# More on the nature of prediction: geometric integration

Jason Frank

9 November 2016

*Mijnheer de Rector van de Universiteit Utrecht,*
*Waarde Collega's van het Mathematisch Instituut, de Bètafaculteit Utrecht,*
*en de Nederlandse universiteiten,*
*studenten, vrienden en familie,*

## 1 Prediction

Since antiquity people have attempted to predict the future. Soothsayers, oracles, prophets, bookies . . . and meteorologists. Countless adages pertain to the weather, indicating that people have always tried to find some order in why it changes as it does, to anticipate and respond to it.

The current political climate in the Netherlands stresses the need for science to demonstrate its relevance to society, industry and government. Predicting complex systems like the weather is one way that science provides such a service to society. Society also spends great effort trying to predict things like stock prices, the outcomes of sporting events, and even . . . political elections.

In most scientific disciplines prediction plays a subtle role, as a means of validating a new theory. One reads, "our theory predicts that ...", and then experiments are constructed, first to verify and later to falsify—that is—break the theory. If we are unable to break it, then that provides at least empirical evidence of its truth. Nevertheless we see that in science, prediction provides a *means* for checking if our *understanding* is correct, rather than being a goal itself.

Falsification has no role in formal mathematics. A proper mathematical theorem is true because it has been proven so, and no experiments are needed to test its validity or lack thereof.

I suspect that most scientists are less motivated by the idealistic goal of furthering human knowledge, than by the somewhat more selfish desire to *personally understand* the object of their study. There is a exhileration that comes from finally putting the pieces together and reaching a new understanding of something complex. If indeed, scientists are motivated by personal understanding, then relevance to society is a by-product of this motivation, and it may also play a role in deciding which topics are worth investing time in.

Of course, some mathematicians are motivated otherwise: Ph.D. student John Urschel of MIT—who divides his time between research at the boundary of numerical linear algebra and graph theory and his “day job” as an offensive lineman for a professional football team—states his motivation as, “I love hitting people. There’s a rush you get when you go out on the field, lay everything on the line and physically dominate the player across from you.” (Well, actually that’s his motivation for football not mathematics, or else I don’t want to meet him at a conference).

Many mathematicians find it challenging to express the connection between their research and its application to problems from society, industry or other fields. Mathematics deals with abstract concepts and the relationships between these. Its power lies precisely in that abstraction and its ability to uncover deep truths that pertain to a wide range of problems. One colleague has lamented, “I wish it were sufficient to argue: this research is important because it is important for mathematics”.

My own research subject, numerical analysis, can rather easily be connected to applications, because we develop methods for solving concrete problems computationally. My predecessor in the chair of Numerical Analysis was Prof. Henk van der Vorst, one of the world’s foremost experts in developing efficient methods for solving large systems of algebraic equations. One of Henk’s articles, that introduced a new numerical method called Bi-CGStab, was the single most cited mathematics article in the whole decade of the 1990s. The methods developed by Henk are applicable to a wide class of problems encountered in different fields, although the papers themselves treat the problems in abstract form, with little reference to specific applications.

On April 21, 2011, I gave my inaugural address at the University of Amsterdam, entitled “The Nature of Prediction (and the Prediction of Nature)”. At that time I explained the computational mechanisms of prediction. Today I want to re-examine some aspects of prediction and talk about some more nuanced details and the relation to *geometry*. For me, geometry stands for the underlying spatial structure of the problems I study, as related to constraints imposed by conservation laws and symmetries. At the same time, I associate geometry with the strong, fundamental work of a number of my colleagues.

Within the Mathematical Institute in Utrecht we have a concentration of mathematicians working in Geometry. This is the branch of mathematics that deals with the properties of space. For example, topology tells us which of these knots can be deformed into one of the others without breaking the loop.

Much of my research in numerical analysis has addressed methods for preserving geometric structure while simulating dynamical systems. This topic is called “geometric integration”.

In this lecture I want to emphasize:
- how geometry is applied in numerical analysis, and
- how numerical analysis is applied to prediction of complex systems.

## 2 Complex Systems Science

A research area that holds promise for societal impact is *Complex Systems Science* or simply *Complexity.* The science agendas of the Royal Dutch Academy of Sciences and NWO, as well as the National Science Agenda have all recently included Complex Systems Science as strategic research theme. Utrecht University has identified Fundamentals of Complex Systems as one its Focus Areas, intended to promote cross-disciplinary research within the organization.

This relatively recent field addresses the collective behavior of large numbers of interacting "agents", where the latter may be:
- cells or other organisms in a biological or ecological system,
- traders in an economic market,
- packets or travelers in a transport system,
- opinions expressed through social media,
- molecules in a gas or fluid, etc.
A complex system is one in which the interactions of a great number of individuals at one level gives rise to some coordinated aggregate behavior at another level. Simply put, "the whole is greater than the sum of its parts".

Complexity distinguishes between dynamics at a *microcopic* level which addresses the interactions of individual agents, and those at the *macroscopic* level associated to collective behavior. The induced behavior at the macroscopic level is referred to as emergent behavior. For example, large numbers of atoms interact to form complex proteins, large numbers of proteins interact to form cells, large numbers of brain cells interact to form neural networks, whose signals in turn interact in a complex manner to give rise to a conscious being such as a theoretical physicist, and a large number of physicists interact to think up string theory to explain atoms. At each level there emerge new dynamics that are hard to surmise from the underlying levels. From the mathematical perspective, one tries to understand the macroscopic behavior of the system by analogy with the known dynamics of relatively simple systems.

Complex systems science relies on computational models to simulate the behavior, decisions and strategies of individual agents. The science of complex systems adapts analysis techniques from mathematics and physics to try to understand such systems, with the promise of giving us a handle on problems that have in the past been hard to penetrate with the formalisms of exact sciences.

The climate is a complex system comprised of a very large number of interacting subsystems whose feedbacks and feed-forwards are hard to quantify in isolation. The cloud system, just to name one, blocks radiation both from the sun to the earth and from the earth to space, absorbs and releases moisture and enhances vertical transfer of heat, etc.

Other examples of complex systems are the investment mechanisms of financial and economic systems, in which the investment actions of a large number of agents give rise to observable macroeconomic variables like stock market prices and averages;

Or social media networks, where a post to Facebook or Twitter may "go viral" or be forgotten, and where the spread of information may influence what people buy or *how they vote*.

Here again, our primary motivation has been to *understand* complexity.

But society has only limited use for "pure understanding": knowledge becomes relevant when it leads to action. Generally this means controlling or at least predicting complex systems to some degree.

Pure knowledge and understanding are fascinating. That's why many newspapers include a science section. Yet understanding alone goes little further than that. People really only derive benefit from scientific knowledge when it can be used to take action concerning the objects of study, either to influence them or at least to predict them and respond.

Understanding financial markets is satisfying, but when that understanding helps us regulate markets to prevent crashes, the knowledge gains societal relevance.
Understanding climate is intriguing, but when we can predict climate, we know how high to build the dikes.

## 3 Chaos

How predictable are complex systems? Some are clearly less predictable than others. Over the last months we have witnessed the difficulty of predicting how a nation will decide a referendum or an election, for instance.

Yet some complex systems exhibit a certain degree of predictability. Most of us trust the weather predictions enough to adapt our plans to them several days in advance. Nevertheless complex systems generally allow only a limited predictability horizon.

To provide a concrete system to illustrate this lecture, *as well as for your amusement*, this summer I prepared, with the aid of my father, a demonstration. I would like to ask my daughters Helen and Rosa to assist me.

You see here two double pendula. Each consists of a pendulum with, attached to its tip, a second pendulum. It is easy to derive a mathematical model of a double pendulum. It satisfies Newton's laws for mechanical systems, which are well understood. From

Newton's laws the state of the pendulum is perfectly constrained if we know the position and velocity of each arm. For instance, if we release the pendulum from rest, then its initial velocity is zero, and we only have to state its position. A different initial condition leads to a different motion.

I want you to imagine that the pendulum on your left, held by Rosa, is the true physical pendulum we wish to predict, while the pendulum on your right is an approximate model, complying with our best scientific understanding of the pendulum on the left. For this system, the uncertainty is due to:
- the imprecision of the machining done by my father and myself,
- the friction of the bearings,
- and the uncertainty in the initial condition.

I will ask Helen to prepare her pendulum as accurately as she can to reflect the initial state of the pendulum held by Rosa. When they release the pendula we will see how predictable Rosa's pendulum is using Helen's model. The longer the motion of Helen's pendulum matches that of Rosa's, the more predictable is the pendulum, since the two are then insensitive to the errors we make in manufacturing and initializing the system. Ready, 3 - 2 - 1, release!

Unfortunately, we see there is really very little predictability with this system. (Thank you, girls.)

To illustrate a bit better we move to the computer, where we can study the pendulum in a clean environment, bereft of annoyances like friction, wobbly tables, and machining errors. This, coupled with the relative speed and low cost with which one can build a model on a computer compared to manufacturing one, is the reason that computers are so prominent in experimental science. On the computer we can make the pendula perfectly identical, up to a small difference in initial state. Namely, you will see here 10 identical pendula, whose only difference is that the angles of the second arms are initially perturbed by less than a 100 millionth of a degree. Initially the motion appears identical but suddenly the differences seem to blow up. Mathematically we say their motions diverge at an exponential rate. This phenomenon is referred to as "sensitive dependence on initial conditions" and it is one of the hallmarks of a *chaotic system*.

## 4 Geometric integration

This is an appropriate moment to say something about one role of geometry here. The motion of the double pendulum, though hardly meeting the criteria for a complex system, is still too complex to be described exactly with a function (if it could be so described, it would also be predictable). Instead I have solved the equations approximately using a computer. By definition, an approximation is inexact, and consequently subject to error.

One of the first steps in model construction is the identification of an appropriate set of variables. The power of geometry is to recognize that the underlying spatial structure of a given model is independent of this choice of variables. Consider our double pendulum. As already said, the laws of Newton dictate that the system is completely determined if we know the position and velocity of each point at some initial time. If we define a coordinate system with the pivot A at the its origin, then we can specify the system configuration in terms of the $(x, y)$ coordinates of the points B and C, and the velocity vectors of those points. There are two coordinates for each point and two coordinates for each velocity vector, so in total 8 numbers to specify the system.

We can think of those 8 values as the coordinates of a point in another, 8 dimensional space. As the pendulum swings, all 8 values change, and the point traces out a curve in the 8 dimensional space. Don't panic if you cannot visualize this, no one else can either.

Geometry allows us to simplify this picture significantly. For instance, the point B is always the same distance from A. Its motion is on a circle. Instead of the two coordinates $(x_B, y_B)$ we can simply express the state of the first arm by its angle $\theta_1$ with respect to the vertical. Similarly for the second arm $\theta_2$. The velocities can also be expressed as angular rates $\alpha_1$ and $\alpha_2$ in degrees per second. Now we have 4 dimensions instead of 8. We can do one better. Without friction, the double pendulum satisfies the law of conservation of energy. That means there is some 3-dimensional subset of our now 4-dimensional space that contains only those states of the system consistent with a given value of energy, and the motion is constrained to this subspace. Using geometry we have reduced the motion from 8 dimensions to 3.

On the other hand, describing motion in terms of 3 variables makes the equations extremely complicated. The original description in 8 dimensions was easy by comparison. From the point of view of computation, it is easier to work with 8 dimensions, yet it is imperative that our numerical approximation respect the fact that the actual motion is in 3 dimensions. Otherwise we get silly results where, for instance, the pendulum arms grow, or energy gets artificially pumped into the system.

*Geometric Integration methods* are tailored to a problem to guarantee that its underlying geometry is preserved by the approximation. However, this comes at a price: namely, we cannot easily design generic "black-box" numerical methods, since the method has to be told about the underlying geometry.

I would now like to discuss two more subtle aspects of prediction. Here I have to apologize to the master students present, because they have heard me tell this a number of times this semester. The first point is the following:

## 5 Shadowing

We saw with the double pendulum that some systems are highly sensitive to small perturbations. Consequently we know that our attempts to predict complex systems are doomed to failure because of exponential divergence of motions following from approximation errors. We may then ask, what is the reliability of numerically computed solutions over long times, such as those utilized in climate studies.

Chaotic systems are sensitive to perturbations. However, not all perturbations are created equal. We can be more precise by stating that most complex systems have directions in which they may be stably perturbed and directions in which perturbations are unstable. The space in which our system is described can in fact be split at each point into a component along which perturbations grow exponentially, a component that decays exponentially, and the rest. We refer to the first of these as the unstable space and the second as the stable space.

With a bit of work we can compute these directions and construct moving coordinate axes at each point along the solution. Doing so is itself an exercise in geometric integration: the coordinate axes are obtained by solving systems of differential equations on something called a Stiefel manifold. One of my scientific collaborators, Erik van Vleck of the University of Kansas, is an expert on computing such moving coordinate systems.

Small perturbations in the *unstable* directions will cause the motions to *diverge* from each other exponentially fast, whereas perturbations in the *stable* directions will de damped out. We will see that this picture is a very useful way to think about predictability.

The decomposition into stable and unstable spaces is important because perturbations that *grow* in forward time *decay* when we look back in time and vice-versa. This allows us to apply an elegant trick to say something about the validity of our simulations, for times long beyond the time that predictability is lost. Put simply, for some systems we can show that although the computed prediction diverges from the true motion of the system, there is some other true motion of the system that stays near to our computed prediction for a very long time. It is just not the solution that corresponds to our initial state.

Lecture Notes of Michael Cross of the California Institute of Technology contain a nice schematic illustration of this.

Suppose these points are a computed prediction of a chaotic system, and suppose that the error encountered in going from one step to the next is no bigger than $d$. Construct a box around the first point with sides of length $2d$, such that the sides of the box are aligned with stable and unstable coordinate axes. Draw a curve in the box that passes through the ends perpendicular to the unstable direction.

We think of each point in the box as a distinct initial state of our system. The exact motion of our system over a short time moves each of these points to a new location. But they still define a connected set of points which still looks rather like a box. However, the new set is stretched in the unstable direction and compressed in the stable direction. In particular, it contains the second approximation point and the now stretched image of the curve. Now we construct a new box of size $2d$ around that point and note that it contains a part of the image of the first box. The rest we ignore. We can repeat this construction along the whole numerically approximate motion.

In the last box we select a point on the curve and examine its inverse image all the way back to the initial time. The exact motion through this point , by construction, is a shadow solution that remains near our computed motion along the whole time interval.

As a consequence of shadowing, we can say that while we cannot precisely predict how the motion originating from a given initial state will evolve over long time, we can be confident that our computed motion is representative of *some* solution of the system. For complex systems it is often the case that statistics of different motions are identical within some class. As such, when this shadowing result holds, we can be relatively confident that the *statistics* gleaned from our simulation data are correct.

## 6 Predicting the now

Compared to the challenge presented by chaos, the actual specification of the current state of a system may seem a rather innocuous ingredient in a recipe for prediction. Of course, we have seen that uncertainty in the initial state will lead to divergence and loss of predictability, so this needs to be determined as accurately as possible.

But there is more to it: for most truly complex systems, we are not even able to completely *observe* the instantaneous state of the system. What is the current state of the climate, for instance? Instead we typically have a means of measuring some "observable" quantities. such as averages or local measurements.

To determine the current state of the atmosphere for weather prediction, measurements of the pressure, temperature, wind speed and direction, and more exotic quantities are recorded at a vast array of weather stations on land and to lesser extent at sea. Most of these observations are taken near the surface—at the *bottom* of the atmosphere. Satellites and radar provide remote observations. In any case, these observations are insufficient to precisely pin-down the state of the atmosphere to high resolution. Somehow we have to synthesize all the available data to produce a "state estimation" from which the initial condition for the weather prediction is distilled.

On the other hand, there is a lot more data available if we consider all the observations recorded in the past. This is, in fact, how state-of-the-art weather forecasts are prepared. Instead of trying to fit current observations, weather centers try to find a model solution that best describes the data observed over the recent past.

The data from the past can be considered to yield a stream of incomplete observations of the state of the system, which are also subject to measurement errors: so, a 'noisy signal'. The goal of data assimilation is to synchronize the computated prediction with this noisy signal.

*Synchronization* is a much studied topic in Complex Systems Science. The popular science book *Sync!* by mathematician Steven Strogatz describes many instances of synchronization in physical and biological systems:

Bird flocks, schools of fish, and swarms of insects synchronize their motion, both to save energy and to evade predators. There is safety in numbers because the herd can feed while a few keep watch, and because "you don't have to be faster than the bear, just faster than the other campers".

Other examples of synchronization are our metabolic and sleep cycles, which synchronize to the 24 hour rhythm of the earth's rotation. The re-synchronizing process after flying a great distance to the east or west is what we call Jet-lag, and working-hours that continually perturb the synchronized state can cause health problems.

On You-Tube you can find a number of videos illustrating the synchronization of metronomes placed on a platform that is free to move horizontally due to the momentum of the oscillators. (...) Christiaan Huygens first observed synchronization of oscillators while attempting to design highly accurate clocks to aid in navigation at sea. He noticed that his clock pendula synchronized when hung next to each other from a plank suspended between two chairs. This implied a touchiness that convinced Huygens his clocks would be useless aboard a ship at sea. The same effect led to the swaying of the Millennium Bridge in London when it opened in 2000, as the swaying caused the public to synchronize their walking motion, which in turn reinforced the swaying.

Data assimilation is akin to synchronization. We receive a partial observation of the signal, in the form of a stream of noisy measurements from nature. Then we try to synchronize an imperfect model to that signal. Frank Selten of KNMI has developed a number of clever applications of synchronization in weather and climate prediction.

Divakar Viswanath of the University of Michigan has studied prediction of chaotic systems via symbolic dynamics. To define a symbolic dynamics, one decomposes the motion of the system into a finite number of distinct states, assigning each of them a symbol, such as a letter of the alphabet. As the system evolves, moving from state to state, one records the sequence of letters, as a sort of shorthand description of the motion. For instance, one could label the event that Arm 2 of the pendulum overturns in a clockwise direction as A, the event that it overturns in a counterclockwise direction as B, and keep track of the sequence of As and Bs. Such a description is called *generating* if any two different motions of the system give rise to distinct sequences. Symbolic dynamics offers a means of prediction just by matching the currently evolving sequence against the longest identical sequence from the past. This approach is based purely on data and does not even require a model. A theorem of Mark Kac (kahtz) states that the amount of historical data needed to predict the motion over given period of time grows exponentially with the length of this time period.

An important question in data assimilation is what signal do we need to observe? In terms of weather prediction, where should we put the weather stations, and what data should we record there? It turns out that the state of a chaotic system can be better and better estimated if we can observe its motion in the *unstable space*. To illustrate this I show you two pendula. The blue one moves freely and independently according to its motion under gravity. The orange one is able to observe the unstable directions of the blue one, and it continually tries to correct itself along these directions. Gradually their motions synchronize. The convergence is exponential in time. The longer we record data on a system, the more accurately we can eventually estimate its state.

Ph.D. student Bart de Leeuw at CWI—working with Svetlana Dubinkina, Erik van Vleck and myself—is combining shadowing and synchronization into a practical data assimilation algorithm for estimating system states given past observations.

Here we see a nice example where modelling a problem from an application (how do coupled systems synchronize?) gives us ideas for mathematics (how to identify the state of a system given only partial observations). Such exchange makes multidisciplinary science productive.

## 7 Geometric integration, broadly construed.

Finally, I want to return briefly to the dialog on societal relevance of fundamental science.

I think it is only fair that science, including mathematics, must justify its merit to the society that foots the bill. It would be ridiculous to proudly proclaim that one's work has no relevance to society.

And knowledge for knowledge sake—extending the totality of human knowledge and understanding—while ultimately valuable, may not be the most efficient way to confront real pressing problems, possibly justifying a secondary prioritization.

On the other hand, mathematics truly is a very relevant discipline. Perhaps naively, I envision the subject as a large directed graph, where the nodes represent truths and the connections represent logical arguments. In the *core* of this graph are...some statements...that we have to accept on faith..., and at the *outskirts* are connections to other fields. Mathematics as a whole is highly embedded in society, with broad-reaching impact. *That impact* in turn is highly dependent on the links in the interior. There are many connections in the interior that are only conjectured or are yet to be discovered, and sometimes the addition of a well-placed link can have tremendous consequences for the whole structure. Fundamental work of high importance is considered so because it fills gaps and establishes links related to the internal consistency of the subject. And these links are critical. They allow us to be more certain of our results and make knowledge from one area of the graph applicable in another. Ultimately some fundamental results also find direct tangible application outside of mathematics, and others help strengthen the existing structure. Even the fundamental work is *applied*, but the field of application is again mathematics.

So it *is* acceptable to argue that a piece of fundamental work is important, *because it is important for mathematics.* But one has to argue at the same time that mathematics as a whole is important for society.

And that seems evident too. We surely are not training hundreds of students per year to continue a career developing mathematics at a fundamental level. Society simply could not support that. The fact that there is an increasing demand for mathematically trained graduates is evidence of the greater practical value of mathematics.

Mathematicians would be well served to build up collaborations that connect this graph in the radial direction, that is, spanning topics from the highly fundamental to the highly applied. This opens up doors to funding via large scale programs, since it involves a wide range of expertise, and large teams. Furthermore, just as the exchange of ideas in multidisciplinary work can be synergetic to both disciplines, such radial collaboration within a discipline can also lead to synergy.

For my own part, there is a wealth of knowledge in Utrecht on geometry that would be beneficial for developing geometric integrators to improve predictions. In turn, to learn and apply that knowledge, requires a different kind of "Geometric Integration". Namely, my personal *integration among the geometers.*

Just a few weeks ago a first step was taken: one our top geometers found his way downstairs to my floor of the building to get some help with applied knot theory: ...

*I helped him tie his necktie.*

**Acknowledgements.** In closing, I would like to take this opportunity to thank Erik van den Ban and Gunther Cornelissen on behalf of the Mathematical Insitute for the trust and counsel and cooperation, to Henk van der Vorst for encouraging me to give a second inaugural lecture, to Jim Frank for help in the fabrication of the double pendula, to Erik van Vleck, Frank Selten and Bart de Leeuw for providing indispensible references to scientific literature and inspiring discussions. Additionally, I thank Helen & Rosa and my wife Linda and her parents, for patiently enduring various annoyances during our recent vacation in the Austrian Alps, where I wrote the first draft of this lecture. Finally I thank you all for your attendance and attention.

*Ik heb gezegd.*