On the Love Affair between Computing and Maths

The 2017 IMA Lighthill Lecture, was given by Professor Beth Wingate at the British Applied Mathematics Colloquium, University of Surrey. The lecture is given in memory of Sir James Lighthill, founder President of the IMA. This article is based on Beth's Lighthill Lecture.

ENIAC and numerical weather prediction

The first grand challenge and one of the most famous breakthroughs in scientific computing occurred just after World War II when a group of mathematicians and scientists came together to create the world's first numerical weather prediction on the Electronic Numerical Integrator and Computer (ENIAC), an early computer. One of the most important mathematical lessons learned from this endeavour was that there is an intimate relationship between the underlying mathematical structure of the governing equations and their numerical approximation.

> … Peter Lynch recreated the first numerical weather simulations on his mobile phone, which he called the PHONIAC …

The problem that had to be addressed before numerical simulations of the governing equations were made possible was the oscillatory stiffness inherent in numerical weather prediction (see [1–3]).

By making key mathematical approximations based on observed velocity and time scales available through weather maps, Charney derived reduced equations, called the Quasi-Geostrophic (QG) equations, that represented the large scales of interest to planetary scale dynamics and filtered out the fast waves that, at the time, could not be resolved numerically. The first simulations were published in the widely cited paper [4].

An interesting account of that time period, including descriptions of the grand challenge project to use the ENIAC for numerical weather prediction, and a re-creation of the numerical results was published in *The ENIAC Forecasts* [5] in 2008 by Peter Lynch from University College Dublin. The original weather prediction researchers were contemporaries of Sir James Lighthill, whose many contributions to wave theory persist today, and whose book, *Waves in Fluids* [6] is still an important reference text for practitioners.

Mathematically, PDEs of the type faced by Charney have the following form:

$$\frac{\partial \mathbf{u}}{\partial t} + \frac{1}{\epsilon}\mathcal{L}\mathbf{u} + \mathcal{N}(\mathbf{u},\mathbf{u}) = \mathcal{D}\mathbf{u}, \;\; \mathbf{u}(0) = \mathbf{u}_0, \qquad (1)$$

where the linear operator $\mathcal{L}$ has pure imaginary eigenvalues, the non-linear term is of quadratic type, the operator $\mathcal{D}$ is some form of dissipation with real eigenvalues, and ϵ is a small non-dimensional parameter. In this equation we also define $\mathbf{u}(t)$ to be the spatial (vector-valued) function $\mathbf{u}(t,\cdot) = (u_1(t,\cdot),\ u_2(t,\cdot),\ldots)$.

The operator $\epsilon^{-1}\mathcal{L}$ results in oscillations on an order $\mathcal{O}(\epsilon)$ time scale, and requires small timesteps if standard explicit numerical integrators are used. Even implicit integrators need to use small timesteps if accuracy is required. One of the interesting effects of the non-linearity is that it acts like a phase scrambler and creates dynamics that is slow relative to the fast $\mathcal{O}(\epsilon)$ oscillations. Each application, such weather prediction or magnetic field dynamos, will have a particular *native* timestep that has to do with the types of waves that exist in the system. The notion that both slow and fast time scales exist simultaneously in the solution is called a multiscale system and is associated with the phenomenon of time-scale separation. If there is more than one type of native frequency (e.g. buoyancy, rotation) in a system, this gives rise to even more complex mathematical structure.

Even though there are slow dynamics available in the solutions, the fact that the fastest time scales are highly oscillatory means that as the spatial resolution increases, the timestep must decrease. For example, contemporary atmosphere models with a resolution of 300 km will use a timestep of 20 mins. If the resolution is refined to 1 km, this would require a timestep of 4 s, making the cost of doing century scale runs expensive.

This means there are trade-offs between horizontal resolution, optimal distribution on parallel processors, and the timestep, and this is affected by the oscillations of the system.

The numerical issues first identified by these earlier mathematicians, in particular the timestep limitation due to dispersive waves, has been an active topic of research ever since those early years and an issue that has to be addressed in every field that uses computations of partial differential equations.

Despite the timestep limitations, the years since the ENIAC simulations have seen computational science become the third pillar of scientific discovery alongside theory and experiment. During this era computational science has provided major gains in our understanding of the physical and biological world, including, for example, anthropogenic climate change, tsunami prediction, and the simulation of supernovae.

Beyond the silicon limitation

One of the most important reasons why computational science has flourished in our lifetime is that Moore's Law, an observation made in 1960 by Gordon Moore, has held true these last 40 years. To get an idea of how far computers have advanced, Peter Lynch recreated the first numerical weather simulations on his mobile phone, which he called the PHONIAC (Portable Hand-Operated Numerical Integrator And Computer) – https://maths.ucd.ie/plynch/eniac/phoniac.html. The exponential processor speedups, which led to the most advanced computer of its age becoming no more powerful than a mobile phone today, are at an end. This is due to physical limitations in the manufacture of transistors and their subsequent power consumption. For practical purposes Moore's Law ended in 2005 and unless some other way is found to advance processor speeds, such as quantum computing, this is the end of unending computer processor speeds.

This has not dampened the creativity of computer architects who have been working steadily on creating new types of computers that are different from those of the past and that could be considered the grand challenge computers of our time. In particular, they have been developing new types of silicon architectures that replace processor speed with the possibility of doing hundreds of millions more calculations at the same time, called *concurrency*. This means that rather than relying on faster computers to help us answer questions about how to make wind power more efficient, or to help us decide where continental boundaries might change due to sea level rise, we are asked to find many more things that can be computed *at the same time*. It is expected that, in the next five years, machines will be delivered capable of 100-200-million-way parallelism. Just this year, to prepare UK science for the jump to concurrency, the Engineering Physical Sciences Research Council (EPSRC) funded five new Tier 2 Computing Centres, three of which have new architectures of the kind we can expect in the future.

The urgency to prepare scientific simulation methods for the post-Moore shift to exascale computer architectures has already initiated major projects in every nation that relies on high performance computing (HPC). For numerical weather prediction alone examples include the Accelerated Climate Modeling for Energy (ACME) in the USA, the Met Office/NERC/STFC GungHo (next-generation numerics) in the UK, the Centre of Excellence in Simulation of Weather and Climate in Europe (ESiWACE) and the Energy-efficient Scalable Algorithms for Weather Prediction at Exascale (ESCAPE) project in Europe. These efforts are required to address the immediate difficulties with today's sophisticated computer models that will have to be run on early versions of these post-Moore architectures expected in the next 3–5 years.

The return of the native timestep limitation

In this new era of post-Moore computing, the question is, can we continue to advance science on new computer architectures where we do not have faster processors, but we do have the potential to do hundreds of millions more computations at the same time? These new computers are the ENIACs of our time and making best use of them is one of the important mathematical challenges of our time. What is standing in the way of computational science continuing in the same way it has the last 40 years? It is our old friend the timestep limitation that was confronted in the early days of numerical weather prediction.

This challenge invites us to re-examine and rethink the mathematical structure of the equations. In particular, examining the structure of the time domain has been proposed by the parallel-in-time numerical analysis community (www.parallelintime.org). There are different flavours of time-parallelism, but for an introduction see the review article by Martin Gander [7]. One idea of interest to me is one that illuminates the mathematical structure of the low frequencies available in the PDE. This method, the topic of the Lighthill Lecture, relies on trying to make gains in parallelism by taking advantage of the frequency domain.

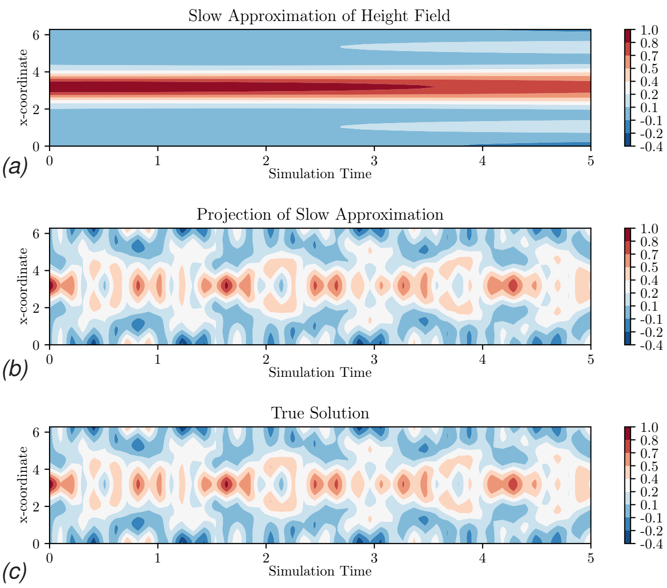

Figure 1: The time evolution of a 1D Gaussian dam-break problem using the psuedo-1D shallow water equations used in [5]. The figure is for $\epsilon = 0.005$; (a) shows the evolution of the frequency-filtered twist equation (3); (b) is the result of rotating (a) back using the matrix exponential mapping (2); and (c) the time-evolution of the dam-break problem computed with a 4th order runge-kutta timestepping method. Courtesy Adam Peddle.

Our approach uses a mapping, the matrix exponential, to transform the unknowns into the space of the oscillations, thereby exposing the frequency content of the non-linearity to mathematical analysis and modelling. The mapping looks like,

$$\mathbf{u}(t) = e^{-t/\epsilon\mathcal{L}}\,\mathbf{v}(t), \qquad (2)$$

$$\frac{\partial \mathbf{v}}{\partial t} + e^{t/\epsilon\mathcal{L}} N(e^{-t/\epsilon\mathcal{L}}\mathbf{v}(t), e^{-t/\epsilon\mathcal{L}}\mathbf{v}(t)) = 0, \quad \mathbf{v}(0) = e^{t/\epsilon\mathcal{L}}\,\mathbf{u}_0 \qquad (3)$$

where we have dropped the dissipation, $\mathcal{D}$, to focus on the oscillations in non-linearity. Here the operator $\mathrm{e}^{-t/\epsilon\mathcal{L}}$ is the semi-group operator associated with $\mathcal{L}$, also called the matrix exponential. The mapping, Equation (2), continuously twists the unknowns in what I like to think of as a generalised helix in time, allowing the non-linear solution to unfold in Equation (3).

Inspired by the use of this mapping in multi-scale theorems of the 1990s, Terry Haut (Lawrence Livermore National Laboratory) and I [8] showed that finite frequency averaging combined with the formulation of Equations (2, 3) could lead to significant parallel speedups for a range of time-scale separations. While the parallel speedups we obtained for the simple problem studied were interesting, we were only able to prove super-linear convergence in the case where epsilon goes to infinity.

There has been plenty of evidence (see examples from [9–11] and those who cite them) before our work that the case for finite epsilon is at least as interesting as the case when epsilon goes to infinity, an assumption commonly used in mathematics. For example, though the Earth rotates once per day, mathematics often assumes that the Earth rotates infinitely fast to make progress with proofs. This assumption omits the bandwidths of frequencies that contribute to the low frequency dynamics shown in Figure 1 through phase scrambling. Therefore, the immediate need is for methods and theorems for the more realistic case of finite epsilon, and therefore finite time-scale separation. This means there is a great deal more to be done and understood before methods that rely on time-scale separation can be used for realistic applications like weather prediction.

Grand challenge problems

In this section I propose some conceptual grand challenges for mathematics and computing. These suggestions are biased toward my own area of expertise and the list will be different for applications beyond weather and climate, such as plasma physics, tsunamis, magnetohydrodynamics, wave propagation, and other application areas whose governing equations are PDEs in the form of Equation (1).

Finite time-scale separation:

As mentioned above, there has been persistent evidence that the finite time-scale separation is just as important as the infinite case [9-11]

(a) PDEs analysis: Proofs of regularity, existence and uniqueness for PDEs like Equation (1) with finite time-scale separation could help guide numerical methods designers. Even knowing the assumptions under which theorems can be proved could be useful.

(b) Numerical methods design and analysis: How far can we take time-parallelism? Can we do novel integrations of the evolution equations in the frequency domain?

(c) **Mathematics of fluid dynamics and physics:** If the low frequencies matter, there should be structural treasure to find. Are there new theories and scaling laws for finite time-scale separation?

Solution space of PDEs as images

Consider the solution space of Equation (1) over some long, finite time T to be an image. Are there mathematical notions that allow the reconstruction of the solution space from an initial guess? Will it be faster on new machines to drive snapshots of data from one parameter regime to another rather than beginning every solution at time zero? As an example, in Ingrid Daubechies Lecture at the 2017 BAMC, she spoke about the application of wavelets to art reconstruction. To fill in damaged spaces of a painting the applied mathematician considers 'where the paint stroke came from'. Can we use ideas from the structure of the finite-frequency PDEs, or image compression and reconstruction, to reconstruct solution spaces of PDEs on finite time intervals? This could allow different parameter regimes to be explored with much more concurrency.

Performance models as mathematical maps

One commonly used strategy for advancing the computational performance of complex applications is to construct a *performance model* that is then used to understand and transform how the more complex simulation will perform on a new architecture. Can we develop this further to construct mathematical maps between the entire solution space of Equation (1) over some long-time T, and different configurations (memory, bandwidth, concurrency) of new computer architectures? If we can view these types of performance models as mappings, analysis may be able to tell us how far we can take the solution into concurrency. If successful this could help modify existing algorithms and architectures in the direction of best performance.

Ensembles Averages and Machine learning

While increasing model resolution may be challenging on new architectures, running many variations of our current model resolutions, which would use significant degrees of concurrency, may bring on an era of ensemble-driven science. Though it is unclear whether models of this type can answer some of the pressing questions about climate change, it seems possible that having richer data sets and ensembles could be useful. What are the new science questions we can answer with richer data sets of ensemble averages? Where can the growing fields of machine learning and data analytics take us in answering questions of interest to life on earth?

Conclusion

In David Keyes' opening address at the 2017 SIAM CS&E meeting in Atlanta he reminded the attendees of an important moment after World War I. At the International Mathematical Congress in Bologna, 1928, after the 1920 and 1924 Congresses had excluded mathematicians representing the countries defeated in World War I, Keyes said: 'Mathematics knows no races or geographic boundaries; for mathematics, the whole cultural world is one country' [12]. It seems to me that other challenges of our time could be met by working together to see how far we can go with the ENIACS of our time.

Beth A. Wingate
University of Exeter

References

1 Charney, J.G. (1947) On the dynamics of long waves in a baroclinic westerly current, *J. Meteorology*, vol. 4, pp. 135–162.

2 Charney, J.G. (1948) On the scale of atmospheric motions, *Geofysiske Publikasjoner*, vol. 17, pp. 3–17.

3 Charney, J.G. (1949) On a physical basis for numerical prediction of large-scale motions in the atmosphere, *Journal of Meterology*, vol. 6, pp. 371–385.

4 Charney, J.G., Fjörtoft, R. and von Neumann, J. (1950) Numerical Integration of the Barotropic Vorticity Equation, *Tellus*, vol. 2, pp. 237–254

5 Lynch, P. (2008) The ENIAC Forecasts: A Re-creation, *Bull. Amer. Met. Soc.*, vol. 89, pp. 45–55.

6 Lighthill, M.J. (1978) *Waves in Fluids,* Cambridge University Press.

7 Gander, M.J. (2015) 50 Years of Time Parallel Time Integration in Carraro, T. et al. (Eds.) *Multiple Shooting and Time Domain Decomposition*, Springer-Verlag.

8 Haut, T. and Wingate, B. (2014) An Asymptotic Parallel-in-Time Method for Highly Oscillatory PDEs, *SIAM J. Sci. Comput.*, vol. 36, pp. A693–A713.

9 Newell, A.C. (1969) Rossby wave packet interactions, *J. Fluid Mechanics*, vol. 35, pp. 255–271.

10 Smith, L.M. and Lee, Y. (2005) On near resonances and symmetry breaking in forced rotating flows at moderate Rossby number, *J. Fluid Mechanics*, vol. 535, pp. 111–142.

11 Babin, A., Mahalov A. and Nicolaenko, B. (1999) Global regularity of 3D rotating Navier-Stokes equations for resonant domains," *Indiana University Mathematics Journal* , vol. 48 pp. 1133–1176.

12 Eves, H.W. (1971) *Mathematical circles revisited: a second collection of mathematical stories and anecdotes*, Prindle, Weber & Schmidt.