A probabilistic model to integrate chip and microarray data

Guido Sanguinetti
Department of Computer Science
Regent Court
211 Portobello Road
Sheffield, S1 4DP
U.K.

Magnus Rattray
School of Computer Science
University of Manchester
Oxford Road
Manchester, M13 9PL
U.K.

Neil D. Lawrence
Department of Computer Science
Regent Court
211 Portobello Road
Sheffield, S1 4DP
U.K.

10th January 2006

1 Introduction

The purpose of this note is to introduce and solve in detail a probabilistic model for integrating microarray data with location data (typically obtained via Chromatine Immunoprecipitation, *ChIP*). Experimental validation and biological insights obtained using this model will be published elsewhere (Sanguinetti et al. [2006]).

The problem we are addressing is a key one in bioinformatics. Cellular processes are assumed to be initiated by the transcription of genes into mRNA and its successive translation into proteins. Transcription is regulated by a complex network of biochemical processes, entailing the binding of transcription factor proteins to the promoter regions of the genes. Given the structure of the network

(i.e. which transcription factors bind which genes) and some gene expression measurements, can we deduce quantitative estimates of the strength of the transcription factors-genes interactions? This would include the quantitative effect of a transcription factor in regulating its targets (whether it's a promoter or a repressor), the relative strength with which different transcription factors regulate their targets, and what are the correlations between transcription factors.

Data about the structure of transcriptional networks is becoming increasingly available. The regulatory network of yeast has now been studied in detail using Chromatine Immunoprecipitation (ChIP) by Lee et al. [2002]; some information about the transcriptional regulatory network in higher organisms is also becoming available using motif conservation studies (Xie et al. [2005]).

Studies so far on how to integrate this information with microarray data have focused on modified forms of regression ([Liao et al., 2003, Gao et al., 2004, Alter and Golub, 2004, Boulesteix and Strimmer, 2005], and see the introduction section of Boulesteix and Strimmer [2005] for a review of these methods). While these methods provide a worthwhile global picture, they only infer a generic transcription factor activity (TFA) for each transcription factor, assumed constant across genes. This is often an unrealistic assumption, as it is well known that different genes may respond differently to the same concentration of transcription factor protein (for example, due to the presence of other transcription factors binding the same gene). Also, none of these methods is probabilistic and it is hard to see how to assign credibility intervals to their predictions.

We propose a probabilistic model that models individual gene-specific TFAs for each transcription factor. The resulting explosion in the number of parameters is dealt with by placing a prior distribution on the TFAs (shared by all genes) and marginalising. The model is computationally tractable and can be further modified to propagate known uncertainties in the microarray data (following Sanguinetti et al. [2005]).

2 Model

The logged gene expression measurements are collected in a design matrix $\mathbf{Y} \in \Re^{N \times d}$, where N is the number of genes and d the number of experiments. We assume the rows of the design matrix to be centred, so that each gene expression oscillates about zero. The connectivity measurements are collected in a binary matrix $\mathbf{X} \in \Re^{N \times q}$, where q is the number of transcription factors; element (i, j) of $\mathbf{X}$ is one if transcription factor j binds gene i, zero otherwise.

We assume that the TFAs can be obtained by regressing the gene expressions using the connectivity information, giving the following linear model

$$\mathbf{y}_n = \mathbf{B}_n \mathbf{x}_n + \boldsymbol{\epsilon}_n. \tag{1}$$

Here $n = 1, \ldots, N$ indexes the gene, $\mathbf{y}_n = \mathbf{Y}(n,:)^T$, $\mathbf{x}_n = \mathbf{X}(n,:)^T$ and $\boldsymbol{\epsilon}_n$ is an error term. The matrix $\mathbf{B}_n$ has d rows and q columns, and models the gene-specific TFAs; each column contains the TFA of a certain transcription factor relative to gene n. The crucial difference between our model and other models

previously proposed is that the regression coefficients (the TFAs) are allowed to be different from gene to gene.

Allowing different TFAs for each gene leads to an explosion in the number of model parameters. We can deal with this large parameter space through marginalization by placing a prior distribution on the rows of $\mathbf{B}_n$ (the gene specific TFAs at a certain experimental point, denoted as $\mathbf{b}_{nt}$). We will make two physically plausible assumptions in selecting the prior distribution. Firstly, we assume that the gene specific TFA at time t depends solely on the gene specific TFA at time $t-1$ (mathematically, this means that the sequence $\mathbf{b}_{nt}$ has the *Markov property*). This is a simplifying assumption but should be sufficient to capture the main correlations between time points. Secondly, we assume the prior distribution to be *stationary* in time, i.e. no time point is *a priori* special. Mathematically, this amounts to requiring the distributions obtained by marginalising all but one of the time points to be the same.

There are two obvious limiting cases of prior distributions satisfying these conditions. The first is when all the $\mathbf{b}_{nt}$ are assumed to be identical, so that

$$\begin{aligned} \mathbf{b}_{n1} &\sim \mathcal{N}(\boldsymbol{\mu}, \Sigma), \\ \mathbf{b}_{n(t+1)} &\sim \mathcal{N}(\mathbf{b}_{nt}, 0). \end{aligned} \tag{2}$$

This would be an appropriate model when the experimental data set consists of replicates of a condition. The second limiting case is when all the $\mathbf{b}_{nt}$ are assumed to be independent and identically distributed,

$$\mathbf{b}_{nt} \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma). \tag{3}$$

This is a static model which could be of use when the data set consists of independent samples drawn from conditions without a temporal order.

In general, we expect a realistic model of time-series data to be somewhere in between these two extremes. There are infinite possible choices for such a model; we will make the simplest possible choice of a linear combination of the two models, as this combines computational tractability with simplicity in interpreting the results. We therefore model the gene specific TFAs as

$$\mathbf{b}_{n(t+1)} \sim \mathcal{N}\left(\gamma \mathbf{b}_{nt} + (1-\gamma)\boldsymbol{\mu}, \left(1-\gamma^2\right)\Sigma\right) \tag{4}$$

for $t = 1, \ldots, d-1$ and

$$\mathbf{b}_{n1} \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma).$$

$\gamma \in [0,1]$ is a parameter measuring the global variability of the TFAs: $\gamma = 1$ returns the replicates model (2), while $\gamma = 0$ returns the static model with all TFAs independent (3).

3 Likelihood computations

The joint likelihood is factorized as

$$p\left(y_1,\ldots,y_d,\mathbf{b}_1,\ldots,\mathbf{b}_d|\beta,\tau,\boldsymbol{\mu},\Sigma\right) = \\ \mathcal{N}\left(\mathbf{b}_1|\boldsymbol{\mu},\Sigma\right)\Pi_{t=1}^{d}\mathcal{N}\left(y_t|\mathbf{b_t^T}\mathbf{x},\beta^{-2}\right)\Pi_{t=2}^{d}\mathcal{N}\left(\mathbf{b_t}|\gamma\mathbf{b}_{t-1}+\left(1-\gamma\right)\boldsymbol{\mu},\left(1-\gamma^2\right)\Sigma\right). \tag{5}$$

To compute the marginal likelihood, we first compute the following integral

$$\int\mathcal{N}\left(y|\mathbf{b^T}\mathbf{x},\beta^{-2}\right)\mathcal{N}\left(\mathbf{b}|\boldsymbol{\mu},\Sigma\right)=\mathcal{N}\left(y|\boldsymbol{\mu}^T\mathbf{x},\beta^{-2}+\mathbf{x}^T\Sigma\mathbf{x}\right). \tag{6}$$

We then proceed to integrate eq (5); we notice that $\mathbf{b}_T$ appears only in one factor in equation (**??**), therefore we start off by integrating it first to obtain

$$\begin{aligned}&\int\Pi_{t=1}^{d-2}\mathcal{N}\left(y_t|\mathbf{b_t^T}\mathbf{x},\beta^{-2}\right)\Pi_{t=2}^{d-2}\mathcal{N}\left(\mathbf{b_t}|\gamma\mathbf{b}_{d-1}+\left(1-\gamma\right)\boldsymbol{\mu},\left(1-\gamma^2\right)\Sigma\right)\\&\times\mathcal{N}\left(\mathbf{b}_1|\boldsymbol{\mu},\Sigma\right)\mathcal{N}\left(y_{d-1}|\mathbf{b_{d-1}^T}\mathbf{x},\beta^{-2}\right)\\&\times\mathcal{N}\left(y_d|\gamma\mathbf{b_{d-1}^T}\mathbf{x}+\left(1-\gamma\right)\boldsymbol{\mu}^T\mathbf{x},\beta^{-2}+\left(1-\gamma^2\right)\mathbf{x}^T\Sigma\mathbf{x}\right)\end{aligned} \tag{7}$$

where we have used eq. (6) with $\boldsymbol{\mu}=\gamma\mathbf{b}_{d-1}+\left(1-\gamma\right)\boldsymbol{\mu}$ and $\Sigma=\left(1-\gamma^2\right)\Sigma$. We notice that the second line in eq. (7) can be rearranged using

$$\frac{\left(y_{d-1}-\mathbf{b_{d-1}^T}\mathbf{x}\right)^2}{\beta^{-2}}+\frac{\left(y_d-\gamma\mathbf{b_{d-1}^T}\mathbf{x}-\left(1-\gamma\right)\boldsymbol{\mu}^T\mathbf{x}\right)^2}{\beta^{-2}+\left(1-\gamma^2\right)\mathbf{x}^T\Sigma\mathbf{x}}=\frac{\left(k_{d-1}-\mathbf{b_{d-1}^T}\mathbf{x}\right)^2}{\alpha_{d-1}^{-2}}+\iota_{d-1}$$

where

$$\begin{aligned}\alpha_{d-1}^2&=\beta^2+\gamma^2\left[\beta^{-2}+\left(1-\gamma^2\right)\mathbf{x}^T\Sigma\mathbf{x}\right]^{-1}\\\frac{k_{d-1}}{\alpha_{d-1}^{-2}}&=\frac{y_{d-1}}{\beta^{-2}}+\gamma\frac{y_d-\left(1-\gamma\right)\boldsymbol{\mu}^T\mathbf{x}}{\beta^{-2}+\left(1-\gamma^2\right)\mathbf{x}^T\Sigma\mathbf{x}}\\\iota&=\frac{\left(\frac{y_d-(1-\gamma)\boldsymbol{\mu}^T\mathbf{x}}{\gamma}-y_{d-1}\right)^2}{\beta^{-2}+\frac{\beta^{-2}+(1-\gamma^2)\mathbf{x}^T\Sigma\mathbf{x}}{\gamma^2}}.\end{aligned} \tag{8}$$

In other words, the precision is replaced with the sum of the precisions and the observed variable is replaced with the precision-weighted sum of the observed variables. The factor ι is independent of the $\mathbf{b}$s and can be taken out of the integral.

We can then proceed to marginalise $\mathbf{b}_{d-1}$; equation (8) provides us with a handle to compute the general solution recursively. Therefore we get

$$p\left(y_1,\ldots,y_d|\beta,\tau,\boldsymbol{\mu},\Sigma\right)=\Pi_{t=2}^{d}\mathcal{N}\left(y_{t-1}|k_t,\phi_t^{-1}\right)\times\mathcal{N}\left(k_1|\mathbf{x}^T\boldsymbol{\mu},\phi_1^{-1}\right) \tag{9}$$

where

$$
\begin{aligned}
k_d &= \frac{y_d - (1-\gamma)\,\boldsymbol{\mu}^T\mathbf{x}}{\gamma} \\
\phi_d &= \left(\beta^{-2} + \frac{\beta^{-2} + \left(1-\gamma^2\right)\mathbf{x}^T\Sigma\mathbf{x}}{\gamma^2}\right)^{-1} \\
\phi_{d-1} &= \left(\beta^{-2} + \frac{\alpha_{d-1}^{-2} + \left(1-\gamma^2\right)\mathbf{x}^T\Sigma\mathbf{x}}{\gamma^2}\right)^{-1}
\end{aligned}
$$

α_{d-1} and k_{d-1} were defined in eq (8) and finally

$$
\begin{aligned}
\alpha_{d-2}^2 &= \beta^2 + \gamma^2\left(\alpha_{d-1}^{-2} + \left(1-\gamma^2\right)\mathbf{x}^T\Sigma\mathbf{x}\right)^{-1} \\
\frac{k_{d-2}}{\alpha_{d-2}^{-2}} &= \frac{y_{d-2}}{\beta^{-2}} + \gamma^2\frac{k_{d-1}}{\alpha_{d-1}^{-2} + (1-\gamma^2)\,\mathbf{x}^T\Sigma\mathbf{x}} \\
\phi_1 &= \left(\alpha_1^{-2} + \mathbf{x}^{\mathbf{T}}\Sigma\mathbf{x}\right)^{-1}
\end{aligned}
\tag{10}
$$

(the last update is different because of the different prior).

Having computed the marginal likelihood we now need the gradients in order to optimise it. Again, we can exploit equation (10) to obtain recursive formulae for the gradient with respect to the various parameters. Defining $\alpha_d = \beta$,

$$
\begin{aligned}
&\frac{\partial\alpha_d^2}{\partial\beta} = 2\beta \\
&\frac{\partial\alpha_d}{\partial\gamma} = \frac{\partial\alpha_d}{\partial\Sigma} = \frac{\partial\alpha_t}{\partial\boldsymbol{\mu}} = 0 \\
&\frac{\partial\alpha_t^2}{\partial\beta} = 2\beta - \gamma^2\alpha_{t+1}^{-4}\left(\alpha_{t+1}^{-2} + \left(1-\gamma^2\right)\mathbf{x}^T\Sigma\mathbf{x}\right)^{-2}\frac{\partial\alpha_{t+1}^2}{\partial\beta} \\
&\frac{\partial\alpha_t^2}{\partial\gamma} = \gamma^2\left(\alpha_{t+1}^{-2} + \left(1-\gamma^2\right)\mathbf{x}^T\Sigma\mathbf{x}\right)^{-2}\left(2\gamma\mathbf{x}^T\Sigma\mathbf{x} - \alpha_{t+1}^{-4}\frac{\partial\alpha_t^2}{\partial\gamma}\right) + \\
&2\gamma\left(\alpha_{t+1}^{-2} + \left(1-\gamma^2\right)\mathbf{x}^T\Sigma\mathbf{x}\right)^{-1} \\
&\frac{\partial\alpha_t^2}{\partial\Sigma} = -\gamma^2\left(1-\gamma^2\right)\left(\alpha_{t+1}^{-2} + \left(1-\gamma^2\right)\mathbf{x}^T\Sigma\mathbf{x}\right)^{-2}\mathbf{x}\mathbf{x}^T - \gamma^2\alpha_{t+1}^{-4}\left(\alpha_{t+1}^{-2} + \left(1-\gamma^2\right)\mathbf{x}^T\Sigma\mathbf{x}\right)^{-2}\frac{\partial\alpha_{t+1}^2}{\partial\Sigma} \\
&\frac{\partial\phi_{t-1}^{-1}}{\partial\beta} = 2\beta + \frac{1}{\gamma^2}\frac{\partial\alpha_{t-1}^{-2}}{\partial\beta} \\
&\frac{\partial\phi_{t-1}^{-1}}{\partial\gamma} = -2\left[\frac{\alpha_{d-1}^{-2} + \left(1-\gamma^2\right)\mathbf{x}^T\Sigma\mathbf{x}}{\gamma^3} + \frac{\mathbf{x}^T\Sigma\mathbf{x}}{\gamma}\right] \\
&\frac{\partial\phi_{t-1}^{-1}}{\partial\Sigma} = \frac{\left(1-\gamma^2\right)}{\gamma^2}\mathbf{x}\mathbf{x}^T
\end{aligned}
$$

$$\frac{\partial k_d}{\partial \beta} = \frac{\partial k_d}{\partial \Sigma} = 0$$

$$\frac{\partial k_d}{\partial \gamma} = -\frac{y_d - (1-\gamma)\,\boldsymbol{\mu}^T\mathbf{x}}{\gamma^2} + \frac{\boldsymbol{\mu}^T\mathbf{x}}{\gamma}$$

$$\frac{\partial k_d}{\partial \boldsymbol{\mu}} = -\frac{(1-\gamma)}{\gamma}\mathbf{x}$$

$$\frac{\partial k_{t-1}}{\partial \beta} = -\frac{k_{t-1}}{\alpha_{t-1}^2}\frac{\partial \alpha_{t-1}^2}{\partial \beta} + \frac{\gamma^2\alpha_{t-1}^{-2}k_t}{\alpha_t^4\left[\alpha_t^{-2} + (1-\gamma^2)\,\mathbf{x}^T\Sigma\mathbf{x}\right]^2}\frac{\partial \alpha_t^2}{\partial \beta} +$$
$$2\beta\alpha_{t-1}^{-2}y_{t-1} + \frac{\gamma^2\alpha_{t-1}^{-2}}{\left[\alpha_t^{-2} + (1-\gamma^2)\,\mathbf{x}^T\Sigma\mathbf{x}\right]}\frac{\partial k_t}{\partial \beta}$$

$$\frac{\partial k_{t-1}}{\partial \gamma} = -\frac{k_{t-1}}{\alpha_{t-1}^2}\frac{\partial \alpha_{t-1}^2}{\partial \gamma} + \frac{\gamma^2\alpha_{t-1}^{-2}k_t}{\alpha_t^4\left[\alpha_t^{-2} + (1-\gamma^2)\,\mathbf{x}^T\Sigma\mathbf{x}\right]^2}\frac{\partial \alpha_t^2}{\partial \gamma} +$$
$$\frac{\gamma^2\alpha_{t-1}^{-2}}{\left[\alpha_t^{-2} + (1-\gamma^2)\,\mathbf{x}^T\Sigma\mathbf{x}\right]}\frac{\partial k_t}{\partial \gamma} + \frac{2\gamma^3\mathbf{x}^T\Sigma\mathbf{x}\alpha_{t-1}^{-2}k_t}{\left[\alpha_t^{-2} + (1-\gamma^2)\,\mathbf{x}^T\Sigma\mathbf{x}\right]^2} +$$
$$\frac{2\gamma\alpha_{t-1}^{-2}k_t}{\left[\alpha_t^{-2} + (1-\gamma^2)\,\mathbf{x}^T\Sigma\mathbf{x}\right]}$$

$$\frac{\partial k_{t-1}}{\partial \boldsymbol{\mu}} = \frac{\gamma^2\alpha_{t-1}^{-2}}{\left[\alpha_t^{-2} + (1-\gamma^2)\,\mathbf{x}^T\Sigma\mathbf{x}\right]}\frac{\partial k_t}{\partial \boldsymbol{\mu}}$$

$$\frac{\partial k_{t-1}}{\partial \Sigma} = -\frac{k_{t-1}}{\alpha_{t-1}^2}\frac{\partial \alpha_{t-1}^2}{\partial \Sigma} + \frac{\gamma^2\alpha_{t-1}^{-2}k_t}{\alpha_t^4\left[\alpha_t^{-2} + (1-\gamma^2)\,\mathbf{x}^T\Sigma\mathbf{x}\right]^2}\frac{\partial \alpha_t^2}{\partial \Sigma} +$$
$$\frac{\gamma^2\alpha_{t-1}^{-2}}{\left[\alpha_t^{-2} + (1-\gamma^2)\,\mathbf{x}^T\Sigma\mathbf{x}\right]}\frac{\partial k_t}{\partial \Sigma} + \frac{\gamma^2\left(1-\gamma^2\right)\alpha_{t-1}^{-2}k_t\mathbf{x}\mathbf{x}^T}{\left[\alpha_t^{-2} + (1-\gamma^2)\,\mathbf{x}^T\Sigma\mathbf{x}\right]^2}.$$

4 Posterior estimation

Once the model parameters have been optimised, gene-specific TFAs can now be estimated from the posterior distribution over the $\mathbf{b}_n$s. This is obtained by applying Bayes' rule and has the form

$$p\left(\mathbf{b}_{n1}, \ldots, \mathbf{b}_{nd} | \sigma, \gamma, \boldsymbol{\mu}, \Sigma, \mathbf{X}, \mathbf{Y}\right) = \mathcal{N}\left(\bar{\mathbf{b}}_n, \Sigma_{\mathbf{b}_n}\right) \tag{11}$$

where the posterior covariance is given by

$$\Sigma_{\mathbf{b}_n} = \begin{pmatrix} A_1 & B & 0 & 0 \\ B & A & \ldots & 0 \\ 0 & B & \ldots & B \\ 0 & 0 & \ldots & A_d \end{pmatrix}^{-1} \tag{12}$$

where

$$A_1 = A_d = \sigma^{-2}\mathbf{x}_n\mathbf{x}_n^T + \left(1-\gamma^2\right)^{-1}\Sigma^{-1}$$
$$A = \sigma^{-2}\mathbf{x}_n\mathbf{x}_n^T + \left(1+\gamma^2\right)\left(1-\gamma^2\right)^{-1}\Sigma^{-1}$$
$$B = -\gamma\left(1-\gamma^2\right)^{-1}\Sigma^{-1},$$

and the posterior mean is given by

$$\bar{\mathbf{b}}_n = \Sigma_{\mathbf{b}_n}\begin{bmatrix} \sigma^{-2}y_1\mathbf{x} + \frac{1}{1+\gamma}\Sigma^{-1}\boldsymbol{\mu} \\ \sigma^{-2}y_2\mathbf{x} + \frac{1-\gamma}{1+\gamma}\Sigma^{-1}\boldsymbol{\mu} \\ \vdots \\ \sigma^{-2}y_d\mathbf{x} + \frac{1}{1+\gamma}\Sigma^{-1}\boldsymbol{\mu} \end{bmatrix}.$$

Notice that the posterior mean is a dq dimensional vector and the posterior covariance a $dq \times dq$ matrix. These numbers for a genome-wide study are quite large (in the thousands) and inverting the matrix in equation (12) in a careless way can lead to severe computational costs. We can speed it up considerably by exploiting the special structure of the matrix (12) using a banded LU decomposition (see e.g. Golub and van Loan [1996] sect 4.5).

Then the LU decomposition of C yields

$$C = FG = \begin{pmatrix} I & 0 & 0 & 0 \\ L_1 & I & 0 & 0 \\ 0 & \cdots & \cdots & 0 \\ 0 & 0 & L_{d-1} & I \end{pmatrix}\begin{pmatrix} U_1 & B & 0 & 0 \\ 0 & U_2 & B & 0 \\ 0 & 0 & \cdots & B \\ 0 & 0 & 0 & U_d \end{pmatrix}$$

with the Ls and Us defined recursively as

$$U_1 = D_1$$
$$L_{i-1} = -\gamma\left(1-\gamma^2\right)^{-1}\Sigma^{-1}U_{i-1}^{-1}$$
$$U_i = D_i + L_{i-1}\gamma\left(1-\gamma^2\right)^{-1}\Sigma^{-1}$$

where D_i are the diagonal blocks in (12). The computation of the inverses of the U_is is particularly simple using the Sherman-Morrison formula (Golub and van Loan [1996] sect. 2.1.3)

$$\left(\kappa\Sigma^{-1} + \alpha\mathbf{x}\mathbf{x}^T\right)^{-1} = \kappa^{-1}\Sigma - \alpha\frac{\mathbf{y}\mathbf{y}^T}{1+\lambda}$$

where $\mathbf{y} = \kappa^{-1}\Sigma\mathbf{x}$ and $\lambda = \alpha\mathbf{x}^T\left(\kappa^{-1}\Sigma\right)\mathbf{x}$. Notice that L_i contains terms involving the identity matrix and $\mathbf{x}\mathbf{y}^T$, while U_i always contains terms involving only Σ^{-1} and $\mathbf{x}\mathbf{x}^T$. The inverse of C is easily obtained by noticing the following property of banded triangular matrices

$$\begin{pmatrix} I & 0 & 0 & 0 \\ L_1 & I & 0 & 0 \\ 0 & L_2 & I & 0 \\ 0 & 0 & L_3 & I \end{pmatrix}^{-1} = \begin{pmatrix} I & 0 & 0 & 0 \\ -L_1 & I & 0 & 0 \\ L_2L_1 & -L_2 & I & 0 \\ -L_3L_2L_1 & L_3L_2 & -L_3 & I \end{pmatrix}$$

and similarly for upper triangular matrices

$$\begin{pmatrix} U_1 & -F & 0 & 0 \\ 0 & U_2 & -F & 0 \\ 0 & 0 & U_3 & -F \\ 0 & 0 & 0 & U_4 \end{pmatrix}^{-1} =$$
$$\begin{pmatrix} U_1^{-1} & U_1^{-1}FU_2^{-1} & U_1^{-1}FU_2^{-1}FU_3^{-1} & U_1^{-1}FU_2^{-1}FU_3^{-1}FU_4^{-1} \\ 0 & U_2^{-1} & U_2^{-1}FU_3^{-1} & U_2^{-1}FU_3^{-1}FU_4^{-1} \\ 0 & 0 & U_3^{-1} & U_3^{-1}FU_4^{-1} \\ 0 & 0 & 0 & U_4^{-1} \end{pmatrix}$$

We have described the 4×4 case; generalisation to the $d \times d$ case is trivial.

References

O. Alter and G. H. Golub. Integrative analysis of genome-scale data using pseudoinverse projection predicts novel correlation between dna replication and rna transcription. *Proceedings of the National Academy of Sciences USA*, 101(47):16577–16582, 2004.

A.-L. Boulesteix and K. Strimmer. Predicting transcription factor activities from combined analysis of microarray and ChIP data: a partial least squares approach. *Theor. Biol. Med.Model.*, 2(23):1471–16582, 2005.

F. Gao, B. C. Foat, and H. J. Bussemaker. Defining transcriptional networks through integrative modeling of mRNA expression and transcription factor binding data. *BMC Bioinformatics*, 5(31):1471–16582, 2004.

G. H. Golub and C. F. van Loan. *Matrix Computations*. Johns Hopkins, Baltimore, 1996.

T. I. Lee, N. J. Rinaldi, F. Robert, D. T. Odom, Z. Bar-Joseph, G. K. Gerber, N. M. Hannett, C. T. Harbison, C. M. Thompson, I. Simon, J. Zeitlinger, E. G. Jennings, H. L. Murray, D. B. Gordon, B. Ren, J. J. Wyrick, J.-B. Tagne, T. L. Volkert, E. Fraenkel, D. K. Gifford, and R. A. Young. Transcriptional regulatory networks in Saccharomyces cerevisiae. *Science*, 298(5594):799–804, 2002.

J. C. Liao, R. Boscolo, Y.-L. Yang, L. M. Tran, C. Sabatti, and V. P. Roychowdhury. Network component analysis: Reconstruction of regulatory signals in biological systems. *Proceedings of the National Academy of Sciences USA*, 100(26):15522–15527, 2003.

G. Sanguinetti, M. Milo, M. Rattray, and N. D. Lawrence. Accounting for probe-level noise in principal component analysis of microarray data. *Bioinformatics*, 21(19):3748–3754, 2005.

G. Sanguinetti, M. Rattray, and N. D. Lawrence. A probabilistic dynamical model for quantitative inference of the regulatory mechanism of transcription. Submitted to *Bioinformatics*, 2006.

X. Xie, J. Lu, E. J. Kulbokas, T. R. Golub, V. Mootha, K. Lindblad-Toh, E. S. Lander, and M. Kellis. Systematic discovery of regulatory motifs in human promoters and 3' UTRs by comparison of several mammals. *Nature*, 434: 338–345, 2005.