

Finite Fields are Better Booleans

Humberto Ortiz-Zuazaga

High Performance Computing facility
University of Puerto Rico
PO BOX 23334
San Juan PR 00931-3334
E-mail: humberto@hpcf.upr.edu

Reinhard Laubenbacher

Virginia Bioinformatics Institute
1880 Pratt Drive
Blacksburgh, VA 24061
E-mail: reinhard@vbi.vt.edu

María Alicia Aviño-Díaz

Department of Physics-Mathematics
University of Puerto Rico
Cayey, PR 00736
E-mail: m_avino@cayey1.upr.clu.edu

Oscar Moreno

Department of Mathematics and Computer Science
University of Puerto Rico at Rio Piedras
Rio Piedras, PR 00931
E-mail: o_moreno@upr1.upr.clu.edu

Abstract

Microarrays allow researchers to measure the expression of every gene in simple organisms. Understanding the biological processes underlying gene expression changes requires solving the reverse-engineering problem, where one determines a set of control functions that relate measurements of the genes in one state to the next state. Many models of gene networks have been proposed to serve as bases for the reverse engineering problem. Examples are Boolean networks, generalized logical networks, linear networks, Bayesian networks, differential equations, and neural networks. The best understood of these models are the Boolean networks, for which the complexity of several problems is known.

We show that Boolean genetic networks are Finite Dynamical Systems, and we will extend Boolean networks to networks over finite fields. This preserves most of the theoretical machinery of Boolean networks, but allows simulation of networks where genes have more than 2 states, and also brings a new set of tools from coding theory, engineering, and modern algebra to bear on the problem of reverse-engineering genetic networks from microarray data. We are currently exploring how these tools are best employed to reverse-engineer Finite Dynamical Systems from microarray data.

1 Introduction.

Microarray technology allows researchers to measure the expression of hundreds to thousands of genes simultaneously. It is currently possible to measure the expression of *every* gene in certain simple organisms such as yeast. Making sense of this data, however, is a difficult proposition. One technique for data abstraction is to reverse-engineer a set of control functions that describe the changes in gene expression based on gene expression measurements.

Many different models of genetic regulation have been proposed, and used to create genetic networks in the reverse-engineering problem. Hidde de Jong presents most of the genetic network models in his recent review [dJ02]. The model system best studied formally is the Boolean network model. These have been studied in the context of microarrays since at least 1996 [SS96], and in the context of genetic networks since 1973 [Tho73]. The complexity of several important problems in reverse engineering Boolean networks have been shown by [AKMM98]. In addition, there are robust algorithms for solving the reverse engineering problem in Boolean networks when certain restrictions about the number of genes that may influence another are observed [ITK00, LFS98].

We will show that these Boolean network models are examples of Finite Dynamical Systems [MBAD02] (FDS), and we will generalize Boolean networks to finite fields. We also show two different methods to construct these FDS mod-

els over finite fields from microarray experimental data. Our generalization, however, allows for a more natural treatment of microarray data than Boolean variables that have only two possible values, to a full range of discrete values. In addition, finite fields allow for a more natural algebraic treatment than Boolean variables, and also form vector spaces, which simplify the computations performed over them.

2 Deterministic Boolean Network Models

Various research groups have described genetic regulatory networks using Boolean variables to represent gene expression levels or stimuli [AKMM98, SS96, ITK00, LFS98]. Taking the model definition from [ITK00], for example, we can describe a genetic network as:

1. A graph consisting of N numbered nodes and, $1 \leq n \leq N$.
2. A set of directed edges between nodes.
3. A Boolean function f_n for each node.

An edge from a node to another represents an influence of the first node on the expression of the second.

3 The Boolean Genetic Network Model

In [OZAnDCB⁺] we define Boolean genetic network model (BGNM) as:

- A *Boolean variable* takes the values 0, 1.
- A *Boolean function* is a function of Boolean variables, using the operations and, or, not.

A *Boolean genetic network model* (BGNM) is:

- An n -tuple of Boolean variables $(x_1, \dots, x_n)$ associated with the genes
- An n -tuple of Boolean control functions $(f_1, \dots, f_n)$, describing how the genes are regulated

4 Finite Fields Correspond to Boolean Variables

The operations $+$ and $\cdot$ in Z_2 are defined as follows:

x	y	$x \cdot y$	$x + y$
1	1	1	0
0	1	0	1
1	0	0	1
0	0	0	0

Z_2 is a field with those two operations. Note that for any two Boolean variables x, y we have:

$$\begin{aligned}
 x \text{ and } y &= x \cdot y \\
 x \text{ or } y &= x + y + x \cdot y \\
 \text{not } x &= 1 + x \\
 x \text{ xor } y &= x + y
 \end{aligned}$$

We can now show that any function over Boolean variables can be realized instead as a function over the finite field Z_2 . In a natural manner, we can extend this to sets of Boolean functions, such as those contained in a Boolean network model. The advantage of finite fields and vector spaces is that they allow tools developed for cryptanalysis and communications theory to be applied to microarray data.

5 Boolean Networks are Finite Dynamical Systems

A finite dynamical system (FDS) is a pair (V, f) where V is the set of vectors over a finite field $\text{GF}(p^n)$ and $f : V \rightarrow V$.

A set of n Boolean functions such as those defining a Boolean network can be realized as a set of n functions over Z_2 instead. A FDS over Z_2 can also be decomposed into a set of n functions (in n variables constructed from the components of the vectors in V) from Z_2^n to Z_2 . In fact, we can show that these two descriptions are equivalent.

An FDS is also a finite state system (FSS, see [LN97]), a fact that we can exploit to simplify the computations described in the next section.

6 Reverse Engineering Finite Field Network Models

We are currently developing procedures for reverse engineering networks over finite fields extending the results in [ITK00]. In [LSS03], the problem

of reverse engineering genetic networks from time-series data is studied using a Gröbner basis method. We can use the results developed in [MBAD02] to change the multi-variable polynomials of [LSS03] into a one-variable polynomials which can be computed over a finite field. These computations may be more efficient than computations over Gröbner bases.

7 Conclusions

In this presentation we show that Boolean network models are finite dynamical systems and also finite state systems. We generalize the models in [AKMM98, ITK00] to finite fields. We provide two methods for reverse engineering these generalized models, one by extending the procedures in [ITK00] and another by simplifying the procedures in [LSS03].

References

- [AKMM98] T. Akutsu, S. Kuahara, O. Maruyama, and S. Miyano. Identification of gene regulatory networks by strategic gene disruptions and gene overexpressions. In H. Karloff, editor, *Proceedings of the 9th ACM-SIAM Symposium on Discrete Algorithms*. ACM Press, 1998.
- [dJ02] H. de Jong. Modeling and simulation of genetic regulatory systems: A literature review. *Journal of Computational Biology*, 9(1):67–103, 2002.
- [ITK00] T. E. Ideker, V. Thorsson, and R. M. Karp. Discovery of regulatory interactions through perturbation: Inference and experimental design. In *Pacific Symposium on Biocomputing*, number 5, pages 302–313, 2000.
- [LFS98] S. Liang, S. Fuhrman, and R. Somogyi. REVEAL: A general reverse engineering algorithm for inference of genetic network architectures. In *Pacific Symposium on Biocomputing*, number 3, pages 18–29, 1998.
- [LN97] R. Lidl and H. Niederreiter. *Finite Fields*. Cambridge University Press, 1997.
- [LSS03] R. Laubenbacher, J. Shah, and B. Stigler. Simulation of polynomial systems. To appear in *Simulation in the Health and Medical Sciences*, 2003.
- [MBAD02] O. Moreno, D. Bollman, and M. Aviño-Díaz. Finite dynamical systems, linear automata, and finite fields. In *WSEAS Transactions*, 2002.
- [OZAnDCB⁺] H. Ortiz-Zuazaga, M. A. Aviño Díaz, C. J. Corrada Bravo, R. Laubenbacher, S. Peña-de Ortiz, and Moreno O. Applications of finite fields to the study of microarray expression data. Submitted to the 11th International Conference on Intelligent Systems for Molecular Biology (ISMB 2003), June 29 to July 3, 2003, Brisbane, Australia.
- [SS96] R. Somogyi and C. Sniegowski. Modeling the complexity of genetic networks: Understanding multigenic and pleiotropic regulation. *Complexity*, (1):45–63, 1996.
- [Tho73] R. Thomas. Boolean formalization of genetic control circuits. *Journal of Theoretical Biology*, (42):563–585, 1973.