

All of my graduate research could fall under the umbrella of Mathematical Biology. The obvious connection to real world issues drives me. Not only do I appreciate the ability to convey the application of my research to non-mathematicians, but I also appreciate the chance to collaborate with researchers in other fields. Moreover, though inspired by biological problems, the underlying approaches can be applied to many other areas.

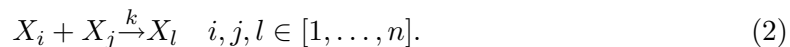
## 1 research experience

My thesis will be comprised of two main components that I worked on under the direction of Professor Shlomo Ta'asan. The problem that unites the parts is the ongoing quest to understand the complexities of the immune system. This requires integrating information across many different aspects of the immune system including gene expression, cell surface receptor expression, protein secretion, physiological parameters, and more. Our ultimate goal is to paint a more complete picture of the complicated system by creating dynamic models from the vast amount of data that measures these aspects. We are now faced with challenges that originate from interpreting the data collected of mRNA concentration, protein multiplex assays, and flow cytometry and successfully weaving this information together in order to build on our understanding. Each model built to describe the dynamics of a portion of the immune system response has the potential to greatly impact modern medicine by their individual contributions and as a whole.

**SYSTEM IDENTIFICATION** This was the first topic that I tackled as a graduate student. Our goal is to obtain a reasonable model of the interactions within the immune system based on temporal data, including but not limited to gene interaction models and protein interaction models. The simplest class of model involves unary reactions that are expressed by diagrams of the form,



where  $X_i$  and  $X_j$  represent the genes or proteins, the arrow indicates the relationship between them, and  $n$  is the total number of objects in the network. The  $k$  is the rate that  $X_i$  reacts with  $X_j$ . This represents all unary reactions between objects in the network. A more realistic model will also involve binary reactions of the form,



Our network will be comprised of both unary and binary reactions. We have two types of data collected via multiplex assays, one that measures mRNA concentrations and one that measures protein concentration. The mRNA data helps us determine gene expression while the protein data gives us insight into the secretion of specific molecules in the immune response. Both types of data will help us build dynamical network models of distinct systems.

**Modeling Gene Activation and Suppression using Probabilistic Models** From the data we have which contains time plots of the growth of specific proteins produced by classified immune system genes, we wanted to determine the activating and suppressive interactions between those genes. We were specifying these interactions as stochastic in nature. The probabilities we attempted to determine were defined by the likelihood that a gene affects another gene ( $k$ ), positively or negatively. In our attempt to realize these interactions, we produced various models that show protein production. Here we tried to solve the forward problem: given a specified number of probabilities, create a time plot of the protein productions. These models consider single genes producing one protein, simple two gene interactions producing two proteins, and random  $n$  gene interactions producing  $n$  proteins. After creating these models, we used them to solve the backward problem: given a time plot of protein productions, find the inhibiting and activating interactions between specific genes and their associated probabilities. Unfortunately, as our network becomes more complex, we could no longer recover these interaction probabilities.

**Models with Binary Interactions** We returned to the system identification problem with the idea that immune system interactions should contain binary components which could be modeled by quadratic systems. The simplified system (1) that only contains unary reactions can be described by the following equations,

$$\begin{cases} \dot{x}_i &= -kx_i, \\ \dot{x}_j &= kx_i. \end{cases}$$

All reactions of this form can be combined into the linear system,

$$\dot{x} = Ax,$$

where  $x$  is a vector of length  $n$  and  $A$  is a matrix whose entries are the appropriate interaction rates,  $k$ . However, in biology systems are rarely linear. Often two reactants are involved in the interaction as described by the binary reaction above. So we have a new system of equations that describes the relationships of the objects from system (2),

$$\begin{cases} \dot{x}_i &= -kx_ix_j, \\ \dot{x}_j &= -kx_ix_j, \\ \dot{x}_l &= kx_ix_j. \end{cases}$$

All reactions of this form can be combined into the quadratic system,

$$\dot{x} = Lxx^T,$$

where  $L$  is a third order tensor that associates the appropriate rate,  $k$ , to the quadratic combination of  $x_i$ 's. In order to create a system that could contain both unary and binary

reactions we must combine these two types of equations. This leads us to a quadratic problem,

$$\dot{x} = Ax + Lxx^T.$$

We must first discretize the equation before minimizing in order to utilize our temporal data,

$$\Delta x_{n+1} = x_{n+1} - x_n = (Ax_n + Lx_n x_n^T) \Delta t,$$

and reformulate this into a least squares problem,

$$\min_{A,L} \sum_n \|\Delta x_{n+1} - (Ax_n + Lx_n x_n^T) \Delta t\|^2.$$

Now we analytically recover rates that fit the data by converting the above equation into a single matrix equation and using eigenvalue decomposition. Unfortunately, these rates are not unique. Currently we are attempting to solve the non-uniqueness issue by introducing restrictions on the rates so they represent reactions. For example, when considering only unary reactions, the linear equations from system (1) can not have negative rates when  $i \neq j$ , only when  $i = j$ . We also know that unary reactions create an  $A$  where the columns sum to zero. Using these restrictions we transform our linear system into a constrained minimization problem,

$$\begin{aligned} \min_A \quad & \sum_n \|\Delta x_{n+1} - Ax_n \Delta t\|^2, \\ \text{s.t.} \quad & a_{ij} \leq 0 \quad i = j, \\ & -a_{ij} \leq 0 \quad i \neq j, \\ & \mathbf{1}^T A = 0. \end{aligned}$$

Using the primal-dual method, we then solve numerically for the unique minimizing  $A$ . We will place similar restrictions on  $A$  and  $L$  in the nonlinear case involving binary reactions.

**INTERVENTION DESIGN** Intervention is the process of specifying an additional input to a dynamic system to change the expected output. Commonly, this method is used for biological purposes for drug testing. However, our goal is not only to determine the correct amount of drug to administer, but also to determine the design of the drug. We have data that represents the immune system response of mice, and in it we observe a change in the immune system response of elderly mice to the response of healthy adult mice. When presented with a virus, the immune system of an elderly mouse has a slower, less condense response which often results in death. The implication is that if we can intervene with a drug to improve the immune system response in elderly mice then they could potentially survive virus that previously would have killed them. We want to limit the amount of virus found in the mouse while controlling the other physiological parameters, such as weight

loss and other indicators of disease severity measured in the data,  $x$ . Our problem can be formulated as,

$$\min_{\|B\|=1, u} \frac{1}{2} \int_0^{\infty} x^2 + \alpha u^2 dt,$$

while satisfying the constraint,

$$\dot{x} = Ax + Bu.$$

Take  $x$  and  $u$  to be time dependent variables and  $A$  and  $B$  to be time independent matrices. In this equation  $x$  is our immune response data,  $u$  is the amount of drug to give,  $B$  determines which drugs to administer, and  $A$  describes the dynamics of our system. Note that for simplicity, we assume that our system is linear. We wrote an algorithm which uses the Riccati Equation to find  $u$  and  $B$ . Yet in order for these values to be useful, they must be practical in terms of drug administration. Along with restricting the size of  $B$ , we also want to impose another restriction on the type drugs we can administer. There are certain aspects of the state of the mouse that we cannot change, for example we cannot give a drug that will directly increase the weight of the mouse. We added the following restriction on  $B$ ,

$$B^T C = 0,$$

where  $C$  represents the aspects of the mouse that we cannot directly change. There are still some open issues with this problem that we are working on.

## 2 research direction

**MATHEMATICAL BIOLOGY** I am excited about continuing research in Mathematical Biology. I am planning on solving the non-uniqueness issue with the system identification problem, as well as applying the intervention design problem to topics in partial differential equations. I am also interested in approaching new topics in Biology like Epidemiology and Cell Structure and Growth. As Mathematical Biology is a vast subject, I have discovered many mathematical interests through this field. Just in the last few years I approached biological problems with techniques arising from Probability, Control Theory, and Optimization, to name a few. I will continue to seek new ways to describe biological phenomena with mathematics.

**STUDENT RESEARCH** I had the opportunity to work on research projects as an undergraduate student. Some under the direction of professors at my own college and one through a summer program at Carnegie Mellon. These were a valuable experience for me and I look forward to the opportunity to introduce undergraduate students to mathematical research. Giving students a chance to apply the mathematics they have learned to real and significant problems is important for undergraduates not only for the

sake of research experience, but to excite them and to spur their interest in the material. I believe Mathematical Biology is the perfect topic for undergraduates to research. There are many problems that are accessible to a beginning mathematician, and the transparency of the direct real world applications give great motivation for students.