Solving PDE's with FEniCS

Pitfalls

Chapter 27

Introduction to

Automated Modeling

with FEniCS

by L. Ridgway Scott

Pitfalls

Systems of PDEs provide the basis for some of the most powerful models of physical and social phenomena.

The formalism of such models allows a remarkable level of automation of the process of simulating complex systems.

Leveraging this potential for automation has been developed to the greatest extent by the FEniCS Project.

However, using PDE models, and numerical methods to solve them, involves numerous *pitfalls*.

We highlight many of these pitfalls and discuss ways to circumvent them.

Well posed equations

It is simply not possible to provide a solution to all systems of differential equations.

Any given differential equation may be *ill posed*, meaning that it does not make sense to talk about the solution for one reason or another.

At the moment, there is no simple criterion to determine if a system of differential equations is well posed.

Thus, it is not possible to provide software that solves all systems of differential equations automagically.

The first step, then, is to determine if a system being studied is *well posed*.

Failure modes: nonexistence

There are several ways in which a differential equation can fail to be well posed.

The most serious potential pitfall for a differential equation is the lack of a solution regardless of any boundary conditions or initial conditions.

Although it may be quite rare, such a pitfall does exist if one starts trying to solve arbitrary systems of partial differential equations.

We give an example of such an equation.

Fortunately, there is a general criterion that can be applied to identify such behavior.

Failure modes: nonuniqueness

If a physical system is supposed to have a unique state given suitable determining conditions, then a mathematical model having multiple solutions is seriously flawed.

Nonlinear problems can have different solutions that are separated from each other.

But if a PDE had a continuum of solutions, this would probably be bad

Thus local unqiueness is a key feature that should hold for a good model.

All nonlinear models studied here have this property.

Boundary conditions

The typical cause of a system of partial differential equations to have too many solutions is a lack of boundary conditions.

It is not at all trivial to determine what the right number of boundary conditions might be for an arbitrary system of partial differential equations, and getting it wrong could lead to either too many solutions or too few!

We present a case where both of these can be seen.

Continuous dependence

Equally damaging, but often more subtle to detect, is the lack of continuous dependence of the solution on the data of a mathematical model, at least when the physical problem should have this property.

Continuous dependence of the solution on the data is verified in Sobolev spaces for many systems of PDEs.

However, not always required that a physical problem have this property in standard Sobolev spaces.

One such “ill-posed” problem is considered subsequently.

Well posedness

All of these shortcomings of models can be summarized as a lack of *well posedness*.

Coercivity and continuity provide a way to determine if a particular differential equation is well-posed, but more general techniques are also available.

Examples:

- the transport equation,
- non-Newtonian fluid models.

Moreover, some ill-posed models (e.g., the backwards heat equation) are useful.

Numerical stability

Equally difficult to insure, even for well-posed PDEs, is the stability and consistency (and equivalently, convergence) of numerical approximations.

There is no automatic way to define a discrete approximation scheme that will always converge to the solution of a PDE as the approximation is refined.

We discuss various pitfalls and examine in depth the most critical.

However, we make no attempt to be exhaustive.

Ideal case

Ultimately, it is assumed that the reader is trying to solved a well-posed problem with a stable and consistent numerical approximation.

The language of variational formulations of differential equations is a powerful approach that allows a simple proof of well-posedness in many cases.

Moreover, it leads to stable and consistent numerical schemes via the Galerkin method.

In this way, finite element methods, spectral methods, spectral element methods, etc., can be derived and analyzed with respect to stability and consistency.

Right-sizing BCs

Differential equations typically have too many solutions of the equations themselves to specify a solution in any reasonable sense.

A unique solution, required by most physical models, is typically determined by boundary conditions and, for time-dependent problems, initial conditions.

Consider the Laplace equation

$$-\Delta u = f \tag{1}$$

Then for $f \equiv 0$ the solutions are *harmonic* functions, and the real part of any complex analytic function in the plane (in two space dimensions) is harmonic.

How many BCs?

For any solution to (1), we can get another by adding any harmonic function.

Thus there are way too many solutions to (1) without any further restrictions.

This is an example of extreme nonuniqueness, where a continuum of solutions exist.

But specifying the value of u on the boundary of some open set Ω makes the solution of (1) unique in Ω:

$$-\Delta u = f \text{ in } \Omega, \qquad u = g \text{ on } \partial\Omega, \tag{2}$$

has a unique solution, under suitable smoothness conditions on f, g and the boundary $\partial\Omega$.

How many BCs?

There is no unique type of boundary condition that is appropriate for a given system of differential equations.

For example, the system

$$-\Delta u = f \text{ in } \Omega, \qquad \frac{\partial u}{\partial n} = g \text{ on } \partial\Omega \tag{3}$$

also has a solution provided that

$$\int_\Omega f(\mathbf{x})\, d\mathbf{x} + \oint_{\partial\Omega} g(s)\, ds = 0. \tag{4}$$

For any solution u to (3), $u + c$ is also a solution for any constant c, but that is the limit of nonuniquenss:

solutions modulo constants are unique.

Too many BCs?

If some is good, then one might think more is better.

However, it is easy to see that the system of equations

$$\begin{aligned} -\Delta u &= f \text{ in } \Omega \\ u = g_0 \text{ on } \partial\Omega, &\quad \frac{\partial u}{\partial n} = g_1 \text{ on } \partial\Omega \end{aligned} \tag{5}$$

has too many boundary conditions.

Since the condition $u = g_0$ on $\partial\Omega$ already uniquely determines the solution u, it will only be a miracle that $\frac{\partial u}{\partial n} = g_1$ also holds on $\partial\Omega$.

More precisely, there is a linear mapping A defined on functions on $\partial\Omega$ such that (5) has a solution if and only if $g_1 = Ag_0$ (exercise).

How many BCs?

Similarly, the system

$$\begin{aligned} -\Delta u &= f \text{ in } \Omega \\ \nabla u &= \mathbf{g} \text{ on } \partial\Omega \end{aligned} \tag{6}$$

is over-specified.

It is closely related to (5) if we observe that the second equation says that the tangential derivative of u is equal to that of g_0.

The over-determined boundary value problem (6) appears in a non-local compatibility condition for the Navier-Stokes equations.

Numerical Stability

The simplest differential equation to solve is an ordinary differential equation

$$\frac{du}{dt} = f(u,t) \tag{7}$$

with initial value

$$u(0) = u_0 \tag{8}$$

where we solve on some interval $[0, T]$.

The definition of the derivative as a limit of difference quotients suggests a method of discretization:

$$\frac{du}{dt}(t) \approx \frac{u(t+\Delta t) - u(t)}{\Delta t} \tag{9}$$

where Δt is a small positive parameter.

This suggests an algorithm for generating a sequence of values $u_n \approx u(n\Delta t)$ given by (for example)

$$u_n = u_{n-1} + \Delta t f(u_n, t_n) \tag{10}$$

where $t_n = n\Delta t$.

The algorithm (10) is called the **implicit Euler** method, and it can be shown that it generates a sequence with the property that

$$|u(t_n) - u_n| \leq C_{f,T}\Delta t \quad \forall t_n \leq T \tag{11}$$

provided that we solve the implicit equation (10) for u_n exactly and we compute with exact arithmetic.

Numerical error

The issue of solving the nonlinear equation at each step is important but not a show-stopper.

However, requirement of using finite-precision arithmetic means that best error behavior we could expect is

$$|u(t_n) - u_n| \leq C_{f,T}\Delta t + n\epsilon \quad \forall t_n \leq T \tag{12}$$

where ϵ measures the precision error that occurs at each step in (10).

Useful to re-write (12) using the fact that $n = t_n/\Delta t$ as

$$|u(t_n) - u_n| \leq C_{f,T}\Delta t + \frac{t_n\epsilon}{\Delta t}. \tag{13}$$

(13) shows that the error reaches a minimum and cannot be reduced by reducing Δt.

One way to increase the accuracy in (12) is to use a more accurate approximation of the derivative than (9), such as given by the backwards differentiaion formulæ (BDF)

$$\frac{du}{dt}(t) \approx \frac{1}{\Delta t} \sum_{i=0}^{k} a_i u_{n-i} \tag{14}$$

where the coefficients $\{a_i \ : \ i = 0, \ldots k\}$ are chosen so that (14) is exact for polynomials of degree k.

The BDF for $k = 1$ is the same as implicit Euler.

Higher order schemes

Using the approximation (14), we get an algorithm of the form

$$\sum_{i=0}^{k} a_i u_{n-i} = \Delta t f(u_n, t_n) \tag{15}$$

which can be solved for u_n provided $a_0 \neq 0$.
In this case, the final error estimate would be

$$|u(t_n) - u_n| \leq C_{f,T,k} \Delta t^k + \frac{t_n \epsilon}{\Delta t}. \tag{16}$$

Ultimate accuracy is still limited, but smaller absolute errors (with larger Δt) can be achieved with higher values of k.

Round-off error

For example, suppose that

- $\epsilon = 10^{-6}$ (which corresponds to single precision on a 32-bit machine)
- $T = 1$ and
- (for the sake of argument) $C_{f,T,k} = 1$.

Then with implicit Euler ($k = 1$) the smallest error we can get is 10^{-3} with $\Delta t = 10^{-3}$.

But with $k = 2$ we get an error of size 10^{-4} with $\Delta t = 10^{-2}$.

Not only is error smaller but less work needs to be done to achieve it.

BDF methods

In practice, the constant $C_{f,T,k}$ would depend on k and the exact error behavior would likely be different in detail, but the general conclusion that a higher-order scheme may be better still holds.

The BDF methods for $k = 2$ and 3 are extremely popular schemes.

We see that higher-order schemes can lead to more managible errors and potentially less work for the same level of accuracy.

Thus it seems natural to ask whether there are limits to choosing the order to be arbitrarily high.

BDF methods

Unfortunately, not all of the BDF schemes are viable.

Beyond degree six, they become **unconditionally unstable**.

Let us examine the question of stability via a simple experiment.

Suppose that, after some time T_0, it happens that $f(u,t) = 0$ for $t \geq T_0$.

Then the solution u remains constant after T_0, since $\frac{du}{dt} \equiv 0$.

What happens in the algorithm (15) is that we have

$$\sum_{i=0}^{k} a_i u_{n-i} = 0 \tag{17}$$

for $n \geq T_0/\Delta t$.

But, this does not necessarily imply that u_n would tend to a constant.

Let us examine what the solutions of (17) could look like.

Consider the sequence $u_n := \xi^{-n}$ for some number ξ.

Root condition

Plugging into (17) we find

$$0 = \sum_{i=0}^{k} a_i \xi^{-n+i} = \xi^{-n} \sum_{i=0}^{k} a_i \xi^i \tag{18}$$

If we define the polynomial p_k by

$$p_k(\xi) = \sum_{i=0}^{k} a_i \xi^i \tag{19}$$

we see that we have a null solution to (17) if and only if ξ is a root of p_k.

If there is a root ξ of p_k where $|\xi| < 1$ then we get solutions to (17) which grow like

$$u_n = \xi^{-n} = \left(\frac{1}{\xi}\right)^{t_n/\Delta t}. \tag{20}$$

Not only does this blow up exponentially, the exponential rate goes to infinity as $\Delta t \to 0$.

This clearly spells disaster.

On the other hand, if $|\xi| > 1$, then the solution (20) goes rapidly to zero, and more rapidly as $\Delta t \to 0$.

Numerical Stability

For roots ξ with $|\xi| = 1$ the situation is more complicated, and $\xi = 1$ is always a root because the sum of the coefficients a_i is always zero.

Instability occurs if there is a multiple root on the unit circle $|\xi| = 1$.

In general, one must consider all complex (as well as real) roots ξ.

Given this simple definition of the general case of BDF, it is hard to imagine what could go wrong regarding stability.

Numerical Stability

Unfortunately, the condition that $|\xi| \geq 1$ for roots of $p_k(\xi) = 0$ restricts k to be six or less for the BDF formulæ.

In particular, $p_7(0.0735 \pm \iota\, 0.9755) = 0$, and the complex modulus $|0.0735 \pm \iota\, 0.9755| \approx 0.9783 < 1$.

The inf-sup condition for mixed methods is another type of numerical stability condition that effects the choice of finite element spaces suitable for fluid flow problems.

BDF roots

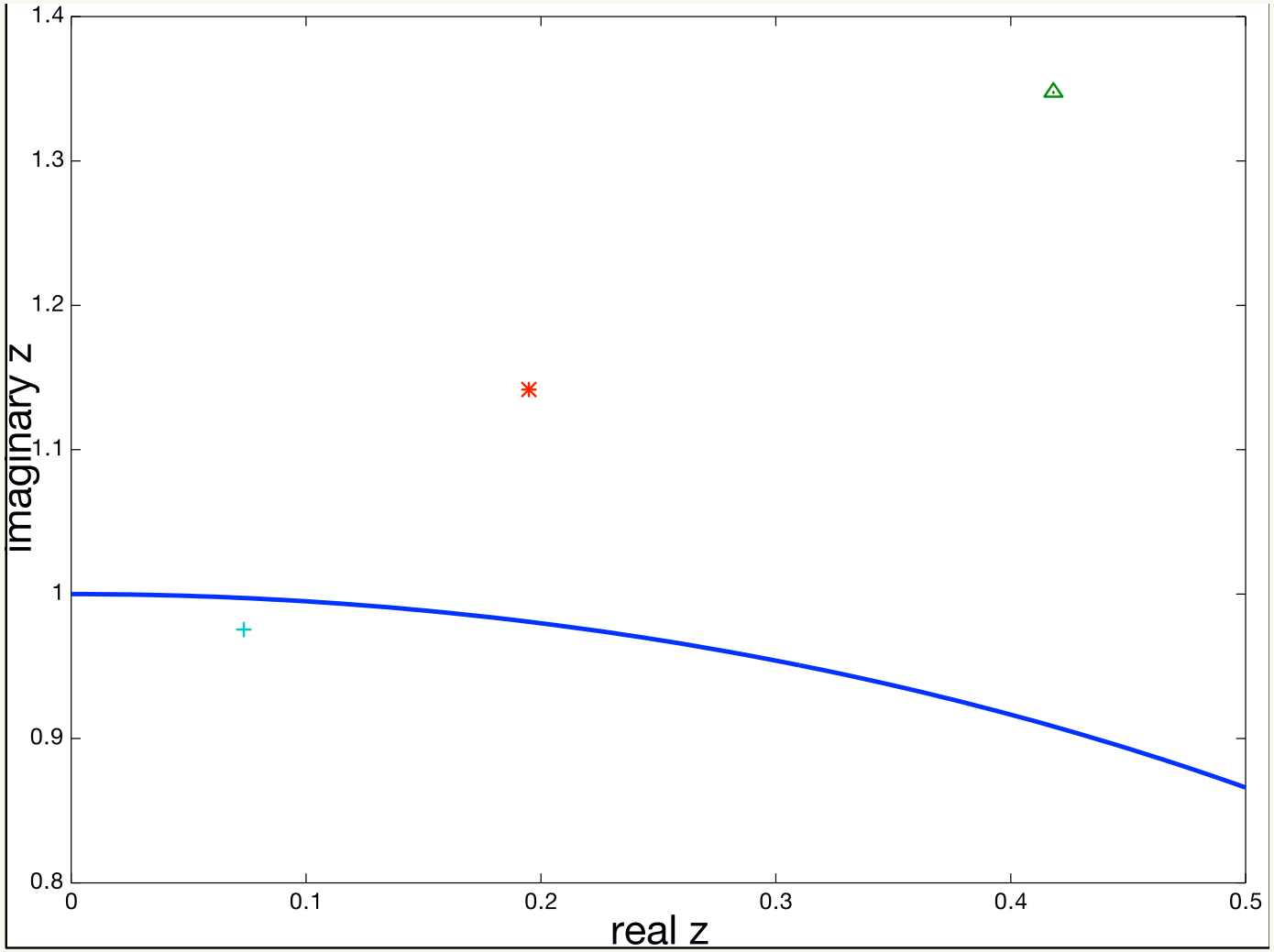

Figure 1: Roots of polynomials (19) with the smallest modulus are plotted for degrees $k = 5$ (triangle), $k = 6$ (asterisk), and $k = 7$ (plus). The solid line indicates the unit circle in the complex plane.

The right box

The variational formulation of PDEs is not only a convenient way to describe a model, it also provides a way to ensure that a model is well posed.

One of the critical ingredients in the variational formulation is the space of functions in which one seeks the solution.

We can see that this is just the right size to fit the needs of most modeling problems.

We can think of the space of functions in the variational formulation of PDEs as a box in which we look for a solution.

The right box

If the box is too big,

- we might get spurious solutions.

If the box is too small,

- then we may get no solutions.

So we need a box that is **just right** to have the right number of solutions.

A box too small

A box too small is easy to describe.

The space C^m of functions whose derivatives up through order m are continuous is natural if the number of derivatives in the PDE is less than or equal to m.

Then all derivatives are classically defined, and the equation itself makes sense as a equation among numbers at all points in the model domain.

However, many problems do not fit into this box.

For example, when the geometry of the boundary of a domain in two-dimensions has an interior angle that is greater than π (and hence the domain is not convex), a singularity arises even for the Laplace equation.

A box too big

A box too big can be described already in one-dimension.

Cantor “middle-thirds” function defined as follows on $[0, 1]$:

$$C(x) := \begin{cases} \frac{1}{2} & \frac{1}{3} \leq x < \frac{2}{3} \\ \frac{1}{2}C(3x) & 0 \leq x < \frac{1}{3} \\ \frac{1}{2}(1 + C(3x - 2)) & \frac{2}{3} \leq x < 1 \end{cases} \tag{21}$$

Recursive nature of definition means that it is easily computed.

Any programming language with recursion is suitable.

A box too big

By definition, the derivative of C is zero at almost every point (a notion made precise by Lebesgue measure theory [2].)

This is problematic, since we expect solutions of linear PDEs to be unique in most cases.

But the simplest equation $u' = f$ would have $u + aC$ as a solution for any given solution u, for any real number a.

This would violate our principle that solutions should be isolated.

The Cantor function

However, derivative of C not a Lebesgue integrable function.

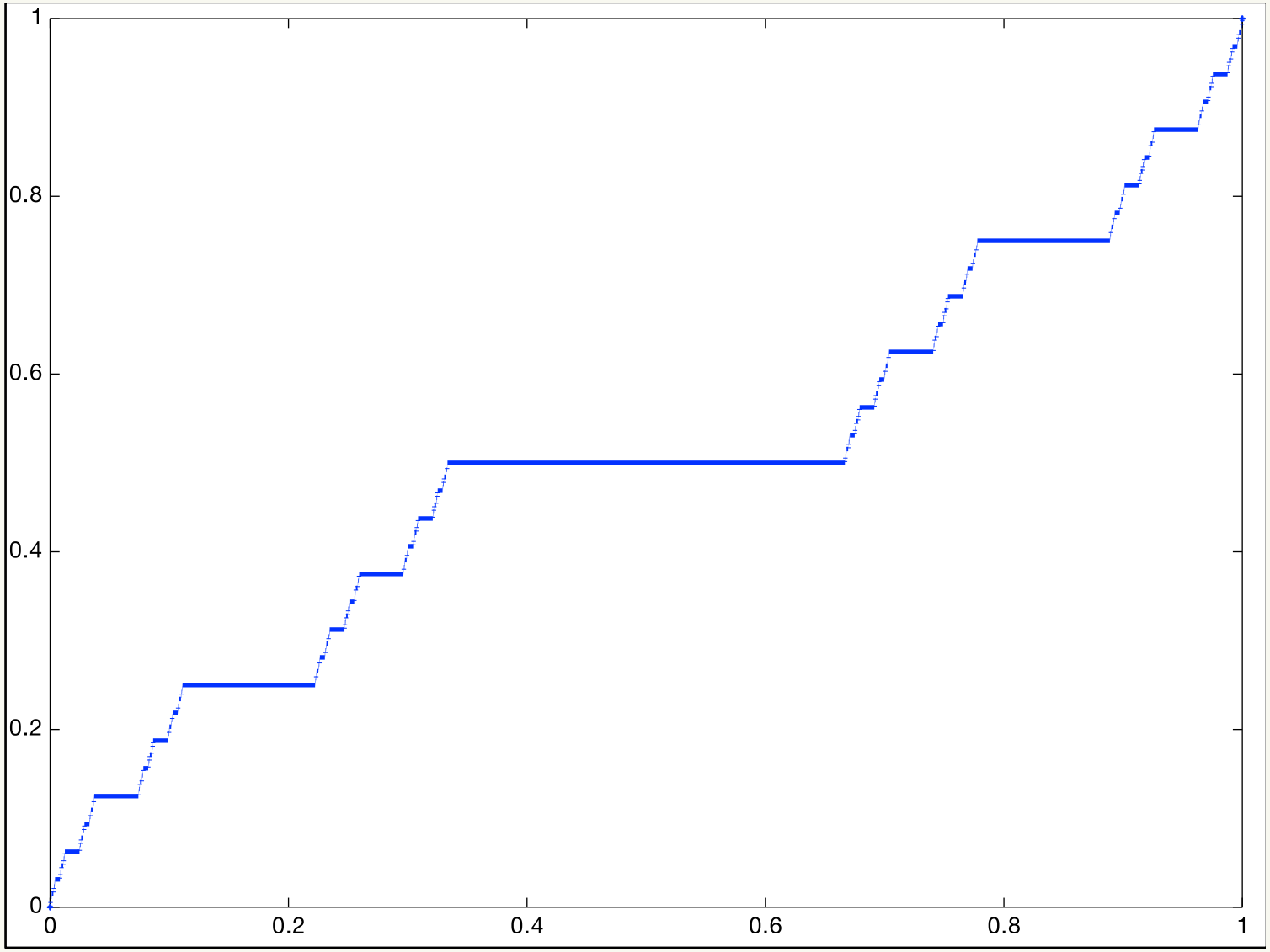

Figure 2: Cantor function (21).

The right box

Thus the choice of spaces V of functions whose derivatives are Lebesgue integrable functions provides just the right size box in the variational formulation.

These spaces are the workhorses of the variational approach to PDEs.

These are called Sobolev spaces, and they are reviewed at the end of the notes.

An equation such as $\Delta u = f$ presumes that it is possible to specify u, at least locally, by giving a combination of its derivatives ($u_{,11} + u_{,22} + \cdots$).

What is it that makes this possible?

That is, should we always assume that an arbitrary combination of derivatives can be specified without any internal consistency required?

It is easy to see one kind of partial differential equation that would make little sense:

$$\frac{\partial^2 u}{\partial x\, \partial y} = -\frac{\partial^2 u}{\partial y\, \partial x}, \tag{22}$$

Lewy equation

Since we know that for smooth functions, the order of cross derivatives does not matter, we see that (22) has an internal contradiction.

Thus (22) corresponds to an equation of the form $t = -t$ and has only the zero solution.

There are some differential equations that simply have no solution even locally, independent of any boundary conditions.

A famous example is due to Hans Lewy:

Lewy equation

Lewy's equation is

$$\frac{\partial u}{\partial x_1} - \iota \frac{\partial u}{\partial x_2} + 2(\iota x_1 - x_2)\frac{\partial u}{\partial x_3} = f, \tag{23}$$

where ι is the imaginary unit, $\iota = \sqrt{-1}$.

Then for most infinitely differentiable functions f there is no solution of this equation in *any* open set in three-space.

Note that this has nothing to do with boundary conditions, just with satisfying the differential equation.

This equation is a complex equation ($\iota = \sqrt{-1}$) but it can be written as a system of two real equations for the real and imaginary parts of u respectively.

Multi-index notation

There is a general condition that must be satisfied in order that linear partial differential equations have a local solution.

This condition is known as the **local solvability condition**.

To explain the condition, we need to introduce some notation.

Let $D = -\iota\left(\frac{\partial}{\partial x_1}, \ldots, \frac{\partial}{\partial x_j}, \ldots, \frac{\partial}{\partial x_d}\right)$ stand for the vector of complex partial derivatives, and

let $\alpha = (\alpha_1, \ldots, \alpha_j, \ldots, \alpha_d)$ be a **multi-index** (i.e., a vector of non-negative integers), so that

More index notation

$$D^\alpha u := (-\iota)^{|\alpha|} \frac{\partial^{\alpha_1}}{\partial x_1^{\alpha_1}} \cdots \frac{\partial^{\alpha_j}}{\partial x_j^{\alpha_j}} \cdots \frac{\partial^{\alpha_d}}{\partial x_d^{\alpha_d}} u \,, \tag{24}$$

where $|\alpha| := \alpha_1 + \cdots \alpha_j + \cdots \alpha_d$.

For any d-dimensional variable ξ, we can form the monomial

$$\xi^\alpha := {\xi_1}^{\alpha_1} \cdots {\xi_j}^{\alpha_j} \cdots {\xi_d}^{\alpha_d} \tag{25}$$

so that D^α is the same as ξ^α with the substitution $\xi_j = -\iota \partial / \partial x_j$.

In this notation, the Lewy equation (23) becomes

$$-\iota D_1 u + D_2 u - 2(x_1 + \iota x_2) D_3 u = f. \tag{26}$$

More index notation

The reason for the factor $-\iota$ in the definition of D is so that the Fourier transform of D works out nicely; if $\hat{u}$ denotes the Fourier transform of u, then

$$\widehat{D^\alpha u}(\xi) = \xi^\alpha \hat{u}(\xi).$$

Suppose that we want to consider linear partial differential equations of the form

$$P(\mathbf{x}, D)u = \sum_{|\alpha| \leq m} a_\alpha(\mathbf{x}) D^\alpha u = f \tag{27}$$

for some f.

Symbol of a PDO

We can form the corresponding **symbol** $P(\mathbf{x}, \xi)$ of the linear partial differential operator

$$P(\mathbf{x}, \xi) = \sum_{|\alpha| \leq m} a_\alpha(\mathbf{x}) \xi^\alpha \,, \tag{28}$$

Define the **principal part** of the symbol, P_m, by

$$P_m(\mathbf{x}, \xi) = \sum_{|\alpha| = m} a_\alpha(\mathbf{x}) \xi^\alpha \,, \tag{29}$$

and correspondingly its complex conjugate $\overline{P}_m$ by

$$\overline{P}_m(\mathbf{x}, \xi) = \sum_{|\alpha| = m} \overline{a_\alpha(\mathbf{x})} \xi^\alpha \,. \tag{30}$$

Also define the following partial derivatives of the principal symbol:

$$P_m^{(j)}(\mathbf{x},\xi) := \frac{\partial P_m}{\partial \xi_j}(\mathbf{x},\xi)\ , \quad P_{m,j}(\mathbf{x},\xi) := \frac{\partial P_m}{\partial x_j}(\mathbf{x},\xi) \tag{31}$$

and define their complex conjugates analogously.

Finally, define the **commutator** $C_{2m-1}(\mathbf{x},\xi)$ of the principal part of the symbol via

$$C_{2m-1}(\mathbf{x},\xi) = \iota \sum_{j=1}^{d} \Big(P_m^{(j)}(\mathbf{x},\xi)\overline{P}_{m,j}(\mathbf{x},\xi) - \overline{P}_m^{(j)}(\mathbf{x},\xi) P_{m,j}(\mathbf{x},\xi) \Big). \tag{32}$$

No solutions

The commutator is a polynomial of degree $2m-1$ in ξ with real coefficients.

Finally, the local solvability theorem.

Theorem 0.1 *If the differential equation (27) has a solution in a set Ω for every smooth f that vanishes near the boundary of Ω, then*

$$C_{2m-1}(\mathbf{x},\xi)=0 \ \textit{ for all } \ \xi \ \textit{ and all } \ x\in\Omega \ \textit{ such that } \ P_m(\mathbf{x},\xi)=0. \tag{33}$$

Meaning: if (33) does not hold, there are no (even local) solutions.

No solutions

Here, notion of “solution” is very weak; need not be a smooth solution.

Thus the result provides a very stringent condition on the symbol in order to expect any sort of solution at all.

For a complete description of this and other examples see [3]; see [1] for more recent results and references.

The most striking feature of the local solvability condition (33) is that it is a “closed” condition.

Otherwise said, “non-solvability” is an open condition:

if $C_{2m-1}(\mathbf{x},\xi) \neq 0$ then small perturbations would not be expected to make it vanish.

No solutions

Moreover, even if (33) holds for one set of coefficients a_α, it may fail to hold for a small perturbation.

Finally, we will be interested in nonlinear partial differential equations; if these have a solution, then the solution can be viewed as solutions to linear partial differential equations with appropriate coefficients (which depend on the particular solution).

No solutions

Despite the pessimism implied by the local solvability condition (33), we will see that there are indeed broad classes of nonlinear partial differential equations which can be proved to have solutions.

But this should not be taken for granted in general.

In proposing a new model for a new phenomenon,

the first question to ask is whether it

makes sense at this most fundamental level.

References

[1] Antonio Bove and Tatsuo Nishitani. Necessary conditions for local solvability for a class of differential systems. *Communications in Partial Differential Equations*, 27:1301 – 1336, 2002.

[2] S. C. Brenner and L. R. Scott. *The Mathematical Theory of Finite Element Methods*. Springer-Verlag, third edition, 2008.

[3] Lars Hörmander. *Linear Partial Differential Operators*. Springer-Verlag, 1969.