Physica A 392 (2013) 1038–1049

Contents lists available at SciVerse ScienceDirect

Physica A

journal homepage: www.elsevier.com/locate/physa

Optimization-based structure identification of dynamical networks

Tao He [a], Xiliang Lu [a], Xiaoqun Wu [a,b,*], Jun-an Lu [a], Wei Xing Zheng [b]

[a] School of Mathematics and Statistics, Wuhan University, Hubei 430072, China

[b] School of Computing, Engineering and Mathematics, University of Western Sydney, Penrith NSW 2751, Australia

ARTICLE INFO

Article history:
Received 6 March 2012
Received in revised form 17 August 2012
Available online 20 November 2012

Keywords:
Dynamical networks
Structure identification
Optimization problem
Projected conjugate gradient method

ABSTRACT

The topological structure of a dynamical network plays a pivotal part in its properties, dynamics and control. Thus, understanding and modeling the structure of a network will lead to a better knowledge of its evolutionary mechanisms and to a better cottoning on its dynamical and functional behaviors. However, in many practical situations, the topological structure of a dynamical network is usually unknown or uncertain. Thus, exploring the underlying topological structure of a dynamical network is of great value. In recent years, there has been a growing interest in structure identification of dynamical networks. As a result, various methods for identifying the network structure have been proposed. However, in most of the previous work, few of them were discussed in the perspective of optimization. In this paper, an optimization algorithm based on the projected conjugate gradient method is proposed to identify a network structure. It is straightforward and applicable to networks with or without observation noise. Furthermore, the proposed algorithm is applicable to dynamical networks with partially observed component variables for each multidimensional node, as well as small-scale networks with time-varying structures. Numerical experiments are conducted to illustrate the good performance and universality of the new algorithm.

1. Introduction

The last decade had witnessed the birth and explosive growth of the field of complex dynamical networks, triggered by the seminal works of Watts and Strogatz [1] on small-world networks in 1998 and of Barabási and Albert [2] on scale-free networks in 1999. Nowadays, dynamical networks have become ubiquitous as models of many real-world systems. Examples range from the Internet, the World Wide Web, power grids [3], transportation networks, social and economic networks [4] to a plethora of examples in systems biology like gene regulatory networks [5], metabolic networks [6] and so on.

In the initial phase of the field's development, the focus was mainly placed upon the study of statistical properties, collective behaviors and control of dynamical networks with predefined topological structures [7], which has provided a tremendous insight into the interplay between the structures and functions. As is known, the topological structure of a dynamical network plays a pivotal role in determining its properties, dynamics and control [8]. Thus, understanding and modeling the structure of a dynamical network will lead to a better knowledge of its evolutionary mechanisms, and to a better control on its dynamical as well as functional behaviors [9]. However, in reality, it is invariably the case that the topological structure of a dynamical network is unknown or uncertain. So it is very important to investigate the underlying topological structure of a network. Indeed, active research during recent years has resulted in various approaches for the identification of network structures.

* Corresponding author at: School of Mathematics and Statistics, Wuhan University, Hubei 430072, China. Tel.: +86 27 87229812.
E-mail address: xqwu@whu.edu.cn (X. Wu).

0378-4371/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
doi:10.1016/j.physa.2012.11.014

For interacting deterministic systems, some new approaches have been presented for topology identification [10–15] based on adaptive control and outer synchronization between two networks [16,17]. For simultaneously observed time series, some techniques based on measuring the cross correlation or partial correlation among time series have been proposed. For example, a method based on the dynamical correlation was proposed by Ren et al. [18] to predict the network structure. Granger causality is a widely employed technique based on linear regression for causal link detection, particularly in neuroscience and economics [19]. Many extensions have also been made to generalize the linear Granger causality to the nonlinear case [20]. Inferring directed connections among observed time series using Bayesian networks is also widely used, where the Bayesian models [21] represent probabilistic relationship between multiple interacting entities. Some interesting techniques based on the theory of recurrences [22] have been proposed. In 2007, Romano et al. introduced a method to uncover directional coupling between two interacting systems based on the mean conditional probabilities of recurrence, which is applicable to both weak and strong couplings [23]. Later on, by generalizing partial phase synchronization, Nawrath et al. proposed partial recurrence-based synchronization analysis for inferring the interactions of oscillators with multiple time scales [24]. Very recently, a permutation-based measure named inner composition alignment was introduced to identify relations between subsystems [25].

However, in the synchronization-based methods, the interacting systems and observed data have to be noise free, which usually does not conform to practical cases [10–15]. The correlation-based methods are incapable of distinguishing between direct and indirect interactions, which in many situations do not provide very satisfactory results [18]. Those techniques based on Granger causality [19,20] or Bayesian networks [21] are usually computationally expensive and require a huge amount of memory and time for computation. The methods based on recurrence properties [23,24] can only deal with very small-scale networks, while the performance of the permutation-based method [25] depends on the density of the considered network. Moreover, most of the approaches cannot tell the coupling strength among nodes [18–21,23–25].

On the other hand, the structure identification problem is a typical parameter identification problem if the configuration matrix is taken as a parameter, and optimization methods are widely used in such parameter identification problems, see Refs. [26–28]. Based on the above discussions, in this paper, a technique for inferring the topology of a weighted dynamical network topology is proposed based on optimization. The proposed optimization based algorithm is also robust to observation noise. In particular, the new algorithm works well when only some component variables of each multidimensional node are observable. Interestingly, the method is applicable to small-scale networks with time-varying topological structures as well.

The rest of the paper is organized as follows. Theoretical modeling and mathematical analysis of network structure identification based on optimization are presented in Section 2. The detailed algorithm based on the projected conjugate gradient method is given in Section 3. Five illustrative examples are provided in Section 4 for verification and demonstration. Finally, some conclusions and an outlook are summarized in Section 5.

2. Problem formulation and analysis

Consider the following dynamical network consisting of N linearly-coupled systems:

$$\dot{\mathbf{x}}_i(t) = \mathbf{f}_i(\mathbf{x}_i(t), t) + \sum_{j=1}^{N} c_{ij} A \mathbf{x}_j(t), \quad i = 1, 2, \ldots, N, \tag{1}$$

where $\mathbf{x}_i(t) = (x_{i1}(t), x_{i2}(t), \ldots, x_{in}(t))^\top \in \mathbb{R}^n$ is the state vector of the i-th node, and $\mathbf{f}_i : \mathbb{R}^n \times \mathbb{R}^+ \mapsto \mathbb{R}^n$ is a smooth vector field governing the dynamics of the isolated i-th node. The configuration matrix $C = (c_{ij})_{N\times N}$ describes the coupling topology of the network, in which $c_{ij} \neq 0$ if there is a coupling from node j to node i ($j \neq i$); otherwise $c_{ij} = 0$. Moreover, the configuration matrix satisfies the constraints that the sum of entries of each row is zero, i.e., $\sum_{j=1}^{N} c_{ij} = 0$, for $i = 1, 2, \ldots, N$. Matrix $A \in \mathbb{R}^{n\times n}$ is the inner-coupling matrix. For simplicity, A is assumed to be an identity matrix, i.e., $A = I_{n\times n}$. Thus, the network model can be written as

$$\dot{\mathbf{x}}_i(t) = \mathbf{f}_i(\mathbf{x}_i(t), t) + \sum_{j=1}^{N} c_{ij} \mathbf{x}_j(t), \quad i = 1, 2, \ldots, N. \tag{2}$$

Introduce $X = (x_{11}, x_{12}, \ldots, x_{1n}, \ldots, x_{N1}, x_{N2}, \cdots, x_{Nn})^\top \in \mathbb{R}^{Nn}$ and $F = (f_{11}, f_{12}, \ldots, f_{1n}, \ldots, f_{N1}, f_{N2}, \ldots, f_{Nn})^\top \in \mathbb{R}^{Nn}$. Assume that the initial value of $X(t)$ at $t = 0$ is X_0. Then network (2) can be rewritten as

$$\begin{cases} \dot{X} = F(X) + C \otimes I_{n\times n} \cdot X, \\ X(0) = X_0, \end{cases} \tag{3}$$

where $\otimes$ denotes the Kronecker product.

With a given configuration matrix C, the nodal dynamics $X(t)$ is thus determined by (3). Define a map $M : \mathbb{R}^{N\times N} \mapsto \mathbb{R}^{Nn}$ as $C \mapsto X(t) = M(C)$, and the admissible set $\mathcal{B}$ for possible configuration matrix C as

$$\mathcal{B} = \left\{ C \in \mathbb{R}^{N\times N} \,\middle|\, \sum_{j=1}^{N} c_{ij} = 0,\ i = 1, 2, \ldots, N \right\}. \tag{4}$$

Our goal is to recover the configuration matrix C from observation data X_v, which are obtained from network (3). It is known that this configuration identification is a typical parameter identification problem, which is subjected to the strong ill-posedness in the sense of Hadamard.

This ill-posedness can be viewed from the following three aspects.

(a) Since the number of observation data is larger than the number of unknown parameters ($N \times (N-1)$ in this case), generally one does not have the existence of a strong solution in the classical sense. This also reflects the problem setting: to identify the network configuration, partial data (information either from partial component variables or from short observation time) may be enough.

(b) If the network becomes synchronized after a certain time, using the information after synchronization may lead to the non-uniqueness of the configuration matrix. This is the already known problem: Synchronization is an obstacle to structure identification [29].

(c) Stability is a more important issue for numerical computation. One may use a simple example to see the instability for identification problem. Suppose the governing equation is a scalar ordinary differential equation as follows,

$$\dot{x} = cx, \quad x(0) = 1.$$

If $x(t) \equiv 1$, then $c = 0$. But if we slightly change observation data x to $x(t) = 1 + \delta \sin(nt)$, then $c = \frac{\delta n \cos(nt)}{1+\delta \sin(nt)}$, which implies a high frequency observation noise may cause a large error in the solution. On the other hand, since the classical solution for this problem does not exist generally, to define a stable solution with respect to noisy observation data is a key step to identify the network configuration.

To overcome the above mentioned difficulties, the problem of estimating the elements of the configuration matrix C is formulated as the minimal optimization problem with an objective function described by

$$\min_{C\in\mathcal{B}} J_0(C) = \frac{1}{2}\int_0^T \|M(C) - X_v\|^2 \,\mathrm{d}t. \tag{5}$$

Moreover, To increase the numerical stability, a classical Tikhonov regularization [30] is introduced in the function J_0. The regularized problem reads as

$$\min_{C\in\mathcal{B}} J_\alpha(C) = \frac{1}{2}\int_0^T \|M(C) - X_v\|^2 \,\mathrm{d}t + \frac{\alpha}{2}\|C\|^2 \tag{6}$$

where α is the regularization parameter to avoid the large deviation from the optimal solution.

The first-order derivative of the map M can be obtained by the following technique. Let $\varepsilon\Delta C$ be a small increment and $(C + \varepsilon\Delta C) \in \mathcal{B}$, then

$$\frac{\delta M}{\delta C}\Delta C = \lim_{\varepsilon\to 0}\frac{M(C+\varepsilon\Delta C) - M(C)}{\varepsilon} = \frac{\mathrm{d}}{\mathrm{d}\varepsilon}M(C+\varepsilon\Delta C)|_{\varepsilon=0}.$$

For $i = 1, 2, \ldots, N$,

$$\begin{cases} \dot{\mathbf{x}}_i(C+\varepsilon\Delta C) = \mathbf{f}_i(\mathbf{x}_i(C+\varepsilon\Delta C), t) + \sum_{j=1}^{N}(c_{ij} + \varepsilon(\Delta c)_{ij})\mathbf{x}_j(C+\varepsilon\Delta C), & \text{(a)} \\ \dot{\mathbf{x}}_i(C) = \mathbf{f}_i(\mathbf{x}_i(C), t) + \sum_{j=1}^{N} c_{ij}\mathbf{x}_j(C). & \text{(b)} \end{cases} \tag{7}$$

Subtracting (7)(b) from (7)(a) yields

$$\begin{aligned} \dot{x}_{ik}(C+\varepsilon\Delta C) - \dot{x}_{ik}(C) &= f_{ik}(x_{i1}(C+\varepsilon\Delta C), \ldots, x_{in}(C+\varepsilon\Delta C)) \\ &\quad - f_{ik}(x_{i1}(C), \ldots, x_{in}(C)) + \sum_{j=1}^{N}[(c_{ij} + \varepsilon(\Delta c)_{ij})\cdot x_{jk}(C+\varepsilon\Delta C) - c_{ij}\cdot x_{jk}(C)] \\ &= \frac{\partial f_{ik}}{\partial x_{i1}}(x_{i1}(C+\varepsilon\Delta C) - x_{i1}(C)) + \cdots + \frac{\partial f_{ik}}{\partial x_{in}}(x_{in}(C+\varepsilon\Delta C) - x_{in}(C)) \\ &\quad + \sum_{j=1}^{N} c_{ij}(x_{jk}(C+\varepsilon\Delta C) - x_{jk}(C)) + \varepsilon\sum_{j=1}^{N}(\Delta c)_{ij}\cdot x_{jk} + o(\varepsilon). \end{aligned}$$

Let $w_{ik} = \frac{\mathrm{d}}{\mathrm{d}\varepsilon}x_{ik}(C+\varepsilon\Delta C)|_{\varepsilon=0}$, and take the limit $\varepsilon \to 0$, then

$$\dot{w}_{ik} = \frac{\partial f_{ik}}{\partial x_{i1}}w_{i1} + \cdots + \frac{\partial f_{ik}}{\partial x_{in}}w_{in} + \sum_{j=1}^{N} c_{ij}w_{jk} + \sum_{j=1}^{N}(\Delta c)_{ij}x_{jk}.$$

Define $W = (w_{11}, \ldots, w_{1n}, \ldots, w_{N1}, \ldots, w_{Nn})^\top \in \mathbb{R}^{Nn}$, and

$$\nabla \mathbf{f}_i = \begin{pmatrix} \dfrac{\partial f_{i1}}{\partial x_{i1}} & \cdots & \dfrac{\partial f_{i1}}{\partial x_{in}} \\ \vdots & \cdots & \vdots \\ \dfrac{\partial f_{in}}{\partial x_{i1}} & \cdots & \dfrac{\partial f_{in}}{\partial x_{in}} \end{pmatrix}_{n\times n}.$$

Then the corresponding system for the first-order derivative of map M can be obtained as follows:

$$\begin{cases} \dot{W} = \operatorname{diag}(\nabla \mathbf{f}_1, \ldots, \nabla \mathbf{f}_N) \cdot W + C \otimes I_{n\times n} \cdot W + \Delta C \otimes I_{n\times n} \cdot X, \\ W(0) = 0. \end{cases} \tag{8}$$

Let the Lagrangian functional $\mathcal{L}(X, C, P, \lambda)$ be defined by

$$\mathcal{L}(X, C, P, \lambda) = J_\alpha(C) + \int_0^T (\dot{X} - F - C \otimes I_{n\times n} \cdot X) \cdot P \, \mathrm{d}t + \sum_{i=1}^N \lambda_i \left(\sum_{j=1}^N c_{ij} \right), \tag{9}$$

where P is the adjoint variable and $\lambda_i (i = 1, 2, \ldots, N)$ are Lagrange multipliers associated with the constraints $\sum_{j=1}^N c_{ij} = 0$ $(i = 1, 2, \ldots, N)$. Taking the derivative of $\mathcal{L}(X, C, P, \lambda)$ with respect to P and λ_i, the primal problem is thus equal to the governing equations (3) and the constraints $C \in \mathcal{B}$.

To get the adjoint equation, the first-order derivative with respect to X should satisfy $\frac{\delta \mathcal{L}}{\delta X} Y = 0, \forall\, Y$, such that $X + \varepsilon Y$ satisfies the governing dynamical system (3). Since

$$\begin{aligned} \frac{1}{\varepsilon}(\mathcal{L}(X + \varepsilon Y, C, P, \lambda) - \mathcal{L}(X, C, P, \lambda)) &= \frac{1}{2\varepsilon} \int_0^T (|X + \varepsilon Y - X_v|^2) \, \mathrm{d}t - \frac{1}{2\varepsilon} \int_0^T (|X - X_v|^2) \, \mathrm{d}t \\ &\quad + \int_0^T (\dot{Y} - C \otimes Y \cdot Y) P \, \mathrm{d}t - \frac{1}{\varepsilon} \int_0^T (F(X + \varepsilon Y) - F(X)) P \, \mathrm{d}t \\ &= \int_0^T Y(X - X_v) \, \mathrm{d}t + \int_0^T \left[\frac{\mathrm{d}}{\mathrm{d}t}(YP) - Y\dot{P} - (C^T \otimes I_{n\times n} \cdot P) Y \right] \mathrm{d}t \\ &\quad - \int_0^T (\operatorname{diag}(\nabla \mathbf{f}_1, \ldots, \nabla \mathbf{f}_N) \cdot P) Y \, \mathrm{d}t + O(\varepsilon), \end{aligned}$$

we have

$$0 = YP \Big|_0^T + \int_0^T (-\dot{P} - \operatorname{diag}(\nabla \mathbf{f}_1, \ldots, \nabla \mathbf{f}_N) \cdot P - C^\top \otimes I_{n\times n} \cdot P + X - X_v) \cdot Y \, \mathrm{d}t.$$

From $(X + \varepsilon Y)(0) = X(0)$, and $Y(0) = 0$, it follows that

$$Y(T)P(T) + \int_0^T (-\dot{P} - \operatorname{diag}(\nabla \mathbf{f}_1, \ldots, \nabla \mathbf{f}_N) \cdot P - C^\top \otimes I_{n\times n} \cdot P + X - X_v) \cdot Y \, \mathrm{d}t = 0.$$

For the arbitrary Y, the adjoint equation reads as

$$\begin{cases} -\dot{P} = \left(\operatorname{diag}(\nabla \mathbf{f}_1, \ldots, \nabla \mathbf{f}_N) + C^\top \otimes I_{n\times n}\right) P - (X - X_v), \\ P(T) = 0. \end{cases} \tag{10}$$

To arrive at a complete set of the first-order necessary conditions, the optimality condition is needed by applying a similar argument. The first-order derivative with respect to parameter C is zero, i.e., $\frac{\delta \mathcal{L}}{\delta C} D = 0, \forall\, D$. Then

$$\begin{aligned} \frac{1}{\varepsilon}(\mathcal{L}(X, C + \varepsilon D, P, \lambda) - \mathcal{L}(X, C, P, \lambda)) &= \frac{\alpha}{2\varepsilon}(\|C + \varepsilon D\|^2 - \|C\|^2) - \int_0^T (D \otimes I_{n\times n} \cdot X) \cdot P \, \mathrm{d}t + \lambda_i \sum_{j=1}^N d_{ij} \\ &= \alpha \sum_{i,j} c_{ij} d_{ij} - \int_0^T \sum_{ij} d_{ij} \mathbf{p}_i \cdot \mathbf{x}_j \, \mathrm{d}t + \sum_i \lambda_i \sum_j d_{ij} + \frac{\alpha\varepsilon}{2} \|D\|^2 \\ &= \sum_{ij} d_{ij} \left(\alpha c_{ij} - \int_0^T \mathbf{p}_i \cdot \mathbf{x}_j \, \mathrm{d}t + \lambda_i \right) + \frac{\alpha\varepsilon}{2} \|D\|^2. \end{aligned}$$

Taking the limit $\varepsilon \to 0$ and noticing the arbitrarily chosen d_{ij}, the optimality condition reads as

$$\alpha c_{ij} - \int_0^T \mathbf{p}_i \cdot \mathbf{x}_j \, \mathrm{d}t + \lambda_i = 0 \quad i, j = 1, 2, \ldots, N. \tag{11}$$

The following theorem is a combination of the primal equation, the adjoint equation and the optimality condition.

Theorem 2.1. *The first-order optimality system (K.K.T. condition) for problem (6) is as follows:*

$$\begin{cases} \text{primal equation } \dot{X} = F(X) + C \otimes I_{n\times n} \cdot X, \qquad X(0) = X_0, \quad C \in \mathcal{B}, \\ \text{adjoint equation } -\dot{P} = (\operatorname{diag}(\nabla \mathbf{f}_1, \ldots, \nabla \mathbf{f}_N) + C^{\top} \otimes I_{n\times n})P - (X - X_v), \quad P(T) = 0 \\ \text{optimality condition } \alpha c_{ij} - \displaystyle\int_0^T \mathbf{p}_i \cdot \mathbf{x}_j \, \mathrm{d}t + \lambda_i = 0 \quad i, j = 1, 2, \ldots, N. \end{cases}$$

To apply the conjugate gradient method which will be presented in details in the next section, we need the information of the gradient of the objective function $\nabla J_\alpha(C)$. For any matrix ΔC, we have

$$\nabla J_\alpha(C)\Delta C = \frac{\mathrm{d}}{\mathrm{d}\varepsilon} J_\alpha(C + \varepsilon \Delta C)|_{\varepsilon=0} = \lim_{\varepsilon \to 0} \frac{J_\alpha(C + \varepsilon \Delta C) - J_\alpha(C)}{\varepsilon}.$$

Since

$$\begin{aligned} J_\alpha(C + \varepsilon \Delta C) - J_\alpha(C) &= \frac{1}{2} \int_0^T |M(C + \varepsilon \Delta C) - X_v|^2 + \frac{\alpha}{2} \parallel (C + \varepsilon \Delta C) \parallel_2^2 \\ &\quad - \frac{1}{2} \int_0^T |M(C) - X_v|^2 - \frac{\alpha}{2} \parallel C \parallel_2^2 \\ &= \varepsilon \int_0^T W(X - X_v) \, \mathrm{d}t + \alpha \varepsilon C \cdot \Delta C + o(\varepsilon), \end{aligned}$$

where W is the first-order derivative of map M, see Eq. (8), the gradient of the objective function is given by

$$\nabla J_\alpha(C)\Delta C = \int_0^T W(X - X_v) \, \mathrm{d}t + \alpha C \cdot \Delta C.$$

Using the adjoint equation (10) and the derivative of map M in (8), we obtain

$$\begin{aligned} \nabla J_\alpha(C)\Delta C &= \int_0^T W(\dot{P} + \operatorname{diag}(\nabla \mathbf{f}_1, \ldots, \nabla \mathbf{f}_n) \cdot P + C^{\top} \otimes I_{n\times n} \cdot P) \, \mathrm{d}t + \alpha C \cdot \Delta C \\ &= WP \Big|_0^T + \int_0^T (-\dot{W} + \operatorname{diag}(\nabla \mathbf{f}_1, \ldots, \nabla \mathbf{f}_N) W + C \otimes I_{n\times n} \cdot W) P \, \mathrm{d}t + \alpha C \cdot \Delta C \\ &= \int_0^T (-\Delta C \otimes I_{n\times n} \cdot X) P \, \mathrm{d}t + \alpha C \cdot \Delta C. \end{aligned}$$

Then the matrix ΔC can be eliminated and the gradient can be described as follows:

$$[\nabla J_\alpha(C)]_{ij} = -\int_0^T \mathbf{p}_i^{\top} \mathbf{x}_j \, \mathrm{d}t + \alpha c_{ij}, \quad (j = 1, 2, \ldots, N). \tag{12}$$

3. The optimization algorithm based on the projected conjugate gradient method

In this section, we propose an optimization algorithm based on the projected conjugate gradient method with the Fletcher-Reeves formula to solve the structure identification problem formulated in the preceding section.

The detailed steps of the optimization algorithm are described as follows.

Optimization algorithm

Step 1: Set the initial value $z_0 = C_0$ and $k = 0$.

Step 2: Solve the governing equation (3) and the dual equation (10) for $C = C_k$ by the fourth-order Runge–Kutta method. Compute the value of the objective function $J_\alpha(z_k)$ and its gradient $g_k = \nabla J_\alpha(z_k)$ by (12).

Step 3: If $\|g_k\| < \varepsilon$, then $z^* = z_k$ and stop. Otherwise go to Step 4.

Step 4: Determine a searching direction d_k:

$$d_k = \begin{cases} -g_k + \beta_{k-1} d_{k-1} & k \geq 1 \\ -g_k & k = 0, \end{cases}$$

where $\beta_{k-1} = \frac{g_k^{\top} g_k}{g_{k-1}^{\top} g_{k-1}}$ (Fletcher–Reeves formula).

Step 5: Choose the step length α_k by line search with the Armijo–Goldstein rule:

$$\begin{cases} J_\alpha(z_k + \alpha_k d_k) \le J_\alpha(z_k) + \rho\alpha_k \nabla J_\alpha(z_k)^\top d_k \\ J_\alpha(z_k + \alpha_k d_k) \ge J_\alpha(z_k) + (1-\rho)\alpha_k \nabla J_\alpha(z_k)^\top d_k, \end{cases}$$

where $\rho \in (0, 0.5)$, and $\rho = 0.1$ in general.

Step 6: Compute the new point $z_{k+1} = z_k + \alpha_k d_k$, and project it to the admissible set $\mathcal{B}$, i.e., $z_{k+1} = \mathcal{P} z_{k+1}$, where the projection operator $\mathcal{P}$ is mapped from $\mathbb{R}^{N\times N}$ to its subspace $\mathcal{B}$. Set $k = k + 1$ and go to Step 2.

To complete the algorithm, the formula for the projection $\mathcal{P}$ in Step 6 is needed. By definition,

$$\mathcal{P}(C) = \arg\min_{D\in\mathcal{B}} \|C - D\|^2.$$

Since the constraints are independent for each row, this projection can be computed row by row. For this purpose, the following subproblem needs to be solved analytically:

$$\min_{\sum b_i = 0} \frac{1}{2}\sum_{i=1}^{N} (b_i - a_i)^2,$$

where $a_1, \ldots, a_N$ are real numbers. Its optimal solution can be obtained by the standard technique as

$$b_i = a_i - \frac{a_1 + \cdots + a_N}{N}.$$

According to the above subproblem, the formula of the projection $\mathcal{P}$ is given by

$$(\mathcal{P}(C))_{ij} = c_{ij} - \frac{\sum_{j=1}^{N} c_{ij}}{N}.$$

Remark 3.1. In each iteration of the projected conjugate gradient algorithm, most computational cost is spent in solving the governing equation (3) and the dual equation (10) (in Step 2). The total computational time depends on how to solve Eqs. (3) and (10), and how many iterations are needed. As mentioned before, usually only partial observation data or data from short observation times are needed to identify the network structure, which can dramatically reduce the computational cost in solving ordinary differential equations. On the other hand, a good initial guess may dramatically decrease the number of iteration steps. According to our empirical experience, usually 400 to 800 conjugate gradient iterations are needed to find an optimal solution.

4. Numerical experiments

In this section, five numerical examples are presented to illustrate the effectiveness and robustness of the proposed optimization-based topology identification method. According to Chen et al. [29], synchronization is an obstacle to topology identification. Therefore, it is important to ensure that different nodes of a dynamical network are not in any kind of synchronization, otherwise it is impossible to differentiate one from another, which immediately leads to identification failure.

Consider the weighted dynamical network consisting of N linearly coupled oscillators as shown in (1). The well-known chaotic Lorenz system is taken as the dynamics of every node in the network, and is described by

$$\begin{cases} f_{i1} = -ax_{i1} + ax_{i2}, \\ f_{i2} = cx_{i1} - x_{i2} - x_{i1}x_{i3}, \\ f_{i3} = -bx_{i3} + x_{i1}x_{i2}, \end{cases} \tag{13}$$

where $a = 10$, $b = 8/3$, $c = 28$, and $i = 1, 2, \ldots, N$.

In the following numerical simulations, the regularization parameter is $\alpha = 10^{-4}$, $\rho = 0.1$, and the maximum number of iterations is set to be 800. The Runge–Kutta fourth order method is employed to solve the ordinary differential equations of the dynamical network (3) in the time interval $[0, 0.1]$ with a time step 10^{-4}. The initial values of $\mathbf{X}_0$ in the dynamical network are randomly assigned. Unless otherwise specified, the data used for numerical experiments are sampled with a time step 10^{-3}. That is, 100 sets of data are used for structure identification.

4.1. Structure identification of a weighted directed network

A general directed network model with three Lorenz oscillators is considered, whose weighted coupling matrix $C = (c_{ij})_{N\times N}$ is given by

$$C = \begin{pmatrix} -4 & 2 & 2 \\ 1 & -2 & 1 \\ 0 & 1 & -1 \end{pmatrix}. \tag{14}$$

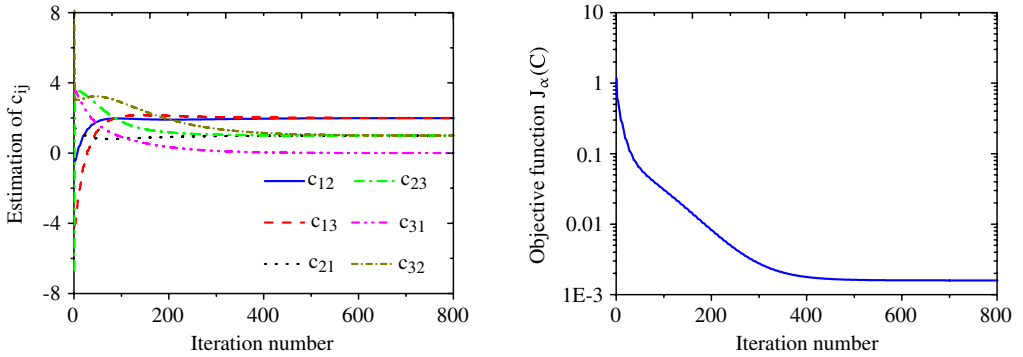

Fig. 1. (Color online) Left: Identification of the topological structure for a directed network; Right: The logarithmic value of the objective function $J_\alpha(C)$.

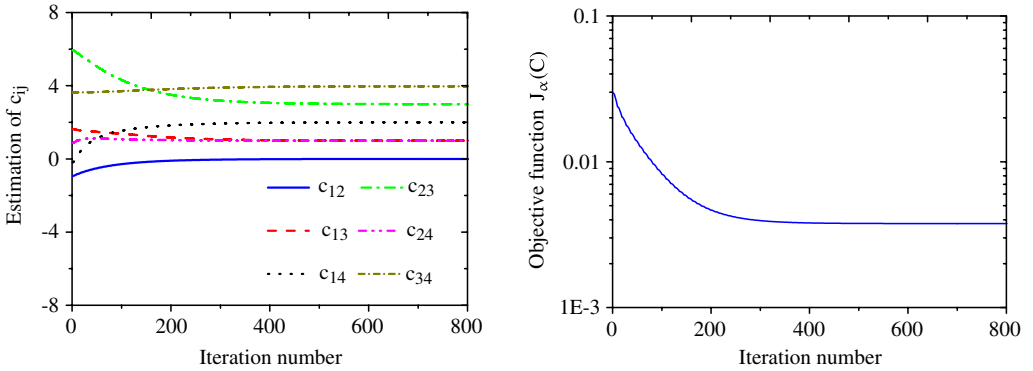

Fig. 2. (Color online) Left: Identification of the topological structure for an undirected network with observation noise, where $X_v = (1+1.5\%\,\mathrm{rand}(1))X(t)$; Right: The logarithmic value of the objective function $J_\alpha(C)$.

Since the initial nodal values of the original networks and C_0 of the objective function are randomized, the optimization-based algorithm is repeated for ten times and the estimation values of the coupling matrix are averaged at each iteration. The results of structure identification are shown in the left panel of Fig. 1. It is obvious that the network structure is correctly identified after approximate 400 iterations. The average evolutionary value of objective function $J_\alpha(C)$ at each iteration step is displayed in the right panel of Fig. 1. It can be seen from the panel that the objective function quickly goes to zero once the topological structure is correctly identified.

4.2. Structure identification with observation noise

In real-world applications, observation noise commonly exists. Thus, in order to test the robustness of the proposed optimization-based structure identification technique, 1.5% random relative noise is added as observation noise to the observed time series $X(t)$. That is to say, $X_v = (1 + 1.5\%\,\mathrm{rand}(1))X(t)$, other than $X_v = X(t)$, is used as the observed data in the objective function for identification.

Consider an undirected network consisting of four Lorenz oscillators, whose coupling matrix is given by

$$C = \begin{pmatrix} -3 & 0 & 1 & 2 \\ 0 & -4 & 3 & 1 \\ 1 & 3 & -8 & 4 \\ 2 & 1 & 4 & -7 \end{pmatrix}. \tag{15}$$

The left panel of Fig. 2 shows the results of structure identification with observation noise on node dynamics. Since the coupling matrix is symmetric, only the elements of the upper triangular matrix are displayed. One can see that all the couplings are correctly identified. The right panel of Fig. 2 displays the corresponding objective function $J_\alpha(C)$ varying with the number of iterations, which descends below 4×10^{-3} after the structure is successfully recovered. It can be seen that the objective function $J_\alpha(C)$ does not achieve the lower value as that of Fig. 1, where the discrepancy is due to the effect of observation noise.

As is known, the synchronization-based methods have gathered wide attention and great interest in the field of network topology identification [10–15]. Fig. 3 displays the identification results obtained from noisy observed data according to this technique. Specifically, a network composed of four coupled Lorenz oscillators is employed here as the unknown network,

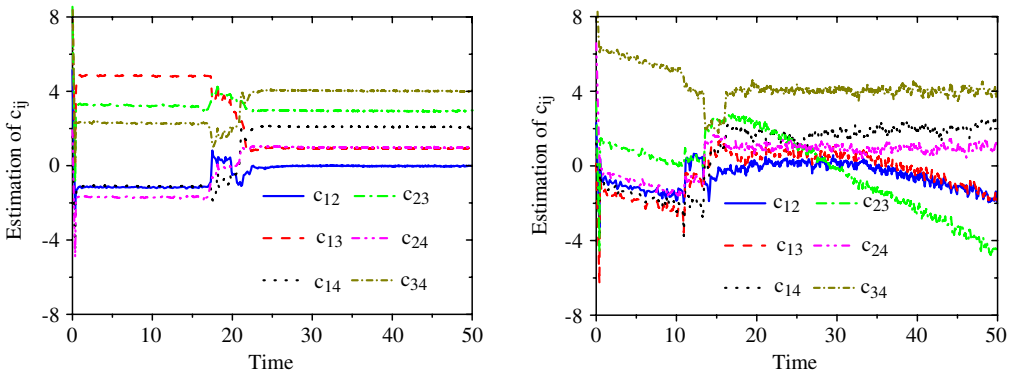

Fig. 3. (Color online) Synchronization-based topology identification with observation noise. Left: $X_v = (1 + 0.3\%\,\mathrm{rand}(1))X(t)$; Right: $X_v = (1 + 1.5\%\,\mathrm{rand}(1))X(t)$.

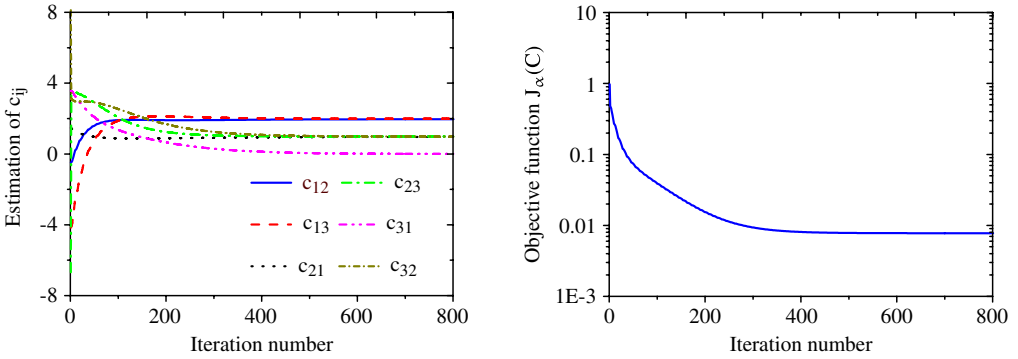

Fig. 4. (Color online) Left: Identification of the topological structure for a directed network with partially observable variables; Right: The logarithmic value of the objective function $J_\alpha(C)$.

with the topological structure (15) to be recovered. According to the synchronization-based technique [10–15], a response network is correspondingly constructed and some adaptive controllers and updating laws are designed to recover the unknown structure using the observed node dynamics of the unknown network. In the left panel of Fig. 3, 0.3% random relative noise is added as observation noise to the observed time series $X(t)$, that is, $X_v = (1 + 0.3\%\,\mathrm{rand}(1))X(t)$. It is seen from this panel that the structure is roughly recovered. In the right panel of Fig. 3, 1.5% random relative noise is added. It is obvious that when the level of observation noise is increased, the synchronization-based technique fails to recover the underlying network structure. In contrast, the optimization-based technique is more robust to observation noise, as shown in Fig. 2.

4.3. Structure identification with partially observable variables

In real-world dynamical networks such as neural networks, the behavior of an individual node is usually described by multidimensional variables. However, in many practical situations, only some component variables of a node are observable. Therefore, in this subsection, such a practical situation is considered.

The network model as considered in Section 4.1 is employed again. But it is assumed that only the second and third component variables of each Lorenz oscillator are observable. That is, for each three-dimensional node $\mathbf{x}_i(t) = (x_{i1}, x_{i2}, x_{i3})^\top$, only the component variables $(x_{i2}, x_{i3})^\top$ are observed. Fig. 4 displays the identification results, which again verify the effectiveness of the proposed optimization-based structure identification algorithm. It can be seen from the right panel that the objective function $J_\alpha(C)$ decreases to around 7×10^{-3} after 400 iterations. The reason why the objective function $J_\alpha(C)$ does not achieve a lower value as that in Fig. 1 is the effect of unobservable component variables. However, the topological structure of the network is correctly inferred.

4.4. Structure identification of a 20-node network

In this subsection, the proposed method is tested on a network consisting of 20 nodes. The Watts–Strogatz (WS) [1] algorithm is employed here to generate a small-world network. Specifically, start from a ring-shaped network with 20 nodes, with each node connecting to its 4 nearest neighbors. Then, rewire each edge in such a way that the beginning end

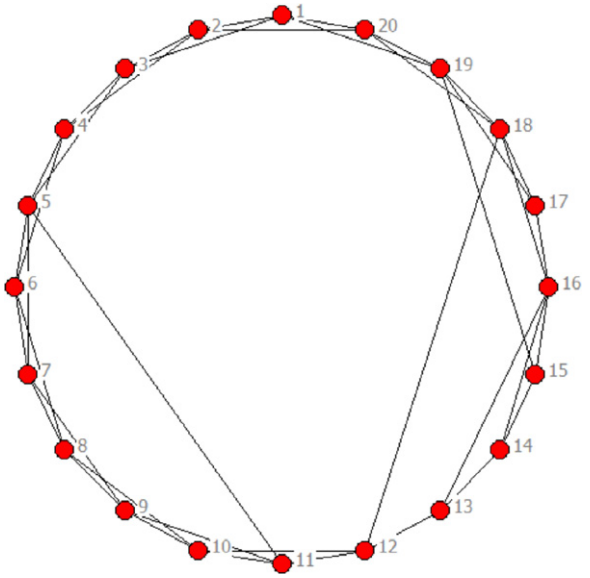

Fig. 5. (Color online) A 20-node network generated with WS small-world algorithm, where the rewiring probability $p = 0.1$.

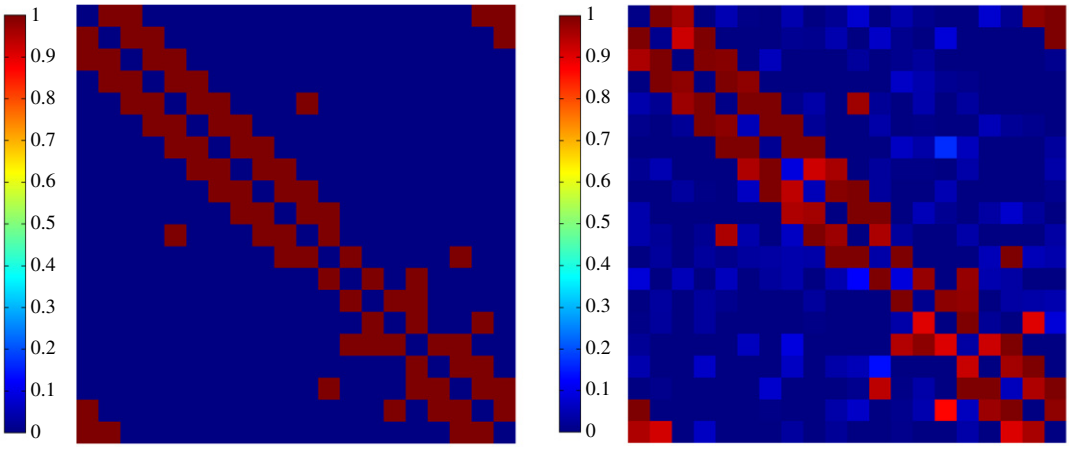

Fig. 6. (Color online) Left: The adjacency matrix of the small-world network as shown in Fig. 5; Right: The reconstructed adjacency matrix.

of the edge is kept but the other end is disconnected with probability p and reconnected to another node randomly chosen from the network. A WS small-world network generated with rewiring probability $p = 0.1$, as shown in Fig. 5, is used here for illustration. For simplicity, the weight for every existent edge is supposed to be 1.

The network structure can be completely described by giving the adjacency (or connectivity) matrix $\mathscr{A}$, a $N \times N$ square matrix whose entry $\mathscr{A}_{ij}(i, j = 1, \ldots, N)$ is equal to 1 when the edge from j to i exists, and zero otherwise. The diagonal elements of $\mathscr{A}$ are all zeros. The colormap in the left panel of Fig. 6 shows the adjacency matrix $\mathscr{A}$, where the red blocks represent existent edges ($\mathscr{A}_{ij} = 1$) and blue blocks for non-existent ones ($\mathscr{A}_{ij} = 0$). The estimated structure is displayed in the right panel. There are slight discrepancies between the two colormaps, which might be due to local optimal solutions caused by the large number of unknown parameters. However, it is seen that the existent edges are clearly separated from the non-existent ones. Therefore, the underlying structure of the 20-node small-world network can be correctly recovered within an allowable margin of relative error.

To demonstrate the applicability and robustness of the proposed method to real data, different levels of observation noise are considered and the corresponding ROC (receiver operating characteristic) curves for structure identification are provided in Fig. 7. For the noise free data, a 100% true positive rate (the fraction of correctly estimated edges out of all existent edges) and 100% true negative rate (the fraction of correctly estimated non-existent edges out of all non-existent edges, or 1-false positive rate) can be rendered. For data with different observation noise, the proposed method also yields satisfactory results. Here, the AUC (area under an ROC curve), which is an important measure of structure identification, is

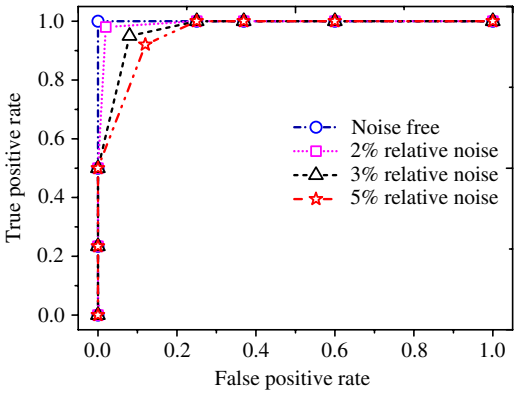

Fig. 7. (Color online) ROC curves for structure identification with different levels of observation noise.

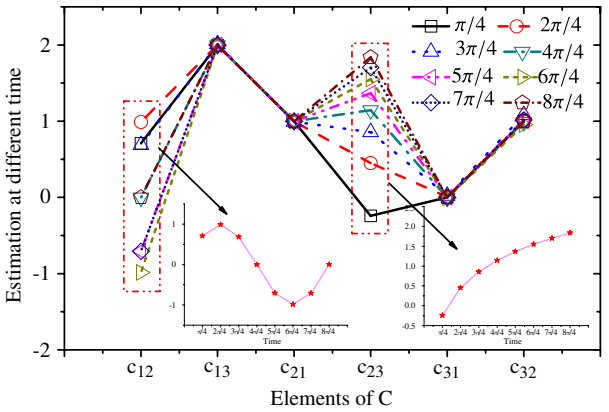

Fig. 8. (Color online) Structure identification of the time-varying dynamical network at different time points. The two insets show the estimation of C_{12} (left) and C_{23} (right) with respect to time.

calculated. If the area under an ROC curve is equal to 1.0, such as that for the noise free data as shown in Fig. 7, it means 100% accuracy of structure identification, that is, a 100% true positive rate and 100% true negative rate. The closer the AUC is to 1, the better the identification result. Particularly, the areas under the ROC curves for data with 2%, 3% and 5% relative observation noise are respectively 0.9925, 0.9737 and 0.96. Thus the actual structure of this 20-node network can be mostly recovered, even though the data contain high levels of observation noise.

4.5. Structure identification of a time-varying dynamical network

In reality, many dynamical networks usually have time-varying structures [31]. Therefore, an interesting topic is to identify structures for time-varying networks, which has also been known as a very challenging task. In this subsection, the proposed optimization-based technique is applied to a 3-node weighted and directed network with a time-varying structure, whose coupling matrix is given by

$$C(t) = \begin{pmatrix} -2-\sin t & \sin t & 2 \\ 1 & -1-\ln t & \ln t \\ 0 & 1 & -1 \end{pmatrix}. \tag{16}$$

To estimate the structure at time t, observed data are sampled from the interval $[t-0.02, t]$ with a sampling rate 0.001. That is to say, only 21 sets of data are adopted for identification. Fig. 8 displays the structure identification results at the time points from $\pi/4$ to 2π with a time step $\pi/4$. The two insets show respectively the estimation results of the time-varying couplings C_{12} and C_{23} with respect to time. The estimation error for each edge, which is defined as the average of the difference between estimated values and real values over the 8 different time points, is shown in the left panel of Fig. 9. It is obvious that the estimation errors reach around zero after 400 iterations and the network structure is correctly identified. The right panel of Fig. 9 displays the corresponding objective function $J_\alpha(C)$ at each iteration step. It is seen from the panel that the objective function quickly goes below 1×10^{-3} once the topological structure is correctly inferred. Hence, one can see that the proposed method is valid for structure identification of small-scale time-varying networks. Furthermore,

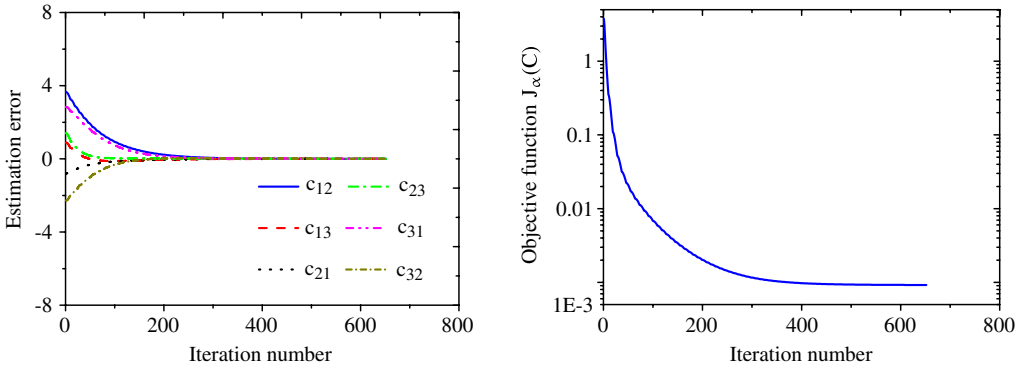

Fig. 9. (Color online) Left: Structure estimation error of the time-varying dynamical network; Right: The logarithmic value of the objective function $J_\alpha(C)$.

only a small number of sets of observed data are needed (≈20), which is quite practical for data collection and real-world applications.

5. Conclusion and outlook

It is of practical importance to identify the topological structure of a dynamical network. By transforming the structure identification problem into an optimization problem through rigorous mathematical analysis, an optimization algorithm based on the nonlinear conjugate gradient method has been proposed to recover the underlying topological structure of a dynamical network from observed nodal data. Five numerical examples have been presented to illustrate the validity and robustness of the new algorithm. Particularly, the algorithm works effectively when there is observation noise or when only some component variables of each multidimensional node are observable. In addition, the proposed algorithm has been further employed on a small-scale time-varying network and good identification results have been rendered.

The proposed algorithm is efficient for inferring the topological structures of small- and medium-scale networks. For a large-scale dynamical network, the adopted objective function tends to usually have a large number of local minima in the region of interest. Though the proposed algorithm is still effective to identify topological structures for large-scale networks by finding a global minimizer, it might be very computationally expensive by directly adopting this algorithm. Thus, an interesting topic for future work would be to probe into structure identification of large-scale dynamical networks. Furthermore, our result on a 3-node network with a time-varying structure demonstrates the applicability of the proposed method to small-scale time-varying networks. Therefore, another interesting aspect would be the application of optimization-based technique to identifying structures for medium- or large-scale time-varying dynamical networks, which has been known as a very challenging task. Our analysis and experiments have given some hints on how to do so, however, a more systematic approach is a part of our future work as well.

Acknowledgments

The authors would like to thank the anonymous reviewers for their valuable comments and suggestions. This work was supported in part by the National Natural Science Foundation of China (Grant Nos. 11101316, 11172215, 61174028), a research grant from the Australian Research Council, and the Doctoral Fund of Ministry of Education of China (Grant No. 200804861072).

References

[1] D. Watts, S. Strogatz, Collective dynamics of "small-world" networks, Nature 393 (6684) (1998) 440–442.
[2] A.L. Barabási, R. Albert, Emergence of scaling in random networks, Science 286 (5439) (1999) 509–512.
[3] S. Mei, X. Zhang, M. Cao, Power Grid Complexity, Springer (Jointly published by Tsinghua University Press), 2011.
[4] M.O. Jackson, A. Watts, The evolution of social and economic networks, J. Econom. Theory 106 (2) (2002) 265–295.
[5] H. de Jong, Modeling and simulation of genetic regulatory systems: a literature review, J. Comput. Biol. 9 (1) (2002) 67–103.
[6] H. Jeong, B. Tombor, R. Albert, Z.N. Oltvai, A.L. Barabasi, The large-scale organization of metabolic networks, Nature 407 (6804) (2000) 651–654.
[7] J. Zhou, J. Lu, J. Lü, Pinning adaptive synchronization of a general complex dynamical network, Automatica 44 (2008) 996–1003.
[8] M.E.J. Newman, The structure and function of complex networks, SIAM Rev. 45 (2) (2003) 167–256.
[9] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, D.U. Hwang, Complex networks: structure and dynamics, Phys. Rep. 424 (4) (2006) 175–308.
[10] D. Yu, M. Righero, L. Kocarev, Estimating topology of networks, Phys. Rev. Lett. 97 (18) (2006) 188701.
[11] J. Zhou, J. Lu, Topology identification of weighted complex dynamical networks, Physica A 386 (1) (2007) 481–491.
[12] X. Wu, Synchronization-based topology identification of weighted general complex dynamical networks with time-varying coupling delay, Physica A 387 (4) (2008) 997–1008.
[13] H. Liu, J. Lu, J. Lü, D.J. Hill, Structure identification of uncertain general complex dynamical networks with time delay, Automatica 45 (8) (2009) 1799–1807.
[14] J. Zhou, W. Yu, X. Li, M. Small, J. Lu, Identifying the topology of a coupled FitzHugh–Nagumo neurobiological network via a pinning mechanism, IEEE Trans. Neural Netw. 20 (10) (2009) 1679–1684.

[15] J. Zhao, L. Qin, J. Lu, Z.P. Jiang, Topology identification of complex dynamical networks, Chaos 20 (2) (2010) 023119.
[16] C. Li, W. Sun, J. Kurths, Synchronization between two coupled complex networks, Phys. Rev. E 76 (4) (2007) 046204.
[17] X. Wu, W.X. Zheng, J. Zhou, Generalized outer synchronization between complex dynamical networks, Chaos 19 (1) (2009) 013109.
[18] J. Ren, W. Wang, B. Li, Y.C. Lai, Noise bridges dynamical correlation and topology in coupled oscillator networks, Phys. Rev. Lett. 104 (5) (2010) 058701.
[19] C.W.J. Granger, Investigating causal relations by econometric models and cross-spectral methods, Econometrica 37 (3) (1969) 424–438.
[20] D. Marinazzo, M. Pellicoro, S. Stramaglia, Kernel method for nonlinear Granger causality, Phys. Rev. Lett. 100 (14) (2008) 144103.
[21] R. Jansen, H. Yu, D. Greenbaum, Y. Kluger, N.J. Krogan, S. Chung, A. Emili, M. Snyder, J.F. Greenblatt, M. Gerstein, A Bayesian networks approach for predicting protein-protein interactions from genomic data, Science 302 (5466) (2003) 449–453.
[22] N. Marwan, M.C. Romano, M. Thiel, J. Kurths, Recurrence plots for the analysis of complex systems, Phys. Rep. 438 (5) (2007) 237–329.
[23] M.C. Romano, M. Thiel, J. Kurths, C. Grebogi, Estimation of the direction of the coupling by conditional probabilities of recurrence, Phys. Rev. E 76 (3) (2007) 036211.
[24] J. Nawrath, M.C. Romano, M. Thiel, I.Z. Kiss, M. Wickramasinghe, J. Timmer, J. Kurths, B. Schelter, Distinguishing direct from indirect interactions in oscillatory networks with multiple time scales, Phys. Rev. Lett. 104 (3) (2010) 038701.
[25] S. Hempel, A. Koseska, J. Kurths, Z. Nikoloski, Inner composition alignment for inferring directed networks from short time series, Phys. Rev. Lett. 107 (5) (2011) 054101.
[26] H.T. Banks, K. Kunisch, Estimation Techniques for Distributed Parameter Systems, Birkhäuser, Boston, 1989.
[27] H.W. Engl, M. Hanke, A. Neubauer, Regularization of Inverse Problems, Kluwer, Dordrecht, 1996.
[28] C. Vogel, Computational Methods for Inverse Problems, SIAM, Philadelphia, 2002.
[29] L. Chen, J. Lu, C.K. Tse, Synchronization: an obstacle to identification of network topology, IEEE Trans. Circuits Syst. II 56 (4) (2009) 310–314.
[30] N. Tikhonov, V.Y. Arsenin, Solutions of Ill-Posed Problems, John Wiley & Sons, New York, 1977.
[31] J. Lü, G. Chen, A time-varying complex dynamical network model and its controlled synchronization criteria, IEEE Trans. Automat. Control 50 (6) (2005) 841–846.