Acta Polytechnica **61**(SI):49–58, 2021
DOI:10.14311/AP.2021.61.0049

available online at https://ojs.cvut.cz/ojs/index.php/ap

MODIFIED EQUATION FOR A CLASS OF EXPLICIT AND IMPLICIT SCHEMES SOLVING ONE-DIMENSIONAL ADVECTION PROBLEM

Tomáš Bodnár[a,b,*], Philippe Fraunié[c], Karel Kozel[a]

[a] *Czech Technical University in Prague, Faculty of Mechanical Engineering, Karlovo Náměstí 13, 121 35 Prague, Czech Republic*

[b] *Czech Academy of Sciences, Institute of Mathematics, Žitná 25, 115 67 Prague, Czech Republic*

[c] *Université de Toulon, Mediterranean Institute of Oceanography - MIO, BP 20132 F-83957 La Garde cedex, France*

[*] corresponding author: Tomas.Bodnar@fs.cvut.cz

Abstract. This paper presents the general modified equation for a family of finite-difference schemes solving one-dimensional advection equation. The whole family of explicit and implicit schemes working at two time-levels and having three point spatial support is considered. Some of the classical schemes (upwind, Lax-Friedrichs, Lax-Wendroff) are discussed as examples, showing the possible implications arising from the modified equation to the properties of the considered numerical methods.

Keywords: Modified equation, finite difference, advection equation.

1. Introduction

Numerical solution of differential equations became a standard tool in many disciplines of theoretical science as well as in applied sciences and engineering. Numerical and computational methods brought the possibility to solve non-trivial problems described by ordinary and partial differential equations for which it was impossible to obtain an analytical solution by standard mathematical methods. A wide range of numerical methods was developed during the years for specific problems.

Typically the physical problem is first described mathematically, so the mathematical model is created. Then this mathematical model is solved numerically. The obtained numerical solution is an approximation to the solution of the mathematical model, which itself is just an approximation of the physical problem. We keep aside the error introduced by the inaccuracies and approximations made when developing the mathematical model from the physical problem. Here the focus is on the difference between the discrete numerical solution of the mathematical model and its exact (analytical) solution. The numerical solution strongly depends on the method used to obtain it. So the properties and quality of the numerical solution may differ from the solution of the original mathematical model.

The aim of this paper is to show that the numerical solution of certain problems (equations) is much closer to the solution of some modified equation, rather than to the solution of the original problem. Many important properties of the numerical solution can than be easily seen (and a-priori expected) from the behavior of the known solution of the modified problem.

This rather general principle will be demonstrated on a finite-difference approximation of the solution of one-dimensional advection equation.

The structure of the paper is as follows. First the advection, diffusion and dispersion equations are introduced and discussed. Then several explicit schemes for advection equation are presented and analyzed at the discrete level. Modified equation is first derived for upwind scheme and then extended for a general class of explicit and implicit schemes. Finally the modified equations are discussed from the point of view of numerical diffusion and dispersion.

2. Model problems

In order to be able to explain the properties of modified equations and the underlying numerical schemes three model problems are introduced. They describe the physical phenomena of advection, diffusion and dispersion. All these problems can be mathematically formulated using a linear evolutionary partial differential equation. In all cases the unknown function $u(x,t)$ is the sought subject to the initial data $u(x,t=0)=\eta(x)$. For a sketch of an example of initial data see Fig. 1.

The initial value problem can be thus solved analytically using the Fourier transform method. The Fourier transform (in space) needed to obtain the analytical solutions is defined by:

$$\hat{u}(\xi,t)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}u(x,t)e^{-\boldsymbol{i}\xi x}dx\ ,$$

while the inverse is

$$u(x,t)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}\hat{u}(\xi,t)e^{\boldsymbol{i}\xi x}d\xi\ .$$

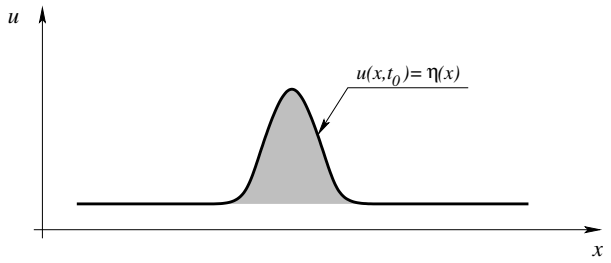


FIGURE 1. Initial condition.

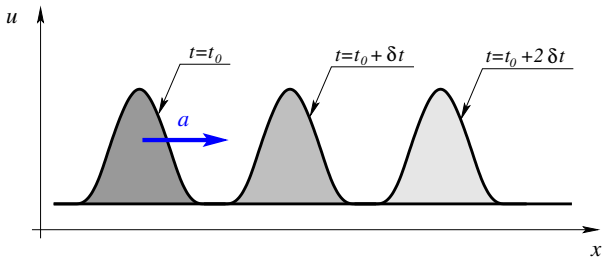


FIGURE 2. Advection equation solution evolution.

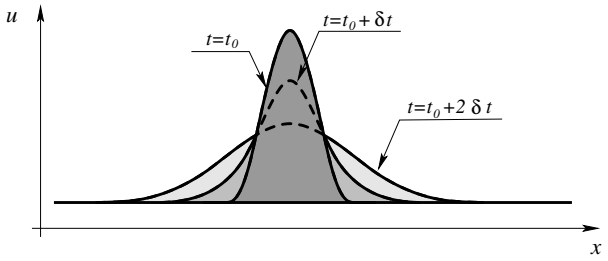


FIGURE 3. Diffusion equation solution evolution.

Using this transformation, the analytical solutions of model linear problems (discussed further) can be derived. The details can be found for example in the appendix of the book [1].

2.1. ADVECTION EQUATION

The advection problem can be described by a first order PDE of the form:

$$u_t + a u_x = 0$$

In this paper we consider the advection velocity $a > 0$, but for $a \leq 0$ the solution can be obtained as well. This advection equation is supplemented by the initial data

$$u(x, t = 0) = \eta(x) \ .$$

Applying the Fourier transform we can get the expression for the Fourier image of the solution

$$\hat{u}(\xi, t) = e^{-\boldsymbol{i}\xi a t} \hat{\eta}(\xi) \ .$$

Using the inverse transform, the solution can be written in the form

$$u(x,t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\boldsymbol{i}\xi a t} \hat{\eta}(\xi) e^{\boldsymbol{i}\xi x} d\xi$$

and finally

$$u(x,t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \hat{\eta}(\xi) e^{\boldsymbol{i}\xi(x-at)} d\xi$$

It's not difficult to see that the solution corresponds in fact to the initial data $\eta(x)$ shifted along the x-axis at velocity a, i.e.

$$u(x,t) = \eta(x - at) \ .$$

An illustration of such a situation can be seen in Fig. 2.

When the advection equation is solved numerically the discrete solution differs from the exact one, depending on the numerical method used. The numerical solution often exhibits some non-physical oscillations (due to dispersion) or smeared solution gradients (due to diffusion). This is why the diffusion and dispersion equations are briefly presented hereafter.

2.2. DIFFUSION EQUATION

The diffusion equation contains second spatial derivative, multiplied by a diffusion coefficient b.

$$u_t + b u_{xx} = 0$$

Also this linear PDE can easily be solved analytically using the Fourier transform to obtain

$$\hat{u}(\xi, t) = e^{-\boldsymbol{i}^2 \xi^2 b t} \hat{\eta}(\xi)$$

and after the inverse transform the solution

$$u(x,t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{\xi^2 b t} \hat{\eta}(\xi) e^{\boldsymbol{i}\xi x} d\xi \ .$$

A straightforward comparison with the solution of the advection equation reveals that now the individual Fourier modes found in the initial data $\eta(x)$ do not change their position. They only change their amplitude by the factor $e^{\xi^2 b t}$, which means the exponential decay for $b < 0$ and growth for $b > 0$. This is why only the decaying case with $b < 0$ is usually physically acceptable leading to an asymptotically stable evolution in time. It should also be noted that the decay depends on the square of wave number ξ and thus the rapidly oscillating modes decay much faster. An illustration of the diffusion process and diffusion equation solution is shown in Fig. 3.

2.3. DISPERSION EQUATION

The linear dispersion equation can be written as

$$u_t + c u_{xxx} = 0$$

where c is the dispersion coefficient. The Fourier image of the solution is

$$\hat{u}(\xi, t) = e^{-\boldsymbol{i}^3 \xi^3 c t} \hat{\eta}(\xi)$$

from which follows the solution

$$u(x,t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{i\xi^3 ct} \hat{\eta}(\xi) e^{i\xi x} d\xi$$

that can be rearranged to

$$u(x,t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \hat{\eta}(\xi) e^{i\xi(x+c\,\xi^2 t)} d\xi$$

This form is very similar to the solution of the advection equation

$$u(x,t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \hat{\eta}(\xi) e^{i\xi(x-at)} d\xi \ .$$

Comparing the corresponding expressions in the exponential, it can be seen that during the dispersive evolution the Fourier modes are also shifted along the x-axis, but each mode is shifted at a different velocity depending on the wave number as $-c\,\xi^2$. The minus sign indicates that for $c > 0$ the modes propagate against the sense of the x-axis. In general it means that the initial data are decomposed into individual modes and each of them propagates at a different speed, proportional to ξ^2.

2.4. Advection-Diffusion-Dispersion equation

The above described model problems involving the advection, diffusion and dispersion can be combined into advection-diffusion or advection-dispersion equations. The solutions of these combined equations can also be derived analytically. They will share the properties and behavior of both original equations solutions. These combinations will be important later in this paper while discussing the properties of the modified equations, where often the diffusive and dispersive term appear due to the discretization of advection equation. See the discussions in the sections 4.1–5.

3. Discrete analysis

The problem of numerical solution of advection equation is one of the classical topics in numerical mathematics. There exists a wide range of numerical schemes to discretize this problem. Here we only show a few classical methods as an illustration.

Further we will consider advection in one spatial dimension

$$u_t + a u_x = 0 \qquad a > 0 \ ,$$

where $u(x,t)$ is the sought solution and $u_i^n \approx u(x_i, t_n)$ is its discrete numerical approximation at point $x_i = i\Delta x$ and time instant $t_n = n\Delta t$, i, n being integer indices in spatial and temporal coordinates respectively.

The numerical approximation of the solution can be constructed e.g. using the finite-difference discretization, replacing the derivatives in the original equation by the corresponding divided differences formulas developed from the Taylor expansions. The final formulas for few classical schemes are shown in the Table 1. Some details concerning the process of derivation of some of these schemes will be shown in the following sections. More details can be found e.g. in the classical textbooks [2], [3] or [4].

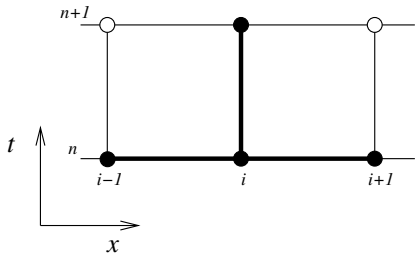


FIGURE 4. Initial condition.

3.1. Up-wind & Down-wind decomposition

A family of simple explicit schemes working on a three-point spatial stencil and two time levels (see Fig. 4) can be derived using forward and backward difference approximations of the spatial derivative. When a general (convex) combination of forward and backward differences is used, with weighting factor α, the explicit finite difference approximation will take the form

$$\frac{u_i^{n+1} - u_i^n}{\Delta t} + a \left[\alpha \left(\frac{u_{i+1}^n - u_i^n}{\Delta x} \right) + \right.$$
$$\left. + (1-\alpha) \left(\frac{u_i^n - u_{i-1}^n}{\Delta x} \right) \right] = 0$$

This family of schemes can be formally written using the forward (down-wind) difference $\{\mathbf{D}\}$ and backward (up-wind) difference $\{\mathbf{U}\}$ as

$$\frac{u_i^{n+1} - u_i^n}{\Delta t} + a\big[\alpha\{\mathbf{D}\} + (1-\alpha)\{\mathbf{U}\}\big] = 0 \ .$$

It's evident, that the classical upwind and downwind schemes can be recovered for special choices of $\alpha = 0$ and $\alpha = 1$ respectively. The simple central scheme can be obtained for a symmetric choice represented by $\alpha = 1/2$. A little bit less apparent, but still easy to verify is the representation of Lax-Friedrichs and Lax-Wendroff schemes, that also belong to this family. Values of the upwind/downwind blending coefficient α for all these schemes are summarized in the Tab. 2.

3.2. Central & Upwind decomposition

In the same way as every explicit scheme with a three-point stencil can be written in the form of convex combination of forward and backward differences, it can also be formally rewritten as a weighted sum of central and upwind difference approximations. The central approximation $\{\mathbf{C}\}$ can simply be expressed

Scheme		Formula
Up-wind	**[U]**	$u_i^{n+1} = u_i^n - \frac{a\Delta t}{\Delta x}(u_i^n - u_{i-1}^n)$
Down-wind	**[D]**	$u_i^{n+1} = u_i^n - \frac{a\Delta t}{\Delta x}(u_{i+1}^n - u_i^n)$
Central	**[C]**	$u_i^{n+1} = u_i^n - \frac{a\Delta t}{2\Delta x}(u_{i+1}^n - u_{i-1}^n)$
Lax-Friedrichs	**[LF]**	$u_i^{n+1} = \frac{1}{2}(u_{i+1}^n + u_{i-1}^n) - \frac{a\Delta t}{2\Delta x}(u_{i+1}^n - u_{i-1}^n)$
Lax-Wendroff	**[LW]**	$u_i^{n+1} = u_i^n - \frac{a\Delta t}{2\Delta x}(u_{i+1}^n - u_{i-1}^n) + \frac{a^2\Delta t^2}{2\Delta x^2}(u_{i+1}^n - 2u_i^n + u_{i-1}^n)$

TABLE 1. Examples of some simple classical discretizations for advection equation.

Scheme	Coefficient α
[U]	0
[D]	1
[C]	$\frac{1}{2}$
[LF]	$\frac{1}{2} - \frac{\Delta x}{2a\Delta t}$
[LW]	$\frac{1}{2} - \frac{a\Delta t}{2\Delta x}$

TABLE 2. Coefficient of up-wind/down-wind decomposition of classical schemes.

Scheme	Coefficient α	ϵ	μ
[U]	0	1	$\frac{a\Delta x}{2}$
[D]	1	-1	$-\frac{a\Delta x}{2}$
[C]	$\frac{1}{2}$	0	0
[LF]	$\frac{1}{2} - \frac{\Delta x}{2a\Delta t}$	$\frac{1}{\gamma}$	$\frac{\Delta x^2}{2\Delta t}$
[LW]	$\frac{1}{2} - \frac{a\Delta t}{2\Delta x}$	γ	$\frac{a^2\Delta t}{2}$

TABLE 3. Numerical viscosity coefficients for some classical schemes.

as an average of upwind and downwind (backward and forward) differences as

$$\{\mathbf{C}\} = \frac{\{\mathbf{D}\} + \{\mathbf{U}\}}{2} \quad \Longrightarrow \quad \{\mathbf{D}\} = 2\{\mathbf{C}\} - \{\mathbf{U}\} \ .$$

The family of explicit schemes can thus be rewritten using the central difference approximation $\{\mathbf{C}\}$ and additional upwinding $\{\mathbf{U}\}$, where the blending parameter now has the value 2α.

$$\frac{u_i^{n+1} - u_i^n}{\Delta t} + a\left[2\alpha\{\mathbf{C}\} + (1-2\alpha)\{\mathbf{U}\}\right] = 0$$

3.3. CENTRAL & VISCOUS DECOMPOSITION

By taking a divided difference of forward and backward differences (approximations of u_x), the central approximation of the second derivative u_{xx} can be constructed. This is often used in approximation of diffusive (or viscous terms) containing second derivatives. This viscous-like term $\{\mathbf{V}\}$ can be formally written as

$$\{\mathbf{V}\} = \frac{\{\mathbf{D}\} - \{\mathbf{U}\}}{\Delta x} \quad \Longrightarrow \quad \{\mathbf{U}\} = \{\mathbf{C}\} - \frac{\Delta x}{2}\{\mathbf{V}\} \ .$$

Using this definition of $\{\mathbf{V}\}$, the whole family of schemes can be written as a combination of the central approximation part and a viscous (diffusive) part.

$$\frac{u_i^{n+1} - u_i^n}{\Delta t} + a\left[\{\mathbf{C}\} - (1-2\alpha)\frac{a\Delta x}{2}\{\mathbf{V}\}\right] = 0$$

3.3.1. NUMERICAL VISCOSITY

By a simple rearrangement of this formula, it's easy to see that all the schemes from the considered explicit family can be written in the form, where only the central approximation of the first derivative is kept on left, while the remaining viscous-like part is moved to the right-hand side

$$\frac{u_i^{n+1} - u_i^n}{\Delta t} + a\{\mathbf{C}\} = (1-2\alpha)\frac{a\Delta x}{2}\{\mathbf{V}\} \ ,$$

or shortly

$$\frac{u_i^{n+1} - u_i^n}{\Delta t} + a\{\mathbf{C}\} = \mu\{\mathbf{V}\} \ .$$

It's not difficult to see that this discrete equation corresponds to a finite-difference approximation of the advection-diffusion equation rather than just to the original advection equation.

$$u_t + au_x = 0 \qquad \longrightarrow \qquad u_t + au_x = \mu\, u_{xx}$$

The extra term on the right hand side corresponds to the numerical diffusion (viscosity), where the coefficient μ depends on the method being used.

$$\mu = \underbrace{(1-2\alpha)}_{\epsilon}\frac{a\Delta x}{2}$$

Values of this numerical viscosity coefficient for few classical explicit methods are listed in the table 3.

In this context, each scheme within this explicit family can be considered as a simple central scheme with different amounts of added numerical viscosity (or upwinding, alternatively). When taking into account the definition of the non-dimensional Courant-Friedrichs-Lewy parameter $\gamma = \frac{a\Delta t}{\Delta x}$, which is positive (and bounded by stability conditions), the non-dimensional viscosity coefficient ϵ can be defined.

The value of the numerical viscosity coefficient determines the essential behavior and properties of the specific numerical method. Just by changing the parameter μ, the schemes can become more diffusive or dispersive, stable or unstable and also their order of accuracy will change. These properties for some schemes are summarized in the table 4.

The schemes in the table 4 are sorted from top to bottom according to increasing numerical viscosity. It can be seen (as expected) that when the numerical viscosity coefficient is negative, the scheme is unstable. By increasing its value, the scheme becomes stable and also can improve its accuracy. However by further increase of the numerical viscosity, the behavior of the method becomes more diffusive and the formal order of accuracy drops down again.

In summary, the identification of the amount of numerical viscosity embedded in a numerical scheme is the key point in understanding its behavior. Here it was done at the discrete level, by identifying the discrete diffusive term in the scheme. Similarly even more detailed analysis can be performed at the continuous level, leading to so called *modified equation*.

4. Modified equation

The modified equation approach is well known, often used to assess the order of accuracy for finite-difference schemes. When doing the Taylor expansions to develop the finite-difference approximations, it is possible to go beyond just finding the order of the leading term in the truncated Taylor series. It's possible to find analytically the form of the leading term, add it to (keep it in) the original equation and study the behavior of the modified equation that includes this extra term introduced by the discretization of the problem. The properties of the modified problem solution will be close to the properties of the numerical solution of the original problem.

4.1. Modified equation for upwind scheme

When solving the the advection equation

$$u_t + a u_x = 0 \qquad a > 0$$

the Up-wind scheme can be written as

$$\frac{u_i^{n+1} - u_i^n}{\Delta t} + a \left[\frac{u_i^n - u_{i-1}^n}{\Delta x} \right] = 0 \ ,$$

using the explicit Euler discretization in time and backward difference in space (i.e. upwind when the advection velocity $a > 0$).

Considering a sufficiently smooth interpolant $u(x_i, t_n) = u_i^n$, the Taylor expansions can be used to derive the corresponding approximation formulas.

$$\frac{u(x_i,t_{n+1}) - u(x_i,t_n)}{\Delta t} + a \left[\frac{u(x_i,t_n) - u(x_{i-1},t_n)}{\Delta x} \right] = 0$$

The terms $u(x_i, t_{n+1})$ and $u(x_i, t_{n+1})$ appearing in this formula are then obtained from (truncated) the Taylor series

$$u(x_i,t_{n+1}) = u(x_i,t_n) + \Delta t u_t(x_i,t_n) + \\ + \frac{\Delta t^2}{2} u_{tt}(x_i,t_n) + \frac{\Delta t^3}{6} u_{ttt}(x_i,t_n) + \mathcal{O}(\Delta t^4)$$

$$u(x_{i-1},t_n) = u(x_i,t_n) - \Delta x u_x(x_i,t_n) + \\ + \frac{\Delta x^2}{2} u_{xx}(x_i,t_n) - \frac{\Delta x^3}{6} u_{xxx}(x_i,t_n) + \mathcal{O}(\Delta x^4) \ .$$

Using these values, the corresponding difference approximations can be obtained for temporal and spatial derivatives.

$$\frac{u(x_i,t_{n+1}) - u(x_i,t_n)}{\Delta t} = u_t(x_i,t_n) + \\ + \frac{\Delta t}{2} u_{tt}(x_i,t_n) + \frac{\Delta t^2}{6} u_{ttt}(x_i,t_n) + \mathcal{O}(\Delta t^3)$$

$$\frac{u(x_i,t_n) - u(x_{i-1},t_n)}{\Delta x} = u_x(x_i,t_n) - \\ - \frac{\Delta x}{2} u_{xx}(x_i,t_n) + \frac{\Delta x^2}{6} u_{xxx}(x_i,t_n) + \mathcal{O}(\Delta x^3)$$

When these difference approximations are put together as in the numerical scheme, the original advection equation appears on the right hand side, together with some extra terms, representing the leading order terms in the remainder of the truncated Taylor series.

$$\frac{u(x_i,t_{n+1}) - u(x_i,t_n)}{\Delta t} + a \left[\frac{u(x_i,t_n) - u(x_{i-1},t_n)}{\Delta x} \right] = \\ = u_t(x_i,t_n) + a u_x(x_i,t_n) + \ldots\ldots\ldots = 0$$

It means that although the difference scheme approximates the advection equation, some extra (higher order) terms appear on the right hand side due to discretization.

$$u_t(x_i,t_n) + a u_x(x_i,t_n) = \\ - \frac{\Delta t}{2} u_{tt}(x_i,t_n) - \frac{\Delta t^2}{6} u_{ttt}(x_i,t_n) + \mathcal{O}(\Delta t^3) + \\ + \frac{a \Delta x}{2} u_{xx}(x_i,t_n) - \frac{a \Delta x^2}{6} u_{xxx}(x_i,t_n) + \mathcal{O}(\Delta x^3)$$

Partial derivatives with respect to time and also space, both appear on the right hand side. The derivatives with respect to time can be converted to spatial derivatives using the original advection equation (or corresponding scheme Taylor's expansions):

$$u_t + a u_x = 0 \quad \Longrightarrow \quad \frac{\partial \cdot}{\partial t} = -a \frac{\partial \cdot}{\partial x}$$

$$u_{tt} = a^2 u_{xx} \qquad \& \qquad u_{ttt} = -a^3 u_{xxx}$$

This leads to

$$u_t(x_i,t_n) + a u_x(x_i,t_n) = \\ - \frac{a^2 \Delta t}{2} u_{xx}(x_i,t_n) + \frac{a^3 \Delta t^2}{6} u_{xxx}(x_i,t_n) + \mathcal{O}(\Delta t^3) + \\ + \frac{a \Delta x}{2} u_{xx}(x_i,t_n) - \frac{a \Delta x^2}{6} u_{xxx}(x_i,t_n) + \mathcal{O}(\Delta x^3) \ .$$

Scheme	ϵ	μ	Accuracy	Stability	Behavior
[D]	-1	$-\frac{a\Delta x}{2} < 0$	$(\Delta t)^1/(\Delta x)^1$	Unstable	Dispersive
[C]	0	0	$(\Delta t)^1/(\Delta x)^2$	Unstable	Dispersive
[LW]	$\gamma < 1$	$\frac{a^2\Delta t}{2}$	$(\Delta t)^2/(\Delta x)^2$	Stable	Dispersive
[U]	1	$\frac{a\Delta x}{2}$	$(\Delta t)^1/(\Delta x)^1$	Stable	Diffusive
[LF]	$\frac{1}{\gamma} > 1$	$\frac{\Delta x^2}{2\Delta t}$	$(\Delta t)^1/(\Delta x)^1$	Stable	Diffusive

TABLE 4. Properties and behavior of selected explicit numerical schemes.

When only the leading order term is kept, the modified equation takes the form:

$$u_t(x_i,t_n) + a u_x(x_i,t_n) = \\ = \left(\frac{a\Delta x}{2} - \frac{a^2\Delta t}{2}\right) u_{xx}(x_i,t_n) + \mathcal{O}(\Delta t^2;\Delta x^2)\ .$$

In the approximate version, the extra term containing second (spatial) derivative u_{xx} appears on the right hand side.

$$u_t(x_i,t_n) + a u_x(x_i,t_n) \doteq \frac{a\Delta x}{2}(1-\gamma) u_{xx}(x_i,t_n)$$

It means that while solving the advection equation by the Up-wind scheme, we obtain rather the solution of the advection-diffusion equation with the diffusion coefficient corresponding to the numerical viscosity μ

$$u_t + a u_x = 0 \ \longrightarrow\ u_t + a u_x = \underbrace{\frac{a\Delta x}{2}(1-\gamma)}_{\mu} u_{xx}$$

In more detail, instead of the first order approximation of the advection equation we have obtained a second order approximation of the advection-diffusion equation.

1^{st} order approximation of original equation

$$u_t + a u_x = 0 + \mathcal{O}(\Delta t;\Delta x)$$

2^{nd} order approximation of *modified equation*

$$u_t + a u_x = \frac{a\Delta x}{2}(1-\gamma) u_{xx} + \mathcal{O}(\Delta t^2;\Delta x^2)$$

So the numerical solution of the original (advection) problem is in fact much closer to the solution of the modified (advection-diffusion) problem, with all the consequences it may have on the solution behavior.

From the form of the modified equation we can see the order of accuracy of the approximation of the original problem, the diffusive (or possibly dispersive) character of the modified equation, representing the behavior of the numerical solution. It's also possible to see e.g. how the numerical viscosity coefficient μ scales with time-step Δt and spatial step Δx. From the requirement of positivity of the diffusion coefficient $\mu > 0$ it's possible to estimate the stability of the underlying numerical scheme, i.e. the limitations for the CFL parameter γ (evidently $\gamma < 1$ is required for the Up-wind scheme).

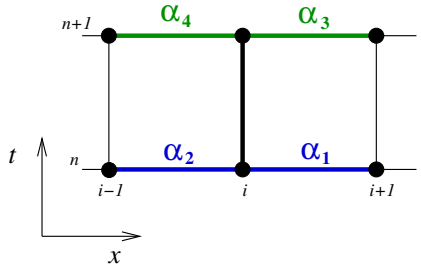


FIGURE 5. Computational stencil for general implicit and explicit schemes.

4.2. GENERAL MODIFIED EQUATION

The procedure of deriving the modified equation can be applied to all explicit schemes discussed so far. In fact it can be applied to a much larger family including also implicit schemes. Further we will work with a family of explicit and implicit schemes, where the approximation of the spatial derivative is obtained as a linear combination of the forward and backward differences at time levels n and $n+1$. The whole family of such schemes can be written in the form:

$$\frac{u_i^{n+1} - u_i^n}{\Delta t} + \\ + a\left[\alpha_1\left(\frac{u_{i+1}^n - u_i^n}{\Delta x}\right) + \alpha_2\left(\frac{u_i^n - u_{i-1}^n}{\Delta x}\right)\right] + \\ + a\left[\alpha_3\left(\frac{u_{i+1}^{n+1} - u_i^{n+1}}{\Delta x}\right) + \alpha_4\left(\frac{u_i^{n+1} - u_{i-1}^{n+1}}{\Delta x}\right)\right] = 0\ .$$

The computational stencil for such family of schemes is shown in the Fig. 5, introducing also the weights α_1 – α_4 used for blending the individual differences in the linear combination. Due to consistency the condition $\alpha_1 + \alpha_2 + \alpha_3 + \alpha_4 = 1$ should be verified, moreover obviously $|\alpha_3| + |\alpha_4| = 0$ for explicit schemes, while the opposite $|\alpha_3| + |\alpha_4| \neq 0$ characterizes the implicit schemes. For simplicity, the coefficients of the explicit part are marked in blue, while the implicit part is green.

The coefficients α_1 – α_4 for some examples of classical explicit and implicit schemes are shown in the Tab. 5. The last two rows in the table correspond to the Wendroff scheme denoted by [W] and Crank-Nicolson scheme denoted by [CN]. These two schemes

Scheme	α_1	α_2	α_3	α_4
[U]	0	1	0	0
[D]	1	0	0	0
[C]	$\frac{1}{2}$	$\frac{1}{2}$	0	0
[LF]	$\frac{1}{2}-\frac{\Delta x}{2a\Delta t}$	$\frac{1}{2}+\frac{\Delta x}{2a\Delta t}$	0	0
[LW]	$\frac{1}{2}-\frac{a\Delta t}{2\Delta x}$	$\frac{1}{2}+\frac{a\Delta t}{2\Delta x}$	0	0
[W]	$\frac{1}{2}$	0	0	$\frac{1}{2}$
[CN]	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$

TABLE 5. Coefficients of some explicit and implicit schemes with three-point stencil.

Scheme	Modified equation
[U]	$u_t + au_x = (1-\gamma)\frac{a\Delta x}{2}u_{xx}$
[D]	$u_t + au_x = -(1+\gamma)\frac{a\Delta x}{2}u_{xx}$
[C]	$u_t + au_x = -\gamma\frac{a\Delta x}{2}u_{xx}$
[LF]	$u_t + au_x = (\frac{1}{\gamma}-\gamma)\frac{a\Delta x}{2}u_{xx}$
[LW]	$u_t + au_x = -(1-\gamma^2)\frac{a\Delta x^2}{6}u_{xxx}$
[W]	$u_t + au_x = -(2+3\gamma+\gamma^2)\frac{a\Delta x^2}{12}u_{xxx}$
[CN]	$u_t + au_x = -(2+\gamma^2)\frac{a\Delta x^2}{12}u_{xxx}$

TABLE 6. Modified equations for some explicit and implicit schemes.

are examples of implicit schemes, using the difference approximations at both time levels n and $n+1$. For each of these schemes it's possible to derive the modified equation similarly as in the case of the Up-wind scheme. These modified equations are listed in the Tab. 6

Using the same procedure a general modified equation (with up to 3^{rd} order terms) for the whole family of considered schemes can be obtained in the form:

$$u_t + au_x = \epsilon_2 u_{xx} + \epsilon_3 u_{xxx} \ .$$

This is formally an advection-diffusion-dispersion equation with the numerical diffusion coefficient ϵ_2 and dispersion coefficient ϵ_3. These coefficients can be derived analytically to the form of functions depending on the blending weights $\alpha_1 - \alpha_4$.

$$\epsilon_2 = -\frac{a\Delta x}{2}\Big\{(\alpha_1-\alpha_2)+(\alpha_3-\alpha_4)+ \\ +\gamma\left[(\alpha_1+\alpha_2)^2-(\alpha_3+\alpha_4)^2\right]\Big\}$$

$$\epsilon_3 = -\frac{a\Delta x^2}{6}\Big\{1+3\gamma\left[(\alpha_1^2-\alpha_2^2)-(\alpha_3^2-\alpha_4^2)\right]+ \\ +2\gamma^2\left[(\alpha_1+\alpha_2)^3+(\alpha_3+\alpha_4)^3\right]\Big\}$$

5. NUMERICAL DIFFUSION AND DISPERSION

The general modified equation for the whole explicit/implicit family of schemes can be rewritten in the form

$$u_t + au_x = \epsilon_2 u_{xx} + \epsilon_3 u_{xxx} = \\ = \tilde{\epsilon}_2\frac{a\Delta x}{2}u_{xx} + \tilde{\epsilon}_3\frac{a\Delta x^2}{6}u_{xxx}$$

where the non-dimensional coefficients $\tilde{\epsilon}_2$ and $\tilde{\epsilon}_3$ are functions of the CFL parameter γ. This dependence of numerical diffusion and dispersion coefficients is obvious from the Tab. 7.

As already noted in the case of the discrete analysis of classical schemes (end of section 3), the amount of added numerical viscosity (diffusion) is responsible for the essential properties and behavior of each scheme.

5.1. DIFFUSIVE SCHEMES

The schemes, for which the leading order term on the right-hand side of the modified equation is the *diffusive* term $\tilde{\epsilon}_2\frac{a\Delta x}{2}u_{xx}$ are of the *first order* of accuracy. If the sign of the numerical viscosity coefficient $\tilde{\epsilon}_2$ is positive, the scheme is stable and it's behavior is diffusive. The negative sign of $\tilde{\epsilon}_2$ indicates that the scheme is unstable. The typical behavior of first order schemes is shown in the Fig. 6 for Up-wind and in Fig. 7 for Lax-Friedrichs scheme. The same initial value problem is solved for piece-wise constant initial data. The analytical solution corresponds to the "shifted" original data. The discrete numerical solution obtained using each scheme is marked by circles in the corresponding plots.

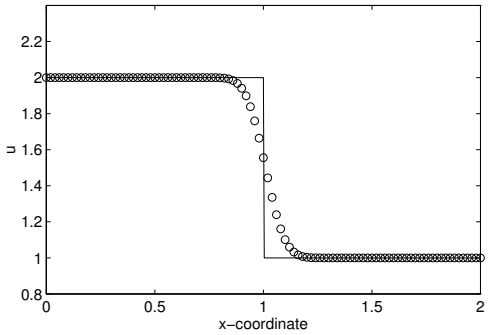
While the sign of $\tilde{\epsilon}_2$ determines the stability of the numerical method, the size of $\tilde{\epsilon}_2$ is responsible for the amount of added numerical diffusion. The higher the coefficient, the more diffused (smeared) the solution will be. By comparing the Lax-Friedrichs and the upwind schemes modified equations it is clear that for given CFL parameter γ the numerical viscosity coefficient $\tilde{\epsilon}_2 = \left(\frac{1}{\gamma}-\gamma\right)$ of the [LF] scheme is always greater than the $\tilde{\epsilon}_2 = (1-\gamma)$ of the [U] scheme. So the Lax-Friedrichs will be more diffusive than the Up-wind scheme. This is also well visible in the corresponding numerical solutions in the Figs. 6 and 7.

5.2. DISPERSIVE SCHEMES

For some schemes the first-order diffusive term vanishes and the dominant role in the modified equation is played by the dispersive term $\tilde{\epsilon}_3\frac{a\Delta x^2}{6}u_{xxx}$. Due to this, such schemes are of a *second order* and their behavior is *dispersive*. The numerical solution behaves much more like the solution of the advection-dispersion equation. It means that the advected initial data are decomposed into individual Fourier modes and each of them propagates at a different velocity, depending on the corresponding wave-number. This kind of behavior can be observed for Lax-Wendroff

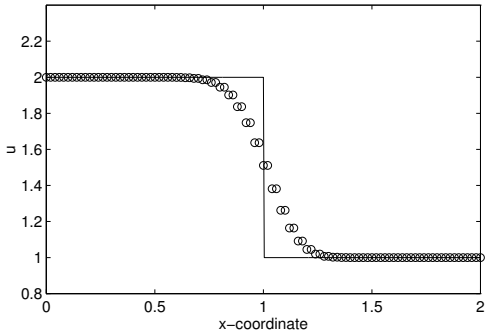
Scheme	$\tilde{\epsilon}_2$	$\tilde{\epsilon}_3$		Behavior
[U]	$1-\gamma$	$\cdots$	$\epsilon_2 > 0$	Diffusive
[D]	$-(1+\gamma)$	$\cdots$	$\epsilon_2 < 0$	Unstable
[C]	$-\gamma$	$\cdots$	$\epsilon_2 < 0$	Unstable
[LF]	$\frac{1}{\gamma}-\gamma$	$\cdots$	$\epsilon_2 > 0$	Diffusive
[LW]	0	$-(1-\gamma^2)$	$\epsilon_3 \neq 0$	Dispersive
[W]	0	$-\frac{1}{2}(2+3\gamma+\gamma^2)$	$\epsilon_3 \neq 0$	Dispersive
[CN]	0	$-\frac{1}{2}(2+\gamma^2)$	$\epsilon_3 \neq 0$	Dispersive

TABLE 7. Coefficients of numerical diffusion and dispersion for selected schemes.



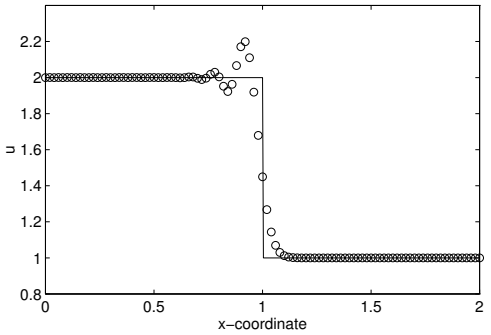
$$u_t + a u_x = (1-\gamma)\frac{a\Delta x}{2} u_{xx}$$

FIGURE 6. Up-wind [U] scheme solution and modified equation.



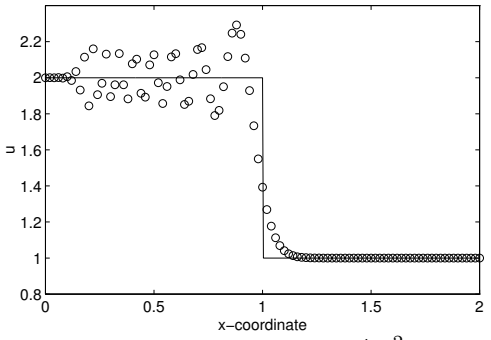
$$u_t + a u_x = \left(\frac{1}{\gamma}-\gamma\right)\frac{a\Delta x}{2} u_{xx}$$

FIGURE 7. Lax-Friedrichs [LF] scheme solution and modified equation.



$$u_t + a u_x = -(1-\gamma^2)\frac{a\Delta x^2}{6} u_{xxx}$$

FIGURE 8. Lax-Wendroff [LW] scheme solution and modified equation.



$$u_t + a u_x = -(2+3\gamma+\gamma^2)\frac{a\Delta x^2}{12} u_{xxx}$$

FIGURE 9. Wendroff [W] scheme solution and modified equation.

and Wendroff schemes in the Fig. 8 and 9 respectively. Again, for a given CFL parameter γ the numerical dispersion coefficient $\tilde{\epsilon}_3 = -(1-\gamma^2)$ of the [LW] scheme is smaller (in the absolute value) than the coefficient $\tilde{\epsilon}_3 = -(2+3\gamma+\gamma^2)/2$ for the [W] scheme. This results in a less dispersive (less oscillatory) solution obtained using the Lax-Wendroff scheme.

6. LIMITATIONS AND EXTENSIONS

In this paper the presentation was limited to the finite-difference approximation of linear advection, using schemes with three-point stencil operating at two time levels. Most of these limiting assumptions can be removed or relaxed. Larger stencil and more time levels can be used to construct the scheme, it will only make the analysis of the scheme and its modified equation derivation more difficult (time consuming), however it can easily be done. The only important limitation is that the schemes should be linear in the sense that their coefficients can't depend on the solution. The non-linear schemes case will be significantly more complicated, although some kind of local linearization (frozen coefficients) would probably give at least some basic information.

Multidimensional schemes can be analyzed in a very similar way, at least on regular, non-distorted grids.

For finite-volume schemes such analysis can be performed on arbitrary grids, based on the expression and splitting of the numerical flux as a sum of the simple average central flux and a dissipative stabilizing part [5, 6]. Based on the extra dissipation, the properties of the scheme can be shown.

7. Remark on applications

Based on a detailed knowledge of the diffusive/dispersive behavior of numerical schemes, suitable strategy can be chosen to improve their properties, namely the stability and resolution. The classical way is to modify the embedded numerical viscosity of the scheme. It can be done for example by the following approaches:

(1.) Modify the coefficients – As it was shown in section 3, each scheme can be written as a sum of the central (non-diffusive) part and additional internal dissipative part (or sum of down-wind and up-wind parts alternatively). The coefficient of internal numerical viscosity embedded in the scheme can be modified and adjusted by varying e.g. the blending parameter α and balancing the amount of forward and backward differences in the approximation. For modified Lax-Friedrichs scheme with reduced numerical viscosity see e.g. [7].

(2.) Build a composite scheme – Two schemes are used, one with higher accuracy (typically dispersive) and one with lower accuracy (typically diffusive). During the time-stepping process, several steps are performed using the higher order (but usually oscillatory) scheme, followed by a "smoothing" step performed using a lower order diffusive scheme. The ratio of high/low order steps can be tuned to optimize the performance of the combined composite method. An example of this technique using a combination of Lax-Wendroff and Lax-Friedrichs scheme can be found in [8].

(3.) Use artificial viscosity – The splitting of numerical methods into the central and viscous (diffusive) part offers the possibility to use always the (accurate but unstable) central scheme, followed by a separate stabilizing step applying an artificial viscosity term. In this way, the numerical viscosity is separated and no longer relies on the form embedded (hidden) in the numerical discretization. The separate, stand alone, numerical viscosity can be designed and tuned for the specific problem and best performance. Typically the second and fourth order damping terms were used proportional to second and fourth order derivatives $\epsilon_2 u_{xx}$ and $\epsilon_4 u_{xxxx}$. More efficient non-linear numerical viscosities can be constructed easily by considering variable, solution dependent, numerical viscosity coefficients $\epsilon_2(u)$, $\epsilon_4(u)$. When these coefficients are kept small on the smooth solution and only locally increased where the solution has high gradients or is oscillatory, a good resolution properties can be preserved while the stability and robustness of the method is improved. See e.g. [9] for artificial viscosity example and application or [10] or alternative filtering techniques.

8. Conclusions and Remarks

The discrete analysis based on a decomposition of the scheme into a central and diffusive (or upwind) part was shown for a family of explicit schemes. A general modified equation was developed for a large family of explicit and implicit schemes. This led to a discussion concerning the diffusive and dispersive properties of numerical schemes. Most of the partial conclusions were already formulated in the above sections. Let's just point out again some key points.

The numerical solution of the advection equation is "much closer" to the solution of advection-diffusion or advection-dispersion equation, depending which is the leading order term in the discretization error.

The behavior and quality of the numerical solution heavily depends on the coefficients of the modified equation. Their knowledge can help to assess a-priori the behavior of a numerical method.

The detailed knowledge of the structure of the leading order terms on the right-hand side of the modified equation can be used to construct the "high(er) resolution" numerical methods.

Acknowledgements

T. Bodnár is greatful for the support provided by the European Regional Development Fund-Project "Center for Advanced Applied Science" No.CZ.02.1.01/0.0/0.0/16 019/0000778 and partly by the Czech Science Foundation under the grant No. P201-19-04243S.

References

[1] R. J. LeVeque. *Finite difference methods for ordinary and partial differential equations : steady-state and time-dependent problems*. SIAM, 2007.

[2] C. Hirsch. *Numerical computation of internal and external flows*, vol. 1,2. John Willey & Sons, 1988.

[3] J. D. Anderson. *Computational Techniques for Fluid Dynamics*, vol. 1-2 of *Springer Series in Computational Physics*. Springer-Verlag Berlin Heidelberg, 2nd edn., 1991.

[4] C. A. J. Fletcher. *Computational Fluid Dynamics - The Basics with Applications*. McGraw-Hill, 1995.

[5] R. J. LeVeque. *Numerical Methods for Conservation Laws*. Lectures in Mathematics. Birkhäuser Verlag, 1990.

[6] R. J. LeVeque. *Finite Volume Methods for Hyperbolic Problems*. Cambridge Texts in Applied Mathematics. Cambridge University Press, 2002.

[7] R. Dvořák, K. Kozel. *Mathematical modeling in aerodynamics (in Czech)*. Vydavatelství ČVUT, 1996.

[8] R. Liska, B. Wendroff. Composite schemes for conservation laws. *SIAM Journal on Numerical Analysis* **35**(6):2250–2271, 1998. DOI:10.1137/S0036142996310976.

[9] T. Bodnár, L. Beneš, K. Kozel. Numerical simulation of flow over barriers in complex terrain. *Il Nuovo Cimento C* **31**(5–6):619–632, 2008. doi:10.1393/ncc/i2008-10323-4.

[10] A. Sequeira, T. Bodnár. On the filtering of spurious oscillations in the numerical simulations of convection dominated problems. *Vietnam Journal of Mathematics* **47**:851–864, 2019. doi:10.1007/s10013-019-00369-z.