Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Theoretical Biology 232 (2005) 427–441
Journal of Theoretical Biology
www.elsevier.com/locate/yjtbi

Inference of signaling and gene regulatory networks by steady-state perturbation experiments: structure and accuracy

Michael Andrec[a,b,*], Boris N. Kholodenko[c], Ronald M. Levy[a,b], Eduardo Sontag[d,b]

[a]*Department of Chemistry and Chemical Biology, Rutgers, The State University of New Jersey, Piscataway, NJ 08854-8087, USA*
[b]*BioMaPS Institute for Quantitative Biology, Rutgers, The State University of New Jersey, Piscataway, NJ 08854-8087, USA*
[c]*Department of Pathology, Anatomy and Cell Biology, Thomas Jefferson University, Philadelphia, PA 19107, USA*
[d]*Department of Mathematics, Rutgers, The State University of New Jersey, Piscataway, NJ 08854-8087, USA*

Received 15 April 2004; received in revised form 19 August 2004; accepted 24 August 2004
Available online 8 October 2004

Abstract

One of the fundamental problems of cell biology is the understanding of complex regulatory networks. Such networks are ubiquitous in cells and knowledge of their properties is essential for the understanding of cellular behavior. In earlier work (Kholodenko et al. (PNAS 99: 12841), it was shown how the structure of biological networks can be quantified from experimental measurements of steady-state concentrations of key intermediates as a result of perturbations using a simple algorithm called "unravelling". Here, we study the effect of experimental uncertainty on the accuracy of the inferred structure (i.e. whether interactions are excitatory or inhibitory) of the networks determined using the unravelling algorithm. We show that the accuracy of the network structure depends not only on the noise level but on the strength of the interactions within the network. In particular, both very small and very large values of the connection strengths lead to large uncertainty in the inferred network. We describe a powerful geometric tool for the intuitive understanding of the effect of experimental error on the qualitative accuracy of the inferred network. In addition, we show that the use of additional data beyond that needed to minimally constrain the network not only improves the accuracy of the inferred network, but also may allow the detection of situations in which the initial assumptions of unravelling with respect to the network and the perturbations have been violated. Our ideas are illustrated using the mitogen-activated protein kinase (MAPK) signaling network as an example.

Keywords: Reverse engineering; Unravelling; Systems biology; Black box; Module

1. Introduction

One of the fundamental problems of cell biology is to infer and quantify interconnections in complex regulatory networks. Such networks are ubiquitous in cells and critical for their function and knowledge of their properties is essential for the understanding of cellular behavior. These regulatory networks include signaling cascades such as mitogen-activated protein kinase (MAPK) systems (Chang and Karin, 2001; Hazzalin and Mahadevan, 2002; Huang and Ferrell, 1996; Kholodenko, 2000; Widmann et al., 1999) and other networks which transmit information about the external environment to the inside of the cell (Brivanlou and Darnell, 2002). In addition, the networks created by the mutual activation and repression of the expression of genes by each other's products play a critical role in the development of multicellular organisms and other important cellular processes (Brazhnik et al., 2002).

Such networks can be studied experimentally and theoretically using either a "bottom-up" or a "top-down" strategy. The bottom-up approach relies on the numerical simulation of a mechanistic model of the

*Corresponding author. Department of Chemistry, Rutgers University, 610 Taylor Road, Piscataway, NJ 08854-8087, USA. Tel.: +1 732 445 5159; fax: +1 732 445 5958.
E-mail address: andrec@lutece.rutgers.edu (M. Andrec).

0022-5193/$ - see front matter © 2004 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jtbi.2004.08.022

network and quantitatively comparing the results of such simulations to experimental observations (Hasty et al., 2001; Kholodenko et al., 1999; Moehren et al., 2002; Shvartsman et al., 2002; Smolen et al., 1998; Tyson et al., 2001; von Dassow et al., 2000). While such a strategy can provide a wealth of detail, it by necessity requires the modeling of a large number of molecular processes, some of which may in fact be unknown. Therefore, the bottom-up approach is less applicable to poorly characterized networks, since it will by its very nature miss interactions and regulatory feedbacks that still await discovery. For such "network discovery" purposes, a top-down approach is more appropriate (Brand, 1996; de la Fuente et al., 2002; Kholodenko et al., 1997, 2002; Stark et al., 2003). Here, the architecture of the network is inferred (or "reverse engineered") based on observed global responses of the system to a series of signals or experimental perturbations, such as hormones, growth factors,neurotransmitters, or experimental interventions (such as chemical inhibitors). The global responses that could be measured include changes in the phosphorylation states or activities of proteins, mRNA levels, or transcription rates. From these global responses, top-down analysis methods attempt to recover the local interactions between components of regulatory networks, which in turn form a "map" of the network. The top-down approach has the great advantage that in principle it can be applied to regulatory networks of arbitrary complexity.

The top-down approach achieves this scaling ability by making extensive use of the concept of a network module. The concept of modules in biology has recently received much attention as part of the focus on the fundamentally multi-component aspects of molecular biological function known as "systems biology" (Hartwell et al., 1999; Lauffenburger, 2000). In this paper, we will use the term "module" in the specific technical sense used in the previous work of Kholodenko et al. (2002), namely, a group of genes, enzymes, or other cellular components of arbitrary complexity connected by physical and/or chemical interactions which together perform one or more identifiable tasks. These tasks usually depend on one or a small number of components of the module (which we call "communicating intermediates") which alone interact with components of other modules. In a top-down approach, the network modules are considered to be "black boxes", and only the concentrations or chemical states of the communicating intermediates will be used in the analysis without explicit reference to any internal structure that the modules may have.

As an example, consider the MAPK signaling network shown in Fig. 1. It is clear that in this model only the doubly phosphorylated forms of the kinases have any interactions with components on other "levels" of the network, while the unphosphorylated and singly

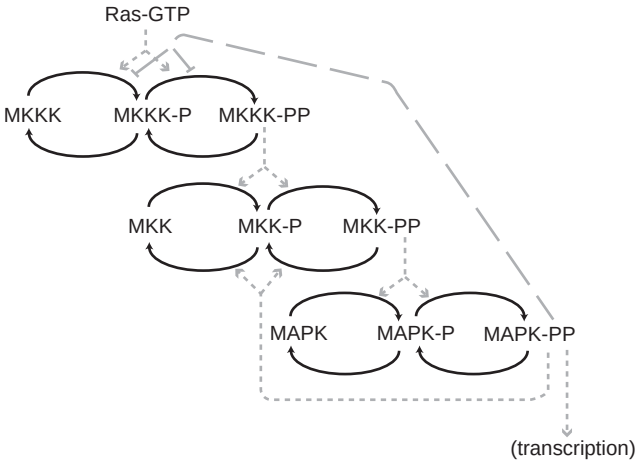

Fig. 1. A prototypical example of a MAPK cascade incorporating negative feedback (Kholodenko, 2000). Dotted lines indicate activation and long-dashed lines indicate repression. Note that MAPK-PP activates the *degradation* of MKK-PP. For purposes of generating synthetic data, the detailed kinetic model and parameters given in Kholodenko et al. (2002) were used and are provided as electronic supplementary material. All simulated data based on the MAPK system was generated by setting each rate equation equal to zero and solving for the concentrations of all of the species. Perturbations were performed by modifying the kinetic parameters by the fractions indicated.

phosphorylated forms only interact with components on the same "level", e.g. MKKK-P only interacts with MKKK and MKKK-PP, and not any of the forms of MKK or MAPK. Therefore, although there are a total of nine components in this prototypical network, we can treat it from a modular point of view as consisting of three modules (corresponding to the "MKKK", "MKK", and "MAPK" levels), which are represented by the communicating intermediates MKKK-PP, MKK-PP, and MAPK-PP, respectively, thereby reducing the effective dimensionality of the problem by a factor of three. For networks having modules of more complex internal structure the savings in terms of dimensionality reduction would be correspondingly greater. Thus, by introducing the concept of the module, one can gain many of the advantages of more phenomenological network modeling approaches such as Bayesian or Boolean networks (de Jong, 2002) while still retaining a basis in chemical kinetics.

In earlier work (Kholodenko et al., 2002; Kholodenko and Sontag, 2002), it was shown how interactions between modules can be quantified from the concentrations of the communicating intermediates following perturbations of each module using a simple algorithm which we call "unravelling". In contrast to other approaches (Gardner et al., 2003; Ronen et al., 2002), unravelling only requires steady-state concentrations, thereby obviating the need for the quantitation of rates, and assumes nothing regarding the functional form of the rate equations beyond the assumption that the

perturbations used do not directly effect a given module. Since unravelling depends on measuring changes in steady-state concentration of key intermediates in response to perturbations, it would appear that it would not be applicable to systems which exhibit "total adaptation", i.e. systems which eventually return to their original steady state after a perturbation. However, even such systems exhibit pseudo-steady-state behavior which differs from the original steady state, and which would allow the measurement of concentrations and the determination of the network topology.

A key consideration in any reverse engineering method is its sensitivity to experimental error. In this paper, we investigate in detail how noise effects our ability to correctly reconstruct the interactions between modules and how statistical methods could be used to determine the validity of the model used for the reconstruction. In particular, we will emphasize the qualitative accuracy of the inferred network, i.e. whether the identification of the pattern of interactions between modules as activating or repressing is correct. We will demonstrate how the accuracy depends on the degree of redundancy of the data and analyse how the accuracy improves with increasing amounts of data.

2. Methods

2.1. General aspects of the unravelling formalism

In this section, we review the fundamentals of the unravelling methodology, and discuss some of the practical aspects of its implementation. Let us assume that we have a network which obeys deterministic kinetics given by the system of L differential equations

$$\frac{\mathrm{d}z_i}{\mathrm{d}t} = g_i(z_1,\ldots,z_L,\boldsymbol{p}), \tag{1}$$

where z_i $(i=1,\ldots,L)$ are the concentrations of each chemical species in the network and $\boldsymbol{p}$ is a vector of kinetic parameters (rate constants, Michaelis constants, etc.). If we only consider the steady-state behavior of the system, then the concentrations are determined by a system of L algebraic equations

$$g_i(z_1,\ldots,z_L,\boldsymbol{p}) = 0. \tag{2}$$

Let us suppose that the L chemical components are organized into $N<L$ modules, such that components within a given module interact only with each other with the exception of a single communicating intermediate. In that case, the variables z_i may be partitioned into two disjoint sets: those representing the communicating intermediates, and those representing species "internal" to their module. The kinetic system given by Eq. (1) can then be rewritten in the form

$$\begin{aligned}
\frac{\mathrm{d}x_i}{\mathrm{d}t} &= g_{0,j}(\boldsymbol{y}_j,\boldsymbol{x},\boldsymbol{p}),\\
\frac{\mathrm{d}y_{1,j}}{\mathrm{d}t} &= g_{1,j}(\boldsymbol{y}_j,\boldsymbol{x},\boldsymbol{p}), \qquad j=1,\ldots,N,\\
&\vdots\\
\frac{\mathrm{d}y_{l_j,j}}{\mathrm{d}t} &= g_{l_j,j}(\boldsymbol{y}_j,\boldsymbol{x},\boldsymbol{p}),
\end{aligned} \tag{3}$$

where x_j represents the concentration of the communicating intermediate of module j, the $y_{i,j}$ variables represent the concentrations of the l_j species which are "internal" to the jth module, and

$$\boldsymbol{y}_j = (y_{1,j},\ldots,y_{l_j,j}),$$

$$\boldsymbol{x} = (x_1,\ldots,x_N). \tag{4}$$

If we assume that the Jacobian of the functions $g_{i,j}$ with respect to $y_{i,j}$ is non-singular when evaluated at the steady state corresponding to $\boldsymbol{p}$, then the "intramodular" variables $\boldsymbol{y}_j$ can be viewed as implicit functions $\boldsymbol{h}_j(\boldsymbol{x},\boldsymbol{p})$ of the communicating intermediates $\boldsymbol{x}$ and parameters $\boldsymbol{p}$ (Kholodenko and Sontag, 2002). If we define

$$f_j(\boldsymbol{x},\boldsymbol{p}) = g_{0,j}(\boldsymbol{h}_j(\boldsymbol{x},\boldsymbol{p}),\boldsymbol{x},\boldsymbol{p}), \tag{5}$$

then the steady-state behavior of the communicating intermediates can be described by the reduced set of $N<L$ algebraic equations

$$f_j(\boldsymbol{x},\boldsymbol{p}) = 0. \tag{6}$$

This top-down analysis "black-boxes" the modular organization of the molecular network, explicitly considering communicating intermediates only. If the number of modules is small compared to the total number of species in the system, then Eq. (6) represents a considerable simplification over the full problem given in Eq. (2). This formalism is easily generalized to cases where modules have multiple communicating intermediates, as is shown in Kholodenko et al. (2002).

The direct effect of module j on module i at the steady state corresponding to $\boldsymbol{p}$ can be defined as $\partial f_i/\partial x_j$. However, this definition is ambiguous, since a multiplicative scaling of the rate equations can scale $\partial f_i/\partial x_j$ while leaving the steady states unchanged. Therefore, we can infer the derivatives $\partial f_i/\partial x_j$ only up to a constant factor (Kholodenko et al., 2002; Kholodenko and Sontag, 2002). In view of this, we will describe the interaction between modules in terms of the "connection coefficients" r_{ij}, which are defined as

$$r_{ij} \equiv -\left(\frac{\partial f_i}{\partial x_j}\right)\Big/\left(\frac{\partial f_i}{\partial x_i}\right) \tag{7}$$

and which are obviously invariant with respect to multiplicative scaling of the rate equations. The connection coefficients r_{ij} are the elements of the

Jacobian matrix "normalized" by its diagonal elements. In biological terms, the connection coefficient r_{ij} tells us how much x_i will change in response to a causative change in x_j, when all other modules are kept constant, while module i is relaxing to its new steady state. Conceptually, we can imagine "disconnecting" module i from the network, and then "perturbing" the concentration x_j while observing the resulting change in x_i after a new "steady state" local to module i has been achieved. The ratio of these changes in the limit of small "perturbation" is simply the derivative of the steady-state condition $f_i(\boldsymbol{x},\boldsymbol{p})$ with respect to x_j at constant x_k ($k \neq i$ or j), which equals r_{ij}. If the modules are stable (i.e. $\partial f_i/\partial x_i < 0$), then $r_{ij} > 0$ corresponds to the case where module j activates module i, $r_{ij} < 0$ corresponds to the repression of module i by module j, and $r_{ij} = 0$ corresponds to module j having no effect on module i.

2.2. Analysis of ideal data

Suppose that we have N parameters $p_1, \ldots, p_N$ (a subset of the components of vector $\boldsymbol{p}$) such that p_i is known a priori to have a direct effect only on module i. The change in the steady-state concentration of each communicating intermediate to an infinitesimal perturbation of p_i defines the global response of the system with respect to that perturbation

$$R_{ij} \equiv \frac{\partial x_i}{\partial p_j}. \tag{8}$$

Unlike the local connection coefficients r_{ij}, the global response R_{ij} can be measured experimentally by means of a finite difference approximation. In practice, the quantity R_{ij} is usually defined as the logarithm of the ratio of the steady-state concentration of communicating intermediate i after perturbation j to its pre-perturbation concentration (Kholodenko et al., 1997, 2002). However, all of the key results below are true for the derivative of the logarithm as well, which can be interpreted as a fractional change of the steady-state concentration. Since we have chosen p_j such that its only direct effect is on module j, it follows that $\partial f_i/\partial p_j = 0$ for any $i \neq j$. However, we can also use the chain rule to write

$$\frac{\partial f_i}{\partial p_j} = \sum_k \frac{\partial f_i}{\partial x_k}\frac{\partial x_k}{\partial p_j} = -\frac{\partial f_i}{\partial x_i}\sum_k r_{ik}R_{kj} \tag{9}$$

for each pair of modules $i \neq j$. Provided that $\partial f_i/\partial x_i \neq 0$, this leads to

$$\sum_k r_{ik}R_{kj} = 0. \tag{10}$$

Suppose for the moment that we are only interested in the vector $\boldsymbol{r}_i = (r_{i1}, r_{i2}, \ldots, r_{iN})^{\mathrm{T}}$, i.e. the direct effect of each of the modules $1 \ldots N$ on one particular module i, and we define the "data vector" $\boldsymbol{R}_j = (R_{1j}, R_{2j}, \ldots, R_{Nj})$ as the response of all modules to a perturbation of parameter p_j effecting only module j. Then the $N-1$ data vectors $\boldsymbol{R}_j$ ($j \neq i$) are all perpendicular to the vector $\boldsymbol{r}_i$, and define an $N-1$-dimensional hyperplane in $\Re^N$ (the "data plane") having a normal vector $\boldsymbol{r}_i$ (provided that the data vectors are linearly independent). The connection coefficients $\boldsymbol{r}_i$ can be determined by recognizing that $r_{ii} = -1$ and that Eq. (10) can be rewritten as the linear system

$$\sum_k r_{ik}R_{kj}(1-\delta_{ik}) = R_{ij} \quad (j \neq i), \tag{11}$$

which can be solved by matrix inversion. Eq. (11) represents a case where there is just sufficient data to determine $\boldsymbol{r}_i$, i.e. the number of "parameters" exactly equals the number of "data", and therefore we can always find a vector $\boldsymbol{r}_i$ which gives an exact fit, i.e. strict equality in Eq. (11).

The reverse engineering strategy described above can be easily extended to the overdetermined case where more than $N-1$ perturbations are performed (Kholodenko and Sontag, 2002). Suppose that we perform $M > N-1$ perturbations which do not have any direct effect on module i and measure the corresponding responses $\boldsymbol{R}_j$ ($j = 1, \ldots, M$) (note that the subscript now does not refer to a module, but to the "perturbation number"). These perturbations may be replicate measurements or could correspond to perturbations which simultaneously effect several modules other than i. Based on the above logic, the resulting M vectors $\boldsymbol{R}_j$ must all be perpendicular to the vector $\boldsymbol{r}_i$:

$$\sum_k r_{ik}R_{kj} = 0 \quad (j = 1, \ldots, M). \tag{12}$$

Therefore, in the ideal case all of the $\boldsymbol{R}_j$ vectors will lie in an $N-1$ dimensional hyperplane and therefore the matrix $\boldsymbol{R}$ (defined as the matrix with row vectors $\boldsymbol{R}_j$) must have rank $N-1$, provided that the associated perturbations do not directly effect module i and that all modules $j \neq i$ are perturbed either singly or in combination. If the latter is not the case, then the matrix $\boldsymbol{R}$ will have a rank less than $N-1$.

2.3. Analysis of noise corrupted data

The above results are strictly true only for ideal data. In practice, the data will not be ideal due to numerical error arising from the use of a finite-difference approximation to the derivative defining R_{ij}, as well as due to experimental measurement error. For the purposes of this paper, we will assume that any "systematic error" arising from numerical approximation is negligible compared to the size of the measurement error. In fact, any systematic error will tend to zero with decreasing perturbation size, whereas the relative

amplitude of the noise will only increase. If we neglect systematic error, then we can view the observed data $\boldsymbol{R}$ as consisting of the ideal data plus a noise matrix $\boldsymbol{\varepsilon}$, the elements of which are random variables. Let us first consider the "exactly determined" case corresponding to Eq. (11). Since the components of the data vectors $\boldsymbol{R}_i$ are random variables, the $N-1$-dimensional hyperplane which they define is now also random, as is its normal vector and consequently the vector $\boldsymbol{r}_i$. However, in all cases the fit will remain "exact", in that equality in Eq. (11) will always be achieved.

In the overdetermined case, it was shown above that the ideal data matrix remains rank $N-1$ regardless of the number of perturbations made. In the presence of noise, however, the data matrix $\boldsymbol{R}$ becomes full rank, since the noise matrix $\boldsymbol{\varepsilon}$ is full rank. Therefore, the noisy data vectors $\boldsymbol{R}_i$ will not lie exactly in an $N-1$ dimensional hyperplane, and therefore we will not be able to find a vector of connection coefficients which results in strict equality in Eq. (12). However, if the magnitude of the noise is not large, we expect that Eq. (12) will be "almost" satisfied, and we would like to find an approximate solution which minimizes the error in Eq. (12) according to some measure. Since Eq. (12) can be rewritten as a linear system of the form $\boldsymbol{A}\boldsymbol{x} \approx \boldsymbol{b}$ in the same way as Eq. (10) can be rewritten in the form of Eq. (11), one is lead to a "least-squares" solution for the connection coefficients. However, the standard least-squares solution $\boldsymbol{x}_{\mathrm{LS}} = (\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-1}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{b}$ which minimizes the norm of $\boldsymbol{A}\boldsymbol{x}-\boldsymbol{b}$ is not the best solution of our problem, because it implicitly assumes that the uncertainty is entirely in $\boldsymbol{b}$ while $\boldsymbol{A}$ is known perfectly. This can be seen more easily if one recognizes that finding $\boldsymbol{x}_{\mathrm{LS}}$ is equivalent to finding a "perturbed" vector $\boldsymbol{b}'$ which gives an exact equality $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}'$ (i.e. $\boldsymbol{b}'$ is in the column-space of $\boldsymbol{A}$) and which also minimizes the norm of $(\boldsymbol{b}-\boldsymbol{b}')$, and then finding the value $\boldsymbol{x}_{\mathrm{LS}}$ which solves $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}'$ (van Huffel and Vandewalle, 1991). Therefore, the standard least-squares solution does not treat $\boldsymbol{A}$ and $\boldsymbol{b}$ equivalently. However, in our case the elements of $\boldsymbol{A}$ and $\boldsymbol{b}$ are all experimental measurements which are equally corrupted by noise. Therefore, it would make more sense for us to solve for $\boldsymbol{x}$ in a way that allows for "perturbation" of both $\boldsymbol{A}$ *and* $\boldsymbol{b}$.

A solution to the linear system $\boldsymbol{A}\boldsymbol{x} \approx \boldsymbol{b}$ which treats $\boldsymbol{A}$ and $\boldsymbol{b}$ equivalently is known in the numerical analysis and engineering literatures as "total least squares" (TLS) (Golub and van Loan, 1980; van Huffel and Vandewalle, 1991). Specifically, TLS attempts to find a perturbed matrix $\boldsymbol{A}'$ and a perturbed vector $\boldsymbol{b}'$ such that $\boldsymbol{A}'\boldsymbol{x} = \boldsymbol{b}'$ gives exact equality ($\boldsymbol{b}'$ is in the column-space of $\boldsymbol{A}'$) subject to the constraint that the norm of $[\boldsymbol{A}' : \boldsymbol{b}'] - [\boldsymbol{A} : \boldsymbol{b}]$ is minimized (where $[\boldsymbol{A} : \boldsymbol{b}]$ represents the matrix formed by appending the column-vector $\boldsymbol{b}$ to the right side of matrix $\boldsymbol{A}$). The solution of $\boldsymbol{A}'\boldsymbol{x} = \boldsymbol{b}'$ is known as the TLS solution $\boldsymbol{x}_{\mathrm{TLS}}$.Alternatively, we can adopt a more statistical view of Eq. (12) to obtain the same result. In particular, if the elements of the noise matrix $\boldsymbol{\varepsilon}$ are independent and identically distributed normal (Gaussian) random variables, then the values of the connection coefficients which maximize the likelihood of the data are those which minimize the square deviation to the data. In other words, we must find the hyperplane which best fits the $\boldsymbol{R}_j$ in a least-squares sense, or equivalently, the "perturbed" matrix $\boldsymbol{R}'$ of rank $N-1$ such that the norm of the difference $\boldsymbol{R}-\boldsymbol{R}'$ is minimized. The resulting solution can be seen to be identical to $\boldsymbol{x}_{\mathrm{TLS}}$, since the "augmented" matrix $[\boldsymbol{A} : \boldsymbol{b}]$ is identical to $\boldsymbol{R}$ (van Huffel and Vandewalle, 1991).

The TLS solution is readily obtained using singular value decomposition (SVD) (Golub and van Loan, 1989; Press et al., 1992), which is defined for an $M \times N$ matrix $\boldsymbol{A}$ as an $(M \times N)$ column-orthonormal matrix $\boldsymbol{U}$, an $(N \times N)$ orthonormal matrix $\boldsymbol{V}$, and a non-negative $(N \times N)$ diagonal matrix $\boldsymbol{\Sigma}$ such that $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^{\mathrm{T}}$. The diagonal elements σ_i ($\sigma_1 \geqslant \sigma_2 \geqslant \cdots \geqslant \sigma_N$) of $\boldsymbol{\Sigma}$ are known as the singular values of $\boldsymbol{A}$, and the number of non-zero singular values equals the rank of $\boldsymbol{A}$. The column-vectors $\boldsymbol{u}_i$ of $\boldsymbol{U}$ and the column-vectors $\boldsymbol{v}_i$ of $\boldsymbol{V}$ are known as the left and right singular vectors of $\boldsymbol{A}$, respectively. According to the Eckart–Young Matrix Approximation Theorem (Eckart and Young, 1936; van Huffel and Vandewalle, 1991), the matrix $\boldsymbol{R}'$ of rank $N-1$ which best approximates the rank N matrix $\boldsymbol{R}$ in the least-squares sense is $\boldsymbol{R}' = \boldsymbol{U}\boldsymbol{\Sigma}'\boldsymbol{V}^{\mathrm{T}}$, where the matrices $\boldsymbol{U}$, $\boldsymbol{\Sigma}$, and $\boldsymbol{V}$ are given by the SVD, and $\boldsymbol{\Sigma}'$ is the matrix $\boldsymbol{\Sigma}$ with the smallest singular value set to zero. The best-fit hyperplane is defined by its normal vector, which is simply the right singular vector corresponding to the smallest singular value σ_N, and σ_N itself is equal to the norm of $\boldsymbol{R}-\boldsymbol{R}'$ (van Huffel and Vandewalle, 1991). Therefore, in the context of unravelling in the presence of experimental error, the connection coefficients can be found by performing the SVD of the matrix of global responses, and scaling the Nth right singular vector such that $r_{ii} = -1$ (note that the N such normalizations are needed for the determination of the complete network, one for each linear system given in Eq. (12)).

The straightforward least-squares solution in terms of the SVD of the data matrix given above assumes that all elements of the noise matrix $\boldsymbol{\varepsilon}$ are independent and identically distributed normal random variables. If instead the elements of $\boldsymbol{\varepsilon}$ are independent but have different variances, then one must find a weighted total least-squares solution. If the variance of noise matrix element $\boldsymbol{\varepsilon}_{ij}$ can be written in the form $a_i b_j$, then it is still possible to find a solution using the ordinary SVD algorithm (Golub and van Loan, 1980). Generalizations of the SVD or other non-linear methods are required when variances are independent but proportional to the magnitudes of the measurements (de Moor, 1993) or do

not follow any pattern (Gerhold, 1969; Premoli and Rastello, 2002). Problems where the elements of $\boldsymbol{\varepsilon}$ are not independent or are non-normal also require alternative algorithms. In this paper, we consider the case where the errors are normal, independent, and identically distributed.

The computational results described below were all obtained by Monte Carlo using simulated data. Ideal data matrices were determined by non-linear root-finding using rate expressions from a detailed kinetic model (e.g. Fig. 1) using the FindRoot function of Mathematica (Wolfram, 1991) or by a geometric construction (see below). Normally distributed noise realizations were added to each element of the ideal data matrix to generate the noisy data matrix $\boldsymbol{R}$, and the connection coefficients were determined by solving the linear system of Eq. (11) or by performing the SVD of $\boldsymbol{R}$ (Eq. (12)) using the SSVDC subroutine from LINPACK (Dongarra et al., 1979). In order to reduce visualization problems, we will limit ourselves to three modules (e.g. Fig. 1). Many of the results given below generalize to an arbitrary number of modules. However, the visualization of the results can be quite problematic as the number of modules increases, and this is a topic of current investigation. To study the effects of noise independently of errors due to the finite difference approximation and other numerical errors arising from kinetic models, it will be useful to generate completely fictitious data matrices by geometric construction. In particular, given "true" values for the connection coefficients $\boldsymbol{r}_i$, one can constructas many vectors as desired in the plane normal to $\boldsymbol{r}_i$. These vectors can be of arbitrary "length" (corresponding to the absolute magnitudes of the measured data relative to the noise) and arbitrary orientation relative to each other (representing the degree of redundancy in the data).

Unlike ordinary least squares, the analysis of error propagation for the TLS problem is not straightforward. The few studies of related topics in the literature are focused on the "small errors" regime (Durando and Mana, 2002; Lefebvre et al., 2000), which may not be relevant in our case, since the DNA microarray and quantitative Western blot data from which we obtain estimates of the global responses are often quite noisy. In the results below, we will focus on the implications of substantial experimental uncertainty on the qualitative correctness of the inferred reverse engineered biological network. Although one could study the effect of experimental uncertainty on the connection coefficients directly, one can gain more insight if one considers the propagation of error in two stages: (1) the propagation of error in the data matrix into uncertainty in the direction of the normal vector of the data plane, and (2) the propagation of uncertainty in the normal vector direction into uncertainty in the connection coefficients r_{ij}.

Geometrically, the vector of connection coefficients $\boldsymbol{r}_i$ representing the direct effect of each module $1 \ldots N$ on a particular module i can be represented as a point on the surface of a unit sphere in $\Re^N$. To see this, consider the MAPK example of Fig. 1. Suppose that the vector of connection coefficients $\boldsymbol{r}_1$ (for the MKKK level) are given by $(-1, 0, 1.1)$. We can normalize this vector to obtain $(-0.6727, 0, 0.7399)$, which is a point on the unit sphere in $\Re^3$ and can be represented by the spherical coordinates $(\theta,\ \phi) = (42.3^\circ, 180^\circ)$. Conversely, a point on the unit sphere (corresponding, for example, to the direction of the vector normal to the data plane) can be converted into a vector of connection coefficients by scaling the vector such that the appropriate element is equal to -1. For example, the point represented by the spherical coordinates $(\theta,\ \phi) = (30^\circ, 90^\circ)$ is$(0, \frac{1}{2}, \sqrt{3}/2)$. If this point was obtained as a result of perturbations that did not directly effect module 2, then the corresponding vector of connection coefficients must have -1 as the second element. Therefore, to convert the vector $(0, \frac{1}{2}, \sqrt{3}/2)$ into connection coefficients, we must scale it such that the second element is -1, giving us $\boldsymbol{r}_2 = (0, -1, -\sqrt{3})$. Given this one-to-one mapping between points on unit spheres and vectors of connection coefficients, one can make use of the methods of probability and statistics of spherical distributions (Fisher et al., 1987) to describe the joint distribution of the connection coefficients given a (spherical) distribution of directions of the normal vector to the data plane. Alternatively, one can use this mapping as an intuitive aid in understanding the qualitative aspects of error propagation in unravelling.

3. Results and discussion

3.1. Exactly determined case

We will begin by considering the case described by Eq. (11), namely, where the connection coefficients are just determined by data. Furthermore, we will limit ourselves for the moment to the case of three modules and consider only the connection coefficients for the direct effect of modules 1 and 2 on module 3, which for notational convenience we will call r_1 and r_2. Let $\boldsymbol{R}_1$ and $\boldsymbol{R}_2$ be the vectors of global responses to two different perturbations which do not effect module 3 directly. For example, if we perturb the MAPK network shown in Fig. 1 from the reference steady state given in the caption by reducing [Ras-GTP] by 10% (perturbation of the MKKK module only), we obtain a noiseless global response vector $\boldsymbol{R}_1 = (-0.0797, -0.0613, -0.1393)$, while reduction of k_{cat5} and k_{cat6} by 15% (perturbation of the MKK module only) results in a noiseless global response $\boldsymbol{R}_2 = (0.1026, -0.0462, -0.1051)$, where modules 1, 2 and 3 correspond to the MKKK, MKK, and

MAPK levels, respectively. The vectors $\boldsymbol{R}_1$ and $\boldsymbol{R}_2$ form a plane defined by the normalvector $\boldsymbol{n} = (-2.811 \times 10^{-4}, 0.9156, -0.4021)$. To convert this normal vector to connection coefficients, we recognize that since module 3 was not directly perturbed, $\boldsymbol{n}$ must be scaled such that the third element is -1, giving us $r_1 = -0.0007$ and $r_2 = 2.2770$, in good agreement with the theoretically determined values of $r_1 = 0$ and $r_2 = 2$ (Kholodenko et al., 2002) (the small discrepancy arising from numerical approximation). Intuitively, we expect that the precision with which we can estimate r_1 and r_2 will depend on the angle formed by the two vectors $\boldsymbol{R}_1$ and $\boldsymbol{R}_2$. In particular, we expect the most favorable case to be when $\boldsymbol{R}_1$ and $\boldsymbol{R}_2$ are perpendicular to each other, and worst when they are collinear, in which case they no longer uniquely define a plane. In the above example, the angle between $\boldsymbol{R}_1$ and $\boldsymbol{R}_2$ is 69°. While this is certainly smaller than 90°, it is still fairly large. By contrast, consider an alternative perturbation scheme in which we first perturb the MAPK network by simultaneously reducing [Ras-GTP] by 10% *and* k_{cat5} and k_{cat6} by 10%, giving a noiseless global response of $\boldsymbol{R}_1 = (-0.0157, -0.0928, -0.2099)$, and then reduce only k_{cat5} and k_{cat6} by 15%, resulting in a noiseless global response $\boldsymbol{R}_2 = (0.1026, -0.0462, -0.1051)$ (as before). Although the resulting values of the connection coefficients are still correct ($r_1 = -0.0055$ and $r_2 = 2.2634$), the angle between $\boldsymbol{R}_1$ and $\boldsymbol{R}_2$ is now only 46°, which we expect will be reflected in larger uncertainties in the connection coefficients. If we repeatedly add noise to each component of $\boldsymbol{R}_1$ and $\boldsymbol{R}_2$ and find a unit vector perpendicular to the noise-perturbed data plane, we will obtain a distribution of points on the unit sphere $P(\theta, \phi)$, where θ and ϕ are spherical coordinates. Since the vectors (x, y, z) and $(-x, -y, -z)$ are both normal to the same plane, the distribution $P(\theta, \phi)$ will always possess inversion symmetry with respect to the origin. Given the one-to-one mapping between points on the sphere and the connection coefficients described above, the distribution $P(\theta, \phi)$ completely describes the uncertainty in r_1 and r_2.

In order to further strengthen our intuition about the relationship between the connection coefficients r_1 and r_2 and the data vectors $\boldsymbol{R}_1$ and $\boldsymbol{R}_2$, suppose that the "true values" of the connection coefficients are $r_1 = r_2 = 0$ (i.e. both module 1 and 2 have no direct effect on module 3). In that case, any perturbation that directly effects only module 1 or module 2 (singly or in combination) cannot have any effect on module 3 (either directly or indirectly). Therefore, the elements of the resulting noiseless vectors $\boldsymbol{R}_1$ and $\boldsymbol{R}_2$ vectors corresponding to module 3 must be zero, and consequently must lie in the x–y plane (where x, y, and z represent the modules 1, 2, and 3, respectively) with the vector $\boldsymbol{r} \equiv (r_1, r_2, -1) = (0, 0, -1)$ parallel to the z-axis. Note that if only one of r_1 and r_2 is zero, then this will not be reflected in any obvious way in $\boldsymbol{R}_1$ and $\boldsymbol{R}_2$.

3.1.1. Orthogonal data vectors

Let us assume for the moment that the noiseless $\boldsymbol{R}_1$ and $\boldsymbol{R}_2$ are both unit vectors and are perpendicular to each other, e.g. $\boldsymbol{R}_1 = (1, 0, 0)$ and $\boldsymbol{R}_2 = (0, 1, 0)$. Although this will not be the case for general experimental results (as has been seen above), this does represent a "best-case scenario", and we will consider the more general case in Section 3.1.2 below. Adding Gaussian noise to each component of $\boldsymbol{R}_1$ and $\boldsymbol{R}_2$ in a Monte Carlo fashion gives us a representation of distribution of data plane orientations $P(\theta, \phi)$, as shown in Fig. 2. Clearly, as the variance of the noise increases, the variability in the data plane orientation increases, approaching a uniform distribution over the sphere for very large noise levels. In all cases, however, the distribution remains axially symmetric about the z-axis. One obvious consequence of this axial symmetry is that the absolute orientations of the data vectors have no effect on the error propagation when the data vectors are orthogonal.

As was described above, every point in the scatter plots of Fig. 2 corresponds to one noise-corrupted realization of r_1 and r_2, since if the unit normal vector to

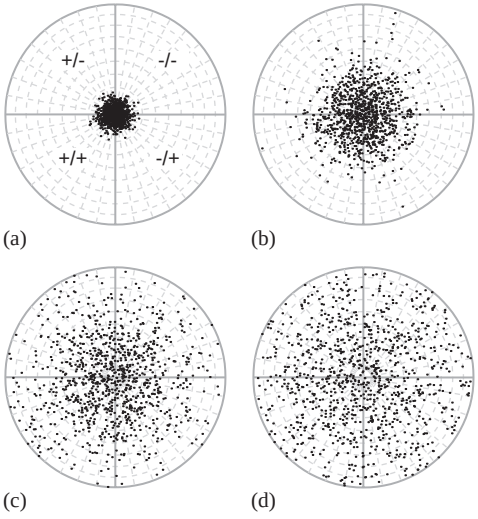

Fig. 2. Scatter plots of data plane orientations for orthogonal data vectors of length one lying in the x–y plane to which 1000 realizations of noise of standard deviation 0.1 (a), 0.3 (b), 0.5 (c), and 0.7 (d) were added. The plots are equal-area Lambert projections (Fisher et al., 1987) of the $z \geqslant 0$ hemisphere oriented such that the z-axis is at the center of the plot, and the symmetry-related points in the opposite hemisphere have been inverted into the $z \geqslant 0$ hemisphere. The equal-area projection was used to maintain the true density of the scatter plot. The dashed lines represent latitude and longitude in 10° increments. The red circles represent the location of the noiseless data vectors in the x–y plane. The regions bounded by the heavy gray lines correspond to data plane orientations which give connection coefficients having the same pattern of signs (e.g. "±" corresponding to $r_1 > 0$ and $r_2 < 0$).

the data plane is given by

$$(x, y, z) = (\cos\phi \sin\theta, \sin\phi \sin\theta, \cos\theta),$$

then the connection coefficients are $r_1 = -x/z = -\cos\phi \tan\theta$ and $r_2 = -y/z = -\sin\phi \tan\theta$. Although $P(\theta, \phi)$ completely determines $P(r_1, r_2)$, the latter probability density may unfortunately be quite complicated and pathological, since it is the distribution of the ratios of random variables. Even for two normal random variables a and b, the probability density of their ratio $c = a/b$ is not simple (Fieller, 1932; Marsaglia, 1965), and can be bimodal and/or thick tailed, especially if the random variable b has significant probability density at zero. In our case, the density will become more non-normal as the probability density $P(\theta, \phi)$ at $\theta = \pi/2$ increases, becoming thick-tailed and causing the moments of r_1 and r_2 to become infinite. Regardless of the complexity of the distribution $P(r_1, r_2)$, the probability distribution for the algebraic signs of r_1 and r_2 (i.e. the structure of the network) can be very simple, and is determined by fraction of probability mass $P(\theta, \phi)$ lying in each the octant, as indicated in Fig. 2. Since the distribution of the data plane orientations is axially symmetric about z when the true values of r_1 and r_2 are 0 and the data vectors are orthogonal, it is easy to see that each pattern of algebraic signs of the estimated r_1 and r_2 is equally likely.

Next, consider the case where the true values of the connection coefficients are $r_1 < 0$ and $r_2 < 0$ (i.e. modules 1 and 2 both repress module 3). The normal vector to the data plane must be proportional to $(r_1, r_2, -1)$ in the absence of noise, and therefore will lie in either the $(x > 0, y > 0, z > 0)$ or the $(x < 0, y < 0, z < 0)$ octant (due to inversion symmetry). The distributions $P(\theta, \phi)$ for three examples (corresponding to $r_1 = r_2 = -0.5$, $r_1 = r_2 = -1$, and $r_1 = r_2 = -4$) for orthogonal data vectors $\boldsymbol{R}_1$ and $\boldsymbol{R}_2$ of length one at a noise standard deviation of $\sigma = 0.2$ are shown in Fig. 3. Since this problem is equivalent to the $r_1 = r_2 = 0$ case up to a rotation, it follows that the resulting distributions will also be axially symmetric. Although $P(r_1, r_2)$ is quite complicated, in practice the experimentalist may be interested in a much simpler question, namely, how qualitatively accurate is the inferred network? In particular, is the structure of the network correct, i.e. are the activating interactions inferred to have positive connection coefficients and the repressive ones to be negative? The probability that the estimated value of r_1 or r_2 does not have the same sign as the true value is shown as a function of r_1, r_2, and σ in Fig. 4. Although the probability increases with noise level as expected, the magnitude varies quite strongly with the true values of r_1 and r_2, particularly at lower noise levels.

This variability the probability of qualitative error can be understood in terms of the distribution $P(\theta, \phi)$. If the true values of r_1 and r_2 are both negative, then the probability of qualitative error is simply the probability that a point in the scatter plots of Fig. 3 ends up in an octant other than $(x > 0, y > 0, z > 0)$ or $(x < 0, y < 0, z < 0)$. One can immediately see, for example, that if the true values of r_1 and r_2 are small and of the same sign, then one is more likely to mis-estimate the sign of either r_1 or r_2 (but not both), while if r_1 and r_2 are large, then one is more likely to simultaneously mis-estimate both signs. In general, the probability of qualitative error must increase as the connection coefficients become very small or very large, since in either case the center of the $P(\theta,\phi)$ distribution approaches the boundary which defines the sign of r_1 or r_2. One can ask what values of r_1 and r_2 minimize this probability. The empirical answer to this question is shown in Fig. 5, which indicates that the optimum occurs near $r_1 = r_2 = 1$.

Theoretically, the minimum occurs when the center of the $P(\theta, \phi)$ distribution is simultaneously maximally far

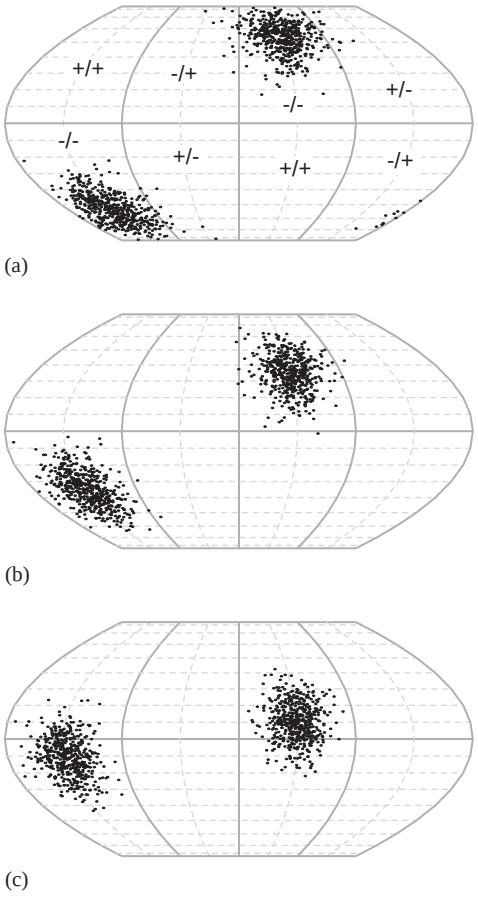

Fig. 3. Scatter plots of data plane orientations for orthogonal data vectors of length one corresponding to true values $r_1 = r_2 = -0.5$ (a), $r_1 = r_2 = -1$ (b), and $r_1 = r_2 = -4$ (c) to which 1000 realizations of noise of standard deviation 0.2 were added. The plots are pseudo-cylindrical equal-area Putnins P4' projections (Canters and Decleir, 1989), and latitude and longitude lines are shown in 10° and 45° increments, respectively, with the positive x-axis at the center of the plot. The regions bounded by the heavy gray lines correspond to data plane orientations which give connection coefficients having the same pattern of signs (as in Fig. 2).

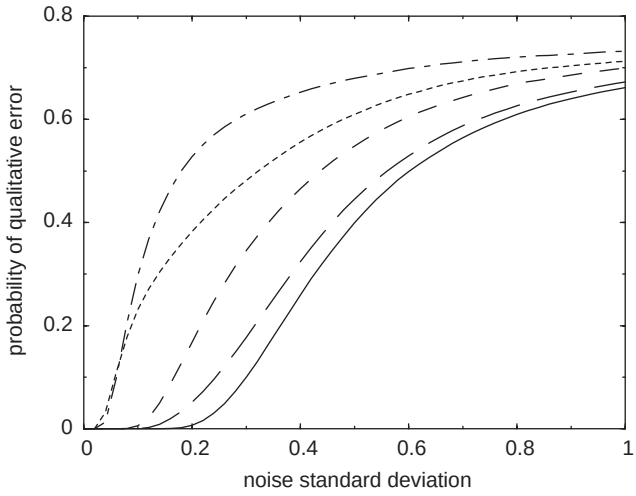

Fig. 4. Dependence of the probability of qualitative error (i.e. the probability of misestimating at least one of the signs of r_1 or r_2) as a function of the noise standard deviation for various values of r_1 and r_2. Probabilities were estimated in a Monte Carlo fashion using 10^5 noise realizations. The true values of (r_1, r_2) are $(-1,-1)$ (solid), $(-2,-2)$ (long-dashed), $(-0.3,-0.3)$ (short-dashed), $(-0.2,-2.5)$ (dotted), and $(-0.1,-0.1)$ (dot-dashed).

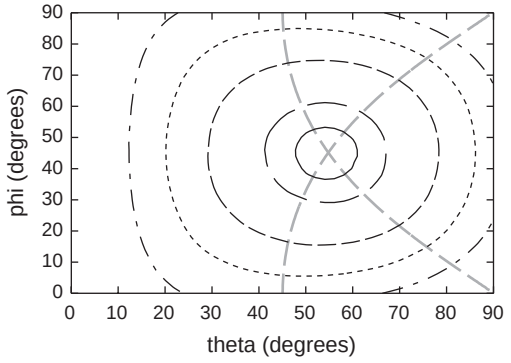

Fig. 5. Contour plot of the probability of qualitative error (i.e. the probability of misestimating at least one of the signs of r_1 or r_2) as a function of the spherical coordinates ϕ and θ of the normal vector to the data plane for the case of orthogonal data vectors of length one. Probabilities were estimated in a Monte Carlo fashion using 10^5 noise realizations of standard deviation 0.4. The error contours correspond to probabilities of 0.27 (solid), 0.3 (long-dashed), 0.4 (short-dashed), 0.5 (dotted), and 0.6 (dot-dashed), while the heavy dashed gray lines correspond to the values of ϕ and θ which give $r_1 = r_2 = -1$.

away from the edges of the octant $(x>0, y>0, z>0)$ (for simplicity we will neglect the inversion-symmetric possibility). Imagine for a moment that $P(\theta,\phi)$ is uniform on the spherical segment defined by a small circle of radius ρ centered at Cartesian coordinates $\boldsymbol{p}_c = (x_c, y_c, z_c)$ $(x_c>0, y_c>0, z_c>0)$, and that ρ sufficiently large that the circle cannot be fit within the spherical triangle defined by the coordinate axes. Then the position $\boldsymbol{p}_c$ which minimizes the probability of qualitative error is that which minimizes the area of the circle lying outside the triangle. This is minimized when the great-circle distances from $\boldsymbol{p}_c$ to the each of the three sides are minimized. Since the latter are given by $\cos^{-1}\sqrt{x_c^2+y_c^2}$, $\cos^{-1}\sqrt{x_c^2+z_c^2}$, and $\cos^{-1}\sqrt{y_c^2+z_c^2}$, the sum of these three distances is minimized when $x_c = y_c = z_c = 3^{-1/2}$, corresponding to connection coefficients $r_1 = r_2 = 1$. Furthermore, this argument remains valid for any distribution $P(\theta,\phi)$ provided that it is axially symmetric, and therefore has contours of equal probability which are small circles. An alternative derivation of the same result for a two-dimensional case is given in the appendix. We expect that the probability density of the connection coefficient itself will become strongly non-normal as the true value of r_1 becomes large even if the noise level is quite small, since the probability density $P(\theta,\phi)$ at $\theta = \pi/2$ increases as the center of the distribution approaches the x–y plane irrespective of the noise level.

3.1.2. Non-orthogonal data vectors

Let us next consider the more general and experimentally more relevant case where the noiseless data vectors are not orthogonal. First, consider the case where the true values of the connection coefficients are zero. Examples of scatter plots of samples from the $P(\theta,\phi)$ distribution of data plane orientations are shown in Fig. 6. Unlike the orthogonal case, the resulting distributions are no longer axially symmetric, and the degree of deviation from axial symmetry depends on the angle between the data vectors, becoming a "girdle-shaped" distribution when the data vectors are coincident (corresponding to a singular matrix in Eq. (11)). Since $P(\theta,\phi)$ is no longer symmetric when $\eta \neq \pi/2$, the absolute orientations of the data vectors *will* have an effect on the error propagation. For example, in Fig. 6(c) one can see that if the true data vectors lie in the $(+,+)$ or $(-,-)$ quadrants of the x–y plane and $\eta < \pi/2$, then it is considerably more likely that the inferred values of r_1 and $r_2$will be of opposite sign rather than the same sign. For the general case where the true values of r_1 and r_2 are not zero, the magnitude of the dependence of the qualitative error rate (probability of mis-estimating at least one of the signs of r_1 or r_2) depends strongly on the true values of r_1 and r_2, as can be seen in Fig. 7 for the case where the angle between the data vectors is 45°. The error rate follows approximately the same pattern as for the orthogonal case: one generally observes smaller error rates when the true values of r_1 and r_2 are close to 1 then when those values are larger or smaller. Unlike in the orthogonal data vector case, however, $r_1 = r_2 = 1$ does not strictly minimize the error rate, since we see that $r_1 = r_2 = 2$ gives a lower error rate than $r_1 = r_2 = 1$ for a small number of data vector orientations (Fig. 7). Furthermore, by comparing each solid curve to the corresponding horizontal line representing the error rate corresponding to orthogonal vectors, we can see that, for the most part, the error rate is larger when the angle between the data vectors is

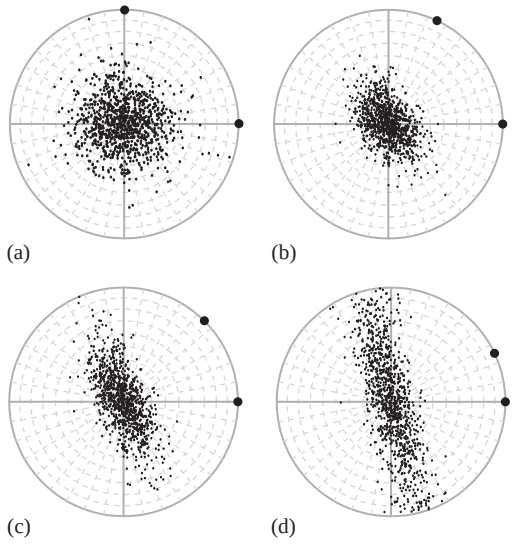

Fig. 6. Scatter plots of data plane orientations for non-orthogonal data vectors of length one lying in the x–y plane in the presence of noise level $\sigma = 0.2$. The scatter plot for orthogonal data is shown in panel (a) for references, and panels (b)–(d) correspond to angle between the noiseless data vectors of 65°, 45°, and 25°, respectively. The red circles represent the location of the noiseless data vectors in the x–y plane. Each plot shows 1000 noise realizations. The plots are equal-area Lambert projections (Fisher et al., 1987) of the $z \geqslant 0$ hemisphere oriented such that the z-axis is at the center of the plot, and the symmetry-related points in the opposite hemisphere have been inverted into the $z \geqslant 0$ hemisphere. The dashed lines represent latitude and longitude in 10° increments. The regions bounded by the heavy gray lines correspond to data plane orientations which give connection coefficients having the same pattern of signs (as in Fig. 2).

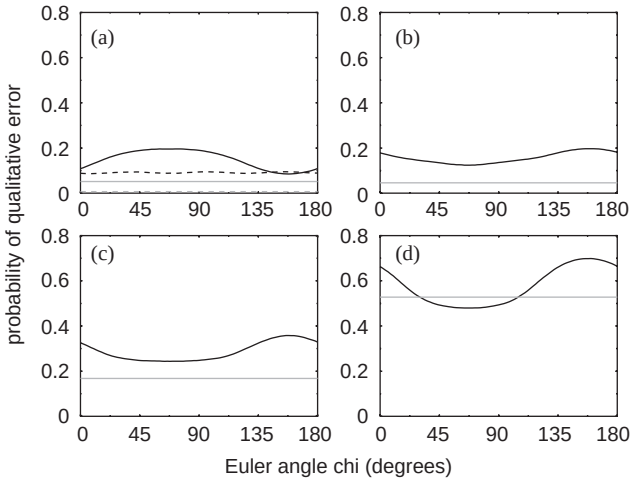

Fig. 7. Dependence of the qualitative error rate (the probability of misestimating at least one of the signs of r_1 or r_2) as a function of absolute vector orientation and the true values of r_1 and r_2. The results shown are for the case where the angle between the noiseless data vectors is 45°. The Euler angle χ is defined as follows: the noiseless data vectors $\boldsymbol{R}_1$ and $\boldsymbol{R}_2$ were initially taken to be equal to $(1, 0, 0)$ and $(2^{-1/2}, 2^{-1/2}, 0)$, respectively, and then rotated first through an angle $-\chi$ about the z-axis, then by an angle $-\theta$ about the y-axis, and finally by an angle $-\phi$ about the z-axis, where θ and ϕ are the spherical coordinates of the unit vector in the direction of $(r_1, r_2, -1)$. The probabilities were estimated from 10^5 noise realizations of standard deviation 0.2. Results are shown for (a) for $r_1 = r_2 = -2$, and $r_1 = r_2 = -1$ (solid and dotted lines, respectively), $r_1 = r_2 = -0.5$ (b), $r_1 = r_2 = -0.3$ (c), and $r_1 = r_2 = -0.1$ (d). The horizontal gray lines correspond to the qualitative error rate for two orthogonal data vectors at the same noise standard deviation (e.g. Fig. 4).

45° than when they are orthogonal. Again, there is an exception when the true values of r_1 and r_2 are quite small, in which case some orientations lead to a lower error rate than the orthogonal case.

3.2. Overdetermined case

Finally, let us consider the case where we have more than two measured data vectors for our hypothetical three-module network. As discussed in the Methods section above, as long as the perturbations do not perturb module 3, the resulting data vectors $\boldsymbol{R}_1, \ldots, \boldsymbol{R}_M$ will all lie in a plane in the zero-noise limit, and the normal vector to that plane will be proportional to $(r_1, r_2, -1)$. First, we would like to know how much accuracy in the inferred network is gained by adding additional data. Let us first consider the "best case scenario" where the data vectors are maximally spread out, i.e. with the M data vectors pointing toward each of the vertices of a regular M-gon. This corresponds to the orthogonal data vector case in the exactly determined case, and results in axially symmetric distributions $P(\theta, \phi)$ (data not shown). Increasing the number of data vectors should increase the accuracy of the inferred connection coefficients, as is confirmed in Fig. 8. Furthermore, it shows that when the true values of r_1 and r_2 are 1 and the noise level is relatively low, increasing the number of data vectors even modestly leads to a significant improvement in the qualitative error rate. Unfortunately, this improvement in error rate is not nearly as strong at higher noise levels, and also decreases when r_1 and r_2 are not near 1.

As in the two-vector case, there is a degradation in performance as the data vector orientations deviate from optimality. To describe the degree of non-optimality of a set of data vectors of unit length, we make use of the orientation matrix, a commonly used technique in the study of circular and spherical distributions (Fisher et al., 1987). For points on a circle, the orientation matrix is given by

$$M^{-1} \sum_{i=1}^{M} \begin{pmatrix} x_i^2 & x_i y_i \\ x_i y_i & y_i^2 \end{pmatrix}, \tag{13}$$

where x_i and y_i are the Cartesian coordinates of the ith data vector. The eigenvalues of the orientation matrix are μ and $1 - \mu$, where

$$\mu = \frac{1}{2}\left[1 + M^{-1}\sqrt{\left(\sum_{i=1}^{M} \cos 2\theta_i\right)^2 + \left(\sum_{i=1}^{M} \sin 2\theta_i\right)^2}\right] \tag{14}$$

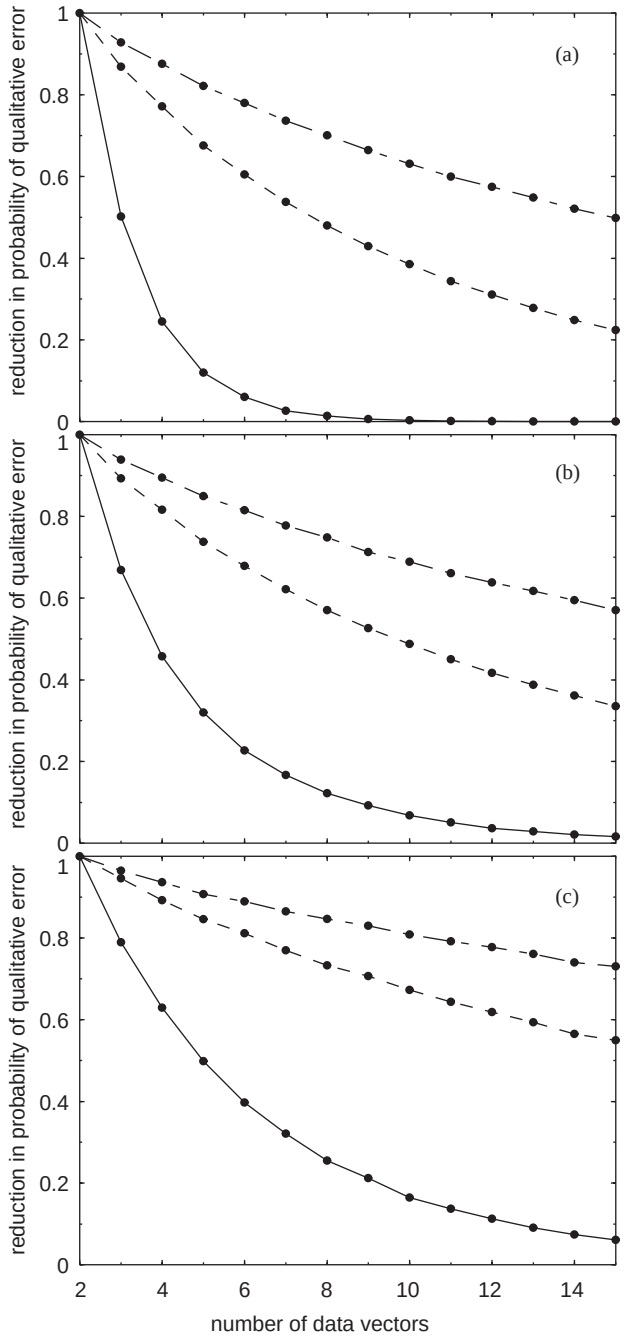

Fig. 8. The reduction in the probability of qualitative error (i.e. the probability of misestimating at least one of the signs of r_1 or r_2) due to additional data vectors relative to the two-vector case. Results are shown for noise levels of 0.3 (solid lines), 0.6 (dashed lines), and 0.8 (dot-dashed lines). The true values of the connection coefficients were $r_1 = r_2 = -1$ (a), $r_1 = r_2 = -2$ (b), and $r_1 = r_2 = -0.3$ (c), and the relative orientations were chosen to be optimal (i.e. orthogonal for the two-vector case, and pointing to the vertices of a regular M-gon for $M > 2$ vectors). Probabilities were estimated in a Monte Carlo fashion using 10^5 noise realizations.

ranges from $\frac{1}{2}$ when the M vectors are maximally spread out to 1 when they are linearly degenerate. The correlation of μ with the probability of qualitative error is shown in Fig. 9 for the case of three randomly oriented data vectors. The relationship is essentially deterministic when $r_1 = r_2 = 1$ (Fig. 9(a)), and degrades somewhat when $r_1 = r_2 \neq 1$. The latter observation is not at all unexpected, since μ contains no information at all about the absolute orientation of the data vectors. Nonetheless, these plots do provide some idea of the degree of degradation in the probability of qualitative error when the data vectors are oriented suboptimally.

To get a better idea of the practical impact of additional data vectors, let us consider an example based on the MAPK system of Fig. 1. Suppose for the moment that we wish to determine the values of r_{21} and r_{23}, i.e. the direct effect of the MKKK and MAPK modules on the MKK module, the theoretically determined values of which are 1.9 and -0.6, respectively (Kholodenko et al., 2002). In order to do that, we must perform perturbations which directly effect only modules 1 and 3. A minimal data set, for example, might perturb the system by decreasing the concentration of Ras-GTP by 6%, and $V_{\max 11}$ and $V_{\max 12}$ by 10%, leading to data vectors $\boldsymbol{R}_1 = (-0.462, -0.357, -0.813)$ and $\boldsymbol{R}_2 = (-0.350, -0.902, 0.348)$, respectively (data vectors have been scaled to be approximately length one). Using this minimal data set, one obtains qualitative error rates as a function of noise level as shown in the solid curve of Fig. 10(a). In addition, one can achieve similar perturbations by decreasing the total concentration of MKKK by 8%, and k_{cat9} and k_{cat10} by 10%,giving two additional data vectors $\boldsymbol{R}_3 = (-0.446, -0.344, -0.784)$ and $\boldsymbol{R}_4 = (0.347, 0.901, -0.350)$. While these data vectors are essentially "redundant", in that the pairs $(\boldsymbol{R}_1, \boldsymbol{R}_3)$ and $(\boldsymbol{R}_2, \boldsymbol{R}_4)$ are nearly co-linear (Fig. 10(b)), one does obtain a significant reduction in the qualitative error rate, as can be seen by comparing the long-dashed and solid curves of Fig. 10(a). Rather than reducing $[\mathrm{MKKK}]_{tot}$, k_{cat9} and k_{cat10} in two separate experiments, one could instead perturb them simultaneously, resulting in the data vector $\boldsymbol{R}_5 = (-0.152, 0.319, -0.926)$. The qualitative error rates for the data set $(\boldsymbol{R}_1, \boldsymbol{R}_2, \boldsymbol{R}_5)$ is shown as the short-dashed curve in Fig. 10(a), and shows that this data set results in somewhat higher error rates than the $(\boldsymbol{R}_1, \boldsymbol{R}_2, \boldsymbol{R}_3, \boldsymbol{R}_4)$ data set, but still represents an improvement over the minimal$(\boldsymbol{R}_1, \boldsymbol{R}_2)$ set. Finally, the use of all five data vectors gives the error rates shown in the dotted curve of Fig. 10(a). This data set gives the smallest error rates of the four data sets considered here, and at $\sigma = 0.4$ represents nearly a factor of 2 reduction in the error rate over the minimal two-vector data set.

Obtaining enough data for the problem to be overdetermined is useful not only because it improves the accuracy of the inferred network, but also because it allows us to assess the goodness-of-fit of the model. Up

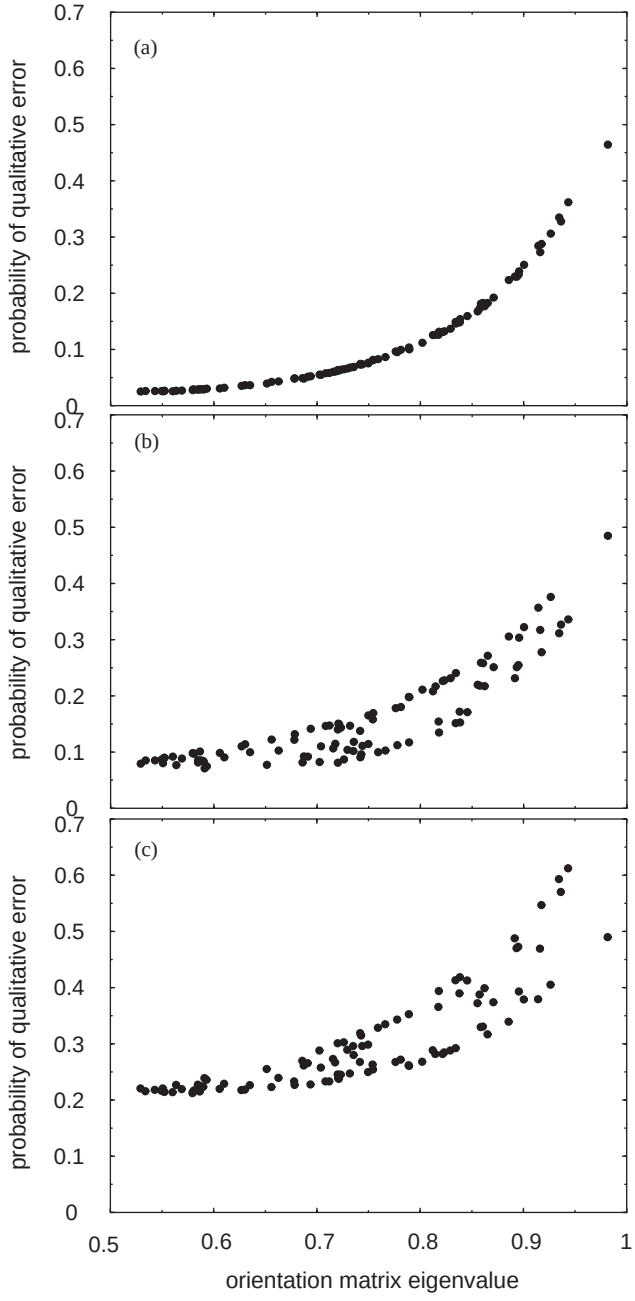

Fig. 9. Dependence of the qualitative error rate (the probability of misestimating at least one of the signs of r_1 or r_2) as a function of the largest eigenvalue of the orientation matrix of the data vectors given by Eqs. (11) and (12). For all plots the data vectors were of unit length and the noise level was 0.3. The true values of the connection coefficients were $r_1 = r_2 = -1$ (a), $r_1 = r_2 = -2$ (b), and $r_1 = r_2 = -0.3$ (c). Each of the 100 points in each panel corresponds to a different set of data vectors generated randomly with a uniform distribution on the unit circle. Qualitative error probabilities were estimated in a Monte Carlo fashion using 10^5 noise realizations.

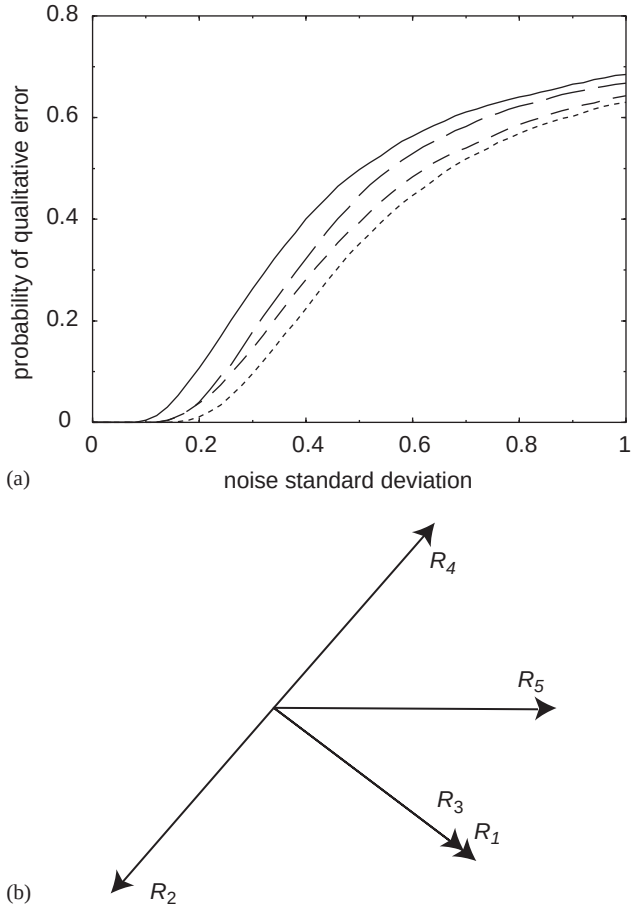

Fig. 10. (a) Dependence of the probability of qualitative error (i.e. the probability of misestimating at least one of the signs of r_1 or r_2) as a function of the noise standard deviation for various combinations of data vectors obtained by numerical solution of the MAPK model given in Fig. 1. Results are shown for data sets $(\boldsymbol{R}_1, \boldsymbol{R}_2)$ (solid), $(\boldsymbol{R}_1, \boldsymbol{R}_2, \boldsymbol{R}_3, \boldsymbol{R}_4)$ (long-dashed), $(\boldsymbol{R}_1, \boldsymbol{R}_2, \boldsymbol{R}_5)$ (short-dashed), and $(\boldsymbol{R}_1, \boldsymbol{R}_2, \boldsymbol{R}_3, \boldsymbol{R}_4, \boldsymbol{R}_5)$ (dotted) (see text for details). Probabilities were estimated in a Monte Carlo fashion using 10^4 noise realizations. (b) Relative lengths and orientations of the five data vectors $\boldsymbol{R}_1, \boldsymbol{R}_2, \boldsymbol{R}_3, \boldsymbol{R}_4$, and $\boldsymbol{R}_5$ in the data plane.

to this point, we have assumed that, apart from the additive noise, the data fit the model perfectly, i.e. there is no systematic error. In real-world problems, however, this may not be the case. In particular, the unravelling strategy contains several key assumptions, the validity of which may not be known with certainty. For example, one must make assumptions about the number of communicating intermediates, which chemical species can be treated as communicating intermediates, and which modules are (not) directly effected by a given perturbation. We would like to be able to detect when our observed data is not consistent with our assumed model. Measures of the consistency of the data with the model are known as the "goodness-of-fit", and most commonly take the form of the sum of the squared deviations of the data from the values predicted by the best-fit model (the "residuals"). However, if we only have just enough data to fit the model, we cannot hope to find such inconsistencies based on goodness-of-fit, since in all cases we will obtain zero residuals. If we have

more data than parameters, then the sum of the squared residuals will not be zero in general, but will have contributions from the additive noise and from any deviations of the data from the model. A standard way of assessing the magnitude of the residuals is via goodness-of-fit testing, where given an estimate of the magnitude of the additive noise we can estimate the probability of seeing residuals of the observed magnitude if the model were correct.

As an example of how such goodness-of-fit testing might be used in a real-world situation, consider the MAPK cascade of Fig. 1. Let us suppose that the details of this system were unknown, and that in fact we did not know that the doubly phosphorylated forms of the kinases were the communicating intermediates, but assumed (wrongly) that the unphosphorylated forms were the relevant communicating intermediates. If we perturb module 2 by decreasing k_{cat5} and k_{cat6} by 10%, and perturb module 3 by decreasing k_{cat9} and k_{cat10} by 10%, then the resulting data vectors are $\boldsymbol{R}_1 = (-1.07, 0.19, 3.13)$ and $\boldsymbol{R}_2 = (-0.54, -0.62, 1.64)$. These lead to connection coefficients r_{21} and r_{31} of -0.02 and -0.34, which are in substantial disagreement with the theoretical values of 0 and -1.1 (Kholodenko et al., 2002). Suppose that we extended the data set by including four more perturbations: (1) decreasing $V_{\max 7}$ and $V_{\max 8}$ by 10%, (2) decreasing $V_{\max 11}$ and $V_{\max 12}$ by 10%, (3) decreasing the total concentration of MKK by 5%, and (4) decreasing the total concentration of MAPK by 5%. The resulting matrix of fractional responses is

$$\boldsymbol{R} = \begin{pmatrix} -1.07 & 0.19 & 3.13 \\ -0.54 & -0.62 & 1.64 \\ 1.01 & -0.19 & -3.37 \\ 0.53 & 0.57 & -1.71 \\ -0.39 & -5.40 & 1.18 \\ -0.16 & -0.18 & -7.15 \end{pmatrix}$$

and gives the (still incorrect) connection coefficients $r_{21} = 0.06$ and $r_{31} = -0.10$ and a smallest singular value (equal to the square root of the sum of the squared residuals) $\varsigma = 1.42$. One can now ask what is the probability that we would have obtained such a value ς by chance under independent and identically distributed normal noise if the unravelling model were correct? Since ς^2/σ^2 is a χ^2 random variable of MN degrees of freedom (where σ^2 is the noise variance), the probability in question can be readily determined. Specifically, we find that the probability that we will obtain $\varsigma = 1.42$ is less than 5% if the noise standard deviation is 0.25 or smaller. Thus, if the noise is sufficiently small, then we could detect that the assumptions underlying unravelling were not satisfied by means of a significance test strategy.

4. Conclusions

In earlier work (Kholodenko et al., 2002; Kholodenko and Sontag, 2002), we developed a reverse engineering strategy for the determination of the structure of biological networks based on the effect of perturbations on the steady-state concentrations of key intermediates. The results above describe the effect that experimental uncertainty has on the accuracy of the networks obtained by that procedure. One important measure of the accuracy is the probability that the inferred network is qualitatively incorrect, i.e. that one or more of the inferred connection coefficients are of the wrong sign and an inhibitory interaction has been mistaken for an activating one (or vice versa). We have found that in the case of independent and identically distributed normal errors, the probability of qualitative error is minimized when the true values of the connection coefficients are near 1 and when the data vector directions are oriented maximally far apart from each other (orthogonal or pointing toward the vertices of a regular polygon). These results can be understood in an intuitive manner by considering the distribution of the normal vector to the plane determined by the data vectors. Increasing the number of data vectors does, of course, increase the accuracy, though the degree of reduction is reduced if the noise level is high, the true values of the connection coefficients are not near 1, or the data vector directions have a high degree of collinearity. Furthermore, we show how the residual from an overdetermined unravelling problem could be used as part of a significance test to detect if the assumptions of the unravelling strategy have been violated.

It should be remembered that in some sense these results represent a ''best case scenario'', since realistic data will have noise that is not normally distributed. We are currently investigating more realistic noise models and their effect on the network accuracies, as well as alternative inference methods that could reduce the overall probability of qualitative error. Furthermore, we are investigating the possibility of using the ideas of this work (namely the co-planarity of data vectors associated with a given module) as a basis for a techniques to mine biological data when the underlying modules, communicating intermediates, and perturbations are not known with certainty.

Acknowledgements

MA, RML, and ES acknowledge support by Grant GM068699 from the National Institutes of Health and by the BioMaPS Institute for Quantitative Biology at Rutgers University. BNK acknowledges support by Grant GM59570 from the National Institutes of Health.

Appendix A

Here, we present an alternative derivation for the two-module case of the fact that the error in the connection coefficient is minimized when the absolute value of the connection coefficient is 1. Suppose that the connection coefficient describing the effect of one module on the other is a. Therefore, the unit-length data vector $\boldsymbol{R}$ in the absence of noise will be orthonormal to $\boldsymbol{r} = (a, -1)$:

$$\boldsymbol{R} = \left(\frac{1}{\sqrt{1+a^2}}, \frac{a}{\sqrt{1+a^2}}\right). \tag{A.1}$$

Adding the noise vector $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2)$ to $\boldsymbol{R}$, and solving for $\hat{\boldsymbol{r}} = (x, -1)$ such that $(\boldsymbol{R} + \boldsymbol{\varepsilon}) \cdot \hat{\boldsymbol{r}} = 0$, we find that

$$x = \frac{a + \varepsilon_2\sqrt{1+a^2}}{1 + \varepsilon_1\sqrt{1+a^2}}. \tag{A.2}$$

We wish to find the value of a which gives the smallest relative error

$$E = \frac{x-a}{a} = \frac{\sqrt{1+a^2}(\varepsilon_2 - a\varepsilon_1)}{(1+\varepsilon_1\sqrt{1+a^2})a}. \tag{A.3}$$

Let us assume that the random variables ε_1 and ε_2 are independent and have zero mean and small but equal variances. The last assumption allows us to replace the factor in parentheses in the denominator of Eq. (A.3) with 1, giving

$$E \approx \frac{\varepsilon_2\sqrt{1+a^2}}{a} - \varepsilon_1\sqrt{1+a^2}. \tag{A.4}$$

Since ε_1 and ε_2 have mean zero, the mean relative error is zero (to within the approximation of Eq. (A.4)), and is independent of the value of a. Therefore, we instead find the value of a which minimizes the mean-*squared* relative error, which in turn is simply the variance of E. Since ε_1 and ε_2 are uncorrelated and have equal variance, the variance of E is proportional to $a^2 + a^{-2} + 2$, which is minimized when $a = \pm 1$.

Appendix B. Supplementary material

The online version of this article contains additional supplementary material. Please visit doi: 10.1016/j.jtbi.2004.08.022.

References

Brand, M.D., 1996. Top down metabolic control analysis. J. Theor. Biol. 182, 351–360.

Brazhnik, P., de la Fuente, A., Mendes, P., 2002. Gene networks: how to put the function into genomics. Trends Biotech. 20, 467–472.

Brivanlou, A.H., Darnell, J.E.J., 2002. Signal transduction and the control of gene expression. Science 295, 813–818.

Canters, F., Decleir, H., 1989. The World in Perspective: A Directory of World Map Projections. Wiley, Chichester.

Chang, L., Karin, M., 2001. Mammalian MAP kinase signalling cascades. Nature 410, 37–40.

de Jong, H., 2002. Modeling and simulation of genetic regulatory systems: a literature review. J. Comput. Biol. 9, 67–103.

de la Fuente, A., Brazhnik, P., Mendes, P., 2002. Linking the genes: inferring quantitative gene networks from microarray data. Trends Genet. 18, 395–398.

de Moor, B., 1993. Structured total least squares and L_2 approximation problems. Linear Algebra Appl. 188–189, 163–205.

Dongarra, J.J., Bunch, J.R., Moler, C.B., Stewart, G.W., 1979. LINPACK Users' Guide. SIAM Press, Philadelphia.

Durando, G., Mana, G., 2002. Propagation of error analysis in a total least-squares estimator in absolute gravimetry. Metrologia 39, 489–494.

Eckart, G., Young, G., 1936. The approximation of one matrix by another of lower rank. Psychometrica 1, 211–218.

Fieller, E.C., 1932. The distribution of the index in a normal bivariate population. Biometrika 24, 428–440.

Fisher, N.I., Lewis, T., Embleton, B.J.J., 1987. Statistical Analysis of Spherical Data. Cambridge University Press, Cambridge.

Gardner, T.S., di Bernardo, D., Lorenz, D., Collins, J.J., 2003. Inferring genetic networks and identifying compound mode of action via expression profiling. Science 301, 102–105.

Gerhold, G.A., 1969. Least-squares adjustment of weighted data to a general linear equation. Am. J. Phys. 37, 156–161.

Golub, G.H., van Loan, C.F., 1980. An analysis of the total least squares problem. SIAM J. Numer. Anal. 17, 883–893.

Golub, G.H., van Loan, C.F., 1989. Matrix Computations. The Johns Hopkins University Press, Baltimore.

Hartwell, L.H., Hopfield, J.J., Leibler, S., Murray, A.W., 1999. From molecular to modular cell biology. Nature 402, C47–C52.

Hasty, J., McMillen, D., Isaacs, F., Collins, J.J., 2001. Computational studies of gene regulatory networks: in numero molecular biology. Nat. Rev. Genet. 2, 268–279.

Hazzalin, C.A., Mahadevan, L.C., 2002. MAPK-regulated transcription: a continuously variable gene switch? Nat. Rev. Mol. Cell Biol. 3, 30–40.

Huang, C.F., Ferrell, J.E.J., 1996. Ultrasensitivity in the mitogen-activated protein kinase cascade. Proc. Natl Acad. Sci. USA 93, 10078–10083.

Kholodenko, B.N., 2000. Negative feedback and ultrasensitivity can bring about oscillations in the mitogen-activated protein kinase cascades. Eur. J. Biochem. 267, 1583–1588.

Kholodenko, B.N., Sontag, E.D., 2002. Determination of functional network structure from local parameter dependence data. arXiv:-physics/0205003.

Kholodenko, B.N., Hoek, J.B., Westerhoff, H.V., Brown, G.C., 1997. Quantification of information transfer via cellular signal transduction. FEBS Lett. 414, 430–434.

Kholodenko, B.N., Demin, O.V., Moehren, G., Hoek, J.B., 1999. Quantification of short term signaling by the epidermal growth factor receptor. J. Biol. Chem. 274, 30169–30181.

Kholodenko, B.N., Kiyatkin, A., Bruggeman, F.J., Sontag, E., Westerhoff, H.V., Hoek, J.B., 2002. Untangling the wires: a strategy to trace functional interactions in signaling and gene networks. Proc. Natl Acad. Sci. USA 99, 12841–12846.

Lauffenburger, D.A., 2000. Cell signaling pathways as control modules: complexity for simplicity? Proc. Natl Acad. Sci. USA 97, 5031–5033.

Lefebvre, M., Keeler, R.K., Sobie, R., White, J., 2000. Propagation of errors for matrix inversion. Nucl. Insrum. Methods: Phys. Res. A 451, 520–528.

Marsaglia, G., 1965. Ratios of normal variables and ratios of sums of uniform variables. J. Am. Stat. Assoc. 60, 193–204.

Moehren, G., Markevich, N., Demin, O., Kiyatkin, A., Goryanin, I., Hoek, J.B., Kholodenko, B.N., 2002. Temperature dependence of the epidermal growth factor receptor signaling network can be accounted for by a kinetic model. Biochemistry 41, 306–320.

Premoli, A., Rastello, M.L., 2002. The parametric quadratic form method for solving TLS problems with elementwise weighting. In: van Huffel, S., Lemmerling, P. (Eds.), Total Least Squares and Errors-in-Variables Modeling. Kluwer Academic Publishers, Dordrecht, pp. 67–76.

Press, W.H., Teukolsky, S.A., Vetterling, W.T., Flannery, B.P., 1992. Numerical Recipes in C: The Art of Scientific Computing. Cambridge University Press, Cambridge.

Ronen, M., Rosenberg, R., Shraiman, B.I., Alon, U., 2002. Assigning numbers to the arrows: parametrizing a gene regulation network by using accurate expression kinetics. Proc. Natl Acad. Sci. USA 99, 10555–10560.

Shvartsman, S.Y., Hagan, M.P., Yacoub, A., Dent, P., Wiley, H.S., Lauffenburger, D.A., 2002. Autocrine loops with positive feedback enable context-dependent cell signaling. Am. J. Physiol. 282, C545–C559.

Smolen, P., Baxter, D.A., Byrne, J.H., 1998. Frequency selectivity, multistability, and oscillations emerge from models of genetic regulatory systems. Am. J. Physiol. 274, C531–C542.

Stark, J., Callard, R., Hubank, M., 2003. From the top down: towards a predictive biology of signalling networks. Trends Biotech. 21, 290–293.

Tyson, J.J., Chen, K., Novak, B., 2001. Network dynamics and cell physiology. Nat. Rev. Mol. Cell. Biol. 2, 908–916.

van Huffel, S., Vandewalle, J., 1991. The Total Least Squares Problem: Computational Aspects and Analysis. SIAM, Philadelphia.

von Dassow, G., Meir, E., Munro, E.M., Odell, G.M., 2000. The segment polarity network is a robust developmental module. Nature 406, 188–192.

Widmann, C., Gibson, S., Jarpe, M.B., Johnson, G.L., 1999. Mitogen-activated protein kinase: conservation of a three-kinase module from yeast to human. Physiol. Rev. 79, 143–180.

Wolfram, S., 1991. Mathematica: A System for Doing Mathematics by Computer. Addison-Wesley, Reading, MA.