



Solving Equations on Discrete Dynamical Systems

Alberto Dennunzio, Enrico Formenti, Luciano Margara, Valentin Montmirail,
Sara Riva

► To cite this version:

Alberto Dennunzio, Enrico Formenti, Luciano Margara, Valentin Montmirail, Sara Riva. Solving Equations on Discrete Dynamical Systems. 2019. hal-02271381

HAL Id: hal-02271381

<https://hal.univ-cotedazur.fr/hal-02271381>

Preprint submitted on 26 Aug 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Solving Equations on Discrete Dynamical Systems

Alberto Dennunzio

DISCo - Università degli Studi di Milano-Bicocca, Italy
dennunzio@disco.unimib.it

Enrico Formenti

Université Côte d’Azur, CNRS, I3S, Nice, France
enrico.formenti@univ-cotedazur.fr

Luciano Margara

Università degli Studi di Bologna, Campus di Cesena, Cesena, Italy
margara@cs.unibo.it

Valentin Montmirail

Université Côte d’Azur, CNRS, I3S, Nice, France
valentin.montmirail@univ-cotedazur.fr

Sara Riva

Université Côte d’Azur, CNRS, I3S, Nice, France
sara.riva@univ-cotedazur.fr

Abstract

Boolean automata networks, genetic regulation networks, and metabolic networks are just a few examples of biological modeling by discrete dynamical systems (DDS). A major issue in modeling is the verification of the model against the experimental data or inducing the model under uncertainties in the data. Equipping finite discrete dynamical systems with an algebraic structure of commutative semiring provides a suitable context for hypothesis verification on the dynamics of DDS. Indeed, hypothesis on the systems can be translated into polynomial equations over DDS. Solutions to these equations provide the validation to the initial hypothesis. Unfortunately, finding solutions to general equations over DDS is undecidable. In this article, we want to push the envelop further by proposing a practical approach for some decidable cases in a suitable configuration that we call the *Hypothesis Checking*. We demonstrate that for many decidable equations all boils down to a “simpler” equation. However, the problem is not to decide if the simple equation has a solution, but to enumerate all the solutions in order to verify the hypothesis on the real and undecidable systems. We evaluate experimentally our approach and show that it has good scalability properties.

2012 ACM Subject Classification

Keywords and phrases Boolean Automata Networks, Discrete Dynamical Systems, Decidability

Digital Object Identifier 10.4230/LIPIcs...

Acknowledgements Sara Riva is the main contributor.

1 Scientific Background

Boolean automata networks have been heavily used in the study of systems biology [2, 6]. The main drawback of the approach by automata network is in the very first step, namely when one induces the network from the experiments. Indeed, most of the time the knowledge of about the network is partial and hypothesis are made about its real structure. Those hypotheses must be verified either by further experiments or by the study of the dynamical evolution of the network compared to the expected behaviour provided by the experimental evidence.

In [3], Dennunzio *et al.* propose an abstract algebraic setting for representing the dynamical evolution of finite discrete dynamical systems have been proposed. The basic idea is to



© Dennunzio, Formenti, Margara, Montmirail and Riva;
licensed under Creative Commons License CC-BY

Leibniz International Proceedings in Informatics

LIPICs Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

identify a discrete dynamical system with the graph of its dynamics (finite graphs having out-degree exactly 1) and then define operations $+$ and \cdot which compose dynamical systems to obtain larger ones. Using dynamical systems to represent biological regulatory network is not new and we redirect the reader to [7, 9] for more information.

Indeed, a discrete dynamical system (DDS) is a structure $\langle X, f \rangle$ where X is a finite set called **the set of states** and $f : X \rightarrow X$ is a function called the **next state map**. Any DDS $\langle X, f \rangle$ can be identified with its **dynamics graph** which is a structure $G \equiv \langle V, E \rangle$ where $V = X$ and $E = \{(a, b) \in V \times V, f(a) = b\}$. From now on, when speaking of a DDS, we will always refer to its dynamics graph.

Given two DDS $G_1 = \langle V_1, E_1 \rangle$ and $G_2 = \langle V_2, E_2 \rangle$ their **sum** $G_1 + G_2$ is defined as $\langle V_1 \cup V_2, E_1 \cup E_2 \rangle$. The **product** $G_1 \cdot G_2$ is the structure $\langle V', E' \rangle$ where $V' = V_1 \times V_2$ and $E' = \{((a, x), (b, y)) \in V' \times V', (a, b) \in E_1 \text{ and } (x, y) \in E_2\}$. It is easy to see that $F \equiv \langle X, +, \cdot \rangle$ is a commutative semiring in which $\langle \emptyset, \emptyset \rangle$ is the neutral element *w.r.t.* $+$ and $\langle \{a\}, \{(a, a)\} \rangle$ is the neutral element *w.r.t.* \cdot operation.

Now, consider the semiring $R[X_1, X_2, \dots, X_n]$ of polynomials over R in the variables X_i , naturally induced by R . Let us go back to our initial motivation. Assume that some parts of the overall dynamics a_1, a_2, \dots, a_k are known, then the following equation represents a hypothesis on the overall structure of the expected dynamical system C on the basis of the known data a_1, \dots, a_k .

$$a_1 \cdot X_1 + a_2 \cdot X_2 + \dots + a_k \cdot X_k = C \quad (1)$$

The hypotheses are verified whenever the previous equation admits a solution, therefore providing a way to solve such equation can be used to check hypotheses against a given dynamical system. For the sake of clarity, we denote our unknown variables as X_i , whereas they, in fact, represent any polynomial equation of the shape $x_i^{w_i}$. The following fundamental result states that solving polynomial equations over DDS is not an easy task.

► **Theorem 1** (Dennuzio et al. [3]).

Given two polynomials $P(X_1, \dots, X_n)$ and $Q(X_1, \dots, X_n)$ over $R[X_1, \dots, X_n]$, consider the following equation

$$P(X_1, \dots, X_n) = Q(X_1, \dots, X_n). \quad (2)$$

The problem of finding a solution to Equation 2 is undecidable. Moreover, if Equation 2 is linear or quadratic, then finding a solution is in NP. Finally, when $P(X) = \text{const}$, where the polynomial is in a single variable and all its coefficients are systems consisting of self-loops only, the equation is solvable in polynomial time.

According to Theorem 1, solving polynomial equations of the type $P(X) = \text{const}$ is in NP even for quadratic polynomials and no efficient algorithm is known unless $P = NP$. In order to overcome this issue, one can follow at least two strategies: either further constrain the polynomials or solve approximated equations which can provide information on the real solution.

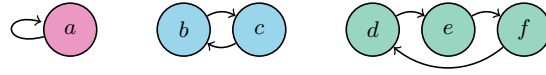
In this article, we follow the second option. Indeed, we focus on strongly connected components (SCC) of the dynamics graph. Recall that SCC represents a very important feature in finite DDS since they are the attracting sets. These sets contain the asymptotic information about system evolution.

2 Methods

In the dynamics graph, each component of a system can be divided in two parts: the transient part and the periodic part, see [5] for more details. A point $x \in X$ of a dynamical system $\langle X, f \rangle$ belongs to a cycle if there exists a positive number $p \in \mathbb{N}$ such that $f^p(x) = x$. The smallest p is the period of the cycle, and x is periodic. The periodic part is the set of nodes periodic. All the others nodes are transient, but in this work, X is a finite set hence any state x is **ultimately periodic** and in each component of the graph there is only one cycle of length at least 1.

Every finite DDS can be described as a sum of single components, and every component can be described, for our purposes, with the length of its period (strongly connected components in dynamics graphs are cycles). The transient part of a component is not relevant for the result of the sum and product operations when the equation is over SCC.

A single component of period p is denoted C_p^1 , while C_p^n means that there are n components of period p in the system. Therefore, if a system is composed by n components, each of period p_i with $i \in \{1, \dots, n\}$, then $\bigoplus_{i=1}^n C_{p_i}^1$ completely describes the system where \bigoplus denotes the sum of components since each component has only one period (see Figure 1).



■ **Figure 1** a DDS with three components: $(C_1^1 \oplus C_2^1 \oplus C_3^1)$ in our notation.

► **Remark 2.** When a system has several components with the same period, then their representation can be added. As an example, we have $C_2^1 \oplus C_2^1 = C_2^2$. Otherwise, the sum \oplus consists of a concatenation of components.

3 Contributions

From now on, \bar{R} will indicate the restriction of R to systems made by strongly connected components only. First, we need to adapt the definition of product between two DDS in terms of components and their period.

► **Definition 3.** For a system composed by m components C_p^m , multiplied by a system with n components C_q^n , the result of the product operation depends only on the length of the periods of the components involved according to the following formula

$$\forall n, m, p, q \in \mathbb{N} \setminus \{0\} \quad C_p^m \odot C_q^n = C_{\text{lcm}(p,q)}^{m \times n \times \text{gcd}(p,q)} . \quad (3)$$

One can also simplify the parameter of a component. The following definition provides a formula to compact the notation of a DDS with n identical components.

► **Definition 4.** Consider a single component C_p^m , then $\forall n, m, p \in \mathbb{N} \setminus \{0\}$ it holds

$$C_p^{mn} = n \cdot C_p^m . \quad (4)$$

Let us remind that each X_i represents, in fact, a variable $x_i^{w_i}$. Therefore, it is necessary to know how we can retrieve the solutions for the original x_i . To do so, we will use the following lemma:

XX:4 Solving Equations on Discrete Dynamical Systems

► **Lemma 5.** Given $S \equiv C_{p_1}^1 \oplus C_{p_2}^1 \oplus \dots \oplus C_{p_m}^1$ with m components of period $p_i \in \mathbb{N} \setminus \{0\}$, $\forall i \in \{1, \dots, m\}$, let $g(p_1, p_2, \dots, p_m, k_1, k_2, \dots, k_m)$ the gcd between the p_i for which $k_i \neq 0$ and $l(p_1, p_2, \dots, p_m, k_1, k_2, \dots, k_m)$ the lcm between the p_i for which $k_i \neq 0$.

$$(S)^n = \bigoplus_{i=1}^m C_{p_i}^{p_i^{n-1}} \oplus \bigoplus_{\substack{k_1+k_2+\dots+k_m=n \\ 0 \leq k_1, k_2, \dots, k_m < n}} \binom{n}{k_1, k_2, \dots, k_m} C_l^{g \cdot \prod_{t=1}^m p_t^{k_t-1}}.$$

Proof.

Using the multinomial theorem one finds

$$\begin{aligned} (S)^n &= (C_{p_1}^1 \oplus C_{p_2}^1 \oplus \dots \oplus C_{p_m}^1)^n = \bigoplus_{k_1+k_2+\dots+k_m=n} \binom{n}{k_1, k_2, \dots, k_m} \bigodot_{t=1}^m (C_{p_t}^1)^{k_t} = \\ &= \bigoplus_{i=1}^m (C_{p_i}^1)^n \oplus \bigoplus_{\substack{k_1+k_2+\dots+k_m=n \\ 0 \leq k_1, k_2, \dots, k_m < n}} \binom{n}{k_1, k_2, \dots, k_m} \bigodot_{t=1}^m (C_{p_t}^1)^{k_t} \end{aligned} \quad (5)$$

The resulting formula 5 is obtained by extrapolating the cases in which a $k_i = n$. Another transformation is possible according with the Lemma 3.

$$\begin{aligned} &\bigoplus_{i=1}^m (C_{p_i}^1)^n \oplus \bigoplus_{\substack{k_1+k_2+\dots+k_m=n \\ 0 \leq k_1, k_2, \dots, k_m < n}} \binom{n}{k_1, k_2, \dots, k_m} \bigodot_{t=1}^m (C_{p_t}^1)^{k_t} = \\ &= \bigoplus_{i=1}^m (C_{p_i}^1)^n \oplus \bigoplus_{\substack{k_1+k_2+\dots+k_m=n \\ 0 \leq k_1, k_2, \dots, k_m < n}} \binom{n}{k_1, k_2, \dots, k_m} C_{l(p_1, p_2, \dots, p_m, k_1, k_2, \dots, k_m)}^{g(p_1, p_2, \dots, p_m, k_1, k_2, \dots, k_m) \cdot \prod_{t=1}^m p_t^{k_t-1}} = \\ &= \bigoplus_{i=1}^m C_{p_i}^{p_i^{n-1}} \oplus \bigoplus_{\substack{k_1+k_2+\dots+k_m=n \\ 0 \leq k_1, k_2, \dots, k_m < n}} \binom{n}{k_1, k_2, \dots, k_m} C_{l(p_1, p_2, \dots, p_m, k_1, k_2, \dots, k_m)}^{g(p_1, p_2, \dots, p_m, k_1, k_2, \dots, k_m) \cdot \prod_{t=1}^m p_t^{k_t-1}}. \end{aligned}$$

◀

For k equal to 0 we assume that $(S)^0$ is equal to C_1^1 , the neutral element of the product operation. Let us go back to Equation 1 which is the problem that we want to solve. It can be rewritten as follows:

$$\left(\bigoplus_{j=1}^{S_1} C_{p_{1j}}^1 \odot X_1 \right) \oplus \left(\bigoplus_{j=1}^{S_2} C_{p_{2j}}^1 \odot X_2 \right) \oplus \dots \oplus \left(\bigoplus_{j=1}^{S_k} C_{p_{kj}}^1 \odot X_k \right) = \bigoplus_{j=1}^m C_{q_j}^{m_j} \quad (6)$$

with S_i , the number of different components in the system i , p_{ij} is the value of the period of the j^{th} component in the system i . In the right term, there are m different periods, where for the j^{th} different period, n_j is the number of components, and q_j the value of the period. However, Equation (6) is still hard to solve. We can simplify it performing a **contraction step** which consists in cutting Equation (6) into two simpler equations: $(C_{p_{11}}^1 \odot X_1) = W$, where $W \subseteq \bigoplus_{i=1}^m C_{q_i}^{n_i}$ and $((C_1^1 \odot Y) = \bigoplus_{i=1}^m C_{q_i}^{n_i} \setminus W)$ with $Y = (\bigoplus_{i=2}^{S_1} C_{p_{1i}}^1 \odot X_1) \oplus (\bigoplus_{j=1}^{S_2} C_{p_{2j}}^1 \odot X_2) \oplus \dots \oplus (\bigoplus_{j=1}^{S_k} C_{p_{kj}}^1 \odot X_k)$. By applying recursively a contraction step on all the partitions

of W and on the second equation obtained (*i.e.* the one containing Y) one finds that, solving Equation (6) boils down to solving multiple times the following type of equation:

$$C_p^1 \odot X = C_q^m. \quad (7)$$

If the variable X presents a power different from one, it is possible use the Lemma 5 in order to study the squared by the power.

However, equations of the shape of Equation 7 will be numerous therefore an efficient practical algorithm able to enumerate all its solutions is needed. In fact, we can propose the following bounds to know how many times equations of the shape Equation 7 are solved with the following lemma:

► **Lemma 6.** *Let us denote by Z the number of times that we will solved equation of the shape Equation 7, we have the following: $\prod_{i=1}^m \binom{n_i + \sum_{j=1}^k S_j - 1}{\sum_{j=1}^k S_j - 1} \cdot m \leq Z \leq \prod_{i=1}^m \binom{n_i + \sum_{j=1}^k S_j - 1}{\sum_{j=1}^k S_j - 1} \cdot m \cdot \sum_{j=1}^k S_j$.*

The intuition is as follows: the contraction step is necessary in order to study all possible way for produce the right term with the components in the left part of the equation. Accordingly, it is necessary understand the number of possible decompositions of the right term in order to discover the bounds for the number of the executions of the **colored-tree** method (a decomposition corresponds to assign a subset of the components of the right part to a product operation between a variable and a known component). For each period a Star and Bars decomposition is applied (we redirect the reader unfamiliar with the Star and Bars decomposition to [4]).

Proof. In general for a fixed q_i , the n_i components are divided in $\sum_{j=1}^k S_j$ groups, in this case there are $\binom{n_i + \sum_{j=1}^k S_j - 1}{\sum_{j=1}^k S_j - 1}$ different ways for divide the components. Therefore we can rewrite the lemma as follows: $m \leq \frac{Z}{\prod_{i=1}^m \binom{n_i + \sum_{j=1}^k S_j - 1}{\sum_{j=1}^k S_j - 1}} \leq m \cdot \sum_{j=1}^k S_j$. And now, toward a

contradiction for the lower bound. Let us assume that we can solve less than m equations. This implies that we solve less equations than the number of different periods on the right term. Contradiction, we need at least all of them (not necessary all their combination) to determine the solution of the equation. And now, toward a contradiction again to prove the upper-bound. Firstly, we know for all the components in the right term there are $\prod_{i=1}^m \binom{n_i + \sum_{j=1}^k S_j - 1}{\sum_{j=1}^k S_j - 1}$ feasible divisions. Now, let us assume that in the worst case, for each coefficient the product operation must produce more than one components of each possible periods in the right term. This is a contradiction from the definition of the equation, where all the components must all have a different period. The second possibility to go beyond this bounds is that it would exists more S_i than the one present in the equation, again a contradiction by definition of the equation. Therefore, we know that we have: $\prod_{i=1}^m \binom{n_i + \sum_{j=1}^k S_j - 1}{\sum_{j=1}^k S_j - 1} \cdot m \leq Z \leq \prod_{i=1}^m \binom{n_i + \sum_{j=1}^k S_j - 1}{\sum_{j=1}^k S_j - 1} \cdot m \cdot \sum_{j=1}^k S_j$, for Z being the number of times that we will solved equation of the shape Equation 7. ◀

4 The Colored-Table Method

First of all, let us formally define the problem and analyze its complexity.

► **Definition 7** (DSECP). *The (finite) Discrete Dynamical Systems Solving Equations on Components Problem is a problem which takes in input C_p^1 and C_q^n and outputs the list of all the solutions X to the equation $C_p^1 \odot X = C_q^n$.*

Solving DSECP is hard but still tractable. Indeed, the following lemma classifies our problem in EnumP. Recall that EnumP is the complexity class of enumeration problems for which a solution can be verified in polynomial time [8]. It can be seen as the enumeration counterpart of the NP complexity class.

► **Lemma 8.** DSECP is in EnumP.

Proof. One just needs to be able to check if a given value is a solution in polynomial time. This can be done in linear time using Lemma 3. ◀

4.0.0.1 Notation.

For any $n, p, q \in \mathbb{N}^*$, let $T_{p,q}^n$ denote the set of solutions of Equation (7) and $S_{p,q}^n$ the set of solutions returned by the `colored-tree` method.

The `colored-tree` method is pretty involved, we prefer start to illustrate it by an example.

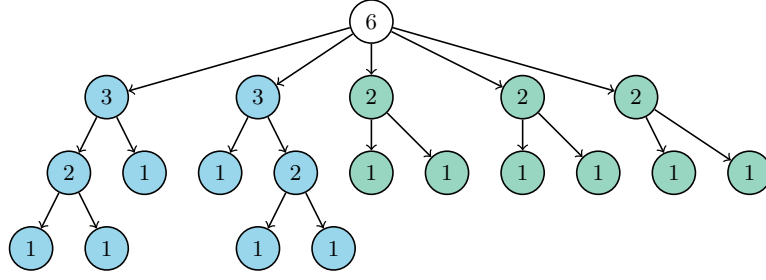
► **Example 9.** Consider the following equation $C_2^1 \odot X = C_6^6$. The algorithm consists in two distinct phases: tree building and solution aggregation. In the first phase, the algorithm enumerates all the divisors \mathcal{D} of 6 i.e. $\{6, 3, 2, 1\}$. It then applies a making-change decomposition algorithm (MCDA) [1] in which the total sum is 6 and the allowed set of coins is $\mathcal{D}' = \mathcal{D} \setminus \{6\}$. MCDA decomposes 6 as $3 + 3$ (which is an optimal decomposition). MCDA is then applied recursively (always using $\mathcal{D} \setminus \{i\}$ as the set of coins to decompose i). We obtain $(6 = 3 + 3)$, $(3 = 2 + 1)$ and $(2 = 1 + 1)$ as reported in Table 1. At this point, a check is

Node	Splits	Node solution	Subtree solutions set
6	[3,3][2,2,2]	C_6^1	$\{C_6^1, C_3^2, C_1^1 + C_2^1 + C_3^1, C_3^1 + C_1^3, C_2^1 + C_1^4, C_1^6, C_2^3, C_1^2 + C_2^2\}$
3	[2,1]	C_3^1	$\{C_3^1, C_1^1 + C_2^1, C_1^3\}$
2	[1,1]	C_2^1	$\{C_1^2, C_1^1\}$
1	\emptyset	C_1^1	$\{C_1^1\}$

■ **Table 1** Final data-structure storing all the decompositions, each solution for each value and at each step, the set of all solutions for a given value.

performed to ensure that all possible ways of decomposing 6 using \mathcal{D}' are present in the tree. In our case, we already have [3,3] found by the first run of MCDA. We also found: [3, 2, 1], [2, 2, 1, 1], [1, 1, 2, 1, 1], [1, 1, 1, 1, 1, 1] by the recursive application of MCDA. By performing the check, we discover that the decomposition of 6 as [2, 2, 2] is not represented in the current tree. For this reason, [2, 2, 2] is added to the set of decompositions of 6 as illustrated in Figure 2, it is assigned a new color and a recursive application of MCDA is started on the newly added nodes. A new check ensures that all decompositions are present. This ends the building phase. The resulting tree is reported in Figure 2.

After this first phase of construction of the tree, the aggregation of solutions starts. Remark that each node m represents the equation $C_p^1 \odot X = C_q^m$ that we call the **node equation**. The single component solution is called the **node solution** and it is obtained



■ **Figure 2** The colored tree for the equation $C_2^1 \odot X = C_6^6$ after the completeness check.

thanks to Lemma 3, $C_{\frac{q}{p} \times m}^1$ whenever a **feasible solution** exists *i.e.* if $\gcd(p, \frac{q}{p} \times m) = m$ and $\text{lcm}(p, \frac{q}{p} \times m) = q$. For example, for $m = 3$ one finds $x = C_3^1$. To find all the solutions for the current node one must also take the Cartesian product of the solutions sets in the subtrees of the same color and then the union of the solution sets of nodes of different colors (different splits). All the solution can be found in Table 1.

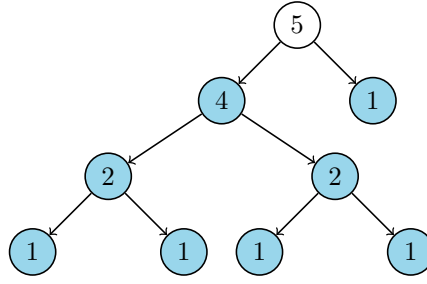
► **Example 10.** Consider the equation $C_2^1 \odot X = C_4^5$. In the first phase, the algorithm enumerates all the divisors \mathcal{D} of 4 *i.e.* $\{4, 2, 1\}$. It then applies a making-change decomposition algorithm (MCDA) [1]. MCDA decomposes 5 as $4 + 1$ (which is an optimal decomposition). MCDA is then applied recursively always using $\mathcal{D} \setminus \{i\}$ as the set of coins to decompose i . We obtain $(5 = 4 + 1)$, $(4 = 2 + 2)$ and $(2 = 1 + 1)$ as reported in Table 2. At this point, a

Node	Splits	Node solution	Subtree solutions set
5	[4,1]	$\{\}$	$\{\}$
4	[2,2]	$\{\}$	$\{C_4^2\}$
2	[1,1]	C_4^1	$\{C_4^1\}$
1	\emptyset	$\{\}$	$\{\}$

■ **Table 2** Final data-structure storing all the decomposition, each solution for each value and at each step, the set of all solutions for a given value.

check is performed to ensure that all possible ways of decomposing 5 using $\mathcal{D} \setminus \{i\}$ as the set of coins to decompose i . In our case, we already have $[4, 1]$ found by the first run of MCDA. We also found: $[2, 2, 1]$, $[2, 1, 1, 1]$, $[1, 1, 1, 1, 1]$ by the recursive application of MCDA. By performing the check, we discover that all the possible decompositions of 5 are represented in the current tree. This ends the building phase. The resulting tree is reported in Figure 3. After this first phase of construction of the tree, the aggregation of solutions starts. In this case the tree presents only one color. Remark that if in the cartesian product a empty set is involved, the result of the operation is the empty set. For example, for $m = 2$, one has that the node solution is C_4^1 . From the subtrees of the node one finds a empty set, but with the union of the solution of the node, the subtree solutions set for $m = 2$ is $\{C_4^1\}$. Moreover, the final solution set for the node 5 is the empty set, in fact in the Cartesian product $m = 1$ is involved (empty set) but also the solution node for $m = 5$ is empty. In this case the method return a empty set of solutions, that represents the impossibility of the equation.

► **Example 11.** Consider the equation $C_2^1 \odot X = C_6^{12}$. In the first phase, the algorithm enumerates all the divisors \mathcal{D} of 6 *i.e.* $\{6, 3, 2, 1\}$. It then applies a making-change decomposition algorithm (MCDA) [1]. MCDA decomposes 12 as $6 + 6$ (which is an optimal decomposition).



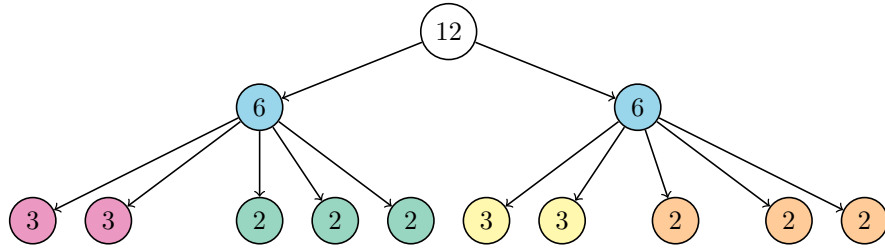
■ **Figure 3** The tree represented in the table for $C_2^1x = C_4^5$, after the check of completeness.

MCDA is then applied recursively always using $\mathcal{D} \setminus \{i\}$ as the set of coins to decompose i . We obtain $(12 = 6 + 6)$, $(6 = 3 + 3)$, $(3 = 2 + 1)$ and $(2 = 1 + 1)$ as reported in Table 3.

Node	Splits	Node solution	Subtree solutions set
12	[6,6]	$\{\}$	$\{C_3^4 + C_6^4, C_3^{12}, C_6^6, C_3^6 + C_6^3, C_6^2 + C_3^8, C_3^2 + C_6^5, C_6^1 + C_3^{10}\}$
6	[3,3] [2,2,2]	$\{\}$	$\{C_3^6, C_6^2 + C_3^2, C_3^4 + C_6^1, C_6^3\}$
3	[2,1]	$\{\}$	$\{C_3^3, C_6^1 + C_3^1\}$
2	[1,1]	C_6^1	$\{C_6^1, C_3^2\}$
1	\emptyset	C_3^1	$\{C_3^1\}$

■ **Table 3** Final data-structure storing all the decomposition, each solution for each value and at each step, the set of all solutions for a given value.

At this point, a check is performed to ensure that all possible ways of decomposing 12 using \mathcal{D}' is present in the tree. In our case, the decomposition of 6 in $[2, 2, 2]$ is added in each occurrence of 6. This ends the building phase. The resulting tree is reported in Figure 4.



■ **Figure 4** The first two levels of the tree represented in the table for $C_2^{12}x = C_6^{12}$, after the check of completeness.

After this first phase of construction of the tree, the aggregation of solutions starts. To find the solutions for the current node one must also take the Cartesian product of the solutions sets in the subtrees of the same color and then the union of the solution sets of nodes of different colors (different splits). For example, for $m = 12$ (*i.e.* the root node), the cartesian product between 6 and 6 is computed, but for $m = 6$ (in each occurrence) two cartesian operations and a union are necessary. Therefore, the final solution set for the node 12 is $\{C_3^4 + C_6^4, C_3^{12}, C_6^6, C_3^6 + C_6^3, C_6^2 + C_3^8, C_3^2 + C_6^5, C_6^1 + C_3^{10}\}$.

Although we can describe our algorithm with a pseudocode, and then we can sketch some proofs about its soundness, completeness and termination.

■ **Listing 1 Colored-Tree** - Complete algorithm for the enumeration problem.

```

1 procedure Colored-Tree(p, n, q):
2   // input 'p,q,n': the parameters of the equation
3   // enumerate all the solutions of the equation
4   node,splits,nodeSolution,SubTreeSolutions=[]
5   D=divisors(q)
6   node.add(n,1)
7   for i in node.length do
8     if (node[i]!=1) then
9       splits[i]=MCDA(node[i],D \ node[i])
10      generateNewNodes(splits[i])
11      SubTreeSolutions[i].add(nodeSolutions[i])
12    end
13  end
14  checkRepresented()
15  for i in node.length do
16    nodeSolution[i]=computeSingleSolution(node[i])
17  end
18  IncreaseOrder()
19  for i in node.length do
20    if (node[i]!=1) then
21      solutionsSplits=[]
22      for j in splits[i] do
23        solutionsSplits.add(cartesian(splits[i][j]))
24      end
25      SubTreeSolutions[i].add(union(solutionsSplits))
26    end
27  end
28  return SubTreeSolutions[node.length]

```

The Listing 1 presents the procedure using some particular functions:

- **generateNewNodes** adds the elements of the split, the node necessary in order to decompose but not yet represented as nodes in the nodes set.
- **MCDA** computes the optimal solutions of the making-change problem for a node value and a set of coins.
- **computeSingleSolution** returns the node solution for a node equation represented with a node.
- **checkRepresented** check if all the possible decompositions of the root are represented, otherwise add the corresponding sub-tree.
- **IncreaseOrder** permutes the row of the table in the increasing order according to the value of the nodes.

Now we can sketch some proofs about its soundness, completeness and termination.

► **Lemma 12 (Soundness).** *For all $n, p, q \in \mathbb{N}^*$, $S_{p,q}^n \subseteq T_{p,q}^n$.*

Proof. Let us prove the soundness by induction on the depth of the tree from leaves to root. Induction base: if there is only one step, we know by Lemma 3, that a solution found is feasible iff $\gcd(p, \frac{q}{p} \times m) = m$ and $\text{lcm}(p, \frac{q}{p} \times m) = q$, and because there is only one leaf in the base, we therefore obtain all the solution. Induction hypothesis: let us assume that we have all the solution possible at a depth n and let us show that we can obtain all the solution at a depth $n + 1$. Induction step: It is easy to see that a solution exists if and only if it comes from a decomposition. Thus, by performing a Cartesian product between the set of solutions at depth n (which is true by IH) and the node solution (which is true by

XX:10 Solving Equations on Discrete Dynamical Systems

Induction base, since the node can be seen as a leaf), we know that we will obtain all the solution coming from the possible decomposition in the sub-tree. If a solution is coming from another sub-tree, since we perform an exhaustive check where we assign a different color to the other sub-tree, we know again, by IH and because we are taking the union of all the solutions possible, that we have all the solution possible at a depth $n + 1$. ◀

► **Lemma 13** (Completeness). *For all $n, p, q \in \mathbb{N}^*$, $T_{p,q}^n \subseteq S_{p,q}^n$.*

Proof. Toward a contradiction. Let us assume that it exists a solution $r \in T_{p,q}^n$ and that $r \notin S_{p,q}^n$. This means that **colored-tree** method does not return it. This implies that it exists a decomposition of n , which leads to r , such that this decomposition is not in the tree. Contradiction, an exhaustive check is performed to assure that all the decompositions are there. Therefore, all solutions are returned. ◀

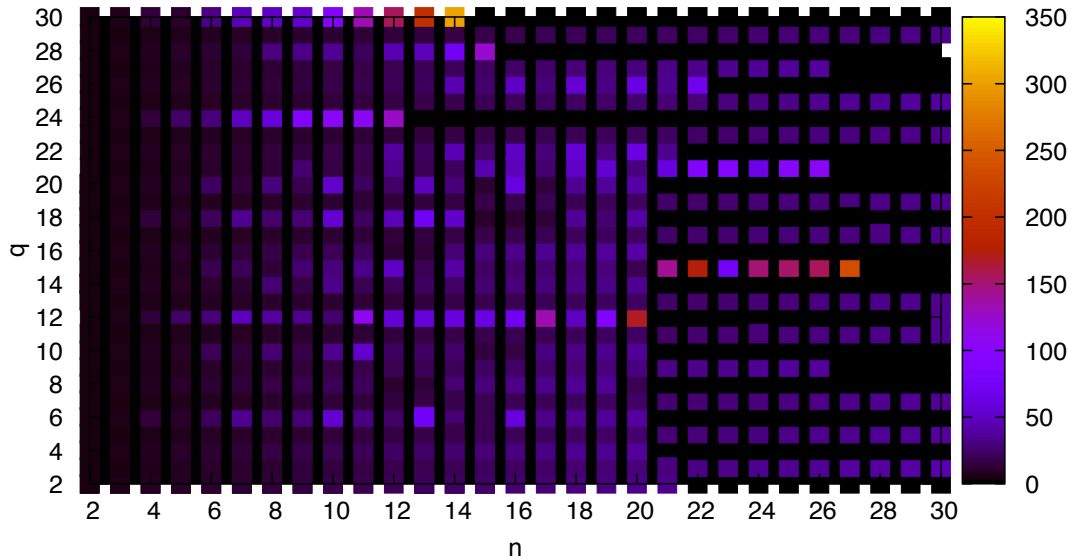
► **Lemma 14** (Termination). *The **colored-tree** method always terminates.*

Proof. The building phase always terminates since the colored-tree has maximal depth \mathcal{D}' and the number of different possible colors is bounded by 2^k where k is the size of the multiset containing n/p_i copies of the divisor p_i per each divisor in \mathcal{D}' . The aggregation phase always terminates since it performs a finite number of operations per each node of the colored tree. ◀

Now that we have defined the problem, its complexity and a sound and complete algorithm to solve it. It is time to experimentally evaluate it in order to study its scalability.

5 Experimental Evaluations

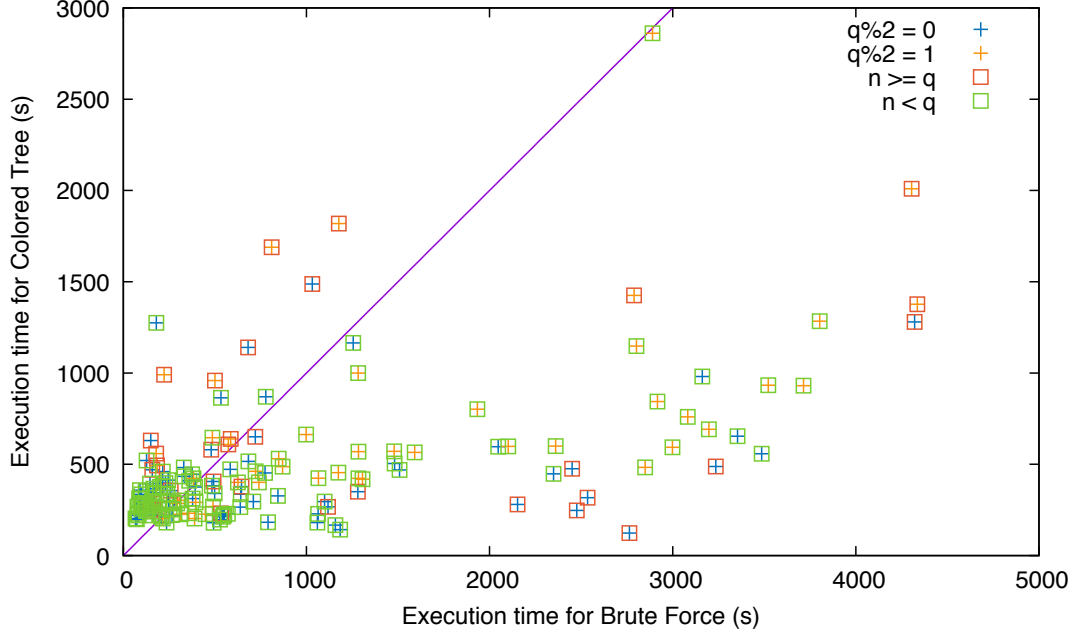
The **colored-tree** method provides a complete set of solutions of simple equations of type Equation 7. Its complexity can be experimentally measured counting the number of nodes in the colored tree.



■ **Figure 5** The number of nodes in the colored tree as a function of n and q .

Figure 5 shows how the complexity grows as a function of n and q . For this case, we set $p = q$ to ensure that we always have at least one solution and therefore a tree-decomposition.

Notice that, in some cases, the complexity is particularly high due to specific analytical relations between the input parameters that we are going to study in the future. Notice also that our method seems to have a weakness when q is an even number. This is easily explained: in many cases, all the divisors can be expressed by the other ones. Therefore the check that ensures that all the decompositions are present is particularly time- and memory-consuming.



■ **Figure 6** The brute force approach vs. `colored-tree` method *w.r.t.* execution time (in seconds).

Since there is no other competitor algorithm at the best of our knowledge, we compared the `colored-tree` method to a brute force algorithm. We test our algorithm with n from 1 to 20, p is also from 1 to 20 and at any time, $p = q$. Results are reported in Figure 6. As expected, the `colored-tree` method outperforms the brute force solution, sometimes with many orders of magnitude faster. However, when the input equation has small coefficients, the `colored-tree` method performs worse. This can be explained considering that building the needed data structures requires a longer time than the execution of the brute force algorithm.

6 Conclusion

Dennuzio *et al.* in [3] introduced a formalism to study polynomial equations over finite DDS as boolean automata networks used biological modelling for genetic regulatory networks and metabolic networks.

They argued that polynomial equations are a convenient tool for the analysis of the dynamics of a system. However, algorithmically solving such equations is an unfeasible task. In this article, we propose a practical way to partially overcome those difficulties using a couple of approaches which aims at studying separately the number of component (*i.e.* the number of attractors) and the length of their periods. This paper proposes an algorithm for the number of components of the solution of a polynomial equation over finite DDS.

One of the core routines of the algorithm uses a brute force check for the make-change

problem which clearly affects the overall performances. Therefore, a natural research direction consists in finding a better performing routine. Another way would be to parallelise the whole computation, indeed it seems to have a nice parallelization property. Another interesting research direction consists of better understanding the computational complexity of the DSECP.

We are still hard-working to improve the performances of the algorithm to have strong scalability properties in the perspective of providing a handy tool which can be exploited by bioinformaticians to actually solve the *Hypothesis Checking* problem in their context.

Acknowledgments

This work has been partly funded by IDEX UCA^{JEDI}.

References

- 1 Anna Adamaszek and Michal Adamaszek. Combinatorics of the change-making problem. *Eur. J. Comb.*, 31(1):47–63, 2010. URL: <https://doi.org/10.1016/j.ejc.2009.05.002>, doi:10.1016/j.ejc.2009.05.002.
- 2 Stefan Bornholdt. Boolean network models of cellular regulation: prospects and limitations. *Journal of The Royal Society Interface*, 5:85–94, 2008. URL: <http://doi.org/10.1098/rsif.2008.0132.focus>, doi:<http://doi.org/10.1098/rsif.2008.0132.focus>.
- 3 Alberto Dennunzio, Valentina Dorigatti, Enrico Formenti, Luca Manzoni, and Antonio E. Porreca. Polynomial equations over finite, discrete-time dynamical systems. In *Proc. of ACR18*, pages 298–306, 2018. URL: https://doi.org/10.1007/978-3-319-99813-8_27, doi:10.1007/978-3-319-99813-8_27.
- 4 Oscar Levin. *Discrete Mathematics: an open introduction*. <https://github.com/oscarlevin/discrete-book/>, 2017.
- 5 Henning S. Mortveit and Christian M. Reidys. *An Introduction to Sequential Dynamical Systems*. Universitext. Springer, 2008.
- 6 Sylvain Sené. *On the bioinformatics of automata networks*. Habilitation à diriger des recherches, Université d’Evry-Val d’Essonne, November 2012. URL: <https://tel.archives-ouvertes.fr/tel-00759287>.
- 7 Heike Siebert. Dynamical and structural modularity of discrete regulatory networks. In *COMPMOD*, volume 6 of *EPTCS*, pages 109–124, 2009.
- 8 Y. Strozecki. *Enumeration complexity and matroid decomposition*. PhD thesis, Université Paris Diderot - Paris 7, 2010.
- 9 Asli Ürgüplü. *Contributions to Symbolic Effective Qualitative Analysis of Dynamical Systems; Application to Biochemical Reaction Networks. (Contributions à l’analyse qualitative symbolique effective des systèmes dynamiques; l’application aux réseaux de réactions biochimiques)*. PhD thesis, Lille University of Science and Technology, France, 2010.