Reconstructing the topology of sparsely connected dynamical networks

Domenico Napoletani[1,2] and Timothy D. Sauer[1,*]
[1]*Department of Mathematical Sciences, George Mason University, Fairfax, Virginia 22030, USA*
[2]*Center for Applied Proteomics and Molecular Medicine, George Mason University, Manassas, Virginia 20110, USA*
(Received 14 September 2007; revised manuscript received 7 December 2007; published 8 February 2008)

Given a general physical network and measurements of node dynamics, methods are proposed for reconstructing the network topology. We focus on networks whose connections are sparse and where data are limited. Under these conditions, common in many biological networks, constrained optimization techniques based on the $L1$ vector norm are found to be superior for inference of the network connections.

DOI: 10.1103/PhysRevE.77.026103 PACS number(s): 89.75.Hc, 05.45.Tp, 05.45.Jn, 05.45.Xt

Network dynamics have recently become ubiquitous as models of physical and biological behavior [1]. Examples range from Josephson junction arrays [2], power grids [3], networks of neurons [4], models of infectious disease [5], and social networks [6] to a plethora of current examples in systems biology like genetic networks [7], protein interaction nets [8], and metabolic networks [9].

Because of the strong emphasis on network models, a great need exists for methods of inferring network structure from data. Topological information has influence over the dynamical behavior [10] and, in particular, has been used to great effect in the study of network synchronization [11–14].

We will assume that time series measurements are available at some or all of the network nodes. The first step toward understanding the system dynamics is to decipher the network topology by detecting the links between nodes. Topological information for an N-node network is contained in the *adjacency matrix*, the $N \times N$ matrix A whose (i,j) entry is 1 if there is a link from node i to node j and 0 otherwise. Inferring the network topology from data is in general an ill-posed problem. Factors that increase the difficulty of this task are nonlinearity, large noise, and lack of data. Each of these factors is notably present in biophysics and system biological applications.

Methods for detecting links based on observed data have been proposed in the recent literature. A method based on chaotic synchronization was proposed and demonstrated through examples in [15]. In [16], a method based on perturbing the dynamics was proposed and applied to a phase oscillator network. Neither of these methods was evaluated in the presence of noisy signals. In our opinion, noise will be pervasive in any proposed application of these methods and the handling of noise will be an important factor in a method that successfully disentangles the network structure from observed data.

In this article we introduce a method for reconstructing dynamic network topology in the presence of significant noise and low data availability, under the assumption of sparse connectivity. By sparse we mean around 10% of possible connections between nodes, although the method will often succeed with lesser accuracy for higher connectivity rates. Figure 1 shows a typical 10%-connected sparse network of discrete-time dynamics that undergoes sustained chaotic behavior.

We apply techniques that have proved useful for time series analysis of chaotic systems. Takens [17] showed that the phase space of autonomous systems can be reconstructed from time series measurements, using what has come to be known as the delay-coordinate method (see also [18,19]). It has been widely exploited for noise reduction [20] and time series prediction [21] of chaotic time series. In the present context, we assume a network of autonomous systems to which a sparse set of connections has been added. Further, we assume that a generic observable of each system is available, creating a multivariate time series. Our work on the present problem is complementary to previous work on more complete reconstruction problems in small networks with high data coverage [22]; in fact, the use of delay coordinates is not a requirement for the method presented in this article.

To deal with the nonlinearity of the system dynamics, we will work locally in phase space, using linearization of the dynamics around the center of a neighborhood of data points. Small neighborhoods in high-dimensional spaces contain relatively few points. Therefore, for low data rates it is essential to replace the least-squares methodology of those methods with recent ideas from nonlinear signal processing involving minimization with respect to the $L1$ norm. We will explain why this methodology is appropriate for sparse net-

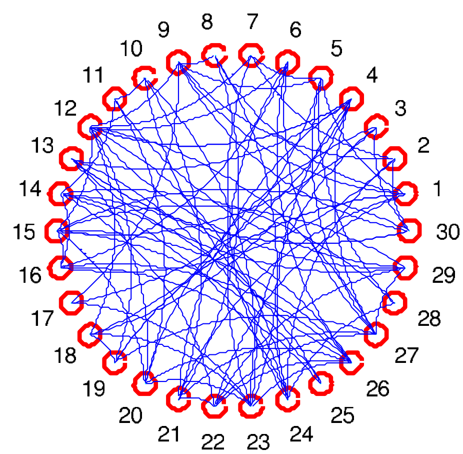

FIG. 1. (Color online) Connection diagram of system. The system is sparsely connected; each node has three incoming connections from other nodes.

*tsauer@gmu.edu

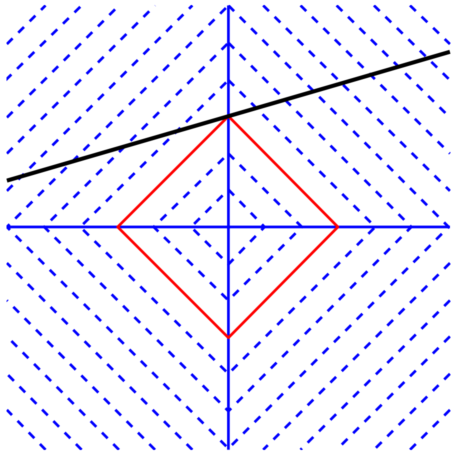

FIG. 2. (Color online) Isoclines of the $L1$ norm. The minimum-norm point on a linear space tends to have relatively few nonzero coordinates, making $L1$ minimization effective for computing sparse solutions.

work inference and demonstrate its superiority to the least-squares approach.

We begin by investigating the properties of the $L1$ minimization in the simplest possible case. Let X be an $m \times n$ matrix, c an n vector, and $b = Xc$. Add observation noise to X and b, and consider the problem of estimating the value of c from the noisy $\tilde{X}$ and $\tilde{b}$. To correspond to the stated context, we will assume that (i) c is a sparse vector—i.e., most of its entries are zero—and (ii) the number m, corresponding to the number of observations, is small. We do not assume known which entries of c are nonzero.

The method of $L2$ minimization, using the ordinary least-squares (OLS) or total least-squares (TLS) methodology, is the classical answer to this problem. If $X = \tilde{X}$ and the noise is only in b, then the maximum likelihood unbiased estimate is $\bar{c} = X^{\dagger} b$, where $X^{\dagger}$ is the pseudoinverse [23] of the rectangular matrix X. (The computation of $\bar{c}$ is best done indirectly using techniques of orthogonal matrices that avoid the explicit construction of the pseudoinverse.) If X is noisy as well, the so-called "errors-in-variables" scenario, then the TLS method is preferred to the OLS method in theory and finds a more accurate solution in the limit of large m. However, in practice, and in particular in the case of limited data, the OLS method often outperforms the TLS method [24,25] even in situations involving errors in variables.

Next we show that there are even better alternatives if c is a sparse vector. First, we increase the number of variables in the system by adding columns of random numbers to X, enough to make the system underdetermined ($m < n$). The augmented system $X'c' = b$ now has infinitely many solutions composing a linear hyperplane in n-dimensional space. For the solution $\bar{c}$, choose the point on this set of minimum $L1$ norm. The $L1$ norm of a vector is the sum of the absolute values of its coordinates. Figure 2 shows that in the plane, the point on a one-dimensional linear space of minimum $L1$ norm generally lies on a coordinate axis—i.e., has at least one coordinate equal to zero. More generally, on a linear space defined by a general m equations in n-dimensional space, the point of the minimum $L1$ norm must have at least

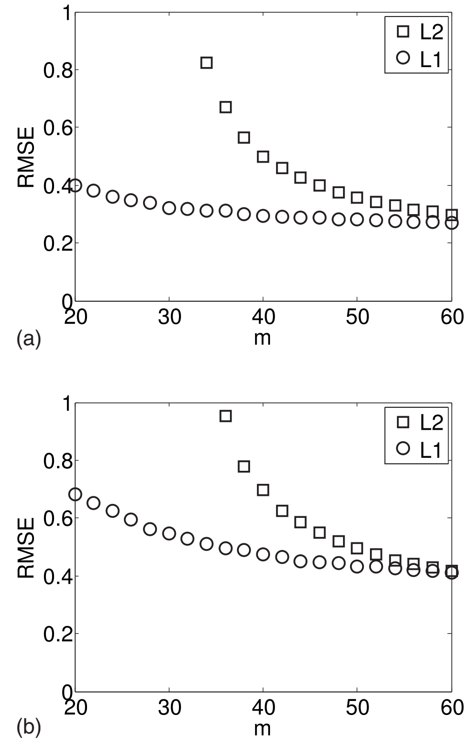

FIG. 3. Error in reconstructing coefficients versus the number of data points: (a) 10% sparsity and (b) 20% sparsity.

m zero coordinates. This is the reason that the method tends to favor sparse vectors as solutions. Once the solution c' of the augmented system is found, the extra coordinates of c' are ignored to give the solution c.

Calculation of a minimum $L1$ norm point can be done using general-purpose optimization software as follows. Given the system of equations $X'c' = b$, consider the alternate system

$$[X' \ -X'] \begin{bmatrix} c_+ \\ c_- \end{bmatrix} = b. \quad (1)$$

Here c_+ (c_-) is playing the role of the positive (negative) part of c', so that the sum of the coordinates of c_+ and c_- is equal to the sum of the absolute values of the coordinates of c, the $L1$ norm. Solve the classical optimization problem of minimizing the sum of the coordinates of the vector $[c_+^T c_-^T]$ with respect to equality constraints (1) and inequality constraints $c_+ \geq 0, c_- \geq 0$. Then the minimum $L1$ norm point is $\bar{c} = c_+ - c_-$. Several algorithms exist for constrained optimization; in the following calculations, we use LIPSOL [28], a primal-dual interior point method in general distribution.

Figure 3 shows a comparison of the error in reconstructing the original vector c when using the classical $L2$ least-squares (OLS) method and the proposed $L1$ minimization method. After $b = Xc$ is calculated, 50% Gaussian noise is added to both X and b. Then 120 random vector columns are added to augment the matrix. These entries are Gaussian random numbers of standard deviation 0.5. Then LIPSOL is used as above to solve for c. The root-mean-squared error (RMSE) is averaged over 1000 realizations of random m

$\times 30$ matrices X of normal random numbers of unit standard deviation. (The standard error bars are within the size of the symbol shown.) The results in Fig. 3 show that for sparse c and limited data, the $L1$ method is far superior to the OLS method. The results using the TLS method are markedly inferior for these relatively small values of m and are not shown. As the number of nonzero entries in c grows and as the amount of available data grows, the $L2$ method becomes competitive and will eventually dominate.

The alternative method depending on $L1$ minimization was motivated by recent work on sparse representation of signals using overcomplete systems [26,27]. In their work, a large basis of possible signal terms is used; in our application, the matrix X is given to us from the time series data, and we explicitly add extra random vectors to the basis to absorb noise and model error in the data.

Now we turn to the application of these ideas to the chaotic network of Fig. 1. Assume that time series can be recorded from the network nodes. We will work in the proxy phase space of the network, reconstructed from time series. Choose one data point x_*^t from the phase space at time t and form the neighborhood of N nearest data points from the collected multivariate time series. For the ith coordinate, let the dynamics at time step t be expressed in a Taylor series expansion around x_*^t as

$$x_i^{t+1} = F_i(x_*^t) + DF_i(x_*^t)(x_i^t - x_*^t) + O(\|x_i^t - x_*^t\|^2). \qquad (2)$$

The higher-order terms are of size $\|x_i^t - x_*^t\|^2$ and therefore will depend sensitively on the size of the neighborhood. The smaller the neighborhood, the smaller the higher-order terms and the more accurate the linear terms will be in approximating the dynamics F_i, but the fewer points will be available for fitting the linear terms accurately. This is the reason that the fitting method must be as efficient as possible and why the $L1$ method described above is essential in low-data situations. We begin by estimating the constant term and linear coefficients in (2) by the linear system $Xc=b$, where c represents the unknown coefficients and the number of equations (rows of the matrix X) is the number of data points in the neighborhood. Then we apply the method of $L1$ minimization to solutions of Eq. (1) after augmenting the matrix with random columns, as described above.

The computation of the ith column of the adjacency matrix A is done by thresholding the estimated linear coefficients for F_i in the local neighborhoods. An estimated linear coefficient whose absolute value exceeds the threshold in more than 25% of the neighborhoods is accepted as a link in the network. The threshold can be chosen using any of a number of heuristic arguments. In the studies below the threshold was set to make the false positive rate approximately 5% for purposes of comparison of the alternative approaches.

In Fig. 4 we display results for the discrete dynamical system

$$x_i^{t+1} = c_i + \sum_{j=1}^{N} a_{ij} \cos b_{ij} x_j^t \qquad (3)$$

for $i=1,\ldots,N$. For this test, the number of nodes is $N=30$ and 10% of the adjacency matrix entries a_{ij} were set to 1; the remainder were set to 0. The coefficients c_i were chosen randomly from the standard normal distribution $N(0,1)$ and fixed for the simulation run. The b_{ij} were chosen similarly from $N(0.88, 0.4)$, to avoid nongeneric results caused by symmetries in the dynamics. Under these conditions, the network develops a chaotic attractor whose Lyapunov dimension [20] is approximately 8.2. The initial conditions for the trajectory that provided data were also chosen randomly. On top of the chaotic dynamics, observational noise of σ times the standard deviation of the time series was added, where σ ranged from 0 to 0.25, as shown in the figure. The method was applied to noisy trajectories of length 200 and 400.

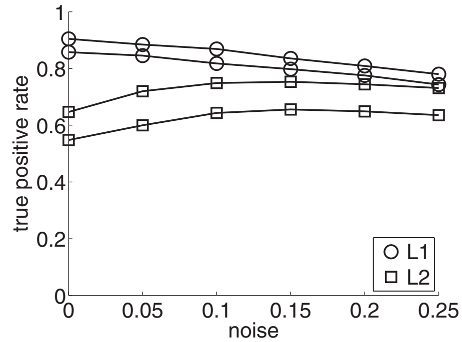

FIG. 4. True positive rate for example system (3). The upper pair of traces use single trajectories of length 200 (400) points and apply the $L1$ optimization technique from the text. The mean TPR over 40 trajectories is shown. The lower pair of traces are the same, but using the ordinary least-squares approach.

The true positive rate (TPR) is defined to be the number of correct detected links divided by the total links in the network. The false positive rate (FPR) is the number of incorrectly detected links divided by the total number of nonlinks. Figure 4 compares the TPR for varying amounts of time series data, while the FPR is held constant near 5% by choice of threshold. The data points of this simulation are averages of 40 simulations. In each simulation, 40 local neighborhoods were used to group points and carry out the linearizations.

Note that the TPR declines as the noise increases and the availability of data declines. Figure 4 shows the relative ineffectiveness of the ordinary least-squares method for this purpose, as predicted by the above discussion and Fig. 3. Again, the results of link detection using the total least-squares method were inferior to the results for the ordinary least-squares method.

As the data availability increases, the effectiveness of the $L1$ optimization method and the ordinary least-squares method grow similar, and true positive rates of both approach 100% for sufficient data.

We note that although the method is illustrated with a single trajectory from a chaotic, ergodic system, the fact that the dynamics is chaotic is not a requirement for the method to be successful. In fact, several shorter trajectories can be used. The method will be more accurate to the degree that data acquired are representative of the phase space of the system whether the system is ergodic or not.

Figure 5 shows corresponding results for an example with higher-dimensional dynamics. Each node consists of the Hénon-like [29] two-dimensional system

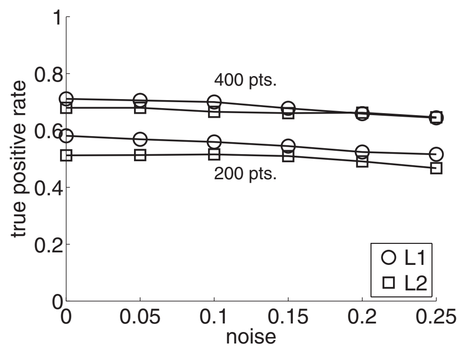

FIG. 5. True positive rate for example system (4). Settings are similar to those in Fig. 4. The circles denote the use of the $L1$ technique described in the text. The squares denote the use of the ordinary least-squares approach.

$$x_i^{t+1} = \cos(1.2x_i^t + 1.35y_i^t + \sum_{j=1}^{N} a_{ij}b_{ij} \sin(x_j^t + y_j^t)),$$

$$y_i^{t+1} = c_i(x_i^t)^2 + d_i y_i^t, \tag{4}$$

where the b_{ij} are chosen randomly (but fixed for each trajectory) from the normal distribution $N(0.18, 0.1)$ and c_i and d_i are chosen from $N(0.75, 0.03)$ and $N(-0.65, 0.03)$, respectively. For purposes of comparison, the same adjacency matrix A, represented in Fig. 1, was used in the network. There is again a chaotic attractor present in the 60-dimensional phase space, of much higher dimension than system (3). Noise was added to the signal x measured from each of the nodes as in Fig. 4. The trajectories of a few hundred data points are clearly inadequate for complete resolution of the adjacency matrix in this example. Even in this case, however, the best results are given by the $L1$ method.

The application of $L1$ techniques to sparse signal processing problems is under intense development due to the pioneering work of [26,27]. We have customized the basic idea to the context of network dynamics and found that the use of extra random vectors enhances the resolution of the adjacency matrix. It is interesting to speculate on the effect of these random vectors, which are added to the matrix prior to the $L1$ optimization. In addition to reducing the condition number of the calculation, they evidently take up the noise and distribute it in small quantities among the random vectors.

The method we have outlined for detecting links in a sparsely connected chaotic dynamical network is surprisingly robust in the case of small data sets. Asymptotically, as the data coverage increases, the difference between the $L1$ optimization and $L2$ regression disappears. We expect the method to have great value for large networks, in particular metabolic and genomic pathways, where data are expensive and difficult to gather.

This research was supported by the National Science Foundation.

[1] R. Albert and A.-L. Barabasi, Rev. Mod. Phys. **74**, 47 (2002); S. N. Dorogovtsev and J. F. F. Mendes, Adv. Phys. **51**, 1079 (2002); M. E. J. Newman, SIAM Rev. **45**, 167 (2003); S. H. Strogatz, Nature (London) **410**, 268 (2001).
[2] P. Hadley, M. R. Beasley, and K. Wiesenfeld, Phys. Rev. B **38**, 8712 (1988); K. Wiesenfeld, Physica B **222**, 315 (1996).
[3] A. G. Phadke and J. S. Thorp, *Computer Relaying for Power Systems* (Wiley, New York, 1988).
[4] L. F. Abbott and C. van Vreeswijk, Phys. Rev. E **48**, 1483 (1993); J. J. Hopfield and A. V. M. Herz, Proc. Natl. Acad. Sci. U.S.A. **92**, 6655 (1995).
[5] M. Kretzchmar and M. Morris, Math. Biosci. **133**, 165 (1996).
[6] D. J. Watts and S. H. Strogatz, Nature (London) **393**, 440 (1998).
[7] T. S. Gardner, S. Shimer, and J. J. Collins, ASM News **70**, 121 (2004); T. S. Gardner, D. Bernardo, D. Lorenz, and J. J. Collins, Science **301**, 102 (2003).
[8] C. L. Tucker, J. F. Gera, and P. Uetz, Trends Cell Biol. **11**, 102 (2001).
[9] H. Jeong, B. Tombor, R. Albert, Z. N. Oltvai, and A.-L. Barabasi, Nature (London) **407**, 651 (2000); H. Jeong, S. P. Mason, A.-L. Barabasi, and Z. N. Oltvai, *ibid.* **411**, 41 (2001).
[10] M. E. J. Newman, Phys. Rev. Lett. **89**, 208701 (2002).
[11] L. M. Pecora and T. L. Carroll, Phys. Rev. Lett. **80**, 2109 (1998).
[12] M. Barahona and L. M. Pecora, Phys. Rev. Lett. **89**, 054101 (2002).
[13] T. Nishikawa, A. E. Motter, Y. C. Lai, and F. C. Hoppensteadt, Phys. Rev. Lett. **91**, 014101 (2003).
[14] L. Donetti, P. I. Hurtado, and M. A. Munoz, Phys. Rev. Lett. **95**, 188701 (2005).
[15] D. Yu, M. Righero, and L. Kocarev, Phys. Rev. Lett. **97**, 188701 (2006).
[16] M. Timme, Phys. Rev. Lett. **98**, 224101 (2007).
[17] F. Takens, in *Dynamical Systems and Turbulence*, edited by D. A. Rand and L.-S. Young, Lecture Notes in Mathematics, Vol. 898 (Springer-Verlag, New York, 1981), p. 366.
[18] T. Sauer, M. Casdagli, and J. A. Yorke, J. Stat. Phys. **65**, 579 (1991).
[19] L. Pecora, L. Moniz, J. Nichols, and T. Carroll, Chaos **17**, 013110 (2007).
[20] E. Ott, T. Sauer, and J. Yorke, *Coping with Chaos: Analysis of Chaotic Data and the Exploitation of Chaotic Systems* (Wiley Interscience, New York, 1994); H. Kantz and T. Schreiber, *Nonlinear Time Series Analysis* (Cambridge University Press, Cambridge, England, 1997).
[21] J. D. Farmer and J. J. Sidorowich, Phys. Rev. Lett. **59**, 845 (1987); A. Weigend and N. Gershenfeld, *Time Series Prediction* (Addison-Wesley, Reading, MA, 1993).
[22] T. D. Sauer, Phys. Rev. Lett. **93**, 198701 (2004).
[23] T. Sauer, *Numerical Analysis* (Addison Wesley, Boston, MA, 2006).

[24] H. Voss, J. Timmer, and J. Kurths, Int. J. Bifurcation Chaos Appl. Sci. Eng. **14**, 1905 (2004).
[25] E. J. Kostelich, Phys. Rev. E **64**, 016213 (2001).
[26] S. S. Chen, D. L. Donoho, and M. A. Saunders, SIAM J. Sci. Comput. **20**, 33 (1998); D. L. Donoho, Commun. Pure Appl. Math. **59**, 797 (2006).
[27] R. Candes, J. Romberg, and T. Tao, Commun. Pure Appl. Math. **59**, 1207 (2006).
[28] Y. Zhang (unpublished).
[29] M. Hénon, Commun. Math. Phys. **50**, 69 (1976).