J. Math. Biol. (2010) 60:285–312
DOI 10.1007/s00285-009-0263-x

Mathematical Biology

Identification of feedback loops embedded in cellular circuits by investigating non-causal impulse response components

Chao-Yi Dong · Tae-Woong Yoon · Declan G. Bates · Kwang-Hyun Cho

Received: 25 November 2008 / Revised: 3 March 2009 / Published online: 31 March 2009

Abstract Feedback circuits are crucial dynamic motifs which occur in many biomolecular regulatory networks. They play a pivotal role in the regulation and control of many important cellular processes such as gene transcription, signal transduction, and metabolism. In this study, we develop a novel computationally efficient method to identify feedback loops embedded in intracellular networks, which uses only time-series experimental data and requires no knowledge of the network structure. In the proposed approach, a non-parametric system identification technique, as well as a spectral factor analysis, is applied to derive a graphical criterion based on non-causal components of the system's impulse response. The appearance of non-causal components in the impulse response sequences arising from stochastic output perturbations is shown to imply the presence of underlying feedback connections

Electronic supplementary material The online version of this article (doi:10.1007/s00285-009-0263-x) contains supplementary material, which is available to authorized users.

C.-Y. Dong · T.-W. Yoon (✉)
School of Electrical Engineering, Korea University, Seoul 136-713, Korea
e-mail: twy@korea.ac.kr

C.-Y. Dong
School of Information and Engineering, Inner Mongolia University of Technology,
010051 Huhhot, China

D. G. Bates
Systems Biology Lab, Department of Engineering, University of Leicester,
Leicester LE1 7RH, UK

C.-Y. Dong · K.-H. Cho (✉)
Department of Bio and Brain Engineering,
Korea Advanced Institute of Science and Technology (KAIST),
335 Gwahangno, Yuseong-gu, 305-701 Daejeon, Republic of Korea
e-mail: ckh@kaist.ac.kr
URL: http://sbie.kaist.ac.kr

within a linear network. In order to extend the approach to nonlinear networks, we linearize the intracellular networks about an equilibrium point, and then choose the magnitude of the output perturbations sufficiently small so that the resulting time-series responses remain close to the chosen equilibrium point. In this way, the impulse response sequences of the linearized system can be used to determine the presence or absence of feedback loops in the corresponding nonlinear network. The proposed method utilizes the time profile data from intracellular perturbation experiments and only requires the perturbability of output nodes. Most importantly, the method does not require any a priori knowledge of the system structure. For these reasons, the proposed approach is very well suited to identifying feedback loops in large-scale biomolecular networks. The effectiveness of the proposed method is illustrated via two examples: a synthetic network model with a negative feedback loop and a nonlinear caspase function model of apoptosis with a positive feedback loop.

Keywords Biomolecular regulatory networks · Feedback loops · Nonparametric identification · Spectral factor analysis · Signaling pathways · Systems biology

Mathematics Subject Classification (2000) 92-08 · 92B05 · 93B30 · 93E12 · 93E10 · 93E24

1 Introduction

Various complex control and regulation mechanisms appear within or between reacting biomolecular species that are coordinated in the context of cell functions such as growth, differentiation, and apoptosis. As a result, two of the central themes in current systems biology research are the identification of regulatory network motifs, and the development of a quantitative understanding of regulation, control and coordination in intracellular networks (Wolkenhauer et al. 2003; Cho et al. 2003). Feedback circuits are key regulatory motifs in many biological systems. It is believed that positive feedback loops determine diverse cellular processes including development, cell proliferation, apoptosis, and the response to stress (Eisen et al. 1967; Wolpert and Lewis 1975; Kim et al. 2007a,b), whereas negative feedback loops contribute to maintaining homeostasis of biological systems under internal and external changes (Laub and Loomis 1998; Strogatz 2000a; Thomas and Kaufman 2001; Maeda et al. 2004). Moreover, many complicated functional modules in biological systems are composed of multiple feedback loops. For instance, many types of chaotic behavior in biological systems are generated by combinations of positive and negative feedback loops (Tufillaro et al. 1992; Glendinning 1994; Strogatz 2000b; Kwon and Cho 2008a,b; Kim et al. 2008; Shin et al. 2009), while interlinked fast and slow positive feedback loops (Brandman et al. 2005) have been shown to result in a "dual-time" switch that is rapidly inducible as well as resistant to noise in the upstream signaling system. Further connections between feedback circuits and particular types of system behaviour were made by Thomas (1981) who proposed an intriguing conjecture stating that a positive feedback loop is necessary for multistationarity (i.e., multiple equilibria) and a negative feedback loop is necessary for a stable periodic behavior. Gouze (1998) later proved this conjecture through a strict mathematical deduction.

As feedback mechanisms are so pivotal in various cellular functions, the problem of identifying the existence (or absence) of an intracellular feedback circuit from experimental data has recently been the subject of intensive research. The likelihood ratio test method (LRTM) proposed by Caines and Chan (1975) aimed at identifying the existence of a feedback loop, but it required very detailed a priori knowledge on the system order and structure. The acquisition of such a priori knowledge is difficult in practice for biological systems, particularly for large-scale regulatory networks. Schnider et al. (1989) developed a direct coherence method to detect feedbacks in the central nervous system in the frequency domain. Subsequent studies developed and extended this method to other physiological applications: for example, the detection of bi-directional hippocampal interactions between the CA3 and CA1 region (Baccala and Sameshima 1999) and the quantification of the linear causal strength in closed interacting cardiovascular variability signals (Porta et al. 2002). The direct coherence approach used in the above studies requires the determination of a multi-dimensional parametric structure for the system before the computation of the frequency coherence. Thus, it usually has a high computational complexity. Vance et al. (2002) investigated the effect of a pulse change in the concentration of chemical species in a reaction network. A graphical characteristic of temporal responses was used to identify control pathways, including feedback regulations and feedforward regulations, from other generic interactions. However, this graphical characteristic could not distinguish feedback regulations from feedforward regulations. Recently, a novel method was proposed (Dong et al. 2008) by applying intermittent step perturbations to excite the biomolecular species and using a prominent circular causal property to infer the existence of a feedback loop. This approach was applicable to nonlinear networks without a priori knowledge of the system structure, but included a requirement on the perturbability of all reacting species that makes it very difficult to apply to real biomolecular regulatory networks.

In this study, we apply correlation identification and spectral factor analysis to the intracellular feedback loop identification problem. The main idea is to identify any underlying feedback loops by investigating non-causal components of the system's impulse response sequences caused by output perturbations. According to Granger's definition of causality (Granger 1962), an impulse response is said to be causal if its negative time parts are equal to zero. This means that the present output is only determined by the present and past input values. If this is not the case, the impulse response is deemed to be non-causal. The negative time parts of non-causal impulse response are called its non-causal components, whereas the zero and positive time parts are termed its causal components. Let us consider a single input and single output (SISO) linear system in which the input and output are both stationary stochastic processes. From correlation identification, the impulse response of the system can be identified using the input and output time-series data. In the case of no feedback loop between the input and output, the identified impulse response sequences are causal. On the contrary, if a feedback loop exists between the input and output, some non-causal components will appear in the identified impulse response sequences. This prominent characteristic provides a straightforward criterion with which to identify the existence of a feedback loop. In order to further extend this criterion to nonlinear systems, we ensure that the magnitude of the output stochastic perturbation is sufficiently small so

that the resulting time-responses stay within small neighborhoods of their equilibria. By then applying the feedback identification procedure to the linearized system, we can extend our graphical criterion to nonlinear intracellular networks. Importantly, the proposed method does not require any explicit knowledge about the network order or structure, and it is also applicable to those cases where not all of the network's nodes are perturbable. Hence, it overcomes almost all of the significant limitations associated with the previous approaches cited above.

The organization of the paper is as follows. We start by describing the correlation algorithm for identifying the non-causal impulse response of a SISO linear time-invariant system (Sect. 2). Spectral factor analysis is applied in Sect. 3 to derive the non-causal components criterion for identifying the existence of a feedback loop. In Sects. 4–6 this criterion is described as an identification algorithm which can be applied to intracellular perturbation experiments. Factors affecting the precision of the identification algorithm are detailed in Sect. 7. Finally, the application of the proposed method is illustrated using a synthetic negative feedback model and a caspase function model in apoptosis with a positive feedback loop in Sects. 8 and 9, respectively. The results show the efficacy of the proposed method for feedback loop identification under nonlinear reacting dynamics.

2 Correlation identification method

Correlation identification is an important nonparametric identification method, which has proved very useful in identifying the impulse response sequences of a system in the presence of noise. For this reason, it is widely used in a diverse range of applications, including stochastic processes in which the probability density distributions of random signals evolve over time—see, for example, its application to power converter identification (Miao et al. 2005) and X-ray CT imaging system analysis (Doré et al. 1997), etc. In this study, we apply the correlation identification method for the identification of non-causal impulse response sequences. Some preliminary definitions and nomenclatures on stochastic processes are provided in the Supplementary Information.

Let us consider a SISO linear time-invariant system described as follows:

$$y(k) = \sum_{\sigma=0}^{\infty} u(k-\sigma)\, h(\sigma) + v(k)\,, \tag{1}$$

where k and σ denote the discrete time indices, $y(k)$ is the output, $u(k)$ is the input, and $v(k)$ is the output noise, which is uncorrelated with $u(k)$.

The purpose of identification is to determine the impulse response sequence $h(k)$ from two data measurements: the deterministic input $u(k)$ and the noise-mixed output $y(k)$. This can be done by computing the cross-covariance function between $u(k)$ and $y(k)$ as follows:

$$R_{uy}(\tau) = \sum_{k=0}^{\infty} h(k) R_{uu}(\tau - k) + R_{uv}(\tau)\,. \tag{2}$$

In Eq. (2), $R_{uu}(\tau)$ is the auto-covariance function of $u(k)$ and $R_{uv}(\tau)$ is the cross-covariance function of $u(k)$ and $v(k)$. If the system is asymptotically stable, $h(k)$ converges to zero as k goes to infinity. We also note that $R_{uv}(\tau)$ approaches zero since $v(k)$ is uncorrelated with $u(k)$. Hence, for some finite number of discrete time steps M, we have

$$R_{uy}(\tau) \approx \sum_{k=0}^{M} h(k) R_{uu}(\tau - k). \tag{3}$$

Equation (3) is known as a discrete *Wiener-Hopf* equation (Godfrey 1980), and it suggests a computationally feasible way of computing the causal impulse response sequences $h(k)$ from the covariance functions. Note that the covariance functions in Eq. (3) can be estimated accurately by using a large data length $L >> M_1 + M_2 + 1$ as follows:

$$R_{uu}(\tau) \cong \frac{1}{L} \sum_{k=1}^{L} u(k) u(k+\tau), \tag{4}$$

$$R_{uy}(\tau) \cong \frac{1}{L} \sum_{k=1}^{L} u(k) y(k+\tau). \tag{5}$$

The fast Fourier transform (FFT) technique can be applied to compute the estimates of the covariance functions in a computationally efficient way (Fang and Xiao 1988). For a non-causal system,

$$R_{uy}(\tau) \approx \sum_{k=-M_1}^{M_2} g(k) R_{uu}(\tau - k), \tag{6}$$

where $g(k)$ represents a non-causal impulse response sequence. As we focus here on the non-causal components induced by output noises, we will use Eq. (6) to compute the non-causal impulse response sequences. In particular, a multivariable linear regression algorithm (Gauss 1963; Rao 1973; Rabiner et al. 1978; Draper and Smith 1981; Hunter and Kearney 1983; Dayal and MacGregor 1996; Westwick and Kearney 1997) is employed to solve $g(k)$. Here we use the method of Hunter and Kearney (1983) to identify the two-sided linear filter $g(k)$. According to this algorithm, the least-squares estimate of $g(k)$, denoted by $\hat{g}(k)$, can be calculated from the following vector representation:

$$\mathbf{C_{uy}} = \mathbf{C_{uu}} \hat{\mathbf{g}}_{\mathbf{LS}}, \tag{7}$$

where $\mathbf{C_{uy}} = \left[R_{uy}(-M_1), R_{uy}(-M_1+1), \ldots, R_{uy}(0), \ldots, R_{uy}(M_2)\right]'$ is an $M_1 + M_2 + 1$-dimensional column vector; $\hat{\mathbf{g}}_{\mathbf{LS}} = [\hat{g}(-M_1), \hat{g}(-M_1+1), \ldots, \hat{g}(0), \ldots, \hat{g}(M_2-1), \hat{g}(M_2)]'$ is a $M_1 + M_2 + 1$-dimensional column vector; and

$\mathbf{C_{uu}} = \begin{bmatrix} R_{uu}(0) & R_{uu}(-1) & \ldots R_{uu}(-M_1 - M_2) \\ R_{uu}(1) & R_{uu}(0) & \ldots R_{uu}(-M_1 - M_2 + 1) \\ \ldots & \ldots & \ldots\ldots \\ R_{uu}(M_2 + M_1) & R_{uu}(M_2 + M_1 - 1) & \ldots R_{uu}(0) \end{bmatrix}$ is a $M_1 + M_2 + 1$-dimensional symmetric square matrix, since $R_{uu}(\tau) = R_{uu}(-\tau)$. Note $\mathbf{C_{uu}}$ is a Toeplitz matrix since $\mathbf{C_{uu}}(i, j) = \mathbf{C_{uu}}(i-1, j-1)$, and hence Levinson's algorithm can also be applied to the symmetric Toeplitz system to calculate $\hat{\mathbf{g}}_{\mathbf{LS}}$ recursively (Levinson 1947; Durbin 1960; Golub and van Loan 1996). Usually, when $M_1 + M_2 + 1$ is small, it is unnecessary to perform a very provisional computation as Levinson's algorithm does and we can directly compute $\hat{\mathbf{g}}_{\mathbf{LS}}$ by finding the inverse of $\mathbf{C_{uu}}$ (Hunter and Kearney 1983):

$$\hat{\mathbf{g}}_{\mathbf{LS}} = \mathbf{C}_{\mathbf{uu}}^{-1}\mathbf{C_{uy}}. \tag{8}$$

Through the investigation of the statistical properties of $\hat{\mathbf{g}}_{\mathbf{LS}}$ (Ljung 1987), we then obtain

$$\frac{\hat{g}(k) - g_0(k)}{\sqrt{\mathbf{P}_{\mathbf{L}}^{(kk)}}} \in N(0, 1), \tag{9}$$

where $g_0(k)$ is the supposed true value of $g(k)$, $k = \{-M_1, -M_1 + 1, \ldots, M_2\}$, and $\mathbf{P}_{\mathbf{L}}^{(kk)}$ is the kth diagonal element of $\mathbf{P_L}$ (the covariance matrix of $\hat{\mathbf{g}}_{\mathbf{LS}}$). The computation of $\mathbf{P_L}$ is shown in Eq. (10), under the assumption that all observations are normally distributed (Ljung 1987):

$$\mathbf{P_L} = \frac{1}{L\,[L - (M_1 + M_2 + 1)]} \left\{ \sum_{k=1}^{L} \left[y(k) - \sum_{k=-M_1}^{M_2} \hat{g}(\tau) u(k-\tau) \right]^2 \right\} (\mathbf{C_{uu}})^{-1}. \tag{10}$$

Even when the observations are not normally distributed, it is often the case that the distribution of $\hat{\mathbf{g}}_{\mathbf{LS}}$ approaches the normal distribution as L increases to infinity. This follows from the application of central limit theorems to the sum of random variables that constitutes the estimates (Ljung 1987). For example, the probability that $g_0(k)$ deviates from $\hat{g}_0(k)$ by more than $Z_{(1/2+\alpha/2)}\sqrt{\mathbf{P}_{\mathbf{L}}^{(kk)}}$ is $1 - \alpha$. Herein, $Z_{(1/2+\alpha/2)}$ is the $(1/2 + \alpha/2)$ quantile of the normal distribution, which is available in standard statistical tables. Then, a credential region of $g_0(k)$ with the significance level α can be constructed as $\left|g_0(k) - \hat{g}_0(k)\right| \geq Z_{(1/2+\alpha/2)}\sqrt{\mathbf{P}_{\mathbf{L}}^{(kk)}}$. Therefore, the components of $\hat{g}_0\,(k)$ falling out of this region are considered to have a low credential, and they are not accepted as significant.

3 Spectral factor analysis

Just like the frequency analysis of deterministic processes, the concept of spectral density function can be defined for stochastic processes (see Supplementary Information for details). The advantage of introducing spectral density functions is that we can study a stationary stochastic process in terms of deterministic functions. A spectral factor analysis of stochastic processes is employed here to derive a deterministic criterion for feedback identification.

Let us consider a discrete-time linear system with a sampling interval T_s as shown in Fig. 1. In this case, the dynamical couplings between input $u(k)$ and output $y(k)$ are as follows:

$$Y(z) = H(z)U(z) + L(z)V_0(z), \tag{11}$$

$$U(z) = F(z)Y(z) + N(z)W_0(z), \tag{12}$$

where $U(z)$, $Y(z)$, $V_0(z)$, and $W_0(z)$ represent the Z transforms (Oppenheim et al. 1997) of the signals: input $u(k)$, output $y(k)$, input noise source $w_0(k)$, and output noise source $v_0(k)$ respectively. $H(z)$, $L(z)$, $F(z)$, and $N(z)$ are transfer functions with the frequency variable $z = e^{sT_s}$. For instance, $H(z) = \sum_{k=-\infty}^{\infty} h(k)z^{-k}$ where $h(k)$ is the impulse response and $H(z)$ is the Z transform of $h(k)$. Without loss of generality, we can make the following assumptions (Schnider et al. 1989):

(i) Both $v_0(k)$ and $w_0(k)$ are uncorrelated white noises, $WN(0, 1)$.
(ii) The forward transfer function $H(z)$, the feedback transfer function $F(z)$, the feedforward noise model $L(z)$, and the feedback noise model $N(z)$ are causal, i.e. $H(\infty)$, $F(\infty)$, $L(\infty)$, and $N(\infty)$ should be finite.
(iii) $L(z)$ and $N(z)$ are analytic on and outside of the unit circle; i.e., they are asymptotically stable.

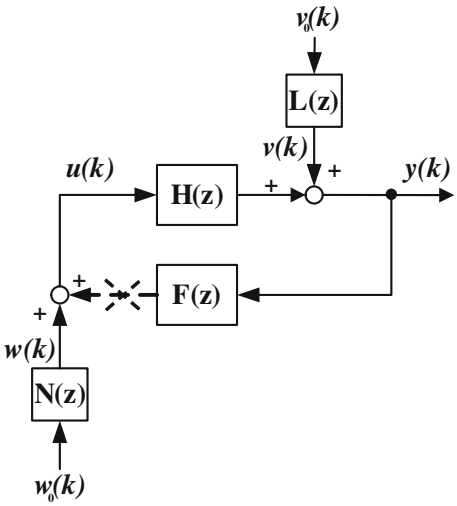

Fig. 1 A mathematical framework for feedback identification. The 'cut point' indicates the two possible cases of feedback or feedback-free

(iv) $H(z)$ is asymptotically stable.
(v) The closed-loop transfer function $\frac{1}{1-H(z)F(z)}$ is asymptotically stable.

Based on these assumptions, formal definitions of feedback-free and feedback systems were proposed by Caines and Chan (1975) as follows: Given a system described by Eqs. (11) and (12), satisfying assumptions (i)–(iv), the system is defined as a "feedback-free system" if $F(z) = 0$. Otherwise (and with a further assumption (v)), it is a "feedback system". We note that assumptions (i)–(iii) and (v) imply that the signals $u(k)$, $y(k)$, $v(k)$, and $w(k)$ are all bounded stationary stochastic variables with rational spectra; the same holds for assumptions (i)–(iv) if there is no feedback, i.e. $F(z) = 0$ and $u(k) = w(k)$.

To identify the existence of a feedback loop between $u(k)$ and $y(k)$, we first need to perform factor decompositions of the spectral density functions. Let $R_{uu}(\tau)$ and $R_{uy}(\tau)$ represent the auto-covariance function of $u(k)$ and the cross-covariance function between $u(k)$ and $y(k)$, respectively. The corresponding discrete spectral density functions are defined by the following two DTFT (See Sect. 2 in Supplementary Information for a strict definition of discrete spectral density function):

$$S_{uu}(z) = \sum_{k=-\infty}^{\infty} R_u(k) z^{-k}, \tag{13}$$

$$S_{uy}(z) = \sum_{k=-\infty}^{\infty} R_{uy}(k) z^{-k}. \tag{14}$$

Then, the non-casual Wiener filter $G(z)$ (Ljung 1987) can be derived as follows:

$$G(z) = S_{uy}(z) S_{uu}^{-1}(z). \tag{15}$$

Finally, we have the following major theorem.

Theorem 1 *Given the closed-loop system described in Eqs.* (11) *and* (12) *with the assumptions* (i)–(v)*,* $G(z)$ is causal if and only if there is no feedback, i.e. $F(z) = 0$.

A complete proof of Theorem 1 is given in the Supplementary Information. In this proof, we also note that if there is no feedback noise (i.e., $w_0(k) = 0$)), $F(z)$ should have at least a one step time-delay to utilize the above concept for identification of a feedback loop. Therefore, we have another theorem which may be used in this case:

Theorem 2 *Given the closed-loop system described in Eqs.* (11) *and* (12) *satisfying* $w_0(k) = 0$, $v_0(k) \sim WN(0, 1)$*, and the assumptions:* (ii)–(v)*, then*:

(1) $G(z) = 1/F(z)$;
(2) $G(z)$ *is non-causal if and only if* $F(z)$ *is strictly proper.*

Proof The proof is straightforward by substituting $w_0(k) = 0$ into Eqs. (11) and (12).

A transfer function is said to be strictly proper, if and only if the degree of its numerator is less than the degree of its denominator. Moreover, a transfer function is said to be biproper, if and only if the degree of its numerator is equal to the degree of

its denominator. For a discrete time transfer function, it requires no time-forward part to be biproper and at least one step time-delay to be strictly proper.

Theorem 1 and 2 imply that there can still exist a feedback loop even if $G(z)$ $(\neq H(z))$ is causal; however, this only happens when $w_0(k) = 0$ and $F(z)$ is biproper. Taken together, the above results provide us with an instructive criterion that can be used to identify the existence of a feedback loop in a SISO linear system regardless of the presence of feedback noise. The criterion also tells us that the appearance of non-causal components in the impulse response sequences indicates the existence of a feedback loop unless $w_0(k) = 0$ and $F(z)$ is biproper. Such a situation is, however, highly implausible in the context of biomolecular interactions which very often involve various scales of time-delay. For the remainder of this paper, we refer to this graphical criterion as the non-causal components criterion (NC-criterion). In the following, we employ it to develop a new identification method for feedback loops in intracellular networks based on linearized cellular dynamics.

4 Linearization of nonlinear dynamics of intracellular networks

The dynamic behavior of an intracellular network is often described as a set of nonlinear differential equations:

$$\mathrm{d}\mathbf{x}/\mathrm{d}t = \mathbf{f}(t, \mathbf{x}, \lambda) + \mathrm{v}, \tag{16}$$

where $\mathbf{x}(t) = (x_1(t), x_2(t), \ldots x_n(t))$, $\lambda = (\lambda_1, \lambda_2, \ldots \lambda_p)$, and $\mathrm{v} = (v_1(t), v_2(t), \ldots v_n(t))$.

In Eq. (16), n is the number of state variables; $\mathbf{x}(t)$ is the n dimensional state vector representing the concentrations or activity levels of the network nodes, i.e., the reacting biomolecules; $\mathbf{f}$ is the n dimensional vector field describing the nonlinear dynamics of the network; λ is the parameter vector denoting external or internal conditions such as rate constants, pH values, and temperatures (as the nominal interaction structure of a network is considered time-invariant, λ is assumed as a constant vector); p is the number of parameters; $\mathbf{v}$ is the additive external noise or perturbation vector such as an impulse, step, ramp, sinusoid, or stochastic function.

In the context of the biological system under investigation, we assume that such perturbations can be realized by an active substance released under control, and that they are applied to some specified nodes in the network. The concept of a feedback loop in nonlinear biological systems is based on the Jacobian matrices of the above system description in Eq. (16). Suppose that the function vector $\mathbf{f}$ is continuously differentiable in its biologically feasible domain, then the Jacobian matrix $\mathbf{A}$ at the steady state $\mathbf{x}^* = (x_1^*, x_2^*, \ldots x_n^*)$ is defined as $\mathbf{A}|_{\mathbf{x}^*} = \partial\mathbf{f}/\partial\mathbf{x}|_{\mathbf{x}^*}$. If the element $\partial f_i/\partial x_j$ of the Jacobian matrix $\mathbf{A}$ is nonzero, this means that x_j has an interaction with x_i. In particular, there is a feedback loop if some components of the system interact with each other in a circular manner. In this case, one set of nonzero elements $\{a_{r_1c_1}, a_{r_2c_2}, \ldots, a_{r_dc_d}\}$ in $\mathbf{A}$ is said to form a d-element feedback loop if and only if the column index set $\{c_1, c_2, \ldots, c_d\}$ is a cyclic permutation of the row index set $\{r_1, r_2, \ldots, r_d\}$ (Thomas and Kaufman 2001).

Since a nonlinear system can be linearized in a small neighborhood of its equilibrium, we can estimate the impulse response sequences from the measurement of time-series data near the equilibrium. In this linearized intracellular network, one artificially perturbed node is chosen as the output node of a SISO system and another node as the input node, while considering the remaining nodes as intermediate variables between the input and output nodes. The impulse response sequences can be obtained by computing the auto- and cross-covariance functions of the input and output data according to Eqs. (4) and (5). Thus, we can determine the existence of a feedback loop between the two nodes by investigating the non-causal components involved.

5 Identification algorithm

In this section, we propose an algorithm for feedback loop identification in linearized intracellular networks. The algorithm has been implemented in Matlab 7.1 (*Mathworks, Inc.*). Let us consider an experimental setup where the state (activity) change of each node can be measured under perturbation. We assume that the perturbation applied to node j can be modeled as $v(k)$ in Fig. 1 and the feedback noise of node i as $w(k)$. So, nodes i and j can be considered as the input and output nodes of a SISO system respectively. The identification steps are as follows:

Step 1. Apply a random perturbation that has a direct influence on a given node j.
Step 2. Measure the time-series profiles $x_i(t)$ and $x_j(t)$ of network nodes i and j with the sampling time T_s.
Step 3. Estimate $R_{uu}(\tau)$ and $R_{uy}(\tau)$ using Eqs. (4) and (5).
Step 4. Estimate the non-causal impulse response sequences $g(k)$ as shown in Eq. (8). Simultaneously, a credential region of the estimated impulse response sequences is determined according to Eq. (9).
Step 5. Apply the NC-criterion. To do this, check the non-causal components of the identified im

pulse response sequences in the negative time axis. If the non-causal components are found to be significant (within the credential region), then a feedback loop exists between nodes i and j, otherwise, no feedback loop exists.
Step 6. Repeat Steps 1–5 for any two nodes of the network.

The complexity of this algorithm can be analyzed as follows: (i) Direct computations of (4) and (5) have a computational complexity $O(L \times (M_1 + M_2 + 1))$, where $M_1 + M_2 + 1$ is the required maximum delay for $R_{uu}(\tau)$. It is required that at least $L \geq M_1 + M_2 + 1$ to compute all needed $R_{uu}(\tau)$ and $R_{uy}(\tau)$. With the use of the FFT algorithm, the complexity is reduced to $O(L \times (1 + \log_2(M_1 + M_2 + 1)))$ (Fang and Xiao 1988). Considering that $M_1 + M_2 + 1$ is always greater than 2, we obtain $O(L \times \log_2(M_1 + M_2 + 1))$ for the computation of the covariance functions. (ii) Straightforward application of Gaussian elimination to solve Eq. (7) is time consuming with complexity $O((M_1 + M_2 + 1)^3)$, since it does not exploit the strong structures present in the Toeplitz system. The Levinson-Durbin recursion can reduce the complexity of solving Eq. (7) to $O((M_1 + M_2 + 1)^2)$ (Golub and van Loan 1996). (iii) To estimate the credential region, we need $O(L \times (M_1 + M_2 + 1))$ flops. Finally,

since $L > M_1 + M_2 + 1$, the total complexity to identify $\hat{\mathbf{g}}_{\mathbf{LS}}$ will be dominated by $O(L \times (M_1 + M_2 + 1)$, and thus it appears that the proposed approach has the potential to be applied to networks of realistic size.

6 Perturbation method

The dynamical behaviors of intracellular networks are usually dominated by the kinetics of biochemical reactions between various biological molecules such as proteins, genes, and RNAs. For example, we can describe biochemical kinetics using the mass action law, introduced by Guldberg and Waage (1879). It is stated that the reaction rate is proportional to the probability of a collision of the reactions. This is in turn proportional to the concentration of reactants to the power of the number in which they enter the specific reaction (Klipp et al. 2005). To investigate the biochemical kinetics, it is necessary to trace the variations of biological molecular concentrations with time, in a situation where the biochemical system is perturbed. Thus, the measurements of bimolecular concentrations in a given time interval form the time series that are to be used for network modeling and identification. In practical experimental conditions, two kinds of perturbations are commonly applied to a biochemical network: (i) parameter perturbations due to changes in external or internal conditions such as rate constants, pH values, and temperatures, (ii) direct concentration perturbations by the release of active molecules under control. The following discussions are focused on the latter, i.e., direct concentration perturbations.

The applied perturbation for the identification can be a random change in the concentration of the node j. Note that the output stochastic perturbation or noise is essential for the identification of non-causal components in our approach. Several different technologies are now available which generate such perturbations in biological systems, e.g. the flash photolysis technology that was developed in cell biology and biochemistry over the last decade (Corrie et al. 1992; Nerbonne 1986).

With flash photolysis technology, the incorporation and photolysis of caged components (biologically relevant molecules) into living cells can be realized through controlled flash sequences. The inactive caged compound precursor is usually formed through the introduction of photolabile chemical groups into certain of the active molecules. Many species of molecules, for example, ATP, neurotransmitters, second messengers, chelators, ionophores, proteins, enzyme inhibitors or activators, and fluorescent dyes, can now be treated with such photochemical preparation (Adams and Tsien 1993; Giovannardi et al. 1998). Then, the active molecules are released by cleaving the photolabile chemical groups with a pulse of intense light in the near ultraviolet (350–360 nm) range. This perturbation technique brings many advantages: for example, a particular intra- or extracellular region can be preloaded with an exactly determined amount of substance in its inactive form (usually in much larger quantity than that released by a single flash); subsequently, it can be controlled and activated at a very precise time. In this way, perturbation delays due to diffusion of substances into the preparation and spatial and temporal inhomogeneities can be greatly attenuated. Therefore, with the flash pulse sequences, one can perturb intracellular networks schematically and delineate their kinetics in a rather precise manner (Adams and Tsien 1993).

Flash light resources widely used include xenon flash lamps, the frequency-doubled ruby laser, or nitrogen laser, and Nd-YAG laser. The pulse width of xenon flash lamps, such as those used in photography, is in the microsecond time scale. For faster reactions, specially designed lasers must be used that have pulse widths in the nanosecond range. Using ultra-fast pulsed lasers allows processes in the picosecond or femtosecond time scale to be studied (Mialocq et al. 1982; Okamura et al. 1991; Miyasaka et al. 1992; Sension et al. 1993). Some PC-controlled Laser Flash Photolysis Spectrometers can provide flexible pulse sequence setup, hardware control, data acquisition, and analysis software for the perturbation designs. For example, a complicated composition of pulse sequences can be formed by a random triggering from a high frequency laser stimulator. Since the width of one pulse is very short, it is calculated that the released amount of active substance by one pulse can be considered proportional to its duration (Rapp and Güth 1988; Brown et al. 1999). Thus, we can photolyze some amount of caged proteins with a high frequency flash pulse sequence to approximate a desired perturbation profile (Ludwig and Ehrhardt 1995; McClung and Hellwarth 2004; Bernardinelli et al. 2005).

The perturbation inputs are commonly designed according to the minimum possible pulse width provided by flash sources as well as the kinetic time constants of the reactions. In Fig. 2, we illustrate how a random binary perturbation (Hull and Dobell 1962) can be approximated by a width-adjustable high frequency flash pulse sequence. Herein, the minimum duration of a single rectangular pulse τ determines the bandwidth of the generated perturbation signals as $f_B = 1/\tau$ (Oppenheim et al. 1997). In addition, for a biochemical process with multiple reactions, the time scale of reactions is often related to the cut-off frequency of the reactions. Suppose that all studied reactions are exponential decay processes and the shortest half life (corresponding to the fastest velocity) of the reactions is $t_{1/2}$. Then, the associated time constant of the reactions can be computed as $\mu = t_{1/2}/\ln 2$. Thus, we have that the approximated cut-off frequency of the fastest reaction is $f_{\text{cutoff}} = \frac{\ln 2}{2\pi}\frac{1}{t_{1/2}} \approx 0.11\frac{1}{t_{1/2}}$. In order to fully excite the reactive system, we require $f_B > f_{\text{cutoff}}$, i.e. we need to guarantee sufficient input energy at frequencies where the system gain is large (Ljung 1987).

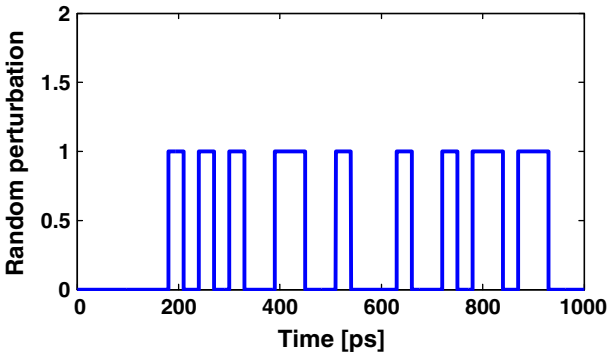

Fig. 2 A scheme of random binary perturbation constructed by flash pulse sequences. The minimum width of rectangular pulse is set as 30 ps (corresponding to 33.3 GHz band-width)

Since pulsed lasers with widths in various time scales (from ms to fs) are available, the kinetic characteristics of most biochemical reactions can be investigated using perturbation signals with sufficient bandwidth. Even in cases where the bandwidth of the perturbation signals is unsatisfactory due to experimental limitations, some other methods can still be employed to obtain a robust estimation–see the next section for a detailed discussion of such methods for band-limited input signals.

7 Factors affecting the identification precision

In the preceding section, we have presented the primary scheme to identify intracellular feedback loops using the NC-criterion. However, it was noted that computational imprecision in the estimation of the covariance functions might influence the identification precision of $g(k)$ significantly, owing to the use of a limited data length L (Rabiner et al. 1978). In cases where the data length L is fixed, the magnitudes of the perturbations and the gains of the transfer functions become the critical factors which influence the identification precision. In order to illustrate this, let us consider a simple case of a linear SISO system in Fig. 3. In this figure, all the transfer functions were transformed into scale-free forms. Gain 1, Gain 2, Gain 3, and Gain 4 represent the gains of the feedback noises, forward transfer function, feedback transfer function, and output noises, respectively. The forward transfer function includes a one step time-delay ($T_s = 1$ min); whereas the feedback transfer function includes a two-step time-delay. In addition, we set the significance level α as 0.0027, which can be regarded as a relatively demanding statistical test level. The same significance level was also adopted in the other example identifications presented in this paper.

Initially, we investigated the identification problem in the open-loop case and studied the underlying factors which influence identification precision. Rabiner et al. (1978) showed that the increase of signal and noise ratio (SNR) and data length L can improve

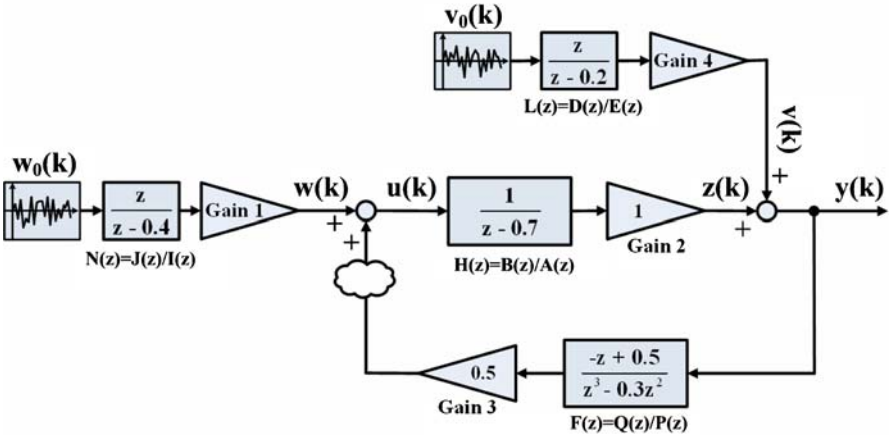

Fig. 3 A simulation model of a linear SISO system to illustrate the NC-criterion. A cloud sign is placed to differentiate the feedback-free cases from the feedback cases

estimates in the open-loop case. The SNR is defined as

$$S/N = 10\log_{10}\left[\frac{\sigma_z^2}{\sigma_v^2}\right], \tag{17}$$

where σ_z^2 is the variance of the signal $z(k)$ and σ_v^2 is the variance of the noise signal $v(k)$. Thus, a high ratio of Gain 1·Gain 2/Gain 4 is beneficial for achieving a high precision in the open-loop identifications—see Fig. 4 for one illustration of this conclusion. Moreover, although the computational errors introduced by $R_{uv}(\tau)$ can be efficiently alleviated with a long data length ($L \gg M_1 + M_2 + 1$), we can still obtain a reliable result with limited data lengths, under the condition that the SNR is not extremely small. In Fig. 5, if only 40 or 80 data points are available for computations, the significant causal components can still be observed, even when the gain ratio is 10. Thus the NC-criterion does not in general require a long data length, making it cost efficient both in terms of experiments and computations.

Next, we investigated the situation with a feedback loop, where the broken line was reconnected by removing the cloud sign in Fig. 3. The result showed that the ratio of (Gain 1·Gain 2)/(Gain 3·Gain 4) is now the crucial factor that influences the identification of $g(k)$. For simplicity, we just give the cases where only Gain 1 and Gain 4 are adjusted. In Fig. 6a–e, the non-causal component with a two step time-delay can be correctly identified under the condition that $L = 300$. With an increase of the ratio of Gain 1/Gain 4, the identified non-causal component becomes gradually indiscernible. Fig. 6f–h showed that the non-causal components cannot be identified successfully when the ratio (Gain 1/Gain 4) was as large as 10, 20, and ∞. This result further strengthened the previous conclusion that a certain magnitude of output perturbations are necessary to yield the non-causal components in $g(k)$. Additionally, following the

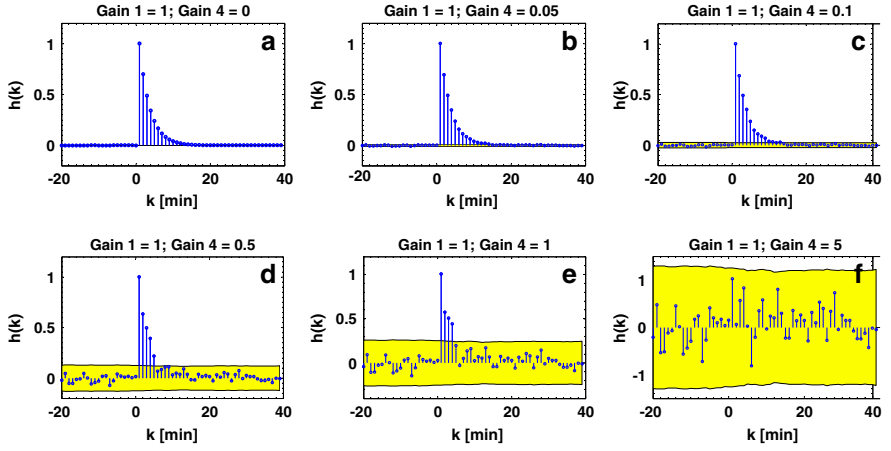

Fig. 4 The influence of the ratio (Gain 1–Gain 4) on the identification precisions in the open-loop cases. The data length used for the identification is equal to 300. The index of $g(k)$ is set to $k = -20, -19, \ldots, 40$. **a** Gain 1 = 1 and Gain 4 = 0. **b** Gain 1 = 1 and Gain 4 = 0.05. **c** Gain 1 = 1 and Gain 4 = 0.1. **d** Gain 1 = 1 and Gain 4 = 0.5. **e** Gain 1 = 1 and Gain 4 = 1. **f** Gain 1 = 1 and Gain 4 = 5 (the failure case of the identification)

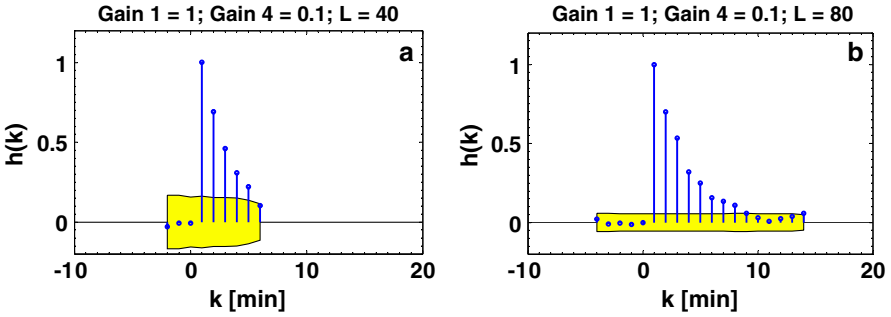

Fig. 5 The identifications using two short data length in the open-loop cases. The gain ratio (Gain 1 · Gain 2 − Gain 4) is kept as 10. **a** $L = 40$. The index of $g(k)$ is set as $k = -2, -1, \ldots, 6$. **b** $L = 80$. The index of $g(k)$ is set to $k = -4, -3, \ldots, 14$

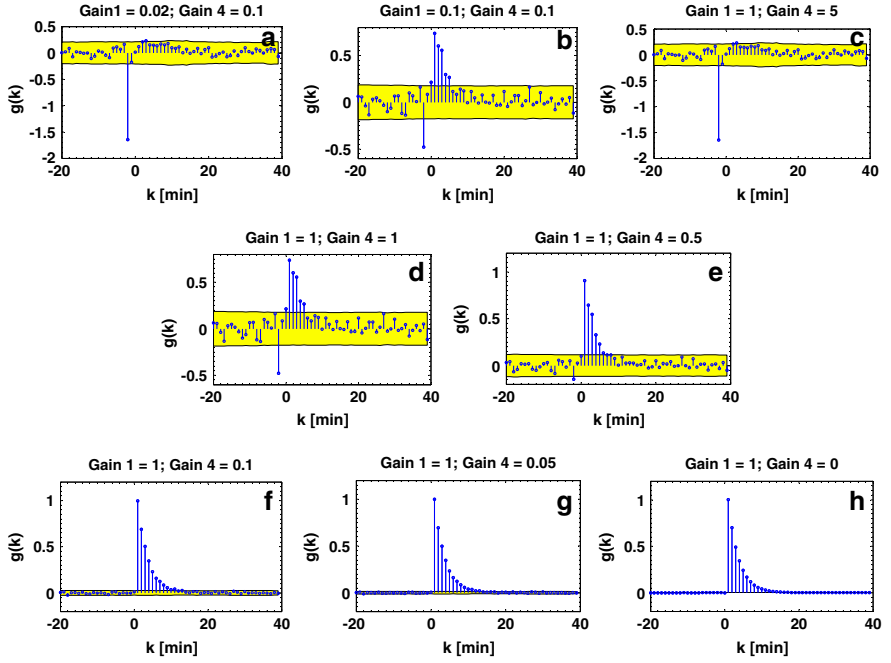

Fig. 6 The influence of the ratio (Gain 1–Gain 4) to the identification precisions in the closed-loop cases. The data length used for the identification is equal to 300. The index of $g(k)$ is set to $k = -20, -19, \ldots, 40$. **a** Gain 1 = 0.02 and Gain 4 = 0.1. **b** Gain 1 = 0.1 and Gain 4 = 0.1. **c** Gain 1 = 1 and Gain 4 = 5. **d** Gain 1 = 1 and Gain 4 = 1. **e** Gain 1 = 1 and Gain 4 = 0.5; **f** Gain 1 = 1 and Gain 4 = 0.1 (the failure case). **g** Gain 1 = 1 and Gain 4 = 0.05 (the failure case). **h** Gain 1 = 1 and Gain 4 = 0 (the failure case)

same reasoning as in the previous open-loop cases, increasing the data length used for identification can also improve the precision of the identifications in these closed-loop cases. Figure 7a and b took two long data lengths: 4000 and 400,000 min; it was noted that the precision of identifications were improved by the increase of L (compare Fig. 7a and b with Fig. 6f). It is often difficult to ensure stationarity of time series

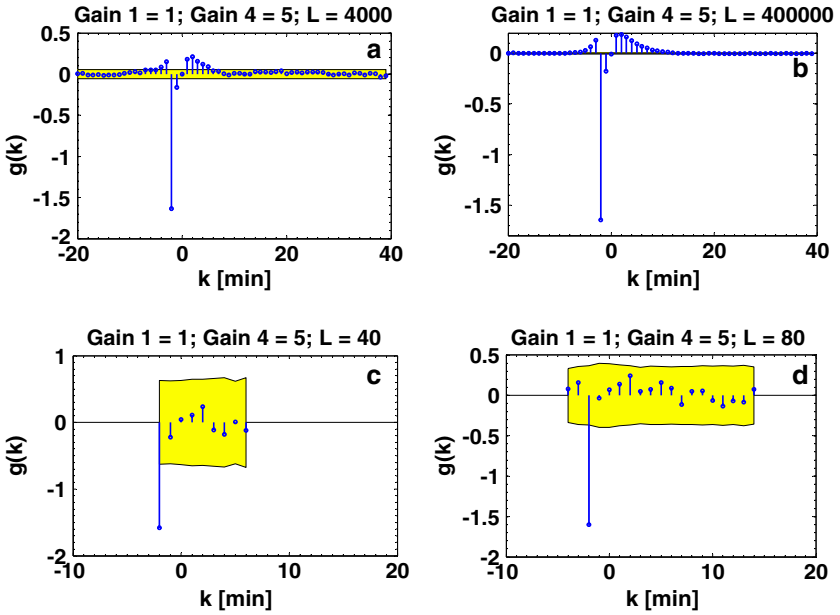

Fig. 7 The identifications using various data length in the closed-loop cases. The gain ratio (Gain 1–Gain 4) is kept as 0.2. **a** $L = 4000$. The index of $g(k)$ is set to $k = -20, -19, \ldots, 40$. **b** $L = 400000$. The index of $g(k)$ is set to $k = -20, -19, \ldots, 40$; **c** $L = 40$. The index of $g(k)$ is set to $k = -2, -1, \ldots, 6$. **d** $L = 80$. The index of $g(k)$ is set to $k = -4, -3, \ldots, 14$

when collecting long data records from biological preparations. Thus, the data length that can be applied for identification is usually limited. In practice, however, a shorter data length is often enough for accurate identification of the non-causal components, under the condition that the gain ratio (Gain 1·Gain2)/(Gain 3·Gain 4) is not extremely large. Figure 7c and d showed the results of applying two short data lengths of 40 and 80 min, respectively; the simulations showed one two-step non-causal component of $g(k)$ was successfully identified in each of the two cases.

We also illustrated the situation addressed by Theorem 2, where $w_0(k) = 0$ and the other system structure was maintained as in Fig. 3. The influence of the properness of $F(z)$ on the non-causal component identification was investigated by assigning both a strictly proper and a proper transfer function to $F(z)$. The two cases, $F(z) = (-z + 0.5) / (z^3 - 0.3z^2)$ (strictly proper) and $F(z) = (-z + 0.5) / (z - 0.3)$ (biproper), were simulated and the results were compared in Fig. 8. The only difference between these two feedback transfer functions is that the former one is delayed by 2 time-steps with respect to the later one. The two non-causal components of $g(k)$ appeared in Fig. 8a, showing the existence of the feedback loop. In contrast, $g(k)$ was causal in Fig. 8b. The algorithm thus failed to deduce the existence of the feedback loop due to the biproperness of $F(z)$. However, if we include the feedback noise in Fig. 3, the non-causal components were again present (see Fig. 8c; Gain 1 was reset to one). This alteration made the simulation return to the situation dealt with in Theorem 1, i.e., although $F(z)$ is bi-proper, there still exist some identified non-causal components when $w_0(k) \neq 0$.

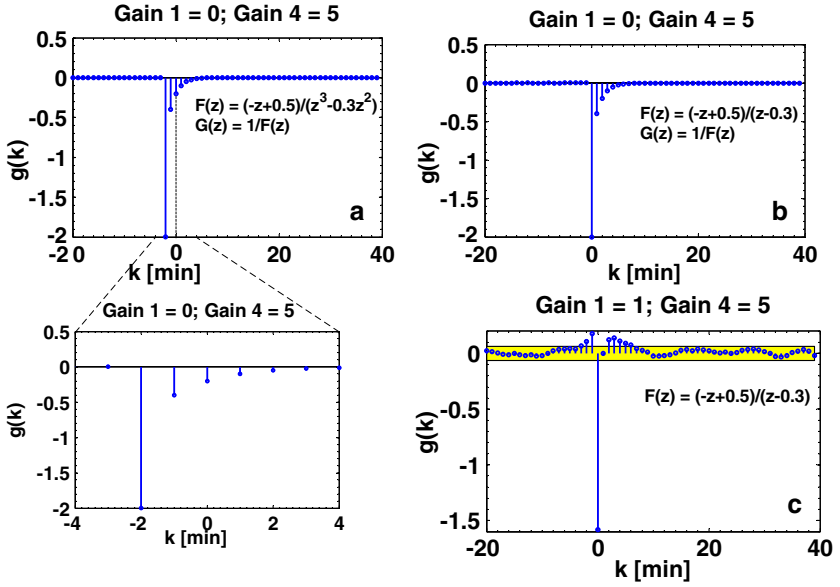

Fig. 8 The influence of the properness of the feedback transfer function on the identifications. The index of $g(k)$ is set as $k = -20, -19, \ldots, 40$. The data length used for identification equals 300. **a** Gain1 = 0, Gain 4 = 5, $F(z) = (-z + 0.5) / \left(z^3 - 0.3z^2\right)$ (strictly proper), and the inlet below (bf a) shows the amplified region $k = -4, -3, \ldots, 0, \ldots, 4$. **b** Gain 1 = 0, Gain 4 = 5, and $F(z) = (-z + 0.5) / (z - 0.3)$ (proper). **c** Gain 1 = 1, Gain 4 = 5, and $F(z) = (-z + 0.5) / (z - 0.3)$ (proper)

Under the assumption of limited data length, it was the ratios between the different signal gains rather than their absolute values which most influence the identification precision. This means that perturbation signals of small magnitudes are admissible for achieving a high precision of identification without driving the system far from its equilibria. This therefore provides a reliable theoretical basis for applying small perturbation signals near the equilibria of the nonlinear intracellular networks. As shown in the above illustrations, for the identifications about $h(k)$ in the open-loop cases, a high ratio of input signal to output noise was preferential; whereas, for the identifications about $g(k)$ in the closed-loop cases, a high ratio of output noise feedback signal to input signal should be given a priority. In the context of intracellular networks, we are usually interested in those feedback regulations that have significant gains both in the feedforward and feedback channels. Therefore, we can adopt an appropriate output perturbation to satisfy the requirements on the gain ratio in both types of identifications by just adjusting the magnitudes of the biological perturbations applied. In this way, the above failure cases in the identification procedure can be avoided through an appropriate choice of biological experimental setup.

Another critical factor influencing identification accuracy is the bandwidth of the input signals. As addressed in many references (Rabiner et al. 1978; Golub and van Loan 1996; Dayal and MacGregor 1996), a high condition number of $\mathbf{C_{uu}}$ is unfavorable for solving $\hat{\mathbf{g}}_{\mathbf{LS}}$ in a linear equation like Eq. (7). When $u(k)$ is colored (band limited), the condition number of $\mathbf{C_{uu}}$ typically increases significantly. Thus, the spectrum

of $u(k)$ would seriously influence the numerical stability of Eq. (7). Many methods, for example, ridge regression (Hoerl and Kennard 1970a,b), the partial least squares method (Geladi and Kowalski 1986), and the pseudo-inverse method (Westwick and Kearney 1997) have been developed to deal with such a problem. Here, we shall use ridge regression and pseudo-inverse method to illustrate this issue for identifications with colored inputs. For ridge regression, a biased estimate of $\hat{\mathbf{g}}_{\mathbf{LS}}$, which is defined as $\hat{\mathbf{g}}_{\mathbf{RR}}$, can be found by

$$\hat{\mathbf{g}}_{\mathbf{RR}} = (\mathbf{C}_{\mathbf{uu}} + k\mathbf{I})^{-1}\,\mathbf{C}_{\mathbf{uy}}, \tag{18}$$

where k is a scalar, and $\mathbf{I}$ is the identity matrix. This is equivalent to adding a constraint on the magnitude and length of the estimated parameters, so that the sum of the variances can decrease very rapidly with an increase in k (Hoerl and Kennard 1970a). The scalar k is selected using the starting point of a stable region of the ridge trace (Hoerl and Kennard 1970b). For the pseudo-inverse method, a singular value decomposition (SVD) is performed as follows:

$$\mathbf{C}_{\mathbf{uu}} = [\mathbf{V_1} \quad \mathbf{V_2}] \begin{bmatrix} \mathbf{S_1} & \mathbf{0} \\ \mathbf{0} & \mathbf{S_2} \end{bmatrix} \begin{bmatrix} \mathbf{V_1^T} \\ \mathbf{V_2^T} \end{bmatrix}, \tag{19}$$

where the subscript "1" represents the elements that are related to the larger singular values, and the subscript "2" denotes the elements that are associated with the insignificant singular values. During the construction of the pseudo-inverse, the elements which are distinguishable from the effects of noise $v(k)$ are retained (Westwick and Kearney 1997). Then, a pseudo-inverse can be formed as

$$\mathbf{C}_{\mathbf{uu}}^{+} = \mathbf{V_1}\mathbf{S_1^{-1}}\mathbf{V_1^T}, \tag{20}$$

Replacing $\mathbf{C}_{\mathbf{uu}}$ with $\mathbf{C}_{\mathbf{uu}}^{+}$, the impulse response estimate is computed as

$$\hat{\mathbf{g}}_{\mathbf{PI}} = \mathbf{C}_{\mathbf{uu}}^{+}\mathbf{C}_{\mathbf{uy}}, \tag{21}$$

We show how these two methods can handle the case of colored inputs via a numerical example—see the open-loop structure in Fig. 1. A white noise signal $w_0(k)$ ($L = 1000$; sample frequency $f_s = 100\,\text{Hz}$) was applied to the input of $N(z)$, which is a 3rd order Chebyshev type I digital filter. This filter has 0.1 dB peak-to-peak ripple in the pass-band and 0.3 normalized cut-off frequency (the normal cut-off frequency is half of the sample frequency; all frequencies are normalized by this value). Thus, $u(k)$ (identical to $w(k)$) becomes a colored stochastic process with a cut-off frequency of 15 Hz. The test system $H(z)$ is constructed as a fifth order Chebyshev type I digital filter, which has 0.1 dB peak-to-peak ripple and 0.5 normalized cut-off frequency. The output of $H(z)$ is added with a 7 dB SNR white noise signal to form the noise-corrupted output $y(k)$. The identification of the impulse response $h(k)$ ($= g(k)$) is shown in Fig. 9a, where the relative error of the estimate is evaluated by

(Rabiner et al. 1978)

$$Q = 10 \log_{10} \left(\frac{\sum_{k=-M_1}^{M_2} \left(g(k) - \hat{g}(k) \right)^2}{\sum_{k=-M_1}^{M_2} g^2(k)} \right). \tag{22}$$

A large value of Q signifies a low accuracy of the estimate. As shown in Fig. 9a, the exact inverse method produces a quite sensitive estimate with an error $Q_{EI} = -10.1022$. The ridge regression method has $Q_{RR} = -13.8232$, and thus produces an estimate with much higher accuracy than Q_{EI}. The pseudo-inverse method provides

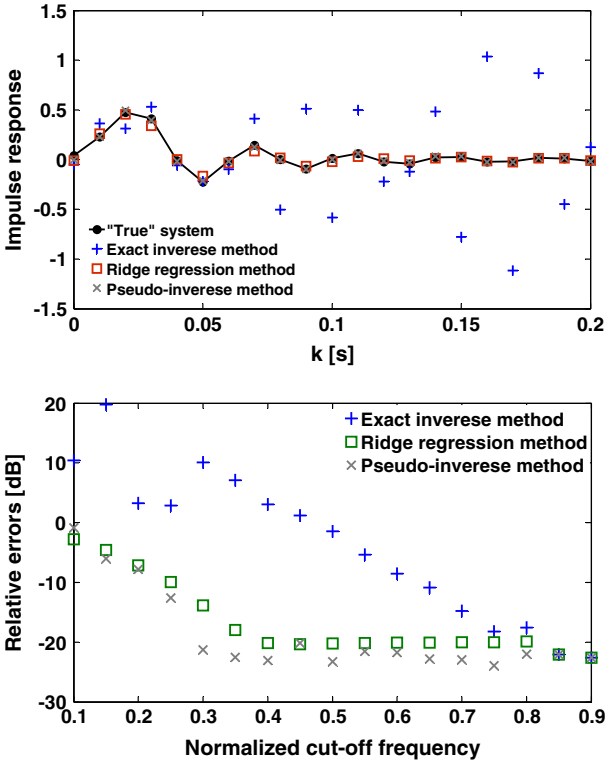

Fig. 9 The identification of impulse response for colored inputs by the three methods: exact inverse method, ridge regression method, and pseudo-inverse method. A 5th order Chebyshev type I digital filter is taken as a test system, which has 0.1 dB peak-to-peak ripple and 0.5 normalized cut-off frequency. **a** The identified impulse response when a colored input is used to excite the test system. The colored input is obtained as a white noise passing through a 3rd order Chebyshev type I digital filter with 0.1 dB peak-to-peak ripple and 0.3 normalized cut-off frequency. The relative errors of three identification methods are: $Q_{EI} = -10.1022$; $Q_{RR} = -13.8232$; $Q_{PI} = -21.2986$. The latter two methods greatly improve the identification accuracies. **b** The change of identification errors over a range of input bandwidth. The bandwidth of the colored input has been changed from 0.1 to 0.9 normalized frequencies. In the low and middle frequency bands much higher identification accuracies are observed for the ridge regression method and the pseudo-inverse method than for the exact inverse method. However, at high frequencies (e.g. 0.85 and 0.9) all three methods reach nearly the same accuracies, which are all quite high

an even better performance with $Q_{PI} = -21.2986$. We also showed the influence of the bandwidth of the colored $u(k)$ on the errors in the estimates. In Fig. 9b, the normalized cut-off frequency of $N(z)$ varies from 0.1 to 0.9 by an increment of 0.05. It is shown that, for most band-limited inputs, the ridge regression method and pseudo-inverse method have improved accuracy over the exact inverse method. When the bandwidth of $u(k)$ is greater than 0.85 times the normal frequency, the three methods almost reach the same accuracy. Thus we see that there is no significant difference in the errors of the estimates if the bandwidth of the input signal is much greater than that of the test system. However, for specific physiological conditions where this is not the case, we are able to use the ridge regression method and pseudo-inverse method to obtain an improved identification performance for colored input signals.

8 Identification of negative feedback loops in a synthetic network

In this section, the NC-criterion was applied to a synthetic model containing a negative feedback loop. We constructed this in silico feedback model using the kinetic mechanisms and parameters for a set of protein interactions obtained from the literature (Szallasi et al. 2005). In Fig. 10, the kinase X phosphorylates Y, and the phosphorylated $Y(Y_P)$ activates the degradation of X. In this case, two ODEs (ordinary differential equations) were used to describe the system behavior:

$$\frac{\mathrm{d}X}{\mathrm{d}t} = k_1 S - k_2 Y_P X + w_0(t), \tag{23}$$

$$\frac{\mathrm{d}Y_P}{\mathrm{d}t} = \frac{k_3 X (Y_T - Y_P)}{K_{m3} + Y_T - Y_P} - \frac{k_4 E Y_P}{K_{m4} + Y_P} + v_o(t), \tag{24}$$

where Y and Y_P satisfy the conservation relation and their total amount is denoted by Y_T (the units are as follows: $[X] = [Y_P] = [S] = [E] = nM, [k_1] = [k_3] = [k_4] = \mathrm{min}^{-1}$, $[k_2] = \mathrm{nM}^{-1}\mathrm{min}^{-1}$, and $[K_{m3}] = [K_{m4}] = \mathrm{nM}$). The equation for X describes constant synthesis (proportional to S) minus degradation (proportional to

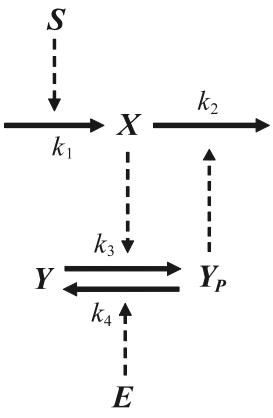

Fig. 10 A synthetic example network containing a feedback loop. Parameter values are assumed as $k_1 = k_2 = k_3 = k_4 = 1$, $S = K_{m3} = K_{m4} = 0.1$, $Y_T = 1$, and $E = 0.5$

$Y_P \cdot X$), while the equation for Y_P adopts Michaelis-Menten kinetics to describe the saturating synthesis and degradation rates.

The identification procedure was as follows. Let us choose X as the input of the network and Y_P as the output. During a deterministic simulation (free from noise), the states of the system are to converge to its stable equilibrium. Fig. 11a shows the temporal profiles of X and Y_P from the initial conditions ($X_0 = 0$ and $Y_{P0} = 0$) until its stable equilibrium ($X^* = 0.4036$, $Y_P^* = 0.2478$). To reflect the actual intracellular environment, we added continuous feedback noise $w_0\,(t)$ and feedforward noise $v_o\,(t)$ to the network, which are uncorrelated white noises $WN(0, 0.001)$ and $WN(0, 0.01)$, respectively. Then we simulated this continuous-time model and took samples for the time-series data of X and Y_P. Figure 11b shows the stochastic simulation near the system's equilibrium ($X^* = 0.4036$, $Y_P^* = 0.2478$). The non-causal impulse response sequence $g(k)$ (shown in Fig. 11c) was obtained by applying the proposed algorithm in Sect. 5. Our analysis correctly inferred the existence of a feedback loop, since one significant non-causal component was observed at $k = -T_s$. The other significant components observed at $k = 0$ and $k = T_s$ min show the causal part of $g(k)$.

To compare the impulse response sequence of the feedback loop case with that of the feedback-free case, we removed the feedback regulation in Eq. (23) by substituting

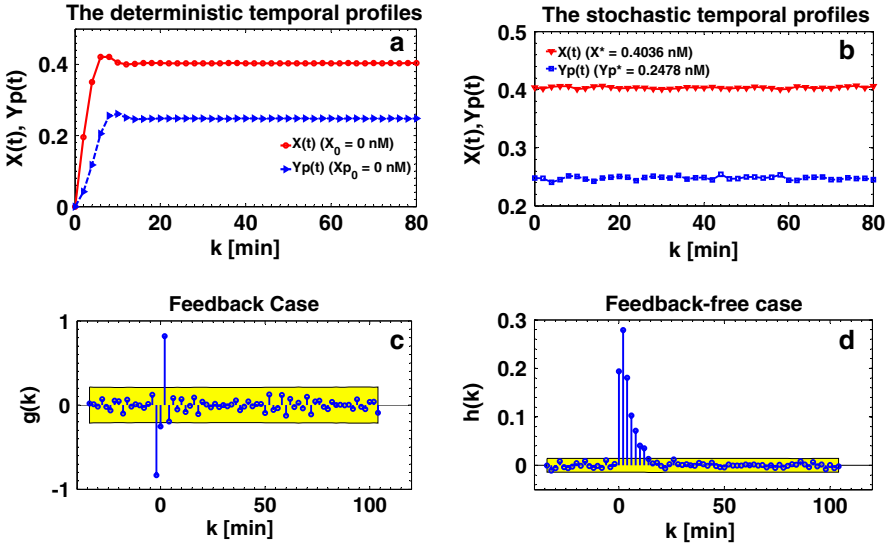

Fig. 11 Feedback loop identification in the synthetic biological network. $T_s = 2$ min; $\alpha = 0.0027$; $L = 300$. **a** The deterministic simulation in the system containing a negative feedback loop. The initial condition is $X_0 = 0$ and $Y_{P0} = 0$ nM. The steady state is $X^* = 0.4036$, $Y_P^* = 0.2478$ nM. **b** The stochastic responses of X and Y_P in the system containing a negative feedback loop. In this case, the feedback noise $w_0\,(t) = WN\,(0, 0.001)$ and the feed forward noise $v_o\,(t) = WN\,(0, 0.01)$ provide the required small disturbances to X and Y_P. **c** The identified non-causal impulse response sequences $g(k)$ from the temporal profile of Fig. 9b. The one step non-causal component at $k = -T_s$ indicates the existence of a feedback loop. **d** The identified causal impulse response sequences $h\,(k)$ of the feedback-free model. The feedback noise $w_0(t) = WN\,(0, 0.01)$ and the feed forward noise $v_o\,(t) = WN\,(0, 0.001)$ are applied. The identified impulse response sequence $h(k)$ is causal, implying that the system is feedback-free

Y_P with its equilibrium value (to set the same initial conditions). For the feedback-free model, $w_0(t) \sim WN(0, 0.01)$ and $v_o(t) \sim WN(0, 0.001)$ were used, and the impulse response sequences were identified as illustrated in Fig. 11d. As there was no significant non-causal component in the impulse response sequences, our algorithm correctly infers that there is no feedback loop present in this case.

9 Identification of a positive feedback loop in the caspase cascade signaling model

Apoptosis is a type of programmed cell death which is regulated in an orderly manner by a series of signal cascades. It plays a crucial role in embryonic development, immune responses and the elimination of virally infected or transformed cells. A misregulation of apoptosis often results in severe pathological alterations. For instance, many cancers are difficult to be eradicate because cells lose the ability to respond to apoptotic signals. Conversely, neurodegenerative disorders such as Parkinson's, Alzheimer's, and Huntington's diseases are characterized by excessive apoptotic activity in certain classes of neurons (Thompson 1995; Haass 1999). Therefore an in-depth understanding of the underlying mechanisms of apoptosis is of great medical interest.

A caspase cascade involving a family of caspases (cysteine-containing aspartate-specific proteases) forms a crucial part of the apoptotic signal transduction pathway. Quantitative modeling of the caspase cascade was first undertaken by Eissing et al. (2004, 2005) who elucidated the multiple equilibria (bistability) and switch-like properties of caspase cascade signal transduction pathways. Another kinetic model was proposed by Bullinger (2005) to analyze the long lag phase of the activated caspase 3 ($C3^*$) responses. In this model, an apoptopic decision is made by a positive feedback regulation between the activated caspase 8 ($C8^*$) and the activated caspase 3 ($C3^*$). Here, we employed Bullinger's model to generate time-dependent responses of the network variables, and then used these data to illustrate the proposed NC-criterion. The network is shown in Fig. 12 and all the equations and parameters are detailed in Table 1. In this system description, we regard $C3^*$ and $C8^*$ as the input and output of the system, corresponding to $u(t)$ and $y(t)$ in Fig. 1, respectively. All the other molecules involved in the indirect interactions between $C3^*$ and $C8^*$ constitute inner state variables between the input and output.

Through steady state analysis, we found three equilibria of the system: E1 (normal steady state), E2 (transitory state), and E3 (apoptotic steady state). E1 and E3 are stable equilibria while E2 is an unstable saddle manifold. It was also found that apoptotic steady state E3 has a relatively larger stable region compared to normal steady state E1 (Bullinger 2005). This characteristic of E3 is favorable for applying perturbations while maintaining the stability of systems. We thus chose E3 (9132.37, 74380.1, 18.97, 5161.68, 264.16, 2999.32, 20.54, 3446.51) as an initial condition for the simulation. Then, we applied $w_0(t) \sim WN(0, 0.001)$ and $v_0(t) \sim WN(0, 5)$ as the appropriate excitation signals, while keeping the fluctuations of $C3^*$ and $C8^*$ within small ranges. For this particular example, we can estimate the forward gain (corresponding to Gain 2 in Fig. 3) as $k_2 \cdot [C8]_E = 1.0 \times 10^{-5} \times 9132.37 = 0.09132$, where $[C8]_E$ is the concentration of $C8$ at $E3$; the feedback gain (corresponding to Gain 3 in Fig. 3) as

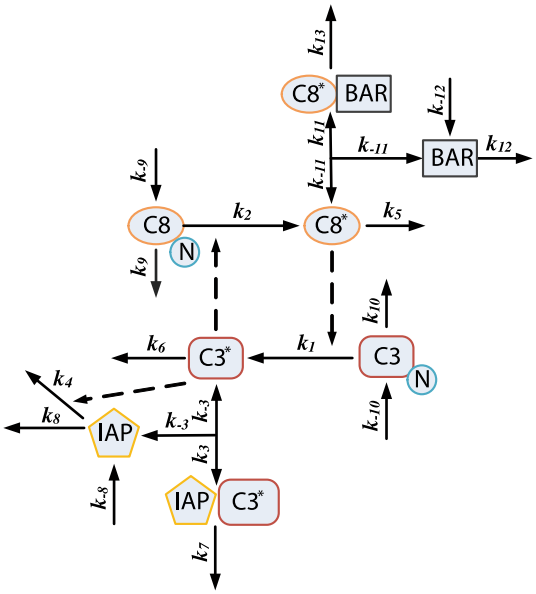

Fig. 12 A diagram of the caspase signaling network. $C8$ is caspase-8, $C8^*$ is activated caspase-8; $C3$ is caspase-3, $C3^*$ is activated caspase-3, *IAP* is the apoptosis inhibitor protein, *CARP* is caspase-associated ring protein, $C8^* \sim IAP$ is the complex of $C8^*$ and *IAP*, and $C3^* \sim CARP$ is the complex of $C3^*$ and *CARP*. The solid arrows indicate the directions of reaction fluxes; the dash arrows show the directions of the control functions. All molecules are assumed to be degraded with a constant rate and four molecules $C8$, $C3$, *IAP*, and *CARP* are assumed to be produced with a constant rate. All reaction parameters are detailed in Table 1

$k_1 \cdot [C3]_E = 5.8 \times 10^{-5} \times 18.97 = 0.0011$, where $[C3]_E$ is the concentration of $C3$ at $E3$. From the above computations, we saw that the forward regulation gain is nearly 83 times that of the feedback regulation gain. In the following, we will show that this choice of excitation signals works even though the gain ratio is 83. Under the excitations, the kinetic model produces near-linear network responses. In particular, the temporal profiles of $C3^*$ and $C8^*$ were obtained (Fig. 13a, b) and, based on these input and output data, we can identify $g(k)$ in Fig. 13c. As shown in the figure, for $g(k)$ two significant non-causal components appear in the negative time axis. This implied the existence of a feedback loop between $C3^*$ and $C8^*$. As with the previous synthetic model example, we tested the reliability of our algorithm by removing the feedback regulation from the output to the input by substituting $C8^*$ and $C3$ with their steady state values. For this case, we applied $w_0(t) \sim WN(0, 10)$ and $v_0(t) \sim WN(0, 0.1)$ as the excitation signals. With this feedback-free model, the impulse response of the open-loop system was again identified (Fig. 13d). In this case, there was no non-causal component in the negative time axis, correctly implying the absence of any feedback loop and clearly demonstrating the accuracy of our proposed identification algorithm.

10 Concluding remarks

We have addressed the problem of identifying the presence or absence of unknown intracellular feedback loops from time-series measurements of intact cell responses for given perturbations. By using correlation identification and spectral factor analysis, we have proposed a novel identification criterion called the NC-criterion, which provides a theoretical as well as an experimental basis for the identification of feedback loops in intracellular networks. The proposed identification method was applied

Table 1 The ODE model describing the caspase signal transduction network

Equations	Rate constants
$\mathrm{d}[C8]/\mathrm{d}t = -v_2 - v_9 = -k_2[C3^*][C8] - (k_9[C8] - k_{-9})$	
$\mathrm{d}[C8^*]/\mathrm{d}t = v_2 - v_5 - v_{11} = k_2[C3^*][C8] - k_5[C8^*] - (k_{11}[C8^*][BAR] - k_{-11}[C8^* \sim BAR]) + v_0(t)$	k_1 = 5.8e−5; k_2 = 1.0e−5;
	k_3 = 5.0e−4; k_4 = 3.0e−4;
$\mathrm{d}[C3]/\mathrm{d}t = -v_1 - v_{10} = -k_1[C8^*][C3] - (k_{10}[C3] - k_{-10})$	k_5 = 5.8e−3;k_6 = 5.8e−3;
	k_7 =1.73e−2; k_8 = 1.16e−2;
$\mathrm{d}[C3^*]/\mathrm{d}t = v_1 - v_3 - v_6 = k_1[C8^*][C3] - (k_3[C3^*][IAP] - k_{-3}[C3^* \sim IAP]) - k_6[C3^*] + w_0(t)$	k_9= 3.9e−3; k_{10} = 3.9e−3;
	k_{11} = 5.0e−4; k_{12}= 1.0e−3;
$\mathrm{d}[IAP]/\mathrm{d}t = -v_3 - v_4 - v_8 = -(k_3[C3^*][IAP] - k_{-3}[C3^* \sim IAP]) - k_4[C3^*][IAP] - (k_8[IAP] - k_{-8})$	k_{13} = 1.16e−2;
$\mathrm{d}[C3^* \sim IAP]/\mathrm{d}t = v_3 - v_7 = (k_3[C3^*][IAP] - k_{-3}[C3^* \sim IAP]) - k_7[C3^* \sim IAP]$	k_{-3} = 0.21; k_{-8} = 464;
$\mathrm{d}[BAR]/\mathrm{d}t = -v_{11} - v_{12} = -(k_{11}[C8^*][BAR] - k_{-11}[C8^* \sim BAR]) - (k_{12}[BAR] - k_{-12})$	k_{-9} = 507; k_{-10} = 81.9;
$\mathrm{d}[C8^* \sim BAR]/\mathrm{d}t = v_{11} - v_{13} = (k_{11}[C8^*][BAR] - k_{-11}[C8^* \sim BAR]) - k_{13}[C8^* \sim BAR]$	k_{-11} = 0.21; k_{-12} = 40

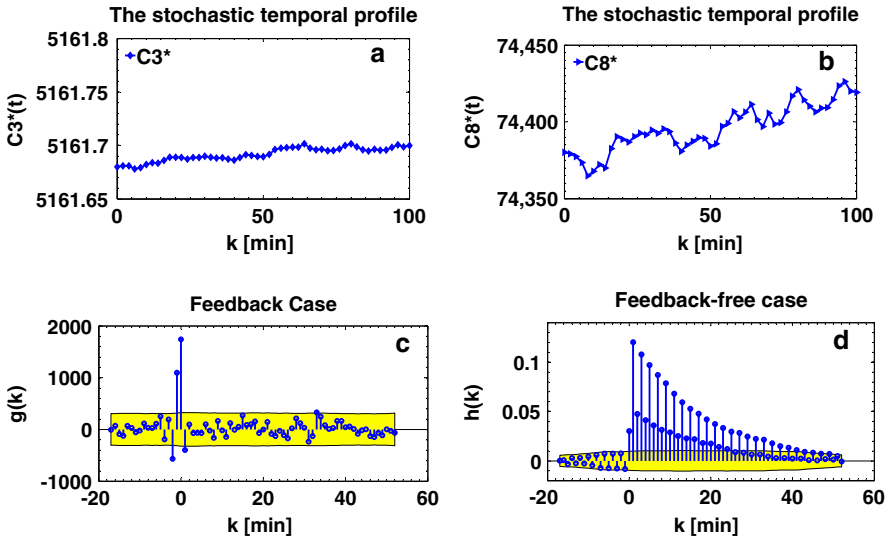

Fig. 13 Feedback loop identification in the Caspase cascade signaling model. $T_s = 2$ min; $\alpha = 0.0027$; $L = 300$. **a** The segment response of $C3^*$ in the network containing a positive feedback loop. The feedback noise $w_0(t) = WN(0, 0.001)$ and the feedforward noise $v_o(t) = WN(0, 5)$. **b** The segment response of $C8^*$ in the same network containing the positive feedback with $w_0(t) = WN(0, 10)$ and $v_o(t) = WN(0, 0.1)$ **c** The identified non-causal impulse response sequences $g(k)$ of the closed-loop system from the temporal data as shown in Fig. 5a and b. **d** The identified causal impulse response sequences $h(k)$ after the activation from $C8^*$ to $C3^*$ is eliminated. (i.e., feedback-free case). Herein, $w_0(t) = WN(0, 10)$ and $v_o(t) = WN(0, 0.1)$

to a synthetic negative feedback model and a caspase signal transduction model with a positive feedback loop. In these examples, significant non-causal components of the impulse response sequences were observed in the negative time axis, correctly implying the existence of a feedback loop. Such non-causal components were no longer observed when the feedback loops were removed from the models, thus demonstrating the ability of the proposed approach to distinguish between feedback and feedback-free systems.

We employed a noise-driven linearized system as the mathematical framework for our approach. Thus, the measurement noises, usually considered as external noises which are uncorrelated with the intrinsic regulations, can be significantly attenuated by the correlation computations. Hence, the NC-criterion is very suitable for application to feedback identification in noisy biological networks. In addition, it only requires the measurement of dynamic expression profiles of the input and output network nodes. Various experimental methods including real-time PCR, immunofluorescence, and microarray technologies can now provide such time-series data in a cost efficient manner.

We noted that perturbations near the output port can give rise to non-causal components in the impulse responses. Thus, to apply the NC-criterion, we only need output noises or perturbations—the perturbability of the input node is not required. Since the feedback structure is often symmetric and therefore either of the two nodes can

arbitrarily be chosen as the output node, the requirement of perturbablity can often be further relaxed in practice.

A significant advantage of the NC-criterion is the lack of a requirement for any information about the structure of the network. Such a priori information is usually difficult to obtain in biological systems. Hence, a relatively simple nonparametric identification approach is often more appropriate for the identification of feedback loops.

It has been shown that the NC-criterion is applicable to nonlinear cellular dynamics near equilibria. The range of perturbations should be maintained within the neighborhood of equilibria such that the nonlinear system can be approximated as a linearized system and thereby the correlation analysis can capture non-causal components between input and output. In this way, the existence of a feedback loop can be identified by investigating those non-causal components. Note however that only the linear dynamics of a given intracellular network are preserved when we apply the NC-criterion. To overcome this limitation, other approaches, such as using the Wiener-Volterra series extension of nonlinear systems, could be investigated in future extensions of the approach.

Acknowledgments This work was supported by the Korea Science and Engineering Foundation (KOSEF) grant funded by the Korea Ministry of Education, Science & Technology through the Systems Biology grant (M10503010001-07N030100112), the Nuclear Research grant (M20708000001-07B0800-00110), the 21C Frontier Microbial Genomics and Application Center Program (Grant MG08-0205-4-0), and the WCU grant (R32-2008-000-10218-0). This work was also supported by the "Systems biology infrastructure establishment grant" provided by Gwangju Institute of Science & Technology in 2009. Tae-Woong Yoon is grateful to a Grant from Korea University.

References

Adams SR, Tsien RY (1993) Controlling cell chemistry with caged compounds. Annu Rev Physiol 55:755–784. doi:10.1146/annurev.ph.55.030193.003543

Baccala LA, Sameshima K (1999) Direct coherence: a tool for exploring functional interactions among brain structures. Methods for neural ensemble recordings. CRC press LLC, Boca Raton

Bernardinelli Y, Haeberli C, Chatton JY (2005) Flash photolysis using a light emitting diode: an efficient, compact, and affordable solution. Cell Calcium 37:565–572. doi:10.1016/j.ceca.2005.03.001

Brandman O, Ferrell JE, Li R, Meyer T (2005) Interlinked fast and slow positive feedback loops drive reliable cell decisions. Science 310:496–498. doi:10.1126/science.1113834

Brown EB, Shear JB, Adams SR, Tsien RY, Webb WW (1999) Photolysis of caged calcium in femtoliter volumes using two-photon excitation. Biophys J 76:489–499. doi:10.1016/S0006-3495(99)77217-6

Bullinger E (2005) System analysis of a programmed cell death model. In: Proceedings of the IEEE conference on decision and control. Seville, Spain, pp 7994–7999

Caines PE, Chan CW (1975) Feedback between stationary stochastic processes. IEEE Trans AC 20: 498–508

Cho K-H, Shin S-Y, Lee H-W, Wolkenhauer O (2003) Investigations into the analysis and modeling of the TNFα mediated NF-κB signaling pathway. Genome Res 13(11):2413–2422. doi:10.1101/gr.1195703

Corrie JET, Katayama Y, Reid GP, Anson M, Trentham DR (1992) The development and application of photosensitive caged compounds to aid time-resolved structure determination of macromolecules. Philos Trans R Soc London Ser A 340:233–243

Dayal BS, MacGregor JF (1996) Identification of finite impulse response models: methods and robustness issues. Ind Eng Chem Res 35:4078–4090. doi:10.1021/ie960180e

Dong CY, Cho K-H, Yoon TW (2008) Identification of intra-cellular feedback loops by intermittent step perturbation method. In: Proceedings of the 17th IFAC world congress (IFAC2008), Seoul, Korea, pp 289–294

Doré S, Kearney RE, De Guise JA (1997) Experimental correlation-based identification of X-ray CT point spread function. Part 1: method and experimental results. Med Biol Eng Comput 35:0118–0140
Draper DR, Smith H (1981) Applied regression analysis, 2nd edn. Wiley, New York
Durbin J (1960) The fitting of time series models. Rev Int Stat Inst 23:233–244. doi:10.2307/1401322
Eisen H, Brachet P, da Silva PL, Jacob F (1967) Regulation of repressor inhibition in lambda. Proc Natl Acad Sci USA 66:855–862. doi:10.1073/pnas.66.3.855
Eissing T, Allgower F, Bullinger E (2005) Robustness properties of apoptosis models with respect to parameter variations and intrinsic noise. IEE Proc Sys Biol 152:221–228. doi:10.1049/ip-syb:20050046
Eissing T, Conzelmann H, Gilles ED, Allgower F, Bullinger E, Scheurich P (2004) Bistability analyses of a caspase activation model for receptor-induced apoptosis. J Biol Chem 279:36892–36897. doi:10.1074/jbc.M404893200
Fang CZ, Xiao DY (1988) System identification. Tsinghua University Press, Beijing
Gauss KF (1963) Theoria motus corporum celestium. english translation: theory of the motion of the heavenly bodies. Dover, New York
Geladi P, Kowalski BR (1986) Partial least-squares regression: a tutorial. Anal Chim Acta 185:1–17. doi:10.1016/0003-2670(86)80028-9
Giovannardi S, Lando L, Peres A (1998) Flash photolysis of caged compounds: casting light on physiological processes. News Physiol Sci 13:251–255
Glendinning P (1994) Stability, instability and chaos: an introduction to the theory of nonlinear differential equations. Cambridge University Press, Cambridge
Godfrey KR (1980) Correlation methods. Automatica 16:527–534. doi:10.1016/0005-1098(80)90076-X
Golub G, van Loan C (1996) Matrix computations. The Johns Hopkins University Press, London
Gouze J (1998) Positive and negative circuits in dynamical systems. J Biol Syst 6:11–15. doi:10.1142/S0218339098000054
Granger CWJ (1962) Economic processes involving feedback. Princeton University, New Jersey
Guldberg CM, Waage P (1879) Über die chemische Affinität. J Prakt Chem 19:69. doi:10.1002/prac.18790190111
Haass C (1999) Apoptosis: dead end for neurodegeneration. Nature 399:204–207. doi:10.1038/20314
Hoerl AE, Kennard RW (1970a) Ridge regression: biased estimation for nonorthogonal problems. Technometrics 42:80–86. doi:10.2307/1271436
Hoerl AE, Kennard RW (1970b) Ridge regression: applications to nonorthogonal problems. Technometrics 12:69–82. doi:10.2307/1267352
Hull TE, Dobell AR (1962) Random number generater. Soc Ind Appl Math Rev 4:230–254
Hunter IW, Kearney RE (1983) Two-sided linear filter identification. Med Biol Eng Comput 21:203–209. doi:10.1007/BF02441539
Kim D, Kwon Y-K, Cho K-H (2007a) Coupled positive and negative feedback circuits form an essential building block of cellular signaling pathways. Bioessays 29(1):85–90. doi:10.1002/bies.20511
Kim D, Rath O, Kolch W, Cho K-H (2007b) A hidden oncogenic positive feedback loop caused by crosstalk between Wnt and ERK pathways. Oncogene 26(31):4571–4579. doi:10.1038/sj.onc.1210230
Kim J-R, Yoon Y, Cho K-H (2008) Coupled feedback loops form dynamic motifs of cellular networks. Biophys J 94(2):359–365. doi:10.1529/biophysj.107.105106
Klipp E, Herwig R, Kowald A, Wierling C, Lehrach H (2005) Systems biology in practice: concepts, implementation and application. Wiley-VCh Verlag GmbH & Co. KGaA, Berlin
Kwon Y-K, Cho K-H (2008a) Quantitative analysis of robustness and fragility in biological networks based on feedback dynamics. Bioinformatics 24(7):987–994. doi:10.1093/bioinformatics/btn060
Kwon Y-K, Cho K-H (2008b) Coherent coupling of feedback loops: A design principle of cell signaling networks. Bioinformatics 24(17):1926–1932. doi:10.1093/bioinformatics/btn337
Laub MT, Loomis WF (1998) A molecular network that produces spontaneous oscillations in excitable cells of dictyostelium. Mol Biol Cell 9:3521–3532
Levinson N (1947) The Wiener RMS error criterion in filter design and prediction. J Math Phys 25:261–278
Ljung L (1987) System identification: theory for the user. Prentice Hall, New Jersey
Ludwig R, Ehrhardt A (1995) Turn-key-ready wavelength-, repetition rate-and pulse width-tunable femtosecond hybrid modelocked semiconductor laser. Electron Lett 31:1165–1167. doi:10.1049/el:19950796
Maeda M, Lu S, Shaulsky G, Miyazaki Y, Kuwayama H, Tanaka Y, Kuspa A, Loomis WF (2004) Periodic signaling controlled by an oscillatory circuit that includes protein kinase ERK2 and PKA. Science 304:875–878. doi:10.1126/science.1094647

McClung FJ, Hellwarth RW (2004) Giant optical pulsations from ruby. J Appl Phys 33:828–829. doi:10.1063/1.1777174

Mialocq JC, Amouyal E, Bernas A, Grand D (1982) Picosecond laser photolysis of aqueous indole and tryptophan. J Phys Chem 86:3173–3177. doi:10.1021/j100213a022

Miao B, Zane R, Maksimovic D (2005) System identification of power converters with digital control through cross-correlation methods. IEEE Trans. Power Electron 20(5):1093–1099. doi:10.1109/TPEL.2005.854035

Miyasaka H, Nagata T, Kiri M, Mataga N (1992) Femtosecond-picosecond laser photolysis studies on reduction process of excited benzophenone with N-methyldiphenylamine in acetonitrile solution. J Phys Chem 96:8060–8065. doi:10.1021/j100199a042

Nerbonne JM (1986) Design and application of photolabile intracellular probes. Soc Gen Physiol Ser 40:417–445

Okamura T, Sancar A, Heelis PF, Begley TP, Hirata Y, Mataga N (1991) Picosecond laser photolysis studies on the photorepair of pyrimidine dimers by DNA photolyase. 1. Laser photolysis of photolyase-2-deoxyuridine dinucleotide photodimer complex. J Am Chem Soc 113:3143–3145. doi:10.1021/ja00008a050

Oppenheim AV, Willsky AS, Nawab SH (1997) Signals & systems. Prentice-Hall, Englewood Cliffs

Porta A, Furlan R, Rimoldi O, Pagani M, Malliani A, van de Borne P (2002) Quantifying the strength of the linear causal coupling in closed loop interacting cardiovascular variability signals. Biol Cybern 86:241–251. doi:10.1007/s00422-001-0292-z

Rabiner L, Crochiere R, Allen J (1978) FIR system modeling and identification in the presence of noise and with band-limited inputs, acoustics, speech, and signal processing. IEEE Trans Signal Process 26:319–333. doi:10.1109/TASSP.1978.1163113

Rao CR (1973) Linear statistical inference and its applications. Wiley, New York

Rapp G, Güth K (1988) A low cost high intensity flash device for photolysis experiments. Pflugers Arch Eur J Physiol 411:200–203. doi:10.1007/BF00582315

Schnider SM, Kwong RH, Lenz FA, Kwan HC (1989) Detection of feedback in the central nervous system using system identification techniques. Biol Cybern 60:203–212. doi:10.1007/BF00207288

Sension RJ, Repinec ST, Szarka AZ, Hochstrasser RM (1993) Femtosecond laser studies of the cis-stilbene photoisomerization reactions. J Chem Phys 98:6291–6315. doi:10.1063/1.464824

Shin S-Y, Rath O, Choo S-M, Fee F, McFerran, Kolch W, Cho K-H (2009) Positive and negative feedback regulations coordinate the dynamic behavior of the Ras/Raf/MEK/ERK signal transduction pathway. J Cell Sci 122(3):425–435. doi:10.1242/jcs.036319

Strogatz SH (2000a) From kuramoto to crawford: exploring the onset of synchronization in populations of coupled oscillations. Physica D 143:1–20. doi:10.1016/S0167-2789(00)00094-4

Strogatz SH (2000b) Nonlinear dynamics and chaos: with applications to physics, biology, chemistry and engineering. Perseus Publishing, Reading

Szallasi Z, Stelling J, Periwal V (2005) System modeling in cell biology from concepts to nuts and bolts. The MIT Press, Cambridge

Thomas R (1981) On the relation between the logical structure of systems and their ability to generate multiple steady states or sustained oscillation. Springer Ser Synergetics 9:180–193

Thompson CB (1995) Apoptosis in the pathogenesis and treatment of disease. Science 267:456–1462. doi:10.1126/science.7878464

Thomas R, Kaufman M (2001) Multistationarity, the basis of cell differentiation and memory. I. structural conditions of multistationarity and other nontrivial behavior. Chaos 11:170–179. doi:10.1063/1.1350439

Tufillaro NB, Abbot T, Reilly J (1992) An experimental approach to nonlinear dynamics and chaos. Perseus Publishing, Chambridge

Vance W, Arkin A, Ross J (2002) Determination of causal connectivities of biomolecular species in reaction networks. Proc Natl Acad Sci USA 99:5816–5821. doi:10.1073/pnas.022049699

Westwick DT, Kearney RE (1997) Identification of physiological systems: a robust method for non-parametric impulse response estimation. Med Biol Eng Comput 35:83–90. doi:10.1007/BF02534135

Wolkenhauer O, Kitano H, Cho K-H (2003) Systems biology: Looking at opportunities and challenges in applying systems theory to molecular and cell biology. IEEE Contr Syst Mag 23(4):38–48. doi:10.1109/MCS.2003.1213602

Wolpert L, Lewis JH (1975) Towards a theory of development. Fed Proc 34:14–20