Research Statement — David F. Anderson

Summary of research interests. My main mathematical interests are nonlinear dynamical systems, probability theory, stochastic processes, scientific computing, and numerical analysis. My main scientific interests lie in applying mathematical methods to models that arise in biology, especially cell biology and biochemistry. My professional goal is to continue to develop and apply mathematical tools to problems arising in biology with an emphasis on collaborative and interdisciplinary research.

Most of my research has involved the analysis of deterministically or stochastically modeled biochemical reaction systems. Typically, if the abundances of the constituent molecules of a biochemical system are low, then the randomness that arises from the making and breaking of chemical bonds is important to the system dynamics, and the abundances are modeled stochastically as a continuous time Markov chain. However, if the abundances are so high that such randomness is averaged out, then the chemical concentrations are modeled with deterministic dynamics. In my thesis work [1] and the publications [2, 3] I developed *fluctuation theory*, which uses stochastic forcing to gain insight into the emergent properties of deterministically modeled systems. In [4] we applied the principles of fluctuation theory to a model of the folate and methionine cycles that allowed a better understanding of the mechanisms that stabilize the DNA-methylation rate. In [5, 6, 7] and [8] we developed new mathematical tools for understanding the *qualitative properties* of deterministically and stochastically modeled biochemical systems, respectively, without the need for detailed information about the system parameter values, which are oftentimes unknown.

Over the previous decade biologists and biochemists have become increasingly aware of the importance of modeling processes with low chemical abundances stochastically. These scientists have found Monte Carlo simulation to be an effective tool. Therefore, in [9, 10, 11] we developed and analyzed simulation methods for the numerical computation of sample paths of stochastically modeled biochemical systems that achieve significant gains in both efficiency and stability over existing methods.

In other recent work [12] we developed a weakly second order method for the simulation of sample paths of stochastic differential equations driven by Brownian motion. This method is applicable to a wide class of systems that naturally incorporates many biochemical systems.

In the next section I will explain my past and current projects in more detail, and in the last section I will highlight some directions for future research.

Past and Current Research Projects

1. Fluctuation Theory. Stochasticity is a fundamental part of biology. From the randomness in the making and breaking of chemical bonds to interactions with other (stochastic) biochemical processes to changes in the extra-cellular environment, stochasticity affects every level of biochemical systems. Understanding how these systems have evolved to either use this stochasticity for the benefit of the organism (such as through increased phenotypic diversity) or to mitigate its deleterious effects is an important, and difficult, question. Because it is the complexity of biochemical networks that leads to stable biological function in a stochastic environment, insight is gained by first translating that complexity into the language of mathematics and then developing the correct analytical tools.

My thesis work [1] consisted of developing fluctuation theory for deterministically modeled biochemical systems. In fluctuation theory one wants to investigate the response to external stochasticity of a large biochemical system or network motif (a common network structure found in biological systems), and in particular to study the emergent properties of the system as the random fluctuations propagate through the network. Here the randomness is a tool used to study the out-of-equilibrium dynamics of the biochemical system, which, because of the stochastic environment of the cell, is oftentimes the common state for such systems. Typically, we are interested in perturbing randomly a single or small number of inputs

or parameters. Hence the object of study is a set of differential equations that is forced by a stochastic process in one or a small number of components. The goal of this analysis is to understand how the stochastic fluctuations spread through the different system components and how the spreading depends on the *topology* of the biochemical network.

The differential equations governing the dynamics of deterministically modeled systems are nearly always nonlinear. However, the reaction network uniquely (up to parameter values) identifies the system of equations. Thus, understanding the effects that different network structures have on noise propagation has led to robust theorems pertaining to large classes of dynamical systems. In [2] we proved general theorems relating network structure to noise propagation under the simplifying assumption that the corresponding unperturbed system is linear. In [3] we extended some of these results to the nonlinear case. In [4] we applied the principles of fluctuation theory to a model of the folate and methionine cycles. Through *in silico* experimentation (e.g. removing particular reactions), we took the system apart piece by piece to discover the regulatory mechanisms that stabilize the DNA-methylation rate, which is a critical factor in the control of gene expression.

2. Qualitative properties of biochemical systems. There are potentially millions of biochemical reaction systems that may warrant study at one time or another. These systems are typically complex and highly nonlinear. Further, there is oftentimes little knowledge about exact values of the different system parameters, and, worse still, these system parameter values may vary from cell to cell. However, as mentioned in the previous section, the network structure of a given system induces the corresponding equations governing its dynamics. This fact may be exploited to infer qualitative properties of large classes of biochemical systems and, most importantly, to learn *which properties are independent of the details of the system parameters.*

In [5] I proved that for a large class of deterministically modeled systems (those with a network that is weakly reversible and has a deficiency of zero) there exists a natural, and easily checked, condition that guarantees that trajectories are *persistent.* In other words, letting $\phi(t,\xi)$ be a trajectory with initial condition $\xi \in \mathbb{R}^N_{>0}$, the set of ω-limit points, $\omega(\xi) = \{x \in \mathbb{R}^N |\ \phi(t_n,\xi) \to x \text{ as } n \to \infty, \text{ for some } t_n \to \infty\}$, does not intersect the boundary of the positive orthant. The condition shown in [5] implies that no ω-limit point can exist on a zero-dimensional face of an invariant polytope. However, I also showed in [5] how to characterize the faces of the invariant polytopes into those that can and cannot admit ω-limit points. Therefore, the condition of [5] guarantees persistence for *all* systems for which ω-limit points can only be found on zero-dimensional faces. Further, this condition guarantees persistence for *any choice of rate constants* (key system parameter values). It had been known since the 1970s that the systems considered in [5] admitted equilibria that were locally asymptotically stable, and that global asymptotic stability would hold so long as trajectories did not converge to the boundary of the positive orthant. Therefore, the results of [5] prove that these equilibria are, in fact, globally asymptotically stable. In [6] we proved results that are complementary to those of [5] by showing that no trajectory of a weakly reversible system can have an ω-limit point on a facet of an invariant polytope. As in [5], this result gives easily checked conditions that guarantee global asymptotic stability of the equilibria for a wide range of systems.

In [7] we showed how the qualitative dynamical properties of a deterministically modeled system can be understood by considering the dynamical properties of certain natural sub-systems. By explicitly characterizing the different subsets of the chemical reactants that, in some sense, determine whether or not trajectories are persistent, we proved that the original system is persistent so long as the sub-dynamical systems associated with these subsets satisfy certain network or dynamical properties. The results presented in [7] represent a large generalization over existing results that require very strong assumptions on the associated sub-networks.

In [8] we considered stochastically modeled chemical reaction systems and proved that a *product-form stationary distribution* exists for each closed, irreducible subset of the state space if the corresponding

deterministically modeled system, with mass-action kinetics (rate functions), admits a specific type of equilibrium value, called a "complex balanced" equilibrium. Further, we characterized the stationary distribution in terms of the specific stochastic kinetics chosen. Feinberg's deficiency zero theorem, which says a complex balanced equilibrium exists for a deterministically modeled system as long as the network is weakly reversible and has a deficiency of zero, then implies that such a distribution exists, *regardless of the specific system parameter values or choice of rate functions* (such as stochastic mass-action, Michaelis-Menten, or Hill kinetics), for the stochastically modeled system so long as the network satisfies those same conditions. Therefore, in [8] we have characterized the long-term dynamics, via the stationary distribution, for a large and natural class of stochastically modeled biochemical reaction systems and have done so without the need of a detailed description of the system parameter values or even the specific form of the stochastic rate functions (intensities) governing the dynamics.

3. Simulation methods for stochastically modeled systems. Many cellular processes, including all processes that include gene transcription and translation, involve constituent molecules that are present in low abundances. For example, active genes are sometimes present in zero to one copy, mRNA molecules can be equally rare, and many proteins are present in less than 100 molecules per bacterial cell. As biologists and biochemists have become increasingly aware of the importance of modeling such processes stochastically, the need for robust and efficient simulation methods for such models (in order to perform numerical experimentation using Monte Carlo methods) has increased accordingly.

In [9] I developed a robust algorithm for the numerical computation of statistically exact sample paths of stochastically modeled biochemical systems. The algorithm relies on the fact that such systems can be represented via a random time change of independent, unit-rate Poisson processes. Explicitly, if $X(t) \in \mathbb{Z}_{\geq 0}^N$ is the state of the system at time t, $\nu_k', \nu_k \in \mathbb{Z}^N$ are vectors representing the number of molecules gained and lost, respectively, in one instance of the kth reaction, $\lambda_k : \mathbb{Z}_{\geq 0}^N \to \mathbb{R}_{\geq 0}$ is the intensity (or propensity in the biology and biochemistry literature) function for the kth reaction, and Y_k are independent unit-rate Poisson processes, then

$$X(t) = X(0) + \sum_k Y_k \left(\int_0^t \lambda_k(X(s))ds \right) (\nu_k' - \nu_k). \tag{1}$$

By tracking the integrals in (1), the algorithm developed in [9] translates the jump times of the Poisson processes into the reaction times of the biochemical system. This algorithm is then easily adaptable to such typical situations as time varying (even stochastically driven) rate constants, delays in the initiation and completion of reactions (such as is common with gene transcription and/or translation models), and non-exponential waiting times between reaction events. Existing algorithms, such as Gillespie's algorithm or the next reaction method of Gibson and Bruck, are not always amenable to such situations.

In [10] I developed an algorithm for the numerical computation of approximate sample paths of stochastically modeled biochemical systems that is significantly more stable than existing algorithms. Further, I clarified the relation of this class of methods (called "tau-leaping" algorithms in the biology and biochemistry communities) to simulation methods of deterministic systems. I showed that standard tau-leaping can be viewed as an Euler approximation to the integrals $T_k(t) = \int_0^t \lambda_k(X(s))ds$ in equation (1) and then demonstrated that by tracking and storing the values $T_k(t)$ and $Y_k(T_k(t))$ one can reject steps (or "leaps") without biasing the statistics of the sample paths generated. Being able to reject leaps allows for an algorithm that chooses step sizes adaptively (based upon whether the last step was rejected or accepted) and guarantees stability. Such stability is in contrast to the previously devised adaptive tau-leaping methods that could not reject a leap without biasing the sample paths generated and for which there was no guarantee that a leap would not take the state into a non-physical domain (such as negative abundances).

In [11] we performed the first rigorous *error analysis* on tau-leaping methods using an appropriate scaling. Previous error analyses of such methods proved different strong and weak convergence rates as

the step size, τ, converged to zero. However, as the step size converges to zero, the expected number of reactions that take place per step goes to zero. This is not the domain in which tau-leaping methods are useful. Therefore, such a scaling is an inappropriate choice for error analysis. However, tau-leaping methods are typically used when the abundances of the constituent molecules are large. Letting V be the order of magnitude of a typical abundance, we showed how to carry out both strong and weak error analyses on different tau-leaping methods in a systematic manner as $V \to \infty$, $\tau \to 0$, and the expected number of reactions per step remains large, which is the domain in which tau-leaping methods are useful. Using this scaling we demonstrated that a midpoint type tau-leaping method does achieve a better order of convergence (both strong and weak) than standard tau-leaping, a result that is in contrast to error analyses using only the $\tau \to 0$ scaling.

Future Research Projects

1. Fluctuation theory and stochastically modeled systems. Due to the small abundances of some cellular molecules (e.g. active genes, mRNAs, certain proteins), many of the most important biochemical processes, including many gene regulatory networks, are now modeled, or have sub-systems that are modeled, stochastically. It is theorized that the complexity of the network structure of such systems plays an important role in ensuring robust functionality. Therefore, extending the principles of fluctuation theory to these systems is a future goal that will meet a pressing problem in biology.

Many stochastically modeled systems (including some gene regulatory networks) have *multiple scales* that arise when some of the chemical components are present in much higher abundances than others, and/or when the reaction rate constants vary over several orders of magnitude. Approximation and multi-scale methods using the random time-change representation (1) have been developed for such systems that break the system into analytically tractable pieces. Explicitly, by appropriately scaling each abundance and rate constant so that each new scaled abundance and intensity function is of order 1, averaging techniques and limit theorems may be used, as the scale of the system increases, to break the system into tractable pieces. These methods can show the mechanisms causing the emergent behavior of a system. I will both utilize such methods and develop new tools in my extension of fluctuation theory to stochastically modeled systems.

2. Qualitative properties of dynamical systems. There is still significant progress to be made in understanding (*i*) how the qualitative properties of a dynamical system are determined by its network structure, and (*ii*) which of those properties depend, and which do not, on a detailed knowledge of system parameter values. For example, it is still unknown in the most general case if the equilibria guaranteed to exist (regardless of the choice of rate constants) in Feinberg's deficiency zero theorem are globally asymptotically stable.

There is an especially large gap in such results related to stochastically modeled biochemical systems. Our result in [8] is, to the best of my knowledge, the first of its kind relating the qualitative, long-term behavior of a stochastically modeled system to its network structure alone. Extending known results relating qualitative behavior to network structure in the deterministic setting to the stochastic setting, or deriving completely new results in the stochastic setting, seems very promising and relevant.

3. Simulation methods, numerical analysis, and software development. Open source software for the simulation method I developed in [10] is in development. A future project will involve developing software for the efficient implementation of the exact method developed in [9].

The methods for my future research described in the subsection "Fluctuation theory and stochastically modeled systems" include approximation methods for multi-scale systems. These methods lead naturally to systems that incorporate both discrete-stochastic and absolutely continuous (with respect to time) components. Therefore, hybrid numerical methods will need to be developed and analyzed.

References

[1] David F. Anderson, *Stochastic Perturbations of Biochemical Reaction Systems*, Duke University PhD Thesis, 2005.

[2] David F. Anderson, Jonathan C. Mattingly, H. Frederik Nijhout, Michael C. Reed, *Propagation of Fluctuations in Biochemical Systems, I: Linear SSC Networks*, Bulletin of Mathematical Biology, **69**(6), 1791 - 1813, 2007.

[3] David F. Anderson, Jonathan C. Mattingly, *Propagation of Fluctuations in Biochemical Systems, II: Nonlinear chains*, IET Systems Biology, **1**(6), 313 - 325, 2007.

[4] H. Frederik Nijhout, Michael C. Reed, David F. Anderson, Jonathan C. Mattingly, S. Jill James, Cornelia M. Ulrich, *Long-Range Allosteric Interactions between the Folate and Methionine Cycles Stabilize DNA Methylation Reaction Rate*, Epigenetics, **1**(2), 81 - 87, 2006.

[5] David F. Anderson, *Global asymptotic stability for a class of nonlinear chemical equations*, SIAM Journal of Applied Mathematics, **68**(5), 1464 - 1476, 2008.

[6] David F. Anderson, Anne Shiu, *Persistence of deterministic population processes and the Global Attractor Conjecture*, in preparation (to be submitted by January 2009).

[7] David F. Anderson, Gheorghe Craciun, *Reduced reaction networks and persistence of chemical systems*, in preparation (to be submitted by January 2009).

[8] David F. Anderson, Gheorghe Craciun, Thomas G. Kurtz, *Product-form stationary distributions for deficiency zero chemical reaction networks*, submitted, arXiv:0803.3042.

[9] David F. Anderson, *A modified Next Reaction Method for simulating chemical systems with time dependent propensities and delays*, Journal of Chemical Physics, **127**(21), 214107, 2007.

[10] David F. Anderson, *Incorporating postleap checks in tau-leaping*, Journal of Chemical Physics, **128**(5), 054103, 2008.

[11] David F. Anderson, Arnab Ganguly, Thomas G. Kurtz, *Error Analysis of the tau-leap simulation method for stochastically modeled chemical reaction systems*, in preparation (to be submitted by January 2009).

[12] David F. Anderson, Jonathan C. Mattingly, *A weakly second-order method for the simulation of stochastic differential equations*, in preparation (to be submitted by January 2009).