

NONLINEAR PREDICTION AND DYNAMICS

NORBERT WIENER

THE MASSACHUSETTS INSTITUTE OF TECHNOLOGY

A problem which is both mathematically fascinating and practically important arises in meteorology. We can conceive of the various constants which make up the weather on earth as part of a vast dynamical system which must be extended ultimately to include the rotation of the earth and the change of seasons, the variations in radiation from the sun, as well as those from the surface of the earth into empty space, and perhaps other driving forces. The ideally Newtonian way of building up a scientific meteorology would be to set up equations for the course of all the meteorological variables in time, and these equations would naturally assume the form either of partial differential or integral equations or combinations thereof. In order to make an effective use of such equations, we should have a very complete knowledge of the course of our variables in the past and at least of their values at some instant in the past. This completeness of knowledge, which would lead to the justification of a purely dynamical meteorological prediction, is absurdly far from what has actually been given us. In fact, we only have meteorological data from a few hundred or a few thousand stations all over the surface of the globe, and not all of these data have been collected continuously, but rather at certain stated intervals. Moreover, they are in no significant sense absolutely precise. If the thermometer in the weather bureau station in Boston reads 35° Fahrenheit, it is scarcely conceivable that this reading will characterize the effective temperature over the Boston area, which it is meant to represent by closer than 1° ; and it is highly probable that this is too precise an estimate.

Thus the data on which meteorological prediction is to be done represent a very sketchy sampling of the true data which include every local gust of wind and every cool spot or warm spot in every area. Perhaps it may be possible to maintain that these local fluctuations are unimportant in the development of the weather. It is quite conceivable that the general outlines of the weather give us a good, large picture of its course for hours or possibly even for days. However, I am profoundly skeptical of the unimportance of the unobserved part of the weather for longer periods. To assume that these factors which determine the infinitely complicated pattern of the winds and the temperature will not in the long run play their share in determining major features of weather, is to ignore the very real possibility of the self-amplification of small details in the weather map. A tornado is a highly local phenomenon, and apparent trifles of no great extent may determine its exact track. Even a hurricane is probably fairly local where it starts, and phenomena of no great importance there may change its ultimate track by hundreds of miles. Meteorology is a living exemplification of the old proverb:

“For want of a nail, the shoe was lost;
 For want of a shoe, the horse was lost;
 For want of a horse, the rider was lost;
 For want of a rider, the battle was lost;
 For want of a battle, the kingdom was lost!”

What I am looking for then is a rational account of the sort of meteorological prediction which can be carried out on the basis of observations which are sparse in space, sparse in time, and incomplete in precision. The basis of all this is ergodic theory.

It was familiar long before Gibbs that in a conservative dynamical system there is an invariant measure. In case the quantities p_n stand for the momenta of such a system and the quantities q_n for the coordinates, the element of the invariant measure is $dp_n dq_n$. This invariant measure may be finite or infinite for the total universe of discourse concerned. In many cases we do not know whether it will be finite or infinite. In all cases, this invariant measure generates a conditional invariant measure over each surface in phase space which is of constant energy.

If we take the system of a sun and a planet with the planet rotating around the sun according to Newtonian laws in an elliptical orbit, it is quite clear that for that energy level the total invariant measure will be finite. If on the other hand, we consider a comet moving in a hyperbolic orbit, we should naturally expect the total invariant measure to be infinite and therefore quite unusable as a probability. When it comes to the solar system as a whole, we probably have an infinite total invariant measure, because the gravitational potential and kinetic energy of the solar system, when combined, are enough to reject a smaller planet like Mercury to infinity.

Whether a system have finite or infinite total invariant measure, it may be that this invariant measure may be represented as a sum or integral of invariant measures over smaller sets, and that these partial invariant measures may be finite. What the actual state of affairs in the solar system is we do not know. The question we are asking here is one way of putting the question of the stability of the solar system. If there are no smaller invariant measures than the invariant measure which belongs to the known invariants of the solar system, which are energy, momentum, and moment of momentum, it is quite clear that the solar system is unstable and that sooner or later it will blow up on the basis of its purely gravitational forces, and will reject some planet to infinity. It is conceivable although implausible that there is in fact a more restricted invariant measure and that the solar system is stable.

For the purposes of this discussion, let me now confine my consideration to dynamical systems with a finite invariant measure.

This invariant measure may or may not have its origin in the fact that we are working with a conservative system. For in addition to the invariant measure which derives itself in the Gibbsian manner from the dynamics of the conservative systems, there are also nonconservative systems which either actually or very plausibly contain an invariant measure. When, for example, we have a wind tunnel in which the entering air is blown through a screen by a fan and in which the air leaving the tunnel is water-cooled to a steady temperature before it arrives at the fan, we are very far from a conservative system, for the power input of the fan is large and all the energy eventually leaves the wind tunnel in the cooling water. Nevertheless, it is highly plausible that we can treat the state of the air in the wind tunnel as in

statistical equilibrium. Some of the basis for discussing nonconservative systems in statistical equilibrium, or what is the same thing, with an invariant measure, has been given by the Russians, Kryloff and Bogolyuboff.

For the present let us then consider that a stable dynamical system is simply one which has a finite measure invariant with the time, and let us suppose that the dynamics of meteorology is of this sort. In the theory of systems with an invariant finite measure, the fundamental theorem is that of Birkhoff. This theorem known as the ergodic theorem states that if we have a measure-preserving transformation T of the segment $0 \leq \alpha \leq 1$ into itself, then if f is a measurable function of α belonging to L , we shall have that

$$(1) \quad \lim_{n \rightarrow \infty} \frac{f(\alpha) + f(T\alpha) + \dots + f(T^n\alpha)}{n + 1}$$

will almost everywhere be a function $f^*(\alpha)$ which belongs to L . We shall furthermore have

$$(2) \quad \lim_{n \rightarrow \infty} \frac{f(\alpha) + f(T^{-1}\alpha) + \dots + f(T^{-n}\alpha)}{n + 1} = f^*(\alpha)$$

almost everywhere.

This result is perfectly general. However, there will be certain cases in which for every function f of class L , $f^*(\alpha)$ will almost everywhere be equal to $\int_0^1 f(\beta) d\beta$. This case is known as the ergodic or metrical transitive case, and the necessary and sufficient conditions for metrical transitivity will be that if there is any measurable set of values for α which is invariant under the transformation T , this set will have measure either 0 or 1. By no means are all measure-preserving transformations metrical transitive, and we cannot even say in a single clear, unambiguous sense whether metrical transitivity is the norm or the exception. It has been shown by various writers that this depends on the particular topology in which we work. However, there is a theorem which shows that in a very special sense metrical transitivity is the norm. There is the theory of von Neumann (see p. 617, [1]) which asserts that if we have any measure-preserving transformation of the segment $(0, 1)$ into itself, then this segment will be divisible into what we may call components. It will be metrical transitive with respect to the measures on the components which include almost all of the segment $(0, 1)$. It will then be possible to build up a measure on the whole segment $(0, 1)$ as a sum or integral of measures on the components. In other words, if S is any measurable set on $(0, 1)$, and if we take the partial measures of S on the components, which is equivalent to saying that we take the measure on each component on the intersection of S with that component, the total measure of S may be built up as a linear combination of the measure of the components with a positive weighting. This means that if we take the $f^*(\alpha)$ which we have already defined, it will almost always be equal to the integral of $f(\beta)$ over the component to which α belongs, with the measure belonging to that component. A restatement of this is the following: whenever we have given a measure-preserving transformation T of the segment $0 \leq \alpha \leq 1$ and whenever we form $f^*(\alpha)$, we can replace the segment by a subset containing α over which there will be a metrical transitive invariant measure. In this, of course, we make an exception of a set of values of α of zero measure. Thus if we work backward from the definition of $f^*(\alpha)$, and if we

have the entire past history of $f^*(\alpha)$, we have a unique metrically transitive dynamics in which α averages may be replaced by time averages. Time here pertains to the exponent of the transformation T , and these time averages can almost always be obtained by considering the past history.

This means that if we know that a process has an invariant measure, and if we know the entire past history of the process in a particular case, then we gain nothing from the standpoint of prediction by knowing its dynamics as well. This is to say that the past history together with the knowledge that there is a dynamics gives us that dynamics in full.

Suppose then that we consider the way in which a limited set of past observations determines a quantity in the future. These observations tell us that certain parameters dependent on α belong in certain ranges, or in other words, they give us a set of functions $\phi_n(\alpha)$ which are characteristic functions assuming only the values 1 and 0. The determination of the distribution of some quantities depending on α in the future amounts then to the determination of the conditional distribution of this future parameter when the past imperfectly accurate observations are given. These conditional probabilities amount to this: we know the measure of the intersection of every region for the future parameters which is integrable in α when we know the measure of that part of the region of α consistent with the past measurements. This compound measure we obtain by translating the whole situation backward in time until all the measurements in it, the future ones as well as the past ones, come to lie in the past. We then average over past history and we obtain our probability for the future conjunction of measurements. This represents all that our data can possibly give us concerning the future, and we gain nothing except convenience and speed by the assumption of any specific dynamic hypothesis. Even in the case in which the full dynamics involves not merely the integral powers T^n of the transformation T , but the complete group T^λ , this approach gives us everything which we may obtain by approximate observations at discrete, equally spaced intervals of time.

Thus the theory of prediction even in the nonlinear case gives us the distributions for the future as linear combinations of observed distributions in the past. From the distribution point of view, the nonlinear feature of nonlinear prediction lies not in the nonlinearity of the method of combination of past data, because this combination has now become linear; but rather in the fact that the distributions of the past, on which we predict the future, are not exclusively the distributions of observed quantities at one time in the past, but combined distributions at various times. This suggests to us what the function of hypothesis is in prediction theory. The purpose of hypothesis is to cut down the interminable work of projecting against every possible combination of past observations and to replace it by a projection on a smaller set of combinations of past observations. In other words, hypothesis is fundamentally a statement that after certain elementary projections upon the past have been accomplished, the further advantage of projecting on other past combinations of data is small or zero. I have said that this can be observed to some extent, but of course a full observation of this would be tantamount to a complete use of the entire past. Thus hypothesis consists in a guess, which is only useful because it is not fully justified, that certain sorts of past projection do not add information to that which we already possess on the basis of other sorts of past projections.

In connection with the use of hypothesis, I want to point out a certain duality running through the entire texture of science which enables us to make a significant separation between the methodology of the highly exact sciences such as celestial mechanics, and the semi-exact sciences such as meteorology or econometrics. The traditional Newtonian way of predicting in the highly exact sciences had the method which is used in constructing a nautical almanac or a set of ballistic tables. That is, to assume the validity of a dynamics and then to start with a certain set of initial data as if they were precisely known. Having then obtained a fairly precise account of the constants of the system at a future time, we do a separate investigation to determine how much a shift of initial conditions will affect the final conditions. In a set of ballistic systems, for example, we indicate the derivatives of the range, etc., with respect to small shifts in the angle of elevation, the muzzle velocity, and the other initial constants. If the ensemble of initial positions taken is sufficiently small and compact, these derivatives will give us quite a reasonably good account of the distribution of final positions and momenta in terms of the distributions of initial positions and momenta.

If, however, the distribution of initial positions and momenta is broad, the approximation which we get to the distribution of final positions and momenta will no longer be valid, as the first derivative terms in their expansion will only form a small part of their development. Thus when our observational errors are large, we can no longer consider the method of straight dynamics plus variation as giving us a precise estimate of final positions. It is therefore not enough to study the history by which a precise point in phase space gives rise to a precise point in another part and the first order corrections of the results of such an investigation; but we must take seriously the point of view of Willard Gibbs, according to which we consider a genuine flow between a somewhat broad ensemble of initial positions in phase space and the corresponding final positions in phase space. The differential equation gives way to the integral equation.

Behind this is a fundamental philosophical fact associated with instruments and methods of computation. Indeed there is no really deep distinction between a predicting instrument which follows the present and past and indicates the future, and a computational method which enables us to do the same sort of thing numerically either by a hand process of computation or by a sequence of operations adapted to digital or analog computing machines. Precise instruments or methods of computation are delicate and can be thrown off very badly by irregular modifications in the input data. A very fine galvanometer will go into serious oscillation if it is exposed to an input which varies in a widely irregular manner. The more delicate it is, the longer it will take for this oscillation to die out or damp down, and the less accurately the reading of the machine or process will correspond at a particular time to the exact result which we could get if there were no such oscillation. Thus for rough data, it may and generally will pay to use a more stable and consequently less precise instrument. The final errors of an instrument or computational method represent a compromise between the errors of inaccuracy belonging to the instrument or method, and the errors of instability; and the best result will be obtained by a proper balance between these errors. This balance can only be reached on a statistical basis.

Any attempt to throw one sort of error to zero, throws the other sort of error to

infinity. It is only a statistical knowledge of the ensemble of inputs for which the instrument or method is designed which can lead us to optimal results. This statistical element and this duality of errors belong to all computation, but I have already called attention to it in connection with the sort of prediction problem which arises in anti-aircraft fire (see p. 70, [4]). In other words, it is bad technique to apply the sort of scientific method which belongs to the precise, smooth operations of astronomy or ballistics to a science in which the statistics of errors is wide and the precision of observations is small. In the semi-exact sciences, in which observations have this character, the technique must be more explicitly statistical and less dynamic than in astronomy.

Accordingly, there is a grammar for the semi-exact sciences differing considerably from that which is appropriate in the case of the exact sciences. The errors of observation are part of the observation itself, and any separation between precise observations and errors will lead in the case of the semi-exact sciences to a methodology which actually is less accurate for the results we really desire than a broader, less obviously precise method. This even applies to many branches of engineering, such as the study of oscillation in buildings, in which the traditional methods have been purely dynamic.

There is a further closely related method of linearizing statistics of time series which needs further investigation as to its relation with the method I have already given. Let $f_n(\alpha)$ be a set of measurable functions of α , such that all of their powers are Lebesgue integrable. Let T be the measure-preserving transform of the segment $(0, 1)$ into itself. Now form all the combinations $f_1(T^{m_1}\alpha)f_2(T^{m_2}\alpha) \cdots f_r(T^{m_r}\alpha)$. This whole system of nonlinear combinations of the f 's may be regarded as a larger set of $g_n(\alpha)$. In this set, $g_n(\alpha)$, we can develop a linear theory of multiple prediction. It then requires no great effort to transfer this theory to a nonlinear theory of the prediction of the functions $f_n(\alpha)$ themselves. The autocorrelations demanded are expressions of the function

$$\int_0^1 f_1(T^{m_1}\alpha)f_2(T^{m_2}\alpha) \cdots f_r(T^{m_r}\alpha)d\alpha$$

provided that all the functions $f_n(\alpha)$ are taken as real. Here again we develop a nonlinear theory of predictions which is only nonlinear because it is a linear prediction in terms of nonlinear combinations of the various time series $f_n(\alpha)$. To develop the details of this theory we may then transfer our entire attention to the theory of multiple linear prediction.

REFERENCES

- [1] J. VON NEUMANN, "Zur Operatorenmethode in der Klassischen Mechanik," *Annals of Math.*, Vol. 33 (1932), pp. 587-642.
- [2] P. R. HALMOS, "The decomposition of measures," *Duke Math. Jour.*, Vol. 8 (1941), pp. 386-392.
- [3] P. R. HALMOS, "On a theorem of Dieudonné," *Proc. Nat. Acad. Sci.*, Vol. 35 (1949), pp. 38-42.
- [4] N. WIENER, *The Extrapolation, Interpolation, and Smoothing of Stationary Time Series*, New York, The Technology Press and John Wiley and Sons, Inc., 1949.