**Proceedings of the 2007 IEEE Symposium on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB 2007)**

# Inference of Gene Regulatory Networks using S-System: A Unified Approach

Haixin Wang and Lijun Qian
Department of Electrical Engineering
Prairie View A&M University
Prairie View, Texas 77446
Email: HWang, LiQian@pvamu.edu

Edward Dougherty
Computational Biology Division
Translational Genomics Research Institute (TGen)
Phoenix, AZ 85004
Email: edward@ee.tamu.edu

***Abstract*— In this paper, a unified approach to infer gene regulatory networks using the S-system model is proposed. In order to discover the structure of *large-scale* gene regulatory networks, a simplified S-system model is proposed that enables fast parameter estimation to determine the major gene interactions. If a detailed S-system model is desirable for a subset of genes, a two-step method is proposed where the range of the parameters will be determined first using Genetic Programming and Recursive Least Square estimation. Then the exact values of the parameters will be calculated using a multi-dimensional optimization algorithm. Both downhill simplex algorithm and modified Powell algorithm are tested for multi-dimensional optimization. Simulation results using both synthetic data and real microarray measurements demonstrate the effectiveness of the proposed methods.**

## I. Introduction

The advances of DNA microarray technologies and gene chips have allowed biologists to analyze the genetic behaviors among different genes. After image processing of the DNA microarray photos, it is possible to discover gene regulatory networks (GRNs) which are complex and nonlinear in nature. Specifically, the increasing existence of microarray time-series data makes possible the characterization of dynamic nonlinear regulatory interactions among genes.

Because GRN models are difficult to deduce solely by means of experimental techniques, computational and mathematical methods are indispensable. Biochemical systems such as GRNs are commonly modeled by systems of ordinary differential equations (ODEs). Much research has been done on GRN modeling by *linear* differential equations using time-series data. However, nonlinear differential equation models, such as an S-system [1], can model much more complicated GRN behavior [2]. In general, modeling GRNs may be considered as a nonlinear identification problem. Assume that there are $N$ genes of interest, define $x_i$ as the state (such as the gene expression level) of the $i^{th}$ gene, then the dynamics/interactions of the GRN may be modeled as

$$\frac{dx_i}{dt} = f_i(x_1, x_2, \cdots, x_N) \tag{1}$$

where the nonlinear functions $f_i$ need to be determined from time-series microarray measurements.

Inference of GRNs using S-system model from time-series microarray measurement data has attracted a lot of attentions recently. The S-system model is given by:

$$\frac{dx_i}{dt} = \alpha_i \prod_{j=1}^{N} x_j^{g_{i,j}} - \beta_i \prod_{j=1}^{N} x_j^{h_{i,j}},\ (i = 1, ..., N) \tag{2}$$

where $x_i$ is the state variable. $\alpha_i$ and $\beta_i$ are the positive rate constants. $g_{i,j}$ and $h_{i,j}$ are the exponential parameters called kinetic orders. If $g_{i,j} > 0$, gene $j$ will induce the expression of gene $i$. On the contrary, gene $j$ will inhibit the expression of gene $i$ if $g_{i,j} < 0$. $h_{i,j}$ will have the opposite effects on controlling gene expressions compared to $g_{i,j}$. S-system is a quantitative model which is characterized by power-law functions. It has the rich structure capability of capturing various dynamics in many biochemical systems [3]. In addition, the S-system model has been proven to be successful in modeling GRNs [4], [5], [6], [7], [8], [9]. Hence, the S-system model is adopted for modeling GRNs in this paper.

In order to solve the nonlinear parameter estimation problem, or equivalently the nonlinear optimization problem, evolutionary algorithms are applied by many studies. In [7], genetic algorithm and a crossover method called Simplex Crossover (SPX) are used to solve the optimization problem. In addition, a gradual optimization strategy is applied to increase the number of predictable parameters. The authors successfully inferred the dynamics of a small genetic network constructed with 60 parameters for 5 network variables and feedback loops. A Memetic Algorithm (MA) is applied in [4] to enhance the optimization process. It is shown that MA performs much better than the standard evolution strategies. Other improvements over standard evolutionary algorithms include the differential evolution algorithm employed by [5] and the cooperative coevolutionary algorithm proposed in [6].

The identification of the S-system requires the estimation of $2N(N+1)$ parameters simultaneously. Although the proposed schemes in the literature have successfully inferred small scale GRNs, they can not be directly applied to the inference of large scale GRNs because of their high computational complexity. In this paper, a simplified S-system model that captures the essential gene interactions is proposed. Although the proposed simplified S-system model is still nonlinear, the corresponding parameter estimation problem becomes *linear*. Hence, parameter estimation algorithms with very low computational complexity, such as the Recursive Least Square (RLS) [18] algorithm, may be applied to infer the simplified S-system

model.

After parsing the entire GRN using the proposed simplified S-system model, a detailed S-system model may be obtained for a subset of genes that are of special interests. We propose a two-step method to infer the detailed S-system model. Firstly, a range search procedure using genetic programming [13] and RLS estimation is applied to determine the range of the parameters. We would like to point out that the range of the parameters are assumed to be known in most previous works [5], [6], [7], [8], which is not realistic. Then a multi-dimensional optimization algorithm is needed to further pin-point the values of the parameters. In this paper, both downhill simplex algorithm [11] and modified Powell algorithm [10] are tested for multi-dimensional optimization. A decomposition procedure that allows to investigate the genes one-at-a-time is applied to S-system model in [8]. In this study, we also employ the same decomposition procedure to reduce the dimensions of the optimization problem.

The remainder of the paper is organized as follows: The proposed simplified S-system model and the corresponding identification scheme are illustrated in Section II. Section III presents a two-step method to provide the parameter estimation for an exact S-system model and the simulation results are given in Section IV. Section V contains some concluding remarks.

## II. Simplified S-System

The task of identifying GRNs may be considered as an optimization problem. The goal is to minimize the identification error and keep the model as simple as possible, which may be achieved by minimizing the following fitness function

$$\text{fitness} = \sum_{\mathrm{i=1}}^{\mathrm{N}}[\sum_{\mathrm{k=1}}^{\mathrm{M}}(\mathrm{x_i(k)} - \mathrm{x_i^{tar}(k)})^2 + \mathrm{C_i}] \qquad (3)$$

where $M$ is the number of data points, $x_i^{tar}$ is the target time series and $x_i$ is the obtained time series given by the obtained S-system model. $C_i$ is a penalty term that may be set to be proportional to the complexity of the model. For instance, $C_i$ may be chosen as $C_i = w\sum_j[|g_{i,j}| + |h_{i,j}|]$, where $w$ is a design parameter.

In this section, a simplified S-System model will be derived from the standard S-system model. Note that equation (2) may be re-written as

$$\frac{dx_i}{dt} = \alpha_i \prod_{j=1}^{N} x_j^{g_{i,j}} * (1 - \frac{\beta_i \prod_{j=1}^{N} x_j^{h_{i,j}}}{\alpha_i \prod_{j=1}^{N} x_j^{g_{i,j}}}) \qquad (4)$$

In a S-system , all $\alpha_i$ and $\beta_i$ and the state variables are always positive. Hence, a logarithm function can be performed on both sides of the equation (4):

$$\log(\frac{dx_i}{dt}) = log(\alpha_i) + \sum_{j=1}^{j=N} g_{i,j} * log(x_j) + log(1 - \bar{x}_i) \quad (5)$$

where

$$\bar{x}_i = \frac{\beta_i \prod_{j=1}^{N} x_j^{h_{i,j}}}{\alpha_i \prod_{j=1}^{N} x_j^{g_{i,j}}} \qquad (6)$$

The different values of $\bar{x}_i$ may lead to different solutions as explained in the follows.

1) When $\bar{x}_i \ll 1$
   From the Taylor-series expansion of the logarithm function $log(1-x) = -x - \frac{x^2}{2} + ...$, we can simplify equation (4) to the following approximation
   $$\log(\frac{dx_i}{dt}) \approx log(\alpha_i) + \sum_{j=1}^{j=N} g_{i,j} * log(x_j) \qquad (7)$$
2) When $\bar{x}_i \gg 1$
   Using the same Taylor-series expansion we can get the following approximation
   $$\log(\frac{-dx_i}{dt}) \approx log(\beta_i) + \sum_{j=1}^{j=N} h_{i,j} * log(x_j) \qquad (8)$$
3) When $\bar{x}_i \approx 1$
   In this case, $\frac{dx_i}{dt} \approx 0$. The gene $i$ stays at a steady state and its steady state value may be deduced directly from the measurements of the experiment.

The biological explanation of the assumptions made above is that during protein synthesis and gene expression, the active and thus interesting genes are either activated or inhibited. Therefore, the main activities of the interesting genes can be covered by the proposed simplified S-system model. Although the simplified S-system model may not be exact, the major activities of the genes can be modeled with reasonable accuracy. If indeed an exact S-system model of certain subset of genes is needed, a two-step method may be used as described in detail in the next section of the paper.

Note that although the proposed simplified S-system model is still nonlinear, the corresponding parameter estimation problem becomes *linear*. Hence, parameter estimation algorithms with very low computational complexity, such as the recursive least square (RLS) algorithm, may be applied to infer the simplified S-system model. The procedures are outlined as follows. Here a simple example is given to illustrate the proposed procedures. Suppose the original S-system model is known

| |
|---|
| Step 1, Initialization: Preprocess the raw time-series data $x_i$ and make sure there is no negative value. |
| Step 2, Apply RLS to the system, and solve the parameters : $g_{i,j}$ and $h_{i,j}$. |
| step 3, Apply RLS to the solved system ( not the raw data) to estimate the parameters $\alpha_i$ and $\beta_i$. |

TABLE I
PARAMETER ESTIMATIONS USING RLS FOR SIMPLIFIED S-SYSTEM MODEL.

$$\begin{aligned} \dot{x_1} &= x_1^{1.5}x_2^{1.2} - x_2^{0.2} \\ \dot{x_2} &= x_1^{0.1} - x_1 x_2^{0.5} \end{aligned}$$

and the raw time-series data are generated with initial conditions $x_1 = 1;\ x_2 = 0.1$. Before we apply the suggested

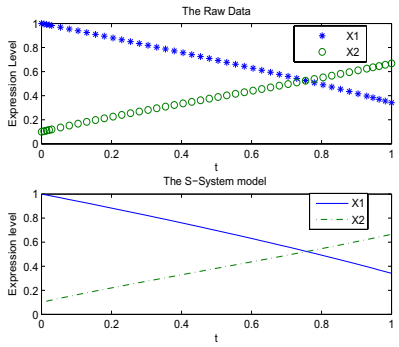


Fig. 1. Trajectories of the original S-system and the obtained simplified S-system.

method, we calculate $\bar{x}$ to verify the conditions of the simplified S-system model are met. In fact, $\bar{x}_1$ is around $(0.1, 0.2)$ and $\bar{x}_2$ is around $(2.27, 3.16)$.

The obtained simplified S-system model is given by

$$\begin{aligned} \dot{x_1} &= -0.6035 x_1^{-0.274} x_2^{0.0284} \\ \dot{x_2} &= 0.347 x_1^{0.45} x_2^{-0.282} \end{aligned}$$

The corresponding trajectories of the original S-system and the obtained simplified S-system are shown in Fig. 1. It is observed that the trajectories from the two models are almost identical. In addition, it is clear that gene 2 inhibits gene 1, as expected.

## III. Inference of the Exact S-System: A Two-Step Method

In this section, a two-step method is proposed to infer the exact S-system model for a relatively small group of genes that may be of special interests.

### A. Optimization Range Search

In most of the previous works [5], [6], [7], [8], the range of the parameters are assumed to be known a priori and are set manually. However, the range of the parameters are usually not known in realistic environment. Hence, the first step in inference of the exact S-system model would be to determine the range of the parameters.

Genetic programming (GP) [13] and RLS estimation algorithm are applied to solve the optimization range. The general parameters in GP are defined as follows.

- $\phi$ : the initial population randomly generated by computer program
- $f$: the fitness function for each individual
- $\gamma$: the fitness threshold to terminate the loops
- $\delta$: the size of $\phi$
- $\mu$: the crossover factor of $\phi$ in each generation
- $\nu$: the mutation rate
- $\lambda$: alternate termination threshold

The GP algorithm can be described by a function $GP(f, \gamma, \delta, \mu, \nu, \lambda)$. The goal of the whole process is to find the generation that minimize the fitness function and get the minimize value of $GP(f, \gamma, \delta, \mu, \nu, \lambda)$.

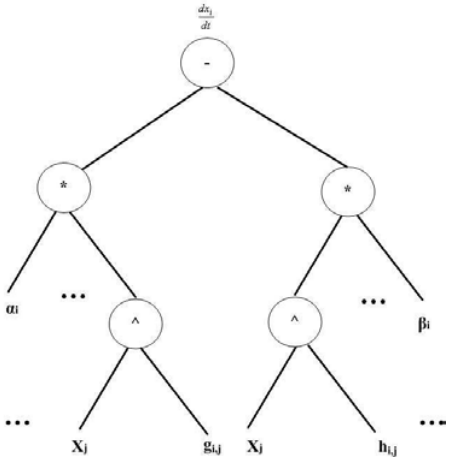


Fig. 2. The data structure of the S-System model

*1) Data Structure:* In order to identify a S-System, the right-hand side of the differential equations of each individual can be described by the following data structure shown in Fig. 2.

*2) Fitness Function Definition:* The fitness function of each individual is defined as follows.

$$\text{fitness}(\mathrm{P_j}) = \sum_{\mathrm{i}=1}^{\mathrm{T}-1} [(\mathrm{x_j^{'}}(\mathrm{t_0} + \mathrm{i}\Delta\mathrm{t}) - (\alpha_j \prod_{k=1}^{N} x_j(t_0 + i\Delta t)^{g_{j,k}} - \beta_j \prod_{k=1}^{N} x_j(t_0 + i\Delta t)^{h_{j,k}}))^2 + C_j] \quad (9)$$

where

- $P_j = (\alpha_j, g_{j,1}, ..., g_{j,N}, \beta_j, h_{j,1}, ..., h_{j,N})$
- $t_0$ : the starting time
- $\Delta t$ : the step size
- $T$: the number of the data points
- $N$: the number of the genes
- $i$: the current gene
- $x_j^{'}(t_0 + i\Delta t)$ : the given data series

In this fitness function, all parameters in equation (4) are trained using all the time-series data. The individual with the least value of the fitness function is selected as the best individual to fit for the given data.

*3) The framework:* In this study, GP and RLS are used to search for the optimization range. GP has the ability to get the global optimization range and RLS will make sure the search converges locally. Fig. 3 shows the general process for global range search.

### B. Exact Parameters Calculation

After the range of the parameters are determined, the exact values of the parameters need to be calculated using a multi-

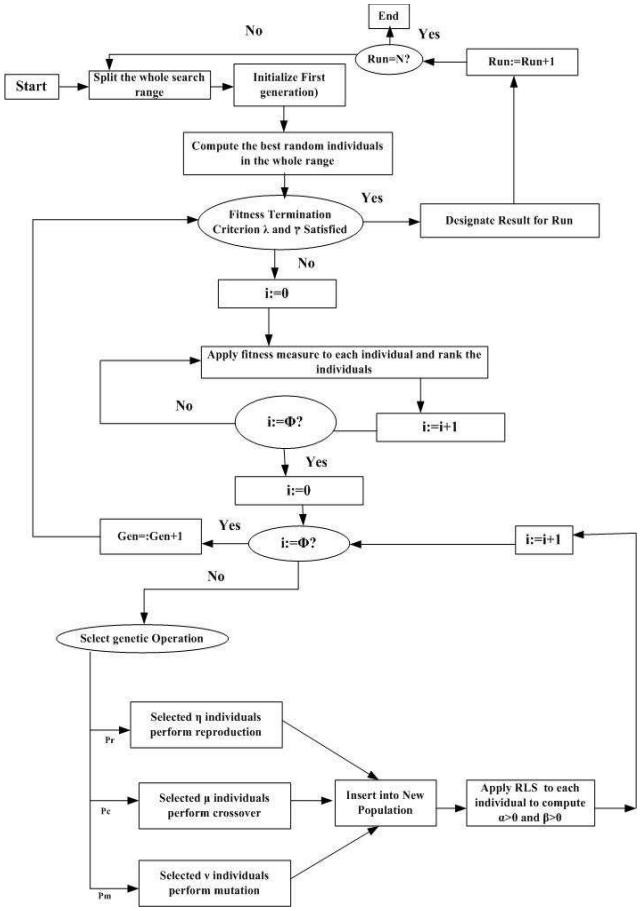


Fig. 3. The global optimization range search by genetic programming and RLS estimation

dimensional optimization algorithm. Both downhill simplex algorithm [11] and modified Powell algorithm [10] are tested for multi-dimensional optimization.

*1) Downhill Simplex Method:* Downhill simplex method is a direct calculation method based on heuristic ideas. The multidimensional downhill methods try to find the minimum of a function of more than one variable, which is not analogous to the one-dimensional problem. It is easy to implement and it does not need the derivatives which are not easily written as analytic expressions or solved in simple terms. The disadvantages of this method are that it requires large number of iterations and it may not converge to the global minimum. Powerful computers and some optimal programming may be used to solve the first problem. And GA is applied to avoid the local minimum.

Initialization of the $N+1$ starting points is very important because the downhill simplex process may fail if they are not selected appropriately. In our case, the optimization range search during the previous step already guarantees that the initial starting points of the downhill simplex are near the global minimum. And applying GA can further guarantee that the local minimum will be avoided. Therefore, the initial points for the downhill simplex process can be any value in the range that we obtained in the optimization range search.

A simplex is a geometric figure of $n$ dimensions, with $n+1$ vertices, interconnecting lines and polygonal faces. The key equation in downhill simplex method is the following equation.

$$P_i = P_0 + \lambda_i e_i,\ i = 1, ..., n \tag{10}$$

where $P_0$ is the initial guess. $e_i$ is the unit vector. $\lambda_i$ is the characteristic length scale which can be a constant or a variable. The downhill simplex method takes a series of steps to find the minimum of the function. It involves moving the vertex of the simplex where the function evaluation is the largest through the opposite face of the simplex to a lower point. This process is called reflection. Amoeba search [12] is also used here to search the valley of the simplex. The entire process is listed in the following pseudo-code.

| |
|---|
| Produce the initial points $P_0$ |
| COMPUTE the fitness for $x_i$ |
| WHILE $\lambda$ and $\nu$ are not satisfied |
| FOR i=1 to number of the simplex points |
| DO: |
| Compute $P_i : P_i = P_0 + \lambda_i e_i, i = 1, ..., n$ |
| Determine the highest ($x_h$), next-highest($x_{nl}$) and |
| lowest points ($x_l$) by f |
| END |
| Compute the range $x_h - x_l$ |
| IF Range $<$ Tolerance $\varepsilon$ |
| RETURN $x_l$ in simplex |
| END |
| Extrapolate $x_h$ in the simplex through the opposite face |
| IF Reflected point $<$ Current $x_l$ |
| Try an extrapolation by a factor of 2 |
| ELSE IF Reflected Point $>$ Current $x_l$ |
| Do a one-dimensional contraction from $x_h$ |
| IF Contracted Point $> x_h$ |
| contract around the $x_l$ |
| END IF |
| END IF |
| Apply reproduction operation |
| Apply crossover operation |
| Apply mutation operation |
| Keep the best individual |
| APPLY RLS to compute coefficients $\alpha_i$ and $\beta_i$ |
| END WHILE |

TABLE II
PSEUDO-CODE OF THE DOWNHILL SIMPLEX METHOD.

*2) Modified Powell Algorithm:* Powell method is also a direct search algorithm. If the tolerance is large, downhill simplex method is a better choice than the Powell algorithm. However, Powell's method is faster in most applications.

Powell's algorithm is based on line minimizations. If we start with a point $P$ in $n$-dimensional space with a new direction $u$, any function $f$ of $n$ variables can be optimized along $u$ using one-dimensional method. The key process of the Powell's algorithm is called the Linmin process (given in Table III). The major step during the Linmin process is to find $u$. Powell's algorithm is one of the suggested methods. The pseudo-code for the combined Powell's algorithm, GA and RLS estimation is as follows.

| linmin: Given as input the vectors $P$ , $n$ and function $f$, find the scalar $\lambda$ that minimizes $f(P+\lambda n)$. Replace $P$ by $P+\lambda n$ and $n$ by $\lambda n$. |
|---|

TABLE III
THE LINMIN PROCESS.

| Given the starting position $P_0$ |
|---|
| Produce the first generation with $\phi$ individuals |
| Compute fitness value $f$ for each individual |
| WHILE fitness $f < \varepsilon$ |
| Rank the individuals according to the fitness evaluation |
| Apply reproduction operation |
| Apply crossover operation |
| Apply mutation operation |
| FOR i=1 TO $\phi$ |
| DO: |
| for i=1 to number of dimensions |
| DO: |
| Move $P_i$ to the minimum along direction |
| $u_i$ and call this point $P_i$. |
| Set $u_i \leftarrow u_{i+1}$ |
| END FOR |
| Set $u_N \leftarrow P_N - P0$ |
| Move $P_N$ to the minimum along direction |
| $u_N$ and call this point $P_0$ |
| END WHILE |

TABLE IV
PSEUDO-CODE OF THE COMBINED POWELL'S ALGORITHM, GA AND RLS ESTIMATION.

## IV. SIMULATION RESULTS

The two-step method for the inference of the exact S-system model is tested using both synthetic data and microarray measurements.

### A. Optimization Range Search

In order to examine the effectiveness of the proposed procedures for parameter range search, a synthetic S-System model is used. The original S-system model is given as follows.

$$\begin{aligned} \dot{x_1} &= x_1^{0.268}x_2^{-2.26} - x_1^{0.469}x_2^{0.359} \\ \dot{x_2} &= x_1^{2.739}x_2^{0.155} - x_1^{0.197}x_2^{0.281} \end{aligned} \quad (11)$$

The parameters' range of the target model is given in the following table. The original raw data are with initial condition $[1, 1.5]$. The range of the parameters obtained from

| Items | $\alpha_i$ | $g_{i1}$ | $g_{i2}$ | $\beta_i$ | $h_{i1}$ | $h_{i2}$ |
|---|---|---|---|---|---|---|
| $x_1$ | 1.0 | 0.268 | -2.26 | 1.0 | 0.469 | 0.359 |
| $x_2$ | 1.0 | 2.739 | 0.155 | 1.0 | 0.197 | 0.281 |

TABLE V
PARAMETERS OF THE ORIGINAL S-SYSTEM.

the proposed range search process are given in the following table. From those two tables, it is observed that the proposed optimization range search process captures the correct range for the parameters. However, GP plus RLS can not converge to the exact solution. It is necessary to use either downhill simplex method or the modified Powell algorithm to calculate the exact values of the parameters.

| Items | $\alpha$<br>$\beta$ | $g_{i1}$<br>$h_{i1}$ | $g_{i2}$<br>$h_{i2}$ |
|---|---|---|---|
| $x_1$ | (-0.152,3.90)<br>(1.023,3.9) | (-1.08,1.72)<br>(-5.07,4.21) | (-10.23,6.45)<br>(-2.14,3.90) |
| $x_2$ | (0.188,3.38)<br>(0.62,3.36) | (-5.06,2.78)<br>(-2.32,3.34) | (-0.98,1.72)<br>(0.06,4.36) |

TABLE VI
OBTAINED RANGE OF THE PARAMETERS USING THE PROPOSED OPTIMIZATION RANGE SEARCH.

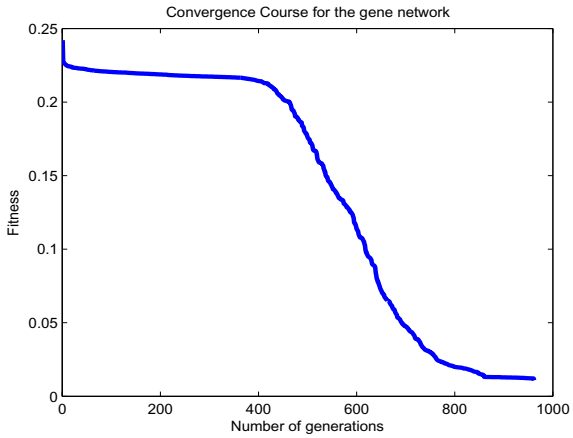


Fig. 4. Convergence course of genetic programming during parameter range search.

The convergence of the genetic programming is also shown in Fig.4.

### B. Downhill Simplex Method

In this part of the simulation, the same synthetic S-system model (equation (11)) is used. The obtained S-system model is given by

$$\begin{aligned} \dot{x_1} &= x_2^{-2.49} - x_1^{0.12} \\ \dot{x_2} &= x_1^{2.51} - x_2^{0.12} \end{aligned}$$

The value of the fitness function is $0.1038$. The trajectories of the original S-system model and the obtained model are compared in Fig. 5. It can be observed that the trajectories are very close to each other. The branch pathway of the GRN model is also given in Fig. 6.

### C. Modified Powell Method

The same synthetic S-system model (equation (11)) is used. The obtained S-system model using modified Powell method is given by

$$\begin{aligned} \dot{x_1} &= x_2^{-2.426} - x_1^{0.06}x_2^{0.008896} \\ \dot{x_2} &= x_1^{2.535} - x_1^{0.02}x_2^{0.084774} \end{aligned}$$

The fitness evaluation is $0.014$. The modified Powell method achieves better accuracy compared to the downhill simplex

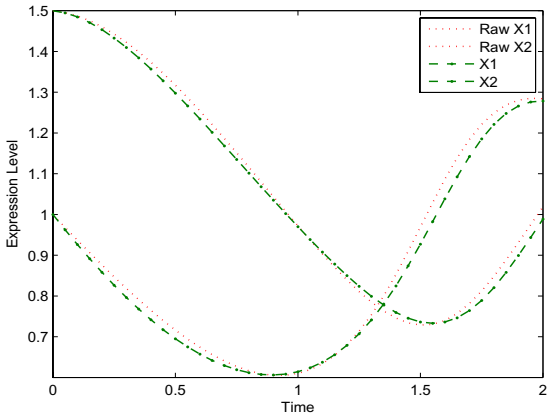


Fig. 5. The dynamics of the S-system model by downhill simplex method.

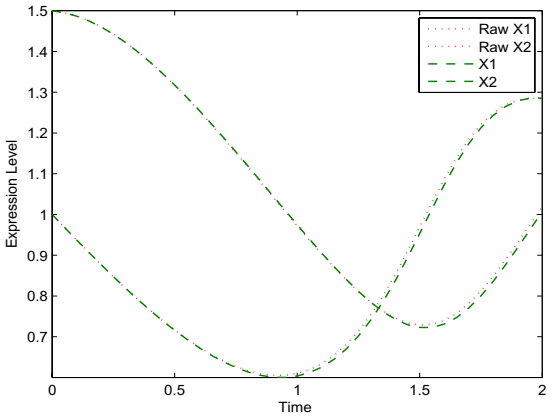


Fig. 7. The dynamics of the S-System model by Powell method

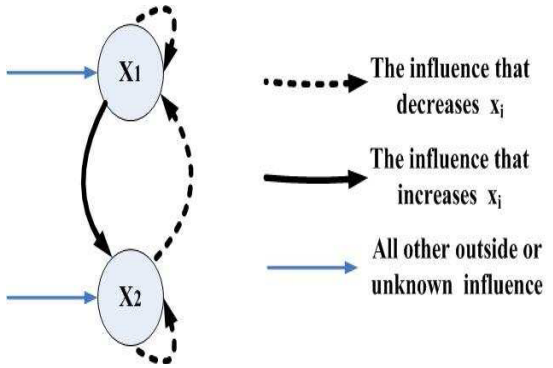


Fig. 6. The branch pathway of the 2 dimensional S-system network.

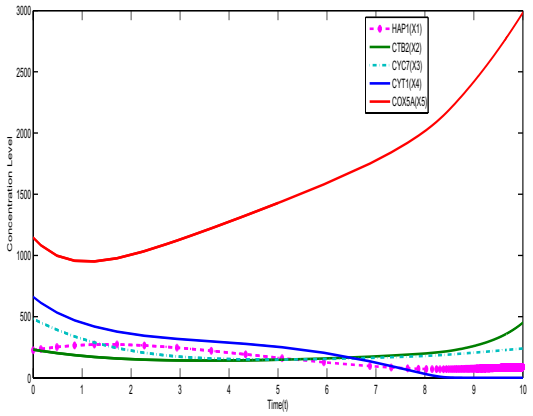


Fig. 8. The dynamics of expression level of the 5 genes in yeast.

method. The trajectories compared to the original S-system is also given in Fig. 7. It is demonstrated similar satisfactory results as in the case of downhill simplex method.

### *D. Microarray measurements of yeast*

During this part of the simulation, we consider time-series gene-expression data corresponding to yeast protein synthesis. Five genes (HAP1($x_1$), CYB2($x_2$), CYC7($x_3$), CYT1($x_4$), COX5A($x_5$)) are selected because the relations among them have been revealed by biological experiments.

Z-score [14] is applied to the results from each generation to calculate the robustness of the term and parameters. Z - score is defined by $Z(i)_j^k = (X(i)_j^k - \mu(i)_j)/\sigma(i)_j$, where $k$ is the rank index of Z-score. $\mu$ is the mean of the population. $\sigma$ is the standard deviation of the population. We use Z-score to evaluate the parameters and the one with the lowest value is chosen as the candidate. Z-score expresses the divergence of the experimental result $x_i^j$ from the most probable result $\mu$ as a number of standard deviations $\sigma$. The larger the value of $Z_i^j$, the less probability of the experimental result is due to chance.

Table VII contains the results from the proposed optimization range search. Using this table, downhill simplex method is employed to further pinpoint the values of the parameters (given in Table VIII). The relationships among the 5 genes are shown by both their trajectories given in Fig. 8 and the branch pathway model given in Fig. 9.

We observe that HAP1 represses CYC7, and CYB2 activates CYC7. It is also observed that HAP1 activates COX5A and CYT1. These observations are in agreement with the biological experiment findings in [15], [16].

## V. Conclusions and Future Work

In this paper, a unified approach to infer gene regulatory networks using the S-system model is proposed. In order to discover the structure of *large-scale* gene regulatory networks, a simplified S-system model is proposed that enables fast parameter estimation to determine the major gene interactions. If a detailed S-system model is desirable for a small group of genes, a two-step method is proposed where the range of the parameters will be determined first and the exact values of the parameters will be searched using a multi-dimensional optimization algorithm. Both downhill simplex algorithm and

| Items | $\alpha_i$ | $g_{i1}$ | $g_{i2}$ | $g_{i3}$ | $g_{i4}$ | $g_{i5}$ |
|---|---|---|---|---|---|---|
| $x_1$ | (-2.81,2.23) | (-2.86,2.17) | (-2.05,2.55) | (-2.29,2.65) | (-2.85,2.54) | (-0.152,1.31) |
| $x_2$ | (-3.04,3.38) | (-2.88,2.83) | (-2.86,0.0338) | (-2.83,0.996) | (-2.98,2.53) | (-0.26,2.01) |
| $x_3$ | (-2.46,2.19) | (-2.89,1.84) | (-3.17,2.47) | (-2.99,2.9) | (-2.87,2.14) | (-0.68,1.4) |
| $x_4$ | (-1.95,21.78) | (-2.97,1.75) | (-3.09,1.87) | (-3.14,2.56) | (-2.95,1.87) | (-0.02,1.6) |
| $x_5$ | (0,6.08) | (0,0.62) | (-0.2,0.78) | (0,0.81) | (-2.78,1.53) | (-1.27,2.09) |
| Items | $\beta_i$ | $h_{i1}$ | $h_{i2}$ | $h_{i3}$ | $h_{i4}$ | $h_{i5}$ |
| $x_1$ | (-0.107,2.13) | (-0.03,2.2) | (-0.23,0.2) | (0.12,0.5) | (0,5.11) | (0,2.08) |
| $x_2$ | (-2.79,0.605) | (-1.39,1.447) | (-0.20,0.63) | (0.1,0.7) | (0,40) | (0,1364) |
| $x_3$ | (-0.09,0.05) | (-2.84,0.96) | (0,3.24) | (-0.30,0.7) | (0,1472) | (0,845) |
| $x_4$ | (-0.03,1.08) | (-2.00,0.28) | (0,0.308) | (0,0.48) | (0,0.204) | (0,2.004) |
| $x_5$ | (0,1.24) | (0,0.25) | (-0.12,1.24) | (0.62,1.98) | (-0.28,0.52) | (0,3.28) |

TABLE VII

OPTIMIZATION RANGE SEARCH RESULTS OF THE 5-GENE NETWORK IN YEAST

| Items | $\alpha_i$ | $g_{i1}$ | $g_{i2}$ | $g_{i3}$ | $g_{i4}$ | $g_{i5}$ |
|---|---|---|---|---|---|---|
| $x_1$ | 0.113 | -0.053 | 0.099 | 1.30 | 0.098 | -0.044 |
| $x_2$ | 0.00229 | 0.5053 | -0.6515 | 0.533 | -0.232 | 1.166 |
| $x_3$ | 0.0163 | 0.688 | 0.7505 | -0.486 | 0.074 | 0.5569 |
| $x_4$ | 1883.07 | 0.00408 | -0.7939 | -0.099 | -0.104 | 0.338 |
| $x_5$ | 5.49 | 0.00024 | 0.118 | 0.6213 | 0.0047 | 0.0865 |
| Items | $\beta_i$ | $h_{i1}$ | $h_{i2}$ | $h_{i3}$ | $h_{i4}$ | $h_{i5}$ |
| $x_1$ | 0.00026 | 0.4099 | 0.113 | 0.2793 | 0.251 | 1 |
| $x_2$ | $1.58\times10^{-6}$ | 1.243 | 0.5372 | 0.5471 | 0.6375 | 0.0617 |
| $x_3$ | 0.000264 | 0.878 | 1,512 | -0.2535 | 0.154 | 0.2446 |
| $x_4$ | $3.699\times10^{-5}$ | 0.1713 | 0.9767 | 0.3245 | 0.1426 | 1 |
| $x_5$ | 0.00839 | 0.02866 | 0.4458 | 1 | 0.2955 | 0.1891 |

TABLE VIII

THE PARAMETERS OF THE EXACT S-SYSTEM MODEL OF THE 5-GENE NETWORK IN YEAST

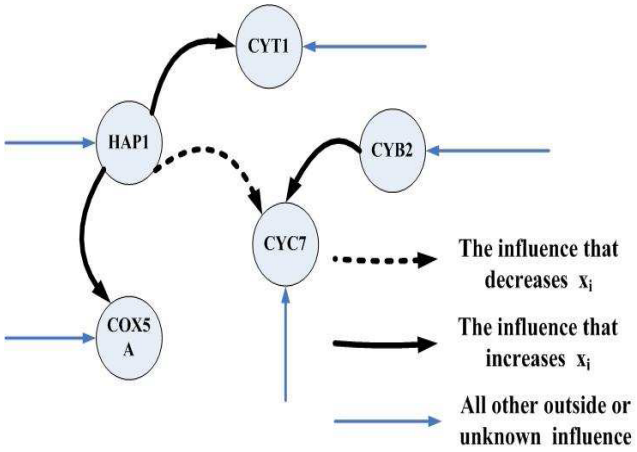


Fig. 9. The branch pathway model of the 5 genes in yeast.

modified Powell algorithm are tested for multi-dimensional optimization. Simulation results using both synthetic data and real microarray measurements demonstrate the effectiveness of the proposed methods.

Note that noise in the data complicates the parameter estimation and often leads to local minima in the search space, as well as to unwanted redundancies in inference [17]. Kalman filtering [18] may be applied to mitigate the effects of noise. And this will be one of our future efforts.

## REFERENCES

[1] M. A. Savageau, "Rules for the Evolution of Gene Circuitry", *Pac. Symp. Biocomputing*, 3: 54 - 65, 1998.

[2] L.F.A. Wessels, E.P. Van Someren, and M.J.T. Reinders, "A Comparison of Genetic Network Models", *Pac. Symp. Biocomputing*, 6: 508 - 519, 2001.

[3] E. Voit, *Computational analysis of biochemical systems*, Cambridge University Press, 2000.

[4] C.Spieth, et. al "A Memetic Inference Method for Gene Regulatory Networks Based on S-Systems", *Proceedings of the IEEE Congress on Evolutionary Computation (CEC 2004)*, pp.152-157, 2004.

[5] N. Norman and H. Iba, "Inference of Gene Regulatory Networks Using S-system and Differential Evolution", *Proceedings of GECCO*, pp.439-446, 2005.

[6] S. Kimura, et. al, "Inference of S-system models of genetic networks using a cooperative coevolutionary algorithm", *Bioinformatics*, 21(7), pp.1154-1163, 2005.

[7] S. Kikuchi, et. al, "Dynamic modeling of genetic networks using genetic algorithm and s-system", *Bioinformatics*, 19(5), pp.643-650, 2003.

[8] S. Kimura, et. al, "Inference of S-system models of genetic networks from noisy time-series data", *Chem-Bio Informatics Journal*, 4(1), pp.1-14, 2004.

[9] J. Almeida and E. Voit, "Neural-network-based parameter estimation in s-system models of biological networks", *Genome Informatics*, vol.14, pp.114-123, 2003.

[10] J. Bernon, et. al, "A comparative study of Powell's and Downhill Simplex algorithms for a fast multimodal surface matching in brain imaging", *Computerized Madical Imaging and Graphics*, 25(2001), 287-297.

[11] P. Bangert, "Downhill Simplex Methods for Optimizing Simulated Annealing are Effective", *Proceedings of ALGORITMY*, pp.341-347, 2005.

[12] W. Press, et. al, *Numerical Recipes in C*, 2nd. Ed, Cambridge University Press, 1992.

[13] J.R. Koza, *Genetic Programming: On the Programming of Computers by Means of Natural Selection*, MIT press, 1992.

[14] H. Abdi, "Z-scores", *in Neil Salkind (Ed.) (2007). Encyclopedia of Measurement and Statistics*, Thousand Oaks (CA): Sage.

[15] P. Woolf and Y. Wang, "A fuzzy logic approach to analyzing gene expression data", *Physiol. Genomics*, 3:9-15, 2000.

[16] J. Schneider and L. Guarente, "Regulation of the Yeast CYTI Gene Encoding Cytochrome cl by HAP1 and HAP2/3/4", *Molecular and Cellular Biology*, 11(10): 4934-4942, 1991.

[17] Z. Kutalik, W. Tucker, and V. Moulton, "S-system parameter estimation for noisy metabolic profiles using Newton-flow analysis", *preprint*, available at: http://www.math.uu.se/ warwick/main/papers/ktmArt.pdf

[18] S. Haykin, *Adaptive Filter Theory*, Prentice Hall, 1992.