uses all available information; it is related to the problem of reconstructing a 3D structure from a 2D projection. The operation has been fully described and is now available in Mathematica.

The generalized inverse also enables one to handle redundant axes in quasicrystals, but usually the interesting problems are nonlinear. Other inverse problems include the following.

(i) Finding the arrangement of atoms that gives rise to the observed scattering patterns of X-rays or electrons from a crystal.

(ii) Reconstructing a 3D image from 2D projections in microscopy or X-ray tomography.

(iii) Reconstructing the geometry of a molecule given probable interatomic distances (and perhaps bond angles and torsion angles).

(iv) Finding the way in which a protein molecule folds to give an active site, given the sequence of constituent amino acids.

(v) Finding the pathway to producing a molecule synthetically, given that it occurs in nature.

(vi) Finding the sequence of rules that generate a membrane or a plant or another biological object, given that it takes a certain shape.

Some questions of this type do not have unique answers. For example, the classic question as to whether the shape of a drumhead can be determined from its vibration spectrum (can you hear the shape of a drum?) has been answered in the negative: two vibrating membranes with different shapes may have the same spectrum. It was thought that this ambiguity might also be the case for crystal structures. Linus Pauling suggested that there might be two different crystal structures that were *homometric* (that is, giving the same diffraction pattern), but no definite example has been found.

5 Conclusion

As the examples in this article show, mathematics and chemistry have a symbiotic relationship, with developments in one often stimulating advances in the other. Many interesting problems, including several that we have mentioned here, are still waiting to be solved.

Further Reading

Cotton, F. A. 1990. *Chemical Applications of Group Theory.* New York: Wiley Interscience.

Hollas, J. M. 2003. *Modern Spectroscopy.* New York: John Wiley.

Hyde, S., S. Andersson, K. Larsson, Z. Blum, T. Landh, S. Lidin, and B. W. Ninham. 1997. *The Language of Shape. The Role of Curvature in Condensed Matter: Physics, Chemistry and Biology.* Amsterdam: Elsevier.

Parr, R. G., and W. Yang. 1989. *Density-Functional Theory of Atoms and Molecules.* Oxford: Oxford University Press.

Thomas, J. M. 2003. Poetic suggestion in chemical science. *Nova Acta Leopoldina NF* 88:109–39.

Wales, D. J. 2004. *Energy Landscapes.* Cambridge: Cambridge University Press.

Wells, A. F. 1984. *Structural Inorganic Chemistry.* Oxford: Oxford University Press.

Wolfram, S. 2002. *A New Kind of Science.* Champaign, IL: Wolfram Media.

VII.2 Mathematical Biology
Michael C. Reed

1 Introduction

Mathematical biology is an extremely large and diverse field. It studies objects ranging from molecules to global ecosystems and the mathematical methods come from many of the subdisciplines of the mathematical sciences: ordinary and partial differential equations, probability theory, numerical analysis, control theory, graph theory, combinatorics, geometry, computer science, and statistics. The most that one short article can do is to illustrate by selected examples this diversity and the range of new mathematical questions that arise naturally in the biological sciences.

2 How Do Cells Work?

From the simplest point of view, cells are large biochemical factories that take inputs and manufacture lots of intermediate products and outputs. For example, when a cell divides, its DNA must be copied and that requires the biochemical synthesis of large numbers of adenine, cytosine, guanine, and thymine molecules. Biochemical reactions are usually catalyzed by enzymes, proteins that facilitate a reaction but are not used up by it. Consider, for example, a reaction in which chemical A is converted to chemical B with the help of an enzyme E. If $a(t)$ and $b(t)$ are the respective concentrations of A and B at time t, then one typically writes down a differential equation for $b(t)$, which takes the form

$$b'(t) = f(a, b, E) + \cdots - \cdots .$$

Here, f is the rate of production, which typically depends on a, b, and E. Of course B may be produced

by other reactions (which would lead to additional positive terms $+ \cdots$) and may be used as a substrate itself in still other reactions (which would lead to additional negative terms $- \cdots$). So, given a particular cell function or biochemical pathway, we can just write down the appropriate set of nonlinear coupled ordinary differential equations for the chemical concentrations and solve it by hand or by machine computation. However, this straightforward approach is often unsuccessful. First of all, there are a lot of parameters (and variables) in these equations and measuring them in the context of real living cells is difficult. Second, different cells behave differently and may have different functions, so we would expect the parameters to be different. Third, cells are alive and change what they are doing, so the parameters may themselves be functions of time. But the greatest difficulty is that the particular pathway under study is not really isolated. Rather, it is embedded in a much larger system. How do we know that our model system will continue to behave in the same way when embedded in this larger context? We need new theorems in dynamical systems that answer questions such as this, not for general "complex systems" but for the particular kinds of complex systems that arise in important biological problems.

Cells continue to accomplish many basic tasks even though their environments (i.e., their inputs) are constantly changing. A brief example of this phenomenon, which is known as *homeostasis*, will illustrate the problem of "context." Let us suppose that the chemical reaction above is one step in the pathway for making the thymines necessary for cell division. If the cell is a cancer cell, we would like to turn off this pathway, and a reasonable way to try to do this would be to put into the cell a compound X that binds to E, thereby reducing the amount of free enzyme available to make the reaction run. Two homeostatic mechanisms immediately come into play. First, a typical reaction is inhibited by its product: that is, f decreases as b increases. This makes biological sense because it ensures that B is not overproduced. So, when the amount of free E is reduced and the rate f declines, the resulting decrease in b drives the rate up again. Second, if the rate f is lower than usual, the concentration a typically rises since A is not being used up as quickly, which also drives the rate f up again since f increases as a increases. Given the network in which A and B are embedded, one can imagine calculating how much f will drop if we put a certain amount of X into the cell. In fact, f may drop even less than we calculate because of another homeostatic

mechanism that is not even in our network. The enzyme E is a protein produced by the cell via instructions from a gene. It turns out that sometimes the concentration of free E inhibits the messenger RNA that codes for the production of E itself. Then, if we introduce X and reduce free E, the inhibition is removed and the cell automatically increases its rate of production of E, thus raising the amount of free E and with it raising the reaction rate f.

This illustrates a fundamental difficulty in studying cell biochemistry, indeed a difficulty in studying many biological systems. These systems are very large and very complex. To gain understanding, it is natural to concentrate on particular relatively simple subsystems. But one always has to be aware that the subsystems exist in a larger context that may contain variables (excluded by the simplification) that are crucial for understanding the behavior and biological function of the subsystem itself.

Although cells exhibit remarkable homeostasis, they also undergo spectacular changes. For example, cell division requires unzipping of the DNA, synthesis of two new complementary strands, the movement apart of the two new DNAs, and the pinching off of the mother cell to produce two daughters. How does a cell do all this? In the case of yeast cells, which are comparatively simple, the actions of the biochemical pathways are quite well understood, partly because of the mathematical work of John Tyson. But as our brief discussion makes clear, biochemistry is not all there is to cell division; an important additional feature is motion. Materials are being transported all the time throughout cells from one specific place to another (so their motion is not just diffusion), and indeed, cells themselves move. How does this happen? The answer is that materials are transported by special molecules called molecular motors that turn the energy of chemical bonds into mechanical force. Since bonds are formed and broken stochastically (that is, some randomness is involved), the study of molecular motors leads naturally to new questions in STOCHASTIC ORDINARY AND PARTIAL DIFFERENTIAL EQUATIONS [IV.24]. A good introduction to the mathematics of cell biology is Fall et al. (2002).

3   Genomics

To understand the mathematics that was involved in sequencing the human genome it is useful to start with the following simple question. Suppose that we cut up a line segment into smaller segments and are presented

with the pieces. If we are told the order in which the pieces came in the original segment, then we can put them back together and reconstruct the segment. In general, since there are many possible orders, we cannot reconstruct the segment without extra information of this kind. Now suppose that we have cut up the segment in *two different ways*. Think of the line segment as an interval I of real numbers, and let the pieces be $A_1, A_2, \ldots, A_r$ when you cut it up the first way, and $B_1, B_2, \ldots, B_s$ when you cut it up the other way. That is, the sets A_i form a partition of the interval I into subintervals, and the sets B_j form another partition. For simplicity, assume that no A_i shares an endpoint with any B_j, except for the two endpoints of I itself.

Suppose that we know nothing about the order in which the pieces A_i and B_j come in I. In fact, suppose that all we know about them is which A_i overlap with which B_j: that is, which of the intersections $A_i \cap B_j$ are nonempty. Can we use this information to work out the original order of the pieces A_i and thereby reconstruct the interval I (or its reflection)? The answer will sometimes be yes and sometimes no. If it is yes, then we would like to find an efficient algorithm for doing the reconstruction, and if it is no, then we would like to know how many different reconstructions are consistent with the given information. This so-called *restriction mapping problem* is really a problem in GRAPH THEORY [III.34]: the vertices of the graph correspond to the sets A_i or B_j, and there is an edge between A_i and B_j if $A_i \cap B_j \neq \varnothing$.

A second problem is whether we can find the original order of the A_i (or the B_j) if what we are told is the length of each set A_i and each set B_j, and the set of all the lengths of the intersections $A_i \cap B_j$. The catch is that we are not told which length corresponds to which intersection. This is called the *double digest problem*. Again one would like to be able to tell when there is only one solution, or to place an upper bound on the number of possible reconstructions if there is more than one.

Human DNA is, for our purposes here, a word of length approximately 3×10^9 over a four-letter alphabet A, G, C, T. That is, it is a sequence of length 3×10^9 in which each entry is A, G, C, or T. In the cell, the word is bound letter by letter to the "complementary" word, which is determined by the rule that A can only be bound to T, and C can only be bound to G. (For example, if the word is ATTGATCCTG, then the complementary word is TAACTAGGAC.) In this brief discussion we will ignore the complementary word.

Since DNA is so long (it would be approximately two meters if one stretched it out into a straight line) it is very hard to handle experimentally, but the sequence of letters in short segments of approximately five hundred letters can be determined by a process called gel chromatography. There are enzymes that cut DNA wherever specific very short sequences occur. So if we digest a DNA molecule with one of these enzymes and digest another copy with a different enzyme, we can hope to determine which fragments from the first digestion overlap fragments from the second digestion and then use techniques from the restriction mapping problem to reconstruct the original DNA molecule. The interval I corresponds to the whole DNA word, and the sets A_i to the fragments. This involves sequencing and comparing the fragments, which has its own difficulties. However, *lengths* of fragments are not so hard to determine, so another possibility is to digest with the first enzyme and measure lengths, digest with the second and measure lengths, and finally digest with both and measure lengths. If one does this, then the problem one obtains is essentially the double digest problem.

To completely reconstruct the DNA word one takes many copies of the word, digests with enzymes, and selects at random enough fragments that together they have a high probability of covering the word. Each of the fragments is cloned, in order to get enough mass, and then sequenced by gel chromatography. Both processes can introduce errors, so one is left with a very large number of sequenced fragments with known error rates for the letters. These need to be compared to see if they overlap: that is, to see if the sequence near the end of one fragment is the same as (or very similar to) the sequence at the beginning of another. This alignment problem is itself difficult because of the large number of possibilities involved. So, in the end we have a very large restriction mapping problem except that we can only say that given fragments overlap with probabilities that are themselves hard to estimate. A further difficulty is that DNA tends to have large blocks that repeat in different parts of the word. As a result of these complications, the problem is much harder than the restriction mapping problem described earlier. It is clear that graph theory, combinatorics, probability theory, statistics, and the design of algorithms all play central roles in sequencing a genome.

Sequence alignment is important in other problems as well. In phylogenetics (see below) one would like a way of saying how similar two genes or genomes

are. When studying proteins, one can sometimes predict protein three-dimensional structure by searching databases for known proteins with the most similar amino acid sequence. To illustrate how complex these problems are, consider a sequence $\{a_i\}_{i=1}^{1000}$ of one thousand letters from our four-letter alphabet. We wish to say how similar it is to another sequence $\{b_i\}_{i=1}^{1000}$. Naively, one could just compare a_i with b_i and define a METRIC [III.56] like $d(\{a_i\}, \{b_i\}) = \sum \delta(a_i, b_i)$. However, DNA sequences have evolved typically by insertions and deletions as well as by substitutions. Thus if the sequence ACACAC $\cdots$ lost its first C to become AACAC $\cdots$, the two sequences would be very far apart in this metric even though they are very similar and related in a simple way. The way around this difficulty is to allow sequences to include a fifth symbol, –, which stands for the place of a deletion or a place opposite an insertion. Thus, given two sequences (of perhaps different lengths), we wish to find how they can be augmented with dashes to give the minimum possible distance between them. A little thought will convince the reader that it is not feasible to use a brute-force search for a problem like this, even for the fastest computers— there are so many potential augmentations that the search would take far too long. Serious and thoughtful algorithm development is required. Two excellent introductions to the material discussed in this section are Waterman (1995) and Pevzner (2000).

4 Correlation and Causality

The central dogma of molecular biology is DNA → RNA → proteins. That is, information is stored in DNA, it is transferred out of the nucleus by RNA, and the RNA is then used in the cell to make proteins that carry out the work of the cell through the metabolic processes discussed in section 2. Thus DNA directs the life of the cell. Like most things in biology, the true situation is much more complicated. Genes, which are segments of DNA that code for the manufacture of particular proteins, are sometimes turned on and sometimes turned off. Usually, they are partially turned on; that is, the protein they code for is manufactured at some intermediate rate. This rate is controlled by the binding (or lack of binding) of small molecules or specific proteins to the gene, or to the RNA that the gene codes for. Thus genes can produce proteins that inhibit (or excite) other genes; this called a gene network.

In a way, this was obvious all along. If cells can respond to their environments by changing what they

do, they must be able to sense the environment and signal the DNA to change the protein content of the cell. Thus, while sequencing DNA and understanding specific biochemical reactions are important first steps in understanding cells, the hard and interesting work to come is to understand *networks* of genes and biochemical reactions. It is these networks, in which proteins control genes and genes control proteins, that carry out and control specific cellular functions. The mathematics will be ordinary differential equations for chemical concentrations and variables that indicate to what extent a gene is turned on. Since transport into and out of the nucleus occurs, partial differential equations will be involved. And, finally, since some of the molecular species occur in very small numbers, concentration (molecules per unit volume) may not be a useful approximation for computations about chemical binding and dissociation: they are probabilistic events.

Two kinds of statistical data can give hints about the components of these gene networks. First, there are large numbers of population studies that correlate specific genotypes to specific phenotypes (such as height, enzyme concentration, cancer incidence). Second, tools known as *microarrays* allow us to measure the relative amounts of a large number of different messenger RNAs in a group of cells. The amount of RNA tells us how much a particular gene is turned on. Thus, microarrays allow us to find correlations that may indicate that certain genes are turned on at the same time or perhaps in a sequence. Of course, correlation is not causality and a consistent sequential relationship is not necessarily causal either (sure, football causes winter, a sociologist once said). Real biological progress requires understanding the gene networks discussed above; they are the mechanisms by which the genotypes play out in the life of the cell.

A nice discussion of the relationship between population correlations and mechanisms occurs in Nijhout (2002), from which we take the following simple example. Most phenotypic traits depend on many genes; suppose that we consider a trait that depends on only two genes. Figure 1 depicts a surface that shows how the trait in an individual depends on how much each of the genes is turned on. All three variables are scaled from 0 to 1. Suppose that we study a population whose members have a genetic makeup that puts the individuals near the point X on the graph. If we do a statistical analysis of the population, we will find that gene B is highly statistically correlated to the trait, but gene A is
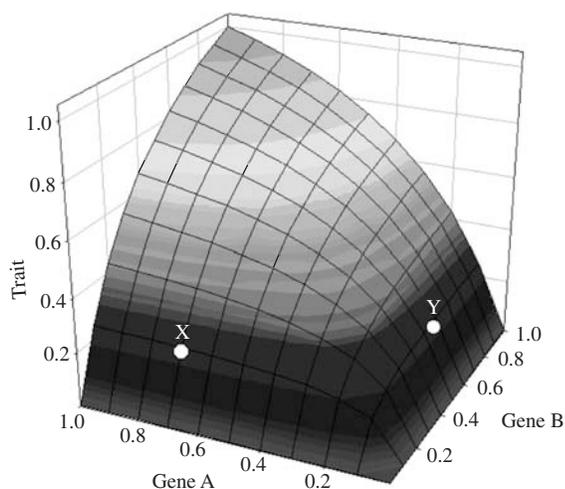
Figure 1 A phenotypic surface.

not. On the other hand, if the individuals in the population all live near the point Y on the surface, we will discover in our population study that gene A is highly statistically correlated to the trait, but gene B is not. More detailed examples with specific biochemical mechanisms are discussed in Nijhout's paper. Similar examples can be given for microarray data. This does not mean that population studies or microarray data are unimportant. Indeed, in studying hugely complex biological systems, statistical information can suggest where to look for the mechanisms that will ultimately give biological understanding.

5   The Geometry and Topology of Macromolecules

To illustrate the natural geometric and topological questions that arise when one studies macromolecules, we will briefly discuss molecular dynamics, protein–protein interactions, and the coiling of DNA. Genes code for the manufacture of proteins, which are large molecules made up of sequences of amino acids. There are twenty amino acids, each coded by a triplet of base pairs, and a typical protein might have five hundred amino acids. Interactions among the amino acids cause the protein to fold up into a complicated three-dimensional shape. This three-dimensional structure is crucial for the function of the protein since the exposed groups and the nooks and crannies in the shape govern the possible chemical interactions with small molecules and other proteins. Three-dimensional structures of

proteins can be approximately determined by X-ray crystallography and nontrivial inverse scattering calculations. The forward problem—namely, given the sequence of amino acids, predict the three-dimensional structure of the protein—is important not only for understanding existing proteins, but also for the pharmacological design of new proteins to accomplish specific tasks. Thus, in the past twenty years a large field called *molecular dynamics* has arisen, in which one uses classical mechanical methods.

Suppose we have a protein that consists of N atoms. Let x_i denote the position (specified by three real coordinates) of the ith atom, and let $\boldsymbol{x}$ denote the vector formed from all these coordinates (which belongs to $\mathbb{R}^{3N}$). For each pair of atoms, one attempts to write down a good approximation to the potential energy, $E_{i,j}(x_i, x_j)$, due to their pairwise interaction. This could be the electrostatic interaction, for example, or the van der Waals interaction, which is a classical mechanical formulation of quantum effects. The total potential energy is $E(\boldsymbol{x}) \equiv \sum E_{i,j}(x_i, x_j)$ and Newton's equations of motion take the form

$$\dot{\boldsymbol{v}} = -\nabla E(\boldsymbol{x}), \quad \dot{\boldsymbol{x}} = \boldsymbol{v},$$

where $\boldsymbol{v}$ is the vector of velocities. Starting with some initial conditions one can try to solve these equations to follow the dynamics of the molecule. Note that this is a very high-dimensional problem. A typical amino acid has twenty atoms, so that is sixty coordinates right there, and if we are looking at a protein made up of five hundred amino acids, then $\boldsymbol{x}$ will be a vector with thirty thousand coordinates. Alternatively, one could assume that the protein will fold to the configuration that has the minimum potential energy. Finding this configuration would mean finding the roots of $\nabla E(\boldsymbol{x})$, by NEWTON'S METHOD [II.4 §2.3] say, and then checking to see which root gives the lowest energy. Again this is an enormous computational task.

It is not surprising that molecular dynamics calculations have been only moderately successful and have predicted the shapes of only relatively small molecules and proteins. The numerical problems are substantial and the choice of energy terms is somewhat speculative. Even more importantly, context matters, as it does in many biological problems. The way proteins fold depends on properties of the solution in which they sit. Many proteins have several preferred configurations and switch from one to the other depending on interactions with small molecules or other proteins. Finally, it has recently been discovered that proteins do

not fold up by themselves from their linear configuration to their three-dimensional shape, but are helped and guided by other proteins called chaperones. It is natural to ask whether there are quantifiable geometrical units larger than points (atoms) that could reasonably form the basis for a good approximation to the dynamics of large molecules.

A start has been made in this direction by research groups studying the interactions of proteins with small molecules and other proteins. These interactions are fundamental to cell biochemistry, cell-transport processes, and cell signaling, and so progress is vital to understanding how cells work. Suppose one has two large proteins that are bound to each other. The first thing one would like to do is describe the geometry of the binding region. One could do this as follows. Consider an atom in either protein that is at point x. Given another atom at point y, there is a plane that divides $\mathbb{R}^3$ into two open half-spaces: the points closer to x and the points closer to y. Now let R_x denote the intersection of all such open half-spaces as y ranges over the positions of all other atoms: that is, R_x consists of those points that are closer to x than to any other atom. The union of the boundaries, $\bigcup_x \partial(R_x)$, called a *Voronoi surface*, consists of triangles and pieces of planes and has the property that each point on the surface is equidistant from at least two atom positions. To model the binding region between the two proteins, we discard all pieces of the Voronoi surface that are equidistant from two atoms that belong to the same protein and keep just the ones that are equidistant from two atoms that are in different proteins. This surface goes off to infinity, so we clip off the parts that are not "close" to either protein. The result is a surface with a boundary made up of polyhedral faces that is a reasonable approximation of the interaction interface between the two proteins. (This is not quite an accurate description: in the actual construction, "distance" is weighted in a way that depends on the atoms involved.) Now choose colors representing the twenty amino acids and color each side of each polyhedral piece with the color of the amino acid that the closest atom is in. This divides each side of the surface into large colored patches corresponding to nearness of a particular amino acid on that side. The coloring of the two sides of the boundary surface will be different, of course, and the placement of the patches gives information about which amino acids in one protein are interacting with which amino acids in the other. In particular, one amino acid in one protein may interact with

several in the other. This gives a way of using geometry to classify the nature of the particular protein–protein interaction.

Finally, let us touch on questions involving the packaging of DNA. The basic problem is easy to see. As mentioned earlier, the human DNA double helix when stretched out linearly is about two meters long. A typical cell has a diameter of about one-hundredth of a millimeter and its nucleus has a diameter of about one-third that size. All of that DNA has to be packed into the nucleus. How is this done?

At least the first stages are well understood. The DNA double helix is wound around proteins called *histones*, which consist of about two hundred base pairs each, yielding chromatin, which is a sequence of such DNA-wrapped histones connected by short segments of DNA. Then the chromatin is itself wrapped up and compacted; the geometrical details are not completely understood. It is important to understand the packing and the mechanisms that create it, because the life of the cell requires unpacking! When the cell divides, the entire DNA helix must be unzipped to form two separate strands, which are the templates on which the two new copies of DNA will be built. Clearly this cannot be done all at once but must involve local unwinding of the DNA off the histones, local unzipping, synthesis, and then local repacking.

It is equally challenging to understand the sequence of events that occurs when a protein is synthesized from a gene. Transcription factors diffuse into the nucleus and bind to specific short segments of DNA (of about ten base pairs) in the regulatory region of the gene. Of course, they will randomly bind wherever they see the same segment. Typically, one needs the binding of several different transcription factors in the regulatory region along with RNA polymerase to start transcription of a gene. That process involves the unwinding of the gene-coding region from the histones so that it can be transcribed, the transport of the resulting RNA out of the nucleus, and the recompactification of the DNA. To understand these processes fully, one will have to solve problems in partial differential equations, geometry, combinatorics, probability theory, and topology. DeWitt Sumners is the mathematician who brought the topological problems in the study of DNA (links, twists, knots, supercoiling) to the attention of the mathematics community. A good reference for molecular dynamics and the general mathematical issues posed by biological macromolecules is Schlick (2002).

6   Physiology

When one first studies human physiological systems, they seem almost miraculous. They accomplish enormous numbers of tasks simultaneously. They are robust but capable of quick changes if the situation warrants. They are made up of large numbers of cells that actively cooperate so that the tasks of the whole can be done. It is the nature of many of these systems that they are complex, controlled by feedback, and integrated with each other. It is the job of mathematical physiology to understand how they work. We will illustrate some of these points by discussing problems in biological fluid dynamics.

The heart pumps blood throughout a circulatory system that consists of vessels of diameter as large as 2.5 cm (the aorta) and as small as 6×10^{-4} cm (the capillaries). Not only are the vessels flexible, but many are surrounded by muscle and can contract to exert local force on the blood. The main force-generating mechanism (the heart!) is approximately periodic, but the period can change. The blood itself is a very complicated fluid. About 40% of its volume is made up of cells: red blood cells carry most of the oxygen and CO_2; white blood cells are immune system cells that hunt bacteria; and platelets are part of the blood clotting process. Some of these cells have diameters that are larger than the smallest capillaries, which raises the nice question of how they get through. You notice that we are very far away from most of the simplifying assumptions of classical fluid dynamics.

Here is an example of a circulatory-system question. In a significant number of people, the mitral valve, which is the inflow valve to the left side of the heart, becomes defective. It is common to replace the valve by an artificial one and this leads to an important question: how should one design the artificial valve so that the resulting flow in the left heart chamber has as few stagnant points as possible, since clots tend to form at these points? Charles Peskin did the pioneering work on this problem. Here is another question. The white blood cells are not carried in the middle of the fluid but tend to roll along the walls. Why do they do that? It is a good thing that they do, because their job is to sniff out inflammation outside the blood vessel and, when they find it, to stop and burrow through the blood vessel wall to get to the inflamed site. Another circulatory fluid dynamics question is discussed in section 10.

The circulatory system is connected to many other systems. The heart has its own pacemaker cells, but its frequency of contraction is regulated by the autonomic nervous system. Through the *baroreceptor reflex*, the sympathetic nervous system tightens blood vessels to avoid a dramatic drop in blood pressure when we stand. Overall average blood pressure is maintained by a complicated regulatory feedback mechanism involving the kidneys. It is worthwhile remembering that all these things are being accomplished by living tissues whose parts are always decaying and being replaced. For example, the gap junctions that transmit current at very low resistance between heart muscle cells have a half-life of approximately one day.

As a final example, we consider the lung, which has a fractal branching structure that terminates after twenty-three levels in about 600 million air sacs called *alveoli*, in which oxygen and CO_2 are exchanged with the circulating blood. The Reynolds number of the air flow varies by about three orders of magnitude between the large vessels near the throat and the tiny vessels near the alveoli. Premature infants often have respiratory difficulty because they lack surfactants that reduce surface tension on the inner surfaces of the alveoli. The high surface tension makes the alveoli collapse, which makes breathing difficult. One would like the infants to breathe in air that includes tiny aerosol drops of surfactant. How small should the drops be so that as much surfactant as possible makes it to the alveoli?

The mathematics of physiology consists mostly of ordinary and partial differential equations. However, there is a new feature: many of these equations have time delays. For example, the rate of respiration is controlled by a brain center that senses the CO_2 content of blood. It takes almost fifteen seconds for blood to go from the lungs to the left heart and from there to the brain center. This time delay is even longer in patients with weak hearts and often these patients display Cheyne–Stokes breathing: very rapid breathing alternates with periods of little or no breathing. Such oscillations in control systems are well-known as the time delay gets longer. Since partial differential equations are often involved, new mathematical results are needed that go well beyond the standard theory of ordinary differential equations with delay, which was initiated by Bellman in the 1950s. An excellent reference for the applications of mathematics to physiology is Keener and Sneyd (1998).

7   What's Wrong with Neurobiology?

The short answer is that there is not enough theory. This may seem an odd thing to say, since neurobiology

is the home of the Hodgkin–Huxley equations, which are often cited as a triumph of mathematics in biology. In a series of papers in the early 1950s, Hodgkin and Huxley described several experiments, and gave a theoretical basis for explaining them. Building on the work of physicists and chemists (for example, Walter Nernst, Max Planck, and Kenneth Cole), they discovered the relationship between certain ionic conductances and the trans-membrane electrical potential, $v(x,t)$, in the axons of neurons, and they formulated a mathematical model:

$$\frac{\partial v}{\partial t} = \alpha \frac{\partial^2 v}{\partial x^2} + g(v, y_1, y_2, y_3),$$
$$\frac{\partial y_i}{\partial t} = f_i(v, y_i), \quad i = 1, 2, 3.$$

Here the y_i are related to the membrane conductances of various ions. The equations have solutions that are pulses that keep their shape and travel at constant velocity in a way that corresponds to the observed behavior of action potentials in real neurons. The ideas, both explicit and implicit, in these discoveries form the basis of much single-neuron physiology. Of course, mathematicians should not be too proud about this since Hodgkin and Huxley were biologists. The Hodgkin–Huxley equations were part of the stimulus for interesting work by mathematicians on traveling waves and pattern formation in reaction–diffusion equations.

However, not everything can be explained at the level of just one neuron. Watch your hand as it reaches out gracefully to pick up an object. Think about the so-called ocular–vestibular reflex in which motions of the head are automatically compensated for by motions of the eyes so that your gaze can remain fixed. Consider the fact that you are looking at stereotypical black marks on a page and they mean something inside your head. These are *system properties*, and the systems are large indeed. There are approximately 10^{11} neurons in the central nervous system and on average each makes about one thousand connections to other neurons. These systems will not be understood just by examining their parts (the neurons) and, for obvious reasons, experimentation is limited. Thus, experimental neurobiology, like experimental physics, needs input from deep and imaginative theorists.

The lack of a large theory community interacting robustly with experimentalists is to some extent a historical accident. Grossberg asked how groups of (quite simple) model neurons, if they were connected in the right ways, could accomplish various tasks such as pattern recognition and decision making, or could exhibit certain "psychological" properties (Grossberg 1982). He also asked how these networks could be trained. At about the same time it was shown that networks of neuron-like elements connected in the right way could automatically compute good solutions of large, difficult problems like the TRAVELING SALESMAN PROBLEM [VII.5 §2]. These and other factors, including the great interest in software engineering and artificial intelligence, led to the emergence of a large community of researchers studying "neural networks." The members of this community were mostly computer scientists and physicists, so it was natural for them to concentrate on the design of devices, rather than biology. This was noticed, of course, by experimental neurobiologists, who lost interest in collaborating with these theorists.

This brief history is of course an oversimplification. There are mathematicians (and physicists and computer scientists) who are essentially theoreticians for neuroscience. Some of them work on hypothetical networks, typically either very small networks or networks with strong homogeneity properties, to discover what are the emergent behaviors of the systems. Others work on modeling real physiological neural networks, often collaboratively with biologists. Usually, the models consist of ordinary differential equations for the firing rates of the individual neurons or mean-field models that involve integral equations. These mathematicians have made real contributions to neurobiology.

But much more is needed, and to see why, it is useful to think about just how difficult these problems really are. First, there is no one-to-one correspondence between the cells of the central nervous system in different members of the same species (except in special cases like *C. elegans*). Second, neurons in the same animal differ widely in their anatomy and physiology. Third, the details of a particular network may well depend on the life history of the animal. Fourth, most neurons are somewhat unreliable devices in that they give different outputs under repeated trials with the same input. Finally, one of the prime characteristics of neural systems is that they are plastic, adaptable, and ever changing. After all, if you remember anything of what is written here, then your head is different from when you began. Between the level of the single neuron and the psychological level, there are probably twenty levels of networks, each network feeding into and being controlled by networks at other levels. The mathematical objects that will enable us to classify, analyze, and

understand how this all works have probably not yet been discovered.

8   Population Biology and Ecology

Let us begin with a simple example. Imagine a large orchard of equally spaced trees and suppose that one tree has a disease. The disease can be transmitted only to nearest neighbors, and is transmitted with probability p. What is $E(p)$, the expected percentage of trees that will be infected? Intuitively, if p is small, $E(p)$ should be small, and if p is large, $E(p)$ should be close to 100%. In fact, one can prove that $E(p)$ changes very rapidly from being small to being large as p passes through a small transition region around a particular critical probability p_c. One would expect p to decrease as the distance, d, between trees increases; farmers should choose d in such a way that p is less than the critical probability, in order to make $E(p)$ small. We see here a typical issue in ecological problems: how does behavior on the large scale (tree epidemic or not) depend on behavior at the small scale (the distance between trees). And, of course, the example illustrates that understanding the biological situation requires mathematics. For other examples of sharp global changes in probabilistic models, see PROBABILISTIC MODELS OF CRITICAL PHENOMENA [IV.25].

Suppose that we now widen our gaze to consider forests—let us say the forests on the East coast of the United States. We would like to understand how they have come to be as they are. Most of them were not planted in neat rows, so that is already a complication. But there are two other really new features. First, there is not one species but many, and each species of tree has different properties: shape, seed dispersal, need for light, and so forth. The species are different, but their properties affect each other because they are living in the same space. Second, the species, and the interactions between the species, are affected by the physics of the environment. There are physical parameters that vary on long timescales, like average temperature, and there are other parameters that vary on very short timescales, like wind speed (for seed dispersal). Certain properties of forests may depend on the fluctuations in these parameters as much as on the values themselves. Finally, one might have to take into account the reaction of the ecosystem to catastrophic events such as hurricanes or prolonged drought.

The difficulties are similar to those we have seen for other problems in mathematical biology. One would like to understand the emergent behavior on the large scale. To do this one creates mathematical models that relate the behavior on the small scale to the large scale. However, on the small scale one is overwhelmed by the biological details. Which of these details should be in the model? Of course, there is no simple answer to this because, in fact, this is the heart of what we want to know. Which of the bewildering variety of local properties or variables give rise to the large-scale behavior and by what mechanisms? Furthermore, it is not obvious what kinds of model are best. Should we model each individual and its interactions, or should we use population densities? Should we use deterministic models or stochastic models? These are also hard questions, and the answers depend on the system being studied and the questions being asked. A nice discussion of these different modeling choices can be found in Durrett and Levin (1994).

Let us focus again on a simple model: the so-called *SIRS model* for the spread of a disease in a population. A crucial parameter is the *infectious contact number*, σ, which represents the average number of new infections that an infected individual creates in the susceptible population. For a serious disease one would like to bring the value of σ down to below 1 (so that an epidemic will be unlikely) by vaccination, which takes individuals from the susceptible category and puts them in the removed category. Since vaccination is expensive and it is difficult to vaccinate high percentages of the population, it is an important public-health problem to know how much vaccination is needed to bring σ to below 1. A little reflection shows us how difficult this problem really is. First of all, the population is not well mixed, so one may not be able to ignore spatial separation, as is done in the SIRS model. Even more important, σ depends on the social behavior of individuals and the subclasses of the population to which they belong (as anyone with small children in school will attest). Thus, we see a genuinely new issue here: if an ecological problem involves animals, then the social behavior of the animals may affect the biology.

In fact, the issues are even deeper. Individuals in groups, or species, or subpopulations, vary and it is just this variation on which natural selection acts. So, to understand how an ecosystem got to where it is today, one may have to take this individual variability into account. Social behavior is also transmitted from generation to generation, both biologically and culturally, and therefore also evolves. For instance, there are many examples of plant and animal species

in which the biology of the plants and the sociology of the animals clearly coevolved, to the benefit of both. Game-theory models have been used to study the evolution of certain human behaviors such as altruism. Therefore, ecological problems, which sometimes seem simple at first, are often very deep, because the biology and its evolution are connected in complicated ways to both the physics of the environment and the social behavior of the animals. A good introductory review of these questions can be found in Levin et al. (1997).

9   Phylogenetics and Graph Theory

Since Darwin, a deep ongoing problem in biology has been to determine the history of the evolution of species that has brought us to our current state. It is natural when thinking about such questions to draw directed GRAPHS [III.34] in which the vertices, V, are species (past or present) and an edge from species v_1 to species v_2 indicates that v_2 evolved directly from v_1. Indeed, Darwin himself wrote down such graphs. To explain the mathematical issues, we will consider a simple special case. A connected graph with no cycles is called a *tree*. If we distinguish a particular vertex, ρ, and call it the *root*, then the tree is called *rooted*. The vertices of the tree that have degree one (i.e., have only one attached edge) are called *leaves*. We will assume that ρ is not a leaf. Notice that, because there are no cycles, there is exactly one path in the tree from ρ to each vertex v. We say that $v_1 \leqslant v_2$ if the path from ρ to v_2 contains v_1 (see figure 2). The problem is to determine which trees with a given set of leaves X (current species) and a given root vertex ρ (a hypothesized ancestral species) are consistent with experimental information and theoretical assumptions about the mechanisms of evolution. Such a tree is called a *rooted phylogenetic X-tree*. One can always add extra intermediate species, so typically one imposes the additional restriction that the phylogenetic trees be as simple as possible.

Suppose that we are interested in a certain characteristic, the number of teeth, for example. We can use it to define a function f from X, the set of current species, to the nonnegative integers: given a species x in X, we let $f(x)$ be the number of teeth of members of x. In general, a *character* is a function from X to a set C of possible values of a particular characteristic (having or not having a gene, the number of vertebrae, the presence or absence of a particular enzyme, etc.). It is characters such as these that are measured by biologists in current
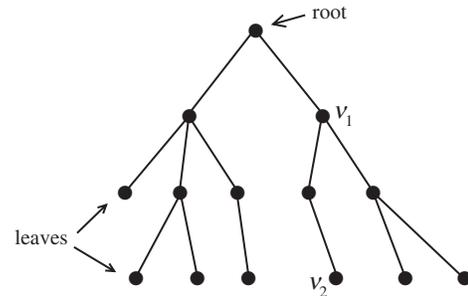

Figure 2 A rooted tree.

species. In order to say something about evolutionary history, one would like to extend the definition of f from X to the larger set V of all the vertices in a phylogenetic tree. To do this, one specifies some rules for how characters can change as species evolve. A character is called *convex* if f can be extended to a function $\bar{f}$ from V to C in such a way that for every $c \in C$, the subset $\bar{f}^{-1}(c)$ of V is a connected subgraph of the tree. That is, between any two species x and y with character value c there should be a path back in evolutionary history from x and forward again to y such that all the species in between have the same value c. This essentially forbids new values from arising and then reverting back and forbids two values evolving separately (in different parts of the tree). Of course, we have the current species and lots of characters. What is unknown is the phylogenetic tree, that is, the collection of intermediate species and the relations between them that link the current species to a common ancestor. A collection of characters is called *compatible* if there exists a phylogenetic tree on which they are all convex. Determining when this is the case and finding an algorithm for constructing such a tree (or a minimal such tree) is called the *perfect phylogeny problem*. This problem is understood for collections of characters with binary values, but not in general.

An alternative problem is the following. Note that we have been treating all the edges alike when in fact some may represent longer or shorter evolutionary steps. Suppose that we have a function w that assigns a positive number to each edge. Then, since there is a unique shortest path between any two vertices in the tree, w induces a distance function d_w on $V \times V$, and in particular on X. Now, suppose that we are given a distance function δ on $X \times X$ that tells us how far apart current species are. The question is whether there exists a phylogenetic tree and a weighting function w so that

$\delta(x, y) = d_w(x, y)$ for all $x, y \in X$. If so, one would like an algorithm to construct the tree and the weights. If not, one would like to construct a family of trees that satisfy the relation approximately.

Finally, we note that there is a blossoming field of Markov processes on trees where the partial order on V forms the basis for the Markov condition. Not only are there wonderful mathematical questions relating the geometry of the tree to the processes, but there are important issues for phylogenetics. Suppose that one starts with characters defined only at the root and then allows them to "evolve" down the tree by (possibly different) Markov processes. Then, given the distribution of characters on the leaves, when can we reconstruct the tree? These questions have even given rise to problems in algebraic geometry.

Phylogenetics is useful not only for determining our past but also for controlling our present and future: see Fitch et al. (1997), where you can find a phylogenetic reconstruction for the influenza A virus. An excellent recent graduate text in this field is Semple and Steel (2003).

10 Mathematics in Medicine

It is clear that an improved understanding of biological systems leads, at least indirectly, to improved medical care. However, there are many cases in which mathematics has a direct impact on medicine. We give two brief examples.

Charles Taylor is a biomedical engineer at Stanford who works on the fluid dynamics of the cardiovascular system. He wants to use fast simulations of flows as part of the medical decision-making process. Suppose that a patient presents with leg weakness and is found on magnetic resonance imaging (MRI) to have an arterial constriction in the thigh. Typically, the surgical group will meet and consider a variety of options including shunting blood from other vessels to a point below the constriction or shunting blood around the constriction with vessels removed from some other site in the patient's body. Among a fairly large number of possible choices, the surgical group chooses based on what they have been taught and on their own experience. The characteristics of the flow after the graft are important not just for recovery of function but to prevent the formation of possibly destructive clots. An important difficulty is that patients treated successfully are rarely seen again, so one does not know the actual characteristics of the flow after the operation. Taylor wants to be in on the discussion with the surgical team with immediate fluid dynamical simulations based on the patient's actual vasculature (as revealed by the MRI) for each proposed graft suggested. And he wants followup on each patient to check how well his simulations predicted the actual postoperative flow.

David Eddy is an applied mathematician who has worked on health policy for thirty years. He first became prominent when he published *Screening for Cancer: Theory, Analysis and Design* (Eddy 1980), which grew out of his Ph.D. thesis. Because of this book, the American Cancer Society changed its recommendation for the frequency of Pap smears from once a year to once every three years, since Eddy's modeling showed that the change would have little effect on the life expectancy of the average American woman. A short calculation easily estimates the amount of money saved in an economy that spends 15% of its gross domestic product (GDP) on health care. Throughout his career Eddy has criticized both the indiscriminate use of diagnostic tests and the incorrect use of the results by physicians and policy boards often ignorant of the basic facts of conditional probability. He has criticized specific health-policy guidelines as based on seat-of-the-pants guesswork instead of quantitative analysis. In a classic case he distributed questionnaires to physicians at a conference on colorectal cancer. The physicians were asked to estimate the percentage drop in mortality from colorectal cancers if all Americans over age fifty were to have the two most common diagnostic tests each year: fecal blood smear and flexible sigmoidoscopy. The answers were approximately uniformly distributed in a range from 2% to 95%. Even more startling was the fact that the physicians did not even know that they disagreed so dramatically. He has used mathematical models to analyze the costs and benefits of new and existing surgeries, medical treatments, and drugs, and he has participated robustly in debates on the current health-policy crisis. Throughout, he has pointed out that a hefty percentage of GDP is spent on devices, drugs, and procedures with almost no mathematical analysis of which are effective.

For more on the interrelations between mathematics and medicine, see MATHEMATICS AND MEDICAL STATISTICS [VII.11].

11 Conclusions

Mathematics and mathematicians have played important roles in many fields of biology that this brief

article has not had the space to cover: immunology, radiology, developmental biology, and the design of medical devices and synthetic biomaterials, to name just a few of the most obvious omissions. Nevertheless, this collection of examples and introductory discussions allows us to draw a few conclusions about mathematical biology. The range of biological problems needing explanation by mathematics is enormous and techniques from many different branches of mathematics are important. It is not so easy in mathematical biology to extract simple, clear mathematical questions to work on, because biological systems typically operate in a complex environment where it is difficult to decide what should be counted as the system and what as the parts. Finally, biology is a source of new, interesting, and difficult questions for mathematicians, whose participation in the biological revolution is necessary for a full understanding of the biology itself.

Further Reading

Durrett, R., and S. Levin. 1994. The importance of being discrete (and spatial). *Theoretical Population Biology* 46: 363–94.

Eddy, D. M. 1980. *Screening for Cancer: Theory, Analysis and Design.* Englewood Cliffs, NJ: Prentice-Hall.

Fall, C., E. Marland, J. Wagner, and J. Tyson. 2002. *Computational Cell Biology.* New York: Springer.

Fitch, W. M., R. M. Bush, C. A. Bender, and N. J. Cox. 1997. Long term trends in the evolution of H(3) HA1 human influenza type A. *Proceedings of the National Academy of Sciences of the United States of America* 94:7712–18.

Grossberg, S. 1982. *Studies of Mind and Brain: Neural Principles of Learning, Perception, Development, Cognition, and Motor Control.* Boston, MA: Kluwer.

Keener, J., and J. Sneyd. 1998. *Mathematical Physiology.* New York: Springer.

Levin, S., B. Grenfell, A. Hastings, and A. Perelson. 1997. Mathematical and computational challenges in population biology and ecosystems science. *Science* 275:334–43.

Nijhout, H. F. 2002. The nature of robustness in development. *Bioessays* 24(6):553–63.

Pevzner, P. A. 2000. *Computational Molecular Biology: An Algorithmic Approach.* Cambridge, MA: MIT Press.

Schlick, T. 2002. *Molecular Modeling and Simulation.* New York: Springer.

Semple, C., and M. Steel. 2003. *Phylogenetics.* Oxford: Oxford University Press.

Waterman, M. S. 1995. *Introduction to Computational Biology: Maps, Sequences, and Genomes.* London: Chapman and Hall.

VII.3 Wavelets and Applications
Ingrid Daubechies

1 Introduction

One of the best ways to understand a function is to expand it in terms of a well-chosen set of "basic" functions, of which TRIGONOMETRIC FUNCTIONS [III.92] are perhaps the best-known example. Wavelets are families of functions that are very good building blocks for a number of purposes. They emerged in the 1980s from a synthesis of older ideas in mathematics, physics, electrical engineering, and computer science, and have since found applications in a wide range of fields. The following example, concerning image compression, illustrates several important properties of wavelets.

2 Compressing an Image

Directly storing an image on a computer uses a lot of memory. Since memory is a limited resource, it is highly desirable to find more efficient ways of storing images, or rather to find *compressions* of images. One of the main ways of doing this is to express the image as a function and write that function as a linear combination of basic functions of some kind. Typically, most of the coefficients in the expansion will be small, and if the basic functions are chosen in a good way it may well be that one can change all these small coefficients to zero without changing the original function in a way that is visually detectable.

Digital images are typically given by large collections of *pixels* (short for *picture elements*; see figure 1).

The boat image in figure 1 is made up of 256×384 pixels; each pixel has one of 256 possible gray values, ranging from pitch black to pure white. (Similar ideas apply to color images, but for this exposition, it is simpler to keep track of only one color.) Writing a number between 0 and 255 requires 8 digits in binary; the resulting 8-bit requirement to register the gray level for each of the $256 \times 384 = 98\,304$ pixels thus gives a total memory requirement of $786\,432$ bits, for just this one image.

This memory requirement can be significantly reduced. In figure 2, two squares of 36×36 pixels are highlighted, in different areas of the image. As is clear from its blowup, square A has fewer distinctive characteristics than square B (a blowup of which is shown in figure 1), and should therefore be describable with fewer bits. Square B has more features, but it too contains