

Lecture Notes on Computational Fluid Dynamics

Instructor: Prof. Dongwook Lee (dlee79@ucsc.edu)
MWF, 2:00 pm – 3:10 pm at Kresge classroom 319
Winter, 2015

<https://courses.soe.ucsc.edu/courses/ams260/Winter15/01>
https://users.soe.ucsc.edu/~dongwook/?page_id=335

Chapter 1

Fundamentals of CFD

1. What is CFD? Why do we study CFD?

Let's begin our first class with a couple of interesting scenarios.

Scenario 1: See Fig. 1. Consider you're a chief scientist in a big aerospace research lab. You're given a mission to develop a new aerospace plane that can reach at hypersonic speed ($>$ Mach 5) within minutes after taking off. Its powerful supersonic combustion ramjets continue to propel the aircraft even faster to reach to a velocity near 26,000 ft/s (or 7.92 km/s, or Mach 25.4 in air at high altitudes, or a speed of NY to LA in 10 min), which is simply a low Earth orbital speed. This is the concept of transatmospheric vehicle the subject of study in several countries during the 1980s and 1990s. When designing such extreme hypersonic vehicles, it is very important to understand full three-dimensional flow field over the vehicle with great accuracy and reliability. Unfortunately, ground test facilities – wind tunnels – do not exist in all the flight regimes around such hypersonic flight. We neither have no wind tunnels that can simultaneously simulate the higher Mach numbers and high flow field temperatures to be encountered by transatmospheric vehicles.

Scenario 2: See Fig. 2. Consider you're a theoretical astrophysicist who tries to understand core collapse supernova explosions. The theory tells us that very massive stars can undergo core collapse when the core fail to sustain against its own gravity due to unstable behavior of nuclear fusion. We simply cannot find any ground facilities that allow us to conduct any laboratory experiments in such highly extreme energetic astrophysical conditions. It is also true that in many astrophysical circumstances, both temporal and spatial scales are too huge to be operated in laboratory environments.

Scenario 3: See Fig. 3. Consider you a golf ball manufacturer. Your goal is to understand flow behaviors over a flying golf ball in order to make a better golf ball design (and become a millionaire!) Although you've already collected a wide range of the laboratory experimental data on a set of golf ball shapes (i.e., surface dimple design), you realize that it is very hard to analyze the data and understand them because the data are all nonlinearly coupled and can't



Figure 1. DARPA's Falcon HTV-2 unmanned aircraft can max out at a speed of about 16,700 miles per hour – Mach 22, NY to LA in 12 minutes.

be isolated easily. To keep your study in a better organized way, you wish to perform a set of parameter studies by controlling flow properties one by one so that you can also make reliable flow prediction for a new golf ball design.

As briefly hinted above, in practice there are various levels of difficulties encountered in real experimental setups. When performing the above mentioned research work, CFD therefore can be the major player that leads you to success because you obtain mathematical controls in numerical simulations. Let us take an example how numerical experiment via CFD can elucidate physical aspects of a real flow field. Consider the subsonic compressible flow over an airfoil. We are interested in answering the differences between laminar and turbulent flow over the airfoil for $Re = 10^9$. For the computer program (assuming the computer algorithm is already well established, validated and verified!), this is a straightforward matter – it is just a problem of making one run with the turbulence model switched off (for the laminar setup), another run with the turbulence model switched on (for the turbulent flow), followed by a comparison study of the two simulation results. In this way one can mimic Mother Nature with simple knobs in the computer program – something you cannot achieve quite readily (if at all) in the wind tunnel. Without doubt, however, in order to achieve such success using CFD, you'd better to know what you do exactly when it comes to numerical modeling – the main goal of this course.

We are now ready to define what CFD is. CFD is a scientific tool, similar to experimental tools, used to gain greater physical insights into problems of interest. It is a study of the numerical solving of PDEs on a discretized system that, given the available computer resources, best approximates the real

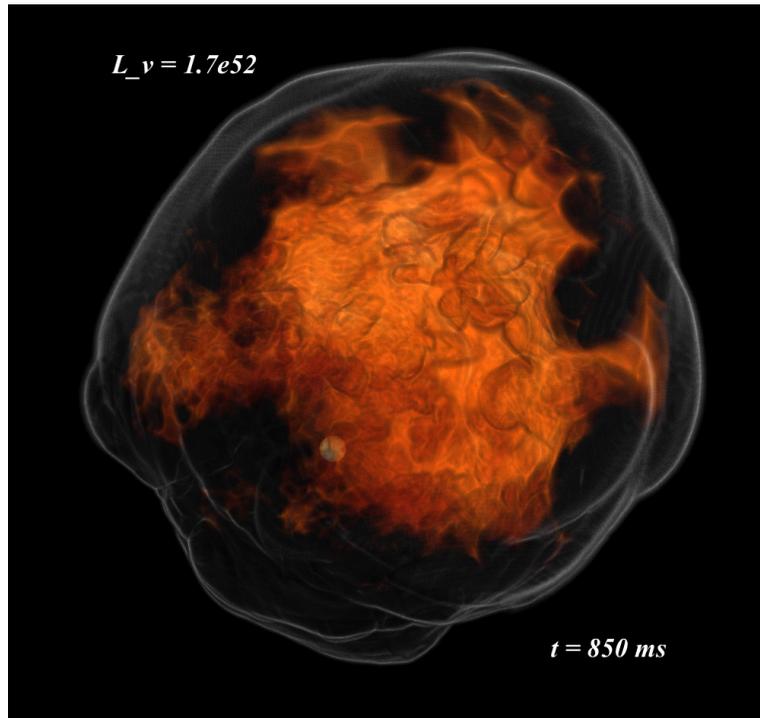


Figure 2. FLASH simulations of neutrino-driven core-collapse supernova explosions. Sean Couch (ApJ, 775, 35 (2013)).

geometry and fluid flow phenomena of interests. CFD constitutes a new “third approach” in studying and developing the whole discipline of fluid dynamics. A brief history on fluid dynamics says that the foundations for *experimental* fluid dynamics began in 17th century in England and France. In the 18th and 19th centuries in Europe, there was the gradual development of *theoretical* fluid dynamics. These two branches – experiment and theory – of fluid dynamics have been the mainstreams throughout most of the twentieth century. However, with the advent of the high speed computer with the development of solid numerical studies, solving physical models using computer simulations has revolutionized the way we study and practice fluid dynamics today – the approach of CFD. As sketched in Fig. 4, CFD plays a truly important role in modern physics as an equal partner with theory and experiment, in that it helps bringing deeper physical insights in theory as well as help better designing experimental setups.

The real-world applications of CFD are to those problems that do *not* have known analytical solutions; rather, CFD is a scientific vehicle for solving flow problems that cannot be solved in any other way. In this reason – the fact that we use CFD to tackle to solve those *unknown* systems – we are strongly encouraged to learn thorough aspects in *all* three essential areas of study: (i) numerical theories, (ii) fluid dynamics, and (iii) computer programing skills.

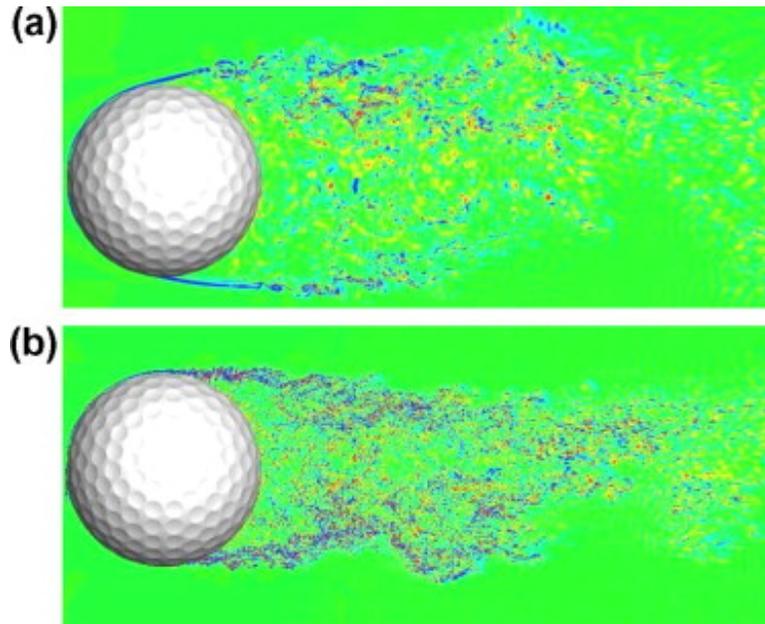


Figure 3. Contours of azimuthal velocity over a golf ball: (a) $Re = 2.5 \times 10^4$; (b) $Re = 1.1 \times 10^5$. C. E. Smith et al. (Int. J. Heat and Fluid Flow, 31, 262-273 (2010)).

2. About Homework Problems, Mid-term Exam, and Final-Term Project

Something needs to be said about what is expected in homework problems and one final computer project in this course. Most importantly, as early as possible, you need to choose your preferred programming languages and visualization tools in order to demonstrate your academic progress throughout the coursework.

2.1. Scientific languages

Fortran, C, C++, etc. (compiled languages); python, java, ruby, idl, matlab, GNU Octave, etc. (interpreted languages)

2.2. Visualization tools

gnuplot, idl, matlab, python, matplotlib (matplotlib.org), yt (yt-project.org), VisIt (<https://wci.llnl.gov/simulation/computer-codes/visit/>), techplot (www.techplot.com), etc.

For beginners, matlab can be a good tool to start with for both programming (relatively slower in performance but easier in programming) and visualization (convenient!) purposes. For skillful students, you're very welcome to use a compiled language (faster in performance and a better educational investment in scientific programming). If you haven't done any scientific computing yet, please consult with me by the end of the first week, Jan 9, 2015.

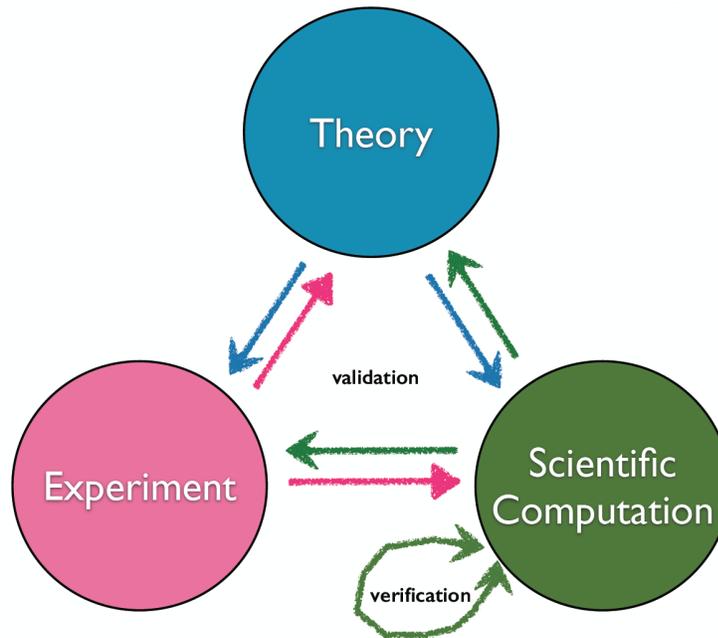


Figure 4. Three healthy cyclic relationship in fluid dynamics.

2.3. Homework Submission

There are total of 4 homework problem sets on both mathematical theories and computer programming in every two weeks. They take 30% toward your total grade. The purpose of the assignments is to provide you with opportunities in exploring mathematical concepts and using them to conduct numerical calculations. In this course you will learn extensively how to discretize mathematical equations and visualize numerical solutions to them. There is a policy on any late homework submission that you are going to receive a maximum of 80% if late by less than a day; 50% if late by more than a day. Students are strongly encouraged to submit their homework electronically in pdf (no word documents).

2.4. Mid-term – Feb 18, 2015

One in-class exam will be counted 30% toward your total grade.

2.5. Final-term computer programming project – Mar 20, 2015

A final project will be written up in a professional style using either latex or any word documents and submitted as a pdf file. It is expected that the quality of this project, of the students choosing in consultation with the instructor), will go past the material required for the other computer assignments. At the end of the quarter, each student will present his/her results to the class (15 min oral presentation), and the rest of your classmates will participate in evaluating your presentation. The project will take 40% of your total grade.

3. Course materials

3.1. Main resources:

- (1) Class notes and handouts,
- (2) Numerical methods for conservation laws – R. J. LeVeque (Birkhäuser)

3.2. Other textbook materials:

- (3) Finite volume methods for hyperbolic problems – R. J. LeVeque (Cambridge),
- (4) Riemann solvers and numerical methods for fluid dynamics – E. F. Toro (Springer),
- (5) Computational fluid dynamics – J. D. Anderson (McGraw-Hill),
- (6) Computational gasdynamics – C. B. Laney (Cambridge), and
- (7) Fundamentals of computational fluid dynamics – H. Lomax, T. H. Pulliman, D. W. Zingg (Springer).

3.3. Non-textbook reading materials:

- (8) journal papers on computational methods are provided when needed.

4. The Governing Equations

In this chapter, we discuss fundamental principles in fluid dynamics and derive their governing equations, their physical meaning, and their mathematical forms particularly appropriate in CFD.

4.1. The fundamental equations of fluid dynamics

In modeling fluid motion, there are always following philosophy we need to consider. First is to choose the appropriate fundamental physical principles from the law of physics that are:

- (a) Mass is conserved,
- (b) $\mathbf{F} = \mathbf{ma}$ (Newton's second law), and
- (c) energy is conserved.

We apply these physical principles to an appropriate flow model of our interest, and extract the needed mathematical equations which embody such physical principles. As we are interested in physical behaviors of a continuum fluid (or gas dynamics) in this course (rather than those of solid body, i.e., fluid mechanics rather than solid mechanics), we can construct one of the four models in modeling fluid motion:

- (F1) finite control volume approach fixed in space,
- (F2) finite control volume approach moving with the fluid,
- (F3) infinitesimal fluid element fixed in space, and finally,
- (F4) infinitesimal fluid element fixed moving along a streamline.

The first two cases based on finite control volume (FVC) are illustrated in Fig. 5, whereas the last two cases of infinitesimal fluid element (IFE) are shown in Fig. 6. Let's now consider each of the four different approaches and derive the related mathematical relations.

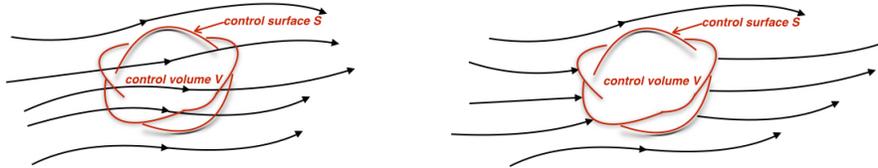


Figure 5. Finite control volume approach. Left: (F1) Finite control volume \mathcal{V} fixed in space with the fluid moving through it. Right: (F2) Finite control volume moving \mathcal{V} with the fluid with the same number of fluid particles kept in the same control volume \mathcal{V} .

4.1.1. General Remarks on FCV (F1 & F2):

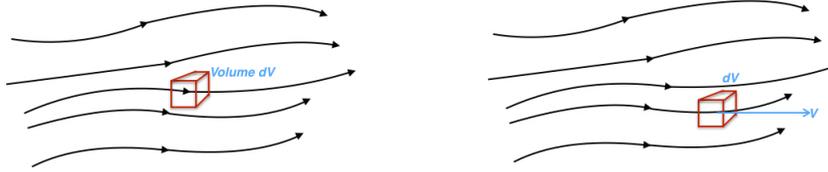


Figure 6. Infinitesimal fluid element approach. Left: (F3) Infinitesimal fluid element dV fixed in space with the fluid moving through it. Right: (F4) Infinitesimal fluid element dV moving along a streamline with the local velocity \mathbf{V} equal to the local flow velocity at each point.

- We conceptually define ‘FCV’ a reasonably large closed region of the flow with a finite volume \mathcal{V} and call its surface a ‘control surface’ S .
- FCV can be put in two different cases: (1) fixed in space with the fluid moving through it – this approach gives rise to the *conservative form* of the governing equations in *integral form*; (2) moving with the fluid such that the same fluid particles are always inside it – this results in the *nonconservative form* of the governing equations *integral form*.
- With the FCV approach, we limit our attention to just the fluid in the finite region of the volume itself (that is, we apply the law of physics to V) instead of looking at the whole flow field at once.

4.1.2. General Remarks on IFE (F3 & F4):

- In this approach we consider an infinitesimally small fluid element in the flow with a differential volume dV .
- The fluid element is infinitesimal in the same sense as differential calculus and is large enough to contain a huge number of molecules (i.e., a continuous medium).
- As in FCV, two approaches are available wherein (3) IFE is fixed in space with the fluid moving through it – *conservative form* in *differential form* of the governing equations; and (4) moving along a streamline with a velocity vector \mathbf{V} equal to the flow velocity at each point – *nonconservative form* of the *differential form* of the governing equations.

Note: We can possibly think of another approach that is based on the fundamental physics applied directly to the atoms and molecules – this is called the *kinetic theory* that solves the Boltzmann equations for individual particle using their distribution functions f_α . Notice that this approach has a microscopic view point in fluid motions, whereas FCV and IFE have a macroscopic view point.

4.2. Two important mathematical relations: D/Dt and $\nabla \cdot \mathbf{V}$

Before we start deriving the above mentioned mathematical relations, let’s first take a moment to refresh our physical insights into two important mathematical

relations: (i) the substantial derivative D/Dt , and (ii) the divergence of velocity fields, $\nabla \cdot \mathbf{V}$.

(i) The substantial derivative D/Dt : Consider adopting the flow model described in F4, which is shown again in Fig. 7 in two different incidents in space and time in Cartesian space. Let's take a velocity vector $\mathbf{V} = u\mathbf{i} + v\mathbf{j} + w\mathbf{k}$, where each component is a function of both space and time,

$$u = u(x, y, z, t), \quad (1.1)$$

$$v = v(x, y, z, t), \quad (1.2)$$

$$w = w(x, y, z, t). \quad (1.3)$$

We denote the scalar density field by

$$\rho = \rho(x, y, z, t). \quad (1.4)$$

The density of the *same* fluid at the two different locations of space and time

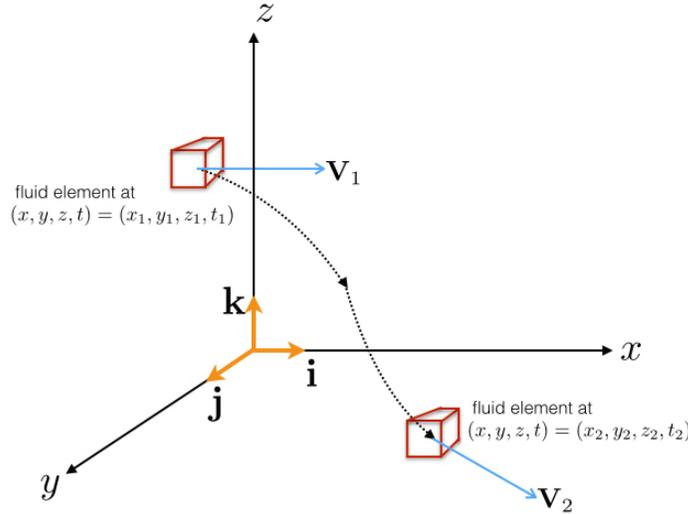


Figure 7. Illustration for the substantial derivative for a fluid element moving in the flow

can be written as $\rho_1 = \rho(x_1, y_1, z_1, t_1)$ and $\rho_2 = \rho(x_2, y_2, z_2, t_2)$, where we can further expand the density function about point 1 as follows:

$$\rho_2 = \rho_1 + \left(\frac{\partial \rho}{\partial x}\right)_1 (x_2 - x_1) + \left(\frac{\partial \rho}{\partial y}\right)_1 (y_2 - y_1) + \left(\frac{\partial \rho}{\partial z}\right)_1 (z_2 - z_1) + \left(\frac{\partial \rho}{\partial t}\right)_1 (t_2 - t_1) + H.O.T \quad (1.5)$$

Dividing by $t_2 - t_1$ and ignoring high-order terms (H.O.T), we get

$$\frac{\rho_2 - \rho_1}{t_2 - t_1} = \left(\frac{\partial \rho}{\partial x}\right)_1 \frac{x_2 - x_1}{t_2 - t_1} + \left(\frac{\partial \rho}{\partial y}\right)_1 \frac{y_2 - y_1}{t_2 - t_1} + \left(\frac{\partial \rho}{\partial z}\right)_1 \frac{z_2 - z_1}{t_2 - t_1} + \left(\frac{\partial \rho}{\partial t}\right)_1 \quad (1.6)$$

Take a look at the LHS of Eq. 1.6 and we notice that this is physically the ‘average’ time rate of change in density of the fluid element as it moves from point 1 to point 2. in the limit of $t_2 \rightarrow t_1$, we get

$$\lim_{t_2 \rightarrow t_1} \frac{\rho_2 - \rho_1}{t_2 - t_1} \equiv \frac{D\rho}{Dt} \quad (1.7)$$

By definition, the symbol is called the substantial derivative D/Dt and it has its physical meaning that measures the time rate of change of a given quantity (density in our current example) of the given fluid element as it moves from one location to another in both space and time.

Note: Notice that there is a clear difference between D/Dt and $\partial/\partial t$ in that the latter is called the *local derivative* which represents the time rate of change at a ‘fixed’ point – our eyes are locked on the stationary point 1; whereas for the first, our eyes are locked on the fluid element as it moves watching its density change as it passes through point 1.

Now, taking the limit of Eq. 1.6 as $t_2 \rightarrow t_1$, we can further cast the relation into

$$\frac{D\rho}{Dt} = u \frac{\partial\rho}{\partial x} + v \frac{\partial\rho}{\partial y} + w \frac{\partial\rho}{\partial z} + \frac{\partial\rho}{\partial t} \quad (1.8)$$

Finally, we can obtain an expression for the substantial derivative in Cartesian coordinate system:

$$\frac{D}{Dt} = \frac{\partial}{\partial t} + u \frac{\partial}{\partial x} + v \frac{\partial}{\partial y} + w \frac{\partial}{\partial z} = \frac{\partial}{\partial t} + \mathbf{V} \cdot \nabla, \quad (1.9)$$

where we have introduced

$$\nabla \equiv \mathbf{i} \frac{\partial}{\partial x} + \mathbf{j} \frac{\partial}{\partial y} + \mathbf{k} \frac{\partial}{\partial z}. \quad (1.10)$$

Quick summary:

- D/Dt is called the *substantial derivative* (or, also called *material derivative*),
- $\partial/\partial t$ is called the *local derivative*, and
- $\mathbf{V} \cdot \nabla$ is called the *convective derivative*.

Note: Recall that the substantial derivative is nothing but a total derivative with respect to time, d/dt . In other words, from differential calculus, we easily see that

$$d\rho = \frac{\partial\rho}{\partial x} dx + \frac{\partial\rho}{\partial y} dy + \frac{\partial\rho}{\partial z} dz + \frac{\partial\rho}{\partial t} dt, \quad (1.11)$$

which yields

$$\frac{d\rho}{dt} = u \frac{\partial\rho}{\partial x} + v \frac{\partial\rho}{\partial y} + w \frac{\partial\rho}{\partial z} + \frac{\partial\rho}{\partial t} \quad (1.12)$$

Example: You are entering an ice cave with a friend of yours. You will experience a temperature decrease as you walk deeper in to the cave – this is analogous to the convective derivative. As you keep walking in to the cave, your friend throws a snowball at you and you feel an additional instantaneous temperature drop when the snowball hits you – this effect is analogous to the local derivative. Notice that the substantial derivative is the sum of the two effects.

(ii) **The divergence of the velocity fields $\nabla \cdot \mathbf{V}$:** Consider a finite control volume (FCV) moving from one place to another depicted as in Fig. 8. In this example, the FCV consists of the same fluid particles when moving, therefore keeping its mass fixed in time. However, its volume \mathcal{V} and its control surface S can vary with time as it moves to a different location of the flow where different density occupies. That is, the control volume keeps changing its volume and shape depending on the characteristic of the flow.

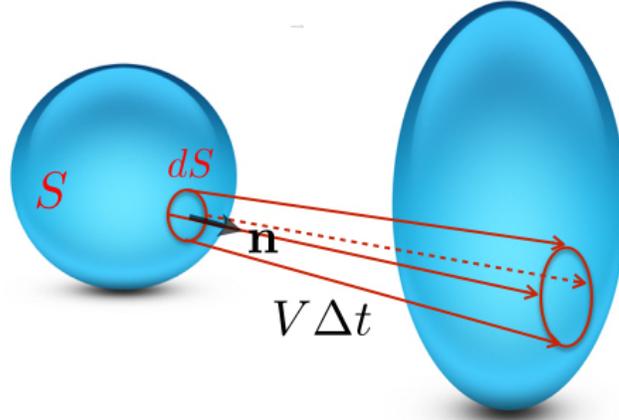


Figure 8. Moving control volume for the physical interpretation of the divergence of the velocity fields

Let us now focus on an infinitesimal element of surface dS moving at the local velocity \mathbf{V} along the normal direction \mathbf{n} which is perpendicular to dS . The change in the volume $\Delta\mathcal{V}$ of the control volume due to the movement of dS over Δt is available by inspecting the volume of the long, thin cylinder with the base area of dS and the height $\mathbf{V}\Delta t \cdot \mathbf{n}$. That is,

$$\Delta\mathcal{V} = \mathbf{V}\Delta t \cdot \mathbf{n}dS = \mathbf{V}\Delta t \cdot \mathbf{dS}, \quad (1.13)$$

where $\mathbf{n}dS = \mathbf{dS}$. In the limit of $dS \rightarrow 0$, the total change in volume of the whole control volume is

$$\int \int_S \mathbf{V}\Delta t \cdot \mathbf{dS}. \quad (1.14)$$

After dividing Eq. 1.14 by Δt and subsequently applying the divergence theorem, we obtain its physical meaning of ‘the time rate of change of the control volume’,

denoted by $\frac{D\mathcal{V}}{Dt}$ (note here that we used the substantial derivative notation of \mathcal{V} as we wish to define the time rate of change of the control volume *as the volume moves along with the flow*):

$$\frac{D\mathcal{V}}{Dt} = \frac{1}{\Delta t} \int \int_S \mathbf{V} \Delta t \cdot d\mathbf{S} = \int \int_S \mathbf{V} \cdot d\mathbf{S} = \int \int \int_{\mathcal{V}} \nabla \cdot \mathbf{V} d\mathcal{V}. \quad (1.15)$$

By keeping continuously shrink \mathcal{V} to $\delta\mathcal{V}$ in such a way that $\delta\mathcal{V}$ is so small enough to treat $\nabla \cdot \mathbf{V}$ as constant in $\delta\mathcal{V}$. Then in the limit of $\delta\mathcal{V} \rightarrow 0$, we can rewrite Eq. 1.15 as

$$\frac{D(\delta\mathcal{V})}{Dt} = \int \int \int_{\delta\mathcal{V}} \nabla \cdot \mathbf{V} d\mathcal{V} = \nabla \cdot \mathbf{V} \delta\mathcal{V}, \quad (1.16)$$

or

$$\nabla \cdot \mathbf{V} = \frac{1}{\delta\mathcal{V}} \frac{D(\delta\mathcal{V})}{Dt} \quad (1.17)$$

Quick summary:

- $\nabla \cdot \mathbf{V}$ physically means the time rate of change of the volume of a moving fluid element per unit volume.

4.3. The Continuity Equation

We are now ready to apply the philosophy discussed in Sec. 4.1. to all four of the flow models illustrated in Figs. 5 and 6. Let's begin with the first principle:

- (a) Mass is conserved.

We are going to derive the continuity equation in four different ways and see they are all related mathematically.

(F1) FCV fixed in space:

Let us examine the principle of the mass conservation by considering a small control volume \mathcal{V} surrounded by its control surface S as depicted in the left panel of Fig. 5. Then the mass conservation law can be stated as:

The net mass flow 'out' of \mathcal{V} through surface S = The time rate of 'decrease' of mass inside \mathcal{V}

In order to obtain a mathematical expression for LHS, we write the mass flow of a moving fluid with fluid velocity \mathbf{V} across any fixed surface. The elemental mass flow across the area dS normal to \mathbf{n} becomes

$$\rho \mathbf{V} \cdot \mathbf{n} dS = \rho \mathbf{V} \cdot d\mathbf{S} \quad (1.18)$$

Recall that by convention, the direction of the flow is 'out' of \mathcal{V} because $d\mathbf{S}$ points in a direction 'out' of \mathcal{V} , hence the mass inside \mathcal{V} 'decreases' in the above statement. By taking the surface integral of Eq. 1.18, we obtain the net mass flow out of the entire control volume \mathcal{V} – the expression for LHS:

$$\int \int_S \rho \mathbf{V} \cdot d\mathbf{S} \quad (1.19)$$

The expression for RHS is the time rate of ‘decrease’ of the total mass $\int \int \int_{\mathcal{V}} \rho d\mathcal{V}$ inside \mathcal{V} , that is,

$$-\frac{\partial}{\partial t} \int \int \int_{\mathcal{V}} \rho d\mathcal{V} \quad (1.20)$$

Equating the two, we finally get a mathematical relation for the mass conservation:

$$\frac{\partial}{\partial t} \int \int \int_{\mathcal{V}} \rho d\mathcal{V} + \int \int_S \rho \mathbf{V} \cdot d\mathbf{S} = 0 \quad (1.21)$$

Note: We emphasize that Eq. 1.21 is an *integral form of the continuity equation*. The ‘finite’ aspect of the control volume is why the equation is obtained directly in integral form. The fact that the control volume was ‘fixed in space’ resulted in the specific integral form given by Eq. 1.21, which is called the *conservation form*.

(F2) FCV moving with the fluid: As seen earlier, we can write another mathematical expression for the mass conservation law using the substantial derivative which perfectly describes behavior of the time rate of change of any property of a fluid element moving with the flow. That is to say, the mass conservation law is simply put into a form

$$\frac{D}{Dt} \int \int \int_{\mathcal{V}} \rho d\mathcal{V} = 0 \quad (1.22)$$

Note: We remark that Eq. 1.22 is also an *integral form of the continuity equation* which is different from the previous result – this is now called the *non-conservation form*. Comparing with the previous conservation form, we can see that the nonconservative form is a result of considering the control volume *moving* with the fluid.

(F3) IFE fixed in space: For convenience we adopt an infinitesimal fluid element fixed in space in a Cartesian coordinate system shown in Fig. 9. What we want to calculate is the net mass flow through all surrounding six faces with the elemental areas of $dx dy$, $dy dz$ and $dx dz$. As illustrated in Fig. 9, we consider each individual net flow in each coordinate direction. They are

(a) the net outflow in x -direction:

$$\left(\rho u + \frac{\partial \rho u}{\partial x} dx\right) dy dz - (\rho u) dy dz = \frac{\partial \rho u}{\partial x} dx dy dz, \quad (1.23)$$

(b) the net outflow in y -direction:

$$\left(\rho v + \frac{\partial \rho v}{\partial y} dy\right) dx dz - (\rho v) dx dz = \frac{\partial \rho v}{\partial y} dx dy dz, \quad (1.24)$$

(c) the net outflow in z -direction:

$$\left(\rho w + \frac{\partial \rho w}{\partial z} dz\right) dx dy - (\rho w) dx dy = \frac{\partial \rho w}{\partial z} dx dy dz. \quad (1.25)$$

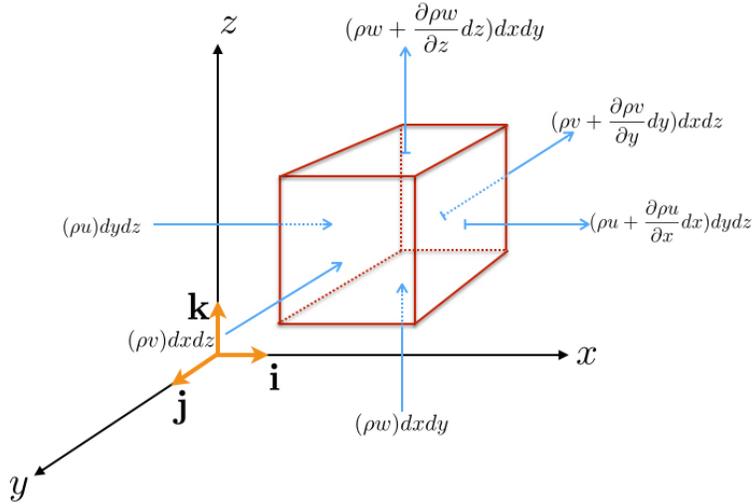


Figure 9. Model of the infinitesimal fluid element fixed in space and mass fluxes through various faces of the element

Hence, the net mass flow out of the element in all directions is given by summing all of the above relations:

$$\left(\frac{\partial \rho u}{\partial x} + \frac{\partial \rho v}{\partial y} + \frac{\partial \rho w}{\partial z} \right) dx dy dz, \quad (1.26)$$

which should be equal to the time rate of decrease of the total mass $\rho dx dy dz$ in the infinitesimal element of volume $dx dy dz$:

$$-\frac{\partial \rho}{\partial t} dx dy dz \quad (1.27)$$

Equating the two we get yet another form describing the mass conservation

$$\frac{\partial \rho}{\partial t} + \left(\frac{\partial \rho u}{\partial x} + \frac{\partial \rho v}{\partial y} + \frac{\partial \rho w}{\partial z} \right) = \frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{V}) = 0 \quad (1.28)$$

Note: We call Eq. 1.28 the *differential form of the continuity equation in conservation form*. The ‘infinitesimal’ aspect of the small element lead to the differential form of the equation, and, as before, the fact that the fluid element was ‘fixed in space’ resulted in the *conservation form*.

(F4) IFE moving with the fluid: We remind ourselves that although the mass of an IFE is conserved when it moves with the fluid, its elemental volume $\delta \mathcal{V}$ varies. Since the mass in the IFE is invariant, invoking the physical meaning of the substantial derivative and using the chain rule, we have

$$0 = \frac{D\rho \delta \mathcal{V}}{Dt} = \delta \mathcal{V} \frac{D\rho}{Dt} + \rho \frac{D\delta \mathcal{V}}{Dt} \quad (1.29)$$

Combining the definition of the divergence of the velocity fields in Eq. 1.17, this can be rewritten as

$$\frac{D\rho}{Dt} + \rho \nabla \cdot \mathbf{V} = 0 \quad (1.30)$$

Note: We call Eq. 1.30 the *differential form of the continuity equation in non-conservation form*. The ‘infinitesimal’ aspect of the small element lead to the differential form of the equation, while the fact that the fluid element was ‘moving with the fluid’ resulted in the *nonconservation form* as in (F2).

Homework 1. Often times, the condition for incompressible flows is given by $\nabla \cdot \mathbf{V} = 0$. Why?

Homework 2. Derive the integral form of the momentum equation in conservation form

$$\frac{\partial}{\partial t} \int \int \int_{\mathcal{V}} \rho \mathbf{V} d\mathcal{V} + \int \int_S (\rho \mathbf{V} \cdot \mathbf{dS}) \mathbf{V} = \int \int \int_{\mathcal{V}} \rho \mathbf{f} d\mathcal{V} - \int \int_S p \mathbf{dS} \quad (1.31)$$

using the Newton’s second law applied to a fluid flow. Ignore any viscous effect (Hint: $\mathbf{F} = \frac{d}{dt}(m\mathbf{V})$).

Homework 3. Derive the integral form of the energy equation in conservation form using the energy conservation law for adiabatic inviscid flows:

$$\frac{\partial}{\partial t} \int \int \int_{\mathcal{V}} \rho \left(e + \frac{V^2}{2} \right) d\mathcal{V} + \int \int_S \rho \left(e + \frac{V^2}{2} \right) \mathbf{V} \cdot \mathbf{dS} = \int \int \int_{\mathcal{V}} \rho \mathbf{f} \cdot \mathbf{V} d\mathcal{V} - \int \int_S p \mathbf{V} \cdot \mathbf{dS} \quad (1.32)$$

Homework 4. Show that all four approaches discussed in (F1)-(F4) for the continuity equation are in fact all the same. That is, one of them can be obtained from any of the others. (Hint: You can show that there are equivalent relationships in circle: (F1) \Rightarrow (F2) \Rightarrow (F4) \Rightarrow (F3) \Rightarrow (F1))

Chapter 2

Reviews on PDEs

1. Properties of PDEs

In this chapter, we study the key defining properties of partial differential equations (PDEs). First of all, there are more than one ‘independent’ variables t, x, y, z, \dots . Associated to these is so called a ‘dependent’ variable u (of course there could be more than one dependent variables) which is a function of those independent variables,

$$u = u(x, y, z, t, \dots) \tag{2.1}$$

We now provide a bunch of basic definitions and examples on PDEs.

Definition: A PDE is a relation between the independent variables and the dependent variable u via the partial derivatives of u .

Definition: The order of PDE is the highest derivative that appears.

Example: $F(x, y, u, u_x, u_y) = 0$ is the most general form of first-order PDE in two independent variables x and y .

Example: $F(t, x, y, u, u_t, u_{xx}, u_{xy}, u_{yy}) = 0$ is the most general form of second-order PDE in three independent variables t, x and y .

Example: $u_t - u_{xx} = 0$ is a second-order PDE in two independent variables t and x .

Example: $u_{xxxx} + (u_y)^3 = 0$ is a fourth-order PDE in two independent variables x and y .

Definition: \mathcal{L} is called a linear operator if $\mathcal{L}(u+v) = \mathcal{L}u + \mathcal{L}v$ for any functions u and v .

Definition: A PDE $\mathcal{L}u = 0$ is called a linear PDE if \mathcal{L} is a linear derivative operator.

Definition: A PDE $\mathcal{L}u = g$ is called an inhomogeneous linear PDE if \mathcal{L} is a linear derivative operator and if $g \neq 0$ is a given function of the independent variables. If $g = 0$, it is called a homogeneous linear PDE.

Example: The following PDEs are homogeneous linear:

$u_x + u_y = 0$ (transport); $u_x + yu_y = 0$ (transport); $u_{xx} + u_{yy} = 0$ (Laplace's equation)

Example: The following PDEs are homogeneous nonlinear:

$u_x + uu_y = 0$ (shock wave); $u_{tt} + u_{xx} + u^3 = 0$ (wave with interaction);
 $u_t + uu_x + u_{xxx} = 0$ (dispersive wave);

Example: The following PDEs are inhomogeneous linear:

$\cos(xy^2)u_x - y^2u_y = \tan(x^2 + y^2)$

2. Well-posedness of PDEs

When solving PDEs, one often encounters a problem that has more than one solution (non-uniqueness) if few auxiliary conditions are imposed. Then the problem is called underdetermined. On the other hand, if too many conditions are given, there may be no solution at all (non-existence) and in this case, the problem is overdetermined.

The well-posedness property of PDEs is therefore required in order for us to enable to solve the given PDE system successfully. Well-posed PDEs of proper initial and boundary conditions follows the following fundamental properties:

1. Existence: There exists at least one solution $u(x, t)$ satisfying all these conditions,
2. Uniqueness: There is at most one solution,
3. Stability: The unique solution $u(x, t)$ depends in a stable manner on the data of the problem. This means that if the data are changed a little, the corresponding solution changes only a little as well.

3. Classifications of Second-order PDEs

PDEs arise in a number of physical phenomena to describe their natures. Some of the most popular types of such problems include fluid flows, heat transfer, solid mechanics and biological processes. These types of equations often fall into one of three types, (i) hyperbolic PDEs that are associated with advection, (ii) parabolic PDEs that are most commonly associated with diffusion, and (iii) elliptic PDEs that most commonly describe steady states of either parabolic or hyperbolic problems.

In reality, not many problems fall simply into *one* of these three types, rather most of them involve combined types, e.g., advection-diffusion problems.

Mathematically, however, we can rather easily determine the type of a general second-order PDEs, which we are going to briefly discuss here.

In general, let's consider the PDE of form with nonzero constants a_{11} , a_{12} , and a_{22} :

$$a_{11}u_{xx} + 2a_{12}u_{xy} + a_{22}u_{yy} + a_1u_x + a_2u_y + a_0u = 0, \quad (2.2)$$

which is a second-order linear equation in two independent variables x and y with six constant coefficients.

Theorem: By a linear transformation of the independent variables, the equation can be reduced to one of three forms:

1. Elliptic PDE: if $a_{12}^2 < a_{11}a_{22}$, it is reducible to

$$u_{xx} + u_{yy} + L.O.T = 0 \quad (2.3)$$

where $L.O.T$ denotes all the lower order terms (first or zeroth order terms).

2. Hyperbolic PDE: if $a_{12}^2 > a_{11}a_{22}$, it is reducible to

$$u_{xx} - u_{yy} + L.O.T = 0 \quad (2.4)$$

3. Parabolic PDE: if $a_{12}^2 = a_{11}a_{22}$ (the condition for parabolic is in between those of elliptic and hyperbolic), it is reducible to

$$u_{xx} + L.O.T = 0 \quad (2.5)$$

Remark: Notice the similarity between the above classification and the one in analytic geometry. We know from analytic geometry that, given (again assuming nonzero constants a_{11} , a_{12} , and a_{22})

$$a_{11}x^2 + 2a_{12}xy + a_{22}y^2 + a_1x + a_2y + a_0 = 0, \quad (2.6)$$

Then Eq. 2.6 becomes

1. Ellipsoid if $a_{12}^2 < a_{11}a_{22}$
2. Hyperbola if $a_{12}^2 > a_{11}a_{22}$
3. Parabola if $a_{12}^2 = a_{11}a_{22}$.

Note again that parabola is in between ellipsoid and hyperbola. See Fig. 1 for an illustration.

Example: $u_{xx} - 5u_{xy} = 0$ is hyperbolic; $4u_{xx} - 12u_{xy} + 9u_{yy} + u_y = 0$ is parabolic; $4u_{xx} + 6u_{xy} + 9u_{yy} = 0$ is elliptic.

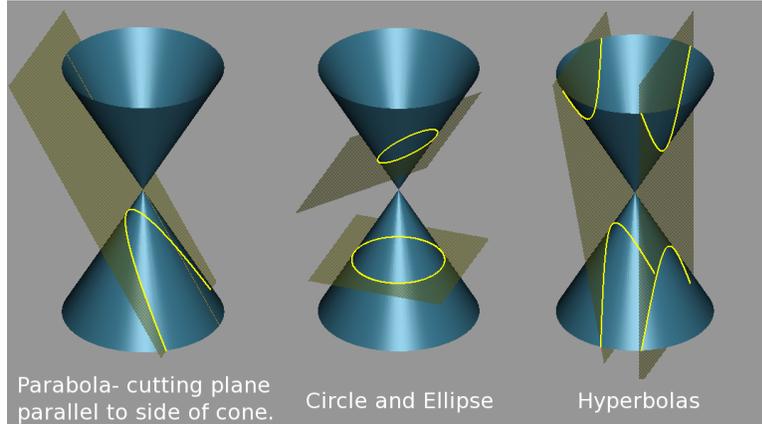


Figure 1. Three major types of conic section from analytic geometry –
Image source: Wikipedia

Example: The wave equation is one of the most famous examples in hyperbolic PDEs. We write the wave equation as

$$u_{tt} = c^2 u_{xx} \text{ for } -\infty < x < \infty, c \neq 0. \quad (2.7)$$

Factoring the derivative operator, we get

$$\left(\frac{\partial}{\partial t} - c\frac{\partial}{\partial x}\right)\left(\frac{\partial}{\partial t} + c\frac{\partial}{\partial x}\right)u = 0 \quad (2.8)$$

Considering the characteristic coordinates $\xi = x + ct$ and $\eta = x - ct$, we obtain

$$0 = \left(\frac{\partial}{\partial t} - c\frac{\partial}{\partial x}\right)\left(\frac{\partial}{\partial t} + c\frac{\partial}{\partial x}\right)u = \left(-2c\frac{\partial}{\partial \xi}\right)\left(2c\frac{\partial}{\partial \eta}\right)u \quad (2.9)$$

Hence, we conclude that the general solution must have a form $u(x, t) = f(x + ct) + g(x - ct)$, the sum of two functions, one (g) is a wave of any shape traveling to the the *right* at speed c , and the other (f) with another arbitrary shape traveling to the the *left* at speed c . We call the two families of lines, $x \pm ct = \text{constant}$, the characteristic lines of the wave equation.

Example: One very simple and famous example in the parabolic PDEs is so called the diffusion equation

$$u_t = ku_{xx}, \text{ with } k \text{ constant and } (x, t) \in D \times T \quad (2.10)$$

One of the important properties in the diffusion equations is to have the maximum principle. Recall that the maximum principle says if $u(x, t)$ is the solution of Eq. 2.10 on $D \times T = [x_{min}, x_{max}] \times [T_0, T_1]$ in space-time, then the maximum value of $u(x, t)$ is assumed only on the initial and domain boundary of $D \times T$. That is, the maximum value only occurs either initially at $t = T_0$ or on the sides

$x = x_{min}$ or $x = x_{max}$.

Remark: The fundamental properties of the two types of PDEs can be briefly compared in the following table. The physical meanings in Table 1 are also illustrated in Fig. 2 and Fig. 3.

Table 1. Comparison of Waves and Diffusions: Fundamental properties of the wave and diffusion equations are summarized.

Property	Waves	Diffusions
(1) speed of propagation	finite ($\leq c$)	∞
(2) singularities for $t > 0$?	transported along characteristics (with speed = c)	lost immediately
(3) well-posed for $t > 0$?	yes	yes (at least for bounded solutions)
(4) well-posed for $t < 0$?	yes	no
(5) maximum principle?	no	yes
(6) behavior as $t \rightarrow \infty$	energy is constant so does not decay (i.e., simple advection without diffusion)	decays to zero
(7) information	transported	lost gradually

4. Finite difference scheme for 1D advection

Consider a simple advection equation with constant speed $c > 0$:

$$u_t + cu_x = 0, \text{ with } u(x, 0) = \sin(x), x \in [0, 2\pi] \quad (2.11)$$

with a periodic boundary condition. In order to discretize the system, we first subdivide both spatial and temporal domains as

$$x_i = i\Delta x \text{ and } t^n = n\Delta t, \quad (2.12)$$

where i and n are integers. $\Delta x > 0$ and $\Delta t > 0$ are respectively, a spatial grid spacing and a time step. Let us denote our discrete data at each (x_i, t^n) :

$$u_i^n = u(x_i, t^n) \quad (2.13)$$

The forward difference scheme writes

$$u_x(x, t) = \frac{u(x + \Delta x, t) - u(x, t)}{\Delta x} + O(\Delta x), \quad (2.14)$$

$$u_t(x, t) = \frac{u(x, t + \Delta t) - u(x, t)}{\Delta t} + O(\Delta t). \quad (2.15)$$

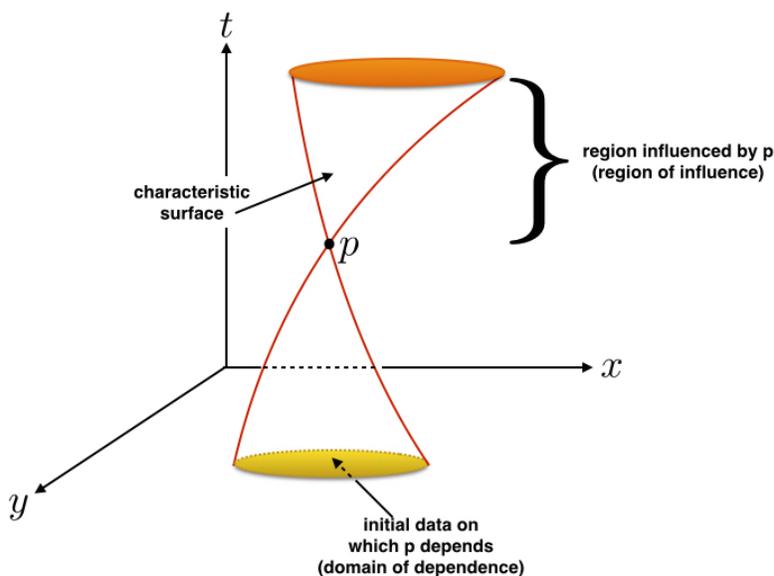


Figure 2. Domain and boundaries for the solution of hyperbolic PDEs in 2D. Note that any information or disturbance introduced at p is going to affect *only* the region called the ‘region of influence’ but nowhere. Such information is propagated with the finite advection speed along the characteristic surface which forms the conic region of influence. On the other hand, if the characteristic surface can be extended backward in time to the place where the initial data is imposed. This also forms another conic section on the lower part of the figure which is called the ‘domain of dependence’.

Dropping the truncation error terms $O(\Delta x)$ and $O(\Delta t)$ yields a simple first-order difference scheme that approximates the advection PDE. As a result, we arrive at a first-order accurate discrete difference equation from an analytic differential equation:

$$\frac{u_i^{n+1} - u_i^n}{\Delta t} + c \frac{u_{i+1}^n - u_i^n}{\Delta x} = 0, \quad (2.16)$$

which gives a temporal update scheme of u_i^{n+1} in terms of the known data at $t = t^n$:

$$u_i^{n+1} = u_i^n - c \frac{\Delta t}{\Delta x} (u_{i+1}^n - u_i^n) \quad (2.17)$$

On the other hand, if we use a backward difference scheme for u_x

$$u_x(x, t) = \frac{u(x, t) - u(x - \Delta x, t)}{\Delta x} + O(\Delta x), \quad (2.18)$$

we arrive at another first-order difference equation

$$u_i^{n+1} = u_i^n - c \frac{\Delta t}{\Delta x} (u_i^n - u_{i-1}^n). \quad (2.19)$$

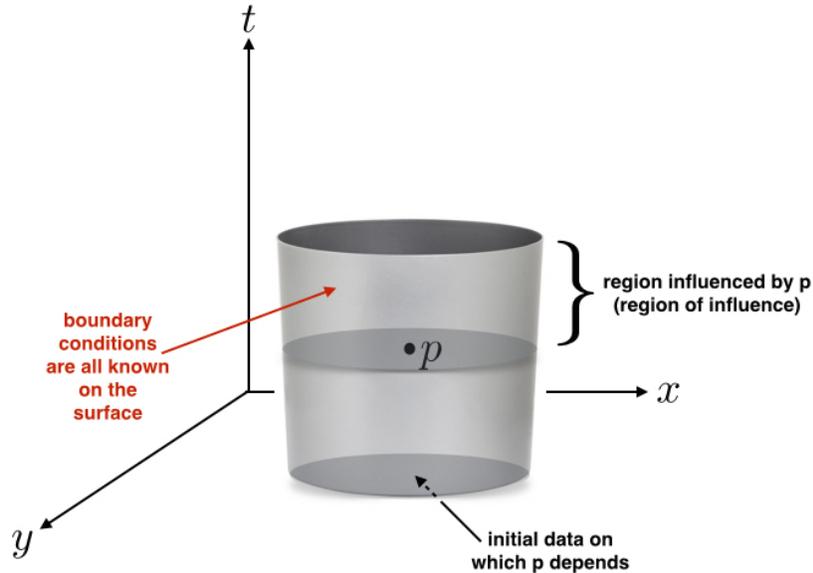


Figure 3. Domain and boundaries for the solution of parabolic PDEs in 2D. Note that from a given point p in the mid plane, there is only one physically meaningful direction that is positive in t . Therefore, any information at p influences the entire region onward from p , called the 'region of influence'. Such information can only march forward in time under the assumption that all boundary conditions around the surface and the initial condition are known.

Let us choose Δt small enough that

$$|c|\Delta t \leq \Delta x \quad (2.20)$$

Homework 1. Write a simple MATLAB program (or use any other scientific language) in order to numerically solve Eq. 2.17 and Eq. 2.19. Please make sure your code satisfies the condition in Eq. 2.20. Choose $t = t_{max}$ in such that the initial sinusoidal wave makes two complete cycles over the domain (we conveniently assume the cgs unit system – e.g., *cm* in length, *sec* in time, *gram* in mass.).

- Use the grid sizes of 16, 32, 64, 128 and 256 and compare your results.
- First solve for $c > 0$. Which scheme is better between Eq. 2.17 and Eq. 2.19?
- What happens if $c < 0$?
- What happens if your Δt fails to satisfy Eq. 2.20 for your choices of c and Δx ?
- Plot your numerical solutions at $t = t_{cycle1}$ and $t = t_{cycle2}$ on a grid size of 32 using $c > 0$ and the scheme in Eq. 2.19. What do you observe?

5. Numerical Solutions of 1D Diffusion

Consider a temporal evolution of solving the classical homogeneous heat equation (or diffusion equation) of the form

$$u_t = \kappa u_{xx} \quad (2.21)$$

with $\kappa > 0$ (Note if $\kappa < 0$ then Eq. 2.21 would be a “backward heat equation”, which is an ill-posed problem. See Table 1). Along with this equation, let us impose an initial condition at $t = 0$,

$$u(x, 0) = f(x) \quad (2.22)$$

and also the Dirichlet boundary condition on a bounded domain $0 \leq x \leq 1$

$$u(0, t) = g_0(t) \text{ and } u(1, t) = g_1(t), \text{ for } t > 0. \quad (2.23)$$

Use the discretization technique we used in the previous example of the 1D advection finite difference scheme in order to discretize your temporal and spatial domains (i.e., Eq. 2.12 and Eq. 2.13). As before, we choose the forward difference scheme for temporal discretization as in Eq. 2.15. For a spatial discretization, we adopt the standard second-order central difference difference scheme,

$$u_{xx}(x, t) = \frac{u(x + \Delta x, t) - 2u(x, t) + u(x - \Delta x, t)}{\Delta x^2} + O(\Delta x^2), \quad (2.24)$$

which gives a final discrete form of our explicit finite difference scheme for the heat equation:

$$u_i^{n+1} = u_i^n + \kappa \frac{\Delta t}{\Delta x^2} (u_{i+1}^n - 2u_i^n + u_{i-1}^n) \quad (2.25)$$

Similar to the 1D advection case, we choose Δt satisfying

$$\kappa \Delta t \leq \frac{\Delta x^2}{2} \quad (2.26)$$

Homework 2. Write a simple MATLAB program (or use any other scientific language) in order to numerically solve Eq. 2.21. The boundary condition is given so as to hold the temperature u to be zero at $x = 0$ and 100°F at $x = 1$ for $t \geq 0$ (i.e., $g_0 = 0^\circ \text{F}$ and $g_1 = 100^\circ \text{F}$). Your numerical scheme solves three different temporal evolutions for three materials:

- (i) iron with $\kappa = 0.230 \text{cm}^2/\text{sec}$,
- (ii) aluminum $\kappa = 0.975 \text{cm}^2/\text{sec}$, and
- (iii) copper with $\kappa = 1.156 \text{cm}^2/\text{sec}$.

Choose $t = t_{max}$ in each so that each material reaches to a steady state solution. Your initial condition in all three cases is to describe a same initial temperature profile

$$f(x) = 0^\circ \text{F for } 0 \leq x < 1; f(x) = 100^\circ \text{F for } x = 1 \quad (2.27)$$

(a) Use the grid sizes of 16, 32, 64, 128 and 256 and compare your results. What

can you say about the grid resolution study in the diffusion equation as compared to the case of the advection equation?

- (b) What happens if your Δt fails to satisfy Eq. 2.26 for each κ ?
- (c) What are your values of t_{max} for three different materials?

Chapter 3

Scalar Conservation Laws - Theories

In many practical applications of CFD, one mostly tackles physical phenomena described by ‘systems’ of (nonlinear) equations such as the Euler or Navier-Stokes equations. Solving such systems is more complicated than solving a scalar equation (linear or nonlinear) in both mathematical and computational aspects.

However, we often gain rich insights in our understandings of the more complicated systems from studying the simpler systems first. In this chapter, we seek for a good understanding of the linear and nonlinear scalar advection equations, whereby it will enlighten us in achieving our bigger goals in studying the systems of (nonlinear) conservation laws later.

1. Linear scalar equations

We consider two types of linear scalar advection equations, one with a constant velocity a , and the other with a variable velocity $a(t)$. Let’s first take a look at the 1D linear scalar advection equation for $t \geq 0$ written as

$$u_t + au_x = 0 \tag{3.1}$$

with a constant advection velocity a , and together with initial conditions on \mathbb{R} ,

$$u(x, 0) = u_0(x). \tag{3.2}$$

As shown in the previous chapter, we know the solution is given by

$$u(x, t) = u_0(x - at) \tag{3.3}$$

for $t \geq 0$. Recall that $x - at = x_0$ is called the characteristic line with a given constant x_0 and with the propagation velocity a . Depending on the sign of a , the initial data $u_0(x)$ is advected (or transported) – hence the name ‘advection equation’ – to the right (if $a > 0$) or left (if $a < 0$). Note that there are infinitely many characteristic lines in the x - t plane as there are infinite choices of $x_0 \in \mathbb{R}$. See Fig. 1.

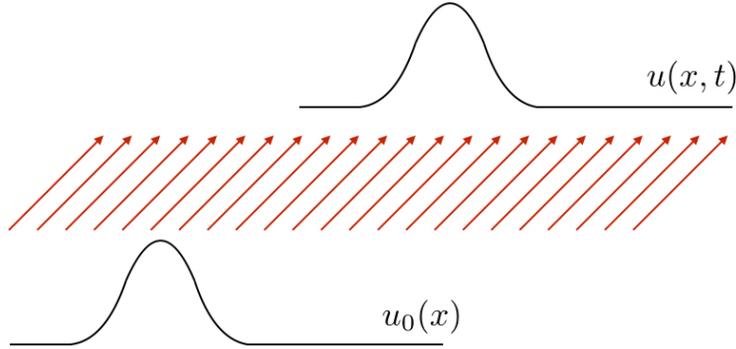


Figure 1. Characteristic curves and the advection of the solution. All information is simply advected to the later time solution $u(x, t)$ along the characteristic curves in the x - t plane *without* any shape changes from the initial condition $u_0(x)$.

In general, the characteristics are curves (or simply ‘the characteristics’) in the x - t plane satisfying the ODEs

$$x'(t) = a \text{ and } x(0) = x_0. \quad (3.4)$$

One very important property on the characteristics is that the solution $u(x, t)$ of the constant velocity a remains as constant along the characteristics. To see this,

$$\frac{d}{dt}u(x(t), t) = \frac{\partial}{\partial t}u(x(t), t) + \frac{\partial}{\partial x}u(x(t), t)x'(t) = u_t + au_x = 0, \quad (3.5)$$

confirming the claim.

In the more general case of the scalar equation with the variable velocity $a(x(t))$, we consider

$$u_t + \left(a(x(t))u\right)_x = 0. \quad (3.6)$$

In this case, the characteristics are no longer straight lines satisfying

$$x'(t) = a(x(t)) \text{ and } x(0) = x_0, \quad (3.7)$$

and the solution $u(x, t)$ is no longer constant along the characteristics. This can be easily verified if we rewrite Eq. 3.6 as

$$u_t + a(x(t))u_x = -a'(x(t))u, \quad (3.8)$$

therefore we obtain

$$\frac{d}{dt}u(x(t), t) = -a'(x(t))u \neq 0. \quad (3.9)$$

In both cases of the constant and variable velocities, the solution can be easily determined by solving sets of ODEs.

Remark: In words, the characteristic curves track the motion of material particles.

Remark: We can see that if $u_0(x) \in C^k(\mathbb{R})$ then $u(x, t) \in C^k(\mathbb{R}) \times (0, \infty)$.

Remark: So far, we have assumed differentiability of $u(x, t)$ in manipulating the above relations. Note that this assumption makes it possible to seek for a classical solution $u(x, t)$ of the differential equations.

1.1. Domain of dependence & Range of influence

We now make an important observation in solutions to the linear advection equations:

The solution $u(x, t)$ at any point (\bar{x}, \bar{t}) depends only on the initial data u_0 only at a *single* point, namely \bar{x}_0 such that (\bar{x}, \bar{t}) lies on the characteristic through \bar{x}_0 .

This means that the solution $u(\bar{x}, \bar{t})$ will remain unchanged no matter how we change the initial data at any points other than \bar{x}_0 . We now define two related regions, the first is called the domain of dependence, and the second is called the range of influence.

Definition: The set $\bar{\mathcal{D}}(\bar{x}, \bar{t}) = \{\bar{x} - \lambda_m \bar{t} : m = 1, 2, \dots, p\}$ is called the domain of dependence of the point (\bar{x}, \bar{t}) , where p is the total number of characteristic velocities (or the number of equations of hyperbolic PDE systems). See Fig. 2 for an illustration.

Remark: For convenience, let us assume $\lambda_1 \leq \dots \leq \lambda_m \leq \dots \leq \lambda_p$. Note that $p = 1$ for *scalar* hyperbolic equations, whereas $p > 1$ for *systems* of hyperbolic equations. For instance, $p = 3$ for the systems of 1D Euler equations (1 continuity equations, 1 momentum equation, and 1 energy equation).

Note: What are the values of p for the systems of 2D Euler and 3D Euler equations?

Definition: The region $\mathcal{R} = \{x : \lambda_1 t \leq x - x_0 \leq \lambda_p t\}$ is called the range of influence of the point x_0 . See Fig. 3 for an illustration.

Note: One can always find a bounded set $\mathcal{D} = \{x : |x - \bar{x}| \leq \lambda_p t\}$ such that $\bar{\mathcal{D}}(\bar{x}, \bar{t}) \subset \mathcal{D}$. The existence of $\bar{\mathcal{D}}$ and \mathcal{R} are the consequence of the fact that hyperbolic equations have *finite propagation speed*; information can travel with speed at most

$$\max\{|\lambda_m| : m = 1, \dots, p\}$$

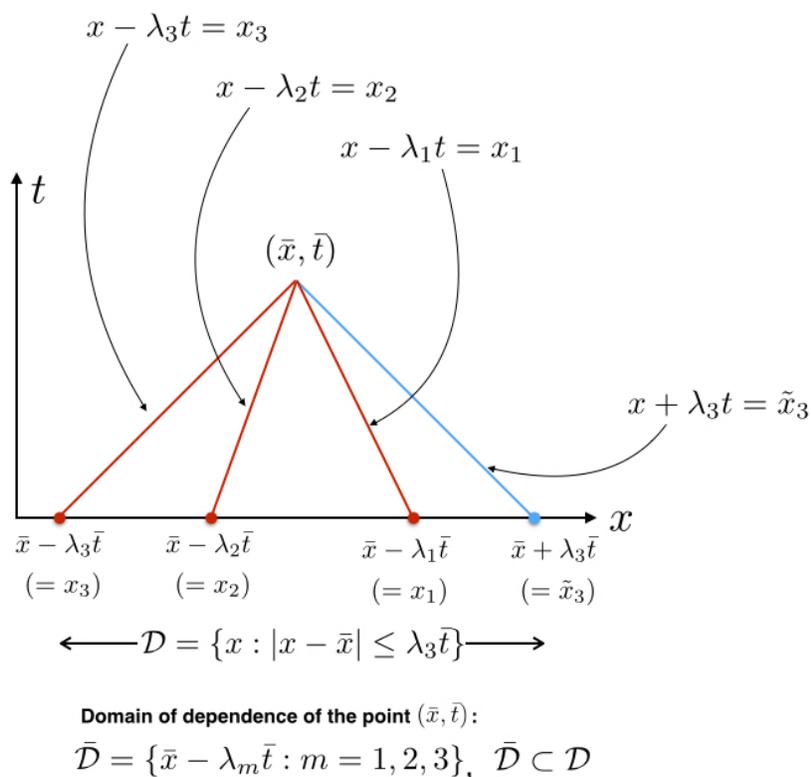


Figure 2. The domain of dependence of the point (\bar{x}, \bar{t}) for a typical hyperbolic system of three equations with $\lambda_1 < 0 < \lambda_2 < \lambda_3$. Note that one can always find a *bounded* domain \mathcal{D} such that $\bar{\mathcal{D}} \subset \mathcal{D}$ because of the fact that the propagation velocities (or characteristic velocities) of hyperbolic PDEs are always finite.

1.2. Non-smooth data

Consider for a moment what happens if $u_0(x)$ has a singularity at some point x_0 (i.e., a discontinuity in u_0 or some derivatives). In this case, the resulting $u(x, t)$ will have a singularity of the same order along the characteristic curve through x_0 . This is a fundamental property of *linear* hyperbolic equations in which singularities are simply advected only along characteristics (Also see Fig. 2). This is because the solution, $u(x, t) = u_0(x - at)$, along a characteristic curve $x - at = x_0$, only depends on the one and only value $u_0(x_0)$, thus allowing a non-smooth “solution” to the PDE even if $u_0(x)$ is not smooth.

Such non-smooth solution, although it is no longer a classical solution of the differential equation everywhere, *does* satisfy the integral form of the conservation law, which continues to make sense for non-smooth u as long as u is an integrable function.

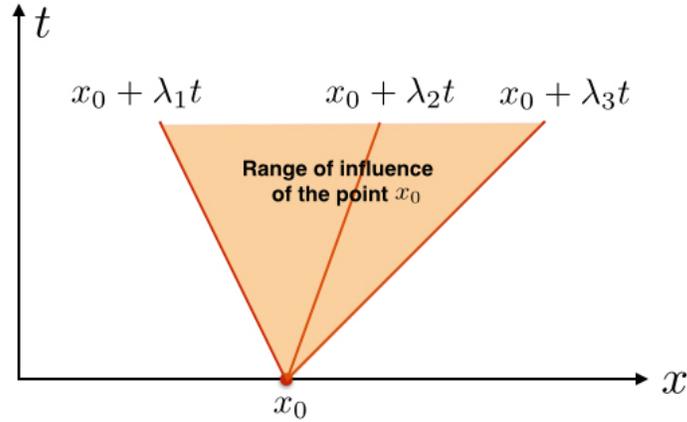


Figure 3. The range of influence $\mathcal{R} = \{x : \lambda_1 t \leq x - x_0 \leq \lambda_3 t\}$ of the point x_0 of the same problem in Fig. 2. Notice that the conic region \mathcal{R} is a symmetric image of \mathcal{D} with respect to (\bar{x}, \bar{t}) , shifted to $t = 0$ axis.

Therefore, it may sound like a perfect idea to accept this concept – i.e., integrating along characteristics regardless of the regularity of $u_0(x)$ – in order to achieve a generalized solution $u(x, t)$. Unfortunately, we can no longer simply integrate along characteristics when solving the *nonlinear* equations (yes, the linear case is relatively too easy!) because the nonlinear characteristic curves often converge (collide) each other to form a shock, losing their characteristic information for good. The nonlinear equations also can develop singularities even from a smooth initial data $u_0(x)$.

One working idea that can be generalized to both linear and nonlinear equations, is to leave the initial data alone but modify the PDE by adding a small diffusive term ϵu_{xx} and take the limit of the diffusive term as $\epsilon \rightarrow 0$. The solution obtained this way is called the *vanishing viscosity* solution. Mathematically, one writes an advection-diffusion equation

$$u_t + au_x = \epsilon u_{xx} \quad (3.10)$$

as an approximation to the advection equation for very small $\epsilon > 0$. Notice that we can always find the solution $u^\epsilon \in C^\infty(\mathbb{R}) \times \mathbb{R}^+$ to Eq. 3.10 even if $u_0(x)$ is not smooth, because Eq. 3.10 is a parabolic equation (why? See Homework 1). We can therefore obtain a generalized solution $u(x, t)$ by

$$\lim_{\epsilon \rightarrow 0} u^\epsilon(x, t) = u(x, t). \quad (3.11)$$

Homework 1 Use a change of variables to follow the characteristics and set

$$v^\epsilon(x, t) = u^\epsilon(x + at, t). \quad (3.12)$$

First show that v^ϵ satisfies the heat equation

$$v_t^\epsilon(x, t) = \epsilon v_{xx}^\epsilon(x, t). \quad (3.13)$$

Now show that, using the well-known solution to the heat equation to solve for $v^\epsilon(x, t)$, show that we have

$$u^\epsilon(x, t) = v^\epsilon(x - at, t), \quad (3.14)$$

where

$$\lim_{\epsilon \rightarrow 0} u^\epsilon(x, t) = u_0(x - at). \quad (3.15)$$

2. Nonlinear scalar equations

We now move on to consider the nonlinear scalar equation

$$u_t + \left(f(u) \right)_x = 0 \quad (3.16)$$

where $f(u)$ is a nonlinear function of u and is called the flux function. There are two types of flux functions that give rise to different solution behaviors, especially involving solution structures containing shocks and/or rarefaction waves:

1. $f(u)$ is a convex function – i.e., $f''(u) > 0, \forall u$ (or, equally well, f is concave with $f''(u) < 0, \forall u$): e.g., the Burger’s equation, the Euler equations, the Navier-Stokes equations.
2. $f(u)$ is a non-convex function: e.g., the Buckley-Leverett equation, magnetohydrodynamics (MHD) equations.

Remark: Later, we will see that the “convexity” assumption in the nonlinear scalar equation corresponds to a “genuinely nonlinearity” assumption for systems of equations.

Definition: If we rewrite Eq. 3.16 in nonconservation form, we get $u_t + f'(u)f_x(u) = 0$. The derivative of the flux function

$$\lambda(u) = f'(u) = \frac{df}{du} \quad (3.17)$$

is called the *characteristic speed*.

Remark:

1. In the system case,

$$0 = \mathbf{U}_t + \mathbf{F}(\mathbf{U})_x = \mathbf{U}_t + \frac{\partial \mathbf{F}}{\partial \mathbf{U}} \frac{\partial \mathbf{U}}{\partial x}, \quad (3.18)$$

the characteristic speed corresponds to the eigenvalues of the Jacobian matrix $\frac{\partial \mathbf{F}}{\partial \mathbf{U}}$.

2. For the linear scalar advection case, we already saw that $\frac{df}{du} = a$.

By far the most famous and popular example in the nonlinear scalar equations is Burgers' equation, in which the flux function is given as

$$f(u) = \frac{u^2}{2}, \quad (3.19)$$

hence resulting in the equation in the nonconservation form as

$$u_t + uu_x = 0. \quad (3.20)$$

We now take a look at the its mathematical properties from two different perspectives: (i) for small t , and (ii) for large t .

2.1. Burgers's equation for small t_s

Let's first assume that the initial data $u_0(x)$ is smooth and no singularity is observed for $0 < t \leq t_s$. In this case, we can conveniently follow characteristics

$$x'(t) = u(x(t), t) \quad (3.21)$$

along which the solution $u(x, t)$ is constant, since

$$\frac{d}{dt}u(x(t), t) = \frac{\partial}{\partial t}u(x(t), t) + \frac{\partial}{\partial x}u(x(t), t)x'(t) = u_t + uu_x = 0. \quad (3.22)$$

This also tells us that the slope $x'(t)$ is constant, and so the characteristics are straight lines, determined by the initial data. See Fig. 4.

Therefore, if the initial data $u_0 = u(\xi, 0)$ is smooth, and if t_s is chosen small enough so that the characteristics do not cross each others, we can solve the equation

$$x = \xi + u(\xi, 0)t_s \quad (3.23)$$

for ξ , and thus we obtain a well-defined solution

$$u(x, t_s) = u(x - u(\xi, 0)t_s, 0). \quad (3.24)$$

2.2. Burgers's equation for large t_b : Shock formation

For large $t = t_b$ at or after which the characteristics cross, Eq. 3.23 may not have a unique solution. This indeed will occur if $u'_0(\xi) < 0$ at any point ξ – that is, if $u_0(\xi)$ is a monotone decreasing function of ξ then the characteristics $x(t) = \xi(t) + u_0(\xi(t))t$ eventually cross at some finite time $t = t_b$ at which the wave will break and develop into a shock. When this first happens at $t = t_b$, the function $u(x, t)$ has an infinite slope, beyond which there is no classical solution of the PDE, and the (weak) solution becomes discontinuous. See Fig. 5.

Homework 2 Given a smooth initial data $u_0(\xi)$ for Burgers' equation with its slope $u'_0(\xi) < 0$ at some point ξ_0 . Show that the wave break time t_b is written as

$$t_b = \frac{-1}{u'_0(\xi_0)}. \quad (3.25)$$

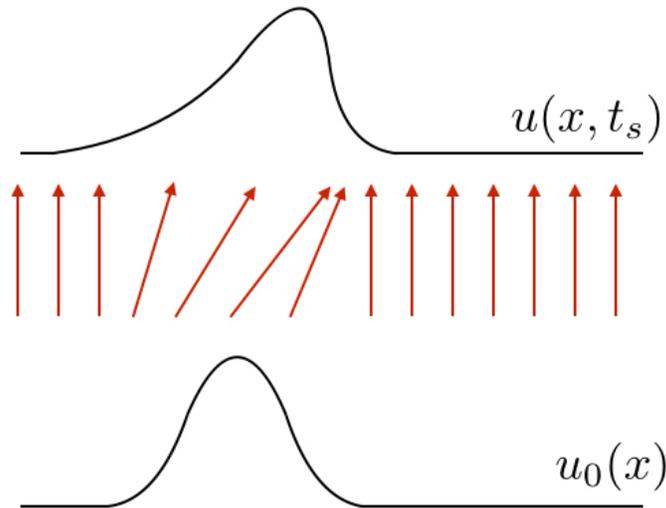


Figure 4. Characteristics and solution for Burgers' equation for small $t = t_s$.

Recall that in the case of the linear scalar advection, in which the characteristic speed is constant, $df/du = a$, the solution is simply a translated form of the initial data with speed a without any distortion (see Fig. 1). In the nonlinear case the characteristic speed is a function of the solution $u(x, t)$ itself – e.g., $df/du = u$ for Burgers' equation, therefore, distortions are inevitably produced. This is a distinctive feature of nonlinear problem.

To see the wave distortion phenomenon – also referred to as ‘the wave steepening’ – we refer to the initial condition $u_0(x)$ depicted as in Figs. 4 & 5. First, note that the flux function for Burgers' equation is convex (i.e., $f''(u) = 1 > 0$), and therefore, its characteristic speed (i.e., $f'(u) = u$) is an increasing function of u – the characteristic speed of Burgers' equation is u itself. The behavior of the characteristic speed therefore depends on the behavior of u . Specifically, the initial characteristics $x_m(t)$ emanating from the initial points \bar{x}_m , $m = 1, \dots, p$ have the form (see also Fig. 3)

$$x_m(t) = \bar{x}_m + u_0(\bar{x}_m)t. \quad (3.26)$$

We see that depending on how $u_0(x)$ increases or decreases as a function of x , the initial characteristic speeds vary, and the characteristic curves can cross. We can think of two intervals I_e and I_c on the x -axis where distortions are more evident. See Fig. 6 for an illustration. If we let \bar{x}_0 to be a point where $u'_0(\bar{x}_0) = 0$ (i.e., \bar{x}_0 is a local maximum point of u_0), and take

$$I_e = [\bar{x}_+, \bar{x}_0], I_c = [\bar{x}_0, \bar{x}_-], \quad (3.27)$$

where \bar{x}_+ and \bar{x}_- are the points where u_0 starts to increase and stops to decrease as x , respectively, as shown in Fig. 6. We say that I_e is an expansive region

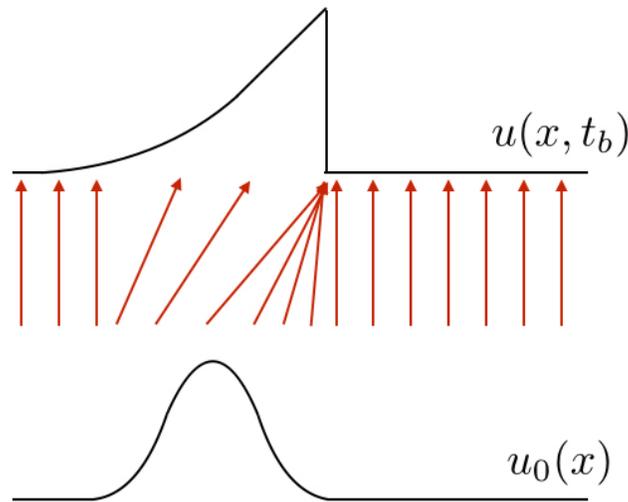


Figure 5. Characteristics and solution for Burgers' equation for large $t = t_b$. The characteristics cross and a shock is formed as a result.

where the characteristic speed keeps increasing as x increases. On the contrary, I_c is a compressive region where the characteristic speed decreases with x . It is easy to see that the characteristics from I_e and I_c will eventually cross each others, generating a sharp discontinuous profile of $u(x, t)$ for $t > t_b$, although the initial data $u_0(x)$ was smooth to begin with.

For times $t > t_b$ some of the characteristics have crossed. When this happens, there are points x where there are three characteristics leading back to $t = 0$. The solution u at such a time is a triple-valued function as seen in Fig. 7. Although there exist some cases that this makes sense, such as a breaking ocean wave modeled by the shallow water equations, in most physical situations, this doesn't make sense. For instance, the density of a gas cannot be triple valued at a given point.

As seen in **Homework 1**, one way to determine the correct physical behavior can be achieved by adopting the vanishing discontinuity approach. There is yet another approach that results in a differential *integral* formulation that is often more convenient to work with. This approach is available by considering so-called the *weak solutions* and this is discussed in the next section in more detail.

2.3. Weak solutions

In order to successfully seek for physically meaningful solutions $u(x, t)$ of PDEs that are relevant to various physical phenomena, it would be much more desirable if we can relax those mathematical constraints on smoothness in $u(x, t)$.

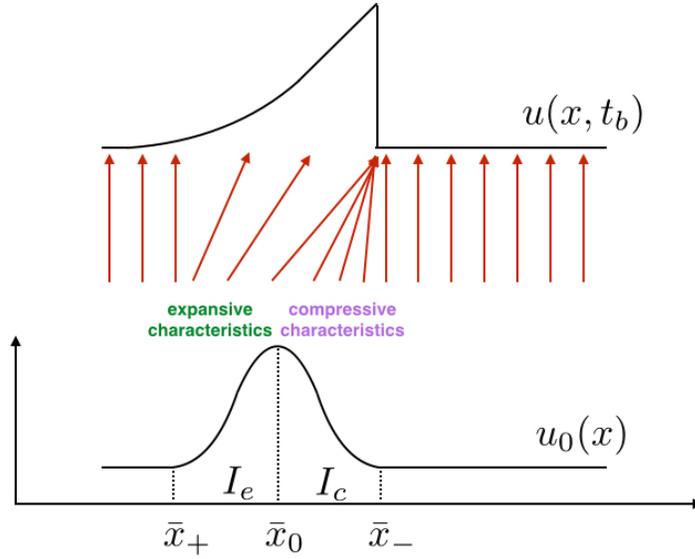


Figure 6. Characteristics crossing for Burgers' equation for large $t > t_b$.

In other words, we wish to come up with a mathematical technique that can be applied more generally to rewrite a differential equation in a form where less regularity is required to define a 'solution'. The weak solution approach is, in that sense, is one such technique we are now considering. The basic idea is to take the PDE, multiply by a *smooth* "test function", integrate one or more times over some domain, and then use integration by parts to move derivatives off the function u and onto the smooth test function. The outcome is an equation involving fewer derivatives on u , and hence requiring less smoothness.

Definition: The function $u(x, t)$ is called a *weak solution* of the scalar conservation law $u_t + f_x = 0$ if it satisfies the following condition for all test functions $\phi(x, t) \in C_0^1(\mathbb{R} \times \mathbb{R}^+)$:

$$\int_{\mathbb{R}^+} \int_{\mathbb{R}} [\phi_t u + \phi_x f(u)] dx dt = - \int_{\mathbb{R}} \phi(x, 0) u(x, 0) dx. \quad (3.28)$$

Note: C_0^1 is the space of functions that are continuously differentiable (C^1) with compact support.

Note: $f \in C_0(\mathbb{R})$ iff $f = 0$ in outside of some bounded sets and the support of f lies in a compact set. The support of f , $\text{supp}(f) = \{x \in X : f(x) \neq 0\}$.

Remark: One can obtain Eq. 3.28 by multiplying ϕ to $u_t + f_x = 0$ and then integrate over space and time,

$$\int_{\mathbb{R}^+} \int_{\mathbb{R}} [\phi u_t + \phi f(u)_x] dx dt = 0. \quad (3.29)$$

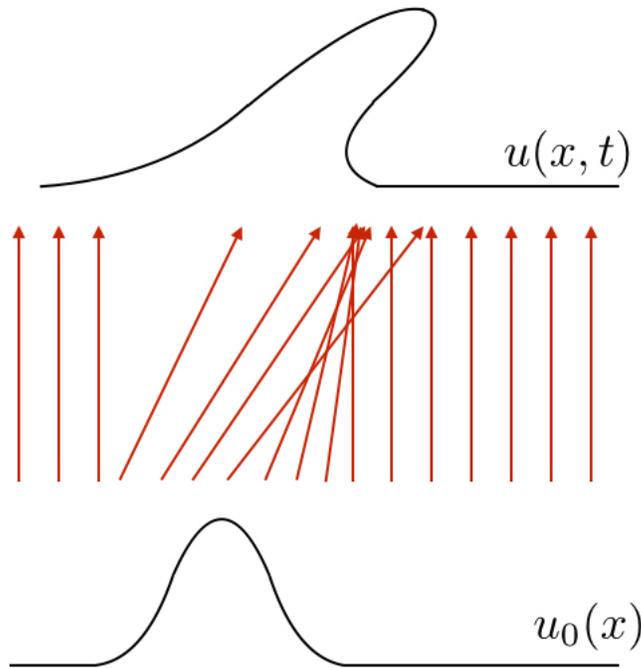


Figure 7. Triple-valued solution to Burgers' equation for large $t > t_b$.

Finally, integrating Eq. 3.29 by part gives the definition of a weak solution in Eq. 3.28. Notice that nearly all the boundary terms which normally arise through integration by parts drop out because ϕ has compact support, hence becomes zero outside of some bounded region of the x - t plane. The RHS in Eq. 3.28 bears the initial conditions of the PDE which cannot be ignored in the weak formulation.

Quick summary: A nice feature of Eq. 3.28 is that the derivatives are on the smooth test function ϕ , and no longer on u and $f(u)$. This enables Eq. 3.28 to take some discontinuous u as a solution in this weak sense.

Remark: If u is a weak solution, then u also satisfies the original integral conservation law, and vice versa.

Remark: With the help of weak solutions, can we say we are now happy about solving nonlinear scalar conservation laws? The answer is not really yet, unfortunately. One of the reasons is that weak solutions are often not unique and therefore, we need some criteria to choose a physically correct weak solution among choices. To do this, we will consider a condition called the '*entropy condition*' at the end of this chapter.

2.4. The Riemann problem

The conservation law together with piecewise constant data separated by a single discontinuity is known as the Riemann problem (RP). There are two physically admissible types of solutions, (i) shock solution, and (ii) rarefaction solution, which we will consider here in detail.

The RP involves a PDE with piecewise constant initial data,

$$u(x, 0) = \begin{cases} u_l & \text{if } x < 0 \\ u_r & \text{if } x > 0 \end{cases} \quad (3.30)$$

and the form of the solution, as will be shown, closely depends on the relation between u_l and u_r .

- **Case I:** $u_l > u_r$ In this case, there is a unique weak solution

$$u(x, t) = \begin{cases} u_l & \text{if } x < st \\ u_r & \text{if } x > st \end{cases} \quad (3.31)$$

where s is a shock speed, the speed at which the discontinuity travels. We are going to study how to compute a general expression for the shock speed in the next section. The characteristics in each left and right regions where u is constant (i.e., either u_l or u_r) go *into* the shock as time advances. See Fig. 8 and Fig. 9.

Note: We note for the RP with a shock solution, the characteristic speeds $f'(u)$ satisfy the following converging characteristic condition:

$$f'(u_l) > s > f'(u_r) \quad (3.32)$$

where s is a shock speed.

- **Case II:** $u_l < u_r$ In this case there are infinitely many weak solutions, therefore, we need to choose a physically correct weak solution. Our first attempt is to apply the exact same idea as in **Case I** in which the discontinuity propagates with speed s . This now allows the characteristics *go out* of the shock as illustrated in Fig. 10. This type of weak solution is called the entropy violating solution and needs be rejected. One crucial reason for rejecting this solution as a physical solution is because the solution is not stable to perturbation (also recall the three requirements for well-posed PDEs we studied in Chapter 2). This means that small perturbations of the initial data lead to large changes in the solution. For example, if the data is smeared out little bit, or if a small amount of viscosity is added to the equation, the solution changes completely.

Another weak solution is the rarefaction wave

$$u(x, t) = \begin{cases} u_l & \text{if } x < u_l t \\ x/t & \text{if } u_l t \leq x \leq u_r t \\ u_r & \text{if } x > u_r t \end{cases} \quad (3.33)$$

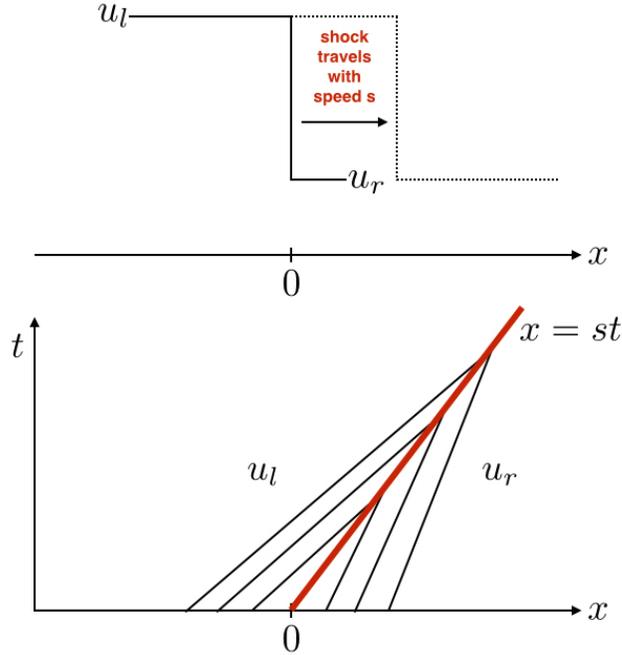


Figure 8. Weak solution of shock wave to the Riemann problem $u_l > u_r$.

This solution is stable to perturbation and is in fact the physically correct weak solution satisfying the vanishing viscosity approach.

Note: We note for the RP with a rarefaction solution, the characteristic speeds $f'(u)$ satisfy the following diverging characteristic condition:

$$f'(u_l) < f'(u_r). \quad (3.34)$$

Remark: Before proceeding to the next section, we briefly study four variants of the integral form of conservation laws $u_t + f_x = 0$. Recalled that we already have studied this in Chapter 1. This time, we choose a control volume $\mathcal{V} = [x_L, x_R] \times [t_1, t_2]$ on the $x-t$ plane.

1. *Integral form I:*

$$\frac{d}{dt} \int_{x_L}^{x_R} u(x, t) dx = f(u(x_L, t)) - f(u(x_R, t)) \quad (3.35)$$

2. *Integral form II:* Integrating *Integral form I* in time gives

$$\int_{x_L}^{x_R} u(x, t_2) dx - \int_{x_L}^{x_R} u(x, t_1) dx = \int_{t_1}^{t_2} f(u(x_L, t)) dt - \int_{t_1}^{t_2} f(u(x_R, t)) dt \quad (3.36)$$

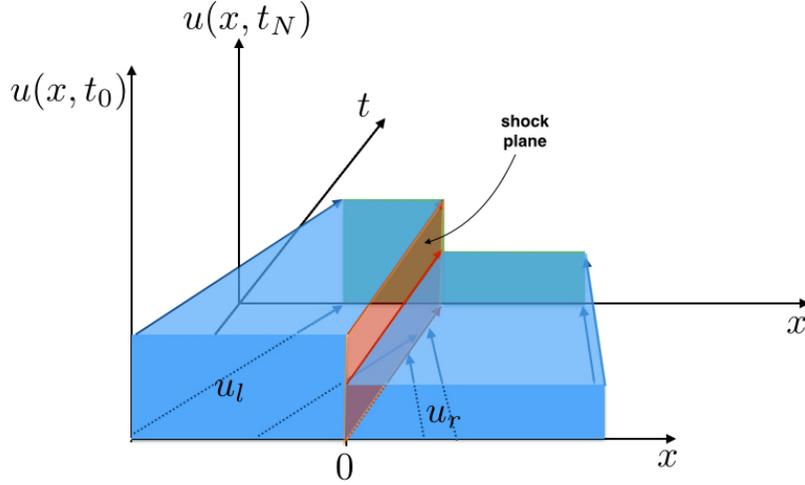


Figure 9. Weak solution of shock wave to the Riemann problem $u_l > u_r$. The characteristic curves are drawn in blue in the x - t plane. The dark orange shaded plane is the shock plane due to the crossing of the characteristics from the two discontinuous initial data u_l and u_r . The shock plane travels with the shock speed s which will be studied by considering the Rankine-Hugoniot jump condition in the next section.

3. *Integral form III:* Integrating $u_t + f_x = 0$ in any domain \mathcal{V} in the x - t plane and using Green's theorem, we obtain

$$\oint_{\partial\mathcal{V}} [udx - f(u)dt] = 0 \quad (3.37)$$

4. *Integral form IV:* The last variant is the integral relation that the weak or generalized solution u satisfies (see also Eq. 3.28) for all test function $\phi(x, t) \in C_0^1(\mathbb{R} \times \mathbb{R}^+)$:

$$\int_{\mathbb{R}^+} \int_{\mathbb{R}} [\phi_t u + \phi_x f(u)] dx dt = - \int_{\mathbb{R}} \phi(x, 0) u(x, 0) dx. \quad (3.38)$$

2.5. Shock speed: the Rankine-Hugoniot jump condition

The propagating shock solution in Eq. 3.31 is a weak solution only with a proper value of the shock speed s . In fact, a correct shock speed s can be determined by considering conservation – called the Rankine-Hugoniot jump condition.

Consider a solution $u(x, t)$ such that $u(x, t)$ and $f(u)$ and their derivatives are continuous everywhere except on a line $S = S(t)$ on the x - t plane across which $u(x, t)$ has a jump discontinuity. Choose two fixed points x_L and x_R such that $x_L < S(t) < x_R$. Adopting *Integral form I* on the control volume $[x_L, x_R]$,

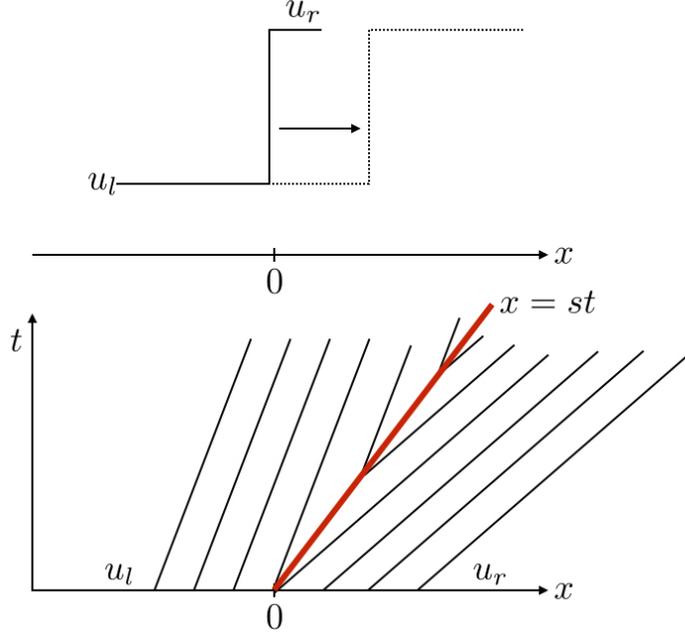


Figure 10. Entropy-violating shock and should be rejected.

we have

$$f(u(x_L, t)) - f(u(x_R, t)) = \frac{d}{dt} \int_{x_L}^{S(t)} u(x, t) dx + \frac{d}{dt} \int_{S(t)}^{x_R} u(x, t) dx, \quad (3.39)$$

which becomes

$$f(u(x_L, t)) - f(u(x_R, t)) = (u(S_L, t) - u(S_R, t)) \frac{dS}{dt} + \int_{x_L}^{S(t)} u_t(x, t) dx + \int_{S(t)}^{x_R} u_t(x, t) dx, \quad (3.40)$$

where

$$u(S_L, t) = \lim_{x \uparrow S(t)} u(x, t), \quad (3.41)$$

$$u(S_R, t) = \lim_{x \downarrow S(t)} u(x, t). \quad (3.42)$$

Note the two integrals in Eq. 3.40 become

$$\int_{x_L}^{S(t)} u_t(x, t) dx = - \int_{x_L}^{S(t)} f_x(u(x, t)) dx = f(u(x_L, t)) - f(u(S_L, t)), \quad (3.43)$$

$$\int_{S(t)}^{x_R} u_t(x, t) dx = - \int_{S(t)}^{x_R} f_x(u(x, t)) dx = f(u(S_R, t)) - f(u(x_R, t)). \quad (3.44)$$

After canceling $f(u(x_L, t)) - f(u(x_R, t))$ from both sides, we finally obtain

$$f(u(S_L, t)) - f(u(S_R, t)) = (u(S_L, t) - u(S_R, t))s, \quad (3.45)$$

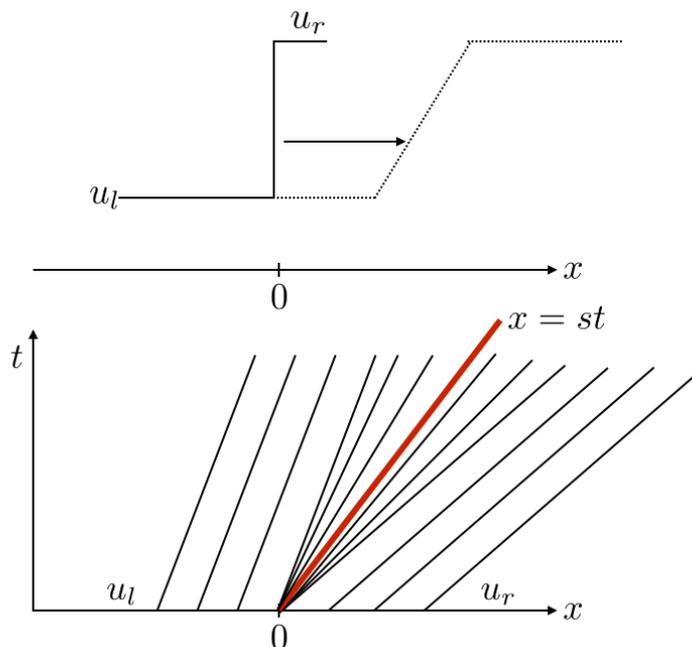


Figure 11. Entropy satisfying weak solution – the rarefaction wave.

where we introduced $s = dS/dt$ the speed of the discontinuity.

Definition: The relation in Eq. 3.45 is called the *Rankine-Hugoniot jump condition* (RH condition) and it provides a relation between the shock speed s and the states $u_l = u(S_L, t)$ and $u_r = u(S_R, t)$. We often denote shock speed s in the RH condition using brackets as follow:

$$s = \frac{[f]}{[u]} \equiv \frac{\lim_{x \downarrow S(t)} f(u(x, t), t) - \lim_{x \uparrow S(t)} f(u(x, t), t)}{\lim_{x \downarrow S(t)} u(x, t) - \lim_{x \uparrow S(t)} u(x, t)}. \quad (3.46)$$

Homework 3 Consider Burgers' equation

$$u_t + \left(\frac{u^2}{2}\right)_x = 0 \quad (3.47)$$

(a) By multiplying the equation by $2u$, show that you can derive a new conservation law for u^2 . What is the new flux function?

(b) Show that the original Burgers' equation and the new derived equation have different weak solutions (Hint: It suffices to show that there exist two different shock speeds from the two equations for the Riemann problem with $u_l > u_r$).

2.6. Entropy conditions

As demonstrated in **Homework 3** above, there are situations in which the weak solution is not unique. It is therefore natural to ask for an additional condition

to pick out the physically relevant solution. Recall that we've already seen there is an obvious condition for the characteristic speeds in Eq. 3.32. A shock should have characteristics *going into* the shock as time evolves. We are now ready to state it and call it the entropy condition:

Definition: A discontinuity propagating with speed s given by Eq. 3.45 (or equivalently, Eq. 3.46) – that is, the two data states u_l and u_r are connected through a single discontinuity with its speed s – satisfies the *entropy condition* if

$$f'(u_l) > s > f'(u_r), \quad (3.48)$$

or equivalently,

$$\lambda(u_l) > s > \lambda(u_r). \quad (3.49)$$

Remark: On the other hand, if the two data states u_l and u_r are connected through a smooth transition – i.e., rarefaction wave – the divergence relation of the characteristics holds:

$$f'(u_l) < f'(u_r) \quad (3.50)$$

or equivalently,

$$\lambda(u_l) < \lambda(u_r). \quad (3.51)$$

Example: Let's consider Burgers' equation on \mathbb{R} with the following initial conditions:

$$u(x, 0) = \begin{cases} 0 & \text{if } x < 0 \\ 1 & \text{if } x > 0 \end{cases} \quad (3.52)$$

We can first try to obtain an entropy violating solution, and we know that this solution needs to be rejected anyway as it is ill-posed. But we are going to find this solution to practice what we already learned in this chapter. If we apply the RH condition to this problem – which *is wrong to do so* – to compute the shock speed s , we get

$$s = \frac{f(u_r) - f(u_l)}{u_r - u_l} = \frac{1}{2}. \quad (3.53)$$

This results in the following entropy violating self-similar solution

$$u(x, t) = \begin{cases} 0 & \text{if } \frac{x}{t} < \frac{1}{2} \\ 1 & \text{if } \frac{x}{t} > \frac{1}{2} \end{cases} \quad (3.54)$$

which is shown in Fig.12.

Let us try again to get the correct weak solution this time. Consider the following self-similar solution

$$u(x, t) = \begin{cases} 0 & \text{if } \frac{x}{t} < 0 \\ x/t & \text{if } 0 < \frac{x}{t} < 1 \\ 1 & \text{if } \frac{x}{t} > 1. \end{cases} \quad (3.55)$$

We can check that the wave diagram for this solution is plotted in Fig. 13. It is also easy to check if this solution, especially the part in the expansion region,

satisfies the Burgers' equation. To see this,

$$\frac{\partial u}{\partial t} + u \frac{\partial u}{\partial x} = \frac{\partial}{\partial t} \left(\frac{x}{t} \right) + \frac{x}{t} \frac{\partial}{\partial x} \left(\frac{x}{t} \right) = -\frac{x}{t^2} + \frac{x}{t} \frac{1}{t} = 0. \quad (3.56)$$

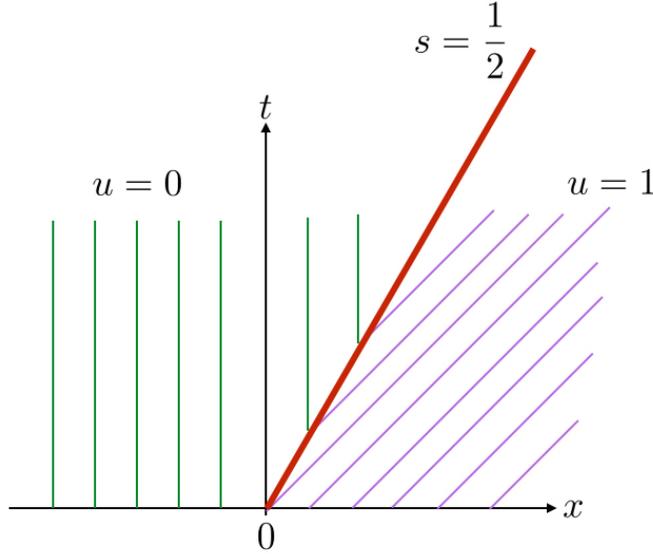


Figure 12. Wave diagram for the wrong entropy violating weak solution.

Example: Let's consider Burgers' equation on \mathbb{R} with the following initial conditions for $t \leq 4/3$:

$$u(x, 0) = \begin{cases} 1 & \text{if } |x| < 1/3 \\ 0 & \text{if } |x| > 1/3 \end{cases} \quad (3.57)$$

We see that the jump at $x = -1/3$ creates a rarefaction wave solution; the jump at $x = 1/3$ creates a shock solution. For $t \leq 4/3$ the shock and the rarefaction fan do not intersect each other and therefore, we can seek for the exact piecewise-linear solution as follows.

Let us first compute the shock speed using RH with $u_l = 1$ and $u_r = 0$:

$$s = \frac{f(u_r) - f(u_l)}{u_r - u_l} = \frac{1}{2}, \quad (3.58)$$

which gives the characteristic curve (the red thick line in Fig. 14) for shock $x - 1/2t = 1/3$.

We also consider the first characteristic curve right next to the rarefaction region – this is the left most purple line in Fig. 14. Since the characteristic slope is $f'(u)|_{u=1} = 1$, we obtain the relation $x - t = -1/3$. From these we easily obtain

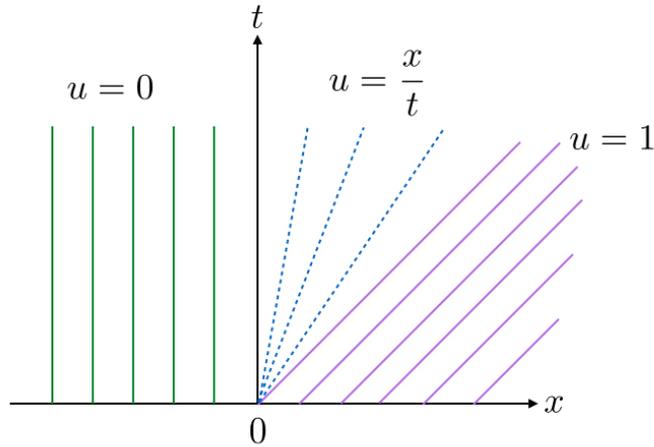


Figure 13. Wave diagram for the correct rarefaction wave weak solution.

the exact weak solution as follows:

$$u(x, t) = \begin{cases} 0 & \text{if } x < -\frac{1}{3} \\ \frac{x+1/3}{t} & \text{if } -\frac{1}{3} < x < t - \frac{1}{3} \\ 1 & \text{if } t - \frac{1}{3} < x < \frac{1}{2}t + \frac{1}{3} \\ 0 & \text{if } x > \frac{1}{2}t + \frac{1}{3} \end{cases} \quad (3.59)$$

Note that at $t = \frac{4}{3}$ we get $\frac{1}{2}t + \frac{1}{3} = t - \frac{1}{3}$, and as a result, the shock and the rarefaction solutions intersect for $t > \frac{4}{3}$.

Homework 4 Solve Burgers' equation on \mathbb{R} for small enough $t \leq t_b$ that allows the exact piecewise-linear weak solution with the following initial conditions:

$$u(x, 0) = \begin{cases} 2 & \text{if } |x| < 1/2 \\ -1 & \text{if } |x| > 1/2 \end{cases} \quad (3.60)$$

Find the time t_b when the two waves first intersect. Draw a wave diagram for the weak solution.

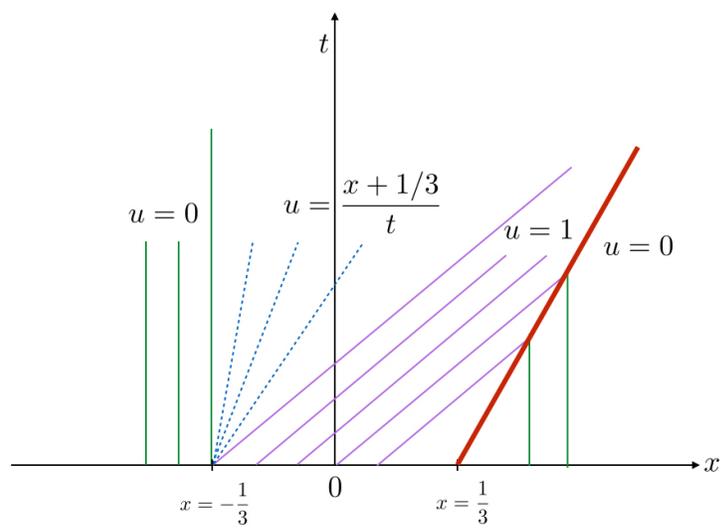


Figure 14. Wave diagram of the weak solution for $t \leq 4/3$.

Chapter 4

Discrete Numerical Approaches

We review several key ideas on numerical methods that discretize PDEs and provide approximated solutions to numerical PDE models derived from the analytical PDEs. Three major methods are briefly described along with the principal advantages and drawbacks in each method. Three solution schemes include:

- Finite difference method (FD)
- Finite volume method (FV)
- Finite element method (FE)

Some other approaches also used in many CFD applications include:

- Discontinuous Galerkin (DG) (or, discontinuous FE as compared to the standard ‘continuous’ FE)
- Spectral element (SE)

In general, a proper choice of numerical approaches strongly depends on various components of your problem, including especially the following factors:

- Flow regimes – e.g., compressible (FV) vs. incompressible (FD), high Mach number (FV) vs. low Mach number (low Mach number scheme), turbulent (subgrid models) vs. laminar (boundary layer), advection dominated (FV, FD, DG) vs. diffusion dominated (FE)
- Physics of flows – e.g., macroscopic (fluid models: FV, FD, FE) vs. microscopic (kinetic models: PIC – particle-in-cell), hydrodynamics vs. magnetohydrodynamics vs. rad-hydro, single-fluid (single bulk velocity) vs. multi-fluid (multiple bulk velocity), advection dominated (explicit) vs. diffusion dominated (implicit) vs. combined (explicit & implicit via operator split), gravitational flow (elliptic solver)
- Geometry of flows – e.g., rectangular domain (FD) vs. engineering flow (complicated physical boundaries such as bridges, airplane, airfoils, cars, buildings – mostly FE, but also FV), localized dynamics (AMR – adaptive mesh refinements; stretched grid) vs. global dynamics (UG – uniform grid)

- Numerical issues – ease of high-order implementation (FD, FE, DG) vs. difficulty in high-order implementation (FV), ease of multi-dimensional extension (FD) vs. difficulty in multi-dimensional extension (FV)

Our primary interest in this course lies in studying the first two methods, FD and FV. Later, we are going to use FD and FV approaches to solve linear advection equation and linear hyperbolic systems. Such fundamental ideas of solving linear hyperbolic PDEs will be extended to the nonlinear cases.

For the rest of the study in this chapter, we are going to refer to a short article by Joaquim Peiró and Spencer Sherwin which provides a nice overview and comparison of three discrete finite approaches, FD, FE, and FV:

- “Finite difference, finite element and finite volume methods for partial differential equations” (enclosed below).

8.2

FINITE DIFFERENCE, FINITE ELEMENT AND FINITE VOLUME METHODS FOR PARTIAL DIFFERENTIAL EQUATIONS

Joaquim Peiró and Spencer Sherwin

Department of Aeronautics, Imperial College, London, UK

There are three important steps in the computational modelling of any physical process: (i) problem definition, (ii) mathematical model, and (iii) computer simulation.

The first natural step is to define an idealization of our problem of interest in terms of a set of relevant quantities which we would like to measure. In defining this idealization we expect to obtain a well-posed problem, this is one that has a unique solution for a given set of parameters. It might not always be possible to guarantee the fidelity of the idealization since, in some instances, the physical process is not totally understood. An example is the complex environment within a nuclear reactor where obtaining measurements is difficult.

The second step of the modeling process is to represent our idealization of the physical reality by a mathematical model: the governing equations of the problem. These are available for many physical phenomena. For example, in fluid dynamics the Navier–Stokes equations are considered to be an accurate representation of the fluid motion. Analogously, the equations of elasticity in structural mechanics govern the deformation of a solid object due to applied external forces. These are complex general equations that are very difficult to solve both analytically and computationally. Therefore, we need to introduce simplifying assumptions to reduce the complexity of the mathematical model and make it amenable to either exact or numerical solution. For example, the irrotational (without vorticity) flow of an incompressible fluid is accurately represented by the Navier–Stokes equations but, if the effects of fluid viscosity are small, then Laplace’s equation of *potential flow* is a far more efficient description of the problem.

After the selection of an appropriate mathematical model, together with suitable boundary and initial conditions, we can proceed to its solution. In this chapter we will consider the numerical solution of mathematical problems which are described by partial differential equations (PDEs). The three classical choices for the numerical solution of PDEs are the finite difference method (FDM), the finite element method (FEM) and the finite volume method (FVM).

The FDM is the oldest and is based upon the application of a local Taylor expansion to approximate the differential equations. The FDM uses a topologically square network of lines to construct the discretization of the PDE. This is a potential bottleneck of the method when handling complex geometries in multiple dimensions. This issue motivated the use of an integral form of the PDEs and subsequently the development of the finite element and finite volume techniques.

To provide a short introduction to these techniques we shall consider each type of discretization as applied to one-dimensional PDEs. This will not allow us to illustrate the geometric flexibility of the FEM and the FVM to their full extent, but we will be able to demonstrate some of the similarities between the methods and thereby highlight some of the relative advantages and disadvantages of each approach. For a more detailed understanding of the approaches we refer the reader to the section on suggested reading at the end of the chapter.

The section is structured as follows. We start by introducing the concept of conservation laws and their differential representation as PDEs and the alternative integral forms. We next discuss the classification of partial differential equations: elliptic, parabolic, and hyperbolic. This classification is important since the type of PDE dictates the form of boundary and initial conditions required for the problem to be well-posed. It also, permits in some cases, e.g., in hyperbolic equations, to identify suitable schemes to discretise the differential operators. The three types of discretisation: FDM, FEM and FVM are then discussed and applied to different types of PDEs. We then end our overview by discussing the numerical difficulties which can arise in the numerical solution of the different types of PDEs using the FDM and provides an introduction to the assessment of the stability of numerical schemes using a Fourier or Von Neumann analysis.

Finally we note that, given the scientific background of the authors, the presentation has a bias towards fluid dynamics. However, we stress that the fundamental concepts presented in this chapter are generally applicable to continuum mechanics, both solids and fluids.

1. Conservation Laws: Integral and Differential Forms

The governing equations of continuum mechanics representing the kinematic and mechanical behaviour of general bodies are commonly referred

to as *conservation laws*. These are derived by invoking the conservation of mass and energy and the momentum equation (Newton's law). Whilst they are equally applicable to solids and fluids, their differing behaviour is accounted for through the use of a different constitutive equation.

The general principle behind the derivation of conservation laws is that the rate of change of $u(\mathbf{x}, t)$ within a volume V plus the flux of u through the boundary A is equal to the rate of production of u denoted by $S(u, \mathbf{x}, t)$. This can be written as

$$\frac{\partial}{\partial t} \int_V u(\mathbf{x}, t) dV + \int_A \mathbf{f}(u) \cdot \mathbf{n} dA - \int_V S(u, \mathbf{x}, t) dV = 0 \quad (1)$$

which is referred to as the *integral* form of the conservation law. For a fixed (independent of t) volume and, under suitable conditions of smoothness of the intervening quantities, we can apply Gauss' theorem

$$\int_V \nabla \cdot \mathbf{f} dV = \int_A \mathbf{f} \cdot \mathbf{n} dA$$

to obtain

$$\int_V \left(\frac{\partial u}{\partial t} + \nabla \cdot \mathbf{f}(u) - S \right) dV = 0. \quad (2)$$

For the integral expression to be zero for any volume V , the integrand must be zero. This results in the *strong* or differential form of the equation

$$\frac{\partial u}{\partial t} + \nabla \cdot \mathbf{f}(u) - S = 0. \quad (3)$$

An alternative *integral* form can be obtained by the method of weighted residuals. Multiplying Eq. (3) by a *weight* function $w(\mathbf{x})$ and integrating over the volume V we obtain

$$\int_V \left(\frac{\partial u}{\partial t} + \nabla \cdot \mathbf{f}(u) - S \right) w(\mathbf{x}) dV = 0. \quad (4)$$

If Eq. (4) is satisfied for any weight function $w(\mathbf{x})$, then Eq. (4) is equivalent to the differential form (3). The smoothness requirements on \mathbf{f} can be relaxed by applying the Gauss' theorem to Eq. (4) to obtain

$$\int_V \left[\left(\frac{\partial u}{\partial t} - S \right) w(\mathbf{x}) - \mathbf{f}(u) \cdot \nabla w(\mathbf{x}) \right] dV + \int_A \mathbf{f} \cdot \mathbf{n} w(\mathbf{x}) dA = 0. \quad (5)$$

This is known as the *weak* form of the conservation law.

Although the above formulation is more commonly used in fluid mechanics, similar formulations are also applied in structural mechanics. For instance, the well-known principle of virtual work for the static equilibrium of a body [1], is given by

$$\delta W = \int_V (\nabla \boldsymbol{\sigma} + \mathbf{f}) \cdot \delta \mathbf{v} \, dV = 0$$

where δW denotes the virtual work done by an arbitrary virtual velocity $\delta \mathbf{v}$, $\boldsymbol{\sigma}$ is the stress tensor and \mathbf{f} denotes the body force. The similarity with the method of weighted residuals (4) is evident.

2. Model Equations and their Classification

In the following we will restrict ourselves to the analysis of one-dimensional conservation laws representing the transport of a scalar variable $u(x, t)$ defined in the domain $\Omega = \{x, t : 0 \leq x \leq 1, 0 \leq t \leq T\}$. The convection–diffusion–reaction equation is given by

$$\mathcal{L}(u) = \frac{\partial u}{\partial t} + \frac{\partial}{\partial x} \left(au - b \frac{\partial u}{\partial x} \right) - r u = s \quad (6)$$

together with appropriate boundary conditions at $x=0$ and 1 to make the problem well-posed. In the above equation $\mathcal{L}(u)$ simply represents a linear differential operator. This equation can be recast in the form (3) with $f(u) = au - \partial u / \partial x$ and $S(u) = s + ru$. It is linear if the coefficient a, b, r and s are functions of x and t , and non-linear if any of them depends on the solution, u .

In what follows, we will use for convenience the convention that the presence of a subscript x or t under an expression indicates a derivative or partial derivative with respect to this variable, for example

$$u_x(x) = \frac{du}{dx}(x); \quad u_t(x, t) = \frac{\partial u}{\partial t}(x, t); \quad u_{xx}(x, t) = \frac{\partial^2 u}{\partial x^2}(x, t).$$

Using this notation, Eq. (6) is re-written as

$$u_t + (au - bu_x)_x - ru = s.$$

2.1. Elliptic Equations

The steady-state solution of Eq. (6) when advection and source terms are neglected, i.e., $a=0$ and $s=0$, is a function of x only and satisfies the Helmholtz equation

$$(bu_x)_x + ru = 0. \quad (7)$$

This equation is elliptic and its solution depends on two families of integration constants that are fixed by prescribing boundary conditions at the ends of the domain. One can either prescribe Dirichlet boundary conditions at both ends, e.g., $u(0) = \alpha_0$ and $u(1) = \alpha_1$, or substitute one of them (or both if $r \neq 0$) by a Neumann boundary condition, e.g., $u_x(0) = g$. Here α_0 , α_1 and g are known constant values. We note that if we introduce a perturbation into a Dirichlet boundary condition, e.g., $u(0) = \alpha_0 + \epsilon$, we will observe an instantaneous modification to the solution throughout the domain. This is indicative of the elliptic nature of the problem.

2.2. Parabolic Equations

Taking $a = 0$, $r = 0$ and $s = 0$ in our model, Eq. (6) leads to the heat or diffusion equation

$$u_t - (b u_x)_x = 0, \quad (8)$$

which is parabolic. In addition to appropriate boundary conditions of the form used for elliptic equations, we also require an initial condition at $t = 0$ of the form $u(x, 0) = u_0(x)$ where u_0 is a given function.

If b is constant, this equation admits solutions of the form $u(x, t) = A e^{\beta t} \sin kx$ if $\beta + k^2 b = 0$. A notable feature of the solution is that it decays when b is positive as the exponent $\beta < 0$. The rate of decay is a function of b . The more diffusive the equation (i.e., larger b) the faster the decay of the solution is. In general the solution can be made up of many sine waves of different frequencies, i.e., a Fourier expansion of the form

$$u(x, t) = \sum_m A_m e^{\beta_m t} \sin k_m x,$$

where A_m and k_m represent the amplitude and the frequency of a Fourier mode, respectively. The decay of the solution depends on the Fourier contents of the initial data since $\beta_m = -k_m^2 b$. High frequencies decay at a faster rate than the low frequencies which physically means that the solution is being smoothed. This is illustrated in Fig. 1 which shows the time evolution of $u(x, t)$ for an initial condition $u_0(x) = 20x$ for $0 \leq x \leq 1/2$ and $u_0(x) = 20(1 - x)$ for $1/2 \leq x \leq 1$. The solution shows a rapid smoothing of the slope discontinuity of the initial condition at $x = 1/2$. The presence of a positive diffusion ($b > 0$) physically results in a smoothing of the solution which stabilizes it. On the other hand, negative diffusion ($b < 0$) is de-stabilizing but most physical problems have positive diffusion.

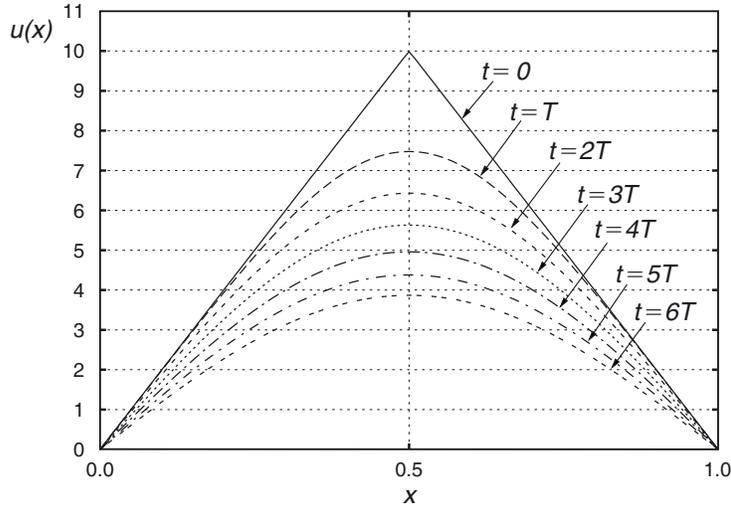


Figure 1. Rate of decay of the solution to the diffusion equation.

2.3. Hyperbolic Equations

A classic example of hyperbolic equation is the linear advection equation

$$u_t + a u_x = 0, \quad (9)$$

where a represents a constant velocity. The above equation is also clearly equivalent to Eq. (6) with $b = r = s = 0$. This hyperbolic equation also requires an initial condition, $u(x, 0) = u_0(x)$. The question of what boundary conditions are appropriate for this equation can be more easily be answered after considering its solution. It is easy to verify by substitution in (9) that the solution is given by $u(x, t) = u_0(x - at)$. This describes the propagation of the quantity $u(x, t)$ moving with speed “ a ” in the x -direction as depicted in Fig. 2. The solution is constant along the *characteristic line* $x - at = c$ with $u(x, t) = u_0(c)$.

From the knowledge of the solution, we can appreciate that for $a > 0$ a boundary condition should be prescribed at $x = 0$, (e.g., $u(0) = \alpha_0$) where information is being fed into the solution domain. The value of the solution at $x = 1$ is determined by the initial conditions or the boundary condition at $x = 0$ and cannot, therefore, be prescribed. This simple argument shows that, in a hyperbolic problem, the selection of appropriate conditions at a boundary point depends on the solution at that point. If the velocity is negative, the previous treatment of the boundary conditions is reversed.

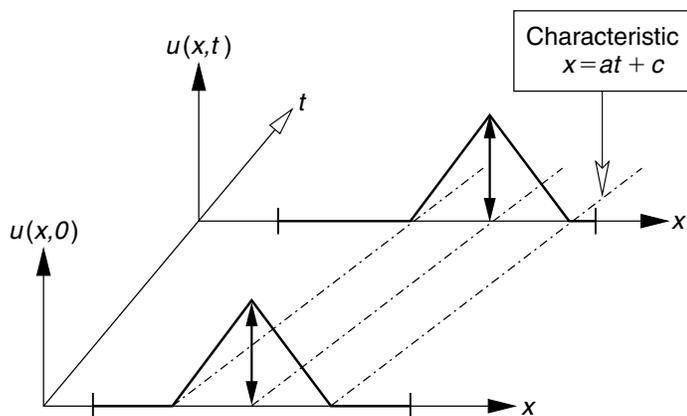


Figure 2. Solution of the linear advection equation.

The propagation velocity can also be a function of space, i.e., $a = a(x)$ or even the same as the quantity being propagated, i.e., $a = u(x, t)$. The choice $a = u(x, t)$ leads to the non-linear inviscid Burgers' equation

$$u_t + u u_x = 0. \quad (10)$$

An analogous analysis to that used for the advection equation shows that $u(x, t)$ is constant if we are moving with a local velocity also given by $u(x, t)$. This means that some regions of the solution advance faster than other regions leading to the formation of sharp gradients. This is illustrated in Fig. 3. The initial velocity is represented by a triangular “zig-zag” wave. Peaks and troughs in the solution will advance, in opposite directions, with maximum speed. This will eventually lead to an overlap as depicted by the dotted line in Fig. 3. This results in a non-uniqueness of the solution which is obviously non-physical and to resolve this problem we must allow for the formation and propagation of discontinuities when two characteristics intersect (see Ref. [2] for further details).

3. Numerical Schemes

There are many situations where obtaining an exact solution of a PDE is not possible and we have to resort to approximations in which the infinite set of values in the continuous solution is represented by a finite set of values referred to as the *discrete* solution.

For simplicity we consider first the case of a function of one variable $u(x)$. Given a set of points $x_i; i = 1, \dots, N$ in the domain of definition of $u(x)$, as

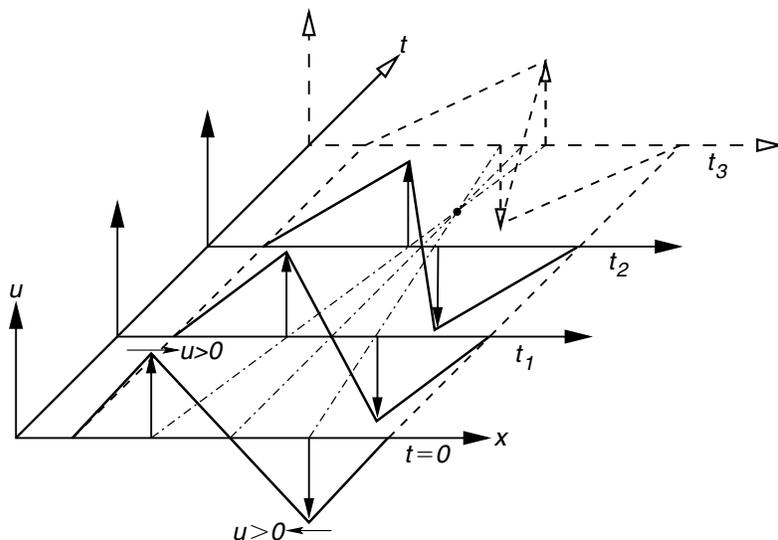


Figure 3. Formation of discontinuities in the Burgers' equation.

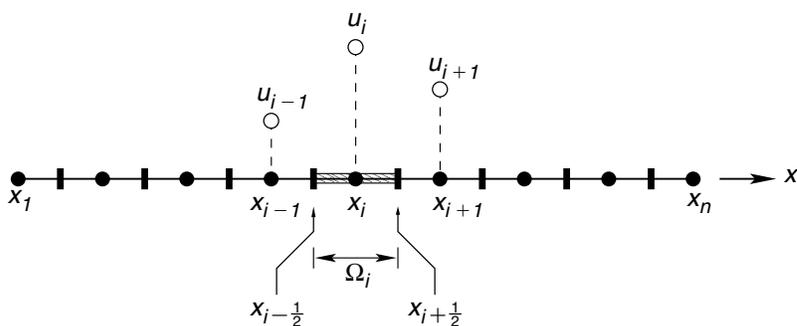


Figure 4. Discretization of the domain.

shown in Fig. 4, the numerical solution that we are seeking is represented by a discrete set of function values $\{u_1, \dots, u_N\}$ that approximate u at these points, i.e., $u_i \approx u(x_i)$; $i = 1, \dots, N$.

In what follows, and unless otherwise stated, we will assume that the points are equally spaced along the domain with a constant distance $\Delta x = x_{i+1} - x_i$; $i = 1, \dots, N - 1$. This way we will write $u_{i+1} \approx u(x_{i+1}) = u(x_i + \Delta x)$. This partition of the domain into smaller subdomains is referred to as a *mesh* or *grid*.

3.1. The Finite Difference Method (FDM)

This method is used to obtain numerical approximations of PDEs written in the strong form (3). The derivative of $u(x)$ with respect to x can be defined as

$$\begin{aligned} u_x|_i = u_x(x_i) &= \lim_{\Delta x \rightarrow 0} \frac{u(x_i + \Delta x) - u(x_i)}{\Delta x} \\ &= \lim_{\Delta x \rightarrow 0} \frac{u(x_i) - u(x_i - \Delta x)}{\Delta x} \\ &= \lim_{\Delta x \rightarrow 0} \frac{u(x_i + \Delta x) - u(x_i - \Delta x)}{2\Delta x}. \end{aligned} \quad (11)$$

All these expressions are mathematically equivalent, i.e., the approximation converges to the derivative as $\Delta x \rightarrow 0$. If Δx is small but finite, the various terms in Eq. (11) can be used to obtain approximations of the derivative u_x of the form

$$u_x|_i \approx \frac{u_{i+1} - u_i}{\Delta x} \quad (12)$$

$$u_x|_i \approx \frac{u_i - u_{i-1}}{\Delta x} \quad (13)$$

$$u_x|_i \approx \frac{u_{i+1} - u_{i-1}}{2\Delta x}. \quad (14)$$

The expressions (12)–(14) are referred to as forward, backward and centred finite difference approximations of $u_x|_i$, respectively. Obviously these approximations of the derivative are different.

3.1.1. Errors in the FDM

The analysis of these approximations is performed by using Taylor expansions around the point x_i . For instance an approximation to u_{i+1} using $n + 1$ terms of a Taylor expansion around x_i is given by

$$\begin{aligned} u_{i+1} &= u_i + u_x|_i \Delta x + u_{xx}|_i \frac{\Delta x^2}{2} + \cdots + \frac{d^n u}{dx^n} \Big|_i \frac{\Delta x^n}{n!} \\ &\quad + \underline{\frac{d^{n+1} u}{dx^{n+1}}(x^*) \frac{\Delta x^{n+1}}{(n+1)!}}. \end{aligned} \quad (15)$$

The underlined term is called the remainder with $x_i \leq x^* \leq x_{i+1}$, and represents the error in the approximation if only the first n terms in the expansion are kept. Although the expression (15) is exact, the position x^* is unknown.

To illustrate how this can be used to analyse finite difference approximations, consider the case of the forward difference approximation (12) and use the expansion (15) with $n = 1$ (two terms) to obtain

$$\frac{u_{i+1} - u_i}{\Delta x} = u_x|_i + \frac{\Delta x}{2} u_{xx}(x^*). \quad (16)$$

We can now write the approximation of the derivative as

$$u_x|_i = \frac{u_{i+1} - u_i}{\Delta x} + \epsilon_T \quad (17)$$

where ϵ_T is given by

$$\epsilon_T = -\frac{\Delta x}{2} u_{xx}(x^*). \quad (18)$$

The term ϵ_T is referred to as the *truncation error* and is defined as the difference between the exact value and its numerical approximation. This term depends on Δx but also on u and its derivatives. For instance, if $u(x)$ is a linear function then the finite difference approximation is exact and $\epsilon_T = 0$ since the second derivative is zero in (18).

The *order* of a finite difference approximation is defined as the power p such that $\lim_{\Delta x \rightarrow 0} (\epsilon_T / \Delta x^p) = \gamma \neq 0$, where γ is a finite value. This is often written as $\epsilon_T = O(\Delta x^p)$. For instance, for the forward difference approximation (12), we have $\epsilon_T = O(\Delta x)$ and it is said to be first-order accurate ($p = 1$).

If we apply this method to the backward and centred finite difference approximations (13) and (14), respectively, we find that, for constant Δx , their errors are

$$u_x|_i = \frac{u_i - u_{i-1}}{\Delta x} + \frac{\Delta x}{2} u_{xx}(x^*) \Rightarrow \epsilon_T = O(\Delta x) \quad (19)$$

$$u_x|_i = \frac{u_{i+1} - u_{i-1}}{2\Delta x} - \frac{\Delta x^2}{12} u_{xxx}(x^*) \Rightarrow \epsilon_T = O(\Delta x^2) \quad (20)$$

with $x_{i-1} \leq x^* \leq x_i$ and $x_{i-1} \leq x^* \leq x_{i+1}$ for Eqs. (19) and (20), respectively.

This analysis is confirmed by the numerical results presented in Fig. 5 that displays, in logarithmic axes, the exact and truncation errors against Δx for the backward and the centred finite differences. Their respective truncation errors ϵ_T are given by (19) and (20) calculated here, for lack of a better value, with $x^* = x^* = x_i$. The exact error is calculated as the difference between the exact value of the derivative and its finite difference approximation.

The slope of the lines are consistent with the order of the truncation error, i.e., 1:1 for the backward difference and 1:2 for the centred difference. The discrepancies between the exact and the numerical results for the smallest values of Δx are due to the use of finite precision computer arithmetic or round-off error. This issue and its implications are discussed in more detail in numerical analysis textbooks as in Ref. [3].

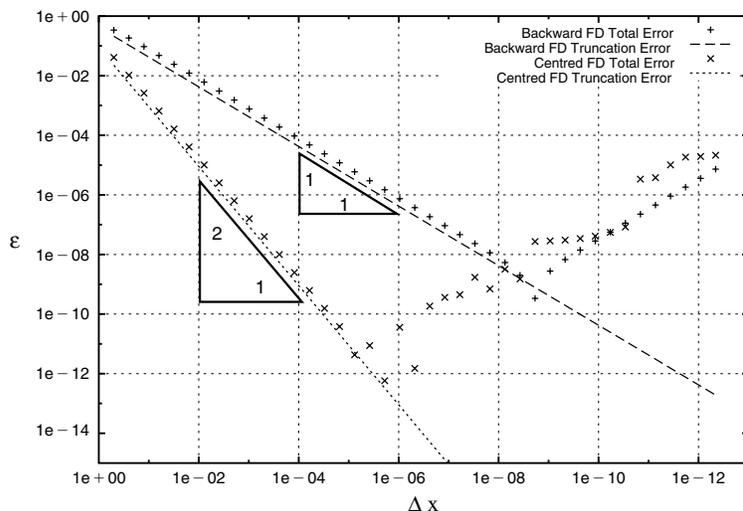


Figure 5. Truncation and rounding errors in the finite difference approximation of derivatives.

3.1.2. Derivation of approximations using Taylor expansions

The procedure described in the previous section can be easily transformed into a general method for deriving finite difference schemes. In general, we can obtain approximations to higher order derivatives by selecting an appropriate number of interpolation points that permits us to eliminate the highest term of the truncation error from the Taylor expansions. We will illustrate this with some examples. A more general description of this derivation can be found in Hirsch (1988).

A second-order accurate finite difference approximation of the derivative at x_i can be derived by considering the values of u at three points: x_{i-1} , x_i and x_{i+1} . The approximation is constructed as a weighted average of these values $\{u_{i-1}, u_i, u_{i+1}\}$ such as

$$u_x|_i \approx \frac{\alpha u_{i+1} + \beta u_i + \gamma u_{i-1}}{\Delta x}. \quad (21)$$

Using Taylor expansions around x_i we can write

$$u_{i+1} = u_i + \Delta x u_x|_i + \frac{\Delta x^2}{2} u_{xx}|_i + \frac{\Delta x^3}{6} u_{xxx}|_i + \dots \quad (22)$$

$$u_{i-1} = u_i - \Delta x u_x|_i + \frac{\Delta x^2}{2} u_{xx}|_i - \frac{\Delta x^3}{6} u_{xxx}|_i + \dots \quad (23)$$

Putting (22), (23) into (21) we get

$$\begin{aligned} \frac{\alpha u_{i+1} + \beta u_i + \gamma u_{i-1}}{\Delta x} &= (\alpha + \beta + \gamma) \frac{1}{\Delta x} u_i + (\alpha - \gamma) u_x|_i \\ &+ (\alpha + \gamma) \frac{\Delta x}{2} u_{xx}|_i + (\alpha - \gamma) \frac{\Delta x^2}{6} u_{xxx}|_i \\ &+ (\alpha + \gamma) \frac{\Delta x^3}{12} u_{xxxx}|_i + O(\Delta x^4) \end{aligned} \quad (24)$$

We require three independent conditions to calculate the three unknowns α , β and γ . To determine these we impose that the expression (24) is consistent with increasing orders of accuracy. If the solution is constant, the left-hand side of (24) should be zero. This requires the coefficient of $(1/\Delta x)u_i$ to be zero and therefore $\alpha + \beta + \gamma = 0$. If the solution is linear, we must have $\alpha - \gamma = 1$ to match $u_x|_i$. Finally whilst the first two conditions are necessary for consistency of the approximation in this case we are free to choose the third condition. We can therefore select the coefficient of $(\Delta x/2) u_{xx}|_i$ to be zero to improve the accuracy, which means $\alpha + \gamma = 0$.

Solving these three equations we find the values $\alpha = 1/2$, $\beta = 0$ and $\gamma = -(1/2)$ and recover the second-order accurate centred formula

$$u_x|_i = \frac{u_{i+1} - u_{i-1}}{2\Delta x} + O(\Delta x^2).$$

Other approximations can be obtained by selecting a different set of points, for instance, we could have also chosen three points on the side of x_i , e.g., u_i, u_{i-1}, u_{i-2} . The corresponding approximation is known as a one-sided formula. This is sometimes useful to impose boundary conditions on u_x at the ends of the mesh.

3.1.3. Higher-order derivatives

In general, we can derive an approximation of the second derivative using the Taylor expansion

$$\begin{aligned} \frac{\alpha u_{i+1} + \beta u_i + \gamma u_{i-1}}{\Delta x^2} &= (\alpha + \beta + \gamma) \frac{1}{\Delta x^2} u_i + (\alpha - \gamma) \frac{1}{\Delta x} u_x|_i \\ &+ (\alpha + \gamma) \frac{1}{2} u_{xx}|_i + (\alpha - \gamma) \frac{\Delta x}{6} u_{xxx}|_i \\ &+ (\alpha + \gamma) \frac{\Delta x^2}{12} u_{xxxx}|_i + O(\Delta x^4). \end{aligned} \quad (25)$$

Using similar arguments to those of the previous section we impose

$$\left. \begin{aligned} \alpha + \beta + \gamma &= 0 \\ \alpha - \gamma &= 0 \\ \alpha + \gamma &= 2 \end{aligned} \right\} \Rightarrow \alpha = \gamma = 1, \beta = -2. \quad (26)$$

The first and second conditions require that there are no u or u_x terms on the right-hand side of Eq. (25) whilst the third condition ensures that the right-hand side approximates the left-hand side as Δx tends to zero. The solution of Eq. (26) lead us to the second-order centred approximation

$$u_{xx}|_i = \frac{u_{i+1} - 2u_i + u_{i-1}}{\Delta x^2} + O(\Delta x^2). \quad (27)$$

The last term in the Taylor expansion $(\alpha - \gamma)\Delta x u_{xxx}|_i/6$ has the same coefficient as the u_x terms and cancels out to make the approximation second-order accurate. This cancellation does not occur if the points in the mesh are not equally spaced. The derivation of a general three point finite difference approximation with unevenly spaced points can also be obtained through Taylor series. We leave this as an exercise for the reader and proceed in the next section to derive a general form using an alternative method.

3.1.4. Finite differences through polynomial interpolation

In this section we seek to approximate the values of $u(x)$ and its derivatives by a polynomial $P(x)$ at a given point x_i . As way of an example we will derive similar expressions to the centred differences presented previously by considering an approximation involving the set of points $\{x_{i-1}, x_i, x_{i+1}\}$ and the corresponding values $\{u_{i-1}, u_i, u_{i+1}\}$. The polynomial of minimum degree that satisfies $P(x_{i-1}) = u_{i-1}$, $P(x_i) = u_i$ and $P(x_{i+1}) = u_{i+1}$ is the quadratic Lagrange polynomial

$$\begin{aligned} P(x) &= u_{i-1} \frac{(x - x_i)(x - x_{i+1})}{(x_{i-1} - x_i)(x_{i-1} - x_{i+1})} + u_i \frac{(x - x_{i-1})(x - x_{i+1})}{(x_i - x_{i-1})(x_i - x_{i+1})} \\ &\quad + u_{i+1} \frac{(x - x_{i-1})(x - x_i)}{(x_{i+1} - x_{i-1})(x_{i+1} - x_i)}. \end{aligned} \quad (28)$$

We can now obtain an approximation of the derivative, $u_x|_i \approx P_x(x_i)$ as

$$\begin{aligned} P_x(x_i) &= u_{i-1} \frac{(x_i - x_{i+1})}{(x_{i-1} - x_i)(x_{i-1} - x_{i+1})} + u_i \frac{(x_i - x_{i-1}) + (x_i - x_{i+1})}{(x_i - x_{i-1})(x_i - x_{i+1})} \\ &\quad + u_{i+1} \frac{(x_i - x_{i-1})}{(x_{i+1} - x_{i-1})(x_{i+1} - x_i)}. \end{aligned} \quad (29)$$

If we take $x_i - x_{i-1} = x_{i+1} - x_i = \Delta x$, we recover the second-order accurate finite difference approximation (14) which is consistent with a quadratic

Homework 1 Use Taylor expansion to derive expressions for $u_{x,i}$ of orders $\mathcal{O}(\Delta x^3)$ and $\mathcal{O}(\Delta x^4)$ using five-point stencil, $\{u_n : i - 2 \leq n \leq i + 2\}$. (Hint: You're going to get one backward differencing (BD) and one forward differencing (FD) that are of orders $\mathcal{O}(\Delta x^3)$, and one centered-differencing (CD) that is of $\mathcal{O}(\Delta x^4)$.)

Homework 2 Use Taylor expansion to derive expressions for $u_{xx,i}$ of order $\mathcal{O}(\Delta x^4)$ using five-point stencil, $\{u_n : i - 2 \leq n \leq i + 2\}$. (Hint: You're going to get one centered-differencing (CD) that is of $\mathcal{O}(\Delta x^4)$.)

Homework 3 Consider $f(x) = \sin(x)$ over $\mathcal{D} = [0, 2\pi]$. Assume period BC.

(a) Use formulas (all three of them) you obtain in **Homework 1** with step sizes $h = 0.1, 0.01, 0.001$, and 0.0001 and calculate approximations for $f'(x)$ over \mathcal{D} . Carry nine decimal places in all of your calculations.

(b) Compare with the exact result $f'(x) = \cos(x)$, $\forall x \in \mathcal{D}$. What you wish to do is to produce log-log error norm plots of $\|E\|_1$, where the discrete 1-norm of error E is defined by

$$\|E\|_1 = h \sum_j^N \left| F_{exact}(x_j) - f'(x_j) \right|, \quad (4.1)$$

versus the grid scales, $1/h$. Here, F_{exact} is your analytical solution to which your numerical solution $f'(x_j)$ approximates. See Fig. 5 in the article. Please over-plot your numerical solutions with analytical convergence rates you expect to see (i.e., the dotted lines in Fig. 5). Discuss the convergence behaviors.

(c) Repeat the same exercises for $f''(x)$ using the result you get from **Homework 2**.

1. Polynomial Interpolations

We know that there exists a unique N -th order polynomial passing through any set of $N + 1$ number of sample points. Depending on where it is evaluated, this polynomial is either an interpolation polynomial or an extrapolation polynomial. Two very popular interpolation polynomial choices frequently adopted in discrete approximations are the *Lagrange form* and the *Newton form*.

1.1. Lagrange Form

See Section 3.1.4 of the attached article.

Remark: The Lagrange form is easy to derive and easy to remember but may be difficult to work with (for example, integration and differentiation are difficult in Lagrange form as they are in rational form).

Using similar arguments to those of the previous section we impose

$$\left. \begin{aligned} \alpha + \beta + \gamma &= 0 \\ \alpha - \gamma &= 0 \\ \alpha + \gamma &= 2 \end{aligned} \right\} \Rightarrow \alpha = \gamma = 1, \beta = -2. \quad (26)$$

The first and second conditions require that there are no u or u_x terms on the right-hand side of Eq. (25) whilst the third condition ensures that the right-hand side approximates the left-hand side as Δx tends to zero. The solution of Eq. (26) lead us to the second-order centred approximation

$$u_{xx}|_i = \frac{u_{i+1} - 2u_i + u_{i-1}}{\Delta x^2} + O(\Delta x^2). \quad (27)$$

The last term in the Taylor expansion $(\alpha - \gamma)\Delta x u_{xxx}|_i/6$ has the same coefficient as the u_x terms and cancels out to make the approximation second-order accurate. This cancellation does not occur if the points in the mesh are not equally spaced. The derivation of a general three point finite difference approximation with unevenly spaced points can also be obtained through Taylor series. We leave this as an exercise for the reader and proceed in the next section to derive a general form using an alternative method.

3.1.4. Finite differences through polynomial interpolation

In this section we seek to approximate the values of $u(x)$ and its derivatives by a polynomial $P(x)$ at a given point x_i . As way of an example we will derive similar expressions to the centred differences presented previously by considering an approximation involving the set of points $\{x_{i-1}, x_i, x_{i+1}\}$ and the corresponding values $\{u_{i-1}, u_i, u_{i+1}\}$. The polynomial of minimum degree that satisfies $P(x_{i-1}) = u_{i-1}$, $P(x_i) = u_i$ and $P(x_{i+1}) = u_{i+1}$ is the quadratic Lagrange polynomial

$$\begin{aligned} P(x) &= u_{i-1} \frac{(x - x_i)(x - x_{i+1})}{(x_{i-1} - x_i)(x_{i-1} - x_{i+1})} + u_i \frac{(x - x_{i-1})(x - x_{i+1})}{(x_i - x_{i-1})(x_i - x_{i+1})} \\ &\quad + u_{i+1} \frac{(x - x_{i-1})(x - x_i)}{(x_{i+1} - x_{i-1})(x_{i+1} - x_i)}. \end{aligned} \quad (28)$$

We can now obtain an approximation of the derivative, $u_x|_i \approx P_x(x_i)$ as

$$\begin{aligned} P_x(x_i) &= u_{i-1} \frac{(x_i - x_{i+1})}{(x_{i-1} - x_i)(x_{i-1} - x_{i+1})} + u_i \frac{(x_i - x_{i-1}) + (x_i - x_{i+1})}{(x_i - x_{i-1})(x_i - x_{i+1})} \\ &\quad + u_{i+1} \frac{(x_i - x_{i-1})}{(x_{i+1} - x_{i-1})(x_{i+1} - x_i)}. \end{aligned} \quad (29)$$

If we take $x_i - x_{i-1} = x_{i+1} - x_i = \Delta x$, we recover the second-order accurate finite difference approximation (14) which is consistent with a quadratic

interpolation. Similarly, for the second derivative we have

$$P_{xx}(x_i) = \frac{2u_{i-1}}{(x_{i-1} - x_i)(x_{i-1} - x_{i+1})} + \frac{2u_i}{(x_i - x_{i-1})(x_i - x_{i+1})} + \frac{2u_{i+1}}{(x_{i+1} - x_{i-1})(x_{i+1} - x_i)} \quad (30)$$

and, again, this approximation leads to the second-order centred finite difference (27) for a constant Δx .

This result is general and the approximation via finite differences can be interpreted as a form of Lagrangian polynomial interpolation. The order of the interpolated polynomial is also the order of accuracy of the finite difference approximation using the same set of points. This is also consistent with the interpretation of a Taylor expansion as an interpolating polynomial.

3.1.5. Finite difference solution of PDEs

We consider the FDM approximation to the solution of the elliptic equation $u_{xx} = s(x)$ in the region $\Omega = \{x : 0 \leq x \leq 1\}$. Discretizing the region using N points with constant mesh spacing $\Delta x = (1/N - 1)$ or $x_i = (i - 1/N - 1)$, we consider two cases with different sets of boundary conditions:

1. $u(0) = \alpha_1$ and $u(1) = \alpha_2$, and
2. $u(0) = \alpha_1$ and $u_x(1) = g$.

In both cases we adopt a centred finite approximation in the interior points of the form

$$\frac{u_{i+1} - 2u_i + u_{i-1}}{\Delta x^2} = s_i; \quad i = 2, \dots, N - 1. \quad (31)$$

The treatment of the first case is straightforward as the boundary conditions are easily specified as $u_1 = \alpha_1$ and $u_N = \alpha_2$. These two conditions together with the $N - 2$ equations (31) result in the linear system of N equations with N unknowns represented by

$$\begin{bmatrix} 1 & 0 & \dots & & & & 0 \\ 1 & -2 & 1 & 0 & \dots & & 0 \\ 0 & 1 & -2 & 1 & 0 & \dots & 0 \\ & & \ddots & \ddots & \ddots & & \\ 0 & \dots & 0 & 1 & -2 & 1 & 0 \\ 0 & & \dots & 0 & 1 & -2 & 1 \\ 0 & & & \dots & & 0 & 1 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ \vdots \\ u_{N-2} \\ u_{N-1} \\ u_N \end{bmatrix} = \begin{bmatrix} \alpha_1 \\ \Delta x^2 s_2 \\ \Delta x^2 s_3 \\ \vdots \\ \Delta x^2 s_{N-2} \\ \Delta x^2 s_{N-1} \\ \alpha_2 \end{bmatrix}.$$

Example: Find the Lagrange form of the interpolation polynomial passing through $(-1, 1)$, $(0, 2)$, $(3, 101)$, $(4, 246)$.

$$P_3(x) = 1 \frac{x}{-1-0} \frac{x-3}{-1-3} \frac{x-4}{-1-4} + 2 \frac{x+1}{0+1} \frac{x-3}{0-3} \frac{x-4}{0-4} + 101 \frac{x+1}{3+1} \frac{x-0}{3-0} \frac{x-4}{3-4} + 246 \frac{x+1}{4+1} \frac{x-0}{4-0} \frac{x-3}{4-3} \quad (4.2)$$

$$= -\frac{1}{20}x(x-3)(x-4) + \frac{2}{12}(x+1)(x-3)(x-4) - \frac{101}{12}(x+1)x(x-4) + \frac{246}{20}(x+1)x(x-3) \quad (4.3)$$

1.2. Newton Form

The Newton form of a polynomial is defined as

$$P_N(x) = a_0 + \sum_{i=1}^N a_i \prod_{j=0}^{i-1} (x - x_j) \quad (4.4)$$

$$= f[x_0] + \sum_{i=1}^N f[x_0, \dots, x_i] \prod_{j=0}^{i-1} (x - x_j), \quad (4.5)$$

where the Newton divided differences $f[x_i, \dots, x_{i+n}]$ is recursively defined as

$$f[x_i, \dots, x_{i+n}] = \frac{f[x_{i+1}, \dots, x_{i+n}] - f[x_i, \dots, x_{i+n-1}]}{x_{i+n} - x_i}, \quad (4.6)$$

with the zeroth Newton divided difference is

$$f[x_i] = f(x_i). \quad (4.7)$$

The Newton form is easily understood if we solve the same example we had in the Lagrange form:

Example: Find the Newton form of the interpolation polynomial passing through $(-1, 1)$, $(0, 2)$, $(3, 101)$, $(4, 246)$.

We first get the zeroth one: $f[x_0] = f[-1] = f(-1) = 1$.

Next, let's compute three of the first Newton divided differences:

$$f[x_0, x_1] = \frac{2-1}{0-(-1)} = 1 \quad (4.8)$$

$$f[x_1, x_2] = \frac{101-2}{3-0} = 33 \quad (4.9)$$

$$f[x_2, x_3] = \frac{246-101}{4-3} = 145 \quad (4.10)$$

Next, we obtain two of the second Newton divided differences:

$$f[x_0, x_1, x_2] = \frac{33 - 1}{3 - (-1)} = 8 \quad (4.11)$$

$$f[x_1, x_2, x_3] = \frac{145 - 33}{4 - 0} = 28 \quad (4.12)$$

Lastly, the fourth Newton divided difference is:

$$f[x_0, x_1, x_2, x_3] = \frac{28 - 8}{4 - (-1)} = 4 \quad (4.13)$$

As a result, we get the following third-degree interpolation polynomial,

$$P_3(x) = 1 + 1(x + 1) + 8(x + 1)x + 4(x + 1)x(x - 3). \quad (4.14)$$

Note: One should notice that this result is the same as the one found using the Lagrange form as shown above.

Remark: We will see later that the Newton form is very handy and useful in constructing some of the high-order interpolation and reconstruction schemes such as ENO, WENO, PPM, etc.

interpolation. Similarly, for the second derivative we have

$$P_{xx}(x_i) = \frac{2u_{i-1}}{(x_{i-1} - x_i)(x_{i-1} - x_{i+1})} + \frac{2u_i}{(x_i - x_{i-1})(x_i - x_{i+1})} + \frac{2u_{i+1}}{(x_{i+1} - x_{i-1})(x_{i+1} - x_i)} \quad (30)$$

and, again, this approximation leads to the second-order centred finite difference (27) for a constant Δx .

This result is general and the approximation via finite differences can be interpreted as a form of Lagrangian polynomial interpolation. The order of the interpolated polynomial is also the order of accuracy of the finite difference approximation using the same set of points. This is also consistent with the interpretation of a Taylor expansion as an interpolating polynomial.

3.1.5. Finite difference solution of PDEs

We consider the FDM approximation to the solution of the elliptic equation $u_{xx} = s(x)$ in the region $\Omega = \{x : 0 \leq x \leq 1\}$. Discretizing the region using N points with constant mesh spacing $\Delta x = (1/N - 1)$ or $x_i = (i - 1/N - 1)$, we consider two cases with different sets of boundary conditions:

1. $u(0) = \alpha_1$ and $u(1) = \alpha_2$, and
2. $u(0) = \alpha_1$ and $u_x(1) = g$.

In both cases we adopt a centred finite approximation in the interior points of the form

$$\frac{u_{i+1} - 2u_i + u_{i-1}}{\Delta x^2} = s_i; \quad i = 2, \dots, N - 1. \quad (31)$$

The treatment of the first case is straightforward as the boundary conditions are easily specified as $u_1 = \alpha_1$ and $u_N = \alpha_2$. These two conditions together with the $N - 2$ equations (31) result in the linear system of N equations with N unknowns represented by

$$\begin{bmatrix} 1 & 0 & \dots & & & & 0 \\ 1 & -2 & 1 & 0 & \dots & & 0 \\ 0 & 1 & -2 & 1 & 0 & \dots & 0 \\ & & \ddots & \ddots & \ddots & & \\ 0 & \dots & 0 & 1 & -2 & 1 & 0 \\ 0 & & \dots & 0 & 1 & -2 & 1 \\ 0 & & & \dots & & 0 & 1 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ \vdots \\ u_{N-2} \\ u_{N-1} \\ u_N \end{bmatrix} = \begin{bmatrix} \alpha_1 \\ \Delta x^2 s_2 \\ \Delta x^2 s_3 \\ \vdots \\ \Delta x^2 s_{N-2} \\ \Delta x^2 s_{N-1} \\ \alpha_2 \end{bmatrix}.$$

This matrix system can be written in abridged form as $A\mathbf{u} = \mathbf{s}$. The matrix A is non-singular and admits a unique solution \mathbf{u} . This is the case for most discretization of well-posed elliptic equations.

In the second case the boundary condition $u(0) = \alpha_1$ is treated in the same way by setting $u_1 = \alpha_1$. The treatment of the Neumann boundary condition $u_x(1) = g$ requires a more careful consideration. One possibility is to use a one-sided approximation of $u_x(1)$ to obtain

$$u_x(1) \approx \frac{u_N - u_{N-1}}{\Delta x} = g. \quad (32)$$

This expression is only first-order accurate and thus inconsistent with the approximation used at the interior points. Given that the PDE is elliptic, this error could potentially reduce the global accuracy of the solution. The alternative is to use a second-order centred approximation

$$u_x(1) \approx \frac{u_{N+1} - u_{N-1}}{\Delta x} = g. \quad (33)$$

Here the value u_{N+1} is not available since it is not part of our discrete set of values but we could use the finite difference approximation at x_N given by

$$\frac{u_{N+1} - 2u_N + u_{N-1}}{\Delta x^2} = s_N$$

and include the Neumann boundary condition (33) to obtain

$$u_N - u_{N-1} = \frac{1}{2}(g\Delta x - s_N\Delta x^2). \quad (34)$$

It is easy to verify that the introduction of either of the Neumann boundary conditions (32) or (34) leads to non-symmetric matrices.

3.2. Time Integration

In this section we address the problem of solving time-dependent PDEs in which the solution is a function of space and time $u(x, t)$. Consider for instance the heat equation

$$u_t - bu_{xx} = s(x) \text{ in } \Omega = \{x, t : 0 \leq x \leq 1, 0 \leq t \leq T\}$$

with an initial condition $u(x, 0) = u_0(x)$ and time-dependent boundary conditions $u(0, t) = \alpha_1(t)$ and $u(1, t) = \alpha_2(t)$, where α_1 and α_2 are known

functions of t . Assume, as before, a mesh or spatial discretization of the domain $\{x_1, \dots, x_N\}$.

3.2.1. Method of lines

In this technique we assign to our mesh a set of values that are functions of time $u_i(t) = u(x_i, t)$; $i = 1, \dots, N$. Applying a centred discretization to the spatial derivative of u leads to a system of ordinary differential equations (ODEs) in the variable t given by

$$\frac{du_i}{dt} = \frac{b}{x^2} \{u_{i-1}(t) - 2u_i(t) + u_{i+1}(t)\} + s_i; \quad i = 2, \dots, N - 1$$

with $u_1 = \alpha_1(t)$ and $u_N = \alpha_2(t)$. This can be written as

$$\frac{d}{dt} \begin{bmatrix} u_2 \\ u_3 \\ \vdots \\ u_{N-2} \\ u_{N-1} \end{bmatrix} = \frac{b}{\Delta x^2} \begin{bmatrix} -2 & 1 & & & \\ 1 & -2 & 1 & & \\ & \ddots & \ddots & \ddots & \\ & & & 1 & -2 & 1 \\ & & & & 1 & -2 \end{bmatrix} \begin{bmatrix} u_2 \\ u_3 \\ \vdots \\ u_{N-2} \\ u_{N-1} \end{bmatrix} + \begin{bmatrix} s_2 + \frac{ba_1(t)}{\Delta x^2} \\ s_3 \\ \vdots \\ s_{N-2} \\ s_{N-1} + \frac{ba_2(t)}{\Delta x^2} \end{bmatrix}$$

or in matrix form as

$$\frac{d\mathbf{u}}{dt}(t) = \mathbf{A} \mathbf{u}(t) + \mathbf{s}(t). \quad (35)$$

Equation (35) is referred to as the *semi-discrete* form or the method of lines. This system can be solved by any method for the integration of initial-value problems [3]. The numerical stability of time integration schemes depends on the eigenvalues of the matrix \mathbf{A} which results from the space discretization. For this example, the eigenvalues vary between 0 and $-(4\alpha/\Delta x^2)$ and this could make the system very *stiff*, i.e., with large differences in eigenvalues, as $\Delta x \rightarrow 0$.

3.2.2. Finite differences in time

The method of finite differences can be applied to time-dependent problems by considering an independent discretization of the solution $u(x, t)$ in space and time. In addition to the spatial discretization $\{x_1, \dots, x_N\}$, the discretization in time is represented by a sequence of times $t^0 = 0 < \dots < t^n < \dots < T$. For simplicity we will assume constant intervals Δx and Δt in space and time, respectively. The discrete solution at a point will be represented by

$u_i^n \approx u(x_i, t^n)$ and the finite difference approximation of the time derivative follows the procedures previously described. For example, the forward difference in time is given by

$$u_t(x, t^n) \approx \frac{u(x, t^{n+1}) - u(x, t^n)}{\Delta t}$$

and the backward difference in time is

$$u_t(x, t^{n+1}) \approx \frac{u(x, t^{n+1}) - u(x, t^n)}{\Delta t}$$

both of which are first-order accurate, i.e., $\epsilon_T = O(\Delta t)$.

Returning to the heat equation $u_t - bu_{xx} = 0$ and using a centred approximation in space, different schemes can be devised depending on the time at which the equation is discretized. For instance, the use of forward differences in time leads to

$$\frac{u_i^{n+1} - u_i^n}{\Delta t} = \frac{b}{\Delta x^2} (u_{i-1}^n - 2u_i^n + u_{i+1}^n). \quad (36)$$

This scheme is *explicit* as the values of the solution at time t^{n+1} are obtained directly from the corresponding (known) values at time t^n . If we use backward differences in time, the resulting scheme is

$$\frac{u_i^{n+1} - u_i^n}{\Delta t} = \frac{b}{\Delta x^2} (u_{i-1}^{n+1} - 2u_i^{n+1} + u_{i+1}^{n+1}). \quad (37)$$

Here to obtain the values at t^{n+1} we must solve a tri-diagonal system of equations. This type of schemes are referred to as *implicit* schemes.

The higher cost of the implicit schemes is compensated by a greater numerical stability with respect to the explicit schemes which are stable in general only for some combinations of Δx and Δt .

3.3. Discretizations Based on the Integral Form

The FDM uses the strong or differential form of the governing equations. In the following, we introduce two alternative methods that use their integral form counterparts: the finite element and the finite volume methods. The use of integral formulations is advantageous as it provides a more natural treatment of Neumann boundary conditions as well as that of discontinuous source terms due to their reduced requirements on the regularity or smoothness of the solution. Moreover, they are better suited than the FDM to deal with complex geometries in multi-dimensional problems as the integral formulations do not rely in any special mesh structure.

These methods use the integral form of the equation as the starting point of the discretization process. For example, if the strong form of the PDE is $\mathcal{L}(u) = s$, the integral form is given by

$$\int_0^1 \mathcal{L}(u)w(x) \, dx = \int_0^1 s w(x) \, dx \quad (38)$$

where the choice of the weight function $w(x)$ defines the type of scheme.

3.3.1. The finite element method (FEM)

Here we discretize the region of interest $\Omega = \{x : 0 \leq x \leq 1\}$ into $N - 1$ subdomains or elements $\Omega_i = \{x : x_{i-1} \leq x \leq x_i\}$ and assume that the approximate solution is represented by

$$u^\delta(x, t) = \sum_{i=1}^N u_i(t) N_i(x)$$

where the set of functions $N_i(x)$ is known as the expansion basis. Its *support* is defined as the set of points where $N_i(x) \neq 0$. If the support of $N_i(x)$ is the whole interval, the method is called a *spectral method*. In the following we will use expansion bases with compact support which are piecewise continuous polynomials within each element as shown in Fig. 6.

The global shape functions $N_i(x)$ can be split within an element into two local contributions of the form shown in Fig. 7. These individual functions are referred to as the *shape functions* or *trial functions*.

3.3.2. Galerkin FEM

In the Galerkin FEM method we set the weight function $w(x)$ in Eq. (38) to be the same as the basis function $N_i(x)$, i.e., $w(x) = N_i(x)$.

Consider again the elliptic equation $\mathcal{L}(u) = u_{xx} = s(x)$ in the region Ω with boundary conditions $u(0) = \alpha$ and $u_x(1) = g$. Equation (38) becomes

$$\int_0^1 w(x) u_{xx} \, dx = \int_0^1 w(x) s(x) \, dx.$$

At this stage, it is convenient to integrate the left-hand side by parts to get the weak form

$$-\int_0^1 w_x u_x \, dx + w(1) u_x(1) - w(0) u_x(0) = \int_0^1 w(x) s(x) \, dx. \quad (39)$$

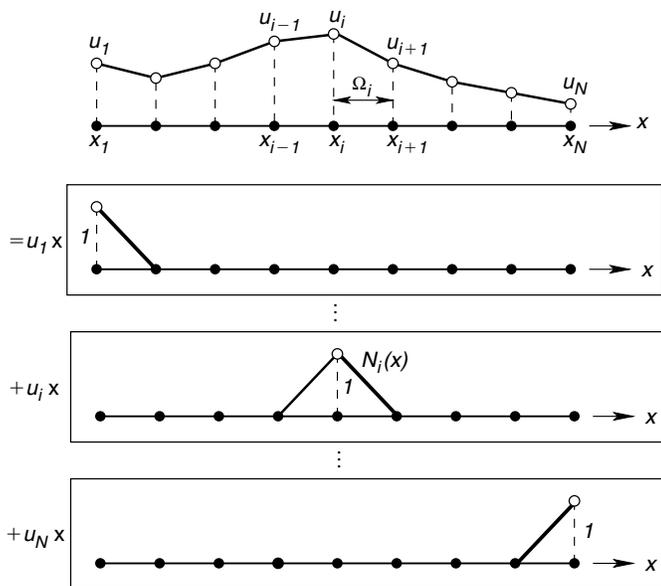


Figure 6. A piecewise linear approximation $u^\delta(x, t) = \sum_{i=1}^N u_i(t) N_i(x)$.

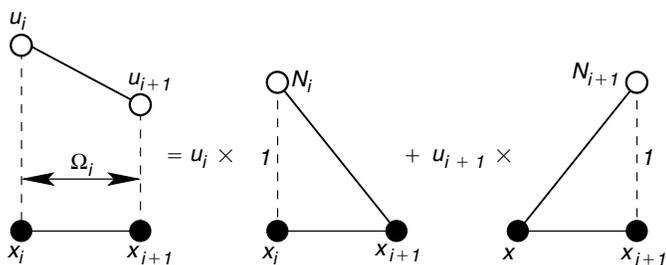


Figure 7. Finite element expansion bases.

This is a common technique in the FEM because it reduces the smoothness requirements on u and it also makes the matrix of the discretized system symmetric. In two and three dimensions we would use Gauss' divergence theorem to obtain a similar result.

The application of the boundary conditions in the FEM deserves attention. The imposition of the Neumann boundary condition $u_x(1) = g$ is straightforward, we simply substitute the value in Eq. (39). This is a very natural way of imposing Neumann boundary conditions which also leads to symmetric

matrices, unlike the FDM. The Dirichlet boundary condition $u(0) = \alpha$ can be applied by imposing $u_1 = \alpha$ and requiring that $w(0) = 0$. In general, we will impose that the weight functions $w(x)$ are zero at the Dirichlet boundaries.

Letting $u(x) \approx u^\delta(x) = \sum_{j=1}^N u_j N_j(x)$ and $w(x) = N_i(x)$ then Eq. (39) becomes

$$-\int_0^1 \frac{dN_i}{dx}(x) \sum_{j=1}^N u_j \frac{dN_j}{dx}(x) dx = \int_0^1 N_i(x) s(x) dx \quad (40)$$

for $i=2, \dots, N$. This represents a linear system of $N - 1$ equations with $N - 1$ unknowns: $\{u_2, \dots, u_N\}$. Let us proceed to calculate the integral terms corresponding to the i th equation. We calculate the integrals in Eq. (40) as sums of integrals over the elements Ω_i . The basis functions have compact support, as shown in Fig. 6. Their value and their derivatives are different from zero only on the elements containing the node i , i.e.,

$$N_i(x) = \begin{cases} \frac{x - x_{i-1}}{\Delta x_{i-1}} & x_{i-1} < x < x_i \\ \frac{x_{i+1} - x}{\Delta x_i} & x_i < x < x_{i+1} \end{cases}$$

$$\frac{dN_i(x)}{dx} = \begin{cases} \frac{1}{\Delta x_{i-1}} & x_{i-1} < x < x_i \\ -\frac{1}{\Delta x_i} & x_i < x < x_{i+1} \end{cases}$$

with $\Delta x_{i-1} = x_i - x_{i-1}$ and $\Delta x_i = x_{i+1} - x_i$. This means that the only integrals different from zero in (40) are

$$\begin{aligned} & -\int_{x_{i-1}}^{x_i} \frac{dN_i}{dx} \left(u_{i-1} \frac{dN_{i-1}}{dx} + u_i \frac{dN_i}{dx} \right) - \int_{x_i}^{x_{i+1}} \frac{dN_i}{dx} \left(u_i \frac{dN_i}{dx} + u_{i+1} \frac{dN_{i+1}}{dx} \right) dx \\ & = \int_{x_{i-1}}^{x_i} N_i s dx + \int_{x_i}^{x_{i+1}} N_i s dx \end{aligned} \quad (41)$$

The right-hand side of this equation expressed as

$$F = \int_{x_{i-1}}^{x_i} \frac{x - x_{i-1}}{\Delta x_{i-1}} s(x) dx + \int_{x_i}^{x_{i+1}} \frac{x_{i+1} - x}{\Delta x_i} s(x) dx$$

can be evaluated using a simple integration rule like the trapezium rule

$$\int_{x_i}^{x_{i+1}} g(x) dx \approx \frac{g(x_i) + g(x_{i+1})}{2} \Delta x_i$$

and it becomes

$$F = \left(\frac{\Delta x_{i-1}}{2} + \frac{\Delta x_i}{2} \right) s_i.$$

Performing the required operations in the left-hand side of Eq. (41) and including the calculated value of F leads to the FEM discrete form of the equation as

$$-\frac{u_i - u_{i-1}}{\Delta x_{i-1}} + \frac{u_{i+1} - u_i}{\Delta x_i} = \frac{\Delta x_{i-1} + \Delta x_i}{2} s_i.$$

Here if we assume that $\Delta x_{i-1} = \Delta x_i = \Delta x$ then the equispaced approximation becomes

$$\frac{u_{i+1} - 2u_i + u_{i-1}}{\Delta x} = \Delta x s_i$$

which is identical to the finite difference formula. We note, however, that the general FE formulation did not require the assumption of an equispaced mesh.

In general the evaluation of the integral terms in this formulation are more efficiently implemented by considering most operations in a standard element $\Omega_{st} = \{-1 \leq x \leq 1\}$ where a mapping is applied from the element Ω_i to the standard element Ω_{st} . For more details on the general formulation see Ref. [4].

3.3.3. Finite volume method (FVM)

The integral form of the one-dimensional linear advection equation is given by Eq. (1) with $f(u) = au$ and $S = 0$. Here the region of integration is taken to be a *control volume* Ω_i , associated with the point of coordinate x_i , represented by $x_{i-(1/2)} \leq x \leq x_{i+(1/2)}$, following the notation of Fig. 4, and the integral form is written as

$$\int_{x_{i-(1/2)}}^{x_{i+(1/2)}} u_t \, dx + \int_{x_{i-(1/2)}}^{x_{i+(1/2)}} f_x(u) \, dx = 0. \quad (42)$$

This expression could also be obtained from the weighted residual form (4) by selecting a weight $w(x)$ such that $w(x) = 1$ for $x_{i-(1/2)} \leq x \leq x_{i+(1/2)}$ and $w(x) = 0$ elsewhere. The last term in Eq. (42) can be evaluated analytically to obtain

$$\int_{x_{i-(1/2)}}^{x_{i+(1/2)}} f_x(u) \, dx = f(u_{i+(1/2)}) - f(u_{i-(1/2)})$$

and if we approximate the first integral using the mid-point rule we finally have the semi-discrete form

$$u_t|_i (x_{i+(1/2)} - x_{i-(1/2)}) + f(u_{i+(1/2)}) - f(u_{i-(1/2)}) = 0.$$

This approach produces a *conservative* scheme if the flux on the boundary of one cell equals the flux on the boundary of the adjacent cell. Conservative schemes are popular for the discretization of hyperbolic equations since, if they converge, they can be proven (Lax-Wendroff theorem) to converge to a weak solution of the conservation law.

3.3.4. Comparison of FVM and FDM

To complete our comparison of the different techniques we consider the FVM discretization of the elliptic equation $u_{xx} = s$. The FVM integral form of this equation over a control volume $\Omega_i = \{x_{i-(1/2)} \leq x \leq x_{i+(1/2)}\}$ is

$$\int_{x_{i-(1/2)}}^{x_{i+(1/2)}} u_{xx} \, dx = \int_{x_{i-(1/2)}}^{x_{i+(1/2)}} s \, dx.$$

Evaluating exactly the left-hand side and approximating the right-hand side by the mid-point rule we obtain

$$u_x(x_{i+(1/2)}) - u_x(x_{i-(1/2)}) = (x_{i+(1/2)} - x_{i-(1/2)}) s_i. \quad (43)$$

If we approximate $u(x)$ as a linear function between the mesh points $i-1$ and i , we have

$$u_x|_{i-(1/2)} \approx \frac{u_i - u_{i-1}}{x_i - x_{i-1}}, \quad u_x|_{i+(1/2)} \approx \frac{u_{i+1} - u_i}{x_{i+1} - x_i},$$

and introducing these approximations into Eq. (43) we now have

$$\frac{u_{i+1} - u_i}{x_{i+1} - x_i} - \frac{u_i - u_{i-1}}{x_i - x_{i-1}} = (x_{i+(1/2)} - x_{i-(1/2)}) s_i.$$

If the mesh is equispaced then this equation reduces to

$$\frac{u_{i+1} - 2u_i + u_{i-1}}{\Delta x} = \Delta x s_i,$$

which is the same as the FDM and FEM on an equispaced mesh.

Once again we see the similarities that exist between these methods although some assumptions in the construction of the FVM have been made. FEM and FVM allow a more general approach to non-equispaced meshes (although this can also be done in the FDM). In two and three dimensions, curvature is more naturally dealt with in the FVM and FEM due to the integral nature of the equations used.

4. High Order Discretizations: Spectral Element/ p -Type Finite Elements

All of the approximations methods we have discussed this far have dealt with what is typically known as the h -type approximation. If $h = \Delta x$ denotes the size of a finite difference spacing or finite elemental regions then convergence of the discrete approximation to the PDE is achieved by letting $h \rightarrow 0$. An alternative method is to leave the mesh spacing fixed but to increase the polynomial order of the local approximation which is typically denoted by p or the p -type extension.

We have already seen that higher order finite difference approximations can be derived by fitting polynomials through more grid points. The drawback of this approach is that the finite difference stencil gets larger as the order of the polynomial approximation increases. This can lead to difficulties when enforcing boundary conditions particularly in multiple dimensions. An alternative approach to deriving high order finite differences is to use compact finite differences where a Padé approximation is used to approximate the derivatives.

When using the finite element method in an integral formulation, it is possible to develop a compact high-order discretization by applying higher order polynomial expansions within every elemental region. So instead of using just a linear element in each piecewise approximation of Fig. 6 we can use a polynomial of order p . This technique is commonly known as p -type finite element in structural mechanics or the *spectral element* method in fluid mechanics. The choice of the polynomial has a strong influence on the numerical conditioning of the approximation and we note that the choice of an equi-spaced Lagrange polynomial is particularly bad for $p > 5$. The two most commonly used polynomial expansions are Lagrange polynomial based on the Gauss–Lobatto–Legendre quadratures points or the integral of the Legendre polynomials in combination with the linear finite element expansion. These two polynomial expansions are shown in Fig. 8. Although this method is more

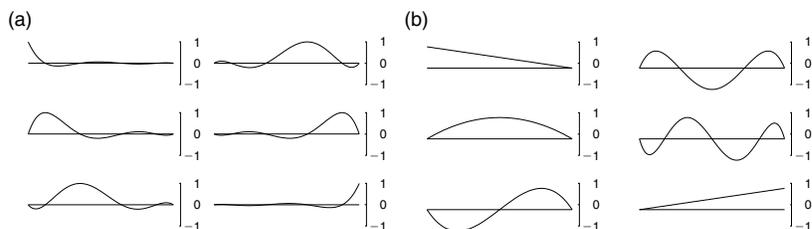


Figure 8. Shape of the fifth order ($p = 5$) polynomial expansions typically used in (a) spectral element and (b) p -type finite element methods.

involved to implement, the advantage is that for a smooth problem (i.e., one where the derivatives of the solution are well behaved) the computational cost increases algebraically whilst the error decreases exponentially fast. Further details on these methods can be found in Refs. [5, 6].

5. Numerical Difficulties

The discretization of linear elliptic equations with either FD, FE or FV methods leads to non-singular systems of equations that can easily solved by standard methods of solution. This is not the case for time-dependent problems where numerical errors may grow unbounded for some discretization. This is perhaps better illustrated with some examples.

Consider the parabolic problem represented by the diffusion equation $u_t - u_{xx} = 0$ with boundary conditions $u(0) = u(1) = 0$ solved using the scheme (36) with $b = 1$ and $\Delta x = 0.1$. The results obtained with $\Delta t = 0.004$ and 0.008 are depicted in Figs. 9(a) and (b), respectively. The numerical solution (b) corresponding to $\Delta t = 0.008$ is clearly unstable.

A similar situation occurs in hyperbolic problems. Consider the one-dimensional linear advection equation $u_t + au_x = 0$; with $a > 0$ and various explicit approximations, for instance the backward in space, or upwind, scheme is

$$\frac{u_i^{n+1} - u_i^n}{\Delta t} + a \frac{u_i^n - u_{i-1}^n}{\Delta x} = 0 \Rightarrow u_i^{n+1} = (1 - \sigma)u_i^n + \sigma u_{i-1}^n, \quad (44)$$

the forward in space, or downwind, scheme is

$$\frac{u_i^{n+1} - u_i^n}{\Delta t} + a \frac{u_{i+1}^n - u_i^n}{\Delta x} = 0 \Rightarrow u_i^{n+1} = (1 + \sigma)u_i^n - \sigma u_{i+1}^n, \quad (45)$$

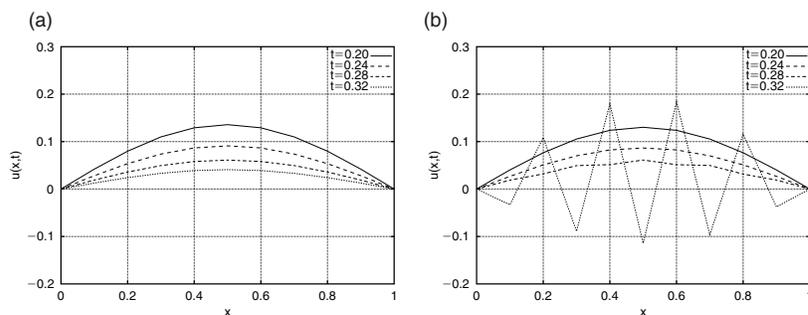


Figure 9. Solution to the diffusion equation $u_t + u_{xx} = 0$ using a forward in time and centred in space finite difference discretization with $\Delta x = 0.1$ and (a) $\Delta t = 0.004$, and (b) $\Delta t = 0.008$. The numerical solution in (b) is clearly unstable.

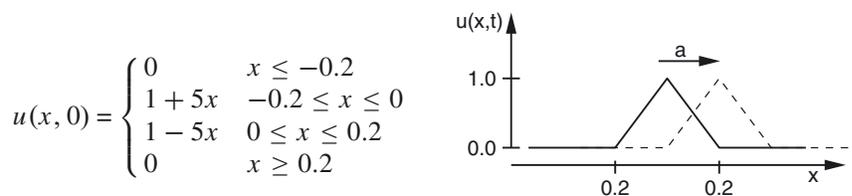


Figure 10. A triangular wave as initial condition for the advection equation.

and, finally, the centred in space is given by

$$\frac{u_i^{n+1} - u_i^n}{\Delta t} + a \frac{u_{i+1}^n - u_{i-1}^n}{2\Delta x} = 0 \quad \Rightarrow \quad u_i^{n+1} = u_i^n - \frac{\sigma}{2}(u_{i+1}^n - u_{i-1}^n) \quad (46)$$

where $\sigma = (a \Delta t / \Delta x)$ is known as the *Courant number*. We will see later that this number plays an important role in the stability of hyperbolic equations. Let us obtain the solution of $u_t + au_x = 0$ for all these schemes with the initial condition given in Fig. 10.

As also indicated in Fig. 10, the exact solution is the propagation of this wave form to the right at a velocity a . Now we consider the solution of the three schemes at two different Courant numbers given by $\sigma = 0.5$ and 1.5 . The results are presented in Fig. 11 and we observe that only the upwinded scheme when $\sigma \leq 1$ gives a stable, although diffusive, solution. The centred scheme when $\sigma = 0.5$ appears almost stable but the oscillations grow in time leading to an unstable solution.

6. Analysis of Numerical Schemes

We have seen that different parameters, such as the Courant number, can effect the stability of a numerical scheme. We would now like to set up a more rigorous framework to analyse a numerical scheme and we introduce the concepts of *consistency*, *stability* and *Convergence* of a numerical scheme.

6.1. Consistency

A numerical scheme is consistent if the discrete numerical equation tends to the exact differential equation as the mesh size (represented by Δx and Δt) tends to zero.

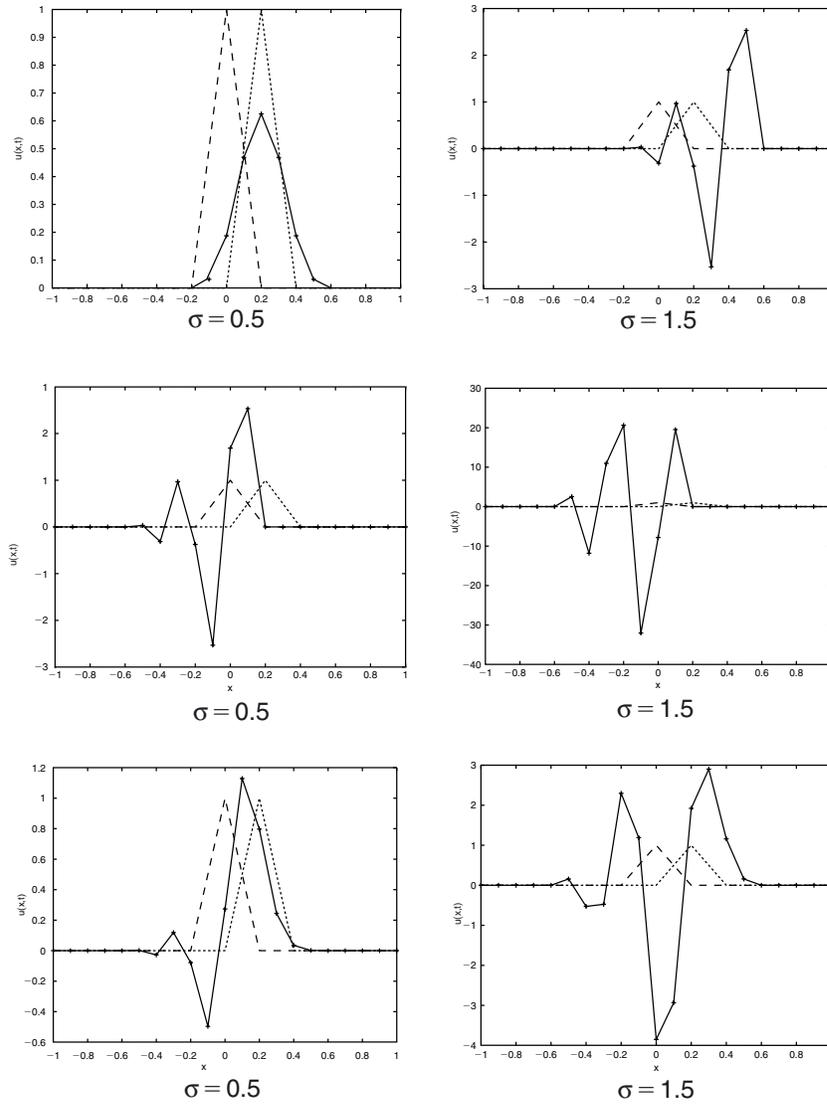


Figure 11. Numerical solution of the advection equation $u_t + au_x = 0$. Dashed lines: initial condition. Dotted lines: exact solution. Solid line: numerical solution.

Consider the centred in space and forward in time finite difference approximation to the linear advection equation $u_t + au_x = 0$ given by Eq. (46). Let us consider Taylor expansions of u_i^{n+1} , u_{i+1}^n and u_{i-1}^n around (x_i, t^n) as

$$u_i^{n+1} = u_i^n + \Delta t u_t|_i^n + \frac{\Delta t^2}{2} u_{tt}|_i^n + \dots$$

$$u_{i+1}^n = u_i^n + \Delta x u_x|_i^n + \frac{\Delta x^2}{2} u_{xx}|_i^n + \frac{\Delta x^3}{6} u_{xxx}|_i^n + \dots$$

$$u_{i-1}^n = u_i^n - \Delta x u_x|_i^n + \frac{\Delta x^2}{2} u_{xx}|_i^n - \frac{\Delta x^3}{6} u_{xxx}|_i^n + \dots$$

Substituting these expansions into Eq. (46) and suitably re-arranging the terms we find that

$$\frac{u_i^{n+1} - u_i^n}{\Delta t} + a \frac{u_{i+1}^n - u_{i-1}^n}{2\Delta x} - (u_t + au_x)|_i^n = \epsilon_T \quad (47)$$

where ϵ_T is known as the *truncation error* of the approximation and is given by

$$\epsilon_T = \frac{\Delta t}{2} u_{tt}|_i^n + \frac{\Delta x^2}{6} au_{xxx}|_i^n + O(\Delta t^2, \Delta x^4).$$

The left-hand side of this equation will tend to zero as Δt and Δx tend to zero. This means that the numerical scheme (46) tends to the exact equation at point x_i and time level t^n and therefore this approximation is *consistent*.

6.2. Stability

We have seen in the previous numerical examples that errors in numerical solutions can grow uncontrolled and render the solution meaningless. It is therefore sensible to require that the solution is stable, this is that the difference between the computed solution and the exact solution of the discrete equation should remain bounded as $n \rightarrow \infty$ for a given Δx .

6.2.1. The Courant–Friedrichs–Lewy (CFL) condition

This is a necessary condition for stability of explicit schemes devised by Courant, Friedrichs and Lewy in 1928.

Recalling the theory of characteristics for hyperbolic systems, the *domain of dependence of a PDE* is the portion of the domain that influences the solution at a given point. For a scalar conservation law, it is the characteristic passing through the point, for instance, the line PQ in Fig. 12. The *domain of dependence of a FD scheme* is the set of points that affect the approximate solution at a given point. For the upwind scheme, the numerical domain of dependence is shown as a shaded region in Fig. 12.

The *CFL criterion* states that a *necessary* condition for an explicit FD scheme to solve a hyperbolic PDE to be stable is that, for each mesh point, the domain of dependence of the FD approximation contains the domain of dependence of the PDE.

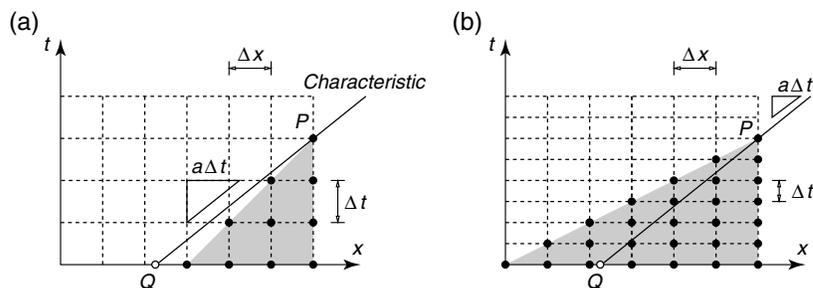


Figure 12. Solution of the advection equation by the upwind scheme. Physical and numerical domains of dependence: (a) $\sigma = (a\Delta t/\Delta x) > 1$, (b) $\sigma \leq 1$.

For a Courant number $\sigma = (a\Delta t/\Delta x)$ greater than 1, changes at Q will affect values at P but the FD approximation cannot account for this.

The CFL condition is necessary for stability of explicit schemes but *it is not sufficient*. For instance, in the previous schemes we have that the upwind FD scheme is stable if the CFL condition $\sigma \leq 1$ is imposed. The downwind FD scheme does not satisfy the CFL condition and is unstable. However, the centred FD scheme is unstable even if $\sigma \leq 1$.

6.2.2. Von Neumann (or Fourier) analysis of stability

The stability of FD schemes for hyperbolic and parabolic PDEs can be analysed by the von Neumann or Fourier method. The idea behind the method is the following. As discussed previously the analytical solutions of the model diffusion equation $u_t - b u_{xx} = 0$ can be found in the form

$$u(x, t) = \sum_{m=-\infty}^{\infty} e^{\beta_m t} e^{I k_m x}$$

if $\beta_m + b k_m^2 = 0$. This solution involves a Fourier series in space and an exponential decay in time since $\beta_m \leq 0$ for $b > 0$. Here we have included the complex version of the Fourier series, $e^{I k_m x} = \cos k_m x + I \sin k_m x$ with $I = \sqrt{-1}$, because this simplifies considerably later algebraic manipulations.

To analyze the growth of different Fourier modes as they evolve under the numerical scheme we can consider each frequency separately, namely

$$u(x, t) = e^{\beta_m t} e^{I k_m x}.$$

A discrete version of this equation is $u_i^n = u(x_i, t^n) = e^{\beta_m t^n} e^{I k_m x_i}$. We can take, without loss of generality, $x_i = i \Delta x$ and $t^n = n \Delta t$ to obtain

$$u_i^n = e^{\beta_m n \Delta t} e^{I k_m i \Delta x} = \left(e^{\beta_m \Delta t} \right)^n e^{I k_m i \Delta x}.$$

The term $e^{I k_m i \Delta x} = \cos(k_m i \Delta x) + I \sin(k_m i \Delta x)$ is bounded and, therefore, any growth in the numerical solution will arise from the term $G = e^{\beta_m \Delta t}$, known as the *amplification factor*. Therefore the numerical method will be stable, or the numerical solution u_i^n bounded as $n \rightarrow \infty$, if $|G| \leq 1$ for solutions of the form

$$u_i^n = G^n e^{I k_m i \Delta x}.$$

We will now proceed to analyse, using the von Neumann method, the stability of some of the schemes discussed in the previous sections.

Example 1 Consider the explicit scheme (36) for the diffusion equation $u_t - b u_{xx} = 0$ expressed here as

$$u_i^{n+1} = \lambda u_{i-1}^n + (1 - 2\lambda) u_i^n + \lambda u_{i+1}^n; \quad \lambda = \frac{b \Delta t}{\Delta x^2}.$$

We assume $u_i^n = G^n e^{I k_m i \Delta x}$ and substitute in the equation to get

$$G = 1 + 2\lambda [\cos(k_m \Delta x) - 1].$$

Stability requires $|G| \leq 1$. Using $-2 \leq \cos(k_m \Delta x) - 1 \leq 0$ we get $1 - 4\lambda \leq G \leq 1$ and to satisfy the left inequality we impose

$$-1 \leq 1 - 4\lambda \leq G \quad \implies \quad \lambda \leq \frac{1}{2}.$$

This means that for a given grid size Δx the maximum allowable timestep is $\Delta t = (\Delta x^2 / 2b)$.

Example 2 Consider the implicit scheme (37) for the diffusion equation $u_t - b u_{xx} = 0$ expressed here as

$$\lambda u_{i-1}^{n+1} + -(1 + 2\lambda) u_i^{n+1} + \lambda u_{i+1}^{n+1} = -u_i^n; \quad \lambda = \frac{b \Delta t}{\Delta x^2}.$$

The amplification factor is now

$$G = \frac{1}{1 + \lambda(2 - \cos \beta_m)}$$

and we have $|G| < 1$ for any β_m if $\lambda > 0$. This scheme is therefore unconditionally stable for any Δx and Δt . This is obtained at the expense of solving a linear system of equations. However, there will still be restrictions on Δx

and Δt based on the accuracy of the solution. The choice between an explicit or an implicit method is not always obvious and should be done based on the computer cost for achieving the required accuracy in a given problem.

Example 3 Consider the upwind scheme for the linear advection equation $u_t + au_x = 0$ with $a > 0$ given by

$$u_i^{n+1} = (1 - \sigma)u_i^n + \sigma u_{i-1}^n; \quad \sigma = \frac{a \Delta t}{\Delta x}.$$

Let us denote $\beta_m = k_m \Delta x$ and introduce the discrete Fourier expression in the upwind scheme to obtain

$$G = (1 - \sigma) + \sigma e^{-I\beta_m}$$

The stability condition requires $|G| \leq 1$. Recall that G is a complex number $G = \xi + I\eta$ so

$$\xi = 1 - \sigma + \sigma \cos \beta_m; \quad \eta = -\sigma \sin \beta_m$$

This represents a circle of radius σ centred at $1 - \sigma$. The stability condition requires the locus of the points (ξ, η) to be interior to a unit circle $\xi^2 + \eta^2 \leq 1$. If $\sigma < 0$ the origin is outside the unit circle, $1 - \sigma > 1$, and the scheme is unstable. If $\sigma > 1$ the back of the locus is outside the unit circle $1 - 2\sigma < 1$ and it is also unstable. Therefore, for stability we require $0 \leq \sigma \leq 1$, see Fig. 13.

Example 4 The forward in time, centred in space scheme for the advection equation is given by

$$u_i^{n+1} = u_i^n - \frac{\sigma}{2}(u_{i+1}^n - u_{i-1}^n); \quad \sigma = \frac{a \Delta t}{\Delta x}.$$

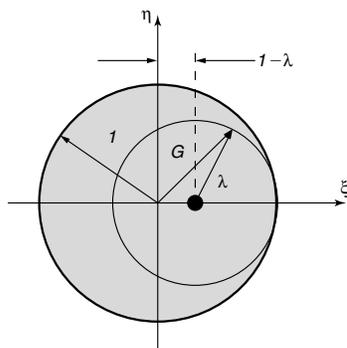


Figure 13. Stability region of the upwind scheme.

The introduction of the discrete Fourier solution leads to

$$G = 1 - \frac{\sigma}{2}(e^{I\beta_m} - e^{-I\beta_m}) = 1 - I\sigma \sin \beta_m$$

Here we have $|G|^2 = 1 + \sigma^2 \sin^2 \beta_m > 1$ always for $\sigma \neq 0$ and it is therefore unstable. We will require a different time integration scheme to make it stable.

6.3. Convergence: Lax Equivalence Theorem

A scheme is said to be *convergent* if the difference between the computed solution and the exact solution of the PDE, i.e., the error $E_i^n = u_i^n - u(x_i, t^n)$, vanishes as the mesh size is decreased. This is written as

$$\lim_{\Delta x, \Delta t \rightarrow 0} |E_i^n| = 0$$

for fixed values of x_i and t^n . This is the fundamental property to be sought from a numerical scheme but it is difficult to verify directly. On the other hand, consistency and stability are easily checked as shown in the previous sections.

The main result that permits the assessment of the convergence of a scheme from the requirements of consistency and stability is the equivalence theorem of Lax stated here without proof:

Stability is the necessary and sufficient condition for a *consistent* linear FD approximation to a well-posed linear initial-value problem to be *convergent*.

7. Suggestions for Further Reading

The basics of the FDM are presented a very accessible form in Ref. [7]. More modern references are Refs. [8, 9].

An elementary introduction to the FVM can be consulted in the book by Versteeg and Malalasekera [10]. An in-depth treatment of the topic with an emphasis on hyperbolic problems can be found in the book by Leveque [2].

Two well established general references for the FEM are the books of Hughes [4] and Zienkiewicz and Taylor [11]. A presentation from the point of view of structural analysis can be consulted in Cook *et al.* [11]

The application of p -type finite element for structural mechanics is dealt with in book of Szabo and Babuška [5]. The treatment of both p -type and spectral element methods in fluid mechanics can be found in book by Karniadakis and Sherwin [6].

A comprehensive reference covering both FDM, FVM and FEM for fluid dynamics is the book by Hirsch [13]. These topics are also presented using a more mathematical perspective in the classical book by Quarteroni and Valli [14].

References

- [1] J. Bonet and R. Wood, *Nonlinear Continuum Mechanics for Finite Element Analysis*. Cambridge University Press, 1997.
- [2] R. Leveque, *Finite Volume Methods for Hyperbolic Problems*, Cambridge University Press, 2002.
- [3] W. Cheney and D. Kincaid, *Numerical Mathematics and Computing*, 4th edn., Brooks/Cole Publishing Co., 1999.
- [4] T. Hughes, *The Finite Element Method: Linear Static and Dynamic Finite Element Analysis*, Dover Publishers, 2000.
- [5] B. Szabo and I. Babuška, *Finite Element Analysis*, Wiley, 1991.
- [6] G.E. Karniadakis and S. Sherwin, *Spectral/hp Element Methods for CFD*, Oxford University Press, 1999.
- [7] G. Smith, *Numerical Solution of Partial Differential Equations: Finite Difference Methods*, Oxford University Press, 1985.
- [8] K. Morton and D. Mayers, *Numerical Solution of Partial Differential Equations*, Cambridge University Press, 1994.
- [9] J. Thomas, *Numerical Partial Differential Equations: Finite Difference Methods*, Springer-Verlag, 1995.
- [10] H. Versteeg and W. Malalasekera, *An Introduction to Computational Fluid Dynamics. The Finite Volume Method*, Longman Scientific & Technical, 1995.
- [11] O. Zienkiewicz and R. Taylor, *The Finite Element Method: The Basis*, vol. 1, Butterworth and Heinemann, 2000.
- [12] R. Cook, D. Malkus, and M. Plesha, *Concepts and Applications of Finite Element Analysis*, Wiley, 2001.
- [13] C. Hirsch, *Numerical Computation of Internal and External Flows*, vol. 1, Wiley, 1988.
- [14] A. Quarteroni and A. Valli, *Numerical Approximation of Partial Differential Equations*, Springer-Verlag, 1994.