We propose a method of Genetic Regulatory Network (GRN) model identification using mathematical programming and global optimization techniques. The problem consists in the estimation of the unknown parameters of a GRN model such that the asymptotic dynamics of the model closely match a set of experimental observations. This problem can be naturally cast as an optimization problem that minimizes a given distance between a set of observed expression patterns and estimated values of the parameters, subject to constraints derived from the algebraic equations that describe the dynamics of the biological system. We apply this approach to the inference of GRNs controlling early flower organ development in the model plant Arabidopsis thaliana.

The GRN model identification problem (GRNMIP): \[
\min_{x} f(x), \quad \text{subject to } g_i(x) \leq 0, \quad (2)
\]
where \( x \in \mathbb{R}^n \) are decision variables, \( f : \mathbb{R}^n \to \mathbb{R} \) is the objective function and \( g_i : \mathbb{R}^n \to \mathbb{R}^m \) is a set of constraints including variable ranges and integerality constraints.

Given \( G, \alpha \) and the sought-for steady configurations \( \phi_i \in \mathbb{P} \), given by the expression data, the GRNMIP consists in finding \( \alpha \) such that \( (G, \alpha, \phi_i, \theta) \) collectively minimize the hamming distance \( D_H(\phi_i) = \sum_{i=1}^n |\phi_i - \phi_k| \) from steady states to data and satisfy the set of steadiness constraints:

\[
\forall i \in V \quad \text{and} \quad j \in \delta^-(i) : \quad \phi_i = \sum_{k \in R} f_k(i)
\]

\[
\pi_k(i)h_j - |V|(1-b_{ij}) \leq \phi_i \leq (\tau_j(i)-1)(1-b_{ij}) + |V|b_{ij}
\]

\[
\theta_{k}\phi_i - \pi_k(i)h_j - |V|(1-b_{ij}) \leq \phi_i \leq (\tau_j(i)-1)(1-b_{ij}) + |V|b_{ij}
\]

Reformulation and Solution: the GRNMIP is a non-convex Mixed Integer Nonlinear Program (MINLP); after standard mathematical manipulations non-linearities reduce to product terms of binary and/or real variables that can be reformulated exactly to a linear one (MILP) [7]. We use the modelling language AMPL [7] to describe the reformulated problem and solve it with CPLEX [8].

The same methodology is used to model other GRNMIP using different reconstruction criteria and/or gene activity conceptual models; in the particular case where the same conceptual model is used but genes states take binary qualitative values, a similar formalization applies but the constraints are simplified.

3. Results

Early stages of plant lateral organ development The method is used to reconstruct GRNs controlling early flower organ development in Arabidopsis. A qualitative boolean model for lateral organ polarity has been identified for which the steady states match precisely observed expression data.

The GRN conceptual model: Let the directed graph \( G = (V,A) \) be the interaction structure, \( V \) being the set of genes and \( A \) the set of pair-wise interactions, \( w : A \to \mathbb{R}_+ \), \( \alpha : A \to [-1,1] \) and \( \theta : A \to \mathbb{R}_+ \), denote the interaction strengths, types and thresholds, respectively, and \( \delta^-(i) = \{j \in V : (j,i) \in A\} \) the input and output neighborhoods. The qualitative activity state of gene \( i \) is denoted by \( \phi_i \in \mathbb{P} \), where \( \mathbb{P} = \{1 \ldots Q_i\} \) and \( Q_i \in \mathbb{N} \) is the number of interactions with different threshold values having gene \( i \) as source. The GRN verifies the state transition condition:

\[
q_i^{t+1} = \sum_{(x,y) \in A} \sum_{\phi_i \in \mathbb{P}} \alpha_{ij} |H(q_i, \pi_j(i), \phi_i - \phi_j)| \quad (1)
\]

\[H(x,\theta) = \{0 \text{ if } x \geq \theta; 0 \text{ if } x < \theta\} \quad \text{and} \quad \tau_j(i) = \sum_{l \in \delta^-(i)} \pi_l(i) \]

\( \theta_{k}\phi_i - \pi_k(i)h_j - |V|(1-b_{ij}) \leq \phi_i \leq (\tau_j(i)-1)(1-b_{ij}) + |V|b_{ij}\)

4. Discussion

The approach we propose focuses on problem modelling rather than on algorithms and has the advantage of generality. The identification method presented doesn’t depend on the particular form of the GRN model used here and can be extended to more intricate gene activation mechanisms. Different inverse problems can be modelled, reformulated and solved with reliable and fast general-purpose algorithms.