that the reconstruction scheme has the following properties:

$u_j^C(x)$ is of degree K and takes on the value $u_{j-1/2}^{\mathrm{com}}$ at $x_{j-1/2}$ and $u_{j+1/2}^{\mathrm{com}}$ at $x_{j+1/2}$;

$\{v_j(x) = (u_j^C)_x(x)\}$ is a discontinuous piecewise polynomial of degree $K-1$;

$v_j^C(x)$ is of degree K and takes on the value $v_{j-1/2}^{\mathrm{com}}$ at $x_{j-1/2}$ and $v_{j+1/2}^{\mathrm{com}}$ at $x_{j+1/2}$;

$\{(v_j^C)_x\}$ is a discontinuous piecewise polynomial of degree $K-1$.

If $u_{j,k}$ are the solution of the boundary value problem, then, in each cell, the corresponding $(v_j^C)_x$, which is of degree $K-1$, satisfies, for $k = 1,...,K$,

$$(v_j^C)_x(x_{j,k}) = s_{j,k}\,. \tag{A.23}$$

Since $s_j(x)$ is the (unique) polynomial of degree $K-1$ interpolating $s_{j,1}$, …, $s_{j,K}$, the above implies $(v_j^C)_x(x) = s_j(x)$. Therefore, recalling that $v_{j-1/2}^{\mathrm{com}} = (u_x)_{j-1/2}^{\mathrm{com}}$,

$$v_j^C(x) = (u_x)_{j-1/2}^{\mathrm{com}} + \int_{x_{j-1/2}}^{x} s_j(z)\,dz\,. \tag{A.24}$$

Consequently,

$$(u_x)_{j+1/2}^{\mathrm{com}} = (u_x)_{j-1/2}^{\mathrm{com}} + \int_{x_{j-1/2}}^{x_{j+1/2}} s_j(z)\,dz\,. \tag{A.25}$$

Concerning the above integral, since $s_j = \mathrm{Pr}_{K-1}(s)$, we have

$$\int_{x_{j-1/2}}^{x_{j+1/2}} s_j(x)\,dx = \int_{x_{j-1/2}}^{x_{j+1/2}} s(x)\,dx\,. \tag{A.26}$$

We now march from $j=1$ to $j=J$. For $j=1$, at the left interface of E_1, namely, at $x_{1/2} = a$, the quantity $(u_x)_{1/2}^{\mathrm{com}} = \hat{\alpha}$ is known and is exact. We can calculate $(u_x)_{3/2}^{\mathrm{com}}$ by applying (A.25) with $j=1$. Equations (A.26) and (A.3) then imply that $(u_x)_{3/2}^{\mathrm{com}}$ is exact. Continuing for $j=2$, …, J, respectively, we conclude that all common interface derivatives $(u_x)_{j+1/2}^{\mathrm{com}}$ are exact regardless of the method: for all j,

$$(u_x)_{j+1/2}^{\mathrm{com}} = (v_{\mathrm{exact}})_{j+1/2}\,. \tag{A.27}$$

Note that in the proof, the key ingredient is $(v_j^C)_x = s_j$; we do not need v_j.

To show part (b), i.e., the exactness of the common interface values $u_{j+1/2}^{\mathrm{com}}$, we integrate from right to left and make use of the same result in the local coordinate, namely, (A.12). Note first that by (A.24), $v_j^C(x)$ is independent of the method chosen; $v_j(x)$, however, is method dependent. Therefore, we wish to obtain $u_{j-1/2}^{\mathrm{com}} - u_{j+1/2}^{\mathrm{com}}$ by integrating $v_j^C(x)$ instead of $v_j(x)$. To this end, recall that by definition, $v_j(x) = (u_j^C)_x(x)$. In addition, at all interfaces, $u_j^C(x_{j+1/2}) = u_{j+1/2}^{\mathrm{com}}$. Thus, for any x on E_j,

$$u_j^C(x) = u_{j+1/2}^{\mathrm{com}} + \int_{x_{j+1/2}}^{x} v_j(y)\,dy\,.$$

With $x = x_{j-1/2}$,

$$u_{j-1/2}^{\mathrm{com}} - u_{j+1/2}^{\mathrm{com}} = \int_{x_{j+1/2}}^{x_{j-1/2}} v_j(y)\,dy\,. \tag{A.28}$$

In the local coordinate, $v_j^C(\xi) = v_j(\xi) + c_1 g^{\mathrm{LB}}(\xi) + c_2 g^{\mathrm{RB}}(\xi)$ for some constants c_1 and c_2. Now, suppose the correction functions for the solution points satisfy, for $g = g^{\mathrm{LB}}$ and $g = g^{\mathrm{RB}}$,

$$\int_{-1}^{1} g(\xi)\, d\xi = 0 . \tag{A.29}$$

Then,

$$\int_{x_{j+1/2}}^{x_{j-1/2}} v_j(y)\, dy = \int_{x_{j+1/2}}^{x_{j-1/2}} v_j^C(y)\, dy . \tag{A.30}$$

As a consequence, if (A.29) holds, then, by (A.28),

$$u_{j-1/2}^{\mathrm{com}} - u_{j+1/2}^{\mathrm{com}} = \int_{x_{j+1/2}}^{x_{j-1/2}} v_j^C(y)\, dy . \tag{A.31}$$

To show that these interface values are exact, we employ (A.12). Starting with the last cell E_J, the boundary quantity $u_{J+1/2}^{\mathrm{com}} = \beta$ is exact. The polynomial v_J^C of degree K plays the role of $\widetilde{v}$ in (A.9). Therefore, (A.12) implies that $u_{J-1/2}^{\mathrm{com}}$ is exact. By using this argument successively for $j = J-1,...,1$, i.e., marching from right to left, we conclude that the interface values $u_{j+1/2}^{\mathrm{com}}$ are exact provided that $K \geq 2$ and (A.29) holds.

Finally, assuming that $K \geq 3$, we wish to employ (A.13) to show that the cell average values of the solution are exact. The problem here is that $\widetilde{u}$ of (A.13) is of degree $K+1$ whereas u_j is of degree $K-1$. To bridge this gap, we first observe that since $u_j^C = u_j + c_3 g^{\mathrm{LB}} + c_4 g^{\mathrm{RB}}$ for some constants c_3 and c_4, and the averages of g^{LB} and g^{RB} are zero by our assumption, the average of u_j on E_j is the same as that of u_j^C (degree K). Therefore, we only need to bridge the gap between u_j^C (degree K) and $\widetilde{u}$ (degree $K+1$). To this end, recall that we integrate s_j, which is of degree $K-1$, to obtain v_j^C, which is of degree K; here, v_J^C plays the role of $\widetilde{v}$. Next, we integrate v_j and v_j^C to obtain u_j^C, and $\widetilde{u}$, respectively: for any x on E_j,

$$u_j^C(x) = u_{j+1/2}^{\mathrm{com}} + \int_{x_{j+1/2}}^{x} v_j(y)\, dy ,$$

and

$$\widetilde{u}(x) = u_{j+1/2}^{\mathrm{com}} + \int_{x_{j+1/2}}^{x} v_j^C(y)\, dy .$$

Since $v_j^C = v_j + c_1 g^{\mathrm{LB}} + c_2 g^{\mathrm{RB}}$, the problem reduces to the following. With $p = g^{\mathrm{LB}}$ or $p = g^{\mathrm{RB}}$, set

$$P(\eta) = \int_{-1}^{\eta} p(\xi)\, d\xi . \tag{A.32}$$

We wish to show that

$$\int_{-1}^{1} P(\eta)\, d\eta = 0 . \tag{A.33}$$

If we can show the above, then u_j^C, and $\widetilde{u}$ have the same averages on E_j.

In fact, we will show a more general and simpler to state result: if p is orthogonal to $\boldsymbol{P}^1$, i.e.,

$$\int_{-1}^{1} p(\xi)\, d\xi = 0 \quad \text{and} \quad \int_{-1}^{1} p(\xi)\, \xi\, d\xi = 0 , \tag{A.34a,b}$$

then P defined by (A.32) satisfies (A.33). Indeed, by (A.34a) and definition (A.32) of P, we have $P(1) = 0$. Next, using integration by parts and (A.34b),

$$\int_{-1}^{1} P(\xi)\, d\xi = \left[P(\xi)\,\xi\right]_{-1}^{1} - \int_{-1}^{1} P'(\xi)\,\xi\, d\xi = P(1) + P(-1) - \int_{-1}^{1} p(\xi)\,\xi\, d\xi = 0 . \tag{A.35}$$

This completes the proof.