Numerical Methods II, Spring 2018,
Courant Institute, New York University, Jonathan Goodman
http://www.math.nyu.edu/faculty/goodman/teaching/NumericalMethodsII2018/index.html

Section 4
Wave propagation, modes, artifacts

1 Motivation

A *wave* is a feature in a time dependent field that moves (*propagates*) with time. The function $u(x,t) = v(x-st)$ is a wave, in that it moves to the right (toward larger x values) with speed s without changing size (*amplitude*) or shape. The function u is constant on lines $x - st = x_0$, which are lines in (x,t) space that move to the right with speed s. This is in one dimension (one *space* dimension, two dimensions with time). In more than one dimension, we say $\hat{k}$ is a *unit vector* if

$$\left|\hat{k}\right| = \left(\sum_j k_j^2\right)^{\frac{1}{2}} = 1 \; .$$

A *plane wave* (or *planar* wave) in the direction $\hat{k}$ is a function

$$u(x,t) = v(\hat{k}^t x - st).$$

This looks like the one dimensional wave if you move in the $\hat{k}$ direction and it is constant on the planes $\hat{k}^t x - st = const.$ If the *wave form* $v(r)$ is a bump (such as $v(r) = e^{-r^2}$), then u is a bump along the plane normal to $\hat{k}$.

There are more complicated waves. The wave form can change over time, replacing $v(r)$ by $v(r,t)$. The wave fronts may be curved. For example, a *spherical* wave (in three dimensions, *circular* in two dimensions) might take the form

$$u(x,t) = v(|x| - st, t) \; .$$

Waves can spread if the wave form $v(r,t)$ becomes wider as a function of r as t increases. *Dispersive* waves (described below) typically do this. Waves can *dissipate* if the height of $v(r,t)$ decreases with t. A more subtle form of dissipation is that the wave form $v(r,t)$ becomes smoother as a function of r as t increases.

A partial differential equation is a *wave equation* if its solutions (or many of its solutions) look like waves. Among the many partial differential equations that do this, there is "the" wave equation

$$\partial_t^2 u = c^2 \triangle u \; . \tag{1}$$

The parameter c is the *wave speed.* In one dimension, the wave equation has solutions

$$u(x,t) = v_1(x+ct) + v_2(x-ct) \; .$$

That is, the solution is a *superposition* (sum) of a simple wave with parameter $s = -c$ and another simple wave with parameter $s = c$. These are the left propagating mode ($s = -c$) and right propagating mode ($s = c$). We will see that every solution to the wave equation (1) in one dimension has this form. More complicated linear wave equations have solutions that may be written as a superposition of modes. Nonlinear partial differential equations do not have a superposition principle. Nevertheless, modes help design solution algorithms.

Wave equations may be understood in two ways. For *hyperbolic* partial differential equations (definition below), there is a finite *propagation speed.* This is the fastest "information" can travel. For a linear equation, a finite propagation speed c means that if $u(x,0)=0$ for $|x|>r$, then $u(x,t)=0$ for $|x|>r+ct$. The region where $u\neq 0$ spreads at speed c and no faster. For "the" wave equation (1), the parameter c in the equation is the propagation speed. The heat equation is different from the wave equation in that it has infinite propagation speed. If $u(x,0)\geq 0$ for all x and $u(x,0)>0$ for some x and $u(x,0)$ is continuous, then $u(x,t)>0$ for all x as long as $t>0$. Propagation speed is important in designing time stepping methods for hyperbolic wave propagation problems. It is the basis of the famous *CFL* condition, named for Richard Courant, Kurt Friedrichs, and Hans Lewy. Courant and Friedrichs were important professors at the Courant Institute.

Linear wave equations (not just hyperbolic ones) also may be understood as a large or infinite collection of oscillators. For "the" wave equation, try a solution of the form

$$u(x,t)=A(t)e^{ikx}\ . \tag{2}$$

Solutions like this are sometimes called *separation of variables* solutions, a term that applies to solutions that are products of functions of fewer variables. This one is a function of t multiplying a function of x. The separation of variables *ansatz* satisfies the wave equation (1) if

$$\ddot{A}=-c^2\left|k\right|^2 A\ . \tag{3}$$

This is the equation for a harmonic oscillator $\ddot{A}=-\omega^2 A$, where the oscillation rate is

$$\omega(k)=\pm c\left|k\right|\ . \tag{4}$$

The solution may be written in real form as $A(t)=\alpha\cos(\omega t)+\beta\sin(\omega t)$, or in complex form as $A(t)=\alpha e^{i\omega t}+\beta e^{-i\omega t}$. As the wave vector k takes all real values, the *dispersion relation* (4) gives the corresponding oscillation frequencies. This one is homogeneous of degree one in k, which means that if $r>0$ is a positive real "scale factor", then $\omega(rk)=r\omega(k)$. It is a defining feature of hyperbolic partial differential equations (linear, homogeneous ones) that the dispersion relation is homogeneous of degree one. We will see other examples. Using Fourier analysis, the initial data may be represented as a sum or integral of "simple plane waves" e^{ikx} (slightly different meaning of "plane wave"). Therefore, the solution may be represented as a sum or integral of oscillating plane waves (2). Viewed this way, the wave equation is an infinite collection of oscillators, one for each wave number k.

A finite difference approximation to the wave equation (1) has plane wave separation of variables solutions of the form

$$\left.\begin{array}{lll} U_j(t)=e^{i\omega(k)t}e^{ikx_j} & x_j=j\Delta x & \text{(semidiscrete)} \\ U_j^n=e^{i\omega(k)t_n}e^{ikx_j} & t_n=n\Delta t & \text{(fully discrete)} \end{array}\right\}\ . \tag{5}$$

The accuracy of the scheme is determined by the order of agreement between the continuous and discrete dispersion relation as $k \to 0$. This is the *long wave* limit and is the limit in which the Fourier mode e^{ikx} is well represented by the discrete counterpart e^{ikx_j}. Let ω_c be the dispersion relation for the PDE (c for continuous) and ω_d the dispersion relation for the finite difference approximation (d for discrete) The finite difference approximation has accuracy of order p if

$$\omega_d(k) = \omega_c(k) + O(|k|^{p+1}) \; . \tag{6}$$

This formula determines the order of accuracy without Taylor series expansion of the solution. Instead you ask what the discrete scheme does to a Fourier mode and compare that to what the PDE does to the same Fourier mode.

In real applications, you usually are at the edge of resolution. The grid is barely able to resolve the smallest scale features in the solution.[1] In this situation, the finite difference approximation probably has *artifacts*, or features that are not in the true solution. We saw this with Fourier interpolation – an under-resolved interpolating function can have overshoots and oscillations. Two common artifacts are *artificial dissipation* and *artificial dispersion.* Artificial dissipation occurs when the discrete dispersion relation is not real, and it becomes possible that

$$\left| e^{i\omega_d(k)t} \right| < 1 \; .$$

This "removes" high wave number modes from the discrete approximation (gives them less weight, likely much less). The result is that sharp edges are rounded and small sharp features have reduced amplitude.

If ω_d is real and $\left| e^{i\omega_d(k)t} \right| = 1$, the difference between $e^{i\omega_d(k)t}$ and $e^{i\omega_c(k)t}$ is called *phase error* or *dispersive error.* In $e^{i\theta}$, θ is the *phase.* Every Fourier mode has the right amplitude but the wrong phase. Dispersion comes from the fact that phase errors make waves move at the wrong speed. Local features are made from a superposition of Fourier modes. If the feature moves with constant speed as $v(x - st)$ all the phases have to change in the right way to keep the feature intact. Phase errors can make local structures come apart. The result can be oscillations similar to the ones from under-resolved Fourier interpolation.

We can understand dissipation and dispersion using explicit calculation in simple one dimensional examples. These may be linear scalar (u has one component) homogeneous (PDE independent of x). We call these *model problems.* They are not realistic but they allow calculations that elucidate phenomena. More realistic problems are nonlinear, have multi-components, and have x dependent coefficients. We may not be able to understand dissipation and dispersion as precisely in real problems, but we can see their effects and understand their origins.

[1] To understand why this happens imagine that you get a new more powerful computer that can compute with more grid points. You have a choice between doing an existing calculation more accurately – with higher resolution – or doing a new calculation that your old computer didn't have the resolution for. A mathematician might choose the existing calculation with higher resolution, but most engineers would rather do a bigger problem at the same resolution.

The simplest model problem for wave propagation is the one dimensional linear homogeneous advection (*advection* defined below) equation. The equation, sometimes called the *Kreiss equation* after master theoretical numerical analyst Heinz Kreiss, is

$$\partial_t u + \partial_x u = 0 \; . \tag{7}$$

Figure 1 shows the result of three approximation schemes. The first order upwind scheme (all schemes described below) has the most artificial dissipation (informally called *smearing*). The sharp edges of the square wave on the right are smeared. The bump on the left has less amplitude than it should. Actually (see below) the total mass of the bump is correct, it has less height because it is smeared out. The approximation converges to the correct answer as $\Delta x \to 0$ and $\Delta t \to 0$, but not very quickly. Lots grid points and computer time are needed to get an accurate approximation.

The Lax Wendroff method is higher order (second order) than the simple upwind scheme so it gets the bump much better. But its artificial dispersion leads to overshoots and oscillations near what should be simple discontinuities. The oscillations are so severe that Lax and Wendroff advocated using extra artificial dissipation. The "pure" Lax Wendroff method is not what Lax or Wendroff advocated. The oscillations become more pronounced for small λ because the dissipation is less. It is possible to derive a *modified equation* that explains these phenomena. The modified equation is a PDE whose solutions are closer numerical approximation than the original advection equation (7). It consists of (7) with dissipation and dispersion terms added, both with coefficients that depend on Δx. The oscillations are large because the leading order error term is dispersive (producing oscillations). Dissipation (which reduces oscillations) is present only at higher order in Δx.

The third order asymmetric scheme is most accurate for the smooth bump, having the highest order of accuracy. It has less artificial dispersion than the Lax Wendroff because the leading term in its modified equation is dissipative, not dispersive. But even the dissipative scheme produces some overshoots. The second order diffusion equation preserves monotonicity and does not produce any extra oscillation. But higher order dissipative equations such as $\partial_t u = -\partial_x^4 u$ (part of the Kuramoto Sivashinsky equation) do produce small overshoots.

2 Wave propagation equations

This section is a partial crash course partial differential equations. Courant Institute graduate courses have students with quite different backgrounds. Many will not need this section, and others will. Good methods for solving wave propagation problems depend on understanding the wave phenomena you are trying to compute, at least qualitatively. Methods for analyzing the PDE can be adopted to understanding and designing computational methods.

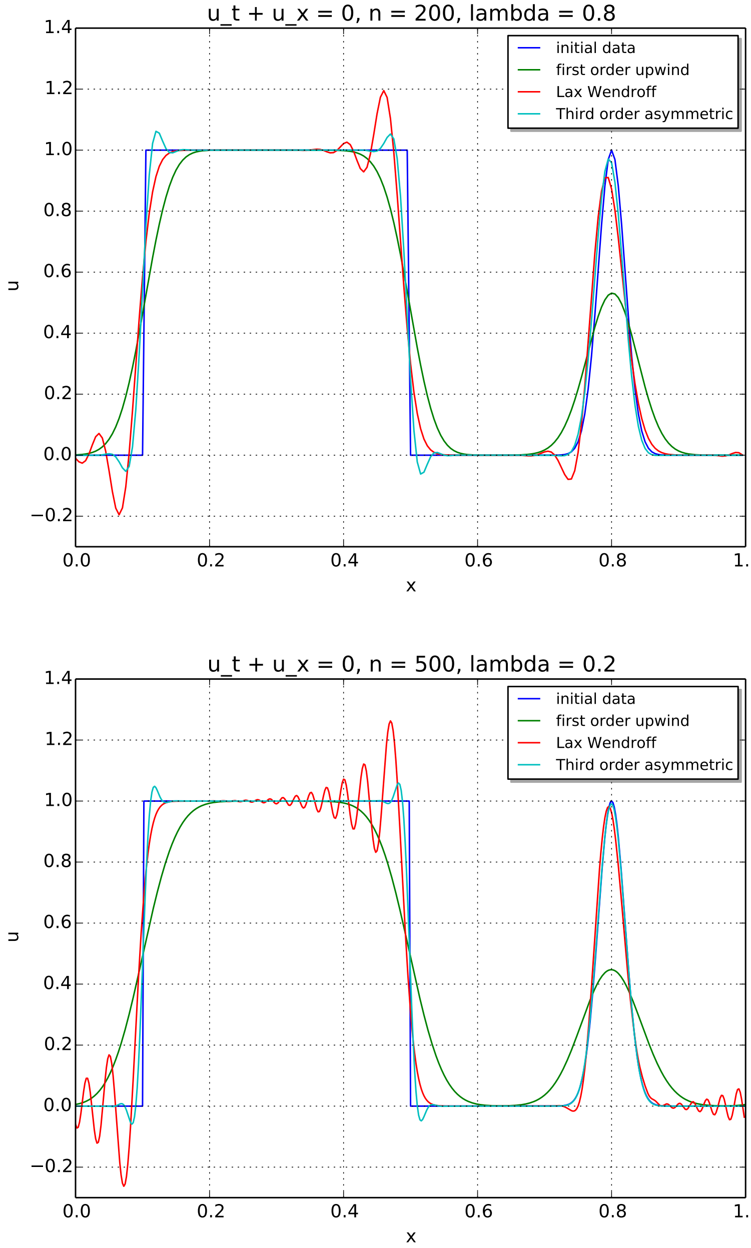


Figure 1: Three approximate solutions for (7) with periodic boundary conditions $u(x+1,t) = u(x,t)$. The solution moves to the right with unit speed and without changing shape. The approximate solutions at $T = 1$ (one full revolution) are plotted, along with the initial conditions. The exact solution is equal to the initial condition. The higher order schemes (Lax Wendroff and the third order scheme) capture the smooth structure well but have overshoots at discontinuities. The first order scheme smears the discontinuities and gets the peak of the smooth structure quite wrong. For the first order method, smearing is worse at small λ. For Lax Wendroff, oscillations are worse at small λ. Note that the lower plot has much higher resolution, yet the first and second order schemes are still poor. `codes/DissipationDispersion.tar`.

2.1 Gas dynamics, local conservation, first order systems

Gas dynamics here refers to a PDE system that describes a compressible gas. The system given here simplifies the physics, particularly the thermodynamics. The space variable x will be in d dimensions with $d = 1, 2$, or 3. The density at point x at time t is $\rho(x,t)$. The pressure is $p(x,t)$. Here, we model the pressure as being a function of the density[2]: $p(x,t) = p(\rho(x,t))$. We assume that the pressure is a strictly increasing function of density (equivalently, density is a strictly increasing function of pressure), which is

$$c^2(\rho) = \frac{dp(\rho)}{d\rho} > 0 \; . \tag{8}$$

A dimensional analysis[3] shows that $c(\rho)$ has units [speed] $= L/T$. This turns out to be the speed of sound in the gas. The gas velocity is $v(x,t)$. In d dimensions, this is $v = (v_1(x,t), \ldots, v_d(x,t)$. For $d = 1$, we write $v(x,t)$ instead of $v_1(x,t)$.

Here is a "conservative" derivation of the gas dynamics PDE system. A quantity $q(x,t)$ is *locally conserved* if the amount of q inside a "volume" V changes only by *flux* (also called *current*) of q crossing the boundary of V. Let Γ be the boundary of V (also written $\Gamma = \partial V$). Let $dA(x)$ be an area element on Γ and $n(x)$ the unit outward normal to Γ at x. The flux for q is $F(x,t) = (F_1 \ldots, F_d)$. Saying F is the flux for q is the same as saying that for "every" volume,

$$\frac{d}{dt}\int_V q(x,t)\,dx = -\int_\Gamma F(x,t)\cdot n(x)\,dA(x) \; . \tag{9}$$

You can understand the sign on the right by thinking of V as a sphere and F pointing out. Then the integral over the sphere (on the right) is positive and the amount of q inside the sphere is decreasing. If F is tangential to Γ (i.e., $F \cdot n = 0$ on Γ), then the amount of q inside V does not change. If there is lots of flux at the boundary, none of it crosses Γ. The divergence theorem allows us to state this in an equivalent differential form

$$\partial_t q(x,t) + \mathrm{div}(F(x,t)) = 0 \; , \;\; \mathrm{div}(F) = \nabla\cdot F = \sum_{j=1}^{d} \partial_{x_j} F_j(x,t) \; . \tag{10}$$

In one dimension, these formulas simplify to (assuming $b > a$)

$$\frac{d}{dt}\int_a^b q(x,t)\,dx = -F(b,t) + F(a,t) \; , \;\; \partial_t q(x,t) + \partial_x F(x,t) = 0 \; . \tag{11}$$

[2] A fluid (gas or liquid) is *incompressible* if pressure changes do not influence the density. No fluid is exactly incompressible, but many fluid problems are accurately modeled as incompressible. The incompressible fluid model does not have sound waves, which are the point of this section.

[3] Pressure is force per unit area. Force has units of mass times acceleration. Density is mass per unit volume. $dp/d\rho$ has units of p/ρ, which are (([force]/[area])/(M/[volume])). In d dimensions, [area] $= L^{d-1}$ and [volume] $= L^d$. Also [force] $= [M \cdot \text{acceleration}] = ML/T^2$. Together, this gives $[dp/d\rho] = ((ML/T^2)/L^{d-1})/(M/L^d) = (L/T)^2$.

Suppose there are n locally conserved quantities $q = (q_1, \ldots, q_n)$ with corresponding fluxes $F_1, \ldots, F_n)$. A *constitutive relation* is a formula for $F_k(x,t)$ as a function of $q(x,t)$. If there is a constitutive relation for every flux, then the local conservation equation (10) becomes a system of partial differential equations

$$\partial_t q_k(x,t) + \nabla \cdot F_k(q(x,t)) = 0 \ , \quad k = 1, \ldots, n \ . \tag{12}$$

More explicitly,

$$\partial_t q_k(x,t) + \sum_{j=1}^{d} \partial_{x_j} F_{j,k}(q(x,t)) = 0 \ , \quad k = 1, \ldots, n \ . \tag{13}$$

There are two ways to make vectors out of flux functions $F_{j,k}$. In the local conservation formula for q_k (12), F_k is the $d-$component vector of fluxes for quantity q_k. In some formulas below, F_j is the $n-$component vector of fluxes for direction j.

In some models, the constitutive relation involves space derivatives of the q_j. In the scalar diffusion equation, for example, $F = -D\nabla q$, which (check this) turns (12) into $\partial_t q = D \triangle q$. The models in this section are "first order" conservation law systems, which means that $F(q)$ depends only on q, not derivatives of q. For a first order conservation law system, we can define the *coefficient matrix*, $A_j(q)$, which is the $n \times n$ derivative matrix of the fluxes for direction j,

$$A_j(q) = F_j'(q) \quad , \quad A_{j,kl}(q) = \frac{\partial F_{j,k}(q)}{\partial q_l} \ . \tag{14}$$

(This would have been written $F_{jk,l}$ in the notation used to derive Runge Kutta methods.) We use the chain rule to write the conservation law system (13) so that it displays it explicitly the first spatial derivatives of the locally conserved quantities:

$$\begin{aligned} &\partial_t q_k + \sum_{j=1}^{d} \sum_{l=1}^{n} \left(\frac{\partial F_{j,k}(q)}{\partial q_l} \right) \partial_{x_j} q_l \\ &\partial_t q + \sum_{j=1}^{d} A_j(q) \partial_{x_j} q \ . \end{aligned} \tag{15}$$

This form of the dynamical equations is called *quasilinear*, which means not linear but something like linear. More precisely, it means that derivatives appear linearly ($\partial_t q$ and $\partial_{x_j} q$) while the coefficient matrix A depends on q. An example of a *fully nonlinear* equation is

$$\partial_t u + (\partial_x u)^2 = 0 \ .$$

The equivalent formulations (12) and (15) are the *conservative* and *qualilinear* versions respectively.

Conservation law reasoning is helpful in deriving gas dynamics and other PDE systems because

- It involves simple direct physical principles and reasoning.
- Some solution methods use approximations to F.
- Solutions of conservation law systems can develop discontinuities, where derivatives are not defined. The conservation law formulation allows us to find physically correct solutions in such cases.

Here is the gas dynamics model in one dimension. There are $n = 2$ locally conserved quantities, which are related to mass and momentum ($x-$momentum if the velocity is in the x direction). The mass in a small interval dx is $\rho(x)dx$. The "mass density" is what we just call density. The momentum in the interval is mass · velocity $= \rho(x)v(x)dx$. The momentum density, the momentum per unit length, is $m(x) = \rho(x)v(x)$. The original fields ρ and v are called *primitive variables.* The mass and momentum densities are *conserved variables.* Either can be expressed in terms of the other, for example $v(x) = m(x)/\rho(x)$.

The rate at which mass crosses a point a is the density multiplied by the velocity, so the mass flux is $F_\rho(a) = \rho(a)v(a)$. Momentum crosses a point a in two ways. One is *streaming* or *advection*; the material crossing a carries momentum with it. The flux from this source is the momentum density times the velocity, which is $[\rho(a)v(a)]\, v(a) = \rho(a)v^2(a)$. Momentum also crosses a because of pressure force – the left of a pushes on the right of a with a force equal to the pressure. The total momentum flux is $F_m(a) = \rho(a)v^2(a) + p(\rho(a)) = m^2(x)/\rho(x) + p(a)$. The conservation of mass and momentum equations are

$$\begin{aligned}\partial_t\rho(x,t) + \partial_x F_\rho(x,t) &= 0\\ \partial_t m(x,t) + \partial_x F_m(x,t) &= 0\end{aligned}$$

Substituting the expressions for the fluxes, we get the conservative formulation of first order gas dynamics:

$$\left.\begin{array}{l}\partial_t\rho(x,t) + \partial_x m(x,t) = 0\\ \partial_t m(x,t) + \partial_x\left[\, m^2(x,t)/\rho(x,t) + p(\rho(x,t))\right] = 0\end{array}\right\} . \qquad (16)$$

You might be more familiar with the equivalent quasilinear formulation in terms of primitive variables. The mass conservation may be rewritten as

$$\begin{aligned}\partial_t\rho + \partial_x\left(\rho v\right) &= 0\\ \partial_t\rho + v\partial_x\rho + \rho\partial_x v &= 0\ .\end{aligned}$$

Next, we manipulate the momentum conservation equation. We use the first form of mass conservation in the next to last line to cancel some terms. We use the sound speed formula (8) in the end.

$$\begin{aligned}\partial_t\left(\rho v\right) + \partial_x\left(\rho v^2 + p(\rho)\right) &= 0\\ v\partial_t\rho + \rho\partial_t v + v\partial_x\left(\rho v\right) + \left(\rho v\right)\partial_x v + p'(\rho)\partial_x\rho &= 0\\ \rho\partial_t v + \left(\rho v\right)\partial_x v + c^2(\rho)\partial_x\rho &= 0\ .\end{aligned}$$

This leads to a common quasilinear form in primitive variables (using subscripts for derivatives):

$$\rho_t + \rho v_x + v\rho_x = 0 \tag{17}$$

$$v_t + vv_x + \frac{c^2(\rho)}{\rho}\rho_x = 0\ . \tag{18}$$

In matrix form, this is

$$\begin{pmatrix}\rho_t\\ v_t\end{pmatrix} + A(\rho,v)\begin{pmatrix}\rho_x\\ v_x\end{pmatrix} = 0 \tag{19}$$

with coefficient matrix

$$A(\rho,v) = \begin{pmatrix} v & \rho \\ \frac{c^2(\rho)}{\rho} & v\end{pmatrix}\ . \tag{20}$$

In three dimensions, the $n = 4$ conserved quantities form a vector

$$q(x,y,z,t) = \begin{pmatrix}\rho\\ m_x\\ m_y\\ m_z\end{pmatrix}\ .$$

Here $m_x = \rho v_x$ is the $x-$momentum density, etc. The flux has x, y, and z components: $F = (F_x, F_y, F_z)$. The four components of F_x correspond to the rates of the local conserved quantities crossing a surface normal to the x direction. These are

$$F_x = \begin{pmatrix}\rho v_x\\ \rho v_x^2 + p(\rho)\\ \rho v_x v_y\\ \rho v_x v_z\end{pmatrix}\ .$$

The ρ and m_x components of F_x are familiar from the one dimensional problem. The m_y flux in the x direction is only advection $\rho v_x v_y = v_x m_y$, because pressure across an x interface pushes only in the x direction. The reasoning that led to (19) in one dimension gives the three dimensional quasilinear gas dynamics system (check this). We write partial derivatives explicitly and use subscripts only for components.

$$\begin{aligned}
&\partial_t\rho + \partial_x(\rho v_x) + \partial_y(\rho v_y) + \partial_z(\rho v_z) = 0\\
&\partial_t v_x + \frac{c^2(\rho)}{\rho}\partial_x\rho + v_x\partial_x v_x + v_y\partial_y v_x + v_z\partial_z v_x = 0\\
&\partial_t v_y + \frac{c^2(\rho)}{\rho}\partial_y\rho + v_x\partial_x v_y + v_y\partial_y v_y + v_z\partial_z v_y = 0\\
&\partial_t v_z + \frac{c^2(\rho)}{\rho}\partial_z\rho + v_x\partial_x v_z + v_y\partial_y v_z + v_z\partial_z v_z = 0
\end{aligned}$$

The vector of primitive variables is

$$u(x,y,z,t) = \begin{pmatrix} \rho \\ v_x \\ v_y \\ v_z \end{pmatrix} \ .$$

The quasilinear form above may be written in matrix form

$$\partial_t u + A_x(u)\partial_x u + A_v(u)\partial_y u + A_z(u)\partial_z u = 0 \ .$$

The coefficient matrices are (entries not shown are zero)

$$A_x(u) = \begin{pmatrix} v_x & \rho & & \\ \frac{c^2(\rho)}{\rho} & v_x & & \\ & & v_x & \\ & & & v_x \end{pmatrix} \ , \ A_y(u) = \begin{pmatrix} v_y & & \rho & \\ & v_y & & \\ \frac{c^2(\rho)}{\rho} & & v_y & \\ & & & v_y \end{pmatrix} \ , \ \text{etc.} \tag{21}$$

2.2 Wave modes, hyperbolic equations, acoustics

We examine solutions of the general system (15) under the hypothesis that there is a constant state q_0 and $u(x,t) = q(x,t) - q_0$ is small. If we suppose $u = O(\epsilon)$, then the leading order approximation to (15) is the linear system

$$\partial_t u + \sum_{j=1}^{d} A_j \partial_{x_j} u \ . \tag{22}$$

We write A_j for $A_j(q_0)$. In one dimension, the linearized system becomes

$$\partial_t u + A\partial_x u = 0 \ . \tag{23}$$

The unknown is $u(x,t)$, which has $n = 2$ components $u = (\rho, v)$ if the linearized system came from gas dynamics. The coefficient matrix A is $n \times n$.

We seek simple wavelike solutions of the one dimensional linear problem (23),

$$u(x,t) = f(x - st) \ .$$

This ansatz has $\partial_t u = -sf'$ and $\partial_x u = f'$. The equation becomes

$$sf'(x - st) = Af'(x - st) \ . \tag{24}$$

This is an eigenvalue problem for the coefficient matrix, A. Only real eigenvalues are relevant here, because $x - st$ must be real. The PDE (23) is called *strongly hyperbolic*[4] if A has n linearly independent real eigenvectors with n corresponding real eigenvalues. The eigenvalues, s_m, are the propagation speeds and the eigenvectors r_m are the *wave modes.* Suppose

$$Ar_m = \lambda_m r_m \ , \ \ m = 1, \ldots, n \ .$$

[4] There also are *strictly* hyperbolic, *weakly* hyperbolic, and *symmetrizable* hyperbolic equations.

The right eigenvector matrix (as usual) is

$$R = \begin{pmatrix} | & | & & | \\ r_1 & r_2 & \cdots & r_n \\ | & | & & | \end{pmatrix}$$

The left eigenvector matrix (as usual) is

$$\begin{pmatrix} - & l_1 & - \\ - & l_2 & - \\ & \vdots & \\ - & l_n & - \end{pmatrix} = L = R^{-1} \ .$$

The eigenvalue/eigenvector representation of A may be written (as usual) as

$$LAR = \Lambda \ .$$

Here, Λ is the diagonal matrix of speeds, which are called s_m instead of λ_m because they turn out to be wave speeds. We write the solution in terms of right eigenvectors and expansion coefficients $w_m(x,t)$:

$$u(x,t) = \sum_{m=1}^{n} w_m(x,t)\, r_m \ .$$

This may be written as

$$u(x,t) = \begin{pmatrix} | & | & & | \\ r_1 & r_2 & \cdots & r_n \\ | & | & & | \end{pmatrix} \begin{pmatrix} w_1(x,t) \\ w_2(x,t) \\ \vdots \\ w_n(x,t) \end{pmatrix} = R\, w(x,t) \ .$$

Since $L = R^{-1}$, we find w from u using $w(x,t) = Lu(x,t)$. This formalism makes it easy to find a diagonal form for the PDE (23). Multiply by L from the left, use the relations $RL = I$ and $\partial_t(Lu) = L\partial_t u$:

$$\begin{aligned} L\partial_t u + LARL\partial_x u &= 0 \\ \partial_t w + \Lambda \partial_x w &= 0 \end{aligned}$$

The second line is a collection of n un-coupled scalar PDEs of the form

$$\partial_t w_m + s_m \partial_x w_m = 0 \ . \tag{25}$$

The solution is

$$x_m(x,t) = f_m(x - s_m t) \ ,$$

The signal in mode m is determined from the initial condition using

$$u(x,0) = \sum_{m=1}^{n} f_m(x) r_m \ .$$

Therefore,

$$f_m(x) = l_m u(x,0) \ .$$

The conclusion is that a linear constant coefficient strongly hyperbolic system in one dimension has n real characteristic speeds s and n corresponding wave modes r_m. The solution is a superposition (sum) of waves propagating at the characteristic speeds without changing shape.

For linear gas dynamics, the coefficient matrix is (20). The eigenvalue problem is

$$\det \begin{pmatrix} v-s & \rho \\ \frac{c^2(\rho)}{\rho} & v-s \end{pmatrix} = 0$$

$$(v-s)^2 = c^2$$

$$s_1 = v + c \ , \quad s_2 = v - c \ .$$

The *sound speed* is

$$c = \sqrt{\frac{dp(\rho)}{d\rho}}$$

The gas is not moving if $v = 0$, in which case the propagation speeds are $s = \pm c$. If the gas is moving, then c is the propagation speed relative to the gas. The flow is *subsonic* if $|v| < c$. In this case, wave with speed $s_1 = v - c$ moves left and the wave with speed $s_2 = v + c$ moves right. If $|v| > c$, then the gas is *supersonic*. In this case, either $s_1 < 0$ and $s_2 < 0$ (both waves move left) or $s_1 > 0$ and $s_2 > 0$ (both waves move right).

Two related concepts are important for numerical solution: domain of influence and domain of dependence. The *domain of influence* at time t of a point x_0 at time $t_0 = 0$ is the set of points x where the value $u(x,t)$ changes if we change $u(x_0, 0)$. This is the set of places that where the solution is influenced by the initial data at x_0. In one dimension for linear constant coefficient hyperbolic systems, the domain of influence is just the n points $x_0 + s_m t$. The line $x_m(t) = x_0 + s_m t$ is the m^{th} *characteristic curve*. In this case, characteristic "curves" are actually straight lines because s_m is constant. If the matrix A is not constant (called *variable coefficient*), then the characteristic curves are not straight. Characteristic curves for nonlinear problems (however they might be defined, see a PDE course) are not straight. The *domain of dependence* of a point x at time t is the set of points x_0 so that changing $u(x_0,0)$ changes $u(x,t)$. The solution $u(x,t)$ depends on the initial data in its domain of influence, but nowhere else. In one dimension for linear constant coefficient hyperbolic systems, the domain of influence is the set of points $x - s_m t$. If B is a set of points (a ball or square or whatever), then the domain of dependence of B is the set of domains of dependence of all the points in B.

Wave propagation is more complicated multi-dimensions ($d > 1$). You get some idea what is possible by looking for *plane wave* solutions in direction $\hat{k}$. These are solutions that depend only on the "distance" of a point $x \in \mathbb{R}^d$ in direction $\hat{k}$. In formulas, we seek solutions of the form

$$u(\hat{k}^t x, t) \; , \;\; u(y,t) \; , \; y = \hat{k}^t x \; .$$

Here, y is the scalar variable that tells you how far $x \in \mathbb{R}^d$ is along the direction $\hat{k}$. If you substitute this plane wave ansatz into the linear constant coefficient system (22), you find

$$\partial_t u + A_{\hat{k}} \partial_y u = 0 \; , \;\; A_{\hat{k}} = \sum_{j=1}^{d} \hat{k}_j A_j \; .$$

The matrix $A_{\hat{k}}$ is the coefficient matrix for plane wave propagation in direction $\hat{k}$. A first order system (22) is called *strongly hyperbolic* if the one dimensional plane wave problems are strongly hyperbolic in every direction. That is, it is strongly hyperbolic if for every real direction vector $\hat{k}$, the matrix $A_{\hat{k}} = \sum_j \hat{k}_j A_j$ has n real eigenvalues and no Jordan blocks. The characteristic speeds and modes in direction $\hat{k}$ are the n eigenvalues and eigenvectors

$$s_m(\hat{k}) r_m(\hat{k}) = A_{\hat{k}} r_m(\hat{k}) \; .$$

Domain of dependence and domain of influence are more complicated in multi-dimensions, but neither of them grow faster than the maximum characteristic speed

$$s_{\max} = \max_{\hat{k}} \max_{m} \left| s_m(\hat{k}) \right| \; .$$

That is, the domain of dependence of a point x is contained in a ball of radius $s_{\max} t$ around the point x. If $u(x,0)$ is different from zero only on a set $B \subseteq \mathbb{R}^d$, then the solution at time t is equal to zero outside the domain of influence. In time t a plane wave cannot go farther than $s_{\max} t$. Therefore, if x is in the domain of influence of B, then which is contained in the set

$$\text{dist}(x, B) \leq s_{\max} t \; .$$

3 Difference schemes for one dimension

Suppose there's a time step Δt and a space step Δx. The numerical solution at $x_j = j\Delta x$ and $t_k = k \Delta t$ is

$$U_{j,k} \approx u(x_j, t_k) \; .$$

The numerical solution at time t_k is $U_k = (U_{j,k})$. A finite difference discretization of the PDE (13) or (15) or (23) uses a finite difference approximation of $\partial_x u$ (or an approximation of $\partial_x F(q)$) and a finite difference approximation of $\partial_t u$ to determine U_{k+1} from U_k (and $U_{k-1}, \ldots$ if it is a multi-step method). A semi-discrete approximation involves finite difference approximations to $\partial_x u$ or $\partial_x F$ but not $\partial_t u$. You can make a finite difference approximation from a semi-discrete approximation using an ODE time stepping method. And there are other direct ways.

There is a good chance that a proposed method will be unstable. This would make it useless in practice. Even a stable method can give approximate solutions with *artifacts*, which are features of the numerical solution that are not in the actual solution. Common artifacts are smearing, overshoots, and oscillations. Much of the work in developing a solution strategy goes into analysis that determines stability and artifacts. We approach stability by von Neumann (Fourier) analysis. We approach artifacts also using Fourier analysis, or by using the method of *modified equations*. The task, for serious professionals, is not to find a scheme that converges, but to find a good scheme among the many possibilities.

The "simplest possible scheme" is a good example of the scheme development and analysis procedure. You replace ∂_x with the second order accurate centered difference:

$$\partial_x u \implies \frac{1}{2\Delta x}\left(U_{j+1,k} - U_{j-1,k}\right) \ .$$

You replace $\partial_t u$ with the first order one sided difference:

$$\partial_t u \implies \frac{1}{\Delta t}\left(U_{j,k+1} - U_{j,k}\right) \ .$$

This replaces the PDE (23) with

$$\frac{1}{\Delta t}\left(U_{j,k+1} - U_{j,k}\right) + A\frac{1}{2\Delta x}\left(U_{j+1,k} - U_{j-1,k}\right) = 0 \ .$$

Some algebra turns this into

$$U_{j,k+1} = U_{jk} - \frac{\Delta t}{2\Delta x}A\left(U_{j+1,k} - U_{j-1,k}\right) \ . \tag{26}$$

This scheme is called *forward Euler in time, centered difference in space*. This is a <u>bad scheme</u>. Never use it, except as an example of a scheme you're not using.

The formal order of accuracy for this scheme, and for most schemes, is found by plugging the exact PDE solution into the finite difference equations to find the order of magnitude of the residual. The residual (for this scheme) is defined by (warning: this definition of R is different from the one we used for ODE solving in that no Δt factor is taken out.)

$$u(x_j,t_{k+1}) = u(x_j,t_k) - \frac{\Delta t}{2\Delta x}A(\,u(x_{j+1},t_k) - u(x_{j-1},t_k)) + R_{jk} \ . \tag{27}$$

The calculation of R involves Taylor series in t and x. First, using the PDE,

$$\begin{aligned} u(x_j,t_{k+1}) &= u(x_j,t_k+\Delta t) \\ &= u(x_j,t_k) + \Delta t\,\partial_t u(x_j,t_k) + \frac{1}{2}\Delta t^2\,\partial_t^2 u(x_j,t_k) + O(\Delta t^3) \\ &= u(x_j,t_k) - \Delta t\,A\partial_x u + \frac{1}{2}\Delta t^2\,\partial_t^2 u(x_j,t_k) + O(\Delta t^3) \ . \end{aligned}$$

Next,

$$u(x_{j+1},t_k) - u(x_{j-1},t_k) = 2\Delta x\, \partial_x u(x_j,t_k) + O(\Delta x^3) \ .$$

We substitute these expansions back into the finite difference time step formula (26) and simplify notation by leaving out the x_j, t_k arguments. For example, we write u for $u(x_j,t_k)$, and $\partial_k u$ for $\partial_x u(x_j,t_k)$, etc. Substituting on the left and right sides gives

$$\begin{aligned} u - \Delta t A \partial_x u + \frac{1}{2}\Delta t^2 \partial_t^2 u + O(\Delta t^3) &= u - \Delta t A \partial_x u + O(\Delta t \Delta x^2) + R_{jk} \\ O(\Delta t^2) + O(\Delta t \Delta x^2) &= R_{jk} \ . \end{aligned}$$

If we take out one factor of Δt, there remains a term of the order of Δt and a terms of the order of Δx^2. This scheme is first order accurate in time and second order accurate in space. That is a natural conclusion, given that we used a first order approximation to $\partial_t u$ and a second order approximation to $\partial_x u$. If the scheme were stable (it isn't), it would be first order accurate.

For the von Neumann stability analysis we assume the PDE is strongly hyperbolic and use the wave propagation mode analysis above. In place of $w = Lu$, we have (written in various ways)

$$\begin{aligned} W_k &= LU_k \ \ \text{(approximation at time } t_k\text{)} \\ W_{jk} &= LU_{jk} \ \ \text{(approximation at point } x_j \text{ and time } t_k\text{)} \\ W_{m,jk} &= l_m U_{jk} \ \ \text{(approximation of mode } m \text{ at point } x_j \text{ and time } t_k\text{).} \end{aligned}$$

We multiply the difference scheme (26) by L and calculate as we did for the PDE using $LAR = \Lambda$. The result is

$$W_{j,k+1} = W_{jk} + \frac{\Delta t}{2\Delta x}\Lambda(W_{j+1,k} - W_{j-1,k}) \ .$$

We can look at mode m of this equation (the modes are uncoupled), or we can multiply the difference scheme (26) by l_m and calculate. Either approach gives the evolution of the approximation of mode m:

$$W_{m,j,k+1} = W_{m,jk} + \frac{\lambda_m}{2}\left(W_{m,j+1,k} - W_{m,j-1,k}\right) \ , \qquad (28)$$

with *CFL* (for Courant Friedrichs, Levy) ratio

$$\lambda_m = \frac{s_m \Delta t}{\Delta x} \ . \qquad (29)$$

The scalar difference equation (28) is what you get when you apply the general finite difference approximation (26) to the single mode linear constant speed advection equation (25). In this context, the single mode PDE is sometimes called the *Kreiss equation* in honor of Heinz Kreiss who made important philosophical and technical contributions to stability theory.

We drop the mode index m and let W_k be the numerical approximation at time t_k. The finite difference update formula (28) is a linear update formula for the vector W_k:

$$W_{k+1} = MW_k \; .$$

The matrix M has entries 1 on the diagonal and $\pm\frac{\lambda}{2}$ on the first off-diagonals. The eigenvectors of M are discrete Fourier modes (because M is translation invariant). The corresponding eigenvalues are found by a *symbol* calculation like calculations we've done before. It is traditional to call the discrete wave number θ, so the eigenvector is V_θ, with

$$V_{\theta,j} = e^{i\theta j} \; .$$

We will calculate

$$MV_\theta = m(\theta)V_\theta \; .$$

The eigenvalue for V_θ is $m(\theta)$. This is the *symbol* of the matrix M. Since V_θ is periodic in θ, we can consider $-\pi < \theta \le \pi$ or an equivalent range. We see from (28) that

$$MV_{\theta,j} = V_{\theta,j} + \frac{\lambda}{2}\left(V_{\theta,j+1} - V_{\theta,j-1}\right) \; .$$

The eigenvalue relation is

$$\begin{aligned}
m(\theta)V_{\theta,j} &= V_{\theta,j} + \frac{\lambda}{2}\left(V_{\theta,j+1} - V_{\theta,j-1}\right) \\
m(\theta)e^{i\theta j} &= e^{i\theta j} + \frac{\lambda}{2}\left(e^{i\theta(j+1)} - e^{i\theta(j-1)}\right) \\
m(\theta) &= 1 + \frac{\lambda}{2}\left(e^{i\theta} - e^{-i\theta}\right) \\
m(\theta) &= 1 + i\lambda\sin(\theta) \; . \qquad (30)
\end{aligned}$$

We immediately see that the symbol is outside the unit circle, as

$$\left|m(\theta)\right|^2 = 1 + \lambda^2\sin^2(\theta) > 1 \; , \quad \text{if } \sin(\theta) \neq 0.$$

This implies that the method is unstable. If you program it, the numerical solution will "blow up".

3.1 The Lax Wendroff method

The scheme just described – forward Euler in time and centered differencing in space – has two drawbacks. It is only first order accurate in time and it is unstable. Lax and Wendroff showed that fixing the order of accuracy drawback makes the method stable. This would be called "killing two birds with one stone".

Here is one of the many derivations of this scheme. It starts with a Taylor series in time up to second order (to get second order accuracy in time):

$$u(x, t + \Delta t) = u(x, t) + \Delta t\, \partial_t u(x, t) + \frac{\Delta t^2}{2}\partial_t^2 u(x, t) + O(\Delta t^3) \; .$$

The PDE (23) gives

$$\partial_t u = -A \partial_x u \; .$$

It also can be used to find $\partial_t^2 u$. The derivation uses the fact that $\partial_t \partial_x u = \partial_x \partial_t u$.

$$\begin{aligned}
\partial_t^2 u &= \partial_t \left(\partial_t u \right) \\
&= \partial_t \left(-A \partial_x u \right) \\
&= -A \partial_t \partial_x u \\
&= -A \partial_x \left(\partial_t u \right) \\
&= -A \partial_x \left(-A \partial_x u \right) \\
\partial_t^2 u &= A^2 \partial_x^2 u \; .
\end{aligned}$$

Therefore, if u satisfies the PDE (23), then

$$u(x, t + \Delta t) = u(x,t) - \Delta t \, A \partial_x u(x,t) + \frac{\Delta t^2}{2} A^2 \partial_x^2 u(x,t) + O(\Delta t^3) \; .$$

The Lax Wendroff scheme is to use this formula, with second order centered difference formulas for the derivatives on the right:

$$U_{j,k+1} = U_{jk} - \frac{\Delta t}{2\Delta x} A \left(U_{j+1,k} - U_{j-1,k} \right) + \frac{\Delta t^2}{2 \Delta x^2} A^2 \left(U_{j+1,k} - 2U_{jk} + U_{j-1,k} \right) \; . \tag{31}$$

This scheme is second order accurate, it applies (when suitably generalized) to a wide range of important problems, and it is stable (see below). Even today, decades after it was introduced, this method is the basis for many large scale computations.

The von Neumann analysis for the Lax Wendroff scheme (31) starts with the scheme applied to the Kreiss equation written in terms of the CFL ratio

$$W_{j,k+1} = W_{jk} - \frac{\lambda}{2} \left(W_{j+1,k} - W_{j-1,k} \right) + \frac{\lambda^2}{2} \left(W_{j+1,k} - 2W_{jk} + W_{j-1,k} \right) \; .$$

The symbol is (applying the scheme to $W_{jk} = e^{ij\theta}$ and calculating)

$$m(\theta) = 1 - i\lambda \sin(\theta) + \lambda^2 \left(\cos(\theta) - 1 \right) \; .$$

We look for instability by calculating $\left| m(\theta) \right|^2$. The calculation seems like a quagmire at first, but you can find a way through. We use the notation $c = \cos(\theta)$ and the relation $\sin^2(\theta) = 1 - c^2$. This allows us to express $\left| m(\theta) \right|^2$ as a quadratic polynomial in c.

$$\begin{aligned}
\left| m(\theta) \right|^2 &= \left(1 - \lambda^2 (1 - \cos(\theta)) \right)^2 + \lambda^2 \sin^2(\theta) \\
&= \left(1 - \lambda^2 (1 - c) \right)^2 + \lambda^2 \left(1 - c^2 \right) \\
&= 1 - \lambda^2 + \lambda^4 + 2\lambda^2 \left(1 - \lambda^2 \right) c - \lambda^2 \left(1 - \lambda^2 \right) c^2 \; .
\end{aligned}$$

We look for a min or max by setting the derivative with respect to c to zero. This gives $c_* = 1$ (the minimizer or maximizer). We learn whether $c = 1$ is a min or max from the sign of the c^2 term. This is negative if $|\lambda| < 1$ and positive otherwise. But $|\lambda| > 1$ is not interesting, because the geometric CFL condition requires $|\lambda| < 1$. The Lax Wendroff scheme is exact for $\lambda = 1$ (check this). Therefore, we are only interested in the case where c_* is a local max.

Since $c = \cos(\theta)$, we know that $|c| \leq 1$. The max of the quadratic is at $c = 1$ so the min, in this range, must be at $c = -1$. When $c = -1$, we know $\sin(\theta) = 0$, and $\cos(\theta) - 1 = -2$, so so $m(\theta) = 1 - 2\lambda^2$. This has $|m| \leq 1$ if $|\lambda| < 1$. This reasoning shows that $|m(\theta)| \leq 1$ for all θ if $|\lambda| \leq 1$. That is, von Neumann analysis shows that the Lax Wendroff method is stable.

3.2 First order upwind

4 Stability and convergence

We have seen stability and consistency arguments before. If a scheme is stable, then the error for the scheme is bounded by the residual. The residual is determined by what happens when you apply the finite difference formulas to the exact solution of the PDE. These ideas are not new. We describe them here in a notation that is more general and abstract than before.

An explicit time stepping finite difference method for the hyperbolic system (22) may be written

$$U_{j,k+1} = \sum_l M_l U_{j-l,k} \; . \tag{32}$$

The matrices M_l are formed from the PDE matrices A and the finite difference scheme. They also depend on the time step Δt and the space step Δx. For example in one dimension the Lax Wendroff method has

$$\begin{aligned} U_{j,k+1} = \quad & \left[\frac{\Delta t}{2\Delta x} A + \frac{\Delta t^2}{2\Delta x^2} A^2 \right] U_{j-1,k} \\ + & \left[I - \frac{\Delta t^2}{\Delta x^2} A^2 \right] U_{jk} \\ + & \left[-\frac{\Delta t}{2\Delta x} A + \frac{\Delta t^2}{2\Delta x^2} A^2 \right] U_{j+1,k} \; . \end{aligned}$$

Comparing to the general formula (32), we have

$$\begin{aligned} M_1 &= \frac{\Delta t}{2\Delta x} A + \frac{\Delta t^2}{2\Delta x^2} A^2 \\ M_0 &= I - \frac{\Delta t^2}{\Delta x^2} A^2 \\ M_{-1} &= \frac{-\Delta t}{2\Delta x} A + \frac{\Delta t^2}{2\Delta x^2} A^2 \; . \end{aligned}$$

The residual (also called truncation error, or local truncation error) is defined as before (the explicit Δt factor makes the convergence theorem simpler)

$$u(x_j, t_{k+1}) = \sum_l M_l u(x_{j-l}, t_k) + \Delta t \, R_{jk} \; . \tag{33}$$

As before, you find R_{jk} using Taylor expansions of u about (x_j, t_k) and using the fact that u satisfies the PDE to cancel terms up to a certain point. For hyperbolic equations, we assume that the ratio of Δt to Δx is fixed as $\Delta t \to 0$. The method has order of accuracy p if $R_{jk} = O(\Delta t^p)$. If the method is stable (definition below), then the error also is of order Δt^p.

A scheme is stable in the discrete norm $\|\cdot\|$ if there is a C so that for any k

$$\|U_{\cdot,k}\| \leq C \, \|U_{\cdot,0}\| \; . \tag{34}$$

If we are using von Neumann analysis on the scalar model (the Kreiss equation) (25), then it is the l^2 norm that goes with Fourier analysis. If the symbol satisfies $|m(\theta)| \leq 1$ for all θ, then (because $\widehat{W}_{\alpha,k} = m(\theta_\alpha)^k \widehat{W}_{\alpha,0}$)

$$\sum_j W_{m,jk}^2 \leq \sum_j W_{m,j0}^2 \; . \tag{35}$$

You can apply the scalar l^2 bound component by component (i.e., separately for components $m = 1, \ldots, n$). But the eigenvector matrix R is not an orthogonal matrix (usually). Therefore

$$\|U_{j,k}\|_{l^2}^2 = \sum_{m=1}^n U_{m,j,k}^2 \neq \|W_{j,k}\|_{l^2}^2 = \sum_{m=1}^n W_{m,j,k}^2 \; .$$

However it is an "easy exercise" to see that the norms are equivalent in the sense that there is a κ (relative condition number of norms) so that

$$\frac{1}{\kappa} \|U_{j,k}\|_{l^2} \leq \|W_{j,k}\|_{l^2} \leq \kappa \, \|U_{j,k}\|_{l^2} \; .$$

If the scheme is von Neumann stable for each m, then "it is an exercise" to see that

$$\|U_{\cdot,k}\|_{l^2} \leq \kappa^2 \, \|U_{\cdot,0}\|_{l^2} \; .$$

If a scheme for a hyperbolic system is stable, and if it's in more than one dimension and not a scalar, or if it's more than first order accurate, then it's stable only in norms related to l^2. No scheme like this can be stable in any other l^p norm, in particular, not the max norm or the l^1 norm. It't von Neumann analysis or nothing.

If a scheme is stable and has order of accuracy p, then the error is also order p. The proof is as usual (in different notation, also as usual) Let $E_{j,k} = U_{jk} - u(x_j, t_k)$ be the error. Denote the time step abstractly as S, so

$$U_{\cdot,k+1} = S \, U_{\cdot,k} \; .$$

Then

$$E_{\cdot,k+1} = S\,E_{\cdot,k} + R_{\cdot,k} \ .$$

We fix a time T and look at a time $t_n \leq T$. Assume that $\|R_{\cdot,k}\| \leq r$. Then (we saw this before)

$$\begin{aligned}
\|E_{\cdot,n}\| &\leq \sum_{k=1}^{n} \left\| S^{n-k} R_{\cdot,k} \right\| \\
&\leq C\Delta t \sum_{k=1}^{n} \|R_{\cdot,k}\| \\
&\leq Crn\Delta t \\
&= rCt_n \ .
\end{aligned}$$

If $r = O(\Delta t^p)$ then $\|E_{\cdot,n}\| = O(\Delta t^p)$. This is the Lax version of the proof that stability and accuracy implies convergence.

5 Phase error, disperson and dissipation

Phase error refers to the fact that a complex number that should be z is in fact $e^{i\phi}z$. The real number ϕ is the phase error. Figure 2 shows what can happen if you make extreme phase errors in the Fourier coefficients of a nice function. To make the pictures, I started with a "bump function" (the blue curve) that has the form $f(x) = e^{-x^2/(2r^2)}$, so r sets the scale of the bump. I calculated the Fourier coefficients $\widehat{f}_\alpha$ and multiplied them by a random phase:

$$\widehat{g}_\alpha = e^{i\phi_\alpha} \widehat{f}_\alpha \ .$$

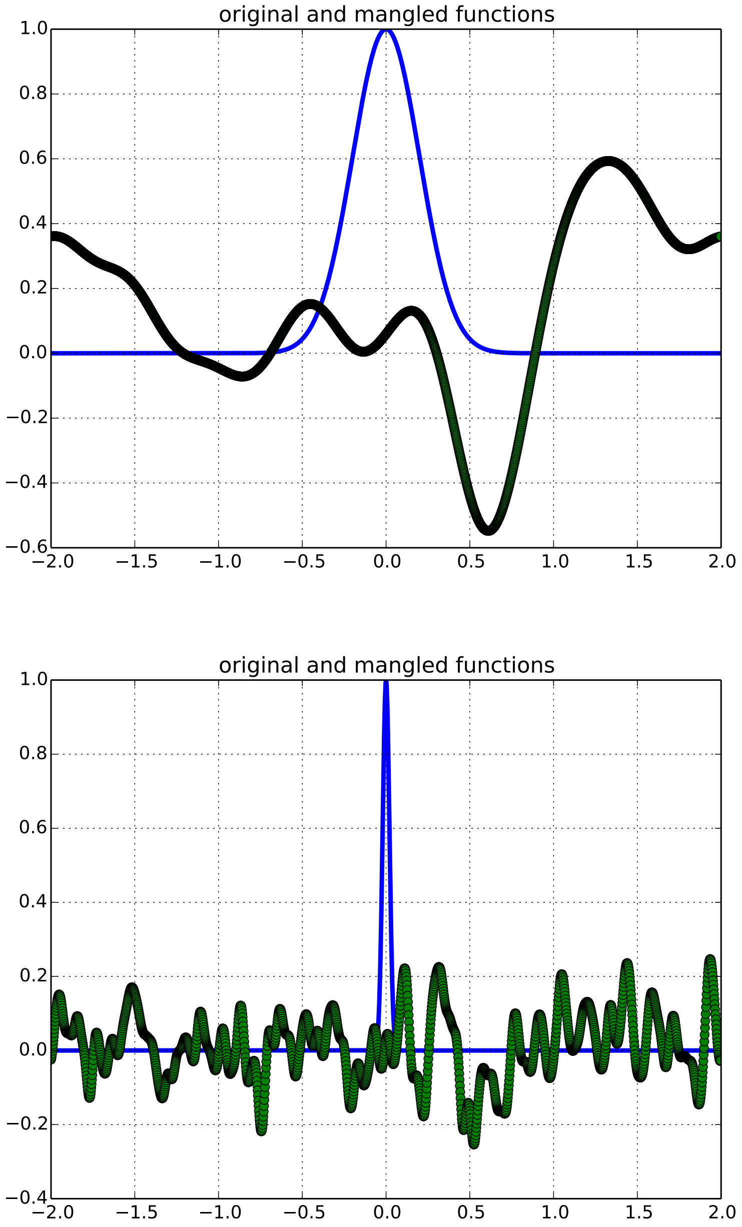
The phases ϕ_α were chosen random (uniformly distributed) in the interval $[0, 2\pi]$, except that I took $\phi_{-\alpha} = 1\phi_\alpha$ so that

$$g(x) = \sum_\alpha \widehat{g}_\alpha e^{ik_\alpha x}$$

is real. I actually did this using the FFT and a uniformly spaced set of grid points x_j.

Roughly speaking, the Fourier coefficients of f are "important" (far enough from zero to effect the behavior of f) if k_α is not much bigger than $1/r$. This is the same as saying the length scale of $e^{ik_\alpha x}$ is not much smaller than r. Since we change the phase but not the magnitude, the same Fourier coefficients are important in g. This means that g moves on a length scale of r but not faster. You can see this in the plots. The "fat" bump function leads to a g that has big wiggles but not small ones. The "thin" bump function has a g with faster wiggles. What people call the *frequency content* of f and g are the same, even though the functions look different.

The difference between f and g is that in f the plane wave phases are "coherent" so that they add up to a coherent structure, which is the bump.


original and mangled functions

original and mangled functions

Figure 2: A local "bump function" (in blue) what what happens when you multiply the Fourier coefficients by random phases (green). The coherent structure of the original function is lost.

When you mangle (change) the phases at random, this coherence is lost. The local structure in f is replaced by a delocalized incoherent g. You also can see that the amplitude of g (the maximum height) is smaller for the thin function than for the fat one. If you believe that g is relatively homogeneous, then the smaller height is a consequence of the fact that f is thinner. The energy of f and g is the same (the Parseval relation), so a thinner f means a g with less energy.

For PDE solving, finite difference schemes introduce phase errors. A coherent structure like a thin bump or a sharp step will get mangled by phase errors. Not as severely as here (see Figure 1), but mangled some.

6 Exercises

1. Consider the linearized gas dynamics problem with coefficient matrices (21) in three dimensions. Show that the characteristic speeds in direction $\hat{k}$ are $v_{\hat{k}}$ with multiplicity 2 and $v_{\hat{k}} \pm c$. Here $v_{\hat{k}} = \sum_j \hat{k}_j v_j$ is the velocity in the $\hat{k}$ direction. Show that the corresponding modes are: (a) two *shear* modes with no density/pressure disturbance but *transverse* velocity disturbance (perpendicular to $\hat{k}$), and (b) two acoustic modes with density disturbance and *longitudinal* velocity disturbances (i.e., velocity disturbances only in the direction of $\hat{k}$.

2. The *Friedrichs scheme* (also called *Lax Friedrichs* because Lax suggested that it be used in practice while Friedrichs only used it to prove something) is a stabilized version of the unstable centered difference in space, forward Euler in time scheme. Instead of U_{jk}, it uses the average $\frac{1}{2}(U_{j-1,k} + U_{j+1,k})$. The scheme is

$$U_{j,k+1} = \frac{1}{2}(U_{j-1,k} + U_{j+1,k}) - \frac{\Delta t}{2\Delta x} A \left(U_{j+1,k} - U_{j-1,k} \right) \ .$$

   (a) Show that this scheme is first order accurate.

   (b) Do the von Neumann analysis to show that the scheme is stable if the PDE is strongly hyperbolic (full family of real linearly independent eigenvectors) and the geometric CFL condition is satisfied.

   (c) Fun fact: the scheme is inconsistent if $\Delta t = \Delta x^2$ as $\Delta x \to 0$. That is, $\|R_k\|$ does not go to zero as $\Delta t \to 0$ with $\Delta x = \Delta t^{1/2}$.

3. The *scalar wave equation* with variable propagation speed $c(x)$ is

$$\partial_t^2 u = c^2(x) \triangle u \ . \qquad (36)$$

   This second order equation can be converted to a first order system, or it can be solved with schemes specific to second order problems. The *leap* *frog* scheme[5] in two dimensions is

$$\frac{1}{\Delta t^2}\left(U_{i,j,k+1} - 2U_{i,j,k} + U_{i,j,k-1}\right)$$
$$= c^2(x_i, y_j)\frac{1}{\Delta x^2}\left(U_{i+1,j,k} + U_{i-1,j,k} + U_{i,j+1,k} + U_{i,j-1,k} - 4U_{i,j,k}\right) \ .$$

(a) Show that the scheme is second order accurate, for smooth solutions.

(b) Assume that c is constant. Show that the PDE (36) has plane wave solutions that move with speed c in any direction $\hat{k}$. (This justifies calling c the "sound speed" or "propagation speed".)

(c) Assume that c is constant. Show that the scheme is stable (assume periodic boundary conditions in space. Use von Neumann analysis in space and recurrence relation type analysis in time.

4. Write a code to solve the variable speed scalar wave equation (36) with propagation speed

$$c(x, y) = 1 - (1 - c_{min})e^{-(x^2+y^2)/(2r^2)} \ .$$

This wave speed has a maximum $c = 1$ for large (x, y) and a minimum $c_{min} > 0$ at $(x, y) = (0, 0)$. The parameter r sets the length scale of the region where waves go slower. Choose initial data a plane wave moving to the right in the x direction but starting to the left of the slow region (you decide the precise parameter values and wave shape). Note that you have to choose initial data $u(x, y, 0)$ and $\partial_t u(x, y, 0)$. These must work together to get the desired plane waving moving right. Do this in a square box with *Neumann* boundary conditions (Carl Neumann is not related to John von Neumann). The are $\partial_x u = 0$ at vertical boundaries and $\partial_y = 0$ at horizontal boundaries. Show that the Neumann boundary condition can be enforced by symmetry like the Dirichlet boundary condition. But the Neunann condition symmetry is even rather than odd symmetry. Use a natural discretization of the Neumann boundary condition that comes from discrete symmetry. Make a movie or some contour plots to show what happens to the plane wave at the slow patch. Do a grid refinement study to demonstrate that you have computed the solution accurately. You will observe *wave focusing* as the wave crosses the slow region if r is enough bigger than the length scale of the plane wave initial data.

[5] The name comes from a kids' game where kids take turns jumping over each other. The scheme does a leap from t_{k-1} to t_{k+1}.