INTERNATIONAL JOURNAL FOR NUMERICAL METHODS IN FLUIDS
*Int. J. Numer. Meth. Fluids* 2004; **44**:1119–1162 (DOI: 10.1002/fld.690)

# High resolution Godunov-type schemes with small stencils

Vincent Guinot*,†

*Université Montpellier 2, Maison des Sciences de l'Eau, 34095 Montpellier Cedex 5, France*

SUMMARY

Higher-order Godunov-type schemes have to cope with the following two problems: (i) the increase in the size of the stencil that make the scheme computationally expensive, and (ii) the monotony-preserving treatments (limiters) that must be implemented to avoid oscillations, leading to strong damping of the solution, in particular linear waves (e.g. acoustic waves). When too compressive, limiting procedures may also trigger the instability of oscillatory numerical solutions (e.g. in advection–dispersion phenomena) via the artificial amplification of the shorter modes. The present paper proposes a new approach to carry out the reconstruction. In this approach, the values of the flow variable at the edges of the computational cells are obtained directly from the reconstruction within these cells. This method is applied to the MUSCL and DPM schemes for the solution of the linear advection equation. The modified DPM scheme can capture contact discontinuities within one computational cell, even after millions of time steps at Courant numbers ranging from 1 to values as low as $10^{-4}$. Linear waves are subject to negligible damping. Application of the method to the DPM for one-dimensional advection–dispersion problems shows that the numerical instability of oscillatory solutions caused by the over compressive, original DPM limiter is eliminated. One- and two-dimensional shallow water simulations show an improvement over classical methods, in particular for two-dimensional problems with strongly distorted meshes. The quality of the computational solution in the two-dimensional case remains acceptable even for mesh aspect ratios $\Delta x/\Delta y$ as large as 10. The method can be extend to the discretization of higher-order PDEs, allowing third-order space derivatives to be discretized using only two cells in space. Copyright © 2004 John Wiley & Sons, Ltd.

KEY WORDS: Godunov-type schemes; reconstruction; reduced stencils; conservation laws; advection–dispersion; shallow water equations

## 1. INTRODUCTION

The dissipative character of Godunov-type schemes makes them difficult to use for problems involving both strong shocks and quasi-linear waves (e.g. acoustic waves). This can be overcome to some extent by increasing the size of the stencil to gain accuracy in variable reconstruction (see e.g. Van Leer's MUSCL [1], Ben-Artzi and Falcowitz's scheme [2], Colella

*Correspondence to: V. Guinot, Maison des Sciences de l'Eau, Université Montpellier 2, 34095 Montpellier Cedex 5, France.
†E-mail: guinot@msem.univ.montp2.fr

*Received September 2001*
*Revised 22 July 2003*

and Woodward's PPM [3]). This however leads to other problems, in particular (i) the time-consuming character of the scheme, induced by the complexity of the formulae used for the reconstruction (see e.g. Reference [3], that is most probably one of the most accurate, but also complex, Godunov-type schemes known to date), and (ii) the spatial extent of the stencil that leads to extra complexity in the treatment of boundary conditions. Moreover, increasing the size of the stencil is not a solution when discontinuities are present in the solution, because the values of the flow variable at the interfaces between the computational cells are computed using a continuous interpolation across cells where the profile is discontinuous. Also, high-order interpolation yields valuable results when applied to smooth profiles, but may induce oscillations in the presence of discontinuities. Limiters must be used to prevent such undesirable behaviour (see References [1–3] and [4, 5] for recent developments on limiters). Although satisfactory when purely hyperbolic or parabolic PDEs are to be solved, classical profile reconstruction and limiting may lead to unacceptable degradation of the quality of the solution when PDEs with oscillatory solutions (such as advection–dispersion equations) are to be solved. A first source of inaccuracy is the strong numerical diffusion induced by most limiters in the neighbourhood of steep fronts that leads to a very strong damping of the oscillations. Another source of inaccuracy is the over-compression (that is, the steepening of shocks and contact discontinuities) induced by some limiters. The artificial sharpening of the profile caused by over-compression leads to amplifying the shorter modes in the oscillations, sometimes causing instability when dispersion terms are present in the PDEs to be solved.

A well-known approach for developing high resolution schemes with reduced stencils is the so-called compact approach introduced in the nineteen-eighties. In a first class of compact schemes (see References [6–8]), the higher-order space derivatives in the truncation error of the scheme are estimated by deriving the conservation law to be solved with respect to each direction of space successively. This allows to express third-order derivatives in a given direction as a combination of first- and second-order cross-derivatives with respect to time and the other directions of space. Since such derivatives can be estimated using a centred, three-point stencil, removing them from the truncation error of the scheme does not yield any increase in the stencil of the scheme. In a second class of compact schemes, higher-order accuracy is obtained by expressing the Taylor series expansions between the computational points as combinations of the first-order derivatives at several neighbouring points. The coefficients for such combinations can be found using e.g. Pade approximants (see Reference [9] for recent examples and references). This yields systems of algebraic equations, to be solved for the (unknown) derivatives.

The type of scheme proposed in the present paper does not belong to these classes of schemes. Higher-order accuracy is achieved by using two unknown variables, namely the average of the sought variables over the computational cells and the point values at the cell interfaces. This approach stems from the observation that (i) the estimate of the values at the interface of the computational cell is of key importance for the accuracy of Godunov-type schemes, and (ii) very accurate results can be achieved using schemes with a reduced stencil, but that make better use of the (already available) information than the point (or cell-averaged) values of the variable. This principle is used by two schemes derived from completely different approaches: The Holly–Preissmann scheme [10] and Chang's scheme [11]. The Holly–Preissmann scheme is a finite difference, characteristic-based scheme designed at the end of the nineteen-seventies. It solves the advection equation, not only for the variable, but also for its first space derivative. Two points are sufficient to derive a third-order interpolation,

 

using both the variable and its derivative at each point. The scheme has been applied to a number of scalar, linear and non-linear advection problems [12–14]. Chang's scheme is based on the space–time conservation formulation. This approach consists in integrating the PDEs to be solved over a domain that extends not only in space but also in time. The solution is sought over a mesh that is staggered in both space and time for both the flow variable and its derivative. A three-points stencil is sufficient to obtain very sharp profiles in the presence of both shocks and contact discontinuities. Oscillations are eliminated by the local application of artificial viscosity. The scheme has been applied to multi-dimensional gas dynamics [15], to the one-dimensional shallow water equations [16] and to one-dimensional advection–diffusion problems [17].

These two examples support the argument that making better use of the information related to the variable within limited space brings more benefit than complex interpolation procedures using large stencils. The work presented here follows this line. It consists of solving the flow equations not only for the average value of the variable over a computational cell, but also for its point values at the cell interfaces. This provides a much more accurate basis for the reconstruction of the profile. This approach was explored by Van Leer in the early developments of the MUSCL scheme [1] but seems to have been abandoned. One of the reasons may be that it was not clear at this time how the method should be generalized to hyperbolic systems of conservation laws. Another reason may be that the publication did not make clear how the PDEs describing other processes (such as the diffusion equation, source terms, etc.) could be solved within the proposed framework.

The present paper aims to generalize the approach to one- and multi-dimensional hyperbolic systems of conservation laws and to explain how it can be used to solve other types of PDEs. Section 2 presents the principle of higher-order Godunov-type schemes. Section 3 the principle of the approach for scalar conservation laws and hyperbolic systems of conservations laws. Section 4 explains how it can be extended to the solution of partial differential equations other than hyperbolic systems of conservation laws. Section 5 provides computational examples and Section 6 is devoted to concluding remarks.

## 2. HIGHER-ORDER GODUNOV-TYPE SCHEMES

This section presents the principle of higher-order Godunov-type schemes for the solution of one-dimensional hyperbolic systems of conservation laws. Such systems can be written in conservation form as

$$\frac{\partial \mathbf{U}}{\partial t} + \frac{\partial \mathbf{F}}{\partial x} = 0 \tag{1}$$

where $\mathbf{U}$ is the vector that contains the conserved variable, $\mathbf{F}$ is the flux vector, $x$ is the space co-ordinate and $t$ is time. The flux $\mathbf{F}$ is a function of $\mathbf{U}$. Note that Equation (1) can be rewritten in non-conservation, or characteristic form, as

$$\frac{\partial \mathbf{U}}{\partial t} + \mathbf{A}\frac{\partial \mathbf{U}}{\partial x} = 0 \tag{2}$$

where $\mathbf{A} = \partial \mathbf{F}/\partial \mathbf{U}$ is the Jacobian matrix of $\mathbf{F}$ with respect to $\mathbf{U}$. The characteristic form is used in the development of higher-order Godunov methods, as shown below. Section 2.1

 

presents the general principle of higher-order Godunov-type schemes. Section 2.2 details the reconstruction process as it is used in classical schemes. Section 2.3 outlines the methodology used to convert the General Riemann Problems (GRPs) at the cell interfaces to equivalent Riemann problems (ERPs) that are solved to compute the intercell fluxes.

### 2.1. The six steps of Godunov-type algorithms

Higher-order Godunov-type algorithms can be decomposed into the following six steps.

(1) Space is discretized into volumes, also called cells, over which the average value of $\mathbf{U}$ is assumed to be known at a given time $t^n$. The average value of $\mathbf{U}$ over the cell $i$ at time level $n$ is denoted by $\mathbf{U}_i^n$.

(2) The second step, called reconstruction, consists of 'guessing' the variations of $\mathbf{U}$ within each cell from the average values in the neighbouring cells. The reconstructed profile at the abscissa $x$ in the cell $i$ at time level $n$ is denoted by $\tilde{\mathbf{U}}_i^n(x)$. The reconstruction step is essential to the accuracy of the scheme. It is described in deeper detail in Section 2.2.

(3) The reconstructed profiles are used to construct Generalized Riemann Problems (GRPs) at the interfaces between the computational cells. Such GRPs are defined as

$$\mathbf{U}_{i+1/2}(x,t^n)=\begin{cases}\tilde{\mathbf{U}}_i^n(x) & \text{for } x<x_{i+1/2}\\ \tilde{\mathbf{U}}_{i+1}^n(x) & \text{for } x>x_{i+1/2}\end{cases} \tag{3}$$

where $x_{i+1/2}$ denotes the abscissa of the interface $i+\frac{1}{2}$ between the cells $i$ and $i+1$.

(4) The GRPs at the cell interfaces are converted to ERPs

$$\mathbf{U}_{i+1/2}(x,t^n)=\begin{cases}\mathbf{U}_{i+1/2,\mathrm{L}}^{n+1/2} & \text{for } x<x_{i+1/2}\\ \mathbf{U}_{i+1/2,\mathrm{R}}^{n+1/2} & \text{for } x>x_{i+1/2}\end{cases} \tag{4}$$

where the (constant) left and right states $\mathbf{U}_{i+1/2,\mathrm{L}}$ and $\mathbf{U}_{i+1/2,\mathrm{R}}$ are such that the solution of the ERP (4) yields the same average flux as the solution of the GRP (3) between the time levels $n$ and $n+1$. The next subsection explains how the ERPs can be determined from the GRPs.

(5) The solution of the Riemann problem of Equation (4) depends only on the ratio $(x-x_{i+1/2})/(t-t^n)$ [18, 19]. Therefore the solution $\mathbf{U}_{i+1/2}^{n+1/2}$ of the ERP (4) is constant at the location of the initial discontinuity, that is, at the interface $i+\frac{1}{2}$. The flux $\mathbf{F}_{i+1/2}^{n+1/2}$ at the interface $i+\frac{1}{2}$ between the time levels $n$ and $n+1$ is computed as

$$\mathbf{F}_{i+1/2}^{n+1/2}=\mathbf{F}(\mathbf{U}_{i+1/2}^{n+1/2}) \tag{5}$$

(6) The fluxes at the interfaces between the cells are used to compute the solution at the next time level $n+1$ as

$$\mathbf{U}_i^{n+1}=\mathbf{U}_i^n+\frac{\Delta t}{\Delta x_i}(\mathbf{F}_{i-1/2}^{n+1/2}-\mathbf{F}_{i+1/2}^{n+1/2}) \tag{6}$$

where $\Delta t$ is the computational time step and $\Delta x_i$ is the width of the cell $i$.

 

### 2.2. The reconstruction step

The reconstruction step is a key factor to scheme accuracy. The nature of the reconstruction function is defined *a priori*. The basic constraint on the reconstructed profile is that it should satisfy mass conservation:

$$\frac{1}{\Delta x_i}\int_{x_{i-1/2}}^{x_{i+1/2}} \tilde{\mathbf{U}}_i^n(x)\,\mathrm{d}x = \mathbf{U}_i^n \tag{7}$$

However, this condition is not sufficient to guarantee the uniqueness of the reconstruction function. For instance the MUSCL/PLM schemes [1, 20] use a linear reconstruction

$$\tilde{\mathbf{U}}_i^n(x) = (x - x_i)\mathbf{a}_i^n + \mathbf{b}_i^n \tag{8}$$

where $x_i$ is the abscissa of the centre of the cell $i$. Equation (7) yields the following condition on $\mathbf{b}_i^n$

$$\mathbf{b}_i^n = \mathbf{U}_i^n \tag{9}$$

and the vector slope $\mathbf{a}_i^n$ is still to be determined. In the classical MUSCL scheme $\mathbf{a}_i^n$ is taken as a combination of the average values of $\mathbf{U}$ in the neighbouring cells

$$\mathbf{a}_i^n = 2\,\frac{\mathbf{U}_{i+1}^n - \mathbf{U}_{i-1}^n}{\Delta x_{i-1} + 2\Delta x_i + \Delta x_{i+1}} \tag{10}$$

The PPM [3] uses a parabolic reconstruction

$$\tilde{\mathbf{U}}_i^n(x) = (x - x_i)^2\mathbf{a}_i^n + (x - x_i)\mathbf{b}_i^n + \mathbf{c}_i^n \tag{11}$$

Substituting Equation (8) into Equation (7) leads to the following condition

$$\frac{\Delta x_i^2}{12}\,\mathbf{a}_i^n + \mathbf{c}_i^n = \mathbf{U}_i^n \tag{12}$$

Two additional conditions must be supplied to guarantee the uniqueness of the reconstruction. These conditions are supplied by prescribing the value taken by $\tilde{\mathbf{U}}_i^n$ at the left- and right-hand edges of the cell

$$\begin{aligned} \tilde{\mathbf{U}}_i^n(x_{i-1/2}) &= \mathbf{U}_{i,\mathrm{L}}^n \\ \tilde{\mathbf{U}}_i^n(x_{i+1/2}) &= \mathbf{U}_{i,\mathrm{R}}^n \end{aligned} \tag{13}$$

where $\mathbf{U}_{i,\mathrm{L}}^n$ and $\mathbf{U}_{i,\mathrm{R}}^n$ are given by (for a uniform cell width):

$$\begin{aligned} \mathbf{U}_{i,\mathrm{L}}^n &= \tfrac{1}{12}\,(-\mathbf{U}_{i-2}^n + 7\mathbf{U}_{i-1}^n + 7\mathbf{U}_i^n - \mathbf{U}_{i+1}^n) \\ \mathbf{U}_{i,\mathrm{R}}^n &= \tfrac{1}{12}\,(-\mathbf{U}_{i-1}^n + 7\mathbf{U}_i^n + 7\mathbf{U}_{i+1}^n - \mathbf{U}_{i+2}^n) \end{aligned} \tag{14}$$

These formulae are obtained by fitting a cubic polynomial in $x$ in such a way that it verifies the conservation condition (7) over the two cells on the left- and right-hand side of a given interface [3]. The fitting of the polynomial and the reconstruction within the cell are illustrated by Figure 1.

 

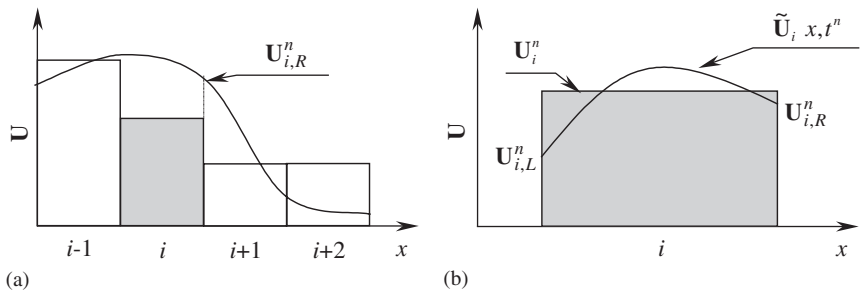


Figure 1. Principle of the PPM reconstruction. Fitting of a cubic polynomial for the computation of the edge values (a) and parabolic interpolation within the cell (b).

Note that in these examples, the reconstructed profile may be further modified in such a way that the monotony of the solution is ensured. Monotony is achieved by adjusting the reconstructed profile in the cell $i$ in such a way that the following conditions are verified: (i) $\tilde{\mathbf{U}}_i^n(x)$ lies between $\mathbf{U}_{i-1}^n$ and $\mathbf{U}_{i+1}^n$, (ii) $\tilde{\mathbf{U}}_i^n(x)$ is monotone within the cell $i$, and (iii) $\mathrm{d}\tilde{\mathbf{U}}_i^n(x)/\mathrm{d}x$ has the same sign as $\mathbf{U}_{i+1}^n - \mathbf{U}_{i-1}^n$. Note that more elaborate procedures have been introduced for one- and multi-dimensional MUSCL-type algorithms [4, 5]. The common point between all higher-order Godunov-types schemes is that the conditions of the type (10, 13, 14) used to characterize the reconstruction within a given cell make use of the average values of the variable $\mathbf{U}$ over the neighbouring cells. The proposed method is based on a different approach, as shown in Section 3.

### 2.3. Determination of the ERP

The method for the determination of the ERP is explained in detail in other publications (see e.g. References [3, 21] or [22]). However, its principle is recalled herein because it forms the basis for the reconstruction method proposed in Section 3. Consider two neighbouring cells $i$ and $i+1$ where $\mathbf{U}$ has been reconstructed. The two profiles $\tilde{\mathbf{U}}_i^n(x)$ and $\tilde{\mathbf{U}}_{i+1}^n(x)$ form the GRP given by Equation (3) at the interface $i+\frac{1}{2}$. The purpose is to transform this GRP into an ERP of the form (4). To do so, the characteristic form (2) of the conservation law is used. In a first step, $\mathbf{U}$ is rewritten as a combination of the eigenvectors of the characteristic matrix $\mathbf{A}$:

$$\mathbf{U}(x,t) = \sum_p \alpha^{(p)}(x,t)\mathbf{K}^{(p)} = \mathbf{K}\boldsymbol{\alpha}(x,t) \tag{15}$$

where $\mathbf{K}^{(p)}$ is the $p$th eigenvector of $\mathbf{A}$, $\mathbf{K}$ is the matrix formed by the eigenvectors $\mathbf{K}^{(p)}$, $\alpha^{(p)}$ is the $p$th wave strength and $\boldsymbol{\alpha}$ is the vector formed by the components $\alpha^{(p)}$.

It is now explained how to determine the left state of the ERP. The first relation (15) is substituted into Equation (2) to give:

$$\frac{\partial}{\partial t}\mathbf{K}\boldsymbol{\alpha} + \mathbf{A}\frac{\partial}{\partial x}\mathbf{K}\boldsymbol{\alpha} = 0 \tag{16}$$

 

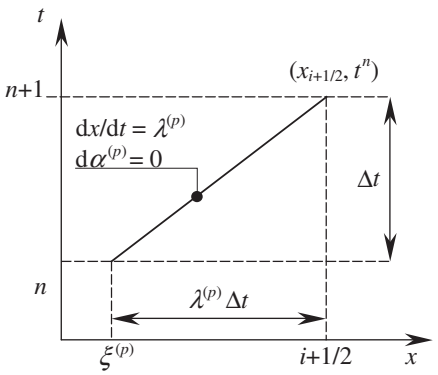


Figure 2. Invariance of the wave strengths along the characteristic lines in the phase space.

Left-multiplying Equation (16) by the matrix $\mathbf{K}^{-1}$ leads to:

$$\mathbf{K}^{-1}\frac{\partial}{\partial t}\mathbf{K}\boldsymbol{\alpha} + \mathbf{K}^{-1}\mathbf{A}\frac{\partial}{\partial x}\mathbf{K}\boldsymbol{\alpha} = 0 \tag{17}$$

$\mathbf{K}$ is approximated with the constant matrix $\mathbf{K}_i^n$ that can be taken out of the derivation operators to give:

$$\mathbf{K}_i^{n-1}\mathbf{K}_i^n\frac{\partial}{\partial t}\boldsymbol{\alpha} + \mathbf{K}_i^{n-1}\mathbf{A}_i^n\mathbf{K}_i^n\frac{\partial}{\partial x}\boldsymbol{\alpha} = 0 \tag{18}$$

Note that the matrix product $\mathbf{K}^{-1}\mathbf{A}\mathbf{K}$ is equal to the diagonal matrix $\boldsymbol{\Lambda}$ formed by the eigenvalues of $\mathbf{A}$, Equation (18) becomes

$$\frac{\partial\boldsymbol{\alpha}}{\partial t} + \boldsymbol{\Lambda}\frac{\partial\boldsymbol{\alpha}}{\partial x} = 0 \tag{19}$$

Equation (19) is equivalent to the following set of relationships

$$\frac{\mathrm{d}\alpha^{(p)}}{\mathrm{d}t} = 0 \quad \text{along} \quad \frac{\mathrm{d}x}{\mathrm{d}t} = \lambda^{(p)} \tag{20}$$

where $\lambda^{(p)}$ is the $p$th eigenvalue of $\mathbf{A}$. The invariance of the wave strengths along the characteristic lines is used to determine the value of $\mathbf{U}$ on the left-hand side of the interface $i+\frac{1}{2}$ (Figure 2). From Equation (15), one can write

$$\mathbf{U}_{i+1/2,\mathrm{L}}(t) = \sum_p \alpha_{i+1/2,\mathrm{L}}^{(p)}(t)\mathbf{K}^{(p)} \tag{21}$$

The summation in Equation (21) is split into two parts according to the sign of the eigenvalues

$$\mathbf{U}_{i+1/2,\mathrm{L}}(t) = \sum_{p\geqslant 0} \alpha_{i+1/2,\mathrm{L}}^{(p)}(t)\mathbf{K}^{(p)} + \sum_{p<0} \alpha_{i+1/2,\mathrm{L}}^{(p)}(t)\mathbf{K}^{(p)} \tag{22}$$

The value of $\alpha^{(p)}$ with positive eigenvalues at the interface at a time $t$ is deduced from the reconstructed value at time level $n$. The invariance property as stated by Equation (20) is

 

used to trace the $p$th wave strength backward in time in the cell $i$:

$$\alpha^{(p)}(x_{i+1/2},t)=\alpha^{(p)}(\xi^{(p)},t^n)=\tilde{\alpha}^{(p)}(\xi^{(p)}) \quad \text{for } \lambda^{(p)}\geqslant 0 \tag{23}$$

where $x^{(p)}=x_{i+1/2}-(t-t^n)\lambda^{(p)}$ is the abscissa of the foot of the $p$th characteristic line issued form the interface $i+\frac{1}{2}$ at time $t$ (see Figure 2). Since the wave strengths with negative eigenvalues travel in the direction opposite to the interface $i+\frac{1}{2}$ they do not influence the solution. Therefore, they can be taken equal to any arbitrary value. The simplest possible solution is to use the cell average

$$\alpha^{(p)}(x_{i+1/2},t)=\alpha^{(p)}_i \quad \text{for } \lambda^{(p)}<0 \tag{24}$$

The average value of $\mathbf{U}$ on the left-hand side of the interface is equal to

$$\begin{aligned}\mathbf{U}^{n+1/2}_{i+1/2,\mathrm{L}} &= \frac{1}{\Delta t}\int_{t^n}^{t^{n+1}}\mathbf{U}_{i+1/2,\mathrm{L}}(t)\,\mathrm{d}t=\frac{1}{\Delta t}\int_{t^n}^{t^{n+1}}\alpha^{(p)}_{i+1/2,\mathrm{L}}\mathbf{K}^{(p)^n_i}(t)\,\mathrm{d}t\\ &=\sum_p \alpha^{(p)^{n+1/2}_{i+1/2,\mathrm{L}}}\mathbf{K}^{(p)^n_i}\end{aligned} \tag{25}$$

where $\alpha^{(p)^{n+1/2}_{i+1/2,\mathrm{L}}}$ denotes the average value of the $p$th wave strength on the left-hand side of the interface between time levels $n$ and $n+1$. From Equations (23) and (24), $\alpha^{(p)^{n+1/2}_{i+1/2,\mathrm{L}}}$ is given by

$$\alpha^{(p)^{n+1/2}_{i+1/2,\mathrm{L}}}=\begin{cases}\dfrac{1}{\lambda^{(p)^n_i}\Delta t}\displaystyle\int_{x_{i+1/2}-\lambda^{(p)^n_i}\Delta t}^{x_{i+1/2}}\tilde{\alpha}^{(p)^n_i}(x)\,\mathrm{d}x & \text{if } \lambda^{(p)^n_i}\geqslant 0\\ \alpha^{(p)^n_i} & \text{if } \lambda^{(p)^n_i}<0\end{cases} \tag{26}$$

Reasoning by symmetry, the right state of the ERP on the right-hand side of the interface is given as

$$\mathbf{U}^{n+1/2}_{i+1/2,\mathrm{R}}=\sum_p \alpha^{(p)^{n+1/2}_{i+1/2,\mathrm{R}}}\mathbf{K}^{(p)^n_{i+1}} \tag{27}$$

where

$$\alpha^{(p)^{n+1/2}_{i+1/2,\mathrm{R}}}=\begin{cases}\dfrac{1}{\lambda^{(p)^n_{i+1}}\Delta t}\displaystyle\int_{x_{i+1/2}-\lambda^{(p)^n_{i+1}}\Delta t}^{x_{i+1/2}}\tilde{\alpha}^{(p)^n_{i+1}}(x)\,\mathrm{d}x & \text{if } \lambda^{(p)^n_{i+1}}\geqslant 0\\ \alpha^{(p)^n_{i+1}} & \text{if } \lambda^{(p)^n_{i+1}}<0\end{cases} \tag{28}$$

The determination of the left and right states of the ERP from the GRP can be summarized as follows: the wave strengths that travel towards the interface are averaged over the domain of dependence of the interface. The wave strengths that travel in the direction opposite to the interface can be set equal to any arbitrary value, say the wave strength computed from the cell average on the corresponding side of the interface.

 

## 3. PROPOSED METHOD

The proposed method can be seen as a generalization to hyperbolic systems of conservation laws of a reconstruction method originally proposed by Van Leer [1] for the solution of scalar conservation laws. In the development of the MUSCL scheme Van Leer experimented various methods for profile reconstruction. Quite ironically, the method that became popular as embedded in the MUSCL scheme is the least accurate of the three approaches tested by Van Leer. Section 3.1 presents the method as proposed by Van Leer [1] for the solution of the scalar, linear advection equation. Section 3.2 generalizes this approach to hyperbolic systems of conservation laws. Section 3.3 gives an outline of the method in algorithmic form.

### *3.1. Scalar conservation laws*

The present subsection focuses on the principle of the method for the solution of scalar laws. Consider first the scalar linear advection equation

$$\frac{\partial U}{\partial t} + \frac{\partial}{\partial x}(\lambda U) = 0 \tag{29}$$

the characteristic form of which is (under the assumption of a constant $\lambda$)

$$\frac{\partial U}{\partial t} + \lambda \frac{\partial U}{\partial x} = 0 \tag{30}$$

As mentioned in Section 2.2, the degrees of freedom in higher-order reconstructions are eliminated by forcing the reconstructed profiles to take specific values at the edges of the computational cells. In classical higher-order Godunov-type schemes these edge values are computed using the average values of the variable in the cells neighbouring the cell of interest, see e.g. Equations (10), (13) and (14). However, Van Leer had experimented in Reference [1] an approach where the values of $U$ at the edges of the cell $i$ were computed as

$$\begin{aligned} U^n_{i,\mathrm{L}} &= U(x_{i+1/2} - \lambda\Delta t, t^{n-1}) = \begin{cases} \tilde{U}^{n-1}_{i-1}(x_{i-1/2} - \lambda\Delta t) & \text{if } \lambda \geqslant 0 \\ \tilde{U}^{n-1}_{i}(x_{i-1/2} - \lambda\Delta t) & \text{if } \lambda \leqslant 0 \end{cases} \\ U^n_{i,\mathrm{R}} &= U(x_{i+1/2} - \lambda\Delta t, t^{n-1}) = \begin{cases} \tilde{U}^{n-1}_{i}(x_{i+1/2} - \lambda\Delta t) & \text{if } \lambda \geqslant 0 \\ \tilde{U}^{n-1}_{i+1}(x_{i+1/2} - \lambda\Delta t) & \text{if } \lambda \leqslant 0 \end{cases} \end{aligned} \tag{31}$$

The formulae of Equation (31) are justified by the invariance of $U$ along the characteristic lines as expressed by Equation (30). Travelling backwards in time along the characteristic $\mathrm{d}x/\mathrm{d}t = \lambda$ allows the value at the cell interfaces at the time level $n$ to be related to the reconstructed profile at the time level $n-1$ (Figure 2).

The approach is now extended to non-linear scalar laws, that is, to equations of the form

$$\frac{\partial U}{\partial t} + \frac{\partial F}{\partial x} = 0 \tag{32}$$

 

where the flux $F$ is a non-linear function of $U$. The characteristic form of Equation (32) is Equation (30) where $\lambda$ is given by

$$\lambda = \frac{\mathrm{d}F}{\mathrm{d}U} \tag{33}$$

The non-linearity of the flux function $F$ induces a dependence of $\lambda$ on the conserved variable $U$. The variability of $\lambda$ in space and time may trigger the appearance of shocks or rarefaction waves in the solution. Consequently the profile is not necessarily continuous at the interface $i+\frac{1}{2}$ at the beginning of the time step between the time levels $n$ and $n+1$. Assume that the values $U^n_{i+1/2,\mathrm{L}}$ and $U^n_{i+1/2,\mathrm{R}}$ on the left- and right-hand sides of the interface $i+\frac{1}{2}$ can be determined at $t=t^n$ (how they are to be calculated is described in the next paragraph). These two values are used to construct a Riemann problem, the solution $U^n_{i+1/2}$ of which is the value of $U$ at the interface for a time $t$ infinitely close to $t^n$. In the present approach, $U^n_{i+1/2}$ is used as the right-hand edge value of the reconstructed profile in the cell $i$ and as the left-hand edge value of the reconstructed profile in the cell $i+1$,

$$\begin{aligned} \tilde{U}^n_i(x_{i+1/2}) &= U^n_{i,\mathrm{R}} = U^n_{i+1/2} \\ \tilde{U}^n_{i+1}(x_{i+1/2}) &= U^n_{i,\mathrm{L}} = U^n_{i+1/2} \end{aligned} \tag{34}$$

It is now explained how the left and right states $U^n_{i+1/2,\mathrm{L}}$ and $U^n_{i+1/2,\mathrm{R}}$ of the Riemann problem should be determined. Extending Van Leer's idea to non-linear scalar laws, it is chosen to compute $U^n_{i+1/2,\mathrm{L}}$ and $U^n_{i+1/2,\mathrm{R}}$ using the reconstructed profiles $\tilde{U}^{n-1}_i$ at the previous time level rather than the cell averages $U^n_i$. This is done using the property of invariance of $U$ along the characteristic lines. Applying Equation (30) along the characteristic line issued from the left-hand side of the interface $i+\frac{1}{2}$ at $t=t^n$ gives:

$$\begin{aligned} U^n_{i+1/2,\mathrm{L}} &= U(x_{i+1/2} - \lambda^{n-1}_i \Delta t, t^{n-1}) \\ &= \tilde{U}^{n-1}_i(x_{i+1/2} - \lambda^{n-1}_i \Delta t) \end{aligned} \tag{35}$$

As mentioned in Section 2.3, Equation (35) is valid only if the characteristic line $\mathrm{d}x/\mathrm{d}t = \lambda^{n-1}_i$ travels towards the interface $i+\frac{1}{2}$ (that is, if $\lambda^{n-1}_i$ is positive). If $\lambda^{n-1}_i$ is negative, the characteristic travels in the opposite direction to the interface and the profile in the cell $i$ does not influence the value of $U$ at the interface. Therefore, $U^n_{i+1/2,\mathrm{L}}$ can be taken equal to any arbitrary value. As in Section 2.3, the simplest possible is to use the average value $U^n_i$ over the cell. Applying a similar reasoning to the right-hand side of the interface leads to the following relationships

$$\begin{aligned} U^n_{i+1/2,\mathrm{L}} &= \begin{cases} \tilde{U}^{n-1}_i(x_{i+1/2} - \lambda^{n-1}_i \Delta t) & \text{if } \lambda^{n-1}_i \geqslant 0 \\ U^{n-1}_i & \text{if } \lambda^{n-1}_i < 0 \end{cases} \\ U^n_{i+1/2,\mathrm{R}} &= \begin{cases} \tilde{U}^{n-1}_{i+1}(x_{i+1/2} - \lambda^{n-1}_{i+1} \Delta t) & \text{if } \lambda^{n-1}_{i+1} \leqslant 0 \\ U^{n-1}_{i+1} & \text{if } \lambda^{n-1}_{i+1} > 0 \end{cases} \end{aligned} \tag{36}$$

 

### 3.2. Generalization to hyperbolic systems of conservation laws

The principle of the approach remains the same as in the previous subsection: the edge values $\mathbf{U}^n_{i,\mathrm{L}}$ and $\mathbf{U}^n_{i,\mathrm{R}}$ used for the reconstruction of $\mathbf{U}$ in the cell $i$ are taken as

$$\begin{aligned}\mathbf{U}^n_{i,\mathrm{L}} &= \mathbf{U}^n_{i-1/2}\\ \mathbf{U}^n_{i,\mathrm{R}} &= \mathbf{U}^n_{i+1/2}\end{aligned} \tag{37}$$

where $\mathbf{U}^n_{i+1/2}$ is the solution of the Riemann problem $(\mathbf{U}^n_{i+1/2,\mathrm{L}}, \mathbf{U}^n_{i+1/2,\mathrm{L}})$ at the interface $i+\frac{1}{2}$. As in the previous subsections, the left and right states of the Riemann problems $(\mathbf{U}^n_{i+1/2,\mathrm{L}}, \mathbf{U}^n_{i+1/2,\mathrm{L}})$ are determined using the reconstructed profiles at the previous time level $n-1$ rather than the cell averages at the current time level $n$. The method used for the determination of the left state $\mathbf{U}^n_{i+1/2,\mathrm{L}}$ is detailed hereafter.

The left state $\mathbf{U}^n_{i+1/2,\mathrm{L}}$ is determined by applying the decomposition (15) on the left-hand side of the interface $i+\frac{1}{2}$

$$\mathbf{U}^n_{i+1/2,\mathrm{L}} = \mathbf{K}_{i+1/2,\mathrm{L}}\boldsymbol{\alpha}_{i+1/2,\mathrm{L}} \tag{38}$$

As in the linear case, the proposed method consists in approximating the matrix $\mathbf{K}$ and the wave strengths $\boldsymbol{\alpha}$ using the reconstructed profile at the previous time level $n-1$. As in Section 2.3, the matrix of eigenvectors is taken constant over the cell $i$

$$\mathbf{K}_{i+1/2,\mathrm{L}} = \mathbf{K}^{n-1}_i \tag{39}$$

and the wave strengths are determined using Equation (20) that expresses the invariance of the $p$th wave strength along the $p$th characteristic line

$$\alpha^{(p)}_{i+1/2,\mathrm{L}} = \tilde{\alpha}^{(p)^{n-1}_i}(x_{i+1/2} - \lambda^{(p)^{n-1}_i}\Delta t) \tag{40}$$

Note again that Equation (40) is applied only if the $p$th characteristic travels towards the interface, that is, if $\lambda^{(p)^{n-1}_i}$ is positive. If $\lambda^{(p)^{n-1}_i}$ is negative the $p$th wave strength does not influence the solution and can be taken equal to $\alpha^{(p)^{n-1}_i}$. Reasoning by symmetry for the right state of the Riemann problems leads to

$$\begin{aligned}\alpha^{(p)}_{i+1/2,\mathrm{L}} &= \begin{cases}\tilde{\alpha}^{(p)^{n-1}_i}(x_{i+1/2} - \lambda^{(p)^{n-1}_i}\Delta t) & \text{if } \lambda^{(p)^{n-1}_i} \geqslant 0\\ \alpha^{(p)^{n-1}_i} & \text{if } \lambda^{(p)^{n-1}_i} < 0\end{cases}\\ \alpha^{(p)}_{i+1/2,\mathrm{R}} &= \begin{cases}\tilde{\alpha}^{(p)^{n-1}_{i+1}}(x_{i+1/2} - \lambda^{(p)^{n-1}_{i+1}}\Delta t) & \text{if } \lambda^{(p)^{n-1}_{i+1}} \leqslant 0\\ \alpha^{(p)^{n-1}_{i+1}} & \text{if } \lambda^{(p)^{n-1}_{i+1}} > 0\end{cases}\\ \mathbf{K}_{i+1/2,\mathrm{L}} &= \mathbf{K}^{n-1}_i\\ \mathbf{K}_{i+1/2,\mathrm{R}} &= \mathbf{K}^{n-1}_{i+1}\end{aligned} \tag{41}$$

 

### 3.3. Outline of the algorithm

The algorithm is provided for hyperbolic systems of conservation laws. It is assumed that the variable $\mathbf{U}$ has been reconstructed in all the cells.

(1) For each interface $i+\frac{1}{2}$, compute the left and right states $\mathbf{U}^{n+1/2}_{i+1/2,\mathrm{L}}$ and $\mathbf{U}^{n+1/2}_{i+1/2,\mathrm{R}}$ of the ERP using Equations (26) and (28).
(2) Use the solution $\mathbf{U}^{n+1/2}_{i+1/2}$ of the ERP at each interface $i+\frac{1}{2}$ to compute the flux $\mathbf{F}^{n+1/2}_{i+1/2}$ between the cells $i$ and $i+1$.
(3) Compute the cell average at the next time level using the balance equation (6).
(4) For each interface $i+\frac{1}{2}$, compute the wave strengths and eigenvectors for the left and right states of the Riemann problem at the time level $n+1$ by rewriting Equation (41) between $t^n$ and $t^{n+1}$:

$$\begin{aligned}
\alpha^{(p)}_{i+1/2,\mathrm{L}} &= \begin{cases} \tilde{\alpha}^{(p)^n_i}(x_{i+1/2} - \lambda^{(p)^n_i}\Delta t) & \text{if } \lambda^{(p)^n_i} \geqslant 0 \\ \alpha^{(p)^n_i} & \text{if } \lambda^{(p)^n_i} < 0 \end{cases} \\
\alpha^{(p)}_{i+1/2,\mathrm{R}} &= \begin{cases} \tilde{\alpha}^{(p)^n_{i+1}}(x_{i+1/2} - \lambda^{(p)^n_{i+1}}\Delta t) & \text{if } \lambda^{(p)^n_{i+1}} \leqslant 0 \\ \alpha^{(p)^n_{i+1}} & \text{if } \lambda^{(p)^n_i} > 0 \end{cases} \\
\mathbf{K}_{i+1/2,\mathrm{L}} &= \mathbf{K}^n_i \\
\mathbf{K}_{i+1/2,\mathrm{R}} &= \mathbf{K}^n_{i+1}
\end{aligned} \tag{42}$$

and by rewriting $\mathbf{U}$ in the form (15) as a combination of the wave strengths

$$\begin{aligned}
\mathbf{U}^{n+1}_{i+1/2,\mathrm{L}} &= \mathbf{K}_{i+1/2,\mathrm{L}}\boldsymbol{\alpha}_{i+1/2,\mathrm{L}} \\
\mathbf{U}^{n+1}_{i+1/2,\mathrm{R}} &= \mathbf{K}_{i+1/2,\mathrm{R}}\boldsymbol{\alpha}_{i+1/2,\mathrm{R}}
\end{aligned} \tag{43}$$

(5) Solve the Riemann problem $(\mathbf{U}^{n+1}_{i+1/2,\mathrm{L}}, \mathbf{U}^{n+1}_{i+1/2,\mathrm{R}})$ at each interface $i+\frac{1}{2}$. This yields the solution $\mathbf{U}^{n+1}_{i+1/2}$.
(6) Reconstruct the profile in each cell $i$ at the time level $n+1$ using the cell average $\mathbf{U}^{n+1}_i$ obtained from the steps (1)–(3) and the values $\mathbf{U}^{n+1}_{i-1/2}$ computed in steps (4)–(5) to eliminate the remaining degrees of freedom in the reconstruction by enforcing the following conditions

$$\begin{aligned}
\tilde{\mathbf{U}}^{n+1}_i(x_{i-1/2}) &= \mathbf{U}^{n+1}_{i-1/2} \\
\tilde{\mathbf{U}}^{n+1}_i(x_{i+1/2}) &= \mathbf{U}^{n+1}_{i+1/2}
\end{aligned} \tag{44}$$

The reconstructions may be further modified via profile limiting in order to enforce the monotony of the solution.

 

## 4. EXTENSION TO OTHER TYPES OF PARTIAL DIFFERENTIAL EQUATIONS

This section shows how the proposed approach can be extended to PDEs involving source term and second- or higher-order space derivatives. The principle of the extension is explained using the linear dispersion equation as an example. Both the conservative and non-conservative version of the equation are considered. The present section is complemented with Appendix B that provides a stability analysis for the explicit versions of the discretizations introduced hereafter.

### *4.1. Conservative PDEs involving higher-order derivatives*

The extension to conservative PDEs involving higher-order derivatives is explained on the example of the dispersion equation in conservation form

$$\begin{aligned} \frac{\partial U}{\partial t} + \frac{\partial F}{\partial x} &= 0 \\ F &= k \frac{\partial^2 U}{\partial x^2} \end{aligned} \tag{45}$$

for negative $k$. Assuming a uniform cell size $\Delta x$, Equation (45) can be discretized for both the cell averages and the cell interfaces as

$$\begin{aligned} U_i^{n+1} &= U_i^n + \frac{\Delta t}{\Delta x}(F_{i-1/2}^{n+1/2} - F_{i+1/2}^{n+1/2}) \\ U_{i+1/2}^{n+1} &= U_{i+1/2}^n + 2\frac{\Delta t}{\Delta x}(F_i^{n+1/2} - F_{i+1}^{n+1/2}) \end{aligned} \tag{46}$$

where $F_i^{n+1/2}$ is the flux evaluated at $x = x_i$. Note that $F_i^{n+1/2}$ being a point value, it is not necessarily equal to $F(U_i^{n+1/2})$. Note that in the case of a uniform cell width $\Delta x$, the classical finite volume and finite difference approaches yield the following semi-discretizations for the derivatives:

$$\begin{aligned} F_{i+1/2}^{n+1/2} &\approx k \frac{U_{i-1} - 2U_i + U_{i+1}}{\Delta x^2} \\ F_i^{n+1/2} &\approx k \frac{U_{i-3/2} - 2U_{i-1/2} + U_{i+1/2}}{\Delta x^2} \end{aligned} \tag{47}$$

Note that this semi-discretization is not the only possible one, but that off-centring the estimate of the derivatives in the direction of negative $x$ allows the solution to be conditionally stable for explicit discretizations. Substituting Equation (47) into Equation (46) leads to the following relationships

$$\begin{aligned} U_i^{n+1} &= U_i^n + k \frac{\Delta t}{\Delta x^3}(U_{i-2} - 3U_{i-1} + 3U_i - U_{i+1}) \\ U_{i+1/2}^{n+1} &= U_{i+1/2}^n + k \frac{\Delta t}{\Delta x^3}(U_{i-3/2} - 3U_{i-1/2} + 3U_{i+1/2} - U_{i+3/2}) \end{aligned} \tag{48}$$

In this case the cell averages and the point values at the cell interfaces are computed independently from each other. The stencil of the scheme involves four cells. The proposed approach

 

consists in using the following off-centred estimates for the second-order derivatives

$$\left.\frac{\partial^2 U}{\partial x^2}\right)_{i+1/2} \approx \frac{aU_{i-1/2} + bU_{i+1/2} + cU_{i+1}}{\Delta x^2} \tag{49a}$$

$$\left.\frac{\partial^2 U}{\partial x^2}\right)_{i} \approx \frac{dU_{i-1} + eU_{i} + fU_{i+1/2}}{\Delta x^2} \tag{49b}$$

The coefficients $a$ to $f$ are determined as follows. $U(x)$ is written in terms of Taylor series expansions starting from $x_i$ and $x_{i+1/2}$

$$U^{(1)}(x) = U_{i+1/2} + (x - x_{i+1/2}) \left.\frac{\partial U}{\partial x}\right)_{i+1/2} + \left.\frac{(x - x_{i+1/2})^2}{2} \frac{\partial^2 U}{\partial x^2}\right)_{i+1/2} + \theta(x - x_{i+1/2})^3 \tag{50a}$$

$$U^{(2)}(x) = U(x_i) + (x - x_i) \left.\frac{\partial U}{\partial x}\right)_{i} + \left.\frac{(x - x_i)^2}{2} \frac{\partial^2 U}{\partial x^2}\right)_{i} + \vartheta(x - x_i)^3 \tag{50b}$$

Equation (50a) allows $U_{i+1/2}$ and $U_{i+1}$ to be expressed as functions of $U_{i-1/2}$ and the first- and second-order derivatives of $U$ with respect to $x$

$$\begin{aligned} U_{i-1/2} &= U^{(1)}(x_{i-1/2}) = U_{i+1/2} - \Delta x \left.\frac{\partial U}{\partial x}\right)_{i+1/2} + \left.\frac{\Delta x^2}{2} \frac{\partial^2 U}{\partial x^2}\right)_{i+1/2} + \theta_1(\Delta x^3) \\ U_{i+1} &= \frac{1}{\Delta x} \int_{x_{i-1/2}}^{x_{i+1/2}} U^{(1)}(x)\,\mathrm{d}x = U_{i+1/2} + \left.\frac{\Delta x}{2} \frac{\partial U}{\partial x}\right)_{i+1/2} + \left.\frac{\Delta x^2}{6} \frac{\partial^2 U}{\partial x^2}\right)_{i+1/2} + \theta_2(\Delta x^3) \end{aligned} \tag{51}$$

Substituting Equation (51) into Equation (49a) yields

$$\left.\frac{\partial^2 U}{\partial x^2}\right)_{i+1/2} \approx \frac{a+b+c}{\Delta x^2} U_{i+1/2} + \left.\frac{c-2a}{2\Delta x} \frac{\partial U}{\partial x}\right)_{i+1/2} + \left(\frac{a}{2} + \frac{c}{6}\right) \left.\frac{\partial^2 U}{\partial x^2}\right)_{i+1/2} + \theta_4(\Delta x) \tag{52}$$

Comparing Equations (49a) and (52) yields the following conditions on $a$, $b$ and $c$

$$\begin{aligned} a + b + c &= 0 \\ c - 2a &= 0 \\ \frac{a}{2} + \frac{c}{6} &= 1 \end{aligned} \tag{53}$$

Solving the system (53) for $a$, $b$ and $c$ gives

$$\left.\frac{\partial^2 U}{\partial x}\right)_{i+1/2} \approx \frac{6}{5\Delta x^2} (U_{i-1/2} - 3U_{i+1/2} + 2U_{i+1}) \tag{54}$$

 

Equation (50b) provides the expressions for $U_{i-1}$, $U_i$ and $U_{i+1/2}$ to be used in Equation (49b):

$$\begin{aligned} U_{i-1} &= \int_{x_{i-3/2}}^{x_{i-1/2}} U^{(2)}(x)\,\mathrm{d}x = U(x_i) - \Delta x \left.\frac{\partial U}{\partial x}\right)_i + \frac{13}{24}\Delta x^2 \left.\frac{\partial^2 U}{\partial x^2}\right)_i + \vartheta_1(\Delta x^3) \\ U_i &= \int_{x_{i-1/2}}^{x_{i+1/2}} U^{(2)}(x)\,\mathrm{d}x = U(x_i) + \frac{1}{24}\Delta x^2 \left.\frac{\partial^2 U}{\partial x^2}\right)_i + \vartheta_2(\Delta x^3) \\ U_{i+1/2} &= U^{(2)}(x_{i+1/2}) = U(x_i) + \frac{\Delta x}{2}\left.\frac{\partial U}{\partial x}\right)_i + \frac{1}{8}\Delta x^2 \left.\frac{\partial^2 U}{\partial x^2}\right)_i + \vartheta_3(\Delta x^3) \end{aligned} \tag{55}$$

Substituting Equation (55) into Equation (49b) yields

$$\left.\frac{\partial^2 U}{\partial x^2}\right)_i = \frac{d+e+f}{\Delta x^2} U(x_i) + \frac{f-2d}{2\Delta x}\left.\frac{\partial U}{\partial x}\right)_i + \frac{13d+e+3f}{24}\left.\frac{\partial^2 U}{\partial x^2}\right)_i + \vartheta_4(\Delta x) \tag{56}$$

Comparing Equations (49b) and (56) leads to the following conditions on $d$, $e$ and $f$

$$\begin{aligned} d+e+f &= 0 \\ f-2d &= 0 \\ 13d+e+3f &= 24 \end{aligned} \tag{57}$$

Solving the system (57) leads to the following expression for the second-order derivative at $x_i$

$$\left.\frac{\partial^2 U}{\partial x^2}\right)_i \approx \frac{3}{2\Delta x^2}(U_{i-1} - 3U_i + 2U_{i+1/2}) \tag{58}$$

Substituting Equations (54) and (58) into Equation (46) gives

$$\begin{aligned} U_i^{n+1} &= U_i^n + \frac{6}{5}\frac{k\Delta t}{\Delta x^3}(U_{i-3/2} - 4U_{i-1/2} + 2U_i + 3U_{i+1/2} - 2U_{i+1}) \\ U_{i+1/2}^{n+1} &= U_{i+1/2}^n + \frac{3}{2}\frac{k\Delta t}{\Delta x^3}(U_{i-1} - 4U_i + 2U_{i+1/2} + 3U_{i+1} - 2U_{i+3/2}) \end{aligned} \tag{59}$$

In contrast with the classical discretization (48), only three cells are needed in the discretization (59). Note that Equation (59) should be modified when used in combination with Godunov-type schemes. The reason is that the limiters embedded in most reconstruction procedures lead to profiles that are discontinuous across the cell edges. The interface values $U_{i+1/2}$ on the right-hand side of Equation (59) should be replaced with the average of the values on

 

each side of the interface:

$$\begin{aligned} U_i^{n+1} &= U_i^n + \frac{6}{5}\frac{k\Delta t}{\Delta x^3}\left[\frac{U_{i-2,\mathrm{R}}+U_{i-1,\mathrm{L}}}{2} - 2(U_{i-1,\mathrm{R}}+U_{i,\mathrm{L}}) + 2U_i 3\,\frac{U_{i,\mathrm{R}}+U_{i+1,\mathrm{L}}}{2} - 2U_{i+1}\right] \\ U_{i,\mathrm{R}}^{n+1} &= U_{i+1,\mathrm{L}}^{n+1} = \frac{U_{i,\mathrm{R}}^n + U_{i+1,\mathrm{L}}^n}{2} \\ &\quad + \frac{3}{2}\frac{k\Delta t}{\Delta x^3}(U_{i-1} - 4U_i + U_{i,\mathrm{R}} + U_{i+1,\mathrm{L}} + 3U_{i+1} - U_{i+1,\mathrm{R}} - U_{i+2,\mathrm{L}}) \end{aligned} \tag{60}$$

### 4.2. Non-conservative PDEs involving higher-order derivatives

Consider now the non-conservative version of Equation (45)

$$\frac{\partial U}{\partial t} + k\frac{\partial^3 U}{\partial x^3} = 0 \tag{61}$$

where $k$ is assumed to be negative as in the previous subsection. Since Equation (60) is not in conservation form, it is not necessary to approximate the third-order derivative as the differential of a flux. This allows more compact approximations to be used. Retaining the off-centring approach used in the previous subsection and using third-order Taylor series expansions, the following approximations can be proposed

$$\begin{aligned} \left.\frac{\partial^3 U}{\partial x^3}\right)_i &\approx \frac{6}{\Delta x^3}(-U_{i-1} + 4U_{i-1/2} - 5U_i + 2U_{i+1/2}) \\ \left.\frac{\partial^3 U}{\partial x^3}\right)_{i+1/2} &\approx \frac{6}{\Delta x^3}(-2U_{i-1/2} + 5U_i - 4U_{i+1/2} + U_{i+1}) \end{aligned} \tag{62}$$

The semi-discrete form of Equation (61) becomes

$$\begin{aligned} U_i^{n+1} &= U_i^n + 6\frac{k\Delta t}{\Delta x^3}(U_{i-1} - 4U_{i-1/2} + 5U_i - 2U_{i+1/2}) \\ U_{i+1/2}^{n+1} &= U_{i+1/2}^n + 6\frac{k\Delta t}{\Delta x^3}(2U_{i-1/2} - 5U_{i-1} + 4U_{i+1/2} - U_{i+1}) \end{aligned} \tag{63}$$

This formulation uses only two cells in space. The modified expression of Equation (63) to be used with Godunov-type schemes with profile limiters is:

$$\begin{aligned} U_i^{n+1} &= U_i^n + 6\frac{k\Delta t}{\Delta x^3}[U_{i-1} - 2(U_{i-1,\mathrm{R}} + U_{i,\mathrm{L}}) + 5U_i - U_{i,\mathrm{R}} - U_{i+1,\mathrm{L}}] \\ U_{i,\mathrm{R}}^{n+1} &= U_{i+1,\mathrm{L}}^{n+1} = \frac{U_{i,\mathrm{L}}^n + U_{i+1,\mathrm{R}}^n}{2} \\ &\quad + 6\frac{k\Delta t}{\Delta x^3}[U_{i-1,\mathrm{R}} + U_{i,\mathrm{L}} - 5U_{i-1} + 2(U_{i,\mathrm{R}} + U_{i+1,\mathrm{L}}) - U_{i+1}] \end{aligned} \tag{64}$$

 

### 4.3. Discretization of source terms

In the proposed approach, source terms must be used to update both the cell edges and the cell averages. Consider the following PDE

$$\frac{\partial U}{\partial t} = S(U, x, t) \tag{65}$$

The explicit discretized form of Equation (63) is

$$\begin{aligned} U_i^{n+1} &= U_i^n + \Delta t S(U_i^n, x_i, t^n) \\ U_{i+1/2}^{n+1} &= U_{i+1/2}^n + \Delta t S(U_{i+1/2}^n, x_{i+1/2}, t^n) \end{aligned} \tag{66}$$

In some particular cases, an analytical expression can be found for the solution. This is the case of the degradation equation with kinetics of arbitrary order

$$\frac{\partial U}{\partial t} = -\rho U^{\beta}(x, t) \tag{67}$$

where $\rho$ is the degradation constant and $\beta$ is the order of the degradation kinetics. Equation (67) admits the following analytical solution

$$U(t) = \begin{cases} U_0 \exp(-\rho t) & \text{if } \beta = 1 \\ [U_0^{1-\beta} + (1-\beta)\rho t] & \text{if } \beta \neq 1 \end{cases} \tag{68}$$

Applying this solution between $t^n$ and $t^{n+1}$ to both the cell average and the cell edge value yields the following expressions

$$\begin{aligned} U_i^{n+1} &= \begin{cases} U_i^n \exp(-\rho \Delta t) & \text{if } \beta = 1 \\ [{U_i^n}^{1-\beta} + (1-\beta)\rho \Delta t]^{\frac{1}{1-\beta}} & \text{if } \beta \neq 1 \end{cases} \\ U_{i+1/2}^{n+1} &= \begin{cases} U_{i+1/2}^n \exp(-\rho \Delta t) & \text{if } \beta = 1 \\ [{U_{i+1/2}^n}^{1-\beta} + (1-\beta)\rho \Delta t]^{\frac{1}{1-\beta}} & \text{if } \beta \neq 1 \end{cases} \end{aligned} \tag{69}$$

Applying the analytical solution (69) eliminates the stability problems associated with explicit discretizations of the type (66).

## 5. COMPUTATIONAL EXAMPLES

### 5.1. The scalar advection equation

The present section illustrates the application of the proposed approach to the scalar advection equation (29). The new reconstruction algorithm was applied to the MUSCL scheme [1] and to

 

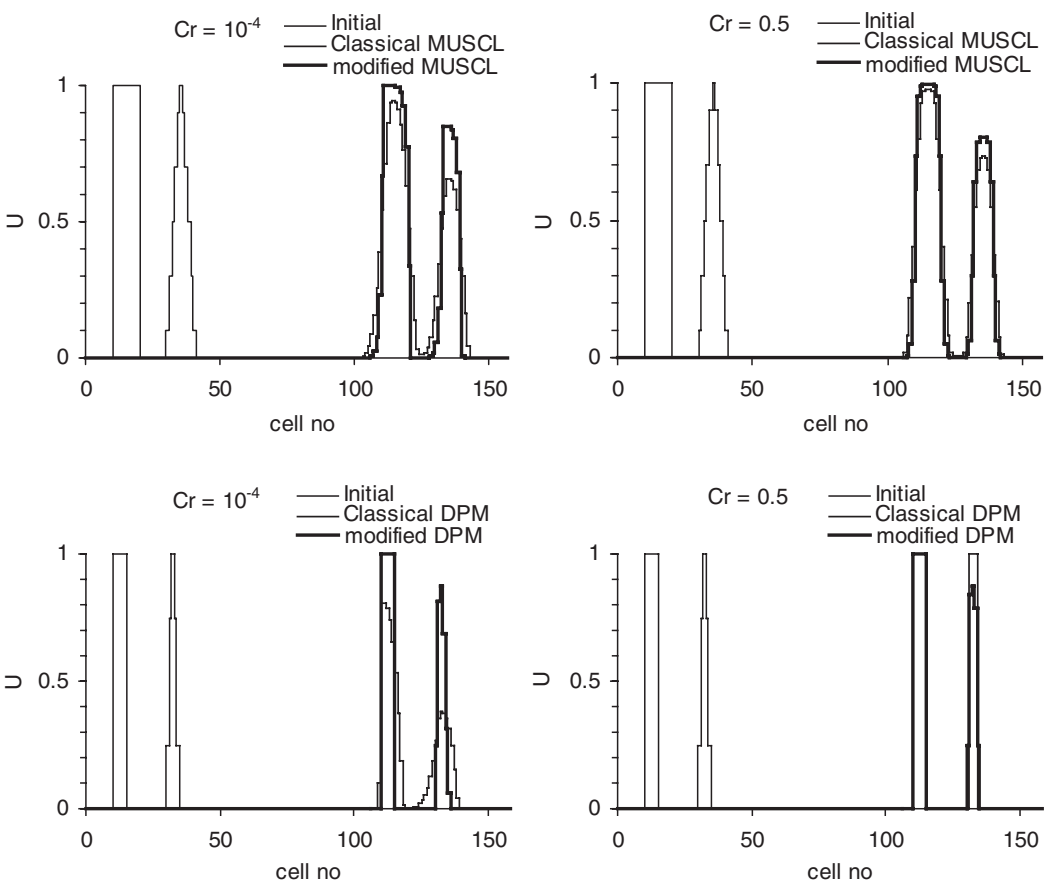


Figure 3. Advection of an 11-cells-wide square and a triangular profile computed by the classical and modified MUSCL (top) and DPM (bottom) schemes for Courant numbers of $10^{-4}$ and 0.5.

the DPM scheme [21]. The modified schemes were used to simulate the advection of a square and a triangular profiles in a uniform velocity field on a regular grid. In a first numerical experiment, the initial width of both profiles was set to 11 computational cells. In a second experiment, the initial width of the profiles was set to 5 cells. The results were compared to those given by the classical versions of the schemes. Figure 3 shows the numerical solutions obtained after transporting the 11-cells-wide profiles over 100 cells. Figure 4 shows the results obtained with the 5-cells-wide profiles. In both cases the values chosen for the Courant number were $10^{-4}$ and 0.5. These values of the Courant number were chosen because they represent very unfavourable conditions for the schemes (see Appendix A). For small values of Cr the phase error is maximum and the global amplification factor is minimum (however the global amplification factor of the modified MUSCL is much larger than that of the original MUSCL scheme). A Courant number of 0.5 corresponds to minimum values of the amplification factor for all wave numbers. The solutions are displayed in the form of horizontal

 

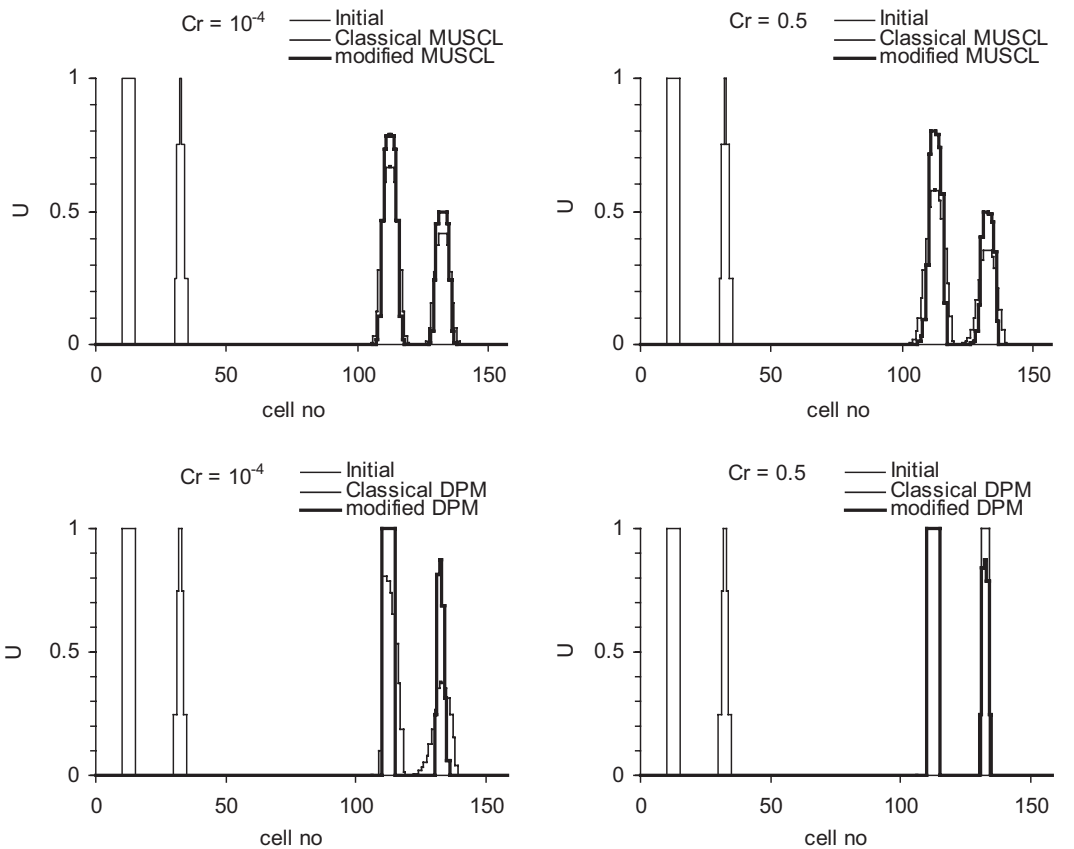


Figure 4. Advection of a 5-cells-wide square and a triangular profile computed by the classical and modified MUSCL (top) and DPM (bottom) schemes for Courant numbers of $10^{-4}$ and 0.5.

segments that give the average value of the computed variable over the cells. This type of representation allows local extrema or small oscillations in the computed profiles to be identified easily.

The computational results show that the profile limiters allow the phase error to be eliminated for both the original and modified versions of the MUSCL and DPM schemes. As expected from the amplitude portrait of the MUSCL scheme, the modified versions of the MUSCL and DPM preserve better the shape of the solutions than the original schemes. The modified DPM is seen to preserve the solutions better than the modified MUSCL. In particular the square profile is preserved for all values of the Courant number. Although damped to some extent, the triangular profile is much less affected by numerical diffusion with the modified DPM than in the modified MUSCL scheme. Also note that the original DPM scheme causes an artificial compression of the triangular profile for Courant numbers smaller than 0.5. This behaviour, that could be seen as valuable for the resolution of steep fronts, may turn out to be a strong drawback when oscillatory solutions are to be modelled (see Section 5.3).

 

Table I. Parameters for the advection–degradation test.

| Symbol | Meaning | Value |
|---|---|---|
| $L$ | Length of the computational domain | 200 m |
| $U_0$ | Initial condition | 0 everywhere |
| $U_b$ | Left-hand boundary condition | 1 |
| $\lambda$ | Advection velocity | 0.5 m s$^{-1}$ |
| $\rho$ | Degradation coefficient | $5 \times 10^{-2}$ s$^{-1}$ |
| $\Delta x$ | Cell width | 1 m |
| $\Delta t$ | Computational time step | 0.01, 0.1 and 1 s |

### 5.2. The scalar advection–degradation equation

This example shows an application of the modified DPM scheme to the scalar advection–degradation equation

$$\frac{\partial U}{\partial t} + \frac{\partial}{\partial x}(\lambda U) = -\rho U \tag{70}$$

where $\rho$ is the (positive, constant) degradation rate. The source term $-\rho U$ in Equation (70) is accounted for using a classical time splitting procedure, in which the advection terms are solved using the modified DPM and the degradation step is applied as in Equation (69). The parameters of the test are specified in Table I. The initial condition is the 11-cell-wide triangular profile used in the pure advection test. The results for $t = 100$ s are shown in Figure 5. Except for the slight damping that affects the maximum of the profile, the numerical solution is close to the analytical solution. Note that the smoothing of the peak in the numerical profile is introduced by the DPM scheme (as shown by the pure advection test in Section 5.1) and not by the discretization of the source terms.

### 5.3. The scalar advection–dispersion equation

This subsection presents an application to the linear advection–dispersion equation in conservation form

$$\frac{\partial U}{\partial t} + \frac{\partial}{\partial x}\left(\lambda U + k\frac{\partial^2 U}{\partial x^2}\right) = 0, \quad k < 0 \tag{71}$$

The advection and dispersion terms are taken into account separately using a classical time splitting procedure. The dispersion terms are discretized using the explicit version of the semi-discretization (60). The parameters of the test are specified in Table II. Figures 6 to 12 compare the numerical results obtained from all possible combinations between the classical and modified MUSCL and DPM schemes for the advection part, and the classical and modified discretization of the dispersion terms. The solutions are compared to a so-called 'converged solution' carried out using a cell size 100 times as small as that used for the comparisons. These numerical experiments allow the following conclusions to be drawn.

(a) The classical MUSCL scheme yields a strong damping of the oscillations for Courant numbers lower than 0.5 (see Figure 6). This damping is due to the profile limiter of

 

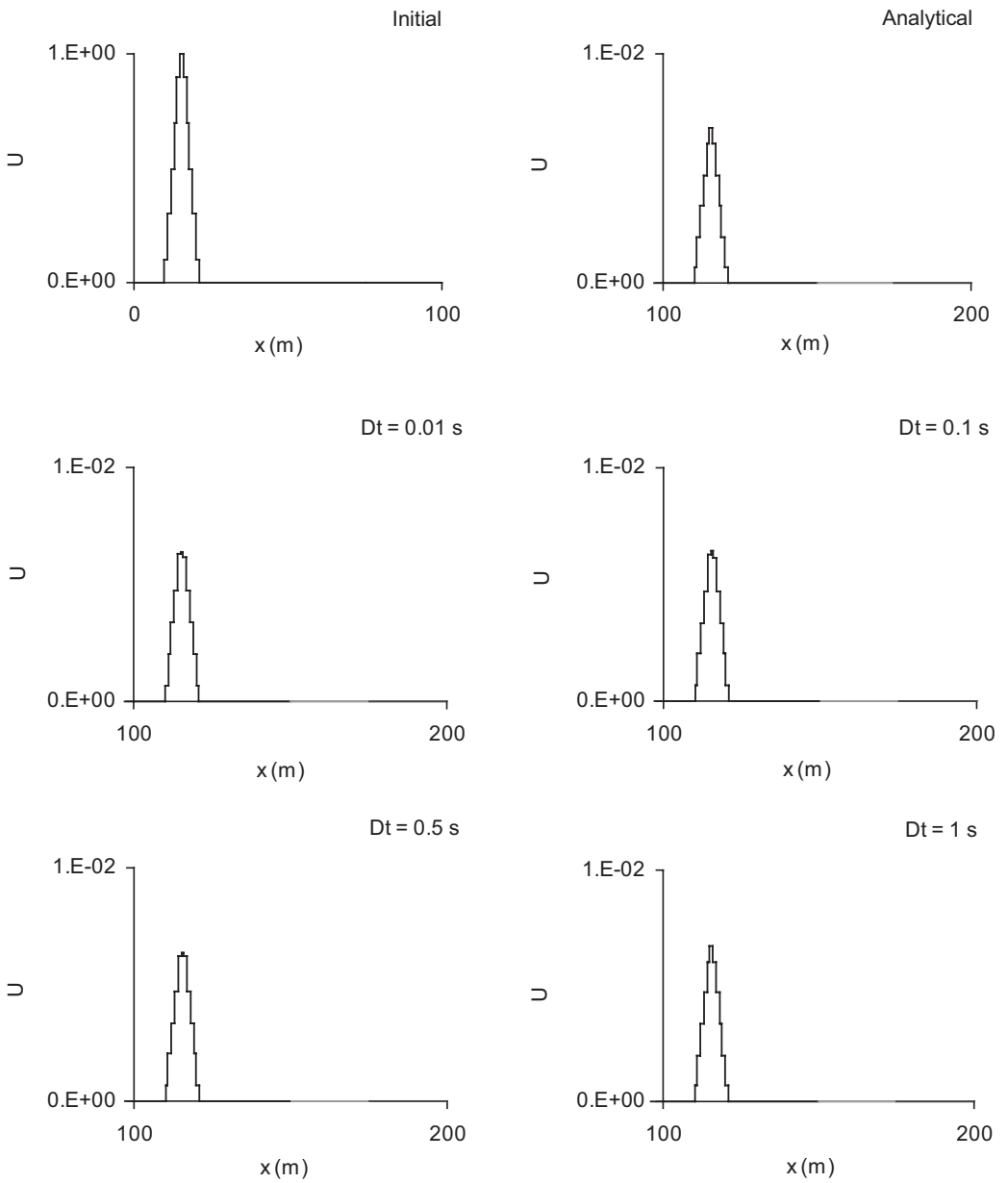


Figure 5. Linear advection–degradation simulation using the modified DPM scheme for the advection term and the discretization (69) for the degradation term.

 

Table II. Parameters for the advection–dispersion test.

| Symbol | Meaning | Value |
|---|---|---|
| $k$ | Dispersion coefficient | $10^{-2}\ \mathrm{m^3\ s^{-1}}$ |
| $L$ | Length of the computational domain | 200 m |
| $\lambda$ | Advection velocity | $1\ \mathrm{m\ s^{-1}}$ |
| $\Delta x$ | Cell width | 1 m (0.01 m for the converged solution) |
| $\Delta t$ | Computational time step | 0.01, 0.1, 0.5 and 1 s |

the scheme, the role of which is precisely to eliminate possible oscillations resulting from the higher-order reconstruction.

(b) The modified MUSCL scheme allows the damping of the oscillations to be reduced to some extent (see e.g. the profiles obtained for $\Delta t = 0.01$ s in Figure 7). The behaviour of the numerical solution is less dependent on the Courant number than when the classical MUSCL scheme is used.

(c) Using the modified discretization for the dispersion terms in combination with the original or modified MUSCL scheme is not seen to introduce any significant improvement (Figures 8 and 9). Moreover, the explicit version of the semi-discretization (60) being less stable than the original discretization (48), computational time steps larger than 0.5 s cannot be used in these simulations (see Appendix B).

(d) The combination of the classical DPM and the classical discretization (48) of the dispersion terms produces unstable solutions for time steps smaller than 0.5 (Figure 10). This is due to the over-compressive character of the original DPM scheme for values of the Courant number smaller than 0.5. Note that artificial compression was already observed in the pure advection simulations (Figures 3 and 4). This over-compressive character of the original DPM, that may be valuable for the accurate resolution of shocks and contact discontinuities, causes problems when oscillatory behaviours are to be modelled. Profile compression indeed leads to an artificial amplification of the shorter waves, which may result in unstable solutions when dispersion-like equations are to be solved.

(e) The modified DPM allows overcompression to be eliminated to a large extent and the stability of the numerical solution to be restored (Figure 11). Although the amplitude of the shorter waves is exaggerated for small time steps, this was not observed to result in unstable behaviours. Note that the explicit version of the semi-discretization (60) for the dispersion terms is less stable than its classical counterpart and that time steps larger than 0.5 produced unstable solutions (see Appendix B for a stability analysis).

(f) The combination of the classical DPM and the modified discretization for the dispersion terms yields unstable solutions for all possible values of $\Delta t$. This was expected from (d) because the original DPM causes instability of the solution for time steps smaller than 0.5 s, while the modified discretization of the dispersion terms is unstable for $\Delta t > 0.5$ s.

(g) The modified DPM and the modified semi-discretization of the dispersion terms yield a better solution than in (e) for the smaller time steps (Figure 12). However the oscillations become asymmetrical for larger time steps.

 

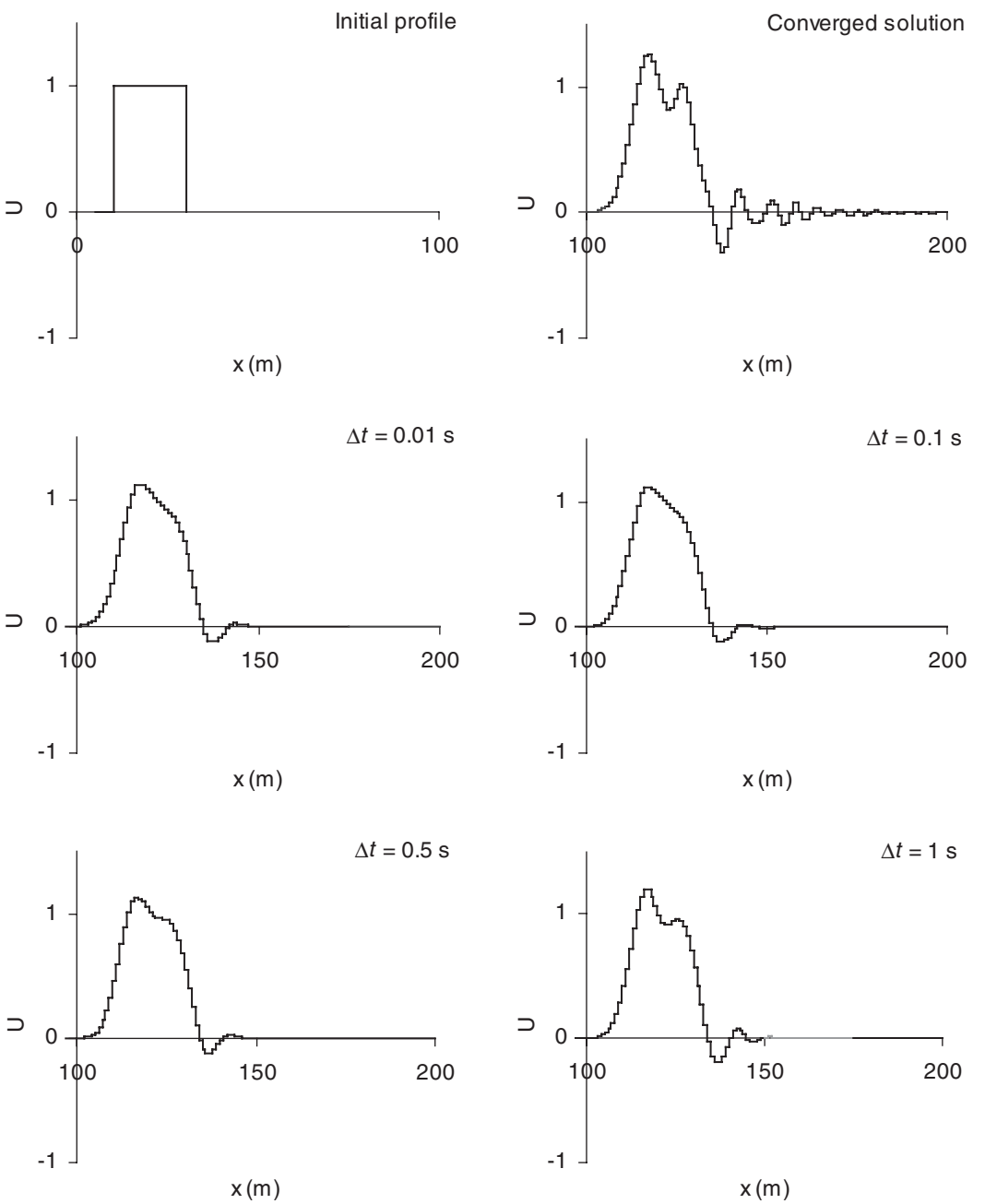


Figure 6. Linear, conservative advection–dispersion of a square profile. Solutions at $t = 100$ s using the classical MUSCL scheme and the classical discretization for the dispersion terms.

 

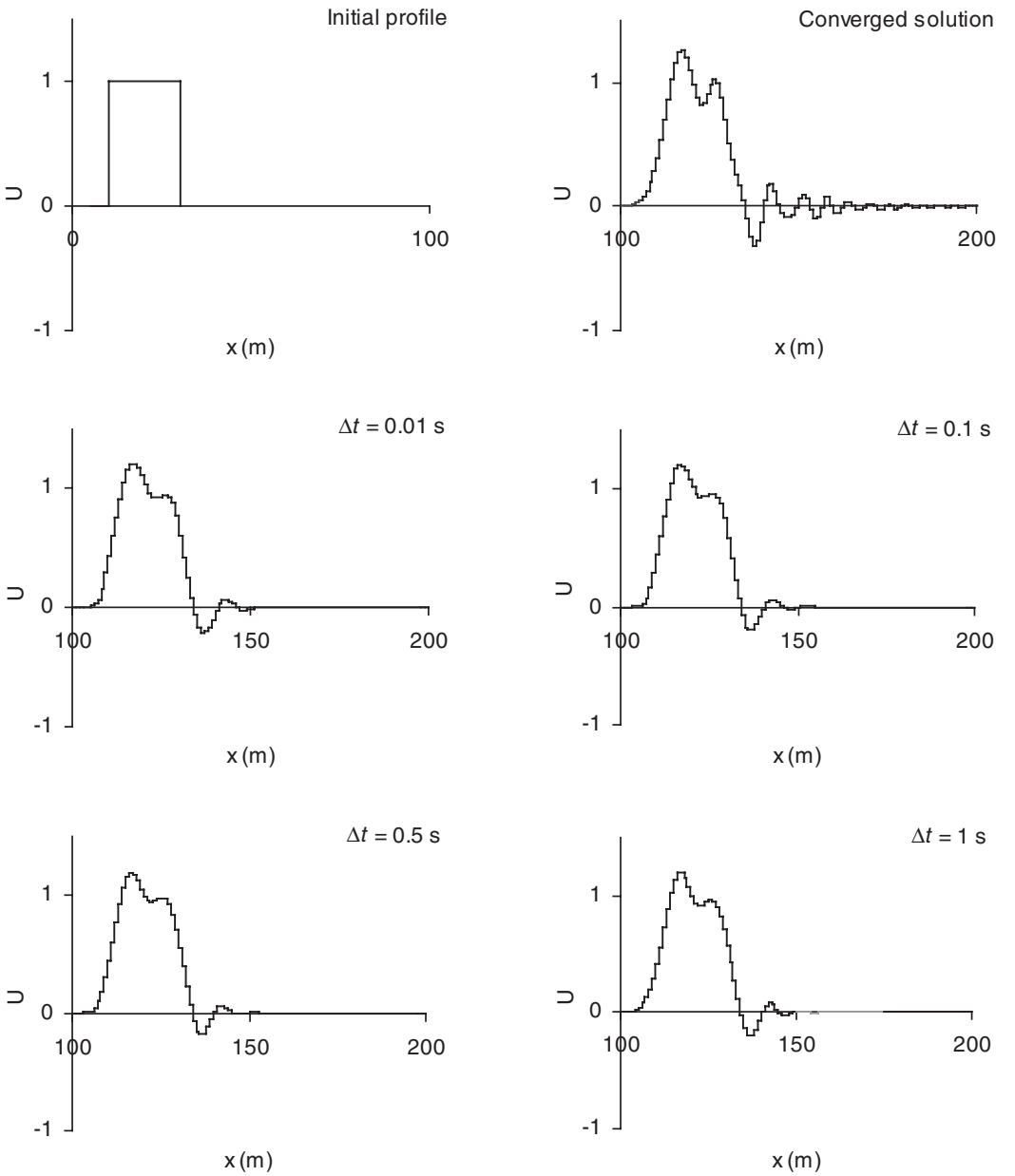


Figure 7. Linear, conservative advection–dispersion of a square profile. Solutions at $t = 100$ s using the modified MUSCL scheme and the classical discretization for the dispersion terms.

 

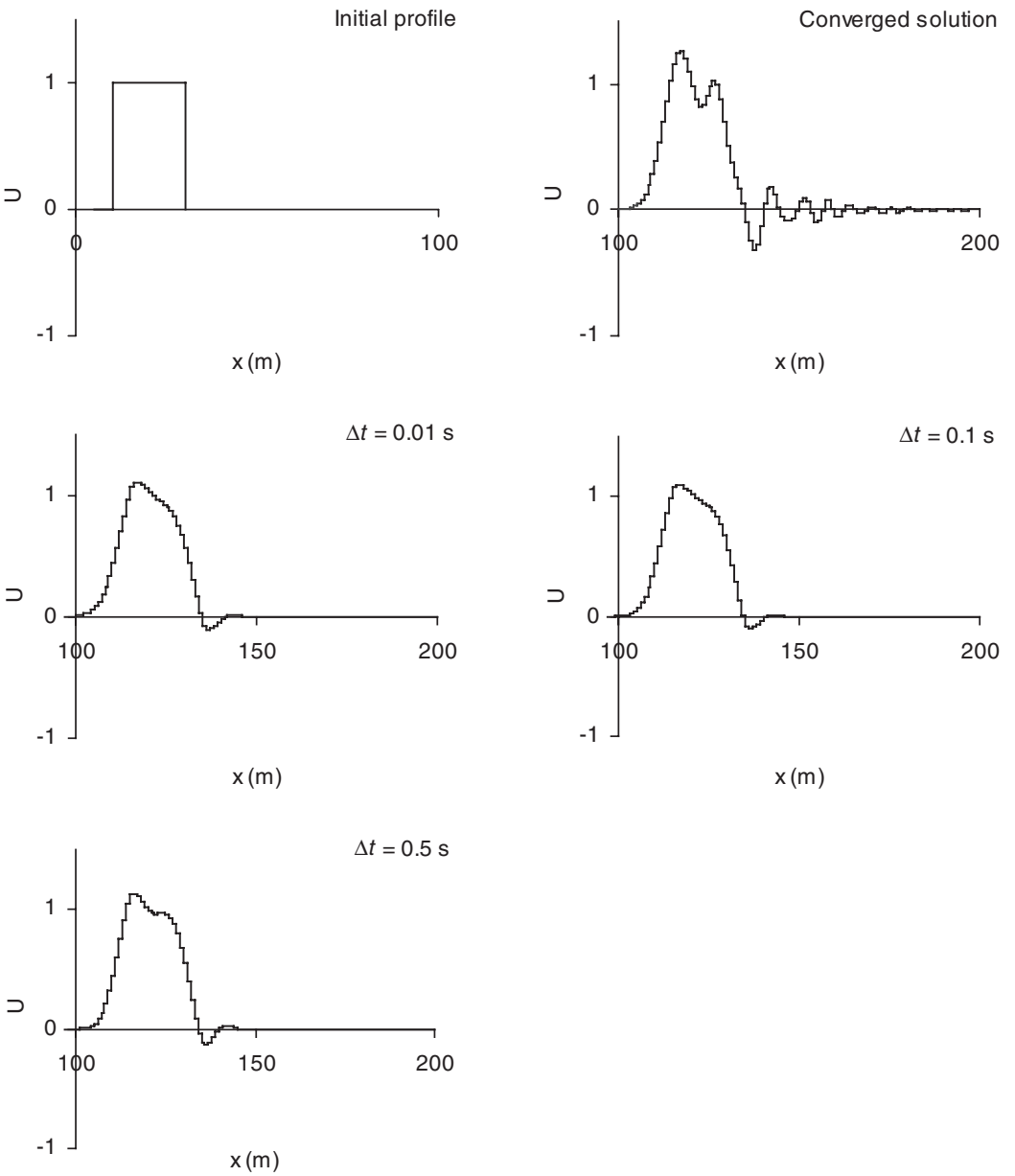


Figure 8. Linear, conservative advection–dispersion of a square profile. Solutions at $t = 100$ s using the classical MUSCL scheme and the modified discretization for the dispersion terms. The explicit discretization of the dispersion terms is unstable for $\Delta t = 1$ s.

 

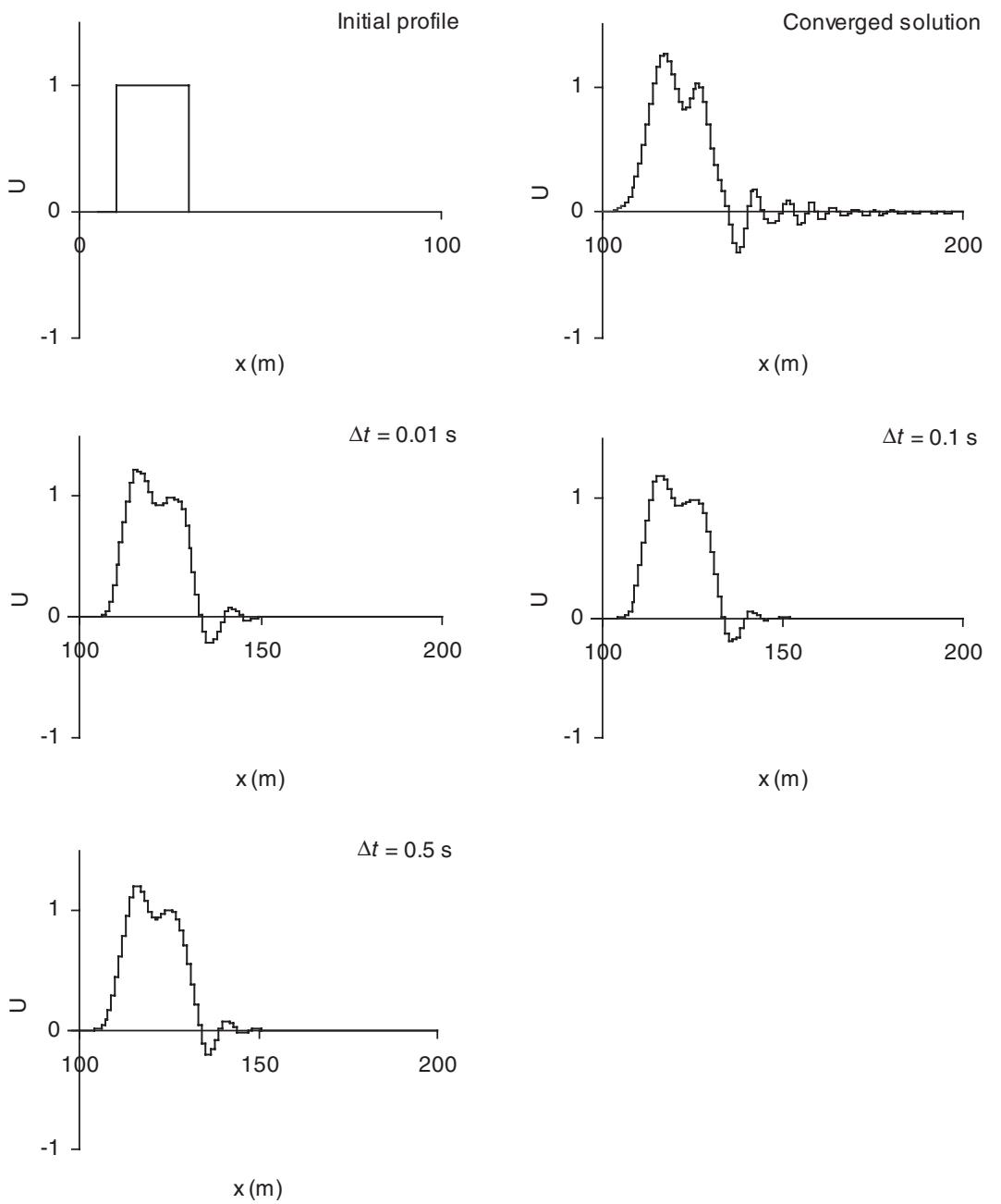


Figure 9. Linear, conservative advection–dispersion of a square profile. Solutions at $t = 100$ s using the modified MUSCL scheme and the modified discretization for the dispersion terms. The explicit discretization of the dispersion terms is unstable for $\Delta t = 1$ s.

 

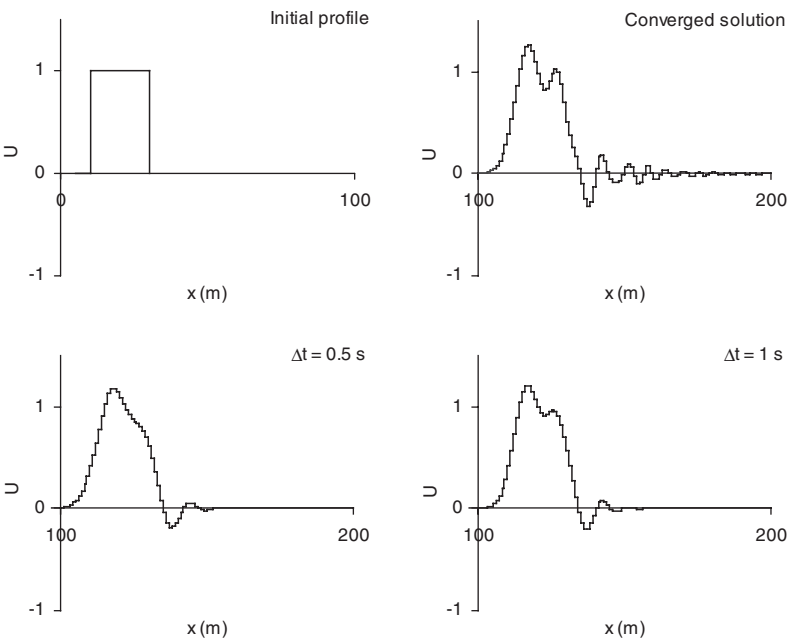


Figure 10. Linear, conservative advection–dispersion of a square profile. Solutions at $t = 100$ s using the classical DPM and the classical discretization for the dispersion terms. The solution is unstable for Courant numbers smaller than 0.5.

Form these numerical experiments it can be concluded that the modified approach for profile reconstruction allows the over-compressive behaviour of the DPM scheme to be eliminated to a large extent. The modified DPM gives better numerical solutions than the MUSCL scheme. Although the modified discretization of the dispersion terms brings a slight improvement over the classical discretization, the most influential factor on the quality of the numerical solution is the discretization of the advective terms in the PDE.

### *5.4. The one-dimensional shallow water equations*

In this subsection the application of the proposed approach to the one-dimensional shallow water equations is presented. In the absence of friction and bottom slope these equations can be written in the form (1) by defining **U** and **F** as

$$\mathbf{U} = \begin{bmatrix} h \\ q \end{bmatrix}, \quad \mathbf{F} = \begin{bmatrix} q \\ q^2/h + gh^2/2 \end{bmatrix} \tag{72}$$

where $g$ is gravitational acceleration, $h$ is the water depth and $q$ is the discharge per unit width. Note that introducing friction and a bottom slope leads to introducing a source term in the right-hand side of Equation (1). The Jacobian matrix **A** in Equation (2) and the matrix

 

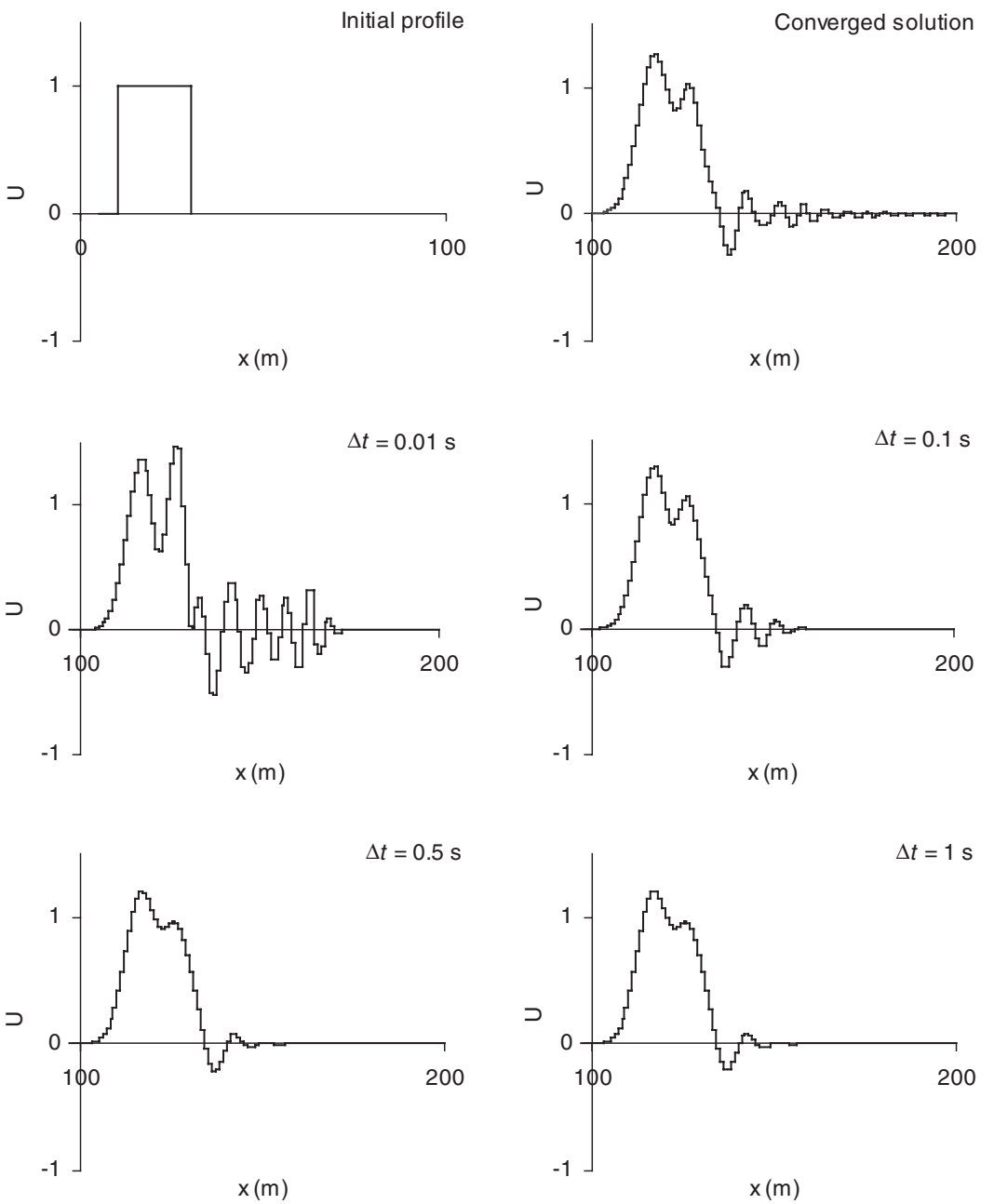


Figure 11. Linear, conservative advection–dispersion of a square profile. Solutions at $t = 100$ s using the modified DPM and the classical discretization for the dispersion terms.

 

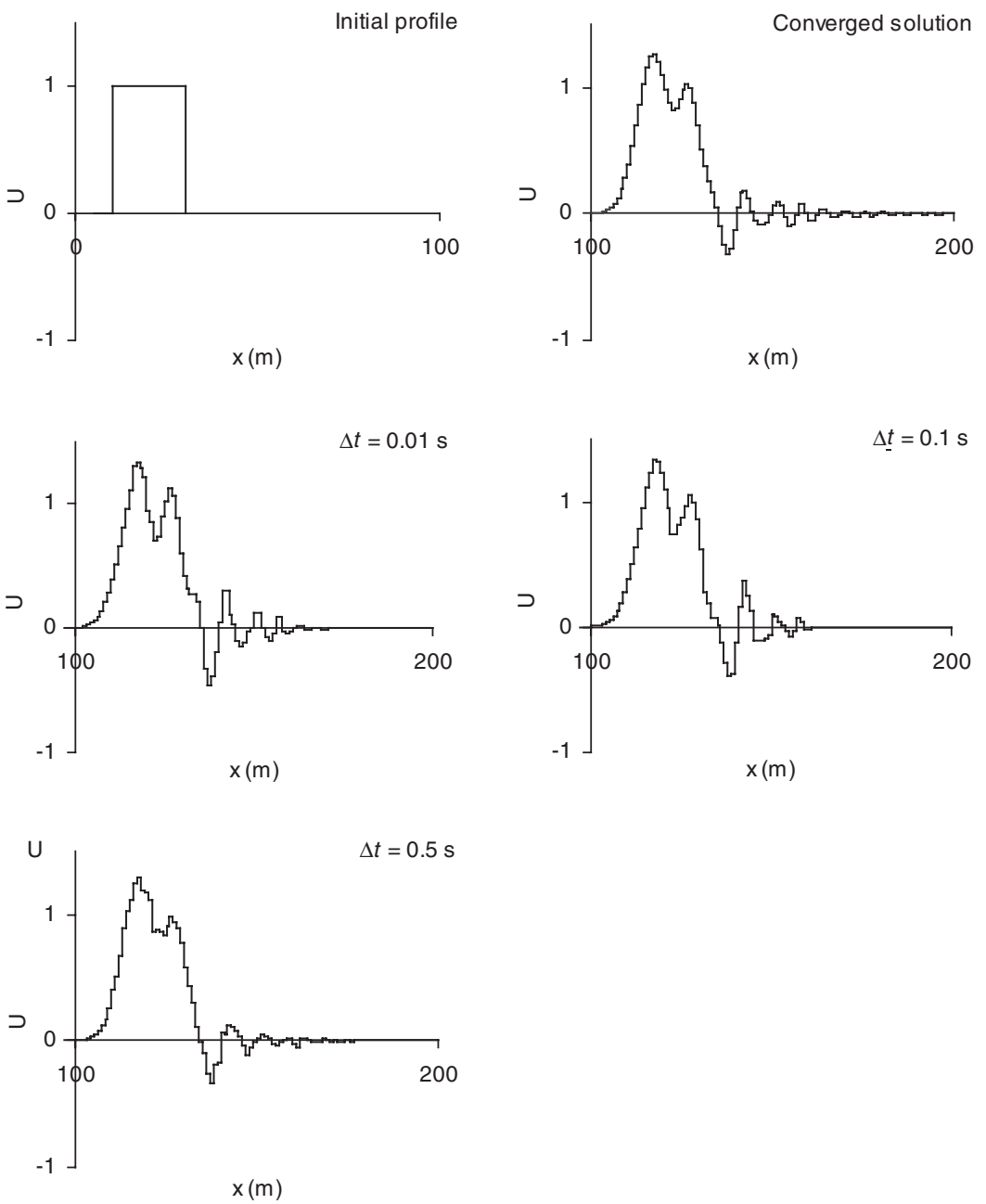


Figure 12. Linear, conservative advection–dispersion of a square profile. Solutions at $t = 100$ s using the modified DPM and the modified discretization for the dispersion terms. The explicit discretization of the dispersion terms is unstable for $\Delta t > 0.5$ s.

 

Table III. Parameters for the one-dimensional dambreak test.

| Symbol | Meaning | Value |
|---|---|---|
| $g$ | Gravitational acceleration | 9.81 m s$^{-2}$ |
| $h_L$ | Initial water depth on the left-hand side of the dam | 50 m |
| $h_R$ | Initial water depth on the right-hand side of the dam | 1 m |
| $L$ | Length of the computational domain | 1000 m |
| $q_0$ | Initial unit discharge | 0 m$^2$ s$^{-1}$ |
| $x_0$ | Initial location of the dam | 500 m |
| $\Delta t$ | Computational time step | $10^{-2}$, $10^{-3}$ and $10^{-4}$ s |
| $\Delta x$ | Cell width | 10 m |

**K** of its eigenvectors are

$$\mathbf{A} = \begin{bmatrix} 0 & 1 \\ c^2 - u^2 & 2u \end{bmatrix}, \quad \mathbf{K} = \begin{bmatrix} 1 & 1 \\ \lambda^{(1)} & \lambda^{(2)} \end{bmatrix} \tag{73}$$

where $c = (gh)^{1/2}$ is the celerity of pressure waves in still water, $u = q/h$ is the velocity and $\lambda^{(p)}$ $(p = 1, 2)$ are the eigenvalues of **A**:

$$\begin{aligned} \lambda^{(1)} &= u - c \\ \lambda^{(2)} &= u + c \end{aligned} \tag{74}$$

It is easy to check that the wave strengths are given by

$$\alpha^{(1)} = \alpha^{(2)} = h \tag{75}$$

The performance of the classical and modified DPM schemes is compared on a dambreak simulation. The parameters of the test are given in Table III. Figure 13 shows the water depth profiles obtained at $t = 10$ s for various values of the computational time step $\Delta t$. Note that the maximum permissible time step in this simulation is $\Delta t = 0.3$ s, which corresponds to a maximum Courant number of 1 at the location of the initial discontinuity. The time steps of $10^{-4}$, $10^{-3}$ and $10^{-2}$ s used in the simulations correspond to Courant numbers of $3.3 \times 10^{-4}$, $3.3 \times 10^{-3}$ and $3.3 \times 10^{-2}$, respectively. It can be seen from Figure 13 that the classical DPM scheme allows for a better representation of the shock than the modified DPM for small values of the time step (Figure 13, upper right). This is due to the (previously mentioned) over-compressive character of the original DPM. For larger values of $\Delta t$ however, the modified DPM leads to a steeper representation of the shock than the original scheme. Numerous simulations using various values of $\Delta t$ show that the modified DPM usually captures 90% of the height of the shock within 3 cells. Note that the modified DPM introduces a small, artificial depression in the computed water depth at the junction between the intermediate region of constant state and the head of the rarefaction wave. This non-monotony of the profile was not seen to exert any influence on the stability of the solution.

 

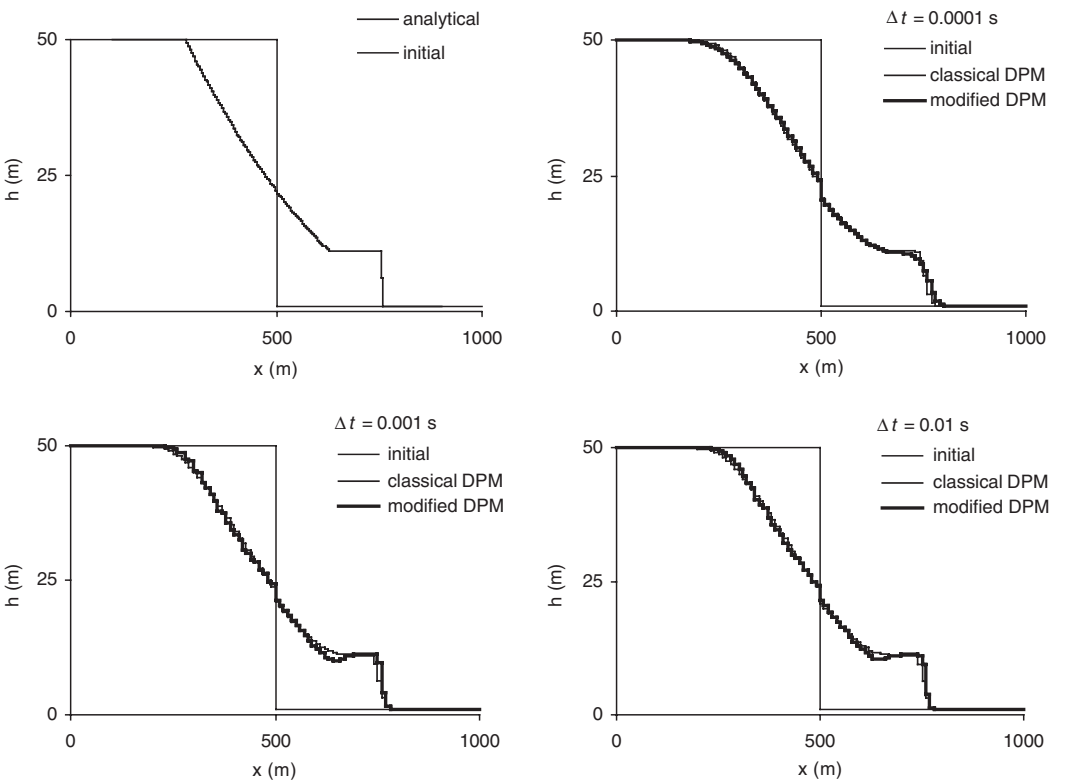


Figure 13. Dambreak simulation. Solutions obtained at $t = 10$ s using the classical and modified DPM for various values of the time step.

### 5.5. *The two-dimensional shallow water equations*

The last example consists of an application of the classical and modified DPM schemes to the two-dimensional shallow water equations.

$$\frac{\partial \mathbf{U}}{\partial t} + \frac{\partial \mathbf{F}}{\partial x} + \frac{\partial \mathbf{G}}{\partial y} = 0 \tag{76}$$

where $\mathbf{U}$, $\mathbf{F}$ and $\mathbf{G}$ are given by

$$\mathbf{U} = \begin{bmatrix} h \\ q \\ r \end{bmatrix}, \quad \mathbf{F} = \begin{bmatrix} q \\ q^2/h + gh^2/2 \\ qr/h \end{bmatrix}, \quad \mathbf{G} = \begin{bmatrix} r \\ qr/h \\ r^2/h + gh^2/2 \end{bmatrix} \tag{77}$$

where $q$ and $r$ are the unit discharges in the $x$- and $y$-directions, respectively. In the present test, the contributions of the $x$- and $y$-fluxes to the solution were taken into account separately via a time splitting procedure. Consequently, the profiles were reconstructed independently from each other in the $x$- and $y$-directions. This was intentional because the independence

 

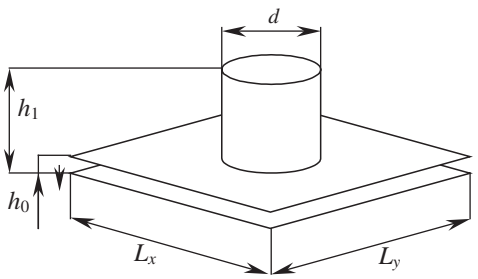


Figure 14. Geometry of the circular dambreak problem.

Table IV. Characteristics of the simulations for the two-dimensional dambreak test.

| Symbol | Meaning | Value in simulation 1 | Value in simulation 2 |
|---|---|---|---|
| $d$ | Diameter of the circular region | 32 m | 32 m |
| $g$ | Gravitational acceleration | 9.81 m s$^{-2}$ | 9.81 m s$^{-2}$ |
| $h_0$ | Initial water depth outside the circular region | 1 m | 1 m |
| $h_1$ | Initial water depth inside the circular region | 1.5 m | 1.5 m |
| $L_x$ | Domain size in the $x$-direction | 4000 m | 4000 m |
| $L_y$ | Domain size in the $x$-direction | 4000 m | 4000 m |
| $\Delta x$ | Cell size in the $x$-direction | 40 m | 40 m |
| $\Delta y$ | Cell size in the $x$-direction | 40 m | 4 m |

between the $x$- and $y$-reconstructions usually leads to undesirable phenomena, such as polarization of the flow variables in the direction of the mesh and other manifestations of anisotropy in the behaviour of the solution. The purpose of the test was to investigate whether the modified reconstruction could contribute to reduce such anisotropy effects. The numerical solution of the idealised so-called circular dambreak problem illustrated by Figure 14 was investigated. Consider a domain of dimensions $L_x \times L_y$, where water is initially immobile, at a uniform depth $h_0$, except in a circular region of diameter $d$ where the depth is $h_1$. Since the problem has a radial symmetry, the solution is also expected to exhibit such symmetry. More considerations about the solution of this problem can be found in References [23]. The parameters of the test can be found in Table IV.

Figures 15 and 16 show the results obtained using the first-order, original Godunov scheme [24], the original DPM and the modified DPM method. These solutions are compared to a converged solution, obtained on a square grid of width 4 m. Tables V and VI summarize the results obtained for the two simulations. The shock amplitude denotes the difference between the water depth at the crest of the shock and the initial water depth $h_0$. Owing to the splitting between the two directions $x$ and $y$, the computational solution is anisotropic, and therefore the shock amplitude cannot be expected to be the same in all directions. Tables V and VI provide the minimum and maximum shock amplitudes that were obtained for the three schemes. It also provides the 'amplitude ratio' for each scheme, that is the ratio of the maximum amplitude observed in the solution to that of the converged solution. The isotropy ratio is also indicated, giving the ratio of the minimum to the maximum shock amplitude. An isotropy ratio of 1 indicates a totally isotropic solution, whereas a ratio of 0 indicates a total collapsing of the

 

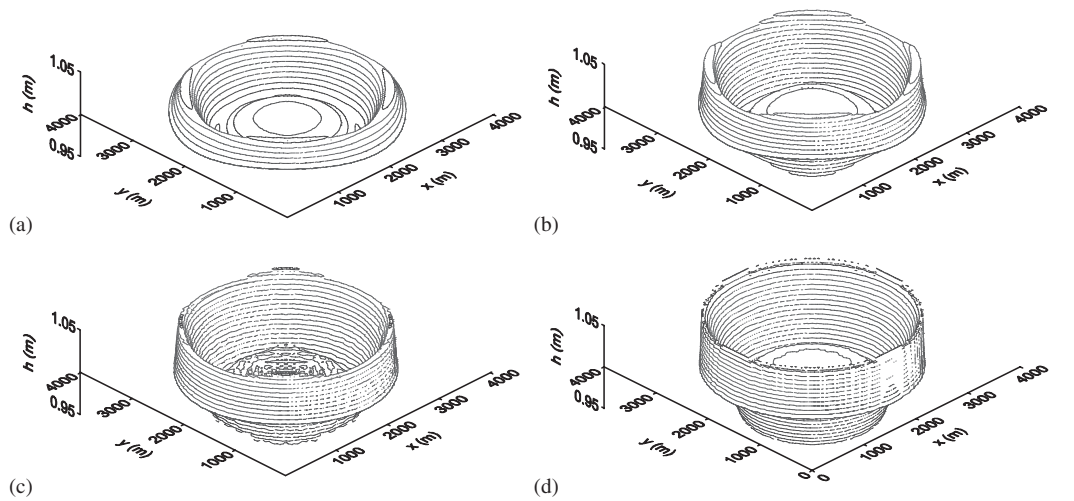


Figure 15. Simulation 1. Contour lines of the water depth after 400 s computed using the original Godunov scheme (a), the classical DPM (b) and the DPM in combination with the proposed approach (c) and compared to a converged solution (d).

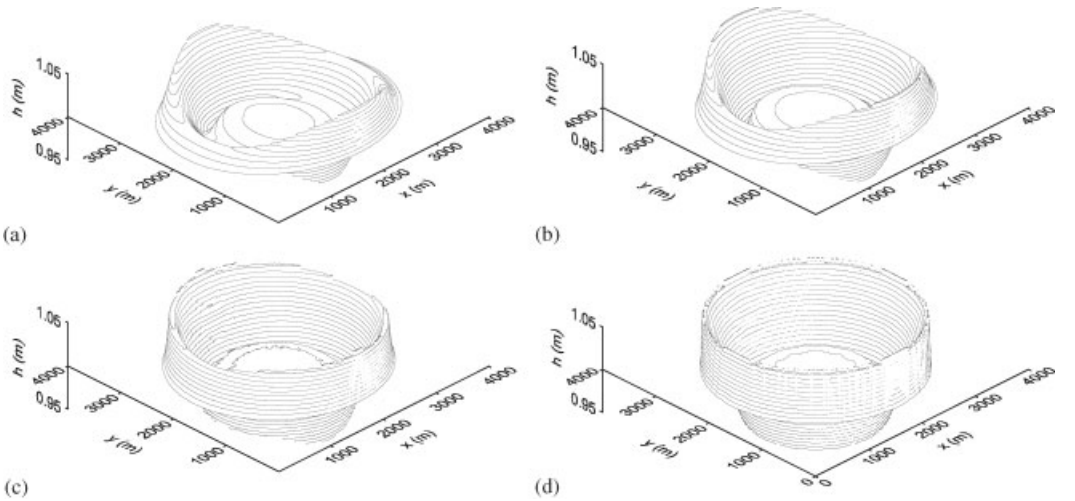


Figure 16. Simulation 2. Contour lines of the water depth after 400 s computed using the original Godunov scheme (a), the classical DPM (b) and the DPM in combination with the proposed approach (c) and compared to the converged solution (d).

wave in one particular direction. In simulation 1, the minimum shock amplitude is found in the $x$- and $y$-directions from the centre of the wave ring, whereas the lines of maximum amplitude follow the diagonals of the grid. In simulation 2, the large aspect ratio $\Delta x/\Delta y$ introduces substantial numerical diffusion in the $x$-direction (where the grid is coarser) and

 

Table V. Simulation 1. Minimum and maximum shock amplitude. The minimum amplitude is found along the diagonal of the grid from the centre of the wave. The maximum shock amplitude is reached along the $x$- and $y$-directions from the centre of the wave. The amplitude ratio indicates the ratio of the maximum amplitude of the solution with that of the converged one. The isotropy ratio is that of the minimum shock amplitude to the maximum shock amplitude.

| Solution | Minimum shock amplitude | Maximum shock amplitude | Amplitude ratio | Isotropy ratio |
|---|---|---|---|---|
| Converged | 0.073 | 0.073 | 1.0 | 1.0 |
| Godunov | 0.026 | 0.035 | 0.48 | 0.74 |
| Original DPM | 0.041 | 0.060 | 0.82 | 0.68 |
| DPM modified | 0.051 | 0.064 | 0.88 | 0.80 |

Table VI. Simulation 2. Minimum and maximum shock amplitude. The minimum amplitude is found in the $x$-direction from the centre of the wave, the maximum is reached in the $x$-direction. The amplitude ratio indicates the ratio of the maximum amplitude of the solution with that of the converged one. The isotropy ratio is that of the minimum shock amplitude to the maximum shock amplitude.

| Solution | Minimum shock amplitude | Maximum shock amplitude | Amplitude ratio | Isotropy ratio |
|---|---|---|---|---|
| Converged | 0.073 | 0.073 | 1.0 | 1.0 |
| Godunov | 0.018 | 0.062 | 0.85 | 0.29 |
| Original DPM | 0.027 | 0.064 | 0.88 | 0.42 |
| DPM modified | 0.048 | 0.066 | 0.90 | 0.73 |

the minimum amplitude is found along the $x$-direction, while the maximum is located along the $y$-direction from the centre of the circular wave.

Table V shows that for simulation 1 the Godunov scheme induces strong damping in the solution, while the isotropy ratio remains satisfactory. As shown by Table VI, the amplitude ratio increases in simulation 2, whereas the isotropy ratio decreases substantially. The original DPM yields a better solution in as much as the shock is captured within fewer cells and the minimum shock amplitude is larger than with the Godunov scheme. Still, the isotropy ratio of 0.42 obtained in simulation 2 cannot be seen as satisfactory. The best results are obtained with the DPM in combination with the proposed approach, as shown by the figures and tables. It should be noted however that the splitting between the $x$- and $y$-directions, together with the divergent flow in the central region resulting from the collapsing of the dam, induce small wiggles in the central region of the wave (visible in Figure 15(c)). The wiggles are not present in the solution of calculation 2 (see Figure 16(c)), because the slight numerical diffusion induced by the coarse mesh in the $x$-direction was sufficient to eliminate them.

In order to improve further the performance of the scheme, a genuinely unsplit multi-dimensional version should be designed. However, even in its split version, the proposed method can be seen to yield a dramatic improvement over usual schemes.

## 6. CONCLUDING REMARKS

A method has been presented for increasing the accuracy of Godunov-type schemes. The key factor for the accuracy of these schemes is the way the reconstruction is carried out and

 

more particularly, how the values of the flow variables at the edges of the computational cells are computed. Classical higher-order Godunov methods use the average values over several cells in each direction of space. The proposed method uses the point value of the (already reconstructed) variable within the computational cells immediately neighbouring the interface under consideration. This yields a substantial increase in scheme performance without a need for an increasing the stencil of the scheme. Applications to the MUSCL and DPM schemes for one-dimensional linear advection simulations show that damping of linear or quasi-linear waves, that is generally mentioned to be a drawback of Godunov-type schemes, is limited considerably by the new approach. When applied to the DPM reconstruction, the proposed approach is also seen to reduce considerably the over-compression of steep fronts compared to the classical reconstruction method. Such over-compression is often a source of numerical instability when the PDEs to be solved include dispersion terms. This is because the shorter waves in the (oscillatory) solution are amplified artificially by over-compressive limiters. The modified reconstruction technique proposed here allows this drawback to be eliminated. The method is extended to PDEs containing higher-order derivatives, such as advection–dispersion equations, and to hyperbolic systems of conservation laws. Numerical experiments indicate a gain of accuracy for the numerical solutions. Implementation of the method for two-dimensional shallow water simulations show that the quality of numerical solutions in the presence of weak shocks (that are prone to fast damping) also considerably increases, even though the reconstruction procedure is split with respect to space. The quality of the solution remains acceptable even when the computational grid is considerably distorted. Nonetheless, the presence of small wiggles in regions of strong flow divergence indicates that an unsplit reconstruction method should give better results. The coupling of the present method with genuinely unsplit, multi-dimensional methods such as that developed in Reference [25] is the subject of ongoing research.

## APPENDIX A. PHASE AND AMPLITUDE PORTRAITS OF THE CLASSICAL AND MODIFIED MUSCL SCHEMES

This appendix is devoted to the phase and amplitude portraits of the original and compact MUSCL schemes applied to the linear, scalar advection equation

$$\frac{\partial U}{\partial t} + \frac{\partial}{\partial x}(\lambda U) = 0$$

where the advection velocity $\lambda$ is a constant. Note that such portraits are valid for linear schemes with constant coefficients. Therefore, the analysis is carried out for a uniform cell size on the scheme without slope limiter. In what follows, the advection velocity is assumed to be positive.

### *A1. Original MUSCL scheme*

In the original MUSCL scheme the profile $\tilde{U}_i^n(x)$ of the variable $U$ is reconstructed over the cell $i$ at time level $n$ using a linear profile

$$\tilde{U}_i^n(x) = U_i^n + (x - x_i)\beta_i^n \tag{A1}$$

 

where $U_i^n$ is the average value of $U$ over the cell $i$ at $t^n$, $x_i$ is the abscissa of the centre of the cell $i$ and $\beta_i^n$ is the slope of the reconstructed profile, given by

$$\beta_i^n = \frac{U_{i+1}^n - U_{i-1}^n}{x_{i+1} - x_{i-1}} = \frac{U_{i+1}^n - U_{i-1}^n}{2\Delta x} \tag{A2}$$

The average flux $F_{i+1/2}^{n+1/2}$ between $t^n$ and $t^{n+1}$ at the interface $i + \frac{1}{2}$ between the cells $i$ and $i+1$ is given by

$$F_{i+1/2}^{n+1/2} = \lambda U_{i+1/2}^{n+1/2} \tag{A3}$$

where $U_{i+1/2}^{n+1/2}$ is the average value of $U$ at the interface $i + \frac{1}{2}$ between $t^n$ and $t^{n+1}$. $U_{i+1/2}^{n+1/2}$ is computed as the average of the reconstructed profile over the domain of dependence of the interface $i + \frac{1}{2}$, that is between $x_{i+1/2} - Cr\Delta x$ and $x_{i+1/2}$

$$U_{i+1/2}^{n+1/2} = \int_{x_{i+1/2} - Cr\Delta x}^{x_{i+1/2}} \tilde{U}_i^n(x)\,\mathrm{d}x \tag{A4}$$

Substituting Equations (A1) and (A2) into Equation (A4) and noting that $x_{i+1/2} = x_i + \Delta x$, yields

$$U_{i+1/2}^{n+1/2} = \frac{Cr - 1}{4} U_{i-1}^n + U_i^n + \frac{1 - Cr}{4} U_{i+1}^n \tag{A5}$$

Substituting Equation (A5) into Equation (A3) leads to the following expression for the flux

$$F_{i+1/2}^{n+1/2} = \left( \frac{Cr - 1}{4} U_{i-1}^n + U_i^n + \frac{1 - Cr}{4} U_{i+1}^n \right) \lambda \tag{A6}$$

The average value $U_i^{n+1}$ of $U$ over the cell $i$ at time $t^{n+1}$ is given by the classical balance over the cell

$$U_i^{n+1} = U_i^n + \frac{\Delta t}{\Delta x}(F_{i-1/2}^{n+1/2} - F_{i+1/2}^{n+1/2}) \tag{A7}$$

Substituting Equation (A6) into Equation (A7) gives the following expression

$$\begin{aligned} U_i^{n+1} &= -\alpha U_{i-2}^n + (Cr + \alpha) U_{i-1}^n + (1 - Cr + \alpha) U_i^n - \alpha U_{i+1}^n \\ \alpha &= \frac{1 - Cr}{4} Cr \end{aligned} \tag{A8}$$

The phase and amplitude portraits are determined by seeking a solution in the form

$$U_i^n = U_0 \exp(\mathrm{i}\sigma\Delta x + n\mu\Delta t) \tag{A9}$$

where $\sigma$ is a pure imaginary number and $\mu$ is a complex number with a real part $\mu_r$ and an imaginary part $\mu_i$. Substituting Equation (A9) into Equation (A8) and dividing by $U_i^n$ yields

 

the following equation on $\sigma$ and $\mu$:

$$\exp(\mu\Delta t) = -\alpha\,\exp(-2\sigma\Delta x) + (Cr+\alpha)\exp(-\sigma\Delta x) \\ +1 - Cr + \alpha - \alpha\exp(\sigma\Delta x) \tag{A10}$$

The (complex) quantity $A_N = \exp(\mu\Delta t)$, also called the amplification factor, is the factor by which $U_i^n$ is multiplied form one time step to the next. If the modulus of $A_N$ is smaller than unity and the solution is stable. Equation (A10) can be rewritten as

$$A_N(Cr,\sigma\Delta x) = -\alpha\exp(-2\sigma\Delta x) + (Cr+\alpha)\exp(-\sigma\Delta x) \\ +1 - Cr + \alpha - \alpha\exp(\sigma\Delta x) \tag{A11}$$

It can be observed with Van Leer [1] that the following equality holds:

$$A_N(1-Cr,\sigma\Delta x) = \exp(-\sigma\Delta x)\overline{A_N(Cr,\sigma\Delta x)} \tag{A12}$$

where the upper bar indicates the complex conjugate. Equation (A12) implies in particular that

$$|A_N(1-Cr)| = |A_N(Cr)| \tag{A13}$$

Therefore the minimum of the modulus of the amplification factor is obtained for $Cr = \frac{1}{2}$. It is equal to unity for $Cr = 0$ and 1 and larger than unity for $Cr < 0$ or $> 1$. These are the classical stability properties of the MUSCL scheme.

Another quantity of interest is the celerity factor $C_N$ defined as

$$C_N = -\frac{\mathrm{Im}(A_N)}{\sigma\Delta x Cr} \tag{A14}$$

that expresses the ratio of the propagation celerity of the numerical solution to the analytical solution. Eventually, the performance of the scheme can be analysed in the light of the global amplitude convergence factor $A_G$ defined as [26]

$$A_G = |A_N|^{1/Cr} \tag{A15}$$

The global amplitude convergence factor indicates the evolution of the scheme performance with the computational time step for a given discretization of the computational domain. Assuming that the cell size $\Delta x$ is known, the number $M$ of time steps needed to compute the solution over a given time interval is proportional to the inverse of the time step, that is, to the inverse $1/Cr$ of the Courant number. Therefore the final amplitude of the numerical solution is equal to the initial amplitude multiplied by $|A_N|^M$, which is proportional to $A_G$ as given by Equation (A15).

 

*A2. Modified MUSCL scheme*

The reconstruction of the modified MUSCL scheme is also given by Equation (A1). The difference with the original MUSCL lies in the method used for the calculation of $U$ at the interfaces of the cell $i$. In the present approach, $U_{i+1/2}^{n+1/2}$ is computed by tracing the Riemann invariant $U$ backwards in time along the characteristic line $\mathrm{d}x/\mathrm{d}t = \lambda$:

$$\begin{aligned} U_{i,R}^{n+1} = U_{i+1,L}^{n+1} &= \tilde{U}_i^n(x_{i+1/2} - \lambda\Delta t) \\ &= \tilde{U}_i^n(x_{i+1/2} - Cr\Delta x) \end{aligned} \tag{A16}$$

Substituting Equation (A1) into Equation (A16) and noticing that $x_{i+1/2} = x_i + \Delta x/2$ yields

$$U_{i+1,L}^{n+1} = U_{i,R}^{n+1} = U_i^n + \left(\frac{1}{2} - Cr\right)\beta_i^n\Delta x \tag{A17}$$

The slope $\beta_i^{n+1}$ is computed as

$$\beta_i^{n+1} = \frac{U_{i,R}^{n+1} - U_{i,L}^{n+1}}{\Delta x} \tag{A18}$$

Substituting Equation (A17) into Equation (A18) leads to

$$\beta_i^{n+1} = \frac{U_i^n - U_{i-1}^n}{\Delta x} + \left(\frac{1}{2} - Cr\right)(\beta_i^n - \beta_{i-1}^n) \tag{A19}$$

The average $U_{i+1/2}^{n+1/2}$ at the interface $i + \frac{1}{2}$ is obtained by substituting Equation (A1) into Equation (A4)

$$U_{i+1/2}^{n+1/2} = U_i^n + (1 - Cr)\frac{\Delta x}{2}\beta_i^n \tag{A20}$$

Substituting Equation (A20) into Equations (A3) and (A7) gives the following expression

$$U_i^{n+1} = CrU_{i-1}^n + (1 - Cr)U_i^n + (1 - Cr)Cr\frac{\Delta x}{2}(\beta_{i-1}^n - \beta_i^n) \tag{A21}$$

Equations (A19) and (A21) can be rewritten in vector form as

$$\mathbf{V}_i^{n+1} = \mathbf{E}\mathbf{V}_i^n \tag{A22}$$

where

$$\mathbf{V}_i^n = \begin{bmatrix} U_i^n \\ \Delta x\beta_i^n \end{bmatrix}, \quad \mathbf{E} = \begin{bmatrix} Cr\delta^{-1} + 1 - Cr & (1 - Cr)\dfrac{Cr}{2}(\delta^{-1} - 1) \\ 1 - \delta^{-1} & \left(\dfrac{1}{2} - Cr\right)(1 - \delta^{-1}) \end{bmatrix} \tag{A23}$$

where $\delta$ is the forward shift operator defined as

$$\delta U_i^n = U_{i+1}^n \tag{A24}$$

 

The solution of Equation (A22) is sought in the form

$$\left.\begin{aligned} U_i^n &= U_0 \exp(\mu n \Delta t + \mathrm{i}\sigma \Delta x) \\ \beta_i^n &= \beta_0 \exp(\mu n \Delta t + \mathrm{i}\sigma \Delta x) \end{aligned}\right\} \tag{A25}$$

From (A24), the forward shift operator $\delta$ is given by

$$\delta = \exp(\sigma \Delta x) \tag{A26}$$

and $\mathbf{E}$ becomes

$$\mathbf{E} = \begin{bmatrix} Cr\exp(-\sigma\Delta x) + 1 - Cr & \dfrac{1}{2} - Cr \\ 1 - \exp(-\sigma\Delta x) & \left(\dfrac{1}{2} - Cr\right)[1 - \exp(-\sigma\Delta x)] \end{bmatrix} \tag{A27}$$

The eigenvalues of $\mathbf{E}$ are

$$\begin{aligned} \kappa^{(1)} &= \frac{e_{11} + e_{22}}{2} - \frac{1}{2}[(e_{11} + e_{22})^2 + 4(e_{21}e_{12} - e_{11}e_{22})]^{1/2} \\ \kappa^{(2)} &= \frac{e_{11} + e_{22}}{2} + \frac{1}{2}[(e_{11} + e_{22})^2 + 4(e_{21}e_{12} - e_{11}e_{22})]^{1/2} \end{aligned} \tag{A28}$$

where $e_{pq}$ denotes the element on the $p$th row and $q$th column of $\mathbf{E}$. These eigenvalues are amplification factors for the eigenvectors of the matrix $\mathbf{E}$. The variable vector $\mathbf{V}$ can be viewed as the sum of two independent signals propagating at different speeds. The amplification factor of the $p$th signal is $A_N^{(p)} = \kappa^{(p)}$ $(p = 1, 2)$ and its celerity factor is given by

$$C_N^{(p)} = -\frac{\mathrm{Im}(A_N^{(p)})}{\sigma \Delta x Cr}, \quad p = 1, 2 \tag{A29}$$

It can be noticed that the properties (A12) and (A13) also hold for both signals:

$$\left.\begin{aligned} A_N^{(p)}(1 - Cr, \sigma\Delta x) &= \exp(-\sigma\Delta x)\overline{A_N^{(p)}(Cr, \sigma\Delta x)} \\ |A_N^{(p)}(1 - Cr)| &= |A_N^{(p)}(Cr)| \end{aligned}\right\} \quad p = 1, 2 \tag{A30}$$

Consequently the minimum of the modulus each amplification factor is obtained for $Cr = \frac{1}{2}$ and the numerical solution is stable for Courant numbers between 0 and 1.

Figure A1 shows the variations of the amplification factor, the celerity factor and the global amplitude convergence factor (defined by Equation (A15)) as functions of the wave number $M = 2\pi/(\sigma\Delta x)$ for various values of the Courant number. These plots can be interpreted as follows:

- The celerity factor of the first signal in the compact MUSCL scheme is negative. It means that this signal travels in the opposite direction to that of the analytical solution.
- However, the amplification factor of this signal converges to a value smaller than unity as the wave number $M$ tends to infinity. This means that the signal is damped within a few time steps and does not induce a large error in the numerical solution.

 

- Both the amplitude and celerity factor of the second signal in the compact MUSCL scheme tend to unity as the wave number tends to infinity. This indicates convergence of the numerical solution to the analytical solution.
- The modulus of the amplification factor of the second signal is always larger than the modulus of the amplification factor of the original MUSCL scheme. Therefore, the numerical solution can be less affected by damping with the compact than with the classical MUSCL scheme.

It should be remembered however that these conclusions are derived for the unlimited versions of both the classical and compact MUSCL scheme (because such versions verify the assumption of linear schemes with constant coefficients, while the addition of limiters invalidates this assumption). Incorporating limiters to the schemes may cause the numerical solutions to behave in a slightly different way from what can be concluded from the amplitude and phase portraits.

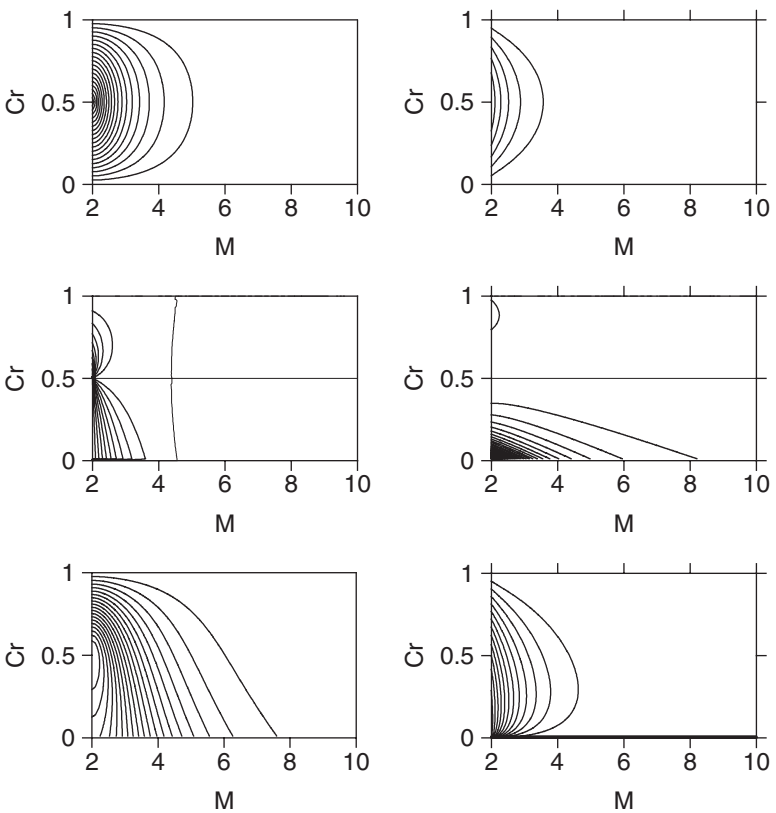


Figure A1. Amplification factor (top), celerity convergence factor (middle) and global amplification factor (bottom) of the classical (left) and compact (right) MUSCL schemes. The contour lines for the amplification factors start from 0.95 with a spacing of 0.05. The thick contour line on the celerity convergence factor corresponds to unity.

 

## APPENDIX B. STABILITY ANALYSIS OF THE CLASSICAL AND MODIFIED, EXPLICIT DISCRETIZATIONS OF THE DISPERSION EQUATION

### *B1. Stability analysis of the classical discretization*

The explicit version of the classical semi-discretization (48) is

$$U_i^{n+1} = U_i^n + (U_{i-2}^n - 3U_{i-1}^n + 3U_i^n - U_{i+1}^n)D \tag{B1}$$

where $D = k\Delta t/\Delta x^3$ is a dimensionless number playing the same role as the Courant number plays for the advection equation. Note that Equation (B1) can be applied indifferently to the cell averages $U_i$ and to the interface values $U_{i+1/2}$ by redefining the convection for indexing. Also note that $k$ is assumed to be negative. The solution of Equation (B1) is sought in the form given by Equation (A9). Substituting Equation (A9) into Equation (B1) yields

$$\begin{aligned} A_N &= 1 + [\exp(-2\mathrm{i}\sigma\Delta x) - 3\exp(-\mathrm{i}\sigma\Delta x) + 3 - \exp(\mathrm{i}\sigma\Delta x)]D \\ &= 1 + [\exp(-\mathrm{i}\sigma\Delta x) - 2 + \exp(\mathrm{i}\sigma\Delta x)][\exp(-\mathrm{i}\sigma\Delta x) - 1]D \end{aligned} \tag{B2}$$

where $A_N = U_i^{n+1}/U_i^n$ is again the amplification factor of the numerical solution. Denoting by $z$ the complex number

$$z = \exp(-\mathrm{i}\sigma\Delta x) - 1 \tag{B3}$$

Equation (B2) can be rewritten as

$$A_N = 1 + (z + \bar{z})zD \tag{B4}$$

where the overbar denotes the complex conjugate. $z$ is located on a circle of unit radius, the centre of which is $-1$ in the complex plane. Therefore, $A_N$ is located on a circle of radius $(z+\bar{z})D$ tangent to the point $+1$ in the complex plane (Figure B1). The centre of this circle is the point $1-(z+\bar{z})D$. The circle intersects the axis of real numbers at the point $1-2(z+\bar{z})D$. For the modulus of $A_N$ to be smaller than or equal to unity, the following condition must be

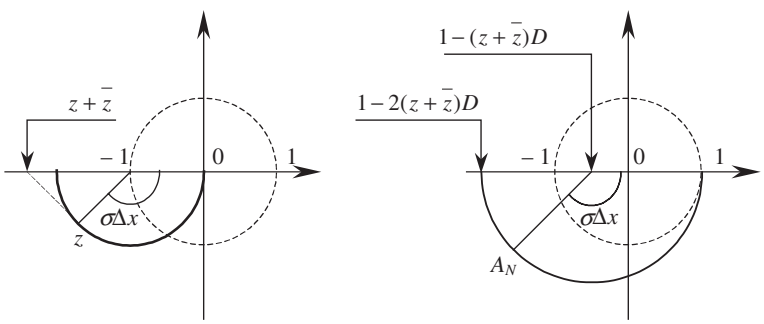


Figure B1. Graphical representation in the complex plane of the amplification factor of the explicit version of the semi-discretization (48). Location of $z$ (left) and $A_N$ (right). The unit circle is represented by a dashed line.

 

satisfied

$$-1 \leqslant 1 - 2(z + \bar{z})D \leqslant 1 \tag{B5}$$

The real part of $z$ is negative and varies between $-2$ and 0. For Equation (B5) to be satisfied, $D$ must satisfy the following conditions:

$$\begin{aligned} D &\leqslant 0 \\ D &\geqslant -\tfrac{1}{4} \end{aligned} \tag{B6}$$

The first condition implies that $k$ should be negative, which is true by assumption. The second condition leads to a stability constraint between $k$, $\Delta t$ and $\Delta x$.

*B2. Stability analysis of the modified, conservative discretization*

The present subsection deals with the stability of the explicit version of the modified, conservative discretization (59). Equation (59) can be rewritten in the vector form (A22) by defining $\mathbf{E}$ and $\mathbf{V}$ as follows:

$$\mathbf{E} = \begin{bmatrix} 1 + \dfrac{12D}{5}(1-\delta) & \dfrac{6D}{5}(\delta^{-2} - 4\delta^{-1} + 3) \\ \dfrac{3D}{2}(\delta^{-1} - 4 + 3\delta) & 1 + 3D(1-\delta) \end{bmatrix}, \quad \mathbf{V} = \begin{bmatrix} U_i^n \\ U_{i+1/2}^n \end{bmatrix} \tag{B7}$$

$\mathbf{E}$ can be rewritten as

$$\mathbf{E} = \begin{bmatrix} 1 + \dfrac{12D}{5}(1-\delta) & \dfrac{6D}{5}(\delta^{-2} - 3\delta^{-1})(1-\delta) \\ \dfrac{3D}{2}(\delta^{-1} - 3)(1-\delta) & 1 + 3D(1-\delta) \end{bmatrix} \tag{B8}$$

where $\delta$ is the forward shift operator. The eigenvalues $\lambda^{(p)}$ $(p = 1, 2)$ of $\mathbf{E}$ are given by

$$\left.\begin{aligned} \lambda^{(1)} &= a - (a^2 - b)^{1/2} \\ \lambda^{(2)} &= a + (a^2 - b)^{1/2} \end{aligned}\right\} \tag{B9}$$

where $a$ and $b$ are defined as

$$\begin{aligned} a &= 1 + \frac{27D}{10}(1-\delta) \\ b &= \frac{\delta}{5}[3(\delta^{-2} - 3\delta^{-1})(1-\delta)D]^2 \end{aligned} \tag{B10}$$

The discretization is stable if the modulus of both $\lambda^{(1)}$ and $\lambda^{(2)}$ is smaller than unity. Owing to the complexity of the expressions (B10), relating this condition to a necessary and sufficient condition on $D$ is very difficult. This has not been done so far. However, a sufficient condition can be derived. Observing that instability first appears for the smaller wave numbers (i.e. for

 

$\delta = \exp(-i\pi) = -1$), the discretization is stable if the modulus of $A_N$ is stable for $\delta = -1$. From the expressions of $a$ and $b$, this yields the following condition on $D$:

$$D \geqslant \frac{135 - 25\sqrt{259}}{5746} \approx -4.65 \times 10^{-2} \tag{B11}$$

which is five to six times as small as the permissible range for $D$ in the case of the classical discretization.

### *B3. Stability analysis of the non-conservative discretization*

A sufficient condition for the stability of the non-conservative discretization (63) is now derived. Equation (63) is rewritten in the vector form (A22) by defining **E** and **V** as

$$\mathbf{E} = \begin{bmatrix} 1 + 6(\delta^{-1} + 5)D & -12(1 + 2\delta)D \\ -6(5\delta^{-1} + 1)D & 1 + 12(\delta^{-1} + 2)D \end{bmatrix}, \quad \mathbf{V} = \begin{bmatrix} U_i^n \\ U_{i+1/2}^n \end{bmatrix} \tag{B12}$$

The eigenvalues of **E** are given by Equation (B9), where $a$ and $b$ are defined as

$$\begin{aligned} a &= 1 + 9(\delta^{-1} + 3)D \\ b &= 72(5\delta^{-1} + 1)(1 + 2\delta)D^2 \end{aligned} \tag{B13}$$

Following the same reasoning as in the previous subsection, the following sufficient condition is obtained for $D$

$$D \geqslant \frac{3 - \sqrt{33}}{48} \approx -5.71 \times 10^{-2} \tag{B14}$$

## REFERENCES


1. Van Leer B. Toward the ultimate conservative difference scheme, IV: a new approach to numerical convection. *Journal of Computational Physics* 1977; **23**:276–299.
2. Ben-Artzi M, Falcovitz J. A second-order Godunov-type scheme for compressible fluid dynamics. *Journal of Computational Physics* 1984; **55**:1–32.
3. Colella P, Woodward PR. The piecewise parabolic method (PPM) for gas-dynamical simulations. *Journal of Computational Physics* 1984; **54**:174–201.
4. Billet G, Louedin O. Adaptive limiters for improving the accuracy of the MUSCL approach for unsteady flows. *Journal of Computational Physics* 2001; **170**:161–183.
5. Hubbard ME. Multidimensional slope limiters for MUSCL-type finite volume schemes on unstructured grids. *Journal of Computational Physics* 1999; **155**:54–74.
6. Abarbanel S, Kumar A. Compact high-order schemes for the Euler equations. *SIAM Journal on Scientific and Statistical Computing* 1988; **3**:275.
7. Lele SK. Compact finite difference schemes with spectral-like resolution. *Journal of Computational Physics* 1992; **103**:16–42.
8. Yee HC. Explicit and implicit multidimensional compact high-resolution shock-capturing methods: formulation. *Journal of Computational Physics* 1997; **131**:216–232.
9. Lerat A, Corre P. Residual-based compact schemes for multidimensional hyperbolic systems of conservation laws. *Computers & Fluids* 2002; **31**:639–661.
10. Holly Jr FM, Preissmann A. Accurate calculation of two-dimensional advection. *Journal of the Hydraulics Division* (*ASCE*) 1997; **98**:1259–1277.
11. Chang S-C. The method of space-time conservation element and solution element—a new approach for solving the Navier-Stokes and Euler equations. *Journal of Computational Physics* 1995; **119**:295–324.
12. Toda K, Holly Jr FM. Hybrid numerical method for linear advection diffusion. *Microsoftware for Engineers* 1987; **3**(4):199–205.


 

13. Toda K, Holly Jr FM. Hybrid numerical method for nonlinear advection diffusion. *Journal of Hydroscience and Hydraulic Engineering* 1998; **6**(1):1–11.
14. Komatsu T, Ohgushi K, Asai K, Holly Jr FM. Accurate simulation of scalar advective transport. *Journal of Hydroscience and Hydraulic Engineering* 1989; **7**(1):63–73.
15. Chang S-C, Wang XY, Chow CY. The space-time conservation element and solution element method: a new high-resolution and genuinely multidimensional paradigm for solving conservation laws. *Journal of Computational Physics* 1999; **156**:89–136.
16. Molls T, Moll F. Space-time conservation method applied to Saint-Venant equations. *Journal of Hydraulic Engineering* (*ASCE*) 1999; **124**(5):891.
17. Chang S-C, Wang X-Y, To W-M. Application of the space-time conservation element and solution element method to one-dimensional convection–diffusion problems. *Journal of Computational Physics* 1999; **165**: 189–215.
18. Lax PD. Hyperbolic systems of conservation laws, II. *Communications in Pure and Applied Mathematics* 1957; **10**:537–566.
19. LeVeque RJ. *Numerical Methods for Conservation Laws*. Birkhaüser: Basel, 1992.
20. Colella P. A direct Eulerian MUSCL method for gas dynamics. *SIAM Journal on Scientific and Statistical Computing* 1985; **6**:104–117.
21. Guinot V. The Discontinuous Profile Method (DPM) for simulating two-phase flow in pipes using the single-component approximation. *International Journal for Numerical Methods in Fluids* 2001; **37**:341–359.
22. Guinot V. *Godunov-Type Schemes*: *an Introduction for Engineers*. Elsevier: Amsterdam, 2003.
23. Toro EF. *Shock-Capturing Methods for Free-Surface Shallow Flows*. Wiley: New York, 2001.
24. Godunov SK. A difference method for numerical calculation of discontinuous equations of hydrodynamics. *Matematicheski Sbornik* 1959; **47**:271–300.
25. Guinot V. Riemann solvers and boundary conditions for two-dimensional shallow water simulations. *International Journal for Numerical Methods in Fluids* 2003; **41**:1191–1219.
26. Guinot V. Boundary condition treatment for $2 \times 2$ systems of propagation equations. *International Journal for Numerical methods in Engineering* 1988; **42**:647–666.

 