CHAPTER 1

Introduction

Many interesting problems in the physical, biological, engineering and social sciences are modeled by a simple paradigm: Consider a domain $\Omega \subset \mathbb{R}^n$ and a quantity of interest $\mathbf{U}$, defined for all points $\mathbf{x} \in \Omega$. The quantity of interest $\mathbf{U}$ may be the temperature of a rod, the pressure of a fluid, the concentration of a chemical or a group of cells or the density of a human population. The evolution (in time) of this quantity of interest $\mathbf{U}$ can be described by a simple phenomenological observation:

The time rate of change of $\mathbf{U}$ *in any fixed sub-domain* $\omega \subset \Omega$ *is equal to the total amount of* $\mathbf{U}$ *produced or destroyed inside* ω *and the flux of* $\mathbf{U}$ *across the boundary* $\partial\omega$.

The above observation says that the change in $\mathbf{U}$ is due to two factors: the *source* or *sink*, representing the quantity produced or destroyed, and the *flux*, representing the amount of $\mathbf{U}$ that either goes in or comes out of the sub-domain, see Figure 1.1. This observation is mathematically rendered as

$$\frac{d}{dt}\int_\omega \mathbf{U}\, d\mathbf{x} = \underbrace{-\int_{\partial\omega} \mathbf{F}\cdot\nu\, d\sigma(\mathbf{x})}_{\text{flux}} + \underbrace{\int_\omega \mathbf{S}\, d\mathbf{x}}_{\text{source}}, \tag{1.1}$$

where ν is the unit outward normal, $d\sigma(\mathbf{x})$ is the surface measure, and $\mathbf{F}$ and $\mathbf{S}$ are the flux and the source respectively. The minus sign in front of the flux term is for convenience. Note that (1.1) is an *integral* equation for the evolution of the total amount of $\mathbf{U}$ in ω.

We simplify (1.1) by using integration by parts (or the Gauss divergence theorem) on the surface integral to obtain

$$\frac{d}{dt}\int_\omega \mathbf{U}\, d\mathbf{x} + \int_\omega \operatorname{div}(\mathbf{F})\, d\mathbf{x} = \int_\omega \mathbf{S}\, d\mathbf{x}. \tag{1.2}$$

Since (1.2) holds for all sub-domains ω of Ω, we can use an infinitesimal ω to obtain the following *differential* equation:

$$\mathbf{U}_t + \operatorname{div}(\mathbf{F}) = \mathbf{S} \qquad \forall\ (\mathbf{x}, t) \in (\Omega, \mathbb{R}_+). \tag{1.3}$$

The differential equation (1.3) is often termed as a *balance law* as it is a statement of the fact that the rate of change in $\mathbf{U}$ is a balance of the flux and the source. Frequently, the only change in $\mathbf{U}$ is from the fluxes and the source is set to zero. In such cases, (1.3) reduces to

$$\mathbf{U}_t + \operatorname{div}(\mathbf{F}) = 0 \qquad \forall\ (\mathbf{x}, t) \in (\Omega, \mathbb{R}_+). \tag{1.4}$$

Equation (1.4) is termed as a *conservation law*, as the only change in $\mathbf{U}$ comes from the quantity entering or leaving the domain of interest.

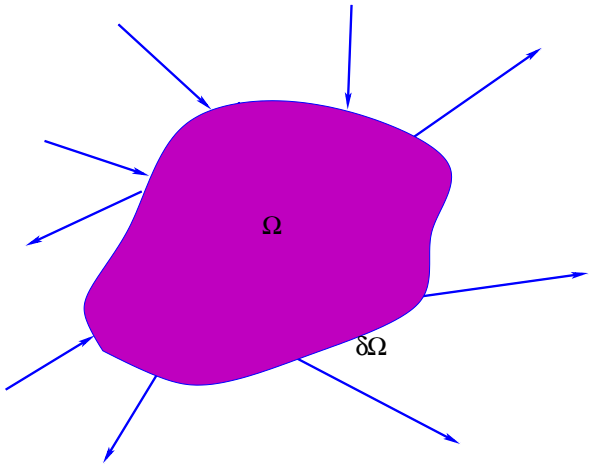


FIGURE 1.1. An illustration of conservation in a domain with the change being determined by the net flux.

The discussion so far is very general. We have not yet specified the explicit forms of $\mathbf{U}, \mathbf{F}$ and $\mathbf{S}$. In fact, the conservation law (1.4) and the balance law (1.3) are generic to a very large number of models. Explicit forms of the quantity of interest, flux and source depend on the specific model being considered. The modeling of the flux $\mathbf{F}$ is the core function of a physicist, biologist, engineer or other domain scientists. We will provides several examples to illustrate conservation laws.

1.1. Examples for conservation laws.

For simplicity of the exposition, we begin with scalar examples, i.e, the quantity of interest $\mathbf{U}$ is a scalar U.

Scalar transport equation. Let $\mathbf{U} = U$ denote the concentration of a chemical (for example, a pollutant in a river). Assume that the river flows with a velocity field $\mathbf{a}(\mathbf{x}, t)$ and we know the velocity field at all points in the river. The pollutant will clearly be transported in the direction of the velocity and so the flux in this case is $\mathbf{F} = \mathbf{a}U$. Since there is no production or destruction of the pollutant during the flow, the source term in (1.3) is set to zero. Consequently, the conservation law (1.4) takes the form

$$U_t + \operatorname{div}(\mathbf{a}(\mathbf{x}, \mathbf{t})U) = 0. \tag{1.5}$$

This equation is linear. In the simple case of one space dimension and a constant velocity field $\mathbf{a}(\mathbf{x}, t) \equiv a$, (1.5) reduces to

$$U_t + aU_x = 0. \tag{1.6}$$

The scalar one-dimensional equation (1.6) is often referred to as the transport or *advection* equation.

The heat equation. Another illustrative example of a conservation law is provided by heat conduction. Assume that a hot material (like a metal block) is heated at one end and is left to cool afterwards, without providing any additional source of heat. It is a common observation that the heat spreads or *diffuses* out

and the temperature of the material becomes uniform after some time. Let U be the temperature of the material. Diffusion of heat is governed by Fourier's or Fick's law

$$\mathbf{F}(U) = -\mathbf{k}\nabla U.$$

Here, $\mathbf{k}$ is the conductivity tensor for the medium. The minus sign is due to the fact that heat flows from hotter to cooler zones. Substituting Fourier's law into the conservation law (1.4), we obtain the *heat* equation

$$U_t - \operatorname{div}(\mathbf{k}\nabla U) = 0. \tag{1.7}$$

If the conductivity is assumed to be unity and the material is one-dimensional (like a rod), (1.7) reduces to the well-known one-dimensional heat equation

$$U_t - U_{xx} = 0. \tag{1.8}$$

The scalar transport equation (1.5) and the heat equation (1.7) are both linear equations and deal with the evolution of a single scalar quantity. As nature is too complicated to be described by scalar linear equations, their utility is limited. Next, we present a *nonlinear system* of conservation laws.

Euler equations of gas dynamics. A gas (as an example consider air) consists of a large number of molecules. The motion of each molecule can be tracked individually. This description is termed as the particle description and leads to a very large number of ODEs. The resulting system of ODEs is too large to be computationally feasible. Instead, a more *macroscopic description* is used. In a macroscopic model, the key variables of interest are: the density ρ, the velocity field $\mathbf{u}$ and the gas pressure p. All these quantities can be measured experimentally. The relevant conservation laws are

- *Conservation of mass:* It is well-known in fluid dynamics that the total mass of the gas is conserved. Mathematically, using Kelvin's theorem, this translates into

$$\rho_t + \operatorname{div}(\rho\mathbf{u}) = 0.$$

- *Conservation of momentum:* By Newton's second law of motion, the rate of change of momentum equals force. In the absence of external forces, the gas pressure is the only force acting on the gas. The resulting conservation law is

$$(\rho\mathbf{u})_t + \operatorname{div}(\rho\mathbf{u}\otimes\mathbf{u}) + \nabla p = 0.$$

  Note that the above conservation laws implies that the rate of change of the *advective* (material) derivative of the momentum equals the gradient of pressure. This is a consequence of the following observation: Gas flows from high to low pressure.

  The symbol $\otimes$ is the tensor product, i.e, for any two vectors $\mathbf{a} = (a_1, a_2, a_3)$ and $\mathbf{b} = (b_1, b_2, b_3)$, we have

$$\mathbf{a}\otimes\mathbf{b} = \begin{pmatrix} a_1b_1 & a_1b_2 & a_1b_3 \\ a_2b_1 & a_2b_2 & a_2b_3 \\ a_3b_1 & a_3b_2 & a_3b_3 \end{pmatrix}.$$

- *Conservation of energy:* The total energy of a gas is a sum of its kinetic and internal (potential) energy. The kinetic energy has the standard expression

$$E_k = \frac{1}{2}\rho|\mathbf{u}|^2,$$

whereas the internal energy is determined by an equation of state. If the gas is an ideal gas, then the equation of state is

$$E_i = \frac{p}{\gamma - 1},$$

where γ is the gas constant. It takes the values 5/3 and 7/5 for monoatomic and diatomic gases, respectively. Hence, the total energy of an ideal gas is

$$E = \frac{p}{\gamma - 1} + \frac{1}{2}\rho|\mathbf{u}|^2. \tag{1.9}$$

The rate of change of total energy is computed as:

$$E_t + \operatorname{div}((E + p)\mathbf{u}) = 0.$$

All the three conservation laws are combined together and written in divergence form to obtain the *Euler equations* of gas dynamics:

$$\begin{aligned} \rho_t + \operatorname{div}(\rho\mathbf{u}) &= 0, \\ (\rho\mathbf{u})_t + \operatorname{div}(\rho\mathbf{u} \otimes \mathbf{u} + p\mathbf{I}) &= 0, \\ E_t + \operatorname{div}((E + p)\mathbf{u}) &= 0, \end{aligned} \tag{1.10}$$

where $\mathbf{I}$ denotes the 3×3 identity matrix. The above system is an example of a multi-dimensional *nonlinear system* of conservation laws. This derivation of the Euler equations was very brief and details can be found in fluid dynamics textbooks like [**6**]. We ignore fluid viscosity effects and heat conduction in the gas while deriving (1.10).

The above examples already reveal a multitude of diverse physical phenomena that can be modeled in terms of conservation laws. The flux $\mathbf{F}$ in (1.4) is often a function of $\mathbf{U}$ and its derivatives,

$$\mathbf{F} = \mathbf{F}(\mathbf{U}, \nabla\mathbf{U}, \nabla^2\mathbf{U}, \dots)$$

For simplicity of the analysis, it is common to neglect the role of the higher than first-order derivatives. Hence, the flux is of the form:

$$\mathbf{F} = \mathbf{F}(\mathbf{U}, \nabla\mathbf{U}).$$

If it is of the form $\mathbf{F} = \mathbf{F}(\mathbf{U})$, then the conservation law (1.4) is a first-order PDE. It is usually classified as *hyperbolic*. The notion of hyperbolicity will be described in detail in the sequel. The scalar transport equation (1.5) and the Euler equations of gas dynamics (1.10) are examples for hyperbolic equations.

If we have $\mathbf{F} = \mathbf{F}(\nabla\mathbf{U})$, then the conservation law (1.4) is a second-order PDE and is often classified as *parabolic*. The heat equation (1.7) is an example of a parabolic equation. When the flux $\mathbf{F}$ depends on both the function $\mathbf{U}$ and its first derivative, the conservation law (1.4) is termed as a *convection-diffusion* equation. In these notes, we will consider hyperbolic equations and convection-diffusion equations with the convection dominating the diffusion.

1.1.1. Other examples. Examples for conservation laws of both the hyperbolic and convection-diffusion type abound in nature. In these notes, we will consider the scalar Burgers equation, the Buckley-Leverett equation (modeling flows in oil and gas reservoirs), the wave equation, the shallow water equations of meteorology and oceanography, the equations for linear and nonlinear elastic waves that arise in materials science and the equations of magnetohydrodynamics (MHD) from plasma physics.

1.2. Content and scope of these notes

The reason for studying conservation laws extensively is obvious: They arise in many models in the sciences, ranging from the design of aircraft (Euler equations) to the study of supernovas in astrophysics (MHD equations). Since interesting conservation laws like the Euler equations are nonlinear, it is not possible to obtain explicit solution formulas. Hence, numerical methods need to be developed for *approximating* or *simulating* the solutions of conservation laws. The design and implementation of efficient numerical methods is the main focus of these notes.

In order to design efficient numerical methods, we need to understand the analytical structure of the solutions of conservation laws. Therefore, we will briefly discuss theoretical properties of the solutions that are relevant for the design and analysis of numerical schemes.

We begin with the study of one-dimensional scalar problems. Both linear and nonlinear equations are considered, and efficient numerical schemes are described for them. Then, the focus shifts to linear and nonlinear systems like the Euler equations of gas dynamics. Finally, we consider the multi-dimensional versions of systems of conservation laws and describe efficient numerical schemes for them.

CHAPTER 2

Linear Transport Equations

In this chapter we consider the one-dimensional version of the linear transport equation,

$$U_t + a(x,t)U_x = 0 \qquad \forall\ (x,t) \in \mathbb{R} \times \mathbb{R}_+. \tag{2.1}$$

The simplest case of the scalar transport equation arises when the velocity field is constant, that is, $a(x,t) \equiv a$. The resulting transport equation is

$$U_t + aU_x = 0. \tag{2.2}$$

The rather simple equation (2.2) has served as a crucible for designing highly efficient schemes for much more complicated systems of equations. We concentrate on it for the rest of this chapter.

2.1. Method of characteristics

The initial value problem (or Cauchy problem) for (2.1) consists of finding a solution of (2.1) that also satisfies the initial condition

$$U(x,0) = U_0(x) \qquad \forall\ x \in \mathbb{R}. \tag{2.3}$$

It is well known that the solution of the initial value problem can be constructed by using the *method of characteristics*. The idea underlying this method is to reduce a PDE like (2.1) to an ODE by utilizing the structure of the solutions. As an ansatz, assume that we are given some curve $x(t)$, along which the solution U is constant. This means that

$$\begin{aligned} 0 &= \frac{d}{dt}U(x(t),t) && \text{(as U is constant along } x(t)) \\ &= U_t(x(t),t) + U_x(x(t),t)x'(t) && \text{(chain rule).} \end{aligned}$$

We also know that $U_t(x(t),t) + U_x(x(t),t)a(x(t),t) = 0$, since U is assumed to be a solution of (2.1). Therefore, if $x(t)$ satisfies the ODE

$$\begin{aligned} x'(t) &= a(x(t),t) \\ x(0) &= x_0, \end{aligned} \tag{2.4}$$

then $x(t)$ is precisely such a curve. The solution $x(t)$ of this equation is called a *characteristic curve*. From ODE theory, we know that solutions of (2.4) exist provided that a is Lipschitz continuous in both arguments. It may or may not be possible to find an explicit solution formula for (2.4).

The importance of characteristic curves lies in the property that U is constant along them:

$$U(x(t),t) = U(x(0),0) = U_0(x_0).$$

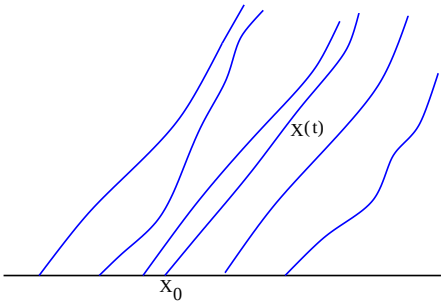


FIGURE 2.1. Characteristics curves $x(t)$ for (2.1)

The initial data $U_0(x)$ is already known, so if we can find characteristic curves that go through all points $(x,t) \in \mathbb{R} \times \mathbb{R}_+$, then we have found the solution U at all points in the plane. (See Figure 2.1) for an illustration.)

In the simple case of a constant velocity field $a(x,t) \equiv a$, the characteristic equation (2.4) is explicitly solved as

$$x(t) = x_0 + at.$$

Therefore, given some point (x,t), the unique characteristic that goes through (x,t) (so that $x(t) = x$) has initial value $x_0 = x - at$. Hence, the solution of (2.2) is

$$U(x,t) = U_0(x_0) = U_0(x - at) \tag{2.5}$$

for any $(x,t) \in \mathbb{R} \times R_+$. The solution formula (2.5) implies that the initial data is transported with the velocity a.

In the more general case of (2.1), the characteristic equation (2.4) may not be possible to solve explicitly. Hence, it is essential that we obtain some information about the structure of solutions of (2.1) from the equation itself. This is done by means of the following *a priori* energy estimate:

Lemma 2.1. *Let $U(x,t)$ be a smooth solution of (2.1) which decays to zero at infinity, i.e, $\lim_{|x|\to\infty} U(x,t) = 0$ for all $t \in \mathbb{R}_+$, and assume that $a \in C^1(\mathbb{R}, \mathbb{R}_+)$. Then U satisfies the energy bound*

$$\int_{\mathbb{R}} U^2(x,t)dx \leqslant e^{\|a\|_{C^1} t} \int_{\mathbb{R}} U_0^2(x)dx \tag{2.6}$$

for all times $t > 0$.

PROOF. The proof of the estimate (2.6) is based on multiplying (2.1) with U on both sides:

$$UU_t + a(x,t)UU_x = 0 \qquad \text{(multiplying (2.1) by } U\text{)}$$

$$\left(\frac{U^2}{2}\right)_t + a(x,t)\left(\frac{U^2}{2}\right)_x = 0 \qquad \text{(chain rule)}$$

$$\left(\frac{U^2}{2}\right)_t + \left(a(x,t)\frac{U^2}{2}\right)_x = a_x(x,t)\frac{U^2}{2} \qquad \text{(product rule)}$$

$$\frac{d}{dt}\int_{\mathbb{R}} \left(\frac{U^2}{2}\right) dx + \int_{\mathbb{R}} \left(a(x,t)\frac{U^2}{2}\right)_x dx = \int_{\mathbb{R}} a_x(x,t)\frac{U^2}{2} dx \qquad \text{(integrating over space)}$$

$$\frac{d}{dt}\int_{\mathbb{R}}\left(\frac{U^2}{2}\right)dx = \int_{\mathbb{R}} a_x(x,t)\,\frac{U^2}{2}dx \quad \text{(decay to zero at infinity)}$$
$$\leqslant \|a\|_{C^1}\int_{\mathbb{R}}\frac{U^2}{2}dx \quad \text{(regularity of } a\text{)}.$$

The last inequality can be used together with Gronwall's inequality (Theorem A.1) to obtain the bound (2.6). □

The quantity $\int U^2/2$ is commonly called the *energy* of the solution. The above lemma shows that the energy of the solutions to the transport equation (2.1) are bounded. The energy estimate is going to be used for designing robust schemes for the transport equation. We remark that the restriction that U decays to zero at infinity may be relaxed by considering a different energy functional.

2.2. Finite difference schemes for the transport equation

It may not be possible to obtain an explicit formula for the solution of the characteristic equation (2.4). For example, the velocity field $a(x,t)$ might have a complicated nonlinear expression. Hence, we have to devise numerical methods for approximating the solutions of (2.1). For simplicity, we consider $a(x,t) \equiv a > 0$ and solve (2.2). It is rather straightforward to extend the schemes to the case of a more general velocity field.

Discretization of the domain. The first step in any numerical method is to discretize both the spatial and temporal parts of the domain. Since $\mathbb{R}$ is unbounded, we have to truncate the domain to some bounded domain $[x_l, x_r]$. This truncation implies that suitable boundary conditions need to be imposed. We discuss the problem of boundary conditions later on.

For the sake of simplicity, the domain $[x_l, x_r]$ is discretized uniformly with a mesh size Δx into a sequence of $N+1$ points x_j such that $x_0 = x_l$, $x_N = x_r$ and $x_{j+1} - x_j = \Delta x$ for all j. A non-uniform discretization can readily be considered.

For the temporal discretization, we choose some terminal time T and divide $[0,T]$ into M points $t^n = n\Delta t$ $(n = 0,\dots,M)$. The space-time mesh is shown in Figure 2.2. Our aim is obtain an approximation of the form $U_j^n \approx U(x_j, t^n)$. To get from the initial time step t^0 to the terminal time step t^M, we first set the initial data $U_j^0 = U_0(x_0)$ for all j. Then the solution U_j^1 at the next time step is computed

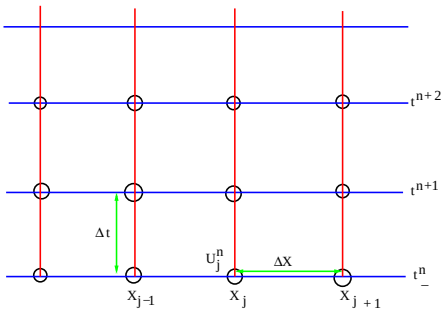


FIGURE 2.2. A representation of the mesh in space-time

using some update formula, again for all j. This process is reiterated until we arrive at the final time step $t^M = T$ with our final solution U_j^M.

A simple centered finite difference scheme. On the mesh, we need to approximate the transport equation (2.2). We do so by replacing both the spatial and temporal derivatives by finite differences. The time derivative is replaced with a forward difference and the spatial derivative with a central difference. This combination is standard (see schemes for the heat equation in standard textbooks like [**7**]). The resulting scheme is

$$\frac{U_j^{n+1} - U_j^n}{\Delta t} + \frac{a(U_{j+1}^n - U_{j-1}^n)}{2\Delta x} = 0 \qquad \text{for } j = 1, \ldots, N-1. \tag{2.7}$$

Some special care must be taken when defining the boundary values. We have a consistent discretization of (2.2) that is very simple to implement. We test it on the following numerical example.

A numerical example. Consider the linear transport equation (2.2) in the domain $[0, 1]$ with initial data

$$U_0(x) = \sin(2\pi x). \tag{2.8}$$

Since the data is periodic, it is natural to assume periodic boundary conditions. We implement this numerically by letting

$$U_0^n = U_{N-1}^n, \quad U_N^n = U_1^n.$$

The exact solution is calculated by (2.5) as $U(x, t) = \sin(2\pi(x - at))$. We set $a = 1$ and compute the solutions with the central scheme (2.7) with 500 mesh points, and plot the solution at time $t = 0.3$ in Figure 2.3. The figure clearly shows that, despite being a consistent approximation, the scheme is unstable, with very large oscillations.

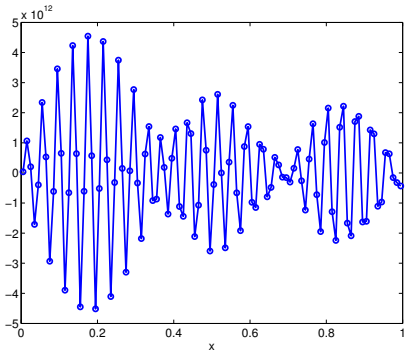


FIGURE 2.3. Approximate solution for (2.2) with the central scheme (2.7) at time $t = 3$ with 100 mesh points. `[central.m]`

A physical explanation. Why do the solutions computed with the central scheme (2.7) blow up? After all, the central scheme seems a reasonable approximation of the transport equation. A *physical* explanation can be deduced from the following argument: The exact solution moves to the right (as $a > 0$) with a fixed speed. Therefore, information goes from left to right. However, the central scheme (see Figure 2.4) takes information from both the left and the right, violating the physics. Consequently, the solutions are unstable. This explanation seems intuitive but has to be backed by solid mathematical arguments. We proceed to do so below.

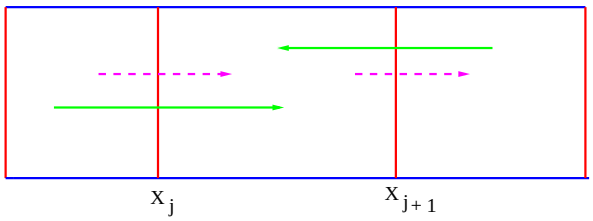


FIGURE 2.4. The central scheme (2.7). Green arrows indicate numerical propagation and magenta arrows physical propagation.

A mathematical explanation. The observed instability of the central scheme can be explained mathematically in terms of estimates. We recall that the exact solutions have a bounded energy (see estimate (2.6)). It is reasonable to require that the scheme is *energy stable* like the exact solution, that is, a discrete version of energy remains bounded. For a given Δx, we define the discrete version of energy as

$$E^n = \frac{1}{2}\Delta x \sum_j \left(U_j^n\right)^2. \tag{2.9}$$

Note that the integral in the energy for the continuous problem has been replaced with a Riemann sum.

Lemma 2.2. *Let U_j^n be the solutions computed with the central scheme* (2.7)*. Then the following estimate holds:*

$$E^{n+1} = E^n + \frac{\Delta x}{2} \sum_j \left(U_j^{n+1} - U_j^n\right)^2. \tag{2.10}$$

Consequently, the energy grows at every time step for any choice of $\Delta x, \Delta t$, and so the scheme is unconditionally unstable.

PROOF. We mimic the steps of continuous energy estimate (Lemma 2.1) and multiply both sides of the scheme (2.7) by U_j^n to obtain

$$U_j^n\left(U_j^{n+1} - U_j^n\right) + \frac{a\Delta t}{\Delta x}\left(U_j^n U_{j+1}^n - U_j^n U_{j-1}^n\right) = 0. \tag{2.11}$$

We have the following elementary identity:

$$d_2(d_1 - d_2) = \frac{(d_1)^2}{2} - \frac{(d_2)^2}{2} - \frac{1}{2}(d_1 - d_2)^2 \tag{2.12}$$

for any two numbers d_1, d_2. We denote

$$H_{j+1/2} = a\frac{U_j^n U_j^{n+1}}{2}$$

to reduce (2.11) to

$$\frac{\left(U_j^{n+1}\right)^2}{2} = \frac{(U_j^n)^2}{2} + \frac{1}{2}\left(U_j^{n+1} - U_j^n\right)^2 - \frac{\Delta t}{\Delta x}(H_{j+1/2} - H_{j-1/2}). \tag{2.13}$$

Summing (2.13) over all j and using zero (or periodic) boundary conditions, the flux term H vanishes by cancellation and we obtain the estimate (2.10). □

Although we assumed zero or periodic boundary conditions in the proof of this lemma, a variant of the lemma holds for more general boundary conditions, as for the continuous setting in Lemma 2.1.

The above lemma provides a mathematical justification for our physical intuition. The central scheme leads to a growth of energy at every time step and is unstable. We need to find schemes that posses a discrete version of the energy estimate. This use of rigorous mathematical tools like energy analysis to justify physical reasoning will be an essential ingredient of these notes.

2.3. An upwind scheme

The central scheme (2.7) does not respect the direction of propagation of information for the transport equation (2.2). Hence, we must include the correct direction of information propagation and hope that it stabilizes the scheme. This entails using one-sided differences instead of a central difference to approximate the linear transport equation (2.2).

If $a > 0$ and the direction of information propagation is from left to right, then we can use a backward difference in space to obtain the scheme

$$\frac{U_j^{n+1} - U_j^n}{\Delta t} + \frac{a(U_j^n - U_{j-1}^n)}{\Delta x} = 0 \qquad \text{for } j = 1, \ldots, N-1, \tag{2.14}$$

and if $a < 0$, we can use the forward difference to obtain:

$$\frac{U_j^{n+1} - U_j^n}{\Delta t} + \frac{a(U_{j+1}^n - U_j^n)}{\Delta x} = 0 \qquad \text{for } j = 1, \ldots, N-1. \tag{2.15}$$

Using the notation

$$a^+ = \max\{a, 0\}, \quad a^- = \min\{a, 0\}, \quad |a| = a^+ - a^-,$$

(2.14) and (2.15) can be written together as

$$\frac{U_j^{n+1} - U_j^n}{\Delta t} + \frac{a^+(U_j^n - U_{j-1}^n)}{\Delta x} + \frac{a^-(U_{j+1}^n - U_j^n)}{\Delta x} = 0. \tag{2.16}$$

The above scheme takes into account the direction of propagation of information – information is "carried with the wind". Hence, this scheme is termed as the *upwind* scheme.

Using the definition of the absolute value and some simple algebraic manipulations, the upwind scheme (2.16) can be recast as

$$\frac{U_j^{n+1} - U_j^n}{\Delta t} + \frac{a(U_{j+1}^n - U_{j-1}^n)}{2\Delta x} = \frac{|a|}{2\Delta x}(U_{j+1}^n - 2U_j^n + U_{j-1}^n) \tag{2.17}$$

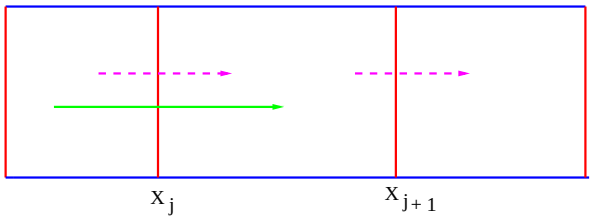
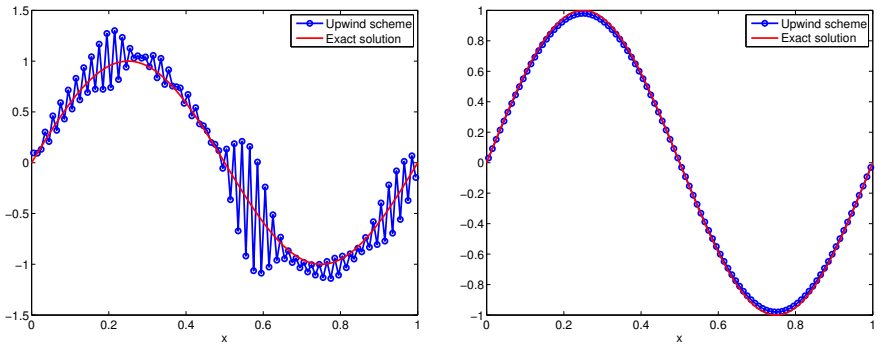


FIGURE 2.5. The upwind scheme (2.16). Green arrows indicate numerical propagation and magenta arrows physical propagation.

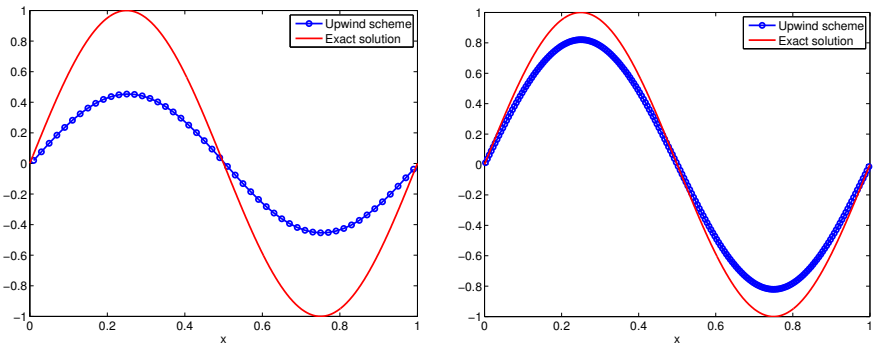
(compare to (2.7)). Note that in the above form, the spatial derivatives are the central term and a *diffusion* term. The right hand side of (2.17) approximates $\frac{\Delta x|a|}{2}U_{xx}$. Hence, the upwind scheme (2.17) adds *numerical viscosity* or *diffusion* to the unstable central scheme (2.7). Numerical viscosity is going to play a crucial role later on.

Since the upwind scheme incorporates the correct direction of propagation of information (see Figure 2.5), we expect it to be more stable than the central scheme. This is endorsed by the numerical experiment with initial data (2.8). We take $a = 1$ and compute approximate solutions for the linear transport equation (2.2) on a uniform mesh with 100 mesh points up to $t = 1$. We use two different timesteps: $\Delta t = 1.3\Delta x$ and $\Delta t = 0.9\Delta x$. As seen in Figure 2.6, the results with $\Delta t = 1.3\Delta x$ are still oscillatory and the scheme continues to be unstable. In spite of the upwinding, stability stills seems to be elusive. However, results with $\Delta t = 0.9\Delta x$ are stable. The approximation appears to be good in this case. Much better results are obtained by refining the mesh, while keeping the ratio $\Delta t/\Delta x$ fixed, as is presented in Figure 2.7.



(a) $\Delta t = 1.3\Delta x$ (b) $\Delta t = 0.9\Delta x$

FIGURE 2.6. Solution with initial data (2.8) at $t = 1$. The ratio $\Delta t/\Delta x$ is important for stability. `[upwind_cfl.m]`



(a) 50 mesh points (b) 200 mesh points

FIGURE 2.7. Solution with initial data (2.8) at $t = 10$. Refining the mesh gives a more accurate solution. `[upwind_refinement.m]`

2.4. Stability for the upwind scheme

The numerical results indicate that stability for the upwind scheme is subtle. It is not unconditionally unstable as the central scheme (2.7); instead, stability depends on the parameters $\Delta x, \Delta t$. Numerical results indicate the crucial role played by the ratio $\frac{\Delta t}{\Delta x}$. It seems that one must not only take into account the correct direction of propagation, but also the correct magnitude.

The quantification of stability will involve energy analysis as in the last section. We have the following stability result:

Lemma 2.3. *Let the mesh parameters satisfy the condition*

$$|a|\frac{\Delta t}{\Delta x} \leqslant 1. \tag{2.18}$$

Then solutions computed with the upwind scheme (2.17) *satisfy the energy estimate*

$$E^{n+1} \leqslant E^n, \tag{2.19}$$

where the energy is defined as in (2.9)*. The upwind scheme is thus* conditionally stable.

PROOF. For the sake of simplicity, we assume that $a > 0$. Hence the upwind scheme (2.17) reduces to

$$\frac{U_j^{n+1} - U_j^n}{\Delta t} + \frac{a(U_{j+1}^n - U_{j-1}^n)}{2\Delta x} = \frac{a}{2\Delta x}(U_{j+1}^n - 2U_j^n + U_{j-1}^n). \tag{2.20}$$

It is also equivalent to the scheme (2.14). As in the proof of the estimate (2.10) we multiply both sides of the scheme (2.20) by U_j^n to obtain

$$\begin{aligned} U_j^n(U_j^{n+1} - U_j^n) = &-\frac{a\Delta t}{2\Delta x}(U_j^n U_{j+1}^n - U_j^n U_{j-1}^n) \\ &+ \frac{a\Delta t}{2\Delta x}(U_j^n(U_{j+1}^n - U_j^n)) + \frac{a\Delta t}{2\Delta x}(U_j^n(U_{j-1}^n - U_j^n)). \end{aligned} \tag{2.21}$$

Now we use elementary identity (2.12) a couple of times and rewrite (2.21) as

$$\begin{aligned}\frac{(U_j^{n+1})^2}{2} &= \frac{(U_j^n)^2}{2} + \frac{(U_j^{n+1}-U_j^n)^2}{2} - \frac{a\Delta t}{2\Delta x}(U_j^n U_{j+1}^n - U_j^n U_{j-1}^n) \\ &\quad + \frac{a\Delta t}{4\Delta x}\left((U_{j+1}^n)^2 - (U_j^n)^2\right) - \frac{a\Delta t}{4\Delta x}\left((U_j^n)^2 - (U_{j-1}^n)^2\right) \\ &\quad - \frac{a\Delta t}{4\Delta x}(U_{j+1}^n - U_j^n)^2 - \frac{a\Delta t}{4\Delta x}(U_j^n - U_{j-1}^n)^2.\end{aligned} \tag{2.22}$$

Denoting

$$K_{j+1/2} = \frac{a}{2}(U_j^n U_{j+1}^n) - \frac{a}{4}\left((U_{j+1}^n)^2 - (U_j^n)^2\right),$$

we may rewrite (2.22) as

$$\begin{aligned}\frac{(U_j^{n+1})^2}{2} &= \frac{(U_j^n)^2}{2} + \frac{(U_j^{n+1}-U_j^n)^2}{2} - \frac{a\Delta t}{\Delta x}(K_{j+1/2} - K_{j-1/2}) \\ &\quad - \frac{a\Delta t}{4\Delta x}(U_{j+1}^n - U_j^n)^2 - \frac{a\Delta t}{4\Delta x}(U_j^n - U_{j-1}^n)^2.\end{aligned} \tag{2.23}$$

Summing (2.23) over all j and using the definition of discrete energy (2.9) and either zero or periodic boundary conditions, we obtain

$$E^{n+1} \leqslant E^n + \frac{\Delta x}{2}\sum_j (U_j^{n+1} - U_j^n)^2 - \frac{a\Delta t}{2}\sum_j (U_j^n - U_{j-1}^n)^2. \tag{2.24}$$

Using the definition of the upwind scheme (2.14) in (2.24) yields

$$E^{n+1} \leqslant E^n + \left(\frac{a^2\Delta t^2}{2\Delta x} - \frac{a\Delta t}{2}\right)\sum_j (U_j^n - U_{j-1}^n)^2. \tag{2.25}$$

Since the term in the sum in (2.25) is positive, we obtain the energy bound (2.19), provided

$$\frac{a^2\Delta t^2}{\Delta x} \leqslant a\Delta t,$$

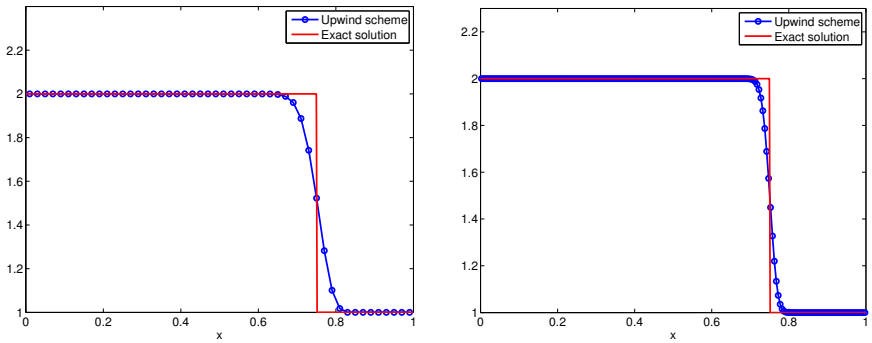
which is precisely the condition (2.18). □

The stability condition (2.18) is termed the *CFL condition* after Courant, Friedrichs and Lewy who first proposed it. The conditional stability of the upwind scheme is confirmed in numerical experiments.

Numerical experiment: Discontinuous data. Consider the transport equation (2.2) with $a = 1$ in the domain $[0, 1]$ and initial data

$$U_0(x) = \begin{cases} 2 & \text{if } x < 0.5 \\ 1 & \text{if } x > 0.5. \end{cases} \tag{2.26}$$

The initial data and consequently the exact solution (2.5) are discontinuous. We compute with the upwind scheme using 50 and 200 mesh points and display the results in Figure 2.8. The results show that the upwind scheme approximates the solution quite well, at least at a fine resolution. However the errors on a coarse mesh are somewhat large. This issue will be addressed in later sections.



(a) 50 mesh points (b) 200 mesh points

FIGURE 2.8. The upwind scheme (2.14) for advection (2.2) with discontinuous initial data (2.26). Both results are at time $t = 0.25$. `[upwind_disc_refinement.m]`

APPENDIX A

Results from real analysis

Theorem A.1 (Gronwall's inequality)**.** *Let $\beta(t)$ be continuous and $u(t)$ be differentiable on some interval $[a, b]$, and assume that*

$$u'(t) \leqslant \beta(t)u(t) \qquad \forall\ t \in (a, b).$$

Then

$$u(t) \leqslant u(a) \exp\left(\int_a^t \beta(t)\right) \qquad \forall\ t \in [a, b].$$

Bibliography

1. E. Godlewski and P.A. Raviart, *Hyperbolic systems of conservation laws*, Ellipses, 1991.
2. S. K. Godunov, *A difference scheme for numerical solution of discontinuous solution of hydrodynamic equations*, Math. Sbornik **47** (1959), 271–306.
3. Sigal Gottlieb, Chi-Wang Shu, and Eitan Tadmor, *Strong stability-preserving high-order time discretization methods*, SIAM Review **43** (2001), no. 1, 89–112.
4. Ernst Hairer, Syvert P. Nørsett, and Gerhard Wanner, *Solving ordinary differential equations*, 1987.
5. S N Kruzkov, *First order quasilinear equations in several independent variables*, Mathematics of the USSR-Sbornik **10** (1970), no. 2, 217–243.
6. L. D. Landau and E. M. Lifschitz, *Fluid mechanics*, Butterworth-Heinemann, 1987.
7. Aslak Tveito and Ragnar Winther, *Introduction to partial differential equations; a computational approach*, second ed., vol. 29, Springer-Verlag, 2009, 392 pages.