RESEARCH

Reverse engineering gene regulatory network based on complex-valued ordinary differential equation model

Bin Yang[1], Wenzheng Bao[2*] , Wei Zhang[1], Haifeng Wang[1], Chuandong Song[1], Yuehui Chen[3] and Xiuying Jiang[1]

*Correspondence:
baowz55555@126.com
[2] School of Information and Electrical Engineering, Xuzhou University of Technology, Xuzhou 221018, China
Full list of author information is available at the end of the article

Abstract

Background: The growing researches of molecular biology reveal that complex life phenomena have the ability to demonstrating various types of interactions in the level of genomics. To establish the interactions between genes or proteins and understand the intrinsic mechanisms of biological systems have become an urgent need and study hotspot.

Results: In order to forecast gene expression data and identify more accurate gene regulatory network, complex-valued version of ordinary differential equation (CVODE) is proposed in this paper. In order to optimize CVODE model, a complex-valued hybrid evolutionary method based on Grammar-guided genetic programming and complex-valued firefly algorithm is presented.

Conclusions: When tested on three real gene expression datasets from *E. coli* and Human Cell, the experiment results suggest that CVODE model could improve 20–50% prediction accuracy of gene expression data, which could also infer more true-positive regulatory relationships and less false-positive regulations than ordinary differential equation.

Keywords: Gene regulatory network, Complex-valued ordinary differential equation, Grammar-guided genetic programming, Firefly algorithm

Background

The aim of gene regulatory network (GRN) research is to explain comprehensively the generation process and complex regulatory relationships of genes and their products in biological tissues from the perspective of system scale [1–3]. GRN is an important way to cope with various internal and external stimuli by synergizing the functions of genes. GRN is a complex continuous and dynamic system, in which gene regulation is a dynamic event that changes with time and environment [4–6]. Gene expression data

could reflect the regulatory orders of genes and the important information of regulatory objects [7]. With the development of DNA microarray or chip technology, a large amount of gene expression profiles have been generated, which provide a condition for identifying gene regulatory networks using computational models [8–10].

Some computational approaches have been developed to describe biology networks [11–17], especially the regulatory relationships between genes with gene expression time series, such as directed graph [18], Boolean network [3, 19], Bayesian network [20, 21], differential equation [22], neural network [23–25], stochastic equation [26, 27], etc. Differential equation model is more conducive to describing the concentration evolution of biological macromolecules such as RNA and protein over time. Therefore, this model has been widely utilized in pharmacokinetics, enzymology and gene regulatory network construction [28–33]. For each target gene, its corresponding differential equation is identified, in which the dependent variables (genes) indicate the corresponding regulatory factors of target gene. Hohm and Zitzler resolved the issue of how to optimize the parameters of ordinary differential equation (ODE) for GRN inference [34]. Tian et al. proposed stochastic delay differential equations to describe the time-delayed regulatory relationships in GRN [35]. Gebert and Jong proposed piecewise linear differential equation to identify the most relevant regulating interactions in GRN [36, 37]. Zhang et al. proposed single-index ODE model and clipped absolute deviation penalty penalized function to infer network structure [38]. Matsumoto et al. proposed ODE models to identify GRN with single-cell RNA-Seq data [39]. Some special nonlinear ODE models also have been proposed to infer GRN, such as S-system model [40–42].

Recently the complex versions of many models have been proposed due to their potential to optimize more easily, better generalization characteristics, faster learning and higher noise-robust memory mechanisms [43–45]. You and Hong presented multilayer feedforward neural network (MFNN) with complex-valued activation function to model QAM signals of different constellation sizes [46]. Deng et al. also proposed complex-valued radial basis function neural network (CVRBFNN) to deal with QAM signals and the results showed that CVRBFNN performed better than functional link artificial neural network [47]. Hu et al. discussed the global stability of the complex-valued and delayed version of recurrent neural network [48]. Trabelsi et al. proposed complex-valued convolutional feed-forward networks [49]. Yang et al. proposed a novel complex-valued method based on mathematical expression to resolve real-valued prediction and classification problems [50].

In order to forecast gene expression data and identify GRN accurately, complex-valued ordinary differential equation (CVODE) is presented in this paper. Grammar-guided genetic programming (GGGP) is utilized to evolve the structure of CVODE and complex-valued firefly algorithm (CFA) is proposed to search the optimal complex-valued parameters of model. Three real gene expression datasets are applied to test the inference performances of our proposed methods.

Results and discussion

Three real gene expression data sets are utilized to test the performance of our proposed method. All the real gene expression data need to be normalized, which is defined as follows.

$$g' = \frac{g - g_{\min}}{g_{\max} - g_{\min}} \tag{1}$$

where $g_{\min}$ and $g_{\max}$ are the minimum and maximum of dataset, respectively.

In our experiment, ODE and CVODE models are both applied to predict gene expression data and infer gene regulatory networks with three real gene expression datasets. Our proposed hybrid evolutionary algorithm is utilized to infer ODE and CVODE models and the parameters are the same. In GGGP, population size is set to 50, crossover rate is set to 0.9 and mutation rate is set to 0.1. In CFA, population size is set as 100, and step size is set as 0.02. The parameters are selected according to previous research results [51–53]. Sensitive (S_n) and Specificity (S_p) are utilized to evaluate the inferred gene regulatory networks. S_n represents the proportion of the positive samples to be identified correctly in all positive ones. S_p denotes the proportion of the negative samples to be identified correctly in all negative ones.
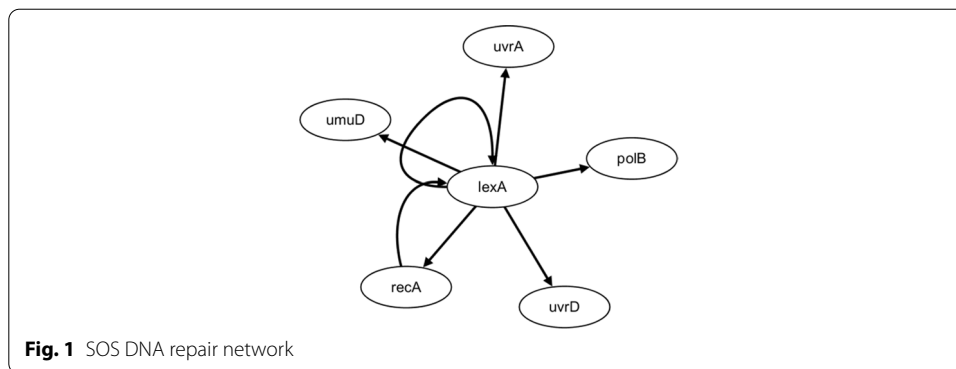
SOS DNA repair network

SOS DNA damage repair network is selected as the first real gene regulatory network, which contains six main genes (**uvrD**, **lexA**, **umuD**, **recA**, **uvrA** and **polB**). The real structure of SOS DNA damage repair network is depicted in Fig. 1. The gene expression data used in this experiment are from Uri Alon [54, 55]. The DNA repair process of *E. coli* is observed by ultraviolet irradiation. The expression levels of six genes at different time points are measured.

The datasets from the first three experiments are utilized to optimize the CVODE models. The last dataset is utilized to test. With hybrid evolutionary algorithm the corresponding CVODE models of six genes are described as follows. Variables $x_1, x_2, \ldots, x_6$ represent six genes: **uvrD**, **lexA**, **umuD**, **recA**, **uvrA** and **polB**, respectively.

$$\frac{dx_1}{dt} = (-0.1523 - 0.1507i)\frac{x_1}{x_5} + (-2.4023 + 0.0498i) \times (x_1 - x_6) + (0.2349 + 0.1709i)x_2$$

$$\frac{dx_2}{dt} = (0.3382 + 3.3836i) \times (1 - x_1) + (4.4767 + 0.0309i)x_6 + (0.0773 - 0.0837i) \times \cos(x_4 - x_2)$$

$$\frac{dx_3}{dt} = (0.222 + 0.8159i)x_3x_1 + (0.906 + 1.638i)\cos(x_3) + (0.6876 - 0.514i) \times (x_2 + x_1)$$

$$\frac{dx_4}{dt} = (4.396 + 2.799i) \times (2x_2 - x_6) + (-9.0003 + 2.4687i) \times (\frac{x_1}{x_4})^2 + (-7.023 + 10.339i)$$
$$\times (\cos(x_3) - x_3) + (-3.2559 + 2.3771i) \times (x_4 - \frac{x_6}{x_4}) + (0.4655 - 0.1199\,i)\cos(x_1 - x_5)$$

$$\frac{dx_5}{dt} = (-0.2329 + 0.7256i)\cos(\sin(x_1)) + (0.7061 + 0.0325i)\cos(x_5 - x_2) + (0.0241 - 0.5961i)\cos(\cos(x_3))$$

$$\frac{dx_6}{dt} = (0.0461 + 0.0237i)x_3x_2 - (0.2843 + 0.0345i)x_6^3$$

$$\tag{2}$$

ODE models are also utilized to infer SOS network. The predicted results of ODE and CVODE for six genes are depicted in Fig. 2. From Fig. 2, it could be seen that the results of CVODE predicted are closer to real ones than ODE. In order to compare clearly, the prediction RMSE values of ODE and CVODE for six genes are listed in Table 1, which reveals that CVODE model has smaller prediction RMSE values than ODE for six genes. On average CVODE could decrease 34.4% RMSE value.
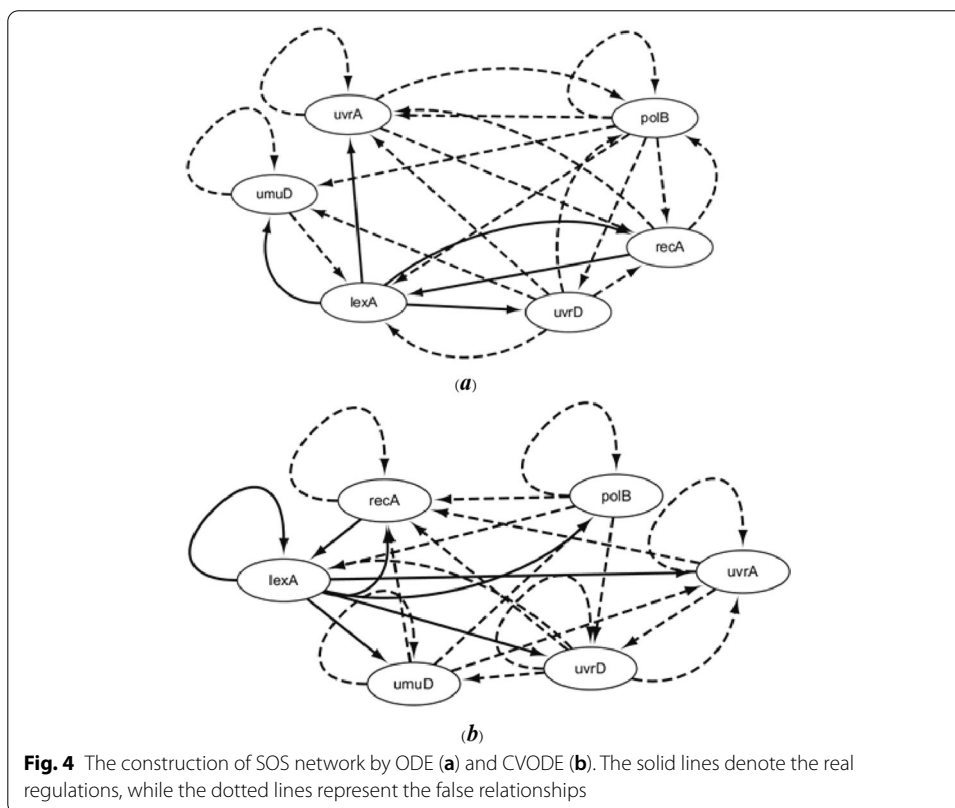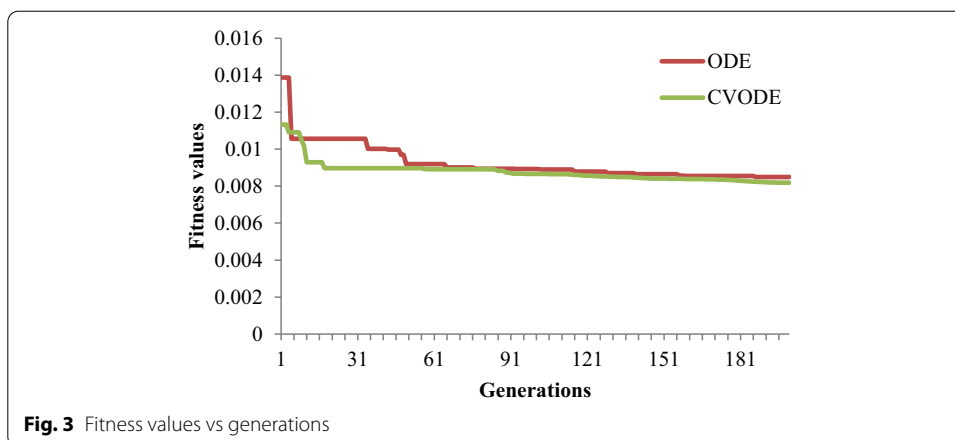
Fig. 1 SOS DNA repair network
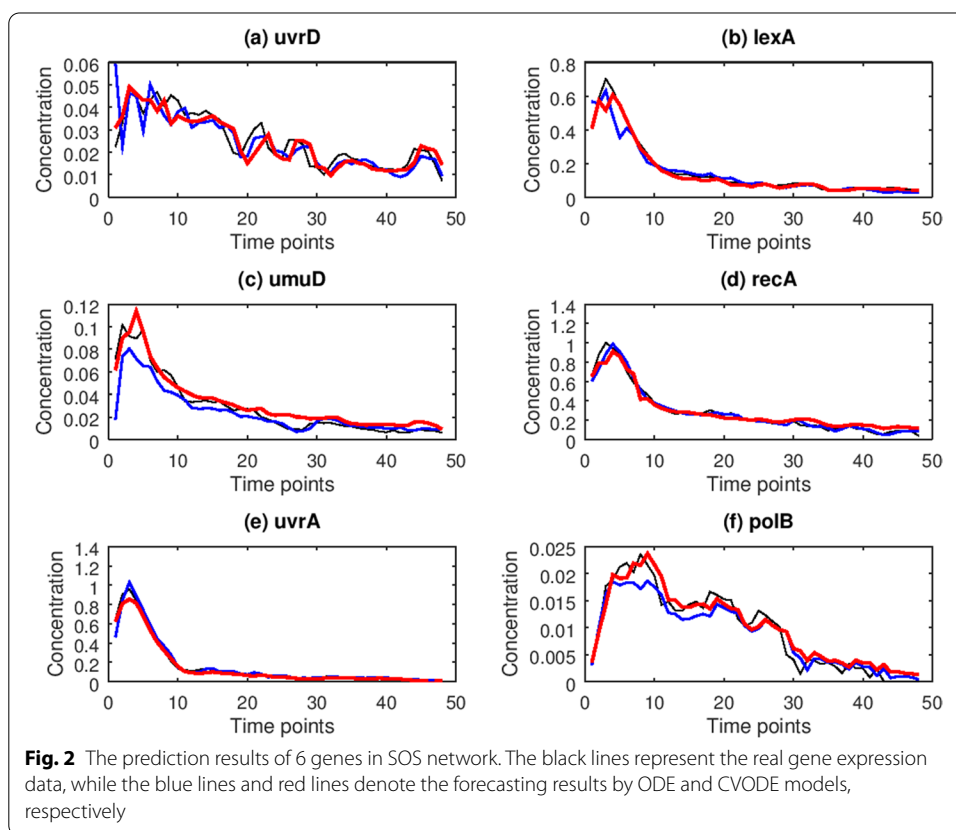
Fig. 2 The prediction results of 6 genes in SOS network. The black lines represent the real gene expression data, while the blue lines and red lines denote the forecasting results by ODE and CVODE models, respectively

Table 1 Prediction RMSE values of ODE and CVODE for six genes

Genes	ODE	CVODE
uvrD	0.014345	**0.007601**
lexA	0.075643	**0.072955**
umuD	0.017476	**0.008536**
recA	0.164121	**0.076576**
uvrA	0.168299	**0.122481**
polB	0.004601	**0.003229**
Averaged	0.074081	**0.048563**

The 'bold values' represent the better results between two methods

Fig. 3 Fitness values vs generations

(*a*)

(*b*)

Fig. 4 The construction of SOS network by ODE (**a**) and CVODE (**b**). The solid lines denote the real regulations, while the dotted lines represent the false relationships

Table 2 Performances of ODE and CVODE for construction of SOS network

Methods	S_n	S_p
CVODE	**1**	**0.4138**
ODE	0.7143	0.3793

The 'bold values' represent the better results between two methods

In order to test the modeling performance of CVODE further, the fitness values of ODE and CVODE against generations are plotted in Fig. 3. From Fig. 3, we can see

that the hybrid evolutionary algorithm could search the optimal CVODE model with about 20 generations, while the optimal ODE model is found with about 50 generations. Thus CVODE model has smaller fitness value and gains the optimal solution with fewer generations than ODE model.

The inferred SOS networks by ODE and CVODE are depicted in Fig. 4. Compared Fig. 1 with Fig. 4, it could be seen that CVODE model could identify 7 true-positive (real) regulations and 17 false-positive regulatory relationships, while ODE model could identify 5 true-positive relationships and 18 false-positive relationships. The performances of S_n and S_p of SOS networks inferred by ODE and CVODE are listed in Table 2. In terms of S_n, CVODE is 39.99% higher than ODE. In terms of S_p, our method is 9.1% higher than ODE. The results reveal that our method could identify all true-positive regulatory relationships and less false-positive regulations than ODE model.

Human cell time-series data

Cell cycle refers to the whole the process from the beginning of one cell division to the end of the next division. The second real gene expression time series data are from the genes periodically expressed, which were supplied by Whitfield [56]. The experiment includes five periods and 1134 genes. The sub dataset extracted in this part includes five genes and 46 time points, in which the first 40 time points are utilized to search the optimal CVODE models and the rest time points are utilized to test the performance. By applying our method, we have obtained the following CVODE models. Variables $x_1, x_2, \ldots, x_5$ represent five genes, respectively.

$$\frac{dx_1}{dt} = (0.2366 + 4.2126i) \times (x_5 + x_2) - (1.0667 - 0.0343i)x_4^2 - (4.9572 - 2.8702i)$$
$$x_3x_5 + (6.5323 - 0.7549i)x_3 - (4.3955 + 0.5311i) \times (x_1 - x_4)$$

$$\frac{dx_2}{dt} = (1.5138 + 11.009i) \times (x_5 - x_4) - (6.201 + 5.4258i)x_3 + (2.0535 + 0.2076i)$$
$$\times (x_1 + x_3) - (3.7539 - 4.6864i)\cos(x_3)$$

$$\frac{dx_3}{dt} = (2.7072 + 3.5088i)\frac{x_4}{x_1} - (3.3836 + 0.1783\,i) \times (x_4 - x_3) - (0.941 + 0.5496i)$$
$$\times \cos(x_3) + (2.8053 + 3.0456i)x_4$$

$$\frac{dx_4}{dt} = (-0.029 - 2.7968i)\sin(x_5) + (4.5367 + 5.986i)\sin(x_1) - (4.9878 + 0.757i) \times (x_4 - x_3)$$

$$\frac{dx_5}{dt} = (-4.0444 + 0.7446i) \times (x_1 - x_4) - (0.9768 + 2.0349i) \times (x_3 - x_2) + (3.5815$$
$$+ 1.2342\,i)\sin(x_1) + (-1.0076 + 3.716i) \times (x_5 - x_1) + (-1.0549 - 1.8469i) \times 2x_2$$
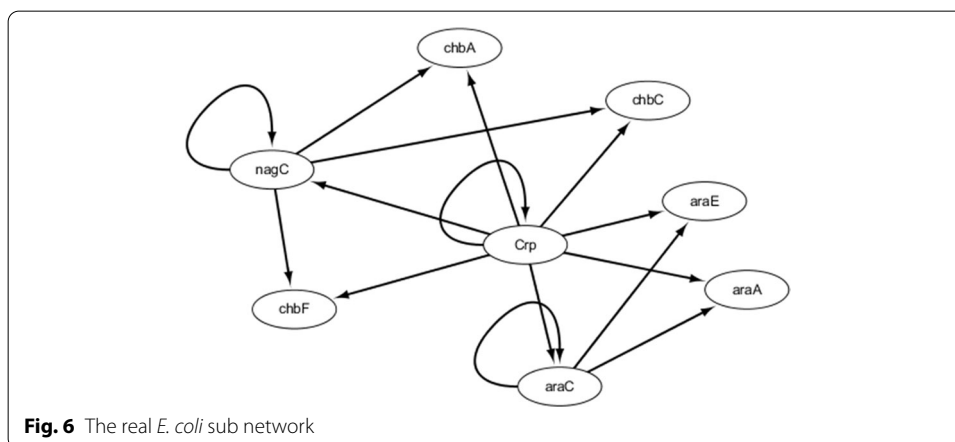$$\tag{3}$$

The prediction errors of ODE and CVODE are depicted in Fig. 5, which reveals that the prediction errors of CVODE are closer to zero than ones of ODE. The predicted RMSE values of ODE and CVODE are listed in Table 3. For the predicted results of five genes, CVODE has smaller prediction RMSE values than ODE, which show that CVODE could more accurately predict gene expression data. On average CVODE could decrease 49.07% RMSE value.
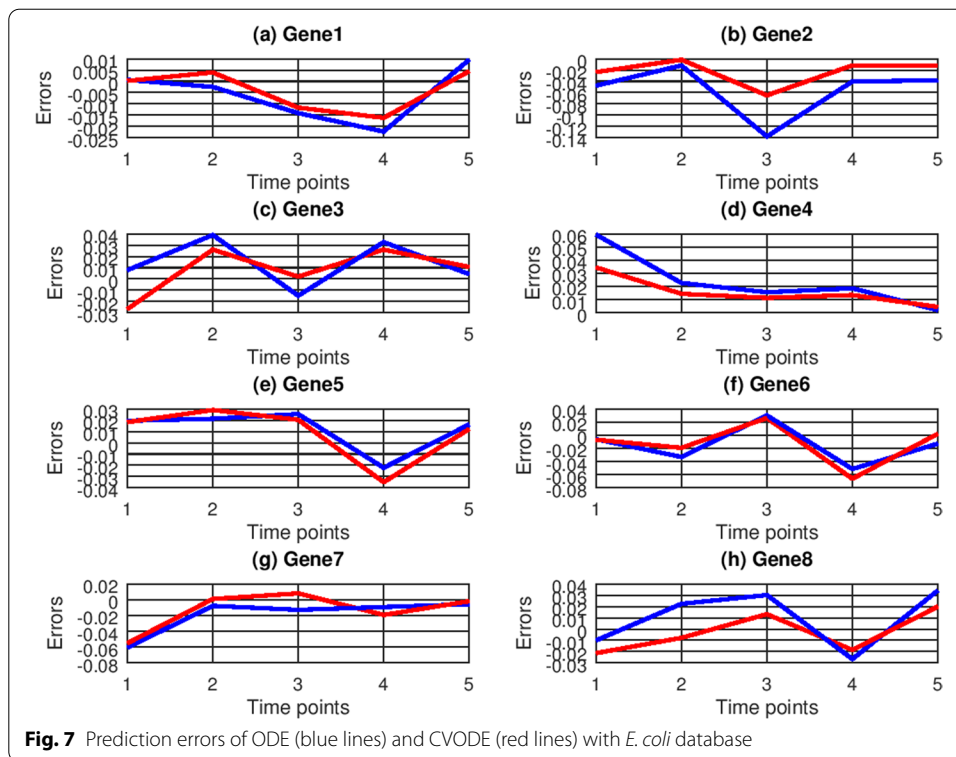
Fig. 5 Prediction errors of ODE (blue lines) and CVODE (red lines) with human cell time-series data
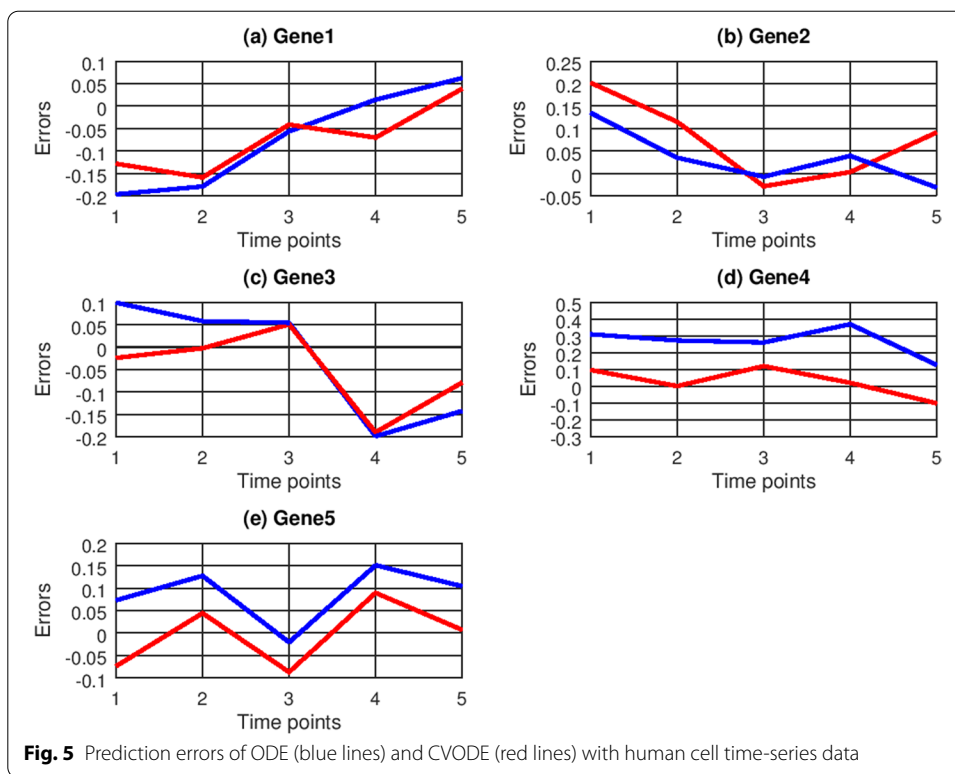
Table 3 Prediction RMSE values of ODE and CVODE for six genes

Genes	ODE	CVODE
Gene 1	0.227851	**0.125126**
Gene 2	0.11358	**0.066308**
Gene 3	0.363917	**0.123348**
Gene 4	0.340177	**0.2797**
Gene 5	0.328782	**0.10551**
Averaged	0.274861	**0.139998**

The 'bold values' represent the better results between two methods

Fig. 6 The real *E. coli* sub network

Fig. 7 Prediction errors of ODE (blue lines) and CVODE (red lines) with *E. coli* database

Table 4 Prediction RMSE values of ODE and CVODE for eight genes

Genes	ODE	CVODE
Crp (Gene 1)	0.013349	**0.010013**
araC (Gene 2)	0.058062	**0.031771**
nagC (Gene 3)	**0.01153**	0.020923
chbC (Gene 4)	0.041527	**0.018565**
araE (Gene 5)	**0.008005**	0.024344
araA (Gene 6)	**0.023222**	0.033139
chbA (Gene 7)	0.028299	**0.015565**
chbF (Gene 8)	0.031919	**0.016927**
Averaged	0.026989	**0.021406**

The 'bold values' represent the better results between two methods

E. coli database

The third real gene expression dataset is extracted from polymicrobial probe database (version 4 build 6), which contains 4297 genes and 907 biology experiments [57]. The used dataset contains the first 35 experiments, in which the first 30 time points are utilized to search the optimal CVODE models and the rest time points are utilized as testing data. Eight genes (**Crp**, **araC**, **nagC**, **chbC**, **araE**, **araA**, **chbA** and **chbF**) are chosen and the real network structure is from RegulonDB [58] containing 15 regulations, which is described in Fig. 6.

Fig. 8 The construction *E. coli* network by ODE (**a**) and CVODE (**b**). The solid lines denote the real regulations, while the dotted lines represent the false relationships
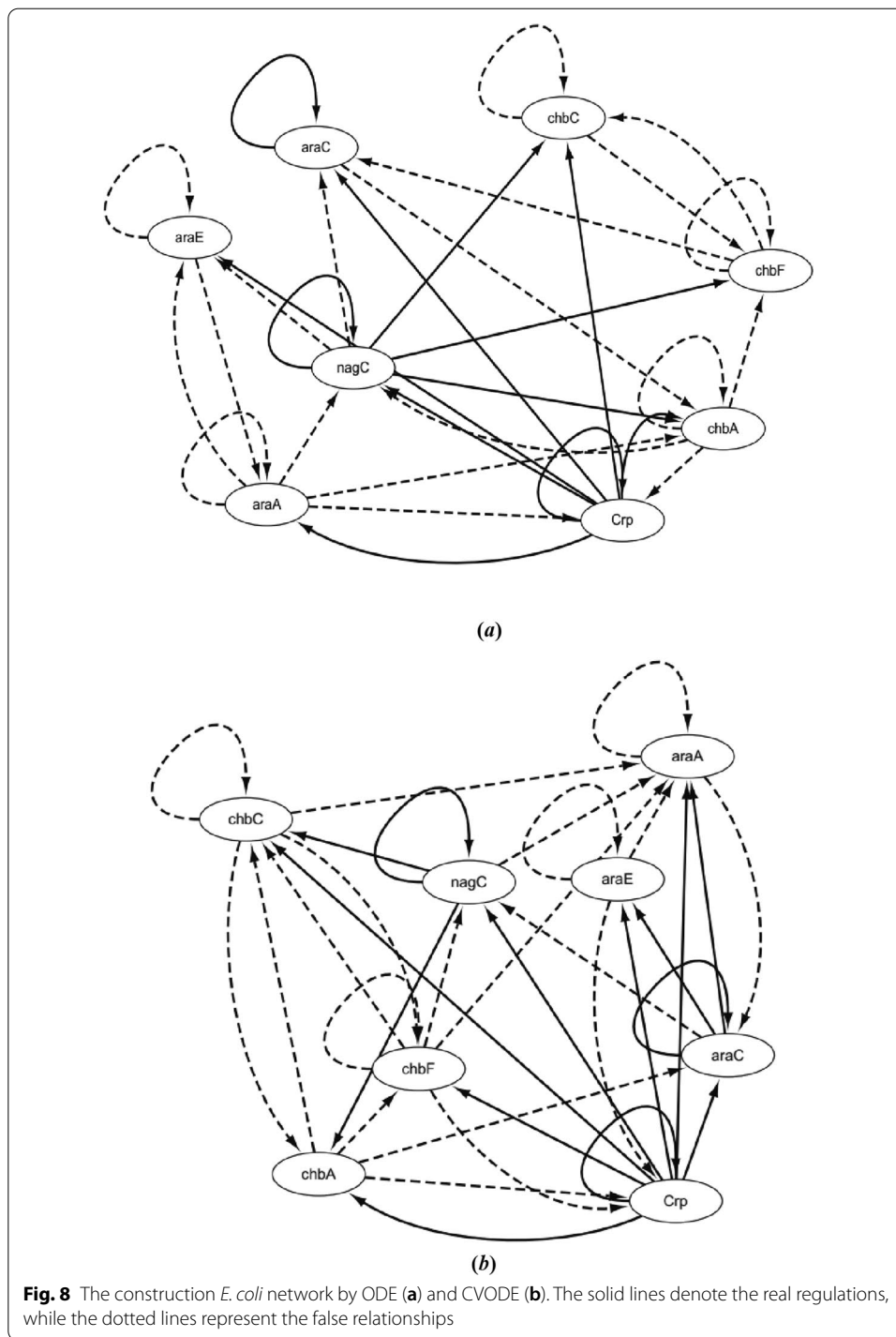
Through our proposed hybrid evolutionary method, we could obtain the CVODE models as Eqs. (4). Variables $x_1, x_2, \ldots, x_8$ represent genes **Crp**, **araC**, **nagC**, **chbC**, **araE**, **araA**, **chbA** and **chbF**, respectively.

Table 5 Performances of ODE and CVODE for constructing *E. coli* network

Methods	S_n	S_p
CVODE	**0.933333**	0.591837
ODE	0.8	**0.612245**

The 'bold values' represent the better results between two methods

$$\frac{dx_1}{dt} = (-4.1015 + 5.371i) \times (x_8 + \frac{x_1}{x_8}) + (2.5219 - 0.3343i)\cos(x_5^2) + (2.0703 + 2.9986i) \times (x_8 + x_7 \times x_8)$$

$$\frac{dx_2}{dt} = (-7.9767 + 9.8189i)x_7^2 + (5.5059 - 5.2522i)x_2x_7 + (3.5167 + 10.3982i) \times (x_2 - x_6)$$
$$- (16.939460 - 13.448812i) \times (x_6 - x_1)$$

$$\frac{dx_3}{dt} = (-0.1369 + 1.0964i)\cos(x_2) + (1.5904 + 0.2761i)x_3^2 + (0.2941 + 1.5517\ i)x_2x_3 +$$
$$- (2.4638 + 1.6326i)\frac{x_1}{x_8} + (0.8683 + 0.6514i)\frac{x_8}{x_2}$$

$$\frac{dx_4}{dt} = (-5.0047 - 0.6521i)\cos(\sin(x_8)) + (0.79 + 0.6943i)\cos(x_3 - x_1) - (3.9688 + 0.779i)$$
$$\times (2x_1 - x_7) + (1.9222 + 0.8736i)\frac{\cos(x_4)}{x_4} + (3.234 - 0.0632i)x_1$$

$$\frac{dx_5}{dt} = (-1.9009 - 0.8821i)\cos^2(x_5) + (0.5571 + 0.9842i)(\frac{x_2}{x_5})^2 + (2.0521 - 0.7683i)x_1$$

$$\frac{dx_6}{dt} = (-0.648 - 0.2905i)x_6 + (-2.2871 + 0.5342i)\frac{x_4}{x_8} - (1.3399 + 4.0571i)\frac{x_6}{x_3}$$
$$- (0.1901 + 2.2337i)x_1^2 + (2.1551 - 0.3921i) \times (x_5 - x_2)$$

$$\frac{dx_7}{dt} = (-3.1357 - 1.2992i)\cos(x_1) + (-1.8814 + 2.4155i)x_3^2 + (0.7295 + 0.7151i)x_4x_1$$
$$+ (-1.4544 - 0.4791i)\sin(x_4)$$

$$\frac{dx_8}{dt} = (-1.7613 - 3.4795i) \times (x_1 - x_6) + (0.1275 + 1.415i)x_1 + (3.3253 - 4.9547i)\sin(x_8)$$
$$+ (-2.8165 + 1.5682i)\frac{x_7}{x_1} + (-0.6489 + 1.2929i)x_6^2$$

$$(4)$$

The predicted errors of ODE and CVODE are depicted in Fig. 7, which reveal that the predicted errors of CVODE are closer to zero than ones of ODE. The predicted RMSE values of ODE and CVODE are listed in Table 4. For Gene1, Gene2, Gene4, Gene7 and Gene8, CVODE has smaller predicted RMSE values than ODE. For Gene3, Gene5 and Gene6, ODE performs better than CVODE. On average CVODE could improve 20.69% prediction accuracy and more accurately forecast gene expression data than ODE.

According to the optimal CVODE models, the inferred *E. coli* sub-network is described in Fig. 8b. ODE models are also utilized to infer *E. coli* and the result is also depicted in Fig. 8a. Compared with real network structure, we can see that CVODE model could identify 14 true-positive regulations and ODE model could identify 12 true-positive relationships. The performances of S_n and S_p of *E. coli* networks inferred by ODE and CVODE are listed in Table 5. From the inferred results, it could be seen that CVODE could identify more accurately gene regulatory network than ODE model.

Conclusions

In this paper, we have presented complex-valued ODE model to identify the regulations among genes for gene regulatory network inference. Grammar-guided genetic programming and complex-valued firefly algorithm are proposed to search the optimal structure and parameters of CVODE. Three real gene expression datasets are utilized and the results reveal that CVODE model could not only improve 20%-50% prediction accuracy of gene expression data, but also identify more true-positive regulatory relationships than ODE.

Our proposed method has some advantages including (1) compared with ODE, CVODE has the complex-valued structures, constants and coefficients, which could improve the modeling ability; (2) GGGP overcomes the shortcomings of GP and CFA has more population diversity and faster convergence than the traditional firefly algorithm, so our proposed hybrid optimization method based on GGGP and CFA could search the optimal model faster; (3) because of the equipment and sample tissues, the expression data may contain noise during the data collection process. Complex-valued methods have higher noise-robust memory mechanisms, so CVODE model has better performance than ODE.

In the further, CVODE will be applied for real large-scale GRN inference. The optimization of complex-valued model needs many computing resources, so the parallel technologies will improve the learning speed of model in the following work.

Methods

Complex-valued ordinary differential equation model

Complex-valued ordinary differential equation (CVODE) model is a variant of ODE model, whose coefficients and functions are complex-valued. Its expression is given as followed [59].

Fig. 9 The derived tree of CVODE model $\sin z + \cos z - z$
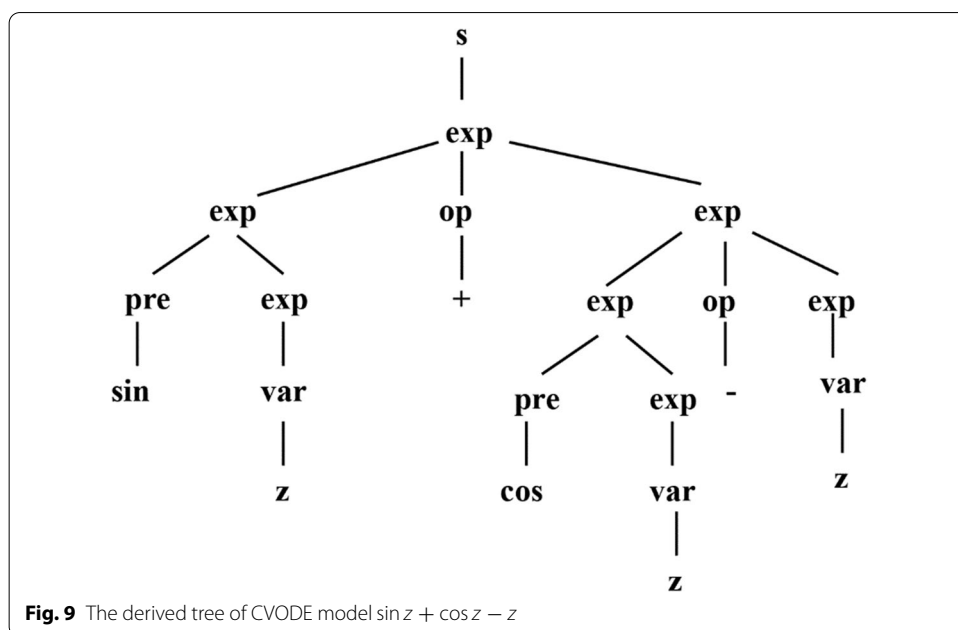
$$\frac{dZ}{dt} = \beta \cdot F(Z, t). \tag{5}$$

where Z is complex-valued independent variable, β is complex-valued coefficient vector, t is the time point and $F(\cdot)$ is complex-valued unknown function.

Structure optimization

Grammar-guided genetic programming (GGGP) is an improved version of genetic programming (GP), which overcomes the shortcomings of GP, such as the generation and preservation of valid programs, small search space, and complex genetic operators [51–60]. GGGP is presented to evolve the structure of CVODE model. In GGGP, context-free grammar (CFG) model is utilized to guide the evolutionary process of GP in order to obtain the optimal solution faster.

CFG model contains a quadruple, which is represented as $G = \{N, T, P, \sum\}$, where N is non-terminal symbol set, T is terminal symbol set, P is production rule set and $\sum$ is beginning symbol set. Four sets satisfy the conditions: $N \cap T = \phi$ and $\sum \in N$. An element in production rule set is represented as $x \rightarrow y$, where $x \in N$, and $y \in N \cup T$. In order to represent an example of CVODE model $\sin z + \cos z - z$, four sets of CFG model are defined in advance. $N = \{s, exp, op, pre, var\}$, $T = \{\sin, \cos, +, -, z\}$, $\sum = \{s\}$, and P is given as

$$
\begin{aligned}
&s \rightarrow \exp \\
&\exp \rightarrow \exp \; op \; \exp \\
&\exp \rightarrow pre \; \exp \\
&\exp \rightarrow var \\
&pre \rightarrow \sin \,|\, \cos \\
&op \rightarrow +\,|- \\
&var \rightarrow z.
\end{aligned}
\tag{6}
$$

The derived tree of CVODE model $\sin z + \cos z - z$ is obtained by generating sentence through context-free grammar, which is depicted in Fig. 9.

In GGGP, the individuals come from the derived trees of the grammar models. As the same as GP, GGGP also has genetic operators to evolve the derived trees, which contain selection, replication, crossover and mutation. Selection and replication mechanisms are consistent with GP. For crossover mechanism, two derived trees are randomly selected, and two sub-trees with the same source characteristics are selected to be crossed, whose root nodes have the same non-terminal symbols. For mutation mechanism, an internal node is selected randomly. If the node is a non-terminal node, the node is retained and the sub-tree with this node as the root node is deleted and replaced by a sub-tree created by syntax rules. If this node is the terminal node, it is replaced by a new terminal node created randomly.

Parameter optimization

Complex-valued firefly algorithm (CFA) is an efficient complex-valued evolutionary algorithm, which is based on the mutual attraction and movement processes of complex-valued fireflies [52]. Compared with many traditional evolutionary algorithms, CFA has some advantages, such as simple design, few parameters, strong robustness, high population diversity and fast convergence.

In CFA, the positions of firefly populations are complex, so the real parts and imaginary parts of the positions of firefly populations need to be optimized in parallel. In CVODE model, complex-valued constants and coefficients need to be optimized. In this paper, CFA is presented to evolve the parameters of CVODE. The detail optimization process of parameters of CVODE with CFA is given in Algorithm 1.

Algorithm 1 Parameter optimization of CVODE with complex-valued firefly algorithm.

Count the number of constants and coefficients as n;

Generate m complex-valued firefly individuals:

$$[X_{1,R} + X_{1,I}\, i, X_{2,R} + X_{2,I}\, i, \ldots, X_{m,R} + X_{m,I}\, i];$$

for $i = 1$; $i \le m$; $i{+}{+}$ **do**

    for $j = 1$; $j \le m$; $j{+}{+}$ **do**

        $r_{ij} \leftarrow$ calculate complex-valued distance with $X_{i,R} + X_{i,I}\, i$ and

        $X_{j,R} + X_{j,I}\, i$;

for $i = 1$; $i \le m$; $i{+}{+}$ **do**

    $f_i \leftarrow$ calculate the fitness value of firefly i;

$t = 1$;

while $t \le T$ **do**

    for $i = 1$; $i \le m$; $i{+}{+}$ **do**

        Select firefly j;

        $\gamma \leftarrow$ coefficient of light absorption;

        $B_{i,R} = B_{i,0} * e^{-\gamma r_{ij,R}}$;   // Real part of the brightness

        $B_{i,I} = B_{i,0} * e^{-\gamma r_{ij,I}}$;   // Imaginary part of the brightness

        $\alpha \leftarrow$ created from the range [0, 1];

        $X_{i,R}(t+1) = X_{i,R}(t) + B_{i,R}(X_{j,R}(t) - X_{i,R}(t)) + \alpha \varepsilon_{i,R}$; // Real part

        of the position

        $X_{i,I}(t+1) = X_{i,I}(t) + B_{i,I}(X_{j,I}(t) - X_{i,I}(t)) + \alpha \varepsilon_{i,I}$; // Imaginary

        part of the position

    $t{+}{+}$;

The flowchart of GRN inference with CVODE

In this paper, CVODE is utilized to infer gene regulatory network, in which GGGP and CFA are proposed to optimize CVODE model. Suppose that gene expression data $[D_1, D_2, \ldots, D_m]$ contains m genes and each gene contains n time points $(D_i = [D_i^1, D_i^2, \ldots, D_i^n])$. The inference flowchart of GRN with m genes by our proposed algorithm is described in Fig. 10 The detailed process is given as follows.

(1) Due to that the input data of CVODE model are complex-valued, gene expression data need to convert into complex data using Algorithm 2 before GRN inference.
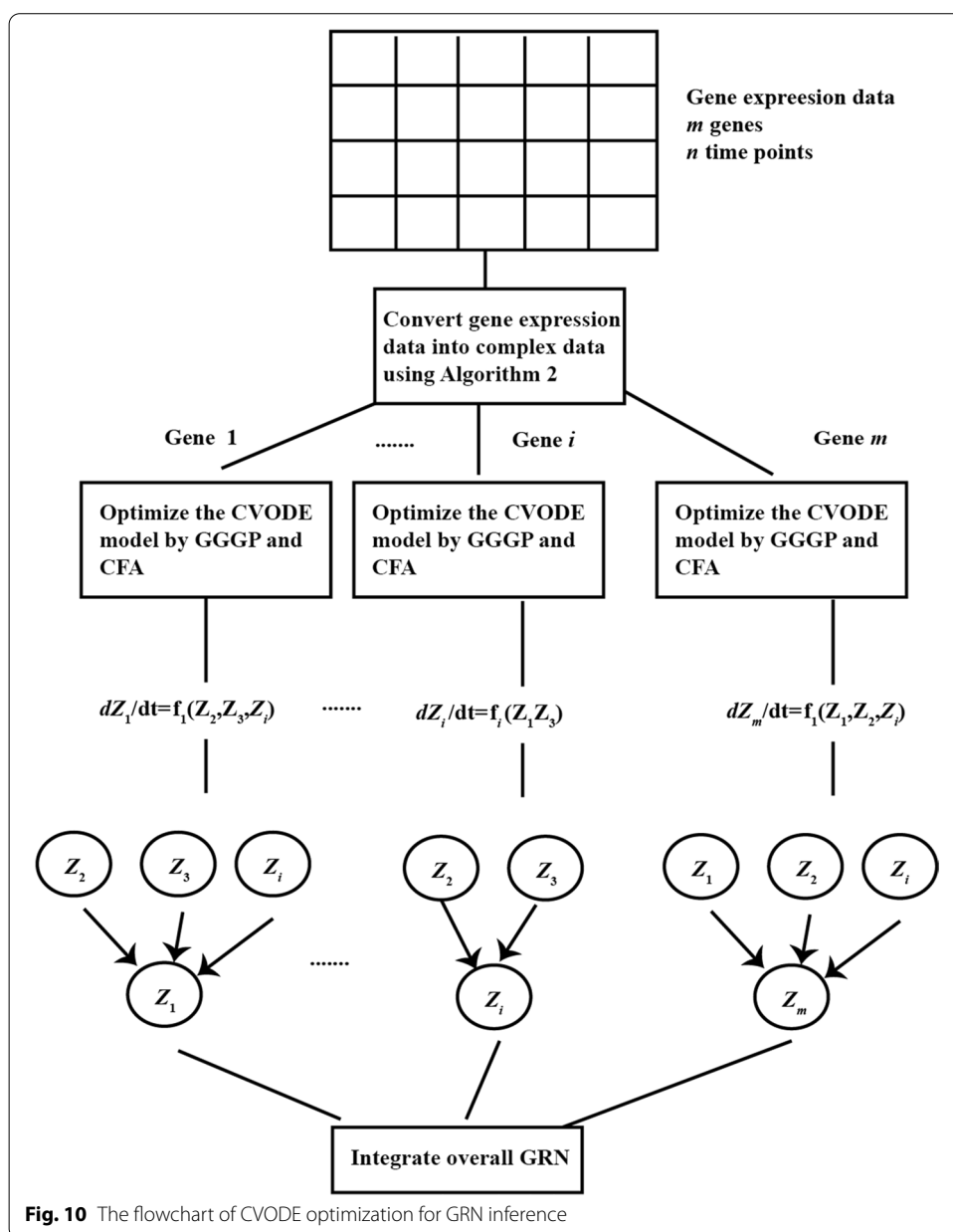
Fig. 10 The flowchart of CVODE optimization for GRN inference

(2) The regulatory relationships of each target gene are inferred by CVODE model independently. For gene *i*, with gene expression data, the process of searching the optimal CVODE model is introduced as followed.

1. Initialize *N* CVODE individuls.
2. Calculate the fitness values of *N* CVODE individuals. The complex outputs need to be converted into real values by Algorithm 3.
3. GGGP is applied to search the optimal structure of CVODE, which is introduced in Sect. 2.2. At some iterations, CFA is utilized to search the optimal complex-valued parameters of CVODE, whose structure is fixed.
4. If the satisfied condition is achieved, the optimal process stops; otherwise, go to step 2).

(3) If the optimal CVODE contains independent variables, the corresponding genes of independent variables could regulate the target gene. For example gene *j* is included in the optimal CVODE $\frac{dZ_i}{dt} = f(Z_j)$), which means that gene *i* is regulated by gene *j*.

(4) $i++$. If the regulatory relationships of all genes have been identified, create overall gene regulatory network; otherwise go to step (2).

Algorithm 2 Pseudo code of real number converted into complex number.

Input: Real-valued input vector $[g_1, g_2, \ldots, g_m]$ (*m* is the number of sample

  points)

Output: Complex-valued output vector $[z_1, z_2, \ldots, z_m]$

Count the maximum and minimum values of input vector (g_{max} and g_{min});

for $k = 1; \ k \leq m \ ; k++$ **do**

$$\varphi_k = \frac{g_k - g_{min}}{g_{max} - g_{min}}(2\pi - \delta); \ // \ \delta \ \text{is a shift angle}$$

$$z_k = e^{i\varphi_k}$$

end

Algorithm 3 Pseudo code of complex number converted into real number.

Input: Complex-valued input vector $[z_1, z_2, \ldots, z_m]$, maximum and minimum

values of input data g_{max} and g_{min}

Output: Real-valued output data $[y_1, y_2, \ldots, y_m]$

for $k = 1;\ \ k \leq m\ ; k{+}{+}$ **do**

$\quad \arg z_k = \varphi_k;$

$$y_i = \frac{\varphi_k (g_{max} - g_{min})}{2\pi - \delta_k} + g_{min};$$

end

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Author details
[1]School of Information Science and Engineering, Zaozhuang University, Zaozhuang 277160, China. [2]School of Information and Electrical Engineering, Xuzhou University of Technology, Xuzhou 221018, China. [3]School of Information Science and Engineering, University of Jinan, Jinan 250022, China.

References
1. Marcel S, Ramon MMJ, Panagiota M, Wittbrodt B, Wittbrodt J. A global survey identifies novel upstream components of the Ath5 neurogenic network. Genome Biol. 2009;10(9):R92.
2. Israel JW, Martik ML, Byrne M, Raff EC, Raff RA, McClay DR, Wray GA. Comparative developmental transcriptomics reveals rewiring of a highly conserved gene regulatory network during a major life history switch in the sea urchin genus heliocidaris. PLoS Biol. 2016;14(3):e1002391.
3. Taou NS, Corne DW, Lones MA. Investigating the use of boolean networks for the control of gene regulatory networks. J Comput Sci. 2018;26:147–56.
4. Jin ZC, Cheng WU, Gao QB, Jiang Y. Gene regulatory network models based on time series gene expression data: recent progress. Acad J Second Mil Univ. 2008;28(9):1106–9.
5. Wu H, Lu T, Xue H, Liang H. Sparse additive ordinary differential equations for dynamic gene regulatory network modeling. J Am Stat Assoc. 2014;109(506):700–16.
6. Ortizgutiérrez E, Garcíacruz K, Azpeitia E, Castillo A, de la Paz Sánchez M, Álvarez-Buylla ER. A dynamic gene regulatory network model that recovers the cyclic behavior of arabidopsis thaliana cell cycle. PLoS Comput Biol. 2015;11(9):e1004486.
7. Zheng CH, Huang DS, Zhang L, Kong XZ. Tumor clustering using nonnegative matrix factorization with gene selection. IEEE Trans Inf Technol Biomed. 2009;13(4):599–607.
8. Modrák M, Vohradský J. Genexpi: a toolset for identifying regulons and validating gene regulatory networks using time-course expression data. BMC Bioinform. 2018;19(1):137.
9. Thomas T. Approximate inference of gene regulatory network models from RNA-Seq time series data. BMC Bioinform. 2018;19:127.
10. Thomas R, Mehrotra S, Papoutsakis ET, Hatzimanikatis V. A model-based optimization framework for the inference on gene regulatory networks from DNA array data. Bioinformatics. 2004;20(17):3221–35.
11. Huang DS, Yu HJ. Normalized feature vectors: a novel alignment-free sequence comparison method based on the numbers of adjacent amino acids. IEEE/ACM Trans Comput Biol Bioinform. 2013;10(2):457–67.
12. Zheng CH, Zhang L, Ng VT, Shiu SC, Huang DS. Molecular pattern discovery based on penalized matrix decomposition. IEEE/ACM Trans Comput Biol Bioinform. 2011;8(6):1592–603.
13. Huang DS, Zheng CH. Independent component analysis based penalized discriminant method for tumor classification using gene expression data. Bioinformatics. 2006;22(15):1855–62.
14. Zhu L, You ZH, Huang DS, Wang B. t-LSE: a novel robust geometric approach for modeling protein-protein interaction networks. PLoS ONE. 2013;8(4):e58368.
15. Huang DS, Zhang L, Han K, Deng S, Yang K, Zhang H. Prediction of protein-protein interactions based on protein-protein correlation using least squares regression. Curr Protein Pept Sci. 2014;15(6):553–60.
16. Bao WZ, Huang ZH, Yuan CA, Huang DS. Pupylation sites prediction with ensemble classification model. Int J Data Mining and Bioinformatics. 2017;18(2):91–104.
17. Bao W, Wang D, Chen Y. Classification of protein structure classes on flexible neutral tree. IEEE/ACM Trans Comput Biol Bioinform. 2017;14(5):1122–33.
18. Pietro Z, Sandro M, Michele C. TimeDelay-ARACNE: Reverse engineering of gene networks from time-course data by an information theoretic approach. BMC Bioinform. 2010;11:154.
19. Liu Z, He Q. A novel boolean network for analyzing the p53 gene regulatory network. Curr Bioinform. 2016;11(1):13–21.
20. Njah H, Jamoussi S. Weighted ensemble learning of Bayesian network for gene regulatory networks. Neurocomputing. 2015;150:404–16.
21. Xing L, Guo M, Liu X, Wang C, Zhang L. Gene regulatory networks reconstruction using the flooding-pruning hill-climbing algorithm. Genes. 2018;9(7):342.
22. Tao L, Min W. Investigate data dependency for dynamic gene regulatory network identification through high-dimensional differential equation approach. Commun Stat Simul Comput. 2014;45(7):2377–91.
23. Huang DS. Systematic theory of neural networks for pattern recognition. Beijing: Publishing House of Electronic Industry of China; 1996.
24. Huang DS. Radial basis probabilistic neural networks: Model and application. Int J Pattern Recognit Artif Intell. 1999;13(7):1083–101.
25. Yang B, Chen Y, Jiang M. Reverse engineering of gene regulatory networks using flexible neural tree models. Neurocomputing. 2013;99(1):458–66.
26. Hegland M, Burden C, Santoso L, MacNamara S, Booth H. A solver for the stochastic master equation applied to gene regulatory networks. J Comput Appl Math. 2007;205(2):708–24.
27. Munsky B, Khammash M. The finite state projection approach for the analysis of stochastic noise in gene networks. IEEE Trans Autom Control. 2008;53(Special Issue):201–14.
28. Ding H, Luo LF. Kinetic model of the lysogeny/lysis switch of phage λ. Chin Phys Lett. 2009;26(9):098701.
29. Ji ZW, Su J, Wu D, Peng HM, Zhao WL, Zhao BN, Zhou XB. Predicting the impact of combined therapies on myeloma cell growth using a hybrid multi-scale agent-based model. Oncotarget. 2017;8(5):7647–65.

30. Ding H, Luo LF, Lin H. Entropy production rate changes in lysogeny/lysis switch regulation of bacteriophage lambda. Commun Theor Phys. 2011;55(2):371–5.
31. Ji ZW, Zhao WL, Lin HK, Zhou XB. Systematically understanding the immunity leading to CRPC progression. PLoS Comput Biol. 2019;15(9):e1007344.
32. Feng W, Ding H, Lin H, Luo LF. The lysogeny/lysis switch and entropies of stationary states in λ phage. Acta Phys Sin. 2012;61(16):168701.
33. Ji ZW, Yan K, Li WY, Hu HG, Zhu XL. Mathematical and computational modeling in complex biological systems. Biomed Res Int. 2017;2017:5958321.
34. Hohm T, Zitzler E. Multicellular pattern formation: parameter estimation for ordinary differential equation-based gene regulatory network models. IEEE Eng Med Biol Mag Q Mag Eng Med Biol Soc. 2009;28(4):52.
35. Tian T, Burrage K, Burrage PM, Carletti M. Stochastic delay differential equations for genetic regulatory networks. J Comput Appl Math. 2007;205(2):696–707.
36. Gebert J, Radde N, Weber GW. Modeling gene regulatory networks with piecewise linear differential equations. Eur J Oper Res. 2007;181(3):1148–65.
37. Jong HD, Page M. Search for steady states of piecewise-linear differential equation models of genetic regulatory networks. IEEE/ACM Trans Comput Biol Bioinform. 2008;5(2):208–22.
38. Zhang Q, Yu Y, Zhang J, Liang H. Using single-index ODEs to study dynamic gene regulatory network. PLoS ONE. 2018;13(2):e0192833.
39. Matsumoto H, Kiryu H, Furusawa C, Ko MSH, Ko SBH, Gouda N, Hayashi T, Nikaido I. SCODE: an efficient regulatory network inference algorithm from single-cell RNA-Seq during differentiation. Bioinformatics. 2017;33(15):2314–21.
40. Lee WP, Hsiao YT. An adaptive GA-PSO approach with gene clustering to infer s-system models of gene regulatory networks. Comput J. 2011;54(9):1449–64.
41. Chowdhury AR, Chetty M, Evans R. Stochastic S-system modeling of gene regulatory network. Cogn Neurodyn. 2015;9(5):535–47.
42. Chowdhury AR, Chetty M. Network decomposition based large-scale reverse engineering of gene regulatory network. Neurocomputing. 2015;160(21):213–27.
43. Nitta T. On the critical points of the complex-valued neural network. In: 9th international conference on neural information processing. 2002; p. 1099–1103.
44. Hirose A, Yoshida S. Generalization characteristics of complex-valued feedforward neural networks in relation to signal coherence. IEEE Trans Neural Netw Learn Syst. 2012;23(4):541–51.
45. Wisdom S, Powers T, Hershey JR, Roux JL, Atlas L. Full-capacity unitary recurrent neural networks. Adv Neural Inf Process Syst. 2016;29:4880–8.
46. You C, Hong D. Nonlinear blind equalization schemes using complex-valued multilayer feedforward neural networks. IEEE Trans Neural Netw. 1998;9(6):1442–55.
47. Deng JP, Sundararajan N, Saratchandran P. Communication channel equalization using complex-valued minimal radial basis function neural networks. IEEE Trans Neural Netw. 2002;13(3):687–96.
48. Hu J, Wang J. Global stability of complex-valued recurrent neural networks with time-delays. IEEE Trans Neural Netw Learn Syst. 2012;23(6):853–65.
49. Trabelsi C, Bilaniuk O, Zhang Y, Serdyuk D, Subramanian S. Deep complex networks. In: ICLR. 2018; p. 1–19.
50. Yang B, Chen YH. A new complex-valued polynomial model. Neural Process Lett. 2018;50:1–18.
51. Wu P, Chen Y. Grammar guided genetic programming for flexible neural trees optimization. Lect Notes Comput Sci. 2007;4426:964–71.
52. Song C. A complex-valued firefly algorithm. Intell Comput Theor Appl. 2019;11644:700–7.
53. Yang XS, Hosseini SSS, Gandomi AH. Firefly Algorithm for solving non-convex economic dispatch problems with valve loading effect. Appl Soft Comput. 2012;12(3):1180–6.
54. Raza K, Alam M. Recurrent neural network based hybrid model for reconstructing gene regulatory network. Comput Biol Chem. 2016;64:322–34.
55. Ronen M, Rosenberg R, Shraiman BI, Alon U. Assigning numbers to the arrows: parameterizing a gene regulation network by using accurate expression kinetics. Proc Natl Acad Sci. 2002;99(16):10555–60.
56. Whitfield ML, Sherlock G, Saldanha AJ, Murray JI, Ball CA, Alexander KE, Matese JC, Perou CM, Hurt MM, Brown PO, Botstein D. Identification of genes periodically expressed in the human cell cycle and their expression in tumors. Mol Biol Cell. 2002;13:1977–2000.
57. Faith JJ, Driscoll ME, Fusaro VA, Cosgrove EJ, Hayete B, Juhn FS, Schneider SJ, Gardner TS. Many Microbe Microarrays Database: uniformly normalized Affymetrix compendia with structured experimental metadata. Nucleic Acids Res. 2008;36:D866–70.
58. Alberto S, Heladia S, Socorro G, et al. RegulonDB v 10.5: tackling challenges to unify classic and high throughput knowledge of gene regulation in *E. coli* K-12. Nucleic Acids Res. 2019;47(1):212–20.
59. Yang B, Bao WZ. Complex-valued ordinary differential equation modeling for time series identification. IEEE Access. 2019;7:41033–42.
60. Luna JM, Romero JR, Ventura S. Design and behavior study of a grammar-guided genetic programming algorithm for mining association rules. Knowl Inf Syst. 2012;32(1):53–76.

Publisher's Note