Available online at www.sciencedirect.com
ScienceDirect
Computational Biology and Chemistry 31 (2007) 239–245
Computational Biology and Chemistry
www.elsevier.com/locate/compbiolchem

Inferring gene regulatory networks from temporal expression profiles under time-delay and noise

Shinuk Kim [a], Junil Kim [b], Kwang-Hyun Cho [a,c,*]

[a] *Bio-MAX Institute, Seoul National University, Gwanak-gu, Seoul 151-818, Republic of Korea*
[b] *Interdisciplinary Program in Bioinformatics, Seoul National University, Gwanak-gu, Seoul 151-747, Republic of Korea*
[c] *College of Medicine, Seoul National University, Jongno-gu, Seoul 110-779, Republic of Korea*

Received 15 January 2007; accepted 30 March 2007

Abstract

Ordinary differential equations (ODE) have been widely used for modeling and analysis of dynamic gene networks in systems biology. In this paper, we propose an optimization method that can infer a gene regulatory network from time-series gene expression data. Specifically, the following four cases are considered: (1) reconstruction of a gene network from synthetic gene expression data with noise, (2) reconstruction of a gene network from synthetic gene expression data with time-delay, (3) reconstruction of a gene network from synthetic gene expression data with noise and time-delay, and (4) reconstruction of a gene network from experimental time-series data in budding yeast cell cycle.

Keywords: Gene networks; Inference; Optimization; Ordinary differential equations; Noise; Time-delay

1. Introduction

While the recent developments of the high throughput measurement technologies such as DNA microarrays have made it possible to obtain a large volume of gene expression profiles, it is not answered yet how to assemble the data and construct a predictive model on the gene network. In this regard, inferring the interaction structure of a gene network is an imperative task in order to advance our understanding on the molecular mechanism of cellular functioning at a genome level.

Various approaches have been proposed to tackle such a gene network identification problem (Bansal et al., 2006; Brazhnik, 2005; Kholodenko et al., 2002; Kimura et al., 2005; Wahde and Hertz, 2000; Yeung et al., 2002). Although those approaches have been applied with some success, they are mostly demanding on the data along with a certain degree of *a priori* information. For instance, a popular approach using steady state gene expression data has been proved to be highly effective in simplifying the mathematical formulations to infer small microbial gene networks and thereby minimizing unknown parameters (Andrec et al., 2005; Gardner et al., 2003; Kholodenko et al., 2002; Sontag et al., 2004). However, the approach requires *a priori* knowledge on the genes involved in the network as well as the measurement of gene expressions at a steady state after transcriptional perturbations. On the other hand, an approach using time-series gene expression data can provide a clearer picture on the functional interaction between genes and the corresponding network. However, this approach also requires a perturbation of each gene of interest and subsequent measurements of the gene expression profiles at multiple time points.

Even though the measurement technologies have been revolutionized, noise and time-delay still have to be considered to infer gene networks under various experimental situations. However, such noise and time-delay have not been properly considered in previous studies (Dasika et al., 2004; Liu and Cao, 2006; Smolen et al., 2000; Xu et al., 2005). Hence, we propose in this paper a numerical scheme by which we can infer a gene network in consideration of such time-delay and noise.

To develop a mathematical formulation on inference of gene networks, we employ ordinary differential equations (ODEs) (Gardner et al., 2003; de Jong, 2002; Tegner et al., 2003). Based on this framework, we formulate the inference problem as a mathematical optimization problem and apply the

* Corresponding author. Present address: Bio-MAX Institute, Seoul National University, 3rd Floor, IVI, San 4-8, Bongcheon 7-dong, Gwanak-gu, Seoul 151-818, Republic of Korea. Tel.: +82 2 887 2650; fax: +82 2 887 2692.
E-mail address: ckh-sb@snu.ac.kr (K.-H. Cho).
URL: http://systemsbiology.snu.ac.kr (K.-H. Cho).

1476-9271/$ – see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.compbiolchem.2007.03.013

Broyden–Fletcher–Goldfarb–Shanno (BFGS) method (Press et al., 1992).

In particular, we consider the following two datasets: synthetic data produced from an artificial example (Sontag et al., 2004) and cell cycle experimental data (Bahler, 2005; Spell man et al., 1998). Especially, the proposed numerical scheme was applied to example networks of four nodes from G2/M and M/G1 transition phases and eight nodes representing the clusters of a budding yeast cell cycle. The remainder of this paper is organized as follows. Section 2 presents a mathematical model and proposes a numerical analysis scheme. Section 3 shows two examples illustrating the proposed scheme. Section 4 provides a summary of the method and discussions.

2. Model and method

Mathematical formulation and a schematic outline of the proposed method for inferring a dynamic gene network are described in this section.

2.1. Mathematical model

To analyze a gene regulatory system, the following nonlinear dynamic equation is employed:

$$\frac{dx}{dt} = f(x) \tag{1}$$

where $x = [x_1, x_2, \ldots, x_n]$ represents a set of gene expression levels.

If we take the first-order Taylor approximation, then

$$f(x) = f(x_0) + \frac{df(x)}{dx}(x - x_0) \tag{2}$$

where x_0 denotes the initial value of x. Based on this, we can have the following linear dynamic system (Chen et al., 1999):

$$f(x) = Ax \tag{3}$$

where A is the Jacobian matrix describing $df(x)/dx$. The component of A (i.e., $a_{ij} = df_i(x)/dx_j$) represents the gene interaction between x_i and x_j that is to be inferred. By substituting (3) into (1), we have

$$\frac{dx_i}{dt} = \sum_{j=1}^{n} a_{ij}x_j \quad i = 1, \ldots, n \tag{4}$$

where a_{ij} represents the influence of gene j on gene i. If we further consider the noise and time-delay that are usually involved in the interactions, we can rewrite this as follows:

$$\frac{dx_i(t)}{dt} = \sum_{j=1}^{n} a_{ij}(x_j(t - \tau_j) + n_{ij}) \tag{5}$$

where τ_j indicates the time-delay accompanied with the influence of x_j and n_{ij} denotes the intrinsic noise of the interaction between x_j and x_i.

2.2. Numerical scheme

In this section, we present a numerical scheme to infer gene regulatory networks. The proposed scheme consists of three main components: the direct solver, the optimization routine, and the objective function. Each component is described below in detail and the flow of overall algorithm is illustrated in Fig. 1.

2.2.1. Direct solver

We call given gene expression data 'reference data' and the data produced by solving the ODE system 'numerical data'. To solve (4), finite difference approximations are given as follows:

$$x_i(t_{k+1}) = x_i(t_{k-1}) + dt(a_{i1}x_1)(t_{k-1}) + a_{i2}x_2(t_{k-1}) + \ldots + a_{in}x_n(t_{k-1})) \tag{6}$$

where $i = 1, \ldots, n$ and $k = 1, \ldots, k_{max}$.

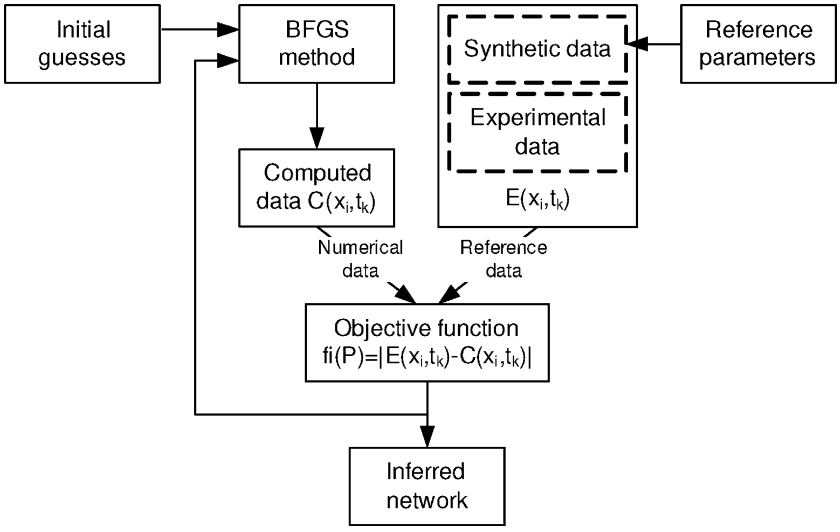

Fig. 1. A schematic diagram of the proposed numerical scheme.