ADVANCES IN FOOD AND NUTRITION RESEARCH, VOL. 40

Chapter 20

DYNAMIC SYSTEMS AND NEURAL NETWORKS: MODELING IN PHYSIOLOGY AND MEDICINE

SAMIR I. SAYEGH

Physics Department
Purdue University
Ft. Wayne, Indiana 46805

I. INTRODUCTION

This paper deals with a general introduction to dynamic systems as well as a brief introduction to neural networks. In particular, it is intended for those involved in modeling physiological and biochemical systems where the questions of appropriate dimension and existence of a model are important considerations. Other complementary approaches to modeling should be considered depending on the problem at hand (Collins, 1992; Webber and Zbilut, 1994; van Rossum *et al.*, 1989).

One of the main goals of the following paragraphs is to emphasize the importance for scientists involved in modeling to understand the implications and limitations of dimensionality, nonlinearity, and the related nature of the problem they are attempting to solve. The choice of a modeling approach is strongly determined by such considerations. One of the great advantages of neural networks and other techniques of multivariate analysis is that they are applicable to phenomena with a limited level of definition (Benigni and Giuliani, 1994). Advantages of dimensionality reduction include "escaping chaos" and allowing for the qualitative analysis of the system at hand.

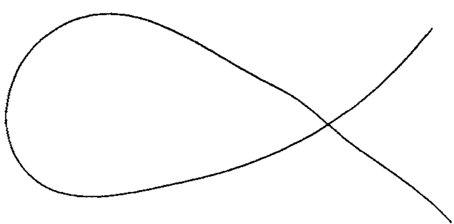
II. LINEAR SYSTEMS MODELING

Whether one is dealing with a system by using dynamic systems or neural network approaches, the level of difficulty involved is quite high for nonlinear systems and relatively trivial for linear systems. The most general linear system can be represented by the equation $dx/dt = x$ or $dx/dt = \alpha x$. Here x can represent any vector of variables that depends on the independent variable which in general stands for time. Such a linear system is limited in its capability to represent complex phenomena and exhibits only a couple of behaviors. The first kind of behavior is exponential decay. Indeed a function that is its own derivative is clearly an exponential function. With a negative exponent, one gets exponential decay. With a complex exponent or a pure imaginary exponent, one gets an oscillatory behavior which typically translates into a limit cycle. So if we talk in terms of attractors, namely, a structure that will attract the motion in the sense that the evolution of a system will end up being such a portion of a phase space, then the only attractors that are associated with linear systems will be point attractors, i.e., fix points, and limit cycles, i.e., periodic behavior. Of course, one could also get a positive exponentially developing solution, but that would require an infinite amount of energy that is not assumed to be available to the systems we usually model. Therefore, in a sense, linear systems can be considered to be trivial. From a computational point of view all one needs to do is to compute an exponential. That could be an exponential of a constant multiplied by the time as it develops or a sine or cosine function. More generally, if the dimension of the system is higher than one, one is dealing with a constant which is a matrix and in order to exponentiate a matrix, the standard approach is to diagonalize the matrix and exponentiate the eigenvalues, again a standard procedure that is built into a number of commercial software packages. The motion can then be essentially reconstructed by these techniques. Though these techniques can display some subtleties, one can still in a general sense classify these as very simple techniques in comparison to the complexity that one encounters with nonlinear systems.

III. NONLINEAR SYSTEMS, CHAOS, AND FRACTIONAL DIMENSIONS

As soon as one deals with nonlinear systems, one encounters a richness that is absent from linear systems. One of the richer aspects of nonlinear systems has been called chaos and one of the characteristics of chaos is the sensitive dependence on initial conditions. What could be discussed at this point is a definition of chaos and some development of this definition, particularly in relationship to the notion of the dimension of the system. The main goal of dimensional considerations should be the characterization of the system at hand, the identification of the most appropriate modeling approach, and attempts at the reduction of the dimensionality. There are at least two reasons to try and reduce the dimensionality of a system. One reason is to avoid chaos or to show that the underlying system is nonchaotic, which quite often is a very important result to establish. The other reason is to be able to perform a qualitative analysis of the system and to gain insight into the workings of the dynamics.

The working definition of chaos is essentially a sensitive dependence on initial conditions, which means that slightly different initial conditions lead to very different final conditions or very different states after a finite time of evolution of a system. In order to understand the geometric interpretation of dimensionality considerations, one has to look at the concept of the phase space in which only the dependent variables are represented. Figure 1 shows such a construct. Points in phase space represent the state of the system at a given time. As time evolves and the state of the system changes, the point traces a curve. As Fig. 1 shows again, this curve is not allowed to intersect itself if the system is to be considered deterministic. The reason is that if the system is deterministic, the complete specification of the state of the system given by a point should predict all future evolution of the system. At an intersection point, we would clearly have two different evolutions of the system that are diverging from the same point, that is, the same initial state is leading to different futures, in violation of the assumption of determinism. Therefore, one cannot assume any intersection of the curve tracing the evolution of the system. Now since chaos has been defined as "the exponential separation of trajectories leading to sensitive dependence on initial conditions," and if one is in the presence of a system with finite



No Determinism

FIG. 1.

energy, or, better yet, a dissipative system that is losing energy, as the system loses energy trajectories have to become restricted to a limited region of phase space. The evolution in time cannot lead to explosive behavior and has to remain finite or in a finite boundary. The curves have to return therefore to the region in phase space from which they originated. However, as they return, they are not allowed to intersect, in order to preserve determinism. This leads to the contradiction between a low-dimensional system, i.e., a two-dimensional system, and the existence of chaos.

The conclusion, therefore, is that a two-dimensional system cannot display chaotic behavior, i.e., there is not enough room in two dimensions to allow for the exponential separation of the trajectories while preserving the constraint of no intersection that preserves determinism (Fig. 2). One can see how in a higher dimensional system, if one is allowed a little more than two dimensions, one could preserve both chaotic behavior and determinism. That is illustrated in Fig. 3 where the trajectory is returning to the general region of interest where it originated. Instead of intersecting, it crosses over and it crosses over because it has been given a certain amount of extra "space" in the form of increased dimensionality.

In a three dimensional system that is dissipative, i.e., that loses energy, the dimensionality of the system collapses. That is, the potential dimension of the motion is going to be less than three because the system is losing energy. However, if the system is a chaotic one, the eventual motion cannot be on a two-dimensional manifold, for the reasons explained above. The conclusion therefore, is that, the eventual structures on which the trajectory ends up evolving has dimension less than three but more than two (Fig.

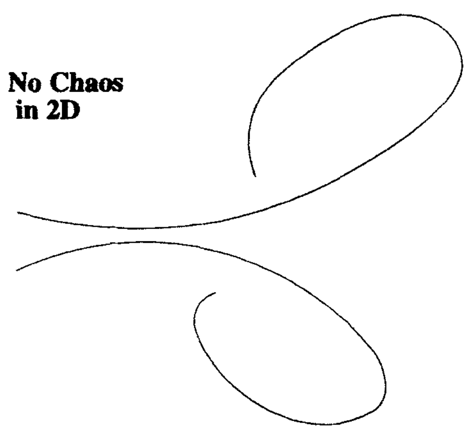


FIG. 2.

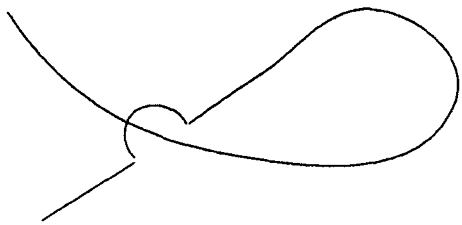


FIG. 3.

4). In a three-dimensional differential equation the system is going to collapse on a structure that is strictly less than three in dimension, but strictly more than two. Such a strange observation leads us immediately to the concept of fractional dimensionality and the attractor associated with such a fractional dimensionality is called a strange attractor, due to its unusual nature. The fractional dimensionality is also described as a "fractal."

A fractal object such as a "C curve" may have some unusual properties. The properties are that it has a fractal dimension but this fractal dimension is not a fraction. In the case of the C curve it is equal to two. The reason this object is still a fractal relates to a definition of fractal dimension. First, one defines two concepts of dimensions: the topological dimension, which corresponds to our usual concept of a dimension, and a so-called Hausdorff–Besicovic dimension. If for a given object the two dimensions defined are different, the object is said to have a fractal dimension. In the case of the

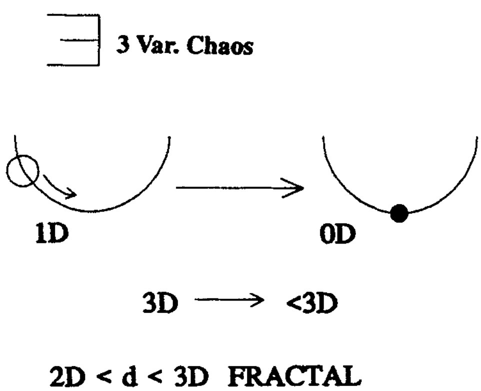


FIG. 4.

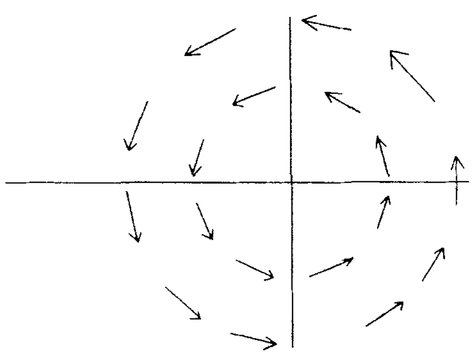
C curve, the topological dimension is equal to one and the Hausdorff–Besicovic dimension is equal to two and that is why it is considered fractal, despite the fact that neither of these dimensions is itself fractional. In most cases, however, a fractal has a fractional dimension.

IV. GEOMETRIC INTERPRETATION OF MODELS AND DYNAMICAL SYSTEMS

The following discussion will concentrate on a geometric definition of differential equations and on actual reduction of a dimensionality. As Fig. 5 shows, a differential equation can be considered to be a collection of arrows. As Fig. 6 shows, the solution to the differential equation consists of a collection of trajectories which are everywhere tangent to the collection of arrows.

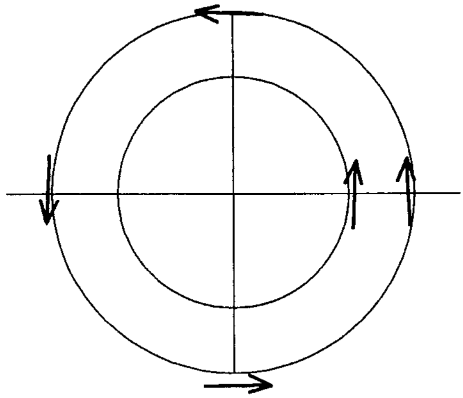
Figures 7 and 8 show another example of a differential equation, with a collection of arrows, in phase space, and their solution, the trajectories, that are tangent to the arrows. In Fig. 9, the two differential equations are displayed simultaneously. The intent of that figure is to show that these two particular differential equations have a very special property, namely, that they are orthogonal, in the sense that the arrows of one differential equation take the arrows of the other differential equation into each other. That is, it maps arrows of one equation into arrows of the other. As a consequence, if one were to draw a solution to one of the differential equations, then the arrows of the other differential equation would take

How to reduce d?



A Differential Equation (U)

FIG. 5.



A Solution

FIG. 6.

this solution into another solution. This is illustrated in Fig. 10. Having this property, if one could display just one solution to a differential equation, the arrows of the other equation would generate all other solutions. One can intuitively see that such a construct is going to collapse/reduce the dimension of a differential equation by one. It turns out that the mathematical condition for such a situation to arise is rather simple and is expressed by the commutator of the arrows corresponding to the two differential

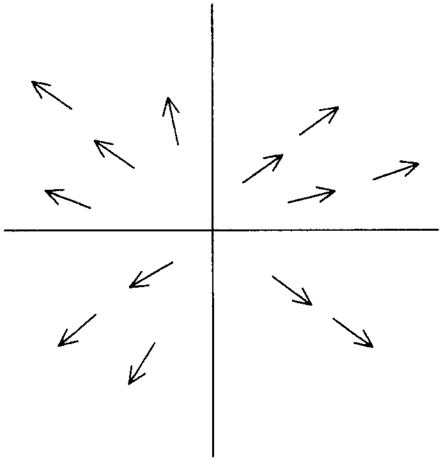


FIG. 7.

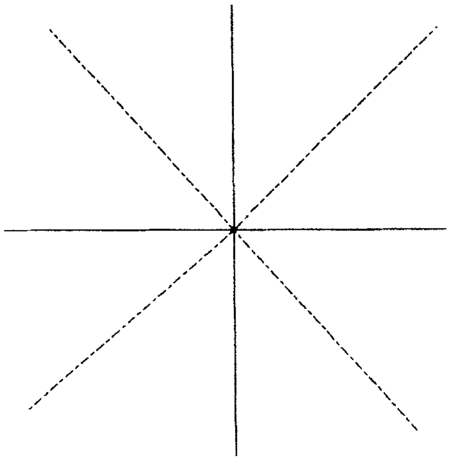


FIG. 8.

equations being zero. The computation of the commutator is very inexpensive and can be performed by hand or by any symbolic manipulation software such as Reduce, Maple, or Mathematica. There is also a slightly more general formulation where the commutator can be simply equal to a multiple of one of the differential equations. It is worthwhile to note that the commutator condition expressed, which might be unfamiliar to some

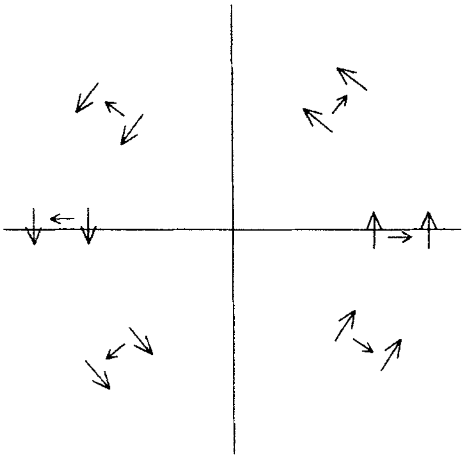


FIG. 9.

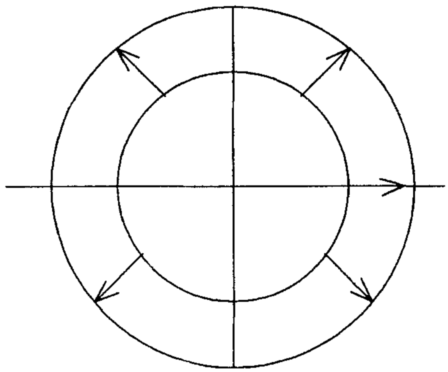


FIG. 10.

readers, encompasses most techniques of integration of elementary differential equations. The importance of such reduction techniques in low-dimensional models is that they can make the difference between the need to deal with chaotic system and the possibility of avoiding them altogether (Sayegh and Jones, 1986).

We now examine another aspect of dimensionality reduction, namely, the one that provides for a qualitative analysis of a differential equation. A very good example, physiologically of great interest, is that of the Hodgkin–Huxley equations. Basically the Hodgkin–Huxley equations give rise to behavior of the voltage across a membrane as a function of time. So one has an axon and the voltage across the axon is a function of time. One particular time behavior is that of an action potential development and subsequent evolution of that action potential as a function of time. The equations that were written by Hodgkin and Huxley can basically be understood in the following simplified or modified description. Namely, one variable will represent voltage, and several other variables will describe the conductances of the ion channels across which sodium and potassium ions flow. These conductances themselves are functions of voltage and time and need therefore be defined in terms of a differential equation.

While historically things were described a little differently, all in all, Hodgkin and Huxley had one equation for the membrane potential and three more for describing conductance. They had a total of four differential equations. Four differential equations is a manageable number and one that might benefit from reduction because two is very close to four. Indeed, two prevents chaos and allows two dimensional qualitative analysis of the differential equation. The reduction to two dimensions is precisely what Nagumo *et al.* (1962) and Fitzhugh (1961) have done. They have done a two-dimension analysis of a Hodgkin and Huxley equation. So, essentially by eliminating the phase dynamic and by lumping two of the variables together, Fitzhugh and Nagumo were able to write a two-dimensional system that is represented by

$$\frac{dv}{dt} = v(a - v)(v - 1) - w$$

$$\frac{dw}{dt} = bv - yw.$$

Figure 11 illustrates the $dv/dt = 0$ and the $dw/dt = 0$ curves which partition the two-dimensional phase space into several regions, each region having a certain sign for dv/dt and dw/dt. For example, the lower right quadrant below the sine-like $dv/dt = 0$ curve would have $dv/dt > 0$ and $dw/dt > 0$. One can start with a zero potential through the injection of a current raise the potential to a certain value. Through simple qualitative analysis that takes into account only the signs one can trace the evolution of the system under different excitation conditions. Through such an analysis, one first establishes the existence of a threshold; i.e., if the system is not pushed hard enough, there is no action potential that results and the system settles back to the point and to the voltage that represents the resting condition. However, if the system is pushed hard, current is injected in such a way as to exceed a certain threshold and an action potential results. This can be traced qualitatively and it does display all the known characteristics of an action potential as represented in Fig. 12.

The technique of dimensionality reduction is particularly powerful but obviously has some limitations. One obvious limitation is in cases where

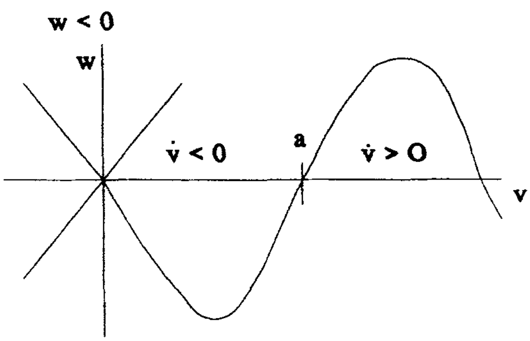


FIG. 11.

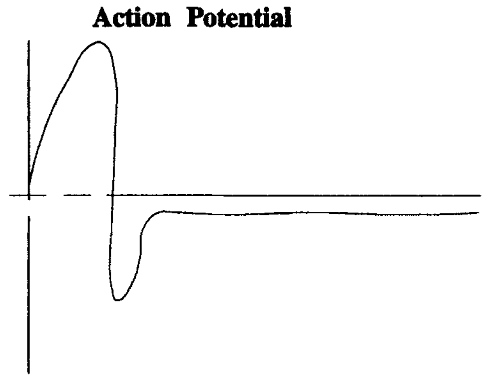


FIG. 12.

the number of variables is very large. An example which we have been interested in recently is that of modeling a portion of the hippocampus in the place cell phenomenon as well as the onset of epilepsy (Traub and Miles, 1991; Jaboori *et al.*, 1995). An assembly of 10,000 neurons each is represented by about 20 compartments and each compartment represented by about five variables. Thus one has on the order of a million variables to deal with, and the reduction of the system by one variable is useless. Other techniques, perhaps from statistical physics, should be brought to bear on such high-dimensional problems. It is not unusual, however, that realistic good models relating to physiology have a lower dimension that can be possibly amenable to the techniques presented above.

V. NEURAL NETWORKS

The remaining is a discussion of some historical and introductory aspects of neural networks. Figure 13 represents a brain as one of the most naive and minimal models of a learning system consisting of two neurons and one synaptic weight. Figure 14 is an extension showing the same two neurons and a synaptic weight. The underlying principle of a weight implementing a mapping between input x and output y through a simple multiplication of the input by a weight w. The problem of mapping or learning as represented is the problem of finding the perfect w such that $y = wx$ where x and y are known. Clearly one can find w through simple division, and the problem is solved.

The above problem of learning can be generalized in at least three different directions. The first direction for generalization simply involves

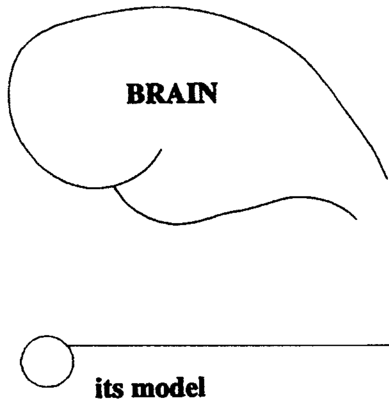


FIG. 13.

an increase in the number of input and output lines in order to represent a more complex mapping of patterns or vectors to be learned. The second direction is that of an increase in the number of patterns where a large number of patterns becomes allowed, possibly largely exceeding the number of weights or parameters in the learning system. In the case of linear problems, it turns out that the solution to these generalized problems is still essentially given by a "generalized division" or matrix inversion. The above two generalizations, however, are restricted to solving a class of relatively simple problems. The third direction of generalization is that of a more complex network that is needed to handle more complex problems. The network is more complex both in the sense that it has at least an additional layer of nodes and weights (a so-called hidden layer) and the fact that it needs to introduce a degree of nonlinear processing in order to allow for the learning of the more complex tasks.

The standard procedure that is then used to find the unknown weights is that of gradient descent, i.e., a smooth continuous motion in the direction

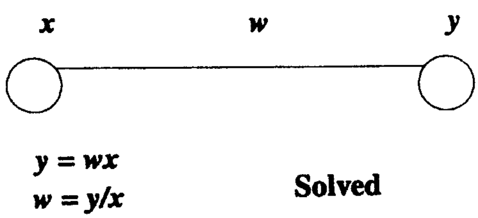


FIG. 14.

that minimizes an error function. Another way of seeing the search procedure that is needed is simply to think in terms of feedback. If one is to start from the arbitrary value of the weight w, produce an output using that weight w, and then change the w via feedback in proportion to the error produced, one does get the same procedure as the one dictated by gradient descent. Notice one must also change w in proportion to the input in order to give credit to different inputs, i.e., if an input has caused a large error, one would like to correct proportionally to that input line. Now the feedback equation can be turned into a differential equation where the change in weight is turned into dw and time is introduced in the form of dt. When one looks at this equation, one sees that the derivative of w is now proportional to w itself plus some extra constant term. As indicated in the beginning of this article, these equations have simple exponential behavior. In this particular case, the initial weight w exponentially can evolve toward the best solution w^*. w^* can be evaluated and turns out to be the same as the minimum of the error function as previously discussed.

The history of neural networks, at least in its popular version, has its angels and demons. One of them is Marvin Minsky, referred to as the devil. Minsky and Papert wrote a book entitled Perceptrons (1969) in which they showed, among other things, some of the limitations of the neural networks that were popular at the time. This was interpreted by some as a very restrictive limit on all kinds of neural networks and it was interpreted by some as being the end of an era of funding for the then popular neural networks in order to make room for the more classical AI approaches promoted by Minsky and others.

In 1982, John Hopfield published a paper that dealt with associative memory that had a compelling analogy to a system of electron spins and could therefore be treated in a reasonably rigorous fashion. It is believed by some that the combination of Hopfield's reputation and the timing of the publication of the paper has given a new impetus to the field of neural networks that was reinforced by the error backpropagation algorithm (Rumelhart and McClelland, 1986) or, perhaps, the reintroduction of an algorithm that was discovered in the early 1970s but remained unnoticed (Werbos, 1974). Here one wants to think of neural networks as universal mapping devices taking any input sequence to any output sequence of patterns. The true angel was Kolmogorov, (1963) who has proved a theorem in the theory of function of several variables that translate to the universality of a class of neural networks. The neural networks that are universal, however, need to be nonlinear. Linear networks do suffer from the limitations that were outlined by Minsky and others and briefly discussed above. It is therefore necessary to apply nonlinear networks to hard problems. The disadvantage of such an approach is the appearance of local minima that have plagued search techniques in the past (Fig. 15). In other words as one uses more powerful techniques allowing one to handle tougher problems, one is no longer guaranteed that the solutions found will be optimal.

The following summarize the advantages of neural networks.

(1) Any method can be implemented, no matter how complex, provided one has a complex enough network to perform such a task. A complex enough network usually means a network with at least three layers of nodes and nonlinear units and an undefined number of hidden units in the intermediate layer between input and output.
(2) Neural networks can generalize from the input–output pattern that have been shown. That is, through training, not only do they fit the given input and output, they fit them in such a way that new, previously unseen patterns can be predicted correctly in a large number of cases. An underlying physiological model need not be present.
(3) This implies a certain robustness of neural network to perturbation of the patterns that were originally shown as well as a robustness to perturbation of the network itself in the form of destruction or modification or some of the nodes.
(4) Neural networks are universal and friendly. The universality aspects usually mean the same as the implementation of arbitrary mapping as in (1) while the friendliness stems from a generalized architecture that is generally provided nowadays in a large number of commercially available software packages which come with a very similar interface. The uninitiated user can quickly adapt to such systems and switch comfortably between different systems.
(5) One of the great advantages of neural networks is that the problem is formulated directly on its architecture, i.e., one can solve the problem

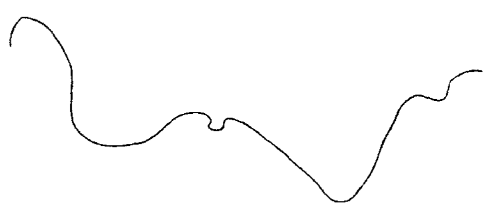


FIG. 15.

and formulate the corresponding parallel architecture simultaneously. This is to be contrasted with algorithmic approaches where the problem is first solved and, as the problem is solved, a fast, probably parallel architecture is sought to efficiently implement the solution.

VI. CONCLUSION

Once again one can simply conclude that linear systems, although they enjoy a mathematical apparatus of great simplicity and power, have their limitations as to the richness of representation and the limited class of problems they can address. It is only when one deals with nonlinear problems and nonlinear techniques that the wealth of realistic physiological and physical phenomena can be tackled. The use of statistical techniques should be reserved for problems with a very large number of variables. However, for the more common physiologically realistic situations a combination of dynamical system techniques and neural network processing can yield powerful, useful, and elegant results.

REFERENCES

Benigni, R., and Giuliani, A. (1994). Quantitative modeling and biology: The multivariate approach. *Am. J. Physiol.* **266**(5, Pt. 2), R1697–R1704.

Collins, J. C. (1992). Resources for getting started in modeling. *J. Nutr.* **122,** Suppl. 3, 695–700.

Fitzhugh, R. (1961) Impulses and physiological states in theoretical models of nerve membrane. *Biophys. J.* **1,** 445–466.

Hopfield, J. J. (1982). Neural networks and physical systems with emergent collective computational properties. *Proc. Natl. Acad. Sci. U.S.A.* **79.**

Jaboori, S., Sampat, P., and Sayegh, S. (1995). Analyzing the hippocampal place cell phenomenon by modeling the visual pathway. *In* "The Neurobiology of Computation: Proceedings of the Third Annual Computation and Neural Systems Conference" (J. Bower, ed.). Kluwer Academic Publishers, New York.

Kolmogorov, A. N. (1963). On the representation of continuous functions of many variables by superposition of continuous functions of one variable and addition. *Dokl. Akad. Nauk SSSR* **144,** 679–681. Ame. Math. Soc. (*Engl. Transl.*) **28,** 55–59.

Minsky, M., and Papert, S. (1969). "Perceptrons." MIT Press, Cambridge, MA.

Nagumo, J. S., Arimoto, S., and Yoshizawa, S. (1962). An active pulse transmission line simulating nerve axon. *Proc IRE* **50,** 2061–2071.

Rumelhart, D. E., and McClelland, J. L. (1986). "Parallel Distributed Processing." MIT Press, Cambridge, MA.

Sayegh, S. I., and Jones, G. L. (1986). Symmetries of differential equations. *J. Phys. A: Math. Gen.* **19,** 1793–1800.

Traub, R. D., and Miles, R. (1991). "Neuronal Networks of the Hippocampus." Cambridge Univ. Press, New York.

van Rossum J. M., de Bie J. E., van Lingen, G., and Teeuwen, H. W. (1989). Pharmacokinetics from a dynamical systems point of view. *J Pharmacokinet. Biopharm.* **17**(3), 365–392.

Webber C.L., Jr., and Zbilut, J. P. (1994). Dynamical assessment of physiological systems and states using recurrence plot strategies. *J. Appl. Physiol.* **76**(2), 965–973.

Werbos, P. (1974). Beyond regression: New tools for prediction and analysis in the behavioral sciences. Ph.D. Thesis, Harvard University Committee on Applied Mathematics, Cambridge, MA.