



On the design of high order residual-based dissipation for unsteady compressible flows

A. Lerat*, K. Grimich, P. Cinnella

DynFluid Lab., Arts et Metiers ParisTech, 151 Boulevard de l'Hopital, 75013 Paris, France

ARTICLE INFO

Article history:

Received 16 May 2012

Received in revised form 25 October 2012

Accepted 28 October 2012

Available online 22 November 2012

Keywords:

Residual-based compact scheme

High order

Unsteady Euler equations

Numerical dissipation

ABSTRACT

The numerical dissipation operator of residual-based compact (RBC) schemes of high accuracy is identified and analysed for hyperbolic systems of conservation laws. A necessary and sufficient condition (χ -criterion) is found that ensures dissipation in 2-D and 3-D for any order of the RBC scheme. Numerical applications of RBC schemes of order 3, 5 and 7 to a diagonal wave advection and to a converging cylindrical shock problem confirm the theoretical results.

© 2012 Elsevier Inc. All rights reserved.

1. Introduction

Classical methods for calculating compressible flows on a structured mesh rely on a directional approach in which space derivatives are approximated independently direction by direction. In the present paper, we study compact approximations that provide high accuracy not for each space derivative treated apart but for the complete residual r , i.e. the sum of all of the terms in the governing equations. For steady problems solved by time marching, r is the residual at steady state; it involves space derivatives only. For unsteady problems, r also includes the time derivative. Schemes of this type are referred-to as residual-based compact (RBC). They have been developed in the last ten years and applied to realistic flow configurations in aerodynamics and aeroacoustics (see [1–8]). Related schemes are the Residual Distribution Schemes (see [9–12]), developed on unstructured meshes, in which the residuals are distributed to the nodes of triangles or tetrahedrons following suitable design principles. A special feature of the RBC schemes is the use of a numerical dissipation term also constructed from the complete residual r . This unusual dissipation gives to the RBC schemes special properties that have not been fully analyzed so far. In practice, the RBC schemes are robust for steady flow computations, but some of them may have difficulties (described in Section 2.5) for unsteady problems, apparently due to a weak instability. Here, we present a comprehensive study of the residual-based dissipation term of high-order RBC schemes for the unsteady Euler equations. The study provides a deeper insight of the dissipation mechanism, provides a mathematical criterion (called χ -criterion) characterizing the dissipation for 2-D and 3-D problems and restores the stability of RBC schemes for unsteady problems. Given the importance of numerical dissipation in computational fluid dynamics, it is also hoped that the present work could help the development of other classes of high-order schemes.

The paper is organized as follows. Section 2 reminds the principles of high-order RBC schemes for solving a hyperbolic system of conservation laws on a 2-D Cartesian mesh. Such schemes are dissipative and compact (for instance, order 7 in

* Corresponding author.

E-mail address: alain.lerat@ensam.eu (A. Lerat).

space is achieved using 5×5 -points only). Compact schemes for compressible flows have been mainly developed as centered approximations in space (see [13–16] for instance) relying on the use of artificial viscosities, numerical filters or limiters for shock capturing. Upwind compact schemes have also been proposed in [17,18]. A peculiarity of the present RBC schemes is to use three independent compact approximations of the residual (four in 3-D). One applies to the usual residual at the current location j, k (main residual). The two others are involved in the dissipation and defined at location $j + \frac{1}{2}, k$ or $j, k + \frac{1}{2}$ (mid-point residuals). A correct choice of the discretization of the mid-point residuals is essential to ensure dissipation for all flow conditions. This is why we try here to identify the effective dissipation operator, which is not obvious for a RBC scheme since the dissipation operator comes from high order expansions of the mid-point residuals. This work is done in two stages, in Section 3 and 4 respectively, for RBC schemes of order $2p - 1 = 3, 5$ and 7. In Section 5, the dissipation operator is cast into a general form (2-D partial differential operator of degree $2p$) and a necessary and sufficient condition, the χ -criterion, is found for this operator be always dissipative. Application of this criterion to the RBC schemes gives the correct coefficients to use in the high-order approximations of the mid-point residuals. A complete extension of the analysis to three space-dimension is presented in Section 6. Finally, numerical experiments are presented in Section 7 to confirm the relevance of the χ -criterion.

2. High-order RBC schemes

2.1. Concept of residual-based scheme

Let us consider an initial-value problem for the hyperbolic system of conservation laws:

$$w_t + f_x + g_y = 0 \tag{1}$$

where t is the time, x and y are Cartesian space coordinates, w is the state vector and $f = f(w), g = g(w)$ are flux components depending smoothly on w . The Jacobian matrices of the flux are denoted $A = df/dw$ and $B = dg/dw$. System (1) is approximated in space on a uniform mesh ($x_j = j\delta x, y_k = k\delta y$) with steps δx and δy of the same order of magnitude, say $\mathcal{O}(h)$, using a residual-based scheme. Such a scheme can be expressed only in terms of approximations of the exact residual, i.e. of the left-hand side of System (1). More precisely this scheme is a discrete form of

$$w_t + f_x + g_y = \frac{\delta x}{2} [\Phi_1(w_t + f_x + g_y)]_x + \frac{\delta y}{2} [\Phi_2(w_t + f_x + g_y)]_y \tag{2}$$

The coefficients Φ_1 and Φ_2 are numerical viscosity matrices that depend only on the eigensystems of the Jacobian matrices A and B and on the step ratio $\delta x/\delta y$. They use no tuning parameters or limiters. Their construction is presented in Section 2.4.

In System (2), the exact residual

$$r = w_t + f_x + g_y$$

is everywhere approximated in a space-centered way, but not at the same location or at the same order. The discrete form of (2) can be written as

$$\tilde{r}_{j,k} = \tilde{d}_{j,k} \tag{3}$$

where $\tilde{r}_{j,k}$ is a space-centered approximation of r , called the *main residual* and $\tilde{d}_{j,k}$ is the *residual-based dissipation* defined as:

$$\tilde{d}_{j,k} = \frac{1}{2} [(\Phi_1 \tilde{r}_1)_{j+\frac{1}{2},k} - (\Phi_1 \tilde{r}_1)_{j-\frac{1}{2},k}] + \frac{1}{2} [(\Phi_2 \tilde{r}_2)_{j,k+\frac{1}{2}} - (\Phi_2 \tilde{r}_2)_{j,k-\frac{1}{2}}] \tag{4}$$

where $(\tilde{r}_1)_{j+\frac{1}{2},k}$ and $(\tilde{r}_2)_{j,k+\frac{1}{2}}$ are space-centered approximations of r , called the *mid-point residuals* - see Fig. 1.

Despite appearance, the order of magnitude of the residual-based dissipation is not simply $\mathcal{O}(h)$ as it could seem from (2), but much smaller because the mid-point residuals approximate the exact residual which is everywhere null. Since centered differencing always leads to even order of accuracy, let the mid-point residuals be discretized so that

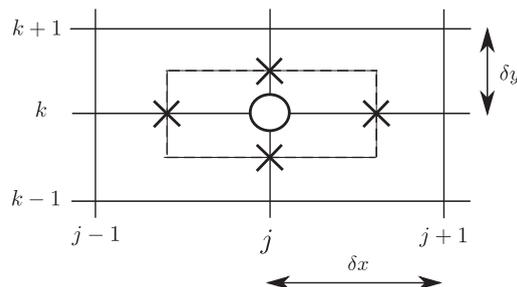


Fig. 1. Location of the discrete residuals, \circ : main residual \tilde{r} , \times : mid-point residual \tilde{r}_1 or \tilde{r}_2 .

$$\begin{aligned}(\tilde{r}_1)_{j+\frac{1}{2},k} &= r_{j+\frac{1}{2},k} + \mathcal{O}(h^{2p-2}) + \mathcal{O}(\Delta t^l) = \mathcal{O}(h^{2p-2}) + \mathcal{O}(\Delta t^l) \\(\tilde{r}_2)_{j,k+\frac{1}{2}} &= r_{j,k+\frac{1}{2}} + \mathcal{O}(h^{2p-2}) + \mathcal{O}(\Delta t^l) = \mathcal{O}(h^{2p-2}) + \mathcal{O}(\Delta t^l)\end{aligned}$$

where p and l are integers ($p \geq 2$, $l \geq 1$). Then the residual-based dissipation verifies

$$\tilde{d}_{j,k} = \mathcal{O}(h^{2p-1}) + \mathcal{O}(h\Delta t^l). \quad (5)$$

If in addition the main residual is such that

$$\tilde{r}_{j,k} = r_{j,k} + \mathcal{O}(h^{2p}) + \mathcal{O}(\Delta t^l) = \mathcal{O}(h^{2p}) + \mathcal{O}(\Delta t^l)$$

then the truncation error of the scheme (3) is

$$\varepsilon_{j,k} = \mathcal{O}(h^{2p}) + \mathcal{O}(\Delta t^l) + \mathcal{O}(h^{2p-1}) + \mathcal{O}(h\Delta t^l) = \mathcal{O}(h^{2p-1}) + \mathcal{O}(\Delta t^l) \quad (6)$$

and the scheme is accurate at order $2p - 1$ in space. Since the spatial approximation of \tilde{r} , \tilde{r}_1 and \tilde{r}_2 will be made using compact formulas, the scheme (3) is said to be *residual-based compact* of order $2p - 1$ and denoted as RBC_{2p-1} .

2.2. Main residual

Let us first introduce the way we denote and handle high order approximations in the main residual. Any discrete formula can easily be expressed from combinations of a difference and an average operator over one mesh interval in each space direction, namely

$$\begin{aligned}(\delta_1 v)_{j+\frac{1}{2},k} &= v_{j+1,k} - v_{j,k} & (\delta_2 v)_{j,k+\frac{1}{2}} &= v_{j,k+1} - v_{j,k} \\(\mu_1 v)_{j+\frac{1}{2},k} &= \frac{1}{2}(v_{j+1,k} + v_{j,k}) & (\mu_2 v)_{j,k+\frac{1}{2}} &= \frac{1}{2}(v_{j,k+1} + v_{j,k})\end{aligned}$$

where j and k are integers or half integers. All these discrete operators commute. For instance:

$$\begin{aligned}\left(\frac{\delta_1 \mu_1 f}{\delta x}\right)_{j,k} &= \left(\frac{\mu_1 \delta_1 f}{\delta x}\right)_{j,k} = \frac{f_{j+1,k} - f_{j-1,k}}{2\delta x}, \\(\delta_1^2 f)_{j,k} &= (\delta_1(\delta_1 f))_{j,k} = f_{j+1,k} - 2f_{j,k} + f_{j-1,k}.\end{aligned} \quad (7)$$

With the above notations, the 8th-order centered approximation at (j, k) of a first derivative of a smooth function ($f \in \mathcal{C}^9$) can be written as:

$$f_x = \left(I - \frac{1}{6}\delta_1^2 + \frac{1}{30}\delta_1^4 - \frac{1}{140}\delta_1^6\right) \frac{\delta_1 \mu_1 f}{\delta x} + \mathcal{O}(\delta x^8) \quad (8)$$

where I is the identity operator and the subscripts (j, k) are omitted. This formula has a 9-point stencil (from $j - 4, k$ to $j + 4, k$). The advantage of the writing (8) is to use 3 coefficients only ($-\frac{1}{6}, \frac{1}{30}, -\frac{1}{140}$) and to display embedded formulas of lower order: dropping the δ^6 term leads to the 6th-order approximation, dropping also the δ^4 term leads to the 4th-order and dropping in addition the δ^2 term gives the 2nd-order formula (7).

Compact approximations can be described similarly by using *Pade fractions* of difference operators. For instance, a 8th-order centered approximation at (j, k) of a first derivative on a 5-point stencil can be written formally as:

$$f_x = \frac{I + \frac{5}{42}\delta_1^2}{I + \frac{2}{7}\delta_1^2 + \frac{1}{70}\delta_1^4} \frac{\delta_1 \mu_1 f}{\delta x} + \mathcal{O}(\delta x^8) \quad (9)$$

The meaning of the above formula is

$$\left(I + \frac{2}{7}\delta_1^2 + \frac{1}{70}\delta_1^4\right) f_x = \left(I + \frac{5}{42}\delta_1^2\right) \frac{\delta_1 \mu_1 f}{\delta x} + \mathcal{O}(\delta x^8) \quad (10)$$

where both sides are defined on a 5-point stencil. In other words, the “denominator” in (9) denotes an operator inversion. An important point is the following one: by expanding the inverse of the denominator $(I + \varepsilon_1)^{-1}$ in terms of the operator $\varepsilon_1 = \frac{2}{7}\delta_1^2 + \frac{1}{70}\delta_1^4 = \mathcal{O}(h^2)$ and taking into account the numerator $I + \frac{5}{42}\delta_1^2$, we recover a non-compact formula which is nothing but (8). Conversely, this procedure is useful to determine the coefficients of a compact formula from the knowledge of a non-compact one.

In a compact formula like (10), the derivative is usually found by solving a linear algebraic system on each horizontal mesh line $k = cte$. In the present residual-based approach we do not follow this procedure, as we will see below. Finally, it should be emphasized that a compact formula reduces the truncation error with respect to a non-compact one of the same order of accuracy. For instance the remainder $\mathcal{O}(\delta x^8)$ in (8) and (10) can be expressed as¹:

¹ In this paper, f_{qx} denotes the q th-derivative $\frac{\partial^q f}{\partial x^q}$.

$$C\delta x^8 f_{9x}(x_j + \theta\delta x), \quad \text{with} \quad -4 \leq \theta \leq 4$$

where the constant C is $-\frac{1}{630}$ for the non compact formula (8) and $-\frac{1}{44100}$ for the compact formula (10), which is precisely 70 times smaller. Fourier analysis also shows that compactness improves accuracy in all the spectrum (see Lele [13]).

We now describe the compact treatment of the main residual $\tilde{r}_{j,k}$ in the RBC $_{2p-1}$ scheme. As we have seen in Section 2.1, the main residual should be approximated at order $2p$. This is done by using the Pade fractions:

$$\begin{aligned} (f_x)_{j,k} &= \left(\frac{\overline{N}_1}{\overline{D}_1} \frac{\delta_1 \mu_1 f}{\delta x} \right)_{j,k} + \mathcal{O}(\delta x^{2p}) \\ (g_y)_{j,k} &= \left(\frac{\overline{N}_2}{\overline{D}_2} \frac{\delta_2 \mu_2 g}{\delta y} \right)_{j,k} + \mathcal{O}(\delta y^{2p}) \end{aligned} \tag{11}$$

where \overline{N}_m and \overline{D}_m are formal polynomials of second difference operators:

$$\overline{N}_m = I + \bar{a}\delta_m^2, \quad \overline{D}_m = I + \bar{b}\delta_m^2 + \bar{c}\delta_m^4, \quad m = 1, 2, \tag{12}$$

The degrees of these polynomials are chosen so that the scheme stencil does not exceed 5×5 points. Both space directions are treated similarly, i.e. the polynomials coefficients \bar{a} , \bar{b} and \bar{c} are the same for $m = 1$ and 2.

The denominators in (11) are eliminated by applying the operator $\overline{D}_1 \overline{D}_2$ to all the terms. Then, the main residual is defined as:

$$\tilde{r}_{j,k} = \left(\overline{D}_1 \overline{D}_2 w_t + \overline{D}_2 \overline{N}_1 \frac{\delta_1 \mu_1 f}{\delta x} + \overline{D}_1 \overline{N}_2 \frac{\delta_2 \mu_2 g}{\delta y} \right)_{j,k}$$

which is really of order $2p$, since

$$\tilde{r}_{j,k} = (\overline{D}_1 \overline{D}_2)_{j,k} [w_t + f_x + g_y + \mathcal{O}(h^{2p})] = (\overline{D}_1 \overline{D}_2)_{j,k} \mathcal{O}(h^{2p}) = [I + \mathcal{O}(h^2)] \mathcal{O}(h^{2p}) = \mathcal{O}(h^{2p}).$$

The detailed expression of the main residual is

$$\tilde{r}_{j,k} = [(I + \bar{b}\delta_1^2 + \bar{c}\delta_1^4)(I + \bar{b}\delta_2^2 + \bar{c}\delta_2^4)w_t + (I + \bar{b}\delta_2^2 + \bar{c}\delta_2^4)(I + \bar{a}\delta_1^2) \frac{\delta_1 \mu_1 f}{\delta x} + (I + \bar{b}\delta_1^2 + \bar{c}\delta_1^4)(I + \bar{a}\delta_2^2) \frac{\delta_2 \mu_2 g}{\delta y}]_{j,k} \tag{13}$$

Several accuracy orders $2p$ can be reached with this expression.

(a) Order $2p = 4$ is achievable on a 3×3 -point stencil by the choice:

$$\bar{a} = 0, \quad \bar{b} = \frac{1}{6}, \quad \bar{c} = 0. \tag{14}$$

which reduces (13) to

$$\tilde{r}_{j,k} = \left[\left(I + \frac{1}{6}\delta_1^2 \right) \left(I + \frac{1}{6}\delta_2^2 \right) w_t + \left(I + \frac{1}{6}\delta_2^2 \right) \frac{\delta_1 \mu_1 f}{\delta x} + \left(I + \frac{1}{6}\delta_1^2 \right) \frac{\delta_2 \mu_2 g}{\delta y} \right]_{j,k} \tag{15}$$

(b) Order $2p = 6$ can be obtained using

$$\bar{a} = \frac{1}{30} + 6\bar{c}, \quad \bar{b} = \frac{1}{5} + 6\bar{c}$$

which gives a family of the 6th-order approximations depending on the coefficient \bar{c} . The choice $\bar{c} = 0$ is the simplest one, but we retain the one made in [4,5]:

$$\bar{a} = \frac{1}{10}, \quad \bar{b} = \frac{4}{15}, \quad \bar{c} = \frac{1}{90} \tag{16}$$

because it is more suitable for the extension to the compressible Navier–Stokes equations.

(c) Finally, order $2p = 8$ (the highest one for a 5×5 -point stencil) can be obtained only with

$$\bar{a} = \frac{5}{42}, \quad \bar{b} = \frac{2}{7}, \quad \bar{c} = \frac{1}{70} \tag{17}$$

which corresponds to the example of Pade fractions (9). Concerning now the time approximation, it is treated independently of the space approximation, so that various choices can be made. In the present work, we use a linear multistep method (LMM). In this implicit method, the residual is taken at the new time level $(n + 1)\Delta t$ and

$$(w_t)^{n+1} = \frac{1}{\Delta t} \left(\Delta w + \frac{1}{2} \Delta^2 w + \frac{1}{3} \Delta^3 w + \dots + \frac{1}{l} \Delta^l w \right)^{n+1} + \mathcal{O}(\Delta t^l) \tag{18}$$

where

$$(\Delta w)^{n+1} = w^{n+1} - w^n, \quad \Delta^2 w = \Delta(\Delta w), \quad \Delta^3 w = \Delta(\Delta^2 w) \dots$$

For $l = 2$, we get the popular approximation

$$(w_t)^{n+1} = \frac{1}{2\Delta t} (3w^{n+1} - 4w^n + w^{n-1}) + \mathcal{O}(\Delta t^2). \tag{19}$$

LMM methods are A-stable at order $l = 1$ and 2 and A(α)-stable at higher orders (see [19–21]). A-stable methods are unconditionally stable when the spatial approximation is dissipative.

2.3. Mid-point residuals

We now consider the dissipation term (4). Using the discrete operators introduced in the previous section, it can be rewritten as:

$$\tilde{d}_{j,k} = \frac{1}{2} [\delta_1(\Phi_1 \tilde{r}_1) + \delta_2(\Phi_2 \tilde{r}_2)]_{j,k} \tag{20}$$

In theory, the mid-point residuals $(\tilde{r}_1)_{j+\frac{1}{2},k}$ and $(\tilde{r}_2)_{j,k+\frac{1}{2}}$ are constructed similarly as the main residual. In practice, the treatment is a bit more intricate. First, to define a mid-point residual, say $(\tilde{r}_1)_{j+\frac{1}{2},k}$, we have to approximate f_x and to average $w_t + g_y$ at the mid location $x = (j + \frac{1}{2})\delta x$, using new types of Pade fractions based on the 2-point difference and average operators δ_1 and μ_1 . As mentioned in Section 2.1, the mid-point residuals are discretized at a lower order than the main residual ($2p - 2$ instead of $2p$). So we introduce:

$$\begin{aligned} (f_x)_{j+\frac{1}{2},k} &= \left(\frac{N_1^\delta}{D_1^\delta} \frac{\delta_1 f}{\delta x} \right)_{j+\frac{1}{2},k} + \mathcal{O}(\delta x^{2p-2}) \\ (g_y)_{j,k+\frac{1}{2}} &= \left(\frac{N_2^\delta}{D_2^\delta} \frac{\delta_2 g}{\delta y} \right)_{j,k+\frac{1}{2}} + \mathcal{O}(\delta y^{2p-2}) \end{aligned} \tag{21}$$

and for any mesh function v :

$$\begin{aligned} (v)_{j+\frac{1}{2},k} &= \left(\frac{N_1^\mu}{D_1^\mu} \mu_1 v \right)_{j+\frac{1}{2},k} + \mathcal{O}(\delta x^{2p-2}) \\ (v)_{j,k+\frac{1}{2}} &= \left(\frac{N_2^\mu}{D_2^\mu} \mu_2 v \right)_{j,k+\frac{1}{2}} + \mathcal{O}(\delta y^{2p-2}) \end{aligned} \tag{22}$$

where $N_m^\delta, D_m^\delta, N_m^\mu$ and D_m^μ for $m = 1, 2$ are formal polynomials of second difference operators:

$$\begin{aligned} N_m^\delta &= I + a^\delta \delta_m^2, & N_m^\mu &= I + a^\mu \delta_m^2 \\ D_m^\delta &= I + b^\delta \delta_m^2 + c^\delta \delta_m^4, & D_m^\mu &= I + b^\mu \delta_m^2 + c^\mu \delta_m^4. \end{aligned} \tag{23}$$

Another point is that the dissipation has to be defined on a 5×5 -point stencil as the main residual. Therefore, the formula (20) shows that the mid-point residuals should have smaller stencils. Namely, $(\tilde{r}_1)_{j+\frac{1}{2},k}$ and $(\tilde{r}_2)_{j,k+\frac{1}{2}}$ should only use 4×5 and 5×4 points, respectively. Fortunately, this is possible because orders $2p - 2 = 2, 4$ and 6 can be achieved with Pade fractions (21) having the same denominators, more precisely:

$$D_1^\delta = D_1^\mu, \quad D_2^\delta = D_2^\mu \tag{24}$$

Finally, for the approximations of g_y at $j + \frac{1}{2}, k$ in \tilde{r}_1 (respectively of f_x at $j, k + \frac{1}{2}$ in \tilde{r}_2), we need the average (22) in the x -direction (resp. in the y -direction), but also the classical Pade fractions for a derivative in the y -direction (resp. in the x -direction). These Pade fractions are analogous to the formulas (11) used in the main residual, but they have different coefficients because they require a lower accuracy order ($2p - 2$ instead of $2p$). So we introduce new formal polynomials N_m, D_m and define:

$$\begin{aligned} (f_x)_{j,k} &= \left(\frac{N_1}{D_1} \frac{\delta_1 \mu_1 f}{\delta x} \right)_{j,k} + \mathcal{O}(\delta x^{2p-2}) \\ (g_y)_{j,k} &= \left(\frac{N_2}{D_2} \frac{\delta_2 \mu_2 g}{\delta y} \right)_{j,k} + \mathcal{O}(\delta y^{2p-2}) \end{aligned} \tag{25}$$

where

$$N_m = I + a\delta_m^2, \quad D_m = I + b\delta_m^2 + c\delta_m^4, \quad m = 1, 2. \tag{26}$$

Here the denominators are compatible with the stencils of the mid-point residuals (5 points allowed in the y -direction for \tilde{r}_1 and in the x -direction for \tilde{r}_2).

A direct discretization of the exact residual at $j + \frac{1}{2}, k$ gives:

$$(w_t + f_x + g_y)_{j+\frac{1}{2},k} = \left[\frac{N_1^\mu}{D_1^\mu} \mu_1 \left(w_t + \frac{N_2}{D_2} \frac{\delta_2 \mu_2 g}{\delta y} \right) + \frac{N_1^\delta}{D_1^\delta} \frac{\delta_1 f}{\delta x} \right]_{j+\frac{1}{2},k} + \mathcal{O}(h^{2p-2})$$

By applying the operator $D_1^\mu D_2 = D_1^\delta D_2$ to all the terms, we obtain the first mid-point residual:

$$(\tilde{r}_1)_{j+\frac{1}{2},k} = \left[N_1^\mu \mu_1 \left(D_2 w_t + N_2 \frac{\delta_2 \mu_2 g}{\delta y} \right) + N_1^\delta D_2 \frac{\delta_1 f}{\delta x} \right]_{j+\frac{1}{2},k} \tag{27}$$

Similarly, we get the second mid-point residual:

$$(\tilde{r}_2)_{j,k+\frac{1}{2}} = \left[N_2^\mu \mu_2 \left(D_1 w_t + N_1 \frac{\delta_1 \mu_1 f}{\delta x} \right) + N_2^\delta D_1 \frac{\delta_2 g}{\delta y} \right]_{j,k+\frac{1}{2}}. \tag{28}$$

These residuals are respectively defined on a 4×5 and a 5×4 -point stencil. Both can be $\mathcal{O}(h^{2p-2})$ for $p = 2, 3, 4$. Their detailed expressions can be written as:

$$\begin{aligned} \tilde{r}_1 &= (I + b\delta_2^2 + c\delta_2^4) \left[(I + a^\mu \delta_1^2) \mu_1 w_t + (I + a^\delta \delta_1^2) \frac{\delta_1 f}{\delta x} \right] + (I + a^\mu \delta_1^2)(I + a\delta_2^2) \frac{\delta_2 \mu_2 \mu_1 g}{\delta y} \\ \tilde{r}_2 &= (I + b\delta_1^2 + c\delta_1^4) \left[(I + a^\mu \delta_2^2) \mu_2 w_t + (I + a^\delta \delta_2^2) \frac{\delta_2 g}{\delta y} \right] + (I + a^\mu \delta_2^2)(I + a\delta_1^2) \frac{\delta_1 \mu_1 \mu_2 f}{\delta x}. \end{aligned} \tag{29}$$

They depend on the five coefficients a^μ, a^δ, a, b and c , the values of which will be given in Section 3 and 4. Note that the time approximation in the mid-point residuals is the same as in the main residual.

2.4. Numerical viscosity matrices

The numerical viscosity matrices Φ_1 and Φ_2 in the residual-based dissipation were designed for the solution of steady flow problems. For these problems, the time derivative is cancelled into the mid-point residuals (29) and the scheme becomes a discrete form of

$$w_t + f_x + g_y = d^{\text{steady}} \tag{30}$$

where

$$d^{\text{steady}} = \frac{\delta x}{2} [\Phi_1 (f_x + g_y)]_x + \frac{\delta y}{2} [\Phi_2 (f_x + g_y)]_y \tag{31}$$

The time evolution being reduced to a numerical procedure to reach a steady-state solution, we simply use the backward Euler approximation in time (LMM method of order 1) with a large time step. During the convergence to the steady-state, note that the residual-based scheme is also first-order in space because of the lack of w_t in (31). This ensures robustness to the scheme in the convergence process. At steady-state, the scheme recovers the high accuracy order of the spatial approximation.

It is clear that some conditions should be satisfied by Φ_1 and Φ_2 in order that (31) be actually dissipative. The term d^{steady} can be rewritten as:

$$d^{\text{steady}} = \frac{\delta x}{2} (\Phi_1 A w_x + \Phi_1 B w_y)_x + \frac{\delta y}{2} (\Phi_2 A w_x + \Phi_2 B w_y)_y \tag{32}$$

Consider the case where A, B, Φ_1 and Φ_2 are scalar constants. Then (32) reduces to

$$d^{\text{steady}} = \mathcal{D}^{\text{steady}} w$$

with the linear partial differential operator of second order:

$$\mathcal{D}^{\text{steady}} = \frac{\delta x}{2} \Phi_1 A \frac{\partial^2}{\partial x^2} + \frac{1}{2} (\delta x \Phi_1 B + \delta y \Phi_2 A) \frac{\partial^2}{\partial x \partial y} + \frac{\delta y}{2} \Phi_2 B \frac{\partial^2}{\partial y^2}$$

which contains the highest derivatives of (32). The Fourier symbol of $\mathcal{D}^{\text{steady}}$ is:

$$\widehat{\mathcal{D}}^{\text{steady}} = -\frac{1}{2} [\delta x \Phi_1 A \xi^2 + (\delta x \Phi_1 B + \delta y \Phi_2 A) \xi \eta + \delta y \Phi_2 B \eta^2] \tag{33}$$

where ξ and η are the wave numbers (Fourier variables).

By dissipation (in the broad sense), we mean:

$$\forall \xi \in \mathbb{R}, \quad \forall \eta \in \mathbb{R}, \quad \widehat{\mathcal{D}}^{\text{steady}}(\xi, \eta) \leq 0,$$

that is d^{steady} damps any Fourier mode.

For $\xi = 0$, the above dissipation condition leads to $\Phi_2 B \geq 0$. For $\xi \neq 0$, we can rewrite (33) as:

$$\widehat{\mathcal{D}}^{\text{steady}}(\xi, \eta) = -\frac{\xi^2}{2} \mathcal{D}(\xi, \eta)$$

with

$$\mathcal{D}(\xi, \eta) = \delta x \Phi_1 A + (\delta x \Phi_1 B + \delta y \Phi_2 A) \left(\frac{\eta}{\xi} \right) + \delta y \Phi_2 B \left(\frac{\eta}{\xi} \right)^2$$

Clearly, \mathcal{D} is always positive if and only if

$$\Phi_2 B \geq 0 \quad \text{and} \quad \Delta \leq 0$$

where

$$\Delta = (\delta x \Phi_1 B + \delta y \Phi_2 A)^2 - 4 \delta x \Phi_1 A \delta y \Phi_2 B = (\delta x \Phi_1 B - \delta y \Phi_2 A)^2.$$

Thus, the necessary and sufficient conditions to get dissipation are $\Phi_2 B \geq 0$ and $\Delta = 0$, which can also be expressed as:

$$\begin{aligned} \Phi_1 A &\geq 0, \quad \Phi_2 B \geq 0 \\ \delta x \Phi_1 B &= \delta y \Phi_2 A \end{aligned} \quad (34)$$

Introducing the notations

$$\begin{aligned} \Phi_1 &= \text{sgn}(A)\Phi, \quad \Phi_2 = \text{sgn}(B)\Psi \\ \alpha &= \frac{\delta x |B|}{\delta y |A|}, \end{aligned} \quad (35)$$

the dissipation conditions (34) become:

$$\Phi \geq 0, \quad \Psi = \alpha \Phi. \quad (36)$$

Note that the parameter α characterizes the local advection direction with respect to the mesh. For $\alpha = 1$, the advection takes place along the mesh diagonal. For $\alpha < 1$, it takes place between the x -direction and a mesh diagonal and for $\alpha > 1$ between the y -direction and a mesh diagonal. Various choices of Φ satisfying (36) are possible. An optimal one proposed in [1] is:

$$\Phi = \min\left(1, \frac{1}{\alpha}\right), \quad \Psi = \alpha \Phi = \min(1, \alpha). \quad (37)$$

These functions make use of no tuning parameter and only depend on the local advection direction α . For hyperbolic systems of conservation laws, the matrix functions Φ_1 and Φ_2 are defined through a direct extension of the scalar case: the eigenvectors of Φ_1 are those of the Jacobian matrix A – evaluated at some suitable intermediate state –, the eigenvectors of Φ_2 are those of B and the eigenvalues of Φ_1 and Φ_2 are deduced from the above scalar definitions. More precisely, let T_A (respectively T_B) be a matrix the columns of which are the right eigenvectors of A (resp. B) and let $a^{(i)}$ (resp. $b^{(i)}$) be the eigenvalues of A (resp. B). Matrices Φ_1 and Φ_2 are then defined as

$$\Phi_1 = T_A \text{Diag}[\phi_1^{(i)}] T_A^{-1}, \quad \Phi_2 = T_B \text{Diag}[\phi_2^{(i)}] T_B^{-1}$$

with

$$\begin{aligned} \phi_1^{(i)} &= \text{sgn}(a^{(i)}) \phi^{(i)}, \quad \phi_2^{(i)} = \text{sgn}(b^{(i)}) \psi^{(i)} \\ \phi^{(i)} &= \min\left(1, \frac{\delta y |a^{(i)}|}{\delta x m(B)}\right), \quad \psi^{(i)} = \min\left(1, \frac{\delta x |b^{(i)}|}{\delta y m(A)}\right) \end{aligned}$$

where $\text{Diag}[d^{(i)}]$ denotes a diagonal matrix with diagonal entries $d^{(i)}$ and $m(A) = \min_i |a^{(i)}|$, $m(B) = \min_i |b^{(i)}|$.

2.5. Dissipation mechanism for steady and unsteady computations

For steady flow problems, it is clear that the operator d^{steady} in (31) with the above numerical viscosity matrices Φ_1 and Φ_2 ensures dissipation. As a matter of fact, its use has led to successful calculations of various steady flow problems (see [2,8] for instance).

For unsteady simulations, the time derivative is included in the dissipation operator as in (2) and the viscosity matrices are kept unchanged. This has allowed the solution of the Euler equations using RBC schemes of order 5 and 7 for shock-vor-

tex interaction [5], the computation of some turbomachinery flows using RBC3 [7] and a 3D propagation of spinning acoustic modes in an aeroengine inlet using RBC7 [8]. However for the later case, a 10th-order azimuthal filtering was necessary. Other unsteady applications to turbomachinery flows were difficult, even with the RBC scheme of order 3, so that a weak instability of the unsteady scheme can be suspected. This is the reason why a deeper insight of the dissipation mechanism is needed.

For the unsteady scheme, the discrete operator $\tilde{d}_{j,k}$ represents the differential operator:

$$d = \frac{\delta x}{2} [\Phi_1(w_t + f_x + g_y)]_x + \frac{\delta y}{2} [\Phi_2(w_t + f_x + g_y)]_y \quad (38)$$

which cannot be viewed as the real dissipation since it is identically null! So the effect of $\tilde{d}_{j,k}$ must be identified by expanding it further in space and time. For a RBC_{2p-1} scheme, the dissipation $\tilde{d}_{j,k}$ is of order h^{2p-1} in space –see (5). To identify it, the centered term $\tilde{d}_{j,k}$ must be expanded up to a remainder $\mathcal{O}(h^{2p+1})$. Concerning the expansion in time, the expression (6) of the truncation error of the global scheme shows that the leading term in space comes from the dissipation and the leading term in time from the main residual. Therefore, the dissipation operator can finally be studied by expanding $\tilde{d}_{j,k}$ in space only, that is by keeping the time derivative continuous in the dissipation. In other words, a semi-discrete analysis of the dissipation (based on the expression (29) of the mid-point residuals) is sufficient for studying its effect on the scheme. For a better understanding of the role of the different contributions in the dissipation, this semi-discrete analysis will be done in two stages: in the following section, we restrict our attention to the role of the x -discretization in \tilde{r}_1 and of the role of the y -discretization in \tilde{r}_2 and, in the section after, we complete the analysis by adding the effect of the remaining spatial discrete terms.

3. Partial analysis of the RBC dissipation

3.1. Partial residual \tilde{r}_1^x

Here, we discretize at $(j + \frac{1}{2}, k)$ the x -derivative as in (21) and the x -average as in (22), but we keep continuous the y -derivative (we do not use (25)). Then the mid-point residual \tilde{r}_1 defined in (27) takes the partial form:

$$(\tilde{r}_1^x)_{j+\frac{1}{2},k} = \left[N_1^\mu \mu_1(w_t + g_y) + N_1^\delta \frac{\delta_1 f}{\delta x} \right]_{j+\frac{1}{2},k} = \left[(I + a^\mu \delta_1^2) \mu_1(w_t + g_y) + (I + a^\delta \delta_1^2) \frac{\delta_1 f}{\delta x} \right]_{j+\frac{1}{2},k} \quad (39)$$

We now carry out a Taylor expansion of \tilde{r}_1^x around $(j + \frac{1}{2})\delta x$. Provided the exact residual is sufficiently smooth, we obtain after some algebra:

$$N_1^\mu \mu_1 v = v + \frac{\delta x^2}{8} (1 + 8a^\mu) v_{xx} + \frac{\delta x^4}{384} (1 + 80a^\mu) v_{4x} + \frac{\delta x^6}{46080} (1 + 728a^\mu) v_{6x} + \mathcal{O}(\delta x^8), \quad (40)$$

where $v = w_t + g_y$ and,

$$N_1^\delta \frac{\delta_1 f}{\delta x} = f_x + \frac{\delta x^2}{24} (1 + 24a^\delta) f_{xxx} + \frac{\delta x^4}{1920} (1 + 240a^\delta) f_{5x} + \frac{\delta x^6}{322560} (1 + 2184a^\delta) f_{7x} + \mathcal{O}(\delta x^8). \quad (41)$$

For brevity, the subscript $(j + \frac{1}{2}, k)$ has been omitted. Summing (40) and (41) gives:

$$\begin{aligned} \tilde{r}_1^x = r + \frac{\delta x^2}{8} (1 + 8a^\mu) r_{xx} + \delta x^2 \left(a^\delta - a^\mu - \frac{1}{12} \right) f_{xxx} + \frac{\delta x^4}{384} (1 + 80a^\mu) r_{4x} + \frac{\delta x^4}{24} \left(3a^\delta - 5a^\mu - \frac{1}{20} \right) f_{5x} \\ + \frac{\delta x^6}{46080} (1 + 728a^\mu) r_{6x} + \frac{\delta x^6}{5760} \left(39a^\delta - 91a^\mu - \frac{3}{28} \right) f_{7x} + \mathcal{O}(\delta x^8) \end{aligned} \quad (42)$$

with the exact residual $r = w_t + f_x + g_y = v + f_x$. Since r is null everywhere for an exact unsteady solution, \tilde{r}_1^x at $(j + \frac{1}{2}, k)$ reduces to:

$$\tilde{r}_1^x = \delta x^2 \left(a^\delta - a^\mu - \frac{1}{12} \right) f_{xxx} + \frac{\delta x^4}{24} \left(3a^\delta - 5a^\mu - \frac{1}{20} \right) f_{5x} + \frac{\delta x^6}{5760} \left(39a^\delta - 91a^\mu - \frac{3}{28} \right) f_{7x} + \mathcal{O}(\delta x^8) \quad (43)$$

which no longer contains t and y -derivatives.

3.2. Partial residual \tilde{r}_2^y

Similarly, we restrict the mid-point residual \tilde{r}_2 to the partial form:

$$(\tilde{r}_2^y)_{j,k+\frac{1}{2}} = \left[N_2^\mu \mu_2(w_t + f_x) + N_2^\delta \frac{\delta_2 g}{\delta y} \right]_{j,k+\frac{1}{2}} = \left[(I + a^\mu \delta_2^2) \mu_2(w_t + f_x) + (I + a^\delta \delta_2^2) \frac{\delta_2 g}{\delta y} \right]_{j,k+\frac{1}{2}} \quad (44)$$

Carrying out similar Taylor expansions at $(j, k + \frac{1}{2})$ as above, we obtain

$$\tilde{r}_2^y = \delta y^2 \left(a^\delta - a^\mu - \frac{1}{12} \right) g_{yyy} + \frac{\delta y^4}{24} \left(3a^\delta - 5a^\mu - \frac{1}{20} \right) g_{5y} + \frac{\delta y^6}{5760} \left(39a^\delta - 91a^\mu - \frac{3}{28} \right) g_{7y} + \mathcal{O}(\delta y^8) \quad (45)$$

3.3. Partial dissipation $\tilde{d}^{x,y}$

Inserting the partial residuals (39) and (44) in the dissipation (20) gives the partial dissipation term:

$$\tilde{d}_{j,k}^{x,y} = \frac{1}{2} [\delta_1(\Phi_1 \tilde{r}_1^x) + \delta_2(\Phi_2 \tilde{r}_2^y)]_{j,k} \quad (46)$$

(a) For $a^\delta - a^\mu \neq \frac{1}{12}$, the partial mid-point residuals are $\mathcal{O}(h^2)$.

If

$$a^\delta = a^\mu = 0 \quad (47)$$

then these residuals reduce to 2-point formulas and from the expansions (43) and (45), we obtain:

$$\tilde{r}_1^x = -\frac{\delta x^2}{12} f_{xxx} + \mathcal{O}(\delta x^4), \quad \tilde{r}_2^y = -\frac{\delta y^2}{12} g_{yyy} + \mathcal{O}(\delta y^4) \quad (48)$$

Therefore the partial dissipation involves only 3×3 -points and expands as:

$$\tilde{d}^{x,y} = -\frac{1}{24} [\delta x^3 (\Phi_1 f_{xxx})_x + \delta y^3 (\Phi_2 g_{yyy})_y] + \mathcal{O}(h^5) \quad (49)$$

(b) For

$$a^\delta - a^\mu = \frac{1}{12} \quad \text{and} \quad a^\mu \neq \frac{1}{10} \quad (50)$$

the mid-point residuals are $\mathcal{O}(h^4)$ and expand as:

$$\tilde{r}_1^x = \frac{\delta x^4}{12} \left(\frac{1}{10} - a^\mu \right) f_{5x} + \mathcal{O}(\delta x^6), \quad \tilde{r}_2^y = \frac{\delta y^4}{12} \left(\frac{1}{10} - a^\mu \right) g_{5y} + \mathcal{O}(\delta y^6) \quad (51)$$

so that the partial dissipation is represented by:

$$\tilde{d}^{x,y} = \frac{1}{24} \left(\frac{1}{10} - a^\mu \right) [\delta x^5 (\Phi_1 f_{5x})_x + \delta y^5 (\Phi_2 g_{5y})_y] + \mathcal{O}(h^7) \quad (52)$$

(c) Finally, for

$$a^\mu = \frac{1}{10} \quad \text{and} \quad a^\delta = \frac{11}{60} \quad (53)$$

we obtain:

$$\tilde{r}_1^x = -\frac{\delta x^6}{2800} f_{7x} + \mathcal{O}(\delta x^8), \quad \tilde{r}_2^y = -\frac{\delta y^6}{2800} g_{7y} + \mathcal{O}(\delta y^8) \quad (54)$$

and

$$\tilde{d}^{x,y} = -\frac{1}{5600} [\delta x^7 (\Phi_1 f_{7x})_x + \delta y^7 (\Phi_2 g_{7y})_y] + \mathcal{O}(h^9) \quad (55)$$

In this first stage of the analysis, we have obtained the dissipation contribution due to the approximation in the main direction of each mid-point residual. This intermediate result is useful to simplify the global analysis, but above all it will be important for the interpretation of the χ -criterion for dissipation in Section 5.

4. Full analysis of the RBC dissipation

4.1. Residuals \tilde{r}_1 and \tilde{r}_2

We consider the complete mid-point residuals (27), (28). A Taylor expansion of \tilde{r}_1 can easily be obtained from the one of \tilde{r}_1^x by noting that (25) yields

$$N_2 \frac{\delta_2 \mu_2 g}{\delta y} = D_2 (g_y + \varepsilon_2), \quad \varepsilon_2 = \mathcal{O}(\delta y^{2p-2}) \quad (56)$$

so that \tilde{r}_1 can be related to \tilde{r}_1^x as

$$\tilde{r}_1 = D_2(\tilde{r}_1^x + N_1^\mu \mu_1 \varepsilon_2).$$

Since N_1^μ and D_2 are consistent with the identity plus second order terms, whereas \tilde{r}_1^x and ε_2 are $\mathcal{O}(h^{2p-2})$ for $p \geq 2$, we get

$$\tilde{r}_1 = \tilde{r}_1^x + \mu_1 \varepsilon_2 + \mathcal{O}(h^{2p}). \tag{57}$$

Similarly, using

$$N_1 \frac{\delta_1 \mu_1 f}{\delta x} = D_1(f_x + \varepsilon_1), \quad \varepsilon_1 = \mathcal{O}(\delta x^{2p-2}) \tag{58}$$

we get

$$\tilde{r}_2 = \tilde{r}_2^y + \mu_2 \varepsilon_1 + \mathcal{O}(h^{2p}). \tag{59}$$

The simple relations (57) and (59) are used below to obtain the full dissipation term \tilde{d} of the RBC_{2p-1} schemes, for $p = 2, 3$ and 4.

4.2. Dissipation for RBC3

The RBC3 scheme is constructed on a 3×3 -point stencil from the 4th-order main residual (15) and from a 3rd-order dissipation based on 2nd-order mid-point residuals defined by (29) with

$$a^\delta = a^\mu = a = c = 0, \tag{60}$$

that is

$$\begin{aligned} \tilde{r}_1 &= (I + b\delta_2^2) \left(\mu_1 w_t + \frac{\delta_1 f}{\delta x} \right) + \frac{\delta_2 \mu_2 \mu_1 g}{\delta y} \\ \tilde{r}_2 &= (I + b\delta_1^2) \left(\mu_2 w_t + \frac{\delta_2 g}{\delta y} \right) + \frac{\delta_1 \mu_1 \mu_2 f}{\delta x} \end{aligned} \tag{61}$$

the stencils of which use 2×3 and 3×2 -points, respectively. Thus, the RBC3 dissipation depends on the parameter b only. Its simplest form corresponds to $b = 0$. This choice was made in the first paper on residual-based schemes [1], focussed on the calculation of steady compressible flows. As we discussed in Section 2, any consistent approximation of d^{steady} is dissipative with the numerical viscosity matrices given in Section 2.4. However for unsteady problems, the RBC3 dissipation with $b = 0$ is not always dissipative as we will prove through the study of the dissipation for any b .

Let us now complete the expansion of the mid-point residuals (61) using the relations (57), (59) and the expansions (48) of \tilde{r}_1^x and \tilde{r}_2^y for $a^\delta = a^\mu = 0$. A classical expansion of the Pade approximations (56), (58) with $a = c = 0$ gives the error terms:

$$\varepsilon_1 = \left(\frac{1}{6} - b \right) \delta x^2 f_{xxx} + \mathcal{O}(\delta x^4), \quad \varepsilon_2 = \left(\frac{1}{6} - b \right) \delta y^2 g_{yyy} + \mathcal{O}(\delta y^4),$$

so that we obtain:

$$\begin{aligned} \tilde{r}_1 &= -\frac{\delta x^2}{12} f_{xxx} + \left(\frac{1}{6} - b \right) \delta y^2 g_{yyy} + \mathcal{O}(h^4) \\ \tilde{r}_2 &= -\frac{\delta y^2}{12} g_{yyy} + \left(\frac{1}{6} - b \right) \delta x^2 f_{xxx} + \mathcal{O}(h^4). \end{aligned}$$

Inserting these expansions in the definition (20) of \tilde{d} , we find the effective RBC3 dissipation term for unsteady problems:

$$\tilde{d} = -\kappa[\Phi_1(\delta x^3 f_{xxx} + \chi \delta x \delta y^2 g_{yyy})]_x - \kappa[\Phi_2(\delta y^3 g_{yyy} + \chi \delta y \delta x^2 f_{xxx})]_y + \mathcal{O}(h^5) \tag{62}$$

with the coefficients

$$\kappa = \frac{1}{24}, \quad \chi = 2(6b - 1).$$

The dissipative nature of this term depends on the parameter b and will be studied in the general framework of Section 5.

4.3. Dissipation for RBC5

The RBC5 scheme is constructed on a 5×5 -point stencil from the 6th-order main residual (13) with (16) and a 5th-order dissipation based on 4th-order mid-point residuals defined by (29) with a^δ and a^μ related by (50). For the error terms ε_1 and ε_2 in (56), (58) be $\mathcal{O}(h^4)$, we prescribe

$$b - a = \frac{1}{6} \tag{63}$$

which precisely gives:

$$\varepsilon_1 = \left(\frac{b}{6} - c - \frac{1}{30}\right) \delta x^4 f_{5x} + \mathcal{O}(\delta x^6)$$

and a similar expression for ε_2 . By plugging these results in Eqs. (57), (59) and using the expansions (51), we obtain:

$$\begin{aligned} \tilde{r}_1 &= \frac{1}{12} \left(\frac{1}{10} - a^\mu\right) \delta x^4 f_{5x} + \left(\frac{b}{6} - c - \frac{1}{30}\right) \delta y^4 g_{5y} + \mathcal{O}(h^6) \\ \tilde{r}_2 &= \frac{1}{12} \left(\frac{1}{10} - a^\mu\right) \delta y^4 g_{5y} + \left(\frac{b}{6} - c - \frac{1}{30}\right) \delta x^4 f_{5x} + \mathcal{O}(h^6) \end{aligned} \quad (64)$$

Inserting (64) in (20), we get the effective RBC5 dissipation term for unsteady problems:

$$\tilde{d} = \kappa[\Phi_1(\delta x^5 f_{5x} + \chi \delta x \delta y^4 g_{5y})]_x + \kappa[\Phi_2(\delta y^5 g_{5y} + \chi \delta y \delta x^4 f_{5x})]_y + \mathcal{O}(h^7) \quad (65)$$

with the coefficients

$$\kappa = \frac{1}{24} \left(\frac{1}{10} - a^\mu\right), \quad \chi = \frac{1}{2\kappa} \left(\frac{b}{6} - c - \frac{1}{30}\right).$$

and $a^\mu < 1/10$.

4.4. Dissipation for RBC7

The RBC7 scheme is constructed on a 5×5 -point stencil from the 8th-order main residual (13) with (17) and a 7th-order dissipation based on 6th-order mid-point residuals defined by (29) with a^ν and a^μ given by (53). For the error terms ε_1 and ε_2 in (56), (58) be $\mathcal{O}(h^6)$, we impose

$$b - a = \frac{1}{6}, \quad \frac{b}{6} - c = \frac{1}{30} \quad (66)$$

so that

$$\varepsilon_1 = \frac{1}{30} \left(\frac{1}{70} - c\right) \delta x^6 f_{7x} + \mathcal{O}(\delta x^8)$$

and similarly for ε_2 . By using the relations (57), (59) and the expansion (54), we obtain:

$$\begin{aligned} \tilde{r}_1 &= -\frac{\delta x^6}{2800} f_{7x} + \frac{1}{30} \left(\frac{1}{70} - c\right) \delta y^6 g_{7y} + \mathcal{O}(h^8) \\ \tilde{r}_2 &= -\frac{\delta y^6}{2800} g_{7y} + \frac{1}{30} \left(\frac{1}{70} - c\right) \delta x^6 f_{7x} + \mathcal{O}(h^8) \end{aligned} \quad (67)$$

Inserting (67) in (20), we get the effective RBC7 dissipation term for unsteady problems:

$$\tilde{d} = -\kappa[\Phi_1(\delta x^7 f_{7x} + \chi \delta x \delta y^6 g_{7y})]_x - \kappa[\Phi_2(\delta y^7 g_{7y} + \chi \delta y \delta x^6 f_{7x})]_y + \mathcal{O}(h^9) \quad (68)$$

with the coefficients

$$\kappa = \frac{1}{5600}, \quad \chi = \frac{280}{3} \left(c - \frac{1}{70}\right).$$

5. The χ -criterion for dissipation

5.1. Dissipation criterion

The effective dissipation term \tilde{d} induced by the discretization of the second-order partial differential operator (38) has been identified through the above expressions (62) for RBC3, (65) for RBC5 and (68) for RBC7. Owing to the residual-based structure of the dissipation, these expressions contain no time derivative. In some sense, the time derivatives have been replaced by space derivatives because the exact residual r and its derivatives are null everywhere.

The dissipation expressions (62), (65) and (68) can be cast in a general form. Consider a RBC $_{2p-1}$ scheme and denote partial derivatives as

$$f_{qx} = \frac{\partial^q f}{\partial x^q}, \quad g_{qy} = \frac{\partial^q g}{\partial y^q}, \quad q = 2p - 1,$$

the dissipation term is of the form:

$$\tilde{d} = d_q + \mathcal{O}(h^{q+2}) \tag{69}$$

with

$$d_q = (-1)^{p-1} \kappa \left\{ \delta x [\Phi_1 (\delta x^{q-1} f_{qx} + \chi \delta y^{q-1} g_{qy})]_x + \delta y [\Phi_2 (\delta y^{q-1} g_{qy} + \chi \delta x^{q-1} f_{qx})]_y \right\} \tag{70}$$

where $\kappa > 0$ and χ are two constant coefficients depending on the order q of the scheme.

To determine whether the multidimensional operator (70) is really dissipative or not, we proceed as in Section 2.4 for the operator d^{steady} . Considering the linear scalar case, (70) reduces to

$$d_q = \mathcal{D}_q w$$

with the linear partial differential operator:

$$\mathcal{D}_q = (-1)^{p-1} \kappa \left(\delta x^{2p-1} \Phi_1 A \frac{\partial^{2p}}{\partial x^{2p}} + \chi \delta x \delta y^{2p-2} \Phi_1 B \frac{\partial^{2p}}{\partial x \partial y^{2p-1}} + \delta y^{2p-1} \Phi_2 B \frac{\partial^{2p}}{\partial y^{2p}} + \chi \delta y \delta x^{2p-2} \Phi_2 A \frac{\partial^{2p}}{\partial y \partial x^{2p-1}} \right)$$

which contains the highest derivatives in (70).

All the derivatives in \mathcal{D}_q being even, its Fourier symbol is real. It is denoted by $\widehat{\mathcal{D}}_q(\xi, \eta)$, where ξ and η are the wave numbers (Fourier variables).

Again, by dissipation (in the broad sense), we mean:

$$\forall \xi \in \mathbb{R}, \quad \forall \eta \in \mathbb{R}, \quad \widehat{\mathcal{D}}_q(\xi, \eta) \leq 0. \tag{71}$$

Theorem 5.1 (χ -criterion). *The operator (70) is dissipative for any order $q = 2p - 1$ ($p \geq 2$), any advection direction (A, B) and any functions Φ_1, Φ_2 satisfying the conditions (34) if and only if $\chi = 0$.*

Proof. Since the Fourier symbol of a derivative like $\partial^{2p} / \partial x \partial y^{2p-1}$ is $(-1)^p \xi \eta^{2p-1}$, we get:

$$\widehat{\mathcal{D}}_q = -\kappa \delta x^{2p-1} \left[\Phi_1 A \xi^{2p} + \chi \left(\frac{\delta y}{\delta x} \right)^{2p-2} \Phi_1 B \xi \eta^{2p-1} + \chi \left(\frac{\delta y}{\delta x} \right) \Phi_2 A \xi^{2p-1} \eta + \left(\frac{\delta y}{\delta x} \right)^{2p-1} \Phi_2 B \eta^{2p} \right].$$

- For $A = B = 0$ (no advection), then $\widehat{\mathcal{D}}_q = 0$.
- For $A = 0$ and $B \neq 0$ (1-D advection), then from (20) $\Phi_1 = 0, \Phi_2 B \geq 0$ and

$$\widehat{\mathcal{D}}_q = -\kappa \delta y^{2p-1} \Phi_2 B \eta^{2p} \leq 0.$$

- Now for $A \neq 0, \widehat{\mathcal{D}}_q$ can be written as:

$$\widehat{\mathcal{D}}_q = -\kappa \delta x^{2p-1} \Phi_1 A \xi^{2p} \mathcal{D}$$

with

$$\mathcal{D} = 1 + \chi \left(\frac{\delta x}{\delta y} \frac{B}{A} \lambda^{2p-1} + \frac{\Phi_2}{\Phi_1} \lambda \right) + \frac{\delta x}{\delta y} \frac{\Phi_2 B}{\Phi_1 A} \lambda^{2p}$$

where $\lambda = \frac{\delta y \eta}{\delta x \xi}$.

Using the conditions (34) with the definition (35) of α , we have

$$\frac{\delta x B}{\delta y A} = \frac{\Phi_2}{\Phi_1} = \text{sgn}(AB) \alpha = \check{\alpha}$$

so that

$$\mathcal{D}(\lambda, \check{\alpha}) = 1 + \chi \check{\alpha} \lambda (\lambda^{2p-2} + 1) + \check{\alpha}^2 \lambda^{2p}$$

Since $\Phi_1 A$ is always positive, $\widehat{\mathcal{D}}_q$ is negative if and only if

$$\forall \check{\alpha} \in \mathbb{R}, \quad \forall \lambda \in \mathbb{R}, \quad \mathcal{D}(\lambda, \check{\alpha}) \geq 0.$$

Note that λ is a reduced wave number ratio and $\check{\alpha}$ characterizes the advection direction with respect to the mesh. It is not easy to discuss the sign of \mathcal{D} viewed as a polynomial of degree $2p$ in λ . It is better to begin with the dependency on $\check{\alpha}$. Considering \mathcal{D} as a polynomial of degree 2 in $\check{\alpha}$, we compute its discriminant

$$\Delta = \chi^2 \lambda^2 (\lambda^{2p-2} + 1)^2 - 4 \lambda^{2p} = \lambda^2 [\chi (\lambda^{2p-2} + 1) + 2 \lambda^{p-1}] [\chi (\lambda^{2p-2} + 1) - 2 \lambda^{p-1}]$$

\mathcal{D} is positive if and only if $\Delta \leq 0$. For $\lambda = 0, \Delta = 0$. For $\lambda \neq 0, \Delta \leq 0$ is tantamount to:

$$-\frac{2|\lambda|^{p-1}}{1+\lambda^{2p-2}} \leq \chi \leq \frac{2|\lambda|^{p-1}}{1+\lambda^{2p-2}}$$

Since the lower and upper bounds of χ tend to zero as λ tends to infinity, the above condition can be satisfied for all real numbers λ if and only if $\chi = 0$. \square

5.2. Interpretation of the χ -criterion

We now try to understand the meaning of the dissipation condition $\chi = 0$. When it holds, the mixed derivatives vanish and the dissipation operator reduces to

$$d_q = (-1)^{p-1} \kappa [\delta x^q (\Phi_1 f_{qx})_x + \delta y^q (\Phi_2 g_{qy})_y], \quad q = 2p - 1 \tag{72}$$

which is precisely the result obtained in the partial analysis of Section 3 where we only took into account the x -discretization in \tilde{r}_1 and the y -discretization in \tilde{r}_2 . This means that the effect of the complementary discretization comes at a higher order in the expansion of \tilde{d} . To ensure dissipation, the derivative g_y in the mid-point residual \tilde{r}_1 (resp. f_x in \tilde{r}_2) should be approximated at a higher order than necessary to get the accuracy required for the mid-point residual. Practically, the mid-point residual \tilde{r}_1 of RBC q scheme ($q = 2p - 1$) uses approximations of order $2p - 2$ for f_x and of order $2p$ for g_y . The latter is nothing but the one used in the main residual \tilde{r} . For the mid-point residual \tilde{r}_2 , the roles of f_x and g_y are exchanged.

Fortunately, this extra accuracy for half of the dissipation terms is achieved without extending the scheme stencil.

5.3. Application of the χ -criterion to RBC3

The mid-point residuals of the RBC3 scheme are defined by (61) and depend on the parameter b . Since $\chi = 2(6b - 1)$ for RBC3, the condition $\chi = 0$ for dissipation requires

$$b = \frac{1}{6} \tag{73}$$

For this value of b , g_y in \tilde{r}_1 and f_x in \tilde{r}_2 are approximated at order 4. Thus, the simplest choice $b = 0$ should not be used for unsteady problems. However, the correct choice (73) does not extend the scheme stencil.

5.4. Application of the χ -criterion to RBC5 and RBC7

The mid-point residuals of the RBC5 and RBC7 schemes are defined by (29) with

$$a^\delta = a^\mu + \frac{1}{12} \tag{74}$$

and

$$b - a = \frac{1}{6}.$$

For RBC5, $a^\mu \neq 1/10$. In addition, the condition $\chi = 0$ for dissipation gives

$$\frac{b}{6} - c = \frac{1}{30},$$

that is

$$a = \frac{1}{30} + 6c, \quad b = \frac{1}{5} + 6c. \tag{75}$$

When (75) holds, g_y in \tilde{r}_1 and f_x in \tilde{r}_2 are approximated at order 6. Note that the correct dissipation depends on two parameters (a^μ and c), but the dissipation coefficient κ in (65) depends only on a^μ (that should be strictly lower than $1/10$).

In the previous works [4,5], the following coefficients were used:

$$a^\mu = \frac{1}{12}, \quad a^\delta = \frac{1}{6}, \quad a = \frac{1}{10}, \quad b = \frac{4}{15}, \quad c = \frac{1}{90}. \tag{76}$$

They happen to satisfy (74) and (75).

Another correct choice satisfying the same conditions and producing the same dissipation (same a^μ) for RBC5 is:

$$a^\mu = \frac{1}{12}, \quad a^\delta = \frac{1}{6}, \quad a = \frac{1}{30}, \quad b = \frac{1}{5}, \quad c = 0. \tag{77}$$

It is simpler because the operators δ_1^4 and δ_2^4 vanish in \tilde{r}_1 and \tilde{r}_2 .

Finally for RBC7, the coefficients are given by (74) and (75) along with $a^\mu = 1/10$. Here the condition $\chi = 0$ for dissipation gives $c = 1/70$, and a unique set of coefficients is found, leading to a dissipative RBC7 scheme. This is:

$$a^\mu = \frac{1}{10}, \quad a^\delta = \frac{11}{60}, \quad a = \frac{5}{42}, \quad b = \frac{2}{7}, \quad c = \frac{1}{70}. \tag{78}$$

Here, g_y in \tilde{r}_1 and f_x in \tilde{r}_2 are approximated at order 8. Note that the coefficients are different from those used in [4,5] for RBC7, that were:

$$a^\mu = \frac{1}{10}, \quad a^\delta = \frac{11}{60}, \quad a = \frac{1}{30}, \quad b = \frac{1}{5}, \quad c = 0. \tag{79}$$

Actually, the coefficients (79) do not ensure dissipation to RBC7 for all flow conditions.

6. Extension to 3-D

6.1. Dissipation criterion in 3-D

Consider now the three-dimensional hyperbolic system

$$w_t + f_x + g_y + e_z = 0 \tag{80}$$

where $e = e(w)$ is the flux component in the z-direction and $C = de/dw$. Using difference and average operators δ_3 and μ_3 on the discrete axis $z_l = l\delta z$, the residual-based scheme reads:

$$\tilde{r}_{j,k,l} = \tilde{d}_{j,k,l} \tag{81}$$

where the main residual \tilde{r} is a space-centered approximation of the exact residual r (the left-hand side of (80)) and \tilde{d} is the dissipation term

$$\tilde{d}_{j,k,l} = \frac{1}{2} [\delta_1(\phi_1 \tilde{r}_1) + \delta_2(\phi_2 \tilde{r}_2) + \delta_3(\phi_3 \tilde{r}_3)]_{j,k,l} \tag{82}$$

which approximates:

$$d = \frac{\delta x}{2} (\phi_1 r)_x + \frac{\delta y}{2} (\phi_2 r)_y + \frac{\delta z}{2} (\phi_3 r)_z \tag{83}$$

For steady problems, the time derivative is cancelled into (82) and \tilde{d} is consistent with d^{steady} defined from (83) after replacing r by

$$r^{steady} = f_x + g_y + e_z$$

Conditions on Φ_1, Φ_2 and Φ_3 should be satisfied for the operator d^{steady} be actually dissipative in the broad sense. For a scalar problem, these conditions are found to be

$$\begin{aligned} \Phi_1 A \geq 0, \quad \Phi_2 B \geq 0, \quad \Phi_3 C \geq 0, \\ \delta x \Phi_1 B = \delta y \Phi_2 A, \quad \delta x \Phi_1 C = \delta z \Phi_3 A. \end{aligned} \tag{84}$$

With the following notations:

$$\begin{aligned} \Phi_1 = \text{sgn}(A)\phi, \quad \Phi_2 = \text{sgn}(B)\psi, \quad \Phi_3 = \text{sgn}(C)\zeta, \\ \alpha = \frac{\delta x|B|}{\delta y|A|}, \quad \beta = \frac{\delta x|C|}{\delta z|A|}, \end{aligned}$$

the dissipation conditions (84) for d^{steady} become

$$\phi \geq 0, \quad \psi = \alpha\phi, \quad \zeta = \beta\phi. \tag{85}$$

An optimal choice is

$$\begin{aligned} \phi &= \min\left(1, \frac{1}{\alpha}, \frac{1}{\beta}\right) = \min\left(1, \frac{\delta y|A|}{\delta x|B|}, \frac{\delta z|A|}{\delta x|C|}\right), \\ \psi &= \min\left(1, \alpha, \frac{\alpha}{\beta}\right) = \min\left(1, \frac{\delta x|B|}{\delta y|A|}, \frac{\delta z|B|}{\delta y|C|}\right), \\ \zeta &= \min\left(1, \beta, \frac{\beta}{\alpha}\right) = \min\left(1, \frac{\delta x|C|}{\delta z|A|}, \frac{\delta y|C|}{\delta z|B|}\right). \end{aligned} \tag{86}$$

For hyperbolic systems, Φ_1, Φ_2 and Φ_3 are matrix functions defined from the above relations by following a process perfectly similar to the one described in Section 2.4 for the bidimensional case.

To construct and analyse the dissipation term (82), we discretize the mid-point residuals \tilde{r}_1 , \tilde{r}_2 and \tilde{r}_3 in two stages as in the 2-D case. First, we discretize \tilde{r}_p in the p^{th} -direction only. For instance for \tilde{r}_1 , this leads to the partially discrete residual

$$(\tilde{r}_1^x)_{j+\frac{1}{2},k,l} = \left[N_1^\mu \mu_1 (w_t + g_y + e_z) + N_1^\delta \frac{\delta_1 f}{\delta x} \right]_{j+\frac{1}{2},k,l}$$

where the operator polynomials N_1^μ and N_1^δ are still defined by (23) with $m = 1$. Clearly, the expansions made in Section 3.1 remains valid with the new definition of v :

$$v = w_t + g_y + e_z.$$

As a result the expansions of \tilde{r}_1^x and \tilde{r}_2^y are unchanged in 3-D and we just have to introduce a similar expansion for \tilde{r}_3^z .

Then in the second stage, we proceed as in Section 4 to complete the space discretization, which requires to introduce a Pade approximation for e_z :

$$N_3 \frac{\delta_3 \mu_3 e}{\delta z} = D_3 (e_z + \varepsilon_3), \quad \varepsilon_3 = \mathcal{O}(\delta z^{2p-2}) \quad (87)$$

where N_3 and D_3 are defined by (26) for $m = 3$ and ε_3 is the error term.

After applying the operator product $D_2 D_3$ to all the terms involved in \tilde{r}_1 , we define:

$$(\tilde{r}_1)_{j+\frac{1}{2},k,l} = \left[N_1^\mu \mu_1 \left(D_2 D_3 w_t + D_3 N_2 \frac{\delta_2 \mu_2 g}{\delta y} + D_2 N_3 \frac{\delta_3 \mu_3 e}{\delta z} \right) + N_1^\delta D_2 D_3 \frac{\delta_1 f}{\delta x} \right]_{j+\frac{1}{2},k,l} \quad (88)$$

and similarly:

$$(\tilde{r}_2)_{j,k+\frac{1}{2},l} = \left[N_2^\mu \mu_2 \left(D_3 D_1 w_t + D_3 N_1 \frac{\delta_1 \mu_1 f}{\delta x} + D_1 N_3 \frac{\delta_3 \mu_3 e}{\delta z} \right) + N_2^\delta D_3 D_1 \frac{\delta_2 g}{\delta y} \right]_{j,k+\frac{1}{2},l} \quad (89)$$

$$(\tilde{r}_3)_{j,k,l+\frac{1}{2}} = \left[N_3^\mu \mu_3 \left(D_1 D_2 w_t + D_2 N_1 \frac{\delta_1 \mu_1 f}{\delta x} + D_1 N_2 \frac{\delta_2 \mu_2 g}{\delta y} \right) + N_3^\delta D_1 D_2 \frac{\delta_3 e}{\delta z} \right]_{j,k,l+\frac{1}{2}} \quad (90)$$

A Taylor expansion of \tilde{r}_1 can be easily obtained by noting that the formulas (56) and (87) yield

$$D_3 N_2 \frac{\delta_2 \mu_2 g}{\delta y} = D_2 D_3 (g_y + \varepsilon_2), \quad D_2 N_3 \frac{\delta_3 \mu_3 e}{\delta z} = D_2 D_3 (e_z + \varepsilon_3)$$

so that \tilde{r}_1 is related to \tilde{r}_1^x by

$$\tilde{r}_1 = D_2 D_3 [\tilde{r}_1^x + N_1^\mu \mu_1 (\varepsilon_2 + \varepsilon_3)].$$

With the same arguments as in the 2-D case, we can reduce this relation to

$$\tilde{r}_1 = \tilde{r}_1^x + \mu_1 (\varepsilon_2 + \varepsilon_3) + \mathcal{O}(h^{2p})$$

and similar expressions for \tilde{r}_2 and \tilde{r}_3 , the error ε_3 being quite similar to the error ε_1 and ε_2 .

Finally, we easily find the general form of the dissipation in 3-D. With the same notations as in Section 5.1, the dissipation of a RBCq scheme ($q = 2p - 1$) is given by (69) with

$$d_q = (-1)^{p-1} \kappa \{ \delta x [\Phi_1 (\delta x^{q-1} f_{qx} + \chi \delta y^{q-1} g_{qy} + \chi \delta z^{q-1} e_{qz})]_x + \delta y [\Phi_2 (\delta y^{q-1} g_{qy} + \chi \delta z^{q-1} e_{qz} + \chi \delta x^{q-1} f_{qx})]_y + \delta z [\Phi_3 (\delta z^{q-1} e_{qz} + \chi \delta x^{q-1} f_{qx} + \chi \delta y^{q-1} g_{qy})]_z \} \quad (91)$$

and exactly the same coefficients κ and χ as in 2-D (these coefficients depend only of the order q of the RBCq scheme).

The definition (71) of dissipation is unchanged except by the adding of a third wave number.

Theorem 6.1 (χ -criterion). *The operator (91) is dissipative for any order $q = 2p - 1$ ($p \geq 2$), any advection direction (A, B, C) and any functions Φ_1, Φ_2, Φ_3 satisfying the conditions (84) if and only if $\chi = 0$.*

Proof. The condition $\chi = 0$ is sufficient for dissipation since when it holds, the operator d_q has no crossed derivatives and reads

$$d_q = (-1)^{p-1} \kappa \{ \delta x^q (\Phi_1 f_{qx})_x + \delta y^q (\Phi_2 g_{qy})_y + \delta z^q (\Phi_3 e_{qz})_z \} \quad (92)$$

which is always dissipative provided $\Phi_1 A \geq 0$, $\Phi_2 B \geq 0$ and $\Phi_3 C \geq 0$.

The condition $\chi = 0$ is also necessary since it does in a 2-D situation as we have shown in Section 5.

The above χ -criterion can be interpreted similarly as in 2-D: a RBCq scheme ($q = 2p - 1$) satisfying this condition has a mid-point residual \tilde{r}_1 (resp. \tilde{r}_2, \tilde{r}_3) in which the derivative f_x (resp. g_y, e_z) is approximated at order $2p - 2$ and the two other derivatives are approximated at order $2p$. \square

6.2. Application of the χ -criterion to RBC3 in 3-D

In 3-D, the RBC3 dissipation involves $3 \times 3 \times 3$ -points and is given by (82) with a $2 \times 3 \times 3$ -point residual $(\tilde{r}_1)_{j+\frac{1}{2},k,l}$ defined as:

$$\tilde{r}_1 = (I + b\delta_2^2)(I + b\delta_3^2) \left(\mu_1 w_t + \frac{\delta_1 f}{\delta x} \right) + (I + b\delta_3^2) \frac{\delta_2 \mu_2 g}{\delta y} + (I + b\delta_2^2) \frac{\delta_3 \mu_3 e}{\delta z} \tag{93}$$

along with a $3 \times 2 \times 3$ -point residual $(\tilde{r}_2)_{j,k+\frac{1}{2},l}$ and a $3 \times 3 \times 2$ -point residual $(\tilde{r}_3)_{j,k,l+\frac{1}{2}}$ deduced from (93) by circular permutations of the subscripts 1,2,3, of the flux components f, g, e and of the space steps $\delta x, \delta y, \delta z$.

Since the coefficient χ is the same in 3-D as in 2-D, the dissipation condition is again (73), which sets the only parameter b .

6.3. Application of the χ -criterion to RBC5 and RBC7 in 3-D

The $5 \times 5 \times 5$ -point RBC5 and RBC7 schemes are defined by (82) with

$$\begin{aligned} \tilde{r}_1 = & (I + b\delta_2^2 + c\delta_3^4)(I + b\delta_3^2 + c\delta_2^4) \left[(I + a^\mu \delta_1^2) \mu_1 w_t + (I + a^\delta \delta_1^2) \frac{\delta_1 f}{\delta x} \right] \\ & + (I + a^\mu \delta_1^2) \mu_1 \left[(I + b\delta_3^2 + c\delta_2^4)(I + a\delta_2^2) \frac{\delta_2 \mu_2 g}{\delta y} + (I + b\delta_2^2 + c\delta_3^4)(I + a\delta_3^2) \frac{\delta_3 \mu_3 e}{\delta z} \right] \end{aligned} \tag{94}$$

and similar expressions for \tilde{r}_2 and \tilde{r}_3 deduced by circular permutations of the subscripts, the flux components and the space steps.

The coefficients corresponding to the RBC5 dissipation satisfying the condition $\chi = 0$ are given by (74) and (75). They still depend on the two parameters a^μ and c . The analogous coefficients for RBC7 are uniquely set by (78).

7. Numerical experiments

In the following, we carry out some numerical tests to check the effect of satisfying or not the χ -criterion. We consider the RBC3 scheme with a dissipation operator corresponding to $b = 0$ (χ -criterion violated) or $b = 1/6$ (χ -criterion satisfied) and the RBC7 scheme with the coefficients (79) associated to $c = 0$ (χ -criterion violated) or the coefficients (78) associated to $c = 1/70$ (χ -criterion satisfied).

For the present computations, the time derivative in the main residual and in the mid-point residuals is discretized by a linear multistep method of order two, which is A-stable -see Formula (19). Such a method being fully implicit, it is solved by using a dual-time stepping approach.

7.1. Diagonal advection of a sine wave

In the proof of Theorem 5.1, we have found that the multidimensional dissipation property is equivalent to the positivity of the function $\mathcal{D}(\lambda, \check{\alpha})$. Consider the simple case of an advection directed in the first mesh diagonal, that is $\check{\alpha} = 1$ and consider any sinusoidal wave propagating along this diagonal, that is $\delta y \eta = \delta x \xi$, or $\lambda = 1$. Then

$$\mathcal{D}(1, 1) = 2(1 + \chi)$$

for any order of the RBC scheme.

Clearly $\mathcal{D}(1, 1)$ is positive for $\chi = 0$, but it is negative for the values corresponding to RBC3 ($b = 0$) and RBC7 ($c = 0$) as summarized in Table 1. So this diagonal advection is an interesting situation to investigate numerically.

We thus consider the initial-value problem:

$$\begin{cases} w_t + w_x + w_y = 0 \\ w(x, y, 0) = \sin(2\pi(x + y)), \quad -1 \leq x \leq 1, \quad -1 \leq y \leq 1, \end{cases}$$

with periodic boundary conditions. The initial condition is shown on Fig. 2. In the diagonal direction, the wavelength is $\sqrt{2}/2$ and the advection speed is $\sqrt{2}$. The computational domain $[-1, 1]^2$ is discretized by 25×25 square cells ($\delta x = \delta y = 0.08$), which corresponds to 12.5 points per wavelength. The time step is rather small: $\Delta t / \delta x = 0.05$. The time evolution of the amplitude is notably amplified when the χ -criterion is violated as shown in Fig. 3. This amplification is faster with RBC7 than with RBC3. On the contrary the sine wave is damped out when the χ -criterion is satisfied. The damping is very small for RBC7: 1.6% after 5000 time-iterations ($t = 20$), corresponding to a diagonal advection over a distance of 40 wavelengths.

7.2. Converging cylindrical shock

When the χ -criterion is violated in an RBC scheme, the lack of dissipation occurs in some oblique flow directions. So, we consider a test case involving all the flow directions and a large range of wave numbers, that is a 2-D simulation of a con-

Table 1

Dissipation characteristics for a diagonal advection.

	κ	χ	$\mathcal{D}(1,1)$
RBC3 ($b = 0$)	$\frac{1}{24}$	-2	-2
RBC3 ($b = \frac{1}{6}$)	$\frac{1}{24}$	0	2
RBC7 ($c = 0$)	$\frac{1}{5600}$	$-\frac{4}{3}$	$-\frac{2}{3}$
RBC7 ($c = \frac{1}{70}$)	$\frac{1}{5600}$	0	2

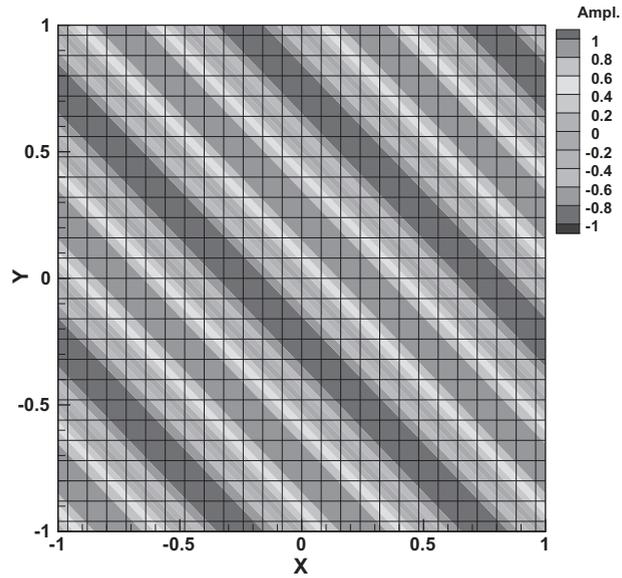
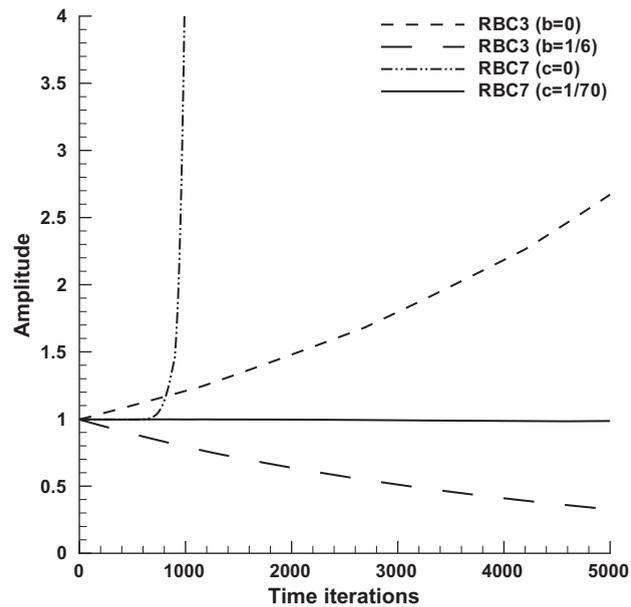
Fig. 2. Initial wave on the 25×25 mesh.

Fig. 3. Wave amplitude versus time-iterations.

verging cylindrical shock on a uniform Cartesian mesh. Of course, this axisymmetric problem could be solved more easily as a 1-D problem in polar coordinates. Here, the 2-D Euler equations, for a perfect gas with a specific heat ratio $\gamma = 1.4$, are solved in a square domain $[-0.5, 0.5]^2$.

At time $t = 0$, a cylindrical shock (satisfying the Rankine-Hugoniot relations) lies on a circle of center $(x, y) = (0, 0)$ and radius $r_0 = 0.25$. Inside the cylindrical shock (state 0), the fluid is at rest and at pressure p_0 . The pressure just behind the shock is $p_1 = 2.4p_0$ at $t = 0$. Outside the cylindrical shock, the initial state corresponds to a steady converging flow, i.e. the flow at a radius $r > r_0$ is related to the state 1 just behind the shock by the conservation of mass ($\rho Vr = \rho_1 V_1 r_0$ where ρ is the density and V the radial velocity), the conservation of total enthalpy and of the entropy. For improving the initial representation of the shock on the Cartesian mesh, the vector w of conservative variables is defined as follows in the mesh cells intersecting the shock:

$$w_* = (1 - \theta)w_0 + \theta w_1, \quad 0 \leq \theta \leq 1$$

where $\theta \delta x \delta y$ is the cell area fraction in state 1.

During the evolution, the cylindrical shock increases in strength as it converges towards the axis. When the shock reaches the axis, it is reflected as a divergent shock. At the very instant of reflection, the pressure at the axis becomes infinite in the

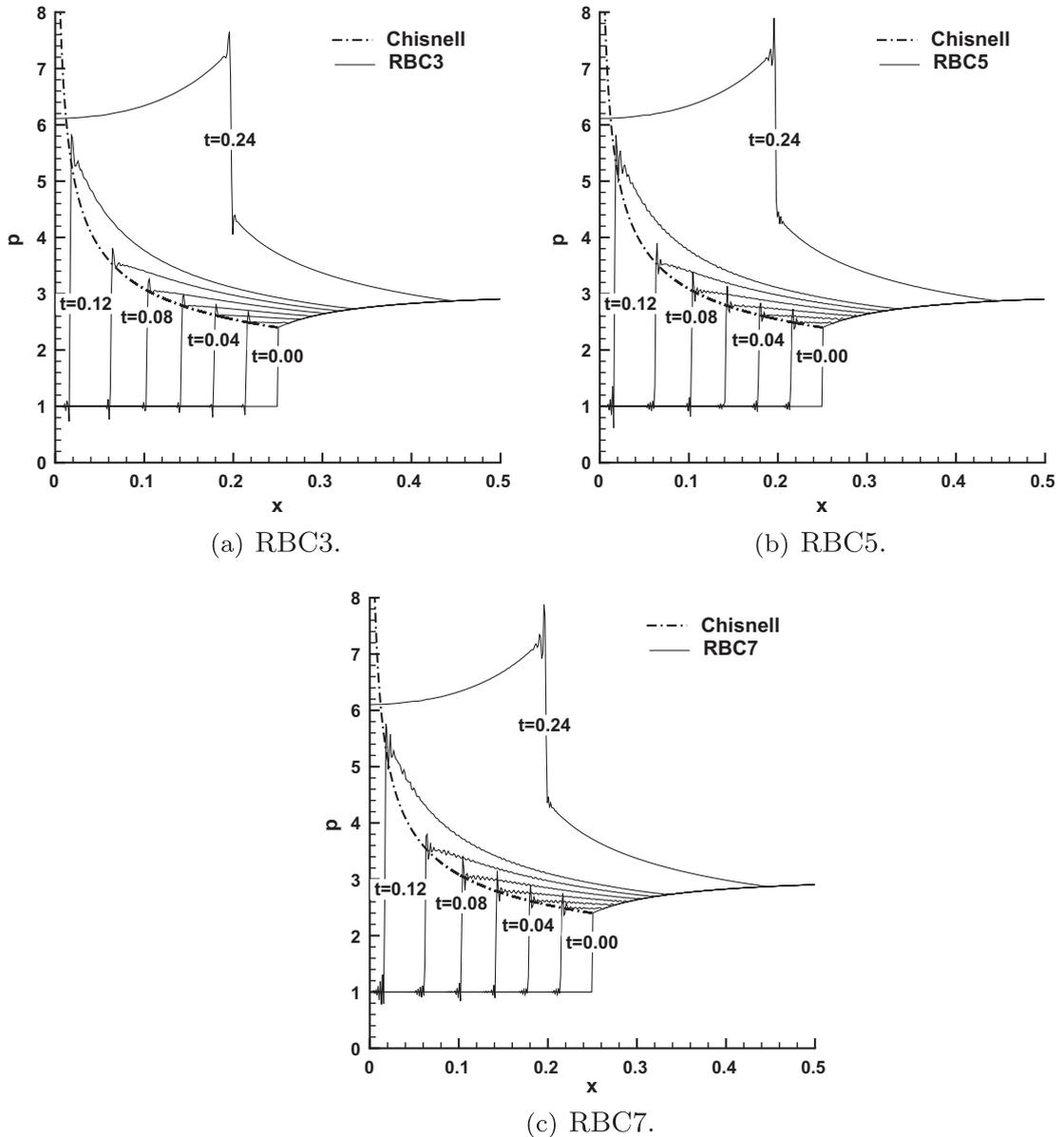


Fig. 4. Pressure along the x-axis at different times for RBC schemes satisfying the χ -criterion.

Euler model. To avoid a numerical difficulty, the Cartesian mesh is set so that the axis corresponds to a cell vertex and not to a cell center. This prevents the computation of any unphysical quantity on the axis. Note also that the outside boundary is not affected by the perturbations coming from the shock motion in the duration of the present simulation.

Chisnell [22] gave in 1957 an analytical estimation of the pressure behind a moving cylindrical shock, the theoretical arguments of which were improved by Whitham [23]. According to this theory, the Mach number M of the shock wave (relative to the fluid at rest) at radius r is solution of the differential equation:

$$\frac{dM}{dr} = -\frac{1}{r} \frac{(M^2 - 1)K(M)}{2M} \quad (95)$$

where

$$K(M) = 2 \left[\left(1 + \frac{2}{\gamma + 1} \frac{1 - \mu^2}{\mu} \right) \left(2\mu + 1 + \frac{1}{M^2} \right) \right]^{-1}$$

$$\mu = \left[\frac{(\gamma - 1)M^2 + 2}{2\gamma M^2 - (\gamma - 1)} \right]^{\frac{1}{2}}$$

For $\gamma = 1.4$, the function $K(M)$ decreases slowly from 0.5 for $M = 1$ to $14/(17 + 7\sqrt{7}) \approx 0.394$ for $M \rightarrow \infty$. Starting from the initial condition $M_0 = M(r_0)$, the Eq. (95) can easily be solved numerically with a high accuracy. An exact solution is also available [22], but its expression is very complicated and defined in the form $r = r(M)$.

The pressure behind the shock in motion is deduced from $M = M(r)$ using the Rankine–Hugoniot relations:

$$p_1 = \frac{2\gamma M^2 - (\gamma - 1)}{\gamma + 1} p_0 \quad (96)$$

The converging cylindrical shock problem is solved by the RBC schemes on a 800×800 Cartesian mesh with $\Delta t/\delta x = 0.21$. When the χ -criterion is violated, the computation fails after a few time iterations (one iteration for RBC3 with $b = 0$ and 26 for RBC7 with $c = 0$). When the χ -criterion is satisfied, the computation succeeds, even after the shock reflection on the axis. In this case, the pressure profiles along the x -axis are shown on Fig. 4 for the RBC3, RBC5 and RBC7 schemes at different times, together with the analytical pressure behind the shock deduced from (95), (96). The agreement between the numerical solution and the Chisnell theory is very good. The shape of the converging shock computed by the RBC7 scheme at different times is shown on Fig. 5. This shape has been defined as the isobar lines of level $\frac{1}{2}(p_1 + p_0)$ at each time. The converging shock appears to be perfectly circular on the Cartesian mesh. Clearly, Fig. 4 reveals the oscillatory nature of the shock profiles computed by the present high order schemes, specially by RBC5 and RBC7. It should be noted that the computations have been achieved by a strict use of the method described in the present paper: there is no limiter, no entropy correction, no filtering or other additive. In these conditions, it appears that a good design of the dissipative operator allows the calculation of a difficult test case, even if the discrete shock is oscillatory and could be improved.

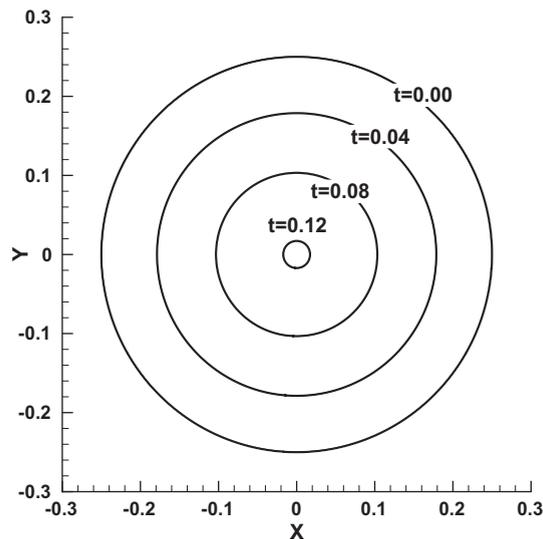


Fig. 5. Shape of the converging shock at different times computed by RBC7.

8. Conclusion

A comprehensive study of the dissipation properties of a family of residual-based compact schemes has been presented for 2-D and 3-D hyperbolic systems of conservation laws. The residual-based numerical dissipation operator has been shown to be the counterpart of a high-order differential operator based on pure and mixed derivatives of even order. A general criterion (Theorem 5.1 and 6.1) has been established for this operator to be dissipative.

Numerical tests confirm the theoretical results and demonstrate the importance of a well-designed dissipation operator for numerical simulations in gas dynamics. Specifically, the present residual-based formulation ensures controlled damping of sine waves propagating along any direction with respect to the computational mesh and with any advection speed. It also allows the computation of unsteady multidimensional flows with strong shocks without any treatment for shock capturing. Further work is in progress to investigate the need for a suitable residual-based correction to suppress spurious oscillations around strong flow discontinuities.

In future works, it might be interesting to consider a more general form of residual-based dissipation based on a space-time approach. In the two-dimensional case, this consists in adding the new term

$$\frac{\Delta t}{2} [\Phi_0(w_t + f_x + g_y)]_t$$

to the present dissipation (38) and in discretizing it with a new mid-point residual $(\bar{r}_0)_{jk}^{n+\frac{1}{2}}$. The resulting scheme would be more complicated, but could have some advantages, notably for increasing time accuracy.

Acknowledgements

This research has been done within the framework of the European project IDIHOM (Industrialization of High Order Methods) which aims to promote the use of high-order numerical methods by the European aerospace industry.

References

- [1] A. Lerat, C. Corre, A residual-based compact scheme for the compressible Navier–Stokes equations, *Journal of Computational Physics* 170 (2001) 642–675.
- [2] A. Lerat, C. Corre, Residual-based compact schemes for multidimensional hyperbolic systems of conservation laws, *Computers and Fluids* 31 (2002) 639–661.
- [3] C. Corre, G. Hanss, A. Lerat, A residual-based compact scheme for the unsteady compressible Navier–Stokes equations, *Computers and Fluids* 34 (2005) 561–580.
- [4] A. Lerat, C. Corre, Higher order residual-based compact schemes on structured grids, in: 34th Computing Fluid Dynamics Course, von Karman Institute for Fluid Dynamics, VKI LS 2006-1, pp. 1–111.
- [5] C. Corre, F. Falissard, A. Lerat, High-order residual-based compact schemes for compressible inviscid flows, *Computers and Fluids* 36 (2007) 1567–1582.
- [6] C. Corre, A. Lerat, High-order residual-based compact schemes for advection–diffusion problems, *Computers and Fluids* 37 (2008) 505–519.
- [7] B. Michel, P. Cinnella, A. Lerat, Multiblock residual-based compact schemes for the computation of complex turbomachinery flows, *International Journal of Engineering Systems Modelling and Simulation* 3 (2011).
- [8] A. Lerat, P. Cinnella, B. Michel, F. Falissard, High-order residual-based compact schemes for aerodynamics and aeroacoustics, *Computers and Fluids* 61 (2012) 31–38.
- [9] M. Ricchiuto, A. Csik, H. Deconinck, Residual distribution for general time dependent conservation laws, *Journal of Computational Physics* 209 (2005) 249–289.
- [10] R. Abgrall, Residual distribution schemes: current status and future trends, *Computers and Fluids* 35 (2006) 641–669.
- [11] H. Deconinck, M. Ricchiuto, Residual distribution schemes: foundation and analysis, *Encyclopedia of Computational Mechanics*, John Wiley and Sons Ltd, 2007.
- [12] R. Abgrall, A. Larat, M. Ricchiuto, C. Tavé, A simple construction of very high order non-oscillatory compact schemes on unstructured meshes, *Computers and Fluids* 38 (2009) 1314–1323.
- [13] S.K. Lele, Compact finite difference schemes with spectral-like resolution, *Journal of Computational Physics* 103 (1992) 16–42.
- [14] B. Cockburn, C.-W. Shu, Nonlinear stable compact schemes for shock calculations, *SIAM Journal of Numerical Analysis* (1994) 607–627.
- [15] H.C. Yee, Explicit and implicit multidimensional compact high-resolution shock-capturing methods: formulation, *Journal of Computational Physics* 131 (1997) 216–232.
- [16] M.R. Visbal, D.V. Gaitonde, High-order accurate methods for complex unsteady subsonic flows, *AIAA Journal* 37 (1999) 1231–1239.
- [17] A.I. Tolstykh, *High Accuracy Non-centered Compact Difference Schemes for Fluid Dynamics Applications*, World Scientific, Singapore, 1994.
- [18] D. Fu, Y. Ma, A high order accurate difference scheme for complex flow fields, *Journal of Computational Physics* 134 (1997) 1–15.
- [19] G. Dahlquist, A special stability problem for linear multistep methods, *BIT* 3 (1963) 27–43.
- [20] O.B. Wildlund, A note on unconditionally stable linear multistep methods, *BIT* 7 (1967) 65–70.
- [21] J.D. Lambert, *Numerical Methods for Ordinary Differential Systems: The Initial Value Problem*, Wiley Edition, 1991.
- [22] R.F. Chisnell, The motion of a shock wave in a channel with applications to cylindrical and spherical shock waves, *Journal of Fluid Mechanics* 2 (1957) 286–298.
- [23] G.B. Whitham, On the propagation of shock waves through regions of non uniform area or flow, *Journal of Fluid Mechanics* 4 (1958) 337–360.