
AMS Short Course

Modeling and Simulation of Biological Networks

**Organized by: Reinhard Laubenbacher
Virginia Bioinformatics Institute
and Mathematics Department
Virginia Polytechnic Institute and State University
San Antonio, Texas, January 10–11, 2006**

It is planned that lecture notes will be available to those who register for this course. Advance registration fees are: member of the AMS–US\$87; nonmember–US\$115; student, unemployed, emeritus–US\$38. On-site fees are: member of the AMS–US\$118; nonmember–US\$148; student, unemployed, emeritus–US\$57. Registration and housing information can be found in this issue of the *Notices*; see the section “Registering in Advance and Hotel Accommodations” in the announcement for meetings in Atlanta. The registration form is at the back of this issue.

General Introduction

One of the major challenges for biology in the 21st century is an understanding of the organizational principles of biological systems. Mathematics, statistics, computer science, and engineering will play key roles in meeting this challenge. Recent years have seen tremendous progress in our understanding of genetic information in organisms, and its fundamental role in cell metabolism. Mathematical and statistical tools have been instrumental in this progress. Beyond an understanding of the processes at these scales, integrating them with the organism and ecosystem scales presents many mathematical challenges. Thus, biology will be one of the major drivers of the development of new mathematics.

The aim of this Short Course is to present the state of the art and the mathematical challenges at these different scales, from encoded genomic information all the way to the ecosystem level. Particular emphasis will be placed on “nontraditional” mathematical approaches. Two panel discussions will explore the role mathematics and mathematicians can play in life sciences and biomedical research.

Phylogenetics

Elizabeth S. Allman, University of Maine

Understanding evolutionary relationships between species is a fundamental issue in biology. For many years, this was accomplished through painstaking comparisons

of morphological or other features, perhaps with reference to the fossil record. As biological sequence data has become readily available in recent years, however, this new data source has allowed increasingly mathematical approaches to inferring *phylogenetic trees*.

This talk will begin with a survey of the many ideas that have been used to construct phylogenetic trees from sequence data. Approaches range from the primarily combinatorial, to probabilistic model-based methods appropriate for developing statistical viewpoints. Strengths and weaknesses of the various mathematical approaches currently employed, as well as points where further development is needed, will be discussed.

The final part of the talk will discuss a thread of research in which algebraic methods have been adopted to understand some of the probabilistic models used in phylogenetics. Recent progress on understanding the set of possible probability distributions arising from a model as an algebraic variety has helped provide new theoretical results, and may point toward improved approaches to phylogenetic inference.

References

- [1] E. ALLMAN and J. RHODES, Phylogenetic ideals and varieties for the general Markov model, preprint, <http://arXiv:math.AG/0410604>.
- [2] J. FELSENSTEIN, *Inferring Phylogenies*, Sinauer Associates, Sunderland, MA, 2004.
- [3] O. GASCUEL, ed., *Mathematics of Evolution and Phylogeny*, Oxford University Press, 2005.
- [4] L. PACTER and B. STURMFELS, eds., *Algebraic Statistics for Computational Biology*, Cambridge University Press, 2005, to appear.
- [5] C. SEMPLE and M. STEEL, *Phylogenetics*, vol. 24 of *Oxford Lecture Series in Mathematics and Its Applications*, Oxford University Press, 2003.

Optimal Control of Population and Disease Models

Suzanne Lenhart, University of Tennessee and Oak Ridge National Laboratory

Optimal control theory has been used successfully to make management or strategy decisions involving biological or medical models. The desired goal of the control actions depends on the particular scenario. Making reasonable choices requires understanding and quantifying the trade-offs between competing goals and features, including their future relevant consequences.

This talk gives an introduction to optimal control theory for ordinary differential equations. The applications emphasize biological models, especially immunology disease models. The background of an iterative numerical method used to solve the optimality systems of these problems will be presented.

References

- [1] S. P. SETHI and G. L. THOMPSON, *Optimal Control Theory: Applications to Management Science and Economics*, Kluwer, 2nd Edition, 2000.
- [2] D. E. KIRSCHNER, S. LENHART, and S. SERBIN, Optimal control of the chemotherapy of HIV, *J. Math Biology* 35 (1997), 77–96.
- [3] <http://www.math.utk.edu/~lenhart/smb2003.v2.html> (webpage for NIH sponsored short course on this topic).

Interaction-Based Computing Approach to Modeling and Simulations of Large Biological and Socio-Technical Systems

Madhav Marathe, Virginia Bioinformatics Institute and Department of Computer Science, Virginia Polytechnic Institute and State University

The lecture will describe an interaction-based modeling and simulation approach for understanding large biological, information, social, and technological (BIST) systems.

Such systems consist of large numbers of interacting components that together produce a global system with properties that are the result of interactions among the representations of the local system elements. Examples of biological systems that are suitable for representation and analysis using the above paradigm include: bio-signaling systems, ecologies, gene regulatory networks, the public health system, and contagious disease economics.

The interaction-based computer simulations are based on a mathematical and computational theory called Sequential Dynamical Systems (SDS). SDS theory forms the formal basis for describing complex simulations, by composing simpler ones. SDS is a new class of discrete, finite dynamical systems. SDS-based “formal simulations” potentially provide a rigorous, useful new setting for a theory of interaction-based computation. The setting is natural for comprehension of distributed systems characterized by interdependent, but separately functioning sub-parts.

I will focus on the mathematical and computational aspects of SDS. Applicability of these concepts will be described in the context of large-scale biological and epidemiological simulations being developed by our group at the Virginia Bioinformatics Institute.

Modeling and Simulation of Biochemical Networks

Pedro Mendes, Virginia Bioinformatics Institute

Cellular behavior and function passes through an intricate network of enzyme-catalyzed reactions. These networks and the phenomena that depend on them can be modeled through several mathematical frameworks. The best-established one uses the machinery of calculus, especially differential equations. A review of this framework for modeling biochemical networks will be made, starting

with its origins in physical chemistry and enzymology. Particular emphasis will be placed on the biochemical concepts and their mathematical counterparts. Illustrations of the concepts discussed will be made using biochemical modeling software.

Reconstructing Ancestral Genomes

Lior Pachter, Department of Mathematics, University of California, Berkeley

Recent advances in high-throughput genomics technologies have resulted in the sequencing of large numbers of (near) complete genomes. These genome sequences are being mined for important functional elements, such as genes, however and they are also being compared and contrasted in order to identify sequences that have been conserved over time. Such sequences frequently point to biologically important elements. In cases where DNA sequences from different organisms can be determined to have originated from a common ancestor, it is natural to try to infer the ancestral sequences. The reconstruction of the ancestral genome can lead to insights about genome evolution, and the origin and diversity of functional elements. There are a number of interesting mathematical questions associated with reconstructing ancestral genomes. What are the appropriate statistical models for evolution that allow for making inferences about ancestral sequences? How many present day genomes are necessary to reconstruct an ancient genome? How can insertions and deletions be accounted for? We will review progress on these questions, and highlight interesting mathematical results and problems.

References

- [1] M. BLANCHETTE, E. D. GREEN, W. MILLER, and D. HAUSSLER, Reconstructing large regions of an ancestral mammalian genome in silico, *Genome Research*, 14:2412–2423, 2004.
- [2] L. PACTHER and B. STURMEELS, *Algebraic Statistics for Computational Biology*, Cambridge University Press, 2005.
- [3] _____, The mathematics of phylogenomics, <http://arxiv.org/abs/math.ST/0409132>, 2005.
- [4] L. PACTHER and S. SNIR, Phylogenetic profiling of insertions and deletions in vertebrate genomes, 2005.

A Computational Algebra Approach to Systems Biology

Brandilyn Stigler, Mathematical Biosciences Institute

Synopsis: One goal of systems biology is to predict and modify the behavior of biological networks by accurately monitoring and modeling their responses to certain types of perturbations. The construction of mathematical models based on observation of these responses, referred to as reverse engineering, is an important step in elucidating the structure and dynamics of such networks. Continuous models, described by systems of differential equations, are an example of a framework that has been used to reverse engineer biochemical networks. Of increasing interest is the use of discrete models, given by systems of polynomials defined over finite fields, which may provide a qualitative description of the network.

About the Cover

The Riverwalk, San Antonio, Texas. San Antonio is the site of the 2006 Joint Mathematics Meetings, January 12–15, 2006. Photograph by Robert Herzik, courtesy of the San Antonio Convention and Visitors Bureau.



Conferences

In this talk I will provide an overview of existing reverse-engineering methods, as well as discuss some modeling issues that arise from biological systems. A discrete modeling approach, rooted in computational algebra, to reverse engineer networks from experimental time series data will also be introduced. The discrete method uses algorithmic tools, including Gröbner-basis techniques, to build the set of all polynomial models that fit time series data and to select minimal models from this set. The effectiveness of the algorithm will be demonstrated on simulated networks. As it is important to identify the type of data set that are best suited to build accurate models, properties of data that make them suitable for the algebraic method will also be discussed.

Reading List

- [1] P. D'HAESELEER, S. LIANG, and R. SOMOGYI, Genetic network inference: From co-expression clustering to reverse engineering, *Bioinformatics* **16** (2000) 707–726.
- [2] H. DE JONG, Modeling and simulation of genetic regulatory systems: A literature review, *J. Computational Biology* **9** (2002) 67–103.
- [3] H. KITANO, Systems biology: A brief overview, *Science* **295** (2002), 1662–1664.
- [4] R. LAUBENBACHER and B. STIGLER, A computational algebra approach to the reverse engineering of gene regulatory networks, *J. Theoretical Biology* **229** (2004) 523–537.
- [5] R. MAY, Uses and abuses of mathematics in biology, *Science* **303** (2004), 790–793.