Published in IET Systems Biology
Received on 19th July 2011
Revised on 28th December 2011
doi: 10.1049/iet-syb.2011.0052

ISSN 1751-8849

Comparison of statistical and optimisation-based methods for data-driven network reconstruction of biochemical systems

B. Asadi[1,2] *M.R. Maurya*[2] *D.M. Tartakovsky*[1] *S. Subramaniam*[2,3]

[1]*Department of Mechanical and Aerospace Engineering, University of California, San Diego, La Jolla, CA, USA*
[2]*Department of Bioengineering, University of California, San Diego, La Jolla, CA, USA*
[3]*Department of Chemistry and Biochemistry, Cellular and Molecular Medicine and Graduate Program in Bioinformatics, University of California, San Diego, La Jolla, CA, USA*
E-mail: mauryam@yahoo.com

Abstract: Data-driven reconstruction of biological networks is a crucial step towards making sense of large volumes of biological data. Although several methods have been developed recently to reconstruct biological networks, there are few systematic and comprehensive studies that compare different methods in terms of their ability to handle incomplete datasets, high data dimensions and noisy data. The authors use experimentally measured and synthetic datasets to compare three popular methods – principal component regression (PCR), linear matrix inequalities (LMI) and least absolute shrinkage and selection operator (LASSO) – in terms of root-mean-squared error (RMSE), average fractional error in the value of the coefficients, accuracy, sensitivity, specificity and the geometric mean of sensitivity and specificity. This comparison enables the authors to establish criteria for selection of an appropriate approach for network reconstruction based on a priori properties of experimental data. For instance, although PCR is the fastest method, LASSO and LMI perform better in terms of accuracy, sensitivity and specificity. Both PCR and LASSO are better than LMI in terms of fractional error in the values of the computed coefficients. Trade-offs such as these suggest that more than one aspect of each method needs to be taken into account when designing strategies for network reconstruction.

1 Introduction

Inferring the topology of networks from experimental data is a central endeavour in modern biology [1]. It is essential for meaningful interpretation and use of biological data. Although there has been progress in developing kinetic models of biochemical networks with known topology and interactions, the number of such networks is not large [1]. Recent advances in high-throughput techniques have led to the collection of different types of data related to signalling, gene regulatory and metabolic networks. Various data types, for example, cellular read outs, obtained via alternative experimental techniques, have distinct observational scales (e.g. high against low intensity in microarray datasets) and accuracy. Integration of different data types, although technically challenging, can lead to dramatic improvements in our ability to identify the connectivity of different segments of a network and to predict events within a cellular system. Some recent studies, in which data integration was used to reconstruct biochemical networks and predictive models from large-scale datasets, can be found in [2–4].

Given the importance of network reconstruction, various methods have been introduced and are being developed to reconstruct static, dynamic and static–dynamic networks [5, 6]. Some of the drawbacks of different approaches can be compensated by the others [7]. Examples of well-developed network reconstruction techniques include optimisation-based approaches (e.g. least-squares methods [8]), dimensionality reduction methods (e.g. statistical significant tests combined with either principal component regression (PCR), or partial least-squares (PLS) [4, 9]), partial-correlation-related analyses [10], Bayesian networks [3, 11], hybrid methods (e.g. linear matrix inequalities (LMI) [12, 13] and least absolute shrinkage and selection operator (LASSO) [2, 14]) and matrix-based approaches [15, 16]. Through appropriate formulation these approaches can be tailored to static or temporal (dynamic) data. When yielding conflicting predictions, such alternative approaches provide opportunities to propose additional hypothesis and experiments. A detailed review of various network-reconstruction methods can be found in [1].

These methods highlight the importance and benefits of employing systems approaches to decipher and reconstruct cellular networks from high-throughput data. For example, a PLS approach has been used to identify the interaction of apoptotic and pro-survival signals in cellular apoptosis [3]. Deciphering the connections in this network is complicated since any given stimulus activates more than one pathway, and any given pathway can also be activated by several stimuli. In turn, the signalling pathways act together as a

module or network to extract specific responses. In [3] the authors used high-throughput data on signalling activations and apoptotic or survival phenotypes for reconstructing a canonical network to link signalling to apoptosis. An alternative algorithm proposed in [4] enables one to perform an input–output mapping by utilising steady-state or time-averaged data. The algorithm was used to identify lumped networks from signalling pathways to cytokines in macrophages. It employed the levels of activation of the signalling pathway acting as inputs and the levels of cytokine release acting as outputs.

To the best of our knowledge, there have been a few systematic efforts to compare the performance of various network-reconstruction methods in terms of their ability to deal with data patterns characterised by different amounts of missing data, dimension of the dataset and type and level of noise. For example, the Dialogue for Reverse Engineering Assessments and Methods (DREAM) is an extensive effort towards optimal design, application and assessment of models in systems biology [17]. Several prediction and model assessment metrics have been used in DREAM that are common with the current work. There are also a number of worthy reviews available in the literature for comparing the network reconstruction methods such as [18, 19]. However, as pointed out by Cho *et al.* [19], there is still a need for improvement of our understanding regarding the fundamental idea presented by each method and to consider properties of input data, constraints and specific applications for selecting an appropriate network reconstruction method. The ability to effectively deal with different levels of noise is especially important for biological datasets because different types of measurements, such as proteomic data [4] against gene-expression data [20], inherently have different measurement errors and other uncertainties. For example, in gene-expression measurements, the relative level of noise in fold-change data is higher if the corresponding intensities are in the low-value regime [21]. In the present work, we provide a systematic comparison of the ability of three methods – PCR, LMI and LASSO to reconstruct networks from experimentally measured biological data and synthetic data. Comparison criteria are the robustness in identifying a 'true' network and the accuracy of the estimated model coefficients to different levels of missing/unavailable data and different types and levels of noise in the data. When appropriate, the comparison also includes results from a standard least squares (SLS; abbreviation 'LS' also used) approach. This study is motivated by the fact that although a systematic comparison is theoretically intractable, it is amenable to computational exploration. Additional novelty of the current work also includes a more accurate approach to statistical significance analysis of the model coefficients in the PCR approach than that used in [4]. Further, previous application of LMI to network reconstruction [12] used an *ad hoc* approach to decide the threshold. We have improved this approach by using a cross-validation procedure to compute the threshold in the LMI method automatically.

In Section 2, we briefly describe the four network-reconstruction methods as mentioned above. These methods are used in Section 3 to reconstruct biological networks from an experimentally measured dataset (phosphoprotein [PP] signalling and cytokine measurements in RAW 264.7 cells provided by the Alliance for Cellular Signalling [AfCS] [22]) and from simulated data for which the true network (or model coefficients) is known. In this section, the effect of noise level, noise type and the size of dataset on the performance of different methods will be investigated. The effect of the amount of missing data will also be studied. This section ends with a comparison of computation time used by these methods. Section 4 presents a discussion and comparative analysis of the different methods for non-collinear and collinear input datasets. Section 5 concludes the article with some directions for future efforts.

2 Methods

In this study, we compare four methods for network reconstruction, namely, LS, PCR, LASSO and LMI. The selection of these methods is based on the conceptual differences in the approaches they employ for network reconstruction. LS is based on minimisation of the sum of squared errors between predicted response and true response, PCR is based on dimensionality reduction, LASSO is based on shrinkage of the model coefficients by minimising the sum of squared errors [2, 14] and LMI is based on the minimisation of L-infinity norm of the prediction error and shrinkage heuristics [12, 13]. Here we introduce the well-studied method of SLS. Detailed description of the other three methods, PCR, LASSO and LMI, are presented in Supplementary material (Sections 1.2–1.4).

The SLS method, one of the oldest techniques in modern statistics [23], estimates unknown coefficients (or parameters) of a linear model by minimising the sum of squares of model deviations from an observed response. Consider a system that consists of n input variables (predictors) and p outputs. Let m denote the number of data points, $\boldsymbol{X}$ denote an $m \times n$ matrix of input data (each column normalised to zero-mean and unit standard deviation) and $\boldsymbol{Y}$ denote an $m \times p$ matrix of the corresponding (mean-centred) observed response (outputs). For simplicity, we assume that $p = 1$ (else repeat the procedure on each output individually) and hence denote the data by $\boldsymbol{y}$ as opposed to $\boldsymbol{Y}$. Let $\hat{\boldsymbol{b}}$ be an estimate of the coefficient vector $\boldsymbol{b}$ in the linear affine system, so that, $\boldsymbol{y} = \boldsymbol{Xb}$. Specifically

$$\boldsymbol{y} = \boldsymbol{X}\hat{\boldsymbol{b}} + \boldsymbol{\varepsilon} \tag{1}$$

where $\boldsymbol{\varepsilon}$ is the residual vector. SLS computes $\hat{\boldsymbol{b}}$ as

$$\hat{\boldsymbol{b}} = \arg\min\{\mathrm{e}^2 = (\boldsymbol{y} - \boldsymbol{Xb})^{\mathrm{T}}(\boldsymbol{y} - \boldsymbol{Xb})\} \tag{2}$$

The least squares solution to (2) is

$$\hat{\boldsymbol{b}} = (\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})^{-1}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{y} \tag{3}$$

The predicted values of the output data, $\boldsymbol{y}_{\mathrm{p}}$, can be computed as

$$\boldsymbol{y}_{\mathrm{p}} = \boldsymbol{X}\hat{\boldsymbol{b}}$$

More details, such as calculation of root-mean-squared error (RMSE) for prediction, are described in Supplementary material Section 1.1. Two different metrics (fractional error in the estimated coefficients, and sensitivity, specificity and related metrics) have been utilised to evaluate the performance of the methods. They are described in Supplementary material, Section 1.5.

doi: 10.1049/iet-syb.2011.0052

3 Results

In this section, we present the results with input data containing independent columns. The case of collinear data is presented in the next section. Relative performance of the four network-reconstruction methods is analysed below.

3.1 Performance on experimental data (PP/cytokine data)

This case study deals with network reconstruction for cytokine release in RAW 264.7 cells using the data from the AfCS [22]. In response to molecules that bind to cell surface receptors, these macrophage cells utilise a cascade of cellular signalling events to alter gene expression resulting in release of cytokines. The cytokines are a hallmark of inflammation and innate immune response in physiology [24]. The large scale measurements of PPs and cytokines as a function of stimulus and time in these cells serve as an excellent testbed for network reconstruction methods. Although these individual macromolecule measurements and cell phenotypes are well-characterised, the knowledge of precise pathways and context-specific networks is incomplete. Prior work has established some of these networks and hence serves as a good metric for comparison and validation of network reconstruction methods [24, 25]. We use this as the experimental case study to analyse the reconstruction methods described above. The PP/cytokine dataset consists of 22 inputs and 6 outputs (G-CSF, IL-1α, IL-6, IL-10, MIP-1α, TNFα). Both the input and output data are time averaged since the time scales of the input (PP) data are in minutes (measurements taken at 1, 3, 10 and 30 min) whereas that of the output (cytokines) data are in hours (measurements taken at 2, 3 and 4 h). The PPs were measured using western blots (AfCS protocols #PP00000177 and #PP00000181) and cytokines were measured using multiplex suspension arrays (AfCS protocols #PP00000209 and #PP00000223 [22]). More details about the experiments can be found on the AfCS website [22] and the procedure for pre-processing the data is explained elsewhere [4]. RMSE of the network reconstructed with each of the four methods was calculated for all outputs. Fig. 1 shows a scatter plot of the predicted outputs against corresponding experimental values for the LS and PCR methods. The dashed and dotted lines represent the $\sigma = \text{RMSE}_{\text{LS}}$ and 2σ bands, respectively.

As discussed in Section 2 and Supplementary material, Sections 1.1–1.4, each method employs a different strategy to identify the most informative/significant variables contributing to network identification. The number of significant variables depends on a selection criterion (tuning parameter). For PCR, $0.8 <$ Fraction of the cumulative variance (FCV) < 0.95 was used to capture 80–95% of the variance in the input data, and the significance was based on the average ratio for t-test to obtain a stable estimate. The t-test was also employed in the LS method to identify significant predictors ('LS_sig' in Table S1; LS refers to SLS without any significance testing). In LASSO, the criterion was set to $t = 0.66$ which translates into about 33% ($= 1-0.66$) of the coefficients becoming zero in the resulting model. For LMI, the threshold $r_{\text{LMI}} = 0.3$ was used. A detailed discussion on the selection of appropriate thresholds for the LASSO and LMI methods using cross-validation is presented in Supplementary material, Section 3.1. For each method the number of significant input variables identified and used to build the model is listed in

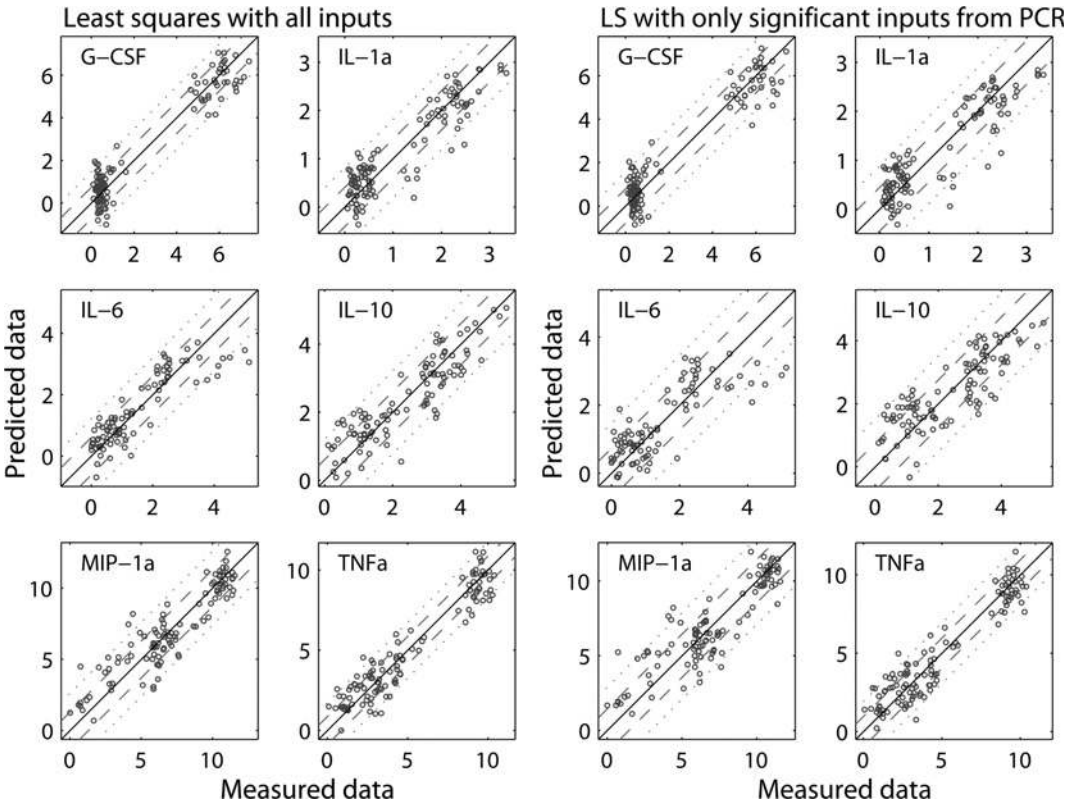

Fig. 1 *Actual response (experimental data) against response predicted with the LS and PCR methods for PP/cytokine data*

Dashed and dotted lines represent the σ and 2σ bands, respectively. IL: interleukin; TNF: tumor necrosis factor; G-CSF: granulocyte colony stimulating factor, MIP: macrophage inflammatory protein

Table 1 Number of significant inputs (TP, FP) for each output of PP/cytokine data

Output	LS_sig.	PCR	LASSO	LMI
G-CSF	3 (1,2)	11 (2,9)	6 (3,3)	12 (3,9)
IL-1α	4 (1,3)	12 (2,10)	6 (3,3)	15 (3,12)
IL-6	1 (1,0)	5 (3,2)	2 (2,0)	8 (1,7)
IL-10	3 (2,1)	7 (4,3)	7 (4,3)	8 (3,5)
MIP-1α	2 (1,1)	11 (3,8)	6 (3,3)	16 (3,13)
RANTES	1 (1,0)	9 (2,7)	4 (2,2)	9 (3,6)
TNFα	4 (2,2)	12 (4,8)	6 (4,2)	13 (3,10)

Table 1 along with the number of true and false positive (TP, FP) connections. The method of LS with significance testing retains the smallest number of inputs (predictors). PCR comes second, whereas LASSO tends to retain the largest number of inputs (depending on the value of t). Overall, LASSO and LMI are comparable.

3.2 Performance on synthetic noisy data

Next, we apply the four methods to reconstruct a network from a synthetic data consisting of 25 inputs and 1 output. In an effort to make the synthetic networks as compatible as possible with experimental datasets, several factors such as network size, noise variance, granularity, sparse network topology and appropriate correlation range among columns of input matrix have been taken into account in generating 'synthetic' data for network reconstruction and analysis. Synthetic datasets, which are generated with a model whose parameters are known, facilitate the assessment of the overall accuracy of network reconstruction performed by the four methods. A portion (about a third) of true coefficients for the inputs are intentionally made zero to test the ability of these methods to identify them as insignificant. The input data can be made collinear based on a fraction of co-linearity (zero means all input columns are independent). The algorithm for generating the synthetic dataset has been described in Supplementary material (Section 2).

3.2.1 Effect of noise level: The first study is designed to investigate the effect of increasing noise in the output data. Four sets of outputs corresponding to 5, 10, 20 and 40% noise levels, respectively, were generated from the noise-free (true) output. The size of the input data matrix is $m = 100$, $n = 25$. Supplementary material Fig. S1 shows the fit of predicted against supplied (noisy) output data for LMI, as a representative of all the methods. Increase in the noise is evident. In terms of RMSE (Supplementary material Table S2), LS performs better than PCR, LMI and LASSO. Since the inputs were independent, RMSE_{LS} is comparable with the standard deviation of the noise added to the outputs. Overall, a similar pattern was observed for the training set as well. RMSE_{LMI} is very close to RMSE_{LS} for the training set.

Table 2 demonstrates how the sensitivity, specificity, accuracy and geometric mean of sensitivity and specificity (G) of PCR, LASSO and LMI vary as the noise level increases. In terms of these metrics, LMI exhibits the best overall performance, and PCR the worst. For example, at 10% noise level, $G = 0.71$, 0.90 and 0.97 for PCR, LASSO and LMI, respectively. However, at higher levels of noise, the performance of the three methods becomes similar. At

Table 2 Accuracy, sensitivity and specificity of methods for white noise ($m = 100$, $n = 25$) (based on average of 50 runs)

Noise, %	5	10	20	40
PCR				
ACC.	0.71	0.71	0.70	0.69
Sense.	0.72	0.70	0.66	0.57
Spec.	0.71	0.73	0.77	0.87
G	0.71	0.71	0.71	0.70
LASSO				
ACC.	0.88	0.89	0.88	0.83
Sense.	0.81	0.83	0.83	0.84
Spec.	0.99	0.99	0.96	0.81
G	0.89	0.90	0.89	0.82
LMI				
ACC.	0.97	0.97	0.93	0.79
Sense.	0.95	0.94	0.93	0.90
Spec.	1.00	1.00	0.92	0.63
G	0.97	0.97	0.92	0.75

40% noise level, $G = 0.70$, 0.82 and 0.75, for PCR, LASSO and LMI, respectively. The other metrics (accuracy, sensitivity and specificity) follow a similar trend. For example, both sensitivity and accuracy decrease with increasing noise level for all the three methods, whereas specificity increases for PCR and decreases for LASSO and LMI. We also note that with increasing noise level G of LMI and LASSO decreases more rapidly with increasing noise level than G of PCR does.

The second study is designed to analyse the effect of different types (distribution) of noise in the data. In terms of fractional error in estimating the coefficients, performance of all methods decreases with increase in the level of noise with LASSO exhibiting the smallest decrease. For datasets with white noise, accuracy of LMI and LASSO decreases with the noise level and for datasets with t-distributed and uniformly distributed noise, the accuracy of PCR increases with the noise level whereas it decreases for LMI and LASSO. The details are presented in Supplementary material, Section 3.1.

3.2.2 Variability between realisations of data with white noise: The three network reconstruction methods – PCR, LASSO and LMI – are used to identify significant predictors for 1000 different realisations of white noise added to the output (described in Supplementary material, Section 3.2) and the resulting spectra of the coefficient values for the significant inputs are compared for the different methods (Fig. 2). Lowest standard deviation and closest mean value to the true values from LMI (error $<3\%$, Supplementary material, Table S5) suggest that overall LMI is the most robust method for detecting the significant predictors and computing their values.

3.2.3 Effect of the size of the dataset: Datasets extracted from many biological experiments have different sizes both in the number of data points (cases or experiments) and in the number of inputs. One important measure of the performance of network-reconstruction methods is to evaluate the reliability of each method in terms of the dataset structure (the number of predictors and the ratio of the number of data points to the number of predictors). A method's performance on datasets of different sizes can be used to understand their applicability to datasets of a

doi: 10.1049/iet-syb.2011.0052

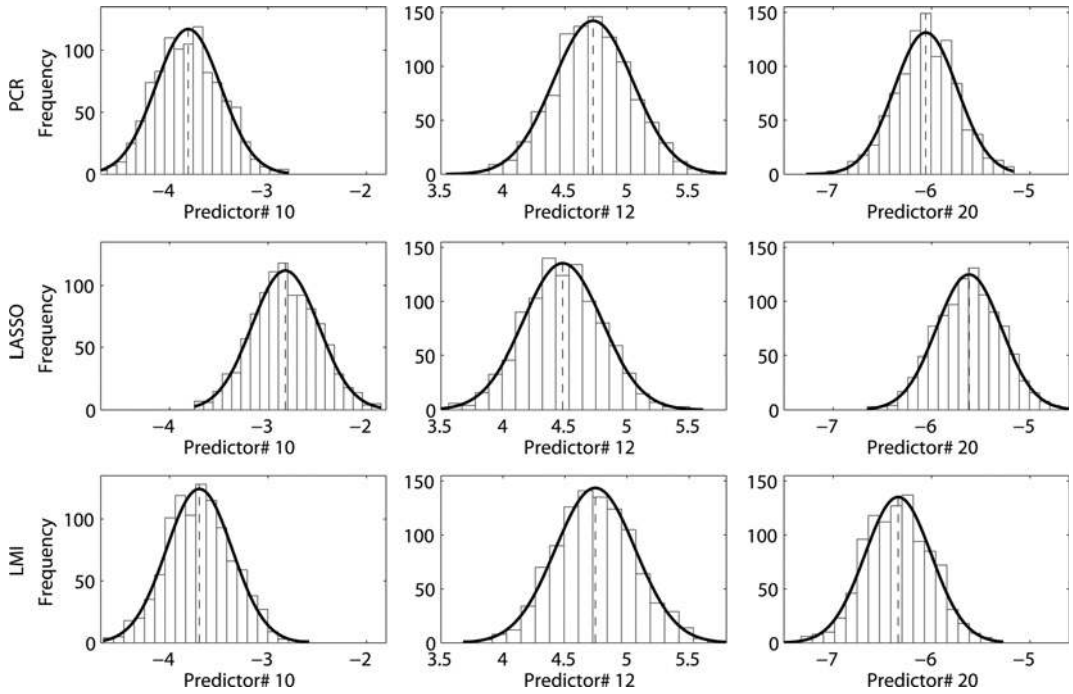

Fig. 2 *Histograms of the three selected significant coefficients from PCR, LASSO and LMI (based on 1000 realisations)*

particular size and structure. In this section, we evaluate the performance of the four methods when (i) the ratio of the number of samples (the number of rows in the input/output data matrices) to the number of inputs decreases, whereas the number of samples remains fixed (comparison between the datasets with $n = 25$ and 50; $m = 100$ in both cases); and (ii) the number of both inputs and samples increase whereas the ratio m/n is kept fixed (comparison between the datasets with $m = 100$, $n = 25$, and $m = 400$ and $n = 100$).

Table 3 shows that in case (a) for both the datasets ($n = 25$ and $= 50$), the fractional error in estimation of coefficients increases with the noise level for PCR and LMI but decreases for LASSO. On the other hand, the geometric mean G decreases with the noise level for all methods and all dataset sizes. PCR exhibits the slowest rate of decrease in G for increasing noise, but it is more sensitive as compared to other methods to the number of inputs. Some more details are presented in the Supplementary material, Section 3.3, case (a). The main result from case (b) is that the performance of the methods are similar when the ratio, $r = m/n$, is comparable. Supplementary material, Section 3.3, case (b) provides more details for this conclusion.

Table 3 Fractional error in estimation of parameters for datasets of different sizes (based on average of 50 runs)

Noise, %	5	10	20	40
$m = 100$, $n = 25$ (same as in Supplementary material Table S3 for white noise)				
PCR	0.15	0.16	0.21	0.26
LASSO	0.48	0.46	0.44	0.44
LMI	0.12	0.15	0.22	0.35
$m = 100$, $n = 50$				
PCR	0.33	0.35	0.36	0.43
LASSO	0.50	0.48	0.46	0.43
LMI	0.16	0.23	0.39	0.71
$m = 400$, $n = 100$				
PCR	0.17	0.18	0.21	0.27
LASSO	0.48	0.47	0.44	0.43
LMI	0.10	0.13	0.20	0.35

3.3 Effect of missing data

The effect of the amount of training data used on the prediction accuracy (through RMSE), both for the experimentally measured data and the synthetic data, has been studied in different scenarios.

For testing the effect of missing data, we re-analysed experimentally measured and synthetic datasets. To this end, up to 60% of data, in increments of 5%, were assumed to be missing. The average computed validation RMSE confirms that a larger amount of missing data leads to less accurate predictions. Detailed procedure and results are presented in Supplementary material, Sections 3.4.1 and 3.4.2.

In terms of computing time, with increasing data sizes, PCR scales very well (sub-linear time complexity), LASSO scales nearly linearly (or sub-linearly) and LMI scales in a quadratic manner (Supplementary material, Section 3.5).

4 Discussion

In this section we provide an overall summary of the results and discuss the important conclusions of our analyses. We discuss the results and analysis of the methods in using the experimentally measured and synthetic data with correlated inputs.

4.1 Comparison of different methods on the experimentally measured data

We found that the different methods identified unique sets of common and distinct predictors for each output. As an example, consider the output interleukin (IL)-6 of the experimental data (PP/cytokine). Fig. 3 illustrates a comparison of the common and distinct sets of predictors

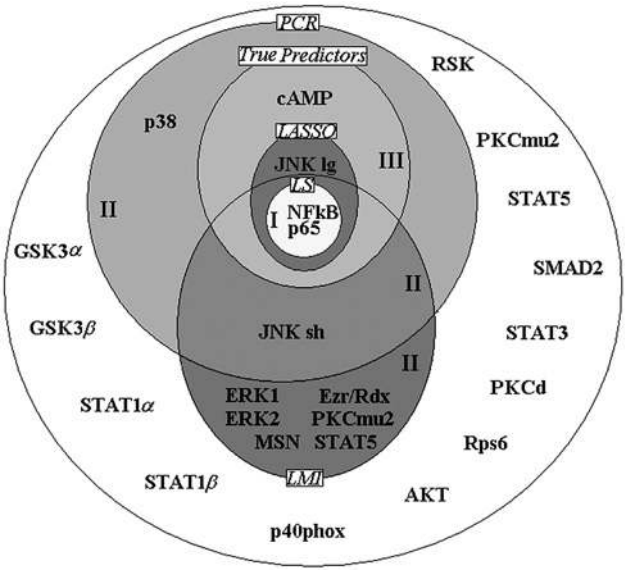

Fig. 3 *Graphical illustration of methods PCR, LASSO and LMI in detection of significant predictors for output IL-6 in PP/cytokine experimental dataset*

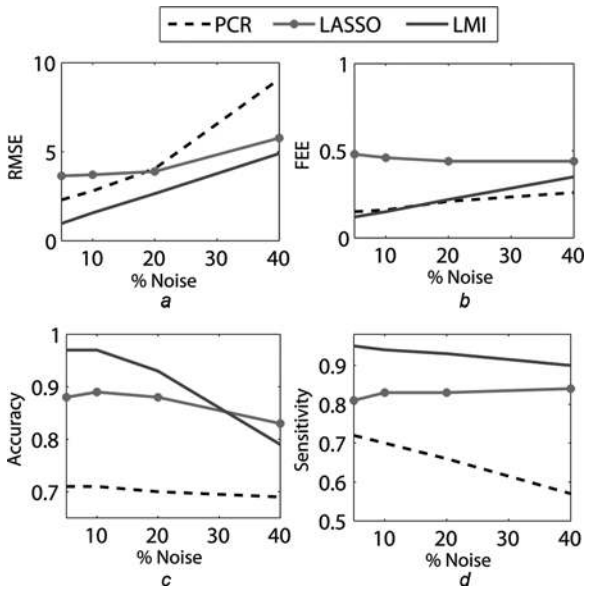

Fig. 4 *Variation of RMSE*

a Fractional error in estimation of coefficients (FEE) (Supplementary material Table S2)
b Accuracy (Supplementary material Table S3)
c Sensitivity (Table 2 and Supplementary material Table S4)
d Table 2 and Supplementary material Table S4 against noise level

identified for IL-6 by these methods. In the Venn diagram of Fig. 3, the sizes of the areas labelled by each method are proportional to the number of significant predictors detected by the corresponding method and the largest area represents all the 22 inputs. The names of the true inputs are known from literature and are listed in the diagram in the labelled areas [4]. Zone I contains the true inputs that are detected by all methods, zone II represents the predictors that are detected by different methods as being significant, but are not true in reality (related to false positive) and zone III corresponds to the true inputs that are detected by only a specific method. For IL-6, the true inputs are: cAMP, JNK lg and NFkB p65. Only the PCR method detects the true input cAMP. Therefore zone III for PCR contains cAMP. On the other hand, all the methods identified NFkB p65 as a significant predictor which belongs to zone I. In summary, zone I provides validation and highlights the common output of all the methods. Zone III highlights the unique outcomes from individual methods. In terms of comparing the different methods on the experimentally measured dataset, PCR had the best overall performance.

In Supplementary material, Section 3.6, a detailed comparison of different methods on the experimental data is also presented with respect to whether or not one of the ligands applied was lipopolysaccharide (LPS). LPS activates Toll-like receptors resulting in very large response, and thus masking the effect of other ligands.

4.2 Comparison of the performance of methods on the synthetic non-collinear data

Fig. 4 provides a summary of the performance of the methods on synthetic data with respect to several criteria. The variation of RMSE and fractional estimation error (FEE) of the parameters against noise level for different methods are shown in Figs. 4*a* and *b* (Supplementary material Tables S2 and S3), respectively. Based on RMSE and FEE, LMI is the best method. PCR is the next best method. LASSO performs least well based on these criteria. The trend of FEE for the LASSO method is opposite to that for PCR and LMI. However, it can be noted that FEE values for the LASSO method are much larger than for PCR and LMI for all noise levels. Mathematical reasoning of either of these trends appears intractable.

Based on accuracy and sensitivity criteria, Figs. 4*c* and *d* (Tables 2 and Supplementary material Table S4) (accuracy value for 20% noise) show that LMI provides maximum accuracy and sensitivity. LASSO is the next best method in this aspect and PCR performs least well although it is not significantly poor.

In this work, we have tried to provide a broad comparison of the methods by using the following metrics – RMSE, fractional error in the estimated coefficients, sensitivity, specificity and related metrics. For example, although the LMI study [12] claims that LMI is a leading method based on true and negative occurrence in reconstruction of the networks, or the LASSO study [14] claims that it is a prominent tool in network reconstruction based on error in finding the coefficients, the metrics are not unique. Therefore even if one method outperforms others based on a certain metric score, it might suffer from scores based on another metric. Table 4 summarises the performance of the methods with respect to different criteria such as ratio of the number of data points to the number of inputs (dimensional ratio), data size, noise level and type of noise contaminating the data and the amount of missing data. The normalised score of each method is described along with how the normalised score is computed to provide a semi-quantitative picture of how the different methods perform. The data reported in Table 4 are based on the analysis of synthetic data. Some more details of comparison with respect to various metrics are presented in Supplementary material, Section 3.7. In summary, it can be concluded that LMI is the best method overall, and is closely followed by LASSO. The method of PCR is comparable with the method of LASSO with respect to several of the above criteria, although it performs poorly with respect to a few other criteria such as accuracy and fractional error in parameter estimation.

doi: 10.1049/iet-syb.2011.0052

Table 4 Summary of the performance of different methods on non-collinear data according to different criteria

Methods/criteria	PCR	LASSO	LMI
Increasing noise			
RMSE (Supplementary material Table S2) score = (average RMSE across different noise levels for LS)/(average RMSE across different noise levels for the chosen method)	✓✓/0.68 degrades gradually with level of noise	✓/0.56	✓✓✓/0.94
standard deviation and error in mean of coefficients. (Supplementary material Table S5): score = 1 − average (fractional error in mean (10, 12, 20) + (std(10, 12, 20)/|true associated coefficients|))	✓✓/0.53	✓✓/0.47	✓✓✓/0.55
Acc./*G* (Table 2 and Supplementary material Table S4) score = average accuracy across different noise levels for chosen method (white noise)	✓/0.70	✓✓/0.87	✓✓✓/0.91 at high noise all similar
fractional error (Table 3 and Supplementary material Table S3) score = 1 − average fractional error in estimating the coefficients across different noise levels for chosen method (white noise)	✓✓/ 0.81	✓/0.55	✓✓/0.78
Types of noise			
fractional error (Supplementary material Table S3) score = 1 − average fractional error in estimating the coefficients across different noise levels and different noise types (20% noise level)	✓✓/0.80	✓/0.56	✓✓/0.79
accuracy and *G* (Table 2 and Supplementary material Table S4) score = average accuracy across different noise levels and different noise types	✓/0.71	✓✓/0.87	✓✓✓/0.91
Dimension ratio/size			
fractional error (Table 3) score = 1 − average fractional error in estimating the coefficients across different noise levels and different ratios (m/n = 100/25, 100/50, 400/100)	✓✓/0.77	✓/0.53	✓✓/0.75
accuracy and *G* (Table 2 and Supplementary material Tables S4, S6 and S7) score = average accuracy across different white noise levels and different ratios (m/n = 100/25, 100/50, 400/100)	✓✓/0.66	✓✓✓/0.83	✓✓✓/0.90
Missing data			
overall (Supplementary material Table S9) let A = Max RMSE of all methods across all level of missing data score = (A − (max RMSE + fractional max deviation in RMSE + fractional std deviation in RMSE)/3)/A	✓✓/0.68	✓✓/0.67	✓✓✓/ 0.71

4.3 Comparison with respect to amount of collinearity in the input dataset

In biological data inputs are not necessarily independent of each other. For example, in the PP/cytokine data, extracellular signal-regulated kinase (ERK) 1 (ERK1) and ERK2 are highly correlated. To compare the performance of the methods in presence of collinearity, a synthetic dataset with the same size and same levels of noise as used in Section 3.2.1 ($m = 100$, $n = 25$) has been used such that 10–30% of the columns are linear combinations of the remaining independent columns. Fig. 5 illustrates the outcome of applying the different methods on noisy datasets with different fractions of dependence among the input columns (see also Supplementary material Tables S12–S16). The surface plots shown in Figs. 5*a* and *b* suggest that the rate of increase of FEE for increase in the noise level is smallest for PCR and largest for LMI; Figs. 5*a*–*c* also show similar trend for FEE with increasing level of collinearity. As shown in Figs. 5*d*–*i*, the accuracy and G metrics show similar trend in that they deteriorate with increasing fraction of collinearity in the datasets for all the methods studied. Further, Figs. 5*d*–*i* suggest that accuracy and G metrics deteriorate (decrease) with increasing noise level for LASSO and LMI; however, they increase for PCR with increasing noise (especially when the fraction of collinearity is considerably high [20–30%]) which is counter-intuitive. Analysis of the number of TP

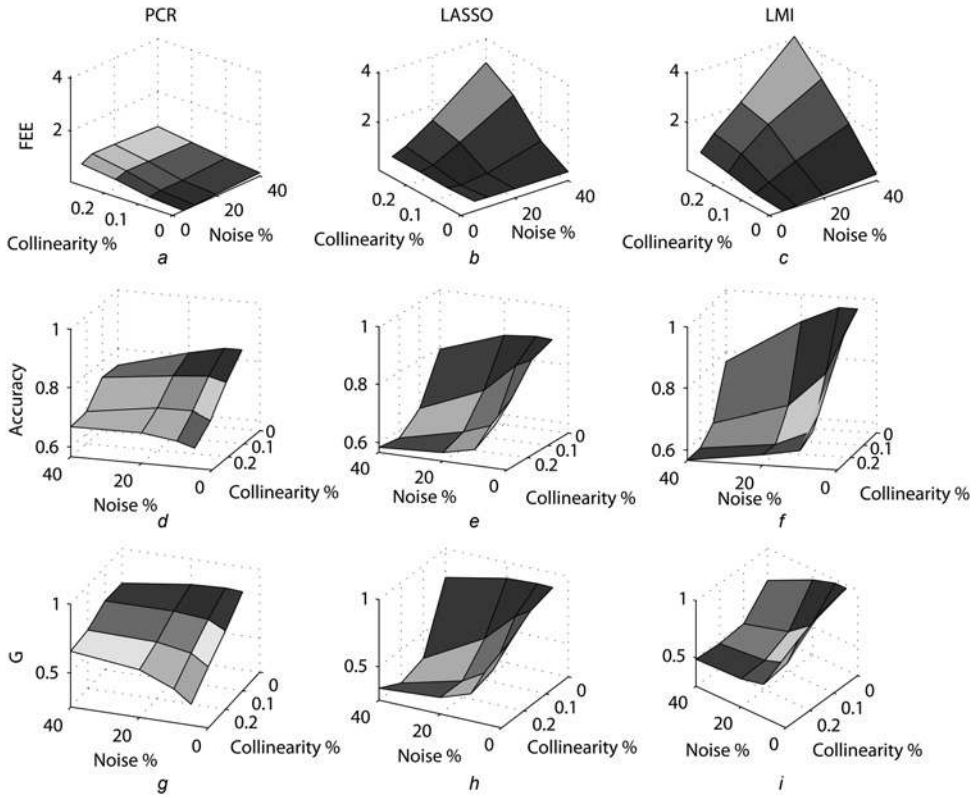

Fig. 5 *Fractional error in estimating the coefficients (Panels a, b, c), accuracy (Panels d, e, f) and geometric mean of sensitivity and specificity (Panels g, h, i) of each method against level of noise and percentage of collinearity in input data*

Table 5 Summary of the performance of different methods on collinear data according to different criteria

Methods/criteria	PCR	LASSO	LMI
Accuracy and *G* (Tables S14 and S15) score = average accuracy across different levels of collinearity for 40% noise level for chosen method (white noise)	✓✓✓/ 0.71	✓✓/0.65	✓✓/0.63 at low noise all similar
Fractional error for the 40% white noise case (Table S13) since fractional error is more than one for LASSO and LMI, the score (= 1 − fractional error) would be negative, and hence it is not reported	✓✓✓/0.43	✓/1.72	✓/2.30

and TN (Supplementary material Table S17) offers help to resolve this anomaly. For PCR, the number of TP decreases with increasing noise as expected. At the same time, the number of TN increases. LMI and PCR exhibit largest and smallest decrease, respectively, in accuracy and *G* value with increase in collinearity across different noise levels. Hence, PCR is the most robust method in dealing with collinear input data with respect to the accuracy and *G* metrics. It is no surprise that PCR deals with collinear data much more effectively than LMI and LASSO, especially when compared with their ability to deal with data with independent input columns. Table 5 summarises the performances of the methods in terms of accuracy, geometric mean of sensitivity and specificity, and fractional error in estimating the parameters for collinear data.

5 Conclusions

Four methods for reconstruction of networks (LS, PCR, LASSO and LMI) described in this work have been applied to two different types of datasets: an experimental dataset and a synthetically constructed dataset. The first dataset comprised of experimental data on PPs/cytokines and the second set contained different artificially synthesised datasets. The least-squares method is seen to perform the best in terms of goodness of fit, but the other three methods are better in capturing most of the true inputs/predictors for the outputs. Various case studies were designed on the artificially synthesised datasets to evaluate the performance of the methods for increasing level of noise, noise type, dimension ratio (size of dataset), missing data (availability of data for building the models) and amount of collinearity in input dataset. LMI performed better than other methods for synthetic data with increasing levels of noise. With regard to the robustness of the methods for different types of the noises, method of LMI exhibited the best performance for reconstruction of the network. LASSO and PCR followed LMI with respect to least fractional error in estimation of coefficients and detection of true predictors, respectively. Based on the dimensional ratio and size of the dataset, PCR and LMI were shown to be the most robust methods in terms of fractional error in estimating the coefficients and LMI and LASSO in terms of accuracy in detection of true predictors. The investigation of the

doi: 10.1049/iet-syb.2011.0052

missing data case demonstrates that the LASSO technique is the most robust method for both the experimentally measured data and synthetic data and LMI results in the lowest prediction error (RMSE) with low/medium percentage of missing data and for medium level of noise. Not surprising is the fact that LS and PCR are the fastest and LMI is the slowest in computational performance. The result of comparisons on datasets with different amount of collinearity in input dataset showed that PCR performs better than other methods overall. This can be owed, at least in part, to the fact that PCR can inherently handle the correlation among the inputs. There are several unanswered questions such as why LASSO performed poorly on synthetic data with increasingly more missing data. Nevertheless, the results presented for the three noise types suggest that all three methods perform similarly overall to a large degree. It is imperative that the user needs to use more than one method and evaluate their pros and cons on the dataset of interest before deciding method of choice.

6 Acknowledgments

This research was supported by the National Heart, Lung and Blood Institute (NHLBI) grant 5 R33 HL087375-02 (SS), National Science Foundation (NSF) grant DBI-0641037 (SS), the NSF collaborative grant DBI-0835541 (SS) and the NSF collaborative grant STC-0939370 (SS). Asadi and Maurya taken equal effort.

7 References

1 Maurya, M., Subramaniam, S.: 'Computational challenges in systems biology', in Liu, E., Lauffenburger, D. (Eds.): 'Systems biomedicine: concepts and perspectives' (Academic Press, San Diego, 2009, 1st edn.), pp. 177–223
2 Bonneau, R., Reiss, D.J., Shannon, P., *et al.*: 'The Inferelator: an algorithm for learning parsimonious regulatory networks from systems-biology data sets de novo', *Genome Biol.*, 2006, **7**, (5), p. R36
3 Janes, K.A., Albeck, J.G., Gaudet, S., Sorger, P.K., Lauffenburger, D.A., Yaffe, M.B.: 'Systems model of signaling identifies a molecular basis set for cytokine-induced apoptosis', *Science*, 2005, **310**, (5754), pp. 1646–53
4 Pradervand, S., MauryaM.R., Subramaniam, S.: 'Identification of signaling components required for the prediction of cytokine release in RAW 264.7 macrophages', *Genome Biol.*, 2006, **7**, (2), pp. R11, 1–14
5 Uckun, S.: 'Model-based reasoning in biomedicine', *Crit. Rev. Biomed. Eng.*, 1992, **19**, (4), pp. 261–292
6 Maurya, M.R., Rengaswamy, R., Venkatasubramanian, V.: 'A systematic framework for the development and analysis of signed digraphs for chemical processes. 1. Algorithms and analysis', *Ind. Eng. Chem. Res.*, 2003, **42**, (20), pp. 4789–4810
7 Hwang, D., Smith, J.J., Leslie, D.M., *et al.*: 'A data integration methodology for systems biology: experimental verification', *Proc. Natl Acad. Sci. USA*, 2005, **102**, (48), pp. 17302–17307
8 Maurya, M.R., Bornheimer, S.J., Venkatasubramanian, V., Subramaniam, S.: 'Mixed-integer nonlinear optimisation approach to coarse-graining biochemical networks', *IET – Syst. Biol.*, 2009, **3**, (1), pp. 24–39
9 Gupta, S., Maurya, M.R., Subramaniam, S.: 'Identification of crosstalk between phosphoprotein signaling pathways in RAW 264.7 macrophage cells', *PLoS Comput. Biol.*, 2010, **6**, (1), p. e1000654
10 Camacho, D., Vera Licona, P., Mendes, P., Laubenbacher, R.: 'Comparison of reverse-engineering methods using an in silico network', *Ann. NY Acad. Sci.*, 2007, **1115**, pp. 73–89
11 Sachs, K., Perez, O., Pe'er, D., Lauffenburger, D.A., Nolan, G.P.: 'Causal protein-signaling networks derived from multiparameter single-cell data', *Science*, 2005, **308**, (5721), pp. 523–529
12 Cosentino, C., Curatola, W., Montefusco, F., Bansal, M., di Bernardo, D., Amato, F.: 'Linear matrix inequalities approach to reconstruction of biological networks', *IET Syst. Biol.*, 2007, **1**, (3), pp. 164–173
13 Montefusco, F., Cosentino, C., Amato, F.: 'CORE-Net: exploiting prior knowledge and preferential attachment to infer biological interaction networks', *IET Syst. Biol.*, 2010, **4**, (5), pp. 296–310
14 Tibshirani, R.: 'Regression shrinkage and selection via the Lasso', *J. Roy. Stat. Soc. B Met.*, 1996, **58**, (1), pp. 267–288
15 Famili, I., Mahadevan, R., Palsson, B.O.: 'k-cone analysis: determining all candidate values for kinetic parameters on a network scale', *Biophys. J.*, 2005, **88**, (3), pp. 1616–1625
16 Karnaukhov, A.V., Karnaukhova, E.V., Williamson, J.R.: 'Numerical matrices method for nonlinear system identification and description of dynamics of biochemical reaction networks', *Biophys. J.*, 2007, **92**, (10), pp. 3459–3473
17 Prill, R.J., Marbach, D., Saez-Rodriguez, J., *et al.*: 'Towards a rigorous assessment of systems biology models: the DREAM3 challenges', *Plos One*, 2010, **5**, (2), p. e9202
18 Bansal, M., Belcastro, V., Ambesi-Impiombato, A., di Bernardo, D.: 'How to infer gene networks from expression profiles', *Mol Syst Biol.*, 2007, **3**, p. 78
19 Cho, K.H., Choo, S.M., Jung, S.H., Kim, J.R., Choi, H.S., Kim, J.: 'Reverse engineering of gene regulatory networks', *Syst. Biol., IET*, 2007, **1**, (3), pp. 149–163
20 Lu, R., Markowetz, F., Unwin, R.D., *et al.*: 'Systems-level dynamic analyses of fate change in murine embryonic stem cells', *Nature*, 2009, **462**, (7271), pp. 358–362
21 Hsiao, A., Worrall, D.S., Olefsky, J.M., Subramaniam, S.: 'Variance-modeled posterior inference of microarray data: detecting gene-expression changes in 3T3-L1 adipocytes', *Bioinformatics*, 2004, **20**, (17), pp. 3108–3127
22 The Alliance for Cellular Signaling (AfCS): Available from http://www.signaling-gateway.org
23 Geladi, P., Kowalski, B.R.: 'Partial least-squares regression – a tutorial', *Anal. Chim. Acta*, 1986, **185**, pp. 1–17
24 Barnes, P.J., Karin, M.: 'Nuclear factor-kappaB: a pivotal transcription factor in chronic inflammatory diseases', *N. Engl. J. Med.*, 1997, **336**, (15), pp. 1066–1071
25 Ozato, K., Tsujimura, H., Tamura, T.: 'Toll-like receptor signaling and regulation of cytokine gene expression in the immune system', *Biotechniques*, 2002, **33**, (66–8), p. 70, 2 passim

Supplementary Material

Comparison of Statistical and Optimization-Based Methods for Data-Driven Network Reconstruction of Biochemical Systems

Behrang Asadi [a,b,1], Mano Ram Maurya[b,1], Daniel M. Tartakovsky[a,2] and Shankar Subramaniam [b,c,3]

[a]Department of Mechanical and Aerospace Engineering
[b]Department of Bioengineering
[c]Departments of Chemistry & Biochemistry, Cellular and Molecular Medicine and Graduate Program in Bioinformatics
University of California, San Diego, La Jolla, CA, USA

E-mail addresses: basadi@ucsd.edu, mano@sdsc.edu, dmt@ucsd.edu and shankar@ucsd.edu

[1] Equal effort.
[2] Corresponding author; E-mail: dmt@ucsd.edu, Phone: (858) 534-1375, Fax: (858) 534-7599.
[3] Corresponding author, E-mail: shankar@ucsd.edu, Phone: (858) 822-0986, Fax: (858) 822-3752.

1. Methodology

Details of the four methods of network reconstruction are described in this section as following.

1.1 Standard Least Squares

The Least Squares method minimizes the root-mean-squared error (RMSE) on the given (training) dataset where RMSE can be computed as

$$RMSE_{LS} = \sqrt{\frac{1}{m}\sum_{i=1}^{m}(y_i - y_{i,p})^2} = std(\mathbf{y} - \mathbf{y}_p) \times \sqrt{(m-1)/m}\,.$$

To test whether an element of the parameter vector $\hat{\mathbf{b}}$ is statistically different from zero, a two-tailed *t*-test can be performed on the coefficient (1). First, the standard deviation of the model parameters σ_b is computed as

$$\sigma_{\mathbf{b},LS} \approx diag((\mathbf{X}^T\mathbf{X})^{-1})^{1/2} \times RMSE_{LS} \times (m/v)^{1/2}; v = m-n-1 \qquad \text{(S1)}$$

The ratio $r_j = b_j / \sigma_{B,LS,j}$ is computed for the *j*th input. The predictor is considered significant if this ratio is greater than $r_j = \text{tinv}(1-\alpha/2, v)$, where tinv denotes the inverse of cumulative *t*-distribution, and $\alpha = 0.01$ for a confidence level of 0.99 (1, 2).

1.2 Principal Component Regression (PCR)

PCR is based on the principal component analysis and is employed when $\mathbf{X}^T\mathbf{X}$ is (nearly) singular, so that one or more of its eigenvalues are (close to) zero. Then, the principal components corresponding to only the first several eigenvalues (starting with the largest) are used.

PCR consists of the follow steps.

1) Given normalized input data $\mathbf{X}$ and mean-centered output data $\mathbf{y}$, let γ_j be jth eigen value (j = 1, 2, …, n) and $\mathbf{v}_j$ be the corresponding eigen vector of the covariance matrix $\mathbf{C} = \mathbf{X}^T\mathbf{X}/(m-1)$, arranged in the order of decreasing eigen values. Let $\Gamma_k = \{\gamma_j, j = 1,...k\}$ be a set of k largest eigenvalues and $\mathbf{V}_k = [\mathbf{v}_1,...\mathbf{v}_j,...,\mathbf{v}_k]$ be a set of corresponding eigenvectors. Calculate the matrix of latent variables $\mathbf{T}_k$,

$$\mathbf{T}_k = \mathbf{X} \times \mathbf{V}_k . \quad \text{(S2)}$$

2) Create a PCR model based on the k latent variables as

$$\mathbf{b}_k = \mathbf{V}_k \times \mathbf{\Gamma}_k^{-1} \times \mathbf{T}^T \times \mathbf{y}, \quad \mathbf{y}_p = \mathbf{X} \times \mathbf{b}_k,$$

$$RMSE_{PCR} = std(\mathbf{y} - \mathbf{y}_p) \times ((m-1)/m)^{1/2}. \quad \text{(S3)}$$

The number of latent variables in PCR can be determined either by cross-validation or by specifying the fraction of the cumulative variance (FCV) captured (say, $0.8 < FCV_k < 0.95$).

Partial least squares (PLS) is a method similar to PCR, with the difference being that both matrix **X** and vector **y** (rather than only **X**) are used to construct the set of linear combinations of significant inputs for regression. A detailed description of PLS can be found in (3, 4). Statistical significance of the model parameters **b** determined with either of these two methods can be tested by first estimating their standard deviation σ_B and then using a two-tailed *t*-test to compare their ratios. For PCR,

$$\sigma_{b,k} \approx diag(\mathbf{V}_k \times \mathbf{\Gamma}_k^{-1} \times \mathbf{V}_k^T)^{1/2} \times RMSE_{PCR} \times (m/v)^{1/2}; v = m-k-1$$

and $r_{j,k} = b_{j,k} / \sigma_{B,j,k}$ for the *j*th input when *k* latent vectors are used. Average of $r_{j,k}$ over k, such that fractional cumulative variance (FCV), $0.8 < FCV_k < 0.95$, is computed. The input is considered significant if $r_{j,k} > \text{tinv}(1-\alpha/2, v)$, where $v = m-k-1$ and $\alpha = 0.01$ for a significance level of 0.99 (2).

1.3 Least Absolute Shrinkage and Selection Operator (LASSO)

LASSO recasts the problem of network reconstruction into an optimization problem analogous to (2), but with an additional nonlinear constraint. An abstract formulation of LASSO is given by

$$\hat{\mathbf{b}} = \arg\min\{e^2 = (\mathbf{y} - \mathbf{Xb})^T(\mathbf{y} - \mathbf{Xb})\} \quad s/t \quad \sum_j \left|\hat{b}_j\right| \le t. \tag{S4}$$

The parameter *t* determines the amount of "shrinkage" in the estimation of $\hat{\mathbf{b}}$. The constraint in (S4) leads to shrinkage in absolute value of the parameters. For certain values of *t* ($t \le t_{LS}$, $t_{LS} = \sum_j \left|\hat{b}_j\right|$ obtained from LS*)* the algorithm shrinks some of the larger parameter-values and sets some of the parameters to zero, thus identifying a parsimonious model. The constrained optimization problem (S4) is solved by employing a quadratic programming approach (e.g., the interior-point method used in this paper) (5).

1.4 Linear Matrix Inequalities (LMI)

The basic idea of LMI is to convert a nonlinear optimization problem into a linear optimization problem (6). This method has been used to reconstruct and minimize dynamic networks (7-9). To facilitate the comparison of the four network-reconstruction approaches, the LMI method is also applied to static data. Equations in this section are for $p \geq 1$. LMI replaces optimization problem (2) with

$$\min_{B \in \mathbb{R}^{n \times p}}(e) \; s/t \; (\mathbf{y} - \mathbf{Xb})(\mathbf{y} - \mathbf{Xb})^T < e\mathbf{I}_{m \times m} . \tag{S5}$$

The constraint in (S5) is nonlinear with respect to $\hat{\mathbf{B}}$. A congruence transformation of (8) yields an LMI representation

$$\begin{pmatrix} -e\mathbf{I}_{m \times m} & \mathbf{y} - \mathbf{X}\hat{\mathbf{b}} \\ (\mathbf{y} - \mathbf{X}\hat{\mathbf{b}})^T & -\mathbf{I}_{p \times p} \end{pmatrix} < 0 . \tag{S6}$$

Prior knowledge (e.g., $b_{12} > 0$ or $b_{31} = 0$) can be added as LMI constraints,

$$v_i^T \mathbf{B} u_j + u_j^T \mathbf{B}^T v_i = (><)0 , \text{ (general form of constraints for multiple outputs)} \tag{S7}$$

where $v_i = \begin{cases} v_r = 0, r \neq i \\ v_r = 1, r = i \end{cases}$ and $u_i = \begin{cases} u_r = 0, r \neq i \\ u_r = 1, r = i \end{cases}$ are respectively $n \times 1$ and $p \times 1$ column-vectors constraining the element $\hat{b}_{ij}$. The LMI assembled by equations (S6) and (S7) can be solved simultaneously to obtain optimal estimate for $\hat{\mathbf{b}}$ (10).

The normalized matrix of parameters $\overline{\hat{\mathbf{b}}}$ is calculated by dividing each element by the L_2-norm of its row and column,

$$\overline{\hat{b}}_{ij} = \left|\hat{b}_{ij}\right| / \left\|\hat{b}_{i..}\right\|_2 \left\|\hat{b}_{:j}\right\|_2 \quad . \tag{S8}$$

If $\overline{\hat{b}}_{ij}$ becomes smaller than a threshold (say, r_{LMI}), then the corresponding parameter is nullified (deemed insignificant). Since all analyses have been carried out based on a single output at a time, constraint described in Eqn. S7 can be determined by only one vector (**v** or **u**) and the matrix of parameters morphs to a column vector. A further discussion of this method can be found in (8).

1.5 Metrics for comparing the methods

We used two data sets to evaluate the performance of the four network-reconstruction methods presented in above Supplementary Material Sections 1.1-1.4. The first set is experimental data measured in macrophage cells stimulated with a defined set of ligands; for the network construction, we use Phosphoproteins (PP) as inputs and Cytokines as outputs (11). The data is made available by the Alliance for Cellular Signaling (11). The second set consists of synthetic data. We reconstruct the networks from 80% of each dataset (called a training set; randomized but pairs of rows in the input-output maintained) and then use the remaining portion (20%, called a test set) to validate the reconstruction. We have found that the models based on 100% of the data and 80% of the data are similar. Root-Mean-Squared-Error (RMSE) on the test set, and the number and the identity of the significant predictors (network parameters) selected is used as the basic metric to evaluate the performance of each method.

Once the network parameters have been identified with each method, we utilized the two different metrics described below to evaluate both the overall performance of the methods and the validity of the computed parameters.

1.5.1. Fractional error in the parameter values

The fractional error of the estimated coefficients with respect to the true synthetic parameters serves as the first measure. It can be computed as

$$\Delta b_{frac,j} = mean\left(\left|\frac{b_{method,j}}{b_{true,j}} - 1\right|\right) \tag{S9}$$

where $\mathbf{b}_{\text{method}}$ and $\mathbf{b}_{\text{true}}$ are the estimated and true values, respectively, of the parameters for a chosen output. The "mean" is computed over the coefficients for all the n inputs.

In order to avoid the effect of very small true value of the coefficients in the denominator, the parameters smaller than 10% of the standard deviation of all parameter values were set to 0 when generating the synthetic data.

1.5.2. Sensitivity, specificity and related metrics

The second measure used in our analysis is an aggregate of the several measures of model evaluation that are described below. The latter rely on the following definition of *positive and negative occurrences* in a reconstructed network (with respect to one output at a time):

a) True Positive (TP): the number of true predictors identified as significant in an estimated model (found in the correct network; significant in the estimated model).

b) False Positive (FP): the number of false predictors identified as significant in the estimated model (not found in the correct network; significant in the estimated model).

c) True Negative (TN): the number of predictors not present in the correct network and identified as insignificant in the estimated model as well.

d) False Negative (FN): the number of predictors present in the actual network but identified as insignificant in the estimated model.

The performance of each network-reconstruction method can be evaluated according to their accuracy, sensitivity, and specificity, which are defined as

$$Accuracy: \frac{TN+TP}{TN+TP+FN+FP}$$
$$Sensitivity: \frac{TP}{TP+FN}$$
$$Specificity: \frac{TN}{TN+FP}$$

In other words, accuracy is the fraction of relevant (the corresponding coefficient in the "exact", noise-free model is non-zero) and irrelevant (the corresponding coefficient in the 'exact", noise-free model is zero) predictors detected correctly. Sensitivity is the fraction of relevant predictors detected correctly. Specificity is the fraction of irrelevant predictors detected correctly.

Since the measure of accuracy is meaningful only if the numbers of relevant and irrelevant predictors is about the same, the accuracy metric and the geometric mean of sensitivity and specificity ($G = \sqrt{sensitivity \times specificity}$) can be used as a robust measure to benchmark the methods (12). The definitions of sensitivity and specificity imply, that, $0 \le G \le 1$. The closer G is to 1, the higher the percentage of correct detections.

2. Algorithm for generating synthetic data

The synthetic datasets referred to in the main text were generated as follows.

- Generate a random zero-mean, unit standard deviation $m \times n$ matrix $\mathbf{X}$.
- Generate a random zero-mean, unit standard deviation $m \times n'$ matrix $\underline{\mathbf{X}}$ where n' = (1 - fraction of co-linearity) × n.
- Generate a random ($n' \times (n-n')$) collinearity coefficient matrix ($\boldsymbol{C}$). Set $\boldsymbol{X}' = \mathbf{X} \times \boldsymbol{C}$. X' is of size $m \times (n-n')$ and its columns are dependent on $\boldsymbol{X}$.
- Augment $\boldsymbol{X}'$ to $\underline{\boldsymbol{X}}$ to create the final input matrix $\boldsymbol{X}$ of size $m \times n$. 5% noise can be added to $\boldsymbol{X}$ if desired.
- Generate a random $n \times p$ ($p = 1$) coefficient matrix, and set the (randomly selected) one-third of these coefficients to 0 ($\mathbf{b}$).
- Multiply $\mathbf{X}$ and $\mathbf{b}$ to obtain matrix of true output data $\mathbf{y}_0$.
- Generate a zero-mean, unit standard deviation $m \times p$ "raw" noise matrix from the following distributions:
    - $N(0,1)$ for white noise,
    - $t(\nu) / (\nu / (\nu - 2))^{1/2}; \nu = 10$ for T-distributed noise, and
    - $(U[0,1] - 0.5) / (1/12)^{1/2}$ for shifted-uniform noise;
- Scale the raw noise matrix with the factor $f \times std(\mathbf{y}_0)$ (set $f = 0.05$ for 5% noise level);
- Add the resulting noise matrix to $\mathbf{y}_0$ to obtain the output data $\mathbf{y}$.

Table S1. RMSE on training set for different methods (PP/cytokine data). The LS method has the smallest RMSE due to the absence of constraints in the optimization problem, but RMSEs of the other three methods (PCR, LMI and LASSO) are also comparable.

Output	**G-CSF**	**IL-1α**	**IL-6**	**IL-10**	**MIP-1α**	**TNFα**
LS	0.73	0.41	0.61	0.61	1.30	0.99
LS_sig	1.16	0.56	0.80	0.80	2.30	1.27
PCR	0.79	0.44	0.73	0.76	1.45	1.06
LASSO	0.92	0.56	0.79	0.72	1.84	1.27
LMI	0.76	0.41	0.67	0.67	1.34	1.09

Table S2. RMSE on all data: methods vs. noise level (synthetic data).

Noise %	**5**	**10**	**20**	**40**
On training set				
LS_sig	0.94	1.5	2.47	4.22
PCR	2.28	2.79	3.99	6.86
LASSO	3.46	3.52	3.91	5.12
LMI	0.57	1.31	2.39	4.23
On validation set				
LS_sig	0.95	1.64	2.66	4.82
PCR	2.3	2.8	4.05	9.08
LASSO	3.64	3.71	3.89	5.77
LMI	0.97	1.57	2.65	4.9

3 Results

In support of the results presented in Section 3 of main text, here we present the details of the analysis of effect of noise distribution, size of the dataset, and missing data in comparing the different methods.

3.1 Effect of noise distributions of (type of noise in) the output data

In the first study (main text Section 3.2.1), the effect of increasing noise in the output data has been investigated. The second study (this study) is designed to elucidate the effect of different types of noise on the performance of different methods in reconstructing the network, i.e., on identification and estimation and of significant predictors. We considered biologically relevant white noise (noise type 1), *t*-distributed noise (noise type 2), and shifted uniform noise (noise type 3). These were added to the noise-free output as described in Supplementary Material Section 2.

Table S3 shows the fractional error (defined in Supplementary Material Section 1.5.1, Eqn. S9) in the computed values of the coefficients relative to the true values. These results are based on the average from 50 runs. The different types of noises result in a similar (and obvious) trend: the fractional error in estimating the coefficients by the three methods increases with the level of noise. LASSO exhibits the smallest increase.

Table S4 lists the values of sensitivity, specificity, and other relevant indicators of a method's performance for different types of noises. For noise types 2 and 3, only the values for 5% and 20% are listed, which serves adequately for comparison purposes. The relevant results for noise type 1 (white noise) are discussed in main text Section 3.2.1 (Table 2). Here we focus on the geometric mean of sensitivity and specificity, *G*, and accuracy as two main metrics for comparing the performance of the three methods. The results reveal that for the output data with the type 1 noise, both the accuracy and *G* of LMI and LASSO decrease with the noise level. For the type 2 and 3 noise, both the accuracy and *G* of PCR increase with the noise level, but they decrease for LMI and LASSO.

Table S3. Effect of noise type**:** Fractional error in estimating the coefficients [based on average of 50 runs].

Noise %	5%	10%	20%	40%
White noise				
PCR	0.15	0.16	0.21	0.26
LASSO	0.48	0.46	0.44	0.44
LMI	0.12	0.15	0.22	0.35
Noise from t-distribution				
PCR	0.16	0.17	0.20	0.28
LASSO	0.48	0.47	0.43	0.44
LMI	0.13	0.16	0.21	0.37
Uniform noise with mean 0				
PCR	0.16	0.16	0.18	0.26
LASSO	0.48	0.47	0.44	0.45
LMI	0.11	0.13	0.21	0.38

Table S4. Accuracy, Sensitivity, and Specificity of methods for noise with t-distribution (noise-type = 2) and uniform distribution (noise-type = 3) (m = 100, n = 25) [based on average of 50 runs].

	Noise type = 2		Noise type = 3	
Noise %	5	20	5	20
PCR				
ACC.	0.74	0.73	0.72	0.72
Sense.	0.73	0.69	0.74	0.68
Spec.	0.75	0.81	0.70	0.78
G	0.73	0.74	0.71	0.73
LASSO				
ACC.	0.89	0.89	0.89	0.89
Sense.	0.82	0.84	0.82	0.85
Spec.	0.99	0.95	1.00	0.95
G	0.90	0.89	0.90	0.90
LMI				
ACC.	0.97	0.92	0.98	0.93

Sense.	0.94	0.94	0.97	0.95
Spec.	1.00	0.89	1.00	0.91
G	0.97	0.91	0.98	0.93

The tuning parameters in LASSO and LMI (the threshold parameters t and r_{LMI}, respectively) were identified through k-fold (with k = 10) cross-validation, for different noise levels in the synthetic dataset. The optimal value of the tuning parameter was computed by analyzing the dependence of the prediction error from the ten-fold cross-validation estimation on the tuning parameter. The tuning parameter in the LASSO method, t, varies between 0 (the null model without predictors) and 1 (LS model with all predictors). Supplementary Material Figure S2 illustrates how the RMSE (its mean and standard deviation across the k-folds) varies over the validation set. For LMI, the upper bound of the tuning parameter (r_{LMI} = 1) corresponds to the null model and the lower bound (r_{LMI} = 0) corresponds to the full model.

Following (13), we define the optimal parameter value to be the smallest value of the tuning parameter that is within one standard deviation from the lowest RMSE on the RMSE vs. tuning-parameter curve (e.g., Supplementary Material Figure S2). This definition provides a conservative estimate of the tuning parameter that prevents over-fitting. Applying this algorithm to obtain the optimal values of the tuning parameters in LASSO and LMI, we obtain $t \approx 2/3$ and $r_{\mathrm{LMI}} \approx 1/3$, respectively. These values of t and r_{LMI} did not vary much with the noise level, which indicates their robustness.

This procedure is repeated for different types of noise in the output data (synthetic data case) and used in the subsequent studies as appropriate.

3.2 Variability between realizations of data with white noise

The data set is generated as follows. First the true input-output (noise free) data are generated. About one third of the coefficients are set to 0. Then 1000 realizations of the noisy output data are generated by adding white noise. Four levels of noise (5, 10, 20 and 40%) were considered, but only representative results from the 20% noise set are presented below.

The shape of the distributions of the coefficient values for the predictors 10, 12, and 20 is similar. These distributions are approximately Gaussian, with different means but similar standard deviations (Supplementary Material Table S5). For a selected input, the discrepancy in the means is due to the fact that different methods give rise to different sets of significant input predictors. The similarity of the standard deviations is a reflection of the consistency in the level of noise in the predictors obtained with the three methods. The LMI and PCR methods result in the lowest standard deviations. This implies that PCR and LMI are relatively more robust in detecting significant predictors for output data with moderate level of noise. While not shown here, for common predictors, we found that higher levels of noise result in higher standard deviations in the values of the coefficients.

Table S5. Mean and standard deviation in the histograms of the coefficients computed with PCR, LASSO, and LMI.

Method	Predictor No.	**10**	**12**	**20**
	True value	-3.40	5.82	-6.95
PCR	Mean	-3.81	4.73	-6.06
	Std.	0.33	0.32	0.32
	Frac. Err. in mean	0.12	0.19	0.13
LASSO	Mean	-2.82	4.48	-5.62
	Std.	0.34	0.32	0.33
	Frac. Err. in mean	0.17	0.23	0.19
LMI	Mean	-3.70	4.74	-6.34
	Std.	0.34	0.32	0.34
	Frac. Err. in mean	0.09	0.18	0.09

3.3 Comparison of methods on the data sets of different sizes

Two cases studies are designed as discussed below.

Case (a):

We compare the methods' performance on two synthetic datasets with m = 100 data points. The number of input columns in the two datasets is n = 25 (dataset 1, the base case with white noise presented in Section 3.2.1) and n = 50 (dataset 2), respectively. Table 2 lists the sensitivity and other metrics for the n = 25 (dataset 1), while Table 3 provides the corresponding fractional error metric. Table 3 also contains these metrics for the n = 50 case. In both cases, the fractional error in estimation of coefficients increases with the noise level for PCR and LMI but decreases for LASSO. On the other hand, the geometric mean G decreases with the noise level for all methods and all dataset sizes. PCR exhibits the slowest rate of decrease in G. Comparing the results in Section 3.2.1 for the type 1 noise (Table 2) with those presented in Supplementary Material

Table S6 reveals that, as the number of inputs increases, *G* of PCR decreases faster than *G* of both LASSO and LMI does. In other words, PCR is more sensitive to the number of inputs. Also, comparison of Tables 2 (dataset 1) and S5 (dataset 2) shows that for the same noise type (i.e., white noise) and the same number of samples (i.e., $m = 100$), both accuracy and *G* decrease as the number of inputs increases from dataset 1 (n = 25) to dataset 2 (n = 50), with PCR exhibiting the smallest rate of degradation (for dataset 1, G and accuracy ranges across noise levels are: PCR = [0.71 - 0.69], LMI = [0.90 - 0.82], LASSO = [0.97 – 0.75]).

Case (b):

To investigate the impact of increasing the number of inputs as well as the number of samples (with same dimension ratio m/n = 4), we created the third dataset with $m = 400$ and $n = 100$ (dataset 3). The fractional error (Table 3), sensitivity and other performance metrics (Table S7) of the network reconstruction methods are consistent with their counterparts observed for the previous two datasets. As the noise level increases, the fractional error in estimation of coefficients increases except for the LASSO which is decreasing and the geometric mean *G* decreases for all methods. A comparison of results from datasets 1 and 3 in Table 3, and between Tables 2 and S7 suggests that the performance of the methods are similar when the ratio, $r = m/n$, is comparable.

For example, comparing the fractional error in estimating the coefficients of dataset 1 and data 3 in Table 3 reveals that LASSO is the most robust methods when the system size increase in a manner that preserves the dimensional ratio $r = m/n$. The comparison of Tables 2 and S7 implies that the change in the number of samples, *m*, does not affect the *G* and accuracy of the methods when the dimensional ratio $r = m/n$ remains unchanged.

Table S6. Accuracy, Sensitivity, and Specificity (m = 100, n = 50).

Noise Type = 1, 50 runs				
PCR				
Noise %	5	10	20	40
ACC.	0.65	0.64	0.63	0.58
Sense.	0.52	0.52	0.47	0.36
Spec.	0.84	0.84	0.87	0.92
G	0.66	0.65	0.64	0.57
LASSO				
ACC.	0.83	0.83	0.79	0.71
Sense.	0.77	0.78	0.79	0.81
Spec.	0.91	0.90	0.78	0.55
G	0.84	0.84	0.79	0.66
LMI				
ACC.	0.96	0.91	0.80	0.69
Sense.	0.93	0.89	0.89	0.85
Spec.	1.00	0.94	0.67	0.44
G	0.96	0.92	0.77	0.61

Table S7. Accuracy, Sensitivity, and Specificity (m = 400, n = 100).

Noise Type = 1, 50 runs				
PCR				
Noise %	5	10	20	40
ACC.	0.71	0.71	0.71	0.68
Sense.	0.71	0.70	0.67	0.56
Spec.	0.71	0.73	0.77	0.86
G	0.71	0.71	0.72	0.69
LASSO				
ACC.	0.87	0.88	0.88	0.83
Sense.	0.79	0.80	0.82	0.81
Spec.	0.99	0.99	0.98	0.86
G	0.89	0.89	0.89	0.83
LMI				
ACC.	0.97	0.97	0.92	0.79
Sense.	0.95	0.95	0.93	0.90
Spec.	1.00	1.00	0.91	0.62
G	0.97	0.97	0.92	0.74

3.4 Effect of missing data

Detailed analysis of effect of the amount of training data used on the prediction accuracy (through RMSE) is studied in this section for both the experimentally measured and synthetic data.

3.4.1 Experimentally measured dataset

To test the effect of missing data, we analyzed the output GCSF from the experimentally measured data set (11). This analysis consists of the following steps. First, 5-60% data (in increments of 5%) were assumed to be missing. Second, the remaining data were used for parameter estimation, and RMSE was computed on the test data (20% of the data). Third, this was repeated 100 times by randomly choosing the selected fraction of missing data. Finally, an average RMSE was computed which is shown in Supplementary Material Figure S3. As expected, the larger the amount of missing data, the less accurate (higher RMSE) the predictions are. Table S8 lists the corresponding fractional standard deviation, std($RMSE_{0\text{-}60\%}$) / $RMSE_{0\%}$, and fractional maximum deviation, max($RMSE_{x\%}$ - $RMSE_{0\%}$) / $RMSE_{0\%}$, for the four methods. Table S8 suggests that PCR and LASSO are more robust than LS and LMI for increases in the fraction of missing data, especially when less than half the data is used for training the model. However it can be noted from Supplementary Material Figure S3 that for missing data fraction less than 45%, both PCR and LASSO result in larger RMSE as compared to the other two methods.

Table S8. Effect of missing data with PP/cytokine data (validation set).

Method	Fractional standard deviation of RMSE across different fractions of missing data	Fractional max deviation of RMSE across different fractions of missing data
LS	0.12	0.40
PCR	0.07	0.20
LASSO	0.05	0.14
LMI	0.10	0.32

3.4.2 Synthetic dataset

We conducted the analysis described above to ascertain the performance of the four network reconstruction methods on the synthetic data with 20% noise. The results are shown in Supplementary Material Figure S4. In terms of RMSE, LMI and LS with significant predictors yield the best performance, while LASSO provides the worst. At the same time, RMSE of LASSO exhibits the slowest rate of increases as the amount of missing data increases. PCR demonstrates a stable performance when the percentage of missing data is low (less than 30%). Table S9 lists the fractional standard deviation and fractional maximum deviation for each method. LASSO is the most robust method in terms of the fractional standard deviation, while other methods are comparable. LMI and LS with significant predictors perform well when the percentage of missing data is small (less than 20%). LASSO performs slightly worse for the experimentally measured data, and PCR for the synthetic data. Overall, given the robust behavior of LASSO to the missing data in the network reconstruction from experimental data, its use is recommended.

Table S9. Effect of missing data with Synthetic data (validation set).

Method	Fractional standard deviation of RMSE across different fractions of missing data	Fractional max deviation of RMSE across different fractions of missing data
LS	0.06	0.21
PCR	0.08	0.21
LASSO	0.02	0.06
LMI	0.08	0.25

3.5 Computational efficiency

The simulations were run on a Dual-Core Intel Pentium IV processor with 2.66GHz processing speed per core, 4MB of cache, and 3 GB of RAM. An estimate of the CPU time used by each method is summarized in Supplementary Material Table S10. The time reported also includes the time taken for some text-output on the screen in Matlab®. A trade-off between robustness and computational time is apparent. LASSO and LMI are slower than LS and PCR by a factor of hundred for the datasets we analyzed. With increasing data sizes, PCR scales very well (sub-linear time-complexity) LASSO scales nearly linearly (or sub-linearly) and LMI scales in a quadratic manner. A plot of computing time versus the data-size is shown in Supplementary Material Figure S5.

Table S10. Processing time vs. methods.

Method	Processing time(sec)		
	(m = 100, n = 25)	(m = 100, n = 50)	(m = 400, n = 100)
PCR	4.67 E-2	4.34 E-1	4.38 E-1
LASSO	5.86 E 0	1.25 E+1	4.36 E+1
LMI	1.97 E+1	9.56 E+1	7.63 E+2

3.6 Comparison on Toll and Non-Toll data sets

In order to demonstrate a practical comparison between the methods given in this paper, we have applied the methods on experimental data sets in two different ways. In the PP/cytokine experiment data (11), both single- and double-ligands (stimuli) were applied. One of the ligands is lipopolysaccharide (LPS). LPS activates Toll-like receptors (TLR) and hence it is also referred to as a TLR-ligand. Since LPS results in very large response, it can mask the effect of other ligands. Hence, in our application, we tested the methods in two ways. In one case, we included data from all ligands (Toll data). In the second case, all the experiments with LPS as one of the ligands were excluded (non-Toll data). Both sets have the same number of inputs (22) and outputs (7).

Supplementary Material Table S11 shows the true and identified predictors from each method for Toll data (all ligands included). As an example, for the output MIP-1α, true inputs have been considered as cAMP, JNK lg, p38, and STAT1α among which different methods could identify some or all of the true predictors. Least squares with only significant predictors was able to identify one, PCR three, LASSO three, and LMI two predictors. Although PCR and LASSO have identified the same set of true predictors (hence same value of sensitivity of 3/4), there is still difference between the numbers of predictors identified as significant in each of these two methods. PCR has identified 11 significant predictors and LASSO has identified 6 among which only three of them are true predictors (TP).

Supplementary Material Table S12 shows the identified predictors from each method for PP/Cytokine dataset when the data from TLR ligand are removed i.e. experiments in which LPS is a ligand are not included. In order to compare the results of each method to others, predictors

identified by PCR are considered as the base for the comparison. For instance, consider the output TNFa. The significant identified predictors for PCR are cAMP, p38, p40Phox, and RSK out of 22 predictors. The method of least squares has identified 3 predictors where 2 of them are in common with the outcome of PCR. LASSO and LMI have identified 9 and 4 significant predictors with 3 and 2 predictors being common with that from PCR, respectively.

Table S11. Identified predictors by PCR, LASSO, and LMI versus true predictors for Toll data.

Output	G-CSF	IL-1α	IL-6	IL-10	MIP-1α	RANTES	TNFα
TRUE	*JNK lg* *p38* *NFkB p65*	*JNK lg* *NFkB p65* *STAT1β*	*JNK lg* *NFkB p65* *cAMP*	*cAMP* *JNK sh* *NFkB p65* *STAT3*	*cAMP* *JNK lg* *p38* *STAT1α*	*JNK lg* *NFkB p65* *STAT1β*	*cAMP* *JNK lg* *p38* *NFkB p65*
LS Sig	*NFkB p65* PKCd PKCmu2	JNK sh MSN *NFkB p65* PKCmu2	*NFkB p65*	Ezr/Rdx *NFkB p65* *STAT3*	*cAMP* PKCmu2	*NFkB p65*	*p38* *NFkB p65* PKCmu2 STAT3
PCR	AKT ERK1 ERK2 *JNK lg* JNK sh p38 *NFkB p65* PKCd PKCmu2 RSK STAT3	cAMP ERK1 ERK2 *JNK lg* JNK sh p38 *NFkB p65* PKCd PKCmu2 RSK STAT1α STAT5	*cAMP* *JNK lg* JNK sh p38 *NFkB p65*	*cAMP* JNK lg *JNK sh* p38 *NFkB p65* PKCmu2 *STAT3*	*cAMP* ERK1 ERK2 *JNK lg* JNK sh *p38* NFkB p65 PKCd PKCmu2 RSK STAT5	ERK1 ERK2 *JNK lg* JNK sh p38 *NFkB p65* PKCd PKCmu2 RSK	*cAMP* ERK1 ERK2 *JNK lg* JNK sh *p38* *NFkB p65* PKCd PKCmu2 RSK STAT3 STAT5
LASSO	ERK2 *JNK lg* JNK sh *p38* *NFkB p65* PKCd	*JNK lg* JNK sh p38 *NFkB p65* *STAT1β* STAT5	*JNK lg* *NFkB p65*	*cAMP* Ezr/Rdx *JNK sh* *NFkB p65* PKCd STAT1*β* *STAT3*	*cAMP* ERK1 *JNK lg* *p38* NFkB p65 STAT5	ERK1 *JNK lg* *NFkB p65* STAT1α	*cAMP* ERK1 *JNK lg* *p38* *NFkB p65* PKCd
LMI	AKT ERK2 Ezr/Rdx GSK3α GSK3*β* *JNK lg* JNK sh MSN *p38* *NFkB p65*	ERK1 ERK2 Ezr/Rdx GSK3*β* *JNK lg* JNK sh MSN p40Phox *NFkB p65* PKCmu2	ERK1 ERK2 Ezr/Rdx JNK sh MSN *NFkB p65* PKCmu2 STAT5	Ezr/Rdx *JNK sh* MSN *NFkB p65* PKCmu2 STAT1α STAT1*β* *STAT3*	*cAMP* ERK1 ERK2 Ezr/Rdx GSK3α GSK3*β* JNK sh MSN *p38* NFkB p65	AKT ERK1 ERK2 Ezr/Rdx *JNK lg* MSN *NFkB p65* PKCd *STAT1β*	*cAMP* ERK1 Ezr/Rdx GSK3α GSK3*β* JNK sh *p38* *NFkB p65* PKCmu2 RSK

	PKCd PKCmu2	RSK SMAD2 *STAT1β* STAT3 STAT5			PKCmu2 RSK *STAT1***α** STAT1*β* STAT3 STAT5		STAT1**α** STAT1*β* STAT3

Table S12. Identified predictors by PCR, LASSO, and LMI versus true predictors for Non-Toll data.

Output	**G-CSF**	**IL-1α**	**IL-6**	**IL-10**	**MIP-1α**	**RANTES**	**TNFα**
LS Sig	p38		STAT1**α** STAT1*β*	STAT1**α** STAT1*β* STAT3	cAMP	STAT1**α**	cAMP p38 STAT3
PCR	p38		cAMP	cAMP AKT ERK1 ERK2 RSK STAT3	cAMP JNK lg p38 STAT1**α** STAT1*β* STAT3	STAT1**α** STAT1*β*	cAMP p38 p40Phox RSK
LASSO	p38	cAMP STAT1**α**	cAMP Rps6 STAT5	cAMP ERK1 ERK2 Ezr/Rdx STAT1*β* STAT3 STAT5	cAMP AKT ERK1 ERK2 GSK3**α** JNK lg JNK sh p38 NFkB p65 STAT3 STAT5	GSK3**α** JNK sh STAT1**α**	cAMP AKT Ezr/Rdx p38 p40Phox PKCmu2 SMAD2 STAT1**α** STAT3

LMI	ERK2 MSN p38 STAT1*β* STAT5	ERK1 ERK2 Ezr/Rdx GSK3*β* MSN STAT1**α** STAT3	ERK1 ERK2 Ezr/Rdx MSN STAT1**α** STAT1*β*	ERK1 ERK2 GSK3α STAT1**α** STAT1*β* STAT3	cAMP ERK1 ERK2 JNK lg NFkB p65 STAT1**α** STAT1*β* STAT3 STAT5	ERK1 ERK2 GSK3**α** p38 STAT1**α** STAT1*β*	cAMP p38 STAT1*β* STAT3

3.7 Overall comparison of the methods

In this section we summarize the analyses and comparison of methods' performance. An investigator can use this summary in selecting an appropriate method, or designing a new method, for network-reconstruction based on the properties of the data.

RMSE for different noise levels: To score the methods according to RMSE, the average RMSE across the four noise levels (5, 10, 20 and 40%) is computed for a chosen method (PCR, LASSO or LMI) and the LS method. Then, their ratio (e.g. average RMSE for LS/ average RMSE for PCR) is taken as the score of the method (i.e. LS gets a score of 1 corresponding to the best/smallest RMSE). With respect to this basis, LMI gets the best score (0.94) and LASSO gets the worst score (0.56). With increasing noise level, accuracy and *G* of LMI and LASSO decreases more rapidly as compared to that for the PCR method (Figure 4(C)).

Comparison with respect to the distribution of estimated coefficients: For a better method, standard deviation and the difference of the mean value from the true value of the coefficient

should be smaller. Hence the score as defined in Table 4 is used. Data from Supplementary Material Table S5 is used to compute the scores. LMI method performs best according to this comparison too. One can say the numerical values of the coefficients obtained from LMI method are least affected due to the presence of noise as compared to other methods.

Comparison on the basis of accuracy and related metrics: While at the highest noise level all the methods have similar accuracy, LMI receives the best score for accuracy (average for 5, 10, 20, and 40 percent noise level) followed by LASSO. This comparison encourages the application of LMI in presence of medium level of noise. LMI performs better well (similar to PCR) for different types of noises as well.

Comparison with respect to fractional error in the coefficients estimates: Fractional error in estimating the coefficients is the average deviation of each coefficient value from the corresponding true value. The corresponding score is such that smaller fractional deviation results in a higher score. PCR receives the highest score for different levels and types of noise followed by LMI.

Comparison with respect to performance on datasets of different sizes and dimensional ratios: The effect of dimension of data has been considered in two different aspects: The size of dataset, and the dimensional ratio of dataset. It is shown that with increase in the size of the dataset the fractional error in estimating the coefficients decreases. This is not surprising because if the structure of the model (e.g. linear mapping) is correctly chosen then, generally, more data provides a better estimate of the model parameters. The method of PCR has the highest average

score for different dimensional ratios and different sizes followed by the method of LMI. However, in terms of detecting the true inputs, LASSO and LMI, in that order, are better than PCR. Overall, this comparison suggests the use of LMI when the size of data is not very large.

Comparison with respect to the amount of missing data: For the case of missing data in which different fractions of dataset are not available for reconstruction of the network, simulations are run on two different datasets: experimental (PP/cytokine) dataset and synthetic dataset with 20% noise. The performance of each method has been evaluated with respect to RMSE on the validation set (20% of the data assumed to be available; 80% of the data assumed to be available is used for training the model). A performance score is calculated as described in Table 4 and assigned to each method. Based on scores assigned to the methods, LMI is assessed as the best method with respect to smallest validation RMSE, and sensitivity to different fractions of missing data, overall.

Table S13..Fractional error in estimating the coefficient versus fraction of collinearity and noise level for different methods.

PCR				
Noise/Collinearity	5%	10%	20%	40%
0 collinearity	0.08	0.10	0.14	0.21
0.1 collinearity	0.17	0.23	0.24	0.33
0.2 collinearity	0.41	0.54	0.53	0.50
0.3 collinearity	0.57	0.77	0.75	0.68
LASSO				
Noise/collinearity	5%	10%	20%	40%
0 collinearity	0.48	0.46	0.45	0.44
0.1 collinearity	0.42	0.34	0.47	1.07
0.2 collinearity	0.45	0.54	1.11	2.32
0.3 collinearity	0.45	0.69	1.41	3.04
LMI				
Noise/collinearity	5%	10%	20%	40%
0 collinearity	0.09	0.12	0.20	0.34
0.1 collinearity	0.17	0.45	0.97	1.77
0.2 collinearity	0.40	1.00	1.84	2.99
0.3 collinearity	0.62	1.22	2.13	4.10

Table S14. Accuracy versus fraction of collinearity and noise level for different methods.

PCR				
Noise/Collinearity	5%	10%	20%	40%
0 collinearity	0.84	0.83	0.81	0.74
0.1 collinearity	0.77	0.78	0.77	0.75
0.2 collinearity	0.68	0.71	0.71	0.67
0.3 collinearity	0.63	0.65	0.67	0.67
LASSO				
Noise/Collinearity	5%	10%	20%	40%
0 collinearity	0.88	0.88	0.87	0.79
0.1 collinearity	0.85	0.83	0.72	0.63
0.2 collinearity	0.76	0.67	0.63	0.57
0.3 collinearity	0.71	0.62	0.60	0.58
LMI				
Noise/collinearity	5%	10%	20%	40%
0 collinearity	0.97	0.97	0.91	0.77
0.1 collinearity	0.93	0.79	0.68	0.61
0.2 collinearity	0.84	0.64	0.59	0.57
0.3 collinearity	0.74	0.62	0.60	0.57

Table S15. Geometric mean of sensitivity and specificity versus fraction of collinearity and noise level for different methods.

PCR				
Noise/Collinearity	5%	10%	20%	40%
0 collinearity	0.82	0.83	0.82	0.75
0.1 collinearity	0.67	0.73	0.76	0.76
0.2 collinearity	0.54	0.62	0.67	0.68
0.3 collinearity	0.41	0.50	0.60	0.66
LASSO				
Noise/collinearity	5%	10%	20%	40%
0 collinearity	0.89	0.89	0.88	0.79
0.1 collinearity	0.85	0.78	0.56	0.31
0.2 collinearity	0.73	0.50	0.37	0.25
0.3 collinearity	0.66	0.45	0.38	0.34
LMI				
Noise/collinearity	5%	10%	20%	40%
0 collinearity	0.97	0.97	0.91	0.72
0.1 collinearity	0.93	0.76	0.61	0.49
0.2 collinearity	0.83	0.58	0.51	0.47
0.3 collinearity	0.71	0.54	0.50	0.49

Table S16. Sensitivity versus fraction of collinearity and noise level for different methods.

PCR				
Noise/Collinearity	5%	10%	20%	40%
0 collinearity	0.88	0.85	0.76	0.62
0.1 collinearity	0.94	0.89	0.81	0.66
0.2 collinearity	0.92	0.89	0.81	0.62
0.3 collinearity	0.93	0.90	0.82	0.68
LASSO				
Noise/collinearity	5%	10%	20%	40%
0 collinearity	0.81	0.82	0.82	0.81
0.1 collinearity	0.84	0.90	0.92	0.95
0.2 collinearity	0.83	0.90	0.92	0.89
0.3 collinearity	0.84	0.88	0.89	0.89
LMI				
Noise/collinearity	5%	10%	20%	40%
0 collinearity	0.94	0.94	0.92	0.88
0.1 collinearity	0.92	0.86	0.82	0.79
0.2 collinearity	0.86	0.77	0.73	0.73
0.3 collinearity	0.82	0.77	0.78	0.73

Table S17. Specificity versus fraction of collinearity and noise level for different methods.

PCR				
Noise/Collinearity	5%	10%	20%	40%
0 collinearity	0.76	0.81	0.88	0.92
0.1 collinearity	0.51	0.63	0.73	0.87
0.2 collinearity	0.34	0.45	0.57	0.75
0.3 collinearity	0.21	0.31	0.46	0.65
LASSO				
Noise/collinearity	5%	10%	20%	40%
0 collinearity	0.98	0.98	0.95	0.77
0.1 collinearity	0.87	0.72	0.43	0.17
0.2 collinearity	0.66	0.33	0.19	0.11
0.3 collinearity	0.54	0.26	0.19	0.15
LMI				
Noise/collinearity	5%	10%	20%	40%
0 collinearity	1.00	1.00	0.90	0.60
0.1 collinearity	0.94	0.69	0.48	0.33
0.2 collinearity	0.80	0.46	0.39	0.33
0.3 collinearity	0.63	0.41	0.35	0.35

Table S18. Average number of true positives, true negatives, false positives, and false negatives versus noise level for different methods (total number of inputs = 25).

Method	5	10	20	40
PCR				
FN	1	1.5	2.6	4.7
FP	8.2	7.2	5.7	3.6
TP	13.6	13.1	12	9.9
TN	2.2	3.2	4.7	6.8
LASSO				
FN	2.4	1.8	1.6	1.6
FP	4.8	7.7	8.5	8.8
TP	12.2	12.8	13	13
TN	5.6	2.7	1.9	1.6
LMI				
FN	2.6	3.3	3.3	4
FP	3.8	6.1	6.7	6.8
TP	12	11.3	11.3	10.6
TN	6.6	4.3	3.7	3.6

Author Contributions

Research design: SS, DMT, MRM. Implementation of algorithms and methods: BA, MRM. Wrote manuscript: BA, MRM. Revision/supervision: SS, DMT, MRM. BA and MRM contributed equally to this work.

Nomenclature

B, b	Matrix or vector, respectively, of model coefficients
C	Covariance matrix
FP, FN	False positive and false negative, respectively
G	Geometrical mean of sensitivity and specificity
RMSE	Root-mean squared error
T	Matrix of latent variables
TP, TN	True positive and true negative, respectively
V	Set of eigen vectors
X	Matrix of input data
Y, y	Matrix or vector, respectively, of output data
e	L_2 or L_∞ norm
m	Number of rows in **X, Y** and **y**

n	Number of columns in **X**
p	Number of columns in **Y**
std	Standard deviation
t	Shrinkage coefficient, t-distribution
tinv	Inverse of cumulative *t*-distribution

Subscripts

frac	Refers to “fractional”
i, j, k	Generic indices
LS	Values derived from the method of least squares
LS_sig	Least squares with significant predictors only
p	Values derived from prediction

Greek letters

Γ, γ	Set of eigen values and an eigen value, respectively
α	Significance level
ε	Residual vector

v	Degree of freedom
σ	Standard deviation of the model parameters

References

1. Dash S, Maurya MR, Venkatasubramanian V, Rengaswamy R. A novel interval-halving framework for automated identification of process trends. Aiche J. 2004 Jan;50(1):149-62.
2. Pradervand S, Maurya MR, Subramaniam S. Identification of signaling components required for the prediction of cytokine release in RAW 264.7 macrophages. Genome Biology. 2006;7(2):R11.
3. Helland IS. Partial Least-Squares Regression and Statistical-Models. Scand J Stat. 1990;17(2):97-114.
4. Gerlach RW, Kowalski BR, Wold HOA. Partial Least-Squares Path Modeling with Latent-Variables. Anal Chim Acta-Comp. 1979;3(4):417-21.
5. Tibshirani R. Regression shrinkage and selection via the Lasso. J Roy Stat Soc B Met. 1996;58(1):267-88.
6. Vandenberghe L, Boyd S, Wu SP. Determinant maximization with linear matrix inequality constraints. Siam J Matrix Anal A. 1998 Apr;19(2):499-533.
7. Julius A, Zavlanos M, Boyd S, Pappas GJ. Genetic network identification using convex programming. IET Systems Biology. 2009 May;3(3):155-66.
8. Cosentino C, Curatola W, Montefusco F, Bansal M, di Bernardo D, Amato F. Linear matrix inequalities approach to reconstruction of biological networks. IET Systems Biology. 2007 May;1(3):164-73.
9. Montefusco F, Cosentino C, Amato F. CORE-Net: exploiting prior knowledge and preferential attachment to infer biological interaction networks. IET Systems Biology. 2010;4(5):296-310.
10. Lofberg J. Dualize it: software for automatic primal and dual conversions of conic programs. Optim Method Softw. 2009;24(3):313-25.
11. The Alliance for Cellular Signaling (AfCS) Available from: http://www.signaling-gateway.org.
12. Chong IG, Jun CH. Performance of some variable selection methods when multicollinearity is present. Chemometr Intell Lab. 2005 Jul 28;78(1-2):103-12.
13. Hastie T, Tibshirani R, Friedman JH. The elements of statistical learning : data mining, inference, and prediction : with 200 full-color illustrations. New York: Springer; 2001.

Figures:

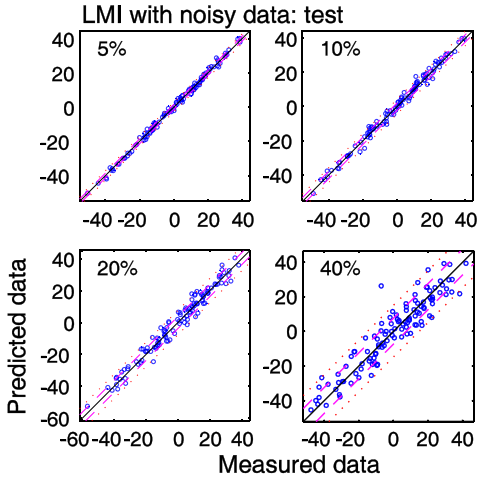

Figure S1. Supplied (noisy) response vs. response predicted with the LMI method for the synthetic noisy data. The dashed and dotted lines represent the σ and 2σ bands, respectively.

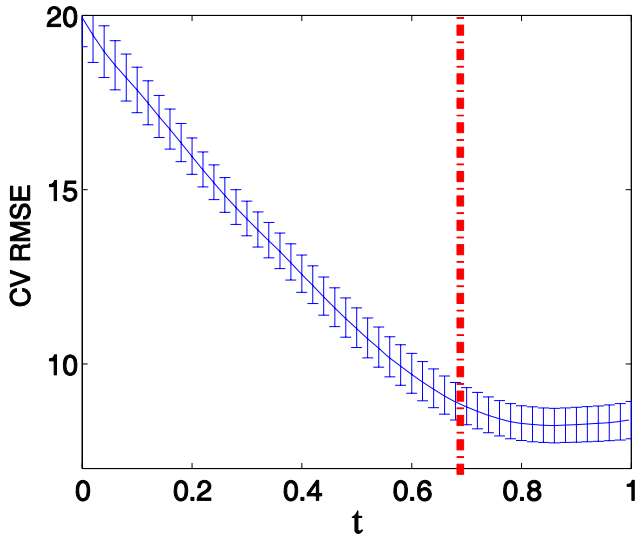

Figure S2. Validation error versus selection threshold *t* for LASSO. Vertical dashed line indicates the optimal tuning parameter (the smallest value of the tuning parameter that is within one standard deviation from the lowest RMSE on curve).

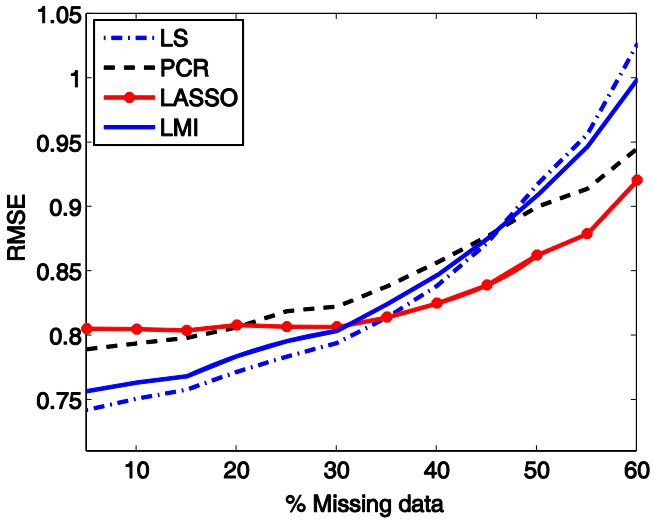

Figure S3. RMSE versus percentage of missing data for different methods on PP/Cytokine data.

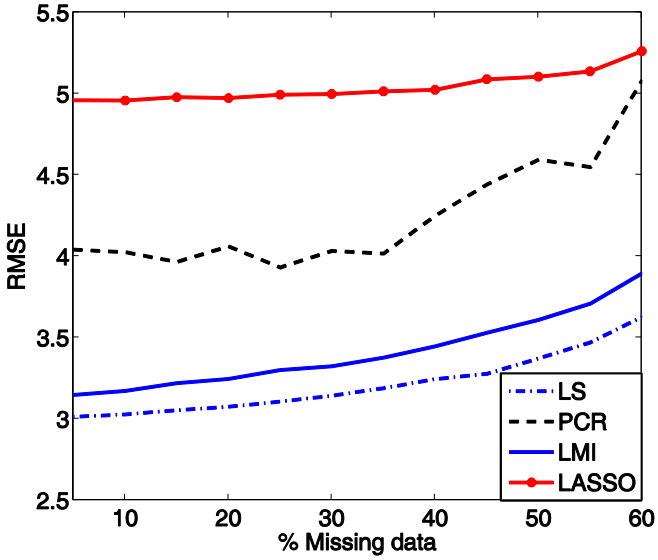

Figure S4. RMSE versus percentage of missing data for different methods on synthetic date with 20% noise level.

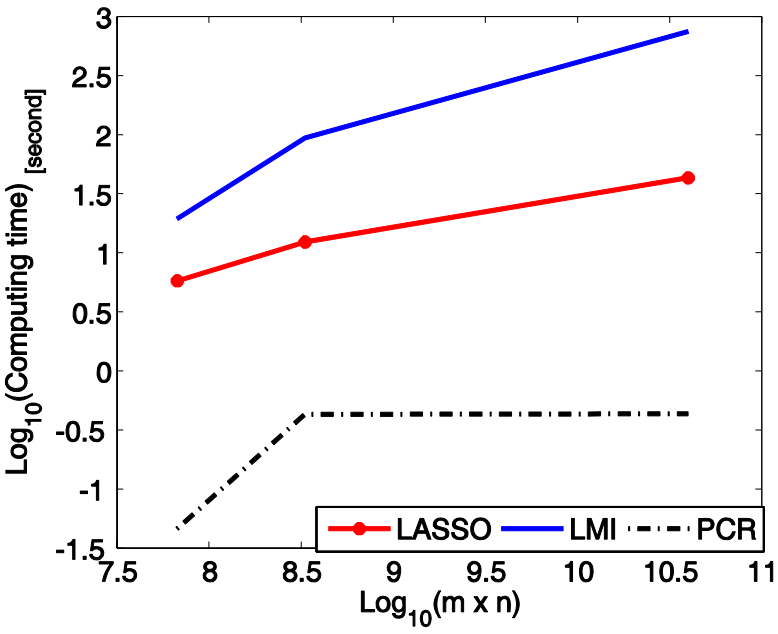

Figure S5. Scaling of computing time with the data-size for different methods.