Canalizing structure of genetic network dynamics: modelling and identification via mixed-integer programming

Eugenio Cinquemani, Riccardo Porreca, John Lygeros and Giancarlo Ferrari-Trecate

Abstract—We discuss the identification of genetic networks based on a class of boolean gene activation rules known as hierarchically canalizing functions. We introduce a class of kinetic models for the concentration of the proteins in the network built on a family of canalizing functions that has been shown to capture the vast majority of the known interaction networks. The simultaneous identification of the structure and of the parameters of the model from experimental data is addressed based on a mixed integer parametrization of the model class. The resulting regression problem is solved numerically via standard branch-and-bound techniques. The performance of the method is tested on simulated data generated by a simple model of Escherichia coli nutrient stress response.

I. INTRODUCTION

Genetic network modelling has been addressed by a number of researchers in the recent times. A number of approaches have been proposed which differ in the modelling approach and in the biological experiments for which they are conceived. An efficient way to automate modelling from the data is to consider qualitative interaction models in the form of Boolean gene activation rules [7]. This approach is well suited for simple experimental data where the (possibly perturbed) system is observed at equilibrium, and is purely discrete. It allows one to draw conclusions about the set of genes that are involved in a certain regulation pattern, but many interactions and the hierarchy of the structure may be obscured. Bayesian networks have also been proposed for the identification of the regulatory logic, for instance in [13].

An alternative approach is to use kinetic models describing the dynamics of the system, see [8] and references therein. When time-course experimental data is available, in principle, a kinetic model can be fitted to the data in the form of a regression, yielding a complete view of the regulation network. The central issue with this approach is that an overwhelming number of models need to be parsed. For nonlinear models, the problem of identifying the network of interactions was reduced to a parametric identification problem via neural network-like models in [21].

A middle ground between purely discrete approaches and kinetic model fitting is provided by linearization methods [16], [17], [18]. The idea is to start from a kinetic model and to fit a linearized version of it to several equilibria of the system obtained in response to a set of known perturbations. Under the assumption that the perturbations are small enough, these methods provide hints on the interactions among genes and alsoUrl and local information about the strength of the regulatory interactions. However, the key assumption that the linearized dynamics are the same at all equilibria appears to be overly restrictive. Hence, the practicality of these methods is somewhat limited. A stochastic approach that exploits intrinsic gene expression noise as an inherent perturbation signal has been proposed in [22].

A promising direction of research is based on the approximation of kinetic modelling via hybrid modelling [19], [20], [1]. The idea is to turn a full-blown kinetic model into a piecewise linear model by replacing sigmoidal activation functions with step functions. This method retains most of the descriptive capability of dynamical models but allows one to reduce identification to the fitting of several linear models over different partitions of the state space. This makes the approach extremely powerful and efficient. Unfortunately, experimental evidence shows that the approximation via step functions is too coarse for certain systems. A stochastic hybrid counterpart has been explored e.g. in [23], where activation functions have been replaced by sigmoidal switching probabilities of genetic regulatory events.

The objective of this work is to develop an efficient approach to genetic network identification via kinetic modelling. The key observation is that, out of all possible dynamical interactions among genes, it is likely that a restricted family of interaction patterns is found in the actual biological systems. In the context of Boolean networks, it was observed in the literature that most of the known gene activation rules fall in a specific class of Boolean functions, namely the Hierarchically Canalizing Functions (HCF) [4], [5], [6]. When translated into a kinetic modelling framework, this means that a restricted set of nonlinearities should be considered for the modelling and identification. Our aim is to exploit this knowledge to devise a strategy for the simultaneous structural and parametric identification of kinetic models of genetic networks. Ideas on how to apply HCF in the context of Boolean networks are discussed e.g. in [10], [11], [12].

In Section II we review the modelling of gene activation via Boolean rules and concentrate on the class of hierarchically canalizing functions. Kinetic models of gene activation via ordinary differential equations are discussed in Section III. In Section IV we state the genetic network identification problem and formalize the class of models.
where a solution is sought. A reformulation of the problem in terms of a mixed-integer optimization problem is developed in Section V. A numerical case study on a simulated model of nutrient stress response in *Escherichia Coli* is presented in Section VI to assess the performance of our identification approach. Conclusions on and perspectives of our work are drawn in Section VII.

II. BOOLEAN RULES: HIERARCHICALLY CANALIZING FUNCTIONS

Boolean network modelling [7] describes the activation of each gene as a boolean function of the expression of the genes in network. Consider a network with *n* genes. For *i* = 1, . . . , *n*, let *x*<sub>*i*</sub> ∈ {0, 1} indicate the activation status of gene *i* at a fixed time instant; *x*<sub>*i*</sub> = 1 means that gene *i* is expressed, while *x*<sub>*i*</sub> = 0 means it is not. The activation of gene *i* at the next time instant is indicated by *x*<sub>*i*</sub><sup>+</sup> and is modelled by

\[ x_{i} = b_{i}(x_{1}, \ldots, x_{n}), \]

where the function \( b_{i} : \{0, 1\}^{n} \rightarrow \{0, 1\} \) is a boolean rule. It is easily seen that any boolean rule can be translated into algebraic relationships. It is sufficient to note the following two equivalences:

\[ x_{i} = \neg x_{j} \equiv x_{i} = 1 - x_{j}, \]

\[ x_{i} = x_{j} \land x_{k} \equiv x_{i} = x_{j} \cdot x_{k}. \]

In particular, in follows that *x*<sub>*i*</sub> = *x*<sub>*j*</sub> ∨ *x*<sub>*k*</sub> is equivalent to

\[ x_{i} = 1 - (1 - x_{j}) \cdot (1 - x_{k}). \]

*Canalizing functions* are a subclass of boolean rules whose output is determined by at least one value of at least one input variable. That is, there exists a *canalizing input* *x*<sub>*j*</sub> such that, if *x*<sub>*j*</sub> takes the *canalizing value* *u* ∈ {0, 1}, then the function \( b_{i}(x_{1}, \ldots, x_{n}) \) takes on the *canalized value* *z* ∈ {0, 1} regardless of the value of the other variables *x*<sub>*k*</sub>, with *k* ≠ *j*. Within this class, Hierarchically Canalizing Functions (HCFs) with *ℓ* ≤ *n* effective inputs are characterized by the existence of an ordered subset \( (x_{j_{1}}, \ldots, x_{j_{\ell}}) \) of the variables *x*<sub>1</sub>, . . . , *x*<sub>*n*</sub> with the following property. Input *x*<sub>*j*</sub><sup>1</sup> is canalizing with canalizing value *u*<sub>*j*</sub><sup>1</sup>. When *x*<sub>*j*</sub><sup>1</sup> takes its non-canalizing value ¬ *u*<sub>*j*</sub><sup>1</sup>, *x*<sub>*j*</sub><sup>2</sup> is canalizing with canalizing value *u*<sub>*j*</sub><sup>2</sup>. When both *x*<sub>*j*</sub><sup>1</sup> and *x*<sub>*j*</sub><sup>2</sup> take on their non-canalizing values, *x*<sub>*j*</sub><sup>3</sup> is canalizing, an so on. Finally, when *x*<sub>*j*</sub><sup>1</sup>, *x*<sub>*j*</sub><sup>2</sup>, . . . , *x*<sub>*j*</sub><sup>*n*</sup> take on their non-canalizing values, the value of the expression is determined by *x*<sub>*j*</sub><sup>1</sup>. It was shown that every HCF \( b_{i}(x_{1}, \ldots, x_{n}) \) with *ℓ* effective inputs can be written as [4]

\[ \sigma^{\pm}(x_{j_{1}}) \circ \left( \sigma^{\pm}(x_{j_{2}}) \circ \left( \ldots \circ \left( \sigma^{\pm}(x_{j_{\ell-1}}) \circ \sigma^{\pm}(x_{j_{\ell}}) \right) \right) \right), \]

where \( \circ \) is either \( \lor \) or \( \land \), and each unary operator \( \sigma^{\pm} \) is either the negation, \( \sigma^{-} = \neg \), or the boolean identity map, \( \sigma^{+} \).

HCFs received much attention since it was observed that the vast majority of the known regulatory interactions between genes can be written in terms of HFC-type rules, see [5], [4], [6]. In particular, two specific subfamilies of HCFs, termed \( S_{0}^{\ell} \) and \( S_{1}^{\ell} \), appear to explain most of the known interactions. By definition, the boolean rule (1) is in \( S_{0}^{\ell} \) if all operators \( \circ \) are equal to \( \land \), while it is in \( S_{1}^{\ell} \) if the rightmost operator \( \circ \) is equal to \( \lor \) and all others are equal to \( \land \). (A generalization of this definition to the classes \( S_{k}^{\ell} \), with \( k = 0, \ldots, \ell \), exists, see [4], but we do not need it here.) In order to express canalizing rules by algebraic relationships, for any *x*<sub>*j*</sub> ∈ {0, 1}, let us overload the notation by defining

\[ \sigma^{+}(x_{j}) = x_{j} \quad \text{and} \quad \sigma^{-}(x_{j}) = 1 - x_{j}. \]

Then we have the following straightforward result.

**Proposition 1**: Any function \( b_{i}(x_{1}, \ldots, x_{n}) \) in the class \( S_{0}^{\ell} \cup S_{1}^{\ell} \) can be written in one of the two forms:

\[ \sigma^{\pm}(x_{j_{1}}) \cdot \ldots \cdot \sigma^{\pm}(x_{j_{\ell}}), \]

\[ \sigma^{\pm}(x_{j_{1}}) \cdot \ldots \cdot \sigma^{\pm}(x_{j_{\ell-1}}) \cdot (1 - \sigma^{\pm}(x_{j_{\ell-1}}) \sigma^{\pm}(x_{j_{\ell}})), \]

where \( \ell \leq n \), \( (j_{1}, \ldots, j_{\ell}) \) is an ordered subset of \( \{1, \ldots, n\} \) and each function \( \sigma^{\pm} \) is either \( \sigma^{+} \) or \( \sigma^{-} \).

III. DYNAMICAL MODELS OF GENETIC NETWORKS

For \( t \in \mathbb{R} \) and \( i = 1, \ldots, n \), let \( x_{i}(t) \in \mathbb{R}^{+} \) denote the concentration at time *t* of the protein encoded by gene *i*. Consider the following dynamical model:

\[ \dot{x}_{i} = -\gamma_{i} x_{i} + g_{i}(x_{1}, \ldots, x_{n}), \]

where \( \gamma_{i} \in \mathbb{R}^{+} \) is the unregulated degradation constant. Function \( g_{i} : \mathbb{R}^{n} \rightarrow \mathbb{R} \) represents a variable synthesis rate that encodes the regulatory effects of all proteins in the network on the expression of gene *i*. In practice, only a subset of these proteins act as transcription factors for gene *i*. In this case, \( g_{i} \) effectively depends on a subset of \( \{x_{1}, \ldots, x_{n}\} \).

Kinetic-type models (see e.g. [8]) express \( g_{i}(x_{1}, \ldots, x_{n}) \) as a weighted combination (sums and products) of sigmoidal functions \( \sigma^{+}(x_{j}, \theta) \) or \( \sigma^{-}(x_{j}, \theta) \), with \( j = 1, \ldots, n \), where \( \sigma^{+}(\cdot, \theta) : \mathbb{R}^{+} \rightarrow [0, 1] \) is increasing, parameter vector \( \theta \) determines the “shape” (e.g. threshold and steepness) of the sigmoid, and \( \sigma^{-} = 1 - \sigma^{+} \). In this case, \( g_{i}(x_{1}, \ldots, x_{n}) \) represents the activation level of the expression of the gene.

**Example 1**: The following model, developed in [3], captures the transition of the bacterium *Escherichia coli* between the stationary phase, due to carbon starvation, and the exponential growth phase, when carbon is available. A starvation (input) signal \( u_{s} : \mathbb{R} \rightarrow \mathbb{R}^{+} \) indicates abundance \( (u_{s} = 0) \) or lack \( (u_{s} = 1) \) of carbon, and determines the activation or deactivation of the response-to-starvation mechanism. Subscribers 1 through 6 stand for proteins Cya, CRP, Fis, GyrAB, TopA and rnm, in the same order. Each \( \theta_{i} \) is a vector of parameters for a sigmoid acting on the *i*-th protein concentration. Different superscripts indicate possibly different parameter values. Coefficients \( \kappa_{i} \in \mathbb{R}^{+} \) quantify the protein synthesis rate when gene *i* is expressed. Different superscripts correspond to alternative activation
paths and/or baseline production rates.
\[
\dot{x}_1 = \kappa_1 + \kappa_5^2 (1 - \sigma^+(x_2, \theta_1^2)) \sigma^+(x_1, \theta_1^1) u_1 - \gamma_1 x_1
\]
(4)
\[
\dot{x}_2 = \kappa_2^2 + \kappa_5^2 \sigma^-(x_3, \theta_2^1) \sigma^+(x_2, \theta_1^2) \sigma^+(x_1, \theta_1^1) u_4 \\
+ \kappa_5^2 \sigma^-(x_3, \theta_2^1) \sigma^-(x_2, \theta_1^2) x_2
\]
(5)
\[
\dot{x}_3 = \kappa_3^2 \sigma^-(x_1, \theta_3^2) (1 - \sigma^+(x_2, \theta_1^2)) \sigma^+(x_1, \theta_1^1) u_4 \\
+ \kappa_5^2 \sigma^-(x_4, \theta_2^1) \sigma^-(x_3, \theta_2^1) \sigma^-(x_3, \theta_3^3) x_3
\]
(6)
\[
\dot{x}_4 = \kappa_4 (1 - \sigma^+(x_4, \theta_2^1)) \sigma^-(x_2, \theta_1^2) - \gamma_4 x_4
\]
(7)
\[
\dot{x}_5 = \kappa_5 \sigma^+(x_4, \theta_1^1) \sigma^-(x_3, \theta_3^3) \sigma^+(x_3, \theta_3^1) - \gamma_5 x_5
\]
(8)
\[
\dot{x}_6 = \kappa_6 + \kappa_5^2 \sigma^+(x_3, \theta_3^1) - \gamma_6 x_6
\]
(9)
A detailed illustration of the model can be found in [1]. It should be clear from Example 1 that the activation functions \(g_i\) encode activation rules that are qualitatively similar to the boolean rules of Section II. We shall exploit this similarity in what follows.

IV. PROBLEM STATEMENT

Given a regulatory network with \(n\) genes, we address the problem of identifying a model for the dynamics of the system in the form (3). We assume that noisy measurements \(y_1(t), \ldots, y_n(t)\) of the concentrations \(x_1(t), \ldots, x_n(t)\) are available for \(t \in \mathcal{T}\), where \(\mathcal{T} = \{t_1, t_2, \ldots, t_m\}\) is a sequence of observation instants. We further assume that noisy measurements \(s_1(t), \ldots, s_n(t)\) of the synthesis rates \(g_1, \ldots, g_n\) are available at the same time instants \(t \in \mathcal{T}\). It is shown e.g. in [15] that time-course concentration measurements and the relevant synthesis rates can be drawn from dedicated experiments under similar experimental conditions. On the other hand, the results below can be extended with some additional complexity to the case where only concentrations are available. For the observations we shall consider a model with multiplicative noise,
\[
y_i(t) = (1 + e_i(t)) x_i(t),
\]
(10)
\[
s_i(t) = (1 + e_i(t)) g_i(t),
\]
(11)
where \(\{e_i(t) : t \in \mathcal{T}, i = 1, \ldots, n\}\) and \(\{e_i(t) : t \in \mathcal{T}, i = 1, \ldots, n\}\) are i.i.d random variables with mean 0 and variance \(\sigma^2 > 0\) and \(\sigma^2 > 0\). This model appears to be well suited to describe nonnegative observations of protein concentrations [24].

For \(i = 1, \ldots, n\), we wish to solve the following regression:
\[
\min J, \quad J = \sum_{h=1}^{m} w_h \left( s_i(t_h) - g_i(y_1(t_h), \ldots, y_n(t_h)) \right)^2,
\]
(12)
where the weights \(w_h \in \mathbb{R}_+^+\) must be chosen based on the relative accuracy of the estimates, i.e. on the noise model, and the minimum is taken with respect to a suitable class of gene activation functions. In light the discussion of Section II, we focus on the class of models where gene activation rules are expressed by functions in the classes \(S_0^\ell\) and \(S_1^\ell\), with \(\ell \leq n\).

Assumption 1: For \(i = 1, \ldots, n\), there exists a nonnegative integer \(\ell \leq n\) such that
\[
g_i(x_1, \ldots, x_n) = \kappa_i^1 + \kappa_i^2 b_i(x_{j_1}, \ldots, x_{j_m}),
\]
where \(\kappa_i^1 \in \mathbb{R}_+\), \(\kappa_i^2 \in \mathbb{R}_+\), \((j_1, \ldots, j_m)\) is an ordered subset of \(\{1, \ldots, n\}\) and \(b_i(x_{j_1}, \ldots, x_{j_m})\) is a combination of sigmoids \(\sigma^\ell(x_j, \theta_{i,j})\) in the form (2).
This assumption allows us to largely reduce the complexity of the identification problem. In summary, we wish to minimize \(J\) with respect to the functions \(b_i \in S_0^\ell \cup S_1^\ell\), with \(0 \leq \ell \leq n\), and all the unknown parameters \(\theta_{i,j}\). By convention, we let \(S_0^\ell\) and \(S_1^\ell\) coincide with the constant function \(b_i \equiv 1\). Identification may result in a poor matching of the data in the few outstanding cases where the gene activation rule is not in \(S_0^\ell \cup S_1^\ell\). This situation will not be addressed here.

A. Choice of the regression weights

Weights attributed to the error terms in \(J\) must reflect the reliability of the data. In principle, weights should account for the uncertainty from \(s_i\) and from \(g_i(y_1(t_h), \ldots, y_n(t_h))\). The noise contribution from \(y_1(t_h), \ldots, y_n(t_h)\) is reshaped by the unknown function \(g_i\) and is difficult to quantify. For the time being we shall ignore this contribution and discuss this approximation in light of numerical results in Section VI. The standard deviation of the \(e_i(t_h)\) is proportional to \(g_i(t_h)\). For \(\sigma^2\) not too large, the latter will be of the same order as \(s_i(t_h)\). Therefore we define the weights as \(w_h \propto s_i^2(t_h)\).

B. Model complexity and overfitting

It is well known that matching data based on a class of arbitrarily complex models may result in overfitting. To counteract this we modify the cost function by including a term that penalizes overly complicated models. Several criteria exist and are accompanied by solid theoretical guarantees. For numerical convenience, we opt for the Final Prediction Error (FPE) principle [9], which results in the modified optimization problem
\[
\min J', \quad J' = \begin{cases} 
\frac{m + p}{m-p} \times J, & \text{if } p < m, \\
+\infty, & \text{otherwise},
\end{cases}
\]
(13)
where \(p\) is the number of the optimization parameters (unknown parameters of the model). This number depends on the parametrization of the model class and on the specific form of the sigmoidal activation functions. Recall that \(m\) is the number of data points.

V. FORMULATION AS A MIXED INTEGER OPTIMIZATION PROBLEM

Direct solution of the regression requires an exhaustive exploration of the class of rules \(S_0^\ell \cup S_1^\ell\) and is too costly. To ameliorate this limitation, we recast the problem in a mixed-integer optimization framework. This allows one to solve the problem by well-established branch-and-bound techniques. While the complexity of the worst-case solution remains unchanged, existing branch-and-bound heuristics tend to
result in a very effective exploration of $S^*_0 \cup S^*_1$, $\ell = 0, \ldots, n$, with enormous savings in terms of computational cost.

Let $\alpha = (\alpha_1, \ldots, \alpha_n)$, $\alpha^* = (\alpha^*_1, \ldots, \alpha^*_n)$ and $\beta = (\beta_1, \ldots, \beta_n)$ be in $\{0, 1\}^n$. Define $\hat{g}_i = \kappa_i^1 + \kappa_i^2 \hat{b}_i$ with

$$
\hat{b}_i = \left[ \prod_{j=1}^{n} \left( 1 - \alpha^*_j \right) + \alpha_j^* \left( \beta_j \sigma_{i,j}^- + (1 - \beta_j) \sigma_{i,j}^+ \right) \right] \times \left[ 1 - \prod_{j=1}^{n} \left( 1 - \alpha^*_j \right) + \alpha_j^* \left( \beta_j \sigma_{i,j}^- + (1 - \beta_j) \sigma_{i,j}^+ \right) + \gamma \right]
$$

where $\sigma_{i,j}^\pm$ is a shorthand for $\sigma^\pm(x_j, \theta_{i,j})$ and

$$
\gamma = (1 - \alpha^*_1) \cdots (1 - \alpha^*_n).
$$

**Proposition 2:** The set of rules $\hat{b}_i$ such that

$$
\alpha_i = \alpha_i^* \leq 1, \hspace{1cm} i = 1, \ldots, n, \hspace{1cm} (14)
$$

$$
0 \leq \alpha_1 + \cdots + \alpha_n \leq 2 \hspace{1cm} (15)
$$

is equal to $(S^*_0 \cup S^*_1) \cup (S^*_1 \cup S^*_0) \cup \cdots \cup (S^*_0 \cup S^*_0)$. Roughly speaking, the first factor of $\hat{b}$ corresponds to a chain of “and” operations, whereas the second factor includes an “or” operation as a special case (compare with Proposition 1). If $\alpha_i = \alpha_i^* = 0$, $x_i$ does not appear in $\hat{b}_i$. If $\alpha_i = 1$, $\sigma^\pm(x_i) \equiv \kappa^i_1 + \kappa^2 \hat{b}_i$. Constraint (15) implies that the rightmost factor is either $1 - \sigma^\pm(x_j)$ (i.e., $\sigma^\pm(x_j)$ is negated and the whole expression is a chain of “and” operations) or $1 - \sigma^\pm(x_j)$ (i.e., $\sigma^\pm(x_j)$ is negated and the whole expression is a chain of “or” operations) $1 - (1 - \sigma^\pm(x_j))(1 - \sigma^\pm(x_k))$. The latter case corresponds to $\alpha_j^* = \alpha_k^* = 1$ and encodes an “or” between (possibly negated) $x_j$ and $x_k$. Finally, $\beta_i = 0$ and $\beta_i = 1$ correspond to having $\sigma^-_i(x_i)$ or $\sigma^+_i(x_i)$, i.e., $\beta_i = 0$ if $x_i$ is negated. The expression of $\gamma$ makes sure that the second factor of $\hat{b}_i$ does not vanish when $\alpha^*_1 = \cdots = \alpha^*_n = 0$. This parameterization is redundant. For example, an “and” factor $\sigma^+(x_j)$ can be equally rewritten as a “degenerate or” factor $1 - \sigma^-(x_j)$. In the first case, $\alpha_j = 1$ and $\alpha_j^* = 0$; in the latter case, $\alpha_j = 0$ and $\alpha_j^* = 1$. Simulations show that the parameterization is suitable for the optimization problem of interest, see Section VI. We may now restate (13) as the following constrained mixed-integer nonlinear optimization problem:

$$
\min \frac{m + p}{m - p} \times \sum_{h=1}^{m} \left( s_h(t_h) - \hat{g}_h(y_1(t_h), \ldots, y_n(t_h)) \right)^2 
$$

s.t. (14), (15).

Minimization is now expressed with respect to $\alpha$, $\alpha^*$, $\beta$, $\kappa^i_1$, $\kappa^i_2$ and the sigmoidal parameters $\theta_{i,j}$, with $j = 1, \ldots, n$. The model complexity $p$ is computed as follows:

$$
p = \sum_{j=1}^{n} \text{size}(\theta_{i,j}) \mathbb{1}_{\alpha_j + \alpha_j^* > 0} + 2,
$$

i.e., $p$ counts the number of parameters of the sigmoids included in the model (indicated by $\mathbb{1}_{\alpha_j + \alpha_j^* > 0}$) plus the two parameters $\kappa^i_1$ and $\kappa^i_2$. $\alpha$, $\alpha^*$ and $\beta$ are the integer variables of the problem that will be handled with branch-and-bound techniques. For fixed values of $\alpha$, $\alpha^*$ and $\beta$, optimization with respect to the parameters $\theta_{i,j}$ is generally nonconvex. Its complexity depends on the analytic expression of $\sigma^+$ and on the number of the unknown parameters $p$.

We are especially interested in the case of Hill functions. Let $\theta = (\eta, d)$, with $\eta \in \mathbb{R}^+$ and $d \in \mathbb{R}^+$, denote the threshold and cooperativity parameters. Take

$$
\sigma^+(x, \theta) = \frac{x^d}{x^d + \eta^d}, \hspace{1cm} \sigma^-(x, \theta) = \frac{\eta^d}{x^d + \eta^d}, \hspace{1cm} \text{(17)}
$$

Threshold $\eta$ is such that $\sigma^+(\eta, \theta) = \sigma^-(\eta, \theta) = 1/2$, whereas $d$ determines how abruptly $\sigma^+$ increases from 0 to 1 ($\sigma^-$ decreases from 1 to 0) as $x$ increases. With this choice, the expression of $\hat{b}_i$ becomes:

$$
\hat{b}_i = \left[ \prod_{j=1}^{n} \left( 1 - \alpha_j^* \right) + \alpha_j^* \left( \beta_j \frac{x_{d,i,j}^+ + (1 - \beta_j) \eta_{d,i,j}^+}{x_{d,i,j}^+ + \eta_{d,i,j}^+} \right) \right] \times \left[ 1 - \prod_{j=1}^{n} \left( 1 - \alpha_j^* \right) + \alpha_j^* \left( \beta_j \frac{x_{d,i,j}^- + (1 - \beta_j) \eta_{d,i,j}^-}{x_{d,i,j}^- + \eta_{d,i,j}^-} \right) + \gamma \right].
$$

Provided the $d_{i,j}$ are not fixed in advance, a statement of the problem with less binary variables (namely, without variables $\beta_j$) may be obtained by observing that

$$
\frac{\eta^d}{\eta^d + x^d} = \frac{x^d}{x^d + \eta^d},
$$

with $d' = -d < 0$. Therefore, without loss of generality one may write all terms of the type $(\beta x^d + (1 - \beta) \eta^d)/(x^d + \eta^d)$, with $d \geq 0$, as $x^d/(x^d + \eta^d)$, with $d \in \mathbb{R}$, where $d < 0$ corresponds to $\beta = 0$ and $d > 0$ to $\beta = 1$. In this case, the number of binary variables is $2n$, and the total number of parameters is $4n + 2$. Otherwise, in general, the problem has $3n$ binary variables and $2n + 2$ continuous variables, for a total of $5n + 2$ unknown parameters.

VI. NUMERICAL RESULTS

We considered the problem of identifying the model of Example 1 from simulated data. The model is endowed with sigmoidal activation functions in the form (17). Each sigmoid is parameterized by a pair $\theta^k = (\eta^k, d^k)$. We focused on the reentry of the system in the exponential growth phase. In this case the starvation signal if off and the system is described by the model obtained by setting $u_0 = 0$ in (4)-(9), namely

$$
\dot{x}_1 = \kappa_1^1 + \kappa_2^1 \sigma - \gamma_1 x_1, \hspace{1cm} (18)
$$

$$
\dot{x}_2 = \kappa_2^1 + \kappa_3^1 \sigma(x_3, \theta^3_1) - \gamma_2 x_2, \hspace{1cm} (19)
$$

$$
\dot{x}_3 = \kappa_3^1 \sigma(x_3, \theta^3_1) + \kappa_4^1 \sigma(x_4, \theta^4_1) \sigma(x_5, \theta^5_1) \sigma(x_3, \theta^3_1) - \gamma_3 x_3, \hspace{1cm} (20)
$$

$$
\dot{x}_4 = \kappa_4 \left( 1 - \sigma(x_4, \theta^4_1) \sigma(x_5, \theta^5_1) \sigma(x_3, \theta^3_1) \right) - \gamma_4 x_4, \hspace{1cm} (21)
$$

$$
\dot{x}_5 = \kappa_5 \sigma(x_4, \theta^4_1) \sigma(x_5, \theta^5_1) \sigma(x_3, \theta^3_1) - \gamma_5 x_5, \hspace{1cm} (22)
$$

$$
\dot{x}_6 = \kappa_6^1 + \kappa_7^1 \sigma(x_3, \theta^3_1) - \gamma_6 x_6. \hspace{1cm} (23)
$$

We assume that the true parameters of the system and the initial conditions after starvation are as in Table I.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\kappa_1$</td>
<td>0.5</td>
</tr>
<tr>
<td>$\kappa_2$</td>
<td>1.0</td>
</tr>
<tr>
<td>$\kappa_3$</td>
<td>1.5</td>
</tr>
<tr>
<td>$\kappa_4$</td>
<td>2.0</td>
</tr>
</tbody>
</table>

The cooperativity (steepness)
coefficients of the Hill functions were chosen based on common biological knowledge and were identically set to \( d = 3 \). Note that all equations fall in a class \( S_i^{t} \cup S_i^{t} \) with \( \ell \leq 6 \) except for Eq. (20).

In our experiment we assumed that the cooperativity coefficients are known and correctly set to \( d = 3 \). Since parameters \( d \) are not part of the optimization problem, from now on they will be dropped from our notation. In particular, we shall write \( \eta \) in place of \( \theta \) for the unknown parameters of the sigmoidal functions. We attempt the estimation of the structure of the system along with the synthesis rates and the thresholds of the sigmoidal activation functions. Identification is performed separately for each of the 6 equations of the model and relies on the numerical solution of (16). The total number of unknown parameters for each equation is \( 4n + 2 \). A basic MATLAB implementation of a nonlinear branch-and-bound optimization algorithm, solving appropriate optimization subproblems by continuous relaxation of the integer variables [25], is used for this purpose. Optimization is performed on data from a single simulation of the model. Measurements of protein concentrations and synthesis rates were collected every \( T = 5 \) min over the time interval \([0, 1200] \) min, i.e. \( t_h = (h - 1)T \), with \( h = 1, \ldots, m \) and \( m = 241 \). They were artificially corrupted by zero-mean Gaussian noise generated at random in accordance with the multiplicative noise model. We set \( \sigma_n = 0.01 \). This corresponds to a noise magnitude effectively concentrated within 3\% of the observed synthesis rate value. For the noise in the observation of \( x_i \) we consider two different cases:

1) \( \sigma_n = 0 \) (i.e. perfect measurements \( y_i = x_i \));
2) \( \sigma_n = 0.01 \) (i.e. noise effectively concentrated within 3\% of the observed concentration value).

The comparison of the two cases will reveal the effect of ignoring \( e \) in the choice of the regression weights (see Section IV-A). Optimization was initialized by setting each threshold value at the mean value of the corresponding concentration measurements. For each \( i \) we initially set \( \kappa_i^{1} = \min \{ s_i(t_h) : h = 1, \ldots, m \} \) and \( \kappa_i^{2} = \max \{ s_i(t_h) : h = 1, \ldots, m \} \). The integer variables \( \alpha_j \) and \( \beta_j \) were initially relaxed and set to 0.5 (i.e. the mean of the admissible values), whereas the \( \alpha_j^{*} \) were all set to 0 so as to fulfill constraint (15).

Simulation of the model (18)–(23) from the initial conditions of Table I reveals that some of the sigmoidal nonlinearities are not sufficiently explored by the data. This is apparent in Figure 1, where it can be seen that the thresholds associated to several sigmoidal functions are not crossed by the protein concentrations they act on. As a result only part of the model can be identified. The identifiable part of the model turns out to be:

\[
g_1 = \kappa_1^1 + \kappa_1^2, \tag{24}
g_2 = \kappa_2^1 + \kappa_2^2 \sigma^-(x_3, \eta_{3,3}), \tag{25}
g_3 \simeq \kappa_3^1 + \kappa_3^2 \sigma^+(x_4, \eta_{4,3}) \sigma^-(x_3, \eta_{3,3}), \tag{26}
g_4 \simeq \kappa_4^2 \sigma^-(x_4, \eta_{4,3}) \sigma^+(x_3, \eta_{3,3}), \tag{27}
g_5 \simeq \kappa_5^2 \sigma^+(x_4, \eta_{4,3}) \sigma^- (x_3, \eta_{3,3}), \tag{28}
g_6 = \kappa_6^1 + \kappa_6^2 \sigma^+(x_3, \eta_{6,3}). \tag{29}
\]

This model was derived “manually” from (18)–(23) by fixing the parameters to the true values and removing all the nonlinearities that have a negligible effect on the predictive power of the model for the given initial conditions. Note that every function (24)–(29) belongs to \( S_i^t \cup S_i^t \) for some \( \ell \leq 6 \).

We noticed that model penalization was made less effective by the convergence to local minima corresponding...
to nonminimal models. To compensate for this, we post-processed the identification results by the following iterative procedure. Given the identified model, take one sigmoid of the model out and refit the parameters of the remaining sigmoids. Do this iteratively until no sigmoid can be taken out of the model with an improvement in the fitting cost. We verified that this procedure provides simpler models at a lower cost in several cases. This indicates clearly that the nonconvexity of the problem is an issue and that the iterative procedure improves optimization effectively.

The interactions identified after postprocessing are
\[
\begin{align*}
\hat{g}_1 &= \hat{\kappa}_1^1 + \hat{\kappa}_1^2 \\
\hat{g}_2 &= \hat{\kappa}_2^1 + \hat{\kappa}_2^2 \sigma^+(x, \hat{\eta}_{2,3}) \\
\hat{g}_3 &= \hat{\kappa}_3^1 + \hat{\kappa}_3^2 \sigma^-(x, \hat{\eta}_{3,3}) \sigma^+(x_4, \hat{\eta}_{3,4}) \\
\hat{g}_4 &= \hat{\kappa}_4^1 + \hat{\kappa}_4^2 \sigma^-(x, \hat{\eta}_{4,4}) \sigma^+(x_4, \hat{\eta}_{4,4}) \\
\hat{g}_5 &= \hat{\kappa}_5^1 + \hat{\kappa}_5^2 \sigma^+(x, \hat{\eta}_{5,3}) \sigma^+(x_4, \hat{\eta}_{5,4}) \sigma^+(x_6, \hat{\eta}_{5,6}) \\
\hat{g}_6 &= \hat{\kappa}_6^1 + \hat{\kappa}_6^2 \sigma^+(x, \hat{\eta}_{6,3})
\end{align*}
\]
for Case 1 (perfect concentration measurements), and
\[
\begin{align*}
\hat{g}_1 &= \hat{\kappa}_1^1 + \hat{\kappa}_1^2 \\
\hat{g}_2 &= \hat{\kappa}_2^1 + \hat{\kappa}_2^2 \sigma^+(x_2, \hat{\eta}_{2,2}) \sigma^-(x_3, \hat{\eta}_{2,3}) \\
\hat{g}_3 &= \hat{\kappa}_3^1 + \hat{\kappa}_3^2 \sigma^+(x_1, \hat{\eta}_{3,1}) \sigma^-(x_3, \hat{\eta}_{3,3}) \times \sigma^+(x_4, \hat{\eta}_{3,4}) \sigma^+(x_6, \hat{\eta}_{3,6}) \\
\hat{g}_4 &= \hat{\kappa}_4^1 + \hat{\kappa}_4^2 \sigma^-(x_4, \hat{\eta}_{4,4}) \sigma^-(x_5, \hat{\eta}_{4,5}) \times \sigma^+(x_2, \hat{\eta}_{4,2}) \sigma^+(x_3, \hat{\eta}_{4,3}) \\
\hat{g}_5 &= \hat{\kappa}_5^1 + \hat{\kappa}_5^2 \sigma^+(x_3, \hat{\eta}_{5,3}) \sigma^+(x_4, \hat{\eta}_{5,4}) \sigma^+(x_6, \hat{\eta}_{5,6}) \\
\hat{g}_6 &= \hat{\kappa}_6^1 + \hat{\kappa}_6^2 \sigma^+(x_3, \hat{\eta}_{6,3})
\end{align*}
\]
for Case 2 (noisy concentration measurements). Superscript "\(^{\cdot}\)" is used to denote estimates. The estimated values of the parameters for the two cases are reported in Tables II and III. Figure 2 shows that the matching between the data and the estimated model predictions is very accurate in all cases.

In general, the estimation of the interactions and of the unknown parameters is quite accurate, but the presence of measurement noise on the concentration values tends to favor the choice of overly complicated models. We interpret this as a consequence of having ignored the contribution of \(e_i\) in the choice of the regression weights \(w_{ij}\). In fact, the regression problem could be better formulated as the identification of an errors-in-variables model [9]. The resulting problem is nonstandard due to the nonlinearity of the model and requires further investigation. In Case 1, the estimated structure of the synthesis rates is exact with the exception of \(\hat{g}_5\), which contains the spurious sigmoid \(\sigma^-(x_6, \hat{\eta}_{5,6})\). Table II shows that the value \(\sigma^-(x_6, \hat{\eta}_{5,6})\) is nearly 1 along the observed state trajectories. An offset in the estimate \(\hat{\kappa}_2^2\) can be observed. This is compensated partly by an offset in the estimate \(\hat{\eta}_{5,4}\) and partly by the presence of \(\sigma^-(x_6, \hat{\eta}_{5,6})\). In both Case 1 and Case 2, there is a clear identifiability problem for the parameters \(\kappa_1^1\) and \(\kappa_1^2\). The reason is clear from the expression of \(g_1\). In both cases, what is estimated accurately is the sum \(\hat{\kappa}_1^1 + \hat{\kappa}_1^2 \approx \kappa_1^1 + \kappa_1^2\). In Case 2, incorrect interactions are introduced in \(\hat{g}_3\) and \(\hat{g}_4\). For \(\hat{g}_3\), Table III suggests that the erroneous presence of \(\sigma^-(x_6, \hat{\eta}_{5,6})\) \(\approx 1\) and of \(\sigma^-(x_1, \hat{\eta}_{3,1})\) mitigates the effect of the overestimation of \(\kappa_3^2\), while the additional error in the estimation of \(\kappa_3^2\) is masked by the
TABLE II

Estimation results for Case 1. For each threshold parameter, the rightmost column indicates the range of values taken on by the corresponding estimated sigmoidal nonlinearity along the system trajectory.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Param</th>
<th>Estimate</th>
<th>True</th>
<th>Sigmoid span</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$\kappa_1^1$</td>
<td>$2.477\times10^{-11}$</td>
<td>$3.034\times10^{-12}$</td>
<td>[0.000, 0.998]</td>
</tr>
<tr>
<td></td>
<td>$\kappa_1^2$</td>
<td>$1.476\times10^{-12}$</td>
<td>$2.317\times10^{-11}$</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>$\kappa_1^1$</td>
<td>$1.552\times10^{-9}$</td>
<td>$1.553\times10^{-9}$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\kappa_2^1$</td>
<td>$1.322\times10^{-8}$</td>
<td>$1.322\times10^{-8}$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\eta_{2,3}$</td>
<td>$4.056\times10^{-8}$</td>
<td>$3.991\times10^{-8}$</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>$\kappa_1^2$</td>
<td>$3.502\times10^{-10}$</td>
<td>$3.401\times10^{-9}$</td>
<td>[0.000, 0.998]</td>
</tr>
<tr>
<td></td>
<td>$\kappa_3$</td>
<td>$8.682\times10^{-9}$</td>
<td>$8.688\times10^{-9}$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\eta_{3,3}$</td>
<td>$1.991\times10^{-6}$</td>
<td>$2.020\times10^{-6}$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\eta_{3,4}$</td>
<td>$4.022\times10^{-8}$</td>
<td>$3.991\times10^{-8}$</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>$\kappa_1^2$</td>
<td>$9.954\times10^{-10}$</td>
<td>$9.938\times10^{-10}$</td>
<td>[0.018, 1.000]</td>
</tr>
<tr>
<td></td>
<td>$\eta_{4,3}$</td>
<td>$7.496\times10^{-7}$</td>
<td>$7.472\times10^{-7}$</td>
<td>[0.011, 0.999]</td>
</tr>
<tr>
<td></td>
<td>$\eta_{4,4}$</td>
<td>$1.883\times10^{-7}$</td>
<td>$1.888\times10^{-7}$</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>$\kappa_2^2$</td>
<td>$1.809\times10^{-9}$</td>
<td>$2.548\times10^{-9}$</td>
<td>[0.000, 0.806]</td>
</tr>
<tr>
<td></td>
<td>$\eta_{5,3}$</td>
<td>$7.684\times10^{-7}$</td>
<td>$7.472\times10^{-7}$</td>
<td>[0.001, 0.131]</td>
</tr>
<tr>
<td></td>
<td>$\eta_{5,4}$</td>
<td>$1.636\times10^{-7}$</td>
<td>$1.888\times10^{-7}$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\eta_{5,6}$</td>
<td>$4.418\times10^{-5}$</td>
<td>$4.022\times10^{-5}$</td>
<td>[0.929, 1.000]</td>
</tr>
<tr>
<td>6</td>
<td>$\kappa_6^2$</td>
<td>$1.496\times10^{-8}$</td>
<td>$1.560\times10^{-8}$</td>
<td>[0.000, 0.974]</td>
</tr>
<tr>
<td></td>
<td>$\kappa_6$</td>
<td>$1.488\times10^{-7}$</td>
<td>$1.488\times10^{-7}$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\eta_{6,3}$</td>
<td>$3.663\times10^{-7}$</td>
<td>$3.663\times10^{-7}$</td>
<td></td>
</tr>
</tbody>
</table>

TABLE III

Estimation results for Case 2. For each threshold parameter, the rightmost column indicates the range of values taken on by the corresponding estimated sigmoidal nonlinearity along the system trajectory. True threshold values in brackets indicate that the corresponding sigmoid was deemed unidentifiable.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Param</th>
<th>Estimate</th>
<th>True</th>
<th>Sigmoid span</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$\kappa_1^1$</td>
<td>$2.477\times10^{-11}$</td>
<td>$3.034\times10^{-12}$</td>
<td>[0.000, 0.998]</td>
</tr>
<tr>
<td></td>
<td>$\kappa_1^2$</td>
<td>$1.476\times10^{-12}$</td>
<td>$2.317\times10^{-11}$</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>$\kappa_1^1$</td>
<td>$1.552\times10^{-9}$</td>
<td>$1.553\times10^{-9}$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\kappa_2^1$</td>
<td>$1.322\times10^{-8}$</td>
<td>$1.322\times10^{-8}$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\eta_{2,3}$</td>
<td>$4.056\times10^{-8}$</td>
<td>$3.991\times10^{-8}$</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>$\kappa_1^2$</td>
<td>$3.502\times10^{-10}$</td>
<td>$3.401\times10^{-9}$</td>
<td>[0.000, 0.998]</td>
</tr>
<tr>
<td></td>
<td>$\kappa_3$</td>
<td>$8.682\times10^{-9}$</td>
<td>$8.688\times10^{-9}$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\eta_{3,3}$</td>
<td>$1.991\times10^{-6}$</td>
<td>$2.020\times10^{-6}$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\eta_{3,4}$</td>
<td>$4.022\times10^{-8}$</td>
<td>$3.991\times10^{-8}$</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>$\kappa_1^2$</td>
<td>$9.954\times10^{-10}$</td>
<td>$9.938\times10^{-10}$</td>
<td>[0.018, 1.000]</td>
</tr>
<tr>
<td></td>
<td>$\eta_{4,3}$</td>
<td>$7.496\times10^{-7}$</td>
<td>$7.472\times10^{-7}$</td>
<td>[0.011, 0.999]</td>
</tr>
<tr>
<td></td>
<td>$\eta_{4,4}$</td>
<td>$1.883\times10^{-7}$</td>
<td>$1.888\times10^{-7}$</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>$\kappa_2^2$</td>
<td>$1.809\times10^{-9}$</td>
<td>$2.548\times10^{-9}$</td>
<td>[0.000, 0.806]</td>
</tr>
<tr>
<td></td>
<td>$\eta_{5,3}$</td>
<td>$7.684\times10^{-7}$</td>
<td>$7.472\times10^{-7}$</td>
<td>[0.001, 0.131]</td>
</tr>
<tr>
<td></td>
<td>$\eta_{5,4}$</td>
<td>$1.636\times10^{-7}$</td>
<td>$1.888\times10^{-7}$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\eta_{5,6}$</td>
<td>$4.418\times10^{-5}$</td>
<td>$4.022\times10^{-5}$</td>
<td>[0.929, 1.000]</td>
</tr>
<tr>
<td>6</td>
<td>$\kappa_6^2$</td>
<td>$1.496\times10^{-8}$</td>
<td>$1.560\times10^{-8}$</td>
<td>[0.000, 0.974]</td>
</tr>
<tr>
<td></td>
<td>$\kappa_6$</td>
<td>$1.488\times10^{-7}$</td>
<td>$1.488\times10^{-7}$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\eta_{6,3}$</td>
<td>$3.663\times10^{-7}$</td>
<td>$3.663\times10^{-7}$</td>
<td></td>
</tr>
</tbody>
</table>

VII. CONCLUSIONS AND PERSPECTIVES

We have developed an approach to the structural and parametric identification of gene regulatory networks. Our method is based on a model for protein concentration dynamics where gene activation rules obey a specific canalizing structure. This choice was motivated by the conjecture that most gene interactions in nature can be explained by these rules [4], [5], [6]. We exploited the canalizing structure of the gene dynamics to reduce the complexity of the model class and to turn identification into a mixed-integer optimization problem.

The numerical analysis of the method on simulated data via standard optimization tools revealed that the approach is promising, however a number of issues remain open. Even for a fixed network of interactions, the estimation of the unknown model parameters is nonconvex. Convergence to local minima appears to be in conflict with the penalization of the model complexity. While ameliorated by suitable postprocessing of the optimization results, the issue remains open. In presence of noise, spurious weak nonlinearities are sometimes introduced with the effect of “explaining” the noise. The penalization of overly complicated models did not solve the problem completely. To address this issue, the reformulation of the problem as the identification of an errors-in-variables model is being considered.

Regardless of the specific identification method, the performance of gene network identification is severely limited by the poor excitation of the system. Nonlinearities that are not fully explored by the system state are not reflected in the data. This raises a nontrivial identifiability issue that requires thorough investigation.

REFERENCES
