Design of efficient numerical schemes and qualitative properties of solutions of ODEs and PDEs

M. VANNINATHAN
IISc–TIFR Mathematics Programme, TIFR Centre, IISc Campus, Bangalore 560 012.

Received on February 19, 1998.

Abstract
The paper presents important steps in the approximation of solutions of differential equations and examines crucial issues involved in the process. In particular, the role played by the regularity of solutions in their approximation is emphasized. After highlighting the progress achieved, it speculates on what may be in store for the future.

Keywords: Differential equations, partial differential equations, approximation, numerical analysis.

1. Introduction
Ordinary differential equations (ODE) and partial differential equations (PDE) arise as models in various situations (physical, natural, etc.). We need not dwell on this point as it is too well known. Most such equations are nonlinear. Linear equations also appear either directly from the modeling processes (e.g. quantum mechanics) or as an approximation to nonlinear ones via the so-called linearization scheme. Thus, ‘solving’ them is imperative if one wishes to understand the various phenomena which they are supposed to incorporate and make quantitative predictions based on them. Looking into the past history, we notice that five broad approaches have been followed to achieve this goal and get a first-hand information on the problem at hand. They are:

(i) obtaining explicit solutions exploiting invariance of the problem (or) by some ingenious ways,
(ii) obtaining approximate and new solutions by perturbation methods around a known solution,
(iii) obtaining asymptotic solutions by the application of wide spectrum of techniques from applied mathematics,
(iv) physical experiments,
(v) numerical computations.

World War II was a watershed in this process. It is indeed impossible to exaggerate the extent to which modern applied mathematics has been shaped and fueled by the availability of computers in the post-war period. These electronic machines offer an alternative to physical experiments which have become either expensive or impossible at times. Let me quote the pro-
Phetic words of John von Neumann: "really efficient high speed computing devices may, in the field of nonlinear PDEs as well as in many other fields which are now difficult or entirely denied of access, provide us with those heuristic hints which are needed in all parts of mathematics for genuine progress... many branches of both pure and applied mathematics are in great need of computing instruments to break the present stalemate created by the failure of purely analytical approach to nonlinear problems".

Having mentioned the importance of computation, let us cite a few remarkable achievements.

Fermi, Pasta and Ulam discovered the remarkable almost periodic behavior of the vibrations of nonlinear chains, and Kruskal and Zabusky the generation and interaction of solitons. The complete integrability of the Toda Lattice became plausible through very careful numerical calculations of Joe Ford. For a description of these developments, see Toda.\(^1\) Mitchell Feigenbaum\(^2,3\) discovered his remarkable universal laws on iterations by analyzing numerical experiments. Numerical studies led Lorenz\(^4\) to the concept of strange attractor; the understanding of chaotic behavior of simple dynamical systems coexisting with islands of stability has been much enhanced by numerical studies.\(^5\)

Our objective here is to examine the various issues involved in this kind of calculations, highlight the progress made and raise a few questions about the future course of action.

The following are the steps which are usually involved in the computation of solutions of ODE/PDE:

- discretization
- solving discrete equations
- convergence

2. Discretization

The rapid rise of computing was made possible by striking improvements and novel ideas introduced in the discretization of the equations. Let us name a few commonly used techniques and discuss them separately.

(i) **Finite-difference method** (FDM): Crudely put, this amounts to replacing differential operators in the equation by difference quotients. Here is a partial list of mathematical ideas that have borne fruit.

The methods of alternating direction and fractional step initiated by Peaceman, Rachford, Douglas, Yanenko and Strang are used universally.\(^5-9\)

High-order difference schemes developed by Lax and Wendroff, Mac Cormack and others have been particularly effective in meteorology and are of use for calculating any smooth flow.\(^9\)

**Implicit methods:** A variety of ideas, introduced by Hirt, Warming, Beam, Hardned, and others have proved effective in both imcompressible and compressible flow calculations, as well as in magnetofluid dynamics.\(^5,8\)
The method of complex coordinates, developed by Garabedian, allows a unified treatment of sub- and supersonic regimes in flows and has been used successfully to design shockless transonic airfoils, compressor blades and turbines.

The challenge of calculating flows with shocks has generated a number of mathematical ideas. One line of thought, shock capturing, started with von Neumann and Richtmyer's notions of artificial viscosity; to this was added the notion of difference equation in conservation form with numerical flux function and Godunov's idea of threading together solutions of the Riemann initial value problem. Glimm's method is also based on solutions of Riemann problems; it employs a sequence of random parameters and has the virtue of calculating entropy production more realistically than other methods that employ an artificial viscosity. Chorin noted that this feature of the method makes it a good candidate for calculating reacting flows.

The far-reaching modification that Van Leer, Colella and Woodward have made of Godunov's method has resulted in astonishingly accurate calculations of very complicated patterns of shocks.

The method of flux-corrected transport, developed by Boris and Book and artificial compression, developed by Harten, are successful in resolving discontinuities, both contact and shock, that develop in flows of compressible, multimedia fluids.

Jameson has developed intricate and rapidly convergent iterative techniques for the calculation of steady transonic flow fields with shocks around complicated aerodynamic shapes.

An alternative to shock capturing is shock tracking, pioneered by Moretti, Richtmyer, Lazarus, greatly advanced by Glimm and Mc Bryan and recently by Colella.

While modeling the physical process, some parameters may be neglected and assumed to be zero. One way of suggesting an approximation involves the reintroduction of the parameters and this process is called unfolding. Novel schemes are born this way. One example is the notion of artificial viscosity mentioned earlier. Another one is mean free path which gives rise to kinetic schemes.

The conclusion of all these is that one has to choose a suitable one from the myriad of possible schemes. If not satisfied, the art is to introduce modifications and novelties to suit one's needs. But one thing is certain; straightforward schemes will not suffice to achieve high accuracy with minimal cost.

(ii) Finite-element method (FEM): This is a significant alternative to FDM and can cope with complicated geometrical configurations. It is based on more rigorous mathematical footing and deals with what is called a weak formulation of the given problem. There may be several formulations of the problem (primal, dual, hybrid, etc.). The idea is then to approximate the spaces involved in the formulation by finite-dimensional spaces consisting of piece-wise polynomials. One of the virtues is that they have a canonical basis of functions with small support. The construction of such spaces uses triangulation of the domain which can be a very nontrivial geometrical problem depending upon the configuration.
(iii) **Spectral method (SM):** The procedure here is more or less the same as in FEM except that we now use trigonometric polynomials (or other more general eigenfunctions associated with certain operators) instead of the usual algebraic ones. The advantage is that one achieves very high accuracy in regions where the solution is smooth.

Spectral methods, pioneered by Leith and put on map by Gottlieb and Orszag, have been made more efficient by the use of fast Fourier transform introduced by Cooley and Tukey; they are of use in calculating space-periodic flows, both smooth and rough.

(iv) **Wavelet method (WM):** Instead of trigonometric basis, we work with wavelet basis. It is a virtue of wavelet basis that the representation is very lacunar; the wavelet coefficients in regions of smoothness are negligible and so we will be able to localize important features of the solution like shocks. Another important advantage is the enormous gain in the storage requirements. Many standard operators are almost diagonalized in wavelet basis and hence a rapid convergence of algorithms is expected. This is yet to be confirmed. However, there are some disadvantages of treating nonlinear equations. Indeed, the computation of wavelet coefficients of nonlinear terms is not straightforward; it has to be done in the physical space. These issues as well as comparison between WM and other methods are discussed in an accompanying paper appearing in this issue.

(v) **Particle methods (PM):** This is based on the approximation of functions by delta measure at a finite number of points. One finds that these points evolve with characteristic speed and the corresponding coefficient satisfies a transport equation. This method, introduced by Harlow, has been very effective in problems where two different media are in contact and exert a force on each other, such as in high-velocity impact.

The vortex method (i.e. particle method applied to vorticity equation) of Chorin generates and propagates vorticity in a very original fashion. The method has been very effective in calculating effects that depend sensitively on vorticity, such as drag at high Reynolds numbers. The method has been used by Peskin to calculate flows around valves, real and artificial, in the beating heart.

Let us point out one important unsettled issue in this domain. Does the vorticity blow-up in classical norms for the 3d Euler incompressible flow? Numerical experiments of Chorin suggest that it is so.

(vi) **Finite-volume method (FVM):** After triangulation of the domain, the given set of equations is integrated on each element of the triangulation. Exploiting the divergence form of the equation, we get an integral identity on the boundary of element. Via a numerical integration scheme, we are then led to a set of algebraic equations which have to be solved. We point out that the theoretical basis of FVM is not as well developed as in the case of FEM.

(vii) **Other methods:** There are other techniques which combine the ones listed earlier and which are based on different formulations of the problem. Let us cite a few: domain decomposition method, parallel computing, nonlinear least squares, operator decomposition, adaptive methods, etc.

The multigrid method suggested by Federenko and Bahvalov and developed by Brandt, is an extremely rapid method for solving elliptic equations with variable coefficients.
The capacitance matrix method of Widlund exploits the fast algorithm for solving Poisson's equation in rectangles, developed by Bunemann and Hockney, to solve Poisson's equation and related ones in more general geometries.

3. Solving discrete equations

We are not going to dwell on this point in this brief write-up. We merely point out that there are plenty of clever algorithms (both direct and iterative) to solve the system of algebraic equations resulting from discretization. One of the main concerns is to minimize the number of operations and thereby computer time. One has to devise ways of exploiting the sparseness of the system which is usually large. We should worry about the condition number of the system. This is essential if we want to check the growth of round-off errors.

4. Convergence

A straightforward discretization of given equations may not converge at all. Even if it does, the limit may not be the solution we seek. This is especially true in the case of nonlinear equations which possess multiple solutions due to the presence of instabilities. Therefore, the question arises: how to believe the numbers churned out by the machine? In other words, is there convergence? If so, how to accelerate it at minimal increase in the cost? Can one estimate the error? These are some of the issues which we take up now.

The answer to the question of convergence is provided by the central theorem of numerical analysis which states that a consistent and stable scheme is convergent. Consistency means that if the approximating sequence is convergent, then the limit is a solution. The order of consistency will usually be found from the local exactness of the scheme on polynomials. It is not entirely obvious to show that Glimm scheme is consistent. Stability means that the approximate solutions are 'bounded'. In classical elliptic problems, this is a consequence of ellipticity. In general, stability will depend on the kind of apriori estimates one can deduce on the exact solution. These are not easy to prove either. These estimates usually imply some weak convergence for a subsequence which is enough to pass to the limit in linear problems. The convergence of the entire sequence depends on the uniqueness of the solution. In the case of nonlinear problems, the above weak convergence may not suffice. This is simply a manifestation of instabilities created by nonlinearities. We require some compactness criterion of Rellich type. This does not always hold and even where it does, it is hard to prove. The compensated compactness result of Murat-Tarter is a far-reaching powerful generalization of Rellich's Theorem. Using this, DiPerna has succeeded in proving the convergence of schemes for systems of conservation laws with two equations. This is considered to be one of the landmarks in this area.

Apart from the central theorem cited above, experience also shows that stability and consistency can serve as valuable design principles for discretizing problems even if they are not well-posed in a strict mathematical sense.

Admitting convergence, the next step is to obtain error estimates. It is well known that the error depends not only on the order of the consistency but also on the regularity of the exact
solution. This explains the success of FEM in the case of linear elliptic problems\textsuperscript{17} and the difficulties in the case of turbulent fluid flows where the velocity field is notoriously irregular.

So far we have been discussing approximation of well-posed problem. Let us now focus our attention on some singular problems which all arise, as a rule, in practice. One of the grand open problems in the field is to suggest schemes which can efficiently compute irregular solutions. One idea to overcome this difficulty is to know the location and the nature of singularities of the solution.\textsuperscript{3}\textsuperscript{4} (Mandelbrot's seminal observation about the fractal character of the singularities of the velocity field in fluid flows cries out for an explanation.) Incorporating these singularities into our approximation scheme, we will then be able to compute 'rough' solutions. In order to localize the singularities, point-wise estimates will be of immense help; or one can try to check the decay of wavelet coefficients. These are the programmes for the future. Thus, we see how qualitative properties of the solution are intimately connected with its approximation.

Another idea is to study the asymptotic behaviour of the solution for large times. This means, in the modern language, to find out the attractor and the inertial manifold.\textsuperscript{3}\textsuperscript{5} Based on this, one can suggest the so-called nonlinear Galerkin approximation\textsuperscript{3}\textsuperscript{6} in which only significant Fourier modes representing the solution are taken into account. The trouble is that there are still too many of them. Here is where wavelet basis may be of immense help. The idea is to group Fourier modes and seek a new representation of the solution in terms of wavelet basis. The research in this direction seems to be full of promise.

Yet another way to treat higher Fourier modes is to set up a turbulence model whose solution is reasonably smooth and lies near the original one. This process is nowadays also called homogenization. In spite of having several tools like H-, T-, and G-convergences, H-measures, semi-classical measures and Wigner measures which had enormous success in homogenizing oscillating coefficients, domains and boundaries and in the analysis of composite materials, justification of a turbulence model still remains a dream.\textsuperscript{3}\textsuperscript{7}\textsuperscript{-}\textsuperscript{4}\textsuperscript{0}

Other unstable phenomena where an increasing number of calculations are done are interface instabilities of Helmholtz and Rayleigh–Taylor, boundary-layer instabilities, turbulent multiphase flows, turbulent combustion, bifurcation problems, etc. Let us mention also that no general approximate scheme exists as of today to the models which incorporate the basic physical laws of conservation of mass, momentum and energy. Construction of finite-element schemes even for scalar conservation laws in multi-dimensions is a hot topic of research.

5. Other issues

So far, we have considered direct approximation of a given model. In some cases, it is possible to 'simplify' it before proceeding to make computations. Presently, we will look at some examples in which once again qualitative properties of the solution play crucial role. We have already mentioned the equations which have rapidly oscillating coefficients or posed on domains with a lot of perforations or oscillating boundaries. Such problems arise as models in the study of composite materials, tall buildings and towers or as an idealization of certain fluid motions. Perforations are made in the structures in order to make them light and they are com-
monly used in space industry. In such cases, it is not wise to proceed with direct computations because the task of triangulation of the domain which can be as such a hard geometrical problem is further compounded by the nature of the domain. Homogenization procedures produce simplified models, in which coefficients, domain or boundaries oscillate no more. Needless to say that it is easy to deal with the homogenized models.

Similarly, if one has three-dimensional bodies which are thin in one direction or multistructures consisting of bodies of different dimensions, it is better to make an asymptotic study which yields (coupled) models of lower dimension that can be handled more easily. In the same spirit, let us mention that long-time integration of equations generally needs the knowledge of the solution and calls for a study of asymptotic behaviour of the solution for large times. Linear hyperbolic equations on unbounded domains arise as models in scattering problems. How are we going to discretize such domains? One idea to overcome this difficulty is to look for a suitable boundary formulation of the problem. If scattering frequencies are to be calculated then we have to worry about diffraction, grazing, etc. Direct computations are quite difficult. One can think of using the ansatz of linear geometrical optics which reduces the problem to a set of ODEs and a simple transport PDE. In this context, let us also mention the problem of localization of a shock. This is a quite difficult problem because the shock is driven by the flow behind it which is enormously complicated. Shock dynamics tries to isolate those mechanisms which are mainly responsible for the shock movements and cooks up a simplified model. Once again geometrical optics techniques are in forefront and are found very useful. We end our remarks on this aspect by mentioning that a wide variety of singular perturbation techniques are also available to simplify a given model before starting computations on it.

Another aspect which we have not yet touched is the following: so far, we were concerned with the computation of the solutions from the data. This is known as direct problem. Inverse problems are the ones where we need to extract information on data (or even determine data) so that the solution has a desired behaviour. Concepts like stabilizability and exact controllability have been introduced in this context. These have to be developed further especially in the case of infinite dimensional nonlinear systems. Computational issues involved in such control approaches are highlighted by Glowinski and Lions.

6. Conclusions

In this brief sketch, we have merely touched upon several aspects of computations which are more or less specific in nature and discussed their connections with various qualitative properties of the solution. Thus, merely proving its existence is not sufficient; indeed, it is only a first step which provides a framework through which further analysis should be pursued. Needless to mention that many more points are left out in our discussion.

We would like to conclude with some observations which are of general nature. Scientific computing with the corresponding supporting mathematical analysis is an independent discipline which is gathering momentum. Having computers is like having telescopes in astronomy and microscopes in biology. The calculations not only aid engineers in their design but also give theoreticians clues about possible structures involved and jog their imagination. For this, it
is essential that the numbers from the computer should have some significance and not be a garbage. How to check this? Rigourous proofs may take a long time to come, if at all. Testing the programme with special and explicit solutions, checking with asymptotic description, verifying the results in simplified form of equations, matching with experiments in laboratories and repeating the computations with several values of parameters involved and checking the convergence are some of the numerical pragmatics that are usually followed. But obviously we need more.

Another direction in which progress is made is the acceleration of various algorithms. Parallel computations serve this purpose apart from specific algorithms like fast Fourier transform (FFT) and fast wavelet transform (FWT).

Our discussion of convergence and error analysis is focused at bringing out several features of solution into the picture. A good error estimate is the one which is both computable and realistic. A bound on the error which overestimates it by a factor of 100 is surely not realistic. Most of the error analysis leads to estimates which are often not realistic. A potential exception is the method of aposteriori error analysis but even this has not been worked out to any significant extent.

There is a large class of calculations in the field of dynamical systems to show that they exhibit chaos. By the very definition of chaos, there is extreme sensitivity to initial conditions and other data in the system. Because of the unavoidable round-off errors in the computer, the question arises as to how one is sure that the calculations really represent the exact situation. This is a very puzzling situation indeed! Of course, there is the so-called shadowing lemma. Is this sufficient in all situations? From the myriad of computational results, can one extract some information on the average behaviour of dynamical systems?

We would like to conclude by giving a rather stunning application of computations. It is not a surprise to use machine calculations to prove results in combinatorics (four-colour theorem). But Fefferman and Lanford have taken the road of using them as an ingredient of a rigourous proof of theorems in analysis. Unbelievable, isnt it? The reader is most welcome to try his own skills.

Acknowledgements

This paper was written by the author as a member of the Programme Advisory Committee of Mathematical Sciences (PAC-MS) of the Department of Science and Technology (DST), Government of India, New Delhi. A condensed version of it is appearing among the 'Vision papers in mathematical sciences' published in a three-part serial in Mathematics Newsletter (MN) brought out by the Ramanujan Mathematical Society. Two parts have already appeared: Part 1: MN, Vol. 6(4), 64-77 (1997) and Part 2: MN, Vol. 7(1), 8-21 (1997). The author gratefully acknowledges the permission granted by DST to publish the article.

References

3. Feigenbaum, M. J.  

4. Lorenz, E. N.  

5. Henon, M.  

6. Iserles, A.  

7. Sod, G. A.  

8. Strikwerda, J. C.  

9. Yar'enko, N. N.  

10. Garabedian, P.  

11. Richtmyer, R. D. and Morton, K. W.  

12. Smoller, J.  

    Finite difference techniques for vectorized fluid dynamics calculations, Springer-Verlag, 1981.

14. Fletcher, C. A. J.  

15. Glimm, J. and Majda, A. J. (Eds)  

16. Godlewski, E. and Raviart, P. A.  

17. Charnet, P. G.  

18. Quarteroni, A. and Valli, A.  

19. Gottlieb, D. and Orszag, S.  

20. Dautray, R. and Lions, J. L.  

21. Cohen, A.  

22. Daubechies, I.  

23. Meyer, Y. (Ed.)  


45. Phoolan Prasad


46. Lions, J. L.


47. Komornik, V.


48. Lions, J. L.


49. Glowinski, R. and Lions, J. L.


50. Isaacson, E. and Keller, H. B.


51. Lanford, O. E. III


53. Lanford, O. E. III