Society for
Mathematical
Biology

Geometric Characterization of Data Sets with Unique Reduced Gröbner Bases

Qijun He[2] · Elena S. Dimitrova[1] · Brandilyn Stigler[3] · Anyu Zhang[3]

Abstract

Model selection based on experimental data is an important challenge in biological data science. Particularly when collecting data is expensive or time-consuming, as it is often the case with clinical trial and biomolecular experiments, the problem of selecting information-rich data becomes crucial for creating relevant models. We identify geometric properties of input data that result in an unique algebraic model, and we show that if the data form a staircase, or a so-called linear shift of a staircase, the ideal of the points has a unique reduced Gröbner basis and thus corresponds to a unique model. We use linear shifts to partition data into equivalence classes with the same basis. We demonstrate the utility of the results by applying them to a Boolean model of the well-studied *lac* operon in *E. coli*.

Keywords Biological data science · Algebraic design of experiments · Gröbner bases · Ideals of points · Staircases of monomial ideals

Mathematics Subject Classification 14 · 92

1 Introduction

Developing predictive models from large-scale experimental data is an important problem in biological data science (Schatz 2015). In certain settings, such as when modeling biological switches (Dalchau et al. 2018), it is advantageous or even necessary to view the input data as discrete (Dimitrova et al. 2010). While there are many classes of functions that can fit data from an underlying network, data over a finite field can be fit

✉ Brandilyn Stigler
    bstigler@smu.edu

1   School of Mathematical and Statistical Sciences, Clemson University, Clemson, SC29634, USA

2   University of Virginia, Charlottesville, VA 22911, USA

3   Department of Mathematics, Southern Methodist University, Dallas, TX 75275, USA

by polynomials (Lidl et al. 1997). A useful consequence is that polynomial models of such data can be written in terms of a monomial basis, where each choice of monomial basis provides a different prediction regarding network structure (Jarrah et al. 2007).

More precisely, when the entries of a given data set are elements of a finite field, they can be viewed as an algebraic variety of a system of polynomials and the corresponding set of polynomials which vanish on the points in the variety is a *zero-dimensional ideal of points* (Cox et al. 1997). A popular computational tool for ideals of points is a *Gröbner basis* (GB), which is a generalization of echelon forms of linear systems (Abbott et al. 2000; Farr and Gao 2006; Dong 2016). Studying GBs, in particular the number of different GBs for a given ideal of points, has impact on any application which depends on GB computation for model construction.

While algebraic geometry literature is vast with regard to finding solutions of polynomial systems, to the best of our knowledge, the only result which connects directly geometric properties of points in a variety V to the uniqueness of reduced GBs of an ideal of points $I(V)$ for any monomial order is found in Dimitrova et al. (2019, Theorem 4.8(b)): it implies that an ideal of points $I(V)$ and the ideal of the set complement of the points $I(V^C)$ have the same number of reduced GBs and thus for any monomial order, $I(V)$ has a unique reduced GB if and only if $I(V^C)$ has a unique reduced GB. The uniqueness of reduced GBs has significance in the following two applications.

In Laubenbacher and Stigler (2004), Gröbner bases were applied to the problem of model selection in systems biology. They were introduced as a tool to select minimal models from a set of polynomial dynamical systems that fit discrete experimental data: for a given set of data points over a finite field, the ideal of points forms a coset representing the space of polynomial dynamical systems that fit the data and a minimal model is selected from the space by computing a reduced GB of the ideal and taking the normal forms of the model equations. While this provides an algorithmic solution to model selection, each choice of monomial order results in a different GB and thus, likely, in a different minimal polynomial dynamical system. To remedy this computational artifact, Dimitrova and Stigler (2014) proposed a systematic way of adding new data points to an existing data set to ensure that the ideal of points has a unique reduced GB, yielding a unique minimal model. They also gave an algebraic characterization of the smallest set of points that need to be added. The method, however, involves solving a large system of polynomial equations even for small data sets.

More recently, neural ideals were introduced in Curto et al. (2013) as an algebraic object that can be used to better understand the combinatorial structure of neural codes. A neural ideal has a special generating set, called its canonical form that encodes a minimal description of the so-called receptive field structure intrinsic to the neural code. Also, for a given monomial order, a neural ideal is generated by its (reduced) Gröbner basis with respect to that monomial order. It was shown in Garcia et al. (2018) that for small dimensions, Gröbner basis computations are faster than canonical form ones and it is thus desirable to be able to identify neural ideals whose canonical forms are Gröbner bases. They proceeded to show that this is the case exactly when the neural ideal has a unique Gröbner basis. However, there is still no known condition on the neural codes themselves that guarantees that the corresponding neural ideal has a unique Gröbner basis.

This paper addresses an important question in algebraic design of experiments: What properties should a discrete data set satisfy in order to uniquely identify an algebraic model? Since the possibility of having multiple models that correspond to a single data set arises from the fact that a polynomial ideal can have multiple reduced GBs, we approach this question by identifying properties of the input data that result in a unique reduced Gröbner basis for any monomial order.

For m data points and n variables, an upper bound for the number of distinct reduced Gröbner bases for an ideal of points is $O(m^{2n(n-1)/(n+1)})$ (Babson et al. 2003); however, in the finite field setting, this upper bound will be much smaller. From an algebraic geometry point of view, the problem is the following: find varieties V in $\mathbb{Z}_p^n$ whose ideals of points $I(V)$ have unique reduced GBs for any monomial order. Alternatively, the problem can be phrased as identifying varieties V for which the universal Gröbner basis of $I(V)$ is itself a reduced Gröbner basis, or for which the Gröbner fan of $I(V)$ consists of a single cone.

We define the concept of a *linear shift*, originally introduced in He (2016) for varieties and extended to rings in Dimitrova et al. (2019). We show that linear shifts are equivalence relations on data sets in $\mathbb{Z}_p^n$ (where $\mathbb{Z}_p$ is the field of integers modulo p) of fixed size. We use linear shifts to partition data into equivalence classes with the same basis. This impacts research in biological data science which uses discrete data, including data storage in the biomedical sciences (Adam et al. 2017) and model selection in functional genomics (Stigler and Chamberlin 2012).

In Sect. 3, we establish a geometric property of the points in $V \subseteq \mathbb{Z}_p^n$ which guarantees that the ideal of the points $I(V)$ has a unique reduced Gröbner basis, regardless of the monomial order: Corollary 3 to the main result, Theorem 2. Finally, we apply the main result to a Boolean model of the well-studied *lac* operon in *E. coli*.

2 Background

Much of the notation and formalization in this section is due to Cox et al. (1997).

Let k be a field and $n \in \mathbb{N}$. Let $I \subseteq k[x_1, \ldots, x_n]$ be an ideal and $LT_\prec(I)$ be the leading term ideal of I with respect to some monomial order $\prec$. Recall that the quotient ring $k[x_1, \ldots, x_n]/I$ is isomorphic to $span\{x^\alpha | x^\alpha \notin LT_\prec(I)\}$ as a k-vector space. In fact, for each choice of monomial order, $\{x^\alpha | x^\alpha \notin LT_\prec(I)\}$ forms a basis for $k[x_1, \ldots, x_n]/I$. Such monomials are called *standard* with respect to $\prec$; we denote the set $\{x^\alpha | x^\alpha \notin LT_\prec(I)\}$ as $SM_\prec(I)$. Note that standard monomials satisfy the following divisibility property: if $x^\alpha \in SM_\prec(I)$ and $x^\beta | x^\alpha$, then $x^\beta \in SM_\prec(I)$.

In the current setting, as all ideals are zero-dimensional, the quotient ring $k[x_1, \ldots, x_n]/I$ is finite dimensional as a vector space. Hence, the set $SM_\prec(I)$ of standard monomials associated with I is finite and can be represented as a *staircase*, which is a set $\lambda \subseteq \mathbb{N}^n$ of nonnegative integer vectors such that if $v \in \lambda$ and $u \leq v$ coordinate-wise, then $u \in \lambda$. Such staircases can be visualized on an integer lattice where a monomial is depicted via its exponent vector: $(m, n) \leftrightarrow x^m y^n$; see Fig. 1.

Finally, recall that if $\prec$ is a monomial order and $I \subseteq k[x_1, \ldots, x_n]$ is nonzero, then a subset $G = \{g_1, \ldots, g_t\}$ is a *Gröbner basis* for I with respect to $\prec$ if $\langle LT_\prec(g_1), \ldots, LT_\prec(g_t) \rangle = \langle LT_\prec(I) \rangle$. Furthermore, a Gröbner basis is *reduced* if
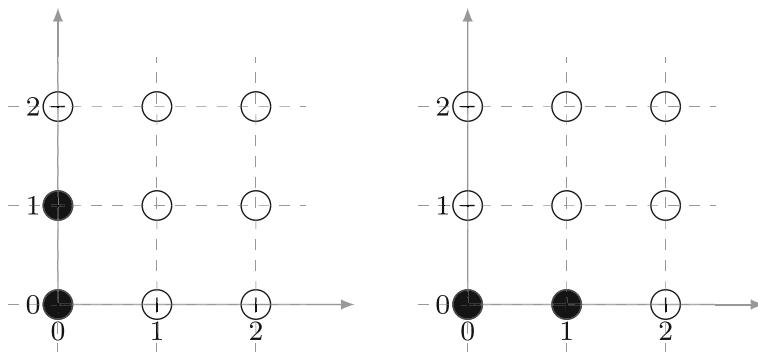
Fig. 1 Staircases of the ideal $I(V)$, where $V = \{(2, 0), (0, 1)\} \subseteq \mathbb{Z}_3^2$. For a given Gröbner basis, each diagram shows the interface between the standard monomials, which are represented by black dots, and the leading terms, which are represented by white dots. Here, $SM_1(I(V)) = \{1, y\}$ and $SM_2(I(V)) = \{1, x\}$

the leading coefficient of each element of the basis is 1 and no monomial in any element of the basis is in the ideal generated by the leading terms of the other elements of the basis. Given a fixed monomial order, an ideal I has a unique reduced Gröbner basis.

The union of all reduced Gröbner bases is *the universal Gröbner basis* of I, which is a Gröbner basis for every monomial order. We note that the universal GB may not be reduced when I has multiple reduced GBs associated with different monomial orders.

Example 1 Let $V = \{(0, 0), (1, 0), (2, 1)\} \subseteq \mathbb{Z}_3$ and $I(V) \subseteq \mathbb{Z}_3[x, y]$ be the ideal of polynomials that vanish on the points in V. Let $\prec_1$ denote the monomial order with weight vector $(1, 1)$ and $\prec_2$ the order with weight vector $(1, 3)$. Then, $I(V)$ has two distinct reduced Gröbner bases, $GB_{\prec_1}(I(V)) = \{y^2 - y, xy + y, x^2 - x + y\}$ and $GB_{\prec_2}(I(V)) = \{x^3 - x, y + x^2 - x\}$. So the universal Gröbner basis for $I(V)$ is $\mathcal{G} = \{y^2 - y, xy + y, x^2 - x + y\} \cup \{x^3 - x, y + x^2 - x\}$. This set has two different leading term ideals: $LT_{\prec_1}(I(V)) = \langle y^2, xy, x^2 \rangle$ and $LT_{\prec_2}(I(V)) = \langle x^3 - x, y + x^2 - x \rangle$. The associated standard monomial bases are $SM_{\prec_1}(I(V)) = \{1, x, y\}$ and $SM_{\prec_2}(I(V)) = \{1, x, x^2\}$.

We note that every ideal of points over a finite field of characteristic $p > 0$ has relations of the form $x^p - x$ for every variable. As such, the degrees of a variable in a polynomial are bounded above by $p - 1$, though polynomials may have high total degree due to many mixed terms.

Example 2 Consider a network of three nodes x, y, z with states in $\mathbb{Z}_3$, where 0 stands for low, 1 for medium, 2 for high. Suppose that z is regulated by x and y and its outputs are displayed in the table below for the input data in Example 1.

Time	x	y	$z = f(x, y)$
t_1	0	0	0
t_2	1	0	0
t_3	2	1	1

As computed above, we know that $I(V)$ has two distinct reduced GBs with corresponding standard monomial bases. Given each basis, f can be written as a linear combination $ax^{\alpha_1} + bx^{\alpha_2} + cx^{\alpha_3}$ of the basis elements. To find the coefficients, one can set up evaluation matrices $\mathbb{X}$ as follows:

(x, y)	1	x	y
(0,0)	1	0	0
(1,0)	1	1	0
(2,1)	1	2	1

and

(x, y)	1	x	x^2
(0,0)	1	0	0
(1,0)	1	1	1
(2,1)	1	2	1

Solving each linear system $\mathbb{X}[a\ b\ c]^T = [0\ 0\ 1]^T$ results in two candidate models for f which are compatible with the given data: $f = y$ and $f = x + 2x^2$.

Note that the first function model is regulated only by gene y but the second model is regulated only by gene x. This presents a problem for experimentalists in that the two models suggest different regulatory relationships. We therefore ask the question: Which data sets generate a unique model? That is, we are interested in varieties for which the universal Gröbner basis is itself a reduced GB and has a single leading term ideal (and therefore single standard monomial basis) regardless of which monomial order is chosen.

3 Geometric Properties of Data Sets Whose Ideals Have Unique Reduced Gröbner Bases

In this section, we give a geometric characterization of $V \subseteq \mathbb{Z}_p^n$ such that $I(V)$ has a unique reduced Gröbner basis, regardless of the monomial order. Unless otherwise stated, definitions and results in this section are found in He (2016).

In Sect. 2, we recalled that exponent vectors of standard monomials of an ideal I form a staircase since any divisor of a standard monomial is again a standard monomial. Such a staircase is called *initial* (Babson et al. 2003); for example, the staircases in Fig. 1 are initial. Furthermore, a staircase λ is *basic* for I if the congruence classes modulo I of the monomials x^v with $v \in \lambda$ form a vector space basis for the quotient space $\mathbb{Z}_p[x_1, \ldots, x_n]/I$ (Babson et al. 2003).

If λ is basic, then the class $[f] = f + I$ of any $f \in \mathbb{Z}_p[x_1, \ldots, x_n]$ can be uniquely represented as a linear combination of elements in $\{x^v \mid v \in \lambda\}$. For a given monomial order, any polynomial $f \in \mathbb{Z}_p[x_1, \ldots, x_n]$ has a unique normal form with respect to I. Hence, an initial staircase of an ideal I is basic. As it will be shown in Theorem 1, we can determine whether a set $\lambda \in \mathbb{Z}_p^n$ is basic by checking the invertibility of the evaluation matrix defined next.

Definition 1 Let $\lambda = \{u^1, \ldots, u^r\}$ be an r-subset of $\mathbb{Z}_p^n$ and let $V = \{v^1, \ldots, v^s\}$ be an s-subset of $\mathbb{Z}_p^n$. The *evaluation matrix* $\mathbb{X}(x^\lambda, V)$ is the s-by-r matrix whose element in position (i, j) is $x^{u^j}(v^i)$, the evaluation of x^{u^j} at v^i.

Example 3 Let $\lambda_1 = \{(0, 0), (1, 0)\}$, $\lambda_2 = \{(0, 0), (0, 1)\}$, and $V = \{(2, 0), (0, 1)\}$ be subsets of $\mathbb{Z}_3^2$. Then, $\mathbb{X}(x^{\lambda_1}, V) = \begin{bmatrix} 1 & 2 \\ 1 & 0 \end{bmatrix}$ and $\mathbb{X}(x^{\lambda_2}, V) = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$.

The following result from Babson et al. (2003) illustrates a connection between basic sets and evaluation matrices, which will be used below.

Theorem 1 (Babson et al. 2003) *Let λ and V be subsets of $\mathbb{Z}_p^n$. Then, λ is basic for $I(V)$ if and only if $\mathbb{X}(x^\lambda, V)$ is invertible.*

Example 4 Let $\lambda_1 = \{(0, 0), (1, 0)\}$, $\lambda_2 = \{(0, 0), (0, 1)\}$, and $V = \{(0, 0), (1, 0)\}$ be subsets of $\mathbb{Z}_3^2$. Then, λ_1 is basic for $I(V)$ since $\mathbb{X}(x^{\lambda_1}, V) = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$ is invertible; however, λ_2 is not basic for $I(V)$ since $\mathbb{X}(x^{\lambda_2}, V) = \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix}$ is not invertible.

Notice that an initial staircase must be basic, while a basic staircase might not be initial; however, if $I(V)$ has a unique initial staircase (and thus a unique reduced Gröbner basis), then $I(V)$ has a unique basic staircase.

Proposition 1 *An ideal $I(V)$ has a unique initial staircase if and only if $I(V)$ has a unique basic staircase.*

Proof We recall that the initial ideals of a polynomial ideal are in a one-to-one correspondence with its reduced Gröbner bases. The result then follows directly from Proposition 2.2 in Babson et al. (2003) and the fact that for every two monomials x^α, x^β with $x^\alpha \nmid x^\beta$, there exists a weight vector γ and monomial order $\prec_\gamma$ such that $x^\beta \prec_\gamma x^\alpha$. □

Based on Proposition 1, if we want to find out whether $I(V)$ has a unique reduced Gröbner basis, we just need to check whether $I(V)$ has a unique basic staircase. In other words, we can check if there exist a unique staircase $\lambda \subseteq \mathbb{Z}_p^n$ such that $\mathbb{X}(x^\lambda, V)$ is invertible.

Next, we present a sufficient condition for $I(V)$ to have a unique reduced Gröbner basis. A construction that will aid in proving this condition is that of a *layer* of a staircase.

Definition 2 Given a staircase $\lambda = \{(u_1, \ldots, u_n) : u_i \in \mathbb{N}\}$, the *ith layer* of λ with respect to the *j*th coordinate is the subset $\{u \in \lambda : u_j = i\} \subseteq \lambda$. Let ℓ be the largest integer such that $\{u \in \lambda : u_j = \ell\} \neq \emptyset$. The *height* of λ in the *j*th coordinate is defined to be $\ell + 1$, denoted as $h_j(\lambda)$.

Note that we identify a layer of a staircase in $\mathbb{N}^n$ with a staircase in $\mathbb{N}^{n-1}$.

Proposition 1 prompts the following natural question. If $\lambda \subseteq \mathbb{Z}_p^n$ is a staircase, which subsets $V \subseteq \mathbb{Z}_p^n$ have λ as their *unique* basic staircase? Let us consider the case when V is itself a staircase.

Theorem 2 *Let λ and V be two staircases in $\mathbb{Z}_p^n$. Then, λ is basic for $I(V)$ if and only if $\lambda = V$.*

Proof We will prove this by induction on the number of variables.

Let $n = 1$. To prove the necessary condition, suppose $\lambda \neq V$. Since there is only one variable, we must have $|\lambda| \neq |V|$. Therefore, $\mathbb{X}(x^\lambda, V)$ is not invertible since it is not a square matrix. By Theorem 1, λ is not basic. For the sufficient condition, assume $\lambda = V$. Then, $\mathbb{X}(x^\lambda, V)$ is a square Vandermonde matrix. Since V is a set of distinct points, $\mathbb{X}(x^\lambda, V)$ is invertible. By Theorem 1, λ is basic.

Assume the inductive hypothesis holds for $n = k$ and consider $n = k + 1$. We will prove the inductive step by induction on the height of the monomial staircase with respect to the first coordinate.

Consider the base case when $h_1(\lambda) = 1$; in other words, for all $u \in \lambda$, we have $u_1 = 0$. That is to say all monomials of x^λ do not involve x_1.

($\Leftarrow$) Suppose $\lambda = V$. Then, the first coordinate of any point in V is 0. Therefore, λ and V are staircases with one fewer variable, x_1. Based on our inductive hypothesis for $n = k$, we have x^λ are basic monomials of $I(V)$.

($\Rightarrow$) Suppose $\lambda \neq V$. Then, we consider three cases.

C1: $|\lambda| \neq |V|$. In this case, $\mathbb{X}(x^\lambda, V)$ is not invertible as it is not square.
C2: $|\lambda| = |V|$ and $h_1(V) \geq 2$. In this case, $\mathbb{X}(x^\lambda, V)$ is not invertible as at least two rows are the same.
C3: $|\lambda| = |V|$ and $h_1(V) = h_1(\lambda) = 1$, while $\lambda \neq V$. In this case, the first coordinate of any point in λ and V is 0. In other words, λ and V are staircases with one fewer variable, x_1. Based our inductive hypothesis for $n = k$, we have that $\mathbb{X}(x^\lambda, V)$ is not invertible.

In each case, the evaluation matrix is not invertible and so λ is not basic, concluding the base case of $h_1(\lambda) = 1$.

Assume the inductive hypothesis holds for all monomial staircases λ with $1 \leq h_1(\lambda) \leq d$. Let us consider a staircase λ with $h_1(\lambda) = d + 1$.

($\Leftarrow$) Suppose $\lambda = V$. Let $\lambda_0 := \{u \in \lambda : u_1 = 0\}$ denote the zeroth layer of λ and $V_0 := \{v \in \lambda : v_1 = 0\}$ denote the zeroth layer of V with respect to the first coordinate. Since $\lambda = V$, we have $\lambda_0 = V_0$. By inductive hypothesis, the evaluation matrix $\mathbb{X}(x^{\lambda_0}, V_0)$ is invertible. Now, let us consider the evaluation matrix $\mathbb{X}(x^\lambda, V)$. We can reorder rows and columns of $\mathbb{X}(x^\lambda, V)$, so that $\mathbb{X}(x^{\lambda_0}, V_0)$ appears as the upper left submatrix of $\mathbb{X}(x^\lambda, V)$. Since the upper left submatrix of $\mathbb{X}(x^\lambda, V)$ is invertible, after elementary row and column operations, we can transform $\mathbb{X}(x^\lambda, V)$ into a block matrix of the form

$$\begin{bmatrix} I & 0 \\ 0 & \mathbb{X}(x^{\lambda \backslash \lambda_0}, V \backslash V_0) \end{bmatrix}.$$

Moreover, x_1 divides all monomials in $x^{\lambda \backslash \lambda_0}$ and for any point $v \in V \backslash V_0$, we have $v_1 \neq 0$. Therefore, $\mathbb{X}(x^{\lambda \backslash \lambda_0}, V \backslash V_0)$ is invertible if and only if $\mathbb{X}(\frac{x^{\lambda \backslash \lambda_0}}{x_1}, V \backslash V_0)$ is invertible. Note that $\frac{x^{\lambda \backslash \lambda_0}}{x_1}$ corresponds to a staircase with height at most d and $V \backslash V_0$ is a linear shift of the same staircase, so X^* is invertible by the inductive hypothesis and Theorem 3. Hence, the original evaluation matrix $\mathbb{X}(x^\lambda, V)$ is also invertible; thus, λ is basic.

($\Rightarrow$) Suppose $\lambda \neq V$. If $|\lambda| \neq |V|$, then $\mathbb{X}(x^\lambda, V)$ is not invertible since it is not a square matrix.

Assume $|\lambda| = |V|$. We can reorder rows and columns so that $\mathbb{X}(x^\lambda, V)$ appears in the form

$$\begin{bmatrix} \mathbb{X}(x^{\lambda_0}, V_0) & 0 \\ A & \mathbb{X}(x^{\lambda \setminus \lambda_0}, V \setminus V_0) \end{bmatrix}.$$

Note that the row space of A is a subspace of the row space of $\mathbb{X}(x^{\lambda_0}, V_0)$.

Suppose $\mathbb{X}(x^{\lambda_0}, V_0)$ is not a square matrix. We then consider two cases.

C1: If $\mathbb{X}(x^{\lambda_0}, V_0)$ has more rows than columns, then the rows of $[\mathbb{X}(x^{\lambda_0}, V_0), 0]$ are linearly dependent.

C2: If $\mathbb{X}(x^{\lambda_0}, V_0)$ has more columns than rows, then the columns of $[\mathbb{X}(x^{\lambda_0}, V_0), A]^T$ are linearly dependent.

In either case, $\mathbb{X}(x^{\lambda_0}, V_0)$ is not invertible.

If $\mathbb{X}(x^{\lambda_0}, V_0)$ is a square matrix but $\lambda_0 \neq V_0$, then $\mathbb{X}(x^{\lambda_0}, V_0)$ is not invertible by the inductive hypothesis. So $\mathbb{X}(x^\lambda, V)$ is not invertible.

If $\mathbb{X}(x^{\lambda_0}, V_0)$ is a square matrix and $\lambda_0 = V_0$, then we must have $\lambda \setminus \lambda_0 \neq V \setminus V_0$. Note that $\mathbb{X}(x^{\lambda \setminus \lambda_0}, V \setminus V_0)$ is invertible if and only if $\mathbb{X}(\frac{x^{\lambda \setminus \lambda_0}}{x_1}, V \setminus V_0)$ is invertible. Since $\lambda \setminus \lambda_0 \neq V \setminus V_0$, $\frac{x^{\lambda \setminus \lambda_0}}{x_1}$ corresponds to a staircase with height at most d, and $V \setminus V_0$ is a linear shift of some other staircase, $\mathbb{X}(\frac{x^{\lambda \setminus \lambda_0}}{x_1}, V \setminus V_0)$ is not invertible by the inductive hypothesis and Theorem 3. Therefore, $\mathbb{X}(x^\lambda, V)$ is also not invertible.

Since the evaluation matrix is not invertible, λ is basic.

Hence, the inductive step holds for $h_1(\lambda) = d + 1$, thereby concluding the inductive proof on the height of λ. Completing the nested inductive proof completes the inductive step for $n = k + 1$. Therefore, the original statement holds for all $n \in \mathbb{N}$. □

Corollary 1 *If $V \subseteq \mathbb{Z}_p^n$ is a staircase, then $I(V)$ has a unique reduced Gröbner basis for any monomial order.*

Proof It follows immediately from Theorem 2. □

Example 5 The set $V = \{(0, 0), (0, 1), (1, 0)\}$ is a staircase in $\mathbb{Z}_3^2$. Its corresponding ideal $I(V)$ has a unique reduced Gröbner basis $\mathcal{G} = \{y^2 - y, xy, x^2 - x\}$ with respect to any monomial order.

Before we state the main theorem of this section, we introduce an equivalence relation on the subsets of $\mathbb{Z}_p^n$ that is useful in the context of staircases and Gröbner bases.

Definition 3 For $V_1, V_2 \subseteq \mathbb{Z}_p^n$ with $|V_1| = |V_2|$, we say that V_1 is a *linear shift* of V_2, denoted $V_1 \overset{L}{\sim} V_2$, if there exists $\phi = (\phi_1, \ldots, \phi_n) : \mathbb{Z}_p^n \to \mathbb{Z}_p^n$ such that $V_1 = \phi(V_2)$ and $\phi_i : \mathbb{Z}_p \to \mathbb{Z}_p$ is defined coordinate-wise as $\phi_i(x_i) = a_i x_i + b_i$ for some $a_i \in (\mathbb{Z}_p)^\times$ and $b_i \in \mathbb{Z}_p$ for $i = 1, \ldots, n$.

It is straightforward to see that linear shift is a bijection between two point sets and defines an equivalence relation on subsets of $\mathbb{Z}_p^n$ of the same size.

Example 6 Consider $V_1, V_2, V_3 \subseteq \mathbb{Z}_3^2$, where $V_1 = \{(0, 0), (0, 1)\}$, $V_2 = \{(1, 1), (1, 2)\}$, and $V_3 = \{(1, 1), (2, 2)\}$. Then, $V_1 \overset{L}{\sim} V_2$ since $V_1 = \phi(V_2)$, where $\phi = (x + 1, x + 1)$. On the other hand, $V_1 \overset{L}{\not\sim} V_3$ since the first coordinates of the points in V_1 are the same, while in V_3 they are different.

We will see next that linear shifts preserve standard monomial bases.

Theorem 3 *If $V_1, V_2 \subseteq \mathbb{Z}_p^n$ and $V_1 \overset{L}{\sim} V_2$, then $I(V_1)$ and $I(V_2)$ have the same leading term ideals and thus the same standard monomial bases.*

Proof Let $V_1 \overset{L}{\sim} V_2$. Then, there is a permutation $\phi(x_1, \ldots, x_n) = (a_1 x_1 + b_1, \ldots, a_n x_n + b_n)$ for some $a_i, b_i \in \mathbb{Z}_p$ with $a_i \neq 0$ for all i, such that $\phi(V_1) = V_2$. Observe that

$$I(V_2) = \{f : f(v) = 0, \forall v \in V_2\}$$
$$= \{f : f(\phi(u)) = (f \circ \phi)(u) = 0, \forall u \in V_1\}.$$

Thus, for any $f \in I(V_2)$, we have $f \circ \phi \in I(V_1)$. Since V_2 is a linear shift of V_1 via ϕ, $(f \circ \phi)(x_1, \ldots, x_n) = f(\phi(x_1, \ldots, x_n)) = f(a_1 x_1 + b_1, \ldots, a_n x_n + b_n)$. So f and $f \circ \phi$ have the same leading monomial with respect to any monomial order $\prec$. Therefore, $LT_\prec(I(V_2)) \subseteq LT_\prec(I(V_1))$. We can show $LT_\prec(I(V_1)) \subseteq LT_\prec(I(V_2))$ by replacing ϕ with ϕ^{-1}. Hence, $LT_\prec(I(V_1)) = LT_\prec(I(V_2))$ with respect to any monomial order. $\square$

Corollary 2 *If $V_1, V_2 \subseteq \mathbb{Z}_p^n$ and $V_1 \overset{L}{\sim} V_2$, then $I(V_1)$ and $I(V_2)$ have the same number of reduced Gröbner bases. In particular, for any monomial order, $I(V_1)$ has a unique reduced Gröbner basis if and only if $I(V_2)$ has a unique reduced Gröbner basis.*

Proof Due to the one-to-one correspondence between initial ideals and reduced Gröbner bases, Theorem 3 implies that $I(V_1)$ and $I(V_2)$ have the same number of reduced Gröbner bases. $\square$

Corollary 3 *If $V \subseteq \mathbb{Z}_p^n$ is a linear shift of a staircase, then $I(V)$ has a unique reduced Gröbner basis for any monomial order.*

Proof This follows from Theorems 2 and 3. $\square$

Example 7 Consider subsets $V_1 = \{(0, 0), (0, 1), (1, 0)\}$ and $V_2 = \{(0, 1), (0, 2), (2, 2)\}$ of $\mathbb{Z}_3^2$. Notice that V_1 is a staircase, while V_2 is not. As $V_2 = \phi(V_1)$, where $\phi = (2x, 2x + 2)$, we have that $V_1 \overset{L}{\sim} V_2$. We see that $I(V_2)$ has a unique reduced Gröbner basis $\mathcal{G} = \{y^2 - 1, xy + x, x^2 + x\}$.

A set of points being a linear shift of a staircase is not a necessary condition for the corresponding ideal of points to have a unique reduced Gröbner basis, as is shown in the following example.

Example 8 Consider the set $V = \{(0, 0, 0), (1, 0, 0), (1, 1, 0), (1, 1, 1)\}$. Since each function in the component-wise definition of a linear shift ϕ is of the form $\phi_i(x_i) = x_i$ or $x_i + 1$ for $i = 1, 2, 3$, there are only 2^3 possible linear shifts, and by inspection, we see that when applied to V none of them results in a staircase. Therefore, V is not a linear shift of a staircase in $\mathbb{Z}_2^3$. However, using Theorem 3.3(b) in Dimitrova et al. (2019), for example, we see that $I(V)$ has this single reduced Gröbner basis $\mathcal{G} = \{z^2 + z, yz + z, y^2 + y, xz + z, xy + y, x^2 + x\}$ regardless of the choice of monomial order.

We conclude this section by noting that $V \subseteq \mathbb{Z}_2^n$ is a simplicial complex if and only if it is a staircase. Thus, Corollaries 1 and 3 generalize Proposition 4.2 in Garcia et al. (2018) for an arbitrary finite field and linear shift.

4 Applications

In this section, we apply our main results to a small Boolean model of the well-studied *lac* operon and highlight a possible improvement on the time to compute multiple Gröbner bases.

4.1 Design of Experiments in the Lac Operon

The *lac* operon is a system of genes which control the transport and metabolism of lactose in many bacteria including *E. coli*. While there are numerous models for the *lac* operon (see, for example, Santillán 2008; Goodwin 1963; Ozbudak et al. 2004; Wong et al. 1997), we consider a Boolean model proposed in (Veliz-Cuba and Stigler 2011). There the authors reduced the system to a core subnetwork consisting of the following four variables: M representing *lac* mRNA, L intercellular lactose, L_e extracellular lactose, and G_e extracellular glucose. The Boolean model for this subnetwork is given by the following Boolean functions, where extraneous variables are introduced to capture intermediate values (L_m, L_{em}) of lactose inside and outside of the cell, respectively: see Section 4.2.2 in Veliz-Cuba and Stigler (2011) for a full description of the model.

$$f_M = \neg G_e \wedge (L \vee L_m)$$
$$f_L = M \wedge L_e \wedge \neg G_e$$
$$f_{L_m} = ((L_{em} \wedge M) \vee L_e) \wedge \neg G_e$$
$$f_{L_e} = L_e$$
$$f_{G_e} = G_e$$
$$f_{L_{em}} = L_{em}$$

For the sake of illustrating the utility of the above results, we reduce this model to only include the four essential variables. To this end, we replace L_{em} with L_e and L_m

with L, and remove all instances of L_{em} and L_m via substitution. Doing so produces

$$f_{L_m} = ((L_e \wedge M) \vee L_e) \wedge \neg G_e = L_e \wedge \neg G_e$$

which we substitute into the function f_M:

$$f_M = \neg G_e \wedge (L \vee (L_e \wedge \neg G_e)) = \neg G_e \wedge (L \vee L_e).$$

This results in the following Boolean network on four variables, with wiring diagram depicted in Fig. 2:

$$f_M = \neg G_e \wedge (L \vee L_e)$$
$$f_L = M \wedge L_e \wedge \neg G_e$$
$$f_{L_e} = L_e$$
$$f_{G_e} = G_e.$$

Boolean functions can be rewritten as polynomial functions over $\mathbb{Z}_2$ using the following translations: the Boolean expression $x \vee y$ can be represented as the polynomial $x + y + xy$, $x \wedge y$ as xy, and $\neg x$ as $x + 1$. Applying these rules to the above functions yields the finite dynamical system $f : \mathbb{Z}_2^4 \to \mathbb{Z}_2^4$ where $f = (f_{x_1}, f_{x_2}, f_{x_3}, f_{x_4})$ and each f_{x_i} is a polynomial in the variables $x_1 := M$, $x_2 := L$, $x_3 := L_e$, and $x_4 := G_e$.

$$f_{x_1} = x_2 x_3 x_4 + x_2 x_3 + x_2 x_4 + x_3 x_4 + x_2 + x_3$$
$$f_{x_2} = x_1 x_3 x_4 + x_1 x_3$$
$$f_{x_3} = x_3$$
$$f_{x_4} = x_4 \tag{1}$$

Consider the first component of the state space of $f = (f_{x_1}, f_{x_2}, f_{x_3}, f_{x_4})$ in Fig. 3:

$$C_1 = \{\{0,0,0,0\}, \{0,1,0,0\}, \{1,0,0,0\}, \{1,1,0,0\}\}.$$

Fig. 2 Wiring diagram for a simplified Boolean model of the *lac* operon in *E. coli*. Directed edges with pointed ends indicate positive regulation, while directed edges with round ends indicate negative regulation. The variables G_e and L_e regulate the operon from outside the cell, represented by a rectangle around M and L
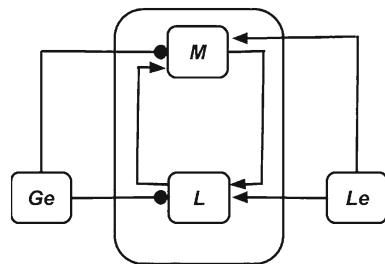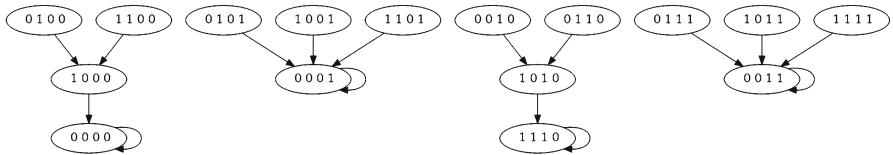
Fig. 3 State space graph for the four-dimensional finite dynamical system given by $f = (f_{x_1}, f_{x_2}, f_{x_3}, f_{x_4})$. Each node is a state (M, L, L_e, G_e) of the network, and a directed edge from state a to state b indicates that $f(a) = b$. Figure created using Hosny et al. (2016)

Note that the data points in C_1 form a staircase. By Corollary 3, the ideal $I(C_1)$ has a unique reduced Gröbner basis for any monomial order, namely

$$G_1 = \{x_1^2 + x_1, x_2^2 + x_2, x_3, x_4\}.$$

In particular, the data set C_1 has the unique leading term ideal $L = \langle x_1^2, x_2^2, x_3, x_4 \rangle$ and standard monomial basis $S = \{1, x_1, x_2, x_1 x_2\}$ for any monomial order.

If we label the other components similarly,

$$C_2 = \{\{0, 0, 0, 1\}, \{0, 1, 0, 1\}, \{1, 0, 0, 1\}, \{1, 1, 0, 1\}\},$$
$$C_3 = \{\{0, 0, 1, 0\}, \{0, 1, 1, 0\}, \{1, 0, 1, 0\}, \{1, 1, 1, 0\}\},$$
$$C_4 = \{\{0, 0, 1, 1\}, \{0, 1, 1, 1\}, \{1, 0, 1, 1\}, \{1, 1, 1, 1\}\},$$

we find that they are linear shifts of C_1, that is, $C_1 \overset{\phi_{12}}{\sim} C_2$, $C_1 \overset{\phi_{13}}{\sim} C_3$, and $C_1 \overset{\phi_{14}}{\sim} C_4$, where

$$\phi_{12} = (x_1, x_2, x_3, x_4 + 1)$$
$$\phi_{13} = (x_1, x_2, x_3 + 1, x_4)$$
$$\phi_{14} = (x_1, x_2, x_3 + 1, x_4 + 1).$$

According to Corollary 2, the data sets C_2, C_3, and C_4 have the same leading term ideal and standard monomial basis as C_1. So each of C_2, C_3, and C_4 also has a unique reduced Gröbner basis. As such, each of the four data sets will identify a unique polynomial involving the monomials in $S = \{1, x_1, x_2, x_1 x_2\}$. The advantage here is that each data set produces only one model; however, since the polynomials in Eq. (1) involve other monomials, we see that none of these data sets would correctly infer the original model.

4.2 Using Linear Shifts to Compute Gröbner Bases

In Dimitrova et al. (2007) and Dimitrova and Stigler (2014), the authors highlight the significance of being able to compute *all* reduced Gröbner bases, as each one potentially corresponds to a distinct model. Here, we show how to use linear shifts to compute other Gröbner bases given one. For example, given the Gröbner basis G_1 and linear shift functions $\phi_{12}, \phi_{13}, \phi_{14}$ from the previous section, we can directly apply

the linear shift functions to produce the generators of the other reduced Gröbner bases explicitly, rather than computing them from the respective ideals.

$$G_2 = GB(I(C_2)) = \{x_1^2 + x_1, x_2^2 + x_2, x_3, x_4 + 1\},$$
$$G_3 = GB(I(C_3)) = \{x_1^2 + x_1, x_2^2 + x_2, x_3 + 1, x_4\},$$
$$G_4 = GB(I(C_4)) = \{x_1^2 + x_1, x_2^2 + x_2, x_3 + 1, x_4 + 1\}.$$

While algorithms (and their corresponding complexities) related to the above theoretical results are not in the scope of the presented work, we close with a note about its potential to greatly reduce the time to compute Gröbner fans of zero-dimensional ideals. The worst-case complexity of computing one Gröbner basis of a zero-dimensional ideal in a general setting is quadratic in the number of variables n and cubic in the number of points m, that is $\mathcal{O}(nm^3 + n^2m^2)$ (Abbott et al. 2000), with various improvements in specialized settings. Computing a Gröbner fan for a zero-dimensional ideal from a given Gröbner basis is proved to be "a polynomial-time algorithm in the size of the output" (Fukuda et al. 2007). In settings where data sets yield Gröbner fans with distinct cones, we can take advantage of linear shifts: from one data set and its calculated fan (set of reduced Gröbner bases), use the linear shift functions to produce the reduced Gröbner bases for the ideals of the linearly shifted points. We expect that finding linear shifts between data sets will have smaller complexity than computing multiple Gröbner fans; this analysis will be performed in future work.

5 Discussion and Future Work

We addressed the problem of characterizing data sets which correspond to ideals with a unique reduced Gröbner basis with respect to all monomial orders. Our results fill in important theoretical gaps in the use of polynomial dynamical systems as models in systems biology by providing a criterion for determining whether a set of data will give rise to a unique set of predictions (i.e., a unique model) without having to compute all reduced Gröbner bases associated with the data. Furthermore, this work decreases the computational cost of modeling using polynomial dynamical systems and has the potential to reduce the financial cost in collecting experimental data by identifying data sets that result in unambiguous hypotheses for future testing. This is especially important in experimental settings since it is common for the number of observations (m) to be much smaller than the number of variables (n) in the system being studied.

For example, the set $\{\{0, 0, 1, 0\}, \{0, 1, 0, 1\}, \{1, 0, 0, 1\}, \{1, 1, 0, 0\}, \{1, 1, 1, 1\}\}$ of five data points in $\mathbb{Z}_2$ has 13 distinct reduced Gröbner bases and 13 corresponding standard monomial bases. If these are considered as inputs into the *lac* operon as described in Sect. 4.1, these standard monomial bases give rise to 4 distinct functions for x_1, 7 for x_2, 3 for x_3, and 5 for x_4, totaling 420 different finite dynamical systems, each with a different set of predictions that must be tested. The number of points (new experiments) that must be added to the existing data set to guarantee a unique model is an astonishing six: that is, the data set must be more than doubled! Instead, our work identifies candidate sets that are linear shifts of a staircase and have unique bases. (In

fact, there are close to 600 sets with unique GBs out of nearly 4400 sets of size five; though there are many such sets, the chances of successfully picking such a set at random is about 14%.)

We note that partitioning data sets of same size into equivalence classes under the linear shift equivalence relation leaves many interesting question to explore. For example, knowing the number and size of equivalence classes as well as the distribution of the number of Gröbner bases across equivalence classes would provide valuable information for design of experiments and model inference.

Finally, while we were able to prove a sufficient condition for a set of points to have an ideal with a unique reduced Gröbner basis (Corollary 3), the condition is not necessary. A necessary and sufficient condition on the set of points which guarantees a unique reduced Gröbner basis for any monomial order would be of great value as it could lead to an algorithm for generating all such sets. These results are likely to also benefit other areas that use Gröbner bases such as neural ideals computation.

References

Abbott J, Bigatti A, Kreuzer M, Robbiano L (2000) Computing ideals of points. J Symb Comput 30(4):341–356

Adam NR, Wieder R, Ghosh D (2017) Data science, learning, and applications to biomedical and health sciences. Ann N Y Acad Sci 1387(1):5–11

Babson E, Onn S, Thomas R (2003) The Hilbert zonotope and a polynomial time algorithm for universal Gröbner bases. Adv Appl Math 30(3):529–544

Cox D, Little J, O'Shea D (1997) Ideals, varieties, and algorithms. Springer, New York

Curto C, Itskov V, Veliz-Cuba A, Youngs N (2013) The neural ring: an algebraic tool for analyzing the intrinsic structure of neural codes. Bull Math Biol 75:1571–1611

Dalchau N, Szép G, Hernansaiz-Ballesteros R, Barnes CP, Cardelli L, Phillips A, Csikász-Nagy A (2018) Computing with biological switches and clocks. Nat Comput 17(4):761–779

Dimitrova E, Stigler B (2014) Data identification for improving gene network inference using computational algebra. Bull Math Biol 76(11):2923–2940

Dimitrova ES, Jarrah AS, Laubenbacher R, Stigler B (2007) A Gröbner fan method for biochemical network modeling. In: Proceedings of 2007 international symposium on symbolic and algebraic computation. ACM, pp 122–126

Dimitrova ES, Licona MP, McGee J, Laubenbacher R (2010) Discretization of time series data. J Comput Biol 17(6):853–868

Dimitrova ES, He Q, Robbiano L, Stigler B (2019) Small Gröbner fans of ideals of points. J Algebra Appl. https://doi.org/10.1142/S0219498820500875

Dong T (2016) A two-dimensional improvement for Farr–Gao algorithm. J Syst Sci Complex 29:1382–1399

Farr JB, Gao S (2006) Computing Gröbner bases for vanishing ideals of finite sets of points. In: Fossorier MPC, Imai H, Lin S, Poli A (eds) Applied algebra, algebraic algorithms and error-correcting codes. AAECC 2006. Lecture notes in computer science, vol 3857. Springer, Berlin, Heidelberg, pp 118–127

Fukuda F, Jensen A, Thomas R (2007) Computing Gröbner fans. Math Comput 76(260):2189–2212

Garcia R, García-Puente LD, Kruse R, Liu J, Miyata D, Petersen E, Phillipson K, Shiu A (2018) Gröbner bases of neural ideals. Int J Algebra Comput

Goodwin B (1963) Temporal organization in cells. Academic Press, Cambridge

He Q (2016) Algebraic geometry arising from discrete models of gene regulatory networks. Ph.D. thesis, Clemson University

Hosny A, Vera-Licona P, Laubenbacher R, Favre T (2016) AlgoRun: a Docker-based packaging system for platform-agnostic implemented algorithms. Bioinformatics 32(15):2396–2398. http://algorun.org/

Jarrah AS, Laubenbacher R, Stigler B, Stillman M (2007) Reverse-engineering of polynomial dynamical systems. Adv Appl Math 39(4):477–489

Laubenbacher R, Stigler B (2004) A computational algebra approach to the reverse engineering of gene regulatory networks. J Theor Biol 229(4):523–537

Lidl R, Niederreiter H, Rota G-C (1997) Finite fields: encyclopedia of mathematics and its applications. Comput Math Appl 33(7):136–136

Ozbudak E, Thattai M, Lim H, Shraiman B, van Oudenaarden A (2004) Multistability in the lactose utilization network of *Escherichia coli*. Nature 427:737–740

Santillán M (2008) Bistable behavior in a model of the lac operon in *Escherichia coli* with variable growth rate. Biophys J 94(6):2065–2081

Schatz MC (2015) Biological data sciences in genome research. Genome Res 25:1417–1422

Stigler B, Chamberlin HM (2012) A regulatory network modeled from wild-type gene expression data guides functional predictions in *Caenorhabditis elegans* development. BMC Syst Biol. https://doi.org/10.1186/1752-0509-6-77

Veliz-Cuba A, Stigler B (2011) Boolean models can explain bistability in the *lac* operon. J Comput Biol 18(6):783–794

Wong P, Gladney S, Keasling J (1997) Mathematical model of the *lac* operon: Inducer exclusion, catabolite repression, and diauxic growth on glucose and lactose. Biotechnol Prog 13(2):132–143