Physica D 241 (2012) 929–938

Contents lists available at SciVerse ScienceDirect

Physica D

journal homepage: www.elsevier.com/locate/physd

The number of multistate nested canalyzing functions[☆]

David Murrugarra [*], Reinhard Laubenbacher

Virginia Bioinformatics Institute, Virginia Tech, USA
Mathematics Department, Virginia Tech, USA

ARTICLE INFO

Article history:
Received 8 August 2011
Received in revised form
25 December 2011
Accepted 13 February 2012
Available online 24 February 2012
Communicated by K. Josic

Keywords:
Nested canalyzing functions
Formula
Multistate
Regulatory logic

ABSTRACT

Identifying features of molecular regulatory networks is an important problem in systems biology. It has been shown that the combinatorial logic of such networks can be captured in many cases by special functions called nested canalyzing in the context of discrete dynamic network models. It was also shown that the dynamics of networks constructed from such functions has very special properties that are consistent with what is known about molecular networks, and that simplify analysis. It is important to know how restrictive this class of functions is, for instance for the purpose of network reverse-engineering. This paper contains a formula for the number of such functions and a comparison to the class of all functions. In particular, it is shown that, as the number of variables becomes large, the ratio of the number of nested canalyzing functions to the number of all functions converges to zero. This shows that the class of nested canalyzing functions is indeed very restrictive. The principal tool used for this investigation is a description of these functions as polynomials and a parametrization of the class of all such polynomials in terms of relations on their coefficients. This parametrization can also be used for the purpose of network reverse-engineering using only nested canalyzing functions.

1. Introduction

A central problem of molecular systems biology is to understand the structure and dynamics of molecular networks, such as gene regulatory, signaling, or metabolic networks. Some progress has been made in elucidating general design principles of such networks. For instance, in [1] it was shown that certain graph theoretic motifs appear far more often in the topology of regulatory network graphs than would be expected at random. In [2,3] it was shown that a certain type of Boolean regulatory logic, encoded by the so-called *nested canalyzing* Boolean functions, has the kind of dynamic properties one would expect from molecular networks. In [4] we showed that the Boolean functions studied there do have a multistate generalization that shows similar dynamic properties. Furthermore, we showed that the large majority of regulatory rules that appear in published models of molecular networks, whether Boolean or multistate, do indeed have this form. Thus, there is evidence that multistate nested canalyzing rules capture key features of molecular regulation and deserve further study.

These rules, the so-called *nested canalyzing* rules, are a special case of *canalyzing* rules, which are reminiscent of Waddington's concept of canalyzation in gene regulation [5]. Nested canalyzing Boolean rules were shown in [6] to be identical with the class of unate cascade functions, which have been studied extensively in computer engineering. They represent exactly the class of Boolean functions that result in binary decision diagrams of shortest average path length [7]. This in itself has interesting implications for information processing in molecular networks. One consequence of this result is that a recursive formula derived earlier for the number of unate cascade functions of a given number of variables [8] applies to give a formula for the number of nested canalyzing Boolean functions, described in [6]. A formula for the number of canalyzing Boolean functions had been given in [9].

Many molecular networks cannot be described using the Boolean framework, since more than one threshold for a molecular species might be required to represent different modes of action. There are several frameworks available for multistate discrete models, such as the so-called logical models, Petri nets, and agent-based models. It has been shown in [10,11] that all these model types can be translated into the general and mathematically well-founded framework of polynomial dynamical systems over a finite number system. In [4] the concept of nested canalyzing logical rule has been generalized to such polynomial systems. It has been shown there, furthermore, that a large proportion of rules in multistate discrete models are indeed nested canalyzing, showing that this concept captures an important feature of the regulatory logic of molecular networks.

As was pointed out in [9,6], knowing the number of nested canalyzing rules for a given number of input variables and for a given number of possible variable states is important because on

☆ This research was supported by the National Science Foundation under Grant Number CMMI-0908201.

* Correspondence to: Virginia Bioinformatics Institute, Virginia Tech, Blacksburg, VA 24061, USA. Tel.: +1 540 231 9264; fax: +1 540 231 2606.
E-mail address: davidmur@vt.edu (D. Murrugarra).

0167-2789/$ – see front matter. Published by Elsevier B.V.
doi:10.1016/j.physd.2012.02.011

the one hand it provides an estimate of how plausible it is that such rules have evolved as regulatory principles and, on the other hand, provides an estimate of how restrictive the set of rules is. The latter is important, for instance, for the reverse-engineering of networks [12]. If the set of rules is sufficiently restrictive, then the reverse-engineering problem, which is almost always underdetermined due to limited data, becomes more tractable when restricted to a smaller model space. Knowing the proportion of regulatory rules that are nested canalyzing gives an estimate of how much one can additionally constrain the problem by limiting the search space of reverse-engineering algorithms to these rules. That this is indeed the case is an important consequence of the results in this paper. We present a formula for the number of nested canalyzing functions in a given number of variables and show that the ratio of nested canalyzing functions and all multistate functions converges to zero as the number of variables increases. We follow the approach in [6] and solve the problem within the framework of polynomial dynamical systems, which makes it possible to frame it as a problem of counting solutions to a system of polynomial equations.

2. Nested canalyzing functions

As mentioned in the previous section, it is possible to view most discrete models within the framework of dynamical systems over a finite number system, or finite field. For our purposes we will use the finite fields $\mathbb{F}_p = \{0, 1, \ldots, p-1\}$, p an arbitrary prime number, otherwise known as $\mathbb{Z}/p$, the integers modulo p. Furthermore, we will assume that $\mathbb{F}_p$ is totally ordered under the canonical order, that is, its elements are arranged in linear increasing order, $\mathbb{F}_p = \{0 < 1 < \cdots < p-1\}$. Let $\mathbb{F} = \mathbb{F}_p$ for some prime p. We first recall the general definition of a nested canalyzing function in variables $x_1, \ldots, x_n$ from [4]. The underlying idea is as follows: a rule is nested canalyzing, if there exists a variable x such that, if x receives certain inputs, then it by itself determines the value of the function. If x does not receive these certain inputs, then there exists another variable y such that, if y receives certain other inputs, then it by itself determines the value of the function, and so on, until all variables are exhausted.

Definition 2.1. Let $S_i \subset \mathbb{F}$, $i = 1, \ldots, n$, be subsets that satisfy the property that each S_i is a proper, nonempty subinterval of $\mathbb{F}$; that is, every element of $\mathbb{F}$ that lies between two elements of S_i in the chosen order is also in S_i. Furthermore, we assume that the complement of each S_i is also a subinterval, that is, each S_i can be described by a threshold s_i, with all elements of S_i either larger or smaller than s_i. Let σ be a permutation on $\{1, \ldots, n\}$.

- The function $f : \mathbb{F}^n \to \mathbb{F}$ is a nested canalyzing function in the variable order $x_{\sigma(1)}, \ldots, x_{\sigma(n)}$ with canalyzing input sets $S_1, \ldots, S_n \subset \mathbb{F}$ and canalyzing output values $b_1, \ldots, b_n, b_{n+1} \in \mathbb{F}$, with $b_n \neq b_{n+1}$, if it can be represented in the form

$$f(x_1, \ldots, x_n) = \begin{cases} b_1 & \text{if } x_{\sigma(1)} \in S_1, \\ b_2 & \text{if } x_{\sigma(1)} \notin S_1, x_{\sigma(2)} \in S_2, \\ \vdots & \\ b_n & \text{if } x_{\sigma(1)} \notin S_1, \ldots, x_{\sigma(n)} \in S_n, \\ b_{n+1} & \text{if } x_{\sigma(1)} \notin S_1, \ldots, x_{\sigma(n)} \notin S_n. \end{cases}$$

- The function $f : \mathbb{F}^n \to \mathbb{F}$ is a nested canalyzing function if it is a nested canalyzing function in some variable order $x_{\sigma(1)}, \ldots, x_{\sigma(n)}$ for some permutation σ on $\{1, \ldots, n\}$.

It is straightforward to verify that, if $p = 2$, that is $\mathbb{F} = \{0, 1\}$, then we recover the definition in [2] of a Boolean nested canalyzing rule. We emphasize that several important classes of multistate discrete models can be represented in the form of a dynamical system $f : \mathbb{F}^n \longrightarrow \mathbb{F}^n$, as mentioned above, so that the concept of a nested canalyzing rule defined in this way has broad applicability.

Example 2.2. Let $\mathbb{F}$ be the field with three elements, i.e. $\mathbb{F} = \{0, 1, 2\}$. The function $f : \mathbb{F}^2 \to \mathbb{F}$ given by

$$f(x_1, x_2) = \begin{cases} 1 & \text{if } x_1 \in \{0, 1\}, \\ 2 & \text{if } x_1 \notin \{0, 1\}, x_2 \in \{2\}, \\ 0 & \text{if } x_1 \notin \{0, 1\}, x_2 \notin \{2\}, \end{cases}$$

is nested canalyzing in the variable order x_1, x_2, with canalyzing input sets $S_1 = \{0, 1\}$, $S_2 = \{2\} \subset \mathbb{F}$, and canalyzing output values $b_1 = 1$, $b_2 = 2$, $b_3 = 0 \in \mathbb{F}$.

3. Polynomial form of nested canalyzing functions

We now use the fact that any function $f : \mathbb{F}^n \to \mathbb{F}$ can be expressed as a polynomial in n variables [13, p. 369]. In this section we determine the polynomial form of nested canalyzing functions. That is, we will determine relationships among the coefficients of a polynomial that make it nested canalyzing. We follow the approach in [6]. Let B_n be the set of functions from $\mathbb{F}^n$ to $\mathbb{F}$, i.e., $B_n = \{f : \mathbb{F}^n \longrightarrow \mathbb{F}\}$. The set B_n is endowed with an addition and multiplication that is induced from that of $\mathbb{F}$, which makes it a ring. Let I be the ideal of the ring of polynomials $\mathbb{F}[x_1, \ldots, x_n]$ generated by the polynomials $\{x_i^p - x_i\}$ for all $i = 1, \ldots, n$, where p is the number of elements in $\mathbb{F}$. There is an isomorphism between B_n and the quotient ring $\mathbb{F}[x_1, \ldots, x_n]/I$ which is also isomorphic to

$$R = \left\{ \sum_{\substack{(i_1,\ldots,i_n) \\ i_t \in \mathbb{F} \\ t=1,\ldots,n}} C_{i_1 \cdots i_n} x_1^{i_1} x_2^{i_2} \cdots x_n^{i_n} \right\}.$$

Now we use this identification to study nested canalyzing functions as elements of R.

Given a subset S of $\mathbb{F}$, we will denote by Q_S the indicator function of the complement of S, i.e., for $x_0 \in \mathbb{F}$, let

$$Q_S(x_0) = \begin{cases} 0 & \text{if } x_0 \in S, \\ 1 & \text{if } x_0 \notin S. \end{cases}$$

We will derive the polynomial form for $Q_S(x)$ in Lemma A.2. The following theorem gives the polynomial form of a nested canalyzing function.

Theorem 3.1. *Let f be a function in R. Then the function f is nested canalyzing in the variable order $x_1, \ldots, x_n$ with canalyzing input sets $S_1, \ldots, S_n$ and canalyzing output values $b_1, \ldots, b_n, b_{n+1}$, with $b_n \neq b_{n+1}$, if and only if it has the polynomial form*

$$f(x_1, \ldots, x_n) = \sum_{j=0}^{n-1} \left\{ B_{n-j} \prod_{i=1}^{n-j} Q_{S_i}(x_i) \right\} + b_1, \tag{3.1}$$

where Q_{S_i} is defined as in Lemma A.2 and $B_{n-j} = (b_{n-j+1} - b_{n-j})$ for $j = 0, \ldots, n-1$.

Proof. Let f be a nested canalyzing function as in Definition 2.1, and let

$$g(x_1, \ldots, x_n) = \sum_{j=0}^{n-1} \left\{ B_{n-j} \prod_{i=1}^{n-j} Q_{S_i}(x_i) \right\} + b_1.$$

Since g has the right form to be in R, we can use the isomorphism between B_n and R, to reduce the proof to showing that

$$g(a_1, \ldots, a_n) = f(a_1, \ldots, a_n)$$

for all $(a_1, \ldots, a_n) \in \mathbb{F}^n$.

If $a_1 \in S_1$, then $Q_{S_1}(a_1) = 0$; therefore

$$g(a_1, \ldots, a_n) = b_1 \quad \text{whenever } a_1 \in S_1.$$

If $a_1 \notin S_1$ and $a_2 \in S_2$, then $Q_{S_1}(a_1) = 1$ and $Q_{S_2}(a_2) = 0$; therefore

$$g(a_1, \ldots, a_n) = (b_2 - b_1) + b_1 = b_2.$$