





Inferring Small-Scale Maximum-Entropy Genetic Regulatory Networks by Using DE Algorithm

Fu Yin¹ , Jiarui Zhou², Zexuan Zhu³, Xiaoliang Ma³, and Weixin Xie¹ 

¹ College of Electronics and Information Engineering, Shenzhen University, Shenzhen, China
{fuyin, wxxie}@szu.edu.cn

² School of Biosciences, University of Birmingham, Birmingham, UK

³ College of Computer Science and Software Engineering,
Shenzhen University, Shenzhen, China

Abstract. Maximum-entropy genetic regulatory networks (GRNs) have been increasingly applied to infer pairwise gene interactions from biological data. Most maximum-entropy GRNs inferring methods estimate the inverse covariance matrix based on the assumption that the network is sparse and the problem can be approximated via convex optimization. However, the assumption might not be true in reality. To address this issue, in this paper, we propose an adaptive differential evolution (DE) algorithm to directly infer the maximum-entropy GRNs, which is formulated as a constrained optimization problem with the maximum entropy being the objective function and the first and second moments being two penalty terms. A GRN inferred by DE is a fully connected network that can reflect the gene regulatory relations. The experimental results on both simulated and real data suggest that the proposed method is robust in inferring the small-scale maximum-entropy GRNs.

Keywords: Maximum-entropy · Genetic regulatory networks · Constrained optimization · DE algorithm

1 Introduction

In recent years, a large number of methods have been proposed for inferring gene regulatory networks (GRNs) from gene expression data. Correlations and other statistical measures that group genes by profile similarity identify functionally related groups of genes [1, 2]. Much effort has been devoted to inferring GRNs using various modeling approaches, ranging from simple Boolean networks to dynamical models of cellular processes [3, 4]. However, correlation measures do not provide direct insight into the identification of the gene interactions that give rise to the observed expression patterns [5]. For this reason, pairwise maximum-entropy probability models have been introduced to infer the GRNs [6]. The logic of such methods is to determine the probability distribution governing the microarray data where the entropy-reducing constraint their pairwise correlations is faithfully encoded [6]. Consequently, the real-valued maximum-entropy

distribution given first and second moments is found as a Boltzmann-like distribution, which is determined by the mean and the covariance matrix C , and the inverse covariance matrix M (also known as the precision or concentration matrix) [7].

Statistical inference methods using partial correlations in the context of Gaussian graphical models (GGMs) have led to similar results. By assuming that the precision matrix is sparse and the data samples are drawn independently from the same distribution, the most commonly used method to infer the maximum-entropy networks are spectral decomposition [6] and graphical lasso [8].

Although the above methods have been successfully used to estimate the regulatory relationships among genes, their performance may be limited in some areas. The inferring performance of the spectral decomposition method is not good enough when the samples are very few. Although the graphical lasso is better than spectral decomposition on small samples, its GRNs can not reflect gene regulatory relations (expression and repression) and it is difficult how to choose the right lasso penalty to fit the sparseness is very difficult [9].

To address the above problems, in this paper, we proposed a new way to infer the maximum-entropy GRNs, which is formulated as a constrained optimization problem, using a differential evolution (DE) algorithm. DE has achieved widely successes on various complex constrained optimization problems [10]. We set maximum entropy as the objective function and its subject to the constraint of first and second moment as the penalty functions based on the pairwise maximum-entropy probability models. To demonstrate the performance of the proposed method, the method is compared with other state-of-the-art methods on two synthetic datasets and four real-world datasets. The GRNs obtained by the DE are fully connected networks that reflect gene regulatory relations to identify GRNs involved in diverse cellular processes. Experiment results demonstrate that our method outperforms the other two state-of-the-art inferring maximum-entropy GRNs methods on synthetic datasets. In the meantime, the real data results suggest that the proposed approach is robust to inferring small-scale GRNs.

The rest of this paper is organized as follows. In the Sect. 2, we introduce the background of the maximum-entropy GRNs. In the Sect. 3, we present our proposed framework in detail. In the Sect. 4, we describe the performance of our method on synthetic and real-world datasets. Finally, Sect. 5 concludes this work.

2 Maximum-Entropy GRNs

Pairwise associations between genes can be determined by gene expression and are commonly estimated by the sample Pearson correlation coefficient computed for each pair of genes. However, the Pearson correlation is a misleading measure for direct dependence as it only reflects the association between two genes while ignoring the influence of the remaining ones. Therefore, the relevance network approach is not suitable to deduce direct interactions from a dataset [11].

To address these problems, maximum-entropy GRNs is proposed. It relies on Boltzmann’s concept of entropy maximization to support statistical inference with minimal reliance on the form of missing information, which can remove the variational effect due to the influence of the remaining variables.

Let the state vector $x = (x_1, \dots, x_N)$ denote the expression levels of the N genes that are probed in a microarray experiment, and a series of T measurements then has associated with it T distinct state vectors. Let $\rho(x)$ denote the probability that the genome is in the arbitrary state x . We determine $\rho(x)$ by maximizing the Shannon entropy.

$$S = -\rho(\vec{x}) \ln(\vec{x}) \tag{1}$$

subject to the $\rho(x)$ is normalized

$$\sum_{\vec{x}} \rho(\vec{x}) = 1 \tag{2}$$

first moment, $\langle x_i \rangle$, and second moment, $\langle x_i, x_j \rangle$

$$\langle x_i \rangle = \sum_{\vec{x}} \rho(\vec{x}) x_i = \frac{1}{T} \sum_{k=1}^T x_i^k \tag{3}$$

$$\langle x_i, x_j \rangle = \sum_{\vec{x}} \rho(\vec{x}) x_i x_j = \frac{1}{T} \sum_{k=1}^T x_i^k x_j^k \tag{4}$$

Equation (2) provides the normalization condition that the probabilities of all observable states sum to 1. Equations (3) and (4) ensure that the distribution $\rho(x)$ preserves the mean expression level of each gene and the correlations between genes. This procedure leads to a Boltzmann-like distribution:

$$\rho(x) \sim e^{-H}$$

where

$$H = \frac{1}{2} \sum_{ij} x_i M_{ij} x_j$$

Note that, because $\langle x_i, x_j \rangle = \langle x_j, x_i \rangle$, the number of constraints is $1 + N + N(N + 1)/2$. In the same reason $M_{ij} = M_{ji}$, the number of M_{ij} should be estimated is $N(N + 1)/2$.

The elements of the matrix M are the effective pairwise gene interactions that reproduce the gene profile covariances exactly while maximizing the entropy of the system. The intensity and type of an element M_{ij} : a positive value denotes expression (facilitation) and a negative value denotes repression while a zero (0) value implies that there is no interaction between i and j [6, 11].

The matrix of M can be obtained by inverting the matrix of their covariances C . However, in the high dimensional setting where the number of features p is larger than the number of observations n , the empirical covariance matrix C is singular and so can not be inverted to yield an estimate of M . If $p \approx n$, then even if C is not singular the estimate for C will suffer from very high variance [5].

Spectral decomposition and graphical lasso are proposed to get around this problem by estimating the inverse covariance matrix based on the assumption that the network is sparse and the problem can be approximated via convex optimization. However, the assumption might not be true in reality. The performance of spectral decomposition is not good enough at solving the small sample problem. Although the graphical lasso has better inferring performance, its GRNs cannot reflect the gene regulatory relations (expression and repression). Also, it is difficult to choose a right lasso penalty to control the sparseness of the inferred networks [9].

3 Method

To address the above problems, in this paper, we proposed a new way to infer the maximum-entropy GRNs, a constrained optimization problem, by using an adaptive differential evolution (DE) algorithm. DE as a nature-inspired method has become a more feasible and popular choice among researchers due to their competitive performance on complex search spaces to address the constrained optimization problems [10].

However, the performance of the classic DE is still entirely dependent on control parameters and mutation strategies to both experimental studies and theoretical analyses [12]. The adaptive and self-adaptive DE algorithms have shown faster and more reliable convergence performance than the classic DE algorithms for many benchmark problems.

For this reason, the main objectives of this work are three-fold. First, we set maximum entropy as the objective function and its subject to the constraint of first and second moment as the penalty functions. Second, the Probability Matching (PM) method is integrated into DE to implement the adaptive strategy selection. Third, the JADE [13] is used to set controls mutation factor F and crossover probability CR in an adaptive manner. Details behind algorithm are elucidated as follows.

3.1 Problem Formulation

The general form of the constrained optimization problem will be expressed as follows:

$$\begin{aligned} & \min f(x) \\ \text{s.t. } & g_i(x) \leq 0, \quad i = 1, 2, \dots, j \\ & h_i(x) = 0, \quad i = j + 1, j + 2, \dots, m \end{aligned} \quad (5)$$

where $x = (x_1, x_2, \dots, x_n)$ are the decision variables of the objective function $f(x)$, g_i is an inequality constraint describing the variable, which role of the inequality constraint is to form the search area in the feasible domain. h_i is equality constraint that forms a boundary value condition in the feasible domain, which role is to control the boundary of the search area. Normally, we define the objective function as $\min f(x)$ and penalty function as

$$G_i(x) = \begin{cases} \max\{0, g_i(x)\}, & 1 \leq i \leq j \\ \max\{0, |h_i(x) - \delta|\}, & j + 1 \leq i \leq m \end{cases}$$

$$G(x) = \sum_{i=1}^m G_i(x) \quad (6)$$

where δ is a small positive tolerance value. The final fit function would be like

$$\text{Fit} = f(x) + \sigma G(x) \quad (7)$$

where σ is the punitive coefficient.

In this paper, Eq. (1) is set to the objective function and its subject to the constraint of first and second moment Eqs. (2–4) are set to the penalty functions. Then the optimization problems (1–4) can convert into the optimization problem (7) that will be minimized by DE. The initial population $\{x_i = (x_{1,i,0}, x_{2,i,0}, \dots, x_{D,i,0}) | i = 1, 2, \dots, NP\}$ is randomly generated according to a uniform distribution $[-1, 1]$, where $D = N(N + 1)/2$ (the number of M_{ij}) is the dimension of the problem and NP is the population size.

3.2 Strategy Selection: Probability Matching

Suppose there are $K > 1$ strategies in the pool $A = \{a_1, \dots, a_K\}$ and a probability $P(t) = \{p_1(t), \dots, p_k(t)\} (\forall t : p_{min} \leq p_i(t) \leq 1; \sum_{i=1}^K p_i(t) = 1)$. In this work, the PM technique is used to adaptively update the probability $p_a(t)$ of each strategy a based on its known performance and updated by the rewards received. Denote $r_a(t)$ as the reward that a strategy a receives after its application at time t . $q_a(t)$ is the empirical estimate of a strategy a , that is updated as follows [14]:

$$q_a(t + 1) = q_a(t) + \alpha[r_a(t) \cdot q_a(t)] \quad (8)$$

where $\alpha \in (0, 1]$ is the adaptation rate. Based on this quality estimate, the PM method updates the probability $p_a(t)$ of applying each operator as follows:

$$p_a(t + 1) = p_{min} + (1 - K \cdot p_{min}) \frac{q_a(t+1)}{\sum_{i=1}^K q_i(t+1)} \quad (9)$$

where $p_{min} \in (0, 1)$ is the minimal probability value of each strategy, used to ensure that no operator gets lost [13].

In order to assign the credit for each strategy, we adopt the relative fitness improvement η_i proposed in [12] as follows:

$$\eta_i = \frac{\delta}{cf_i} \cdot |pf_i - cf_i| \quad (10)$$

where $i = 1, \dots, NP$. δ is the fitness of the best-so-far solution in the population. pf_i and cf_i are the fitness of the target parent and its offspring, respectively. If no improvement is achieved, a null credit is assigned.

Denote S_a as the set of all relative fitness improvements achieved by the application of a strategy a ($a = 1, \dots, K$) during generation t . At the end of the generation, a unique

reward is used to update the quality measure kept by the PM method (Eq. 9). The credit assignment is as follows [14]:

$$r_a(t) = \frac{\sum_{i=1}^{|S_a|} S_a(i)}{|S_a|} \tag{11}$$

where $|S_a|$ is the number of elements in S_a . If $|S_a| = 0$, $r_a(t) = 0$.

In DE, many schemes have been proposed, applying different mutation strategies and/or recombination operations in the reproduction stage [15]. In order to constitute the strategy pool used in this work, we have chosen four strategies: ‘DE/rand/1’, ‘DE/rand/2’, ‘DE/rand-to-best/1’ and ‘DE/current-to-rand/1’.

3.3 Parameter Adaptation

The parameter adaptation is similar to JADE. At each generation g , the crossover probability CR_i of each individual x_i is independently generated according to a normal distribution of mean μ_{CR} and standard deviation 0.1 as

$$CR_i = \text{randn}_i(\mu_{CR}, 0.1) \tag{12}$$

and then truncated to $[0, 1]$ [13]. Denote S_{CR} as the set of all successful crossover probabilities CR_i ’s at generation g . The mean μ_{CR} is initialized to be 0.5 and then updated at the end of each generation as

$$\mu_{CR} = (1 - c) \cdot \mu_{CR} + c \cdot \text{mean}_A(S_{CR}) \tag{13}$$

where c is a positive constant between 0 and 1 and $\text{mean}_A(\cdot)$ is the usual arithmetic mean.

Similarly, at each generation g , the mutation factor F_i of each individual x_i is independently generated according to a Cauchy distribution with location parameter μ_F and scale parameter 0.1 as:

$$F_i = \text{randn}_i(\mu_F, 0.1) \tag{14}$$

and then truncated to be 1 if $F_i \geq 1$ or regenerated if $F_i \leq 0$ [13]. Denote S_F as the set of all successful mutation factors in generation g . The location parameter μ_F of the Cauchy distribution is initialized to be 0.5 and then updated at the end of each generation is as follows:

$$\mu_F = (1 - c) \cdot \mu_F + c \cdot \text{mean}_L(S_F) \tag{15}$$

where $\text{mean}_L(\cdot)$ is the Lehmer mean [13].

3.4 Optimization Work Flow

The probability matching method and JADE are used respectively to select the mutate strategy and set the parameter adaptively. In Algorithm 1, the use of our adaptive DE for inferring Maximum-Entropy GRNs is illustrated.

Algorithm 1 The DE algorithm with adaptive strategy selection and JADE parameter adaptation based on probability matching

Input: search space $[-1, 1]$, objective function f Eq.(1), penalty function G Eqs.(2-4)

Output: optimal solution

```

1: Generate the initial population
2: Evaluate the fitness for each individual
3: For each strategy  $a$ , set  $q_a(t) = 0$  and  $p_a(t) = 1/K$ 
4: while The halting criterion is not satisfied do
5:   for  $i = 1$  to  $NP$  do
6:     Select the strategy  $SI_i$  based on its probability
7:     Select uniform randomly  $r_1 \neq r_2 \neq r_3 \neq r_4 \neq r_5 \neq i$ 
8:      $j_{rand} = \text{randn}(1, D)$ 
9:     Using Eqs. (12), (14) generating  $F_i^a$  and  $CR_i^a$ 
10:    for  $j = 1$  to  $D$  do
11:      if  $\text{randj}[0, 1] < CR_i^a$  or  $j = j_{rand}$  then
12:        Using strategy  $SI_i$  generating  $u_{ij}$ 
13:      end if
14:    end for
15:  end for
16:  for  $i = 1$  to  $NP$  do
17:    Evaluate the offspring  $u_i$ 
18:    if  $f(u_i) \geq f(x_i)$  then
19:      Calculate  $\eta_i$  using Eq.(10)
20:      Replace  $x_i$  with  $u_i$ 
21:    else
22:      Set  $\eta_i = 0$ 
23:    end if
24:     $S_{SI_i} \leftarrow \eta_i$ 
25:  end for
26:  Calculate the reward  $r_a(t)$  for each strategy using Eq.(11)
27:  Update the quality  $q_a(t)$  for each strategy using Eq.(8)
28:  Update the probability  $p_a(t)$  for each strategy using Eq.(9)
29:  Update the  $F_i^a$  and  $CR_i^a$  using Eqs.(13),(15)
30:   $t = t + 1$ 
31: end while

```

4 Experimental Results

Let θ_{ij} and $\hat{\theta}_{ij}$ denote the elements in the true GRNs and the inferred GRNs, respectively. Whether the absolute value of a particular element is 0 or 1 can be evaluated by a threshold defined for the purpose of inclusion of an interaction in a GRN. An edge can be characterized into four types: true positive (TP), false positive (FP), true negative (TN), and false negative (FN), with their definitions as follows:

TP: if $\theta_{ij} = 1$ and $\hat{\theta}_{ij} = 1$; TN: if $\theta_{ij} = 0$ and $\hat{\theta}_{ij} = 0$.

FP: if $\theta_{ij} = 0$ and $\hat{\theta}_{ij} = 1$; FN: if $\theta_{ij} = 1$ and $\hat{\theta}_{ij} = 0$.

Then the metrics based on which the proposed methodology can be evaluated. (1) True Positive Rate (TPR)/Recall: this signifies the fraction of the total number of existing edges in the original network, correctly predicted in the inferred GRNs; (2) False Positive Rate (FPR)/Complimentary Specificity: this signifies the fraction of the total number of nonexistent edges, incorrectly predicted in the inferred GRNs; (3) Positive Predictive Value (PPV)/Precision: this signifies the fraction of the total number of inferred edges, which is correct. (4) F-Score: this signifies the harmonic mean of the precision and recall.

4.1 Simulation Studies

We first build a random ER random network denoted by its adjacency binary matrix M with non-zero element substituted by a uniform distribution value on $[-0.6, -0.3] \cup [0.3, 0.6]$. To ensure the positive definiteness of the covariance matrix, the real precision matrix Θ is set as

$$\Theta = M + \sigma I$$

σ is the absolute value of the eigenvalues of M , and I is an identity matrix. After this procedure, the synthetic gene expression data could be generated with zero means and covariance $C = \Theta^{-1}$. In order to test the performance of our method with the other two state-of-art method which can deal with the small sample problem ($p \approx n$). We generate two small-scale groups of samples (group 1 with $n = 5$ samples and $p = 5$ genes; group 2 with $n = 10$ samples and $p = 10$ genes) to simulate the small-scale GRNs. 10 random datasets are generated for the above two groups. The empirical covariance matrix C in each dataset is singular and so can not be inverted to yield an estimate of M .

The gene interaction network comprising the genes showing the strongest couplings is highly interconnected. For this reason, we choose the top 20% strongest pairwise interaction to identify the most consistent predicted edges for the construction of the final GRNs.

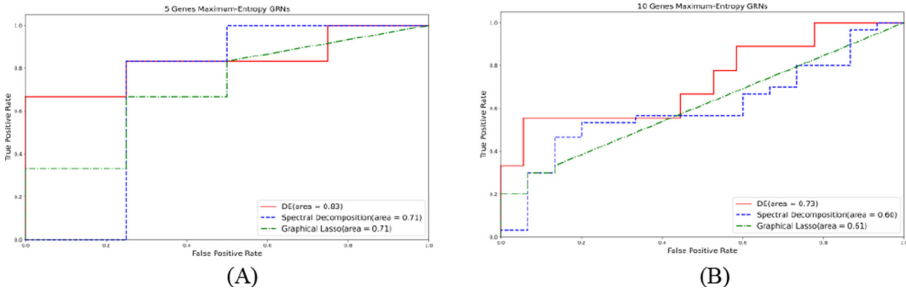


Fig. 1. The experiment results on two small-scale groups of samples: (A) 10 datasets of 5 genes network with 5 samples. (B) 10 datasets of 10 genes network with 10 samples.

Figure 1 presents the average performance on two different scale networks datasets. For graphical lasso, its sparsity-controlling parameter is chosen automatically by cross-validation. In particular, we run 10 times of the DE on each dataset and take the best optimization result as the final inferring result.

We can find that our method dominates the other two methods on inferring maximum-entropy GRNs. As the number of genes increases, there is a degradation on the performance of all three methods, our method could still achieve competitive performance with the other two comparative methods. The results suggest that our approach can effectively solve the small sample problem while meantime reflecting the gene regulatory relations.

4.2 Real Data Analysis

In this section, the proposed method for inferring maximum-entropy GRNs has been employed to identify the causal relationships among the genes from an in vivo (experimental) microarray dataset. The said dataset summarizes the dynamics of the well illustrated transcriptional network involved in the SOS DNA repair mechanism of E. coli studied experimentally by Ronen et al. [16]. The study included eight genes heavily involved in the SOS repair mechanism: *recA*, *lexA* (the master repressor), *uvrA*, *uvrD*, *uvrY*, *umuD*, *ruvA*, and *polB*. The original network has been shown in Fig. 2.

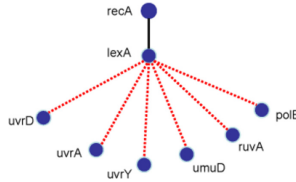


Fig. 2. The original structure of the SOS DNA repair transcriptional network of E.coli. The solid black lines denote activation, and the dashed red lines denote repression.

We choose the absolute value of $M_{ij} > 0.7$ as the inferred Maximum-Entropy GRNs for the 4 datasets. Table 1 displays a comparison of the statistical properties of the inferred GRNs with those presented in recent investigative work [17–19] for different experimental datasets. Table 2 shows the top 3 genes in our inferred GRNs.

Table 1. Comparison of results obtained from the E.coli experiments with those presented in a recent investigative work.

Dataset	Precision				Recall				F-score			
	[17]	[18]	[19]	DE	[17]	[18]	[19]	DE	[17]	[18]	[19]	DE
1	0.23	0.36	0.44	0.50	0.43	0.71	1	0.71	0.30	0.48	0.61	0.59
2	0.58	0.26	0.30	0.45	1	0.57	1	0.71	0.73	0.36	0.46	0.55
3	0.31	0.38	0.41	0.56	0.57	0.71	1	0.71	0.40	0.50	0.58	0.63
4	0	0.27	0.25	0.43	0	0.43	0.57	0.43	0	0.33	0.35	0.43
Mean	0.28	0.32	0.35	0.50	0.5	0.61	0.9	0.64	0.36	0.42	0.49	0.55

Experiment results show that our method performs better in Precision than the other methods in all experiments except the method [17] in experiment 2. However, the method [17] fails to identify any true positive in experiment 4.

We have to concede that the method [19] has a higher Recall than our GRNs, but its Precision is significantly less than our method. When viewed from the aspect of F-score, our method performs better by comprehensive considering of the Precision and Recall in all 4 experiments.

In Table 2, it should be noticed that *lexA* (the master gene) is the node with the highest degree in our inferred GRNs in all four experiments, so our GRNs find the master gene correctly. In the meantime, the *lexA*—*uvrY* and *lexA*—*polB* are not predicted in our GRNs except experiment 2.

Table 2. Top 3 nodes with the highest degree in our inferred GRNs

Dataset	Rank1	Rank2	Rank3
1	<i>lexA</i>	<i>uvrA</i>	<i>recA</i>
2	<i>lexA</i>	<i>uvrA</i>	<i>recA</i>
3	<i>lexA</i>	<i>uvrA</i>	<i>uvrD</i>
4	<i>lexA</i>	<i>recA</i>	<i>ruvA</i>

5 Conclusion

In this paper, we propose a DE based inferring method that can effectively estimate maximum-entropy GRNs. Unlike the other inferring methods of inverse the covariance we consider the inferring maximum-entropy GRNs as a constrained optimization problem and the best individuals searched by our method is the inferred maximum-entropy GRNs.

First, we assume that the maximum-entropy distribution is a Boltzmann-like distribution. Under this assumption, we set maximum entropy as the objective function and its subject to the constraint of first and second moment as the penalty functions. Then the probability matching method and JADE are used respectively to select the mutate strategy and set the parameter adaptively to improve the success rate of the algorithm.

The GRNs resulting from the DE is a fully connected network that fulfills the maximum-entropy GRNs reflecting gene regulatory relations (expression and repression), which graphical lasso can not. For this reason, we can identify connections between genes involved in diverse cellular processes by choosing the different degrees of pairwise interactions.

It outperforms the other two state-of-the-art inferring maximum-entropy GRNs methods on synthetic datasets. In the meantime, the real data results suggest that it can find the master gene, so the proposed approach is robust to inferring small-scale GRNs.

The performance of nature-inspired algorithms often deteriorates rapidly as the dimensionality of the problem increases. Small-scale and large-scale constrained optimization are two completely different problems. There are too few true predictions and a large number of incorrect predictions [19]. Thus, the methodology implemented in this paper needs to be enriched further by studying its performance in larger networks. This provides a vital scope for further research.

References

1. Saldanha, A.J., Brauer, M.J., Botstein, D.: Nutritional homeostasis in batch and steady- state culture of yeast. *Mol. Biol. Cell* **15**, 4089–4104 (2004)
2. Perou, C.M., et al.: Molecular portraits of human breast tumours. *Nature* **406**, 747–752 (2000)
3. Shmulevich, I., Dougherty, E.R., Kim, S., Zhang, W.: Probabilistic Boolean networks: a rule-based uncertainty model for gene regulatory networks. *Bioinformatics* **18**, 261–274 (2002)
4. Akutsu, T., Miyano, S., Kuhara, S.: Algorithms for identifying Boolean networks and related biological networks based on matrix multiplication and fingerprint function. *J. Comput. Biol.* **7**, 331–343 (2000)
5. Danaher, P., Wang, P., Witten, D.M.: The joint graphical lasso for inverse covariance estimation across multiple classes. *J. Roy. Stat. Soc. Ser. B Stat. Methodol.* **76**, 373 (2014)
6. Lezon, T.R., Banavar, J.R., Cieplak, M., Maritan, A., Fedoroff, N.V.: Using the principle of entropy maximization to infer genetic interaction networks from gene expression patterns. *Proc. Natl. Acad. Sci.* **103**, 19033–19038 (2006)
7. Stein, R.R., Marks, D.S., Sander, C.: Inferring pairwise interactions from biological data using maximum-entropy probability models. *PLoS Comput. Biol.* **11**, e1004182 (2015)
8. Friedman, J., Hastie, T., Tibshirani, R.: Sparse inverse covariance estimation with the graphical lasso. *Biostatistics* **9**, 432–441 (2008)
9. Wu, N., et al.: Weighted fused pathway graphical lasso for joint estimation of multiple gene networks. *Front. Genet.* **10**, 623 (2019)
10. Noman, N., Iba, H.: Inferring gene regulatory networks using differential evolution with local search heuristics. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **4**, 634–647 (2007)
11. De Martino, A., De Martino, D.: An introduction to the maximum entropy approach and its application to inference problems in biology. *Heliyon* **4**, e00596 (2018)
12. Qian, W.: Adaptive differential evolution algorithm for multiobjective optimization problems. *Appl. Math. Comput.* **201**, 431–440 (2008)
13. Zhang, J., Sanderson, A.C.: JADE: adaptive differential evolution with optional external archive. *IEEE Trans. Evol. Comput.* **13**, 945–958 (2009)
14. Thierens, D.: An adaptive pursuit strategy for allocating operator probabilities. In: *Proceedings of the 7th Annual Conference on Genetic and Evolutionary Computation*, pp. 1539–1546 (2005)
15. Price, K., Storn, R.M., Lampinen, J.A.: *Differential Evolution: A Practical Approach to Global Optimization*. Springer, Heidelberg (2006). <https://doi.org/10.1007/3-540-31306-0>
16. Ronen, M., Rosenberg, R., Shraiman, B.I., Alon, U.: Assigning numbers to the arrows: parameterizing a gene regulation network by using accurate expression kinetics. *Proc. Natl. Acad. Sci.* **99**, 10555–10560 (2002)
17. Kentzoglanakis, K., Poole, M.: A swarm intelligence framework for reconstructing gene networks: searching for biologically plausible architectures. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **9**, 358–371 (2011)
18. Xu, R., Wunsch, D., II., Frank, R.: Inference of genetic regulatory networks with recurrent neural network models using particle swarm optimization. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **4**, 681–692 (2007)
19. Khan, A., Mandal, S., Pal, R.K., Saha, G.: Construction of gene regulatory networks using recurrent neural networks and swarm intelligence. *Scientifica* **2016** (2016)