Preprint:
Hiroaki Nishikawa
A first-order system approach for diffusion equation. I: Second-order residual-distribution schemes
Journal of Computational Physics 227 (2007) 315–352

A First-Order System Approach for Diffusion Equation. I. Second-Order Residual-Distribution Schemes

Hiroaki Nishikawa

W. M. Keck Foundation Laboratory for Computational Fluid Dynamics,
Department of Aerospace Engineering,
University of Michigan, FXB Building, 1320 Beal Avenue, Ann Arbor, MI 48109-2140, USA

Abstract

In this paper, we embark on a new strategy for computing the steady state solution of the diffusion equation. The new strategy is to solve an equivalent first-order *hyperbolic* system instead of the second-order diffusion equation, introducing solution gradients as additional unknowns. We show that schemes developed for the first-order system allow $O(h)$ *time step* instead of $O(h^2)$ and converge very rapidly toward the steady state. Moreover, this extremely fast convergence comes with the solution gradients (viscous stresses/heat fluxes for the Navier-Stokes equations) simultaneously computed with *the same order of accuracy* as the main variable. The proposed schemes are formulated as residual-distribution schemes (but can also be identified as finite-volume schemes), directly on unstructured grids. We present numerical results to demonstrate the tremendous gains offered by the new diffusion schemes, driving the rise of *explicit* schemes in the steady state computation for diffusion problems.

Key words: diffusion first-order system fast convergence large time step residual distribution unstructured grids

1 Introduction

In this paper, we embark on a new strategy for computing the steady state solution to the diffusion equation,

$$u_t = \nu(u_{xx} + u_{yy}), \tag{1.1}$$

where ν is a positive diffusion coefficient. The new strategy is based on the following first-order system:

$$\begin{array}{rcl} u_t &=& \nu\,(p_x + q_y), \\ p_t &=& (u_x - p)/T_r, \\ q_t &=& (u_y - q)/T_r, \end{array} \tag{1.2}$$

where T_r may be called a relaxation time. This is in fact a relaxation system, often called the hyperbolic heat equations, asymptotically equivalent to the original diffusion equation as $T_r \to 0$ [1, 2, 3]. There have been many attempts to develop numerical methods for such relaxation systems [1, 4, 5, 6], with a particular focus on the stiffness problem: an explicit time step, $\Delta t = O(T_r) \to 0$, is prohibitively restricted due to an extremely small relaxation time; an implicit treatment of the stiff source term could degrade the solution accuracy [7]. Although based on the same equations, a new strategy is radically different from these relaxation methods. The key is to realize the fact that the first-order system is equivalent to the diffusion equation at the steady state ($u_t = p_t = q_t = 0$) for *any* T_r:

$$\begin{array}{rcl} 0 &=& \nu\,(p_x + q_y), \\ 0 &=& (u_x - p)/T_r, \\ 0 &=& (u_y - q)/T_r, \end{array} \quad \to \quad \begin{array}{rcl} 0 &=& \nu\,(p_x + q_y), \\ p &=& u_x, \\ q &=& u_y, \end{array} \quad \to \quad 0 = \nu\,(u_{xx} + u_{yy}). \tag{1.3}$$

Then, as far as the steady state computation is concerned, the relaxation time T_r is a free parameter, and the stiffness is no longer an issue. In short, we gain the freedom to choose T_r to avoid the stiffness by giving up the time accuracy. This is the key idea of the new strategy. And we will see in due course that this simple idea paves the way for the rise of *explicit* schemes in the steady state computation for diffusion problems, and also brings a dramatic change in the way an advection scheme and a diffusion scheme are combined for advection-diffusion problems.

In developing numerical schemes for the first-order diffusion system (1.2), we focus on the residual-distribution (or fluctuation-splitting) method. This is partly because the present study was originally motivated by the need to develop diffusion schemes in the framework of the residual-distribution method, and also because this method has superior features especially for unstructured grids. This is a method based on nodal degrees of freedom and cell-residuals in the same spirit of the cell-vertex schemes [8], but its development has been almost exclusively for triangular unstructured grids. It has been developed extensively for problems dominated by advection and wave propagation because of the ability to reflect multidimensional physics of the governing equations [9, 10, 11, 12, 13, 14]. But on the other hand, its application to diffusion problems had long been almost untouched, apparently because diffusion is an isotropic process and does not benefit particularly from such a multidimensional capability. In fact, it has been a standard practice to discretize the viscous term by the Galerkin method and simply add to the existing residual-distribution Euler code to construct a Navier-Stokes code [15, 16, 17]. It was pointed out in [18] however that such a strategy deteriorated the formal accuracy of the scheme due to an incompatibility of the two discretizations, especially in regions where advection and diffusion effects are equally important. Then, in [18], a first-order system approach was introduced as a basis for developing uniformly accurate schemes for the advection-diffusion problems. But without the time derivatives and the relaxation time, it only discusses the spatial discretization and no details on the method to compute the steady state solution is given. In this paper, we introduce the time derivatives and the relaxation time to write the first-order system as a set of evolution equations as in (1.2), and develop a class of residual-distribution schemes for computing the steady state solution. In so doing, we take full advantage of having an arbitrary relaxation time. We will show in particular that we can develop a class of schemes that allow an $O(h)$ time step, where h is a mesh size, instead of the conventional $O(h^2)$ time step. This is a tremendous gain, and shows a great potential for promoting the use of explicit methods for steady state computations in diffusion problems for which $O(h^2)$ time step has always been the major obstacle for using explicit methods (even for steady calculations) and the motivation for resorting to other methods such as implicit methods. Moreover, this rapid convergence comes with solution gradients computed with the equal order of accuracy as the solution u. This not only eliminates the need of post-processing to compute the physical quantity of interest such as viscous stresses or heat fluxes, but also provides such quantities with excellent accuracy whereas the post-processed quantities often lose the order of accuracy by at least one. We also pay a particular attention to the relation with the Galerkin discretization. The Galerkin discretization does not precisely fit in the framework of residual-distribution (although can be arranged as if it is), but rather surprisingly it is shown to emerge as a special case of the proposed schemes. Although this paper is largely concerned with the residual-distribution method, finite-difference or finite-volume schemes can also be developed based on the same first-order system. We believe that it can be done straightforwardly and the description of the one-dimensional residual-distribution schemes in this paper will provide a guide for developing these schemes.

The first-order system, although in a slightly different form, has often been utilized for developing diffusion schemes in finite-element methods: the mixed finite-element method [19] or the least-squares finite-element method [20]. But the focus there is rather on accuracy, and the method to obtain the steady state solution is not paid a particular attention, which makes it hard to compare the present approach with. Also, in the discontinuous Galerkin method, the first-order system is utilized for a proper discretization of diffusion terms [21]. The same approach was taken also in the spectral finite-volume method [22]. In these methods, because of the discontinuous nature of the numerical solution, the gradient variables are explicitly solved locally and eliminated by direct substitution back into the diffusion term. Therefore, the first-order system disappears at the end of the discretization. In the case of the residual-distribution method, this is not possible because the solution data is continuous, and therefore we end up with a globally coupled system of equations. In effect, we will be solving this global system iteratively by marching in time until convergence. This however should not be taken as a disadvantage because this is how the residual-distribution schemes achieve second-order accuracy at the steady state without reconstruction. Also this makes it possible to achieve a rapid convergence to the steady state with $O(h)$ time step in the first-order system approach with the gradient variables directly available on boundary nodes where such information is particularly valuable (e.g., skin friction/heating rate).

We set out in Section 2 the residual-distribution method in relation to diffusion problems. Describing the

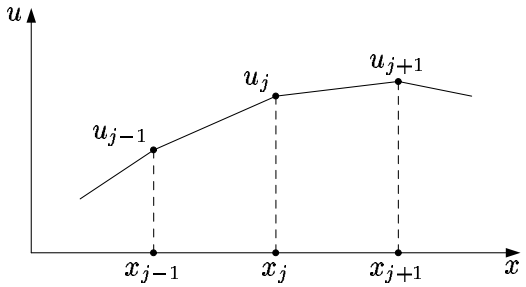

Figure 1: Continuous piecewise linear data representation.

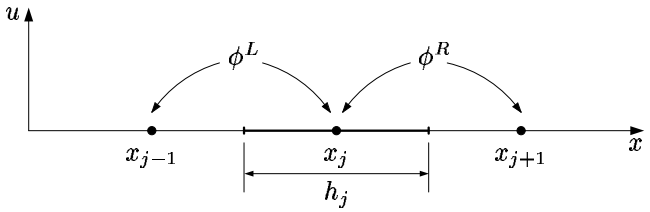

Figure 2: Distribution of cell-residuals and the dual control volume.

difficulties with the diffusion equation, we finally arrive at the first-order system approach. We then begin to develop a class of residual-distribution schemes for the first-order system. In Section 3 we describe the development and the analysis of the new schemes in one dimension. It is then extended to two dimensions in Section 4. In Section 5, we show that the first-order system approach can be used to derive dissipation terms for scalar diffusion schemes. In Section 6, we present numerical results to demonstrate the accuracy and the convergence properties of the new schemes for both one-dimensional and two-dimensional problems.

2 Residual-Distribution Method and Diffusion Equation

2.1 Residual-Distribution Method in One Dimension

We call methods residual-distribution if they can be factored into the two steps, *residual evaluation and distribution*. Consider computing the steady state solution of the one-dimensional conservation law,

$$u_t + f_x = q. \tag{2.1}$$

To discretize, we generate a set of nodes $\{J\}$ with coordinates x_j distributed arbitrarily over the domain of interest, and store the solution at each node $(u_j, p_j),\ \ j \in \{J\}$ assuming the piecewise linear variation over each cell (see Figure 1). This defines a set of cells $\{C\}$ of size $\Delta x_C = x_{j+1} - x_j$. Then, for each cell, we evaluate the cell-residual (or fluctuation) ϕ^C as an integral value of the steady part of the equation,

$$\phi^C = -\int_C (f_x - q)\, dx = -\left(f_{j+1} - f_j\right) + \frac{q_{j+1} + q_j}{2}\left(x_{j+1} - x_j\right), \tag{2.2}$$

where the source term has been evaluated by the trapezoidal rule. Note that the source term approximation has been deliberately chosen to be exact for linear q, in order to be compatible with the accuracy of the other term. This defines a measure of the error in satisfying the steady equation over the cell. If this does not vanish, we must change the nodal solutions to reduce the error. This brings the second step, i.e., distribution. We determine fractions of ϕ^C to be distributed to the nodes on the left and the right, ϕ^C_j and ϕ^C_{j+1} by

$$\phi^C_j = \beta^C_j \phi^C, \quad \phi^C_{j+1} = \beta^C_{j+1}\phi^C, \tag{2.3}$$

where β^C_j and β^C_{j+1} are distribution coefficients that satisfy

$$\beta^C_j + \beta^C_{j+1} = 1 \tag{2.4}$$

for conservation. Having done this for all cells, we have the following semi-discrete equation, with L and R indicating the left and right cells of node j,

$$\frac{du_j}{dt} = \frac{1}{h_j}\left[\phi^L_j + \phi^R_j\right] = \frac{1}{h_j}\left[\beta^L_j \phi^L + \beta^R_j \phi^R\right], \tag{2.5}$$

where $h_j = (x_{j+1} - x_{j-1})/2$, which we integrate until we reach the steady state. The key to construct a successful scheme is, of course, the choice of the distribution coefficient β^C_j. This is where the physics of the

equation plays an important role. For example, for hyperbolic equations, an upwind scheme is constructed by the following distribution coefficients:

$$\beta_j^C = \frac{1}{2}\left(1 - \frac{a^C}{|a^C|}\right), \quad \beta_{j+1}^C = \frac{1}{2}\left(1 + \frac{a^C}{|a^C|}\right), \tag{2.6}$$

where $a^C = (\partial f/\partial u)^C$ which may be evaluated using the Roe linearization, $f_{j+1} - f_j = a^C(u_{j+1} - u_j)$ [23]. In fact, with these coefficients, the semi-discrete equation (2.5) can be written as

$$\frac{du_j}{dt} = -\frac{1}{h_j}\left[F_{j+\frac{1}{2}} - F_{j-\frac{1}{2}}\right] + \hat{q}_j, \tag{2.7}$$

where

$$F_{j+\frac{1}{2}} = \frac{1}{2}(f_{j+1} + f_j) + \frac{|a^C|}{2}(u_{j+1} - u_j) \tag{2.8}$$

$$\hat{q}_j = \frac{1}{h_j}\left[\beta_j^L \frac{q_j + q_{j-1}}{2}\Delta x_L + \beta_j^R \frac{q_{j+1} + q_j}{2}\Delta x_R\right]. \tag{2.9}$$

This can be interpreted as a finite-volume scheme with a rather complicated source term discretization which would be simply $\hat{q}_j = q_j$ in the finite-volume method. Hence, the residual distribution scheme and the finite-volume scheme are identical except for the source term discretization. Note that the scheme is second-order accurate at a steady state. This is true for any bounded distribution coefficients on general non-uniform grids. This is because the nodal residual is a weighted average of cell-residuals that vanish individually for exact linear solutions of the conservation law. This property is called residual property and one of the reasons for the superior accuracy of the residual-distribution schemes on irregular grids. This is particularly advantageous over the finite-difference and the finite-volume schemes for advection-diffusion problems where nonuniform grids are desirable to efficiently resolve narrow transition regions such as boundary layers.

If implemented as a finite-volume scheme with $\hat{q}_j = q_j$, the scheme will be only first-order accurate at a steady state due to the lack of the residual property. To recover the second-order accuracy, the source term must be discretized in such a way that the steady equation

$$f_x = q \tag{2.10}$$

is satisfied with second-order accuracy at a steady state. This can be done by using the residual distribution formulation which gives a proper discretization such as (2.9), or by using other techniques specific to the finite-volume method (see [24] and references therein). In particular, a method in [25] is capable of producing a finite-volume scheme in the form (2.8) with (2.9).

2.2 Residual-Distribution Method in Two Dimensions

Now, in two dimensions, consider again solving the conservation law,

$$u_t + f_x + g_y = q. \tag{2.11}$$

We begin by dividing the domain of interest into a set of triangles $\{T\}$, with a set of nodes $\{J\}$, and store the solution values at nodes. We then proceed as in one dimension, first to evaluate the cell-residual. For each triangular cell $T \in \{T\}$ with vertices $\{i_T\} = \{1, 2, 3\}$, we evaluate a local cell-residual, ϕ^T,

$$\phi^T = -\iint_T (f_x + g_y - q)\, dxdy, \tag{2.12}$$

which becomes, for a piecewise linear approximation of f, g, and q,

$$\phi^T = -(f_x^T + g_y^T)S_T + \frac{q_1 + q_2 + q_3}{3}S_T = -\sum_{i \in \{i_T\}} \frac{1}{2}(f_i, g_i) \cdot \mathbf{n}_i + \frac{q_1 + q_2 + q_3}{3}S_T, \tag{2.13}$$

where f_x^T and g_y^T denote constant derivatives over the triangle, S_T is the area of the triangle, $\{i_T\}$ denotes a set of nodes that form the triangle, and $\mathbf{n}_i$ is the scaled inward normal vector of the edge opposite to node i (see

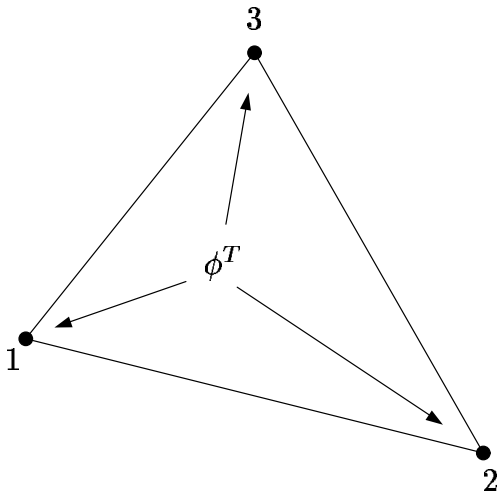

Figure 3: Distribution of a non-zero cell-residual to the set of vertices $\{i_T\} = \{1, 2, 3\}$.

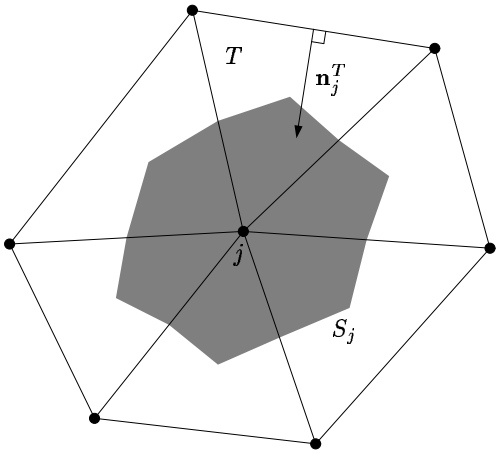

Figure 4: Median dual cell around node j in the set of triangles sharing that node $\{T_j\}$.

Figure 4). Note that the source term approximation has been deliberately chosen, as in one dimension, to be compatible with the accuracy of the other term. We now move on to distribute the cell-residual to the nodes. We determine a fraction ϕ_i^T of ϕ^T to be distributed to node i of triangle T by

$$\phi_i^T = \beta_i^T \phi^T \quad i \in \{i_T\}, \tag{2.14}$$

(see Figure 3) where β_j^T is a distribution coefficient with the property

$$\sum_{i \in \{i_T\}} \beta_i^T = 1 \tag{2.15}$$

for conservation. Again, it is the distribution coefficient that reflects the physics of the governing equations. There has been extensive research work on the distribution coefficients almost exclusively for hyperbolic problems, and today various upwind schemes are available (see [9, 26] for example). Note that the upwind scheme in two dimensions is not unique even for a linear problem, and that residual-distribution schemes cannot always be rephrased as a finite-volume scheme. This means that the residual-distribution schemes are fundamentally different from the finite-volume schemes, and the connection between the two methods begins to blur in higher dimensions.

It is important to note that the cell-residual (2.13) vanishes for exact linear solutions, nothing will be distributed then, and the solution is preserved as a result. So, we have the residual property, and it is independent of the shape of the cell. This is a great advantage especially for unstructured grids. And as in one dimension, the scheme is therefore second-order accurate at the steady state for bounded distribution coefficients [11]. Note that this is no longer true if we evaluate the source term separately by a point value as is done in the finite-volume schemes, and the scheme will then be only first-order accurate. In this study, we do not consider this option.

Finally, accumulating the partial residuals distributed at node j, we arrive at the following semi-discrete form:

$$\frac{du_j}{dt} = \frac{1}{S_j} \sum_{T \in \{T_j\}} \phi_j^T, \tag{2.16}$$

where S_j is the median dual cell area around node j, and $\{T_j\}$ denotes a set of triangles sharing the node (see Figure 4). We then integrate this in time to reach the steady state.

2.3 Galerkin Discretization of Diffusion Equation

In applying the residual-distribution method to the diffusion equation which involves second-order derivatives, we immediately notice that a cell-residual cannot be defined over a cell because it vanishes identically for

piecewise linear solutions. One way to overcome this difficulty is to discretize the diffusion term directly at a node by the Galerkin method, and then write the result as a sum of the contributions from the nearby cells as if it is residual-distribution. Consider the one-dimensional diffusion equation,

$$u_t = \nu u_{xx}. \tag{2.17}$$

We assume a uniform grid $h = x_{j+1} - x_j$, and apply the Galerkin method: multiply the equation by the piecewise linear basis function that takes 1 at node j, and 0 at nodes $j-1$, and $j+1$, and then integrate by parts from $x = x_{j-1}$ to $x = x_{j+1}$. Then, lumping the left hand side, we obtain the following semi-discrete equation:

$$h\frac{du_j}{dt} = \frac{\nu}{h}\left(u_{j+1} - 2u_j + u_{j-1}\right), \tag{2.18}$$

which can be written as

$$\frac{du_j}{dt} = \frac{1}{h}\left[\phi_j^R + \phi_j^L\right] = \frac{1}{h}\left[\frac{\nu\left(u_{j+1} - u_j\right)}{h} - \frac{\nu\left(u_j - u_{j-1}\right)}{h}\right], \tag{2.19}$$

so that we find that the contributions to the nodes within cell C are defined as

$$\phi_j^C = \frac{\nu(u_{j+1} - u_j)}{h}, \quad \phi_{j+1}^C = -\frac{\nu(u_{j+1} - u_j)}{h}. \tag{2.20}$$

In this form, the scheme can be implemented in the residual-distribution framework. However, it is clear that the contributions within a cell sum up to zero: $\phi^C = \phi_j^C + \phi_{j+1}^C = 0$. Hence the cell-residual does not exist, and in this sense the Galerkin scheme is not residual-distribution.

Similarly, the two-dimensional diffusion equation (1.1) can be easily discretized by the Galerkin method. Or equivalently, we can directly integrate the diffusion term over a set of triangles $\{T_j\}$: first convert the integral to the line integral around $\{T_j\}$ by the divergence theorem, and then evaluate it with the constant gradient over each triangular cell. In either way, we arrive at the following discretization:

$$S_j\frac{du_j}{dt} = -\frac{\nu}{2}\sum_{T\in\{T_j\}} \nabla u^T \cdot \mathbf{n}_j^T, \tag{2.21}$$

where $\mathbf{n}_j^T$ is the scaled inward normal vector of the edge opposite to node j of triangle T (see Figure 4). Then, we find from this that the contribution to node i within cell T is defined as

$$\phi_i^T = -\frac{\nu}{2}\nabla u^T \cdot \mathbf{n}_j^T, \tag{2.22}$$

which however again sums up to zero over the cell because $\mathbf{n}_1^T + \mathbf{n}_2^T + \mathbf{n}_3^T = 0$, and therefore no cell-residual exists. This might seem a natural consequence because the diffusion term identically vanishes over the cell for piecewise linear solutions, but in fact, it has been shown that this is true for *any* basis functions [27]. Cell-residuals are necessary for a scheme to be residual-distribution and even vital for the advection-diffusion schemes in which cell-residuals for the entire equation are sought. It seems hopeless to have cell-residuals for the Galerkin scheme, but we will discover later that cell-residuals for the Galerkin scheme do exist; they emerge, rather surprisingly and paradoxically in a way, out of the residual-distribution schemes that we propose in this paper.

2.4 Residual-Distribution for Diffusion Equation

It is possible to evaluate a cell-residual for the diffusion term if the solution gradient is available at nodes. In one dimension, we may reconstruct the gradient at node j, $(u_x)_j$, by a simple finite-difference approximation,

$$(u_x)_j = \frac{u_{j+1} - u_{j-1}}{2h}, \tag{2.23}$$

and evaluate the cell-residual as

$$\phi^C = \int_C \nu u_{xx}\, dx = \nu\left[(u_x)_{j+1} - (u_x)_j\right]. \tag{2.24}$$

This does not vanish identically and therefore can drive the change of the nodal solutions. Similarly in two dimensions, we can reconstruct the gradients at nodes, and then evaluate cell-residuals for the diffusion term. This type of scheme was studied in [12, 27, 28] and in [13] for quadrilateral grids. To distribute the cell-residual, in [13, 18, 27], equal weights are proposed to reflect the isotropic nature of diffusion, and in [12, 28] where the advection-diffusion problems are considered, upwind coefficients are used for the entire cell-residual.

The resulting scheme is genuinely residual-distribution: it has the residual property and can be naturally combined with an advection scheme for the advection-diffusion problems. But the scheme is no longer compact because the stencil has been extended by way of reconstruction. For example, in order for the scheme to be second-order accurate, the cell-residual must be evaluated with second-order accuracy. This requires at least a quadratic reconstruction, thus demanding a very large stencil especially in two dimensions. Even worse, it is pointed out in [13] that these schemes (with bounded distribution coefficients) always suffer from a lack of dissipation for high-frequency error modes for both triangular and quadrilateral grids. Certainly, these schemes need some form of dissipation, but deriving a dissipation term for the scalar diffusion scheme turns out to be a nontrivial task. But we will discover a form of dissipation from the new diffusion schemes we develop in this paper. We will discuss this in more details in Section 5.

2.5 First-Order System Approach

We now propose a new strategy: we carry the gradient p as unknown and solve the first-order system instead,

$$\begin{aligned} u_t &= \nu\, p_x, \\ p_t &= (u_x - p)/T_r, \end{aligned} \tag{2.25}$$

where T_r is a free parameter. This is then equivalent to the diffusion equation, $u_t = \nu\, u_{xx}$, at the steady state where exactly we seek the solution. With the first-order system, since there appear only first-order derivatives, the cell-residuals can be evaluated straightforwardly with second-order accuracy without reconstruction as we store all variables (u, p) at nodes. In short, we can now develop compact schemes. And this is true not only for the residual-distribution schemes but also for finite-difference or finite-volume schemes, simply because we no longer need to discretize the second-order derivative which generally requires an extended stencil. This is one of the advantages of solving the first-order system instead of the second-order diffusion equation. In fact, in general, there are a number of advantages for solving first-order systems in place of equations with higher derivatives: compact stencils, stiffness made local, ease of functional decomposition, and so on. An extensive discussion on the use of first-order systems in computational fluid dynamics is given by Van Leer [29]. Here, we focus on the aspects particular to the first-order diffusion system.

The first-order system (2.25) is identical to the hyperbolic heat equations: asymptotically equivalent to the original diffusion equation as $T_r = O(\nu) \to 0$; correctly modeling the short time behavior of heat flows (a solution to the paradox of the infinite heat propagation) [1, 2, 3]. Difficulties in solving this system lies in the stiff source term, $-\frac{p}{T_r}$, on the right hand side of the second equation. Because T_r is typically an extremely small quantity, an explicit time step, $\Delta t = O(T_r) \to 0$, is prohibitively restrictive. But an implicit treatment of the stiff source term could degrade the solution accuracy unless it is strongly coupled with the flux computation [7, 30, 31]. The same difficulties are shared with other physical models of interest, such as rarefied gas dynamics or radiation hydrodynamics. Hence, numerical methods for solving these relaxation systems have been extensively studied [1, 4, 5, 6], with a particular focus on the same stiffness problems. But the stiffness is not an issue in our case because the system is equivalent to the diffusion equation for any T_r at the steady state, and the steady state solution is exactly what we are interested in. This makes the development of numerical schemes a lot easier than the relaxation methods.

It is interesting to note that the removal of the stiffness comes at the expense of correct transient behavior. This is similar to the local preconditioning technique [32, 33, 34, 35]. In this technique, by altering the transient property of the time-dependent system (losing time accuracy), one attempts to optimize the condition number (the ratio of the maximum to the minimum wave speeds) in order to maximize the effect of error propagation thereby accelerating the convergence toward the steady state. The stiffness here is caused by a large condition number, and this is made to close to 1 as much as possible by multiplying the spatial part of the time-dependent system by a preconditioning matrix. In fact, the first-order system (2.25) can be interpreted as a preconditioned system of the hyperbolic heat equations. Suppose we have the hyperbolic heat equations with the relaxation time $\epsilon \ll 1$,

$$\begin{pmatrix} u \\ p \end{pmatrix}_t = \begin{pmatrix} 0 & \nu \\ 1/\epsilon & 0 \end{pmatrix} \begin{pmatrix} u \\ p \end{pmatrix}_x - \begin{pmatrix} 0 \\ p/\epsilon \end{pmatrix}. \tag{2.26}$$

This is a physically correct time-dependent system. Now, it is easy to see that multiplying the right hand side by the following preconditioning matrix:

$$\begin{pmatrix} 1 & 0 \\ 0 & \epsilon/T_r \end{pmatrix}, \tag{2.27}$$

where T_r is a free parameter, we obtain the first-order system (2.25). In effect, the preconditioning matrix replaces the relaxation time ϵ by a free parameter T_r. The system no longer describes a physically correct evolution of heat flows, but it is not stiff any more and still yields a correct solution at the steady state. Although the meaning of stiffness is slightly different, in both cases, the key idea is that we remove 'stiffness' by discarding correct time-dependent behavior.

It is important to note that although analytically the steady state solution does not depend on ν, the transient solution depends on it. But numerically, the dependency on ν can be eliminated by a suitable definition of the time step. In fact, for scalar schemes directly solving the diffusion equation, such as the Galerkin scheme and the distribution scheme based on the gradient reconstruction, a time integration with time step $\Delta t \propto \frac{1}{\nu}$ will cancel the effect of ν, and the convergence toward the steady state will be independent of ν. Or simply but equivalently, it is always possible in the diffusion equation to eliminate ν by a suitable time scaling. This is a natural and desirable property for steady state computations. In the case of the first-order system, the same can be true if the entire right hand side is proportional to ν. This is possible by setting $T_r \propto \frac{1}{\nu}$, and therefore we set

$$T_r = \frac{L_r^2}{\nu}, \tag{2.28}$$

where the length scale L_r has been introduced for the sake of dimensional consistency. Then, in view of the relaxation approach [1], the solution to the first-order system tends to stay in the frozen limit, i.e., obey the hyperbolic system rather than the diffusion equation for small ν. For large ν, the relaxation time T_r becomes small, but in this case the solution should reach the steady state quickly anyway. This seems to indicate that the relaxation time is adjusted so as to keep the system strongly hyperbolic toward the steady state for arbitrary ν.

As for the value of L_r, we may simply take $L_r = 1$ so that the system becomes symmetric:

$$\begin{pmatrix} u \\ p \end{pmatrix}_t = \begin{pmatrix} 0 & \nu \\ \nu & 0 \end{pmatrix} \begin{pmatrix} u \\ p \end{pmatrix}_x - \begin{pmatrix} 0 \\ \nu p \end{pmatrix}. \tag{2.29}$$

This is a good choice, but certainly may not be the best. We shall see later that the best value of L_r depends on the type of the scheme and also on the purpose for which the scheme is employed.

Note that the equations we are trying to solve should now be completely hyperbolic. But we expect that the solution is smooth because it will satisfy the diffusion equation eventually at the steady state. This can be a great advantage because all techniques developed for hyperbolic problems can be applied without any special mechanisms to capture discontinuities (of course such a mechanism may help when an initial solution contains some irregularity). In other words, we can only focus on the accuracy rather than other qualitative properties such as monotonicity. It may seem, by the way, that the isotropic nature of diffusion seems to have disappeared, but as we shall see later it remains in the disguise of a set of waves traveling isotropically.

We are now ready to develop numerical schemes for the first-order diffusion system. We continue to focus on the residual-distribution method in the rest of the paper, but the first-order system approach can equally apply to other methods. In one dimension, this can be clearly seen in the finite-difference formula arising from the new diffusion schemes we present in the next section.

3 New Diffusion Schemes in One Dimension

In this section, we design a class of residual-distribution schemes for one-dimensional diffusion problems based on the equivalent first-order system. In the first subsection, we define the one-dimensional first-order diffusion system and discuss the property of the system. In the second subsection, we develop a class of residual-distribution schemes for the first-order system. In particular, we will discover that the Galerkin scheme turns out to be a special case of the proposed scheme. In the third subsection, we show that some of the schemes allow $O(h)$ time step for explicit time integration toward the steady state. In the fourth subsection, Fourier analysis follows where L_r is defined to minimize the damping factor of the scheme, and this completes the design of the new schemes. Then, in the following subsection, we show from a truncation error analysis that the scheme is second-order accurate for all variables.

3.1 First-Order Diffusion System

We consider the one-dimensional diffusion problem:

$$u_t = \nu\, u_{xx} \quad \text{in } \Omega = [0,1], \tag{3.1}$$

where $\nu > 0$, and both $u(0)$ and $u(1)$ are given as boundary conditions. Our interest is to obtain the steady state solution to this problem. We then consider solving the following first-order system:

$$\begin{array}{rcl} u_t &=& \nu\, p_x, \\ p_t &=& (u_x - p)/T_r, \end{array} \tag{3.2}$$

or written in the vector form,

$$\mathbf{U}_t + \mathbf{A}\mathbf{U}_x = \mathbf{Q}, \tag{3.3}$$

where

$$\mathbf{U} = [u, p]^t, \quad \mathbf{A} = \begin{bmatrix} 0 & -\nu \\ -1/T_r & 0 \end{bmatrix}, \quad \mathbf{Q} = [0, -p/T_r]^t, \tag{3.4}$$

with $T_r = \frac{{L_r}^2}{\nu}$. It should be remembered that this system is equivalent to the original equation only in the steady state. In fact, the solution behaves very differently in the transient phase. In particular, we find that the eigenvalues of the matrix $\mathbf{A}$ are $\pm\sqrt{\nu/T_r}$ which are real (and called 'frozen speed' in the relaxation system [1]), and therefore we see that the first-order system has an advective character that is not at all present in the original diffusion problem. Indeed, the matrix $\mathbf{A}$ is diagonalizable with the matrix of the right eigenvectors $\mathbf{R}$,

$$\mathbf{R} = \begin{bmatrix} -L_r & L_r \\ 1 & 1 \end{bmatrix} \tag{3.5}$$

as

$$\mathbf{R}^{-1}\mathbf{A}\mathbf{R} = \mathbf{\Lambda} = \begin{bmatrix} \sqrt{\nu/T_r} & 0 \\ 0 & -\sqrt{\nu/T_r} \end{bmatrix}. \tag{3.6}$$

The view has now been totally switched from diffusion to advection, and hence the type of schemes we need are advection schemes rather than central-difference schemes that are generally considered suitable for diffusion. But this does not mean that the isotropic nature of the diffusion equation is totally lost. It manifests itself as a pair of two waves traveling in the opposite directions at the same speed, which is isotropic as a whole.

3.2 Discretization

For simplicity, but without loss of generality, we consider a uniform grid over a domain of interest with the mesh size $h = x_{j+1} - x_j,\ \forall j \in \{J\}$. We store the solution as well as the gradient at each node $(u_j, p_j),\ j \in \{J\}$, and then, with two boundary conditions available for u only, the task is to compute the steady state solution $\{u_j\}$ at the interior nodes and $\{p_j\}$ at all nodes. Note that the number of unknowns is now exactly equal to the number of cell-residuals. If there are N_c cells, we have $2N_c$ cell-residuals, and $2(N_c+1)$ unknowns. But because of the two boundary conditions (whether Dirichlet or Neumann), the actual number of unknowns is $2(N_c+1) - 2 = 2N_c$, i.e., the same as the number of cell-residuals. This means that *all the cell-residuals can be driven to zero exactly at the steady state*, implying the existence of a unique solution for linear problems. This is not possible for scalar schemes which distribute a single cell-residual for νu_{xx} evaluated with reconstructed nodal gradients. This is because in that case we have N_c cell-residuals for $(N_c+1) - 2 = N_c - 1$ degrees of freedom, i.e., always overdetermined.

We begin by evaluating the cell-residual, which is now a vector quantity, over cell $C = [x_j, x_{j+1}]$ as

$$\Phi^C = \int_{x_j}^{x_{j+1}} (-\mathbf{A}\mathbf{U}_x + \mathbf{Q})\, dx. \tag{3.7}$$

Assuming the piecewise linear variation of $\mathbf{U}$ over the cell, we obtain

$$\Phi^C = -\mathbf{A}\Delta\mathbf{U}_C + \overline{\mathbf{Q}}_C h, \tag{3.8}$$

where $\Delta\mathbf{U}_C = \mathbf{U}_{j+1} - \mathbf{U}_j$ and $\overline{\mathbf{Q}}_C = (\mathbf{Q}_{j+1} + \mathbf{Q}_j)/2$. We then distribute this to the nodes, by a distribution coefficient which is now a matrix $\mathcal{B}_j^C$ giving a fraction of Φ^C distributed to node j,

$$\Phi_j^C = \mathcal{B}_j^C \Phi^C, \quad \Phi_{j+1}^C = \mathcal{B}_{j+1}^C \Phi^C, \tag{3.9}$$

where for conservation we must have

$$\mathcal{B}_j^C + \mathcal{B}_{j+1}^C = \mathbf{I}, \quad \mathbf{I} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}. \tag{3.10}$$

The precise form of $\mathcal{B}_j^C$ is left open for a moment. Having done the distribution for all cells, we have the following semi-discrete equation at each node:

$$\frac{d\mathbf{U}_j}{dt} = \frac{1}{h_j}\left[\Phi_j^L + \Phi_j^R\right] = \frac{1}{h_j}\left[\mathcal{B}_j^L \Phi^L + \mathcal{B}_j^R \Phi^R\right], \tag{3.11}$$

where L and R denote the cells on the left and right of node j respectively, and h_j is the measure of the dual control volume centered at x_j which is identical to the mesh size h for uniform grids. We then integrate this in time until we reach the steady state. Note that we can use this scheme directly on non-uniform grids, simply by replacing the mesh size h by the variable mesh size h^C in the definition of the distribution matrices and setting $h_j = (h^L + h^R)/2$.

We now define the distribution matrix $\mathcal{B}_j^C$. The distribution matrix must be defined to reflect the physics of the governing equation: isotropic for diffusion or upwind for advection. In our case, the equations we are solving is not the diffusion equation anymore, but the equivalent first-order system which is hyperbolic with the wave speeds $\pm\sqrt{\nu/T_r}$. We expect also that the solution is smooth because it finally becomes the solution of the diffusion equation, and therefore there is no need to incorporate discontinuity-capturing mechanisms in the scheme. Then, for simplicity, we employ the Lax-Wendroff distribution scheme, also known as Ni's scheme in the context of residual-distribution [36], which is second-order accurate for smooth solutions. The scheme can be derived as follows. Consider the time expansion of the solution

$$\mathbf{U}_j^{n+1} \approx \mathbf{U}_j^n + \Delta t \mathbf{U}_t + \frac{1}{2}\Delta t^2 \mathbf{U}_{tt} = \mathbf{U}_j^n + \Delta t \left(1 + \frac{\Delta t}{2}\partial t\right)\mathbf{U}_t. \tag{3.12}$$

By using the equation itself, but partially ignoring the effect of the source term for simplicity, we can write

$$\mathbf{U}_j^{n+1} \approx \mathbf{U}_j^n + \Delta t \left(1 - \frac{\Delta t}{2}\mathbf{A}\partial x\right)\left(-\mathbf{A}\mathbf{U}_x + \mathbf{Q}\right), \tag{3.13}$$

which is approximated as

$$\begin{aligned} \mathbf{U}_j^{n+1} &\approx \mathbf{U}_j^n + \Delta t \left[\frac{1}{2}\left(\frac{\Phi_j^L}{h} + \frac{\Phi_j^R}{h}\right) - \frac{\Delta t}{2}\mathbf{A}\left(\frac{\Phi_j^R/h - \Phi_j^L/h}{h}\right)\right] && (3.14) \\ &= \mathbf{U}_j^n + \frac{\Delta t}{h}\left[\left(\frac{1}{2} + \frac{\Delta t}{2h}\mathbf{A}\right)\Phi_j^L + \left(\frac{1}{2} - \frac{\Delta t}{2h}\mathbf{A}\right)\Phi_j^R\right]. && (3.15) \end{aligned}$$

This implies that the distribution matrix is defined as

$$\mathcal{B}_j^C = \frac{1}{2}\mathbf{I} - \frac{\tau}{2h}\mathbf{A}, \quad \mathcal{B}_{j+1}^C = \frac{1}{2}\mathbf{I} + \frac{\tau}{2h}\mathbf{A}, \tag{3.16}$$

where Δt has been replaced by a time-like parameter τ which does not have to be equal to the actual time step because we are only interested in the steady state. Even if we take τ to be the actual time step, the scheme will not be time accurate because we have ignored the effect of the source term in the above derivation. Moreover, it is even pointless to develop time accurate schemes for the first-order system because it is not equivalent to the diffusion equation for time dependent problems unless $T_r \to 0$.

We point out that the scheme can be interpreted as a sum of the central distribution and a least-squares minimization term. The dissipation term can be derived by minimizing the residual in the least-squares norm, e.g. following the least-squares finite-element method [20] or based on a discrete minimization formulation [37]. In [26], this type of approach was used to derive a stabilization term in the residual-distribution schemes.

The parameter τ can be thought of as a cell time step, and the scheme will be conservative as long as it is constant over the cell. The simplest choice would then be the ratio of the mesh size h to the wave speed $\sqrt{\nu/T_r}$, giving

$$\tau = k_C \frac{h}{\sqrt{\nu/T_r}}, \tag{3.17}$$

where k_C is a cell CFL number which is taken to be 1 to maximize the effect of error propagation over the cell.

We remark also that in the previous work [18, 27], it is argued that diffusion is an isotropic process and therefore it is natural to distribute the residual with equal weights,

$$\mathcal{B}_j^C = \mathcal{B}_{j+1}^C = \begin{bmatrix} 1/2 & 0 \\ 0 & 1/2 \end{bmatrix}. \tag{3.18}$$

But in practice this scheme is not dissipative enough to damp high frequency errors, and in particular the highest frequency error cannot be damped at all [13]. The proposed scheme overcomes this problem by having a dissipation term added to the isotropic distribution coefficient. Note however that this is not by design but rather a natural consequence of solving the first-order system instead of the diffusion equation. The isotropic nature of diffusion is automatically incorporated by way of applying a suitable advection scheme, which typically comes with some form of dissipation, for the first-order system that is hyperbolic and whose waves travel isotropically. In fact, the proposed scheme can be shown to be an upwind scheme. To see this, consider the distribution matrices (3.16) with $\tau = \frac{h}{\sqrt{\nu/T_r}}$,

$$\mathcal{B}_j^C = \frac{1}{2}\mathbf{I} - \frac{1}{2\sqrt{\nu/T_r}}\mathbf{A}, \;\; \mathcal{B}_{j+1}^C = \frac{1}{2}\mathbf{I} + \frac{1}{2\sqrt{\nu/T_r}}\mathbf{A}. \tag{3.19}$$

Since $\mathbf{A}$ can be diagonalized, we have

$$\mathcal{B}_j^C = \frac{1}{2}\mathbf{I} - \frac{1}{2\sqrt{\nu/T_r}}\,\mathbf{R}\begin{bmatrix} \sqrt{\nu/T_r} & 0 \\ 0 & -\sqrt{\nu/T_r} \end{bmatrix}\mathbf{R}^{-1} = \mathbf{R}\begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}\mathbf{R}^{-1}, \tag{3.20}$$

$$\mathcal{B}_{j+1}^C = \frac{1}{2}\mathbf{I} + \frac{1}{2\sqrt{\nu/T_r}}\,\mathbf{R}\begin{bmatrix} \sqrt{\nu/T_r} & 0 \\ 0 & -\sqrt{\nu/T_r} \end{bmatrix}\mathbf{R}^{-1} = \mathbf{R}\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}\mathbf{R}^{-1}, \tag{3.21}$$

which shows that the solution mode with the negative wave speed is distributed to the left; the mode associated with the positive wave speed is distributed to the right. This is nothing but upwinding. An interesting observation is that the choice $\tau = \frac{h}{\sqrt{\nu/T_r}}$ makes the distribution matrix singular, creating a nullspace that implies one-sided distribution, i.e., upwind. This interpretation applies also in higher dimensions and may be used to check if a given scheme has an upwinding character.

It should be noted however that this is a rather special case where the Lax-Wendroff scheme and the upwind scheme coincide to each other. This is because the eigenvalues of the matrix $\mathbf{A}$ are of the equal magnitude with opposite signs, i.e., equal modulus. In general, an upwind scheme for a system is constructed by defining τ as a matrix such as

$$\tau = h\,|\mathbf{A}|^{-1}. \tag{3.22}$$

If all eigenvalues are equal in magnitude, this reduces to a scalar. This is exactly the case for the first-order diffusion system.

We now show that the new residual-distribution scheme is closely related to the Galerkin scheme. Expand the right hand side of the semi-discrete equation (3.11) with (3.16) for arbitrary τ to get

$$\begin{aligned} \frac{du_j}{dt} &= \frac{1}{2h}(\nu\Delta p_L + \nu\Delta p_R) + \frac{\tau\nu}{2h^2T_r}\left[\Delta u_R - \overline{p}_R h - (\Delta u_L - \overline{p}_L h)\right] \\ &= \nu\left[\left(1 - \frac{\tau}{2T_r}\right)\frac{p_{j+1} - p_{j-1}}{2h} + \frac{\tau}{2T_r}\frac{u_{j+1} - 2u_j + u_{j-1}}{h^2}\right], \end{aligned} \tag{3.23}$$

$$\begin{aligned} \frac{dp_j}{dt} &= \frac{1}{2hT_r}\left[(\Delta u_L - \overline{p}_L h) + (\Delta u_R - \overline{p}_R h)\right] + \frac{\tau}{2h^2T_r}\left[\nu\Delta p_R - \nu\Delta p_L\right] \\ &= \frac{1}{T_r}\left[\frac{1}{2h}\left(u_{j+1} - u_{j-1}\right) - \frac{1}{2}(\overline{p}_L + \overline{p}_R)\right] + \frac{\tau\nu}{2T_r}\frac{p_{j+1} - 2p_j + p_{j-1}}{h^2}, \end{aligned} \tag{3.24}$$

where $\Delta p_L = p_j - p_{j-1}$, $\Delta p_R = p_{j+1} - p_j$, $\overline{p}_L = (p_j + p_{j-1})/2$, and $\overline{p}_R = (p_{j+1} + p_j)/2$; similarly for u. It is then immediate that the choice

$$\tau = 2T_r \tag{3.25}$$

decouples the variables and yields the Galerkin scheme for u_j. Hence, the Galerkin scheme emerges as a special case of our residual distribution scheme. The cell-residual for the Galerkin scheme turns out to be associated not with the original diffusion equation but with the first-order system. Therefore, implemented this way, the Galerkin scheme has the residual property: if the cell-residual Φ^T for the first-order system vanishes, no updates will be sent to the nodes. Because of the decoupling, it is possible to solve for u_j first, and then compute p_j, which means that this scheme is simply the Galerkin scheme for u_j combined with implicit reconstruction or compact differentiation.

Also note from (3.23) and (3.24) that the proposed scheme can be implemented as a three-point finite-difference scheme or even a finite-volume scheme whose interface flux can easily be identified. But as mentioned in Section 2.1, in one dimension, the finite-volume schemes and the residual-distribution schemes are identical except for the treatment of source terms: the finite-volume scheme typically evaluates the source term directly by the cell average while the residual-distribution scheme evaluates the source term by the trapezoidal rule on each cell and weights them by the distribution coefficients. Then, the residual-distribution scheme has the residual property whereas the finite-volume scheme does not. This limits the accuracy of the finite-volume scheme to first-order. To improve the accuracy, methods to ensure the residual property for finite-volume schemes [24, 25] must be employed. In the case of the first-order diffusion system, the source term is inevitable, and therefore the finite-volume scheme will be first-order accurate unless the source term is discretized so as to have the residual property. The scheme above is certainly one of those having this property.

In the rest of the paper, we focus on two choices of τ: $\frac{h}{\sqrt{\nu/T_r}}$ and $2T_r$. The latter implements the Galerkin scheme as a residual-distribution scheme, and may be preferred in some cases. But we will show next that the former has a great advantage over the latter particularly for steady calculations.

3.3 $O(h)$ Time Step

To reach the steady state, we integrate the semi-discrete equation (3.11) in time until the solution stops changing. Any time integration scheme can be employed for this purpose. In any case, the time step is restricted by the maximum modulus of the eigenvalues of the coefficient matrix $\mathbf{C}_j$ of the scheme written in the following form:

$$\frac{d\mathbf{U}_j}{dt} = \mathbf{C}_{j-1}\mathbf{U}_{j-1} + \mathbf{C}_j\mathbf{U}_j + \mathbf{C}_{j+1}\mathbf{U}_{j+1}. \tag{3.26}$$

By expanding the right hand side of (3.11) with (3.16), we find

$$\mathbf{C}_j = \begin{bmatrix} -\dfrac{\tau\nu}{h^2 Tr} & 0 \\ 0 & -\dfrac{\tau\nu/h^2 + 1/2}{T_r} \end{bmatrix}. \tag{3.27}$$

Clearly, the maximum modulus of the eigenvalues is $\frac{\tau\nu/h^2+1/2}{T_r}$. Then, for example, in the case of the forward Euler time integration, the time step Δt is restricted by

$$\Delta t \le \frac{T_r}{\tau\nu/h^2 + 1/2}. \tag{3.28}$$

For the purpose of converging to the steady state, we simply take it as an equality to maximize the time step. For small h, this is approximately

$$\Delta t \le \frac{h^2 T_r}{\tau\nu}, \tag{3.29}$$

and for $\tau = 2T_r$, this will give the well-known severe stability limit for the Galerkin scheme,

$$\Delta t \le \frac{h^2}{2\nu}. \tag{3.30}$$

On the other hand, for the choice $\tau = \frac{h}{\sqrt{\nu/T_r}}$ with $T_r = \frac{{L_r}^2}{\nu}$, we obtain

$$\Delta t \leq \frac{h}{\sqrt{\nu/T_r}} = \frac{hL_r}{\nu}\,. \tag{3.31}$$

This is remarkable. *The time step is proportional to h instead of h^2.* This means that the number of time steps required to reach the steady state increase linearly with the mesh size. This is a great advantage over the conventional schemes. Of course, this is true only if $L_r = O(1)$. But we will see later that there is a case where L_r can be defined as such.

Finally, we point out that the condition (3.31) is nothing but the CFL condition for an advection equation with the advection speed $\sqrt{\nu/T_r}$. As a matter of fact, $O(h)$ time step is typical for advection schemes. This means that $O(h)$ time step is not something special to the residual-distribution schemes but rather special to the first-order system approach, and therefore we certainly can have it also for the finite-difference or the finite-volume schemes.

3.4 Fourier Analysis

Consider a Fourier mode of phase angle (or nondimensional wave number) $\beta \in [0, \pi]$,

$$\mathbf{U}^\beta = e^{i\beta x/h}\mathbf{U}_0, \tag{3.32}$$

where $\mathbf{U}^\beta = (u^\beta, p^\beta)$ and $\mathbf{U}_0 = (u_0, p_0)$. Inserting this into the original diffusion equation (3.1), we obtain

$$\frac{du^\beta}{dt} = \lambda_d u^\beta, \tag{3.33}$$

where

$$\lambda_d = -\frac{\nu}{h^2}\beta^2. \tag{3.34}$$

On the other hand, for the first-order system (3.2), we obtain

$$\frac{d\mathbf{U}^\beta}{dt} = \mathbf{M}_{\text{fos}}\mathbf{U}^\beta, \tag{3.35}$$

where

$$\mathbf{M}_{\text{fos}} = \begin{bmatrix} 0 & \nu\dfrac{i\beta}{h} \\ \dfrac{i\beta}{hT_r} & -\dfrac{1}{T_r} \end{bmatrix}. \tag{3.36}$$

The eigenvalues of this matrix are

$$\lambda_{\text{fos}} = -\frac{1}{2T_r}\left[1 \pm \sqrt{1 - \frac{4\nu T_r}{h^2}\beta^2}\,\right]. \tag{3.37}$$

For small β, we find

$$\lambda_{\text{fos}} = \begin{cases} -\dfrac{\nu}{h^2}\beta^2\left(1 + \dfrac{\nu T_r}{h^2}\beta^2\right) + O(\beta^6), \\ -\dfrac{1}{T_r} + \dfrac{\nu}{h^2}\beta^2 + \dfrac{\nu^2 T_r}{h^4}\beta^4 + O(\beta^6), \end{cases} \tag{3.38}$$

in which the first eigenvalue accurately represents the diffusion operator with second-order accuracy. This shows that the difference between the first-order system and the diffusion equation is of $O(\beta^2)$ for small β. Note that the eigenvalues can be complex. This happens when

$$\beta > \beta_{cr}, \quad \beta_{cr} = \frac{h}{2\sqrt{\nu T_r}}, \tag{3.39}$$

and the Fourier mode with $\beta > \beta_{cr}$ begins to propagate. Recall that we take $T_r = \frac{L_r^2}{\nu}$, then we have

$$\beta_{cr} = \frac{h}{2L_r}, \tag{3.40}$$

and so it is independent of ν.

For the Lax-Wendroff scheme, (3.11) with (3.16), we obtain the following equation:

$$\frac{d\mathbf{U}^\beta}{dt} = \mathbf{M}\mathbf{U}^\beta, \tag{3.41}$$

where

$$\mathbf{M} = \begin{bmatrix} -\dfrac{\tau\nu}{h^2T_r}(1-\cos\beta) & -\dfrac{i\nu}{2hT_r}(\tau-2T_r)\sin\beta \\ \dfrac{i\sin\beta}{hT_r} & -\dfrac{\tau\nu}{h^2T_r}(1-\cos\beta) - \dfrac{1}{2T_r}(1+\cos\beta) \end{bmatrix}. \tag{3.42}$$

The eigenvalues are

$$\lambda = -\frac{2\tau\nu}{h^2T_r}\sin^2\frac{\beta}{2} - \frac{1}{2T_r}\cos^2\frac{\beta}{2} \pm \frac{1}{2T_r}\sqrt{\cos^4\frac{\beta}{2} + \frac{2\nu}{h^2}(\tau-2T_r)\sin^2\beta}. \tag{3.43}$$

For small β, we find

$$\lambda = \begin{cases} -\dfrac{\nu}{h^2}\beta^2 + \dfrac{\nu}{12h^4T_r}\left[h^2T_r - 3\nu(\tau-2T_r)^2\right]\beta^4 + O(\beta^6), \\ -\dfrac{1}{T_r} + \dfrac{1}{T_r}\left[\dfrac{1}{4} - \dfrac{\nu(\tau-T_r)}{h^2}\right]\beta^2 + O(\beta^4). \end{cases} \tag{3.44}$$

which, compared with (3.38), confirms itself that the scheme is indeed second-order accurate for the first-order system, and consequently second-order accurate for the diffusion equation as well.

First we consider the case $\tau = 2T_r$. In this case, the eigenvalues simplify to

$$\lambda = -\frac{4\nu}{h^2}\sin^2\frac{\beta}{2}, \quad -\frac{4\nu}{h^2}\sin^2\frac{\beta}{2} - \frac{1}{T_r}\cos^2\frac{\beta}{2}, \tag{3.45}$$

which are always real and thus the errors are purely damped. The damping property of the scheme depends on the choice of L_r. Suppose that we employ the forward Euler time integration. Then, the eigenvalues g_1 and g_2 of the amplification matrix of the fully discrete equation $\mathbf{G} = \mathbf{I} + \Delta t\,\mathbf{M}_\Delta$ where $\Delta t = \frac{T_r}{\tau\nu/h^2+1/2}$ are given by

$$g_1 = 1 - \frac{8}{4+(h/L_r)^2}\sin^2\frac{\beta}{2}, \quad g_2 = \frac{4-(h/L_r)^2}{4+(h/L_r)^2}\cos^2\frac{\beta}{2}. \tag{3.46}$$

If we compare this with the point Jacobi iteration applied to solving $u_{xx} = 0$

$$u_j^{n+1} = u_j^n + \frac{\omega}{2}\left(u_{j+1}^n - 2u_j^n + u_{j-1}^n\right), \tag{3.47}$$

where ω is a relaxation factor $0 \le \omega \le 1$, whose amplification factor is given by [38]

$$1 - 2\omega\sin^2\frac{\beta}{2}, \tag{3.48}$$

we immediately find

$$\omega = \frac{1}{1+\frac{1}{4}(h/L_r)^2}, \tag{3.49}$$

which we write, introducing $L_r = \frac{h}{2k}$,

$$\omega = \frac{1}{1+k^2}. \tag{3.50}$$

It is well known that $\omega = \frac{2}{3}$ gives the optimal damping for high frequency errors ($\pi/2 \le \beta \le \pi$) and makes the scheme an effective smoother for multigrid [38]. This is achieved in our scheme by taking $k = 1/\sqrt{2}$, giving

$$L_r = \frac{h}{\sqrt{2}}. \tag{3.51}$$

In this case, $|g_1| \le \frac{1}{3}$ is guaranteed for $\pi/2 \le \beta \le \pi$, and it is clear from (3.46) that we have also $|g_2| \le \frac{1}{3}$ for the entire frequency. Therefore, the scheme is a good smoother not only for u_j but also for the other variable. However, if the scheme is used simply to iterate toward the steady state, this is not optimal. We should use the largest possible relaxation factor which corresponds to $k \to 0$. Practically, we may take any small number such as $k = 0.01$. But as we shall see later, if k is too small, we encounter an accuracy problem: the scheme reduces to first-order accurate for p_j. Experimentally, we found that $k = 0.2$ would not suffer from this problem:

$$L_r = \frac{h}{0.4}. \tag{3.52}$$

This means that this scheme is not well suited for iterating toward the steady state.

Now, we consider the case $\tau = \frac{h}{\sqrt{\nu/T_r}}$. In this case, the eigenvalues can be complex, and it is better to be complex. If complex, the eigenvalues are complex conjugates, thus having the same damping factor and propagation speed. There is no possibility that either u_j or p_j will converge much quicker than the other. Also, the damping is much more effective in the complex branch than the real branch that approximates the diffusion operator for low frequency modes. This can be seen in Figure 5 in which the real part of the eigenvalues are plotted against the phase angle. For all schemes and the equations, the eigenvalues are real for low frequency modes and make a second-order contact with the eigenvalue of the exact diffusion operator. This part, being closer to 0 than the complex branch in general, is a reason for slow convergence and we wish to avoid it. We will therefore choose L_r such that the eigenvalues are complex for all discrete error modes ($\beta \ge \pi h$). For $\tau = \frac{h}{\sqrt{\nu/T_r}} = \frac{hL_r}{\nu}$, the expression inside the square root in (3.43) is quadratic in L_r. It is easy to show that this is negative if

$$L_r \ge \frac{h}{4}\left(1 + \frac{1}{\sin\frac{\pi h}{2}}\right), \tag{3.53}$$

where we have set $\beta = \pi h$ to ensure that we have complex eigenvalues for all possible discrete error modes. In fact, in Figure 5, the lowest discrete mode ($\beta = \pi h$) is indicated by the vertical line, and we see that it passes through the branch point as designed. Note that this L_r is not $O(h)$ but $O(1)$ because

$$L_r \ge \frac{h}{4}\left(1 + \frac{1}{\sin\frac{\pi h}{2}}\right) \approx \frac{1}{2\pi} + \frac{h}{4} + O(h^2), \tag{3.54}$$

and so, as we claimed earlier, $O(h)$ time step is guaranteed for L_r that satisfies the condition (3.53). An optimal value of L_r can be derived by minimizing the amplification factor for the fastest convergence. For the forward Euler time integration, the eigenvalues of the amplification matrix of the fully discrete equation are complex conjugates whose magnitude $|g|$ is given by

$$|g|^2 = \frac{[2(L_r/h) + \cos\beta]\,[2(L_r/h) - 1]}{[2(L_r/h) + 1]^2}. \tag{3.55}$$

Let $L_r = \frac{h}{4}\left(1 + \frac{1}{\sin\frac{\pi h}{2}}\right)K$ and $K \ge 1$, then we find for small h,

$$|g| = 1 - \frac{\pi}{K}(3 - \cos\beta)h + O(h^2). \tag{3.56}$$

It is then obvious that $K = 1$ gives the minimum and therefore we set

$$L_r = \frac{h}{4}\left(1 + \frac{1}{\sin\frac{\pi h}{2}}\right), \tag{3.57}$$

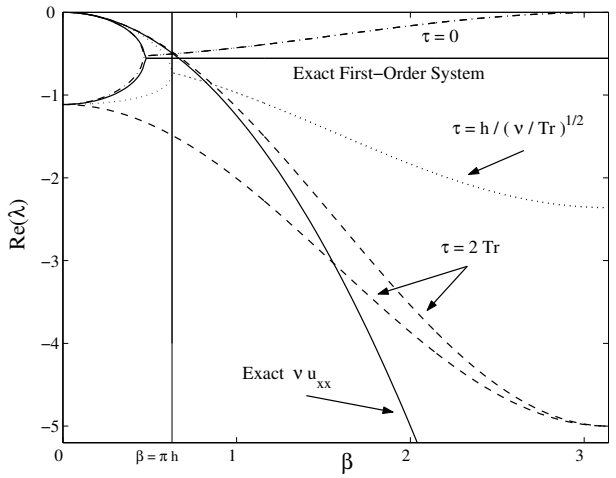

Figure 5: $Re(\lambda)$ for $h = 0.2$ and $\nu = 0.05$. $T_r = \frac{{L_r}^2}{\nu}$ with, for a comparison purpose, $L_r = \frac{h}{4}\left(1 + \frac{1}{\sin\frac{\pi h}{2}}\right)$ for all. $Re(\lambda)$ for $\tau = 0$ eventually becomes 0 at $\beta = \pi$.

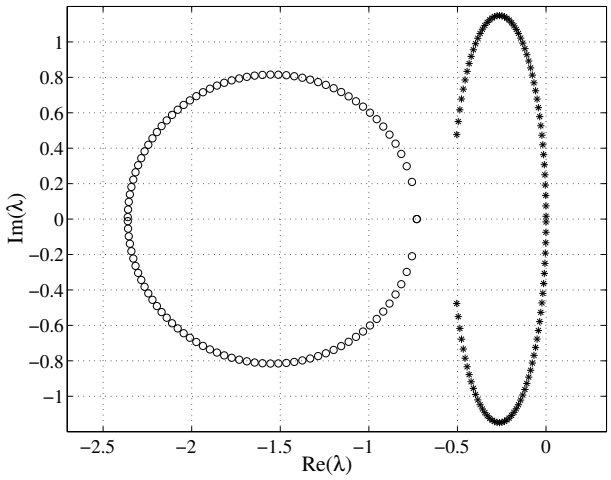

Figure 6: Polar plots of the eigenvalues for $\tau = 0$ (stars) and for $\tau = \frac{h}{\sqrt{\nu/T_r}}$ with the optimal L_r (circles). The eigenvalues were sampled from the range $\pi h \le \beta \le \pi$ with $h = 0.2$ and $\nu = 0.05$.

or we can use the following simple approximation:

$$L_r = \frac{1}{6} + \frac{h}{4}, \tag{3.58}$$

which satisfies the condition (3.53) for $h < \frac{1}{3}$.

Taking advantage of the propagation as an additional means to remove the error, this scheme takes a full advantage of the hyperbolic character of the first-order diffusion system, and it is therefore well suited for iterating toward the steady state. The polar plot of the eigenvalues of this scheme and the purely isotropic scheme is shown in Figure 6. Both schemes allow the error modes to propagate, but the one with nonzero τ (the upwind scheme) has much better damping.

3.5 Truncation Error

Expand smooth functions u and p around node j, and substitute into (3.11) with (3.16) to obtain

$$\frac{d\mathbf{U}_j}{dt} = \left[\mathbf{I} - \frac{\tau}{2}\mathbf{A}\partial_x\right]\mathbf{r} + O(h^2), \tag{3.59}$$

where

$$\mathbf{r} = [\nu p_x, (u_x - p)/T_r]^t, \tag{3.60}$$

or component-wise

$$\frac{du_j}{dt} = \nu\, p_x + \frac{\tau\nu}{2T_r}(u_x - p)_x + O(h^2), \tag{3.61}$$

$$\frac{dp_j}{dt} = (u_x - p)/T_r + \frac{\tau\nu}{2T_r}\,(p_x)_x + O(h^2). \tag{3.62}$$

We remark that the scheme has the residual vector $\mathbf{r}$ as a factor in the truncation error, which vanishes at the steady state and second-order accuracy is obtained. This is a property shared with the residual-based compact scheme [39]. In a way, residual-distribution is an alternative form of implementing the compact schemes.

To get more insight, suppose that the smooth solutions are exact solutions to the discrete equations in the steady state ($\frac{du_j}{dt} = \frac{dp_j}{dt} = 0$). Then, they satisfy

$$0 = \nu\, p_x + \frac{\tau\nu}{2T_r}(u_x - p)_x + O(h^2), \tag{3.63}$$

$$0 = (u_x - p)/T_r + \frac{\tau\nu}{2T_r}\,(p_x)_x + O(h^2). \tag{3.64}$$

For $\tau = 2T_r$, we obtain

$$0 = \nu u_{xx} + O(h^2), \tag{3.65}$$

$$0 = (u_x - p)/T_r + \nu\,(p_x)_x + O(h^2), \tag{3.66}$$

which clearly shows that the solution u converges to the solution of the original diffusion equation with second-order accuracy. We write the second equation by expanding $T_r = \frac{L_r^2}{\nu}$ with $L_r = \frac{h}{2k}$,

$$0 = \frac{4\nu k^2}{h^2}(u_x - p) + \nu\,(p_x)_x + O(h^2). \tag{3.67}$$

For $k = O(1)$, this shows that the numerical solution converges to the solution of $u_x - p = 0$ with second-order accuracy. But if $k = O(h)$, the scheme is not consistent, solving a wrong equation. Also, as $k \to 0$, it converges to the solution of

$$0 = \nu\,(p_x)_x + O(h^2). \tag{3.68}$$

This shows that the scheme is not consistent, not solving $u_x - p = 0$ nor even $\nu p_x = 0$. But fortunately in one dimension, the scheme is in fact consistent but only first-order accurate. This is because for one-dimensional problems, not only the nodal residuals but also the cell-residuals which approximate νp_x vanish at the steady state. This ensures at a node that $p_x = O(h)$, thus the scheme is consistent and first-order accurate. This is the accuracy problem mentioned in the previous subsection.

On the other hand, for $\tau = \frac{h}{\sqrt{\nu/T_r}} = \frac{hL_r}{\nu}$, we obtain

$$0 = \nu\, p_x + \frac{\nu h}{2L_r}(u_x - p)_x + O(h^2), \tag{3.69}$$

$$0 = \frac{\nu}{L_r^2}(u_x - p) + \frac{\nu h}{2L_r}\,(p_x)_x + O(h^2). \tag{3.70}$$

For $L_r = O(1)$ which is the case of (3.57), this shows clearly that the numerical solution converges to the solution of the first-order system (3.2) as $h \to 0$. By eliminating the first-order terms using the equations themselves, we find

$$0 = \nu p_x - \frac{\nu h^2}{4}\,(p_x)_x + O(h^2), \tag{3.71}$$

$$0 = (u_x - p) - \frac{h^2}{4}(u_x - p)_x + O(h^2), \tag{3.72}$$

which shows that the solution converges with second-order accuracy. Finally, we point out that by setting $L_r = \frac{h}{2}$ we recover the Galerkin scheme which corresponds to $\tau = 2T_r$ with $L_r = \frac{h}{2}$, i.e., the two choices of τ are not independent of each other.

3.6 Boundary Conditions

As mentioned earlier, with two boundary conditions, the number of unknowns exactly matches the number of cell-residuals, and thus for a linear problem there exists a unique solution. The boundary conditions can be either the Dirichlet type where u_j is specified or the Neumann type where p_j is specified. In any case, only one value is specified on each boundary. This can be interpreted also as a characteristic condition. Since the first-order system is hyperbolic with two characteristics running to the left and the right, there is always one characteristic coming into the domain from through the boundary, and therefore we need to specify one value on the boundary.

4 New Diffusion Schemes in Two Dimensions

We now consider two-dimensional problems, and develop again a class of residual-distribution schemes for the two-dimensional first-order diffusion system. We shall see that the two-dimensional schemes share many of the remarkable properties of the one-dimensional schemes. But there is also a striking difference. Unlike the one-dimensional problem, all cell-residuals cannot be made to vanish in two dimensions because of a counting problem: the number of elements is not equal to the number of nodes. This brings a consistency problem in the case $L_r \to \infty$.

We remark that only triangular unstructured grids will be considered here. For structured grids, the one-dimensional scheme can be applied as a finite-difference scheme or a finite-volume scheme by decomposing the two-dimensional equation into dimension by dimension one-dimensional equations. This can be done in a straightforward manner (see [36, 40] for example). We point out also that a finite-volume scheme can be developed in a similar manner for unstructured grids by applying a one-dimensional flux function normal to the cell face. Again, it should be remembered that these schemes will be only first-order accurate unless the source term in the first-order system is discretized to guarantee the residual property.

4.1 First-Order Diffusion System

We consider the two-dimensional scalar diffusion problem,

$$u_t = \nu\,(u_{xx} + u_{yy}) \quad \text{in } \Omega, \tag{4.1}$$

where $\nu > 0$ and $u = g(x, y)$ is given as a boundary condition on $\partial\Omega$. Our interest is again to obtain the steady state solution of this problem. As in one dimension, we consider solving the equivalent first-order system instead,

$$\begin{array}{rcl} u_t &=& \nu\,(p_x + q_y), \\ p_t &=& (u_x - p)/T_r, \\ q_t &=& (u_y - q)/T_r, \end{array} \tag{4.2}$$

where $T_r = \frac{{L_r}^2}{\nu}$, or written in the vector form,

$$\mathbf{U}_t + \mathbf{A}\mathbf{U}_x + \mathbf{B}\mathbf{U}_y = \mathbf{Q}, \tag{4.3}$$

where

$$\mathbf{U} = [u, p, q]^t \quad , \qquad \mathbf{Q} = [0, -p/T_r, -q/T_r]^t, \tag{4.4}$$

$$\mathbf{A} = \begin{bmatrix} 0 & -\nu & 0 \\ -1/T_r & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad , \qquad \mathbf{B} = \begin{bmatrix} 0 & 0 & -\nu \\ 0 & 0 & 0 \\ -1/T_r & 0 & 0 \end{bmatrix}. \tag{4.5}$$

Again, this system is equivalent to the diffusion equation only in the steady state. In converging to the steady state, this system behaves like a hyperbolic system. In fact, the matrix $\mathbf{A}_n = \mathbf{A}n_x + \mathbf{B}n_y$ is diagonalizable for any chosen normal vector $\mathbf{n} = (n_x, n_y)$ with the following matrix of right eigenvectors $\mathbf{R}$:

$$\mathbf{R} = \begin{bmatrix} -L_r & L_r & 0 \\ n_x & n_x & -n_y \\ n_y & n_y & n_x \end{bmatrix}, \tag{4.6}$$

as

$$\mathbf{R}^{-1}\mathbf{A}_n\mathbf{R} = \mathbf{\Lambda} = \begin{bmatrix} \sqrt{\nu/T_r} & 0 & 0 \\ 0 & -\sqrt{\nu/T_r} & 0 \\ 0 & 0 & 0 \end{bmatrix}. \tag{4.7}$$

The eigenvalues of the matrix $\mathbf{A}_n$ are $\pm\sqrt{\nu/T_r}$ and 0. The vanishing eigenvalue is associated with the consistency constraint,

$$q_x - p_y = 0, \tag{4.8}$$

under which the system (4.2) needs to be solved. This states that an inconsistent mode, represented as a 'vorticity' of the gradient vector (p, q), must vanish at a steady state. It is easy to show that this satisfies

$$(q_x - p_y)_t = -\frac{1}{T_r}(q_x - p_y). \tag{4.9}$$

This is an ordinary differential equation, showing that the inconsistency is purely damped out at the time scale of T_r. This is what the stationary mode is responsible for: it damps out any inconsistency contained in an initial solution. For this reason, this may be called the inconsistency damping mode.

The other two eigenvalues represent a wave traveling isotropically (a wave speed independent of $\mathbf{n}$ implies a circular wave), giving an alternative description of isotropic diffusion. Hence, as in one dimension, we will consider advection schemes for the first-order diffusion system (4.3).

4.2 Discretization

To discretize the system, we divide the domain into a set of triangles $\{T\}$ and a set of vertices $\{V\}$, and store the solution at each vertex (u_j, p_j), $j \in \{V\}$. Now, the task is to compute the steady state solution $\{u_j\}$ at the interior nodes, and $\{p_j, q_j\}$ at all nodes except for the boundary nodes on which they can be computed from u given on the boundary. Note that this time the number of unknowns is much less than (typically a half of) the number of cell-residuals. Therefore, all cell-residuals cannot be driven to zero at the steady state. Only nodal residuals, which are weighted averages of cell-residuals, can be driven to zero, and these weights are determined by the distribution matrices.

We begin by defining the cell-residual over cell T,

$$\Phi^T = \int_T \left(-\mathbf{A}\mathbf{U}_x - \mathbf{B}\mathbf{U}_y + \mathbf{Q}\right) dxdy. \tag{4.10}$$

Assuming a piecewise linear variation of $\mathbf{U}$ over the cell, we obtain

$$\Phi^T = -\sum_{i=1}^{3} \mathbf{K}_i \mathbf{U}_i + \overline{\mathbf{Q}}_T S_T, \tag{4.11}$$

where

$$\mathbf{K}_i = \frac{1}{2}\left(\mathbf{A}, \mathbf{B}\right) \cdot \mathbf{n}_i, \quad \overline{\mathbf{Q}}_T = \frac{\mathbf{Q}_1 + \mathbf{Q}_2 + \mathbf{Q}_3}{3}, \tag{4.12}$$

and $\mathbf{n}_i$ is the inward scaled normal (see 4). We then distribute this to the nodes by a distribution matrix $\mathcal{B}_i^T$

$$\Phi_i^T = \mathcal{B}_i^T \Phi^T, \tag{4.13}$$

where

$$\Phi^T = \sum_{i=1}^{3} \Phi_i^T, \qquad \sum_{i=1}^{3} \mathcal{B}_i^T = \mathbf{I}, \tag{4.14}$$

and as a result we have the following semi-discrete equation at each node:

$$\frac{d\mathbf{U}_j}{dt} = \frac{1}{S_j} \sum_{T \in \{T_j\}} \mathcal{B}_j^T \Phi^T, \tag{4.15}$$

where S_j is the medial dual cell area (see 4). This is then integrated in time to reach the steady state.

To distribute the cell-residual, we employ the Lax-Wendroff distribution. Consider the time expansion of the solution

$$\mathbf{U}_j^{n+1} \approx \mathbf{U}_j^n + \Delta t \mathbf{U}_t + \frac{1}{2}\Delta t^2 \mathbf{U}_{tt} = \mathbf{U}_j^n + \Delta t \left(\mathbf{I} + \frac{\Delta t}{2}\partial t\right) \mathbf{U}_t. \tag{4.16}$$

By using the equation itself, but partially ignoring the effect of the source term, we obtain

$$\mathbf{U}_j^{n+1} \approx \mathbf{U}_j^n + \Delta t \left[\mathbf{I} - \frac{\Delta t}{2}\left(\mathbf{A}\partial x + \mathbf{B}\partial y\right)\right] \left(-\mathbf{A}\mathbf{U}_x - \mathbf{B}\mathbf{U}_y + \mathbf{Q}\right), \tag{4.17}$$

which can be integrated over the median dual control volume around j as was done in deriving the Galerkin scheme in Section 2.3, resulting

$$S_j \mathbf{U}_j^{n+1} \approx S_j \mathbf{U}_j^n + \Delta t \sum_{T \in \{T_j\}} \left[\frac{1}{3}\mathbf{I} + \frac{\Delta t}{4S_T} (\mathbf{A}, \mathbf{B}) \cdot \mathbf{n}_j^T \right] \Phi^T. \tag{4.18}$$

This implies that the distribution matrix is defined as

$$\mathcal{B}_i^T = \frac{1}{3}\mathbf{I} + \frac{\tau}{2S_T}\mathbf{K}_i, \tag{4.19}$$

which is again the sum of the central distribution and the least-squares dissipation. Here, as in one dimension, Δt has been replaced by τ and it is taken as a free parameter. Taking it as a time-like parameter in particular, we define τ by

$$\tau = k_T \frac{h_T}{\sqrt{\nu / T_r}} \quad h_T = \frac{2S_T}{\max_{i \in \{i_T\}} |\mathbf{n}_i|}, \tag{4.20}$$

where we set $k_T = 1$ to maximize the effect of error propagation. Note that the distribution matrices sum up to the identity matrix over the triangle T as long as τ is constant over the triangle (dissipation terms sum up to zero), and so the scheme is conservative.

In one dimension, the scheme derived this way happens to be upwind. But this is not the case in two dimensions. Recall that the upwind distribution matrix must be singular, implying the existence of a nullspace. It is easy to show that the matrix (4.19) is singular only if

$$\tau = \frac{4}{3} \frac{S_T}{|\mathbf{n}_i|} \sqrt{\frac{T_r}{\nu}}. \tag{4.21}$$

This shows that τ should not be constant but depend on the node for the scheme to be upwind. Therefore, the scheme with τ as in (4.20) can be made to be upwind by taking $k_T = \frac{4}{3}$, but this is true only for one particular node associated with the maximum height. In general, this scheme distributes the residual to all nodes. A full upwind scheme can be obtained by defining τ as a matrix defined by

$$\tau = \frac{2S_T}{3} |\mathbf{K}_i|^{-1}, \tag{4.22}$$

with which the distribution matrix (4.19) becomes

$$\mathcal{B}_i^T = \frac{1}{3}\mathbf{I} + \frac{1}{3} |\mathbf{K}_i|^{-1} \mathbf{K}_i = \frac{1}{3} \mathbf{R}_i \left(\mathbf{I} + |\Lambda_i|^{-1} \Lambda_i \right) \mathbf{R}_i^{-1} = \frac{1}{3} \mathbf{R}_i \left[\mathbf{I} + \text{sign}(\Lambda_i) \right] \mathbf{R}_i^{-1} \tag{4.23}$$

where $\text{sign}(\lambda_i)$ may be set to be zero for the null eigenvalue mode, so that the distribution becomes isotropic for that mode, exactly as is done by the Lax-Wendroff scheme (4.19). For the first-order diffusion system, this upwind matrix can be analytically obtained as follows,

$$\mathcal{B}_i^T = \frac{1}{3} \begin{bmatrix} 1 & -L_r n_i^x & -L_r n_i^y \\ -n_i^x / L_r & 1 & 0 \\ -n_i^y / L_r & 0 & 1 \end{bmatrix} \tag{4.24}$$

where $\mathbf{n}_i = (n_i^x, n_i^y)$. It follows immediately from this that

$$\sum_{i=1}^{3} \mathcal{B}_i^T = \mathbf{I} \tag{4.25}$$

because $\sum_{i=1}^{3} \mathbf{n}_i = 0$, and therefore the scheme is conservative. Note that the upwind scheme is not unique in two dimensions. This is just one example of upwind distribution schemes, and other upwind schemes can also

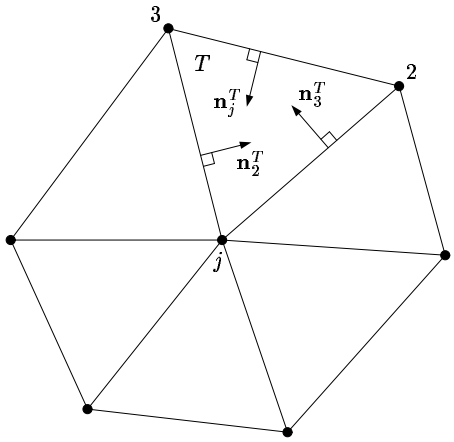

Figure 7: Normals for triangle $T \in \{T_j\}$.

be applied such as the matrix LDA scheme [41]. It should be noted however that the solution is smooth for diffusion problems and therefore the focus should rather be on the accuracy. Therefore, the scheme does not need to be upwind and the Lax-Wendroff distribution scheme is more than adequate in this case. For this reason, we here stick to the simple Lax-Wendroff scheme and do not explore other possibilities. We will have to discuss other possibilities when we consider the advection-diffusion problems for which upwinding can be very important.

We now show that the Galerkin discretization is obtained by taking $\tau = 2T_r$ exactly as in one dimension. Expand the first component of (4.15) with (4.19),

$$S_j \frac{du_j}{dt} = \sum_{T \in \{T_j\}} \left[\frac{\nu}{3}(p_x^T + q_y^T) S_T - \frac{\tau\nu}{4T_r} \left\{ \nabla u^T - (\overline{p}_T, \overline{q}_T) \right\} \cdot \mathbf{n}_j^T \right]. \tag{4.26}$$

Using the following identities:

$$(p_x^T + q_y^T) S_T = \frac{1}{2} \sum_{i=1}^{3} (p_i, q_i) \cdot \mathbf{n}_i^T, \tag{4.27}$$

$$(\overline{p}_T, \overline{q}_T) \cdot \mathbf{n}_j^T = \frac{1}{3} \sum_{i=1}^{3} (p_i, q_i) \cdot \mathbf{n}_j^T, \tag{4.28}$$

where we identify $i = 1$ as j, we obtain

$$S_j \frac{du_j}{dt} = -\frac{\tau\nu}{4T_r} \sum_{T \in \{T_j\}} \nabla u^T \cdot \mathbf{n}_j^T + \frac{\nu}{6} \sum_{T \in \{T_j\}} \sum_{i=1}^{3} (p_i, q_i) \cdot \left(\mathbf{n}_i^T + \frac{\tau}{2T_r} \mathbf{n}_j^T \right). \tag{4.29}$$

The first term is nothing but the Galerkin discretization, and the second term is a coupling term with p and q which can be simplified for $\tau = 2T_r$ as

$$\sum_{i=1}^{3} (p_i, q_i) \cdot \left(\mathbf{n}_i^T + \mathbf{n}_j^T \right) = 2\,(p_j, q_j) \cdot \mathbf{n}_j^T - (p_2, q_2) \cdot \mathbf{n}_3^T - (p_3, q_3) \cdot \mathbf{n}_2^T. \tag{4.30}$$

This all vanishes when summed over a set of triangles $\{T_j\}$ unless node j is on the boundary (see Figure 7), and therefore we are left with the Galerkin part,

$$S_j \frac{du_j}{dt} = -\frac{\nu}{2} \sum_{T \in \{T_j\}} \nabla u^T \cdot \mathbf{n}_j^T. \tag{4.31}$$

So, again, the Galerkin discretization arises as a special case of a residual-distribution scheme. It is a residual-distribution scheme with the residual defined for the first-order diffusion system. We remark that this is similar

to the situation that a least-squares residual-distribution scheme developed in [37] for the Cauchy-Riemann system turned out to be the Galerkin scheme for the associated Laplace's equations. In both cases, the Galerkin scheme arises from a residual-distribution scheme for the associated first-order system. Moreover, note that in both cases the Galerkin discretization comes from the least-squares minimization (the least-squares scheme itself or the least-squares dissipation term in the Lax-Wendroff scheme). In fact, the connection between the least-squares method for the first-order system and the Galerkin discretization for the associated second-order equation has already been pointed out by Jiang [20].

4.3 $O(h)$ Time Step

For time integration, we find a stability condition based on the eigenvalues of the coefficient matrix $\mathbf{C}_j$ for $\mathbf{U}_j$ of the scheme written in the following form:

$$\frac{d\mathbf{U}_j}{dt} = \sum_{i\in\{i_j\}} \mathbf{C}_i \mathbf{U}_i, \tag{4.32}$$

where the sum is over the nodes in the compact stencil: the node j and its immediate neighbors, denoted by $\{i_j\}$. By expanding the right hand side of (4.15) with (4.19), we find the coefficient matrix for $\mathbf{U}_j$ as

$$\mathbf{C}_j = \frac{1}{S_j T_r} \begin{bmatrix} -\sum_{T\in\{T_j\}} \frac{\tau\nu}{8S_T}|\mathbf{n}_j^T|^2 & 0 & 0 \\ 0 & -\sum_{T\in\{T_j\}} \left(\frac{S_T}{9} + \frac{\tau\nu}{8S_T}(n_j^x)^2\right) & -\sum_{T\in\{T_j\}} \frac{\tau\nu}{8S_T} n_j^x n_j^y \\ 0 & -\sum_{T\in\{T_j\}} \frac{\tau\nu}{8S_T} n_j^x n_j^y & -\sum_{T\in\{T_j\}} \left(\frac{S_T}{9} + \frac{\tau\nu}{8S_T}(n_j^y)^2\right) \end{bmatrix}. \tag{4.33}$$

The maximum modulus of the eigenvalues, which is relevant to the stability, is given by

$$|\lambda| = \frac{1}{S_j T_r} \left[\sum_{T\in\{T_j\}} \left[\frac{S_T}{9} + \frac{\tau\nu}{8S_T}|\mathbf{n}_j^T|^2\right] + \frac{\tau\nu}{16} \sqrt{\left(\sum_{T\in\{T_j\}} \frac{(n_j^x)^2 - (n_j^y)^2}{S_T}\right)^2 + 4\left(\sum_{T\in\{T_j\}} \frac{n_j^x n_j^y}{S_T}\right)^2} \right], \tag{4.34}$$

where $\mathbf{n}_j^T = (n_j^x, n_j^y)$. The time step Δt is then restricted locally by

$$\Delta t \le \frac{S_j}{|\lambda|}, \tag{4.35}$$

whose minimum over all nodes will give a global time step condition. For the purpose of converging to the steady state, we simply take it as an equality to maximize the time step,

$$\Delta t = \frac{S_j}{|\lambda|}. \tag{4.36}$$

For a practical purpose, the maximum modulus $|\lambda|$ in (4.34) can be simplified by the Cauchy-Schwarz inequality,

$$\left(\sum_{T\in\{T_j\}} a_T b_T\right)^2 \le \left(\sum_{T\in\{T_j\}} (a_T)^2\right)\left(\sum_{T\in\{T_j\}} (b_T)^2\right), \tag{4.37}$$

with $a_T = (n_j^x + n_j^y)/\sqrt{S_T}$ and $b_T = (n_j^x - n_j^y)/\sqrt{S_T}$, to

$$|\lambda|^* = \frac{1}{S_j T_r} \sum_{T\in\{T_j\}} \left[\frac{S_T}{9} + \frac{3\tau\nu}{16S_T}|\mathbf{n}_j^T|^2\right] \ge |\lambda|. \tag{4.38}$$

Then, we may take

$$\Delta t = \frac{S_j}{|\lambda|^*} \le \frac{S_j}{|\lambda|}. \tag{4.39}$$

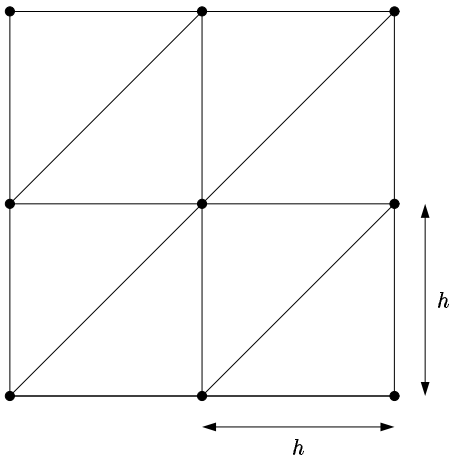

Figure 8: A regular triangular grid.

This is a more severe restriction but simpler to implement.

For a regular triangular grid (see Figure 8), the condition (4.35) simplifies to

$$\Delta t \leq \frac{3T_r}{1 + \dfrac{6\tau\nu}{h^2}}, \tag{4.40}$$

which is approximately, for small h,

$$\Delta t \leq \frac{h^2 T_r}{2\tau\nu} \tag{4.41}$$

and therefore, for $\tau = 2T_r$, this gives the two-dimensional version of the well-known severe stability limit for the Galerkin scheme,

$$\Delta t \leq \frac{h^2}{4\nu}. \tag{4.42}$$

On the other hand, $\tau = \frac{h_T}{\sqrt{\nu/T_r}}$ and $T_r = \frac{{L_r}^2}{\nu}$ gives (note that $h_T = h$ for these regular triangles by definition; see (4.20)),

$$\Delta t \leq \frac{hL_r}{2\nu}. \tag{4.43}$$

Hence, again, *the time step is proportional to h instead of* h^2, for $L_r = O(1)$. The remarkable property of the one-dimensional scheme carries over to two dimensions.

4.4 Fourier Analysis

Consider again a regular triangular grid (see Figure 8), and define a Fourier mode of phase angle $\beta = (\beta_x, \beta_y)$ with $\beta_x,\ \ \beta_y \in [0, \pi]$,

$$\mathbf{U}^\beta = e^{i(\beta_x x/h + \beta_y y/h)}\mathbf{U}_0. \tag{4.44}$$

Inserting this into the original diffusion equation (4.1), we get

$$\frac{du^\beta}{dt} = \lambda_d u^\beta, \tag{4.45}$$

where

$$\lambda_d = -\frac{\nu}{h^2}\beta^2, \tag{4.46}$$

which is identical to the one-dimensional counterpart. In the case of the first-order system (4.2), we obtain

$$\frac{d\mathbf{U}^\beta}{dt} = \mathbf{M}_{\text{fos}}\mathbf{U}^\beta, \tag{4.47}$$

where

$$\mathbf{M}_{\text{fos}} = \begin{bmatrix} 0 & \nu\dfrac{i\beta_x}{h} & \nu\dfrac{i\beta_y}{h} \\ \dfrac{i\beta_x}{hT_r} & -\dfrac{1}{T_r} & 0 \\ \dfrac{i\beta_y}{hT_r} & 0 & -\dfrac{1}{T_r} \end{bmatrix}. \tag{4.48}$$

The eigenvalues are

$$\lambda_{\text{fos}} = \begin{cases} -\dfrac{1}{2T_r}\left(1 \pm \sqrt{1 - \dfrac{4\nu T_r}{h^2}\beta^2}\right), \\ -\dfrac{1}{T_r}, \end{cases} \tag{4.49}$$

where the first two are exactly the same as those in one dimension and therefore we have exactly the same condition as in one dimension for these eigenvalues to be complex,

$$\beta > \beta_{cr}, \quad \beta_{cr} = \frac{h}{2\sqrt{\nu T_r}} = \frac{h}{2L_r}. \tag{4.50}$$

The third eigenvalue corresponds to the inconsistency damping mode.

On the other hand, for the Lax-Wendroff scheme, i.e., (4.15) with (4.19), we find

$$\frac{d\mathbf{U}^\beta}{dt} = \mathbf{M}\mathbf{U}^\beta, \tag{4.51}$$

where

$$\mathbf{M} = \frac{1}{T_r}\begin{bmatrix} -\dfrac{\tau\nu(c_x+c_y)}{h^2} & \dfrac{-i\nu(\tau-2T_r)S_{xy}}{6h} & \dfrac{-i\nu(\tau-2T_r)S_{yx}}{6h} \\ \dfrac{iS_{xy}}{3h} & -\dfrac{1}{3} - \dfrac{\tau\nu c_x}{h^2} - \dfrac{2}{9}C_p & \dfrac{\tau\nu C_m}{2h^2} \\ \dfrac{iS_{yx}}{3h} & \dfrac{\tau\nu C_m}{2h^2} & -\dfrac{1}{3} - \dfrac{\tau\nu c_y}{h^2} - \dfrac{2}{9}C_p \end{bmatrix}, \tag{4.52}$$

where

$$\begin{aligned} c_x &= 1-\cos(\beta_x), \\ c_y &= 1-\cos(\beta_y), \\ S_{xy} &= \sin(\beta_x+\beta_y) + 2\sin(\beta_x) - \sin(\beta_y), \\ S_{yx} &= \sin(\beta_x+\beta_y) + 2\sin(\beta_y) - \sin(\beta_x), \\ C_p &= \cos(\beta_x+\beta_y) + \cos(\beta_y) + \cos(\beta_x), \\ C_m &= \cos(\beta_x+\beta_y) - \cos(\beta_y) - \cos(\beta_x) + 1. \end{aligned}$$

First we consider the case $\tau = 2T_r$. In this case, the eigenvalue associated with the Galerkin discretization can be trivially found and is given by

$$\lambda_1 = -\frac{4\nu}{h^2}\left(\sin^2\frac{\beta_x}{2} + \sin^2\frac{\beta_y}{2}\right). \tag{4.53}$$

For the forward Euler time stepping, the amplification factor $g_1 = 1 + \Delta t\lambda_1$ with Δt defined in (4.40) is given by

$$g_1 = 1 - \frac{1}{1 + \frac{1}{12}(h/L_r)^2}\left(\sin^2\frac{\beta_x}{2} + \sin^2\frac{\beta_y}{2}\right). \tag{4.54}$$

Comparing this with the amplification factor of the point Jacobi iteration [38],

$$g = 1 - \omega \left(\sin^2 \frac{\beta_x}{2} + \sin^2 \frac{\beta_y}{2} \right), \tag{4.55}$$

we find

$$\omega = \frac{1}{1 + \frac{1}{3}k^2}, \tag{4.56}$$

where we have set $L_r = \frac{h}{2k}$. It is well known that $\omega = \frac{4}{5}$ gives the optimal smoothing factor for high frequency modes ($\frac{\pi}{2} \le \beta_x \le \pi$ or $\frac{\pi}{2} \le \beta_y \le \pi$) [38]. This is achieved in our scheme by taking $k = \frac{\sqrt{3}}{2}$, i.e.,

$$L_r = \frac{h}{\sqrt{3}}. \tag{4.57}$$

In this case, $|g_1| \le 0.6$ is guaranteed for high-frequency modes. Unfortunately, unlike the one-dimensional scheme, the amplification factors associated with the other two eigenvalues exceed 0.6 and hence the scheme is not entirely optimal. But the variables are completely decoupled in this case anyway, i.e., the gradient variables can be computed separately as a compact differentiation, and therefore here we do not even attempt to optimize the scheme for all solution modes. On the other hand, for the fastest convergence toward the steady state, we wish to take $\omega = 1$, which is possible by taking

$$k \to 0. \tag{4.58}$$

However, this causes a serious problem: the scheme will not be consistent for p_j and q_j. We discuss this problem in the next subsection. Here, we only say that this is not a suitable explicit scheme for the purpose of iterating toward the steady state. Therefore, we will not discuss this scheme further.

Next, we consider the case $\tau = \frac{h_T}{\sqrt{\nu/T_r}}$. In this case, in principle, the eigenvalues can be found since they are the roots of a cubic equation, but they are too complicated to analyze. We therefore focus on the persistent modes, i.e., low frequency modes, and derive an estimate for L_r for fast convergence toward the steady state. For small β_x and β_y, the amplification matrix (4.52) simplifies to

$$\mathbf{M}_\Delta \approx \frac{1}{T_r} \begin{bmatrix} -\dfrac{\tau\nu(\beta_x^2+\beta_y^2)}{2h^2} & \dfrac{-i\nu(\tau-2T_r)\beta_x}{2h} & \dfrac{-i\nu(\tau-2T_r)\beta_y}{2h} \\ \dfrac{i\beta_x}{h} & -1-\dfrac{\tau\nu\beta_x^2}{2h^2} & \dfrac{\tau\nu\beta_x\beta_y}{2h^2} \\ \dfrac{i\beta_y}{h} & \dfrac{\tau\nu\beta_x\beta_y}{2h^2} & -1-\dfrac{\tau\nu\beta_y^2}{2h^2} \end{bmatrix}. \tag{4.59}$$

It is pleasing that the eigenvalues of this simplified matrix are particularly simple,

$$\lambda_{1,2} = -\frac{1}{2T_r} \left(\frac{\tau\nu}{h^2}\beta^2 + 1 \pm \sqrt{1 + \frac{2\nu(\tau - 2T_r)}{h^2}\beta^2} \right), \tag{4.60}$$

$$\lambda_3 = -\frac{1}{T_r}. \tag{4.61}$$

Note that because the characteristic equation is cubic there is always one real root. This is given by λ_3 for small β_x and β_y. This eigenvalue represents the inconsistency damping mode.

Turning attention to $\lambda_{1,2}$, we require, as in one dimension, these eigenvalues to be complex conjugates which leads to

$$L_r \ge \frac{h}{4} \left(1 + \sqrt{1 + \frac{4}{\beta^2}} \right), \tag{4.62}$$

and, in order to ensure it for all possible discrete error modes (although only approximately this time), we set $\beta = \pi h$ and define

$$L_r = \frac{h}{4} \left(1 + \sqrt{1 + \frac{4}{\pi^2 h^2}} \right) \approx \frac{1}{2\pi} + \frac{h}{4} + O(h^2). \tag{4.63}$$

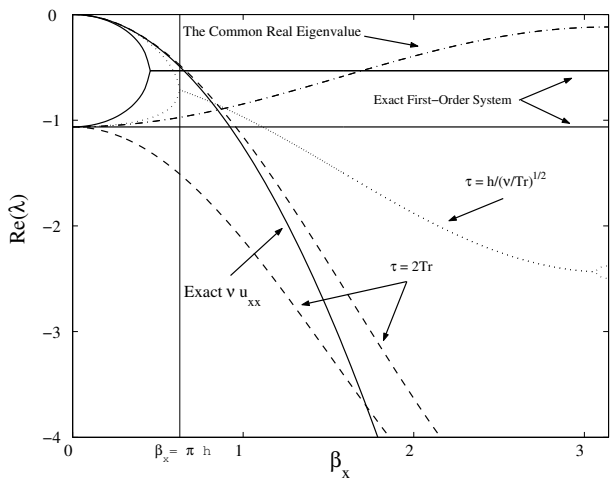

Figure 9: $Re(\lambda)$ for $\beta_y = 0$, $h = 0.2$ and $\nu = 0.05$. $T_r = \frac{{L_r}^2}{\nu}$ with, for a comparison purpose, $L_r = \frac{h}{4}\left(1 + \sqrt{1 + \frac{4}{\pi^2 h^2}}\right)$ for all.

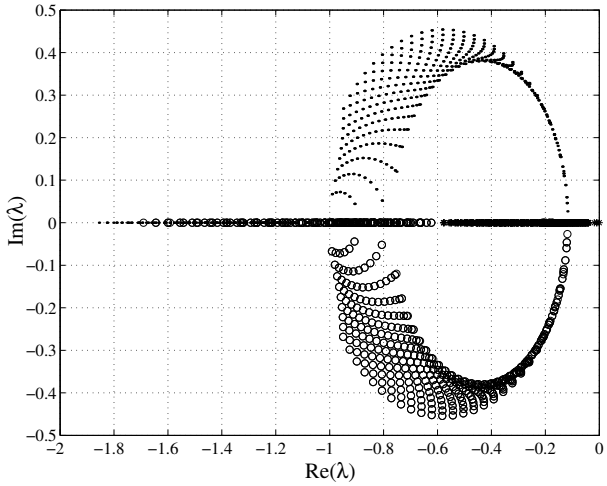

Figure 10: Polar plots of the eigenvalues for the scheme with $\tau = \frac{h_T}{\sqrt{\nu/T_r}}$ and $L_r = \frac{h}{4}\left(1 + \sqrt{1 + \frac{4}{\pi^2 h^2}}\right)$, for $\pi h \le \beta \le \pi$. $h = 0.2$ and $\nu = 0.05$.

This agrees with the one-dimensional version (3.57) up to $O(h^2)$, and so we could use the same approximation as in (3.58).

The real part of the eigenvalue is plotted for $\beta_y = 0$ in Figure 9. They are very similar to the one-dimensional counterparts. However, this time, there is a real eigenvalue common to all schemes. This corresponds to the inconsistency damping mode, i.e., the wave that does not propagate. Fortunately, the damping factor is always less than 1, so that all modes will be damped out. It is also noted that there is a bifurcation point for the scheme with $\tau = \frac{h_T}{\sqrt{\nu/T_r}}$ near $\beta_x = \pi$, beyond which the eigenvalues turn to real. This is because the L_r in (4.63) guarantees complex eigenvalues only approximately for small β, and so the eigenvalues could be real for high-frequency modes. In fact, if we look at the polar plot of the eigenvalues of this scheme as given in Figure 10, in which the three eigenvalues are distinguished by different symbols, we see that there are indeed error modes that are purely damped (apart from the common real eigenvalue indicated by stars).

4.5 Truncation Error

Expand smooth functions u, p, and q over a regular triangular grid (see Figure 8), and substitute them into the semi-discrete equation (4.15) with (4.19) to get

$$\frac{d\mathbf{U}_j}{dt} = \left[\mathbf{I} - \frac{\tau}{2}\left(\mathbf{A}\partial_x + \mathbf{B}\partial_y\right)\right]\mathbf{r} + O(h^2), \tag{4.64}$$

where

$$\mathbf{r} = [\nu(p_x + q_y), (u_x - p)/T_r, (u_y - q)/T_r]^t, \tag{4.65}$$

or component-wise

$$\frac{du_j}{dt} = \nu\left(p_x + q_y\right) + \frac{\tau\nu}{2T_r}\left[(u_x - p)_x + (u_y - q)_y\right] + O(h^2), \tag{4.66}$$

$$\frac{dp_j}{dt} = (u_x - p)/T_r + \frac{\tau\nu}{2T_r}\left(p_x + q_y\right)_x + O(h^2), \tag{4.67}$$

$$\frac{dq_j}{dt} = (u_y - q)/T_r + \frac{\tau\nu}{2T_r}\left(p_x + q_y\right)_y + O(h^2). \tag{4.68}$$

Note again as in one dimension that the scheme has the residual vector $\mathbf{r}$ as a factor in the truncation error, which vanishes at the steady state and second-order accuracy is obtained. The residual-distribution scheme can be thought of as a generalization of the residual-based compact scheme [39] for unstructured triangular grids.

Suppose now that the smooth solutions are exact solutions to the discrete equations in the steady state ($\frac{du_j}{dt} = \frac{dp_j}{dt} = \frac{dq_j}{dt} = 0$), the numerical solutions satisfy

$$\begin{aligned} 0 &= \nu\,(p_x + q_y) + \frac{\tau\nu}{2T_r}\,[(u_x - p)_x + (u_y - q)_y] + O(h^2), && (4.69)\\ 0 &= (u_x - p)/T_r + \frac{\tau\nu}{2T_r}\,(p_x + q_y)_x + O(h^2), && (4.70)\\ 0 &= (u_y - q)/T_r + \frac{\tau\nu}{2T_r}\,(p_x + q_y)_y + O(h^2). && (4.71) \end{aligned}$$

For $\tau = 2T_r$, we obtain

$$\begin{aligned} 0 &= \nu\,(u_{xx} + u_{yy}) + O(h^2), && (4.72)\\ 0 &= (u_x - p)/T_r + \nu\,(p_x + q_y)_x + O(h^2), && (4.73)\\ 0 &= (u_y - q)/T_r + \nu\,(p_x + q_y)_y + O(h^2), && (4.74) \end{aligned}$$

and so u_j converges to the solution of the diffusion equation. We write the other two equations with $T_r = \frac{{L_r}^2}{\nu}$ and $L_r = \frac{h}{2k}$,

$$\begin{aligned} 0 &= \frac{4\nu k^2}{h^2}(u_x - p) + \nu\,(p_x + q_y)_x + O(h^2), && (4.75)\\ 0 &= \frac{4\nu k^2}{h^2}(u_y - q) + \nu\,(p_x + q_y)_y + O(h^2). && (4.76) \end{aligned}$$

The situation is similar to that in one dimension. For $k = O(1)$, this shows that the scheme gives second-order accuracy for p_j and q_j, and for $k = O(h)$, the scheme is not consistent. However, the situation is different in the case $k \to 0$. In this case, the solution converges to the solution of

$$\begin{aligned} 0 &= \nu\,(p_x + q_y)_x + O(h^2), && (4.77)\\ 0 &= \nu\,(p_x + q_y)_y + O(h^2), && (4.78) \end{aligned}$$

and thus the scheme is not consistent. Unfortunately, unlike the one-dimensional case, the cell-residuals do not all necessarily vanish in two dimensions, and so we cannot discuss this any further. We only mention that numerical experiments show that the scheme is indeed inconsistent.

On the other hand, for $\tau = \frac{h_T}{\sqrt{\nu/T_r}} = \frac{hL_r}{\nu}$, we obtain

$$\begin{aligned} 0 &= \nu\,(p_x + q_y) + \frac{\nu h}{2L_r}\,[(u_x - p)_x + (u_y - q)_y] + O(h^2), && (4.79)\\ 0 &= \frac{\nu}{L_r^2}(u_x - p) + \frac{\nu h}{2L_r}\,(p_x + q_y)_x + O(h^2), && (4.80)\\ 0 &= \frac{\nu}{L_r^2}(u_y - q) + \frac{\nu h}{2L_r}\,(p_x + q_y)_y + O(h^2). && (4.81) \end{aligned}$$

For a nonzero finite value of L_r which is the case of (4.63), this shows that the numerical solution converges to the solution of the first-order system (4.3) as $h \to 0$. By eliminating the first-order terms by using the equations themselves, we find

$$\begin{aligned} 0 &= \nu\,(p_x + q_y) - \frac{\nu h^2}{4}\,[(p_x + q_y)_{xx} + (p_x + q_y)_{yy}] + O(h^2), && (4.82)\\ 0 &= (u_x - p) - \frac{h^2}{4}\,[(u_x - p)_x + (u_y - q)_y]_x + O(h^2), && (4.83)\\ 0 &= (u_y - q) - \frac{h^2}{4}\,[(u_x - p)_x + (u_y - q)_y]_y + O(h^2), && (4.84) \end{aligned}$$

thus they converge at the rate of $O(h^2)$.

4.6 Boundary Conditions

In two dimensions, the discrete problem is always overdetermined for triangular grids. Therefore, unlike the one-dimensional case, there are no ways via boundary conditions to equate the number of unknowns and the number of cell residuals. A simple treatment would be that we specify just any values that can be specified. For example, for the Dirichlet conditions, we specify u_j and the gradient along the boundary (p_j or q_j or their combination). Note that the tangential gradient corresponds to zero eigenvalue and so it is irrelevant to the characteristic condition. This means that it suffices to specify one value on the boundary because there is only one characteristic coming out of the boundary. The same is true for the Neumann conditions where we specify only the gradient normal to the boundary.

5 Derived Scalar Schemes

The first-order system approach is useful also in deriving scalar schemes. As mentioned earlier, the scalar isotropic distribution scheme does not provide sufficient dissipation for high-frequency error modes. This means that we need to add a dissipation term in the scheme. However, deriving a dissipation term for the diffusion scheme is not a trivial task especially if we wish to keep the scheme compact. For example, if we apply the Lax-Wendroff time-expansion procedure for the second-order diffusion equation, we immediately face a problem of discretizing second-derivatives of the residual (i.e., the fourth-derivative of the solution) which is not trivial on unstructured grids and certainly cannot be done in a compact manner. Now recall that the system schemes can be thought of as a scalar scheme with an implicit reconstruction of the gradients. Then, it is legitimate to replace the implicit reconstruction by an explicit one. If we decide to do this, we are left with the u_j component of the system schemes. In one dimension, this is given by (3.23) which can be written as

$$h\frac{du_j}{dt} = \left[\frac{1}{2}\phi^L - \frac{\tau\nu}{2T_r}\left(\frac{\Delta u_L}{h} - \overline{p}_L\right)\right] + \left[\frac{1}{2}\phi^R + \frac{\tau\nu}{2T_r}\left(\frac{\Delta u_R}{h} - \overline{p}_R\right)\right], \tag{5.1}$$

where $\phi^L = \int_L \nu u_{xx}\, dx = \nu\Delta p_L$ and $\phi^R = \int_R \nu u_{xx}\, dx = \nu\Delta p_R$. In two dimensions, we have from (4.26)

$$S_j\frac{du_j}{dt} = \sum_{T\in\{T_j\}}\left[\frac{1}{3}\phi^T - \frac{\tau\nu}{4T_r}\left\{\nabla u^T - (\overline{p}_T, \overline{q}_T)\right\}\cdot\mathbf{n}_j^T\right], \tag{5.2}$$

where $\phi^T = \int_T \nu\left(u_{xx} + u_{yy}\right) dxdy = \nu(p_x^T + q_y^T)S_T$. These are now scalar schemes with all (p_j, q_j) evaluated by explicitly reconstructed gradients. As can be seen clearly from these formulas, we have just discovered a form of dissipation: it is constructed over a cell by the difference between the constant gradient of u within the cell and the average of the reconstructed gradients. These terms sum up to zero over the cell, and therefore the schemes remain conservative. Note that these dissipation terms originate from the least-squares part of the distribution matrices, (3.16) and (4.19). This means that we can derive a dissipation term of a scalar scheme for the second-order diffusion equation by applying the least-squares discretization to its equivalent first-order system, in much the same way as is done in [26] for advection schemes . Without going through the first-order system, it would have been almost impossible to derive these dissipation terms.

We now have a family of scalar schemes for the diffusion equation in which the parameter (τ/T_r) may be chosen, for example, to endow the scheme with a property such as positivity [42]. These scalar schemes are, however, under the $O(h^2)$ time step restriction because they are discretizations of second-order derivatives for which $O(h^2)$ geometric factor cannot be avoided. For this reason, we do not consider these schemes further in this paper.

6 Results

6.1 One-Dimensional Problem

We consider the following problem:

$$u_t = \nu\, u_{xx} + \nu\pi^2\sin(\pi x) \quad \text{in } \Omega = [0,1], \tag{6.1}$$

N	ITR	L_1 error of u	Order	L_1 error of p	Order
8	401	2.40E-04		1.29E-03	
16	1601	5.67E-05	2.08	3.24E-04	2.00
32	6394	1.38E-05	2.04	8.12E-05	2.00
64	25546	3.39E-06	2.02	2.03E-05	2.00
128	102112	8.40E-07	2.01	5.08E-06	2.00
256	408372	2.09E-07	2.01	1.27E-06	2.00

Table 1: $\tau = 2T_r$ and $L_r = \dfrac{h}{\sqrt{2}}$

N	ITR	L_1 error of u	Order	L_1 error of p	Order
8	279	2.40E-04		1.29E-03	
16	1109	5.67E-05	2.08	3.24E-04	2.00
32	4432	1.38E-05	2.04	8.12E-05	2.00
64	17710	3.39E-06	2.02	2.03E-05	2.00
128	70796	8.04E-07	2.01	5.08E-06	2.00
256	283133	2.09E-07	2.01	1.27E-06	2.00

Table 2: $\tau = 2T_r$ and $L_r = \dfrac{h}{0.4}$

where $\nu = 1$ and $u(0) = u(1) = 0$. We compute the steady state solution to this problem, by solving the equivalent first-order system

$$\begin{aligned} u_t &= \nu\, p_x + \nu\pi^2 \sin(\pi x), \\ p_t &= (u_x - p)/T_r. \end{aligned} \tag{6.2}$$

The source term in the first equation is evaluated by the trapezoidal rule over the cell and included in the cell-residual, in exactly the same way that p in the second equation is treated. We start from the initial solutions, $u = x(x-1)$ and $p = 2x-1$, integrate in time with a time step defined by (3.28) until convergence, and compare the solutions with the exact steady state solutions: $u = \sin(\pi x)$ and $p = \pi\cos(\pi x)$. We tested new schemes for grids with numbers of cells N = 8, 16, 32, 64, 128, 256. The CFL number is taken to be 0.99 for all cases. A scheme is taken to be converged when the nodal residuals are reduced nine orders of magnitude in the L_1 norm, in order to ensure that the solutions are fully converged. We remark that the steady state solution is independent of ν, and the schemes are designed also to be independent of ν, and therefore all results shown here are valid for any ν. In all results, we show L_1 errors only for brevity. L_2 and L_∞ errors behave similarly.

Shown in Tables 1 to 3 are results for the choice $\tau = 2T_r$ for three different choices of L_r. Table 1 shows results for the scheme with the optimal damping for high-frequency modes. As expected, we see that the number of iterations (indicated by the abbreviation ITR) grows quadratically with the mesh size, and also that the second-order accuracy is obtained for both variables. Table 2 shows results for the scheme with a larger L_r which corresponds to increasing the relaxation parameter ω toward 1 in the point Jacobi iteration for u_j. It converges faster than the previous scheme, but we still have the quadratic increase in the number of iterations. Tables 3 shows results for the scheme with an even greater L_r, corresponding to $\omega \to 1$. We now observe that the accuracy for the gradient variable deteriorates to first-order. This confirms the analysis in Section 3.5.

Next, in Tables 4, 5, and 6, we show results for the choice $\tau = \frac{h}{\sqrt{\nu/T_r}}$ for three different choices of L_r. These schemes allow $O(h)$ time step, and we expect that the number of iterations grows linearly. Table 4 shows results for the optimal L_r in convergence toward the steady state. As can be clearly seen, the scheme converges surprisingly fast, and the number of iterations does grow linearly. Even for the finest grid, it takes only 2,279 iterations while the schemes $\tau = 2T_r$ take about 280,000 iterations (more than 100 times as many) for the same grid. It should be noted that each iteration costs roughly the same for all schemes. Therefore, the gain in the number of iterations is directly translated into CPU time. In the finest grid case, we compared CPU times for two schemes, i.e., the scheme of Table 4 and the scheme of Table 2. The result is that the former took only 3 seconds while the latter took 357 seconds (nearly 6 minutes). The gain is substantial, and will be more and

N	ITR	L_1 error of u	Order	L_1 error of p	Order
8	333	2.40E-04		1.03E-02	
16	1197	5.67E-05	2.08	5.13E-03	1.00
32	4521	1.38E-05	2.04	2.56E-03	1.00
64	17549	3.39E-06	2.02	1.28E-03	1.00
128	69178	8.04E-07	2.01	6.41E-04	1.00
256	275671	2.09E-07	2.01	3.20E-04	1.00

Table 3: $\tau = 2T_r$ and $L_r = \dfrac{h}{0.00001}$

N	ITR	L_1 error of u	Order	L_1 error of p	Order
8	89	2.40E-04		1.29E-03	
16	150	5.67E-05	2.08	3.24E-04	2.00
32	268	1.38E-05	2.04	8.12E-05	2.00
64	561	3.39E-06	2.02	2.03E-05	2.00
128	1022	8.04E-07	2.01	5.08E-06	2.00
256	2279	2.09E-07	2.01	1.27E-06	2.00

Table 4: $\tau = \dfrac{h}{\sqrt{\nu/T_r}}$ and $L_r = \dfrac{h}{4}\left(1 + \dfrac{1}{\sin\frac{\pi h}{2}}\right)$

more substantial as the grid gets finer. Table 5 shows results with the approximate L_r in (3.58). It generally takes more iterations but just a few, showing that it is a good and useful approximation. For the $N = 64$ grid, however, the scheme converges faster than the one with the optimal L_r. Remember that the optimal L_r was derived based on the damping property only, and the error propagation was not taken into account. In particular, the most persistent mode ($\beta = \pi h$) was designed to be purely damped with the optimal L_r. A detailed analysis shows that the most persistent error mode begins to propagate rather than purely damped for the approximate L_r. This could improve the convergence but only marginally, and as we increase L_r from the optimal one the convergence property soon deteriorates because the damping factor grows rapidly. Table 6 shows results with $L_r = 1$. This symmetrizes the first-order diffusion system. Although it takes longer to converge than the previous schemes, the time step is still $O(h)$ and the number of iterations grows linearly. For example, for the finest grid, the number of iterations is only about $\frac{1}{20}$ of those of the schemes $\tau = 2T_r$. Obviously, as far as the iteration toward the steady state is concerned, these schemes offer a great advantage over the conventional schemes with $O(h^2)$ time step.

Finally, we remark that all schemes converge to the same solution (except for the gradient variable in Table 3). This is because the one-dimensional discrete problem has a unique solution as mentioned in 3.2. Therefore, as long as the scheme is consistent, it converges to the same solution.

6.2 Two-Dimensional Problem

We consider the following problem:

$$u_t = \nu(u_{xx} + u_{yy}) \quad \text{in } \Omega = [0,1] \times [0,1], \tag{6.3}$$

where $\nu = 1$ and the boundary conditions $u(x = 0) = 0$, $u(x = 1) = \sin(\pi y)$, $u(y = 0) = 0$, $u(y = 1) = \sin(\pi x)$. We compute the steady state solution to this problem by solving the equivalent first-order system

$$\begin{aligned} u_t &= \nu\,(p_x + q_y), \\ p_t &= (u_x - p)/T_r, \\ q_t &= (u_y - q)/T_r. \end{aligned} \tag{6.4}$$

The exact steady solution is given by

$$u(x,y) = \frac{\sinh(\pi x)\sin(\pi y) + \sinh(\pi y)\sin(\pi x)}{\sinh(\pi)}. \tag{6.5}$$

N	ITR	L_1 error of u	Order	L_1 error of p	Order
8	94	2.40E-04		1.29E-03	
16	157	5.67E-05	2.08	3.24E-04	2.00
32	270	1.38E-05	2.04	8.12E-05	2.00
64	556	3.39E-06	2.02	2.03E-05	2.00
128	1126	8.04E-07	2.01	5.08E-06	2.00
256	2293	2.09E-07	2.01	1.27E-06	2.00

Table 5: $\tau = \dfrac{h}{\sqrt{\nu/T_r}}$ and $L_r = \dfrac{1}{6} + \dfrac{h}{4}$

N	ITR	L_1 error of u	Order	L_1 error of p	Order
8	331	2.40E-04		1.29E-03	
16	702	5.67E-05	2.08	3.24E-04	2.00
32	1383	1.38E-05	2.04	8.12E-05	2.00
64	2973	3.39E-06	2.02	2.03E-05	2.00
128	6152	8.04E-07	2.01	5.08E-06	2.00
256	12753	2.09E-07	2.01	1.27E-06	2.00

Table 6: $\tau = \dfrac{h}{\sqrt{\nu/T_r}}$ and $L_r = 1$

We start from the initial solutions, $u = p = q = 1$ inside the domain. On the boundary, we specify u everywhere, p on the top and bottom boundary, and q on the left and right boundary as they can be evaluated from u given there. We employ the forward Euler time stepping to integrate in time until convergence with the CFL number 0.9, and compare the solutions with the exact steady state solution. The method is taken to be converged when the nodal residuals are reduced nine orders of magnitude in the L_1 norm. This ensures that all numerical solutions are fully converged. New schemes were tested for a series of regular triangular grids: 10×10, 20×20, 40×40, 80×80, 160×160. We remark again as in the one dimensional cases that the steady state solution as well as the schemes are independent of ν, and therefore all results shown here are valid for any ν.

Table 7 shows results for the scheme with $\tau = 2T_r$ (the Galerkin scheme for u_j) and the optimal L_r for high frequency damping. Exactly as expected, we observe a quadratic increase in the number of iterations and second-order accuracy for both variables. Remember that this is the Galerkin scheme for u_j, but we have now the solution gradients of the equal order of accuracy. Table 8 shows results for the scheme with $\tau = \frac{h}{\sqrt{\nu/T_r}}$ with the optimal choice for L_r for the fastest convergence. As can be seen, the number of iterations indeed increases linearly as we expect, and it converges tremendously faster than the previous one for all grids. For example, in the finest grid case, this scheme is about 40 times faster than the previous one. Table 9 shows results for the same scheme with an approximate expression for the optimal L_r which is the same as the one-dimensional version. It only shows a slight increase in the number of iterations. This demonstrates the effectiveness of the approximation. Table 10 shows results for $L_r = 1$, i.e., the symmetric first-order diffusion system. As expected, it takes more iterations to reach the steady state. Nevertheless, the time step remains $O(h)$, and the number of iterations grows linearly with the mesh size. For the finest grid, this scheme converges nearly 8 times faster than the Galerkin scheme. Furthermore, this factor grows linearly as the grid gets finer because the factor in the time steps are $O(h)$. Hence, this scheme still offers a great advantage for the iterative convergence toward the steady state over the conventional schemes with $O(h^2)$ time step.

Note that the numerical solution is not unique in two dimensions. We observe from these results that the errors are generally larger for the Galerkin scheme ($\tau = 2T_r$).

Finally, we mention that there are other iterative methods which show a similar linear convergence property, such as the alternating-direction implicit methods or the preconditioned conjugate gradient methods (see [43]). But these are inherently implicit methods and require a considerable amount of work such as inverting a large matrix at every iteration. This makes them incomparably more expensive than our schemes which are purely explicit and do not require any matrix inversion. Also, the conventional Gauss-Seidel iteration scheme could perform similarly with optimum over-relaxation [44]. But this is not general: it is true only with an optimal

Grids	ITR	L_1 error of u	Order	L_1 error of p	Order	L_1 error of q	Order
10×10	660	2.67E-03		2.60E-02		2.60E-02	
20×20	2534	6.09E-03	2.13	6.48E-03	2.00	6.49E-03	2.00
40×40	9784	1.46E-04	2.06	1.60E-03	2.02	1.60E-03	2.03
80×80	37601	3.56E-05	2.04	3.97E-04	2.01	3.97E-04	2.01
160×160	144542	8.82E-06	2.01	9.94E-05	2.00	9.94E-05	2.00

Table 7: $\tau = 2T_r$ and $L_r = \dfrac{h_T}{\sqrt{3}}$

Grids	ITR	L_1 error of u	Order	L_1 error of p	Order	L_1 error of q	Order
10×10	283	1.39E-03		1.50E-02		1.50E-02	
20×20	590	4.06E-04	1.78	3.22E-03	2.22	3.20E-03	2.23
40×40	969	1.14E-04	1.83	7.48E-04	2.11	7.52E-04	2.09
80×80	2116	3.00E-05	1.93	1.81E-04	2.05	1.81E-04	2.05
160×160	3545	7.67E-06	1.97	4.45E-05	2.03	4.45E-05	2.03

Table 8: $\tau = \dfrac{h_T}{\sqrt{\nu/T_r}}$ and $L_r = \dfrac{h}{4}\left(1 + \sqrt{1 + \dfrac{4}{\pi^2 h_T^2}}\right)$

relaxation factor. On the other hand, there are no subtle tuning parameters in our schemes: any $L_r = O(1)$ will allow $O(h)$ time step. Remember also that our schemes come with solution gradients computed simultaneously with comparable accuracy to the main variable.

7 Concluding Remarks

This paper has introduced a new strategy for computing the steady state solution of the diffusion equation, based on the first-order system that is equivalent to the diffusion equation in the steady state. We developed a class of residual-distribution schemes for the first-order system. Compared with the standard Galerkin scheme, the proposed scheme has remarkable features. First, the new scheme gives second-order accuracy for both the solution and the gradient variables. For practical problems, such as the Navier-Stokes equations, this means that the scheme directly computes the viscous stresses and the heat fluxes in addition to the velocity components with the same order of accuracy. Second, the schemes with $\tau = \frac{h}{\sqrt{\nu/T_r}}$ and $L_r = O(1)$ allow $O(h)$ time step which is significantly larger than the time step of $O(h^2)$ for the conventional schemes. This is a great advantage for steady state computations, motivating the use of explicit time integration schemes for diffusion problems. For time accurate computations, we can employ the dual time stepping technique [45, 46] in which the proposed scheme can be used as a fast iterative method in the inner iteration (see also [47, 48] which are specific to residual-distribution schemes).

In this paper, we studied two types of schemes with $\tau = 2T_r$ and $\tau = \frac{h}{\sqrt{\nu/T_r}}$. The former corresponds to the Galerkin discretization for the main variable, and can be designed so as to have a smoothing property in exactly the same way as the standard scalar scheme. Note that it is identical to the standard Galerkin scheme for the main variable but it comes with equally accurate solution gradients. For the purpose of marching in time toward the steady state, however, this scheme is not well suited for because increasing the relaxation factor ω in the context of the point Jacobi iteration causes accuracy deterioration for the gradient variables. In this case, the other scheme, $\tau = \frac{h}{\sqrt{\nu/T_r}}$, is better suited because this scheme is stable with $O(h)$ time step and converges rapidly to the steady state. We have shown therefore that the first-order system approach works for deriving an effective smoother for multigrid as well as for developing a fast explicit scheme for steady state computations.

We have shown also that the Galerkin scheme, which by itself is not a residual-distribution scheme by definition, arises as a special case of the proposed scheme. It is not residual-distribution by itself, but combined with gradient computations, it is a residual-distribution scheme with cell-residuals defined for the equivalent first-order system. This paper revealed a connection between two methods which had been apparently completely

Grids	ITR	L_1 error of u	Order	L_1 error of p	Order	L_1 error of q	Order
10×10	294	1.41E-03		1.49E-02		1.49E-02	
20×20	610	4.10E-04	1.78	3.21E-03	2.22	3.18E-03	2.23
40×40	1006	1.15E-04	1.84	7.45E-04	2.11	7.49E-04	2.09
80×80	2196	3.01E-05	1.93	1.81E-04	2.04	1.81E-04	2.05
160×160	3703	7.68E-06	1.97	4.45E-05	2.02	4.45E-05	2.02

Table 9: $\tau = \dfrac{h_T}{\sqrt{\nu/T_r}}$ and $L_r = \dfrac{1}{6} + \dfrac{h_T}{4}$

Grids	ITR	L_1 error of u	Order	L_1 error of p	Order	L_1 error of q	Order
10×10	1489	2.62E-03		1.37E-02		1.37E-02	
20×20	3158	5.67E-04	2.21	2.92E-03	2.23	2.84E-03	2.27
40×40	5913	1.33E-04	2.10	6.99E-04	2.06	6.99E-03	2.02
80×80	10480	3.24E-05	2.04	1.74E-04	2.01	1.74E-04	2.01
160×160	18169	7.97E-06	2.02	4.34E-05	2.00	4.34E-05	2.00

Table 10: $\tau = \dfrac{h_T}{\sqrt{\nu/T_r}}$ and $L_r = 1$

different methods, and justifies the use of the Galerkin discretization in the framework of the residual-distribution method.

Although we focused on residual-distribution schemes in this paper, the first-order system approach can apply equally to finite-difference or finite-volume methods. For each scheme employed, an optimal value of L_r may be derived based on a smoothing property or the fastest convergence to a steady state. Or we may simply take $L_r = 1$ to keep the system symmetric. In this case, obviously $L_r = O(1)$, and therefore the resulting scheme will allow $O(h)$ time step. It must be kept in mind however that accuracy is obtained only in the steady state. A rapid convergence with $O(h)$ time step is achieved at the cost of giving up the time accuracy.

This paper has just established a basis for a further development. Yet another remarkable improvement comes in advection-diffusion problems. The first-order system has now an advection term and remains a hyperbolic system, and so we may simply apply *an upwind scheme for the entire system*. Apparently, there is no need any more to 'add' two schemes, an advection scheme and a diffusion scheme, to construct an advection-diffusion scheme. This will be the subject of the subsequent paper.

Acknowledgments

This work has been sponsored by the Space Vehicle Technology Institute, under grant NCC3-989, one of the NASA University Institutes, with joint sponsorship from the Department of Defense. I would like to thank Dr. Mario Ricchiuto(INRIA) for the stimulating discussion from which the idea of extracting a scalar scheme from the first-order system scheme came out, and I thank also Professor Remi Abgrall(the University of Bordeaux I) for the support by which such a discussion was made possible. I would like to thank Mr. Yoshifumi Suzuki for illuminating discussions on the relation between the present approach and the relaxation approach, as well as a number of constructive comments. I also thank Professor P. L. Roe for his valuable comments. Finally, I am grateful to the reviewers for their constructive comments.

References

[1] P. L. Roe, M. Arora, Characteristic-based schemes for dispersive waves I. the method of characteristics for smooth solutions, Numerical Methods for Partial Differential Equations 9 (1993) 459–505.

[2] C. Cattaneo, A form of heat-conduction equations which eliminates the paradox of instantaneous propagation, Ct. R. Acad. Sci., Paris 247 (1958) 431–433.

[3] G. B. Nagy, O. E. Ortiz, O. A. Reula, The behavior of hyperbolic heat equations' solutions near their parabolic limits, J. Math. Phys. 35 (1994) 4334–4356.

[4] R. B. Lowrie, J. E. Morel, Methods for hyperbolic systems with stiff relaxation, International Journal for Numerical Methods in Fluids 40 (2002) 413–423.

[5] S. F. Liotta, V. Romano, G. Russo, Central schemes for balance laws of relaxation type, SIAM Journal on Numerical Analysis 38 (2000) 1337–1356.

[6] S. Jin, C. D. Levermore, Numerical schemes for hyperbolic conservation laws with stiff relaxation terms, Journal of Computational Physics 126 (1996) 449–467.

[7] M. Arora, Explicit characteristic-based high-resolution algorithms for hyperbolic conservation laws, Ph.D. thesis, University of Michigan, Ann Arbor, Michigan (1996).

[8] P. I. Crumpton, J. A. MacKenzie, K. W. Morton, Cell vertex algorithms for the compressible Navier-Stokes equations, Journal of Computational Physics 109 (1993) 1–15.

[9] H. Deconinck, R. Abgrall, Introduction to residual distribution methods, in: 34th VKI CFD Lecture Series Very-High Order Discretization Methods, VKI Lecture Series, 2005.

[10] A. Csik, M. Ricchiuto, H. Deconinck, A conservative formulation of the multidimensional upwind residual distribution schemes for general nonlinear conservation laws, Journal of Computational Physics 179 (2002) 286–312.

[11] R. Abgrall, Toward the ultimate conservative scheme: following the quest, Journal of Computational Physics 167 (2001) 277–315.

[12] D. Caraeni, L. Fuchs, Compact third-order multidimensional upwind scheme for Navier-Stokes simulations, Theoretical and Computational Fluid Dynamics 15 (2002) 373–401.

[13] P. De Palma, G. Pascazio, T. Rubino, M. Napolitano, Residual distribution schemes for advection and advection diffusion problems on quadrilateral cells, Journal of Computational Physics 218 (2006) 159–199.

[14] M. E. Hubbard, A. L. Laird, Achieving high-order fluctuation splitting schemes by extending the stencil, Computers and Fluids 34 (2005) 443–459.

[15] G. T. Tomaich, A genuinely multi-dimensional upwinding algorithm for the Navier-Stokes equations on unstructured grids using a compact, highly-parallelizable spatial discretization, Ph.D. thesis, University of Michigan, Ann Arbor, Michigan (1995).

[16] W. A. Wood, W. L. Kleb, 2-D/axisymmetric formulation of multi-dimensional upwind scheme, in: 15th AIAA Computational Fluid Dynamics Conference, AIAA Paper 2001-2630, Anaheim, 2001.

[17] E. van der Weide, H. Deconinck, E. Issman, G. Degrez, A parallel, implicit, multi-dimensional upwind, residual distribution method for the Navier-Stokes equations on unstructured grids, Computational Mechanics 23 (1999) 199–208.

[18] H. Nishikawa, P. L. Roe, On high-order fluctuation-splitting schemes for Navier-Stokes equations, in: Computational Fluid Dynamics 2004, Springer-Verlag, 2004, pp. 799–804.

[19] O. C. Zienkiewicz, R. L. Taylor, The Finite Element Method, Volume 1, McGraw-Hill Company, 1994.

[20] B.-N. Jiang, The Least-Squares Finite Element Method, Springer, 1998.

[21] B. Cockburn, C.-W. Shu, The local discontinuous Galerkin method for time-dependent convection-diffusion systems, SIAM Journal on Numerical Analysis 35 (1998) 2440–2463.

[22] Y. Sun, Z. J. Wang, Y. Liu, Spectral (finite) volume method for conservation laws on unstructured grids VI: Extension to viscous flow, Journal of Computational Physics 215 (2006) 41–58.

[23] P. L. Roe, Approximate Riemann solvers, parameter vectors, and difference schemes, Journal of Computational Physics 43 (1981) 357–372.

[24] Y. Xing, C.-W. Shu, High order well-balanced finite volume WENO schemes and discontinuous Galerkin methods for a class of hyperbolic systems with source terms, Journal of Computational Physics 214 (2006) 567–598.

[25] M. E. Habbard, P. Garcia Navarro, Flux difference splitting and the balancing of source terms and flux gradients, Journal of Computational Physics 165 (2000) 89–125.

[26] R. Abgrall, Essentially non-oscillatory residual distribution schemes for hyperbolic problems, Journal of Computational Physics 214 (2006) 773–808.

[27] H. Nishikawa, Higher-order discretization of diffusion terms in residual-distribution methods, in: 34th VKI CFD Lecture Series Very-High Order Discretization Methods, VKI Lecture Series, 2005.

[28] H. Paillére, J. Boxho, G. Degrez, H. Deconinck, Multidimensional upwind residual distribution schemes for the convection-diffusion equation, International Journal for Numerical Methods in Fluids 15 (1996) 923–936.

[29] B. van Leer, Computational fluid dynamics: Science or Toolbox?, in: 15th AIAA Computational Fluid Dynamics Conference, AIAA Paper 2001-2520, Anaheim, 2001.

[30] J. A. Hittinger, Foundations for the generalization of the Godunov method to hyperbolic systems with stiff relaxation source terms, Ph.D. thesis, University of Michigan, Ann Arbor, Michigan (2000).

[31] R. B. Pember, Numerical methods for hyperbolic conservation laws with stiff relaxation II. higher-order godunov methods, SIAM Journal on Scientific Computing 14 (1993) 824–859.

[32] E. Turkel, Preconditioning methods for solving the incompressible and low-speed compressible equations, Journal of Computational Physics 72 (1987) 277–298.

[33] J. M. Weiss, W. A. Smith, Preconditioning applied to variable and constant density flows, AIAA Journal 33 (11) (1995) 2050–2057.

[34] B. van Leer, W.-T. Lee, P. L. Roe, Characteristic time-stepping or local preconditioning of the Euler equations, in: 10th AIAA Computational Fluid Dynamics Conference, AIAA Paper 91-1552, Hawaii, 1991.

[35] H. Nishikawa, P. Roe, Y. Suzuki, B. van Leer, A general theory of local preconditioning and its application to the 2D ideal MHD equations, in: 16th AIAA Computational Fluid Dynamics Conference, AIAA Paper 2003-3704, Orlando, 2003.

[36] R.-H. Ni, A multiple-grid scheme for solving the Euler equations, AIAA Journal 20 (11) (1981) 1565–1571.

[37] H. Nishikawa, On grids and solutions from residual minimization, Ph.D. thesis, University of Michigan, Ann Arbor, Michigan (Aug. 2001).

[38] W. L. Briggs, V. E. Henson, S. F. McCormick, A Multigrid Tutorial, 2nd Edition, SIAM, 2000.

[39] A. Lerat, C. Corre, Residual-based compact schemes for multidimensional hyperbolic systems of conservation laws, Computers and Fluids 31 (2002) 639–661.

[40] C.-S. Chou, C.-W. Shu, High order residual distribution conservative finite difference WENO schemes for steady state problems on non-smooth meshes, Journal of Computational Physics 214 (2006) 698–724.

[41] E. van der Weide, H. Deconinck, Positive matrix distribution schemes for hyperbolic systems, with application to the Euler equaitons, in: Computational Fluid Dynamics 1996, Wiley, New York, 1996, pp. 747–753.

[42] M. Ricchiuto, H. Nishikawa, A preliminary look at positive coefficient residual distribution discretizations for advection-diffusion, INRIA Report, 2007.

[43] J. F. Ferziger, M. Perić, Computational Methods for Fluid Dynamics, Springer, 1997.

[44] J. C. Tannehill, D. A. Anderson, R. H. Pletcher, Computational Fluid Mechanis and Heat Transfer, 2nd Edition, Taylor & Francis, 1997.

[45] R. Payret, T. Taylor, Computational Methods for Fluid Flows, Springer, New York, 1983.

[46] A. Jameson, Time dependent calculations using multigrid, with applications to unsteady flows past airfoils and wings, AIAA Paper 91-1596, 1991.

[47] D. Caraeni, L. Fuchs, Compact third-order multidimensional upwind discretization for steady and unsteady flow simulations, Computers and Fluids 34 (2005) 419–441.

[48] G. Rossiello, P. De Palma, G. Pascazio, M. Napolitano, Third-order-accurate fluctuation-splitting schemes for unsteady hyperbolic problems, Journal of Computational Physics 222 (2007) 332–352.