Supplementary Material of “Identification of genetic network dynamics with unate structure”

Riccardo Porreca 1, Eugenio Cinquemani 2, John Lygeros 1, Giancarlo Ferrari-Trecate 3

1 Institut für Automatik, ETH Zürich, 8092 Zürich, Switzerland
{porreca, lygeros}@control.ee.ethz.ch

2 INRIA Grenoble-Rhône-Alpes, Montbonnot, 38334 Saint-Ismier Cedex, France
eugenio.cinquemani@inria.fr

3 Dipartimento di Informatica e Sistemistica, Università di Pavia, 27100 Pavia, Italy
giancarlo.ferrari@unipv.it

1 Algebraic form of kinetic models with unate structure

In the main text, a kinetic model \( b_i(x) \), with \( x = (x_1, \ldots, x_n) \), is obtained from a Boolean expression \( B_i(X) \), with \( X = (X_1, \ldots, X_n) \), via the following substitutions. Each variable \( X_i \) is replaced by \( \sigma^+(x_i) \). Given any two functions \( \tau(x) \) and \( \tau'(x) \) representing the boolean expressions \( T(X) \) and \( T'(X) \), we encode \( \neg T(X) \) by \( 1 - \tau(x) \) and \( T(X) \wedge T'(X) \) by \( \tau(x) \cdot \tau'(x) \).

Proposition 1. For an integer \( n_i > 0 \) and \( l = 1, \ldots, n_i \) let \( J_l \) be a set of pairwise different indices from \( \{1, \ldots, n\} \). Assume that

\[
B_i(X) = \bigwedge_{l=1}^{n_i} T_l(X), \quad T_l(X) = \bigvee_{j \in J_l} \tilde{X}_j, \quad (1)
\]

where \( \tilde{X}_j \) is either \( X_j \) or \( \neg X_j \). Then the model \( b_i(x) \) obtained from \( B_i(X) \) can be written as

\[
b_i(x) = \prod_{l=1}^{n_i} \tau_l(x), \quad \tau_l(x) = 1 - \prod_{j \in J_l} \left( 1 - \sigma^\pm(x_j) \right), \quad (2)
\]

where \( \sigma^\pm(x_j) = \sigma^+(x_j) \) if \( \tilde{X}_j = X_j \) and \( \sigma^\pm(x_j) = \sigma^-(x_j) \) if \( \tilde{X}_j = \neg X_j \).
Proof. Since $X_i$ is replaced by $\sigma^+(x_i)$, if $\bar{X}_i = X_i$ then $\sigma^+(x_i) = \sigma^+(x_i)$, while if $\bar{X}_i = -X_i$ then $\sigma^+(x_i) = 1 - \sigma^+(x_i) = \sigma^-(x_i)$. The expression of each term $\tau_l(x)$ follows from the fact that $T_l(X)$ is a chain of disjunctions, which can be equivalently expressed by means of De Morgan’s laws as

$$T_l(X) = -\bigwedge_{i=1}^{n_l} \neg \bar{X}_i.$$

Equation (2) then follows by converting the conjunction of terms $T_l(X)$ into the product of terms $\tau_l(x)$.

\[\square\]

2 Properties and hierarchy of sign patterns

Using the notation introduced in the main text, let $p = (p_1, \ldots, p_n) \in \{-1, 0, 1\}^n$ be a sign pattern, $g_i(x|p)$ be any kinetic model for gene $i$ with unate structure and sign pattern $p$, $S(p)$ be the set of different model structures corresponding to $p$, and $C(p) = \sum_{i=1}^{n} |p_i| \leq n$ be the complexity of $p$. As in the main text, the subscript $i$ will be dropped from $g_i$ because the results are independent of it. Regardless of the model parameters and of its specific structure, $p$ encodes the monotonicity properties of $g(x|p)$.

Proposition 2. Given any sign pattern $p$ with $C(p) > 0$ and concentration vectors $x^1$, $x^2$, it holds that

$$\left[ p_j(x_j^2 - x_j^1) \geq 0, \; j = 1, \ldots, n \right] \Rightarrow \left[ g(x^2|p) - g(x^1|p) \geq 0 \right].$$

(3)

Proof. Let

$$\sigma^{p_j}(x_j) = \begin{cases} 
\sigma^+(x_j), & \text{if } p_j = 1, \\
\sigma^-(x_j), & \text{if } p_j = -1.
\end{cases}$$

Then $g(x|p) = \kappa_0 + \kappa_1 b(x)$, with $\kappa_0 \geq 0$ and $\kappa_1 \geq 0$, where, from Eq.(2),

$$b(x) = \prod_{l=1}^{n_l} \tau_l(x), \quad \tau_l(x) = 1 - \prod_{j \in J_l} (1 - \sigma^{p_j}(x_j)).$$

(4)

In particular, one has $\bigcup_{l=1}^{n_l} J_l = \{j: \; p_j \neq 0\}$. For every $j$, the assumption $p_j(x_j^2 - x_j^1) \geq 0$ implies that $\sigma^{p_j}(x_j^2) \geq \sigma^{p_j}(x_j^1)$. If this assumption holds for $j = 1, \ldots, n$, it can be verified by inspection of (4) that $b(x^2) \geq b(x^1)$. Since the constants $\kappa_0$ and $\kappa_1$ are nonnegative, it follows that $g(x^2|p) - g(x^1|p) \geq 0$.

\[\square\]

Recall the following definition from the main text.
Definition 1. Consider a set of \( m \) data points \((x^k, g^k), k = 1, \ldots, m\). A sign pattern \( p \) is called inconsistent if
\[
\exists k, l \in \{1, \ldots, m\} \text{ such that } \left[ p_j(x^k_j - x^l_j) \geq 0, \ j = 1, \ldots, n \right] \text{ and } \left[ g^k - g^l < 0 \right]. \tag{5}
\]
A pattern that is not inconsistent is called consistent.

Recall that, given any two sign patterns \( p \) and \( p' \), \( p' \subseteq p \) (\( p' \) subpattern of \( p \)) means that all nonzero entries of \( p' \) are equal to the corresponding entries of \( p \). In particular, for every \( p \) it holds that \( p \subseteq p \) and \((0, 0, \ldots, 0) \subseteq p \). It is easily seen that “\( \subseteq \)” is a partial order relation.

We now derive the hierarchical relationships among consistent and inconsistent sign patterns induced by the partial ordering.

Proposition 3.

i. If \( p \) is inconsistent then any \( p' \subseteq p \) is inconsistent.

ii. If \( p \) is consistent then any \( p' \) such that \( p \subseteq p' \) is consistent.

Proof. (i) By the definition of inconsistent pattern, there exist indices \( k \) and \( l \) such that \( g^k - g^l < 0 \) and \( p_j(x^k_j - x^l_j) \geq 0 \), with \( j = 1, \ldots, n \). Since \( p' \subseteq p \), if \( p'_j \neq 0 \) then \( p'_j = p_j \). Hence, it must hold that \( p'_j(x^k_j - x^l_j) \geq 0, j = 1, \ldots, n \). Therefore \( p' \) is inconsistent. (ii) Follows by contradiction from (i). \( \square \)

Definition 2. A consistent pattern \( p \) is minimal if all patterns \( p' \subseteq p \), with \( p' \neq p \), are inconsistent.

Note that, if the null sign pattern \( p = (0, 0, \ldots, 0) \) is consistent, then it is also minimal consistent, in that it has no subpatterns other than itself. Let \( \bar{P} \) be the set of all patterns \( \bar{p} \) defined as in Section 2.2 of the main text. That is, \( \bar{p} = (\bar{p}_1, \ldots, \bar{p}_n) \in \bar{P} \) if and only if, for some \( 1 \leq k, l \leq m \), \( \bar{p}_j = \text{sign}(x^k_j - x^l_j) \), with \( j = 1, \ldots, n \), and \( g^k_j - g^l_j < 0 \). From Definition 1, \( \bar{p} \) is inconsistent. Let \( p \not\subseteq \bar{p} \) indicate that the sign pattern \( p \) is not a subpattern of \( \bar{p} \).

Proposition 4. A sign pattern \( p \) is:

a. inconsistent if and only if \( p \subseteq \bar{p} \) for some \( \bar{p} \in \bar{P} \);

b. consistent if and only if \( p \not\subseteq \bar{p} \), \( \forall \bar{p} \in \bar{P} \);

Proof. (a) Since \( \bar{p} \in \bar{P} \) is inconsistent, any \( p \subseteq \bar{p} \) is inconsistent from Proposition 3. Conversely, assume that \( p \) is inconsistent. By Definition 1, there exist indices \( k, l \in \{1, \ldots, m\} \) such that \( g^k - g^l < 0 \) and \( p_j(x^k_j - x^l_j) \geq 0 \), with \( j = 1, \ldots, n \). For the same indices \( k \) and \( l \), define \( \bar{p} = (\bar{p}_1, \ldots, \bar{p}_n) \) by \( \bar{p}_j = \text{sign}(x^k_j - x^l_j) \), with \( j = 1, \ldots, n \). It is easy to see that \( p \subseteq \bar{p} \), and also \( \bar{p} \in \bar{P} \). (b) It follows by negation of (a). \( \square \)
In particular, it follows that \( p = (0,0,\ldots,0) \) is consistent if and only if \( \bar{P} = \emptyset \). The next result proves the correctness of the algorithm for the computation of \( P^* \) (Section 2.2 of the main text).

**Proposition 5.** A sign pattern \( p \) is minimal consistent if and only if it is consistent and no \( p' \) such that \( p' \subseteq p \) and \( p' \neq p \) is minimal consistent.

**Proof.** Assume that \( p \) is minimal consistent. By Definition 2, \( p \) is consistent and any \( p' \subseteq p \) with \( p' \neq p \) is not consistent (hence it is not minimal consistent). Conversely, assume that \( p \) is not minimal consistent. If \( p \) is not consistent the claim is proven. If \( p \) is consistent but not minimal consistent, it must be the case that \( p \neq (0,0,\ldots,0) \) (see the comment after Definition 2). Then, by Definition 2 there must exist a \( p' \subseteq p \), with \( p' \neq p \), that is consistent. If \( p' \) is minimal the claim is proven. Otherwise, again by Definition 2 there must exist a \( p'' \subseteq p' \), with \( p'' \neq p' \), that is consistent. Since the set of subpatterns of \( p \) is finite, this argument can be repeated only a finite number of times, until a minimal consistent pattern is found.

Now we show that the iterative computation of \( P^* \) reported in Section 2.2 of the main text yields the set of minimal consistent sign patterns. Throughout the algorithm, the consistency of a sign pattern \( p \) is assessed in Step III on the basis of Proposition 4, i.e. by verifying that no sign pattern in \( \bar{P} \) is a superpattern of \( p \). Define \( \ell = \max\{C(\bar{p}) : \bar{p} \in \bar{P}\} \) and initialize \( P^* \) to the emptyset. In the light of Proposition 4, all patterns \( p \) with \( C(p) > \ell \) are consistent. Therefore the minimal consistent patterns must have complexity less than or equal to \( \min\{n,\ell + 1\} \). For \( \ell = 0 \), the only sign pattern \( p \) with \( C(p) = \ell \) is \( p = (0,\ldots,0) \). Since \( p \) has no subpatterns other than itself, if \( p \) is consistent, then \( p \) is minimal consistent and is included in \( P^* \). At the beginning of a given iteration \( \ell \), with \( 1 \leq \ell \leq \ell + 1 \), assume that \( P^* \) contains all minimal consistent sign patterns with complexity strictly less than \( \ell \). Let \( p_i \), with \( C(p_i) = \ell_i \), be consistent. In light of Proposition 5, in Step IV, \( p_i \) is declared minimal consistent if and only if \( P^* \) does not contain subpatterns of \( p_i \). If this is the case \( p_i \) is included into \( P^* \). Therefore, at the end of the iteration, \( P^* \) contains all minimal patterns with complexity less than or equal to \( \ell \). In particular, at the end of the last iteration \( (\ell = \min\{n,\ell + 1\}) \), \( P^* \) contains all minimal consistent sign patterns.

### 3 Statistical properties of \( \delta \)

We study the statistical properties of the random variables \( \delta \) defined in Eq. (8) of the main text, and introduce approximate statistics that are used for testing model hypotheses in Step 2 of the identification algorithm.

Fix a synthesis rate function \( g_i \). Let \( (x^k, g_i^k) \), with \( x^k = (x_1^k, \ldots, x_n^k) \), \( g_i^k = g_i(x^k) \) and \( k = 1,\ldots,m \), be a set of data points. Let \( (\tilde{x}, \tilde{g}_i^k) \), with \( \tilde{x} = (\tilde{x}_1, \ldots, \tilde{x}_n) \), be the noisy measurements of \( (x^k, g_i^k) \), \( k = 1,\ldots,m \). It is assumed that

\[
\tilde{x}_i = x_i(t_k) + \epsilon_i^k, \quad \tilde{g}_i^k = g_i(x_i(t_k)) + \epsilon_i^k,
\] 

(6)
with \(e_i^k\) and \(\epsilon_i^k\) mutually uncorrelated jointly Gaussian random variables with zero mean and variance \(v_e(x_i^k) = \text{var}(e_i^k)\) and \(v_e(g_i^k) = \text{var}(\epsilon_i^k)\) possibly depending on \(x_i^k\) and \(g_i^k\). For every \(k\) consider the linear approximation of \(g\) around \(x^k\)

\[
g_i(x) \approx g_i(x^k) + G(x^k)(x - x^k) = g_i^k + G(x^k)(x - x^k)
\]

(7)

where \(G(x^k)^T = [\partial g_i(x)/\partial x_1 \ldots \partial g_i(x)/\partial x_n]_{x=x^k}\) is the gradient of \(g_i\) evaluated at \(x^k\).

**Proposition 6.** Assume that (7) holds as an equality for all \(k = 1, \ldots, m\). Define

\[
w_k = \left[ G(x^k)^T \text{diag}(v_e(x_1^k), \ldots, v_e(x_n^k))G(x^k) + v_e(g_i^k) \right]^{-1}
\]

The random variable

\[
\Delta = \sum_{k=1}^m w_k (\tilde{g}_i^k - g_i(x^k))^2
\]

is distributed as a \(\chi^2(m)\) (Chi-square with \(m\) degrees of freedom).

**Proof.** From equality (7) it follows that

\[
\tilde{g}_i^k - g_i(x^k) = \tilde{g}_i^k - g_i^k - G(x^k)^T (\tilde{x}^k - x^k) = \epsilon_i^k - G(x^k)^T [\epsilon_1^k \ldots \epsilon_n^k]^T.
\]

Since this is a linear function of zero-mean jointly Gaussian variables, for \(k = 1, \ldots, m\), \(\tilde{g}_i^k - g_i(x^k)\) is itself zero-mean Gaussian with variance

\[
\text{var}(\tilde{g}_i^k - g_i(x^k)) = v_e(g_i^k) + G(x^k)^T \text{diag}(v_e(x_1^k), \ldots, v_e(x_n^k))G(x^k) = w_k^{-1}.
\]

Therefore \(\Delta\) is a sum of \(m\) squared, mutually independent, normalized, jointly Gaussian random variables, which proves the claim. \(\square\)

In the main text this result is used to test the hypothesis that an estimated model \(g\) is correct. In light of the proposition above, if a given \(g\) is the true model for the data and \(\Delta\) is the corresponding fitting residual, for any \(\tau > 0\) it holds that \(\text{Prob}(\Delta < \tau) = F_m(\tau)\), where \(F_m\) is the probability distribution function of a \(\chi^2(m)\) random variable. Thus, with confidence level \(\alpha \in (0, 1)\), we may reject the hypothesis that \(g\) is correct if \(\Delta \geq \tau\), with \(\tau = F_m^{-1}(\alpha)\). In our application, for every fixed structure of \(g\), a parameter vector \(\theta\) is estimated via (nonlinear) least squares regression (minimization of \(\Delta\)). In other words, the model to be tested is itself a function of the data. To account for this we borrow a result from (linear) least squares regression [3] and modify the definition of the threshold \(\tau\) to \(\tau = F_{m-|\theta|}^{-1}(\alpha)\), where \(|\theta|\) denotes the number of unknown parameters. This modification is exact only for models \(g\) linear in \(\theta\), but it turned out to be acceptable also for the nonlinear models of our concern. Finally, since \(g_i^k\) and \(x^k\) are not observed directly, in the definition of \(w_k\) we replace \((g_i^k, x^k)\) with \((\tilde{g}_i^k, \tilde{x}^k)\). This approximation and approximation (7) are both most accurate for small values of \(v_e(x_i^k)\) and \(v_e(g_i^k)\).
4 Case study model: artificial network based on a repressilator

The network is depicted in Figure 1; recall from the main text the equations describing the system. The cooperativity coefficients $d$ of all sigmoids are all equally set to 2.1 [2]. The threshold parameters and the initial conditions $x_i(0)$, with $i = 1, \ldots, 6$, were chosen in order to produce permanent oscillations and to prevent identifiability problems. Let $\eta_{ij}$ be the threshold parameter of the sigmoid $\sigma^\pm(x_j)$ appearing in the expression of $g_i$. We set

$$\kappa_{0,i} = 0, \quad \kappa_{1,i} = 15, \quad \gamma_{1,i} = 1, \quad i = 1, \ldots, 6; \quad (8)$$

$$\eta_{ij} = \begin{cases} 0.1, & i = 1, 2, 3; \\ 0.5, & i = 4, 5, 6, \end{cases} \quad j = 1, \ldots, 6; \quad (9)$$

$$x_1(0) = 0.3, \quad x_2(0) = 0.1, \quad x_3(0) = 0.2, \quad x_4(0) = x_5(0) = x_6(0) = 0. \quad (10)$$

We simulated the model for 15 time units (the measurement units used for time and parameters are irrelevant for this numerical case study) and collected 90 equally spaced data samples. The (normalized) concentration profiles are shown in Figure 2.

5 In vivo data for the IRMA network

The data reported in [1] are publicly available for download on the publisher’s website as a supplemental file. Those data consist of different time-series measurements obtained by quantitative real-time RT-PCR under two different experimental conditions. Indeed, experiments were performed by shifting cells either from glucose- to galactose-raffinose-containing medium or from galactose-raffinose- to glucose-containing medium, referred to as switch-on or switch-off
Figure 2: Evolution of the normalized protein concentration levels simulated for the artificial network.
experiments, respectively. In both experimental conditions, data from five independent experiments were collected. In particular, datapoints were obtained by sampling every 20 minutes up to 5 hours and 10 minutes up to 3 hours for the switch-on and the switch-off experiments, respectively. As in [1], we used data averaged over the experimental replicates for testing our method, relying on the corresponding standard errors as an estimate of the noise level affecting the data. Moreover, because of washing and medium shift occurring during the first 10 minutes of each experiment, we discarded the first datapoint of both switch-on and switch-off time-series. In summary, the dataset used in our study comprised 15 samples collected every 20 minutes for the switch-on condition and of 20 samples collected every 10 minutes for the switch-off case. The corresponding in silico synthesis data were obtained by evaluating the synthesis rate functions at the averaged concentration levels according to the model in the Supplemental Results of [1].

References

